TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABREVIATIONS

CE	Cross Entropy
CNN	Convolutional Neural Network
DNN	Deep Neural Network
ETS	École de Technologie Supérieure
GPU	Graphical Processing Unit
MLP	Multi Layer Perceptron
MRI	Magnetic Resonance Imaging
ReLU	Rectified Linerar Unit
SGD	Stochastic Gradient Descent
WCE	Weighted Cross Entropy

INTRODUCTION

Problem statement

Driven by the revolution of big data and computing power, deep learning techniques have shown their ability to deliver outstanding performances in different fields where machines become more suitable than humans dealing with a huge amount of data in a comprehensive way. The medical field is one of the fields where deep learning had the largest impact.

Last years have witnessed a revolution in medical imaging that has broken paradigms and seems to open new possibilities in the field. Recent studies estimate that medical images currently account for at least 90 % of medical data, which is an overwhelming amount on human scale. The detection, classification and segmentation of specific medical patterns, e.g organs, tissues, tumors are the main key steps for diagnosis, treatment or monitoring of different diseases progression. For instance, an accurate delineation of one or multiple structures is essential to derive important clinical parameters such as ventricular volume, ejection fraction, and myocardial mass for cardiac images or lesion volume for brain images. However, the manual segmentation of these structures is complex, time consuming and requires multiple expert opinions. In some cases, the annotation may not be possible due to the absence of experts or due to the rarity of the disease or lesion to be segmented. To address this problem, deep neural networks have shown their ability to deliver outstanding performances in such tasks.

Recently, a substantial growth in the number of deep learning techniques has been witnessed. However, while achieving a lot of successes in many complex medical problems, researchers have realized that there are still some challenges to deal with. One of them is the data unbalance problem, which occurs for instance when the size of target foreground region is several orders of magnitude less than the background region size. Such problem occurs mostly in cases of abnormalities detection/segmentation, e.g tumors (M.Havaei *et al.* (2015); Pereira *et al.* (2016);

Menze *et al.* (2014); Hwang & Kim (2016); Bellver *et al.* (2017)), the breast (Kallenberg *et al.* (2016); Wang *et al.* (2016); Hassanien *et al.* (2014)), the lungs (Kalinovsky & Kovalev (2016); Mi *et al.* (2014); Skourt *et al.* (2018)), etc. For example, in the characterization of white matter hyperintensities (WMH) of presumed vascular origin, the foreground composed of target regions may be 500 times smaller than the background (see the typical example in Fig 0.1). In such case, common approaches, assuming that all the patterns are evenly distributed, fail to deliver good segmentation performance. The issue is that they tend to favour the majority class and, then, misclassify the target class.

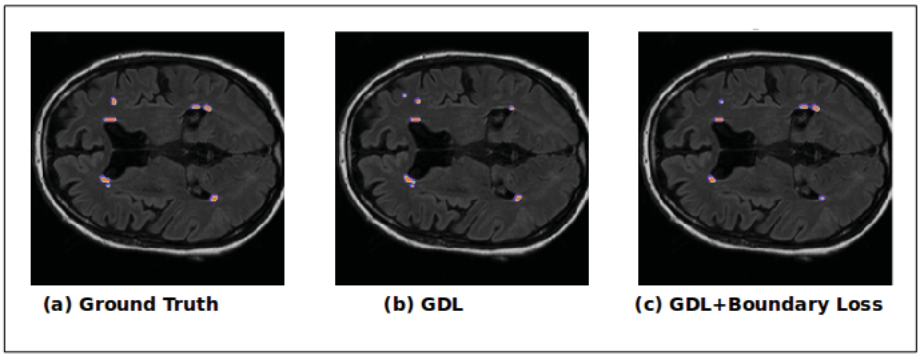


Figure 0.1 A visual comparison that shows the positive effect of our boundary loss on a validation data from the WMH dataset. Our boundary loss helped recovering small regions that were otherwise missed by the generalized Dice loss (GDL). Best viewed in colors.

Research hypotheses and objectives

All common proposed strategies to re-balance class prior distributions are *region-based.* In this work, our objective is to propose a boundary loss that takes the form of a distance metric on the space of contours (or shapes), not regions. We argue that a boundary loss can mitigate the issues related to regional losses in highly unbalanced segmentation problems. Rather than using unbalanced integrals over the regions, a boundary loss uses integrals over the boundary (interface)

between the regions. Furthermore, a boundary loss provides information that is complimentary to regional losses. It is, however, challenging to represent the boundary points corresponding to the regional softmax outputs of a CNN. This difficulty may explain why boundary losses have been mostly avoided in the context of deep segmentation networks. Our boundary loss is inspired by techniques in discrete (graph-based) optimization for computing gradient flows of curve evolution (Boykov et al., 2006). Following an integral approach for computing boundary variations, we express a non-symmetric L2 distance on the space of shapes as a regional integral, which avoids completely local differential computations involving contour points. Our boundary loss is the sum of linear functions of the regional softmax probability outputs of the network. Therefore, it can be easily combined with standard regional losses and implemented with any existing deep network architecture for N-D segmentation.

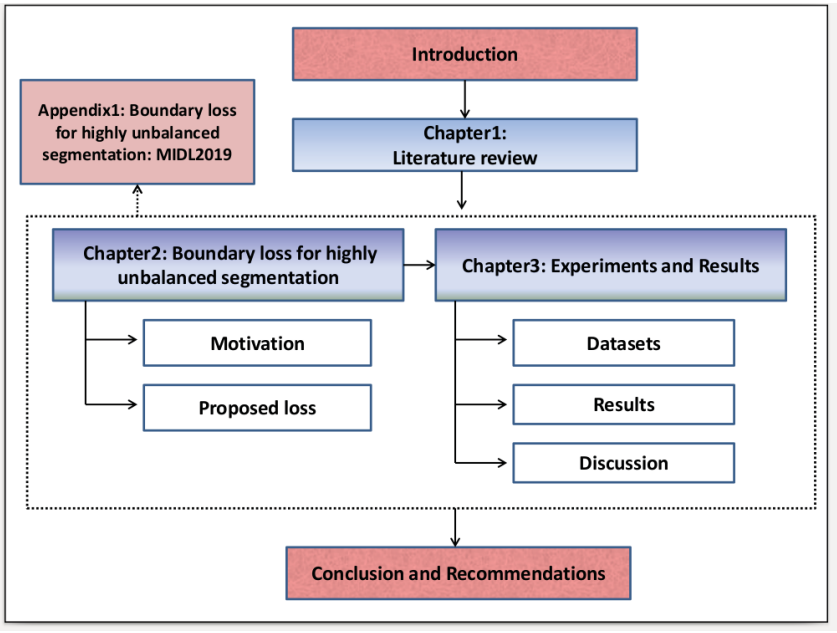


Figure 0.2 Thesis outline

Thesis outline

This document is organized as shown in Fig.0.2 . Chapter one introduces the literature review where we report some basics of CNNs and discuss their major impact on medical image analysis, showing both successes and current challenges. In the second chapter, we propose a new boundary loss for highly unbalanced problems, and explain in details the methodology that we followed. The next chapter is dedicated to the experimental protocol, where we show our results on different datasets. Finally, we conclude with a summary of the work and a discussion of some possible future directions.

CHAPTER 1

LITERATURE REVIEW

In this chapter , we introduce the literature review. First, we discuss some traditional techniques used for medical image segmentation. Then, we report some basics of convolutional neural networks. In the following section, we give an overview of the fast-growing place of deep learning techniques in medical image analysis, showing various successes and also discussing two main challenges faced by current research: lack of annotations and unbalanced problems.

1.1 Traditional approaches for medical image segmentation

Image segmentation is the process of delineating one or multiple semantic structures in the image. It is one of the most active research tasks in medical image processing where it facilitates the extraction of useful informations for diagnostic medicine and biomedical imaging applications such as pathology localization, analysis of the anatomical structure and computer-assisted surgery. Several traditional approaches have been introduced in the literature and can be divided into four main categories: thresholding-based methods, atlas-guided methods, statistical approaches and deformable models.

Based on the assumption that target objects can be identified by their intensity values, thresholding, as a region-based technique, has been largely used to segment medical images. The idea is to apply a threshold value on pixels intensities to extract the target. Typically, this value corresponds to a peak in the histogram that divides an image into two different parts: pixels with intensities greater than a given threshold T correspond to a first class while remaining pixels belong to a second class.

The main advantage of thresholding is its simplicity and rapidity. However, global thresholding assumes that the image has a bimodal histogram, which is not the case in many medical images, especially when the background is not homogeneous or the object presents a lot of intensity variation.

To address that, local thresholding has been introduced (N & S (2016); Aja-Fernández *et al.* (2015)). Authors in (Wu *et al.* (1995); Manikandan *et al.* (2014); Xie & Bovik (2013)) proposed the so-called *Otsu's* method for an automated selection of the threshold. One of the common weaknesses of thresholding approach is its strong dependence to the database. Also, thresholding does not take into account the spatial information, which leads to poor segmentation results especially with noisy images. Another popular approach to segment medical images is the use of atlases. The idea is to build a reference model from some training data and then register the unseen image to the reference model. Registration is the process of aligning the image so that corresponding features can be easily related and distances can be computed. Although they often deliver good performances, atlas-based methods suffer from complex registration steps. Furthermore, it is not easy to cover all the anatomical variability of target objects e.g., rare diseases or anatomical abnormalities. Thus, atlas-based methods can fail to deliver good performances when dealing with huge deformations.

Over the last decades, the explosion of medical datasets called for new solutions. In that context, statistical approaches propose a parametric model, which describes the variability of the target object (shape, texture, size, etc). Training data is used during the learning process. Despite their success in various medical applications, such techniques can lead to an over fitting problem when the size of data is limited comparing to the number of parameters. Unlike statistical techniques, deformable models do not need a training data. They are defined as dynamic contours that evolve to fit a target shape via an energy-minimization strategy. Deformable models can be divided into two main categories; parametric and geometric models. In the first category, we find the well known active contours widely used for medical image segmentation (McInerney & Terzopoulos (1996); Yushkevich *et al.* (2006); Wang *et al.* (2009); He *et al.* (2008). In a different strategy, geometric models use a distance transformation to describe the target shape. In other words, they rely on the so-called *level-set* based shape representation (Zhang *et al.* (2008); Droske *et al.* (2001); Lin *et al.* (2003); Li *et al.* (2011); Mitiche & Ben Ayed (2011)). The most straightforward advantage of deformable models is their flexibility and their usefulness in specific tasks, e.g., object tracking. However, they are

still sensitive not only to the initialization but also to the presence of noise, when the initial boundary is far from the target, they might get stuck in a local minima.

1.2 Convolutional neural networks (CNNs): basic concepts

In image analysis, one common idea between traditional approaches is that they all need to extract features in a separated way before applying another method for classification. A logical next step is to let the machine learn these features in an automatic fashion. This is the key idea behind most deep learning techniques. In simple words, deep neural networks (*DNNs*) are models composed of many successive layers in order to transform an input data to an output while learning increasingly higher level features. Convolutional neural networks (*CNNs*) are among the most successful models for image analysis, in general, and for image segmentation particularly.

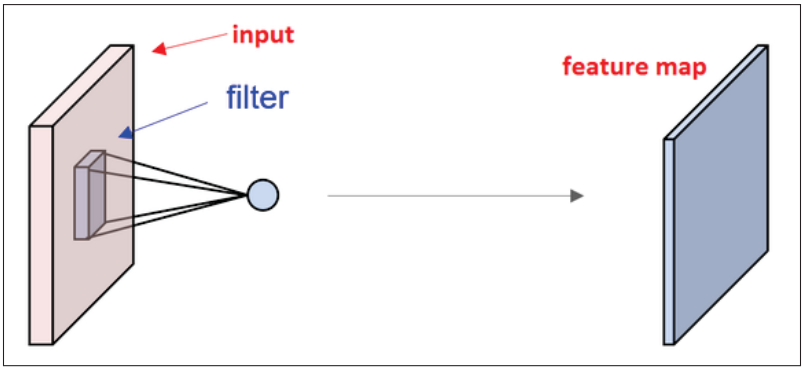


Figure 1.1 Illustration of a feature map
(Taken from: http://https://www.quora.com/
What-is-meant-by-feature-maps-in-convolutional-neural-networks)

A *CNN* is a multilayer perceptron (*mlp*) designed to use a fewer amount of parameters than a fully connected network by taking advantage of the spatial relationships between neighboring pixels in an image. Typically, the network is a set of convolutional and fully connected layers. It can also include the so-called *pooling layers*.

Each convolutional layer is composed of small neuron collections (called *receptive fields*) which process portions of the input image and produce a set of outputs called *feature maps* (see Fig. 1.1). A feature map is obtained by a local convolution operation using a linear filter K, adding a bias term and then applying an activation function.

After introducing discrete convolution, we discuss the basic components of a CNN.

