SEGMENTATION AND CLASSIFICATION OF MULTIMODAL MEDIAL IMAGES BASED ON GENERATIVE ADVERSARIAL LEARNING AND CONVOLUTIONAL NEURAL NETWORKS

Vivek Kumar Singh

UNIVERSITAT ROVIRA I VIRGILI
SEGMENTATION AND CLASSIFICATION OF MULTIMODAL MEDIAL IMAGES BASED ON GENERATIVE ADVERSARIAL LEARNING AND CONVOLUTIONAL NEURAL NETWORKS
Vivek Kumar Singh

UNIVERSITAT ROVIRA I VIRGILI
SEGMENTATION AND CLASSIFICATION OF MULTIMODAL MEDIAL IMAGES BASED ON GENERATIVE ADVERSARIAL LEARNING AND CONVOLUTIONAL NEURAL NETWORKS
Vivek Kumar Singh

Segmentation and Classification of Multimodal Medical Images based on Generative Adversarial Learning and Convolutional Neural Networks

DOCTORAL THESIS

Author:
Vivek Kumar Singh

Advisors:
Dr. Santiago Romaní Also
Dr. Domènec Savi Puig Valls

Departament d'Enginyeria Informàtica i Matemàtiques

UNIVERSITAT ROVIRA I VIRGILI
Tarragona
2019

UNIVERSITAT ROVIRA I VIRGILI
SEGMENTATION AND CLASSIFICATION OF MULTIMODAL MEDIAL IMAGES BASED ON GENERATIVE ADVERSARIAL LEARNING AND CONVOLUTIONAL NEURAL NETWORKS
Vivek Kumar Singh

UNIVERSITAT ROVIRA I VIRGILI

Departament d'Enginyeria Informàtica i Matemàtiques

Av. Paisos Catalans, 27
43007 Tarragona
Tel. +34 977 55 95 95
Fax. +34 977 55 95 97

We STATE that the present study, entitled "Segmentation and Classification of Multimodal Medical Images based on Generative Adversarial Learning and Convolutional Neural Networks", presented by Vivek Kumar Singh, for the award of the degree of Doctor, has been carried out under our supervision at the Departament d'Enginyeria Informàtica i Matemàtiques.

Tarragona, 10th September 2019.

Doctoral Thesis Supervisors,

Dr. Santiago Romaní Also

Dr. Domènec Savi Puig Valls

UNIVERSITAT ROVIRA I VIRGILI
SEGMENTATION AND CLASSIFICATION OF MULTIMODAL MEDIAL IMAGES BASED ON GENERATIVE ADVERSARIAL LEARNING AND CONVOLUTIONAL NEURAL NETWORKS
Vivek Kumar Singh

To my wife Nidhi, my brother, my sister and my parents

UNIVERSITAT ROVIRA I VIRGILI
SEGMENTATION AND CLASSIFICATION OF MULTIMODAL MEDIAL IMAGES BASED ON GENERATIVE ADVERSARIAL LEARNING AND CONVOLUTIONAL NEURAL NETWORKS
Vivek Kumar Singh

Abstract

Medical imaging is an important means for early illness detection in the majority of medical fields, which provides better prognosis to the patients. But properly interpreting medical images needs highly trained medical experts: it is difficult, time-consuming, expensive, and error-prone. It would be more beneficial to have a computer-aided diagnosis (CAD) system that can automatically outline the possible ill tissues and suggest diagnosis to the doctor. Current development in deep learning methods motivates us to improve current medical image analysis systems.

In this thesis, we have considered three different medical diagnosis, such as breast cancer from mammograms and ultrasound images, skin lesion from dermoscopic images, and retinal diseases from fundus images. These tasks are very challenging due to the several sources of variability in the image capturing processes.

Firstly, we propose a method to analyze the breast cancer in mammograms. In a first stage, we utilize the Single Shot Detector (SSD) method to locate the possibly abnormal regions, which are called regions of interest (ROIs). Then, in a second stage we apply a conditional generative adversarial network (cGAN) method to segment possible masses within the ROIs. This network works efficiently with a reduced number of training images. In a third stage, a convolutional neural network (CNN) has been introduced to classify the shape of the masses (round, oval, lobular and irregular). Besides, we also try to classify those masses into four distinct breast cancer molecular subtypes (Luminal-A, Luminal-B, Her-2, and Basal-like), based on its shape and also on the micro-texture rendered in the image pixels. Moreover, for ultrasound image processing, we extended the proposed cGAN model by introducing a novel channel attention and weighting (CAW) block, which improves the robustness of segmentation by fostering the more relevant features of the masses. Some statistical analysis corroborate the accuracy of the segmented masks. Finally, we also performed a classification between benign and malignant tumors based on the shape of the segmented masks.

Second, skin lesion segmentation in dermoscopic images is still challenging due to the low contrast and fuzzy boundaries of lesions. Besides, lesions have high similarity

to healthy regions. To overcome this problems, we introduce a novel layer inside the encoder of the cGAN, called factorized channel attention (FCA) block. It integrates a channel attention mechanism and a residual 1-D kernel factorized convolution. The channel attention mechanism increases the discriminability between the lesion and non-lesion features by taking into account feature channel interdependencies. The 1-D factorized kernels provide extra convolutional layers with a minimal set of parameters and a residual connection that minimizes the impact of image artifacts and irrelevant objects.

Third, segmentation of retinal optic disc in fundus photographs plays a critical role in the diagnosis, screening and treatment of many ophthalmologic diseases. Therefore, we have applied our cGAN method to the task of optic disc segmentation, obtaining promising results with a really short number of training samples (less than twenty).

Experiments with these three kinds of medical image diagnosis have been performed for quantitative and qualitative comparisons with other state-of-the-art methods, to show the advantages of the proposed detection, segmentation and classification techniques.

Keywords: Medical image analysis, Deep learning, Breast cancer, Skin lesion, Retinal fundus image, Convolutional neural network, Conditional generative adversarial network, Detection, Segmentation, Classification.

Acknowledgements

The author was supported by a PhD grant from URV in 2016 (Mart Franqus program). This work was supported by the Spanish Government project: DPI2016-77415-R.

I would like to express my gratitude to my supervisors Dr. Santiago Romaní and Dr. Domenèc Puig for their useful guidance, insightful comments and considerable encouragement to complete this thesis. They have guided me to pursue important problems that will have practical impact and was always available to guide me whenever I approached him. Their extensive knowledge, experience, and exceptional ability to find new approaches for difficult problems were pivotal in this work and my development as a researcher. This work would not have been completed without their encouragement and patience.

My sincere thanks also goes to Dr. Alain Lalande and Dr. Benoit Presles who provided me an opportunity to join their lab (Dijon, France) as an intern, and who gave access to the laboratory and research facilities. Without their precious support it would not be possible to conduct this research.

In addition, because of the research environment sustained by Dr. Puig, I have crossed paths with many graduate students and postdocs who have influenced and enhanced my research. The direction of my research work has been strongly influenced by the members of the IRCV group, notably Dr. Hatem A. Rashwan, Dr. Farhan Akram and Dr. Mohammed Abdel Nasser has led me to discuss ideas, and revise the manuscripts. Besides, I appreciate Dr. Adel Saleh who motivated me to work on deep learning. Also, I thank my fellow labmates in for the stimulating discussions for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last three years of my doctorate research.

Finally, I would have never managed to accomplish this task without God blessings and the loving support of my family. I cordially thank to my wife, parents, brother, sister and uncles. Lastly, I would like to thank to my wife's parents and her uncle who supported and motivated me to think this research from medical perspective.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

This first chapter highlights the details about medical image analysis along with deep learning and explains the major objectives of this thesis. Also, it presents some of the leading publications concluded from the thesis, as an index of its scientific quality, as well as the thesis organization.

1.1 Medical image analysis

Medical image analysis is the ability to examine medical issues based on various imaging modalities. It consists of obtaining images from the human body to help doctors in making an accurate diagnosis of possible illnesses. The most common types of medical images include X-rays, MRI (Magnetic Resonance Image), CT (Computerized Tomography), PET (Positron Emission Tomography), and ultrasounds, which allow obtaining a visualization of the inside of body parts without

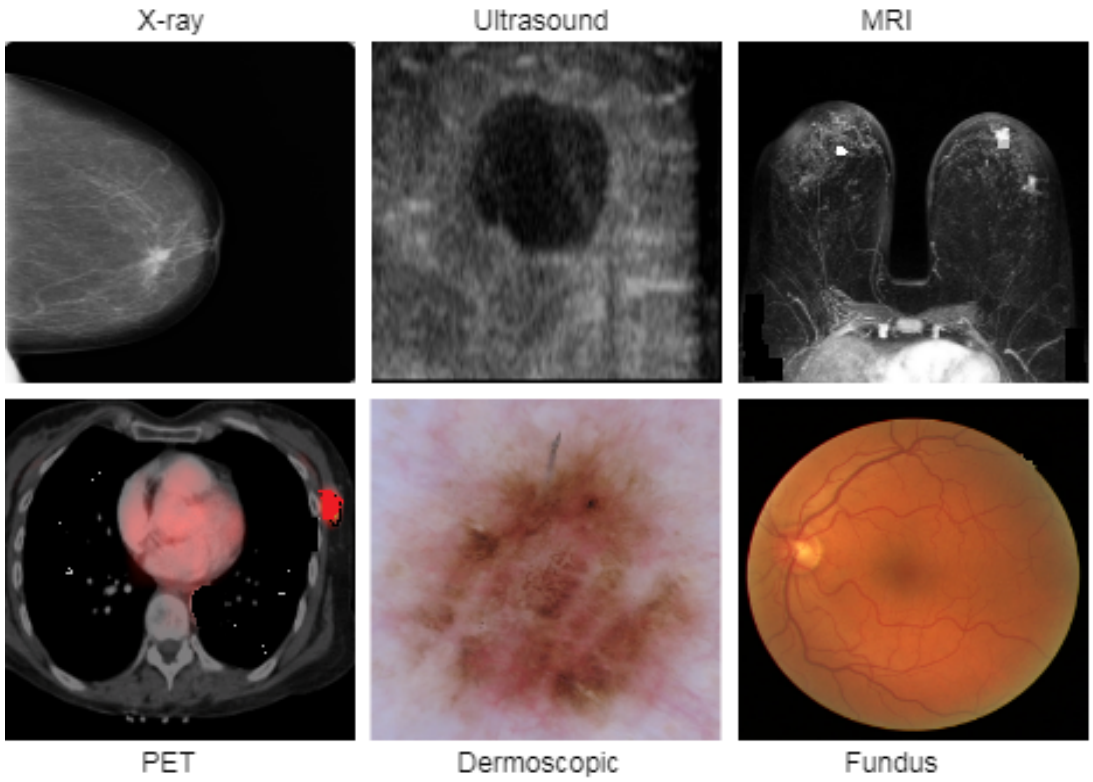


Figure 1.1: Examples of radiological and camera based images.

surgery. Moreover, some medical images use visible light cameras to capture images from outside of the body, like skin images (dermoscopy), eye fundus, endoscopic images and histological images for analyzing the shape of cells. Figure 1.1 shows some examples of these types of medical images.

Radiologists are doctors specialized in analyzing these kinds of medical images. They need a lot of practice to achieve high experience of being capable of providing accurate diagnosis. The current course of action of examining medical images is labor-intensive, time-consuming, expensive, and error-prone. It would be more useful to apply Computer Vision (CV) algorithms capable of extracting high-level information from numerical images. Those algorithms will lead to automatic or semiautomatic systems that can help radiologists by providing clues for more reliable diagnosis and treatment recommendations. This is known as Computer-Aided Diagnosis (CAD) systems.

1.2 Deep learning

Machine learning comprises a collection of algorithms that enables a computer to determine significant patterns from data without human intervention. With rapid advancing of computational power and the availability of large amounts of data, deep learning LeCun et al. (2015) has become the default machine-learning method that is used because it can determine significantly higher complex patterns than traditional machine-learning methods. Deep learning has been a great and strong device to foster Artificial Intelligence (AI) in the recent few years. It has performed remarkable or yet superior human-level performance on image classification He et al. (2016a), speech identification Xiong et al. (2018), and reading knowledge Devlin et al. (2018).

Therefore, this is particularly essential for the field of medical imaging analysis. Recent success in deep learning enables us to rethink clinical diagnostic methods based on medical images Litjens et al. (2017), Maier et al. (2019), Hesamian et al. (2019). In all deep learning approaches, CNNs are of exceptional concern. By utilizing confined connectivity patterns, such as those employed in the ImageNet competition Krizhevsky et al. (2012), CNNs have fast enhanced the state-of-the-art approach for image processing. It is crucial for the current CAD systems to provide accurate and efficient diagnosis and to deal with various types of medical data. Medical image analysis tasks include detection, segmentation and recognition of organs or lesions from images pixels provided in mammograms, ultrasound, CT or MRI images. These are very challenging assignments for traditional Computer Vision algorithms, but they can be efficiently tackled with deep learning methods.

1.3 Motivation

Our main motivation for the thesis is to create an advanced CAD system for any type of medical image modality with high sensitivity and specificity rates based on deep learning techniques. More specifically, we want to improve the automatic method of detection of Regions of Interest (ROI), which are areas of the image that

contain possible ill tissues, as well as segmentation of the findings (delimitation with a boundary), and ultimately, a prediction of a most suitable diagnose (classification). In this thesis, we focus on several topics including mammograms and ultrasound images to diagnose breast cancer, skin lesions analysis in dermoscopic images and retinal fundus images examination to avoid diabetic retinopathy.

1.4 Thesis objectives

The main objectives of this thesis are:

- To classify the breast cancer molecular subtypes only from mammogram images.
- To develop a CAD system for breast cancer diagnosis, able to detect, segment and classify mass regions in mammograms. Moreover, to possibly predict the molecular subtypes of masses based on segmented shape features.
- To segment and classify breast lesions also in breast ultrasound images. The diagnosis of this system can complement the output of the previous system.
- To segment the optic disc from retinal fundus image to address the problem of diabetic retinopathy.
- To provide a fully automatic skin lesion boundary segmentation from dermoscopic images.

1.5 Scientific dissemination

The list below shows the main generated publications depending on the type of dissemination.

1.5.1 Journal articles

1. Vivek Kumar Singh, Hatem A. Rashwan, Santiago Romani, Farhan Akram, Nidhi Pandey, Md Mostafa Kamal Sarker, Adel Saleh, Meritxell Arenas, Miguel Arquez, Domenec puig, Jordina Torrents Barrena, ***"Breast Tumor***

Segmentation and Shape Classification in Mammograms using Generative Adversarial and Convolutional Neural Network", Experts Systems with Applications. Impact Factor: 4.29 **(Q1)**. (Singh et al. (2020))

2. Vivek Kumar Singh, Mohamed Abdel-Nasser, Hatem A. Rashwan, Farhan Akram, Nidhi Pandey, Santiago Romani, Domenec Puig, ***"An Efficient Solution for Breast Tumor Segmentation and Classification in Ultrasound Images using Deep Adversarial Learning"***, IEEE Transaction on Biomedical Engineering. Impact Factor: 4.28 **(Q1)**. (To be submitted).
3. Vivek Kumar Singh, Mohamed Abdel-Nasser, Hatem A. Rashwan, Farhan Akram, Nidhi Pandey, Alain Lalande, Benoit Presles, Santiago Romani, Domenec Puig, ***"FCA-Net: Adversarial Learning for Skin Lesion Segmentation based on Multi-scale Features and Factorized Channel Attention"***, IEEE Access Journal. Impact Factor: 4.09 **(Q1).** (Accepted).

1.5.2 Conference proceedings

1. Vivek Kumar Singh, Santiago Romani, Hatem A Rashwan, Farhan Akram, Nidhi Pandey, Md. Mostafa Kamal Sarker, Saddam Abdulwahab, Jordina Torrents-Barrena, Adel Saleh, Miguel Arquez, Meritxell Arenas, Domenec Puig, ***"Conditional Generative Adversarial and Convolutional Networks for X-ray Breast Mass Segmentation and Shape Classification"***, Proceeding of 21^{st} International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI2018), pp. 833-840, 2018 Springer, Cham. **Core A**. (Singh et al. (2018b))
2. Md. Mostafa Kamal Sarker, Hatem A Rashwan, Farhan Akram, Syeda Furruka Banu, Adel Saleh, Vivek Kumar Singh, Forhad U H Chowdhury, Saddam Abdulwahab, Adel Saleh, Santiago Romani, Petia Radeva, Domenec Puig, ***"SLSDeep: Skin Lesion Segmentation based on Dilated Residual***

and Pyramid Pooling Networks", Proceeding of 21^{st} International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI2018), pp. pp. 21-29, 2018 Springer, Cham. **Core A**. (Sarker et al. (2018))

3. Vivek Kumar Singh, Santiago Romani, Jordina Torrents-Barrena, Farhan Akram, Nidhi Pandey, Md Mostafa Kamal Sarker, Adel Saleh, Meritxell Arenas, Miguel Arquez, Domenec Puig, ***"Classification of Breast Cancer Molecular Subtypes from Their Micro-Texture in Mammograms using a VGGNet-Based Convolutional Neural Network"***, 20^{th} International Conference of the Catalan Association for Artificial Intelligence (CCIA 2017), pp. 76-85, IOS press, 2017. (Singh et al. (2017))
4. Vivek Kumar Singh, Hatem A Rashwan, Farhan Akram, Nidhi Pandey, Md Mostafa Kamal Sarker, Adel Saleh, Saddam Abdulwahab, Najlaa Maaroof, Jordina Torrents-Barrena, Santiago Romani, Domenec Puig, ***"Retinal Optic Disc Segmentation using Conditional Generative Adversarial Network"***, 21^{st} International Conference of the Catalan Association for Artificial Intelligence (CCIA 2018), pp. 373-380, IOS press, 2018. (Singh et al. (2018a))
5. Vivek Kumar Singh, Mohamed Abdel-Nasser, Hatem A. Rashwan, Farhan Akram, Rami Haffar, Nidhi Pandey, Md. Mostafa Kamal Sarker, Sebastian Kohan, Josep Guma, Santiago Romani, Domenec Puig, ***"Mass Detection in Mammograms using a Robust Deep Learning Model"***, 22^{nd} International Conference of the Catalan Association for Artificial Intelligence (CCIA 2019), 2019. (Accepted)
6. Farhan Akram, Miguel Angel Garcia, Vivek Kumar Singh, Nasibeh Saffari, Md. Mostafa Kamal Sarker, Domenec Puig, ***"Image Segmentation using Active Contours Driven by Bias Fitted Image Robust to Intensity Inhomogeneity"***, 20^{th} International Conference of the Catalan Association for Artificial Intelligence (CCIA 2017), pp. 146-155, IOS press, 2017.
7. Farhan Akram, Vivek Kumar Singh, Miguel Angel Garcia, Md. Mostafa

Kamal Sarker, Domenec Puig, ***"Brain MR Image Segmentation using Multiphase Active Contours Based on Local and Global Fitted Images"***, 21^{st} International Conference of the Catalan Association for Artificial Intelligence (CCIA 2018), pp. 325-324, IOS press, 2018.

8. Mohamed Abdel-Nasser, Antonio Moreno, Mohamed A. Abdelwahab, Adel Saleh, Saddam Abdulwahab, Vivek Kumar Singh, Domenec Puig, ***"Matching Tumour Candidate Points in Multiple Mammographic Views for Breast Cancer Detection"***, International Conference on Innovative Trends in Computer Engineering (ITCE), pp. 202-207. IEEE, 2019.

1.5.3 Book chapters

1. Vivek Kumar Singh, Hatem A. Rashwan, Mohamed Abdel-Nasser, Farhan Akram, Rami Haffar, Nidhi Pandey, Sebastian Kohan, Josep Guma, Santiago Romani, Domenec Puig, ***"A Computer-Aided-Diagnosis System for Breast Cancer Molecular Subtypes Prediction in Mammographic Images"***, State of the Art in Neural Networks, Elsevier, 2019. (To be submitted)

1.6 Thesis organization

The thesis is outlined as follows:

In Chapter 2, we design a CAD system able to classify the four breast cancer molecular subtypes just from the image pixels of digital mammography. The proposed method is based on a VGGNet-based deep learning techniques that are able to learn the micro-texture features of image pixels from tumor area. We have collected 716 image samples of 100 $\times$ 100 pixels wide, manually extracted from real tumor image areas that had been labeled in the digital mammography by a radiologist, jointly with the corresponding oncologist diagnose based on histological indicators. Using this ground truth, we have been able to train and test the proposed CNN, which can achieve a promising accuracy rate. The results of the this chapter are published in Singh et al. (2017).

In Chapter 3, we propose a cGAN devised to segment a breast tumor within a region of interest (ROI) in a mammogram. The generative network learns to recognize the tumor area and to create the binary mask that outlines it. In turn, the adversarial network learns to distinguish between real (ground truth) and synthetic segmentations, thus enforcing the generative network to create binary masks as realistic as possible. The cGAN works well even when the number of training samples is limited. As a consequence, the proposed method outperforms several state-of-the-art approaches. Our working hypothesis is corroborated by diverse segmentation experiments performed on INbreast and a private in-house dataset. The proposed segmentation model working on an image crop containing the tumor as well as a significant surrounding area of healthy tissue (loose frame ROI), provides a significant improvement in terms of Dice Coefficient and Intersection over Union (IoU). In addition, a shape descriptor based on a CNN is proposed to classify the generated masks into four tumor shapes: irregular, lobular, oval and round. The proposed shape descriptor outperforms state-of-the-art methods on DDSM dataset. At the end, a study of tumor shape and molecular subtype correlation has been presented. The results of the this chapter are partially published in Singh et al. (2018b) and they will be fully covered in Singh et al. (2020).

In Chapter 4, we propose to add an atrous convolution layer to the cGAN segmentation model to learn tumor features at different resolutions of breast ultrasound images. To automatically re-balance the relative impact of each of the highest level encoded features, we also propose to add a channel-wise weighting block in the network. In addition, the SSIM and L1-norm loss with the typical adversarial loss are combined in an overall loss function to train the model. The propose model outperforms other state-of-the-art segmentation models. In the classification stage, we show that few statistic features extracted from the shape of the boundaries of the predicted masks can properly discriminate between benign and malignant tumors with a promising accuracy. The results of the this chapter are under review in paper 2 from the journal publication list (1.5.1).

