



HAL
open science

Approche multi-critère pour la caractérisation des adventices

Jehan-Antoine Vayssade

► **To cite this version:**

Jehan-Antoine Vayssade. Approche multi-critère pour la caractérisation des adventices. Intelligence artificielle [cs.AI]. Université Bourgogne Franche-Comté, 2022. Français. NNT : . tel-03688127

HAL Id: tel-03688127

<https://theses.hal.science/tel-03688127v1>

Submitted on 3 Jun 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.

Public Domain

**THÈSE DE DOCTORAT DE L'UNIVERSITÉ
BOURGOGNE FRANCHE-COMTÉ
PRÉPARÉE À L'INRAE – UMR AGROÉCOLOGIE**

ÉCOLE DOCTORALE N°554, ENVIRONNEMENT-SANTÉ
DOCTORAT EN INFORMATIQUE

THÈSE PRÉSENTÉE ET SOUTENUE PAR
JEHAN-ANTOINE VAYSSADE
À DIJON LE 08 MARS 2022

**Approche multi-critère
pour la caractérisation des adventices**

Composition du Jury :

Pr. Christian Germain	Bordeaux Sciences Agro	Examineur
Pr. Bruno Tisseyre	Institut Agro Montpellier	Rapporteur
Pr. Laure Tougne-Rodet	Université Lumière - Lyon 2	Rapporteuse
Pr. Christelle Gée	Institut Agro Dijon	Directrice de thèse
Mc. Jean-Noël Paoli	Institut Agro Dijon	Co-encadrant
Mc. Gawain Jones	Institut Agro Dijon	Co-encadrant

Titre : Approche multi-critère pour la caractérisation des adventices

Mots clés : Vision par ordinateur ; Analyse d'image ; Intelligence artificielle ; Statistiques ; Prédiction ; Agriculture de précision

Résumé : L'objectif de cette thèse est de mettre au point un moyen de détecter les adventices dans un champ à l'aide d'images multispectrales, afin de pouvoir déterminer quelles sont les adventices à éliminer pendant le cycle de culture en cours et plus particulièrement aux stades précoces. L'approche multi-critère s'intéresse à la disposition spatiale, à la signature spectrale, à la morphologie et à la texture des plantes présentes dans les parcelles. Ce manuscrit propose une méthode permettant de sélectionner les meilleurs critères pour une discrimination optimale dans un contexte donné. Préalablement à l'extraction de ces critères, un ensemble de méthodes ont été développées afin de corriger les erreurs du dispositif d'acquisition, de détecter précisément la végétation, puis d'identifier au sein de la végétation les individus sur lesquels les différents critères peuvent être extraits. Pour l'étape de détection des individus, il s'est révélé que l'échelle de la feuille était plus adaptée que celle de la plante. La détection de la végétation et l'identification des feuilles s'appuient

sur des méthodes d'apprentissage profond, capables de traiter des feuillages denses. L'introduction de ces méthodes dans une chaîne de traitement usuelle constitue l'originalité de ce manuscrit où chaque partie a fait l'objet d'un article. Concernant le dispositif d'acquisition, une méthode de recalage des bandes spectrales a été développée, et les résultats montrent une précision de l'ordre du pixel. Ensuite, de nouveaux indices de végétation reposant sur de l'intelligence artificielle constituent l'une des avancées scientifiques de cette thèse. A titre indicatif, ces indices permettent d'atteindre 82.19% de mIoU contre 63.93%-73.71% pour des indices standards et fonctionnent en environnement non-contrôlé. Par extension, une méthode de détection des feuilles a été définie. Elle repose sur la détection de leurs contours, et semble avantageuse sur nos données multispectrales. Finalement, les meilleurs couples de propriétés ont été définis pour la discrimination culture/adventices à l'échelle de la feuille, dont les performances atteignent 91% de classification.

Title : A multi-criteria approach to weed characterization

Keywords : Computer vision ; Image analysis ; Artificial intelligence ; Statistics ; Prediction ; Precision agriculture

Summary : The objective of this thesis is to develop a way to detect weeds in a field using multispectral images, in order to determine which weeds should be eliminated during the current crop cycle and more particularly at the early stages. The multi-criteria approach focuses on the spatial arrangement, the spectral signature, the morphology and the texture of the plants located in the plots. This work proposes a method for selecting the best criteria for optimal discrimination for a given setup. Prior to the extraction of these criteria, a set of methods was developed in order to correct the errors of the acquisition device, to precisely detect the vegetation and then to identify within the vegetation the individuals on which the different criteria can be computed. For the individual detection step, it appears that leaf scale is more suitable than plant scale. Vegetation detection

and leaf identification are based on deep learning methods capable of processing dense foliage. The introduction of these methods in a usual processing chain constitutes the originality of this manuscript where each part was the subject of an article. Concerning the acquisition device, a method of spectral band registration was developed. Then, new vegetation indices based on artificial intelligence constitute one of the scientific advances of this thesis. As an indication, these indices offer a mIoU of 82.19% when standard indices ceil at 63.93%-73.71%. By extension, a leaf detection method was defined and is based on the detection of their contours, this method seems advantageous on our multispectral data. Finally, the best property pairs were defined for crop/weed discrimination at leaf level, which classification performances up to 91%.

Contributions de l'auteur

Conférences

- **Two-step multi-spectral registration via key-point detector and gradient similarity** February 2020 - Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications : Valetta (Malta) - Jehan-Antoine Vayssade; Gawain Jones; Jean Noel Paoli; Christelle Gée
- **DeepIndices : Une nouvelle approche des indices de télédétection basée sur l'optimisation et l'approximation de fonctions par DeepLearning. Application aux indices de végétation sur des données non calibrées** July 2021 - RJCIA : Rencontres des Jeunes Chercheurs en Intelligence Artificielle (Bordeaux) - Jehan-Antoine Vayssade; Jean Noel Paoli; Christelle Gée; Gawain Jones

Articles de journaux internationaux

- **DeepIndices : Remote Sensing Indices Based on Approximation of Functions through Deep-Learning, Application to Uncalibrated Vegetation Images** June 2021 - Remote Sensing 13(2261) - Jehan-Antoine Vayssade; Jean Noel Paoli; Christelle Gée; Gawain Jones
- **Dense Leaves Pixelwise Instance Segmentation** February 2022 - Computer and Electronics in Agriculture - Jehan-Antoine Vayssade; Gawain Jones; Christelle Gée; Jean Noel Paoli

Article en cours de publication

- **Weed discrimination at leaf scale : towards the characterization of the vegetation cover** January 2022 - Computer and Electronics in Agriculture - Jehan-Antoine Vayssade, Gawain Jones, Christelle Gée and Jean-Noël Paoli

Collaborations extérieures

- **Automatic activity tracking of goats using drone camera** July 2019 - Computers and Electronics in Agriculture 162 :767-772 - Jehan-Antoine Vayssade; Rémy Arquet; Mathieu Bonneau
- **Goats monitoring at the pasture scale combining neural network and time-lapse cameras** August 2019 - European Conference on Precision Livestock Farming (ECPLF), - Mathieu Bonneau; Jehan-Antoine Vayssade; Willy Troupe; Rémy Arquet
- **Outdoor animal tracking combining neural network and time-lapse cameras** December 2019 - Computers and Electronics in Agriculture 168 - Mathieu Bonneau; Jehan-Antoine Vayssade; Willy Troupe; Rémy Arquet

Préambule

Cadre général : En cultures maraîchères et grandes cultures, la gestion des adventices est une question critique. La limitation de leur prolifération à tous les stades de la culture est cruciale. Les adventices ont des conséquences importantes du point de vue qualitatif et financier mais également sur le plan agroécologique. Les solutions de gestion des cultures peuvent être préventives en contribuant de manière directe ou indirecte à diminuer la nuisibilité des adventices [Caussanel, 1989] dans le temps, leur gestion aux stades de développement les plus précoces est donc essentielle. Ces solutions reposent essentiellement sur des leviers agronomiques (allongement des rotations, réduction du travail du sol : faux-semis, cultures associées, ...). D'autres solutions, dites curatives [cordeau et al., 2016], proposent des solutions techniques à ces questions. Le plus souvent chimiques, on cherche de plus en plus à les réduire en s'orientant vers des solutions alternatives en ayant recours au désherbage mécanique, voire manuel, selon le type et le stade de développement des cultures.

En agriculture conventionnelle, pour des raisons budgétaires et de rentabilité, l'utilisation de produits phytosanitaires est encore privilégiée, au détriment d'autres méthodes jugées plus coûteuses ou moins performantes, même fréquentes et bien raisonnées [Colbach et al., 2010]. L'utilisation massive de ces produits chimiques de synthèse a conduit, entre autre, au cours des dernières décennies à une pollution des eaux ¹ [Reulier, 2015], à un appauvrissement de la biodiversité [Jonathan Storkey, 2007] et à l'apparition de phénomènes de résistance [Délye, 2017]. La réduction des produits phytosanitaires représente donc un des enjeux majeurs du secteur agricole. En effet, dans un objectif "zéro-phyto", il s'agit de réduire l'impact environnemental de cette pratique de désherbage tout en accompagnant l'agriculteur vers de nouvelles stratégies fiables et robustes qui allient agroécologie et nouvelles technologies. Cela repose principalement sur la diversification des méthodes de lutte [Wezel et al., 2014] au-delà des herbicides (choix de variétés, rotation des cultures, prophylaxie, biocontrôle, ...). Cette démarche s'inscrit dans les plans gouvernementaux français tels que Ecophyto, Ecophyto 2 et Ecophyto II+.

Objectif : L'objectif est donc de proposer des outils numériques pour une meilleure gestion des adventices dans le cadre des cultures dites conventionnelles. La démarche envisagée s'inscrit dans le domaine de l'agriculture numérique [Bellon Maurel and HUYGHE, 2019], qui englobe l'agriculture de précision, et consiste en une évaluation multi-critère des adventices détectées par imagerie pour une gestion localisée.

Actuellement les études sur la détection des adventices par imagerie sont difficilement comparables car fortement liées aux systèmes d'imagerie, aux résolutions spatiales et au stade de développement du cycle cultural en cours. En effet, la comparaison des performances des méthodes actuelles de traitement d'images est difficile, puisque leurs études sont basées sur des critères d'appréciation disparates ; c'est-à-dire que les données, les analyses et les métriques ne concordent pas entre les études. Dans la majorité des cas, elles ne permettent donc pas leurs transpositions à une nouvelle situation.

1. Projet PULSE : Paysages, ParticULEs, peSticidEs (2020-2022)

Contexte du projet : L'équipe de recherche "agriculture de précision" (pôle GestAd) de l'UMR Agroécologie de Dijon a pour objectif de produire des connaissances mobilisables pour une gestion durable de la flore adventice, c'est-à-dire une gestion minimisant les intrants, tout en maintenant la production agricole. Dans cette optique, notre projet portera sur l'étude approfondie des différentes méthodes et métriques usuelles afin de définir clairement les meilleures d'entre elles. Cette démarche permettra également de préconiser les conditions d'acquisition optimales, afin de détecter et traiter les adventices directement et localement.



Partenaires : Ces travaux sont menés dans le cadre de deux projets de recherche :

- **H2020 IWMPRAISE (2017-2022) :** La gestion intégrée des mauvaises herbes est la voie à suivre pour une agriculture durable et résiliente. IWMPRAISE (Integrated Weed Management : PRActical Implementation and Solutions for Europe) est un projet européen de type "Horizon 2020" qui soutient et favorise la mise en œuvre de la GIRE en Europe. La gestion des mauvaises herbes en Europe deviendra plus respectueuse de l'environnement, si le concept de gestion intégrée des mauvaises herbes s'impose dans les fermes [Kudsk et al., 2020]. Dans le cadre de ce travail, l'équipe envisagera des solutions couvrant majoritairement les problèmes de « détection - interprétation ».
- **ANR challenge ROSE consortium ROSEAU (2018-2021) :** Le dispositif de financement Challenge ROSE (RObotique aux Service des Ecophyto) vise à susciter des dynamiques de recherche en mobilisant les acteurs scientifiques et industriels autour d'une même problématique. Quatre équipes de recherche se focalisent sur le désherbage de l'intra-rang en cultures maraîchères de plein champ et en grandes cultures à fort écartement. Les solutions envisagées par le consortium ROSEAU (RObotics SEnsorimotor loops to weed AUtonomously) couvrent la totalité de la chaîne « détection - interprétation - décision - action ». Chaque équipe proposera des éléments de recherche dans une ou plusieurs parties de la chaîne, et ces travaux seront mis en commun pour la création d'une machine bloc-outils permettant la gestion des adventices de façon autonome. L'équipe de recherche "Agriculture de précision" envisage de couvrir les problèmes de « détection » au sein du consortium ROSEAU porté par la société SITIA, possédant un robot autonome dénommé PUMA.

Conclusion : Notre cadre général est donc soutenu par *IWMPRAISE H2020*, qui consiste en l'évaluation d'une méthode de détection des adventices. Nous appliquerons nos algorithmes pour un contexte particulier dans le cadre de l'*ANR Challenge Rose* sur des cultures sarclées.



Table des matières

1	Introduction	1
1.1	Agriculture : un avènement	1
1.1.1	Dynamique des systèmes	2
1.1.2	Limites de l'agriculture intensive	3
1.1.3	Modes de production	6
1.1.4	Transition agroécologique	7
1.2	Agriculture de précision	8
1.2.1	La télédétection au service de l'agriculture de précision	9
1.2.2	Vers une agriculture numérique	11
1.3	Le désherbage	12
1.3.1	Place des adventices dans les systèmes de culture	12
1.3.2	La biologie des adventices	14
1.3.3	Fenêtres de traitement	16
1.3.4	Les technologies de désherbage actuelles	17
1.3.5	Les technologies de désherbage en développement	19
1.4	Conclusion de chapitre	21
2	Vision par ordinateur	23
2.1	Systèmes d'imagerie	24
2.1.1	Percevoir la couleur	26
2.1.2	Des capteurs multispectraux	27
2.2	Prétraitement	28
2.2.1	Correction géométrique	28
2.2.2	Rehaussement d'image	29
2.2.3	Indices de télédétection	31
2.3	Segmentation	34
2.3.1	Segmentation sémantique	34
2.3.2	Segmentation en région	36
2.3.3	Segmentation des instances	39
2.4	Extraction de propriétés	41
2.4.1	Propriétés géométriques	41
2.4.2	Propriétés spatiales	42
2.4.3	Propriétés de texture	42
2.4.4	Robustesse des propriétés	43
2.4.5	Apprentissage profond	43
2.5	Classification	44
2.5.1	Classification supervisée	44
2.5.2	Classification non-supervisée	47
2.5.3	Classification semi-supervisée	48
2.6	Métriques d'évaluation	49
2.6.1	Classification	49
2.6.2	Régression et calibration	50

2.7	Applications en agriculture	51
2.7.1	Échelle du pixel et voisinage proche	51
2.7.2	Échelle de la fenêtre	52
2.7.3	Composantes connexes et super-pixels	53
2.7.4	Échelle de la plante	55
2.8	Conclusion de chapitre	56
3	Matériel et données	57
3.1	Les sites expérimentaux	57
3.2	Capteur optique	58
3.3	Jeux de données	59
3.3.1	Campagnes d’acquisition	59
3.3.2	Annotations	60
3.4	Calibration	62
3.4.1	Résolution spatiale	62
3.4.2	Correction géométrique	63
3.5	Recalage des bandes spectrales	65
3.5.1	Introduction	67
3.5.2	Material and Method	68
3.5.3	Results and discussion	72
3.5.4	Conclusion	76
3.6	Conclusion de chapitre	77
4	Indices de végétation	79
4.1	Introduction	81
4.2	Material and Data	83
4.2.1	Instrument Details	83
4.2.2	Image Dataset	83
4.2.3	Data Pre-Processing	84
4.2.4	Training and Validation Datasets	85
4.3	Methodology	85
4.3.1	Existing Spectral Indices	85
4.3.2	Deepindices : Baseline Models	86
4.3.3	Enhancing Baseline Models	88
4.3.4	Last Activation Function	90
4.3.5	Loss Function	90
4.3.6	Performance Evaluation	90
4.3.7	Comparison with Standard Indices	91
4.3.8	Training Setup	91
4.4	Results and Discussion	92
4.4.1	Fixed Models	92
4.4.2	Deepindices	93
4.4.3	Initial Image Processing	95
4.4.4	Discussion	95
4.5	Conclusions	99
4.6	Conclusion de chapitre	102

5	Détection des individus	103
5.1	Introduction	106
5.1.1	Related Works	106
5.1.2	Objectives	107
5.2	Material and data	108
5.2.1	Specific multispectral dataset	108
5.2.2	Online image databases	109
5.3	Methodology	110
5.3.1	Proposed CNN architecture	110
5.3.2	Loss function	114
5.3.3	Refinement with vegetation mask and watershed	116
5.3.4	Training setup	117
5.3.5	Evaluation metrics	117
5.4	Results	117
5.4.1	Komatsuna dataset	117
5.4.2	Leaf Segmentation Challenge dataset	119
5.4.3	Airphen dataset	123
5.5	Discussion	125
5.6	Conclusion	126
5.7	Further research	127
5.8	Conclusion de chapitre	128
6	Extraction de propriétés	129
6.1	Introduction	131
6.2	Material and data	132
6.2.1	Experimental plot	132
6.2.2	Multispectral camera	133
6.2.3	Image acquisition and annotation	133
6.3	Methodology	134
6.3.1	Literature review	134
6.3.2	Optimization	139
6.3.3	Feature Selection	141
6.4	Results and Discussion	141
6.4.1	Feature optimisation	141
6.4.2	Feature selection by type	143
6.4.3	Best of all features	145
6.4.4	Visual results	146
6.5	Conclusion	148
6.6	Further research	149
6.7	Conclusion de chapitre	152
7	Conclusion Générale	153
7.1	Contributions	154
7.2	Perspectives	155

1 Introduction

Ce chapitre présente les principaux modèles de production agricole et leurs évolutions. Il aborde dans une première partie les intérêts et limites de l'agriculture conventionnelle. Dans une seconde partie, il s'intéresse à la transition agroécologique et à l'agriculture de précision. L'intérêt des systèmes d'imagerie et des méthodes de traitement des données numériques sera notamment souligné.

1.1 Agriculture : un avènement

Apparue au Néolithique, l'agriculture a constamment évolué tout au long de l'histoire de l'humanité. Les premiers grands changements remontent à la moitié du XXe siècle où l'agriculture s'est considérablement modernisée. En effet, l'agriculture traditionnelle évolue vers une agriculture mécanisée grâce à l'essor des sciences et techniques telles que la mécanique, l'hydraulique et l'électrotechnique. Cette période s'inscrit dans un contexte global d'accélération de la démographie et de la consommation des ressources naturelles, c'est l'Anthropocène [Federau, 2016]. Elle est caractérisée par l'ensemble des événements géologiques qui se sont produits depuis que les activités humaines ont une incidence globale significative sur l'écosystème terrestre.

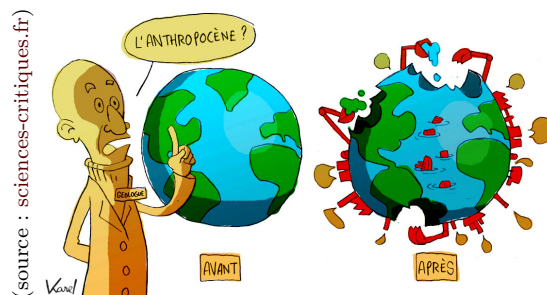


Figure 1.1 – L'Anthropocène

La production agricole a été fortement accrue, mais le modèle actuel (i.e l'agriculture conventionnelle), soulève de nombreuses questions relatives à sa viabilité économique (prix de vente des productions, coût des intrants), à son impact sur la santé des consommateurs (qualités nutritionnelles des produits, résidus de pesticides) et à son empreinte environnementale [Loiret, 2016] (changement climatique, baisse de la biodiversité). La baisse des prix a également engendré une chute des salaires engendrant des conséquences sociales importantes, avec une baisse significative du nombre d'exploitants à travers le monde (suicide, non-reprise, ...) [Chau-meil, 1939, Shiva and Jalees, 2006, Carles, 2007, Nagaraj, 2008, Bossard et al., 2016, Utyasheva et al., 2021]. Dans un souci d'augmenter la production, l'agriculture rurale évolue ainsi vers une production industrielle.

Les solutions étudiées visent à réduire les quantités globales d'intrants employées, et à en éliminer certains. La suite de ce chapitre s'intéresse aux scénarios d'évolution de notre écosystème et à la manière dont ils sont pris en compte par nos politiques publiques. Il détaille ensuite les modes de production usuels, puis s'intéresse aux limites de l'agriculture intensive et à la transition agroécologique.

1.1.1 Dynamique des systèmes

La dynamique des systèmes est une approche pour comprendre le comportement des systèmes complexes dans le temps [Keller, 2018]. Cette approche peut être utilisée par exemple pour modéliser les limites et vulnérabilités de notre système de production agricole. Pour cela, deux facteurs biophysiques sont utilisés : **1) l’empreinte écologique**, c’est l’impact que les activités humaines ont sur l’environnement. Cet impact se divise en trois catégories (i) les ressources prélevées (ii) les pollutions et déchets rejetés et (iii) les dégradations infligées, **2) la biocapacité**, c’est la capacité du système à (i) se régénérer lorsque l’on prélève une ressource (ii) absorber les déchets et pollutions (iii), à réparer les dégâts qui lui ont été infligés. La dynamique des systèmes formalise simplement que l’activité humaine « ne peut pas avoir une empreinte écologique qui dépasse durablement la biocapacité de la planète ». À partir de là, plusieurs scénarios peuvent être envisagés [Martino, 1973, Keller, 2018, Herrington, 2021] :

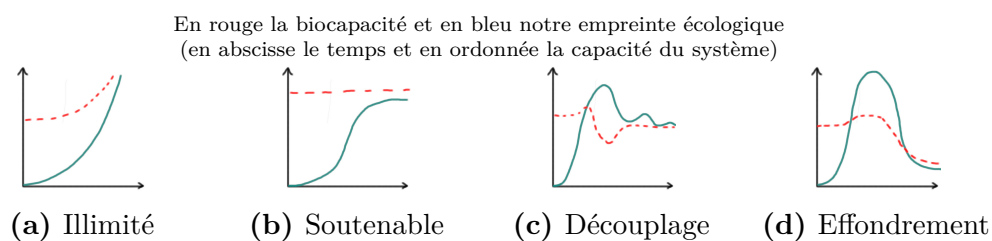


Figure 1.2 – Dynamique des systèmes : les différents imaginaires possibles

Illimité : Ce scénario est antérieur aux prises de conscience environnementale. La PAC de 1962, orientée vers l’augmentation, d’année en année, des productions agricoles en Europe, s’inscrit dans cette vision.

Soutenable : Postérieur aux prises de conscience environnementales, ce scénario considère une biocapacité limitée, mais suffisante pour absorber l’impact de l’activité humaine. Il en a résulté une modification de la PAC en 1999 avec l’apparition d’exigences environnementales devant permettre de rester en dessous de la limite de la biocapacité.

Découplage : Ce scénario, introduit une biocapacité fluctuante, et imagine des dépassements temporaires, avant une convergence progressive entre biocapacité et empreinte. Il s’accompagne d’un découplage de notre croissance économique et de l’empreinte écologique de nos modes de production. Il s’agit d’un modèle de développement durable envisagé par exemple lors des COP (Conference of the Parties), ou de la rédaction du protocole de Kyoto.

Effondrement : Il s’agit de la perte de la capacité du système à se régénérer. S’ensuit un effondrement de la biodiversité, puis de la production sur une très longue période entraînant une chute démographique brute. Les politiques évoquées précédemment visent à éviter ce scénario.

Aujourd’hui, nous savons que les deux premiers univers ne sont pas possibles puisque l’on a déjà dépassé la biocapacité de notre planète [Meadows et al., 1972, Day, 2017]. Et nous observons une baisse significative, aux niveaux mondiaux, de la capacité de production des sols, des ressources halieutiques, du nombre de pollinisateurs et la fonte des glaciers d’altitudes [Perrier-Bruslé et al., 2013, Martin et al., 2017]. Il est donc urgent aujourd’hui de reconsidérer nos modes de production. Et plus largement les bases de notre économie qui représente un défi scientifique et technologique majeur [Carles, 2007, Aberkane, 2015, Jancovici, 2017].

1.1.2 Limites de l'agriculture intensive

1.1.2.1 Impact de la mécanisation

L'intensification de l'agriculture française se développe à partir de 1950 en France à travers le plan Marshall. La mécanisation permet la réduction de la pénibilité du travail physique, l'automatisation de certaines tâches quotidiennes et des débits de chantier élevés, permettant un gain de productivité notable. Le gouvernement français, soutenu par la PAC, lance alors le remembrement des terres agricoles peu avant 1970, pour faciliter l'usage des matériels agricoles et la simplification des interventions culturales. On estime ainsi que 40 % à 80 % des bocages ont disparu (figure 1.3), et le remembrement des terres agricoles continue au fil des générations.

(source : [Boissinot et al., 2014])

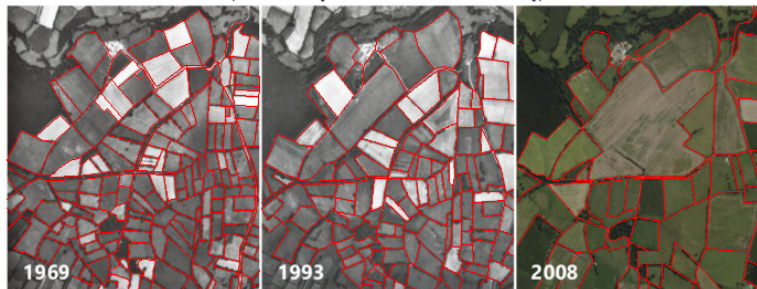


Figure 1.3 – Remembrement des terres agricoles

Avec l'apparition de grandes terres agricoles, les besoins en équipements agricoles évoluent et des progrès techniques considérables y font leurs apparitions afin d'être de plus en plus compétitifs sur les marchés. L'intensification de nouvelles pratiques (labour profond, engrais chimiques) entraînent lentement une disparition de la matière organique des sols [Vitinnov et al., 2017] qui sert d'alimentation à la faune et flore des sols [Pujol, 2014] tel que les lombrics [Pelosi, 2008], les carabes [Carbonne, 2020] et les mycorhizes [Zerbib, 2018].

Ces pratiques intensives, mettent également les sols à nu, impliquant un phénomène de tassement et compactage des sols. Ce dernier freine le développement des racines et la pénétration de l'eau dans les couches inférieures, utiles aux cultures, par exemple comme réserve d'eau en cas de sécheresse ou le stockage de certains éléments comme l'azote et le carbone [dos Reis Martins et al., 2009, Némery et al., 2016, Kan et al., 2020]. Puisque l'eau ne pénètre plus, elle ruisselle en surface et « entraîne une acidification des sols par perte du calcium, lequel sert de pont d'attache entre les argiles et les humus. Sans calcium, le complexe argilo-humique se détruit et les argiles partent en suspension dans l'eau. Enfin, il y a l'érosion » [Anthony and Gratiot, 2012, Yousefi et al., 2016, Cossart and Fressard, 2019, Hinimbio, 2019, Kolli et al., 2021].

Ces deux facteurs menacent ainsi la durabilité de la production par une pollution environnementale (eau, sol, air), un appauvrissement des sols et une réduction de la biodiversité. Aujourd'hui un regain d'intérêt pour les bocages se fait ressentir, ainsi de nombreuses études ont montré qu'à l'échelle du paysage, les habitats peu perturbés comme les haies ou les prairies jouent un rôle important pour le maintien et le déplacement de certains auxiliaires, mais également dans la gestion et le ralentissement du lessivage des sols. À contrario, les petites parcelles qui n'ont pu être remembrées sont abandonnées dû à plusieurs facteurs, comme la pénibilité du travail. Les parcelles abandonnées s'enfrichent [Pic et al., 2020] faisant temporairement office de réserve à la bio-diversité. Après un certain temps, les friches tendent soit vers la forêt, soit à la reprise, soit vers l'urbanisation en fonction des acteurs économiques à proximité.

1.1.2.2 Impact des intrants

Parallèlement au développement de la mécanisation, l'usage de la fertilisation chimique et des traitements phytopharmaceutiques se développe massivement, permettant d'augmenter des rendements et d'alimenter le marché mondial. Leurs utilisations deviennent nécessaires puisque les sols s'appauvrissent avec l'intensification. Lors de fortes pluies et du fait de la suppression des bocages, les polluants sont lessivés puis ruissellent vers les rivières [Tournebize, 2020, Pic et al., 2022]. C'est le cas des fertilisants, visible sur la figure 1.4 pour les excédents d'azote.

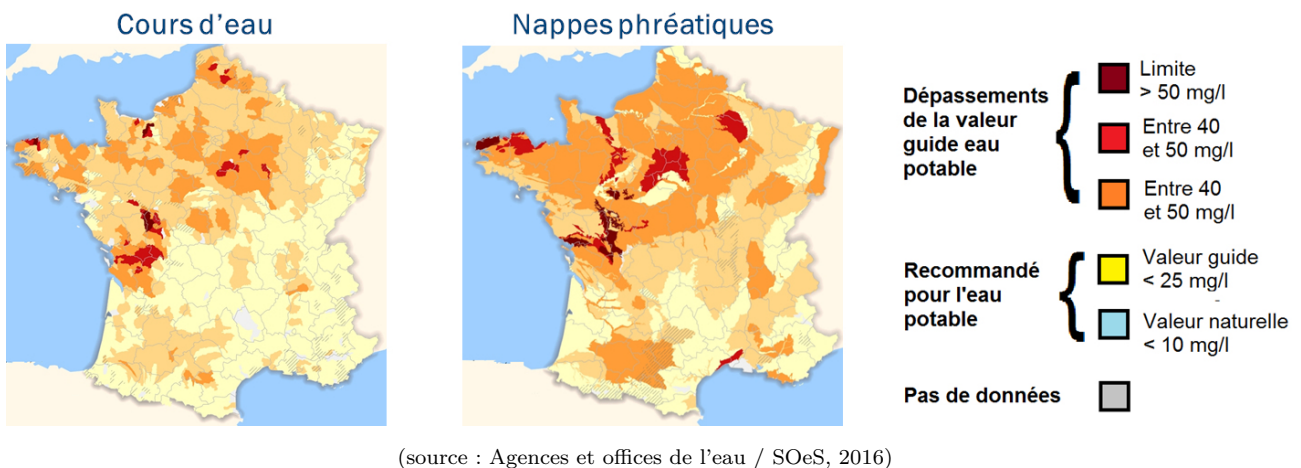


Figure 1.4 – Excédent d'azote dans la majorité des cours d'eau en 2013

L'excès de nutriments dans les rivières induit une prolifération d'algues et de cyanobactéries qui se développent en masse et entraînent à terme la limitation de la pénétration de la lumière dans la colonne d'eau (en particulier dans les barrages). Ensuite, les algues légèrement sous la surface meurent et tombent au fond entraînant une dégradation de matière organique consommant l'oxygène du milieu (désoxygénation), puis à terme l'asphyxie du milieu aquatique (mort des poissons). C'est l'eutrophisation des milieux aquatiques !

L'eutrophisation d'origine anthropique est donc due à l'intensification des modes de productions et l'usage de fortes doses. Le phénomène des algues vertes [Léraud and Van Hove, 2019], comme les Sargasses dans les Caraïbes [Louime et al., 2017] sont alimentées par 95 % des nutriments azotés provenant de la lixiviation des terres agricoles et donc des fertilisants agricoles. De plus, dans le golfe du Mexique une immense « zone morte » grandit le long de la Louisiane et du Texas (estimée à 22 000 kilomètres carrés) provoquée par l'asphyxie, ici du Mississippi. Ces trois exemples importants sont visibles en figure 1.5.



Figure 1.5 – Prolifération d'algues et asphyxie des milieux aquatiques

Ces algues ont également des conséquences importantes sur la santé humaine. Les gaz émis par leur décomposition sont hautement toxiques (hydrogène sulfuré et ammoniac) dont leur inhalation peut entraîner la mort, plusieurs cas ont été reportés depuis 1984 en Bretagne. La loi sur l'eau et les milieux aquatiques du 30 décembre 2006 (LEMA) tente de limiter ces effets.

Comme nous l'avons vu pour les fertilisants, les pesticides sont également emportés par le lessivage des terres agricoles [Tournebize, 2020]. Un rapport du ministère de la Transition écologique et solidaire de juin 2017 fait état des pesticides présents sur presque tout le territoire français, mais à des concentrations variables (figure 1.6).

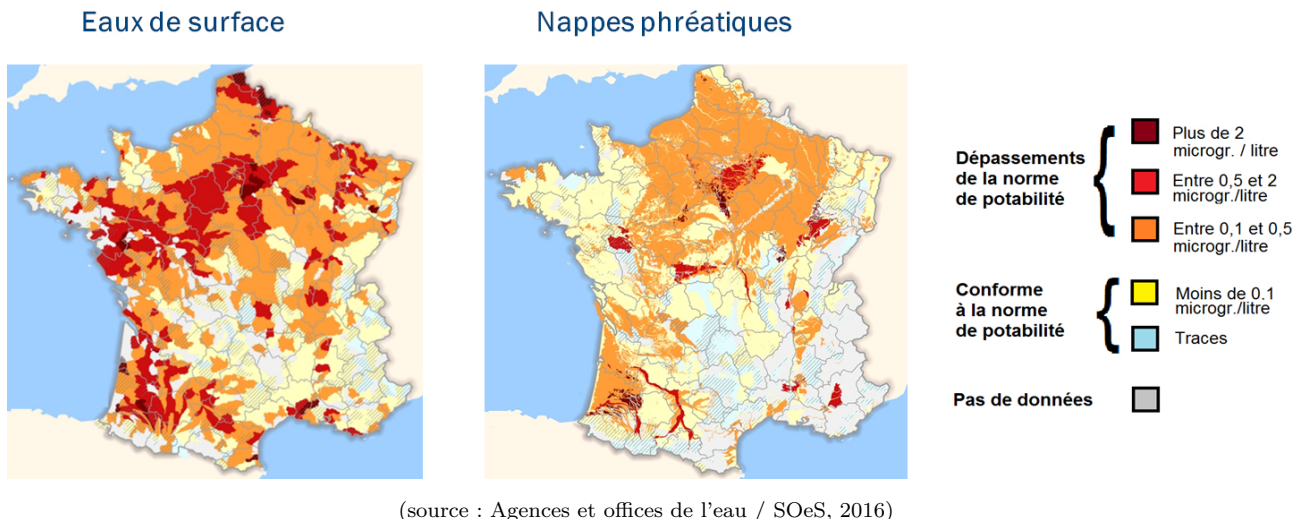


Figure 1.6 – Presence de pesticide dans la majorité des cours d'eau en 2013

Plusieurs publications ont mis en évidence les effets néfastes des pesticides sur l'environnement [McLaughlin and Mineau, 1995, Van Der Werf, 1997, Pesce et al., 2021] impliquant une dégradation des écosystèmes [Aubertot et al., 2005, Bianchi et al., 2006]. D'après [Geiger et al., 2010], sur les 13 composantes de l'intensification agricole qui ont été mesurées, l'utilisation d'insecticides et de fongicides a eu des effets négatifs constants sur la biodiversité. Il est important de noter que les pesticides sont généralement appliqués par pulvérisation, impliquant une forte volatilisation du produit, qui peut s'échelonner de quelques jours à plusieurs semaines. Engendrant un transfert de masse dans l'atmosphère [Bedos et al., 2002] et touchant donc des populations non-ciblées. Ceci implique des effets collatéraux, importants, sur la diversité des espèces de plantes sauvages, de carabes et d'oiseaux [Geiger et al., 2010], et donc sur le potentiel de lutte biologique contre les ravageurs. Les populations d'invertébrés sont également touchées [Beketov et al., 2013, Grozinger and Flenniken, 2019, Wood et al., 2020].

Concernant les espèces humaines et animales, l'exemple de l'utilisation du Chlordécone, pesticide utilisé dans les Antilles Françaises, pour traiter les plants de bananes, a montré un impact négatif [Dubuisson et al., 2007, Maudouit and Rochoy, 2019]. Les conséquences possibles à long terme continuent d'être étudiées, car le Chlordécone contamine encore les sols des Caraïbes. Les effets d'une exposition aiguë sont bien décrits, en revanche les effets à des doses plus diffuses se précisent, même s'ils restent complexes à identifier.

Face à ces éléments (engrais, pesticides, pratiques intensives, ...), « il est clair que nous devons trouver des solutions techniques afin de maximiser l'intégrité de l'écosystème et le bien-être de l'homme sans affecter notre avenir » [Dufour and Piégay, 2009].

1.1.3 Modes de production

Actuellement, l'agriculture dite "conventionnelle" constitue le mode de production de référence des sociétés industrielles. Elle a recours à divers intrants chimiques de synthèse [Kisinyo et al., 2011] pour maintenir l'état sanitaire de la culture face aux différents stress biotiques [Pandey et al., 2017] et abiotiques [Cramer et al., 2011] qu'elle peut subir. Face aux préoccupations environnementales, mais également pour répondre aux attentes des consommateurs (en termes de risque sanitaire [Nougadère et al., 2015] et de qualité nutritive des aliments), différentes évolutions ont été envisagées pour l'agriculture conventionnelle, et des modes de production alternatifs ont été définis, dont les principaux sont :

Agriculture durable : Dérivée de l'agriculture conventionnelle, c'est une agriculture extensive qui s'inscrit dans les perspectives ouvertes par le développement durable défini en 1992 à Rio¹. Ce type d'agriculture vise à « répondre aux besoins du présent sans compromettre la capacité des générations futures à répondre à leurs propres besoins ». Il s'agit donc de moduler nos productions pour concilier les dimensions écologiques, économiques et sociales. Ce concept n'est pas appuyé par une législation précise.

Agriculture biologique : Il s'agit du mode de production alternatif le plus connu. Son objectif est d'intégrer un ensemble de pratiques agricoles plus respectueuses des équilibres écologiques, du bien-être des animaux et de l'autonomie des agriculteurs. Cette agriculture a pour particularité d'exclure l'usage des produits chimiques de synthèse [Wintermantel et al., 2020], des OGM [Bagnoud and Recordon, 2020] et limite l'emploi d'intrants [Etter et al., 2011]. Les produits issus de ce mode de production font l'objet d'une certification (variable en fonction de l'origine du produit).

Permaculture : La permaculture est un ensemble de principes de conception centrés sur une pensée systémique globale au sein d'une petite surface. Ce mode de production se rapproche de l'agriculture biologique, mais conçoit son système autour de l'agriculteur, de la famille et de l'habitat. L'exploitation est alors divisée en plusieurs zones fournissant des services différents (jardin annuel, jardin permanent, verger, fourrages, forêt). La permaculture prend également écho dans le domaine des semences libres et paysannes [Shiva, 2005, Mésini, 2006, Zerbib, 2018, El Amine et al., 2018]. En Autriche, le Krameterhof [Holzer, 2011], la ferme du Bec Hellouin en France [Mésini, 2006] et Fukuoka au Japon [Fukuoka, 1983] sont les exemples les plus connus.

1. https://fr.wikipedia.org/wiki/Agenda_21

1.1.4 Transition agroécologique

Les modes de production alternatifs précédemment énoncés ont pour but de répondre à ces problématiques, tout en satisfaisant les besoins alimentaires des populations. Néanmoins, à l'exception de la permaculture, ces alternatives commencent à avoir recours aux fertilisants pour augmenter les rendements et ainsi répondre à la demande des consommateurs. Certains équilibres peuvent s'avérer difficiles à trouver, entre production et écologie, notamment pour ce qui concerne la gestion des adventices [Bajwa et al., 2019]. Un recours au labour retarde la levée des adventices, mais s'avère préjudiciable à la vie biologique des sols [Lemtiri et al., 2016]. Les approches dites de conservation (abandon du travail du sol profond, couverture permanente du sol, allongement des rotation) reposent sur l'usage d'herbicides de synthèse, qui ont un impact négatif sur les milieux naturels et la santé des consommateurs (comme vu dans la partie 1.1.2) mais permettent de diminuer le compactage des sols. Pour définir une agriculture plus respectueuse de l'environnement, d'autres approches existent comme l'agroécologie et l'agriculture de précision.

L'agroécologie vise à développer une agriculture durable, en considérant les parcelles agricoles comme des agroécosystèmes. Elle s'appuie fortement sur l'utilisation de services écosystémiques [Altieri, 1999, Lamarque et al., 2014, Perrette, 2018, Lefloch, 2018, Altieri, 2018, Haan et al., 2021]. Elle étudie pour cela les interactions au sein de ces écosystèmes (plantes-plantes et plantes-microorganismes) à différents niveaux d'intégration (de la molécule à la communauté) et d'échelles spatio-temporelles (microcosme, parcelle, paysage, cycle de culture, rotation, ...).

L'Agriculture de précision se focalise sur l'utilisation de nouvelles technologies et outils numériques, qui permettent de recueillir des informations intra-parcellaires géo-référencées. La parcelle n'est plus considérée comme une entité homogène, mais présente une variabilité importante qui peut être mesurée et cartographiée. Grâce à un suivi régulier, l'agriculteur a alors les moyens de piloter ses interventions localement, et de réduire ses doses d'intrants chimiques (engrais, produits phytosanitaires) grâce à des machines dotées d'automatismes. L'agriculture de précision peut être une approche pour le développement de systèmes de culture à bas intrants et l'application de nouveaux produits de traitement (tels que les produits de bio-contrôle). À ce titre, elle constitue un levier pour la transition agroécologique.

Plusieurs approches co-existent donc, entre culture en équilibre avec son environnement, et approche technologique visant à maîtriser les volumes d'intrants. Cette thèse s'intéresse à l'agriculture de précision telle qu'étudiée par l'UMR 1347 Agroécologie, promue par les plans Ecophyto et repose sur une gestion localisée des adventices, pour une transition agroécologique.

1.2 Agriculture de précision

Le concept d'agriculture de précision est apparu dans les années 80 aux États-Unis, avec la naissance des systèmes de localisation par satellite (GPS/GNSS), avec comme première application la modulation de la fertilisation en fonction des besoins des cultures [Robert, 2002]. [McBratney et al., 2005, Taylor and Whelan, 2005] définissent le concept d'agriculture de précision comme la mise en place d'une stratégie de gestion de l'ensemble de l'exploitation s'appuyant sur les technologies de l'information dans le but d'améliorer la production agricole tout en limitant l'impact environnemental. Pour ce faire, l'agriculture de précision se base sur l'étude de la variabilité intra-parcellaire, que ce soit au niveau spatial ou temporel [Pierce and Nowak, 1999]. En 2019, la communauté scientifique internationale en agriculture de précision (ISPA) s'est de nouveau concertée pour proposer une nouvelle définition du terme agriculture de précision :

« Precision Agriculture is a management strategy that gathers, processes and analyzes temporal, spatial and individual data and combines it with other information to support management decisions according to estimated variability for improved resource use efficiency, productivity, quality, profitability and sustainability of agricultural production »

Les technologies de l'information occupent une place fondamentale dans l'agriculture de précision [Lherbier, 2005]. La gestion localisée des parcelles peut être représentée par un cycle composé de 4 étapes (figure 1.7) : (i) l'observation des parcelles (collecte de données), (ii) la caractérisation des parcelles (traitement des données), (iii) la préconisation (prise de décision) et (iv) l'action (application au champ). Le géo-référencement (association de coordonnées géographiques aux objets étudiés) relie ces étapes entre elles.

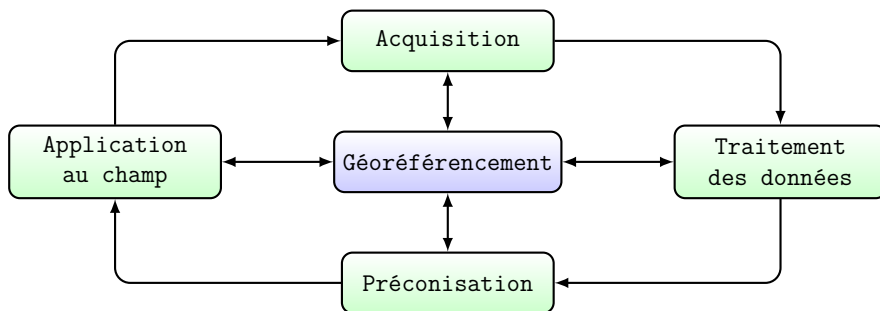


Figure 1.7 – Cycle de l'agriculture de précision composé de 4 étapes

Ainsi, l'agriculture de précision, par le biais des nouvelles technologies [Martos et al., 2021], propose par exemple une gestion intra-parcellaire de la fertilisation des cultures ou encore une gestion spécifique (SSWM Site-Specific Weed Management [López-Granados, 2011]) de la flore adventice au plus près grâce à des systèmes de très haute précision. Les données de la parcelle sont acquises régulièrement permettant également d'utiliser des modèles agronomiques afin de proposer des prédictions pour anticiper la gestion d'une parcelle (prévision de la croissance des adventices [McAllister et al., 2021]). De plus le géo-référencement de l'information et l'utilisation de GPS haute précision, permet un guidage automatique pour moduler les doses en fonction de la variabilité spatiale [Perez-Ruiz et al., 2021].

1.2.1 La télédétection au service de l'agriculture de précision

La photogrammétrie aérienne fait son apparition en 1919. Des applications agricoles sont rapidement envisagées, notamment en sylviculture. En 1929, au Canada, cette technologie permet notamment, de comptabiliser le nombre d'arbres plus rapidement qu'avec un comptage au sol et plus précisément qu'avec une estimation basée sur la surface d'exploitation. Par la suite, le développement de la stéréo-photogrammétrie permet de reconstruire de larges zones d'observations, des courbes de niveau des terrains, et cela, avec une précision métrique. Les possibilités se sont enrichies au fil des années avec le développement des appareils photo numériques et de l'imagerie satellitaire à travers la télédétection :

- En 1962 l'US Air Force développe et lance KH-4. Deux caméras sont embarquées qui permettent de réaliser des images tridimensionnelles avec une résolution spatiale de 60 mètres.
- En 1972 les États-Unis lancent le programme Landsat ses premiers objectifs étaient l'évaluation des volumes de récolte céréalières en URSS et aux États-Unis afin d'anticiper l'évolution des cours. Le dernier, Landsat 9, a été lancé le 27 septembre 2021.
- En 2002, c'est au tour de agence spatiale européenne de lancer son propre satellite d'observation ENVISAT (ENVironment SATellite) dont l'objectif est de mesurer de manière continue à différentes échelles les principaux paramètres environnementaux de la Terre relatifs à l'atmosphère, l'océan, les terres émergées et les glaces. C'est aussi le début de l'utilisation de la spectro-radiométrie permettant d'obtenir des images hyperspectrales.
- En 2014, c'est le programme Copernicus avec les satellites Sentinel, lancé en Europe. L'accès aux données des satellites est libre et gratuit. Le premier modèle fourni des images à 10 mètres de résolution sur plusieurs bandes spectrales. Ces très hautes résolutions permettent d'échantillonner et d'estimer différents indices sur de larges zones agricoles. Le programme Copernicus est prévu jusqu'en 2029.

Parallèlement à l'imagerie satellite, le développement de drones aériens pour l'observation a été initié pour des usages militaires. C'est dans les années 2000 que les premiers usages agricoles des drones font leurs apparitions, dont le gain en autonomie, précision, résolution (<10 cm) et temps d'acquisition sont bien plus compétitifs que les satellites, en particulier à l'échelle de la parcelle [Lherbier, 2005, Bossu, 2007, Koot, 2014, Pretto et al., 2019]. La Figure 1.8 montre l'évolution de la résolution spatiale.

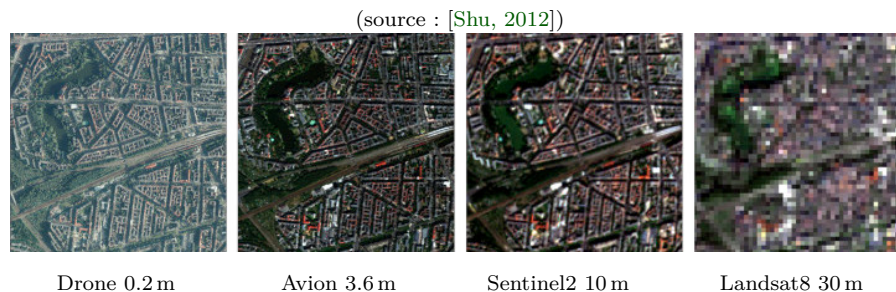


Figure 1.8 – Évolution de la résolution spatiale

Les recherches en cours tendent à ajouter un degré de précision supplémentaire allant jusqu'à une résolution millimétrique. Là où les satellites et les drones aériens présentent une résolution insuffisante, ces travaux se concentrent sur le développement de l'imagerie de proxi-détection, embarquée sur machines agricoles usuelles ou sur drones terrestres [Machleb et al., 2020].

Apport de la géomatique Avec l'apparition de l'informatique, c'est dans les années 1960 que la géomatique fait son entrée et que le premier véritable SIG opérationnel est conçu à Ottawa, au Canada par le ministère des forêts. Afin d'obtenir des informations sur géographiques les sols, l'agriculture, la faune, la flore, et la sylviculture. La géomatique appuyée par des observations agronomiques permet de définir de premiers leviers d'intervention pour l'agriculteur [Lherbier, 2005]. La géomatique, c'est avant tout une méthodologie, en trois points :

1. **L'observation** qui consiste à mettre en évidence et à caractériser la variabilité spatiale inter-parcellaire. Cela consiste à collecter de l'information sur le milieu et les plantes pour mieux connaître les hétérogénéités et les échanges.
2. **L'analyse** des données pour comprendre cette variabilité, étudier l'influence de différents facteurs sur une variable et l'étude des impacts de cette variabilité sur l'opération culturale à réaliser afin d'établir les incidences économiques et environnementales.
3. **L'action**, c'est définir avec l'agriculteur les moyens d'action et moduler ces pratiques en fonction de l'itinéraire technique qu'il souhaite mettre en œuvre.

La géomatique, travail donc a l'échelle du territoire (bassin versant, vallée, région, ...), c'est-à-dire les interactions entre différentes parcelles et les mouvements de flux tels que les polluants, les sédiments ou autres. Tous ces facteurs ont un rôle important à ces grandes échelles par exemple sur les rendements et la biodiversité comme nous l'avons vu précédemment. Ainsi des modèles d'agencement sont modélisables (figure 1.9) permettant d'optimiser les rendements à l'échelle du territoire [Caquet et al., 2020].

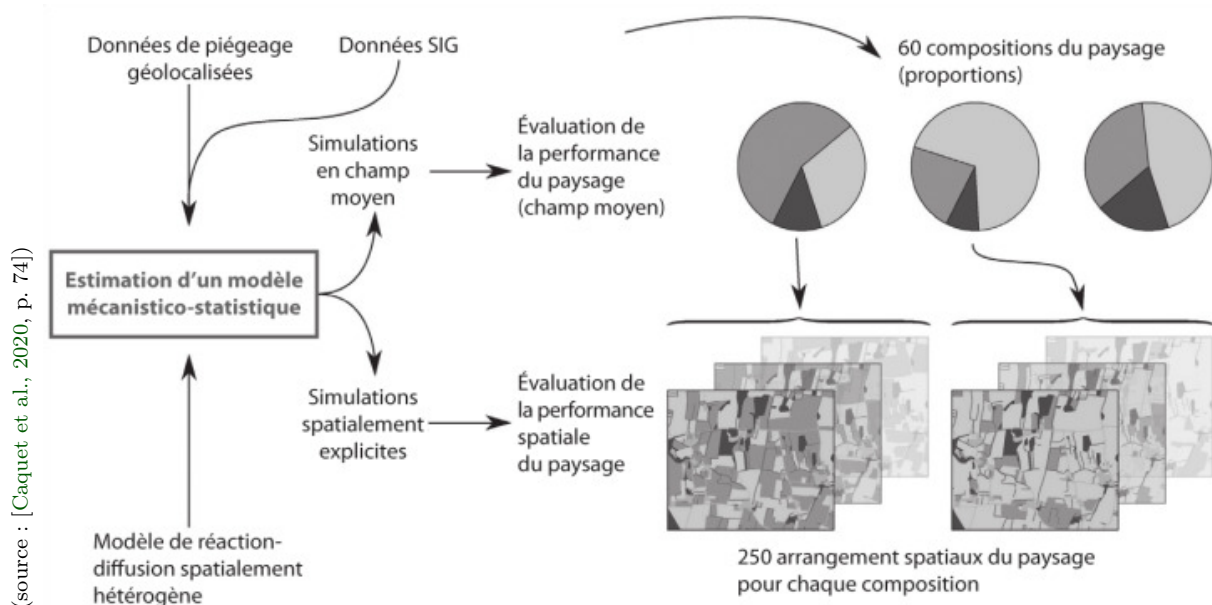


Figure 1.9 – Simulation de l'agencement spatial de paysages

En augmentant la résolution de ces observations, des modèles de simulation des interactions entrent les plantes ont commencé à voir le jour pour montrer l'influence de différents facteurs sur l'ensemble du système (itinéraire technique, adventice, ...) à l'échelle de la parcelle, tel-que FLORSYS [Colbach et al., 2021]. Ces méthodes ont donc permis de passer d'une gestion des doses (intrant, phytosanitaire, ...) de l'exploitation (homogène et uniforme), à un traitement des zones "dégradées" de la parcelle après analyse [Lherbier, 2005].

1.2.2 Vers une agriculture numérique

Parallèlement au développement de l’agriculture de précision, de nouvelles technologies ont récemment émergé : réseaux de télécommunication, internet des objets [Tang et al., 2021], fouille de données, apprentissage profond et intelligence artificielle . L’utilisation et le couplage de l’ensemble de ces outils numériques conduit à l’émergence d’un nouveau domaine : l’agriculture numérique [Birner et al., 2021]. Les applications envisagées reposent sur l’utilisation de multiples capteurs qui peuvent être embarqués dans plusieurs drones [Pretto et al., 2019]. Ainsi, un drone terrestre peut être piloté à partir d’images acquises par des drones aériens. (figure 1.10). Une communication entre modules est alors nécessaire.

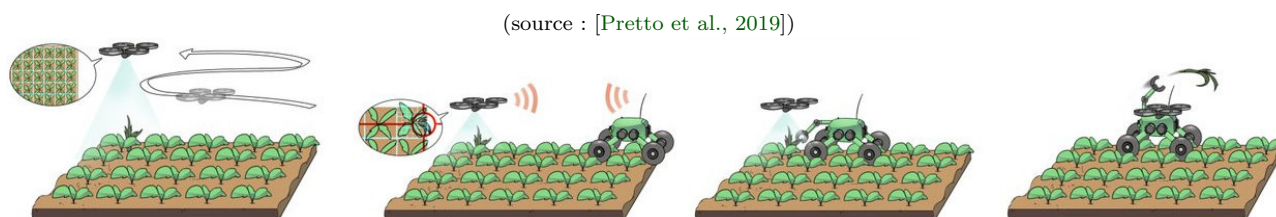


Figure 1.10 – Concept de projet en agriculture numérique

L’ensemble des informations collectées permettent de constituer des outils d’aide à la décision pour accompagner les agriculteurs dans leurs interventions sur les parcelles en temps réel [Bellon-Maurel and Huyghe, 2019, Saddik et al., 2022]. Il est ainsi possible d’envisager : d’identifier des plantes au champs (espèce, stade de développement), d’appliquer des produits phytosanitaires au plus près des cultures, ou d’éliminer les plantes indésirables tout en préservant les plantes présentant des services écosystémiques [Green et al., 2021].

Lien avec l’agroécologie : L’agriculture de précision, et par extension l’agriculture numérique peuvent permettre de réduire l’impact environnemental des différents traitements précédemment présentés (herbicides, fongicides, insecticides, fertilisants, ...), en diminuant les surfaces traitées et les doses appliquées. Les problématiques liées au tassement des sols seront grandement amoindries, puisque ce type d’approche peut être embarqué sur de “petits” robots autonomes, plus légers. Malgré tout, ces approches restent pour l’instant hyper-spécialisées et ne laissent donc que peu de place à l’hétérogénéité des cultures, ce qui implique que les processus agroécologiques (interactions biotiques, services écosystémiques) ne sont pas encore mobilisables [Caquet et al., 2020, p. 81].

Phénotypage : A contrario cette hyper-spécialisation n’est pas un frein dans le domaine du phénotypage à haut débit. Cela permet d’accélérer les progrès dans la sélection des variétés [Cabrera-Bosquet et al., 2012], pour répondre aux attentes des différents acteurs en augmentant la rapidité d’acquisition des mesures, qui sont de plus non-destructives [Jiang and Li, 2020]. C’est le cas par exemple de la culture d’igname, dont le manque de variétés améliorées sur le rendement et les attributs de qualité, entraînent un déclin de la filière aux Antilles et freinent l’adoption de nouvelles variétés en Afrique de l’Ouest. L’objectif est de quantifier le rendement par la dynamique de recouvrement, la qualité “subjective” est associée à la couleur de la chair des tubercules et son oxydation dans le temps. La qualité nutritionnelle des produits à base d’igname, peut être déterminée par la taille et la forme des grains d’amidon, en utilisant l’analyse d’image [Mell, 2018, CORNET et al., 2018].

1.3 Le désherbage

1.3.1 Place des adventices dans les systèmes de culture

Les adventices sont des plantes, non désirées ni semées par l'agriculteur, se développant naturellement dans un système de culture (ou agrosystème). En fonction du mode de production, les adventices font l'objet d'une attention particulière. Nous resterons à partir de maintenant dans le cadre de l'agriculture conventionnelle et de l'agriculture de précision.

Dans ce contexte, depuis le début du XXe siècle, les adventices sont considérées comme indésirables car responsables de la perte de rendement et de qualité [Cordeau et al., 2016]. On parle dans ce cas, de nuisibilité primaire [Caussanel, 1989]. Elle se mesure avec le nombre de pieds d'adventices par m² qui occasionne une baisse de rendement de 5 % en système conventionnel. Cependant, ce paramètre de densité de plantes a été remis en cause, car il ne prend pas en compte le stade de la végétation. C'est pourquoi de nouveaux paramètres ont été proposés comme le ratio de biomasse entre adventices et cultures (BMw/BMc), ou en étudiant le rapport des surfaces foliaires, plus simple à caractériser par des méthodes non-destructives [Gée et al., 2021]. La nuisibilité secondaire concerne les conséquences engendrées sur la parcelle sur le long terme (stock semencier, salissement, etc).

Pour acquérir une compréhension fiable des risques de réduction des récoltes dans le temps, des expériences acquises localement dans des conditions comparables sur 110 essais de contrôle des adventices ont été menées en France entre 1993 et 2015 sur trois grandes cultures annuelles (blé tendre, colza et tournesol). Des réductions de 25 q/ha, 4 q/ha et 3.5 q/ha ont respectivement été observées [Cordeau et al., 2016]. Les adventices rivalisent donc avec leurs plantes voisines par une concurrence accrue pour les nutriments (azote, potasse, phosphore), la lumière et l'eau. La nuisibilité des adventices reste un sujet d'étude par la communauté scientifique. Comprendre la compétition des adventices avec la culture pour les ressources du milieu (lumière, eau, nutriments, ...) est très complexe du fait d'une interaction multi-spécifique et qui dépendra du stade de développement des plantes [Morison et al., 2008, Colbach et al., 2021]. Toutefois, d'autres études, suggèrent qu'il existe un manque de connaissances concernant la "réelle" nuisibilité pour les cultures [Bendaoud, 2020, Jouan et al., 2022].

A travers cette complexité, la conséquence pratique d'une telle gestion devient donc pour les agriculteurs de trouver des solutions globales simples et pratiques de mise en œuvre ; c'est ce que la chimie de synthèse a pu apporter avec des herbicides systémiques (à large spectre) ou spécifiques (pour une espèce particulière). Cependant, certaines adventices offrent par exemple des services écosystémiques (figure 1.11), d'autres sont comestibles (figure 1.12) et enfin certaines sont plus ou moins "toxiques", mais peuvent trouver des usages médicinaux (figure 1.13).

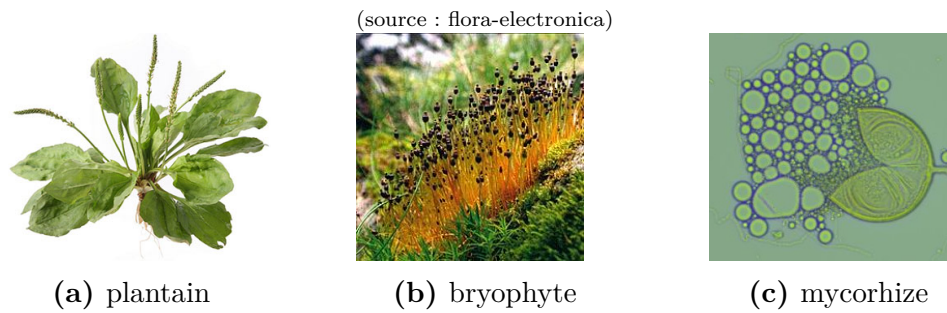


Figure 1.11 – Plantes bio-indicatrices, fixatrices et symbiotiques

Certaines adventices aident par exemple, à la structuration du sol, à la réduction des phénomènes érosifs, à l'absorption des excédents de fertilisation et sont des indicateurs d'un état particulier du milieu comme le plantain. D'autres permettent la fixation de l'azote tels que les bryophytes et des champignons augmentant significativement les rendements avec les mycorrhizes [Zerbib, 2018]. Elles participent également à la diversité écologique du milieu en servant de refuge et en pourvoyant des ressources alimentaires (pollen, nectar) aux auxiliaires des cultures [Schaub, 2010, Morison et al., 2008].

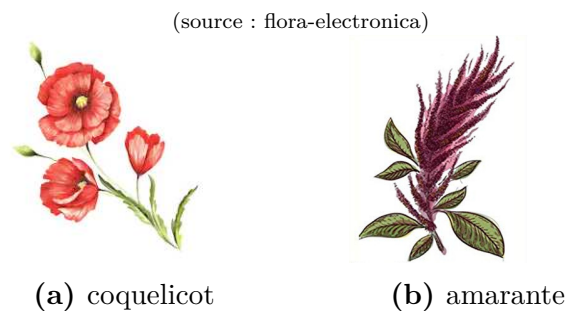


Figure 1.12 – Adventices comestibles

Certaines sont également comestibles et des phénomènes de résistance aux herbicides ont commencé à émerger. C'est le cas du coquelicot en France et de l'amarante aux USA due à la sur-utilisation d'herbicide et entraînant des mutations.

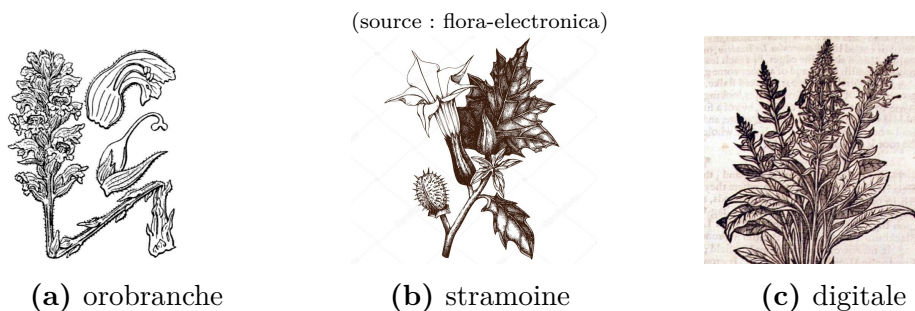


Figure 1.13 – Adventices parasite, psychotrope et médicinale

Certaines adventices ont toutefois des inconvénients importants. C'est le cas de l'orobranche, une plante parasite qui utilise les nutriments de l'hôte, du stramoine plante chamanique fortement présente dans le maïs et qui provoque des délires. Ainsi que de la digitale qui est une plante médicinale à très faible dose.

1.3.2 La biologie des adventices

La connaissance de la biologie des adventices est essentielle pour mettre en œuvre des stratégies de désherbage. Ceci est d'autant plus important pour les systèmes d'agriculture biologique où l'utilisation d'herbicides est interdite. Parmi les principales notions à connaître sur les adventices, il y a : la profondeur de germination (figure 1.14), l'époque préférentielle de germination (figure 1.15) et le taux annuel de décroissance (figure 1.16) de ces adventices.

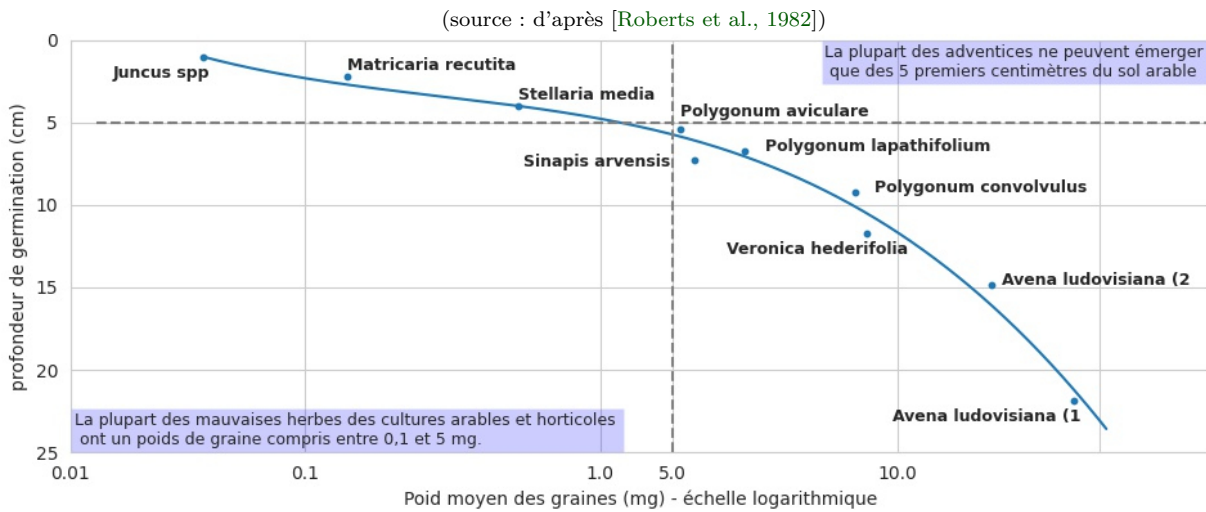


Figure 1.14 – Profondeur de germination en fonction de la masse des graines

Aussi, les adventices se caractérisent par une profondeur de germination préférentielle. Pour la plupart d'entre elles, la germination se produit principalement à une profondeur de moins de 5 cm. La plupart des semences sont incapables de germer au-delà de 10 cm de profondeur. Donc enfouir les graines en profondeur par le labour permet d'empêcher leurs levées, même si elles ont tendance à se conserver plus longtemps. Il y a cependant quelques exceptions, comme la folle avoine, une espèce à grosses graines qui peut germer au-delà de 10 cm.

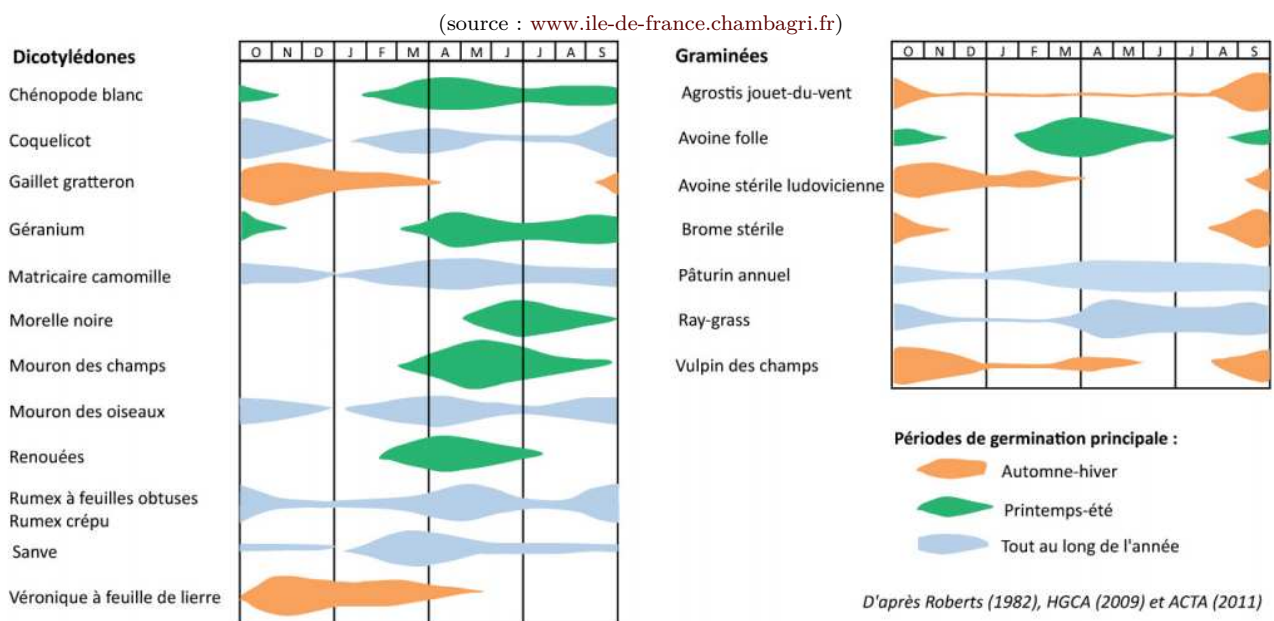


Figure 1.15 – Période préférentielle de germination de certaines adventices

Les adventices annuelles se caractérisent aussi par leur période de germination. Certaines adventices apparaissent sur une période relativement courte (Vulpinia, qui apparaît en automne et en hiver) ou, à l'inverse, sur une longue période, voire toute l'année (pâturin annuel). La concentration des plantations sur une période de temps limitée conduit à une spécialisation de la flore. Par exemple, les systèmes dominés par les cultures d'hiver (colza-blé) favorisent le développement des adventices d'automne-hiver comme Vulpinia.

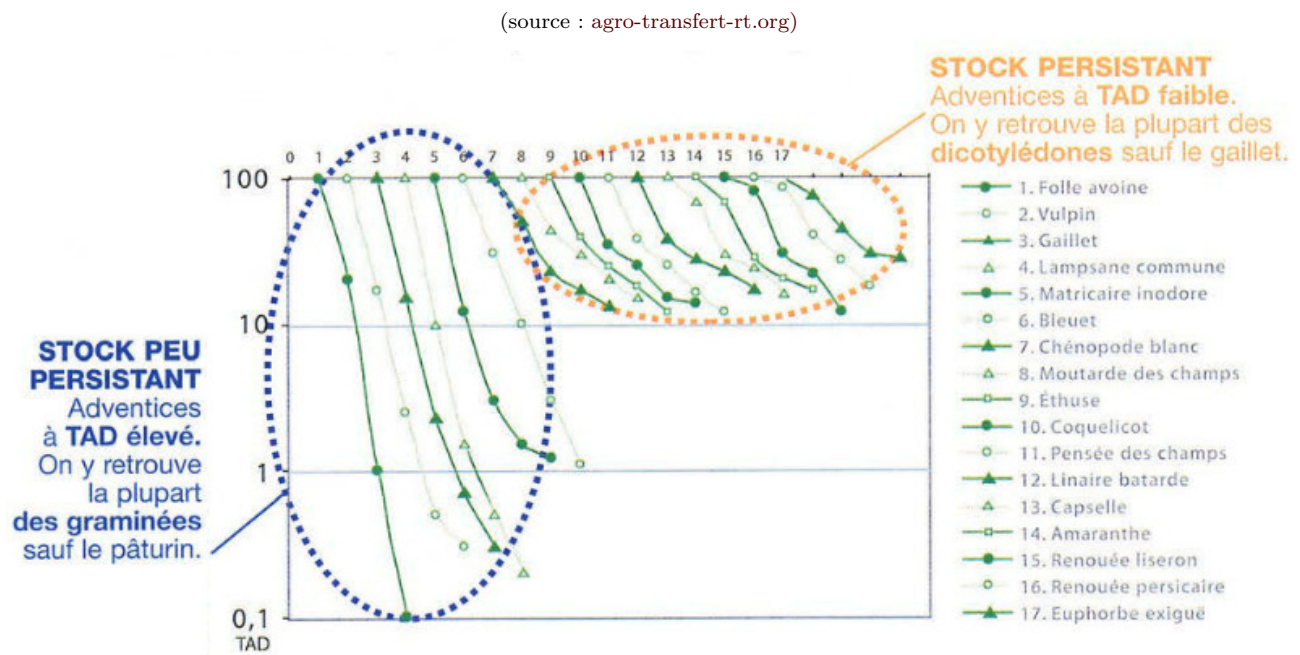


Figure 1.16 – Taux annuel de décroissance du stock semencier des adventices

Finalement, il y a le Taux Annuel de Décroissance (TAD) qui correspond au pourcentage de graines qui disparaissent d'une année à l'autre. Il montre l'évolution de la viabilité des graines d'adventices dans le sol. Le nombre de graines viables diminue chaque année proportionnellement à la valeur TAD. Plus le TAD est élevé, plus la durée de vie des semences est courte. Les adventices à faible TAD sont plus difficiles à contrôler, car le labour est beaucoup moins efficace pour lutter contre ce type d'adventices, puisqu'elles seront encore capables de germer lors du prochain labour. La plupart des graminées ont un TAD élevé, tandis que les graines des dicotylédones telles que le coquelicot, le chénopode ou la renouée ont une durée de vie assez longue. Ainsi, la durée de vie des graines est très variable d'une espèce à l'autre. Les graines des adventices céréalières ont une durée de vie exceptionnellement longue. Certaines on pu germer après 1700 ans [Wesson and Wareing, 1967].

1.3.3 Fenêtres de traitement

Les moments de gestion des adventices peuvent être définis par cinq "fenêtres d'opportunité", déterminées par le cycle de la culture semée (Figure 1.17) et le stade de vie des adventices [Storkey et al., 2021]. (fenêtre 1) Avant le semis d'une culture, les herbicides non-sélectifs et le travail du sol peuvent être utilisés pour contrôler les adventices dans le stock semencier ou les plantules émergentes. (fenêtre 2) Après le semis, mais avant l'émergence de la culture, des herbicides de pré-émergence (PRE-herbicides), des cultures de couverture ou le paillage peuvent être utilisés pour contrôler sélectivement la levée des adventices. (fenêtre 3). Les herbicides sélectifs et la lutte mécanique contre les adventices peuvent être utilisés lorsque les plantules sont suffisamment petites pour être contrôlées et avant qu'elles ne commencent à concurrencer la culture (fenêtre 4). Une fois que les adventices commencent à se disputer les ressources, l'augmentation de la compétitivité du couvert végétal peut engendrer des pertes de rendement et le retour des graines d'adventices. Enfin, les progrès récents dans le contrôle des graines des adventices après la récolte ont également soulevés la possibilité de réduire les populations d'adventices en détruisant les graines fraîches (fenêtre 5).

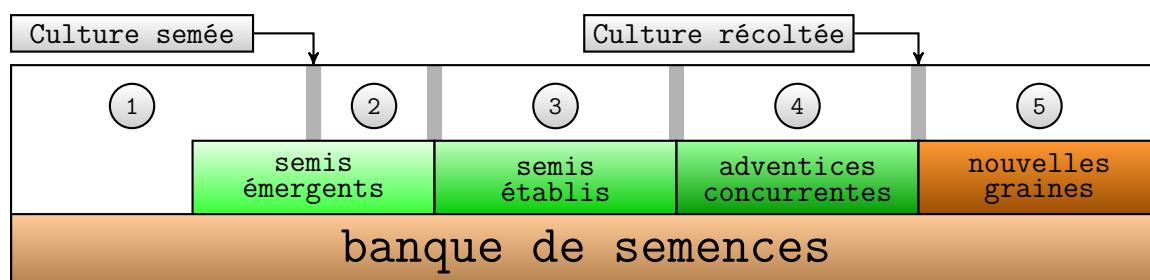


Figure 1.17 – Illustration de cinq fenêtres d'opportunité pour la gestion des adventices

Les solutions de gestion reposent donc sur une connaissance approfondie du cycle de développement et cycle de reproduction [Chauvel et al., 2018]. La rotation des cultures, le désherbage, le travail du sol et les mesures préventives sont les seuls leviers d'influence sur les populations d'adventices des parcelles. Néanmoins, les solutions de gestion des adventices favorisent les espèces qui ont le même cycle de développement et les mêmes caractéristiques (monocotylédone ou dicotylédone) que les cultures. À travers cette thèse, nous proposons un axe de recherche permettant la réduction des produits phytosanitaires. Plus spécifiquement, dans le cadre du Challenge RoSE, nous nous intéressons aux stades de développement les plus précoces des adventices, afin d'en limiter leur développement (fenêtre 1 et 2).