1.2.1 Discrete convolution

Discrete convolution is a powerful mathematical tool used in many areas of math and engineering, including computer vision. Applied to image analysis, the idea is to use a kernel (or filter) K and slide it over the image I. For each position of the kernel, we multiply the overlapping values of the kernel and image together, and add up the results. This sum of products will be the value of the output image (see Fig. 1.2). The discrete convolution of an image I with a

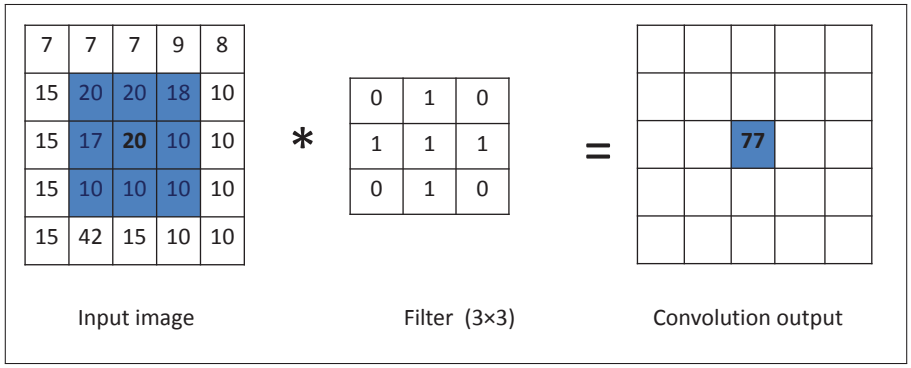


Figure 1.2 Discrete convolution

kernel K of size $(2h_1+1)\times(2h_2+1)$ can be given by:

$$(I*K)_{p,q} := \sum_{u=-h_1}^{h_1}\sum_{v=-h_2}^{h_2} K_{u,v} I_{p+u,q+v} \tag{1.1}$$

Where the filter K is given by:

$$K = \begin{pmatrix} K_{-h1,-h2} & \dots & K_{-h1,h2} \\ \dots & K_{0,0} & \dots \\ K_{h1,-h2} & \dots & K_{h1,h2} \end{pmatrix} \tag{1.2}$$

1.2.2 Distinguishing features of CNNs

The first question that can be asked about CNNs is the following: **What is the difference between a CNN and an MLP?**

A CNN is an MLP with fewer parameters and a special structure that aims to reduce computational complexity and ensure a spatial invariance, i.e., the network perform the input regardless its spatial location in the image. In other words, a car is a car regardless its position in the image. Convolutional neural networks have two main distinguishing features.

Local connectivity: In the case of high-dimensional inputs such as images, a fully connected hidden layer would have unmanageable parameters. Also, computing activation functions in respective neurons would be very expensive. In order to avoid these problems, CNNs take advantage of the local spatial coherence property of images and enforce a local connectivity pattern between neurons of adjacent layers. In other words, neurons in one layer are only connected to neurons in the next layer that are spatially close to them (see Fig. 1.3). Basically, we assume that two distant pixels have less significant relationship that close pixels.

Parameter sharing: The idea behind weights sharing is the convolutional concept of layers. The idea behind this property is to control and reduce the number of parameters by making an assumption: certain neurons are aggregated together and share the parameters matrix (weighted connections) forming, then, a feature map. This concept allows features to be detected regardless of their position in the image. Parameter sharing reduces the number of free parameters, achieving better generalization and computation performance.

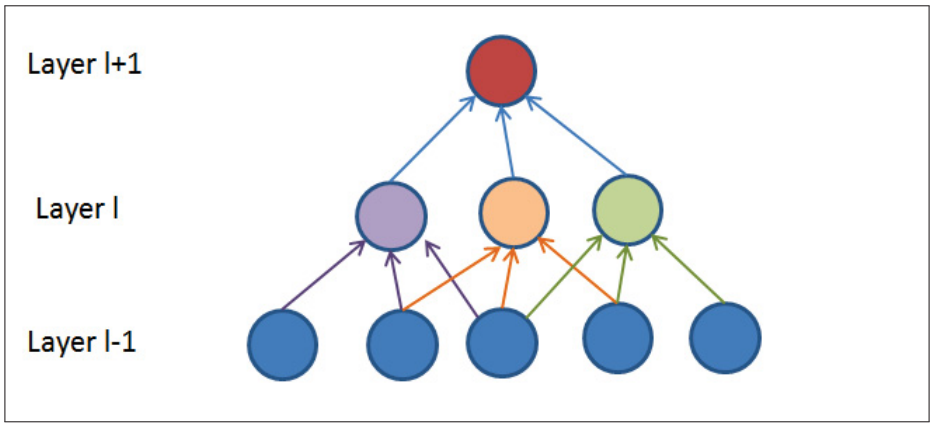


Figure 1.3 Local connectivity in CNN

1.2.3 Layers

Convolutional layer

The convolutional layer is the core building block of a CNN. It is designed to learn hierarchical features from the image. In general, the input is a feature map. For the first layer, it would be the initial image. Let layer l be a convolutional layer. Then, the input layer contains $m_1^{(l-1)}$ feature maps, each of size $m_2^{(l-1)} \times m_3^{(l-1)}$. The output of layer l consists of $m_1^{(l)}$ feature maps, each of size $m_2^{(l)} \times m_3^{(l)}$. The $i^{(th)}$ feature map in layer l, denoted $Y_i^{(l)}$ is computed as:

$$Y_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K_{i,j}^{(l)} * Y_i^{(l-1)} \tag{1.3}$$

Where $B_i^{(l)}$ is the bias matrix introduced as an external input. $K_{i,j}^{(l)}$ is the filter of size $(2h_1^{(l)} + 1) \times (2h_2^{(l)} + 1)$ used to connect the j^{th} feature map in layer $l-1$ with the i^{th} feature map in layer l.

The size of output feature maps is influenced by borders effect. In other words, the discrete convolution can be applied only on pixels where the sum of equation (1.3) is defined properly.

Thus, output feature maps have size

$$m_2^{(l)} = m_2^{(l-1)} - 2h_1 + 2p \text{ and } m_3^{(l)} = m_3^{(l-1)} - 2h_1 + 2p \tag{1.4}$$

Where p denotes the amount of zero padding used.

Pooling and sub-sampling layer

Another important feature in CNNs is the pooling operation. It aims at reducing the size of extracted feature maps via a non linear subsampling function. Max-pooling is the most common used function. As illustrated in Fig. 1.4, max-pooling consists in partitioning the input feature map into a set of non-overlapping rectangles. For each rectangle, the maximum value is retained as output. Max pooling is then, useful for two reasons; first, it reduces computation for the next layers. It assures also a form of local translation invariance by eliminating non-maximal values. In addition to this technique, pooling can also be performed with other functions such as average pooling which takes the average value in each rectangle.

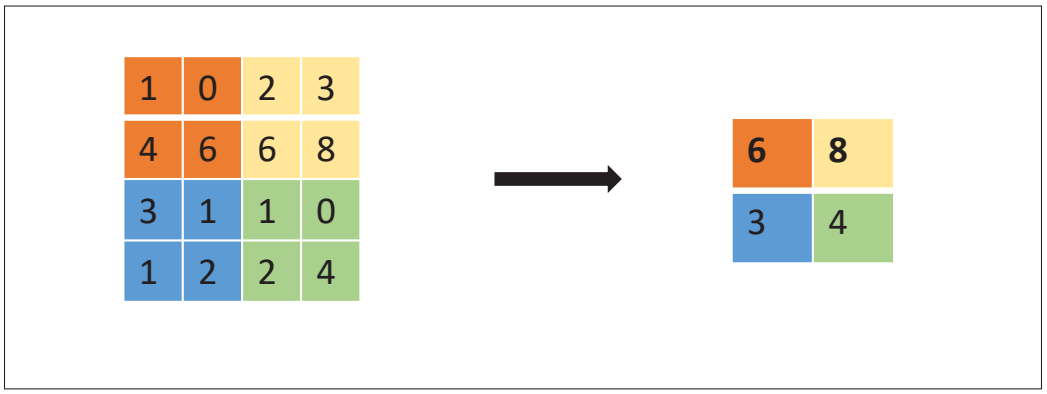


Figure 1.4 Max-Pooling with a 2x2 filter

ReLu layer

In order to improve processing efficiency, a ReLu (Rectified linear Unit) layer is typically used. It is a layer of neurons with the non-saturating activation function $f(z) = max(0,z)$. This

rectifier has been argued to be more practical than many other functions since it adds a non linearity to the network while allowing faster and effective training for deep neural architectures on large and complex datasets.

Fully connected layer

After alternating between several convolutional and pooling layers, the final classification is typically done via fully connected layers. Neurons in a fully connected layer l have connections with all neurons in the previous layer $l-1$, as seen in multi-layer perceptrons. Generally, one or more fully connected layers are used for classification purposes based on computed features in previous convolutional layers.

Dropout layer

Neural networks contain many non linear layers and a lot of parameters to adjust during training. With a limited training data, an overfitting problem can occur. It consists in a poor predictive performance; the system overreacts to small fluctuations in the training data. Thus, the system is not able to perform well on unseen data. To solve this problem, the dropout technique was proposed by (N.Srivastava *et al.* (2014)) as a simple method to prevent neural networks from overfitting. The idea is illustrated in Fig. 1.5. In a given layer, some neurons are dropped out, that is set to zero. For each training example, a different set of neurons to drop is randomly chosen. This technique is used only in the training step. At test time, all units are used.

1.2.4 Activation functions

There is a variety of common activation functions used in neural networks. They can be divided into two main categories:

- Linear functions, e.g, the identity:

$$f(z) = z \tag{1.5}$$

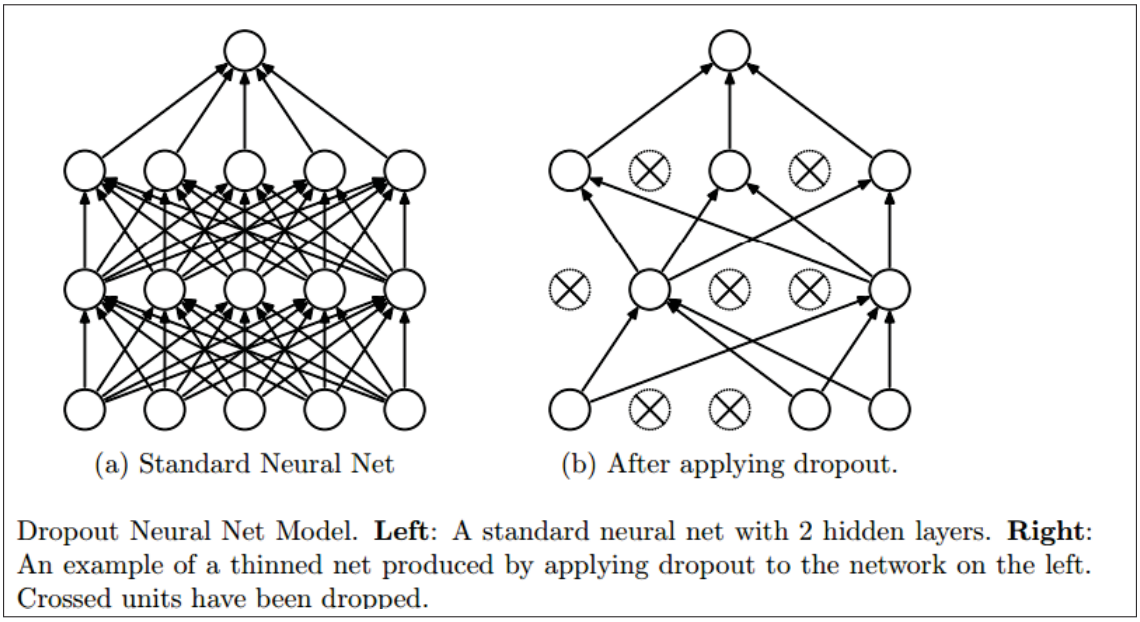


Figure 1.5 Droptout method as introduced by N.Srivastava *et al.* (2014) (Taken from the same paper)

When used in a multilayer architecture, linear functions make the network equivalent to a single-layer model and fail to solve complex non-linear problems.

- Non linear functions: For classification purposes, Sigmoid functions are often used. In fact, a sigmoid function is a bounded differentiable function defined for all real input values and has an 'S' shaped curve. Thus, a large variety of sigmoid functions have been used as activation functions including the logistic and hyperbolic tangent functions which are linearly-related by:

$$2f(z) = 1 + tanh(\frac{z}{2}) \tag{1.6}$$

where f is the logistic function defined as:

$$f(z) = \frac{1}{1+\exp(-z)} \tag{1.7}$$

and tanh is the hyperbolic function:

$$tanh(z) = \frac{1-\exp(-2z)}{1+\exp(-2z)} \tag{1.8}$$

Softsign is another activation function:

$$f(z) = z/(1+|z|) \tag{1.9}$$

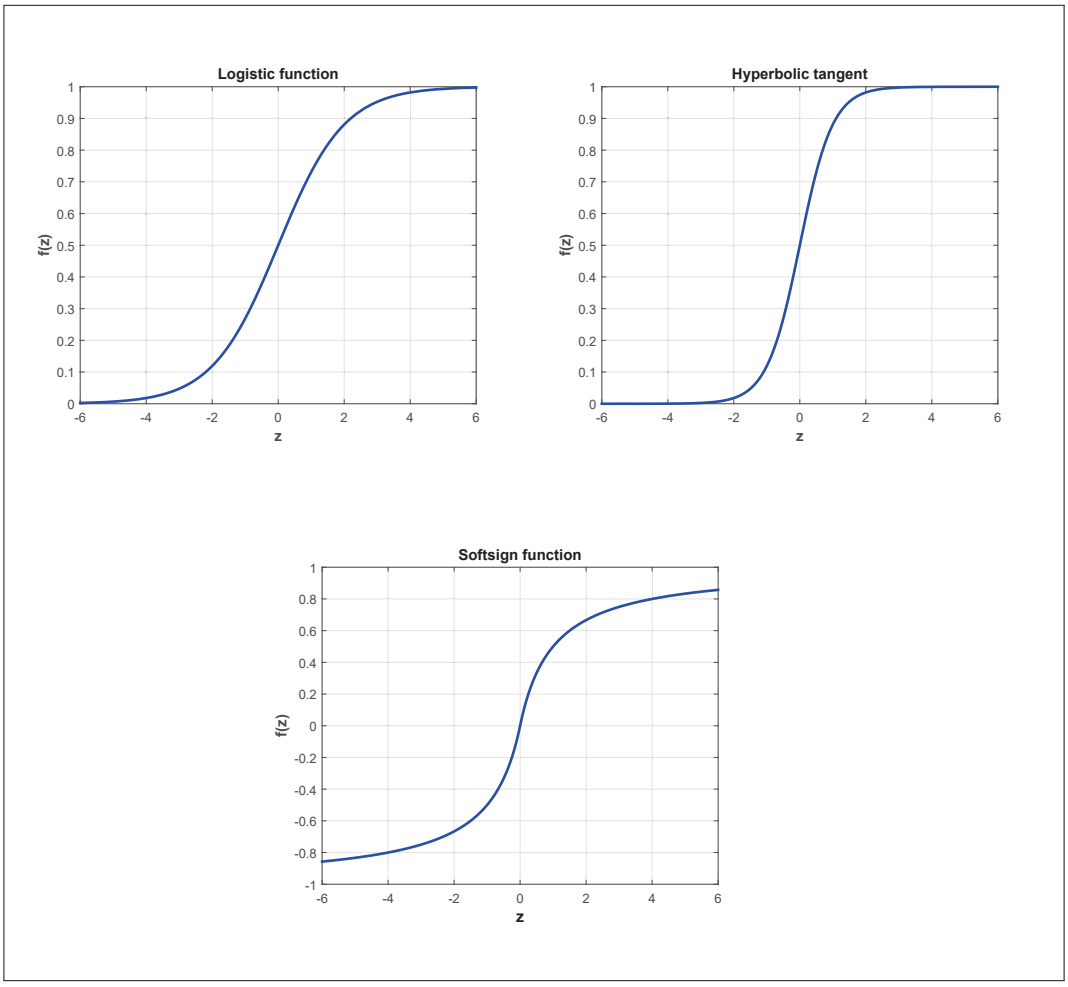


Figure 1.6 Activation functions

In deep learning, the choice of the activation function is a challenging task since it affects the training quality and the hole system performance. A comparative study between activation

functions in deep architectures is presented in (Glorot & Bengio (2010)). Experiments show that logistic and hyperbolic functions are less suited for deep learning due to their fast saturation and their slow convergence. Better performance is reported with softsign function.

The network seems to be more robust to the initialization procedure. Although these smooth non linear functions were mathematically inspired by the biological neuron behavior, most of current methods are using the *ReLU* function f defined by:

$$f(z) = max(0, z) \tag{1.10}$$

Authors in (Skourt *et al.* (2018); Paszke *et al.* (2016)) argue that it achieves better results in deep networks as it makes the activation sparse and more efficient. However, the horizontal line of the function (where $z < 0$) causes some problems when the activation function stacks in this region. In such case, the gradient of corresponding neurons will be zero and corresponding weights would not be adjusted during descent. This is called the *Dying ReLU problem*. To mitigate that, a modified version called Leaky RELU (LReLU) was proposed by Xu *et al.* (2015) introducing a light slope in the negative part of the function.

1.2.5 Basic Loss functions

Network training is an important task in supervised learning. It consists in approximating the network parameters (weights) using a training dataset composed of N labelled samples:

$$T_N := (\mathbf{I_n}, \mathbf{g_n}) : 1 \leq n \leq N \tag{1.11}$$

where $\mathbf{I_n}$ is the network input and $\mathbf{g_n}$ is the desired output.

The objective is to adjust the network parameters θ in order to minimize the so-called **loss function** that evaluates the gap between the network output $\mathbf{s}_{\theta,\mathbf{n}}$ and the desired output $\mathbf{g_n}$ for each training sample $\mathbf{I_n}$. The loss function is used to compute the network errors. Commonly used

mathematical expressions include the mean squared loss (*MSL*), the cross entropy(*CE*) and more recently the Dice loss (*DL*).

The mean squared error measure is given by :

$$MSL(\theta) = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{d}(s_{\theta,n,k} - g_{n,k})^2 \tag{1.12}$$

where d is the number of labels per sample. In practice, the mean squared (or quadratic) measure gives good results. However, if the sigmoid activation function is used, the quadratic loss would converge slowly to the final output. To resolve this problem, the *CE* was proposed:

$$CE(\theta) = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{d} g_{n,k}\log(s_{\theta,n,k}) \tag{1.13}$$

More recently (Milletari *et al.* (2016)), the Dice measure, initially used as a similarity metric to evaluate the segmentation quality, was introduced as a loss function for two-class segmentation problem. It is defined as:

$$DL = 1 - \frac{\sum_{n=1}^{N} s_{\theta,n}g_n + \varepsilon}{\sum_{n=1}^{N} s_{\theta,n} + g_n + \varepsilon} - \frac{\sum_{n=1}^{N}(1 - s_{\theta,n})(1 - g_n) + \varepsilon}{\sum_{n=1}^{N} 2 - s_{\theta,n} - g_n + \varepsilon} \tag{1.14}$$

Where ε is a small constant to avoid the problem of dividing by zero.

1.3 Deep learning for medical image segmentation: Overview, successes and challenges

1.3.1 Deep learning: from natural to medical imaging

The few last decades have witnessed a digital revolution and fast technical advances in medical imaging with the explosion of new imaging modalities, e.g, X ray computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound (US). Hence, experts are no longer able to manually process the huge amount of data and extract useful informations. Searching for new paths, the medical image community has been attracted by the astonishing successes of

deep learning techniques in natural imaging and have applied a lot of these findings to improve medical image analysis for different tasks including classification, detection, segmentation and registration. Image segmentation is a one of the most important tasks in medical image analysis. It consists of partitioning the image into different segments that often correspond to organs, tissue categories, pathologies or other biological structures. An accurate delineation allows efficient quantitative analysis of important clinical measures related to volume or shape as in cardiac or cerebral image analysis. Such measures are very useful for diagnostic medicine and clinical interventions.

Recently, deep learning methods have shown their ability to deliver excellent performances in such task. Many methods have been proposed for segmentation of specific organs or other substructures, e.g, brain (Pereira *et al.* (2016); Brébisson & G.Montana (2015); M.Havaei *et al.* (2015); Chen *et al.* (2018); Ganaye *et al.* (2018)) , heart cavities and muscles (M.Havaei *et al.* (2015); Yu *et al.* (2017a); Avendi *et al.* (2016)), vertebrae (Chen *et al.* (2015)), bone (Minnema *et al.* (2018); Zhang *et al.* (2010); Klein *et al.* (2018)) or abnormalities such as tumors (M.Havaei *et al.* (2015); Pereira *et al.* (2016); Menze *et al.* (2014)) or lesions (Hwang & Kim (2016); Bellver *et al.* (2017)). During the last years, thousands of new papers have been published. However, it is worth to note that most of them suffer from reproducibility problems. This can be either a problem of methodology or also lack of details in the published paper as the use of hyper-parameters or heuristics specific to a particular dataset. In some cases, researchers can validate their method only on some selected scans and omit the rest which biases their results. One way to guarantee a fair and direct comparison of proposed methods on the same dataset and evaluation procedure is to organize public challenges. From 2007, around 36 public challenges[1] and 118 hidden ones have been organized to deal with specific medical image analysis problems, e.g., brain lesion segmentation (BRATS[2](2012-2018), ISLES[3], WMH[4],

[1] https://grand-challenge.org/stats

[2] http://braintumorsegmentation.org/

[3] http://www.isles-challenge.org/

[4] http://wmh.isi.uu.nl

MSSEG[5]), cardiac image segmentation (ACDC[6], LVSC [7], MM-WHS 2017[8], CETUS2014[9]), liver segmentation (CHAOS[10]), etc. Deep neural networks have been the winners of most of the challenges.

The typical idea of a DNN-based approach is to train one or more deep networks to assign each pixel (2d input) or voxel (3D input) to its corresponding anatomical region based on different input features. As the architectural choice is an important key for a good segmentation, dozens of different designs have been introduced. Initially, the common used strategy was *patch-based* (Avendi *et al.* (2016); Pereira *et al.* (2016); Brébisson & G.Montana (2015)). It consists in the use of the neighborhood of each pixel in the image as input (for example: 3x3 or 5x5 window centered on the pixel, called a *patch*). A variety of features was proposed: 2D and 3D intensity patches (Brébisson & G.Montana (2015); Lai (2015); Kleesiek *et al.* (2016)), orthogonal views of MRI scans (A.Prasoon *et al.* (2013)). One straightforward drawback of a *sliding-window* strategy is that it doesn't take into account the global context of the image which is needed to segment medical images. By changing the last fully connected layers by convolutional ones, the DNN becomes able to process the entire image in a single forward pass and produce a probability map. This was the idea of a *fully convolutional neural network* proposed by (J.Long *et al.* (2015)) for natural image segmentation. Influenced by the FCNN idea, Ronneberger et al. (Ronneberger *et al.* (2015a)) have proposed the famous U-net architecture to segment biomedical images. Authors combined a regular FCNN with an up-sampling path where "deconvolutions" are applied to build an output of the same size than the input. The particularity of a U-net is the insertion of "*skip connections*" to connect opposing; convolutional and deconvolutional, layers between the down-sampling and up-sampling paths. To exploit the volumetric context of medical data, an extension to 3D U-Net was suggested (Çiçek *et al.*

[5] https://portal.fli-iam.irisa.fr/msseg-challenge/overview

[6] https://www.creatis.insa-lyon.fr/Challenge/acdc/index.html

[7] http://www.cardiacatlas.org/challenges/lv-segmentation-challenge/

[8] http://www.sdspeople.fudan.edu.cn/zhuangxiahai/0/mmwhs/

[9] https://www.creatis.insa-lyon.fr/Challenge/CETUS/index.html

[10] https://chaos.grand-challenge.org/

(2016a)). Milletari et al. (Milletari *et al.* (2016)) have proposed an other variant of U-Net for 3D prostate segmentation. In this study, authors used a loss function based on the Dice coefficient to mitigate the problem of imbalanced data.

1.3.2 Lack of annotated data

Fully supervised approaches (Çiçek *et al.* (2016b); J.Long *et al.* (2015); Avendi *et al.* (2016); Brébisson & G.Montana (2015); M.Havaei *et al.* (2015); LC.Chen & G.Papandreou (2015)) rely on available ground truth sets derived generally from comprehensive manual annotations by experts. Initially, the majority of researches have focused on supervised deep learning. However, they require huge training dataset, which is not easy to meet in the medical field. The non-availability of large and high-quality labeled data still be the biggest barrier for the performance of deep learning in medical imaging. In fact, an accurate annotation is a difficult and time consuming task. An expert may spend 3 or 4 days to segment a full brain. On the other hand, manual annotations require multiple expert opinions to get over the human error. In some cases, annotation may not be possible due to the absence of experts or due to the rarity of the target disease or lesion. To address this problem, different methods and techniques have been proposed, e.g. transfer learning (Tran (2016); Shin *et al.* (2016); Yu *et al.* (2017a); Ghafoorian *et al.* (2017); Tajbakhsh *et al.* (2019)) and data augmentation (Asperti & Mastronardo (2018); Vasconcelos & Vasconcelos (2017); Tran (2016); Pereira *et al.* (2016); Vasconcelos & Vasconcelos (2017)).

Transfer learning is a common strategy that has shown its usefulness especially in cases of limited labeled data. The basic idea is to use a pre-trained network on a large dataset and re-apply it on another dataset. Typically, a fine-tuning step is performed to adapt the network to the given task of interest. In (Yu *et al.* (2017a)), authors use dynamic CNNs to segment fetal left ventricle in echocardiographic sequences. Initially, a pre-trained model is fine-tuned by a deep tuning using the first annotated frame then a shallow tuning for the remaining frames. Conducted on 51 echocardiographic sequences, dynamic networks achieve good results in terms of Dice accuracy of 94.5%. As another way to mitigate the lack of annotations, recent methods

have focused more on semi and weak supervision. A variety of weak labels such as image tags (Pathak *et al.* (2015); Papandreou *et al.* (2015)), partial labels, bounding boxes (Shah *et al.* (2018); Dai *et al.* (2015); Papandreou *et al.* (2015)) and scribbels (Wang *et al.* (2017a,b); Kervadec *et al.* (2019)) have been used. In a very recent work, Kervadec *et al.* (2019) exploit weak annotations, e.g., scribbles and dots for semantic segmentation. They have proposed new constrained losses based on anatomical priors (shape, size) to guide the learning process during training. Validated on three medical applications, the method showed promising results close to full supervision performances.

1.3.3 Class unbalanced problems

While achieving a lot of successes in many complex medical problems, researchers have realized that there are still some challenges with deep learning techniques to deal with. In addition to lack of annotations discussed previously, another major issue is the data unbalance that occurs when the target is underrepresented in the image. This is very common in health field, for instance, when the task consists in segmenting abnormalities, rare diseases or small regions. For example, to segment a brain cancer, the frequency of the target class (cancer) can be 1000 times less than the majority class (healthy patient). Assuming that all the classes are equally distributed, classic classifiers are highly biased toward the majority class and fail to deliver good segmentation performances. To combat this problem, three main strategies have been suggested; methods operating directly on the data via down-sampling or over-sampling, methods operating on the learning process and methods introducing specific loss functions. A common solution; which falls into the first category, is to re-balance the data by down-sampling frequent labels (Havaei *et al.* (2017); Valverde *et al.* (2017)). Others perform a data augmentation on target samples (Kamnitsas *et al.* (2017); Litjens *et al.* (2016); Pereira *et al.* (2016)).