In Chapter 5, we propose an application of cGAN to segment the optic disc from retinal fundus images. Experiments were performed on two publicly available dataset; DRISHTI GS1 and RIM-ONE. The proposed model outperformed several state-of-the-art methods. The results of the this chapter are published in Singh et al. (2018a).

In Chapter 6, we propose an accurate skin lesion segmentation model based on a modified cGAN. We introduce a new block in the encoder of cGAN called factorized channel attention (FCA), which exploits both channel attention mechanism and residual 1-D kernel factorized convolution. The channel attention mechanism increases the discriminability between the lesion and non-lesion features by taking feature channel interdependencies into account. The 1-D factorized kernel block provides extra convolutions layers with a minimum number of parameters to make computations of higher-order convolutions easier. Besides, we use a multi-scale input strategy to encourage the development of filters which are scale-invariant (i.e., constructing a scale-invariant representation). The proposed model is assessed on three skin challenge datasets. It yields competitive results when compared to several state-of-the-art methods. The results of this chapter will be published as paper 3 from the journal publication list (1.5.1).

In Chapter 7, presents the conclusion of the thesis and some lines of future research.

References are presented at the end of the thesis.

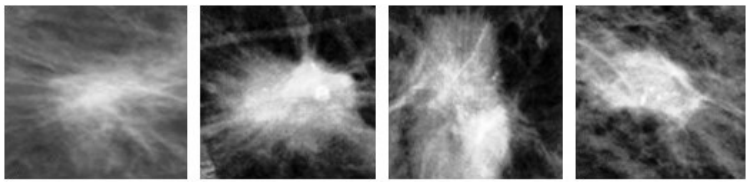
CHAPTER 2

Classification of Breast Cancer Molecular Subtypes from their Micro-Texture in Mammograms

Breast cancer can be detected at early stages by radiologists from periodic screening mammography. However, just by viewing the mammogram they cannot discern the subtype of the cancer, which is a crucial information for the oncologist to decide the appropriate therapy. Consequently, a painful biopsy must be carried out for determining the tumor subtype from cytological and histological analysis of the extracted tissue. This second chapter presents the method to classify the breast cancer molecular subtypes from mammographic images by utilizing deep learning method called VGGNet. The work aims to reduce or avoid the biopsy procedure.

2.1 Introduction

Breast cancer is one of the main causes of high mortality among women Howell et al. (2014). For oncologists, it is critical to correctly identify the malignant breast cancer molecular subtypes (e.g. Luminal A, Luminal B, Her-2+ and Basal-like) to treat them with the appropriate therapy. Digital mammography is the most common acquisition protocol used to detect and locate breast tumors. In Figure 2.1 shows the four breast cancer molecular subtypes in separate regions of interest extracted from digital mammograms. Even for specialized radiologists, it is currently impossible to identify the breast cancer molecular subtypes of a previously detected tumor just by visual inspection of prompts in mammogram images. Hence, a needle biopsy procedure is required to obtain the histological prognostic factors of a suspicious tissue, which provides reliable information to uncover its breast cancer molecular subtypes.



(a) (b) (c) (d)

Figure 2.1: Regions of interest of the four breast cancer molecular subtypes (from left to right): (a) Luminal A, (b) Luminal B, (c) Her-2+ and (d) Basal-like.

2.2 Related work

Up to date, numerous approaches have been proposed to classify the breast cancer tumor subtypes based on histological information. The method designed by Perou et al. (2000) performed a breast cancer classification into certain intrinsic subtypes based on gene expression patterns. Harbeck et al. (2013) presented the guidelines for the breast cancer molecular subtypes categorization based on several immunohistochemistry (IHC) biomarkers such as estrogen receptors (ER), progesterone receptors (PR), human epidermal growth factor receptor 2 (HER-2) and antigen KI-67 (Ki67) has been presented in Table 2.1. Moreover, Spanhol et al. (2016)

avoided the traditional hand-crafted features to propose a deep learning approach characterized by a modified AlexNet CNN architecture. Experiments were conducted using the well-known BreaKHis database[1]. Jeleń et al. (2008) presented a malignant

Table 2.1: Biological markers in the primary tumour including Estrogen receptor (ER), Progesterone receptor (PR), Human epidermal growth factor receptor-2 (HER-2) Falck et al. (2013)

Molecular subtypes	ER	PR	Her-2
Luminal-A	+	+	-
Luminal-B	+	+	+
Her-2	-	-	+
Basal-like	-	-	-

breast cancer classification from cytological images acquired via fine-needle aspiration biopsies. Authors tested the classification performance using several models such as multilayer perceptron (MLP), probabilistic neural networks, learning vector quantization and support vector machines. From the results, they demonstrated the predictive ability of both probabilistic neural networks and support vector machines versus the learning vector quantization and MLP. In addition, Dev et al. (2012) proposed the use of genomic information according to patient treatment to build a CAD system capable of identifying tumor cells. Authors found that their novel fused approach based on both Functional Link Artificial Neural Network (FLANN) and Particle Swarm Optimization (PSO) predict better than other state-of-the-art methods.

On the other hand, Torrents-Barrena et al. (2015b) presented the first work to determine the feasibility of using a CAD system to differentiate among all breast cancer molecular subtypes in mammograms. They hypothesized that computer vision and machine learning algorithms can offer benefits to address the aforementioned problem. Authors designed two classification experiments: Luminal A vs. Luminal B, and Luminal A vs. Luminal B vs. Her-2+ vs. Basal-like. Support Vector Machines (SVM) and Local Binary Patterns (LBP) yielded the best accuracy: 75% and 52.17%, respectively. Moreover, they designed in Torrents-Barrena et al. (2015a) a new methodology based on fractal texture analysis and unsupervised /

[1]http://web.inf.ufpr.br/vri/breast-cancer-database

supervised classifiers. SVM also achieved the best performance (76.48% and 55.67%, respectively). The main drawback of both works was the limited number of Her-2+ and Basal-like samples.

In this chapter, we propose an automated CAD system to classify the four molecular subtypes of breast cancer from full-field digital mammograms (FFDM). A modified VGG_{16} Simonyan and Zisserman (2014) CNN architecture is presented to learn the underlying micro-texture patterns of the mammogram image pixels for each subtypes. Our approach only requires a manual ground truth of the tumor region made by expert radiologists to predict its molecular subtypes without the need of any histological information. Although there is still much room for further improvements, classification results achieved through our methodology is better than our previous methods based on hand-crafted features and that, besides our previous papers, there are no other attempts to solve it.

2.3 Proposed methodology

Recognition of texture patterns in mammograms is challenging due to the high variability of the gray levels of pixels, which correlate to the amount of radiologic energy that has crossed the breast tissue Wong et al. (2012). This variability comes basically from signal noise, from the specific settings of the mammographer, and the tissue features of the breast under inspection. Besides, the micro-texture of the pixels in the tumor area is affected not only by the tumor cells but also by normal tissue around the tumor, as the mammography is a 2D projection of the radioactive beam crossing a full 3D body Bovis and Singh (2000).

Nowadays, Neural Networks have achieved great success in modeling highly complex and unstructured information. Specifically, CNN have shown significant improvements in image recognition tasks, greatly outperforming other classical Computer Vision strategies. Indeed, CNNs have been successfully applied to solve various problems in biomedical image analysis Litjens et al. (2017).

However, the majority of the papers propose to train the CNN to recognize

full objects of interest, e.g., abnormal cells, breast microcalcifications, lung nodules, blood vessels, colon polyps, etc. Materka et al. (1998). In other approaches, the CNN is trained to detect the texture features of the region of interest (ROI) corresponding to body areas suspicious to be ill, like hippocampal sclerosis in brain MRI areas Döhler et al. (2008). Our method belongs to the second strategy, using a VGGNet-based architecture for learning and classifying micro-texture patterns in the ROIs of mammograms corresponding to manually segmented breast tumors.

We use the term micro-texture for referring to similar pixel intensity variations at pixel-wide local areas, corresponding to less than 1 square millimeter, so those similarities are unnoticeable to humans. In contrast, macro-texture refers to repetitions of visible shadings across the ROI. Since macro-texture patterns are not present in the sampled tumors we have seen so far, it is not possible to classify the subtypes of the tumor by mere visual inspection.

Therefore, we hypothesize that a CNN conveniently designed can learn the prototypical underlying micro-textures of each cancer subtypes and that those prototypes are characteristic of each subtype, i.e., they are similar to all samples of the same subtypes but different from the micro-texture prototypes of the other cancer subtypes. Hence, the trained CNN should be able to predict the subtypes of any new breast tumor, given an ROI sample extracted from its corresponding segmented mammography.

The details related to the proposed methodology are discussed below.

2.3.1 VGGNet-based convolutional neural network architecture

The VGGNet architecture was proposed in 2014 for the contest ImageNet Large Scale Visual Recognition Challenge (ILSVRC 2014) Simonyan and Zisserman (2014) of large-scale image classification and also for localizing learning objects within the image. This model demonstrated that the depth of the network (16 or 19 layers) improves the classification performance significantly.

We will base our design on the VGG_{16} architecture, since it uses small area filters

(3×3) that we expect they are well suited for micro-texture prototype learning, in contrast to other CNN architectures (e.g. AlexNet Krizhevsky et al. (2012)) that use larger filters (11×11) to look for edges, macro-textures or other salient features of the objects.

Since our CNN must learn just pixel-wide micro-texture prototypes (not the full tumor shapes) of only four classes of cancer subtypes, we have checked several simplifications of the VGG_{16} original architecture (see Figure 2.2). Concretely, we have defined smaller sets of filters and reduced the number of neuron layers.

For example, in the first convolutional layer, we use just 32 filters, instead of the 64 filters defined by VGG_{16}. We expect that these 32 filters will be enough for representing the most frequent 3×3 pixel configurations of the micro-texture prototypes. This reduction of filters can also be observed in the rest of the convolutional layers.

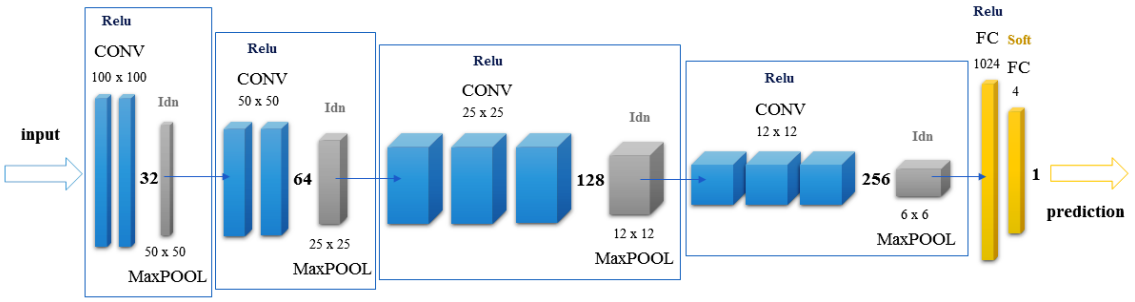


Figure 2.2: Our modified VGG_{16}-based CNN architecture.

For the number of layers, we eliminated the last (higher level) set of 3 convolutional layers, and also removed the second fully connected layer. The first fully connected layer contains 1024 neurons (instead of 4096) and has a dropout ratio of 0.5. The last fully connected layer contains 4 neurons (instead of 1000), using the softmax classifier to assign the final membership degree of the input sample to each class. Thus, we only use 10 convolutional layers and 2 fully connected layers, instead of 13 convolutional layers and 3 fully connected layers proposed in the original VGG_{16}. Once more, the reduction in the number of layers is possible because of the less complexity and fewer number of patterns that the neural network must model.

2.4 Experiments and discussion

2.4.1 Dataset

To prove the feasibility of our proposal, we have conducted several experiments using our own set of mammograms, obtained from patients with breast cancer at the Universitary Hospital Sant Joan de Reus, Spain. The duty of confidentiality and security measures were fully complied, in accordance with the current legislation on the Protection of Personal Data (article 7.1 of the Organic Law 15/1999, 13th of December). The Hospital also wrote an authorization/consent form including all measures to provide this information to the volunteer patients. Our dataset consists of 203 tumors captured with full-field digital mammograms. For each patient, we usually have 4 images, two per breast (CC and MLO), except for patients with mastectomy. So we have selected only images with cancer tumors, with two views per tumor. An expert radiologist has marked the area of each breast tumor, yielding 192 regions of interest (ROIs). Within each ROI, we have manually cropped squared windows of 100×100 pixels, which constitute the input samples to our CNN. Depending on the size of the tumor, we have extracted from one to three image samples per ROI. The samples have been labeled according to its true breast cancer molecular subtypes, which had been diagnosed by an oncologist based on histopathological information obtained from tissues extracted by biopsy. For our study, we ended up with a total of 179 image samples, distributed in 64, 63, 25 and 27 samples for classes Luminal A, Luminal B, Her-2+ and Basal-like, respectively. The last two classes, Her-2+ and Basal-like, are less frequent among the population, so we could not collect as many cases as for the two Luminal classes. Nonetheless, we expected that the shorter number of samples would be significant enough, as we will discuss below.

2.4.2 Pre-processing and data augmentation

To get rid of signal noise, we have applied a soft Gaussian smoothing to the mammography pixels, with sigma equal to 0.75 in image coordinates. Afterward, we

have scaled down the 10-bit or 12-bit values of the pixels of processed mammograms into 8-bit values with a simple linear transformation, to accommodate input data to the range expected by the CNN.

Neural Networks usually need from tens of thousands to millions of training samples for a proper fitting of the weights of all neuron inputs, because the number of neuron inputs can be of that orders of magnitude. Since we just have hundreds of tumor samples, we apply data augmentation techniques to obtain diversified views of the available information. For our experimental framework, we have chosen to rotate the original samples 90°, 180° and 270°, thus multiplying by four the initial set of image samples, thus yielding a total of 716 image samples. Fortunately, image rotations of 90° preserve the spatial scale of the pixel-value variation, while it provides different orientations of the portrayed micro-texture patterns.

2.4.3 CNN model training

When training CNNs, one usually starts from a pre-trained set of weights and then fine tunes the fitting of the parameters with specific training samples of the environment in which the final CNN must work on Tajbakhsh et al. (2016). However, we decided to start from scratch to check the intrinsic ability of the proposed architecture to fit the weights of the neuron inputs for discerning the underlying micro-texture patterns present in the given image samples.

To train our CNN model, we split our dataset into 70% image samples for training, 15% for validation and 15% for testing. Figure 2.3, shows six examples of samples for each of the four subtypes of cancer to be recognized. Some of these samples are used for training, validation or testing: we have chosen the better defined (well contrasted, sharper, no artifacts) to be the training samples, expecting that they would carry essential micro-texture patterns information of each class. For example, we excluded samples with microcalcification's, such as the sixth example (starting from the left) of Luminal A and the third example of Luminal B.

In the definition of training hyperparameters, we found that 0.01 is a proper learning rate. Higher learning rates will decay the loss faster but they can get stuck

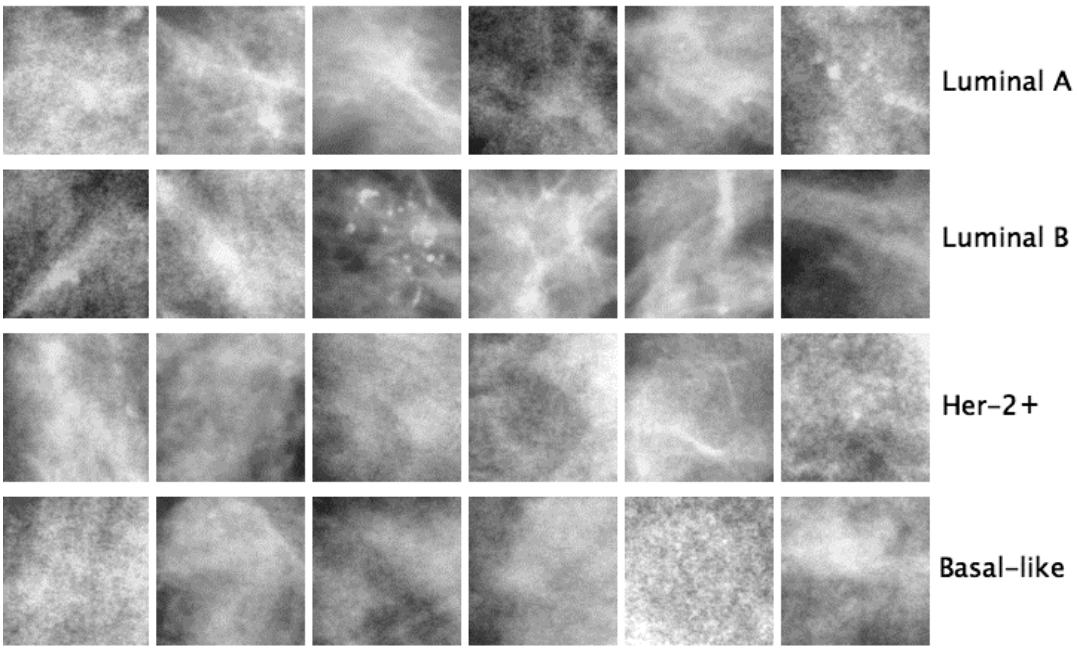


Figure 2.3: 24 examples of input image samples, 6 for each of the 4 molecular subtypes of breast cancer

at worse loss values. This is because there is too much "energy" in the parameters and the optimization can bounce around chaotically. To avoid that, the learning rate is reduced by a factor of 10 at every 5 epochs. Furthermore, the Adadelta optimizer is used with the momentum of 0.9 and we have set up mini-batches of 10 images.

For obtaining the maximum accuracy, we have checked dozens of different network architectures, tweaking the hyperparameters such the number of layers, filters per layer, number of nodes in fully connected layers and others (e.g. learning rate, momentum).

The process of training a CNN is extremely computationally expensive, due to the huge amount of calculations on arrays of data and weights that must be carried out. To perform the experiments, we used a PC with an Intel I3, 64-bit 2.90GHz quad-core CPU with 4GB of memory space, running an Ubuntu 14.04 Linux operating system. We used CPU to simulate the CNN model on the deep learning framework Keras Chollet et al. (2015).

2.4.4 Experimental results

In order to validate our approach, we grouped the test samples according to their ground truth labels and passed them as input to the trained CNN. After comparing the predicted class with the true class, we can obtain an accuracy index of our classification system. More specifically, we have computed confusion matrices.

Firstly, we have checked the performance of our model by training and validating the network with regards to the first two classes, Luminal A and Luminal B, which correspond to the less aggressive cancer subtypes. Table 2.2 depicts the confusion matrix for our best 2-class classification.

Table 2.2: Confusion matrix for the 2-class experiment: each cell shows both the number of samples of each ground truth group (rows) classified to each available class (columns), as well as its corresponding percentage with respect to the total number of test samples of the group.

		Prediction		
		Luminal A	Luminal B	# Test samples
Ground Truth	Luminal A	36 (95%)	2 (5%)	38
	Luminal B	15 (39%)	23 (61%)	38

A significant amount of samples has been correctly classified to their ground truth class, as can be seen in the diagonal of the previous confusion matrix. Our network has performed well on Luminal A samples, achieving a 95% of accuracy. On the other hand, just 61% of Luminal B samples had been correctly classified, while the remaining 39% had been misclassified as belonging to Luminal A. That indicates that the network fitting had slightly biased to the Luminal A class. Nevertheless, our network renders an overall accuracy around 78%, which is quite a good result taking into account the evident lack of visual patterns in the image samples (see fig.2.3).

In addition to confusion matrices, we also present plots rendering the loss and accuracy evolution through the iterative training+validation phase. The evolution of each indicator is shown for both training and validation samples. Theoretically, the loss should tend to 0.0 (no misclassification) and the overall accuracy should tend to 1.0 (100% of accuracy), after a certain number of epochs.

Plots in Figure 2.4 show the evolution of loss and accuracy for the 2-class

experiment. As can be observed, the loss reduces and the accuracy increases as the training procedure evolve through epochs, which indicates that the network parameters are being nicely optimized. The validation indicators are less stable than the training ones, but one can say that they follow the expected tendency, on average.

The second experiment we present here corresponds to the full 4-class classification, i.e., including all breast cancer subtypes. Table 2.3 shows the confusion matrix for the whole 106 test image samples. From the results, Luminal A and

Table 2.3: Confusion matrix for the 4-class classification

		Prediction				
		Luminal A	Luminal B	Her-2+	Basal-like	# Test samples
Ground Truth	Luminal A	31 (82%)	7 (18%)	0 (0%)	0 (0%)	38
	Luminal B	14 (37%)	24 (63%)	0 (0%)	0 (0%)	38
	Her-2+	0 (0%)	0 (0%)	10 (71%)	4 (29%)	14
	Basal-like	2 (13%)	0 (0%)	8 (50%)	6 (37%)	16

Her-2+ have performed reasonably good, taking into account that the complexity of the classification has increased significantly. However, the accuracies for Luminal B are fair and Basal-like are poor. This performance could be induced by a lack of micro-texture templates in Basal-like tumors, or because of a high degree of similarity between tissue density of the Her-2+ and Basal-like, but we also must notice that the number of samples for these classes are around one half of the number of the two other classes, so the network may be biasing it's learning to the micro-texture patterns of Luminal A and B samples. From the individual accuracy, we can obtain

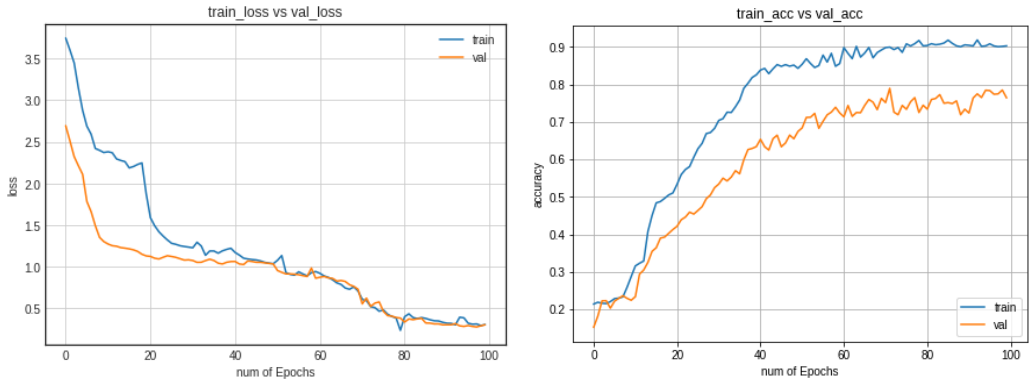


Figure 2.4: Evolution of loss (left plot) and accuracy (right plot) of the 2-class experiment, for both training and validation samples.

an overall accuracy as the weighted average concerning the number of test samples of each class, obtaining a fair 67% of good predictions.