Ainsi, selon les systèmes de culture étudiés, il pourrait être intéressant d'arriver à identifier les espèces les plus nuisibles, leurs périodes de levées préférentielles pour identifier celles qui vont se reproduire le plus vite pour mieux les contrôler.

1.3.4 Les technologies de désherbage actuelles

Le pulvérisateur est la principale machine agricole utilisée pour le désherbage chimique. Ce dernier se compose d'une cuve, qui contient la bouillie (mélange eau+produit phytosanitaire) à appliquer ainsi que le système de mise sous pression du fluide, et d'une rampe d'épandage dont la bouillie sous pression circule vers les buses qui vont fragmenter le jet en gouttelettes (figure 1.18). Un système de régulation permet à chaque instant de contrôler le débit de pulvérisation selon l'avancement de la machine pour maîtriser le volume de bouillie épandu par hectare.



Figure 1.18 – Exemple de pulvérisateurs

Le choix des buses dépend de différents facteurs tels que, la viscosité du fluide, la pression, le débit, la vitesse de pulvérisation, la taille des gouttelettes et la dérive en fonction de la couverture souhaitée (figure 1.19). À noter le développement de buses anti-dérive à injection d'air, permettant de créer des gouttelettes plus grosses, et donc moins sensibles à la dérive [Vulgarakis Minov, 2015]. Leur utilisation permet de réduire la largeur des zones non traitées en bordure des points d'eau (de 20 cm ou 50 cm à 5 m) et, sous certaines conditions, les distances non traitées à respecter vis-à-vis des riverains.

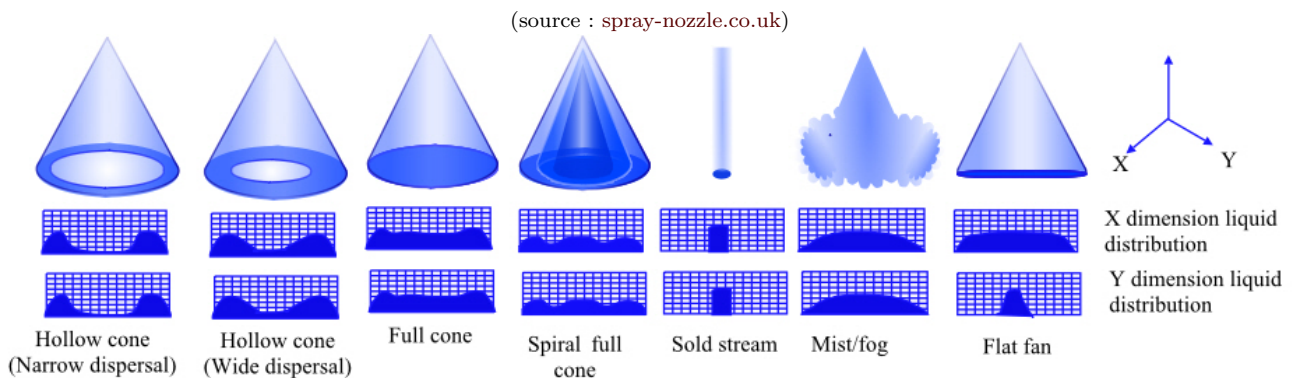


Figure 1.19 – Différentes formes de buses associées

Appuyés par les GPS/GNSS et la géomatique, les systèmes de pulvérisation se sont grandement améliorés afin de prendre en compte la variabilité spatiale. Ainsi, après l'observation et l'analyse des données du champ, une carte géoréférencée (dite de préconisation) est embarquée sur le pulvérisateur permettant d'agir (ON/OFF) localement grâce à des électrovannes installées sur les rampes. Ces dernières sont commandées par un ordinateur intégré (ou terminal isobus) et une communication ISOBUS : c'est ce que l'on appelle la coupure de tronçon. Permettant ainsi d'ouvrir ou non toutes les buses sans modulation. Depuis, il est possible de faire une commutation sur un porte-buse type quadri jet avec une coupure électrique permettant d'intégrer la coupure individuelle de la buse. Enfin, de nouvelles buses à pulsation (reposant sur la technologie Pulse-Width Modulation) ont fait leur apparition, permettant de faciliter le traitement localisé pour aller vers une modulation de dose.

Dans la même lignée de technologie de traitement que la pulvérisation, les systèmes de pulvérisation ponctuelle (spot spraying) ont fait leur apparition. Ce sont des appareils, permettent de traiter “uniquement” l’inter-rang (tels que le WeedSeeker ou le Weed-It). Le dispositif attaché à la rampe de pulvérisation (visible en rouge sur la figure 1.20), permet de détecter la présence de végétation grâce à une photodiode et l’utilisation d’indice de végétation (présentés en section 2.2.3). Lorsque de la végétation est détectée, l’électrovanne la plus proche est ouverte afin de traiter uniquement la plante. Cette solution, bien que limitée à l’inter-rang ou aux sols nus, permet de considérablement réduire l’utilisation des intrants.

(source : avidorhightech.com)



Figure 1.20 – Présentation du WeedSeeker en viticulture

Dans l’optique de s’affranchir des herbicides, il faut s’intéresser à des solutions alternatives de désherbage. Il existe trois grandes catégories, présentées sur la figure 1.21. Par exemple (a) les solutions dites mécaniques, tel-que la houe rotative, va simplement arracher les plantes émergentes. Similaires aux solutions de pulvérisation, les solutions thermiques (b) utilisent des combustibles fossiles, tel que le gaz pour brûler les adventices. Finalement, la dernière alternative (c) est le recours à l’électrique, des bandes conductrices sont appliquées au sol dans l’inter-rang. La capacité à désherber sans herbicide, permet de s’affranchir de l’apparition de résistance chez les adventices ainsi que des problèmes environnementaux liées à leur utilisation.

(source : hatzenbichler.com)



(a) Houe rotative

(source : annuaire-agricole.fr)



(b) Gaz

(source : terre-net.fr)



(c) Électrique

Figure 1.21 – Exemples de méthodes alternatives de désherbage

Actuellement, ces méthodes reposent essentiellement sur un traitement de l’inter-rang. Pour aller plus loin, les recherches actuelles, basées sur du traitement de l’image visent à détecter les plantes dans l’intra-rang. Ces méthodes sont exposées dans la section suivante.

1.3.5 Les technologies de désherbage en développement

La recherche actuelle rentre davantage dans le cadre de l'agriculture numérique et les concepts sont nombreux [Korres et al., 2019, van Mourik et al., 2021, Kakamoukas et al., 2021]. Ainsi, une présentation exhaustive des différentes possibilités ne sera pas faite et le focus sera fait autour du Challenge ROSE. L'objectif du challenge est d'inciter à la collaboration entre équipes de recherche et industries. Ainsi différents consortiums ont été regroupés pour évaluer leurs performances respectives. Quatre consortiums ont été sélectionnés (figure 1.22).

(source : challenge-rose.fr)



De gauche a droite : BIPBIP, PEAD, Roseau et WeedElec

Figure 1.22 – Les différents consortiums de l'ANR Challenge ROSE

Chacun de ces consortiums propose des solutions techniques avec leurs outils respectifs (figure 1.22) dans une ou plusieurs étapes (i) détection des plantes, (ii) interprétation : culture ou adventices ? (iii) décision : faut-il traiter ? et (iv) action : mécanique, chimique, thermique ou électrique. Une brève présentation pour chacun d'entre eux est faite ci-dessous.

Solution BIP BIP : Le consortium BIP-BIP, pour Bloc-outil et Imagerie de Précision pour le Binage Intra-rang Précoce, propose « de développer une solution mécanique pour le désherbage de cultures maraîchères et de grandes cultures au stade précoce. Cette solution repose sur un module de binage autoguidé par imagerie et télémétrie, couvrant un rang de culture. Le module s'appuie sur un système de vision fournissant la position des cultures et adventices transmise à un système décisionnel contrôlant le dispositif mécanique réalisant le binage proprement dit ». Le module est donc composé d'une caméra, d'une unité de calcul et d'un doigt mécanique qui est actionné au besoin en évitant les cultures. Ce dernier gratte la surface du sol permettant d'éliminer les adventices émergentes (figure 1.23).

(source : challenge-rose.fr)



Figure 1.23 – Module de désherbage BIP-BIP

Solution PEAD : Ce consortium propose de gérer les tâches “action - interprétation” grâce à une caméra hyperspectrale spécifique développée par Carbon Bee. Des solutions de vision informatique basées sur l’intelligence artificielle ainsi que des systèmes d’imagerie hyperspectrale sont mis en œuvre. Les applications se déclinent aujourd’hui dans l’agriculture de précision permettant la détection précoce de stress et maladie dans les cultures ainsi que les adventices. La solution repose sur trois modules, une caméra, un contrôleur et un afficheur (figure 1.24)



Figure 1.24 – Les trois modules CarbonBee (caméra hyperspectrale, contrôleur, afficheur)

Solution WeedElec : La solution WeedElec repose également sur du deep-learning pour les tâches “détection/décision”. Ils sont innovateurs sur deux aspects : (i) ils s’appuient sur de la robotique à faible consommation énergétique et légère, leur permettant d’être autonomes en énergie grâce à un panneau solaire et (ii) leurs méthodes d’action s’appuient sur des décharges électriques, une électrode est téléguidée grâce à un bras robotisé (figure 1.25). Cette solution permet la modulation de l’intervention de manière extrêmement précise, puisque en fonction de la taille de l’adventice visée et de la force du choc électrique, l’impact sur cette dernière est différent : entre ralentissement de la croissance et brûlure.



Figure 1.25 – Électrode téléguidée par le bras robotique

Solution Roseau : Le projet ROSEAU vise à développer des outils qui interviennent sur trois composantes (perception/décision/action) via des couches de décision à plusieurs niveaux, allant du réflexe (temps-réel) aux réflexions élaborées (pouvant prendre plusieurs heures). Les différentes couches s’appuient sur des données à différentes échelles spatio-spectro-temporelles (depuis un drone et depuis un robot mobile terrestre) en provenance de caméras sensibles aux domaines du visible et de l’infrarouge. Après la prise de décision le traitement est assuré par un module spécifique, qui correspond actuellement à une roue de binage (figure 1.26).



Figure 1.26 – Outil de binage ROSEAU

1.4 Conclusion de chapitre

Comme nous l'avons vu, d'étroites relations existent entre les différents domaines liés à l'agriculture [Dainese and Martin, 2019] : la richesse des écosystèmes, la simplification des structures paysagères, les services écosystémiques, les pollinisateurs, la gestion des maladies, les flux de sédiments ... Ces liens démontrent l'importance de la connaissance des espèces, de la dynamique des systèmes, de l'interaction entre les éléments (figure 1.27) et de l'impact de nos modes de production [Nicholson et al., 2019] afin d'en restreindre leurs effets néfastes [Modica and Zoboli, 2016]. L'agriculture de précision et l'agriculture numérique sont des voies agroécologiques possibles à la réduction des conséquences de l'agriculture "moderne".

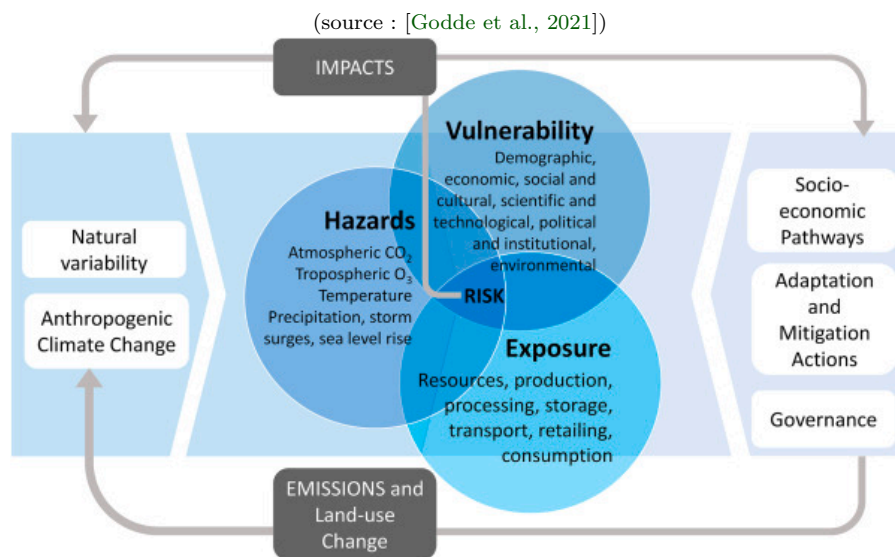


Figure 1.27 – Schéma de l'interaction entre le système climatique physique, l'exposition et la vulnérabilité produisant des risques dans la chaîne de production.

Les solutions de traitement en cours de développement dans l'agriculture de précision, reposent principalement sur l'analyse d'image. Cette thèse s'inscrit dans ce domaine de par l'utilisation de méthodes d'analyses d'images classiques et de réseaux de neurones. Dans le cadre du Challenge RoSE, les images utilisées dans cette étude sont acquises par une caméra multispectrale fixée sur un drone terrestre, visant à fournir une analyse en "temps réel".

2 Vision par ordinateur

Durant ce chapitre, nous aborderons un état de l'art détaillant les différents systèmes optiques et méthodes de détection par imagerie. La revue de [Perez-Sanz et al., 2017] montre les différentes étapes classiques en analyse d'image associées à la détection et classification des adventices. Afin de répondre à nos objectifs, ces différents traitements ont été mis en place et résumés dans la figure 2.1 ci-dessous. La présentation de ces méthodes permettra d'avoir une vision globale des possibilités en analyse d'image et de comprendre les choix qui ont été faits durant cette thèse.

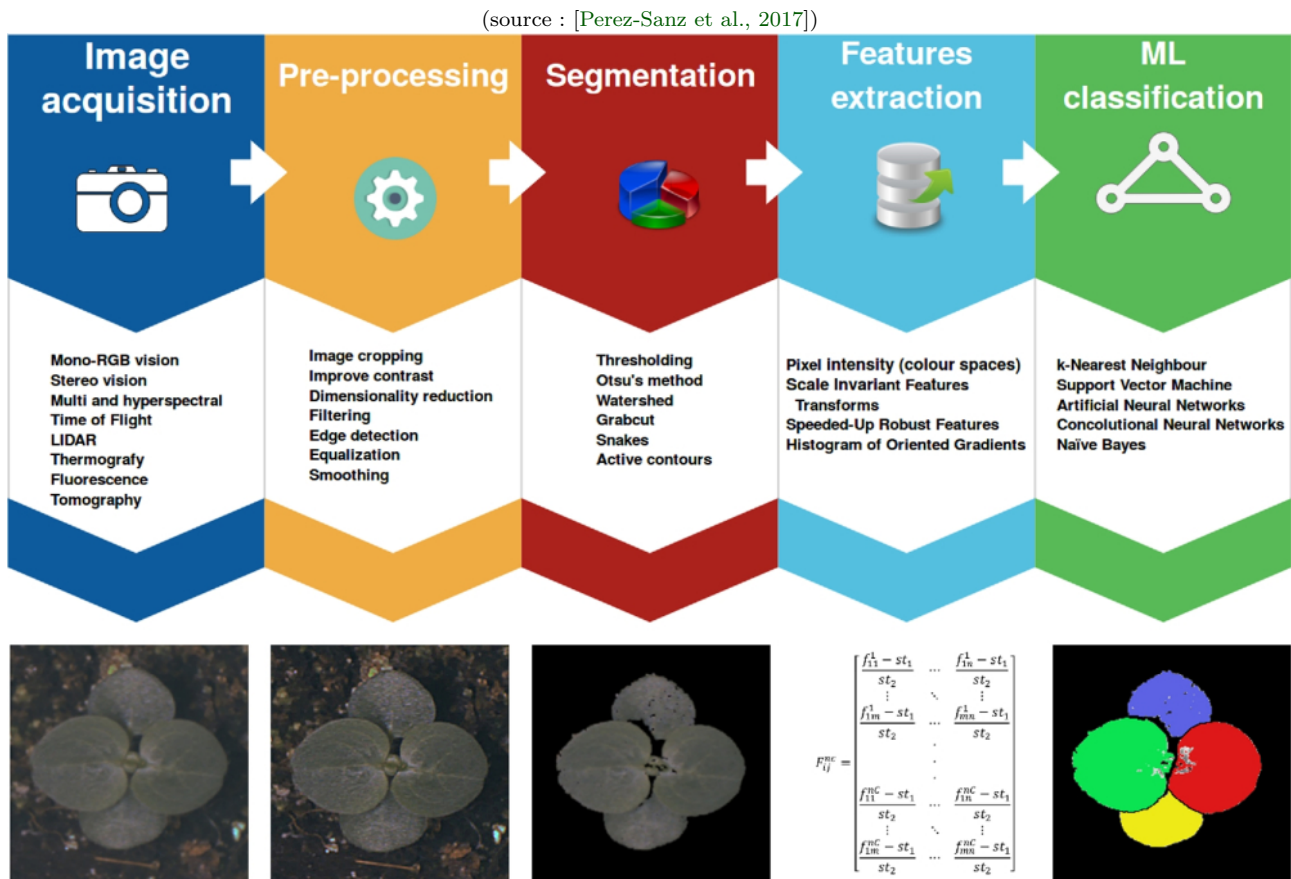


Figure 2.1 – Étapes clés de l'analyse d'images agricoles

Ainsi, ce chapitre est structuré en 5 parties, reprenant ces étapes clés de l'analyse d'image : les différents systèmes d'acquisition, les pré-traitements, la segmentation, l'extraction des paramètres et les méthodes de classification. Finalement, la conclusion permettra de synthétiser les orientations choisies durant cette thèse selon les avantages et inconvénients de chacune des méthodes par rapport au cahier des charges initial.

2.1 Systèmes d'imagerie

Pour construire un outil de vision par ordinateur, il faut comprendre comment se forment les images numériques. Un capteur en contraste noir et blanc fonctionne en estimant la quantité de photons qui frappent chaque pixel dans un laps de temps donné. Une image peut être vue comme une matrice bidimensionnelle de pixels. Ainsi, le système optique intègre donc le spectre de la lumière sur une certaine plage (380-780 pour le spectre visible) que l'on peut aussi écrire $Luminance = \int_{380}^{780} f(\lambda)d\lambda$ transformant le spectre visible $\mathbb{R}^\infty \rightarrow \mathbb{R}$.

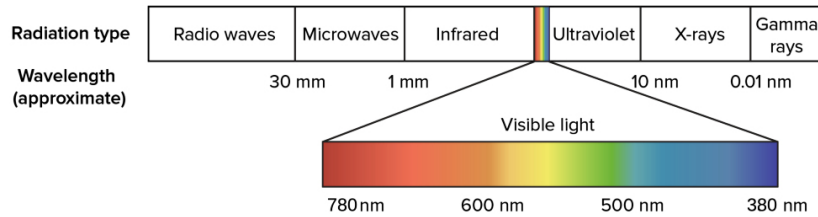


Figure 2.2 – Spectre électromagnétique et spectre visible

Pour acquérir une image, on peut utiliser un capteur CCD (Charge-Coupled Device) ou un CMOS (Complementary Metal Oxide Semiconductor). Ces éléments sont composés d'une matrice de photosites (semi-conducteurs photosensibles). Ces derniers convertissent l'information lumineuse reçue en tension électrique (analogique) exprimée en volts, puis en un nombre entier à travers un convertisseur analogique-numérique (CAN). Notons que les valeurs numériques ont une précision limitée, i.e. l'information est encodée sur un nombre de bit fixe. Avant d'arriver à la matrice de détection, la lumière passe par un système optique, composé d'une ou plusieurs lentilles et d'un filtre pour limiter la plage d'intégration (visible), tel que représenté ci-dessous (figure 2.3).

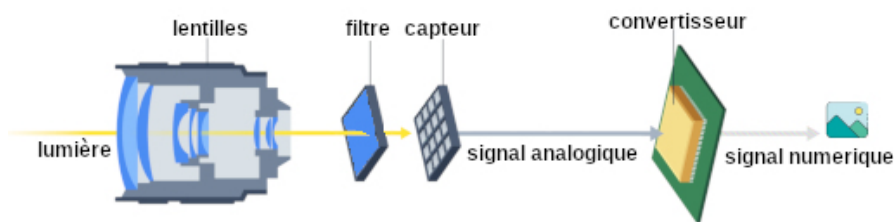


Figure 2.3 – Les différents composants d'une caméra

Dans un système optique, il y a quatre facteurs importants dans la formation de l'image : la distance focale, l'ouverture, le gain et le temps d'intégration. La combinaison de ces trois derniers facteurs contrôle la luminosité moyenne de l'image que l'on appelle également "le triangle d'exposition". Grâce à certaines heuristiques, le meilleur équilibre du triangle d'exposition peut être défini. Chaque constructeur de caméra ou d'appareil photo propose une méthode propriétaire embarquée dans l'appareil pour effectuer un ajustement automatique du triangle d'exposition.

La distance focale : Exprimée en mm, elle représente la distance entre le centre optique et la surface du capteur. Plus la longueur focale est courte, plus l'angle d'observation est large et vice-versa. Par exemple, un objectif de 8 mm peut avoir un champ visuel de 110° et un objectif de 50 mm peut avoir un champ visuel de 32°. Comme l'ouverture, la focale influence la profondeur de champ. Plus la focale est importante, plus la profondeur de champ est faible. Cela avantage les petits capteurs car à champ de vision équivalent, plus le capteur est petit, plus la focale est petite.



Figure 2.4 – Effet de la longueur de la focale sur le champ de vision

L'ouverture : C'est le diamètre d'ouverture du diaphragme de l'objectif. Par exemple, l'objectif peut avoir un diamètre de 70 mm mais seulement 5 mm peuvent collecter la lumière. Une grande ouverture de l'objectif laissera entrer beaucoup de lumière, mais la profondeur de champ sera plus faible, ce qui signifie que seule une petite gamme de distances sera mise au point. Une grande profondeur de champ, donc une petite ouverture est préférable pour la perception robotique, car nous voulons capturer autant de détails que possibles dans la scène.

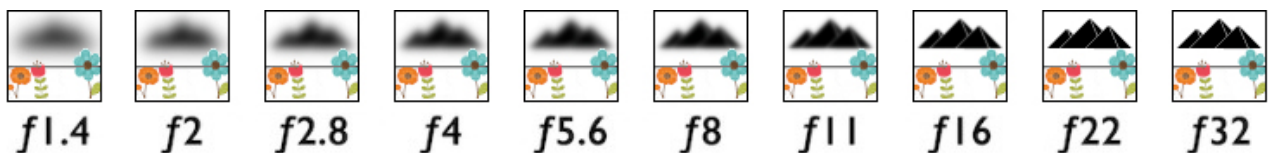


Figure 2.5 – Effet de l'ouverture sur la profondeur de vision

Le gain : Ce paramètre contrôle la luminosité de l'image. Le niveau de tension mesuré par pixel peut être amplifié par un facteur, appelé gain ou sensibilité ISO. Un gain plus élevé signifie une image plus lumineuse et vice-versa. L'amplification du signal induit également l'amplification du bruit. Il faut donc trouver un équilibre entre l'amplification et le bruit. En condition de faible luminosité, ce paramètre est essentiel pour distinguer les objets.

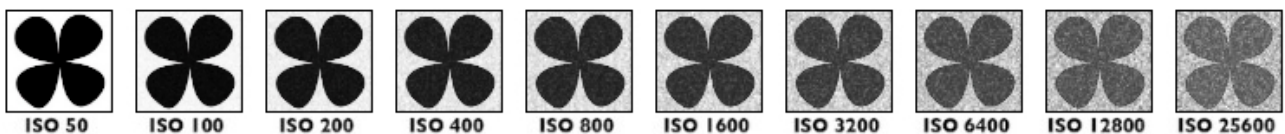


Figure 2.6 – Effet du gain sur le bruit de l'image

Temps d'intégration : Il s'agit de la durée, en secondes, pendant laquelle le capteur collecte la lumière. La tension mesurée sera la somme de tous les photons collectés pendant le temps d'intégration. Une vitesse d'obturation plus longue laissera entrer plus de lumière, mais rendra tout mouvement flou. Ce phénomène est communément appelé le "flou de bougé".



Figure 2.7 – Effet du temps d'intégration (en seconde) sur le flou de bougé

2.1.1 Percevoir la couleur

Pour percevoir la couleur plutôt que les niveaux de gris, le capteur doit être capable de séparer la lumière en différents domaines spectraux ; c'est-à-dire de transformer le spectre visible de $\mathbb{R}^\infty \rightarrow \mathbb{R}^N$, avec N le nombre de couleurs, généralement 3 pour *RGB* (Red, Green, Blue). L'image est alors représentée comme un tableau à trois dimensions de taille *Largeur* \times *Hauteur* $\times N$. Chaque couche de couleur intègre une tranche du spectre (domaine spectral). Plusieurs approches sont possibles :

Filtre colorimétrique L'approche la plus courante et la moins coûteuse est de disposer les pixels selon un motif, tel que le filtre de "Bayer", comme illustré ci-dessous (figure 2.8). Les filtres de Bayer disposent deux fois plus de pixels verts que de pixels rouges ou bleus. Ce qui signifie que certaines couleurs seront plus bruitées que d'autres, ici rouge et bleu. Généralement, le bleu est naturellement moins présent en extérieur et donc plus difficile à estimer. Mais il existe différents types de filtres et certains se concentrent davantage sur le bleu tel que les filtres *RGBE* [Miao et al., 2004]. Dans ce cas, le canal bleu n'est pas le plus bruité. Le désavantage de ces filtres est qu'ils diminuent la résolution physique. On a alors recours à un algorithme d'interpolation pour estimer les couleurs manquantes (Demosaicing) [Rathi and Goyal, 2021]. Certaines méthodes de demosaicing peuvent partiellement corriger le bruit associé à une couleur [Tan et al., 2014] en se basant sur la plus présente, par exemple sur le vert pour les filtres de Bayer.

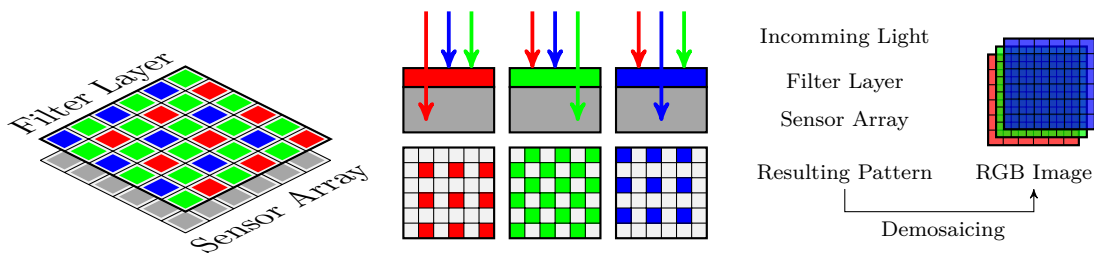


Figure 2.8 – Filtre de Bayer

Diviser la lumière incidente Parmi les techniques de dispersion (réseaux de diffraction, réflexion, ...) de la lumière dans un capteur optique, plusieurs prismes peuvent être utilisés pour diviser la lumière incidente vers des capteurs physiquement séparés qui absorbent les longueurs d'onde différentes (figure 2.9). L'avantage de cette technique est d'obtenir des images couleurs de haute qualité avec une résolution physique complète. Les capteurs peuvent être exposés pendant des temps différents pour gérer le bruit efficacement. La difficulté d'alignement précis des prismes [Galo et al., 2006] rend cette technologie coûteuse.

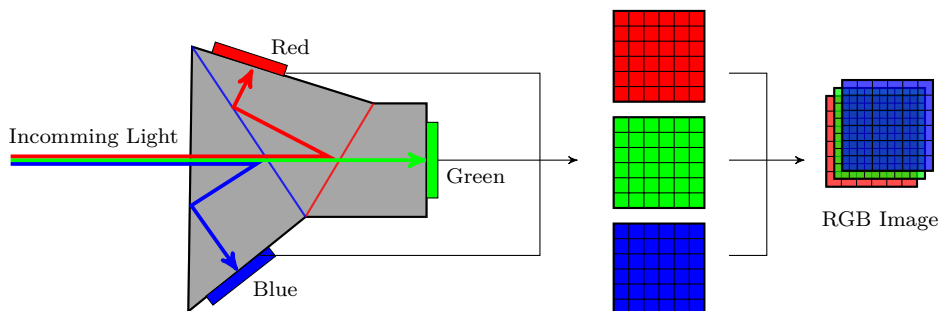


Figure 2.9 – Triple prismes pour diviser la lumière incidente

2.1.2 Des capteurs multispectraux

Nous avons vu précédemment que la première approche a le désavantage de réduire la résolution physique, tandis que la seconde est coûteuse et il devient impossible pour les deux méthodes d'obtenir efficacement plus de couleurs (infrarouge, ultraviolet, ...). Pour pallier ces désavantages, d'autres dispositifs (multispectraux) sont apparus : les systèmes d'acquisition d'images multispectrales qui permettent d'extraire un certain nombre de propriétés sur les surfaces observées [Qiu et al., 2018]. Leurs utilisations se sont diversifiées pour caractériser l'état sanitaire des cultures, la texture des sols, le stress hydrique, ou encore la présence de végétation.

Filtre rotatif Ce type de dispositif, développé au début des années 2000 [Vioix, 2004], associe un système optique à une roue portant les différents filtres colorimétriques. Plusieurs acquisitions sont effectuées pour obtenir les différentes couleurs. L'avantage de ce dispositif est d'être très peu coûteux tout en fournissant des images de grandes résolutions. Mais faire tourner le filtre engendre des problèmes de recalage des images mono-modes d'une même scène. Ce dispositif ne peut donc pas être utilisé efficacement en conditions extérieures en particulier en présence de mouvement.

(source : [Vioix, 2004])

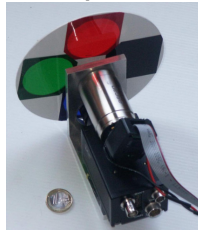


Figure 2.10 – Filtre rotatif à 4 bandes spectrales : bleu, vert, rouge et infrarouge

Capteurs multiples Une autre solution est d'utiliser plusieurs caméras disposées en damier. Chaque caméra dispose d'un filtre permettant de percevoir une section différente du spectre. C'est le cas du dispositif développé par HYPHEN pour l'acquisition d'images depuis un drone aérien, et qui a été utilisé durant cette thèse (Cf. détails en chapitre 3). Cette solution a l'avantage d'avoir des temps d'acquisition très courts, contrairement au dispositif proposé par [Vioix, 2004] et peut également être utilisé en mouvement. Cependant, du fait de la différence de position des capteurs sur le système d'imagerie, leurs champs de vision ne sont pas identiques. Ceci implique l'utilisation d'algorithmes complexes pour superposer les images mono-modes. À cet effet, un algorithme spécifique est proposé dans ce manuscrit (section 3.5).

Vers l'hyperspectral D'autres solutions permettent d'obtenir des images dites hyperspectrales, c'est-à-dire un découpage plus fin du spectre électromagnétique. Par exemple $\mathbb{R}^\infty \rightarrow \mathbb{R}^{100}$, on parle alors d'hyper-cube. La solution proposée par le consortium PEAD du challenge RoSE, repose sur un spectromètre imageur par tomographie (CTIS [Descour and Dereniak, 1995]) développée par CarbonBee. L'imagerie hyperspectrale présente de nombreux avantages et a de nombreux domaines d'application (agriculture, médecine, minéralogie, surveillance, astronomie, chimie, etc). Son prix est encore élevé ce qui limite fortement son utilisation. Ce dispositif permet de définir les longueurs d'onde les plus discriminantes pour un sujet donné, de manière à orienter le choix de capteurs multispectraux, plus abordables. Nous verrons plus tard, que le dispositif multispectral utilisé dans cette thèse repose sur les travaux de [Louargant, 2016].

2.2 Prétraitement

Le système optique engendre par nature des distorsions de l'image. Il existe plusieurs sources d'erreurs qui peuvent amener à des déformations géométriques et/ou colorimétriques. Au niveau colorimétrique, l'objectif du prétraitement des images est d'améliorer le contraste et d'éliminer le bruit afin de mettre en valeur les objets d'intérêt dans une image donnée. Les pré-traitements permettent aussi de stabiliser certaines méthodes de segmentation ainsi que la qualité de l'extraction des caractéristiques. Les sous-sections suivantes expliquent et présentent les corrections possibles.

2.2.1 Correction géométrique

Les premières déformations sont issues de la lentille de l'appareil et de l'alignement du capteur à cette dernière. Ces déformations peuvent se modéliser mathématiquement. Trois sources d'erreurs sont à prendre en considération :

- **Distorsion radiale** : La distorsion radiale se produit lorsque les rayons lumineux se courbent plus près des bords d'une lentille qu'ils ne le font en son centre optique. Plus la lentille est petite, plus la distorsion est importante. La distorsion peut être positive (barrel) ou négative (pincushion)
- **Distorsion tangentielle** : La distorsion tangentielle corrige les inclinaisons du plan de l'image (capteur) après une distorsion radiale.
- **Distorsion prismatique** : La distorsion à prisme mince est due à une légère inclinaison de la lentille ou du réseau de capteurs d'images et provoque des distorsions radiales et tangentielles supplémentaires.

En résumé, les différentes distorsions se cumulent et constituent le modèle de distorsion (figure 2.11). Pour rectifier les déformations d'une image, il faut inverser le modèle, mais il n'existe pas d'expression de forme close pour estimer ce mappage inverse. En pratique, l'évaluation est effectuée par un processus de table de recherche. Dans l'ensemble, ce modèle de caméra avec distorsion de l'objectif comprend jusqu'à onze paramètres intrinsèques en plus des paramètres extrinsèques, avec six degrés de liberté.

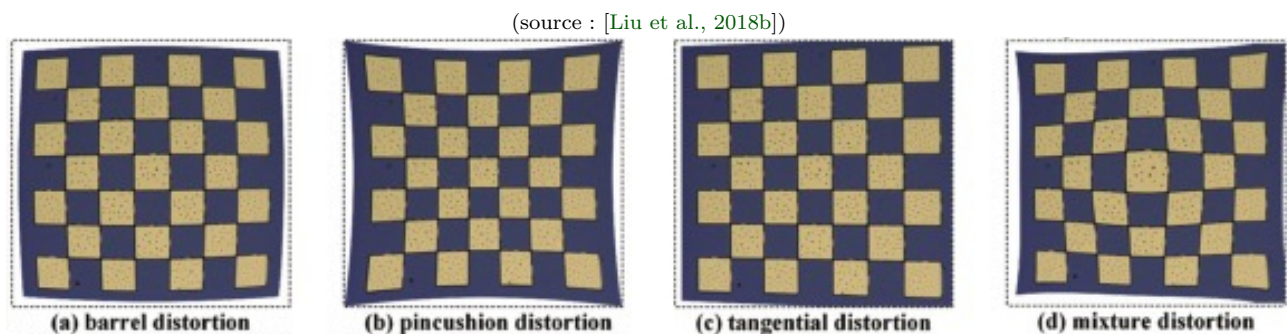


Figure 2.11 – Modèle de distorsion

Cette correction est indispensable si l'on cherche à évaluer la distance d'un objet à l'objectif ou mesurer la taille d'objet [Vayssade, 2018]. En effet, la distorsion géométrique de l'image induit une distorsion des surfaces et des distances. En l'absence de ce type d'analyse, la correction géométrique n'est pas indispensable.

2.2.2 Rehaussement d'image

Le ré-haussement d'images est essentiellement un processus qui permet de faciliter l'interprétation d'une image. Ainsi, rehausser l'image, consiste entre autres à standardiser l'information, comme corriger la luminosité, l'intensité des couleurs ou à supprimer des effets de l'environnement comme l'ombre ou le bruit.

2.2.2.1 Correction colorimétrique et radiométrique

Selon les paramètres de la caméra (triangle d'exposition), de l'illuminant et les filtres spectraux employés, la quantité de photons reçue, et donc l'estimation des couleurs de la scène peut être biaisée. Nous avons vu par exemple dans la section précédente que les filtres de Bayer sous-évalue le bleu [Lukac and Plataniotis, 2005]. L'illuminant joue aussi un rôle prédominant puisqu'il change la distribution des photons émis [Brainard and Wandell, 1992]. L'objectif est alors de déterminer la couleur des objets dans des conditions d'éclairage différentes en supprimant l'effet de décalage des spectres. Les méthodes employées reposent souvent sur l'utilisation d'un dispositif d'étalonnage, on prend alors en photo le dispositif dont les couleurs sont connues a priori. Une équation permet ensuite de corriger les couleurs dans ces mêmes conditions d'acquisition afin d'obtenir des images standardisées. D'autres approches sont toutefois possibles et ne nécessitent pas de dispositif d'étalonnage [Finlayson et al., 2015, Afifi, 2018], ces méthodes ne sont pas encore utilisées en agriculture de précision.

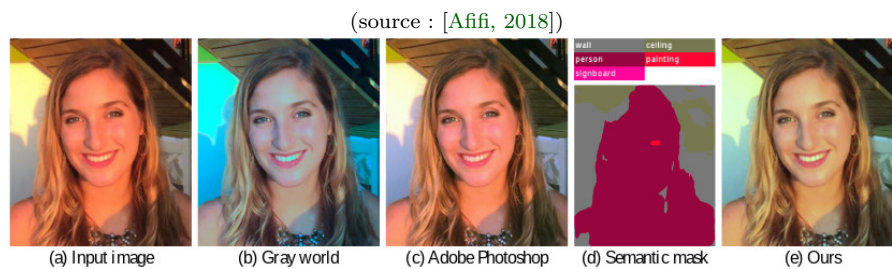


Figure 2.12 – Correction des couleurs par apprentissage profond proposé par Afifi

Calibration radiométrique La correction radiométrique suit les mêmes principes, mais permet d'obtenir des mesures de réflectance [Honkavaara and Khoramshahi, 2018] (figure 2.13). La réflectance se définit comme la proportion de lumière réfléchiée par la surface d'un matériau et s'exprime généralement en pourcentage. La réflectance spectrale est estimée afin de rendre la discrimination robuste aux variations d'illumination. Les indices de végétation reposent sur ces valeurs calibrées. L'étude de [Amziane et al., 2021] permet l'estimation de la réflectance sans dispositif d'étalonnage en proxidtection.

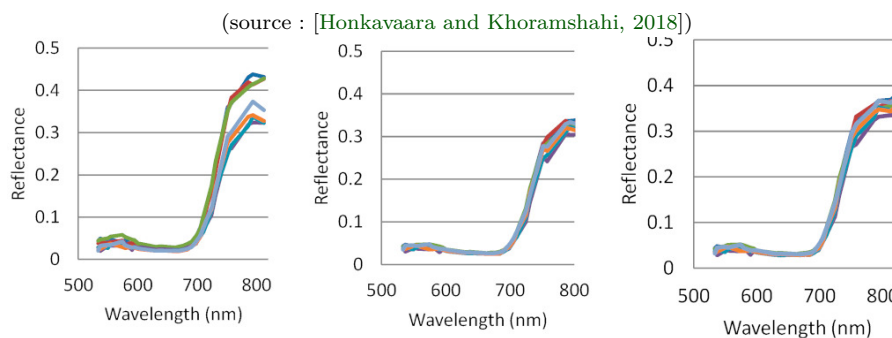


Figure 2.13 – Différentes corrections radiométriques proposé par Honkavaara

2.2.2.2 Correction du bruit

De par l’environnement d’acquisition et la qualité du matériel utilisé, il peut être intéressant d’estimer et de corriger le bruit éventuel des capteurs. Le “bruit d’image” est l’équivalent numérique du grain des caméras analogiques. Pour les images numériques, ce bruit apparaît sous la forme de points aléatoires et peut dégrader considérablement la qualité de l’image. Le bruit augmente avec le réglage de la sensibilité de l’appareil photographique, le temps d’exposition, la température, et varie également d’un modèle d’appareil photographique à l’autre.



Figure 2.14 – Comparaison des méthodes de débruitage

Il n’existe pas de méthode propre à l’agriculture de précision, cependant, il existe de nombreuses méthodes générales [Buades et al., 2005]. Les plus connues pour ce type de traitement sont les filtres bilatéraux [Chaudhury and Dabhade, 2016], les “Non local means”, les “Wavelet-based thresholding denoising” [Hao et al., 2014], “Random Field Denoising” [Barbu, 2009] ou encore les filtres “Total Variation”. Pour autant, toutes ces approches sont gourmandes en temps de calcul et ne permettent pas une utilisation en temps réel. De plus supprimer le bruit d’une image peut également lisser les textures et ainsi induire à une perte d’information utile à la discrimination. Ici encore l’apprentissage profond surpasse les méthodes usuelles [Tian et al., 2020, Zamir et al., 2020].

2.2.2.3 Correction d’ombrage

La présence d’ombres est naturelle dans une scène. Une ombre portée modifie la couleur sur une partie de l’image et ajoute des ruptures de gradient. Ainsi, les méthodes de segmentation sont très impactées par la présence d’ombres. En agriculture, l’exemple de la segmentation sol/végétation illustre facilement cette problématique. En effet, cette segmentation s’appuie généralement sur un indice de végétation couplé à un seuillage simple. Or, le seuil adéquat est différent à l’intérieur et à l’extérieur des zones ombragée, il en résulte une segmentation incorrecte dans l’un des deux cas. L’algorithme “Contrast Limited Adaptive Histogram Equalization” propose de traiter l’image en $N \times N$ sous image en appliquant une correction de couleurs via une égalisation de l’histogramme [Zuiderveld, 1994]. Cependant, une nouvelle approche basée sur l’apprentissage profond permet de détecter et supprimer l’ombre beaucoup plus efficacement [Zhang et al., 2020].



Figure 2.15 – Suppression de l’ombre par apprentissage profond proposé par Zhang

2.2.3 Indices de télédétection

L'idée de base de l'imagerie hyperspectrale repose sur le fait que, pour tout matériau, la quantité de rayonnement réfléchi, absorbé ou émis dépend de la longueur d'onde. Donc la réponse spectrale du matériau varie selon un mode unique lié à sa composition moléculaire [Mercan and Alam, 2011]. Cela signifie que, les signatures spectrales des matériaux, sont uniques et ces signatures peuvent être utilisées efficacement pour distinguer et identifier des surfaces (figure 2.16).

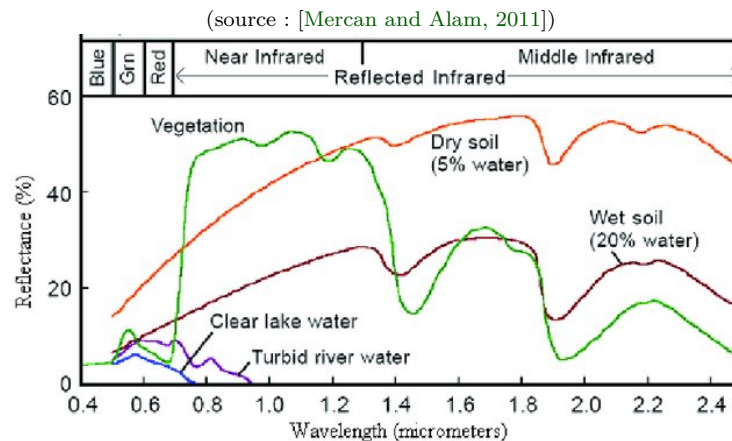


Figure 2.16 – Signatures spectrales de certains matériaux

2.2.3.1 Signature spectrale des plantes

Les plantes se caractérisent, d'une part par leurs compositions moléculaires et d'autre part par leurs structures cellulaires. La figure 2.17 montre la structure cellulaire d'une feuille de dicotylédone (principal groupe des adventices). On y trouve deux épidermes (dessus/dessous) et entre ceux-ci se trouve la couche mésophylle [Ben Abdallah et al., 2016]. La couche supérieure du mésophylle est composée de cellules contenant, entre autres, la chlorophylle (Chl a et Chl b) formées lors du processus de photosynthèse. Cette couche supérieure absorbe en grande partie le spectre visible, elle contient également d'autres pigments comme les carotènes, mais n'interfère pas avec d'autres longueurs d'onde. La couche inférieure, spongieuse, du mésophylle est constituée de cellules irrégulières et d'espaces vides qui eux réfléchissent de manière importante toutes les longueurs d'onde (diffère chez les monocotylédones).

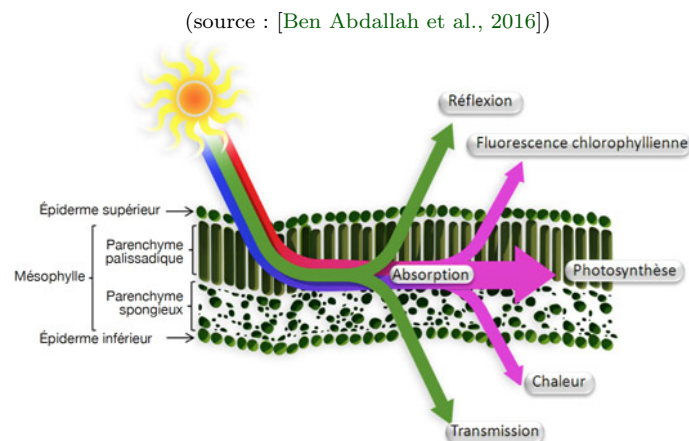


Figure 2.17 – Structure cellulaire d'une feuille et interaction avec le spectre électromagnétique

En conséquence, les plantes ont une faible réflectance pour les longueurs d’onde entre 400 nm (bleu) et 700 nm (rouge) avec un maximum autour de 550 nm (vert). À contrario, les plantes ont une forte réflectance dans le proche infrarouge (entre 850 nm et 1300 nm). La région du red-edge, qui représente la zone de transition entre le rouge et le proche infrarouge (approximativement entre 675 nm et 750 nm), se caractérise par une brusque augmentation de la réflectance. La concentration en chlorophylle déplace la réflectance dans le red-edge. Une diminution de la concentration engendre une pente plus douce dans cette région. De la même façon, une diminution de la concentration d’eau change la couche inférieure, spongieuse, du mésophylle et donc la réflectance dans l’infrarouge. Finalement dans le domaine de l’infrarouge court (short wave infrared, SWIR), le spectre de réflectance présente trois pics d’absorption en 1450 nm, 1950 nm et 2500 nm due à la présence d’eau dans les feuilles. L’ensemble de ces éléments sont visibles sur la courbe verte (végétation) de la figure 2.16.

2.2.3.2 Indices de végétation

Mesurer et interpréter le rayonnement électromagnétique permettent d’établir des indices qui caractérisent les surfaces observées. Ces indices évoluent en fonction de la nature de l’objet (sol, eau, végétation) et sont donc particulièrement utiles pour la segmentation. Il existe une base de données indexdatabase.de qui recense ces différents indices de télédétection. Pour chaque forme d’indice, leurs variantes (en fonction du matériel d’acquisition) sont également mises à disposition. Certains indices permettent par exemple d’estimer la composition des sols [Dhawale et al., 2021].

Observer le spectre de réflectance des plantes a permis d’établir des indicateurs physiologiques, phénologiques et physiques. Le NDVI pour “Normalized Difference Vegetation Index” est probablement l’indice de végétation le plus populaire et repose sur les connaissances empiriques précédemment exposées. Le NDVI se calcule à partir de la réflectance obtenue pour deux bandes spectrales : $NDVI = (N - R) \div (N + R)$ avec N le proche infrarouge (≈ 850 nm) et R le rouge (≈ 650 nm). Les valeurs théoriques du NDVI sont comprises entre -1 et 1 , celles-ci sont bornées entre 0 et 1 dans le cas de la végétation. Pour la végétation, le NDVI renseigne sur l’état sanitaire d’un végétal. Une valeur de NDVI faible induit une faible activité photosynthétique d’une plante et inversement. Il existe une grande variété d’indices de végétation, dont tous reposent sur des combinaisons de réflectance pour différents domaines spectraux. Les articles [Wang et al., 2018a, Xue et al., 2017, Zhang et al., 2017] mettent en exergue les indices ayant été testés pour détecter la présence de végétaux sur des images multispectrales. Ces indices reposent principalement sur une différence de réflectance entre le visible et le proche infrarouge. Ces indices n’utilisent pas plus de deux ou trois bandes spectrales ce qui diminue leurs potentiels de discrimination.

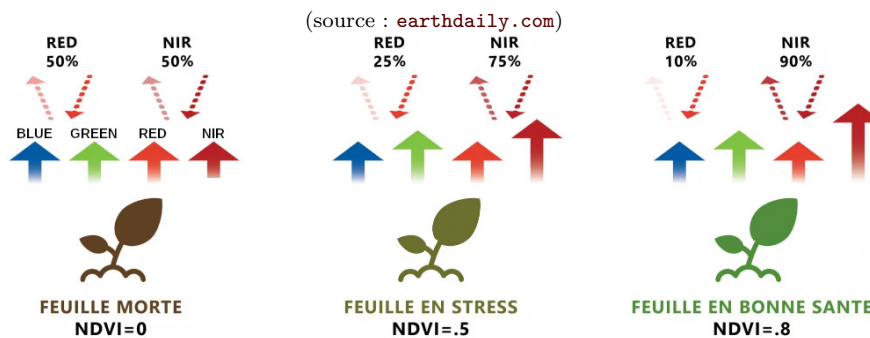


Figure 2.18 – Une faible activité photo-synthétique induit une faible valeur de NDVI

2.2.3.3 Limite des indices de végétation

Bien que ces indices soient largement utilisés, ils ne permettent pas une utilisation optimale en conditions réelles. Leurs natures empiriques, l'influence de nombreux facteurs environnementaux, ainsi que le système d'acquisition sont des éléments qui jouent un rôle dans les performances de tels indices :

- **Bandes spectrales** Pour établir des indices de télédétection, on a recours à l'imagerie hyperspectrale. Selon la scène observée, le choix des “meilleures” bandes spectrales se fait à partir d'un échantillon d'acquisition. Elles seront ensuite sélectionnées pour être utilisées dans des capteurs multi-spectraux. Finalement, les indices sont eux même basés sur une ou deux bandes spectrales et diminuent encore le potentiel de discrimination.
- **Acquisition** Les conditions d'acquisition (jour, nuit, soleil, nuage, ombres, ...) modifient la lumière incidente et donc les performances de ces indices [Zhang et al., 2015]. Il y a également le mix spectral (mélange de signatures spectrales issues de différents matériaux) qui constitue une source d'erreur lorsque la résolution spatiale est insuffisante [Louargant et al., 2017], les petites adventices peuvent ne pas être détectées.
- **Calibration** Les indices de végétation reposent sur la réflectance. Une calibration doit être mise en œuvre et nécessite un dispositif d'étalonnage (section 2.2.2 partie A). Cette calibration peut introduire des erreurs, puisque les résultats ne sont pas stables dans le temps (changement de luminosité au cours de la journée, passage d'un nuage, ...). Et difficilement utilisable en proxi-détection et en temps réel.
- **Plante** Le groupe d'appartenance de la plante est une source d'erreur : la classe de la plante (monocotylédone, dicotylédone, hétérotrophes¹, photoautotrophes²) [Wang et al., 2019b], leurs états sanitaires [Gay and Wery, 2003, Al Naser, 2018, Fournier, 2011] (chaleur, eau, azote, nitrate, phosphate, ..., maladies, ...) [Piedallu et al., 2019], leurs stades de développement (croissance/sénescence) [Palanisamy et al., 2019], sont autant de facteurs qui peuvent modifier la valeur d'un indice de végétation comme le NDVI, si bien qu'une plante peut ne pas être détectée (faux négatif).
- **Sol** Des corps étrangers sur le sol peuvent être détectés (faux positifs). Par exemple, certaines moisissures et bactéries changent la couleur du sol, de même que la présence de métaux et la teneur en eau.

Ces éléments démontrent rapidement que la détection de la végétation n'est pas un problème trivial. Même si ces indices ont été largement utilisés, des études récentes soulignent le faible niveau de discrimination pour l'estimation de la surface foliaire [Zhang et al., 2019b, Niu et al., 2021] et donc leurs performances relatives pour une segmentation sol/végétation.

1. Les végétaux hétérotrophes prélèvent leurs nutriments par symbiose, parasitisme ou prédation

2. Caractérise les végétaux chlorophylliens, faisant intervenir la photosynthèse

2.3 Segmentation

L'extraction d'informations à partir d'images s'effectue par le processus de segmentation. L'objectif de la procédure est d'extraire les composantes de l'image qui présentent un intérêt. La construction des régions d'intérêt est basée sur des caractéristiques de l'image telles que la couleur, la luminosité spectrale, la détection des contours, la similarité des pixels voisins ou la combinaison de ces propriétés, qui sont normalement intégrées à l'aide d'un processus d'apprentissage.

2.3.1 Segmentation sémantique

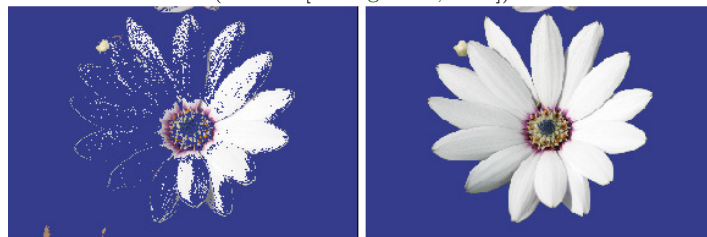
La segmentation sémantique consiste à classer les pixels selon leur classe d'appartenance. L'exemple le plus courant en agriculture de précision est la discrimination de la végétation (première classe) du reste (i.e. le sol, second classe). D'avantage de classes peuvent être définies.

Thresholding La méthode la plus simple consiste à déterminer un seuil entre les classes, à partir de l'indice de végétation pour l'exemple de la segmentation sol/végétation. Le seuil peut être fixé manuellement, ou par apprentissage. Les méthodes de seuillage peuvent aussi se reposer sur des algorithmes de classification. Cette approche est peu robuste aux changements d'illumination.

Otsu C'est une version améliorée du seuillage, consistant en une recherche automatique et exhaustive du seuil de segmentation. L'algorithme recherche le seuil qui minimise la variance intra-classe, définie comme une somme pondérée des variances de deux classes. La méthode montre de bonne performance si l'histogramme a une distribution bi-modale. Ceci est le cas pour un indice de végétation lorsque le sol est suffisamment visible sur l'image. À contrario, lorsque la végétation recouvre majoritairement le sol (ou lorsqu'elle est peu présente), la segmentation échoue. Cette méthode, non-supervisée, à l'avantage de bien gérer les gradients de luminosité contrairement à un seuillage simple. Cette approche est largement utilisée en agriculture [Sa et al., 2017], mais souffre de précision sur les cas limites, tels que les feuilles fines.

Grabcut L'algorithme apprend un modèle de mixture gaussienne pour classer l'information à partir de la zone d'intérêt et d'un masque. Ensuite, un graphe est construit à partir de la distribution des pixels entre foreground / background. Chaque pixel est connecté au nœud correspondant avec un poids d'appartenance en fonction de la probabilité de distribution de l'ensemble des pixels. Finalement, un algorithme de mincut est utilisé pour segmenter le graphe, donc l'image. Une application sur des données spectrales existe [Estrada et al., 2004].

(source : [Göring et al., 2012])



(a) initial segmentation

(b) result after GrabCut

Figure 2.19 – Segmentation Grabcut

Chaîne de Markov [Derrode and Pieczynski, 2003] L'image est transformée en signal 1D par exemple avec un algorithme de « Space-filling curve » comme Hilbert-Peano, Z-order, Sierpiński ou Moore, le graphe linéaire ainsi obtenu (chaîne) est stationnaire, ces probabilités sont déduites des probabilités conjointes qui sont indépendantes. La classe du premier pixel est tirée aléatoirement à partir de la distribution a posteriori. Par la suite, pour chaque nouveau pixel, la probabilité de transition est calculée, la classe du pixel précédent est fixée, et la classe du pixel suivant est obtenue par un prélèvement aléatoire selon cette distribution. La distribution a posteriori peut être estimée et initialisée grâce à l'histogramme de l'image et en utilisant l'algorithme des k-means (moyenne, déviation, écart-type).

Champ de Markov [Rechid et al., 2011] Le principe est similaire aux chaînes de Markov, le graphe de représentation est alors 2D et le nombre de transitions est plus important, ce qui augmente les temps de calcul. Une étude est présentée dans un contexte agronomique [Yue et al., 2016]. En revanche, l'approche est largement utilisée en segmentation d'images médicales. L'étude de [Salzenstein and Collet, 2006] propose une comparaison entre les chaînes de Markov et les champs de Markov appliqués aux données multispectrales, les résultats montrent une meilleure homogénéité des régions segmentées par les champs de Markov, bien que statistiquement, les chaînes de Markov semblent donner de meilleurs résultats. Une application avancée permet de combiner les avantages de l'apprentissage profond et des champs de Markov cachés [Liu et al., 2016].

Apprentissage profond Les réseaux de neurones convolutifs basés sur une structure d'encodeur-décodeur sont très populaires et ont démontrés leurs performances dans la discrimination des adventices [Brilhador et al., 2019, You et al., 2020, Khan et al., 2020]. L'amont du réseau utilise une série de convolutions et de sous-échantillonnages (pooling), cet ensemble est généralement appelé encodeur et génère des informations sur l'objet, telles que sa forme et sa taille à une résolution inférieure. La sortie est ensuite sur-échantillonnée (upsampling) par interpolation bilinéaire ou par une série de convolutions transposées en aval du réseau, ensemble appelé décodeur. Il cherche à distribuer les informations basses résolutions de l'encodeur vers les informations de haute résolution pour obtenir une carte de segmentation dans la résolution initiale. Ces éléments sont présentés dans la figure 2.20.

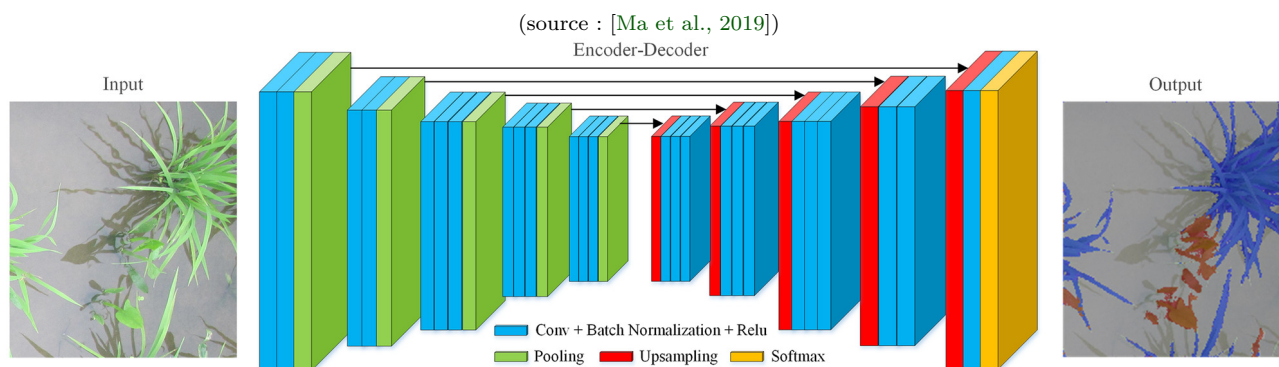


Figure 2.20 – Architecture encodeur-décodeur d'un réseau convolutif

2.3.2 Segmentation en région

Au lieu de ne travailler qu’avec des pixels (segmentation sémantique), on découpe l’image en région. L’utilisation de régions présente deux avantages majeurs (i) calculer des caractéristiques sur des régions plus significatives et (ii) réduire les entités en entrée pour les étapes suivantes.

Découpage régulier L’image est simplement découpée en sous-images de tailles fixes (avec ou sans recouvrement figure 2.21). La taille a un impact important : trop petite, elle manquera d’information discriminante et trop grande, elle pourra contenir des mélanges des classes (culture, adventice, sol) rendant la classification difficile. La discrimination des adventices se résume alors à un problème de classification de texture. Des approches multi-échelles sont aussi possibles [Hu et al., 2020] mais ne permettent pas de détecter les plus petites adventices.

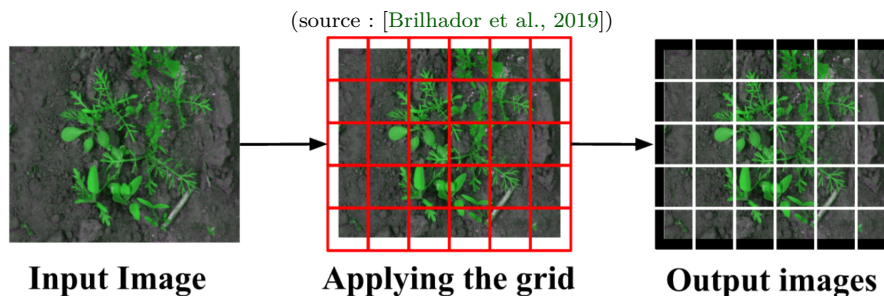


Figure 2.21 – Découpage régulier sans recouvrement

Composantes connexes On peut récupérer les zones distinctes dans l’image à l’aide d’un algorithme, après une première segmentation sémantique (par exemple sol/végétation). Dans le cas de la discrimination des adventices, cette approche ne permet pas de distinguer les feuilles, des plantes, des groupes de plantes. Une composante connexe peut donc être composée de cultures et d’adventices en proportions variables, en fonction du taux de couverture végétale. Extraire des propriétés sur ces zones et essayer de les discriminer est donc peu fiables [Louargant et al., 2018].

Template matching C’est une méthode permettant de rechercher et de trouver l’emplacement d’une image modèle dans une image plus grande. Il suffit de faire glisser l’image modèle sur l’image d’entrée et de comparer le modèle et avec la portion de l’image d’entrée. Cette approche est peu robuste pour de la détection de plante ou de feuille, car ces dernières ont une grande variabilité. De plus, c’est une méthode lourde en temps de calcul bien qu’elle ait été utilisée pour la détection des bananiers [Zhou et al., 2014].

Transformation de Hough Il existe deux grandes variantes de la transformée de Hough qui permettent de détecter des formes fixes dans l’image, par exemple des lignes ou des ellipses. Ainsi cette transformée a été utilisée pour détecter les rangs de culture. Les composantes connexes entre les lignes de culture sont principalement des adventices [Louargant et al., 2018]. La détection des ellipses par transformée de Hough présente aussi des avantages : les formes des feuilles des dicotylédones étant souvent proches d’une ellipse cette transformée permet de les détecter [Kumar and Domnic, 2018, Zhonghua et al., 2020]. Des techniques similaires peuvent également détecter les “blobs” (Laplacian of Gaussian, Difference of Gaussian, Determinant of Hessian, Maximally stable extremal region, ...).

Super-pixel Un super-pixel est un groupe de pixels connectés avec des couleurs ou des niveaux de gris similaires. La segmentation des super-pixels consiste donc, à diviser une image en centaines de super-pixels ne se chevauchant pas. Cette approche a été testée dans le domaine de l'agriculture de précision avec une approche multi-échelle [Bah et al., 2018b] proposant des résultats de classification intéressants $\sim 90\%$. Cette approche, pertinente à haute altitude présente des défauts majeurs en proxi-détection avec un paramètre difficile à déterminer : la taille des super-pixels. Les adventices et les cultures présentent des différences de taille notables (stade de développement) et de texture significatives. Ce qui donne des résultats instables en proxi-détection [Li et al., 2021].

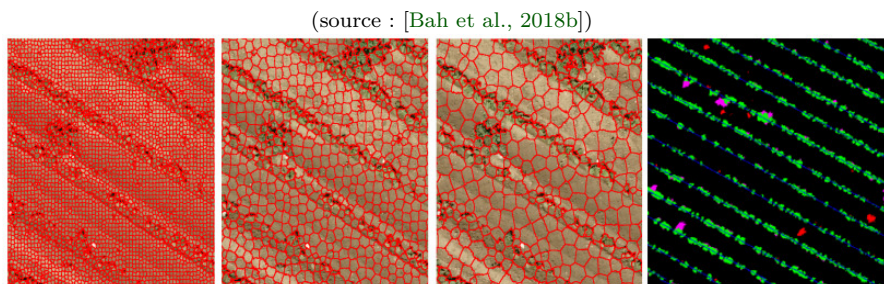
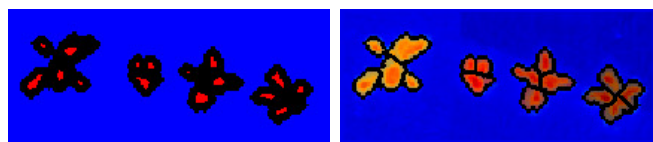


Figure 2.22 – Analyse multi-échelle des super-pixels et classification

Watershed L'algorithme calcule les distances entre les bords des régions segmentées et les différents centres initialisés. Ensuite, l'algorithme distingue autant de sous-régions qu'il y a de points culminants, et alimente chacune d'entre elles par l'incrémention des pixels voisins selon une propagation circulaire de proche en proche. Les régions sont isolées lorsqu'un même ensemble de pixels est candidat pour intégrer deux régions différentes.



les zones d'initialisation à gauche et à droite, en couleur les régions étendues

Figure 2.23 – Exemple de watershed sur des images drones

Ce type d'approche est donc très utile pour agrandir la segmentation, en particulier, on peut déterminer facilement le sol grâce à la méthode de seuillage de Otsu et les feuilles via des descripteurs de texture. L'algorithme Watershed propose alors une division interne des plantes (figure 2.23). Mais le choix des centres reste difficile à déterminer, rendant l'initialisation de ces derniers très dépendants de la résolution spatiale. Ainsi des approches complémentaires sont souvent nécessaires.

Chan-veese Le modèle proposé par Chan et Vese, est l'un des modèles les plus populaires pour la segmentation des images en région. Il combine le modèle Mumford-Shah simplifié et la méthode LevelSet pour résoudre un problème de minimisation de l'énergie. Le modèle est largement appliqué aux images qui doivent être segmentées en deux régions : cible et arrière-plan. Ce type de segmentation a été utilisé avec succès pour la segmentation de feuilles qui se chevauchent en champ proche [Wang et al., 2018b]. Bien que l'approche donne des résultats intéressants dans la littérature, le temps de calcul est important, ce qui est également à prendre en considération.

Active contours Le modèle de contour actif est une méthode d’ajustement d’une courbe aux lignes ou aux bords d’un objet dans une image. L’algorithme fonctionne en minimisant l’énergie définie en partie par l’image (texture, gradient) et en partie par la forme de la courbe (courbure). En ce sens, l’algorithme est similaire aux super-pixels. Les contours actifs sont utilisés pour identifier les plantes par leurs fleurs [Manh et al., 2001, Suta et al., 2012]. Une amélioration pour reconnaître les arbres par leurs feuilles a été proposée par [Cerutti et al., 2011], une contrainte est ajoutée aux contours actifs en utilisant un modèle de feuille polygonale. Cette amélioration suppose une connaissance à priori de la forme dans la scène qui doit être “complète”.

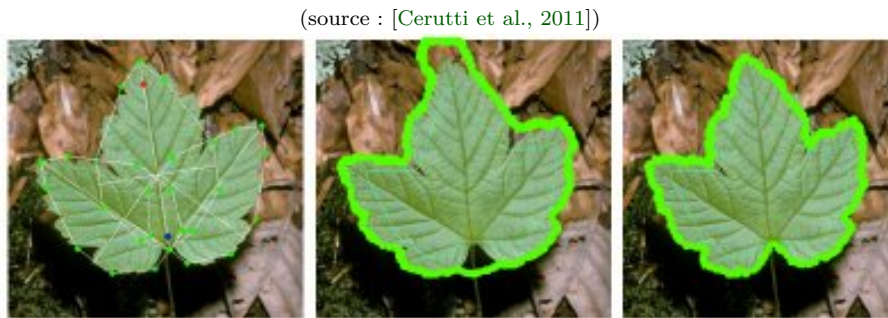


Figure 2.24 – Polygone initial, contour actif, et contour avec contrainte

Apprentissage profond Récemment, des améliorations considérables en matière de détection d’objets ont été obtenues grâce aux réseaux de neurones convolutifs. Le même principe que pour la segmentation sémantique est utilisé. Cependant la partie “décodeur” est remplacée par un réseau de prédiction des boîtes englobantes (bounding-box en anglais). Le réseau proposera par exemple une sortie de taille $(4 + 2) \times 1000$, soit 1000 objets à détecter avec leurs positions $[x, y, width, height]$ ainsi que la classe et le score de confiance en cette prédiction $[class, score]$. Finalement, seules les boîtes englobantes avec un score suffisant sont retenues.

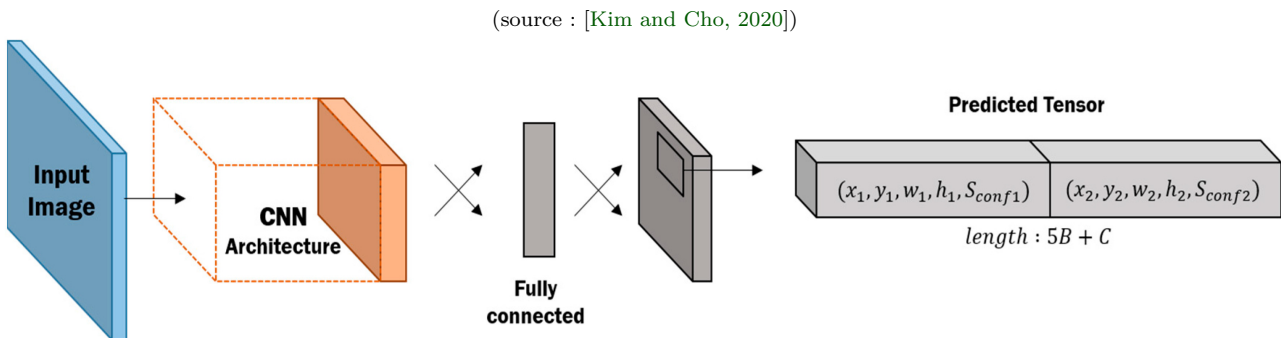


Figure 2.25 – Structure d’un réseau “You Only Look Once” (YOLO)

Ce type d’approche détecte généralement plusieurs fois le même objet avec un décalage, pour pallier cet effet un algorithme de non-max suppression (NMS) est alors utilisé pour combiner les boîtes qui se superposent. Ce dernier point est problématique, car plus la densité d’objets est forte, plus il y aura des fusions et donc des non-détections. L’utilisation des CNN et plus particulièrement l’architecture YOLO (You Only Look Once) a été utilisée pour la discrimination des adventices [Quan et al., 2021] mais n’est donc pas envisageable pour les stades de développement précoces et avancés. En agriculture de précision la revue [Zheng et al., 2019] définit un ensemble de jeux de données (cultures, adventices, fruits, ...) et leurs évaluations à travers différents modèles de détection, dont le modèle YOLO.

2.3.3 Segmentation des instances

La segmentation des instances consiste à la fois à détecter les objets et leurs masques de segmentation individuelle. Ici, seule l'utilisation des réseaux de neurones donne des résultats prometteurs. Il existe plusieurs variantes [Hafiz and Bhat, 2020] à cette démarche, dont :

Détection puis segmentation La première consiste à utiliser une segmentation des objets, puis pour chaque objet détecté une segmentation sémantique est utilisée pour prédire le masque de segmentation de chaque objet. Ce type de méthode nécessite une base d'apprentissage importante. La détection d'objet peut être difficile sur les petits objets, comme nous l'avons vu précédemment, repose sur l'algorithme NMS (non-max suppression) qui occasionne des mauvaises détections en scène dense. Puis le masque est prédit avec un second réseau, dont la taille de sortie est petite et fixe, l'ajustement du masque vers l'objet détecté se fait par étirement ce qui explique la qualité des masques générés (figure 2.26).

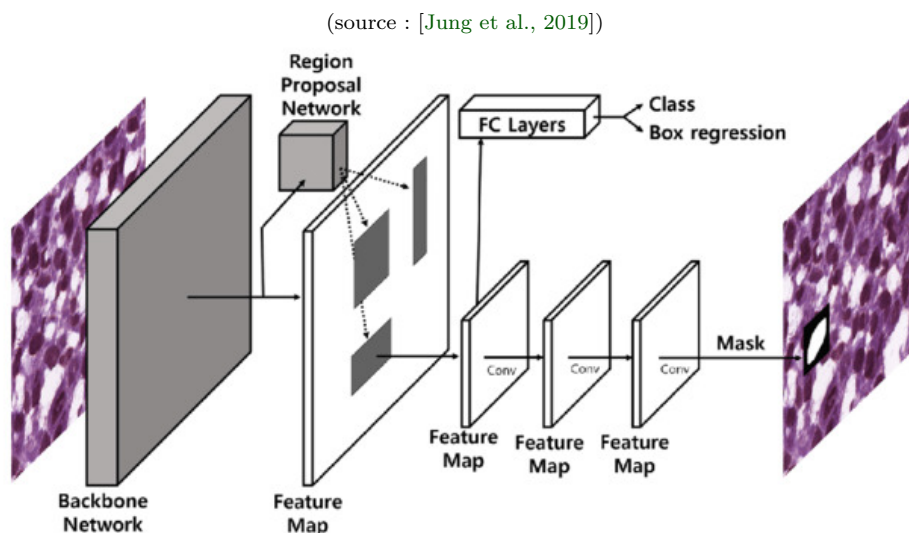
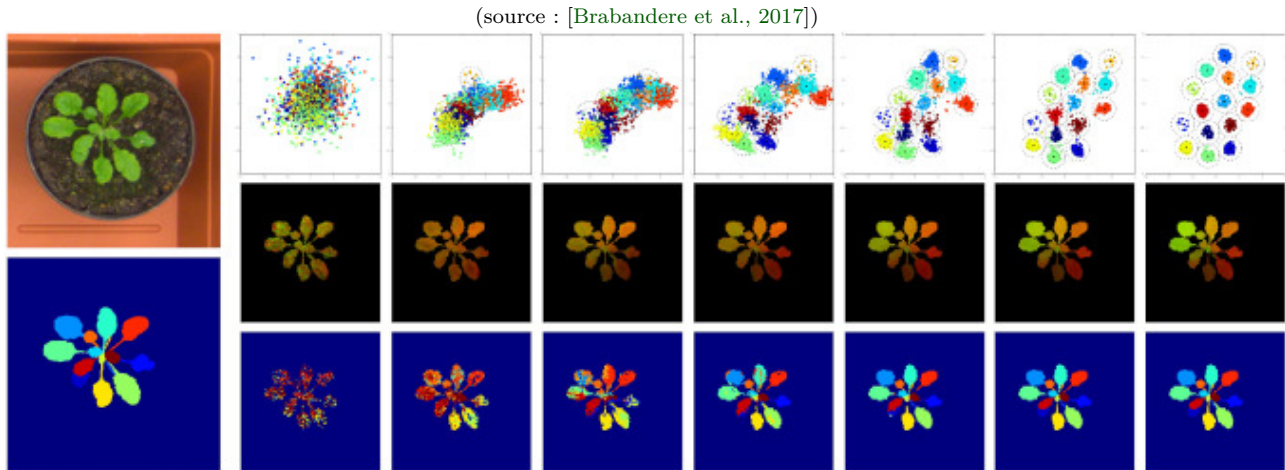


Figure 2.26 – Architecture d'un réseau de type Mask-RCNN

Cette approche est utilisée par le consortium WeedElec [Champ et al., 2020] dans le challenge RoSE, à travers une architecture mask-rcnn (figure 2.26), les “petites” adventices ($< 3.5 \text{ cm}^2$) sont détectées avec une précision moyenne de $\approx 2\%$, les adventices émergentes sont donc inexistantes. Les adventices “moyennes” comprises entre 3.5 et 9.5 cm^2 sont détectées à hauteur de $\approx 35\%$ tandis que les plus grandes sont mieux détectées $\approx 51\%$.

Instance embedding La technique consiste à apprendre une transformation de l’image vers un espace latent. Les pixels qui font partie de la même instance doivent être proches les uns des autres pour former un cluster, tandis que les pixels des autres instances doivent s’éloigner. Cet espace latent est en fait une projection des propriétés des pixels avec leurs instances associées spatialement. Cette approche repose essentiellement sur un réseau de segmentation sémantique, cherchant à minimiser la variance spatiale des pixels des instances (composantes connexes) tout en maximisant la distance des clusters au sein de l’espace latent (2.27). Cette approche a été testée pour la segmentation de feuilles [Brabandere et al., 2017, Wu et al., 2020]. Mais du fait de cet espace latent, la présence d’ombre ou de réflexion spéculaire peuvent être sources d’erreurs.