From a different point of view, authors in (M.Havaei *et al.* (2015)) propose a *two-phase training* for brain tumor segmentation; the first is performed using equi-probable patches to adjust lower layers parameters. Then, in a second step, they re-train only the output layer to take into

account the unbalanced nature of real patches. The same strategy was used in (Minnema *et al.* (2018)) to segment bone in CT scans. A considerable number of recent works has focused on the adaptation of the loss function to alleviate data imbalance issue. A common practice is to assign weights to the different classes, inversely proportional to the frequency of the corresponding labels (Brosch *et al.* (2015); Ronneberger *et al.* (2015b); Kamnitsas *et al.* (2017); Yu *et al.* (2017b)). In (Ronneberger *et al.* (2015b)), the standard cross entropy (CE) loss is modified by adding weights that assign more importance to the rare class than to the majority class. In a distinguished way, Sudre et al. (Sudre *et al.* (2017)) generalized the Dice loss (*GDL*) initially proposed by Milletari et al. (Milletari *et al.* (2016)). To have better trade-off between precision and recall, some recent works(Salehi *et al.* (2017))investigated losses based on the Tversky similarity index. Inspired by the concept of focal loss (Lin et al., 2018), Dice and Tvserky losses have been extended to integrate a focal term, which is parameterized by a value that controls the importance between easy and hard training samples (Abraham & Khan (2018); Wong *et al.* (2018); Hashemi *et al.* (2018)).

CHAPTER 2

BOUNDARY LOSS FOR HIGHLY UNBALANCED SEGMENTATION

In the present chapter, we propose a new method that tackle imbalanced medical segmentation problems. First, we describe the motivation behind the idea of our method. Then, we explain the proposed methodology via a detailed mathematical background.

2.1 Motivation

During the last decades, fully deep neural networks have shown their great potential to deliver the best performances in medical image analysis. However, one of the major challenges encountred is the data unbalance problem where the size of target region is several orders of magnitude less than the background size. Such problem occurs mostly in cases of abnormalities detection/segmentation, e.g tumors segmentation (brain (M.Havaei *et al.* (2015); Pereira *et al.* (2016); Menze *et al.* (2014); Hwang & Kim (2016); Bellver *et al.* (2017)), breast (Kallenberg *et al.* (2016); Wang *et al.* (2016); Hassanien *et al.* (2014)), lung (Kalinovsky & Kovalev (2016); Mi *et al.* (2014); Skourt *et al.* (2018)), etc). For example, to segment a brain cancer, the frequency of the target class (cancer) can be 1000 times less than the majority class (healthy patient). In such cases, common approaches, assuming that all patterns are evenly distributed, fail to deliver good segmentation performance. The issue is that they tend to favour the majority class and, then, misclassify the target class.

As discussed previously in section (1.3.3), similarity loss functions have been proposed to mitigate the unbalance problem, e.g., weighted cross entropy (Ronneberger *et al.* (2015b)) and the generalized dice loss (GDL) (Sudre *et al.* (2017)). Although effective for some unbalanced problems, weighted cross entropy may undergo serious difficulties when dealing with highly unbalanced datasets. The CE gradient computed over the few pixels of infrequent labels is typically noisy, and amplifying this noise with a high class weight may lead to instability. To address this problem, many recent researches (Sudre *et al.* (2017); Milletari *et al.* (2016) relied on GDL as an alternative. Despite the improvements over CE , the dice loss has some

limitations in a highly unbalanced problem. Initially defined as an overlap measure to evaluate the network performance, the dice coefficient accounts only the correctly classified pixels/ voxels without taking into account the spatial distribution. As shown in Fig.2.1 and Fig.2.2, Dice coefficient gives same values in both cases regardless the distance that separates segments. Thus, when used as a loss function, the information of how far is the prediction from the ground truth is missing. Moreover, regional dice losses still be not convenient for small region segmentation such as lesions, tumors or small structures. In fact, the probability of low overlap between two small regions is high. In such case, dice loss would give a small value regardless of how far regions are from each other.

Based on all the above-mentioned limitations of *region*-based losses, we propose a new *boundary* loss that takes the form of a distance metric on the space of contours (or shapes), not regions. We argue that a boundary loss can mitigate the issues related to regional losses in highly unbalanced segmentation problems. Moreover, it provides information that is complimentary to regional losses.

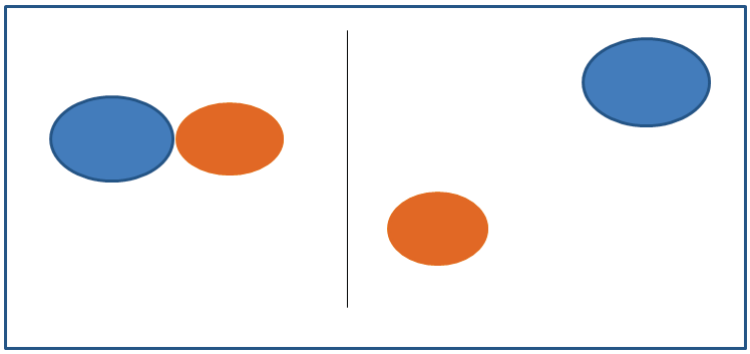


Figure 2.1 Case of zero Dice

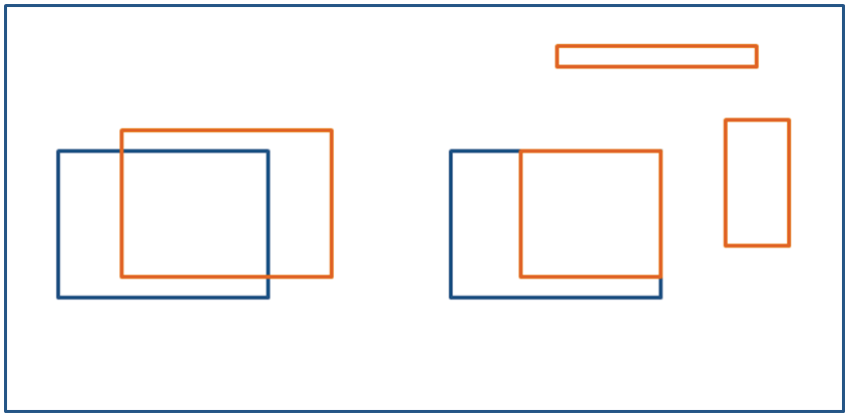


Figure 2.2 Same Dice with spatial variance

2.2 Methodology

2.2.1 Proposed boundary loss

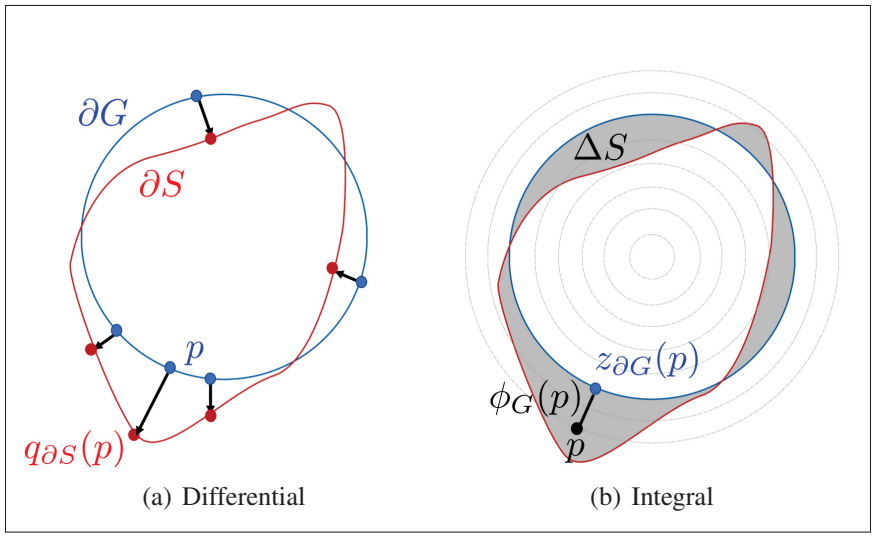
Here, we introduce the formulation of the new proposed loss. Let $I : \Omega \subset \mathbb{R}^{2,3} \rightarrow \mathbb{R}$ be an input image with spatial domain Ω, and $g : \Omega \rightarrow \{0,1\}$ a binary ground-truth segmentation of the image: $g(p) = 1$ if pixel/voxel p belongs to the target region $G \subset \Omega$ (foreground region) and 0 otherwise. Let $s_\theta : \Omega \rightarrow [0,1]$ denotes the softmax probability output of a deep segmentation network, and $S_\theta \subset \Omega$ the corresponding segmentation region:z $S_\theta = \{p \in \Omega \,|\, s_\theta(p) \geq \delta\}$ for some threshold δ. Widely used segmentation loss functions involve a *regional integral* for each segmentation region in Ω, which measures some similarity (or overlap) between the region defined by the probability outputs of the network and the corresponding ground-truth. In the two-region case, we have an integral of the general form $\int_\Omega g(p) f(s_\theta(p)) dp$ for the foreground, and of the form $\int_\Omega (1 - g(p)) f(1 - s_\theta(p)) dp$ for the background. The generalized Dice loss (GDL) (Sudre *et al.* (2017)) involves regional integrals with $f = 1$, subject to some normalization, and is given as follows for the two-region case:

$$\mathscr{L}_{GD}(\theta) = 1 - 2\frac{w_G \int_{p\in\Omega} g(p) s_\theta(p) dp + w_B \int_{p\in\Omega} (1 - g(p))(1 - s_\theta(p)) dp}{w_G \int_\Omega [s_\theta(p) + g(p)] dp + w_B \int_\Omega [2 - s_\theta(p) - g(p)] dp} \tag{2.1}$$

where coefficients $w_G = 1/\left(\int_{p\in\Omega} g(p)dp\right)^2$ and $w_B = 1/\left(\int_\Omega (1-g(p))dp\right)^2$ are introduced to reduce the well-known correlation between the Dice overlap and region size.

Now, our objective is to build a boundary loss $\text{Dist}(\partial G, \partial S_\theta)$, which takes the form of a distance metric on the space of contours (or region boundaries) in Ω, with ∂G denoting a representation of the boundary of ground-truth region G (e.g., the set of points of G, which have a spatial neighbor in background $\Omega \setminus G$) and ∂S_θ denoting the boundary of the segmentation region defined by the network output. On the one hand, a boundary loss should be able to mitigate the above-mentioned difficulties for unbalanced segmentations: rather than using unbalanced integrals within the regions, it uses integrals over the boundary (interface) between the regions.

The main challenge here is how to represent this surface between the two boundaries ∂S_θ and ∂G as a differentiable function that can accommodate standard stochastic optimizers such as SGD. This challenge might explain why *distance*-based losses have been, to the best of our knowledge, largely avoided in the context of deep segmentation networks.



(a) Differential (b) Integral

Figure 2.3 The relationship between *differential* and *integral* approaches for evaluating boundary variation.

Our boundary loss is inspired from discrete (graph-based) optimization techniques for computing gradient flows of curve evolution (Boykov *et al.* (2006)). Similarly to our problem, curve evolution methods aim to evaluate boundary variations. The typical solution to compute the distance between contours is based on a variational method where we should compute each pixel's local differential motion (see illustration in Fig.2.3). However, similarly to any contour distance invoking directly points on boundary ∂S, the computed distance can not be used as a loss function. For that, we propose to use an approximation that expresses the differential boundary variation as a regional integral:

$$\text{Dist}(\partial G, \partial S) = 2 \int_{\Delta S} |\phi_G(p)| dp \tag{2.2}$$

where ΔS denotes the region between the two contours and $\phi_G : \Omega \to \mathbb{R}$ is a *level set* representation of boundary ∂G: $\phi_G(p)$ evaluates a signed distance between point $p \in \Omega$ and the nearest point $z_{\partial G}(p)$ on contour ∂G: $\phi_G(p) = -\|p - z_{\partial G}(p)\|$ if $p \in G$ and $\phi_G(p) = \|p - z_{\partial G}(p)\|$ otherwise. Fig. 2.3.b illustrates this integral framework for evaluating the boundary distance.

Thus, the non-symmetric L_2 distance between contours can be expressed as a sum of regional integrals:

$$\int_S \phi_G(p) dp - \int_G \phi_G(p) dp = \int_\Omega \phi_G(p) s(p) dp - \int_\Omega \phi_G(p) g(p) dp \tag{2.3}$$

where $s : \Omega \to \{0,1\}$ is binary indicator function of region S: $s(p) = 1$ if $p \in S$ belongs to the target and 0 otherwise. Now, for $S = S_\theta$, i.e., replacing binary variables $s(p)$ in Eq. (2.3) by the softmax probability outputs of the network $s_\theta(p)$, we obtain the following boundary loss which, up to a constant independent of θ, approximates boundary distance $\text{Dist}(\partial G, \partial S_\theta)$:

$$\mathscr{L}_B(\theta) = \int_\Omega \phi_G(p) s_\theta(p) dp \tag{2.4}$$

Notice that we omitted the last term in Eq. (2.3) as it is independent of network parameters. The level set function ϕ_G is pre-computed directly from the ground-truth region G. In practice, our boundary loss in Eq. (2.4) is the sum of linear functions of the regional softmax probability

outputs of the network. Therefore, it can be easily combined with standard regional losses and implemented with any existing deep network architecture for N-D segmentation. In the experiments, we tested our boundary loss as a stand alone loss in a first step. Unfortunately, we did not get competitive results. We believe that this is due to the following technical facts. In theory, the global optimum of our boundary loss corresponds to a negative value (when the softmax probabilities correspond to a non-empty foreground). However, an empty foreground (null values of the softmax probabilities almost everywhere) corresponds to low gradients. Therefore, this trivial solution is close a local minimum or a saddle point. To avoid getting stuck in such trivial solution, we propose to use our loss in conjunction with the regional Generalized Dice loss:

$$\alpha \mathscr{L}_{GD}(\theta) + (1-\alpha)\mathscr{L}_{B}(\theta) \tag{2.5}$$

Finally, it is worth noting that our boundary loss uses ground-truth boundary information via pre-computed level-set function $\phi_G(p)$, which encodes the distance between each point p and ∂G. In Eq. (2.4), the softmax for each point p is weighted by the distance function. Such distance-to-boundary information is omitted in widely used regional losses, where all the points within a given region are treated equally, independently of their distances from the boundary.

