Prior-constrained Convolutional Neural Networks for Medical Image Segmentation

Rosana El Jurdi

Normandie Université

THÈSE

Pour obtenir le grade de Docteur de Normandie Université et le grade de Docteur de l'Université Libanaise

Spécialité Informatique

École Doctorale Mathématiques, Information, Ingénierie des Systèmes, France
École Doctorale en Sciences et Technologie, Liban

Prior-constrained Convolutional Neural Networks for Medical Image Segmentation

Apprentissage profond avec contraintes a priori pour la segmentation d'images médicales

Présentée et soutenue par

Rosana EL JURDI

Dirigée par Caroline PETITJEAN, Paul HONEINE et Fahed ABDALLAH

Thèse soutenue publiquement le 7 Octobre 2021 à Rouen Normandie devant le jury composé de

Nicolas Thome	Professeur, CNAM Paris, France	Rapporteur
Diana Mateus	Professeur, Ecole Centrale Nantes, France	Rapporteure
Catherine Achard	Professeur, Sorbonne Université, France	Examinatrice
Veronika Cheplygina	Professeur associé, ITU, Denmark	Examinatrice
Clovis Francis	Professeur, Université Libanaise, Liban	Examinateur
Paul Honeine	Professeur, Université de Rouen Normandie, France	Co-directeur
Caroline Petitjean	Professeur associé, Université de Rouen Normandie, France	Co-directrice
Fahed Abdallah	Professeur, Université Libanaise, Liban	Co-directeur

litis

UNIVERSITÉ DE ROUEN NORMANDIE

MIIS

Abstract

Today, deep convolutional neural networks (CNNs) have demonstrated state-of-the-art performance for medical image segmentation, on various imaging modalities and tasks. Despite early success, segmentation networks may still generate anatomically aberrant segmentations, with holes or inaccuracies near the object boundaries. Moreover, they often require large amounts of labeled training data, which is not easily available within the medical field. To mitigate these limitations, recent research studies have focused on incorporating prior knowledge, such as object shapes or boundary or location, as constraints within the deep learning framework, in order to enforce anatomical plausibility. Constraints via prior knowledge can be incorporated in CNNs either at the level of the network architecture or at the level of the loss function. Whereas structural constraints are rather robust, loss constraints are more generic and can be plugged into any backbone network.

The objective of our research is to investigate and propose new methods to constrain CNNs in order to segment anatomical objects in medical imaging. Our contributions are threefold: 1) We propose a survey on prior-based losses in medical image segmentation, in order to identify limitations of current methods and novel research directions. In addition, we conduct a benchmark where we investigate the performance of some state-of-the-art prior-based losses across segmentation tasks and shed light on the underlying relationship between the prior losses relative to the dataset characteristics. 2) We propose novel architectural constraints with BB-UNet (Bounding Box U-Net), which is a novel U-Net variant that integrates location as well as shape prior through a convolutional layer, introduced at the level of skip connections. We also investigate the CoordConv component, a spatial coordinate conditioned extension of convolutional layers for medical image segmentation. 3) We propose a constraint at the level of the loss function. The objective of this novel prior-based loss is to regularize organ perimeters in order to take into account border and shape irregularities. We test the proposed methods across a variety of datasets of different tasks and modalities including organs, lesion and tumor segmentation via medical challenges such as SegTHOR, Decathlon, ISLES, to name a few.

Résumé

Aujourd'hui, les réseaux de neurones convolutifs profonds (CNN) ont montré de très bonnes performances pour la segmentation des images médicales issues de diverses modalités. Malgré ces premiers succès, les réseaux de segmentation peuvent encore générer des segmentations anatomiquement aberrantes, avec des trous ou des imprécisions près des frontières des objets à segmenter. De plus, ils nécessitent souvent de grandes quantités de données étiquetées pour l'entraînement des CNN, données qui ne sont pas facilement disponibles dans le domaine médical. Pour remédier à ces limitations, des recherches récentes se sont concentrées sur l'incorporation de connaissances a priori, telles que la forme, les caractéristiques ou l'emplacement des objets, utilisées comme contraintes, dans le cadre des CNN, afin de renforcer la plausibilité anatomique. Les contraintes via les connaissances a priori peuvent être incorporées dans les CNNs soit au niveau de l'architecture du réseau, soit au niveau de la fonction de perte. Alors que les contraintes structurelles sont plutôt robustes, les contraintes de perte sont plus génériques et peuvent être intégrées à n'importe quelle architecture.

L'objectif de notre thèse est d'étudier et de proposer de nouvelles méthodes pour contraindre les, CNN afin de segmenter des organes et/ou des lésions en imagerie médicale. Nos contributions sont de trois ordres : 1) Nous avons proposé un état de l'art des a priori intégrés au fonctions de coût dans les CNN de segmentation, afin d'identifier les limites des méthodes actuelles et les nouvelles directions de recherche. En outre, nous avons effectué un benchmarkn dans lequel les performances d'une sélection de fonctions de pertes basées sur des a priori sont étudiés et mises en rapport avec les caractéristiques des jeux de données. 2) Nous avons proposé de nouvelles contraintes architecturales avec BB-UNet (Bounding Box U-Net), qui est une variante de U-Net qui intègre l'emplacement de l'objet à segmenter, ainsi qu'un a priori sur sa forme, par le biais d'une couche convolutive introduite au niveau des skip connections. Nous étudions également le l'apport des couches de type CoordConv, qui sont des convolutions intégrant les coordonnées spatiales des pixels, pour la segmentation des images médicales. 3) Enfin, nous avons proposé une contrainte au niveau de la fonction de pet. L'objectif de cette nouvelle perte basée sur les antériorités est de régulariser les périmètres des organes afin de prendre en compte les irrégularités de bordure et de forme. Nous testons les méthodes proposées sur une variété de jeux de données de différentes tâches et modalités, y compris la segmentation d'organes, de lésions et de tumeurs via des défis médicaux tels que SegTHOR, Decathlon, ISLES, pour n'en citer que quelques-uns.

Acknowledgements

First and foremost, I would like to thank my directors, Prof. Paul Honeine, Prof. Caroline Petitjean, and Prof. Fahed Abdallah for their continuous guidance and availability throughout my PhD journey. Their insightful suggestions, constructive criticism, and unwavering support have allowed the advance of my research till completion, and the expansion of my scientific boundaries and perspectives.

I would like to thank the Agence Universitaire de la Francophonie (AUF), the Lebanese University and the CNRS-Lebanon for funding part of the thesis, and for the European Regional Development Fund (ERDF) and the Normandy Region (DAISI project) for completing this fund. In addition, I would like to thank the CRIANN for providing computational resources used to conduct most of the experiments in this thesis.

I would like to thank the reviewers and examinators, Prof. Diana Mateus, Prof. Nicolas Thome, Prof. Catherine Achard, Prof. Veronika Cheplygina, and Prof. Clovis Francis, for taking the time to review my work. The provided feedback not only enriched my confidence in these thesis contributions, but has also widened my perspective towards the infinite possibilities the proposed ideas could develop through.

Much work during the thesis has been in collaboration with Dr. Veronika Cheplygina who I'd like to thank for her warm welcome during my research visit at the Eindhoven University of Technology. Veronika's unique vision of machine learning and her insistence on success despite hardships made her an inspiring research idol that drove me forward in my research journey.

My heartfelt thanks go to my fellow lab mates and friends, namely Cyprien, Suman, Rachel, Haydar, Achille and Guillaume for their continuous support, lively discussions and for teaching me how to play Tarot. A special thanks goes to Cyprien and Suman for their technical support, stimulating discussions, and the heartfelt distractions that rested my mind away from my research.

I would like to thank my family, my father (Afif), my mother (Noha), my sisters, my niece and nephews (Amin , Tamer and Ranim) and my dearest lifelong friend Dalia. To all, I would like to say that you were the main building blocks shaping who I am today. Thank you for your wise counsel, sympathetic ears and unconditional love. In this all, I am most grateful for my mother who has been my backbone and main motivator. Her ever present support and belief in my own capability for success has been a major force that kept me moving forward despite challenging moments in my life. I can honestly say I am where I am because of her.

My most sincere gratitude goes to Prof. Paul Honeine who has not only been my supervisor, but also a friend, a mentor, and a person I consider family. Throughout my entire PhD, Paul's door was always open whenever I was facing research blocks or stressful ad-

ministrative struggles. Paul's ever present support has given me strength and power and has taught me to be resilient and to believe in my abilities in the face of uncertainty. My interactions with him have profoundly shaped not only how I do research but also who I am as a person and how I interact with life problems and hardships. I can honestly say that his support was one of the main engines that drove me to carry on this thesis to the very end.

My gratitude extends to Caroline (Prof. Caroline Petitjean), Laurent (Prof. Laurent Haute) and Samia (Prof. Samia Ainouz) for the warm welcome they have provided upon my arrival in France. I consider myself lucky to have met such amazing and supportive colleagues and friends who I was able to share my thesis uncertainties and insecurities with ultimate impressionability. You have always believed in my ability to succeed and have been major milestones that allowed me to overcome many barriers in my thesis journey. Aside from work, you have also embraced me with a sense of home away from my home and to that I am very thankful.

My final thanks go to God who has been the ultimate engineer in my life's success. Even when I walk through the darkest valleys of life, I will not be afraid, for you are close beside me. Thank you for the strength and the blessings you have bestowed upon me!

Rosana El Jurdi

Contents

List of Figures

List of Tables

General Introduction

" To every problem there is a solution. If there is no solution, then there is no problem ! "

Medical image segmentation is the process of making per-pixel predictions in an image in order to identify organs or lesions from the background. Generally, medical images are largely versatile in nature, depending on the acquisition process and the type of object to be segmented. Imaging modalities include magnetic resonance imaging (MRI), computed tomography (CT), nuclear medicine functional imaging, ultrasound imaging, microscopy, to name a few. Hence, they vary in characteristics and nature and are broad with regards to the anatomical object of interest. As such, guaranteeing high performance for medical image segmentation can be considered very challenging when compared to other types of images or segmentation tasks. Regardless, segmentation in the medical domain is considered a key step in assisting early disease detection, diagnosis, monitoring treatment and follow up.

In the recent era, deep learning has registered a pivotal milestone in many fields including pattern recognition, object detection, natural language processing, with medical image segmentation being no exception to the rule. Convolutional neural networks (CNNs), a class of deep learning models, have been known to achieve considerable results due to their generalization ability. Since the segmentation process involves indicating not only what is present in an image but also where, medical image segmentation via CNNs considers a trade-off between contextual and spatial understanding. A pioneering approach for segmentation is the U-Net model, which is known for the ability to consider semantic and contextual information while achieving promising performance. U-Net has gained a high-level of success within image segmentation generally, and medical image segmentation particularly, due to its enhanced properties and powerful predictive notions.

Despite undeniable success, segmentation networks for medical images, including U-Net and its variants, may still generate anatomically aberrant segmentations, with holes or inaccuracies near the object boundaries. Thus, such models lack the anatomical plausibility and background that a medical expert has. Moreover, they often require large amounts of annotated training data, which is not easy to obtain in the medical domain. Unannotated or partially labeled data are, rather, more easily available or less computationally expensive.

To mitigate these limitations, recent research studies have focused on incorporating

medical expert information, known as prior knowledge, as constraints within the deep learning framework. Prior knowledge can be information concerning the object shapes, size, topology, boundary or location, and has been known to be useful via variational approaches prior to the deep learning era. The exploitation of prior knowledge allows enforcing anatomical plausibility within segmentations provided by deep networks and can also overcome the need for fully labeled data.

The objective of this thesis is to propose and investigate new methods to constrain CNNs, via prior knowledge, in order to increase the anatomical correctness of automated models, led by U-Net, on medical images. Constraints via prior knowledge can be incorporated in CNNs either at the level of network architecture or at the level of loss function. Whereas structural constraints are rather robust, loss constraints are more generic and can be plugged into any backbone network.

The contributions conducted within this research can be divided into three folds displayed over five chapters. First, a survey that summarizes state-of-the art contributions within the field is proposed, in order to quantify limitations of current methods and novel research directions. Moreover, a benchmark of some state-of-the-art prior-based losses for segmentation tasks is conducted, in order to investigate the underlying relationship between loss performance and dataset characteristics. Secondly, novel architectural constraints are proposed with BB-UNet and CoordConv-Unet. BB-UNet (i.e., Bounding Box U-Net) is a deep learning model that integrates bounding box and shape prior through a novel convolutional layer introduced at the level of skip connections. On the other hand, CoordConv-Unet is a U-Net variant introduced in order to investigate location prior via coordinate conditioned convolutions known as CoordConv components. Finally, a novel loss function is designed in order to regularize organ perimeters and take into account border and shape irregularities. The proposed methods and contributions are validated on a variety of public datasets of different tasks and modalities, including organs, lesion and tumor segmentation via medical challenges in the field. In the following, we present a brief overview of the manuscript chapter materials.

Chapter 1: Fundamentals of Medical Image Segmentation in Deep Learning

This chapter introduces the building blocks of medical image segmentation via deep learning and identifies the different approaches to integrating prior on the deep learning framework. It briefly reviews prior-based medical segmentation methods before the deep learning era and pinpoints the motivation towards integrating such prior onto the deep learning framework. Moreover, it identifies the different approaches to this integration and formulates the process as a constrained optimization problem where the constraints can either be structural or at the level of the loss function. Finally, the chapter sheds light on the different segmentation frameworks and commonly known challenges and public datasets.

Chapter 2: A Survey on High-level Prior-based Loss Functions for Medical Image Segmentation

This chapter reviews state-of-the-art methods relative to an uprising branch of prior knowledge integration methods that target the loss functions known as high-level prior-based losses. High-level prior involves information regarding object shape, or size, or topology, and is external information injected onto the learning framework. In this context, a novel categorization of methods and papers is proposed according to the type of prior they integrate: the object shape, size, topology, and the inter-regions constraints. Finally, the chapter highlights strengths and limitations of current approaches, discusses the challenge related to the design and the integration of prior-based losses, and draws future research directions.

Chapter 3: BB-UNet: U-Net with Bounding Box Prior for Medical Image Segmentation

This chapter presents BB-Unet, which is a novel U-Net variant that exploits bounding box prior inside U-Net layers in order to increase segmentation performance in medical image segmentation. By integrating bounding box information at the internal layers of the network, the segmentation model is then able to learn features relative to the object location and topology. In such a way, the proposed architecture helps in presenting attention kernels onto the neural training in order to guide the model on where to look for the organs. Moreover, it finetunes the encoder layers based on positional constraints. The proposed model is exploited within two main paradigms: as a solo model given a fully supervised framework, and as an ancillary model in a weakly supervised setting. In our contribution, manual bounding boxes are fed at inference and, as such, BB-Unet is exploited in a semi-automatic setting; However, BB-Unet has the potential of being part of a fully automated process, if it relies on a preliminary step of object detection. Results show that the proposed method outperforms state-of-the-art methods in fully supervised learning frameworks and registers relevant results given the weakly supervised domain.

Chapter 4: Investigating CoordConv for Medical Image Segmentation

This chapter introduces CoordConv-Unet, which is a novel U-Net variant that makes use of coordinate-conditioned convolutional layers, known as the CoordConv layers, in order to allow the network to take into consideration the object location. Hence, the CoordConv layers are extensions of convolutional neural network wherein convolution is conditioned on spatial coordinates via extra x and y channels. The proposed architecture demonstrates a dual role relative to prior constrained CNN learning. Thus, it can either stabilize learning while maintaining system performance, thus embracing a regularization role, or it can improve system performance by allowing the learning to be more stable and to evade local minima. Results show that, despite the inadequacy of CoordConv when trained with the regular Dice baseline loss, the proposed CoordConv-Unet structure can significantly improve model performance when trained under anatomically constrained prior losses.

Chapter 5: Effect of Prior-based Losses on Segmentation performance in Medical Imaging

This chapter investigates the performance of four low-level and high-level prior-based losses on a variety of segmentation tasks and public datasets. Moreover, it analyzes the hypothesis concerning the relations existing between loss significance and dataset properties. Generally, prior-based losses are quite interesting because they allow integration of expert knowledge while still being architecture-agnostic, that is to say, they can be plugged into any backbone network. For this, we are able to unify the segmentation network given the same learning environment, while varying the prior-based losses accordingly. The chosen prior-based losses are validated on 8 different datasets from a variety of medical image segmentation challenges including the Decathlon, the ISLES and the WMH challenge.

Chapter 6: Perimeter-based Loss for Border Irregularity

This chapter presents a novel loss constraint that optimizes the perimeter length of the segmented object relative to the ground-truth segmentation. The proposed loss computes the mean squared error between the predicted perimeter length and ground-truth perimeter length. Compared to peer contour loss approaches that are based on distance maps, the novelty of the proposed loss lies in computing the perimeter with a soft approximation of the contour of the probability map via specialized non-trainable layers. This soft optimization of contour boundaries allows the network to take into consideration border irregularities within organs while still being efficient. Relative to distance maps, which tend to underestimate the contour-to-contour distances, the exploited contour map feature emphasizes border characteristics and curvatures. Our experiments on three public datasets (spleen, hippocampus and cardiac structures) show that the proposed method outperforms state-of-the-art boundary losses for both single and multi-organ segmentation.

Research Publications

In the following table, we present the product publications of this manuscript and their relevant chapters.

Table 1 – Table of Contributions

Citation	Reference	Chapters
El Jurdi et al., 2019	R. El Jurdi, C. Petitjean, P. Honeine, and F. Abdallah. "Organ Segmentation in CT Images with Weak Annotations: A Preliminary Study." In proceedings of the 27th GRETSI Symposium on Signal and Image Processing, Lille, France, Aug. 2019b.	Chapter 3
El Jurdi et al., 2020a	R. El Jurdi, C. Petitjean, P. Honeine and F. Abdallah, "BB-UNet: U-Net with Bounding Box Prior," in IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 6, pp. 1189-1198, Oct. 2020	Chapter 3
El Jurdi et al., 2020b	R. El Jurdi, T. Dargent, C. Petitjean, P. Honeine, and F. Abdallah. "Investigating CoordConv for Fully and Weakly Supervised Medical Image Segmentation." In proceedings of the 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA), pages 1–5, 2020.	Chapter 4
El Jurdi et al., 2021a	R. El Jurdi, C. Petitjean, P. Honeine, and F. Abdallah, "CoordConv-Unet: Investigating CoordConv for Organ Segmentation," in Innovation and Research in BioMedical engineering (IRBM), 2021 (in press)	Chapter 4
El Jurdi et al., 2021b	R. EL Jurdi, C. Petitjean, P. Honeine, V. Cheplygina, and F. Abdallah, "High-Level Prior-Based Loss Functions for Medical Image Segmentation: A Survey," in Computer Vision and Image Understanding, 2021	Chapter 2
EL Jurdi et al., 2021	R. El Jurdi, C. Petitjean, P. Honeine, V. Cheplygina, and F. Abdallah, "A Surprisingly Effective Perimeter-based Loss for Medical Image Segmentation," in Proceedings of the Medical Imaging with Deep Learning conference (MIDL), Lübeck, Germany, 7 - 9 July 2021	Chapter 6

Chapter 1

Fundamentals of Medical Image Segmentation in Deep Learning

Contents

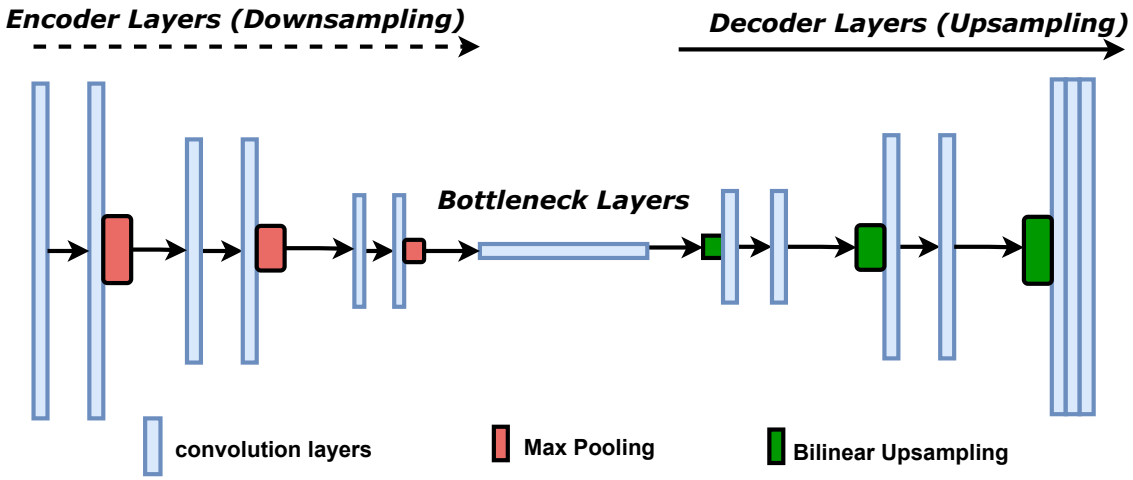


Figure 1.1 – FCN autoencoder structure with its encoder path composed of convolutional and max-pooling layers, and its decoder path of bilinear interpolation layers or transposed convolutions.

1.1 Deep Learning Fundamentals for Image Segmentation

There are two main building blocks to segmentation via deep neural networks: network architecture and loss functions. In What follows is a brief review of choices in medical image segmentation based on each building block.

1.1.1 Segmentation Networks

One of the first CNN architectures to allow automatic end-to-end semantic segmentation is the fully convolutional network (FCN) (Long et al., 2015). FCNs are derived from deep classification models, such as VGG16, AlexNet or GoogLeNet, by removing the corresponding classification layers, i.e., replacing their fully connected layers with convolutional ones, and plugging in an upsampling path that is dedicated to transforming coarse outputs into dense predictions (see Figure 1.1). With its ability to extract multiscale features, FCNs have set a milestone in segmentation approaches and paved the way for encoder-decoder segmentation networks.

To increase depth and precision within the learnt contextual features, many works within the field advocate to go deeper with FCN layers (Simonyan and Zisserman, 2015; Szegedy et al., 2015). However, increasing a model's prediction ability by adding additional layers is not an easy task because, as one goes deeper within the layers, insights on location features are hence lost. As a result, deep FCNs often fail to consider global and spatial information and are prone to producing fuzzy coarse-grained predictions (Ravishankar et al., 2017). Moreover, deepening the convolutional network will often increase the model's complexity, thus subjecting the training to additional challenges such as vanishing gradients. As a result, deep FCNs may suffer from performance saturation or degradation while training. To address these issues, many FCN improved variants have emerged, among which is the very well-known U-Net introduced by Ronneberger et al. (2015).

A pioneering approach is the U-Net model (Ronneberger et al., 2015), especially popu-

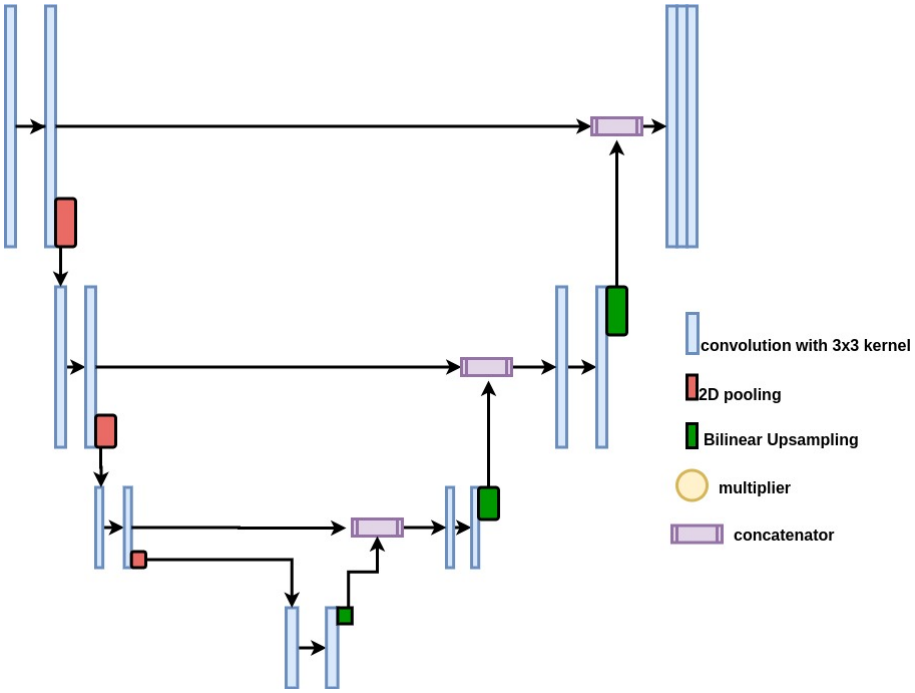


Figure 1.2 – U-Net structure with its encoder path composed of convolutional and max-pooling layers, and its decoder path constituted of bilinear interpolation layers or transposed convolutions. Skip connections are added in order to concatenate features from both downsampling and upsampling paths.

lar in medical imaging. It is characterized by a symmetric encoder/decoder structure with skip connections (See Figure 1.2). The encoder part is a contracting path composed of stacked convolutional and max pooling layers, whereas the decoder part is an expanding path composed of deconvolutional or bilinear upsampling layers. Layers within the encoder are dedicated to capturing contextual information in order to detect objects/classes present in an image. The decoder part is an expanding path composed of deconvolutions, also called transposed convolution, or bilinear upsampling layers, that help precise localization of patterns including contours and boundaries. As an image moves further into the contracting layers of a U-Net, it decreases in size but increases in depth of its learnt contextual features. In contrast, the decoder layers increase its input size but decrease its depth until it reaches its initial resolution. To make use of both contextual and positional features, skip connections between the downsampling (encoder) and upsampling (decoder) paths are added. Skip connections concatenate symmetrical features from opposing convolution and deconvolution layers. Thus, through end-to-end training, the U-Net takes on as input an image of any size and produces a segmentation map of similar dimensions to that of the input and of depth equal to the number of classes considered.

Taking advantage of U-Net's success, multiple variants emerged in order to increase model performance given different tasks (Milletari et al., 2016; Weng et al., 2019). These variants consist in changing the backbone model used for encoding, e.g. VGG or DenseNet, replacing deconvolution layers with super-resolution ones for more concise localization ability (Hu et al., 2019b), or enhancing the architecture with modifications such as attention mechanisms or residual connections (Oktay et al., 2018b; Zhang et al., 2018). In

Table 1.1 – Mathematical notations

Ω	spatial image domain, $\Omega \subset \mathbb{R}^2$ or $\mathbb{R}^3$
$\mathcal{N}_p$	set of neighboring pixels of pixel p
K	number of classes
y	true label map of dimension $\|\Omega\|$
$\widehat{y}$	predicted label (probability) map of dimension $\|\Omega\|$
y_p	true label of pixel p
$\widehat{y}_p$	probability of pixel p of having label y_p
y_p^r	binary value indicating whether pixel p belongs to class r or not
$\widehat{y_p}^r$	predicted probability of pixel p belonging to class r or not

addition, extensions to 3D images have been proposed such as the 3D-UNet model by Çiçek et al. (2016) and the V-Net model by Milletari et al. (2016). One can say that, since U-Net's publication in 2015, it has quickly become a state-of-the-art architecture for medical image segmentation with outstanding performances and results (Isensee et al., 2018).

1.1.2 Common Loss Functions for Image Segmentation

One of the key components of CNN training is the loss function, as it drives the back-propagation of the error between the predicted and the reference label. In what follows, we describe the general optimization process of neural network training under a generic loss function.

Given a training dataset of N images $\{\mathbf{x}_i\}_{i=1}^{\mathrm{N}}$ and their corresponding ground-truth masks $\{\mathbf{y}_i\}_{i=1}^{\mathrm{N}}$, the goal is to train the segmentation network so that it can learn to approximate the "true" function f parameterized by θ, that maps the input image $\{\mathbf{x}\}$ to the predicted label map, i.e., such that $f(\mathbf{x},\theta)$ represents a map with the label probability at each pixel. We define $\widehat{\mathbf{y}}_i = f(\mathbf{x}_i,\theta)$ and rely on Table 1.1 for mathematical notations.

Training the network boils down to finding the network parameters θ that minimize a loss function $\mathscr{L}(\theta)$ reflecting the problem at hand. We call this loss the data-fitting loss $\mathscr{L}_{fit}(\theta)$. For sake of simplicity, we will drop the dependency in θ and denote by $\mathscr{L}_{fit}$ the data-fitting loss function, which takes the following form:

$$\mathscr{L}_{fit} = \sum_{p\in\Omega} \mathrm{L}(\widehat{y}_p, y_p), \tag{1.1}$$

where L is a function that penalizes the discrepancy between the predicted pixel label ($\widehat{y}_p$) and the ground-truth label (y_p) for each pixel $p \in \Omega$. The shape of L defines how the error is computed. It is mainly derived from common norms, such as the Euclidean norm or the log-norm (cross-entropy shape) for example. When based on such norms, the loss is continuous and differentiable, which allows it to be efficiently optimized during back-propagation. Properties of the loss shape are important to translate task specifications. Symmetry, for instance, ensures equal penalization between errors caused by over-segmentations and ones that are caused by under-segmentations (Charoenphakdee et al., 2019).

In the literature, two main losses are widely used for segmentation: the cross-entropy (Ronneberger et al., 2015) and the Dice loss (Milletari et al., 2016) as well as their variants.

The cross-entropy $\mathscr{L}_{\text{CE}}$ is a common standard loss function that is formulated via the Kullback–Leibler divergence and computes the dissimilarity between the predicted probabilistic distribution and its corresponding target binary distribution. Its mathematical expression, given in the case of K classes, is:

$$\mathscr{L}_{\text{CE}} = -\frac{1}{|\Omega|}\sum_{r=1}^{\text{K}}\sum_{p\in\Omega} y_p^r \log(\widehat{y}_p^r), \tag{1.2}$$

where $\widehat{y}_p^r$ and y_p^r are the predicted probability and ground-truth values indicating whether pixel p belongs to class r or not. Since each pixel is handled independently from its neighbors, problems may arise due to class imbalance, as training can be dominated by the most prevalent class. For this reason, multiple works proposed variants of cross-entropy, weighted according to class or pixel imbalance (Baumgartner et al., 2017; Jang et al., 2018). One important cross-entropy variant is the weighted cross-entropy (Milletari et al., 2016), which tackles the cross-entropy sensitivity towards class distributions. It is defined as:

$$\mathscr{L}_{\text{WCE}} = -\frac{1}{|\Omega|}\sum_{r=1}^{\text{K}}\sum_{p\in\Omega} w_r\, y_p^r \log(\widehat{y}_p^r), \tag{1.3}$$

where the weighting factor w_i assigns more weight to recessive classes, thus enforcing a higher penalty on their corresponding errors. Another variant is the focal loss (Lin et al., 2017), which extends upon cross-entropy in order to deal with extreme foreground-background class imbalance in images.

Introduced by Milletari et al. (2016), the Dice loss $\mathscr{L}_{\text{D}ice}$ is a soft approximation of the well-known Dice metric, which penalizes the overlap mismatch between the predicted segmentation map and the corresponding target map. It can be computed in the general case with K classes (Sudre et al., 2017):

$$\mathscr{L}_{\text{D}ice} = 1 - 2\frac{\sum_{r=1}^{\text{K}}\sum_{p\in\Omega} y_p^r \widehat{y}_p^r}{\sum_{r=1}^{\text{K}}\sum_{p\in\Omega} (y_p^r)^2 + \sum_{r=1}^{\text{K}}\sum_{p\in\Omega} (\widehat{y}_p^r)^2}. \tag{1.4}$$

The Dice loss is sensitive to segmentation errors when the segmented object is small. For this reason, some works have aimed at weighting the Dice loss (Sudre et al., 2017; Yang et al., 2018) in order to take into account the class imbalance, or extending upon it by accounting for background pixels, such as the Kappa coefficient inspired loss (Zhang et al., 2020).

1.2 Motivation Towards Integrating Prior Knowledge onto Deep Learning for Medical Image Segmentation

1.2.1 Prior-based Image Segmentation in the pre-Deep Learning era

Among the segmentation methods that existed before the deep learning upsurge (Cremers et al., 2007; Grady, 2012), optimization-based approaches have been hugely popular, due to their versatility and efficiency. They consist in obtaining the segmentation

by optimizing an appropriate energy functional. In the case where the image domain is considered to be continuous, optimization-based approaches have embraced active contours, level sets, and deformable models in general (Xu et al., 2000). On the other hand, spatially discrete approaches consider the image as a graph, and include the well-known graph cut and normalized cuts approaches (Boykov et al., 2001; Shi and Malik, 2000). In order to counteract the effect of the noise, occlusion and low contrast in medical images, and to increase anatomical plausibility, researchers exploited prior information to guide optimization-based segmentation algorithms that allowed the encoding of constraints on segmentation results. In this context, the energy functional to be minimized was composed of at least two terms: a data-fidelity term, related to the image to be segmented, and a regularization term that controls the appearance of the contour such as its smoothness for instance. One convenient way to add prior information was to embed it into an additional term in the objective function. The additional term contains a dissimilarity measure between the model feature and the segmented region feature. Depending on the targeted property, the additional term was designed in a similar manner to the data term or to the smoothness term.

Prior information integrated within optimizational approaches encompassed a breadth of various forms, as distinguished by Nosrati and Hamarneh (2016) in their review on the topic. Thus, they can be based on elementary image properties, such as intensity, color, and texture information, or more elaborate features on the object shape, such as topological and geometrical constraints (Vicente et al., 2008), moment priors (Ayed et al., 2008; Foulonneau et al., 2008), distance and adjacency constraints (Liu et al., 2008), as well as motion and model/atlas-based priors (Lorenzo-Valdés et al., 2002; Rohlfing et al., 2005).

In the literature of prior constrained convolutional neural networks, many advances have been inspired by these past non-deep learning methods and an agreeable analogy could be realized. For example, with regards to shape prior, Mirikharaji and Hamarneh (2018) took inspiration from Veksler (2008a) to design a star-shape prior in order to induce convexity in skin lesion segmentation. Moreover, basing on the Mumphord-Shah functional in Mumford and Shah (1989), which targets regularizing perimeter length of objects, Kim and Ye (2020) proposed a novel loss function for deep learning based image segmentation without or with small labeled data. Alternatively, the Boundary loss proposed in Kervadec et al. (2019a) was derived from techniques in discrete graph-based optimization for computing gradient flows of curve evolution as demonstrated in Boykov and Funka-Lea (2006). On the other hand, the shape template (Rousson and Paragios, 2002; Slabaugh and Unal, 2005) and the popular statistical models (Cootes et al., 1995; Heimann and Meinzer, 2009) were built based on ground-truth segmentation maps and their corresponding images, even though they are not "learning" approaches per se.

Two broad differences can be drawn between optimization-based approaches and deep learning ones for segmentation. First, prior information with regards to image properties, which are represented in optimization-based methods via a fitting term, are already learned within the traditional CNN training, since the learning process is generally based on capturing characteristics relative to image data. Moreover, in order to be easily optimized, newly-designed energy terms have to be convex, which is also a condition for loss functions within CNN training. In addition, there is an evident tradeoff between an accurate modelling of an underlying property and its optimizational cost (Nosrati and

Hamarneh, 2016). In deep segmentation networks, designing new loss functions could also relate to the fidelity-optimizability trade-off by taking into consideration how well the proposed constraints are satisfied while training.

1.2.2 Segmentation Frameworks & Learning Paradigms

Since CNNs are insensitive to image noise, blur, or contrast particularities, they can be considered as excellent tools for segmentation, particularly within the medical domain. Nevertheless, segmentation results seriously depend on high-quality labels and large availability of labeled training data. Within the medical domain, data acquisition and labeling often incur high time and computational costs. For this reason, recent research contributions have targeted making use of incomplete or imperfect labels in order to conduct CNN training. Whereas CNN training under full supervision methods are the most common and extensively adopted, several recent research works aim to make use of these "not-so-accurate" labels in order to derive proper segmentation masks, thus embracing weakly- or semi-supervised learning paradigm. In weakly supervised learning, dataset labels may be incomplete, inexact or inaccurate. Weak labels can come in different forms, such as bounding boxes encompassing the understudied organ (Dai et al., 2015; Khoreva et al., 2017), image tags (Durand et al., 2017), seeds generated from object center of mass (Zotti et al., 2017), randomly or by erosion (Pathak et al., 2015). The objective of weakly supervised segmentation models is to make use of these coarse-grained annotations to derive proper and accurate predictions at the pixel level. Weakly supervised image segmentation can be conducted in two different ways: (i) a two-step iterative approach where initial label estimates are generated from weak labels in the first step, and fine-tuned through a deep learning model in the second step (Dai et al., 2015; Khoreva et al., 2017); and (ii) through a direct modification of the network (e.g. insertion of customized segmentation layers) to take into account weak labels (Durand et al., 2017; Papandreou et al., 2015). Semi-supervised learning consists in having only a subset of the training dataset with full labels, whereas the other subset unlabeled. One of the objectives of the semi-supervised learning paradigm is to learn relations between the different subsets, both labeled and unlabeled, so as to propagate labels from the supervised set to the non-supervised one (Cheplygina et al., 2019; Loog and Jensen, 2015). In this way, weakly- and semi-supervised learning frameworks make use of unlabeled or partially labeled data in order to conduct segmentation with agreeable results, thus alleviating the need for large labeled datasets.

