



HAL
open science

Bayesian deep learning for weak lensing analyses : overlapping galaxies separation and galaxy parameters estimation from blended scenes

Bastien Arcelin

► **To cite this version:**

Bastien Arcelin. Bayesian deep learning for weak lensing analyses : overlapping galaxies separation and galaxy parameters estimation from blended scenes. *Cosmology and Extra-Galactic Astrophysics* [astro-ph.CO]. Université Paris Cité, 2021. English. NNT : 2021UNIP7033 . tel-03553937

HAL Id: tel-03553937

<https://theses.hal.science/tel-03553937v1>

Submitted on 3 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian deep learning for weak lensing analyses

Overlapping galaxies separation and galaxy parameters estimation from blended scenes

Par Bastien ARCELIN

Thèse de doctorat de Physique de l'Univers

Dirigée par Eric ABOURG

*Présentée et soutenue publiquement le 20 Septembre 2021
devant un jury composé de :*

<i>Rapportrice</i>	Shirley HO	Professor,	Flatiron Institute
<i>Rapporteur</i>	Joe ZUNTZ	Reader,	Royal Observatory Edinburgh
<i>Examinatrice</i>	Stéphanie ESCOFFIER	DR-CNRS	Aix-Marseille Université
<i>Examineur</i>	Alan HEAVENS	Professor	Imperial College London
<i>Présidente du jury</i>	Simona MEI	PU	Université de Paris
<i>Examineur</i>	Junpeng LAO	PhD	Google Research
<i>Co-encadrante</i>	Cécile ROUCELLE	MCU	Université de Paris
<i>Directeur de thèse</i>	Eric ABOURG	DR-CEA	Université de Paris

Résumé

Le lentillage gravitationnel faible est l'une des sondes cosmologiques les plus prometteuses pour contraindre les paramètres de l'énergie noire. Il correspond à la modification de l'image d'une source, induite par la courbure de l'espace-temps, et donc du passage des rayons lumineux, générée par la présence de masse le long de la ligne de visée. Cet effet est cependant très faible et ne peut être détecté qu'en mesurant la corrélation des formes d'un ensemble de sources en arrière-plan. La valeur de cette corrélation donne la mesure du cisaillement gravitationnel.

Plusieurs relevés dédiés à son étude vont bientôt démarrer, tels que le Vera Rubin Legacy Survey of Space and Time (LSST), et Euclid. Ces relevés ont été construits avec des exigences particulières, notamment sur le traitement des erreurs systématiques, afin d'atteindre une précision extrême sur les paramètres de l'équation d'état de l'énergie noire.

Cette thèse s'est déroulée dans le contexte de la Dark Energy Science Collaboration (DESC), au sein de l'expérience LSST. Elle se focalise sur la problématique de la superposition de sources. Cet effet génère une incertitude systématique qui est dominante dans la mesure du cisaillement gravitationnel. Cette systématique impacte à la fois les mesures de forme et celles du décalage vers le rouge des galaxies, deux mesures nécessaires à l'étude du lentillage gravitationnel faible.

Nous proposons deux solutions basées sur les méthodes bayésiennes d'apprentissage profond. La première est un algorithme de séparation de galaxies qui utilise un réseau de neurones génératif appelé autoencodeur variationnel. Ce réseau de neurones permet l'apprentissage d'un prior pour la génération d'images de galaxies isolées. Celui-ci est utilisé dans un second réseau qui effectue la séparation de la galaxie centrale sur des images de galaxies simulées. Nous montrons que l'analyse jointe des pixels des données LSST et Euclid permet une diminution de l'erreur médiane de reconstruction des formes des galaxies de 8 à 47%. La superposition de sources étant étroitement liée à l'étape de détection, nous démontrons la robustesse de notre méthode à de petits décentrages. Nous avons aussi testé notre méthode sur des images de vraies galaxies artificiellement superposées, montrant l'intérêt du transfert d'apprentissage à partir d'un réseau de neurones entraîné sur des images de galaxies simulées. Un processus itératif est ensuite mis en place afin de pouvoir séparer toutes les galaxies d'une image en passant par des étapes de détection, classification, et séparation des sources. Enfin, cet algorithme de séparation de galaxies est testé sur des images extraites de la simulation DC2, générée au sein de la collaboration DESC dans le but de préparer à l'analyse des futures images prises par LSST. Nos résultats montrent une amélioration de 70 à 120% sur l'erreur médiane de reconstruction des formes de galaxies comparée à la méthode générique utilisée dans la pipeline LSST actuellement.

Ensuite, inspirés de cette première méthode, nous proposons un réseau de neurones permettant de faire directement l'estimation des paramètres de forme et de décalage vers le rouge d'une galaxie à partir des images DC2, sans appliquer de méthode de séparation de sources. Nous montrons que ce réseau permet des mesures précises de ces paramètres, même lorsque les sources sont superposées, et comparons ces résultats à ceux obtenus avec séparation de source. Par la suite, nous présentons une première application d'un réseau de neurones bayésien à l'estimation de formes des galaxies.

Ce type de réseau permet l'estimation de l'incertitude épistémique qui représente l'incertitude liée à l'échantillon d'entraînement. Celle-ci peut être vue comme un niveau de confiance dans la mesure réalisée par le réseau, une information cruciale à l'utilisation des réseaux de neurones en science et en cosmologie. Elle permet de diminuer l'importance d'une mesure qui a de fortes probabilités d'être erronée dans l'analyse, voire de la rejeter.

Mots-clés : cosmologie - énergie noire - relevés de galaxies - lentillage faible - apprentissage profond - apprentissage profond bayésien - traitement d'images - traitement de données astronomiques - LSST - synergie LSST-Euclid

Abstract

Weak gravitational lensing is one of the most promising probes to constrain dark energy parameters. It corresponds to the distortion of a source image, induced by the bending of space-time, thus of the light path, generated by the presence of mass along the line of sight. This effect is small and can only be detected measuring the correlation of the shapes of a group of background galaxies. The value of this correlation yields the value of the cosmic shear.

Several surveys dedicated to its study are going to start soon, such as the Vera Rubin Legacy Surveys of Space and Time (LSST), and Euclid. These surveys have been built with particular requirements, especially about the treatment of systematic errors, in order to reach an extreme precision on the dark energy equation of state parameters.

This thesis happened in the context of the Dark Energy Science Collaboration (DESC), inside the LSST experiment. It focuses on the issue of blending, the overlap of astronomical sources. The blending of galaxies introduces a dominant systematic uncertainty on the cosmic shear measurement. This systematic effect impacts the shape and the redshift measurements, both necessary for weak gravitational lensing analysis.

We propose two avenues based on Bayesian deep learning methods. The first one is a deblending algorithm which uses a deep generative network called variational autoencoder. This neural network allows to learn a prior for the generation of isolated galaxy images. The latter is used in a second network to perform the deblending of the centred galaxy on images of simulated galaxies. We show that the pixel joint analysis of LSST and Euclid data decreases the median error on galaxy shape reconstruction from 8 to 47%. Blending of source being closely linked to detection, we demonstrate that our method is robust to small decentring. Also we test our method on images of real galaxies artificially blended showing the interest of transfer learning from a neural network trained on simulated galaxy images. An iterative process is then designed in order to separate all the galaxies in an image going through detection, classification and deblending of sources. Finally this deblending algorithm is tested on images extracted from the DC2 simulation, generated within the DESC to prepare for the analysis of futures images taken by LSST. Our results show an improvement of 70 to 120% on the median ellipticity error compared to the generic method used in the current LSST pipeline.

Then, inspired by this first method, we propose a neural network allowing for the direct estimation of galaxy shape and redshift parameters from DC2 images, without going through the deblending. We show that this neural network allows for a precise measurement of these parameters, even when sources are blended, and we compare these results to the ones obtained with deblending. Then we present the first application of a Bayesian neural network to galaxy shapes estimation. This kind of network provides the estimation of the epistemic uncertainty which characterizes the uncertainty coming from the training sample. It can be viewed as a confidence level in the measurement performed by the network, a primordial information for the application of neural networks in science and in cosmology. It can be used to decrease the importance (or reject) of an incorrect measurement in the analysis.

Keywords: cosmology - dark energy - galaxy surveys - weak lensing - deep learning - Bayesian deep learning - image processing - astronomical data processing - LSST - LSST-Euclid synergy

A Norah.

Remerciements

Je voudrais tout d'abord remercier chaleureusement mes deux directeurs de thèse, Eric et Cécile, pour m'avoir fait confiance pendant ces trois années. Ils m'ont toujours laissé une grande autonomie et liberté dans mes choix, tout en étant présents quand nécessaire, ce qui m'a permis d'apprendre beaucoup à leur côtés. Nos discussions, scientifiques ou non, ont été enrichissantes et agréables, et j'ai toujours pris plaisir à venir les voir, pour partager un café ou un whisky. Merci à eux pour leur gentillesse et bienveillance, j'ai beaucoup apprécié ces années à leur côtés.

I would also like to kindly thank all the members of my defense's jury, starting with the two reviewers Shirley Ho and Joe Zuntz. I thank them for taking the time to read and report this manuscript, and for their insightful inputs and questions during the defense. I also thank Stephanie Escoffier, Alan Heavens and Junpeng Lao for reading my manuscript and asking questions leading to very interesting discussion. Finally I would like to thank Simona Mei for presiding this jury.

Le remerciement suivant va à Cyrille Doux, mon troisième "directeur de thèse" officieux. Grâce à lui j'ai énormément appris en cosmologie, analyse de données et programmation, entre autres. Merci à lui de m'avoir accompagné et guidé notamment au début de cette thèse. Les moments et discussions partagés à Philadelphie ou à Paris resteront pour moi de très bons souvenirs, qu'ils soient de travail intense ou beaucoup plus détendus autour d'une bière.

Je tiens également à remercier le groupe LSST de l'APC, et plus particulièrement Françoise Virieux et Alexandre Boucaud avec qui les discussions et moments passés ont toujours été agréables et enrichissants. J'ai particulièrement appris aux côtés d'Alexandre durant cette thèse, entre autres de son expertise en deep learning. Merci à lui pour le temps consacré à nos discussions, aux différentes relectures, et pour son aide précieuse à l'organisation du workshop BDL 2020 notamment. Cela devient redondant, mais les moments partagés avec cette équipe à la zythothèque, pour prendre un exemple au hasard, resteront d'excellents souvenirs.

Je n'oublie évidemment pas les thésards de l'APC et notamment ceux du bureau 427-B (Calum, Louise, Thomas, Guillaume, Raphaël, ...). L'ambiance y a toujours été très agréable et grâce à eux je suis toujours venu au bureau (quand c'était possible...) avec le sourire.

Plus généralement, merci à tous les ingénieurs, chercheurs et enseignant-chercheurs de l'APC, et plus particulièrement à ceux du groupe cosmologie, que j'ai pu croiser et avec qui j'ai pu échanger pendant ces trois années.

Un remerciement spécial va à Antoine Kouchner, David Saclier et Lydie Giot, qui ont tous les trois joué un rôle important dans ma décision de réaliser cette thèse. Par leur bienveillance, leurs conseils et leurs encouragements, ils ont été d'une aide précieuse dans ma démarche.

Enfin, merci à mes amis et à ma famille, de m'avoir toujours soutenu, suivi et encouragé pendant ces trois années. Merci à mes parents de m'avoir transmis leur curiosité et ouverture d'esprit, de m'avoir toujours encouragé et soutenu dans mes choix, cette thèse n'aurait pas été possible sans leur soutien. Évidemment, un merci particulier à Manon. Sa gentillesse, son sourire et son humour ont été des soutiens indispensables pendant ces trois années, notamment durant la rédaction de cette thèse, et sont un bonheur au quotidien. J'ai hâte que l'on commence une nouvelle étape de

notre vie à Rennes avec notre petite Norah.

Je voudrais terminer en adressant un remerciement à toutes celles et ceux qui ont pris du temps pour relire une ou plusieurs parties de ce manuscrit. Merci à Claire, Justine, Alexandre, Axel, Cyrille, Eric et Cécile pour le temps consacré.

Résumé substantiel du manuscrit

Lentillage gravitationnel faible et superposition de sources pour le Legacy Survey of Space and Time (LSST) de l'observatoire Vera Rubin

Modèle standard de la cosmologie et énergie noire

La cosmologie moderne s'appuie sur le modèle standard, Λ CDM, qui peut être décrit par six paramètres et considère une équation d'état de l'énergie noire avec un paramètre w constant : $p = w\rho$, avec $w = -1$. Il est aujourd'hui corroboré par les différents résultats des expériences et relevés cosmologiques passés, notamment par ceux, très précis, de l'expérience Planck.

L'énergie noire représente la majorité ($\approx 70\%$) de l'énergie totale de notre Univers, mais sa nature est à l'heure actuelle encore inconnue. Ainsi, son étude est l'un des grands défis de la cosmologie moderne. Pour tester une éventuelle déviation au modèle standard, il est possible de reparamétriser l'équation d'état de l'énergie noire. Le paramètre w est alors défini comme $w = w_0 + w_a \frac{z}{1+z}$, z étant le décalage vers le rouge, ce qui permet d'appliquer des contraintes observationnelles sur les paramètres w_0 et w_a .

Les progrès effectués sur la précision des contraintes appliquées aux paramètres de l'énergie noire peuvent être quantifiés via une figure de mérite définie comme : $\text{FoM} = \frac{1}{\sigma(w_0) \times \sigma(w_a)}$. Plus la valeur de cette FoM est élevée, plus les incertitudes sur les paramètres de l'énergie noire sont faibles, et donc plus les contraintes sont élevées. L'objectif des futures expériences, dont le Vera Rubin Legacy Survey of Space and Time (LSST), est de multiplier la FoM par un facteur 10 si on la compare aux expériences aujourd'hui terminées.

Différentes sondes cosmologiques sont utilisées dans ce but (BAOs, Clusters, Supernovae par exemple). Cette thèse se concentre sur l'une des plus prometteuses: le lentillage gravitationnel faible. Pour LSST par exemple, la FoM avec le lentillage gravitationnel faible seul devrait permettre d'atteindre, au minimum, une valeur de 52. Cette valeur représente une augmentation d'un facteur deux par rapport aux résultats obtenus avec les expériences en cours d'acquisition de données.

Vera Rubin Legacy Survey of Space and Time

Le Simonyi Survey Telescope qui sera utilisé dans le cadre du relevé LSST est actuellement en construction sur le site du Vera C. Rubin Observatory à Cerro Pachón au Chili. L'observatoire Vera Rubin devrait commencer à prendre des données en 2023. Il utilisera un système à trois miroirs, avec un miroir primaire de 8.4m qui permettra de couvrir un large champ de ciel de 9.6 deg^2 . Il comportera, de plus, une caméra de 3.2 gigapixels permettant d'observer le ciel dans six bandes de fréquences différentes. Grâce à ces caractéristiques, il pourra observer la même partie du ciel environ 100 fois par an atteignant ainsi des magnitudes très élevées, jusqu'à 27.5 dans la bande r . L'observation d'une grande surface du ciel à ces magnitudes élevées permettra d'augmenter

drastiquement le nombre de galaxies utilisables pour l’analyse du lentillage gravitationnel faible et devrait donc permettre d’améliorer la précision des contraintes sur les paramètres de l’énergie noire.

Le lentillage gravitationnel faible

Le lentillage gravitationnel correspond à la modification de l’image d’une source induite par la courbure locale de l’espace-temps, et donc du trajet des rayons lumineux, dûe à la présence de masse le long de la ligne de visée.

Il est possible de définir deux grandeurs qui caractérisent le potentiel de lentillage le long de la ligne de visée : la convergence, κ , et le cisaillement gravitationnel, γ . Ces deux paramètres permettent de décrire la transformation de l’image de la source par l’effet de lentillage gravitationnel dans l’hypothèse de sources petites devant l’échelle angulaire sur laquelle la distribution de masse évolue.

Dans son régime faible, cet effet ne peut cependant pas être détecté sur une seule source. Il est alors nécessaire d’estimer la corrélation des formes d’un ensemble de galaxies d’arrière-plan. La fonction de corrélation à deux points du cisaillement peut en effet être estimée à partir des ellipticités (formes) des galaxies observées et il est ainsi possible de calculer son spectre de puissance. Des contraintes peuvent alors être appliquées sur les paramètres cosmologiques, notamment ceux de l’énergie noire, en utilisant entre autres l’égalité entre le spectre de puissance du cisaillement et celui de la convergence.

Comme mentionné précédemment, le lentillage gravitationnel faible constitue aujourd’hui l’une des sondes les plus prometteuses pour contraindre les paramètres de l’énergie noire et plusieurs relevés dédiés à son étude, dont LSST, vont bientôt débuter. Ils sont spécifiquement conçus pour optimiser les contraintes obtenues avec cet effet : ils ont été construits avec des exigences particulières, notamment sur le traitement des erreurs systématiques, afin d’atteindre une précision extrême sur les paramètres de l’équation d’état de l’énergie noire.

Le problème de la superposition de sources astrophysiques

De nombreux biais peuvent intervenir dans la mesure du cisaillement gravitationnel faible, notamment lors du traitement des images permettant de mesurer les formes des galaxies. En effet, différents éléments tels que la détection de sources, la fonction d’étalement du point (*Point Spread Function* ou PSF) ou la méthode de mesure peuvent influencer et biaiser la mesure finale du cisaillement. Ces effets systématiques doivent être corrigés ou calibrés afin de diminuer leur impact sur l’analyse, ce qui peut souvent être fait en utilisant des simulations ou des algorithmes dédiés (par exemple METACALIBRATION).

Cette thèse se focalise sur une autre incertitude systématique qui est considérée comme dominante dans la mesure du cisaillement gravitationnel pour LSST : la superposition de sources. Cet effet impacte à la fois les mesures de forme et celles du décalage vers le rouge des galaxies, deux mesures nécessaires à l’étude du lentillage gravitationnel faible.

Comme mentionné précédemment, la prochaine génération de relevés de galaxies dont fait partie LSST, va observer le ciel bien plus profondément que les précédents relevés. La densité de galaxies utilisables dans les analyses de lentillage gravitationnel augmente alors drastiquement, mais avec elle la probabilité d’observer des objets superposés. Un exemple peut être donné par le relevé HSC (Hyper Suprime-Cam Subaru Strategic Survey) qui produit aujourd’hui les images les plus proches de celles que va produire LSST. Parmi les sources observées par ce relevé, 58% sont superposés à d’autres objets. Cette expérience observe des objets jusqu’à des magnitudes de l’ordre de 26 en

bande i , alors que LSST va observer jusqu'à une magnitude ≈ 26.8 dans cette même bande (≈ 27.5 en bande r). Le taux d'objets correspondant à des sources astrophysiques superposées devrait alors atteindre environ 62%.

Il n'est donc pas possible d'exclure ces objets des analyses de lentillage gravitationnel, sous peine de diminuer largement le nombre de galaxies utilisables. Il existe alors deux solutions au biais apporté par cet effet systématique : calibrer les mesures en utilisant des simulations ou des algorithmes tel que **METADTECTION**, ou effectuer la séparation de sources à partir des images.

Pour cette seconde solution, de nombreux algorithmes ont déjà été proposés. Une des méthodes les plus utilisées jusqu'ici est **SExtractor** qui permet de faire de la segmentation entre différentes sources (chaque pixel est associé à une source particulière) sur des images mono-bande (une seule bande de fréquences observée). Plus récemment de nouvelles techniques ont été développées pour séparer différentes sources sur des images multi-bandes en associant une portion du flux de chaque pixel à différentes sources (**PHOTO**, algorithmes utilisés pour la Dark Energy Survey - DES, **SCARLET** par exemple). Enfin, des méthodes d'apprentissage profond ont également été appliquées à ce genre de problème, mais principalement dans un objectif d'estimation du flux des sources, ou de classification de ces sources. Seul un premier travail utilisant un algorithme d'apprentissage profond, un GAN (Generative Adversarial Network) en l'occurrence, a cherché à réaliser la séparation de sources pour permettre une mesure de forme et de flux des galaxies.

La première méthode proposée dans cette thèse est basée sur un autre type de réseau de neurones génératif, les autoencodeurs variationnels.

Apprentissage profond pour la cosmologie

Eléments de base de l'apprentissage profond

Les méthodes utilisées dans cette thèse et appliquées aux images issues des relevés cosmologiques sont basées sur des méthodes d'apprentissage profond, et principalement sur les réseaux de neurones artificiels.

Un réseau de neurones est un empilement de couches de neurones artificiels permettant de d'apprendre une tâche spécifique. Ces couches peuvent être de différents types (denses ou convolutionnelles par exemple) et ont chacune une fonction particulière. Pour l'application de réseaux de neurones artificiels à des images (ce qui est notre cas dans ce manuscrit), les couches convolutionnelles sont les plus aptes à extraire les informations pertinentes d'une image, et ce à différentes échelles.

Ce travail est particulièrement basé sur un type de réseau de neurones, les réseaux de neurones génératifs, et notamment les autoencodeurs variationnels (VAE). Un tel réseau est, dans notre cas, constitué d'un encodeur et d'un décodeur composés de couches convolutionnelles. Il permet, à partir d'une image, d'encoder l'information nécessaire à sa reproduction par le décodeur, dans un espace latent de faible dimension. La particularité de l'autoencodeur variationnel est de considérer une couche probabiliste pour l'espace latent : les différents neurones de cette couche ne sont pas fixés à des valeurs singulières mais des distributions de probabilité leur sont plutôt associées. Ces distributions sont échantillonnées à chaque passage d'un nouvel exemple dans le réseau. Une image donnée plusieurs fois en entrée du VAE permettra donc la génération d'images légèrement différentes en sortie du réseau.

Introduction aux réseaux de neurones bayésiens

Ces couches probabilistes permettent donc de prendre en compte et de mesurer une certaine sorte d'incertitude, l'incertitude aléatoire, intrinsèque aux données. En effet, aux données fournies au réseau, une incertitude est associée qui impacte les résultats obtenus en sortie d'un réseau de neurones.

Cependant, elles ne permettent pas d'estimer l'incertitude épistémique, apportée par le réseau de neurones lui-même. Cette incertitude est liée en partie à l'architecture du réseau et, de façon plus importante, aux données sur lesquelles il a été entraîné. Prenons l'exemple d'un réseau entraîné à classer des images de sources astrophysiques dans les catégories *étoiles* ou *galaxies*. Dans le cas où l'on présente une image d'un autre objet (par exemple la trace laissée par le passage d'un satellite), il serait souhaitable que le réseau présente une très forte incertitude puisqu'il n'a jamais rencontré ce genre d'exemples lors de son entraînement. Cependant, ce n'est pas possible avec des réseaux "classiques" ou "déterministes" et il faut pour cela utiliser des *réseaux de neurones bayésiens*. Dans ce type de réseaux de neurones, les poids et les biais ne sont plus fixés à des valeurs précises mais une distribution de probabilité leur est associée et ce sont ces distributions qui sont apprises lors de l'entraînement.

L'entraînement de ces réseaux constitue une difficulté de part leur nature probabiliste. Cependant il est rendu possible, notamment en utilisant des méthodes d'inférence variationnelle, et plusieurs approximations permettent d'obtenir des résultats satisfaisants. Les distributions apprises sur les poids permettent alors d'évaluer l'incertitude épistémique en donnant au réseau plusieurs fois le même exemple et en estimant la variance des mesures (ou prédictions).

Apprentissage profond appliqué à la cosmologie

L'estimation de cette incertitude épistémique paraît d'autant plus importante que l'apprentissage profond est de plus en plus utilisé en astrophysique et plus particulièrement en cosmologie. En effet, elle permet d'éviter d'utiliser des réseaux de neurones pour des cas non adaptés mais également d'apporter une information améliorant l'interprétation des résultats obtenus avec ces réseaux.

Un reproche classique fait aux réseaux de neurones tient justement à ce manque d'interprétabilité. L'incertitude épistémique est donc un outil permettant d'améliorer cet aspect bien que d'autres outils existent pour améliorer la compréhension d'un résultat obtenu avec un réseau de neurones. Par exemple, les "saliency maps" permettent d'identifier quelles parties d'une image ont été utilisées par le réseau pour effectuer sa prédiction. La problématique de l'interprétabilité des réseaux de neurones est un domaine de recherche actif et de plus en plus de méthodes sont développées pour améliorer la compréhension du fonctionnement des réseaux de neurones artificiels. L'argument de la *boîte noire* opposé à ce type d'algorithmes devient donc de moins en moins pertinent.

De plus, il est intéressant de noter qu'utiliser des méthodes d'apprentissage profond en cosmologie peut permettre de produire des avancées dans chacun des deux domaines. Les réseaux de neurones artificiels peuvent parfois produire des résultats plus précis que les méthodes classiques pour des tâches spécifiques sur des données cosmologiques, notamment du fait qu'un réseau de neurones peut, en théorie, apprendre une fonction arbitrairement complexe à partir de ces données. Inversement, les données issues de relevés cosmologiques offrent des cas d'applications particulièrement complexes pouvant aboutir à de nouvelles techniques d'apprentissage profond.

Les résultats présentés dans cette thèse ont été obtenus en appliquant certaines de ces méthodes au problème de la superposition de galaxies.

Méthodes d'apprentissage profond bayésien appliquées à la superposition de galaxies

Deux solutions basées sur les méthodes bayésiennes d'apprentissage profond sont proposées dans cette partie.

Séparation de galaxies à l'aide d'autoencodeurs variationnels

La première est un algorithme de séparation de galaxies qui utilise des VAEs. Dans un premier temps un VAE est entraîné à générer des images de galaxies isolées. Cela permet d'utiliser le décodeur entraîné de ce VAE comme un prior pour la génération d'images de galaxies isolées dans un second réseau, d'architecture similaire, pour faire la séparation de galaxies. Ce second réseau est alimenté par des images simulées de galaxies superposées et a pour objectif d'extraire uniquement l'image de la galaxie centrée sur l'image d'entrée, qui est aussi la plus brillante. Pour cela il utilise le prior entraîné précédemment comme décodeur, avec ses poids fixés. Seul l'encodeur est entraîné pour apprendre à encoder uniquement les informations nécessaires à la reproduction de l'image de la galaxie centrée dans l'espace latent. La séparation de galaxies est donc réalisée lors de cette étape d'encodage.

Nous montrons dans un premier temps que cette méthode fonctionne sur des simulations d'images multi-bandes LSST pour la séparation de galaxie : elle permet une reconstruction précise de la forme et du flux des galaxies. Afin d'évaluer l'apport d'une analyse jointe de données issues de LSST et d'un télescope spatial, ces dernières ne souffrant pas de la PSF due à l'atmosphère, nous avons réalisé l'analyse conjointe des pixels d'images simulées pour LSST et Euclid. Cela permet une diminution de l'erreur médiane de reconstruction des formes des galaxies allant de 8 à 47%. Ensuite, la superposition de sources étant étroitement liée à l'étape de détection, nous démontrons la robustesse de notre méthode à de petits décentrages, en testant deux configurations : une configuration supposant une détection parfaite mais un décentrage dû à la pixelisation de l'image, et une configuration prenant en compte les décentrages obtenus avec un algorithme de détection. Pour finir, et montrer la possibilité d'appliquer notre méthode à de vraies données, nous avons artificiellement superposé des images de vraies galaxies et montré l'intérêt du transfert d'apprentissage à partir d'un réseau de neurones entraîné sur des images de galaxies simulées.

J'ai par la suite encadré le stage de Thomas Sainrat lors duquel nous avons mis au point un processus itératif afin de séparer toutes les galaxies d'une image en passant par des étapes de détection, classification, et séparation des sources. Des réseaux de neurones sont également utilisés pour les étapes de détection et de classification. La mise en place de ce processus itératif a permis d'obtenir des résultats préliminaires sur des champs de galaxies simulées.

Enfin, pour se rapprocher un peu plus des futures images LSST, cet algorithme de séparation de galaxies est testé sur des images extraites de la simulation DC2 (*Data Challenge 2*), générée au sein de la collaboration DESC dans le but de préparer à l'analyse des futures images prises par LSST. Les résultats obtenus montrent une amélioration de 70 à 120% sur l'erreur médiane de reconstruction des formes de galaxies comparée à la méthode générique utilisée dans la pipeline LSST actuellement. Ces résultats préliminaires devront être confrontés aux résultats qui seront obtenus avec l'autre algorithme de séparation de sources récemment implémenté dans la pipeline LSST, SCARLET.

Réseaux de neurones bayésiens pour l'estimation de paramètres sur des galaxies superposées: mesures d'ellipticité et de décalage vers le rouge

Les résultats précédents montrent donc qu'il est possible, à partir d'une image contenant des sources superposées, d'encoder dans un espace latent les informations nécessaires à la reconstruction de la galaxie centrée sur cette image, par un décodeur entraîné. L'espace latent contient donc, entre autres, les informations de forme et de décalage vers le rouge de cette galaxie, les deux paramètres nécessaires à l'analyse du lentillage gravitationnel faible.

Par conséquent, nous proposons un réseau de neurones permettant de faire directement l'estimation combinée de ces paramètres à partir des images DC2, sans appliquer de méthode de séparation de sources. L'architecture du réseau est reprise de celle de l'encodeur précédemment implémenté (réseau de neurones convolutif) et les résultats montrent une estimation précise des paramètres pour des réseaux de neurones déterministes, du même ordre de grandeur, pour la mesure de forme, que ceux obtenus avec l'algorithme de séparation de sources précédemment décrit.

Cet algorithme ne permettant pas d'avoir une estimation de l'incertitude épistémique, nous appliquons ensuite un réseau de neurones bayésiens à l'estimation de formes des galaxies. Ce réseau permet d'évaluer la distribution de probabilité à posteriori des paramètres de forme d'une galaxie, ce qui permet d'obtenir une diminution de 50% de l'erreur médiane pour l'estimation des ellipticités. Par ailleurs, il est possible d'utiliser l'estimation de l'incertitude épistémique pour diminuer l'importance d'une mesure qui a de fortes probabilités d'être erronée dans l'analyse, voire de la rejeter. Ainsi il est possible, par exemple, de diminuer l'erreur moyenne sur la mesure d'ellipticité d'un échantillon de galaxies. Nous montrons donc que cette incertitude apporte une information cruciale à l'utilisation des réseaux de neurones en cosmologie.

La future mesure combinée des distributions de probabilité a posteriori des paramètres d'ellipticité et de décalage vers le rouge des galaxies, à l'aide d'un réseau de neurones bayésiens, permettra donc d'obtenir des mesures précises des deux paramètres nécessaires à l'analyse du lentillage gravitationnel faible.

Les deux méthodes développées dans ce manuscrit montrent l'intérêt des techniques d'apprentissage profond bayésien pour des problématiques spécifiques à la cosmologie comme l'effet systématique de superposition de galaxies, pour lequel elles présentent des résultats probants. L'utilisation croissante de ces techniques en cosmologie montre l'intérêt porté par la communauté scientifique à leur égard et suggère qu'elles seront des outils importants pour relever les défis de la cosmologie moderne.

Contents

Résumé	i
Abstract	iii
Remerciements	vi
Résumé substantiel du manuscrit	viii
Introduction	2
I Weak lensing with the LSST	6
1 The standard cosmological model	7
1.1 A Universe in accelerated expansion	7
1.1.1 General relativity	7
1.1.2 Cosmology	8
1.1.3 Density perturbations and matter power spectrum	9
1.2 Different probes for different epochs	11
1.2.1 Early Universe: the cosmic microwave background	11
1.2.2 Large Scale Structures	13
1.2.3 Supernovae and the H_0 tension	17
1.3 Next generation of surveys	20
1.3.1 LSST	20
1.3.2 Euclid	22
2 Weak lensing	30
2.1 Weak lensing formalism	30
2.1.1 Deflection from a source	30
2.1.2 Deflection of light in the Universe	31
2.1.3 Lensing quantities: Shear and convergence	32
2.1.4 Convergence power spectrum	33
2.2 Shear related probes	34
2.2.1 Shear correlation function	34
2.2.2 3x2-point correlation function	37
2.2.3 Intrinsic alignments: a bias in cosmological parameters estimation	37
2.2.4 Shear tomography	39
2.3 Cosmological constraints with weak lensing	41

2.3.1	Recent results	41
2.3.2	Expectations for LSST	44
2.4	Measuring shear from galaxies shapes	45
2.4.1	From image to ellipticities	45
2.4.2	From ellipticities to the shear	45
2.5	Weak lensing pipeline and associated biases	46
2.5.1	Detection and blending	46
2.5.2	Point-Spread Function modelling	48
2.5.3	Ellipticity measurement	49
2.5.4	Selection bias	51
2.5.5	Shear Calibration	51
3	The blending systematic for weak gravitational lensing	58
3.1	Deep ground-based galaxy surveys: blending as a systematic	58
3.1.1	Deep surveys	58
3.1.2	Blending metrics	59
3.2	Impact of blending	61
3.2.1	Effective number density, n_{eff}	61
3.2.2	Shear biases	62
3.3	Dealing with blending	64
3.3.1	Calibration	65
3.3.2	Deblending	66
3.3.3	Probabilistic cataloging	71
3.3.4	Blending ToolKit	71
3.3.5	Multi-band, multi-instrument analysis	72
II	Deep learning for cosmology	78
4	Elements of deep learning	80
4.1	Components of a neural network	80
4.1.1	Dense layers - fully connected layers	80
4.1.2	Convolutional layers	82
4.1.3	Other types of layer	83
4.2	The Backpropagation algorithm	84
4.2.1	A simple example: single-neuron layers neural network	84
4.2.2	Optimisation with stochastic gradient descent	85
4.2.3	Training issues and examples of regularisation	86
4.3	Generative models	89
4.3.1	Variational Autoencoder (VAE)	89
4.3.2	Generative Adversarial Networks (GAN)	90
4.3.3	Normalizing flows	91
4.4	Example: Generate images from a VAE latent space	92
4.4.1	Direct sampling of the latent space	92
4.4.2	Sampling with Normalizing flows	93
4.4.3	Conclusion	94

5	Introduction to Bayesian neural networks	99
5.1	Uncertainties for neural networks	99
5.1.1	Aleatoric and Epistemic uncertainty	99
5.1.2	Example: fit a dataset	100
5.2	Bayesian neural networks: Definition	101
5.2.1	Bayes' rule	101
5.2.2	Neural networks as statistical models	101
5.3	Approximate inference: Monte-Carlo and Variational Inference	103
5.3.1	Monte-Carlo methods	103
5.3.2	Variational inference	104
5.3.3	Training BNNs	108
6	Cosmology with deep learning	110
6.1	The growing use of deep learning for cosmology	110
6.1.1	Interpretability of neural networks	110
6.1.2	Large and deep surveys	112
6.1.3	Interaction between cosmology and deep learning	112
6.2	Cosmology examples	113
III Bayesian deep learning applied to blended galaxies		118
7	Variational AutoEncoders for deblending	120
7.1	Introduction	120
7.2	Simulated images	121
7.2.1	Datasets	121
7.2.2	Blendedness metrics	125
7.2.3	Image preprocessing	125
7.3	Method	126
7.3.1	Variational autoencoders	126
7.3.2	Implementation and architecture	126
7.3.3	Deblender	128
7.4	Results	128
7.4.1	Reconstruction metrics	128
7.4.2	Prior for single galaxies with a VAE	129
7.4.3	Deblending performance	131
7.4.4	Effect of decentring	134
7.5	Discussion	136
7.5.1	Detection and decentring	136
7.5.2	Challenges with real data	137
7.5.3	Towards a Bayesian weak-lensing analysis	139
7.6	Summary	141
7.7	Deblending pipeline	142
7.7.1	Pipeline structure	142
7.7.2	Detection	144
7.7.3	Classification and residual masking	145
7.7.4	Pipeline results	148
7.7.5	Discussion and Future work	148

7.8	Application to DC2 simulations	150
7.8.1	DC2 simulations	150
7.8.2	DC2 images for training	151
7.8.3	Network architecture choice and training procedure	155
7.8.4	Results	156
7.8.5	Conclusion	160
7.8.6	Future work	161
8	Bayesian neural network for galaxy parameters estimation in blended scenes: ellipticity and redshift measurements	166
8.1	Introduction	166
8.2	Simulated images: DC2 simulation	167
8.3	Parameter estimation from the latent space of a trained deblender	168
8.3.1	Architecture	168
8.3.2	Results	169
8.4	CNN and BNN for galaxy parameters estimation	172
8.4.1	Architecture	172
8.4.2	Results	174
8.5	Ellipticity and redshift estimation	181
8.6	Discussion and future work	183
8.6.1	Application to real data	183
8.6.2	Bias in ellipticity distribution	183
8.6.3	Choice of prior	186
8.7	Conclusion	186
	Conclusion	190
A	Comparison of Graphcore IPUs and Nvidia GPUs for cosmology applications	193
A.1	Introduction	193
A.2	Hardware description	195
A.3	Cosmological use cases	196
A.3.1	Training data	196
A.3.2	Galaxy shape parameter estimation	196
A.3.3	Galaxy image generation	198
A.4	Summary and discussion	200
A.5	Acknowledgements	200
B	A Bayesian Convolutional Neural Network for Robust Galaxy Ellipticity Re- gression	203
C	Is blending always an issue ?	220

Introduction

...; the chance of unfavourable weather is the chief but by no means the only apprehension.

A.S. Eddington, The Observatory, March 1919, Vol 42, pages 119-122

This is full of hope that A.S. Eddington starts the conclusion of his paper detailing the motivations for the two expeditions sent to observe the total solar eclipse of the 29th of May 1919. He finishes by detailing the major consequences that this observation could lead to:

The present eclipse expeditions may for the first time demonstrate the weight of light; or they may confirm Einstein's weird theory of non-Euclidean space; or they may lead to a result of yet more far-reaching consequences - no deflection.

A.S. Eddington, The Observatory, March 1919, Vol 42, pages 119-122

In a paper published in 1920, Dyson, Eddington, and Davidson presented their results, measuring that light coming from stars behind the sun was deflected from an amount in agreement with the predictions of Einstein's theory of General Relativity (and in disagreement with the two other hypotheses).

The first use of gravitational lensing, the deflection of light due to the bending of space-time induced by mass along the line of sight, was to prove General Relativity. Its observation was a turning point in our understanding of the Universe, and it is at the core of this thesis.

Gravitational lensing is one of the major probes to study dark matter and dark energy, the two main components of our standard cosmological model, or Λ CDM model. Today, this model, which can be defined using only six parameters, stands on solid ground. One of the major argument in its favour comes from the measurements on the Cosmic Microwave Background (CMB) by the Planck satellite. Initially discovered by Penzias and Wilson in 1964, the CMB has been measured by several instruments, and in particular by the FIRAS instrument on the COBE satellite, which measured its spectrum to be extremely close to a black-body spectrum, and detected the CMB anisotropies. These anisotropies result from density perturbations in the early Universe that froze at recombination, and they carry a lot of information about the physics of that period. They were measured by Planck with an extremely high precision, allowing to apply strong constraints on the cosmological parameters, particularly through the computation of the temperature and E-modes power spectra. Both these measurements were found in excellent agreement with the Λ CDM model.

The history of the Universe can also be probed by large scale structures. Historically, the expansion of the Universe was discovered by Hubble who, observing galaxies, noticed that the ones which are the further from us are also moving away the fastest. He derived a linear relationship between the recession velocities of galaxies, measured through their redshift, and their distances, measured using the work of Henrietta Leavitt on Cepheids, known as Hubble's law. The acceleration of this expansion was discovered around 70 years later by the High-Z Supernova Search Team,

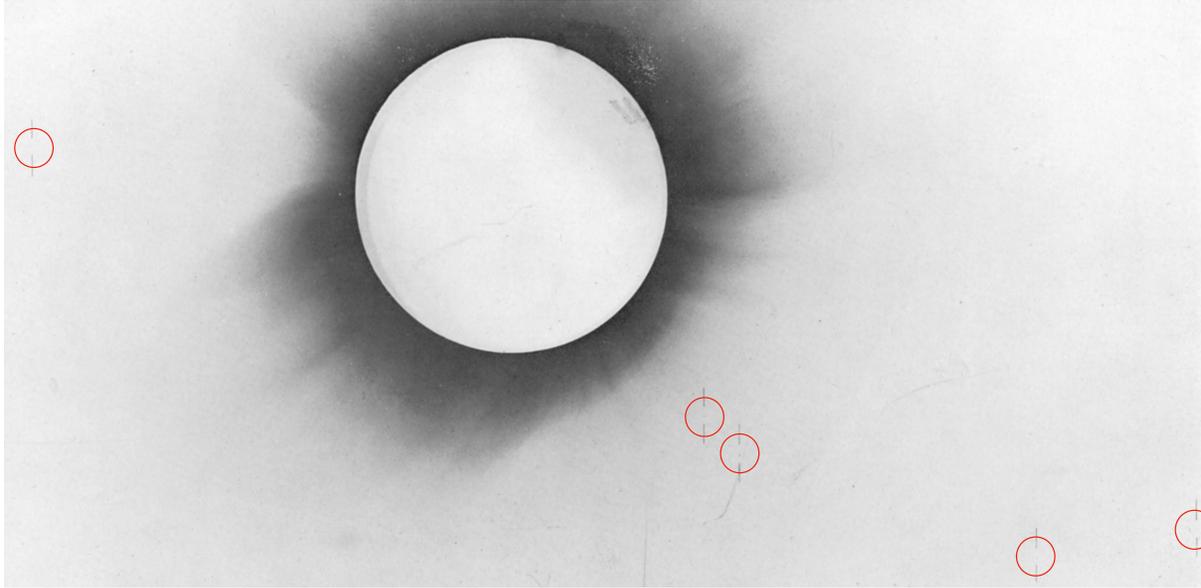


Figure 1: Image of the eclipse of 1919 taken with a photographic plate by one of the two expeditions. The red circles show five of the seven stars used to measure the deflection of light by the sun's gravitational field. Credit: Dyson, Eddington and Davidson (1919)

simultaneously with the The Supernova Cosmology Project, using supernovae, bright explosions of white dwarfs in binary systems¹, to probe higher redshifts. Similarly to the CMB, the study of large scale structures probes, such as galaxy clustering, clusters of galaxies, or gravitational lensing for example, allow to apply constraints on the cosmological parameters. Until now, no observation showed a significant deviation from the standard model.

Nevertheless, and fortunately, there is still a lot that we don't know. First, the origin of the fluctuations observed in the CMB at large angular scales is unknown, and even if inflation is the favoured hypothesis today, no conclusive proof has been observed yet. Finding the B-modes in the polarisation of the CMB would be a major step in that regard, and lots of future surveys are devoted to this task (QUBIC, Simons Array, POLARBEAR). Then, the nature of dark energy, accounting for about 70% of the total energy of the Universe according to the Λ CDM model, is unknown. Again, several large galaxy surveys (LSST, Euclid, Roman) are specifically designed to apply tight constraints on the parameters of the dark energy equation of state in a beyond- Λ CDM model (allowing for a time varying proportionality coefficient in the dark energy equation of state), and they should soon start recording data.

This is where gravitational lensing comes back in the forefront, as a powerful probe to constrain dark energy parameters. In particular, the weak regime of gravitational lensing, weak lensing for short, already showed promising results from surveys like CFHTLenS, KiDS, HSC, or DES. It is one of the major probes for future Stage IV galaxy surveys and one of the four major scientific goals of the Vera C. Rubin Legacy Survey of Space and Time (LSST). These surveys are specifically designed for weak lensing, with wide observed sky areas, and depths never reached before, leading cosmology into the so-called "precision era". With these surveys the main source of uncertainties in the weak lensing analysis will not be the statistical uncertainty any more, it will come from

¹Note that this is the current consensus. Parts of this mechanism are still unknown.



Figure 2: Portion of a tract of LSST after five years of operation simulated in the Data Challenge 2 (DC2) for the LSST Dark Energy Science Collaboration. Numerous objects appear overlapped with other objects, they are *blended*. Credit: Jim Chiang (LSST Dark Energy Science Collaboration (LSST DESC) et al., 2021)

systematic errors.

The work presented in this thesis was done in the context of the LSST Dark Energy Science Collaboration (DESC), and is particularly focused on a dominant systematic effect for weak lensing analysis: the overlap of sources on an image, an effect called *blending* (see fig. 2). Due to its characteristics, the LSST will face an unprecedented and concerning proportion of blended objects.

Cosmological constraints coming from weak lensing rely on statistical measurements of shapes, or galaxy ellipticities, and photometric redshifts². Obviously, measuring the shape or the flux of a galaxy is much more difficult if it is blended and state-of-the-art shape measurement methods, for example, fail when applied to blended objects (see fig. 3). To handle blended sources, one usually turns to deblending algorithms: algorithms that can separate the objects while conserving the shape and flux of each of them, before doing the shape, or flux measurement.

In this thesis, I propose two avenues to this systematic effect, both based on neural networks. The first one is a deblending algorithm that I implemented using a kind of deep generative models, namely variational autoencoders (VAE). This method works in two steps: first, a VAE learns the features of isolated galaxy images and their representation in a latent space. Then, the decoder of the trained VAE is used as a prior in a second VAE-like network, which learns to deblend images of blended scenes by encoding the centred galaxy representation into the latent space. This method naturally generalises to multi-band and multi-instrument data, a major advantage of our algorithm compared to other state-of-the-art methods, that I tested simulating LSST and Euclid-like data.

The second solution that I propose is a galaxy parameters estimation algorithm that can perform as good on isolated galaxies as on blended scenes, without the need for a deblending algorithm. It is inspired from the architecture of the previous algorithm. Since the encoder of the deblender is

²In the case of photometric surveys, the redshift of a galaxy can be obtained by measuring its flux in different bands and by fitting (or learning) a redshifted spectral energy distribution model, a measure called photometric redshift.

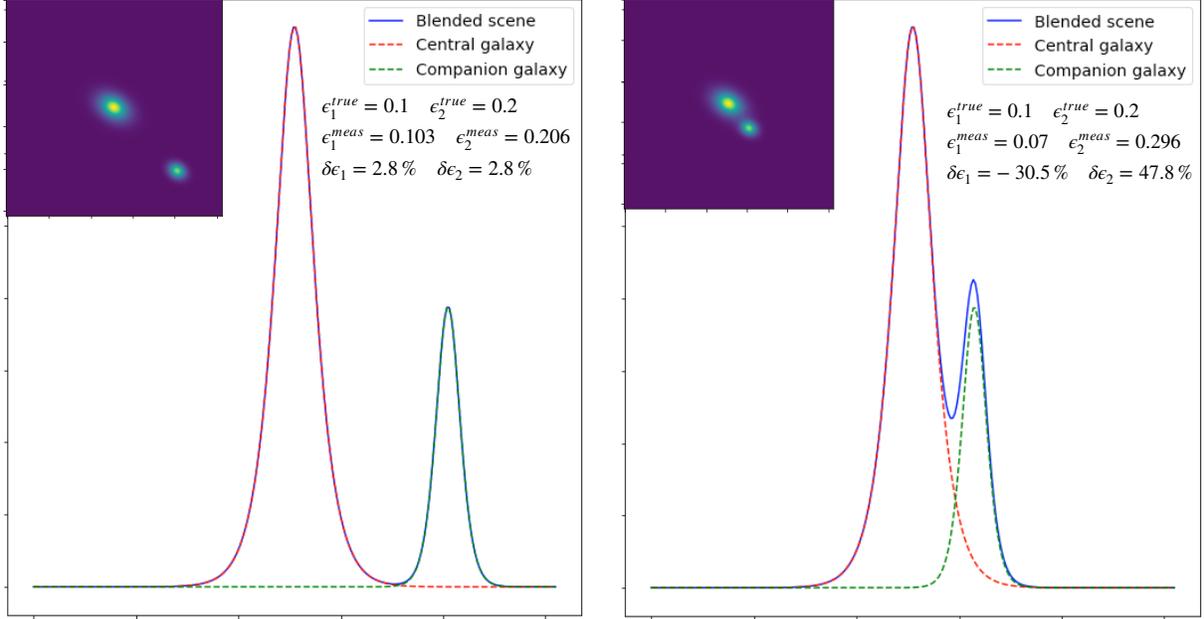


Figure 3: Toy example of the impact of blending on shape measurement. The shape measurement method used here is the KSB algorithm, a moment-based method. On the two panels, the light profile of the two sources is plot, the one from the centred galaxy is in red, the one from it companion is in green, and the total light profile is in blue. In the top left corner, the corresponding scene is showed. On the right is displayed the values of the two ellipticity parameters (ϵ_1, ϵ_2) . $(\epsilon_1^{true}, \epsilon_2^{true})$ are the true values, $(\epsilon_1^{meas}, \epsilon_2^{meas})$ are the values measured by the KSB algorithm, and $(\delta\epsilon_1, \delta\epsilon_2)$ are the errors in percentage. In the case where the galaxies are blended (right panel), the light profiles overlap leading to large errors on the ellipticity parameters estimation. Credit: Thomas Sainrat.

able to extract from a blended scene, the relevant parameters required for the image generation of the centred galaxy, an encoder-like neural network should be able to provide a direct estimation of the weak lensing relevant parameters, namely the ellipticity and redshift of the galaxy. I test this hypothesis using a convolutional neural network (CNN) and a Bayesian neural network (BNN). The latter presents the advantage of providing an estimation of the epistemic uncertainty, the one coming from the model, allowing to yield a reliable estimation of the full predictive posterior distribution of the measured parameters.

This thesis is divided into three parts. The first one is dedicated to cosmology with an introduction to the standard model, a description of weak lensing and how to constrain cosmological parameters using the cosmic shear. A focus is made on blending as a systematic error, since it will be a dominant source of bias for cosmic shear analysis. The second part is dedicated to deep learning and its application to cosmology. I first give some elements of deep learning before introducing to the formalism and methods used to train Bayesian neural networks. I conclude this part by discussing the need for deep learning in cosmology. Finally, my work is described in the third part with two chapters dedicated to each of the two projects presented above.

Part I

Weak lensing with the LSST

Chapter 1

The standard cosmological model

1.1 A Universe in accelerated expansion

This section is inspired from Peter and Uzan (2013), Dodelson (2003), Baumann (2018), Peacock (1999) and Schneider, Kochanek, and Wambsganss (2006).

At the beginning of the twentieth century, physics needed a new theory of gravity since Newtonian physics was failing at different levels. Einstein proposed in 1915 the theory of general relativity (GR), which is at the origin of modern cosmology.

1.1.1 General relativity

This theory describes space-time as a four-dimensional manifold which relies on a metric, $g_{\mu\nu}$, allowing to measure space-time intervals. Two events in space-time are separated by a distance

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad (1.1)$$

Here μ and ν describe the four dimensions of space-time, they can range between 0 and 3. In classical mechanics, the *equivalence principle* states that, if no external force acts on a particle, its trajectory follows a straight line. In general relativity, space-time is curved but this principle holds as particles only subject to gravity move along geodesics, according to the equation:

$$\frac{d^2 X^\alpha}{d\tau^2} + \Gamma_{\mu\nu}^\alpha \frac{dX^\mu}{d\tau} \frac{dX^\nu}{d\tau} = 0 \quad (1.2)$$

where $X^\mu(\tau)$ is the trajectory, τ the proper time, i.e. the time measured in the referential of the object, and $\Gamma_{\mu\nu}^\alpha$ is the Christoffel symbol. The latter is defined as

$$\Gamma_{\mu\nu}^\alpha = \frac{g^{\alpha\beta}}{2} \left[\frac{\partial g_{\mu\beta}}{\partial x^\nu} + \frac{\partial g_{\nu\beta}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\beta} \right] \quad (1.3)$$

Hence the equation of motion of a light particle, which follows geodesics, only depends on the metric g , which defines the geometry of space-time. Light paths follow the curvature of space-time.

The dynamic of the metric is governed by Einstein's field equations, relating the Einstein tensor, $G_{\mu\nu}$, which depends on the metric $g_{\mu\nu}$, to the stress-energy tensor, which describes the energy content of the Universe, $T_{\mu\nu}$:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \quad (1.4)$$

where the cosmological constant Λ arises. In this equation, G is Newton's gravity constant, and c the speed of light in vacuum.

1.1.2 Cosmology

The *cosmological principle* states that on large scales, the Universe is homogeneous and isotropic. Combining this affirmation with eq. (1.1), one can define the FLRW metric (for Friedmann-Lemaître-Robertson-Walker) as

$$ds^2 = -c^2 dt^2 + a^2(t) \left[d\chi^2 + f_K^2(\chi) d\Omega^2 \right] \quad (1.5)$$

where χ is the *comoving distance*, t is the cosmic time, $a(t)$ the *scale factor*, and $f_K(\chi)$ the *comoving angular distance* defined as

$$f_K(\chi) = \begin{cases} \sin(\sqrt{K}\chi) / \sqrt{K} & , \text{ if } K > 0 \\ \chi & , \text{ if } K = 0 \\ \sinh(\sqrt{-K}\chi) / \sqrt{-K} & , \text{ if } K < 0 \end{cases} \quad (1.6)$$

and which depends on the *curvature* of space-time, K . The *comoving distance* is the distance between two observers statics in space, i.e. with comoving coordinates (blind to the expansion of the Universe). The curvature is positive, null, or negative respectively for open, flat, or close space.

The *scale factor*, a , describes how space is expanding or contracting over time. The definition of the physical distance between two objects is then

$$\mathbf{r}(t) = a(t)\boldsymbol{\chi} \quad . \quad (1.7)$$

$\boldsymbol{\chi}$ is the *comoving distance*.

Following the cosmological principle, one can consider the content of the Universe as a perfect, homogeneous and static fluid. The stress-energy tensor then becomes a diagonal matrix with

- $T_{00} = \rho$, ρ being the energy density of the fluid (considering $c = 1$ here and in the following),
- and $T_{ii} = p$, p being the pressure of the fluid.

Then, using Einstein's equations one can derive the Friedmann equations:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{K}{a^2} \quad (1.8)$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \quad (1.9)$$

where $H = \dot{a}/a$ is the Hubble parameter. It represents the expansion rate of our Universe at a particularly time t .

The critical density ρ_c can be defined from eq. (1.8) as

$$\rho_c = \frac{3H^2}{8\pi G} \quad (1.10)$$

and allows for the definition of the density parameters. They can be written, in the case of flat space:

$$\Omega_x = \frac{\rho_x(a)}{\rho_c} \quad (1.11)$$

where $x \in \{rad, \gamma, \nu, m, b, c\}$, with *rad* standing for radiation, γ for photons, ν for neutrinos, m for matter, b for baryonic matter, and c for cold dark matter. Ω_{rad} is the total radiation density,

gathering neutrinos and photons. Ω_m is the total matter density, gathering baryonic and cold dark matter. It is also possible to define $\Omega_\Lambda = \Lambda/3H^2$ and $\Omega_k = -K/a^2H^2$ to re-write the first Friedmann equation, eq. (1.8), as a function of these density parameters only:

$$\Omega_m + \Omega_{rad} + \Omega_\Lambda + \Omega_k = 1 \quad (1.12)$$

This equation can be modified to include the scale factor and the Hubble parameter as:

$$\left(\frac{H}{H_0}\right)^2 = \Omega_{m_0} \left(\frac{a}{a_0}\right)^{-3} + \Omega_{rad_0} \left(\frac{a}{a_0}\right)^{-4} + \Omega_{k_0} \left(\frac{a}{a_0}\right)^{-2} + \Omega_{\Lambda_0} \quad (1.13)$$

with the subscript $_0$ indicating that the values of the density parameters are taken today. In the following this subscript is assumed even if it is not written for density parameters. To obtain eq. (1.13) one uses the relation

$$\rho(t) \propto a(t)^{-3(1+w)} \quad (1.14)$$

for fluids with an equation of state $p = w\rho$, w being constant. It is the solution of the equation

$$\dot{\rho} + 3H(\rho + p) = 0 \quad (1.15)$$

derived from the conservation equation, $\nabla_\mu T^{\mu\nu} = 0$. Note that for a flat Universe, the term in Ω_k vanishes, this hypothesis is considered in the rest.

Equation (1.13) introduces the parameter H_0 called the *Hubble constant*, it is the Universe expansion rate, today. In section 1.2.3, I discuss the current tension that different measurements of this parameter, apart from each other, has generated and ways to solve it.

This equation also presents the Λ CDM parametrisation with a value of $w = 0$ for the radiation, $w = 1/3$ for the matter and $w = -1$ for the cosmological constant, which defines a fluid with negative pressure called *dark energy*.

Other models exist to parametrise the equation of state of dark energy. In the w CDM models for example, w is a free parameter that can be different from -1 . Last results from the Dark Energy Survey give $w = -0.98^{+0.32}_{-0.20}$ (DES Collaboration et al., 2021). It is also possible to consider the w parameter as evolving through time, for example linearly as a function of redshift, with $w = w_0 + w_1z$ (Cooray and Huterer, 1999). However this grows increasingly unsuitable at redshifts larger than one, so other parametrisation have been proposed, the most simple and most used being

$$w = w_0 + w_a \frac{z}{1+z} \quad (1.16)$$

proposed by Chevallier and Polarski (2001) and Linder (2003). Here, z is the *redshift* defined as:

$$1 + z = \frac{a(t_0)}{a(t)} \quad (1.17)$$

Hence the equation of state of dark energy could evolve through time. Constraining the parameters of this equation of state are a major challenge of current cosmology (see section 1.2.2). Nowadays, the Λ CDM model is favoured for its simplicity and as no cosmological result permits to discard it.

1.1.3 Density perturbations and matter power spectrum

Even though the cosmological principle holds at large scales, to allow for the formation of structures in the Universe it is assumed that there were small fluctuations at early times, supposedly originated from quantum fluctuations during the early Universe, that grew by gravitational interaction through

time. They remained small, i.e. the density contrast (defined below) is much smaller than unity, and linear perturbations theory suffices for their description. Here we consider non-relativistic perturbations on scales smaller than the *horizon*¹. The matter fluctuations can be characterized by the density contrast δ ,

$$\delta(\mathbf{x}, a) = \frac{\rho(\mathbf{x}, a) - \bar{\rho}(a)}{\bar{\rho}(a)} \quad (1.18)$$

$\bar{\rho}$ being the mean matter density.

With linear perturbation theory, it is then possible to describe the growth of density perturbations as

$$\delta(a) = \delta_0 a g(a) \quad (1.19)$$

with δ_0 the density contrast linearly extrapolated to present time and g the density-dependent growth function, which depends on the scale factor and the density parameters.

Assuming that the density field is a Gaussian random field, the statistics of the density fluctuations at a position \mathbf{x} can be fully characterized by its *power spectrum*. It can be computed from the *two point correlation function*:

$$\xi_\delta(\mathbf{x}, \mathbf{x}') = \langle \delta(\mathbf{x}) \delta(\mathbf{x}') \rangle \quad (1.20)$$

which can be linked to the power spectrum by going to Fourier space:

$$\langle \delta(\mathbf{k}) \delta^*(\mathbf{k}') \rangle = \delta^{(3)}(\mathbf{k} - \mathbf{k}') P(\mathbf{k}) \quad (1.21)$$

where the power spectrum is defined as

$$P_\delta(k) = \int d^3 r \xi(r) e^{i\mathbf{k}r} \quad (1.22)$$

The latter equation can be inverted to express the correlation function as a function of the power spectrum:

$$\xi(r, t) = \int_0^\infty \frac{dk}{2\pi^2} k^2 P_\delta(k, t) \frac{\sin(kr)}{kr} \quad (1.23)$$

The primordial power spectrum can be predicted from inflation theory for example, leading, for scalar modes, to

$$P_\zeta = A_s \left(\frac{k}{k_p} \right)^{(n_s-1)} \quad (1.24)$$

with A_s the amplitude of the power spectrum, n_s the spectral index which governs the balance between small and large scales, with value close to 1, and k_p a pivot scale. ζ is called the *curvature perturbation* which corresponds, at early times, to the common perturbations for all components of the Universe. It is commonly normalised using the parameter σ_8 representing the root mean square (RMS) of the density contrast within spheres of radius $8 h^{-1} \text{Mpc}$, h the dimensionless Hubble constant (the Hubble constant is defined as $H_0 = 100h \text{ km.s}^{-1}.\text{Mpc}^{-1}$, see section 1.2.3).

The power spectrum can be linked to the cosmological density parameters and its measurement allows to apply constraints on them (see section 2.1.4). The correlation function of density fluctuations is, therefore, a powerful probe, and it can be measured using the different observable phenomenons that the Universe offers.

¹The horizon defines a causally connected region which size is set by the distance that a photon can have travelled in the time t since the Big Bang.

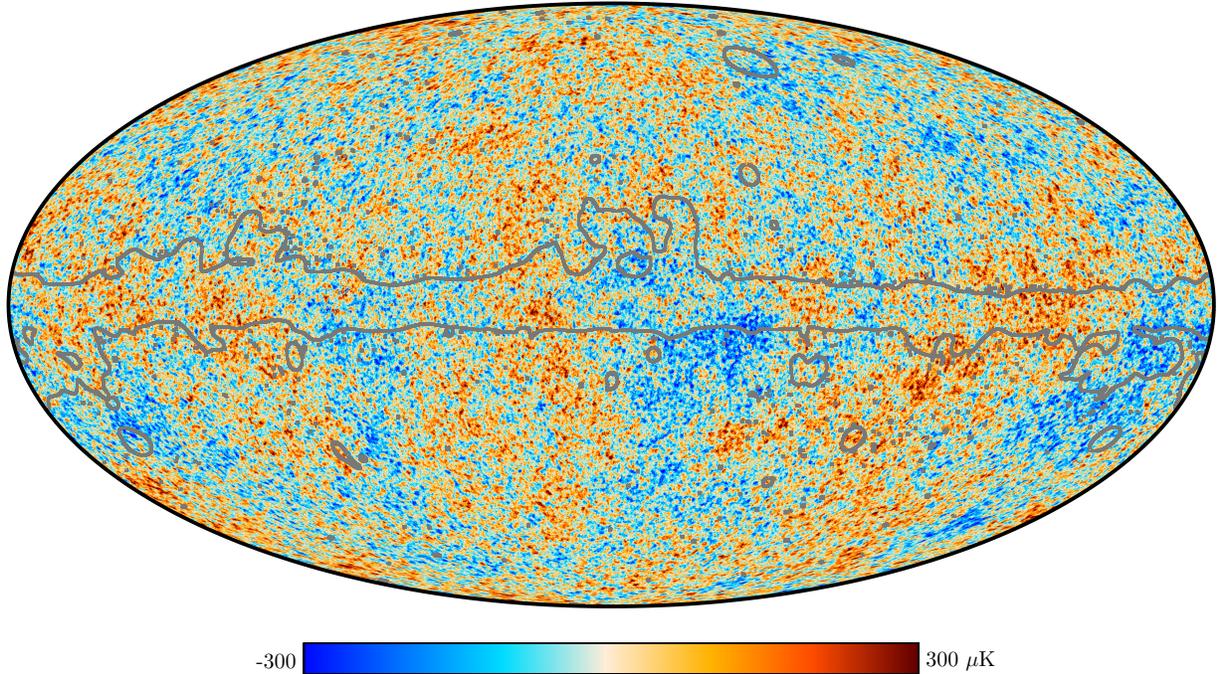


Figure 1.1: Map of the temperature anisotropies of the CMB, showing the density fluctuations in the early Universe. The scale shows the amplitude of these anisotropies around the mean value 2.725K. This map has been masked (grey lines) and inpainted where residuals from foreground emission are expected to be important. Credit: Planck Collaboration et al., 2020a

1.2 Different probes for different epochs

Using eq. (1.14) combined with eq. (1.8) and eq. (1.9), allows to describe the three major era of our Universe for three different values of w :

- the radiation-dominated era for $w = 1/3$: the energy decreases as $\rho(a) \propto a^{-4}$ as photons are diluted with the expansion of the Universe. The scale factor goes as $a(t) \propto \sqrt{t}$.
- the matter-dominated era for $w = 0$: matter starts dominating radiation density around 60 000 years after the Big Bang. The energy density goes as $\rho(a) \propto a^{-3}$ and the scale factor increases as $a(t) \propto t^{2/3}$
- the dark energy-dominated era for $w = -1$: it started about 9 billions years after the Big Bang. The expansion becomes exponential with $a(t) \propto e^{H_0 t}$

Consequently, in theory the expansion rate evolved through times, from the radiation-dominated era to the dark energy-dominated era. To probe those different times, several physical processes produced observable phenomenons that can be used to constrain cosmological parameters. I will now briefly describes few of the main probes that are used nowadays.

1.2.1 Early Universe: the cosmic microwave background

The large scale structures that can be observed today, such as galaxies and galaxy clusters for example, result partially from the propagation of small density perturbations occurring at early

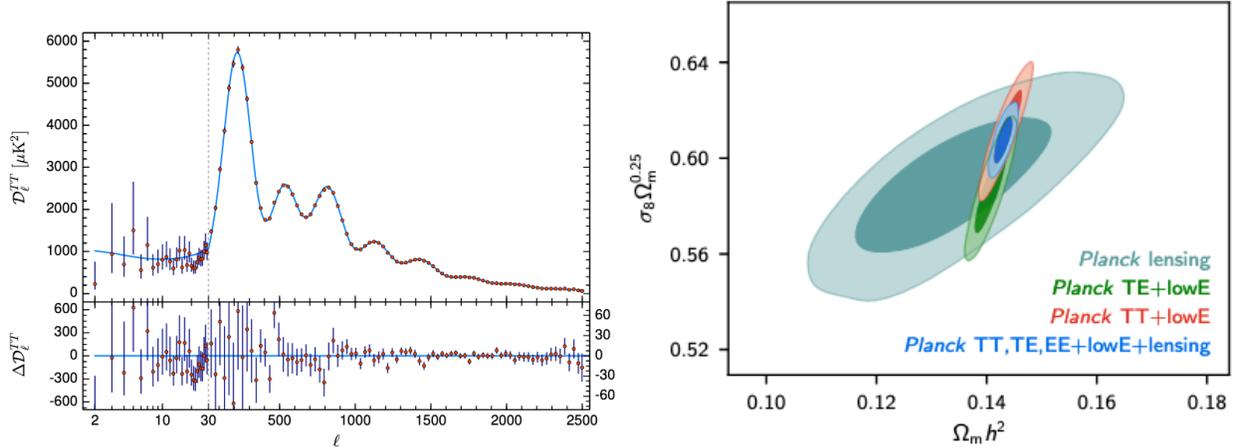


Figure 1.2: (left) Temperature power spectrum obtained by Planck 2018. The flat Λ CDM theoretical best fit to Planck data is shown in blue in the upper panel. Lower panel shows the residuals. (right) Contours obtained using a flat Λ CDM model in the $(\Omega_m, \sigma_8\Omega_m^{0.25})$ plane. Planck TE (Temperature - E-modes), TT (Temperature-Temperature) and lensing likelihoods overlap and the combination give tight constraints. Credit: Planck Collaboration et al., [2020b](#)

times in the Universe. At the beginning of the matter era, 380 000 years after the Big Bang, photons and baryonic matter decoupled as the Universe was expanding. From there, photons started following null geodesics, forming the first electromagnetic radiation in the Universe, a radiation bath called the cosmic microwave background (CMB). They were emitted at a temperature of about 3 000K and reach us at $T_0 = 2.725\text{K}$, leading to $z_{\text{CMB}} \approx 1090$.

The CMB represents a map of the matter density at decoupling, with the density perturbations imprinted on it. These density anisotropies translate in temperature anisotropies of the order of 10^{-5} (see fig. 1.1) and are carrying a lot of information about the primordial Universe.

Most recent results come from the Planck Collaboration (Planck Collaboration et al., [2020b](#)). The left panel of fig. 1.2 shows the power spectrum computed from the auto-correlation function of the temperature, in excellent agreement with the flat Λ CDM model theoretical best fit (in blue). The small discrepancies of the fit with data at small scales are due to cosmic variance (the limited number of modes at low l), preventing to precisely constrain models. The CMB provides the tightest constraints on cosmological parameters today.

Fluctuations in the baryon-photon plasma generated acoustic waves that froze at recombination (baryon-photon decoupling). The wavefronts created over-densities of baryons and photons, followed by under-dense regions. The combination of these density modes at recombination created peaks in the power spectrum. The oscillations correspond to a defined *acoustic angular scale* on the sky $\theta_* = r_*/D_M$, with r_* the sound horizon at recombination and D_M the comoving angular diameter distance. Planck measured the acoustic angular scale to be $100\theta_* = 1.04097 \pm 0.0004$. This value can be compared to the results obtained with *baryonic acoustic oscillations* (BAO, see next section and left pannel of fig. 1.3).

In addition to the temperature, the CMB provides information about the polarization of light. Light is polarized into E and B modes and the E-modes auto-correlation, or its cross-correlations with temperature, can help tighten cosmological constraints as shown on the right panel of fig. 1.2. B modes amplitude is below the systematics of Planck and have not been discovered yet. The search for these modes is a major challenge of nowadays cosmology and most of future CMB surveys such

as Simons observatory (Ade et al., 2019), Qubic (Hamilton et al., 2020) or POLARBEAR-2 (Suzuki et al., 2016) are looking for them. The discovery of these modes would indicate the presence of primordial gravitational waves, another clue in favour of inflation and of the standard cosmological model.

Finally, travelling along geodesics from the last scattering surface, photons from the CMB are *lensed* (see chapter 2) by the large scale structures in the Universe and one can use CMB lensing as a probe of matter distribution (Planck Collaboration et al., 2020c). As shown on the right panel of fig. 1.2, it provides additional information to the temperature and E-modes correlation functions and can be combined with them to decrease the likelihood surface.

1.2.2 Large Scale Structures

With the discovery of CMB’s B-modes, learning about the nature of dark energy is the other major challenge of nowadays’ cosmology. In that sense, the Dark Energy Task Force (DETF) (Albrecht et al., 2006) defined a Figure of Merit (FoM) in order to quantify improvements and set requirements for future surveys that will try to constrain the dark energy equation of state. It is defined as a function of the dark energy parameters as

$$FoM = \frac{1}{\sigma(w_0) \times \sigma(w_a)}. \quad (1.25)$$

The FoM increases as the confidence limits in the (w_0, w_a) plane decrease. Surveys with higher FoM consequently give tighter constraints on the dark energy parameters. For stage IV surveys, surveys that will start in the next few years (LSST, Euclid, Roman), the DETF set the goal of multiplying by ten the FoM compared to stage II surveys, surveys which have ended (CHFTLens, SDSS-II, etc), to reach a value above 500 (for a list of surveys for each stage, see appendix X of Albrecht et al., 2006). However, note that this figure of merit does not allow for the discrimination between all dark energy models, and is particularly not suited to test other hypotheses about dark energy such as modified gravity models for example. It is used here as the simplest Figure of Merit able to point out a divergence from the Λ CDM model and an evolution of the dark energy parameters through time.

Baryonic acoustic oscillations

As explained before, the over-densities produced by acoustic oscillations in the baryon-photon plasma, or *baryonic acoustic oscillations* (BAO), froze at recombination. From this point, photons follow null geodesics while baryons slowly interact with dark matter through gravity until they completely match at low redshift. The temperature power spectrum of the CMB permits to compute the acoustic angular scale, resulting from the preferred scale at which acoustic waves froze.

Since it represents the preferred scale of both temperature and matter density fluctuations at recombination, tracers of the large-scale structures (LSS) should also exhibit a BAO peak in their correlation functions. A scale, corresponding to the CMB θ_* , called $\theta_{drag} = r_{drag}/D_M$ with r_{drag} the comoving sound horizon at the end of the baryonic-drag epoch², can be determined using the galaxy-galaxy correlation function, a method called *galaxy clustering*. This scale can be used as a standard ruler, i.e. a scale that can be compared to the size of the detected BAO peak providing a measurement of cosmological distance. It can be related to the values of the expansion rate, $H(z)$, and to the angular diameter distance $d_A(z) = f_K(\chi(z))/(1+z)$ via the comoving scales measured

²The end of the baryonic-drag epoch is the moment at which the baryons decouple from the drag of the photons, i.e. the baryon optical depth is equal to one.

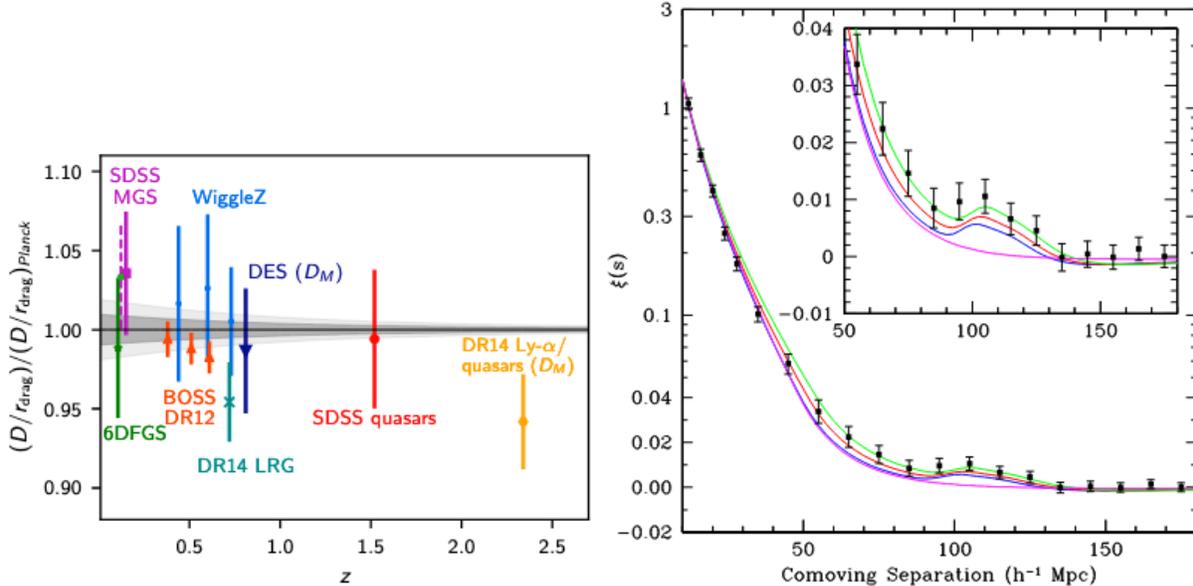


Figure 1.3: (left) Acoustic scale distance measurements. Measurements of each surveys are added with their 1σ error bars as a function of redshift. The grey bands show the 68% and 95% confidence ranges allowed by the combination of Planck TT+TE+EE+lowE+lensing. Credit: Planck Collaboration et al., 2020b. (right) BAO peak detection: measurement of the galaxy-galaxy correlation function as a function of the angular separation. Credit: Eisenstein et al., 2005

along the line of sight and perpendicular to it: $r_{drag\perp} = \theta_{drag} d_A(z)(1+z)$ and $r_{drag\parallel} = \frac{c\Delta z}{H(z)}$. BAOs can consequently probe the expansion rate at different redshifts.

Different tracers of LSS have been used to measure BAOs. It was first detected using galaxy surveys with the Sloan Digital Sky Survey (SDSS, Eisenstein et al., 2005) and 2dF Galaxy Redshift Survey (Cole et al., 2005). The measurement with galaxies at low redshift has been extended to higher redshift using quasars³ (from the extended Baryon Oscillation Survey (eBOSS) for example) and Lyman- α ⁴ spectra of quasars. Planck Collaboration et al. (2020b) present a summary of these measurements at different redshifts and compare them with the prediction from the CMB power spectrum (see fig. 1.3). BAOs obtained from galaxy surveys, quasars and Lyman- α spectra of quasars are in good agreement with the Planck results. The apparent discrepancy between Planck and the Lyman- α spectra of quasars (in orange) is not large enough to conclude to a tension. These results are consistent within 1.7σ (de Sainte Agathe et al., 2019).

Constraints on the dark energy parameters could be improved and next generation of both galaxy and spectroscopic surveys will allow to increase BAO measurements precision. LSST (see section 1.3) will provide large amount of photometric data, expecting a FoM of 14 after 10 years, a nearly 20 times improvement compared to stage III surveys (The LSST Dark Energy Science Collaboration et al., 2018). Euclid is also optimised for BAOs and will provide photometric and spectroscopic data, especially measuring the spectroscopic redshift of 50 millions of galaxies in the redshift range $0.7 < z < 2.1$ (Laureijs et al., 2011). The Dark Energy Spectroscopic Instru-

³Quasars, for quasi-stellar astronomical radiosources, are extremely luminous active galactic nuclei (AGN) with a supermassive black hole in their center. Their large intrinsic luminosity allows to acquire their spectra at high redshift. eBOSS for example, measured the spectra of about 150 000 quasars (Ata et al., 2018).

⁴The absorption of the light from quasars at wavelength corresponding to the rest-frame energy transition of neutral hydrogen in the interstellar medium.

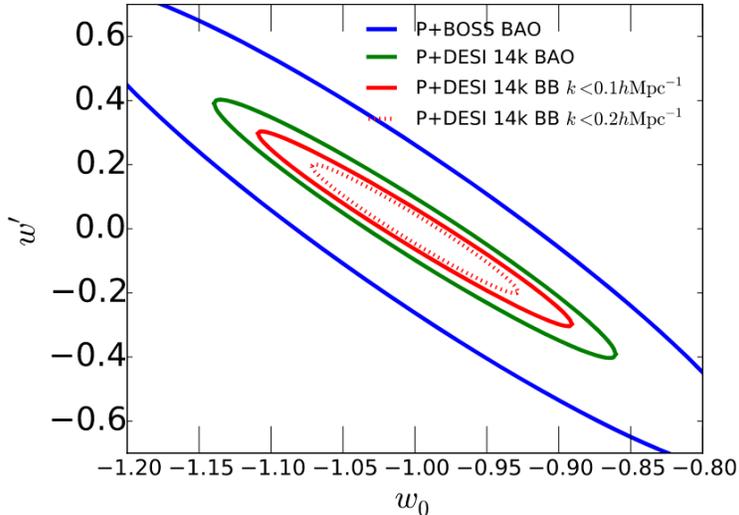


Figure 1.4: Constraints on the (w_0, w') plane of the BOSS measurements and of expected DESI results, including Planck priors. The figure shows the 68% limits of the contours. Credit: DESI Collaboration et al., 2016

ment (DESI, Levi et al., 2019) should measure the spectra of 20 to 30 millions of galaxies and quasars at redshifts $0.5 < z < 3.5$, drastically increasing the number of spectra available currently (BOSS+eBOSS produced 2 millions of spectra in 10 years). This should allow DESI to improve the FoM from more than a factor 3 using BAOs, more than 4 adding Lyman- α forest, and close to 9 including redshift space distortion in the plane (w_0, Ω_k) (DESI Collaboration et al., 2016) (see fig. 1.4). Redshift space distortion is the perturbation of galaxy clustering along the line of sight, due to peculiar velocities of galaxies coming from attraction by large scale structures. This anisotropies can be measured in redshift space and give information about the growth of structure (Reyes et al., 2010).

BAOs are very efficient tracers of the distribution of matter and are expected to apply tight constraints on the dark energy parameters. To tighten even more those constraints, one can combine them with other tracers of the distribution of matter such as gravitational lensing.

Gravitational lensing

Gravitational lensing provides a complementary measurement to the one of BAOs. When the latter traces the distribution of galaxies, weak gravitational lensing probes the complete distribution of matter (dark and baryonic matter) through the Universe. This probe is at the core of this thesis, and will be largely discussed in chapter 2. To summarize, the path traveled by light coming from a background galaxy is modified by the foreground matter. Since photons travel following null geodesics (see eq. (1.2)), the image of the background galaxy is distorted. The gravitational potential acts similarly to an optical lens, that is why this effect is called *lensing*. Two lensing regimes exist: *strong* and *weak* lensing.

The statistics of strong lenses is not large enough nowadays to provide tight constraints on cosmological parameter but first studies already showed that such an analysis can be performed (Wong et al., 2020). It is improved by current galaxy surveys like DES and deep learning algorithms are used to find potential or definite candidates to extend strong lenses catalogs (Jacobs et al., 2019). Using next generation of surveys such as LSST, it is expected to yield a FoM of 9.4 (The LSST

Dark Energy Science Collaboration et al., 2018).

On the contrary, weak lensing has already shown extremely promising results and is a major goal for next generation surveys. It will be much more constraining than strong lensing for dark energy parameters. The Dark Energy Survey just released its "year 3" analysis (DES-Y3) providing tight constraints on cosmological parameters, especially using the 3×2 -point correlation function (improving by more than a factor of 3 the FoM obtained from the shear-shear correlation function, see section 2.3.1). The Rubin Observatory Legacy Survey of Space and Time (LSST, see section 1.3.1 and Abell et al., 2009), Euclid (see section 1.3.2 and Refregier et al., 2010) and the Nancy Grace Roman Space Telescope (Roman, see Spergel et al., 2015) are specifically designed to make the most of this probe. LSST expects a FoM of 52 on dark energy parameters (The LSST Dark Energy Science Collaboration et al., 2018) using weak lensing while Euclid could go up to 180 (Refregier et al., 2010), combining imaging and spectroscopic data (see section 1.3). Since this thesis was done within the LSST Dark Energy Science Collaboration (DESC) and that Euclid-like data are also used in chapter 7, the main characteristics of those two surveys are described in section 1.3.

Clusters

In addition to these measurements, it is possible to use clusters of galaxies, and in particular cluster abundance (number counts of galaxy clusters), to constrain cosmology. Being the largest structures in the Universe, clusters are sensitive to both the expansion history of the Universe and the growth of structures. Clusters lay in dark matter halos, which abundance at a given redshift, and as a function of mass, the *halo mass function*, depends on the density of cosmological parameters. Measuring the mass and redshift of the detected clusters can hence provide constraints on these parameters.

The main difficulty resides in the measurement of the halo masses from the different observable (richness, i.e. count of member galaxies, intensity of the thermal Sunyaev-Zeldovich (SZ) signal, or X-ray luminosity for example). For photometric surveys for example, the mass of clusters can be measured using the mass-richness relationship, which needs to be calibrated. Weak lensing is expected to give a precise calibration of this relationship, as it probes both baryonic and dark matter, and does not depend on the thermodynamic state of the clusters as X-rays or the SZ effect. However, it also includes systematics (see section 2.5) that needs to be taken into account.

It has already been used to apply cosmological constraints with the data from SDSS (Costanzi et al., 2019) or from DES-Y1 (T. M. C. Abbott et al., 2020). The latter presents a tension with other measurements from Planck or the 3×2 -pt analysis from DES-Y1, but it has been found to be due to systematics (Costanzi et al., 2021), namely the modeling of the weak lensing signal.

Stage IV photometric surveys, LSST for example, will produce large catalogs of clusters. Combined with other large scale structure surveys (SKA (Square Kilometre Array Cosmology Science Working Group et al., 2020), or eROSITA (Predehl et al., 2021) for example), it will allow to decrease the impact of systematics on measurement, and to precisely calibrate the observable-mass cluster relationship using gravitational lensing. LSST expects, with cluster alone, to reach a FoM of 22, which will significantly contribute to the joint-probes analysis as shown on fig. 1.10.

However, BAOs, lensing and clusters are not the only tracers of the distribution of matter. In the next section, I briefly present the role of Supernovae cosmology and the tension resulting from the measurement of the Hubble constant using them.

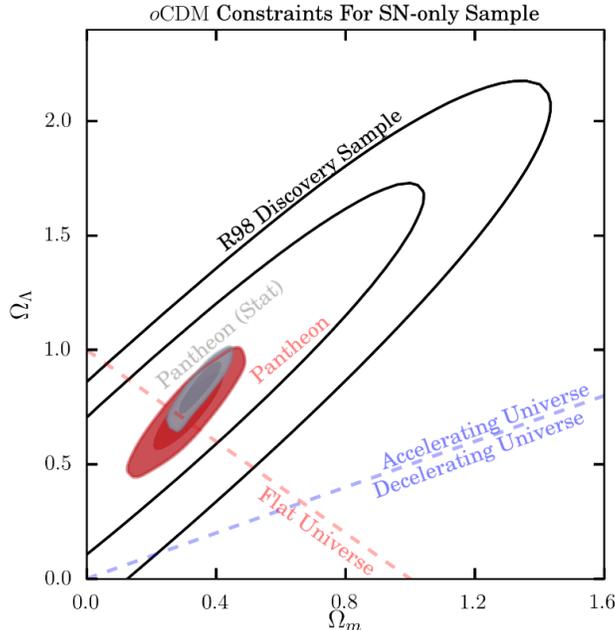


Figure 1.5: Contours at 68% and 95% of SN only constraints for a non-flat Λ CDM model for the Pantheon sample (combining systematics and statistical uncertainties in red, with only statistical uncertainties in grey) and from the sample used to discover the acceleration of the Universe’s expansion (contours in black) (Riess et al., 1998). Credit: Scolnic et al., 2018

1.2.3 Supernovae and the H_0 tension

Supernovae cosmology

Historically, the first probe to demonstrate the accelerated expansion of the Universe is type Ia Supernovae (SNIa) (Perlmutter et al., 1999; Riess et al., 1998). Measuring the relation between luminosity distance and redshift obtained by analysis of the SNIa light curves permits to constrain cosmological parameters. To this effect, SNIa are considered to be standardisable candles, i.e. objects that have the same intrinsic luminosity. Their large intrinsic luminosity allows to probe redshift up to about 2. The Pantheon sample (Scolnic et al., 2018) combined measurements from four different surveys (HST, SDSS, the SuperNova Legacy Survey, SNLS, and Panoramic Survey Telescope and Rapid Response System, Pan-STARRS1) to yield a little bit more than a thousand spectroscopically confirmed light curves. Combining these data with BAOs from SDSS and CMB from Planck yields a FoM of 65 in the (w_0, w_a) plane. Figure 1.5 shows the constraints on the $(\Omega_\Lambda, \Omega_m)$ plane for a non-flat model. Once again, results are consistent with the Λ CDM model.

The future of Supernovae cosmology lies in the utilisation of large field of view galaxy surveys to measure a large number of high-quality light curves. It already benefits from DES (T. M. C. Abbott et al., 2019) and the future LSST, with its large field of view, its six filters and the repeated visits at a each location of the southern sky, is expected to provide a catalogue of around 300 000 SNIa light curves (Abell et al., 2009). Providing a correct classification method, since spectroscopic confirmation will not be possible for such a large sample, these measurements could lead to a FoM of 211 for SNIa only on dark energy parameters (The LSST Dark Energy Science Collaboration et al., 2018). Supernovae is the largest contributor in the increase of the FoM for LSST, providing correct light curve classification. In that regard, promising deep learning methods were proposed, as in the

FINK broker (Möller et al., 2020) developed within the LSST Dark Energy Science Collaboration (DESC).

H_0 tension

SNIa are particularly promising to constrain dark energy equation of state but they can also be used to constrain other cosmological parameters such as H_0 .

To use SNIa as standard candles they need to be calibrated, which can be done using Cepheids. Riess et al. (2019) used the ones of the Large Magellan Cloud (LMC) and measured H_0 to be equal to $74.03 \pm 1.42 \text{ km s}^{-1} \text{ Mpc}^{-1}$, leading to a 4.4σ -tension with the value inferred from the Planck CMB and Λ CDM (Planck Collaboration et al., 2020b). Figure 1.6 shows the discrepancy between the measurements from the early Universe (CMB), and from the late Universe (Cepheids, Supernovae). Several re-analysis and separated measurements (with strong lenses for example, see below) confirmed this tension and no answer has been provided to explain it yet. It leads to two possibilities: either it comes from the systematics, for example in the process of determining the flux of the SNIa by subtracting the flux of the host galaxy (Astier et al., 2013), or more interestingly, from new physics.

The answer will probably come from independent measurements. As we have seen, the measurement of BAOs will drastically improve our knowledge of the expansion rate $H(z)$, especially thanks to DESI. As shown on fig. 1.6, results from DESI should permit to constrain $H(z)$ on a wide range of redshifts, permitting to precisely infer the value of H_0 (DESI Collaboration et al., 2016).

More exotic but quite interesting too, Strong Lensing Time Delay (SLTD), measuring the time delay between multiple images of strongly lensed quasars, are expected to be able to constrain H_0 . The H0LiCOW collaboration already provided a combined measurement with six events presenting a 5.3σ -tension with Planck (Wong et al., 2020). Those result are obviously preliminary but promising, and more events are required to yield a reliable value. As mentioned galaxy surveys as DES are already increasing statistics (Jacobs et al., 2019) and LSST is expected to provide about 400 time delay measurements over its ten years of operation (Abell et al., 2009).

An even more rare process, only one has been recorded yet, is the detection of gravitational waves followed by the detection of its electromagnetic counterpart. GW170817 paved the way for multi-messenger astronomy with a gravitational waves (GW) signal detected by both LIGO (LIGO Scientific Collaboration et al., 2015) and Virgo (Acernese et al., 2015) from the merger of a binary neutron-star system. Two seconds after, a gamma-ray burst was detected in a region of the sky consistent with the GW detection. Astronomical facilities then observed this region finding a transient signal close from the galaxy NGC 4993. Using the gravitational waves signal as a standard siren, it is possible to infer the distance from the GW signal and the redshift can be measured from its electromagnetic counterpart offering a measurement of the distance-redshift relationship, to constrain H_0 (B. P. Abbott et al., 2017) (see fig. 1.7). A single event is obviously not enough to provide significant constraints but with the development of LIGO and Virgo and with the future LISA GW detector (Amaro-Seoane et al., 2017), it can be expected that multi-messenger astronomy will become a independent and very interesting probe for H_0 measurement as well as for cosmology in general.