2.2.2 Scheduling strategy

In our work, we propose to combine the new boundary-loss with the GDL as it has shown good results when dealing with unbalanced problems (Nie *et al.* (2018); Shen *et al.* (2018); Horváth *et al.* (2018); Górriz *et al.* (2018)). The task of finding the ideal combination between the two losses is not obvious. We argue that it is not very convincing to use an empirical fixed weight. For that, we propose to use a simple scheduling strategy. During training, the value of α in Eq. (2.5) is initialized to 1, then decreased by 0.01 after each epoch until reaching 0.01. Thus, we give more importance to the GDL in first epochs then we gradually increase the impact of our boundary loss.

2.2.3 Network architecture

It is worth to note that the choice of the network architecture is important for medical image image segmentation. To validate our method, we employed UNet (Ronneberger *et al.* (2015b)) as a deep learning architecture. In fact, the network is initially designed as a combination of a regular FCNN with an up-sampling path where "deconvolutions" are applied to build an output of the same size than the input. As shown in Fig. 2.4, the particularity of a U-net is the insertion of "*skip connections*" to connect opposing; convolutional and deconvolutional, layers between the down-sampling and up-sampling paths.

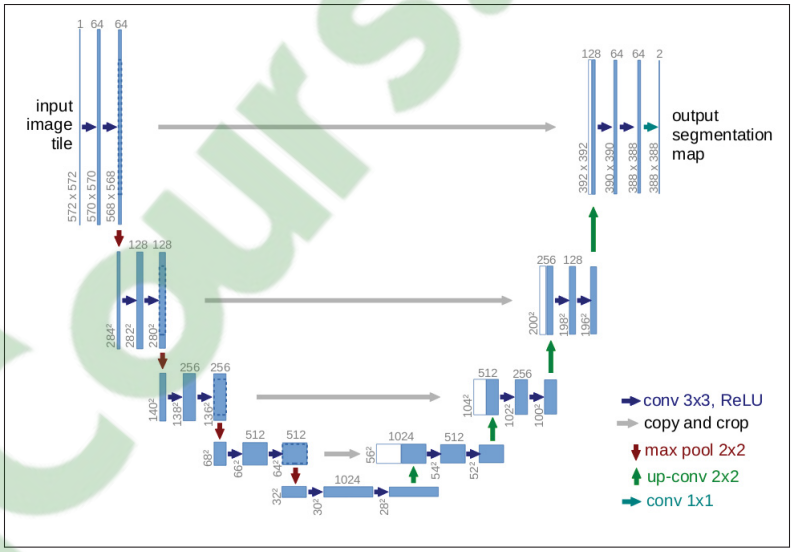


Figure 2.4 Original UNet architecture
(taken from (Ronneberger *et al.* (2015b))

CHAPTER 3

EXPERIMENTS AND RESULTS

3.1 Datasets

To evaluate our proposed method, we selected three challenging public datasets for brain lesion segmentation where classes are highly unbalanced: the ischemic stroke lesion (ISLES), white matter hyperintensities (WMH) and Multi modal brain tumor segmentation (BRATS17). WMH and ISLES segmentation correspond to binary problems while BRATS segmentation is a multi-class problem.

WMH

The public dataset of White Matter Hyperintensities (WMH) [1] MICCAI 2017 challenge is composed of 60 subjects for training and 110 subjects for testing. Images are acquired from five different scanners belonging to three different vendors in three various hospitals. Each subject has 3D T1-weighted scan and 2D multi-slice FLAIR. The ground truth of the 60 subjects of training set is defined only on FLAIR images. From a total of 60 subjects, training was performed on 48 subjects while validation on the 12 remaining ones. Patients were chosen randomly from three scanners and with equal proportions in order to have representative sets. The percentage of lesions in the validation fold ranges from 0.01 % to 0.43% of voxels, with a mean of 0.25% between patients, std of 0.32% and median of 0.04%.

ISLES

The training dataset provided by the ISLES organizers is composed of 94 ischemic stroke lesion multi-modal scans. In our experiments, we split this dataset into training and validation sets containing 74 and 20 examples, respectively. Each scan contains Diffusion maps (DWI)

[1] http://wmh.isi.uu.nl

and Perfusion maps (CBF, MTT, CBV, Tmax and CTP source data), as well as the manual ground-truth segmentation. More details about the data can be found in the ISLES website[2].

BRATS

The public dataset for the Multimodal Brain Tumor Segmentation Challenge 2017 (BRATS17)[3] is composed of multi-modal scans acquired from different 19 institutions. Provided scans are divided into two categories; subjects with *high grade Glioblastoma (HGG)* and subjects with *Low Grade Glioblastoma (LGG)*. Manual annotations cover three different regions; non enhancing tumor (NET), Peritumoral Edema (ED) and Enhacing Tumor (ET). Our experiments were performed on the training set of a total of 285 cases that were split into 80 % for training and 20 % for validation.

3.2 Experimental protocol

3.2.1 Data Pre-processing

Before training the network, a pre-processing step was perfomed. While the scans are provided as 3D images, we process them as a stack of independent 2D images, which are fed into the network. During the pre-processing, all scans are normalized between 0 and 1 and saved as a set of 2D matrices. As mentioned before, WMH scans are from different scanners and with different sizes. Thus, all scans are re-scaled to a uniform size 200×200 pixels via cropping/-padding operations. On the other hand, BRATS scans are normalized and saved as 2D matrices of 240×240 pixels. To test the impact of our loss on a very basic setting, We did not use any data augmentation during experiments.

[2] http://www.isles-challenge.org

[3] https://www.med.upenn.edu/sbia/brats2017/data.html

3.2.2 Implementation details

In our experiments, we employed Adaptive Moment Estimation (Adam) (Kingma & Ba (2014)) optimizer with a batch size of 10 and an initial learning rate equal to 0.001. We trained a UNet model over 200 epochs for WMH and ISLES datasets and 100 epochs for BRATS without an early stopping. Distance maps or level-set functions ϕ_G in Eq. ((2.4)) were computed using standard SciPy functions [4]. In case of slices with only background (no target class), we used a zero-distance map, assuming that the GDL, used alone, is more suitable in those cases.

3.3 Results for binary problems

We trained and validated our network on WMH and ISLES training datasets. Table 3.1 reports the Dice score (DSC) and the Hausdorff distance (HD) performance for our experiments using GDL alone and the loss we proposed in Eq. (2.5). We note that adding our boundary loss jointly with GDL ameliorates efficiently the segmentation performance which is reflected in significant better values of DSC and HD. On the WMH dataset, we had an improvement of around 2% of the DSC while it achieves 8% on the ISLES segmentation task. The same behaviour is observed for the HD metric with a larger gain on ISLES dataset.

Table 3.1 DSC and HD values achieved on ISLES and WMH validation subsets. The values represent the mean performance (and standard deviation) of 2 runs for each setting.

	ISLES		**WMH**	
Loss	**DSC**	**HD (mm)**	**DSC**	**HD (mm)**
$\mathcal{L}_{GD}$	0.575 (0.028)	4.009 (0.016)	0.727 (0.006)	1.045 (0.014)
$\mathcal{L}_{GD} + \mathcal{L}_B$	**0.656 (0.023)**	**3.562 (0.009)**	**0.748 (0.005)**	**0.987 (0.010)**

Figure 3.1 shows the learning process behaviour of validation metrics (DSC and HD) over 200 epochs. The learning curves show the gap in performance of the GDL with respect to the

[4] https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.ndimage.morphology.distance_transform_edt.html

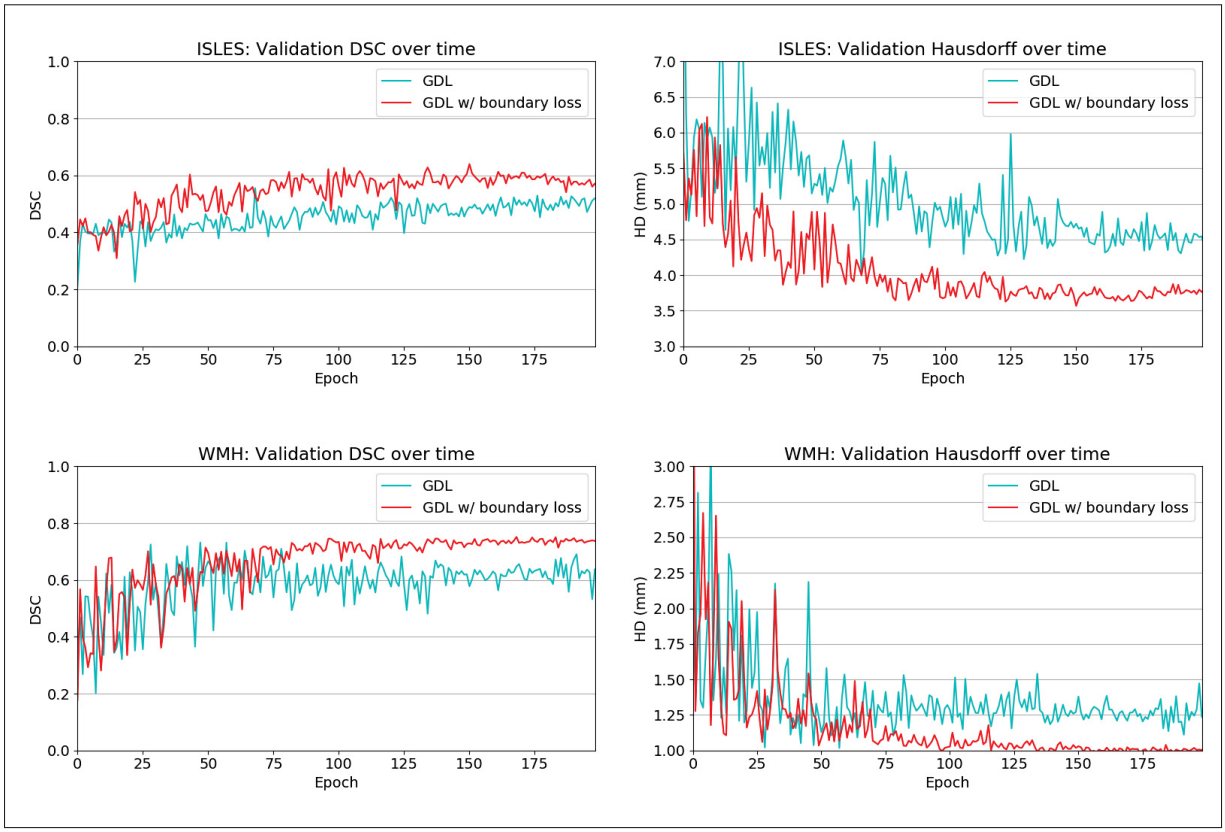


Figure 3.1 Evolution of DSC and HD values on WMH/ISLES validation subsets. The blue curve shows the performance of the network trained using the GDL loss, while the red curve represents the optimization process with the GDL combined with our proposed boundary loss.

GDL with boundary loss, where the difference becomes significant at convergence. In addition to outperforming GDL, we can also observe that the boundary loss term helps stabilizing the training process, yielding a much smoother curve as the network training converges. This behaviour is consistent for both metrics and both datasets, which clearly shows the efficiency of our proposed boundary term.

3.4 Results for multi-class problem

To investigate the behaviour of our boundary loss in a multi-class imbalanced problem, we trained and validated the same network on BRATS database for brain tumour segmentation. Quantitative results are reported in table 3.2. We note that, in a multi-class problem, our

boundary loss jointly with GDL was also able to ameliorate the segmentation performance. Significant better values of the Dice score and the Hausdorff distance are achieved. The mean DSC over classes improves of around 9%. The same behaviour is noted for the Hausdorff distance.

Table 3.2 DSC and HD values achieved on BRATS validation subset. The values represent the mean performance of 2 runs.

	BRATS					
Loss	**DSC**			**HD(mm)**		
	N.enh.	**ED**	**ET**	**N.enh.**	**ED**	**ET**
$\mathcal{L}_{GD}$	0.73	0.66	0.66	1.32	2.23	1.69
$\mathcal{L}_{GD} + \mathcal{L}_{B}$	**0.75**	**0.75**	**0.78**	1.43	**1.98**	**0.97**

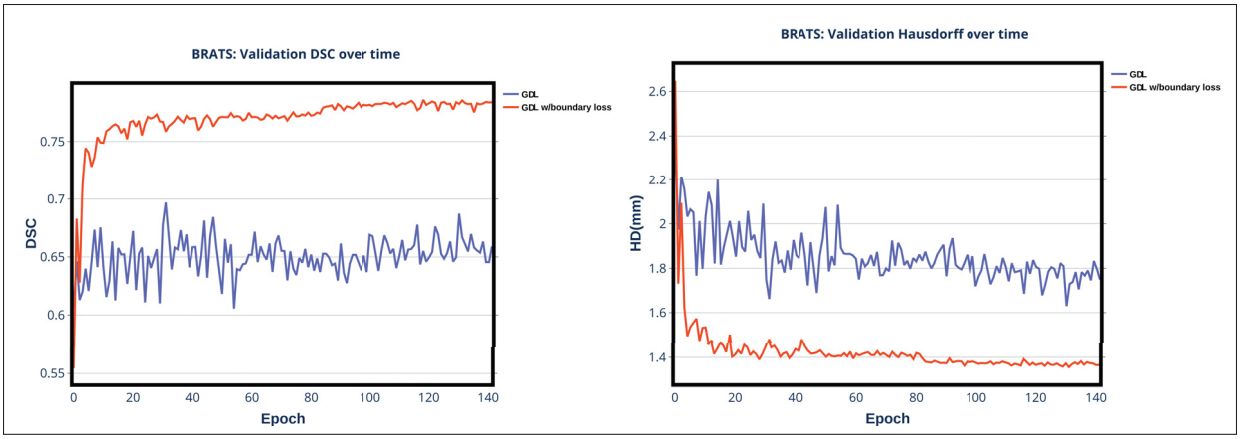


Figure 3.2 Evolution of DSC and HD values on BRATS validation subset. The blue curve shows the performance of the network trained using the GDL loss, while the red curve represents the optimization process with the GDL combined with our proposed boundary loss.