1.2.3 Prior Knowledge in Deep Learning

There is no doubt that, thanks to the complex and powerful architectures such as U-Net and its variants, the segmentation performance has reached a serious breakthrough. Even so, multiple challenges still remain within medical imaging. Thus, these deep networks often require large amounts of annotated training data, which is not easy to obtain given the medical field. Unannotated or partially labeled data are, rather, more easily available or less computationally expensive. Moreover, even with sufficient data, automated systems generally and CNNs particularly still lack the anatomical plausibility that a medical expert has. This is mainly due to the fact that traditional U-Net architectures does not

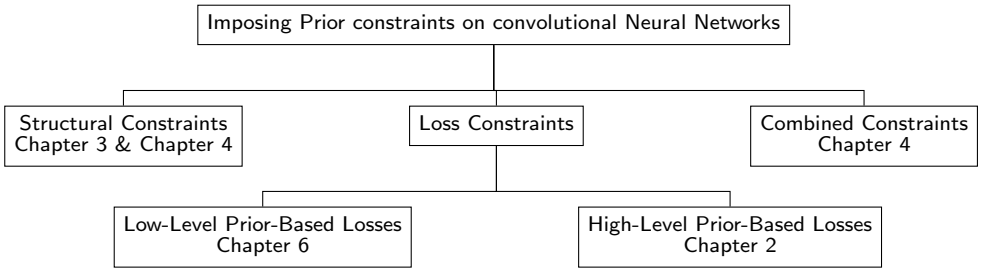


Figure 1.3 – Map of contributions across the different integration methods of Prior and their type.

leverage specific, anatomical or contextual constraints, nor does it exploit spatial relationships between organs. In the same spirit, the cross-entropy and Dice losses ignore high-level features or structures concerning the object of interest, such as their shape or topology. They also penalize all mistakes equally, regardless of their nature. As a result, segmentation produced by regular U-Net architectures via regular fitting losses often result in anatomically aberrant errors.

As previously discussed in Section 1.2.1, prior knowledge, such as the shape of particular organs or their topology, the level of border and shape irregularity have been commonly exploited in variational approaches before deep learning for medical image segmentation. Recent advances in the domain have attempted at integrating such prior information onto CNN training in order to overcome the limitations imposed by traditional deep learning segmentation networks. Thus, many works within the field have embraced the aspect of weakly supervised medical image segmentation by considering prior as incomplete or partial labels (Kervadec et al., 2018, 2020a; Khoreva et al., 2017; Khoreva et al., 2016). In the same essence, prior knowledge has also been used in order to increase anatomical plausibility in segmentations as revealed by the many works in the field (Kervadec et al., 2019a; Mirikharaji and Hamarneh, 2018; Mosinska et al., 2018).

In order to incorporate prior knowledge in a segmentation process, two major questions arise. First, one needs to define the type of prior information and the modeling of the prior. The term "prior knowledge" is quite vague and covers a breadth of notions, as can be observed from the literature. It may refer to spatial or context information, under the form of distance maps or the image gradient, which we denote as low-level prior, to more complex knowledge, regarding the anatomy of the object of interest (such as shape and size for example and the connectivity between regions), which we refer to as high-level prior.

Second, one needs to specify how to embed the prior into the segmentation network. One way to integrate prior constraints is at the level of the network architecture, i.e., modifying the network structure or adding extra layers in order to take into account this external information. Another approach is to incorporate prior at the level of the loss function. Prior-based loss functions offer a versatile way to include constraints at different scales, while maintaining interactions between regions as well as the computational efficiency of the backbone network. In this manuscript, we explore and design both structural and loss prior constraints relative to high- and low-level priors. A map of the contributions over the different chapters according to this categorization is shown in Figure 1.3. In the following section, we formulate the problem of constraining convolutional neural networks via

prior knowledge, and review some of the methods adopted in the state-of-the-art.

1.3 Integration of Prior at the Level of the Loss Function

Incorporating prior at the level of the loss function offers a versatile way to constrain neural network predictions while preserving computational complexity and generality. In this section, we demonstrate the general problem formulation of integrating prior knowledge at the level of the loss function and present the different types of prior that can be integrated into the loss terms.

1.3.1 Problem Formulation

In the general problem formulation of finding the segmentation network parameters by means of optimizing a loss function, only the data-fitting loss is considered. Hence, the network thus boils down to minimizing the loss term of Equation 1.1 as explained in Section 1.1. The integration of a prior onto this optimization problem can be demonstrated through the addition of some constraints to the optimization problem of Equation 1.1 as explained in (Kervadec et al., 2018; Márquez-Neila et al., 2017; Pathak et al., 2015; Peng et al., 2020). The constrained optimization problem can then be formulated according to the minimization of:

$$\mathscr{L}_{fit} \quad s.t. \quad \mathrm{C}_j(\widehat{y}) \leq 0 \quad j = 1,...,\mathrm{I} \tag{1.5}$$

where I is the total number of constraints in the problem. The fitting loss function $\mathscr{L}_{\mathrm{F}it}$ can be any of the common losses such as Dice or cross-entropy, as described in Section 1.1.2, whereas the constraints are mathematical representations of the prior (low-level or high-level). From here on, one can distinguish between a variety of optimization and training strategies. Moreover, optimization can be done either in a continuous domain, where the formulated loss function is mainly derived from soft probabilities (Clough et al., 2019; Kervadec et al., 2018), or a discrete domain, which directly targets hard-label assignments (Peng et al., 2020). One common method for solving constrained CNN training is through the method of Lagrange multipliers, also known as the penalty-based method (Kervadec et al., 2018; Lillo et al., 1993; Nosrati and Hamarneh, 2016). Such method models the constraint as a penalty term or a regularization term $\mathscr{L}_{penalty}$ in the loss function weighted by a parameter λ as follows:

$$\mathscr{L}_{fit} + \lambda \mathscr{L}_{penalty}. \tag{1.6}$$

The additional loss term must be differentiable, convex and produce a value proportional to the degree of constraint violation. The weighting factor λ can be either predefined throughout training (static training) or fine-tuned along training (dynamic training).

1.3.2 Types of Prior-based Losses for Medical Image Segmentation: Low-level vs. High-level Prior

Integrated prior can be low-level, which resembles reformulated ground-truth representation and is extracted from the ground-truth segmentations. For example, distance maps

(Caliva et al., 2019; Kervadec et al., 2019a; Mosinska et al., 2018) and Laplacian filters (Arif et al., 2018). A summary of state-of-the-art methods in low-level prior is demonstrated in Section 1.3.3. Prior could also be high-level representing actual external medical information such as the shape of the organ, compactness or size, and are optimized directly based on ground-truth prior tags (Dolz et al., 2017; Kervadec et al., 2018; Mirikharaji and Hamarneh, 2018). Over the past few years, integrating this type of prior into the loss function has experienced a considerable upsurge as shown by the growing literature on the topic. For this reason, we propose in Chapter 2 a survey on current state-of-the-art methods in high-level prior-based losses and categorize the methods according to the type of prior they exploit.

1.3.3 Low-level Prior-based Losses

One way to improve segmentation consistency is through conducting transformations on the ground-truth in order to obtain representations that are able to reveal geometric and contour attributes of the object of interest. This is what we denote as low-level prior-based losses.

Low-level prior can be in the form of distance map as demonstrated in (Caliva et al., 2019; Karimi and Salcudean, 2019; Kervadec et al., 2019a). In this context, two major contributions are the Boundary loss (Kervadec et al., 2019a) and the Hausdorff loss (Karimi and Salcudean, 2019). The Boundary loss is an approximation of the distance between the real and the estimated boundaries. Through Boykov et al. (2001) graph theories, Kervadec et al. (2019a) derive an equivalent term that fine-tunes the probability distribution via ground-truth distance maps. On the other hand, the Hausdorff loss in Karimi and Salcudean (2019) conducts a direct point-by-point optimization of the predicted and ground-truth contours. Both works tackle the problem of contour optimization between ground-truth and predicted segments, to increase anatomical plausibility in their respective deep learning segmentation models. However, Karimi and Salcudean (2019) insure that the constraints imposed by the distance maps are insured, since they target direct optimization of distance map features between ground-truth and predicted segmentations. On the other hand, Kervadec et al. (2019a) alleviate the high computational load demonstrated by Karimi and Salcudean (2019), resulting from the online computation of the predicted distance maps per each iteration and for all images in the dataset, by simply fine-tuning the probability outputs of the segmentation networks via the ground-truth distance maps. Instead of weighting the probability distributions as in Kervadec et al. (2019a), Caliva et al. (2019) exploit distance maps as weighing factors for a cross-entropy loss term in order to improve extraction of shape characteristics and enable the network to focus on hard-to-segment boundary regions. As a result, the loss in Caliva et al. (2019) gives more weight to pixels lying in close proximity of the segmented anatomical objects than those that are far away.

Instead of distance maps, Yang et al. (2020) exploit Laplacian filters in order to develop a boundary enhanced loss term that invokes the network to generate strong responses around the boundary areas of organs while producing a zero response given pixels that are farther from the peripheries. In the same context, Arif et al. (2018) extend the regular cross-entropy term with an average point to curve Euclidean distance factor between

predicted and ground-truth contours in order to allow the network to take into consideration shape specifications of segmented structures. In addition, Gerl et al. (2020) propose a shape-aware loss that provokes the network to produce smooth boundary regions by penalizing the difference in probability values between neighboring pixels.

Different from the previous approaches, Mosinska et al. (2018) propose to leverage the topological information or shape descriptors present within the internal layers of the VGG16 network, in order to close small gaps in thin structures (neuronal membranes) and alleviate topology mistakes. Kim and Ye (2020) introduce a loss term inspired by the Mumford-Shah functional in order to force each region to have a similar pixel intensity, and to regularize the contour length. Lambert et al. (2020) propose a criterion of edge alignment, based on a weighted total variation term.

1.3.4 Continuous vs Discrete Optimization Strategies

Integration and optimization of prior constraints at the level of the loss function can be carried out either through the continuous domain or through the discrete domain. One very popular and simple method to integrate constraints within a continuous domain is through the penalty-based method. The penalty-based method involves formalizing the constraint as an additional penalty loss term in the main loss function weighted by a parameter λ, which may be statically or dynamically defined through training. The novel loss term, which includes the main per-pixel fitting loss, and the novel penalty loss are then optimized using regular stochastic gradient descent. Many works adopt the penalty-based method while dealing with anatomical prior for its ease of formulation and use (BenTaieb and Hamarneh, 2016; Clough et al., 2019; Hu et al., 2019c; Kervadec et al., 2018; Mirikharaji and Hamarneh, 2018; Shit et al., 2020). Despite this simplicity, penalty-based approaches may not guarantee constraint satisfaction. Moreover, they require careful fine-tuning of their weighting terms, which may not be convenient in the case where multiple constraints occur and where one constraint may overshadow the others.

One way to deal with multiple constraint optimization, demonstrated in (Kervadec et al., 2020a,b), is through Lagrangian optimization with log-barrier extensions. The method involves introducing a standard log-barrier function that avoids the need for dual optimizations and their issues (Boyd and Vandenberghe, 2004). The method then integrates these constraints into the log-barrier function and solves the optimization process in an unconstrained manner via stochastic gradient descent. Unlike the penalty-based approach, the log-barrier approach does not yield null gradients or cause oscillations between competing constraints. It is rather characterized by stable gradients that insure training stability.

Optimization in a discrete domain can also be insightful, given the discrete nature of some types of prior, namely the high-level anatomical prior. Whereas continuous optimization makes assumptions on the soft probabilities in order to estimate differentiable functions, discrete optimization involves extracting the features in their discrete form from model predictions, and optimizing them relative to the ground-truth. One way to perform discrete optimization is through the ADMM method (Peng et al., 2020). The ADMM algorithm generally aims at separating the optimization of the network parameters under SGD from the optimization of discrete constraint segmentation labels. Dis-

cretely optimizing networks generally benefit from the ability to solve sub-problems, either continuous or discrete, separately, and insure global optimum, which can improve solutions within a single gradient step and at higher convergence speed.

1.4 Integrating Prior Constraints at the Level of Network Structures

Structural constraints consist in modifying parts of the network in order to take into consideration external specifications and prior information regarding the anatomical object of interest. Among structural constraint methods, integration of prior can be done either externally in conjunction with the segmentation network (Ghafoorian et al., 2016; Khoreva et al., 2017; Ravishankar et al., 2017; Trullo et al., 2017) or at the intermediate levels (Oktay et al., 2018b; Trullo et al., 2017). Moreover, many works introduce interchangeably both structural and loss constraints in order to enhance segmentation performance (Oda et al., 2018; Ravishankar et al., 2017).

In the context of structural prior constraints, Ghafoorian et al. (2016) propose a cascade of several deep CNN architectures that consider multi-scale patches in order to incorporate anatomical location in their decision making process. Thus, spatial location of patches extracted from the image into a CNN model is injected posterior to the convolutional layers. Similarly, Trullo et al. (2017) demonstrate two collaborative architectures in order to iteratively refine the posterior probability and provide information about neighboring organs. In this work, anatomical constraints are obtained from an auxiliary network that are later used by the segmentation network (U-Net) in order to refine ill-defined organ boundaries. Basing on the same notion, the SR-UNet in Ravishankar et al. (2017) jointly adds an external non-linear shape regularization network to the segmentation one in order to take into consideration the incomplete, over- or under-segmented shape masks provided by the U-Net. However, unlike Trullo et al. (2017) which fine-tunes ill-defined segmentations by U-Net using the constraints obtained from the external network, Ravishankar et al. (2017) maps the segmentations to conform to a manifold of permissible training shapes. Thus, the main function of their adjoint network is to learn projections of arbitrary shapes onto a manifold space. Oktay et al. (2018a) adopt a similar regularization approach to that in Ravishankar et al. (2017). However, they target the decoder layer in their U-Net-like structure and train the up-sampling layers through super resolution ground-truth maps.

Methods in Zotti et al. (2017) and Zotti et al. (2019) extend upon U-Net by proposing a novel structure that learns good features for predicting proper segmentation masks of their understudied organ by properly computing organ center of mass from intermediate U-Net-like layers. In Zotti et al. (2017), a regression model is introduced at the bottleneck level of its U-Net-like structure in order to extract the center of mass corresponding to their understudied organ. The extracted feature map is then merged with that of the decoder layer, then segmentation maps are derived. To avoid anatomically impossible shapes, Zotti et al. (2017) extend upon their previous work to further estimate a probability distribution from the training data with regards to the occurrence probability of the understudied organ.

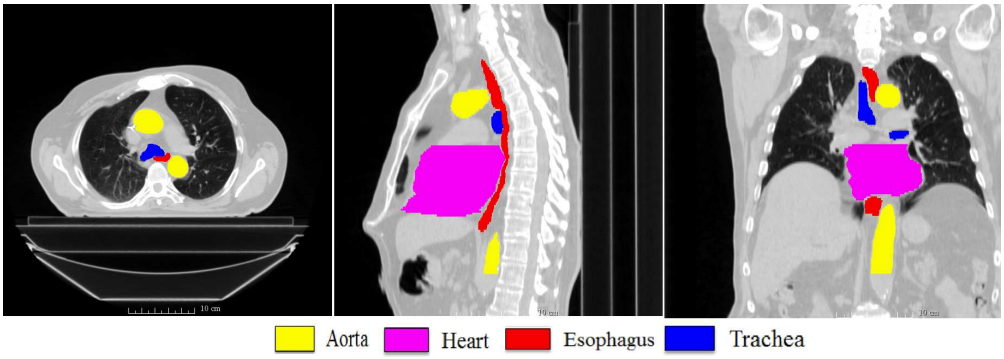


Figure 1.4 – SegTHOR Dataset with its 4 organs: Heart, Aorta, Trachea and Esophagus

In many works, authors introduce interchangeably both structural and loss constraints. For instance, the method of Ravishankar et al. (2017), where a non-linear shape regularization model is trained jointly along U-Net, incorporates a loss function that updates the segmentation network (U-Net) parameters based on the regularized predicted segments, the rough predicted segments as well as the ground-truth labels. Similar to Ravishankar et al. (2017), Tofighi et al. (2018) demonstrate a manifold of permissible nuclei shapes prepared by a domain expert and incorporate this prior information in the form of a regularizing term that encourages detection inside nuclei boundary while simultaneously penalizing false positives. In the same context, Oda et al. (2018) introduce a boundary enhanced loss function similar to that of Caliva et al. (2019) and Arif et al. (2018). However, instead of weighting by the errors through distance maps, Oda et al. (2018) add an extra decoder branch to the U-Net network in order to predict hard to segment boundaries. In the same manner, the method in Zotti et al. (2017) integrates the center of mass and contour prior into the loss function, which was obtained from an encoder/decoder structure trained end-to-end along the segmentation network.

1.5 Segmentation Challenges

Ever since the deep learning upsurge, various medical image datasets and competitions have been launched to promote the development of deep learning methods for computer-aided diagnostics. In what follows, we describe the challenges considered within our research.

1.5.1 SegTHOR Challenge

The SegTHOR challenge addresses the segmentation problem of organs at risk in Computed Tomography (CT) images (Lambert et al., 2020). The goal of the SegTHOR challenge is to segment 4 organs at risk: heart, aorta, trachea and esophagus. The SegTHOR dataset consists of 60 CT scans of patients characterized with non-small cell lung cancer and referred for radiotherapy. The dataset was acquired at the cancer center Centre Henri Becquerel in Rouen, France. The organs at risk in CT images were manually segmented by an expert radiotherapist. The CT images are $512 \times 512 \times (150 \sim 284)$ voxels with in-plane

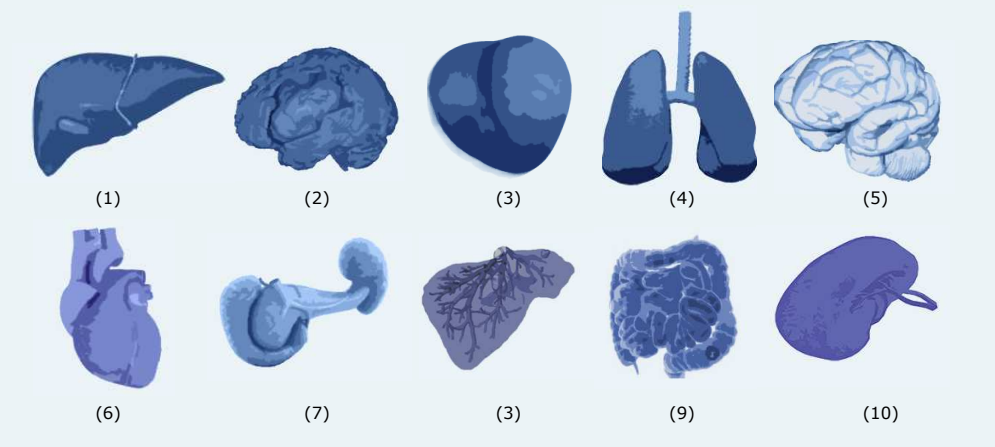


Figure 1.5 – Decathlon Challenge with its 10 datasets: (1) Liver and tumor (2) Brain Tumors: Gliomas segmentation necrotic/active tumor and oedema (3) Prostate: Prostate central gland and peripheral zone (4) Lung and tumors (5) Hippocampus head and body (6) Cardiac: Left Atrium (7) Pancreas: Liver and tumor (8) Hepatic vessels and tumor (9) Colon Cancer Primaries (10) Spleen

resolution varying between 0.90 mm and 1.37 mm per pixel, depending on the patient. The number of slices range from 150 to 284 with a z-resolution between 2 mm and 3.7 mm depending on the patient. The dataset was released publicly in a competition conducted at the 2019 IEEE International Symposium for Biomedical Imaging[1].

1.5.2 Decathlon Challenge

The Decathlon challenge[2] consists of 10 datasets of different modalities and segmentation tasks. The main objective of the challenge is to investigate how well deep learning methods can work well on different tasks each with their own set of particularities such as small data, unbalanced labels, large-ranging object scales, multi-class labels and multi-modal imaging. In this thesis, we have mainly focused on training and validation relative to the training set due to the lack of availability of testing labels for the public outside the challenge environment.

(1) Liver Tumor is a dataset dedicated to the segmentation of the liver organ and the corresponding tumor within Portal venous phase CT images. The dataset is composed of 201 3D volume images from patients with 131 samples for training and 70 samples for testing. One of the main challenges of this dataset is that it is characterized with unbalanced labels relative to the liver (large structure) and the corresponding tumor (small structures).

(2) Brain Tumor is a dataset dedicated to Gliomas segmentation. The task at hand is multi-label where two anatomical objects are to be distinguished: the necrotic/active

[1]The SegTHOR dataset is available at `https://competitions.codalab.org/competitions/21145`.
[2]The Decathlon challenge datasets are available at http://medicaldecathlon.com/tasks

tumor and the oedema. The dataset consists of 750 samples of 4D volumes of multi-parametric MRI divided into 484 samples for training and 266 samples for testing.

(3) Prostate dataset is a multi-parametric MRI dataset dedicated to segmenting two adjacent regions, the prostate central gland (PG) and its corresponding peripheral zone (PZ). The dataset is composed of 48 4D volume samples divided into 32 samples for training and 16 samples for testing. The main challenge of this segmentation task is to distinguish between two adjoint regions that have large inter-subject variations.

(4) Lung Tumor Dataset is a CT dataset dedicated to segmenting the Lung and its corresponding tumor. It is composed of 96 3D volumes divided into 64 training and 32 testing. The main challenge of the segmentation task is to detect a very small target structure (the cancerous tumor) given a very large image.

(5) Hippocampus is a mono-modal brain MRI dataset whose main objective is to simultaneously segment two tissues given low resolution images. The dataset is composed of 394 3D volume samples divided into 263 samples for training and 131 for testing. The challenge presented within this dataset is to simultaneously segment two anatomical entities that are in very close proximity to each other.

(6) Atrium Dataset is a mono-modal cardiac MRI dataset dedicated to segmentation of the left atrium. The dataset is composed of 30 3D volume samples divided into 20 training and 10 testing and is characterized by being small with large variability in organ size.

(7) Pancreas Tumor Dataset is a CT dataset for segmenting 2 anatomical objects simultaneously: the pancreas and its corresponding tumor. It is composed of 420 3D volumes divided into 282 training samples and 139 testing. The main challenge exhibited by this dataset is the large imbalance for label size relative to the moderate size pancreas, the small tumor, and the corresponding large background.

(8) Happatic Vessel Dataset is a CT dataset dedicated to segmenting the hepatic vessel and its corresponding tumor. It is composed of 443 3D volumes divided into 303 training and 140 testing. The main challenge is the need to segment the small tabular structures next to the heterogeneous tumor.

(9) Colon is a CT dataset whose objective is to segment colon cancer primaries. The main challenge of the dataset is the heterogenous appearance of the anatomical structure and its variability relative to its location in the image.

(10) Spleen Dataset is a Decathlon CT dataset whose objective is to segment a single organ (the spleen) with a single instance. The challenge represented by the data is due to the shape of the spleen which is characterized with a largely varying size and mild convexity issues at boundary levels.

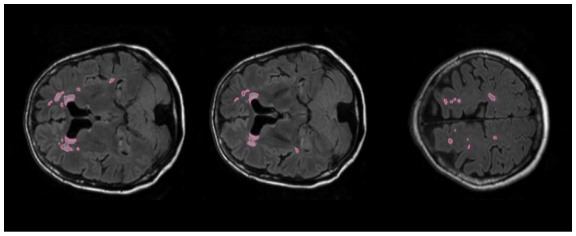


Figure 1.6 – White Matter Hyper-intensity dataset (WMH) from the MICCAI challenge (Kuijf et al., 2019)

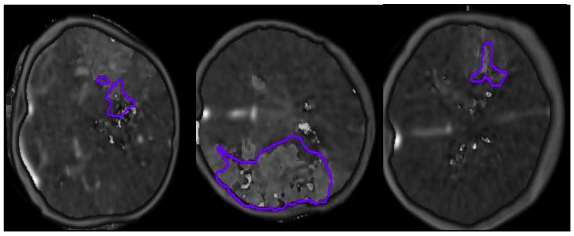


Figure 1.7 – Ischemic Stroke Lesion Segmentation (ISLES) dataset from the MICCAI 2015 challenge (Maier et al., 2017)

1.5.3 ISLES Challenge

The Ischemic Stroke Lesion Segmentation (ISLES) dataset[1] is a multi-spectral MRI public dataset that demonstrates diverse ischemic stroke cases with their corresponding ground-truth segmentations (Maier et al., 2017). The training dataset is composed of 94 multi-modal scans with images of 5 modalities. Images within the dataset are characterized with a spatial resolution that ranges from 0.8 mm × 0.8 mm × 4 mm to 1 mm × 1 mm × 12 mm. Lesions within this dataset are sparse and are scattered over the entire brain area, thus exhibiting a considerable challenge for segmentation networks. This dataset took part of a challenge within MICCAI 2018.

1.5.4 White Matter Hyperintensity Challenge

WMH is a multi-modal MRI dataset that is dedicated to segmenting lesions in the brain (Kuijf et al., 2019). The dataset is characterized with high imbalance with regards to the proportion of slices with and without lesions and also with regards to the background, which is much larger than the segmented anatomical object. The dataset has 2 modalities in MRI images with a number of lesions (connected components) ranging between 1 to 26. The dataset is a public dataset that is part of 2017th MICCAI challenge[2].

1.5.5 ACDC Challenge

ACDC dataset is a cardiac cine MRI dataset consisting of 123 patients. The task at hand is to simultaneously segment three elements of the heart: the left and right ventricular endocardium (LVC and RVC, resp.) and the myocardium (MYO). The segmentation task is rather challenging as the 3 structures are very close and characterized by non-convex shapes. The LVC is ring-shaped, encompasses the simple circular shaped MYO and is bonded to the RVC (See Figure 1.8).

[1] http://www.isles-challenge.org/
[2] https://wmh.isi.uu.nl/

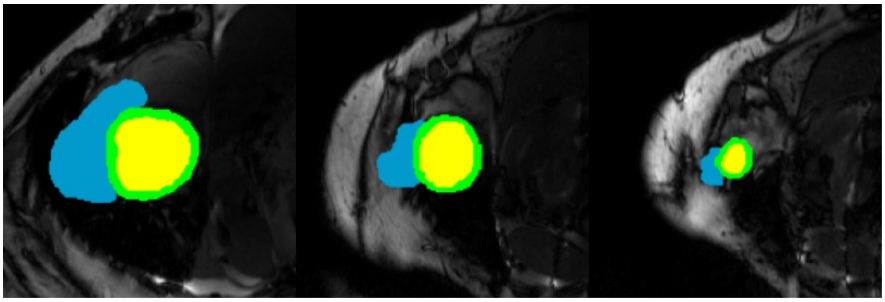


Figure 1.8 – ACDC Dataset: right ventricle (RVC) is in blue, left ventricle is yellow (LVC) and myocardium in green (MYO)

1.6 Conclusion

In this chapter, we introduced the building blocks of medical image segmentation via deep learning. Then, we have clarified the motivation behind exploiting prior knowledge in deep learning frameworks, and identified their different integration strategies. Finally, the chapter sheds light on the challenges and public datasets exploited in this manuscript. In the next chapter, we are going to demonstrate the first contribution of this manuscript that involves a survey on state-of-the-art methods relative to an uprising branch of prior-based losses known as high-level prior-based losses.

Chapter 2

A Survey on High-level Prior-based Loss Functions for Medical Image Segmentation

“ Everyone participates according to his abilities, to the progresses made in his century ”

Denis Diderot

Contents

2.1 Motivation

Recently, the integration of high-level prior concerning anatomical properties has become an increasing trend in automated medical image segmentation, as shown by the growing literature on the topic. For this reason, we dedicate this chapter to summarizing recent contributions that exploit high-level prior at the level of the loss function to the field of prior-constrained convolutional neural networks. The main objective of the conducted survey is to better understand the intuition and mechanism of designing prior-based losses.

We have categorized the articles according to the nature of the prior, that may be the object shape, size, topology, and the inter-regions constraints. We seek to provide the reader with (i) what types of priors exist in the literature and how they are modeled and embedded into the loss function, (ii) the major challenges linked to the design of such prior-based losses, and (iii) their common training and optimization strategies. To do so, Google Scholar was queried for peer-reviewed publications that included tags such as constraint losses, medical image segmentation, prior-based losses, constraint neural networks and anatomical constraints. The papers were then filtered in terms of employing a deep CNN-backbone for segmentation and the novelty present in the design of a new prior loss function.

The rest of the chapter is organized as follows. In Section 2.2, we first review existing surveys in CNN-based medical image segmentation. Section 2.3 contains the proposed categorization of prior-based loss functions, along with a review of the corresponding works. Section 2.4 exhibits the common challenges and weaknesses faced while designing such losses and discusses some future trends and perspectives. Finally, we conclude the survey in Section 2.5.

2.2 Surveys in Medical Image Segmentation

Since the rise of convolutional neural networks in computer vision, various medical image segmentation surveys have been published (Chen et al., 2020; Haque and Neubert, 2020; Havaei et al., 2016; Hesamian et al., 2019; Jiang et al., 2017; Lei et al., 2020; Razzak et al., 2018; Renard et al., 2020; Taghanaki et al., 2019; Zhou et al., 2019).

In one of the most recent surveys presented by Lei et al. (2020), the authors consider contributions according to the level of supervision. Within the fully supervised setting, the paper categorizes the methods according to backbone network, design of network blocks, and loss functions improvement. Some regularization losses via distance-based transforms are briefly reviewed; however, there is no focus on high-level priors for loss functions. Other recent surveys in the field like (Renard et al., 2020) review methods with regards to their source of variability and degree of reproducibility. Zhou et al. (2019) focus on multi-modal fusion strategies and approaches for medical data segmentation. Razzak et al. (2018) present a survey of medical image segmentation with deep learning and categorized methods in terms of convolutional neural network structure and training techniques (weakly-, semi- or fully-supervised). Taghanaki et al. (2019) report contributions in image segmentation for natural and medical images and categorize existing works according to six main categories: model architecture, image modality, loss functions, model

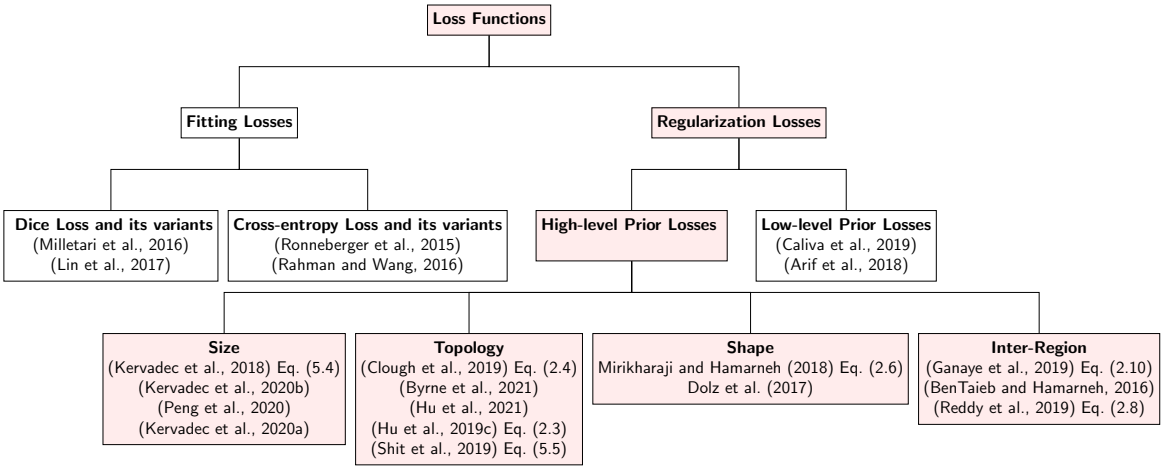


Figure 2.1 – Loss functions categorization tree. Loss functions can be data-fitting loss or regularization loss. Regularization or prior-based losses can be distinguished according to the nature of prior that they incorporate: low-level prior (Section 1.3.3) or high-level prior (Section 2.3). For each paper, the equation refers to the loss function to optimize.

types, supervision strategies and multi-task learning. Haque and Neubert (2020) present an overview of the different deep learning methods used within the biomedical image segmentation domain and categorize them according to the image modality, the segmentation architectures, and their evaluation metrics. Domain specific reviews also exist, focusing on a particular pathology or organ: brain pathology (Chahal et al., 2020; Havaei et al., 2016; Magadza and Viriri, 2021; Meyer et al., 2018), cardiac image segmentation (Chen et al., 2020; Yue et al., 2019) and breast pathology (Debelee et al., 2020; Krithiga and Geetha, 2020).

As far as prior-based loss functions are concerned, Taghanaki et al. (2019) briefly reviews a few works that integrate boundary and topological prior-based losses, in addition to presenting common loss functions, as well as their variants and combinations. Chen et al. (2020) mention the idea of imposing anatomical constraints at the level of the loss function, as a solution that allows the network to learn features representing the underlying anatomical structures. However, this survey focuses only on cardiac image segmentation. Moreover, it includes works that combine structural and loss constraints. Thus, to the best of our knowledge, no survey deals specifically with prior-based loss functions in image segmentation.

Table 2.1 – List of reviewed papers with respect to the category of the prior types: topology, size, shape, inter-regions priors. Evaluation metrics: std refers to the standard evaluation metrics, such as pixel-wise accuracy (pA), dice similarity coefficients (DSC), Hausdorff distance (HD). specif. means that the paper uses prior-specific metrics. SGD: stochastic gradient descent. ADMM: Alternating direction method of multipliers.

Paper	Architecture	Optimized features	Dim.	Class nb	Training strategy	Evaluation metrics std	Evaluation metrics specif	Supervision	Constraint nb	Optimization domain	Optimization method
Size priors											
(Peng et al., 2020)	U-Net, E-Net	Binary segmentation proposals and Lagrangian multipliers	2D 3D	1	N/A	DSC	-	weak	2	Discrete	ADMM
(Kervadec et al., 2020a)	U-net E-Net	Combined tightness and size prior constraints	2D	1	static	DSC	-	weak	3	Continuous	Lagrangian multiplier with Log-barrier extensions
(Kervadec et al., 2018)	U-net E-Net	Lower and upper bounds on organ size	2D 3D	1	dynamic	DSC		weak	1-2	Continuous	Penalty based approach via SGD
Topology priors											
(Byrne et al., 2021)	U-Net	Betti number	2D 3D	3	static	DSC	yes	full	1	Continuous	Penalty based approach via SGD
(Hu et al., 2021)	U-Net, DIVE, Mosin	Morse skeleton	2D 3D	1	static with refinement	pA	yes	full	1	Continuous	Penalty based approach via SGD
(Clough et al., 2020)	U-Net	Betti number	2D 3D	1	static	DSC	yes	semi	1	Continuous	Penalty based approach via SGD
(Hu et al., 2019c)	U-Net, DIVE, Mosin	Birth and death of topological components (persistent homology)	2D 3D	1	static with refinement	pA	yes	full	1	Continuous	Penalty based approach via SGD
(Shit et al., 2019)	U-Net FCN	Skeletons	2D 3D	1	static	DSC, pA	yes	full	1	Continuous	Penalty based approach via SGD
Shape priors											
(Mirikharaji and Hamarneh, 2018)	U-Net ResNet-DUC	Star-shapeness	2D	1	static	DSC, pA, Jaccard, Spec, Sens	-	full	1	Continuous	Penalty based approach via SGD
(Dolz et al., 2017)	FCNN	Ratio of boundary length to area	3D	1	N/A	DSC	-	full	1	Discrete	ADMM
Inter-regions priors											
(Ganaye et al., 2019)	SD-Net	Prohibited and permissible connectivity between organs	2D 2.5D 3D	20-135	dynamic	DSC, HD	yes	full, semi	1	Continuous	Penalty based approach via SGD
(Reddy et al., 2019)	FCN	Boundary smoothness and containment/exclusion properties	2D	3	static	DSC, HD	-	full	2	Continuous	Penalty based approach via SGD
(BenTaieb and Hamarneh, 2016)	Alexnet-FCN FCN-8 U-Net	Boundary smoothness and containment/exclusion properties	2D	2	static	DSC, pA	-	full	2	Continuous	Penalty based approachvia SGD

2.3 Proposed Categorization of High-level Prior-based Loss Functions

In this section, we review the approaches that aim to integrate high-level prior constraints for medical image segmentation at the level of the loss function, listed in Table 2.1. In order to contextualize the high-level prior-based loss functions, we present a categorization of loss functions, for medical image segmentation, in Figure 2.1, we distinguish between data-fitting losses (as described in Section 1.1.2) that model the problem at hand and regularization losses. Prior-based losses are considered to act as regularization losses, and can be classified according to the nature of prior that they incorporate: low-level prior (as already reviewed in the previous chapter) or high-level prior (the current chapter). We have categorized the high-level priors (and subsequently the 13 reviewed papers in this section) according to the nature of the constraint: size, topology, shape and inter-regions constraints.

2.3.1 Size Constraints

The size of an organ is a feature that has a known range of variability. In (Kervadec et al., 2018; Pathak et al., 2015), the idea is to integrate this information into the segmentation process, and to constraint the predicted organ area to be in this known size range. The problem is to estimate the organ size from a soft probability map. Pathak et al. (2015) are the first to propose a size constraint optimized through biconvex Lagrangian dual methods. They formulate the ground-truth as a latent distribution. Then, they alternate between bringing the probability distribution to be as close as possible to the ground-truth distribution, given fixed model parameters on one hand, and optimize model parameters via gradient descent given known latent distribution on the other hand. They experiment their method on natural images. Kervadec et al. (2018) address the problem of medical image segmentation and argue that the dual optimization problem is computationally intractable when applied to neural networks. As a result, it is more convenient to integrate the size prior directly at the level of the loss function, under the form of a differentiable penalty term and optimize model parameters accordingly. Thus, Kervadec et al. (2018) do not threshold the predicted label map, but rather estimate the area with the summation of the probabilities over the whole image domain. The estimated organ size is defined as:

$$\mathrm{A}(\widehat{\mathbf{y}}) = \sum_{p \in \Omega} \widehat{y}_p, \tag{2.1}$$

Then prior knowledge is used to impose a lower bound a and a higher bound b on the organ size. A penalty loss function that integrates these bounds is proposed as follows:

$$\mathscr{L}_{size} = \begin{cases} \left(\mathrm{A}(\widehat{\mathbf{y}}) - a\right)^2 & \text{if } \mathrm{A}(\widehat{\mathbf{y}}) \leq a, \\ \left(\mathrm{A}(\widehat{\mathbf{y}}) - b\right)^2 & \text{if } \mathrm{A}(\widehat{\mathbf{y}}) \geq b, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

The proposed loss $\mathscr{L}_{size}$ is implemented in a weakly supervised setting for cardiac segmentation. In this case, prior knowledge is used in order to overcome the problem of partial label absence.