In conclusion, these are very exciting times for cosmology. Exciting discoveries await at each scale of the Universe with dedicated surveys that will soon start looking for them. The search for the B modes of the CMB will help improving our understanding of the inflation period, while BAOs, Supernovae and weak lensing will apply extremely tight constraints on the dark energy equation of

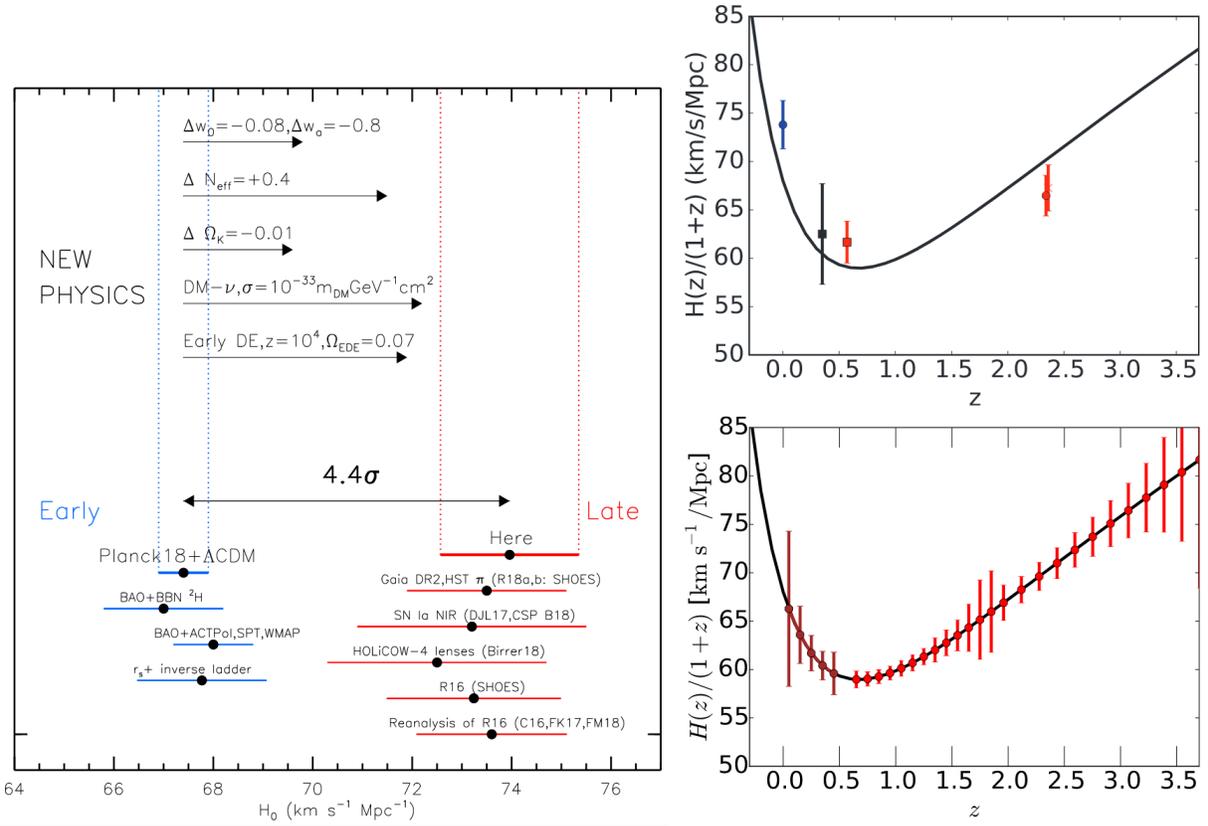


Figure 1.6: (left) The 4.4σ tension between measurements of H_0 using local measurements (mostly Cepheids and SNIa), "Here", and the measurements from Planck and Λ CDM, "Planck18+ Λ CDM". Other results are shown for comparison. Credit: Riess et al., 2019. (right) The expansion rate as a function of the redshift z . The upper panel shows the measurements from SDSS BAO measurement (black), the BOSS galaxy BAO measurement (red square), the BOSS Lyman- α measurement (red circle) and a H_0 measurement using Cepheids and SNIa (blue). The lower panel shows the expected DESI measurement points. Credit: DESI Collaboration et al., 2016

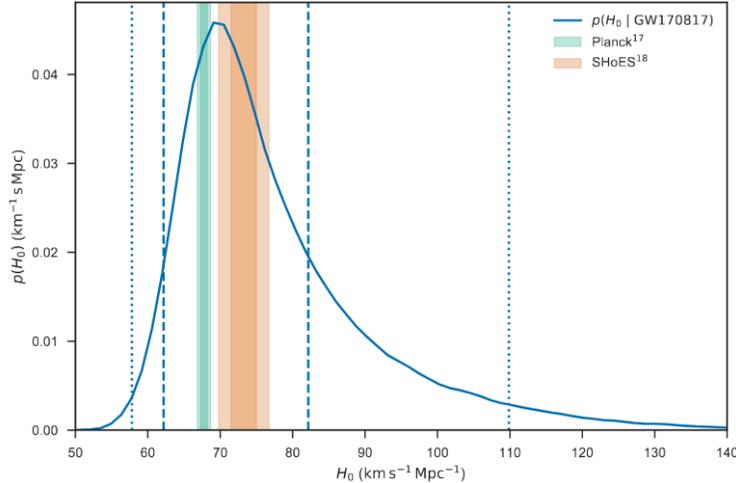


Figure 1.7: Marginal posterior density for H_0 obtained from the analysis of the GW170817 event. The constraints obtained by the SH0ES collaboration (Riess et al., 2016) and from Planck (Planck Collaboration et al., 2016) are shown for comparison. Credit: B. P. Abbott et al. (2017)

state, bringing us a step closer to understanding its nature. As dark energy represents about 70% of the energy budget of the Universe, learning about its nature is of primordial importance. Increased statistics for the already established Supernovae cosmology will be an interesting avenue to solve the H_0 tension but the answer might come from independent measurements, employing probes that became usable only recently thanks to improvements in detection technologies. Gravitational waves astronomy and particularly multi-messenger astronomy seem, to me, extremely promising and I look forward to new measurements to obtain precise constraints on H_0 .

In the following, I will focus on weak gravitational lensing which is the subject of this thesis. In the next section, I start by describing the LSST telescope, since this work was done in the context of the LSST DESC. Then I present the Euclid telescope, which can provide valuable information combined with LSST data for weak gravitational lensing analysis, as shown in chapter 7.

1.3 Next generation of surveys

1.3.1 LSST

Characteristics

The LSST, for Legacy Survey of Space and Time will use the Simonyi Survey Telescope located at the Vera C. Rubin Observaroty at Cerro Pachón in Chile. It will observe the sky through a 3-mirror system, with a 8.4m primary mirror, enabling a large 9.6deg^2 field of view. Such a large field of view will allow to visit the entire night sky (observable from the Southern Hemisphere) every three nights (see fig. 1.8). The 3.2 gigapixels camera is composed of 189 $4k \times 4k$ pixels CCDs and will observe the sky in six different photometric bands, *ugrizy*, in the wavelength range 320 to 1050 nm (see right panel of fig. 1.9). Observing the sky in six filters will provide important information for the estimation of photometric redshifts and the large field of view combined with recurrent visits at every southern sky location will enable the detection of a large number of transients, two important requirements in order to meet the goal set by the Dark Energy Task Force (see section 1.2.2).

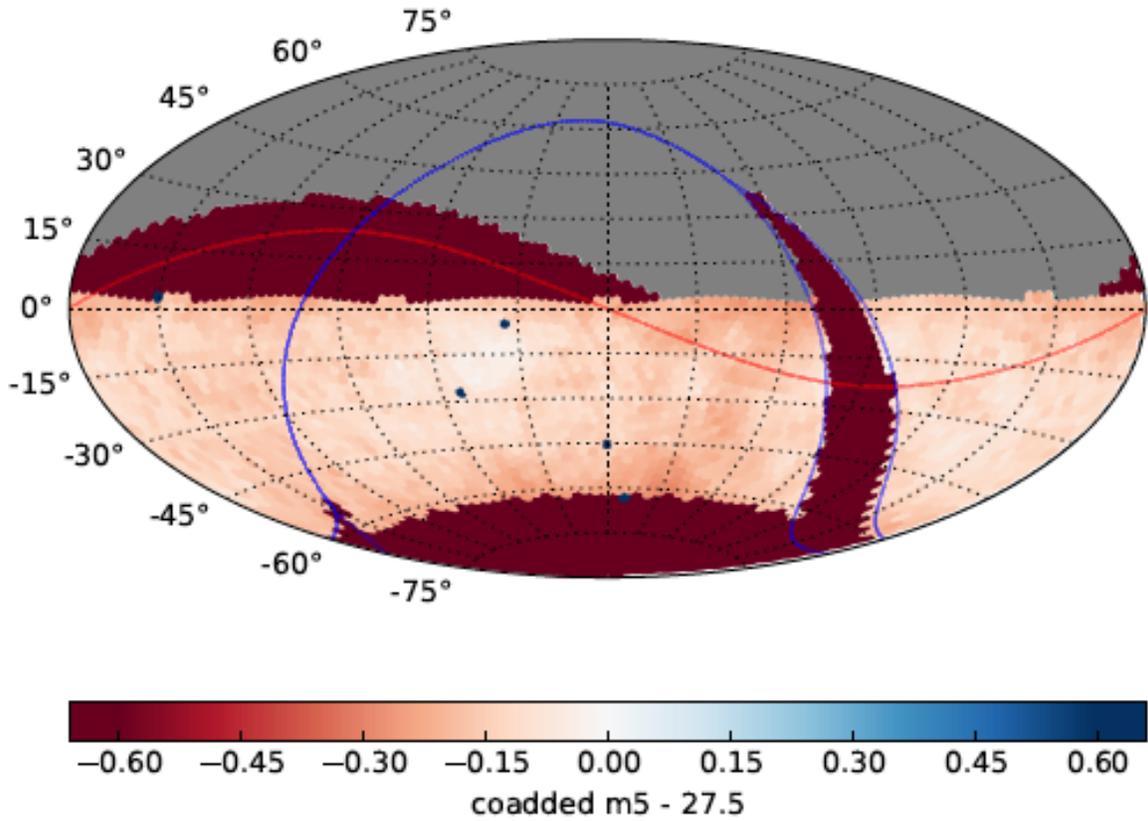
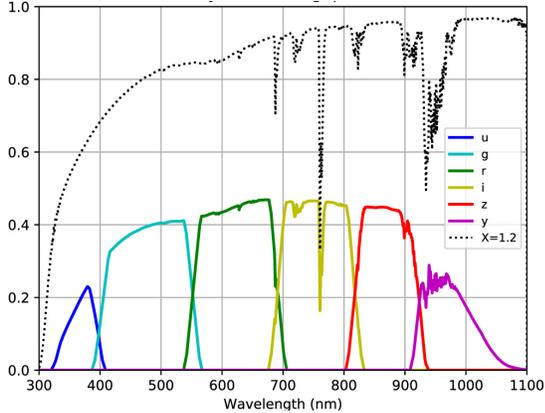


Figure 1.8: Map of the coadded 5σ depth magnitude for point sources in the r -band after ten years. This map is obtained via simulation. The red line is the Ecliptic and the blue one is the Galactic equator. The median value for the Wide, Fast, Deep (WFD) survey (light red) is 27.1. Credit: LSST Science Collaboration et al., [2017](#)



Quantity	Baseline Design Specification
Optical Config.	3-mirror modified Paul-Baker
Mount Config.	Alt-azimuth
Final f-ratio, aperture	f/1.234, 8.4 m
Field of view, étendue	9.6 deg ² , 319 m ² deg ²
Plate Scale	50.9 μm/arcsec (0.2" pix)
Pixel count	3.2 Gigapix
Wavelength Coverage	320 – 1050 nm, <i>ugrizy</i>
Single visit depths, design ^a	23.9, 25.0, 24.7, 24.0, 23.3, 22.1
Single visit depths, min. ^b	23.4, 24.6, 24.3, 23.6, 22.9, 21.7
Mean number of visits ^c	56, 80, 184, 184, 160, 160
Final (coadded) depths ^d	26.1, 27.4, 27.5, 26.8, 26.1, 24.9

Figure 1.9: (left) The six LSST bandpass as a function of wavelength. The vertical axis shows the total throughput. Results include the atmospheric transmission, optics, and the detector sensitivity. (right) LSST survey parameters. Values are set accordingly to the Science Requirements Document, and value for the eighth, ninth and last row are on the AB magnitude scale. Credit: Ivezić et al., 2019

Over its ten years of operation, starting in 2023, LSST is expected to survey $\approx 20\,000\text{ deg}^2$ of the sky visiting each part of the sky about 100 times each year. Each visit will be composed of 15-seconds exposures. This will enable LSST to look deep into the sky, up to a magnitude of ≈ 27.5 in the *r*-band. The main characteristics such as the number of visits and expected coadded depth in each filters, are detailed on the right panel of fig. 1.9 (extracted from Ivezić et al. (2019)).

Scientific objectives

LSST is a stage IV survey, and one of its four main scientific goals is to constrain dark energy. Answering to the DETF objective, the LSST Science Requirements Document (The LSST Dark Energy Science Collaboration et al., 2018) sets the survey requirements to multiply by ten the FoM compared to stage II surveys. To that end, LSST will use galaxy clustering, weak lensing and the 3×2 -point correlation function (see section 2.2.2), galaxy clusters, Supernovae and strong lensing. The contours obtained for the expected precision on the dark energy parameters for each of the five probes as well as the results of the joint analysis are shown on fig. 1.10. LSST is expected to yield a FoM of 711 including Stage III priors, with most of the constraints coming from the Supernovae and weak gravitational lensing correlated with galaxy clustering. The 3×2 -point correlation function alone will improve the stage III surveys (ongoing surveys) value by more than a factor of four, from 20 to 87, with a major contribution from the weak lensing correlation function (shear-shear, see section 2.2.1) which, if used alone, would set the FoM to 52.

1.3.2 Euclid

Euclid (Refregier et al., 2010) is a space telescope which will be launched in 2023. It will carry two instruments, a CCD based optical imaging channel in the VIS instrument, and a near-infrared imaging photometry channel and a near-infrared spectro-photometer in the NISP instrument. It is designed for weak lensing with a very broad visible band (R+I+Z) and a pixel scale of 0.1 arcsec for the VIS instrument (compared to 0.2 for LSST pixels). NISP is composed of three near-infrared (NIR) bands with a 0.3 arcsec pixel scale and a spectro-photometer of similar pixel size and resolution $R = 250$. It will be in operation for five years and survey $20\,000\text{ deg}^2$, i.e. the entire

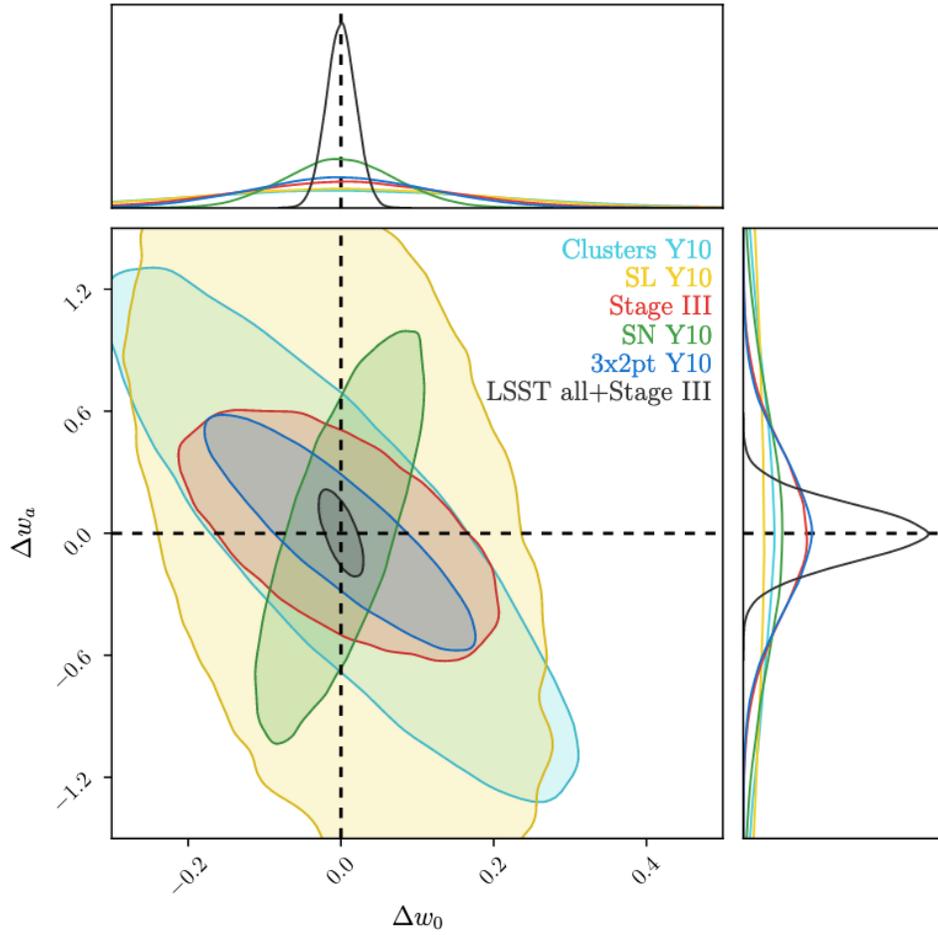


Figure 1.10: The forecast dark energy constraints after 10 years of LSST. The contours show the 68% confidence intervals. Δw_0 and Δw_a are the difference between w_0 and w_a and their fiducial values, -1 and 0 . These contours do not include stage III priors. Credit: The LSST Dark Energy Science Collaboration et al., [2018](#)

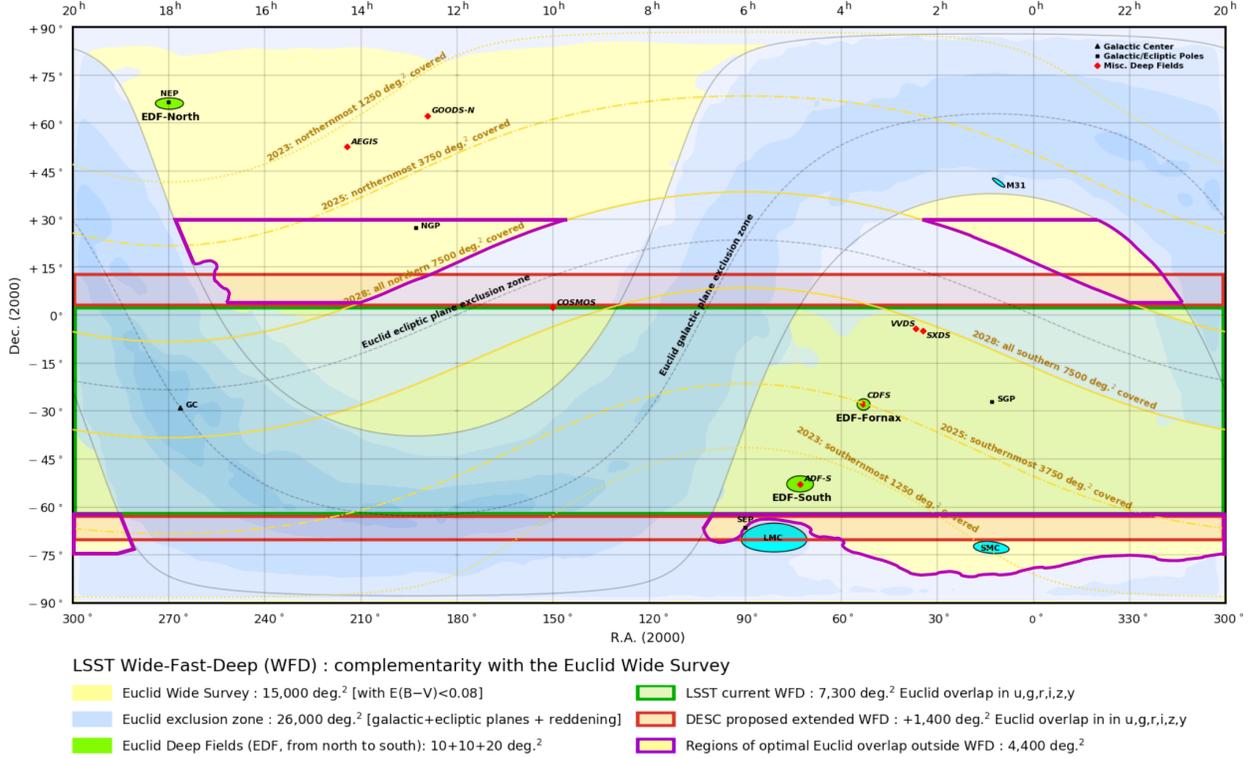


Figure 1.11: Sky coverage of Euclid and LSST surveys. Credit: Capak et al., 2019

extragalactic sky, up to a redshift of 2. The photometric redshifts (see section 2.2.4) of Euclid are supposed to be obtained using the three NIR filters but they will also require ground based data, coming for example from LSST.

The high resolution of the Euclid VIS instrument is a major asset for galaxy shape measurements. Euclid is expected to reach a FoM of 180 with weak lensing only, providing correct measurement of photometric redshifts. All imaging probes combined (adding clustering, clusters and integrated Sachs Wolfe effects), Euclid should be able to reach a FoM of 400 (Refregier et al., 2010). This value could be increased when combined with the results of the Euclid spectroscopic survey which will measure the spectroscopic redshift of 50 millions of galaxies to measure the BAOs. The combination of the imaging and spectroscopic Euclid surveys are expected to yield a FoM of about 1500 (Laureijs et al., 2011).

Interestingly Euclid and LSST sky coverage will overlap for several thousand square degrees (see fig. 1.11) which could allow to combine the photometric redshifts estimated from LSST data and the shape measurements from Euclid. Going further, a joint pixel analysis using data from both surveys can improve shape and flux reconstruction (Schuhmann, Heymans, and Zuntz, 2019; Arcelin et al., 2021), as it will be detailed in chapter 7. Benefiting from strength of both telescopes will increase the FoM of dark energy parameters by several tens of percent, depending of the survey strategy (Capak et al., 2019).

Weak lensing is consequently a leading probe for future surveys to meet requirements sets by the Dark Energy Task Force. In the next chapter, I detail the formalism of weak lensing, the important probes that will help constraining the parameters of dark energy, and how to use observations to

do so.

Bibliography

- Abbott, B. P. et al. (Nov. 2017). “A gravitational-wave standard siren measurement of the Hubble constant”. In: *Nature* 551.7678, pp. 85–88. DOI: [10.1038/nature24471](https://doi.org/10.1038/nature24471). arXiv: [1710.05835](https://arxiv.org/abs/1710.05835) [[astro-ph.CO](#)].
- Abbott, T. M. C. et al. (Feb. 2019). “First Cosmology Results using Type Ia Supernovae from the Dark Energy Survey: Constraints on Cosmological Parameters”. In: *ApJ* 872.2, L30, p. L30. DOI: [10.3847/2041-8213/ab04fa](https://doi.org/10.3847/2041-8213/ab04fa). arXiv: [1811.02374](https://arxiv.org/abs/1811.02374) [[astro-ph.CO](#)].
- Abbott, T. M. C. et al. (July 2020). “Dark Energy Survey Year 1 Results: Cosmological constraints from cluster abundances and weak lensing”. In: *Phys. Rev. D* 102.2, 023509, p. 023509. DOI: [10.1103/PhysRevD.102.023509](https://doi.org/10.1103/PhysRevD.102.023509). arXiv: [2002.11124](https://arxiv.org/abs/2002.11124) [[astro-ph.CO](#)].
- Abell, Paul A et al. (2009). “Lsst science book, version 2.0”. In: *arXiv preprint arXiv:0912.0201*.
- Acernese, F. et al. (Jan. 2015). “Advanced Virgo: a second-generation interferometric gravitational wave detector”. In: *Classical and Quantum Gravity* 32.2, 024001, p. 024001. DOI: [10.1088/0264-9381/32/2/024001](https://doi.org/10.1088/0264-9381/32/2/024001). arXiv: [1408.3978](https://arxiv.org/abs/1408.3978) [[gr-qc](#)].
- Ade, Peter et al. (Feb. 2019). “The Simons Observatory: science goals and forecasts”. In: *J. Cosmology Astropart. Phys.* 2019.2, 056, p. 056. DOI: [10.1088/1475-7516/2019/02/056](https://doi.org/10.1088/1475-7516/2019/02/056). arXiv: [1808.07445](https://arxiv.org/abs/1808.07445) [[astro-ph.CO](#)].
- Albrecht, Andreas et al. (Sept. 2006). “Report of the Dark Energy Task Force”. In: *arXiv e-prints*, astro-ph/0609591, astro-ph/0609591. arXiv: [astro-ph/0609591](https://arxiv.org/abs/astro-ph/0609591) [[astro-ph](#)].
- Amaro-Seoane, Pau et al. (Feb. 2017). “Laser Interferometer Space Antenna”. In: *arXiv e-prints*, arXiv:1702.00786, arXiv:1702.00786. arXiv: [1702.00786](https://arxiv.org/abs/1702.00786) [[astro-ph.IM](#)].
- Arcelin, Bastien et al. (Jan. 2021). “Deblending galaxies with variational autoencoders: A joint multiband, multi-instrument approach”. In: *MNRAS* 500.1, pp. 531–547. DOI: [10.1093/mnras/staa3062](https://doi.org/10.1093/mnras/staa3062). arXiv: [2005.12039](https://arxiv.org/abs/2005.12039) [[astro-ph.IM](#)].
- Astier, P. et al. (Sept. 2013). “Photometry of supernovae in an image series: methods and application to the SuperNova Legacy Survey (SNLS)”. In: *A&A* 557, A55, A55. DOI: [10.1051/0004-6361/201321668](https://doi.org/10.1051/0004-6361/201321668). arXiv: [1306.5153](https://arxiv.org/abs/1306.5153) [[astro-ph.IM](#)].
- Ata, Metin et al. (Feb. 2018). “The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2”. In: *MNRAS* 473.4, pp. 4773–4794. DOI: [10.1093/mnras/stx2630](https://doi.org/10.1093/mnras/stx2630). arXiv: [1705.06373](https://arxiv.org/abs/1705.06373) [[astro-ph.CO](#)].
- Baumann, Daniel (July 2018). “TASI Lectures on Primordial Cosmology”. In: *arXiv e-prints*, arXiv:1807.03098, arXiv:1807.03098. arXiv: [1807.03098](https://arxiv.org/abs/1807.03098) [[hep-th](#)].
- Capak, P. et al. (Apr. 2019). “Enhancing LSST Science with Euclid Synergy”. In: *arXiv e-prints*, arXiv:1904.10439, arXiv:1904.10439. arXiv: [1904.10439](https://arxiv.org/abs/1904.10439) [[astro-ph.IM](#)].
- Chevallier, Michel and David Polarski (Jan. 2001). “Accelerating Universes with Scaling Dark Matter”. In: *International Journal of Modern Physics D* 10.2, pp. 213–223. DOI: [10.1142/S0218271801000822](https://doi.org/10.1142/S0218271801000822). arXiv: [gr-qc/0009008](https://arxiv.org/abs/gr-qc/0009008) [[gr-qc](#)].

- Cole, Shaun et al. (Sept. 2005). “The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications”. In: MNRAS 362.2, pp. 505–534. DOI: [10.1111/j.1365-2966.2005.09318.x](https://doi.org/10.1111/j.1365-2966.2005.09318.x). arXiv: [astro-ph/0501174](https://arxiv.org/abs/astro-ph/0501174) [astro-ph].
- Cooray, Asantha R. and Dragan Huterer (Mar. 1999). “Gravitational Lensing as a Probe of Quintessence”. In: ApJ 513.2, pp. L95–L98. DOI: [10.1086/311927](https://doi.org/10.1086/311927). arXiv: [astro-ph/9901097](https://arxiv.org/abs/astro-ph/9901097) [astro-ph].
- Costanzi, M. et al. (Oct. 2019). “Methods for cluster cosmology and application to the SDSS in preparation for DES Year 1 release”. In: MNRAS 488.4, pp. 4779–4800. DOI: [10.1093/mnras/stz1949](https://doi.org/10.1093/mnras/stz1949). arXiv: [1810.09456](https://arxiv.org/abs/1810.09456) [astro-ph.CO].
- Costanzi, M. et al. (Feb. 2021). “Cosmological constraints from DES Y1 cluster abundances and SPT multiwavelength data”. In: Phys. Rev. D 103.4, 043522, p. 043522. DOI: [10.1103/PhysRevD.103.043522](https://doi.org/10.1103/PhysRevD.103.043522). arXiv: [2010.13800](https://arxiv.org/abs/2010.13800) [astro-ph.CO].
- de Sainte Agathe, Victoria et al. (Sept. 2019). “Baryon acoustic oscillations at $z = 2.34$ from the correlations of Ly α absorption in eBOSS DR14”. In: A&A 629, A85, A85. DOI: [10.1051/0004-6361/201935638](https://doi.org/10.1051/0004-6361/201935638). arXiv: [1904.03400](https://arxiv.org/abs/1904.03400) [astro-ph.CO].
- DES Collaboration et al. (May 2021). “Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing”. In: *arXiv e-prints*, arXiv:2105.13549, arXiv:2105.13549. arXiv: [2105.13549](https://arxiv.org/abs/2105.13549) [astro-ph.CO].
- DESI Collaboration et al. (Oct. 2016). “The DESI Experiment Part I: Science, Targeting, and Survey Design”. In: *arXiv e-prints*, arXiv:1611.00036, arXiv:1611.00036. arXiv: [1611.00036](https://arxiv.org/abs/1611.00036) [astro-ph.IM].
- Dodelson, Scott (2003). *Modern cosmology*. Elsevier.
- Eisenstein, Daniel J. et al. (Nov. 2005). “Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies”. In: ApJ 633.2, pp. 560–574. DOI: [10.1086/466512](https://doi.org/10.1086/466512). arXiv: [astro-ph/0501171](https://arxiv.org/abs/astro-ph/0501171) [astro-ph].
- Hamilton, J. -Ch. et al. (Nov. 2020). “QUBIC I: Overview and Science Program”. In: *arXiv e-prints*, arXiv:2011.02213, arXiv:2011.02213. arXiv: [2011.02213](https://arxiv.org/abs/2011.02213) [astro-ph.IM].
- Ivezić, Željko et al. (Mar. 2019). “LSST: From Science Drivers to Reference Design and Anticipated Data Products”. In: ApJ 873.2, 111, p. 111. DOI: [10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c). arXiv: [0805.2366](https://arxiv.org/abs/0805.2366) [astro-ph].
- Jacobs, C. et al. (July 2019). “An Extended Catalog of Galaxy-Galaxy Strong Gravitational Lenses Discovered in DES Using Convolutional Neural Networks”. In: ApJS 243.1, 17, p. 17. DOI: [10.3847/1538-4365/ab26b6](https://doi.org/10.3847/1538-4365/ab26b6). arXiv: [1905.10522](https://arxiv.org/abs/1905.10522) [astro-ph.GA].
- Laureijs, R. et al. (Oct. 2011). “Euclid Definition Study Report”. In: *arXiv e-prints*, arXiv:1110.3193, arXiv:1110.3193. arXiv: [1110.3193](https://arxiv.org/abs/1110.3193) [astro-ph.CO].
- Levi, Michael et al. (Sept. 2019). “The Dark Energy Spectroscopic Instrument (DESI)”. In: *Bulletin of the American Astronomical Society*. Vol. 51, p. 57. arXiv: [1907.10688](https://arxiv.org/abs/1907.10688) [astro-ph.IM].
- LIGO Scientific Collaboration et al. (Apr. 2015). “Advanced LIGO”. In: *Classical and Quantum Gravity* 32.7, 074001, p. 074001. DOI: [10.1088/0264-9381/32/7/074001](https://doi.org/10.1088/0264-9381/32/7/074001). arXiv: [1411.4547](https://arxiv.org/abs/1411.4547) [gr-qc].
- Linder, Eric V. (Mar. 2003). “Exploring the Expansion History of the Universe”. In: Phys. Rev. Lett. 90.9, 091301, p. 091301. DOI: [10.1103/PhysRevLett.90.091301](https://doi.org/10.1103/PhysRevLett.90.091301). arXiv: [astro-ph/0208512](https://arxiv.org/abs/astro-ph/0208512) [astro-ph].
- LSST Science Collaboration et al. (Aug. 2017). “Science-Driven Optimization of the LSST Observing Strategy”. In: *arXiv e-prints*, arXiv:1708.04058, arXiv:1708.04058. arXiv: [1708.04058](https://arxiv.org/abs/1708.04058) [astro-ph.IM].
- Möller, Anais et al. (Nov. 2020). “FINK, a new generation of broker for the LSST community”. In: MNRAS. DOI: [10.1093/mnras/staa3602](https://doi.org/10.1093/mnras/staa3602). arXiv: [2009.10185](https://arxiv.org/abs/2009.10185) [astro-ph.IM].
- Peacock, John A (1999). *Cosmological physics*. Cambridge university press.

- Perlmutter, S. et al. (June 1999). “Measurements of Ω and Λ from 42 High-Redshift Supernovae”. In: ApJ 517.2, pp. 565–586. DOI: [10.1086/307221](https://doi.org/10.1086/307221). arXiv: [astro-ph/9812133](https://arxiv.org/abs/astro-ph/9812133) [astro-ph].
- Peter, Patrick and Jean-Philippe Uzan (2013). *Primordial cosmology*. Oxford University Press.
- Planck Collaboration et al. (Sept. 2016). “Planck 2015 results. XIII. Cosmological parameters”. In: A&A 594, A13, A13. DOI: [10.1051/0004-6361/201525830](https://doi.org/10.1051/0004-6361/201525830). arXiv: [1502.01589](https://arxiv.org/abs/1502.01589) [astro-ph.CO].
- Planck Collaboration et al. (Sept. 2020a). “Planck 2018 results. I. Overview and the cosmological legacy of Planck”. In: A&A 641, A1, A1. DOI: [10.1051/0004-6361/201833880](https://doi.org/10.1051/0004-6361/201833880). arXiv: [1807.06205](https://arxiv.org/abs/1807.06205) [astro-ph.CO].
- Planck Collaboration et al. (Sept. 2020b). “Planck 2018 results. VI. Cosmological parameters”. In: A&A 641, A6, A6. DOI: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910). arXiv: [1807.06209](https://arxiv.org/abs/1807.06209) [astro-ph.CO].
- Planck Collaboration et al. (Sept. 2020c). “Planck 2018 results. VIII. Gravitational lensing”. In: A&A 641, A8, A8. DOI: [10.1051/0004-6361/201833886](https://doi.org/10.1051/0004-6361/201833886). arXiv: [1807.06210](https://arxiv.org/abs/1807.06210) [astro-ph.CO].
- Predehl, P. et al. (Mar. 2021). “The eROSITA X-ray telescope on SRG”. In: A&A 647, A1, A1. DOI: [10.1051/0004-6361/202039313](https://doi.org/10.1051/0004-6361/202039313). arXiv: [2010.03477](https://arxiv.org/abs/2010.03477) [astro-ph.HE].
- Refregier, A. et al. (Jan. 2010). “Euclid Imaging Consortium Science Book”. In: *arXiv e-prints*, arXiv:1001.0061, arXiv:1001.0061. arXiv: [1001.0061](https://arxiv.org/abs/1001.0061) [astro-ph.IM].
- Reyes, Reinabelle et al. (Mar. 2010). “Confirmation of general relativity on large scales from weak lensing and galaxy velocities”. In: Nature 464.7286, pp. 256–258. DOI: [10.1038/nature08857](https://doi.org/10.1038/nature08857). arXiv: [1003.2185](https://arxiv.org/abs/1003.2185) [astro-ph.CO].
- Riess, Adam G. et al. (Sept. 1998). “Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant”. In: AJ 116.3, pp. 1009–1038. DOI: [10.1086/300499](https://doi.org/10.1086/300499). arXiv: [astro-ph/9805201](https://arxiv.org/abs/astro-ph/9805201) [astro-ph].
- Riess, Adam G. et al. (July 2016). “A 2.4% Determination of the Local Value of the Hubble Constant”. In: ApJ 826.1, 56, p. 56. DOI: [10.3847/0004-637X/826/1/56](https://doi.org/10.3847/0004-637X/826/1/56). arXiv: [1604.01424](https://arxiv.org/abs/1604.01424) [astro-ph.CO].
- Riess, Adam G. et al. (May 2019). “Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond Λ CDM”. In: ApJ 876.1, 85, p. 85. DOI: [10.3847/1538-4357/ab1422](https://doi.org/10.3847/1538-4357/ab1422). arXiv: [1903.07603](https://arxiv.org/abs/1903.07603) [astro-ph.CO].
- Schneider, Peter, Christopher Kochanek, and Joachim Wambsganss (2006). *Gravitational lensing: strong, weak and micro: Saas-Fee advanced course 33*. Vol. 33. Springer Science & Business Media.
- Schuhmann, Robert L., Catherine Heymans, and Joe Zuntz (Jan. 2019). “Galaxy shape measurement synergies between LSST and Euclid”. In: *arXiv e-prints*, arXiv:1901.08586, arXiv:1901.08586. arXiv: [1901.08586](https://arxiv.org/abs/1901.08586) [astro-ph.IM].
- Scolnic, D. M. et al. (June 2018). “The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample”. In: ApJ 859.2, 101, p. 101. DOI: [10.3847/1538-4357/aab9bb](https://doi.org/10.3847/1538-4357/aab9bb). arXiv: [1710.00845](https://arxiv.org/abs/1710.00845) [astro-ph.CO].
- Spergel, D et al. (2015). “Wide-field infrared survey telescope-astrophysics focused telescope assets WFIRST-AFTA 2015 report”. In: *arXiv preprint arXiv:1503.03757*.
- Square Kilometre Array Cosmology Science Working Group et al. (Mar. 2020). “Cosmology with Phase 1 of the Square Kilometre Array Red Book 2018: Technical specifications and performance forecasts”. In: PASA 37, e007, e007. DOI: [10.1017/pasa.2019.51](https://doi.org/10.1017/pasa.2019.51). arXiv: [1811.02743](https://arxiv.org/abs/1811.02743) [astro-ph.CO].
- Suzuki, A. et al. (Aug. 2016). “The Polarbear-2 and the Simons Array Experiments”. In: *Journal of Low Temperature Physics* 184.3-4, pp. 805–810. DOI: [10.1007/s10909-015-1425-4](https://doi.org/10.1007/s10909-015-1425-4). arXiv: [1512.07299](https://arxiv.org/abs/1512.07299) [astro-ph.IM].

- The LSST Dark Energy Science Collaboration et al. (Sept. 2018). “The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document”. In: *arXiv e-prints*, arXiv:1809.01669, arXiv:1809.01669. arXiv: [1809.01669](https://arxiv.org/abs/1809.01669) [[astro-ph.CO](https://arxiv.org/archive/astro-ph)].
- Wong, Kenneth C. et al. (Oct. 2020). “H0LiCOW - XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes”. In: MNRAS 498.1, pp. 1420–1439. DOI: [10.1093/mnras/stz3094](https://doi.org/10.1093/mnras/stz3094). arXiv: [1907.04869](https://arxiv.org/abs/1907.04869) [[astro-ph.CO](https://arxiv.org/archive/astro-ph)].

Chapter 2

Weak lensing

Gravitational lensing is the distortion of the image of an observed source, induced by the bending of space-time, thus of the light path, generated by the presence of mass along the line of sight.

In this chapter, I describe the formalism of weak gravitational lensing, how to estimate the cosmic shear from the measurement of galaxy shapes, recent cosmological results obtained with weak lensing, and the different steps and associated biases of a weak lensing pipeline.

2.1 Weak lensing formalism

This chapter is mainly inspired from Kilbinger (2015), Hoekstra and Jain (2008), Peter Schneider, Kochanek, and Wambsganss (2006), Bartelmann and P. Schneider (2001) and Mandelbaum, 2018.

2.1.1 Deflection from a source

We have seen in section 1.1.3 that fluctuations of the matter density ρ can be parametrised by the density contrast:

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}} \quad (2.1)$$

$\bar{\rho}$ being the mean matter density. Localised density perturbations on scales much smaller than the horizon, with peculiar velocities much smaller than the speed of light, can be described by Newtonian gravity. The density contrast can be related to the gravitational potential through the Poisson equation:

$$\nabla^2 \Phi = 4\pi G a^2 \bar{\rho} \delta \quad (2.2)$$

where Φ is the gravitational potential, a the scale factor and G the gravitational constant. Taking into account the fact that light follows null geodesics, the FLRW metric (eq. (1.5)) can be re-written as a function of the gravitational potential¹:

$$ds^2 = 0 = \left(1 + \frac{2\Phi}{c^2}\right) c^2 dt^2 - a^2(t) \left(1 - \frac{2\Phi}{c^2}\right) dr^2 \quad (2.4)$$

¹Note that this metric is usually written:

$$ds^2 = 0 = \left(1 + \frac{2\Psi}{c^2}\right) c^2 dt^2 - a^2(t) \left(1 - \frac{2\Phi}{c^2}\right) dr^2 \quad (2.3)$$

Here we considered that $\Psi = \Phi$, i.e. the absence of anisotropic stress on large scales

over it along χ' . The unlensed separation vector at a comoving distance χ , $\mathbf{x}(\chi)$, can be written $\mathbf{x}(\chi) = \mathbf{x}_0(\chi) + d\mathbf{x}(\chi)$. Consequently:

$$\mathbf{x}(\chi) = f_K(\chi)\boldsymbol{\theta} - \frac{2}{c^2} \int_0^\chi d\chi' f_K(\chi - \chi') (\nabla_\perp \Phi(\mathbf{x}(\chi'), \chi') - \nabla_\perp \Phi(0, \chi')) \quad (2.9)$$

Also, recalling eq. (2.7), $\boldsymbol{\beta} = \frac{\mathbf{x}(\chi)}{f_K(\chi)}$, we obtain the *lens equation*:

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha} \quad \text{with} \quad \boldsymbol{\alpha} = \frac{2}{c^2} \int_0^\chi d\chi' \frac{f_K(\chi - \chi')}{f_K(\chi)} (\nabla_\perp \Phi(\mathbf{x}(\chi'), \chi') - \nabla_\perp \Phi(0, \chi')) \quad (2.10)$$

being the *scaled deflection angle*.

Finally, lensing conserves surface brightness, so the surface brightness in the lens plane, I , is equal to the surface brightness distribution in the source plane, $I^{(s)}$:

$$I(\boldsymbol{\theta}) = I^{(s)}(\boldsymbol{\beta}(\boldsymbol{\theta})) \quad (2.11)$$

2.1.3 Lensing quantities: Shear and convergence

If the source is much smaller than the angular scale on which the mass distribution changes, the distortion of images can be described by the Jacobian which describes a linear mapping from the lensed coordinates $\boldsymbol{\theta}$ to unlensed coordinates $\boldsymbol{\beta}$:

$$\mathcal{A}(\boldsymbol{\theta}, \chi) = \nabla_{\boldsymbol{\theta}} \boldsymbol{\beta} = \mathbb{1} - \nabla_{\boldsymbol{\theta}} \phi \quad (2.12)$$

giving \mathcal{A} the distortion matrix, and ϕ the *lensing potential* defined as:

$$\phi(\boldsymbol{\theta}, \chi) = \frac{2}{c^2} \int_0^\chi d\chi' \frac{f_K(\chi - \chi')}{f_K(\chi) f_K(\chi')} \Phi(f_K(\chi') \boldsymbol{\theta}, \chi') \quad (2.13)$$

To compute this distortion matrix \mathcal{A} , the integral in eq. (2.9) has been approximated using the *Born approximation*: the potential gradient is integrated along the unperturbed ray, substituting the separation vector \mathbf{x} in the integral by $\mathbf{x}_0 = f_K(\chi)\boldsymbol{\theta}$. This approximation is justified in the weak lensing regime (see distinction between *weak* and *strong* regimes below).

The matrix \mathcal{A} can then be parametrised as a function of the scalar *convergence*, κ and the *shear*, $\gamma = \gamma_1 + i\gamma_2 = |\gamma|e^{2i\varphi}$. φ is the polar angle between the two shear components (see fig. 2.2).

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \quad (2.14)$$

The convergence quantifies an isotropic increase or decrease of the size of an object. The shear component translates into a stretching of the light distribution, going from a circular to an elliptical one (see fig. 2.2).

Since the shear and the convergence can not be dissociated, and are not directly accessible from observation, \mathcal{A} can be reparametrised as a function of the *reduced shear*, g , which can be computed from the shapes of galaxies (see section 2.4.2). The distortion matrix becomes:

$$\mathcal{A} = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix} \quad (2.15)$$

with the *reduced shear*: $g = g_1 + ig_2 = \frac{\gamma}{1 - \kappa}$. It is this observable that is used to compute the cosmic shear from observation (see section 2.4.2).

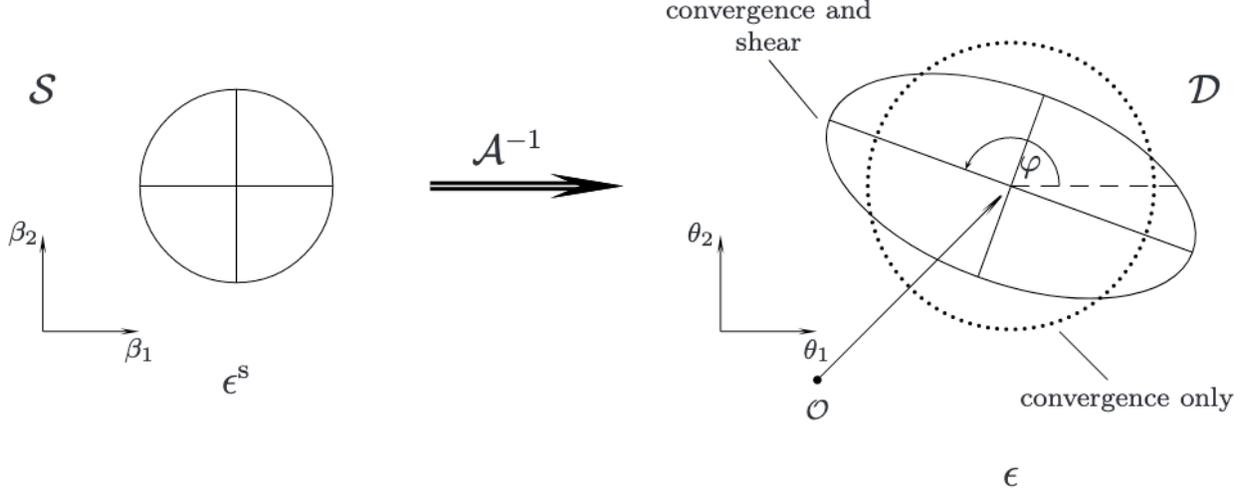


Figure 2.2: Transformation of a circular source (on the left) to an elliptical one by the inverse of the distortion matrix \mathcal{A} . Credit: Peter Schneider, Kochanek, and Wambsganss, 2006

Note: The distinction between *strong* and *weak* regimes comes from the value of the *convergence*, $\kappa(\theta)$. The convergence can be computed from the lensing potential equation and the Poisson equation obtained from the perturbed Einstein equations. It can be expressed as a function of the density contrast:

$$\kappa(\theta, \chi) = \frac{3H_0^2 \Omega_m}{2c^2} \int_0^\chi \frac{d\chi'}{a(\chi')} \frac{f_K(\chi - \chi')}{f_K(\chi)} f_K(\chi') \delta(f_K(\chi') \theta, \chi') \quad (2.16)$$

To obtain the mean convergence, this equation is weighted by the source galaxy distribution $n(\chi)$ in comoving distance. A convergence value higher than 1 implies a strong lensing regime (see right panel of fig. 2.3). The weak regime is when $\kappa \ll 1$, leading to the approximation: $g \approx \gamma$ (see left panel of fig. 2.3). In this regime, the shear and the convergence are much smaller than unity. The lensing of individual galaxy is not detectable and statistical tools are required to detect and measure this effect, as discussed in section 2.2.

2.1.4 Convergence power spectrum

Equation (2.16) shows that the convergence depends on the total matter density Ω_m and also on the density contrast δ . Consequently, it is possible to relate the statistics of the convergence to the one of the density contrast. Similarly to what was done in section 1.1.3 for the density contrast, it is possible to define the convergence power spectrum, in Fourier space, as:

$$\langle \hat{\kappa}(\mathbf{l}) \hat{\kappa}(\mathbf{l}') \rangle = \delta^{(2)}(\mathbf{l} + \mathbf{l}') P_\kappa(l) \quad (2.17)$$

From eq. (2.16), the convergence is the projection of the density contrast and consequently is a homogeneous and isotropic Gaussian field. We can then define a correlation function, ξ_κ :

$$\xi_\kappa(\theta) = \langle \kappa(\boldsymbol{\varphi}) \kappa(\boldsymbol{\varphi} + \boldsymbol{\theta}) \rangle \quad (2.18)$$

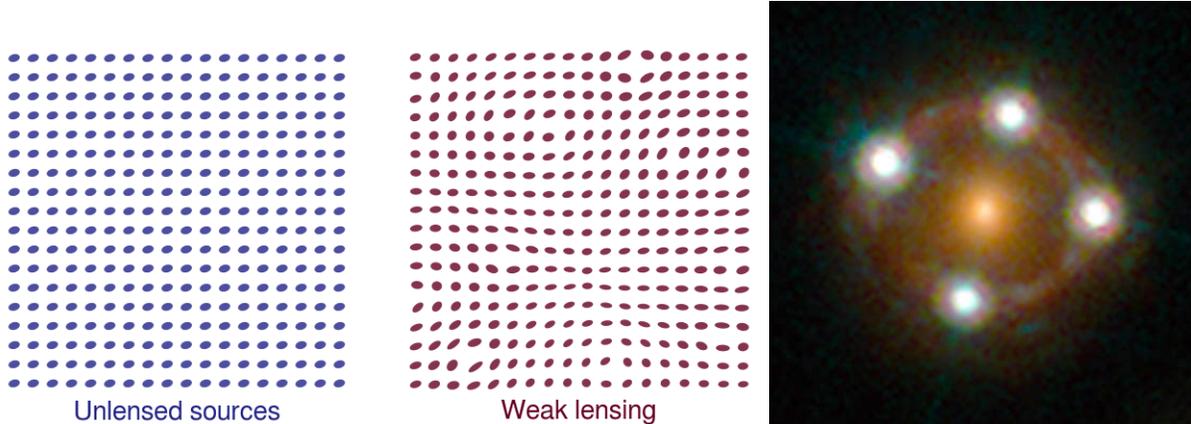


Figure 2.3: (2 panels on the left) These two panels show an example of *weak lensing*. The left panel represents sources regularly placed and oriented similarly whereas the middle panel represents the appearance of the same sources weakly lensed by a cosmic shear field. (right) A quasar being lensed by a galaxy to produce multiple images of it. This is an example of the *strong lensing* regime. Credit: ESA/Hubble, NASA, Suyu et al.

Then, computing this correlation function, one can derive the convergence power spectrum as a function of the density power spectrum (see Peter and Uzan (2013) for a complete derivation):

$$P_\kappa(l) = \frac{9H_0^4}{4c^4} \Omega_m^2 \int_0^{\chi_H} \left(\frac{g(\chi)}{a(\chi)} \right)^2 P_\delta \left(\frac{l}{f_K(\chi)}, \chi \right) d\chi \quad (2.19)$$

with

$$g(\chi) = \int_\chi^{\chi_H} n(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')} \quad (2.20)$$

χ_H is the size of the observable universe.

We saw in section 2.1.3 that the shear and the convergence are not independent since they are defined from the second derivative of the lensing potential ϕ . It is possible to show that the convergence power spectrum is equal to the shear power spectrum (see §6.3 of Peter Schneider, Kochanek, and Wambsganss (2006) for example):

$$P_\gamma(l) = P_\kappa(l) \quad (2.21)$$

P_κ and P_γ are related to Ω_m and P_δ , which is itself related to Ω_Λ and σ_8 . Consequently, we can use the shear correlation functions as observable to probe Ω_m , Ω_Λ and σ_8 .

2.2 Shear related probes

First, I focus on the shear two-point correlation function (2PCF) and then on its cross correlation with the galaxy density.

2.2.1 Shear correlation function

Consider a pair of galaxies located at Θ and $\Theta + \theta$, with ϕ the polar angle of θ (see fig. 2.4). The two components of the shear of each galaxy can be decomposed on a new base (\hat{e}_+ , \hat{e}_\times), to define

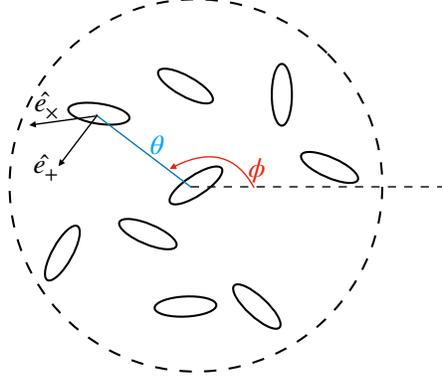


Figure 2.4: The components of the shear correlation function are computed between two galaxies separated by θ . The shear components are decomposed in the new base $(\hat{e}_+, \hat{e}_\times)$ defining the tangential and cross-components of the shear.

the tangential and cross-components, γ_+ and γ_\times , defined as:

$$\gamma_+(\theta) = -\text{Re}(\gamma(\theta e^{-2i\phi})) \quad \text{and} \quad \gamma_\times(\theta) = -\text{Im}(\gamma(\theta e^{-2i\phi})) \quad (2.22)$$

with $\gamma = \gamma_1 + i\gamma_2$. By convention, tangential shear around overdensities gives $\langle \gamma_+ \rangle > 0$, and radial shear around underdensities results in $\langle \gamma_+ \rangle < 0$.

Combining the tangential and cross components, it is possible to build two non-zeros correlators, the two components of the shear correlation (or two-point correlation function):

$$\xi_+(\theta) = \langle \gamma \gamma^* \rangle(\theta) = \langle \gamma_+ \gamma_+ \rangle(\theta) + \langle \gamma_\times \gamma_\times \rangle(\theta) \quad (2.23)$$

$$\xi_-(\theta) = \text{Re}[\langle \gamma \gamma \rangle(\theta) e^{-4i\phi}] = \langle \gamma_+ \gamma_+ \rangle(\theta) - \langle \gamma_\times \gamma_\times \rangle(\theta) \quad (2.24)$$

with the two galaxies separated on the sky from $\theta = |\theta|$ (see fig. 2.4).

An estimator of these components of the 2PCF for pairs of galaxies (i, j) , separated by an angle θ , is (P. Schneider et al., 2002):

$$\xi_\pm(\theta) = \frac{\sum_{ij} w_i w_j (\epsilon_{+,i} \epsilon_{+,j} \pm \epsilon_{\times,i} \epsilon_{\times,j})}{\sum_{ij} w_i w_j} \quad (2.25)$$

Each galaxy has a measured ellipticity ϵ_i and an associated weight w_i which can depend on the measurement uncertainty including shape noise.

This estimator can be linked to cosmology since ξ_\pm can be obtained as Hankel transforms of the convergence power spectrum (see section 2.1.4 and fig. 2.5):

$$\xi_\pm(\theta) = \frac{1}{2\pi} \int l dl J_{0/4}(l\theta) [P_\kappa^{(E)}(l) \pm P_\kappa^{(B)}(l)] \quad (2.26)$$

with J_n the Bessel function of the n-th order.

The components of the 2PCF contain both E- and B-modes power spectra. Lensing should produce only E-modes as it is expected to be a pure gradient field (curl free) and the detection of significant B-modes usually indicates that systematics remain (note however that B-modes can also come from physical effects such as intrinsic alignments, see section 2.2.3). It is possible to perform the E/B-modes decomposition in order to test for systematics (see §3.9 of Kilbinger, 2015).

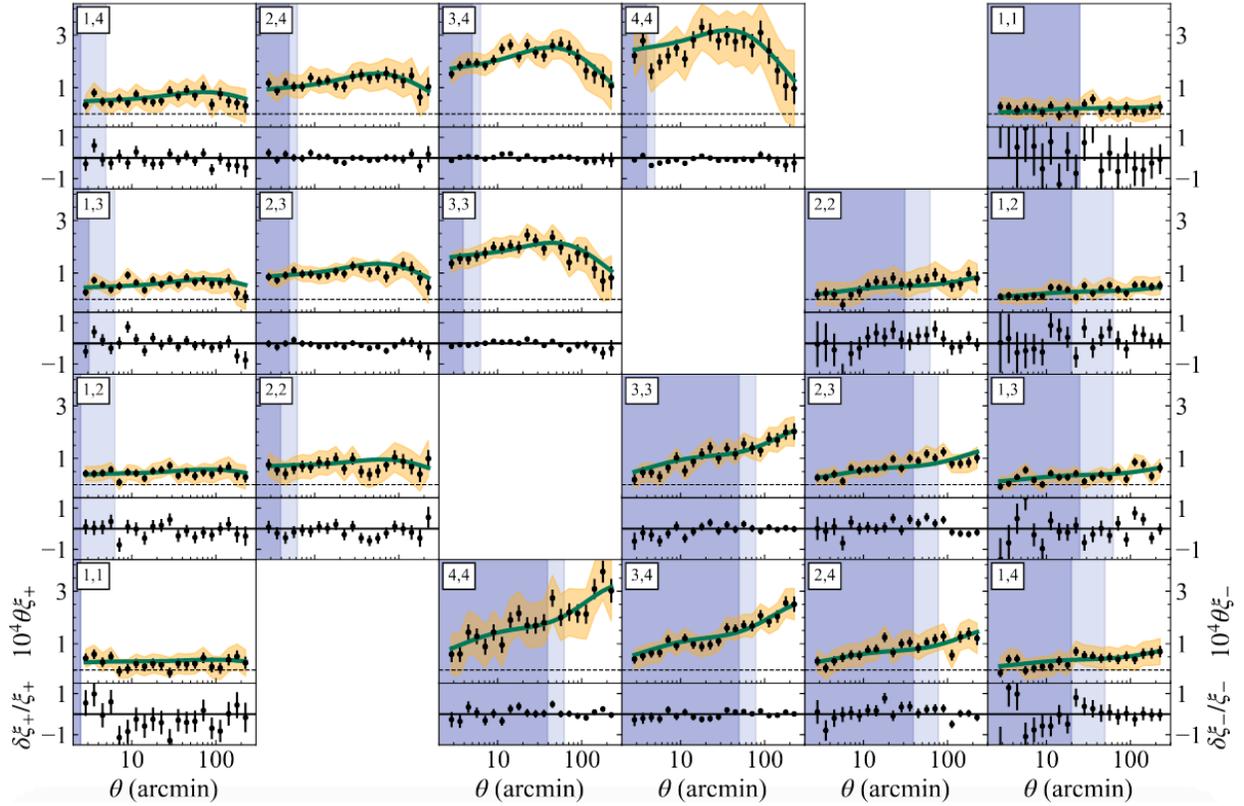


Figure 2.5: Cosmic shear two-point correlation functions measured by DES Y3: $\xi_+(\theta)$ (left) and $\xi_-(\theta)$ (right), scaled by the angular separation, θ . The correlation functions are measured for each redshift bin pair (see section 2.2.4), indicated by the label, and the error bar represents the square root of the diagonal of the analytic covariance matrix. The green line denotes the best-fit Λ CDM theoretical prediction from the cosmic shear-only tomographic analysis. Credit: A. Amon et al., 2021.

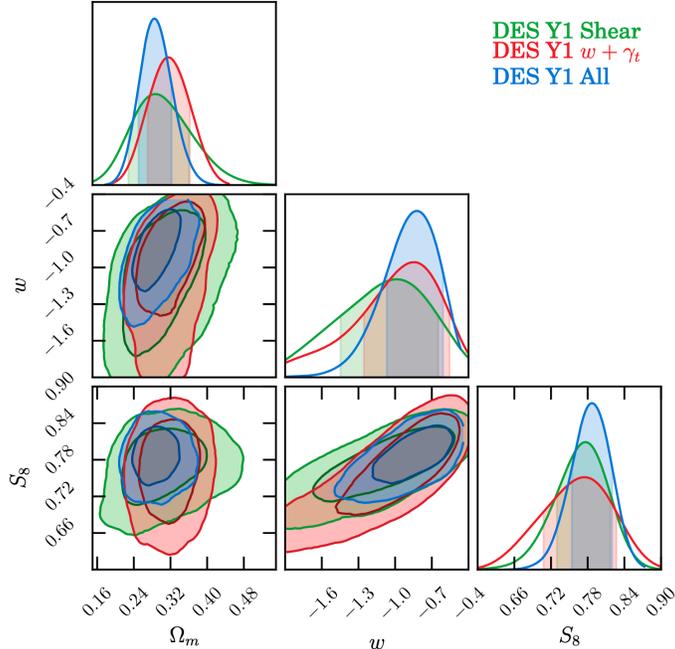


Figure 2.6: The DES-Y1 constraints on the cosmological parameters σ_8 and Ω_m comparing the results obtained with the cosmic shear analysis only (green), the galaxy clustering and galaxy-galaxy lensing (red), and with the 3x2-point correlation function (blue). Contours show the 68% (inner) and 95% (outer) confidence levels. Credit: Abbott et al., 2018

2.2.2 3x2-point correlation function

A powerful application of the cosmic shear is its cross correlation with the galaxy density. In addition to the shear-shear correlation function, the 3×2 -point correlation function takes into account the shear-galaxy (*galaxy-galaxy lensing*) and the galaxy-galaxy correlation function (*galaxy clustering*). Galaxy-galaxy lensing (GGL) computes the correlation of the shapes of high-redshift galaxies with positions of galaxies at lower redshift. This is the cross-correlation of lens galaxy positions and source galaxy shapes. Galaxy clustering is the galaxy two-point correlation function, describing the galaxy spatial distribution.

These functions are not independent but provide different information allowing to break degeneracies between cosmological parameters and systematics producing biases. Consequently the combination of these correlation functions, the 3x2-point correlation function, gives tighter constraints on cosmological parameters than the single shear 2PCF (for example, see fig. 2.6 extracted from Abbott et al., 2018).

2.2.3 Intrinsic alignments: a bias in cosmological parameters estimation

Intrinsic alignments (IA)² is the correlation of shapes due to gravitational interactions between galaxies and the surrounding tidal fields. Consequently, even without lensing, the shapes of galaxies can be correlated. This excess of correlation can induce severe biases on cosmological parameters (see fig. 2.7) and must be taken into account in the computation of the correlation functions. It is particularly difficult to predict IA theoretically and several procedures such as the removal of close

²For a complete review of intrinsic alignments in the context of weak lensing, see Troxel and Ishak (2015).

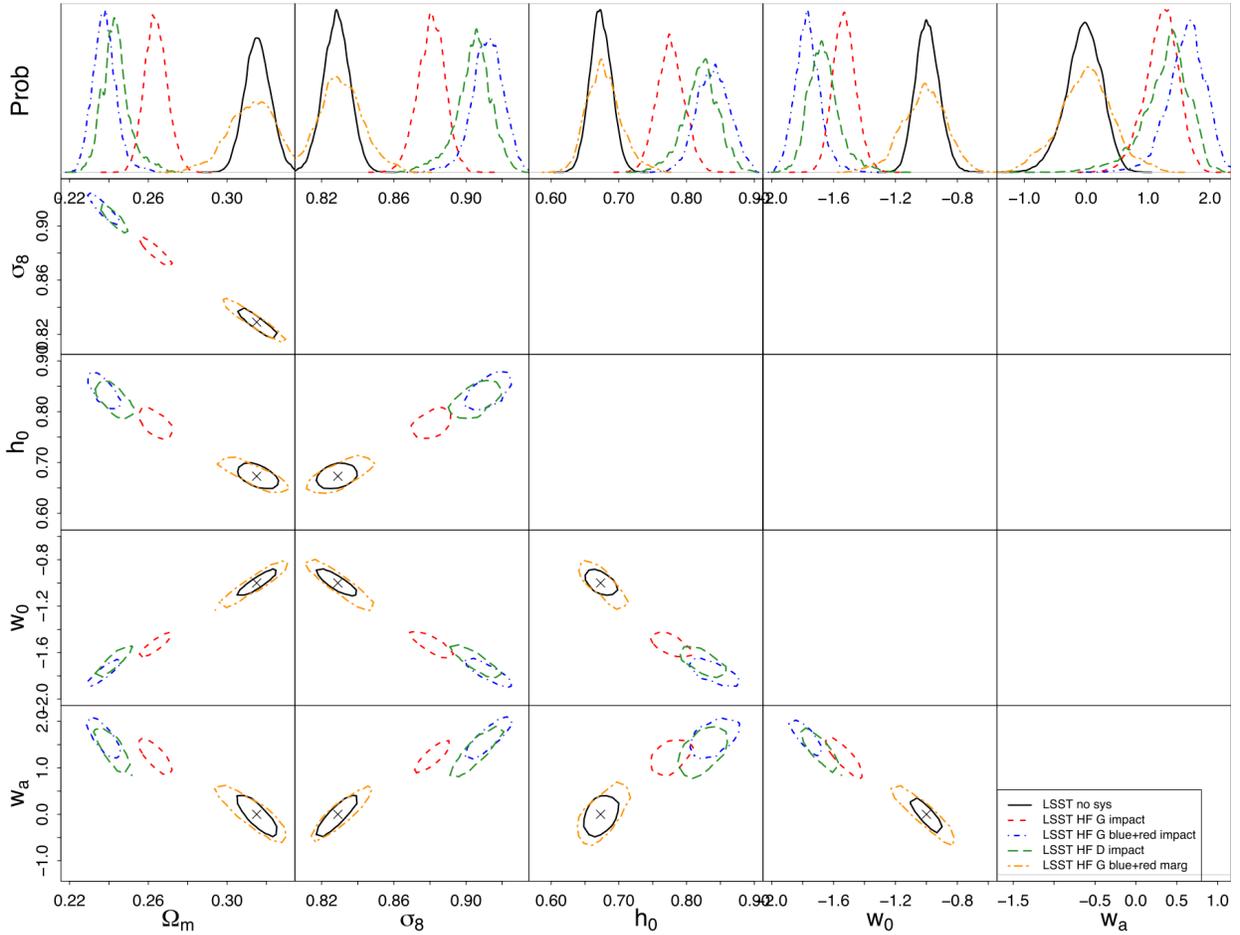


Figure 2.7: Effect of intrinsic alignment on cosmological parameters with weak lensing in LSST, using shear-shear correlation only. The black curves represents the case without intrinsic alignments, the three other colors are different models of intrinsic alignments. Credit: Krause, Eifler, and J. Blazek, 2016

pairs of galaxies, joint modeling or the self-calibration using joint analysis of the three correlation functions, have been proposed to mitigate this effect (for example, Heymans and Heavens (2003) and Yao et al. (2017)). Self-calibration method has recently been applied to KiDS data (Yao et al., 2020) and proves to be able to distinguish IA and lensing signals.

As shown on fig. 2.7, intrinsic alignments can have a major impact on cosmological parameter estimation for LSST survey. However, mitigation strategies are nowadays able to remove biases due to IA for future stage IV surveys (Krause, Eifler, and J. Blazek, 2016). A drawback is that these strategies can degrade the constraining power on cosmological parameters up to several tens of percents in pessimistic cases. More work is currently being done to reduce impact of IA on lensing measurements, for example providing better IA models (J. A. Blazek et al., 2019).

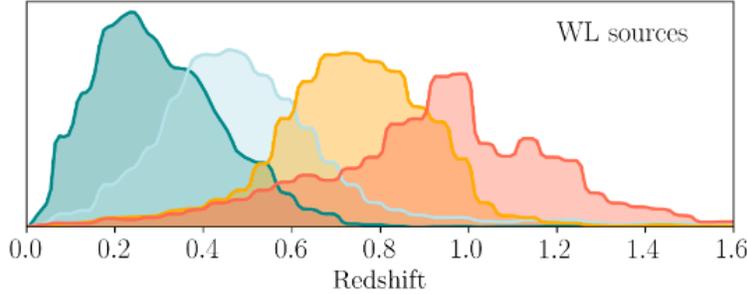


Figure 2.8: Normalized redshift distribution of sources of the weak lensing catalogue divided into four bins for DES-Y3. Credit: Prat et al., 2021

2.2.4 Shear tomography

Redshift binning

As shown on fig. 2.5, in order to measure the evolution of large scale structures, the analysis is usually performed in redshift bins, a technique called *tomography* (Hu, 1999). This provides a lensing field per bin, which gives information about different epochs in the history of the Universe. We obtain a 3D tomographic view of the mass distribution across the Universe permitting to apply constraints on the growth of structure or on time-varying dark energy state parameter $w(z)$ (see section 1.2.2). For a redshift distribution divided into N bins, we can compute a projected overdensity (or convergence, see section 2.1.2) per bin. This yields $\frac{N(N-1)}{2}$ power spectra (see section 2.2.1) $P_{\kappa,ij}$, $1 \leq i \leq j \leq N$, including auto-spectra ($i = j$) and cross-spectra ($i \neq j$).

However, distributing the galaxies into strict uncorrelated redshift bins (see fig. 2.8) is not straightforward. For example, photometric redshift (see next section) errors can cause galaxy distributions in different bins to overlap, leading to an increase of correlation between them. These correlations can alter the correlation functions and the convergence power spectra, leading to incorrect cosmological parameter estimation. Methods re-weighting galaxy redshift distribution have been proposed to obtain uncorrelated bins (for example, Bernardeau, Nishimichi, and Taruya, 2014) and work is currently being done in order to reduce the systematics on photometric redshift estimation (see next section).

Instead of using tomographic bins, 3D-lensing (Heavens, 2003, Castro, Heavens, and T. D. Kitching, 2005, T. D. Kitching, Heavens, and Miller, 2011) proposes to use fully the redshift information, i.e. the full probability distribution $p(z)$, of each galaxy individually. The major advantage of 3D-lensing is that it does not require binning, avoiding the loss of information along the redshift direction that this process implies. The downside is that it relies on a precise measurement of the full posterior distribution of individual photometric redshifts, which is a very difficult task (see below). It has already been used on CFHTLenS data (T. D. Kitching et al., 2014) and LSST might be a good candidate for 3-D lensing since it is expected to reach a good precision on photometric redshift estimation (The LSST Dark Energy Science Collaboration et al., 2018).

Photometric redshifts

There are two ways of measuring the redshift of a galaxy: the most precise consists in measuring the spectrum of the galaxy via a spectroscopic survey, this is the *spectroscopic redshift*, the other one involves using the color information contained in images of a multi-band photometric survey to compute the *photometric redshift*.

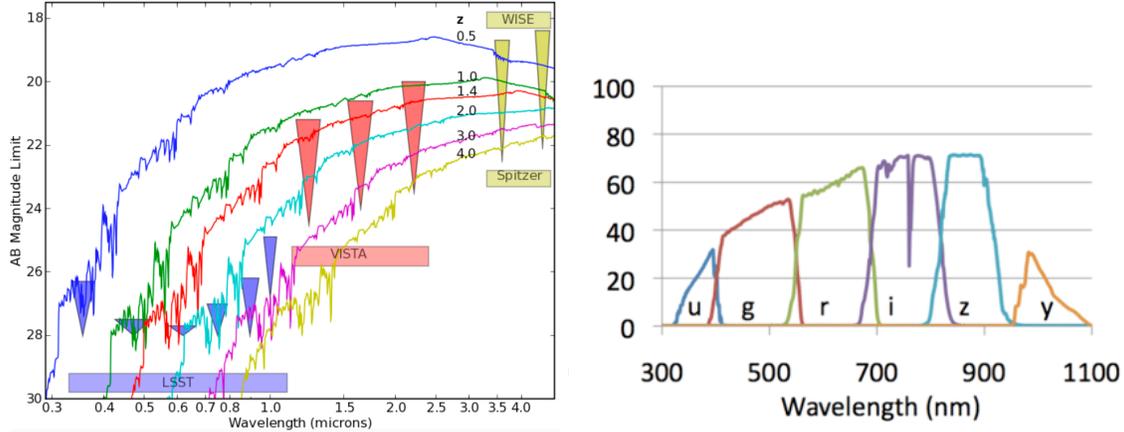


Figure 2.9: (left) Spectral energy distribution of a red sequence galaxy as a function of redshift. (right) Total throughput, in %, of the *ugrizy* LSST filters, accounting for the atmospheric, optical and electronic effects. Credit: Abell et al., 2009

Photometric redshifts are estimated from images using color information (several filters are necessary). The first step is to measure the flux of the galaxy, or of a subset of the galaxy (chosen consistently across the bands), in each filter. This gives an estimation of the Spectral Energy Distribution (SED) of the object (see fig. 2.9). This task is complex as PSF varies with the band and systematics such as seeing or extinction step in. There are mostly two kinds of methods to infer photometric redshifts from measured fluxes: template-fitting (for example, "Le Phare"³ or Brammer, van Dokkum, and Coppi, 2008) or machine learning algorithms (for instance, Pasquet et al., 2019). Template-fitting methods rely on a set of templates for galaxies' SEDs which are used to predict the galaxy photometry as a function of redshift. On the other hand, machine learning methods learn from data the relationship between the photometry and the redshift. Both suffer from almost the same problem: a not good enough representativeness of the real data distribution, whether it comes from incorrect templates, or from incomplete training samples. The methods can lead to "catastrophic" redshift estimations, i.e. errors of a significant amount compared to the true redshifts. These outliers can lead to large biases in the redshift bin distributions. One way to reduce these systematics is to use the full probability distribution $p(z)$ for each galaxy, rather than the most probable redshift (D. Wittman, 2009) (see fig. 2.10). For example, some of the "catastrophic" outliers have a multimodal redshift probability (because of color redshift degeneracies) and considering the complete distribution of redshifts, instead of the most probable value allows to take into account the different modes.

Spectroscopic data is a solution to calibrate the photometric redshift distribution but even this avenue collides with an issue of representativeness. The spectroscopy cannot go as faint as the photometry and even for less deep surveys, the completeness of the sample depends of the type, magnitude and redshift of the sources (Abell et al., 2009). Stage IV surveys will be able to rely on already recorded data, such as the one from the Sloan Digital Sky Survey for example (SDSS, York et al., 2000), which measured around 2 millions of galaxy spectra over a 10 years period (BOSS + eBOSS) (Ahumada et al., 2020). Future wide spectroscopic surveys will also provide valuable information. For example, the Dark Energy Spectroscopic Instrument (DESI, Levi et al., 2019) will give the spectra of 30 million galaxies and quasars covering 14,000 square degrees up to a redshift of 1.7 for galaxies. The future 4-metre Multi-Object Spectroscopic Telescope (4MOST,

³<http://lephare.lam.fr>, Arnouts S. and Ilbert O.

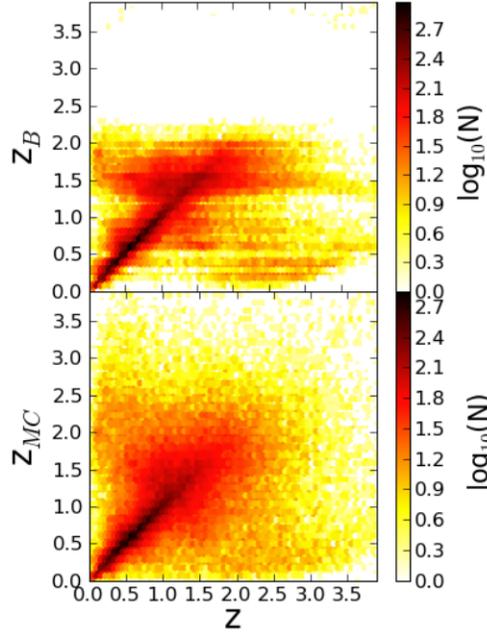


Figure 2.10: Top: each galaxy most probable redshift, z_B , as a function of the spectroscopic redshift, z . Bottom: One Monte-Carlo sample from each galaxy’s redshift distribution, z_{MC} , as a function of the spectroscopic redshift, z . Sampling the redshift distribution cleans up artefact observed when considering the most probable photometric redshift. Credit: D. Wittman, 2009.

R. S. de Jong et al., 2019) has, as primary purpose, to provide the spectroscopic information to large-area surveys like Euclid with more than 20 million spectra. Even if these numbers can seem small compared to the number of galaxies observed by the photometric surveys (100 millions of galaxies in the DES-Y3 weak lensing catalog, ten billions of galaxies expected for LSST), they are sufficient to efficiently calibrate the photometric redshifts (Abell et al., 2009) and the issues mostly come from representativeness.

Another major issue comes from the blending, or overlap, of sources. Recovering the photometry of blended sources is naturally very difficult and even having a spectroscopic counterpart might not be enough to overcome this issue (it is not possible to have spectroscopic data for all blended objects and even in that case, the spectrum would be composed of the SEDs several sources). In chapter 7, we present a deblending algorithm that allows to recover galaxy’s photometry, and in chapter 8, we propose a neural network for galaxy parameters estimation, focused initially on shape parameters estimation, that can provide a posterior distribution for photometric redshifts.

Now that we have seen how to constrain cosmological parameters from weak lensing and some of the major difficulties that can arise, I will review recent results obtained from shear analyses.

2.3 Cosmological constraints with weak lensing

2.3.1 Recent results

The detection of weak lensing was first reported by four different groups at the same time, showing correlations in galaxy ellipticities larger than the expected systematics (Bacon, Refregier, and

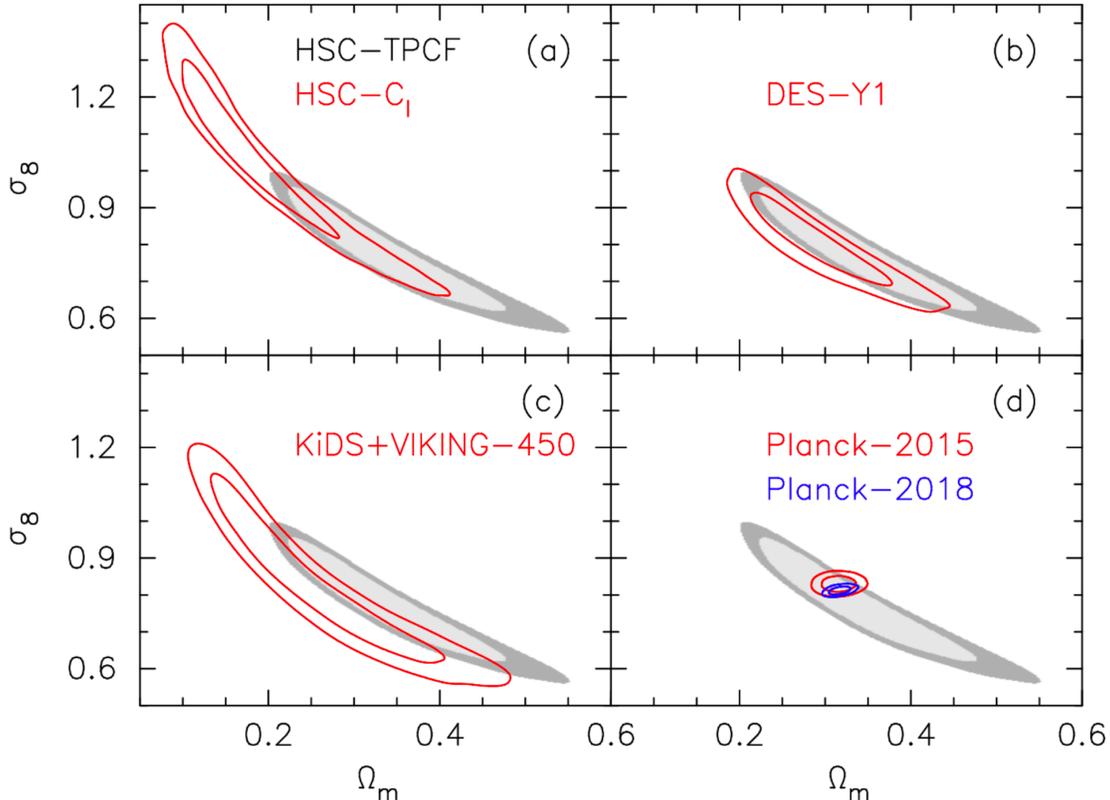


Figure 2.11: The HSC-Y1 constraints on the cosmological parameters σ_8 and Ω_m compared to DES, KiDS and Planck results. Top right shows the comparison with DES-Y1. Contours show the 68% (inner) and 95% (outer) confidence levels. Credit: Hamana et al., 2020

Ellis, 2000; Van Waerbeke et al., 2000; D. M. Wittman et al., 2000; Kaiser, Wilson, and Gerard A. Luppino, 2000). This was the first time weak lensing analyses could apply constraints on cosmological parameters, especially Ω_m and σ_8 .

Currently, stage III surveys are paving the way for next generation of surveys such as LSST, Euclid or Roman. The three main surveys are the Hyper Suprime-Cam Subaru Strategic Program (HSC, Aihara et al., 2018), the Kilo-Degree Survey (KiDS, J. T. A. de Jong et al., 2013) and the Dark Energy Survey Collaboration (DES, The Dark Energy Survey Collaboration, 2005). Figure 2.11 shows the comparison of two recent results obtained by HSC first year of data (HSC-Y1, Hamana et al., 2020) and DES year 1 (DES-Y1, Troxel et al., 2018). Since DES-Y1 covers a much larger area than HSC-Y1 (1321 deg² against 137 deg²), it presents tighter constraints on the cosmological parameters. It is also noticeable that HSC-Y1 constraints at least partially overlap constraints from other surveys (Planck, Planck Collaboration et al. (2020) and KiDS+VIKING-450 Hildebrandt et al. (2020)), finding no particular tension with Planck results in particular. More recent result were published by KiDS recently, KiDS-1000, presenting stronger constraints on $S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$ (Asgari et al., 2021).

The most recent results come from DES-Y3 (including three years of data from DES), which presents an improvement of these constraints. Left panel of fig. 2.12 shows the results obtained for the cosmic shear only analysis. Right panel shows the results obtained with the 3x2-point correlation functions, tightening the constraints on Ω_m and S_8 (A. Amon et al., 2021; DES Collaboration et al.,

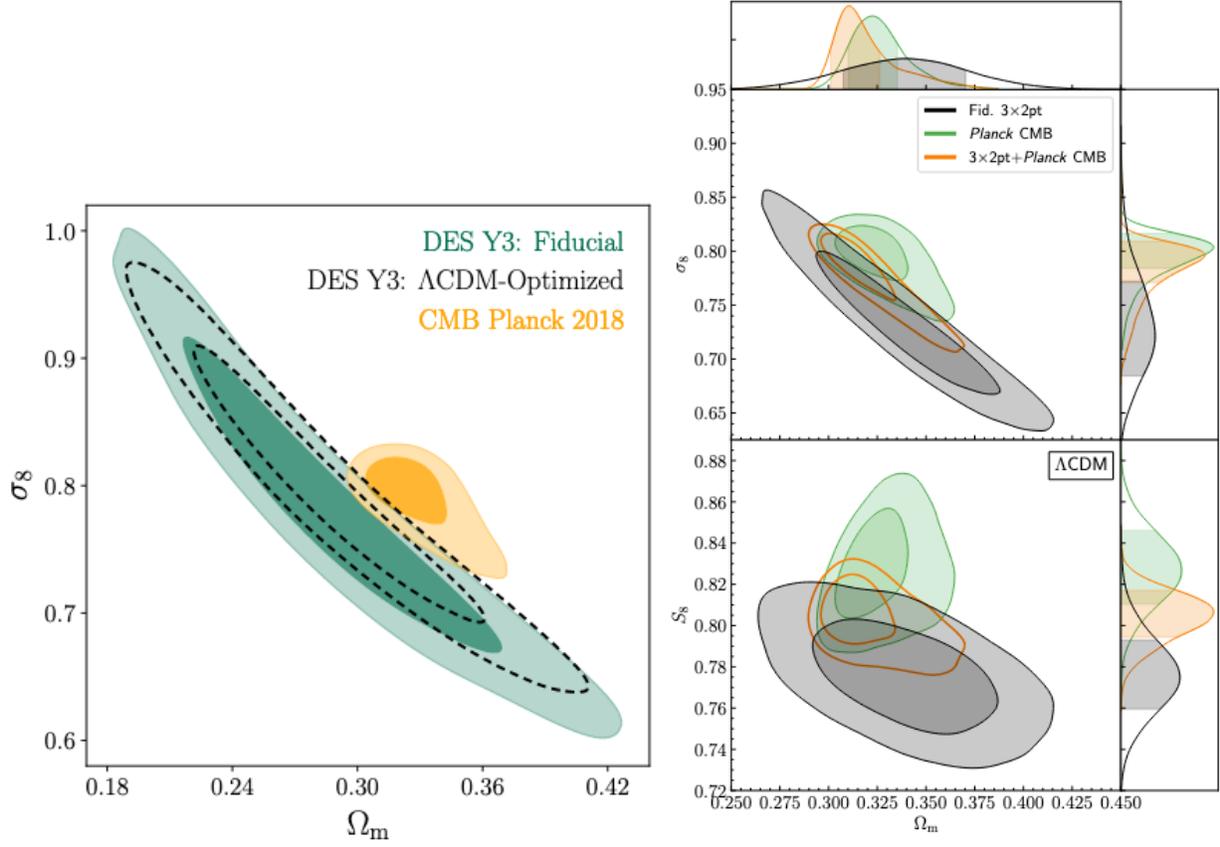


Figure 2.12: (left) DES-Y3 cosmological constraints on the clustering amplitude, σ_8 and the matter density Ω_m in Λ CDM. The contours (inner 68% and outer 95% confidence levels) are shown for the Fiducial DES-Y3 analysis in green and Planck 2018 CMB in yellow. The black dashed contours represent the Λ CDM-Optimized analysis. Credit: A. Amon et al., 2021. (right) Comparison of the marginalized parameter constraints in the Λ CDM model. Results of the 3×2 -points analysis of DES-Y3 are shown in black, the prediction from the CMB results of Planck 2018 are shown in green and the combination if the two are shown in orange. Credit: DES Collaboration et al., 2021

2021). Using the 3×2 -point correlation functions allows to increase the FoM in the (σ_8, Ω_m) plane by a factor of more than 3 compared to the shear-shear correlation function only. The comparison with the prediction from Planck 2018 (Planck Collaboration et al., 2020) is shown on both panels and, once again, no particular tension is found. Results are in overall good agreement with the Λ CDM model.

Stage IV surveys such as LSST will aim tightening even more the constraints on cosmological parameters. Drastic requirements are imposed on systematics in order to achieve the goal of multiplying the FoM by 10 compared to stage II surveys (Albrecht et al., 2006). Some of them are detailed below.

2.3.2 Expectations for LSST

Since results presented by the DES-Y3 series of papers⁴ provide the most updated results in weak lensing, here I compare the LSST and DES requirements for cosmic shear analysis. This will give a hint of the differences between these two surveys and the challenges that LSST will face. The requirements for LSST are listed in the Science Requirements Document (SRD) (The LSST Dark Energy Science Collaboration et al., 2018) and set to meet the goal of increasing by a factor of 10 the FoM on dark energy parameters compared to stage II surveys.

First we have seen that the cosmic shear depends on the estimation of the photometric redshift distribution. LSST requires the redshift distribution to be known to about 0.001 and 0.003 respectively for the mean and scatter, in each redshift bin, whereas DES-Y3 reached a precision of 0.01 (Myles et al., 2020).

The second parameter that needs calibration for cosmic shear analysis is the shear bias. Indeed, a lot of systematics can contaminate the shear estimation (see section 2.5) and decreasing their impact is a major challenge.

Usually shear biases can be characterized, to first approximation, using a multiplicative bias m and an additive bias c . They are linked to the observed and intrinsic ellipticities, and to shear (see next section), via the formula:

$$\epsilon_i^{obs} = \epsilon_i^{int} + (1 + m_i)\gamma_i + c_i \quad \text{with } i \in \{1, 2\} \quad (2.27)$$

However, in the case of LSST, the multiplicative bias is expected to vary as a function of redshift. The LSST SRD considers, for simplicity, that the multiplicative bias depends linearly on redshift:

$$m(z) = m_z \left(\frac{2z - z_{max}}{z_{max}} \right) + m_{avg}, \quad (2.28)$$

where z_{max} is the value at the middle of the highest redshift bin, $m_{avg} = m(z_{max}/2)$ is the average value of m in the range $[0, z_{max}]$. LSST aims at a total systematic uncertainty in the redshift-dependent shear calibration below 0.3%, meaning that in each redshift bin, the multiplicative bias $m(z)$ must be below 0.3%. For comparison, a value of about 2 to 3% was taken into account in the calibration for DES-Y3 (DES Collaboration et al., 2021).

Consequently, LSST has to improve systematics management for both redshift distribution and shear bias calibration. One of the systematics which is particularly difficult to calibrate, as it impacts both redshifts and shear biases, is *blending*, the overlap of sources on an image (see chapter 3). Interestingly, the DES-Y3 analysis shed light on the impact of blending on both shear and photometric redshift calibration arguing that they can't be dissociated (MacCrann et al., 2020). They argue that for future surveys, instead of the usual multiplicative shear bias, m , (see section 2.5) the requirements should be set on the effective redshift distribution n_γ :

$$n_\gamma = \bar{R}(z)n_{\text{eff}}(z) \quad (2.29)$$

with $\bar{R}(z)$ the mean shear response in case of shear bias (see section 2.5.5), and $n_{\text{eff}}(z)$, the density distribution of galaxy usable for weak lensing.

This is a very crucial point for LSST for example. With its increased depth, the survey will yield a density of galaxies n_{eff} of $\approx 40 \text{ arcmin}^{-2}$ (compared to 5.59 arcmin^{-2} for DES-Y3), having to deal with much higher levels of blending. Work is being done to address this systematic (see chapter 3) and the work presented in this thesis focuses on possible avenues using deep learning algorithms (see chapter 7 and chapter 8).

⁴<https://www.darkenergysurvey.org/des-year-3-cosmology-results-papers/>

However, before discussing blending more thoroughly, let's see how to go from images to the estimation of cosmic shear.

2.4 Measuring shear from galaxies shapes

As mentioned before, the cosmic shear can be computed from the galaxy shapes. The quantity measured is the *ellipticity* and can be computed directly from the image using quadrupole moments of the surface brightness.

2.4.1 From image to ellipticities

Let be $I(\boldsymbol{\theta})$ the brightness distribution of a source, it is possible to define the tensor of second brightness moments (or quadrupole moments):

$$Q_{i,j} = \frac{\int d^2\theta I(\boldsymbol{\theta}) q_I(I(\boldsymbol{\theta})) (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)}{\int d^2\theta I(\boldsymbol{\theta}) q_I(I(\boldsymbol{\theta}))} \quad i, j \in \{1, 2\} \quad (2.30)$$

with $q_I(I(\boldsymbol{\theta}))$ a chosen weight function, and $\hat{\theta}$, the center of the source. The weight function is chosen in order to set a range of integration for the surface brightness (see section 2.4.2). It can be the Heaviside step function, the light profile itself or a gaussian function for example (Kaiser, Squires, and Broadhurst, 1995a).

Several types of ellipticity parameters exist (see Bonnet and Mellier, 1995). Here, I present two as they are the most common in cosmology and are used in chapter 7 and chapter 8. They are both called *complex ellipticity* and the choice of which is the most convenient depends on the context. The two parameters are e and ϵ , and are defined as follows:

$$e = \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}} \quad \text{and} \quad \epsilon = \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22} + 2(Q_{11}Q_{22} - Q_{12}^2)^{1/2}} \quad (2.31)$$

The two definitions have the same phase, 2φ , such that:

$$e = |e| \exp 2i\varphi \quad \text{and} \quad \epsilon = |\epsilon| \exp 2i\varphi \quad (2.32)$$

The only difference is the module. With $\frac{b}{a}$, the minor-to-major axis ratio (see fig. 2.13) they can be expressed as:

$$|e| = \frac{1 - (\frac{b}{a})^2}{1 + (\frac{b}{a})^2} \quad \text{and} \quad |\epsilon| = \frac{1 - \frac{b}{a}}{1 + \frac{b}{a}} \quad (2.33)$$

2.4.2 From ellipticities to the shear

Each galaxy has an intrinsic ellipticity, ϵ^{int} , which is modified by cosmic shear. It can be related to the observed ellipticity ϵ^{obs} with:

$$\epsilon^{obs} = \frac{\epsilon^{int} + g}{1 + g^* \epsilon^{int}} \quad (2.34)$$

g is the reduced shear (see section 2.1.3) ("*" means complex conjugate). In the weak lensing regime, one can consider that $\kappa \ll 1$ and $|\gamma| \ll 1$. Equation (2.34) can be approximated to

$$\epsilon^{obs} \approx \epsilon^{int} + g \approx \epsilon^{int} + \gamma \quad \text{with} \quad g \approx \gamma \quad (2.35)$$

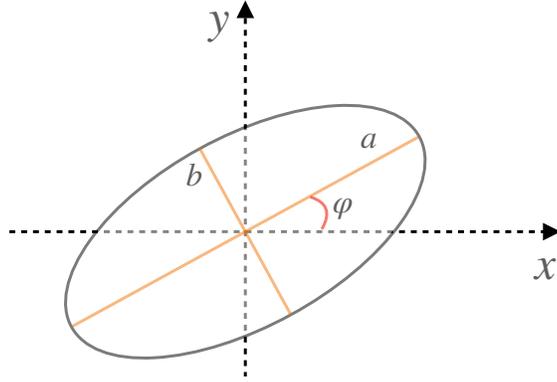


Figure 2.13: Minor and major axis of a galaxy (respectively a and b) and the polar angle φ .