À gauche : Image initial et vérité terrain. La rangée du milieu montre la sortie brute du réseau. La rangée supérieure montre l’espace latent avec les couleurs de chaque instance. La dernière rangée montre le résultat de la mise en cluster des embeddings par seuillage autour de leur centre de cluster. Chaque colonne montre l’apprentissage de l’espace latent.

Figure 2.27 – Espace latent et instances associées aux clusters

Pixelwise instance La segmentation sémantique peut être utilisée pour détecter des instances d’une autre manière, on parle alors de “pixelwise instance segmentation”. L’idée repose essentiellement sur la détection des contours, donc la manipulation des classes de sortie (figure 2.28). Si cette approche n’est pas possible pour la détection des plantes, elle apporte néanmoins des possibilités pour la détection des feuilles [Morris, 2018] et des nucléus [Eschweiler et al., 2018]. La définition est simple et des fonctions de pertes avancées peuvent être mises en place afin d’en améliorer les performances.

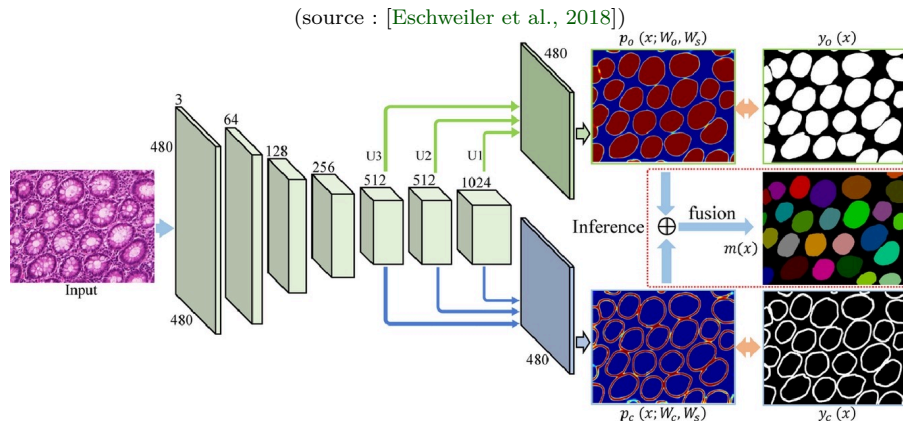


Figure 2.28 – Segmentation des instances par détection des contours des nucléus

2.4 Extraction de propriétés

L'objectif de l'extraction de propriétés est de caractériser les éléments distincts qui ont été détectés par la phase de segmentation, avec des propriétés discriminantes. Pour chaque individu, un ensemble de caractéristiques est extrait et placé dans un "vecteur de caractéristiques" qui synthétise l'objet. Associé à une méthode de classification, le vecteur est utilisé pour classer l'objet dans une catégorie (sol, culture, adventice). L'extraction repose sur des informations différentes, visibles sur la figure 2.29.

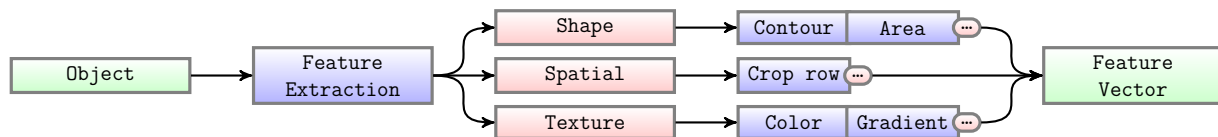


Figure 2.29 – Types de caractéristiques

On distingue les propriétés basées sur la géométrie (shape), celles qui se basent sur la texture et celles qui reposent sur une analyse spatiale des objets. Cette nomenclature n'est pas fixe, certaines études distinguent les propriétés de couleur, des propriétés spectrales au sein des propriétés de texture. Il existe bon nombre d'autres domaines où ces propriétés sont utilisées (granulométrie, médecine, ...). Quelques exemples de propriété de chaque type sont présentés ci-après.

2.4.1 Propriétés géométriques

Les propriétés géométriques se basent sur différentes représentations du contour ou de l'intérieur de la forme. On y retrouve par exemple, le périmètre, l'aire de la surface convexe ou encore l'aire de la forme. La figure 2.30 montre un petit ensemble de propriétés de forme appliqué à une feuille. Ces propriétés, proposées par [Gao et al., 2013] ont été utilisées pour la discrimination culture/adventices, les résultats de l'étude montre un taux de classification compris entre 87.5 % et 92.5 % sur les stades précoces et dans une parcelle peu infestée. Ces propriétés de forme sont invariantes par rotation, translation et changement de couleur. Mais elles sont néanmoins dépendantes de la résolution spatiale et fortement influencées par l'occultation et les performances de la segmentation.

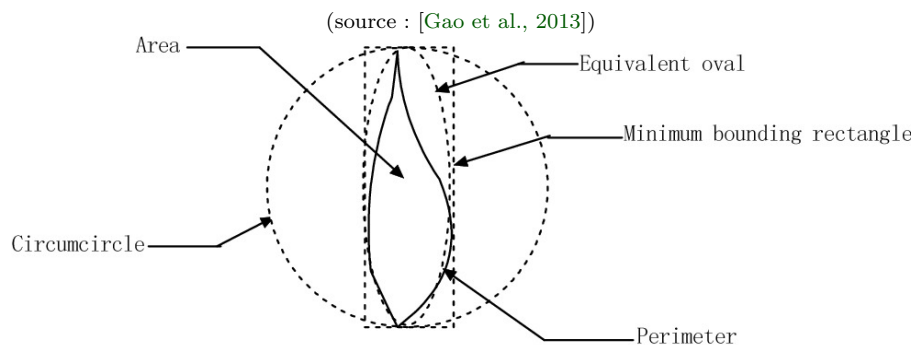


Figure 2.30 – Caractéristiques de la forme d'une feuille

2.4.2 Propriétés spatiales

Les propriétés spatiales, se basent sur la détection d'éléments spécifiques de la scène. Le rang de culture est l'élément le plus remarquable en culture sarclée, cette connaissance a priori peut être utilisée pour discriminer les adventices situées dans l'inter-rang. On calcule alors la distance entre l'individu et son rang le plus proche, cette valeur caractérise l'objet (Figure 2.31). Cette seule propriété est insuffisante pour caractériser et discriminer les adventices intrarang. Le rang peut être défini automatiquement en plaçant des points GPS/GNSS durant le semis [Griepentrog et al., 2005] ou par détection sur l'image [Lourgant, 2016]. Lorsque le taux d'infestation est trop important, la détection automatique des rangs de culture devient impossible.

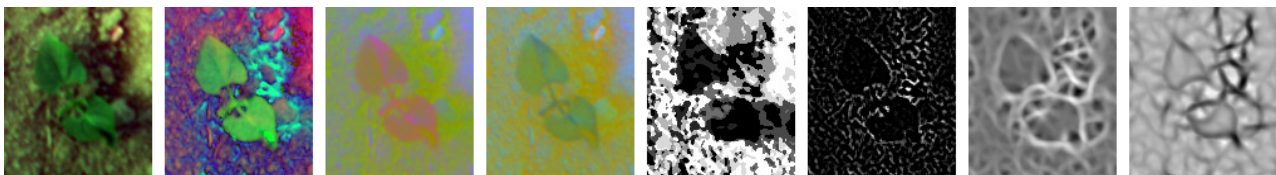
(source : [Griepentrog et al., 2005])



Figure 2.31 – Détection des rangs de culture et distance des adventices

2.4.3 Propriétés de texture

Finalement, les propriétés de texture sont plus vastes et incluent entre autres, des transformations de l'image (Fourier, Ondelettes, espaces colorimétriques, ...), des propriétés spectrales ou encore des histogrammes. L'article [Mekhalfa and Yacef, 2021], teste les propriétés de texture et de couleur sur des images RGB de soja. Les caractéristiques de couleur proposées, sont les moyennes et les écart-types des trois bandes de l'image des espaces colorimétriques RGB et HSV. En termes de propriétés de texture, c'est-à-dire une analyse des distributions spatiales des couleurs, l'article se repose sur les caractéristiques GLCM (Gray-Level Cooccurrence Matrix), d'Haralick et de LBP (Local Binary Pattern). D'après l'étude, les résultats reposant sur ces propriétés ont donné des taux de classification entre 89.43 % et 96.17 %. La figure 2.32 ci-après montre différentes transformations qui peuvent être appliquées. À partir de ces nouvelles données des propriétés statistiques sont extraites, comme l'histogramme, les moyennes, etc.



De gauche à droite, les espaces colorimétriques RGB, HSV, YUB, et LAB, puis les transformations LBP, Laplacien et les valeurs propre minimum et maximum de la matrice hessienne

Figure 2.32 – Exemple de transformation d'une image de haricot

2.4.4 Robustesse des propriétés

Les articles de revue de [Wu et al., 2021a] et de [Su, 2020] recensent les propriétés qui ont été utilisées pour la discrimination des adventices. Ces articles montrent également, que pour obtenir des propriétés discriminantes, ces dernières doivent si possible être invariantes ; c'est-à-dire qu'elles ne changent pas de valeur face à des transformations de l'entrée, par exemple, la translation, la rotation, le changement d'échelle ou la couleur. Chaque type de propriété présente donc des avantages et des inconvénients [Wu et al., 2021a]. Le tableau suivant 2.1 illustre les avantages et sources d'erreur de ces propriétés.

T=Translation ; R=Rotation ; S=Scale ; C=Color

Propriété	Invariance	Source d'erreur
Texture	T+S	Illumination ; Ombres ; Bruit ; Mélange spectral
Shape	T+R+C	Segmentation ; Occlusion ; Résolution
Spatial	R+S+T+C	Intra-rang ; Taux d'infestation

Table 2.1 – Invariance et sources d'erreurs pour ces trois types de propriétés

La similitude entre les adventices et les cultures rend difficile leur discrimination à partir de telles propriétés. C'est pourquoi, les chercheurs mélangent différentes sources d'information pour tirer les avantages de chacune d'elles, afin d'améliorer la précision et la robustesse du résultat. D'avantage de propriétés sont exposées dans le chapitre 6 qui est dédié à l'extraction, l'étalonnage et la sélection de propriétés.

2.4.5 Apprentissage profond

Dans l'approche par apprentissage profond, les réseaux de neurones convolutifs (CNN) sont populaires pour la reconnaissance et le traitement des images [Hasan et al., 2021]. Un ensemble de filtres de convolution est utilisé pour transformer l'image et en extraire des caractéristiques discriminantes (textures, bords, formes). Pendant l'apprentissage, le CNN optimise les valeurs des filtres de convolution. Cette solution qui intègre la classification, permet de s'affranchir de l'extraction "manuelle" des propriétés telles que précédemment évoquée (texture, spatial, forme). Cette approche est de plus en plus utilisée [Hasan et al., 2021] pour la discrimination des adventices. L'article de [Peteinatos et al., 2020] montre des taux de classification entre 97.5 % et 97.8 % pour l'identification des adventices dans une culture de maïs. Des architectures plus avancées [Reedha et al., 2021] montrent des taux de classification entre 99.6 % et 99.8 %.

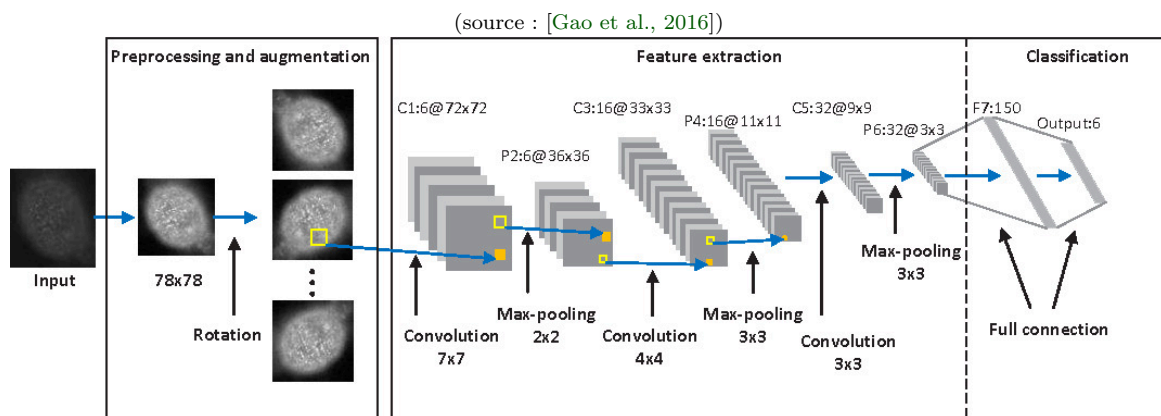


Figure 2.33 – Exemple d'un CNN pour la classification de cellule

2.5 Classification

La classification vise à identifier la catégorie à laquelle appartient un objet en se basant sur un ensemble de propriétés qui le décrit. Plus les données d'entrée sont variées (les descriptions de l'objet), plus il est "facile" pour la machine de trouver des modèles et plus le résultat est précis. De nombreux algorithmes d'apprentissage (machine learning), existent et se différencient en plusieurs domaines, dont [Yuan et al., 2019] en proposent une taxonomie (figure 2.34). Le but de cette section est simplement de montrer des exemples d'approches et non une présentation exhaustive des algorithmes. La plupart de ces méthodes de classification sont accessibles sous différents outils ou bibliothèques de programmation (shogun, mlpack, sklearn [Pedregosa et al., 2011], matlab ...).

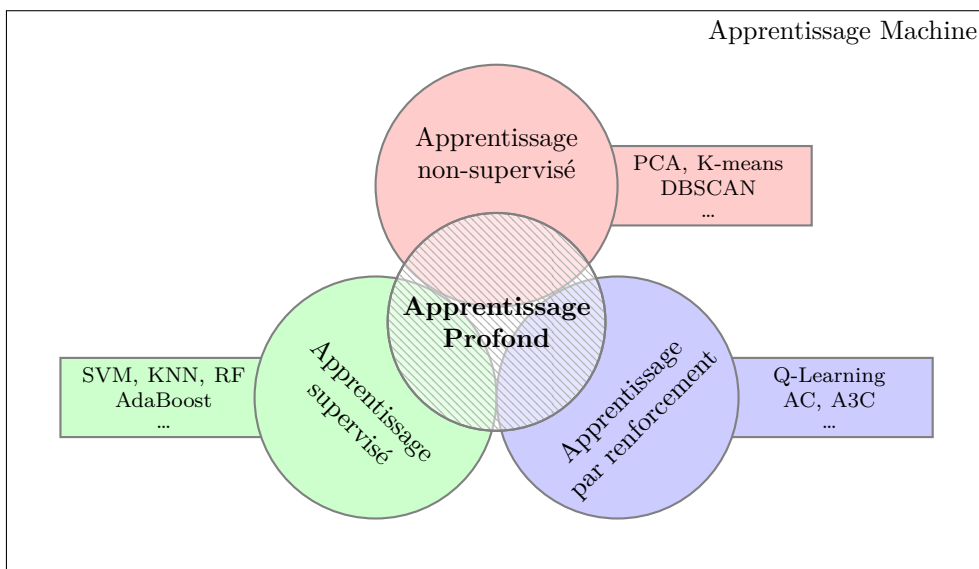


Figure 2.34 – Taxonomie des algorithmes d'apprentissage machine

2.5.1 Classification supervisée

L'objectif de la classification supervisée est de définir des règles qui permettent de classer les objets à partir de variables qualitatives ou quantitatives. Les règles de classification sont apprises en utilisant un échantillon d'apprentissage, contenant les variables discriminantes et leurs classements connus. La classification supervisée est une approche standard et privilégiée par la majorité des études, tout comme en agriculture de précision [Lourgant, 2016, p. 37] (knn, tree, svm, ...). Les réseaux de neurones [Lottes et al., 2018, Perumal and Bhaskaran, 2010, Koot, 2014] (ann, cnn, ...) font partie des méthodes de classification supervisée. La difficulté de cette méthode réside principalement dans la conception du jeu de données d'apprentissage.

2.5.1.1 Approche Bayésienne

Ici le but est de trouver en amont la distribution des classes afin de prédire l'appartenance de nouvelles données à partir d'une fonction de densité de probabilité connue. L'objectif est de réduire la probabilité d'erreur de classement [Dunsmore, 1966]. Lorsqu'une donnée est recouverte par plusieurs classes, c'est la classe dont la probabilité est la plus forte qui prend l'avantage. En cas d'égalité de probabilité, il n'y a pas de méthodes pour lever l'incertitude ce qui provoque une erreur de classification, comme dans le diagramme ci-dessus (figure 2.35).

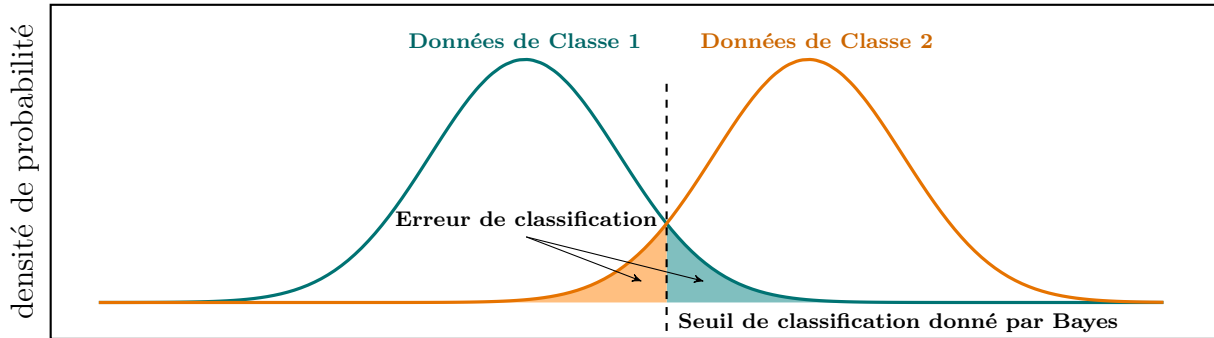


Figure 2.35 – Règle de décision de Bayes

La règle de décision de Bayes constitue en réalité la limite optimale de tout système de classification. Cette limite est théorique : face à un problème réel, les distributions des classes sont inconnues, ainsi que les probabilités a posteriori. Les différentes méthodes de résolution peuvent seulement en fournir des estimations. Cette erreur théorique constitue une borne infranchissable qui représente d'une certaine manière la difficulté intrinsèque du problème.

2.5.1.2 K Plus Proche Voisins

L'algorithme des KPPV ou KNN (*K Nearest Neighbor*) est un algorithme de classification et de régression. Dans cette méthode, on calcule la distance entre le point à classer et l'ensemble des autres points de toutes les classes d'une base de données de référence, on prend les k plus proches et la classe majoritaire identifie l'objet. C'est une méthode de référence lorsque les connaissances préalables sur la distribution des données sont insuffisantes [Peterson, 2009]. La figure 2.36 montre la classification du point noir en fonction de K . Quand K est pair il n'y a pas de classe majoritaire, ce qui implique une frontière d'indécision (visible en blanc, image central) et une heuristique supplémentaire peut être utilisée, par exemple en utilisant un voisin de moins (image de droite). Le choix de K et de l'heuristique est difficile. Si K est trop petit, la classification est trop sensible au bruit (image de gauche $K=1$). À l'inverse, si K est trop grand, la classification peut-être erronée par une trop grande zone de décision.

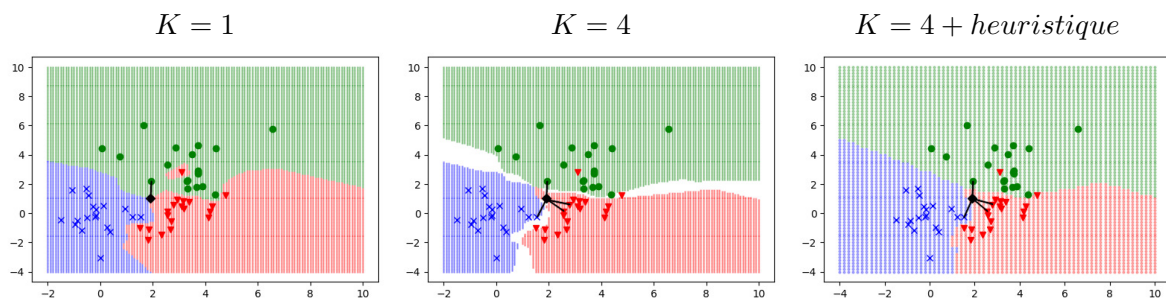


Figure 2.36 – Classification KNN : influence de K et de l'heuristique

2.5.1.3 Perceptron

Le perceptron est le modèle de base des réseaux de neurones [Rosenblatt, 1958]. Cet algorithme aide à trouver une droite de séparation entre deux sets de données par descente de gradient afin d'établir une classification des données entrées. Le perceptron recherche l'hyperplan séparateur entre deux classes de points (-1 et $+1$). L'hyperplan est défini par $f(x) = \sum_{i=1}^n w_i x_i + b = \langle w, x \rangle + b$ avec w le vecteur orthogonal à l'hyperplan, b le déplacement à l'origine et n le nombre de dimensions. Lorsque davantage de classes sont évaluées, il est possible d'obtenir la classification grâce à plusieurs droites de séparation, une par paire de classe. Visible sur la figure 2.37, dont les axes correspondent à des propriétés quelconques.

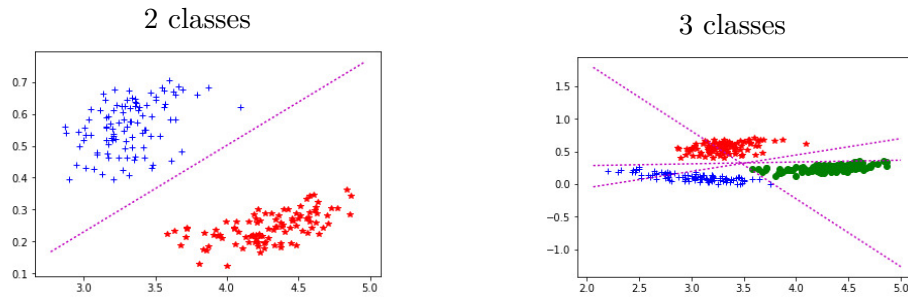


Figure 2.37 – Perceptron

2.5.1.4 Arbre de décision

L'objectif est de créer un modèle qui prédit la valeur d'une variable cible en apprenant des règles de décision simples déduites des caractéristiques des données. En combinant ces règles de décision, une structure arborescente est construite. Les feuilles finales représentent les classes et les branches représentent les conditions sur les caractéristiques qui mènent aux classes. Dans la figure suivante 2.38, en s'appuyant sur de la bibliographie, un arbre de décision peut-être manuellement construit pour classer des pixels en 3 classes (eau, sol, végétation) à partir d'un indice de végétation tel que le NDVI (voir section 2.2.3). Si celui-ci est inférieur à 0 alors c'est de l'eau, sinon, c'est "autre chose" et on continue. Cette approche peut être automatisée, pour chaque variable à disposition, une métrique est utilisée pour définir celle qui divise au mieux les ensembles (Gini Impurity / Information Gain / Variance / ...). Ensuite, la valeur de seuillage peut-être apprise par régression pour obtenir la meilleure séparation de deux ensembles. Cette solution présente l'avantage de donner l'importance aux propriétés et aux décisions associées, ce qui permet d'expliquer facilement la démarche de classification.

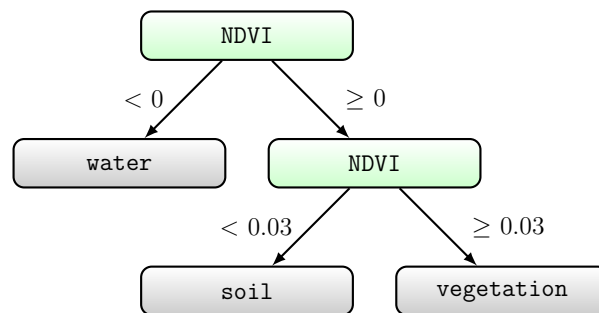


Figure 2.38 – Arbre de décision simple en agriculture de précision

2.5.2 Classification non-supervisée

Les méthodes de classification non-supervisées, sont souvent basées sur des mesures de distance pour agglomérer les individus proches, sans avoir besoin de connaître la classe de l'individu a priori. Cette approche propose deux grands intérêts, en agriculture de précision (i) l'absence de base d'apprentissage annotée et (ii) de s'abstraire des conditions d'acquisition, qui sont alors plus "uniformes", par exemple au sein de l'image. Dû à la faible séparabilité des propriétés extraites entre culture et adventice, cette solution n'est pas encore utilisable pour cette tâche. Cependant, elle reste applicable à la séparation sol/végétation, dont le seuillage *Otsu* présenté précédemment en est un exemple de classification binaire.

K-means L'algorithme des k-means permet de regrouper les individus au sein de groupes (clusters) distincts en utilisant une mesure de distance. Pour ce faire, on fixe des centres de groupes de départ aléatoirement en quantité déterminée. Ce sont des informations à priori que nous avons sur les données. Pour chaque donnée, on associe le centre le plus proche puis on recalcule les nouveaux centres des clusters ainsi formés. Cette étape est ré-itérée tant qu'il n'y a pas convergence, c'est-à-dire tant que les centres de chaque groupe ne sont pas stabilisés (avec une tolérance). À l'issue de son exécution, l'algorithme des k-means a alors associé chaque donnée à un groupe.

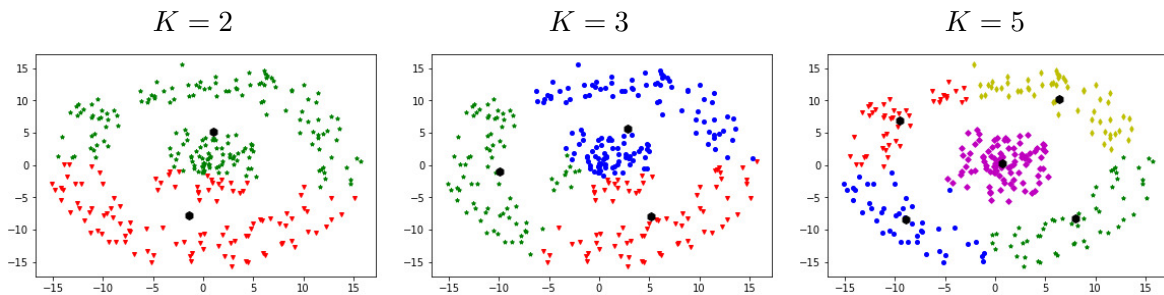


Figure 2.39 – Algorithme de classification K-Mean

DBSCAN L'algorithme considère les clusters comme des zones de haute densité séparées par des zones de faible densité. Deux paramètres, N et ϵ , définissent formellement ce que signifie "dense". On prend le premier individu et on vérifie combien de voisins sont autour de lui à une distance maximale de ϵ , si le nombre de voisins est inférieur à N c'est un "bruit", sinon il fait partie d'un cluster. On crée alors une catégorie en y associant les points par propagation. On répète l'opération pour chaque individu non traité. Donc prendre un N plus élevé ou un ϵ plus faible, indiquent une plus grande densité nécessaire pour former un cluster. La méthode, illustrée sur la figure 2.40 montre en orange le bruit, en vert les premiers individus testés et en bleu ceux associés par propagation.

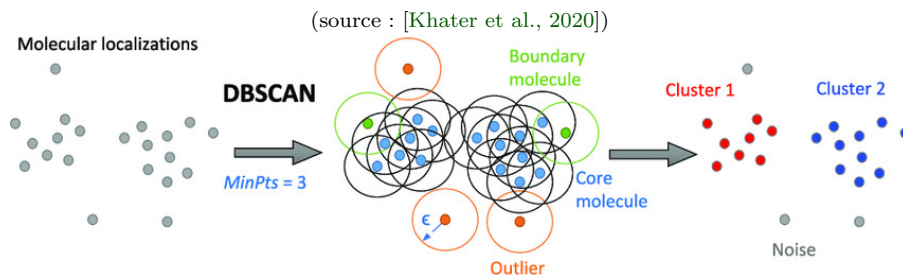


Figure 2.40 – Illustration de la méthode de clustering DBSCAN

2.5.3 Classification semi-supervisée

Un exemple est l'apprentissage d'un classificateur sur une petite base de données annotée. Ce dernier pourrait alors être optimisé dans le temps, à chaque nouvelle observation. Une littérature exhaustive sur les modèles d'apprentissage semi-supervisés a été effectuée par [Zhu, 2006]. Un autre exemple d'apprentissage semi-supervisé est le co-apprentissage, dans lequel deux classificateurs apprennent un ensemble de données, mais en utilisant chacun un ensemble de caractéristiques différentes, idéalement indépendantes.

Connaissance a priori L'étude [Lottes and Stachniss, 2017] met en place une méthode de détection d'adventices semi-supervisée (80–95 %) dans laquelle ils exploitent le fait que la plupart des cultures sont plantées en rangées avec un espacement similaire le long du rang, ce qui peut être utilisé pour améliorer un classificateur existant. Cette même connaissance a été utilisée par [Louargant et al., 2018]. Une segmentation *Otsu* est utilisée pour séparer le sol de la végétation. Une détection de rang permet ensuite de créer un sous-ensemble d'apprentissage pour la discrimination culture/adventice, en utilisant des informations spectrales. Finalement, la classe de chaque composante connexe est déduite à partir de ces deux classifications (rang+spectre). Les résultats (75–85 %) sont fortement dépendants de la résolution des images qui implique un fort mélange spectral et homogénéise les composantes connexes. Cette approche semble difficilement applicable en présence d'une densité importante de plantes.

Espace latent À partir d'un ensemble minimal de connaissances, la construction d'un espace latent qui représente ces individus peut être construite. On cherche alors à diminuer la variance intra-classe tout en maximisant la distance inter-classe, utilisée comme fonction de perte. À chaque nouvelle observation, l'individu est projeté dans l'espace latent et associé au cluster le plus proche. Quand suffisamment de nouvelles connaissances sont apportées, on peut renforcer la représentation de cette espace latent en minimisant une nouvelle fois la fonction de perte. Pour construire l'espace latent, différentes méthodes peuvent être utilisées, telles que l'apprentissage profond [Kamnitsas et al., 2018]. En agriculture de précision, on peut apprendre l'espace latent à partir d'un petit échantillon et à chaque nouvelle image, les cultures et les adventices précédemment classées sont utilisées pour optimiser l'espace latent, ce qui permet de prendre en considération les changements d'illumination ou de stade de développement des cultures.

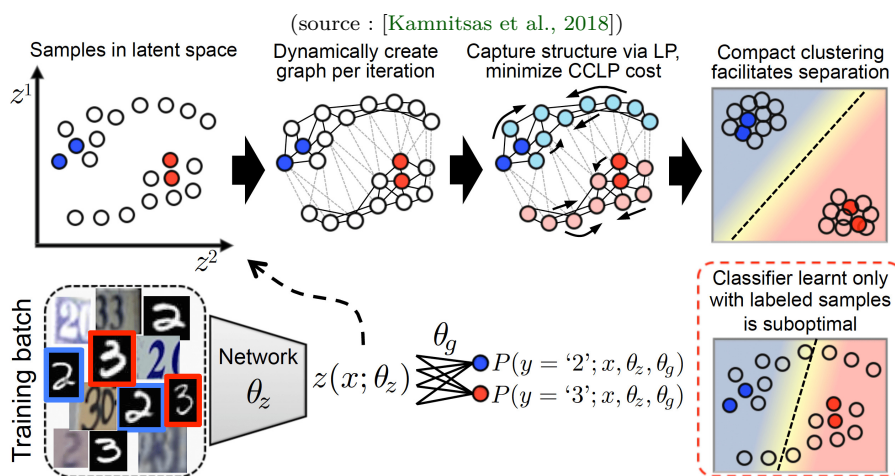


Figure 2.41 – Optimisation de l'espace latent

2.6 Métriques d'évaluation

Pour évaluer une méthode (classification, régression, ...), de nombreuses métriques sont utilisées dans la littérature [Hossin and Sulaiman, 2015, Di Gregorio et al., 2008]. En fonction du problème étudié et de la nature de l'information traitée, les métriques utilisables pour estimer les performances d'un modèle sont différentes (Figure 2.42). Ainsi dans cette section chaque partie s'intéressera à un problème spécifique et présentera un ensemble de métriques permettant de le traiter.

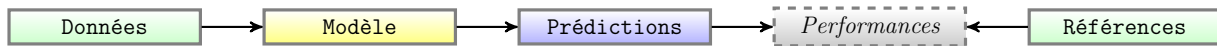


Figure 2.42 – Evaluation des performances d'un modèle après apprentissage

2.6.1 Classification

Pour faciliter la compréhension, nous resterons au cas binaire et nous noterons $y \in \{0, 1\}$ la vérité terrain et $p \in [0, 1]$ la prédiction. Ainsi, les métriques de performance pour la classification binaire sont conçues à partir de quatre quantités de bases de population. Définies par une matrice de confusion, ces quatre quantités sont : les vrais positifs (TP), les faux positifs (FP), les vrais négatifs (TN) et les faux négatifs (FN). Finalement pour obtenir ces quatre quantités fondamentales, un seuillage est appliqué à la prédiction p tel que $p > 0.5$. Deux variables complémentaires sont introduites, données par $\bar{y} = 1 - y$ et $\bar{p} = 1 - p$. La matrice de confusion C est donnée par :

$$C = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} = \sum_{i=0}^N \begin{bmatrix} y_i p_i & y_i \bar{p}_i \\ \bar{y}_i p_i & \bar{y}_i \bar{p}_i \end{bmatrix} \quad (2.1)$$

Il existe 21 métriques populaires que l'on peut calculer à partir de la matrice de confusion C [Tharwat, 2020]. Tel que démontré par [Koyejo et al., 2014], la plupart des métriques issues de ces quatre quantités peuvent être dérivées de la fonction \mathcal{L} définie ci-dessous :

$$\mathcal{L} = \frac{a_0 + a_{11}TP + a_{01}FP + a_{00}TN + a_{10}FN}{b_0 + b_{11}TP + b_{01}FP + b_{00}TN + b_{10}FN} \quad (2.2)$$

Parmi toutes ces métriques, les plus utilisées dans la littérature, sont : (i) le pourcentage de prédictions correctes (**Accuracy**), (ii) la précision (**Precision**) qui est la proportion des items pertinents (TP) parmi l'ensemble des items proposés ($TP + FP$), (iii) le rappel (**Recall**) qui est la proportion des items pertinents proposés (TP) parmi l'ensemble des items pertinents ($TP + FN$). Une mesure qui combine la précision et le rappel est leur moyenne harmonique (**F-score**), donnée par la fonction générique \mathcal{F}_β définie ci-dessous. À titre d'exemple, après simplification, la fonction \mathcal{F}_1 (aussi appelée Dice) est également proposée :

$$\mathcal{F}_\beta = \frac{(1 + \beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall} \quad \mathcal{F}_1 = \frac{2TP}{2TP + FP + FN} \quad (2.3)$$

Cependant, les métriques issues de la matrice de confusion peuvent être trompeuses, surtout en présence de déséquilibre important dans la population des classes. Pour corriger ce problème, une pondération (par la population respective) des éléments de la matrice de confusion peut être effectuée [Tripicchio et al., 2020].

2.6.2 Régression et calibration

Les méthodes de régression visent à apprendre un modèle qui transforme des variables continues (x_i) vers un vecteur dans un espace continu de dimension \mathbb{R}^N [Di Gregorio et al., 2008]. Pour la suite, on notera y une observation attendue et p une prédiction du modèle. Les indices y_i et p_i se réfèrent aux dimensions des variables.

Fonction de perte Une fonction de perte mesure la distance entre deux vecteurs, que l'on appelle aussi une fonction de dissimilarité, d'erreur ou de distance. Dans ce cas, plus la fonction $f(y, p)$ diminue, moins l'erreur est importante. On cherche donc à minimiser cette fonction. Il existe différentes mesures de distance dont la plus courante est la distance euclidienne. Cette distance euclidienne se note L^2 et est dérivée de la norme L^p (ou distance de Minkowski) définie ci-dessous. Elle synthétise une partie des mesures de distance. Il existe deux cas particuliers : L^0 qui n'est pas considérée comme une norme et L^∞ également définie ci-dessous.

$$L^\alpha = \sqrt[\alpha]{\sum_{i=0}^N |y_i - p_i|^\alpha} \quad L^\infty = \max_{i=0, \dots, N} (|y_i - p_i|) \quad (2.4)$$

À partir de cette norme L^p , différentes variantes ont été utilisées dans la littérature ; notamment l'erreur absolue (L^1), l'erreur quadratique ($(L^2)^2$), ou encore la distance de Mahalanobis (L^2 avec prise en compte d'une matrice de covariance). D'autres variantes existent pour des cas particuliers, ce qui signifie que cette base doit être adaptée en fonction de la nature de l'information et du problème. Par exemple la fonction ΔE_{00} calcule la distance perceptuelle entre deux couleurs, en prenant en considération différents espaces colorimétriques. Elle est donc utile pour la calibration d'une caméra à l'aide d'une mire.

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \frac{\Delta C'}{k_C S_C} \frac{\Delta H'}{k_H S_H}} \quad (2.5)$$

Fonction objectif Une fonction objectif mesure la ressemblance entre deux vecteurs, que l'on appelle aussi fonction de similarité. Dans ce cas, plus la fonction $f(y, p)$ diminue, plus l'erreur est importante. On cherche donc à maximiser cette fonction. La fonction de similarité du cosinus est la plus connue et est souvent utilisée pour mesurer la distance entre des vecteurs lorsque la magnitude (taille) des vecteurs n'a pas d'importance. C'est souvent le cas lorsque l'on travaille avec des données textuelles, où la seule chose qui compte est la fréquence d'apparition de chaque mot dans chaque texte. Cette similarité est toujours comprise entre 0 et 1, où 1 correspond à une similarité parfaite et 0 à aucune similarité.

Réciproque Une fonction de perte peut être définie par une fonction objectif (et inversement) telle-que : $g(x) = 1 - f(x)$ si $f(x) \in [0, 1]$. D'autres fonctions existent, par exemple, en utilisant l'équation 2.2 mais sans seuillage sur p , on retrouve alors les fonctions de perte utilisées en segmentation d'image, telles-que "Dice" (equation 2.3) ou $mIoU = 1 - yp \div (yp + \bar{y}p + y\bar{p})$.

Observations multiples Les distances précédemment introduites permettent de calculer l'erreur entre deux points, pour l'étendre à l'ensemble des observations, une somme ou une moyenne entre toutes les observations est alors utilisée. On retrouve alors les métriques suivantes : Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Sum Absolute Error (SAE), etc.

2.7 Applications en agriculture

En agriculture de précision, deux grands axes se dessinent pour la segmentation des plantes [Hu et al., 2021]. Soit (i) par l'utilisation d'une chaîne de traitement, telle que précédemment présentée avec ses différentes étapes [Perez-Sanz et al., 2017, Wang et al., 2019b]. Cette approche a été et est encore largement utilisée dans de nombreux secteurs. L'un des outils les plus connus pour cette tâche, en agriculture de précision est peut-être PlantCV [Fahlgren et al., 2015, Gehan et al., 2017]. Cet outil est encore en développement actif, récemment [Gutierrez Ortega et al., 2021] ont proposé des améliorations pour la segmentation de plantes se superposant. Soit (ii) par le développement d'approches par apprentissage profond, qui se sont largement répandues ces dernières années, les revues de [Kamilaris and Prenafeta-Boldú, 2018, Hu et al., 2021, Hasan et al., 2021, Osco et al., 2021] et [Wang et al., 2022] ont fait état de l'utilisation du deep-learning en agriculture de précision. Dans les parties suivantes, une rapide comparaison entre les méthodes classiques (chaîne de traitement) et modernes (apprentissage profond) sera présentée pour donner des exemples et aider à en comprendre le principe. Toutefois, des comparaisons plus détaillées ont été proposées par [Wu et al., 2021b] et [Salisu Abubakar et al., 2021].

2.7.1 Échelle du pixel et voisinage proche

L'utilisation d'une chaîne de traitement traditionnelle est peu répandue lorsque l'on souhaite travailler à l'échelle du pixel. En effet, l'utilisation de propriétés de textures (qui nécessitent la définition d'un élément structurant englobant plusieurs pixels et l'étude d'objets de taille plus importante) améliore rapidement les performances de classification. Une étude a cependant été effectuée sur des données RGB, par [R.S and Palanisamy, 2010]. L'algorithme proposé est divisé en deux étapes, la première consiste à séparer les plantes de l'arrière-plan (segmentation sol/végétation) à l'aide de l'indice de végétation ExcessGreen sur lequel un seuillage est appliqué. La deuxième étape consiste à séparer les adventices des plantes d'intérêt. Les auteurs se reposent sur le fait que les pixels qui sont liés aux cultures ont une composante verte plus élevée. Une opération logique est alors définie pour détecter les cultures, telle que si les valeurs du pixel respectent l'équation $(Green > Red) * (Green > Blue)$, il sera classifié comme culture. Pour détecter les adventices, un ensemble d'opérations morphologiques est appliqué sur le masque des cultures avant de le soustraire au masque de végétation. Étant donné que les données RGB ont un potentiel de discrimination limité, d'autres études utilisant des données multispectrales [Zwiggelaar, 1998] ou hyperspectrales ont permis d'obtenir de meilleures performances. C'est ce que propose l'étude de [Herrmann et al., 2013] avec une correction radiométrique et une analyse de type *Partial Least Squares Discriminant Analysis*. Cette dernière permet à la fois de détecter les bandes spectrales les plus utiles pour la discrimination et de définir un modèle de classification de forme linéaire. Chaque pixel est ensuite classé en fonction de ce modèle.

Les approches deep-learning ont par contre été largement exploitées en agriculture de précision pour détecter de la végétation. L'approche la plus simple consiste à utiliser et à apprendre une architecture U-Net (précédemment présentée), c'est ce que proposent [Ma et al., 2019, Brilhador et al., 2019] et [Fawakherji et al., 2019]. Cette approche comporte toujours de l'intérêt, [Singh et al., 2022] a testé trois types de réseaux convolutifs (VGG16, InceptionV3 et ResNet34) pour détecter les boules de coton dans les champs de culture. D'autres modèles, on été testés par [Khan et al., 2020] pour une discrimination des adventices avec les modèles SegNet, U-Net, FCN-8s, DeepLabv3 et CED-Net.

2.7.2 Échelle de la fenêtre

Dans cette approche, le principe consiste à découper l'image en N sous-images (Figure 2.21). L'article de [Rasti et al., 2019] propose une comparaison entre des approches standards et une approche deep-learning, pour la discrimination culture/adventice dans la mâche. Concernant l'approche standard, un ensemble de propriétés ont été extraites afin d'évaluer leurs potentiels de discrimination. L'algorithme de décomposition en ondelettes est comparé avec des techniques à une seule échelle et à plusieurs échelles telles que LBP, GLCM, le filtre de Gabor et le réseau de neurones convolutifs (visible en figure 2.43).

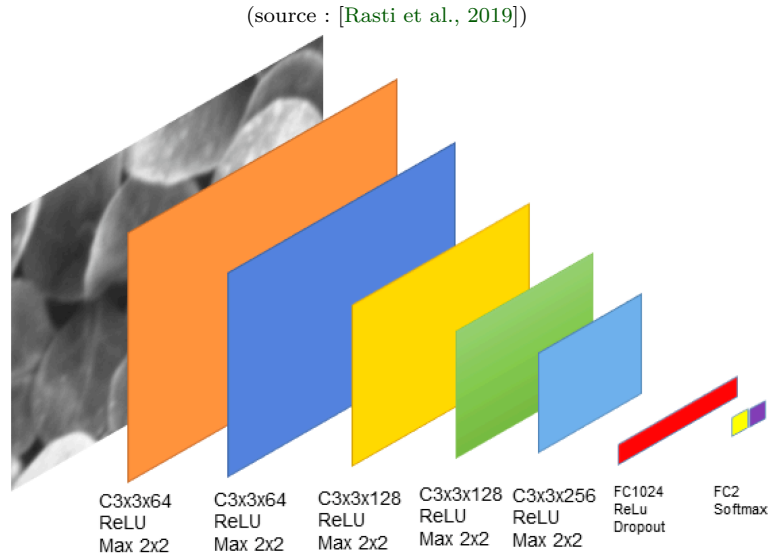


Figure 2.43 – Architecture CNN proposée

Pour les approches standards, les résultats expérimentaux de cette étude ont révélé que l'algorithme de décomposition en ondelettes permettait de mieux détecter les adventices que les autres méthodes proposées. La variété des méthodes testées contribue à confirmer l'efficacité de l'utilisation de cet algorithme sur cette tâche de discrimination culture/adventice. En revanche, l'utilisation d'un CNN donne de meilleurs résultats lorsque la taille de l'ensemble d'apprentissage augmente considérablement, montré sur la figure 2.44 ci-dessous :

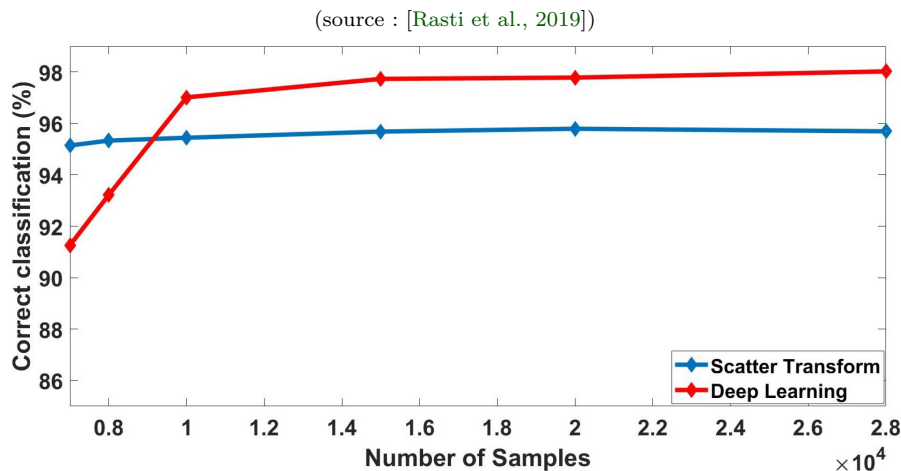


Figure 2.44 – Comparaison de la précision de reconnaissance entre décomposition en ondelettes et apprentissage profond, en fonction du nombre d'échantillons

2.7.3 Composantes connexes et super-pixels

Actuellement, la plupart des cultures sont semées en rangs réguliers, séparés par un espace défini qui dépend du type de culture et de la conduite choisie. En général, les plantes qui poussent en dehors des rangs sont considérées comme des adventices, communément appelées adventices inter-rangs. L'utilisation de composantes connexes semble souvent exploiter cette spécificité. C'est ce que proposent les deux approches de [Louargant et al., 2018] et [Bah et al., 2018a], qui sont très similaires, bien que la première utilise une chaîne de traitement et la seconde de l'apprentissage profond.

2.7.3.1 Composantes connexes

Concernant [Louargant et al., 2018] une chaîne de traitement est proposée et se compose d'une première étape de prétraitement sur les données. Une correction radiométrique et géométrique est utilisée. Ensuite, un algorithme de détection des rangs (transformée de Hough) est utilisé, ce qui permet d'obtenir les rangs de culture qui seront utilisés par la suite. Une segmentation sémantique est utilisée pour discriminer le sol de la végétation. Pour ce faire, un seuillage utilisant la méthode d'Otsu est appliqué sur un indice de végétation (NDVI). À partir de cette segmentation, les composantes connexes sont extraites. En fonction de la distance au rang de culture et des propriétés de forme des composantes, ces dernières sont pré-classées entre culture et adventice, puisque les adventices poussent majoritairement entre les rangs de culture. Finalement, ce pré-classement est utilisé pour initialiser un SVM utilisant des propriétés spectrales depuis les composantes connexes. Ce qui permet de reclasser les adventices intra-rangs. Ces étapes sont synthétisées sur la figure 2.45. Une approche similaire est proposée par [De Castro et al., 2018].

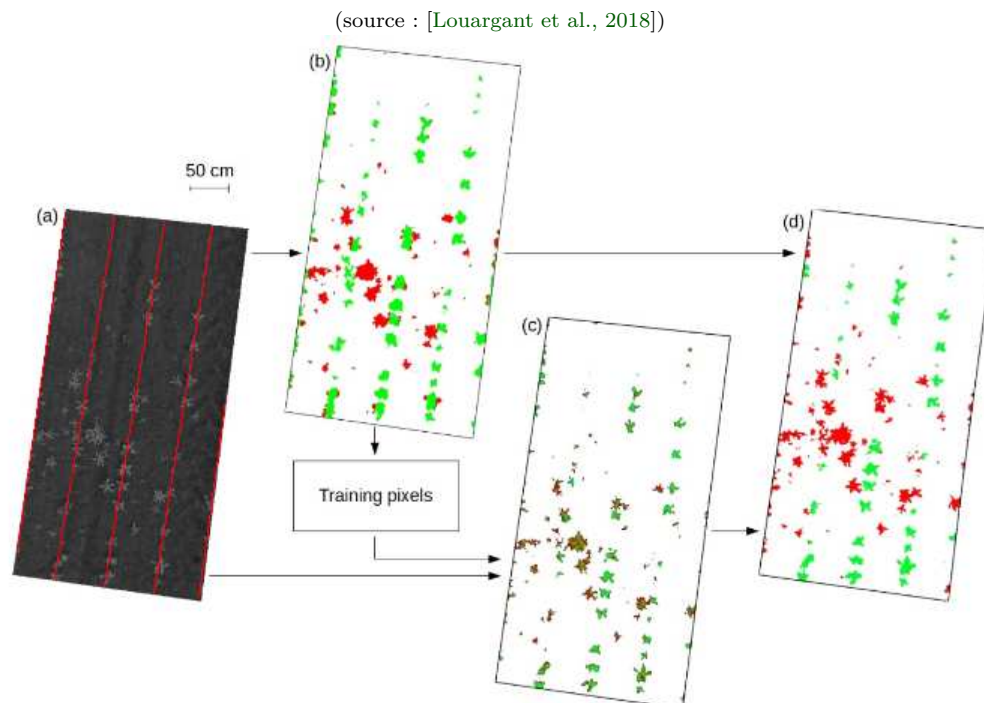


Figure 2.45 – Algorithme proposé par [Louargant et al., 2018], (a) détection des rangs (b) constitution d'une base d'apprentissage (c) classification spectrale et (d) reclassement des adventices intra-rangs

2.7.3.2 Super-pixels

[Bah et al., 2018a] proposent deux solutions, l’une basée sur une approche chaîne de traitement et l’autre sur l’utilisation d’un réseau de neurones. Ce qui permet de comparer ces deux approches sur un même jeu de données. Tout d’abord, une détection des rangs (Hough) de culture est utilisée. Ensuite, un algorithme de segmentation en super-pixel permet d’obtenir des entités. Les entités croisant les rangs visibles sont considérées comme de la culture, celles plus éloignées sont considérées comme adventices, ce qui permet encore une fois de constituer un jeu de donnée d’apprentissage. À partir de ce jeu d’apprentissage – de la même façon que [Louargant et al., 2018] – la dernière étape permet de reclasser les adventices intra-rangs à partir de ce premier pré-classement utilisant les rangs de culture. Pour ce faire, deux approches sont comparées. Soit l’utilisation de critères classiques (HoG, LBP, géométrique, ...) et l’utilisation d’un algorithme de classification (SVM). Soit l’utilisation d’apprentissage profond avec un CNN pour la classification de textures (ici ResNet) tel que synthétisé sur la figure suivante 2.46.

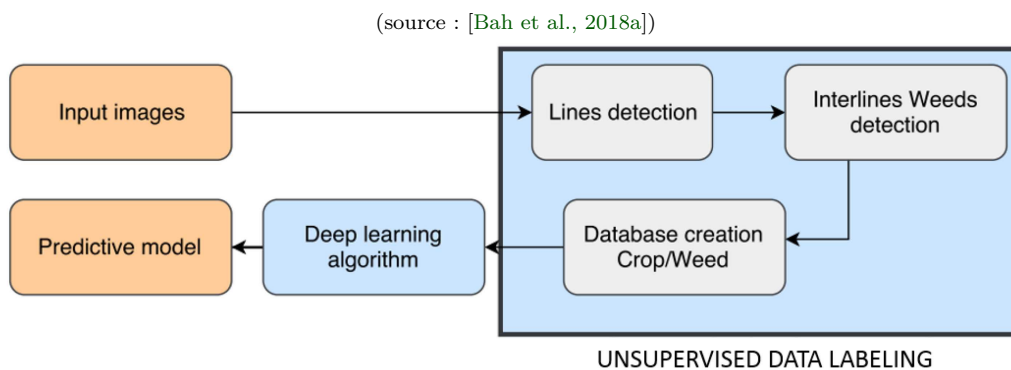


Figure 2.46 – Schéma de la méthode proposée par [Bah et al., 2018a]

Cette approche de segmentation en super-pixel a également été proposée par [Gée et al., 2021] avec l’utilisation d’une chaîne de traitement classique. La première étape consiste à séparer le sol de la végétation. Ici, plusieurs indices de végétation (ExG, ExR, CIVE, ...) sont combinés via un algorithme de classification (VotingClassifier), combinaison dénommée “MetaIndex”. Basée sur les travaux de [Bah et al., 2018a], une segmentation en super-pixels est proposée (SLIC) mais appliquée sur le “MetaIndex”. À partir de chaque super-pixel, une fenêtre est extraite (128×128) et des propriétés sont calculées grâce à l’extraction de points clés et l’apprentissage d’un algorithme nommé “Bag of Visual Word”. Ces éléments sont finalement classés via une classification SVM.

2.7.4 Échelle de la plante

La dernière possibilité consiste à détecter des formes plus précises dans l'image. L'étude de [Osorio et al., 2020] est intéressante dans le sens où elle compare une méthode classique (chaîne de traitement) à deux méthodes de deep-learning (Yolo et mask-RCNN). L'ensemble est appliqué à des données multispectrale de laitue, pour la détection des adventices. Un exemple de résultats de ces trois méthodes est montrées ci-dessous, en figure 2.47.

(source : [Osorio et al., 2020])

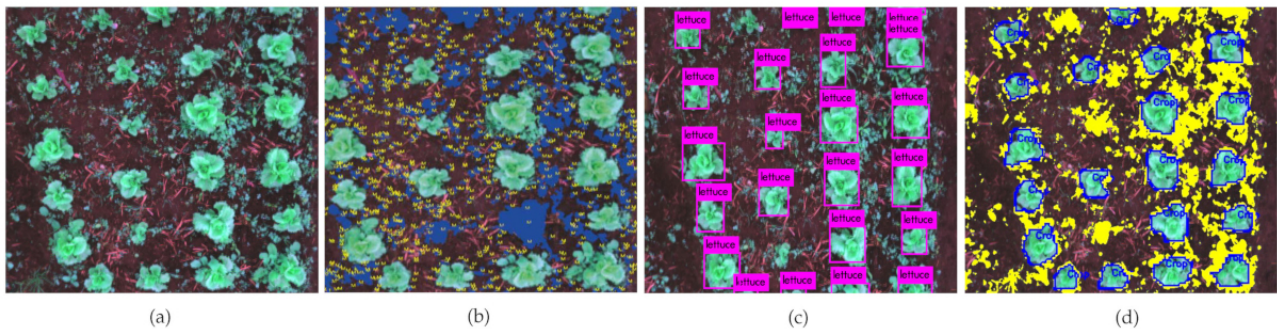


Figure 2.47 – Images résultantes : (a) Image prétraitée. (b) Méthode 1, chaîne de traitement. (c) Méthode 2, Yolo. Et (d) Méthode 3, mask-RCNN.

Concernant la première méthode, les auteurs se reposent sur un histogramme de gradients orientés (HOG). Cette technique a été utilisée pour extraire les caractéristiques de chaque objet de l'image, calculées à partir d'un indice de végétation (NDVI). Ensuite une classification par SVM permet de séparer ces objets entre "laitue" (culture) et "non-laitue" (adventices). Cette première classification est utilisée pour détecter les masques des objets "laitue", qui sont alors soustraits au masque de végétation (obtenu par une segmentation Otsu sur le NDVI). Le résultat est donc un masque de segmentation sémantique "adventices". La deuxième méthode proposée par les auteurs, consiste à utiliser un réseau de neurones de type YOLO. Ce dernier permet de détecter les fenêtres qui contiennent les laitues sur les données RGB. Comme précédemment, ces détections sont utilisées pour générer des masques de segmentation, qui sont soustraits à une segmentation Otsu appliquée au NDVI. Il en résulte encore une fois un masque de segmentation sémantique "adventices". Le même principe est appliqué sur la dernière méthode, en utilisant les masques générés par un mask-RCNN.

Les résultats de cette étude montrent que seule la deuxième méthode (YOLO) obtient des vitesses de calcul suffisantes pour un traitement temps réel. En effet l'utilisation d'un mask-RCNN est souvent coûteuse en temps de calculs pour détecter les masques des cultures. La première méthode n'est pas optimisée sur GPU, ce qui explique probablement les temps de calcul plus long, en particulier pour la phase de classification SVM. Dans tous les cas, les méthodes basées sur des réseaux de neurones (Yolo et mask-RCNN) donnent les mêmes taux de classification, tandis que la première est correcte mais moins performante. A titre indicatif, les résultats sont de 79 %, 89 % et 89 %, respectivement pour les méthodes 1, 2 et 3.

Les auteurs concluent que, le modèle YOLO est attrayant, en raison de sa rapidité et des performances démontrées. Cependant, le modèle mask-RCNN s'est distingué dans cette étude par sa capacité à détecter les contours précisément. Il a aussi été démontré que la méthode 1 (HOG-SVM) a obtenu de très bons résultats, et compte tenu du fait qu'elle nécessite moins de capacité de traitement, elle constitue une très bonne alternative pour les solutions IoT.

2.8 Conclusion de chapitre

À travers ce chapitre, nous avons vu que les possibilités sont nombreuses. Dans la chaîne de traitements classiques, on a des traitements de base sur l'image, puis une segmentation des individus et enfin, on a un ensemble de critères sur ces individus qui peuvent être utilisés pour les classer. Parallèlement, l'apprentissage profond est une méthode qui se développe beaucoup [Kamilaris and Prenafeta-Boldú, 2018], pour passer directement d'une image brute à une image classée. Elle offre de très bonnes performances, mais peu interprétables et nécessite des bases de données d'apprentissage conséquentes. Nous avons également vu que chaque étape de la chaîne de traitement classique peut reposer sur des solutions d'apprentissage profond. Cette approche n'est pas encore utilisée en agriculture de précision, il nous a paru intéressant d'introduire certaines de ces méthodes. Ainsi, cette thèse établit des recherches selon un nouvel axe d'investigation, tout en conservant une architecture standard, visible sur la figure suivante 2.48. Le chemin 1, en bleu, correspond à la chaîne de traitement standard, le chemin 3, en orange à l'approche "tout deep-learning" et le chemin 2, en vert, représente l'approche développée.

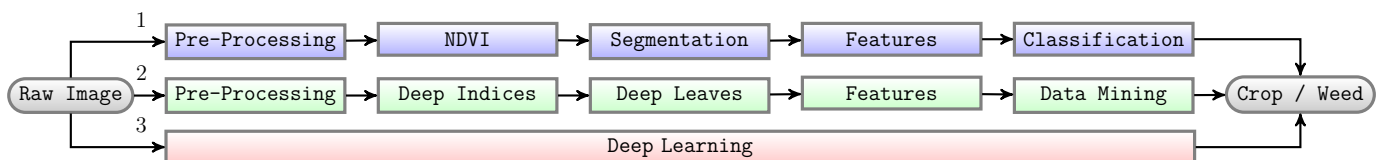


Figure 2.48 – Chaînes de traitement

Pré-traitement En terme de méthode de pré-traitement, une correction géométrique a été réalisée et une méthode d'enregistrement des bandes spectrales a été proposée (présentation à la conférence VISAPP). L'utilisation d'un dispositif d'étalonnage, étant inadéquate dans nos conditions d'acquisition, aucune correction radiométrique n'a été utilisée.

De nouveaux indices Durant cette thèse, une réflexion a été conduite afin de proposer des indices de végétation optimaux ne nécessitant pas de calibration radiométrique. Ces nouveaux indices reposent sur un apprentissage et des approximateurs de fonction, tout en utilisant l'ensemble des bandes spectrales du dispositif d'acquisition.

Segmentation Détecter les plantes est difficile, en particulier en présence de recouvrement. Nous proposons un travail à une échelle plus fine, celui de la feuille, afin de caractériser le couvert végétal avec précision. Ce travail repose sur une extension des indices de végétation précédemment développés. Les algorithmes de segmentation de feuilles proposés ici fonctionnent en feuillage dense et avec un éclairage non contrôlé.

Propriétés Dans cette thèse, nous avons voulu évaluer un ensemble de critères standards ; ensemble beaucoup plus vaste que ce qui est proposé dans la littérature. Tout d'abord, ces critères ont été étalonnés pour optimiser leurs potentiels de discrimination. Il s'agira ensuite de définir parmi ces critères, ceux qui sont clairement les plus discriminants.

Classification Ici aucune avancée fondamentale n'est proposée. Cependant, notre choix s'est porté sur des techniques de fouilles de données. Ce qui permet de définir les propriétés les plus discriminantes, tout en proposant un modèle de sortie "explicable".

3 Matériel et données

3.1 Les sites expérimentaux

3.1.0.1 Expérimentations en micro-parcelles

Une parcelle expérimentale a été mise en place à AgroSup Dijon (semée le 01/06/2019), afin de réaliser des acquisitions à différents stades de développement des cultures et des adventices (figure 3.1 gauche) et de préparer le challenge. Cette parcelle expérimentale a été découpée en 4 micro-parcelles de 9 m par 60 cm pour 4 cultures (de bas en haut : maïs, orge, féverole, trèfle) et trois rangs de cultures ont été semés pour chaque bande. Les adventices ont poussé naturellement et chaque bande a également été divisée en 3 modalités de désherbage de 3 m (faible < 20 %, partiel 40 – 60 % et aucun), dans le but de montrer l'impact de la segmentation sur le potentiel de discrimination. La parcelle a été préparée et semée manuellement.

3.1.0.2 Expérimentations en parcelles agricoles

Des acquisitions en parcelles agricoles ont été réalisées dans le cadre du challenge RoSE, avec le consortium ROSEAU. La parcelle expérimentale est gérée par OPEROSE [Avrin et al., 2020], équipe s'occupant des aspects opérationnels (entretiens, itinéraires techniques, ...) de l'ANR Challenge RoSE, localisée sur le site INRAE, Domaine des Palaquins, Montoldre dans l'Allier (figure 3.1 droite). Cette équipe organise les campagnes d'évaluations, permettant de mesurer la progression des performances et la maturité technologique des équipes au sein de l'ANR RoSE.

La parcelle expérimentale est divisée en plusieurs rangs de tests, dont un rang de maïs (inter-rang 75 cm – 80 cm) et un rang de haricot (inter-rang 15 cm – 37 cm) dédiés à la tâche de détection. Des adventices naturelles sont présentes, comme la matricaire, la renouée des oiseaux, le chénopode, la sétaire, la morelle, ou la digitale. Des adventices "modèles" ont été semées dans l'intra-rang, qui comporte de la moutarde, de la vesce, du trèfle, des lentilles, du ray-grass, de l'avoine, de la fétuque, et des dactyles.

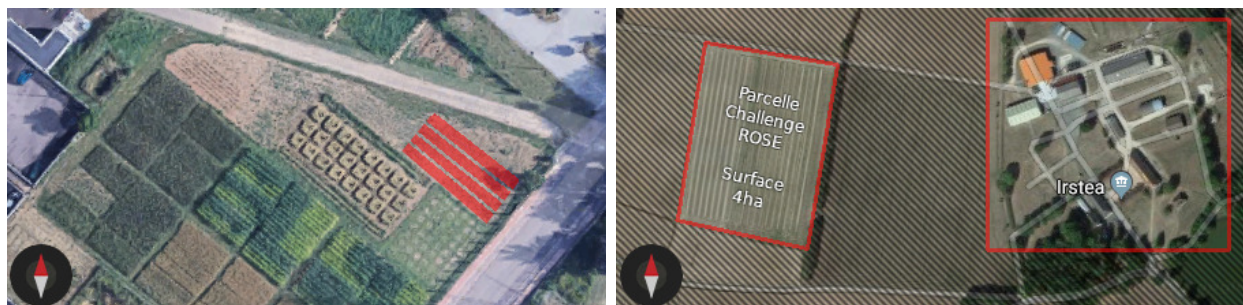


Figure 3.1 – Parcelle pour les premières expérimentations (rouge) à gauche et parcelles utilisées par le consortium ROSEAU (droite)

3.2 Capteur optique

Toutes les données ont été acquises avec la caméra multispectrale Airphen de la société Hiphen (Avignon, France). Ceci permettra de comparer plus facilement le niveau de discrimination de nos critères, en fonction des situations. Le système optique dispose de 6 capteurs optiques disposés en damier (figure 3.2). Chacun capteur se compose d'un CCD 4/3" (Taille : 4.8 mm × 3.6 mm) et d'une focale de 8 mm et est équipé d'un filtre interférentiel de 10 mm. Les 6 images acquises ont une taille de 1280 × 960 px et sont enregistrées au format TIFF avec un codage entier sur 12 bits.



Figure 3.2 – Système optique de la caméra Airphen (Société Hiphen, France)

Pour caractériser la végétation, le choix des bandes spectrales retenues pour les filtres interférentiels a été défini lors d'une étude précédente, conduite durant la thèse de [Lourgant, 2016]. Le travail effectué consistait à chercher les longueurs d'onde optimales pour la discrimination culture/adventices. Plusieurs méthodes ont été utilisées et une synthèse a permis de sélectionner les filtres les plus discriminants. Ils sont respectivement centrés sur 450, 570, 675, 710, 730 et 850 nm (figure 3.3). Ces longueurs d'onde seront utiles pour le calcul des indices de végétation, la segmentation et l'extraction des informations pour la discrimination culture/adventices.

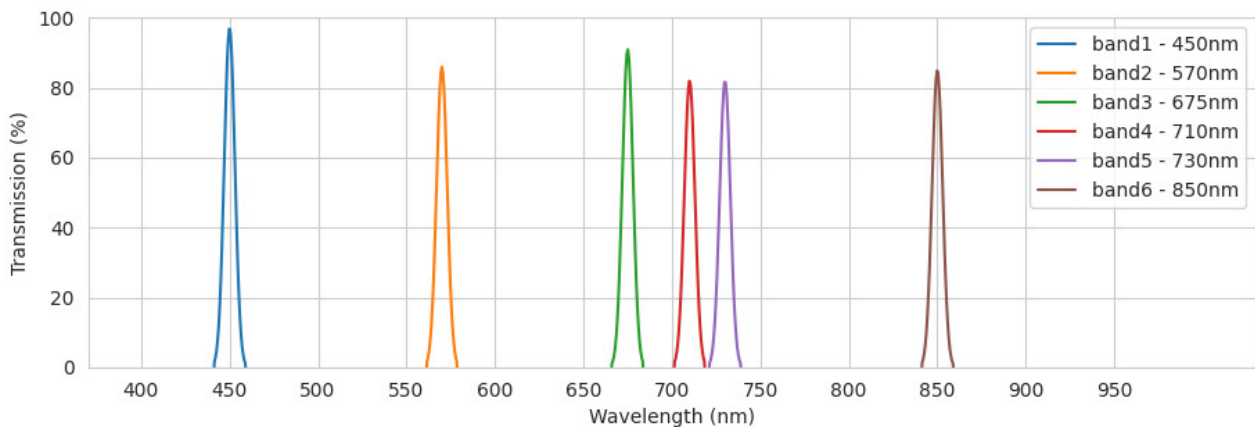


Figure 3.3 – Réponse spectrale des 6 filtres interférentiels placés devant les capteurs de la caméra multispectrale

3.3 Jeux de données

3.3.1 Campagnes d'acquisition

Micro-parcelles Un premier jeu de données a été acquis afin de tester et d'élaborer les différentes étapes de la chaîne fonctionnelle. La caméra multispectrale est embarquée sur une brouette thermique modifiée afin d'être dans les mêmes conditions expérimentales que le challenge ROSE. Les acquisitions ont été réalisées au mois de Juillet à quatre dates espacées d'une semaine. Le tableau 3.1 synthétise ces acquisitions. Cependant, durant cette période, la sécheresse a beaucoup impacté les cultures et les adventices ne permettant pas d'utiliser ces données pour montrer l'impact de la segmentation sur le potentiel de discrimination. Certaines de ces données ont néanmoins été utilisées pour calibrer la caméra et établir des indices de végétation optimaux. Notons que le maïs, qui est absent du tableau 3.1, n'a simplement pas levé. Finalement, les féveroles à la quatrième acquisition étaient en sénescence.

Culture	10/07/2019	17/07/2019	24/07/2019	31/07/2019
Orge 1	61	65	53	40
Orge 2	74	58	70	64
Féverole 1	51	66	51	–
Féverole 2	90	68	70	–
Trèfle 1	68	78	48	48
Trèfle 2	62	73	81	69

Table 3.1 – Dates d'acquisition, cultures et nombre d'images acquises

Challenge RoSE Pour les acquisitions réalisées avec le consortium ROSEAU, la caméra multispectrale Airphen a été embarquée sur le robot PUMAGRI (Figure 3.4) de la société SITIA. C'est un robot autonome destiné à la gestion localisée des adventices disposant de différents outils pour le désherbage (mécanique ou chimique). Dans le cadre de cette thèse, seule la partie détection des adventices est évaluée. Le challenge s'est déroulé sur trois ans entre 2018 et 2021, avec une campagne d'acquisition la première année, et deux les suivantes. Le nombre d'acquisitions par culture est synthétisé sur le tableau 3.2. Cette thèse ayant commencé en Décembre 2018, nous n'avons pas participé au premier challenge. De plus, le challenge 5 devant se dérouler en juin a finalement du être repoussé à cause des conditions météorologiques.

Culture	25/09/2019	10/06/2020	23/10/2020	13/09/2021
Haricot	403	438	458	434
Maïs	405	390	482	433

Table 3.2 – Nombre d'images obtenues lors des campagnes d'acquisition du challenge RoSE

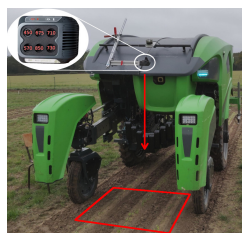


Figure 3.4 – Robot et caméra utilisés par le consortium ROSEAU

3.3.2 Annotations

3.3.2.1 Micro-parcelles

Les données acquises durant les expérimentations en micro-parcelles représentent un total de 1408 images multispectrales. Chaque image a été annotée individuellement et manuellement pour la discrimination sol/végétation. Les annotations sont définies par des masques de segmentation binaire (noir = sol ou blanc = végétation). À titre d'illustration, la figure 3.5 montre à gauche une reconstruction en fausses couleurs d'une culture de maïs (dans un champ avec diverses mauvaises herbes et la présence d'ombres dans les coins de l'image) et montre à droite le masque de segmentation associé.

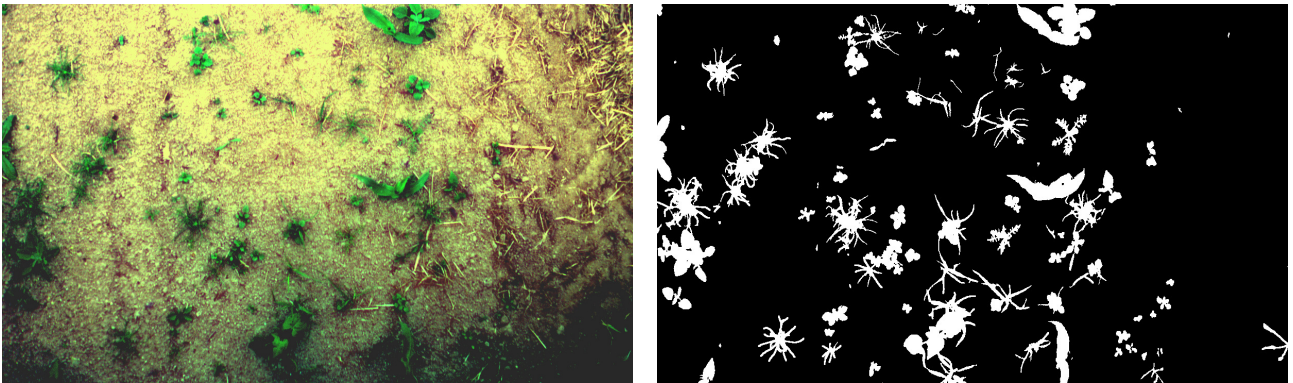


Figure 3.5 – Fausses couleurs à gauche et vérité terrain manuelle correspondante à droite

Ces données n'ont pas été annotées pour la discrimination culture/adventices. En effet, au moment de leur acquisition, le choix du niveau de segmentation nécessaire et la définition des individus (pixel, composante connexe, plante, ou feuille) étaient en cours de discussion.

3.3.2.2 Challenge RoSE

Les acquisitions faites durant les challenges RoSE représentent 3443 images multispectrales. Dû à cette quantité et au temps nécessaire à leurs annotations seule une partie a été annotée. Les conditions propices (beau temps, sans vent) du challenge 2 et 3 ont permis de définir une base de données annotée de 300 images de haricots et de 92 images de maïs, pour la détection des feuilles. Ce qui n'a pas été le cas lors des challenges 4 et 5 (pluie et vent). En revanche, les mauvaises conditions ont été utilisées pour renforcer “nos” indices de végétations, que l'on présentera plus loin dans le chapitre 4. Les annotations des feuilles ont été sous-traitées par la société Infokls (Inde) pour la partie haricot, et les annotations du maïs ont été faites en interne avec le logiciel VIA [Dutta and Zisserman, 2019]. Ainsi, chaque feuille est associée à un polygone et à une étiquette (culture/adventice) comme représenté sur la figure suivante 3.6.

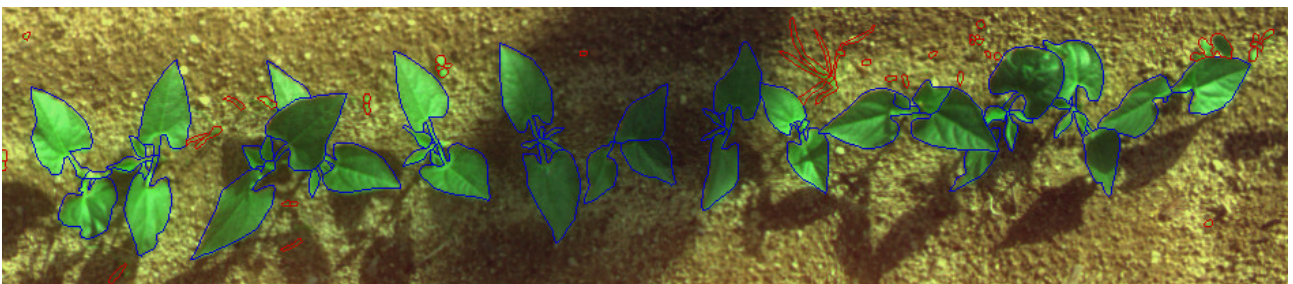


Figure 3.6 – Annotation pour la discrimination des feuilles, entre culture et adventices

3.3.2.3 Autres jeux de données

Les algorithmes mis en place dans la suite de la thèse et notamment pour l'article intitulé « Pixelwise Instance Segmentation Of Leaves In Dense Foliage », ont nécessité des bases de données supplémentaires.

Le premier jeu de données et a été fourni par le Réseau international sur le phénotypage des plantes (INPP) pour le Leaf Segmentation Challenge (LSC [Scharr et al., 2017]). Le jeu de données est en open-access. Il est composé d'images RGB de plantes Arabidopsis Thaliana (783 images) et de plantes de Rosette (27 images) segmentées en plusieurs feuilles (Figure 3.7). Les auteurs indiquent que les images ont été collectées à partir de plusieurs endroits dans une chambre de croissance expérimentale. Cette base de donnée est également acquis avec quatre caméras différentes et résolutions différentes.