Nevertheless, the majority of misleading predictions of Her-2+ and Basal-like classes are located within this group of two classes, and the misleading predictions of Luminal A and B are also located within the group of the former two classes. That indicates that the network is distinguishing well between samples of the two groups. Therefore, one could try another experiment for classifying samples to the group of Luminal A and B classes, which are less harmful cancers (malignant but with better prognosis), or to the group of Her-2+ and Basal-like classes, which are more severe cancers (worse prognosis).

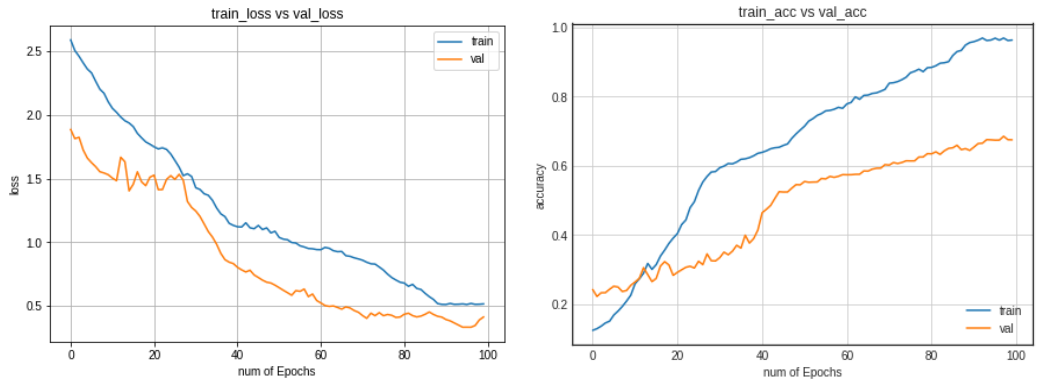


Figure 2.5: Evolution of loss (left plot) and accuracy (right plot) of the 4-class experiment.

Figure 2.5 shows the evolution of loss and overall accuracy for the 4-class experiment. The plots show the same behavior than in the 2-class experiment, i.e. properly stable tendencies, which indicates that the training process has done a good optimization in about 100 epochs.

2.4.5 Discussion

From the experimental results shown above, we can deduce (cautiously) that our hypothesis seems to be true, i.e., different subtypes of malignant cancers depict intrinsic (and reduced) sets of micro-texture patterns in their mammographic projections.

Our CNN architecture has been simplified to learn such micro-texture templates.

In our framework, CNN does not have to learn full objects. Therefore, the network just can track for similarities in the pixel intensity distribution within tiny areas of the samples: all convolutional filters are applied in windows of 3×3 positions of the feature maps, which is ideal to account for micro-texture. Indeed, from the second convolutional layer, the sensitivity of filters spread across larger areas of the input image (5×5, 7×7, etc.), due to the chaining of the neuron layers.

We have reduced the original VGGNet architecture to end up with a network of about 11 million parameters, which is a small fraction of the 138 million parameters of the original VGG_{16} network. Nevertheless, to successfully train a neural network from scratch, one had to feed a number of training samples of one order of magnitude above the number of parameters. So, how is it possible that our network has been properly fitted with less than one thousand samples? Our explanation is the following: notice that each image sample of 100×100 pixels actually contains 98x98 windows of 3×3 pixels, for convolutional filters moving at a stride of 1 pixel in each direction. Since the subsequent convolutional layers account for wider windows, we can add 96×96 windows of 5×5 pixels per sample, and so on. At the end, the number of examples of micro-texture per sample is very high, approaching to the order of 10^5. Thus, multiplying by 10^3 samples, we get a total of 10^8 examples of micro-texture, which can guide the fitting of 10^7 parameters.

2.5 Conclusion

In this chapter, we have presented a supervised breast cancer molecular subtypes classification method based on a CNN that analyse manually selected areas of breast tumors found in DICOM images of mammograms. To the best of our knowledge, this is the first effort to predict the molecular subtypes of malignant tumors just from image excerpts of digital mammograms using CNNs. Before, we tried other approaches to the same problem using classical texture descriptors (Uniform Local Binary Patterns, Histogram of Gradients, Gabor filters, Fractal dimension), but with less degree of accuracy ([7]: 75% — 52%; [8]: 76% — 56%; current approach: 78%

— 67%). Other authors have only focused on automatic detection of tumors and determining if the tumor is benign or malignant.

The obtained results suggest that the proposed CNN architecture is able to learn the intrinsic micro-texture patterns of Luminal A and Luminal B image samples, giving a good prediction rate of about 78%. Although the individual accuracies for Her-2+ and Basal-like classes cannot be accepted as good, we have found that the textural features of the samples of these two classes seems to be very different from the ones of the two Luminal classes, which is a good hint to continue our research in this direction. Our proposed method does not need any histopathological data or gene test to classify the cancer subtypes of breast tumors.

CHAPTER 3

Breast Tumor Segmentation and Classification in Mammograms

Mammogram inspection in search of breast tumors is a tough assignment that radiologists must carry out frequently. Therefore, image analysis methods are needed for the detection and delineation of breast tumors, which portray crucial morphological information that will support reliable diagnosis. This third chapter presents the method of fully automated CAD system for breast diagnosis which involves the detection of the mass region, segmentation of ROI and tumor shape classification from mammograms. Also, to find out the malignancy of the mass, we provide a correlation study between tumor shape and molecular subtypes.

3.1 Introduction

Mammography is a world recognized tool that has been proven effective to reduce the mortality rate, since it allows early detection of breast diseases Lauby-Secretan et al. (2015). Breast masses are the most important findings among diverse types of breast abnormalities, such as micro-calcification and architectural distortion. All these findings may point out the presence of carcinomas Rangayyan et al. (2010). Moreover, morphological information of tumor shape (irregular, lobular, oval and round) and margin type (circumscribed, ill defined, spiculated and obscured) also play crucial roles in the diagnosis of tumor malignancy Tang et al. (2009).

CAD systems are highly recommended to assist radiologists in detecting breast tumors and outlining their borders. However, breast tumor segmentation and classification are still challenges due to low signal-to-noise ratio and variability of tumors in shape, size, appearance, texture and location. Recently, many studies based on deep representation of breast images and combining features have been proposed to improve performance on breast mass classification Jiao et al. (2018) .

In addition, based on mammographic images, it is very complicated for an expert radiologist to discern the molecular subtypes, i.e., Luminal-A, Luminal-B, HER-2 and Basal-like (triple negative), which are key for prescribing the best oncological treatment Cho (2016), Liu et al. (2016a), Tamaki et al. (2011). However, recent studies point out some loose correlations between visual tumor features (e.g., texture and shape) and molecular subtypes. In the previous chapter, we have introduced a CNN to classify molecular subtypes using texture patches extracted from mammography Singh et al. (2017), which yielded an overall accuracy of 67%. However, depending only on texture feature is not sufficient to classify the breast cancer molecular subtypes from mammograms Tamaki et al. (2011).

In this chapter, we propose a method based on two main stages, one for breast tumor segmentation and another for tumor shape classification, as shown in Figure 3.1. Before applying our segmentation approach, the SSD Liu et al. (2016b) is used to locate the tumor and then our method computes the proper coordinates to crop the ROI. Afterwards, the first stage segments the breast tumor, contained in

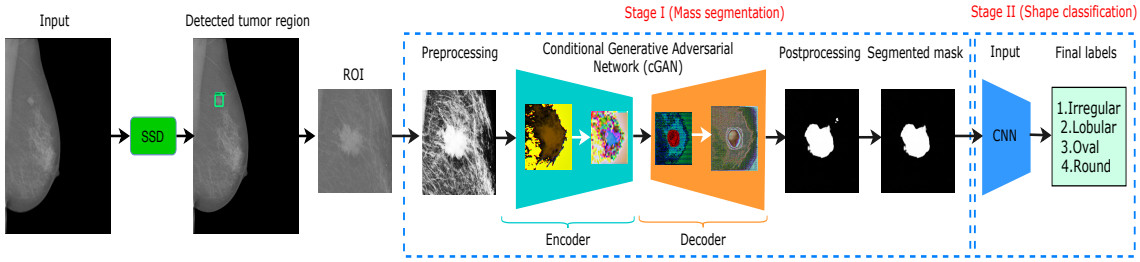


Figure 3.1: Automatic workflow for our breast tumor segmentation and shape classification system.

the ROI, as a binary mask. In the second stage, the binary mask is classified to a shape type (irregular, lobular, oval and round). Unlike traditional object classifiers Kisilev et al. (2015), Kim et al. (2018) that use texture, intensity or edge information, our method is forced to learn only morphological features from the binary masks. The current proposal is a thorough improvement of our previous work Singh et al. (2018b). The major contributions of this chapter are as follows:

1. We believe this is the first adaptation of cGAN in the area of breast tumor segmentation in mammograms. The adversarial network yields more reliable learning than other state-of-the-art algorithms since training data is scarce (*i.e.,* mammograms with labeled breast tumor boundaries), while it does not increase the computational complexity at prediction time.

2. The application of a multi-class CNN architecture to predict the four breast tumor shapes (*i.e.,* irregular, lobular, oval and round) using the binary mask segmented in the previous stage (cGAN output).

3. An in-depth evaluation of our system's performance using two public (1,274 images) and one private (300 images) databases. The obtained results outperform current state-of-the-art in both tumor segmentation and shape classification.

4. A study of the correlation between the tumor shape predicted by our automatic method with respect to the ground-truth molecular subtypes of breast cancer, which reasonably matches with other clinical studies like Boisserie-Lacroix et al. (2013).

3.2 Related work

In the following paragraphs we point out some works mainly focused on breast tumor segmentation and shape classification in mammography, as well as generic image analysis methods highly related with our field of interest.

3.2.1 Tumor segmentation background

CNNs can automatically learn features from the given images to represent objects at different scales and orientations. By increasing the number of layers (depth of CNN model) more detailed features can be obtained, which play crucial part in solving different computer vision problems, such as object detection, classification and segmentation. Thus, numerous methods has been proposed to solve the image segmentation problem based on deep learning approaches Schmidhuber (2015).

One of the well-known architectures for semantic segmentation is the Fully Convolutional Network (FCN) Long et al. (2015), which is based on encoding (convolutional) and decoding (deconvolutional) layers. This approach gets rid of the fully connected layers of CNNs to convert the image classification networks into image filtering networks. An improvement of this scheme was proposed by the U-Net architecture Ronneberger et al. (2015), where skip connections between encoding and decoding layers are added to retain significant information from the input features. Later on, a new variation of FCN was proposed Badrinarayanan et al. (2017) named SegNet, which consists of hierarchy of decoders, each one corresponding to each encoder. The decoder network uses the max-pooling indices received from the corresponding encoder to perform non-linear upsampling of their input feature maps.

Since semantic segmentation has achieved great progress with deep learning, there is recent popularity in applying such models to medical imaging, such as for skin lesions segmentation (Litjens et al. (2017), Sarker et al. (2018)), and for fundus photography of the rear of an eye (Fu et al. (2018a), Singh et al. (2018a)).

For breast tumor detection, segmentation and classification, many medical image analysis methods have been proposed so far, such as Yassin et al. (2018)

and Hamidinekoo et al. (2018). A tumor classification and segmentation method was proposed Rouhi et al. (2015) using an automated region growing algorithm whose threshold was obtained by a trained Artificial Neural Network (ANN) and Cellular Neural Network (CeNN). In turn, to reduce the computational complexity and increase the robustness, a quantized and non-linear CeNN for breast tumor segmentation was proposed in Liu et al. (2018). After segmenting the breast tumor region, a Multilayer Perceptron Classifier was used for tumor classification as benign or malignant.

Furthermore, Dhungel et al. (2015b) segmented breast tumors using Structured Support Vector Machines (SSVM) and Conditional Random Fields (CRF). Both graphical models minimize a loss function build on pixel probabilities provided by a CNN and Deep Belief Network, a Gaussian Mixture Model (GMM) and shape prior. The SSVM is based on graph cuts and the CRF relies on tree re-weighted belief propagation with truncated fitting training Dhungel et al. (2015a). Cardoso et al. (2015, 2017) tackled the same problem by employing a closed contour fitting in the mammogram and minimizing a cost function depending on the radial derivative of the tumor contour. A measure of regularity of the gray pixel values inside and outside the tumor was also included in Cardoso et al. (2017).

In turn, Zhu et al. (2018) proposed an FCN concatenated to a CRF layer to impose the compactness of the segmentation output taking into account pixel position. This approach was trained end-to-end, since the CRF and FCN can exchange data in the forward-backward propagation. An adversarial term was introduced to prevent the samples with the worst perturbation in the loss function, which reduced the overfitting and provided a robust learning with few training samples. In addition, Al-antari et al. (2018) proposed a CAD system consisting of three deep learning stages for detecting, segmenting and classifying the tumors in mammographic images. To locate tumors in a full mammogram, the You Only Look Once (YOLO) network proposed in Redmon et al. (2016) was used. A Full resolution Convolutional Network (FrCN) was then used for segmenting the located tumor region. Finally, a CNN network was used for classifying segmented ROI as

either benign or malignant.

We believe that Yang et al. (2017) is the first work that exploits GAN Goodfellow et al. (2014) for medical image segmentation. In particular, they performed three-dimensional (3D) liver segmentations using abdominal Computerized Tomography (CT) scans. In Singh et al. (2018b), we adapted a cGAN image-to-image translation algorithm Isola et al. (2017) to address the tumor segmentation in two-dimensional (2D) mammograms. With that method, we achieved state-of-the-art performance on both public and private databases.

3.2.2 Shape classification background

In the literature, many approaches used traditional computer vision techniques to extract hand-craft features and subsequently classify them. For instance, Matos et al. (2018) applied the Bag of Features (BoF) paradigm on local feature descriptors, such as Scale-Invariant Feature Transform (SIFT), Speed Up Robust Feature (SURF) or Local Binary Patterns (LBP), achieving very high accuracy of 99% in classifying tumors as malignant or benign.

Recently, deep learning architectures have been designed for 2D and 3D shape classification Kurnianggoro et al. (2018). For example, topological data analysis (TDA) using deep learning was proposed in Hofer et al. (2017) to extract relevant 2D/3D topological and geometrical information. In turn, a CNN model was formulated, which used spectral graph wavelets in conjunction with the BoF paradigm to target the shape classification problem Masoumi and Hamza (2017).

In addition, the authors in Fang et al. (2015) proposed a CNN based shape descriptor for retrieving the 3D shapes. A deep neural network named PointNet was proposed Qi et al. (2017), which directly consumes point cloud for object classification, localized and global semantic segmentation. Moreover, a deep learning framework for efficient 3D shape classification Luciano and Hamza (2018) used geodesic moments by inheriting various properties from the geodesic distance, like the intrinsic geometric structure of 3D shapes and the invariance to isometric deformations.

To date, numerous shape classification methods are applied for medical image analysis Singh et al. (2018b), and Kim et al. (2018). An automated method for textual description of anatomical breast tumor lesions was proposed by Kisilev et al. (2015), which performs joint semantic estimation from image measurements to classify the tumor shape. In addition, Kisilev et al. (2016) also presented a multi-task fast region-based CNN Ren et al. (2015) to classify three tumor shapes: irregular, oval and round. Furthermore, the work in Kim et al. (2018) utilized a GAN to diagnose and classify tumors in mammograms into four shapes: irregular, lobular, oval and round. Previously, Singh et al. (2018b) proposed a multi-class CNN to categorize the tumor shapes into four classes as in Kim et al. (2018) from the public dataset DDSM[1].

3.3 Proposed methodology

The proposed CAD system shown in Fig. 3.1 is divided into two stages: breast tumor segmentation and shape classification.

3.3.1 Obtaining and processing ROIs

Before feeding an image to the first stage, our optimal workflow applies the SSD Liu et al. (2016b) to locate the tumor position and fit a bounding box around it. Based on these bounding coordinates, our method computes new coordinates containing the tumor (*vide infra*), and then uses these new coordinates to crop the mammogram, thus obtaining the Region of Interest (ROI). We evaluated different detectors based on deep learning models, such as SSD Liu et al. (2016b), YOLO Redmon et al. (2016) and Faster R-CNN Ren et al. (2015). Empirically, the SSD detector yields the best results since it is able to detect small tumor regions and provides an overall accuracy of 97%. We are not targeting object sizes less than 7×7 pixels because those objects are really hard to be identified as tumors. Indeed,

[1]https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM

they may correspond to other types of findings, such as calcifications. We are considering only mammograms with tumors, since our main goal is tumor shape classification following tumor segmentation. Therefore, we have not applied SSD on normal mammograms (no tumor), although the SSD method is capable of dealing with this case as well.

To obtain the proper cropping area, our best framing method, so-called loose frame, expands the original bounding box coordinates by adding extra space around, so that the cropped ROI always encompasses the tumor as well as some surrounding area containing healthy tissue (30% and 70% for tumor and healthy tissues, respectively). The computed coordinates are shifted to make the ROI frame fit inside the mammogram image. Besides, both sides of the frame are set equal in order to preserve the original aspect ratio of tumors. Last adjustments required to make the image square sometimes cause the tumor be out of the ROI center. However, this does not preclude the segmentation and classification due to the position-independent nature of convolutional filters.

Moreover, ROI images are scaled to 256×256 pixels, which is the optimal cGAN input size found experimentally. After scaling, they are pre-processed for noise removal as proposed in Kshema et al. (2017) (Gaussian filter with $\sigma = 0.5$ yields the best segmentation results) and then contrast is enhanced using histogram equalization, similarly to Cheng et al. (2003). Then, we apply a normalization for rescaling the pixel values between [0,1].

The prepared data is then fed to the cGAN to obtain a binary mask of the breast tumor, which is post-processed using morphological operations (we used filter sizes of 3×3 for closing, 2×2 for erosion, and 3×3 for dilation) to remove small speckles, as proposed in Hazarika and Mahanta (2018). Fig. 3.2 shows a couple of examples of these small speckles, enclosed in red boxes, which are filtered out after post-processing.

In the second stage, the output binary mask is downsampled into 64×64 pixels, which is then fed to a multi-class CNN shape descriptor to categorize it into four classes: irregular, lobular, oval and round. The reason of this downsampling is that

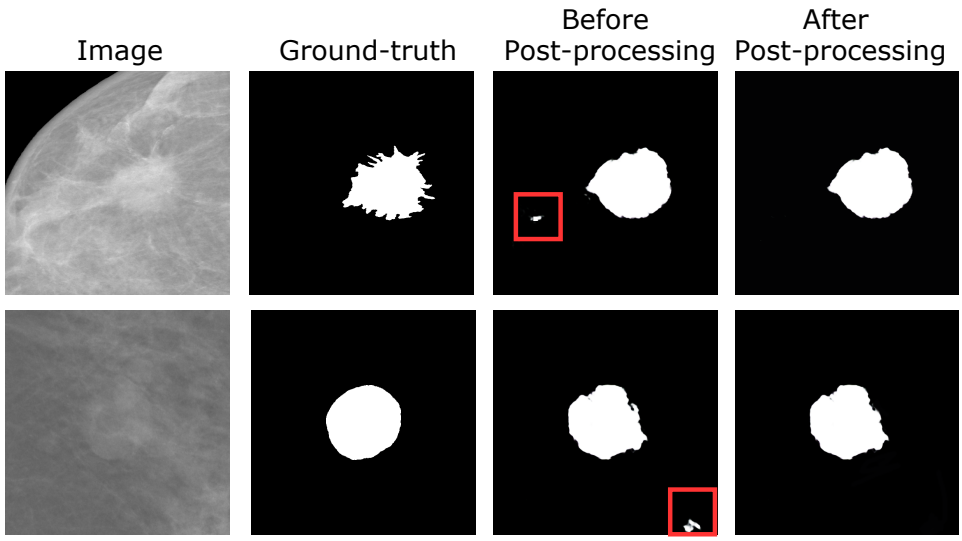


Figure 3.2: Two examples of the effect of morphological post processing after the segmentation.

our shape classification CNN does not need a high resolution image to extract the core morphological features for each class, since the tumors are represented with flat white areas in front of a black background. Hence, changes in the image present very low frequencies.

Our work presented in Singh et al. (2018b) demonstrates the feasibility of applying the cGAN image-to-image translation approach Isola et al. (2017) to breast tumor segmentation, since it can be adapted to our problem in the following senses:

1. The Generator G network of the cGAN is an FCN composed of encoding and decoding layers, which learn the intrinsic features (gray-level, texture, gradients, edges, shape, etc.) of healthy and unhealthy (tumor) breast tissue, and generate a binary mask according to these features.

2. The Discriminative D network of the cGAN assesses if a given binary mask is likely to be a realistic segmentation or not. Therefore, including the adversarial score in the computation of the generator loss strengthens its capability to provide a correct segmentation.