If the intrinsic ellipticity of galaxies have their orientations randomly distributed, then: $\langle \epsilon^{int} \rangle = 0$. The observed ellipticity is then an unbiased estimator of the reduced shear:

$$\langle \epsilon^{obs} \rangle \approx \frac{\langle e^{obs} \rangle}{2} \approx g \approx \gamma \quad (2.36)$$

Unfortunately, there are parts of the sky where the intrinsic ellipticity does have preferred orientations, for example in the presence of intrinsic galaxy alignments, but mitigation strategies are nowadays able to remove biases due to IA for stage IV surveys (see section 2.2.3).

We have seen that, to estimate the reduced shear, and hence the shear, one needs to measure ellipticities of galaxies. This measurement is one particular stage of a weak lensing pipeline.

2.5 Weak lensing pipeline and associated biases

In this part, I briefly describe the different steps of a weak lensing pipeline and the associated shear biases that they can introduce. We have seen in section 2.3.2 that shear biases could be parametrised by a multiplicative bias m , and an additive bias c .

Most of methods used nowadays can generate a multiplicative bias between 10 and 1% of the shear estimate whereas stage IV surveys requirements aim for a mean shear bias $\approx 10^{-3}$. A good understanding of the different analysis stages at which biases can occur is key to a success in meeting these requirements.

Note that the processing steps described below can be performed on co-adds (stacked images of the different exposures), or on individual exposures and that the choice between the two is not trivial, each option having its pros and cons.

2.5.1 Detection and blending

The first step is the detection and identification of sources. It is necessary to separate stars from galaxies since the former are required to estimate the Point-Spread-Function (PSF) and the latter to measure the cosmic shear.

The most used detection software is `SExtractor` (Bertin and Arnouts, 1996), which looks at pixels above a certain threshold defined either as a function of the noise level, or at a fixed value: if a fixed number of contiguous pixels are detected above the threshold, an object is detected. It was particularly used in the recent DES-Y3 analysis (M. Gatti et al., 2020).

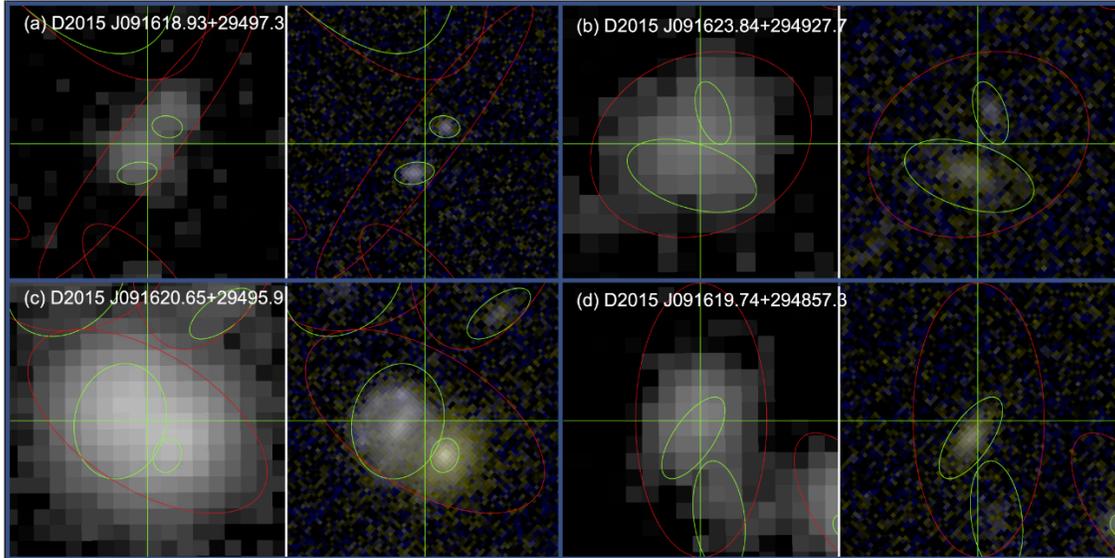


Figure 2.14: Ambiguous blends in the Musket Ball Cluster seen by the Subaru (first and third column) and HST (second and fourth column). Subaru is a ground telescope and suffers from the atmospheric PSF whereas HST is a space telescope. Therefore it benefits a much better resolution. In red are drawn the objects as detected on the Subaru images, in green the ones detected on the HST images. The close blends of galaxies are detected as one single object on ground telescope images. Credit: Dawson et al., 2016

More recent and more complex solutions are proposed either using a convolutional neural networks (CNN) (Paillassa, 2020), or embedded in astronomical image generative model (Regier et al., 2015; Liu, McAuliffe, and Regier, 2021), but none of them has been used in the context of cosmological analysis yet. They seem quite promising but the ones embedded in astronomical image generative model are, among other things, too computationally costly today to be used in a processing pipeline. Even though it is quite preliminary, the CNN method could be particularly interesting since it should be fast and, more interestingly, since CNN simply adapt to multi-band images, allowing for the use of color information from multi-band data. This kind of method might help to deal with some of the biases described below (blending in particular).

The detection can introduce a bias because of several effects. First, the detection is shear-dependent. Indeed, since shear conserve the surface brightness, a galaxy with intrinsic ellipticity perpendicular to the shear will appear rounder, which will increase its probability to be detected if the detector uses a circular kernel for detection (such as `SExtractor`) (Hoekstra, Kannawadi, and Thomas D. Kitching, 2021). An alignment, or misalignment, with the PSF can produce the same effect. Then, a selection bias (see section 2.5.4) can arise since some cuts must be set to ensure that detected objects can be well-measured, typically on SNR. Also, the Point-Spread Function (PSF, see section below) can impact the detection as shown on fig. 2.14. Objects appearing separated on data taken from space (HST images on the second and fourth columns) can be detected as a single one, on images taken from a ground telescope (Subaru images on the first and third columns). This difference is due to the fact that space data do not suffer from the atmospheric PSF. The superposition of sources on an image is an effect called *blending*, and if the objects appear so close that the pipeline detects one object instead of two, it becomes an *unrecognised (or ambiguous) blend*. A large PSF, or objects aligned with the PSF are likely to produce unrecognised blends and to introduce a bias. Indeed, doing shape measurement on these objects would lead to a wrong

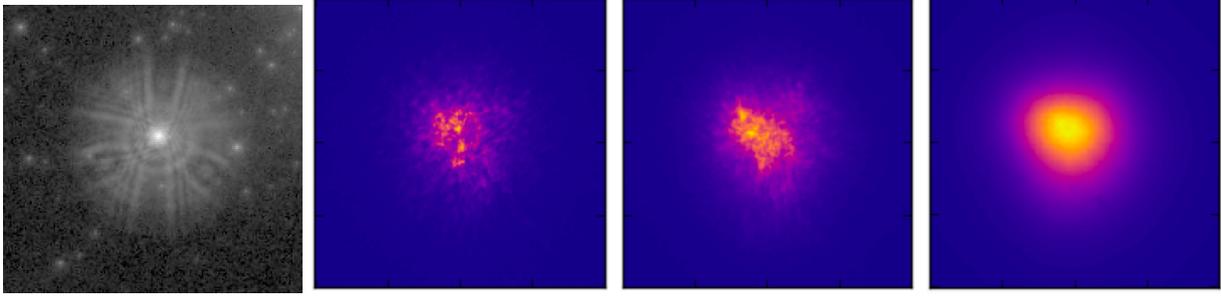


Figure 2.15: (left) The point spread function of Hubble Space Telescope’s WFPC camera. Credit: NASA. (3 panels on the right) Atmospheric PSF of the integrated flux in a set of the Differential Speckle Survey Instrument (DSSI) exposures, from left to right: 0.06s, 0.50s, and 62s. Credit: Hébert, Macintosh, and Burchat, 2018.

ellipticity estimation. This can be a dominant systematic for shape measurement as 14% of detected objects in LSST are expected to be unrecognised blends (Dawson et al., 2016), leading to an increase of 14% in shear noise.

If the different objects of a blend are detected then it is a *recognised blend*. It is then possible to discard the scene, or to apply a deblending algorithm in order to separate the different objects. Deblending permit to obtain higher statistics for shear measurement, but it can introduce a bias of a few percent, if the deblending algorithm produces objects with a preferred orientation (Arcelin et al., 2021). Calibrating such bias could however be possible using the METACALIBRATION algorithm (see section 2.5.5) (E. Huff and Mandelbaum, 2017; Sheldon and E. M. Huff, 2017) or simulations. I will describe more thoroughly the issue of blending and the different existing avenues in chapter 3. In the recent DES-Y3 results, a shear bias of about 2% was attributed to blending (MacCrann et al., 2020).

Notably, these biases due to blending are only a problem if the overlapped objects are located a different redshifts. In that case they are subject to different amounts of shear and must be separated to be classified in the correct redshift bins. If they are at a similar redshift and falling in the same redshift bin, blends would only introduce noise or model bias (see below) and could be probably calibrated via the METACALIBRATION algorithm. Biases introduced by blending are discussed in section 3.2.

As mentioned, this step is highly dependent on the survey resolution and consequently on the PSF, which can also induce shear bias, as described below.

2.5.2 Point-Spread Function modelling

The Point-Spread Function (PSF) is the response of an imaging system to a point source object. Each element through which the light travels generates a different response. For example, the optics of the instrument as well as the atmosphere for ground based experiments generate a specific PSF (see fig. 2.15). These PSFs can alter the objects shapes and corrections must be applied in order to take this effect into account.

As mentioned above, the PSF can impact the detection process, but it can also induce a bias during the measurement of galaxy ellipticities. Indeed, a sheared PSF could modify galaxies shapes and, if not accurately taken into account, introduce a bias in the direction of the PSF’s shear, particularly for small galaxies. This bias can reach a few percents (Mandelbaum et al., 2018b).

All methods designed to remove the impact of the PSF on shear estimates comes with the

hypothesis that the PSF is known. Consequently a good PSF modelling is mandatory. Stars are natural point sources that can be used to create a PSF model to interpolate at any position in the sky. However, the selection of usable stars does not come without issues, the brighter-fatter effect⁵ (Antilogus et al., 2014) for example, and the geometry of the camera must be taken into account (discontinuities between CCDs).

The optical PSF is already quite difficult to model as it needs to consider a lot of different effects from sensors and from the optical elements that compose the telescope (effect of temperature, tree rings, etc). The atmospheric PSF adds another layer of difficulty as it varies from an exposure to the other (see fig. 2.15). In case of surveys with multiple exposures of the same sky area, the PSF is usually modeled for each exposure and then combined rather than modeled from the co-add images of stars (Miller et al., 2013; Bosch et al., 2018). Indeed, variation between different exposures as well as dithers to fill the gaps between CCDs makes the PSF modelling from a co-add almost impossible (Bosch et al., 2018).

There are many ways of modeling the PSF. Most modeling algorithms employ polynomials both to fit the PSF models and for the interpolation part. For example, PSFEx (Bertin, 2013), used for both DES and HSC (M. Jarvis et al., 2016; Bosch et al., 2018), uses this kind of modeling. Other methods are developed and attention is recently turning to the problem of chromatic PSF: the PSFs evolve as a function of wavelength. DES-Y3 used the PIFF (PSFs In the Full FOV) algorithm (Jarvis et al., 2021) which relies on different models to fit the PSF and different interpolation schemes. It promisingly shows very small systematic errors, drastically improving the results compared to DES-Y1. Contrary to other systematics, PSF modeling presents the advantage of having well defined null tests that can point out modeling errors, allowing for corresponding corrections.

In most of the cases, the PSF correction is performed within the ellipticity measurement methods, which are described in the next section.

2.5.3 Ellipticity measurement

Once the objects are identified and the PSF modeled, one can perform the measurement of galaxy ellipticities. The measurement of the shear depends highly on the estimation of the galaxy shapes but a precise measurement of individual shapes is not particularly necessary as the goal is to obtain an unbiased estimator for a sample of galaxies. Mostly two kinds of methods have been developed to measure galaxy ellipticities from images: direct estimation of ellipticity on the image (which are mostly moment-based), and model-fitting methods. For this second kind of methods, the ellipticity estimation is performed using the fitted models.

Moment-based methods

As seen previously it is possible to define the ellipticity as a function of quadrupole moments of the surface brightness. Those quadrupole moments depends on the integration range set by the surface brightness of the object through the weight function $q_I(I(\theta))$ (see eq. (2.30)). This function is difficult to define for real data since neighbouring galaxies can modify the shape of the integration range and the discreteness of pixels makes its definition non-linear. It is usually defined as a function of the galaxy image coordinates and the galaxy’s size. Once this function is defined, it is possible to measure weighted second moments on the image previously deconvolved by the measured PSF. This was firstly done by Kaiser, Squires, and Broadhurst (1995b) (KSB) and improved in several

⁵The brighter-fatter effect is the linear increase of the size of point sources, mostly stars, as a function of their flux.

ways, for example in G. Luppino and Kaiser (1997), Hoekstra et al. (1998) and G. M. Bernstein and M. Jarvis (2002).

Moment-based methods are used for cosmological surveys: the HSC survey implemented the re-Gaussianization PSF correction method (Hirata and Seljak, 2003) and used it to create their first-year catalogue (Bosch et al., 2018; Mandelbaum et al., 2018a). The same method is implemented in the LSST science pipeline and used in chapter 7 and chapter 8.

Model-fitting methods

Model-fitting methods are using analytic models (for example 2D-gaussian, exponential, or Sérsic (1963)) to forward fit galaxies. The galaxy light profile, convolved with a PSF profile, is fitted with the ellipticity as parameter, to approximate the true ellipticity. Most of known model-fitting methods use two profiles, one for the bulge (central bright part of the galaxy) and one for the disk of the galaxy (external and more diffuse part).

Several surveys already used model-fitting methods: the CFHTLenS survey used the `lensfit` method (Miller et al., 2007; T. D. Kitching et al., 2008; T. D. Kitching et al., 2008) and DES used the `ngmix` software⁶ (M. Gatti et al., 2020).

Bayesian inference methods

More recently Bayesian inference techniques were proposed to realise the shear measurement. Contrary to previously presented methods, they do not need to go through individual shape measurements.

M. D. Schneider et al. (2015) presented a method to do hierarchical probabilistic inference of cosmic shear. It is a forward modeling approach with the idea to construct a shear posterior by marginalizing over all intrinsic galaxy properties that contribute to the pixel data. Nevertheless, it needs to assume some models for the galaxy profile that can be Sérsic (1963) for example.

A second example is the Bayesian Fourier Domain (BFD, Gary M. Bernstein and Armstrong, 2014). This is a model-free algorithm which does not use individual galaxy shape estimates to measure cosmic shear. Instead the image is compressed into vectors of moments in order to estimate the probability of the observed data to be produced for a certain amount of distortion along the line of sight.

None of these methods has been applied to real data yet, but the BFD method is very promising and is expected to be applied to the DES-Y6 (six years of data from DES) weak lensing catalog.

Shear bias from measurement methods

However, each type of methods can introduce a shear bias. For example, a model-fitting algorithm might fail at representing the entire distribution of galaxy shapes if the proposed models are not complex enough, which can lead to a bias called *model bias*. This impacts also moment-based methods via the window function, or the hierarchical probabilistic shear inference method through the choice of models for galaxy profiles. All these techniques can suffer from model parametrisation: the choice of analytic model and number of parameters is crucial to be flexible enough in order to capture the entire range of possible light profiles. Not enough complex models can lead to underfitting and too many parameters can lead to overfitting. In both cases the shear will suffer model biases (Voigt and Bridle, 2010, G. M. Bernstein, 2010).

⁶<https://github.com/esheldon/ngmix>

Another bias that affects all measurement methods is the *noise bias*. It comes from the data and does not depend on the measurement method: noise captured by the instrument can interfere with the shape measurement creating noise in the shear estimation. As most of weak lensing signal is supported by high magnitude, or low SNR, galaxies, and as a large amount of galaxies will be blended in LSST data (see section 6.1.2), this bias can have a high impact on cosmic shear estimation. Noise bias can be of the order of tenth to few percents (for example, see Melchior and Viola, 2012 or Sanchez et al., 2021).

Nevertheless, both noise and model biases can be calibrated by the METACALIBRATION algorithm (E. Huff and Mandelbaum, 2017; Sheldon and E. M. Huff, 2017) described below.

Finally, most of these methods perform their measurement on isolated galaxy images and rely on deblending techniques (see section 3.3.2) to tackle the overlapping of sources. In chapter 8 we propose a new method based on a Bayesian neural network (BNN) (see part II) to perform galaxy shape measurements directly from the images, whether the galaxies are blended or not.

2.5.4 Selection bias

Selection bias can be introduced at different step of the pipeline. As we have seen, at the detection stage for example, the removal of close pairs of galaxies can introduce a selection bias.

In the end of the pipeline, to select objects that were well-measured and can enter the weak lensing analysis, cuts must often be done on the processed data. They can also be a source of *selection bias*. For example, if the selection probability depends on the galaxy’s size, its alignment (or misalignment) with the shear will increase (or decrease) the probability of entering the sample, creating a bias. This bias can have a significant impact of the 1 to 5% of several percent as shown by Mandelbaum et al. (2018b) for the HSC survey. This bias can fully be calibrated with simulations.

2.5.5 Shear Calibration

To calibrate this biases several solutions exist. First, it is possible to evaluate some of these biases using simulation. For example, testing a shear measurement pipeline on simulations of realistic galaxies can give an estimate of the model bias and provide the corresponding calibration. Selection bias as well as blending biases can also be characterized and calibrated this way (Mandelbaum et al., 2018b; MacCrann et al., 2020).

Then, METACALIBRATION (E. Huff and Mandelbaum, 2017; Sheldon and E. M. Huff, 2017) allows for the calibration of model and noise biases by computing the shear response of the shape measurement algorithm, using the observed data. Given an ensemble of measurements e_i and response R_{γ_i} , an estimated mean shear can be written

$$\langle \gamma^{est} \rangle \approx \langle R_{\gamma} \rangle^{-1} \langle e \rangle \quad (2.37)$$

with the average taken over the ensemble of measurements. The shear $\langle \gamma^{est} \rangle$ is a weighted mean of the measured ellipticities e , as a function of the shear response defined as

$$R_{\gamma_{i,j}} = \frac{e_i^+ - e_i^-}{\Delta \gamma_j} \quad (2.38)$$

with e_i^+ and e_i^- the i -th components of the ellipticities measured on images sheared by an artificial shear with j -th component equal to $\pm \Delta \gamma_j$. To perform this shearing, the initial image is deconvolved by the PSF, sheared and reconvolved by the PSF. The response can then be integrated in the computation of the shear correlation functions for instance. If selection cuts are applied, it is

possible to calibrate the selection bias selecting on sheared measurements and re-computing the new response $\langle R_s \rangle$. The response matrix $\langle R_s \rangle$ becomes $\langle R \rangle = \langle R_\gamma \rangle + \langle R_s \rangle$. The mean shear can then be computed as in eq. (2.37) replacing $\langle R_\gamma \rangle$ by $\langle R \rangle$.

METACALIBRATION is nowadays widely recognised as very effective to correct for shear biases and has been used in weak lensing analyses of DES-Y1 and DES-Y3 for example (Zuntz et al., 2018; DES Collaboration et al., 2021). METADETECTION (Sheldon et al., 2020) is an extension of METACALIBRATION and adds a correction for the detection bias and specific cases of blending (see section 3.3.1). Both METACALIBRATION and METADETECTION are very powerful and work extremely well on isolated galaxy images. As mentioned, METADETECTION does a first step toward the calibration of the blending bias but as discussed in section 3.3.1, it is not sufficient to deal with blended objects at different redshifts (or in different redshift bins).

Lots of solutions already exist to address the different kinds of biases that can be introduced by a weak lensing pipeline. However, LSST asks for an unprecedented level of precision and will be subject to some systematics, especially *blending*, at a level never seen before. Efforts are particularly necessary on the latter as it can be the dominant systematic for shear measurement for LSST (Dawson et al., 2016). Next chapter will describe more deeply the impact of blending for future cosmological surveys as well as the different solutions that exist nowadays to address it.

Bibliography

- Abbott, T. M. C. et al. (Aug. 2018). “Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing”. In: *PRD* 98.4, 043526, p. 043526. DOI: [10.1103/PhysRevD.98.043526](https://doi.org/10.1103/PhysRevD.98.043526). arXiv: [1708.01530](https://arxiv.org/abs/1708.01530) [[astro-ph.CO](#)].
- Abell, Paul A et al. (2009). “Lsst science book, version 2.0”. In: *arXiv preprint arXiv:0912.0201*.
- Ahumada, Romina et al. (July 2020). “The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra”. In: *ApJS* 249.1, 3, p. 3. DOI: [10.3847/1538-4365/ab929e](https://doi.org/10.3847/1538-4365/ab929e). arXiv: [1912.02905](https://arxiv.org/abs/1912.02905) [[astro-ph.GA](#)].
- Aihara, Hiroaki et al. (Jan. 2018). “The Hyper Suprime-Cam SSP Survey: Overview and survey design”. In: *PASJ* 70, S4, S4. DOI: [10.1093/pasj/psx066](https://doi.org/10.1093/pasj/psx066). arXiv: [1704.05858](https://arxiv.org/abs/1704.05858) [[astro-ph.IM](#)].
- Albrecht, Andreas et al. (Sept. 2006). “Report of the Dark Energy Task Force”. In: *arXiv e-prints*, astro-ph/0609591, astro-ph/0609591. arXiv: [astro-ph/0609591](https://arxiv.org/abs/astro-ph/0609591) [[astro-ph](#)].
- Amon, A. et al. (May 2021). “Dark Energy Survey Year 3 Results: Cosmology from Cosmic Shear and Robustness to Data Calibration”. In: *arXiv e-prints*, arXiv:2105.13543, arXiv:2105.13543. arXiv: [2105.13543](https://arxiv.org/abs/2105.13543) [[astro-ph.CO](#)].
- Antilogus, Pierre et al. (2014). “The brighter-fatter effect and pixel correlations in CCD sensors”. In: *Journal of Instrumentation* 9.03, p. C03048.
- Arcelin, Bastien et al. (Jan. 2021). “Deblending galaxies with variational autoencoders: A joint multiband, multi-instrument approach”. In: *MNRAS* 500.1, pp. 531–547. DOI: [10.1093/mnras/staa3062](https://doi.org/10.1093/mnras/staa3062). arXiv: [2005.12039](https://arxiv.org/abs/2005.12039) [[astro-ph.IM](#)].
- Asgari, Marika et al. (Jan. 2021). “KiDS-1000 cosmology: Cosmic shear constraints and comparison between two point statistics”. In: *A&A* 645, A104, A104. DOI: [10.1051/0004-6361/202039070](https://doi.org/10.1051/0004-6361/202039070). arXiv: [2007.15633](https://arxiv.org/abs/2007.15633) [[astro-ph.CO](#)].
- Bacon, David J., Alexandre R. Refregier, and Richard S. Ellis (Oct. 2000). “Detection of weak gravitational lensing by large-scale structure”. In: *MNRAS* 318.2, pp. 625–640. DOI: [10.1046/j.1365-8711.2000.03851.x](https://doi.org/10.1046/j.1365-8711.2000.03851.x). arXiv: [astro-ph/0003008](https://arxiv.org/abs/astro-ph/0003008) [[astro-ph](#)].
- Bartelmann, M. and P. Schneider (Jan. 2001). “Weak gravitational lensing”. In: *Phys. Rep.* 340.4-5, pp. 291–472. DOI: [10.1016/S0370-1573\(00\)00082-X](https://doi.org/10.1016/S0370-1573(00)00082-X). arXiv: [astro-ph/9912508](https://arxiv.org/abs/astro-ph/9912508) [[astro-ph](#)].
- Bernardeau, Francis, Takahiro Nishimichi, and Atsushi Taruya (Dec. 2014). “Cosmic shear full nulling: sorting out dynamics, geometry and systematics”. In: *MNRAS* 445.2, pp. 1526–1537. DOI: [10.1093/mnras/stu1861](https://doi.org/10.1093/mnras/stu1861). arXiv: [1312.0430](https://arxiv.org/abs/1312.0430) [[astro-ph.CO](#)].
- Bernstein, G. M. (2010). “Shape measurement biases from underfitting and ellipticity gradients”. In: *Monthly Notices of the Royal Astronomical Society* 406.4, pp. 2793–2804.
- Bernstein, G. M. and M. Jarvis (Feb. 2002). “Shapes and Shears, Stars and Smears: Optimal Measurements for Weak Lensing”. In: *AJ* 123.2, pp. 583–618. DOI: [10.1086/338085](https://doi.org/10.1086/338085). arXiv: [astro-ph/0107431](https://arxiv.org/abs/astro-ph/0107431) [[astro-ph](#)].
- Bernstein, Gary M. and Robert Armstrong (Feb. 2014). “Bayesian lensing shear measurement”. In: *MNRAS* 438.2, pp. 1880–1893. DOI: [10.1093/mnras/stt2326](https://doi.org/10.1093/mnras/stt2326). arXiv: [1304.1843](https://arxiv.org/abs/1304.1843) [[astro-ph.CO](#)].

- Bertin, Emmanuel (2013). “PSFEx: Point Spread Function Extractor”. In: *Astrophysics Source Code Library*, ascl-1301.
- Bertin, Emmanuel and Stephane Arnouts (1996). “SExtractor: Software for source extraction”. In: *Astronomy and astrophysics supplement series* 117.2, pp. 393–404.
- Blazek, Jonathan A. et al. (Nov. 2019). “Beyond linear galaxy alignments”. In: *Phys. Rev. D* 100.10, 103506, p. 103506. DOI: [10.1103/PhysRevD.100.103506](https://doi.org/10.1103/PhysRevD.100.103506). arXiv: [1708.09247](https://arxiv.org/abs/1708.09247) [[astro-ph.CO](#)].
- Bonnet, H and Y Mellier (1995). “Statistical analysis of weak gravitational shear in the extended periphery of rich galaxy clusters.” In: *Astronomy and Astrophysics* 303, p. 331.
- Bosch, James et al. (2018). “The Hyper Suprime-Cam software pipeline”. In: *Publications of the Astronomical Society of Japan* 70.SP1, S5.
- Brammer, Gabriel B., Pieter G. van Dokkum, and Paolo Coppi (Oct. 2008). “EAZY: A Fast, Public Photometric Redshift Code”. In: *ApJ* 686.2, pp. 1503–1513. DOI: [10.1086/591786](https://doi.org/10.1086/591786). arXiv: [0807.1533](https://arxiv.org/abs/0807.1533) [[astro-ph](#)].
- Castro, P. G., A. F. Heavens, and T. D. Kitching (July 2005). “Weak lensing analysis in three dimensions”. In: *PRD* 72.2, 023516, p. 023516. DOI: [10.1103/PhysRevD.72.023516](https://doi.org/10.1103/PhysRevD.72.023516). arXiv: [astro-ph/0503479](https://arxiv.org/abs/astro-ph/0503479) [[astro-ph](#)].
- Dawson, William A. et al. (Jan. 2016). “The Ellipticity Distribution of Ambiguously Blended Objects”. In: *ApJ* 816.1, 11, p. 11. DOI: [10.3847/0004-637X/816/1/11](https://doi.org/10.3847/0004-637X/816/1/11). arXiv: [1406.1506](https://arxiv.org/abs/1406.1506) [[astro-ph.CO](#)].
- de Jong, J. T. A. et al. (Dec. 2013). “The Kilo-Degree Survey”. In: *The Messenger* 154, pp. 44–46.
- de Jong, R. S. et al. (Mar. 2019). “4MOST: Project overview and information for the First Call for Proposals”. In: *The Messenger* 175, pp. 3–11. DOI: [10.18727/0722-6691/5117](https://doi.org/10.18727/0722-6691/5117). arXiv: [1903.02464](https://arxiv.org/abs/1903.02464) [[astro-ph.IM](#)].
- DES Collaboration et al. (May 2021). “Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing”. In: *arXiv e-prints*, arXiv:2105.13549, arXiv:2105.13549. arXiv: [2105.13549](https://arxiv.org/abs/2105.13549) [[astro-ph.CO](#)].
- Gatti, M. et al. (Nov. 2020). “Dark Energy Survey Year 3 Results: Weak Lensing Shape Catalogue”. In: *arXiv e-prints*, arXiv:2011.03408, arXiv:2011.03408. arXiv: [2011.03408](https://arxiv.org/abs/2011.03408) [[astro-ph.CO](#)].
- Hamana, Takashi et al. (Feb. 2020). “Cosmological constraints from cosmic shear two-point correlation functions with HSC survey first-year data”. In: *PASJ* 72.1, 16, p. 16. DOI: [10.1093/pasj/psz138](https://doi.org/10.1093/pasj/psz138). arXiv: [1906.06041](https://arxiv.org/abs/1906.06041) [[astro-ph.CO](#)].
- Heavens, A. F. (Aug. 2003). “3D weak lensing”. In: *MNRAS* 343.4, pp. 1327–1334. DOI: [10.1046/j.1365-8711.2003.06780.x](https://doi.org/10.1046/j.1365-8711.2003.06780.x). arXiv: [astro-ph/0304151](https://arxiv.org/abs/astro-ph/0304151) [[astro-ph](#)].
- Hébert, Claire-Alice, Bruce Macintosh, and Patricia R. Burchat (July 2018). “Characterization of atmospheric turbulence for the Large Synoptic Survey Telescope”. In: *Ground-based and Airborne Telescopes VII*. Ed. by Heather K. Marshall and Jason Spyromilio. Vol. 10700. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 107005E. DOI: [10.1117/12.2314311](https://doi.org/10.1117/12.2314311). arXiv: [1807.09337](https://arxiv.org/abs/1807.09337) [[astro-ph.IM](#)].
- Heymans, Catherine and A. F. Heavens (Mar. 2003). “Weak gravitational lensing: reducing the contamination by intrinsic alignments”. In: *MNRAS* 339.3, pp. 711–720. DOI: [10.1046/j.1365-8711.2003.06213.x](https://doi.org/10.1046/j.1365-8711.2003.06213.x). arXiv: [astro-ph/0208220](https://arxiv.org/abs/astro-ph/0208220) [[astro-ph](#)].
- Hildebrandt, H. et al. (Jan. 2020). “KiDS+VIKING-450: Cosmic shear tomography with optical and infrared data”. In: *A&A* 633, A69, A69. DOI: [10.1051/0004-6361/201834878](https://doi.org/10.1051/0004-6361/201834878). arXiv: [1812.06076](https://arxiv.org/abs/1812.06076) [[astro-ph.CO](#)].
- Hirata, Christopher and Uroš Seljak (2003). “Shear calibration biases in weak-lensing surveys”. In: *Monthly Notices of the Royal Astronomical Society* 343.2, pp. 459–480.

- Hoekstra, Henk and Bhuvnesh Jain (Nov. 2008). “Weak Gravitational Lensing and Its Cosmological Applications”. In: *Annual Review of Nuclear and Particle Science* 58.1, pp. 99–123. DOI: [10.1146/annurev.nucl.58.110707.171151](https://doi.org/10.1146/annurev.nucl.58.110707.171151). arXiv: [0805.0139](https://arxiv.org/abs/0805.0139) [astro-ph].
- Hoekstra, Henk, Arun Kannawadi, and Thomas D. Kitching (Feb. 2021). “Accounting for object detection bias in weak gravitational lensing studies”. In: *A&A* 646, A124, A124. DOI: [10.1051/0004-6361/202038998](https://doi.org/10.1051/0004-6361/202038998). arXiv: [2010.04178](https://arxiv.org/abs/2010.04178) [astro-ph.CO].
- Hoekstra, Henk et al. (1998). “Weak lensing analysis of CL 1358+ 62 using Hubble Space Telescope observations”. In: *The Astrophysical Journal* 504.2, p. 636.
- Hu, Wayne (1999). “Power spectrum tomography with weak lensing”. In: *The Astrophysical Journal Letters* 522.1, p. L21.
- Huff, Eric and Rachel Mandelbaum (Feb. 2017). “Metacalibration: Direct Self-Calibration of Biases in Shear Measurement”. In: *arXiv e-prints*, arXiv:1702.02600, arXiv:1702.02600. arXiv: [1702.02600](https://arxiv.org/abs/1702.02600) [astro-ph.CO].
- Jarvis, M. et al. (Aug. 2016). “The DES Science Verification weak lensing shear catalogues”. In: *MNRAS* 460.2, pp. 2245–2281. DOI: [10.1093/mnras/stw990](https://doi.org/10.1093/mnras/stw990). arXiv: [1507.05603](https://arxiv.org/abs/1507.05603) [astro-ph.IM].
- Jarvis, M et al. (2021). “Dark Energy Survey year 3 results: point spread function modelling”. In: *Monthly Notices of the Royal Astronomical Society* 501.1, pp. 1282–1299.
- Kaiser, Nick, Gordon Squires, and Tom Broadhurst (Aug. 1995a). “A Method for Weak Lensing Observations”. In: *ApJ* 449, p. 460. DOI: [10.1086/176071](https://doi.org/10.1086/176071). arXiv: [astro-ph/9411005](https://arxiv.org/abs/astro-ph/9411005) [astro-ph].
- (1995b). “A Method for Weak Lensing Observations”. In: *The Astrophysical Journal* 449, p. 460.
- Kaiser, Nick, Gillian Wilson, and Gerard A. Luppino (Mar. 2000). “Large-Scale Cosmic Shear Measurements”. In: *arXiv e-prints*, astro-ph/0003338, astro-ph/0003338. arXiv: [astro-ph/0003338](https://arxiv.org/abs/astro-ph/0003338) [astro-ph].
- Kilbinger, Martin (July 2015). “Cosmology with cosmic shear observations: a review”. In: *Reports on Progress in Physics* 78.8, 086901, p. 086901. DOI: [10.1088/0034-4885/78/8/086901](https://doi.org/10.1088/0034-4885/78/8/086901). arXiv: [1411.0115](https://arxiv.org/abs/1411.0115) [astro-ph.CO].
- Kitching, T. D., A. F. Heavens, and L. Miller (June 2011). “3D photometric cosmic shear”. In: *MNRAS* 413.4, pp. 2923–2934. DOI: [10.1111/j.1365-2966.2011.18369.x](https://doi.org/10.1111/j.1365-2966.2011.18369.x). arXiv: [1007.2953](https://arxiv.org/abs/1007.2953) [astro-ph.CO].
- Kitching, T. D. et al. (Oct. 2008). “Bayesian galaxy shape measurement for weak lensing surveys - II. Application to simulations”. In: *MNRAS* 390.1, pp. 149–167. DOI: [10.1111/j.1365-2966.2008.13628.x](https://doi.org/10.1111/j.1365-2966.2008.13628.x). arXiv: [0802.1528](https://arxiv.org/abs/0802.1528) [astro-ph].
- Kitching, T. D. et al. (Aug. 2014). “3D cosmic shear: cosmology from CFHTLenS”. In: *MNRAS* 442.2, pp. 1326–1349. DOI: [10.1093/mnras/stu934](https://doi.org/10.1093/mnras/stu934). arXiv: [1401.6842](https://arxiv.org/abs/1401.6842) [astro-ph.CO].
- Krause, Elisabeth, Tim Eifler, and Jonathan Blazek (Feb. 2016). “The impact of intrinsic alignment on current and future cosmic shear surveys”. In: *MNRAS* 456.1, pp. 207–222. DOI: [10.1093/mnras/stv2615](https://doi.org/10.1093/mnras/stv2615). arXiv: [1506.08730](https://arxiv.org/abs/1506.08730) [astro-ph.CO].
- Levi, Michael et al. (Sept. 2019). “The Dark Energy Spectroscopic Instrument (DESI)”. In: *Bulletin of the American Astronomical Society*. Vol. 51, p. 57. arXiv: [1907.10688](https://arxiv.org/abs/1907.10688) [astro-ph.IM].
- Liu, Runjing, Jon D. McAuliffe, and Jeffrey Regier (Feb. 2021). “Variational Inference for Deblending Crowded Starfields”. In: *arXiv e-prints*, arXiv:2102.02409, arXiv:2102.02409. arXiv: [2102.02409](https://arxiv.org/abs/2102.02409) [astro-ph.IM].
- Luppino, GA and Nick Kaiser (1997). “Detection of weak lensing by a cluster of galaxies at $z=0.83$ ”. In: *The Astrophysical Journal* 475.1, p. 20.
- MacCrann, N. et al. (Dec. 2020). “DES Y3 results: Blending shear and redshift biases in image simulations”. In: *arXiv e-prints*, arXiv:2012.08567, arXiv:2012.08567. arXiv: [2012.08567](https://arxiv.org/abs/2012.08567) [astro-ph.CO].

- Mandelbaum, Rachel (Sept. 2018). “Weak Lensing for Precision Cosmology”. In: ARA&A 56, pp. 393–433. DOI: [10.1146/annurev-astro-081817-051928](https://doi.org/10.1146/annurev-astro-081817-051928). arXiv: [1710.03235](https://arxiv.org/abs/1710.03235) [[astro-ph.CO](#)].
- Mandelbaum, Rachel et al. (Jan. 2018a). “The first-year shear catalog of the Subaru Hyper Suprime-Cam Subaru Strategic Program Survey”. In: PASJ 70, S25, S25. DOI: [10.1093/pasj/psx130](https://doi.org/10.1093/pasj/psx130). arXiv: [1705.06745](https://arxiv.org/abs/1705.06745) [[astro-ph.CO](#)].
- Mandelbaum, Rachel et al. (Dec. 2018b). “Weak lensing shear calibration with simulations of the HSC survey”. In: MNRAS 481.3, pp. 3170–3195. DOI: [10.1093/mnras/sty2420](https://doi.org/10.1093/mnras/sty2420). arXiv: [1710.00885](https://arxiv.org/abs/1710.00885) [[astro-ph.CO](#)].
- Melchior, P. and M. Viola (Aug. 2012). “Means of confusion: how pixel noise affects shear estimates for weak gravitational lensing”. In: MNRAS 424.4, pp. 2757–2769. DOI: [10.1111/j.1365-2966.2012.21381.x](https://doi.org/10.1111/j.1365-2966.2012.21381.x). arXiv: [1204.5147](https://arxiv.org/abs/1204.5147) [[astro-ph.IM](#)].
- Miller, L. et al. (Nov. 2007). “Bayesian galaxy shape measurement for weak lensing surveys - I. Methodology and a fast-fitting algorithm”. In: MNRAS 382.1, pp. 315–324. DOI: [10.1111/j.1365-2966.2007.12363.x](https://doi.org/10.1111/j.1365-2966.2007.12363.x). arXiv: [0708.2340](https://arxiv.org/abs/0708.2340) [[astro-ph](#)].
- Miller, L. et al. (Mar. 2013). “Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey”. In: MNRAS 429.4, pp. 2858–2880. DOI: [10.1093/mnras/sts454](https://doi.org/10.1093/mnras/sts454). arXiv: [1210.8201](https://arxiv.org/abs/1210.8201) [[astro-ph.CO](#)].
- Myles, J. et al. (Dec. 2020). “Dark Energy Survey Year 3 Results: Redshift Calibration of the Weak Lensing Source Galaxies”. In: *arXiv e-prints*, arXiv:2012.08566, arXiv:2012.08566. arXiv: [2012.08566](https://arxiv.org/abs/2012.08566) [[astro-ph.CO](#)].
- Paillassa, Maxime (Oct. 2020). “Robust detection of astronomical sources using convolutional neural networks”. Theses. Université de Bordeaux. URL: <https://tel.archives-ouvertes.fr/tel-03161521>.
- Pasquet, Johanna et al. (Jan. 2019). “Photometric redshifts from SDSS images using a convolutional neural network”. In: A&A 621, A26, A26. DOI: [10.1051/0004-6361/201833617](https://doi.org/10.1051/0004-6361/201833617). arXiv: [1806.06607](https://arxiv.org/abs/1806.06607) [[astro-ph.IM](#)].
- Peter, Patrick and Jean-Philippe Uzan (2013). *Primordial cosmology*. Oxford University Press.
- Planck Collaboration et al. (Sept. 2020). “Planck 2018 results. VI. Cosmological parameters”. In: A&A 641, A6, A6. DOI: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910). arXiv: [1807.06209](https://arxiv.org/abs/1807.06209) [[astro-ph.CO](#)].
- Prat, J. et al. (May 2021). “Dark Energy Survey Year 3 Results: High-precision measurement and modeling of galaxy-galaxy lensing”. In: *arXiv e-prints*, arXiv:2105.13541, arXiv:2105.13541. arXiv: [2105.13541](https://arxiv.org/abs/2105.13541) [[astro-ph.CO](#)].
- Regier, Jeffrey et al. (2015). “Celeste: Variational inference for a generative model of astronomical images”. In: *International Conference on Machine Learning*. PMLR, pp. 2095–2103.
- Sanchez, Javier et al. (Mar. 2021). “Effects of overlapping sources on cosmic shear estimation: Statistical sensitivity and pixel-noise bias”. In: *arXiv e-prints*, arXiv:2103.02078, arXiv:2103.02078. arXiv: [2103.02078](https://arxiv.org/abs/2103.02078) [[astro-ph.CO](#)].
- Schneider, Michael D. et al. (July 2015). “Hierarchical Probabilistic Inference of Cosmic Shear”. In: ApJ 807.1, 87, p. 87. DOI: [10.1088/0004-637X/807/1/87](https://doi.org/10.1088/0004-637X/807/1/87). arXiv: [1411.2608](https://arxiv.org/abs/1411.2608) [[astro-ph.CO](#)].
- Schneider, P. et al. (Dec. 2002). “Analysis of two-point statistics of cosmic shear. I. Estimators and covariances”. In: A&A 396, pp. 1–19. DOI: [10.1051/0004-6361:20021341](https://doi.org/10.1051/0004-6361:20021341). arXiv: [astro-ph/0206182](https://arxiv.org/abs/astro-ph/0206182) [[astro-ph](#)].
- Schneider, Peter, Christopher Kochanek, and Joachim Wambsganss (2006). *Gravitational lensing: strong, weak and micro: Saas-Fee advanced course 33*. Vol. 33. Springer Science & Business Media.
- Sérsic, J. L. (Feb. 1963). “Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy”. In: *Boletín de la Asociación Argentina de Astronomía La Plata Argentina* 6, pp. 41–43.

- Sheldon, Erin S. and Eric M. Huff (May 2017). “Practical Weak-lensing Shear Measurement with Metacalibration”. In: *ApJ* 841.1, 24, p. 24. DOI: [10.3847/1538-4357/aa704b](https://doi.org/10.3847/1538-4357/aa704b). arXiv: [1702.02601](https://arxiv.org/abs/1702.02601) [[astro-ph.CO](#)].
- Sheldon, Erin S. et al. (Oct. 2020). “Mitigating Shear-dependent Object Detection Biases with Metacalibration”. In: *ApJ* 902.2, 138, p. 138. DOI: [10.3847/1538-4357/abb595](https://doi.org/10.3847/1538-4357/abb595). arXiv: [1911.02505](https://arxiv.org/abs/1911.02505) [[astro-ph.CO](#)].
- The Dark Energy Survey Collaboration (Oct. 2005). “The Dark Energy Survey”. In: *arXiv e-prints*, astro-ph/0510346, astro-ph/0510346. arXiv: [astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346) [[astro-ph](#)].
- The LSST Dark Energy Science Collaboration et al. (Sept. 2018). “The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document”. In: *arXiv e-prints*, arXiv:1809.01669, arXiv:1809.01669. arXiv: [1809.01669](https://arxiv.org/abs/1809.01669) [[astro-ph.CO](#)].
- Troxel, M. A. and Mustapha Ishak (Feb. 2015). “The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology”. In: *Phys. Rep.* 558, pp. 1–59. DOI: [10.1016/j.physrep.2014.11.001](https://doi.org/10.1016/j.physrep.2014.11.001). arXiv: [1407.6990](https://arxiv.org/abs/1407.6990) [[astro-ph.CO](#)].
- Troxel, M. A. et al. (Aug. 2018). “Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear”. In: *PRD* 98.4, 043528, p. 043528. DOI: [10.1103/PhysRevD.98.043528](https://doi.org/10.1103/PhysRevD.98.043528). arXiv: [1708.01538](https://arxiv.org/abs/1708.01538) [[astro-ph.CO](#)].
- Van Waerbeke, L. et al. (June 2000). “Detection of correlated galaxy ellipticities from CFHT data: first evidence for gravitational lensing by large-scale structures”. In: *A&A* 358, pp. 30–44. arXiv: [astro-ph/0002500](https://arxiv.org/abs/astro-ph/0002500) [[astro-ph](#)].
- Voigt, L. M. and S. L. Bridle (May 2010). “Limitations of model-fitting methods for lensing shear estimation”. In: *MNRAS* 404.1, pp. 458–467. DOI: [10.1111/j.1365-2966.2010.16300.x](https://doi.org/10.1111/j.1365-2966.2010.16300.x). arXiv: [0905.4801](https://arxiv.org/abs/0905.4801) [[astro-ph.CO](#)].
- Wittman, D. (Aug. 2009). “What Lies Beneath: Using $p(z)$ to Reduce Systematic Photometric Redshift Errors”. In: *ApJ* 700.2, pp. L174–L177. DOI: [10.1088/0004-637X/700/2/L174](https://doi.org/10.1088/0004-637X/700/2/L174). arXiv: [0905.0892](https://arxiv.org/abs/0905.0892) [[astro-ph.CO](#)].
- Wittman, David M. et al. (May 2000). “Detection of weak gravitational lensing distortions of distant galaxies by cosmic dark matter at large scales”. In: *Nature* 405.6783, pp. 143–148. DOI: [10.1038/35012001](https://doi.org/10.1038/35012001). arXiv: [astro-ph/0003014](https://arxiv.org/abs/astro-ph/0003014) [[astro-ph](#)].
- Yao, Ji et al. (Oct. 2017). “Effects of self-calibration of intrinsic alignment on cosmological parameter constraints from future cosmic shear surveys”. In: *J. Cosmology Astropart. Phys.* 2017.10, 056, p. 056. DOI: [10.1088/1475-7516/2017/10/056](https://doi.org/10.1088/1475-7516/2017/10/056). arXiv: [1707.01072](https://arxiv.org/abs/1707.01072) [[astro-ph.CO](#)].
- Yao, Ji et al. (July 2020). “Separating the intrinsic alignment signal and the lensing signal using self-calibration in photo-z surveys with KiDS450 and KV450 Data”. In: *MNRAS* 495.4, pp. 3900–3919. DOI: [10.1093/mnras/staa1354](https://doi.org/10.1093/mnras/staa1354). arXiv: [1911.01582](https://arxiv.org/abs/1911.01582) [[astro-ph.CO](#)].
- York, Donald G. et al. (Sept. 2000). “The Sloan Digital Sky Survey: Technical Summary”. In: *AJ* 120.3, pp. 1579–1587. DOI: [10.1086/301513](https://doi.org/10.1086/301513). arXiv: [astro-ph/0006396](https://arxiv.org/abs/astro-ph/0006396) [[astro-ph](#)].
- Zuntz, J. et al. (Nov. 2018). “Dark Energy Survey Year 1 results: weak lensing shape catalogues”. In: *MNRAS* 481.1, pp. 1149–1182. DOI: [10.1093/mnras/sty2219](https://doi.org/10.1093/mnras/sty2219). arXiv: [1708.01533](https://arxiv.org/abs/1708.01533) [[astro-ph.CO](#)].

Chapter 3

The blending systematic for weak gravitational lensing

The superposition of astronomical sources on an image, an effect called *blending*, (see fig. 3.1) is a major systematic for weak gravitational analysis. In this chapter, I first explain why blending becomes a challenge and how to quantify it. Then, I showcase the different biases that blending can introduce and I conclude by presenting some of the current avenues to this issue.

3.1 Deep ground-based galaxy surveys: blending as a systematic

3.1.1 Deep surveys

As mentioned in section 1.3, the next generation of ground surveys will look deeper into the sky, and on broader sky areas, than any other survey before. Those two characteristics will increase the density of galaxies usable for weak lensing analysis, n_{eff} (it can also be defined as a function of redshift, $n_{\text{eff}}(z)$). As n_{eff} grows, so does the probability of observing overlapped objects and blending becomes a dominant source of systematics. LSST will observe to limiting magnitudes of $r \leq 27.5$ and $i \leq 26.8$ (Ivezić et al., 2019), whereas HSC, which is the current ground survey that is producing data that looks the most alike the future LSST data, will observe up to magnitude $i \approx 26$. Bosch et al. (2018) states that 58% of the detected objects in HSC will be blended. This number should reach 62% for LSST (Sanchez et al., 2021).

Other current galaxy surveys such as KiDS (de Jong et al., 2013) or DES (The Dark Energy Survey Collaboration, 2005) have lower limiting magnitudes (respectively $r \leq 25.2/i \leq 24.2$ and $i \leq 24$) and are consequently less impacted by blending.

This systematic mainly concerns ground surveys since space telescopes do not suffer from the atmospheric PSF (see section 2.5.1). Euclid and Roman should be only marginally impacted by this systematic.

Blending is consequently a major issue for deep ground-based surveys. Considering their statistics, discarding all blended objects is not a option and solutions must be provided. To propose appropriate methods it is first necessary to know what biases blending can introduce. We have already mentioned some of them on section 2.5.1 and we will detail them in section 3.2. First, I describe the existing metrics allowing to quantify blending in photometric data.

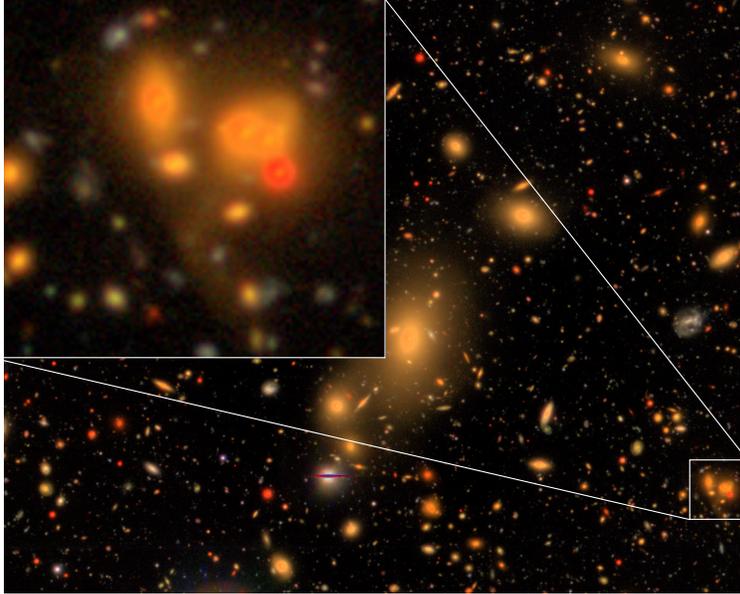


Figure 3.1: Part of the COSMOS field observed with HSC. The top left panel is a zoom of the small patch on the bottom right. Blends of bright and faint galaxies are visible. Bosch et al. (2018) estimated that 58% of detected objects are blended in HSC. Credit: Princeton University/HSC Project

3.1.2 Blending metrics

To define the percentages present above and to characterize blends, metrics are required. Several metrics can be found in the literature, I list them below. Unless stated otherwise, these metrics are defined for simulated data.

The distance between the studied galaxy and its neighbours can be a first metric to verify if a galaxy is blended. For example, Chang et al. (2013) (Chang et al., 2015) studied the impact of blending on n_{eff} for LSST. An object is considered blended when companions have their center in a circle of radius d around the center of the object. They used different distance criteria of 1, 2 and 3 arcseconds to estimate the percentage of blended galaxies that would be observed as a function of the raw galaxy number density, n .

Another way to define an object as blended is to compute a flux ratio to evaluate how much neighbours contribute to the observed galaxy flux. Working on simulated data allows to have an exact estimation of this quantity having access to the true galaxy light profile. For example, a source’s *purity* can be defined as (Sanchez et al., 2021):

$$\rho_i = \frac{\sum_p s_{ip} \cdot s_{ip}}{\sum_p (s_{ip} \sum_j s_{jp})} \quad (3.1)$$

with s_{ip} the signal value for the studied galaxy of index i in the pixel p . The sum over p represents the sum over all the pixels of the overlapping group, and the sum over j is the sum over all the sources in the overlapping group. This group is composed of the galaxy i and all sources that overlap with it (there can be none). Purity is then the ratio of the flux of the target galaxy i and the flux of the overlapping group. In our work, we defined a different metric using the definition proposed in P. Melchior et al. (2018), that can be related to purity. We called it the galaxy’s *blend*

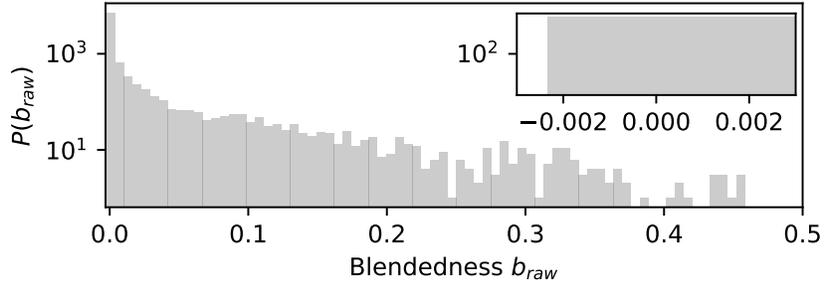


Figure 3.2: Blendedness distribution from a sample of image extracted from the DC2 simulation (LSST Dark Energy Science Collaboration et al., 2021). The top right panel shows a zoom around the zero value. Some values of blendedness are negative, implying an error in the deblending process.

rate (Arcelin et al., 2021):

$$\begin{aligned}
 b_i &= 1 - \frac{\sum_p s_{ip} \cdot s_{ip}}{\sum_p (s_{ip} \sum_j s_{jp})} \\
 &= 1 - \rho_i
 \end{aligned}
 \tag{3.2}$$

By construction the purity and the blend rate lay in the interval $[0,1]$. The purity is equal to 1 when the source is isolated, the blend rate is then null. On the contrary when the galaxy is blended the purity decreases and the blend rate increases.

Working on real data, the definition must change since the true light profile is unknown. For the HSC survey (Bosch et al., 2018), a metric called *blendedness* was used. It is equivalent to the blend rate for real data. It is defined as:

$$b_{raw} = 1 - \frac{g(z_c, \mathbf{Q})}{g(z_p, \mathbf{Q})}
 \tag{3.3}$$

with:

$$g(z, \mathbf{Q}) = \sum_{\mathbf{r}} \frac{z(\mathbf{r})}{k(\mathbf{Q})} e^{-\frac{1}{2} \mathbf{r}^T \mathbf{Q} \mathbf{r}}
 \tag{3.4}$$

$z(\mathbf{r})$ is the image, centred on the galaxy centroid detected at position \mathbf{r} , \mathbf{Q} is the source moment matrix, computed from the deblended image of the central galaxy, and $k(\mathbf{Q})$ is a normalization constant which disappears in the blendedness computation; z_c and z_p respectively represent the image of the source deblended, i.e. with neighbours replaced with noise, and z_p is the original image, possibly with neighbour sources. The blendedness behaves similarly to the blend rate.

Note that this metric depends on the deblending and shape measurement algorithms. Here, \mathbf{Q} is computed as the second-moments of the deblended source image without PSF correction using a Gaussian weight function. The flux coming from z_c and z_p depends on \mathbf{Q} , which can induce outliers blendedness values in case of failure of the deblending algorithm for example (Bosch et al., 2018).

The same metric is implemented in the LSST processing pipeline and used to measure blendedness on DC2 simulated data (LSST Dark Energy Science Collaboration et al., 2021). Figure 3.2 shows the blendedness distribution for a galaxy sample extracted from DC2 (see section 7.8). Some values are negative which implies that the deblending process from the LSST pipeline, used to create the image of the deblended source, failed. This was not further explored but it would be interesting to verify how blendedness changes running a different deblender on the data.

Then, to flag an object as blended, one has to define a criterion based on one of the previously defined metrics.

Criterion for blending

To evaluate the amount of blended objects in LSST, Chang et al. (2013) (Chang et al., 2015) set a rejection criterion for every object within a 2" circle around a detected object. This value of 2" is considered as the limit at which deblender would be able to properly separate objects. This is questionable for several reasons. First it does not take into account the cases of very large and bright galaxies, which could make deblending impossible even at larger scales, and second, recent deblending algorithms may be able to perform correctly at smaller distances in some cases. This could be verified for recent deblenders such as SCARLET (P. Melchior et al., 2018) or the deblender proposed in this work (see chapter 7), for example using the Blending Tool Kit (see section 3.3.4). In any case, fixing a hard limit on the distance from the galaxy center seems not flexible enough, as deblending performance would depend on the method and on the galaxy's properties.

Sanchez et al. (2021) evaluated the effect of blended galaxies considering an object as blended if a least 1% of the flux in its pixels is due to overlapping sources. It corresponds to a purity below 99% or a blend rate above 1%. This is a strict definition which has the advantage of being independent from the galaxies properties, but is restricted to simulated galaxy images.

Since the HSC pipeline (Bosch et al., 2018) is working on real data, and that the measurement of blendedness might not be extremely reliable, the criterion to define a blended object is based on the output of the detection algorithm (see section 3.3.2). The detection algorithm defines **Footprints**, one per detected peak, which are regions above the detection threshold on the initial image convolved with a smoothing filter. These **Footprints** are then dilated by the Root Mean Square (RMS) of the PSF to make sure that they represent correctly the region occupied by the object. Initially there is one **Footprint** per object but they can be merged if they overlap so that, in the end, the **Footprints** in the final set are non-overlapping. If a **Footprint** contains multiple peaks, it is considered as a blend. The deblending algorithm, described in section 3.3.2, works from the images of these **Footprints**. This procedure is also implemented in the LSST pipeline and used to produce the catalogs from the DC2 simulations (Korytov et al., 2019) (see chapter 7 and chapter 8).

Finally, Dawson et al. (2016) focused on unrecognised blends. They consider a blend to be unrecognised when only one object is detected while there were several. A blend is then unrecognised as long as the distance between the two objects stays below a deblending threshold (or detection threshold) fixed by the detection algorithm (here **SExtractor**, Bertin and Arnouts, 1996). The results depends on the detection algorithm and on the set of fixed parameters, which means that improving the deblending algorithm could decrease the number of unrecognised blends.

3.2 Impact of blending

In this part I discuss the different biases that can arise due recognised or unrecognised blends. Two primordial parameters are impacted by blending: the effective projected number density of lensed galaxies, n_{eff} , and the shear.

3.2.1 Effective number density, n_{eff}

Comparing ground and space based surveys allows for a pretty accurate quantification of the proportion of unrecognised blends in ground based data. Dawson et al. (2016) compared the HSC

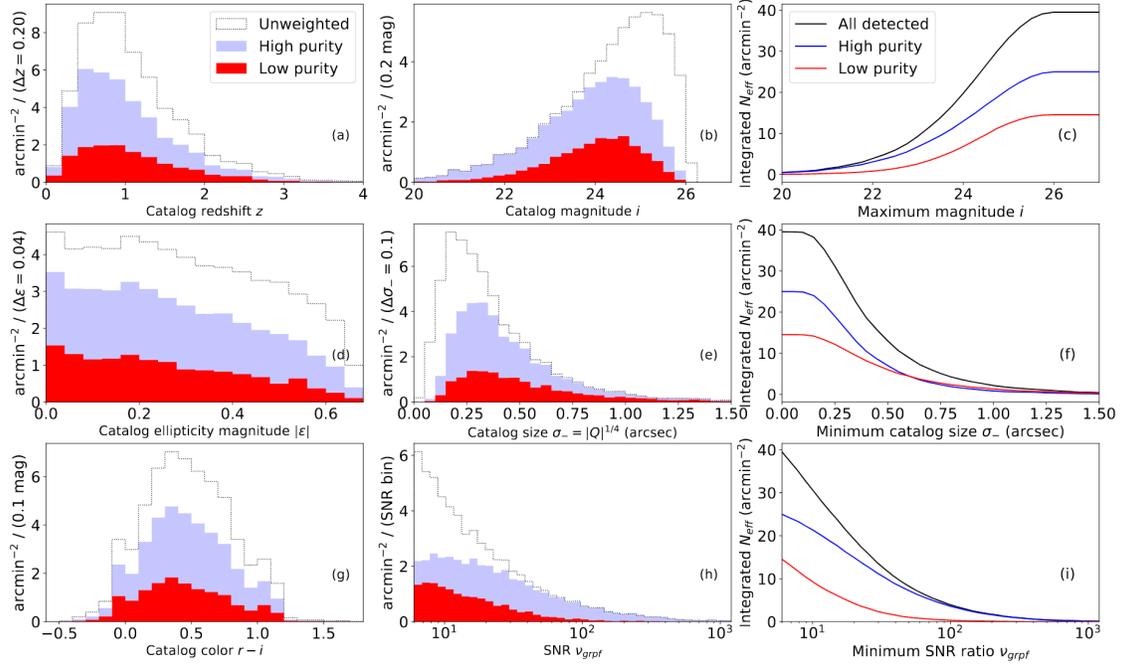


Figure 3.3: Distributions of catalogue parameters for detectable galaxies in a simulated LSST full-depth i -band exposure. The main take-away here is that for magnitudes higher than 24, or similarly at low SNR, the contribution of low-purity objects is very high. Credit: Sanchez et al., 2021

data, a ground based telescope with a seeing close to the one of LSST, to HST ACS observations, a space telescope, on the same sky field. They estimated that 14% of detected objects in LSST data will be unrecognised blends at a magnitude of $i \approx 27$. These unrecognised blends, implying undetected galaxies, lead to a decrease in n_{eff} , of $\approx 12\%$.

Blending consequently impacts n_{eff} via unrecognised blends. However, in cases where a rejection threshold is set to discard recognised blends which are considered too blended to be properly separated, those can also influence this parameter. This has been evaluated for LSST using $2''$ as the limit at which a deblender would be able to properly separate objects, rejecting every galaxy with neighbours closer than this distance (Chang et al., 2013; Chang et al., 2015). Sanchez et al. (2021) used an other method, discarding low-purity galaxies ($\rho_i < 0.98$), mostly located at magnitudes fainter than 24 in the r -band (see fig. 3.3). As expected, in both cases blending decreases significantly n_{eff} , of the order of $\approx 20\%$.

Sanchez et al. (2021) also studied the impact of blending on n_{eff} for DES and HSC and recovered similar results as previous studies (Jarvis et al., 2016 and Mandelbaum et al., 2018a). The decrease is of $\approx 5\%$ for DES and $\approx 15\%$ for HSC, which is coherent considering the characteristics of each survey.

3.2.2 Shear biases

Blending also has an impact on shear estimation. It is for example, the dominant source of detection bias and even the dominant source of shear bias for DES-Y3 (MacCrann et al., 2020). As explained in section 2.5.1, when several blended objects are detected as a single one, this creates a detection bias. This comes from the fact that detection is shear-dependent (Sheldon et al., 2020; Hoekstra, Kannawadi, and Kitching, 2021). As shown on fig. 3.4, a blend of objects which is detected as

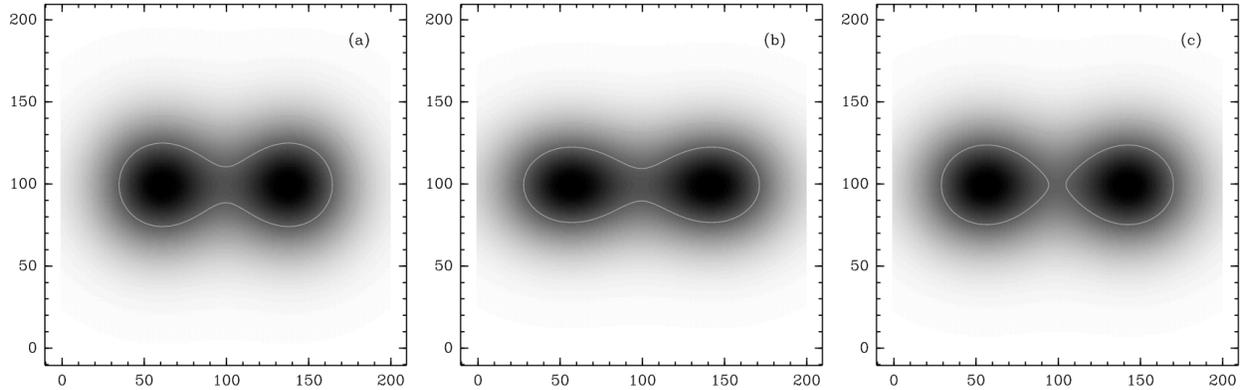


Figure 3.4: Toy example of shear-dependent object detection in the presence of a PSF. In panel (a) two objects are present, convolved by a PSF with no shear. Contours represent constant brightness. In panel (b) the objects are sheared by $\gamma = (0.0, 0.1)$ after the PSF convolution. In panel (c) the shear is applied before the PSF convolution, which mimics real sky images. Credit: Sheldon et al., 2020

one object on a non-sheared image (left panel) can, if sheared before PSF convolution, present two detected objects (right panel). The opposite effect is then of course possible. This can impact the ellipticity distribution: blended objects, if recognised as a single one, present a broader light profile leading to a large estimated ellipticity. The ellipticity distribution of unrecognised blends has a larger intrinsic dispersion compared to objects that can be deblended, leading to a 14% increase in shear noise for LSST (Dawson et al., 2016).

Figure 3.5 shows that a negative bias appears when pairs of galaxies are getting closer and closer (Sheldon et al., 2020; MacCrann et al., 2020; Hartlap et al., 2011). To avoid this, a solution is to set a limit of distance between the centers of the two galaxies, below which one discard the pair. Instead of having a hard limit in terms of distance in the sky (Chang et al., 2013; Chang et al., 2015), this limit can be parametrised as a function of the half-light radius of the galaxy to take into account different galaxy morphologies (Hoekstra, Kannawadi, and Kitching, 2021). However, doing so introduces a selection bias. First, blending can lead to a shift in the location of galaxies since the detected peak could be, in rare cases, in the overlapping region (Hoekstra, Kannawadi, and Kitching, 2021). Galaxies would then appear closer for the detection algorithm, leading to their removal. Also, as we have seen above, the distance measured by a detection algorithm between two peaks is shear-dependent, leading to a higher probability of removing pairs or galaxies with intrinsic ellipticities aligned with the shear (see fig. 3.4). Finally, such removal strategy creates a bias in the redshift distribution, as galaxies in the foreground have a higher probability of being blended, hence of being removed, which is also density-dependent. The number density of sources in the background gets correlated with the density field in the foreground, less galaxies being selected if an overdensity field is in the foreground, and vice versa. This creates a scale-dependent bias (Hartlap et al., 2011). These selection effects can result in a shear bias of several percent to about 10%, largely exceeding stage IV requirements ($< 0.3\%$).

Consequently, avoiding having to remove these pairs of galaxies would be a major step forward to deal with this systematic. The solution that we propose in chapter 7 is to deblend a field galaxy per galaxy optimizing an iterative procedure of detection, classification and deblending. This could, in theory, help reducing the number of very close pairs of galaxies, and maybe some of

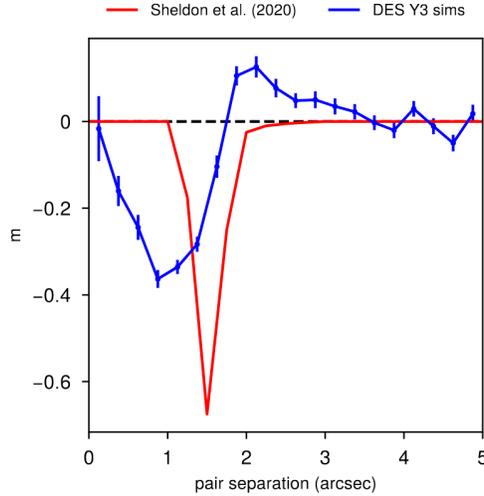


Figure 3.5: Multiplicative bias for detections corresponding to pairs of objects separated by a given angle. Comparison of results from MacCrann et al. (2020), “DES Y3 sims”, and “Sheldon et al. (2020)”. Credit: MacCrann et al., 2020

the unrecognised blends.

However, using a deblending algorithm can also introduce a bias. An incorrect flux assignment for each source might modify the shape and flux distribution (Mandelbaum et al., 2018b). Assigning extra flux from a bright galaxy to its neighbour might allow faint galaxies to pass above a magnitude cut, and to be selected for shear analysis, or it could make some galaxies appear more elliptical. This is why deblending methods must be as precise as possible (even if they cannot be perfect).

In chapter 8, we propose a step forward the resolution of this problem with a method which bypasses the deblending step and perform galaxy parameters estimation on the images, whether the galaxies are blended or not.

Finally, blending increases noise bias (see section 2.5.3) as pairs of galaxies get closer (Sanchez et al., 2021), but the role of blending for this particular bias is small compared to the detection and selection biases discussed above. It also impacts the measurements of photometric redshifts, modifying the distribution by making faint galaxies path the magnitude cut as explained before (Mandelbaum et al., 2018b), or simply rendering almost impossible a correct photometric redshift estimation since the SED of the blended objects are mixed (see section 2.2.4).

We have seen that blending could be a dominant systematic for shear estimation and accurate methods to decrease and mitigate this bias are crucial to reach the stage IV requirements. In the next section I discuss some of the solutions that exist today, and their limits, to provide context and motivations for the avenues explored in chapter 7 and chapter 8.

3.3 Dealing with blending

Most of the studies presented in the previous section stressed the need for efficient algorithms to deal with the issue of blending. In this part, I discuss some of the proposed solutions which I split in two categories: calibration and deblending for shape and redshift measurement.

3.3.1 Calibration

METADETECTION

METADETECTION (Sheldon et al., 2020) is an update of METACALIBRATION (see section 2.5.5, E. Huff and Mandelbaum, 2017 and Sheldon and E. M. Huff, 2017) which includes detection. The idea is to run a detection and measurement step on each sheared image produced by METACALIBRATION. We have seen previously (section 3.2.2) that the detection is shear dependent. As a result, running detection on images sheared differently will generate several catalogs with a different number of sources. The catalogs shear statistics can be then corrected using the linear response computed following the METACALIBRATION process. Finally, the ensemble mean ellipticities is computed over the different catalogs. Doing so allows for the reduction of the multiplicative shear bias, whether a deblending is performed or not, confirming that this bias came from detection. In case where a sufficiently accurate model for the point spread function is available, using METADETECTION could lead to a large decrease in average multiplicative biases for DES or LSST (Sheldon et al., 2020).

However, it is important to note that this method works on objects that are supposedly at the same redshift, i.e. which are sheared by the same amount. It applies the same shear to the entire image and cannot take into account cases where the blended galaxies would be sheared differently. Since a large portion of blends comes from the superposition of foreground objects with background galaxies, METADETECTION is not sufficient and other methods are required.

Simulations

For example, MacCrann et al. (2020) showed that, for two blended galaxies, shearing the galaxy in the foreground will entail a non-zero shear response (defined as in METACALIBRATION, see section 2.5.5) from the galaxy in the background. This has a large impact on $n_\gamma(z)$, the effective redshift distribution for lensing defined as $n_\gamma(z) = \bar{R}(z)n_{\text{eff}}(z)$ (see section 2.3.2), with \bar{R} the shear response. Using realistic simulations they could reproduce the biases introduced by blending and establish a protocol to correct for them. They managed to calibrate $n_\gamma(z)$ allowing for the joint calibration of redshift distribution and shear biases. As mentioned in section 2.3.2, this might be crucial for LSST since it will yield a very large galaxy density.

Most of recent weak lensing analyses such as HSC (Mandelbaum et al., 2018b) and DES Y1 (Samuroff et al., 2018) also used simulations for calibration. Simulations allow to recover blending-induced shear biases, such as the broader ellipticity distribution, or the difference in magnitude distribution mentioned earlier. Methods can then be implemented to correct for them, so that the resulting bias fits the survey requirements. An overall multiplicative bias of $\approx 10\%$ (at maximum) due to nearby galaxies and unrecognised blends is observed but it can be calibrated to meet the surveys' requirements (of the order of $\approx 1\%$ for stage III surveys).

To correctly calibrate the shear biases, it is crucial for the simulation to be as realistic as possible, especially that it includes nearby objects in the images. However, even if realistic simulations allow for a significant decrease in multiplicative bias, all studies affirm that the complementary use of a calibration scheme, such as METACALIBRATION (E. Huff and Mandelbaum, 2017; Sheldon and E. M. Huff, 2017) or METADETECTION (Sheldon et al., 2020), is required. Also, even if this calibration methods are used, the remaining bias is still of the order of 1% while it is supposed to be of the order of 0.3% for stage IV surveys. One avenue could be to create more realistic simulations using for example more realistic galaxy models. Bretonnière et al. (2021) proposed a method, based on the work of François Lanusse et al. (2021), to generate realistic galaxy images from a deep generative neural network, for the Euclid collaboration. Since the prior for galaxy image generation is learned from true galaxy images, the method generates more realistic galaxy profiles. I briefly show a way

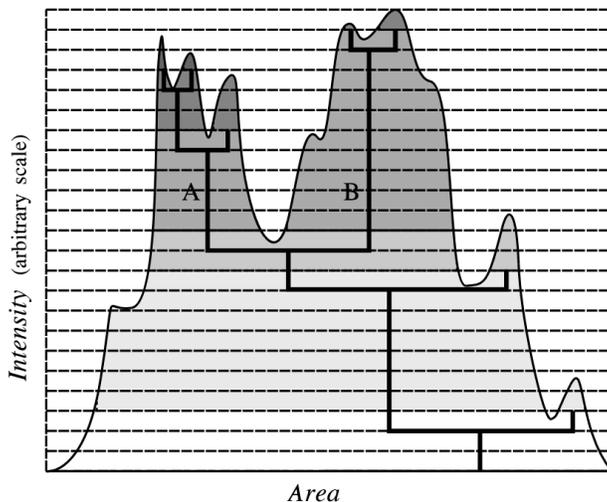


Figure 3.6: Schematic image of the deblending algorithm of **SExtractor**. The light profile of the object (smooth curve) is described as a tree structure. Determining if a branch should be considered as a component of the blend depends to its relative integrated density (tinted area). Here the object is split into two objects A and B. Credit: Bertin and Arnouts, 1996

of doing this in section 4.4 using variational autoencoders. The other avenue is, in addition to the correction of biases introduced via the processing pipeline, to try to decrease their amount using more precise methods. In the next section I describe the efforts that have been done in that sens for deblending algorithms.

3.3.2 Deblending

In this section I define a *deblending* algorithm as a method which accurately separates overlapped objects. In this context, *accurately* means with a precise reproduction of the flux and shape of each object.

SExtractor

SExtractor (Bertin and Arnouts, 1996) is probably the most used algorithm for detection nowadays. It is a peak detection and deblending algorithm. It starts by identifying local maxima above a pixel count threshold. Each of this peak is then re-thresholded at 30 levels exponentially spaced between the primary count threshold and its peak value. Then going downward from the maximum value of the image, a peak is considered as a component of a blend if, at a junction¹ (see fig. 3.6), the integrated pixel intensity of the peak is higher than a fixed fraction of the total intensity of the composite object, and that at least another peak satisfies the previous condition.

SExtractor has an embedded deblending algorithm. Once a blend is detected, the algorithm allocates each pixel to a particular object by assigning a probability for each pixel to be a part of each source. The morphology and flux of each object is then defined by its list of pixels. The deblending is performed on single band images, band per band if a multi-band image need to be processed.

¹Here a junction is defined as the moment when the level reached, while descending the 30 levels, is lower than the pixel value separating two peaks.

By definition this algorithm does not consider that several objects can contribute to the flux in a pixel, this is a *segmentation* approach, which is questionable. This approach can be sufficient for space surveys or for surveys with a very low amount of blended objects, but for ground surveys, and particularly for the most recent and future surveys, this approach is not satisfactory anymore since a lot of objects will be blended, with several sources contributing to same pixels. Results obtained on HSC images, and shown on the third column of fig. 3.7, illustrate this point.

SDSS PHOTO deblender

The SDSS PHOTO pipeline includes a deblender (R. Lupton, [in prep.](#)) designed for multi-band images (SDSS has five bands). It uses as input the peak positions and assumptions about the symmetry of objects to deblend stars and galaxies.

The procedure starts with identifying all of the peaks in a blend, these peaks define the children. For each *child* a template is constructed by looking at pixel pairs symmetrically located around the peak, and replacing both pixels by the lowest value of the two. A weight is then assigned to each *child* template, and optimized to minimize the difference between the reconstructed image from the templates, and the initial image of the blend. Finally, *child* images are constructed from the weighted templates. To generalise to multi-band data, the children are defined by the union of the peaks found in each band. The templates are defined independently in the different bands.

This algorithm is based on several assumptions namely that the astronomical objects are symmetrical, have non-negative fluxes, and are optically thin. The hypothesis of thin objects might be true for the disk of a galaxy (diffuse, external part) but is questionable for the central part which is usually much brighter. Also assuming only symmetrical objects works well most of the time but fails on galaxy with irregular morphologies. Finally, and probably because of the symmetry hypothesis, it proves to struggle with situations where a central object is symmetrically surrounded by neighbours.

It is however an interesting approach as it enables several object contributions in a pixel with few assumptions. The PHOTO deblender was used for SDSS and an improved version of it is used for HSC Bosch et al. (2018). Some results of its application to HSC data are shown on the right column of fig. 3.7.

DES algorithms

For DES Y1 (J. Zuntz et al., 2018) a combination of two deblending algorithms was used: the first one masks pixels close to neighbours (Jarvis et al., 2016), avoiding the possibility to have several contributions to the flux in one pixel, and the other one which is a multi-object fitting (MOF) algorithm (Drlica-Wagner et al., 2018). The MOF algorithm is a forward-modelling method fitting bulge+disk parametric models to each source in the blend. It uses the `ngmix` “multi-epoch” mode to fit a model to all available epochs and bands, and combines it with an iterative procedure, removing the light of fitted models until convergence. The sources in the blend are found using a friend-of-friend algorithm and they are simultaneously fitted in the g, r, i, z bands. Once all the sources are modelled, and the convergence is reached, a fraction of the flux in each pixel can be assigned to a particular object. For DES, this algorithm was used inside the METACALIBRATION pipeline.

Improvement have probably been done for DES Y3 but the presentation of the deblending procedure hasn’t been released yet. It should be detailed in Sevilla-Noarbe ([in prep.](#)).

As already pointed out in section 2.5.3, using parametric models to fit galaxies might quickly introduce a bias since the models that are used might not represent the entirety of the shapes that

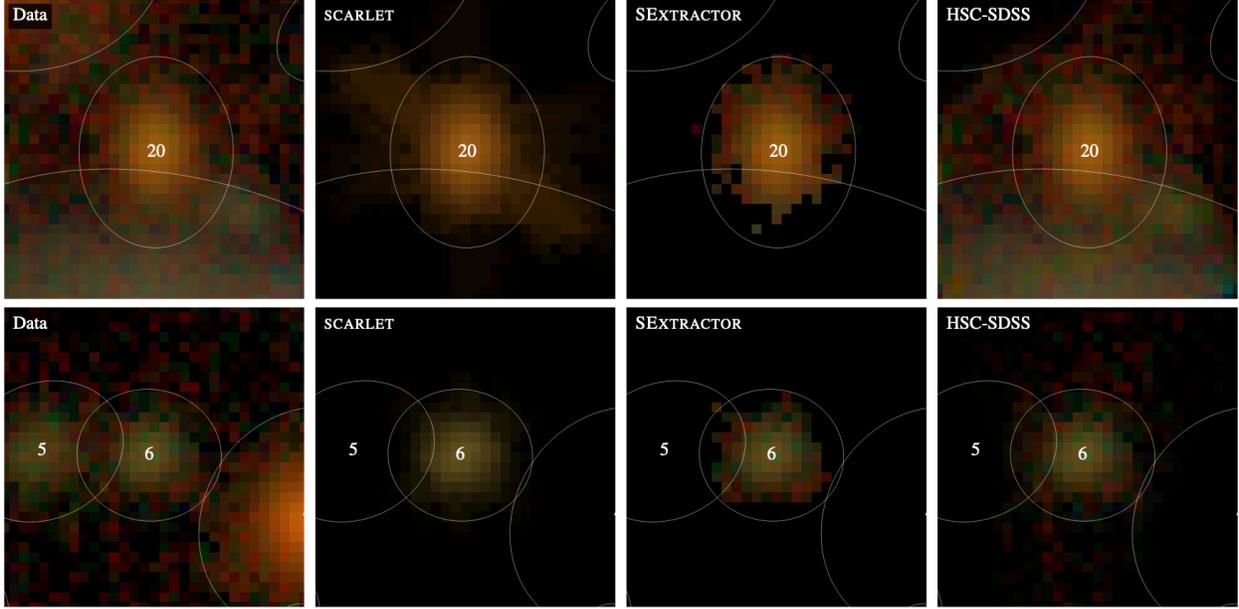


Figure 3.7: Comparison of deblending for several deblenders. The left column shows galaxies detected on an HSC image. The second column from the left presents the results obtained with **SCARLET**, the third shows results obtained with **SExtractor** and the column on the right shows the object deblended by the SDSS/HSC **PHOTO** deblender. Credit: P. Melchior et al., 2018

galaxies can take. Same as for **PHOTO**, an irregular or a spiral galaxy would not be fitted properly.

SCARLET

The baseline algorithm that is being implemented within the LSST science pipeline is the **SCARLET** algorithm (P. Melchior et al., 2018). It is based on constrained non-negative matrix factorisation. It can process multi-band images and performs deblending describing the observed scene as a mixture of components. The multi-band model M can be written as:

$$M = \sum_{k=1}^K A_k^T \times S_k = AS \quad (3.5)$$

with A_k the spectral energy distribution (SED) of the component k across all bands and S_k its shape.

Then, the scene is fitted minimizing the log-likelihood of the model. In addition to the non-negativity constraint they apply others, that they use as priors on the solution, turning the maximum likelihood estimate into a maximum a posteriori estimate. The **SCARLET** algorithm requires a central symmetry of the sources around the peak pixel (inspired from the **PHOTO** deblender), declining monotonicity from the peak pixel, and sparsity on the pixel domain.

Other priors and constraints can also be included. For example the possibility of having a shape prior learned from a deep generative model has already been proposed and tested (Francois Lanusse, Peter Melchior, and Moolekamp, 2019).

This method has several advantages. It permits to use multi-band images, it uses a non-parametric constrained morphological model, and it introduces the monotonicity constraint, which

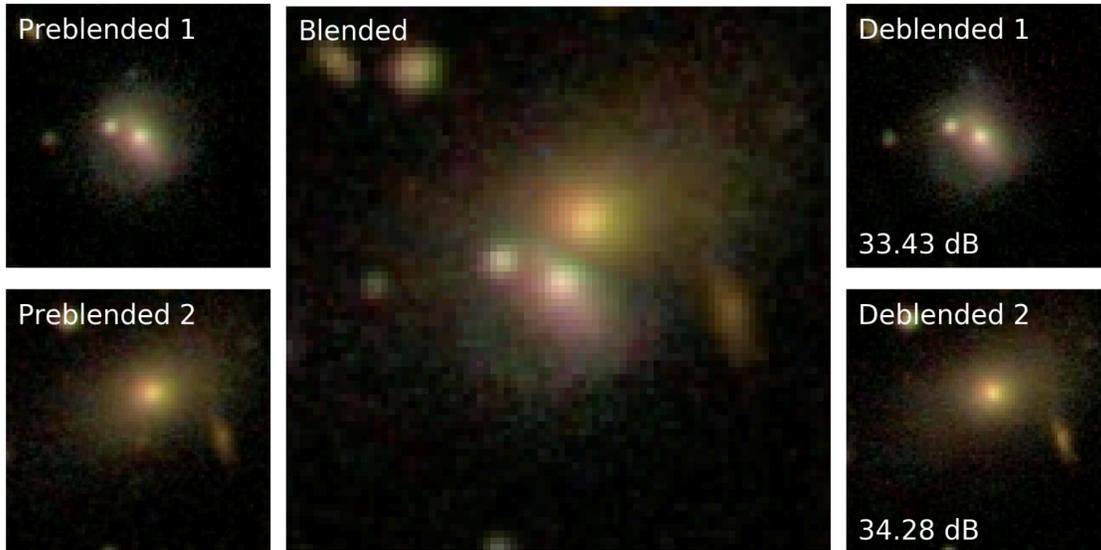


Figure 3.8: On the left are presented two images of galaxies from Galaxy Zoo which are combined to create the central blended input image. The predictions from the network are shown on the right. The image of the blend (center) shows a strange diagonal line coming from the unrealistic blending procedure: selecting the pixelwise maximum of two preblended images. Credit: Reiman and Göhre, 2019.

permits to avoid the current failure mode from PHOTO discussed above. Also it enables to take into account additional constraints that can come, for example from deep generative models. However it struggles with cases where an undetected source is close from the one that it tries to deblend, as illustrated by the first row of fig. 3.7, which shows an undetected source on the lower right of the object labelled 20.

Indeed, all previously presented deblenders require the position of all sources in order to perform the deblending. In chapter 7, we present an iterative procedure which deblends objects one by one requiring only the knowledge of the brightest galaxy position on the image. This iterative process might help to retrieve galaxies which would not be detected by a single detection step. We build our algorithm on a kind of deep generative models (see section 4.3), namely variational autoencoder, since they can learn complex features from galaxy images without making any assumptions, while naturally generalising to multi-band images.

Deblending with deep generative model

Before that, Reiman and Göhre (2019) proposed an approach based on Generative Adversarial Networks (GANs) for deblending multi-band SDSS images taken from GALAXY ZOO (Lintott et al., 2008). This approach was an interesting first step but the blend construction was a bit unrealistic (selecting the pixelwise max of two galaxy images, see fig. 3.8) and the galaxy sample was not representative of what future surveys will encounter (LSST will look at much higher magnitude than SDSS). Also, it was trained on real images, which helps to match a realistic shape distribution, but has downsides such as the difficulty to construct a clean training sample.

In our work presented in chapter 7, we tried to answer several of these issues, starting by using a more realistic blend construction (adding images of isolated galaxies), and working on simulations

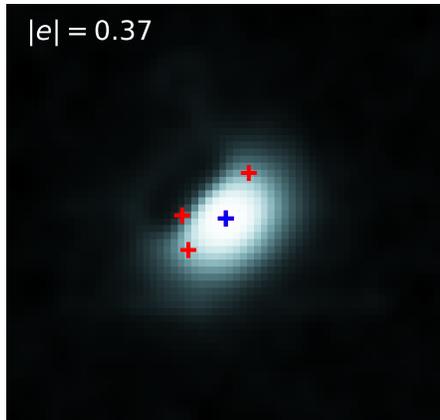


Figure 3.9: Image of a galaxy from the HST COSMOS catalogue with a residual from segmentation clearly visible on the top left side of the galaxy. The blue cross is the center of the galaxy and the red crosses correspond to the positions of galaxies that were added when we tested the deblending of real images with the deblending algorithm developed in chapter 7.

which reach the magnitudes that LSST will observe, and permit to have a clean training sample. We then work with real images of isolated galaxies taken from the HST COSMOS catalog, but even in that case we were faced with a corrupted training sample, i.e. blends that were considered as isolated galaxies or residuals of image processing altering the shapes of the objects (see fig. 3.9).

Other deep learning approaches

Other works using deep learning include a *deblending* step. I mention them here even though the goal of the deblending is always different from what we were trying to achieve previously, i.e. an accurate recovery of shapes and fluxes. In most of these approaches, the deblending algorithm uses segmentation.

For instance Boucaud et al. (2020) specifically target the segmentation and photometry measurement of blended pairs of galaxies. They tested two networks, a simple CNN and a U-Net architecture (Ronneberger, Fischer, and Brox, 2015). This work done on single band images in the context of the future Euclid survey showed promising results. They use a more realistic blending procedure than the one used by Reiman and Göhre (2019), similar to the one that we used in chapter 7: a pair of blended galaxies is created by adding the pixels of two isolated galaxies from the CANDELS fields (Dimauro et al., 2018), shifting one of the two of a small amount.

Most of the other algorithms including a deblending task are designed for source classification. For example, another work based on U-Net, MORPHEUS (Hausen and Robertson, 2020), aims at source detection, segmentation and classification. The main difference with the previous architecture is that it adds a classification step to output feature maps for each classes (*spheroid*, *disk*, *irregular*, *point source/compact*, and *background*), which is consistent with their different objective.

ASTERISM (Tramacere et al., 2016) used DBSCAN (Ester et al., 1996) for detection of sources and DENCLUE (Hinneburg, Keim, et al., 1998) for deblending, which performs segmentation by associating each pixel to a particular source. This algorithm, particularly focused on galaxies, was trained on GALAXY ZOO galaxies (SDSS) similarly to the method proposed by González, Muñoz, and Hernández (2018). Based on the YOLO algorithm (Redmon et al., 2015), the latter performs detection and classification of sources, but without the deblending step.

Burke et al. (2019) proposed another source classification algorithm based on the Mask R-CNN (He et al., 2017) image processing framework, using residual neural networks (ResNets). The algorithm performs instance segmentation at the final stage: each pixel is assigned to a class (star, galaxy or noise) which can be used to mask sources and deblend cutouts from the images. This method interestingly allows for several sources contribution in a single pixel.

Finally, and on a different problematic, Madireddy et al. (2019) proposed a deep learning pipeline for galaxy-galaxy strong lensing detection, deblending and modelling. They also used ResNet-based architecture for the denoising and deblending modules. The detection and modelling tasks are based on variational information bottleneck (VIB) (Alemi et al., 2016) framework, which allows for uncertainty estimation and accurate classification using latent space representation.

Except this last work, all the previously presented algorithms use a segmentation approach. Even if segmentation proved to be effective to recover main shape features and fluxes it is probably not appropriate for astronomical images. Indeed, astrophysical sources are shallow objects and using segmentation almost inevitably implies a loss of information. Moreover, using segmentation to separate sources that overlap implies that the shape of at least one of the objects is going to be strongly modified. For example, fig. 3.9 shows an example of a real galaxy image from the HST COSMOS catalogue where a residual from segmentation, performed with SExtractor, is clearly visible. In the context of weak lensing, i.e. where shapes and fluxes of objects must be recovered accurately, the use of segmentation is questionable.

Also, as we have seen *deblending* is not always used in the same purpose, which makes deblending algorithms comparisons very difficult. Each method develops its own metric and no performance comparison is done even for algorithms that aim for the same objective (except the "by eye" comparison presented on fig. 3.7). The Blending ToolKit tries to fill this gap, see section 3.3.4.

3.3.3 Probabilistic cataloging

Another avenue to the blending issue could be probabilistic cataloging. Contrary to traditional cataloging methods, creating a single catalogue, the idea is to propose a statistical model consisting of a likelihood for the observed image given a catalogue, and a prior distribution over possible catalogues. A posterior distribution over the set of possible catalogues is then created, which gives access to uncertainties. PCAT (Portillo et al., 2017; Feder et al., 2020) and StarNet (Liu, McAuliffe, and Regier, 2021) are two methods of probabilistic cataloging that were tested on crowded starfields. They respectively used MCMC and Variational inference (VI) in order to produce the posterior distribution, VI showing, as expected, a much faster convergence. These methods have the advantage of presenting a fully Bayesian method for deblending, including detection and modelling.