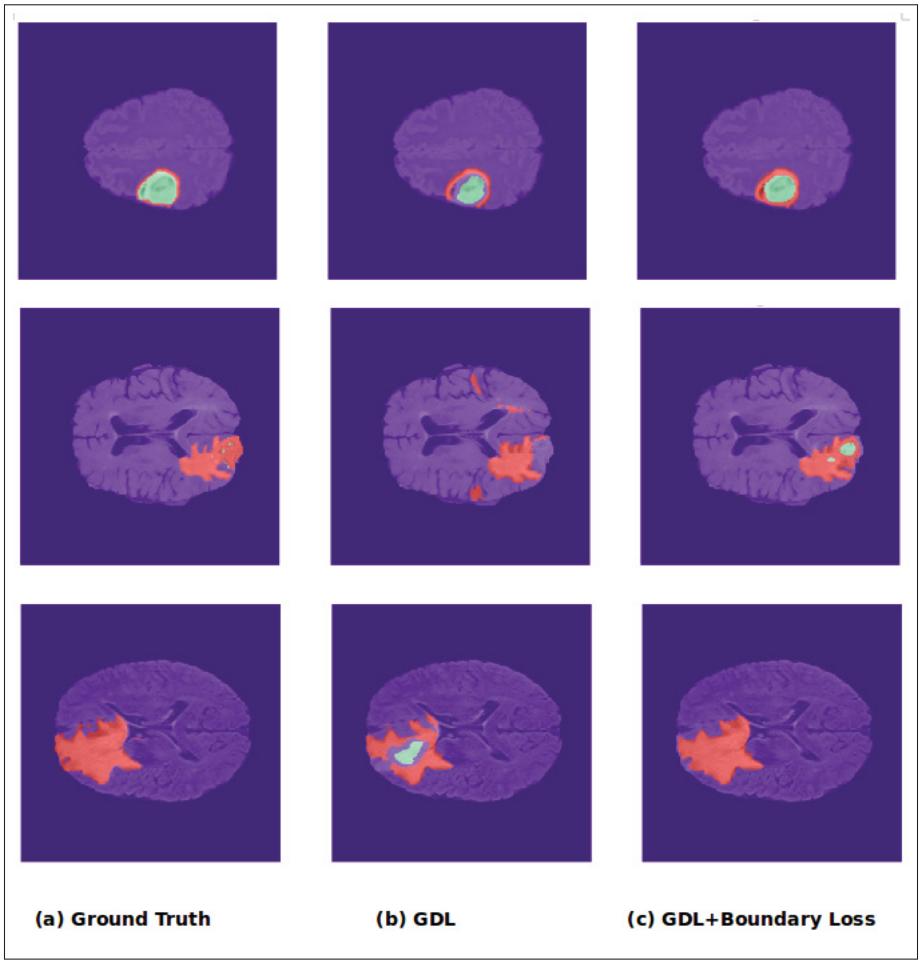


Figure 3.3 Qualitative segmentation results on BRATS validation set. Colors correspond to different tumour regions: Red for edema (ED), green for Non enhancing tumor (NET). Best viewed in colors.

3.5 Discussion

Qualitative evaluation

Inspecting qualitative results (for both binary and multi-class problems) illustrated by Fig. 3.5 and Fig.3.3, we note that there are two main types of improvements when using our boundary loss. First, as mentioned previously, region-based losses (e.g. GDL) do not include the information of how far is the prediction from the ground truth. Thus, a close detected object is treated equally than a totally missed one. However, as our boundary loss is based on the dis-

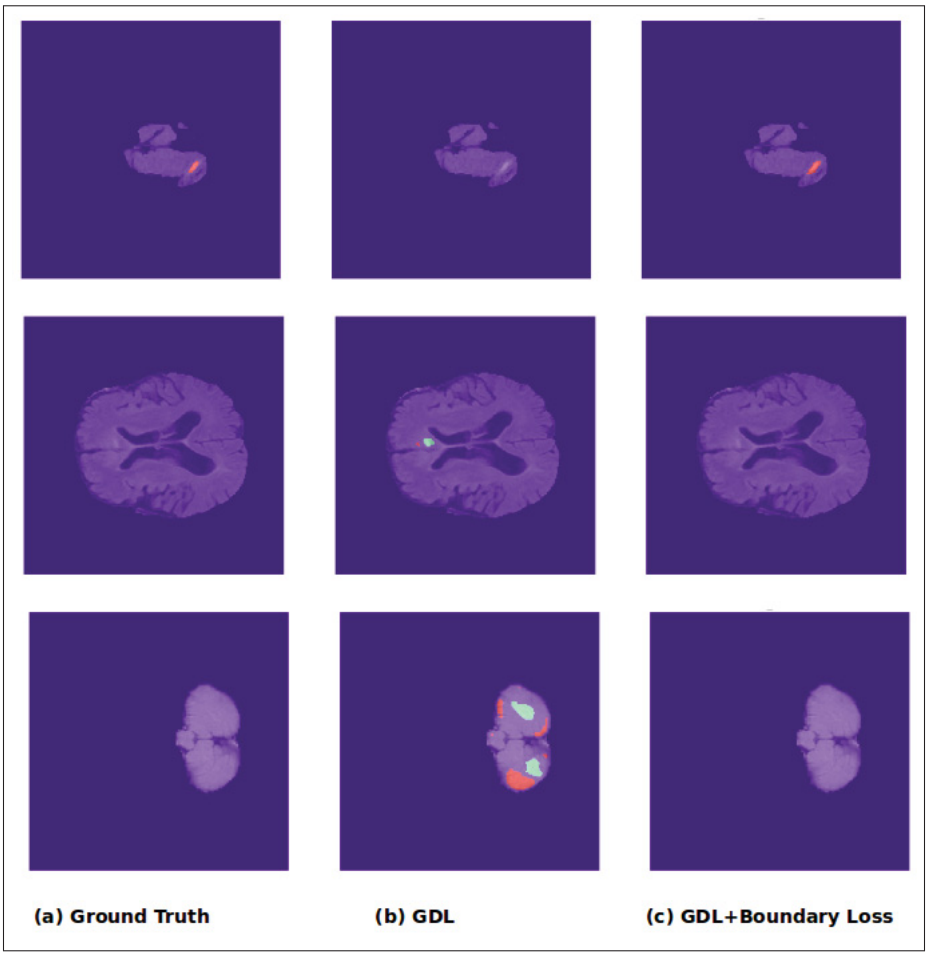


Figure 3.4 Effect of our boundary loss on false positives/ negatives. Best viewed in colors.

tance map from the ground-truth boundary ∂G, it will penalize much more such cases, helping to recover small and far regions. This effect is best illustrated in Fig. 0.1, Fig.3.5 (third row) and Fig 3.4. False positives (first row in Fig. 3.5, second and third row in 3.4) will be far away from the closest foreground, getting a much higher penalty than with the GDL alone. This helps in reducing the number of false positives.

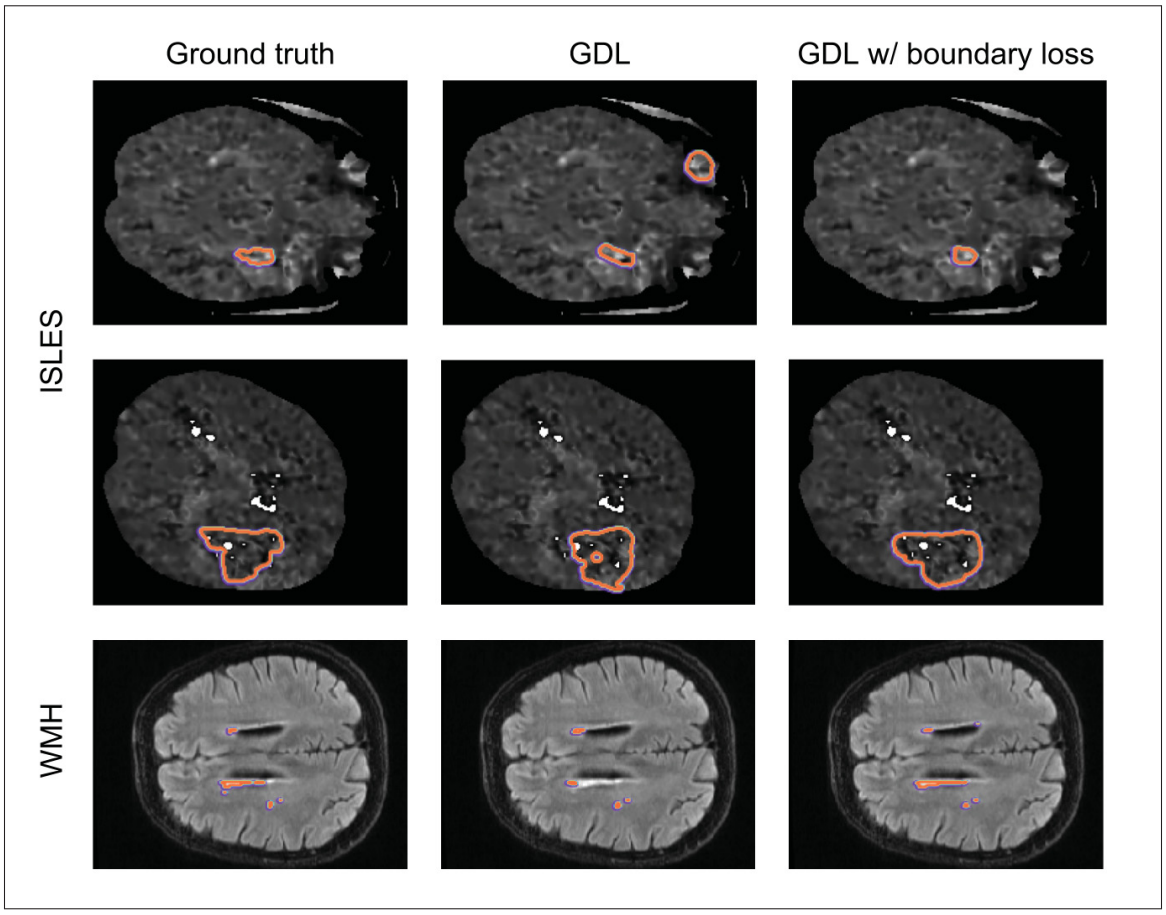


Figure 3.5 Visual results from WMH and ISLES validation sets..

Behaviour of boundary loss alone

As mentioned in 2.2.1, the boundary loss (used alone) is not able to give competitive results in a highly imbalanced problem. To investigate its behaviour on a less imbalance level, we trained the network only on positive samples (where the target is present). In that way, we decreased the imbalance rate of the whole dataset. Preliminary results, depicted in Fig. 3.6 show that, when trained only on positive samples, our boundary loss gives smooth contours. We might explain this behaviour with the following fact. In theory, distance maps used in our boundary loss give, on the contrary of typical ground truth images, non binary values but real ones. In

other words, pixels which are close to the foreground boundary are less penalized than far ones. This results in a more regular segmentation.

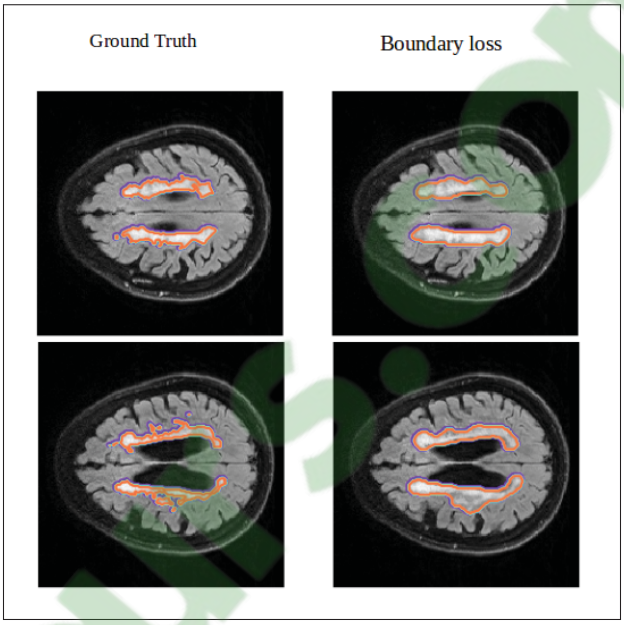


Figure 3.6 Visual results for our boundary loss as a stand alone loss trained only on positive samples on WMH dataset

CONCLUSION AND RECOMMENDATIONS

Summary of contributions

In this work, we tackled a common issue in medical image processing, the class imbalance problem. This problem occurs mostly when there is an unequal class distribution. In chapter 1, we briefly presented the main classic approaches used for medical image segmentation. We then reported some basics of deep CNNs and discussed both their successes and current challenges. In chapter 2, we introduced our methodology. We proposed a new boundary loss that takes the form of a distance metric on the space of contours, not regions. We argue that a distance-based loss can mitigate the issues related to regional losses in highly unbalanced segmentation problems. Rather than using unbalanced integrals over the regions, a boundary loss uses integrals over the boundary (interface) between the regions. Also, our loss provides a new information which is missing in regional losses.