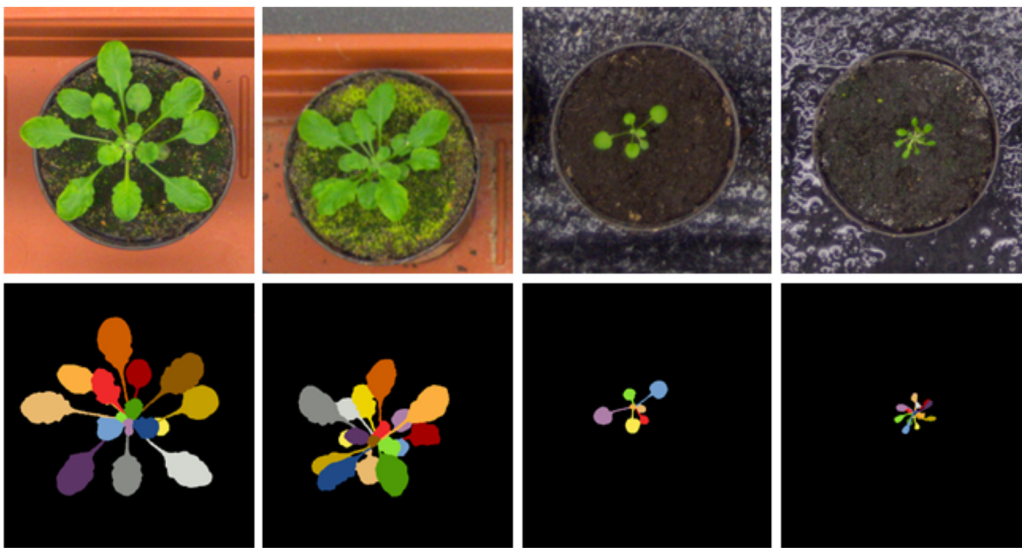


Figure 3.7 – Exemples issues de la base de données CVPPP (haut) et vérité terrain (bas)

Le deuxième jeu de données, Komatsuna, consiste en des images RGB de 137 plantes prises en vue de dessus. Ce jeu de données comprend 900 images de taille 480×480 pixels et contient un grand nombre de stades de croissance des plantes (138). Il a été conçu pour résoudre le problème du phénotypage 3D [Uchiyama et al., 2017].

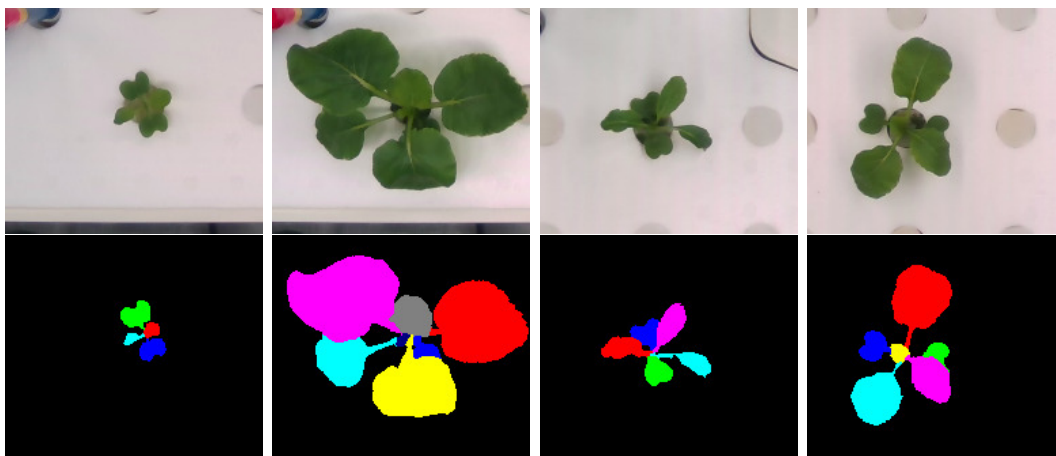


Figure 3.8 – Exemples issues de la base de données Komatsuna (haut) et vérité terrain (bas)

3.4 Calibration

3.4.1 Résolution spatiale

La caméra multispectrale Airphen est habituellement dédiée à des applications drone avec une optique optimisée pour des vols supérieurs à 25 m d'altitude. Dans le cadre du challenge RoSE, cette caméra est destinée à être fixée en amont du robot terrestre. La fauchée ou emprise au sol de l'image, représentée par le rectangle projeté, sur la figure 3.9, est fonction de l'altitude, de la taille du capteur et de la distance focale.

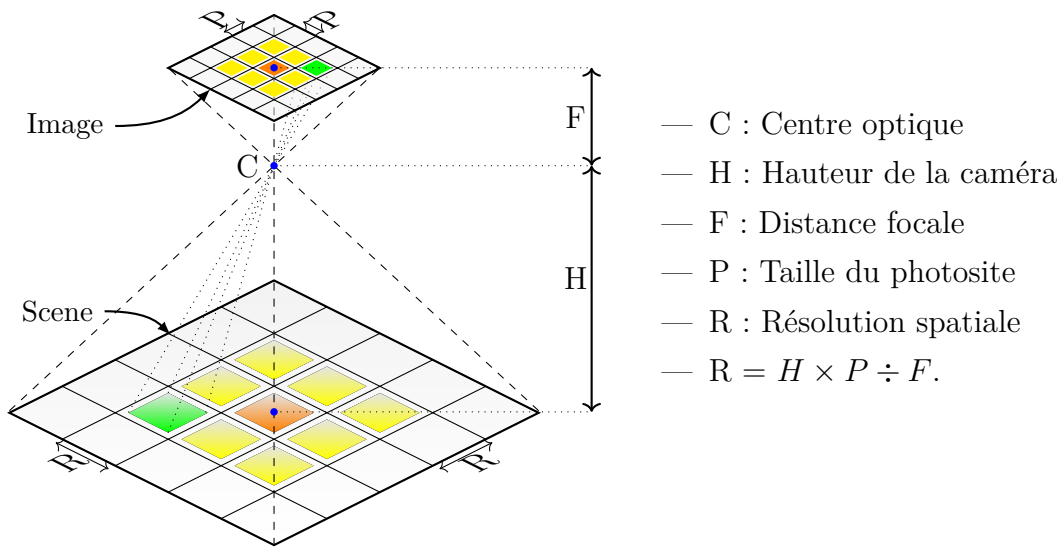


Figure 3.9 – Relation entre paramètres du monde réel et ceux du monde virtuel

À titre indicatif, le tableau suivant 3.3 montre les largeurs et longueurs de la zone observée par cette caméra en fonction de la hauteur. La résolution spatiale des pixels en millimètres est également renseignée. À basse altitude, la résolution des images est haute, a contrario plus la hauteur augmente, plus la zone observée est grande et plus la résolution spatiale des pixels est faible. Ainsi, à 1.80 m d'altitude, la zone observée est de 1.08 m × 0.81 m avec une résolution spatiale des pixels de 0.84 mm²/px. Ce qui permet d'observer les rangs de culture, tout en prenant une marge d'erreur sur les bords, nécessaire entre autres au recalage des images multispectrales.

Hauteur	1.00	1.20	1.40	1.60	1.80	2.00	2.20	2.40	2.60	2.80	3.00	Unité
Longueur	0.45	0.54	0.63	0.72	0.81	0.90	0.99	1.08	1.17	1.26	1.35	m
Largeur	0.60	0.72	0.84	0.96	1.08	1.20	1.32	1.44	1.56	1.68	1.80	m
Résolution	0.47	0.56	0.66	0.75	0.84	0.94	1.03	1.12	1.22	1.31	1.41	mm ² /px

Table 3.3 – Résolution spatiale des images à différentes hauteurs

3.4.2 Correction géométrique

Comme nous l'avons vu dans le chapitre précédent, à la section 2.2.1, l'une des premières source d'erreur est issue de l'alignement du capteur et de la lentille. Le fournisseur de la caméra Airphen ne propose pas de calibration des lentilles, une correction de ces dernières (6 capteurs optiques) a été faite pour corriger les déformations visuellement perceptibles. Chaque capteur dispose d'une lentille dont les spécifications sont identiques entre elles, mais peuvent légèrement différer en sortie de fabrication. Pour détecter les déformations, un échiquier calibré a été photographié sous différents angles de vue (figure 3.10). Ainsi 204 images ont été acquises pour chaque longueur d'onde.



Figure 3.10 – Échiquier de calibration géométrique

Le modèle standard de distorsion est obtenu par l'ajout de la distorsion tangentielle aux modèles de distorsion radiale et d'un modèle de distorsion prismatique [Wang et al., 2008]. La fonction Φ ci-dessous (3.1) définit le mappage polynomial $2D \rightarrow 2D$ du modèle de distorsion.

$$\Phi \begin{pmatrix} x_d \\ y_d \end{pmatrix} = \begin{pmatrix} x_d (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 x_d y_d + p_2 (r^2 + 2x_d^2) + s_1 r^2 + s_2 r^4 \\ y_d (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_2 x_d y_d + p_1 (r^2 + 2y_d^2) + s_3 r^2 + s_4 r^4 \end{pmatrix} \quad (3.1)$$

Où $r^2 = x_d^2 + y_d^2$, $(x_d, y_d)^t$ sont les coordonnées distordues et normalisées. La fonction Φ donne en sortie les coordonnées image normalisées corrigées, notées $(x_u, y_u)^t$. Les coordonnées image normalisées sont calculées par translation du centre optique et divisées par la profondeur de la focale en pixel. Ainsi, $(x_u, y_u)^t$ et $(x_d, y_d)^t$ sont indépendants de la dimension de l'image. Nous pouvons aussi réécrire l'équation de façon plus explicite, pour séparer les différents modèles :

$$\begin{aligned} x_u &= x_d + \underbrace{x_d(k_1 r^2 + k_2 r^4 + k_3 r^6)}_{\Delta_{radial}} + \underbrace{2p_1 x_d y_d + p_2 (r^2 + 2x_d^2)}_{\Delta_{tangential}} + \underbrace{s_1 r^2 + s_2 r^4}_{\Delta_{prism}} \\ y_u &= y_d + \underbrace{y_d(k_1 r^2 + k_2 r^4 + k_3 r^6)}_{\Delta_{radial}} + \underbrace{2p_2 x_d y_d + p_1 (r^2 + 2y_d^2)}_{\Delta_{tangential}} + \underbrace{s_3 r^2 + s_4 r^4}_{\Delta_{prism}} \end{aligned} \quad (3.2)$$

Il existe plusieurs outils pour chercher et estimer cette fonction, afin d'effectuer la rectification de la distorsion. La calibration a ainsi été réalisée avec OpenCV, et ce, pour chaque lentille de façon indépendante. Il existe donc 6 fonctions de mappage Φ . Après calibration, on peut calculer l'erreur de re-projection, exprimée en pixel. Pour ce faire, on calcule la moyenne des distances euclidiennes entre les points détectés et ceux calibrés et reprojétés. Les résultats sont visibles sur le tableau 3.4.

Bande Spectrale	450 nm	570 nm	675 nm	710 nm	730 nm	850 nm
Erreur Résiduelle	0.0115 px	0.0092 px	0.0096 px	0.0094 px	0.0092 px	0.0110 px

Table 3.4 – Erreur résiduelle, en pixel, après correction géométrique

Exemple À titre indicatif, les paramètres de déformation de la fonction Φ estimés pour la bande 450 nm sont visibles sur le tableau 3.5. Et la figure 3.11 montre les différentes corrections. De haut en bas : prismatique, tangentielle et radiale de la bande 450 nm. On constate ici que pour la bande 450 nm la correction prismatique est négligeable. À contrario, les variables pour la correction radiale (k_1 et k_2) et tangentielle (p_2) montrent des rectifications importantes.

Correction	Variables	Valeurs
radial	k_1, k_2, k_3	-0.55181 ; 0.30429 ; -0.00053
prism	s_1, s_2, s_3, s_4	0.03471 ; 0.01294 ; 0.0025 ; -0.00183
tangential	p_1, p_2	0.00267 ; -0.43599

Table 3.5 – Correction de la lentille pour la bande spectral 450 nm

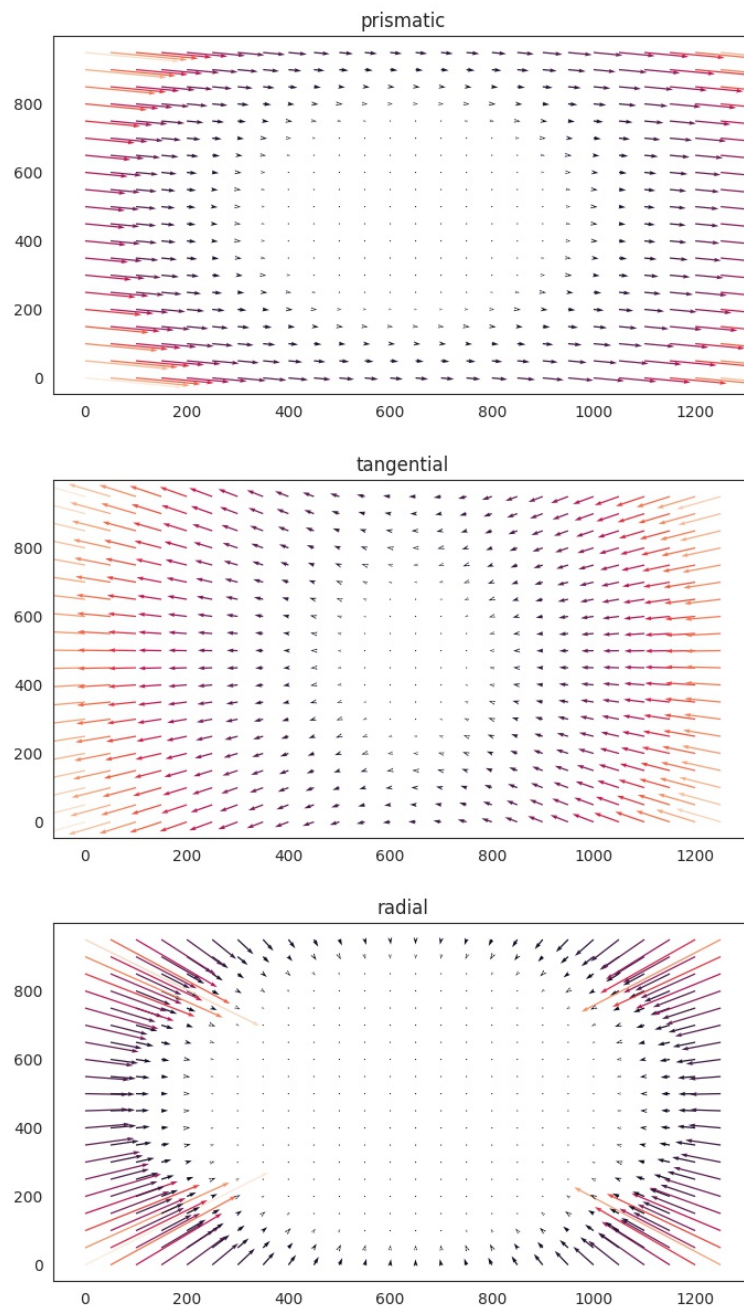


Figure 3.11 – Visualisation de la correction géométrique pour la bande 450 nm

3.5 Recalage des bandes spectrales

Comme nous l'avons vu précédemment, notre caméra est composée de 6 capteurs optiques disposés en damier. Cette conception fait que chaque capteur ne perçoit pas exactement la même zone au sol. Le décalage observé entre chaque bande est inversement proportionnel à la distance au sol ; c'est-à-dire que plus la caméra est proche du sol, plus le décalage est important, en nombre de pixels. La figure 3.12 montre le décalage des trois premières bandes spectrales (RGB) de deux fleurs, l'une jaune à gauche et l'autre blanche à droite sur une image prise à 1.60 m du sol avec la caméra Airphen.

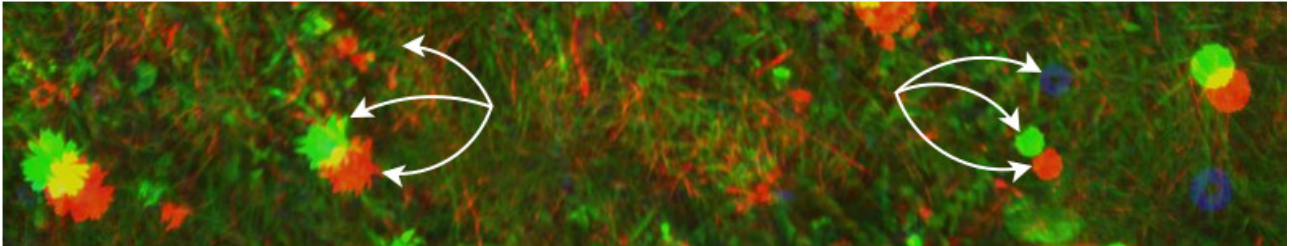


Figure 3.12 – Image non recalée, prise à 1.60 m

Dans le cadre de ce travail de thèse, il a donc été primordial de corriger ce décalage. Le processus de recalage des images consiste à transformer différentes images afin de leur affecter un système de coordonnées commun. Dans le domaine multispectral, l'approche traditionnelle pour le recalage consiste à désigner un canal de référence et à déplacer tous les autres vers celui-ci. Actuellement, seul [Junior et al., 2019] propose une méthode de sélection de la meilleure référence pour des images prises à 50 m. Dans la majorité des cas, le proche infrarouge (850 nm) ou la référence spectrale de moyenne gamme (vert pour RGB) sont conventionnellement utilisés. Les études existantes proposent principalement des méthodes basées sur la mise en correspondance de points clés, et optimisées à haute altitude. Toutefois, ce problème est complexe et encore ouvert, en particulier en champ proche, où ce type de matériel n'est pas couramment employé.

Two-step multi-spectral registration via key-point detector and gradient similarity. Application to agronomic scenes for proxy-sensing

Authors Jehan-Antoine VAYSSADE, Gawain Jones, Jean-Noël Paoli and Christelle Gée

Publication 21/02/2020 – VISAPP conference at Valetta in Malta – Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications – Volume 4 : VISAPP, pages 103-110

Abstract The potential of multi-spectral images is growing rapidly in precision agriculture, and is currently based on the use of multi-sensor cameras. However, their development usually concerns aerial applications and their parameters are optimized for high altitudes acquisition by drone (UAV \approx 50 meters) to ensure surface coverage and reduce technical problems. With the recent emergence of terrestrial robots (UGV), their use is diverted for nearby agronomic applications. Making it possible to explore new agronomic applications, maximizing specific traits extraction (spectral index, shape, texture ...) which requires high spatial resolution. The problem with these cameras is that all sensors are not aligned and the manufacturers' methods are not suitable for close-field acquisition, resulting in offsets between spectral images and degrading the quality of extractable informations. We therefore need a solution to accurately align images in such condition.

In this study we propose a two-steps method applied to the six-bands Airphen multi-sensor camera with (i) affine correction using pre-calibrated matrix at different heights, the closest transformation can be selected via internal GPS and (ii) perspective correction to refine the previous one, using key-points matching between enhanced gradients of each spectral bands. Nine types of key-point detection algorithms (ORB, GFTT, AGAST, FAST, AKAZE, KAZE, BRISK, SURF, MSER) with three different modalities of parameters were evaluated on their speed and performances, we also defined the best reference spectra on each of them. The results show that GFTT is the most suitable methods for key-point extraction using our enhanced gradients, and the best spectral reference was identified to be the band centered on 570 nm for this one. Without any treatment the initial error is about 62 px, with our method, the remaining residual error is less than 1 px, where the manufacturer's involves distortions and loss of information with an estimated residual error of approximately 12 px.

Keywords Registration, Multi-spectral imagery, Precision farming, Feature descriptor

3.5.1 Introduction

Modern agriculture is changing towards a system that is less dependent on pesticides [Lechenet et al., 2014] (herbicides remain the most difficult pesticides to reduce) and digital tools are of great help in this matter. The development of imaging and image processing have made it possible to characterize an agricultural plot [Sankaran et al., 2015] (crop health status or soil characteristics) using non-destructive agronomic indices [Jin et al., 2013] replacing traditional destructive and time-consuming methods. In recent years, the arrival of miniaturized multi-spectral and hyper-spectral cameras on Unmanned Aerial Vehicles (UAVs) has allowed spatio-temporal field monitoring. These vision systems have been developed for precise working conditions (flight height 50 m). Although, very practical to use, they are also used for proxy-sensing applications. However, the algorithms offered by manufacturers to co-register multiple single-band images at different spectral range, are not optimal for low heights. It thus requires a specific close-field image registration.

Image registration is the process of transforming different images of one scene into the same coordinate system. The spatial relationships between these images can be rigid (translations and rotations), affine (shears for example), homographic, or complex large deformation models (due to the difference of depth between ground and leaves) [Kamoun, 2019]. The main difficulty is that multi-spectral cameras have low spectral coverage between bands, resulting in a loss of characteristics between them. Which is caused by (i) plant leaves have different aspect depending on the spectral bands (ii) there are highly complex and self-similar structures in our images [Douarre et al., 2019]. It therefore affects the process of detecting common characteristics between bands for image registration. There are two types of registration, feature based and intensity based [Zitová and Flusser, 2003]. Feature based methods works by extracting point of interest and use feature matching, in most cases a brute-force matching is used, making those techniques slow. Fortunately these features can be filtered on the spatial properties to reduce the matching cost. A GPGPU implementation can also reduce the comparisons cost. Intensity-based automatic image registration is an iterative process, and the metrics used are sensitive to determine the numbers of iteration, making such method computationally expensive for precise registration. Furthermore multi-spectral implies different metrics for each registered bands which is hard to achieve.

Different studies of images alignment using multi-sensors camera can be found for acquisition using UAV at medium (50 – 200 m) and high (200 – 1000 m) distance. Some show good performances (in term of number of key-points) of feature based [Dantas Dias Junior et al., 2019, Vakalopoulou and Karantzalos, 2014] with strong enhancement of feature descriptor for matching performances. Other prefer to use intensity based registration [Douarre et al., 2019] on better convergence metrics [Chen et al., 2018] (in term of correlation), which is slower and not necessarily robust against light variability and their optimization can also fall into a local minimum, resulting in a non-optimal registration [Vioix, 2004].

Traditional approach to multi-spectral image registration is to designate one channel as the target channel and register all the others on the selected one. Currently, only [Dantas Dias Junior et al., 2019] show a method for selecting the best reference, but there is no study who as defined the best spectral reference in agronomic scene. In all cases NIR (850 nm) or middle range spectral reference are conventionally used without studying the others on precision agriculture. In addition those studies mainly propose features matching without large methods comparison [Dantas Dias Junior et al., 2019](less than 4) of their performance (time/precision), without

showing the importance of the spectral reference and the interest of normalized gradients transformation (like in Intensity-based methods).

However, despite the growing use of UGVs and multi-spectral imaging, the domain is not very well sourced, and no study has been found under agricultural and external conditions in near field of view (less than 10 meter) for multi-spectral registration.

Thus, this study propose a benchmark of popular feature extractors inside normalized gradients transformation and the best spectral reference was defined for each of them. Moreover a pre-affine registration is used to filter the feature matching, evaluated at different spatial resolutions. So this study shows the importance of the selection of the reference and the features extractor on normalized gradients in such registration.

3.5.2 Material and Method

3.5.2.1 Material

Camera The multi-spectral imagery is provided by the six-band multi-spectral camera Airphen developed by HiPhen. Airphen is a scientific multi-spectral camera developed by agronomists for agricultural applications. It can be embedded in different types of platforms such as UAV, phenotyping robots, etc.

Airphen is highly configurable (bands, fields of view), lightweight and compact. The camera was configured using interferential filter centered at 450/570/675/710/730/850 nm with FWHM¹ of 10 nm, the position of each band is referenced on figure 3.13. The focal lens is 8 mm for all wavelength. The raw resolution for each spectral band is 1280 × 960 px with 12 bit of precision. Finally the camera also provides an internal GPS antenna that can be used to get the distance from the ground.



Figure 3.13 – Disposition of each band on the Airphen multi-sensors camera

1. Full Width at Half Maximum

Datasets Two datasets were taken at different heights with images of a chessboard (use for calibration) and of an agronomic scene. We used a metallic gantry for positioning the camera at different heights. The size of the gantry is $3 \times 5 \times 4$ m. Due to the size of the chessboard (57×57 cm with 14×14 square of 4 cm), the limited focus of the camera and the gantry height, we have bounded the acquisition heights from 1.6 to 5 m with 20 cm steps, which represents 18 acquisitions.

The first dataset is for the calibration. A chessboard is taken at different heights The second one is for the alignment verification under real conditions. One shot of an agronomic scene is taken at different heights with a maximum bias set at 10 cm.

3.5.2.2 Methods

Alignment is refined in two stages, with (i) affine registration approximately estimated and (ii) perspective registration for the refinement and precision. As example the figure 3.14 shows each correction step, where the first line is for the (i) affine correction (section 3.5.2.2), the second is for (ii) perspective correction. More precisely the second step is per-channel pre-processed where feature detectors are used to detect key-points (section 3.5.2.2). Each channel key-points are associated to compute the perspective correction through homography, to the chosen spectral band (section 3.5.2.2). These steps are explained on specific subsections.

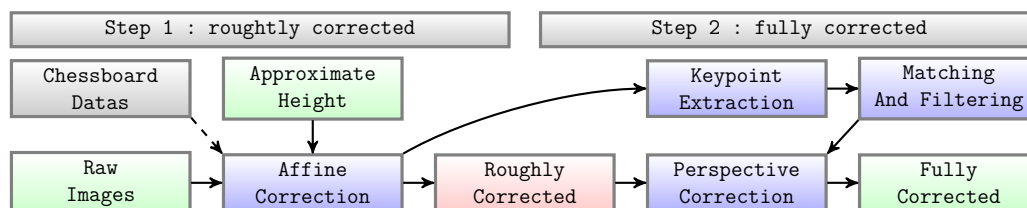


Figure 3.14 – Each step of the alignment procedure, with (step 1) roughly corrected from affine correction and (step 2) enhancement via key-points and perspective

Affine Correction We make the assumption that closer we take the snapshot, the bigger the distance between each spectral band is. On the other hand, at a distance superior or equals to 5 m, the initial affine correction become stable. A calibration is used to build a linear model based on that assumption, which will allow the affine correction to work at any height. The main purpose of this step is to reduce the offset between each spectral band, which allows the similarity between key-points to be spatially delimited within a few pixels, making feature matching more effective.

Calibration : Based on that previous assumption a calibration is run over the chessboard dataset. We detect the chessboard using the opencv calibration toolbox [Bouguet, 2001] on each spectral image (normalized by $I = (I - \min(I)) / \max(I)$ where I is the spectral image) at different heights (from 1.6 m to 5 m). We use the function `findChessboardCorners` how attempts to determine whether the input image is a view of the chessboard pattern and locate the internal chessboard corners. The detected coordinates are roughly approximated. To determine their positions accurately we use the function `cornerSubPix` as explained in the documentation².

2. https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html

Linear model : Using all the points detected for each spectral band, we calculate the centroid grid (each point average). The affine transform from each spectral band to this centroid grid is estimated. Theoretically, the rotation and the scale (A, B, C, D) do not depend on the distance to the ground, but the translation (X, Y) does. Thus a Levenberg-Marquardt curve fitting algorithm with linear least squares regression [Moré, 1978] can be used to fit an equation for each spectral band against X and Y independently to the centroid grid. We adjust the following curve $t = \alpha h^3 + \beta h^2 + \theta h + \gamma$ where h is the height, t is the resulted translation and factors $\alpha, \beta, \theta, \gamma$ are the model parameters.

Correction : Based on the model estimated on the chessboard dataset, we transpose them to the agronomic dataset. To make the affine matrix correction, we used the rotation and scale factors at the most accurate height (1.6 m where the spatial resolution of the chessboard is higher), because it does not theoretically depend on the height. For the translation part, the curve model is applied for each spectral band at the given height provided by the user. Each spectral band is warped using the corresponding affine transformation. Finally, all spectral bands are cropped to the minimal bounding box (minimal and maximal translation of each affine matrix). This first correction is an approximation. It provides some spatial properties that we will use on the second stage.

Perspective correction Each spectral band has different properties and values by nature but we can extract the corresponding similarity by transforming each spectral band into its absolute derivative, to find similarities in gradient break among them. As we can see in figure 3.15 gradients can have opposite direction depending on the spectral bands, making the absolute derivative an important step for matching between different spectral band.

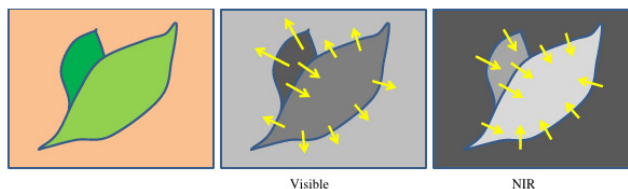


Figure 3.15 – Gradient orientation in spectral band [Rabatel and Labbe, 2016]. Orientation of the gradient is not the same depending to the spectral band.

The affine correction attempts to help the feature matching by adding properties of epipolar lines (close). Thus, the matching of extracted features can be spatially bounded, (i) we know that the maximum translation is limited to a distance of a few pixels (less than 10px thanks to affine correction), and (ii) the angle between the initial and the matched one is limited to $[-1, 1]$ degree.

Computing the gradient : To compute the gradient of the image with a minimal impact of the light distribution (shadow, reflectance, specular, ...), each spectral band is normalized using Gaussian blur [Sage and Unser, 2003], the kernel size is defined by $\text{next_odd}(\text{image_width}^{0.4})$ (19 in our case) and the final normalized images are defined by $I/(G + 1) * 255$ where I is the spectral band and G is the Gaussian blur of those spectral bands. This first step minimizes the impact of the noise on the gradient and smooth the signal in case of high reflectance. Using this normalized image, the gradient $I_{grad}(x, y)$ is computed with the sum of absolute Sharr filter [Seitz, 2010] for horizontal S_x and vertical S_y derivative, noted $I_{grad}(x, y) = \frac{1}{2}|S_x| +$

$\frac{1}{2}|S_y|$. Finally, all gradients $I_{grad}(x, y)$ are normalized using CLAHE [Zuiderveld, 1994] to locally improve their intensity and increase the number of key-points detected.

Key-points Extractor : A key-point is a point of interest. It defines what is important and distinctive in an image. Different types of key-point extractors are available and the following are tested :

(ORB) Oriented FAST and Rotated BRIEF [Rublee et al., 2011], (AKAZE) Fast explicit diffusion for accelerated features in nonlinear scale spaces [Alcantarilla and Solutions, 2011], (KAZE) A novel multi-scale 2D feature detection and description algorithm in nonlinear scale spaces [Ordonez et al., 2018], (BRISK) Binary robust invariant scalable key-points [Leutenegger et al., 2011], (AGAST) Adaptive and generic corner detection based on the accelerated segment test [Mair et al., 2010], (MSER) maximally stable extremal regions [Donoser and Bischof, 2006], (SURF) Speed-Up Robust Features [Bay et al., 2006], (FAST) FAST Algorithm for Corner Detection [Trajković and Hedley, 1998] and (GFTT) Good Features To Track [Shi et al., 1994].

These algorithms are largely described across multiple studies [Dantas Dias Junior et al., 2019, Tareen and Saleem, 2018, Zhang et al., 2016, Ali et al., 2016], they are all available and easily usable in OpenCV. Thus we have studied them by varying the most influential parameters for each of them with three modalities, the table 3.6 in appendix shows all modalities.

Table 3.6 – list of algorithms with 3 modalities of their parameters

ABRV	Parameters	Modality 1	Modality 2	Modality 3
ORB	nfeatures	5000	10000	15000
GFTT	maxCorners	5000	10000	15000
AGAST	threshold	71	92	163
FAST	threshold	71	92	163
AKAZE	nOctaves, nOctaveLayers	(1, 1)	(2, 1)	(2, 2)
KAZE	nOctaves, nOctaveLayers	(4, 2)	(4, 4)	(2, 4)
BRISK	nOctaves, patternScale	(0, 0.1)	(1, 0.1)	(2, 0.1)
SURF	nOctaves, nOctaveLayers	(1, 1)	(2, 1)	(2, 2)
MSER	None	None	None	None

Key-point detection : We use one of the key-point extractors mentioned above between each spectral band gradients (all extractors are evaluated). For each detected key-point, we extract a descriptor using ORB features. We match all detected key-points to a reference spectral band (all bands are evaluated). All matches are filtered by distance, position and angle, to eliminate a majority of false positives along the epipolar line. Finally we use the function *findHomography* between the key-points detected/filtered with RANSAC [Fischler and Bolles, 1981], to determine the best subset of matches to calculate the perspective correction.

Correction : The perspective correction between each spectral band to the reference is estimated and applied. Finally, all spectral bands are cropped to the minimum bounding box, which is obtained by applying a perspective transformation to each corner of the image.

3.5.3 Results and discussion

Firstly the results will focus on affine corrections and then on the effects of the perspective correction. Figure 3.16 shows a closeup inside at 1.6 m (3.16a) raw images acquisition, (3.16c & 3.16d) registered image of each correction steps and (3.16b) the manufacturer results.

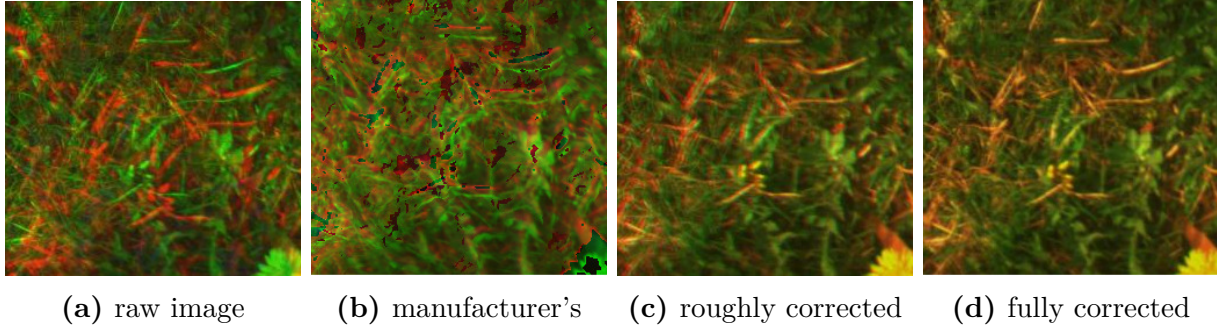


Figure 3.16 – Example of each correction and the manufacturers results

3.5.3.1 Affine correction

The affine correction model is based on the calibration dataset (where the chessboard are acquired). The 6 coefficients (A, B, C, D, X, Y) of the affine matrix were studied according to the height of the camera in order to see their stability. It appears that the translation part (X, Y), depends on the distance to the field (appendix figure 3.17) according to the initial assumption. On this part the linear model is used to estimate the affine correction from an approximated height.

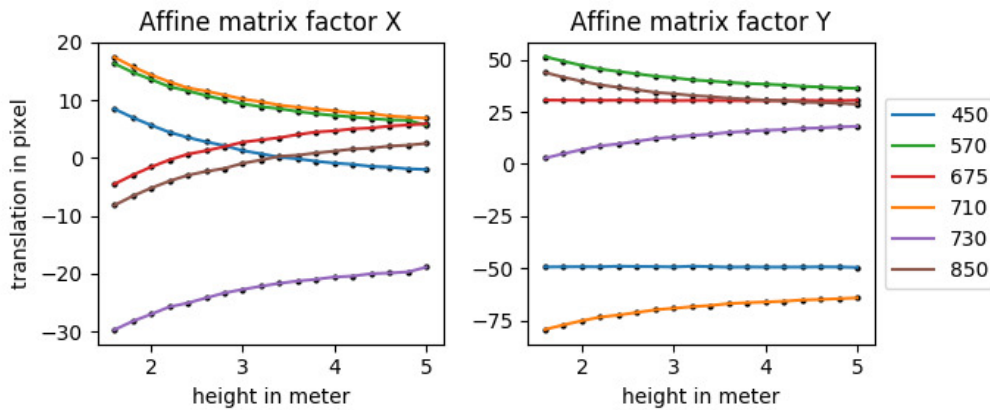


Figure 3.17 – Affine matrix value by height

Rotation and scale do not depend on the ground distance (figure 3.18) according to the theory. These factors (A, B, C, D) are quite stable and close to identity, as expected (accuracy depends on the spatial resolution of the board). As result, single calibration can be used for this part of the matrix, and the most accurate are used (i.e where the chessboard has the best spatial resolution).

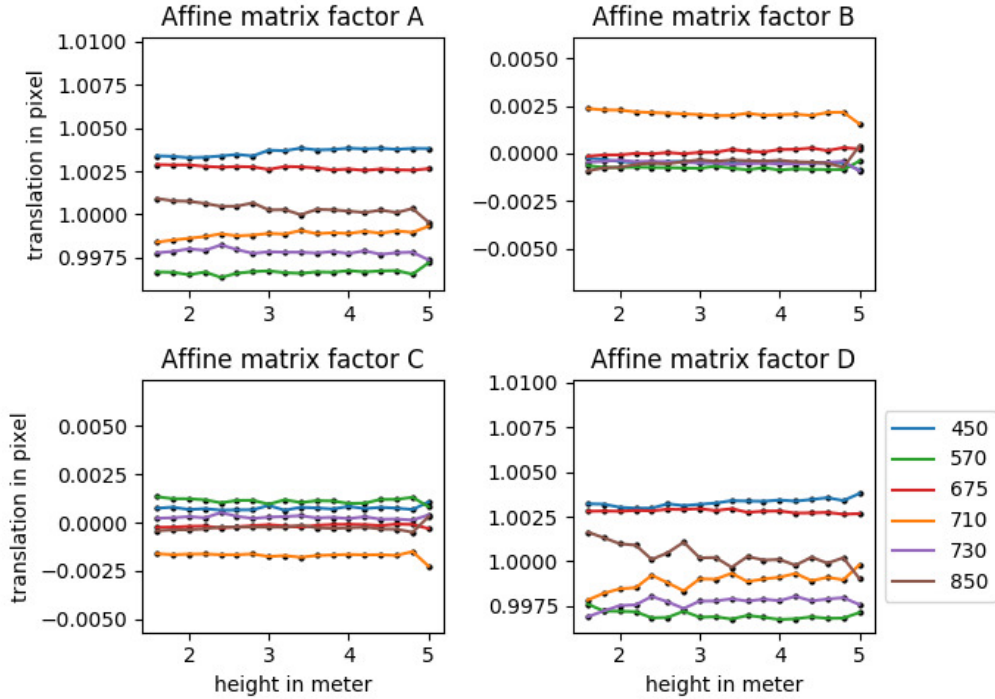


Figure 3.18 – Affine matrix value by height

After the affine correction, the remaining residual distances have been extracted, it is computed using the detected, filtered and matched key-point to the reference spectral band, figure 3.21 (up) shows an example using 570 nm as reference before the perspective correction. The remaining distance between each spectral band to the reference varies according to the distance between the real height and the nearest selected (through linear model). Remember that a bias of $\pm 10\text{cm}$ was initially set to show the error in the worst case, so the difference of errors between each of them are due to the difference of sensors position in the array to the reference and the provided approximate height.

3.5.3.2 Perspective correction

The figures 3.19 shows the numbers of key-points after filtering and homographic association (minimum of all matches) as well as the computation time and performance ratio (matches/time) for each method. The performance ratio is used to compare methods between them, bigger he is, greater is the method (balanced between time and accuracy), making lower of them unsuitable.

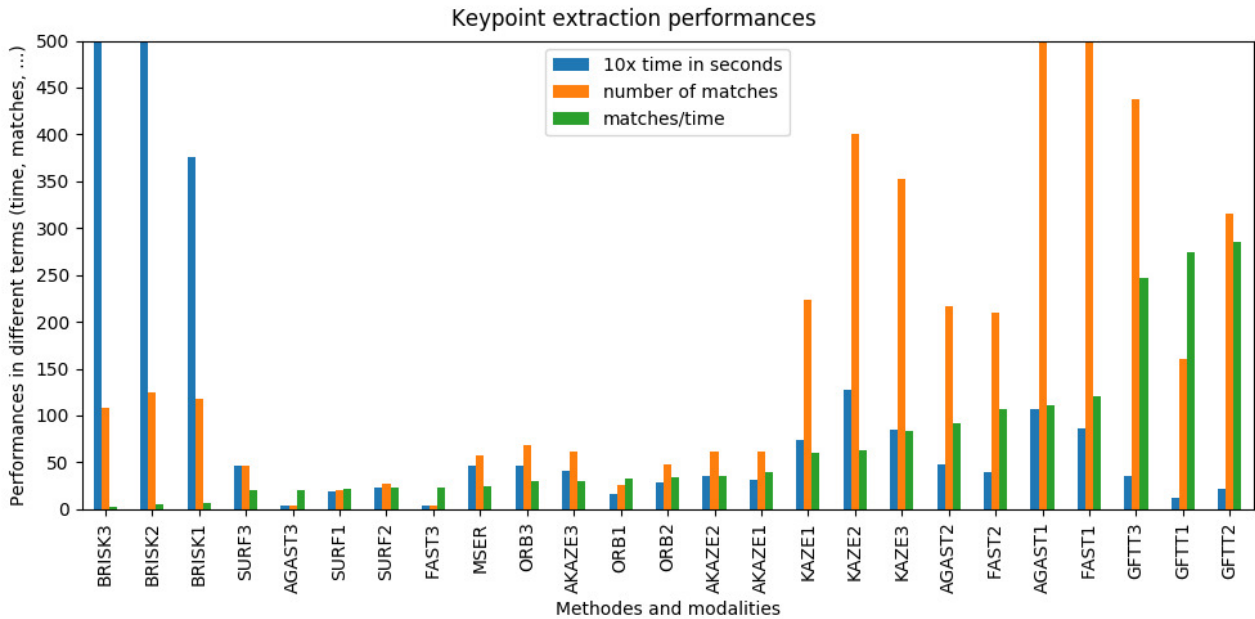


Figure 3.19 – features extractor performances after filtering and homography association

All these methods offer interesting results, the choice of method depends on application needs between computation time and accuracy, three methods stand out in all of these modalities :

- GFTT shows the best overall performance in both computation time and number of matches
- FAST and AGAST1 are quite suitable too, with acceptable computation time and greater matches performances.

The other ones did not show improvement in term of time or matches (especially compared to GFTT), some of them show a small number of matches which can be too small to ensure the precision. Increasing the number of key-points matched allows a slightly higher accuracy [Dantas Dias Junior et al., 2019]. For example, switching from SURF (30 results) to FAST (130 results) reduces the final residual distances from ≈ 1.2 to ≈ 0.9 px but increases the calculation time from ≈ 5 to ≈ 8 seconds.

All methods show that the best spectral band is 710 nm (in red), with an exception for SURF and GFTT which is 570 nm. The figure 3.20 shows the minimum number of matches between each reference spectrum and all the others, for each relevant methods and modalities (KAZE, AGAST, FAST GFTT). Choosing the right spectral reference is important, as we can see, no correspondence is found in some cases between 675-850 nm, but correspondences are found between 675-710 nm and 710-850 nm, making the 710 nm more appropriate, the same behavior can be observed for the other bands and 570 nm as the more appropriate one. This is visible on the figure for all methods, 570 nm and 710 nm have the best minimum number of matches where all the other are quite small.

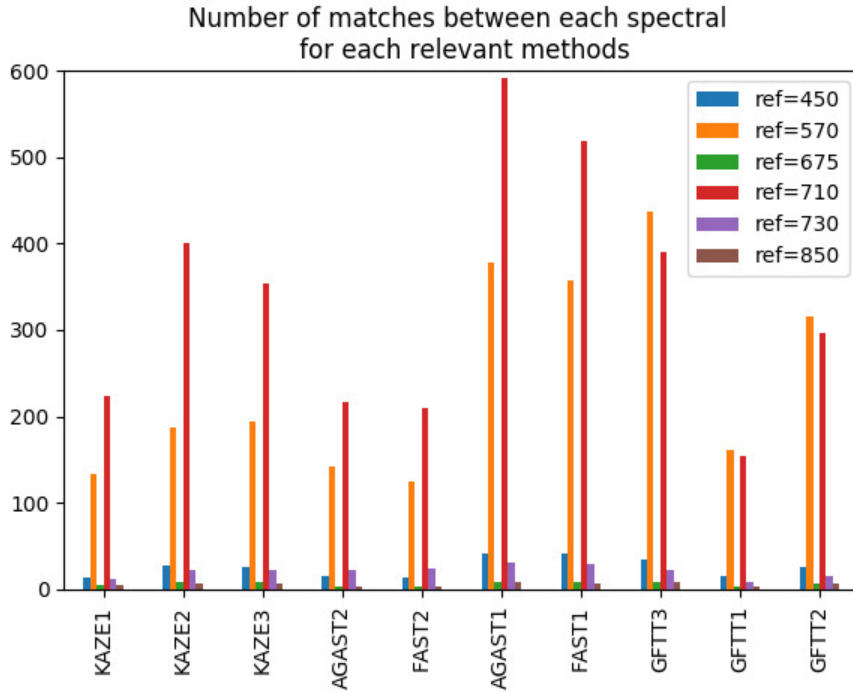


Figure 3.20 – key-point extractor performances

Residues of the perspective correction show that we have correctly registered each spectral band, the figure 3.21 (down) shows the residual distance at different ground distances. In comparison the affine correction error are between $[1.0 - 4.8]$ px where the combination of affine and perspective correction the residual error are between $[0.7 - 1.0]$ px. On average the perspective correction enhance the residual error by $(3.5 - 0.9)/3.5 \approx 74\%$.

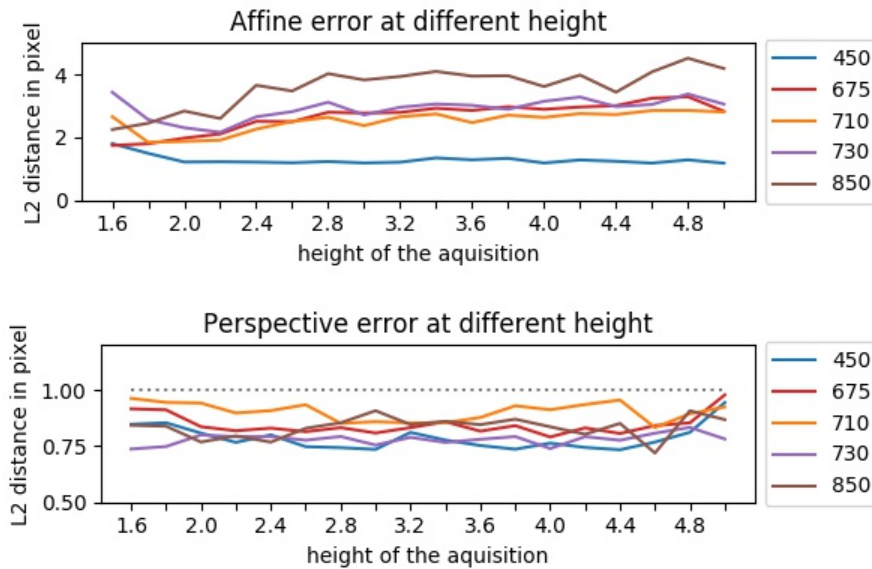


Figure 3.21 – (up) The mean distance of detected key-point before perspective correction with 570 nm as spectral reference (down) Perspective re-projection error with GFTT using the first modality and 570 nm as reference

3.5.3.3 General discussion

Even if the relief of the scene is not taken into account due to the used deformation model, in our case, with flat ground, no difference arise. However, more complex deformation models [Lombaert et al., 2012, Bookstein, 1989] could be used to improve the remaining error. But could also, in some case, create large angular deformations caused by the proximity of key-points, of course, it's possible to filter these key-points, which would also reduce the overall accuracy.

Further research can be performed on each parameter of the feature extractors, for those who need specific performance (time/precision), we invite anyone to download the dataset and test various combinations. Otherwise feature matching can be optimized, at this stage, we use brute-force matching with post filtering, but a different implementation that fulfill our spatial properties should greatly enhance the number of matches by reducing false positives.

3.5.4 Conclusion

In this work, the application of different techniques for multi-spectral image registration was explored using the Airphen camera. We have tested nine type of key-points extractor (ORB, GFTT, AGAST, FAST, AKAZE, KAZE, BRISK, SURF, MSER) at different heights and the number of control points obtained. As seen in the method, the most suitable method is the GFTT (regardless of modalities 1, 2 or 3) with a significant number of matches 150 – 450 and a reasonable calculation time 1.17 s to 3.55 s depending on the modality.

Furthermore, the best spectral reference was defined for each method, for example 570 nm for GFTT. We have observed a residual error of less than 1 px, supposedly caused by the difference of sensors nature (spectral range, lens).

Acknowledgments

The authors acknowledge support from European Union through the project H2020 IWM-PRAISE³ (Integrated Weed Management : PRActical Implementation and Solutions for Europe) and from ANR Challenge ROSE through the project ROSEAU⁴ (RObotics SEnsorimotor loops to weed AUtonomously).

We would like to thanks Combaluzier Quentin, Michon Nicolas, Savi Romain and Masson Jean-Benoit for the realization of the metallic gantry that help us positioning the camera at different heights.

3. <https://iwmpraise.eu/>

4. <http://challenge-rose.fr/en/projet/roseau-2/>

3.6 Conclusion de chapitre

La partie pré-traitement de notre chaîne fonctionnelle correspond à trois étapes. Premièrement, la calibration en hauteur de la caméra a permis d'obtenir la zone que nous souhaitons observer. Deuxièmement, les corrections géométriques des lentilles nous ont permis de limiter les erreurs entre les capteurs. Et troisièmement, le recalage des images multispectrales a permis de superposer les différentes images issues des six capteurs de la caméra Airphen.

Les deux premières corrections ont été présentées durant le chapitre précédent. Tandis que ce chapitre était dédié au recalage des images multispectrales. Cette étude a fait l'objet d'un article de conférence (VISAPP), présenté en Février 2020 à Malte (Rép. de Malte). La méthode proposée offre les performances nécessaires à l'exploitation des images multispectrales acquises en proxidtection, durant nos expérimentations en micro-parcelles et pendant les challenges RoSE. Toutes les images multispectrales acquises avec la caméra Airphen ont donc été transformées à l'aide de cette méthode avant leur annotation et utilisation.

Ce chapitre marque également la fin de la partie pré-traitement des données. Maintenant, que nous sommes capables d'exploiter les données images, nous allons nous intéresser à leur utilisation. La seconde étape de notre chaîne fonctionnelle, vise à détecter les surfaces et objets visibles d'une scène agronomique. L'approche traditionnelle s'appuie sur l'utilisation d'un indice de végétation sur lequel est appliqué une méthode de segmentation, afin d'obtenir les éléments à discriminer. Cette approche a été conservée, mais de nouveaux indices basés sur l'apprentissage profond sont présentés dans le chapitre suivant 4. Concernant la segmentation des objets, une nouvelle méthode a également été proposée. Elle concerne spécifiquement la détection des feuilles. Elle est présentée dans le chapitre 5. Pour rappel, la figure ci-dessous illustre notre chaîne fonctionnelle et où nous nous situons.



Rappel de l'architecture de notre chaîne fonctionnelle
ici la première étape en vert est terminée et en jaune l'étape suivante étudiée

4 Indices de végétation

Dans le chapitre précédent, nous avons vu comment pré-traiter les données en vue de leur exploitation. La seconde étape de notre chaîne de traitement vise à répondre à la question suivante : comment discriminer la végétation du reste de la scène ? Pour répondre à cette question, l'approche classique repose sur l'utilisation d'indices de végétation, sur lesquels une méthode de segmentation simple est appliquée, telle qu'un seuillage. Comme nous l'avons vu dans le chapitre 2, la définition de tels indices (parmi lesquels le NDVI est le plus connu) repose essentiellement sur des connaissances empiriques [Bannari et al., 1995] (section 2.2.3). Ces indices utilisent souvent des données calibrées en réflectance, censées garantir leur stabilité.

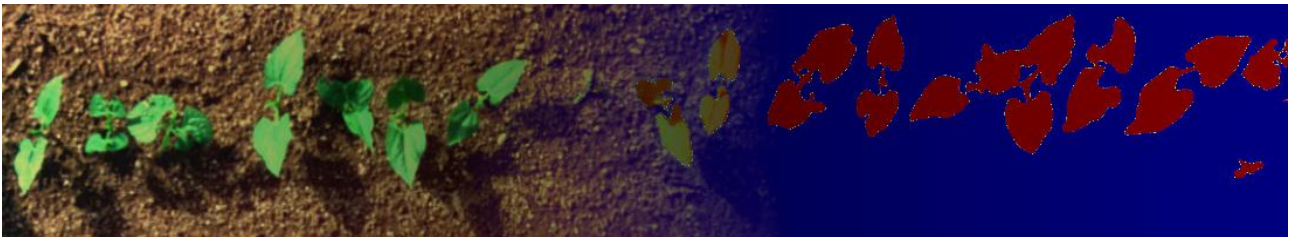
Or de nombreux facteurs influencent de tels indices de télédétection. La calibration radiométrique et le capteur sont deux facteurs importants dans l'estimation de ces indices. [Lee et al., 2021] montre leurs effets au cours de la journée sur le NDVI. Cet indice peut varier d'un facteur de 20 % après correction de ces deux paramètres. D'autres facteurs, comme les conditions atmosphériques, l'humidité de la végétation, l'humidité du sol, la couverture végétale globale, l'ombre et les différences de type de sol, jouent aussi des rôles importants dans l'estimation du couvert végétal. Par exemple, la lumière réfléchiée par le sol et l'ombre obscurcissant partiellement une image, peuvent influencer les valeurs du NDVI dans une large mesure [Zhang et al., 2015], rendant difficile l'utilisation d'un tel indice.

Ces limitations ont été la raison du développement de nombreux indices de végétation, ajustés au sol (par exemple, Soil-Adjusted Vegetation Index, Modified Soil-Adjusted Vegetation Index) ou à l'atmosphère (Atmospherically Resistant Vegetation Index) [Bannari et al., 1995]. Ces modèles, plus robustes, restent néanmoins fortement limités et influencés par tous ces facteurs. L'utilisation d'un seuillage fixe est donc peu pertinente car il faut s'attendre à des valeurs différentes pour une même zone d'observation au fil du temps, en raison des changements environnementaux affectant l'observation (saisonniers, phénologiques, météorologiques, ...).

Pour augmenter la performance de cette étape de traitement, certains utilisent une méthode de classification faisant appel à plusieurs indices usuels [Shishir and Tsuyuzaki, 2018, Gée et al., 2021]. D'autres en revanche, utilisent des algorithmes génétiques afin d'apprendre l'équation idéale permettant une segmentation optimale [Kabiri et al., 2011, Albarracín et al., 2020]. Dans les deux cas, cela revient à prendre en considération une plus large quantité d'information, ce qui augmente le potentiel de discrimination. En revanche, l'optimisation de ces indices utilise des méthodes, soit coûteuses en temps, soit ne garantissant pas un optimum, les rendant peu applicables à de larges bases de données.

Face à ce constat, une nouvelle méthodologie est proposée dans ce chapitre pour l'apprentissage de tels indices. Elle repose sur l'apprentissage d'approximateurs de fonctions. Ainsi, on cherche une fonction qui peut décrire toutes les autres. Cette approche peut être mise en œuvre en apprentissage profond. Ce qui permet d'exploiter le processeur graphique, permettant une évaluation rapide de larges bases de données et une optimisation d'équations (ou de réseau de neurones) par le mécanisme de rétropropagation. L'apprentissage profond peut-être vu comme une méthode de régression générique. Les indices qui ont été définis permettent de s'abstraire des conditions d'acquisition et de la calibration.

DeepIndices : A new approach to remote sensing indices based on optimization and approximation of functions through deep-learning, with application to vegetation surfaces over uncalibrated data



Authors Jehan-Antoine Vayssade, Jean-Noël Paoli, Christelle Gée and Gawain Jones

Publication 09/06/2021 – Remote Sensing Journal – Volume 13 Issue 12 Pages 2261

Abstract The form of a remote sensing index is generally empirically defined, whether by choosing specific reflectance bands, equation forms or its coefficients. These spectral indices are used as preprocessing stage before object detection/classification. But no study seems to search for the best form through function approximation in order to optimize the classification and/or segmentation. The objective of this study is to develop a method to find the optimal index, using a statistical approach by gradient descent on different forms of generic equations. From six wavebands images, five equations have been tested, namely : linear, linear ratio, polynomial, universal function approximator and dense morphological. Few techniques in signal processing and image analysis are also deployed within a deep-learning framework. Performances of standard indices and DeepIndices were evaluated using two metrics, the dice (similar to f1-score) and the mean intersection over union (mIoU) scores. The study focuses on a specific multispectral camera used in near-field acquisition of soil and vegetation surfaces. These DeepIndices are built and compared to 89 common vegetation indices using the same vegetation dataset and metrics. As an illustration the most used index for vegetation, NDVI (Normalized Difference Vegetation Indices) offers a mIoU score of 63.98 % whereas our best models gives an analytic solution to reconstruct an index with a mIoU of 82.19 %. This difference is significant enough to improve the segmentation and robustness of the index from various external factors, as well as the shape of detected elements.

Keywords image ; precision agriculture ; spectral indices ; multi-spectral ; deep-learning ; vegetation segmentation

4.1 Introduction

An important advance in the field of earth observation is the discovery of spectral indices, they have proved their effectiveness in surface description. Several studies have been conducted using remote sensing indices, often applied to a specific field of study like evaluations of vegetation cover, vigor, or growth dynamics [Jinru and Su, 2017, Mezera Jiří, 2019, Tanriverdi et al., 2020, Elbeltagi et al., 2021] for precision agriculture using multi-spectral sensors. Some spectral indices have been developed using RGB or HSV color space to detect vegetation from ground cameras [Lee et al., 2020, Milioto et al., 2017, Hassanein et al., 2018]. Remote sensing indices can also be used for other surfaces analysis like water, road, snow [Dixit et al., 2019] cloud [Zhai et al., 2018] or shadow [Henrich et al., 2009].

There are two main problems with these indices. Firstly they are almost all empirically defined, although the selection of wavelengths comes from observation, like NDVI for vegetation indices. It is possible to obtain better spectral combinations or equations to characterize a surface with specific acquisitions parameters. It is important to optimize upstream the index, as the data transformation leads to a loss of essential information and features for classification [Zhang et al., 2015]. Most studies have tried to optimize some parameters of existing indices. For example, an optimization of NDVI $(NIR - Red)/(NIR + Red)$ was proposed by [Gitelson, 2004] under the name of WDRVI (Wide Dynamic Range Vegetation Index) $(\alpha NIR - Red)/(\alpha NIR + Red)$. The author tested different values for α between 0 and 1. The ROC curve was used to determine the best coefficient for a given ground truth. Another optimized NDVI was designed and named EVI (Enhanced Vegetation Index). It takes into account the blue band for atmospheric resistance by including various parameters $G(NIR - Red)/(NIR + C_1 Red - C_2 Blue + L)$, where G, L are respectively the gain factor and the canopy background adjustment, in addition the coefficients C_1, C_2 are used to compensate for the influence of clouds and shadows. Many other indices can be found in an online database of indices (IDB : www.indexdatabase.de accessed August 10, 2019) [Henrich et al., 2009] including the choice of wavelengths and coefficients depending on the selected sensors or applications. But none of the presented indices are properly optimized. Thus, in the standard approach, the best index is determined by testing all available indices against the spectral bands of the selected sensor with a Pearson correlation between these indices and a ground truth [Liu et al., 2017, Kokhan and Vostokov, 2020]. Furthermore, correlation is not the best estimator because it neither considers the class ratio nor the shape of the obtained segmentation and may again result in a non-optimal solution for a specific segmentation task. Finally, these indices are generally not robust because they are still very sensitive to shadows [Zhang et al., 2015]. For vegetation, until recently, all of the referenced popular indices were man-made and used few spectral bands (usually NIR, red, sometimes blue or RedEdge).

The second problem with standard indices is that they work with reflectance-calibrated data. Three calibration methods can be used in proximal-sensing. (i) The first method use an image taken before acquisition containing a color patch as a reference [Yahui et al., 2019, Minařík et al., 2019], and is used for correction. The problem with this approach is that if the image is partially shaded, the calibration is only relevant on the non-shaded part. Moreover ideally the reference must be updated to reduce the interference of weather change on the spectrum measurement, which is not always possible since it's a human task. (ii) An other method is the use of an attached sunshine sensor [Gilliot et al., 2018], which also requires calibration but does not allow to correct a partially shaded image. (iii) The last method is the use of a controlled lighting environment [Chebrolu et al., 2017, Wu et al., 2019], e.g., natural light is suppressed by a curtain and replaced by artificial lighting. All of these approaches are sometimes difficult to implement for automatic, outdoor use, and moreover in real time like detecting vegetation while a tractor is driving through a crop field.

In recent years, machine learning algorithms have been increasingly used to improve the definition of presented indices in the first main problem. Some studies favor the use of multiple indices and advanced classification techniques (RandomForest, Boosting, DecisionTree, etc.) [Oldeland et al., 2010, Peña-Barragán et al., 2011, Nguy-Robertson et al., 2012, Shishir and Tsuyuzaki, 2018, Elbeltagi et al., 2021, Lu et al., 2021]. Another study has proposed to optimize the weights in an NDVI equation form based on a genetic algorithm [Kabiri et al., 2011] but does not optimize the equation forms. An other approach has been proposed to automatically construct a vegetation index using a genetic algorithm [Albarracín et al., 2020]. They optimize the equation forms by building a set of arithmetic graphs with mutations, crossovers and replications to change the shape of each equation during learning but it does not take into account the weights, since it's use calibrated data. Finally, with the emergence of deep learning, current studies try to adapt popular CNN architectures (UNet, AlexNet, etc.) to earth observation applications : [Lv et al., 2018, Gaetano et al., 2018, Fu et al., 2018, Kussul et al., 2017].

However there is no study that optimize both the equation forms and spectral bands weights. The present study explicitly optimize both of them by looking for a form of remote sensing indices by learning weights in functions approximators. These functions approximators will then reconstruct any equation forms of the desired remote sensing index for a given acquisition system. To solve the presented second problem, this study evaluates the functions approximators on an uncalibrated dataset containing various acquisition conditions. This is not a common approach but can be found in the literature [Bajwa and Tian, 2002, Bareth et al., 2016]. This will lead to creating indices that do not require data calibration. The deep learning framework has been used as a general regression toolkit. Thus, several CNN function approximators architectures are proposed. DeepIndices is presented as a regression problem, which is totally new, as is the use of signal and image processing.

4.2 Material and Data

4.2.1 Instrument Details

The images were acquired with the Airphen (Hyphen, Avignon, France) six-band multi-spectral camera (Figure 4.1). This is a multi-spectral scientific camera developed by agronomists for agricultural applications. It can be integrated into different types of platforms such as drones, phenotyping robots, etc.



Figure 4.1 – AIRPHEN camera composed of 6 sensors

The camera has been configured using the 450/570/675/710/730/850 nm bands with a 10 nm FWHM respectively denoted from λ_0 to λ_5 . These spectral bands have been defined by a previous study [Louargant et al., 2017] for crop/weed discrimination. The focal length of each lens is 8 mm. The raw resolutions for each spectral band is 1280×960 px with 12 bit precision. Finally, the camera is equipped with an internal GPS antenna.

4.2.2 Image Dataset

The dataset were acquired on the site of INRAe in Montoldre (Allier, France) within the framework of the “RoSE challenge” founded by the French National Research Agency (ANR) and in Dijon (Burgundy, France) within the site of AgroSup Dijon. Images of bean and corn, containing various natural weeds (yarrows, amaranth, geranium, plantago, etc) and sowed ones (mustards, goosefoots, mayweed and ryegrass) with very distinct characteristics in terms of illumination (shadow, morning, evening, full sun, cloudy, rain, etc) were acquired in top-down view at 1.8 m from the ground. The Table 4.1 synthesis the dataset.

Source	Year	Corn	Bean	Illumination
Dijon	2019	-	9	full sun, evening
Montoldre	2019	20	22	shadow, sunny, cloudy
Montoldre	2020	18	22	morning, cloudy, rainy
Total		38	53	=91

Table 4.1 – Acquisition sources and global illumination.

Manual annotation takes about 4 h per image to obtain the best quality of ground truth, which is necessary for use in regression algorithms. Thus the ground truth size is small and defined with very distinctive illumination condition. To simulate light variations effect on the ground truth images a random brightness (20%) and a random saturation (5%) are added to each spectral band during the training phase. As illustration the Figure 4.2 shows a false color reconstruction of corn crop in the field with various weeds and shadows on the corners of the image (not vignetting).

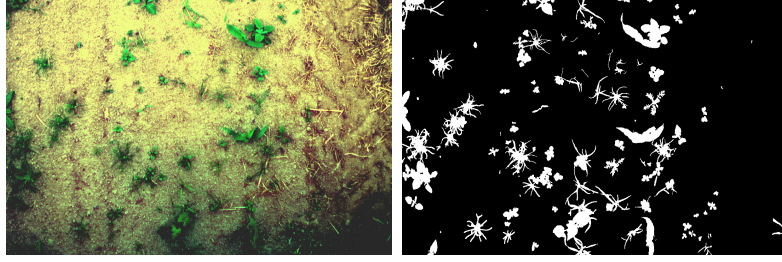


Figure 4.2 – False color in the left and corresponding manual ground truth on the right.

4.2.3 Data Pre-Processing

4.2.3.1 Images Registration

Due to the nature of the camera (Figure 4.1), a spectral band registration is required and performed with a registration method based on previous work [Vayssade et al., 2020] (with a sub-pixel registration accuracy). The alignment is refined in two steps, with (i) a rough estimation of the affine correction and (ii) a perspective correction for the refinement and accuracy through the detection and matching of key points. The result shows that *GoodFeatureToTrack* (GFTT) algorithm is the best key-point detector considering the λ_{570} nm band as spectral reference for the registration. After the registration, all spectral images are cropped to 1200×800 px and concatenated to channel-wise denoted λ where each dimension denoted λ_d refers to each of the six spectral bands.

4.2.3.2 Images Normalization

Spectral bands inherently have a high noise associated with the CCD sensor, which is a potential problem during normalization [Khanna et al., 2017]. To overcome this effect, 1% of the minimum and maximum signal is suppressed by calculating the quantiles, the signal is clipped on the given range and each band is rescaled in the interval $[0, 1]$ using min-max normalization to obtain ρ_d :

$$\rho_d = 0 \leq \frac{\lambda_d - \min(\lambda_d)}{\max(\lambda_d) - \min(\lambda_d)} \leq 1 \quad (4.1)$$

The method also reduces the lighting variation. According to [Blackburn and Vignola, 2012] a little variation is observed in the spectral correction factors between clear and cloudy days. Thus, the correction has a limited impact on the scaling factor and should be managed by this equation. However, the displacement factor could not be estimated, thus the output images are not calibrated in reflectance.

4.2.3.3 Enriching Information

In order to enrich the pool of information, some spectral band transformations are added, which allow to take into account spatial gradients and spectral mixing [Milioto et al., 2017] in the image. The choice is oriented towards seven important information in different respects :

The standard deviation between spectral band, noted ρ_{std} can help to detect the spectral mixture. For example between two different surface like ground and leave which have opposite spectral radiance the spectral mixing make a pixel with linear combination, thus the standard deviation tend to zero [Louargant et al., 2017]. Three Gaussian derivatives on different orientations are computed G_{xx} , G_{xy} and G_{yy} over the standard deviation ρ_{std} which give an important spatial information about the gradients breaks corresponding to the outer limits of

surfaces. These Gaussian derivatives are computed with a fixed $\text{Sigma} = 1$. The Laplacian computed over the standard deviation ρ_{std} , the minimum and maximum eigenvalues of the Hessian matrix (obtained by Gaussian derivation G_{xx} , G_{xy} and G_{yy}), also called ridge are included. These transformations could improve the detection of fine elements [Lin et al., 2014] such as monocotyledons for vegetation images.

All these transformations are concatenated to the channel-wise normalized spectral band input and build the final input image. In total seven transformations are added to the six spectral images for a final image of 13 channels, which will probably help the convergence.

4.2.4 Training and Validation Datasets

The input dataset is composed by spectral images \mathbf{I} of size $1200 \times 800 \times 13$ (or 6 if the “Enriching information” part is disabled) and a manual ground truth \mathbf{p} of size $1200 \times 800 \times 1$ where $p \in \{0, 1\}$. The desired output \hat{p} is a probability vegetation map of size $1200 \times 800 \times 1$ where $\hat{p} \in [0, 1]$. This input dataset is randomly split into two sub-sets respectively training (80%) and validation (20%). All random seed is fixed at the start-up to keep the same training/validation dataset across all trained models which help to compare them. Keeping the same random seed also results in the same starting point between different new runs, making results reproducible on the same hardware.

4.3 Methodology

4.3.1 Existing Spectral Indices

From the indices database, 89 vegetation indices have been identified (Table 4.2) as compatible with the wavelengths used in this study (as near as possible), they will be tested and compared to the designed DeepIndices. Five forms of simple equations have been extracted from this database (a wide variety of indices are derived from these forms, generally a combination of 2 or 3 bands) :

$$\text{band reflectance} = \rho_i \quad (4.2)$$

$$\text{two bands difference} = \rho_i - \rho_j \quad (4.3)$$

$$\text{two bands ratio} = \rho_i \div \rho_j \quad (4.4)$$

$$\text{normalized difference two bands} = (\rho_i - \rho_j) \div (\rho_i + \rho_j) \quad (4.5)$$

$$\text{normalized difference three bands} = (2\rho_i - \rho_j - \rho_k) \div (2\rho_i + \rho_j + \rho_k) \quad (4.6)$$

By analyzing these five equations we can synthesize them into two generic equations (Linear combination and Linear ratio) which take into account all spectral bands. Three other models can generalize any function : the polynomial fitting, the continuous function approximations by Taylor development, and the piecewise continuous function approximations through morphological operators. These forms are interesting to optimize because they can approximate any function. This optimization will lead to automatically defining new indices (DeepIndices). The following subsections present these different models.

4.3.2 Deepindices : Baseline Models

4.3.2.1 Linear Combination

To synthesize Equations (4.2) and (4.3), a simple linear equation such as $y = \sum_{d=0}^N \alpha_d \rho_d$ can be defined. This equation can be generalized to the 2D domain using a 2D convolution allowing consider the neighboring pixels. For a pixel at the position $[i, j]$ the convolution is defined by :

$$y[i, j] = \sum_{d=0}^D \sum_{h=0}^N \sum_{w=0}^N \rho_d [i - N/2 + h, j - N/2 + w] * H[h, w, d] \quad (4.7)$$

where H defines neighborhood weights (corresponding to α_i). D is the number of dimensions (6 spectral bands + 5 transformations) and N is the kernel size. The linear combination is given by $N = 1, D = 12$. The kernel weights are initialized by a truncated normal distribution centered on zero [Jang and Son, 2019], weights are updated during the training of the CNN through back-propagation and unnecessary bands should be set to zero. The interesting part is that increasing the kernel size N allows to take into account the neighborhood of a pixel and should estimate more accurately the spectral mixing [Louargant et al., 2017]. Figure 4.3 shows the corresponding network.

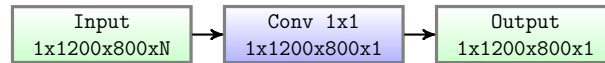


Figure 4.3 – Linear combination model.

4.3.2.2 Linear Ratio

To generalize Equations (4.4)–(4.6), a simple model based on the division of two linear combination is set. In the same way, this form is generalizable to the 2D domain and then corresponds to two 2D convolutions, one for the numerator, the other for the denominator. When the denominator is zero, the result is set to zero as well, to leverage the “not a number” output. The Figure 4.4 shows the corresponding network.

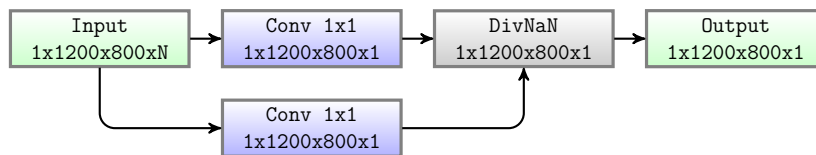


Figure 4.4 – Linear ratio model.

4.3.2.3 Polynomial

According to the Stone-Weierstrass theorem any continuous function defined on a segment can be uniformly approximated by a polynomial function. Thus all forms of color indices can be approximated by a polynomial $y = \sum_{d=0}^N \alpha_d \rho_d^{\delta_d}$ of degree N . Setting the degree is a difficult task which may imply under-fitting or over-fitting. In addition un-stability can be caused by near-zero δ_d . But since the segment is restricted to the domain $[0, 1]$ the Bernstein polynomials are a common demonstration and the equation can be written as a weighted sum of Bernstein basis polynomials $B_{N,i} = (1 - \rho)^i \rho^{N-i}$ which are more stable during the training. Moreover Bernstein Neural Network can solve partially differentiable equations [Sun et al., 2018]. For implementation reasons, two different layers are defined in the network (visible in the Figure

4.5). One for the Bernstein expansion limited to $B_{11,11}$ which takes the input image and produces different Bernstein basis polynomial, then each Bernstein basis is concatenated to the channel-wise and the linear combination is defined by a 2D convolution.

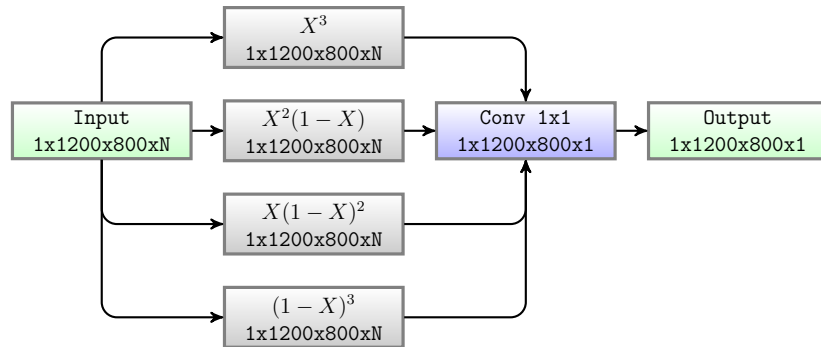


Figure 4.5 – Polynomial model with Bernstein expansions between $B_{4,1}$ and $B_{4,4}$.

4.3.2.4 Universal Function Approximation

The Gaussian color space model proposed by [Geusebroek et al., 2000] shows that the spatio-spectral energy distribution of the incident light E is the weighted integration of the spectrum ρ_d denoted $E(\rho_d)$. Where E can be described as a Taylor series and the energy function is convolved by different derivatives of a Gaussian kernel or structured receptive fields [Jacobsen et al., 2016]. This important point shows that Taylor expansions can decompose any function $f(x)$, especially for color decomposition and remapping, into :

$$f(x) = f(0) + f'(x)x + \frac{1}{2!}f''(x)x^2 + \frac{1}{3!}f'''(x)x^3 + o(x^3) \quad (4.8)$$

Here, the signature of the incident energies distribution of a remote sensing index associated to a surface can be reconstructed. An approach to learn this form of development is proposed by [Huang et al., 2016] which is commonly called DenseNet and then corresponds to the sum of the concatenation of the signal and these spatio-spectral derivatives

$$\mathbf{x} \rightarrow [\mathbf{x}, f_1(\mathbf{x}), f_2(\mathbf{x}, f_1(\mathbf{x})), \dots] \quad (4.9)$$

Various convolutions allow to learn receptive fields and derivatives in spectral domain when the kernel size k is 1, and in spatio-spectral domain when k is higher. Batch-Normalization are used to reduce the covariate shift across convolution output by re-scaling it and speed up the convergence. Finally the Sigmoid activation function is used and defined by

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (4.10)$$

Sigmoid function allows to learn more complex structures and non-linearity of the reconstructed function. The number of derivative and receptive field are configurable with two parameters. The *depth* which corresponds to the number of layers in the network. And the *width* which refers to the number of outputs for each convolution. By default, the *depth* is fixed to 3 and the *width* is fixed to 5. The Figure 4.6 shows the corresponding universal function approximator network.