From a SDSS image, StarNet was able to recover a number of stars very close from the truth, set by the number of stars seen by HST. The application of this method to the blending of galaxies could be interesting even if this method is still preliminary and computationally costly. They plan to include galaxies to their method but the priors, for galaxy shapes especially, will have to be chosen with care. In addition, they foresee to use a prior learned from a variational autoencoder, which is the kind of generative models that we use in chapter 7.

3.3.4 Blending ToolKit

As mentioned above, it is difficult to compare the performance of different deblending algorithms. Simultaneously because they are tested on different datasets, with different performance metrics and do not aim for the same objective.



Figure 3.10: Portion of the DEEP2-3 field as seen by HSC. Credit: HSC³

The Blending ToolKit² (BTK) is a framework developed to test deblending algorithms on identical datasets, offering metrics to evaluate the algorithm performance. It is developed within DESC and was initiated by Sowmya Kamath together with Cyrille Doux, François Lanusse, Pat Burchat, Cécile Roucelle and Eric Aubourg. Ismael Mendoza took over the development recently and is especially helped by Thomas Sainrat, Hironao Miyatake, Maxime Paillassa, Alexandre Boucaud, and Rémy Joseph.

A lot of work has recently been put into making the framework easy to use and making sure that it contains every metrics necessary to test and validate a deblending algorithm. Since *deblending* can mean several things depending on who is using it, the framework pays a particular attention to offer performance metrics for each of the possible definition: detection, segmentation, and shape reconstruction. An extensive review of the literature has been done to this end. A version 1.0 of this package should be available around July 2021.

3.3.5 Multi-band, multi-instrument analysis

Looking at multi-band images clearly suggests that color can help to discriminate between objects, even if they are overlapped (see fig. 3.10). Designing a deblending algorithm able to work on multi-band images allows to make the most of the color information, and is nowadays the preferred solution for deblending.

For example, SCARLET or the DES multi-object fitting algorithm perform deblending on multi-band images.

However, as already explained, "classical" methods needs to make assumptions about the light profile of the sources they are trying to separate, which can induce model bias. Also they do not naturally generalise to multi-instrument images, combining images from a ground and a space telescope for example. Implementing a method able to work on images of different resolutions is extremely complex and none of the deblenders presented in section 3.3.2, using classical methods, has implemented this.

Deep learning provides a very simple solution to combine information from several bands and to build an arbitrary complex model from the data. It also naturally generalises to multi-band and multi-instrument images. This is a major advantage since it consequently allows for the combination

²<https://github.com/LSSTDESC/BlendingToolkit>

³<https://hsc-release.mtk.nao.ac.jp/hscMap2/>

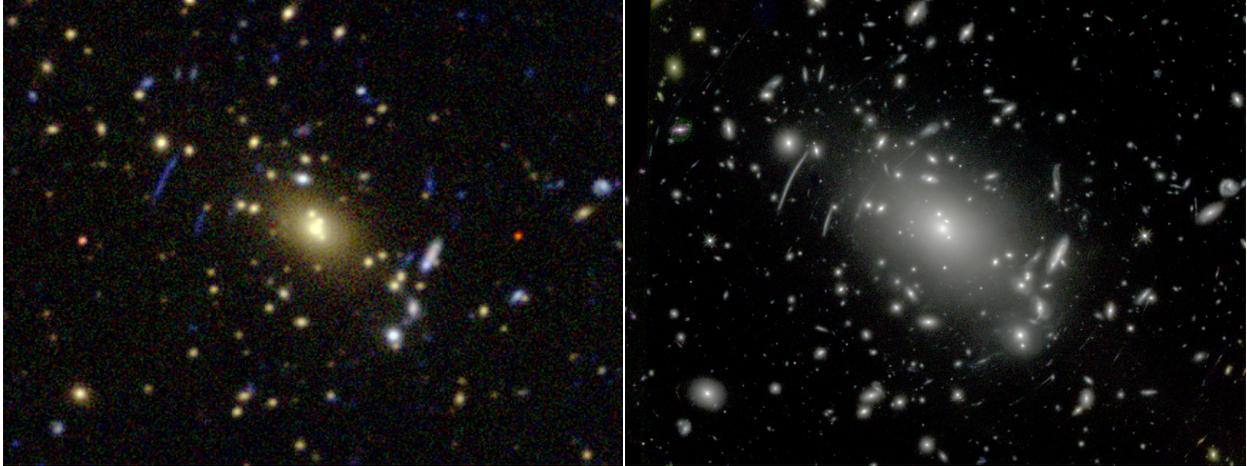


Figure 3.11: (left) Galaxy cluster Abell S1063 as observed by DES, a ground-based telescope. Credit: DES data from P. Melchior et al. (2017) (right) Galaxy cluster Abell S1063 as observed by HST, a space telescope. Credit: CLASH WFC3/IR data, image by Dan Coe

of space and ground data. Combining space and ground data permits to use the high resolution of space data (space data has usually a higher resolution than ground data due to the atmospheric PSF, see fig. 2.14 and fig. 3.11), and the color information from ground data (images from ground data are usually multi-band). Since Euclid and LSST sky coverage will overlap for several thousand square degrees (see fig. 1.11), synergies between the two surveys are expected (Rhodes et al., 2017; Jain et al., 2015; Capak et al., 2019). It has already been shown that LSST can benefit from Euclid data, up to the depth of Euclid, for shape reconstruction (Schuhmann, Heymans, and Joe Zuntz, 2019), and benefits are also expected for deblending.

Combining images from LSST and Euclid to feed a deblending algorithm based on neural networks can consequently, potentially, help improve the network performance. In chapter 7, we show that it permits a large increase in the precision of galaxy shape reconstruction using variational autoencoders.

Thanks to the much higher resolution of Euclid data, source detection is also mentioned as a task which can benefit from the synergy between Euclid and LSST. Nowadays source detection is mainly performed on single band images and to my knowledge, no algorithm performs detection on multi-band images. In section 3.3.2, I presented several deep learning algorithms which perform detection as one of their processing step, and allow for the use of multi-band images but none of them was dedicated to detection. Deep learning allows to make the most of the color information from multi-band images and could probably help to perform accurate detection. It could probably lead to a decrease in the number of unrecognised blends, potentially decreasing their impact on shear noise and effective projected number density of lensed galaxies (see section 2.5.1). A test of this hypothesis is reserved for future work.

In the next part I present some elements of deep learning including deep generative networks that are used in chapter 7. I briefly introduce Bayesian neural networks as a way to estimate epistemic uncertainty, the uncertainty coming from the lack of representativeness of the training sample, and I discuss the need for deep learning in cosmology.

Bibliography

- Alemi, Alexander A. et al. (Dec. 2016). “Deep Variational Information Bottleneck”. In: *arXiv e-prints*, arXiv:1612.00410, arXiv:1612.00410. arXiv: [1612.00410 \[cs.LG\]](#).
- Arcelin, Bastien et al. (Jan. 2021). “Deblending galaxies with variational autoencoders: A joint multiband, multi-instrument approach”. In: *MNRAS* 500.1, pp. 531–547. DOI: [10.1093/mnras/staa3062](#). arXiv: [2005.12039 \[astro-ph.IM\]](#).
- Bertin, Emmanuel and Stephane Arnouts (1996). “SExtractor: Software for source extraction”. In: *Astronomy and astrophysics supplement series* 117.2, pp. 393–404.
- Bosch, James et al. (2018). “The Hyper Suprime-Cam software pipeline”. In: *Publications of the Astronomical Society of Japan* 70.SP1, S5.
- Boucaud, Alexandre et al. (Jan. 2020). “Photometry of high-redshift blended galaxies using deep learning”. In: *MNRAS* 491.2, pp. 2481–2495. DOI: [10.1093/mnras/stz3056](#). arXiv: [1905.01324 \[astro-ph.GA\]](#).
- Bretonnière, H. et al. (May 2021). “Euclid preparation: XVI. Forecasts for galaxy morphology with the Euclid Survey using Deep Generative Models”. In: *arXiv e-prints*, arXiv:2105.12149, arXiv:2105.12149. arXiv: [2105.12149 \[astro-ph.GA\]](#).
- Burke, Colin J. et al. (Dec. 2019). “Deblending and classifying astronomical sources with Mask R-CNN deep learning”. In: *MNRAS* 490.3, pp. 3952–3965. DOI: [10.1093/mnras/stz2845](#). arXiv: [1908.02748 \[astro-ph.IM\]](#).
- Capak, P. et al. (Apr. 2019). “Enhancing LSST Science with Euclid Synergy”. In: *arXiv e-prints*, arXiv:1904.10439, arXiv:1904.10439. arXiv: [1904.10439 \[astro-ph.IM\]](#).
- Chang, C. et al. (Sept. 2013). “The effective number density of galaxies for weak lensing measurements in the LSST project”. In: *MNRAS* 434.3, pp. 2121–2135. DOI: [10.1093/mnras/stt1156](#). arXiv: [1305.0793 \[astro-ph.CO\]](#).
- (Feb. 2015). “Erratum: The effective number density of galaxies for weak lensing measurements in the LSST Project”. In: *MNRAS* 447.2, pp. 1746–1748. DOI: [10.1093/mnras/stu2553](#).
- Dawson, William A. et al. (Jan. 2016). “The Ellipticity Distribution of Ambiguously Blended Objects”. In: *ApJ* 816.1, 11, p. 11. DOI: [10.3847/0004-637X/816/1/11](#). arXiv: [1406.1506 \[astro-ph.CO\]](#).
- de Jong, J. T. A. et al. (Dec. 2013). “The Kilo-Degree Survey”. In: *The Messenger* 154, pp. 44–46.
- Dimauro, Paola et al. (2018). “A catalog of polychromatic bulge-disc decompositions of 17.600 galaxies in CANDELS”. In: *Monthly Notices of the Royal Astronomical Society* 478.4, pp. 5410–5426.
- Drlica-Wagner, A. et al. (Apr. 2018). “Dark Energy Survey Year 1 Results: The Photometric Data Set for Cosmology”. In: *ApJS* 235.2, 33, p. 33. DOI: [10.3847/1538-4365/aab4f5](#). arXiv: [1708.01531 \[astro-ph.CO\]](#).
- Ester, Martin et al. (1996). “A density-based algorithm for discovering clusters in large spatial databases with noise.” In: *Kdd*. Vol. 96. 34, pp. 226–231.

- Feder, Richard M. et al. (Apr. 2020). “Multiband Probabilistic Cataloging: A Joint Fitting Approach to Point-source Detection and Deblending”. In: *AJ* 159.4, 163, p. 163. DOI: [10.3847/1538-3881/ab74cf](https://doi.org/10.3847/1538-3881/ab74cf). arXiv: [1907.04929](https://arxiv.org/abs/1907.04929) [[astro-ph.IM](#)].
- González, R. E., R. P. Muñoz, and C. A. Hernández (Oct. 2018). “Galaxy detection and identification using deep learning and data augmentation”. In: *Astronomy and Computing* 25, pp. 103–109. DOI: [10.1016/j.ascom.2018.09.004](https://doi.org/10.1016/j.ascom.2018.09.004). arXiv: [1809.01691](https://arxiv.org/abs/1809.01691) [[astro-ph.IM](#)].
- Hartlap, J. et al. (Apr. 2011). “A bias in cosmic shear from galaxy selection: results from ray-tracing simulations”. In: *A&A* 528, A51, A51. DOI: [10.1051/0004-6361/201015850](https://doi.org/10.1051/0004-6361/201015850). arXiv: [1010.0010](https://arxiv.org/abs/1010.0010) [[astro-ph.CO](#)].
- Hausen, Ryan and Brant E. Robertson (May 2020). “Morpheus: A Deep Learning Framework for the Pixel-level Analysis of Astronomical Image Data”. In: *ApJS* 248.1, 20, p. 20. DOI: [10.3847/1538-4365/ab8868](https://doi.org/10.3847/1538-4365/ab8868). arXiv: [1906.11248](https://arxiv.org/abs/1906.11248) [[astro-ph.GA](#)].
- He, Kaiming et al. (Mar. 2017). “Mask R-CNN”. In: *arXiv e-prints*, arXiv:1703.06870, arXiv:1703.06870. arXiv: [1703.06870](https://arxiv.org/abs/1703.06870) [[cs.CV](#)].
- Hinneburg, Alexander, Daniel A Keim, et al. (1998). “An efficient approach to clustering in large multimedia databases with noise”. In: *KDD*. Vol. 98, pp. 58–65.
- Hoekstra, Henk, Arun Kannawadi, and Thomas D. Kitching (Feb. 2021). “Accounting for object detection bias in weak gravitational lensing studies”. In: *A&A* 646, A124, A124. DOI: [10.1051/0004-6361/202038998](https://doi.org/10.1051/0004-6361/202038998). arXiv: [2010.04178](https://arxiv.org/abs/2010.04178) [[astro-ph.CO](#)].
- Huff, Eric and Rachel Mandelbaum (Feb. 2017). “Metacalibration: Direct Self-Calibration of Biases in Shear Measurement”. In: *arXiv e-prints*, arXiv:1702.02600, arXiv:1702.02600. arXiv: [1702.02600](https://arxiv.org/abs/1702.02600) [[astro-ph.CO](#)].
- Ivezić, Željko et al. (Mar. 2019). “LSST: From Science Drivers to Reference Design and Anticipated Data Products”. In: *ApJ* 873.2, 111, p. 111. DOI: [10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c). arXiv: [0805.2366](https://arxiv.org/abs/0805.2366) [[astro-ph](#)].
- Jain, B. et al. (Jan. 2015). “The Whole is Greater than the Sum of the Parts: Optimizing the Joint Science Return from LSST, Euclid and WFIRST”. In: *arXiv e-prints*, arXiv:1501.07897, arXiv:1501.07897. arXiv: [1501.07897](https://arxiv.org/abs/1501.07897) [[astro-ph.IM](#)].
- Jarvis, M. et al. (Aug. 2016). “The DES Science Verification weak lensing shear catalogues”. In: *MNRAS* 460.2, pp. 2245–2281. DOI: [10.1093/mnras/stw990](https://doi.org/10.1093/mnras/stw990). arXiv: [1507.05603](https://arxiv.org/abs/1507.05603) [[astro-ph.IM](#)].
- Korytov, Danila et al. (Dec. 2019). “CosmoDC2: A Synthetic Sky Catalog for Dark Energy Science with LSST”. In: *ApJS* 245.2, 26, p. 26. DOI: [10.3847/1538-4365/ab510c](https://doi.org/10.3847/1538-4365/ab510c). arXiv: [1907.06530](https://arxiv.org/abs/1907.06530) [[astro-ph.CO](#)].
- Lanusse, Francois, Peter Melchior, and Fred Moolekamp (Dec. 2019). “Hybrid Physical-Deep Learning Model for Astronomical Inverse Problems”. In: *arXiv e-prints*, arXiv:1912.03980, arXiv:1912.03980. arXiv: [1912.03980](https://arxiv.org/abs/1912.03980) [[astro-ph.IM](#)].
- Lanusse, François et al. (July 2021). “Deep generative models for galaxy image simulations”. In: *MNRAS* 504.4, pp. 5543–5555. DOI: [10.1093/mnras/stab1214](https://doi.org/10.1093/mnras/stab1214). arXiv: [2008.03833](https://arxiv.org/abs/2008.03833) [[astro-ph.IM](#)].
- Lintott, Chris J. et al. (Sept. 2008). “Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey”. In: *MNRAS* 389.3, pp. 1179–1189. DOI: [10.1111/j.1365-2966.2008.13689.x](https://doi.org/10.1111/j.1365-2966.2008.13689.x). arXiv: [0804.4483](https://arxiv.org/abs/0804.4483) [[astro-ph](#)].
- Liu, Runjing, Jon D. McAuliffe, and Jeffrey Regier (Feb. 2021). “Variational Inference for Deblending Crowded Starfields”. In: *arXiv e-prints*, arXiv:2102.02409, arXiv:2102.02409. arXiv: [2102.02409](https://arxiv.org/abs/2102.02409) [[astro-ph.IM](#)].
- LSST Dark Energy Science Collaboration et al. (Mar. 2021). “The LSST DESC DC2 Simulated Sky Survey”. In: *ApJS* 253.1, 31, p. 31. DOI: [10.3847/1538-4365/abd62c](https://doi.org/10.3847/1538-4365/abd62c). arXiv: [2010.05926](https://arxiv.org/abs/2010.05926) [[astro-ph.IM](#)].

- Lupton, R. (in prep.). “SDSS Image Processing I: The Deblender”. In: <https://www.astro.princeton.edu/~rhl/photomisc/deblender.pdf>.
- MacCrann, N. et al. (Dec. 2020). “DES Y3 results: Blending shear and redshift biases in image simulations”. In: *arXiv e-prints*, arXiv:2012.08567, arXiv:2012.08567. arXiv: [2012.08567 \[astro-ph.CO\]](#).
- Madireddy, Sandeep et al. (Nov. 2019). “A Modular Deep Learning Pipeline for Galaxy-Scale Strong Gravitational Lens Detection and Modeling”. In: *arXiv e-prints*, arXiv:1911.03867, arXiv:1911.03867. arXiv: [1911.03867 \[astro-ph.IM\]](#).
- Mandelbaum, Rachel et al. (Jan. 2018a). “The first-year shear catalog of the Subaru Hyper Suprime-Cam Subaru Strategic Program Survey”. In: PASJ 70, S25, S25. DOI: [10.1093/pasj/psx130](#). arXiv: [1705.06745 \[astro-ph.CO\]](#).
- Mandelbaum, Rachel et al. (Dec. 2018b). “Weak lensing shear calibration with simulations of the HSC survey”. In: MNRAS 481.3, pp. 3170–3195. DOI: [10.1093/mnras/sty2420](#). arXiv: [1710.00885 \[astro-ph.CO\]](#).
- Melchior, P. et al. (Aug. 2017). “Weak-lensing mass calibration of redMaPPer galaxy clusters in Dark Energy Survey Science Verification data”. In: MNRAS 469.4, pp. 4899–4920. DOI: [10.1093/mnras/stx1053](#). arXiv: [1610.06890 \[astro-ph.CO\]](#).
- Melchior, P. et al. (July 2018). “SCARLET: Source separation in multi-band images by Constrained Matrix Factorization”. In: *Astronomy and Computing* 24, 129, p. 129. DOI: [10.1016/j.ascom.2018.07.001](#). arXiv: [1802.10157 \[astro-ph.IM\]](#).
- Portillo, Stephen K. N. et al. (Oct. 2017). “Improved Point-source Detection in Crowded Fields Using Probabilistic Cataloging”. In: AJ 154.4, 132, p. 132. DOI: [10.3847/1538-3881/aa8565](#). arXiv: [1703.01303 \[astro-ph.IM\]](#).
- Redmon, Joseph et al. (June 2015). “You Only Look Once: Unified, Real-Time Object Detection”. In: *arXiv e-prints*, arXiv:1506.02640, arXiv:1506.02640. arXiv: [1506.02640 \[cs.CV\]](#).
- Reiman, David M. and Brett E. Göhre (May 2019). “Deblending galaxy superpositions with branched generative adversarial networks”. In: MNRAS 485.2, pp. 2617–2627. DOI: [10.1093/mnras/stz575](#). arXiv: [1810.10098 \[astro-ph.IM\]](#).
- Rhodes, Jason et al. (Dec. 2017). “Scientific Synergy between LSST and Euclid”. In: ApJS 233.2, 21, p. 21. DOI: [10.3847/1538-4365/aa96b0](#). arXiv: [1710.08489 \[astro-ph.IM\]](#).
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.
- Samuroff, S. et al. (Apr. 2018). “Dark Energy Survey Year 1 results: the impact of galaxy neighbours on weak lensing cosmology with IM3SHAPE”. In: MNRAS 475.4, pp. 4524–4543. DOI: [10.1093/mnras/stx3282](#). arXiv: [1708.01534 \[astro-ph.CO\]](#).
- Sanchez, Javier et al. (Mar. 2021). “Effects of overlapping sources on cosmic shear estimation: Statistical sensitivity and pixel-noise bias”. In: *arXiv e-prints*, arXiv:2103.02078, arXiv:2103.02078. arXiv: [2103.02078 \[astro-ph.CO\]](#).
- Schuhmann, Robert L., Catherine Heymans, and Joe Zuntz (Jan. 2019). “Galaxy shape measurement synergies between LSST and Euclid”. In: *arXiv e-prints*, arXiv:1901.08586, arXiv:1901.08586. arXiv: [1901.08586 \[astro-ph.IM\]](#).
- Sevilla-Noarbe, I. et al. (in prep.). In:
- Sheldon, Erin S. and Eric M. Huff (May 2017). “Practical Weak-lensing Shear Measurement with Metacalibration”. In: ApJ 841.1, 24, p. 24. DOI: [10.3847/1538-4357/aa704b](#). arXiv: [1702.02601 \[astro-ph.CO\]](#).

- Sheldon, Erin S. et al. (Oct. 2020). “Mitigating Shear-dependent Object Detection Biases with Metacalibration”. In: *ApJ* 902.2, 138, p. 138. DOI: [10.3847/1538-4357/abb595](https://doi.org/10.3847/1538-4357/abb595). arXiv: [1911.02505](https://arxiv.org/abs/1911.02505) [[astro-ph.CO](#)].
- The Dark Energy Survey Collaboration (Oct. 2005). “The Dark Energy Survey”. In: *arXiv e-prints*, astro-ph/0510346, astro-ph/0510346. arXiv: [astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346) [[astro-ph](#)].
- Tramacere, A. et al. (Dec. 2016). “ASTERIsM: application of topometric clustering algorithms in automatic galaxy detection and classification”. In: *MNRAS* 463.3, pp. 2939–2957. DOI: [10.1093/mnras/stw2103](https://doi.org/10.1093/mnras/stw2103). arXiv: [1609.06728](https://arxiv.org/abs/1609.06728) [[astro-ph.IM](#)].
- Zuntz, J. et al. (Nov. 2018). “Dark Energy Survey Year 1 results: weak lensing shape catalogues”. In: *MNRAS* 481.1, pp. 1149–1182. DOI: [10.1093/mnras/sty2219](https://doi.org/10.1093/mnras/sty2219). arXiv: [1708.01533](https://arxiv.org/abs/1708.01533) [[astro-ph.CO](#)].

Part II

Deep learning for cosmology

In the work presented in chapter 7 and chapter 8, I use Bayesian deep learning algorithms to address the issue of blending of galaxies. Consequently, in the next three chapters, I introduce the concepts and methods which are required for this work. I start by providing some elements of deep learning, especially discussing deep generative models, used in chapter 7. Then I present an introduction to Bayesian neural networks (BNN) and briefly explain some methods to train them, since I use a BNN in chapter 8. Finally, the last chapter is a discussion about the interest of using deep learning for cosmology today.

Chapter 4

Elements of deep learning

In this chapter I briefly describe the mathematical developments behind neural networks, how to overcome some common training issues that I ran on, and a particular class of neural networks, namely generative models. I finish this chapter with an example of galaxy image generation using the trained decoder of a variational autoencoder. This chapter is largely inspired from I. Goodfellow, Bengio, and Courville (2016).

4.1 Components of a neural network

A neural network is usually a stack of different *layers* which can be formed of several neurons. They can be of different types, each of them being suited for specific tasks. In the following I describe the two most commonly used layer kinds to give a hint of what layers are composed of, but much more possibilities exist.

4.1.1 Dense layers - fully connected layers

Fully connected layers or *dense layers* are a combination of activation layers and linear feed-forward layers, which are themselves a combination of bias and linear layers. Here, I will take the example of a single neuron layer, i.e. a neuron.

Neurons

A linear feed-forward neuron is the combination of a bias neuron, $z = b$, which can learn an offset or the mean of a sample, and a linear neuron, $z = x \times w$, which can learn an average rate of correlation between the network input and output. Here, z is the neuron's output, b is the bias, x is the neuron's input, and w is the average rate, or the weight. Combined together, a linear feed-forward neuron can be written as:

$$z = x \times w + b \tag{4.1}$$

It provides a way to learn about a bias and a correlation rate between the neuron's input and the output. It can then be used to learn a linear relationship between the input or the output, or to scale data before feeding a neural network for example.

The problem with this kind of neuron is that, per construction, it does not deal with non-linearity and its applications are consequently very limited. To add non-linearity in the process, the feed-forward neuron output needs to be transformed by a non-linear function called the *activation function*.

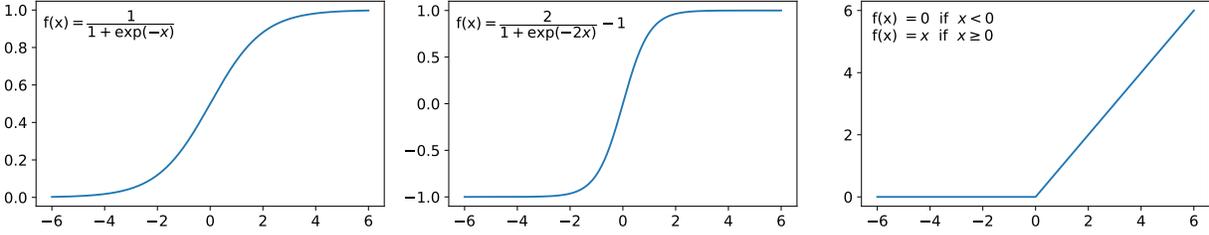


Figure 4.1: Example of activation functions: sigmoid (left), hyperbolic tangent (center) and ReLU function (right).

Activation functions

The activation function, a , applies non-linear transformation to the output of a neuron. In the case of the linear feed-forward neuron, the output becomes:

$$z = a(x \times w + b) \tag{4.2}$$

Training the network still consists in learning w and b , but thanks to the activation function it is possible to learn non-linear relationship between the neuron's input and output (respectively x and z). A lot of different activation functions can be used and some of the most common are shown on fig. 4.1.

Layers

Until now, I only described single neuron layers. However, layers usually contain a lot of neurons, and each of them outputs a different value. For neural networks composed of dense layers, every neuron of each layer is connected to every neuron of the previous and of the next layer (see fig. 4.2). This is why these layers are also called *fully connected layers*. In the example of the 3 layers neural network shown on fig. 4.2, the output y would be computed as:

$$z_1^3 = a^2 \left(z_1^2 \times w_{1,1}^{2,3} + z_2^2 \times w_{2,1}^{2,3} + z_3^2 \times w_{3,1}^{2,3} + z_4^2 \times w_{4,1}^{2,3} + z_5^2 \times w_{5,1}^{2,3} + b_1^2 \right) \tag{4.3}$$

where a^l is the activation function of layer l , z_i^l the value of neuron i in layer l , $w_{i,j}^{l,l+1}$ the weight applied between the neuron i in layer l and the neuron j of layer $l + 1$, and b_j^l , the bias applied between the layer l and the neuron j of layer $l + 1$. An activation function is usually applied on the last neuron value to yield the output: $y = a^3(z_1^3)$.

Here, I specifically detailed the computation for a very small neural network. However, this is a lot of subscripts and superscripts to handle. We can write this more clearly as a matrix operation: $Z_3 = a_2(W_2 \times Z_2 + B_2)$, with $W_2 \times Z_2$ being a matrix multiplication. Since neural networks are often composed of a lot of layers with a lot of neurons, computations are done with matrix operations. Stacking dense layers can quickly become computationally costly. Here, there are already 21 parameters to learn for a 3 layers network respectively composed of two, five, and one neurons. This number increases very quickly with wider layers (more neurons in each layer) and deeper networks (more layers).

To train these networks, one consequently needs specialised hardware. One of the reasons of the success of Graphical Processing Units (GPU) to train neural networks, are their specialisation in doing parallel processing which allows them to perform extremely fast matrix multiplications.

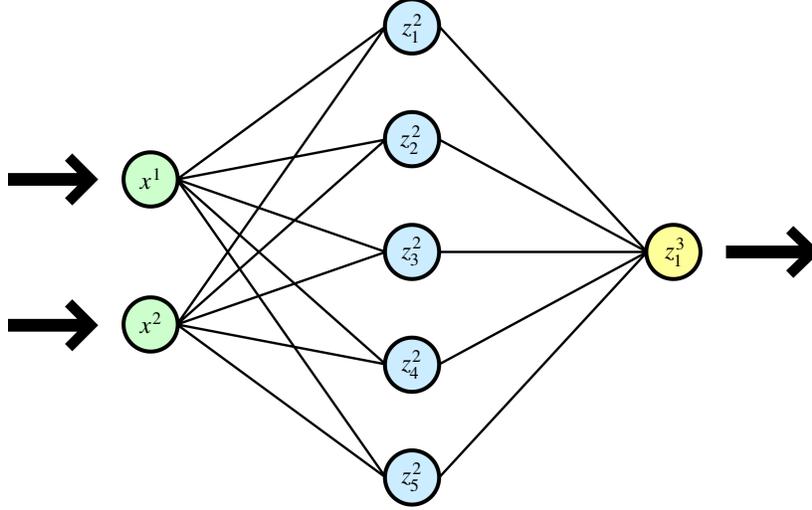


Figure 4.2: Three dense layers neural network. One input layer (green), one hidden layer (blue) and the output layer (yellow). They are respectively composed of two, five and one neurons.

In appendix A, I present a comparison of the performance of a GPU with an other hardware specialised for deep learning applications, the Intelligence Processing Units (IPU). I tested their performance for two cosmological applications: the generation of galaxy images and the training a neural network for galaxy shape parameters estimation.

4.1.2 Convolutional layers

Dense layers are not well-suited for every task. To work on images for example, one usually turns to an other type of layers, *convolutional layers*.

The use of convolutional layers first appears in Fukushima (1979) for pattern recognition but LeCun et al. (1999) are the first to stack several layers to create a convolutional neural network (CNN) and to show that it can be trained to produce high performance in pattern recognition.

As expected, convolutional layers apply convolution operations on their inputs instead of matrix multiplications. The aim, is to learn convolutional kernels (or filters) which will be applied to the data and will learn its specific features.

As we mostly use 2D-convolution, I focus on the example of convolving an image I with a 2D-kernel K . This would give an output S defined as:

$$\begin{aligned}
 S(i, j) &= (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \\
 &= \sum_m \sum_n I(i - m, j - n)K(m, n)
 \end{aligned}
 \tag{4.4}$$

as convolution is commutative. (i, j) is the pixel coordinates of S and (m, n) , the coordinates of I or K (respectively for the first and second line). Usually the second expression is implemented to have smaller variation ranges for m and n .

Here, the aim is to learn the bias and weights of the convolution kernels K . For example for a 3×3 pixels kernel, ten parameters (9 weights + 1 bias) would have to be learned. Those will assign importance to different features in images preserving spatial relationship between pixels and creating a *feature map*. An example is shown on fig. 4.3 going from a 2D matrix (in blue) to a feature map (in red) through the 3×3 kernel (or filter) matrix.

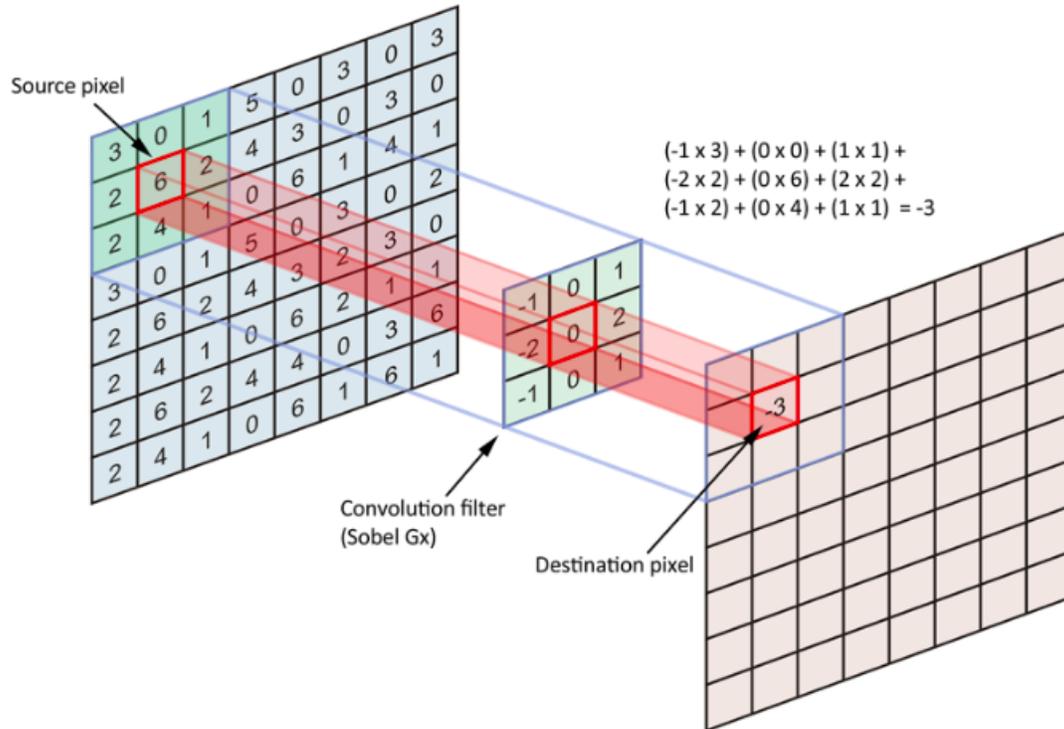


Figure 4.3: Example of 3×3 kernel being applied to a 2D-matrix.

Different values in the kernel matrix will produce different feature maps and the training consist in finding the values which reveal the most meaningful features to propagate through the network. For each convolution layer, it is possible to use several kernels, generating several feature maps from the input. Also, the learned kernels are invariant by translation: one kernel can be applied to any part of the image. Thus convolutional layers contain less trainable parameters than dense layers and are, in addition to being more suited to extract important features, less computationally costly when applied to images.

Convolutional layers are often used in combination with *pooling* layers. The latter allow to reduce the 2D matrix dimensions, giving access to main features at different scales. Different type of pooling layers exist, depending on the feature one wants to propagate: the maximum value in a portion of the image (*Max Pooling*), or the average value (*Average Pooling*) for example.

Finally, it is interesting to note that images are not the only type of data with a known, grid-like topology, for which convolutional layers are particularly suited. Time-series data, for example, can also be processed very efficiently with 1D-convolution layers as in Gabbard et al. (2019) who used it to estimate the source parameters of a gravitational waves event.

4.1.3 Other types of layer

There exists lots of different kind of neural network layers, each of them with a particular purpose: layers to process data (see Batchnormalization, Ioffe and Szegedy, 2015), to add non-linearity with specific activation functions (see PReLU, He et al., 2015b), or to realise operation between layers (addition, multiplication, concatenation...).

Similarly to CNN, there are also layers suited to particular kind of data such as *Long short term memory (LSTM)* network (Hochreiter and Schmidhuber, 1997) particularly suited for time

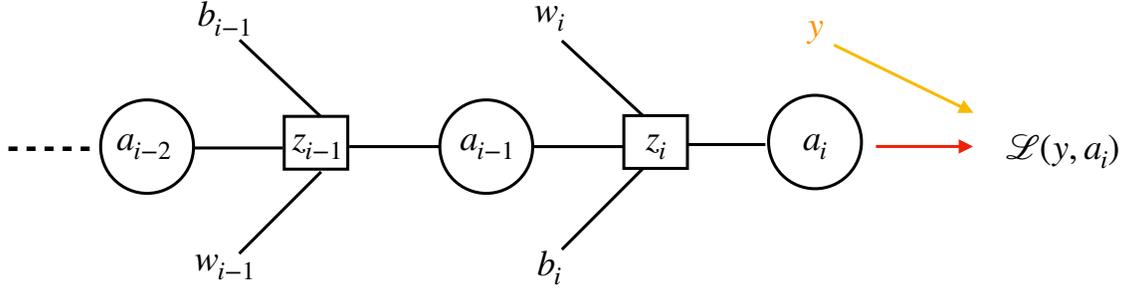


Figure 4.4: Diagram of the last layers, of a neural network composed of i single neuron layers. The loss \mathcal{L} is computed from the output of the network, $a_i(z_i)$, and the target y .

series. It is a type of recurrent neural network (RNN) and is broadly used in Natural Language Processing (NLP) for example (Sak, Senior, and Beaufays, 2014, for instance). More recently, *transformers* (Vaswani et al., 2017) using *attention layers* (Bahdanau, Cho, and Bengio, 2014) have been proposed and improved state-of-the-art results in NLP (Devlin et al., 2018 or Radford et al., 2019, for example). Attention layers can also be used for other tasks such as image generation (Zhang et al., 2019).

To summarize, multiple kinds of layers exist, each particularly good at a specific task. It is not possible to be exhaustive on such topic but most of current algorithms, and particularly the ones I implemented in part III, are based on the two types detailed previously.

Stacking more and more layers, more and more complex mathematical models can be generated and more and more complex mathematical functions can be fitted. The "deep" in *deep-learning* comes from this notion of increased complexity by adding layers. Now that we have an idea of what artificial neural networks are made of, it is necessary to understand how they are trained. In the next section, I briefly describe the backpropagation algorithm which allows the neural networks to learn from the data. Then, I describe some common issues to train neural networks, that I have encountered.

4.2 The Backpropagation algorithm

4.2.1 A simple example: single-neuron layers neural network

Here, we consider a neural network composed of i single neuron layers (see fig. 4.4). This is a simpler example than the one presented previously which is chosen for the sake of clarity of the following explanation.

Here, an artificial neuron z_k from layer k receives information from different sources. We will focus on the neuron in the last layer of the network, z_i , which receives information from z_{i-1} , a weight w_i and a bias b_i . The output of the network is $a_i(z_i)$: the value of the neuron z_i transformed by an activation function a_i .

The value of the neuron z_i is equal to:

$$z_i = a_{i-1}(z_{i-1}) \times w_i + b_i \quad (4.5)$$

If we consider $a_i(z_i)$ as the output of our network, and y the label that needs to be recovered by

the network, then \mathcal{L} , the loss (or cost¹) function used to train the network, is computed from these two values, $\mathcal{L}(y, a_i(z_i))$. This function can be the mean squared error, the binary cross-entropy, or any other cost function. The choice mostly depends on the targeted task (regression, classification, etc.) and on the range of output values.

Since the cost depends on $a_i(z_i)$, it also depends on the values of w_i , b_i and $a_{i-1}(z_{i-1})$. The value of $a_{i-1}(z_{i-1})$ in turn depends on the values of w_{i-1} , b_{i-1} and $a_{i-2}(z_{i-2})$. If the network is composed of N layers, the output and hence, the cost, depends on the values of the weights and biases of the N layers. To minimize the cost, biases and weights need to be updated and optimized using stochastic gradient descent (SGD). During training, the gradient is computed and the trainable parameters (the weights and biases) are updated, using the learning rate (as shown in eq. (4.11) and eq. (4.12)), after each data sample (or batch of examples).

4.2.2 Optimisation with stochastic gradient descent

To use SGD, it is necessary to compute the gradient of the cost, $\nabla\mathcal{L}$, as a function of each parameter of the neural network. To do so, each term of the gradient can be decomposed using the chain rule. When training a neural network this is obviously a large matrix computation. Here, for the sake of simplicity, I will continue with the example of a neural network composed of i single neuron layers. We can start by computing the term of the gradient as a function of the weight w_i for example. For a training example k , it can be decomposed as:

$$\frac{\partial\mathcal{L}_k}{\partial w_i} = \frac{\partial\mathcal{L}_k(y, a_i(z_i))}{\partial a_i(z_i)} \frac{\partial a_i(z_i)}{\partial z_i} \frac{\partial z_i}{\partial w_i} \quad (4.6)$$

Replacing each component by its expression in the right term of the equation:

$$\frac{\partial\mathcal{L}_k}{\partial w_i} = \mathcal{L}'_k(a_i(z_i)) a'_i(z_i) a_{i-1}(z_{i-1}) \quad (4.7)$$

Similarly you can compute the derivative of the loss, \mathcal{L}_k , as a function of the bias and the previous neuron output $a_{i-1}(z_{i-1})$:

$$\frac{\partial\mathcal{L}_k}{\partial b_i} = \mathcal{L}'_k(a_i(z_i)) a'_i(z_i) \quad (4.8)$$

$$\frac{\partial\mathcal{L}_k}{\partial a_{i-1}(z_{i-1})} = \mathcal{L}'_k(a_i(z_i)) a'_i(z_i) w_i \quad (4.9)$$

This is the derivative of the cost for a single training example k , to compute the derivative of the full cost function the only thing to do is to average over all the training examples. For the derivative as a function of w_i , it becomes:

$$\frac{\partial\mathcal{L}}{\partial w_i} = \frac{1}{N} \sum_{k=0}^{k=N-1} \frac{\partial\mathcal{L}_k}{\partial w_i} \quad (4.10)$$

Remember that this is only for a network composed of single neuron layers. This derivative of the full cost function as a function of the weight w_i is only one element of the gradient of the cost function, $\nabla\mathcal{L}$. The latter is composed of all the derivatives of the cost as a function of each weight and bias of the neural network.

To obtain all these derivatives, it is possible to iterate this chain rule idea used for w_i or b_i backward to compute how sensitive is the cost to every previous trainable parameters (w_{i-1} , b_{i-1} ...) and update them accordingly. The algorithm performing the automatic differentiation for

¹I will equivalently use *cost function* or *loss function* as well as *cost* or *loss* through this manuscript.

calculating gradients in a neural network is called the *backpropagation algorithm*. The network can then be optimized, updating the parameters, through SGD.

To modify the parameter's value, a hyperparameter called the *learning rate*, l , is used. It is a coefficient applied to the gradient in order for the parameter's value to change of a reasonable amount. It is set arbitrarily before training, a common choice is 10^{-3} . We previously showed the computation of the gradient for one example but it is usually done per batch of examples. The update of the weights and biases, between the batch $(k - 1)$ and (k) , can then be computed as:

$$w_i^{(k)} = w_i^{(k-1)} - l \frac{\partial \mathcal{L}}{\partial w_i^{(k-1)}} \quad (4.11)$$

$$b_i^{(k)} = b_i^{(k-1)} - l \frac{\partial \mathcal{L}}{\partial b_i^{(k-1)}} \quad (4.12)$$

It is possible to extrapolate these formula for networks with more complicated architectures but the main idea remains the same.

As we have seen, this backpropagation process depends on a lot of parameters. The choice of the loss function, the number of parameters to train, the choice of activation functions (among other things), can all have an impact on the training of a neural network. I briefly describe two of issues that can arise, as well as corresponding solutions, in the next part.

4.2.3 Training issues and examples of regularisation

Vanishing gradient

As previously explained, the update of a parameter depends directly on the corresponding gradient value. A problem can occur if the gradient becomes extremely small: the weights are not updated any more and, in the worst case, the neural network may stop learning.

This issue arises mainly from using specific activation functions such as the hyperbolic tangent or sigmoid function (see fig. 4.1). The derivative of these functions can be almost equal to zero in some intervals. If several layers use one of these activation functions for example, the function derivative arises several time in the computation of the gradient of the cost function. Consequently, if it is close to zero, the gradient becomes exceedingly small and *vanishes*. The weights or biases are not updated any more and the network stops learning.

A simple solution is to use other activation functions such as the Rectified Linear Unit (ReLU) which does not cause small derivative. Another common avenue is to use residual neural networks (ResNet He et al., 2015a). In this particular architecture, skip or residual connection are used: it connects an early layer to a deeper one, skipping one or two layers in the process (see fig. 4.5). It is then possible to propagate the gradient information straight to earlier layers without going through many activation functions, preventing the gradient from vanishing.

Consequently, this problem can be easily dealt with once it is detected.

Overfitting

Another common issue when training neural networks (and for machine learning in general) is *overfitting*. Usually, when we train a neural network, the goal is to make predictions or measurements on data for which we do not have the truth. We hope that the network will be able to generalise to previously unseen data, i.e. not seen during training. The ability of a network to generalise can be monitored during training by splitting the dataset, for which we know the truth, in three samples: the training, validation and test samples. Only the training and validation samples are used during

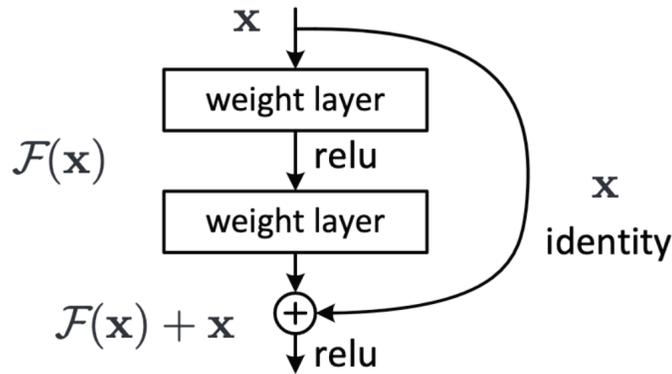


Figure 4.5: Block with a skip connection from an early layer to a deeper one, skipping two layers. Source: He et al., 2015a.

training, and the training sample is the one used for backpropagation. The gradient of the loss function is computed from the performance of the network on this sample. We need to assume that these samples are *independently* drawn from the same data-generating process, which has a defined probability distribution (the samples are independent and identically distributed, i.i.d.). At each epoch, i.e. when the network saw all the training sample once, the loss is also computed on the validation sample (without updating the weights). Consequently, at the end of an epoch we have two values for the training loss and the validation loss, that can be compared. The problem of *overfitting* appears when a network learns too precisely the training data so that it loses its generalisation power.

An example is presented on the left panel of fig. 4.6. At first the network learns features from the training data and is able to generalise: the training and validation losses decrease at the same rate. Then, the network becomes too specialised on learning features of the training data. It is not able to generalise anymore and the validation loss stops decreasing and starts increasing (after about 30 epochs on fig. 4.6).

The result of overfitting is illustrated on the right panel of fig. 4.6: overfitting takes the network away from the optimum solution. On the contrary, *underfitting* is when the network is not able to learn enough features to give correct predictions.

The ability of a model to overfit or underfit depends on its *capacity*, which itself depends on the network's architecture, the number of trainable parameters, and the number of examples in the training sample. If the capacity of the model is too low, the model will underfit. On the contrary if it is too high, it will overfit. Trying to reduce overfitting is equivalent to reduce the network's capacity. Studying the capacity of models helped justifying the fact that machine learning algorithm such as neural networks can work. Important results in statistical learning theory show that the discrepancy between training error and generalisation error, which characterises overfitting, is bounded from above by a quantity that grows as the model capacity increases but shrinks as the number of training examples increases (V. N. Vapnik and Chervonenkis, 2015; V. Vapnik, 1982; Blumer et al., 1989; V. Vapnik, 2013).

All the strategies in machine learning which are explicitly designed to overcome this issue, are known as regularisation. I briefly present two examples of regularisation that I used in my work: data augmentation and dropout.

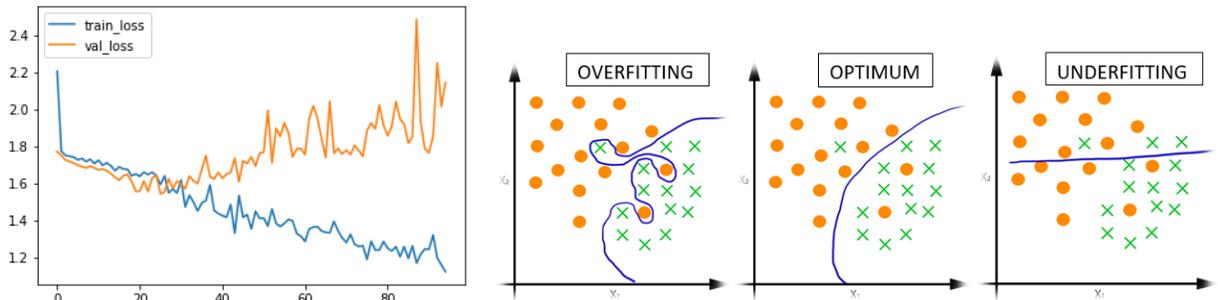


Figure 4.6: Example of overfitting. (left) The evolution of the training and validation loss in case of overfitting after about 30 epochs. (right) Illustration of what happens in case of overfitting, optimum learning and underfitting.

Data augmentation

As mentioned above, overfitting shrinks as the number of independent training examples increases. That is pretty straightforward since, to generalise, the network needs to learn a larger part of the probability distribution of the data-generating process. Learning on larger samples obviously permits to cover a larger part of this distribution.

Since the amount of available data is limited, one solution could be to create new data from simulation. However this solution requires the knowledge of the probability distribution that is learned by the network, which is not always the case. Another solution is to apply transformation to the data for which the response of the desired output is known. For instance, if the desired output is the ellipticity parameter e_1 of a galaxy (see section 2.4.1), it is invariant if the galaxy image is flipped horizontally. One can use this property to increase the number of galaxy images in the training sample. This technique is commonly used and I personally use it in part III.

Dropout regularisation

Another strategy to mitigate overfitting is to use dropout (Srivastava et al., 2014). At first approximation, dropout can be seen as making bagging, i.e. train multiple neural network models, evaluate them on a test sample and combine the outputs. Bagging can become very computationally consuming, and dropout provides an inexpensive approximation of it by randomly removing, *dropping*, non-output units (neurons) from an underlying neural network model. This removal is most commonly done by multiplying some neuron’s outputs by zero. Figure 4.7 shows different sub-networks, created from one underlying architecture. Also, bagging requires training different neural networks on different samples, which is realized in the dropout approximation by working on minibatch of data: the removed neuron outputs are randomly changed after every minibatch provided to the network. The main difference from bagging is that the models are not independent, all different sub-networks share parameters. Dropout can also be seen as regularising every units in order to select features which are relevant in most context, which is highly important when aiming for generalisation.

Srivastava et al. (2014) showed that dropout is very efficient, computationally cheap, and works well with numerous different types of model (for an example on RNN see Bayer et al., 2013). However, dropout can also lead to underfitting as it may reduce too much the effective capacity of a model trained on a small training sample. It often results in an increased complexity of the

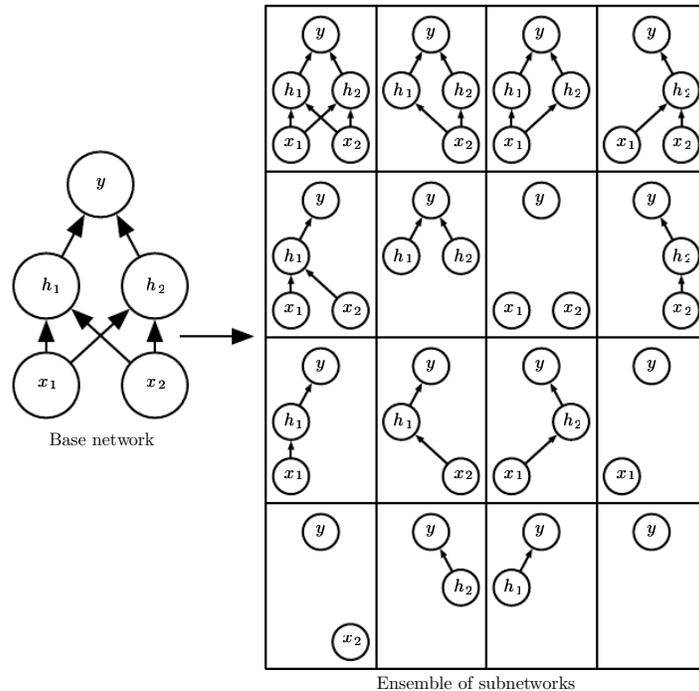


Figure 4.7: Ensemble of sub-networks from a single network architecture which can be created and trained using dropout by randomly removing non-output units. Credit: I. Goodfellow, Bengio, and Courville, 2016

neural network, with more layers and more neurons, to compensate.

In the work presented in chapter 7 and chapter 8, I faced several times overfitting that I managed to overcome with data augmentation or decreasing the capacity of my network (reducing the number of trainable parameters) for example. Discussing training issues here aimed at giving a hint of the difficulties that can arise with deep learning, but also at understanding a little bit more deeply how neural networks are actually trained. Deep learning can be very powerful, but knowing what to monitor, and the different solutions to avoid issues when training your network, is key to use it correctly.

4.3 Generative models

In this part, I describe some of the generative models which exist in deep learning literature. The main advantage over classical generation methods is that these generative models are directly learned from data, and allow for more complex models, closer to the data. I will describe in detail Variational Autoencoder (VAE), as it is at the core of one of my projects, and briefly discuss two other kinds of generative models: Generative Adversarial Networks (GAN) and normalizing flows.

4.3.1 Variational Autoencoder (VAE)

Variational autoencoders (VAE) (Kingma and Welling, 2014) are generative models that learn probabilistic mappings between a data space, where samples usually have a very complex distribution, and a latent space of well-distributed variables. As the work presented in this manuscript is mainly

oriented toward using cosmological surveys images, I focus on the case of image data sets, even if VAEs have been successfully applied to various types of data (for examples, see Wang et al., 2019; Roche et al., 2018).

More formally, image samples are denoted x and latent variables are denoted z with a known prior $p(z)$ chosen to be a standard multivariate gaussian (a common choice in the literature, see Kingma and Welling, 2014; Semeniuta, Severyn, and Barth, 2017, for examples). The generative model is a conditional distribution $p_\theta(x|z)$ (the likelihood) which can be parametrised by a dense or convolutional neural network with weights θ called the *decoder*. The latter computes parameters of this distributions from input latent variables z . The marginal likelihood $p_\theta(x)$ (or evidence) needs to be as close as possible to the real data distribution, so the parameters θ are trained to maximise:

$$p_\theta(x) = \int p(z)p_\theta(x|z)dz \quad (4.13)$$

Unfortunately this integral is most of the time intractable, as well as the true posterior density $p_\theta(z|x) = p_\theta(x|z)p_\theta(z)/p_\theta(x)$. To estimate this posterior, Monte Carlo methods could be used, but it is computationally costly. To get around this issue, the idea proposed by Kingma and Welling (2014) is to use variational inference. They introduce a second neural network (see fig. 4.8), the *encoder*, with weights denoted ϕ , to approximate the intractable posterior $p_\theta(z|x)$ by another distribution $q_\phi(z|x)$, which is trained using the reparameterisation trick (see section 5.3.2).

The VAE then maximises the evidence of the training sample $p_\theta(x)$, which cannot be computed exactly, but can be bounded from below by the Evidence Lower BOund (ELBO, see section 5.3.2 and Blei, Kucukelbir, and J. D. McAuliffe, 2016),

$$\log p_\theta(x) \geq -D_{\text{KL}}(q_\phi(z|x)||p(z)) + \mathbb{E}_{q_\phi}(\log p_\theta(x|z)), \quad (4.14)$$

which defines the variational loss function of the VAE. This loss receives two contributions. The first one is the Kullback-Leibler (KL) divergence between the approximate posterior and the prior. If this prior $p(z)$ is a normal distribution $p(z) \sim \mathcal{N}(0, 1)$, a common choice in the literature, the KL divergence ensures that the latent variables are following the gaussian distribution, and acts as a regularisation term. This divergence can be viewed as a measure of the information loss. If this part of the loss becomes null, the approximate posterior matches perfectly both the real posterior and the prior and then, with Baye's rule: $p(x) = p(x|z)$. This is exactly what one would wish from a generative model.

The second loss term is given by the expectation value of the log-likelihood and is a reconstruction term. It requires sampling latent variables according to the approximate posterior specified by the encoder.

When training a VAE with galaxy images (see chapter 7), the KL term outweighs the reconstruction term and the network is unable to learn important features (galaxies shapes in our case). The network struggles to learn features due to the strong constraints coming from the prior distribution, an effect that appears rapidly with complex datasets (Higgins et al., 2016). It can be necessary to move to a β -VAE (Higgins et al., 2016) which adds a β coefficient (lower than 1) in front of the KL term to reduce its impact during learning.

The work presented in part III and in Arcelin et al. (2021) is based on this kind of generative models and uses the β -VAE.

4.3.2 Generative Adversarial Networks (GAN)

GANs (I. J. Goodfellow et al., 2014) are also latent space based generative models but the learning process is different. It is also composed of two neural networks (see fig. 4.8): a generative network,

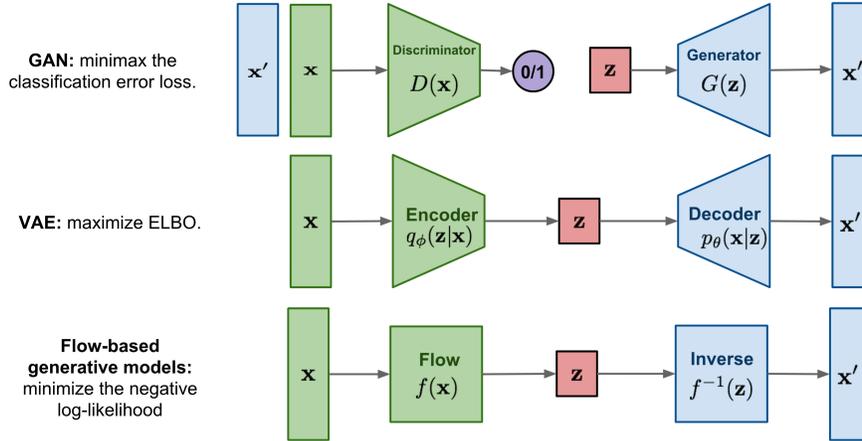


Figure 4.8: Architectures of the three presented kind of generative models. From top to bottom: Generative Adversarial Network, Variational AutoEncoder and generative models based on normalizing flows. Credit: Lilian Weng

creating a image from a latent space, and a discriminative network, which aims at distinguishing between data generated from the generative network and data from the true data distribution. The objective of the generative model is to increase the error rate of the discriminative network. Both networks are competing during training and the convergence is reached when the discriminator can not distinguish between samples from the generator and real data any more.

The discriminator gives a probability $D(x)$, that x comes from real data and the generator generates an image $G(z)$, z being the latent space. The generative network is trained to maximise the probability that the discriminator fails to distinguish between true and generated data, i.e. minimise $\log(1 - D(G(z)))$, while the discriminator aims at maximising the probability of assigning the correct label. This is a minimax game with value function $V(D,G)$:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (4.15)$$

Similarly to VAE, most of the time the latent space, z , is considered to be a multivariate Gaussian distribution.

GAN show great performance at generating all kinds of data and it was the first kind of deep generative models used to be used for the deblending of galaxies, as mentioned in section 3.3.2 (Reiman and Göhre, 2019).

4.3.3 Normalizing flows

Finally, normalizing flows (see Kobyzev, Prince, and Brubaker, 2019 or Papamakarios et al., 2019 for reviews) are stacks of invertible and differentiable transformations, f , which map a simple probability distribution into a more complex one. The new, complex density can then be sampled from by sampling the initial simple density, and applying transformations. A major benefit of normalizing flows is that likelihood can be exactly computed, contrary to VAEs where it is bounded from below. The trick is to make a neural network invertible by using *coupling functions*. The neural network is then called a *conditioner*, and the whole process is a *coupling flow* (Kobyzev, Prince, and Brubaker, 2019). The power of this method is that the conditioner can then be arbitrarily complex and can map any complex function.

Auto-regressive normalizing flows add the auto-regressive constraint: each output only depends on the previously observed data and not on future inputs. Stated differently, the probability of observing x_i is only conditioned by x_1, x_2, \dots, x_{i-1} , i.e. it is a regression of the variable against itself. Auto-regressive networks are mostly used for time series but van den Oord, Kalchbrenner, and Kavukcuoglu (2016) established a method that proves to be highly performing at processing images. The difference with coupling flows is that the conditioner is now an auto-regressive neural network (see for examples Papamakarios, Pavlakou, and Murray, 2017 or Kingma et al., 2016).

In the next section I present an example where I used normalising flows to learn the latent space distribution of a trained VAE in order to generate galaxy images.

4.4 Example: Generate images from a VAE latent space

The work presented in this section was done in the context of a study comparing the performance of Nvidia Graphical Processing Units (GPUs) and Graphcore Intelligence Processing Units (IPUs) for cosmological applications. To know more about the hardware comparison, please refer to appendix A.

The generation of galaxy images using deep generative models is more and more considered as a credible option for image simulation (Regier, J. McAuliffe, and Prabhat, 2015; Ravanbakhsh et al., 2016; Lanusse et al., 2021). Recently, Bretonnière et al. (2021) used a VAE coupled with normalising flows to propose a galaxy image generative model to simulate the sky as it will be seen by Euclid. This is why I chose the generation of galaxy images using deep generative models as example to compare the performance of the hardware for inference.

In this section, I use the VAE presented in section 7.3 with corresponding results shown in section 7.4.2. I do not present the architecture of the network or its training procedure in detail, as I focus on the generation process here. To summarize, both the encoder and the decoder of the VAE are convolutional neural networks and the VAE is trained using the β coefficient in front of the KL divergence as explained above (see section 4.3.1).

In this example, the generation process requires the sampling of the approximate posterior distribution, $q_\phi(z|x)$, over the latent space of the trained VAE. This sample is then fed to the decoder to generate the likelihood $p_\theta(x|z)$. For our VAE we consider the output distribution as the output image of noiseless galaxy generated by the network.

The first subsection presents how the latent space dimensions are distributed after the VAE training and some outputs from the decoder when the latent space is sampled using normal distributions, which is the prior used for training. The second subsection shows that using normalizing flows permits to learn the latent space distribution and to sample it properly.

4.4.1 Direct sampling of the latent space

In the case of a VAE trained to reproduce noiseless isolated galaxy images, sampling the latent space to feed the trained decoder should allow for the production of noiseless, isolated galaxy images. To train this VAE we considered a normal distribution $p(z) \sim \mathcal{N}(0, 1)$ as prior for the approximate posterior distribution of each of the latent space dimensions. However, because of the β coefficient in front of the KL divergence, even though the distributions of the latent space dimensions resemble the normal distribution, they are not exactly normally distributed (see upper panel of fig. 4.9). The β coefficient in front of the KL term in the loss decreases the influence of the prior sufficiently for the distribution to be slightly different, i.e. mainly with a larger variance than the normal distribution.

However, to generate galaxy images, one must sample the approximate posterior distribution over the latent space $q_\phi(z|x)$ to feed the decoder. Since we do not know the exact distribution, the

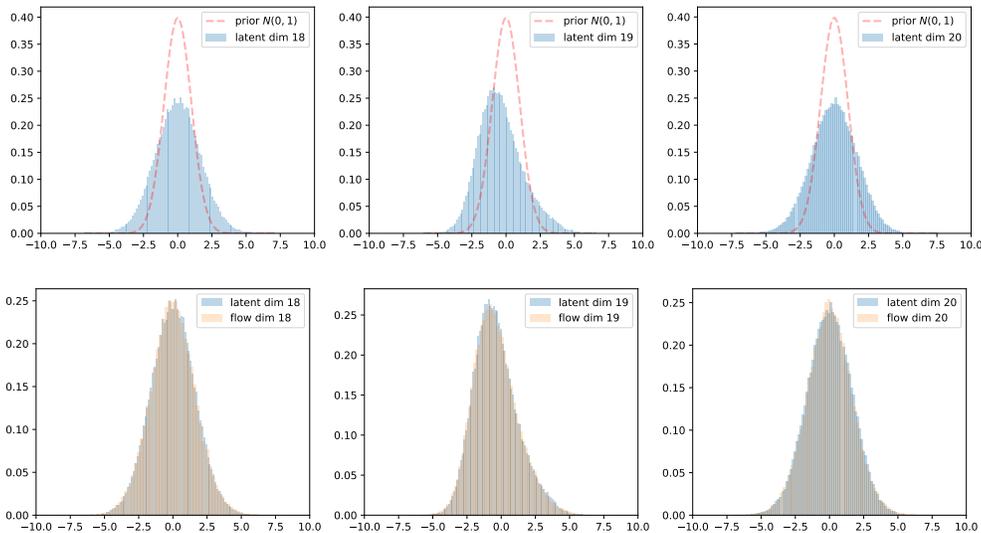


Figure 4.9: (Upper panel) Distribution obtained via sampling the dimensions 18, 19 and 20 of the latent space of a trained VAE (in blue). In red the normal distribution used as prior to train the VAE is drawn. (Lower panel) Distribution obtained via sampling the dimensions 18, 19 and 20 of the latent space of a trained VAE (in blue). In orange, the distribution obtained via sampling the normalizing flow. Both distribution overlap for latent space dimensions 18, 19 and 20, showing that the normalizing flow accurately learned the mapping from the normal distribution to the distribution of the latent space dimensions.

closest thing that we can do, since the prior $p(z)$ is a normal distribution, is sample a multivariate Gaussian distribution (normally distributed) which dimension fits the dimension of the latent space, and feed the sample to the decoder. The upper panel of fig. 4.10 shows images generated by the trained decoder when fed with a 32-vector, with each term sampled from a normal distribution. Since the distributions of latent space dimensions are not exactly normal, the decoder produces weirdly shaped galaxy images. Of course, these images can not be used to simulate real galaxies for a cosmological analysis.

4.4.2 Sampling with Normalizing flows

Lanusse et al. (2021) faced a similar issue and used normalizing flows to learn the approximate posterior distribution over latent space, and sample from it. As described in section 4.3.3, normalizing flows are able to learn how to map any distribution into another thanks to a series of small invertible transformations.

Based on this work, I use Masked Autoencoder for Distribution Estimation (MADE, Germain et al., 2015) to build the normalizing flow. It is trained to learn the mapping between a multivariate Gaussian distribution of size 32 and the variational posterior distribution $q_\phi(z|x)$ of the latent space. From there, it is possible to sample the latent space distribution via the normalizing flow and use the decoder to generate galaxy images.

The lower panel of fig. 4.9 shows the distributions learned by the normalising flow compared to the corresponding distributions over three selected dimensions of the latent space. It shows that the normalising flow accurately learned how to map the multivariate Gaussian distribution into the variational posterior distribution for 3 of the 32 dimensions. The lower panel of fig. 4.10

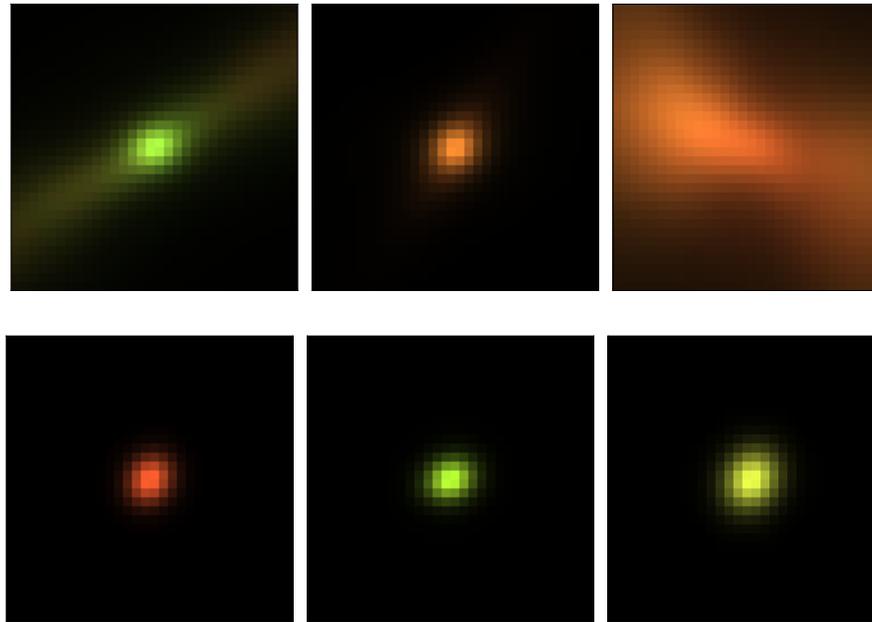


Figure 4.10: (Upper panel) Images obtained via feeding the decoder of a trained VAE with a vector from which each term is sampled from the normal distribution. (Lower panel) Images obtained via feeding the decoder of a trained VAE with the sample produced by the normalizing flow.

presents three galaxy images generated via feeding the trained decoder with a sample from the latent space obtained through the normalising flow. The galaxy shapes look much more reasonable than previously, qualitatively showing that this step is necessary if this kind of methods is used for cosmological simulations in the future. Bretonnière et al. (2021) use the same process for their galaxy image generation for Euclid.

4.4.3 Conclusion

This work aimed at designing a toy model for galaxy image generation using a trained VAE for hardware performance comparison (see appendix A). It presents a very simple example of how galaxies can be simulated from the latent space of a trained VAE using normalizing flows.

As already mentioned, generating galaxy images with deep generative models is a more and more considered option as deep learning offers the possibility of fast inference, permitting the generation of a lot of galaxy images in a very short amount of time. The recent example of Bretonnière et al. (2021), based on the work of Lanusse et al. (2021), who implemented a method for galaxy image simulation for the Euclid survey, is the first to be implemented within a cosmological survey simulation pipeline.

In this chapter, we saw some of the basics elements of deep learning, how to train a model, and different kinds of deep generative models. Deep learning is a very vast field of research and I only gave a tiny glimpse of what can be done with it. In the next chapter, we will see what Bayesian neural networks are, how to train them, and most importantly what additional information they

can provide.

Bibliography

- Arcelin, Bastien et al. (Jan. 2021). “Deblending galaxies with variational autoencoders: A joint multiband, multi-instrument approach”. In: *MNRAS* 500.1, pp. 531–547. DOI: [10.1093/mnras/staa3062](https://doi.org/10.1093/mnras/staa3062). arXiv: [2005.12039](https://arxiv.org/abs/2005.12039) [[astro-ph.IM](#)].
- Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2014). “Neural machine translation by jointly learning to align and translate”. In: *arXiv preprint arXiv:1409.0473*.
- Bayer, Justin et al. (2013). “On fast dropout and its applicability to recurrent networks”. In: *arXiv preprint arXiv:1311.0701*.
- Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe (Jan. 2016). “Variational Inference: A Review for Statisticians”. In: *arXiv e-prints*, arXiv:1601.00670, arXiv:1601.00670. arXiv: [1601.00670](https://arxiv.org/abs/1601.00670) [[stat.CO](#)].
- Blumer, Anselm et al. (Oct. 1989). “Learnability and the Vapnik-Chervonenkis Dimension”. In: *J. ACM* 36.4, pp. 929–965. ISSN: 0004-5411. DOI: [10.1145/76359.76371](https://doi.org/10.1145/76359.76371). URL: <https://doi.org/10.1145/76359.76371>.
- Bretonnière, H. et al. (May 2021). “Euclid preparation: XVI. Forecasts for galaxy morphology with the Euclid Survey using Deep Generative Models”. In: *arXiv e-prints*, arXiv:2105.12149, arXiv:2105.12149. arXiv: [2105.12149](https://arxiv.org/abs/2105.12149) [[astro-ph.GA](#)].
- Devlin, Jacob et al. (2018). “Bert: Pre-training of deep bidirectional transformers for language understanding”. In: *arXiv preprint arXiv:1810.04805*.
- Fukushima, Kunihiko (1979). “Neural network model for a mechanism of pattern recognition unaffected by shift in position-Neocognitron”. In: *IEICE Technical Report, A* 62.10, pp. 658–665.
- Gabbard, Hunter et al. (Sept. 2019). “Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy”. In: *arXiv e-prints*, arXiv:1909.06296, arXiv:1909.06296. arXiv: [1909.06296](https://arxiv.org/abs/1909.06296) [[astro-ph.IM](#)].
- Germain, Mathieu et al. (Feb. 2015). “MADE: Masked Autoencoder for Distribution Estimation”. In: *arXiv e-prints*, arXiv:1502.03509, arXiv:1502.03509. arXiv: [1502.03509](https://arxiv.org/abs/1502.03509) [[cs.LG](#)].
- Goodfellow, Ian J. et al. (June 2014). “Generative Adversarial Networks”. In: *arXiv e-prints*, arXiv:1406.2661, arXiv:1406.2661. arXiv: [1406.2661](https://arxiv.org/abs/1406.2661) [[stat.ML](#)].
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- He, Kaiming et al. (Dec. 2015a). “Deep Residual Learning for Image Recognition”. In: *arXiv e-prints*, arXiv:1512.03385, arXiv:1512.03385. arXiv: [1512.03385](https://arxiv.org/abs/1512.03385) [[cs.CV](#)].
- (Feb. 2015b). “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”. In: *arXiv e-prints*, arXiv:1502.01852, arXiv:1502.01852. arXiv: [1502.01852](https://arxiv.org/abs/1502.01852) [[cs.CV](#)].
- Higgins, Irina et al. (2016). “beta-vae: Learning basic visual concepts with a constrained variational framework”. In:
- Hochreiter, Sepp and Jürgen Schmidhuber (1997). “Long short-term memory”. In: *Neural computation* 9.8, pp. 1735–1780.

- Ioffe, Sergey and Christian Szegedy (Feb. 2015). “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *arXiv e-prints*, arXiv:1502.03167, arXiv:1502.03167. arXiv: [1502.03167 \[cs.LG\]](#).
- Kingma, Diederik P. and M. Welling (2014). “Auto-Encoding Variational Bayes”. In: *CoRR* abs/1312.6114.
- Kingma, Diederik P. et al. (June 2016). “Improving Variational Inference with Inverse Autoregressive Flow”. In: *arXiv e-prints*, arXiv:1606.04934, arXiv:1606.04934. arXiv: [1606.04934 \[cs.LG\]](#).
- Kobyzev, Ivan, Simon J. D. Prince, and Marcus A. Brubaker (Aug. 2019). “Normalizing Flows: An Introduction and Review of Current Methods”. In: *arXiv e-prints*, arXiv:1908.09257, arXiv:1908.09257. arXiv: [1908.09257 \[stat.ML\]](#).
- Lanusse, François et al. (July 2021). “Deep generative models for galaxy image simulations”. In: *MNRAS* 504.4, pp. 5543–5555. DOI: [10.1093/mnras/stab1214](#). arXiv: [2008.03833 \[astro-ph.IM\]](#).
- LeCun, Yann et al. (1999). “Object Recognition with Gradient-Based Learning”. In: p. 319.
- Papamakarios, George, Theo Pavlakou, and Iain Murray (May 2017). “Masked Autoregressive Flow for Density Estimation”. In: *arXiv e-prints*, arXiv:1705.07057, arXiv:1705.07057. arXiv: [1705.07057 \[stat.ML\]](#).
- Papamakarios, George et al. (Dec. 2019). “Normalizing Flows for Probabilistic Modeling and Inference”. In: *arXiv e-prints*, arXiv:1912.02762, arXiv:1912.02762. arXiv: [1912.02762 \[stat.ML\]](#).
- Radford, A. et al. (2019). “Language Models are Unsupervised Multitask Learners”. In:
- Ravanbakhsh, Siamak et al. (Sept. 2016). “Enabling Dark Energy Science with Deep Generative Models of Galaxy Images”. In: *arXiv e-prints*, arXiv:1609.05796, arXiv:1609.05796. arXiv: [1609.05796 \[astro-ph.IM\]](#).
- Regier, Jeffrey, J. McAuliffe, and Prabhat (2015). “A deep generative model for astronomical images of galaxies”. In:
- Reiman, David M. and Brett E. Göhre (May 2019). “Deblending galaxy superpositions with branched generative adversarial networks”. In: *MNRAS* 485.2, pp. 2617–2627. DOI: [10.1093/mnras/stz575](#). arXiv: [1810.10098 \[astro-ph.IM\]](#).
- Roche, Fanny et al. (June 2018). “Autoencoders for music sound modeling: a comparison of linear, shallow, deep, recurrent and variational models”. In: *arXiv e-prints*, arXiv:1806.04096, arXiv:1806.04096. arXiv: [1806.04096 \[eess.AS\]](#).
- Sak, Haşim, Andrew Senior, and Françoise Beaufays (Feb. 2014). “Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition”. In: *arXiv e-prints*, arXiv:1402.1128, arXiv:1402.1128. arXiv: [1402.1128 \[cs.NE\]](#).
- Semeniuta, Stanislau, Aliaksei Severyn, and Erhardt Barth (Feb. 2017). “A Hybrid Convolutional Variational Autoencoder for Text Generation”. In: *arXiv e-prints*, arXiv:1702.02390, arXiv:1702.02390. arXiv: [1702.02390 \[cs.CL\]](#).
- Srivastava, Nitish et al. (Jan. 2014). “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *J. Mach. Learn. Res.* 15.1, pp. 1929–1958. ISSN: 1532-4435.
- van den Oord, Aaron, Nal Kalchbrenner, and Koray Kavukcuoglu (Jan. 2016). “Pixel Recurrent Neural Networks”. In: *arXiv e-prints*, arXiv:1601.06759, arXiv:1601.06759. arXiv: [1601.06759 \[cs.CV\]](#).
- Vapnik, Vladimir (1982). *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics)*. Berlin, Heidelberg: Springer-Verlag. ISBN: 0387907335.
- (2013). *The nature of statistical learning theory*. Springer science & business media.
- Vapnik, Vladimir N and A Ya Chervonenkis (2015). “On the uniform convergence of relative frequencies of events to their probabilities”. In: *Measures of complexity*. Springer, pp. 11–30.
- Vaswani, Ashish et al. (2017). “Attention is all you need”. In: *arXiv preprint arXiv:1706.03762*.
- Wang, Wenlin et al. (Mar. 2019). “Topic-Guided Variational Autoencoders for Text Generation”. In: *arXiv e-prints*, arXiv:1903.07137, arXiv:1903.07137. arXiv: [1903.07137 \[cs.CL\]](#).

Zhang, Han et al. (2019). “Self-attention generative adversarial networks”. In: *International conference on machine learning*. PMLR, pp. 7354–7363.

Chapter 5

Introduction to Bayesian neural networks

As described in the previous section, deep neural networks are a powerful tool to create models of high complexity enabling the execution of complicated tasks. However, the process described above does not provide a way to estimate the uncertainty introduced by the choice of the neural network architecture or the composition of the training sample, on the output. This part presents Bayesian deep learning, a way to characterise both the uncertainty introduced by the architecture and the training sample, the *epistemic* uncertainty, and the one intrinsic of the data, the *aleatoric* uncertainty.

This chapter is mostly inspired from Charnock, Perreault-Levasseur, and Lanusse (2020) and Valentin Jospin et al. (2020).

5.1 Uncertainties for neural networks

To model the PSF of an astronomical image, one needs to identify the stars on this image (see section 2.5.2). Imagine that we use a neural network to classify stars and galaxies by feeding patches of the image centred on sources detected by a detection algorithm. The neural network would provide a probability for each object to be in both categories, either a star or a galaxy. Now what happens if the network is fed with a patch that contains a satellite trail ? Ideally, one would want the network to tell us that it does not know what it is seeing, it never saw a similar example in its training sample and it is *uncertain* about its prediction. Most of neural networks are not equipped to provide information about outlier data, but Bayesian neural networks (BNN), fill this gap by providing the measurement of this uncertainty.

5.1.1 Aleatoric and Epistemic uncertainty

Uncertainties can be categorised in two classes: aleatoric and epistemic uncertainties. The former is the uncertainty which is intrinsic of the data, it comes from the random nature of the data-generating process and the way we perform experiments.

The latter, epistemic uncertainty, is coming from everything we ignore but could, in principle, learn about. By increasing our knowledge, for example via a better theoretical understanding of a data-generating process or a model more suited to the process that is analysed, we could in principle decrease this uncertainty.

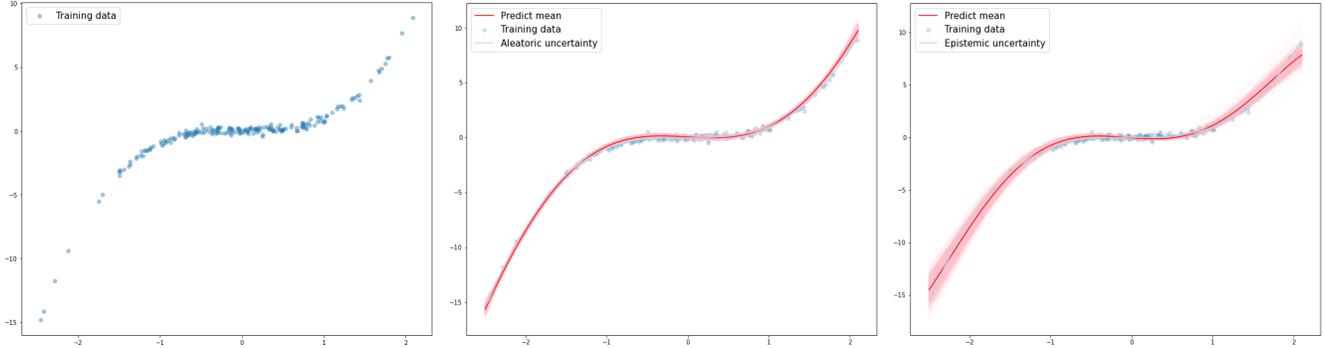


Figure 5.1: (left) A dataset generated by a model that we are trying to learn with neural networks. (center) The results from a point estimate neural network predicting a Gaussian distribution. The mean of this distribution is plotted in red, and its uncertainty (here the standard deviation) is shown in light red. This network only learns the aleatoric uncertainty. Even though the model is equally noisy everywhere, the network is only able to learn this uncertainty correctly in the center, where there are enough data points. (right) Results obtained with a Bayesian neural network predicting a Gaussian distribution. Its mean is plotted in red. The uncertainty (light red) shown on this panel is the epistemic uncertainty. It is composed of the different means that predict the network when fed several times with the test sample. It shows that the network is able to provide information about the confidence level in its network’s prediction. As expected, the uncertainty is high (low confidence level) where there are few data points. Credit: Justine Zeghal.

Training a neural network is equivalent to fitting a model to realise a specific task, consequently the choice of the neural network architecture, all the hyperparameters coming with it (cost function, size of the network, choice of activation layers...), and most importantly the parameters space covered by the training sample (for example the size, shapes, or redshifts of galaxies), are taking part in the epistemic uncertainty. It provides information about the ability of the neural network to generalise (see section 4.2.3). If an element from the test sample is outside the parameters space covered by the training sample, the network will measure a high epistemic uncertainty. On the contrary, it should be small if the test sample lays in the training sample parameters space.

5.1.2 Example: fit a dataset

This is illustrated on fig. 5.1. The dataset shown on the left panel is fitted with two neural networks, one "classical"¹ and one Bayesian, i.e. with distributions over the weights instead of point estimates (see below). They both output a Gaussian distribution, $\mathcal{N}(\mu, \sigma)$.

On the center panel, we show the results from the point estimate neural network. The mean of the predicted distributions for a test sample is shown in red, and the uncertainty measured by the network in light red. Here, since it is a point estimate neural network, this uncertainty is only coming from the data, it is the aleatoric uncertainty. The data-generating process generates a similar amount of noise on each portion of the data set but, as we can see, the network is only able to correctly measure this uncertainty where there is a large enough number of data points. Where the density is lower, it predicts an incorrectly small uncertainty, not large enough to represent the

¹In the rest of this manuscript I will equivalently use the terms "deterministic", "classical", or "point estimate" network to refer to a neural network which has point estimate weights.

predictive error. For example, the points around $x \approx 1.75$ are outside the 1σ confidence interval.

This is not acceptable if one wants to use the output of the neural network in a scientific analysis, in cosmology for example. This is why we need to estimate the epistemic uncertainty, shown on the right panel. Here, we plotted the mean of the distributions predicted by the Bayesian neural network when fed several time with the test sample. In darker red is the mean of these samples. The network measures a high uncertainty when there are few data points and small uncertainty where there are a lot, showing that this uncertainty can provide information about when to trust, or not, the predictions of the network.

To use neural network predictions for scientific studies, it is necessary to fully characterise and understand this uncertainty. The Bayesian neural network formalism described below provides means to estimate and study it.

5.2 Bayesian neural networks: Definition

In their review, Valentin Jospin et al. (2020) define *Bayesian neural networks* as stochastic neural networks trained using Bayesian inference. In the next sections, I will try to describe both parts of this definition. This section will explain how we can consider neural networks as statistical models, and in the next section, I will describe some methods to train stochastic neural networks.

5.2.1 Bayes' rule

The point of Bayesian statistics is to be able to update our prior state of knowledge to a more accurate posterior one, using data. This principle can be summarised in the Bayes' theorem formula:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (5.1)$$

$P(Y|X)$ is called the posterior, $P(Y)$ is the prior, $P(X|Y)$ is the likelihood, and $P(X)$ is the evidence. In the context of neural networks, X would be the data and Y would be the target prediction of the model.

As mentioned earlier, training a deep neural network is optimising the set of parameters θ , using the data X , to predict Y . In the last chapter, we considered those weights and biases as point estimates and trained them with the backpropagation algorithm (see section 4.2). This is the standard approach to train a network, but doing so, all other weights and biases parametrisations are discarded. Also, in theory, the fact that the architecture chosen for the neural network allows to learn the target distribution is not guaranteed. Consequently, the probability, $p(\alpha)$, that the architecture of the neural network allows for the reproduction of the target distribution, α , must be considered.

5.2.2 Neural networks as statistical models

Taking these elements into account, and if we do not discard all possible parametrisations of the weights and biases, but rather try to learn Bayesian posterior distributions over these parameters (see fig. 5.2), it is possible to consider neural networks as statistical models.

To obtain this posterior, we need to chose a prior distribution $P_\alpha(\theta)$ for the weights, which depends on the distribution α which can be modeled by the chosen neural network architecture. Then, for a training set $D = (X, Y)$, X being the input and Y the network target prediction, the

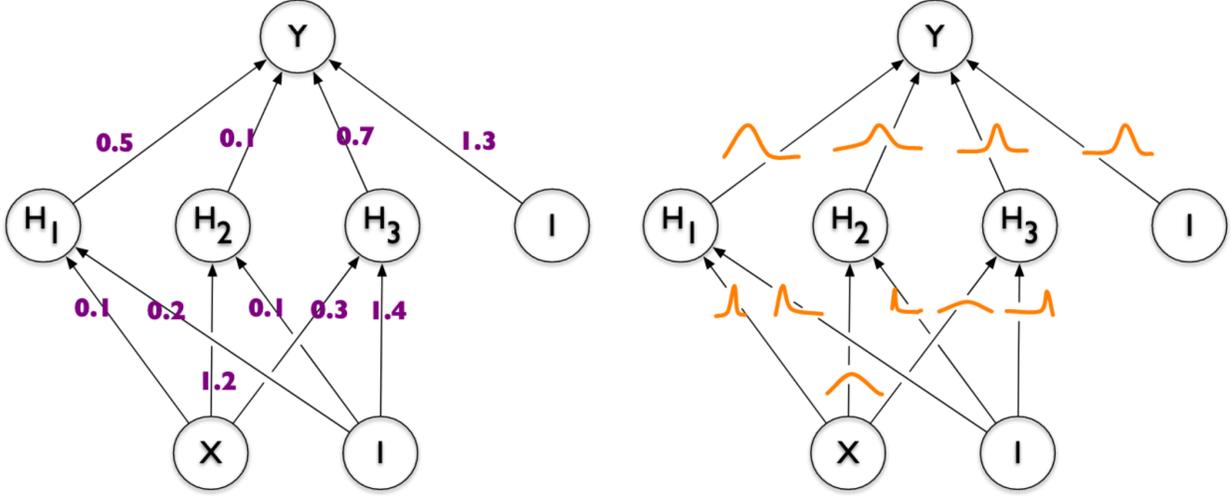


Figure 5.2: (left) Point estimate neural network, the weights have a fixed value learned from backpropagation. (right) A Bayesian neural network, each weights has an learned probability distribution. Credit: Blundell et al. (2015)

Bayesian posterior distribution on the weights can be written as:

$$P(\theta|D) = \frac{P(Y|X, \theta)P_\alpha(\theta)}{P(D)} \quad (5.2)$$

$P(\theta|D)$ represents the posterior distribution of the parameters knowing the data, and $P(D|\theta) = P(Y|X, \theta)$ is the corresponding likelihood.

Given this definition, it becomes possible to evaluate the posterior predictive distribution, $P(y|x, D)$, $x \in X^1 \neq X$ and $y \in Y^1 \neq Y$, from a test set (X^1, Y^1) , and characterise the epistemic uncertainty of our model. To do so, it is necessary to marginalise over all possible distributions that can be fitted by a neural network, i.e. over all possible architectures and over all weights and biases possible parametrisation:

$$P(y|x, D) = \int_A \int_{\Theta} d\theta d\alpha P(y|x, \theta) P_\alpha(\theta|D) P(\alpha) \quad (5.3)$$

with A and Θ the ensemble of values that α and θ can take.

Here, an issue arises: in theory, it would be necessary to marginalise over all possible neural network architectures to accurately estimate the epistemic uncertainty, but this is obviously not feasible. In practice, only one particular model is chosen depending on the task that is targeted. This is a common approximation as neural networks are, in general, able to fit arbitrarily complex distribution. Equation (5.3) then becomes:

$$\begin{aligned} P(y|x, D) &= \int_A \int_{\Theta} d\theta d\alpha P(y|x, \theta) P_\alpha(\theta|D) \delta(\alpha - \alpha') \\ &= \int_{\Theta} d\theta P(y|x, \theta) P_\alpha(\theta|D) \end{aligned} \quad (5.4)$$

However, even after choosing one particular architecture for the neural network, the marginal distribution $P(y|x, D)$ in eq. (5.2) is still intractable. Indeed, the posterior distribution $P(\theta|D)$ is

most of the time not tractable because computing the evidence $P(D)$ is very difficult. Instead, it is possible to approximate the posterior distribution $P_\alpha(\theta|D) \sim p(\theta|D)$ by a tractable distribution q using variational inference, or to approximate the integral using Monte Carlo methods.

Note: to go back to point estimate neural networks, considering only one model and one version of the weights in the previous integral yields:

$$P(y|x, D) = \int_A \int_{\Theta} d\theta d\alpha P(y|x, \theta) \delta(\theta - \theta') \delta(\alpha - \alpha') = \hat{P}_{|\alpha'=\alpha}(y|x, \theta') \quad (5.5)$$

Maximizing this probability is equivalent to computing the maximum likelihood estimate of the parameters distribution. The set of parameters which best describes the event $(y|x)$ is:

$$\hat{\theta} = \operatorname{argmax}_{\theta} P(y|x, \theta). \quad (5.6)$$

However, to do Bayesian inference, the aim is to find the set of parameters which yields the most likely point estimate from the posterior distribution, and not from the likelihood, which is (with eq. (5.2)):

$$\hat{\theta} = \operatorname{argmax}_{\theta} P(\theta|D) = \operatorname{argmax}_{\theta} \frac{P(y|x, \theta)P(\theta)}{P(D)} = \operatorname{argmax}_{\theta} P(y|x, \theta)P(\theta) \quad (5.7)$$

$P(\theta)$ being the prior distribution of the parameters. This is the maximum a posteriori (MAP) estimate. Going to log-likelihood formulation this expression becomes:

$$\hat{\theta} = \operatorname{argmax}_{\theta} P(y|x, \theta)P(\theta) = \operatorname{argmin}_{\theta} (-\log(P(y|x, \theta)) - \log(P(\theta))) = \operatorname{argmin}_{\theta} \operatorname{loss}_{y,x}(\theta) + \operatorname{reg}(\theta) \quad (5.8)$$

loss being the loss function and *reg* being a regularisation function. This familiar formulation results from applying a regularisation with a prior of the form: $P(\theta) \propto e^{-\operatorname{reg}(\theta)+cst}$.

5.3 Approximate inference: Monte-Carlo and Variational Inference

Here, I will mostly describe how to approximate $p(\theta|D)$ thanks to variational inference (VI) and Monte Carlo methods. I will emphasize more on VI avenues as these are the methods that I used in my work. The part on Monte-Carlo methods mostly comes from Goodfellow, Bengio, and Courville (2016) and the one on VI is based on Blei, Kucukelbir, and McAuliffe (2016) and Charnock, Perreault-Levasseur, and Lanusse (2020).