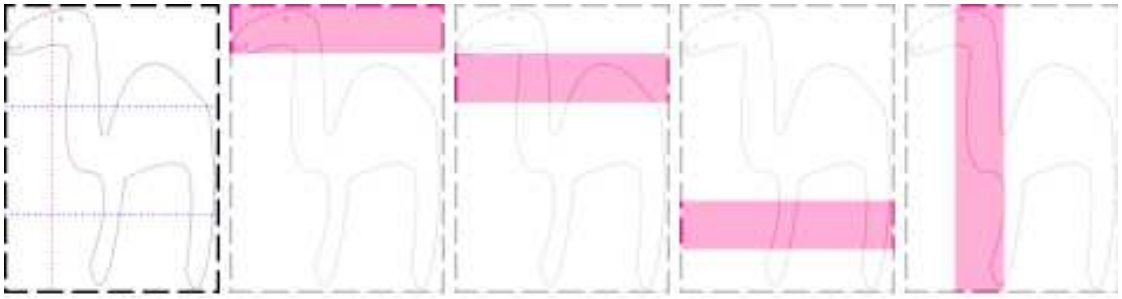


Figure 2.2 – Tightness prior. (i) Camel image. Given that any segment (pink stripe) is made up of w lines and that each line crosses the camel at least once, any segment (horizontal or vertical) of width w crosses the camel in at least w pixels, as illustrated in (ii), (iii), (iv) and (v). Figures from (Kervadec et al., 2019b).

Methods of Kervadec et al. (2018) and Pathak et al. (2015) explore the optimization of neural networks under constraints, given a continuous domain. However, conducting optimization from a continuous perspective may not guarantee that the discrete constraints are satisfied. The issue of the discrete nature of anatomical constraints have led to discrete optimization of neural networks in Peng et al. (2020), where the authors investigated the alternating direction method of multipliers algorithm (ADMM) in order to incorporate boundary smoothness and size constraints. The ADMM method is a variant of the augmented Lagrangian scheme, which allows the decoupling of the continuous optimization of neural network parameters by gradient descent, from the discrete optimization of size constraints.

Both (Kervadec et al., 2020a) and (Kervadec et al., 2019b) aim to exploit a bounding box prior as means of extracting size and tightness constraints. With bounding box annotations, they are able to constrain the organ size and location, inside the bounding box. While Kervadec et al. (2019b) integrate a bounding box prior to conduct fully supervised medical image segmentation, their work in Kervadec et al. (2020a) demonstrate the ability of exploiting this prior in order to permit segmentation under weak supervision with performance close to full supervision. To integrate a bounding box prior, the authors argue that the segmented region should be sufficiently close to the sides of the bounding boxes. Thus, each horizontal or vertical line that is parallel to the sides of the bounding box is to intersect the target segment at least once. As a result, the sum of pixels along the line should be greater than the sum of pixels belonging to the label (Figure 2.2). To integrate the multiple constraints, authors adopted a Lagrangian optimization method with log-barrier extensions (Kervadec et al., 2019b). The method involves introducing a standard log-barrier function (Boyd and Vandenberghe, 2004) that evades the need for dual optimizations and their issues. Optimization under the log-barrier extensions have been introduced previously (Chouzenoux et al., 2019); however, it is still a novel research direction in medical image segmentation.

2.3.2 Topology constraints

Topology is concerned with the properties of spatial objects by abstracting their connectivity, while ignoring their detailed form (Ségonne and Fischl, 2015). In this section, we

present works that are based on explicit topology modeling, through the use of Betti numbers, a measure of topological structures (Clough et al., 2019; Hu et al., 2019c, 2021), and skeletonization (Shit et al., 2019).

Betti numbers are topological invariants. They are determined for a dimension k: Betti number β_k is the number of k-dimensional features on a topological surface. For example, β_0 represents the number of connected components and β_1 the number of holes (in 2D Images, only these two Betti numbers are useful). Betti numbers are discrete, obtained on thresholded binary predictions, and as such cannot be used directly for CNN training. Instead, Hu et al. (2019c) and Clough et al. (2020) have exploited the notion of persistent homology to integrate this prior onto the loss term. Persistent homology is a transformation that encodes the evolution of topological structures of nested spaces. In our case, the persistent homology consists in thresholding the prediction map with a linear sequence of threshold values and summarizing this information in concise format. In particular, depending on the threshold, some structures that may be 0-dimensional or 1-dimensional are born or die in the image.

Hu et al. (2019c) exploit homology via persistent diagrams ($\mathrm{D}gm$). Each persistent diagram contains a finite number of 2D dots ($p = (b, d)$) corresponding to a topological structure that is born at threshold b and killed at a threshold d. Then, the goal is to find the best one-to-one correspondence noted as γ, between the sets of dots corresponding to the ground-truth persistent diagram ($\mathrm{D}gm(g)$) and to the predicted persistent diagram ($\mathrm{D}gm(f)$), by minimizing the squared distance between them as follows:

$$\mathscr{L}_{\mathrm{T}opo1} = \sum_{k} \sum_{p_l^k \in \mathrm{D}gm(f)} \left(b(p_l^k) - b(\gamma(p_l^k))\right)^2 + \left(d(p_l^k) - d(\gamma(p_l^k))\right)^2. \tag{2.3}$$

Following an idea similar to persistence diagrams, topological structures can be represented through persistence barcodes, as in (Clough et al., 2019) and (Byrne et al., 2021) (Figure 2.3). Here, the authors make the hypothesis that Betti numbers for the segmented object, denoted as β_k^*, are known. The idea is to consider that longest bars have the most meaningful topological features in the data, the length being computed as the difference between birth and death. Ideally, longest bars should have a length reaching 1 indicating the persistence of the topological feature throughout the entire threshold space. Hence, their loss aims to maximize the longest bars (first term in Eq. 2.4), and to get rid of transient components, corresponding to shortest bars (second term in Eq. 2.4) as follows:

$$\mathscr{L}_{\mathrm{T}opo2} = \sum_{k} \Big(\sum_{l=1}^{\beta_k^*} \left(1 - |b(p_l^k) - d(p_l^k)|^2\right) + \sum_{l=\beta_k^*+1}^{\infty} |b(p_l^k) - d(p_l^k)|^2 \Big). \tag{2.4}$$

Whereas Clough et al. (2019) investigate this loss within the binary segmentation setting, Byrne et al. (2021) extend this work for multi-class segmentation of cardiac MR Images.

As an alternative to persistent homology, Hu et al. (2021) exploit notions of discrete Morse theories (Milnor, 2016) in order to propose a novel approach that identifies critical topological structures and preserves desired Betti numbers. Morse theories rely on the assumptions that network outputs or probability maps are rather terrain functions characterized with ridges and valleys, representing critical topological structures. Their proposed loss identifies these structures and enforces higher penalties along them.

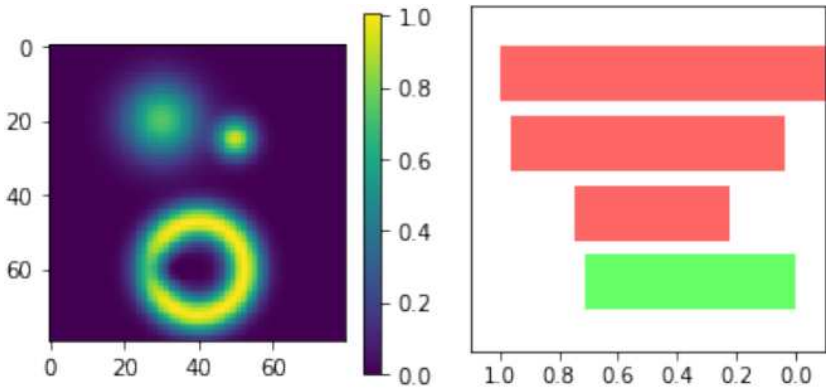


Figure 2.3 – Persistent barcode diagram. The probability map on the left contains three visible regions of high intensity, which correspond to the three persistent 0-dimensional features shown as red bars in the right diagram, with threshold values corresponding to birth and death of a topological feature on the x-axis. The map also contains a loop of high intensity, corresponding to the one persistent 1-dimensional feature, shown here as a green bar on the barcode diagram. Figures from (Clough et al., 2019).

Another important concept that reveals topological properties of objects is the skeleton. Skeletonization is the process of obtaining compact representations of images and objects while still preserving topological properties. The idea of (Shit et al., 2019) is to constrain the skeleton of the predicted map to match the skeleton of the ground-truth map. This prior is used in the segmentation of vessels and neurons in both 2D and 3D. While the skeleton of a shape can be obtained with various approaches, the authors underline that using a discrete operation, such as the Euclidean distance transform, does not allow to obtain a differentiable approximation. Thus, they propose to use morphological thinning, a sequence of dilations and erosions. To handle the soft probability map values, erosions and dilations are replaced by their grey-scale equivalent (min and max filters), giving rise to 'soft-skeletonization'. In the CNN, iterative min-pooling and max-pooling is applied as a proxy for morphological erosion and dilation. Once the skeleton is computed, an appropriate prior loss term can be designed.

Let $\mathbf{s}$ and $\widehat{\mathbf{s}}$ be the ground-truth and the predicted skeleton respectively, of size $|\Omega|$. The sensitivity (or recall) between the predicted segmentation and ground-truth skeleton is introduced as $\mathrm{T}_{sens}(s,\widehat{y}) = |s \cap \widehat{y}|/|s|$. Likewise, the precision between the ground-truth mask y and the predicted skeleton $\widehat{s}$ is defined as: $\mathrm{T}_{prec}(\widehat{s},y) = |\widehat{s} \cap y|/|\widehat{s}|$. The clDice is defined as the F1-score between precision T_{prec} and sensitivity T_{sens} as follows:

$$\mathscr{L}_{cl\mathrm{Dice}} = 2\frac{\mathrm{T}_{prec}(\widehat{s},y)\,\mathrm{T}_{sens}(s,\widehat{y})}{\mathrm{T}_{prec}(\widehat{s},y) + \mathrm{T}_{sens}(s,\widehat{y})}. \tag{2.5}$$

Interestingly, for all three approaches (Clough et al., 2019; Hu et al., 2019c; Shit et al., 2019) results show the topological prior do not fully outperform the no-prior approaches, as measured with regional metrics such as Dice, but they increase specific topological metrics, such as the clDice accuracy in Shit et al. (2019).

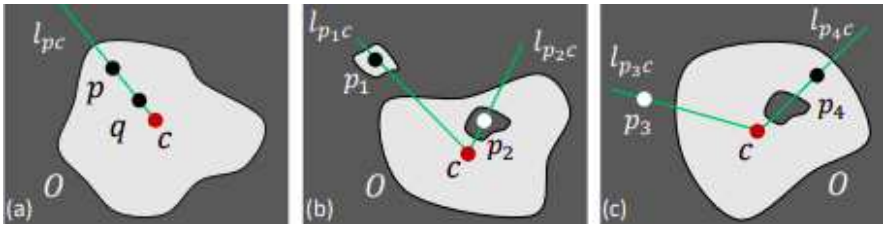


Figure 2.4 – Star shape prior. (a) Star shape object O w.r.t. the supplied object center c (red dot). (b) Example of star shape constraint violation. (c) Example when the star-shape prior loss is required. Figures from (Mirikharaji and Hamarneh, 2018).

2.3.3 Shape Constraints

There are numerous shape descriptors, such as geometric features, moments, shape transforms, or based on scale-space theory or polygonal approximation. These descriptors are usually computed on shapes which are represented by a set of point coordinates or with a binary map. The difficulty of integrating such descriptors into the loss terms stems from the fact that they must be computed on real-valued probability maps, i.e., the network output, instead of binary ones. Inspired by (Veksler, 2008b), Mirikharaji and Hamarneh (2018) propose a loss that forces the segmented region to have a star shape, for the task of segmenting dermoscopic skin lesions. An organ is said to have a star shape if, for any point p inside the object, all the pixels q lying on the straight-line segment connecting p to the object center c, are inside the object (Figure 2.4). Let l_{pc} be the line segment connecting pixel p to the object center c, and q be any pixel incident on line l_{pc}. The proposed loss is expressed as:

$$\mathscr{L}_{star} = \sum_{p\in\Omega}\sum_{q\in l_{pc}} \delta_{p,q}.\left|y_p - \widehat{y_p}\right|\left|\widehat{y_p} - \widehat{y_q}\right|, \tag{2.6}$$

where $\delta_{p,q}$ is the Kronecker symbol defined as:

$$\delta_{p,q} = \begin{cases} 1 & \text{if } y_p = y_q; \\ 0 & \text{otherwise.} \end{cases}$$

Star shape prior is a way to promote convexity for organ shapes. The star-shape loss registers significance improvement on segmentation performance, given a variety of networks such as U-Net and ResNet-DUC (Wang et al., 2018).

In Dolz et al. (2017), the idea is to enforce compactness. This property is represented by the ratio of the square of the perimeter to the shape area, that is required to be as small as possible. Here the boundary length requires to estimate a discrete form of this ratio functional, not with the usual continuous variable $\mathbf{y}$, but with a discrete binary vector $\widehat{\mathbf{z}} \in \{0,1\}^{|\Omega|}$:

$$\mathscr{L}_{compact} = \frac{\mathrm{P}(\widehat{\mathbf{z}})^2}{\mathrm{A}(\widehat{\mathbf{z}})}, \tag{2.7}$$

where $\mathrm{A}(\widehat{\mathbf{z}})$ and $\mathrm{P}(\widehat{\mathbf{z}})$ represent the predicted organ area and boundary respectively. The area is computed according to Equation 2.1 and boundary is proportional to the number

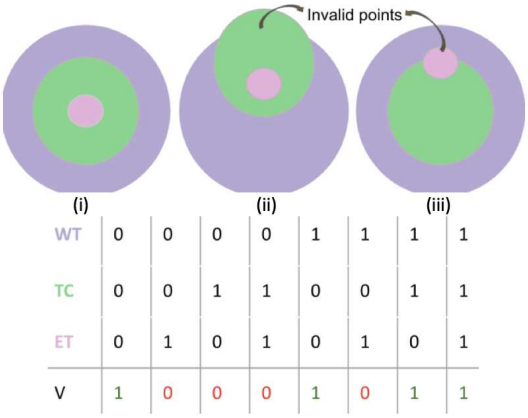


Figure 2.5 – Inter-region constraint prior. Three anatomical objects, (WT, violet), (TC, green) and (ET, pink), have $2^3 = 8$ possible combinations of existence. Given the correct anatomical topology specified in (i), the validity map V can then be derived for each of the 8 cases as shown in the last row of the table. Figure from (Reddy et al., 2019).

of neighboring pixels with different labels, and thus computed as:

$$\mathrm{P}(\widehat{\mathbf{z}}) \propto \sum_{p \in \Omega} \sum_{q \in \mathcal{N}_p} \left| \widehat{z}_p - \widehat{z_q} \right|.$$

The proposed loss is dimensionless, unbiased and position independent. However, due to the discrete nature of the prior involved, the optimization of this compactness-based loss comes with great challenges. For this reason, Dolz et al. (2017) alternate between optimizing the network parameters with SGD and optimizing the discretely-constrained segmentation labels, via ADMM.

2.3.4 Inter-regions Constraints

In the case of multi-label segmentation, specific interactions between regions, known a priori, can be authorized or forbidden: adjacency relations between organs are handled in (Ganaye et al., 2019), while BenTaieb and Hamarneh (2016) propose solutions to enforce regions exclusion and inclusion.

Focusing on gland histology images, BenTaieb and Hamarneh (2016) have identified that the cell and the object to be segmented, is made of two nested regions: one region (the lumen) is contained in another one (the epithelium). The authors integrated this spatial property by introducing a binary validity indicator map. A validity indicator map V returns 1 if a given label y_p corresponds to a topologically-valid assignment, and zero otherwise. In this way, the network not only penalizes incorrect label assignment per pixel, but also penalizes incorrect label hierarchy. Their loss term is based on a V-weighted cross-entropy, defined as follows:

$$\mathscr{L}_{inter} = -\sum_{p \in \Omega} \sum_{r=1}^{\mathrm{K}} y_p^r \log(\widehat{y}_p^r)\mathrm{V}, \tag{2.8}$$

where

$$\mathrm{V} = \begin{cases} 1 & \text{if topology is in accordance;} \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, their method exploits a boundary smoothness term that takes into consideration the difference between probabilities of pixels corresponding to the same labels, to the proposed loss. The two constraints are combined via penalty-based optimization. This approach has been applied to brain tumor segmentation in 3D MR images (See Figure 2.5) (Reddy et al., 2019). We note that the inclusion-exclusion loss (and the boundary smoothness loss) are not optimized alone, but are added to a fitting loss, which may be cross-entropy (BenTaieb and Hamarneh, 2016) or Dice (Reddy et al., 2019), and the total loss function, which is the sum of all three losses, is optimized through regular stochastic gradient descent.

Both publications mentioned above validate the importance of the proposed prior loss in different tasks and modalities. However, the use of the prior loss does not compensate for the need of fitting losses such as Dice or cross-entropy. The presence of such losses is generally necessary for the convergence of the segmentation framework. Moreover, the method still depends on penalty-based optimization to balance out the two constraints, which does not accommodate the interplay and relations between the different constraints (Kervadec et al., 2020a).

In (Ganaye et al., 2019), the authors propose a loss that takes into consideration the relationships between neighboring anatomical objects. From the ground-truth label maps, one can define an adjacency matrix $\mathbf{A}$ of general binary term A_{ij} between regions, that represents whether two regions can be adjacent or not. Then the set of forbidden label connections can be defined as follows: $\mathrm{F} = \{(i, j) | \mathrm{A}_{ij} = 0\}$.

However, an adjacency measure computed from the outputs of the CNN, which are probability maps and not label maps, is more difficult to define. When two regions i and j should not be connected, i.e., $(i, j) \in \mathrm{F}$, then the probability for a pixel and its neighbors to belong to i and j must be close to zero. Let $\widehat{y}_p(i)$ (resp. $\widehat{y}_q(j)$) be the probability of pixel p (resp. q) to belong to class i (resp. j), Ganaye et al. (2019) propose to model the constraint by the product $\widehat{y}_p(i)\widehat{y_q}(j)$. The adjacency measure is then:

$$a_{ij} = \sum_{p \in \Omega} \sum_{q \in \mathcal{N}_p} \widehat{y}_p(i) \widehat{y}_q(j). \tag{2.9}$$

Following this, the proposed loss consists of forcing all the forbidden adjacency relationships, with respect to the relations defined in the set of impossible transitions, F:

$$\mathscr{L}_{adj} = \sum_{(i,j) \in \mathrm{F}} a_{ij}. \tag{2.10}$$

The authors then solve the constrained optimization problem via the penalty-based method mentioned in Section 1.3.1. In the experiments, datasets with a number of regions ranging from 20 to 135 are investigated. Interestingly, as the shape and size constraints, the proposed adjacency loss does not require the ground-truth segmentation – it just requires the definition F, thus the segmentation method can be evaluated in a semi-supervised framework. Model performance in terms of the Dice metric shows no significant improvement, when compared to the baseline method (i.e., the same segmentation method without the

proposed loss term). However, incorporating adjacency constraints registers considerable improvement with regards to the distance metrics (Hausdorff and Mean Distance Metric). These results are consistent over all datasets (in 2D and 3D) and settings, i.e., with full or semi supervision. Depending on the applications, one limitation of the approach may be the assumption that all patients have the same inter-organ connectivity.

2.4 Discussion

In addition to the common challenges in deep network training, such as overfitting, scarcity of annotated data, class imbalance, and gradient divergence, which are extensively discussed in (Havaei et al., 2016; Hesamian et al., 2019; Litjens et al., 2017) for example, there are particular challenges when dealing with a prior-based term and its incorporation into the loss function. In this section, we summarize and highlight a number of key aspects of embedding a prior-based loss function into a segmentation network.

2.4.1 The Nature of the Prior

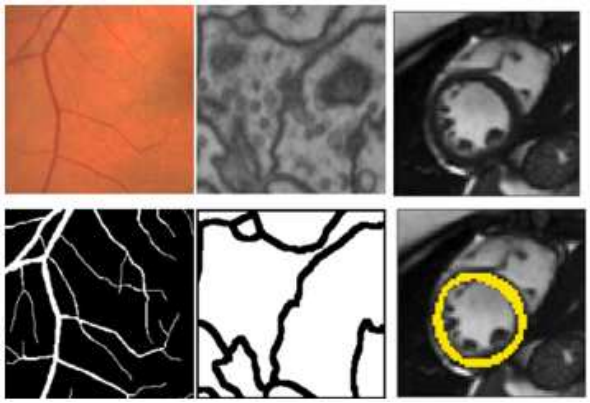
The high-level prior, as we defined it, expresses high-level features regarding the object of interest, with interpretable insight with respect to the organ geometry or anatomy. This prior can stem from medical knowledge (e.g. organ size range, organ connectivity) and as such, can be used in a weakly and semi-supervised learning context to improve performance. Sometimes, the prior has to rely on features extracted from the ground-truth label maps, see for example the Betti numbers or the skeletonization process. In this case, its usage is restricted to full supervision.

2.4.2 The Challenge of Soft Probability Maps

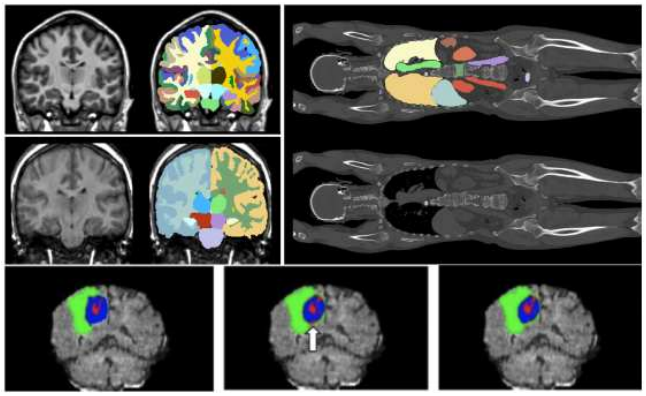
One major challenge is to compute features from soft probability maps. A binary map expressing the object shape is much easier to characterize with usual shape features (e.g. circularity, compactness, isoperimetric ratio, skeleton). However, thresholding the probability map to make it binary can render the loss function non-differentiable. Some parameters can be estimated from probability maps, e.g. the predicted organ size in (Kervadec et al., 2018). Other features require resorting to a discrete optimization scheme, such as the predicted organ boundary length (Dolz et al., 2017) (see Section 1.3.4). Another way of dealing with the soft probability maps is to impose a series of thresholds, to monitor topological changes (Clough et al., 2019; Hu et al., 2019a). However, their method is not generic and cannot be applied to all prior properties. These issues become more complicated as the type of handled prior becomes more complex, and loss functions often end up being non-convex or hard to optimize.

2.4.3 Relationship between Organs and Loss Design

The studied papers address various segmentation problems, as listed in Table 2.2 that shows the targeted organs or objects to be segmented, and the datasets used in each paper. Size and shape constraints mostly concern single instance organs that have convex



(a) Topological priors in retinal vessels (left), neuronal membrane (middle), and myocardium of the left ventricle (right). Top: Images, bottom: ground-truth.



(b) Inter-region prior: adjacency constraint in brain and full body regions (top) and inclusion-exclusion relationships (bottom).

Figure 2.6 – Examples of targeted segmentation objects that can benefit from topological priors (a), inter-region priors (b). Figures from (Clough et al., 2019; Ganaye et al., 2019; Hu et al., 2019a; Reddy et al., 2019).

or a blob shape, such as the prostate, the cardiac ventricles, the aorta, the esophagus, or skin lesions (see for example the Promise[1], ACDC and ISIC[2] datasets). Topology priors are mostly used for thin, curvilinear objects such as neuron membranes, vessels or, at a higher scale, the myocardium of the left ventricle, which has a ring shape (Figure 5.3a). Inter-regions priors can help in problems of multiclass segmentation. The adjacency constraints have shown efficiency in full body segmentation (120 regions in Anatomy3[3] dataset), and for multiple regions segmentation in brain segmentation (20 and 35 regions MICCAI12 dataset). Exclusion and inclusion priors are helpful whenever there is a hierarchy in region membership. Their use is illustrated in applications, both at the microscopic and macroscopic levels. In microscopy Images, cells can be composed of several layers: for example, gland cells (found in the GlaS dataset) are made of the inside region, called the glandular lumen, and the outer region, identified as the epithelial boundary. In the BraTs[4] dataset, brain tumors in MRI are made of an enhancing tumor (the deepest level) that is surrounded by the tumor core, itself surrounded by a region identified as the whole tumor (Figure 5.3c).

[1]https://promise12.grand-challenge.org/
[2]https://www.isic-archive.com/!/topWithHeader/wideContentTop/main
[3]https://www.smir.ch/VISCERAL/Start
[4]https://www.med.upenn.edu/sbia/brats2018/data.html

Table 2.2 – Targeted segmentation objects and datasets used in the presented papers. UKb: UK Biobank. MIC12: MICCAI 2012 workshop on multi-atlas labeling. Anat3: Anatomy3. p: proprietary dataset.

		Size			Topology					Shape		Inter-regions		
		(Peng et al., 2020)	(Kervadec et al., 2019b)	(Kervadec et al., 2018)	(Hu et al., 2021)	(Byrne et al., 2021)	(Clough et al., 2019)	(Hu et al., 2019c)	(Shit et al., 2019)	(Mirikharaji and Hamarneh, 2018)	(Dolz et al., 2017)	(BenTaieb and Hamarneh, 2016)	(Ganaye et al., 2019)	(Reddy et al., 2019)
CT	Full body												Anat3	
	Aorta/eso										p			
MRI	Brain												MIC12	BraTs
	Prostate	Promise	Promise	Promise										
	Cardiac	ACDC		ACDC		ACDC	UKb				p			
Photo	Skin lesion									ISIC				
Microscopy					ISBI12 ISBI13 DRIVE			ISBI12 ISBI13 DRIVE	DRIVE			GlaS		

2.5 Conclusion

In this chapter, we presented a survey of the current state-of-the-art methods regarding high-level prior-based losses, for medical image segmentation. We have proposed a categorization where we grouped these methods according to the type of prior that they incorporate as well as, the type of features they optimized, the architecture they use, optimization strategy and the anatomical object that they target.

In the next part of the thesis, we will demonstrate two main contributions that we introduce to the field of prior-constrained convolutional neural networks and that are conducted at the level of the network architecture.

Part I

Integration of Prior Constraints at Network Structure

Chapter 3

BB-UNet: U-Net with Bounding Box Prior for Medical Image Segmentation

" You have to trust your intuition because if you just let them go, you will be losing an opportunity... "

Yann Le Cun

Contents

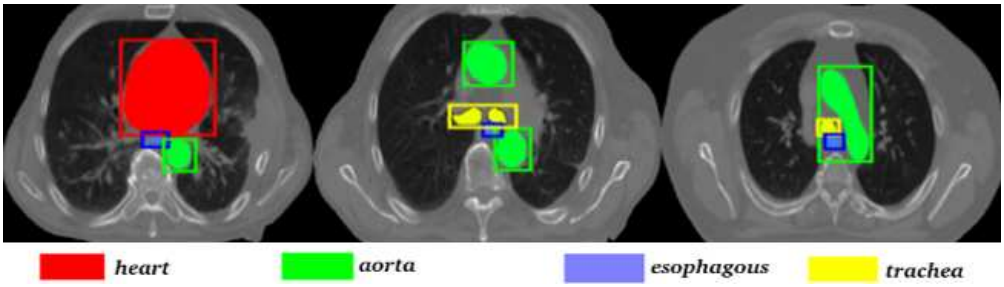


Figure 3.1 – CT images from the SegTHOR dataset with manual segmentation and bounding boxes overlaid on multiple organs (Lambert et al., 2020).

3.1 Motivation

One important type of prior is the bounding box that encompasses an understudied organ. A Bounding box prior can be seen as constraints revealing estimated location areas indicating where in the image an organ is present. Moreover, it gives an idea about the estimated size and whether an organ is found or not within an image. Hence, such prior, if properly exploited, can present valuable information that increases system performance and alternatively compensate for the lack of a large fully annotated dataset.

In this chapter, we integrate a bounding box prior into the network structure in order to improve the performance of segmentation models for medical data in both weakly and fully supervised settings. To do so, we propose a novel deep learning model inspired by U-Net, that integrates bounding box information in-between local and global features. We call the proposed model BB-UNet (Bounding Box U-Net) since it incorporates a bounding box prior via a novel convolutional layer introduced at the level of skip connections. In doing so, it allows to take advantage of positional and shape features as means of guiding the neural network to find consistent organ contours and to distinguish between the different connected components of an organ. Consequently, BB-UNet helps in presenting positional constraints onto the neural training within model structure in order to guide the model on where to look for the organs and fine-tunes the encoder layers accordingly.

The proposed model is incorporated in two frameworks. The first framework is a fully supervised framework in order to increase anatomical plausibility. To this end, the proposed model is validated on two segmentation tasks. The first task is multi-label segmentation in computed tomography (CT) imaging where experiments are conducted on the SegTHOR dataset (See Section 1.5.1). The objective is to segment four organs of interest: the heart, aorta, esophagus and trachea given that they are of variable shapes and share the same gray-scale intensity values with neighboring tissues, as shown in Figure 3.1. The second problem is segmenting a single-label multi-component complex organ in magnetic resonance imaging (MRI) scans where experiments are conducted on the Atrium dataset from the Decathlon challenge (See Section 1.5.2). The main intuition behind using this dataset is to investigate the ability of BB-UNet in capturing topological properties of the dataset, such as the organ's shape or its composition of connected components.

In addition to its role in improving anatomical plausibility given fully annotated datasets, we investigate BB-UNet's significance in compensating for the need for large annotated datasets usually required to insure the generalization ability of deep learning models. In

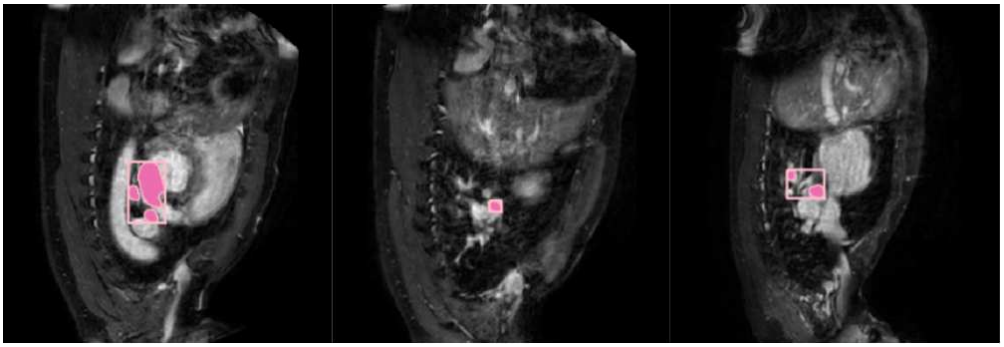


Figure 3.2 – MRI images from the Atrium dataset with manual segmentation and bounding boxes segmentation overlaid on the left atrium (Simpson et al., 2019).

order to do so, we make use of "not-so-accurate" labels (i.e., bounding boxes), thus embracing BB-UNet within a weakly supervised learning paradigm. Consequently, we validate the proposed semi-weakly supervised method on segmentation of the heart organ in the SegTHOR dataset.

The main contributions of this chapter can be summarized as follows:

1. We propose a novel deep learning model that integrates bounding box constraints into the network architecture in order to overcome segmentation ambiguities.

2. We shed light on the role of embedding prior knowledge onto model training relative to data augmentation and post-processing alternatives given fully supervised segmentation frameworks.

3. We incorporate the proposed novel model in a semi-weakly supervised framework where only bounding box tags are present.

The rest of this chapter is organized as follows. Section 3.2 elaborates on the proposed BB-UNet model as well and its different paradigms. Section 3.3 demonstrates the significance of the proposed model given a fully supervised framework. Section 3.4 evaluates the robustness of BB-UNet performances across bounding box size variations and dataset distributions. In Section 3.5, we show that BB-UNet exhibits promising behavior with regards to learning under weak labels. Finally, Section 3.6 concludes with a summary of the chapter.

3.2 BB-UNet: U-Net with Bounding Box Prior

In this section, we describe the BB-UNet model. We first elaborate on its different building blocks (Section 3.2.1), the main principle (Section 3.2.2), and then clarify the different prior information used in Section 3.2.3. Finally, a comparison is presented in Section 3.3.1 between the proposed BB-UNet and related state-of-the-art models.

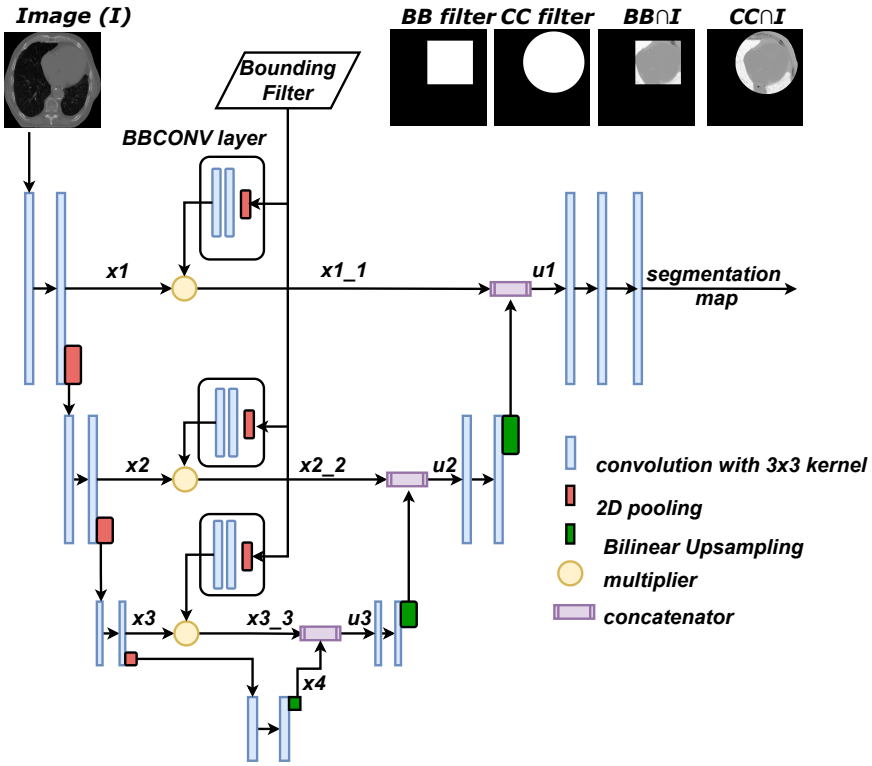


Figure 3.3 – BB-UNet structure with a bounding filter inserted at the BB-ConV layers. Four possible bounding filters are illustrated: BB: Bounding box filter, CC: Circular filters, BB∩I: intersection between bounding box and image, CC∩I: intersection between circular filter and image.

3.2.1 BB-UNet Architecture

In the proposed architecture, we extend U-Net to include not just global and local features, but also ones related to position as well as shape priors. As previously stated in Chapter 1, U-Net is a symmetric encoder/decoder structure with equivalent distribution of convolutional and de-convolutional layers. In order to make use of local and global information, U-Net exploits skip connections that concatenate down-sampling features from the contracting path with up-sampling ones from the expanding path. Our main contribution within the BB-UNet structure lies at skip connection levels. Thus, instead of directly concatenating the features from both paths as in U-Net, a third component layer that takes into consideration the bounding box prior is introduced. This layer is called the BB-ConV layer as shown in Figure 3.4. A BB-ConV layer is composed of a 2D-max pooling layer followed by two consecutive convolutional ones. The input to this layer is a bounding map (i.e., bounding filter) representing a coarse-grained area where the organs are supposedly located. In summary, the BB-ConV layer takes as input a bounding filter and outputs a feature map that allows the network to enhance its estimation to where an organ can be (See Figure 3.4). This learnt feature map can be considered as a per-pixel weighting factor, enhancing discriminative features over non-significant ones. Overall, BB-UNet is taking two inputs, the medical image and bounding filter. Since the filters are inserted at the link between contextual and location information, we are able to adapt

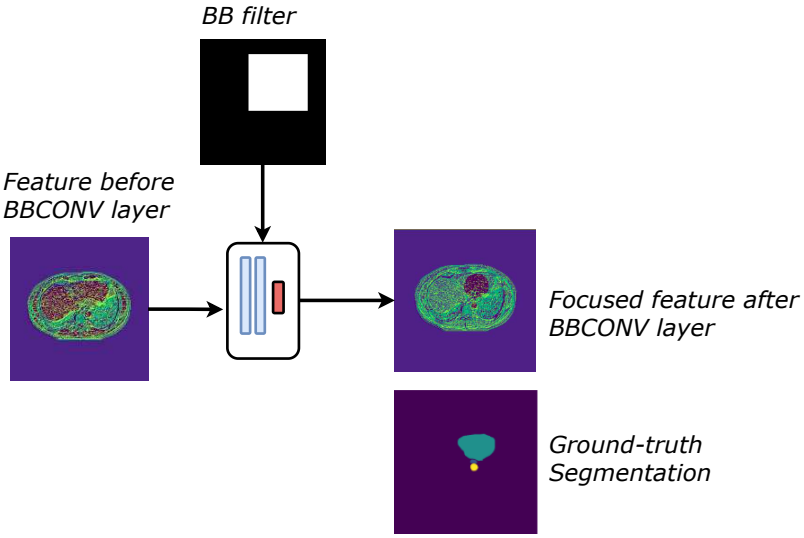


Figure 3.4 – Principle of the BB-ConV layer. We realize that the addition of the BB-ConV layer allowed extraction of focused features around the organ location.

what the model learns, focusing on the attention areas that we are yielding, i.e., enhancing features detection concerning shape and topology within particular sections of the image.