The combination of G and D networks allows robust learning with few training samples. Since the ROI image is a conditioning input for both G and D, the segmentation result is better fitted to the tumor appearance. Otherwise, regular

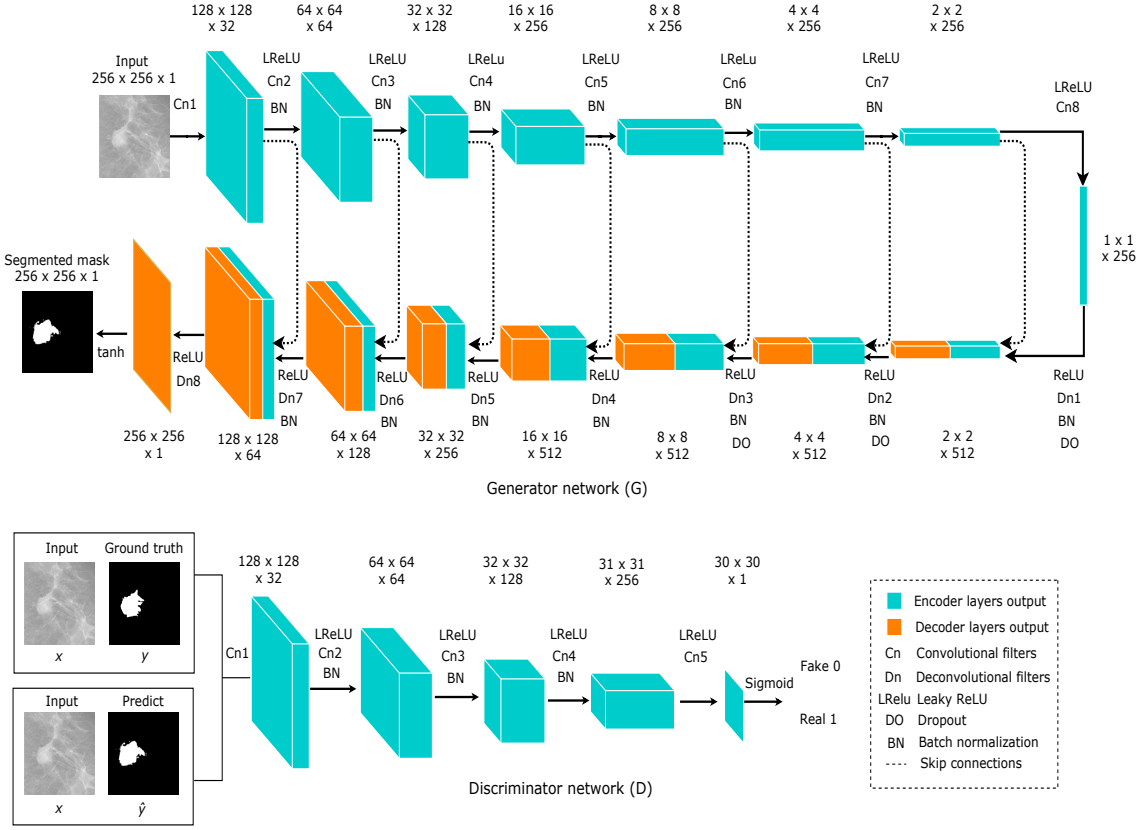


Figure 3.3: Proposed cGAN architecture: generator G (top), and discriminator D (down).

(unconditional) GAN Goodfellow et al. (2014) will infer the segmentation just from random noise, which will require more training iterations compared to the cGAN to obtain an acceptable segmentation result.

3.3.2 Tumor segmentation model (cGAN)

Fig. 3.3 represents the suggested architectures for G and D. The former consists of several encoding and decoding layers (see Fig. 3.3-top). Encoding layers are composed of a set of convolutional filters followed by batch normalization and the leaky ReLU (slope 0.2) activation function. Similarly, decoding layers are composed of a set of deconvolutional filters followed by batch normalization, dropout and ReLU.

Convolutional and deconvolutional filters are defined with a kernel of 4×4 and stride of 2×2, which respectively downsample and upsample the activation maps by a factor of 2. Batch normalization is not applied after the first and the last convolutional filters (Cn_1 and Cn_8). After Cn_8, the ReLU activation function is

applied instead of leaky ReLU. Dropout is applied only at the first three decoding layers (Dn_1, Dn_2 and Dn_3). There is no skip connection in the last decoding layer (Dn_8), after which the *tanh* activation function is applied to generate a binary mask of the breast tumor.

The architecture of D shown in Fig. 3.3 at down consists of five encoding layers with convolutional filters with a kernel of 4×4, stride 2×2 at the first three layers and stride 1×1 at 4^{th} and 5^{th} layers. Batch normalization is applied after Cn_2, Cn_3 and Cn_4 and a leaky ReLU (slope 0.2) is applied after each layer except for the last one. The sigmoid activation function is used after the last convolutional filter (Cn_5). The network input is the concatenation of the ROI and the binary mask to be evaluated (ground truth or predicted). The output segmentation is an array of 30×30 values, each one from 0.0 (completely fake) to 1.0 (perfectly plausible or real). Each output value is the degree of proper segmentation likelihood of a crop of the binary mask and the input image, which corresponds to a 70×70 receptive field for each value.

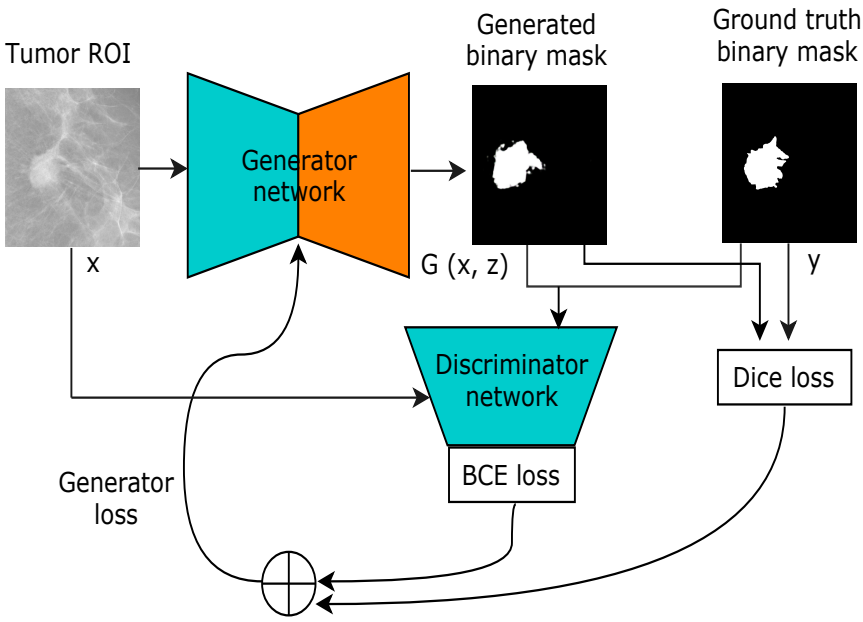


Figure 3.4: Proposed cGAN framework based on Dice and BCE losses.

Let x be a tumor ROI, y the ground truth mask, z a random variable, λ an empirical weighting factor, $G(x, z)$ and $D(x, G(x, z))$ the outputs of G and D,

respectively. Then, the loss function of G is defined as:

$$\ell_{Gen}(G, D) = \mathbb{E}_{x,y,z}(-\log(D(x, G(x, z)))) + \lambda \mathbb{E}_{x,y,z}(\ell_{Dice}(y, G(x, z))), \tag{3.1}$$

where z is introduced as dropout in the decoding layers Dn_1, Dn_2 and Dn_3 at both training and testing phases, which provides stochasticity to generalize the learning processes and avoid overfitting.

The optimization process of G will try to minimize both expected values, *i.e.*, the D values should approach to 1.0 (correct tumor segmentations), and the Dice loss ℓ_{Dice} should approach to 0.0 (generated masks are equal to ground truth). Both terms of generator loss enforce the proper optimization of G: the Dice loss term fosters a rough prediction of the mask shape (central tumor area) while the adversarial term fosters an accurate prediction of the mask outline (tumor borders). Neglecting one of the two terms may lead to either very poor segmentation results or slow learning speed.

In addition, $\ell_{Dice}(y, G(x, z))$ is the Dice loss of the predicted mask with respect to ground truth, which is defined as:

$$\ell_{Dice}(y, z) = 1 - \frac{2|y \circ G(x, z)|}{|y| + |G(x, z)|}, \tag{3.2}$$

where $\circ$ is the pixel wise multiplication of the two images and $|.|$ is the total sum of pixel values of a given image. If inputs are binary images, then each pixel can be considered as a boolean value (white is 1 / black is 0). The formulation in (3.2) is equivalent to the Dice coefficient *i.e.,* $2 \times \frac{TP}{TP+FN+TP+FP}$, but it must be subtracted from 1.0 because the loss function will be minimized. Let A be the ground truth of the ROI and B the segmented region. Then the true positive degree (TP) is defined as $TP = A \cap B$, which is the area of the segmented region common in both A and B. The false positive degree (FP) is defined as $\overline{A} \cap B$, which is the segmented area not belonging to A. Similarly, the false negative degree (FN) is defined as $A \cap \overline{B}$, which is the true area missed by the proposed segmentation method.

In our previous methodology proposed in Singh et al. (2018b), the generator

network loss was formulated by combining the logistic Binary Cross Entropy (BCE) loss and the $L1$-norm. In the methodology proposed in this chapter, we replace the $L1$-norm loss with the Dice loss as shown in Fig. 3.4. $L1$-norm loss minimizes the sum of absolute differences between the ground truth label y and the estimated binary mask $G(x, z)$ obtained from the generator network, which takes all pixels into account. In turn, Dice loss is highly dependent on TP predictions, which is the most influential term in foreground segmentation. Fig. 3.5 shows that the Dice loss achieves lower values (more optimal) than the $L1$-norm loss.

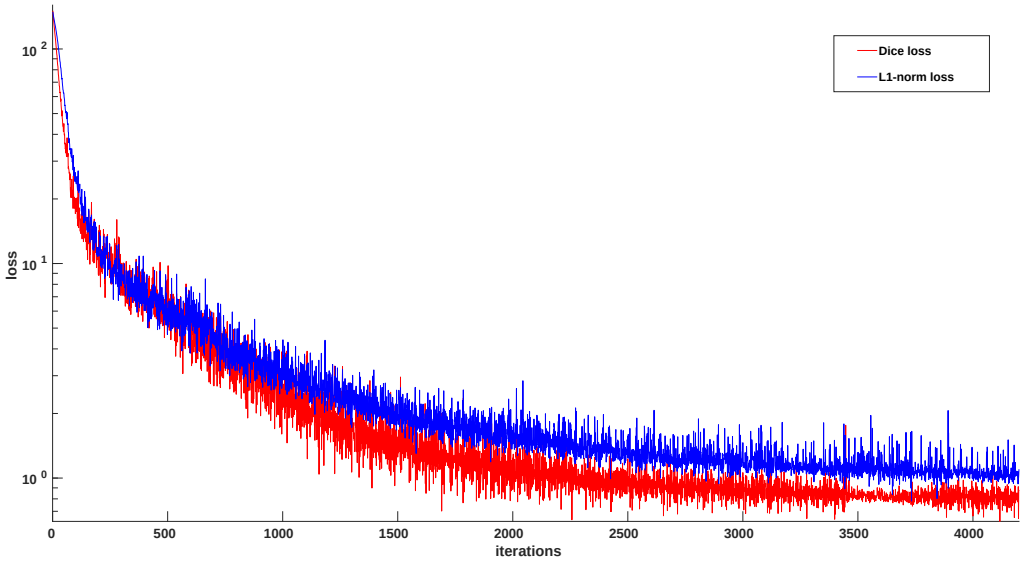


Figure 3.5: Dice and $L1$-norm loss comparison over iterations.

Moreover, the loss function of D is defined in (6.8):

$$\ell_{Dis}(G, D) = \mathbb{E}_{x,y,z}(-\log(D(x, y))) + \mathbb{E}_{x,y,z}(-\log(1 - D(x, G(x, z)))) \tag{3.3}$$

The optimizer will fit D to maximize the loss values for ground truth masks (by minimizing $-\log(D(x, y))$) and minimize the loss values for generated masks (by minimizing $-\log(1 - D(x, G(x, z)))$. These two terms compute BCE loss using both masks, assuming that the expected class for ground truth and generated masks is 1 and 0, respectively.

The optimization of G and D is done concurrently, *i.e.*, one optimization step for both networks at each iteration, where G learns how to compute a valid

tumor segmentations and D learns how to differentiate between synthetic and real segmentations.

We have experimented on different hyper-parameters to improve the segmentation accuracy of our contribution in Singh et al. (2018b). Besides introducing the Dice loss, we have reduced the number of filters of each network from 64 to 32. We also explored different learning rates and loss optimizers (SGD, AdaGrad, Adadelta, RMSProp and Adam), finding the best combination at Adam with $\beta_1 = 0.5$, $\beta_2 = 0.999$ and initial learning rate = 0.0002 with batch size 8. In (3.1), the Dice loss weighting factor $\lambda = 150$ was found to be the best choice. Finally, the best results were achieved by training both G and D from scratch for 150 epochs.

3.3.3 Shape classification model

In the literature, various approaches for tumor shape classification have found that texture and intensity features are relevant for their proposals. However, in this proposal we attempt to use only shape context to classify the tumor shapes. Specifically, we propose a multi-class CNN architecture for breast tumor shape classification (*i.e.*, irregular, lobular, oval and round) using the binary masks obtained from the cGAN. Methods attempting to directly categorize the shape using breast tumor intensity, texture, boundary, etc. include Kisilev et al. (2015, 2016); Ren et al. (2015); Kim et al. (2018), but they all render high computational complexity. We simplify the problem by extracting morphological features only from binary masks.

As shown in Fig. 3.6, our model consists of three convolutional layers with kernel sizes 9×9, 5×5 and 4×4, respectively, and two fully connected (FC) layers. The first two convolutional layers are followed by 4×4 max-pooling with stride 4×4. The output of the last convolutional layer is flattened and then fed into the first FC layer with 128 neurons. These four layers use ReLU as activation function. A dropout of 0.5 is used to reduce overfitting in the first FC layer. Finally, the last FC layer with 4 neurons applies the softmax function to generate the final membership degree of the input binary mask to each class. A weighted categorical cross-entropy loss is

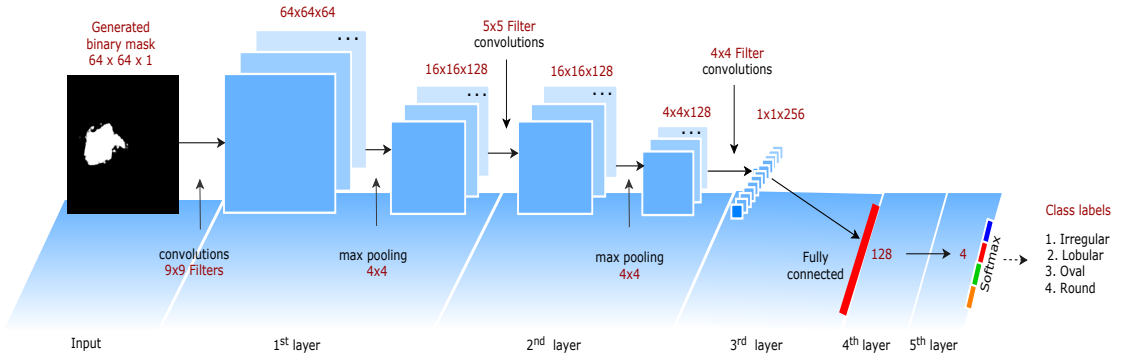


Figure 3.6: CNN architecture for tumor shape classification.

used to avoid the problem of unbalanced dataset. The class weight is one minus the ratio of samples per class to the total number of samples.

The RMSProp is employed for optimizing the model with learning rate = 0.001, momentum = 0.9 and batch size = 16. The network is trained from scratch and the weights of five layers are randomly initialized. During training, we experimentally found the best architecture, number of layers, filters per layer, and number of neurons in FC layers.

3.4 Experiments and discussion

We have evaluated the performance of proposed models on two public mammography datasets and one private dataset:

INbreast dataset [2]

It is a publicly available database containing a total of 115 cases (410 mammograms), which include: masses, calcifications, asymmetries and distortions. However, only 106 out of 410 mammograms have their corresponding ground truth of binary masks. Thus, we only used this 106 mammograms to test our detection and segmentation model.

[2]http://medicalresearch.inescporto.pt/breastcancer/index.php/Get_INbreast_Database/

DDSM dataset

It is a publicly available digital database for screening mammography containing 2,620 mammography studies. In this work, 1,168 cases of breast tumors with their corresponding ground truths are used for shape classification, where 504, 473, 115 and 76 tumors are labeled as irregular, lobular, oval and round, respectively. The remaining images are excluded since they do not provide shape labels. We have used 75% of the images for training and the rest for testing the tumor shape classification model.

Hospital Sant Joan de Reus dataset

It is our private dataset that contains 300 malignant tumors (123 Luminal-A, 107 Luminal-B, 33 Her-2 and 37 Basal-like) with their respective ground truth binary masks obtained by radiologists. The SSD detector and proposed cGAN segmentation model is trained and tested using 220 and 80 images, respectively.

The proposed method was implemented using Python with PyTorch[3] running on a 64-bit Ubuntu operating system using a 3.4 GHz Intel Core-i7 with 16 GB of RAM and NVIDIA GTX 1070 GPU with 8 GB of video RAM.

3.4.1 Tumor detection experiments

In order to localize the tumor in the input mammographies, we compared different common deep learning detectors, such as Dhungel et al. (2017), Kozegar et al. (2013), Faster R-CNN Ren et al. (2015) , YOLO Redmon et al. (2016), and SSD Liu et al. (2016b). The tested detectors were trained with the Hospital Sant Joan de Reus dataset and tested with the INbreast dataset. Table 3.1 presents a quantitative comparison in terms of True Positive Rate (TPR) and False Positive Rate (FPR) with respect to the degree of overlapping between predicted and ground truth bounding boxes containing the tumor. To consider a true positive prediction, we require at least 60% of area overlapping.

[3]https://pytorch.org/

Table 3.1: Mass detection accuracy of proposed method compared with the existing state-of-the-art methods.

Dataset	Method	TPR (%)	FPR(%)
INbreast	Dhungel et al. (2017)	96.00	1.20
	Kozegar et al.(2013)	87.00	3.67
	Faster R-CNN Ren et al. (2015)	96.00	2.94
	YOLO Redmon et al. (2016)	96.35	2.40
	SSD Liu et al. (2016b)	**97.00**	**1.10**

The SSD method yields the best results, with the highest TPR and lowest FPR. In turn, YOLO, Faster R-CNN and Dhungel et al. (2017) models have properly detected masses in the input mammograms, but with slightly worse quantitative results. Consequently, we have chosen the SSD model in order to locate tumors in mammograms.

3.4.2 Tumor segmentation experiments

Evaluation Metrics : Assume A is the ground truth and B is the segmented region (using a segmentation model). True positive (TP), False Positive (FP) and False Negative (FN) rates have been defined above, when introducing equation 3.2. In turn, the true negative rate is defined as $TN = \overline{A} \cap \overline{B}$, which is the area not belonging to any of the two masks A and B. Below, we present the mathematical expression of the accuracy (ACC), Dice coefficient (Dice), IoU (Jaccard Index), sensitivity (SEN), and specificity (SPE).

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \tag{3.4}$$

$$Dice = \frac{2.TP}{2.TP + FP + FN} \tag{3.5}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{3.6}$$

$$SEN = \frac{TP}{TP + FN} \tag{3.7}$$

$$SPE = \frac{TN}{TN + FP} \tag{3.8}$$

The proposed breast tumor segmentation method is compared with the state-of-the-art methods and evaluated both quantitatively and qualitatively. For the quantitative analysis, segmentation accuracy is computed using Dice and IoU. For the qualitative analysis, segmentation results with the their respective ground truth binary masks are compared visually. These experiments have been carried using three different framing of the tumor ROI: full mammogram, loose and tight frames (see Fig. 3.7). The ideal CAD system should be able to automatically segment the breast tumor from a full mammogram. However, this is a very difficult task due to high similarity between gray level pixel distributions of healthy and tumorous tissue. Therefore, removing most of non-ROI portions of the image logically helps the model on learning the visual features that differentiate breast tumor from non-tumor areas. As mentioned in the methodology section, for computing the loose and tight frames we rely on the initial tumor delimitation provided by the SSD method Liu et al. (2016b). The loose frame provides a convenient proportion between healthy and tumorous pixels. The tight frame is a square shrunk on the tumor, as it is intended to evaluate the behavior of the segmentation model when the majority of ROI contains tumor pixels. (see Fig. 3.9).

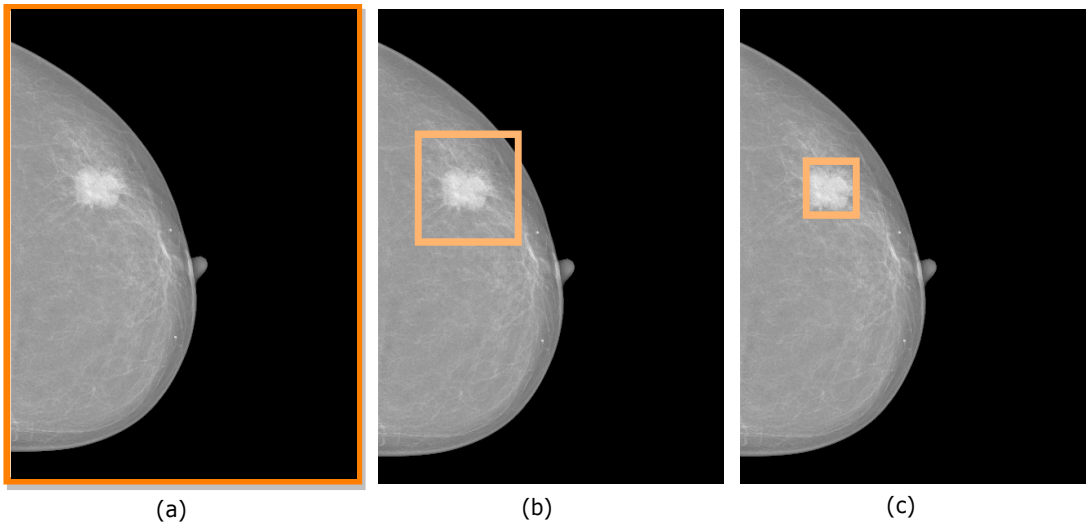


Figure 3.7: Three cropping strategies: (a) full mammogram, (b) loose frame, (c) tight frame.