5.3.1 Monte-Carlo methods

The principle of Monte-Carlo methods is to sample distributions in order to approximate integrals over these distributions. In our case it is used to approximate the intractable integral over the posterior distribution $p(\theta|D)$. The idea is to view the integral as an expectation under some distribution and to approximate the expectation by a corresponding average:

$$s = \int p(\theta)f(\theta)d\theta = \mathbb{E}_p[f(\theta)] \quad (5.9)$$

with p the probability distribution over θ . To refer to eq. (5.4), here s corresponds to $P(y|x, D)$, $p(\theta)$ to $p_\alpha(\theta|D)$ and $f(\theta)$ to $P(y|x, \theta)$.

It is then possible to approximate s by drawing n samples $(\theta_1, \theta_2, \dots, \theta_n)$ from p and averaging:

$$\hat{s}_n = \frac{1}{n} \sum_{i=1}^n f(\theta_i) \quad (5.10)$$

If θ_i are independent and identically distributed (i.i.d.), and if the variance of individual terms, $\text{Var}[f(\theta_i)]$, is bounded, the sum converges to the expected value s as n increases. The central limit theorem states that the distribution of \hat{s}_n converges to a Gaussian distribution with mean s and variance $\frac{\text{Var}[f(\theta)]}{n}$. It is then possible to have an estimation $\hat{p}_n(y|x, D)$ of $p(y|x, D)$ with corresponding asymptotic confidence intervals.

However, to do so we must be able to sample $p(\theta)$ (or $p(\theta|D)$ in our case), which is not always possible. It is then possible to use Monte-Carlo Markov Chain, for example. The principle is to form a sequence of estimators that converge toward the distribution of interest. I won't go into more details as I am not using this kind of methods in my work but Koller and Friedman (2009) give a detailed description of Monte Carlo methods for machine learning.

5.3.2 Variational inference

While the posterior distribution might be approximated by Monte-Carlo methods, it is computationally costly. In variational inference, the idea is to approximate the intractable posterior distribution $p(\theta|D)$ with a chosen, tractable, variational distribution $q_\phi(\theta)$. One aims at finding the variational parameters ϕ which minimise the difference between $p(\theta|D)$ and $q_\phi(\theta)$. To quantify this difference, the Kullback-Leibler divergence (KL divergence, D_{KL}) is commonly used in VI. The cost function now includes this new term (see for example Blei, Kucukelbir, and McAuliffe, 2016):

$$ELBO = -D_{KL}(q_\phi(\theta)||p(\theta|D)) + \log p(D) \quad (5.11)$$

As in section 4.3.1, the KL term measures the loss of information between the two distributions. The objective is to find parameters ϕ for the distribution $q_\phi(\theta)$ as close as possible to the true posterior $p(\theta|D)$. As the evidence, $p(D)$ does not depend on ϕ , maximising the ELBO (for Evidence Lower Bound, as it is a lower bound of the evidence) is equivalent to minimising the KL divergence.

The ELBO can also be written as:

$$ELBO = \mathbb{E}_{q_\phi(\theta)}(\log p(D|\theta)) - D_{KL}(q_\phi(\theta)||p(\theta)) \quad (5.12)$$

with $p(\theta)$, the prior distribution for the variational posterior, $q_\phi(\theta)$. Most of the time these terms can not be computed analytically and are estimated through Monte-Carlo sampling during training:

$$ELBO \approx \frac{1}{L} \sum_{l=1}^L [\log p(D|\theta^{(l)}) - \log q_\phi(\theta^{(l)}) + \log p(\theta^{(l)})] \quad (5.13)$$

with $\theta^{(l)} \sim q_\phi(\theta)$ being the l -th Monte-Carlo sample drawn from the variational posterior.

Now, to learn the posterior, one needs to backpropagate using the gradient of the loss that we just computed, with respect to each parameter. However, taking the gradient of the Monte-Carlo estimate of the loss with respect to ϕ , and especially computing $\nabla_\phi q_\phi(\theta)$, is most of the time not possible. Stochasticity stops backpropagation from working.

To circumvent this issue, several methods exist and I describe few of them below. Note that these methods are only a small part of the methods that can be used. They are however among the most popular, and I present these ones in particular since I used them in my work or in works with collaborators.

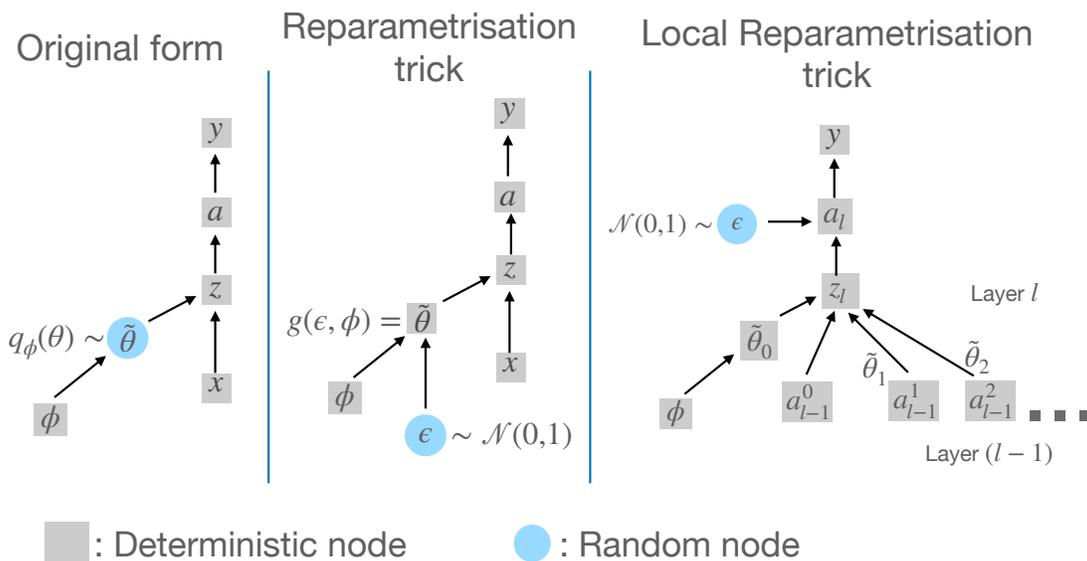


Figure 5.3: Diagram explaining the reparametrisation trick and the local reparametrisation trick. The main idea is to shift the randomness to another distribution by reparametrising the weight distributions or the implied activation distributions.

Reparameterisation trick

The method known as *Bayes-by-backprop* is based on the reparameterisation trick. It was presented by Kingma and M. Welling (2014), initially to train variational autoencoder (see section 4.3.1, I use it in chapter 7 to train a galaxy deblender based on variational autoencoders). Applied to Bayesian neural network in Blundell et al. (2015), the idea here is to “move” the randomness from the weight distribution, to an other distribution that parametrises the weight distribution (see middle panel of fig. 5.3). It is this parametrising distribution that is sampled during training and backpropagation, allowing to backpropagate until the parameters of the weight distribution.

More formally, it is possible to reparametrise the variable $\tilde{\theta} \sim q_{\phi}(\theta)$ (or $\tilde{z} \sim q_{\phi}(z|x)$ for VAE) using a differentiable deterministic function $g(\epsilon, \phi)$ with ϵ being an independent and non-variational noise variable.

The most common example is to reparametrise a Gaussian distribution with parameters $\phi = (\mu, \sigma)$:

$$\begin{aligned} \tilde{\theta} \sim q_{\phi}(\theta) &= \mathcal{N}(\mu, \sigma) \\ &= \mu + \sigma \epsilon \end{aligned} \tag{5.14}$$

with $\epsilon \sim \mathcal{N}(0, 1)$ an independent noise variable. Here, θ is a random variable parameter of the neural network (or a latent space for a VAE). This reparameterisation is done for each parameter and it is then possible to maximise the ELBO via backpropagation.

An issue with *Bayes-by-backprop* is that it is very costly to draw different parameters for each example in the training data. Consequently, the training data is usually split into batches of a fix number of elements and a single sample of the weight distributions is done per batch. However,

since elements in a batch might not well represent the variability of the parameters distribution, using a single sample of the weights to backpropagate often leads to high variance of the gradient estimate.

The local reparameterisation trick

The local reparameterisation trick (Kingma, Salimans, and Max Welling, 2015) aims for improvement in speed and for variance reduction. The idea is to transform the perturbation on parameters to a perturbation on the activation matrix \mathbf{A} , which has a much lower dimension than the parameters matrix. For example, if we assume Gaussian posterior distribution for the weights and biases, then the posterior on the activation is also a factorized Gaussian. Sampling this implied Gaussian distribution allows to reduce the drastically the computational cost.

To illustrate this, let's look at the right part of fig. 5.3. Here, and in the case of a neural network composed of fully connected layers, the last layer l is composed of only one neuron, but it receives input from the N neurons from the layer $(l - 1)$. If the reparametrisation trick is used, a sample must be drawn for each weight involved in the computation of the value of z_l , that is N times. On the contrary, if the local reparametrisation trick is applied, only one sample of the activation layer a_l is required.

Putting this into equation, for a Gaussian posterior on the weights and biases, the posterior on the activation, conditioned on the layer input \mathbf{Z} , is also a factorized Gaussian.

$$f(\theta_{l,j}) = \mathcal{N}(\mu_{l,j}, \sigma_{l,j}^2), \quad \forall \theta_{l,j} \in \Theta \implies f(a_{m,j} | \mathbf{Z}) = \mathcal{N}(\gamma_{m,j}, \delta_{m,j})$$

$$\gamma_{m,j} = \sum_{l=1} z_{m,l} \mu_{l,j}, \quad \text{and} \quad \delta_{m,j} = \sum_{l=1} z_{m,l}^2 \sigma_{l,j}^2 \tag{5.15}$$

with j the j -th parameter of the l -th layer and m the m -th element of the input minibatch. And it can then be sampled as:

$$a_{m,j} = \gamma_{m,j} + \sqrt{\delta_{m,j}} \epsilon_{m,j}, \quad \epsilon_{m,j} \sim \mathcal{N}(0, 1) \tag{5.16}$$

The number of operation is then highly reduced compared to a case where the sampling is done for each weight and bias. It also leads to an estimator of lower variance of the gradient than the reparametrisation trick.

Flipout

Flipout, presented in Wen et al. (2018), is a different way to overcome this issue of high variance in gradient estimate, by perturbing the parameters quasi-independently within a batch. Flipout comes back to the perturbations on the weights, written the distribution of the weights as a mean μ to which a perturbation $\delta\theta$, centred around zero, is added. This perturbation is sampled once per batch (until here, this looks a lot like the reparametrisation trick). Then, to have quasi-independent perturbation within a batch, the idea is to multiply the perturbation by a pseudo-random value, sampled in $\{-1, 1\}$, for each example of the batch.

Formally, if the parameter distribution, θ , can be decomposed as a sum of a mean value, μ , and a perturbation, $\delta\theta$, which has a symmetric distribution around zero:

$$\theta = \mu + \delta\theta \tag{5.17}$$

Then the parameter perturbation can be multiplied by a pseudo-random sign matrix (which leaves the perturbation distribution invariant), which creates a different perturbation matrix for each training example:

$$\delta\theta_i = \delta\theta * r_i s_i^\top \tag{5.18}$$

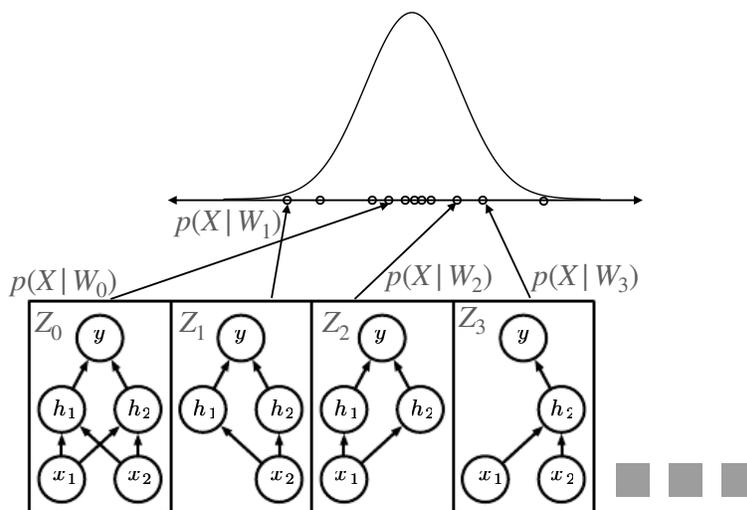


Figure 5.4: Illustration of the sampling process to generate a distribution using MC-Dropout. Each dropout realisation generates an output and the resulting distribution can be interpreted as a Gaussian distribution.

r_i and s_i being random vectors whose entries are sampled uniformly from ± 1 , $\delta\theta$ is sampled only once for the whole batch. Flipout allows for a large decrease of the variance in the stochastic gradient by a factor of $\sim 1/n_{elems}$, n_{elems} the number of elements in a batch. Thanks to matrix multiplication this operation is quick even though it is more computationally costly than methods with shared parameters value in a batch. This is the method that I use in chapter 8 to train a BNN for galaxy parameters estimation.

MC-Dropout

Contrary to previous methods which add a perturbation term on parameters, MC-Dropout uses a different approach: it multiplies them by a mask matrix.

Dropout is initially presented as a regularisation method for regularisation during training (Srivastava et al., 2014) (see section 4.2.3). The idea is to apply multiplicative noise to the output of the previous layer. Most of the time Bernoulli noise is used, but other kinds such as Gaussian noise can also be used.

Instead of using it only during training, Gal and Ghahramani (2015) proposed to use it also during prediction to generate a probability distribution of the output. On fig. 5.4, we show a simplified view of this process. N predictions are being performed with N different dropout configurations, i.e. different neurons are dropped every time. This produces different outputs which distribution can be interpreted as a Gaussian posterior distribution. Indeed, Gal and Ghahramani (2015) showed that using dropout and the l_2 regularisation on the weights is equivalent to minimising the Kullback–Leibler divergence between an approximate distribution and a normal prior.

This method is called Monte-Carlo Dropout or MC-Dropout and allows for an approximation of variational inference.

The approximation of the variational posterior can be parametrised as, for each layer i , with weights matrix W_i of size $K_i \times K_{i-1}$:

$$W_i = \Theta_i \cdot \text{diag}([z_{i,j}]_{j=1}^{K_i}) \quad \text{with } z_{i,j} \sim \text{Bernoulli}(p) \quad (5.19)$$

with $z_{i,j}$ the random activation or inactivation coefficients with associated probability p , Θ_i the weights matrix before dropout and W_i the weights matrix afterward. As mentioned, with this formulation it is possible to approximate the KL divergence, with respect to a normal prior, as a simple l_2 regularisation term. It is then possible to maximise the ELBO and perform variational inference very efficiently.

This method presents the advantages of being quite straightforward to implement, since it only requires dropout layers, and cheap to compute. On the downside, to obtain well-calibrated uncertainty, the dropout rate must be adapted as a variational parameter to the data. Nevertheless, since it works quite well, several works have already used it in cosmology (for example, Walmsley et al., 2020), and we used it in Theobald, Arcelin et al. (2021) to train a BNN for galaxy shape estimation. This paper can also be found on appendix B.

5.3.3 Training BNNs

Previously presented methods work well in theory but training a neural network to predict epistemic uncertainty is often difficult. One point that I did not mention yet is the choice of prior. Indeed, I mainly talk about Gaussian priors previously but the parameters distribution is likely to be non-trivial and there is no guarantee that the Gaussian prior, which is the most commonly used, is relevant. In practice, priors are set empirically, which can make the training quite difficult.

Sticking to Gaussian priors, and assuming mean-field variational inference, a common approximation saying that the distributions of the weights are independent from each other $q_\phi(\theta) = \prod_{i=1}^m q_{\phi_i}(\theta_i)$, might lead the KL divergence term of the loss function to outweigh the likelihood term. This prevent the network from learning, and indicates that the chosen prior distribution might not be the most relevant. This issue can be overcome by several tricks which require to adapt the training procedure: it is for example possible to use *KL-annealing*. This method works in two steps. First, the network is trained cancelling the KL term, i.e. adding a null coefficient in front of the KL term in the loss, keeping only the likelihood. It is equivalent to training a point estimate neural network. Then, the coefficient in front of the KL term is gradually increased from 0 to 1 in order to learn the epistemic uncertainty. This method is quite empirical but proved to work most of the time. This is also a possible avenue to the problem described in section 4.3.1 to train VAEs, and that we solved using a β -VAE (Higgins et al., 2016) in chapter 7.

Another lead is to use other kinds of variational approximations such as restricted variational distributions. The k -tied normal distribution is based on an experimental founding that converged posterior standard deviations under Gaussian mean-field VI consistently display a strong low-rank structure (Swiatkowski et al., 2020). By factorizing the parametrization of posterior standard deviations, it allows for equivalent model performance compared to Gaussian mean-field VI while decreasing the number of trainable parameters and improving signal-to-noise ratio of stochastic gradients, leading to faster convergence. It is used as prior for the BNN presented in chapter 8 for galaxy parameters estimation. In the end, a training procedure can always be established to train a BNN and the methods presented previously work most of the time.

In this chapter we saw the motivations for using BNNs and that it was possible to train them quite efficiently. They can provide a valuable amount of information which is crucial for the utilisation of neural networks in cosmology. However, the training difficulties previously described combined with the struggle to fully grasp the complexity of a neural network model can lead to question the necessity and relevance of using deep learning for cosmology. In the next part I discuss this question.

Bibliography

- Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe (Jan. 2016). “Variational Inference: A Review for Statisticians”. In: *arXiv e-prints*, arXiv:1601.00670, arXiv:1601.00670. arXiv: [1601.00670 \[stat.CO\]](#).
- Blundell, Charles et al. (May 2015). “Weight Uncertainty in Neural Networks”. In: *arXiv e-prints*, arXiv:1505.05424, arXiv:1505.05424. arXiv: [1505.05424 \[stat.ML\]](#).
- Charnock, Tom, Laurence Perreault-Levasseur, and François Lanusse (June 2020). “Bayesian Neural Networks”. In: *arXiv e-prints*, arXiv:2006.01490, arXiv:2006.01490. arXiv: [2006.01490 \[stat.ML\]](#).
- Gal, Yarin and Zoubin Ghahramani (June 2015). “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning”. In: *arXiv e-prints*, arXiv:1506.02142, arXiv:1506.02142. arXiv: [1506.02142 \[stat.ML\]](#).
- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Higgins, Irina et al. (2016). “beta-vae: Learning basic visual concepts with a constrained variational framework”. In:
- Kingma, Diederik P., Tim Salimans, and Max Welling (June 2015). “Variational Dropout and the Local Reparameterization Trick”. In: *arXiv e-prints*, arXiv:1506.02557, arXiv:1506.02557. arXiv: [1506.02557 \[stat.ML\]](#).
- Kingma, Diederik P. and M. Welling (2014). “Auto-Encoding Variational Bayes”. In: *CoRR* abs/1312.6114.
- Koller, Daphne and Nir Friedman (2009). *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press. ISBN: 0262013193.
- Srivastava, Nitish et al. (Jan. 2014). “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *J. Mach. Learn. Res.* 15.1, pp. 1929–1958. ISSN: 1532-4435.
- Swiatkowski, Jakub et al. (2020). “The k-tied Normal Distribution: A Compact Parameterization of Gaussian Mean Field Posteriors in Bayesian Neural Networks”. In: *arXiv preprint arXiv:2002.02655*.
- Theobald, Claire et al. (Apr. 2021). “A Bayesian Convolutional Neural Network for Robust Galaxy Ellipticity Regression”. In: *arXiv e-prints*, arXiv:2104.09970, arXiv:2104.09970. arXiv: [2104.09970 \[cs.LG\]](#).
- Valentin Jospin, Laurent et al. (July 2020). “Hands-on Bayesian Neural Networks – a Tutorial for Deep Learning Users”. In: *arXiv e-prints*, arXiv:2007.06823, arXiv:2007.06823. arXiv: [2007.06823 \[cs.LG\]](#).
- Walmsley, Mike et al. (2020). “Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning”. In: *Monthly Notices of the Royal Astronomical Society* 491.2, pp. 1554–1574.
- Wen, Yeming et al. (Mar. 2018). “Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches”. In: *arXiv e-prints*, arXiv:1803.04386, arXiv:1803.04386. arXiv: [1803.04386 \[cs.LG\]](#).

Chapter 6

Cosmology with deep learning

6.1 The growing use of deep learning for cosmology

In 2020 in the *astro-ph* section of *arxiv*¹, 463 paper were mentioning "neural networks" and/or "deep learning" in their abstract, 109 in the cosmology sub-section. It was 319 in 2019 and 142 in 2018 (respectively 78 and 22 in the cosmology sub-section). There clearly is a growing trend in using this tool for cosmology and astrophysical applications but a fair question to ask would be: is this really justified ? Is deep learning really necessary and can it be used to increase knowledge in cosmology ?

First, let's not forget that neural networks are used for a long time in cosmology, for example Bertin and Arnouts (1996) used a small neural network for galaxy/star classification in their SExtractor algorithm.

Then, Ntampaka et al. (2019) and Kremer et al. (2017), from which this part is partially inspired, present two main reasons for this recent increase in interest: the growing data production from large surveys and the fact that both fields can benefit from each other. Indeed, future cosmological surveys will generate large amounts of data (see section 6.1.2), often requiring fast processing, and cosmology will look for more and more complex features in data, offering new challenging and demanding problems for deep learning.

However, deep learning is often disregarded, especially for science applications, because it is viewed as a "black box", i.e. it is not possible to trace back which feature leads to which prediction. This lack of *interpretability* entails a lack of confidence in the inference produced by a deep learning algorithm. So first, let see if this argument is relevant.

6.1.1 Interpretability of neural networks

This argument is probably the most common regarding neural networks and is absolutely not specific to its application to cosmology. However it has a particular impact in science, as it is mandatory to control and justify every step of the analysis to generate reliable results. Even if the section title asks for a yes or no answer, it is obviously not that simple and I describe below some of the pros and cons arguments.

Nowadays *interpretability* or *explainability* is an open research area in the machine learning community. One could think that since we know how backpropagation works, then we know how a neural network learns, so neural networks are not black boxes. Yet, this argument can obviously not be satisfying: if the model contains several millions of parameters, it is not possible to interpret

¹<https://arxiv.org/>

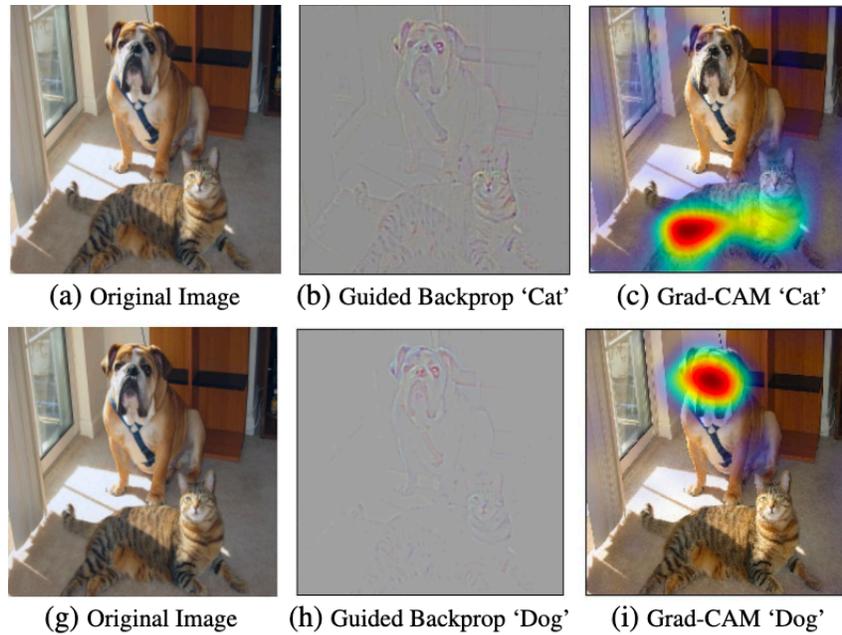


Figure 6.1: (a) and (g): original images from a cat and a dog. (b) and (h): results obtained with Guided Backpropagation (Springenberg et al., 2014), a methods which highlights contributions features. (c) and (i): results obtained with Grad-CAM which localizes class-discriminative regions of the image. Credit: Selvaraju et al., 2016

and explain the behavior of each of them. Studying interpretability is necessary and it is, moreover, not only a problem of social acceptance, it can help prevailing biases and rendering models more accurate.

Several approaches exist to model interpretability and I will not describe them in details here, as I am not familiar with them. Alongside image analysis, I can just mention that methods exist in computer vision to visualise which part of the input image is the most significant for the network to make its predictions via so-called *saliency maps* (Simonyan, Vedaldi, and Zisserman, 2013) (see fig. 6.1 for example, from Buhrmester, Münch, and Arens, 2019 and Selvaraju et al., 2016). Closer to cosmology, Cranmer et al. (2020a) developed a technique using graph neural networks learning a dataset, from which they extract symbolic equation describing the physical laws embedded in the dataset. They provide an analogy between the different part of the graph neural network and Newtonian dynamics offering an other way of interpreting neural networks.

As described in section 5.1, BNNs provide another kind of information about the network, allowing for the quantification of the epistemic uncertainty. The estimation of epistemic uncertainty is a necessary step for the use of neural networks measurements in a science pipeline. It provides information about the confidence level of the network in the prediction and it is perfectly suited to identify outliers and anticipates high predictive errors (see chapter 8, Theobald, Arcelin et al. (2021) and appendix B). From the perspective of interpretability, it is also possible to use saliency maps to probe a BNN and see from where, on a test image, the uncertainty is mostly coming from (Kendall and Gal, 2017 and Murugamoorthy and Khan, 2020 for examples). This can be used to identify images to label and to add to the training sample in order to improve the network’s ability to generalise, a process known as *active learning* (see Prince, 2004).

These examples show that even if much still remains to be done, the “black box” argument is

less and less relevant as the interpretability of neural networks is a more and more explored research subject. There exist methods that can provide meaningful insight on how artificial neural networks work. Particularly and more interestingly for my work, which features of an image are relevant for prediction.

Now let's focus on arguments in favour of the use of deep learning in cosmology, for future surveys.

6.1.2 Large and deep surveys

Several large surveys such as Hyper-Suprime Cam Subaru Strategic Program survey (HSC-SSP²) or Dark Energy Survey (DES³) are already generating a lot of data: HSC is producing around 300 gigabytes (GB) of data per night (Furusawa et al., 2008) and DES can go up to 2.5 terabytes (TB) (*DES Data Management webpage n.d.*).

However these numbers are small compared to what is expected of future surveys. The Rubin Observatory Legacy Survey of Space and Time (LSST) will produce a volume of 20 terabytes (TB) of raw data per night and will collect around 60 petabytes (PB) of data over its ten years of operation⁴. Also, the Euclid survey (Refregier et al., 2010) will produce up to 26 PB per year of operation (de Teodoro, Nieto, and Altieri, 2017), over the 6 years of operation. The Nancy Grace Roman Space Telescope (former WFIRST, Spergel et al., 2013) will produce 20 PB of data over the whole survey⁵ (five years). These large amount are partially due to the quality and nature of the data recorded: large sky images.

Of course, to process these large amounts of data, deep learning, with its ability to perform fast inference, is an appropriate and common tool to turn to. For example, for LSST, a nightly pipeline will look for transients. It will have to send an alert to the community within 60 seconds of image readout. The alert system is supposed to produce around 10 million alerts per night and deep learning is proposed to accelerate parts of the process such as the classification of events (Möller et al., 2020).

Also, deep learning can be used to generate mock data to test processing pipelines or analysing techniques (Bretonnière et al., 2021). Lanusse et al. (2021) use Variational AutoEncoder (VAE) as generative model to simulate galaxy images with more complex shapes that what can be achieved with classical methods.

Nowadays, there is no discussion about the speed up that deep learning can bring to processing or generating large amount of data, providing that the training is done. This is clearly a major point in favour of the usage of deep learning for future cosmological surveys.

6.1.3 Interaction between cosmology and deep learning

The other argument in favour of deep learning is that it can perform better, or at least as good as traditional methods with a shorter time to execute (see Ntampaka et al., 2019 for a detailed list of examples). Regarding galaxy modeling for deblending or image generation for instance, traditional methods used mathematical models and known properties of galaxies. Deep neural networks can create galaxy models based on data, including complex features which cannot be captured in usual parametric models (Regier, McAuliffe, and Prabhat, 2015, Ravanbakhsh et al., 2016, Lanusse et al.,

²<https://hsc.mtk.nao.ac.jp/ssp/survey/>

³<https://www.darkenergysurvey.org/>

⁴<https://www.lsst.org/about/dm>

⁵<https://svs.gsfc.nasa.gov/13667>

2021, Arcelin et al., 2021 or Bretonnière et al., 2021). This allows for model bias reduction and more realistic galaxy image simulations.

Not only neural network methods can lead to improvements in cosmology, but the inverse also applies. For instance, Kremer et al. (2017) present several examples of their work where cosmological data analysis can trigger advancements in machine learning. An example being the use of nearest neighbour approach to estimate weights to apply in an importance-weighting procedure, in order to alleviate selection bias for photometric redshift estimation (Kremer et al., 2015). Cranmer et al. (2020b) is another example of this relationship between cosmology and deep learning. Here, authors present a new type of neural network: *Lagrangian neural networks*. These networks aim at parametrising Lagrangians using neural networks to create symmetric and energy conserving models.

Cosmology create new challenges which can be quite different from the usual test cases of deep learning. It is consequently a field of application that could lead to interesting developments. Difficulties generated by cosmology comes from different sources: there is only one Universe, so only one experiment, we do not have a complete knowledge of the truth, making supervised learning difficult, and we are dealing with unusual, transients, and sometimes shallow, objects that can prove to be very difficult to model.

Those interactions combined with the benefits brought by Bayesian deep learning lead me to think that using these tools for cosmology is extremely promising. In parallel of the progresses made in interpretability of neural networks, works have already shown that BNN can produce promising results for cosmology.

6.2 Cosmology examples

Selecting examples of deep learning use cases for cosmology is inevitably arbitrary: here, I will mention examples related to key points highlighted previously.

As mentioned in section 6.1.2: an alert system is necessary for LSST to produce one alert within 60s of image readout, and deep learning is a great avenue to accelerate parts of the process. The FINK broker⁶ was designed in that purpose and includes deep learning and adaptative learning technologies (Möller et al., 2020). Especially, it uses SuperNNova (Möller and de Boissière, 2020), an algorithm designed as a recurrent neural network (RNN), to perform supernova light-curve classification using photometric information only. They achieved high accuracy performances even for light-curves which do not have reached their maximum peak. In this work they used MC-Dropout and Bayes-by-backprop (see section 5.3.2) to implement a Bayesian RNN. It was then used to estimate uncertainties reflecting the confidence in predictions, exhibiting an expected behaviour of epistemic uncertainties.

Another example of MC-Dropout used to approximate Bayesian posterior is the work of Walsley et al. (2020). They used a Bayesian convolutional neural network to improve classification of galaxies using the Galaxy Zoo dataset. This algorithm obtains reliable results in estimating the probability of a galaxy morphology label (see fig. 6.2). Interestingly they showed that, using active learning, less labelled data would be required than when using classical supervised learning to perform an identical level of performance.

Even if MC-Dropout is the most frequent solution for probabilistic modeling using neural networks, it is also possible to use Flipout or the reparameterisation trick. In their work, Hortúa et al. (2020) compare the performances of these three probabilistic modeling approach for cosmological

⁶<https://fink-broker.org/>

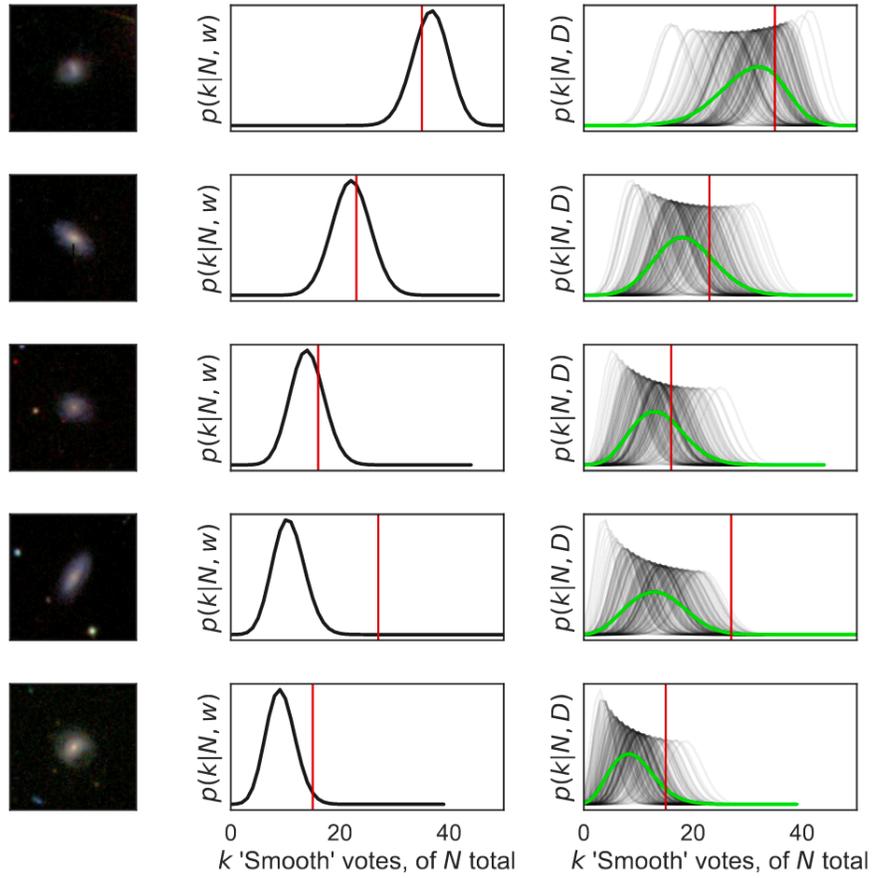


Figure 6.2: Posterior distribution for k of N volunteers answering 'smooth' to the question 'smooth or featured' regarding the galaxy morphology. The left column is the galaxy as input of the network, the center column is the posterior distribution predicted by a single network, and the right column presents the posterior marginalised (averaged) over 30 MC-dropout samples (or 'networks') (in green) as well as each posterior for each 'network' (in grey). The red line is the actual k measured for $N = 40$ volunteers. Credit: Walmsley et al., 2020

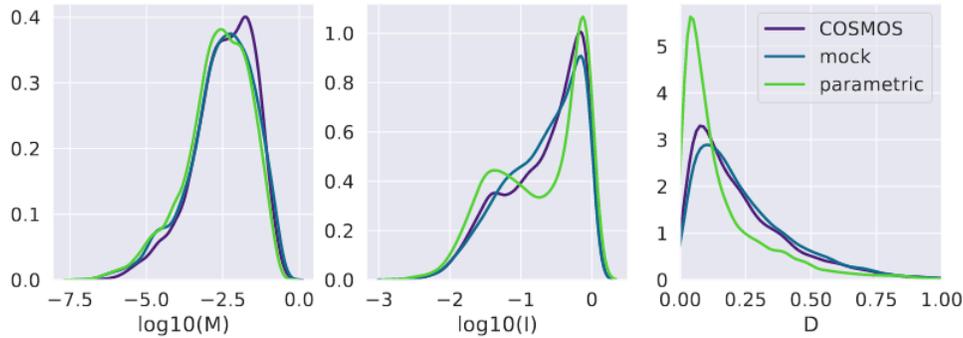


Figure 6.3: Comparison of marginal Multi-mode (M) (left), intensity (I) (center) and deviation (D) (right) statistics evaluated on parametric galaxies, on real COSMOS galaxies and on a sample from the generative model. Credit Lanusse et al., 2021

parameters estimation from Cosmic Microwave Background (CMB) temperature and polarisation maps. They compare Dropout, Dropconnect, Flipout and the reparameterisation trick (RT) finding that Flipout outperforms other methods regardless of network architecture, providing tighter constraints on parameters. In chapter 8, this is what we use to implement a BNN which performs galaxy parameters estimation.

Finally, Lanusse et al., 2021 demonstrate the ability of generative models to generate realistic galaxy image with accurate parameter distribution, outperforming state-of-the-art methods such as parametric fits (see fig. 6.3). The architecture proposed is a *Flow-VAE*: it is based on VAE (see section 4.3.1) and completed with a normalising flow (see section 4.3.3). As mentioned in section 4.3.1, when training a VAE, the matching between posterior and prior distribution is, most of the time not perfect. This is not always problematic, but a requirement for the generation of galaxy images from a latent space is that it must be easily sampled, the latent space distribution must be known. To do so, authors used a conditional normalising flow in order to sample conditionally to relevant parameters, a multivariate Gaussian latent space. Even if this method does not use Bayesian neural network as defined in section 5.2, VAE training is often considered as a Bayesian method since variational inference is used via the reparameterisation trick to learn an approximated posterior distribution over the latent space.

Following these examples, the next chapters will describe the work that I did using Bayesian deep learning to address the issue of blending for weak gravitational lensing studies.

Bibliography

- Arcelin, Bastien et al. (Jan. 2021). “Deblending galaxies with variational autoencoders: A joint multiband, multi-instrument approach”. In: *MNRAS* 500.1, pp. 531–547. DOI: [10.1093/mnras/staa3062](https://doi.org/10.1093/mnras/staa3062). arXiv: [2005.12039](https://arxiv.org/abs/2005.12039) [[astro-ph.IM](#)].
- Bertin, Emmanuel and Stephane Arnouts (1996). “SExtractor: Software for source extraction”. In: *Astronomy and astrophysics supplement series* 117.2, pp. 393–404.
- Brettonnière, H. et al. (May 2021). “Euclid preparation: XVI. Forecasts for galaxy morphology with the Euclid Survey using Deep Generative Models”. In: *arXiv e-prints*, arXiv:2105.12149, arXiv:2105.12149. arXiv: [2105.12149](https://arxiv.org/abs/2105.12149) [[astro-ph.GA](#)].
- Buhrmester, Vanessa, David Münch, and Michael Arens (Nov. 2019). “Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey”. In: *arXiv e-prints*, arXiv:1911.12116, arXiv:1911.12116. arXiv: [1911.12116](https://arxiv.org/abs/1911.12116) [[cs.AI](#)].
- Cranmer, Miles et al. (June 2020a). “Discovering Symbolic Models from Deep Learning with Inductive Biases”. In: *arXiv e-prints*, arXiv:2006.11287, arXiv:2006.11287. arXiv: [2006.11287](https://arxiv.org/abs/2006.11287) [[cs.LG](#)].
- Cranmer, Miles et al. (Mar. 2020b). “Lagrangian Neural Networks”. In: *arXiv e-prints*, arXiv:2003.04630, arXiv:2003.04630. arXiv: [2003.04630](https://arxiv.org/abs/2003.04630) [[cs.LG](#)].
- de Teodoro, P., S. Nieto, and B. Altieri (June 2017). “Data Management in the Euclid Science Archive System”. In: *Astroinformatics*. Ed. by Massimo Brescia et al. Vol. 325, pp. 385–388. DOI: [10.1017/S1743921316012874](https://doi.org/10.1017/S1743921316012874).
- DES Data Management webpage* (n.d.). <https://www.darkenergysurvey.org/the-des-project/survey-and-operations/data-management/>. Accessed: 2020-12-24.
- Furusawa, Hisanori et al. (2008). “Hyper Suprime-Cam: data analysis and management system”. In: *Observatory Operations: Strategies, Processes, and Systems II*. Vol. 7016. International Society for Optics and Photonics, 70161F.
- Hortúa, Héctor J. et al. (Nov. 2020). “Parameter estimation for the cosmic microwave background with Bayesian neural networks”. In: *PRD* 102.10, 103509, p. 103509. DOI: [10.1103/PhysRevD.102.103509](https://doi.org/10.1103/PhysRevD.102.103509). arXiv: [1911.08508](https://arxiv.org/abs/1911.08508) [[astro-ph.IM](#)].
- Kendall, Alex and Yarin Gal (Mar. 2017). “What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?” In: *arXiv e-prints*, arXiv:1703.04977, arXiv:1703.04977. arXiv: [1703.04977](https://arxiv.org/abs/1703.04977) [[cs.CV](#)].
- Kremer, Jan et al. (2015). “Nearest neighbor density ratio estimation for large-scale applications in astronomy”. In: *Astronomy and Computing* 12, pp. 67–72.
- Kremer, Jan et al. (2017). “Big universe, big data: machine learning and image analysis for astronomy”. In: *IEEE Intelligent Systems* 32.2, pp. 16–22.
- Lanusse, François et al. (July 2021). “Deep generative models for galaxy image simulations”. In: *MNRAS* 504.4, pp. 5543–5555. DOI: [10.1093/mnras/stab1214](https://doi.org/10.1093/mnras/stab1214). arXiv: [2008.03833](https://arxiv.org/abs/2008.03833) [[astro-ph.IM](#)].

- Möller, Anaïs and T. de Boissière (Jan. 2020). “SuperNNova: an open-source framework for Bayesian, neural network-based supernova classification”. In: *MNRAS* 491.3, pp. 4277–4293. DOI: [10.1093/mnras/stz3312](https://doi.org/10.1093/mnras/stz3312). arXiv: [1901.06384](https://arxiv.org/abs/1901.06384) [[astro-ph.IM](#)].
- Möller, Anaïs et al. (Nov. 2020). “FINK, a new generation of broker for the LSST community”. In: *MNRAS*. DOI: [10.1093/mnras/staa3602](https://doi.org/10.1093/mnras/staa3602). arXiv: [2009.10185](https://arxiv.org/abs/2009.10185) [[astro-ph.IM](#)].
- Murugamoorthy, Gayathiri and Naimul Khan (Oct. 2020). “Interpreting Uncertainty in Model Predictions For COVID-19 Diagnosis”. In: *arXiv e-prints*, arXiv:2010.13271, arXiv:2010.13271. arXiv: [2010.13271](https://arxiv.org/abs/2010.13271) [[eess.IV](#)].
- Ntampaka, Michelle et al. (2019). “The role of machine learning in the next decade of cosmology”. In: *arXiv preprint arXiv:1902.10159*.
- Prince, Michael (2004). “Does active learning work? A review of the research”. In: *Journal of engineering education* 93.3, pp. 223–231.
- Ravanbakhsh, Siamak et al. (Sept. 2016). “Enabling Dark Energy Science with Deep Generative Models of Galaxy Images”. In: *arXiv e-prints*, arXiv:1609.05796, arXiv:1609.05796. arXiv: [1609.05796](https://arxiv.org/abs/1609.05796) [[astro-ph.IM](#)].
- Refregier, A. et al. (Jan. 2010). “Euclid Imaging Consortium Science Book”. In: *arXiv e-prints*, arXiv:1001.0061, arXiv:1001.0061. arXiv: [1001.0061](https://arxiv.org/abs/1001.0061) [[astro-ph.IM](#)].
- Regier, Jeffrey, J. McAuliffe, and Prabhat (2015). “A deep generative model for astronomical images of galaxies”. In:
- Selvaraju, Ramprasaath R. et al. (Oct. 2016). “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization”. In: *arXiv e-prints*, arXiv:1610.02391, arXiv:1610.02391. arXiv: [1610.02391](https://arxiv.org/abs/1610.02391) [[cs.CV](#)].
- Simonyan, Karen, Andrea Vedaldi, and Andrew Zisserman (2013). “Deep inside convolutional networks: Visualising image classification models and saliency maps”. In: *arXiv preprint arXiv:1312.6034*.
- Spergel, D. et al. (May 2013). “Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA Final Report”. In: *arXiv e-prints*, arXiv:1305.5422, arXiv:1305.5422. arXiv: [1305.5422](https://arxiv.org/abs/1305.5422) [[astro-ph.IM](#)].
- Springenberg, Jost Tobias et al. (Dec. 2014). “Striving for Simplicity: The All Convolutional Net”. In: *arXiv e-prints*, arXiv:1412.6806, arXiv:1412.6806. arXiv: [1412.6806](https://arxiv.org/abs/1412.6806) [[cs.LG](#)].
- Theobald, Claire et al. (Apr. 2021). “A Bayesian Convolutional Neural Network for Robust Galaxy Ellipticity Regression”. In: *arXiv e-prints*, arXiv:2104.09970, arXiv:2104.09970. arXiv: [2104.09970](https://arxiv.org/abs/2104.09970) [[cs.LG](#)].
- Walmsley, Mike et al. (2020). “Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning”. In: *Monthly Notices of the Royal Astronomical Society* 491.2, pp. 1554–1574.

Part III

Bayesian deep learning applied to blended galaxies

In this part, I present the work that I did during the three years of this thesis. The first chapter focuses on designing a deblending algorithm using Variational AutoEncoders (VAE) (chapter 7). In the second chapter, I propose a solution for shape and redshift parameter estimation from images of galaxies, whether they are blended or not (chapter 8).

Chapter 7

Variational AutoEncoders for deblending

This chapter describes the work that I did during the first part of my thesis. It led to a publication in the *MNRAS* (Arcelin et al., 2021). This chapter presents the main ideas from this paper (until section 7.6) and some further improvements and developments (from section 7.7).

7.1 Introduction

Main motivations for this work have already been presented through chapter 3 and chapter 6. To summarize, blending is a major challenge for current and especially future galaxy surveys. Existing deblending solutions all rely on assumptions for the shape or light profiles of objects, possibly leading to model bias. Also even though some approaches are multi-band, none of them provides the ability to combine data from several instruments. Variational AutoEncoders (VAEs) (section 4.3.1) based on CNNs provide both the ability to learn complex galaxy shapes from the data and to use multi-channels images. The latter naturally allows for the use of multi-band images from several instruments.

The method is divided in two steps:

1. First, we train a VAE to learn the features of isolated galaxy images and their representation in a latent space. The aim of this network is to build a generative model for isolated galaxy images directly from the (here simulated) data. The images are generated with LSST and Euclid bandpass filters with fixed PSF (see section 7.2.1).
2. Second, we use this trained VAE to create a second neural network, with a similar architecture, in charge of performing the deblending task. The generative component of the VAE is used with fixed parameters (i.e. non-trainable) and the recognition model is retrained, though with noisy images of blended galaxies as input. This method allows us to use the generative model previously built as a prior for isolated galaxies images and consequently to perform deblending. The artificially blended images are also generated with LSST and Euclid bandpass filters.

We evaluate the performance of the method in terms of errors introduced on intrinsic ellipticities (deconvolved from the PSF) and fluxes (or magnitudes) per band, which are relevant parameters for weak lensing studies. We first perform a multi-band analysis using the six LSST bandpass filters jointly. Then, based on the idea initially developed by Schuhmann, Heymans, and Zuntz (2019), we combine the LSST and Euclid filters in ten-band images to quantify the extent to which LSST can benefit from Euclid in a joint-pixel analysis, for the sake of deblending galaxies.

The analysis presented in this chapter is realised on images where, in the fiducial case, the target galaxy is perfectly centred on isolated galaxy images as well as on blended images. However this ideal case is very unlikely in reality as noise, pixelisation or geometric properties of the galaxy (e.g. asymmetry) and its neighbours, for example, impact the target galaxy centring even in the case of an extremely accurate detection algorithm. Consequently, we also estimate the impact of decentring with two other cases: the first one supposing a perfect detection algorithm and taking into account only noise and pixelisation, and the second one using a basic detection algorithm to centre our stamps, as could be performed on real data (see fig. 7.2).

Note that only galaxy-galaxy separation is studied in this paper as star-galaxy separation is considered in the literature (Melchior et al., 2018, SCARLET) to be less difficult (shapes and light profiles of stars being arguably simpler). Nevertheless, as our method focuses on recovering the centred galaxy and removing all other components, we expect it should generalise to star-galaxy separation, provided stars are included in the training sample.

7.2 Simulated images

7.2.1 Datasets

This work is based on simulated images generated with `GalSim`¹ (Rowe et al., 2015). They are based on parametric models fitted to real galaxies observed by Hubble Space Telescope (HST) in the COSMOS field. The observed images and the corresponding models are recorded in the HST COSMOS catalog containing 81 500 images with $F814W < 25.2$ (Mandelbaum et al., 2012). The fits were done for the third Gravitational Lensing Accuracy Testing (GREAT3) Challenge (Mandelbaum et al., 2014, see Appendix E.2). Each COSMOS galaxy is fitted twice – once with a Sérsic profile (with index, n , free), and once with a de Vaucouleurs ($n = 4$) bulge profile plus an exponential ($n = 1$) disk profile – and the best of the two fits is kept. In addition to central position and ellipticities, the free parameters for Sérsic profiles are effective radius R_{eff} , intensity $I_{1/2}$ and index n . `GalSim` can generate multi-band images with PSF models and pixel noise chosen to simulate LSST-like and Euclid-like images, based on parametric profiles fitted to observations.

Two sets of images have been generated for the training phase: a training (100 000 images) and a validation sample (20 000 images). A third independent test sample (10 000 images) has been generated with 5000 COSMOS galaxies that were not used for the training and validation samples. The performance of our networks is measured on this test sample. For each sample, we simulate both noisy and noiseless images, the former being used as inputs and the latter as targets (i.e. images to reproduce, used for computation in the loss function) of the neural networks.

Point-Spread-Function and noise

For LSST bandpass filters, we use a fixed Kolmogorov point-spread function (PSF), implemented in `GalSim`, using a full width at half maximum (FWHM) of $0.65''$, which is the median value of the expected FWHM distribution, as shown in Ivezić et al. (2008). This is an approximation as our LSST images are a stack of a large number of exposures in each filter. In practice, the PSF is varying, due to changing weather conditions on each frame, and potentially anisotropic. Therefore, the effective PSF on a stacked image would not exactly be the median of the distribution and would vary from one stacked image to another. We reserve this question for future work but note that, for practical applications, two options exist: images could be deconvolved from the measured PSF

¹<https://github.com/GalSim-developers/GalSim>

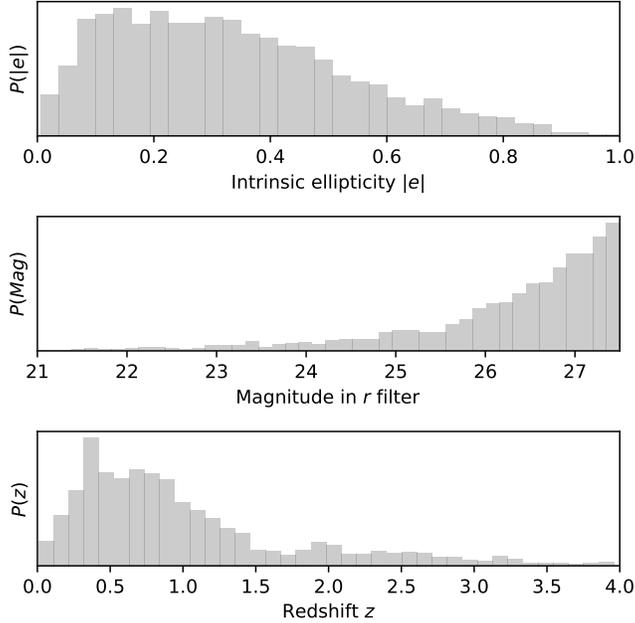


Figure 7.1: Distributions of observed ellipticity, r -band magnitude (a cut is applied at 27.5) and redshift for galaxies in the test sample.

and reconvolved with the fixed PSF we use, before being passed into the deblender network or the training sample could be augmented with multiple realisations of the PSF for each scene.

For Euclid instruments, the PSF is much more stable as it is a space-based experiment and it can be considered fixed to a very good approximation. The PSF is applied to the filters using the Moffat model from `GalSim` with an FWHM of, respectively, $0.22''$ and $0.18''$ (Refregier et al., 2010) for the NISP and VIS instruments.

For each of the ten bandpass filters, Poisson noise is added, taking into account the background sky level² (appropriately scaled to account for exposure time and instrument properties). The expected value of the Poisson distribution is then subtracted from the image in order to centre the noise distribution around 0. We note that noise in the actual Euclid and LSST stacks may significantly be correlated from pixel to pixel due to image resampling, an effect that we do not consider in this work.

Training set 1: isolated galaxies

Using the `galsim.COSMOSCatalog` class, we generate images of isolated galaxies from the parametric models that have been fitted to the COSMOS catalogue. During image generation, the dataset of 81 500 galaxies is randomised by applying rotations before rendering and drawing Poisson noise realisations. In addition, during training and validation runs, we randomly flip images horizontally and vertically and rotate them by 90° on the fly since these operations require no resampling. The spectral energy distributions (SED), profiles and shapes of each galaxy are directly extracted from the catalogue (see fig. 7.1).

Images are generated for the six LSST bandpass filters (u , g , r , i , z and y) as defined in `GalSim` and for the four Euclid bandpass filters³ (three for the Near Infrared Spectrometer and Photometer,

²Data is available at <https://smtn-002.lsst.io/> for LSST.

³Available at <http://svo2.cab.inta-csic.es/svo/theory/fps/index.php?mode=browse&gname=Euclid>.

NISP, one for the visible instrument, VIS). Note that since we use parametric profiles, channels only differ by flux and noise realisations, i.e. we do not simulate colour gradients in this work although it might help improving the deblending performance. Image size is fixed to 64×64 pixels for each instrument which corresponds to, respectively, $19.2''$, $12.8''$ and $6.4''$ -wide images for the NIR filters, for the LSST filters and for the VIS instrument, given their resolutions of 0.3, 0.2 and 0.1 arcsecond per pixel. This fixed stamp size was chosen to work with simple network architectures, although in future work, several stamp sizes might be used for each instrument.

For each galaxy, we consider a stack of four 450 s exposures for Euclid (see *Images Simulations* in Refregier et al., 2010), and a stack of 824 15-second exposures for LSST, split unevenly between bands, though constant between samples, to match the projected full LSST survey mean exposure (specifically, we use respectively 56, 80, 184, 184, 160 and 160 exposures for the *ugrizy* bands as in Ivezić et al., 2008, table 1). The flux of each galaxy is rescaled from HST COSMOS observations to account for exposure time, number of exposures, pixel size and primary mirror area of each instrument. Finally, we apply a cut in the LSST *r*-band magnitude of 27.5, the LSST fiducial depth⁴.

Training set 2: blended galaxies

The deblender network is trained and validated on samples of artificially blended galaxies obtained by adding images of isolated galaxies, produced as described above. Thus we neglect galaxy opacity, as done in most of the deblending literature at this point, which could impact our method by decreasing the amount of color information usable for the networks. This effect will have to be taken into account in future studies. The number of galaxies on each image is randomly chosen between 1 and 4 with uniform probabilities. From these galaxies, the one with the smallest magnitude is centred in the blended image and is the network’s target for reconstruction (see next section). As for images of single galaxies, a cut of $r < 27.5$ is applied to all galaxies added to the blend. Non-centred galaxies are randomly positioned in an annulus around the centred galaxy, with inner radius equal to half of the expected median LSST PSF full-width-at-half-maximum (that is, $0.325''$) and outer radius equal to $2''$. This choice allows us to obtain highly blended galaxies while avoiding cases where non-centred galaxies completely overlap the centred one.

Note that for these samples, signal-to-noise ratio (SNR, see section 7.4.1) and magnitude distributions are slightly shifted to higher SNR/smaller magnitudes since we always choose the smallest magnitude galaxies as targets, but the ranges of SNR and magnitudes remain unchanged. This might impact the performance of our method at low SNR since the statistics is reduced in this range. Results presented here might be improved by increasing the number of examples at low SNR, this analysis is reserved for future work.

Finally, as for isolated galaxy images generation, we apply randomisation to the dataset before rendering, and data augmentation (random flips and 90° rotations) during training.

Decentring

In the fiducial case, considered as reference, target galaxies are perfectly centred on the stamp (at the intersection of the four central pixels). In order to evaluate the impact of decentring, we created two additional configurations (see fig. 7.2), which are defined as follows.

1. The first one aims at simulating a perfectly accurate peak detection algorithm where decentring is only due to pixelisation. It considers that target galaxies are identified as the ones

⁴See, for instance, <https://www.lsst.org/scientists/keynumbers>

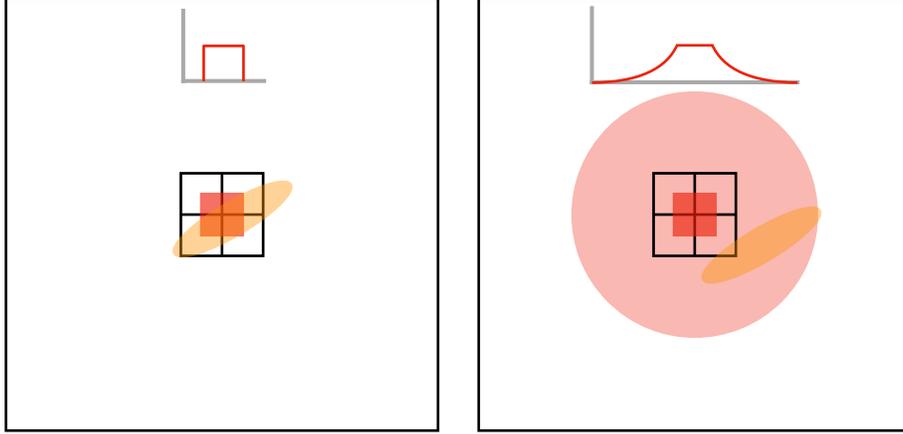


Figure 7.2: Two configurations of decentring to simulate isolated galaxy images. In black are represented the four central pixels of the image. On the left, configuration 1 simulates a perfect detection and accounts only for decentring due to pixelisation. The central galaxy is shifted in a square of size one pixels around the center of the image (in red). On the right, configuration 2 simulates the decentring due to the pixelisation as well as the error coming from the detection algorithm. In that case the galaxy is shifted twice, once in the square of size one pixel around the image center, and a second time with a value sampled from a distribution fitting centring errors in the blending sample. This gives a circular distribution around the center of the image, decreasing exponentially from the border of the square used to represent pixelisation errors. On the top are drawn a simplified view of the decentring distributions.

with smallest magnitudes. Their centres are shifted uniformly around the stamp centre within a square of size $0.2''$, i.e. the size of LSST’s pixels. The shift is applied to isolated galaxies images as well as for blended galaxies images.

2. In the second configuration, we aim at simulating a more realistic case, taking into account decentring from pixelisation and coming from errors of a detection algorithm. We start by producing blended images as in the first configuration, albeit with 128×128 pixels, and subsequently process them with the peak detection algorithm from the `photutils`⁵ library. We identify the galaxy to be deblended as the galaxy with the brightest detected centroid in the r -bandpass filter. We then determine the closest pixel intersection and crop 64×64 images centred on this point. For isolated galaxy images, the galaxies are shifted from the stamp centre a first time, as in the first configuration, for pixelisation. They are then shifted a second time with a value sampled from a distribution fitting centring errors in the blended sample (in practice, a beta prime distribution Johnson, Kotz, and Balakrishnan, 1995), before images are finally drawn. This case is meant to mimic more closely a detection-deblending pipeline that could be applied to real data where we do not know *a priori* which galaxy is the brightest within a blended scene. It is fairly conservative in terms of decentring as we used a basic peak detection algorithm on the r bandpass filter only. Note, however, that this procedure slightly redefines the objective of the deblender network as we are now targeting the galaxy with the brightest detected centroid which is different from the one with the smallest magnitude in about 20% of cases in our samples. We discuss these points in greater detail in section 7.5.1.

⁵<https://photutils.readthedocs.io/en/stable/>

7.2.2 Blendedness metrics

To characterise how much galaxies are blended we defined several metrics, all of which are computed in the LSST r -band (on which we will perform ellipticity and magnitude measurements) on PSF-convolved images.

1. First the blending rate of the centred galaxy with its closest neighbour is defined as

$$B_c = \frac{\langle I_{\text{centred}}, I_{\text{closest}} \rangle}{\sqrt{\langle I_{\text{centred}}, I_{\text{centred}} \rangle \langle I_{\text{closest}}, I_{\text{closest}} \rangle}} \quad (7.1)$$

where the dot product is defined as

$$\langle I, I' \rangle = \sum_{p \in \{\text{pixels}\}} I_p I'_p. \quad (7.2)$$

I_{centred} represents the noiseless image of the centred galaxy and I_{closest} represents the noiseless image of the closest (centre-wise) blended galaxy.

2. Second, we use the *total* blending rate as defined in section 3.1.2:

$$B_{\text{tot}} = 1 - \frac{\langle I_{\text{centred}}, I_{\text{centred}} \rangle}{\langle I_{\text{centred}}, I_{\text{total}} \rangle}. \quad (7.3)$$

These quantities characterise the blendedness of the centred galaxy within a blended scene and have values between 0 and 1. A blend rate close to 0 indicates almost no blending and a blend rate close to 1 indicates that the (closest) neighbouring galaxies and the centred one overlap almost completely. The shifts applied to neighbouring galaxies (see section 7.2.1) was purposefully chosen to obtain a significant amount of highly blended scenes (under those metrics). Note that the hypotheses for images generation are not meant to reproduce the distribution of real blends but rather to allow to train and test our network on relevant blended scenes, albeit in a simplified framework (fixed PSF, simulated parametric images). In section 7.8, we apply this method on a much more realistic distribution of blends, using images extracted from a simulation produced in the Data Challenge 2 (DC2) for LSST Dark Energy Science Collaboration, (Korytov et al., 2019; LSST Dark Energy Science Collaboration (LSST DESC) et al., 2021). This simulation aims at producing images as close as possible from what the future LSST data will look like, especially using realistic galaxy distribution.

7.2.3 Image preprocessing

The networks are fed with batches of images randomly chosen from the training sample, that are also randomly flipped and rotated by 90° on the fly, as explained previously. Both noisy and noiseless images were preprocessed by applying, in each band b , the following normalisation,

$$x_b = \tanh\left(\sinh^{-1}\left(\beta \frac{x_{\text{raw},b}}{\langle \max(x_{\text{raw},b}) \rangle_b}\right)\right), \quad (7.4)$$

where $\langle \max(x_{\text{raw},b}) \rangle_b$ is the mean of the distribution of maximum pixel values in the b band of input images (therefore a constant defined for each bin). β is an arbitrary constant set to 2.5 for our study allowing our training sample to have maxima well-distributed in the range $[0,1]$. This processing is necessary because of the large dynamic range of astronomical images, to ensure that images of bright galaxies do not cause numerical instability during training. Even though this normalisation constrains the choice of likelihood distribution (see next section), we have experimented with non-normalised images and more general distributions and ran into numerical issues, i.e. the loss function diverged.

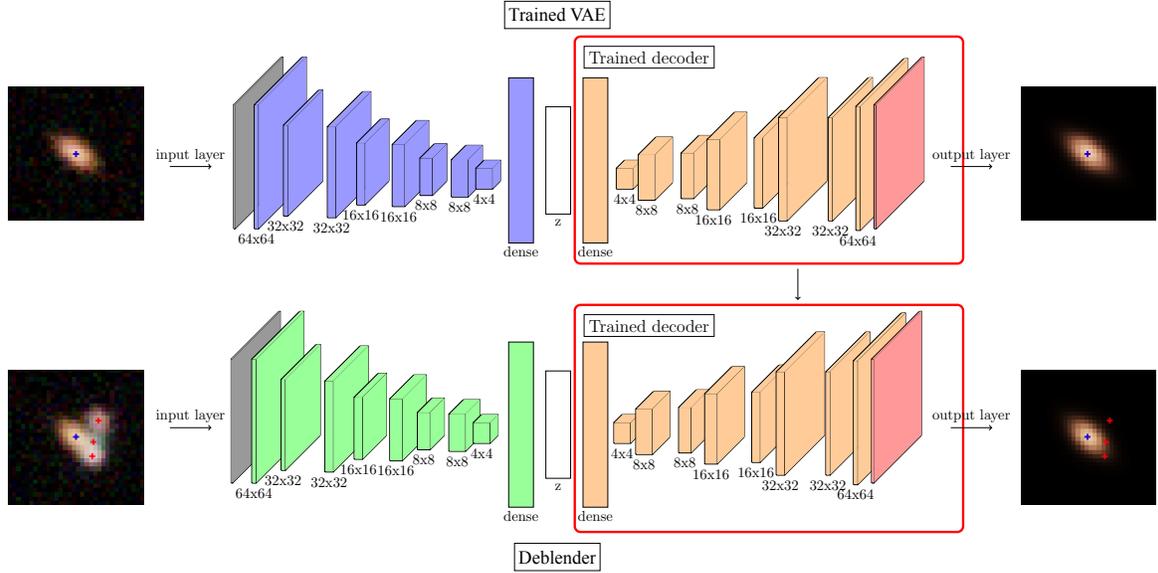


Figure 7.3: Architectures of the VAE and the deblender. The weights of the VAE’s trained decoder are loaded and fixed in the deblender network before training.

7.3 Method

7.3.1 Variational autoencoders

For a detailed explanation of VAEs (Kingma and Welling, 2014), see section 4.3.1. To summarize, a VAE is divided into a latent space, z , a decoder, or generative model, with conditional distribution $p_\theta(x|z)$ (the likelihood), and an encoder, or likelihood model, which approximates the intractable posterior $p_\theta(z|x)$ by another distribution $q_\phi(z|x)$ (see fig. 7.3).

In this work, the latent variables are parametrised with a known prior $p(z)$ chosen to be a standard multivariate gaussian. The generative model and the encoder are parametrised by CNNs, respectively with weights θ and ϕ .

Recall that the VAE maximises the evidence of the training sample $p_\theta(x)$, which cannot be computed exactly, but can be bounded from below,

$$\log p(x) \geq -D_{\text{KL}}(q_\phi(z|x)||p(z)) + \mathbb{E}_{q_\phi}(\log p_\theta(x|z)), \quad (7.5)$$

which defines the variational loss function of the VAE.

In our case we use a denoising VAE. It only differs from usual VAE by the fact that the input \tilde{x} is a noisy version of the target image x . In that case, \tilde{x} is used in the regularisation term (as the encoder input) and x is used in the reconstruction, but the variational loss does not change otherwise (Im et al., 2017).

7.3.2 Implementation and architecture

In our implementation, the approximate posterior $q_\phi(z|x)$ is modelled by a product of univariate Gaussians with mean $\mu_\phi(x)$ and variance $\sigma_\phi^2(x)$, computed by the encoder (with trainable weights ϕ), so that the KL divergence can be computed analytically. The decoder then maps latent variables z onto the expectation values of independent Bernoulli distributions (in each pixel and for each band) which is interpreted as the output image to be compared to the target. This is technically an

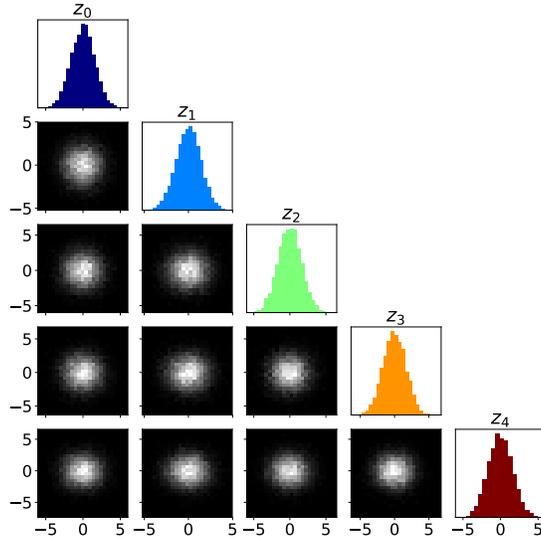


Figure 7.4: Corner plot of the five first dimensions of the latent space of a VAE trained on images composed of the LSST *ugrizy* filters. All components show similar trends.

abuse as the Bernoulli distribution is formally defined for binary variables rather than continuous variables in the range $[0,1]$ where our rescaled images live (see section 7.2.3) and it formally breaks the variational objective in eq. (7.5). However, this is an empirical choice that is commonly found in the literature, including highly cited papers (Kingma and Welling, 2014; Larsen et al., 2016; Sønderby et al., 2016; Jiang et al., 2017; Dilokthanakul et al., 2016). The reader can take a look at Loaiza-Ganem and Cunningham (2019) for a detailed study of the consequences of this implementation and section 7.5 for a discussion about possible improvements.

In order to optimise the reconstruction performance, we used a β -VAE (Higgins et al., 2017), i.e. we decrease the weight of the Kullback-Leiber divergence in the loss by scaling it with a coefficient β . Without this modification, the regularisation term prevails over the reconstruction term and prevents the network to learn to reproduce the input, a longstanding issue with vanilla VAEs. In principle, this further entails the loss of the generative property of our model as this modification relaxes the regularisation on latent variables. However, we found that a value $\beta = 10^{-2}$ yields acceptable reconstruction quality while, given the flexibility of the model, allowing latent variables to have a distribution very close to the Gaussian prior as shown on fig. 7.4 (that could be further improved with regaussianisation techniques, see section 4.4).

Our implementation is based on the Keras library⁶ and the TensorFlow framework⁷ (version 1.13.1) and uses a sequential network architecture presented in fig. 7.3 with four sets of two convolutional layers in the encoder and symmetrically in the decoder (with transposed convolutional layer). In each set, both layers have the same number of filters but they have strides of, respectively (1,1) and (2,2), which allows us not to use pooling layers and thus reduces information loss, while downgrading the image (Springenberg et al., 2014). The convolutional layers have respectively 32, 64, 128 and 256 filters and all use 3x3 kernels. We added hidden dense layers of 256 units in both the encoder and the decoder. All these layers use parametric rectified linear unit (PReLU) activations. Finally, the encoder ends with two parallel dense layers ($\mu_\phi(x)$ and $\sigma_\phi^2(x)$) with 32

⁶<https://keras.io/>

⁷<https://www.tensorflow.org/>

units, corresponding to the dimension of the latent space. In total, the encoder and the decoder have respectively 2.5 million and 3.3 million trainable parameters for the models trained on LSST *ugrizy* bands and on all LSST and Euclid bands.

To create the multi-band images, filters are simply combined in a three dimensional array without further processing, making the images multi-resolution. As the multi-band images processed by the network are multi-resolution (they are composed of the six LSST filters and four Euclid filters), the information from each band is extracted by a particular set of convolutional filters for each convolutional layer. It is then concatenated when the last convolutional layer is flattened, just before the hidden dense layer.

7.3.3 Deblender

The architecture of the deblender network is chosen to be the same as that of the VAE. Before training, we will load decoder weights θ from a pretrained VAE and hold them fixed during training, i.e. only the encoder weights ϕ' are trainable. The encoder learns a mapping from blended images to the posterior parameters $\mu_{\phi'}$ and $\sigma_{\phi'}$ used to draw latent variables. Those are then used as input to the fixed decoder, which outputs the likelihood parameters, needed to evaluate the reconstruction term in the loss, together with the target noiseless image of the isolated central galaxy. Therefore the encoder learns to separate the central galaxy from every other part of the input image (noise and neighbouring galaxies) and to map it into the latent space. The encoding must be done in a way that permits the decoder to reproduce the centred galaxy accurately. As a consequence, our method can be considered as operating deblending in latent space with a mapping created from a prior generative model of isolated galaxies.

For the deblender, we also keep the Kullback-Leiber divergence and the reconstruction term in the loss, which reads

$$L(\phi', x_{\text{in}}, x_{\text{target}}) = \beta D_{\text{KL}}(q_{\phi'}(z|x_{\text{in}})||p(z)) - \mathbb{E}_{q_{\phi'}}[\log p_{\theta}(x_{\text{target}}|z)], \quad (7.6)$$

but here we minimise over the mapper weights ϕ' . Here x_{in} is the noisy input and blended image and x_{target} is the noiseless target image of the central galaxy. Note that we tried to initialise those weights either randomly or from the trained encoder and obtained the same results.

7.4 Results

7.4.1 Reconstruction metrics

Since weak lensing is our main focus, we assess the performance of our method by measuring how well shape parameters and magnitudes can be recovered by the deblender (see definitions below). We first measure ellipticities and magnitudes of the noiseless, denormalised, target images in the test sample, described in section 7.2.1. Then, when training is finished, we run the VAE or the deblender on the test sample and repeat those measurements on the denormalised output of the network. Finally, we compute errors on ellipticities and magnitudes between the target and output images and use them to compute performance metrics. We present the distribution of the measured errors as functions of other relevant quantities, such as signal-to-noise ratio, in order to gain intuition about the assets and limitations of our method. More precisely, we measure:

1. the PSF-corrected ellipticity defined as the reduced shear estimator $|e| \equiv (a-b)/(a+b)$, where a and b are the semi-major and semi-minor radii. The measurement is performed in the LSST r -band from PSF-convolved images using the HSM module in `GalSim` and the `EstimateShear()`

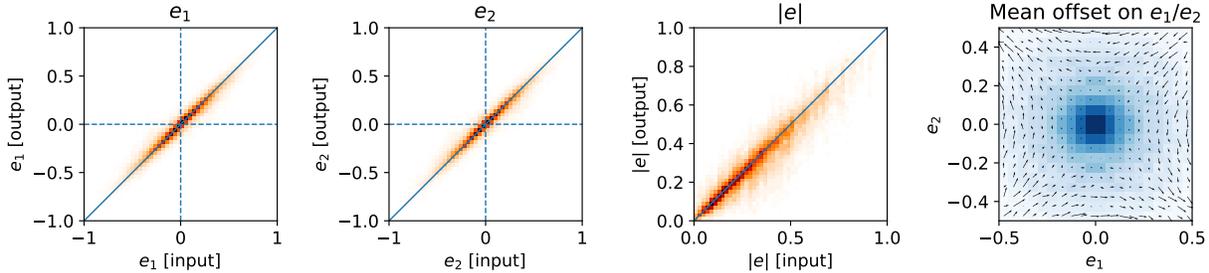


Figure 7.5: Distributions of output shape parameters as a function of parameters measured on target images (e_1 , e_2 and $|e|$ from the leftmost panel to the third one). The rightmost panel shows the mean offset on e_1 and e_2 in the (e_1, e_2) plane with arrows pointing to the average output values of these components (with the same scale as the axes).

function with the Kaiser-Squires-Broadhurst method (KSB, Kaiser, Squires, and Broadhurst, 1995), to which we provide the fixed PSF used to generate the images.

2. the magnitude in the r -band, computed from the total flux, itself obtained by simply summing the number of photons from every pixel. We have verified that we obtain very similar results in all bands.

In order to demonstrate and quantify the benefits of using multiple bands and multiple instruments, we repeat the training of both networks and the analysis for images consisting of

1. the six LSST bandpass filters (*ugrizy*, 6 bands), and
2. all LSST and Euclid bandpass filters together (10 bands).

Finally, we also evaluate the robustness of our networks to signal-to-noise ratio (SNR) as it is expected to have a significant impact on the performance. We define the SNR as

$$S/N = \sqrt{\sum_{p \in \text{pixels}} \frac{I_p^2}{\sigma^2(I_p)}}, \quad (7.7)$$

where I_p is the intensity of the signal, i.e. the flux from the noiseless galaxy image, in pixel p in the r -band and the variance $\sigma^2(I_p)$ is the sum of the flux from the noiseless galaxy and from the sky background. This ratio is measured on isolated galaxy images and this definition is therefore valid for individual objects. Nevertheless, we use the same values when objects are blended for our shape and magnitude reconstruction tests, which over-evaluates the signal-to-noise ratios that could be reached. Therefore, figures concerning deblending performance involving SNR bins present conservative results.

7.4.2 Prior for single galaxies with a VAE

The first network, the *prior model*, is intended to learn a generative model of multi-band images of isolated galaxies. As a probabilistic model, the output is conditioned on a random draw in latent space, therefore we do not compare measurements on individual galaxies but rather the statistical properties of errors on the test sample.

In fig. 7.5, we first show errors for both ellipticity components (e_1, e_2) in the prior model, finding no particular difference in reconstruction between the ellipticities along the pixel grid (e_1) and at

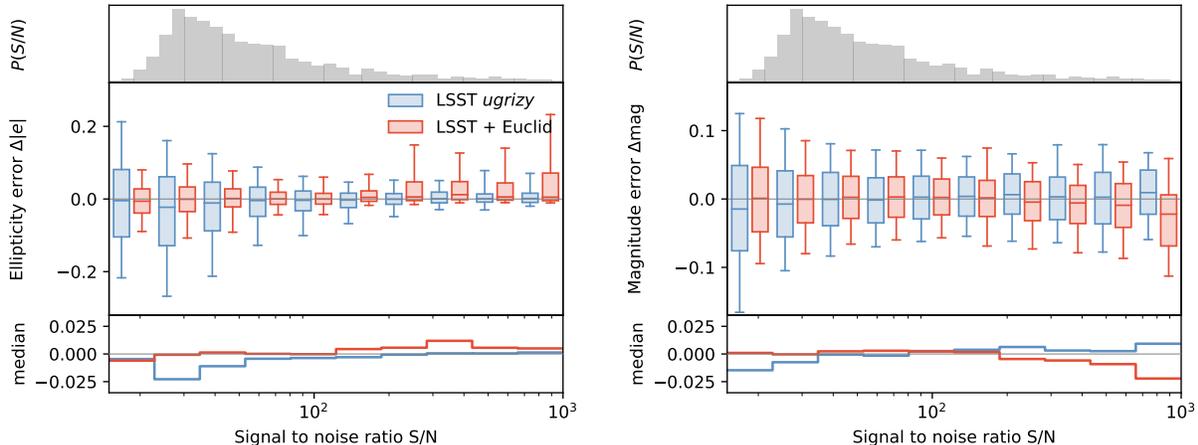


Figure 7.6: Distributions of differences in absolute observed ellipticity $|e|$ (left) and r -band magnitude (right) between target and output images from the test sample. Results for the VAE trained on LSST bands only are shown in blue and those obtained when including Euclid bands are shown in red. The distributions are computed on subsets of the test sample split in bins of signal-to-noise ratio S/N (left) and r -band magnitude (right). In each plot, the top panel shows the distribution over the test sample of the quantity on the x-axis, the middle panel shows the distribution of errors when the test sample is split in bins of this particular quantity, and the bottom panel shows the median error. In the middle panel, the boxes show the median and $\pm 1\sigma$ percentiles and the whiskers show the $\pm 2\sigma$ percentiles.

45° (e_2), which is not guaranteed *a priori*. This figure also shows on the right the mean offset on e_1 and e_2 in the plane (e_1, e_2) with arrows pointing to the output values of these components (with the same scale as the axes). We find offsets to be very small around the centre (round galaxies) and to increase for more elliptical galaxies, which are also less represented in the training sample. More surprisingly, we find an asymmetry in the (e_1, e_2) plane for more elliptical galaxies, and that the network outputs galaxy images with ellipticities slightly biased towards the directions $(-1, 1)$ and $(1, -1)$. Note, however, that results are symmetric by swapping components.

Since we have shown that the network has similar reconstruction error on the two components, we now focus on the errors on the absolute ellipticity $|e| = \sqrt{e_1^2 + e_2^2}$ and r -band magnitude, in particular on the median error and the width of the distributions. Figure 7.6 shows those distributions in both configurations when splitting the test sample in ten bins of SNR. Boxes show the median and percentiles corresponding to $\pm 1\sigma$ of a gaussian, the whiskers show $\pm 2\sigma$ percentiles. We find the distributions of ellipticity errors to be centred around zero with small deviations as the median is contained within ± 0.025 . Increased spread around the median arises at low SNR, typically below 50, particularly for the VAE trained on LSST bandpass filters only, which is likely due to the blurring of object edges in the presence of noise. The network produces on average an output image that is slightly rounder (closer to the average image) than the input image, which creates a small negative bias. The median and spread then reduce as SNR increases, even in the high SNR region which is sparsely populated. We find similar trends for both tested configurations, although the median and spread of the error distributions decrease by, respectively, 22% and 47% on average, when using Euclid bandpass filters (multi-instruments approach) in addition to the six LSST bandpass filters. The improvement is even more significant at low SNR where the median (respectively spread) is reduced by 81% (59%) on average for galaxies with $S/N < 100$. The right

panel of fig. 7.6 presents the error distributions on the reconstruction of the r -band magnitude as a function of SNR. We find that our models reproduce magnitudes with errors below 0.2 at 2σ and median within ± 0.025 , reducing in the middle range of magnitudes and increasing for very bright objects that are outliers in the training sample. Adding Euclid bandpasses reduces the median and spread on the magnitude error by respectively 72% and 21% on average for galaxies with $S/N < 100$. However, on the entire sample, the spread on the magnitude error is reduced by 12% but we do not observe any improvement on the median error, probably due to very bright outliers.

From these results, we conclude that our model is able to learn features from noisy images and accurately reproduce parameters relevant for weak lensing studies. They also show that the VAE is able to integrate information from observations in multiple bands and instruments to learn tighter posterior and likelihood distributions, leading to more accurate results across all bands. It is likely that extra information provided by Euclid filters mainly comes from the VIS instrument, which collects information on a very broad bandpass with higher resolution than LSST ($0.1''/\text{pixel}$ for VIS *vs.* $0.2''/\text{pixel}$ for LSST), yielding significant reductions of errors on galaxy shapes, especially at low SNR. These results are consistent with the findings of Schuhmann, Heymans, and Zuntz (2019), i.e. that a pixel-level joint analysis of LSST and Euclid data yields significant improvements of shape measurements.

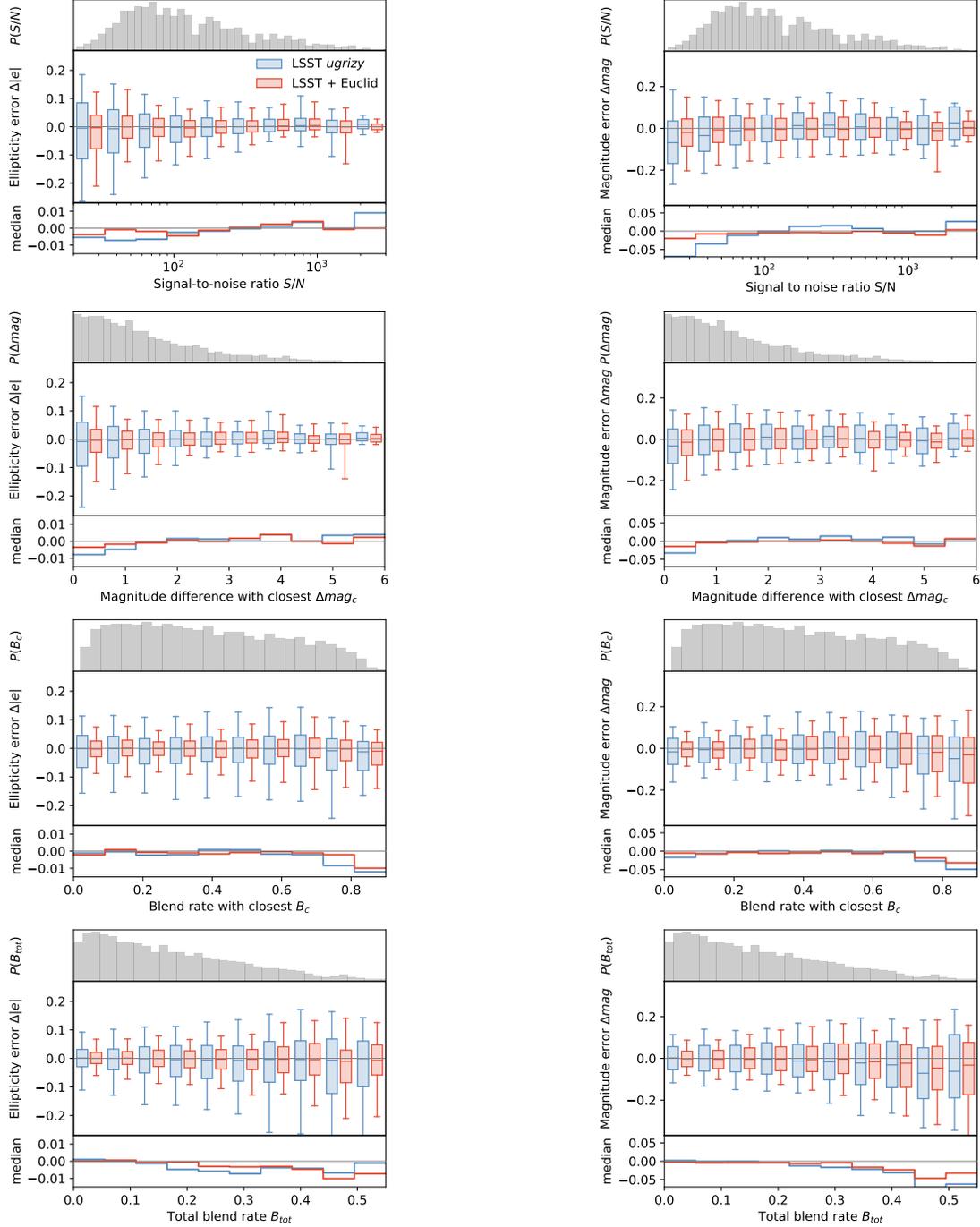
7.4.3 Deblending performance

As explained in section 7.3.3, we now instantiate a second neural network to perform deblending, with the same architecture, and reusing the generative model trained and tested in the previous section. The decoder weights are loaded from the trained prior models and held fixed during training so that only the deblender’s encoder parameters are optimised. In this section, we present results obtained on images using the same two previous configurations where target galaxies are perfectly centred. A random sample of the deblender’s input and output images is shown in fig. 7.8.

Figure 7.7 shows the distributions of errors on target galaxy ellipticity and magnitude as a function of signal-to-noise ratio (S/N), magnitude difference with the closest neighbour galaxy (Δmag_c) and the two blendedness metrics defined in section 7.2.2 (B_c and B_{tot}). For both configurations, we find the median errors to be within ± 0.015 and ± 0.08 for ellipticity and magnitude respectively, across the entire test sample. As expected, we find the largest median errors and spread at low SNR, small difference in magnitude with the closest neighbour and high blend rates. In these regimes, median errors are negative, meaning that the network produces rounder and brighter images, picking up flux from neighbouring galaxies. However, median errors remain limited even in critical cases. For example, at blend rates B_c of 80%, median errors on ellipticity and magnitude are respectively below 0.008 and 0.035.

Comparing both configurations, we observe some significant improvements when including Euclid filters. The largest observed effect is the reduction of the median error on ellipticities by 40 to 47% and of the spread by about 33% on average, which is visible when the test sample is split in bins of SNR, magnitude difference and blend rate with closest neighbour (three top rows in fig. 7.7). A smaller improvement is observed when the data is split in bins of total blend rate B_{tot} , with gains around 8% on the mean and 36% on the spread of the error distribution. Magnitude errors are also reduced across blend rates, SNR and magnitude difference: 27 to 63% for the median and 17 to 33% for the spread of the distribution.

We also characterise the impact of our deblending method on shear measurements. To do so, we generate five test samples of 10 000 input and target images. Each sample is made of the same sets of galaxies (identical profiles, positions and rotations) and noise realisations, but with different amounts of constant shear applied, namely $(0, 0)$, $(+0.01, 0)$, $(-0.01, 0)$, $(0, +0.01)$ and $(0, -0.01)$.



a) Error in ellipticity

b) Error in magnitude

Figure 7.7: Distributions of errors on absolute ellipticity $|e|$ (left) and r -band magnitude (right) for the deblenders trained on LSST (blue) and LSST+Euclid (red) images. These distributions are shown as a function of the signal-to-noise ratio S/N of the target galaxy, the magnitude difference with the closest galaxy Δmag_c , the blendedness with the closest neighbour B_c and the total blendedness B_{tot} . In each plot, the top panel shows the distribution over the test sample of the quantity on the x-axis, the middle panel shows the distribution of errors when the test sample is split in bins of this particular quantity, and the bottom panel shows the median error. In the middle panel, the boxes show the median and $\pm 1\sigma$ percentiles. The whiskers show the $\pm 2\sigma$ percentiles. In each column, all plots have the same vertical scales to facilitate comparison.



Figure 7.8: Random sample of normalised images processed by the deblender trained on both LSST and Euclid bandpass filters, where the target galaxy is perfectly centred on the stamp. We show two examples per row, with the first image showing the noisy input image, the second the output of the deblender and the third the noiseless target. These images are all shown in the *gri* bands using the normalisation defined in section 7.2.1. On each image, the blue cross indicates the centroid of the target galaxy and the red ones those of other galaxies in the scene. Images have been cropped to 50% for improved visualisation.

We decompose the ellipticities measured on output images e_i^{out} , for each component $i = 1, 2$, as a function of the target image ellipticity (considered as the intrinsic ellipticity of the galaxy) and of the true shear γ applied to the sample, as

$$e_i^{\text{out}} \approx (1 + \alpha_i)e_i^{\text{in}} + (1 + m_i)\gamma_i + c_i, \quad (7.8)$$

where α_i is the mean ellipticity bias when zero shear is applied (its measurement is presented in section 7.4.4), m_i is the multiplicative shear bias, and c is the additive shear bias. In order to measure the shear biases m_i , we start by computing the average ellipticity, denoted $\langle e_i^{\text{out}} \rangle_{\pm}$, of the samples where a constant shear $\gamma_i = \pm 0.01$ has been applied (with the other component unsheared). Shear biases can then be obtained by differentiating eq. (7.8), where intrinsic input ellipticities e_i^{in} are equal for all samples, leading to

$$m_i \approx \frac{\langle e_i^{\text{out}} \rangle_+ - \langle e_i^{\text{out}} \rangle_-}{\Delta\gamma} - 1 \quad (7.9)$$

with $\Delta\gamma = 0.02$. This method resembles the metacalibration algorithm (E. Huff and Mandelbaum, 2017; Sheldon and E. M. Huff, 2017) to quantify shear responses, except that we apply shear directly to the image model and not to a noisy image. We find uncalibrated multiplicative shear biases of the order of 1 to 4% for both LSST and LSST+Euclid configurations on deblended images. These values are to be compared with biases of 4 to 6% measured on target images, characterising the bias due to the shear estimator alone on our test sample. We interpret this result as a percent-level negative bias introduced by the deblending method, consistent with previous observations. The observed additive bias, measured on the (0,0)-sheared sample, is consistent with zero (up to the sampling error). Note that these are uncalibrated biases, measured on a sample of highly blended galaxies scenes, unrepresentative of the full sample. As such, they are not directly comparable to the values set by the LSST Science Requirements Document (The LSST Dark Energy Science Collaboration et al., 2018). Nevertheless, following the principle used here for this measurement, the calibration of this multiplicative bias could be achieved with the metacalibration algorithm, which we reserve for future work.

7.4.4 Effect of decentring

In this section, we study the impact of decentring on deblending performance. We compare results obtained in the previous section on perfectly centred images to the two decentring configurations discussed in section 7.2.1, i.e. either a shift within a square of side $0.2''$ around the centre or a shift applied after centroid location obtained with a basic peak detection algorithm. In this part, images are composed of the 6 LSST bandpass filters only. We first train VAE on isolated galaxies in a similar fashion to our fiducial case, except that they are now off-centre. We have obtained error distributions qualitatively and quantitatively very similar to those presented in section 7.4.2 and fig. 7.6, with median errors and distribution width only degraded by a few percents in both cases. We therefore directly present results on deblending performance.