Conducted on both binary and multi-class highly imbalanced problems, our experiments showed the great impact of including a distance based loss. Results demonstrated a good improvement not only on segmentation performance but also on the training stability.

Limitations and recommendations

Even though experiments have shown good results on three challenging medical problems, we argue that there are still some limitations and recommendations left for a future work. Thus, we suggest the following recommendations for the next steps in our research problem:

- In our main experiments, we used the boundary loss only as a joint loss. Motivated by preliminary results illustrated by Fig.3.6, it would be interesting to investigate the loss behaviour, when used alone, in other problems with different levels of imbalance and with careful initialisations.

- Although our experiments were limited to 2-D segmentation problems, the proposed framework can be trivially extended to 3-D, which could further improve the performance of deep networks, as more context is analyzed. We just need to compute distance transform functions in 3D and use it with any choice of 3-D framework.
- In order to compare to other state-of-the-art methods for these particular applications, future works will evaluate on testing sets of the challenges.

BIBLIOGRAPHY

IBM Unveils Watson-Powered Imaging Solutions at RSNA. Accessed: 2018-04-25.

Neural Networks. Accessed: 2016-07-29.

Multi-layer Perceptron Tutorial. Accessed: 2016-07-29.

Abraham, N. & Khan, N. M. (2018). A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation. *CoRR*, abs/1810.07842. Consulted at http://arxiv.org/abs/1810.07842.

Aja-Fernández, S., Curiale, A. H. & Vegas-Sánchez-Ferrero, G. (2015). A local fuzzy thresholding methodology for multiregion image segmentation. *Knowledge-Based Systems*, 83, 1 - 12. doi: https://doi.org/10.1016/j.knosys.2015.02.029.

A.Prasoon, K.Petersen, C.Igel, F.Lauze, E.Dam & M.Nielsen. (2013). Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pp. 246-253.

Artaechevarria, X., Munoz-Barrutia, A. & Ortiz-de Solorzano, C. (2009). Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. *Medical Imaging, IEEE Transactions on*, 28(8), 1266–1277. doi: 10.1109/tmi.2009.2014372.

Asperti, A. & Mastronardo, C. (2018). The Effectiveness of Data Augmentation for Detection of Gastrointestinal Diseases from Endoscopical Images. *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, pp. 199-205. doi: 10.5220/0006730901990205.

Avendi, M., Kheradvar, A. & Jafarkhani, H. (2016). A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Medical Image Analysis*, 30, 108-119.

Baur, C., Albarqouni, S. & Navab, N. (2017). Auxiliary Manifold Embedding for Fully Convolutional Networks. *CoRR*, abs/1703.06000. Consulted at http://arxiv.org/abs/1703.06000.

Bellver, M., Maninis, K., Pont-Tuset, J., Giró i Nieto, X., Torres, J. & Gool, L. V. (2017). Detection-aided liver lesion segmentation using deep learning. *CoRR*, abs/1711.11069.

Boykov, Y., Kolmogorov, V., Cremers, D. & Delong, A. (2006). An Integral Solution to Surface Evolution PDEs Via Geo-cuts. *Computer Vision – ECCV 2006*, pp. 409–422.

Brosch, T., Yoo, Y., Tang, L. Y. W., Li, D. K. B., Traboulsee, A. & Tam, R. (2015). Deep Convolutional Encoder Networks for Multiple Sclerosis Lesion Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 3–11.

Brébisson, A. & G.Montana. (2015). Deep Neural Networks for Anatomical Brain Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 20-28.

Chen, H., Shen, C., Qin, J., Ni, D., Shi, L., Cheng, J. C. Y. & Heng, P.-A. (2015). Automatic Localization and Identification of Vertebrae in Spine CT via a Joint Learning Model with Deep Neural Networks. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 515–522.

Chen, H., Dou, Q., Yu, L., Qin, J. & Heng, P.-A. (2018). VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, 170, 446 - 455. doi: https://doi.org/10.1016/j.neuroimage.2017.04.041. Segmenting the Brain.

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. (2014). Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *CoRR*, abs/1412.7062. Consulted at http://arxiv.org/abs/1412.7062.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. (2016a). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pp. 424–432.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. (2016b). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *MICCAI (2)*, 9901(Lecture Notes in Computer Science), 424–432.

Cireşan, D., Meier, U., Masci, J., Gambardella, L. M. & Schmidhuber, J. (2011). Flexible, high performance convolutional neural networks for image classification. *International Joint Conference on Artificial Intelligence(IJCAI)*, pp. 1237-1242.

Crum, W. R., Camara, O. & Hill, D. L. G. (2006). Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis. *IEEE Transactions on Medical Imaging*, 25(11), 1451-1461. doi: 10.1109/TMI.2006.880587.

Dai, J., He, K. & Sun, J. (2015). BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation. *CoRR*, abs/1503.01640. Consulted at http://arxiv.org/abs/1503.01640.

D.E.Rumelhart, G.E.Hinton & R.J.Wiliams. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536.

Dolz, J., Desrosiers, C. & Ayed, I. B. (2016a). 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study.

Dolz, J., Desrosiers, C. & Ayed, I. B. (2016b). 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *CoRR*, abs/1612.03925.

Dolz, J., Ayed, I. B., Yuan, J. & Desrosiers, C. (2017). HyperDense-Net: A hyper-densely connected CNN for multi-modal image semantic segmentation. *CoRR*, abs/1710.05956. Consulted at http://arxiv.org/abs/1710.05956.

Droske, M., Meyer, B., Rumpf, M. & Schaller, C. (2001). An Adaptive Level Set Method for Medical Image Segmentation. *Information Processing in Medical Imaging*, pp. 416–422.

F.Rosenblatt. (1958). THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN. *Psychological review*, 65(6), 386-408.

Ganaye, P., Sdika, M. & Benoit-Cattin, H. (2018, April). Towards integrating spatial localization in convolutional neural networks for brain image segmentation. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 621-625. doi: 10.1109/ISBI.2018.8363652.

Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., Guttmann, C. R. G., de Leeuw, F., Tempany, C. M., van Ginneken, B., Fedorov, A., Abolmaesumi, P., Platel, B. & III, W. M. W. (2017). Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation. *CoRR*, abs/1702.07841. Consulted at http://arxiv.org/abs/1702.07841.

Glorot, X. & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Artificial Intelligence and Statistics conference*, pp. 249-256.

Górriz, M., Aparicio, A., Raventós, B., Vilaplana, V., Sayrol, E. & López-Codina, D. (2018). Leishmaniasis Parasite Segmentation and Classification Using Deep Learning. *Articulated Motion and Deformable Objects*, pp. 53–62.

Habas, P., Kim, K., Rousseau, F., Glenn, O., Barkovich, A. & Studholme, C. (2009). A Spatio-temporal Atlas of the Human Fetal Brain with Application to Tissue Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009* (ch. 36, pp. 289–296). doi: 10.1007/978-3-642-04268-3_36.

Han, X., Hoogeman, M., Levendag, P., Hibbard, L., Teguh, D., Voet, P., Cowen, A. & Wolf, T. (2008). Atlas-Based Auto-segmentation of Head and Neck CT Images Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008. In Metaxas, D., Axel, L., Fichtinger, G. & Székely, G. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008* (vol. 5242, ch. 52, pp. 434–441). Berlin, Heidelberg: Springer Berlin / Heidelberg. doi: 10.1007/978-3-540-85990-1_52.

Hashemi, S. R., Mohseni Salehi, S. S., Erdogmus, D., Prabhu, S. P., Warfield, S. K. & Gholipour, A. (2019a). Asymmetric Loss Functions and Deep Densely-Connected Networks for Highly-Imbalanced Medical Image Segmentation: Application to Multiple Sclerosis Lesion Detection. *IEEE Access*, 7, 1721-1735. doi: 10.1109/ACCESS.2018.2886371.

Hashemi, S. R., Mohseni Salehi, S. S., Erdogmus, D., Prabhu, S. P., Warfield, S. K. & Gholipour, A. (2019b). Asymmetric Loss Functions and Deep Densely-Connected Networks for Highly-Imbalanced Medical Image Segmentation: Application to Multiple Sclerosis Lesion Detection. *IEEE Access*, 7, 1721-1735. doi: 10.1109/ACCESS.2018.2886371.

Hashemi, S. R., Salehi, S. S., Erdogmus, D., Prabhu, S., Warfield, S. & Gholipour, A. (2018). Asymmetric Loss Functions and Deep Densely Connected Networks for Highly Imbalanced Medical Image Segmentation: Application to Multiple Sclerosis Lesion Detection. *IEEE Access*, 7, 1721-1735. doi: 10.1109/ACCESS.2018.2886371.

Hassanien, A. E., Moftah, H. M., Azar, A. T. & Shoman, M. (2014). MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier. *Applied Soft Computing*, 14, 62 - 71. doi: https://doi.org/10.1016/j.asoc.2013.08.011. Special issue on hybrid intelligent methods for health technologies.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M. & Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*, 35, 18 - 31. doi: https://doi.org/10.1016/j.media.2016.05.004.

He, K., Zhang, X., Ren, S. & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, (ICCV '15), 1026–1034. doi: 10.1109/ICCV.2015.123.

He, L., Peng, Z., Everding, B., Wang, X., Han, C. Y., Weiss, K. L. & Wee, W. G. (2008). A comparative study of deformable contour methods on medical image segmentation. *Image and Vision Computing*, 26(2), 141 - 163. doi: https://doi.org/10.1016/j.imavis.2007.07.010.

H.Kervadec, J.Dolz, M.Tang, E.Granger, Y.Boykov & Ayed, I. (2018). Size-constraint loss for weakly supervised CNN segmentation. *International conference on Medical Imaging with Deep Learning*.

Horváth, A., Tsagkas, C., Andermatt, S., Pezold, S., Parmar, K. & Cattin, P. C. (2018). Spinal Cord Gray Matter-White Matter Segmentation on Magnetic Resonance AMIRA Images with MD-GRU. *CoRR*, abs/1808.02408. Consulted at http://arxiv.org/abs/1808.02408.

Huang, G., Liu, Z. & Weinberger, K. Q. (2016). Densely Connected Convolutional Networks. *CoRR*, abs/1608.06993. Consulted at http://arxiv.org/abs/1608.06993.

Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. A. (1993). Comparing Images Using the Hausdorff Distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(9), 850–863. doi: 10.1109/34.232073.

Hwang, S. & Kim, H.-E. (2016). Self-Transfer Learning for Weakly Supervised Lesion Localization. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 239–246.

Jia, Z., Huang, X., Chang, E. I. & Xu, Y. (2017). Constrained Deep Weak Supervision for Histopathology Image Segmentation. *CoRR*, abs/1701.00794. Consulted at http://arxiv.org/abs/1701.00794.

J.Long, E.Shelhamer & D.Darrell. (2015). Fully convolutional networks for semantic segmentation. *CVPR*, pp. 3431-3440.

Kalinovsky, A. & Kovalev, V. (2016, 10). Lung Image Segmentation Using Deep Learning Methods and Convolutional Neural Networks.

Kallenberg, M., Petersen, K., Nielsen, M., Ng, A. Y., Diao, P., Igel, C., Vachon, C. M., Holland, K., Winkel, R. R., Karssemeijer, N. & Lillholm, M. (2016). Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring. *IEEE Transactions on Medical Imaging*, 35(5), 1322-1331. doi: 10.1109/TMI.2016.2532122.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D. & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36, 61 - 78. doi: https://doi.org/10.1016/j.media.2016.10.004.

Kervadec, H., Dolz, J., Tang, M., Granger, E., Boykov, Y. & Ayed, I. B. (2019). Constrained-CNN Losses for Weakly Supervised Segmentation. *Medical Image Analysis*. doi: https://doi.org/10.1016/j.media.2019.02.009.

K.Fukushima. (1980). Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, 36, 193-202.

Kingma, D. P. & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *CoRR*, abs/1412.6980. Consulted at http://arxiv.org/abs/1412.6980.

Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M. & Biller, A. (2016). Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. *NeuroImage*, 129, 460 - 469. doi: https://doi.org/10.1016/j.neuroimage.2016.01.024.

Klein, A., Warszawski, J., Hillengass, J. & Maier-Hein, K. (2018). Automatic bone segmentation in whole-body CT images. *International Journal of Computer Assisted Radiology and Surgery*, 14. doi: 10.1007/s11548-018-1883-7.

Krähenbühl, P. & Koltun, V. (2011). Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. & Weinberger, K. Q. (Eds.), *Advances in Neural Information Processing Systems*

24 (pp. 109–117). Curran Associates, Inc. Consulted at http://papers.nips.cc/paper/4296-efficient-inference-in-fully-connected-crfs-with-gaussian-edge-potentials.pdf.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q. (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 1097–1105). Curran Associates, Inc. Consulted at http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Lai, M. (2015). Deep learning for medical image segmentation. *arXiv: 1505.02000*.

LC.Chen & G.Papandreou. (2015). Semantic image segmentation with deep convolutional nets and fully connected CRFs. *CVPR*, pp. 3431-3440.

LeCun, Y., Kanter, I. & Solla, S. (1991). Second order properties of error surfaces: Learning time and generalization. *Advances in Neural Information Processing Systems (NIPS)*, 3, 918–924.

Li, B. N., Chui, C. K., Chang, S. & Ong, S. (2011). Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation. *Computers in Biology and Medicine*, 41(1), 1 - 10. doi: https://doi.org/10.1016/j.compbiomed.2010.10.007.

Lin, D., Dai, J., Jia, J., He, K. & Sun, J. (2016). ScribbleSup: Scribble-Supervised Convolutional Networks for Semantic Segmentation. *CoRR*, abs/1604.05144.