The three cropping strategies are evaluated on our cGAN and eleven baseline segmentation models, referred as FCN, FCN-ResNet101, UNet, UNet-VGG16,

SegNet, SegNet-VGG16, CRFCNN, SLSDeep, cGAN-ResNet101, cGAN-ResNet101 (Dice Loss) and proposed cGAN (without post-processing). FCN, UNet, SegNet, CRFCNN and proposed cGAN are trained from scratch. FCN-ResNet101, UNet-VGG16, SegNet-VGG16 and cGAN-ResNet101 (with and without Dice loss) are modifications of the original models, where the filters of the starting encoding layers are replaced by the starting convolutional layers of the well-known VGG (16 layers) and ResNet (101 layers) models, which were pre-trained on the ImageNet database. Thus, we loaded the pre-trained weights and fine tuned the network. When using cGAN-ResNet101 Isola et al. (2017), we replaced the $L1$-norm loss with the Dice loss in the generator loss function to see how the base line model will behave under such change. We called this model cGAN-ResNet101 (Dice loss) to compare the segmentation results with our proposal. The results depicted in Table 3.2 are divided in two sections, one for our private dataset and another for the INbreast dataset. Note that all models are trained on the private dataset, and then tested using our private dataset as well as the INbreast dataset without fine tuning.

According to the results, our method outperforms the compared state-of-the-art methods in all cases except for the IoU computed on tight crops in our private dataset. The SLSDeep approach yielded the best IoU (79.93%), whereas our method yielded the second best result (79.87%) with a very small difference of 0.06%. The post-processing improved the results of our model by 1% with the three framing inputs.

All models yielded their worst segmentation results with full mammograms compared to other framing inputs, which is logical taking into account the difficulties stated earlier in this section. Most of the models have obtained their best results for the tight frame crops except for CRFCNN and our proposal, which yielded their best results for loose frame crops. However, the good results for tight crops may be due to the imbalance of tumor/non-tumor pixels, since the former class is present in more than 90% of the image area. The learning can be biased towards this class, which makes rough solutions (almost everything is tumor) to provide very high ranks of performance. Loose frame crops, on the contrary, have a more balanced proportion

Table 3.2: Dice and IoU metrics obtained with the proposed model with/without post-processing and ten alternatives evaluated on the testing sets of our private and INbreast datasets, for the three cropping strategies. Best results are marked in bold. Dashes (-) indicate that results are not reported in referred papers.

Dataset	Methods	Dice(%)			IoU(%)		
		Full	Loose	Tight	Full	Loose	Tight
Private	FCN	59.06	74.94	80.20	39.92	62.21	78.89
	FCN-ResNet101	59.21	77.42	82.78	40.26	68.16	77.32
	UNet	63.69	78.03	83.15	46.73	68.36	78.81
	UNet-VGG16	59.27	78.57	83.71	42.13	69.71	79.42
	SegNet	59.87	80.26	82.33	42.79	70.07	76.17
	SegNet-VGG16	61.59	81.09	81.41	41.61	68.19	77.82
	CRFCNN	53.21	71.33	63.52	41.38	65.24	54.28
	SLSDeep	59.64	71.10	84.28	43.89	60.16	**79.93**
	cGAN-ResNet101	58.37	80.11	86.22	42.12	71.91	76.62
	cGAN-ResNet101 (Dice Loss)	61.49	86.57	86.37	45.90	76.32	77.26
	Proposed cGAN (without post-processing)	65.17	88.42	87.77	48.45	80.67	78.22
	Proposed cGAN (with post-processing)	**66.38**	**89.99**	**88.12**	**49.68**	**81.81**	79.87
INbreast	FCN	54.36	66.12	81.74	36.88	49.38	77.33
	FCN-ResNet101	51.76	83.80	82.38	38.49	74.12	78.09
	UNet	55.58	77.92	80.76	38.46	70.83	77.97
	UNet-VGG16	56.79	78.02	80.89	39.65	68.32	78.13
	SegNet	53.33	79.06	81.11	36.36	65.37	77.02
	SegNet-VGG16	56.27	80.17	81.75	39.46	69.79	78.68
	CRFCNN	52.96	73.25	65.41	40.41	67.14	57.69
	SLSDeep	60.35	75.90	85.53	44.63	65.16	80.26
	cGAN-ResNet101	54.69	87.19	89.17	37.94	77.51	82.26
	cGAN-ResNet101 (Dice Loss)	59.72	88.89	90.42	44.89	82.58	82.95
	Proposed cGAN (without post-processing)	67.55	93.64	91.47	50.05	86.29	83.58
	Proposed cGAN (with post-processing)	**68.69**	**94.07**	92.11	**52.31**	**87.03**	84.55
	Dhungel et al. (2015b)	–	–	90.00	–	–	–
	Cardoso et al. (2017)	–	–	90.00	–	–	–
	Zhu et al. (2018)	–	–	90.97	–	–	–
	Al-antari et al. (2018)	–	–	**92.69**	–	–	**86.37**

of pixels for both classes, which makes them ideal to learn and evaluate the model on a realistic situation: it is more convenient for radiologists to provide a fast frame drawing around the breast tumor rather than a tight frame.

Comparing the general results for both datasets, most methods performed better on INbreast rather than on private dataset with loose and tight framing. This effect can be explained by the fact that INbreast provides more detailed ground truths, which leads to better testing results, despite all network training has been conducted on our private dataset.

In general, our proposal, with and without post-processing, has performed well in terms of both Dice and IoU metrics. For private dataset, in Dice/Loose frame column, our model with post-processing score (89.99%) is almost 9% above the

second best model, SegNet-VGG16 (81.09%). In the IoU/Loose frame column, our model percentage (81.81%) is almost 10% above the second best model, cGAN-ResNet101 (71.91%). For INbreast dataset, our loose frame results for DIC and IoU are again the best (94.07%, 87.03%), where cGAN-ResNet101 is the second best model for both metrics (87.19%, 77.51%). Thus, our model provides an improving of 7% and 10%, respectively. The fact that the second best results are obtained by the cGAN-ResNet101 model indicates that the adversarial network really helps in training the generative network. In turn, the results obtained by the cGAN-ResNet101 (Dice Loss) mixture model are in-between the cGAN-ResNet101 and our proposal, since the Dice loss term substitution improves the accuracy of tumor segmentations.

For the INbreast dataset, we have included the results mentioned in four related papers Dhungel et al. (2015b), Cardoso et al. (2017), Zhu et al. (2018) and Al-antari et al. (2018). For these methods, we could not compute the metrics for all columns, since they have not released their source code. Our method outperformed the first three papers under similar framework conditions. However, Al-antari et al. (2018) yielded better results for Dice (92.69%) and IoU (86.37%) than our model in the Tight frame columns. Nevertheless, our results in the Loose frame columns surpass their results. For a fair comparison, however, it should be checked how the referenced methods would perform on loose frame crops.

The box-plot in Figure 3.8 shows Dice and IoU values obtained for the 106 testing samples from INbreast dataset with loose frames using FCN-ResNet101, Unet-VGG16, SegNet-VGG16, SLSDeep, cGAN-ResNet101 and proposed cGAN. The two models based on cGAN provide small ranges of Dice and IoU values. For instance, the proposed cGAN is in the range 0.89 to 0.93 for Dice and 0.80 to 0.91 for IoU values, while other deep segmentation methods, SLSDeep, Unet-VGG16 and FCN-ResNet101, show a wider range of values. Moreover, there are many outliers in the results for the segmentation based on the cGAN using pre-trained ResNet101 layers, while using our cGAN trained from scratch there are few number of outliers.

The high Dice and IoU metrics obtained by our model empirically support our

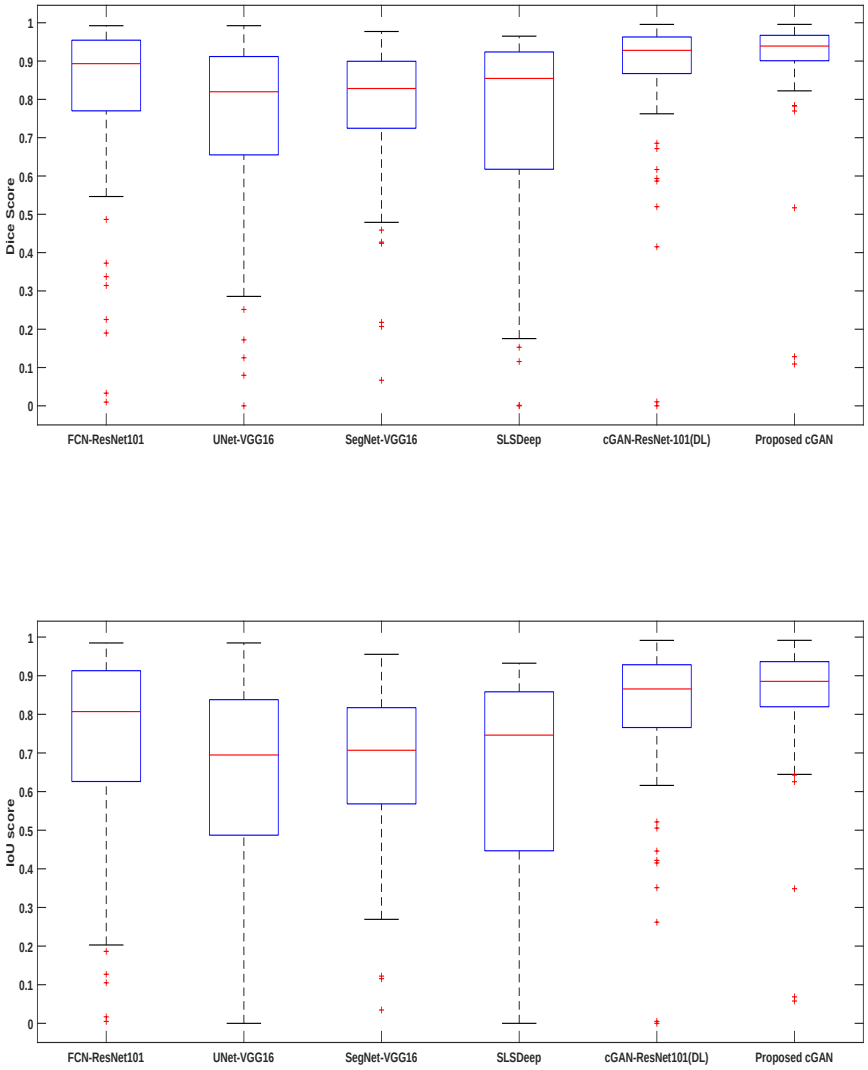


Figure 3.8: Boxplot of Dice (Top) and IoU (Bottom) score over five models compared to our method on loose frames of the test subset of INbreast dataset (106 samples). Blue boxes indicate the interquartile range (Q3-Q1) of the metrics distribution, the red line inside each box represents the median value, the whiskers extend 1.5 times the length of Q1 and Q3, and (+) indicate outlier values, i.e. metrics out of the whiskers.

hypothesis that it achieves accurate tumor segmentation. In Fig. 3.9, we show some examples of our model segmentations using two tumors from the INbreast dataset by applying all three cropping strategies. For each experiment, we show the original ROI image and the comparison of predicted and ground truth mask, color coded to mark up the true positives (TP:yellow), false negatives (FN:red), false positives (FP:green)

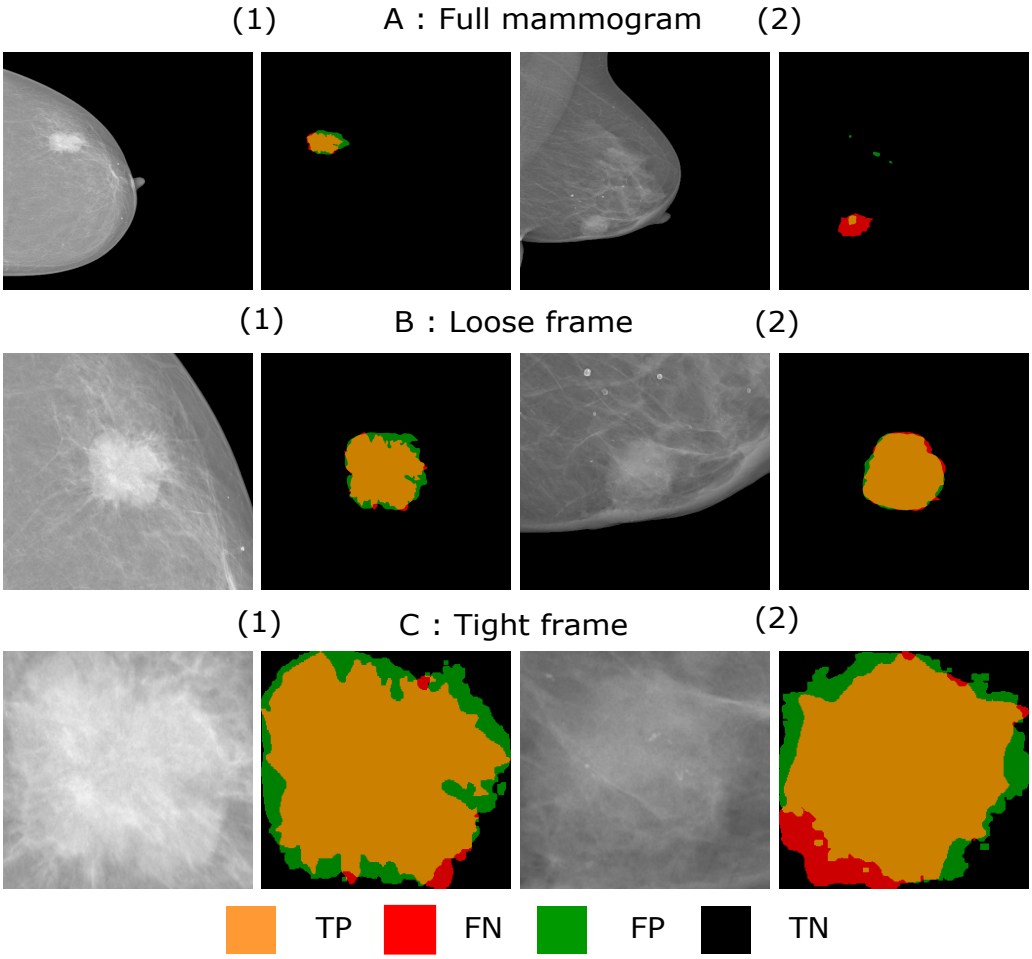


Figure 3.9: Segmentation results of two testing samples extracted from the INbreast dataset with the three cropping strategies.

and true negatives (TN:black). For the full mammogram, the ROI image (1) is an example of good segmentation, since yellow and black pixels depict a high degree of confidence between predicted and real masks. On the contrary, the ROI image (2) is an example of poor segmentation, since red pixels mark up a high portion of the breast tumor area that has been misclassified as healthy area (FN). At the same time, a tiny region of green pixels shows the mis-classification of healthy tissue as breast tumor area (FP). Nevertheless, even in this second segmentation, there is a very high rate of black pixels (TN), which indicates that the model easily recognizes non-tumor areas. In the loose frame segmentations (middle row), specially with example (2), the results contain very few FN and FP pixels. For example (1), a modest amount of green pixels indicate that our model expands the tumor segmentation beyond

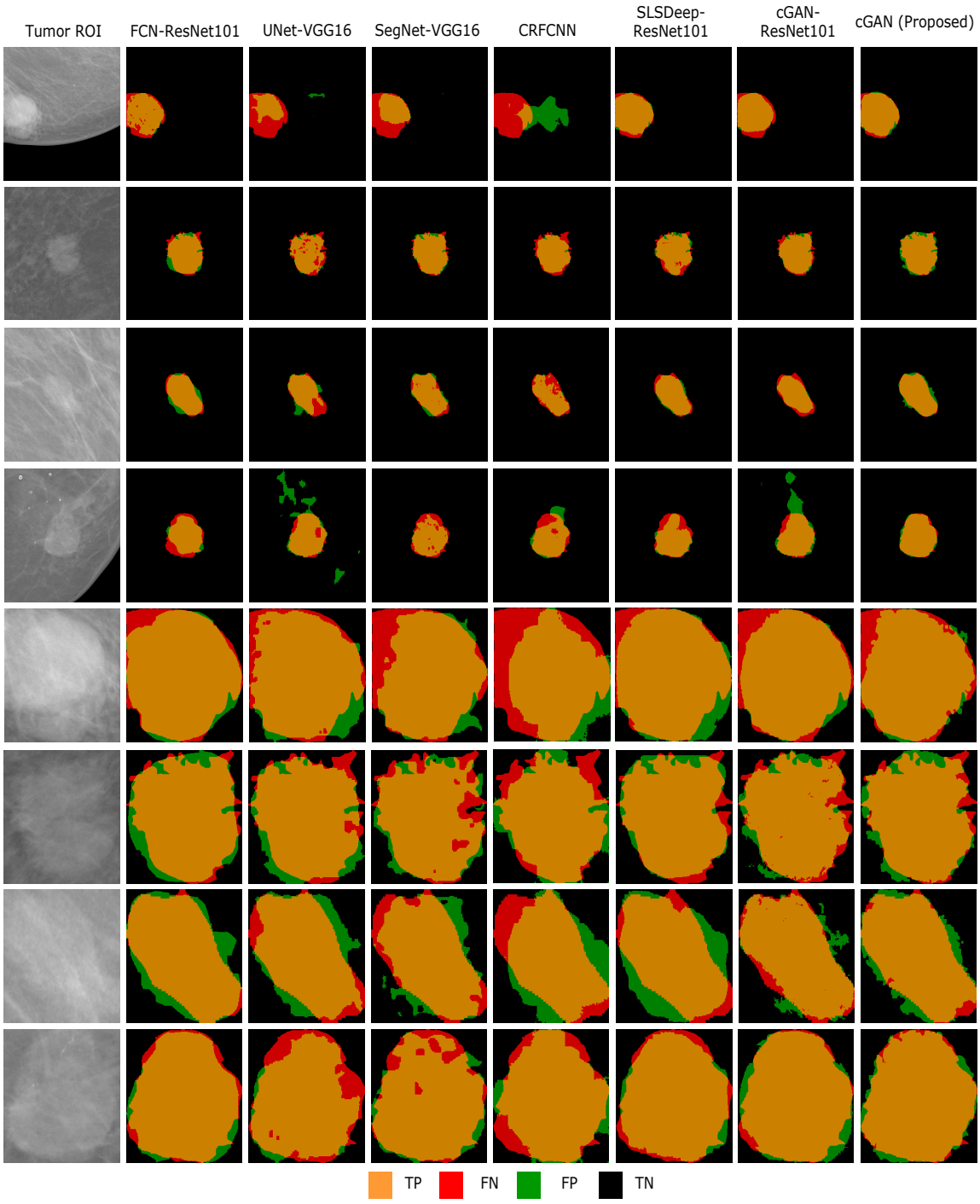


Figure 3.10: Segmentation results of seven models with the INbreast dataset and two cropping strategies: loose frame (the first four rows) and tight frame (the last four rows). (Col 1) original images, (Col 2) FCN-ResNet101, (Col 3) UNet-VGG16, (Col 4) SegNet-VGG16, (Col 5) CRFCNN, (Col 6) SLSDeep, (Col 7) cGAN-ResNet101, and (Col 8) proposed cGAN.

its respective ground truth. In the tight frame crops (bottom row), besides the green areas, our model also has missed some tumor areas *i.e.*, the red pixels (FN). The mistaken areas (red and green) are mostly around the tumor borders, since these areas have a mixture of healthy and unhealthy cells. At the same time, the inner part of the tumor as well as the image regions outside of tumors are properly classified, which indicates the stability of our model.

Fig. 3.10 shows a comparison between our and other six segmentation models, which worked on loose and tight frame crops using four tumors from the INbreast dataset. For the loose frame cases (four top rows), our method clearly outperforms the rest for all tumors except for the second one, where the majority of models provided a similar degree of accuracy. In these four tumors, UNet-VGG16 and CRFCNN provided the worst results. Moreover, cGAN-ResNet101 also performed bad in the fourth example.

For the tight frame cases (four bottom rows), our method also provides the lowest degrees of FN and FP compared to the rest of the models. Our cGAN and the cGAN-ResNet101 model yield irregular borders compared to FCN-ResNet101 and SLSDeep, since GAN models strive for higher accuracy on edges. However, in the third tight frame sample (seventh row), both cGAN-ResNet101 and our proposal generated an irregular border that slightly differs from the smooth ground truth border, which results in lower segmentation accuracy around the edges. Although the rest of the models generate smoother borders, the resulting segmentations may differ from the ground truth significantly.

From the experimental results, it can be concluded that the proposed breast tumor segmentation method is the most effective to date compared to the currently available state-of-the-art methods. However, our method needs a loose crop around the tumor to obtain a proper segmentation, which can be done by the SSD model. Our segmentation model contains about $13,607,043$ parameters for tuning the generator part in the cGAN network. In addition, our method is fast in both training *i.e.*, around 30 seconds per epoch (220 loose frames) and predicting, around 7 images per second. That is 7 to 8 times faster than the segmentation method proposed in

Al-antari et al. (2018) and 10 to 15 times faster than the FCN model.

3.4.3 Shape classification experiments

For validating the tumor shape classification performance, we computed the confusion matrix and the overall classification accuracy on the test set of the DDSM dataset. This set contains 292 images divided into 126, 117, 31 and 18 for irregular, lobular, oval and round classes, respectively.

Table 3.3: Confusion matrix of the tumor shape classification of testing samples of the DDSM dataset.

Prediction / Ground Truth	Irregular	Lobular	Oval	Round	Total
Irregular	96 (76%)	30	0	0	126
Lobular	33	83 (71%)	1	0	117
Oval	0	1	26 (84%)	4	31
Round	0	1	1	16 (89%)	18

However, The DDSM dataset does not have the ground truth binary masks for the breast tumor segmentation. Thus, we applied active contours Akram et al. (2015), which was also used in our work Singh et al. (2018b), to generate the ground truths of the breast tumor regions. Previously, Kisilev et al. (2015) also used active contours Lankton and Tannenbaum (2008) to generate the ground truths in a similar fashion. These ground truth masks are verified by expert radiologists of the hospital of Sant Joan de Reus. In addition, for reliable performance results, we used a stratified 5 fold cross validation with 50 epochs per fold.