Figure 7.9 shows the impact of decentring on the distributions of ellipticity and magnitude errors as a function of SNR and total blend rate B_{tot} . In the first decentring configuration, the impact is relatively limited. Quantitatively, the ellipticity and magnitude median errors increase by a factor 1.4 to 3.5 and 1.4 to 2.8 respectively, but mostly remaining within ± 0.02 and ± 0.05 (except for the two last bins in total blend rate and first bin in signal-to-noise-ratio) and the spread of error distributions increase by 12% and 22% on average respectively for ellipticity and magnitude errors. In the second decentring configuration, we observe a general increase of the width of error

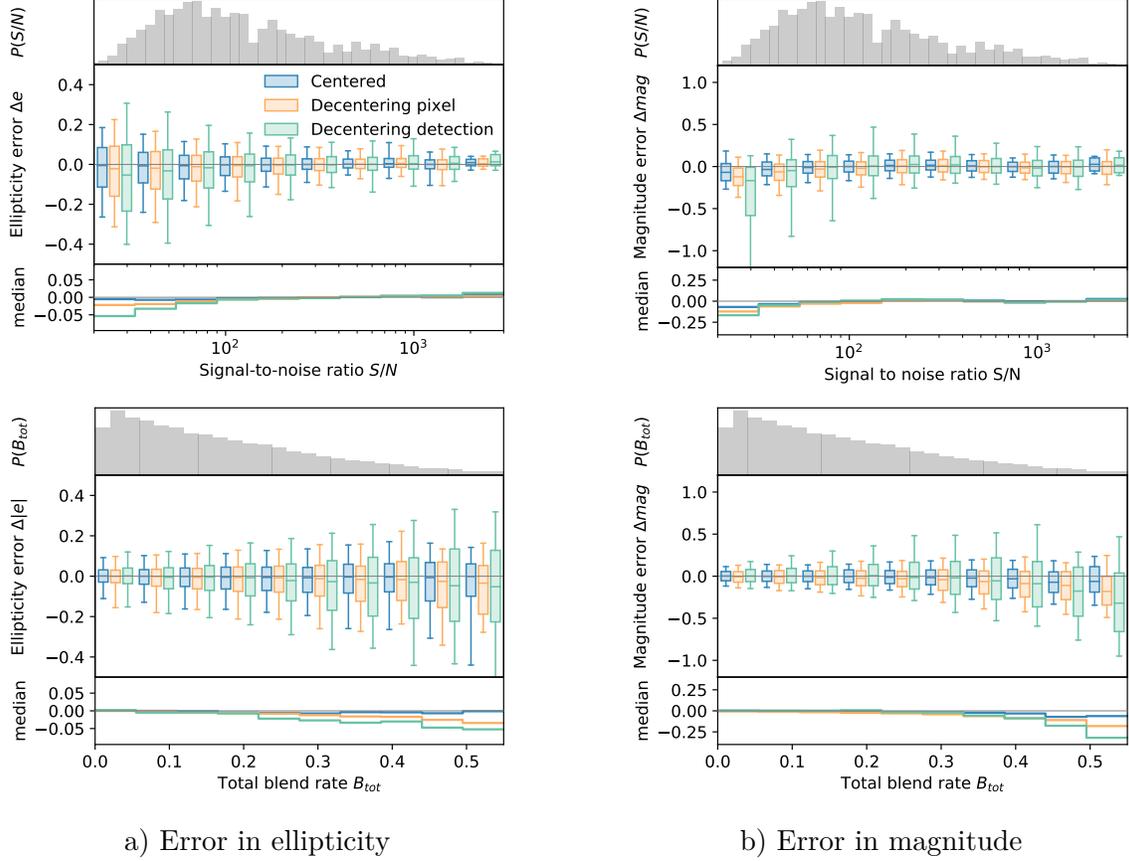


Figure 7.9: Comparison of results for LSST deblenders trained with three different centring configurations: the first one is perfectly centred on the stamp, the second one has its centre in a square of size $0.2''$ around the stamp centre, and the third one has its centre computed by a simple peak detection algorithm (from the `photutils` library) on blended galaxies images (see section 7.2.1). The poor performance of the peak finder hamper those of the deblender network, which, in contrast, seems robust to decentering only due to pixelisation.

distributions. While it performs similarly at $S/N \gtrsim 60$ or blend rates below 25%, we observe significantly degraded performance at low SNR or high blend rates. Overall, the median errors on the whole test sample are increased by a factor of 2 to 6.5 and 1.6 to 3.2 respectively for ellipticity and magnitude. The spread is then increased by 39% for ellipticity and 53% for magnitude in average.

We further evaluate the impact of decentering by computing mean multiplicative ellipticity biases α_i for $i = 1, 2$, defined in eq. (7.8), by fitting a linear regression between output and input ellipticities at zero shear. We compute these biases introduced by our deblender as a function of SNR, which is shown on fig. 7.10 for the four tested configurations. We find negative biases that decrease in amplitude for higher SNR and increase as we probe more complex decentering procedures. In the case of perfectly centred galaxies, we obtain an overall bias of -0.056 for LSST alone, which is significantly reduced to -0.016 for LSST+Euclid. On the decentred configurations, we measure biases of -0.081 and -0.084 .

This confirms that the deblender network can exploit colour information from the six LSST bands to perform deblending to a reasonable accuracy. In addition, it can benefit from additional

information provided by Euclid filters to improve shape reconstruction. Finally, we have assessed the robustness of our method to decentring, whether it is due to pixelisation or to the accuracy of the peak detection algorithm. These results thus emphasise the importance of an accurate detection pipeline for our method (as for any deblending technique) to work properly. In particular, the second case, which is strongly impacted by unrecognised blends, provides conservative performance tests. Our measurements might reveal the poor quality of the detection algorithm in case of blends rather than the actual performance of our deblender, as discussed further in section 7.5.1.

7.5 Discussion

7.5.1 Detection and decentring

As mentioned in section 7.4.4, deblending is in practice closely tied to the detection pipeline. Indeed, blending affects both the pipeline’s ability to detect multiple components and to locate centroids. In severe cases, two (or more) blended objects may be detected as a single one (*unrecognised blends*), in which case the centroid is likely off by a significant amount to any true centroids (see Dawson et al., 2016, for a study using the HST and the Subaru telescope). In this paper, we have proposed a method for deblending, which like most deblenders, requires some information about centroids. In our case, we only need the centroid of the galaxy to be deblended –but not the number of objects or their centres, which is a strong asset of our method.

However, centroids necessarily include some error. We therefore studied the impact of decentring by considering two cases: an optimistic case including pixelisation where errors have a uniform distribution of about one LSST pixel, and a more conservative case based on a simple, single-band peak detector used to both detect the brightest galaxy within a blended scene and locate its centroid. The first configuration leads to only slightly degraded errors, demonstrating that our method can deal with sub-pixel decentring. The second configuration results in more significant impacts on the spread of reconstruction errors, especially for critically blended scenes and at low SNR. However, several comments can be made. First, we underline that we considered a simple peak detection algorithm and one might naturally assume that a more elaborate detection algorithm making use of multiple bands (and/or instruments) could perform better in blended scenes, leading to improved deblending performance. Second, we also used this algorithm to generate the training samples (to the exception of rejecting galaxies too close to each other), which is not optimal as difficult cases may hamper optimisation. In particular, unrecognised blends often entail large decentring, yet, they form a large portion of the training sample, especially at low SNR, due to the poor performance of the detection algorithm. At last, another caveat is that the network’s objective is also modified to deblend the brightest detected centroid galaxy in the r -bandpass filter instead of the smallest magnitude one, which we found to differ in 20% of the training sample. Taking these different considerations into account, we suspect that the performance of our deblending method are indirectly degraded by those of the detection algorithm. We expect that they would be improved by an algorithm suited to deal with blended scenes, potentially one based on neural networks as well. We therefore consider those results to be somewhat of a lower bound on the performance of our method.

Furthermore, we envisage the optimal way to isolate every galaxy within a blended scene with our method would involve an iterative procedure of peak detection and deblending. With this procedure, one does not need to make assumptions about the number of objects in the first pass, unlike most other deblending algorithms, such as SCARLET, for instance. This process might be helpful in unrecognised blend cases, where faint, undetected galaxies could be identified after a few iterations. We explore this avenue in section 7.7.

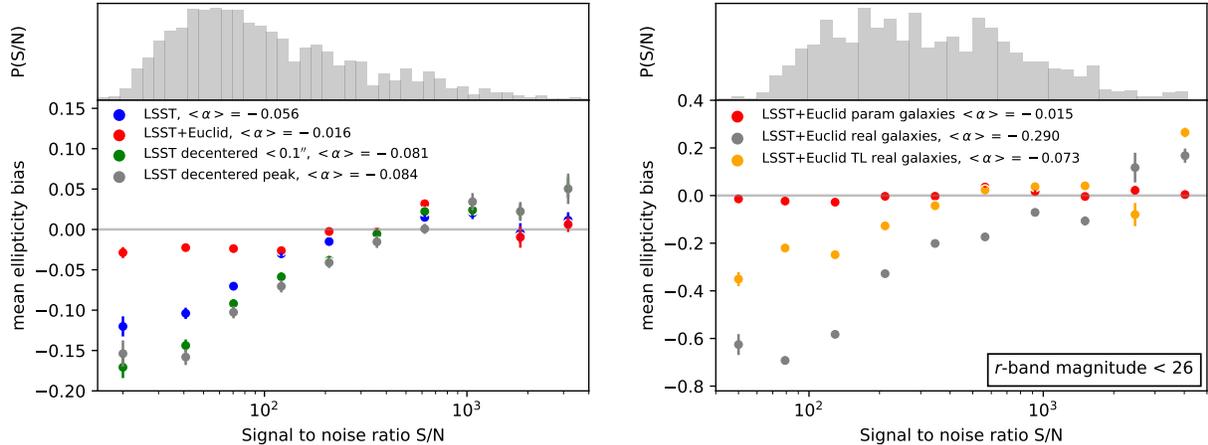


Figure 7.10: (left) Multiplicative ellipticity bias (averaged over the two ellipticities components) between target and deblenders’ output images, as a function of SNR in the four tested configurations: LSST and LSST+Euclid filters on perfectly centred target galaxies and two decentering configurations (simulating pixelisation and a simple peak finder algorithm, see section 7.2.1). (right) Multiplicative ellipticity bias (averaged over the two ellipticity components) between target and deblenders’ output images, as a function of SNR in three tested configurations: LSST+Euclid filters trained and tested on parametric images (in red), LSST+Euclid filters trained on parametric images and tested on real images (in grey), and LSST+Euclid filters first trained on parametric images, then on real images (applying *transfer learning*) and tested on real images (in orange). This figure is realised for galaxies with a minimal magnitude of 26 in the r -band filter.

Finally, we have made the assumption that all galaxies in our sample have a counterpart in both LSST and Euclid, regardless of their magnitude or other properties. For galaxies between the Euclid and LSST magnitude limits, Euclid images are noise-dominated and therefore do not carry much information. However, Euclid images could be used in blended scenes to obtain centroids for galaxies above the Euclid magnitude threshold and facilitate deblending, thus making the most of the high resolution of the VIS instrument for a joint, pixel-level analysis.

7.5.2 Challenges with real data

In this work, we have used simulated images allowing us to compare our results to a ground truth, a common practice to calibrate and validate algorithms in weak lensing studies (see, for instance, Mandelbaum et al., 2018b). However, the driving motivation in employing machine-learning techniques in general is to use real data to build more realistic models that reproduce the diversity of galaxy morphologies (including colour gradients) and observational systematic effects.

We have therefore tested our deblender network on more realistic images. In concrete terms, we have generated images with `GalSim` from the COSMOS catalogue, but this time using the real images of galaxies to capture more complex morphologies, which we then rescaled in different bands to reproduce the flux of their parametric counterparts. We input those to trained deblenders and obtained promising results (see fig. 7.11). We also trained deblender networks on a sample of “real” images, using networks that were pretrained on parametric images, a technique known as *transfer learning* (see, for instance, Domínguez Sánchez et al., 2019, for an application in galaxy classification).

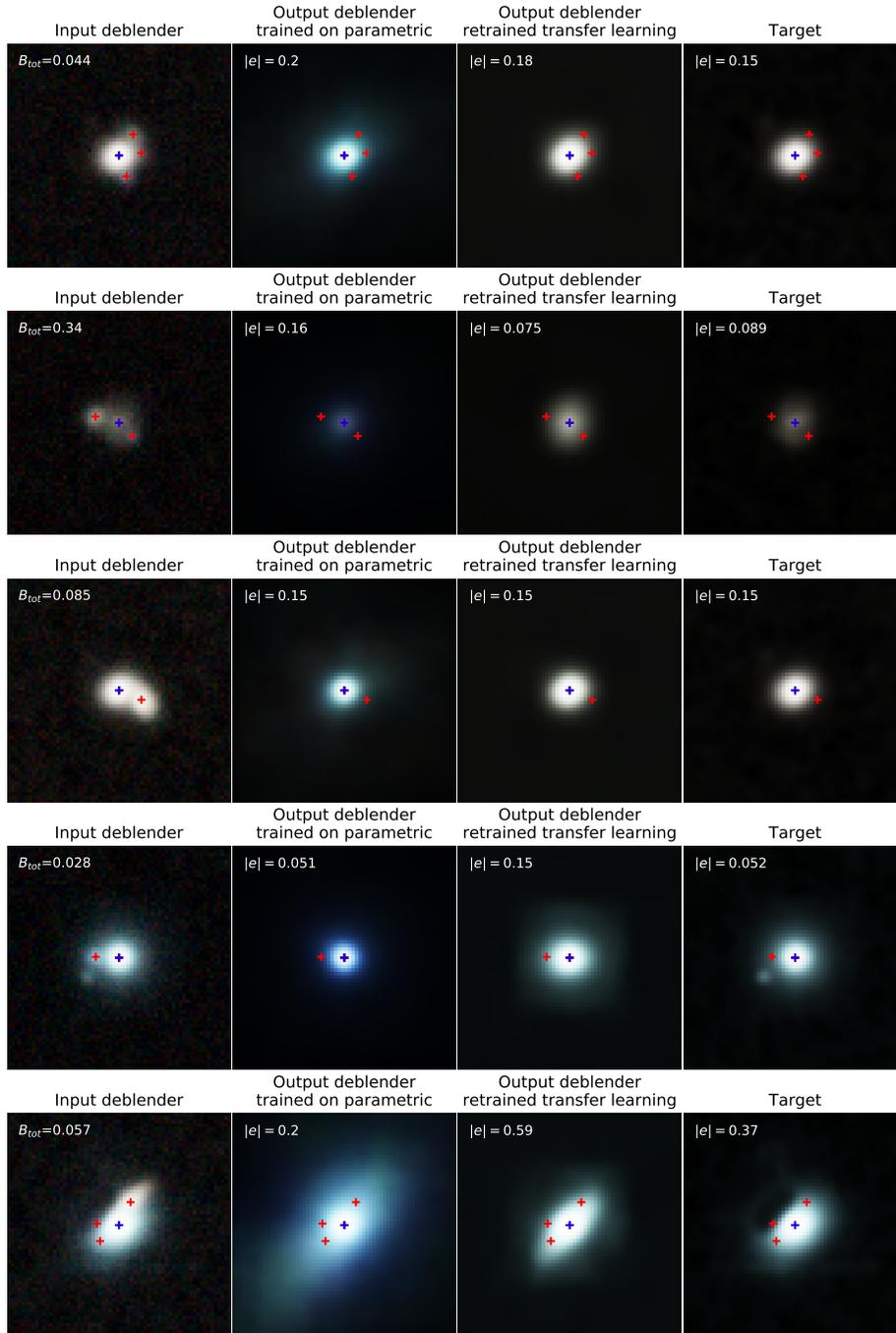


Figure 7.11: Sample of images processed with the LSST+Euclid deblender networks, before and after applying transfer learning. The first column shows the noisy input image and the fourth one the target galaxy image. As mentioned in section 7.5.2 correlated noise and residuals of images processing appear in target images. The middle columns show the output of the network when only trained on simulated images (second from left) and after retraining on a sample including 20% of real images (third from left).

Transfer learning significantly improves reconstruction, decreasing the multiplicative ellipticity bias by a factor of about 2 on the entire test sample (r -band magnitude < 27.5) and by a factor of about 4 when considering only galaxies with r -band magnitude < 26 (see fig. 7.10). However, a major obstacle is that HST images have a limiting depth of 25.2 (i -band), making it difficult to produce effectively noiseless target images, and they present correlated pixel noise that cannot be straightforwardly corrected for with `GalSim`, making adaptation difficult. Moreover, this sample still contains clear blends and residuals of image processing, despite exclusion cuts already available in `GalSim` (see, for instance, the last two rows of fig. 7.11).

This test suggests that assembling a clean and complete training sample from observations is challenging, as deep field data, which have higher SNR, also present more blending in the first place. Moreover, selecting isolated objects could introduce selection biases, for instance isolated galaxies are harder to find in high density regions such as galaxy clusters. Nevertheless, a plausible avenue will be to first assemble a small training sample of images of confirmed isolated galaxies, possibly through visual inspection or another network specialised in detecting blends. Then, this could be combined with a larger sample of realistic simulated images, in order to properly apply transfer learning. Large, realistic image simulations including all known instrumental and observational effects (e.g. quantum efficiency of CCDs, electronic read-noise, brighter-fatter effect, persistence of IR detectors, asteroids, cosmic rays) are already being produced for upcoming surveys (as part of the Data Challenge 1 and 2 simulations for LSST Dark Energy Science Collaboration, see Sánchez et al. (2020) and Korytov et al. (2019)), from which it will be possible to compile high-quality training samples for blending-related algorithms. I investigate deblending with VAEs on DESC Data Challenge 2 (DC2) images in section 7.8.

Finally, we note that another possible avenue to test the deblender on more realistic scenes consists in injecting simulated galaxy images in real ones. This method does not solve the problem of training a generative model on real galaxies, but it would assess the ability of the network to recover the learned parametric model from images including other real galaxies, but also stars and any observational defects.

7.5.3 Towards a Bayesian weak-lensing analysis

As explained in section 7.3.1, the variational autoencoder output consists of a distribution over image space (usually taken as the product of independent distributions for each pixel in each band). The input should therefore be compared to samples drawn from the output distribution. However, as explained in section 7.3.2, we interpreted, as it is common practice, the networks' output as an image. We therefore commit an abuse as the Bernoulli distribution does not correspond to our data which is distributed in the range $[0,1]$.

The output of our method is probabilistic in the sense that we do sample latent variables according to the approximate posterior. We show in fig. 7.12 examples of such sampling which allow us to obtain a posterior distribution on the ellipticity and magnitude of a target, deblended galaxy. Note, however, that the approximate posterior is computed from blended galaxy images while the decoder computes a posterior from a prior trained on isolated galaxy images. This difference breaks the Bayesian property of the loss function and consequently, authors warn that using the deblender's output to feed a Bayesian pipeline could lead to misleading results.

A solution to both aforementioned issues, which will be studied in subsequent work, is to directly generate multivariate posteriors on shape and flux parameters. This path is explored in chapter 8.

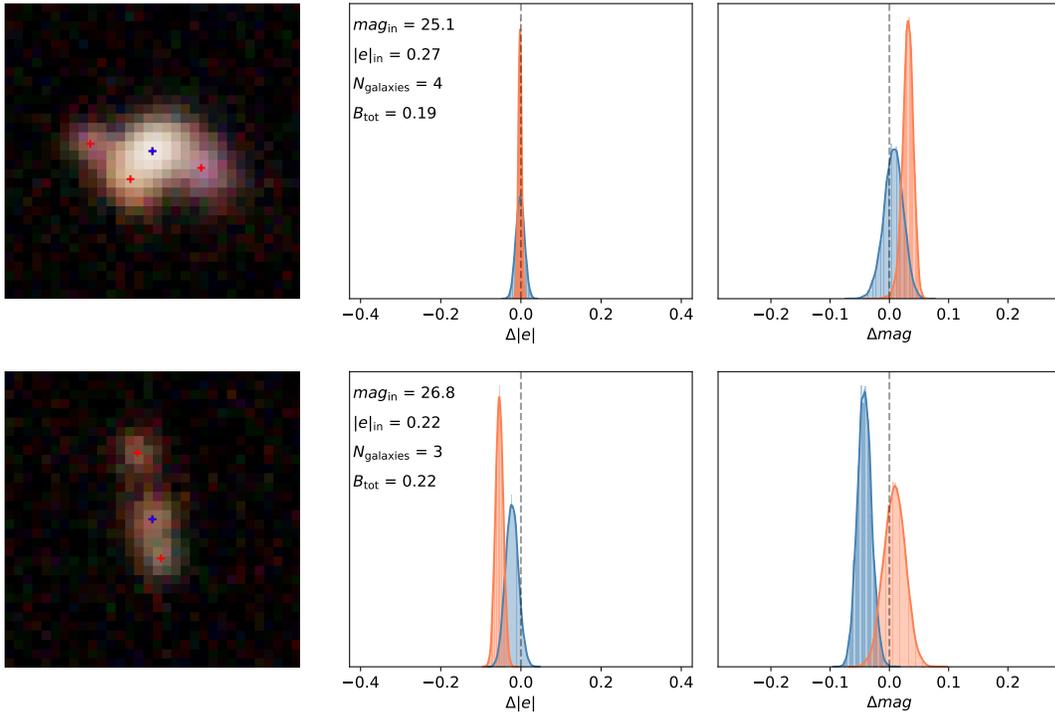


Figure 7.12: Distributions of ellipticity and magnitude errors for individual images (shown on the left) passed 10 000 times in the LSST and LSST+Euclid deblenders. For each image, the encoder parametrises the approximate posterior from which latent variables are sampled.

7.6 Summary

Blending of galaxies will have a major impact on weak lensing studies of future surveys, especially deep, ground-based ones such as LSST. In the previous sections, we have investigated a new method based on deep, probabilistic neural networks, namely variational autoencoders (VAE), to learn a generative model of isolated galaxy images, that we then use as a prior to performing deblending itself. For this second step, we create another network, similar to the VAE, where the generative model is kept fixed. This amounts to perform deblending in a latent space where data is embedded in a probabilistic manner. Our study is meant as a proof of concept and used simulated images emulating observations from the LSST and the Euclid surveys. They are generated with the `GalSim` software and based on a public catalogue from observations of the COSMOS field by the Hubble Space Telescope.

We found that VAEs parametrised by convolutional neural networks (CNN) provide powerful models able to learn features of galaxy images from a noisy sample with sufficient accuracy to recover ellipticities and magnitudes. Variational autoencoders have been used in the literature as denoisers, a feature of potential interest for weak lensing measurement pipelines that proves to work on isolated galaxy images. Then, the second network learns to deblend input images by recovering posterior parameters, that is, the distribution of latent variables encoding the target galaxy. We evaluated the performance of the deblender neural network in terms of statistical reconstruction errors on ellipticities and magnitudes, and demonstrated that the network was able to isolate galaxies and recover their properties with low biases. In particular, we found that the deblender network trained on LSST *ugrizy* images was able to reach a median error on shapes contained within ± 0.01 and on *r*-band magnitudes within ± 0.05 , stable across signal-to-noise ratios spanning the range 10 to 3000. We were able to further decrease the shape median error by 8 to 47% and of the widths of error distributions by about 33% when using 10-band images using all of LSST and Euclid filters (including visible and near-infrared filters). We also measured multiplicative ellipticity biases of respectively 5.6% and 1.6% for LSST and LSST+Euclid images, and uncalibrated multiplicative shear bias of the order of 1 to 4% for both configuration, averaged over the full test sample of blended images. Therefore, we have shown that our deblender networks are able to make the most of multi-band and multi-instrument images to retrieve tighter posterior and likelihood distributions. The higher resolution of the Euclid VIS instrument provides additional information to LSST bandpass filters, yielding significant reductions of errors and biases on galaxy shapes even at high blend rates and low SNR. We hope our results can encourage collaborations between future photometric surveys and the development of joint-pixel analysis and simulation tools.

We then studied the impact of decentring due to pixelisation and errors in peak detection and centroid measurement. We considered two cases, one optimistic, addressing mostly pixelisation and supposing accurate centroid localisation, and a more conservative case where we applied a simple peak finder on highly blended scenes. Even in the conservative case, our method produces a mean ellipticity bias on the measured ellipticities of 8.4%, improving with signal-to-noise ratio, which is slightly reduced to 8.1% in the optimistic case. The latter yields accurate ellipticities $|e|$ and magnitude with median errors below 0.03 and 0.2 (in absolute value), stable with signal-to-noise ratio and blending rates. Our method is consequently robust to pixelisation-related and/or modest decentring. It degrades, particularly for the spread of the distributions, with larger decentring errors caused by the associated peak detection algorithm, but would likely improve along with a more accurate detection pipeline suited to blended images. A potential avenue to measure and mitigate those biases, which we explore in section 7.4.2, is to apply a technique resembling the metacalibration (E. Huff and Mandelbaum, 2017; Sheldon and E. M. Huff, 2017) and metadetection algorithms (Sheldon et al., 2020) once training is complete. Moreover, the envisaged procedure

to isolate every galaxy within a blended scene would involve iterations of peak detection and deblending. This procedure, contrary to most other deblending algorithms, presents the major advantage of not needing to make assumptions about the number of objects in the first pass. The first steps in the implementation of this solution are presented in section 7.7.

Next, we need to adapt our method to real images. We have outlined challenges in assembling a clean training sample without unrecognised or unprocessed blends. We nonetheless conducted preliminary tests by applying transfer learning techniques with COSMOS data, as described in section 7.5.2, and obtained promising results. We therefore strongly recommend to explore this direction for learning-based deblending algorithms. A first step toward more realistic images is explored in section 7.8 with the test of this method on DC2 simulations.

Finally, another avenue, that is explored in chapter 8, is to use an encoder-like network to predict the joint posterior probability of the ellipticities and redshift, and potentially other quantities such as the magnitude in each band.

Please note that, from this point, next sections and chapter are ongoing work based on the method presented in previous sections and published in Arcelin et al., 2021.

7.7 Deblending pipeline

As explained in section 7.5.1, the optimal way to isolate every galaxy within a blended scene with our deblending algorithm would involve an iterative procedure. A major advantage of this kind of procedure is that we do not need to make assumptions about the number of objects in a blend since the deblending is done one galaxy after the other. Such procedure could be especially interesting in the case of unrecognised blends and might be able to reduce their impact on shear estimation. This part presents the first attempt at implementing it. Same as in previous sections, stars are not yet considered in this work.

This work was carried out with Thomas Sainrat, intern at the APC between July and December 2020, whom I mainly supervised with the help of Eric Aubourg, Cécile Roucelle and Alexandre Boucaud.

7.7.1 Pipeline structure

General description

The structure of the pipeline is presented on the left side of fig. 7.13. When a field of an arbitrary size is fed to the pipeline, the first step is to detect the lowest magnitude galaxy on the image (see section 7.7.2). Then a patch of the field, of size 64×64 pixels, and centred on this galaxy is cut. The centred galaxy is then deblended with our network and subtracted to the initial field image. This process is performed several times until no galaxies are detected any more.

If the detection and deblending steps were perfect, this procedure would be sufficient. However, both of them are not (see section 7.4 and next section), and it is possible that the subtraction of a deblended galaxy generates residuals on the image (see right side of fig. 7.13). If they are brighter than the noise level, those residuals could be detected instead of galaxies at some step in the process. Consequently, a procedure to deal with these residuals is necessary. We designed one including the classification of the detected object and the masking of residuals. I describe in section 7.7.3 several explored avenues for these steps.

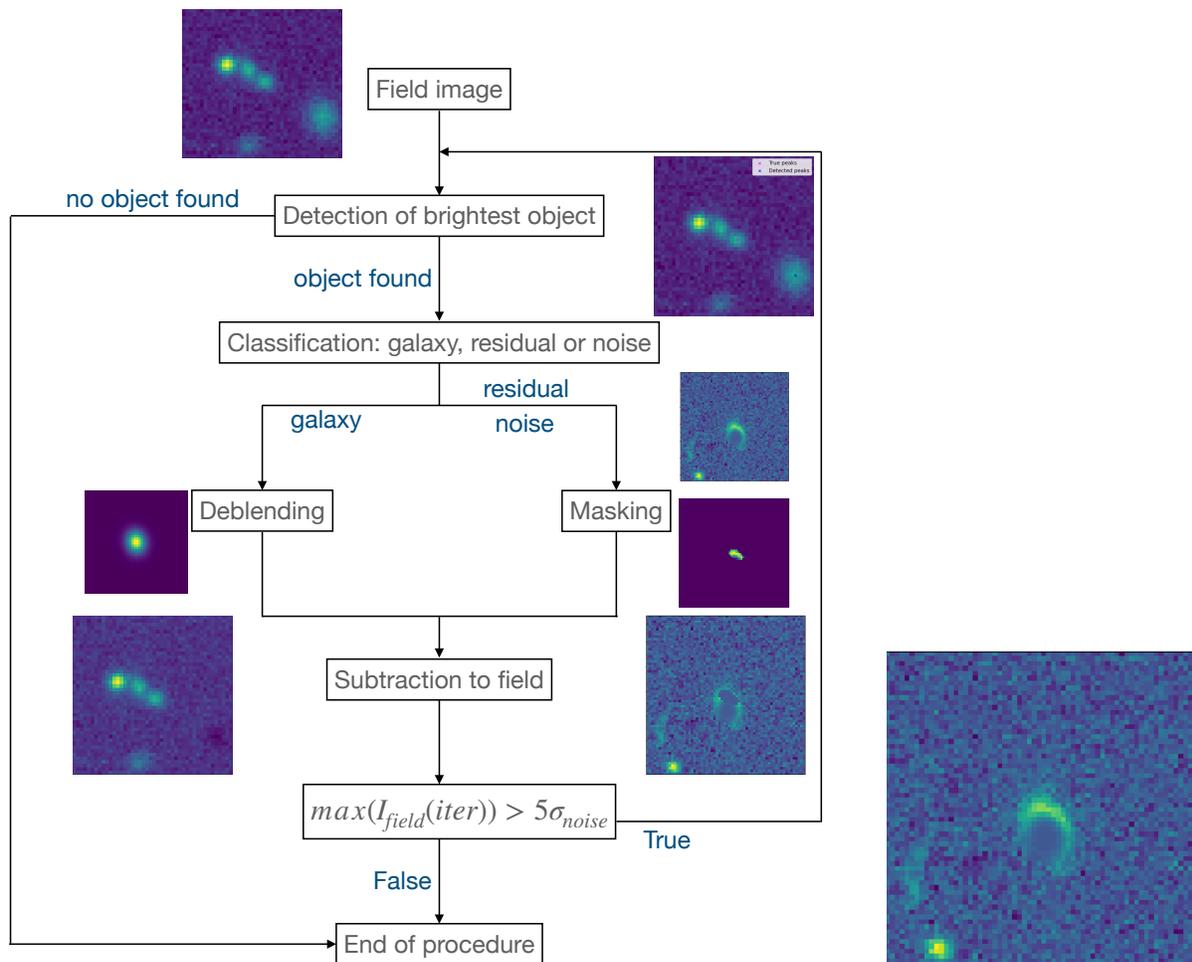


Figure 7.13: (left) Deblending pipeline including detection, classification of detected object, and deblending or masking. (bottom right) Example of residual image after the removal of a deblended galaxy.

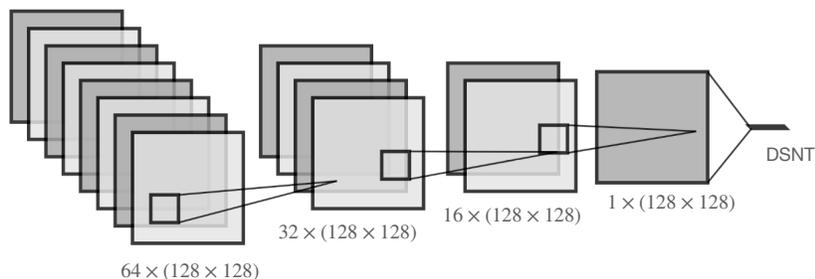


Figure 7.14: Architecture of the detection algorithm. It is composed of four convolutional layers of kernel sizes (5, 3, 3, 3), with strides of (1, 1), and number of filters (64, 32, 16, 1), followed by a DSNT layer.

The method provides two stopping criteria. First, if at the detection step, no peak is detected, the deblending process stops. Then, an arbitrary threshold of $5\sigma_{noise}$ is set as the minimum value at which a peak can be detected, with σ_{noise} the square root of the noise level. Consequently if the maximum pixel value in the field image, I_{field} , is below $5\sigma_{noise}$, the deblending process also stops. The first criterion allows to avoid having to go through each step in the case of an empty field, and the second one prevents the detection algorithm to look for galaxies with SNR lower than 5, as it only sees galaxy with a higher SNR during training (see section 7.7.2). Note that $SNR > 5$ is different from having the maximum pixel value of an image above $5\sigma_{noise}$. However, it is a sufficient approximation in our case, where this criterion is not limiting for the procedure.

Simulated images

Images used in this section are produced similarly to what is described in section 7.2.1. The only difference is that galaxies are positioned randomly on the images. They are all shifted from a random amount within half the image size, to avoid having galaxies partially cut at the image edges. The blending process consists in adding shifted images of isolated galaxies and we use a fixed Kolmogorov PSF using a FWHM of $0.65''$. We did not take it into account in our procedure and reserve the inclusion of PSF treatment in the pipeline for future work. Details about the images used to train the detection and classification algorithms are respectively presented in section 7.7.2 and section 7.7.3.

7.7.2 Detection

This section presents the implementation of the detection algorithm. To make the most of the color information, a neural network is used since, as explained previously, it allows very simply for the use of multi-channel images. Even though most used detection algorithms are not NN based, we saw that all the detection algorithms presented in section 3.3.2 are based on neural networks which demonstrates the growing interest in using them for detection.

To perform deblending, the deblender algorithm requires an image centred on the galaxy of lowest magnitude. Consequently the target for the detection network is set to be the center of this particular galaxy in the field. This is a very clear and simple target for the network to learn but note that it might be too simple for specific cases: the network might get confused in cases like

very close galaxies of same SED and at same redshift. Yet, this target is considered sufficiently precise for our purpose with this procedure.

Architecture and implementation

To implement the detection network, convolutional layers are a natural choice since we are dealing with images. It is composed of a CNN and a differentiable spatial to numerical transform (DSNT) layer (Nibali et al., 2018) as output. This layer adds no trainable parameter and allows for regression on the coordinate values, which is an improvement compared to other detection algorithms (using dense layers or heatmap matching).

The detection network is composed of a first layer of Dropout, with a rate fixed at 0.1, followed by four convolutional layers of kernel sizes (5, 3, 3, 3), with strides of (1, 1), and number of filters (64, 32, 16, 1). The last layer is the DSNT layer (see fig. 7.14).

Training, validation and test sets

The training, validation and test sets are simulated with `GalSim`. They are LSST-like, multiband images (u, g, r, i, z, y). The generated galaxies used for training all have a SNR (defined as in section 7.4.1) higher than 5, which sets our detection threshold. The galaxy models, PSF, noise and blending process are simulated similarly to what is presented in section 7.2.1. We trained this algorithm on artificially blended scenes of size 128×128 pixels, with the center of the lowest magnitude galaxy as target. There are four galaxies per scene, uniformly distributed around the center in a square of 64×64 pixels. The training set is composed of 1000 scenes, same for validation and test sets.

Detection results

The detection algorithm gives satisfying results, the network is able to detect the center of the brightest galaxy around 76% of the time and in more than 91% of the time it is one of the two brightest galaxies (see fig. 7.15). Most of the confusions between the two brightest galaxies occur when the magnitude values of those galaxies are close from each other. The network also struggles as expected on highly blended scenes. However, the mean precision of the detection compared to the closest galaxy is below half a pixel. Since we already showed that our deblending method is robust to small decentring (see section 7.4.4), these performance are considered to be sufficient for a first step and we did not spend more time on this step, even though it could probably be improved.

The detection algorithm outputs the coordinates of the detected peak. Then, to perform the deblending, an image of size 64×64 pixels is cut using the closest intersection from this peak as center. Once the deblending is done, the deblended galaxy image is subtracted to the image used for peak detection.

7.7.3 Classification and residual masking

After the subtraction of the brightest galaxy, residuals can remain. To check if the next detected object is a galaxy or a residual we add a classification step using, once again, a neural network. The network is composed of a first layer of Dropout, with a rate fixed at 0.1, followed by eight convolutional layers split into groups of 2, of kernel sizes (5, 5, 3, 3), strides of (1, 1), and number of filters (128, 64, 32, 16) (see fig. 7.16). The network ends with two dense layers. It looks at *patches*

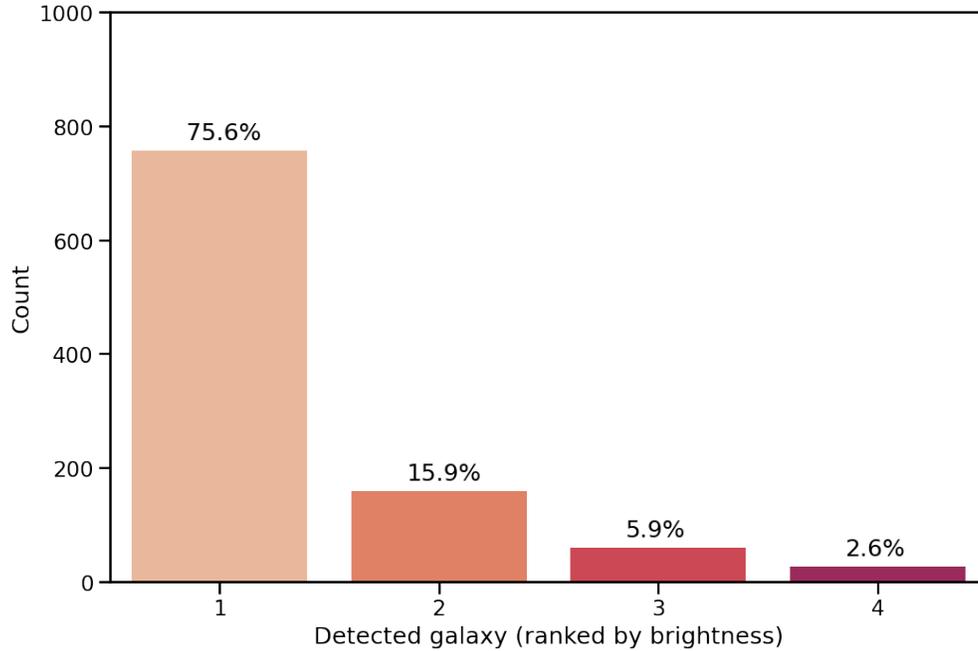


Figure 7.15: Results for the detection algorithm. In 91.5% of the time, one of the two brightest galaxies, within the four galaxies distributed uniformly on the image, are detected. Credit: Thomas Sainrat.

of 16×16 pixels around the detected center and outputs the probability for it to be a galaxy, a residual or just noise. We defined three classes: *galaxy*, *residual* and *noise*.

Labelling training data

In order to train the network we had to construct and label patches for each categories. To do so we performed the galaxy deblending of 1000 fields without the classification step, i.e. considering each detection as the detection of a galaxy. This process ensures to obtain residuals in the image. If we define a step of the deblending process as: starting from the field image to deblending a galaxy; then for each step a patch of size 16×16 pixels is created, centred on the detected peak. To label the patch, we used the remaining field image and the corresponding *segmentation maps* as in most of the classification algorithms mentioned in section 3.3.2. The segmentation maps of the blended

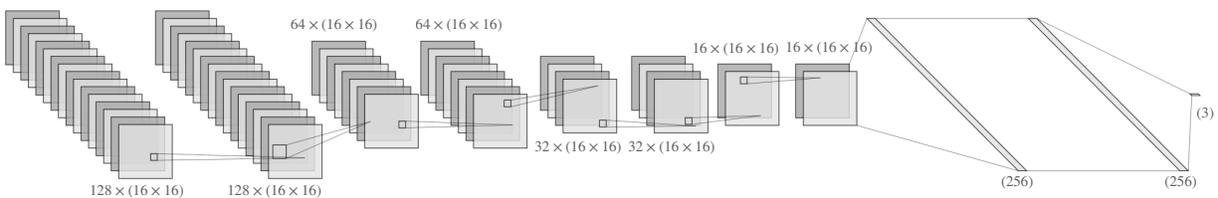


Figure 7.16: Architecture of the classification neural network. It is composed of eight convolutional layers followed by three dense layers.

scenes are created from the light profile of the noiseless galaxies.

The patch is labelled as:

- *galaxy* if the detected peak is at less than 1.5 pixel from the center of a true galaxy. Then the segmentation map of this galaxy is subtracted to the field one for the next step.
- *residual* if:
 - the detected peak is outside of the segmentation map,
 - or if the detected peak is in a pixel with intensity above $5\sigma_{noise}$ after removal of the true remaining galaxy images from the field image at this step.
- *noise* if the detected peak is inside the segmentation map but further than 1.5 pixel from the closest galaxy center and with a pixel value below $5\sigma_{noise}$

The patches extracted are always centred on the detected peak. However we needed more patches of *noise* and *residual* in order to avoid having a majority of galaxy patches. We had to create *noise* patches artificially by taking patches centred on pixels at a 2 pixel distance from a true galaxy center, and *residual* patches by drawing patches centred on pixels with intensity higher than $5\sigma_{noise}$ on the remaining image after the complete deblending process.

Overall 30000 patches are used to train the classification algorithm: around 4000 of galaxies, 6500 of residuals and the rest containing only noise. We tested different proportion of the three classes in the training data and this is the one that gave the best results.

Classification results

To classify a patch as *galaxy*, *residual* or *noise* in the deblending process, we set criteria on the probability predicted by the network. We define C_{galaxy} and $C_{residual}$ as the probability above which a patch is respectively classified as a *galaxy* or a *residual*. The classification is done by first verifying if $p_{residual} > C_{residual}$, in that case it is classified as a residual. If not, one check if $p_{galaxy} > C_{galaxy}$, in that case it is classified as a galaxy. If $p_{galaxy} < C_{galaxy}$ and $p_{residual} < C_{residual}$, the patch is classified as *noise* and is masked as explained below. This classification scheme is probably not the most relevant at it looks at the predicted probabilities one by one instead of assigning conditions considering all of them. This entailed difficulties to obtain accurate results, as explained below.

Setting values for these criteria is not a simple task. First tests integrating this process in the deblending pipeline led to the results presented in table 7.1. On the two sets of criteria tested the ($C_{galaxy} = 0.5; C_{residual} = 0.2$) configuration presents much more acceptable results allowing for less than 15% of errors for the *galaxy* class and less than 20% of errors for the *residual* class. Here the percentage of error is defined as the percentage of objects wrongly classified in a class. The issue with the (0.5;0.5) configuration comes from the classification scheme described above. Since almost none of the patches had a probability of *residual* classification higher than 0.5, they were proposed at the next condition, for the *galaxy* class. Most of the true residuals also yield a high probability of being classified as a *galaxy*, residuals and galaxies look alike, so most of them were. For true galaxies, almost none of them have a probability of being a classified as a *residual* higher than 0.5 and consequently they were almost all classified as *galaxy*.

Decreasing the criterion value for the *residual* class helped correctly classifying more of the true residuals. However, some true galaxies probably fell into the *residual* class since, as explained, residuals and galaxies look alike and some galaxies might yield a high probability of being classified as a *residual*. The reason why true residuals do not present high probability of being classified as a *residual* needs to be explored further but could be explained by the resemblance with galaxies.

$(C_{galaxy}; C_{residual})$	Error for the class <i>galaxy</i>	Error for the class <i>residual</i>
(0.5; 0.5)	1%	> 90%
(0.5; 0.2)	14.5%	19.9 %

Table 7.1: Results for the classification algorithm included into the deblending pipeline for different criteria. Here, 14.5% of error for the class *galaxy* means that 14.5% of the objects classified as galaxies are not galaxies. Setting the threshold at a probability of 0.2 rather than 0.5 for the classification of residuals allows for much better results.

I did not present results for the *noise* class as we did not find large differences between the two tested configurations, the percentage of error was below 1%.

Masking strategy

Once an object is classified as *residual* or *noise*, it needs to be masked in order for the deblender procedure to continue. Several avenues have been tested such as setting the value of the detected pixel or a group of pixels of a specific shape around it at zero. This has two major downsides: it is impossible to find a shape for all the residuals since they all are different, and setting pixels to 0 might cancel the flux of a galaxy blended with the residual.

The masking strategy that we chose consists in using the classification algorithm to detect all the pixels belonging to the residual: starting from the peak classified as *residual*, we look at all the neighbouring pixels and run the classification algorithm on them. If they are also classified as *residual* we continue around these pixels until no other pixel is classified as *residual*.

Instead of setting the pixel value to 0, we subtract a fraction, here 50%, of each pixel intensity to avoid losing too much information. If the detected peak is classified as *noise*, the value of this pixel only is modified.

7.7.4 Pipeline results

As explained, steps in this pipeline require more work before being able to assess performance on shapes or fluxes recovery. However, we managed to implement a pipeline with all the required steps in order to do deblending on a large field. We managed to make it work on small fields (see fig. 7.17) and were able to point out steps requiring improvements. We especially showed that the classification step is primordial for this process and that the stopping criterion of $\max(I_{field}) < 5\sigma_{noise}$ is not the most relevant since several iterations occur detecting only noise if we use it. Setting the detection of noise as a stopping criterion instead decreases drastically the number of non-necessary iterations.

7.7.5 Discussion and Future work

As mentioned before, the performance of the deblending process really depends on the results of the classifier. We will consequently focus on this particular step for future work. The definition of classification criteria might require refinement but other classification techniques are also considered. For example we could use the representation of galaxies and residuals in the latent space of a VAE to separate these two categories. Works to use the latent space representation for classification have already shown conclusive results (Higgins et al., 2016 or Norlander and Sopasakis, 2019 for examples).

Also, even though the masking strategy is now working correctly, it could certainly be improved. An interesting path could be to train a VAE to reproduce residuals and to use it for masking.

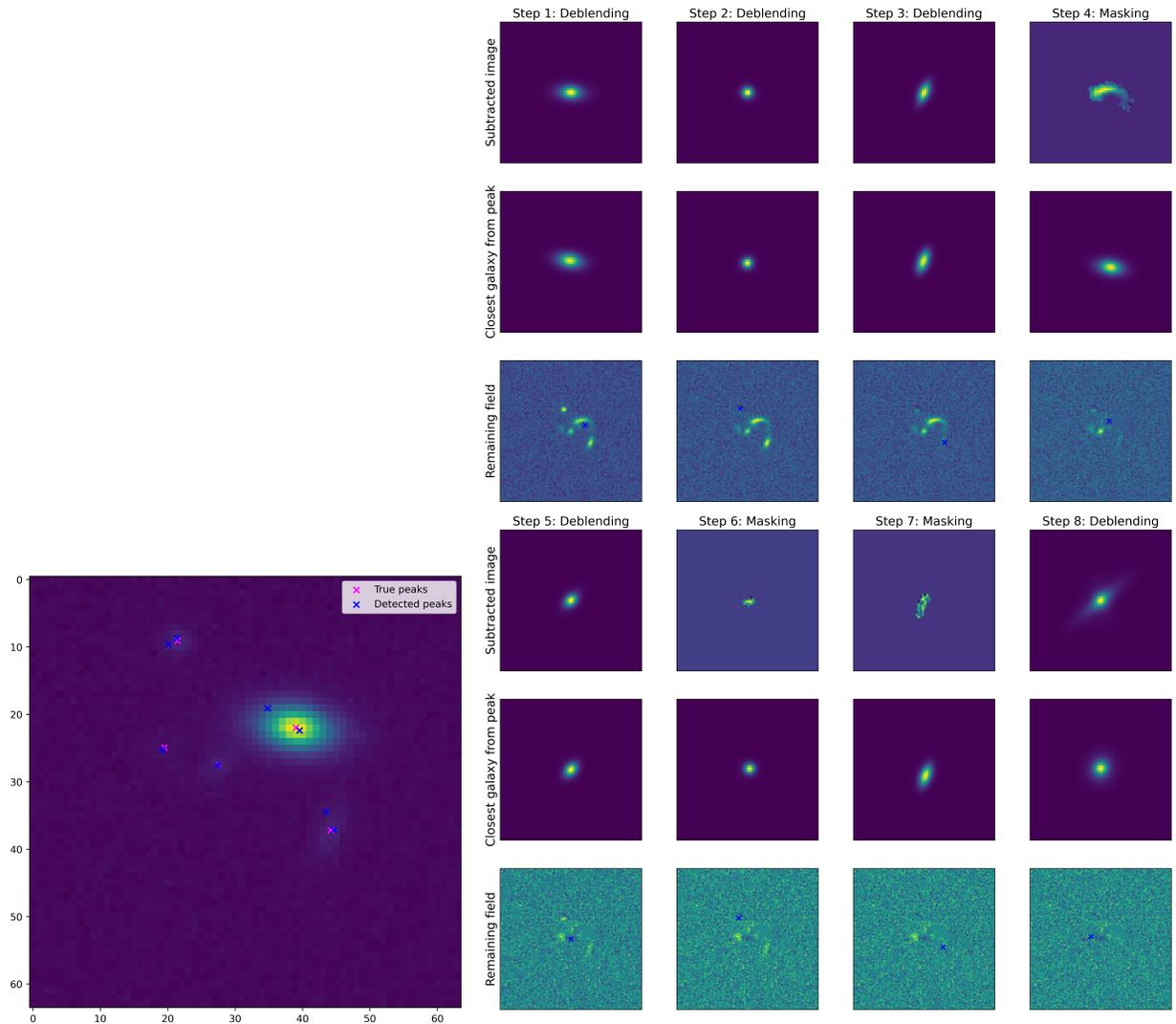


Figure 7.17: (left) Field to deblend with different peaks detected during the iterative process (blue crosses). The pink crosses are the true center of the galaxies. (right) Description of the eight steps necessary to obtain the complete deblending of the scene. Five of them are deblending steps, the other three are masking of residuals. The first and fourth rows represent the subtracted image, the second and fifth rows the image for the closest galaxy from the detected peak, and the third and last rows are the remaining field after subtraction. At step eight (bottom right), the remaining field is empty. However all deblending steps are not perfect as shown on the last step for example: the deblended galaxy is quite different from the target.

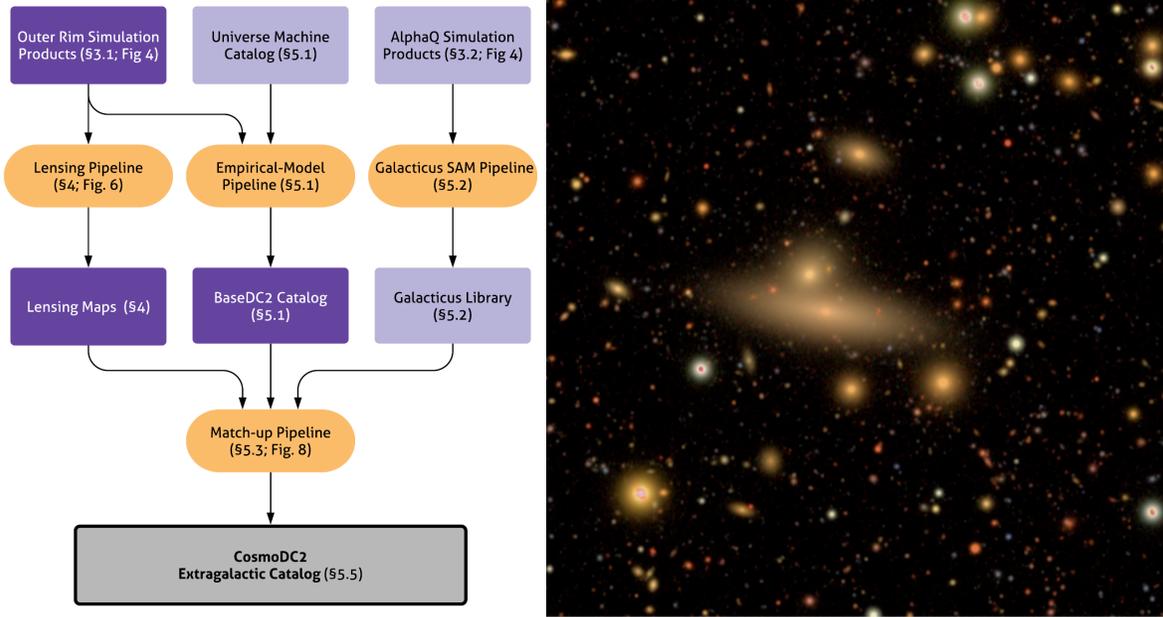


Figure 7.18: (left) Overview of the pipeline for the production of the DC2 Extragalactic Catalog. Credit: Korytov et al., 2019. (right) A zoom in the upper quadrant of tract 3828 in *gri*, from the DC2 simulated sky survey. Credit: LSST Dark Energy Science Collaboration (LSST DESC) et al., 2021

Obviously the PSF should be taken into account in this pipeline. A possibility is to take the measured PSF as input using well established PSF measurement methods.

Finally, and since using deblended objects that produces residuals above the noise level is not conceivable for weak lensing studies, we consider doing optimisation of the deblending process. The iterative procedure would be repeated several times until convergence to an optimal solution. An example of convergence criterion could be that no masking step is required during the deblending process.

These possible improvements will mainly be explored in the future by other members of the Astrodeep team at APC. Biswajit Biswas, PhD student starting in October 2021 in the team, will focus on this part.

7.8 Application to DC2 simulations

In section 7.5.2 we also mention the importance of focusing on more realistic data in order for our algorithm to be used for cosmological surveys. In this part I explored the adjustment of the method to images produced in the Data Challenge 2 (DC2) for the LSST Dark Energy Science Collaboration, (Korytov et al., 2019; LSST Dark Energy Science Collaboration (LSST DESC) et al., 2021).

7.8.1 DC2 simulations

DC2 simulations are based on a N-body (gravity only) simulation (the Outer Rim run carried out using HACC from Habib et al., 2016), and empirical (using UniverseMachine from Behroozi et al., 2019) and semi-analytic models (using AlphaQ simulation) of galaxy formation for the construction

of the catalog (see left side of fig. 7.18 for an overview of the pipeline). The combination of these different simulations produces a catalog: the Extragalactic catalog. It covers 440 deg^2 of sky area up to a redshift of $z = 3$ and to a magnitude of 28 in r -band. It contains all the properties of the galaxies that are required for image simulation (Korytov et al., 2019).

This catalog serves as input for the generation of the DC2 images which covers an area of 300 deg^2 of the Wide-Fast-Deep (WFD) area in the six LSST optical bands (LSST Dark Energy Science Collaboration (LSST DESC) et al., 2021). Data are simulated as if they were the product of the first 5 years of the planned 10-year survey. The images are produced using the `imSim`⁸ software package which itself calls on `GalSim` (Rowe et al., 2015) for astronomical object rendering (see right side of fig. 7.18 for an example). This package was designed to be able to produce simulation of the sky as it will be seen by LSST. The simulation takes into account information from the several LSST software libraries to simulate realistic images. For example, for each visit, the `OpSim` database gives information about the seeing conditions and the sky brightness, the atmospheric and optical PSF are precisely simulated, and defaults due to CCDs are also taken into account (brighter-fatter or tree rings for example).

From this simulation it is possible to extract *cutouts* centred on the galaxies detected by the LSST pipeline (section 7.8.2) (Bosch et al., 2018). However, the noiseless images of these galaxies are not available and I need to simulate them using `GalSim` (section 7.8.2). Since galaxies identified on the DC2 simulation by the LSST pipeline are registered in the catalog coming out of the LSST pipeline, the `Object catalog`, and that the true parameters given as input for the DC2 simulation are registered in an other catalog, the `Extragalactic catalog`, it is necessary to match both catalogs to have the correct information on the observed galaxy. Both are accessible through the python package `GCRCatalogs`⁹.

7.8.2 DC2 images for training

Cutouts in DC2 and catalog matching

To extract cutouts from the DC2 sky simulation, the sky coordinates from the target galaxy are required. For this work, two samples of target galaxies are defined, respectively with cuts in the r -band magnitude at 24.5 and 26.5. First the method is applied to galaxies with a magnitude below 24.5. It will be applied to fainter galaxies in a second step, once the method is set.

The (ra, dec) coordinates of the objects detected with a magnitude lower than the defined threshold are first extracted from the `Object catalog`. In order to avoid obtaining very large output, the objects are required to come from a particular tract (part of the simulated sky, see fig. 7.19), which number is specified when extracting the parameters. It also defines intervals in ra and dec which are used as limits to extract all the source's parameters, for sources respecting the magnitude criterion, from the Extragalactic catalog.

Then the catalogs are matched in order to extract, from the Extragalactic catalog, the parameters of the galaxies extracted from the DC2 sky simulation (from the `Object catalog`). To do so, the algorithm `FoFCatalogMatching`¹⁰ (Mao et al., 2021), based on the friend-of-friend method, is used. Finally, all the parameters required to simulate the noiseless images of the selected galaxies, are extracted from the Extragalactic catalog (see section 7.8.2). Similarly, other parameters are extracted from the `Object catalog` such as the SNR in the r -band, the blendedness (cf. definition presented in section 3.1.2) and the ellipticity parameters measured by the LSST pipeline with HSM

⁸<https://github.com/LSSTDESC/imSim>

⁹<https://github.com/LSSTDESC/gcr-catalogs>

¹⁰<https://github.com/yymao/FoFCatalogMatching>

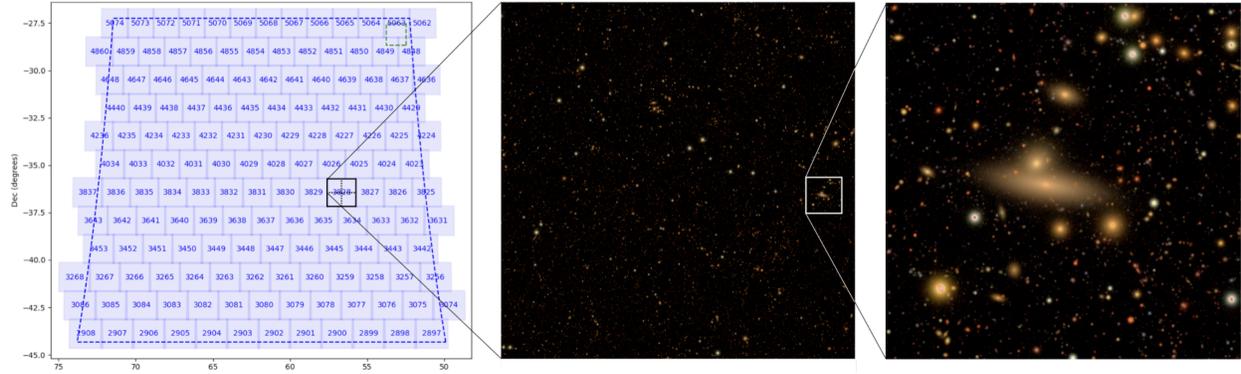


Figure 7.19: Illustration of the area of the DC2 simulation. The left panel shows the different tracts on the simulated area. The middle panel shows the upper quadrant of the tract 3828 in *gri*. The right panel shows a zoom on the extracted image. Credit: LSST Dark Energy Science Collaboration (LSST DESC) et al., 2021

(for the pipeline see Bosch et al., 2018, for the HSM algorithm see Hirata and Seljak, 2003). Those parameters are used to assess the performance of the deblending algorithm.

Once the catalogs are matched, the (ra, dec) coordinates of the matched objects are used to extract cutouts from the DC2 sky centred on these objects. Both a stamp centred on the galaxy and a stamp of the PSF at this point, measured by the LSST pipeline, are extracted. Indeed, the PSF varies for each image and in each band here, contrary to what has been done in section 7.2.1. The deblender architecture presented in section 7.3.2 might not be sufficient to take into account this effect. This is why I also require stamps of PSF, to test a different architecture which could be fed with both images of galaxies and the corresponding PSF. I present this architecture in section 7.8.3.

The training sample is built generating 100000 cutouts centred on galaxies from ten different tracts. The validation and test samples are respectively composed of 20000 cutouts from two tracts and 10000 cutouts from one tract. The tracts used to generate validation and test samples are different from the one used for the training sample. Figure 7.20 shows the distribution of the absolute ellipticity, $|e|$, the *r*-band magnitude, and the redshift, z , for the sample with a magnitude cut at 24.5 in *r*-band magnitude.

This method of image extraction prevents a strong assumption that we had set in section 7.2.1: the galaxy in the center might not be the brightest one on the image. Consequently, this adaptation on DC2 images also proposes a generalisation of our method. Note also that the relaxation of this assumption would entail a modification of the deblending procedure (see section 7.7). It could be easily modified by defining a step in the procedure as 1) a detection run, looking for every objects, 2) the classification of these objects (galaxy, star, residual, noise) and 3) the deblending or masking of each of these objects.

Noiseless galaxy images simulation

In order to train the VAE and the deblender, target noiseless and non-blended images of the galaxies detected on the DC2 images are required. To simulate them, the galaxy parameters were collected from the Extragalactic catalog and the PSF was extracted from the cutouts. *GalSim* is used to reproduce the galaxy image. The galaxy is simulated using bulge+disk models and the

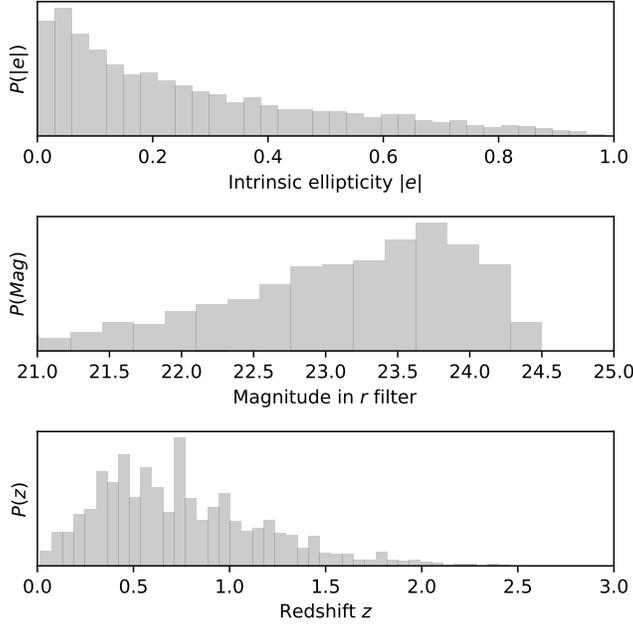


Figure 7.20: Distributions of observed ellipticity, r-band magnitude (a cut is applied at 24.5) and redshift for galaxies in the test sample from the DC2 simulation.

properties from the Extragalactic catalog, especially bulge and disk half-light radii, ellipticities, and the flux in each bands. It is then convolved with the measured PSF. The image generation is visually checked by subtracting the generated isolated galaxy image to the initial image (see fig. 7.21). The code to collect the required parameters and generate the image is available here: https://github.com/BastienArcelin/dc2_img_generation.

Note that the PSF used for the image generation is measured by the LSST pipeline. It is not the exact PSF used in the DC2 simulation (which was unavailable). The noiseless isolated galaxy images produced are then slightly different from the ones used as input for the DC2 image generation. However, since the weak lensing parameters (galaxies shapes and fluxes) are used to simulate the noiseless images, which are the target for the deblender, these small differences are assumed to have a negligible impact on the deblending performance.

Finally, the galaxy is perfectly centred on the simulated noiseless images. It is not necessarily the case for the detected galaxy on the DC2 cutouts since it depends of the performance of the detection and astrometry algorithms used within the LSST pipeline. It should not be a problem to ask the decoder to generate the image of the galaxy on the center of the stamp but an additional step might be necessary in order to slightly shift the deblended image before subtracting it to the field in the deblending procedure.

Data augmentation

The data augmentation is done similarly to what is presented in section 7.2.3: random flips of images horizontally and vertically and rotation by 90 deg on the fly. The normalisation is slightly different, in each band b , both noisy and noiseless images are normalised following,

$$x_b = \tanh\left(\sinh^{-1}(x_{\text{raw},b})\right), \quad (7.10)$$

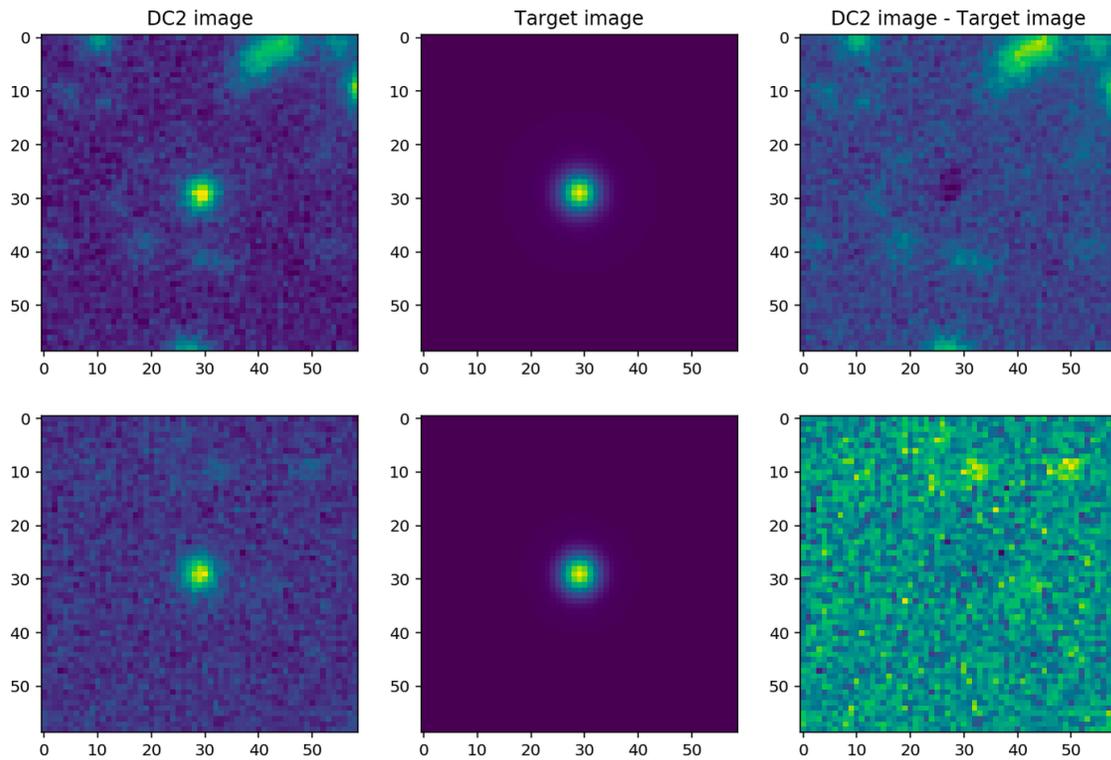


Figure 7.21: Examples of cutouts from the DC2 simulation (left), the corresponding noiseless images of the centred galaxy simulated with `GalSim` used as target for the networks (center), and the subtractions of the simulated noiseless galaxies from the cutouts (right).

with $x_{\text{raw},b}$ the raw noisy or noiseless image.

7.8.3 Network architecture choice and training procedure

Network architecture

As mentioned previously, the deblender architecture presented in section 7.3 might not be sufficient to take into account varying PSF. This is why two different architectures, based on the one used in section 7.3, are tested:

- the first one adds a similar encoder in parallel of the first one, to take into account the varying PSF. One is fed with the image of the galaxy and the other one with the image of the corresponding PSF. The two encoders merge via a concatenation layer which is flattened to feed the dense layer located before the latent space (see fig. 7.22). The decoder does not change from the one presented in section 7.3.3. The idea behind this architecture is that the network could learn information from the PSF image, how to encode it in the latent space, and how to use it to help the decoder reproducing an accurate deblended galaxy image. The architecture is called the DEVAE (for Double Encoder VAE) in the following.
- the second one is similar to the one presented in section 7.3.3.

The DEVAE is designed with two encoders to take into account the PSF and encode necessary information into the latent space. However, this second encoder might not be necessary as the VAE architecture could be sufficient to extract this information directly from the DC2 images. The loss function used for the VAE and DEVAE is eq. (7.5) The performance of those two architectures are compared at the VAE step, results are presented in section 7.8.4.

The two deblenders are implemented similarly to the VAE and DEVAE and, as in section 7.3, the trained decoder from the VAE and DEVAE are used as decoders of the deblenders. Their weights are fixed during training, the deblenders only learning to map the parameters from the centred galaxy into the latent space. The equation used for the deblenders is eq. (7.6).

Training procedure

The training procedure is almost the same as the one presented in section 7.2.1 and section 7.3. The difference is that here, the VAE and the DEVAE are first trained on noiseless images of isolated galaxies and not on noisy images. This allows the encoder part to learn very accurate prior without being disturbed by the noise.

The idea of learning the galaxy image prior from noisy images of isolated galaxies came from the will of being able to learn from real data. It avoids denoising data which is not a simple task, and might entail a loss of information and impact galaxy shapes. Here the hypothesis is different: it is assumed that the true distribution of galaxy shapes and fluxes is fully represented by the models used to generate the noiseless images. This is true in our case since the **Extragalactic catalog** is used as input of the noiseless image generation. In the case of real data it would obviously be wrong and two options could be explored to build the training set:

- using more realistic galaxy image simulations such as *Illustris* (Vogelsberger et al., 2014) for example, could provide noiseless images of galaxies with complex shapes. They could be used to learn a prior for isolated galaxy images which includes more diverse and complex shapes than the ones from the DC2 images,

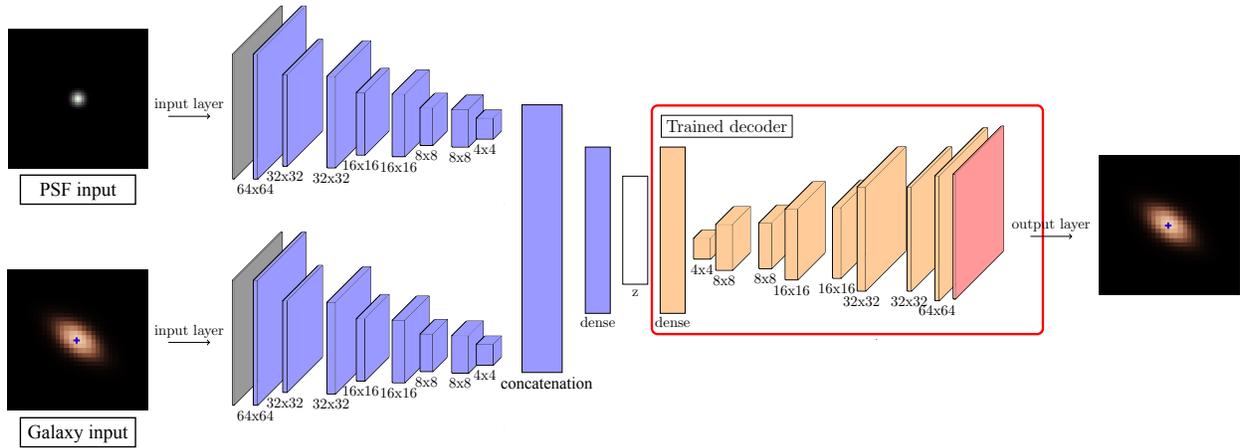


Figure 7.22: Architecture of the DEVAE for DC2 images. Two encoders are working in parallel, one for the galaxy image and the other one for the corresponding PSF image. Both images are in the six LSST bands. The outputs of these two encoders are concatenated, and then flattened to feed a dense layer.

- using a denoising algorithm (for example Starck, Donoho, and Candès, 2003, or more recently with deep learning, Vojtekova et al., 2021) on true images of isolated galaxies in order to obtain a training set. Variational autoencoders have also been used as denoiser in the literature and could maybe also be used in the context of astronomical images.

Comparing both approaches, i.e. learning from noiseless images and denoising DC2 images, could be interesting and is reserved for future work.

The deblender, on the other hand, is trained similarly to what is presented in section 7.3.3. It takes the galaxy image extracted from the DC2 simulation (and the corresponding PSF as input for the DEVAE deblender), and tries to reproduce the noiseless image of the centred galaxy. The only difference, as explained before, is that this target galaxy is not always the brightest on the image.

7.8.4 Results

The results presented in this section are preliminary and more refinements might be required. The first subsection presents the results of the two architecture on noiseless images of isolated galaxies, allowing to chose between the two. Then, the performance of the deblender are presented.

Prior for single galaxy

Similarly to section 7.4.2, I present here the statistical properties of errors on the test sample. On fig. 7.23, the error on the absolute ellipticity $|e| = \sqrt{e_1^2 + e_2^2}$ and on the r -band magnitude are shown splitting the test sample in bins of r -band magnitude of the noiseless target galaxy and of input absolute ellipticity. Again, boxes show the median and percentiles corresponding to $\pm 1\sigma$ of a gaussian, the whiskers show $\pm 2\sigma$ percentiles.

Distribution of ellipticity errors are centred around zero with small deviations and the median is contained with ± 0.01 . As expected, the spread increases as the magnitude increases, objects

becoming fainter. The network not taking into account the PSF produces on average galaxies rounder than expected at low magnitudes and the network taking it into account produces more elliptical objects than expected at high magnitudes, struggling with faint objects. As a function of input ellipticity, both networks produces galaxies more elliptical at low ellipticity and performance of the network taking into account the PSF worsen with the increase in ellipticity.

The errors in magnitude are stable across ellipticity range. Slight difference are visible for low magnitude galaxies where the network trained with PSF seems to provide a better reconstruction of the flux at low magnitude but a worse one a high magnitude.

For both ellipticities and magnitudes, the differences in performance between the networks with and without PSF do not permit to conclude about the best architecture. Both provide similarly accurate results. The DEVAE was proposed to verify if it would be able to take into account information coming from the PSF image, learns the convolution relationship between the PSF image and the galaxy image, and hence generalise to images of galaxies convolved with a PSF that it never saw before. Consequently, we ran a test on a sample generated with a fixed PSF with a larger FWHM (full width at half maximum) than the ones used for DC2 (accessible via the `GRCCatalogs`). It shows that both networks perform poorly when fed with images convolved with unknown PSF. Both the VAE and the DEVAE are in fact learning from the galaxy image, which includes the PSF information, but are not able to generalise to unknown models of PSF. The DEVAE is not using the information coming from the provided PSF image. Note that we also tried with more than one dense layer after concatenation (indeed, one layer is probably not enough to take into account the PSF information), or going to Fourier space to perform the deconvolution, founding each time similar results. This is unexpected, and maybe suggesting that something else was wrong in the test of these architectures. More work on these avenues is required. In the following, I stick to the architecture presented in section 7.3 for both the VAE and deblender.

A better way to take into account the PSF could be to add a first layer taking as input the image of the galaxy and the image of the PSF, and performing the deconvolution of the galaxy image by the PSF. The encoder would then learn to map the representation of the deconvolved galaxy image in the latent space. The decoder could then reproduce the deconvolved galaxy image that we would reconvolved with the PSF, using the PSF image before, at the last layer of the network. This would, in theory allow the network to be insensitive to the PSF model.

These results lead to the conclusion that the VAE is able to learn a tight prior on both shape and magnitude distribution for isolated galaxy images. Comparing with the results presented in section 7.4.2, the spread and mean or the errors in both ellipticity an magnitude seem to be smaller. This could indicate that learning from noiseless images allows to yield a tighter prior than learning from noisy ones. However, the results presented here are obtained for galaxies with magnitude in the r -band below 24.5, which could explain the increase in performance. More work to test these models on a sample of galaxies selected with magnitude in the r -band below 26.5 for example, is necessary to conclude. Comparison with a prior model trained on images denoised with denoising algorithms might be also an interesting avenue to explore. Another solution to conclude on this could be to learn the single galaxy image prior from noisy images of isolated galaxies extracted from DC2, discarding images with a blendedness above a fixed threshold. This will be explored in future work.

Deblender

In the following, the deblender algorithm is called **DebVAd**er, for **De**blending with **Variational Autoencoder**.

The performance of DebVAd

er is presented on fig. 7.24. It shows the distribution of errors on

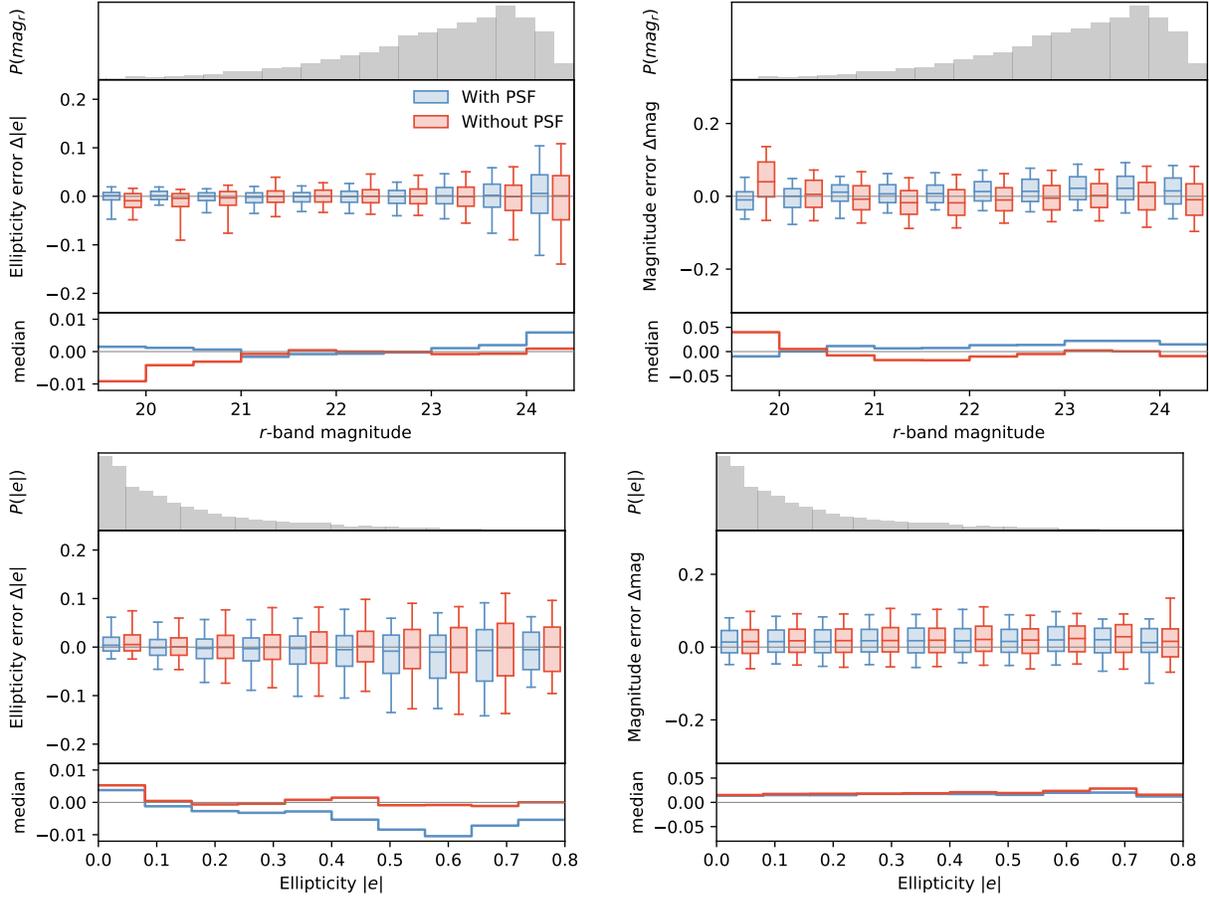


Figure 7.23: Distributions of differences in absolute observed ellipticity $|e|$ (left) and r -band magnitude (right) between target and output images from the test sample. Results for both VAE architectures with (in blue) and without (in red) taking into account the PSF. Networks are trained on the six LSST bands. The distributions are computed on subsets of the test sample split in bins of r -band magnitude (top) and input absolute ellipticity $|e|$ (bottom). In each plot, the top panel shows the distribution over the test sample of the quantity on the x-axis, the middle panel shows the distribution of errors when the test sample is split in bins of this particular quantity, and the bottom panel shows the median error. In the middle panel, the boxes show the median and $\pm 1\sigma$ percentiles and the whiskers show the $\pm 2\sigma$ percentiles.

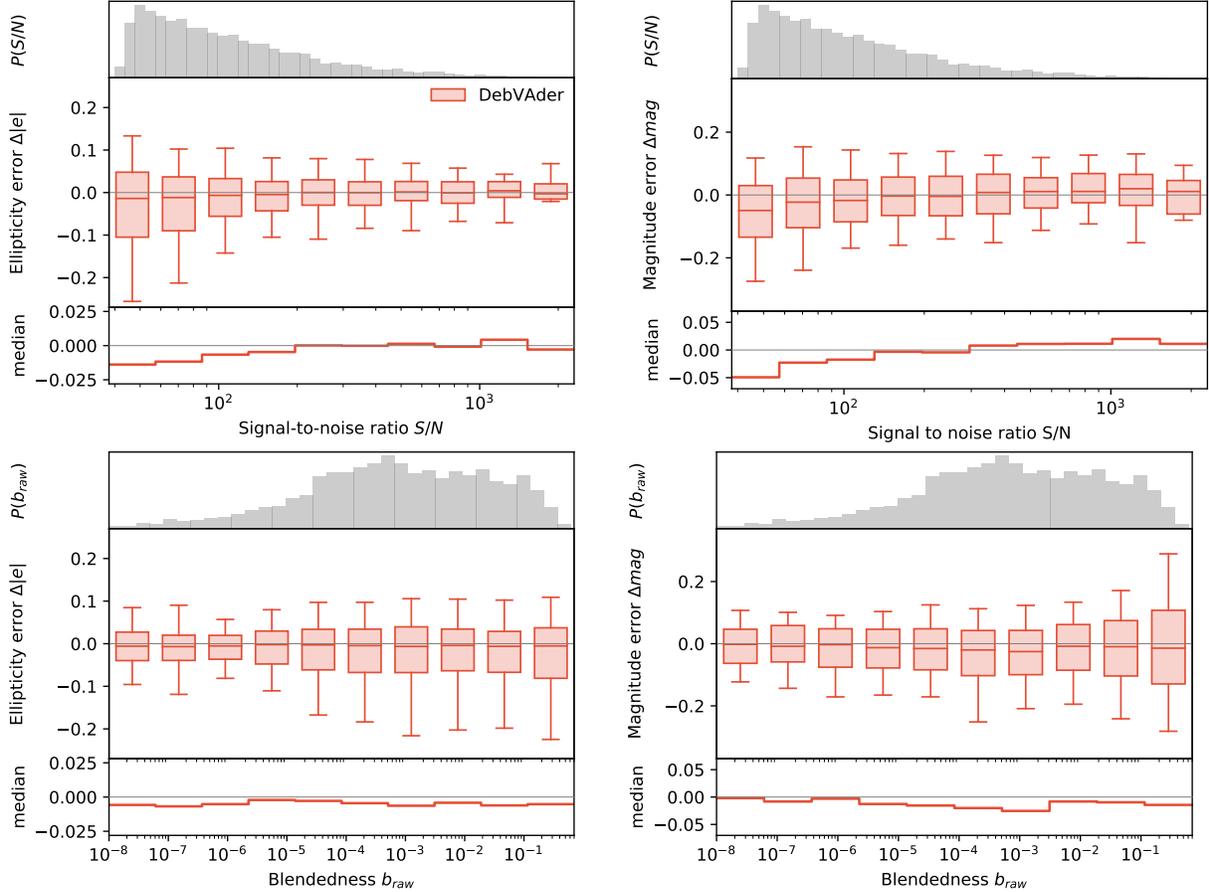


Figure 7.24: Distributions of differences in absolute observed ellipticity $|e|$ (left) and r -band magnitude (right) between target and output images from the test sample. Results for DebVADER trained on the six LSST bands. The distributions are computed on subsets of the test sample split in bins of signal-to-noise ratio S/N (top) and blendedness (bottom).

target galaxy ellipticity and magnitude as a function of signal-to-noise ratio, S/N , and blendedness, b_{raw} , as defined in section 3.1.2. Results show median errors within ± 0.025 and ± 0.05 for ellipticity and magnitude respectively. As expected, the largest spread and median errors appear at low SNR for both ellipticity and magnitude errors. The underestimation of the ellipticity of faint galaxies can be explained by the blurring of the image due to noise. As in section 7.4.2, the network produces a slightly rounder galaxy image, closer to the average image. Similar underestimation is observed for the flux, the network probably confusing part of the galaxy’s flux with noise.

Interestingly, contrary to the spread that increases with the blendedness, the median error in both ellipticity and magnitude is not impacted by this parameter. This means that DebVADER performs as well on an isolated galaxy as on a blended scene.

Finally, comparing the results to section 7.4.3, both median errors and spreads are at a similar order of magnitude. The initial assumption to deblend the brightest galaxy on the image is consequently not necessary to yield accurate shape and flux reconstruction.

Comparison with LSST pipeline measurement

The LSST pipeline provides an ellipticity measurement on the detected galaxies. This measurement is realised after a deblending step using an improved version of the PHOTO algorithm as presented in Bosch et al. (2018) and section 3.3.2. It is done in the *reference band*, which differs for each object and is chosen as a function of whether or not the object is detected in this band, the SNR and the success of the PSF and photometry algorithms. Bosch et al. (2018) indicates that for HSC, the *i*-band is favoured at 57%. In the final shear analysis they use the *i* band for every galaxy with a SNR larger than 10 in this band (Mandelbaum et al., 2018a). Consequently, the shape measurement is performed in the *i*-band for the sake of comparison with the LSST pipeline measurement (it is performed in the *r*-band in the other sections).

Until now, the ellipticity definition used was $\epsilon = \frac{1-\frac{b}{a}}{1+\frac{b}{a}} \exp(2i\varphi)$, since the KSB implementation in GalSim, used in the processing for shape reconstruction, uses this definition. The LSST pipeline uses the REGAUSS implementation of the GalSim package (Bosch et al., 2018). On fig. 7.25 the comparison of the results obtained by the LSST pipeline are compared with the results obtained by the networks presented previously are shown, using the REGAUSS implementation of GalSim in the results processing for shape reconstruction. It uses the second ellipticity definition, $e = \frac{1-(\frac{b}{a})^2}{1+(\frac{b}{a})^2} \exp(2i\varphi)$, which explains the differences in median errors and spreads magnitude. Similarly to section 7.4.2, the ellipticity error is defined as the difference between the measurement obtained from the pipeline and the measurement obtained using the REGAUSS implementation in GalSim on the noiseless image of the galaxy.

The results show that DeVADER performs better than the LSST pipeline deblender, in median errors and in spread. This is particularly true at low SNR and for large blendedness, where galaxies are considered to be more elliptical than they are in reality, by the pipeline. At low SNR, the increase in performance probably comes from the fact that DeVADER also act as a denoising algorithm, allowing for an easier shape measurement. On the complete test sample, it improves performance, compared to the LSST pipeline, of 70% and 120% for the median of the distribution of errors in ellipticity as a function of SNR and blendedness respectively.

Note that the REGAUSS shape measurement method used by the LSST pipeline fails on the deblended images in some cases as already pointed out in Bosch et al. (2018). It was the case for less than 1% (24 galaxies) of the galaxies in our test sample. Also, for a slightly larger fraction of the sample, 1.4% (141 images), the REGAUSS shape measurement method gave non physical ellipticity values, i.e. larger than one in absolute value. For all these objects our algorithm performs as well as for any other galaxy image.

7.8.5 Conclusion

As mentioned earlier, the work presented in section 7.8 is in progress and some improvements are required. However, the results presented in this section permit to draw several conclusions.

First both VAE architectures presented in section 7.8.3 are able to learn features of galaxy images from noiseless sample with very high accuracy. It allows for a precise reconstruction of both ellipticities and magnitudes. Median errors and spreads in ellipticities and magnitudes seem to be smaller than when the prior is learned from noisy images (see section 7.4.2). If this result is confirmed with the galaxy sample generated with a *r*-band magnitude cut at 26.5, paths should be explored to generate noiseless training samples, using realistic galaxy image simulations or denoising algorithms.

A test using different PSFs from the ones of the DC2 training sample showed that the DEVAE

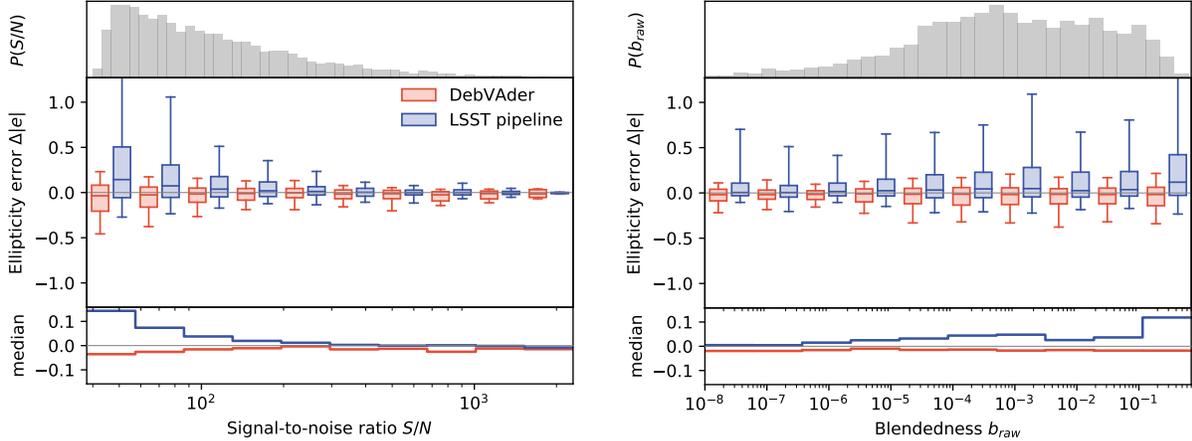


Figure 7.25: Distributions of differences in absolute observed ellipticity $|e|$ between target and output images from the test sample. Results for DebVAdAder are shown (in red), the measurements from the LSST pipeline (in blue). The distributions are computed on subsets of the test sample split in bins of signal-to-noise ratio S/N (left) and blendedness (right).

was unable to take into account information from the PSF image. Both the VAE and DEVAE learn the PSF effect on an image from the galaxy image meaning that augmenting the PSF variation in the training sample would allow for more generalization. Since the DEVAE does not increase performance and is more computationally costly, the VAE architecture should be favoured. As mentioned in section 7.2.1, another avenue to take into account PSF unseen during the training, could be to deconvolve images from the measured PSF and reconvolved with a fixed PSF we learned, before being passed into the deblender network.

Then, DebVAdAder is able to deblend input images recovering the latent space distribution encoding the target galaxy. It shows almost null biases as a function of SNR and blendedness, as well as performance similar to what was obtain in section 7.4.3. This proves that the assumption of deblending first the lowest magnitude galaxy can be relaxed without decreasing the performance level of our method. The performance of our method would allow for the use of all galaxies in the test sample, as they can all be accurately deblended. This is an improvement compared to what was done for the HSC calibration with simulation for example, which discarded all galaxies with a blendedness higher than $10^{-0.375} \sim 0.42$ (Mandelbaum et al., 2018a; Bosch et al., 2018).

Finally DebVAdAder performance are compared to the measurements provided by the LSST pipeline. Figure 7.25 shows that the method presented in this part performs better than the deblender implemented in the LSST pipeline (improved version of the PHOTO algorithm (Bosch et al., 2018)) both at large blendedness as well as at low SNR. The latter is an extra advantage of our method since the networks also act as denoising algorithms.

Note that in order to confirm these results similar tests should be run on galaxy samples generated with a r -band magnitude cut at 26.5, but we reserve this for future work.

7.8.6 Future work

Here I describe what I believe could be the next steps to improve DebVAdAder.