3.2.2 BB-UNet Main Principle

The proposed BB-UNet considers two inputs, the image and the bounding filter. Whereas the image is fed to the encoder layers in the contracting path for contextual feature extraction as is done within a regular U-Net, the bounding filter is fed independently to the BB-ConV layer for shape and location feature extraction. Within each skip connection, the intersection between the unpooled map from a level contracting layer and the location feature map from the BB-ConV layer is then obtained, and further concatenated with the features from the up-sampling layers. The bounding filter provided to the BB-ConV layer is a binary map indicating the attention area corresponding to the position of the organ(s) under consideration. For single-organ segmentation, a single channel indicating the possible area where the organ may be located is provided to the BB-ConV layer. Thus, in the case of multi-component organs, a unified bounding box encompassing all organ elements is taken into consideration. For the multi-organ segmentation, separate channels are fed to the BB-ConV layer relative to each organ independently. This multi-channel map is then convolved within the BB-ConV layers for feature extraction. The output of BB-UNet is a segmentation mask derived from learnt relations between the bounding filters as well as the image. With respect to the BB-UNet output, we distinguish generally between the classes under consideration and the background. For a single-organ segmentation task, the BB-UNet output is two channels corresponding to the segmented organ and the background. For a multi-organ segmentation case, the output is $m+1$ channels corresponding to the m organs to be segmented as well as the background.

3.2.3 Prior Information

We have designed several types of bounding filters. We have considered rectangular bounding boxes denoted as **BB** for the bounding box filter. Noting that the objects of interest (i.e., organs) do not have corners or edges, we also considered a circular filter, denoted as **CC**. The **CC** filters are the smallest circles encompassing the bounding boxes. Since organs may share similar intensity values, filters that include the intersection between the bounding mask and the input image (I). Henceforth, the **BB**$\cap$**I** and **CC**$\cap$**I** filters are introduced in Figure 3.4(top right).

3.2.4 Generating Bounding Boxes

Bounding boxes can be generated through 2D or 3D approaches. In 2D, bounding boxes can be generated either through region proposal approaches (He et al., 2017) or through regression/classification based approaches (Liu et al., 2016). 3D-bounding box generation can be done by training an end-to-end convolutional network for 3D-object extraction (Lu et al., 2019), or by extrapolation from 2D bounding box generation techniques as surveyed in (Jiao et al., 2019). In our implementation, we have considered the case where bounding boxes were generated automatically from the ground-truth and defined as the smallest bounding box encompassing the understudied organ. Despite the fact that we have used manually-obtained bounding boxes for both our training and inference tasks, automatically obtaining coarse grained bounding areas could be considered given current object detection advances in medical imaging (Araújo et al., 2018). In this case, the BB-UNet performance will depend on the shift between detected vs ground-truth bounding box distributions. Ideally in the automatic framework, detected bounding boxes would have to be used for training and testing.

3.3 BB-UNet for Segmentation under Full Supervision

In order to study how well the proposed model performs as a standalone structure, the BB-UNet is firstly trained within a fully supervised framework. Labels used are the ground-truth segmentations provided for each organ respectively. End-to-end training is conducted, guided by the Dice loss function.

3.3.1 Connections to Fully Supervised State-of-the-art Methods

Models most relevant to our work are those developed by Ravishankar et al. (2017) and Oktay et al. (2018b). The SR-UNet in Ravishankar et al. (2017) introduces regularization factors by jointly adding an external network to the U-Net model. The main objective of this network is to take into consideration the incomplete, over- or under-segmented shape masks provided by the U-Net and map it to a manifold of training shapes. Despite the importance of regularization, the addition of an adjoint complex model while training increases model complexity and thus affects the model performance. For this reason, we aim in this work, to add the regularization structure within the same model, denoted as

Table 3.1 – Slice and Size distribution of the SegTHOR dataset.

	Train	Validation	Evaluation	Organ Size (pixels)		
Patients	36	4	20	Average	Min	Max
Heart	1444	155	726	9574	245	23588
Aorta	3363	391	1824	1023	81	6336
Trachea	1767	220	953	340	72	1244
Esophagus	3510	410	1862	226	60	2528
Total	4153	497	2281			

the BB-ConV layer, thus fine-tuning the encoder spatial considerations internally without the need for a manifold space or an external network.

The Attention-UNet in (Oktay et al., 2018b) is most similar to ours in terms of adding attention blocks at the skip-connection level. Thus, both BB-UNet and Attention-UNet focus on imposing convolutional filters midway between encoder and decoder paths. In doing so, both models are able to distinguish between relevant and irrelevant features while training. However, the means to which each model obtains the constraints to these attention maps differ considerably. Whereas Attention-UNet aims at exploiting coarse-grained features obtained from U-Net bottleneck as input constraints to the convolutional layers at skip connection levels, our model imposes external activation inputs based on prior knowledge of the dataset. One can think of both models as functioning in different directions within the skip connections. Whereas BB-UNet imposes external activation inputs to guide the network on to where to look and move downwards through the network until bottleneck, Attention-UNet exploits inputs provided by the bottleneck output and moving upward through the skip connections. Generally, obtaining bounding boxes is a relatively easy task as discussed in Section 3.2.4

3.3.2 Datasets

SegTHOR dataset: Following the SegTHOR challenge, training considered 40 patients subdivided into a training set of 36 patients and a validation set of 4 patients. The test set includes the remaining 20 patients. The patient and slice distributions are shown in Table 3.1. The average, minimum and maximum organ sizes are also provided, on average over the entire set of patients.

Using body contours provided for each image slice, images were cropped and resized to a resolution of 512×512 pixels. Image intensities were bound to values between -1000 and 3000, normalized by subtraction of the mean and division by standard deviation at image level. Slices were filtered to keep only images with at least one organ present.

Taking a closer look at the organ slice frequencies in Table 3.5, one can notice that the slice organ distribution suffers from high class imbalance. Thus, the heart as well as the trachea have a small number of slices (≈ 1000 slices) compared to the aorta and esophagus (≈ 3000 slices). Taking a closer look at the organ size, the trachea and esophagus are the smallest in size (200 to 340 pixels in average) relative to the heart and aorta (≈ 10000 pixels in average). Since target organs are very close to each other, bounding filters used at the BB-ConV layer suffer from a high overlap, as shown in Figure 3.1.

Atrium Dataset: The Atrium dataset consists of 20 mono-modal MRI patient scans divided into 10 patients (670±5 slices) for training, 4 patients (66±1 slices) for validation and 6 patients (416 slices) for testing (See Section 1.5.2). MR image slices are kept at 320×320 pixels and normalized by subtraction of the mean and division by standard deviation at image level. Overall, the dataset is characterized by being small, i.e., 1351 slices in total, with large variability relative to the atrium size.

Aside from the low contrast that these images are characterized with, the segmentation process is particularly challenging due to many factors, among which the high organ size imbalance over the different slices. Thus, the atrium has a large size range that varies from 3 to 1921 pixels with up to 3 connected components. That means that, whereas some slices have a considerably large atrium, others contain a segment which is very small. Moreover, the atrium is composed of multiple components within the bounding box. Therefore, the model must learn to distinguish between the different parts of the atrium present in the same bounding box provided to the network.

3.3.3 Model Training and Architecture.

The BB-UNet has two main components: The base U-Net model and the BB-ConV layer. The U-Net implementation in this work is the one provided by a PyTorch implementation of the original U-Net (Perone et al., 2018). Feature dimensions extend till 256 feature maps within the bottleneck that is composed of 2 simple convolutional layers. The BB-ConV is composed of a 2D-pooling layer followed by two consecutive convolutions with batch normalization (momentum = 0.1) and dropout (factor = 0.4). Bounding boxes relative to different organs are fed independently through a multi-channel input onto the BB-ConV layer. Moreover, to overcome the size imbalance between organs versus the background, we distinguish between classes corresponding to the understudied organs (4 organs for SegTHOR, 1 organ for atrium) and the background class. To guide the training, a loss approximation of the Dice similarity factor as elaborated in (Milletari et al., 2016) was adopted. Moreover, we use the Adam optimizer with an initial learning rate of 10^{-3} and a cosign annealing scheduler.

3.3.4 SegTHOR Multi-Organ Segmentation

In this section, we present fully supervised training results for both SegTHOR and Atrium datasets. SegTHOR results are compared relative to 3 fully supervised segmentation models: the original U-Net (Ronneberger et al., 2015), U-Net with data augmentation (Lambert et al., 2020), and VB-Net that is the winner of the ISBI SegTHOR challenge (Han et al., 2019). VB-Net is a V-Net variant that exploits a multi-resolution strategy to robustly localize organs while reducing the GPU memory cost. Since bounding box prior is required at test time in the proposed model, it is logical to impose such prior on the reference model (U-Net) as well, and compare the obtained results to ours. For this reason, a post-processing step is conducted on U-Net performance by filtering predicted segments with a bounding box prior. Filtering via bounding boxes is conducted by multiplying the predicted segmentations with the bounding box map. In this way, any structure or noise segmented outside of the bounding box area where the organ is located, gets removed.

Table 3.2 – Average Dice ratio for fully supervised multi-organ segmentation of the SegTHOR dataset. First rows represent non-filtered model performance and last rows show results after post processing using bounding boxes

	Heart	Aorta	Trachea	Esophagus
State of the Art				
VB-Net (Han et al., 2019)	94.46	93.88	91.29	84.02
U-Net (+ data augmentation) (Lambert et al., 2020)	93.00	92.00	86.00	81.00
Baselines				
U-Net (+ background) (Ronneberger et al., 2015)	50.37 ± 15.28	85.17 ± 2.21	82.48 ± 5.27	76.56 ± 0.10
U-Net (+ background) + Post.	96.89 ± 1.38	93.20 ± 0.62	97.59 ± 0.24	84.76 ± 1.09
Proposed Models				
BB-UNet-BB	**98.32 ± 0.29**	**96.02 ± 0.46**	**97.82 ± 0.10**	**91.56 ± 0.12**
BB-UNet-BB + Post.	98.59 ± 0.10	96.02 ± 0.46	97.82 ± 0.10	91.56 ± 0.12
BB-UNet-BB ∩ I	97.57 ± 1.52	95.95 ± 0.31	**97.82 ± 0.34**	**91.74 ± 0.17**
BB-UNet-BB ∩ *I* + Post.	98.63 ± 0.10	95.95 ± 0.31	97.97 ± 0.30	91.74 ± 0.17
BB-UNet-CC	93.07 ± 5.80	95.54 ± 0.10	93.526 ± 0.17	90.01 ± 0.28
BB-UNet-CC + Post.	98.07 ± 0.27	95.77 ± 0.23	97.74 ± 0.25	90.29 ± 0.31
BB-UNet-CC ∩I	82.89 ± 4.80	95.30 ± 0.25	93.41 ± 0.24	89.79 ± 0.34
BB-UNet-CC ∩ I + Post.	98.10 ± 0.06	95.54 ± 0.26	96.65 ± 0.14	90.07 ± 0.37

We denote this as U-Net + Post.

Dice Accuracy for the SegTHOR dataset is benchmarked in Table 3.2 respectively. From the results in Table 3.2, we observe that the proposed model outperforms regular U-Net (3rd row) by about 15 % on the esophagus and trachea, 11 % for the aorta, and by a large margin on the heart. This indicates the ability of the proposed model to learn discriminative features both specific for the organs at hand and also relative to their location. With respect to previous leading work that utilized data augmentation (1st and 2nd row), BB-UNet admits comparable results with respect to both VB-Net, the ISBI challenge winner and U-Net with data augmentation. In fact, model performance between them varies mildly with BB-UNet taking the lead for the esophagus, the trachea, and the aorta given all proposed experiments and in 3 out of 4 experiments in the case of the heart. These results shed light on the role of prior embedded structures in obscuring the need for data augmentation.

In comparison to the state-of-the-art models, bounding boxes within U-Net are required in both training and inference phase. By filtering the segmentations using the bounding boxes within U-Net, we obtain the post-processed results in row "U-Net (+background) + Post" from Table 3.2. Comparing these results with ours, we realize that indeed the proposed models provide comparable results (2% to 3% higher) than the post-processed U-Net segmentations, while outperforming the "U-Net + Post" in the case of the esophagus. This indicates the importance of properly integrating prior knowledge onto the model structure while training.

By comparing the different proposed structures in Table 3.5, we gather that the BB-UNet with bounding box filters, regardless of whether it is solely the bounding box or the bounding box intersected with the image, perform better than circular ones contrary to the binary case. This is due to the fact that circular filters have larger shared areas gen-

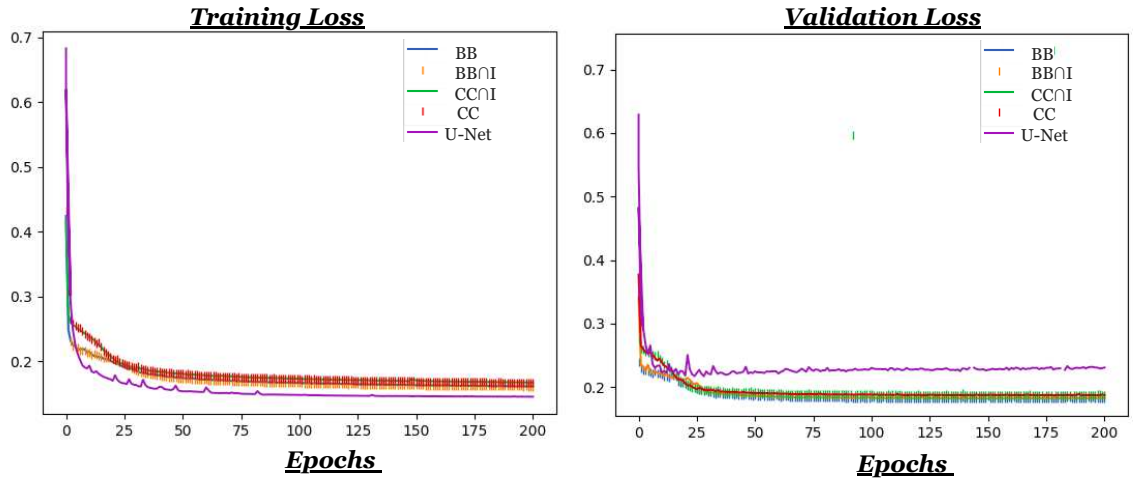


Figure 3.5 – Evolution curve of the training and validation loss on the proposed models and U-Net when trained on the SegTHOR dataset. Same legend as Figure 3.4

erated due to the closeness of organs with respect to each other. Thus, attention areas are more overlapping in the case of circular filters than that of bounding boxes. The variation between model performance given bounding box filters (BB and BB∩I) relative to circular filters (CC and CC∩I) is significant to note, since it opens up the discussion of the dependency of models' performance relative to the approximate area where the organ is estimated to exist in. For example, given segmentation of the aorta which, according to Table 3.1, is the second largest organ (1023 pixels) and the second most common organ present in slice distribution (3363 slices), the variation in the approximated prior area has no effect on model performance which is stable at about 0.96.

Taking a closer look at the evolution of the training and validation Dice error over the number of epochs, we derive the role of prior with respect to the model convergence as given in Figure 3.5. It is evident that BB-UNet models tend to converge onto lower losses than regular U-Net during validation, that seems to be losing its generalization ability or sustaining its limited performance with epoch evolution.

3.3.5 Cardiac Multi-component Organ Segmentation

With regards to Atrium dataset, we compare to regular U-Net as well as the U-Net + Post-processing step. Since the Atrium dataset is composed of multi-connected components, we were also interested in examining how well is the proposed model able to account for the number of connected components the Atrium structure is composed of. To this end, we calculate the mean absolute error between the ground-truth number of connected components and the predicted number of connected components. We present the results in the bar-graph in Figure 3.6.

Results with respect to the Atrium are benchmarked in Table 3.3. Given a closer look, we realize that the utilization of BB-UNet has a direct effect on segmentation quality. Thus, BB-UNet in all its modalities outperform regular U-Net by over 12 % with BB ∩ I model registering highest Dice accuracy scores relative to its peers. The significance of

Table 3.3 – Model Performance for fully supervised organ segmentation of the atrium: Av. Dice: average Dice Accuracy, Av. Hausdorff: average Hausdorff distance, Conn. Comp.: average absolute error between predicted and ground-truth connected components

	Av. Dice (%)	Av. Hausdorff (mm)	Conn. Comp.
State of the Art			
U-Net + (background)	77.95 ± 1.38	2.51 ± 0.11	0.43 ± 0.01
U-Net + (background) + Post	82.59 ± 0.61	2.33 ± 0.04	0.47 ± 0.01
Proposed Models			
BB-UNet-BB	89.94 ± 1.54	2.17 ± 0.13	0.21 ± 0.0397
BB-UNet-BB +Post	89.94 ± 1.26		
BB-UNet-BB∩I	**91.83 ± 0.22**	**2.05 ± 0.01**	**0.13 ± 0.00**
BB-UNet-BB∩I +Post	89.40 ± 0.78		
BB-UNet-CC	88.97 ± 0.35	2.10 ± 0.01	0.241 ± 0.127
BB-UNet-CC +Post	89.59 ± 0.18		
BB-UNet-CC∩I	88.82 ± 0.26	2.10 ± 0.02	0.165 ± 0.004
BB-UNet-CC∩I +Post	89.26 ± 0.28		

BB-UNet is relatively evident when comparing with respect to U-Net post-processing results (2nd row). Thus, even when we integrate bounding box filters at inference time onto U-Net predictions, BB-UNet still has a leading increase in about 6 % in Dice accuracy. This is also verified by the Hausdorff distances of U-Net and U-Net with post-processing relative to the proposed models. Thus, BB-UNet models register a decrease in Hausdorff distances relative to regular U-Net by about 13.54 % in the worst case scenarios (BB-UNet-BB: 2.51 → 2.17) and by about 18 % in the best case scenarios (BB-UNet-BB ∩ I: 2.51 → 2.05). After post-processing with bounding boxes onto U-Net segments, the proposed models register a decrease in Hausdorff distance by about 7 % in worst case scenarios (BB-UNet-BB: 2.33 → 2.17) and about 12 % in best case scenarios BB-UNet-BB: 2.33 → 2.05. The above results re-signify the importance of the BB-ConV layer and bounding box prior in preserving the shape of the target anatomical structures given segmentation problems. In addition to preserving location and shape attributes, the proposed BB-UNet also plays a role in maintaining topological properties, such as the number of connected components constituting the entire atrium. A closer look at the 3rd column of Table 3.3, we realize that the exploitation of the BB-ConV layer has allowed the network to take into consideration topological characteristics of the dataset not evident within the U-Net baseline model (regular or postprocessed).

Comparing the proposed models in Table 3.3 and Figure 3.6, we realize that despite the fact that the change is slight, the utilization of intersection filters with bounding boxes (BB-UNet-BB ∩ I) is slightly more beneficial. This is explained by the fact that with the atrium organ, segments are often composed of multiple components that vary in size and position. Since the bounding box utilized is a unified box that includes all the organ components, the utilization of the intersection filters allows the model to distinguish between the different components while still posing attention on the particular area where the

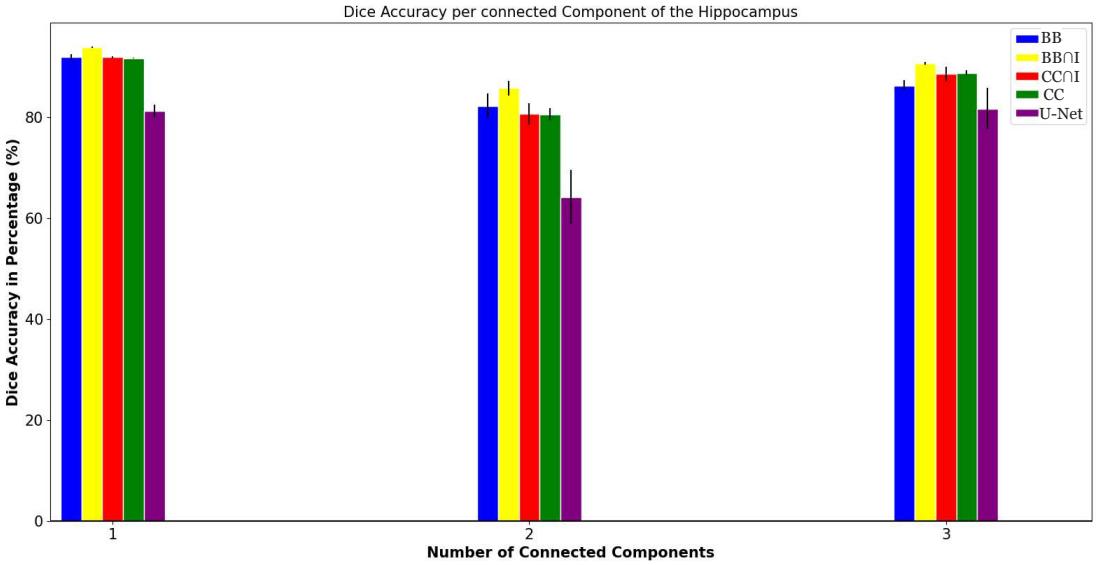


Figure 3.6 – Model performance for the Atrium Dataset per number of connected components (1,2 or 3)

components can be.

3.4 Robustness Assessment of BB-UNet Performance

From the previous results, we have shown that the utilization of the BB-ConV layer at the level of the skip connections allows the network to learn intrinsic properties relative to the target organs' location, shape, and topological characteristics. In this section, we assess the robustness of the BB-UNet performance through two steps. Firstly, we study the invariance of BB-UNet performance given its different modalities when imposing Bounding Box Filtering. Secondly, we conduct a sensitivity analysis with regards to the effect of varying the bounding box size on model performances.

3.4.1 Post-Processing Comparison via Bounding Box Filtering

In this experiment, we apply the same post-processing step that we did with U-Net in the previous section onto the different BB-UNet modalities. The main objective is to determine whether the BB-ConV layer eliminates the need for post-processing. Results are benchmarked relative to both SegTHOR and Atrium in Table 3.2 and Table 3.3 under the name (Model +Post.). An ideal case would be a zero gap between the BB-UNet model performance vs BB-UNet +Post. This would mean that imposing bounding box post-processing will not affect the BB-UNet performances across its different modalities. From Table 3.3, we gather that post-processing indeed resulted in little to no variation in Dice accuracy for the Atrium dataset. This means that even given the variation of the shape/type of utilized filter, the BB-UNet still maintained its agreeable performance and does not require Post-processing. With regards to SegTHOR (see Table 3.2), the same con-

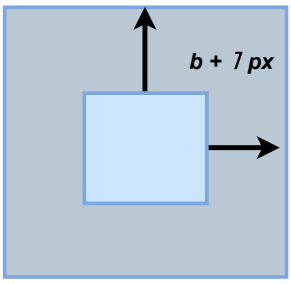


Figure 3.7 – Bounding box variation according to $\pm l\,px$ pixel: an increase in l pixels from each side of the bounding boxes

Table 3.4 – Effect of bounding box size variation on Dice accuracy. Each side of the bounding box is increased by l pixels, as shown in column **BB Variation**. The new bounding box area is $\times h$ greater than the initial bounding box as indicated in **BB Area Increase**. The resulting new Dice accuracy in % is in column **Dice Accuracy**. Comparison is conducted relative to a lower and upper baseline corresponding to UNet without post-processing, column **UNet** and BB-UNet performance.

Organ	UNet	BB Variation	BB Area Increase	Dice Accuracy	BB-UNet
Heart	67.89	$bb+50px$	× 4.11	72.01	98.13
		$bb+10px$	× 1.44	94.05	
		$bb+5px$	× 1.20	98.32	
Aorta	87.60	$bb+10px$	× 2.67	92.69	95.78
		$bb+5px$	× 1.72	94.58	
		$bb+1px$	× 1.12	96.74	
Trachea	87.42	$bb+10px$	× 4.05	95.68	97.82
		$bb+5$ px	× 2.27	97.68	
		$bb+1px$	× 1.21	98.07	
Esophagus	76.47	$bb+10px$	× 5.29	60.10	91.65
		$bb+5px$	× 2.73	78.97	
		$bb+1px$	× 1.27	92.91	
Atrium	78.67	$bb+10px$	× 4.50	57.98	88.42
		$bb+5px$	× 2.30	75.46	
		$bb+1px$	× 1.20	86.29	

clusion can be drawn when comparing relative to the aorta and the esophagus. Thus for both organs, the gap between BB-UNet vs BB-UNet + Post. is almost null. This is not the case for the heart and trachea. Thus, a considerable gap (BB-UNet-CC∩I: 16 %, BB-UNet-CC: 6 %) is registered relative to the heart and about 3 to 4 % for both models relative to the trachea. Going back to Table 3.5 that presents the slice distribution per organ and dataset size, the heart and the trachea are the smallest in slice size. On the other hand, the esophagus as well as the aorta are the largest ones. This provides us with intuition regarding the impact of the organ size on the BB-ConV layer.

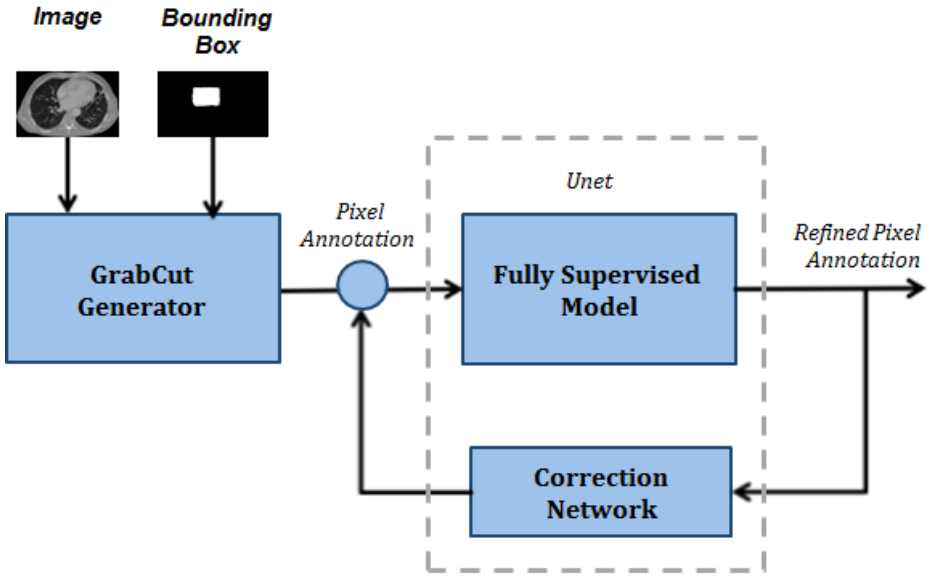


Figure 3.8 – State-of-the-art Weakly supervised segmentation process. Initial label estimates are generated through the GrabCut algorithm and updated via an iterative process of Neural Network training and label refinement.

3.4.2 Sensitivity Analysis

One way to validate the proposed model is through determining the effect of bounding box variation on model performance. To do so, we vary the size of the bounding box with respect to its initial size, which is the smallest bounding box that encompasses the organ. We increase the boundaries of each side of the bounding box by lpx pixels denoted by $+lpx$ in Table 3.4. Table 3.4 shows that given small variations in bounding boxes (increasing bounding box size by 20 % heart, 12 % aorta, 20 % trachea and 27 % esophagus) did not reveal any significant change in model performance. In fact, these small variations may have slightly improved the already present accuracies. Conducting further variations for up to 50 % of initial bounding box size resulted in slight variation in model performance while still outperforming the U-Net results by a considerable margin.

3.5 Towards Weakly Supervised Segmentation via BB-UNet

BB-UNet may also be implemented within a weakly-supervised framework. Labels within this framework are the bounding box prior provided for all images, and the full labels provided for a small proportion of the training dataset.

In the state of the art, numerous methods exploit bounding boxes in order to guide segmentation under weak labels (Dai et al., 2015; Hsu et al., 2019; Kervadec et al., 2020a; Khoreva et al., 2017). One significant approach to weak or semi-weak segmentation is via 2-step iterative processes that gradually improve generated ground-truth segmentation. Such approaches consist of 2 iterative steps. In the first step, initial label estimates are generated via intensity based computer vision methods such as GrabCut (Rother et al., 2004), GrabCut like algorithms (Khoreva et al., 2017) or Unsupervised region proposals

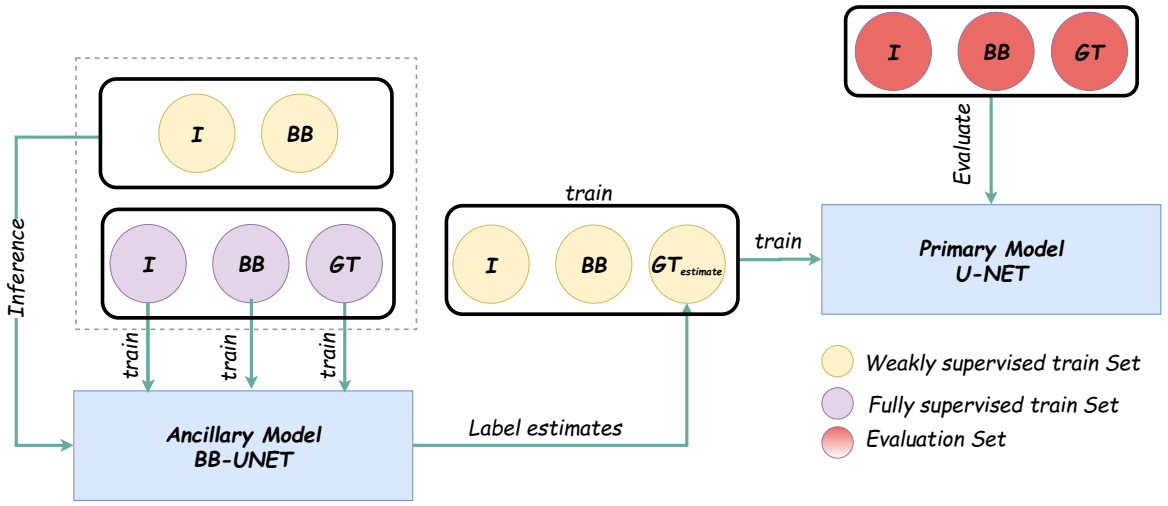


Figure 3.9 – Weakly supervised segmentation process with BB-UNet. BB: Bounding Box, *I*: I image. GT: ground-truth segmentations (full annotation).

(Dai et al., 2015). In the second step, machine learning models are trained on the initial label estimates from the first step and fine-tuned via correction networks that make use of prior cues (See Figure 3.8).

Intuitively, exploiting intensity based algorithms for natural images is logical due to the large information and intensity variations present in RGB images as demonstrated by Khoreva et al. (2017). Segmentation in the medical domain is however, more challenging given that images are rather monotonous either being Gray scale and low in contrast, as the case of CT images, or bombarded with noise and common colored areas, in the case of multi-modal images. Moreover, they are diverse by nature, depending on the region under study and the imaging equipment (e.g. CT, PET scans, X-Rays...). For this reason, exploiting intensity based algorithms as is usually done in the state of the art is rather unfavorable when it comes to medical tasks.

To address this problem, we propose a method that incorporates BB-UNet as a proficient substitute to state-of-the-art intensity based algorithms in order to sustain suitable initial label estimates for the second step of the 2-step iterative process. Exploiting a negligible partition of fully supervised data, that we refer to as the ancillary set, we allow BB-UNet to infer from bounding boxes proper initial label estimates for a much larger weakly supervised dataset called the Primary training set. In the second step, a U-Net model is then trained on the label estimates provided for the Primary-train set. The process is described in Figure 3.9.

3.5.1 Connections to Weakly Supervised State-of-the-art Segmentation Methods

Models most relevant to our work are those developed by Wei et al. (2017) and Khoreva et al. (2017). Wei et al. (2017) elaborate on a simple to complex (STC) framework, where

Table 3.5 – Slice and Size distribution of the SegTHOR dataset for weakly supervised segmentation

	Train	Train Ancillary	Train Primary	Validation	Evaluation
Patients	36	6	30	4	20
Heart	1444	219	1225	155	726

an initial deep CNN is learnt on simple images and their corresponding saliency maps. An enhanced CNN is learnt on the output of the initial ancillary model as well as the image label. Our work shares similarity with STC in that both methods train a primary U-Net based on the predictions of an ancillary model, which is the BB-UNet in our work. However, the BB-UNet does not use saliency maps or simple data, but rather our intuition lies in the idea of developing a robust model that can make full use of the information given a tiny subset of dataset that is fully supervised, so as to derive good initial label estimates for the much larger weakly supervised framework. Whereas STC's main contribution is the use of simple images to infer labels for a much larger weakly supervised dataset, our contribution is to make use of a very small amount of data in order to perform the aforementioned task.

As in (Khoreva et al., 2017), we aim to generate good initial label estimates within just one generation step. However, SDI uses $M \cap G+$ as an initial label estimator. $M \cap G+$ is an intensity-based estimator representing the intersection between multiscale combinatorial grouping segment proposals (Pont-Tuset et al., 2017) and GrabCut (Rother et al., 2004). As we have clarified earlier, intensity-based algorithms such as $M \cap G$ for example, does not provide proper label representatives. Instead, we aim to generate suitable initial label segments by exploiting a network (ancillary model) trained on a very tiny sample of a fully supervised dataset.

3.5.2 Experimental Settings

We validate the weakly supervised proposed method on a single label segmentation task targeting the Heart in SegTHOR. To this end, we extract slices where the organ of interest is present. Additionally, the training set was further divided into a much smaller fully supervised set called train-ancillary, consisting of 6 patients (219 slices), and a primary training set called train-primary, containing 30 patients (1225 slices). While the ancillary set has the full labels, we consider only bounding box labels in the primary set, as shown in Table 3.5.

3.5.3 Weakly Supervised Segmentation Results

In this section, we present results for weakly supervised segmentation of the heart in SegTHOR. Results are compared relative to 3 lower baselines. First, a GrabCut baseline without training where we simply compare the segmentation maps provided by the GrabCut algorithms relative to the ground-truth labels (GrabCut no training). Moreover, a naive circular baseline is also established as a starting point for our implementations. Given an image and a bounding box, a circular shape encompassed within the bounding

Table 3.6 – Average Dice results (in %) for weakly supervised single-organ (heart) segmentation.

	mean	std	max	min
Baselines				
GrabCut (no training)	9.925	29.15	98.28	0.0
U-Net with Circular Labels	54.43	22.46	77.91	0.0
U-Net with GrabCut Labels	64.37	32.67	92.93	0.0
State of the art				
Simple Does It with GrabCut (Khoreva et al., 2017)	84.67	3.94	90.57	67.86
EM-Adapt without CRF (Papandreou et al., 2015)	25.97	9.61	50.72	0.3
Proposed models				
BB-UNet-BB	83.19	13.32	96.36	9.09
BB-UNet-BB∩I	84.47	5.97	95.96	60.62
BB-UNet−CC	**91.69**	**11.27**	**98.77**	**23.82**
BB-UNet-CC∩I	86.79	15.03	98.54	5.12
Full supervision baseline				
U-Net Full Supervision (Perone et al., 2018)	91.53	11.12	98.79	10.67
BB-UNet Full Supervision	**95.29**	**3.51**	**98.55**	**75.65**

box is considered as the label estimate to our model. In addition, we train the baseline U-Net model based on GrabCut label estimates (U-Net with GrabCut Labels). In addition, comparison is done with other common weakly or semi-weakly supervised state-of-the-art methods, including Simple Does It (SDI) (Khoreva et al., 2017), and EM-Adapt (Papandreou et al., 2015). This later deploys an expectation-minimization algorithm where initial label estimates are generated from the weak bounding boxes in the first step and optimization of the network parameters based on the label estimates is conducted in the second.

In an optimal scenario, weakly supervised segmentation networks generate results that are close to full supervision. For this reason, we compare the proposed benchmark relative to 3 upper fully supervised baselines where all labels as well as the bounding boxes are present. We study two fully supervised scenarios: one, a regular U-Net is trained using fully annotated segmentation maps (Perone et al., 2018); two, a fully supervised framework enhanced with prior bounding box knowledge, where we train the BB-UNet model with circular filters.

From the first row of Table 3.6, we verify our conviction relative to the significance of GrabCut for medical image segmentation. Thus, despite the significance of GrabCut (no training) for natural images, it rather yields very poor performance when applied to medical tasks. This is mainly because GrabCut bases its algorithm on intensity interactions within an image not persistent in the medical case, for example CT images. This causes us to believe that intensity based measures, such as GrabCut, may not provide good representations of dataset labels.

Moreover, a closer look at the performance of the proposed models, we gather that the

proposed models outperform state-of-the-art methods by a considerable margin. More precisely, BB-UNet with circular filters takes the lead by an increase in performance of 6% with respect to GrabCut initial label estimates and other state-of-the-art methods. As a result, this leads us to believe that the proposed model is a viable solution when compared to the fully supervised framework. In this way one can avoid expensively annotating large datasets by making use of only a small partition of full annotation to conduct training.

3.6 Conclusion

In this chapter, we have investigated the significance of using bounding boxes in order to enhance segmentation performance given fully supervised setting. To this end, we proposed a new model, the BB-UNet model, that is inspired by U-Net and integrates shape and location prior by incorporating bounding areas as filters within the middle of skip connections. The proposed model outperforms the state-of-the-art in both multi-organ and multi-component segmentation settings. We further integrated the proposed BB-UNet within a weakly supervised framework. Promising results indicate the relevance of the proposed method relative to its peers within the state of the art.

Future works may include steps to relieve the BB-UNet dependency on bounding areas at inference. This can be done through addressing the feature distribution shift resulting from the augmented BB-ConV layer. Moreover, diagnostic as well as interventional imagery often consist of 3D images. Hence, exploration of the possibility of developing a BB-VNet that can perform 3D segmentation is also an aim that we hope to achieve.

Within the weakly supervised approach, future work includes developing training methods suitable for weakly supervised learning using only the BB-UNet model and independent of ancillary vs primary training. This may be done through training the BB-UNet within an unsupervised framework or through an Expectation maximization setting. Moreover, a thorough study should be carried out in order to find suitable loss functions that infer relations between the bounding boxes and the corresponding label segments.