In Table 3.3, the proposed method yielded around 73% of classification accuracy for irregular and lobular classes. The relatively high degree of confusion between these two classes is logical, since both shapes have similar irregular boundaries. In turn, our model yielded classification accuracies of 84% and 89% for oval and round shape classes, respectively. For a quantitative comparison, we compared three state-of-the-art tumor shape classification methods Singh et al. (2018b); Kisilev et al. (2015); Kim et al. (2018) with three variations of our shape classification model: one is fed with a binary mask with the ground-truth, the second is fed with a binary

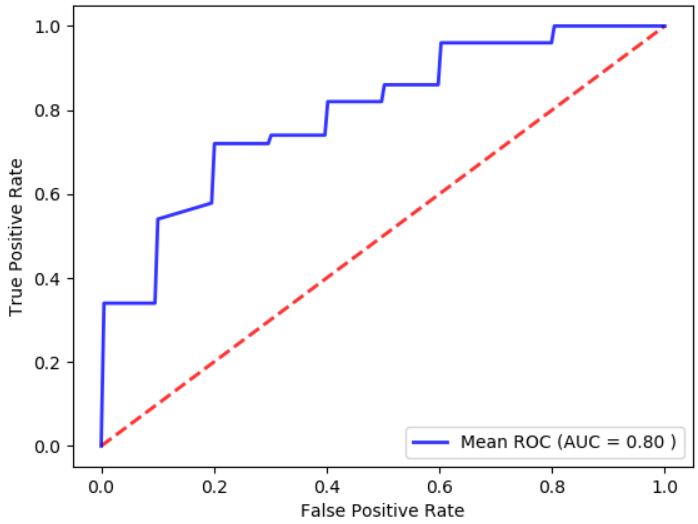


Figure 3.11: Mean ROC curve of 5 folds, for TPR and FPR from shape classification result of 292 test images from DDSM dataset.

mask generated by our segmentation stage and the third is fed with the original ROI image masked with the segmented area (with pixel-wise multiplication). The five methods were evaluated on the DDSM dataset.

We have computed the overall accuracy of each method by averaging the correct predictions (i.e., true positive) of the four classes, weighted with respect to the number of samples per class. As shown in Table 3.4, our classifier based only on binary masks yields an overall accuracy of 80%, outperforming the second best results Kim et al. (2018); Singh et al. (2018b) by 8%. The 83% obtained with our method fed with the original ground truth cannot be considered as a valid result for comparison, since it is the training data accuracy. We provide this result only to show the low degree of overfitting achieved by our network. In turn, the proposed method fed with the masked ROI images provided 70% of overall accuracy. This experiment indicates that gray-level variations inside the segmented area is somehow confusing our shape classification network. In another hand, the multi-task CNN proposed in Kim et al. (2018) based on a pre-trained VGG-16 yielded the worst overall accuracy (66%), probably because the input mammograms are gray-scale images, while the VGG-16 network was trained on color-scale images. In addition, Fig. 3.11 shows

ROC curve illustrating that our model attained area under the curve (AUC) about 0.8.

Furthermore, the proposed shape descriptor contains 767,684 parameters, which can be trained in less than a second per epoch, and predict in about 6 milliseconds per image.

Table 3.4: Shape classification overall accuracy with the DDSM dataset resulting from Kisilev et al. (2015); Kim et al. (2018); Singh et al. (2018b) and our model. Best result is marked in bold.

Methods	Test samples	Accuracy (%)
(SSVM) Kisilev et al. (2015)	515	71
(Multi-task CNN) Kim et al. (2018)	218	66
(ICADx) Kim et al. (2018)	218	72
Singh et al. (2018b)	113	72
Proposed (with ground-truth masks)	292	83
Proposed (generated masks)	292	**80**
Proposed (masked ROI images)	292	70

3.4.4 Shape features correlation to breast cancer molecular subtypes

Table 3.5: Distribution of breast cancer molecular subtypes samples from the hospital dataset with respect to its predicted mask shape.

Shape Classes / Molecular Subtypes	Irregular	Lobular	Oval	Round	Total
Luminal-A	67	29	10	17	123
Luminal-B	58	24	14	11	107
Her-2	6	4	8	15	33
Basal-like	5	10	9	13	37

Tumor shape could play an important role to predict the breast cancer molecular subtypes Tamaki et al. (2011). Thus, we have computed the correlation between breast cancer molecular subtypes classes of our in-house private dataset with the four shape classes. As shown in Table 3.5, most of Luminal-A and -B samples (i.e., 96/123 and 82/107 for Luminal-A and -B, respectively) are mostly assigned to irregular and lobular shape classes. In turn, oval and round tumors give indications to the Her-2 and Basal-like samples, (i.e., 23/33 and 22/37 for Her-2 and Basal-like,

respectively). Moreover, some images related to Basal-like are moderately assigned to the lobular class. Afterwards, from the visual inspection, if the tumor shape is irregular or lobular then radiologist can suspect that it belongs to the Luminal group. In turn, if the tumor shape is round or oval then it is more probable that the tumor is a Her-2 or Basal-like Tamaki et al. (2011). Therefore, this study shows the importance of tumor shape, which can be considered as a key feature to distinguish between different malignancies of breast cancer.

3.4.5 Limitations

For the segmentation stage, our model has only one significant limitation. If there are two tumors (i.e., one is with complete shape and the other is incomplete) in the loose framing, the proposed segmentation methods will be able to properly segment the complete tumor and it will fail to segment the incomplete one.

As shown in Fig. 3.12, we found three samples that are mis-segmented because they contained two tumors, the one in the center, which is properly segmented, and another that is shown partially in the left-down border of the image, which is wrongly ignored as non-tumor region (FN). When the bigger tumor is located in the center of the crop, nevertheless, it is correctly segmented. To classify the tumor

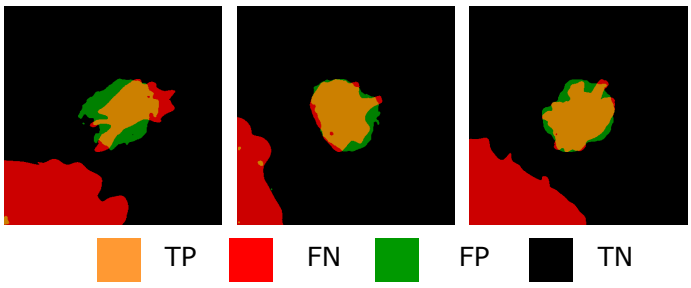


Figure 3.12: Three mis-segmented tumor of non-full tumor shapes with INbreast dataset. The red part in the down-left border.

shape, we depend only on the DDSM dataset to train our model, since it is the only public dataset that has the shape classification information. Thus, more databases containing more samples are required to improve the classification accuracy of four shape classes.

To study the molecular subtypes of breast cancer, Her-2 and Basal-like classes have less samples compared to the other two classes, Luminal-A and Luminal-B. Indeed, we used a weighted loss function to train our shape classification model in order to make a balance between the four classes. However, we anticipate that, by increasing the samples related to the Her-2 and Basal-like classes, we will improve the prediction of molecular subtypes from tumor shape information.

3.5 Conclusion

In this chapter, we propose a two stage breast tumor segmentation and classification method, which first segments the breast tumor ROI using a cGAN and then classify its binary mask using a CNN based shape descriptor.

The segmentation results reveal the importance of the adversarial network in the optimization of the generative network. cGAN-ResNet101 shows an improvement of about 1% to 3% in both Dice and IoU metrics in comparison to the other non-GAN methods. In turn, the proposed method yields an increment of about 10% over the results of cGAN-ResNet101 by training our model from scratch, and replacing the $L1$-norm with the Dice loss using loose frame crop on the given datasets. The breast tumor segmentation from full-mammograms yields low segmentation accuracy for all models including the proposed cGAN. For the tight frame crop, the proposed cGAN yields similar or better segmentation accuracy compared to the other methods.

The classification results show that our second stage properly infers the tumor shape from the binary mask of the breast tumor, which was obtained from the first stage (cGAN segmentation). Hence, we have empirically shown that our CNN is focusing its learning on the morphological structure of the breast tumor, while the rest of approaches (Kisilev et al. (2015), Kim et al. (2018), Kisilev et al. (2016), Ren et al. (2015)) rely on the original pixel variations of the input mammogram to make the same inference. Moreover, in Al-antari et al. (2018) they used a hybrid strategy in which they include the pixel variability within the mask of breast tumor region to retain the intensity and texture information. However, the higher performance

obtained by our method supports our initial idea that the second stage CNN can reliably recognize the tumor shape based only on morphological information.

Furthermore, this chapter provided a study of correlation between the tumor shape and the molecular subtypes of the breast cancer. Most samples of the Luminal-A and -B group are assigned to irregular shapes. In turn, the majority of Her-2 and Basal-like samples are assigned to regular shapes (e.g., oval and round shapes). That gives an indication that the tumor shape can be considered for inferring the molecular subtype of the tumor.

CHAPTER 4

Breast Tumor Segmentation and Classification in Ultrasound Images using Adversarial Learning

This fourth chapter presents a method of adversarial training to segment breast lesions in ultrasound images. Also, to show the efficacy of the proposed segmentation model, statistical features has been extracted from the generated mask in order to perform the classification of the lesions into benign and malignant.

4.1 Introduction and related work

Ultrasound has been recommended as a powerful adjunct screening tool for detecting breast cancers that may be occluded in mammographies Lauby-Secretan et al. (2015)

(e.g., in the case of dense breasts). CAD systems are widely used to detect, segment and classify masses in breast ultrasound (BUS) images. One of the main steps of BUS CAD systems is tumor segmentation.

Over the last two decades, several BUS image segmentation methods have been proposed, which can be categorized into semi-automated and fully automated according to the degree of human intervention. In Abdel-Nasser et al. (2017), a region growing based algorithm was used to automatically extract the regions that contain the tumors, and image super-resolution and texture analysis methods were used to discriminate benign tumors from the malignant ones. Recently, some deep learning based models have been proposed to improve the performance of breast tumor segmentation methods. In Xu et al. (2019), two CNN architectures have been used to segment BUS images into the skin, mass, fibro-glandular, and fatty tissues (90% of accuracy). Hu et al Hu et al. (2019) combined a dilated FCN with a phase-based active contour model to segment breast tumors, achieving a Dice score of 88.97%.

Although these methods and others proposed in the literature do provide useful techniques, there are still challenges due to the high degree of speckle noise present in the ultrasound images, as well as to the high variability of tumors in shape, size, appearance, texture, and location. In this chapter, we propose an efficient solution for breast tumor segmentation and classification in BUS images using deep adversarial learning.

The main contribution of this chapter is to develop an efficient deep model for segmenting the breast tumor in BUS by combining an atrous convolution network (AC) and channel attention with channel weighting (CAW) in a cGAN model in order to enhance the discriminability of feature representations at multi-scale. Besides, we demonstrate that the proposed segmentation model can be used for characterizing accurate shape features from the segmented mask to distinguish between benign and malignant tumors.

4.2 Proposed methodology

4.2.1 Integration of channel attention and channel weighting (CAW) block

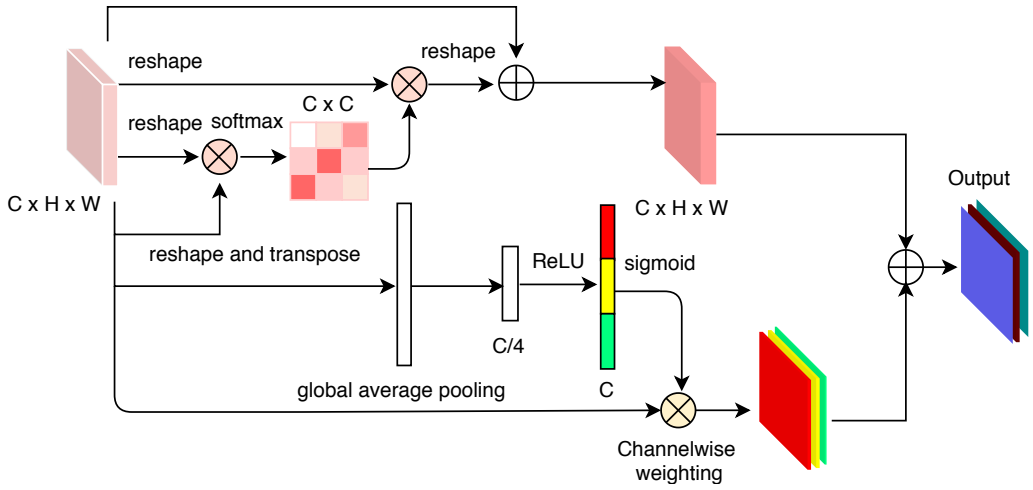


Figure 4.1: The proposed integration of channel attention and channel weighting module

Fig. 4.1 represents the two processes performed by this block, the channel attention process Fu et al. (2018b) (top branch) and the channel weighting process Hu et al. (2018) (bottom branch). Since we have placed our CAW block after the last encoder layer, the processed activation map has spatial dimensions ($H \times W$) of 1×1: indeed, it is a vector of C= 512 scalars. Hence, the method works only on the channel feature space.

The attention mechanism computes a feature correlation matrix of $C \times C$ elements, as the multiplication of the input vector by its transposed. Then, the input vector is multiplied by the transposed of this matrix, in order to enhance the relevance of features that show similar values for a given image. The enhanced version of the vector is then multiplied by a learnable scalar parameter and summed to the original vector.

We define the channel attention map $X \in R^{C \times C}$ from the original features $\gamma \in R^{C \times H \times W}$. Specifically, we reshape γ to $\mathbb{R}^{C \times N}$, then perform a matrix multiplication between γ and the transpose of γ. Finally, we apply a softmax layer to make the

channel attention map $X \in \mathbb{R}^{C\times C}$:

$$x_{ji} = \frac{exp(\gamma_i.\gamma_j)}{\sum_{i=1}^{C} exp(\gamma_i.\gamma_j)} \tag{4.1}$$

where x_{ji} estimate the i^{th} channels impact on the j^{th} channel. Moreover, we implement a matrix multiplication between the transpose of X and γ and reshape their result to $\mathbb{R}^{C\times H\times W}$. Later we multiply the result by a scale parameter β and do an element-wise sum operation with γ to get the final output $E \in \mathbb{R}^{C\times H\times W}$:

$$E = \alpha\sum_{i=1}^{C}(x_{ji}\gamma_i) + \gamma_j \tag{4.2}$$

where α is weight factor.

In channel weighting, the weighting mechanism starts with a global average pooling to transform each channel map into a single value (squeeze), but we can omit this step since we already have one value per channel. The next step is based on two fully connected layers, the first one with C/4 neurons and the second one with C neurons, which learn to output C weights (one per channel), which multiply the original vector values in order to dynamically re-calibrate the importance of the features for each sample (excitation).

At the end, the output vectors of the two branches are summed. With this structure, in the training phase the back-propagated gradient is allowed to flow through both branches, which both are intended to boost the features that are more relevant to the final segmentation: the attention method promotes the high-level features that repeat their values for given input image patterns, which may indicate that these shared features are core for the target classes; the excitation method looks for the optimal non-linear re-calibration of the high-level features that tends to provide better inferences of the final output. In this manner, we have increased the representational power of the highest level features of the generator network, which turns out in a clear improvement from baseline (BL) model and achieved a Dice and IoU scores about 9% and 11% respectively of the breast tumor segmentations in ultrasound images.

4.2.2 Network architecture

The proposed BUS image segmentation technique is based on generative adversarial training, which involves two interdependent networks: a generator G and a discriminator D. (Fig. 4.2). The generator generates a fake example from input noise z, while discriminator determines the probability that the fake example is from training data rather than generated by the generator.

Generator: The generator network incorporates an encoder section, made of seven convolutional layers (En1 to En7), and a decoder section, made of seven deconvolutional layers (Dn1 to Dn7) layers. We have modified the plain encoder-decoder structure by inserting an atrous convolution block Yu and Koltun (2015) between En3 and En4, in addition to a CAW block between En7 and Dn1. The CAW block is an aggregation of a channel attention module Fu et al. (2018b) with channel weighting block Hu et al. (2018). In turn, the CAW block increases the representational power of the highest level features of the generator network, which turns out in a clear improvement of the accuracy of the breast tumor segmentation in ultrasound images.

By including the atrous convolutional block in-between encoder layers En3 and

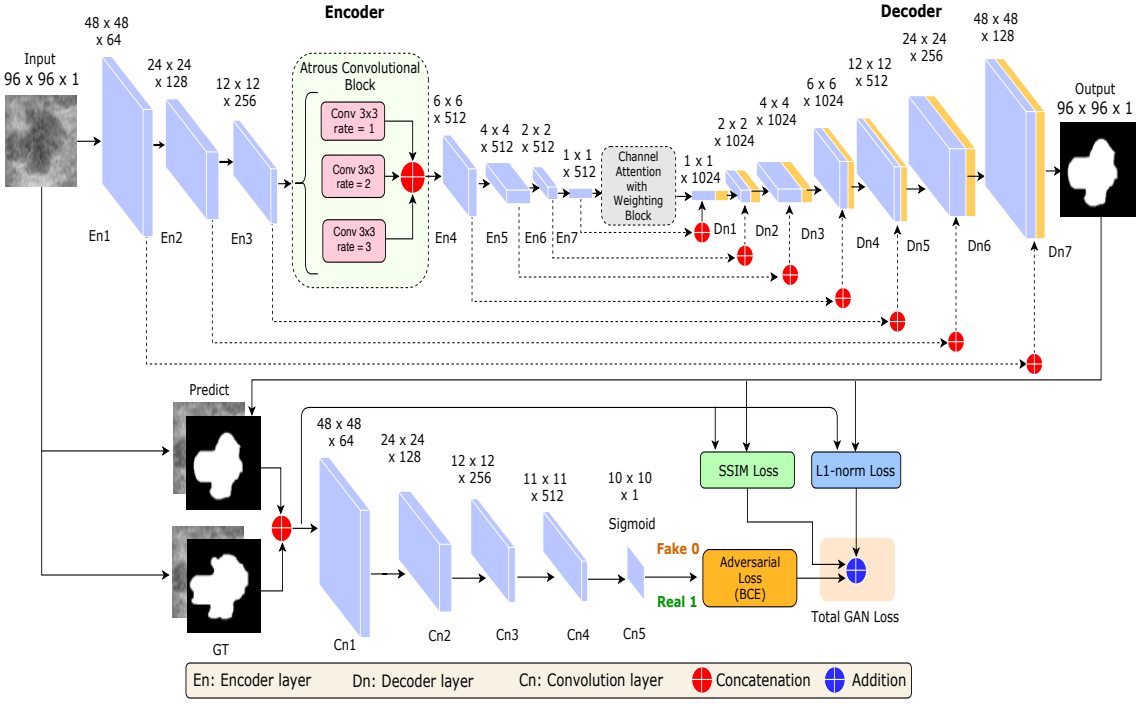


Figure 4.2: The architecture of the proposed segmentation model for BUS images.

En4, the generator network is enabled to characterize features at different scales and also to expand the actual receptive field of the filters. As a consequence, the network is more aware of contextual information without increasing the number of parameters or the amount of computation. Fig 4.3 presents the different atrous convolutional network rates. The atrous convolution helps to manage the resolution of the feature responses computed from the CNN. To incorporate context, without increasing parameters, it also helps to enlarge the field of view of filters and find the best trade-off between small and large field-of-view.

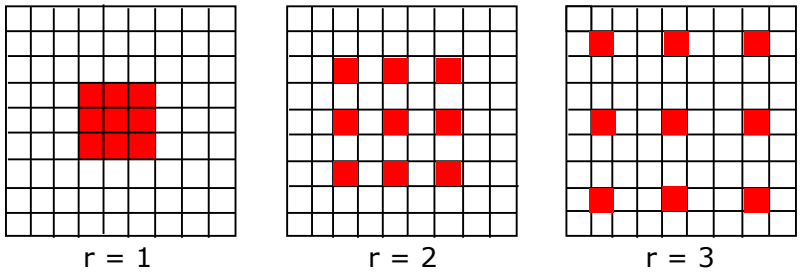


Figure 4.3: The different rates of atrous convolution (r = 1, 2 and 3).

Each layer in the encoder section is followed by batch normalization (except for En1 and En7) and *LeakyReLU* with slope 0.2, except for En7, where the regular non-linearity *ReLU* activation function is used. The decoder section is a sequence of transposed-convolutional layers followed by batch normalization, dropout with rate 0.5 (only in Dn1, Dn2, and Dn3) and *ReLU*. The filters of the convolutional and deconvolutional layers are defined by a kernel of 4×4 and they are shifted with a stride of 2. We add padding of 2 after En4, yielding a $4 \times 4 \times 512$ output feature map. We also add skip connection between the corresponding layers in the encoder and decoder sections, which improve the features in the output image by merging deep, coarse, semantic information and simple, fine, appearance information. After the last decoding layer (Dn7), the *Tanh* activation function is used as a non-linear output of the generator, which is trained to generate a binary mask of the breast tumor.

Discriminator: It is a sequence of convolutional layers applying kernels of size 4×4 with a stride of 2, except for C_{n4} and C_{n5} where the stride is 1. Batch normalization is employed after Cn2 to Cn4. *LeakyReLU* with slope 0.2 is the

non-linear activation function used after Cn1 to Cn4, while the sigmoid function is used after Cn5. The input of the discriminator is the concatenation of the BUS image and a binary mask marking the tumor area, where the mask can either be the ground truth or the one predicted by the generator network. The output of the discriminator is a 10×10 matrix having values varying from 0.0 (completely fake) to 1.0 (real).

Loss Functions: Assume x is a BUS image containing a breast tumor, y is the ground truth mask of that tumor within the image, $G(x, z)$ and $D(x, G(x, z))$ are the outputs of the generator and the discriminator, respectively. The loss function of the generator G comprises three terms: adversarial loss (binary cross entropy loss), L1-norm to boost the learning process, and SSIM loss Wang et al. (2004) to improve the shape of the boundaries of segmented masks:

$$\begin{aligned} \ell_{Gen}(G, D) = \mathbb{E}_{x,y,z}(-\log(D(x, G(x, z)))) + \\ \lambda \mathbb{E}_{x,y,z}(\ell_{L1}(y, G(x, z))) + \alpha \mathbb{E}_{x,y,z}(\ell_{SSIM}(y, G(x, z))) \end{aligned} \tag{4.3}$$

where z is a random variable and λ and α are empirical weighting factors. The variable z is introduced as a dropout in the decoding layers $Dn1$, $Dn2$ and $Dn3$ at both training and testing phases, which helps to generalize the learning processes and avoid overfitting. If the generator network is properly optimized, the values of $D(x, G(x, z))$ should approach 1.0, meaning that discriminator cannot distinguish generated tumor masks from ground truth masks, while L1 and SSIM losses should approach to 0.0, indicating that every generated mask matches the corresponding ground truth both in overall pixel-to-pixel distances (L1) and in basic statistic descriptors (SSIM).