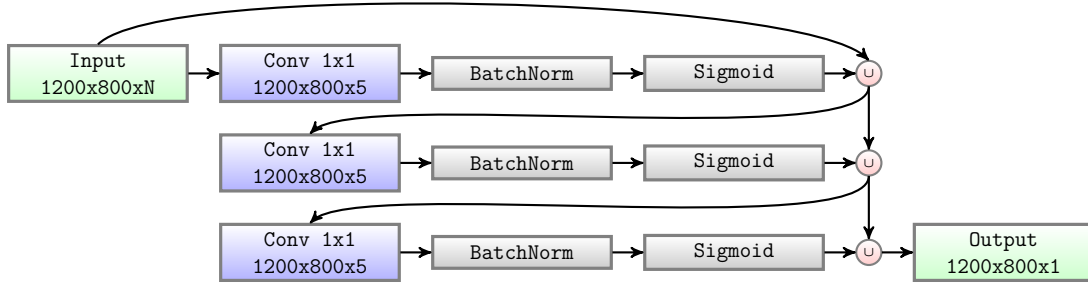


Figure 4.6 – Universal function approximation model (depth = 3, width = 5).

4.3.2.5 Dense Morphological Function Approximation

As for the Taylor series, an approximation of any piecewise continuous function can be established by morphological operators such as dilatation and erosion [Mondal et al., 2019], respectively denoted $\rho \oplus s$ and $\rho \ominus s$ where s are the corresponding erosion or dilatation coefficients. Several erosion and dilatation are defined for each spectral band i , then the expanded layer is defined as the channel concatenation of z_i^+ and in the same way for the erosion layer via z_i^- . Both are defined by

$$z_i^+ = \rho \oplus s_i = \max_k(\rho_k - s_{k,i}, 0) \quad (4.11)$$

$$z_i^- = \rho \ominus s_i = \max_k(s_{k,i} - \rho_k, 0) \quad (4.12)$$

To obtain the output $I = \sum_{i=0}^N z_i^+ w_i^+ + \sum_{i=0}^N z_i^- w_i^-$ of which the w_i^+ and the w_i^- are the linear combination coefficients obtained by a 2D convolution. We chose to set the number of dilatation and erosion neurons at 6. The Figure 4.7 shows the corresponding network.

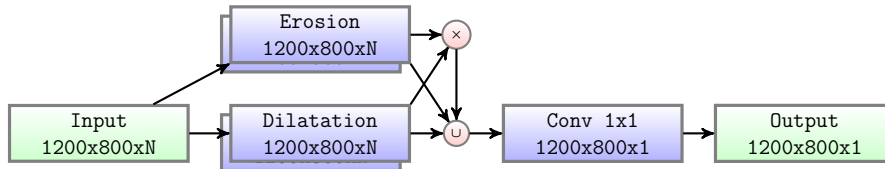


Figure 4.7 – Dense-morphological model.

4.3.3 Enhancing Baseline Models

4.3.3.1 Input Band Filter (IBF)

To remove parts of the signal that may be dispensable, the addition of a low-pass, high-pass and band-pass filter upstream of the network are studied. A good example is provided by vegetation indices, only the high values in the green and near infra-red, and the low values in the red and blue characterize the vegetation.

This is the principle of the NDVI index. Due to the internal structure, the leaves reflect a lot of light in the near infrared, which is in sharp contrast to most non-vegetable surfaces. When the plant is dehydrated or stressed, the spongy layer collapse and the leaves reflect less light in the near-infrared, reaching red values in the visible range [Joshi et al., 2020]. Thus, the mathematical combination of these two signals can help to differentiate plants from non-plant objects and healthy plants from diseased plants. However, this index is then less interesting when detecting only vegetation and is strongly influenced by shade or heat.

We will therefore add a filter in the previous equations to remove undesirable spectral energies of each ρ_d by using two thresholds a and b, which will also be learned. If it turns out that the whole signal is interesting, these two parameters will not change and their values will be a=0 and b=1. To apply the low-pass filter the equation $z = \max(\rho - a, 0) \div (1 - a)$ is used and thus allows to suppress low values. For the high-pass filter the equation $w = \max(b - \rho, 0) \div b$ is applied to suppress high values. The band-pass filter it's the product of low and high-pass filters $y = z * x$. The output layer is the concatenation in the channel-wise of the input images, the low-pass, the high-pass and the band-pass filter which produce $4 \times 13 = 52$ channels. Finally to reduce the output data for the rest of the network, a bottleneck is inserted using a convolution layer, and generate a new image with 6 channels. This image is used by the rest of the network defined previously in Section 4.3.2. The Figure 4.8 shows the corresponding module inserted upstream of the network.

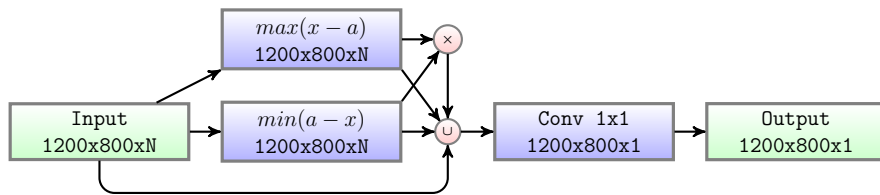


Figure 4.8 – Input Band Filter inserted at the beginning of the model.

4.3.3.2 Spatial Pyramid Refinement Block (SPRB)

To take into account different scales in the image, the addition of a “Spatial Pyramid Refinement Block” at the downstream part of the network is studied. [Liu et al., 2015] showed that fusing the low to high-level features improved the segmentation task. It consists in the sum of different 2D convolutions whose core sizes have been set to 3, 5, 7 and 9. The results of all convolutions are concatenated and the final image output is given by a 2D convolution. The Figure 4.9 shows the corresponding module inserted downstream of the network.

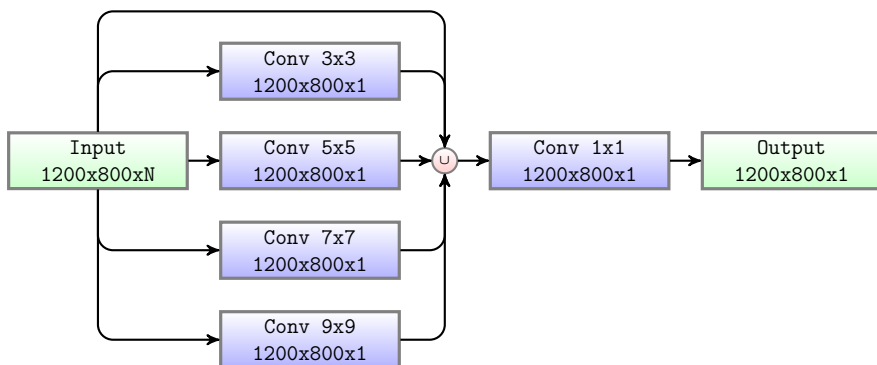


Figure 4.9 – Spatial refinement block inserted at the end of a model.

4.3.4 Last Activation Function

To obtain an index and facilitate convergence, we will only be interested in the values between 0 and 1 at the output of the last layer with the help of an activation function of type clipped ReLU defined by

$$\text{ClippedReLU}(x) = \begin{cases} 1 & \text{if } x > 1 \\ x & \text{if } 0 < x < 1 \\ 0 & \text{if } x < 0 \end{cases} \quad (4.13)$$

where x is a pixel of the output image. Each negative or null pixel will then be the unwanted class, greater or equal to 1 will be the searched class. The indecision border is the values between 0 and 1 which will be optimized. And then correspond to the probability that the pixel is the searched surface $P(Y = 1)$ or not $P(Y = 0)$. This is valid for the output prediction denoted $\hat{p} \in [0, 1]$ and the ground truth denoted $p \in \{0, 1\}$.

4.3.5 Loss Function

A wide variety of loss functions have been developed during the emergence of deep-learning (MSE, MAE, Hinge, Tversky, etc). A cross-entropy loss function is usually used when optimizing binary classification [Bokhovkin and Burnaev, 2019]. This loss function is not optimized for the shape. Recently, with deep neural network and for semantic segmentation [Rahman and Wang, 2016] has proposed a solution to optimize an approximation of the mean intersection over union (mIoU) and defined by

$$\text{mIoU_Loss} = 1 - \frac{p\hat{p}}{p + \hat{p} - p\hat{p}} \quad (4.14)$$

The performance of this loss function seems more efficient than previous methods [Zhou et al., 2019, van Beers et al., 2019, Jadon, 2020]. We will then use it as a loss function.

4.3.6 Performance Evaluation

Commonly, accuracy and Pearson correlation are used to quantify the performance of remote sensing indices [Liu et al., 2017, Kokhan and Vostokov, 2020]. However this type of metrics does not take into account either the class ratio nor the shape of the segmentation. Correlation is also highly sensitive to non-linear relationship, noise, subgroups and outliers [Aggarwal and Ranganathan, 2016, Armstrong, 2019] making incorrect evaluation. According to [Shamir et al., 2019, Choi et al., 2019], the dice score and the mean intersection over union (mIoU) are more adapted to evaluate the segmentation mask. Defined by :

$$\text{Dice} = \frac{2p\hat{p}}{p + \hat{p}} \quad (4.15) \quad \text{mIoU} = \frac{p\hat{p}}{p + \hat{p} - p\hat{p}} \quad (4.16)$$

We will then used these two metrics for the performance evaluation. Prior to quantization, a threshold of 0.5 is applied to the output of the network to transform the probability into a segmentation mask. When \hat{p} is lower than 0.5, it is considered as the background, otherwise it is considered as the object mask we are looking for. Other metrics are not considered because they are not always appropriate in case of segmentation or use in unbalanced data.

4.3.7 Comparison with Standard Indices

In order to make a fair comparison it is necessary to optimize each standard index. A minimal neural network is used to learn a linear regression. The network is thus composed of the spectral index, followed by a normalization $x = (x - \min)/(\min - \max)$, then a 2D convolution with a kernel size of $k = 1$ is used for the linear regression. To perform the classification in the same way as our method, a ClippedReLU activation function is used. This tiny network is presented in the next Figure 4.10. Obviously the same metrics and loss function are used.

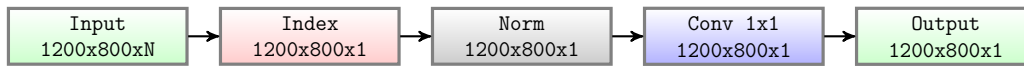


Figure 4.10 – Optimized model for standard indices

4.3.8 Training Setup

The training is done through Keras module within Tensorflow 2.2.0 framework. All computation is done on an NVidia GTX 1080 which have 8111MiB of memory, this limits the number of simultaneous layers on the memory and so the size of the model. Each model is compiled with Adam optimizer. This optimization algorithm is primarily used with lookahead mechanism proposed by [Zhang et al., 2019a]. It iteratively updates two sets of weights : the search directions for the **fast weights** are chosen by inner optimizer, while the **slow weights** are updated every k steps based on the direction of the **fast weights** and the two sets of weights are synchronized. This method improves the learning stability and lowers the covariance of its inner optimizer. The initial learning rate is fixed to 2^{-3} . Batch size is fixed to 1 due to memory limitation. And the learning rate is decreased using *ReduceLROnPlateau* with $factor = 0.2, patience = 5, min_lr = 2e^{-6}$. The training is done through 300 iterations. Finally an *EarlyStopping* callback is used to stop the training when there is no improvement in the training loss after 50 consecutive epochs.

4.4 Results and Discussion

4.4.1 Fixed Models

All standard vegetation models have been optimized using the same training and validation datasets. Each of them has been optimized using a min-max normalization followed by a single 1×1 2D convolution layer and a last clipped ReLU activation function is used like the generic models implemented. The top nine standard indices are presented in Table 4.2. Their respective equations are available in Table 4.8 in Appendix 4.5.

Standard Index	Used ρ	mIoU	Dice
Modified Triangular Vegetation Index 1	3	73.71	83.23
Modified Chlorophyll Absorption In Reflectance Index 1	3	73.68	83.22
Enhanced Vegetation Index 2	2	67.94	79.20
Soil Adjusted Vegetation Index	2	67.28	78.65
Soil And Atmospherically Resistant VI 3	2	65.86	77.61
Enhanced Vegetation Index 3	2	65.05	77.07
Global Environment Monitoring Index	2	65.04	77.01
Adjusted Transformed Soil Adjusted VI	3	64.96	77.00
NDVI	2	63.98	75.97

Table 4.2 – Synthesized standard indices performances : the nine best models are presented.

It is interesting to note that most of them are very similar to NDVI indices in their form. This shows that according to all previous studies, these forms based on a ratio of linear combination are the most stable against light variation. For example the following NDVI based indices are tested and show very different performances, highlighting the importance of weight optimization :

$$\begin{aligned}
 \text{NDVI} &= (\rho_5 - \rho_2) \div (\rho_5 + \rho_2) \\
 \text{EnhancedVegetationIndex} &= 2.5 * (\rho_5 - \rho_2) \div (\rho_5 + 6 * \rho_2 - 7.5 * \rho_0 + 1) \\
 \text{EnhancedVegetationIndex 2} &= 2.4 * (\rho_5 - \rho_2) \div (\rho_5 + \rho_2 + 1) \\
 \text{EnhancedVegetationIndex 3} &= 2.5 * (\rho_5 - \rho_2) \div (\rho_5 + 2.4 * \rho_2 + 1) \\
 \text{SoilAdjustedVegetationIndex} &= (\rho_5 - \rho_2) \div (\rho_5 + \rho_2 + 1) * 2 \\
 \text{SoilAndAtmosphericallyResistant VI 3} &= 1.5 * (\rho_5 - \rho_2) \div (\rho_5 + \rho_2 + 0.5)
 \end{aligned}$$

The *Modified Triangular Vegetation Index 1* is given by $vi = 1.2 * (1.2 * (\rho_5 - \rho_1) - 2.5 * (\rho_2 - \rho_1))$ which shows that a simple linear combination can be as much efficient as NDVI like indices by taking one additional spectral band ($\rho_2 = \text{green}$) and more adapted coefficients. However, the other 80 spectral indices do not seem to be stable against of light variation and saturation. It is thus not relevant to present them.

4.4.2 Deepindices

Finally, each baseline model such as *linear*, *linear ratio*, *polynomial*, *universal function approximation* and *dense morphological function approximation* are evaluated with 4 different modalities of each kernel size $N = 1$, $N = 3$, $N = 5$ and $N = 7$. In addition *input band filter (ibf)* and *spatial pyramid refinement block (sprb)* are put respectively at the upstream and downstream of the network. Figure 4.11 shows that network synthesis. To deal with lighting variation and saturation a BatchNormalization is put in the upstream of the network in all cases. The *ibf* and *sprb* modules are optional and can be disabled.

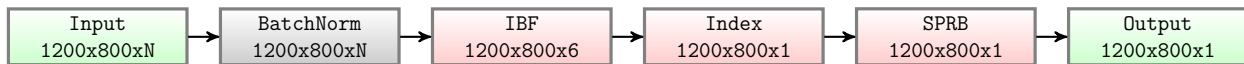


Figure 4.11 – Network synthesis with *ibf*, *evaluated index equation*, and *sprb*.

When the input band filter (*ibf*) is enabled, the incoming tensor size of $1200 \times 800 \times 13$ is transformed to a tensor of size $1200 \times 800 \times 6$ and passed to the generic equation. When it is not, the generic equations get the raw input tensor of size $1200 \times 800 \times 13$. In all cases the baseline model output a tensor of shape $1200 \times 800 \times 1$. The spatial pyramid refinement block transforms the output tensor of the baseline model to a new tensor of the same size.

All models are evaluated with two metrics, respectively the dice and mIoU score. For each kernel size, the results are presented in Tables 4.3–4.6. All models are also evaluated with and without *ibf* and *sprb* for each kernel size.

Model	mIoU				dice			
	baseline	ibf	sprb	ibf+sprb	baseline	ibf	sprb	ibf+sprb
linear	78.58	79.63	78.88	78.12	87.56	88.34	87.57	86.93
linear-ratio	79.01	78.86	77.73	79.67	87.85	87.87	86.55	88.28
polynomial	70.08	80.03	74.47	79.32	80.53	88.61	84.07	88.03
universal-function	78.39	76.59	79.04	80.15	87.27	85.36	87.63	88.53
dense-morphological	76.15	78.86	75.96	80.00	85.26	87.80	85.15	88.54
Stats	–	2.35	0.78	3.01	–	1.90	0.50	2.37

Table 4.3 – Score of DeepIndices with/without *ibf* and *sprb* for a kernel size of 1

Model	mIoU				dice			
	baseline	ibf	sprb	ibf+sprb	baseline	ibf	sprb	ibf+sprb
linear	78.89	78.21	78.53	79.76	87.66	87.16	87.35	88.36
linear-ratio	76.63	78.21	74.90	78.17	85.49	87.37	83.89	86.92
polynomial	72.83	79.31	73.20	79.13	83.06	88.13	82.78	87.82
universal-function	76.67	79.63	77.81	81.08	85.57	88.28	86.67	89.22
dense-morphological	76.54	79.39	75.65	80.29	85.43	88.17	84.40	88.66
Stats	–	2.64	-0.29	3.37	–	2.38	-0.42	2.75

Table 4.4 – Score of DeepIndices with/without *ibf* and *sprb* for a kernel size of 3

Model	mIoU				dice			
	baseline	ibf	sprb	ibf+sprb	baseline	ibf	sprb	ibf+sprb
linear	77.80	78.83	78.92	79.92	86.91	87.67	87.61	88.24
linear-ratio	75.72	77.94	77.36	80.08	84.87	87.26	86.33	88.43
polynomial	73.11	79.92	73.69	80.67	83.29	88.58	83.31	88.83
universal-function	77.60	80.63	80.31	80.63	86.38	89.02	88.53	88.71
dense-morphological	74.89	79.74	76.04	81.92	83.84	88.42	85.09	89.80
Stats	–	3.59	1.44	4.82	–	3.13	1.12	3.74

Table 4.5 – Score of DeepIndices with/without ibf and sprb for a kernel size of 5

Model	mIoU				dice			
	baseline	ibf	sprb	ibf+sprb	baseline	ibf	sprb	ibf+sprb
linear	79.08	80.29	79.25	81.49	87.75	88.57	87.80	89.42
linear-ratio	78.43	80.58	77.85	81.35	87.04	88.78	86.68	89.45
polynomial	72.49	80.79	74.14	81.21	82.92	88.99	83.77	89.27
universal-function	78.49	80.20	80.21	80.36	87.38	88.72	88.35	88.70
dense-morphological	75.70	80.35	76.34	82.19	84.48	88.70	85.61	89.94
Stats	–	3.60	0.72	4.48	–	2.84	0.53	3.44

Table 4.6 – Score of DeepIndices with/without ibf and sprb for a kernel size of 7

For all baseline models, the results (in term of mIoU) show that increasing the kernel size also increases performances. The gain performance between best models in kernel size 1 and 7 are approximately 2% and then correspond to the influence of spectral mixing. So searching for spectral mixing 3 pixels farther (kernel size 7) still increases performance. It could also be possible that function approximation allows to spatially reconstruct some missing information.

For all kernel sizes, the **ibf** module enhance the mIoU score up to 3.6%. So the **ibf** greatly prune the unneeded part of the input signal which increases the separability and the performances of all models. The **sprb** module allows to smooth the output by taking into account neighborhood indices, but their performance are not always better or generally negligible when it is used alone with the baseline model.

The baseline **polynomial** model is probably over-fitted, because it’s hard to find the good polynomial order. But enabling the **ibf** fixes this issue. However further study should be done to setup the order of Bernstein expansion.

The **dense morphological** with a kernel size of 5 and 7 using both **ibf** and **sprb** modules is the best model in term of dice ($\approx 90\%$) and mIoU score ($\approx 82\%$). Followed by **universal function approximator** with a kernel size of 1 or 3 with both **ibf** and **sprb** modules (dice up to 89% and mIoU up to 81%). Further studies on the width of the universal function approximator could probably increase performance. According to [Mondal et al., 2019] it seems normal that the potential of **dense morphological** is higher although the hyper-parameters optimization of **universal function approximator** could increase their performance.

4.4.3 Initial Image Processing

To show the importance of the initial image processing, each model has been trained without the various input transformations, such as ρ_{std} , Gxx, Gxy, Gyy filters, Laplacian filter, minimum and maximum Eigen values. Table 4.7 shows the score of DeepIndices considering only kernel size of 1 in different model.

Model	mIoU				dice			
	baseline	ibf	sprb	ibf+sprb	baseline	ibf	sprb	ibf+sprb
linear	72.34	74.29	72.94	76.97	83.15	84.66	83.03	86.50
linear-ratio	73.72	70.51	73.30	71.55	84.10	82.36	83.19	81.57
polynomial	74.33	74.14	77.88	76.42	85.07	84.49	87.19	85.94
universal-function	74.24	74.42	75.46	76.25	84.36	84.49	85.16	85.86
dense-morphological	72.04	73.72	71.03	74.69	82.27	84.00	81.33	84.72
Stats	–	0.08	0.79	1.84	–	0.21	0.19	1.13

Table 4.7 – Score of DeepIndices in different modalities for a kernel size of 1 without initial image processing

The results shows that none of optimized models outperforms the previous performance with the initial image processing (best mIoU at 80.15%). The maximum benefit is approximately 6% for mIoU score depending on the model and module, especially when using combination of **ibf**, **sprb** and small kernel size. Meaning that signal processing is much more important than spectral mixing and texture.

4.4.4 Discussion

Further improvements can be set on hyper-parameters of the previously defined equations, such as the degree of the polynomial (set to 11), the CNN depth and width for Taylor series (set to 3) and the number of operations in morphological network (set to 10). In particular the learning of 2D convolution kernel of Taylor series may be replaced by a structured receptive field [Jacobsen et al., 2016]. In addition it would be interesting to transpose our study with new data for other surfaces such as shadows, waters, clouds or snows.

The training dataset is randomly split with a fixed seed, which is used for every learned models. As previously noted, this is important to ensure reproducible results but could also favor specific models. Further work to evaluate the impact of varying training datasets could be conducted.

4.4.4.1 Model Convergence

Another way to estimate the robustness of a model against its initialization is to compare the model's convergence speed. Models with faster convergence should be less sensitive to the training dataset. As an example, the convergence speed of few different models is shown in Figure 4.12. The baseline model convergence is the same, as well as **sprb** module. However the speed of convergence also increases with the size of the kernel but does not alter subsequent observations. For greater readability only models with **ibf** are presented.

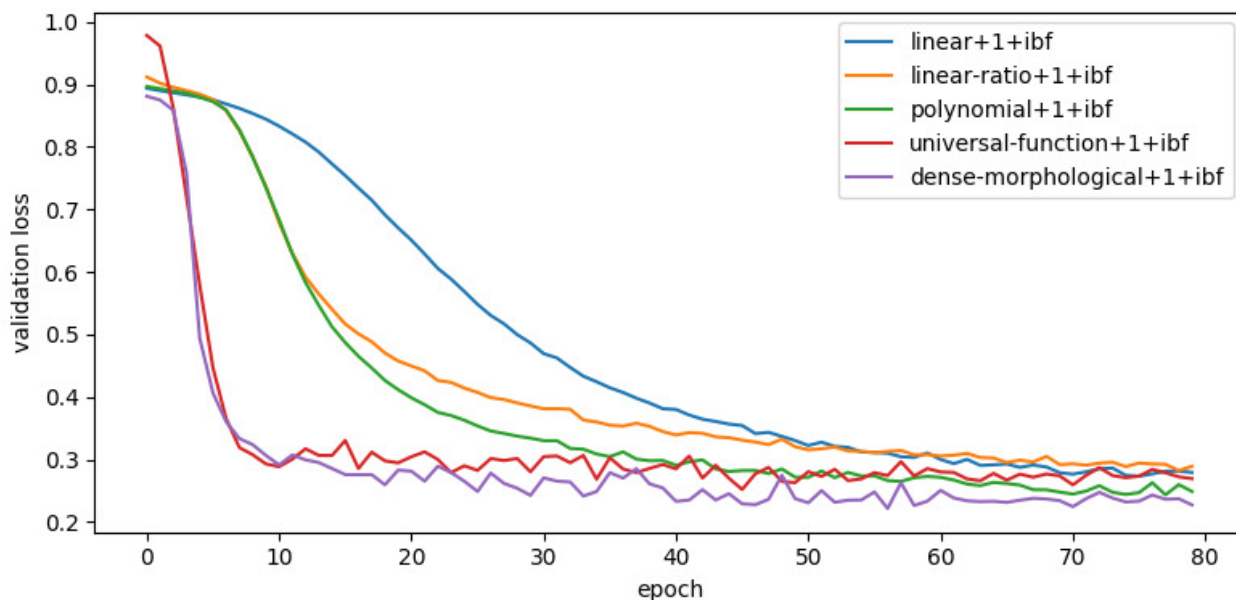


Figure 4.12 – first 80 epochs of loss of generic models with **ibf** in kernel size of 1.

An important difference in the speed of convergence between models is observed. An analysis of this figure allows the aggregation of model types and speed :

- Slow converging models : polynomials models converge slowly as well as the majority of linear or linear-ratio models.
- Fast converging models : universal-functions and dense-morphological are the fastest to converge (less than 30 iterations)

A subset of slow and fast converging models could be evaluated in term of sensitivity against initialization. It shows that the dense morphological followed by universal function approximator convergence faster than the other. Regardless of the used module nor kernel size.

4.4.4.2 Limits of Deepindices

Shadows can be a relatively hard problem to solve in image processing, the proposed models are able to correctly separate vegetation from soil even with shadowy images, as shown in Figure 4.13. In addition, the Figure 4.16 in the Appendix 4.5 shows the impacts of various acquisition factors, such as shadow, noise, specular or thin vegetation features.

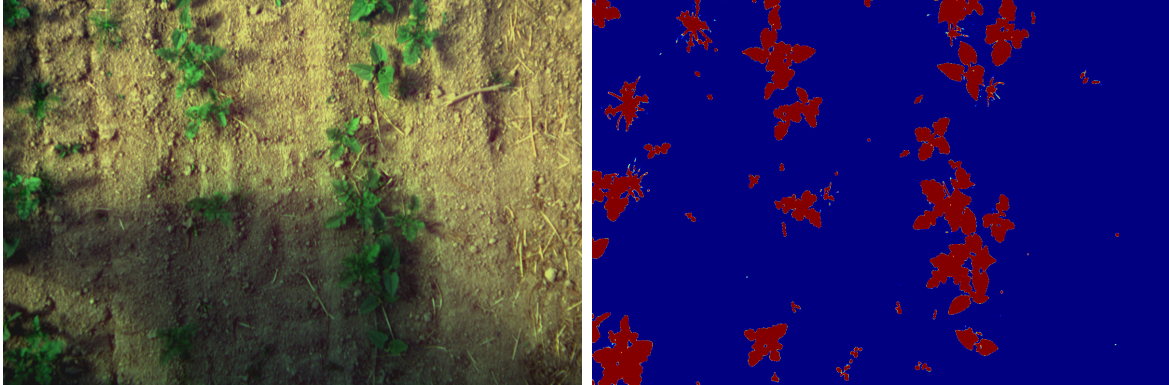


Figure 4.13 – Correct vegetation/soil discrimination despite shadows.

Some problem occurs when there are abrupt transitions between shadowed and light areas of an image as shown in Figure 4.14.

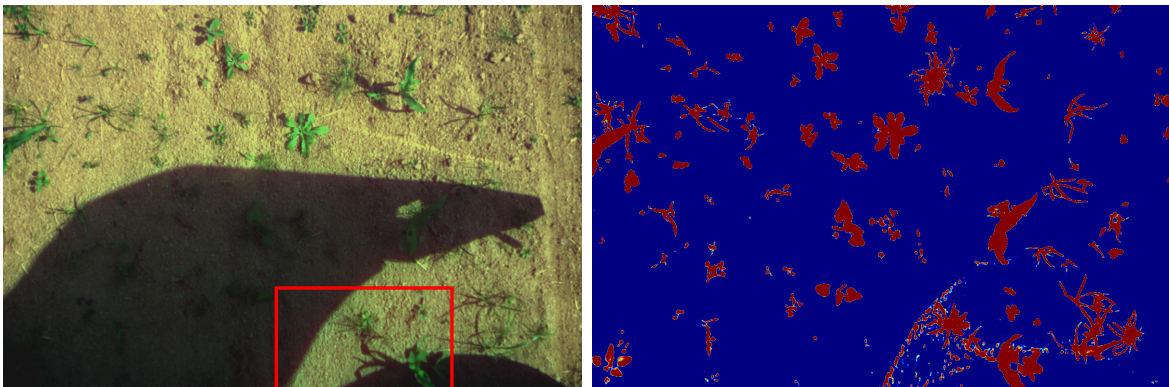


Figure 4.14 – Vegetation/soil discrimination issue with abrupt transition between shadow and light.

It appears that the discrimination error appears where the shadow is cast by a solid object, resulting in edge diffraction that creates small fringes on the soil and vegetation. A lack of such images in the training dataset could explain the model failure. Data augmentation could be used to obtain a training model containing such images, from cloud shadows to solid objects shadows. Further work is needed to estimate the benefit of such a data augmentation on the developed models.

The smallest parts of the vegetation (less than 1 pixel, such as small monocotyledon leaves or plant stems) cannot be detected because of a strong spectral mixture. This limitation is due to the acquisition conditions (optics, CCD resolution and elevation) and should be considered as is. As vegetation with a width over 1 pixel is correctly segmented by our approach, the acquisition parameters should be chosen so that the smallest parts of vegetation that are required by an application are larger than 1 pixel in the resulting image.

A few spots of specular light can also be observed on images, particularly on leaves. These spots are often unclassified (or classified as soil). This modifies the shape of the leaves by creating holes inside them. This problem can be seen on Figure 4.15. Leaves with holes are visible on the left and the middle of the top bean row. It would be interesting to train the network to detect and assign them to a dedicated class.

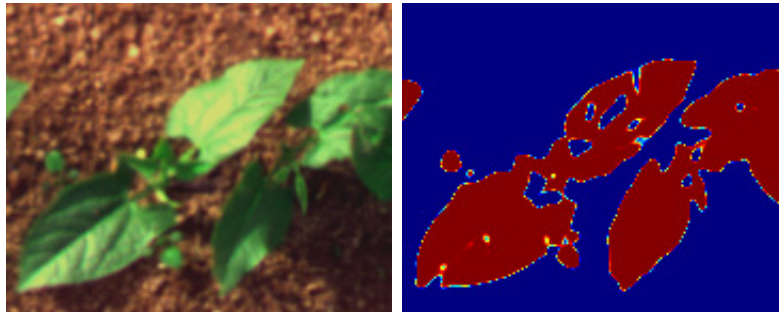


Figure 4.15 – Vegetation/soil discrimination issue caused by specular lights on leaves.

Next the location of the detected spots could be studied to re-assign them to two classes : specular-soil and specular-vegetation. To perform this step, a semantic segmentation could be set up to identify the surrounding objects of the holes specifically. It would be based on the UNet model, which performs a multi-scale approach by calculating, treating and re-convolving images of lower resolutions.

More generally, the quality of the segmentation between soil and vegetation strongly influences the discrimination between crop and weed, which remains a major application following this segmentation task. Three categories of troubles have been identified : the plants size, the ambient light variations (shades, specular light spots), and the morphological complexity of the studied objects.

The size of the plants mainly impacts their visibility on the acquired images. It is not obviously related to the ability of the algorithm to classify them. However, it leads to the absence of essential elements such as monocotyledon weeds at an early vegetation stage. A solution is proposed by setting the acquisition conditions to let the smallest vegetation part be over 1 pixel.

Conversely, the variations of ambient light should be treated by the classification algorithm. As previously mentioned, shadow management needs an improvement of the learning base, and specular light spots could be treated by a multi-scale approach. Their influence on the discrimination step should be major. Indeed, they influence the shape of the objects classified as plants, which is a useful criterion to discriminate crops from weeds. The morphological complexity of the plants can be illustrated by the presence of stems. In our case, bean stems are similar to weed leaves. This problem should be treated by the discrimination step. The creation of a stem class (in addition to the weed and crop classes) will be studied in particular.

4.5 Conclusions

In this work, different standard vegetation indices have been evaluated as well as different methods to estimate new DeepIndices through different types of equations that can reconstruct the others. Among the 89 standard vegetation indices tested, the MTVI (Modified Triangular Vegetation Index 1) gives the best vegetation segmentation. Standard indices remain sub-optimal even if they are downstream optimized with a linear regression because they are usually used on calibrated reflectance data. The results allow us to conclude that any simple linear combination is just more efficient (+4.87% mIoU) than any standard indices by taking into account all spectral bands and few transformations. The results also suggest that un-calibrated data can be used in proximal sensing applications for both standard indices and DeepIndices with good performances.

We therefore agree that it is important to optimize both the arithmetic structure of the equation and the coefficients of the spectral bands, that is why our automatically generated indices are much more accurate. The best model is much more efficient by +8.48% compared to the best standard indices and by +18.21% compared to NDVI. Also the two modules **ibf**, **sprb** and the initial image transformation show a significant improvement. The developed DeepIndices allow to take into account the lighting variation within the equation. It makes possible to abstract from a difficult problem which is the data calibration. Thus, partially shaded images are correctly evaluated, which is not possible with standard indices since they use spectrum measurement that change with shades. However, it would be interesting to evaluate the performance of standard indices and DeepIndices on calibrated reflectance data.

These results suggest that deep learning algorithms are a useful tool to discover the spectral band combinations that identify the vegetation in multi spectral camera. Another conclusion from this research is about the genericity of the methodology developed. This study presents a first experiment employed in field images with the objective of finding deep vegetation indices and demonstrates their effectiveness compared to standard vegetation indices. This paper's contribution improves the classical methods of vegetation index development and allows the generation of more precise indices (i.e., DeepIndices). The same kind of conclusion may arise from this methodology applied on remote sensing indices to discriminates other surfaces (roads, water, snow, shadows, etc).

Author Contributions Conceptualization, Vayssade, J.-A.; data curation, Vayssade, J.-A.; formal analysis, Vayssade, J.-A.; funding acquisition, Jones, G and Paoli, J.-N.; investigation, Vayssade, J.-A.; methodology, Vayssade, J.-A.; project administration, Paoli, J.-N. and Jones, G; resources, Paoli, J.-N. and Jones, G; software, Vayssade, J.-A.; supervision, Paoli, J.-N. and Jones, G; validation, Paoli, J.-N.; Gée, C. and Jones, G.; visualization, Vayssade, J.-A.; writing—original draft preparation, Vayssade, J.-A.; writing—review and editing, Vayssade, J.-A.; Paoli, J.-N.; Gée, C. and Jones, G.; All authors have read and agreed to the published version of the manuscript.

Funding This project is funded by ANR Challenge RoSE and the Horizon 2020 project IWM-PRAISE.

Data Availability Data in this study is publicly available at <https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/DSQC8N>, using Creative Common CC0 1.0 Public Domain Dedication licence

Acknowledgments We would like to thank Masson Jean-Benoit for the realization of the metal gantry which allowed us to position the camera at different heights, it was used in particular for the calibration of the camera and the band registration. We also thank Djemai Mehdi for the spelling correction of the English. And we thank Aubry Clément and Cozic Thibault of the company SITIA for their help in interfacing the camera with the used robot “Trecktor”.

Conflicts Of Interest The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendices and supplementary material

$$b = \rho_0, g = \rho_1, r = \rho_2, e = \rho_3, u = \rho_4, n = \rho_5$$

Model	Equation
Modified Triangular Vegetation Index 1	$1.2 * (1.2 * (n - g) - 2.5 * (r - g))$
Modified Chlorophyll Absorption In Reflectance Index 1	$1.2 * (2.5 * (n - r) - 1.3 * (n - g))$
Enhanced Vegetation Index 2	$2.4 * (n - r) / (n + r + 1)$
Soil Adjusted Vegetation Index	$1.5 * (n - r, n + r + 0.5)$
Soil And Atmospherically Resistant VI 3	$1.5 * (n - r, n + r)$
Enhanced Vegetation Index 3	$2.5 * (n - r) / (n + 2.4 * r + 1)$
Global Environment Monitoring Index	$\frac{2 * (n^2 - r^2) + 1.5 * n + 0.5 * r}{n + r + 0.5} * (1 - n/4) - \frac{r - 0.125}{1 + r}$
Adjusted Transformed Soil Adjusted VI	$a * \frac{n - a * r - 0.03}{a * n + r - a * 0.03 + 0.08 * (1 + a) * * 2} \quad a = 1.22$
NDVI	$(n - r) / (n + r)$

Table 4.8 – Top optimized fixed vegetation model equations

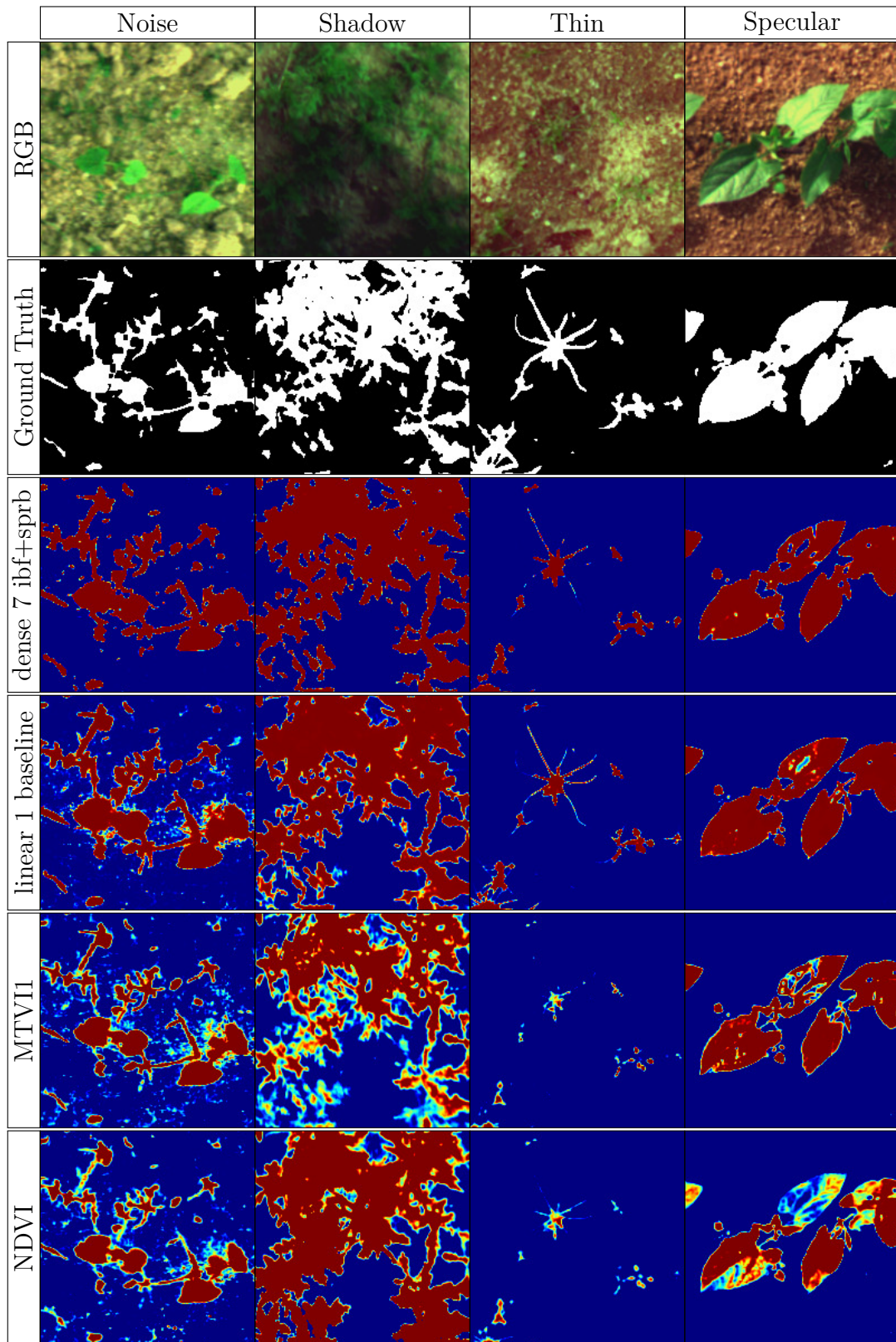


Figure 4.16 – Visual comparison between some relevant models. NDVI (63.98 mIoU), MTVI1 (73.71 mIoU), linear 1 baseline (78.58 mIoU), dense 7 ibf-sprb (82.19 mIoU). Blue indicates sure soil, red indicates sure vegetation, and the other colors indicate uncertainty.

4.6 Conclusion de chapitre

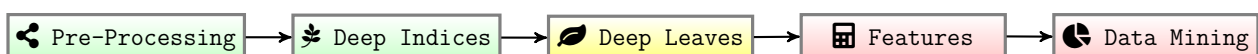
Dans cet article, de nouveaux indices de végétation (Deep Indices) ont été proposés. En partant de la définition d'une large quantité d'équations d'indices de végétation (89), il a été possible de déterminer deux équations simples qui généralisent la majorité d'entre elles, tout en considérant l'ensemble des bandes spectrales. Pour aller plus loin, des méthodes reposants sur des approximateurs de fonctions (développements de Taylor, polynômes de Bernstein, et opérateurs morphologiques) ont également été testées, de même que l'ajout de transformations des données d'entrée (déviation standard, matrice hessienne, ...) ou de traitement du signal en amont des fonctions à optimiser (filtre passe-bas, passe-haut et passe-bande).

Ces indices ont été développés au travers d'un outil d'apprentissage profond permettant des temps de traitement minimes. De plus, une fonction de perte adaptée permet d'optimiser ces indices pour obtenir la segmentation la plus proche de celle souhaitée. Un ensemble conséquent de 89 indices standards a également été optimisé en suivant cette démarche. Cela permet de comparer indices standards et indices proposés avec des degrés d'optimisations similaires.

Ces indices ont été appris sur un ensemble de 91 images non calibrées et contenant des variations importantes des conditions d'acquisition. La taille du jeu d'apprentissage ne nécessite pas d'être plus conséquente, en revanche le choix des images et la représentation des différentes conditions d'acquisition dans des proportions spécifiques jouent un rôle important. Pour ce type de démarche, il est nécessaire de sous-représenter les images simples, acquises dans de bonnes conditions. L'utilisation de techniques d'augmentation de données (bruit de Perlin) renforce également la robustesse des indices appris.

Les résultats montrent que certains indices standards peuvent proposer des résultats intéressants en termes de segmentation sol/végétation après optimisation. Le MTVI (Modified Triangular Vegetation Index 1) donne ainsi les meilleurs résultats (73.71 mIoU). A contrario, le NDVI, l'indice le plus utilisé dans la littérature, montre des résultats moyens (63.98 mIoU) pour cette tâche. Ces indices standards, sont facilement dépassés par une simple régression linéaire (78.58 mIoU). Et ils sont largement moins performants que les indices proposés dans notre étude (82.19 mIoU). Certains artefacts ne sont toutefois pas traités par les Deep-Indices. C'est le cas de la réflexion spéculaire (figure 4.15), des aberrations chromatiques visibles en bordure d'ombre (figure 4.14) et des pixels présentant un mélange spectral (figure 4.16), souvent en bordure de feuille, ou sur les feuilles fines.

Grâce à ces nouveaux indices, ne nécessitant pas de calibration, nous avons répondu efficacement à la première question de l'étape de segmentation : comment discriminer la végétation du reste de la scène ? Ceci constitue un des apports scientifiques de cette thèse. Maintenant que la première étape de segmentation est achevée, la seconde question est : comment différencier les individus au sein d'un couvert végétal ? Pour cela, plusieurs échelles sont possibles : composantes connexes, plantes, pixels, Les méthodes utilisables dépendent de ce choix et le chapitre suivant, en jaune sur la figure ci-dessous, propose une réponse à cette problématique.



Rappel de l'architecture de notre chaîne fonctionnelle
ici les premières étapes en vert sont terminées, et en jaune l'étape suivante étudiée

5 Détection des individus

Après avoir séparé la végétation du reste de l'image, l'étape suivante vise à isoler des individus au sein de la végétation. Les algorithmes pouvant être utilisés pour cette opération ont été introduits dans la section 2.3. Toutefois, le choix de l'algorithme adéquat dépend du type d'entité recherchée. Il est en effet possible de s'intéresser à une composante connexe, une plante, une feuille, un super-pixel ou encore à un simple pixel.

Dans le cadre du challenge RoSE, la densité des plantes est importante. Il existe souvent un recouvrement entre les plantes (figure 5.1). Si l'on définit l'entité comme étant une composante connexe, alors cette entité correspondra parfois à un amas de plantes (quand il y a recouvrement) et parfois à une plante isolée. Cela induira des erreurs d'étiquetage pour les amas de plantes, donc une classification grossière et l'impossibilité de détecter les adventices intra-rang, ce qui est l'objectif initial du challenge. A contrario, si nous définissons une segmentation au niveau du pixel, alors nous manquerons de propriétés pour une classification cultures/adventices performante.

Les problématiques de segmentation, de caractérisation et de classification sont fortement liées. En réalité, la définition d'une méthode permettant d'obtenir un niveau de segmentation adéquat pour la classification culture/adventices constitue l'un des verrous scientifiques majeurs de cette thèse.



Figure 5.1 – Mélange de haricot (culture) et de moutarde (adventice) dans le challenge RoSE

Ainsi, dans notre contexte d'utilisation, en proxidtection, l'échelle de la composante connexe et celle du pixel sont à rejeter. L'échelle du superpixel pourrait être intéressante, en revanche, les algorithmes permettant une telle segmentation, sont longs et peu fiables, car fortement influencés par un paramètre essentiel, celui de la taille des superpixels. En effet, trop petits, les superpixels sont trop nombreux et ne contiennent pas suffisamment d'information discriminante. Trop grands, ils sont alors constitués de mélanges des classes culture et adventice.

En agriculture de précision, pour la détection des adventices, l'entité de la plante est fréquemment utilisée. En effet, d'un point de vue agronomique, la plante est l'entité la plus fine utilisée pour caractériser, entre autres, l'état sanitaire et la pression des adventices.

Les approches historiques de segmentation ne permettent pas de segmenter une entité plante de manière stable en conditions réelles car les images présentent de fortes variations. De plus, les tailles et formes des plantes varient fortement entre individus au sein d'une même espèce, ce qui rend tout processus d'optimisation préalable impossible sur des algorithmes tels que les superpixels. C'est pourquoi, aujourd'hui, l'essentiel de la communauté a recours à des algorithmes reposant sur de l'apprentissage profond.

Les algorithmes d'apprentissage profond pris en considération en agriculture de précision utilisent des réseaux de neurones convolutifs (CNN) associés à une proposition des régions d'intérêts. Ces algorithmes de proposition de régions examinent une image d'entrée à travers un ensemble de convolutions qui maximisent l'identification de l'emplacement potentiel d'un objet. Cette méthode repose sur une régression de "bounding box", telle que proposée par [Girshick et al., 2014] (RCNN) ou [Kim and Cho, 2020] (YOLO).

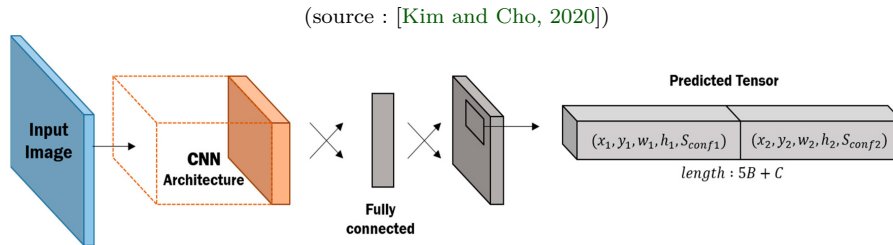


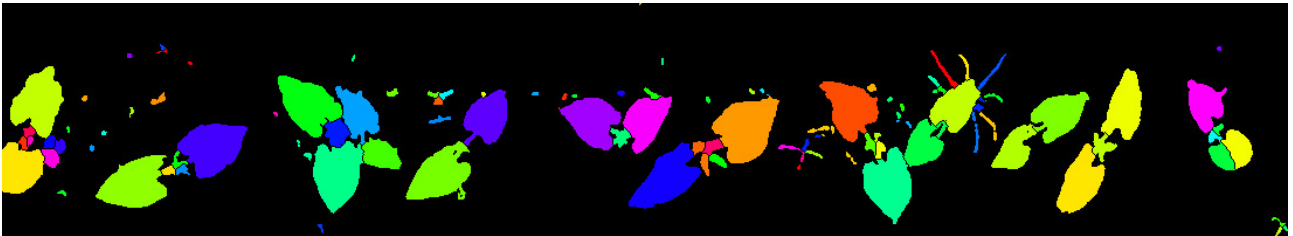
Figure 5.2 – Structure d'un réseau "You Only Look Once" (YOLO)

Les premiers travaux de [Girshick et al., 2014] utilisaient ensuite ces régions d'intérêt pour extraire les caractéristiques de sortie du réseau de neurones, afin de les introduire dans une machine à vecteurs de support (SVM) pour la classification finale. Au fil des évolutions, la classification SVM a été remplacée par un réseau de neurones dédié à l'identification de la classe de l'objet [Girshick, 2015], permettant un apprentissage dit de "bout en bout". Cette régression des régions d'intérêt a toutefois deux inconvénients : (1) elle propose une grande quantité de faux positifs et (2) des sorties multiples, mais légèrement décalées, pour un même objet. Pour pallier ces problèmes, le "score detection" est utilisé pour supprimer la majorité des faux positifs, ainsi qu'un algorithme de suppression des non-maximum (NMS). Le NMS permet d'éviter la détection répétée de la même instance en combinant les régions (bounding box) qui se chevauchent en une seule boîte englobante pour chaque objet.

Hélas, ce type d'approche montre ses limites en environnement dense, du fait de l'algorithme NMS. Ainsi, détecter des plantes par cette méthode ne permet pas un résultat optimal dans le cadre du challenge RoSE. En réalité, de par les nombreuses occultations et recouvrements inter-individus, l'échelle de la plante est discutable. De ce constat, il a été décidé de travailler sur une entité plus petite : la feuille. De plus, nous proposons ici, le développement d'un réseau de neurones ne s'appuyant pas sur l'algorithme NMS pour contourner ces limitations.

En allant plus loin sur les réseaux de neurones définis pour les indices de végétation et la segmentation sol/végétation, il est possible de proposer un réseau de segmentation sémantique pour la segmentation des instances des feuilles. L'approche repose sur la détection des contours des feuilles [Morris, 2018]. Cette approche est peu utilisée mais elle peut être améliorée de différentes façons. La suite de ce chapitre porte sur les travaux qui ont été menés dans ce domaine pour développer une méthode de segmentation de feuille en couvert dense (chevauchement d'objets) et est l'objet d'une publication internationale dans la revue *Computers and Electronics in Agriculture* (travail en cours de révision).

Pixelwise Instance Segmentation Of Leaves In Dense Foliage



Authors Jehan-Antoine Vayssade, Gawain Jones, Christelle Gée and Jean-Noël Paoli

Publication Sent on 20/08/2021 – Computer and Electronics in Agriculture – under-review

Abstract Detecting and identifying plants using image analysis is a key step for many applications in precision agriculture (from phenotyping to site specific weed management). Instance segmentation is usually carried on to detect entire plants. However, the shape of the detected objects changes between individuals and growth stages. A relevant approach to reduce these variations is to narrow the detection on the leaf. Nevertheless, segmenting leaves is a difficult task, when images contain mixes of plant species, and when individuals overlap, particularly in an uncontrolled outdoor environment. To leverage this issue, this study based on recent Convolutional Neural Network mechanisms, proposes a pixelwise instance segmentation to detect leaves in dense foliage environment. It combines “deep contour aware” (to separate the inner of big leaves from its edges), “Leaf Segmentation through classification of edges” (to separate instances with a specific inner edges) and “Pyramid CNN for Dense Leaves” (to consider edges at different scales). But the segmentation output is also refined using a Watershed and a method to compute optimized vegetation indices (DeepIndices). The method is compared to others running the leaf segmentation challenge (provided by the International Network on Plant Phenotyping) and applied on an external dataset of Komatsuna plants. In addition, a new multispectral dataset of 300 images of bean plants is introduced (with dense foliage, individuals overlapping, mixes of species and natural lighting conditions). The ground truth (e.g. the leaves boundaries) is defined by labelled polygons and can be used to train and assess the performance of various algorithms dedicated to leaf detection or crop/weed classification. On the usual datasets, the performances of the proposed method are similar to those of the usual methods involved in the leaf segmentation challenges. On the new dataset, their results are strongly better than those of the usual RCNN method. Remaining errors are bad fusion between neighboring areas and over segmentation of multi-foliolate leaves. Structural analysis methods could be studied in order to overcome these deficiencies.

Keywords precision agriculture ; remote sensing ; leaf segmentation ; dense foliage ; boundary detection ; semantic segmentation ; CNN ; multispectral

5.1 Introduction

In precision agriculture, one of the hardest tasks is the detection of crops and weeds by imagery. Imaging systems will play a significant role in the new generation of agriculture, from the genetic selection in phenotyping [Omari et al., 2020] to site-specific weed management [Louargant et al., 2017]. They are also used to record frequently and accurately the plant growth, crop yield, leaf area, etc [Gée et al., 2021]. These data are then used to quantify and evaluate the quality of production.

In proximal detection, this work is mainly done at the plant level and gives important agronomic information once reported at the plot level (number of weeds, crops and weeds locations, diseases, stress ...). The plant level is also needed for the new agricultural revolution, such as robotic weed management. Thus, studies try to detect crop and weeds plants by using a wide variety of techniques, which have been reviewed by [Wang et al., 2019a]. The instance segmentation is a key-step used before the task of classifying plants as crop and weeds. In CNN field, it is mostly based on a major class approaches [Hafiz and Bhat, 2020], such as Mask-RCNN. However, detecting the entire plants has 3 main limitations (i) when plants are too numerous, it is hard to distinguish individuals when they overlap, one of them is undetected or both are merged which is due to the underlying non-max suppression algorithm [Bonneau et al., 2020], (ii) small and thin elements are undetected and (iii) the number of variations for a plant is infinite : the number of leaves, their orientations, their sizes and other differences that radically change between individuals and growth stages, implies a large amount of data for the training process. These limitations logically impact the detection rate and may cause a significant quantity of miss-detection. To solve this problem, one approach that should be relevant is to base the detection not on the whole plant, but on the leaf. For this purpose, a pixel-wise instance segmentation is proposed. The idea is to separate the instances by detecting the edges of the leaves whose projected shadow gives an interesting gradient break, particularly in an uncontrolled outdoor environment.

5.1.1 Related Works

Some recent works on leaf segmentation were found in controlled illumination environments with limited occlusion between individuals. Especially on an open dataset of Arabidopsis Thaliana [Scharr et al., 2016]. Other studies related to biomedical imagery and nuclei segmentation were also found with pixel-wise instance segmentation. These studies show the important of defining one or more edges classes and a relevant loss function. These two factors are both related to the weight given to each sample of the training dataset in the estimated error for the optimization algorithm. But the way of how parts of the network are dedicated to the edge classification and thus on how the network focuses on instance separation. Edge detection therefore plays a significant role in pixelwise instance segmentation task.

The first studies were dedicated to the definition of specific class of edges to separate instances. Thus [Chen et al., 2016] proposed a novel approach named “Deep Contour-Aware” based on the semantic segmentation of two classes, one class is dedicated to the inside of glands, while the second is for the segmentation of glands edges. A bit later, [Morris, 2018] was the first to define a pixelwise instance segmentation for dense leaf detection. They proposed the “Pyramid CNN for Dense Leaves” architecture which is similar to U-Net (the most common CNN used for biomedical images segmentation). The network is dedicated to the detection of leaf boundaries. To facilitate the learning of edges at different scales, an auxiliary loss function is placed at each sub-scale of the pyramid. Finally, the instances are retrieved by using a superpixel algorithm.

[Cui et al., 2019] enhance the “Deep Contour-Aware” model proposed in 2016 by using a real U-Net architecture and data augmentation. Concerning edge classes based instances segmentation, [Bell and Dee, 2019] study shows the importance of separating edges into two classes. As the outer edges are dominant in the samples, the corresponding error on contiguous edges are less important. Thus, the edges of leaves are separated into two classes, one for the outer edges and one for inner edges (when leaves are touching or overlapping). The multi-scale loss function is still used and proposed by [Xie et al., 2020] for nucleus instance segmentation. They show that multi-scale loss helps to regularize the network and narrow down the perceptual distances and enlarge the semantic dissimilarity between individuals. In addition, a count ranking loss is used on the last feature layer of a fully-connected layer. This count ranking loss enforces the network to focus on the learning of samples containing crowded nuclei. This technique results in an implicitly regularized trained network, to be aware of individuals quantity.

5.1.2 Objectives

All these studies show that pixelwise instance segmentation technique is viable but still limited. They also underline the importance of choosing a loss function adapted to the defined semantic classes. Based on these related works, this paper proposes to merge most recent advances in the field of pixelwise instance segmentation.

First, the proposed method combines “deep contour aware” (to separate the inner of big leaves from its edges), “Leaf Segmentation through classification of edges” (to separate instances with a specific inner edges) and “Pyramid CNN for Dense Leaves” (to consider edges at different scales). Second, a new loss function is also introduced to limit under and over-segmentation. Third output is refined using a specific vegetation index based on previous work [Vayssade et al., 2021] and a watershed algorithm.

This method is applied to dense leaf segmentation, on images containing mixes of plant species and acquired in natural light. A specific multispectral dataset has been acquired, it is presented in the next section 5.2.1 and released publicly. The method is also evaluated on two common online leaf RGB image databases [Scharr et al., 2016], presented in section 5.2.2.

The proposed method is much more robust compared to previous pixel-wise instance segmentation methods and solves the issue of methods based on bounding-box regression : all small and thin elements are detected.

5.2 Material and data

5.2.1 Specific multispectral dataset

5.2.1.1 Experimental plot

The data were acquired at the site of INRAE (Figure 5.3) in Montoldre (Allier, France, at 46°20'30.3"N 3°26'03.6"E) within the framework of the “RoSE challenge” founded by the French National Research Agency (ANR) in 2019. The aim of the Challenge is to objectively compare the solutions proposed by participants for intra-row weed control ([Avrin et al., 2020]). Within this context, the challenge provides to contestants an evaluation plan and a set of experimental plots of bean and maize plants. In addition various natural weeds (yarrows, amaranth, geranium, plantago, etc) and sown ones (mustards, goosefoots, mayweed and ryegrass) are managed to compare performances.



Figure 5.3 – Aerial view of the experimentation plot located in Montoldre (now INRAE)

5.2.1.2 multispectral camera

The images were acquired with the Airphen (Hyphen, Avignon, France) six-band multispectral camera (visible on the upper-left of the Figure 5.4). This is a scientific camera developed by agronomists for agricultural applications. The camera embeds six sensors using six bandpass (450/570/675/710/730/850 nm) filters with a 10 nm FWHM each. The focal length of each lens is 8 mm. The raw resolutions for each spectral band are 1280 × 960 px with 12 bit precision. Due to the conception of the camera, spectral images are not aligned. Based on previous work [Vayssade et al., 2020], a method for registration has been developed with a registration accuracy down to sub-pixel. After the registration, all spectral images are cropped to 1200*800 px and concatenated to channel-wise where each dimension refers to a spectral band.

5.2.1.3 Image acquisition and annotation

From the presented experimental plots, a set of images were acquired. The camera is attached in front of an hybrid autonomous tractor called “TREKTOR” launched by SITIA Company (Bouguenais, France) in 2019. The camera is setup to have a top-down view of crop rows, thus it is mounted on a pole in front of the platform allowing to remove visible part of the robot and at 1.8 m from the ground. The Figure 5.4 below shows the arrangement of the elements. Crops and weeds were between phenological state 3 and 4 which means they have between 2 and 6 leaves. The ground truth is defined on images by experts with polygons around each leaf boundary. In addition, polygons contain a label for their classification between crop and weed (not used in this study). The annotation was defined using the VIA annotation software [Dutta and Zisserman, 2019] and a total of 300 images of bean were annotated, 170 from acquisition made in June and 130 in October. These dataset is freely available at <https://doi.org/10.15454/JMKP9S>.



Figure 5.4 – The experimental set-up : the multispectral camera and the robotic platform

5.2.2 Online image databases

5.2.2.1 Additional data from the Computer Vision Problems in Plant Phenotyping dataset (CVPPP)

To compare the method proposed in this study with others, an additional dataset was used. This dataset is proposed for the Leaf Segmentation Challenge (LSC [Scharr et al., 2017]) provided by the International Network on Plant Phenotyping (INPP), a very popular challenge for data scientists. It is composed of RGB images of *Arabidopsis Thaliana* (783 images) and Rosette (27 images) plants segmented into several leaves. The authors state that the images were collected from multiple locations in a growth chamber experiment and divided into four groups, named A1 through A4. In addition, the dataset is composed of various image sizes (respectively 530×530 , 565×565 , 2048×2048 , 441×441 for each sub-dataset), which have been resize to $512 \times 512 \times 3$.

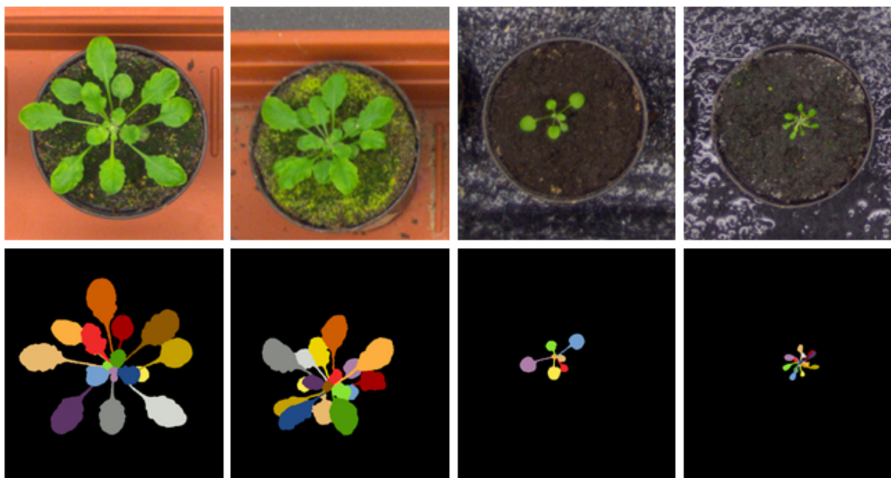


Figure 5.5 – Example of images from the CVPPP dataset (top) and their corresponding ground truth (bottom)

5.2.2.2 Additional data from Komatsuna dataset

Similar to the LSC dataset, Komatsuna dataset consists in RGB images of plants taken in a top view. This dataset contains a large number of plant growth stages, it was designed to solve the problem of 3D phenotyping [Uchiyama et al., 2017]. It is used here primarily as a test set due to its similarity to the CVPPP dataset (training set), as shown in Figure 5.6. This dataset includes 900 images of size $480 \times 480 \times 3$.

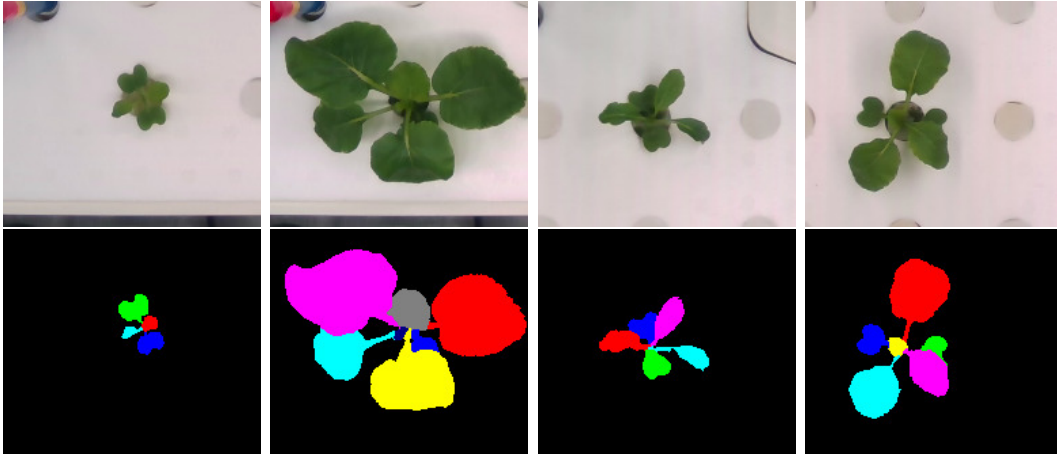


Figure 5.6 – Example of RGB images from the Komatsuna dataset (top) and their corresponding ground truth (bottom)

5.3 Methodology

We consider the leaf segmentation as a binary segmentation problem of boundaries as proposed by [Morris, 2018]. The main idea is to detect the sharp edges of leaves or to follow the projected shadow of a leaf on the one below it. This methodology section is split into three sub-tasks (5.3.1) the proposed CNN architecture to detect and separate leaves, (5.3.2) the specific loss function defined to limit under and over-segmentation, and (5.3.3) a simple watershed algorithm which takes the CNN output and a vegetation mask to refine the segmentation.

5.3.1 Proposed CNN architecture

Unlike recent CNN architectures, the proposed approach is slightly more decomposed like standard biomedical and agricultural computer vision pipelines [Perez-Sanz et al., 2017, Lottes and Stachniss, 2017]. Thus, the architecture (Figure 5.7) is composed of three **upstream modules** (IIT, IBF, UFA) that improves the input data and eliminates unnecessary information. This step composed of 3 **upstream** modules, was proposed in a previous work to construct a vegetation index ([Vayssade et al., 2021]). It is used to identify relevant spectral features on the input data to exploit the inter-channel relationships. After this stage, a **core network** is used to consider spatial information at different scales on the image, the core network returns four down-scaled feature maps. Finally, at the end of the network, three **downstream modules** (CoordConv, UFA, Classification) are proposed to fuse spectral and spatial information. Sigmoid activation function is used at the end of all convolution layers to learn more complex structures and allows non-linearity of the reconstructed function.

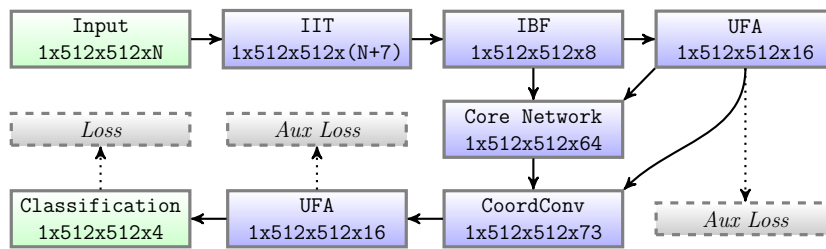


Figure 5.7 – Diagram of the CNN network architecture and losses (dotted). Multiple arrow show concatenation as input layer.

The network takes an input image of size $512 \times 512 \times 6$, thus the learning and computation are done by a sliding window within the registered images of the Airphen camera of size $1200 \times 800 \times 6$. The output layer is defined as a semantic segmentation of four classes. One class is dedicated to the inner of individuals, while three classes are dedicated to the detection of edges to keep aware of leaves instance. As mentioned by [Bell and Dee, 2019], one class is dedicated to outer of leaf (touching soil texture), while the second is dedicated to inner edges (touching/overlapping another leaf). Within our dataset, a third class appears with thin leaves which can be considered as a kind of edge. The ground truth is a set of polygons, drawn with the respect of these classes. Edge classes were empirically drawn with three pixel thickness. The following figure shows the input ground truth.

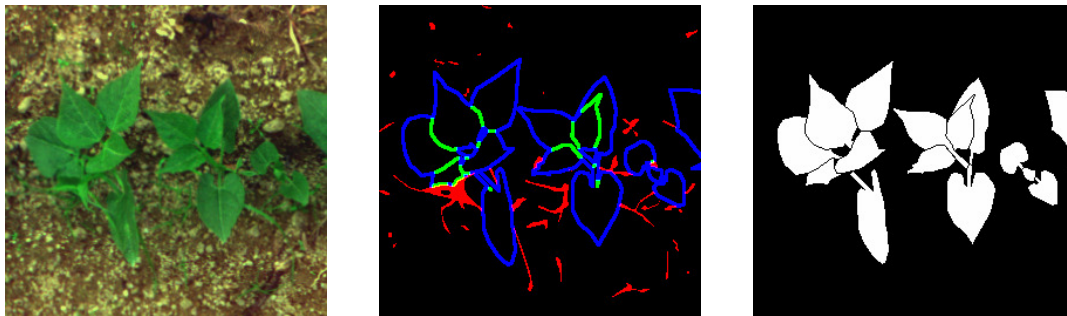


Figure 5.8 – Example of input data : from left to right, the first three bands (RGB) of the Airphen multispectral camera, three edge classes and the inner of individual class

5.3.1.1 Upstream of the network

It is composed of three **upstream** modules : Initial Image Transforms (IIT), Input Band Filter (IBF) and Universal Function Approximator (UFA).

Initial Image Transforms (IIT) In order to enrich the pool of information, spectral band transformations are added to take into account specific spatial gradients in the image and spectral mixing. Seven important transformations are considered. The standard deviation between spectral bands, noted ρ_{std} can help to detect the spectral mixture. For example, between two different surfaces like ground and leaf (which have opposite spectral radiance), the spectral mixing makes a pixel with linear combination, thus the standard deviation tends to zero [Louargant et al., 2017]. Three Gaussian derivatives on different orientations are computed. Gxx, Gxy and Gyy filters on ρ_{std} give an important spatial information about the breaks of gradient, and therefore about the outer limits of surfaces. The Laplacian, the minimum and maximum eigenvalues (of the Hessian matrix) also called ridge of the ρ_{std} seems to easily detect fine elements [Lin et al., 2014], such as monocotyledons for vegetation images. All these transformations are concatenated to the channel-wise normalized spectral band input and build the final input image. In the end we have six spectral images plus seven transformations for a final image composed of 13 channels.

Input Band Filter (IBF) To remove unneeded parts of the signal, low-pass, high-pass and band-pass filters are added. To apply the low-pass filter we use the equation $z = \max(x - a, 0)/(1 - a)$ and thus it allows to suppress low values. For the high-pass filter we apply the equation $w = \max(b - x, 0)/b$ to suppress the high values. The band-pass filter is the product of low-pass and high-pass filters defined by $y = z * w$. The output layer is the concatenation in the channel-wise of the input images, the low-pass, the high-pass and the band-pass filter which produce $4 * 13 = 52$ channels. Finally, to reduce the output data for the rest of the network, a bottleneck is inserted using a 3×3 convolution layer, and it generates a new tensor with 16 channels.

Universal Function Approximator (UFA) To separate efficiently leaves from the background and to learn spectral features, a Universal Function Approximator is defined on the upstream of the network (Figure 5.9). The UFA is based on Taylor expansion theorem, an approach to learn this form of development in deep-learning is called DenseNet and then corresponds to the sum of the concatenation of the signal with these spatio-spectral derivatives. This was successfully used for vegetation segmentation [Vayssade et al., 2021]. Three parameters, such as the *depth* (number of convolutions), the *width* (number of filters denoted W) and k (kernel size) configure the network and were empirically fixed to $depth = 3$, $width = 16$ and $k = 1$. An auxiliary output is used here to maximize the class similarity on the upstream of the network and to extract pure spectral information.

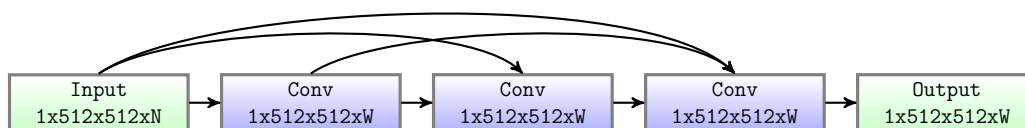


Figure 5.9 – Universal function approximator based on DenseNet. Multiple arrows shows concatenation as input layer

5.3.1.2 Core network model

The proposed core CNN architecture is based on an advanced U-Net architecture named MFP-Unet (multi-feature pyramid U-net) proposed by [Moradi et al., 2019]. It is a neural network composed by several 2D convolution layers between different sub-scales. At each sub-scale, the spatial size is divided by two and the number of filters is multiplied by two. Sub-scales are obtained by Max-Pooling layers. Then, to retrieve the original size a 2D UpSampling layer is used. In this study the depth of the U-Net model is fixed to three down-scale (size 512, 256, 128, 64). The specificity of MFP-Unet is that all sub-scale feature maps are directly up-sampled to the initial size, concatenated to the channel-wise and then used for the classification (Figure 5.10). In addition according to [Morris, 2018] and [Xie et al., 2020] an auxiliary loss function is put at each sub-scale feature layer and it enforces the learning of edges at different scale, making the network more robust to spatial resolution. Losses at each prediction also shorten the back propagation path leading to faster convergence. All convolution layers use a kernel size of 3×3 and are followed by a Batch Normalization and a sigmoid activation function ([Moradi et al., 2019, Nwankpa et al., 2018]).

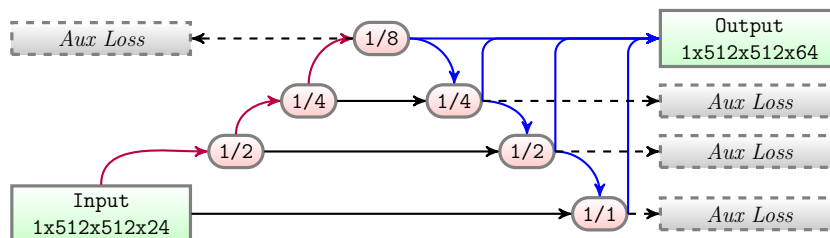


Figure 5.10 – Synthesis of the core network based on MFP-Unet. Red arrows shows Max-Pooling. Blue arrows shows Conv+UpSampling. Black arrows shows Conv. Sub-scale ratio are labelled on corresponding layer. The final output is the concatenation of features UpSampling. Multiple arrows shows concatenation as input layer

5.3.1.3 Downstream of the network

It is composed of three **downstream** modules : CoordConv and Universal Function Approximator (UFA) and Classification.

CoordConv The core network model produces a concatenation of 4 layers of 16 features ($4 \times 512 \times 512 \times 16$) which results of a layer of size $1 \times 512 \times 512 \times 64$. A coordinate layer [Liu et al., 2018a] is also concatenated and allows to consider the mapping between the coordinates in (x,y) Cartesian space to one-hot pixel space. Three variables are appended, the normalized x and y coordinate and the radial coordinate $\sqrt{(x - 0.5)^2 + (y - 0.5)^2}$. This module improves the results removing noise, ground moisture and it fixes few small holes in the segmentation mask.

Universal Function Approximator (UFA) This UFA – also presented on the upstream – is used to accurately mix features coming from various scales as well as the Cartesian coords. This module reconstructs the mapping function from the Cartesian space to a spatio-spectral feature of size $1 \times 512 \times 512 \times 16$. However, in contrast with the upstream UFA, the kernel size k was fixed to 3 to take into account neighboring.

Classification and Auxiliary output The classification is done through a small network composed of two 1×1 convolution layers. Followed by a “Pyramid Pooling Module” to consider different scales before the outputs and smooth the boundary prediction. [Zhao et al., 2017] showed that fusing the low to high-level features improved the segmentation task. It consists in the sum of different 2D convolutions whose kernel sizes have been set to 3, 5, 7 and 9. The number of filters is the same as the number of classes : 4 (defined in 5.3.1). The result of each convolution is concatenated and the final image output is given by a 2D convolution. In addition all convolutions are followed by a Batch Normalization and a sigmoid activation function. The figure 5.11 shows that sub-network.

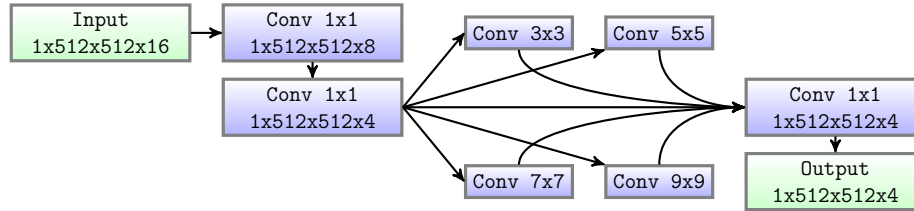


Figure 5.11 – Auxiliary output and classification module. Multiple arrows shows concatenation as input layer

5.3.2 Loss function

A wide variety of loss functions have been developed during the emergence of deep-learning. Recently, [Rahman and Wang, 2016] proposed a solution to optimize an approximation of the mean Intersection over Union (mIoU) which seems to be optimal for binary segmentation [Zhou et al., 2019]. The loss function using ground truth (p) and the prediction (\hat{p}) is defined by :

$$\text{mIoU}(p, \hat{p}) = 1 - \frac{p\hat{p}}{p + \hat{p} - p\hat{p}} \quad (5.1)$$

This loss function was used on each auxiliary. In addition the loss is computed separately on each class, weighted (denoted W_C) and summed. The result of this function is :

$$\text{Aux}(p, \hat{p}) = \sum_{C=0}^4 W_C \times \text{mIoU}(p_C, \hat{p}_C) \quad (5.2)$$

In the above equation, the weight W were empirically set to [0.175, 0.526, 0.211, 0.09]. Meaning that the second class (inner edges) is prioritized (to separate inner instance). Then it is outer edges that allow to separate the “big” leaves from small or thin leaves. Finally the thin leaves (mostly spectral mixing) and inner big leaves (essentially vegetation minus boundaries) have the smallest weight values because these classes should be easier to learn.