In the next Chapter, we are going to present CoordConv-Unet that is yet another method for integrating prior at the level of the network structure by conditioning the convolutions in a U-Net on their coordinate position.

Chapter 4

Investigating CoordConv for Medical Image Segmentation

“ Even when you feel like you are going down, Never go down without a fight! ”

Contents

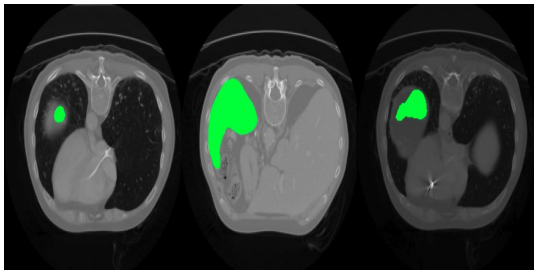


Figure 4.1 – CT images from the Spleen dataset from Decathlon challenge showing the organ having large size variability and convexity issues at boundary level

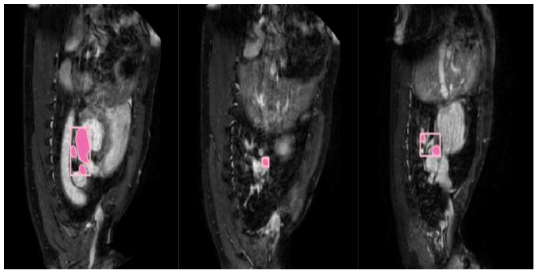


Figure 4.2 – MR images from the Cardiac dataset with manual segmentation and bounding boxes segmentation overlaid on the left Atrium.

4.1 Motivation

As discussed in Chapter 3, integrating prior knowledge at the level of network architecture may compensate for the need for large amounts of annotated training data, required to ensure good generalization or anatomical plausibility of segmentation networks. The CoordConv layers are extensions of convolutional neural networks wherein convolution is conditioned on spatial coordinates (Liu et al., 2018). The goal of CoordConv is to learn a mapping between coordinates in the Cartesian space and coordinates in the one-hot pixel space. CoordConv has shown promising potential for object localization (Levine et al., 2016; Liu et al., 2018), and has rightfully raised interest for image segmentation (Qi et al., 2019; Yao et al., 2019). However, the CoordConv's added value has not been yet assessed in prior-guided medical image segmentation.

This chapter investigates CoordConv as a proficient substitute to conventional convolutional layers for medical image segmentation tasks when trained under prior-based losses. We explore the effect of CoordConv on model performance and rate of convergence when learning is conducted under prior-based losses, particularly the size loss proposed by Kervadec et al. (2018) and the skeleton loss proposed by Shit et al. (2019). Moreover, we introduce CoordConv-Unet, which is a novel U-Net variant based on the CoordConv layer. We show that CoordConv-Unet can be used to accommodate training under these prior losses. The proposed architecture demonstrates a dual role relative to prior constrained CNN learning: it either demonstrates a regularizing role that stabilizes learning while maintaining system performance, or improves system performance by allowing the learning to be more stable and to evade local minima.

The use of CoordConv with U-Net has been investigated before (Wang et al., 2020; Wang et al., 2019). In this chapter, we investigate the significance of U-Net /CoordConv combinations when trained under prior-based losses which, according to our knowledge, has not been addressed before. The significance of CoordConv is validated on two tasks: segmentation of the Atrium and the Spleen organs from the Decathlon challenge presented in Chapter 1. The Atrium dataset is characterized by being small, with large organ size variability, having the organ composed of multi-instances that are in close proximity to each other. The Spleen dataset on the other hand, is characterized by a largely varying size and shape convexity issues at boundary level in CT images. As a summary, in this chapter:

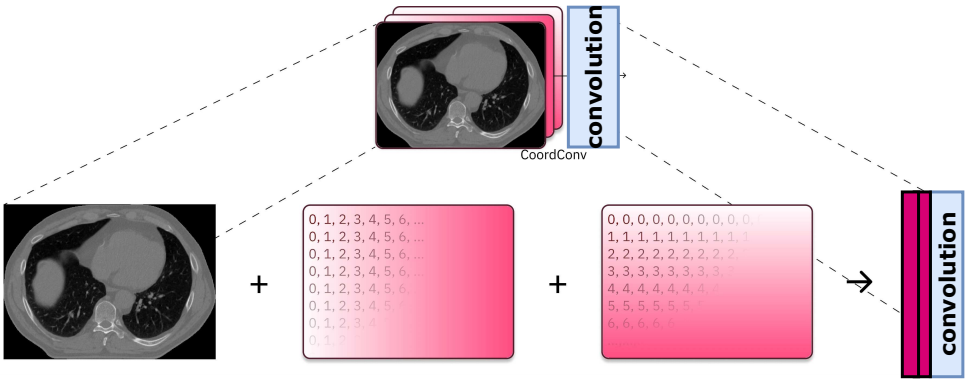


Figure 4.3 – Concept of the CoordConv layer: a combination of the conventional convolutional layer and their own coordinates.

- We investigate the significance of CoordConv solution given organ segmentation under prior-based loss training.
- We propose a novel architecture, the CoordConv-Unet as a proficient substitute to U-Net given prior constrained problems.
- We shed light on the dual role of CoordConv-Unet in increasing and stabilizing system performance.

The rest of the chapter is organized as follows. Section 4.2 elaborates on the concept of the CoordConv component and why we believe it is relevant to our prior-constrained framework. Section 4.3 presents the CoordConv-Unet model as well as the multiple frameworks and paradigms explored. Finally, we present the experimental settings and analyze the results on the two datasets in Section 4.4 and 4.5 respectively.

4.2 Concept of the CoordConv Layer

The CoordConv component is a simple extension of the standard convolutional layer wherein convolution is conditioned by spatial coordinates. The goal is to establish mappings between the Cartesian space and the pixel space, by enabling the filters to know where pixels are located. In a general sense, convolutions are mainly characterized by three specific properties: few training parameters, fast optimization via modern GPUs, and translational invariance. However, given many tasks, there is a controversy with regards to whether translational invariance will truly help model performance or not. A key example where translation invariance is unfavorable, is the supervised coordinate classification task demonstrated by Liu et al. (2018). In this application, a network must pinpoint the corresponding pixel in a pixel space as belonging to a certain considered class given that a set of x and y coordinates are provided at its input. Whereas regular convolutional networks registered a certain error, CoordConv achieves near optimal performance. The significance of CoordConv given the above task sheds light on its ability to learn location features with regards to the absolute position of objects in an image. In doing so, CoordConv ensures the best of both convolutional and spatial features.

Our work is motivated by the role of CoordConv in learning mappings between pixel positions, additionally to the conventional features learned in traditional convolutions. By allowing the convolutions to access their own coordinates as well as the inputs, CoordConv could be thought of as one way that enables constraining convolutional networks via location prior. Generally, dropping translational invariance may endorse overfitting by allowing the network to be hyper-sensitive to any variations within the object locations in the dataset; however, as elaborated by Liu et al. (2018), dedicating a small proportion of network capacity to take into account positional attributes may enable better generalization ability in trainable models as is shown by our experiments later on. Despite the fact that the CoordConv component was not implemented on segmentation within its original paper, however, many relations and connections could be drawn between CoordConv's original implementations and medical image segmentation justifying the relevance of exploiting CoordConv in our prior-constrained framework.

Segmentation is a pixel-level classification task, lying at the intersection between classification problems and object detection. Thus, aside from determining whether the organ is present in an image or not, the CNN should also learn to locate and annotate the corresponding pixels. Despite the fact that CoordConv has no added value on regular classification tasks (which are independent of location), the fact that segmentation is rather classification conditioned on location and object detection raises questions regarding the relevance of CoordConv on segmentation tasks.

4.3 CoordConv-Unet for Medical Image Segmentation

In the proposed architecture, we extend upon U-Net by replacing convolutional blocks with the CoordConv ones. As previously stated, U-Net is a symmetric encoder/decoder structure with equivalent distribution of convolutional and de-convolutional blocks connected via skip connections. Each convolutional block is composed of two consecutive ensembles of convolutional layers and batch normalization, whereas the decoder block adds a bilinear upsampling layer to the previous ensemble. Our main contribution targets the first convolutional layer of the convolutional blocks in the U-Net model. Thus, instead of directly convoluting the input of the convolution layer with that of the one before it, coordinates for each feature are taken into consideration. In doing so, we then allow convolutional filters to have access to pixel coordinates, thus endorsing the ability to pinpoint exact location errors while training. We hypothesize that integrating CoordConv in the proposed way allows taking into consideration spatial and geometric aspects while training. This is important since organs may generally be concentrated within one part of the image more than the other, as shown in Figure 4.1. Hence, by enabling the network to be sensitive to pixel location, we may allow learning of positional features relative to object locations. Moreover, supplying coordinates in one of two convolutional layers per-level, as shown in Figure 4.4, allows alternating between spatially dependent and independent notions. Another aspect is that CoordConv has shown some significance in stabilizing generative model performance in the original implementation. This raises question to whether CoordConv can play the same role in stabilizing CNN training performance under prior-based losses that are known to fall into trivial solutions and local minima.

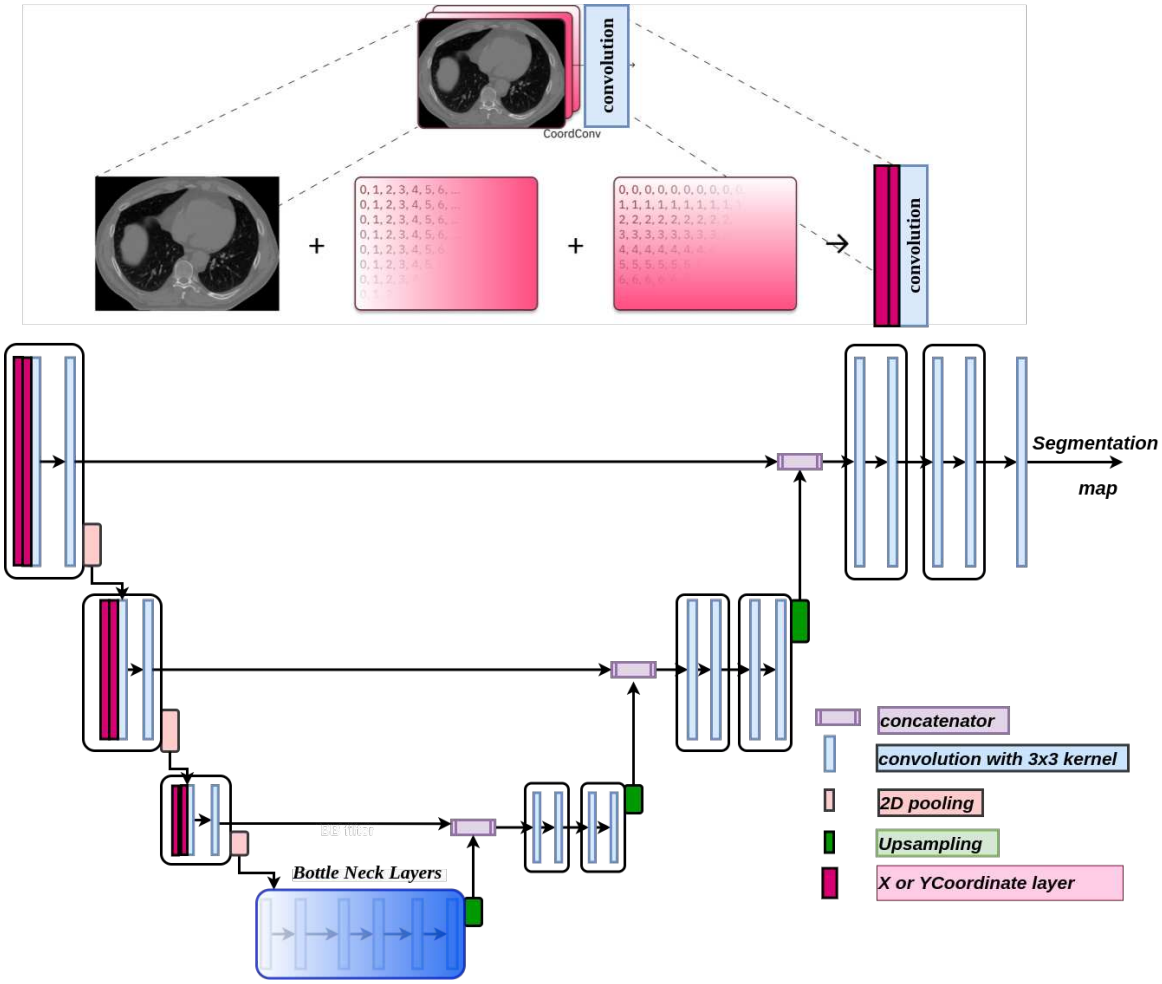


Figure 4.4 – Proposed CoordConv-Unet model. In the top panel, the CoordConv layer consists in concatenating the x-layer and y-layer to the convolutional layer. CoordConv-Unet consists of replacing the first convolutional layer of each stage with the CoordConv layer.

4.3.1 Integration Strategies

We have investigated two integration strategies of the CoordConv block onto the U-Net architecture. The first setting is one that mimics the state of the art, where the x and y channels are added only to the first convolutional layer (Liu et al., 2018; Wang et al., 2019). We call this model CoordConv(+1). The second setting consists of replacing the first convolutional layer of each convolutional block in a regular U-Net (Perone et al., 2018) within the encoding path with the CoordConv layer. We call the proposed structure CoordConv-Unet demonstrated in Figure 4.4.

4.3.2 Loss functions

The CoordConv-Unet is trained with two prior-based losses: the size loss (Kervadec et al., 2018) and the clDice loss (Shit et al., 2019).

The **Size loss** ($\mathscr{L}_{Size}$) is a penalty loss function that integrates size information by com-

puting the mean squared error between the grouping of pixel probabilities indicating predicted organ size and a predefined upper or lower bound indicating ground-truth size. A thorough analysis of the proposed loss and its significance is presented in Chapter 5.

The **clDice**($\mathscr{L}_{cl\mathrm{D}ice}$) loss exploits the topological notion of skeletonization in order to reveal subtle topological properties, such as the shape and connected components of anatomical objects within the dataset as is explained in Chapter 5.

Prior losses essentially suffer from training instability and local solutions (Kervadec et al., 2019a, 2018; Shit et al., 2019). To overcome this, we integrate the size loss (Kervadec et al., 2019a) and clDice loss (Shit et al., 2019) in conjunction with the Dice baseline weighted by a hyperparameter λ according to the following equation

$$\mathscr{L} = (1-\lambda)\mathscr{L}_{\mathrm{D}ice} + \lambda\mathscr{L}_{prior}.$$

Thus, the hyperparameter λ is dynamically updated through training, as follows: starting from a value of $\lambda = 0.01$, λ is increased by a value of 0.01 at each training epoch.

4.3.3 Connections to the State-of-the-art

The CoordConv concept in essence is not a new topic. Introduced by Liu et al. (2018), the CoordConv layer was initially designed to investigate supervised coordinate classification, object detection, supervised coordinate regression, and generative adversarial modeling. The method in (Liu et al., 2018) raises questions with regards to CoordConv's ability to make the training more stable given a coordinate classification task. However, conducted experiments do not address the segmentation problem, which is quite important in current research. To add to this, the method added the CoordConv layer solely at the first convolution of the entire architecture.

Since segmentation is a classification problem done at pixel-level, investigating the efficiency of CoordConv in aiding prior-constrained medical segmentation is an interesting research direction to be explored. In this context, works most relevant to ours are (Wang et al., 2019) and (Wang et al., 2020). Wang et al. (2019) investigate the role of CoordConv in conducting 3D segmentation of pulmonary lobes via a V-Net. Instead of replacing the encoder layers of the network as our proposed CoordConv-Unet model, they perform this interchange at the level of the up-sampling layers. Thus, the method first exploits a 2D automated lung segmentation model followed by the CoordConv embedded V-Net architecture. On the other hand, Wang et al. (2019) propose a 3-stage framework in order to conduct brain mid-line delineation. Within the segmentation step, they introduce the CoordConv component at the input level of the network at the intermediate segmentation step, midway between alignment (first step) and delineation (third step).

Whereas Wang et al. (2020) integrate CoordConv solely at the input level (as is originally implemented in (Liu et al., 2018)), our proposed CoordConv-Unet interchanges the first convolution of each stage of the U-Net architecture with that of the CoordConv layer. Moreover, we do not conduct pre- or post-processing steps as in (Wang et al., 2020). Finally, unlike Wang et al. (2019) where the CoordConv component is incorporated at the level of the upsampling layers, the proposed method does so at the encoder convolution level.

4.3.4 Computational complexity

Given a U-Net, parameters generally involve the number of learnable quantities or weights connecting a convolutional layer with the corresponding precedent or following layer. In order to quantify the computational overhead induced by the CoordConv component, we will adopt the following mathematical notations. Let m represent the width of the convolutional filter and n its height. Let d be the number of filters in the preceding layer and k the number of filters in the current one. Then, the number of parameters involving the conventional convolutional layer is hence computed according to $(mnd+1)k$, where 1 represents the bias term for each filter. Given the CoordConv component, since 2 additional channels are added at the level preceding the convolutional layer, the number of filters in the preceding layer is hence $d+2$, thus resulting in a new number of parameters of $(mn(d+2)+1)k$. In such a way, each convolutional layer adds to the computational complexity $2mnk$ operations.

4.4 Experimental Setting

To ensure reproducibility, we deploy a well-known experimental framework presented by Kervadec et al. (2019a). The U-Net (Milletari et al., 2016), of which we integrate the CoordConv layers onto, is a 3-stage structure composed of convolutional and de-convolutional blocks, bottleneck and skip connections. Each stage within the encoder is composed of convolutional blocks containing an ensemble of convolutional and batch normalization layers. On the other hand, each stage within the decoder path is composed of 2 consecutive convolutional blocks followed by an upsampling layer. The bottleneck is constituted of 2 convolutional blocks separated by a residual block (Zhang et al., 2018).

The implementation of CoordConv is done by concatenating two additional x and y channels to the input channel as shown in Figure 4.4(top figure). In such a way, CoordConv allows the learning of a function characterized by a certain degree of translational dependence, if the weights connecting the coordinate layers of the CoordConv with the convolutional are non-zero or could mimic a regular convolutional layer if they were set to zero. In the proposed experiments, CoordConv is implemented via a PyTorch library[1] where a linear scaling is applied in order to bound the values of the coordinate layers between -1 and 1.

For pre-processing, we have resized the images to a size of 256×256 and normalized them to a pixel value between 0 and 1. Deploying the framework presented by Kervadec et al. (2019a), we have kept negative samples for training. Negative samples are empty images, meaning that the organ of interest is not present. The datasets were split into train and validation based on an 80 % / 20 % partition respectively. Cross-validation was done on three folds of the data based on three Monte-Carlo simulations (Arlot and Celisse, 2010).

Models were evaluated using the Dice index and Hausdorff distance. Training was conducted via the Adam optimizer with a batch size of 8 over 200 epochs. The learning rate was set to 5×10^{-4} and halved each 20 epochs if the validation performance did not improve.

[1] https://github.com/walsvid/CoordConv

Table 4.1 – Average Dice scores ± standard deviation on the Spleen and Atrium datasets. Best results for each architecture are in bold.

		$\mathscr{L}_{Dice}$	$\mathscr{L}_{Dice} + \mathscr{L}_{Size}$	$\mathscr{L}_{Dice} + \mathscr{L}_{clDice}$
Spleen	U-Net	**78.58±5.46**	86.44±15.87	87.15±13.61
	CoordConv(+1)	64.65±5.7	**94.86±1.72**	87.04±9.98
	CoordConv-Unet	65.48±8.48	**94.96±1.59**	**94.54±1.06**
Atruim	U-Net	**83.67±3.66**	84.59±2.62	83.85±2.56
	CoordConv(+1)	82.10±2.51	84.63±1.67	85.35±1.65
	CoordConv-Unet	82.52±2.33	84.48±1.50	**86.15± 1.39**

4.5 CoordConv-Unet Performance under Prior-based Losses

In this section, we present results for the CoordConv-Unet and CoordConv(+1) models under prior-based losses. Training is conducted via the clDice loss and size loss in conjunction with Dice loss according to the dynamic strategy via λ demonstrated in section 4.3.2. The conjunction of prior losses with the Dice baseline is denoted as $\mathscr{L}_{Dice} + \mathscr{L}_{clDice}$ and $\mathscr{L}_{Dice} + \mathscr{L}_{Size}$ respectively dropping the λ notation for simplicity. We compare relative to the regular U-Net baseline under just the Dice loss as well as the prior-based losses. We validate the significance of CoordConv component and CoordConv-Unet via two datasets: the Atrium and the Spleen datasets. We recall that the Atrium is characterized by multi-connected small components that are very close to each other, whereas the Spleen is a complex-shape organ of non-convex curves and edges. Results relative to the Dice accuracy and Hausdorff distances are benchmarked in Table 4.1 and Table 4.2.

From the tables, we observe that, in a general sense, the segmentation performance is improved when the proposed structures are trained on prior-based losses but degraded when trained via the Dice loss baseline. Trained under the clDice loss, CoordConv-Unet increases the Dice accuracy by over 8% from the regular U-Net baseline for the Spleen, and by about 3% on the Atrium. This added value of CoordConv is further verified by the error computed on Hausdorff distance where CoordConv-Unet registers a 20% decrease (from 1.07 to 0.85) relative to the U-Net baseline for Spleen. For the Atrium, CoordConv-Unet under clDice scores second best relative to the other frameworks. This indicates the ability of CoordConv-Unet to learn features regarding the inter-distance position of the pixels. In addition, adding the CoordConv component allows for more stable training and inference as indicated by Figure 4.5-LEFT. Thus, the addition of CoordConv has evaded the undershoot prevalent when the network is trained via regular convolutional layers

Compared to the size loss on the Spleen dataset, CoordConv-Unet improves segmentation performance over Dice accuracy by about 8% and registers a decrease in Hausdorff distance by about 25 % (from 1.02 to 0.76 mm). Intuitively, under full supervision, CoordConv-Unet allows the network to determine the position of border pixels, whereas the size loss constrains the inter-distance or the relative distance between the different border pixels in an organ. We hypothesize that the combination of CoordConv-Unet and the size loss allows the network to make use of both absolute positions of bor-

Table 4.2 – Average Hausdorff distance ± standard deviation results on the Spleen and Atrium datasets. Best results for each architecture are in bold.

		$\mathscr{L}_{Dice}$	$\mathscr{L}_{Dice} + \mathscr{L}_{Size}$	$\mathscr{L}_{Dice} + \mathscr{L}_{clDice}$
Spleen	U-Net	1.30 ± 0.24	1.02±0.56	1.07±0.53
	CoordConv(+1)	1.89± 0.26	**0.76±0.13**	1.11±0.31
	CoordConv-Unet	1.71± 0.32	**0.76±0.12**	**0.85± 0.07**
Atruim	U-Net	1.62±0.16	1.59±0.17	1.64±0.16
	CoordConv(+1)	1.64±0.11	**1.57±0.08**	1.60±0.14
	CoordConv-Unet	1.64±0.11	**1.57±0.10**	**1.59±0.14**

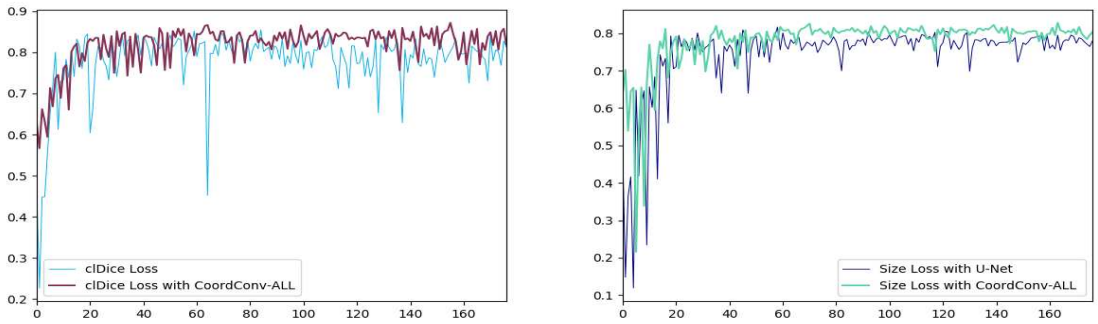


Figure 4.5 – Evolution of the Dice accuracy in validation under Dice +clDice (LEFT), Dice +Size (RIGHT) for the Atrium dataset.

ders provided via CoordConv and relative distances between border pixels provided via the size constraint in order to improve segmentation performance. On the other hand, CoordConv-Unet maintains the system performance on the Atrium. This is logical since the Atrium is generally composed of multi-connected components where the overall size integrated into the loss function resembles the summation of label pixels over all the instances with background gaps in between. Hence, the information provided by the overall organ size does not designate an accurate relative position that can permit enhancing determining boundary pixel location. Nevertheless, the exploitation of CoordConv-UNet with the size loss for the Atrium does not degrade system performance. Based on these results, one would anticipate that the behavior of the CoordConv would be typical to that of a regular convolution. However, a closer look at the evolution of the Dice accuracy over the number of training epochs given folds from Figure 4.5 show us that CoordConv solution is equivalent to that of a regular U-Net. Thus, CoordConv-Unet insures model stability and convergence by evading the undershoot evident when training with regular U-Net. This is yet another significant advantage of using CoordConv-Unet in place of regular U-Net.

Based on the above results, we can realize the dual role that the CoordConv component plays in enabling learning of spatial dependent attributes when needed (case of the clDice) or mimicking typical convolutional functioning with added stability (case of size

loss). Thus, CoordConv can either maintain segmentation performance while regularizing training and stabilizing evolution, or it increases system performance by evading the local solutions that prior losses suffer from.

Despite the significance of CoordConv-Unet as shown by the above study, however, limitations still persist with regards to its performance. When tested against the Dice loss baseline alone (first column of each table), the addition of the CoordConv components degrades the system performance considerably. One could cast the clarification of the phenomena to the complexity/regularizing trade-off relative to CoordConv-Unet. We have clarified earlier 2 main roles of CoordConv: 1) the role of stabilizing the undershoot evident when training against prior losses; 2) the role of enhancing segmentation performance by evading local solutions. Addressing the two attributes relative to the Dice baseline training, we can gather the following: the undershoot revealed in the plot figures are a result of the interchange between the main pixel-wise Dice loss on one hand and the prior loss on the other in prior-constrained training problems. In the Dice baseline, the undershoot is rather non-existent; hence CoordConv cannot play the role of the regularizer that evades training instability and ends up decreasing model performance by adding complexity to the system. The second role of CoordConv is that it increases model performance under prior-based losses by evading local solutions. As we have previously said, prior-based losses are interestingly used because they integrate expert knowledge onto the training. However, designing these loss functions is often tedious and subjected to various differentiability and stability challenges. One of the reasons for these issues is the discrete nature of this prior information vs the continuous (real) soft probability output of the network. With CoordConv-Unet, the network can act as a regular U-Net if needed, i.e., if no stability problems persist or the network could integrate spatial knowledge thus fulfilling the true CoordConv concept. In regular Dice baseline training, prior does not exist, hence, addressing stability is not an issue.

4.6 Conclusion

In this chapter, we have proposed a new model, the CoordConv-Unet model as a proficient substitute to U-Net given prior-constraint tasks. We have exposed the dual role of CoordConv-Unet given these constrained tasks. Thus, CoordConv-Unet can either improve performance by evading stability problems, or can mimic a regular U-Net if the translation invariance attribute is required. In the latter role, CoordConv-Unet still demonstrates a further significance that can help stabilize the network performance under the prior-based losses.

Future work includes designing frameworks that can impose whether the weights connecting the coordinate layers with the convolutional ones are trained or fixed, so as to resolve the problem faced when training CoordConv-Unet under a regular unconstrained Dice loss. Moreover, efficiency of CoordConv-Unet for multi-organ and lesion segmentation is also to be explored.

Despite the fact that structural constraints are robust, loss constraints are more generic and can be plugged into any backbone network. In the next part of the thesis, we will demonstrate our contributions to the field of prior-constrained convolutional neural networks conducted at the level of the loss function.

Part II

Integration of Prior Constraints via the Loss Function

Chapter 5

Effect of Prior-based Losses on Segmentation Performance in Medical Imaging

“ Life is beautiful, so Stop worrying and Smile! ”

Contents

5.1 Motivation

As previously stated in Chapter 2, prior-based losses, whether low-level or high-level, present a rising trend in today's research in semantic image segmentation, particularly in the medical field. Given the diversity of prior-based losses on different medical imaging challenges and tasks, it has become hard to identify what loss works best for which dataset. For this reason, we establish in this chapter, a benchmark of recent prior-based losses for medical image segmentation. Our main objective is to provide intuition onto which losses to choose given a particular task or dataset, based on dataset characteristics and properties.

In (Ma et al., 2021), a benchmark of 20 losses is conducted with a thorough comparison on 4 main segmentation tasks: Liver, Liver Tumor, Pancreas and Multi-Abdominal Organ Segmentation. However, the authors do not address prior-based losses. Instead, they consider regular fitting losses like Dice, Cross entropy and their variants. Their benchmark is limited to only 4 datasets. Another benchmark that targets some low-level prior-losses, is the one proposed by Ma et al., 2020. However, this benchmark is limited to the scope of losses based on distance maps, such as the boundary loss (Kervadec et al., 2019a) or the Hausdorff loss (Karimi and Salcudean, 2019), and do not compare relative to high-level prior losses. Karimi and Salcudean, 2019 also demonstrate results on some structural constraints (i.e., regarding the architecture), that do not lie within the scope of our benchmark. In addition, the benchmark is limited to two datasets: an organ segmentation task of the left atrial structure within MRI images and a liver tumor segmentation task within CT scans. In this work, we target specifically prior-based losses, both high-level and low-level, on 8 datasets of different tasks and modalities. Hence up to our knowledge, there is no benchmark that aims to compare prior-based losses on a number of datasets in order to quantify common trends and limitations.

The main objective of the proposed benchmark is to study the performance of prior-based losses, on a variety of datasets, tasks and modalities. In this way, we provide the readers with intuition onto which losses to choose given a particular task of interest. Prior-based losses are quite interesting because they allow integration of expert knowledge while still being architecture-agnostic, that is to say, they can be plugged into any backbone network. As a result, we are able to unify the segmentation network given the same learning environment, while varying the prior-based losses accordingly. We note that each of the considered losses has been proposed in their respective papers, in order to carry on a particular task. We believe that aside from the initial motive that the considered losses were designed for, additional significance may be drawn on other segmentation tasks and dataset characteristics. For this reason, we validate the chosen prior-based losses on 8 different datasets from a variety of medical image segmentation challenges including the Decathlon, the ISLES and the WMH challenge. The main contributions of this chapter are summarized as follows:

- We present a benchmark of architecture-agnostic prior-based losses for medical image segmentation.
- We attempt to shed light on the underlying relationship between the prior-based losses and some dataset characteristics.

The rest of the chapter is organized as follows. Section 5.2 presents the selected loss functions for the proposed benchmark and elaborate on why the proposed losses were chosen. Section 5.4 illustrates the experimental setting adopted in order to evaluate the considered prior-based losses on the different datasets. In section 5.3, we describe the datasets considered and the meta-features extracted to compare the loss performances. Finally, section 5.5 demonstrates the results and analyzes the loss performances relative to segmentation tasks and dataset characteristics.

5.2 Selected Loss Functions

In this section, we present the chosen prior-based losses for our proposed benchmark. Recalling from Chapter 2 that prior-based losses can be high-level, when the type of prior considered is based on external knowledge (e.g. shape), or low-level, that integrate ground-truth map transformations such as distance or contour maps, in order to reveal geometrical and location properties, the benchmark mainly focuses on 4 recent prior-based losses that have raised interest within the field of medical image segmentation, 2 are low-level, and the other 2 are high-level.

5.2.1 Low-level prior-based losses

Possible low-level prior can be based on distance map as demonstrated in (Karimi and Salcudean, 2019; Kervadec et al., 2019a). In this context, two major contributions are the boundary loss (Kervadec et al., 2019a) and the Hausdorff loss (Karimi and Salcudean, 2019).

The **Boundary loss** $\mathscr{L}_{\mathrm{B}oundary}$ is an approximation of the distance between the real and the estimated boundaries. Based on (Boykov et al., 2001) graph theories, Kervadec et al., 2019a derive an equivalent term that finetunes the probability distribution via ground-truth distance maps and is defined as:

$$\mathscr{L}_{\mathrm{B}oundary} = \sum_{p\in\Omega} \phi_g(p).\hat{y}_p \qquad \text{with } \phi_g(p) = \begin{cases} -\mathrm{D}_{\mathrm{G}}(y_p) & \text{for } p \text{ inside the target region} \\ \mathrm{D}_{\mathrm{G}}(y_p) & \text{else,} \end{cases} \tag{5.1}$$

where $\mathrm{D}_{\mathrm{G}}(p)$ denotes the distance of pixel p to the closest contour (G) point, $\widehat{y}_p$ being the predicted value at pixel p, and Ω the image spatial domain.

The **Hausdorff loss** $\mathscr{L}_{\mathrm{HD}}$ (Karimi and Salcudean, 2019) conducts a direct point-by-point optimization of the predicted and ground-truth contours arriving to the following loss term:

$$\mathscr{L}_{\mathrm{HD}} = \frac{1}{|\Omega|} \sum_{p\in\Omega} (y_p - \hat{y}_p)^2 \Big(\mathrm{D}_{\mathrm{G}}(y_p)^2 + \mathrm{D}_{\mathrm{G}}(\hat{y}_p)^2 \Big). \tag{5.2}$$

The boundary loss has been initially designed in order to segment lesions within the brain, with the WMH and the ISLES datasets, whereas the Hausdorff loss has been tested on 4 different single-organ segmentation tasks, including the prostate, liver and pancreas from the Decathlon and PROMISE challenges. However, these losses were not evaluated

in multi-organ segmentation. Since both losses lie in the same spectrum of low-level prior-based losses, and rely on the distance map, it may be interesting to investigate their performance on the same datasets in order to pinpoint common behaviors. Moreover, we aim to also extend the scope of these losses to the multi-organ case.

5.2.2 High-level prior-based losses

Regarding high-level prior losses, we analyze the performance of the clDice loss (Shit et al., 2020) and the size loss (Kervadec et al., 2018).

The **Size loss** (Kervadec et al., 2018) estimates the organ size from a soft probability map and constrains it, based on higher and lower threshold value of the organ size, according to the following:

$$\mathrm{A}(\widehat{y}) = \sum_{p \in \Omega} \widehat{y}_p, \tag{5.3}$$

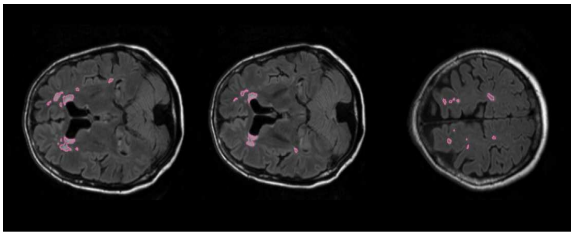
$$\mathscr{L}_{size} = \begin{cases} \big(\mathrm{A}(\widehat{y}) - a\big)^2 & \text{if } \mathrm{A}(\widehat{y}) \leq a, \\ \big(\mathrm{A}(\widehat{y}) - b\big)^2 & \text{if } \mathrm{A}(\widehat{y}) \geq b, \\ 0 & \text{otherwise,} \end{cases} \tag{5.4}$$

where a and b are respectively the upper and lower permissible bounds that the size of the considered object can attain. The size loss was originally designed for weakly supervised learning, to guide the network through the training despite the lack of full label maps. We are particularly interested in studying the effect of the size loss on small structures that are known to be more difficult to segment.

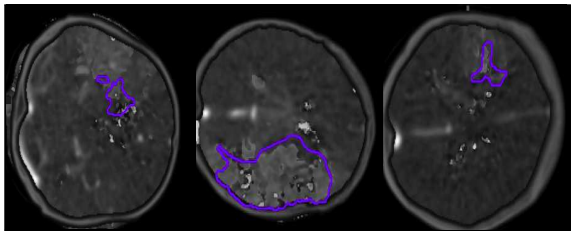
The **clDice loss** (Shit et al., 2019), also called skeleton loss, exploits skeletonization maps that are compact representations of images and objects that preserve topological properties. The objective of this loss is to constrain the skeleton of the predicted map to match the skeleton of the ground-truth map. This prior was used in the segmentation of vessels and neurons in both 2D and 3D. Let s and $\widehat{s}$ be the ground-truth and the predicted skeleton respectively, of size $|\Omega|$. The sensitivity (or recall) between the predicted segmentation and ground-truth skeleton is introduced as $\mathrm{T}_{sens}(s, \widehat{y}) = |s \cap \widehat{y}|/|s|$. Likewise, the precision between the ground-truth mask y and the predicted skeleton $\widehat{s}$ is defined as: $\mathrm{T}_{prec}(\widehat{s}, y) = |\widehat{s} \cap y|/|\widehat{s}|$. The clDice is defined as the F1-score between precision T_{prec} and sensitivity T_{sens} as follows:

$$\mathscr{L}_{clDice} = 2\frac{\mathrm{T}_{prec}(\widehat{s}, y)\,\mathrm{T}_{sens}(s, \widehat{y})}{\mathrm{T}_{prec}(\widehat{s}, y) + \mathrm{T}_{sens}(s, \widehat{y})}. \tag{5.5}$$

The clDice was originally designed to segment vessels; however, due to the nature of the skeletonization feature that they target, we believe that they may be good at distinguishing between different structures lying in close proximity to each other, such as when the organs are made of multiple instances.