The loss function of the discriminator D can be formulated as follows:

$$\ell_{Dis}(G, D) = \mathbb{E}_{x,y,z}(-\log(D(x, y))) + \mathbb{E}_{x,y,z}(-\log(1 - D(x, G(x, z)))) \tag{4.4}$$

The optimizer will fit D to maximize the loss values for ground truth masks (by minimizing $-\log(D(x, y))$) and minimize the loss values for generated masks (by

minimizing $-\log(1-D(x,G(x,z)))$. These two terms compute BCE loss using both masks, assuming that the expected class for ground truth and generated masks are 1 and 0, respectively. G and D networks are optimized concurrently: one optimization step for both networks at each iteration, where G tries to generate a valid tumor segmentation and D learns how to differentiate between the synthetic and real segmentation.

Model training: In the preprocessing step, each BUS images is rescaled to 96×96 pixels, and pixel values are normalized between [0,1]. In the postprocessing step, morphological operations (3×3 closing, 2×2 erosion) are used to suppress most of the outlier predictions (speckled pixels). The hyperparameters of the model were experimentally tuned. We also explored several optimizers, such as SGD, AdaGrad, Adadelta, RMSProp, and Adam with different learning rates. We achieved the best results with Adam optimizer (β_1= 0.5, β_2= 0.999) and learning rate = 0.0002 with a batch size of 8. The SSIM loss and L1-norm loss weighting factors λ and α were set to 10 and 5, respectively. The best results were achieved by training both generator and discriminator from scratch for 40 epochs.

4.2.3 Breast tumor classification

To classify the breast tumor into benign and malignant, we propose to rely on statistic features of the segmented tumor mask. Malignant breast tumors and benign lesions have different shape characteristics: the malignant lesion usually is irregular, speculated or microlobulated. However, benign lesion mainly has smooth boundaries, round, oval or macrolobulated shape Yang et al. (2008).

In Fig 4.4, each BUS image is fed into the trained generative network to obtain the boundary of the tumor, and then we compute 13 statistical features from that boundary: fractal dimension, lacunarity, convex hull, convexity, circularity, area, perimeter, centroid, minor and major axis length, smoothness, Hu moments (6 values) and central moments (order 3 and below). We implemented an *Exhaustive Feature Selection* (EFS) algorithm to select the best set of features. The EFS algorithm indicates that the fractal dimension, lacunarity, convex hull, and centroid

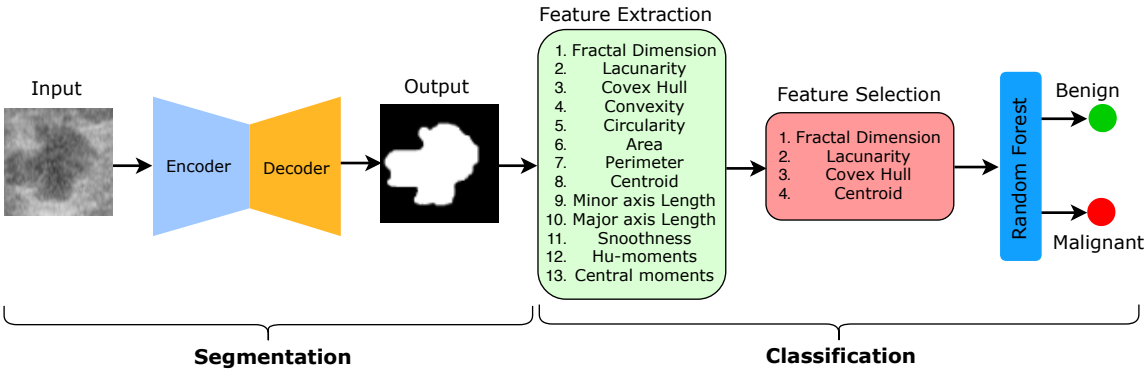


Figure 4.4: Proposed method for Breast tumor classification.

are the 4 optimal features. The selected features are fed into a Random Forest classifier, which is later trained to discriminate between benign and malignant tumors.

4.3 Experiments and discussion

BUS dataset: We evaluated the performance of the proposed model using the Mendeley Data BUS dataset, which is publicly available Rodrigues (2017). This dataset contains 150 malignant and 100 benign tumors contained in BUS images. To train our model, we randomly divided the dataset into the training set (70%), a validation set (10%) and testing set (20%). The dataset does not have a ground truth for tumor segmentation. Thus, cooperative experts have manually segmented the tumors appearing in the BUS images.

Data augmentation: To augment the current set of available examples, we applied different operations: 1) scale the images by factors varying from 0.5 to 2 with steps of 0.25, 2) apply gamma correction on the BUS images by factors varying from 0.5 to 2.5 with steps of 0.5, and 3) flip and rotate the images. These operations yield 8K BUS images.

4.3.1 Breast tumor segmentation results

In Table 4.1, we compare the baseline cGAN model Isola et al. (2017) with three variations of our model: cGAN with atrous convolution (cGAN+AC), cGAN

with channel attention and weighting (cGAN+CAW), and cGAN with AC and CAW (cGAN+AC+CAW). All of these variations are also compared with five state-of-the-art image segmentation methods: FCN Long et al. (2015), UNet Ronneberger et al. (2015) SegNetBadrinarayanan et al. (2017), ERFNet Romera et al. (2018), and Deep Convolutional Generative Adversarial Network (DCGAN) Kim et al. (2017). All methods are evaluated both quantitatively and qualitatively. For the quantitative analysis, we calculate the ACC, Dice, IoU, SEN and SPE metrics.

As shown in Table 4.1, the added AC and CAW blocks improves the results of the baseline cGAN model. In addition, our model (cGAN+AC+CAW) outperforms the rest in all metrics. It achieves Dice and IoU scores of 93.76% and 88.82%, respectively, which are the metrics that better represent the degree of coincidence between predicted and ground truth segmentation. These two results outperform the ones from the second best model in the table, the UNet model, in 5% to 6% absolute points over the full range, which is quite significant taking into account their proximity to the maximum value. The SegNet and ERFNet models yield the worst segmentation results on BUS images. The results of the proposed model have also been compared with other methods evaluated on different datasets. For instance, Hu et al. (2019) yielded an IoU of 85.10% on a private BUS image dataset. In addition, Xu et al. (2019) achieved a Dice score of 89.00%.

Table 4.1: Segmentation results of the proposed model(cGAN+AC+CAW) and compared models FCN Long et al. (2015), SegNetBadrinarayanan et al. (2017), UNet Ronneberger et al. (2015), ERFNet Romera et al. (2018), DCGAN Kim et al. (2017) and cGAN Isola et al. (2017).

Methods	Dice(%)	IoU(%)
FCN	79.73 ± 0.102	66.29 ± 0.116
SegNet	50.35 ± 0.295	41.65 ± 0.274
UNet	88.28 ± 0.090	82.23 ± 0.096
ERFNet	67.02 ± 0.206	53.32 ± 0.185
DCGAN	85.55 ± 0.154	75.28 ± 0.128
cGAN	86.04 ± 0.095	77.56 ± 0.080
cGAN + AC	87.14 ± 0.084	80.77 ± 0.081
cGAN + CAW	90.65 ± 0.075	83.40 ± 0.069
Ours	$\mathbf{93.76 \pm 0.037}$	$\mathbf{88.82 \pm 0.064}$

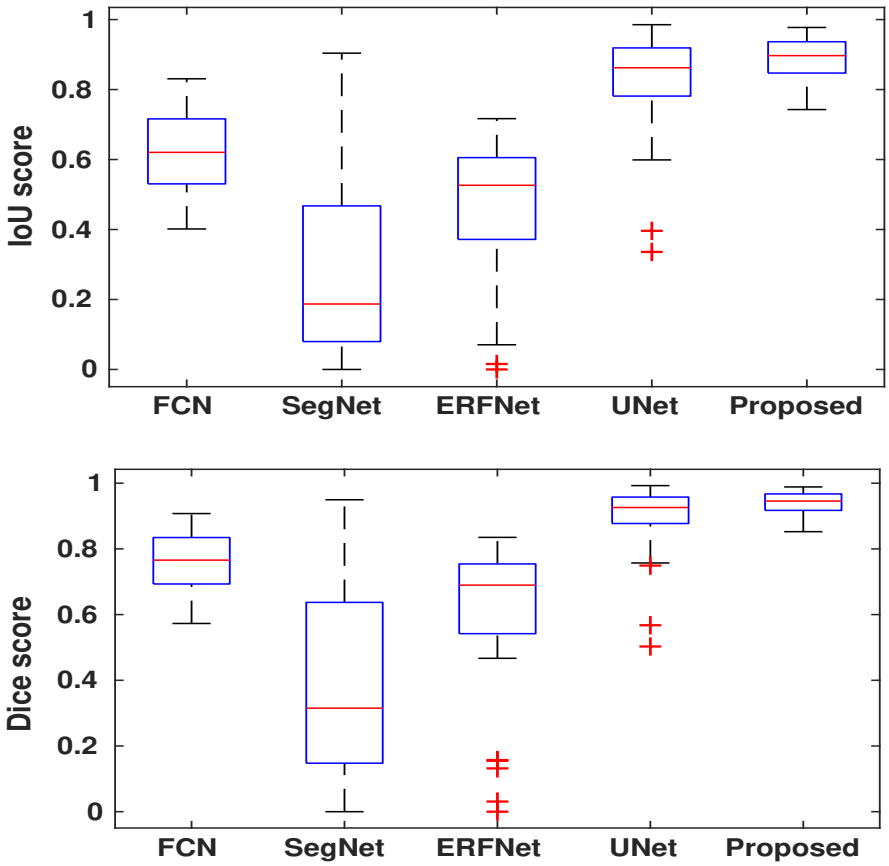


Figure 4.5: Boxplots of IoU and Dice metrics of the proposed model and FCN Long et al. (2015), SegNetBadrinarayanan et al. (2017), ERFNet Romera et al. (2018) and UNet Ronneberger et al. (2015).

Figure 4.5 shows boxplots of Dice and IoU values obtained for the 50 testing samples using FCN, SegNet, ERFNet, UNet and the proposed model. Our model based on cGAN provided the smallest range and highest median of Dice and IoU values. For instance, the mid-half of results (quartiles Q2 and Q3) output by our model is in the range 88% to 94% for Dice and 80% to 89% for IoU, while FCN, SegNet, ERFNet and UNet show less median values and wider ranges of values. Moreover, there are many outliers (results far below the Q1 threshold) in ERFNet and UNet methods, while using our proposal there are none.

Fig 4.6 presents the AUC of the receiver operating characteristic (ROC) curve of different baseline segmentation models and our model with the BUS image dataset. The proposed model gives the best AUC value (97.35%). Figure 4.7 presents a comparison between tumor segmentations obtained with our model and other six models. As shown, SegNet and ERFNet yield the worst results since there are large

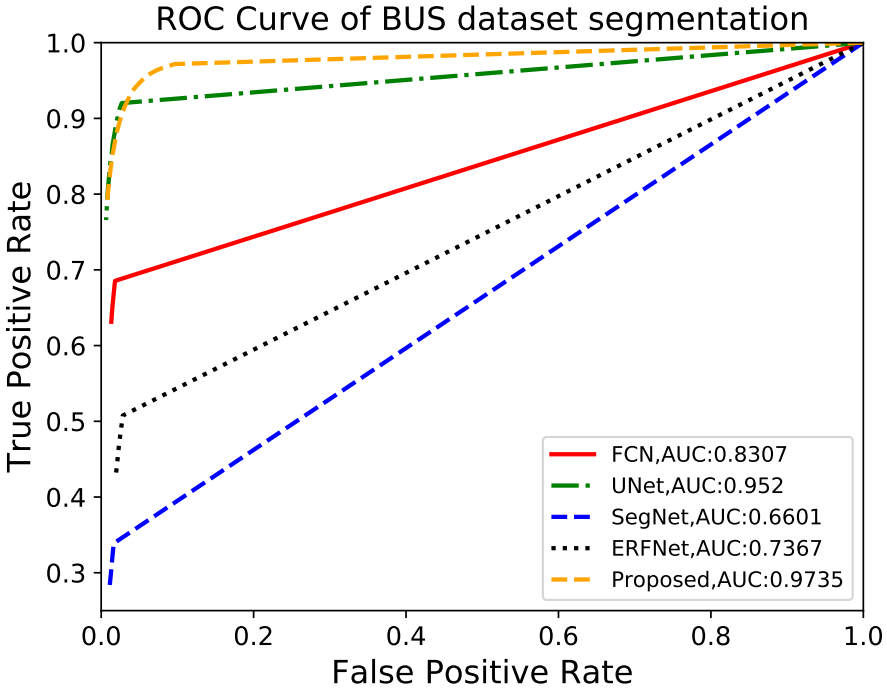


Figure 4.6: The ROC curve of the segmentation models.

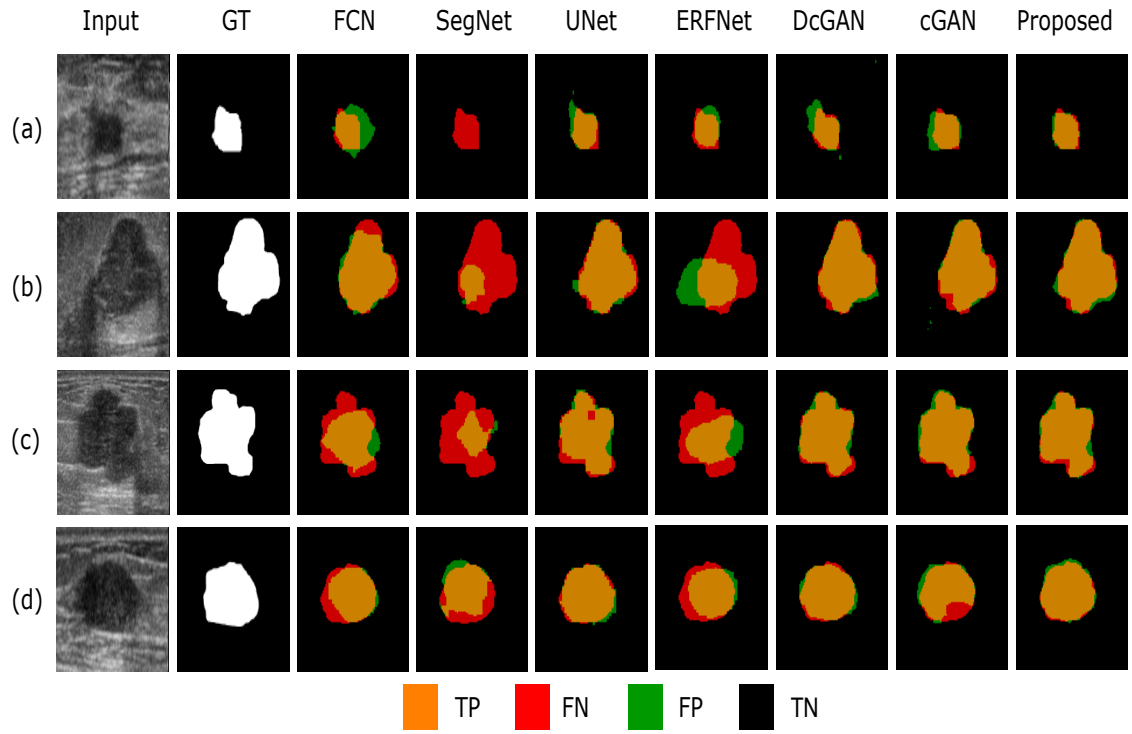


Figure 4.7: Segmentation results on four samples of the BUS dataset. The rows (a) and (b) show benign samples while rows (c) and (d) rows show malignant samples.

false negative areas (in red), as well as some false positive areas (in green). FCN also shows rather significant erroneous areas, although it has fairly segmented the second example (b). In turn, UNet, DCGAN, cGAN provide good segmentation but our model is more accurate in the boundary of breast tumors.

Effect of loss functions and optimizers

Fig 4.8 presents the performance of the proposed model with different combinations of loss functions: BCE, BCE+L1-norm, BCE+SSIM, BCE+Lovasz Hinge and BCE+L1-norm+SSIM loss, achieving the best results with the later combination. We choose this combination because L1-norm loss helps the generator network to provide the sharpen image. In turn, SSIM loss enforces the generator to efficiently learn shape information of BUS image and to provide more accurate segmentation.

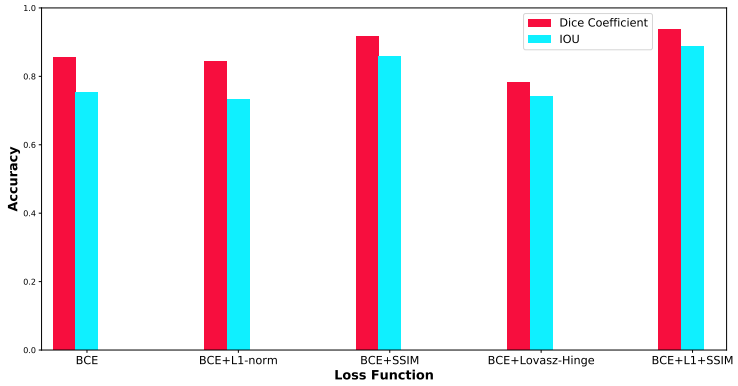


Figure 4.8: The performance of the proposed model with different combinations of loss functions.

Figure 4.9: The Dice (left) and IoU (right) scores of our model with four optimizers: SGD, RMSProp, Adam and Adadelta.

Table 4.2: Breast tumor classification results

Methods	Precision	Recall	Accuracy	F1-Measure
FCN	62.0 ± 0.138	70.0 ± 0.154	69.0 ± 0.128	68.0 ± 0.135
SegNet	51.0 ± 0.216	62.0 ± 0.238	53.0 ± 0.200	51.0 ± 0.229
UNet	70.0 ± 0.098	81.0 ± 0.104	77.0 ± 0.091	77.0 ± 0.087
ERFNet	58.0 ± 0.162	66.0 ± 0.205	61.0 ± 0.146	59.0 ± 0.172
DCGAN	71.0 ± 0.098	82.0 ± 0.081	75.0 ± 0.105	73.0 ± 0.107
cGAN	71.0 ± 0.083	84.0 ± 0.095	78.0 ± 0.079	77.0 ± 0.776
cGAN+AC	73.0 ± 0.060	87.0 ± 0.068	80.0 ± 0.053	81.0 ± 0.059
cGAN+CAW	74.0 ± 0.052	88.0 ± 0.070	82.0 ± 0.056	82.0 ± 0.054
Lee et al. (2018)	78.0	90.0	83.0	83.0
Ours	$\mathbf{81.0 \pm 0.021}$	$\mathbf{92.0 \pm 0.028}$	$\mathbf{85.0 \pm 0.196}$	$\mathbf{84.0 \pm 0.024}$

Fig 4.9 shows the Dice and IoU scores of our model with four optimizers: SGD, RMSProp, Adam and Adadelta, finding that Adam optimizer yielded the best results.

4.3.2 Breast tumor classification results

We have checked our classification strategy (random forest over four optimal tumor shape statistics) with different segmentation methods' output with the leave-one-out cross-validation technique and calculated the precision, recall, accuracy and F1-score metrics. Furthermore, we have also obtained the same metrics from the work of Lee et al. (2018), who proposed a stack denoising autoencoder method to segment and classify breast tumors from the same BUS dataset that we use in this study. As shown in Table 4.2, the proposed breast tumor classification method outperforms Lee et al. (2018), with a total accuracy degree of 85%.

4.4 Conclusion

In this chapter, we have proposed an efficient solution for tumor segmentation and classification in BUS images. We have proposed to add an atrous convolution blocks to the generator network to learn tumor features at different resolutions of BUS images. We also have used a channel-wise weighting block in the generator network to automatically re-balance the relative impact of each of the highest level encoded

features. Our model outperforms the FCN, SegNet, ERFNet, UNet, DCGAN and cGAN segmentation models in terms of Dice and IoU metrics, achieving the top scores of 93.76% and 88.82% respectively. In the classification stage, we used four optimal statistics features extracted from the segmented tumor masks, obtaining an accuracy of 85%, which is 2% over a state-of-the-art related method that uses the same database.

CHAPTER 5

Applying Adversarial Network to Retinal Optic Disc Segmentation

Segmenting the optic disc is a crucial step in designing a structure of reference for diagnosing optic nerve severe problems such as glaucoma. This chapter presents an application of our devised adversarial network from chapter three to segment the optic disc from the retinal fundus image.

5.1 Introduction

Retinal fundus image analysis is very important for doctors to deal with the medical diagnosis, screening and treatment of opthalmologic diseases. The morphology of the optic disc (OD), which is the location where ganglion cell axons exit the eye to form the optic nerve, in which visual information of the photo-receptors is transmitted

to the brain, is an important structural indicator for assessing the presence and severity of retinal diseases, such as diabetic retinopathy, hypertension, glaucoma, hemorrhages, vein occlusion, and neovascularization MacGillivray et al. (2014). Retinal OD segmentation is the first step for a significant investigation of retinal images Almazroa et al. (2015).

The OD appears as a bright yellowish oval region within color fundus images through which the blood vessels enter the eye. The macula is the center of the retina, which is responsible for our central vision. Figure 5.1 shows a sample of a color retinal fundus image with the key anatomical structures denoted. For ophthalmologists and eye care specialists, an automated segmentation and analysis of fundus optic disc plays an important role to diagnose and treat the retinal diseases. Numerous

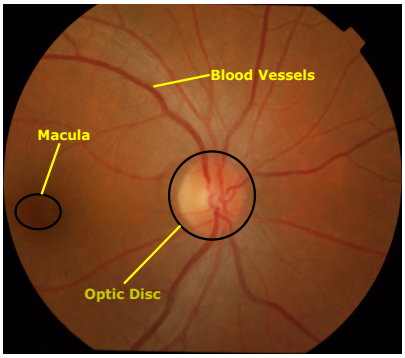


Figure 5.1: Structures in a fundus image.

methods has been proposed to detect and segment the optic disc. For the diagnosis of glaucoma, Chrástek et al. (2005) proposed an automated segmentation algorithm to segment the optic nerve head. They firstly removed the blood vessel by using a distance map algorithm and a morphological operation, and then used an anchored active contour model to segment the optic disc. Lowell et al. (2004) proposed a deformable contour model to segment the optic nerve head boundary of retinal images by using a template matching and a directionally sensitive gradient to discard the interference of vessels. In turn, Welfer et al. (2010) proposed an automated OD segmentation in a fundus image using an adaptive morphological operation. They then used a watershed transform marker to define the OD boundary. In addition, the vessel obstruction is minimized by morphological erosion.