To connect to the training procedure presented above, the GalSim simulation of the noiseless isolated galaxy images might induce limitations. As mentioned earlier, it is not the true galaxy image used as input in the DC2 simulation. One of the main differences being that the PSF used

to convolve the galaxy image comes from the measurement performed by the LSST pipeline, which is not perfect. Differences can be seen by subtracting the target image to the DC2 image: residuals (below the noise) are visible (see fig. 7.21). A procedure to obtain clean targets images from noisy simulated or even real galaxy images should be designed.

Also, the random knots, bright point sources displayed randomly and simulating the star formation regions in a galaxy, were not simulated on the target images and our deblender is not able to deblend them together with their host galaxy. Including random knots on galaxies in the training sample could permit to verify if the color information, contained in these knots, is sufficient for our method to link them to their host galaxy and to deblend them together. To my knowledge, this would be the first deblending algorithm able to extract both the host galaxy and the corresponding random knots.

The approach presented here also allows for the comparison of the performance of the DebVader to other deblenders. Since SCARLET will probably be the preferred deblender in the LSST pipeline it would be interesting to compare performance of both methods on DC2 images and on other datasets. The Blending Tool Kit (see section 3.3.4) is a perfectly suited tool to compare the performance using various metrics. Testing our method on real images, for example from HSC, is another required step to assess the performance of our deblender and validate its architecture.

Finally I did not address the issue raised in section 7.5.3 about the output distribution of the VAE or deblender. Considering the Bernoulli distribution which does not correspond to our data is an abuse that must be fixed. A solution could be to apply a different normalisation on the image allowing for the consideration of another distribution and loss function. Also considering the output as a distribution and not directly as an image could permit to obtain a probabilistic output, with both the mean and standard deviation image, allowing for the inclusion of uncertainties.

Bibliography

- Arcelin, Bastien et al. (Jan. 2021). “Deblending galaxies with variational autoencoders: A joint multiband, multi-instrument approach”. In: *MNRAS* 500.1, pp. 531–547. DOI: [10.1093/mnras/staa3062](https://doi.org/10.1093/mnras/staa3062). arXiv: [2005.12039](https://arxiv.org/abs/2005.12039) [[astro-ph.IM](#)].
- Behroozi, Peter et al. (Sept. 2019). “UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from $z = 0-10$ ”. In: *MNRAS* 488.3, pp. 3143–3194. DOI: [10.1093/mnras/stz1182](https://doi.org/10.1093/mnras/stz1182). arXiv: [1806.07893](https://arxiv.org/abs/1806.07893) [[astro-ph.GA](#)].
- Bosch, James et al. (2018). “The Hyper Suprime-Cam software pipeline”. In: *Publications of the Astronomical Society of Japan* 70.SP1, S5.
- Dawson, William A. et al. (Jan. 2016). “The Ellipticity Distribution of Ambiguously Blended Objects”. In: *ApJ* 816.1, 11, p. 11. DOI: [10.3847/0004-637X/816/1/11](https://doi.org/10.3847/0004-637X/816/1/11). arXiv: [1406.1506](https://arxiv.org/abs/1406.1506) [[astro-ph.CO](#)].
- Dilokthanakul, Nat et al. (Nov. 2016). “Deep Unsupervised Clustering with Gaussian Mixture Variational Autoencoders”. In: *arXiv e-prints*, arXiv:1611.02648, arXiv:1611.02648. arXiv: [1611.02648](https://arxiv.org/abs/1611.02648) [[cs.LG](#)].
- Domínguez Sánchez, H. et al. (Mar. 2019). “Transfer learning for galaxy morphology from one survey to another”. In: *MNRAS* 484.1, pp. 93–100. DOI: [10.1093/mnras/sty3497](https://doi.org/10.1093/mnras/sty3497). arXiv: [1807.00807](https://arxiv.org/abs/1807.00807) [[astro-ph.GA](#)].
- Habib, Salman et al. (Jan. 2016). “HACC: Simulating sky surveys on state-of-the-art supercomputing architectures”. In: *New A* 42, pp. 49–65. DOI: [10.1016/j.newast.2015.06.003](https://doi.org/10.1016/j.newast.2015.06.003). arXiv: [1410.2805](https://arxiv.org/abs/1410.2805) [[astro-ph.IM](#)].
- Higgins, Irina et al. (June 2016). “Early Visual Concept Learning with Unsupervised Deep Learning”. In: *arXiv e-prints*, arXiv:1606.05579, arXiv:1606.05579. arXiv: [1606.05579](https://arxiv.org/abs/1606.05579) [[stat.ML](#)].
- Higgins, Irina et al. (2017). “beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework”. In: *ICLR*.
- Hirata, Christopher and Uroš Seljak (Aug. 2003). “Shear calibration biases in weak-lensing surveys”. In: *MNRAS* 343.2, pp. 459–480. DOI: [10.1046/j.1365-8711.2003.06683.x](https://doi.org/10.1046/j.1365-8711.2003.06683.x). arXiv: [astro-ph/0301054](https://arxiv.org/abs/astro-ph/0301054) [[astro-ph](#)].
- Huff, Eric and Rachel Mandelbaum (Feb. 2017). “Metacalibration: Direct Self-Calibration of Biases in Shear Measurement”. In: *arXiv e-prints*, arXiv:1702.02600, arXiv:1702.02600. arXiv: [1702.02600](https://arxiv.org/abs/1702.02600) [[astro-ph.CO](#)].
- Im, D. et al. (2017). “Denoising Criterion for Variational Auto-Encoding Framework”. In: *AAAI*.
- Ivezić, Željko et al. (May 2008). “LSST: from Science Drivers to Reference Design and Anticipated Data Products”. In: *arXiv e-prints*, arXiv:0805.2366, arXiv:0805.2366. arXiv: [0805.2366](https://arxiv.org/abs/0805.2366) [[astro-ph](#)].
- Jiang, Zhuxi et al. (2017). “Variational Deep Embedding: An Unsupervised and Generative Approach to Clustering”. In: *IJCAI*.

- Johnson, N.L., S. Kotz, and N. Balakrishnan (1995). *Continuous univariate distributions*. Wiley series in probability and mathematical statistics: Applied probability and statistics vol. 2. Wiley & Sons. ISBN: 9780471584940. URL: <https://books.google.fr/books?id=0QzvAAAAMAAJ>.
- Kaiser, Nick, Gordon Squires, and Tom Broadhurst (Aug. 1995). “A Method for Weak Lensing Observations”. In: *ApJ* 449, p. 460. DOI: [10.1086/176071](https://doi.org/10.1086/176071). arXiv: [astro-ph/9411005](https://arxiv.org/abs/astro-ph/9411005) [[astro-ph](https://arxiv.org/abs/astro-ph)].
- Kingma, Diederik P. and M. Welling (2014). “Auto-Encoding Variational Bayes”. In: *CoRR* abs/1312.6114.
- Korytov, Danila et al. (Dec. 2019). “CosmoDC2: A Synthetic Sky Catalog for Dark Energy Science with LSST”. In: *ApJS* 245.2, 26, p. 26. DOI: [10.3847/1538-4365/ab510c](https://doi.org/10.3847/1538-4365/ab510c). arXiv: [1907.06530](https://arxiv.org/abs/1907.06530) [[astro-ph](https://arxiv.org/abs/astro-ph).[CO](https://arxiv.org/abs/astro-ph)].
- Larsen, Anders Boesen Lindbo et al. (2016). “Autoencoding beyond pixels using a learned similarity metric”. In: *International conference on machine learning*. PMLR, pp. 1558–1566.
- Loaiza-Ganem, Gabriel and John P Cunningham (2019). “The continuous Bernoulli: fixing a pervasive error in variational autoencoders”. In: *Advances in Neural Information Processing Systems*, pp. 13287–13297.
- LSST Dark Energy Science Collaboration (LSST DESC) et al. (Mar. 2021). “The LSST DESC DC2 Simulated Sky Survey”. In: *ApJS* 253.1, 31, p. 31. DOI: [10.3847/1538-4365/abd62c](https://doi.org/10.3847/1538-4365/abd62c). arXiv: [2010.05926](https://arxiv.org/abs/2010.05926) [[astro-ph](https://arxiv.org/abs/astro-ph).[IM](https://arxiv.org/abs/astro-ph)].
- Mandelbaum, Rachel et al. (Jan. 2012). *COSMOS real galaxy dataset*. Zenodo. DOI: [10.5281/zenodo.3242143](https://doi.org/10.5281/zenodo.3242143). URL: <https://doi.org/10.5281/zenodo.3242143>.
- Mandelbaum, Rachel et al. (May 2014). “The Third Gravitational Lensing Accuracy Testing (GREAT3) Challenge Handbook”. In: *ApJS* 212.1, 5, p. 5. DOI: [10.1088/0067-0049/212/1/5](https://doi.org/10.1088/0067-0049/212/1/5). arXiv: [1308.4982](https://arxiv.org/abs/1308.4982) [[astro-ph](https://arxiv.org/abs/astro-ph).[CO](https://arxiv.org/abs/astro-ph)].
- Mandelbaum, Rachel et al. (Jan. 2018a). “The first-year shear catalog of the Subaru Hyper Suprime-Cam Subaru Strategic Program Survey”. In: *PASJ* 70, S25, S25. DOI: [10.1093/pasj/psx130](https://doi.org/10.1093/pasj/psx130). arXiv: [1705.06745](https://arxiv.org/abs/1705.06745) [[astro-ph](https://arxiv.org/abs/astro-ph).[CO](https://arxiv.org/abs/astro-ph)].
- Mandelbaum, Rachel et al. (Dec. 2018b). “Weak lensing shear calibration with simulations of the HSC survey”. In: *MNRAS* 481.3, pp. 3170–3195. DOI: [10.1093/mnras/sty2420](https://doi.org/10.1093/mnras/sty2420). arXiv: [1710.00885](https://arxiv.org/abs/1710.00885) [[astro-ph](https://arxiv.org/abs/astro-ph).[CO](https://arxiv.org/abs/astro-ph)].
- Mao, Yao-Yuan et al. (Feb. 2021). “The SAGA Survey. II. Building a Statistical Sample of Satellite Systems around Milky Way-like Galaxies”. In: *ApJ* 907.2, 85, p. 85. DOI: [10.3847/1538-4357/abce58](https://doi.org/10.3847/1538-4357/abce58). arXiv: [2008.12783](https://arxiv.org/abs/2008.12783) [[astro-ph](https://arxiv.org/abs/astro-ph).[GA](https://arxiv.org/abs/astro-ph)].
- Melchior, P. et al. (July 2018). “SCARLET: Source separation in multi-band images by Constrained Matrix Factorization”. In: *Astronomy and Computing* 24, 129, p. 129. DOI: [10.1016/j.ascom.2018.07.001](https://doi.org/10.1016/j.ascom.2018.07.001). arXiv: [1802.10157](https://arxiv.org/abs/1802.10157) [[astro-ph](https://arxiv.org/abs/astro-ph).[IM](https://arxiv.org/abs/astro-ph)].
- Nibali, Aiden et al. (Jan. 2018). “Numerical Coordinate Regression with Convolutional Neural Networks”. In: *arXiv e-prints*, arXiv:1801.07372, arXiv:1801.07372. arXiv: [1801.07372](https://arxiv.org/abs/1801.07372) [[cs](https://arxiv.org/abs/cs).[CV](https://arxiv.org/abs/cs)].
- Norlander, Erik and Alexandros Sopasakis (Nov. 2019). “Latent space conditioning for improved classification and anomaly detection”. In: *arXiv e-prints*, arXiv:1911.10599, arXiv:1911.10599. arXiv: [1911.10599](https://arxiv.org/abs/1911.10599) [[cs](https://arxiv.org/abs/cs).[LG](https://arxiv.org/abs/cs)].
- Refregier, A. et al. (Jan. 2010). “Euclid Imaging Consortium Science Book”. In: *arXiv e-prints*, arXiv:1001.0061, arXiv:1001.0061. arXiv: [1001.0061](https://arxiv.org/abs/1001.0061) [[astro-ph](https://arxiv.org/abs/astro-ph).[IM](https://arxiv.org/abs/astro-ph)].
- Rowe, Barnaby et al. (Apr. 2015). “GALSIM: The modular galaxy image simulation toolkit”. In: *Astronomy and Computing* 10, pp. 121–150. DOI: [10.1016/j.ascom.2015.02.002](https://doi.org/10.1016/j.ascom.2015.02.002). arXiv: [1407.7676](https://arxiv.org/abs/1407.7676) [[astro-ph](https://arxiv.org/abs/astro-ph).[IM](https://arxiv.org/abs/astro-ph)].
- Sánchez, F. Javier et al. (Jan. 2020). “The LSST DESC Data Challenge 1: Generation and Analysis of Synthetic Images for Next Generation Surveys”. In: *arXiv e-prints*, arXiv:2001.00941, arXiv:2001.00941. arXiv: [2001.00941](https://arxiv.org/abs/2001.00941) [[astro-ph](https://arxiv.org/abs/astro-ph).[IM](https://arxiv.org/abs/astro-ph)].

- Schuhmann, Robert L., Catherine Heymans, and Joe Zuntz (Jan. 2019). “Galaxy shape measurement synergies between LSST and Euclid”. In: *arXiv e-prints*, arXiv:1901.08586, arXiv:1901.08586. arXiv: [1901.08586](#) [[astro-ph.IM](#)].
- Sheldon, Erin S. and Eric M. Huff (May 2017). “Practical Weak-lensing Shear Measurement with Metacalibration”. In: *ApJ* 841.1, 24, p. 24. DOI: [10.3847/1538-4357/aa704b](#). arXiv: [1702.02601](#) [[astro-ph.CO](#)].
- Sheldon, Erin S. et al. (Oct. 2020). “Mitigating Shear-dependent Object Detection Biases with Metacalibration”. In: *ApJ* 902.2, 138, p. 138. DOI: [10.3847/1538-4357/abb595](#). arXiv: [1911.02505](#) [[astro-ph.CO](#)].
- Sønderby, Casper Kaae et al. (2016). “Ladder variational autoencoders”. In: *Advances in neural information processing systems*, pp. 3738–3746.
- Springenberg, Jost Tobias et al. (Dec. 2014). “Striving for Simplicity: The All Convolutional Net”. In: *arXiv e-prints*, arXiv:1412.6806, arXiv:1412.6806. arXiv: [1412.6806](#) [[cs.LG](#)].
- Starck, J. L., D. L. Donoho, and E. J. Candès (Feb. 2003). “Astronomical image representation by the curvelet transform”. In: *A&A* 398, pp. 785–800. DOI: [10.1051/0004-6361:20021571](#).
- The LSST Dark Energy Science Collaboration et al. (Sept. 2018). “The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document”. In: *arXiv e-prints*, arXiv:1809.01669, arXiv:1809.01669. arXiv: [1809.01669](#) [[astro-ph.CO](#)].
- Vogelsberger, M. et al. (May 2014). “Properties of galaxies reproduced by a hydrodynamic simulation”. In: *Nature* 509.7499, pp. 177–182. DOI: [10.1038/nature13316](#). arXiv: [1405.1418](#) [[astro-ph.CO](#)].
- Vojtekova, Antonia et al. (May 2021). “Learning to denoise astronomical images with U-nets”. In: *MNRAS* 503.3, pp. 3204–3215. DOI: [10.1093/mnras/staa3567](#). arXiv: [2011.07002](#) [[astro-ph.IM](#)].

Chapter 8

Bayesian neural network for galaxy parameters estimation in blended scenes: ellipticity and redshift measurements

In the previous chapter, we showed that an encoder, implemented as presented in section 7.8.3, was able to encode, into the latent space, all the information required for the reproduction of the image of the centred galaxy on a blended and noisy scene. However, to perform weak lensing analysis, only the shape and redshift of this galaxy are required. In this part, we explore how these parameters can be estimated by a neural network on blended scenes, without requiring deblending.

This work is ongoing and results presented in this chapter are preliminary. It is realised in collaboration with Cyrille Doux, and my supervisors Eric Aubourg and Cécile Roucelle.

8.1 Introduction

As we have seen in chapter 2, to perform weak lensing analysis, we need to measure ellipticities and redshifts of galaxies. However, in future surveys like LSST, blending will be a major issue. It will dominate the systematic error budget for cosmic shear since it can introduce many biases at different steps of the analysis, as described in chapter 3. The common solution to deal with a blended scene nowadays, is to use a deblending algorithm in order to separate the different sources on the image, before performing measurements. This approach is the base of the work presented in the previous chapter. Deblending algorithms can be very efficient but they are also likely to entail a loss of information when extracting the galaxy image, and might introduce biases (see section 7.4.3). To avoid these undesired effects, this work explores a new avenue: measuring the weak lensing parameters (ellipticities and redshift) directly from the images, skipping the deblending step.

Doing these measurements from images of blended scene is particularly difficult for state-of-the-art methods. The multi-object fitting algorithm used by the DES collaboration (see section 3.3.2 and Drlica-Wagner et al., 2018), manages to realise both deblending and shape measurement (via the model fitting) simultaneously but, as explained in section 2.5.3, model-fitting methods are likely to introduce model bias in the cosmic shear analysis. Concerning moment-based methods, a measurement of galaxy ellipticities from a blended object seems unlikely since it relies on the light profile of the image. A deblending step is needed before measurement in that case.

In this work we propose a deep learning algorithm able to perform galaxy parameters estimation from images of isolated or blended galaxies, without requiring a deblending step. The method is inspired from DebVader and from the work presented in chapter 7. We have shown that an encoder, implemented as presented in section 7.8.3, was able, from a blended scene, to extract and encode into the latent space, all the information required for the reproduction of an isolated galaxy image. In this chapter, we use an encoder-like neural network, i.e. a convolutional neural network (CNN), to learn the measurement of weak lensing relevant galaxy parameters from images. Using a CNN allows to learn arbitrarily complex shape and SED models, potentially reducing model bias. It naturally generalises to multi-band, multi-instrument data, which could be particularly useful in a potential joint pixel analysis of LSST and Euclid data.

Also, to use a neural network in a science analysis pipeline, it is necessary to provide a measurement of the uncertainty coming from the model, mostly due to the representativeness of the training sample, called the epistemic uncertainty. As we have seen in chapter 5, Bayesian neural networks (BNN) can provide an estimation of this uncertainty (see section 5.1). Consequently, we apply a BNN to galaxy parameters estimation, and explore how the measurement of epistemic uncertainty could be used in a cosmological analysis.

In this work, we mainly focus on galaxy shape parameters, but we will show in section 8.5, that this algorithm can also perform a precise joint measurement of both ellipticities and redshift of a galaxy, providing a full posterior distribution for each of these parameters. This is the first direct simultaneous measurement of both ellipticities and redshift, the required parameters for weak lensing analysis. It also shows that this method can be extended to other galaxy parameters.

This chapter is organised as follows. First, we show that galaxy parameters can be estimated from the latent space of a trained deblender using dense layers. Then, we investigate the possibility of estimating these parameters directly with a CNN, we present a first implementation of a Bayesian neural network, which provides an estimation of the epistemic uncertainty, and show the benefits of its utilisation. Finally, we show that this algorithm is able to learn to measure both ellipticities and redshift, the parameters relevant for weak lensing, simultaneously.

8.2 Simulated images: DC2 simulation

In this chapter, we use the same training, validation and test samples as the ones presented in section 7.8.2. However, the targets (or labels) are different. Here, the networks aim at measuring the lensed ellipticities, e_1 and e_2 , and the redshift, z .

In the first part of this analysis (until section 8.5), only the ellipticities are measured. The network's labels are consequently the lensed ellipticities. They are not provided by the **Extragalactic catalog** but the unlensed ellipticities, the two shear components, the convergence and the redshift are available. From eq. (2.34), the lensed ellipticities are then computed for each target galaxy as:

$$\epsilon_1 = \text{Re} \left(\frac{\epsilon^{int} + g}{1 + g^* \epsilon^{int}} \right) \approx \left(\frac{\epsilon^{int} + \gamma}{1 + \gamma^* \epsilon^{int}} \right) \quad , \quad \epsilon_2 \approx \text{Im} \left(\frac{\epsilon^{int} + \gamma}{1 + \gamma^* \epsilon^{int}} \right) \quad (8.1)$$

As explained in section 7.8.4, the **Extragalactic catalog** provides an ellipticity defined following the second part of eq. (2.33), $\epsilon = \frac{1-b/a}{1+b/a} \exp 2i\varphi$, but the HSM ellipticity measurement provided by the LSST pipeline uses the other ellipticity definition, $e = \frac{1-(b/a)^2}{1+(b/a)^2} \exp 2i\varphi$.

To compare the measurements from the networks, to the HSM measurements, we use the second definition to define the target ellipticities. The amplitude of the ellipticity given by the

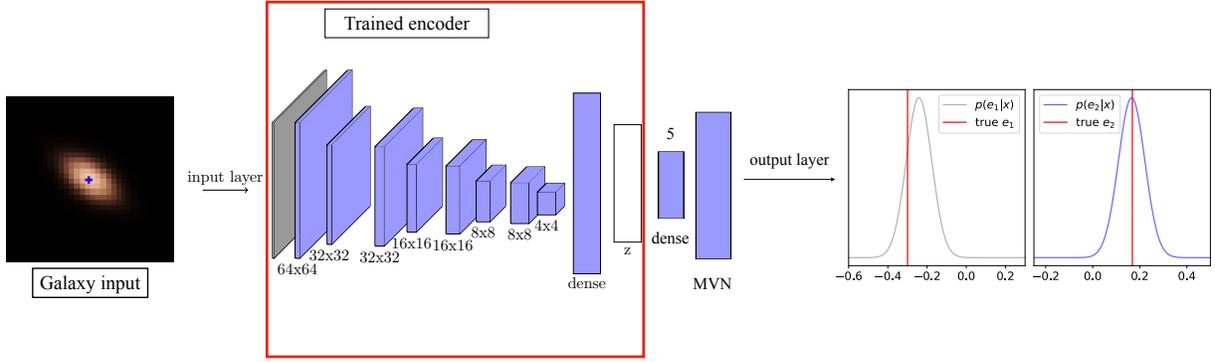


Figure 8.1: Network architecture for parameter estimation from the latent space of a trained VAE. The trained encoder is extracted from the VAE and the weights are fixed. One dense layer and an Multivariate normal distribution layer (MVN) are added to perform the parameter estimation. In the case of ellipticity measurements, the MVN outputs two values for the mean vector, (μ_{e_1}, μ_{e_2}) , and three for the covariance matrix Σ .

Extragalactic catalog needs to be modified following:

$$e = \frac{2\epsilon}{1 + \epsilon^2} \quad (8.2)$$

In section 8.5, we focus on the addition of redshift as a new parameter to measure. True redshifts can be extracted from `Extragalactic catalog` and used as targets.

Images are fed to the network similarly to what is described in section 7.8.2, i.e. with data augmentation, and the signs of the labels, the ellipticity parameters e_1 and e_2 , are modified accordingly.

8.3 Parameter estimation from the latent space of a trained deblender

8.3.1 Architecture

The first step to ensure that it is possible to design a neural network able to do galaxy parameters estimation from blended scenes, is to check if those parameters can be estimated from the encoder and latent space of the deblender trained in section 7.8.4. Indeed, once fed with an image, the encoder is suppose to encode all the required parameters for the generation of the centred galaxy on the image into the latent space. The galaxy ellipticity parameters, for example, should be contained within it.

The network is composed of the encoder and latent space of the deblender as input, with the corresponding weights which are fixed during training. One dense layer is implemented after the latent space, and the network outputs a multivariate normal distribution (MVN) of dimension two, to predict the two ellipticity components (see fig. 8.1). The dense layer is composed 5 neurons, two for the mean of the MVN, (μ_{e_1}, μ_{e_2}) , and three for the covariance matrix Σ .

Considering an input image x , whose complex ellipticities are denoted y , and w the weights of the neural network, the network output is a MVN: $\mathcal{N}(\mu(x, w), \Sigma(x, w))$. The loss function of this

network is defined as the negative log likelihood:

$$\mathcal{L} = -\log(p(y|x, w)) \quad (8.3)$$

with $(x, y) \in D$ the training data. This neural network is a point estimate (or deterministic) neural network and consequently the covariance of the distribution characterises only the aleatoric uncertainty, coming from the input data (see chapter 5).

Only the dense layer and the MVN layer are trained to extract from the latent space, the necessary information for the prediction of the two ellipticity parameters posterior distributions, $p(e_i|D)$, $i \in \{1; 2\}$.

8.3.2 Results

In this section, we present the results for shape estimation obtained with the network described above, and compare them with the measurements from the LSST pipeline, and using the DebVAdAder algorithm presented in section 7.8.4. Thus, here I compare the results from a network that performs ellipticity measurement on the image, without deblending, to results that used a deblending step followed by a shape measurement step, using the REGAUSS implementation of the GalSim package (Bosch et al., 2018).

The ellipticity error, $\Delta|e|$, is defined as the difference between the absolute value of the measured ellipticity and the absolute value of the true lensed ellipticity, computed from the `Extragalactic catalog`. This is different from what was shown in the previous chapter, and this is why the results from the LSST pipeline and from the DebVAdAder might appear slightly dissimilar. To compute the error for the network, I considered the measured ellipticity to be the mean of the MVN distribution.

Comparison with the LSST pipeline and DebVAdAder

The errors on the absolute ellipticity $|e| = \sqrt{e_1^2 + e_2^2}$ are shown splitting the test sample in bins of SNR and of blendedness, as defined in section 3.1.2, and extracted from the `Object catalog`. Again, boxes show the median and percentiles corresponding to $\pm 1\sigma$ of a Gaussian, the whiskers show $\pm 2\sigma$ percentiles.

The comparison of the performance of this network, the LSST pipeline and using the DebVAdAder are shown on fig. 8.2. Since the deblender presented in section 7.8.4 shows better results than the LSST pipeline, we expect the same from the shape measurement network presented her. Indeed, we kept the weights of the encoder fixed and equal to the ones of DebVAdAder. Consequently we expect that the information required for the generation of the centred galaxy image, on the scene fed to the network, is correctly encoded within the latent space. Since measuring two values from the latent space is expected to be a simpler task than generating an galaxy image from it, we expect the results from the network to be at least as good as the ones from the DebVAdAder.

Indeed, it provides more accurate ellipticity measurements, both as a function of SNR and blendedness, than the LSST pipeline. It follows almost exactly the results obtained by the DebVAdAder+REGAUSS method. On the complete test sample, the network decreases the median of the distribution of errors in ellipticity measurement by more than a factor of 6 and 4 compared to the LSST pipeline, respectively as a function of blendedness and SNR.

It performs as well as using the DebVAdAder and REGAUSS, showing slightly better performance if the results are binned as a function of SNR, and slightly worse if they are binned as a function of blendedness. Consequently, we managed to implement a shape measurement neural network that is able to perform as good as a combination of a deblender using generative models, and a state-of-the-art shape measurement method.

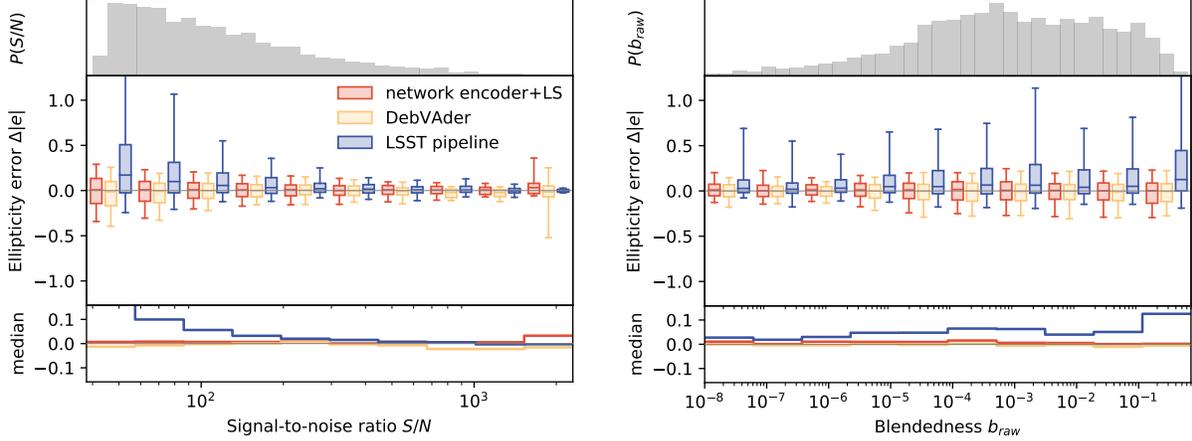


Figure 8.2: Distributions of differences in absolute ellipticity $|e|$ comparing the measurements on DC2 images to the true lensed ellipticities computed from the `Extragalactic catalog`. Comparison of the results from our network using the trained encoder and latent space of DebVAdler, called "network encoder+LS" here, (in red), using DebVAdler and the `REGAUSS` implementation of `GalSim` (in yellow), and from the LSST pipeline (in blue). The distributions are computed on subsets of the test sample split in bins of signal-to-noise ratio S/N (left) and blendedness (right). In each plot, the top panel shows the distribution over the test sample of the quantity on the x-axis, the middle panel shows the distribution of ellipticity errors when the test sample is split in bins of this particular quantity, and the bottom panel shows the median error. In the middle panel, the boxes show the median and $\pm 1\sigma$ percentiles and the whiskers show the $\pm 2\sigma$ percentiles.

Note that the `REGAUSS` shape measurement method used by the LSST pipeline fails, in some cases, on images deblended by the `PHOTO` deblender (as already pointed out in Bosch et al. (2018)). It is the case for only 0.2% (24 galaxies) of our test sample. Also, for a slightly larger fraction of the sample, 1.4% (141 images), the `REGAUSS` shape measurement method gives non physical ellipticity values, i.e. larger than one in absolute value. For all these objects our algorithm performs as well as for any other galaxy image.

Moreover, this deterministic network outputs a MVN distribution, providing a measurement of the aleatoric uncertainty, the uncertainty coming from the data (see top of section 5.1). To verify the behavior of this uncertainty, a first test is to plot it as a function of SNR. This is presented on fig. 8.3 and, as expected, aleatoric uncertainty decreases when SNR increases showing that our network has more difficulties to measure accurately ellipticities in case of noisy data.

To go a step further, when designing the neural network, we have assumed that our measurement could be represented by a Gaussian multivariate distribution. This can be verified, looking at the calibration of the MVN distribution. The network's output can be standardized to see if the resulting distribution follows the standard distribution. We can define, for an image X and a target Y ,

$$Z(X, w) = \Sigma_{pred}(X, w)^{-\frac{1}{2}}(Y - \mu(X, w)) \quad . \quad (8.4)$$

If our hypothesis is correct, the distributions of the two components $z_1 \sim Z(X, w)_1$ and $z_2 \sim Z(X, w)_2$ should follow the standard distribution $\mathcal{N}(0, 1)$. Here Σ_{pred} is the predictive uncertainty, only composed of the aleatoric uncertainty:

$$\Sigma_{pred}(X, w) = \Sigma_{al}(X, w) \quad (8.5)$$

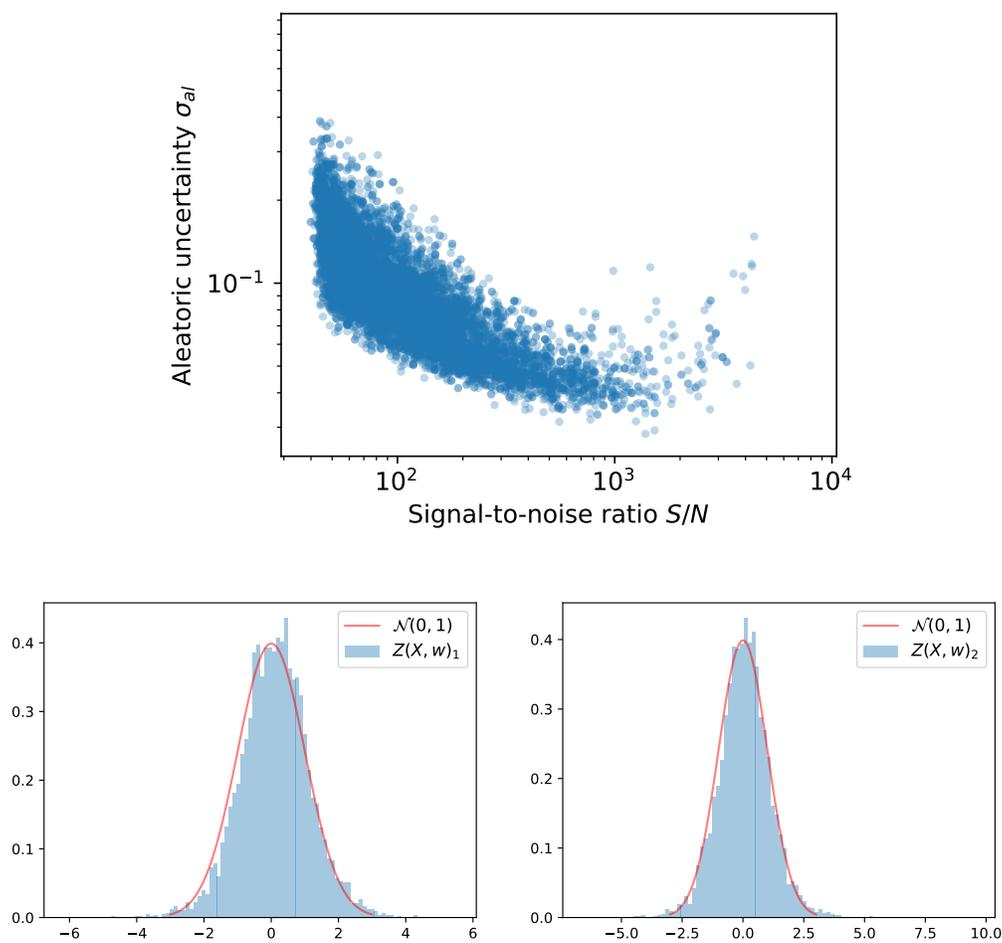


Figure 8.3: (top) Distribution of the standard deviation, σ_{ai} , of the predicted MVN distributions, representing the aleatoric uncertainty of the measurement, as a function of signal-to-noise ratio. As expected the aleatoric uncertainty decreases when SNR increases. (bottom) Standardized distributions of the network’s output, z_1 and z_2 . These distributions follow the normal distribution, showing that the aleatoric uncertainty is well calibrated.

On the bottom of fig. 8.3, we see that the standardized output follow closely the normal distribution, showing that the MVN distribution is well calibrated and that the model neither overestimates or underestimates the predictive uncertainty.

In this section, we showed that a neural network composed of a trained encoder and latent space from a deblender, was able to measure galaxy ellipticity parameters, without requiring deblending, as precisely as using a deblender based on variational autoencoders, and the REGAUSS shape measurement algorithm.

However, this network includes a layer of complexity that might not be necessary. Indeed, the latent space is sampled at each measurement, leading to slightly different results if the network is fed twice with the same image. We saw in section 7.3 that the encoder learns the variational distribution $q_\phi(z|x)$ which is, actually, the multivariate Gaussian latent space. Consequently, the aleatoric uncertainty, shown on the bottom panel of fig. 8.3, is composed of both the aleatoric uncertainty from the encoder, sampled via the latent space, and the one from the two layers network that learns to extract ellipticities from the latent space. This layer of complexity is probably useless, and a single multivariate Gaussian as output of this network should be enough to capture the aleatoric uncertainty. Since parameters are supposed to be encoded in this latent space, the output of the network could be at this depth in the neural network.

In the next section, we use an encoder-like neural network, a CNN, which outputs the measurement at the depth of the latent space in the network presented in this section. This is the architecture of classical CNNs with a MVN distribution layer as output (see fig. 8.4).

8.4 CNN and BNN for galaxy parameters estimation

This section presents the results for ellipticity parameters estimation from a CNN. First, we present the results from a deterministic neural network, i.e. with weights being point estimates ("CNN" in the following). Then we use a Bayesian neural network, which can provide valuable information for weak lensing analyses ("BNN" in the following).

8.4.1 Architecture

Point estimate CNN

Except from the latent space, the architecture of the CNN is similar to what was presented in section 8.3.1. It follows the one from the encoder presented in section 7.3.2. It is composed of 8 convolutional layers, divided into four blocks of two layers. For each block, the first layer has strides of (1, 1) and the second (2, 2). In each block the convolutional layers have respectively 32, 64, 128 and 256 filters and all use 3×3 kernels. Two dense layers of 256 and 5 units are added after the layer that flattens the convolutional output. The output layer of the network is a a MVN distribution layer (see fig. 8.4).

The loss is the same as in section 8.3.1:

$$\mathcal{L} = -\log(p(y|x, w)) \quad , \quad (8.6)$$

with $(x, y) \in D$ the training data. This network, the CNN, is a point estimate neural network.

Bayesian neural network

Instead of point estimates, for a BNN we consider a probability distribution over the weights $p(\theta|D)$, θ being the parameters of the distribution, and D the complete training sample (see chapter 5). It

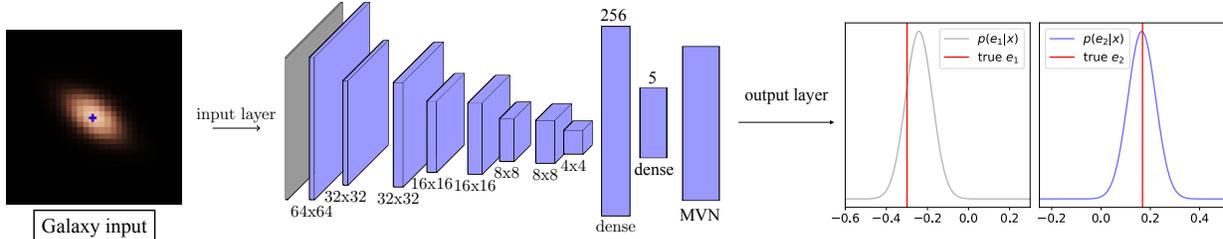


Figure 8.4: Network architecture for parameter estimation without the latent space previously considered. This is a conventional CNN, the latent space is replaced by a 5 neurons dense layer and the MVN distribution layer.

is used to predict the posterior predictive distribution $p(Y|X, D)$, $(X, Y) \in D_{test}$, D_{test} being the test sample, Y the target value of the neural network, here the true lensed ellipticity values, and X the input fed to the network, here an image extracted from DC2.

However, as explained in chapter 5, the posterior distribution $p(\theta|D)$ is often intractable and to approximate this distribution, one can use variational inference (VI). In VI, $p(\theta|D)$ is approximated with a chosen, tractable distribution $q_\phi(\theta|D)$ which depends on a set of parameters ϕ . The loss function (eq. (8.6)) must be modified to take into account the Kullback-Leibler divergence, which measures the loss of information between the two distributions. For $(x, y) \in D$, and similarly to what was presented in section 7.3, the networks maximises the evidence lower-bound (ELBO):

$$\log p(y|x) \geq -D_{\text{KL}}(q_\phi(\theta|x)||p(\theta)) + \mathbb{E}_{q_\phi}(\log p(y|x, \theta)) \quad (8.7)$$

The prior $p(\theta)$, is considered to be a k -tied normal distribution (Swiatkowski et al., 2020), which is a kind of restricted families of distributions, going against the intuitive thinking that more expressive distributions would lead to a better posterior approximation. Gaussian mean-field variational inference (see section 5.3.3 and Blei, Kucukelbir, and McAuliffe (2016)), the usual way of approximating the posterior distribution, implies doubling the number of trainable parameters compared to a point estimate neural network, and can often suffer from over-regularization leading to underfitting (Wenzel et al., 2020). The k -tied normal distribution is a more compact approximation of variational distribution based on an experimental founding that converged posterior standard deviations under Gaussian mean-field VI consistently display a strong low-rank structure. By factorizing the parametrization of posterior standard deviations, k -tied normal distribution allows for equivalent model performance compared to Gaussian mean-field VI, while decreasing the number of trainable parameters and improving signal-to-noise ratio of stochastic gradients, leading to faster convergence.

To learn this approximate posterior distribution we use Flipout (Wen et al., 2018) (see section 5.3.2), which allows for a small variance of the stochastic gradient by perturbing the weights quasi-independently within a mini-batch. This choice was particularly based on the work of Hortúa et al. (2020) who performed tests on different VI methods for cosmological parameters estimation from the Cosmic Microwave Background, and found Flipout to outperform other methods.

Previous work has shown that the epistemic uncertainty is mostly located in the deep layers of a neural network (Gal, Hron, and Kendall, 2017) and that using Bayesian layers only for the deep layers of the network could suffice for a correct estimation of epistemic uncertainty (Theobald, Arcelin et al. (2021)). We follow this idea here by only modifying the two dense layers using Flipout, and convolutional layers are kept as point estimate layers.

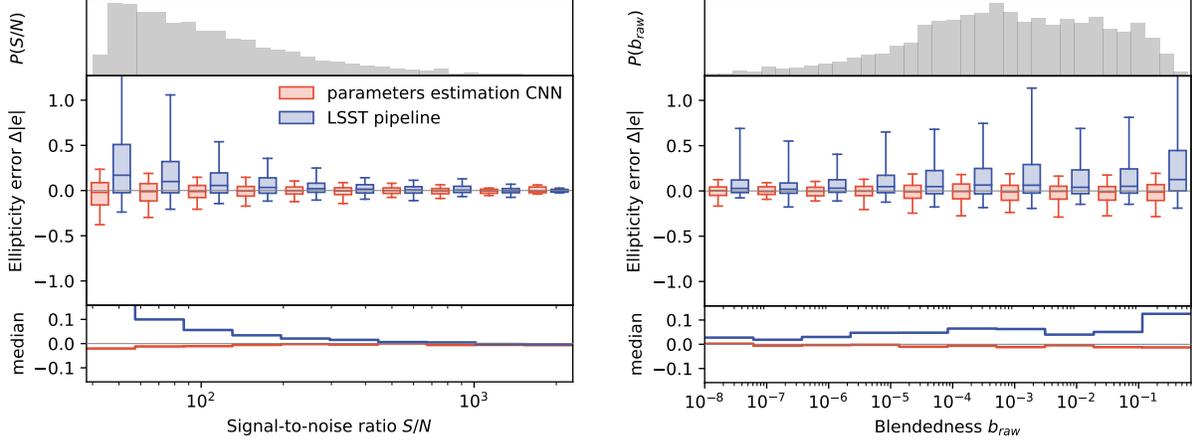


Figure 8.5: Distributions of differences in absolute ellipticity $|e|$ comparing the measurements obtained on DC2 images to the true lensed ellipticities computed from the `Extragalactic catalog`. Comparison of the results from our CNN for parameters estimation (in red), and from the LSST pipeline (in blue). The distributions are computed on subsets of the test sample split in bins of signal-to-noise ratio S/N (left) and blendedness (right).

8.4.2 Results

The results are computed measuring the difference between the measurements and the true lensed ellipticity values calculated from the `Extragalactic catalog`. Unless stated otherwise, the measurements considered for the neural networks presented in this section are the mean of the posterior distribution measured by the networks.

Point estimate CNN

The results obtained with the point estimate CNN, CNN in the following, are presented on fig. 8.5 and compared with the measurements from the LSST pipeline.

Once again, our network performs much better than the measurements from the LSST pipeline with an increase of a factor 8 and 5 in the performance on the median of the distribution of ellipticity errors, respectively as a function of SNR or blendedness. As expected the spread of the errors increases at low SNR and high blendedness. The median of the distribution of errors is slightly biased toward negative errors at low SNR and high blendedness indicating that the network measures rounder galaxies in these cases (see bottom of fig. 8.6). This is expected at low SNR but more surprising at high blendedness where one would expect galaxies to appear more elliptical. It is possible that the network, while extracting the information of the target galaxy, considers the highly blended parts of this object as belonging to the neighbouring galaxies.

However, the results obtained with the CNN are very similar to the ones obtained with the previous network. They are even slightly better if the performance on the median of the distribution of ellipticity errors are compared. Therefore I stick to this architecture in the following. Similarly to the previous network our algorithm performs as well on the galaxies for which the `REGAUSS` shape measurement method fails or gives non-physical measurements, as for any other galaxy image.

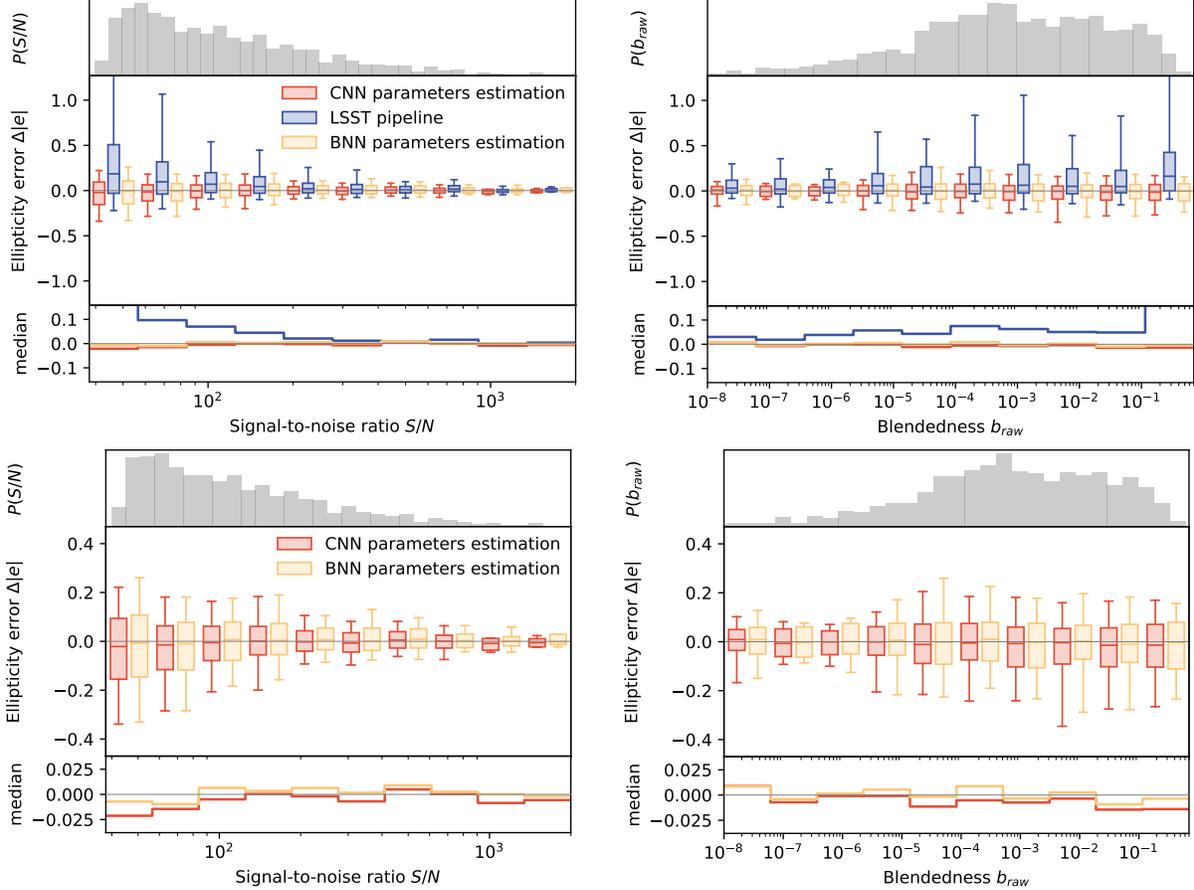


Figure 8.6: Distributions of differences in absolute ellipticity $|e|$ measured on DC2 images and compared to the true lensed ellipticities computed from the **Extragalactic catalog**. (top row) Comparison of the results from our CNN for parameter estimation (in red), from the LSST pipeline (in blue), and from the BNN averaged over 10 samples (yellow). (bottom row) Comparison of the results from the CNN and BNN only, we have reduced the vertical scale on the ellipticity errors, for a better visualisation. The distributions are computed on subsets of the test sample split in bins of signal-to-noise ratio S/N (left) and blendedness (right).

Bayesian neural network

Then, we train a Bayesian neural network, with similar architecture to the CNN, using Flipout on the two dense layers as described in section 8.4.1. In order to have an approximation of the epistemic uncertainty computed by the network, we fed it 10 times with the same 3000 images extracted from DC2. Sampling 10 times the posterior distributions of the weights might not be enough to get a precise estimation of this uncertainty but, as a first step, we restricted ourselves to this number.

First, fig. 8.6 shows the results obtained in terms of errors in ellipticity measurement as a function of SNR and blendedness. For the BNN, we consider the mean of the 10 distribution means measured by the network (for an image X , $\mu(X, w) = \frac{1}{K} \sum_{k=1}^K \mu(X, w_k)$, with $K = 10$) as the ellipticity measurement for a galaxy. This figure presents the results for the ellipticity measurements on 3000 images (from the 10000 images of test sample).

The BNN shows better performance compared to the LSST pipeline in both median and spread

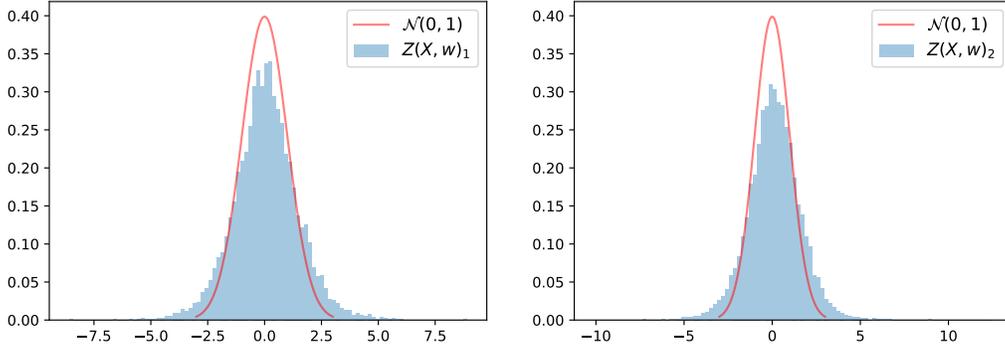


Figure 8.7: Standardized distributions of the network’s output, z_1 and z_2 . These distributions are slightly more spread than the normal distribution, showing the presence of a non-negligible amount of epistemic uncertainty.

of ellipticity error distributions. Again, the spread of the ellipticity errors distribution increases as expected with blendedness and at low SNR. Compared to the CNN, the BNN presents a decrease of the median of ellipticity errors distribution of the order of 50% as a function of both SNR and blendedness. As shown on the bottom row of fig. 8.6, this improvement mostly comes from better ellipticity measurements at low SNR and high blendedness. It can probably be explained by the fact that sampling the weights distribution allows to take into account the epistemic uncertainty from the network, which decreases the impact of large measurement errors. This performance improvement is preliminary and should be confirmed by further work, but it is very promising and shows the benefits of the estimation of epistemic uncertainty for a neural network.

To confirm that epistemic uncertainty plays a role in this measurement, its presence can be seen in the standardised outputs of the network defined as in section 8.3.2. For an image X and target Y ,

$$Z(X, w) = \Sigma_{pred}(X, w)^{-\frac{1}{2}}(Y - \mu(X, w)) \quad , \quad (8.8)$$

with the distributions of the two components $z_1 \sim Z(X, w)_1$ and $z_2 \sim Z(X, w)_2$. Here Σ_{pred} is the predictive uncertainty, computed as the sum of the aleatoric and epistemic uncertainties, for K output samples (here $K = 10$):

$$\begin{aligned} \Sigma_{pred}(X, w) &= \Sigma_{al}(X, w) + \Sigma_{ep}(X, w) \\ &= \frac{1}{K} \sum_{k=1}^K \Sigma(X, w_k) + \sum_{k=1}^K (\mu(X, w_k) - \bar{\mu}(X))(\mu(X, w_k) - \bar{\mu}(X))^T \end{aligned} \quad (8.9)$$

where $\bar{\mu}(X) = \frac{1}{K} \sum_{k=1}^K \mu(X, w_k)$. Figure 8.7 shows that the two distributions are slightly more spread than the normal distribution indicating that the epistemic uncertainty is not negligible.

To test the behavior of this uncertainty we ran a test on one image that we disturbed by two means:

- first we artificially blended the faint central galaxy with two bright galaxies, largely increasing the blendedness (see second row of fig. 8.8),
- then we turned off pixels in a square of size 5×5 on the lower left from the center of the image (see third row of fig. 8.8).

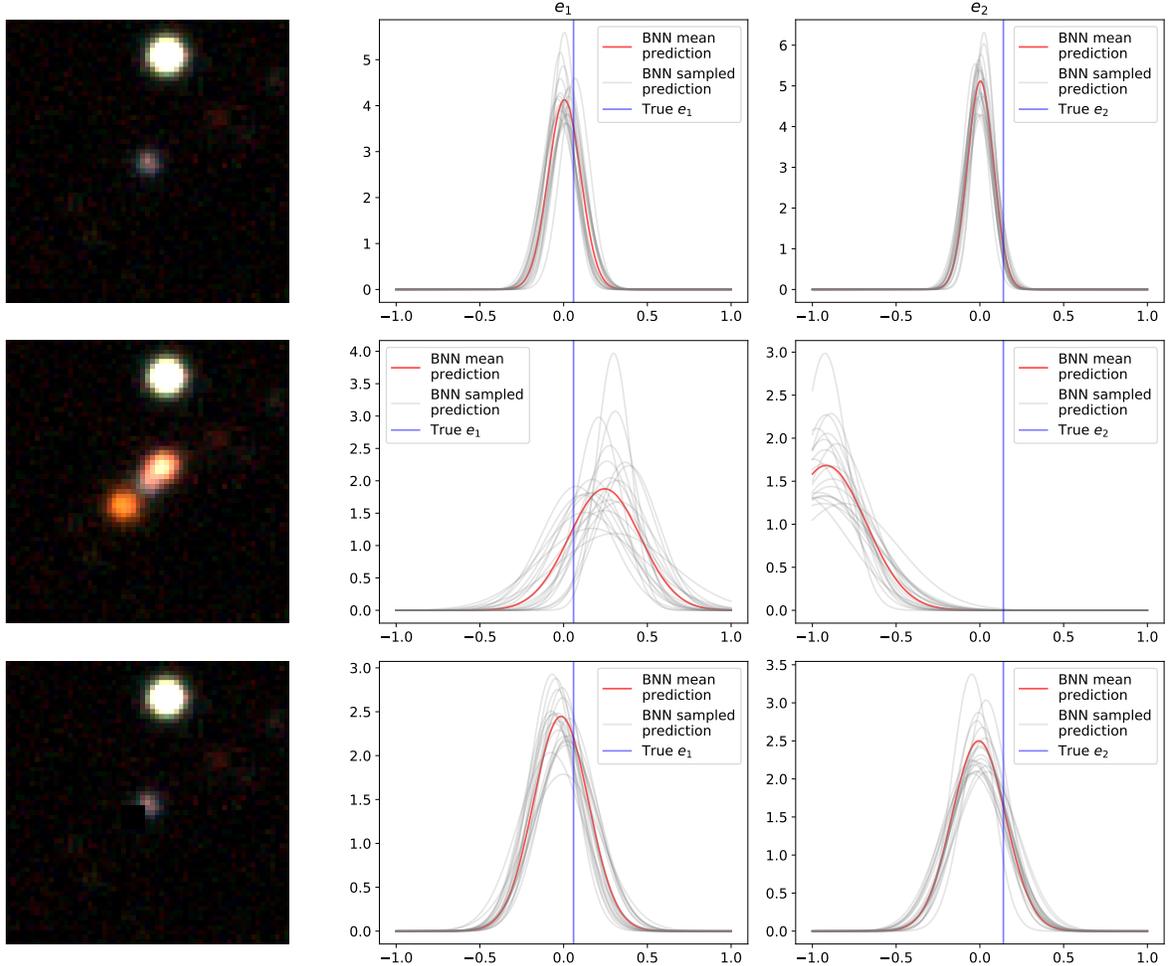


Figure 8.8: Posterior distributions predicted by the BNN in three cases. The first row presents the case of an unperturbed image of a faint galaxy extracted from the DC2 simulation. The second row presents the same image where the central galaxy is blended with two bright galaxies. The third row presents the same image perturbed with turned off pixels in a square of size 5×5 on the lower left from the center of the image. On the center and right columns are shown the measured posterior distributions of e_1 and e_2 respectively. The mean measurement is in red and 30 of the 100 measurements that were performed as shown in grey. The true values of the parameters are shown with a blue vertical line.

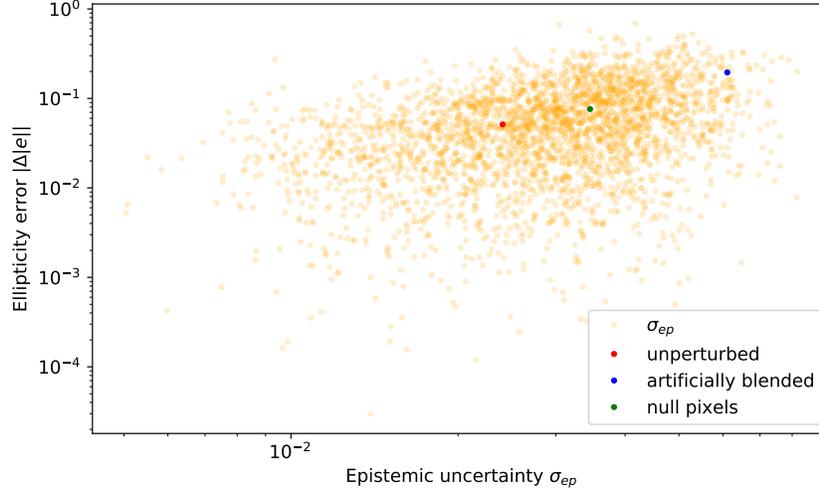


Figure 8.9: Ellipticity error $|\Delta e|$ as a function of the epistemic uncertainty, σ_{ep} . The red, blue and green dots are the measurements performed on the unperturbed image, artificially blended image, and image with null pixels respectively (from fig. 8.8).

These two perturbed images should be seen by the network as outliers, since it did not see such examples in the test sample. We expect the network to estimate a large epistemic uncertainty since it has a large probability of doing an incorrect measurement on these examples.

The epistemic uncertainty on the three cases, unperturbed, artificially blended, and with null pixels, is computed from 100 samples of the weight distributions, i.e. by feeding the network 100 times with each image. The results are shown on fig. 8.8. On the left column are shown the three tested images, the unperturbed image being at the top row, the artificially blended at the middle, and the image with null pixels on the bottom. Next to these images are drawn the e_1 and e_2 posterior distributions measured by the BNN, respectively on the center column and on the right column. The mean posterior distribution (the mean distribution of the 100 measurements) is shown in red. In grey are shown 30 of the 100 measurements and the true ellipticity value are indicated with a vertical blue line.

As expected, the epistemic uncertainty increases in both perturbed cases with samples that are further away from each other compared to the unperturbed case. This is particularly true for the artificially blended case where the measurement is also incorrect. The epistemic uncertainty is measured for these three cases. It increases by a factor 2.6 for the artificially blended image and by a factor 1.3 for the image with null pixels, compared to the unperturbed image. This increase in epistemic uncertainty can be linked to an increase in predictive ellipticity error. On figure fig. 8.9, the measurements on both perturbed images shift to the top right of the plane constructed from epistemic uncertainty and absolute value of the ellipticity error. Interestingly, the perturbation with null pixels does not impact the measurement a lot, and the epistemic uncertainty stays quite low. On the contrary, for the image with the central galaxy artificially blended, the measurement is incorrect and the epistemic uncertainty increases drastically reaching the tail of the ellipticity errors and epistemic uncertainty distributions.

Figure 8.9 also shows a strong correlation between epistemic uncertainty and predictive ellipticity error which can be used to decrease the impact of an incorrect measurement on the weak lensing analysis. This can be done either by taking into account this uncertainty in the approxi-

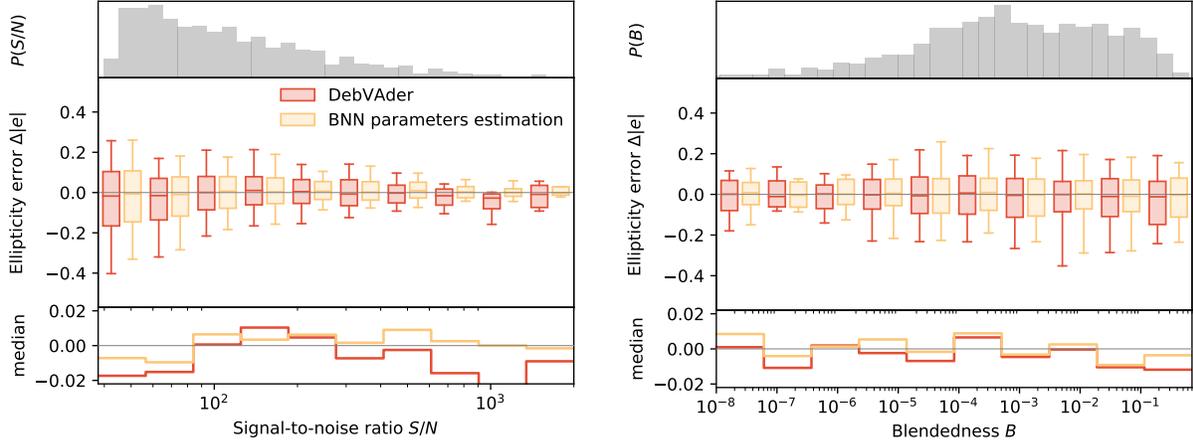


Figure 8.10: Distributions of differences in absolute ellipticity $|e|$ measured on DC2 images and compared to the true lensed ellipticities computed from the `Extragalactic catalog`. Comparison of the results from our BNN for parameters estimation (in yellow), and from DebVADER combined with the `REGAUSS` shape measurement algorithm (in red). The distributions are computed on subsets of the test sample split in bins of signal-to-noise ratio S/N (left) and blendedness (right).

mation of the correlation function, or simply by taking into account the full posterior distribution measured for the ellipticity parameters. Another common avenue is to select measurements that are deemed reliable for the analysis in a weak lensing pipeline, applying cuts on the sample, even though it can introduce selection bias (Mandelbaum et al., 2018). Here, one could use the values of the epistemic uncertainty to select our data, avoiding measurements that have a high probability of being incorrect.

To illustrate the relevance of using epistemic uncertainty instead of aleatoric uncertainty to perform this selection, we applied a cut to large values of aleatoric, epistemic and predictive uncertainties in order to remove 10% (300 images) of the sample and looked at the resulting mean ellipticity error over the cropped sample. The mean ellipticity error decreases by 7% when applying the cut on epistemic uncertainty, and by only 2% when using aleatoric or predictive uncertainty (since aleatoric uncertainty is on average much larger than the epistemic uncertainty). This is another argument showing the crucial need for this measurement if we want to use neural networks in cosmological analyses.

Note that this effect has already been shown in Theobald, Arcelin et al. (2021), where we trained a BNN to perform galaxy ellipticity measurements on simulated images of isolated galaxies and test them on blended galaxy images (see appendix B).

Comparison with DebVADER

To conclude, we can compare the results obtained with the BNN to the ones obtained with DebVADER combined with the `REGAUSS` shape measurement algorithm. Figure 8.10 shows the distributions of errors in ellipticity measurement for both methods. These preliminary results suggest that the BNN performs better as a function of both SNR and blendedness, with particularly better results at low SNR and high blendedness, an improvement probably also coming from the consideration of the epistemic uncertainty. Note that the difference between the results of the BNN and DebVADER at high SNR is probably due to the low statistics since this feature does not appear for DebVADER on fig. 7.25.

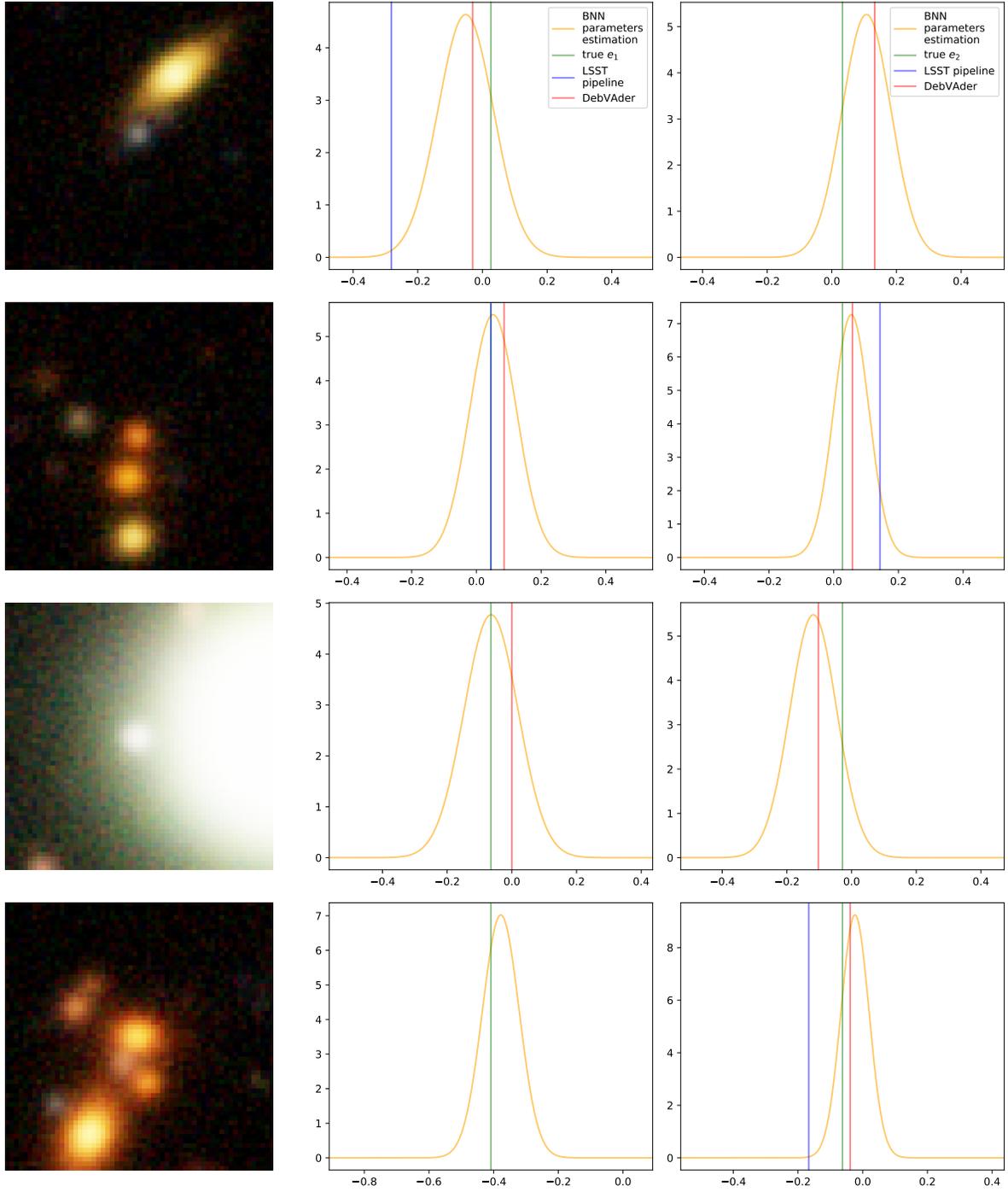


Figure 8.11: Comparison of shape measurement methods. On the left column is shown the image on which the measurement is done. The centred column shows the results for the measurement of e_1 , and the right one the results for e_2 . The methods compared are the BNN for ellipticity parameters estimation (orange), the combination of the DebVader and the REGAUSS shape measurement algorithm (red), and the measurement from the LSST pipeline, i.e. the deblender PHOTO and the REGAUSS shape measurement algorithm (blue). The true value is shown with a vertical green line.

Our BNN is able to measure the ellipticities of galaxies particularly well, probably even better than using the combination of a deblender based on a variational autoencoder and of a state-of-the-art shape measurement method. In any case, it performs much better than the generic deblender implemented in the LSST pipeline (PHOTO, see section 3.3.2). Contrary to usual shape measurement algorithms, it provides the full posterior predictive distributions of the two ellipticity parameters, allowing to take into account the uncertainty associated to each measurement. Also, it provides another piece of information, namely the epistemic uncertainty, that proves to be primordial for the use of neural networks in a scientific analysis pipeline.

Examples of measurements are shown on fig. 8.11 for some of the most blended images in the sub-sample of 3000 examples used for the BNN. The true values of e_1 or e_2 are shown in blue and are consistently within the posterior distribution measured by the BNN, whereas the LSST pipeline is not able to provide a measurement for the image on the third row for example. Measurements provided by the combination of DebVADER and REGAUSS also give more consistent results than the LSST pipeline on average.

8.5 Ellipticity and redshift estimation

In this section, we explore the possibility of adding the other required parameter for weak lensing analysis, i.e. the redshift, as target. We use a deterministic CNN as described in section 8.4 to measure both the two ellipticity parameters and the redshift, z .

Results on ellipticity parameters are not shown in this section, as they match the results obtained previously with the CNN network (section 8.4). Figure 8.12 presents the error on the redshift, splitting the test sample in bins of SNR, blendedness and true redshift.

As a function of SNR and blendedness, distributions of redshift errors are centred around zero with small deviations, and the median is contained within ± 0.025 . Once again, the spread increases at low SNR and high blendedness which is expected as galaxies respectively becomes fainter and receive fluxes contributions from neighbours. We expect the latter to impact drastically redshift estimation but the median of the redshift errors distribution remains quite stable, around ~ -0.01 , at high blendedness. This means that our network is able to perform redshift measurements as good on blended galaxies as on isolated objects.

The lower panel of fig. 8.12, shows the errors in redshift as a function of the true redshift. The median of the distribution of redshift errors is strongly correlated with redshift and spreads on a wider range of values. The network measures higher redshifts than expected at low redshift and conversely at high redshift. This underestimation at high redshifts can probably partially be explained by the low statistics in that range. However, more work is overall required to fully apprehend these results, as this section presents only the first step of the integration of redshift in our galaxy parameters estimation method.

Nonetheless, this is the first method providing a simultaneous measurement of the galaxy ellipticities and redshift posterior distributions, both parts of the required parameters for weak lensing analysis. We have seen in section 2.3.2 that these measurements are not independent and that a joint calibration of shear and redshift is required for future surveys. Providing a method that estimates both parts of the shear analysis might help in that sens.

Regarding the performance of our redshift estimation, lots of methods, using machine learning techniques for example, already exist to measure posterior distributions of photometric redshift (see section 2.2.4), and further work is required to investigate our results and compare our measurements to state-of-the-art methods.

In the next step, this measurement must be performed with a Bayesian neural network, in

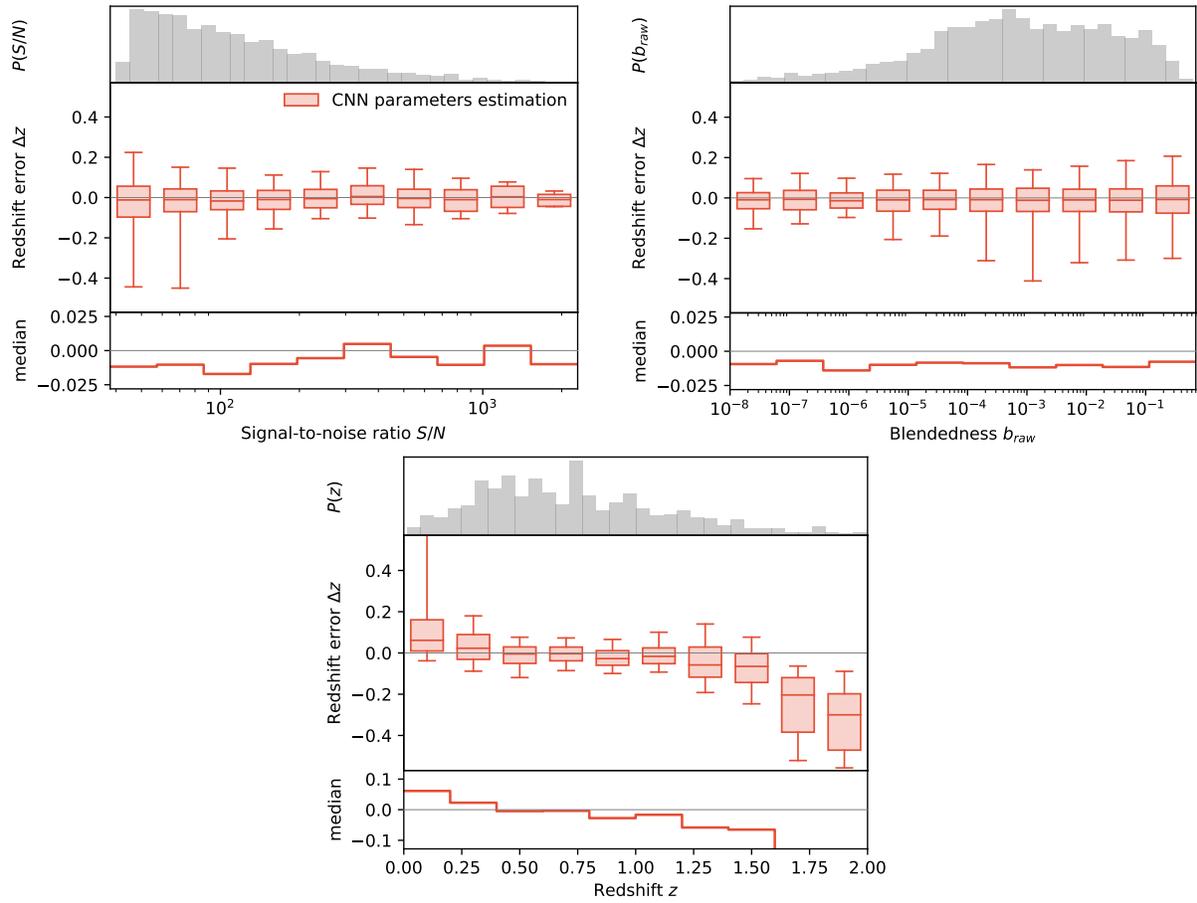


Figure 8.12: Distributions of differences in redshift z measured on DC2 images and compared to the true redshift extracted from the **Extragalactic catalog**. The results presented here were obtained with our CNN for parameters estimation. The distributions are computed on subsets of the test sample split in bins of signal-to-noise ratio S/N (left), blendedness (right) and redshift (bottom).

order to have a complete characterisation of the measurement uncertainty, and to account for it in the following pipeline. Since BNNs provide a full posterior distribution of the measured parameters, they are particularly suited for the estimation of photometric redshifts, which requires the estimation of the full probability density function (see section 2.2.4). The use of a BNN for the common estimation of ellipticities and redshift from DC2 images will be explored in future work.

The example presented in this section is a proof of concept showing that our network is able to perform the joint measurement of ellipticities and redshift. It shows that it is possible to use our galaxy parameters estimation method to perform the simultaneous estimation of the two elements required for weak lensing analysis. It could also probably be applied to other galaxy parameters such as fluxes in the different filters or galaxy size.

8.6 Discussion and future work

8.6.1 Application to real data

Similarly to DebVADER, the main issue of this method is that it is a supervised learning algorithm, meaning that it requires a training sample labelled with the true galaxy parameters to estimate.

On real data, the true values of galaxy ellipticities for example, are never known. Consequently our method would mainly rely on simulations and require a high level of realism to be able to measure ellipticities of objects with more complex shapes than the parametric models used in DC2. The very realistic *Illustris* simulation could be used to that extend (Vogelsberger et al., 2014).

However, the main goal of our method is to be able to avoid the deblending step, i.e. to provide parameters estimation as good on blended scenes as on isolated galaxies. Consequently, the diversity of galaxy shapes in the simulation would be secondary to the blends density and their variety of configurations. The introduction of a small model bias might not be a major issue since it can probably be calibrated using a calibration scheme such as METACALIBRATION (E. Huff and Mandelbaum, 2017; Sheldon and E. M. Huff, 2017).

Another issue is that our method does not take into account information from the PSF. It directly learns the models that are used in the simulation it is trained on. Consequently, an avenue, to apply this method on real data, could be to deconvolve the scene that is going to be fed to the network by the measured PSF on this scene, and to reconvolve it by a larger PSF used in the simulation.

Finally, already mentioned in section 7.5.2, a method to test our method on more realistic scenes could be source injection. It consists in injecting a fake simulated source in a real image, allowing to control each parameter of the simulation while including true galaxies or stars, and possible observational defects. Of course we would still be measuring the shape of an artificially simulated galaxy, but it could make our network robust to potential observational effects that could not be reproduced on a simulation. Work is currently being done within DESC to perform source injection in HSC data, as it is very similar to what LSST data will look like.

8.6.2 Bias in ellipticity distribution

As mentioned, our method is based on supervised learning and therefore is highly dependent on the distribution of the target parameters. To explore this, we can take a look at the distributions of ellipticity parameters, and see that they are peaked around zero and quickly weaken going to ± 1 (see fig. 8.13).

This can introduce a bias in the measurement of highly elliptical galaxies, which is clearly visible on fig. 8.14 (for the network with the "no weighting" label). Going higher in ellipticity, the

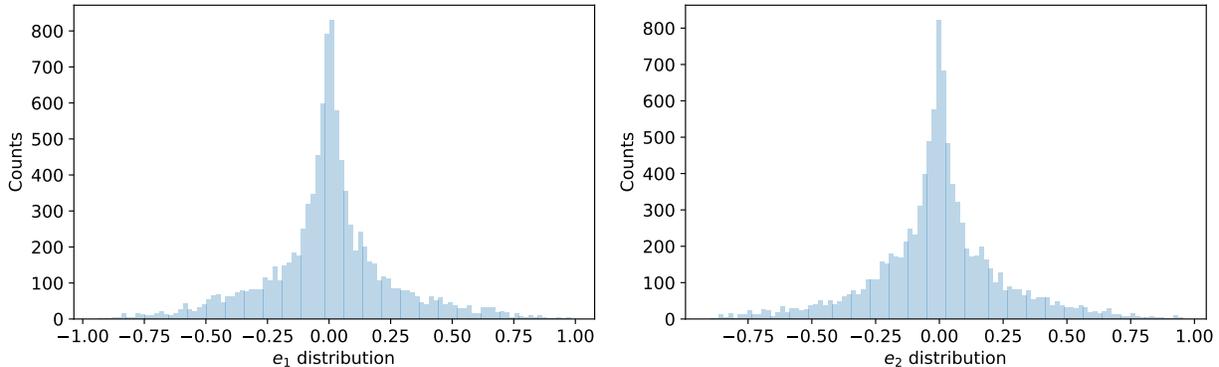


Figure 8.13: Distribution of the two ellipticity parameters e_1 and e_2 for the test sample composed of 10000 images extracted from the DC2 simulation.

negative bias in ellipticity errors increases. The network measures lower ellipticities than the true ones, simply because it sees much more example of galaxies with low ellipticity during training.

This is not necessarily a problem here, since weak lensing relies on ensemble measurement, i.e. the cosmic shear is computed from measurements on a group of galaxies. A high precision on the measurement of low ellipticity galaxies, which represent the majority of the galaxies sample, and consequently carry most of the weak lensing signal, might be more important than a high precision, on average, on the complete range of ellipticities.

To test this hypothesis we trained a network applying weighting to the images. Since the distribution of ellipticities is peaked around 0, we down-weight the galaxies with low ellipticities and increase the weight of galaxies with high ellipticities. The weight is simply written as

$$w_i = \frac{|e_{1i}| + |e_{2i}|}{\sum_i (|e_{1i}| + |e_{2i}|)} \quad (8.10)$$

with (e_{1i}, e_{2i}) the ellipticity parameters of the galaxy i . Note that, in this section, we show the performance of a network implemented as in section 8.3, i.e. with the trained encoder. Similar results would be obtained on the CNN.

Figure 8.14 shows the comparison between the weighting and non-weighting configurations. Results are shown as a function of input absolute ellipticity $|e|$, SNR, and blendedness. First, as a function of ellipticity, weighting improves the measurement accuracy for highly elliptical galaxies, even though it does not fully cancel the bias. However, as a function of SNR or blendedness, the median of errors on ellipticity measurements is smaller for the network without weighting, especially at low SNR and high blendedness. On the complete test sample, the network without weighting performs better than the network with weighting by a factor 2.2 and 2.8 on the median of the ellipticity errors distribution, as a function of SNR and blendedness respectively. This result stresses the importance of precise measurements on low ellipticity galaxies rather than on elliptical galaxies.

More tests should be carried out, for example to compute the shear bias introduced by the two configurations, but one should probably expect similar results as low ellipticity galaxies represent the majority of the test sample, and therefore carry most of the weak lensing signal. This result underlines a major advantage of our method which, since it learns from the data, is able to take into account these statistical effects. It also insists on the importance of a realistic distribution of the parameter that we are trying to learn in the training sample.

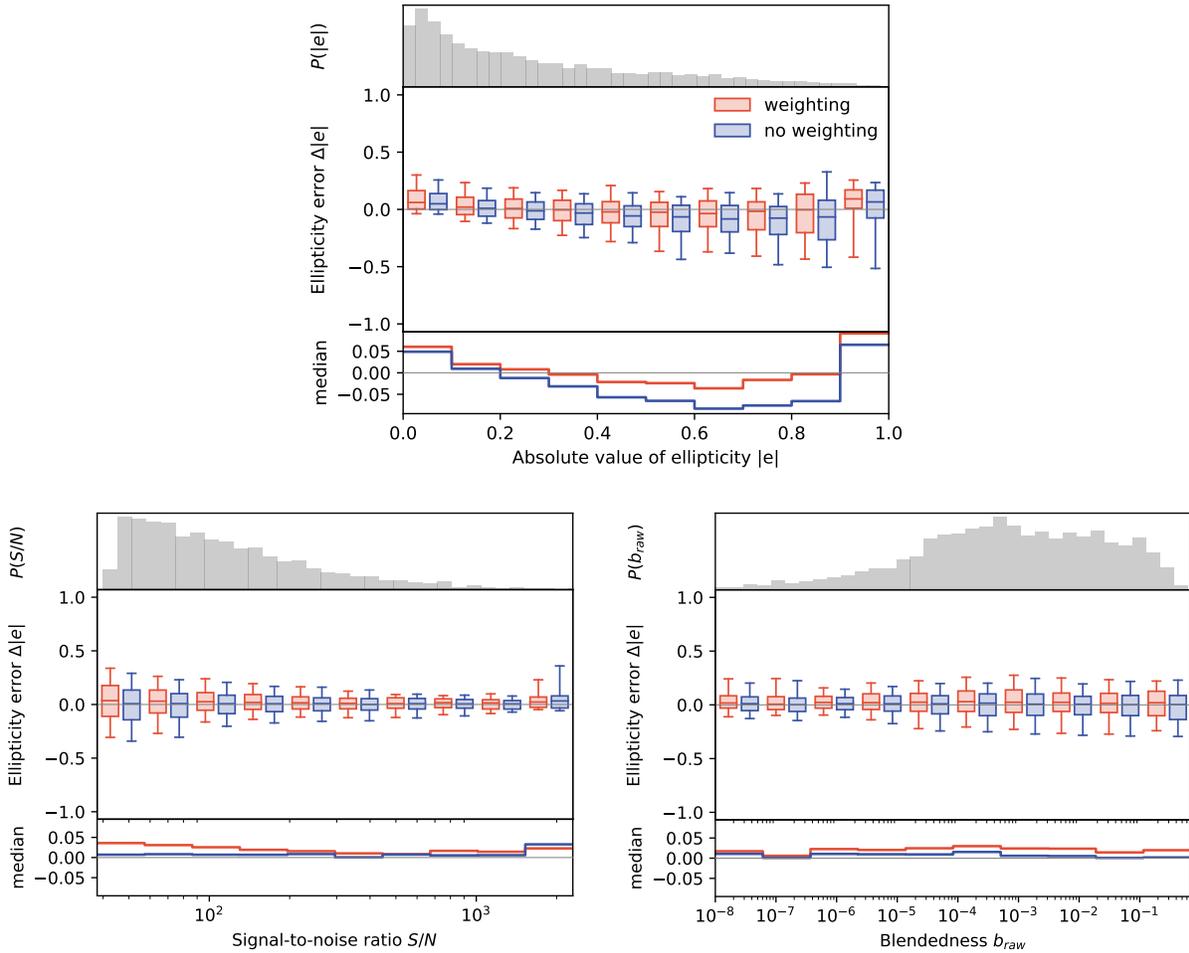


Figure 8.14: Distributions of differences in absolute value ellipticity $|e|$ measured on DC2 images and compared to the true lensed ellipticity computed from the **Extragalactic catalog**. The results presented here compare the performance of two CNNs for parameter estimation, one which was trained without weighting input images (blue), another one trained weighting images as a function of the galaxies ellipticity (red). The distributions are computed on subsets of the test sample split in bins of absolute value of ellipticity $|e|$ (top), SNR (bottom left) and blendedness b_{raw} (bottom right).

Finally, the small ellipticities are always slightly overestimated compared to what they really are. This could have several explanations, of different nature. First it could be linked to the noise in the image, it is difficult to measure a perfectly round object in a noisy image since the noise alters the light profile. However this effect is not seen at low values of SNR for the network without weighting so this hypothesis must be discarded. Secondly, it could be due to some unrecognised blends which have a wider light profile, leading to a larger ellipticity estimation. Since this algorithm rely on inputs from the detection method used by the LSST pipeline, it can be impacted by unrecognised blends. This hypothesis would need to be verified on simulations. Third, it could be linked to the training or our neural network. The network usually starts by predicting the mean of the parameters distribution before learning values in the wings of this distribution. The mean ellipticity of the galaxies in the test sample being slightly non-zero, it is possible that the non-zero values at low ellipticities might be a residual of this early convergence to the mean.

8.6.3 Choice of prior

In section 8.4.1 I set the prior for the weights distribution as the k-tied normal distribution (Swiatkowski et al., 2020), but nothing ensures that this distribution is best fitted for our purpose. As mentioned in section 5.3.3, Gaussian mean-field variational inference, from which the k-tied normal distribution is derived, might not be the most relevant for our tasks and we should try different priors to find the most efficient. An avenue could be to try other families or priors or *trainable* priors. Indeed it could be possible to learn the mean and the variance of the prior distribution during training in order to have better performance of our network. However, this would not make sens as the trainable prior would not really be a prior anymore from a Bayesian perspective. Another path could be to provide more informative priors that could be built from expert knowledge. For example, it could be possible to pre-train the model so that the prior predictive distribution matches a known distribution of the ellipticities and redshift up to a particular redshift. This distribution could come from the measurements performed on the COSMOS deep field for example (Mandelbaum et al., 2014). Hence, prior predictive checking (Gelman et al., 2020) allows to set more informative prior distribution for the weights that comes from measurements done in previous surveys. This is a very interesting path which would be worth pursuing in the future.

8.7 Conclusion

To conclude, we showed that a CNN can be used for precise galaxy parameters estimation on DC2 images, requiring no deblending. It is able to reproduce the same level of precision in ellipticity errors as with DeVADER and the REGAUSS shape measurement algorithm, providing accurate measurements for both isolated and blended galaxies.

We first applied our network for the estimation of galaxy ellipticity parameters and showed that it improved performance on the median of the distribution of ellipticity errors by a factor 8 and 5 compared to the LSST pipeline, as a function of blendedness and SNR respectively. This network also provides an estimation of the aleatoric uncertainty which is found to be well calibrated, i.e. the network does not overestimate neither underestimate the uncertainty. Then we tested a Bayesian neural network using Flipout (Wen et al., 2018) on the two dense layers of the CNN. This allows for the measurement of the epistemic uncertainty which characterises the uncertainty coming from the training sample. This uncertainty is strongly correlated with large measurement errors and can be used to reduce the impact of (or flag) incorrect measurements. The epistemic uncertainty is a crucial information for the application of neural networks in cosmology and this work shows that it

can be measured using a Bayesian neural network for galaxy parameters estimation. The Bayesian neural network presented in this work provides galaxy parameters posterior distributions that can be used in a Bayesian analysis of cosmic shear. Finally, we added the redshift as a parameter to estimate in order to obtain a simultaneous measurement of both ellipticities and redshift, the two key elements of weak lensing analysis. This point estimate CNN is able to measure posterior distributions for both ellipticities and redshift without requiring deblending.

We also discussed the difficulties associated with this method and particularly the fact that this method is based on supervised learning. We underlined the importance of realistic simulations and briefly presented the possible avenue of using fake source injection to get a step closer to real data. We discussed the impact of the parameter distribution of the training sample, which plays a major role on the results obtained with our method. We showed that our algorithm naturally takes into account the statistical properties of the parameters that we are trying to learn, ellipticities and redshifts in our case. It is a powerful advantage of our method for weak lensing analysis, since cosmic shear is computed from measurements on a group of galaxies. However, it also points out the importance for realistic parameter distributions in the simulation, since any bias in this simulation will be embedded in our network's measurements. One of the next steps of this work will consist in testing this method on a sample of galaxies with higher magnitudes extracted from the DC2 simulation, in order to confirm the results presented above.

Work remains to be done. We have already mentioned the necessity of using a BNN for the joint estimation of the posterior distributions of ellipticities and redshifts. This will allow to fully characterise the epistemic uncertainty associated to each measurement, and to decrease the potential impact of those that are incorrect.

Our method provides full posterior distributions of ellipticities which could be combined, for galaxies in a particular redshift bin and close from each other, to obtain a posterior distribution of the cosmic shear at a particular location. This posterior distribution could then be used in a Bayesian weak lensing analysis. One of the next step of this analysis is consequently to explore how to combine these posteriors, and to verify that the resulting distribution represents a correct estimation of the shear posterior distribution.

Before that, it is necessary to characterise the shear bias introduced by this method, and to calibrate it, if necessary, using for example the METACALIBRATION algorithm (Bosch et al., 2018).

Finally these results, even if they are preliminary, are very promising. We showed that our parameters estimation algorithm is able to measure both ellipticities and redshifts on blended or isolated objects, without requiring deblending. We also showed, using a Bayesian neural network, the relevance of the epistemic uncertainty estimation in our context. It is a necessary information to yield reliable measurements with neural networks for weak lensing analysis. We hope that our results can encourage the use of Bayesian neural networks for cosmology.