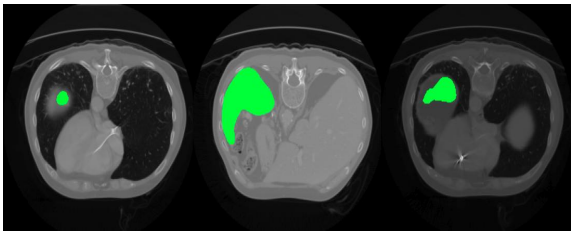


(a) White Matter Hyper-intensity dataset (WMH) from the MICCAI challenge (Kuijf et al., 2019)

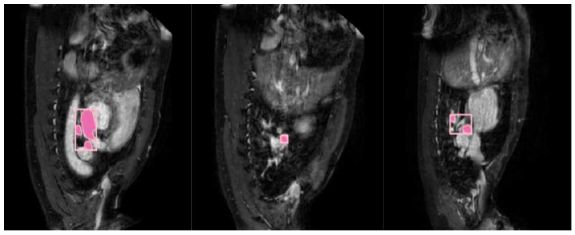


(b) Ischemic Stroke Lesion Segmentation (ISLES) dataset from the MICCAI 2015 challenge (Maier et al., 2017)

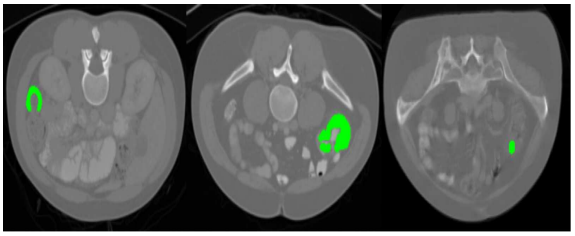
Figure 5.1 – Brain lesion segmentation task



(a) Spleen Dataset from Decathlon challenge showing organ or large size variability and convexity issues at boundary level.



(b) MR images from the Atrium dataset with manual segmentation and bounding boxes segmentation overlaid on the left atrium.

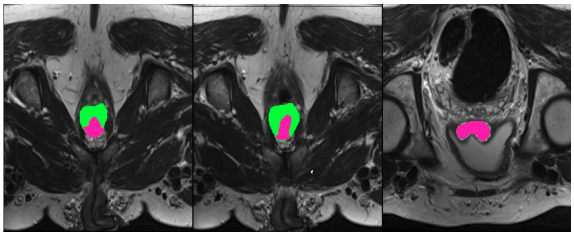


(c) Colon Dataset showing Colon Cancer Primaries of different curvatures and sizes.

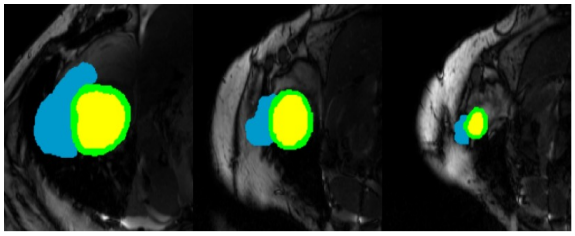
Figure 5.2 – Single-organ segmentation tasks from the Decathlon Challenge (Simpson et al., 2019).

5.3 Datasets and Tasks

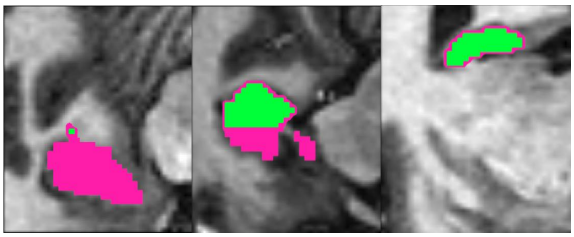
In this section, we present a brief description of the datasets under consideration. The datasets were chosen to cover different tasks, modalities and characteristics. Each dataset encompasses a particular set of challenges the segmentation network must consider while training. A summary of the meta-dataset characteristics is presented in Table 5.1.



(a) Prostate Dataset showing central gland (pink) and peripheral zone (green) (Simpson et al., 2019).



(b) ACDC Dataset with right ventricle structure is in blue, left ventricle structure is yellow and myocardium structure in green (Bernard et al., 2018)



(c) Hippocampus Dataset from Decathlon challenge where two brain tissues lying in close proximity of each other are to be segmented given low resolution images (Simpson et al., 2019).

Figure 5.3 – Multi-Organ Segmentation Tasks.

5.3.1 Brain Lesion Segmentation

To investigate the significance of prior-based losses on Brain lesion segmentation tasks, we mainly focus on the segmentation of white matter hyperintensities (WMH) dataset describe in Section 1.5.5 and the ischemic stroke lesion segmentation dataset (ISLES) described in Section 1.5.3. Both datasets are multi-modal with anatomical objects that are characterized by being sparse and composed of multi-instances (See Figure 5.1).

5.3.2 Single Organ Segmentation

Organs can generally be single-connected of only 1 structure, or multi-connected composed of multi structures that are close to each other. To investigate the segmentation performance of prior-based losses on single-organ segmentation tasks where the organ considered is characterized with multi-connected structures, we targeted the segmentation of the atrium and Colon from the Decathlon Challenge (described in 1.5.2). Alternatively, we target the spleen to investigate the performance of prior-based losses relative to single-label single-connected organs. The spleen and colon are characterized with a largely varying size and mild convexity issues at boundary levels. On the other hand, the atrium is a multi-instance anatomical object with up to 4 elements of varying sizes and lying in close proximity to each other (See Figure 5.2).

Table 5.1 – Dataset Description: # of patients: patient split is 80 % / 20 % on the original dataset; Organ Size: % of pixels occupied by the organ w.r.t. the entire image; # of CC: number of connected components;

		# of patients		Organ Size		# of classes	# of modality	# of
		Train	Test	mean	std			CC
WMH		48	12	0.33	0.56	1	2	0 ~ 26
Isles		74	20	2.11	1.91	1	5	0 ~ 3
Atrium		16	4	0.69	0.43	1	1	0 ~ 4
Colon		100	38	0.6	0.59	1	1	0 ~ 3
Spleen		32	9	1.57	1.03	1	1	0 ~ 1
Hippo.	***H1***	206	54	4.08	3.87	2	1	0 ~ 4
	H2			3.53	2.53			0 ~ 3
Prost.	***CG***	26	6	0.9	0.89	2	2	0 ~ 26
	PZ			3.1	2.98			0 ~ 1
ACDC	***RVC***	99	24	1.29	1.03	3	1	0 ~ 1
	MYO			1.38	0.69			
	LVC			1.28	0.84			

5.3.3 Multi-Organ Segmentation

For multi-organ segmentation, we have targeted the Prostate (Prostate central gland and peripheral zone) and Hippocampus (tissues H1 & H2) datasets from the Decathlon Challenge and the ACDC dataset (Three Cardiac Structures) described in Section 1.5.

5.3.4 Meta-dataset Features

In order to reveal the underlying relationship between loss performance and dataset characteristics, we propose a set of meta-features that describe the datasets. This includes the size of the anatomical object taken as the percentage of occupation from the entire image, the number of connected components, which means how many instances an anatomical object is constituted of, and the number of classes, i.e., whether the segmentation task is single or multi-label.

5.4 Experimental Setting

We deploy the unified U-Net based framework proposed in (Kervadec et al., 2019a, 2018) and modify the loss function accordingly. Training is done using a batch size of 8 and a learning rate of 10^{-3}. The learning rate is halved if the validation performance does not improve during 20 epochs. The U-Net model is trained via each prior-based loss in conjunction with the Dice loss weighted by a parameter λ according to the following equa-

Table 5.2 – Average Dice scores ± standard deviation. Blue (resp. pink) background represents Dice Accuracy superior (resp. inferior) to the corresponding Dice baseline. The bold result is the best Dice score (i.e., the greatest) obtained on the dataset.

Dataset		$\mathscr{L}_{Dice}$	$\mathscr{L}_{Dice}+\mathscr{L}_{Boundary}$	$\mathscr{L}_{Dice}+\mathscr{L}_{HD}$	$\mathscr{L}_{Dice}+\mathscr{L}_{Size}$	$\mathscr{L}_{Dice}+\mathscr{L}_{clDice}$
WMH		74.64 ± 1.34	77.29 ± 0.75	**78.77 ± 0.70**	**78.06 ± 1.61**	66.97 ± 11.48
Isles		53.41 ± 4.61	62.93 ± 2.24	**63.53 ± 1.66**	46.86 ± 7.74	62.53 ± 5.22
Atrium		83.67 ± 3.66	82.80 ± 3.68	**84.57 ± 1.86**	**84.59 ± 2.62**	83.85 ± 2.56
Colon		84.82 ± 1.71	**88.71 ± 0.48**	**88.30 ± 0.78**	**88.71 ± 0.48**	84.52 ± 2.64
Spleen		76.80 ± 7.59	80.38 ± 5.46	**91.79 ± 2.67**	86.44 ± 15.86	87.15 ± 13.61
Hippocampus	*H1*	49.38 ± 0.33	65.20 ± 0.31	**68.54 ± 1.46**	66.24 ± 0.33	**68.39 ± 2.60**
	H2	71.70 ± 1.30	81.33 ± 0.74	**82.12 ± 0.44**	81.84 ± 0.63	**82.82 ± 1.22**
Prostate	*CG*	**45.17 ± 6.41**	44.89 ± 7.09	44.15 ± 5.61	34.12 ± 7.49	42.45 ± 7.03
	PZ	65.13 ± 11.57	**68.99 ± 9.94**	64.38 ± 9.33	29.61 ± 12.07	61.57 ± 11.44
ACDC	*RVC*	80.79 ± 0.95	81.04 ± 0.87	80.54 ± 1.30	41.02 ± 38.39	**83.83 ± 1.39**
	MYO	**83.92 ± 0.13**	**84.16 ± 0.83**	83.91 ± 0.85	83.41 ± 0.72	83.24 ± 0.66
	LVC	**90.26 ± 0.13**	89.53 ± 0.74	88.98 ± 0.90	89.74 ± 0.71	89.56 ± 1.10

Table 5.3 – Average Hausdorff Distances ± standard deviation. Blue (resp. pink) background represent HD inferior (resp. superior) to the corresponding Dice baseline. The bold result is the best (i.e. the smallest) Hausdorff Distance obtained on the dataset.

Data-Set		$\mathscr{L}_{Dice}$	$\mathscr{L}_{Dice}+\mathscr{L}_{Boundary}$	$\mathscr{L}_{Dice}+\mathscr{L}_{HD}$	$\mathscr{L}_{Dice}+\mathscr{L}_{Size}$	$\mathscr{L}_{Dice}+\mathscr{L}_{clDice}$
WMH		0.98 ± 0.13	0.94 ± 0.17	**0.93 ± 0.16**	0.94 ± 0.18	1.16 ± 0.38
Isles		3.75 ± 0.35	**3.05 ± 0.22**	3.07 ± 0.18	3.45 ± 0.79	3.29 ± 0.62
Atrium		1.62 ±, 0.16	1.64 ± 0.16	1.67 ± 0.13	**1.59 ± 0.17**	1.64 ± 0.16
Colon		0.58 ± 0.04	**0.50 ± 0.02**	0.51 ± 0.03	**0.50 ± 0.02**	0.58 ± 0.07
Spleen		1.33±, 0.28	1.34 ± 0.21	**0.92 ± 0.15**	1.27 ± 0.57	1.07 ± 0.53
Hippocampus	*H1*	2.31 ± 0.05	1.99 ± 0.01	1.98 ± 0.03	**1.97 ± 0.02**	1.99 ± 0.04
	H2	3.82 ± 0.14	3.09 ± 0.01	**2.97 ± 0.05**	3.07 ± 0.01	3.20 ± 0.18
Prostate	*CG*	2.80 ± 0.34	**2.77 ± 0.43**	2.88 ± 0.27	3.10 ± 0.26	3.48 ± 0.66
	PZ	3.24 ± 0.35	**2.94 ± 0.27**	3.17 ± 0.47	4.41 ± 0.90	3.45 ± 0.58
ACDC	*RVC*	2.44 ± 0.04	2.41 ± 0.05	**2.33 ± 0.04**	3.88 ± 1.44	2.34 ± 0.08
	MYO	2.60 ± 0.01	**2.57 ± 0.01**	2.65 ± 0.01	2.62 ± 0.00	2.71 ± 0.04
	LVC	1.95 ± 0.02	1.95 ± 0.02	1.98 ± 0.01	1.94 ± 0.01	1.98 ± 0.04

tion:

$$\mathscr{L} = (1-\lambda)\mathscr{L}_{Dice} + \lambda\mathscr{L}_{prior} \tag{5.6}$$

The parameter λ is fine-tuned via the dynamic training strategy in (Kervadec et al., 2019a). Thus, it was set to 0.01 and increased by 0.01 per epoch for 200 epochs.

For pre-processing, we have resized the images to 256 × 256 pixels and normalized them to the range [0, 1]. For multi-modal datasets, we have concatenated the channels at the level of the input. Each dataset was split into train and validation based on an 80 % / 20 % partition respectively, as shown in Table 5.1, and validated via three Monte-Carlo simulations (Arlot and Celisse, 2010).

Table 5.4 – Mean Absolute Error (MAE) on the number of connected components (CC) of the ground truth vs the number of CC of the predicted segmentation map. Blue (resp. pink) background represents an MAE inferior (resp. superior) to the corresponding Dice baseline.

Data-Set	$\mathscr{L}_{Dice}$	$\mathscr{L}_{Dice}+\mathscr{L}_{Boundary}$	$\mathscr{L}_{Dice}+\mathscr{L}_{HD}$	$\mathscr{L}_{Dice}+\mathscr{L}_{Size}$	$\mathscr{L}_{Dice}+\mathscr{L}_{clDice}$
WMH	1.04 ± 0.14	0.98 ± 0.17	1.01 ± 0.23	0.91 ± 0.22	2.14 ± 1.26
Isles	0.69 ± 0.19	0.48 ± 0.04	0.57 ± 0.19	1.34 ± 1.08	0.39 ± 0.10
Atrium	0.25 ± 0.01	0.29 ± 0.02	0.32 ± 0.01	0.28 ± 0.03	0.28 ± 0.03
Colon	0.17 ± 0.02	0.13 ± 0.01	0.13 ± 0.01	0.13 ± 0.01	0.18 ± 0.03
Spleen	0.22 ± 0.05	0.24 ± 0.09	0.09 ± 0.01	0.18 ± 0.13	0.12 ± 0.15
Hippocampus-*H1*	3.76 ± 0.14	1.81 ± 0.14	2.67 ± 0.11	2.88 ± 0.39	1.30 ± 0.96
Hippocampus-*H2*	0.95 ± 0.01	0.23 ± 0.01	0.87 ± 0.10	0.74 ± 0.06	0.10 ± 0.03
Prostate-*CG*	8.96 ± 3.33	9.05 ± 3.33	8.89 ± 3.11	8.78 ± 2.79	8.98 ± 3.32
Prostate-*PZ*	0.36 ± 0.13	0.23 ± 0.09	0.33 ± 0.07	0.80 ± 0.12	0.26 ± 0.11
ACDC-*RVC*	0.18 ± 0.03	0.16 ± 0.00	0.13 ± 0.02	0.11 ± 0.04	0.12 ± 0.03
ACDC-*MYO*	0.04 ± 0.02	0.06 ± 0.01	0.08 ± 0.03	0.07 ± 0.03	0.06 ± 0.02
ACDC-*LVC*	0.06 ± 0.01	0.06 ± 0.01	0.07 ± 0.01	0.07 ± 0.02	0.05 ± 0.02

5.5 Results and Analysis

In this section, we report results of the benchmark datasets relative to the losses under consideration based on the training strategy explained in section 5.4. The segmentation performances are compared via the 2 usual segmentation metrics: the Dice score (DSC) presented in Table 5.2 (Crum et al., 2006), the Hausdorff distance metric (HD) presented in Table 5.3 (Beauchemin et al., 1998). In addition, we have computed the mean absolute error on the number of instances (connected components) presented in Table 5.4.

5.5.1 Added value of prior-based losses over the Dice loss baseline

From the performance tables, we realize that there is always at least one prior-based loss that is superior to the Dice baseline (denoted by cells with blue background in the tables). Thus, the exploitation of prior-based losses generally has enhanced segmentation performance in 10 out of 12 anatomical objects of the 8 datasets. For example, the Hausdorff loss has registered best performances on brain lesion segmentation tasks (WMH, Isles) and single-organ segmentation datasets. On the other hand, the boundary loss registered performances close to the best case performance on lesion tasks (Isles, WMH). The clDice registers best performances in 1 out of 3 multi-organ segmentation datasets and the size loss got good results on a selection of datasets including WMH, Atrium and Colon.

A close look at the Dice baseline performance over the entire set of datasets (first column in the tables), one can observe that the Prostate is quite challenging since it has the lowest Dice baseline performance. On the contrary, the ACDC dataset is the easiest with the highest Dice accuracy, and the problem of cardiac structure segmentation is well known and has been argued to be almost solved (Bernard et al., 2018). Intuitively, an easy dataset would already register good performance given the simple Dice baseline and one would expect the addition of prior-based losses to have no added value, other than adding to the complexity of the training and degrading system performance. Indeed, the results benchmarked on the ACDC dataset registers little to no added value on the performance

relative to the baseline. Alternatively, if the dataset is too complex such as the case of the Prostate (multi-label segmentation, large organ size imbalance, large number of connected components), customized prior-based losses may be needed to accommodate its characteristics: almost no gain is obtained from prior losses for the Prostate dataset.

5.5.2 Low-level vs. High-level Prior-based Losses

Both Hausdorff and Boundary losses register good performances on most datasets and over all segmentation tasks: brain lesions, single-organs and multi-organ segmentation. The Hausdorff loss has a superiority over the Boundary loss in some dataset cases (Spleen, Hippocampus). For example, for the spleen dataset, the Hausdorff loss has registered best case performance in both dice accuracy (added value of 14 %) and reduced the Hausdorff distances by over 30 % in comparison to the Boundary loss. The superiority of the Hausdorff loss over the Boundary loss is mainly due to the fact that the Hausdorff loss extracts distance maps from both predicted and ground-truth contours, and minimizes the error between the two maps accordingly, whereas the Boundary loss simply fine-tunes the probability distribution via the ground-truth distance maps. Based on this, one can say that since Hausdorff targets optimizing the distance map entity directly between predicted and ground-truth labels, it can guarantee a better mapping between predicted outputs and the ground-truth than the Boundary loss. Despite this significance, the Hausdorff loss is very computationally expensive since it consists in computing the predicted distance maps online while training, which directly affects training time. Hence, one may consider that the Boundary loss may represent a reasonable trade-off between good segmentation performance and computational cost.

Regarding the high-level prior-based losses, results are mixed: the size-based loss can either provide great improvement (e.g., WMH, Atrium, Colon), or much worse results (e.g., ISLES). For example, the size loss registers equivalent performance in the case of the WMH dataset relative to the best case segmentation result, but performs poorly on ISLES, despite the similarity in nature between the two datasets. We hypothesize that this may be due to the overall lesion sizes. A closer look at Table 5.1 showing the meta-data characteristic, we can gather that, on small sized organs (e.g.: WMH, Atrium, Colon), the size loss registers performance either better or equivalent to the Dice baseline. Given datasets that have large size variability (e.g., isles, Prostate, or ACDC), the exploitation of the size loss degrades segmentation performance. This is mainly due to the fact that, generally, the exploitation of the size loss allows the network to learn average sizes of the organs. In the same essence, based on the results, one can see that size loss can not accommodate multi-organ segmentations. The above observations are illustrated in Figure 5.4 showing the Dice performance relative to organ sizes. We note that the datasets where the size loss registered degraded results (red dots) are for those whose organ sizes are of large variability or that include multi-label segmentation. Hence, despite the fact that the size loss was initially customized to accommodate weakly supervised segmentation, it may be useful in full supervision, when the anatomical objects under consideration are very small structures, and occupying a tiny percentage of the overall image as in the case of the WMH dataset.

The clDice has a similar behavior but to a lesser extent. It generally registers better

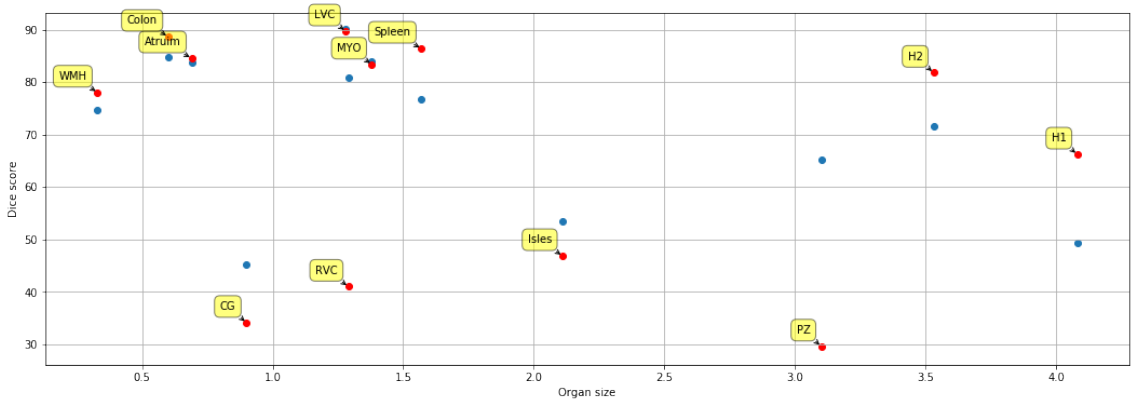


Figure 5.4 – Influence on the organ size on the average Dice score. Each dot represents a dataset. Blue (resp. red) dots show the Dice score obtained with the Dice loss (resp. the Size loss).

performance than the Dice baseline in most single-label segmentation cases and one multi-organ segmentation dataset. However, the clDice loss degraded performance on other datasets such as the WMH and the ACDC. Despite the equivalence in Dice accuracy between the Hausdorff loss and clDice loss on the Hippocampus dataset, the Hausdorff loss outperformed the clDice relative to the Hausdorff distance (Hausdorff loss is about 8 % lower than clDice loss in Hausdorff distance). This indicates its ability to take into consideration shape and border specifications. The degraded performance of clDice on Hausdorff distance can be explained by the fact that the loss is based on the skeleton maps, which tends to blur boundary specifications for the sake of revealing topological properties. This limitation is further verified by the clDice with the Hausdorff distance results on the ACDC dataset. Thus, even when the clDice registered best ranked results relative to the Dice Accuracy, the Hausdorff distance is degraded, even lower than the Dice baseline with regards to the Myocardium, for instance. Given tasks with high border irregularities, such as lesions, failing to consider boundary specifications can hinder overall performance (e.g. case of brain lesion in the WMH dataset).

When studying other meta-data features such as the number of connected components, one can see that the exploitation of high-level prior-based losses does not have a great influence on the results (see Table 5.4). We hypothesize that this may be due to the fact that high-level prior-based losses are rather customized to serve a particular task, or satisfy a particular constraint. If the task at hand does not conform with the dataset characteristics or attributes, the prior-based loss may generally have no added value.

Overall, we can hypothesize that contour-based losses are rather generic, and can be useful for enhancing segmentation performance on any type of dataset. However, if we are aiming at preserving a particular characteristic or anatomical property, a customized high-level prior-based loss may be a feasible solution. Thus, high-level losses may provide improvement; however, they are not very stable and can not be generalized to all datasets and tasks.

5.5.3 Limitations of the Current Proposed Benchmark

Despite our intuitive analysis with regards to some relationships between loss performance and dataset characteristics however, we admit to many limitations. For starters, the proposed benchmark can not be generic, as there are many existent prior-based losses that we fail to include: low-level prior (Caliva et al., 2019; Mosinska et al., 2018; Yang et al., 2018), high-level topological (Clough et al., 2019) or shape prior (Dolz et al., 2017; Mirikharaji and Hamarneh, 2018). Moreover, due to the fact that high-level prior-based losses are customized to target a particular property, providing means of comparison with respect to their effectiveness is subjected to debate. Another key component to take into consideration is their optimization algorithms. Many prior-based losses are discrete in nature; hence, they require particular optimization strategies in order to insure good performances. Our proposed benchmark is based on plugging the considered losses into a penalty-based Lagrangian optimization technique and training via stochastic gradient descent and the ADAM optimizer. On the level of the datasets, despite some similarities between datasets (Lesion task: ISLES, WMH, task: Single vs. Multi), however, the datasets are rather very different, each given a set of characteristics and properties. Hence, there are a lot of variables to take into consideration, which makes the means of comparison often limited. Despite these limitations, presenting a benchmark that can test prior-based losses on different tasks and datasets is important, because it can give the reader an intuitive initial judgment on which loss to choose based on the considered requirements and datasets properties.

5.6 Conclusion

In this chapter, we proposed a benchmark of prior-based losses on medical image segmentation datasets. We provided intuitive explanations on a few existing relationships between prior-based loss significance and dataset characteristics. We summarized the chapter's realizations as follows: the size loss is generally significant when considering datasets of small structures and limited size variability. The contour-based losses generally, and Hausdorff loss particularly, accommodates objects of multi-structures and border irregularities.

Future work includes expanding the proposed benchmark in order to encompass a broader perspective of losses. Moreover, we aim to add other metadata features, in order to better characterize the organ and the task at hand, develop robust similarity feature vectors between datasets for more accurate comparison and conduct meta-learning to predict loss ranks and outputs so as to address the computational complexity issues between losses and their peers. In the next chapter, we will make a proposal for a novel contour-based loss, that allows to take into account contour specifications.

Chapter 6

Perimeter-based Loss for Border Irregularity

“ No problem is hard, it is just a matter of time! ”

Contents

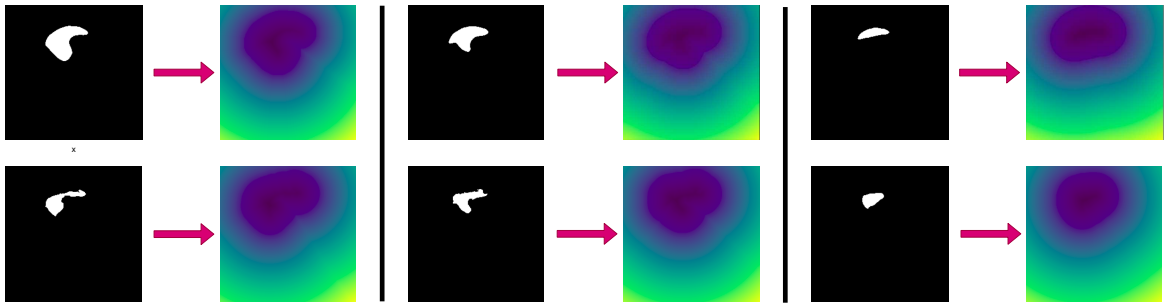


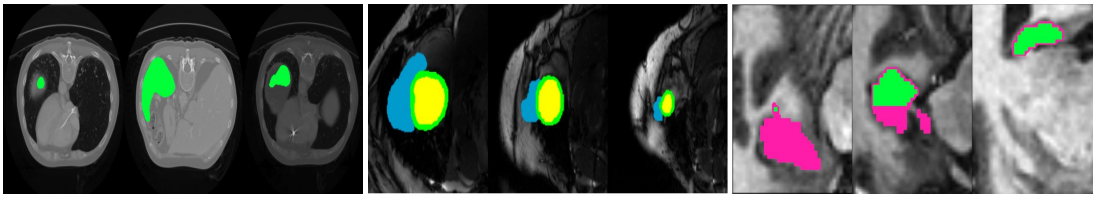
Figure 6.1 – Examples of segmentation maps of organs of different border characteristics and irregularities with their corresponding distance maps. Distance maps are somehow similar even when segmentation maps have different border irregularity specifications.

6.1 Motivation

Incorporating prior information into segmentation encompasses multiple approaches. Thus, it can consist in introducing architecture modifications as demonstrated in Chapter 1, or can be done at the level of the loss function or via a combination of both as in Chapter 4. Compared to architectural constraints, loss constraints have the advantage of being generic with the ability to be plugged into any backbone.

One important branch of loss functions that has taken great interest recently is the contour-based losses. Recent works (Caliva et al., 2019; Karimi and Salcudean, 2019; Kervadec et al., 2019a; Yang et al., 2020) show that contour-based losses could allow for more anatomical plausible segmentation when trained independently or in conjunction with a regional loss, such as the soft Dice approximation (Milletari et al., 2016) or the cross-entropy (Ronneberger et al., 2015). Contour-based losses often aim to minimize directly or indirectly the one-to-one correspondence between points on the predicted and label contours. Therefore, these losses are rather complex in nature and are characterized by hard gradients and high computational cost. Moreover, they often exploit distance maps to represent the change between predicted and ground-truth boundaries (Kervadec et al., 2019a). Distance maps are common transformation tools that have been exploited in many state-of-the-art approaches in order to reveal contour variations between ground-truth and predicted segmentation maps (Caliva et al., 2019; Karimi and Salcudean, 2019; Kervadec et al., 2019a). Despite their popularity, distance maps often tend to underestimate the contour-to-contour distances given that the closest point is chosen systematically. Consequently, they often fail to show differences between border characteristics. The problem becomes particularly challenging when the anatomical object under consideration has a complex shape with concavities or border irregularities. An example of the limitations posed by distance maps is demonstrated in Figure 6.1. Taking a closer look, it is evident that for shapes that have different border irregularity characteristics (e.g. smooth vs deformed surfaces, concavities...), distance maps are somehow similar, varying slightly when the organ is displaced from its position. As a result, the segmentation model may suffer from trivial local solutions or vanishing/exploding gradients (Kervadec et al., 2019a).

In this chapter, a novel contour-based loss is proposed by constraining the perimeter or contour length of the organ to be segmented. Relative to the limitations of current



(a) Spleen (b) ACDC (c) Hippocampus

Figure 6.2 – Sample images from the 3 datasets with ground-truth regions overlaid. (a) Spleen is in green, (b) right ventricle is in blue, left ventricle is yellow and myocardium in green, (c) brain hippocampus with outer (H1) and internal (H2) tissues in pink and green, respectively.

state-of-the-art contour-based losses, the novelty of the proposed loss is two folds. One, it minimizes the overall perimeter-length between predicted and ground-truth contours instead of minimizing the point-by-point distance between each point on the ground-truth contour and its corresponding point on the predicted one. Two, it exploits contour maps, in place of distance maps, that explicitly reveal deformations and irregularities within the organ shapes. Inspired by methods of Shit et al. (2019), we extract contour maps from both the ground-truth and predicted segmentation maps. We then minimize the error between the predicted and ground-truth perimeter-lengths by considering the sum over each respective contour via a mean squared error. Our main motivation behind the proposed loss is that, by targeting the perimeter-length rather than the point-by-point distance, the model will be able to take into consideration border irregularities, such as sudden corners or curvatures within organ shapes. In doing so, it avoids shrunken or expanded anomalies. Moreover, the simplicity of the proposed loss, being the mean squared error between two lengths, may play an important role in allowing the network to learn at a faster rate and with high efficiency.

The proposed loss is evaluated on three public datasets of different tasks and modalities. Recalling from Section 1.5.2, the Spleen dataset is from the Medical Segmentation Decathlon and consists of CT images that target spleen segmentation. The ACDC dataset contains cardiac magnetic resonance images (MRI) and the goal is to segment the two ventricles and the myocardium. The Decathlon Hippocampus dataset consists of segmenting two neighboring small structures in MRI images. These public datasets were chosen in such a way that the structures to be segmented are characterized by particular shapes and non-convexity, as shown in Figure 6.2. We test the significance of the proposed loss in both a single-organ segmentation setting and a multi-organ segmentation setting.

Surprisingly, despite the simplicity of the proposed loss, it outperforms state-of-the-art contour losses for organs with non-convex shapes and maintains segmentation performance for simple shapes such as circles and holes. Moreover, the proposed loss allows for accurate delineation of common boundaries between neighboring organs in the multi-organ segmentation framework.

The rest of the chapter is organized as follows. Section 6.2 elaborates on the proposed loss. Section 6.3 encompasses the experimental settings and analyzes model performance

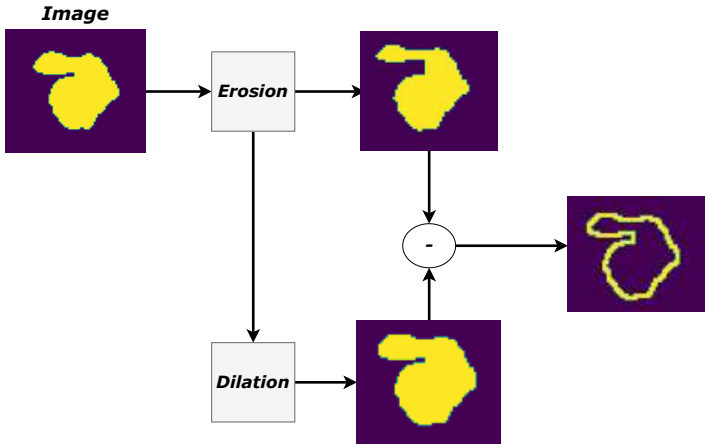


Figure 6.3 – Boundary extraction via morphological gradient: The boundary is obtained via the subtraction of the eroded map of a binary image from the dilated map of the eroded one.

on the three datasets. Finally, Section 6.5 includes the conclusion and future works.

6.2 Perimeter-based Loss

Many works in the field design their losses via distance maps as features in order to integrate geometric and location prior into the learning framework (Caliva et al., 2019; Karimi and Salcudean, 2019; Kervadec et al., 2019a). Alternatively, we exploit contour maps produced via a combination of non-trainable max-pooling and min-pooling layers.

6.2.1 The Contour Function $\mathscr{F}$

In computer vision, boundary extraction is generally conducted via two main morphological operators: dilation and erosion (Rivest et al., 1993). Whereas dilation expands and grows an object in a binary image based on a reference shape, erosion shrinks and thins the object under consideration. Generally, dilation is used in order to fill holes or gaps and enhance features while erosion removes unwanted bridges and protrusions. One way to extract contours in computer vision is through subtraction of the erosion of an image from the dilation of the erosion, as shown in Figure 6.3. This is known as the morphological gradient of an image.

In order to integrate the concept of morphological gradient into the deep learning framework, proper differentiable approximation of the dilation and erosion processes is to be conducted. In this work, we approximate morphological operations by non-trainable neural layers. The implementation of the contour function $\mathscr{F}$, as shown in Figure 6.4 is conducted via subtracting the erosion of the segmentation map from the dilation of the eroded map. Dilation and erosion of the predicted and ground-truth maps are carried out via max-pooling and min-pooling layers, followed by a ReLU layer. We note that $\mathscr{F}$ can extract the contour in any image, whether it is a binary or a probability map. Hence, one can consider the contour function as a simple morphological gradient that can accommodate back-propagation in CNN training.

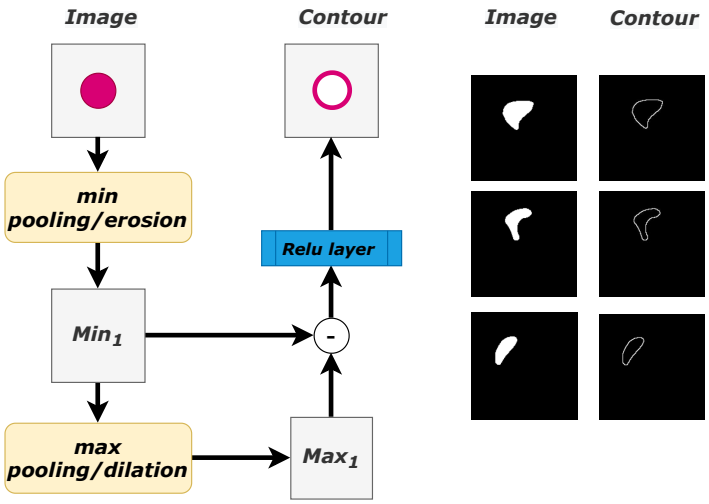


Figure 6.4 – Principle of the contour function $\mathscr{F}$: the difference (-) between the erosion (min-pooling) and the dilation (max-pooling) layers followed by a ReLU layer. Right side: examples of ground-truth segmentation image and respective contour image.

The contour function $\mathscr{F}$ is inspired by the work of Shit et al. (2019) that exploits this extraction strategy in order to integrate skeletonization constraints at the level of the loss function for tabular and vessel segmentation. This extraction process is repeated each time on the transformed image resulting from the previous iteration, consecutively summing over the output of all iterations in order to obtain the desired skeleton maps. After this, they have taken into consideration the intersection over union of the precision and sensitivity between the ground-truth and respective skeleton maps. In our work, we produce the contour maps by considering the subtraction of the erosion of the original segmentation map from the dilation of the eroded, as shown in Figure 6.4. Moreover, we target the perimeter-length of the predicted and ground-truth segmentation maps instead of the point-to-point matching. In order to obtain the perimeter-lengths for both predicted and ground-truth segmentations, summation over the pixels in each of the predicted and ground-truth contour maps is conducted. The proposed loss, to be optimized, is the mean squared error between the predicted perimeter-length and ground-truth perimeter-length. In the following section, we demonstrate the mathematical formulation of the proposed loss and the adopted training strategy.