In recent CNN architectures for instance segmentation, the loss function does not take into account the number of detected instances or the shape of the segmentation. This aspect is only evaluated after learning, e.g., using a symmetric best dice metric. This implies that we can not guaranty that the network is well learned on crowded scene, where instance is generally merged. One problem is that until recently, instances could not be retrieved directly during the learning phase, this is due to the “non-maximal suppression” algorithm required after the CNN or the time required for the association between the detected instances and the ground truth.

In this paper, we introduce a new loss function considered at the downstream of the network. The main idea is to take into account an approximation of the segmentation quality of each leaf. To estimate the segmentation quality, the undetected, under/over segmented and fused objects can be evaluated, through a sorted histogram of the number of pixels associated to each connected component for both prediction and ground truth, as showed in the next figure 5.12 for both prediction (orange) and ground truth (blue).

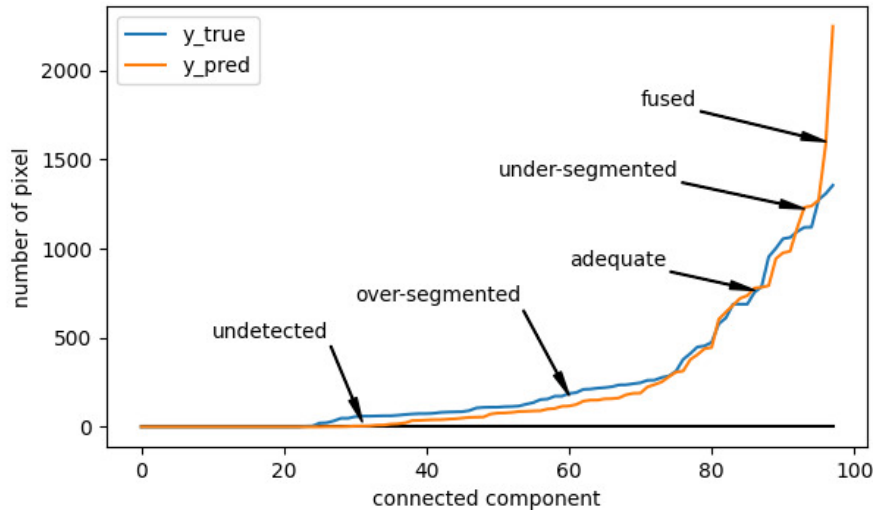


Figure 5.12 – Sorted histogram of the number of pixels associated to each component, the blue (resp. red) line represents the true (resp. predicted) number of pixels for each component.

This figure shows that when the prediction curve is lower (orange) than the ground truth curve (blue), it means that there is over-segmentation. This arise when a bigger element is split in two, when the borders of the shape are trimmed or when zero pixels of the shape are detected (undetected). On the other hand, when the prediction curve is higher than the ground truth, it means that the contours of the shape are roughly detected or if it greatly exceeds the ground truth, then we are in the presence of fused shapes. Based on these curves, a loss function can be construct the deal with over and under segmentation on shape criterion, thus a custom absolute percentage error was defined :

$$\text{Leaf}(p, \hat{p}) = \frac{\sum_{i=0}^N |\text{leaky_relu}(h_i(p) - h_i(\hat{p}))|}{h_{max}(p) + 1} \quad (5.3)$$

$$\text{leaky_relu}(x) = \begin{cases} x & \text{if } x \geq 0 \\ x * 0.2 & \text{if } x < 0 \end{cases} \quad (5.4)$$

On the above equation N is the number of components, $h_i(p)$, and $h_i(\hat{p})$ is respectively the number of pixels of a connected component in the ground truth and on the predicted segmentation within the sorted histogram. While $h_{max}(p)$ is the maximum number of pixels of a component in the ground truth.

The `leaky_relu`, is used to explicitly prioritise the learning on under-segmentation rather than over-segmentation which allows to prioritize the merged objects. This was defined because conventional losses did not give good results in dense vegetation cover, causing a large segmented area that is detected as a single entity. Note also that over-segmentation is less problematic, since it occurs mainly around the borders of the leaves, which are recovered later, through a watershed algorithm (section 5.3.3). It is the first study to suggest this type of loss.

The downstream loss is defined by $\text{DownAux}(p, \hat{p}) = \text{Aux}(p, \hat{p}) + \text{Leaf}(p, \hat{p})$. Finally the global loss considers the upstream auxiliary loss, each of the 4 feature scale auxiliary loss and the downstream loss. Thus the global loss is defined as the weighted sum of all auxiliaries losses (in the same order) where the weights W were empirically defined with $W = [0.01, 1.0, 0.1, 0.1, 0.01, 4.0]$.

5.3.3 Refinement with vegetation mask and watershed

The used U-Net architecture is good at detecting “big” elements on the scene but lacks precision on very small and thin elements. A method to produce an optimized vegetation mask was proposed in a previous work [Vayssade et al., 2021]. Using this mask provides better performance than adding a specific class to the network. Thus our previous work is used here to get the best foreground/background segmentation mask as input of the watershed. It is also learned on a specific dataset with more illumination conditions and should be more robust, especially for thin elements.

The proposed network generates 4 classes. The first two are associated to “big” leaf boundaries (denoted $\text{Edges} = \text{Outer} + \text{Inner}$). The third one is small and thin leaves (denoted Thin), and the fourth is the inner of big leaves denoted Big . The watershed seed is defined with the following equation 5.5 to generate the seed mask and can be see in the figure 5.13 :

$$\text{Seed} = \text{Thin} + \text{Big} - \text{Edges} \quad (5.5)$$

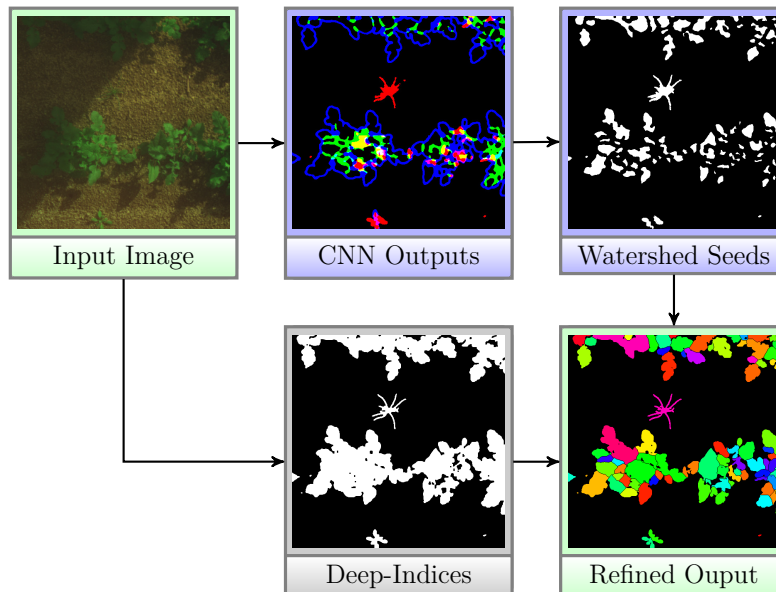


Figure 5.13 – The refinement of the CNN output through watershed and Deep-Indices. The Seed of the equation 5.5 can be seen on the “watershed seed” image.

5.3.4 Training setup

The image dataset is randomly split into a training set (80%) and a validation set (20%). However the initial seed is kept for reproducibility. In addition a random data augmentation is used during the training to increase the dataset variability. A random vertical and horizontal flip is considered as well as a perlin simplex noise [Bae et al., 2018] (of size 512×512), set with 2 modal in range $[0.7 - 1.3]$ which multiplies the number of input images. Low values add shadows, while high values add specular effects. The training is done through Keras module within Tensorflow 2.6.0 framework. All computations are done on an NVidia RTX 3060 which have 12GiB of memory and due to the size of the network only one image is computed at once.

5.3.5 Evaluation metrics

There is a large number of possible evaluation metrics for instance segmentation, which have been reviewed in [Scharr et al., 2016]. However, we decided to keep the evaluation metrics used in the Leaf Segmentation Challenge (LSC) as a reference for comparison [Scharr et al., 2017, Kulikov et al., 2018, Bell and Dee, 2019, Ward and Moghadam, 2020]. Therefore, we use the foreground/background DICE metric to evaluate the separation of large leaves from the ground (denoted *FgBgDice*). We estimate the average accuracy of leaf segmentation by the symmetric best DICE score among all objects (leaves). The best DICE score among all objects (leaves) to estimate the average accuracy of leaf segmentation is denoted *BestDice*. The *AbsDiffFG* estimates how good the algorithm is at segmenting the leaves. And finally, *DiffFG* estimates the efficiency of the algorithm for counting leaves. *SBD* for Symmetric best DICE is extract to estimate the average leaf segmentation accuracy. All these metrics are common and presented by [Scharr et al., 2016]. In addition, to compare datasets results, a new metric is introduced named *NAbsDiffFG* defined by the division of *AbsDiffFG* and the mean of leaves count in the dataset.

5.4 Results

As previously defined, all datasets were split into training and validation sets with a ratio of 80 – 20%. Once training is done on the defined setup and using the loss function, the evaluation metrics are extracted : the *FGBGDice* metric is used to evaluate the soil versus vegetation segmentation, while the *BestDice* metric is used to evaluate the instance segmentation. Each connected component is associated to its best corresponding ground truth based on a dice score. Then the metric is defined by the mean dice score of the best match. Other metrics show relevant information about over and under segmentation. The following subsections are dedicated to each dataset, from the simplest to the most difficult to segment.

5.4.1 Komatsuna dataset

The first one is the Komatsuna dataset. It is composed of RGB data of growing Japanese Mustard. This dataset is interesting for its controlled illumination conditions. In addition, the number of leaves it quite the same for all the growing stages, ranging from 3 to 6 leaves. This is important regarding our loss function which take into account the quantity and quality of each segmented leaf. However, this dataset does not contain small and thin leaves, thus the third class was replaced by a foreground class. The following Table 5.1 shows the evaluation on this dataset.

Metric	Training	Validation
FgBgDice	0.9732	0.9715
BestDice	0.8796	0.8565
SBD	0.8713	0.8517
DiffFG	-0.2639	-0.4444
AbsDiffFG	0.3944	0.5222
Number of images	720	180
Mean leaf count	4.9194	4.9722
NAbsDiffFG	0.0802	0.1050

Table 5.1 – Evaluation metrics for the Komatsuna dataset

The result on this dataset shows a relatively perfect score of the soil and vegetation segmentation with a $FGBGDice$ of 0.97. Few errors remain as shown in figure 5.14, a green bottle is visible on the left side of some images, which seems to indicate that the green component has a major impact on vegetation detection. Since RGB data are used, this error should be visible in most studies. This interesting element also demonstrate that the used *CoordConv* layer does not play its role in the spatial management of this element.

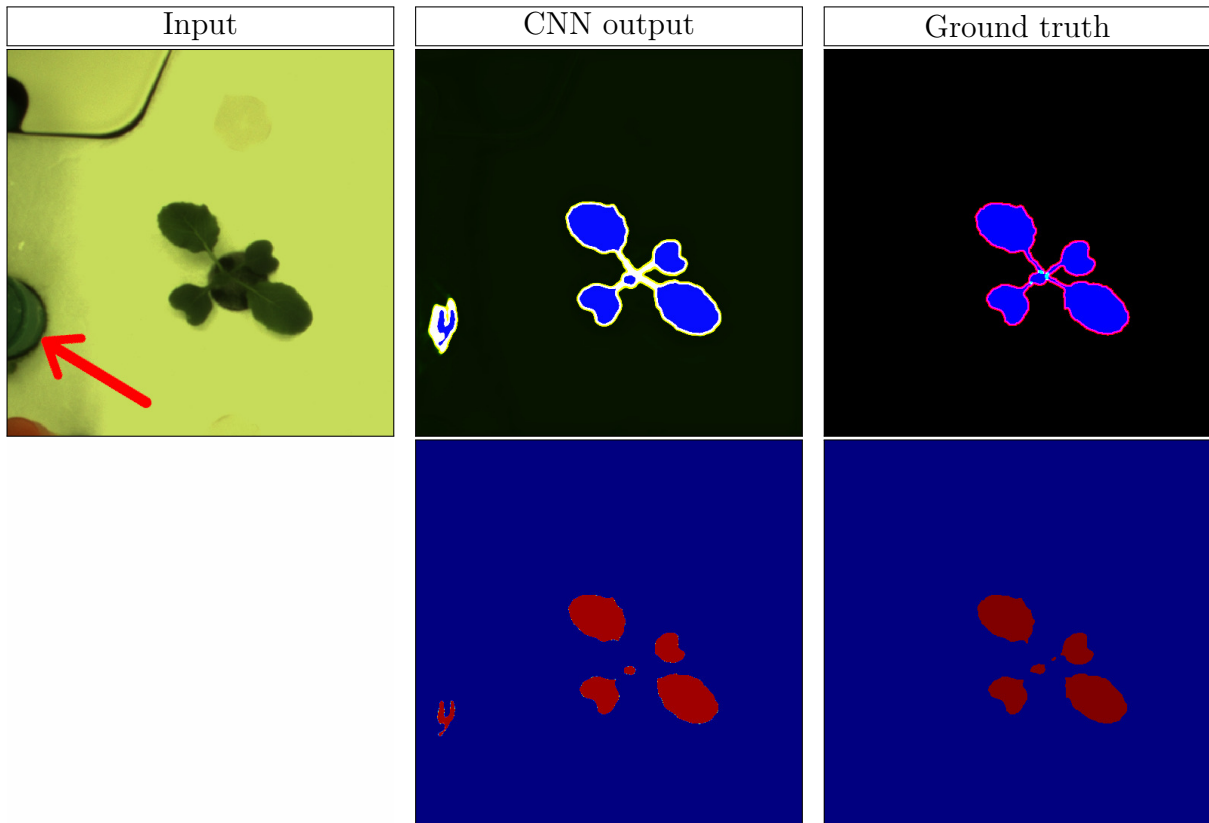


Figure 5.14 – Visual comparison of Komatsuna dataset after the training, the red arrow in the input image shows a green bottle

As shown by the *DiffFG* score, our method mostly detects the right number of leaves. This score is small and negative and shows that few leaves are split, which occurs when a big leaf mostly overlaps a smaller one. A visible bottle also has an impact on the other metrics, shown in figure 5.14. However, most of the errors for the *BestDice* comes from under and over-segmentation. The under-segmentation occurs on the leaves stems. The stem that connects the leaf to the plant is usually not well detected, and this can play a significant role in lowering the *BestDice* and *SymmetricBestDice* scores. The stem can be undetected (under-segmentation) or may be identified as another leaf (over-segmentation). It can also cover some small leaves and divide them in two, which is expected in pixelwise segmentation. To benchmark our study, the following table 5.2 reports the results of few previous studies using the same metrics.

Study	SBD	AbsDiffFG
Our method’s results (2021)	0.8565	0.5222
[Ward and Moghadam, 2020] ward 2020	0.7776	-
[Gomes and Zheng, 2020] gomes 2020	0.7700	0.8800

Table 5.2 – Comparison of the ResNet 101 based on Detectron II framework of facebook and the UPGen based on Mask-RCNN. Our solution is the most effective in both metrics

As comparison our results tackle one of the states of the art ResNet 101 based on Detectron II framework of facebook [Gomes and Zheng, 2020]. And the proposed UPGen methods proposed by [Ward and Moghadam, 2020].

5.4.2 Leaf Segmentation Challenge dataset

The second dataset is interesting for its controlled illumination conditions. Unlike the Komatsuna dataset, this one includes a greater diversity of leaf quantity, between 2 and 16 leaves. This is important to show the robustness of our loss function. In the same regard as the Komatsuna section, this dataset does not contain small and thin leaves. Thus third class was replaced by a foreground class, which is also used by the competition evaluation. In addition, the testing dataset is not publicly available, thus a separated evaluation was performed on the online competition website¹. The training and validation are defined by splitting the publicly available dataset. This step is used to reduce the over-fitting and retrieve stable parameters. The next table summarizes the number of images for each training, validation and test datasets.

Dataset	A1	A2	A3	A4	Total
Training	102	25	22	499	648
Validation	25	6	5	125	161
Test	33	9	56	168	266

Table 5.3 – Number of images for each sub LSC dataset

1. competitions.codalab.org/competitions/18405

The following Table 5.4 shows the overall evaluation on this dataset (merged A1, A2, A3, A4). The test results can be found in the online leader-board. The results are fairly the same between training, validation and test datasets in term of *FGBGDice* and *BestDice*. However a notable divergence with the testing dataset for the others metrics is visible. The *DiffFG*, and *AbsDiffFG* for training and validation indicate an over-segmentation, while the test dataset highlights an under-segmentation. This over-segmentation probably enhances the *SBD* score while the *BestDice* remains stable.

Metric	Training	Validation	Test
FgBgDice	0.9489	0.9522	0.9489
BestDice	0.7659	0.7707	0.7608
SBD	0.7608	0.7655	0.8047
DiffFG	-1.9634	-2.1223	3.5628
AbsDiffFG	2.1707	2.3050	6.1257
Number of images	648	161	266
Mean leaf count	13.8390	13.1463	-
NAbsDiffFG	0.1568	0.1753	-

Table 5.4 – Evaluation metrics of the overall LSC dataset

As discussed in the materials and data section, this database is composed of four sub-databases with different cameras and plants. The learning was done independently on each of them, but they were merged for the presentation in the previous table. The test on the online evaluation platform returns the results for each of them, summarized in the table 5.5. First of all, the sub-datasets A2 and A3 contain a very small amount of images, respectively 25 and 22 for the training sets. Moreover the A3 dataset is composed of images of size 2048×2048 re-scaled to 512×512 . This imply a huge loss of information, especially on the leaf boundaries. In addition, this A3 dataset contains hard shadows. These three factors explain most of the errors for this dataset. The quality, acquisition conditions, and plants of the other three data sets (A1, A2, A4) are similar. Only the background and the amount of images mainly changes, which is also reflected in the evaluation score. The figure 5.18 also shows unlabeled leaves in the center of the plant, moreover they are even defined as background.

Metric	A1	A2	A3	A4
FGBGDice	0.9641	0.9294	0.9692	0.9416
BestDice	0.6905	0.6944	0.6783	0.7964
SBD	0.8047	0.7895	0.7317	0.8294
DiffFG	2.0909	1.7778	21.2857	0.0892
AbsDiffFG	3.0000	3.1111	14.9285	1.6250

Table 5.5 – Evaluation metrics of the independent sub LSC dataset

For all sub-dataset the *FGBGDice* score is slightly less than the Komatsuna dataset. Most pictures of the A1 present wide area of green moisture on the ground, visible on the Input inside the figure 5.15. A2 contains few very small weeds and few small spots of moisture, but the performances of *FGBGDice* is mainly due to the quantity of images for the learning, as showed by the figure 5.16 boundaries are miss-classified. For the same reason, it is also visible for the A3 dataset shown by the figure 5.17. It’s also important to notice that leaves from an outside plant are visible in few images and detected by the algorithm, however these leaves are unlabelled, resulting an over-segmentation.

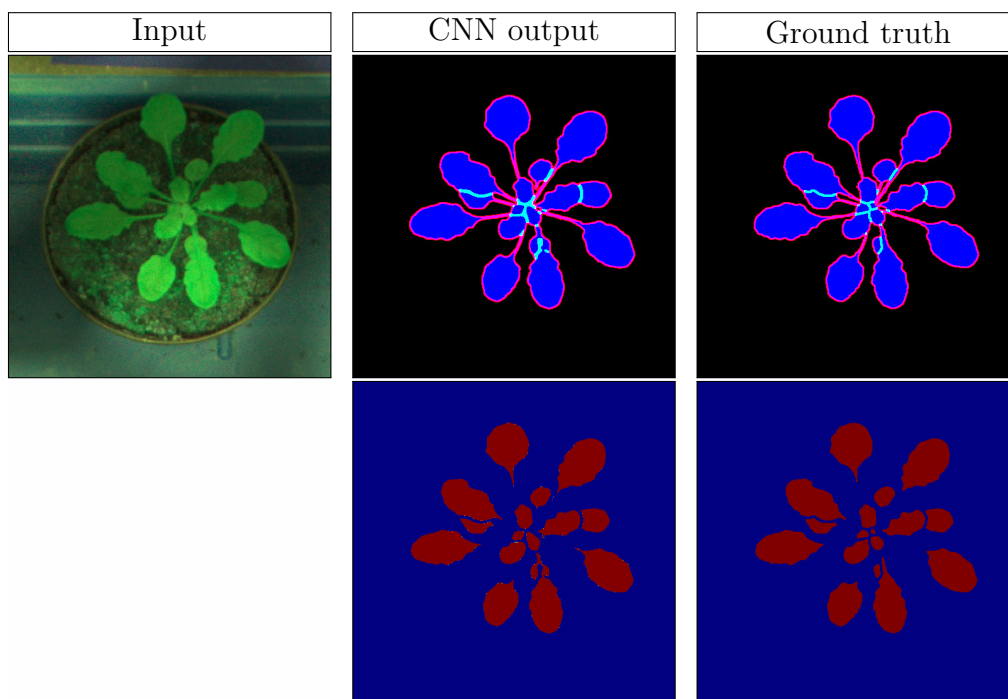


Figure 5.15 – Visual example of LSC results for A1

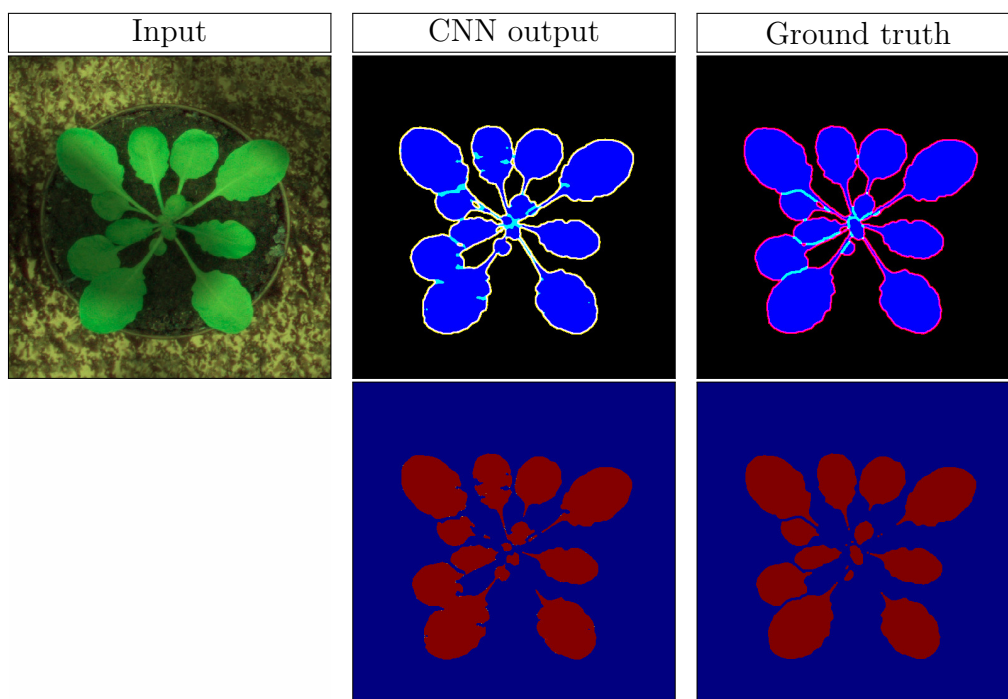


Figure 5.16 – Visual example of LSC results for A2

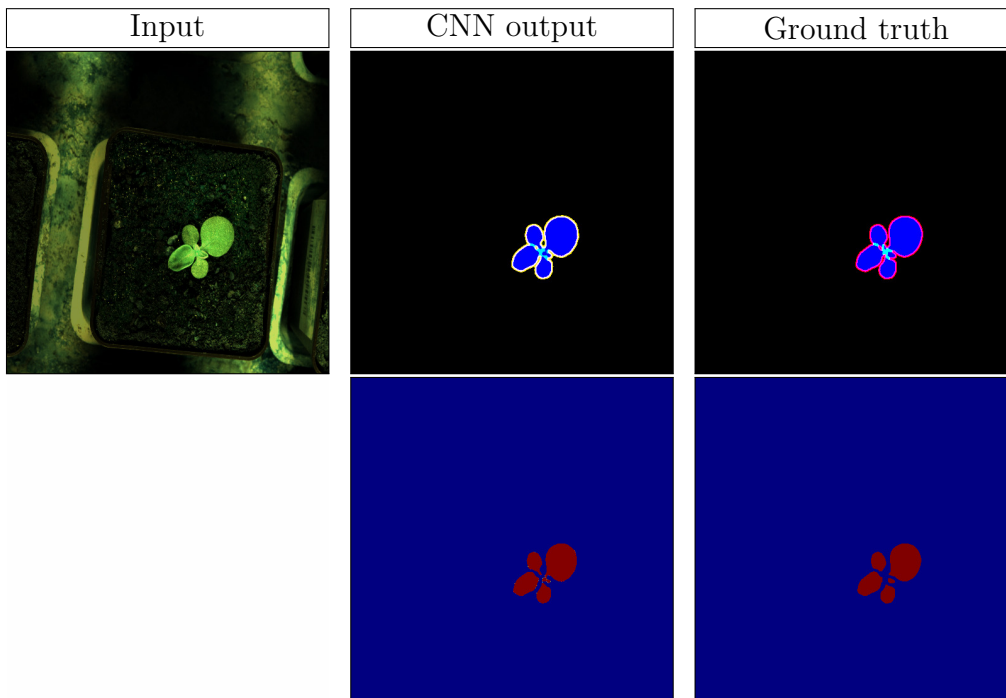


Figure 5.17 – Visual example of LSC results for A3

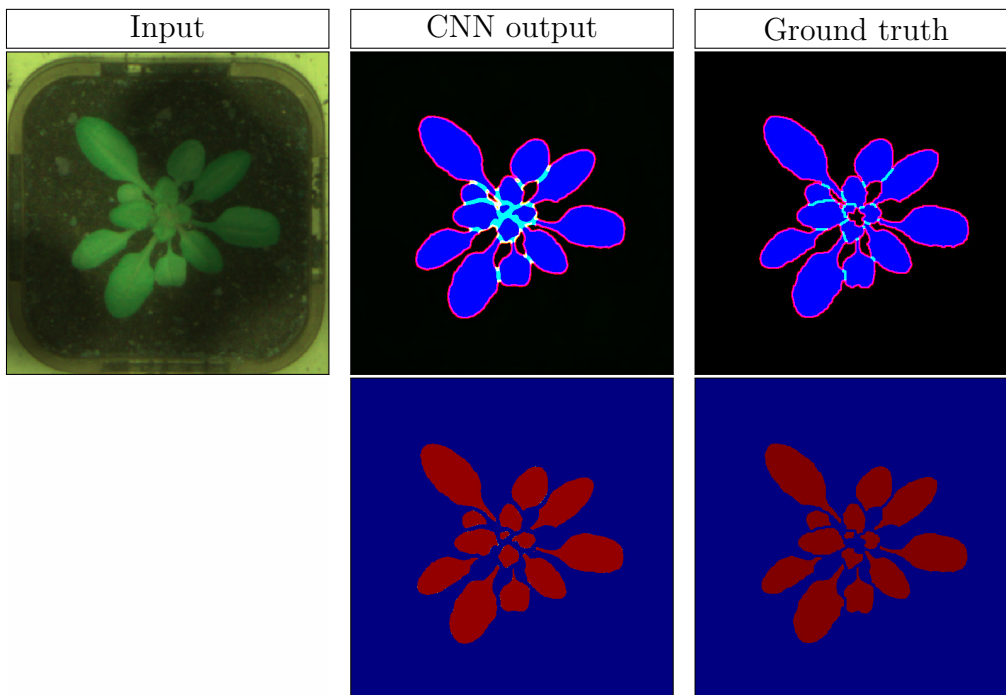


Figure 5.18 – Visual example of LSC results for A4

The following table 5.6 reports method’s results of different previous studies, to benchmark our method’s results on this LSC dataset.

Study	SBD	AbsDiffFG
[Ward and Moghadam, 2020] ward 2020	0.8800	-
[Kulikov et al., 2018] kulikov 2018	0.8040	2.00
[Gomes and Zheng, 2020] gomes 2020	0.7700	0.88
Ours method’s results (2021)	0.7608	6.12
[Pape and Klukas, 2015] pape 2015	0.7440	2.60
[Scharr et al., 2016] scharr 2016	0.6830	3.80

Table 5.6 – Comparison of our solution with state-of-the-art challengers in this dataset.

These results show that the studied method is less efficient than [Ward and Moghadam, 2020, Kulikov et al., 2018] and [Gomes and Zheng, 2020]. Due to the small size of the training sample (102, 25, 22, 499 images for data subsets A1, A2, A3, A4, respectively), it can be assumed that this is due to the increase in data they use in order to expand their data set. It seems the data augmentation based on Perlin noise is insufficient. Nevertheless, our method is better than the following two : [Pape and Klukas, 2015, Scharr et al., 2016], since they don’t use any data augmentation. However it can be noted that the *AbsDiffFG* value for the studied method is much higher than for the others. This implies an important over-segmentation, resulting in particular from the quality of the A3 dataset as it can be seen in the table 5.5.

5.4.3 Airphen dataset

The last one presented is our multispectral dataset, which contains variable acquisition conditions (sunny, cloudy, rainy, etc.), a variable number of leaves (from a few to hundreds), and contains very small leaves that touch or overlap the others. The next table 5.7 show the results of our method applied our dataset.

Metric	Training	Validation
FgBgDice	0.9791	0.9785
BestDice	0.6744	0.6678
SBD	0.6670	0.6638
DiffFG	-41.3583	-49.4667
AbsDiffFG	46.7500	53.2333
Number of images	240	60
Mean leaf count	265.83	295.80
NAbsDiffFG	0.1758	0.1799

Table 5.7 – Evaluation metrics for the Airphen dataset

It can be seen that the *FGBGDice* score shows adequate soil/vegetation segmentation, as demonstrated in our previous study [Vayssade et al., 2021]. The scores *DiffFG* and *AbsDiffFG* show the presence of over-segmentation. This is due to two issues : the presence of small leaves and the advanced stage of mustard development. Indeed, the method sometimes confuses small leaves and large leaves, which implies in some cases under-segmentation for monocotyledons or over-segmentation for dycotyledons, probably generated by an imprecise annotation of classes. In the second case, this is due to the proposed loss function that forces over-segmentation, resulting in the detection of leaflets instead of leaves in the case of advanced mustard develop-

ment, as showed in the figure 5.19. These errors imply that the *BestDice* and *SBD* scores are not as good as on the Komatsuna dataset and LSC Challenge, studied previously.

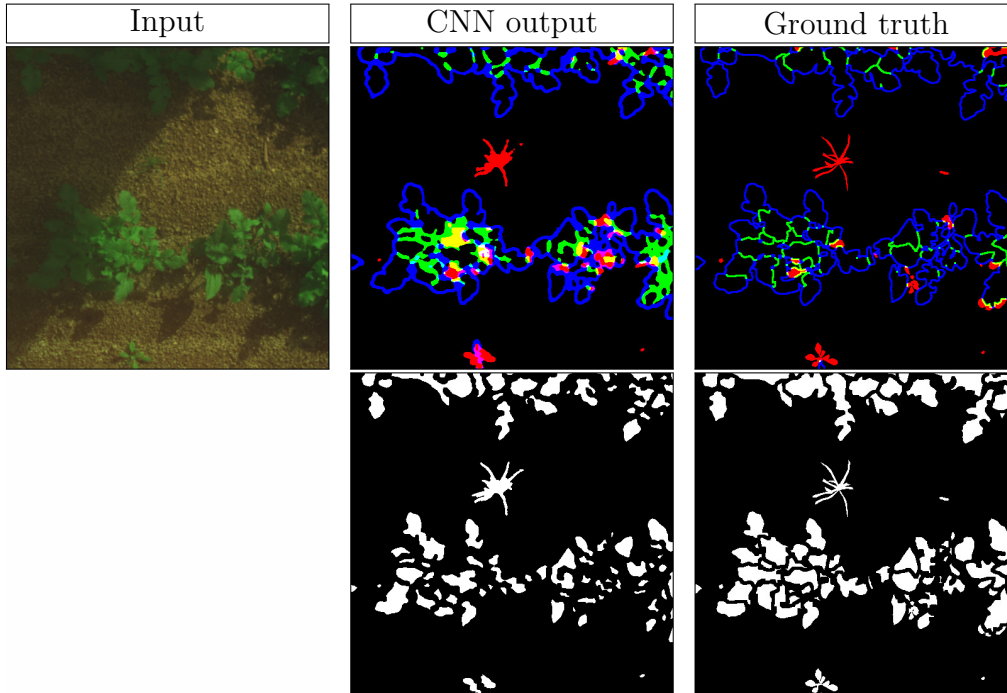


Figure 5.19 – Example of the visual results for the Airphen dataset

To show the difference and complexity of this dataset, the method developed by [Ward and Moghadam, 2020] was also learned. This method is based on the Mask-RCNN. However, the method uses data augmentation based on a library that only supports 8-bit unsigned integers. Thus, data augmentation was disabled because the dataset uses a 32-bit float format. The following table 5.8 shows these results.

Metric	Training	Validation
FgBgDice	0.6266	0.6111
BestDice	0.2271	0.2157
SBD	0.2266	0.2149
DiffFG	-186.0292	-217.0667
AbsDiffFG	186.5542	217.1333
NAbsDiffFG	0.7018	0.7341

Table 5.8 – Evaluation metrics for the Airphen dataset using Mask-RCNN

These results show that the method proposed by [Ward and Moghadam, 2020] does not correctly detect leaves in a dense foliage environment. The *FgBgDice* score shows that the soil/vegetation discrimination is weak. As well as the *BestDice* and *SBD* scores which shows poor detection of individuals segmentation mask. The *DiffFG* and *AbsDiffFG* scores confirm these results and show a large amount of undetected elements. As announced in the introduction, the methods based on object detection uses bounding box regressions and non-max suppression which strongly limit the detection in dense environment. Moreover the part of the network allowing to obtain the segmentation mask uses a fixed low resolution resulting in coarse segmentation masks. As shows in the figure 5.20.

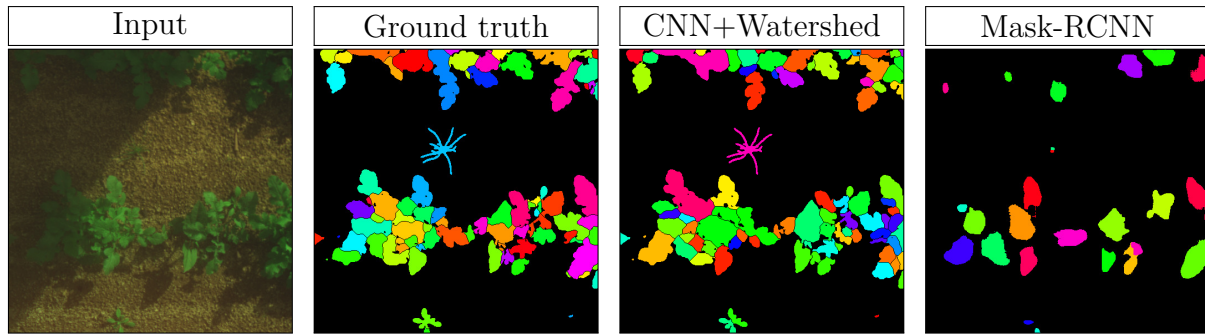


Figure 5.20 – Example of the Watershed output versus Mask-RCNN for Airphen dataset

5.5 Discussion

In this study, we used different datasets. The Komatsuna dataset is composed only of mustard leaves at different stages of development acquired indoors under controlled acquisition conditions with a single camera. There are between 3 and 6 leaves per image for an average of 4.9 leaves. The LSC dataset is decomposed into three sub-datasets of Arabidopsis-Thaliana and one of Rosette plants all at different stages of development. They are acquired indoors under controlled acquisition conditions with their own camera. There are between 2 and 16 leaves per image for an average of 13.5 leaves. Our dataset is composed of bean plants and weeds. Beans are at a single stage of development while weeds are at different stages of development. The acquisitions were made outdoors under uncontrolled conditions. There are between 4 and 777 leaves per image with an average of 271.83 leaves. This dataset is now available online² and may help other studies working on leaves segmentation in natural and complex situations.

We are interested in the metric $NAbsDiffFG$ to compare the performances of our method on the different datasets. We see that for our dataset it is 0.1758, 0.1799 for the LSC dataset and 0.1050 for the Komatsuna dataset. We thus obtain good results for the Komatsuna dataset which is the simplest, we obtain an average result for our dataset which is the most complex and finally weaker results on the LSC dataset. This allows us to deduce that the lighting conditions as well as the number of leaves do not influence the performances. On the other hand, the use of several cameras, as on the LSC dataset, could be at the origin of the decrease in performance due to a variation in spatial resolution from one camera to another. We deduce that our method, designed to be used on complex data such as our dataset, gives results comparable to classical methods on other types of datasets.

The CNN used in our method deliberately gives a coarse segmentation in order to maximize the detection and separation of large leaves. The smaller ones are therefore poorly detected (over-segmentation or undetected). These results are reflected in the fact that there is little over-segmentation and under-segmentation of large leaves due to the loss function introduced which takes into account the number of elements detected and their surface. However two problems are raised with the use of this loss function. First of all, it is subject to error jumps in order to avoid merging objects, which makes its optimization difficult and thus requires several learning sessions for an optimal result.

2. <https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/JMKP9S>

On the other hand, avoiding merging causes over-segmentation in some cases, such as the mustard in our dataset that detects leaflets. At this time we cannot say whether this is problematic. It is possible that the leaflet scale is more relevant in terms of segmentation than the leaf for advanced stages of plant development.

With the watershed used next, we improve the coarse detection provided by the CNN by extending the detected areas to the ideal soil/vegetation segmentation mask defined by the method developed in a previous study [Vayssade et al., 2021]. This step allows the identification of the smallest leaves. Although the smallest elements, with a diameter of 0 – 2 pixels, are still difficult to detect due to spectral mixing [Louargant et al., 2017]. This is an important result as it allows to configure an acquisition system for a specific minimum size of leaves to detect. In this work, the remaining small leaves are already segmented as vegetation and can be considered as stable due to their sizes. Some features are still relevant to be extracted (such as the size of the leaf and its distance from the crop row) and should help in discriminating crop from weeds, other may suffer from the spectral mixing and become irrelevant (spectral signature, texture...).

There is still an important type of error. Indeed, a single pixel can be at the origin of errors leading to a bad fusion between two areas. It would therefore be interesting to study structural analysis methods in order to overcome these deficiencies. A possible method would be to vectorize the contours and look for an algorithm to reconnect or split the contours for instance according to convexity singularities.

The main interest of the proposed method is its efficiency on mixes of plant species, acquired with natural light. It will be integrated in a processing chain dedicated to the discrimination of crops and weeds in agronomic scenes. Indeed, the detected leaves can be classified according to a large set of criteria (spectral signature, morphological characteristics, texture, distance from the crop row). The underlying hypothesis is that these criteria are more stable at the scale of the leaf than at the scale of the plant. However, this approach has certain limitations when leaves overlap others, as the detected shapes would be heterogeneous. Multifoliate leaves could also be difficult to characterize. In that case, detecting leaflets instead of leaves may be more relevant.

5.6 Conclusion

The presented work shows that the CNN network enhances the quality of the segmentation based on multispectral images. Indeed, the background is well removed due to the upstream network with IBF and UFA modules with an accuracy of 95 – 98% of mIoU. Our method is effective in the majority of cases, such as the segmentation of unifoliate leaves like *Arabidopsis-Thaliana* and early developmental stages of plants. However, our method is not effective in the advanced stages of plant development, especially on mustard which has highly segmented leaves. In this case our method detects leaflets instead of leaves. Their identification is nevertheless relevant for the phenotyping or classification of weeds.

We have seen that the developed method presents better SBD performances $+ [1.66 - 7.89]\%$ compared to studies that do not use data augmentation. However, on small datasets, it presents lower SBD performances $- [4.32 - 11.92]\%$ compared to recent studies that use it. To make the method more consistent we will therefore focus on data augmentation. To improve the detection of classes it would be interesting to improve the upstream modules with attention mechanisms.

This method would allow to correct the illumination of the images. As well as the use of a *MIRNET* [Zamir et al., 2020] would allow to eliminate the noise. Finally, to improve the downstream modules, the use of *Selective Kernel Convolution* [Zamir et al., 2020] would allow a better fusion of the multi-scale information instead of *Universal Function Approximator*. We will then try to minimize the under-segmentation detected in our method, by using the *Deep Watershed Transform for Instance Segmentation* method [Bai and Urtasun, 2017].

A small performance loss for the developed method is seen on our dataset compared to Komatsuna and LSC dataset. It would be interesting to evaluate the reason(s) leading to this loss as it may come from the uncontrolled acquisition conditions, the multispectral nature of the images, the size differences between leaves or the important number of leaves in each images.

In all cases, this approach should lead to an enhancement of features extraction which may improve crop/weed classification. These increased performances would lead to a better tracking of the weed flora. These algorithms show promising and robust results in natural acquisition conditions. The segmentation results are obtained fast enough to be used in a real-time crop/weed discrimination setting and could be embedded on a Unmanned Ground Vehicle (UGV) for quick and localised intervention. Applied on images acquired from an Unmanned Aerial Vehicle (UAV) this could be used for tool assisted plot management to help farmer in their decision making.

5.7 Further research

We defined the CNN architecture from the state of the art, adding components in an increasing development cycle. Thus, we have not established the contribution of each block on this paper. The first 3 modules (iit, ibf, ufa) and the last sprb modules have been widely developed in our previous study about vegetation segmentation, each module have showed a contribution. But further research is undergoing to show the impacts of different modules on the upstream and downstream of the network, and the proposed loss function, and loss weights. This task must be automated through an hyper-parameter optimization process which we didn't explore yet. And the watershed algorithm used to refine the output must be added to the neural network, and part of this optimization process. Finally more advanced data augmentation could be explored.

Data availability

The Airphen dataset used in this study have been released and it's publicly available at <https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/JMKP9S>, using Creative Common CC0 1.0 Public Domain Dedication licence

Funds & Conflicts of Interest

This project has received funding from the European Union's Horizon 2020 research and innovation program [grant agreement ID : 727321] (project acronym : IWMPRAISE) and from the French National Research Agency Challenge RoSE [grant agreement ID : ANR-17-ROSE-0002] (project acronym : ROSEAU). The authors declare no conflict of interest.

5.8 Conclusion de chapitre

Les résultats de cette étude montrent que l’approche proposée pour la segmentation des feuilles est intéressante sur plusieurs jeux de données accessibles en ligne et sur nos données. Elle génère d’excellents résultats sur les données de Komatsuna. En revanche, la méthode n’est pas parmi les meilleures lorsqu’elle est appliquée au jeu de données du LSC (Leaf Segmentation Challenge), mais elle reste quand même compétitive. Pour nos données multispectrales, dans le cadre du challenge RoSE, les résultats montrent qu’une approche traditionnelle, telle que les mask R-CNN qui reposent sur une utilisation de l’algorithme NMS ne permettent pas de détecter efficacement les individus en environnement dense. Contrairement à la méthode qui a été développée dans cet article. Cependant, bien que ces résultats représentent une avancée dans le domaine du “pixelwise instance segmentation”, des recherches complémentaires doivent être entreprises.

Concernant l’échelle de segmentation (i.e. la feuille), ces résultats suggèrent qu’une entité légèrement plus fine serait préférable d’un point de vue segmentation. En effet la visualisation des résultats sur les données du challenge RoSE, montre que les feuilles unifoliées/unilobées sont bien détectées, ce qui n’est pas le cas des feuilles multifoliées ou multilobées. Ainsi, l’échelle foliole serait probablement plus opportune.

Le présent chapitre 5 et le chapitre précédent 4 constituent l’étape de segmentation de notre pipeline. Maintenant, que l’on est capable de détecter les individus, nous nous intéresserons à l’extraction de propriétés et la classification, comme nous pouvons l’observer sur la figure ci-dessous.



Rappel de l’architecture de notre chaîne fonctionnelle
ici les premières étapes en vert sont terminées, et en jaune les étapes suivantes étudiées

6 Extraction de propriétés

Ce chapitre aborde le cœur de cette thèse “Approche multicritère pour la caractérisation des adventices”. C’est-à-dire l’extraction et l’exploitation de propriétés permettant la discrimination culture/adventices. Ces dernières, reposent sur l’exploitation de divers descripteurs comme les textures, les formes, ou les propriétés spatiales (présentées en section 2.4). Ces descripteurs sont couramment utilisés [Griepentrog et al., 2005, Gao et al., 2013, Xue et al., 2017, Louargant et al., 2018, Su, 2020, Mekhalfa and Yacef, 2021, Wu et al., 2021a] pour caractériser et classer les plantes détectées dans une image. En revanche, les études utilisant ces descripteurs sont basées sur un petit ensemble de propriétés – souvent moins de 20 – ainsi que sur des cultures et des matériels d’acquisition non concordants entre les études. Ainsi, ces études ne permettent pas une estimation précise des potentiels des différents descripteurs, appliqués à un sujet donné. Il est donc difficile de trouver la meilleure combinaison de critères et leur impact sur la qualité de la classification à partir d’une bibliographie.

Parallèlement, dans le chapitre précédent 5, concernant la segmentation, nous avons défini une approche alternative, reposant sur la détection des feuilles et non des plantes. Cela permet la caractérisation du couvert végétal d’une façon plus fine que ce qui est actuellement établi. L’objectif de l’étude qui suit est donc d’analyser, d’intégrer et d’évaluer les spécificités de chaque critère extractible à l’échelle de la feuille, à partir d’images prises en conditions terrain.

Un ensemble de 25 types de propriétés a été défini, dont 3 spatiaux, 8 de forme et 14 de texture. Chaque type de propriété regroupe un sous-ensemble de propriétés partiellement corrélées entre elles. Par exemple le type “shape particle” contient un sous-ensemble de 18 propriétés de forme, tandis que le type “texture haralick” est composé de 416 propriétés de texture. Le total de propriétés extraites dans cette étude monte à 3545. Étant donnée la quantité de propriétés, deux problèmes se posent :

1) Comment optimiser les propriétés comportant des paramètres ? Par exemple calculer des propriétés comme les DoG (Difference Of Gaussian) nécessite de fixer 4 paramètres préalablement (sig, lm, gm, ps). Ceci est un problème d’optimisation des hyperparamètres et dans l’exemple précédent, les 4 paramètres sont dans un espace continu, ce qui signifie qu’une recherche exhaustive ne peut être établie. Pour réaliser ces optimisations, des algorithmes plus complexes doivent être utilisés pour échantillonner l’espace des paramètres efficacement et automatiquement. C’est le premier point qui est abordé dans l’étude qui suit.

2) Comment sélectionner efficacement le meilleur sous-ensemble ? Pour répondre à cette question, les chercheurs dans le domaine de l’agriculture de précision utilisent principalement une métrique pour estimer dans quelle mesure une caractéristique est corrélée au problème, puis les meilleures caractéristiques sont utilisées. Cependant, il peut exister une meilleure combinaison, par exemple, si l’on combine la distance aux rangs et une caractéristique de texture, l’efficacité pourrait être supérieure à celle de deux critères de texture dont chacun est plus discriminants que le rang, car la distance au rang est probablement complémentaire à une propriété de texture. C’est en tout cas ce que suggèrent les études utilisant une distance au rang [Louargant et al., 2018]. C’est le second point qui est abordé dans l’étude qui suit.

Weed discrimination at leaf scale : towards the characterization of the vegetation cover : A large comparison of shape, spatial and textural features

Authors Jehan-Antoine Vayssade, Gawain Jones and Jean-Noël Paoli

Publication 08/01/2022 – Computer and Electronics in Agriculture – **under-review**

abstract In the context of crop and weeds discrimination, different methods are used to detect and classify plants from an acquisition system. Various estimators and descriptors like textures, shapes, spatial properties, or vegetation indices are commonly used to characterize plants within an image.

However, the available studies are based on disparate criteria, plants, and acquisition materials. The criteria are often defined by parameters set empirically according to the situation, and the processing scale (cluster, plant, leaf, pixel) has a significant impact on classification performance. Thus, these studies do not allow an accurate estimation of the potential of criteria combinations, if applied to a new study. In addition, the detection of plants remains a difficult problem in image processing. Indeed, plant segmentation approaches are often based on unreliable methods when the coverage rate increases. Currently, an emerging solution is leaf segmentation, allowing to characterize dense vegetation cover.

Thus, the objective of this study is to : (1) experimentally evaluate the discrimination potential of each criterion at the leaf scale, using images taken in field condition ; (2) optimize the parameters of these criteria ; and (3) determine the best combination of criteria to use.

A literature review is conducted to determine the set of criteria that could be used. A huge set of 3545 criteria is studied. For criteria defined by parameters, an hyperparameter optimization algorithm is involved. Then, an algorithm is defined to select the best subsets of features. This algorithm evaluates the performance of each subset of criteria on a ground truth dataset, using an appropriate performance metric. Finally, a classification of the vegetation cover is proposed, using the best performing subset. Results show the importance of selecting a smaller set of properties (at most 20 features among the 3545 available) and associating different feature types (for instance spatial with textural and morphological features).

Keywords leaf classification ; crop and weed discrimination ; remote sensing ; multi-spectral

6.1 Introduction

Crop and weed discrimination is a long-standing concern in modern agriculture, as weeds are generally considered a nuisance since they have a possible negative impact on crop yield or harvest quality. To have a reliable knowledge of the risks of yield loss over time, locally acquired experiences on 110 weed control trials under comparable conditions were conducted in France between 1993 and 2015 on three major annual crops (soft wheat, rapeseed and sunflower). A reduction of 25 q/ha, 4 q/ha and 3.5 q/ha was respectively observed [Cordeau et al., 2016]. Weeds are thus competing with their neighbours through competition for nutrients (nitrogen, potassium, phosphorus, carbon), light and water.

To face this problem the main weed control approach is based on the application of herbicide with sprayers. The methods have evolved from uniform chemical weeding of the plot to localized weeding in precision agriculture and the use of new technologies and digital tools. In this impetus, today's precision agriculture is seeking to discriminate even finer elements like the plant in real time. It allows to weed out undesirable individuals, to treat diseased crops or to supply nutrients for crops to reduce the use of synthetic products and the ecological impact of these products.

Current research on crop/weed discrimination are based on criteria extraction and exploitation and provide good discrimination results, up to 77-97 % [Ahmed et al., 2016] depending on the criteria, acquisition conditions and observed plants. The extraction and exploitation of such criteria should be included in a computer vision pipeline that follows the key steps of image analysis including preprocessing, segmentation [Vayssade et al., 2022], feature extraction and classification [Perez-Sanz et al., 2017]. Figure 6.1 illustrates this general workflow.

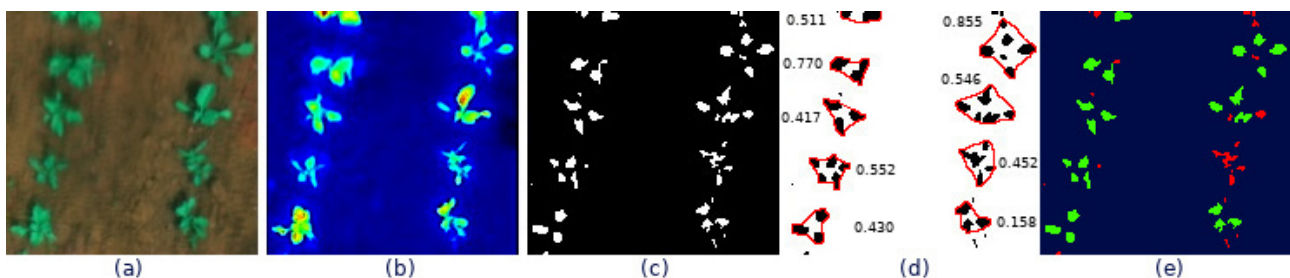


Figure 6.1 – key steps of image analysis for crop/weed discrimination with (a) acquisition enhancement (b) vegetation indices (c) instance segmentation (d) feature extraction (e) classification

Major reviews of weed mapping techniques have been carried out by [Monteiro et al., 2019], [Gao et al., 2018], [Mostajer Kheirkhah and Asghari, 2019] and [Wang et al., 2019a] and show multiple approaches to weed mapping and feature extraction. For example, during the extraction stage, the criteria can be classified according to the nature of the information. We can distinguish (i) shape criteria which are based on different representations of the contours, [Ahmed et al., 2012, Pérez et al., 2000], (2) spatial criteria based on the distance of the individual to another element of the image, such as the distance to the nearest crop row [Louargant et al., 2018], (3) texture features based on the analysis of the spatial distribution of the pixels (may include the soil) [Rojas et al., 2017] and, finally (4) spectral indices at specific points on the surface can also be used [Liu et al., 2019], which could also be considered as a subset of texture features. However, each type of criterion has different effectiveness and limitations. Spectral

and texture features are more affected by illumination variations, while shape features are affected segmentation quality. Moreover, because of these limitations, studies usually combine these types of features to take advantage of all the benefits.

However, although these criteria seem to be efficient, they may lose effectiveness in real conditions because they are limited by the previous stages of the pipeline like in Figure 6.1. In particular by the plant instance segmentation step, which is still a scientific challenge despite the advances in deep-learning. In this context, we have proposed in a previous study to detect not the whole plant but the leaves which gives promising results in terms of instance segmentation and allows to work in dense foliage [Vayssade et al., 2022]. Moreover, the leaf offers more stable criteria : there is less intra-individual variability than with a plant. For example, depending on the growth stage, the plant will contain a variable number of leaves with different orientations and sizes, these leaves may also have a partial overlap between them leading to an infinite number of cases. Thus the extracted criteria are viable for a given stage of development, for given crops and specific acquisition conditions.

We will see that some of these issues do not occur at the leaf level. The extracted criteria are also more understandable at leaf scale : shape criterion, such as length or perimeter is less subject to interpretation on the leaf level than on the plant level. Texture criterion, such as gradients, gives the veins of the leaf, but there's no obvious correspondence at the plant scale. Thus the study of these criteria at the leaf scale could be promising to discriminate leaves between crops and weeds. In addition, such leaf-scale criteria have been used successfully for tree species discrimination, [Cerutti et al., 2013, Cerutti et al., 2014] showing an accuracy up to 90% with only shape criteria at leaf scale for species classification.

Therefore, this study focuses on two points, evaluating criteria that could be used at the leaf level and determining the best criteria combination to discriminate leaves between crops and weeds in field condition. Through this study, we also propose a Python framework allowing the extraction of these criteria available at github.com.

6.2 Material and data

6.2.1 Experimental plot

Data is acquired at the site of INRAE in Montoldre (Allier, France, at 46°20'30.3"N 3°26'03.6"E) within the framework of the ANR Challenge RoSE in 2019, visible on Figure 6.2. The objective of the Challenge is to objectively compare the solutions proposed by participants [Avrin et al., 2020, Regnier et al., 2021]. Within this context, the challenge provides an evaluation plan to contestants and a set of experimental plots of bean and corn plants. In addition various natural weeds (yarrows, amaranth, geranium, plantago, ...) and sowed ones (mustards, goosefoots, mayweed and ryegrass) is managed to compare performances.



Figure 6.2 – Aerial view of the experimentation plot located in Montoldre (now INRAE)

6.2.2 Multispectral camera

The images were acquired with the Airphen (Hyphen, Avignon, France) six-band multispectral camera (Figure 6.3). This is a multispectral scientific camera developed by agronomists for agricultural applications. The camera has been configured using the 450/570/675/710/730/850 nm bands with a 10 nm FWHM. The focal length of each lens is 8 mm. The raw resolutions for each spectral band are 1280x960 px with 12 bit precision. Due to the conception of the camera, spectral images are not aligned in space. A registration method based on previous work for this camera, with a registration accuracy down to sub-pixel is used [Vayssade et al., 2020]. After the registration, all spectral bands are rescaled to 1200×800 px and concatenated to channel-wise where each dimension refers to a spectral band.



Figure 6.3 – The Airphen six-band multispectral camera

6.2.3 Image acquisition and annotation

From the presented experimental plots a set of images were acquired. The camera is attached in front of an hybrid autonomous tractor called “*TREKTOR*” launched by *SITIA* (Bouguenais, France) in 2019. The camera is setup to have a top-down view of crop rows, thus it is placed at the end of a 2 m pole in front of the platform allowing to remove visible part of the robot and at 1.8 m from the ground to see all three crop rows. Crops and weeds were between phenological state 3 and 4 which means they had between 2 and 6 leaves. The ground truth is manually defined on images by experts with polygons around each leaf boundaries. In addition, polygons contain a crop/weed classification label. These annotation are performed using the VIA annotation software [Dutta and Zisserman, 2019] and a total of 300 images of bean were annotated, 170 in June and 130 in October. This dataset is now available online¹ and may help other studies working on leaves classification. In addition the crop rows for each image are also manually set. The total number of leaves is 90 538, with 31 524 for crops and 59 019 for weeds.

1. <https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/JMKP9S&version=1.0>

6.3 Methodology

The objective of this paper is to extract a wide variety of features and evaluate their performances for classifying leaves into crops and weeds, in addition to defining the best subset of feature space. The methodology followed in this study is as follows :

1. In order to evaluate criteria that could be used at the leaf level to discriminate between crop and weed leaves in field condition, a literature review was conducted. These criteria are organized into shape, spatial and textural features and presented in the section 6.3.1.
2. Most of the features have parameters that impact the extraction and classification performances of these features. To solve this problem, we propose to optimize these parameters by using an hyperparameter optimisation algorithm to efficiently sample the parameter space, it is described in section 6.3.2.
3. The results of the literature review indicate that there is no consensus on which of these criteria are the best for discriminating between crop and weed (at plant or leaf scale). In order to determine the best combination of criteria to use in discriminating between crop and weed leaves in field condition, an algorithm to select the best subsets of features is used and described in section 6.3.3.
4. The performance of each method is evaluated on the ground truth dataset, using Area Under the Curve (AUC) as the performance metric. The best performing feature subset is presented to classify leaves into crops and weed. These subsets are extracted for each type of feature (locally) and among all feature (globally), this is available in the results section 6.4.
5. Finally a classification of the vegetation cover is proposed using the best performing features subsets and the results are discussed in section 6.4.3.

6.3.1 Literature review

In computer vision, shape, spatial and textural features are commonly used for describing and classifying objects. In this section we will present a review of these features. To facilitate the extraction of features and their merging, they have been programmed using the OOP² paradigm in Python. Each feature is thus named according to its type (shape / spatial / spectral) and internal properties. Thus the following subsections present each of them according to its type.

6.3.1.1 Shape Criteria

Shape properties are based on a structural analysis of the contours of related regions, it mostly consists of detecting morphological and anatomical traits. The shape can be coded as a set of positions (pixels) that go around the boundary. From this description, various information can be used. This type of analysis is found particularly in granulometry [Hentschel and Page, 2003], geomatics [McGarigal, 1995], leaf classification [Cerutti et al., 2014] or even animal posture detection [Vayssade et al., 2019, Bonneau et al., 2021]. These criterion are the most intuitive, easy to implement, and unaffected by lighting. Each of the following paragraphs explain an extraction method as well as the resulting properties used for discrimination.

2. Object-Oriented Programming

Shape Ellipse [Watcharabutsarakham et al., 2012] have mentioned that the major and minor axis can be used to classify leaves. These properties can be obtained through an elliptical Hough algorithm as described by the article. From the retrieved ellipse, few more information can be compute such as the distance between the ellipse center and the detected shape center. Three factors of eccentricity, the directrix, the angle and the focal distance. Some of these properties were also used by [Lottes et al., 2016] for the classification of plant covers.

Shape Particle Based on the work of [Hentschel and Page, 2003], these shape metrics have been used to characterise and classify particles. [Vayssade et al., 2019] used it to classify goats activities within pasture, [Bonneau et al., 2021] to estimate the sow posture with an accuracy near deep-learning methods (94 % vs 96 %). This set of 18 characteristics are based on a variety of ratios of the Feret diameter, area and perimeter. Some of them are known as : Equivalent Diameter, Eccentricity, Aspect Ratio, Form Factor, Roundness or Convexity. Others are unnamed features. The Feret diameter is defined as (1) the longest distance between two points on the contour, and (2) the shortest distance across the shape. Some of these properties are also used by [Lottes et al., 2016] and [Saha et al., 2017].

Shape Solidity The solidity factor is widely used for weed plant classification [Ahmed et al., 2012]. The solidity factor is defined as a ratio between the segmentation mask area and convex hull area. Thus, it is bound from 0 to 1. This factor indirectly computes the holes area of the regions. When applied to weed discrimination, this factor allows to discriminate some specific species like carrot from others. When carrots are segmented, the segmentation mask is sparse, filled by a large amount of holes, due to leaves structures. This does not occur for dicotyledon plants, and it is less impacted by other monocotyledons weeds. At leaf scale, such criteria provide information about the contour (leaf teeth) as the inner leaf mask should be filled. This approach has been used for species classification [Thyagarajan and Raji, 2019] and weed discrimination [Lottes et al., 2016].

Shape Metrics This set of 8 properties comes from the documentation of the shape metrics tools of the University of Connecticut³, these are used for shape classification of lands [Angel et al., 2010] and building [Pászto et al., 2015]. Few properties, such as Cohesion, Traversal, Range and Viable Interior was not implemented due to time complexity. The others 8 implemented metrics are named : Proximity, Exchange, Spin, Perimeter Index, Depth, Girth, Dispersion and Detour.

Shape Fragstat The book by [McGarigal, 1995] shows some of the features used at the landscape, class, or parcel scale for classification or shape similarity measures. The fragstats documentation shows a small set of criteria⁴ that can be exploited at the leaf scale such as : Perimeter-Area Ratio, Shape Index, Fractal Dimension Index, Linearity Index, Related Circumscribing Circle, Contiguity Index and the Perimeter Area Fractal Dimension. These features have been selected as shape criterion in our study because they reflects shape complexity for a given spatial scale and correspond to : Perimeter-area-ratio, Shape Index, Fractal Dimension Index, Linearity Index, Related Circumscribing Circle, Contiguity Index and Perimeter Area Fractal Dimension.

3. http://clear.uconn.edu/tools/Shape_Metrics/method.htm

4. <http://www.umass.edu/landeco/research/fragstats/documents/Metrics/Shape%20Metrics/SHAPE%20METRICS.htm>

Shape Skeletonize The skeletonization is commonly used in shape recognition techniques. It transforms the input mask (i.e the shape) to a new structural representation of the shape [Erdem and Tari, 2010]. This representation captures part of internal hierarchy. As instance, the [Lottes et al., 2016] study use the skeleton length for weed discrimination. In this study more features are extracted from that structural representation. These properties are defined as : the number of *end_point*, the number of internal *branches*, the number of *intersections* and the total length of the skeleton (*pixel_count*).

Shape Angle The distribution of angles between points around the contours can be exploited to measure the similarities between shapes [Belongie et al., 2002]. This method was used to classify symbols (digits and letters). The procedure is defined as follow : for a point of the contour, we compute the relative angles of all the other points that are in a given radius. The radii and angles are stored in a “shape matrix” of dimensions : *radial_bins* \times *angle_bins*. Each element (*i, j*) of the matrix contains a counter which corresponds, for a given point, to the number of points which are in this bin of radius *i* and in the bin of angle *j*. Finally, the procedure is repeated for each point of the contour to feed the “shape matrix”. This one is used as a discriminating variable.

Shape Hu Moments The Hu Moments have been proposed long time ago by [Hu, 1962] and are part of structural analysis. The contours of the related regions are transformed into “moments”, which are then transformed into Hu Moments. The importance of this extraction comes from its translation, rotation, and scaling invariance. A similar element, e.g. two different bean leaves at various positions, rotation or scale will then be represented identically. This assertion remains true if intra-species and intra-individual variability are not taken into account. Hu Moment have been used for leaf classification challenge [Rhouma et al., 2017].

6.3.1.2 Spatial Criteria

The spatial properties are based on the extraction of global information related to the agricultural plot. These extraction methods are based on a distance which is calculated between the extracted global information and the centroid of the related region. The distribution of the values is strongly correlated to the vegetation density, allowing to discriminate the weeds, mostly between rows.

Spatial Row The distance between the plant and the crop line is one of the most used spatial descriptors when such prior knowledge is observable [Louargant et al., 2018, Koot, 2014, Bah et al., 2018b]. The number of techniques to extract such rows within an agronomic images is large, most of them are based on Hough Transform applied to line detection, as [Louargant et al., 2018]. Recent studies in deep-learning shows that CNN can be used to detect crop row in agricultural acquisition [Bah et al., 2019]. As the objective of this section is to evaluate the impact of the distance between plants and crop row (not the row detection performance), this study uses the crop rows from the ground truth.

Spatial Blob In most cases, plants can be classified into two types : monocotyledons and dicotyledons. At an early stage, when the plant emerges from the ground, its appearance is totally different according to this initial type. Monocotyledons usually looks like a filament, while dicotyledons look like an oval shape when seen from above. With this information, a specific shape detection algorithm can be implemented. Such as Blob detection, which aims to detect groups of connected pixels in an image that share a common property (shape, color, ...). The objective of Blob detection is to identify and mark these regions. These marks are used to calculate the Euclidean distance between the centroid of leaf and the nearest blob. Thus, the distribution of values is strongly correlated to the density of vegetation. This is an hypothesis we want to test and we have not seen any article using this type of approach.

Spatial Corner The detection of points of interest can be used to discriminate weeds. This is what is proposed in the article of [Xu et al., 2020], using an algorithm of detection of corners in the image. From these points of interest, the distance and the angle between the center of the plant and the closest points of interest are used as discriminating properties. In reality, this detection finds the leaves apexes. Here only the distance between the centroid and the nearest points of interest is computed since the angle is already present in Shape Ellipse properties.

6.3.1.3 Textural, Color and Spectral Criteria

Texture properties are vast and include many things, like image transformations (Fourier, Wavelets, color spaces, ...), color, spectral properties or histograms to characterise observed surfaces. The article [Mekhalfa and Yacef, 2021], have tested texture and color properties on RGB images of soybeans. The proposed color characteristics, are the means and standard deviations of the image in two color spaces (RGB and HSV). In terms of texture properties, i.e. an analysis of the spatial distributions of the colors, the paper relies on GLCM (Gray-Level Cooccurrence Matrix), Haralick and LBP (Local Binary Pattern) features. According to the study, the results are based on these properties and gave classification rates between 89.43 % and 96.17 %. Here, we have extended this work by adding their location to the extracted properties. Thus, all the extractions defined hereafter are automatically done on each individual bands, on NDVI and on the standard deviation (std) between spectral bands (which is also an image).

Spectral Signature The spectral signature is defined as the value (at the center of the shape and at the center of its bounding box) of all spectral bands, the NDVI value and the standard deviation between spectral bands.

Spectral Stats As shown by [Lottes et al., 2016], statistical properties of the underlying bounding box (that contains the shape) can be computed and used for weed discrimination. According to this article, a set of nine properties for each spectral bands is extracted separately as well as a vegetation index (NDVI) and the standard deviation (std). These properties are based on color moments (Table 6.1).

min	max	range
mean	std	median
skewness	kurtosis	entropy

Table 6.1 – Synthesis of colors moments

Spectral HoG The HoG descriptors for Histogram of Oriented Gradient was used by [Saha et al., 2017] for weed discrimination. It is often used in computer vision, both for object detection and texture discrimination because of its properties of geometric invariance. The HoG descriptors are based on gradient intensity and edge directions. From this information, a histogram is computed and used as descriptor. Applied to a leaf, these properties allow to characterise leaf veins [Tsolakidis et al., 2014]. These features are generally robust to illumination changes and color differences due to the leaf maturity.

Spectral Hu Moment Moments and Hu Moments are generally applied to segmentation masks or contours. It retrieves shape properties, such as perimeter, centroid, etc. Computing these features on spectral bands and NDVI could include more discriminating information. This is an assumption that we wish to verify and there is no anterior of such use.

Spectral Gabor Gabor filters have been widely used for texture analysis in mono-bands images. By extension, [Jain and Healey, 1998] have proposed to use properties based on the results of Gabor filters applied to multispectral texture classification. They use a set of symmetric and circular Gabor filters, with three octave scales and four orientations to compute properties, such as averages and energy. These filters have been used in the field of weed discrimination by [Ishak et al., 2009]. In the current study, a symmetric and circular Gabor filters are applied with a single (optimized) octave scale and four orientations ([0, 40, 90, 158]) as defined by [Jain and Healey, 1998]. However, extracted properties from such filters are based on color moments (Table 6.1) on each Gabor orientation filter.

Spectral LBP The LBP descriptor, for Local Binary Pattern is a feature that encodes texture information. The general principle of LBP is to compare the level of a pixel luminance with the levels of its neighbours (0 if inferior, and 1 else). Depending on the neighbour position, a weight is applied to the positive value ([1, 2, 4, 8, 16, 32, 64, 128]), the LBP value is then the sum of weighted positives values. The last feature, is the histogram of the leaf texture for each spectral bands and NDVI. This gives an account of information relating to regular patterns in the image, in other words texture.

Spectral CSLBP The CS-LBP operator for Center-Symmetric Local Binary Pattern is an extension of the previously defined LBP operator [Banerji et al., 2012]. For each pixel, the absolute and symmetric difference of the neighborhood is computed, if the value is higher than a fixed threshold $T = 0.1$ the element takes the value 1 (0 otherwise). The set of results is coded as LBP through a weighted sum depending on the neighbour location. These features are considered to be more robust against rotation than standard LBP. This is an assumption that we wish to verify.

Spectral OCLBP The OC-LBP (Opposite Color LBP) operator is an extension of the LBP operation. [Banerji et al., 2012] propose an LBP extension to the colorimetric domain by considering the inter channel relationship instead of local and spatial relationship of a pixel. For each color pair (u, v) , $u > v$ the inter-channel properties are considered by $C = I(x, y, u) - I(x, y, v)$. With I the multispectral image, x, y the pixel position and u, v the color channels to compare. The obtained image is defined in the same way as LBP : a weighted sum of positives values. These properties have been used by [Waghmare et al., 2016] for grape plant diseases detection.

Spectral GLCM This is a square matrix of dimension N_g , where N_g is the number of gray levels in the image. The $[i, j]$ element of the matrix is generated by counting the number of times a pixel of value i is adjacent to a pixel of value j , and then dividing the entire matrix by the total number of comparisons made. Each entry is thus considered as the probability that a pixel of value i is adjacent to a pixel of value j denoted $p(i, j)$, defining for us the co-occurrence matrix G . This has been used by [Bakhshipour et al., 2017] for weed discrimination.

Spectral Haralick The basis of these features is the GLCM (grayscale co-occurrence matrix). From that, Haralick have defined 14 texture descriptors that can be computed. This estimator was tested for canopy discrimination by [Wong et al., 2013].

Spectral Polyfit We introduce a new type of feature based on polynomial fitting. Multispectral images may lack information due to the selection of the spectral domain (The Airphen camera uses 6 spectral bands). To recover the original spectrum distribution across the surface, a simple solution that can be used is curve fitting [Brown, 2006], by using the value of observed pixel and their distances to the center of the shape. The result of the fitting is the polynomial parameter, which can be used as a discriminant variable.

Spectral Shi Tomasi The Shi-Tomasi feather is a kind of key point detector using the eigenvalue of the second moment of the color matrix [Harris et al., 1988]. On these values, a threshold is applied to determine if a pixel can be a corner or not. This algorithm has an interesting step, it transforms the input color image space into a gray scale color space, through this eigenvalue decomposition. Here, instead of finding the local maxima of this transformation, which are the key points. We propose to use a small histogram to describe the observed surface, so the values of corresponding to a corner accumulate in a specific bins of the histogram, while flat surfaces accumulate on other bins, etc. This transformation is applied on the first 3 bands of our data (450 nm, 570 nm, 650 nm) which are nearest to a RGB color space.

6.3.2 Optimization

Some features have parameters, like the number of bins in case of histogram analysis, the *inner* and *outer* radius of “shape angles” and so on. Some parameters are discrete numbers, but others are in a continuous space, like four parameter of the “texture gabor” which requires to set them before using it (sigma, lm, gm, ps). Setting these parameters has a critical impact on classification performances, they can be set empirically. But to optimize the performances of extracted features, these parameters must be optimized properly and automatically. This is known as an hyperparameter optimization problem.

There are many ways to sample the parameter space. One approach is grid search, where the parameter space is divided into a number of discrete regions (or grids) and a search is performed in each point of the grid [Lawler and Wood, 1966]. Another approach is random search, where a random point is chosen in the parameter space and the performance is evaluated [Bergstra and Bengio, 2012]. However, they can also lead to suboptimal solutions, because both methods are brute force approaches that scale poorly with the hyperparameters number. Therefore, it is important to choose an appropriate sampling method for the problem at hand. To achieve this optimization, more complex algorithms must be used to sample the parameter space efficiently.

We propose to use an asynchronous early stop method based on the continuous halving algorithm to solve this challenge, this is the ASHA algorithm (Asynchronous Successive HALving). The ASHA algorithm is a randomized, asynchronous and incremental algorithm for sampling the parameter space. It is based on the continuous halving algorithm, which is a greedy algorithm that divides the parameter space into two halves and selects the best half at each step. The ASHA algorithm uses this approach to iteratively divide the parameter space into smaller subsets. At each step, it selects the best subset based on a performance metric and stops when the subset is small enough that the performance is no longer improving. This approach is more efficient than grid search or random search, because it can find good solutions without exploring the entire parameter space. The asynchronous part of this algorithm allows it to run on multiple processors, which can speed up the search [Hertel et al., 2020].

In addition, to sample the parameter space correctly, each parameter has been manually bounded to limit the search space. The ASHA algorithm has been implemented in the Sherpa software package [Hertel et al., 2020] in Python. Finally, the optimization is done through 10 000 epochs. At each epoch a 2-fold cross-validation is used for training and validating a decision tree classifier. This should enforce the learning of more stable properties and avoid overfitting, because the first fold contains the majority of data obtained in June, while the second fold contain the data obtained in October. The decision tree classifier is chosen to use simple decision criteria, rather than complex model, to extract most relevant features. The performance of the extracted features is therefore the performance of the underlying learned classifier (average of the cross-validation procedure). The Figure 6.4 synthesizes the optimization procedure :

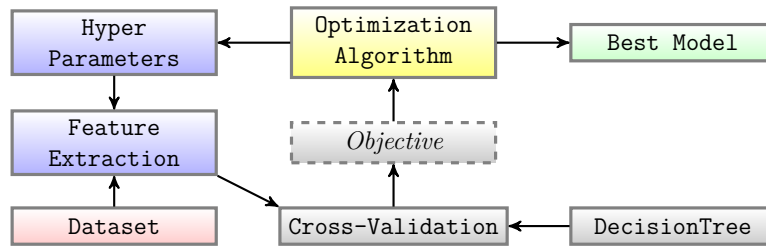


Figure 6.4 – Hyperparameter optimization procedure

As an example, Table 6.2, show 5 epochs for the optimization of spectral Gabor features (parameters : sigma, lm, gm and ps), the objective value is the value to be maximized, it corresponds to the performances of classification model evaluated by the *Area Under the Curve* metric [Fawcett, 2006]. After the optimization process, all features are extracted.

Epoch	sigma	lm	gm	ps	Objective
1616	1.865	5.720	0.409	0.403	0.621
1617	1.838	5.732	0.363	0.421	0.604
1618	1.894	4.960	0.431	0.335	0.556
1619	3.363	8.253	0.083	0.690	0.465
1620	7.705	4.655	0.986	0.090	0.517

Table 6.2 – Five epochs of the optimization procedure through ASHA algorithm. The objective column show the model performance

6.3.3 Feature Selection

Now that all the features are optimized and extracted, the main question of this article can be explored. What are the best features to classify leaves into crops and weeds. Since the number of features is huge (3545), as shown in Table 6.8, a manual estimation of each feature performances cannot be established properly. To answer this question, researchers in this field typically use a metric to estimate how well a feature fits the problem (correlation coefficient or root mean squared error), and then the best features are used. However, a better combination might exist, e.g., merging two high rank criteria might be less accurate than merging a high rank with a low rank. As an example, row distance and a texture characteristic, the efficiency may be greater than that of two texture criteria, because row distance is likely complementary to texture property. This is an assumption we wish to verify.