Lin, N., Yu, W. & Duncan, J. S. (2003). Combinative multi-scale level set framework for echocardiographic image segmentation. *Medical Image Analysis*, 7(4), 529 - 537. doi: https://doi.org/10.1016/S1361-8415(03)00035-5. Medical Image Computing and Computer Assisted Intervention.

Litjens, G., I. Sánchez, C., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen van de Kaa, C., Bult, P., van Ginneken, B. & van der Laak, J. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports*, 6, 26286. doi: 10.1038/srep26286.

Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E. & Joshi, S. (2006). Multi-modal image set registration and atlas formation. *Medical image analysis*, 10(3), 440–451. doi: 10.1016/j.media.2005.03.002.

Manikandan, S., Ramar, K., Iruthayarajan, M. W. & Srinivasagan, K. (2014). Multilevel thresholding for segmentation of medical brain images using real coded genetic algorithm. *Measurement*, 47, 558 - 568. doi: https://doi.org/10.1016/j.measurement.2013.09.031.

Martí, J., Freixenet, J., Muñoz, X. & Oliver, A. (2003). Active Region Segmentation of Mammographic Masses Based on Texture, Contour and Shape Features.

McInerney, T. & Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2), 91 - 108. doi: https://doi.org/10.1016/S1361-8415(96)80007-7.

Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.-A., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, L., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Antonio Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S. J., Riklin-Raviv, T., Reza, S., Ryan, M., Schwartz, L., Shin, H.-C., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Hye Ye, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M. & Van Leemput, K. (2014). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993-2024. doi: 10.1109/TMI.2014.2377694.

M.Havaei, A.Davy, D.Warde-Farley, A.Biard, A.Courville, Y.Bengio, C.Pal, PM.Jodoin & H.Larochelle. (2015). Brain Tumor Segmentation with Deep Neural Networks. *arXiv preprint arXiv:150503540*.

Mi, H., Petitjean, C., Dubray, B., Vera, P. & Ruan, S. (2014, April). Automatic lung tumor segmentation on PET images based on random walks and tumor growth model. *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 1385-1388. doi: 10.1109/ISBI.2014.6868136.

Milletari, F., Navab, N. & Ahmadi, S. (2016, Oct). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565-571. doi: 10.1109/3DV.2016.79.

Milletari, F., Navab, N. & Ahmadi, S. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *CoRR*, abs/1606.04797. Consulted at http://arxiv.org/abs/1606.04797.

Minnema, J., van Eijnatten, M., Kouw, W., Diblen, F., Mendrik, A. & Wolff, J. (2018). CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. *Computers in Biology and Medicine*, 103, 130 - 139. doi: https://doi.org/10.1016/j.compbiomed.2018.10.012.

Minsky, M. & Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. The MIT Press,Cambridge MA.

Mitiche, A. & Ben Ayed, I. (2011). *Variational and Level Set Methods in Image Segmentation*. doi: 10.1007/978-3-642-15352-5.

M.shakeri, S.Tsogkas, E.Ferrante, S.Lippe & S.Kadoury andN.Paragios, I. (2016). Subcortical brain structure segmentation using F-CNN's. *International Symposium on Biomedical Imaging*.

N, S. & S, V. (2016). Image Segmentation By Using Thresholding Techniques For Medical Images. *Computer Science Engineering: An International Journal*, 6, 1-13. doi: 10.5121/cseij.2016.6101.

naceur, M. B., Saouli, R., Akil, M. & Kachouri, R. (2018). Fully Automatic Brain Tumor Segmentation using End-To-End Incremental Deep Neural Networks in MRI images. *Computer Methods and Programs in Biomedicine*, 166, 39 - 49. doi: https://doi.org/10.1016/j.cmpb.2018.09.007.

Nie, D., Gao, Y., Wang, L. & Shen, D. (2018). ASDNet: Attention Based Semi-supervised Deep Networks for Medical Image Segmentation. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pp. 370–378.

N.Srivastava, G.Hinton, A.Krizhevsky, I.Sutskever & R.Salakhutdinov. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research (JMLR)*, 15, 1929-1958.

Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M. P., Bai, W., Caballero, J., Guerrero, R., Cook, S. A., de Marvao, A., Dawes, T., O'Regan, D. P., Kainz, B., Glocker, B. & Rueckert, D. (2017). Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation. *CoRR*, abs/1705.08302.

Oliver, A., Freixenet, J., Martí, J., Pérez, E., Pont, J., Denton, E. R. & Zwiggelaar, R. (2010). A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis*, 14(2), 87 - 110. doi: https://doi.org/10.1016/j.media.2009.12.005.

Papadimitriou, C. H. & Steiglitz, K. (1998). The Max-Flow, Min-Cut Theorem. *Combinatorial Optimization: Algorithms and Complexity*, pp. 120-128.

Papandreou, G., Chen, L., Murphy, K. & Yuille, A. L. (2015). Weakly- and Semi-Supervised Learning of a DCNN for Semantic Image Segmentation. *CoRR*, abs/1502.02734. Consulted at http://arxiv.org/abs/1502.02734.

Paszke, A., Chaurasia, A., Kim, S. & Culurciello, E. (2016). ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. *CoRR*, abs/1606.02147.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. & Lerer, A. (2017). Automatic differentiation in pytorch.

Pathak, D., Krähenbühl, P. & Darrell, T. (2015). Constrained Convolutional Neural Networks for Weakly Supervised Segmentation. *CoRR*, abs/1506.03648.

Pereira, S., Pinto, A., Alves, V., & Silva, C. (2016). Brain Tumor Segmentation using Convolutional Neural Networks in MRI Images. *IEEE Transactions on Medical Imaging*, 35, 1240–1251.

Rajchl, M., Lee, M. C. H., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Kainz, B. & Rueckert, D. (2016). DeepCut: Object Segmentation from Bounding Box Annotations using Convolutional Neural Networks. *CoRR*, abs/1605.07866.

Rezaei, M., Yang, H. & Meinel, C. (2018). Conditional Generative Refinement Adversarial Networks for Unbalanced Medical Image Semantic Segmentation. *CoRR*, abs/1810.03871. Consulted at http://arxiv.org/abs/1810.03871.

Ronneberger, O., Fischer, P. & Brox, T. (2015a). U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR*, abs/1505.04597.

Ronneberger, O., Fischer, P. & Brox, T. (2015b). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241.

Salehi, S. S. M., Erdogmus, D. & Gholipour, A. (2017). Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. *Machine Learning in Medical Imaging*, pp. 379–387.

Shah, M. P., Merchant, S. N. & Awate, S. P. (2018). MS-Net: Mixed-Supervision Fully-Convolutional Networks for Full-Resolution Segmentation. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pp. 379–387.

Shen, C., Roth, H. R., Oda, H., Oda, M., Hayashi, Y., Misawa, K. & Mori, K. (2018). On the influence of Dice loss function in multi-class organ segmentation of abdominal CT using 3D fully convolutional networks. *CoRR*, abs/1801.05912. Consulted at http://arxiv.org/abs/1801.05912.

Shi, J. & Malik, J. (2000). Normalized Cuts and Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8), 888–905. doi: 10.1109/34.868688.

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures,Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*.

Simonyan, K. & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556. Consulted at http://arxiv.org/abs/1409.1556.

Skourt, B. A., Hassani, A. E. & Majda, A. (2018). Lung CT Image Segmentation Using Deep Neural Networks. *Procedia Computer Science*, 127, 109 - 113. doi: https://doi.org/10.1016/j.procs.2018.01.104. PROCEEDINGS OF THE FIRST INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING IN DATA SCIENCES, ICDS2017.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. & Cardoso, M. J. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. *CoRR*, abs/1707.03237.

Sujji, G. E., Lakshmi, Y. V. S. & Jiji, G. W. MRI Brain Image Segmentation based on Thresholding.

Szegedy, C., W.Liu, Jia, Y., P.Sermanet, Reed, S., D.Anguelov, Erhan, D., Vanhoucke, V. & A.Rabinovich. (2014). Going Deeper with Convolutions. *CoRR*, abs/1409.4842. Consulted at http://arxiv.org/abs/1409.4842.

Tajbakhsh, N., Hu, Y., Cao, J., Yan, X., Xiao, Y., Lu, Y., Liang, J., Terzopoulos, D. & Ding, X. (2019). Surrogate Supervision for Medical Image Analysis: Effective Deep Learning From Limited Quantities of Labeled Data. *arXiv e-prints*.

Tajbakhsh, N., Jae Y. Shin Suryakanth R. Gurudu,R. Todd Hurst, C. B. K. M. B. G. & Liang, J. (2017). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans Med Imaging.*, 35, 1299 - 1312.

Tang, M., Perazzi, F., Djelouah, A., Ben Ayed, I., Schroers, C. & Boykov, Y. (2018a). On Regularized Losses for Weakly-supervised CNN Segmentation. *ArXiv e-prints*.

Tang, M., Zhang, Z., Cobzas, D., Jagersand, M. & Jaremko, J. L. (2018b, April). Segmentation-by-detection: A cascade network for volumetric medical image segmentation. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1356-1359. doi: 10.1109/ISBI.2018.8363823.

Tang, M., Ayed, I. B., Marin, D. & Boykov, Y. (2015). Secrets of GrabCut and Kernel K-means. *CoRR*, abs/1506.07439.

Tran, P. V. (2016). A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI. *CoRR*, abs/1604.00494. Consulted at http://arxiv.org/abs/1604.00494.

Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, , Oliver, A. & Llado, X. (2017). Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *NeuroImage*, 155. doi: 10.1016/j.neuroimage.2017.04.034.

Vasconcelos, C. N. & Vasconcelos, B. N. (2017). Increasing Deep Learning Melanoma Classification by Classical And Expert Knowledge Based Image Transforms. *CoRR*, abs/1702.07025. Consulted at http://arxiv.org/abs/1702.07025.

Wachinger, C., Reuter, M. & Klein, T. (2018). DeepNAT: Deep convolutional neural network for segmenting neuroanatomy. *NeuroImage*, 170, 434 - 445. doi: https://doi.org/10.1016/j.neuroimage.2017.02.035. Segmenting the Brain.

Wang, D., Khosla, A., Gargeya, R., Irshad, H. & Beck, A. H. (2016). Deep Learning for Identifying Metastatic Breast Cancer. *arXiv preprint arXiv:1606.05718*.

Wang, G., Li, W., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., Doel, T., David, A. L., Deprest, J., Ourselin, S. & Vercauteren, T. (2017a). Interactive Medical Image Segmentation using Deep Learning with Image-specific Fine-tuning. *CoRR*, abs/1710.04043. Consulted at http://arxiv.org/abs/1710.04043.

Wang, G., Zuluaga, M. A., Li, W., Pratt, R., Patel, P. A., Aertsen, M., Doel, T., David, A. L., Deprest, J., Ourselin, S. & Vercauteren, T. (2017b). DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. *CoRR*, abs/1707.00652. Consulted at http://arxiv.org/abs/1707.00652.

Wang, L., Li, C., Sun, Q., Xia, D. & Kao, C.-Y. (2009). Active contours driven by local and global intensity fitting energy with application to brain MR image segmentation. *Computerized Medical Imaging and Graphics*, 33(7), 520 - 531. doi: https://doi.org/10.1016/j.compmedimag.2009.04.010.

Wong, K. C. L., Moradi, M., Tang, H. & Syeda-Mahmood, T. (2018). 3D Segmentation with Exponential Logarithmic Loss for Highly Unbalanced Object Sizes. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pp. 612–619.

Wu, K., Gauthier, D. & Levine, M. D. (1995). Live cell image segmentation. *IEEE Transactions on Biomedical Engineering*, 42(1), 1-12. doi: 10.1109/10.362924.

Xie, F. & Bovik, A. C. (2013). Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm. *Pattern Recognition*, 46(3), 1012 - 1019. doi: https://doi.org/10.1016/j.patcog.2012.08.012.

Xu, B., Wang, N., Chen, T. & Li, M. (2015). Empirical Evaluation of Rectified Activations in Convolutional Network. *CoRR*, abs/1505.00853. Consulted at http://arxiv.org/abs/1505.00853.

Y.Boykov & Jolly, M.-P. (2001). Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 86, 105-112. doi: 10.1109/ICCV.2001.937505.

Y.LeCun, L.Bottou, Y.Bengio, & Haffner, B. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86, 2278 - 2324.

Yu, L., Guo, Y., Wang, Y., Yu, J. & Chen, P. (2017a). Segmentation of Fetal Left Ventricle in Echocardiographic Sequences Based on Dynamic Convolutional Neural Networks. *IEEE Transactions on Biomedical Engineering*, 64(8), 1886-1895. doi: 10.1109/TBME.2016.2628401.

Yu, L., Yang, X., Chen, H., Qin, J. & Heng, P.-A. (2017b). Volumetric ConvNets with Mixed Residual Connections for Automated Prostate Segmentation from 3D MR Images. *AAAI*.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C. & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31(3), 1116 - 1128. doi: https://doi.org/10.1016/j.neuroimage.2006.01.015.

Zhang, J., Yan, C.-H., Chui, C.-K. & Ong, S.-H. (2010). Fast segmentation of bone in CT images using 3D adaptive thresholding. *Computers in Biology and Medicine*, 40(2), 231 - 236. doi: https://doi.org/10.1016/j.compbiomed.2009.11.020.

Zhang, Y., Matuszewski, B. J., Shark, L. & Moore, C. J. (2008, July). Medical Image Segmentation Using New Hybrid Level-Set Method. *2008 Fifth International Conference BioMedical Visualization: Information Visualization in Medical and Biomedical Informatics*, pp. 71-76. doi: 10.1109/MediVis.2008.12.