6.2.2 Loss Formulation

Let $\Omega \subset \mathbb{R}^2$ be the spatial image domain. Let y be the true label map and $\hat{y}$ the predicted label (probability) map, both of dimension $|\Omega|$. Consider $\mathscr{F}$ to be a function that extracts the contour map of any image, as described in the previous section. The proposed loss is defined as follows:

$$\mathscr{L}_{perim} = \Big(\sum_{p \in \Omega} \widehat{y_p^{\mathscr{F}}} - \sum_{p \in \Omega} y_p^{\mathscr{F}} \Big)^2, \tag{6.1}$$

where $y_p^{\mathscr{F}}$ (resp. $\widehat{y_p^{\mathscr{F}}}$) is the value of pixel p in the map $\mathscr{F}(y)$ (resp. $\mathscr{F}(\hat{y})$), equal to y_p

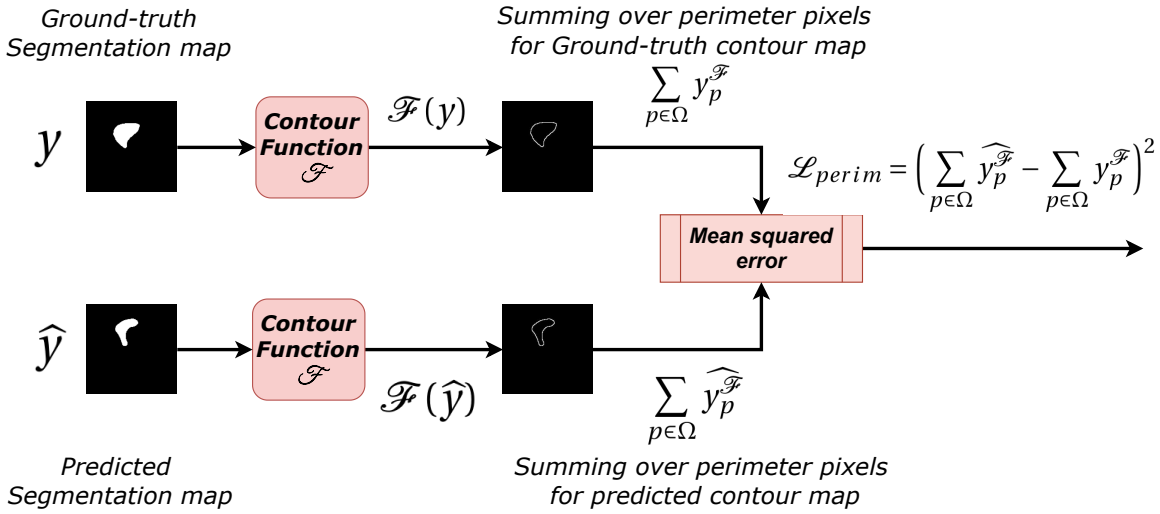


Figure 6.5 – Principle of the perimeter loss function $\mathscr{L}_{perim}$: Contour maps are extracted via the contour function $\mathscr{F}$ for both the predicted and ground-truth segmentation maps. Then, perimeter-lengths are obtained for each criterion by summing over the pixels of the binary contour map. The perimeter-length loss is the mean squared error between the ground-truth perimeter-length and predicted perimeter-length.

(resp. $\widehat{y_p}$), if p belongs to the contour, 0 otherwise. The contour function $\mathscr{F}$ extracts, for each of the predicted and ground-truth segmentation maps, a contour map of the segmented objects. The proposed loss is the mean squared error between the predicted perimeter-length and the ground-truth perimeter-length. The predicted and ground-truth perimeter-lengths are obtained by the summation of pixels in their corresponding contour map represented by the terms $\sum_{p\in\Omega}\widehat{y_p^{\mathscr{F}}}$ and $\sum_{p\in\Omega} y_p^{\mathscr{F}}$ respectively in equation 6.1.

Intuitively, the proposed perimeter-length loss must be trained in conjunction with a regional loss. This is mainly due to the fact that the proposed loss targets the over–all perimeter-length of the organs. Hence, feature-wise, the contour-length is not independently representative of the data. Training the network via the perimeter-length loss alone does not guarantee proper learning and segmentation as will be shown in the chapter's experiments. Therefore, we propose to train the segmentation networks via a convex combination of the Dice loss (Milletari et al., 2016) and the proposed perimeter-based loss weighted by a hyperparameter λ according to the following equation:

$$\mathscr{L} = (1-\lambda)\mathscr{L}_{\mathrm{D}ice} + \lambda\mathscr{L}_{perim}. \tag{6.2}$$

The hyperparameter λ is dynamically updated with values ranging between 0 and 1. At the beginning of the training, and based on the considered dynamic scheduling strategy, the weight corresponding to the perimeter-length loss is first set to 0. In doing so, the network is permitted to learn regional features via the Dice loss. As the training moves forward, the value of λ is successively increased giving more reign to the perimeter-length loss. As a result, shape constraints are imposed, based on the reference perimeter-length a network is supposed to respect while outputting a segmentation of a considered organ. Hence, one can consider the proposed perimeter-based loss as a regularizing term on the

object perimeter, which subsequently affects the organ's shape.

6.2.3 Comparison to the State-of-the-art

In the literature, many research works have attempted to impose constraints at the level of the loss function in segmentation networks. One way to do so is to directly exploit the ground-truth map in order to enhance specific geometric properties, e.g. via distance maps or Laplacian transforms (Bohlender et al., 2021; Ma et al., 2020). With regards to distance maps, two major contributions are the Boundary loss, proposed by Kervadec et al. (2019a), and the Hausdorff loss, proposed by Karimi and Salcudean (2019). Let $\mathrm{D_G}(p)$ denote the distance of pixel p to the closest contour (G) point. The Boundary loss $\mathscr{L}_{Boundary}$ is an approximation of the distance between the real and the estimated boundaries and is defined as:

$$\mathscr{L}_{Boundary} = \sum_{p\in\Omega} \phi_g(p)\hat{y}_p \qquad \text{with } \phi_g(p) = \begin{cases} -\mathrm{D_G}(p) & \text{for } p \text{ inside the target region} \\ \mathrm{D_G}(p) & \text{else} \end{cases} \tag{6.3}$$

with $\hat{y}_p$ being the predicted value at pixel p.

On the other hand, the Hausdorff loss $\mathscr{L}_{\mathrm{HD}}$ in Karimi and Salcudean (2019) conducts a direct point-by-point optimization of the predicted and ground-truth contours, arriving to the following loss term:

$$\mathscr{L}_{\mathrm{HD}} = \frac{1}{|\Omega|} \sum_{p\in\Omega} (y_p - \hat{y}_p)^2 \Big(\mathrm{D_G}(y_p)^2 + \mathrm{D_G}(\hat{y}_p)^2 \Big) \tag{6.4}$$

Both studies tackle the problem of contour optimization between ground-truth and predicted segments, to increase anatomical plausibility in their respective deep learning segmentation models. However, whereas Karimi and Salcudean (2019) conduct a direct point-by-point optimization of the predicted and ground-truth contours, Kervadec et al. (2019a) derive, through Boykov et al. (2001) graph theories, an equivalent term that fine-tunes the probability distribution via ground-truth distance maps. In this way, Kervadec et al. (2019a) alleviate the high computational load demonstrated by Karimi and Salcudean (2019), resulting from the online computation of the predicted distance maps per each iteration and for all images in the dataset. Instead of weighting the probability distributions as in Kervadec et al. (2019a), Caliva et al. (2019) exploit distance maps as weighing factors for a cross-entropy loss term in order to improve extraction of shape bio-markers and enable the network to focus on hard-to-segment boundary regions. As a result, they give more weight to pixels lying in close proximity of the segmented anatomical objects than those that are far away.

All the above approaches exploit distance maps in order to reveal contour variations between predicted and ground-truth segments. Alternatively, we exploit contour maps mimicking morphological gradients, that reveal subtle differences with respect to border smoothness and irregularities, as shown in Figure 6.4 on the left. Moreover, instead of taking into consideration the distance between points on the predicted and ground-truth contours, later on transformed to distance maps formulations, the proposed loss considers the perimeter-length between the predicted and the ground-truth segmentation maps. We believe that the proposed method works well on shapes that have varying

curvatures and boundary particularities, because the proposed loss is based on a contour function that can better express the variations between predicted and ground-truth contours than other representations such as distance maps. Moreover, by directly minimizing the contour-length, the proposed method can accommodate datasets where in fact the perimeter has particularities and specifications.

Alternative to distance maps or the proposed contour maps, Yang et al. (2019) exploit Laplacian filters in order to develop a boundary enhanced loss term that invokes the network to generate strong responses around the boundary areas of organs while producing a zero response given pixels that are farther from the peripheries. In the same context, Arif et al. (2018) extend the regular cross-entropy term with an average point to curve Euclidean distance factor between predicted and ground-truth contours in order to allow the network to consider shape specifications of segmented structures.

6.3 Experiments

6.3.1 Experimental Setting

We deploy the unified U-Net framework proposed in (Kervadec et al., 2019a, 2018, 2020a) and modify the loss function accordingly. Training is done using a batch size of 8 and a learning rate of 10^{-3}. The learning rate is halved if the validation performance does not improve during 20 epochs. In the implementation of the contour function $\mathscr{F}$, max-pooling and min-pooling were carried out via a kernel of size (3,3) and stride of 1. Since the contour extraction function is mainly composed of non-trainable layers, there is no considerable addition to the complexity of the network or the computational cost. Our code is public and available on GitHub [1].

The U-Net is trained with the loss as defined in Eq. 6.2. The hyperparameter λ is dynamically updated via the same scheduling strategy conducted in Kervadec et al. (2019a) and described in Section 6.2.2. Thus, the parameter was set to 0.01 and increased by 0.01 per epoch for 200 epochs for the Spleen and ACDC datasets and for 45 epochs for the Hippocampus dataset.

For pre-processing, we have resized the images to a size of 256×256 and normalized them to the range [0, 1]. We have kept negative samples for training and validation. Each dataset was split into train and validation based on an 80 % / 20 % partition respectively and validated via three Monte-Carlo simulations (Arlot and Celisse, 2010).

The proposed perimeter-based loss is evaluated in two segmentation settings: a single-organ segmentation setting, where we intend to train the network on the Spleen dataset and Right Ventricular structure (RVC) of the heart in the ACDC dataset; and a multi-organ segmentation setting, where simultaneous segmentation of the anatomical objects of the Hippocampus and ACDC datasets is carried out. The proposed method, which consists of training the network via the perimeter loss in conjunction with the Dice loss according to the dynamic strategy in Section 6.3, is compared to two lower baselines. One, the independent training of the Dice loss. Two, the independent training of the proposed perimeter loss. Relative to the state-of-the-art in contour-based losses, we compare to

[1] https://github.com/rosanajurdi/Perimeter_loss

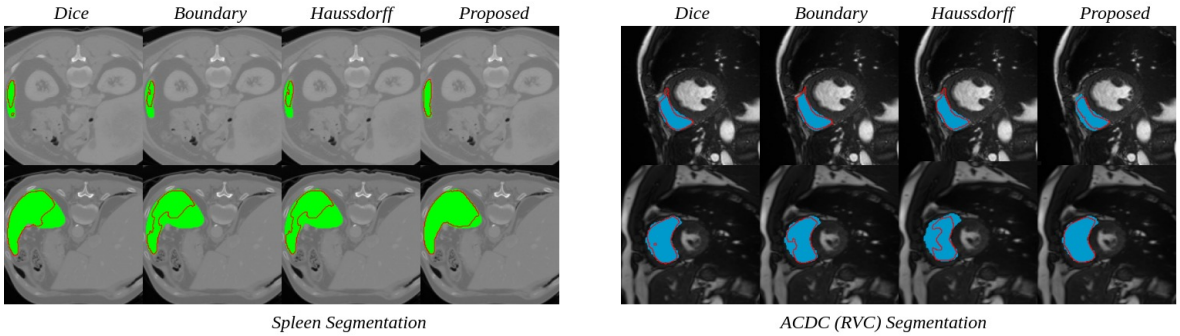


Figure 6.6 – Segmentation results of the proposed loss against the Dice loss baseline and state-of-the-art losses in red with the ground-truth as a region filled with green for spleen and blue for the right ventricle (RVC) segmentation. Each row is a different image.

Table 6.1 – Mean (± std) Dice index (%) and Hausdorff distance (mm) on the Spleen and RVC from ACDC datasets

	Spleen Dataset		***ACDC Dataset (RVC)***	
Loss	Dice index	Hausdorff	Dice index	Hausdorff
$\mathscr{L}_{Dice}$	76.80 ± 7.59	1.33 ± 0.28	81.22 ± 1.05	2.47 ± 0.04
$\mathscr{L}_{perim}$	58.98 ±11.42	1.89 ±0.35	29.34 ±11.83	4.21 ±0.49
$\mathscr{L}_{Dice} + \mathscr{L}_{Boundary}$	80.38 ± 5.46	1.34 ± 0.21	81.73 ± 0.81	2.35 ± 0.01
$\mathscr{L}_{Dice} + \mathscr{L}_{HD}$	91.79 ± 2.67	0.92 ± 0.15	81.47 ± 1.01	2.42 ± 0.05
$\mathscr{L}_{Dice} + \mathscr{L}_{clDice}$	87.15 ± 13.61	1.07 ± 0.53	80.61 ± 1.27	2.51 ± 0.09
$\mathscr{L}_{Dice} + \mathscr{L}_{Size}$	86.44 ± 15.86	1.27 ± 0.57	84.72 ± 1.78	2.24 ± 0.09
$\mathscr{L}_{Dice} + \mathscr{L}_{perim}$	**95.39 ± 1.26**	**0.71 ± 0.07**	**85.67 ± 0.50**	**2.21 ± 0.09**

the Boundary loss in Kervadec et al. (2019a), and the Hausdorff loss in Karimi and Salcudean (2019) in conjunction with the Dice loss according to the same dynamic strategy stated above. On a broader scope of prior-based losses, we also compare relative to shape related losses such as the size loss (Kervadec et al., 2018) and the clDice loss (Shit et al., 2019) under the same training strategy.

6.3.2 Single Organ Segmentation

From a closer look at Table 6.1 (second row), we can verify that the perimeter based loss cannot learn on its own since its main objective is to optimize the perimeter of the organ and regularize the organ boundary. Hence, the optimization criterion and features are rather insufficient to allow proper training. This observation is verified for all organs and datasets.

Training the network via the convex combination of the proposed perimeter-length loss and the Dice loss on the spleen organ, showed that the proposed loss outperforms the lower baselines and the state-of-the-art Boundary losses by a large margin. When compared to the best state-of-the-art performance, the proposed perimeter-based loss

Table 6.2 – Mean (± std) Dice index (%) and Hausdorff distance (mm) on the simpler ACDC structures (MYO: myocardium, LVC: left ventricular cavity) trained independently in single-organ segmentation.

	Myocardium		***Left Ventricular Cavity***	
Loss	Dice index	Hausdorff	Dice index	Hausdorff
L_{Dice}	87.82 ± 1.02	2.54 ± 0.02	91.78 ± 0.89	1.92 ± 0.02
$L_{Dice} + L_{Boundary}$	86.53 ± 0.71	2.55 ± 0.01	91.13 ± 0.51	1.93 ± 0.00
$L_{Dice} + L_{HD}$	84.68 ± 0.59	2.61 ± 0.06	89.91 ± 0.71	1.97 ± 0.05
$L_{Dice} + l_{clDice}$	86.82 ± 0.50	2.61 ± 0.01	89.65 ± 0.81	2.04 ± 0.03
$L_{Dice} + L_{Size}$	86.76 ± 0.90	2.52 ± 0.01	91.40 ± 0.10	1.89 ± 0.02
$L_{Dice} + L_{perim}$	87.76 ± 0.34	2.55 ± 0.03	92.00 ± 0.53	1.92 ± 0.02

has registered an increase by about 4% on the Dice index, and a decrease in about 20% on the Hausdorff distance. Since the spleen is an organ characterized by a concave border, we hypothesize that the significant decrease in Hausdorff distance highlights the ability of the proposed loss in accounting for varying curvature and border irregularities. This is illustrated qualitatively in Figure 6.6, where example images of predictions based on the perimeter loss are able to capture the specifications of the spleen contour much more efficiently than the other contour-based losses.

Our next objective was to pinpoint the significance of the proposed loss on a broader scope of prior losses, particularly, relative to shape related prior losses such as the clDice loss in Shit et al. (2019) and size loss in Kervadec et al. (2018) (third group row entry in Table 6.1). Relative to the tables, the significance of the proposed loss is quite evident. Despite the fact that both clDice and perimeter-length losses use the same contour extraction strategy, however, the proposed loss outperforms the clDice loss under the same dynamic training framework by about 7 %. This is mainly due to the fact that the skeleton maps, derived from contour maps in Shit et al. (2019), do not take into account organ borders, but they target topological characteristics meaning the number of connected components or holes. Similarly, the size loss targets the organ size without taking into account their border specifications. Hence, this verifies the ultimate relationship existing between prior losses and the features they optimize relative to dataset characteristics.

Given that the RVC has a concave shape similar to the spleen shape, we anticipate a similar behavior of the loss performance. Indeed, from Table 6.1, we gather that the proposed loss outperforms the best boundary state-of-the-art loss by 4% in Dice index and by more than 6% (from 2.35 to 2.21) in Hausdorff distances. In comparison to shape-related prior losses, the proposed loss exhibits the same significance over the clDice as the case of the Spleen dataset and comparable results relative to the size loss. From Chapter 5, size losses work best under full supervision if the organ size is rather very small relative to the overall background distribution. Consequently, the RVC organ has a size distribution of $1.48\ \% \pm 1.50$, verifying yet again the relationship existing between the proposed loss and the features optimized relative to dataset characteristics. By conducting further experiments where the other cardiac structures were segmented independently, we observe

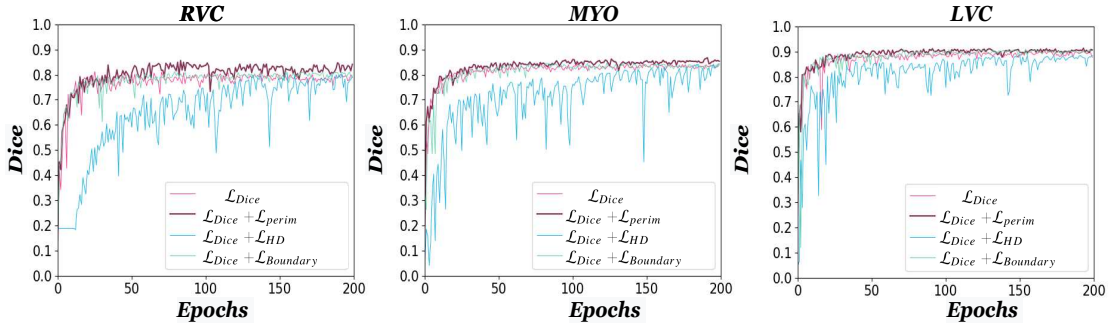


Figure 6.7 – Learning curves of Dice index on the three cardiac structures of the ACDC dataset in the multi-organ segmentation. RVC/LVC: right/left ventricular cavity, MYO: myocardium

Table 6.3 – Dice index and Hausdorff distance (mm) results for ACDC (*simultaneous*) segmentation. RVC: right ventricular cavity, MYO: myocardium, LVC: left ventricular cavity

	Dice index			Hausdorff Distance		
Loss	RVC	MYO	LVC	RVC	MYO	LVC
$\mathcal{L}_{Dice}$	80.79 ± 0.95	83.92 ± 0.13	90.26 ± 0.13	2.44 ± 0.04	2.60 ± 0.01	1.95 ± 0.02
$\mathcal{L}_{perim}$	23.19 ± 13.38	26.53 ± 9.63	9.78 ± 4.07	4.29 ± 0.17	4.59 ± 0.43	4.39 ± 0.05
$\mathcal{L}_{Dice}$ + $\mathcal{L}_{Boundary}$	81.04 ± 0.87	84.16 ± 0.83	89.53 ± 0.74	2.41 ± 0.05	2.57 ± 0.01	1.95 ± 0.02
$\mathcal{L}_{Dice}$ + $\mathcal{L}_{HD}$	80.54 ± 1.30	83.91 ± 0.85	88.98 ± 0.90	2.33 ± 0.04	2.65 ± 0.01	1.98 ± 0.01
$\mathcal{L}_{Dice}$ + $\mathcal{L}_{clDice}$	83.83± 1.39	83.24 ± 0.66	89.56 ± 1.10	2.34 ± 0.08	2.71 ± 0.04	1.98 ± 0.04
$\mathcal{L}_{Dice}$ + $\mathcal{L}_{Size}$	41.02 ± 38.39	83.41 ± 0.72	89.74 ± 0.71	3.88 ± 1.44	2.62 ± 0.00	1.94 ± 0.01
$\mathcal{L}_{Dice}$ + $\mathcal{L}_{perim}$	**84.49 ± 0.57**	**86.22 ± 0.41**	**90.69 ± 0.41**	**2.19 ± 0.03**	**2.55 ± 0.04**	**1.94 ± 0.02**

that the proposed loss maintains state-of-the-art performance when trained to segment organs with simpler shapes like circles (MYO) or holes (LVC) (See Table 6.2).

6.3.3 Multi-organ segmentation

We have benchmarked the performances of the proposed loss on the Hippocampus dataset, which is composed of two neighboring structures, and on the ACDC datasets with the 3 cardiac structures. Multi-label segmentation performance on ACDC, as shown in Table 6.3, reveals that training the model via the perimeter-based loss in conjunction with the Dice loss not only allowed improved segmentation on the irregular shaped RVC, but also on the LVC and MYO as well. This shows that improving segmentation performance on the hard-to-segment LVC has permitted proper delineation of other organs in its neighborhood. These observations were further validated by the evolution plot of the Dice index of the 3 cardiac structures; as shown in Figure 6.7 given the 3 state-of-the-art losses against our proposed loss. By comparing with respect to the Dice accuracy, the proposed loss registers comparable performance relative to the clDice, which may indicate similar significance or behavior on the 3 cardiac structures, however, a closer look at the Hausdorff distance indicates that, even though the Dice accuracy is similar, the proposed loss registers lower Hausdorff distances. This decrease in Hausdorff distance on the

Table 6.4 – Dice index (%) and Hausdorff distance (mm) results for the Hippocampus Dataset. H1: green tissue, H2: pink tissue

	Dice index		Hausdorff Distance	
Loss	H1	H2	H1	H2
$\mathscr{L}_{\text{Dice}}$	49.38 ± 0.33	71.70 ± 1.30	3.82 ± 0.14	2.31 ± 0.05
$\mathscr{L}_{perim}$	16.60 ± 10.06	36.21 ± 1.68	7.97 ± 6.33	3.10 ± 0.12
$\mathscr{L}_{\text{Dice}} + \mathscr{L}_{\text{Boundary}}$	65.20 ± 0.31	81.33 ± 0.74	3.09 ± 0.01	1.99 ± 0.01
$\mathscr{L}_{\text{Dice}} + \mathscr{L}_{\text{HD}}$	68.54 ± 1.46	**82.12 ± 0.44**	2.97 ± 0.05	1.98 ± 0.03
$\mathscr{L}_{\text{Dice}} + \mathscr{L}_{clDice}$	68.39 ± 2.60	**82.82 ± 1.22**	3.20 ± 0.18	1.99 ± 0.04
$\mathscr{L}_{\text{Dice}} + \mathscr{L}_{\text{Size}}$	66.24 ± 0.33	81.84 ± 0.63	3.07 ± 0.01	1.97 ± 0.02
$\mathscr{L}_{\text{Dice}} + \mathscr{L}_{perim}$	**69.06 ± 0.89**	**82.65 ± 0.07**	**3.00 ± 0.04**	**1.95 ± 0.01**

three structures reveals the ability of the proposed loss to preserve shape better than with clDice. As discussed in the previous chapter, the clDice loss is based on skeleton maps that blur boundary specifications, thus justifying the superiority of the perimeter-length loss.

The validity of the perimeter-length loss for multi-organ segmentation is further verified by its performance on the Hippocampus dataset shown in Table 6.4. Thus, the proposed loss produces comparable results for both tissues when compared relative to the state-of-the-art boundary and high-level prior-based losses. Namely, the proposed loss resulted in an increase in Dice accuracy of 4 % on the first hippocampus tissue and about 1 % on the second when compared to the Boundary loss and registers comparable performance with respect to the clDice loss. This verifies the ability of the proposed loss to properly delineate neighboring structures relative to other contour-based losses in the domain. The same conclusion with regards to the difference in Hausdorff distances between the clDice and perimeter-length loss could also be drawn for the Hippocampus dataset. Since one of the tissues of the Hippocampus dataset (tissue H1) is multi-connected, it is significant to study the performance of the proposed loss relative to the different connected components. For this reason, we have constructed a bar-graph (See Figure 6.8) showing the Dice accuracy of the proposed loss and its state-of-the-art peers relative to the number of connected components constituting the dataset. We note that a multi-connected component value of 0 refers to an empty slice where neither hippocampus tissues are present. From Figure 6.8, we can gather that the perimeter-length loss admits an added value even given the multi-connectedness feature. Nevertheless, even with good performance, this does not guarantee that the proposed loss can respect the multi-connected feature of the dataset.

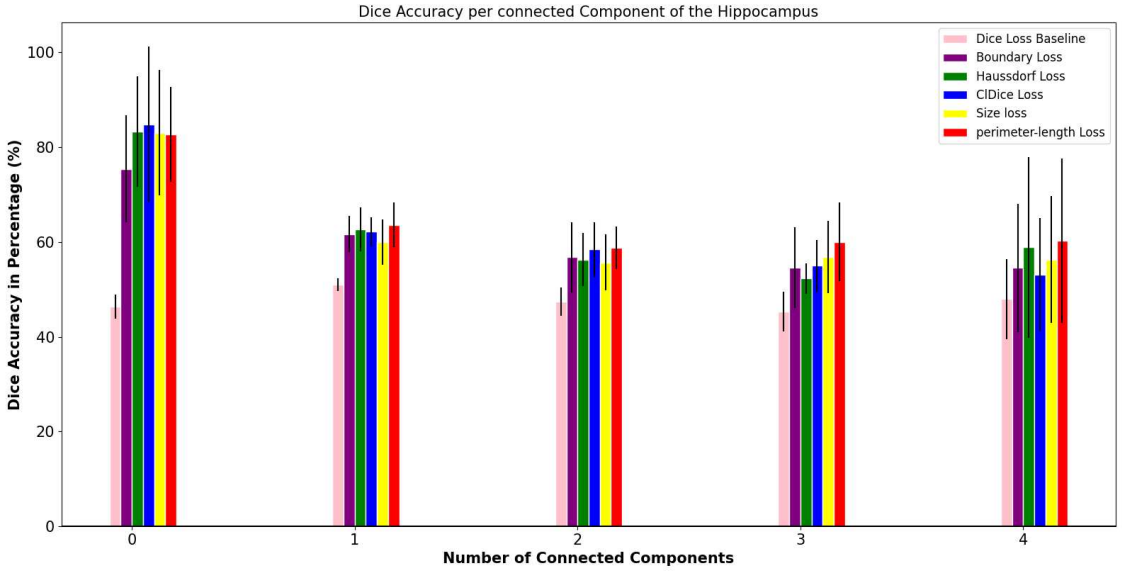


Figure 6.8 – Dice accuracy on the Hippocampus datasets per the number of connected components, where 0 refers to an empty slice.

6.4 Notes on the Behavior of the Perimeter-length Loss

Despite the evident significance of the perimeter-length loss, we note two main observations. Firstly, the perimeter-length loss is most impactful given anatomical objects that have high border and shape irregularities. By computing the compactness ratio (Dolz et al., 2017) between the two tissues, we realize that the H1 tissue, of compactness 12.22 ± 9.75 (average $\pm$ std), is rather more complex in shape relative to the H2 tissue, of compactness 2.36 ± 0.65. The perimeter-length loss had more effect on the anatomical object with higher border irregularity than that of the simpler loss. In the same essence, the proposed loss maintains segmentation performance on the simple ACDC structures (Myocardium and Left Ventricular Cavity), given a single-organ segmentation setting.

Secondly, the perimeter-length loss is sensitive to multi-connected components. This is revealed by comparing the added value of the proposed loss to segmentation performance relative to the ACDC dataset on one hand, and the Hippocampus dataset on the other. The significant, but limited effect of the proposed loss on the Hippocampus dataset can be explained by the fact that the elements within the hippocampus dataset is of multi-connected components (instances in deep learning terms). Thus, by optimizing the perimeter-length, the nature of the perimeter-length loss will generally, favor the segmentation of single compact entities instead of the many expected structures thus limiting segmentation performance.

6.5 Conclusion

In this chapter, we proposed a novel contour-based loss for medical image segmentation. The proposed loss considers the perimeter-length of the segmented organ, instead

of exact boundary matching as it is usually done in present state-of-the-art contour-based losses in the field. We evaluated the proposed loss on three different datasets, that are characterized by varying and non-convex shapes. Despite the simplicity of the proposed loss, it was able to exceed state-of-the-art boundary loss performances by a large margin for both single-organ and multi-organ segmentations.

Future work includes investigating different weighting strategies for the multi-organ segmentation setting under the proposed loss based on the degree of border irregularity of the considered organs. Moreover, a robustness analysis in order to study the effect of perimeter-length variation on system performance. This can be done by fine-tuning the kernel parameter values within the non-trainable layers of the contour function. Finally, the proposed loss is sensitive to anatomical objects with small connected components. Therefore, the proposed loss is to be adapted in order to accommodate multi-connected structures given tasks such as brain lesions.

General Conclusion and Perspectives

" And to every end, there is a better, brighter beginning. "

In this thesis, we have proposed methods of constraining Convolutional Neural Networks via prior knowledge for medical image segmentation. Our main general contribution, that we revisited in each chapter, was showing just how much the integration of prior knowledge into automatic segmentation frameworks can help bridge the gap between artificial network performance and human-level intelligence.

Generally, prior can be integrated at the level of the network architecture, or at the level of the loss function, or via a combination of both. To this end, we have first focused on priors that could be integrated into loss functions, and have proposed to distinguish between low-level priors, that involve pixelwise transformations of the ground-truth labels or the input image, and high-level priors, representing features closer to the spatial or anatomical characteristics of the object to be segmented. In this regard, we have summarized, through a survey in Chapter 2, the state-of-the-art methods concerning high-level prior-based losses and have shed light on the challenges involved, particularly concerning prior modelling and the optimization strategies.

Then, we have proposed two methods that integrate location and shape prior at the level of network structure in Chapters 3 and 4: the BB-UNet and CoordConv-Unet. Both models integrate shape and location prior internally within the model structure. However, whereas BB-UNet does so via bounding boxes, and an additional convolutional block inserted at the level of the skip-connections, CoordConv-Unet conditions selective convolutional layers in a U-Net on coordinate extensions. In this regard, BB-UNet model increased segmentation performance by permitting the extraction of focused topological and shape features with promising tendency to accommodate learning given incomplete labels. On the other hand, CoordConv-Unet plays a dual role in either stabilizing or enhancing segmentation performance under prior-based losses.

In addition to structural constraint, prior can also be integrated at the level of the loss function. In this context, we investigated the performance of state of the art prior-based losses for medical image segmentation tasks via a benchmark presented in Chapter 5. We have also reflected on some intuition regarding the behavior of these losses relative to the datasets' characteristics. Finally, through our analysis of the proposed benchmark, we realized a limitation of current contour-based state-of-the-art losses that includes their high-complexity due to their exact boundary matching property and their tendency to underestimate exact contour particularities due to their reliance on distance maps. To

address these limitations, we have proposed in Chapter 6 a novel loss that targets the perimeter-length of organs instead of direct boundary matching and is based on contour maps instead of distance maps. Despite its simplicity, being the mean squared error between two scalars, the proposed loss allowed production of segmentation results that better satisfied border and shape characteristics while still being efficient. Our proposed loss re-signifies the usefulness of perimeter-based regularizers in deep learning models for segmentation. The contributions and experiments conducted in this thesis leaves open perspectives for future research which we discuss in the following.

Methodology and Design Perspectives

Prior knowledge for weakly supervised learning. Pieces of prior knowledge present powerful tools with regards to their ability to compensate for the need for full annotations as shown in some recent papers (Ganaye et al., 2019; Kervadec et al., 2018, 2020a). We believe that the methods provided in this thesis are thus useful in weakly and semi-supervised segmentation frameworks. For starters, the potential of BB-UNet exceeds the circumferences imposed by our experiments onto weakly supervised setting where only bounding boxes are given. Such extension can be done via an unsupervised learning setting, for example via an Expectation-Maximization process and iterative refining (Papandreou et al., 2015). In such a way, BB-UNet may very well compensate for the need of full labels or the ancillary/primary model dependency, thus allowing inference of labels from bounding boxes with least complexity. Moreover, the significance of the existing relationship between CoordConv-Unet, which provides absolute locations, and the size loss, which provides relative distances, allows us to presume that CoordConv-Unet may reveal promising behavior when trained under a proper distance-based loss that takes into consideration the inter-distance between a seed label and the average organ periphery for a weakly supervised setting. In addition, information revealed on organ border characteristics may provide proper grounds for when boundary labels are for instance incomplete or missing. The problem when using seeds as labels is that shape or border specifications are not captured. The perimeter-length loss may come in handy given incomplete dataset labeling, as it does not require the full ground-truth to function.

Prior knowledge for data augmentation and generation of anatomically plausible medical images. BB-UNet can also be exploited for anatomically plausible medical image generation that are close to real medical images. We believe that the focused feature properties that the BB-UNet acquires via the BBConV layer leaves promising expectations regarding the possibility of producing more anatomically plausible medical images when integrated into a generative adversarial network (GAN) model (Pan et al., 2019). In this way, bounding box prior can enable generation of images with high resolution features and can preserve topological and anatomical properties when integrated into the generator backbone of a generative adversarial network. This is useful for data augmentation in order to compensate for the lack of sufficient data (Shin et al., 2018).

Prior knowledge for label synthesis. Alternatively, the near optimal fully supervised segmentation results of the BB-UNet structure, allows us to believe in its ability to be ex-

ploited for label synthesis. Still in the scope of GANs, synthesized labels that are close to real ones could be obtained by injecting a BB-UNet into the generator section of a GAN network and comparing the distribution of labels generated with a tiny subset of ground-truth distribution (Shamsolmoali et al., 2021). In this way, we would be able to transform bounding boxes into authentic and accurate labels, which may very well compensate for the need for large amounts of labeled training data needed in deep learning.

Meta-Learning to improve the match between datasets and losses. Given the broad and large scope of emerging prior-based losses, each loss has its own characteristics, complexity and behavior. As shown in Chapter 5, finding relations between datasets and loss behavior is very challenging since there are alot of variables to take into account in terms of dataset features and loss properties. Hence, it is important to design automated systems that can predict loss performance given dataset characteristics and provide better meta-features to describe the datasets. In this context, the idea of using meta-learning in order to predict loss performance comes to mind (Cheplygina et al., 2017; van Sonsbeek and Cheplygina, 2020). One way to use meta-learning for loss performance prediction based on the dataset characteristics is to consider each dataset as a sample composed of a set of features. Given a large number of datasets and their corresponding performance on particular losses, a dataset of datasets is composed where each dataset is a sample in the input space along with its label corresponding to the loss value. Training a meta-learner via this dataset of datasets, one would be able to determine the similarity between the datasets on one hand, and anticipate which loss would perform best on the other.

Dynamic weighting of loss functions. Our contributions within the perimeter-length loss strengthen claims on the importance of exploiting the prior knowledge at the level of the loss function particularly to enhance satisfaction of constraints on border characteristics. However, the proposed loss's significance may very well depend on the level of organ's border irregularity. This is particularly interesting given multi-organ segmentation with organs of varying border specifications. The exploitation of a dynamic loss function that initializes the weights of the perimeter-length loss based on the degree of border irregularity comes to mind. The presence of specific indicators of border irregularities is also of vital importance. We believe that with proper border irregularity indicators, we can hence finetune the proposed loss to increase model performance.

Broader Perspectives

On the design of prior-based losses. The design of the prior loss is facing requirements, concerning the differentiability of the loss terms, or at least the computational complexity that must remain reasonable. Lately, as described in Chapter 2, some advances have been made that explored optimization techniques such as ADMM or the barrier functions and that have allowed to incorporate loss terms, that are not directly optimizable by SGD. We believe future progress will originate from using advanced constraint optimization techniques, stemming from both equality or inequality constraints optimization.

Variational and optimization based approaches for image segmentation from the pre-deep learning era can also provide key pointers on how to model prior information regarding object shape and appearance. Researchers can rely on decades of works on the

topic to find inspiration to design losses for their segmentation problem. In this regard, the majority of the priors are only presented in the case of binary segmentation, except for the inter-regions priors obviously. There is still work to be done to address multiple organ segmentation, with priors that are dedicated to organ shape, size or topology.

Performance assessment. Based on our surveillance of state-of-the-art prior losses in Chapter 2 and Chapter 5, we realize that prior losses may not improve the segmentation results, when measured by generic metrics such as the Dice coefficient. In this case, many works propose specific metrics, which are in close relation with the proposed losses, and thus are not generic. However, the prior losses in general generate more plausible regions that are hardly measurable by the generic metrics. There is a wide variety of metrics for medical image segmentation (Taha and Hanbury, 2015) that reviews overlap based, volume based, pair-counting based, information theoretic based, probabilistic based, and spatial distance based metrics - but no metric related to anatomical plausibility is reported. To our knowledge there is no metric that conveys the notion of anatomical plausibility; however we believe this is an important, open topic, to address in the future, to show the benefits of these prior-based losses.

Bibliography

Araújo, T., G. Aresta, A. Galdran, P. Costa, A. M. Mendonça, and A. Campilho (2018). "UOLO - Automatic Object Detection and Segmentation in Biomedical Images". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer International Publishing, pp. 165–173 (cit. on p. 47).

Arif, A., S. M. M. Rahman, K. Knapp, and G. Slabaugh (2018). "Shape-Aware Deep Convolutional Neural Network for Vertebrae Segmentation". In: *Computational Methods and Clinical Applications in Musculoskeletal Imaging*, pp. 12–24 (cit. on pp. 16, 19, 27, 92).

Arlot, S. and A. Celisse (2010). "A Survey of Cross-Validation Procedures for Model Selection". In: *Statistics Surveys* 4.none, pp. 40–79 (cit. on pp. 67, 80, 92).

Ayed, I. B., S. Li, A. Islam, G. Garvin, and R. Chhem (Mar. 2008). "Area Prior Constrained Level Set Evolution for Medical Image Segmentation". In: *Medical Imaging 2008: Image Processing*. Vol. 6914. International Society for Optics and Photonics, p. 691402 (cit. on p. 12).

Baumgartner, C. F., L. M. Koch, M. Pollefeys, and E. Konukoglu (2017). "An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation". In: *STACOM@MICCAI* (cit. on p. 11).