Instead of evaluating each feature through a metric, a data mining technique is proposed, such as “Sequential Forward Selection” (SFS) [Jain and Zongker, 1997]. It is an algorithm that analyzes each feature to check whether it is going to be useful for the classification task. To put it in another way the motivation behind this algorithm is to automatically select the subset of features most relevant to the problem. Thus, it reduces the initial d -dimensional feature space to a k -dimensional feature subspace, where $k < d$. The figure 6.5 shows how the algorithm works :

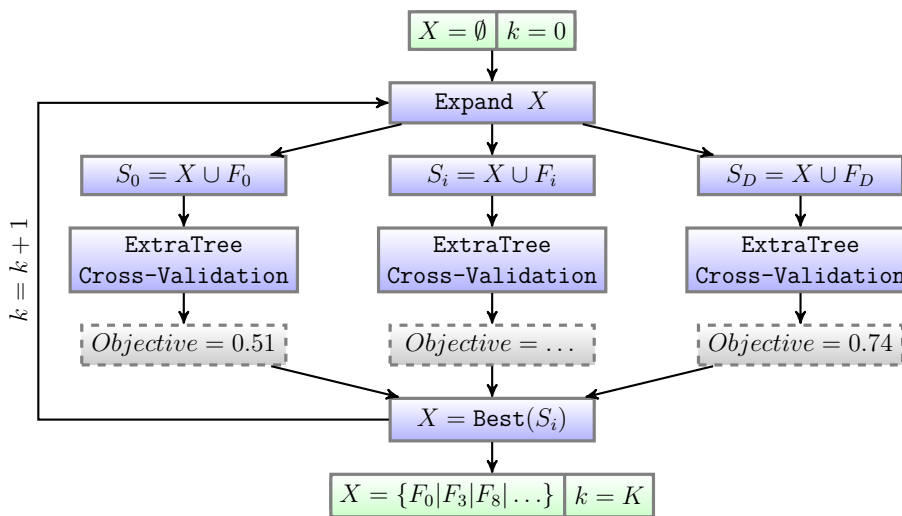


Figure 6.5 – Sequential Feature Selection Algorithm

6.4 Results and Discussion

6.4.1 Feature optimisation

Among the 25 types of properties, only 12 of them have hyperparameters that need to be optimized. For each of them the parameters and the underlying classifier performances have been extracted in the below Table 6.3 (for spatial and shape feature) and Table 6.4 (for spectral features). To show the importance of the feature optimisation process, we added the column **Worst**, which is the least discriminant parameters.

Type	Parameters	Values	Objective	Worst
Shape Angles	bins_radius	2	62.81	50.00
	bins_theta	22		
	radius_inner	0.9827		
	radius_outer	9.6870		
Spatial Corner	blockSize	2	59.52	51.49
	k	0.0995		
	ksize	1		
	threshold	0.09926		
Spatial Blob	area	0	52.00	51.23
	circularity	0		
	inertia	0		
	convexity	1		

Table 6.3 – Hyperparameter optimization through ASHA algorithm, the objective column shows the performance of the underlying classifier

Type	Parameters	Values	Objective	Worst
Spectral DoG	sigma_a	17.7569	72.39	46.75
	sigma_b	0.0475		
Spectral Gabor	gm	37.51	71.68	49.95
	lm	3.8094		
	ps	0.6573		
	sigma	8.7758		
Spectral GLCM	normed	<i>false</i>	68.96	61.55
	symmetric	<i>true</i>		
Spectral Zernike	m	1	60.14	55.12
	n	1		
Spectral Haralick	size	128	69.66	52.99
Spectral HOG	size	256	54.96	51.08
Spectral LBP	nbins	256	69.17	53.16
Spectral Shi-Tomasi	nbins	14	67.23	51.84
Spectral CSLBP	nbins	22	69.53	55.33

Table 6.4 – Feature optimization through ASHA algorithm. To show the importance of the feature optimisation process, we added the column **Worst**, which is the least discriminant parameters. The column **Objective** show the best performance of the model.

Table 6.3 and Table 6.4, show a noticeable improvement of the feature performances through this optimization procedure. Few features are strongly enhanced, from 50% to 62.81% for “Shape Angles”, or from 46.75% to 72.39% for “Spectral DoG”, showing the necessity of hyperparameters optimisation. Irrelevant features start to appear here, for features that have a hyperparameter, this is the case for “Spatial Corner”, “Spatial Blob” or “Spectral HOG” that does not exceed 60% of AUC with a decision tree classifier.

Once the hyperparameters optimization is performed, all features (with or without hyperparameters) can be extracted. Thus, the next step is to use the feature selection algorithm to select the best criteria from all extracted features. In the next section, we will show the results of the feature selection algorithm.

6.4.2 Feature selection by type

For this study, the SFS algorithm is applied independently on each feature type (locally), and on all extracted features (globally). This allows us to check which features are the most relevant for their categories as well as globally. For the SFS algorithm, the number of features included in the subspace is considered a stopping criterion. Here, the stopping criterion have been set to $k = 10$ for each feature type and $k = 20$ for all features. The result of the procedure is available on the next Table 6.5 for spatial feature, in Table 6.6 for shape feature and in Table 6.7 for spectral features. Finally figure 6.6 shows best global criteria. On all of these Tables, the best subspace size is highlighted in bold for each feature type, and the green cells indicate the best feature type to use for a specific feature space size.

Feature Count	1
Spatial Blob	51.80
Spatial Corner	56.59
Spatial Row	66.90

Table 6.5 – Best features for spatial features. Best results for a specific feature count are highlighted

Since these 3 feature types extract only one feature each, Table 6.5 shows only one performance for each feature type. But it also shows that among the 3 spatial feature “Spatial Row”, which is the distance of the leaf to the crop row, has the most relevant performance. This is expected, as it allows to discriminate inter-row weeds. However, the weed coverage in October were high and overlapped the crop row, reducing this feature performance. The worst is “Spatial Blob”, which may be due to a lack in the optimization procedure for this feature, or more likely to the lack of discriminating information in this analysis, as for “Spatial Corner” that offers moderate performances. The next Table 6.6 is for shape criteria.

Feature Count	1	2	3	4	5	6	7	8	9	10	Full
Shape Angles	65.73	65.72	65.73	65.72	65.72	65.73	65.72	65.73	65.73	65.72	68.38
Shape Ellipse	60.84	67.47	69.46	69.76	69.56	69.48	69.28	69.10	68.80	68.54	67.58
Shape Fragstats	67.85	70.10	74.83	73.62	73.10	72.55	–	–	–	–	72.55
Shape Hu-moment	56.87	59.79	63.30	64.87	65.52	65.78	65.74	–	–	–	65.74
Shape Metrics	57.01	62.74	66.88	67.81	68.52	68.76	68.98	68.89	–	–	68.89
Shape Particle	67.88	70.15	74.50	76.08	76.04	75.99	76.03	75.95	75.83	75.73	73.13
Shape Skeletonize	68.81	68.81	68.65	68.52	–	–	–	–	–	–	68.52
Shape Solidity	59.04	–	–	–	–	–	–	–	–	–	59.04

Table 6.6 – Best classification features for each feature type through ASHA algorithm. Each line presents the results for a specific feature type. The feature count is the number of feature(s) combined for a specific feature type. The presented results are the optimal combinations for each feature counts. Column “full” is the result obtained with all the features for a specific feature type. Best results for a type of features are in bold, best results for a specific feature count are highlighted.

As shown in Table 6.6, “Shape Solidity” that have been found relevant at plant scale [Ahmed et al., 2012] is less relevant at leaf scale. The first feature selected for “Shape Skeletonize” is the pixel-count. It refers to the length of the skeleton, which is the most discriminating property among the others of this feature type. Among all shape feature, the “Shape Particle” is the most

important one, the four best features are respectively “inv-area”, “m1”, “m2”, and “form-factor” with a score of 76.08 %. While the second best shape property is “Shape Fragstats” that is used to classify land plot, its 3 best features are respectively the ‘perimeter-area-ratio’, the ‘related-circumscribing-circle’, and the ‘perimeter-area-fractal-dimension’ with a score of 74.83 %. The four first features selected for the “Shape Ellipse” seem quite good, such as “center-distance”, “minor-axis”, “angle”, and “eccentricity 2” with an accuracy score of 69.76 %. The other ones seem to be less relevant for crop and weed discrimination at leaf scale. Finally, Table 6.9 in the supplementary material, shows the name of the ten top best shapes features.

Feature Count	1	2	3	4	5	6	7	8	9	10	Full
Spectral LBP	69.81	70.71	70.72	70.71	70.71	70.71	70.71	70.71	70.71	70.71	72.61
Spectral OCLBP	61.69	65.09	68.26	69.34	69.94	70.34	70.47	70.66	70.56	70.60	73.03
Spectral CSLBP	70.36	70.36	70.36	70.36	70.36	70.36	70.36	70.36	70.36	70.36	68.12
Spectral Zernike	57.26	68.67	72.58	73.42	74.44	75.42	75.80	76.19	76.37	76.60	76.91
Spectral Shi-Tomasi	71.11	67.74	70.02	70.29	71.25	72.01	72.59	72.84	72.94	73.03	72.00
Spectral Signature	66.76	67.37	74.40	75.79	77.25	78.18	79.23	79.35	79.58	79.53	63.65
Spectral Hu Moment	68.29	68.71	70.48	74.21	81.70	83.09	83.15	83.05	83.22	83.19	71.29
Spectral Stats	68.97	72.58	79.09	81.61	83.36	84.37	84.98	85.65	86.29	86.54	74.58
Spectral Polyfit	59.57	62.80	66.01	66.94	67.70	68.40	68.97	69.54	70.48	70.63	70.20
Spectral HOG	55.99	64.73	69.83	72.01	73.93	74.83	75.44	75.95	76.37	76.67	74.40
Spectral DoG	70.54	74.04	78.36	80.76	82.54	83.53	84.64	85.42	86.03	86.33	75.76
Spectral Gabor	69.50	74.48	80.07	82.23	83.53	84.52	85.14	85.60	86.15	86.40	74.92
Spectral GLCM	59.57	66.22	73.84	76.87	78.93	80.12	80.97	81.61	82.13	82.56	79.52
Spectral Haralick	63.80	75.85	78.89	81.16	82.56	83.69	84.62	85.01	85.33	85.51	81.30

Table 6.7 – Best features individually for each feature type for classification through ASHA algorithm. Each line presents the results for a specific feature type. The feature count is the number of feature(s) combined for a specific feature type. The presented results are the optimal combinations for each feature counts. Column “full” is the result obtained with all the features for a specific feature type. Best results for a type of features are in bold, best results for a specific feature count are highlighted.

As shown in Table 6.7, the best features are the ones related to the texture and color of leaves. If a single feature should be used then, the best feature is the “Spectral Shi-Tomasi”, which is related to the roughness of the texture, since it is the first step of corner detection, however the remaining features are less discriminant than the others feature types. Among these spectral properties, the LBP, OCLBP and CSLBP were expected to produce better performances (they max out at around 70 %), this means that the histogram construction is not producing relevant features. The same remarks can be established for the Zernike features even if their performances are slightly better (around 76.60 %).

“Spectral Signature” is commonly used for vegetation classification, it uses the spectral value of the center of the shape (or window) as a criteria. The performances of this feature are better than the features previously discussed but max out at around 80 %. The next feature, which is the “Spectral Hu Moment”, is a transformation of the spectral value of the center of the shape through Hu Moment, this transformation shows a significant increase in the performances of such features, from 79.53 % for “Spectral Signature” to 83.19 % for “Spectral Hu-Moment”, in addition this transformation seems more interesting on spectral properties than shape properties as seen in Table 6.6.

Features “Spectral Polyfit” and “Spectral Stats” are relatively close by definition but show a significant difference on their performances. “Spectral Stats” extracts statistical properties (Table 6.1) of the underlying sub-texture which contains the leaf and the soil around it, while “Spectral Polyfit” uses the spectral information of the sub-texture to reconstruct the spectral distribution of the leaf surface, then the polynomial values are used as discriminating information. The initial assumption was that “Spectral Polyfit” (which is an attempt to reconstruct the initial spectrum distribution from the discrete values) should have been more relevant than “Spectral Stats”.

“Spectral HOG” for Histogram of Oriented Gradient, is generally used to retrieve gradient information. In case of leaf analysis, this transformation detects leaf veins [Larese et al., 2014], which is a criterion to discriminate leaf species. Since this analysis doesn’t produce good enough performances, it is possible that the spatial resolution is too low for this kind of methodology.

The remaining “DoG”, “Gabor”, “GLCM” and “Haralick” features offer good performances as expected. We can notice that “Spectral Gabor” is the best of the four when ten features are used, while “DoG” and “GLCM” show slightly less accurate results. Finally, Table 6.10 in the supplementary material, shows the name of those ten top best spectral features.

6.4.3 Best of all features

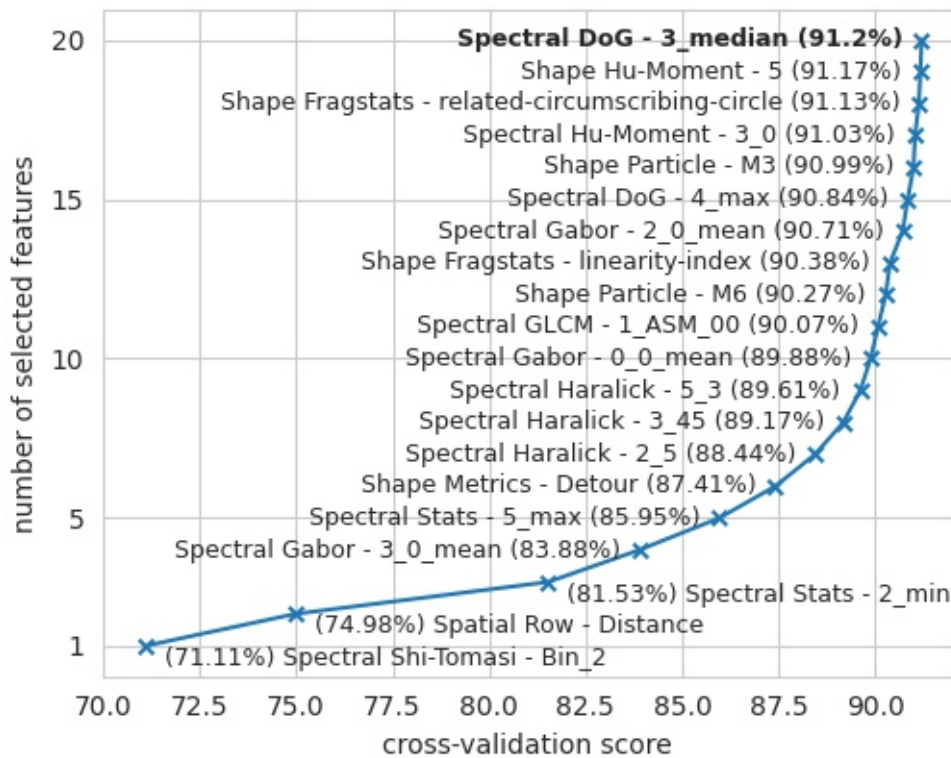


Figure 6.6 – Best features among all extracted feature for classification through ASHA algorithm. Best overall result is highlighted.

As shown previously (Table 6.7), the best starting feature is the second bin of “Spectral Shi-Tomasi” histogram, thus it is selected firstly and its performance level is found in the first point of Figure 6.6 (lower-left). The second selected feature is “Spatial Row”, it is expected that this feature should be on the top as it is a complementary with the color/texture features. At this

stage, we can notice that this combination is less accurate 74.98 % than two texture Haralick feature 75.85 %. The next three features are “Spectral Stats 2_min”, “Spectral Stats 5_max” and “Spectral Gabor 3_0 mean”. Which is the minimum spectral value of the underlying subtexture of the 650 nm spectral band for the first feature. In the same way, it is the maximum for the 850 nm spectral band. The Gabor filter is applied on all spectral bands, the discriminating criterion is one from the 450 nm. This leads to a performance of 85.95 %. The next feature is “Shape Metrics Detour”, with a performance of 87.41 % which beats any previous classification model.

To conclude on this results, the best classification model includes 1 spatial feature, 6 shape features and 13 spectral features. In the previous Figure 6.6, only 20 features are shown, if more features are extracted, the performances begin to decrease (21 features = 91.16 %), this means that the performance reaches a vertical asymptote. To show this a model have been learnt with all features (3545), the cross-validation score is 79.97 %.

6.4.4 Visual results

Figure 6.7 shows two rows with field beans plants. Weeds are located between these two rows and on each of them. Most of plant and weed leaves are correctly classified (plant leaves are colored green, and weed leaves are colored blue). However, this image suggests a problem remains for the smallest leaves that are in the center of the plant. These plants leaves, colored in red, are classified as weed leaves. This classification is induced by shape features, which are less relevant for this category of leaves than for the others.

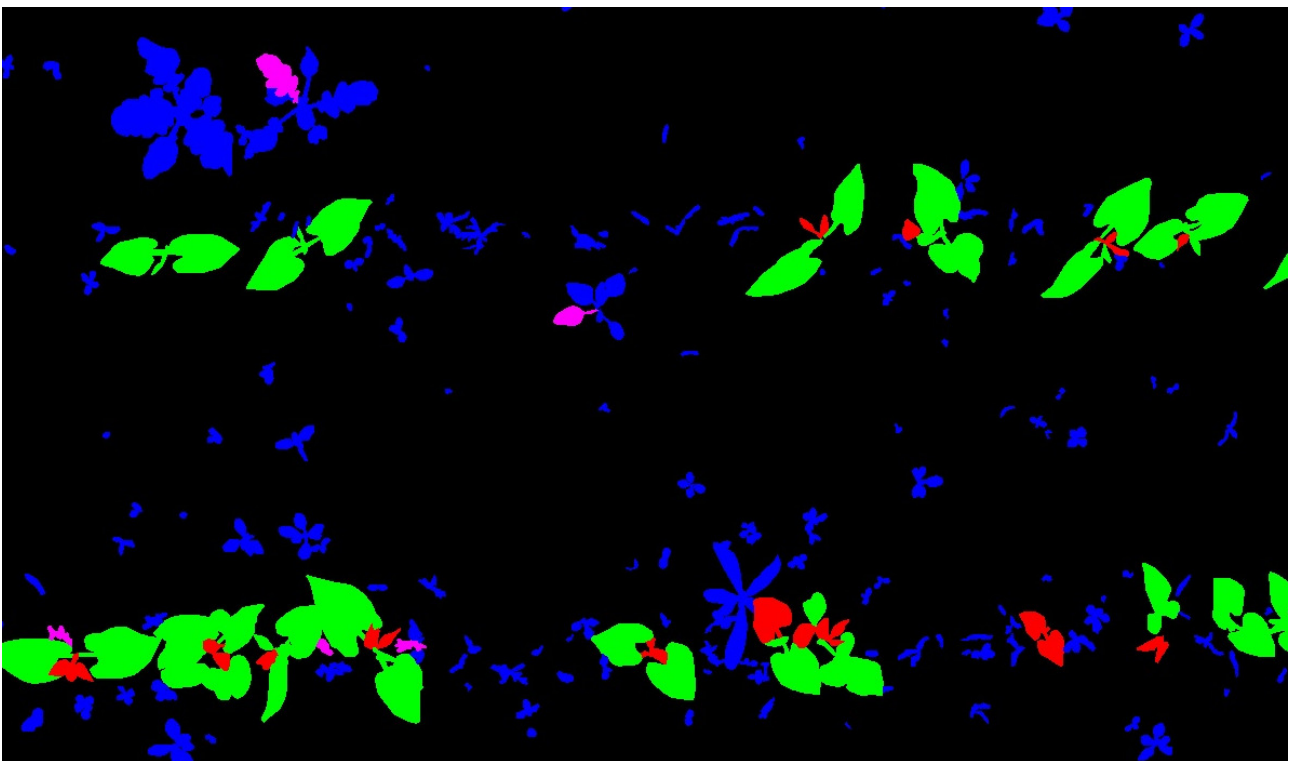


Figure 6.7 – This figure shows the classification of leaves, from an image taken in June 2019. The blue color corresponds to well classified weeds. The green color corresponds to well classified crops. While the purple and red color correspond respectively to weeds and crops poorly classified.

Figure 6.8 focuses on specific parts of plant rows. Weeds on figure 6.8.a are mostly monocots, whereas those visible on figure 6.8.b are mostly dicots. These two figures illustrate the quality of classification, which is equivalent for the two weed groups. Figure 6.8.c presents a dense foliage, with highly developed wild mustard. On this figure, the major part of mustard leaves is correctly classified. The classification of field bean leaves depends on their visibility. Several partially occluded leaves are classified as weed leaves (in red). Conversely, figure 6.8.d presents sparse foliage, where many bean plants have not emerged. In that case, the small bean leaves are also classified as weed leaves.

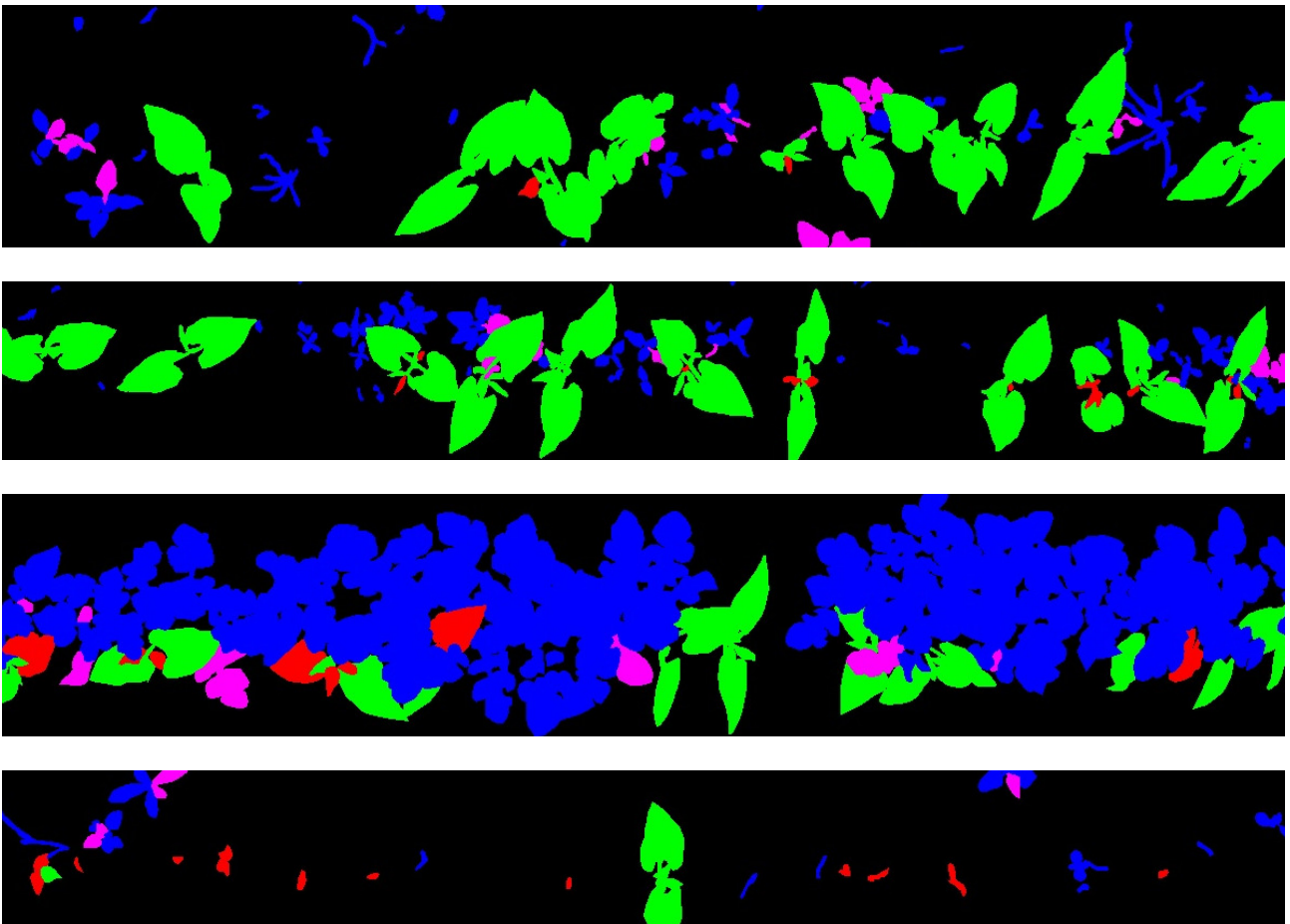


Figure 6.8 – This figure shows the classification of leaves, from parts of four images taken in June 2019. The blue color corresponds to well classified weeds. The green color corresponds to well classified crops. While the purple and red color correspond respectively to weeds and crops poorly classified. Here we see that leaves between crop rows are correctly classified, as are monocots. On the other hand, the emerging leaves of crops and some weed type leaves of dicotyledons are incorrect.

All of these images illustrate the difficulty in classifying leaves which, being in the row of crops, have particular shapes, because they are partially obscured or because they are not sufficiently developed.

6.5 Conclusion

The properties we evaluated were identified from a thorough literature review. We distinguished spatial properties, based on notions of distances at the image scale, shape properties allowing to characterize the morphological features of the individuals. And, finally, properties of textures, colors and spectra were proposed to extract information related to the composition and internal structures of leaves.

From this survey, it came out that some methods have parameters that influence the discrimination potential of these properties. Therefore, an algorithm was proposed to optimize these hyperparameters before the extraction and evaluation of these features. It has been shown that the optimization of these hyperparameters plays a crucial role in the performance of these criteria. For example, “Shape Angles” properties went from 50 % to 62.81 % of discrimination potential after optimization, as well as the “Spectral DoG” properties from 46.75 % to 72.39 %, etc.

After the optimization of these properties, the set of properties was extracted from the dataset for evaluation. In this evaluation, it was shown the importance of selecting a smaller set of properties, the overall score is better with 20 features than all the features (3545). It is also important to notice the the contribution of each feature decreases as the number of features increases, resulting in a less interesting “performance over computing time” ratio. For each property type, the top 10 properties were defined. Then the top 20 among all properties combined were defined. These results show a crop/weed discrimination performance up to 91 percent with 20 properties, while a performance of 79.97 % is observed using the 3545 properties. Indeed, the usual classification techniques do not seem to be able to handle large dimensions of properties. But here, the performances of classifications proposed by smaller sets are superior to the use of all the criteria. Thus, the definition of this minimal set also allows to decrease the computation time.

6.6 Further research

Today, deep learning may be the only competitive alternative that could improve the results of each feature type. Following the approach that has been undertaken in a previous article [Vayssade et al., 2021], an alternative based on function approximators is possible. We could then define the best shape properties from a fixed representation of the contour. The same is true for texture and spatial properties. In reality, a UNet model already synthesizes the set of texture, color, spectrum and a kind of shape features if used in this crop/weed discrimination framework. Moreover, the performance of the extraction of these features does not allow a real time use, which reinforces the need to study deep learning.

In this study, feature selection is only based on classification performances of features. Computation speed is neither estimated nor taken into account. However, depending on the application, a compromise between computation speed and classification performance could be used. The selection of the best features for real-time weed detection and elimination is of particular interest. Figure 6.6 shows that the major classification improvement comes from the 10 first features (from 71 to 90 %), the next 10 features improve the classification result of a bit more than 1%. This observation shows that a study focused on performance over computing time should be relevant. This is also a key step in developing on-board weed control systems in autonomous vehicles where computing power can be restricted and time matters the most.

Considering precision farming applications, this approach could be replicated to select the best features for a specific set of crop(s) and weeds. The objective could change from an application to another and should be linked to the classification accuracy requirements and the time available. A real-time application may focus on computing time with less consideration for classification performance. A UAV approach may allow more computing time as the result is already delayed from the acquisition. In both case, the image resolution may be an important setting : too high and the computing time may be too important, too low and the classification performance may drop too much. This resolution may also impact the feature contribution to the overall classification with spectral or shape feature requiring highest resolution than spatial features.

Supplementary materials

Type	Name	FV	HP	Type	Name	FV	HP
Spatial	Borner	1	4	Spectral	Sig	16	–
Spatial	Row	1	6	Spectral	Stats	72	–
Spatial	Blob	1	4	Spectral	polyfit	58	1
type	name	FV	HP	Spectral	DoG	72	2
Shape	Angles	9	5	Spectral	HOG	72	1
Shape	Ellipse	11	–	Spectral	Gabor	288	4
Shape	Fragstat	6	–	Spectral	Haralick	416	1
Shape	metrics	8	–	Spectral	GLCM	80	2
Shape	Solidity	1	–	Spectral	Hu Moment	56	–
Shape	Particle	18	–	Spectral	Zernike	32	2
Shape	Hu Moment	7	–	Spectral	LBP	2056	1
Shape	Skeletonize	4	–	Spectral	CSLBP	144	1
				Spectral	OCLBP	34	–
				Spectral	Shi-Tomasi	8	1

Table 6.8 – Number of features after hyperparameter optimization (FV) and number of hyperparameters (HP) per feature extraction methods

Shape Angles	bins_8, bins_12, bins_13, bins_14, bins_15, bins_16, bins_17, bins_18, bins_19, bins_20
Shape Ellipse	center-distance, minor-axis, angle, eccentricity_1, eccentricity_2, eccentricity_3, eccentricity_angle, focal-distance, directrice, det
Shape Fragstats	perimeter-area-ratio, shape-index, fractal-dimension_index, linearity-index, related-circumscribing-circle, perimeter-area-fractal-dimension
Shape Hu-Moment	0, 1, 2, 3, 4, 5, 6
Shape Metrics	proximity, spin, area-exchange, perim-index, depth, dispersion, girth, detour
Shape Particle	inv-area, inv-perimeter, m2, m3, m4, m6, m8, aspect-ratio, roundness, convexity
Shape Skeletonize	pixel-count, end-point, itersetion, branch-count

Table 6.9 – Top features for each shape types

Spectral CSLBP	bins_0_21, bins_1_10, bins_1_18, bins_1_19, bins_1_20, bins_1_21, bins_2_10, bins_2_18, bins_2_20, bins_4_21
Spectral DoG	0_median, 2_min, 2_median, 3_max, 3_median, 3_entropy, 4_max, 5_max, 5_mean, std_max
Spectral Gabor	0_0_median, 1_0_mean, 2_0_max, 2_0_mean, 3_0_min, 3_158_entropy, 4_0_min, 5_0_min, std_158_mean, ndvi_0_min
Spectral HOG	0_bins_70, 1_bins_70, 1_bins_130, 2_bins_70, 3_bins_30, 5_bins_110, std_bins_30, std_bins_70, std_bins_150, ndvi_bins_150
Spectral GLCM	0_ASM_00, 1_contrast_90, 3_correlation_90, 3_ASM_90, 4_contrast_90, 4_homogeneity_00, 5_contrast_90, std_contrast_90, std_correlation_90, ndvi_homogeneity_90
Spectral Haralick	1_17, 1_31, 2_24, 2_31, 3_32, 4_32, 5_3, 5_51, std_29, std_44
Spectral Hu-Moment	0_0, 1_1, 2_1, 3_0, 3_2, 4_0, 5_0, 5_4, std_3, ndvi_3
Spectral LBP	bins_0_64, bins_0_209, bins_1_129, bins_2_129, bins_2_209, bins_3_129, bins_4_209, bins_4_250, bins_5_129, bins_std_129
Spectral OCLBP	bins_0, bins_1, bins_2, bins_3, bins_4, bins_9, bins_20, bins_26, bins_32, bins_33
Spectral Polyfit	1_2, 1_6, 2_6, 3_1, 3_6, 4_4, 5_3, 5_4, std_5, ndvi_0
Spectral Shi-tomasi	bins_0, bins_1, bins_2, bins_3, bins_4, bins_6, bins_7, bins_9, bins_11, bins_13
Spectral Signature	shape_0, shape_1, shape_2, shape_3, shape_4, shape_5, shape_ndvi, window_0, window_std, window_ndvi
Spectral Stats	0_min, 1_std, 2_min, 2_median, 3_max, 3_median, 4_std, 4_entropy, 5_max, ndvi_max
Spectral Zernike	0_A, 1_Zr, 1_A, 2_Zr, 3_A, 4_A, 5_A, std_A, std_Phi, ndvi_A

Table 6.10 – Top 10 for each spectral types

Funds & Conflicts of Interest

Funding was received from the European Union’s Horizon 2020 research and innovation program under grant agreement No 727321 (project acronym, IWMPRAISE). It was also funded by the French National Research Agency Challenge RoSE [grant agreement ID : ANR-17-ROSE-0002] (project acronym : ROSEAU). In addition, a collaboration with SITIA company is existing. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

6.7 Conclusion de chapitre

L'un des buts de cette thèse était de définir les meilleures propriétés pour la discrimination culture/adventices. Comme il n'existe pas d'études les définissant clairement (souvent, un petit ensemble de propriétés, dans des conditions spécifiques, est utilisé.), nous avons mené nos propres recherches pour identifier les méthodes les plus prometteuses. De plus, une échelle de segmentation plus fine que celle de la plante a été définie, celle de la feuille. Ce faisant, nous espérons fournir une compréhension plus détaillée de la manière dont ces propriétés peuvent être utilisées à des fins de classification cultures/adventices.

Pour combler cette lacune, nous avons recensé les méthodes existantes et identifiées les propriétés les plus prometteuses pour la discrimination cultures/adventices à l'échelle de la feuille. On y distingue les propriétés spatiales, reposant sur des notions de distances à l'échelle de l'image, et les propriétés de forme permettant de caractériser les traits morphologiques des individus. Des propriétés de texture, de couleur et de spectre ont également été proposées pour extraire des informations liées à la composition et aux structures internes des feuilles (veines, épiderme, ...).

À partir de ce recensement, nous avons vu que certaines méthodes permettant d'extraire des propriétés comportaient des paramètres. Ces paramètres influencent le potentiel de discrimination des propriétés extraites. C'est pourquoi un algorithme a été proposé pour optimiser ces paramètres avant l'extraction et l'exploitation des propriétés concernées. Il a été montré que l'optimisation de ces hyperparamètres jouait un rôle crucial. Par exemple, les propriétés du type ShapeAngles passent de 50 % à 62.81 % de potentiel de discrimination après optimisation, de même que les propriétés SpectralDoG passent de 46.75 % à 72.39 %, etc.

Après cette phase d'optimisation, toutes les propriétés ont été extraites à partir du jeu de données pour évaluation. Au cours de cette évaluation, il a été montré l'importance de la sélection d'un ensemble plus petit de propriétés. Les 10 meilleures ont été sélectionnées pour chaque type de propriété, puis les 20 meilleures parmi toutes les propriétés combinées. Les résultats montrent une performance de discrimination culture/adventices à hauteur de 91 % avec 20 propriétés, tandis qu'une performance de 79.97 % est observée en utilisant l'ensemble des 3545 propriétés. En effet, les techniques de classification usuelles ne semblent pas être capables de gérer de grands ensembles de propriétés. Par conséquent, il est important de définir un ensemble minimal de critères qui permette de diminuer le temps de calcul tout en maximisant les performances de classification.

En conclusion, une étude a été réalisée pour déterminer quelles propriétés sont les plus importantes pour la tâche de discrimination cultures/adventices, puis elles ont été optimisées pour obtenir les meilleurs résultats. Un algorithme a été utilisé pour trouver la meilleure façon d'utiliser ces propriétés pour la classification. Au final, il a été constaté que l'utilisation d'un ensemble plus restreint de propriétés, de différents types, donne les meilleurs résultats.

7 Conclusion Générale

Les travaux de recherche de cette thèse ont été entrepris dans le cadre du projet européen *H2020 IWMPRAISE*, dans le but de proposer des outils pour une meilleure gestion des adventices avec pour principal objectif de diminuer l'utilisation des produits phytosanitaires et leurs impacts écologiques (présentés dans l'introduction). Cette thèse, s'est aussi insérée dans l'*ANR Challenge RoSE*, dont l'objectif était de mobiliser et d'évaluer les performances de différentes équipes de recherche, pour le désherbage de l'intra-rang en grandes cultures et cultures légumières de plein champ.

Les recherches qui ont été effectuées et présentées dans ce manuscrit, concernent la tâche de détection des adventices, qui constitue encore aujourd'hui un verrou scientifique. Pour répondre aux autres tâches évaluées dans le cadre de l'*ANR Challenge RoSE* (telles que l'interprétation, la décision et l'action), un partenariat avec la société *SITIA*, les chambres d'Agriculture de Bretagne et Pays de Loire, et l'*IRSEEM-ESIGELEC* a été établi en amont de cette thèse. L'ensemble constitue le consortium *ROSEAU*. Ainsi, les résultats de ces recherches ont vocation à être implémentés sur le robot « TREKTOR », de la société *SITIA*, pour une gestion automatisée des adventices.

L'objectif principal de cette thèse était de définir les “meilleurs” critères extractibles à partir d'images multispectrales, afin de discriminer les cultures des adventices. Toutefois, il est rapidement apparu que l'extraction et l'évaluation de critères ne présentaient pas de verrou scientifique majeur, mais que la performance et la stabilité des critères de discrimination dépendait de la segmentation des images. Ainsi, le travail s'est orienté vers la définition d'indices de végétation optimaux et d'une méthode de segmentation sémantique des instances à l'échelle de la feuille. La performance des méthodes proposées a été évaluée sur des images de feuillage dense, acquises en lumière naturelle. Au final, l'approche proposée mixe classification usuelle et apprentissage profond.

7.1 Contributions

Comme évoqué dans le chapitre 2, il existe deux grandes voies pour détecter les adventices : une approche standard d’analyse d’image ou l’utilisation du “tout” deep-learning. Notre méthode repose sur une architecture “standard” dont chaque partie a été améliorée par optimisation des hyperparamètres ou par apprentissage profond. La figure 7.1 ci-dessous, montre l’architecture finale qui a été développée ici.

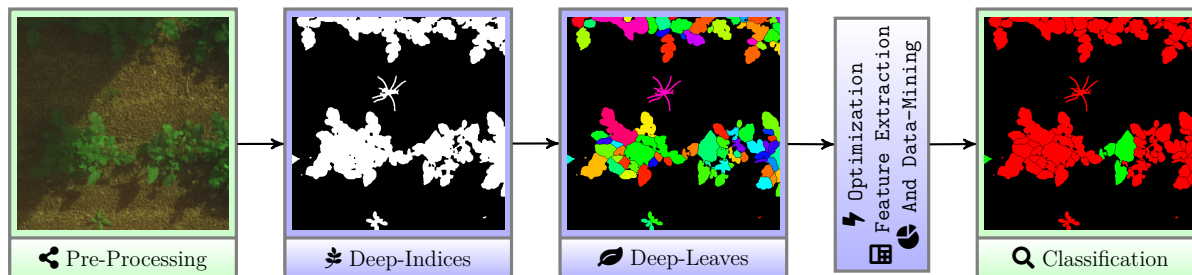


Figure 7.1 – Pipeline final pour la détection et la discrimination des adventices

La première étape du pipeline est le “Pre-Processing” des données, pour l’exploitation des images multispectrales acquises durant nos expérimentations. Elle consiste en la calibration de la caméra et l’utilisation d’un algorithme développé, permettant la superposition des bandes spectrales avec une erreur inférieure au pixel.

La seconde étape, “Deep-Indices”, vise à discriminer la végétation du sol. Cela consiste en l’utilisation de nouveaux indices déduits par apprentissage supervisé sur des approximateurs de fonction. Ces derniers, contrairement aux indices standards, ont l’avantage de ne pas dépendre d’une correction radiométrique. Ces Deep-Indices ont montré une nette amélioration de la qualité de la segmentation (82.19 mIoU) par rapport aux indices standards (63.98–73.71 mIoU).

La troisième étape, “Deep-Leaves”, permet d’identifier les feuilles. Un algorithme a été développé et utilise les Deep-Indices ainsi qu’un CNN qui détecte le contour des feuilles. Cette approche a montré de bonnes performances sur des jeux de données externes et sur nos données acquises en condition de terrain. Cette méthode permet de travailler avec n’importe quelle densité de feuillage et a la capacité de détecter les plus petites adventices même lors de leur émergence.

Enfin, on réalise l’extraction des meilleures caractéristiques pour la discrimination des cultures et des adventices. Pour ce faire, on a recensé plusieurs méthodes permettant l’extraction de différents types de caractéristiques. On utilise ensuite un algorithme d’optimisation des hyperparamètres pour chacune d’elles afin de maximiser leurs performances dans notre cas d’utilisation. On termine en utilisant un algorithme de sélection de propriétés qui permet de déterminer les meilleures combinaisons des propriétés extraites.

Une performance de classification de 91 % a été démontrée. Nous rappelons que notre étude est à l’échelle de la feuille, ce qui est une première dans ce domaine, ainsi qu’en condition terrain. Il n’existe donc pas de points de comparaison dans la littérature. On peut cependant noter qu’à l’échelle de la plante dans différentes conditions d’acquisition, l’état de l’art [Ahmed et al., 2016] donne des performances entre 77 % et 97 % de discrimination culture/adventices par de tels critères. Ce qui place favorablement nos résultats et offre ainsi des perspectives.

7.2 Perspectives

Acquisition Concernant la caméra, de par sa conception multi-capteurs, une méthode de recalage des bandes spectrales est nécessaire. Même si un algorithme a été développé durant cette thèse et même s'il montre d'excellentes performances pour corriger ce problème, la méthode est coûteuse en temps de calcul et il n'existe actuellement pas de solution plus performante pour cet appareil. Ainsi, bien que cette caméra offre une résolution spatiale de très bonne qualité, son utilisation ne permet pas un traitement des données en temps réel. D'autres caméras multispectrales et hyperspectrales ont vu le jour durant cette thèse, et pourraient améliorer les performances (temps de calcul ou précision). De plus, aucune correction radiométrique n'a été utilisée durant cette thèse, principalement du fait des contraintes expérimentales. Les résultats de cette thèse montrent d'excellentes performances sans cette correction radiométrique, mais l'on peut supposer qu'un tel pré-traitement les amélioreraient encore. Des approches par apprentissage profond pourraient être développées, pour proposer une correction ne nécessitant pas de dispositif d'étalonnage [Afifi, 2018, Tian et al., 2020].

Des données synthétiques L'annotation de larges bases de données est longue, coûteuse et souvent sujette à des erreurs d'annotations. Dans le cas de la détection de végétation, comment définir si un pixel contient de la végétation en bordure de feuille? Par exemple, lorsqu'il présente un mélange spectral? Annoter précisément des images dans ce cas de figure est difficile. Certaines techniques d'augmentation de données, permettent d'ajouter de la variabilité pour ce cas de figure. Dans cette thèse, nous avons proposé d'ajouter un bruit de perlin permettant de simuler de l'ombrage. Toutefois, superposer des images ne contenant que de la végétation à des images ne contenant que du sol, en proportion variable, permettrait peut-être d'apprendre des indices encore plus résistants, mais permettrait également d'obtenir la quantité de ce mélange. Concernant les données pour la détection des feuilles, [Ward et al., 2018] a montré qu'il était possible de construire des annotations RGB synthétiques pour la détection de ces individus. Il serait fort intéressant d'étendre cette méthode à des données multispectrales en lien avec notre caméra. Pour réaliser ces données synthétiques, je propose d'utiliser soit un moteur de rendu temps-réel tel que l'UnrealEngine 5 [Qiu and Yuille, 2016], soit un moteur de rendu style path-tracing via Mitsuba 2 [Nimier-David et al., 2019], tous deux basés sur des rendus physiquement réalistes. En utilisant les modèles de BRDF à microfacet et les distributions GGX, différentes conditions d'acquisition et de rugosité de surface peuvent être aisément simulées. Notons de plus, que des méthodes d'informatiques graphiques permettent d'apprendre ces matériaux à partir d'un ensemble d'acquisitions [Dupuy et al., 2015]. L'utilisation de ces technologies permettraient d'obtenir des données synthétiques d'une qualité pour l'instant inexistante dans la littérature scientifique, en agriculture de précision.

Détection des individus La détection des objets est un sujet difficile en particulier en scène dense, tel qu’il a été vu dans cette thèse. L’approche qui a été proposée ici semble prometteuse avec la conception d’une nouvelle architecture CNN, qui s’appuie sur plusieurs recherches antérieures. Bien que les résultats obtenus soient bons, des recherches complémentaires semblent nécessaires. L’échelle de la feuille semble adéquate, mais celle de la foliole serait peut-être plus facile d’un point de vue détection des objets. De plus, d’autres méthodes pourraient être ajoutées à l’architecture CNN proposée. Par exemple, l’algorithme des bassins versants [Bai and Urtasun, 2017] qui est actuellement utilisé en post-traitement (après la sortie du CNN). L’ajout d’un espace latent semble aussi prometteur d’après [Kulikov et al., 2018] pour la segmentation des feuilles. Enfin, comme nous l’avons vu, des améliorations sur la fonction de perte, et des techniques plus avancées d’augmentation des données doivent être explorées. En réalité, la construction de l’architecture d’un CNN est souvent empirique. Or, cette construction pourrait elle aussi être automatisée et améliorée à l’aide d’une optimisation des hyperparamètres. Mais cela nécessite des ressources de calcul qui n’étaient pas disponibles durant cette thèse.

Extraction de propriétés L’un des objectifs initiaux de la thèse était l’extraction et l’évaluation d’un ensemble de propriétés “standards” afin de définir leurs potentiels de discrimination pour la discrimination cultures/adventices. Il s’agissait de déterminer quels étaient les meilleurs d’entre eux. En réalité, ces caractéristiques sont fortement limitées par de nombreux facteurs. Ils ne permettent donc pas une classification parfaite. Aujourd’hui, l’apprentissage profond est la seule alternative compétitive qui pourrait permettre d’améliorer les résultats. En suivant la démarche qui a été entreprise dans cette thèse, une alternative basée sur les approximateurs de fonction est possible. On pourrait ainsi, définir les meilleures propriétés de forme à partir d’une représentation fixe du contour. De même pour les propriétés de texture et les propriétés spatiales. En réalité, un modèle type “UNet” synthétise déjà l’ensemble des caractéristiques de texture, de couleur et de spectre utilisables dans ce cadre de discrimination culture/adventices. De plus, les performances de l’extraction de ces caractéristiques ne permettent pas une utilisation temps réel, ce qui renforce la nécessité d’étudier l’apprentissage profond pour cette dernière étape.

Le numérique pour l’agroécologie - gestion adventices Ces travaux sont d’une grande utilité dans le contexte de transition agroécologique où des systèmes de cultures innovants, sans recours aux herbicides, sont étudiés. En effet, ces systèmes contribueront au maintien de la production agricole et permettront de répondre aux besoins d’une population humaine croissante (9,5 milliards en 2050), tout en préservant les ressources naturelles et en assurant la provision de services associés à la biodiversité des adventices (services de pollinisation, maintien de la biodiversité et des chaînes trophiques basées sur les adventices). Toutefois, ils seront moins robustes face aux aléas climatiques, aux attaques des ravageurs ou parasites, et nécessiteront une grande réactivité. Face à ce problème, les outils non-destructifs issus de la proxidtection permettront de caractériser finement la biodiversité dans les parcelles cultivées. Un suivi spatio-temporel d’une parcelle avec des outils de proxidtection combinés aux outils de télédtection et à la modélisation permettra de prédire ses évolutions et d’anticiper les interventions culturales. Ainsi, l’innovation peut être un levier pour accompagner les agriculteurs dans la transition agroécologique.

Table des figures

1.1	L'Anthropocène	1
1.2	Dynamique des systèmes : les différents imaginaires possibles	2
1.3	Remembrement des terres agricoles	3
1.4	Excédent d'azote dans la majorités des cours d'eau en 2013	4
1.5	Prolifération d'algues et asphyxie des milieux aquatiques	4
1.6	Presence de pesticide dans la majorités des cours d'eau en 2013	5
1.7	Cycle de l'agriculture de précision composé de 4 étapes	8
1.8	Évolution de la résolution spatiale	9
1.9	Simulation de l'agencement spatial de paysages	10
1.10	Concept de projet en agriculture numérique	11
1.11	Plantes bio-indicatrices, fixatrices et symbiotiques	13
1.12	Adventices comestibles	13
1.13	Adventices parasite, psychotrope et médicinal	13
1.14	Profondeur de germination en fonction de la masse des graines	14
1.15	Période préférentielle de germination de certaines adventices	14
1.16	Taux annuel de décroissance du stock semencier des adventices	15
1.17	Illustration de cinq fenêtres d'opportunité pour la gestion des adventices	16
1.18	Exemple de pulvérisateurs	17
1.19	Différentes formes de buses associées	17
1.20	Présentation du WeedSeeker en viticulture	18
1.21	Exemples de méthodes alternatives de désherbage	18
1.22	Les différents consortiums de l'ANR Challenge ROSE	19
1.23	Module de désherbage BIP-BIP	19
1.24	Les trois modules CarbonBee (caméra hyperspectrale, contrôleur, afficheur)	20
1.25	Électrode téléguidée par le bras robotique	20
1.26	Outil de binage ROSEAU	20
1.27	Schéma de l'interaction entre le système climatique physique, l'exposition et la vulnérabilité produisant des risques dans la chaîne de production.	21
2.1	Étapes clés de l'analyse d'images agronomiques	23
2.2	Spectre électromagnétique et spectre visible	24
2.3	Les différents composants d'une caméra	24
2.4	Effet de la longueur de la focale sur le champ de vision	25
2.5	Effet de l'ouverture sur la profondeur de vision	25
2.6	Effet du gain sur le bruit de l'image	25
2.7	Effet du temps d'intégration (en seconde) sur le flou de bougé	25
2.8	Filtre de Bayer	26
2.9	Triple prismes pour diviser la lumière incidente	26
2.10	Filtre rotatif à 4 bandes spectrales : bleu, vert, rouge et infrarouge	27
2.11	Modèle de distorsion	28
2.12	Correction des couleurs par apprentissage profond proposé par Afifi	29
2.13	Différentes corrections radiométriques proposé par Honkavaara	29
2.14	Comparaison des méthodes de débruitage	30
2.15	Suppression de l'ombre par apprentissage profond proposé par Zhang	30

2.16	Signatures spectrales de certains matériaux	31
2.17	Structure cellulaire d'une feuille et interaction avec le spectre électromagnétique	31
2.18	Une faible activité photo-synthétique induit une faible valeur de NDVI	32
2.19	Segmentation Grabcut	34
2.20	Architecture encodeur-décodeur d'un réseau convolutif	35
2.21	Découpage régulier sans recouvrement	36
2.22	Analyse multi-échelle des super-pixels et classification	37
2.23	Exemple de watershed sur des images drones	37
2.24	Polygone initial, contour actif, et contour avec contrainte	38
2.25	Structure d'un réseau "You Only Look Once" (YOLO)	38
2.26	Architecture d'un réseau de type Mask-RCNN	39
2.27	Espace latent et instances associées aux clusters	40
2.28	Segmentation des instances par détection des contours des nucléus	40
2.29	Types de caractéristiques	41
2.30	Caractéristiques de la forme d'une feuille	41
2.31	Détection des rangs de culture et distance des adventices	42
2.32	Exemple de transformation d'une image de haricot	42
2.33	Exemple d'un CNN pour la classification de cellule	43
2.34	Taxonomie des algorithmes d'apprentissage machine	44
2.35	Règle de décision de Bayes	45
2.36	Classification KNN : influence de K et de l'heuristique	45
2.37	Perceptron	46
2.38	Arbre de décision simple en agriculture de précision	46
2.39	Algorithme de classification K-Mean	47
2.40	Illustration de la méthode de clustering DBSCAN	47
2.41	Optimisation de l'espace latent	48
2.42	Evaluation des performances d'un modèle après apprentissage	49
2.43	Architecture CNN proposée	52
2.44	Comparaison de la précision de reconnaissance entre décomposition en ondelettes et apprentissage profond, en fonction du nombre d'échantillons	52
2.45	Algorithme proposé par [Louargant et al., 2018], (a) détection des rangs (b) constitution d'une base d'apprentissage (c) classification spectrale et (d) reclassement des adventices intra-rangs	53
2.46	Schéma de la méthode proposée par [Bah et al., 2018a]	54
2.47	Images résultantes : (a) Image prétraitée. (b) Méthode 1, chaîne de traitement. (c) Méthode 2, Yolo. Et (d) Méthode 3, mask-RCNN.	55
2.48	Chaînes de traitement	56
3.1	Parcelle pour les premières expérimentations (rouge) à gauche et parcelles utilisées par le consortium ROSEAU (droite)	57
3.2	Système optique de la caméra Airphen (Société Hipphen, France)	58
3.3	Réponse spectrale des 6 filtres interférentiels placés devant les capteurs de la caméra multispectrale	58
3.4	Robot et caméra utilisés par le consortium ROSEAU	59
3.5	Fausses couleurs à gauche et vérité terrain manuelle correspondante à droite	60
3.6	Annotation pour la discrimination des feuilles, entre culture et adventices	60
3.7	Exemples issues de la base de données CVPPP (haut) et vérité terrain (bas)	61
3.8	Exemples issues de la base de données Komatsuna (haut) et vérité terrain (bas)	61

3.9	Relation entre paramètres du monde réel et ceux du monde virtuel	62
3.10	Échiquier de calibration géométrique	63
3.11	Visualisation de la correction géométrique pour la bande 450 nm	64
3.12	Image non recalée, prise à 1.60 m	65
3.13	Disposition of each band on the Airphen multi-sensors camera	68
3.14	Each step of the alignment procedure, with (step 1) roughly corrected from affine correction and (step 2) enhancement via key-points and perspective	69
3.15	Gradient orientation in spectral band [Rabatel and Labbe, 2016]. Orientation of the gradient is not the same depending to the spectral band.	70
3.16	Example of each correction and the manufacturers results	72
3.17	Affine matrix value by height	72
3.18	Affine matrix value by height	73
3.19	features extractor performances after filtering and homography association . . .	74
3.20	key-point extractor performances	75
3.21	(up) The mean distance of detected key-point before perspective correction with 570 nm as spectral reference (down) Perspective re-projection error with GFTT using the first modality and 570 nm as reference	75
4.1	AIRPHEN camera composed of 6 sensors	83
4.2	False color in the left and corresponding manual ground truth on the right. . . .	84
4.3	Linear combination model.	86
4.4	Linear ratio model.	86
4.5	Polynomial model with Bernstein expansions between $B_{4,1}$ and $B_{4,4}$	87
4.6	Universal function approximation model (depth = 3, width = 5).	88
4.7	Dense-morphological model.	88
4.8	Input Band Filter inserted at the beginning of the model.	89
4.9	Spatial refinement block inserted at the end of a model.	89
4.10	Optimized model for standard indices	91
4.11	Network synthesis with <i>ibf</i> , <i>evalated index equation</i> , and <i>sprb</i>	93
4.12	first 80 epochs of loss of generic models with ibf in kernel size of 1.	96
4.13	Correct vegetation/soil discrimination despite shadows.	97
4.14	Vegetation/soil discrimination issue with abrupt transition between shadow and light.	97
4.15	Vegetation/soil discrimination issue caused by specular lights on leaves.	98
4.16	Visual comparison between some relevant models. NDVI (63.98 mIoU), MTVI1 (73.71 mIoU), linear 1 baseline (78.58 mIoU), dense 7 <i>ibf</i> - <i>sprb</i> (82.19 mIoU). Blue indicates sure soil, red indicates sure vegetation, and the other colors indicate uncertainty.	101
5.1	Mélange de haricot (culture) et de moutarde (adventice) dans le challenge RoSE	103
5.2	Structure d'un réseau "You Only Look Once" (YOLO)	104
5.3	Aerial view of the experimentation plot located in Montoldre (now INRAE) . .	108
5.4	The experimental set-up : the multispectral camera and the robotic platform .	109
5.5	Example of images from the CVPPP dataset (top) and their corresponding ground truth (bottom)	109
5.6	Example of RGB images from the Komatsuna dataset (top) and their corresponding ground truth (bottom)	110
5.7	Diagram of the CNN network architecture and losses (dotted). Multiple arrow show concatenation as input layer.	111

5.8	Example of input data : from left to right, the first three bands (RGB) of the Airphen multispectral camera, three edge classes and the inner of individual class	111
5.9	Universal function approximator based on DenseNet. Multiple arrows shows concatenation as input layer	112
5.10	Synthesis of the core network based on MFP-Unet.Red arrows shows MaxPooling. Blue arrows shows Conv+UpSampling. Black arrows shows Conv. Sub-scale ratio are labelled on corresponding layer. The final output is the concatenation of features UpSampling. Multiple arrows shows concatenation as input layer . .	113
5.11	Auxiliary output and classification module. Multiple arrows shows concatenation as input layer	114
5.12	Sorted histogram of the number of pixels associated to each component, the blue (resp. red) line represents the true (resp. predicted) number of pixels for each component.	115
5.13	The refinement of the CNN output through watershed and Deep-Indices. The Seed of the equation 5.5 can be seen on the “watershed seed” image.	116
5.14	Visual comparison of Komatsuna dataset after the training, the red arrow in the input image shows a green bottle	118
5.15	Visual example of LSC results for A1	121
5.16	Visual example of LSC results for A2	121
5.17	Visual example of LSC results for A3	122
5.18	Visual example of LSC results for A4	122
5.19	Example of the visual results for the Airphen dataset	124
5.20	Example of the Watershed output versus Mask-RCNN for Airphen dataset . . .	125
6.1	key steps of image analysis for crop/weed discrimination with (a) acquisition enhancement (b) vegetation indices (c) instance segmentation (d) feature extraction (e) classification	131
6.2	Aerial view of the experimentation plot located in Montoldre (now INRAE) . .	133
6.3	The Airphen six-band multispectral camera	133
6.4	Hyperparameter optimization procedure	140
6.5	Sequential Feature Selection Algorithm	141
6.6	Best features among all extracted feature for classification through ASHA algorithm. Best overall result is highlighted.	145
6.7	This figure shows the classification of leaves, from an image taken in June 2019. The blue color corresponds to well classified weeds. The green color corresponds to well classified crops. While the purple and red color correspond respectively to weeds and crops poorly classified.	146
6.8	This figure shows the classification of leaves, from parts of four images taken in June 2019. The blue color corresponds to well classified weeds. The green color corresponds to well classified crops. While the purple and red color correspond respectively to weeds and crops poorly classified. Here we see that leaves between crop rows are correctly classified, as are monocots. On the other hand, the emerging leaves of crops and some weed type leaves of dicotyledons are incorrect.	147
7.1	Pipeline final pour la détection et la discrimination des adventices	154

Liste des tableaux

2.1	Invariance et sources d'erreurs pour ces trois types de propriétés	43
3.1	Dates d'acquisition, cultures et nombre d'images acquises	59
3.2	Nombre d'images obtenues lors des campagnes d'acquisition du challenge RoSE	59
3.3	Résolution spatiale des images à différentes hauteurs	62
3.4	Erreur résiduelle, en pixel, après correction géométrique	63
3.5	Correction de la lentille pour la bande spectral 450 nm	64
3.6	list of algorithms with 3 modalities of their parameters	71
4.1	Acquisition sources and global illumination.	83
4.2	Synthesized standard indices performances : the nine best models are presented.	92
4.3	Score of DeepIndices with/without ibf and sprb for a kernel size of 1	93
4.4	Score of DeepIndices with/without ibf and sprb for a kernel size of 3	93
4.5	Score of DeepIndices with/without ibf and sprb for a kernel size of 5	94
4.6	Score of DeepIndices with/without ibf and sprb for a kernel size of 7	94
4.7	Score of DeepIndices in different modalities for a kernel size of 1 without initial image processing	95
4.8	Top optimized fixed vegetation model equations	100
5.1	Evaluation metrics for the Komatsuna dataset	118
5.2	Comparison of the ResNet 101 based on Detectron II framework of facebook and the UPGen based on Mask-RCNN. Our solution is the most effective in both metrics	119
5.3	Number of images for each sub LSC dataset	119
5.4	Evaluation metrics of the overall LSC dataset	120
5.5	Evaluation metrics of the independent sub LSC dataset	120
5.6	Comparison of our solution with state-of-the-art challengers in this dataset. . . .	123
5.7	Evaluation metrics for the Airphen dataset	123
5.8	Evaluation metrics for the Airphen dataset using Mask-RCNN	124
6.1	Synthesis of colors moments	137
6.2	Five epochs of the optimization procedure through ASHA algorithm. The objective column show the model performance	140
6.3	Hyperparameter optimization through ASHA algorithm, the objective column shows the performance of the underlying classifier	142
6.4	Feature optimization through ASHA algorithm. To show the importance of the feature optimisation process, we added the column Worst , which is the least discriminant parameters. The column Objective show the best performance of the model.	142
6.5	Best features for spatial features. Best results for a specific feature count are highlighted	143

6.6	Best classification features for each feature type through ASHA algorithm. Each line presents the results for a specific feature type. The feature count is the number of feature(s) combined for a specific feature type. The presented results are the optimal combinations for each feature counts. Column “full” is the result obtained with all the features for a specific feature type. Best results for a type of features are in bold, best results for a specific feature count are highlighted. .	143
6.7	Best features individually for each feature type for classification through ASHA algorithm. Each line presents the results for a specific feature type. The feature count is the number of feature(s) combined for a specific feature type. The presented results are the optimal combinations for each feature counts. Column “full” is the result obtained with all the features for a specific feature type. Best results for a type of features are in bold, best results for a specific feature count are highlighted.	144
6.8	Number of features after hyperparameter optimization (FV) and number of hyperparameters (HP) per feature extraction methods	150
6.9	Top features for each shape types	150
6.10	Top 10 for each spectral types	151

Bibliographie

- [Shu, 2012] (2012). Chapter 1 - a systematic view of remote sensing. In Liang, S., Li, X., and Wang, J., editors, *Advanced Remote Sensing*, pages 1 – 31. Academic Press, Boston.
- [Aberkane, 2015] Aberkane, I. J. (2015). *Économie de la connaissance*. Fondapol, Fondation pour l'innovation politique.
- [Afifi, 2018] Afifi, M. (2018). Semantic white balance : Semantic color constancy using convolutional neural network. *CoRR*, abs/1802.00153.
- [Aggarwal and Ranganathan, 2016] Aggarwal, R. and Ranganathan, P. (2016). Common pitfalls in statistical analysis : The use of correlation techniques. *Perspectives in Clinical Research*, 7 :187.
- [Ahmed et al., 2012] Ahmed, F., Al-Mamun, H. A., Bari, A. H., Hossain, E., and Kwan, P. (2012). Classification of crops and weeds from digital images : A support vector machine approach. *Crop Protection*, 40 :98–104.
- [Ahmed et al., 2016] Ahmed, N., Khan, M. U., and Asif, S. (2016). An automatic leaf based plant identification system. *Science International-Lahore*, 28 :427–430.
- [Al Naser, 2018] Al Naser, O. (2018). *Effet des conditions environnementales sur les caractéristiques morpho-physiologiques et la teneur en métabolites secondaires chez Inula montana : une plante de la médecine traditionnelle Provençale*. Theses, Université d'Avignon.
- [Albarracín et al., 2020] Albarracín, J., Oliveira, R., Hirota, M., Santos, J., and Torres, R. (2020). A soft computing approach for selecting and combining spectral bands. *Remote Sensing*, 12 :1–22.
- [Alcantarilla and Solutions, 2011] Alcantarilla, P. F. and Solutions, T. (2011). Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell*, 34(7) :1281–1298.
- [Ali et al., 2016] Ali, F., Khan, S. U., Mahmudi, M. Z., and Ullah, R. (2016). A comparison of fast, surf, eigen, harris, and msr features. *International Journal of Computer Engineering and Information Technology*, 8(6) :100.
- [Altieri, 1999] Altieri, M. A. (1999). The ecological role of biodiversity in agroecosystems. In *Invertebrate biodiversity as bioindicators of sustainable landscapes*, pages 19–31. Elsevier.
- [Altieri, 2018] Altieri, M. A. (2018). *Agroecology : the science of sustainable agriculture*. CRC Press.
- [Amziane et al., 2021] Amziane, A., Losson, O., Mathon, B., Dumenil, A., and Macaire, L. (2021). Reflectance estimation from multispectral linescan acquisitions under varying illumination—application to outdoor weed identification. *Sensors*, 21(11).
- [Angel et al., 2010] Angel, S., Parent, J., and Civco, D. L. (2010). Ten compactness properties of circles : measuring shape in geography. *The Canadian Geographer/Le Géographe canadien*, 54(4) :441–461.
- [Anthony and Gratiot, 2012] Anthony, E. J. and Gratiot, N. (2012). Coastal engineering and large-scale mangrove destruction in guyana, south america : Averting an environmental catastrophe in the making. *Ecological Engineering*, 47 :268–273.

- [Armstrong, 2019] Armstrong, R. A. (2019). Should pearson’s correlation coefficient be avoided? *Ophthalmic and Physiological Optics*, 39(5) :316–327.
- [Aubertot et al., 2005] Aubertot, J.-N., Barbier, J. M., Carpentier, A., Gril, J.-J., Guichard, L., Lucas, P., Savary, S., Savini, I., and Voltz, M. (2005). Pesticides, agriculture et environnement. *Réduire l’utilisation des pesticides et en limiter les impacts environnementaux. Rapport d’expertise scientifique collective, INRA et Cemagref (France)*.
- [Avrin et al., 2020] Avrin, G., Boffety, D., Lardy-Fontan, S., Régnier, R., Rescoussié, R., and Barbosa, V. (2020). Design and validation of testing facilities for weeding robots as part of rose challenge. In *1st International Workshop on Evaluating Progress in Artificial Intelligence (EPAI)*.
- [Bae et al., 2018] Bae, H.-J., Kim, C.-W., Kim, N., Park, B., Kim, N., Seo, J. B., and Lee, S. M. (2018). A perlin noise-based augmentation strategy for deep learning with small data samples of hrct images. *Scientific Reports*, 8.
- [Bagnoud and Recordon, 2020] Bagnoud, A. and Recordon, J. (2020). Le sikkim et la révolution biologique.
- [Bah et al., 2018a] Bah, M. D., Hafiane, A., and Canals, R. (2018a). Deep learning with unsupervised data labeling for weed detection in line crops in uav images. *Remote Sensing*, 10(11).
- [Bah et al., 2018b] Bah, M. D., Hafiane, A., and Canals, R. (2018b). Deep learning with unsupervised data labeling for weeds detection on UAV images. *CoRR*, abs/1805.12395.
- [Bah et al., 2019] Bah, M. D., Hafiane, A., and Canals, R. (2019). Crownnet : Deep network for crop row detection in uav images. *IEEE Access*, 8 :5189–5200.
- [Bai and Urtasun, 2017] Bai, M. and Urtasun, R. (2017). Deep watershed transform for instance segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2858–2866.
- [Bajwa et al., 2019] Bajwa, A. A., Khan, M. J., Bhowmik, P. C., Walsh, M., and Chauhan, B. S. (2019). Sustainable weed management. In *Innovations in Sustainable Agriculture*, pages 249–286. Springer.
- [Bajwa and Tian, 2002] Bajwa, S. and Tian, L. (2002). Multispectral cir image calibration for cloud shadow and soil background influence using intensity normalization. *Applied engineering in agriculture*, 18 :627–635.
- [Bakhshipour et al., 2017] Bakhshipour, A., Jafari, A., Nassiri, S. M., and Zare, D. (2017). Weed segmentation using texture features extracted from wavelet sub-images. *Biosystems Engineering*, 157 :1–12.
- [Banerji et al., 2012] Banerji, S., Verma, A., and Liu, C. (2012). Lbp and color descriptors for image classification. In *Cross Disciplinary Biometric Systems*, pages 205–225. Springer.
- [Bannari et al., 1995] Bannari, A., Morin, D., Bonn, F., and Huete, A. (1995). A review of vegetation indices. *Remote sensing reviews*, 13(1-2) :95–120.
- [Barbu, 2009] Barbu, A. (2009). Training an active random field for real-time image denoising. *IEEE Transactions on Image Processing*, 18(11) :2451–2462.
- [Bareth et al., 2016] Bareth, G., Bolten, A., Gnyp, M. L., Reusch, S., and Jasper, J. (2016). Comparison of Uncalibrated Rgbvi with Spectrometer-Based Ndvi Derived from Uav Sensing Systems on Field Scale. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41B8 :837–843.

- [Bay et al., 2006] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf : Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer.
- [Bedos et al., 2002] Bedos, C., Cellier, P., Calvet, R., Barriuso, E., and Gabrielle, B. (2002). Mass transfer of pesticides into the atmosphere by volatilization from soils and plants : overview. *Agronomie*, 22(1) :21–33.
- [Beketov et al., 2013] Beketov, M. A., Kefford, B. J., Schäfer, R. B., and Liess, M. (2013). Pesticides reduce regional biodiversity of stream invertebrates. *Proceedings of the National Academy of Sciences*, 110(27) :11039–11043.
- [Bell and Dee, 2019] Bell, J. and Dee, H. M. (2019). Leaf segmentation through the classification of edges.
- [Bellon-Maurel and Huyghe, 2019] Bellon-Maurel, V. and Huyghe, C. (2019). Avant-propos. *Sciences Eaux Territoires*, (3) :3–5.
- [Bellon Maurel and HUYGHE, 2019] Bellon Maurel, V. and HUYGHE, C. (2019). Digital agriculture, a (r)evolution in progress in the territories? - Foreword. *Sciences Eaux & Territoires*, (29) :3–5.
- [Belongie et al., 2002] Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, 24(4) :509–522.
- [Ben Abdallah et al., 2016] Ben Abdallah, F., Goffart, J. P., and Philippe, W. (2016). Utilisation de la fluorescence chlorophyllienne pour l'évaluation du statut azoté des cultures (synthèse bibliographique). *Biotechnologie, agronomie, société et environnement*, 20(1) :83–93.
- [Bendaoud, 2020] Bendaoud, A. (2020). *Le semis direct sous couvert végétal vivant : comme une solution alternative aux systèmes de cultures conventionnels en zones arides des steppes algériennes*. PhD thesis, Université Mohamed Khider de Biskra.
- [Bergstra and Bengio, 2012] Bergstra, J. and Bengio, Y. (2012). Random search for hyperparameter optimization. *Journal of machine learning research*, 13(2).
- [Bianchi et al., 2006] Bianchi, F. J., Booij, C., and Tscharrntke, T. (2006). Sustainable pest regulation in agricultural landscapes : a review on landscape composition, biodiversity and natural pest control. *Proceedings of the Royal Society B : Biological Sciences*, 273(1595) :1715–1727.
- [Birner et al., 2021] Birner, R., Daum, T., and Pray, C. (2021). Who drives the digital revolution in agriculture? a review of supply-side trends, players and challenges. *Applied Economic Perspectives and Policy*, 43(4) :1260–1285.
- [Blackburn and Vignola, 2012] Blackburn, G. and Vignola, F. (2012). Spectral distributions of diffuse and global irradiance for clear and cloudy periods. In *World Renewable Energy Forum, Denver, CO*.
- [Boissinot et al., 2014] Boissinot, A., Braconnier, H., Braconnier, J.-C., Braconnier, H., Braconnier, N., Morin-Pinaud, S., and Grillet, P. (2014). *Terres de bocage : concilier nature et agriculture*. Editions Ouest-France.
- [Bokhovkin and Burnaev, 2019] Bokhovkin, A. and Burnaev, E. (2019). Boundary loss for remote sensing imagery semantic segmentation. *CoRR*, abs/1905.07852.
- [Bonneau et al., 2021] Bonneau, M., Benet, B., Labrune, Y., Bailly, J., Ricard, E., and Canario, L. (2021). Predicting sow postures from video images : Comparison of convolutional neural networks and segmentation combined with support vector machines under various training and testing setups. *Biosystems Engineering*, 212 :19–29.

- [Bonneau et al., 2020] Bonneau, M., Vayssade, J.-A., Troupe, W., and Arquet, R. (2020). Outdoor animal tracking combining neural network and time-lapse cameras. *Computers and Electronics in Agriculture*, 168 :105150.
- [Bookstein, 1989] Bookstein, F. L. (1989). Principal warps : thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6) :567–585.
- [Bossard et al., 2016] Bossard, C., Santin, G., and Canu, I. G. (2016). Suicide among farmers in france : Occupational factors and recent trends. *Journal of Agromedicine*, 21(4) :310–315. PMID : 27409004.
- [Bossu, 2007] Bossu, J. (2007). *Segmentation d’images pour la localisation d’adventices : application à la réalisation d’un système de vision pour une pulvérisation spécifique en temps réel*. PhD thesis, Dijon.
- [Bouguet, 2001] Bouguet, J.-Y. (2001). Camera calibration toolbox for matlab.
- [Bourguignon and Bourguignon, 2008] Bourguignon, C. and Bourguignon, L. (2008). *Le sol, la terre et les champs : pour retrouver une agriculture saine*. Lulu.
- [Brabandere et al., 2017] Brabandere, B. D., Neven, D., and Gool, L. V. (2017). Semantic instance segmentation with a discriminative loss function. *CoRR*, abs/1708.02551.
- [Brainard and Wandell, 1992] Brainard, D. H. and Wandell, B. A. (1992). Asymmetric color matching : how color appearance depends on the illuminant. *JOSA A*, 9(9) :1433–1448.
- [Brilhador et al., 2019] Brilhador, A., Gutoski, M., Hattori, L. T., de Souza Inácio, A., Lazaretti, A. E., and Lopes, H. S. (2019). Classification of weeds and crops at the pixel-level using convolutional neural networks and data augmentation. In *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6. IEEE.
- [Brown, 2006] Brown, A. (2006). Spectral curve fitting for automatic hyperspectral data analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6) :1601–1608.
- [Buades et al., 2005] Buades, A., Coll, B., and Morel, J. (2005). A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2) :490–530.
- [Cabrera-Bosquet et al., 2012] Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M. D., and Luis Araus, J. (2012). High-throughput phenotyping and genomic selection : The frontiers of crop breeding converge f. *Journal of integrative plant biology*, 54(5) :312–320.
- [Caquet et al., 2020] Caquet, T., Gascuel, C., and Tixier-Boichard, M. (2020). *Agroécologie : des recherches pour la transition des filières et des territoires*. Quae.
- [Carbonne, 2020] Carbonne, B. (2020). *Le rôle des interactions biotiques dans un système proie-prédateur : le cas de la prédation et de la régulation des graines d’adventices par les carabes*. Theses, Université Bourgogne Franche-Comté.
- [Carles, 2007] Carles, P. (2007). *Volem rien foutre al païs*. CP Productions [éd.].
- [Caussanel, 1989] Caussanel, J. (1989). Nuisibilité et seuils de nuisibilité des mauvaises herbes dans une culture annuelle : situation de concurrence bispécifique. *Agronomie*, 9(3) :219–240.
- [Cerutti et al., 2014] Cerutti, G., Tougne, L., Coquin, D., and Vacavant, A. (2014). Leaf margins as sequences : A structural approach to leaf identification. *Pattern Recognition Letters*, 49 :177–184.
- [Cerutti et al., 2011] Cerutti, G., Tougne, L., Mille, J., Vacavant, A., and Coquin, D. (2011). Guiding active contours for tree leaf segmentation and identification. volume 1177.