Bibliography

- Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe (Jan. 2016). “Variational Inference: A Review for Statisticians”. In: *arXiv e-prints*, arXiv:1601.00670, arXiv:1601.00670. arXiv: [1601.00670 \[stat.CO\]](#).
- Bosch, James et al. (2018). “The Hyper Suprime-Cam software pipeline”. In: *Publications of the Astronomical Society of Japan* 70.SP1, S5.
- Drlica-Wagner, A. et al. (Apr. 2018). “Dark Energy Survey Year 1 Results: The Photometric Data Set for Cosmology”. In: *ApJS* 235.2, 33, p. 33. DOI: [10.3847/1538-4365/aab4f5](#). arXiv: [1708.01531 \[astro-ph.CO\]](#).
- Gal, Yarin, Jiri Hron, and Alex Kendall (May 2017). “Concrete Dropout”. In: *arXiv e-prints*, arXiv:1705.07832, arXiv:1705.07832. arXiv: [1705.07832 \[stat.ML\]](#).
- Gelman, Andrew et al. (Nov. 2020). “Bayesian Workflow”. In: *arXiv e-prints*, arXiv:2011.01808, arXiv:2011.01808. arXiv: [2011.01808 \[stat.ME\]](#).
- Hortúa, Héctor J. et al. (Nov. 2020). “Parameter estimation for the cosmic microwave background with Bayesian neural networks”. In: *PRD* 102.10, 103509, p. 103509. DOI: [10.1103/PhysRevD.102.103509](#). arXiv: [1911.08508 \[astro-ph.IM\]](#).
- Huff, Eric and Rachel Mandelbaum (Feb. 2017). “Metacalibration: Direct Self-Calibration of Biases in Shear Measurement”. In: *arXiv e-prints*, arXiv:1702.02600, arXiv:1702.02600. arXiv: [1702.02600 \[astro-ph.CO\]](#).
- Mandelbaum, Rachel et al. (May 2014). “The Third Gravitational Lensing Accuracy Testing (GREAT3) Challenge Handbook”. In: *ApJS* 212.1, 5, p. 5. DOI: [10.1088/0067-0049/212/1/5](#). arXiv: [1308.4982 \[astro-ph.CO\]](#).
- Mandelbaum, Rachel et al. (Dec. 2018). “Weak lensing shear calibration with simulations of the HSC survey”. In: *MNRAS* 481.3, pp. 3170–3195. DOI: [10.1093/mnras/sty2420](#). arXiv: [1710.00885 \[astro-ph.CO\]](#).
- Sheldon, Erin S. and Eric M. Huff (May 2017). “Practical Weak-lensing Shear Measurement with Metacalibration”. In: *ApJ* 841.1, 24, p. 24. DOI: [10.3847/1538-4357/aa704b](#). arXiv: [1702.02601 \[astro-ph.CO\]](#).
- Swiatkowski, Jakub et al. (Feb. 2020). “The k-tied Normal Distribution: A Compact Parameterization of Gaussian Mean Field Posteriors in Bayesian Neural Networks”. In: *arXiv e-prints*, arXiv:2002.02655, arXiv:2002.02655. arXiv: [2002.02655 \[cs.LG\]](#).
- Theobald, Claire et al. (Apr. 2021). “A Bayesian Convolutional Neural Network for Robust Galaxy Ellipticity Regression”. In: *arXiv e-prints*, arXiv:2104.09970, arXiv:2104.09970. arXiv: [2104.09970 \[cs.LG\]](#).
- Vogelsberger, M. et al. (May 2014). “Properties of galaxies reproduced by a hydrodynamic simulation”. In: *Nature* 509.7499, pp. 177–182. DOI: [10.1038/nature13316](#). arXiv: [1405.1418 \[astro-ph.CO\]](#).

Wen, Yeming et al. (Mar. 2018). “Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches”. In: *arXiv e-prints*, arXiv:1803.04386, arXiv:1803.04386. arXiv: [1803.04386](#) [[cs.LG](#)].

Wenzel, Florian et al. (Feb. 2020). “How Good is the Bayes Posterior in Deep Neural Networks Really?” In: *arXiv e-prints*, arXiv:2002.02405, arXiv:2002.02405. arXiv: [2002.02405](#) [[stat.ML](#)].

Conclusion

In this manuscript, I reported the work that I have carried out between October 2018 and July 2021. The six first chapters set the context for this work. After introducing the standard cosmological model and the different observational probes used to constrain its parameters, I focused on weak lensing, and more particularly on a systematic which will be dominant for future weak lensing surveys, the blending of sources. Then, I took some time to describe the formalism of the methods used in this work, presenting in particular, deep generative models and an introduction to Bayesian neural networks. Finally, these methods were applied to the issue of blending in chapter 7 and chapter 8.

In chapter 7, I proposed a new method to deblend galaxies using variational autoencoders, called DebVAdler. This method, validated on simulated images, showed good performance in shape and flux reconstruction, benefiting largely from the joint pixel analysis of LSST and Euclid data. It was successfully applied to artificially blended images of real galaxies using transfer learning, and to more realistic simulations resembling the future LSST data, showing better results than the generic deblender implemented in the LSST pipeline. Finally, DebVAdler only deblends one galaxy at a time, which led us to design an iterative procedure including detection, classification, deblending and masking steps, a potential solution for the issue of unrecognised blends.

In chapter 8, building on the work done with DebVAdler, I presented a new galaxy parameters estimation algorithm able to predict accurately ellipticities and redshifts from DC2 images, without requiring a deblending step. This algorithm gives precise measurements, similar to the ones obtained with DebVAdler and a state-of-the-art shape measurement method, when applied for ellipticity measurements. Then, using a Bayesian neural network, I showed that epistemic uncertainty is a crucial information to yield reliable ellipticity measurements. It is strongly correlated with measurement errors, allowing to decrease the impact of incorrect estimations in the ensemble measurement for cosmic shear, or even to apply cuts to suppress them. I concluded this chapter by showing that our method can provide a joint measurement of ellipticity parameters and redshifts, the two key elements of weak lensing analysis.

However, much is left to be done, and improvements can be made at several stages of these methods. Concerning the DebVAdler algorithm, and the iterative deblending pipeline presented in chapter 7, I mentioned already that at each stage, refinements are required. The classification scheme for example, is clearly sub-optimal and work is necessary to improve it. The DebVAdler algorithm itself could also be improved. For example, it does not provide a measurement of the network uncertainty. This requires some work but could provide valuable information to the method. Finally, a crucial step for the validation of this method is to compare it to other state-of-the-art methods such as SCARLET, the baseline deblender which will be implemented in the LSST pipeline. This will be soon possible using the Blending ToolKit presented in chapter 3.

Regarding the galaxy parameters estimation method, the first step is to use a Bayesian neural network to perform both ellipticity and redshift measurement, allowing for the full characterisation of the redshift posterior distribution required for weak lensing analysis. One a more technical

point, the network could be made fully Bayesian by imposing distributions over the weights of the convolutional layers. This would provide a complete measurement of the epistemic uncertainty but it might be difficult to implement, especially since adding distributions over each weight in the network might lead to over-regularisation. It would also be necessary to characterise the shear bias introduced by our method, and to see if it can be calibrated by **METACALIBRATION** for example. Finally, this method needs to be validated on real data and compared to state-of-the-art methods for both galaxy ellipticity and redshift measurements.

Both these methods are supervised learning algorithms which require labelled data. To that extent, and particularly for the parameters estimation algorithm, simulations will be key to train these networks. They will need to be as realistic as possible to avoid the introduction of biases in the analysis. As explained in chapter 5, deep generative models can be part of the solution to generate images of galaxies with arbitrarily complex shapes.

The next step of this work would be to go to cosmic shear estimation, which could probably be possible using our parameters estimation algorithm. Indeed, once this estimator calibrated, the combination of the measured ellipticity posterior distributions of galaxies in a particular redshift bin, and close from each other, could provide an estimation of the posterior distribution of the shear at this particular location. This could allow for a Bayesian analysis of cosmic shear. It would be even more relevant to use this method knowing that cosmic shear relies on ensemble measurements and not on the measurement of individual galaxy shapes. Indeed, our method naturally takes into account the statistical properties of the sample it is trained on, and naturally decreases the impact of incorrect measurements by providing a posterior distribution, two characteristics that can be particularly helpful for the estimation of cosmic shear. This path could be explored in future work, for example by applying a neural network trained on simulations, on real data.

In this work, we only looked at blending through the perspective of shape and flux measurements, but it is also tightly linked to detection, and I think that improvements could probably also be carried out at this step of the analysis. For example, I would be very interested in exploring the use of a deep learning algorithm for source detection on multi-band images, to see if it could decrease the number of unrecognised blends. Since neural networks naturally make the most of the color information, this kind of methods might be able to perform particularly well at this task. A first step could be to compare the performance of the detection algorithm presented in chapter 7, with the **SExtractor** algorithm.

At this point, I hope that the reader is convinced that Bayesian deep learning can play an important role to address specific and complex issues for cosmology. The complexity of the models that can be learned by neural networks, combined with the reliable characterisation of the uncertainty provided by methods like Bayesian neural networks, makes it now conceivable to include these methods in an cosmological data analysis pipeline. Moreover, it provides the ability to generalise to multi-band, multi-instrument data, which is a key advantage in the perspective of future surveys as lot of them will produce multi-band data on wide parts of the sky. The sky coverage of these surveys will inevitably overlap, encouraging synergies between them.

With LSST and Euclid starting recording data in 2023, the era of wide and deep surveys ahead of us results in high expectations for new discoveries, and deep learning has a key role to play in that process. As demonstrated by the success of a conference on Bayesian deep learning for cosmology and gravitational waves, that we organised at the APC in March 2020, the scientific community is already well aware of this point and is getting ready to use deep learning to unveil new secrets of the Universe.

Appendices

Appendix A

Comparison of Graphcore IPUs and Nvidia GPUs for cosmology applications

Abstract. This paper represents the first investigation of the suitability and performance of Graphcore Intelligence Processing Units (IPUs) for deep learning applications in cosmology. It presents the benchmark between a Nvidia V100 GPU and a Graphcore MK1 (GC2) IPU on three cosmological use cases: a classical deep neural network and a Bayesian neural network (BNN) for galaxy shape estimation, and a generative network for galaxy images simulation. The results suggest that IPUs could be a potential avenue to address the increasing computation needs in cosmology.

A.1 Introduction

Upcoming imaging galaxy surveys, such as the Legacy Survey of Space and Time (LSST, LSST Science Collaboration et al., 2009) conducted at the future Vera C. Rubin Observatory, ESA’s Euclid satellite (Refregier et al., 2010) or the Nancy-Grace-Roman Space Telescope (Roman Space Telescope, former WFIRST, Spergel et al., 2013) will produce an unprecedented amount of observational data. For instance, LSST will produce 20 Terabytes of data every night and around 60 Petabytes over the its 10 years of service¹. These amounts are mainly due to the quality and the nature of the data recorded: large sky images in different filters (or colors).

In order to prepare and test future data analysis pipelines, it is necessary to simulate a quality and a quantity of data as close as what will be recorded. The pipelines will need to process data with fast and accurate analysis techniques, and to provide reliable uncertainties. When looking for fast data processing techniques, a now common choice is to turn to deep learning algorithms, which allow for very fast inference on data once trained.

As an example, a LSST pipeline will be dedicated to look for transients continuously during the night. It will have to send an alert to the community in within 60 seconds after image readout². The alert system is supposed to produce around 10 million alerts per night. A solution proposes to use deep learning to accelerate parts of the process such as the classification of events (Möller et al., 2020).

¹<https://www.lsst.org/about/dm>

²<https://www.lsst.org/about/dm>

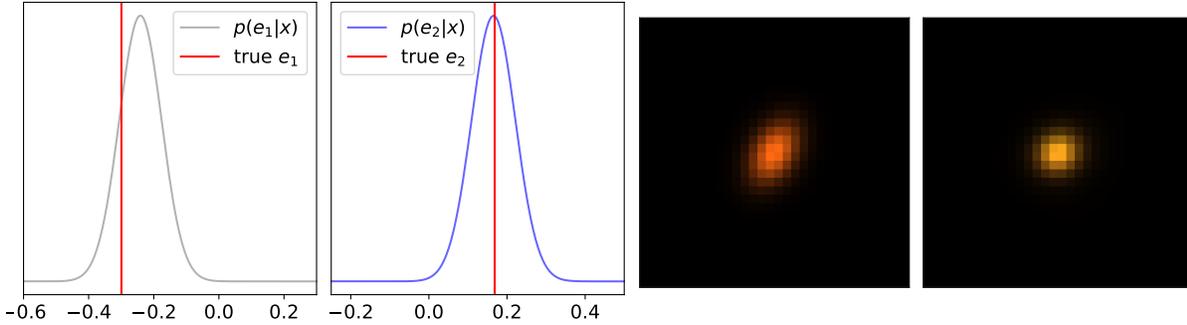


Figure A.1: On the left, the posterior distributions of a galaxy e_1 and e_2 shape parameters estimated with a deterministic deep neural network. On the right, two galaxy images generated via sampling the latent space of a trained VAE.

In order to train deep artificial neural networks, specific hardware are necessary. Graphics Processing Units (GPU) are the most common technology used nowadays. It evolved a lot over the years to reach very high performance at the moment and surpasses Central Processing Units (CPU). Google Tensor Processing Units (TPU), LightOn Optical Processing Units (OPU)³, or Graphcore Intelligence Processing Units (IPU)⁴, are more recent examples of hardware developments aiming for faster computing.

Graphcore’s IPUs have already shown better results than GPUs at language and speech processing, computer vision, probabilistic modeling⁵ and even in some common use cases of particle physics (Mohan et al., 2020).

This paper presents the benchmark between a Nvidia V100 GPU and a Graphcore MK1 (GC2) IPU on three cosmological use cases: a deterministic deep neural network and a Bayesian neural network (BNN) for galaxy shape estimation, and a generative network for galaxy images production. Figure A.1 shows outputs of the previously mentioned use cases: the posterior distribution of galaxy ellipticity parameters, estimated from a deterministic network on the left, and isolated galaxy images generated from a neural network on the right. The code to reproduce this study can be found on GitHub [🔗](#)⁶.

This paper represents a first investigation of the suitability and performance of IPUs in deep learning applications in cosmology. The paper is organised as follows: in appendix A.2, the two tested hardware are described, appendix A.3.2 presents the results obtained to train a deterministic and a Bayesian neural network to learn galaxy shape parameters from isolated galaxy images, appendix A.3.3 compares the performance of the tested hardware in the case of inference, to generate isolated galaxy images, and finally, appendix A.4 concludes and discusses the different results.

³<https://www.lighton.ai/>

⁴<https://www.graphcore.ai/>

⁵<https://www.graphcore.ai/benchmarks>

⁶<https://github.com/BastienArcelin/IPU-GPU>

	Processing Unit	Cores	Memory	Single precision performance	Max Power Consumption
GPU	Nvidia Tesla V100 PCIe	5120	32000Mb	14 TFLOPS	250 W
IPU	Graphcore Colossus™GC2	1216	286Mb	31.1 TFLOPS	120 W

Table A.1: Comparison of IPU⁷ and GPU⁸ specification for machines used for this benchmark.

A.2 Hardware description

In this paper I compare the performance of a single first generation MK1 (GC2) IPU⁷ to the performance of a single Nvidia V100 GPU. Key specifications of the IPU and the GPU⁸ are presented in table A.1.

The IPU is a type of processor with a different architecture from the one of the GPU (Jia et al., 2019). It is specifically designed for machine learning applications as it offers true MIMD (Multiple Instruction, Multiple Data) parallelism. It is designed to adapt to irregular computation and sparse data access. Each processor is composed of 1216 cores, called tiles. Each tile contains 256 KiB of local memory and is complex enough to execute distinct programs. It can also support up to six threads, allowing for 7.296 threads that can be executed in parallel on an IPU. On the contrary, GPU has a SIMD (Single Instruction, Multiple Data) architecture. The 5120 GPU cores are grouped in clusters for which all of the cores will execute the same instruction. As threads are scheduled on clusters, they perform the same operation on independent data.

Two other differences have a direct impact for the user: the memory per processing unit and the single precision performance. The IPU has a much smaller memory than the GPU. This may become a bottleneck in case of large neural network, i.e. with a lot of training parameters, or when processing heavy data, such as imaging survey data.

The other major difference is the single precision performance, IPUs being able to reach 31.1 TFLOPS (TeraFLOPS, Floating-point operation per second), more than twice the GPU value.

Finally, it is significant to notice that the IPU used here reaches the presented level of performance while consuming less than half electrical power consumed by the GPU.

Note than more recent and more powerful versions of both Graphcore and Nvidia technologies have been released: the Colossus™MK2 GC200 at Graphcore, with increased in-processor and exchange memory, and the A100 GPU at Nvidia, with more memory and higher single precision performance.

Both GPUs and IPUs support TensorFlow (Abadi et al., 2016) or PyTorch (Paszke et al., 2019), user-friendly frameworks allowing programmers without specific hardware knowledge to access high-performance computing.

Examples presented in this work are implemented using TensorFlow in its version 2.1.0. Graphcore provided drivers and its Poplar Software Development Kit (SDK) which is used in its version 1.4.0. The GPU is accessed via the IN2P3 Computing Center (CC-IN2P3–Lyon/Villeurbanne - France), and runs with Nvidia CUDA version 10.1.105.

⁷<https://www.graphcore.ai/products/ipu>

⁸<https://www.nvidia.com/en-gb/data-center/v100/>

A.3 Cosmological use cases

A.3.1 Training data

Images generation uses the same hypothesis as what is described in section 2.1 of Arcelin et al., 2021. Images of isolated galaxy are simulated using GalSim⁹ (Rowe et al., 2015) from parametric models fitted to real galaxies from the HST COSMOS catalogue, containing 81 500 images. These fits were realised for the third Gravitational Lensing Accuracy Testing (GREAT3) Challenge (Mandelbaum et al., 2014, see Appendix E.2).

This benchmark is realised on simulations of LSST-like images composed of the six LSST filters (*ugrizy*, LSST Science Collaboration et al., 2009) and convolved with a fixed Point-Spread-Function (PSF). For each filter, Poisson noise is added. Images size is arbitrarily fixed at 64×64 pixels and the pixel size is 0.2 arcsecond. The code for image generation can be found here [🔗](#)¹⁰.

Artificial neural networks presented in the next sections are trained by feeding them by batches. The batch size is not fixed, it is used as a parameter to compare the performance of the two tested hardware. The dimension of the input tensors is (*batchsize*, 6, 64, 64).

A.3.2 Galaxy shape parameter estimation

With future galaxy surveys, various probes will be used to study dark energy. One of the most promising is cosmic shear (Kilbinger, 2015), i.e. the measurement of the coherent distortion of background galaxies by foreground matter through weak gravitational lensing. To obtain a precise measurement of this probe, the shape and redshift (or distance) of the observed galaxies must be accurately measured. Several galaxy shape measurement methods based on model-fitting or moments measurement already exist, but none of them perform accurately on blended objects, i.e. overlapping objects on an image. As galaxy surveys observe further in the sky, dealing with overlapping objects is becoming a major challenge. For example, Bosch et al., 2018 estimated that 58% of detected objects will be blended in the HSC wide survey.

We developed a new technique based on deep neural networks and convolutional layers which permits to measure shape ellipticity parameters on isolated as well as on blended galaxies (Arcelin et al., [in prep](#)). Our network takes as input the isolated galaxy or blended galaxies scene and outputs the posterior distribution of the target galaxy ellipticity parameters. First, we tested a deterministic neural network, and then a BNN. The advantage of BNNs is that they allow for an accurate characterisation of the epistemic uncertainty, which is necessary to do reliable prediction that can be used in scientific studies (Charnock, Perreault-Levasseur, and François Lanusse, 2020).

As this work focuses on hardware performance, the network tested here is a simplified version of the one working on blended scenes and presented in Arcelin et al., [in prep](#). It is trained to retrieve ellipticity parameters distribution from isolated galaxy images only. It is composed of one batchnormalization layer, eight stacked convolutional layers, one dense layer and the output is a multivariate normal distribution layer¹¹ from the TensorFlow Probability library.

Deterministic neural network

The training of a deterministic network performing galaxy shape parameter estimation is the first test presented here. The network is defined as *deterministic* since, once trained, trainable parameters have a fixed value. In other words, the output posterior distributions will not change if

⁹<https://github.com/GalSim-developers/GalSim>

¹⁰https://github.com/BastienArcelin/image_generation_GalSim

¹¹<https://www.tensorflow.org/probability/>

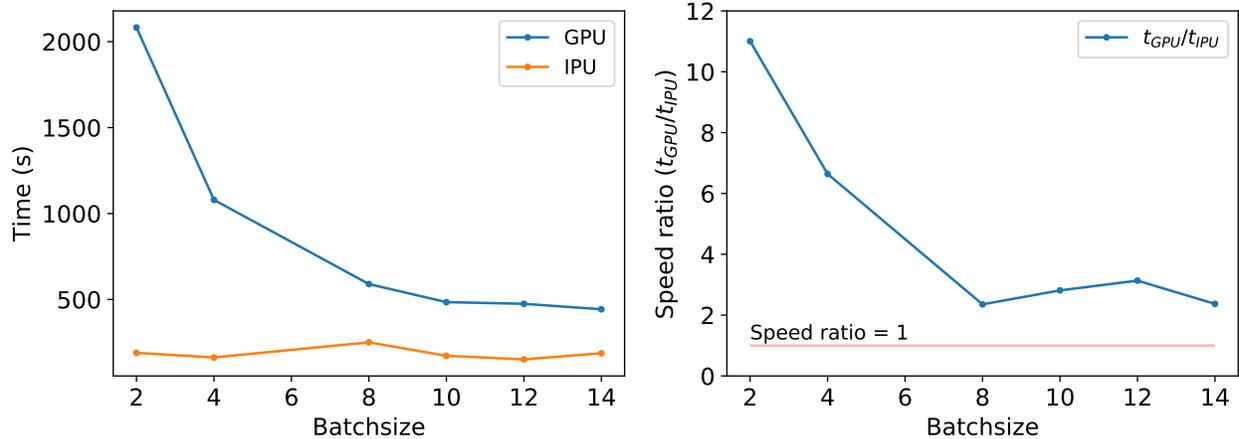


Figure A.2: (left) Training time for a deep deterministic neural network performing galaxy parameter estimation as a function of the batch size. (right) Speed ratio of GPU and IPU as a function of the batch size. The red horizontal line represents the value $R = 1$, i.e. the network takes the same amount of time for training in both cases. The neural network is composed of 1.5M trainable parameters.

the network is fed twice with the same isolated galaxy image. This network is composed of 1.5M parameters.

During training, the Adam optimizer (Kingma and Ba, 2014) is used and the loss is defined as the negative log probability of the output:

$$\mathcal{L} = -\log(p(x)) \quad (\text{A.1})$$

Here p is the posterior distribution, output of the multivariate normal distribution layer and x is the target value.

The training procedure is the same for IPUs and GPUs and the chosen metric is the training time for 100 epochs. Ten epochs are considered as a warm-up phase. The test is realised for different batch sizes: from 2 to 14, 14 being the maximum batch size that can be handled by the IPU before memory is full (see appendix A.2).

The results are presented on fig. A.2. Using IPUs improves the training time compared to GPUs, especially for very small batch sizes. The network does not learn well for batch sizes lower than 4, but even for greater values, the network is trained at least twice as fast on the IPU compared to the GPU. This result confirms the already shown efficiency of IPUs on tasks close to computer vision¹².

Bayesian neural network

Then, the same test is run, but this time with a BNN. Instead of learning fixed values for each parameter, training a BNN consists in learning a posterior distribution over each parameter, knowing the data, through variational inference (Valentin Jospin et al., 2020 and Charnock, Perreault-Levasseur, and François Lanusse, 2020). Here the reparameterization trick is used to perform

¹²<https://www.graphcore.ai/mk2-benchmarks>

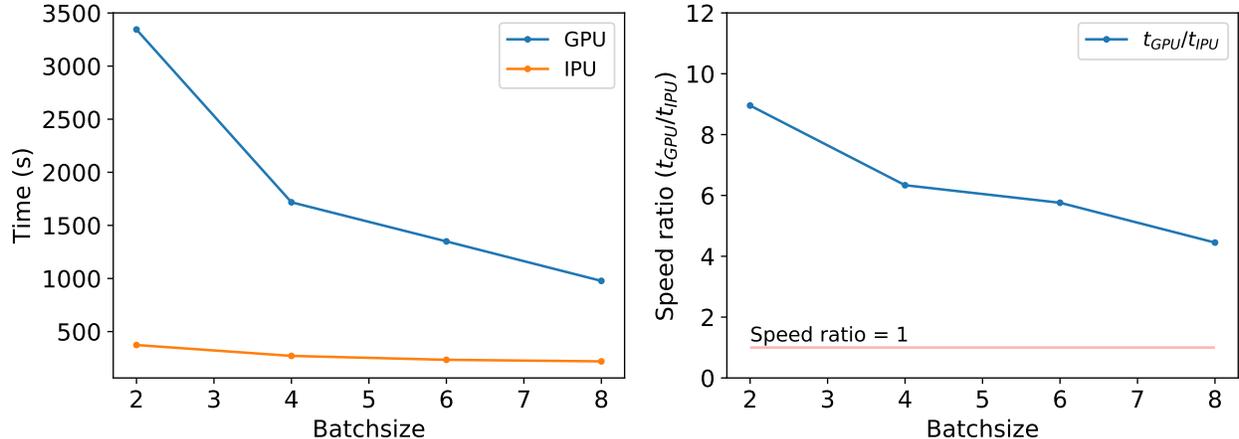


Figure A.3: (left) Training time for a deep BNN performing galaxy parameter estimation as a function of the batch size. (right) Speed ratio of GPU and IPU as a function of the batch size. The red horizontal line represents the value $R = 1$, i.e. the network takes the same amount of time for training in both cases. The neural network is composed of 2.7M trainable parameters.

variational inference (Kingma and M. Welling, 2014, Kingma, Salimans, and Max Welling, 2015 and Blundell et al., 2015).

Since the training consists in learning distribution parameters, the number of trainable parameters increases to 2.7M. Training BNNs also adds a term in the loss: the sum of all the Kullback-Leibler (KL) divergences between the parameter posterior distribution and the prior distribution. Here, this prior distribution is chosen to be the normal distribution, a common choice in literature. The loss is now defined as

$$\mathcal{L} = -\log(p(x)) + \sum_i KL(p(w_i|x)||p(z)) \quad (\text{A.2})$$

Here w_i is the i -th parameter in the network and $p(z)$ is the prior distribution.

Same as previously, this network is trained during 100 epochs on the same images sample, with a 10 epochs warm-up phase.

As expected, the memory size limitation restrains the training of the BNN to lower batch sizes than with the deterministic network. The BNN is trained with batch sizes varying from 2 to 8. IPUs once again outperform GPUs (see fig. A.3) with a training speed at least four time superior.

To summarise, training artificial neural networks on IPUs is much faster than on GPUs. However, the memory size of IPUs becomes a limitation for large networks as it restrains their training to small batch sizes. This might become a problem in some cases, if the batch sizes are too small for the learning to converge. In order to go to higher batch sizes, it would be necessary to split the network over several IPUs.

A.3.3 Galaxy image generation

Generating large amount of mock data is a major step to test and improve analysis pipelines. Until now, most galaxy simulations are based on simple analytic profiles such as Sérsic profiles (see for example Rowe et al., 2015). This is not very fast and is known to increase the risk of introducing model biases (see Mandelbaum et al., 2015).

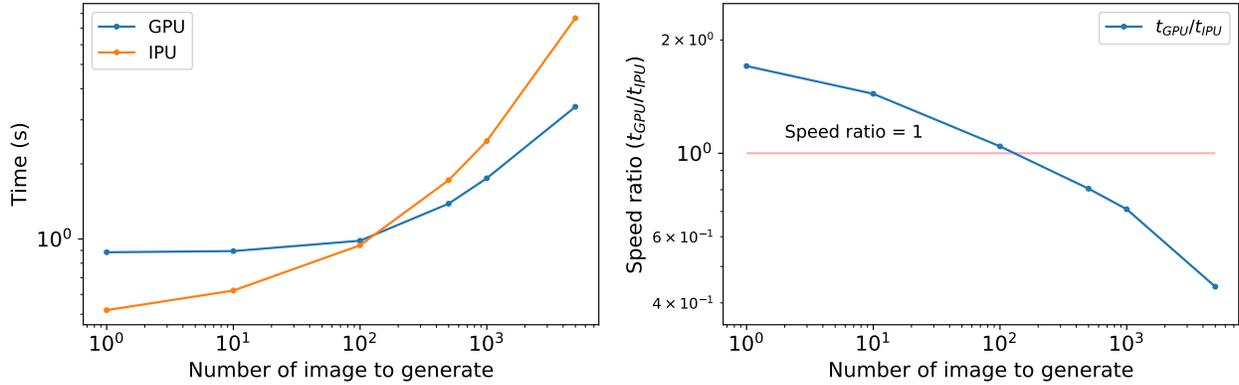


Figure A.4: (left) Inference time, sampling from a trained VAE latent space as a function of number of image to generate. (right) Ratio of inference time of GPU and IPU. The red horizontal line represents the value $R = 1$, i.e. the network takes the same amount of time for inference in both cases.

Deep learning is a natural choice to turn to when looking for fast data generation methods. Several generative models based on neural networks already exist and present interesting performance (Variational Auto-Encoder (Kingma and M. Welling, 2014), Generative Adversarial Network (Goodfellow et al., 2014) or Normalizing flow (Dinh, Krueger, and Bengio, 2014), (Jimenez Rezende and Mohamed, 2015)). In our context, they are also effective tools to learn how to generate complex light profiles and decrease model bias (Francois Lanusse et al., 2020). Using generative neural networks to model galaxy is seriously considered in different cosmological applications (Francois Lanusse et al., 2020, Regier, McAuliffe, and Prabhat, 2015 or Arcelin et al., 2021).

Generative model

In this part, the images generation is done by sampling the latent space distribution of a variational autoencoder (VAE) trained on isolated galaxy images. When training the VAE, variational inference is used to learn a latent space distribution which encodes the galaxy parameters distribution. The prior for the latent space is taken as a normal distribution. However with training, the posterior distribution of the latent space is almost always not exactly normal (see Fig.3 of Arcelin et al., 2021 for instance). It is then possible to use normalizing flows to map this distribution into a normal one thanks to a series of small invertible transformations. From there, it is possible to trivially sample the latent space distribution via the normalizing flows and use the decoder to generate galaxy images. I used Masked Autoencoder for Distribution Estimation (MADE, Germain et al., 2015) to build the normalizing flows network (similarly to Francois Lanusse et al., 2020). Here I compare image generation (or inference) speed using IPUs and GPUs. After a warm-up phase of 100 inferences, the measurement is realised on generating 1 to 5000 isolated galaxy images.

Results

Figure A.4 shows that IPUs perform better than GPUs at generating small batches of images. On the contrary, GPUs outperform IPUs when the number of images to generate increases. The choice of hardware depends on the number of images to produce, which, in turn, can depend on the usage of the generated data. To train a network, generating small batches of data on the fly

can be necessary and IPUs should be used. On the contrary if the goal is to generate a large amount of data in order to test a data processing pipeline for example, generation on GPUs is more appropriate. As an example, the CosmoDC2 (Korytov et al., 2019) catalog released which covers 440 deg^2 of the sky area, contains around 2.26 billion galaxies. To generate this many galaxies with the tested generative model, less than 18 days of computing would be required on a single V100 GPU. Scaling this process on multiple GPUs would decrease this duration even more.

A.4 Summary and discussion

Parameter estimation and simulated images generation are typical examples of tasks that will be required to process and prepare for future imaging galaxy surveys. The increasing use of neural networks and particularly BNNs in cosmology demonstrates the community’s interest in developing fast and accurate tools. These tools require particularly suited hardware. In this work I investigated IPUs and GPUs performance at training neural network and performing inference for cosmology applications.

In this study, we demonstrated that IPUs performed at least twice as fast as GPUs at training neural networks with small batch sizes. The restriction to small batch sizes due to the IPU memory size depends on the number of parameters of the network, leading to greater constraints for Bayesian neural networks. However, these constraints did not prevent the networks presented in this work to learn how to perform accurately.

We also presented performance in generating galaxy images from a trained VAE latent space, a typical inference example. Here, IPUs perform better at small batch sizes but are outperformed by GPUs at larger ones. The hardware choice depends on the task to accomplish: if the data is generated on the fly to train a neural network for example, using IPUs is probably more relevant. On the contrary, if the objective is to generate a large amount of data to test a data processing pipeline, it is more relevant to use GPUs.

To conclude, this first test of IPUs for deep learning applications in cosmology suggests that IPUs perform better than GPUs at training neural networks but, regarding inference, the choice depends on the task to realise. It must be noted that the memory size of IPUs induces a limitation for training, in network size as well as in batch size, which might be a bottleneck for BNNs. This limitation might however be alleviated with the new version of Graphcore’s IPUs. In any case, if these kind of processes are used to analyse cosmological data, scaling them on several IPUs or GPUs would probably be necessary, which might also solve the IPU memory size issue.

A.5 Acknowledgements

I am grateful to Graphcore and to Microsoft for providing cloud access to their IPU through Azure IPU Preview as well as helpful technical support. I would also like to thank the IN2P3 Computing Center (CC-IN2P3–Lyon/Villeurbanne - France), funded by the Centre National de la Recherche Scientifique, for providing access to their GPU farm. This project was partially funded by the AstroDeep project¹³ from the Agence Nationale de Recherche (ANR).

¹³<https://astrodeep.net/>

Bibliography

- Abadi, Martín et al. (Mar. 2016). “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems”. In: *arXiv e-prints*, arXiv:1603.04467, arXiv:1603.04467. arXiv: [1603.04467 \[cs.DC\]](#).
- Arcelin, Bastien et al. (in prep). In: — (Jan. 2021). “Deblending galaxies with variational autoencoders: A joint multiband, multi-instrument approach”. In: *MNRAS* 500.1, pp. 531–547. DOI: [10.1093/mnras/staa3062](#). arXiv: [2005.12039 \[astro-ph.IM\]](#).
- Blundell, Charles et al. (May 2015). “Weight Uncertainty in Neural Networks”. In: *arXiv e-prints*, arXiv:1505.05424, arXiv:1505.05424. arXiv: [1505.05424 \[stat.ML\]](#).
- Bosch, James et al. (2018). “The Hyper Suprime-Cam software pipeline”. In: *Publications of the Astronomical Society of Japan* 70.SP1, S5.
- Charnock, Tom, Laurence Perreault-Levasseur, and François Lanusse (June 2020). “Bayesian Neural Networks”. In: *arXiv e-prints*, arXiv:2006.01490, arXiv:2006.01490. arXiv: [2006.01490 \[stat.ML\]](#).
- Dinh, Laurent, David Krueger, and Yoshua Bengio (Oct. 2014). “NICE: Non-linear Independent Components Estimation”. In: *arXiv e-prints*, arXiv:1410.8516, arXiv:1410.8516. arXiv: [1410.8516 \[cs.LG\]](#).
- Germain, Mathieu et al. (Feb. 2015). “MADE: Masked Autoencoder for Distribution Estimation”. In: *arXiv e-prints*, arXiv:1502.03509, arXiv:1502.03509. arXiv: [1502.03509 \[cs.LG\]](#).
- Goodfellow, Ian J. et al. (June 2014). “Generative Adversarial Networks”. In: *arXiv e-prints*, arXiv:1406.2661, arXiv:1406.2661. arXiv: [1406.2661 \[stat.ML\]](#).
- Jia, Zhe et al. (2019). “Dissecting the Graphcore IPU Architecture via Microbenchmarking”. In: *CoRR* abs/1912.03413. arXiv: [1912.03413](#). URL: <http://arxiv.org/abs/1912.03413>.
- Jimenez Rezende, Danilo and Shakir Mohamed (May 2015). “Variational Inference with Normalizing Flows”. In: *arXiv e-prints*, arXiv:1505.05770, arXiv:1505.05770. arXiv: [1505.05770 \[stat.ML\]](#).
- Kilbinger, Martin (July 2015). “Cosmology with cosmic shear observations: a review”. In: *Reports on Progress in Physics* 78.8, 086901, p. 086901. DOI: [10.1088/0034-4885/78/8/086901](#). arXiv: [1411.0115 \[astro-ph.CO\]](#).
- Kingma, Diederik P. and Jimmy Ba (Dec. 2014). “Adam: A Method for Stochastic Optimization”. In: *arXiv e-prints*, arXiv:1412.6980, arXiv:1412.6980. arXiv: [1412.6980 \[cs.LG\]](#).
- Kingma, Diederik P., Tim Salimans, and Max Welling (June 2015). “Variational Dropout and the Local Reparameterization Trick”. In: *arXiv e-prints*, arXiv:1506.02557, arXiv:1506.02557. arXiv: [1506.02557 \[stat.ML\]](#).
- Kingma, Diederik P. and M. Welling (2014). “Auto-Encoding Variational Bayes”. In: *CoRR* abs/1312.6114.
- Korytov, Danila et al. (Dec. 2019). “CosmoDC2: A Synthetic Sky Catalog for Dark Energy Science with LSST”. In: *APJS* 245.2, 26, p. 26. DOI: [10.3847/1538-4365/ab510c](#). arXiv: [1907.06530 \[astro-ph.CO\]](#).

- Lanusse, Francois et al. (Aug. 2020). “Deep Generative Models for Galaxy Image Simulations”. In: *arXiv e-prints*, arXiv:2008.03833, arXiv:2008.03833. arXiv: [2008.03833 \[astro-ph.IM\]](#).
- LSST Science Collaboration et al. (Dec. 2009). “LSST Science Book, Version 2.0”. In: *arXiv e-prints*, arXiv:0912.0201, arXiv:0912.0201. arXiv: [0912.0201 \[astro-ph.IM\]](#).
- Mandelbaum, Rachel et al. (May 2014). “The Third Gravitational Lensing Accuracy Testing (GREAT3) Challenge Handbook”. In: *APJS* 212.1, 5, p. 5. DOI: [10.1088/0067-0049/212/1/5](#). arXiv: [1308.4982 \[astro-ph.CO\]](#).
- Mandelbaum, Rachel et al. (July 2015). “GREAT3 results - I. Systematic errors in shear estimation and the impact of real galaxy morphology”. In: *MNRAS* 450.3, pp. 2963–3007. DOI: [10.1093/mnras/stv781](#). arXiv: [1412.1825 \[astro-ph.CO\]](#).
- Mohan, Lakshan Ram Madhan et al. (2020). *Studying the potential of Graphcore IPUs for applications in Particle Physics*. arXiv: [2008.09210 \[physics.comp-ph\]](#).
- Möller, Anais et al. (Nov. 2020). “FINK, a new generation of broker for the LSST community”. In: *MNRAS*. DOI: [10.1093/mnras/staa3602](#). arXiv: [2009.10185 \[astro-ph.IM\]](#).
- Paszke, Adam et al. (Dec. 2019). “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *arXiv e-prints*, arXiv:1912.01703, arXiv:1912.01703. arXiv: [1912.01703 \[cs.LG\]](#).
- Refregier, A. et al. (Jan. 2010). “Euclid Imaging Consortium Science Book”. In: *arXiv e-prints*, arXiv:1001.0061, arXiv:1001.0061. arXiv: [1001.0061 \[astro-ph.IM\]](#).
- Regier, Jeffrey, J. McAuliffe, and Prabhat (2015). “A deep generative model for astronomical images of galaxies”. In:
- Rowe, Barnaby et al. (Apr. 2015). “GALSIM: The modular galaxy image simulation toolkit”. In: *Astronomy and Computing* 10, pp. 121–150. DOI: [10.1016/j.ascom.2015.02.002](#). arXiv: [1407.7676 \[astro-ph.IM\]](#).
- Spergel, D. et al. (May 2013). “Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA Final Report”. In: *arXiv e-prints*, arXiv:1305.5422, arXiv:1305.5422. arXiv: [1305.5422 \[astro-ph.IM\]](#).
- Valentin Jospin, Laurent et al. (July 2020). “Hands-on Bayesian Neural Networks – a Tutorial for Deep Learning Users”. In: *arXiv e-prints*, arXiv:2007.06823, arXiv:2007.06823. arXiv: [2007.06823 \[cs.LG\]](#).

Appendix B

A Bayesian Convolutional Neural Network for Robust Galaxy Ellipticity Regression

A Bayesian Convolutional Neural Network for Robust Galaxy Ellipticity Regression

Claire Theobald¹ (✉), Bastien Arcelin³, Frédéric Pennerath², Brieuc Conan-Guez², Miguel Couceiro¹, and Amedeo Napoli¹

¹ Université de Lorraine, CNRS, LORIA, F-54000 Nancy France
{[claire.theobald](mailto:claire.theobald@loria.fr), [miguel.couceiro](mailto:miguel.couceiro@loria.fr), [amedeo.napoli](mailto:amedeo.napoli@loria.fr)}@loria.fr

² Université de Lorraine, CentraleSupélec, CNRS, LORIA, F-57000 Metz France
frederic.pennerath@centralesupelec.fr, brieuc.conan-guez@univ-lorraine.fr

³ Université de Paris, CNRS, Astroparticule et Cosmologie, F-75013 Paris, France
arcelin@apc.in2p3.fr

Abstract. Cosmic shear estimation is an essential scientific goal for large galaxy surveys. It refers to the coherent distortion of distant galaxy images due to weak gravitational lensing along the line of sight. It can be used as a tracer of the matter distribution in the Universe. The unbiased estimation of the local value of the cosmic shear can be obtained via Bayesian analysis which relies on robust estimation of the galaxies ellipticity (shape) posterior distribution. This is not a simple problem as, among other things, the images may be corrupted with strong background noise. For current and coming surveys, another central issue in galaxy shape determination is the treatment of statistically dominant overlapping (blended) objects. We propose a Bayesian Convolutional Neural Network based on Monte-Carlo Dropout to reliably estimate the ellipticity of galaxies and the corresponding measurement uncertainties. We show that while a convolutional network can be trained to correctly estimate well calibrated aleatoric uncertainty, -the uncertainty due to the presence of noise in the images- it is unable to generate a trustworthy ellipticity distribution when exposed to previously unseen data (i.e. here, blended scenes). By introducing a Bayesian Neural Network, we show how to reliably estimate the posterior predictive distribution of ellipticities along with robust estimation of epistemic uncertainties. Experiments also show that epistemic uncertainty can detect inconsistent predictions due to unknown blended scenes.

Keywords: Bayesian Neural Networks · Convolutional Neural Networks · Epistemic uncertainty · Uncertainty calibration · Cosmology.

1 Introduction

One of the goals of large galaxy surveys such as the *Legacy Survey of Space and Time* (LSST, [16]) conducted at the Vera C. Rubin Observatory is to study *dark*

¹ The first author is preparing a PhD thesis at the LORIA Lab in the context of the AstroDeep Research Project (<https://astrodeep.pages.in2p3.fr/website/projects/>) funded by ANR under the grant ANR-19-CE23-0024.

energy. This component of unknown nature was introduced in the current cosmological standard model to explain the acceleration of the Universe expansion. One way to probe dark energy is to study the mass distribution across the Universe. This distribution mostly follows the dark matter distribution, which does not interact with baryonic matter (i.e. visible matter) except through gravitation, as dark matter represents around 85% of the matter in the Universe. Consequently, cosmologists need to use indirect measurement techniques such as *cosmic shear*, which measures the coherent distortion of background galaxies images by foreground matter due to *weak gravitational lensing* [17]. In astrophysics, gravitational lensing is the distortion of the image of an observed source, induced by the bending of space-time, thus of the light path, generated by the presence of mass along the line of sight. The mass acts like a lens, in partial analogy with optical lenses, as illustrated in Fig.1a. The weak gravitational lensing effect is faint (1% of galaxy shape measurement) and only statistical tools provide a way to detect a local correlation in the observed galaxies orientations. This correlation yields a local value at every point of the observable Universe, defining the cosmic shear field. As pictured in Fig.1b, in an isotropic and uniform Universe orientations of galaxies are expected to follow a uniform distribution (left panel). The statistical average of their oriented elongations, hereafter called *complex ellipticities*, is expected to be null. In presence of a lens, a smooth spatial deformation field modifies coherently the complex ellipticities of neighboring galaxies so that their mean is no longer zero (right panel).

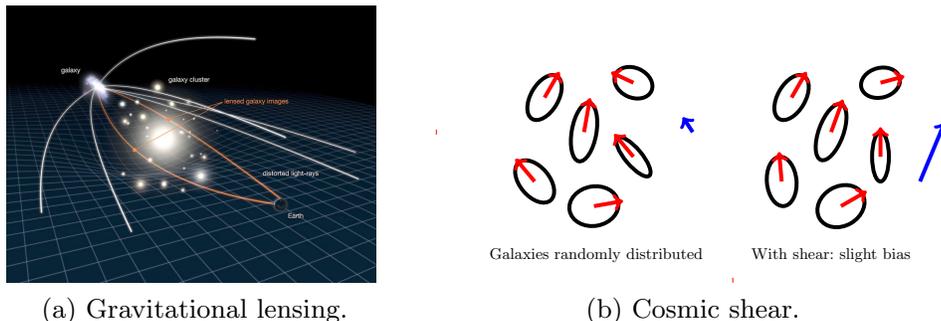


Fig. 1. (a) Effect of gravitational lensing: the mass bends the light and deforms the images of the galaxies. (b) Weak lensing: the correlation between orientations and shapes of neighbour galaxies defines the cosmic shear. In blue: average ellipticity. Left: the expected ellipticity distribution. Right: the observed ellipticity distribution. Image: (a) NASA/ESA

The unbiased measurement of cosmic shear is a major ambition of nowadays cosmology [21]. One avenue to estimate the cosmic shear locally is to combine individual galaxy ellipticity measurements. By looking deeper into the sky, that is to older objects, the next generation of telescopes will allow for the detection

of a very large number of galaxies, potentially leading to very precise shear measurement and resulting in tight constraints on dark energy parameters.

Methods already exist to estimate galaxy ellipticities through direct measurement on images recorded by telescope cameras ([13] for example). This is a complex problem as, among other things, the shear signal is carried by faint galaxies which makes it very sensitive to background noise. Another central issue for current and coming surveys in galaxy shape determination, is the treatment of statistically dominant overlapping objects, an effect called *blending*. A current survey projects that 58% of the detected objects will appear blended [18] and this value is expected to reach around 62% for LSST [19]. To overcome this issue, solutions exist such as deblending [22–24]: the separation of overlapping objects. Yet, they are not perfect and rely on an accurate detection of blended scenes which is also a complex problem. As such, in addition to a precise estimation of the complex ellipticities, a reliable measurement of the uncertainties is crucial in order to discard, or at least decrease the impact of, unreliable and inaccurate measurements avoiding as much as possible the introduction of a bias into the shear estimation.

Classical ellipticity measurement methods usually adopt assumptions about the shape of the galaxies (for example via the shape of the window function in [13]) potentially resulting in model bias. In contrast, *convolutional neural networks* or CNNs [2] make it possible to learn and recognize complex and diverse galaxy shapes directly from data without making any other hypothesis than the representativeness of the training sample. They consequently are appropriate tools to learn the regression of galaxy ellipticities, even in the presence of noise and complex distortions. Yet standard CNNs can only measure the *aleatoric uncertainty*: the one due to the presence of noise in the data. They are unable to estimate the *epistemic uncertainty*, the one due to the limited number of samples a CNN has been trained with and to the model [1, 9]. This second type of uncertainty is essential to detect outliers from the training samples, or formulated accordingly to our problem, to distinguish between reliable or unreliable galaxy ellipticity estimation. It is only accessible by considering neural network weights as random variables instead of constants, that is, by adopting a Bayesian approach. Consequently, we have focused our work on *Bayesian Deep Learning* [11] using Monte Carlo dropout (MC dropout) [1] as the mean to apply Bayesian inference to Deep Learning models.

Foreseeing a Bayesian estimation of the cosmic shear, combining galaxy ellipticity posteriors estimated directly from images (with blends or not) in different filters (or bands), this paper focuses on estimating reliable galaxy ellipticity posteriors from single band images. This is a necessary step to check that the proposed method efficiently estimates a calibrated aleatoric uncertainty and is able to minimize the impact of wrongly estimated ellipticity values due to outliers in the computation of the shear. We compare two networks trained on isolated galaxy images with or without noise in order to test for the calibration of aleatoric uncertainty. Regarding outliers, blended scenes are perfect examples. Note that these are illustrations of aleatoric or epistemic uncertainty sources.

Most of cosmic shear bias sources such as detection, Point-Spread-Function (PSF) treatment, or selection for example [17, 21], can fall in one or the other category. The estimation of galaxy ellipticity posterior from blended scenes in different bands is a harder problem that we will investigate in further work.

The contributions of this article are 1) to propose a Bayesian Deep Learning model that solves a complex multivariate regression problem of estimating the galaxy shape parameters while accurately estimating aleatoric and epistemic uncertainties; 2) to establish an operational protocol to train such a model based on multiple incremental learning steps; and 3) to provide experimental evidences that the proposed method is able to assess whether an ellipticity measurement is reliable. This is illustrated, in this paper, by the accurate differentiation between isolated galaxy or blended scenes, considered here as outliers, and the relationship between epistemic uncertainty and predictive ellipticity error. We also show that this last result could not be obtained with a classical, non Bayesian network.

The rest of the paper is organized as follows. In Section 2 we briefly describe the problem to be solved and comment on some of its peculiarities. We detail our proposed solution in Section 3. We analyse the results obtained on the various experiments we performed in Section 4, and we conclude and give the directions of further research in Section 5.

2 Estimating galaxy ellipticity from images

As mentioned previously, it is possible to estimate cosmic shear combining individual measurements of galaxy shape. This shape information can be quantified by the complex ellipticity, which can be defined in cosmology as in Def. 1.

Definition 1. *Let E be an ellipse with major axis a , minor axis b , and with θ as its position angle. The complex ellipticity of E is defined as:*

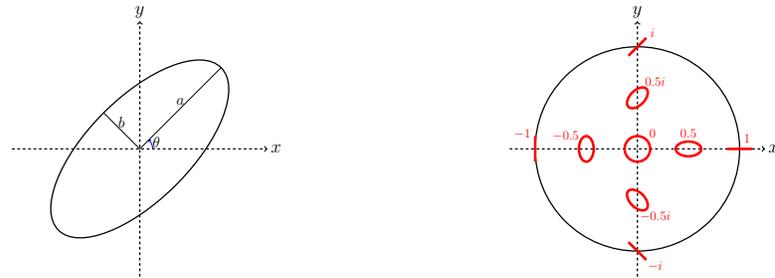
$$\epsilon = \epsilon_1 + \epsilon_2 i = \frac{1 - q^2}{1 + q^2} e^{2i\theta}, \quad (1)$$

where $q = \frac{b}{a}$ is the axis ratio of the ellipse.

An illustration of the ellipticity parameters is shown in Fig. 2a. The complex ellipticity defines a bijection between the orientation and the elongation of the ellipse on one side, and the unit disk on the other side, see Fig. 2b.

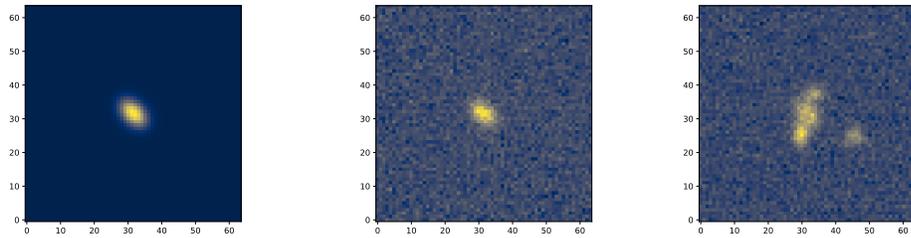
However, the process to achieve an unbiased measurement of cosmic shear, starting with the estimation of ellipticities, is going to be challenging for several reasons. We test the reliability of our networks prediction on noise and blending, two of the many possible bias sources in the cosmic shear estimation. Both of these issues result from the fact that the shear signal is mostly carried by faint galaxies. By definition, these objects have a low signal-to-noise ratio. The noise corrupts the galaxy images, making the shape estimation much harder (see Fig. 3b), and can introduce a bias in shear measurement [17].

Also, a large part of these faint objects will appear blended with foreground galaxies. Even in scenes where objects are only slightly overlapped, the apparent



(a) Ellipticity parameters: major axis a , minor axis b , position angle θ . (b) Bijective mapping between ellipse shapes and complex ellipticities.

Fig. 2. Geometric representation of the complex ellipticity. (a) The ellipse parameters. (b) The complex ellipticity defines a bijection between ellipse shapes and the unit disk. An ellipticity with low magnitude is close to a circle, while one with a high magnitude is closer to a straight line. The argument defines the orientation of the ellipse



(a) Isolated noiseless galaxy (b) Isolated noisy galaxy (c) Blended noisy galaxies

Fig. 3. Three different types of image complexity for the same galaxy: isolated without noise, isolated with noise, blended with noise. Notice how the noise slightly deforms the galaxy (b) and how the blended galaxies makes the ellipticity estimation very difficult (c) when compared to a simple isolated galaxy without noise (a)

shape of the detected object does not correspond to a single galaxy model and an ellipticity measurement on this image could give a completely wrong result. Again, this work is the first step of a longer-term goal. Here, we target a reliable estimation of galaxy ellipticity posterior from single band images. This includes obtaining a well calibrated aleatoric uncertainty, tested here with and without the addition of Poisson noise on images, and an epistemic uncertainty allowing for minimization of the impact of untrustworthy measurement due to outliers (here, blended scenes).

We simulate LSST-like images, allowing us to control the parameters of the scene, e.g., the number of galaxies, their location on the image, and the level and type of noise applied. We consider four categories of simulated data: isolated centered galaxies without noise, isolated centered galaxies with noise, and blended scene with and without noise. Images are 64x64 pixels stamps simulated in the brightest of the six bands corresponding to the LSST filters, each of them selecting

a different part of the electromagnetic spectrum. These images are simulated placing, in their center, a galaxy whose ellipticity is to be measured.

The image generating process relies on the GalSim library [14] and is based on a catalog of parametric models fitted to real galaxies for the third Gravitational Lensing Accuracy Testing (GREAT3) Challenge [20]. It consists in 1) producing an image of a centered noiseless isolated galaxy from a model sampled randomly from the catalog, with its corresponding physical properties (size, shape, orientation, PSF, brightness, redshift, etc) 2) measuring the complex ellipticity of the galaxy with the KSB algorithm [13] on the image and record it as the image label, 3) possibly adding on random image location other galaxy images (from 0 to 5) to generate blended scenes 4) possibly adding Poisson noise (as in [24]). In this study and for sake of interpretability, we only provide as input to our CNN the reference band (the brightest) which we use to define the target ellipticity, making our images two-dimensional. Once again, while using multiple bands is useful for blended galaxies [23, 24], here we focus only on predicting the ellipticity of a single centered galaxy with a correctly estimated uncertainty.

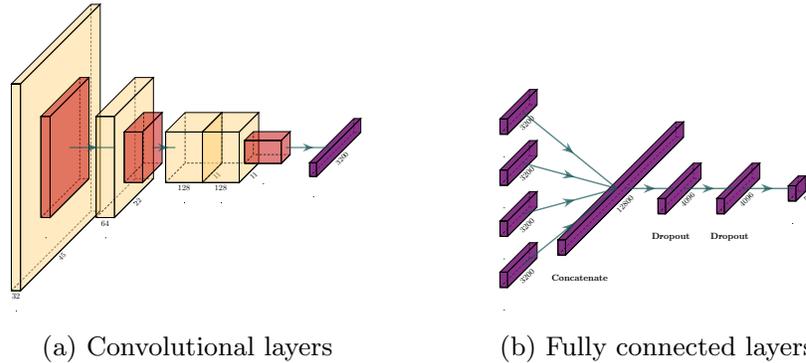


Fig. 4. Convolutional neural network architecture. (a) The input after augmentation has dimensions $45 \times 45 \times 1$. Each convolutional block starts with a batch normalization layer and has a PReLU activation. The first convolutional layer is of dimension $45 \times 45 \times 32$ with a 5×5 kernel size (in yellow), followed by a 2×2 Max-Pooling operation (in orange). The second convolutional layer is $22 \times 22 \times 64$ with kernel size 3×3 , followed by a 2×2 Max-Pooling operation. Then, we add two $11 \times 11 \times 128$ convolutional layers with a 3×3 kernel that ends with a final 2×2 Max-Pooling operation, and the resulting feature maps are flattened into a 3200 fully connected layer (in purple). (b) Each augmented image gives a 3200 fully connected layer convolutional output (all augmented images share the same convolutional layers and filters), which are then concatenated into a 12800 fully connected layer. The two final layers have 4096 neurons in the case of an MVN regression, 2048 else; with Maxout activation [15] and dropout with a rate of 0.5. The output layer has 5 neurons in the case of an MVN regression

3 A method to assess uncertainty in ellipticity estimation

3.1 Estimation of noise related uncertainty

Our first goal is to reliably estimate the first layer of complexity in the galaxy images, the noise. Given the nature of the data, we will be using a CNN [2]. However, training a CNN to solve a standard regression problem with an L2 loss does not allow us to estimate the uncertainty due to the noise. Therefore, in place of a complex scalar output, we predict a 2D *multivariate normal distribution* (MVN) as an output: given an input image X , whose complex ellipticity is denoted Y and given weight parameters w , the network outputs an MVN $Y \sim \mathcal{N}(\mu(X, w), \Sigma(X, w))$. As such, the model is no longer trained on a simple L2 loss but rather on the log-likelihood of the MVN. The mean of the distribution $\mu(X, w)$, which is also the mode, serves as the predicted output, and the covariance matrix $\Sigma(X, w)$, which is also an output of the network, represents the so-called aleatoric uncertainty on the input data X . The model is therefore heteroscedastic, as $\Sigma(X, w)$ depends on the input X [9]. This allows our model to estimate the aleatoric uncertainty for each image individually. The determinant of $\Sigma(X, w)$, denoted $|\Sigma(X, w)|$, is a scalar measure of uncertainty, as it is directly related to the differential entropy, $\ln\left(\sqrt{(2\pi e)^2|\Sigma(X, w)|}\right)$, of an MVN.

The architecture of our network is inspired by the work of Dielman, who proposed a simple model specifically tuned for the Galaxy Zoo challenge, therefore adapted to our data [12]. Each image is augmented in four different parts by cropping thumbnails from high resolution images, centered on spatial modes of light profile. Then each augmented image is fed to the CNN. The complete architecture is explained in Fig. 4. More details on the training process are explained in Section 3.3. Results obtained with this model are given in Section 4.1.

3.2 Estimation of blend related uncertainty

As seen in Section 2, estimating the uncertainty due to the noise in the data is only one part of the problem. An estimated 60% of the images represent blended scenes, for which a direct estimation of ellipticity does not make sense in the context of this work. The uncertainty related to the blended images cannot be estimated simply with the variance of the MVN distribution. Indeed, in the case of a blended scene image, the network is not uncertain because of the noise but rather because this kind of images is not part of the training sample. This can be characterized by the epistemic uncertainty.

This uncertainty can be estimated using a Bayesian Neural Network (BNN), which assumes a probability distribution on the weights W of the network instead of a single point estimate [11]. Given a prior $p(w)$ on W and a set $\mathcal{D} = \{(X_i, Y_i)\}_i$ of observations, the resulting posterior distribution $p(w|\mathcal{D}) \propto p(\mathcal{D}|w)p(w)$ is analytically impossible to compute. A variational Bayes optimization method

is necessary to derive an approximate posterior $q_\theta(w)$ parameterized by hyperparameters θ . In MC dropout [1, 3], the considered search space includes all approximate posteriors resulting from applying *dropout* [7], i.e multiplying every neuron output (of selected layers) by an independent Bernoulli variable. The dropout rate is set to the conventional value of 0.5, as this leads to an approximate posterior that can achieve well calibrated uncertainty estimates [1]. However, there are other ways to define the posteriors such as dropout rate tuning [5], or ensemble methods by training many networks [6]. During training, standard stochastic gradient descent techniques can be used, thanks to the reparameterization trick, to search for an approximate posterior maximizing locally the ELBO [1]. During testing, the *posterior predictive distribution* $p(Y|X, \mathcal{D})$ for some input X can be estimated using Monte Carlo sampling:

$$p(Y|X, \mathcal{D}) \approx \int p(Y|X, w)q_\theta(w)dw \approx \frac{1}{K} \sum_{k=1}^K p(Y|X, w_k), \quad (2)$$

where $(w_k)_{k=1}^K \sim q_\theta(w)$ refer to weights of K independent dropout samples.

In the case of a multivariate regression problem like ours, every distribution $p(Y|X, w_k)$ is a MVN so that the resulting posterior predictive distribution in Eq. 2 is a Gaussian mixture of order K . The uncertainty underlying this mixture can be summarized by its covariance matrix $\Sigma_{pred.}(X, \mathcal{D}) = \text{Cov}(Y|X, \mathcal{D})$. This matrix accounts for both aleatoric and epistemic uncertainties, whose respective contributions can actually be separated in a way that generalizes the variance decomposition described in Depeweg [10]:

$$\Sigma_{pred.}(X, \mathcal{D}) = \Sigma_{aleat.}(X, \mathcal{D}) + \Sigma_{epist.}(X, \mathcal{D}), \quad (3)$$

where the first term represents the aleatoric uncertainty and can be computed as the mean of the covariance matrices for each of the K output samples:

$$\begin{aligned} \Sigma_{aleat.}(X, \mathcal{D}) &= \mathbb{E}_{W|\mathcal{D}}(\text{Cov}(Y|X, W)), \\ &\approx \frac{1}{K} \sum_{k=1}^K \Sigma(X, w_k). \end{aligned} \quad (4)$$

while the second represents the epistemic uncertainty and is estimated as the empirical covariance matrix of the K mean vectors produced as outputs:

$$\begin{aligned} \Sigma_{epist.}(X, \mathcal{D}) &= \text{Cov}_{W|\mathcal{D}}(\mathbb{E}(Y|X, W)), \\ &\approx \frac{1}{K} \sum_{k=1}^K (\mu(X, w_k) - \mu(X))(\mu(X, w_k) - \mu(X))^T. \end{aligned} \quad (5)$$

where $\mu(X) = \frac{1}{K} \sum_{k=1}^K \mu(X, w_k)$.

Matrix $\Sigma_{epist.}(X, \mathcal{D})$ defines the epistemic uncertainty as the covariance matrix of the mean vectors over the posterior. This uncertainty will be high if the sampled predictions from each model vary considerably with respect to W .

This would mean that no consistent answer can be deduced from the model and therefore it would be highly uncertain.

Finally when the context requires to reduce these uncertainty matrices to uncertainty levels so that they can be compared, their determinants are used to define two corresponding scalar quantities:

$$\mathcal{U}_{aleat.}(X, \mathcal{D}) = |\Sigma_{aleat.}(X, \mathcal{D})| \quad \text{and} \quad \mathcal{U}_{epist.}(X, \mathcal{D}) = |\Sigma_{epist.}(X, \mathcal{D})|.$$

3.3 Training protocol

In order to train a BNN with an MVN output, the model needs to learn both the mean and the covariance matrix. The network’s training diverges when trying to learn both at the same time, forcing us to separate the training into two steps. First, we train a simple neural network without an MVN output - we use only two output neurons representing the mean - using a L2 loss. Then, we transfer the filters of the convolutional layers into the model with an MVN output, but reinitialize the fully connected layers. This allows the model to converge smoothly as the mean of the MVN distribution has already been learned, allowing the covariance matrix to be calibrated accordingly.

This protocol works well when training on noiseless images of isolated galaxies but fails when training on noisy images. Indeed, overfitting occurs during the training of the network without MVN. When transferring the filters to the MVN model, the mean of the MVN is not well calibrated enough and the training of the BNN diverges. To fix this, we adjust the protocol for the model without MVN, adding noise incrementally during training: we first submit noiseless images, and modify 5% of the sample, switching from noiseless to noisy images, every 50 epochs for 1000 epochs. This prevents overfitting and allows the MVN model to converge after the transfer.

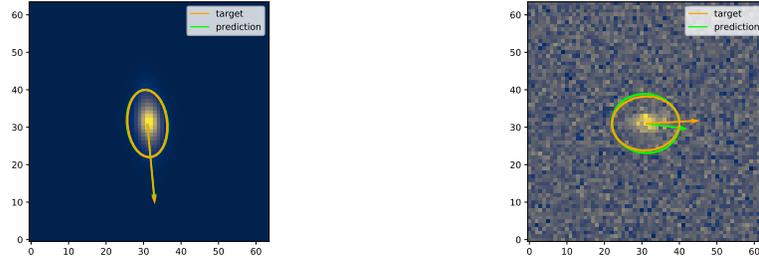
4 Experiments

4.1 Estimation of uncertainty related to noise

In this section we show that using an MVN as an output allows for a reliable and well calibrated estimation of the aleatoric uncertainty, i.e. uncertainty related to the noise in the data.

In order to show that estimating the ellipticity of galaxies in the presence of background noise is complex and can induce incorrect predicted ellipticity values, we first train two simple CNNs without an MVN output: one on noiseless images and one on noisy images, accordingly tested on noiseless and noisy images respectively. Figure 5 shows the images of galaxy with their target complex ellipticity superimposed, as well as the predicted one.

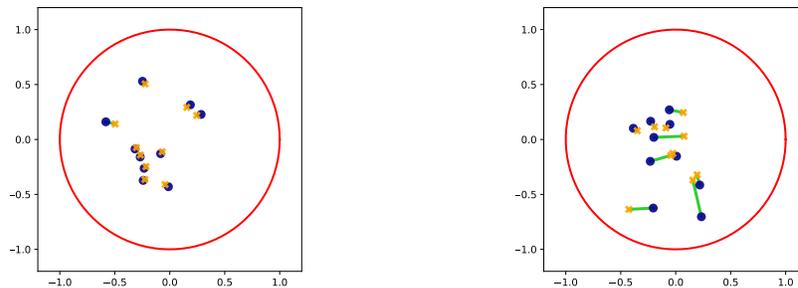
The ellipse represents the estimated shape - with a fixed scale adapted for visualization - and the arrow is the corresponding complex ellipticity - modified with half its argument in order to be aligned with the main axis of the ellipse. On this example, we can qualitatively see that the galaxy ellipticity on the noisy



(a) Predicted ellipticity without noise (b) Predicted ellipticity with noise

Fig. 5. Galaxy images with the predicted ellipticity superimposed on them. The arrow and the corresponding elliptic shape are rendered in an arbitrary scale for visualization purposes. In orange: the true ellipticity. In green: the predicted ellipticity

image is harder to estimate as the noise deforms the shape of the galaxy. Figure 6 generalizes this observation as it shows a sample of the predicted ellipticities on the complex plane within the unit circle, with the target ellipticity and the difference between predicted and targeted values.

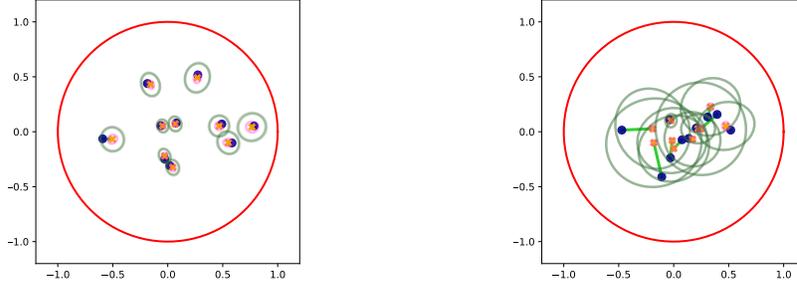


(a) Predicted ellipticities without noise (b) Predicted ellipticities with noise

Fig. 6. Predicted ellipticities on the complex plane. In red: unit circle. In yellow: predicted ellipticities. In blue: target ellipticities. In green: difference between true and predicted values

While the model trained on noiseless data performs really well (Fig. 6a), it cannot achieve the same level of performance when trained on noisy data, losing part of its reliability (Fig. 6b). As such, using a simple CNN without any estimation of aleatoric uncertainty is not satisfying for our application.

We now train two Bayesian Convolutional Neural Networks with an MVN distribution to estimate both epistemic and aleatoric uncertainties, as seen in Section 3.2. Like the simple CNN models, we show in Fig. 7, the ellipticities estimated from the BNNs on the complex plane. We also add the 90% confidence ellipses of both epistemic, aleatoric and predictive uncertainties. We observe that in both cases, the epistemic uncertainty is low if not negligible, meaning that



(a) Predicted ellipticities without noise (b) Predicted ellipticities with noise

Fig. 7. Predicted ellipticities on the complex plane. In red: unit circle. In yellow: predicted ellipticities. In blue: target ellipticities. In light green: difference between true and predicted values. In pink: 90% epistemic confidence ellipse. In dark green: 90% aleatoric confidence ellipse. In grey: 90% predictive confidence ellipse

the model is confident in its predictions. Put another way, all K pairs of outputs $\mu(X, w_k)$ and $\Sigma(X, w_k)$ are roughly equal to their mean, respectively $\mu(X, w)$ and $\Sigma_{aleat.}(X, w)$, so that, according to Eq. 3 and Eq. 5, $\Sigma_{epist.}(X, w) \approx 0$ and $\Sigma_{pred.}(X, w) \approx \Sigma_{aleat.}(X, w)$. The aleatoric uncertainty is low for noiseless images but higher for noisy ones, confirming that the noise corrupting galaxy images makes it more difficult for the model to consistently give an accurate ellipticity estimation.

Finally, in order to see if the MVN distribution is well calibrated, we standardize the output and check if the resulting distribution follows the standard distribution. More precisely, if we define:

$$Z(X, w) = \Sigma_{pred.}(X, w)^{-\frac{1}{2}}(Y - \mu(X, w)) \quad , \quad (6)$$

then the distributions of its two independent components $z_1 \sim Z(X, w)_1$ and $z_2 \sim Z(X, w)_2$ should be equivalent to the standard distribution $\mathcal{N}(0, 1)$. Note that this is true only because all K output MVNs are confounded. Figure 8 shows that the standardized distributions for the model trained on noisy images are indeed well calibrated and therefore the model is neither overestimating nor underestimating the predictive uncertainty.

4.2 Estimation of uncertainty related to blending

In the previous part we showed that our BNNs are well calibrated. Here we submit outliers to the networks in order to study the impact on epistemic uncertainty and whether it can be used to detect them. Our models have only been trained on images of isolated galaxies, but astrophysical images can contain multiple overlapped galaxies. In that case, asking the model to measure a single ellipticity does not make sense. If the epistemic uncertainty behaves as expected, then its measurement would allow us to detect when a predicted ellipticity is incorrect due to the presence of multiple galaxies in the image. We fed images of blended

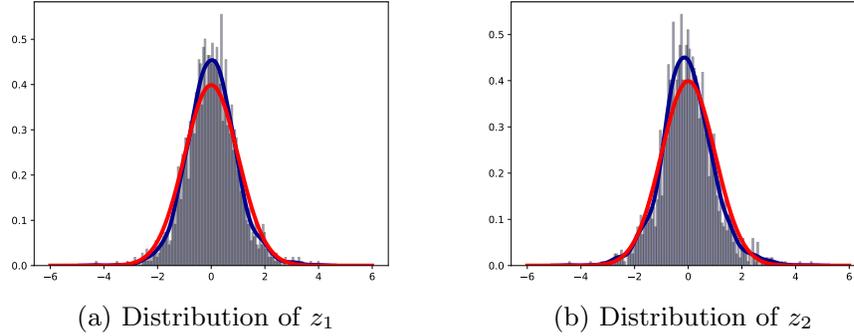


Fig. 8. Histogram of the standardized distributions on the model trained with noisy images. In red: standard bell curve. In blue: histogram of the standardized distribution with the smoothed curve.

scenes to the two models trained on isolated galaxies (with or without noise), adding noise to the blended scenes only for the model trained on noisy images.

Results shown in Fig. 9 demonstrate that in both cases the predictions are particularly inexact when compared to the target ellipticity of the central galaxy. Also, and as expected, the epistemic uncertainty is much higher for these blended scenes than for isolated galaxy images. However, the aleatoric uncertainty gives incoherent values as the model has not been trained to evaluate it on blended images: notice how the aleatoric ellipses are more flattened with a lower area. Figure 10 permits to visualise the behavior of the epistemic uncertainty. It shows how the ellipticities sampled with dropout slightly diverge compared to the mean prediction. Here the model cannot give a consistent answer and therefore its prediction should be deemed untrustworthy.

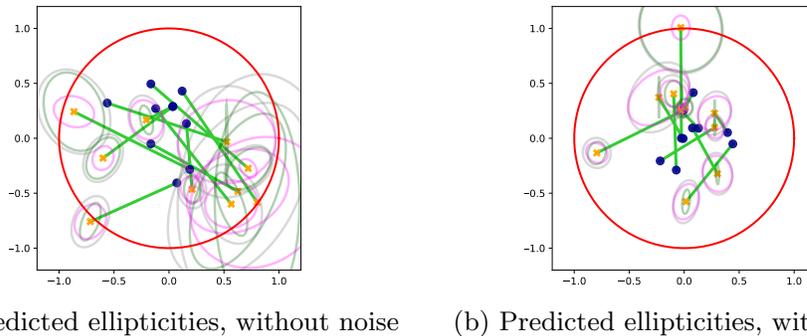


Fig. 9. Predicted ellipticities on the complex plane for blended galaxies images. In red: unit circle. In yellow: predicted ellipticities. In blue: target ellipticities (label of the centered galaxy). In light green: difference between true and predicted values. In pink: 90% epistemic confidence ellipse. In dark green: 90% aleatoric confidence ellipse. In grey: 90% predictive confidence ellipse

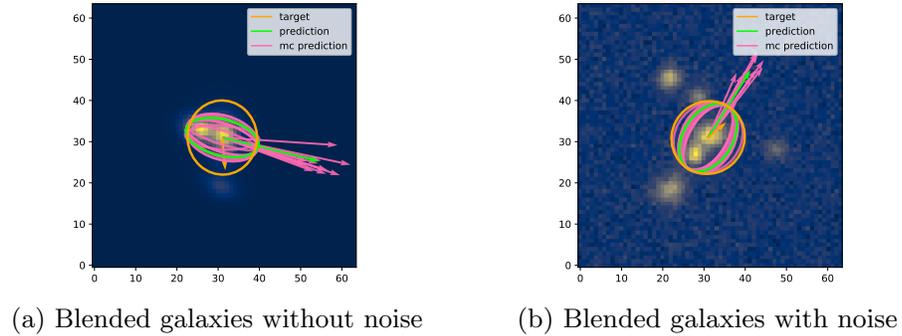
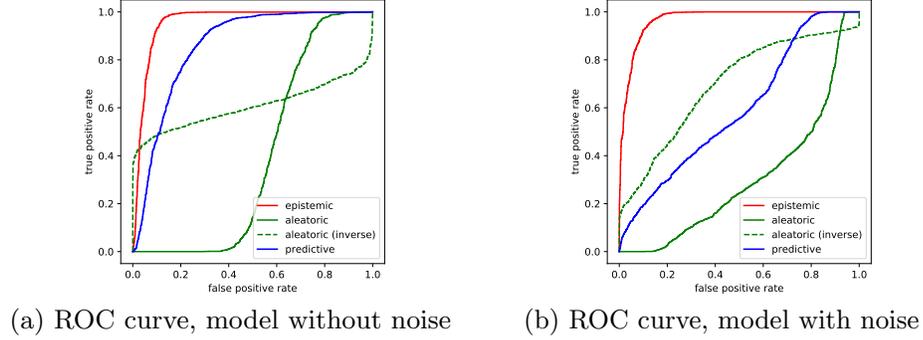


Fig. 10. Blended galaxies images with the predicted ellipticity superimposed on them. The arrow and the corresponding elliptic shape are rendered in an arbitrary scale for visualization purposes. In orange: the true ellipticity (label for the galaxy in the center). In green: the predicted ellipticity. In pink: the individual MC dropout predicted ellipses. The green ellipticity is therefore the mean of the pink ones. On both images the prediction is uncertain as the individual MC samples slightly diverge from the mean

To quantify the quality of the epistemic uncertainty when it comes to detecting incoherent predictions due to outliers, we computed the ROC curves for each uncertainty type. More precisely, we reduce each covariance matrix (aleatoric, epistemic and predictive) to a scalar by computing its determinant. We interpret these estimates as a scoring function to assess whether an image is an outlier, i.e. a blended image: the higher the score, the more likely the image contains a blend. Finally, we compute for each of these scoring functions its ROC curve. We repeat that process for both networks trained with noisy and noiseless data. The results are shown in Fig. 11.

These ROC curves are also summarized by their associated Area Under Curve (AUC) on Tab. 11c. The epistemic uncertainty clearly appears as the most consistent “metric” to detect outliers and therefore to give useful information about the confidence in the model predictions. Even the predictive uncertainty performs worse than the epistemic one. This is especially true in the presence of noise since the aleatoric uncertainty then occupies a more important part of the predictive one compared to the noiseless case. Notice that the aleatoric ROC curve is mostly below the diagonal with an AUC below 0.5, meaning it performs worse than a random classifier. This is due to the fact that the model has not been trained to evaluate aleatoric uncertainty on blended scenes. As seen in Fig. 9, the aleatoric ellipses are more flattened in the blended cases, meaning its determinant is lower. Thus the aleatoric uncertainty is on average lower on blended scenes when compared to isolated ones.

To compensate, results of the complementary classifier for the aleatoric uncertainty are shown. It is still not as satisfying as the epistemic uncertainty. While using epistemic uncertainty to identify inconsistent predictions due to a lack of knowledge is highly effective, we note that few blended images still have



Uncertainty	AUC noiseless	AUC noise
Epistemic	0.956	0.969
Aleatoric	0.394	0.306
Aleatoric (inverse)	0.606	0.694
Predictive	0.856	0.594

(c) AUC values

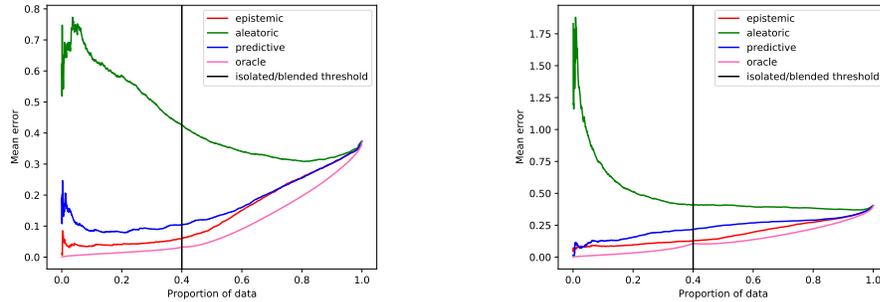
Fig. 11. ROC curves for detecting outliers for aleatoric, epistemic and predictive uncertainty. (a) ROC curve, model without noise. (b) ROC curve model with noise. Since the aleatoric ROC curve gives incoherent answers on outliers (see Fig. 9), we also plot the complementary classifier as a dashed line. (c) AUC values for all uncertainties, for the model with and without noise. Here, the epistemic uncertainty is clearly the best to detect outliers, as its AUC value is close to 1 in both noisy and noiseless datasets

low epistemic uncertainty due, for instance, to a large galaxy that obstructs all of the other ones, making the image actually closer to an isolated galaxy image.

Finally, we evaluate how each type of uncertainty is a reliable representation of the risk of error in ellipticity prediction. Unfortunately, in the presence of blended images, the predictive distribution is no longer a simple MVN but a mixture of K well separated Gaussian distributions. The normalization process that allowed us to obtain the results presented in Fig. 8 is no longer applicable here. It is still possible to study the relationship between the uncertainty and the ellipticity prediction error testing a trivial rule: the higher the uncertainty, the more important we expect the error to be. To do so, we do three sorting of the images according to each uncertainty type, from the lowest uncertainty to the highest, on a scale from 0 to 0.4 for isolated objects, and from 0.4 to 1 for blended scenes. We then compute the mean ellipticity error considering the proportion of the sorted data from 0 to 1. For blended scenes the ellipticity prediction error is computed w.r.t. the ellipticity of the centered galaxy. We repeat this experiment twice, for networks trained on noiseless and noisy data. Finally we add an "oracle" curve where the data is sorted directly according to the ellipticity prediction error which represents a perfect sorting. Results are shown in Fig. 12.

Once again, epistemic uncertainty proves to be best suited to anticipate ellipticity predictive error. The samples with the lowest epistemic uncertainty have the lowest mean ellipticity error and conversely, while samples with low

aleatoric uncertainty can already have high mean error. Consequently, on real astrophysical data, when the predictive ellipticity error is obviously unknown, relying on the epistemic uncertainty to reject, or minimize the impact of, a sample because of its probable predictive error is the best way to go.



(a) Mean error curve, model without noise (b) Mean error curve, model with noise

Fig. 12. Mean error curves w.r.t. data proportion for aleatoric, epistemic and predictive uncertainty. (a) Mean error curve without noise. (b) Mean error curve with noise. In black the threshold between the proportion of isolated galaxies: $[0, 0.4]$ and blended galaxies: $[0.4, 1]$. In pink the oracle curve, where the data is sorted by the predictive error. The closest a curve is to the oracle the better

5 Conclusion

We developed a Bayesian approach to estimate the posterior distribution of galaxy shape parameters using convolutional neural networks and MC-Dropout. In addition to a precise measurement of the ellipticities, this approach provides a calibrated estimation of the aleatoric uncertainty as well as an estimation of the epistemic uncertainty. We showed that the latter is behaving according to expectations when applied to different kind of galaxy images, and is well-suited to identify outliers and to anticipate high predictive ellipticity error. These results confirm the suitability of Bayesian neural networks for galaxy shape estimation and incite us to continue exploring their use to go from ellipticity posterior distributions, estimated from multi-band galaxy images, to cosmic shear estimation.

References

- Gal, Y., Ghahramani, Z.: Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In: ICML Proceedings 2016, JMLR Workshop and Conference Proceedings, vol. 48, pp. 1050–1059.
- LeCun, Y., Bengio, Y. and Hinton, G.: Deep learning. Nature, 521(7553), pp. 436-444 (2015).

3. Gal, Y., 2016, *Uncertainty in Deep Learning*, PhD thesis, University of Cambridge
4. Gal, Y., Hron, J., Kendall, A.: Concrete Dropout. In: NIPS Proceedings 2017, pp. 3581-3590.
5. Lakshminarayanan, B., Pritzel, A., Blundell, A.: Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In: NIPS Proceedings 2017, pp. 6402-6413.
6. Kendall, A., Gal, Y.: What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In: NIPS Proceedings 2017, pp. 5574-5584.
7. Srivastava, N., *et al.*: Dropout: a simple way to prevent neural networks from overfitting. In: J. Mach. Learn. Res., vol. 15, pp. 1929-1958 (2014).
8. Le, V. Q., Smola, J. A., Canu, S.: Heteroscedastic Gaussian process regression. In: ICML Proceedings 2005, vol. 119, pp. 489-496
9. Hüllermeier, E., Waegeman, W.: Aleatoric and Epistemic Uncertainty in Machine Learning: A Tutorial Introduction. CoRR, <http://arxiv.org/abs/1910.09457> (2019)
10. Depeweg, S., *et al.*: Decomposition of uncertainty in bayesian deep learning for efficient and risk-sensitive learning. In: ICML Proceedings 2018, vol. 80, pp. 1192-1201.
11. Hinton, G.E., van Camp, D.: Keeping the neural networks simple by minimizing the description length of the weights. In: Proceedings of the Sixth Annual ACM Conference on Computational Learning Theory, COLT 1993. pp. 5-13.
12. Dieleman S., Willett W. K., Dambre J.: Rotation-invariant convolutional neural networks for galaxy morphology prediction. In: Monthly Notices of the Royal Astronomical Society, Volume 450, Issue 2, 21 June 2015, Pages 1441-1459
13. Kaiser N., Squires G., Broadhurst T.: A Method for Weak Lensing Observations, In: *Astrophys.J.* 449:460-475 (1995)
14. Rowe B., *et al.*: GalSim: The modular galaxy image simulation toolkit, <https://github.com/GalSim-developers/GalSim> (2015)
15. Goodfellow, I.J., *et al.*: Maxout Networks. In: Proceedings of the 30th International Conference on Machine Learning, vol. 28, pp. 1319-1327 (2013)
16. LSST Science Collaboration, Abell P. A., *et al.*: LSST Science Book. 2009, arXiv, arXiv:0912.0201
17. Kilbinger, M.: Cosmology with cosmic shear observations: a review. 2015, *Reports on Progress in Physics* **78**, 086901.
18. Bosch, J., *et al.*: The Hyper Suprime-Cam Software Pipeline. 2018, *Publications of the Astronomical Society of Japan* **70**, S5.
19. Sanchez, J., Mendoza, I., Kirkby, D.P., and Burchat: Effects of overlapping sources on cosmic shear estimation: Statistical sensitivity and pixel-noise bias. 2021, *arXiv e-prints*, arXiv:2103.02078.
20. Mandelbaum, R., *et al.*: The Third Gravitational Lensing Accuracy Testing (GREAT3) Challenge Handbook. 2014, *The Astrophysical Journal Supplement Series* **212**, 5.
21. Mandelbaum, R.: Weak lensing for precision cosmology. 2018, *Annual Review of Astronomy and Astrophysics* **56**, 393.
22. Bertin, E. and Arnouts, S.: SExtractor: Software for source extraction. 1996, *Astronomy and Astrophysics Supplement Series* **117**, 393.
23. Melchior, P., Moolekamp, F., Jerdee, M., Armstrong, R., Sun, A.-L., Bosch, J., and Lupton, R.: SCARLET: Source separation in multi-band images by Constrained Matrix Factorization. 2018, *Astronomy and Computing* **24**, 129.
24. Arcelin, B., Doux, C., Aubourg, E., Roucelle, C., and LSST Dark Energy Science Collaboration: Deblending galaxies with Variational Autoencoders: a joint multi-band, multi-instrument approach. 2021, *Monthly Notices of the Royal Astronomical Society* **500**, 531.

Appendix C

Is blending always an issue ?

At the beginning of my thesis in 2018, during a meal initially planned to discuss my future projects, the discussion shifted a bit and we came at the conclusion that we were all whiskies enthusiasts. As a joke, Eric told me that I should do an appendix where I would list all the whiskies shared during the thesis. Well, Eric, and since we are close from Christmas 2021, this is my gift for you this year. This list might not be completely exhaustive but I did my best, merry Christmas !

Distillery or bottler	Edition	Origin	Comments
France			
Bercloux	<i>single malt</i>	Charente-Maritime	
Bows Distillerie	<i>Segal 60</i>	Montauban	
Ergaster	<i>Nature - 2018 N° 000</i>	Passel	
G. Miclo	<i>small batch #2</i> <i>single cask #3</i> <i>single cask #4</i>	Alsace	fût de cerise à l'eau de vie (cherry cask) fût de Bourgogne tourbé, fût de Bourgogne
Home distillers	<i>Classic pure malt</i>	Auvergne	
Celtic whisky distillerie	<i>Kornog Sauternes cask 2019</i> <i>Kornog Sant Ivy 2019</i>	Larmor Pleubian	

Distillery or bottler	Edition	Origin	Comments
	<i>Kornog Roc'h Hir</i> <i>Teir Gwech</i>		
Michel Couvreur	<i>Overaged</i>		54%
Made in Jura	<i>Rouget de Lisle Pur malt 42°</i> <i>BM signature</i>	Jura	fût de vin de paille fût de Macvin
Twelve	<i>n° 10</i> <i>n° 11</i> <i>Hématite</i>	Laguiole	eau de vie de malt eau de vie de malt
Uberach	<i>'Au-delà de la rivière'</i> <i>XIV edition</i> <i>Biersky</i> <i>X years after</i> <i>Volupte de JB...</i> <i>Rs</i> <i>R8</i> <i>cask jaune</i> <i>cask bleu</i>	Alsace	assemblage eau-de-vie de bière et de malt non filtré non filtré fûts de Banyuls, non filtré mise 2019, non filtré non filtré non filtré
Distillerie du Vercors	<i>Sequoia</i>	Vercors	tourbé
Warenghem	<i>Armorik single malt millésime 2002</i> <i>Armorik Dervenenn</i>	Lannion	
Germany			
Slyrs	<i>single malt 51</i>		
India			
Amrut	<i>Peated cask strength</i>		oak barrels
Ireland			
Writer tears	<i>Copper pot</i>		Cognac cask finish

Distillery or bottler	Edition	Origin	Comments
USA			
Thomas H. Handy	<i>Sazerac</i>	Kentucky	rye Whiskey
Sonoma	<i>Straight bourbon whiskey single barrel reserve</i> <i>Cherrywood rye whiskey</i>	California	
Scotland			
Aldelphi's	<i>Breth of the Highland</i>	Highland	
Ardberg	<i>Dark cove</i> <i>Uigeadail</i>	Islay	
Adrmore	<i>16 years old</i>	Speyside	1990, Whisky magazine
Arran	<i>Sauternes Cask</i>	Highland	
Balblair	<i>vintage 1991</i>	Highland	
Ben Riach	<i>Peated 12 years old</i>	Speyside	dark rum barrel finish
Benriness	<i>20 years old 1995</i> <i>10 years old</i>	Speyside	bottled by The Single Malts of Scotland shiraz wine cask, bottled by Old Particular
Bladnoch	<i>Exotic fruit sorbet 1990</i>	Lowland	Wemyss malt
Bowmore	<i>1994</i> <i>Double the devil</i>	Islay	bottled by Berry bro's and Rudd Ltd limited release III
Bruichladdich	<i>Edition 06.1 1990</i>	Islay	26 years old
Bunnahabain	<i>18 years</i> <i>Maritime memories 1987</i>	Islay	Wemyss malt
Cadenhead's	<i>Blend of Islay</i> <i>Blend of Campbeltown</i> <i>Blend of Highland</i> <i>Blend of Lowland</i>		
	<i>Blend of Campbeltown</i>		sherry cask

Distillery or bottler	Edition	Origin	Comments
	<i>single malt Islay aged 9 years</i> <i>Royal Brackla aged 12 years</i> <i>Miltonduff-Glenlivet aged 9 years</i> <i>12 years old blend</i> <i>Strathclyde aged 28 years</i> <i>Tullibardine aged 24 years</i> <i>Glen-spey Glenlivet aged 28 years</i>	Speyside Lowland Highland Speyside	small batch rum cask Balmenach-Glenlivet, Benrinnes, Dailuaine-Glenlivet, Glenfarclas and Invergordon grain whisky small batch small batch small batch
Coal Ila	<i>8 year, sherry cask</i> <i>18 years</i> <i>'Unpeated style' 15 years old</i>	Islay	bottled by The Single Malts of Scotland
Craigellachie	<i>'A perfect beginner'</i> <i>10 years 2007</i>	Speyside	bottled by distiller's art bottled by Mossburn
Dailuaine	<i>12 years old</i>	Speyside	sheey butt, bottled by Old particular
Deanstone	<i>9 years old 2008</i>		Bordeaux red wine cask matured
Dufftown	<i>12 years old</i>	Speyside	The manager's choice
Edinburgh Whisky Ltd	<i>Highland single malt</i>	Highland	
Glendronach	<i>Allardice 18 years old</i>	Highland	
Glenfiddich	<i>Winter storm</i>		21 years old
Glen Mhor	<i>Against the grain 1982</i>	Highland	closed distillery
Glen Moray	<i>10 years</i>	Speysaide	bottled by Berry bro's and Rudd Ltd

Distillery or bottler	Edition	Origin	Comments
Glen Scotia	<i>Ruby port finish 2008 vintage</i>	Campbeltown	
Imperial	<i>1979</i>	Speyside	closed distillery, bottled by Samaroli, best memory
Laphroaig	<i>25 years old</i>	Islay	
Ledaig	<i>aged 13 years</i>	Highland	Amontillado cask finish
Littlemill	<i>21 years old</i>	Lowland	closed distillery
Longrow	<i>red 11 years</i>	Campbeltown	cabernet franc matured
Post Askaig	<i>19 years</i>	Islay	
Port Ellen	<i>27 years old 1982</i>	Islay	closed distillery, connoisseurs choice
Octomore	<i>08.4 07.3</i>	Islay	
Old Bothwell	<i>single cask malt whisky</i>		
Old Pulteney	<i>18 years old</i>	Highland	
Orkney islands	<i>18 years old</i>	Highland	bottled by Berry Bro's and Rudd Ltd
Rosebank	<i>12 years old rare old - 1990</i>	Lowland	closed distillery bottled by Godron and Macphail
Springbank	<i>The tasting room</i>		
Talisker	<i>30 year old</i>	Isle of Skye	