Beauchemin, M., K. P. B. Thomson, and G. Edwards (1998). "On the Hausdorff Distance Used for the Evaluation of Segmentation Results". In: (cit. on p. 81).

BenTaieb, A. and G. Hamarneh (2016). "Topology Aware Fully Convolutional Networks for Histology Gland Segmentation". English. In: *MICCAI*. Vol. 9901, pp. 460–468 (cit. on pp. 17, 27, 28, 34, 35, 38).

Bernard, O., A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, and P. Jodoin (Nov. 2018). "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?" In: *IEEE Transactions on Medical Imaging* 37.11, pp. 2514–2525 (cit. on pp. 78, 81).

Bohlender, S., I. Oksuz, and A. Mukhopadhyay (Jan. 2021). "A Survey on Shape-Constraint Deep Learning for Medical Image Segmentation". In: *arXiv:2101.07721 [cs, eess]* (cit. on p. 91).

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge: Cambridge University Press (cit. on pp. 17, 30).

Boykov, Y., O. Veksler, and R. Zabih (2001). "Fast Approximate Energy Minimization via Graph Cuts". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.11, pp. 1222–1239 (cit. on pp. 12, 16, 75, 91).

Boykov, Y. and G. Funka-Lea (2006). "Graph Cuts and Efficient ND Image Segmentation". In: *International journal of computer vision* 70.2, pp. 109–131 (cit. on p. 12).

Byrne, N., J. R. Clough, G. Montana, and A. P. King (2021). "A Persistent Homology-Based Topological Loss Function for Multi-Class CNN Segmentation of Cardiac MRI". en. In: *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges*. Ed. by E. Puyol Anton, M. Pop, M. Sermesant, V. Campello, A. Lalande, K. Lekadir, A. Suinesiaputra, O. Camara, and A. Young. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 3–13 (cit. on pp. 27, 28, 31, 38).

Caliva, F., C. Iriondo, A. M. Martinez, S. Majumdar, and V. Pedoia (July 2019). "Distance Map Loss Penalty Term for Semantic Segmentation". In: *International Conference on Medical Imaging with Deep Learning – Extended Abstract Track*. London, UK (cit. on pp. 16, 19, 27, 84, 86, 88, 91).

Chahal, P. K., S. Pandey, and S. Goel (Aug. 2020). "A Survey on Brain Tumor Detection Techniques for MR Images". In: *Multimedia Tools and Applications* 79.29, pp. 21771–21814 (cit. on p. 27).

Charoenphakdee, N., J. Lee, and M. Sugiyama (2019). "On Symmetric Losses for Learning from Corrupted Labels". In: *ICML, PMLR*. PMLR, pp. 961–970 (cit. on p. 10).

Chen, C., C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert (2020). "Deep Learning for Cardiac Image Segmentation: A Review". In: *Frontiers in Cardiovascular Medicine* 7, p. 25 (cit. on pp. 26, 27).

Cheplygina, V., M. [Bruijne], and J. P. Pluim (2019). "Not-so-Supervised: A Survey of Semi-Supervised, Multi-Instance, and Transfer Learning in Medical Image Analysis". In: *Medical Image Analysis* 54, pp. 280–296 (cit. on p. 13).

Cheplygina, V., P. Moeskops, M. Veta, B. Dashtbozorg, and J. P. W. Pluim (2017). "Exploring the Similarity of Medical Imaging Classification Problems". In: *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. Ed. by M. J. Cardoso, T. Arbel, S.-L. Lee, V. Cheplygina, S. Balocco, D. Mateus, G. Zahnd, L. Maier-Hein, S. Demirci, E. Granger, L. Duong, M.-A. Carbonneau, S. Albarqouni, and G. Carneiro. Cham: Springer International Publishing, pp. 59–66 (cit. on p. 101).

Chouzenoux, E., M.-C. Corbineau, and J.-C. Pesquet (2019). "A Proximal Interior Point Algorithm with Applications to Image Processing". In: *Journal of Mathematical Imaging and Vision* (cit. on p. 30).

Çiçek, Ö., A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger (Oct. 2016). "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". In: *MICCAI*. Vol. 9901. LNCS, pp. 424–432 (cit. on p. 10).

Clough, J., N. Byrne, I. Oksuz, V. A. Zimmer, J. A. Schnabel, and A. King (2020). "A Topological Loss Function for Deep-Learning Based Image Segmentation Using Persistent Homology". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1 (cit. on pp. 28, 31).

Clough, J. R., I. Oksuz, N. Byrne, J. A. Schnabel, and A. P. King (Jan. 2019). “Explicit Topological Priors for Deep-Learning Based Image Segmentation Using Persistent Homology”. In: *arXiv:1901.10244 [cs]* (cit. on pp. 15, 17, 27, 31, 32, 36–38, 84).

Cootes, T., D. Cooper, C. Taylor, and J. Graham (1995). “Active Shape Models - Their Training and Application”. In: *Computer Vision and Image Understanding* 61.1, pp. 38–59 (cit. on p. 12).

Cremers, D., M. Rousson, and R. Deriche (Apr. 2007). “A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape”. en. In: *International Journal of Computer Vision* 72.2, pp. 195–215 (cit. on p. 11).

Crum, W. R., O. Camara, and D. L. G. Hill (2006). “Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis”. In: *IEEE Transactions on Medical Imaging* 25.11, pp. 1451–1461 (cit. on p. 81).

Dai, J., K. He, and J. Sun (Dec. 2015). “BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation”. In: *IEEE International Conference on Computer Vision (ICCV)*. Los Alamitos, CA, USA, pp. 1635–1643 (cit. on pp. 13, 55, 56).

Debelee, T. G., F. Schwenker, A. Ibenthal, and D. Yohannes (Mar. 2020). “Survey of Deep Learning in Breast Cancer Image Analysis”. In: *Evolving Systems* 11.1, pp. 143–163 (cit. on p. 27).

Dolz, J., I. Ben Ayed, and C. Desrosiers (2017). “Unbiased Shape Compactness for Segmentation”. In: *MICCAI*, pp. 755–763 (cit. on pp. 16, 27, 28, 33, 34, 36, 38, 84, 97).

Durand, T., T. Mordan, N. Thome, and M. Cord (July 2017). “WILDCAT: Weakly Supervised Learning of Deep ConvNets for Image Classification, Pointwise Localization and Segmentation”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5957–5966 (cit. on p. 13).

El Jurdi, R., C. Petitjean, P. Honeine, and F. Abdallah (2020a). “BB-UNet: U-Net with Bounding Box Prior”. In: *IEEE Journal of Selected Topics in Signal Processing* 14.6, pp. 1189–1198 (cit. on p. 5).

El Jurdi, R., T. Dargent, C. Petitjean, P. Honeine, and F. Abdallah (Nov. 2020b). “Investigating CoordConv for Fully and Weakly Supervised Medical Image Segmentation”. In: *Tenth International Conference on Image Processing Theory, Tools and Applications, IPTA 2020, Paris, France* (cit. on p. 5).

El Jurdi, R., C. Petitjean, P. Honeine, and F. Abdallah (Aug. 2019). “Organ Segmentation in CT Images with Weak Annotations: A Preliminary Study”. In: *27th GRETSI Symposium on Signal and Image Processing*. Lille, France (cit. on p. 5).

— (2021a). “CoordConv-Unet: Investigating CoordConv for Organ Segmentation”. In: *IRBM* (cit. on p. 5).

EL Jurdi, R., C. Petitjean, P. Honeine, V. Cheplygina, and F. Abdallah (2021). “A Surprisingly Effective Perimeter-based Loss for Medical Image Segmentation”. In: *Medical Imaging with Deep Learning* (cit. on p. 5).

El Jurdi, R., C. Petitjean, P. Honeine, V. Cheplygina, and F. Abdallah (2021b). “High-Level Prior-Based Loss Functions for Medical Image Segmentation: ASurvey”. In: *Submitted to Computer Vision and Image Understanding* (cit. on p. 5).

Foulonneau, A., P. Charbonnier, and F. Heitz (Sept. 2008). “Multi-Reference Shape Priors for Active Contours”. en. In: *International Journal of Computer Vision* 81.1, p. 68 (cit. on p. 12).

Ganaye, P.-A., M. Sdika, B. Triggs, and H. Benoit-Cattin (Dec. 2019). “Removing Segmentation Inconsistencies with Semi-Supervised Non-Adjacency Constraint”. In: *Medical Image Analysis* 58, p. 101551 (cit. on pp. 27, 28, 34, 35, 37, 38, 100).

Gerl, S., J. C. Paetzold, H. He, I. Ezhov, S. Shit, F. Kofler, A. Bayat, G. Tetteh, V. Ntziachristos, and B. Menze (2020). “A Distance-Based Loss for Smooth and Continuous Skin Layer Segmentation in Optoacoustic Images”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Ed. by A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz. Cham: Springer International Publishing, pp. 309–319 (cit. on p. 17).

Ghafoorian, M., N. Karssemeijer, T. Heskes, I. Van Uden, C. Sanchez, G. Litjens, F.-E. Leeuw, B. Ginneken, E. Marchiori, and B. Platel (Oct. 2016). “Location Sensitive Deep Convolutional Neural Networks for Segmentation of White Matter Hyperintensities”. In: *Scientific Reports* 7 (cit. on p. 18).

Grady, L. (2012). “Targeted Image Segmentation Using Graph Methods”. In: *Image Processing and Analysis with Graphs*. CRC Press, pp. 111–135 (cit. on p. 11).

Han, M., G. Yao, W. Zhang, G. Mu, Y. Zhan, X. Zhou, and Y. Gao (2019). “Segmentation of CT Thoracic Organs by Multi-Resolution VB-Nets”. In: *Proceedings of the 2019 Challenge on Segmentation of THoracic Organs at Risk in CT Images, SegTHOR@ISBI 2019, April 8, 2019* (cit. on pp. 49, 50).

Haque, I. and J. Neubert (2020). “Deep Learning Approaches to Biomedical Image Segmentation”. In: *Informatics in Medicine Unlocked* 18, p. 100297 (cit. on pp. 26, 27).

Havaei, M., N. Guizard, H. Larochelle, and P.-M. Jodoin (2016). “Deep Learning Trends for Focal Brain Pathology Segmentation in MRI”. In: *Machine Learning for Health Informatics: State-of-the-Art and Future Challenges*. Ed. by A. Holzinger, pp. 125–148 (cit. on pp. 26, 27, 36).

He, K., G. Gkioxari, P. Dollar, and R. Girshick (Oct. 2017). “Mask R-Cnn”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. Los Alamitos, CA, USA: IEEE Computer Society, pp. 2980–2988 (cit. on p. 47).

Heimann, T. and H.-P. Meinzer (Aug. 2009). “Statistical shape models for 3D medical image segmentation: A review”. In: *Medical Image Analysis* 13.4, pp. 543–563 (cit. on p. 12).

Hesamian, M. H., W. Jia, X. He, and P. Kennedy (Aug. 2019). “POURRI - Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges”. English. In: *Journal of Digital Imaging* 32.4, pp. 582–596 (cit. on pp. 26, 36).

Hsu, C.-C., K.-J. Hsu, C.-C. Tsai, Y.-Y. Lin, and Y.-Y. Chuang (2019). “Weakly Supervised Instance Segmentation Using the Bounding Box Tightness Prior”. en. In: *Advances in Neural Information Processing Systems* 32 (cit. on p. 55).

Hu, H., Y. Zheng, Q. Zhou, J. Xiao, S. Chen, and Q. Guan (2019a). “MC-Unet: Multi-Scale Convolution Unet for Bladder Cancer Cell Segmentation in Phase-Contrast Microscopy Images”. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1197–1199 (cit. on pp. 36, 37).

Hu, X., M. A. Naiel, A. Wong, M. Lamm, and P. Fieguth (June 2019b). "RUNet: A Robust UNet Architecture for Image Super-Resolution". en. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Long Beach, CA, USA: IEEE, pp. 505–507 (cit. on p. 9).

Hu, X., F. Li, D. Samaras, and C. Chen (2019c). "Topology-Preserving Deep Image Segmentation". In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett. Vol. 32. Curran Associates, Inc., pp. 5657–5668 (cit. on pp. 17, 27, 28, 31, 32, 38).

Hu, X., Y. Wang, L. Fuxin, D. Samaras, and C. Chen (Mar. 2021). "Topology-Aware Segmentation Using Discrete Morse Theory". In: *arXiv:2103.09992 [cs]* (cit. on pp. 27, 28, 31, 38).

Isensee, F., P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein (2018). "Automatic Cardiac Disease Assessment on Cine-MRI via Time-Series Segmentation and Domain Specific Features". English. In: *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. Lecture Notes in Computer Science, pp. 120–129 (cit. on p. 10).

Jang, Y., Y. Hong, S. Ha, S. Kim, and H. J. Chang (Jan. 2018). "Automatic Segmentation of LV and RV in Cardiac MRI". English. In: *MICCAI Workshop on Statistical Atlases and Computational Models of the Heart*. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Germany: Springer Verlag, pp. 161–169 (cit. on p. 11).

Jiang, F., A. Grigorev, S. Rho, Z. Tian, Y. Fu, W. J. Sori, A. Khan, and S. Liu (2017). "Medical Image Semantic Segmentation Based on Deep Learning". In: *Neural Computing & Applications* 29, pp. 1257–1265 (cit. on p. 26).

Jiao, L., F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu (2019). "A Survey of Deep Learning-Based Object Detection". In: *IEEE Access* 7, pp. 128837–128868 (cit. on p. 47).

Karimi, D. and S. E. Salcudean (Apr. 2019). "Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks". In: *arXiv:1904.10030 [cs, eess, stat]* (cit. on pp. 16, 74, 75, 86, 88, 91, 93).

Kervadec, H., J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. Ben Ayed (July 2019a). "Boundary Loss for Highly Unbalanced Segmentation". In: *Medical Imaging with Deep Learning*. Vol. 102. Proceedings of Machine Learning Research. London, UK: PMLR, pp. 285–296 (cit. on pp. 12, 14, 16, 66, 67, 74, 75, 79, 80, 86, 88, 91–93).

Kervadec, H., J. Dolz, M. Tang, E. Granger, Y. Boykov, and I. B. Ayed (2018). "Constrained-CNN Losses for Weakly Supervised Segmentation". In: *1st Conference on Medical Imaging with Deep Learning (MIDL), Amsterdam, the Netherlands* (cit. on pp. 14–17, 27–30, 36, 38, 62, 65, 66, 76, 79, 92–94, 100).

Kervadec, H., J. Dolz, S. Wang, E. Granger, and I. ben Ayed (2020a). "Bounding Boxes for Weakly Supervised Segmentation: Global Constraints Get Close to Full Supervision". In: *Medical Imaging with Deep Learning* (cit. on pp. 14, 17, 27, 28, 30, 35, 55, 92, 100).

Kervadec, H., J. Dolz, J. Yuan, C. Desrosiers, E. Granger, and I. B. Ayed (Sept. 2019b). "Constrained Deep Networks: Lagrangian Optimization via Log-Barrier Extensions". In: *arXiv:1904.04205 [cs]* (cit. on pp. 30, 38).

— (Apr. 2020b). "Constrained Deep Networks: Lagrangian Optimization via Log-Barrier Extensions". In: *arXiv:1904.04205 [cs]* (cit. on pp. 17, 27).

Khoreva, A., R. Benenson, J. Hosang, M. Hein, and B. Schiele (July 2017). "Simple Does It: Weakly Supervised Instance and Semantic Segmentation". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1665–1674 (cit. on pp. 13, 14, 18, 55–58).

Khoreva, A., R. Benenson, J. H. Hosang, M. Hein, and B. Schiele (2016). "Weakly Supervised Semantic Labelling and Instance Segmentation". In: *CoRR* abs/1603.07485 (cit. on p. 14).

Kim, B. and J. C. Ye (2020). "Mumford–Shah Loss Functional for Image Segmentation With Deep Learning". In: *IEEE Transactions on Image Processing* 29, pp. 1856–1866 (cit. on pp. 12, 17).

Krithiga, R. and P. Geetha (Aug. 2020). "Breast Cancer Detection, Segmentation and Classification on Histopathology Images Analysis: A Systematic Review". In: *Archives of Computational Methods in Engineering* (cit. on p. 27).

Kuijf, H. J., J. M. Biesbroek, J. De Bresser, R. Heinen, S. Andermatt, M. Bento, M. Berseth, M. Belyaev, M. J. Cardoso, A. Casamitjana, D. L. Collins, M. Dadar, A. Georgiou, M. Ghafoorian, D. Jin, A. Khademi, J. Knight, H. Li, X. Lladó, M. Luna, Q. Mahmood, R. McKinley, A. Mehrtash, S. Ourselin, B.-Y. Park, H. Park, S. H. Park, S. Pezold, E. Puybareau, L. Rittner, C. H. Sudre, S. Valverde, V. Vilaplana, R. Wiest, Y. Xu, Z. Xu, G. Zeng, J. Zhang, G. Zheng, C. Chen, W. van der Flier, F. Barkhof, M. A. Viergever, and G. J. Biessels (2019). "Standardized Assessment of Automatic Segmentation of White Matter Hyperintensities and Results of the WMH Segmentation Challenge". In: *IEEE Transactions on Medical Imaging* 38.11, pp. 2556–2568 (cit. on pp. 22, 77).

Lambert, Z., C. Petitjean, B. Dubray, and S. Ruan (2020). "SegTHOR: Segmentation of Thoracic Organs at Risk in CT Images". In: *2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)* abs/1912.05950, pp. 1–6 (cit. on pp. 17, 19, 43, 49, 50).

Lei, T., R. Wang, Y. Wan, B. Zhang, H. Meng, and A. K. Nandi (2020). "Medical Image Segmentation Using Deep Learning: A Survey". In: (cit. on p. 26).

Levine, S., C. Finn, T. Darrell, and P. Abbeel (Jan. 2016). "End-to-End Training of Deep Visuomotor Policies". In: *Journal of Machine Learning Research* 17.1, pp. 1334–1373 (cit. on p. 62).

Lillo, W. E., M. H. Loh, S. Hui, and S. H. Zak (1993). "On Solving Constrained Optimization Problems with Neural Networks: A Penalty Method Approach". In: *IEEE Transactions on Neural Networks* 4.6, pp. 931–940 (cit. on p. 15).

Lin, T.-Y., P. Goyal, R. B. Girshick, K. He, and P. Dollár (2017). "Focal Loss for Dense Object Detection". In: *ICCV*, pp. 2999–3007 (cit. on pp. 11, 27).

Litjens, G., T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez (Dec. 2017). "A Survey on Deep Learning in Medical Image Analysis". English. In: *Medical Image Analysis* 42, pp. 60–88 (cit. on p. 36).

Liu, R., J. Lehman, P. Molino, F. Petroski Such, E. Frank, A. Sergeev, and J. Yosinski (2018). "An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution". In: *Neural Information Processing Systems*, pp. 9605–9616 (cit. on pp. 62–66).

Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg (2016). "SSD: Single Shot MultiBox Detector". In: *Computer Vision – ECCV 2016*. Ed. by B. Leibe, J.

Matas, N. Sebe, and M. Welling. Cham: Springer International Publishing, pp. 21–37 (cit. on p. 47).

Liu, X., O. Veksler, and J. Samarabandu (June 2008). "Graph Cut with Ordering Constraints on Labels and Its Applications". In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (cit. on p. 12).

Long, J., E. Shelhamer, and T. Darrell (June 2015). "Fully Convolutional Networks for Semantic Segmentation". In: *CVPR*, pp. 3431–3440 (cit. on p. 8).

Loog, M. and A. C. Jensen (May 2015). "Semi-Supervised Nearest Mean Classification Through a Constrained Log-Likelihood". In: *IEEE Transactions on Neural Networks and Learning Systems* 26.5, pp. 995–1006 (cit. on p. 13).

Lorenzo-Valdés, M., G. I. Sanchez-Ortiz, R. Mohiaddin, and D. Rueckert (2002). "Atlas-Based Segmentation and Tracking of 3D Cardiac MR Images Using Non-Rigid Registration". en. In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002*. Ed. by T. Dohi and R. Kikinis. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 642–650 (cit. on p. 12).

Lu, W., Y. Zhou, G. Wan, S. Hou, and S. Song (June 2019). "L3-Net: Towards Learning Based LiDAR Localization for Autonomous Driving". In: *CVPR*, pp. 6382–6391 (cit. on p. 47).

Ma, J., J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel (2021). "Loss odyssey in medical image segmentation". In: *Medical Image Analysis* 71, p. 102035 (cit. on p. 74).

Ma, J., Z. Wei, Y. Zhang, Y. Wang, R. Lv, C. Zhu, G. Chen, J. Liu, C. Peng, L. Wang, Y. Wang, and J. Chen (2020). "How Distance Transform Maps Boost Segmentation {CNN}s: An Empirical Study". In: *Medical Imaging with Deep Learning* (cit. on pp. 74, 91).

Magadza, T. and S. Viriri (2021). "Deep Learning for Brain Tumor Segmentation: A Survey of State-of-the-Art". In: *Journal of Imaging* 7.2 (cit. on p. 27).

Maier, O., B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, D. Christiaens, F. Dutil, K. Egger, C. Feng, B. Glocker, M. Götz, T. Haeck, H.-L. Halme, M. Havaei, K. M. Iftekharuddin, P.-M. Jodoin, and e. al. (2017). "ISLES 2015 - A Public Evaluation Benchmark for Ischemic Stroke Lesion Segmentation from Multispectral MRI". In: *Medical Image Analysis* 35, pp. 250–269 (cit. on pp. 22, 77).

Márquez-Neila, P., M. Salzmann, and P. Fua (June 2017). "Imposing Hard Constraints on Deep Networks: Promises and Limitations". In: *arXiv:1706.02025 [cs]* (cit. on p. 15).

Meyer, P., V. Noblet, C. Mazzara, and A. Lallement (May 2018). "Survey on Deep Learning for Radiotherapy". In: *Computers in Biology and Medicine* 98 (cit. on p. 27).

Milletari, F., N. Navab, and S.-A. Ahmadi (2016). "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation". In: *CoRR* abs/1606.04797 (cit. on pp. 9–11, 27, 49, 67, 86, 90).

Milnor, J. (Mar. 2016). *Morse Theory. (AM-51), Volume 51*. en. Princeton University Press (cit. on p. 31).

Mirikharaji, Z. and G. Hamarneh (2018). "Star Shape Prior in Fully Convolutional Networks for Skin Lesion Segmentation". In: *MICCAI*. Vol. 11073, pp. 737–745 (cit. on pp. 12, 14, 16, 17, 27, 28, 33, 38, 84).

Mosinska, A., P. Márquez-Neila, M. Kozinski, and P. Fua (2018). “Beyond the Pixel-Wise Loss for Topology-Aware Delineation”. In: *CVPR*, pp. 3136–3145 (cit. on pp. 14, 16, 17, 84).

Mumford, D. and J. Shah (July 1989). “Optimal approximations by piecewise smooth functions and associated variational problems”. In: *Communications on Pure and Applied Mathematics* 42.5, pp. 577–685 (cit. on p. 12).

Nosrati, M. S. and G. Hamarneh (2016). “Incorporating Prior Knowledge in Medical Image Segmentation: A Survey”. In: *CoRR* abs/1607.01092 (cit. on pp. 12, 15).

Oda, H., H. R. Roth, K. Chiba, J. Sokolić, T. Kitasaka, M. Oda, A. Hinoki, H. Uchida, J. A. Schnabel, and K. Mori (2018). “BESNet: Boundary-Enhanced Segmentation of Cells in Histopathological Images”. In: *MICCAI*, pp. 228–236 (cit. on pp. 18, 19).

Oktay, O., E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvao, T. Dawes, D. P. O‘Regan, B. Kainz, B. Glocker, and D. Rueckert (Feb. 2018a). “Anatomically Constrained Neural Networks (ACNNs): Application to Cardiac Image Enhancement and Segmentation”. In: *IEEE Transactions on Medical Imaging* 37.2, pp. 384–395 (cit. on p. 18).

Oktay, O., J. Schlemper, L. L. Folgoc, M. C. H. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert (2018b). “Attention U-Net: Learning Where to Look for the Pancreas”. In: *Medical Imaging with Deep Learning* (cit. on pp. 9, 18, 47, 48).

Pan, Z., W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng (2019). “Recent Progress on Generative Adversarial Networks (GANs): A Survey”. In: *IEEE Access* 7, pp. 36322–36333 (cit. on p. 100).

Papandreou, G., L.-C. Chen, K. P. Murphy, and A. L. Yuille (2015). “Weakly-and Semi-Supervised Learning of a Deep Convolutional Network for Semantic Image Segmentation”. In: *IEEE ICCV*, pp. 1742–1750 (cit. on pp. 13, 58, 100).

Pathak, D., P. Krähenbühl, and T. Darrell (Dec. 2015). “Constrained Convolutional Neural Networks for Weakly Supervised Segmentation”. In: *ICCV*, pp. 1796–1804 (cit. on pp. 13, 15, 29, 30).

Peng, J., H. Kervadec, J. Dolz, I. B. Ayed, M. Pedersoli, and C. Desrosiers (2020). “Discretely-Constrained Deep Network for Weakly Supervised Segmentation”. In: *Neural networks : the official journal of the International Neural Network Society* 130, pp. 297–308 (cit. on pp. 15, 17, 27, 28, 30, 38).

Perone, C. S., cclauss, E. Saravia, P. L. Ballester, and MohitTare (Nov. 2018). “Perone/Medicaltorch: Release v0.2”. In: (cit. on pp. 49, 58, 65).

Pont-Tuset, J., P. Arbelaez, J. T.Barron, F. Marques, and J. Malik (Jan. 2017). “Multiscale Combinatorial Grouping for Image Segmentation and Object Proposal Generation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.1, pp. 128–140 (cit. on p. 57).

Qi, H., S. Collins, and J. A. Noble (Sept. 2019). “UPI-Net: Semantic Contour Detection in Placental Ultrasound”. In: *Visual Recognition for Medical Images (VRMI), ICCV 2019 Workshop* (cit. on p. 62).

Rahman, M. A. and Y. Wang (2016). “Optimizing Intersection-over-Union in Deep Neural Networks for Image Segmentation”. In: *ISVC* (cit. on p. 27).

Ravishankar, H., R. Venkataramani, S. Thiruvenkadam, P. Sudhakar, and V. Vaidya (2017). "Learning and Incorporating Shape Models for Semantic Segmentation". In: *MICCAI*, pp. 203–211 (cit. on pp. 8, 18, 19, 47).

Razzak, M., S. Naz, and A. Zaib (Jan. 2018). "Deep Learning for Medical Image Processing: Overview, Challenges and the Future". In: *Lecture Notes in Computational Vision and Biomechanics*, pp. 323–350 (cit. on p. 26).

Reddy, C., K. Gopinath, and H. Lombaert (July 2019). "Brain Tumor Segmentation Using Topological Loss in Convolutional Networks". In: *MIDL*. London, UK (cit. on pp. 27, 28, 34, 35, 37, 38).

Renard, F., S. Guedria, N. De Palma, and N. Vuillerme (2020). "Variability and Reproducibility in Deep Learning for Medical Image Segmentation". In: *Scientific Reports* 10.1, pp. 1–16 (cit. on p. 26).

Rivest, J.-F., P. Soille, and S. Beucher (Jan. 1993). "Morphological gradients". In: *J. Electronic Imaging* 2, pp. 326–336 (cit. on p. 88).

Rohlfing, T., R. Brandt, R. Menzel, D. B. Russakoff, and C. R. Maurer (2005). "Quo Vadis, Atlas-Based Segmentation?" en. In: *Handbook of Biomedical Image Analysis: Volume III: Registration Models*. Ed. by J. S. Suri, D. L. Wilson, and S. Laxminarayan. Topics in Biomedical Engineering International Book Series. Boston, MA: Springer US, pp. 435–486 (cit. on p. 12).

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *MICCAI*, pp. 234–241 (cit. on pp. 8, 10, 27, 49, 50, 86).

Rother, C., V. Kolmogorov, and A. Blake (2004). ""GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts". In: *ACM SIGGRAPH 2004 Papers*. SIGGRAPH '04. New York, NY, USA: ACM, pp. 309–314 (cit. on pp. 55, 57).

Rousson, M. and N. Paragios (2002). "Shape Priors for Level Set Representations". In: *In ECCV*. Springer, pp. 78–92 (cit. on p. 12).

Ségonne, F. and B. Fischl (2015). "Integration of Topological Constraints in Medical Image Segmentation". In: *Handbook of Biomedical Imaging: Methodologies and Clinical Research*. Ed. by N. Paragios, J. Duncan, and N. Ayache. Springer US, pp. 245–262 (cit. on p. 30).

Shamsolmoali, P., M. Zareapoor, E. Granger, H. Zhou, R. Wang, M. E. Celebi, and J. Yang (2021). "Image synthesis with adversarial networks: A comprehensive survey and case studies". In: *Information Fusion* 72, pp. 126–146 (cit. on p. 101).

Shi, J. and J. Malik (2000). "Normalized Cuts and Image Segmentation". In: *IEEE Transactions on pattern analysis and machine intelligence* 22.8, pp. 888–905 (cit. on p. 12).

Shin, H.-C., N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski (2018). "Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks". In: *Simulation and Synthesis in Medical Imaging*. Ed. by A. Gooya, O. Goksel, I. Oguz, and N. Burgos. Cham: Springer International Publishing, pp. 1–11 (cit. on p. 100).

Shit, S., J. C. Paetzold, A. Sekuboyina, A. Zhylka, I. Ezhov, A. Unger, J. P. W. Pluim, G. Tetteh, and B. H. Menze (2019). "clDice - a Topology-Preserving Loss Function for Tubular Structure Segmentation". In: *Medical Imaging Meets NeurIPS 2019 Workshop* (cit. on pp. 27, 28, 31, 32, 38, 62, 65, 66, 76, 87, 89, 93, 94).

Shit, S., J. C. Paetzold, A. Sekuboyina, A. Zhylka, I. Ezhov, A. Unger, J. P. W. Pluim, G. Tetteh, and B. H. Menze (2020). "clDice - a Topology-Preserving Loss Function for Tubular Structure Segmentation". In: *ArXiv* abs/2003.07311 (cit. on pp. 17, 76).

Simonyan, K. and A. Zisserman (2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *International Conference on Learning Representations* (cit. on p. 8).

Simpson, A. L., M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. H. Menze, O. Ronneberger, R. M. Summers, P. Bilic, P. F. Christ, R. K. G. Do, M. Gollub, J. Golia-Pernicka, S. Heckers, W. R. Jarnagin, M. McHugo, S. Napel, E. Vorontsov, L. Maier-Hein, and M. J. Cardoso (2019). "A large annotated medical image dataset for the development and evaluation of segmentation algorithms". In: *CoRR* abs/1902.09063 (cit. on pp. 44, 77, 78).

Slabaugh, G. and G. Unal (Sept. 2005). "Graph Cuts Segmentation Using an Elliptical Shape Prior". In: *IEEE International Conference on Image Processing 2005*. Vol. 2, pp. II–1222 (cit. on p. 12).

Sudre, C. H., W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso (2017). "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations". In: *DLMIA/ML-CDS@MICCAI* (cit. on p. 11).

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich (2015). "Going Deeper with Convolutions". In: *CVPR*, pp. 1–9 (cit. on p. 8).

Taghanaki, S. A., K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh (Oct. 2019). "Deep Semantic Segmentation of Natural and Medical Images: A Review". In: *arXiv:1910.07655 [cs, eess]* (cit. on pp. 26, 27).

Taha, A. A. and A. Hanbury (Aug. 2015). "Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool". In: *BMC Medical Imaging* 15, p. 29 (cit. on p. 102).

Tofighi, M., T. Guo, J. K. P. Vanamala, and V. Monga (Oct. 2018). "Deep Networks with Shape Priors for Nucleus Detection". In: *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 719–723 (cit. on p. 19).

Trullo, R., C. Petitjean, S. Ruan, B. Dubray, D. Nie, and D. Shen (2017). "Joint Segmentation of Multiple Thoracic Organs in CT Images with Two Collaborative Deep Architectures". In: *MICCAI'17 workshop Deep Learning in Medical Image Analysis* (cit. on p. 18).

van Sonsbeek, T. and V. Cheplygina (2020). "Predicting Scores of Medical Imaging Segmentation Methods with Meta-learning". English. In: *Interpretable and Annotation-Efficient Learning for Medical Image Computing - 3rd International Workshop, iMIMIC 2020, 2nd International Workshop, MIL3iD 2020, and 5th International Workshop, LABELS 2020, Held in Conjunction with MICCAI 2020, Proceedings*. Ed. by J. Cardoso, W. Silva, R. Cruz, H. Van Nguyen, B. Roysam, N. Heller, P. Henriques Abreu, J. Pereira Amorim, I. Isgum, V. Patel, K. Zhou, S. Jiang, N. Le, K. Luu, R. Sznitman, V. Cheplygina, S. Abbasi, D. Mateus, and E. Trucco. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). LABELS 2020 ; Conference date: 04-10-2020 Through 08-10-2020. Germany: Springer, pp. 242–253 (cit. on p. 101).

Veksler, O. (2008a). “Star Shape Prior for Graph-Cut Image Segmentation”. en. In: *Computer Vision – ECCV 2008*. Ed. by D. Forsyth, P. Torr, and A. Zisserman. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 454–467 (cit. on p. 12).

— (2008b). “Star Shape Prior for Graph-Cut Image Segmentation”. In: *ECCV*, pp. 454–467 (cit. on p. 33).

Vicente, S., V. Kolmogorov, and C. Rother (June 2008). “Graph Cut Based Image Segmentation with Connectivity Priors”. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (cit. on p. 12).

Wang, P., P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell (2018). “Understanding Convolution for Semantic Segmentation”. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1451–1460 (cit. on p. 33).

Wang, S., K. Liang, C. Pan, C. Ye, X. Li, F. Liu, Y. Yu, and Y. Wang (Apr. 2020). “Segmentation-Based Method Combined with Dynamic Programming for Brain Midline Delineation”. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 772–776 (cit. on pp. 62, 66).

Wang, W., J. Chen, J. Zhao, Y. Chi, X. Xie, L. Zhang, and X.-S. Hua (2019). “Automated Segmentation of Pulmonary Lobes Using Coordination-Guided Deep Neural Networks”. In: *CoRR* abs/1904.09106 (cit. on pp. 62, 65, 66).

Wei, Y., X. Liang, Y. Chen, X. Shen, M. Cheng, J. Feng, Y. Zhao, and S. Yan (Nov. 2017). “STC: A Simple to Complex Framework for Weakly-Supervised Semantic Segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.11, pp. 2314–2320 (cit. on p. 56).

Weng, Y., T. Zhou, Y. Li, and X. Qiu (2019). “NAS-Unet: Neural Architecture Search for Medical Image Segmentation”. In: *IEEE Access* 7, pp. 44247–44257 (cit. on p. 9).

Xu, C., D. L. Pham, and J. L. Prince (2000). “Image Segmentation Using Deformable Models”. In: *Handbook of Medical Imaging*. Vol. 2. SPIE, pp. 129–174 (cit. on p. 12).

Yang, D., H. Roth, X. Wang, Z. Xu, A. Myronenko, and D. Xu (2020). “Enhancing Foreground Boundaries for Medical Image Segmentation”. In: *Medical Imaging with Deep Learning* (cit. on pp. 16, 86).

Yang, S., J. Kweon, and Y.-H. Kim (July 2019). “Major Vessel Segmentation on X-Ray Coronary Angiography Using Deep Networks with a Novel Penalty Loss Function”. In: *International Conference on Medical Imaging with Deep Learning – Extended Abstract Track*. London, UK (cit. on p. 92).

Yang, X., C. Bian, L. Yu, D. Ni, and P.-A. Heng (Jan. 2018). “Class-Balanced Deep Neural Network for Automatic Ventricular Structure Segmentation”. In: *STACOM@MICCAI*, pp. 152–160 (cit. on pp. 11, 84).

Yao, X., H. Yang, Y. Wu, P. Wu, B. Wang, X. Zhou, and S. Wang (Jan. 2019). “Land Use Classification of the Deep Convolutional Neural Network Method Reducing the Loss of Spatial Features”. English. In: *Sensors* 19.12, p. 2792 (cit. on p. 62).

Yue, Q., X. Luo, Q. Ye, L. Xu, and X. Zhuang (2019). “Cardiac Segmentation from LGE MRI Using Deep Neural Network Incorporating Shape and Spatial Priors”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Ed. by D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan. Cham: Springer International Publishing, pp. 559–567 (cit. on p. 27).

Zhang, J., C. Petitjean, and S. Ainouz (2020). “Kappa Loss for Skin Lesion Segmentation in Fully Convolutional Network”. In: *IEEE ISBI, Iowa City USA*, pp. 2001–2004 (cit. on p. 11).

Zhang, Z., Q. Liu, and Y. Wang (May 2018). “Road Extraction by Deep Residual U-Net”. In: *IEEE Geoscience and Remote Sensing Letters* 15.5, pp. 749–753 (cit. on pp. 9, 67).

Zhou, T., S. Ruan, and S. Canu (2019). “A Review: Deep Learning for Medical Image Segmentation Using Multi-Modality Fusion”. In: *Array* 3-4, p. 100004 (cit. on p. 26).

Zotti, C., Z. Luo, A. Lalande, and P. Jodoin (May 2019). “Convolutional Neural Network with Shape Prior Applied to Cardiac MRI Segmentation”. In: *IEEE Journal of Biomedical and Health Informatics* 23.3, pp. 1119–1128 (cit. on p. 18).

Zotti, C., Z. Luo, O. Humbert, A. Lalande, and P.-M. Jodoin (2017). “GridNet with Automatic Shape Prior Registration for Automatic MRI Cardiac Segmentation”. In: *Statistical Atlases and Computational Models of the Heart STACOM, Held in Conjunction with MICCAI, Quebec City, Canada*. Vol. 10663. LNCS, pp. 73–81 (cit. on pp. 13, 18, 19).