- [Cerutti et al., 2013] Cerutti, G., Tougne, L., Mille, J., Vacavant, A., and Coquin, D. (2013). Understanding leaves in natural images - a model-based approach for tree species identification. *Computer Vision and Image Understanding*, 117.
- [Champ et al., 2020] Champ, J., Mora-Fallas, A., Goëau, H., Mata-Montero, E., Bonnet, P., and Joly, A. (2020). Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots. *Applications in Plant Sciences*, 8(7) :e11373.
- [Chaudhury and Dabhade, 2016] Chaudhury, K. N. and Dabhade, S. D. (2016). Fast and probably accurate bilateral filtering. *CoRR*, abs/1603.08109.
- [Chaumeil, 1939] Chaumeil, L. (1939). Les travaux et les jours d'un paysan auvergnat. In *Annales de géographie*, volume 48, pages 138–162. JSTOR.
- [Chauvel et al., 2018] Chauvel, B., Darmency, H., Munier-Jolain, N., and Rodriguez, A. (2018). *Gestion durable de la flore adventice des cultures*. Editions Quae.
- [Chebrolu et al., 2017] Chebrolu, N., Lottes, P., Schaefer, A., Winterhalter, W., Burgard, W., and Stachniss, C. (2017). Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. *The International Journal of Robotics Research*, 36 :027836491772051.
- [Chen et al., 2016] Chen, H., Qi, X., Yu, L., and Heng, P.-A. (2016). Dcan : Deep contour-aware networks for accurate gland segmentation.
- [Chen et al., 2018] Chen, S., Shen, H., Li, C., and Xin, J. H. (2018). Normalized total gradient : A new measure for multispectral image registration. *IEEE Transactions on Image Processing*, 27(3) :1297–1310.
- [Choi et al., 2019] Choi, H., Lee, H.-J., You, H.-J., Rhee, S.-Y., and Jeon, W.-S. (2019). Comparative analysis of generalized intersection over union and error matrix for vegetation cover classification assessment. *Sensors and Materials*, 31 :3849.
- [Colbach et al., 2021] Colbach, N., Colas, F., Cordeau, S., Maillot, T., Queyrel, W., Villerd, J., and Moreau, D. (2021). The florsys crop-weed canopy model, a tool to investigate and promote agroecological weed management. *Field Crops Research*, 261 :108006.
- [Colbach et al., 2010] Colbach, N., Kurstjens, D., Munier-Jolain, N., Dalbiès, A., and Doré, T. (2010). Assessing non-chemical weeding strategies through mechanistic modelling of blackgrass (*alopecurus myosuroides huds.*) dynamics. *European Journal of Agronomy*, 32(3) :205–218.
- [cordeau et al., 2016] cordeau, S., Dessaint, F., Deneuil, C., Bonin, L., Vuillemin, F., Delattre, M., Alain, R., Guillemain, J.-P., and Chauvel, B. (2016). La nuisibilité directe des adventices en grandes cultures : quelles réponses nous apportent les essais désherbages ? 23. Conférence du COLUMA - Journées Internationales sur la Lutte contre les Mauvaises Herbes. Poster - SPEEAGESTADAGROSUPINRA.
- [Cordeau et al., 2016] Cordeau, S., Dessaint, F., Denieul, C., Bonin, L., Vuillemin, F., Delattre, M., Rodriguez, A., Guillemain, J., and Chauvel, B. (2016). La nuisibilité direct des adventices en grandes cultures : quelles réponses nous apportent les essais désherbage. In *23e Conférence du COLUMA-journées internationales sur la lutte contre les mauvaises herbes, DIJON*, volume 6.
- [CORNET et al., 2018] CORNET, D., Cormier, F., Arnau, G., Mell, P., Pavis, C., Desfontaines, L., and Tournebize, R. (2018). Recent advances in yam crop physiology and perspectives for breeding strategies. In *2. International Yam Crop Science Workshop (IYCSW)*, Tokyo, Japan. Tokyo University of Agriculture (Tokyo NODAI). Tokyo, JPN.

- [Cossart and Fressard, 2019] Cossart, E. and Fressard, M. (2019). Érosion des sols viticoles : influence de la structuration globale du bassin-versant : stratégies collectives pour casser l’enchaînement des processus érosifs. *Revue des oenologues et des techniques vitivinicoles et oenologiques : magazine trimestriel d’information professionnelle*, 46(171) :24–26.
- [Cramer et al., 2011] Cramer, G. R., Urano, K., Delrot, S., Pezzotti, M., and Shinozaki, K. (2011). Effects of abiotic stress on plants : a systems biology perspective. *BMC plant biology*, 11(1) :1–14.
- [Cui et al., 2019] Cui, Y., Zhang, G., Liu, Z., Xiong, Z., and Hu, J. (2019). A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images. *Medical and Biological Engineering and Computing*, 57(9) :2027–2043.
- [Dainese and Martin, 2019] Dainese, M. and Martin, E. A. (2019). A global synthesis reveals biodiversity-mediated benefits for crop production. *Science Advances*, 5(10).
- [Dantas Dias Junior et al., 2019] Dantas Dias Junior, J., Backes, A., and Escarpinati, M. (2019). Detection of control points for uav-multispectral sensed data registration through the combining of feature descriptors. pages 444–451.
- [Day, 2017] Day, E. O. (2017). Earth overshoot day. *Hentet fra <https://www.overshootday.org>*.
- [De Castro et al., 2018] De Castro, A. I., Torres-Sánchez, J., Peña, J. M., Jiménez-Brenes, F. M., Csillik, O., and López-Granados, F. (2018). An automatic random forest-obia algorithm for early weed mapping between and within crop rows using uav imagery. *Remote Sensing*, 10(2).
- [Délye, 2017] Délye, C. (2017). État des lieux des résistances aux herbicides en France : grandes cultures et maraîchage. In *Journées d’Échanges sur les Résistances aux produits de protection des plantes 2017*, volume Comptes-rendus des Journées d’Échanges sur les Résistances aux produits de protection des plantes 2017, page 49 p., Bordeaux, France. Réseau de Réflexion et de Recherche sur les Résistances aux Pesticides (R4P). FRA.
- [Derrode and Pieczynski, 2003] Derrode, S. and Pieczynski, W. (2003). Segmentation non supervisée d’images par chaîne de markov couple.
- [Descour and Dereniak, 1995] Descour, M. and Dereniak, E. (1995). Computed-tomography imaging spectrometer : experimental calibration and reconstruction results. *Applied optics*, 34(22) :4817–4826.
- [Dhawale et al., 2021] Dhawale, N., Adamchuk, V., Prasher, S., and Viscarra Rossel, R. (2021). Evaluating the precision and accuracy of proximal soil vis-nir sensors for estimating soil organic matter and texture. *Soil Systems*, 5 :48.
- [Di Gregorio et al., 2008] Di Gregorio, R. et al. (2008). Review and comparison of the metrics proposed for measuring the distance between two rigid-body poses. *Quaderni del DIEM-GMA, Atti della Seconda Giornata di Studio Ettore Funaioli*, pages 47–58.
- [Dixit et al., 2019] Dixit, A., Goswami, A., and Jain, S. (2019). Development and evaluation of a new “snow water index (swi)” for accurate snow cover delineation. *Remote Sensing*, 11(23).
- [Donoser and Bischof, 2006] Donoser, M. and Bischof, H. (2006). Efficient maximally stable extremal region (mser) tracking. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, volume 1, pages 553–560. Ieee.
- [dos Reis Martins et al., 2009] dos Reis Martins, M., Corá, J. E., Jorge, R. F., and Marcelo, A. V. (2009). Crop type influences soil aggregation and organic matter under no-tillage. *Soil and Tillage Research*, 104(1) :22–29.

- [Douarre et al., 2019] Douarre, C., Crispim-Junior, C. F., Gelibert, A., Tougne, L., and Rousseau, D. (2019). A strategy for multimodal canopy images registration. In *7th International Workshop on Image Analysis Methods in the Plant Sciences*, Lyon, France.
- [Dubuisson et al., 2007] Dubuisson, C., Héraud, F., Leblanc, J.-C., Gallotti, S., Flamand, C., Blateau, A., Quenel, P., and Volatier, J.-L. (2007). Impact of subsistence production on the management options to reduce the food exposure of the martinican population to chlordecone. *Regulatory Toxicology and Pharmacology*, 49(1) :5–16.
- [Dufour and Piégay, 2009] Dufour, S. and Piégay, H. (2009). From the myth of a lost paradise to targeted river restoration : forget natural references and focus on human benefits. *River research and applications*, 25(5) :568–581.
- [Dunsmore, 1966] Dunsmore, I. (1966). A bayesian approach to classification. *Journal of the Royal Statistical Society : Series B (Methodological)*, 28(3) :568–577.
- [Dupuy et al., 2015] Dupuy, J., Heitz, E., Iehl, J.-C., Poulin, P., and Ostromoukhov, V. (2015). Extracting microfacet-based brdf parameters from arbitrary materials with power iterations. In *Computer Graphics Forum*, volume 34, pages 21–30. Wiley Online Library.
- [Dutta and Zisserman, 2019] Dutta, A. and Zisserman, A. (2019). The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, New York, NY, USA. ACM.
- [El Amine et al., 2018] El Amine, S., Rapior, S., Morel, S., and Alain, D. (2018). *MON JARDIN PERMACULTUREL / MY PERMACULTURAL GARDEN*. PhD thesis.
- [Elbeltagi et al., 2021] Elbeltagi, A., Kumari, N., Dharpure, J. K., Mokhtar, A., Alsafadi, K., Kumar, M., Mehdinejadani, B., Ramezani Etedali, H., Brouziyne, Y., Towfiqul Islam, A. R. M., and Kuriqi, A. (2021). Prediction of combined terrestrial evapotranspiration index (ctei) over large river basin based on machine learning approaches. *Water*, 13(4).
- [Erdem and Tari, 2010] Erdem, A. and Tari, S. (2010). A similarity-based approach for shape classification using aslan skeletons. *Pattern Recognition Letters*, 31(13) :2024–2032. Meta-heuristic Intelligence Based Image Processing.
- [Eschweiler et al., 2018] Eschweiler, D., Spina, T. V., Choudhury, R. C., Meyerowitz, E., Cunha, A., and Stegmaier, J. (2018). Cnn-based preprocessing to optimize watershed-based cell segmentation in 3d confocal microscopy images. *CoRR*, abs/1810.06933.
- [Estrada et al., 2004] Estrada, F. J., Jepson, A. D., and Chennubhotla, C. (2004). Spectral embedding and min cut for image segmentation. In *Bmvc*, pages 1–10.
- [Etter et al., 2011] Etter, B., Tilley, E., Khadka, R., and Udert, K. M. (2011). Low-cost struvite production using source-separated urine in nepal. *Water research*, 45(2) :852–862.
- [Fahlgren et al., 2015] Fahlgren, N., Feldman, M., Gehan, M. A., Wilson, M. S., Shyu, C., Bryant, D. W., Hill, S. T., McEntee, C. J., Warnasooriya, S. N., Kumar, I., et al. (2015). A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in setaria. *Molecular plant*, 8(10) :1520–1535.
- [Fawakherji et al., 2019] Fawakherji, M., Potena, C., Bloisi, D., Imperoli, M., Pretto, A., and Nardi, D. (2019). *UAV Image Based Crop and Weed Distribution Estimation on Embedded GPU Boards*, pages 100–108.
- [Fawcett, 2006] Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8) :861–874. ROC Analysis in Pattern Recognition.
- [Federau, 2016] Federau, A. (2016). *Philosophie de l'Anthropocène : interprétations et épistémologie*. PhD thesis, éditeur non identifié.

- [Finlayson et al., 2015] Finlayson, G. D., Mackiewicz, M., and Hurlbert, A. (2015). Color correction using root-polynomial regression. *IEEE Transactions on Image Processing*, 24(5) :1460–1470.
- [Fischler and Bolles, 1981] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6) :381–395.
- [Fournier, 2011] Fournier, A. (2011). *Influence de la structure des couverts végétaux en télédétection de la fluorescence chlorophyllienne*. Theses, Ecole Polytechnique X.
- [Fu et al., 2018] Fu, T., Ma, L., Li, M., and Johnson, B. A. (2018). Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery. *Journal of Applied Remote Sensing*, 12(2) :1 – 21.
- [Fukuoka, 1983] Fukuoka, M. (1983). La révolution d’un seul brin de paille. *Guy Trédaniel Editions, Paris*.
- [Gaetano et al., 2018] Gaetano, R., Ienco, D., Ose, K., and Cresson, R. (2018). A Two-Branch CNN Architecture for Land Cover Classification of PAN and MS Imagery. *Remote Sensing*, 10(11) :1746.
- [Galo et al., 2006] Galo, M., Hasegawa, J., Tommaselli, A., and Imai, N. (2006). Registration analysis and inner calibration of a three ccd multispectral frame camera. In *EUROCOW 2006 INTERNATIONAL CALIBRATION AND ORIENTATION WORKSHOP*, volume 25, page 27.
- [Gao et al., 2018] Gao, J., Liao, W., Nuyttens, D., Lootens, P., Vangeyte, J., Pižurica, A., He, Y., and Pieters, J. G. (2018). Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. *International Journal of Applied Earth Observation and Geoinformation*, 67 :43–53.
- [Gao et al., 2013] Gao, L., Li, D., and Zhang, Y. (2013). Research of weeds classification system based on shape feature. *J. Softw.*, 8 :1736–1742.
- [Gao et al., 2016] Gao, Z., Wang, L., Zhou, L., and Zhang, J. (2016). Hep-2 cell image classification with deep convolutional neural networks. *IEEE journal of biomedical and health informatics*, 21(2) :416–428.
- [Gay and Wery, 2003] Gay, F. and Wery, J. (2003). Elaboration d’indicateurs pour l’évaluation des risques de pollution nitrique et de l’impact de stress hydrique sur le rendement en cultures irriguées : application au cas de la laitue sur le périmètre des sources perrier.
- [Gehan et al., 2017] Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., Doust, A. N., Feldman, M. J., Gilbert, K. B., Hodge, J. G., et al. (2017). Plantcv v2 : Image analysis software for high-throughput plant phenotyping. *PeerJ*, 5 :e4088.
- [Geiger et al., 2010] Geiger, F., Bengtsson, J., Berendse, F., Weisser, W. W., Emmerson, M., Morales, M. B., Ceryngier, P., Liira, J., Tschardtke, T., Winqvist, C., Eggers, S., Bommarco, R., Pärt, T., Bretagnolle, V., Plantegenest, M., Clement, L. W., Dennis, C., Palmer, C., Oñate, J. J., Guerrero, I., Hawro, V., Aavik, T., Thies, C., Flohre, A., Hänke, S., Fischer, C., Goedhart, P. W., and Inchausti, P. (2010). Persistent negative effects of pesticides on biodiversity and biological control potential on european farmland. *Basic and Applied Ecology*, 11(2) :97–105.
- [Geusebroek et al., 2000] Geusebroek, J.-M., van den Boomgaard, R., Smeulders, A., and Dev, A. (2000). Color and scale : The spatial structure of color images. pages 331–341.

- [Gilliot et al., 2018] Gilliot, J.-M., Michelin, J., Faroux, R., Domenzain, L. M., and Fallet, C. (2018). Correction of in-flight luminosity variations in multispectral UAS images, using a luminosity sensor and camera pair for improved biomass estimation in precision agriculture. In *Conference on Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping!!!*, Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping III.
- [Girshick, 2015] Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448.
- [Girshick et al., 2014] Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587.
- [Gitelson, 2004] Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of plant physiology*, 161(2) :165–173.
- [Godde et al., 2021] Godde, C., Mason-D’Croz, D., Mayberry, D., Thornton, P., and Herrero, M. (2021). Impacts of climate change on the livestock food supply chain ; a review of the evidence. *Global Food Security*, 28 :100488.
- [Gomes and Zheng, 2020] Gomes, D. P. S. and Zheng, L. (2020). Leaf segmentation and counting with deep learning : on model certainty, test-time augmentation, trade-offs.
- [Göring et al., 2012] Göring, C., Fröhlich, B., and Denzler, J. (2012). Semantic segmentation using grabcut. In *VISAPP*.
- [Green et al., 2021] Green, A. G., Abdulai, A.-R., Duncan, E., Glaros, A., Campbell, M., Newell, R., Quarshie, P., KC, K. B., Newman, L., Nost, E., and Fraser, E. D. G. (2021). A scoping review of the digital agricultural revolution and ecosystem services : implications for canadian policy and research agendas. *FACETS*, 6 :1955–1985.
- [Griepentrog et al., 2005] Griepentrog, H. W., Nørremark, M., Nielsen, H., and Blackmore, S. (2005). Seed mapping of sugar beet. *Precision Agriculture*, 6 :157–165.
- [Grozinger and Flenniken, 2019] Grozinger, C. M. and Flenniken, M. L. (2019). Bee viruses : Ecology, pathogenicity, and impacts. *Annual review of entomology*, 64 :205–226.
- [Gutierrez Ortega et al., 2021] Gutierrez Ortega, J. A., Castillo, S. E., Gehan, M., and Fahlgren, N. (2021). Segmentation of overlapping plants in multi-plant image time series. *Earth and Space Science Open Archive*, page 5.
- [Gée et al., 2021] Gée, C., Denimal, E., Merienne, J., and Larmure, A. (2021). Evaluation of weed impact on wheat biomass by combining visible imagery with a plant growth model : towards new non-destructive indicators for weed competition. *Precision Agriculture*, 22 :1–19.
- [Haan et al., 2021] Haan, N. L., Iuliano, B. G., Gratton, C., and Landis, D. A. (2021). Designing agricultural landscapes for arthropod-based ecosystem services in north america. In *Advances in Ecological Research*, volume 64, pages 191–250. Elsevier.
- [Hafiz and Bhat, 2020] Hafiz, A. M. and Bhat, G. M. (2020). A survey on instance segmentation : state of the art. *International Journal of Multimedia Information Retrieval*, 9(3) :171–189.
- [Hao et al., 2014] Hao, Y., Dongyan, Z., Linsheng, H., and Jinling, Z. (2014). Wavelet-based threshold denoising for imaging hyperspectral data. *International Journal of Agricultural and Biological Engineering*, 7(3) :36–42.

- [Harris et al., 1988] Harris, C., Stephens, M., et al. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer.
- [Hasan et al., 2021] Hasan, A. S. M., Soheli, F., Diepeveen, D., Laga, H., and Jones, M. (2021). A survey of deep learning techniques for weed detection from images. *Computers and Electronics in Agriculture*, 184 :106067.
- [Hassanein et al., 2018] Hassanein, M., Lari, Z., and El-Sheimy, N. (2018). A new vegetation segmentation approach for cropped fields based on threshold detection from hue histograms. *Sensors*, 18 :1253.
- [Henrich et al., 2009] Henrich, V., Götze, E., Jung, A., Sandow, C., Thürkow, D., and Gläßer, C. (2009). Development of an online indices database : Motivation, concept and implementation. In *Proceedings of the 6th EARSeL Imaging Spectroscopy SIG Workshop Innovative Tool for Scientific and Commercial Environment Applications, Tel Aviv, Israel*, pages 16–18.
- [Hentschel and Page, 2003] Hentschel, M. L. and Page, N. W. (2003). Selection of descriptors for particle shape characterization. *Particle & Particle Systems Characterization : Measurement and Description of Particle Properties and Behavior in Powders and Other Disperse Systems*, 20(1) :25–38.
- [Herrington, 2021] Herrington, G. (2021). Update to limits to growth : Comparing the world3 model with empirical data. *Journal of Industrial Ecology*, 25(3) :614–626.
- [Herrmann et al., 2013] Herrmann, I., Shapira, U., Kinast, S., Karnieli, A., and Bonfil, D. (2013). Ground-level hyperspectral imagery for detecting weeds in wheat fields. *Precision Agriculture*, 14.
- [Hertel et al., 2020] Hertel, L., Collado, J., Sadowski, P., Ott, J., and Baldi, P. (2020). Sherpa : Robust hyperparameter optimization for machine learning. *SoftwareX*, 12 :100591.
- [Hinimbio, 2019] Hinimbio, T. P. (2019). *Réhabilitation de la fertilité des sols par usage des bioressources (Crotalaria juncea L. et Brachiaria ruziziensis G. & E.) en zone cotonnière de l'Extrême-Nord, Cameroun*. Theses, Université de Maroua (Cameroun).
- [Holzer, 2011] Holzer, S. (2011). *Sepp Holzer's permaculture : a practical guide to small-scale, integrative farming and gardening*. Chelsea Green Publishing.
- [Honkavaara and Khoramshahi, 2018] Honkavaara, E. and Khoramshahi, E. (2018). Radiometric correction of close-range spectral image blocks captured using an unmanned aerial vehicle with a radiometric block adjustment. *Remote Sensing*, 10(2).
- [Hossin and Sulaiman, 2015] Hossin, M. and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, 5(2) :1.
- [Hu et al., 2020] Hu, K., Coleman, G., Zeng, S., Wang, Z., and Walsh, M. (2020). Graph weeds net : A graph-based deep learning method for weed recognition. *Computers and Electronics in Agriculture*, 174 :105520.
- [Hu et al., 2021] Hu, K., Wang, Z., Coleman, G., Bender, A., Yao, T., Zeng, S., Song, D., Schumann, A., and Walsh, M. (2021). Deep learning techniques for in-crop weed identification : A review. *arXiv preprint arXiv :2103.14872*.
- [Hu, 1962] Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2) :179–187.
- [Huang et al., 2016] Huang, G., Liu, Z., and Weinberger, K. Q. (2016). Densely connected convolutional networks. *CoRR*, abs/1608.06993.

- [Ishak et al., 2009] Ishak, A., Hussain, A., and Mustafa, M. (2009). Weed image classification using gabor wavelet and gradient field distribution. *Computers and Electronics in Agriculture - COMPUT ELECTRON AGRIC*, 66 :53–61.
- [Jacobsen et al., 2016] Jacobsen, J.-H., Gemert, J., Lou, Z., and Smeulders, A. (2016). Structured receptive fields in cnns. pages 2610–2619.
- [Jadon, 2020] Jadon, S. (2020). A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–7. IEEE.
- [Jain and Healey, 1998] Jain, A. and Healey, G. (1998). A multiscale representation including opponent color features for texture recognition. *IEEE Transactions on Image Processing*, 7(1) :124–128.
- [Jain and Zongker, 1997] Jain, A. and Zongker, D. (1997). Feature selection : evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2) :153–158.
- [Jancovici, 2017] Jancovici, J.-M. (2017). *L’avenir climatique-Quel temps ferons-nous ?* Média Diffusion.
- [Jang and Son, 2019] Jang, S. and Son, Y. (2019). Empirical evaluation of activation functions and kernel initializers on deep reinforcement learning. In *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1140–1142.
- [Jiang and Li, 2020] Jiang, Y. and Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping : a review. *Plant Phenomics*, 2020.
- [Jin et al., 2013] Jin, X.-l., Diao, W.-y., Xiao, C.-h., Wang, F.-y., Chen, B., Wang, K.-r., and Li, S.-k. (2013). Estimation of wheat agronomic parameters using new spectral indices. *PLOS ONE*, 8(8) :1–9.
- [Jinru and Su, 2017] Jinru, X. and Su, B. (2017). Significant remote sensing vegetation indices : A review of developments and applications. *Journal of Sensors*, 2017 :1–17.
- [Jonathan Storkey, 2007] Jonathan Storkey, D. B. W. (2007). Mini-review managing arable weeds for biodiversity. *Pest Management Science*, (63) :517–523.
- [Jones, 2009] Jones, G. (2009). *Modélisation d’images agronomiques-application a la reconnaissance d’adventices par imagerie pour une pulvérisation localisée*. PhD thesis.
- [Joshi et al., 2020] Joshi, E., Sasode, D. S., Singh, N., and Chouhan, N. (2020). Revolution of indian agriculture through drone technology. *Biotica Research Today*, 2(5 Spl.) :174–176.
- [Jouan et al., 2022] Jouan, J., Jacquet, F., and Huyghe, C. (2022). La recherche pour une agriculture sans pesticides : un cadre disruptif aujourd’hui pour construire les solutions de demain. In *Zéro pesticide : un nouveau paradigme de recherche pour une agriculture durable*. Editions QUAE.
- [Jung et al., 2019] Jung, H., Lodhi, B., and Kang, J. (2019). An automatic nuclei segmentation method based on deep convolutional neural networks for histopathology images. *BMC Biomedical Engineering*, 1.
- [Junior et al., 2019] Junior, J. D. D., Backes, A. R., and Escarpinati, M. C. (2019). Detection of control points for uav-multispectral sensed data registration through the combining of feature descriptors. In *VISIGRAPP*.
- [Kabiri et al., 2011] Kabiri, P., Pandi, M., and Nejat, S. (2011). Ndvi optimization using genetic algorithm. pages 1–5.

- [Kakamoukas et al., 2021] Kakamoukas, G., Sarigiannidis, P., Maropoulos, A., Lagkas, T., Zaralis, K., and Karaiskou, C. (2021). Towards climate smart farming—a reference architecture for integrated farming systems. *Telecom*, 2(1) :52–74.
- [Kamilaris and Prenafeta-Boldú, 2018] Kamilaris, A. and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture : A survey. *Computers and electronics in agriculture*, 147 :70–90.
- [Kamnitsas et al., 2018] Kamnitsas, K., de Castro, D. C., Folgoc, L. L., Walker, I., Tanno, R., Rueckert, D., Glocker, B., Criminisi, A., and Nori, A. V. (2018). Semi-supervised learning via compact latent space clustering. *CoRR*, abs/1806.02679.
- [Kamoun, 2019] Kamoun, E. (2019). Image registration : From sift to deep learning.
- [Kan et al., 2020] Kan, Z.-R., Liu, Q.-Y., He, C., Jing, Z.-H., Virk, A. L., Qi, J.-Y., Zhao, X., and Zhang, H.-L. (2020). Responses of grain yield and water use efficiency of winter wheat to tillage in the north china plain. *Field Crops Research*, 249 :107760.
- [Keller, 2018] Keller, A. (2018). Conférence sur les risques d’effondrement et les chemins de la résilience territoriale.
- [Khan et al., 2020] Khan, A., Ilyas, T., Umraiz, M., Mannan, Z. I., and Kim, H. (2020). Cednet : crops and weeds segmentation for smart farming using a small cascaded encoder-decoder architecture. *Electronics*, 9(10) :1602.
- [Khanna et al., 2017] Khanna, R., Sa, I., Nieto, J., and Siegwart, R. (2017). On field radiometric calibration for multispectral cameras. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6503–6509.
- [Khater et al., 2020] Khater, I., Nabi, I., and Hamarneh, G. (2020). A review of super-resolution single-molecule localization microscopy cluster analysis and quantification methods. *Patterns*, 1 :100038.
- [Kim and Cho, 2020] Kim, J. and Cho, J. (2020). Exploring a multimodal mixture-of-yolos framework for advanced real-time object detection. *Applied Sciences*, 10(2).
- [Kisinyo et al., 2011] Kisinyo, P., Ng’etich, W., Othieno, C., Okalebo, J., and Opile, W. (2011). Épuisement du phosphore : les pays acp doivent-ils s’ en préoccuper. *Quelles sont les perspectives attendues en matière de recherche et de politique*.
- [Kokhan and Vostokov, 2020] Kokhan, S. and Vostokov, A. (2020). Using vegetative indices to quantify agricultural crop characteristics. *Journal of Ecological Engineering*, 21 :120–127.
- [Kolli et al., 2021] Kolli, M. K., Opp, C., and Groll, M. (2021). Estimation of soil erosion and sediment yield concentration across the kolleru lake catchment using gis. *Environmental Earth Sciences*, 80(4) :1–14.
- [Koot, 2014] Koot, B. T. (2014). *Weed detection with Unmanned Aerial Vehicles in agricultural systems*. PhD thesis, Wageningen University and Research Centre. Registration number : 891024465110.
- [Korres et al., 2019] Korres, N. E., Burgos, N. R., Travlos, I., Vurro, M., Gitsopoulos, T. K., Varanasi, V. K., Duke, S. O., Kudsk, P., Brabham, C., Rouse, C. E., and Salas-Perez, R. (2019). Chapter six - new directions for integrated weed management : Modern technologies, tools and knowledge discovery. volume 155 of *Advances in Agronomy*, pages 243–319. Academic Press.
- [Koyejo et al., 2014] Koyejo, O., Natarajan, N., Ravikumar, P., and Dhillon, I. S. (2014). Consistent binary classification with generalized performance metrics. In *NIPS*, volume 27, pages 2744–2752. Citeseer.

- [Kudsk et al., 2020] Kudsk, P., Sønderskov, M., Bonin, L., Gonzalez-Andujar, J., Jensen, J. E., Melander, B., Moonen, A., Riemens, M., Sattin, M., Schaffner, U., and Storkey, J. (2020). Iwmpraise – an eu horizon 2020 project providing integrated weed management solutions to european farmers. *Outlooks on Pest Management*, 31 :152–159.
- [Kulikov et al., 2018] Kulikov, V., Yurchenko, V., and Lempitsky, V. (2018). Instance segmentation by deep coloring.
- [Kumar and Domnic, 2018] Kumar, J. P. and Domnic, S. (2018). Image based plant phenotyping using graph based method and circular hough transform. *Journal of Information Science and Engineering*, 34 :671–686.
- [Kussul et al., 2017] Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14 :778–782.
- [Lamarque et al., 2014] Lamarque, P., Meyfroidt, P., Nettier, B., and Lavorel, S. (2014). How ecosystem services knowledge and values influence farmers’ decision-making. *PLOS ONE*, 9(9) :1–16.
- [Larese et al., 2014] Larese, M. G., Namías, R., Craviotto, R. M., Arango, M. R., Gallo, C., and Granitto, P. M. (2014). Automatic classification of legumes using leaf vein image features. *Pattern Recognit.*, 47 :158–168.
- [Lawler and Wood, 1966] Lawler, E. L. and Wood, D. E. (1966). Branch-and-bound methods : A survey. *Operations research*, 14(4) :699–719.
- [Lechenet et al., 2014] Lechenet, M., Bretagnolle, V., Bockstaller, C., Boissinot, F., Petit, M.-S., Petit, S., and Munier-Jolain, N. M. (2014). Reconciling pesticide reduction with economic and environmental sustainability in arable farming. *PLOS ONE*, 9(6) :1–10.
- [Lee et al., 2021] Lee, K.-d., Ahn, H.-y., Ryu, J.-h., So, K.-h., and Na, S.-i. (2021). Comparison of reflectance and vegetation index changes by type of uav-mounted multi-spectral sensors. *Korean Journal of Remote Sensing*, 37(5_1) :947–958.
- [Lee et al., 2020] Lee, M.-K., Golzarian, M., and Kim, I. (2020). A new color index for vegetation segmentation and classification. *Precision Agriculture*.
- [Lefloch, 2018] Lefloch, C. (2018). Une première approche analytique des services écosystémiques des élevages en race créole bovine, caprine et porcine de guadeloupe. Master’s thesis, Ecole Supérieure d’Agro-Développement International (ISTOM), FRA, FRA.
- [Lemtiri et al., 2016] Lemtiri, A., Degruene, F., Barbieux, S., Hiel, M.-P., Chélin, M., Parvin, N., Vandenbol, M., Francis, F., and Colinet, G. (2016). Crop residue management in arable cropping systems under temperate climate. part 1 : Soil biological and chemical (phosphorus and nitrogen) properties. a review. *Biotechnologie, Agronomie, Société et Environnement*.
- [Léraud and Van Hove, 2019] Léraud, I. and Van Hove, P. (2019). *Algues vertes : L’histoire interdite*. Delcourt.
- [Leutenegger et al., 2011] Leutenegger, S., Chli, M., and Siegwart, R. (2011). Brisk : Binary robust invariant scalable keypoints. In *2011 IEEE international conference on computer vision (ICCV)*, pages 2548–2555. Ieee.
- [Lherbier, 2005] Lherbier, J. (2005). *Valorisation de l’information géographique en agriculture de précision*. PhD thesis.
- [Li et al., 2021] Li, Y., Al-Sarayreh, M., Irie, K., Hackell, D., Bourdot, G., Reis, M. M., and Ghamkhar, K. (2021). Identification of weeds based on hyperspectral imaging and machine learning. *Frontiers in Plant Science*, 11 :2324.

- [Lin et al., 2014] Lin, B., Sun, Y., and Sanchez, J. (2014). Efficient vessel feature detection for endoscopic image analysis. *IEEE transactions on bio-medical engineering*, 62.
- [Liu et al., 2019] Liu, B., Li, R., Li, H., You, G., Yan, S., and Tong, Q. (2019). Crop/weed discrimination using a field imaging spectrometer system. *Sensors*, 19 :5154.
- [Liu et al., 2017] Liu, P., Shi, R., Zhang, C., Zeng, Y., Wang, J., Tao, Z., and Gao, W. (2017). Integrating multiple vegetation indices via an artificial neural network model for estimating the leaf chlorophyll content of spartina alterniflora under interspecies competition. *Environmental Monitoring and Assessment*, 189.
- [Liu et al., 2018a] Liu, R., Lehman, J., Molino, P., Such, F. P., Frank, E., Sergeev, A., and Yosinski, J. (2018a). An intriguing failing of convolutional neural networks and the coordconv solution. *CoRR*, abs/1807.03247.
- [Liu et al., 2015] Liu, W., Rabinovich, A., and Berg, A. C. (2015). Parsenet : Looking wider to see better. *CoRR*, abs/1506.04579.
- [Liu et al., 2018b] Liu, X., Li, Z., Zhong, K., Chao, Y., Miraldo, P., and Shi, Y. (2018b). Generic distortion model for metrology under optical microscopes. *Optics and Lasers in Engineering*, 103 :119–126.
- [Liu et al., 2016] Liu, Z., Li, X., Luo, P., Loy, C. C., and Tang, X. (2016). Deep learning markov random field for semantic segmentation. *CoRR*, abs/1606.07230.
- [Loiret, 2016] Loiret, R. (2016). *Le Bilan écologique. Mesurer la perturbation anthropogénique de l'Ecosphère et de la Biosphère (un bilan de l'Anthropocène). Caractériser les voies du Développement écologique des territoires*. Theses, Université Paris-Saclay.
- [Lombaert et al., 2012] Lombaert, H., Grady, L., Pennec, X., Ayache, N., and Cheriet, F. (2012). Spectral demons – image registration via global spectral correspondence. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *Computer Vision – ECCV 2012*, pages 30–44, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [López-Granados, 2011] López-Granados, F. (2011). Weed detection for site-specific weed management : mapping and real-time approaches. *Weed Research*, 51(1) :1–11.
- [Lottes et al., 2018] Lottes, P., Behley, J., Milioto, A., and Stachniss, C. (2018). Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *CoRR*, abs/1806.03412.
- [Lottes et al., 2016] Lottes, P., Hoferlin, M., Sander, S., Müter, M., Schulze, P., and Stachniss, L. C. (2016). An effective classification system for separating sugar beets and weeds for precision farming applications. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5157–5163.
- [Lottes and Stachniss, 2017] Lottes, P. and Stachniss, C. (2017). Semi-supervised online visual crop and weed classification in precision farming exploiting plant arrangement. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5155–5161.
- [Louargant, 2016] Louargant, M. (2016). *Proximité des adventices par imagerie aérienne : vers un service de gestion par drone*. PhD thesis. Thèse de doctorat dirigée par Gée, ChristelleVillette, SylvainJones, Gawain et Vigneau, Nathalie Instrumentation et informatique de l'image Dijon 2016.
- [Louargant et al., 2018] Louargant, M., Jones, G., Faroux, R., Paoli, J.-N., Maillot, T., Gée, C., and Villette, S. (2018). Unsupervised classification algorithm for early weed detection in row-crops by combining spatial and spectral information. *Remote Sensing*, 10(5).

- [Louargant et al., 2017] Louargant, M., Villette, S., Jones, G., Vigneau, N., Paoli, J., and Gée, C. (2017). Weed detection by uav : simulation of the impact of spectral mixing in multispectral images. *Precision Agriculture*, 18 :1–20.
- [Louime et al., 2017] Louime, C., Fortune, J., and Gervais, G. (2017). Sargassum invasion of coastal environments : a growing concern. *American Journal of Environmental Sciences*, 13(1) :58–64.
- [Lourgant, 2016] Lourgant, M. (2016). *Proximité des adventices par imagerie aérienne : Vers un service de gestion par drone*. PhD thesis, UNIVERSITE DE BOURGOGNE. Agroécologie, AgroSup Dijon, INRA.
- [Lu et al., 2021] Lu, J., Cheng, D., Geng, C., Zhang, Z., Xiang, Y., and Hu, T. (2021). Combining plant height, canopy coverage and vegetation index from uav-based rgb images to estimate leaf nitrogen concentration of summer maize. *Biosystems Engineering*, 202 :42–54.
- [Lukac and Plataniotis, 2005] Lukac, R. and Plataniotis, K. (2005). Color filter arrays : design and performance analysis. *IEEE Transactions on Consumer Electronics*, 51(4) :1260–1267.
- [Lv et al., 2018] Lv, X., Ming, D., Lu, T., Zhou, K., Wang, M., and Bao, H. (2018). A new method for region-based majority voting cnns for very high resolution image classification. *Remote Sensing*, 10(12).
- [Ma et al., 2019] Ma, X., Deng, X., Qi, L., Jiang, Y., Li, H., Wang, Y., and Xing, X. (2019). Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. *PloS one*, 14(4) :e0215676.
- [Machleb et al., 2020] Machleb, J., Peteinatos, G. G., Kollenda, B. L., Andújar, D., and Gerhards, R. (2020). Sensor-based mechanical weed control : Present state and prospects. *Computers and electronics in agriculture*, 176 :105638.
- [Mair et al., 2010] Mair, E., Hager, G. D., Burschka, D., Suppa, M., and Hirzinger, G. (2010). Adaptive and generic corner detection based on the accelerated segment test. In *European conference on Computer vision*, pages 183–196. Springer.
- [Manh et al., 2001] Manh, A.-G., Rabatel, G., Assemat, L., and Aldon, M.-J. (2001). Ae—automation and emerging technologies : Weed leaf image segmentation by deformable templates. *Journal of Agricultural Engineering Research*, 80(2) :139–146.
- [Martin et al., 2017] Martin, P., Pierre-Alain, A., Canovas, I., Cicille, P., F, D., Domergue, J.-M., Douguedroit, A., Grard, N., Kamara, S., Lopez, C., Morardet, S., and Sauvagnargues, S. (2017). Ph-martin-2017-et-al-jistee-1-2017.
- [Martino, 1973] Martino, J. P. (1973). *The limits to growth : A report for the club of rome’s project on the predicament of mankind : Dennis l. meadows et al. universe books, new york. 1972, 205 pages, 6.50; paperback 2.75.*
- [Martos et al., 2021] Martos, V., Ahmad, A., Cartujo, P., and Ordoñez, J. (2021). Ensuring agricultural sustainability through remote sensing in the era of agriculture 5.0. *Applied Sciences*, 11(13).
- [Maudouit and Rochoy, 2019] Maudouit, M. and Rochoy, M. (2019). Systematic review of the impact of chlordecone on human health in the french west indies. *Thérapie*, 74(6) :611–625.
- [McAllister et al., 2021] McAllister, W., Whitman, J., Varghese, J., Davis, A., and Chowdhary, G. (2021). Agbots 3.0 : Adaptive weed growth prediction for mechanical weeding agbots. *IEEE Transactions on Robotics*, pages 1–13.
- [McBratney et al., 2005] McBratney, A., Whelan, B., Ancev, T., and Bouma, J. (2005). Future directions of precision agriculture. *Precision agriculture*, 6(1) :7–23.

- [McGarigal, 1995] McGarigal, K. (1995). *FRAGSTATS : spatial pattern analysis program for quantifying landscape structure*, volume 351. US Department of Agriculture, Forest Service, Pacific Northwest Research Station.
- [McLaughlin and Mineau, 1995] McLaughlin, A. and Mineau, P. (1995). The impact of agricultural practices on biodiversity. *Agriculture, Ecosystems & Environment*, 55(3) :201–212.
- [Meadows et al., 1972] Meadows, D. H., Meadows, D. L., Randers, J., and Behrens, W. W. (1972). The limits to growth. *New York*, 102 :27.
- [Mekhalfa and Yacef, 2021] Mekhalfa, F. and Yacef, F. (2021). Supervised learning for crop/weed classification based on color and texture features. *arXiv preprint arXiv :2106.10581*.
- [Mell, 2018] Mell, P. (2018). Développement de méthodes de phénotypage haut débit basées sur l’analyse d’images pour le suivi de traits d’intérêts sur une collection d’ignames (D. alata) en Guadeloupe. Master’s thesis, INRA, Centre Antilles-Guyanne, Domaine de Duclos, 97127 Petit-Bourg, Guadeloupe.
- [Mercan and Alam, 2011] Mercan, S. and Alam, M. (2011). Anomaly detection in hyperspectral imagery using stable distribution. *Proceedings of SPIE - The International Society for Optical Engineering*, 8049.
- [Mésini, 2006] Mésini, B. (2006). Préservation de la biodiversité et souveraineté alimentaire : réactivation des savoirs paysans et émergence des droits de la terre (2000-2005). *Le retour des paysans ? à l’heure du développement durable*, page 71.
- [Mezera Jiří, 2019] Mezera Jiří, Lukas Vojtěch, E. J. S. V. (2019). Comparison of sentinel-2 and isaria winter wheat mapping for variable rate application of nitrogen fertilizers. *MendelNet*, pages 48–53.
- [Miao et al., 2004] Miao, L., Qi, H., and Snyder, W. E. (2004). A generic method for generating multispectral filter arrays. In *2004 International Conference on Image Processing, 2004. ICIP’04.*, volume 5, pages 3343–3346. IEEE.
- [Milioto et al., 2017] Milioto, A., Lottes, P., and Stachniss, C. (2017). Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns. *CoRR*, abs/1709.06764.
- [Minařík et al., 2019] Minařík, R., Langhammer, J., and Hanuš, J. (2019). Radiometric and atmospheric corrections of multispectral mca camera for uav spectroscopy. *Remote Sensing*, 11(20).
- [Modica and Zoboli, 2016] Modica, M. and Zoboli, R. (2016). Vulnerability, resilience, hazard, risk, damage, and loss : A socio-ecological framework for natural disaster analysis. *Web Ecology*, 16 :59–62.
- [Mondal et al., 2019] Mondal, R., Santra, S., and Chanda, B. (2019). Dense morphological network : An universal function approximator. *CoRR*, abs/1901.00109.
- [Monteiro et al., 2019] Monteiro, A., Von Wangenheim, A., and Júnior, P. (2019). Weed mapping on aerial images - a systematic literature review. Technical Report IN-CoD.LAPIX.01.2019.E, Federal University of Santa Catarina.
- [Moradi et al., 2019] Moradi, S., Ghelich-Oghli, M., Alizadehasl, A., Shiri, I., Oveisi, N., Oveisi, M., Maleki, M., and Dhooge, J. (2019). A novel deep learning based approach for left ventricle segmentation in echocardiography : Mfp-unet.
- [Moré, 1978] Moré, J. J. (1978). The levenberg-marquardt algorithm : Implementation and theory. In Watson, G., editor, *Numerical Analysis*, volume 630 of *Lecture Notes in Mathematics*, pages 105–116. Springer Berlin Heidelberg.

- [Morison et al., 2008] Morison, M., Guichard, L., and Jeuffroy, M.-H. (2008). Comment maîtriser la flore adventice des grandes cultures à travers les éléments de l’itinéraire technique ? *Innovations agronomiques*, 3 :27–41.
- [Morris, 2018] Morris, D. D. (2018). A pyramid CNN for dense-leaves segmentation. volume abs/1804.01646.
- [Mostajer Kheirkhah and Asghari, 2019] Mostajer Kheirkhah, F. and Asghari, H. (2019). Plant leaf classification using gist texture features. *IET Computer Vision*, 13(4) :369–375.
- [Nagaraj, 2008] Nagaraj, K. (2008). *Farmers’ suicides in India : Magnitudes, trends and spatial patterns*. Bharathi puthakalayam.
- [Némery et al., 2016] Némery, J., Gratiot, N., Doan, P., Duvert, C., Alvarado-Villanueva, R., and Duwig, C. (2016). Carbon, nitrogen, phosphorus, and sediment sources and retention in a small eutrophic tropical reservoir. *Aquatic sciences*, 78(1) :171–189.
- [Nguy-Robertson et al., 2012] Nguy-Robertson, A., Gitelson, A., Peng, Y., Viña, A., Arkebauer, T., and Rundquist, D. (2012). Green leaf area index estimation in maize and soybean : Combining vegetation indices to achieve maximal sensitivity. *Agronomy Journal*, 104 :1336–1347.
- [Nicholson et al., 2019] Nicholson, C. F., Stephens, E., Jones, A. D., Kopainsky, B., Parsons, D., and Garrett, J. H. (2019). Setting priorities to address the research gaps between agricultural systems analysis and food security outcomes in low-and middle-income countries.
- [Nimier-David et al., 2019] Nimier-David, M., Vicini, D., Zeltner, T., and Jakob, W. (2019). Mitsuba 2 : A retargetable forward and inverse renderer. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 38(6).
- [Niu et al., 2021] Niu, C., Phinn, S., and Roelfsema, C. (2021). Global sensitivity analysis for canopy reflectance and vegetation indices of mangroves. *Remote Sensing*, 13(13).
- [Nougadère et al., 2015] Nougadère, A., Réty, J., Vial, G., Volatier, J.-L., Leblanc, J.-C., Merlo, M., and Gerault, F. (2015). Surveillance des expositions alimentaires aux résidus de pesticides : des indicateurs de risque pour le suivi et l’orientation des programmes nationaux de surveillance et du plan ecophyto. *Bulletin Épidémiologique, Santé Animale et Alimentation*, 68.
- [Nwankpa et al., 2018] Nwankpa, C., Ijomah, W., Gachagan, A., and Marshall, S. (2018). Activation functions : Comparison of trends in practice and research for deep learning.
- [Oldeland et al., 2010] Oldeland, J., Dorigo, W., Lieckfeld, L., Lucieer, A., and Jürgens, N. (2010). Combining vegetation indices, constrained ordination and fuzzy classification for mapping semi-natural vegetation units from hyperspectral imagery. *Remote Sensing of Environment*, 114(6) :1155–1166.
- [Omari et al., 2020] Omari, M., Lee, J., Faqeerzada, M. . A., Park, E., and Cho, B.-K. (2020). Digital image-based plant phenotyping : a review.
- [Ordonez et al., 2018] Ordonez, A., Arguello, F., and Heras, D. B. (2018). Alignment of hyperspectral images using kaze features. *Remote Sensing*, 10(5).
- [Osco et al., 2021] Osco, L. P., Marcato Junior, J., Marques Ramos, A. P., de Castro Jorge, L. A., Fathollahi, S. N., de Andrade Silva, J., Matsubara, E. T., Pistori, H., Gonçalves, W. N., and Li, J. (2021). A review on deep learning in uav remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 102 :102456.
- [Osorio et al., 2020] Osorio, K., Puerto, A., Pedraza, C., Jamaica, D., and Rodríguez, L. (2020). A deep learning approach for weed detection in lettuce crops using multispectral images. *AgriEngineering*, 2(3) :471–488.

- [Palanisamy et al., 2019] Palanisamy, S., Selvaraj, R., Ramesh, T., and Ponnusamy, J. (2019). Applications of remote sensing in agriculture - a review. *International Journal of Current Microbiology and Applied Sciences*, 8 :2270–2283.
- [Pandey et al., 2017] Pandey, P., Irulappan, V., Bagavathiannan, M. V., and Senthil-Kumar, M. (2017). Impact of combined abiotic and biotic stresses on plant growth and avenues for crop improvement by exploiting physio-morphological traits. *Frontiers in plant science*, 8 :537.
- [Pape and Klukas, 2015] Pape, J.-M. and Klukas, C. (2015). 3-d histogram-based segmentation and leaf detection for rosette plants. In Agapito, L., Bronstein, M. M., and Rother, C., editors, *Computer Vision - ECCV 2014 Workshops*, pages 61–74, Cham. Springer International Publishing.
- [Pászto et al., 2015] Pászto, V., Brychtová, A., and Marek, L. (2015). On shape metrics in cartographic generalization : A case study of the building footprint geometry. In *Modern Trends in Cartography*, pages 397–407. Springer.
- [Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn : Machine learning in python. *Journal of Machine Learning Research*, 12 :2825–2830.
- [Pelosi, 2008] Pelosi, C. (2008). *Modélisation de la dynamique d’une population de vers de terre Lumbricus terrestris au champ. Contribution à l’étude de l’impact des systèmes de culture sur les communautés lombriciennes*. PhD thesis, AgroParisTech.
- [Perez-Ruiz et al., 2021] Perez-Ruiz, M., Martínez-Guanter, J., and Upadhyaya, S. K. (2021). Chapter 15 - high-precision gnss for agricultural operations. In p. Petropoulos, G. and Srivastava, P. K., editors, *GPS and GNSS Technology in Geosciences*, pages 299–335. Elsevier.
- [Perez-Sanz et al., 2017] Perez-Sanz, F., Navarro, P. J., and Egea-Cortines, M. (2017). Plant phenomics : an overview of image acquisition technologies and image data analysis algorithms. *GigaScience*, 6(11). gix092.
- [Perrette, 2018] Perrette, J. (2018). Services écosystémiques de l’élevage en guadeloupe : approche exploratoire. Master’s thesis, Supagro/INRA ; INRA ; Unité de Recherches Zootechniques (URZ).
- [Perrier-Bruslé et al., 2013] Perrier-Bruslé, L. et al. (2013). L’anthropocène et ses victimes. une réflexion terminologique, la géopolitique de l’arctique : sous le signe de la coopération, les régimes internationaux de l’environnement, la tragédie des communaux, «soft power». sens et usages d’un concept incertain, le retour de la revendication maritime bolivienne, la remise en cause des accords de schengen, frontières maritimes dans l’arctique : le droit de la mer est-il un cadre applicable ?, les enjeux politiques autour des frontières maritimes.
- [Perumal and Bhaskaran, 2010] Perumal, K. and Bhaskaran, R. (2010). Supervised classification performance of multispectral images. *CoRR*, abs/1002.4046.
- [Pesce et al., 2021] Pesce, S., Mamy, L., Achard, A. L., Le Gall, M., Le Perchec, S., Rechau-chère, O., Tibi, A., Leenhardt, S., and Sanchez, W. (2021). Scientific literature survey : impacts of pesticides on biodiversity and ecosystem services. *Environmental Science and Pollution Research*, page 7 p.
- [Peteinatos et al., 2020] Peteinatos, G. G., Reichel, P., Karouta, J., Andújar, D., and Gerhards, R. (2020). Weed identification in maize, sunflower, and potatoes with the aid of convolutional neural networks. *Remote Sensing*, 12(24).

- [Peterson, 2009] Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, 4(2) :1883.
- [Peña-Barragán et al., 2011] Peña-Barragán, J. M., Ngugi, M. K., Plant, R. E., and Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, 115(6) :1301–1316.
- [Pic et al., 2022] Pic, J., Cossart, E., and Fressard, M. (2022). Mais où va donc toute cette terre? transferts sédimentaires & structure spatiale des réseaux de drainage sur les versants du beaujolais viticole. In *Théo Quant-Nouvelles approches en Géographie Théorique et Quantitative*.
- [Pic et al., 2020] Pic, J., Le Guen, Y., Cossart, E., and Fressard, M. (2020). La dynamique d'enfrichement, ses expressions paysagères et ses significations : le cas du Beaujolais viticole. *VertigO : La Revue Électronique en Sciences de l'Environnement*.
- [Piedallu et al., 2019] Piedallu, C., Chéret, V., Denux, J., Perez, V., Azcona, J., Seynave, I., and Gégout, J. (2019). Soil and climate differently impact ndvi patterns according to the season and the stand type. *Science of The Total Environment*, 651 :2874–2885.
- [Pierce and Nowak, 1999] Pierce, F. J. and Nowak, P. (1999). Aspects of precision agriculture. *Advances in agronomy*, 67 :1–85.
- [Pretto et al., 2019] Pretto, A., Aravecchia, S., Burgard, W., Chebroly, N., Dornhege, C., Falck, T., Fleckenstein, F., Fontenla, A., Imperoli, M., Khanna, R., Liebisch, F., Lottes, P., Milioto, A., Nardi, D., Nardi, S., Pfeifer, J., Popovic, M., Potena, C., Pradalier, C., and Nieto, J. (2019). Building an aerial-ground robotics system for precision farming.
- [Pujol, 2014] Pujol, J.-L. (2014). Journée mondiale des sols.
- [Pérez et al., 2000] Pérez, A., López, F., Benlloch, J., and Christensen, S. (2000). Colour and shape analysis techniques for weed detection in cereal fields. *Computers and Electronics in Agriculture*, 25(3) :197–212.
- [Qiu et al., 2018] Qiu, R., Wei, S., Zhang, M., Li, H., Sun, H., Liu, G., and Li, M. (2018). Sensors for measuring plant phenotyping : A review. *International Journal of Agricultural and Biological Engineering*, 11(2) :1–17.
- [Qiu and Yuille, 2016] Qiu, W. and Yuille, A. (2016). Unrealcv : Connecting computer vision to unreal engine. In *European Conference on Computer Vision*, pages 909–916. Springer.
- [Quan et al., 2021] Quan, L., Li, H., Li, H., Jiang, W., Lou, Z., and Chen, L. (2021). Two-stream dense feature fusion network based on rgb-d data for the real-time prediction of weed aboveground fresh weight in a field environment. *Remote Sensing*, 13(12).
- [Rabatel and Labbe, 2016] Rabatel, G. and Labbe, S. (2016). Registration of visible and near infrared unmanned aerialvehicle images based on Fourier-Mellin transform. *Precision Agriculture*, 17(5) :564–587.
- [Rahman and Wang, 2016] Rahman, M. and Wang, Y. (2016). Optimizing intersection-over-union in deep neural networks for image segmentation. In *International symposium on visual computing*, volume 10072, pages 234–244. Springer.
- [Rasti et al., 2019] Rasti, P., Ahmad, A., Samiei, S., Belin, E., and Rousseau, D. (2019). Supervised image classification by scattering transform with application to weed detection in culture crops of high density. *Remote Sensing*, 11(3).
- [Rathi and Goyal, 2021] Rathi, V. and Goyal, P. (2021). Convolution filter based efficient multispectral image demosaicking for compact msfas. In *VISIGRAPP (4 : VISAPP)*, pages 112–121.

- [Rechid et al., 2011] Rechid, N., Khelifa, A., Toumi, A., TALEB-AHMED, A., Mimi, M., and Benmahammed, K. (2011). Segmentation non supervisée d’images basée sur les modèles de markov cachés. *Courrier de savoir, No12*, pages 39–34.
- [Reedha et al., 2021] Reedha, R., Dericquebourg, E., Canals, R., and Hafiane, A. (2021). Vision transformers for weeds and crops classification of high resolution uav images. *arXiv e-prints*, pages arXiv–2109.
- [Regnier et al., 2021] Regnier, R., Avrin, G., Barbosa, V., Boffety, D., Kalouguine, A., and Lardy-Fontan, S. (2021). Validation de méthodologies d’évaluation de solutions de désherbage autonomes, dans le cadre des projets Challenge ROSE et METRICS. *Revue Ouverte d’Intelligence Artificielle*, 2(1) :11–32.
- [Reulier, 2015] Reulier, R. (2015). *Structure paysagère et dynamiques spatiales des transferts hydro-sédimentaires. Approche par simulation multi-agents*. Theses, Université de Caen Normandie.
- [Rhouma et al., 2017] Rhouma, M. B. H., Žunić, J., and Younis, M. C. (2017). Moment invariants for multi-component shapes with applications to leaf classification. *Computers and electronics in agriculture*, 142 :326–337.
- [Robert, 2002] Robert, P. C. (2002). Precision agriculture : a challenge for crop nutrition management. In *Progress in Plant Nutrition : Plenary Lectures of the XIV International Plant Nutrition Colloquium*, pages 143–149. Springer.
- [Roberts et al., 1982] Roberts, H., Chancellor, R., and Hill, T. (1982). Weed control handbook : Principles, ed. by ha roberts.
- [Rojas et al., 2017] Rojas, C., Guzmán, L., and Velasco Toledo, N. (2017). Weed recognition by svm texture feature classification in outdoor vegetable crop images. *Ingeniería e Investigación*, 37 :68.
- [Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6) :386.
- [R.S and Palanisamy, 2010] R.S, D. and Palanisamy, V. (2010). Crop and weed discrimination in agricultural field using mrcsf. *International Journal of Signal and Imaging Systems Engineering*, 3.
- [Ruble et al., 2011] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb : An efficient alternative to sift or surf. In *Proceedings of the 2011 International Conference on Computer Vision, ICCV ’11*, pages 2564–2571, Washington, DC, USA. IEEE Computer Society.
- [Sa et al., 2017] Sa, I., Chen, Z., Popovic, M., Khanna, R., Liebisch, F., Nieto, J., and Siegart, R. (2017). Weednet : Dense semantic weed classification using multispectral images and mav for smart farming. *IEEE Robotics and Automation Letters*, PP.
- [Saddik et al., 2022] Saddik, A., Latif, R., El Ouardi, A., Elhoseny, M., and Khelifi, A. (2022). Computer development based embedded systems in precision agriculture : tools and application. *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, pages 1–23.
- [Sage and Unser, 2003] Sage, D. and Unser, M. (2003). Teaching image-processing programming in java. *IEEE Signal Processing Magazine*, 20(6) :43–52. Using “Student-Friendly” ImageJ as a Pedagogical Tool.
- [Saha et al., 2017] Saha, D., Hamer, G., and Lee, J. Y. (2017). Development of inter-leaves weed and plant regions identification algorithm using histogram of oriented gradient and k-means clustering. In *Proceedings of the International Conference on Research in Adaptive and Convergent Systems*, pages 136–139.

- [Salisu Abubakar et al., 2021] Salisu Abubakar, B., Ahmad, A., Uba, H., Yakubu, M., and Ahmad, A. (2021). Systematic literature review and meta-analysis weed detection using machine learning : A systematic literature review. *Systematic Literature Review and Meta-Analysis Journal*, 2.
- [Salzenstein and Collet, 2006] Salzenstein, F. and Collet, C. (2006). Fuzzy markov random fields versus chains for multispectral image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 :1753–1767.
- [Sankaran et al., 2015] Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., Miklas, P. N., Carter, A. H., Pumphrey, M. O., Knowles, N. R., and Pavek, M. J. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping : A review. *European Journal of Agronomy*, 70 :112 – 123.
- [Scharr et al., 2016] Scharr, H., Minervini, M., French, A. P., Klukas, C., Kramer, D. M., Liu, X., Luengo, I., Pape, J.-M., Polder, G., Vukadinovic, D., et al. (2016). Leaf segmentation in plant phenotyping : a collation study. *Machine vision and applications*, 27(4) :585–606.
- [Scharr et al., 2017] Scharr, H., Pridmore, T., and Tsaftaris, S. (2017). Computer vision problems in plant phenotyping, cvppp 2017 : Introduction to the cvppp 2017 workshop papers. pages 2020–2021.
- [Schaub, 2010] Schaub, C. (2010). Mieux connaître les mauvaises herbes pour mieux maîtriser le désherbage. *AgriMieux & Chambre d’Agriculture du Bas-Rhin*.
- [Seitz, 2010] Seitz, H. (2010). Contributions to the minimum linear arrangement problem.
- [Shamir et al., 2019] Shamir, R. R., Duchin, Y., Kim, J., Sapiro, G., and Harel, N. (2019). Continuous dice coefficient : a method for evaluating probabilistic segmentations. *CoRR*, abs/1906.11031.
- [Shi et al., 1994] Shi, J. et al. (1994). Good features to track. In *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, pages 593–600. IEEE.
- [Shishir and Tsuyuzaki, 2018] Shishir, S. and Tsuyuzaki, S. (2018). Hierarchical classification of land use types using multiple vegetation indices to measure the effects of urbanization. *Environmental Monitoring and Assessment*, 190.
- [Shiva, 2005] Shiva, V. (2005). *Globalization’s New Wars : Seed, Water & Life Forms*, volume 3. Women Unlimited.
- [Shiva and Jalees, 2006] Shiva, V. and Jalees, K. (2006). *Seeds of suicide : the ecological and human costs of seed monopolies and globalisation of agriculture*. Navdanya.
- [Singh et al., 2022] Singh, N., Tewari, V., Biswas, P., Dhruw, L., Pareek, C., and Singh, H. D. (2022). Semantic segmentation of in-field cotton bolls from the sky using deep convolutional neural networks. *Smart Agricultural Technology*, 2 :100045.
- [Storkey et al., 2021] Storkey, J., Helps, J., Hull, R., Milne, A. E., and Metcalfe, H. (2021). Defining integrated weed management : A novel conceptual framework for models. *Agronomy*, 11(4).
- [Su, 2020] Su, W.-H. (2020). Advanced machine learning in point spectroscopy, rgb- and hyperspectral-imaging for automatic discriminations of crops and weeds : A review. *Smart Cities*, 3(3) :767–792.
- [Sun et al., 2018] Sun, H., Hou, M., Yang, Y., Zhang, T., Weng, F., and Han, F. (2018). Solving partial differential equation based on bernstein neural network and extreme learning machine algorithm. *Neural Processing Letters*, pages 1–20.

- [Suta et al., 2012] Suta, L., Bessy, F., Veja, C., and Vaida, M. (2012). Active contours : Application to plant recognition. pages 181–187.
- [Tan et al., 2014] Tan, X., Lai, S., Liu, Y., and Zhang, M. (2014). Green channel guiding denoising on bayer image. *The Scientific World Journal*, 2014.
- [Tang et al., 2021] Tang, Y., Dananjayan, S., Hou, C., Guo, Q., Luo, S., and He, Y. (2021). A survey on the 5g network and its impact on agriculture : Challenges and opportunities. *Computers and Electronics in Agriculture*, 180 :105895.
- [Tanrıverdi et al., 2020] Tanrıverdi, C., Fakültesi, Z., Yapılar, T., Bölümü, S., Kahramanmaraş, Tarımda, H., Algılama, U., İndekslerinin, B., and Derlemesi, B. (2020). A review of remote sensing and vegetation indices in precision farming.
- [Tareen and Saleem, 2018] Tareen, S. A. K. and Saleem, Z. (2018). A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pages 1–10.
- [Taylor and Whelan, 2005] Taylor, J. and Whelan, B. (2005). A general introduction to precision agriculture. *Australian Center for Precision Agriculture*.
- [Tharwat, 2020] Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*.
- [Thyagarajan and Raji, 2019] Thyagarajan, K. and Raji, I. K. (2019). A review of visual descriptors and classification techniques used in leaf species identification. *Archives of Computational Methods in Engineering*, 26(4) :933–960.
- [Tian et al., 2020] Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W., and Lin, C.-W. (2020). Deep learning on image denoising : An overview. *Neural Networks*.
- [Tournebize, 2020] Tournebize, J. (2020). *L'ingénierie écologique au service de la qualité de l'eau*. Habilitation à diriger des recherches, Sorbonne Université.
- [Trajković and Hedley, 1998] Trajković, M. and Hedley, M. (1998). Fast corner detection. *Image and vision computing*, 16(2) :75–87.
- [Tripicchio et al., 2020] Tripicchio, P., Camacho-Gonzalez, G., and D'Avella, S. (2020). Welding defect detection : Coping with artifacts in the production line. *The International Journal of Advanced Manufacturing Technology*, 111(5) :1659–1669.
- [Tsolakidis et al., 2014] Tsolakidis, D. G., Kosmopoulos, D. I., and Papadourakis, G. (2014). Plant leaf recognition using zernike moments and histogram of oriented gradients. In *Hellenic conference on artificial intelligence*, pages 406–417. Springer.
- [Uchiyama et al., 2017] Uchiyama, H., Sakurai, S., Mishima, M., Arita, D., Okayasu, T., Shimada, A., and Taniguchi, R.-i. (2017). An easy-to-setup 3d phenotyping platform for komatsuna dataset. pages 2038–2045.
- [Utyasheva et al., 2021] Utyasheva, L., Sharma, D., Ghimire, R., Karunarathne, A., Robertson, G., and Eddleston, M. (2021). Suicide by pesticide ingestion in nepal and the impact of pesticide regulation. *BMC public health*, 21(1) :1–11.
- [Vakalopoulou and Karantzalos, 2014] Vakalopoulou, M. and Karantzalos, K. (2014). Automatic descriptor-based co-registration of frame hyperspectral data. *Remote Sensing*, 6.
- [van Beers et al., 2019] van Beers, F., Lindström, A., Okafor, E., and Wiering, M. A. (2019). Deep neural networks with intersection over union loss for binary image segmentation. In *ICPRAM*, pages 438–445.

- [Van Der Werf, 1997] Van Der Werf, H. M. (1997). Evaluer l'impact des pesticides sur l'environnement. *Le Courrier de l'environnement de l'INRA*, 31(31) :5–22.
- [van Mourik et al., 2021] van Mourik, S., van der Tol, R., Linker, R., Reyes-Lastiri, D., Kootstra, G., Koerkamp, P. G., and van Henten, E. J. (2021). Introductory overview : Systems and control methods for operational management support in agricultural production systems. *Environmental Modelling & Software*, 139 :105031.
- [Vayssade, 2018] Vayssade, J.-A. (2018). Spatial tracking of small ruminants on pasture.
- [Vayssade et al., 2019] Vayssade, J.-A., Arquet, R., and Bonneau, M. (2019). Automatic activity tracking of goats using drone camera. *Computers and Electronics in Agriculture*, 162 :767–772.
- [Vayssade et al., 2022] Vayssade, J.-A., Jones, G., Gée, C., and Paoli, J.-N. (2022). Pixelwise instance segmentation of leaves in dense foliage. *Computers and Electronics in Agriculture*, 195 :106797.
- [Vayssade et al., 2020] Vayssade, J.-A., JONES, G., Paoli, J.-N., and Gée, C. (2020). Two-step multi-spectral registration via key-point detector and gradient similarity. Application to agronomic scenes for proxy-sensing. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, La Valette, Malta.
- [Vayssade et al., 2021] Vayssade, J.-A., Paoli, J.-N., Gée, C., and JONES, G. (2021). DeepIndices : Remote Sensing Indices Based on Approximation of Functions through Deep-Learning, Application to Uncalibrated Vegetation Images. *Remote Sensing*, 13(2261) :1–21.
- [Vioix, 2004] Vioix, J.-B. (2004). *Conception et réalisation d'un dispositif d'imagerie multispectrale embarqué : du capteur aux traitements pour la détection d'adventices*. PhD thesis. Thèse de doctorat dirigée par Truchetet, Frédéric et Douzals, Jean-Paul Informatique et instrumentation de l'image Dijon 2004.
- [Vitinnov et al., 2017] Vitinnov, V. et al. (2017). Journées de l'agroécologie au vignoble.
- [Vulgarakis Minov, 2015] Vulgarakis Minov, S. (2015). *Integration of imaging techniques for the quantitative characterization of pesticide sprays*. PhD thesis. Thèse de doctorat dirigée par Paindavoine, MichelPieters, JanCointault, Frédéric et Nuyttens, David Instrumentation et informatique de l'image Dijon 2015.
- [Waghmare et al., 2016] Waghmare, H., Kokare, R., and Dandawate, Y. (2016). Detection and classification of diseases of grape plant using opposite colour local binary pattern feature and machine learning for automated decision support system. In *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 513–518.
- [Wang et al., 2019a] Wang, A., Zhang, W., and Wei, X. (2019a). A review on weed detection using ground-based machine vision and image processing techniques. *Computers and Electronics in Agriculture*, 158 :226–240.
- [Wang et al., 2022] Wang, D., Cao, W., Zhang, F., Li, Z., Xu, S., and Wu, X. (2022). A review of deep learning in multiscale agricultural sensing. *Remote Sensing*, 14(3).
- [Wang et al., 2019b] Wang, G., Wang, J., Zou, X., Chai, G., Wu, M., and Wang, Z. (2019b). Estimating the fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil from modis data : Assessing the applicability of the ndvi-dfi model in the typical xilingol grasslands. *International Journal of Applied Earth Observation and Geoinformation*, 76 :154–166.

- [Wang et al., 2008] Wang, J., Shi, F., Zhang, J., and Liu, Y. (2008). A new calibration model of camera lens distortion. *Pattern Recogn.*, 41(2) :607–615.
- [Wang et al., 2018a] Wang, X., Singh, D., Marla, S., Morris, G., and Poland, J. (2018a). Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. *Plant Methods*, 14 :53.
- [Wang et al., 2018b] Wang, Z., Wang, K., Yang, F., Pan, S., and Han, Y. (2018b). Image segmentation of overlapping leaves based on chan–vese model and sobel operator. *Information Processing in Agriculture*, 5(1) :1 – 10.
- [Ward and Moghadam, 2020] Ward, D. and Moghadam, P. (2020). Scalable learning for bridging the species gap in image-based plant phenotyping. *Computer Vision and Image Understanding*, 197-198 :103009.
- [Ward et al., 2018] Ward, D., Moghadam, P., and Hudson, N. (2018). Deep leaf segmentation using synthetic data. In *BMVC*.
- [Watcharabutsarakham et al., 2012] Watcharabutsarakham, S., Sinthupinyo, W., and Kiratiratanapruk, K. (2012). Leaf classification using structure features and support vector machines. In *2012 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM2012)*, pages 697–700. IEEE.
- [Wesson and Wareing, 1967] Wesson, G. and Wareing, P. (1967). Light requirements of buried seeds. *Nature*, 213(5076) :600–601.
- [Wezel et al., 2014] Wezel, A., Casagrande, M., Celette, F., jean françois, V., Ferrer, A., and Peigné, J. (2014). Agroecological practices for sustainable agriculture. a review. *Agronomy for Sustainable Development*, 34 :1–20.
- [Wintermantel et al., 2020] Wintermantel, D., Odoux, J.-F., Decourtye, A., Henry, M., Allier, F., and Bretagnolle, V. (2020). Neonicotinoid-induced mortality risk for bees foraging on oilseed rape nectar persists despite eu moratorium. *Science of The Total Environment*, 704 :135400.
- [Wong et al., 2013] Wong, W., Ali, C., Brandon, K., Wee, C. C., and Marriapan, M. (2013). Co-occurrence matrix with neural network classifier for weed species classification : A comparison between direct application of co-occurrence matrix (glcm) and haralick features as inputs. *International Journal of Enhanced Research in Science and Engineering*, 2(2) :1–6.
- [Wood et al., 2020] Wood, T. J., Michez, D., Paxton, R. J., Drossart, M., Neumann, P., Gerard, M., Vanderplanck, M., Barraud, A., Martinet, B., Leclercq, N., et al. (2020). Managed honey bees as a radar for wild bee decline? *Apidologie*, 51(6) :1100–1116.
- [Wu et al., 2019] Wu, X., Aravecchia, S., and Pradalier, C. (2019). Design and implementation of computer vision based in-row weeding system. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4218–4224.
- [Wu et al., 2020] Wu, Y., Chen, L., and Merhof, D. (2020). Improving pixel embedding learning through intermediate distance regression supervision for instance segmentation. In *European Conference on Computer Vision*, pages 213–227. Springer.
- [Wu et al., 2021a] Wu, Z., Chen, Y., Zhao, B., Kang, X., and Ding, Y. (2021a). Review of weed detection methods based on computer vision. *Sensors*, 21(11).
- [Wu et al., 2021b] Wu, Z., Chen, Y., Zhao, B., Kang, X., and Ding, Y. (2021b). Review of weed detection methods based on computer vision. *Sensors*, 21(11).
- [Xie et al., 2020] Xie, X., Chen, J., Li, Y., Shen, L., Ma, K., and Zheng, Y. (2020). Instance-aware self-supervised learning for nuclei segmentation.

- [Xu et al., 2020] Xu, Y., He, R., Gao, Z., Li, C., Zhai, Y., and Jiao, Y. (2020). Weed density detection method based on absolute feature corner points in field. *Agronomy*, 10(1).
- [Xue et al., 2017] Xue, Jinru, S., and Baofeng (2017). Significant Remote Sensing Vegetation Indices : A Review of Developments and Applications. *Journal of Sensors*, 10(1155) :1–1.
- [Yahui et al., 2019] Yahui, G., Senthilnath, J., Wu, W., Zhang, X., Zeng, Z., and Huang, H. (2019). Radiometric calibration for multispectral camera of different imaging conditions mounted on a uav platform. *Sustainability*, 11 :1–24.
- [You et al., 2020] You, J., Liu, W., and Lee, J. (2020). A dnn-based semantic segmentation for detecting weed and crop. *Computers and Electronics in Agriculture*, 178 :105750.
- [Yousefi et al., 2016] Yousefi, S., Pourghasemi, H. R., Hooke, J., Navratil, O., and Kidová, A. (2016). Changes in morphometric meander parameters identified on the karoon river, iran, using remote sensing data. *Geomorphology*, 271 :55–64.
- [Yuan et al., 2019] Yuan, T., Borba, W., Obraczka, K., Barakat, C., Turletti, T., Esteve Rotherberg, C., and da Rocha Neto, W. (2019). Harnessing machine learning for next-generation intelligent transportation systems : A survey.
- [Yue et al., 2016] Yue, H., Cai, K., Lin, H., Man, H., and Zeng, Z. (2016). A markov random field model for image segmentation of rice planthopper in rice fields. *Journal of Engineering Science and Technology Review*, 9 :31–38.
- [Zamir et al., 2020] Zamir, S. W., Arora, A., Khan, S. H., Hayat, M., Khan, F. S., Yang, M., and Shao, L. (2020). Learning enriched features for real image restoration and enhancement. volume abs/2003.06792.
- [Zerbib, 2018] Zerbib, J. (2018). *Relations trophiques ente la plante cultivée et les champignons mycorhiziens à arbuscules. Importance des champignons mycorhiziens à arbuscules sur la production végétale (en termes de quantité et qualité)*. PhD thesis, Bourgogne Franche-Comté. Thèse de doctorat dirigée par Wipf, Daniel Physiologie Bourgogne Franche-Comté 2018.
- [Zhai et al., 2018] Zhai, H., Zhang, H., Zhang, L., and Li, P. (2018). Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144 :235 – 253.
- [Zhang et al., 2017] Zhang, H., Lan, Y., Lacey, R. E., Hoffmann, W. C., Huang, Y., Panda, S. S., Ames, D. P., Panigrahi, S., Schweizer, S., Moura, J. M. F., Miglani, A., Roy, S. S., Min, Y., and Hua, M. (2017). Image processing techniques for vegetation study - retrospective.
- [Zhang et al., 2016] Zhang, H., Wohlfeil, J., and Griebßbach, D. (2016). Extension and evaluation of the agast feature detector.
- [Zhang et al., 2015] Zhang, L., Sun, X., Wu, T., and Zhang, H. (2015). An analysis of shadow effects on spectral vegetation indexes using a ground-based imaging spectrometer. *IEEE Geoscience and Remote Sensing Letters*, 12 :1–5.
- [Zhang et al., 2019a] Zhang, M. R., Lucas, J., Hinton, G. E., and Ba, J. (2019a). Lookahead optimizer : k steps forward, 1 step back. *CoRR*, abs/1907.08610.
- [Zhang et al., 2019b] Zhang, S., Chen, H., Fu, Y., Niu, H., Yang, Y., and Zhang, B. (2019b). Fractional vegetation cover estimation of different vegetation types in the qaidam basin. *Sustainability*, 11(3).
- [Zhang et al., 2020] Zhang, X., Barron, J. T., Tsai, Y.-T., Pandey, R., Zhang, X., Ng, R., and Jacobs, D. E. (2020). Portrait shadow manipulation. volume 39.

- [Zhao et al., 2017] Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network.
- [Zheng et al., 2019] Zheng, Y.-Y., Kong, J.-L., Jin, X.-B., Wang, X.-Y., Su, T.-L., and Zuo, M. (2019). Cropdeep : The crop vision dataset for deep-learning-based classification and detection in precision agriculture. *Sensors*, 19(5).
- [Zhonghua et al., 2020] Zhonghua, M., Xiaoyao, Y., Meihong, X., Chuangxin, H., Nan, L., and Teng, S. (2020). Automatic weed detection method based on fusion of multiple image processing algorithms. *Smart Agriculture*, 2(4) :103.
- [Zhou et al., 2019] Zhou, D., Fang, J., Song, X., Guan, C., Yin, J., Dai, Y., and Yang, R. (2019). Iou loss for 2d/3d object detection.
- [Zhou et al., 2014] Zhou, R., Kaneko, S., Tanaka, F., Kayamori, M., and Shimizu, M. (2014). Disease detection of cercospora leaf spot in sugar beet by robust template matching. *Computers and Electronics in Agriculture*, 108 :58–70.
- [Zhu, 2006] Zhu, X. (2006). Semi-supervised learning literature survey.
- [Zitová and Flusser, 2003] Zitová, B. and Flusser, J. (2003). Image registration methods : A survey. *Image and Vision Computing*, 21 :977–1000.
- [Zuiderveld, 1994] Zuiderveld, K. (1994). Graphics gems iv. chapter Contrast Limited Adaptive Histogram Equalization, pages 474–485. Academic Press Professional, Inc., San Diego, CA, USA.
- [Zwiggelaar, 1998] Zwiggelaar, R. (1998). A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop protection*, 17(3) :189–206.