With the increase of using deep learning models in segmentation tasks, many methods have recently been proposed based on CNN. An automatic optic disc and cup image segmentation has been proposed in Al-Bander et al. (2018) based on a stack of deep UNet models. Each model in a cascade refines the result of the previous one.

In this chapter, we show an application of cGAN model proposed in chapter 3 to perform retinal OD segmentation. To the best of our knowledge, this is the first application of a conditional generative adversarial training for retinal optical disc segmentation. The proposed cGAN network consists of two combined networks: generator and discriminator. The generator network learns the mapping from the input, a fundus image, to the output, a segmented image (binary mask). In turn, the discriminator (i.e, adversarial term) learns a loss function to train this mapping by comparing the ground-truth and the predicted output. Finally, the whole cGAN network optimizes a loss function that combines a conventional binary cross-entropy loss with an adversarial term. The adversarial term encourages the generator to produce output that cannot be distinguished from ground-truth ones.

5.2 Experiments and discussion

We conducted a comprehensive set of experiments to validate the potential of our proposal on two datasets such as DRISHTI-GS1 Sivaswamy et al. (2015) and RIM-ONE Fumero et al. (2011):

DRISHTI-GS1: this dataset is publicly available and comprises 101 images, which are divided into a training and a testing set of images. Training and testing sets consist of 50 and 51 images respectively. These images have their corresponding binary mask as ground truth.

RIM-ONE: this dataset is publicly available and particularly intended for optic nerve head segmentation. It has a total of 169 high resolution images with their corresponding ground truth. We have used 100 images as training and the rest 69 images for testing.

For quantitative assessment of the performance of OD segmentation, we have computed ACC, Dice, IoU score, SEN and SPE as detailed in Table 5.1. We have performed the experiments using the two datasets with three common segmentation methods, FCN Long et al. (2015), UNetRonneberger et al. (2015) and SegNet Badrinarayanan et al. (2017). In addition, we compared our results with three baseline state-of-the-art methods, such as Shankaranarayana et al. (2017), Maninis et al. (2016) and Zilly et al. (2015).

Table 5.1: Evaluation of the cGAN, FCN, SegNet and UNet models, in addition to three baseline methods evaluated on DRISHTI GS1 and RIM-ONE. The best results are marked in bold. Non-reported results are indicated with a dash (-).

Methods	Dataset	ACC	Dice	IoU	SEN	SPE
FCN	DRISHTI GS1	0.93	0.91	0.89	0.92	0.96
	RIM-ONE	0.94	0.92	0.87	0.88	0.95
SegNet	DRISHTI GS1	0.94	0.88	0.83	0.89	0.95
	RIM-ONE	0.93	0.85	0.78	0.86	0.94
UNet	DRISHTI GS1	0.97	0.95	0.90	0.96	0.98
	RIM-ONE	0.94	0.92	0.89	0.93	0.97
Shankaranarayana et al. (2017)	DRISHTI GS1	-	-	-	-	-
	RIM-ONE	-	**0.98**	0.88	-	-
Maninis et al. (2016)	DRISHTI GS1	-	-	-	-	-
	RIM-ONE		0.96	0.89	-	-
Zilly et al. (2015)	DRISHTI GS1	-	**0.97**	0.91	-	-
	RIM-ONE	-	0.94	0.89	-	-
cGAN (our proposal)	DRISHTI GS1	**0.98**	**0.97**	**0.93**	**0.98**	**0.99**
	RIM-ONE	**0.98**	**0.98**	**0.93**	**0.98**	**0.99**

With DRISHTI-GS1, our cGAN model can segment the OD regions with around 98%, 97%, 96%, 97% and 99% of ACC, Dice, IoU, SEN and SPE, respectively, mostly outperforming the rest six tested segmentation models. However, the Maninis et al. (2016) model also provided the top Dice result, however our cGAN model achieved high IoU of 93% as compared to 88%. The UNet model also provided acceptable results and comparable to our's. The three tested baseline methods have only computed the Dice and IoU as shown in Table 5.1. The work proposed in Zilly et al. (2015) yielded feasible scores with around 97% and 91% of the Dice and IoU, respectively.

Furthermore, in order to support the aforementioned results, we evaluated our model on RIM-ONE dataset. The resulted Accuracy, Dice, IoU score, sensitivity and specificity scores with our model were around 98%, 98%, 93%, 97% and

99%, respectively, also outperforming the rest six compared approaches, except for Shankaranarayana et al. (2017) in Dice, since it yielded the top mark (98%). Again, our result in IoU surpasses the one provided by the other method in 2 percentage points. In addition, The UNet model has still provided good results compared the our method.

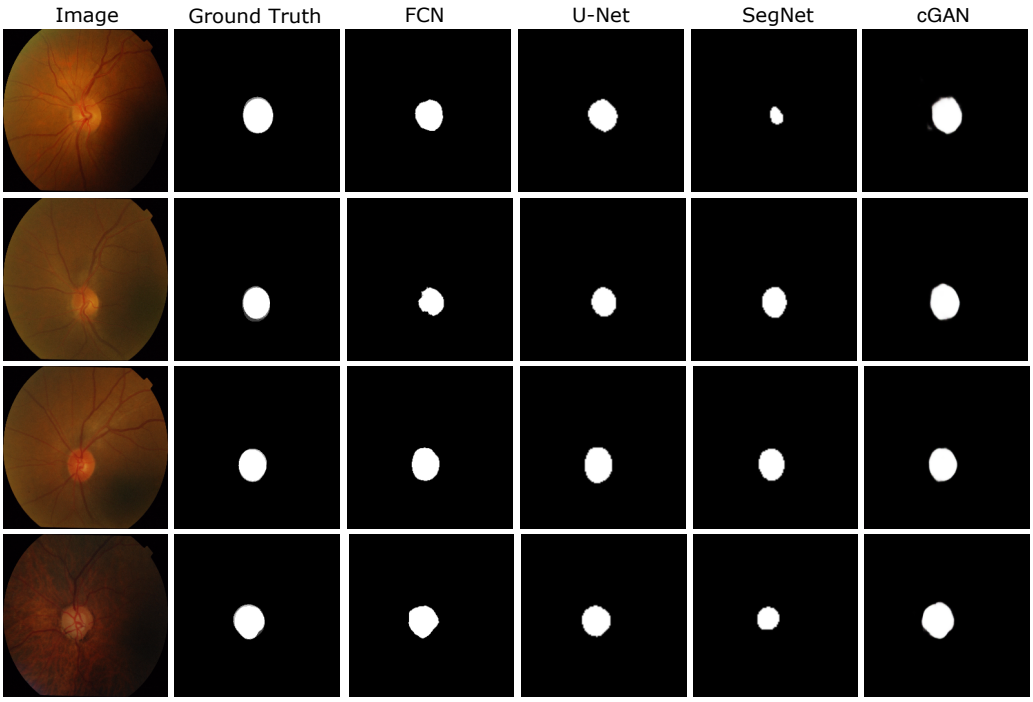


Figure 5.2: Examples of retinal optic disc segmentation : (col 1) retinal images, (col 2) ground-truth masks, (col 3) FCN, (col 4) UNet, (col 5) SegNet and (col 6) generated masks with the cGAN.

A qualitative comparison of segmentation results with the state-of-the-art methods using both retinal optic disc datasets is shown in Figure 5.2. As shown, the OD segmentation with the proposed method is closer to the ground truth with accurate boundaries compared to results of the state-of-the-art methods. The visualization supports our numerical results. The UNet also provided acceptable segmentation. In turn, the SegNet yielded the worst segmentation among the tested methods.

5.3 Conclusion

This chapter proposes a deep learning framework based on cGAN to segment the retinal fundus optic disc. The cGAN consists of two networks: generator and discriminator. The cGAN network does not require a large number of images to be trained properly. In addition, it renders a high segmentation performance without adding extra complexity, since the final segmentation is only achieved with the generator network. Experimental results show that the cGAN slightly outperforms the state-of-the-art OD segmentation methods.

CHAPTER 6

Skin Lesion Segmentation Based on Multi-scale Features and Factorized Channel Attention

Skin lesion segmentation in dermoscopic images is still a challenge due to the low contrast and fuzzy boundaries of lesions. Moreover, lesions have high similarity with healthy regions. In this chapter, we present a fully automated method to segment skin lesions in dermoscopic images.

6.1 Introduction

According to the world health organization (WHO), around 100,000 melanoma skin cancer cases appear every year Stewart et al. (2014). For early diagnosis, different

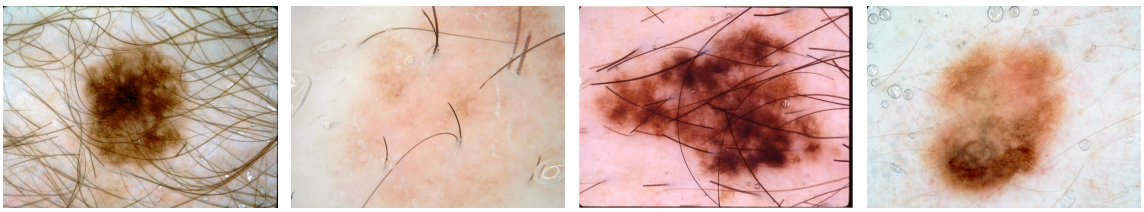


Figure 6.1: Examples of skin lesions with presence of hair, illumination changes, noise, color variations and fuzzy boundaries.

diagnostic algorithms Argenziano et al. (2003), such as the ABCD Dermoscopy Rule and 7-Point Check List, have been utilized. For example, the ABCD rule helps dermatologist to discriminate between benign and malignant tumors by analyzing the following features of the lesion: asymmetry (A), border irregularity (B), color (C) and dermoscopic structures (D).

Nowadays, CAD systems are widely used for an early-stage diagnosis of skin diseases using dermoscopic images. These CAD systems are also used to train inexperienced dermatologists and to devise automated diagnostic procedures. One of the important tasks of these systems is to accurately segment the lesions from dermoscopic images, which helps to follow-up these lesions are growing. Moreover, CAD systems are also designed to extract basic features (e.g., ABCD features) that can be used for in-depth lesion patterns analysis.

Figure 6.1 presents four examples of skin images containing lesions. As can be seen, there are many challenges for skin lesion segmentation methods to properly segregate the observed lesions, such as the presence of hair, illumination changes, noise, color variations, and fuzzy boundaries Day and Barbour (2000). These challenges degrade the performance of automatic segmentation methods.

Several skin lesion segmentation methods have been proposed in the literature Bi et al. (2017), Al-Masni et al. (2018), Yuan (2017), and Berseth (2017), which are based on traditional computer vision, machine learning and/or deep learning techniques. Regarding traditional computer vision techniques, adaptive thresholding, region growing, and contour-based methods have been used to segment skin lesions Rahman et al. (2016). However, these methods yield poor results on low contrast skin images.

Recently, different deep learning techniques have been used to segment biomedical images Du et al. (2018), He et al. (2017), Singh et al. (2018b), Singh et al. (2018a). Several approaches Berseth (2017); Codella et al. (2018) have been proposed to automatically segment skin lesions. These methods yield more accurate results than traditional ones. However, most methods apply several pre-processing (e.g., hair removal, color space transformation and data augmentation) or post-processing techniques (e.g. morphological operations) to improve their results. Among of deep models, cGAN, a cutting-edge idea in image-to-image translation, has been used to segment skin lesions Xue et al. (2018), giving an overall Dice coefficient of 86.7%.

Enhancing the contextual information extracted by some layers of the cGAN model could increase the segmentation accuracy. In order to do so, a dual attention block has been proposed that integrates the spatial and channel long-range dependencies for general scene segmentation Fu et al. (2018b). However, the dual attention block significantly increases the number of training parameters, especially in the spatial attention branch. Therefore, we propose to substitute that branch with a residual connection joined with four layers of 1-D factorized kernel convolution. The factorization method proposed in Romera et al. (2018) further helps in reducing the trainable parameters of the equivalent two layers of 2D kernel convolution. Consequently, for our skin lesion segmentation method, we introduce a novel layer, called FCA, which integrates channel attention and residual 1-D factorized kernel convolutions. On the one hand, we assume that a high cross-channel correlation in activation maps of the encoder layers indicates the presence of relevant cues for distinguishing skin lesion pixels from normal skin pixels. On the other hand, we also assume that the residual convolutions can learn some spatial dependencies in neighboring positions of the feature maps, which lead to a more compact pixel labeling of the lesion regions, but with a minimal set of trained parameters. Consequently, by integrating both methods in Fu et al. (2018b) and Romera et al. (2018), we are able to formulate an efficient and accurate segmentation network.

The main contributions of this chapter are outlined below:

1. We propose a fully automated skin lesion segmentation model based on cGAN,

which can learn more effective features for small-size skin lesions without using pre-processing (e.g., color space transformation) or data augmentation techniques.

2. To model channel and spatial inter-dependencies inside the feature maps of the encoder layers, we introduce the FCA block, which integrates channel attention and residual 1-D factorized convolutions to boost feature discriminability between lesion and non-lesion pixels.
3. We also use a multi-scale input strategy, in which the input images are resized into three different scales of the original size. Thus, the FCA-Net can explicitly deal with variation in resolution, object size and image scale, by encouraging the development of filters which are scale-variant, while constructing a scale-invariant representation.

6.2 Related work

Table 6.1 summarizes some of the skin lesion segmentation methods that have been published recently. These methods include traditional computer vision techniques, CNN and GAN based methods.

Traditional computer vision methods. These methods usually exploit pixel values, color, texture and shape statistics, i.e., hand-crafted features used for the segmentation process. For instance, Rahman et al. (2016) proposed an automatic lesion segmentation using adaptive thresholding and region growing methods to segment skin lesions. These areas were then fed into an extreme learning machine (ELM) to classify skin lesions. The main drawback of thresholding based methods is that they can achieve good results only if there is a high contrast between the lesion area and the surrounding skin region, which is not always the case. Also, Wong et al. (2011) suggested an iterative stochastic region-merging approach, which was employed to segment skin lesions from macroscopic images. In this method, stochastic region merging was initialized on a pixel level, and then on a regional level until convergence. An active contour method (snakes) based on gradient vector

flow (GVF) was proposed in Erkol et al. (2005) for lesion contour extraction. An extension of GVF based on a mean shift method was proposed in Zhou et al. (2013). In Silveira et al. (2009), two contour-based methods were applied to skin images: adaptive snake and active contour. However, contour-based methods usually fail in the presence of hair or air bubbles and if the transition between the lesion and the surrounding skin is smooth.

CNN-Based methods. Nowadays, deep CNN are widely used to analyze natural and medical images for the tasks of detection, segmentation and classification. Several relevant deep learning-based image segmentation methods have been proposed in the last five years. One of the prominent models is the FCN Long et al. (2015) that includes encoder and decoder layers. In Ronneberger et al. (2015), the UNet model was proposed for biomedical image segmentation. It adapted the FCN model by using a skip connection from each encoder layer to the corresponding decoder layer to keep the features extracted from the first layers. Furthermore, the SegNet model was proposed in Badrinarayanan et al. (2017) to improve the accuracy of image segmentation by using a max pooling in the decoder layers that extends from the corresponding encoder layer to achieve a non-linear upsampling of their input feature maps.

Recently, researchers have used state-of-the-art image segmentation based deep learning models to obtain more accurate skin lesion segmentation. In Yu et al. (2017), a CNN-based fully convolutional residual network (FCRN) and multiscale contextual information were proposed to segment skin lesions. However, this method is not able to properly segment low contrast dermoscopic images or those that includes hairs and irregular skin lesion shapes. Also, this method cannot develop the full discrimination capability of a deep CNN with limited training data, according to the authors of the paper. Furthermore, Bissoto et al. (2018) used the UNet network to segment skin lesions. They assessed their model on the ISIC2018 skin lesion dataset and achieved an IoU score of 72.8%. The main limitation of this method is that it requires several pre-processing steps for removing signal noise in dermoscopic images.

Table 6.1: Summary of skin lesion segmentation methods. Dashes (-) indicate that the information is not reported in the referred references.

Study	Dataset	Architecture	Data augmentation/Preprocessing	Input Size	Loss function	Post-processing
Yu et al. (2017)	ISBI2016	FCRN	Rotate, flipping, translation, and random noise	Resize 250×250	-	Thresholding of 0.5
Bissoto et al. (2018)	ISIC2018	UNet	Rotate, illumination, scale, and flipping	Resize 256×256	Binary cross entropy+IoU	Thresholding, hole filling
Bisla et al. (2019)	ISBI2017, ISIC2018	UNet	Random masking	Resize 380×380	Binary cross entropy	Hole filling
Al-Masni et al. (2018)	ISBI2017	FrCN	Rotate and colorspace HSV	Resize 192×256	Cross entropy	-
Mirikharaji et al. (2018)	ISBI2016	UNet	Rotation and flipping	Resize 336×336	Dice	-
Li et al. (2018)	ISBI2017	UNet	Rotation, flipping, and scaling	-	Cross entropy+MSE	Thresholding of 0.5, hole filling
Rahman et al. (2016)	ISBI2016	Adaptive thresholding and region growing	-	-	-	-
Yu et al. (2017)	ISBI2016, ISBI2017	CDNN	Rotate, flipping, shifting, and scaling	Resize 192×192	IoU	Thresholding of 0.5, hole filling
Berseth (2017)	ISBI2017	UNet	Rotation, flipping, and zooming	Resize 192×192	-	CRF
Venkatesh et al. (2018)	ISBI2017	UNet	Rotation, flipping, translation, and scaling	Resize 256×256	Binary cross entropy+IoU	-
Galdran et al. (2017)	ISBI2017	UNet	Rotation, flipping, illumination, scaling, translation, and change RGB to Gray scale	-	IoU	-
Vesal et al. (2018)	ISBI2017	UNet	Rotation, flipping, color shifting, translation, and scaling	Resize 512×512	Dice	-
Sarker et al. (2018)	ISBI2017	UNet	-	Resize 384×384	NLL + EPE	-
Jahanifar et al. (2019)	ISBI2016, ISBI2017	Saliency detection	-	Resize 300×400	-	Thresholding
Bi et al. (2019)	ISBI2016, ISBI2017	FCN	Random cropping and flipping	Resize 1000×1000	Cross entropy	Thresholding, hole filling
Xue et al. (2018)	ISBI2017	GAN	Random cropping, flipping, and illumination change	Resize 128×128	L1-norm	-
Izadi et al. (2018)	DermoFit	GAN	Random cropping, flipping, and elastic deformation	-	Binary cross entropy	-

To accurately segment the lesion boundaries, Vesal et al. (2018) proposed the SkinNet model, which is based on the UNet architecture. The authors replaced standard convolution layers at every level of both the encoder and decoder, with densely connected convolution layers. The SkinNet model was evaluated on ISBI2017 dataset and achieved a Dice of 85.10% and an IoU score of 76.7%. To extract rich features from a dermoscopic image, Sarker et al. (2018) utilized a residual network weighting and a spatial pyramid pooling network. They also proposed the use of a loss function called End Point Error (EPE) to preserve the lesion boundaries. Furthermore, Jahanifar et al. (2019) integrated a multilevel segmentation algorithm, regional contrast, background descriptors, and a random forest regressor to create saliency scores for each region in the image. This method gives poor segmentation results with low contrast dermoscopic images.

To add image appearance information as well as contextual information, Mirikharaji et al. (2018) employed a UNet based method to predict the pixel-wise probability of a skin lesion segmentation. It achieved a Dice of 90.11% and an IoU score of 83.30% on ISBI2016 dataset. Furthermore, Li et al. (2018) proposed a transformation consisting of a self-ensemble model, which enhances the regularization effects by utilizing the unlabeled data. It achieved a Dice of 87.40% and IoU scores 79.87% on ISBI2017 dataset. In turn, Venkatesh et al. (2018) used a UNet model that is based on multi-scale input with a shortcut connection at each block of the UNet. The suggested method has evaluated on ISBI2017 dataset, obtaining a Dice and IoU scores of 85.60% and 76.40%, respectively. Galdran et al. (2017) also exploited an UNet architecture and used color constancy methods to normalize the color throughout the dataset images while retaining the estimated illumination information, enabling them to randomly change the color and illumination of normalized images during the training process. They achieved a Dice of 82.40% on ISBI2017 dataset.

Bi et al. (2019) proposed a FCN-based class-specific training to extract visible features of different kinds of skin lesions and a probability-based step-wise combination of the derived class-specific segmentation maps to guarantee visible

persistence of the segmented regions. Also, Yuan (2017) introduced a FCN-based model, which contains 29 layers to segment skin lesions from dermoscopic images. They have employed an upsampling and deconvolutional layers to compute multi-resolution loss while carrying over the global perspective from pooling layers. However, their model requires several pre- and post-processing operations, such as color space transformations and threshold selection. Moreover, Al-Masni et al. (2018) proposed a full resolution convolutional networks (FrCN), which directly learns the full resolution features of each pixel of the input data without the requirement of pre or post-processing methods.

GAN-based methods. Recently, several GAN-based segmentation models have been applied to medical images. For instance, cGAN has been used in Singh et al. (2018b), Singh et al. (2020) to segment breast cancer sub-types, in which a Dice loss is introduced in the generator network to refine pixel-wise segmentation results. To segment skin lesions, Xue et al. (2018) proposed an adversarial-based model with residual blocks and skip connections. The model was evaluated on ISBI2017 dataset, achieving a Dice of 86.70% and an IoU score of 78.50%. Moreover, Bisla et al. (2019) introduced the DCGAN and ResNet-50 models to jointly segment the skin lesion and classify the lesions into benign and malignant. They exploited pre-processing steps to suppress the artifacts from the skin images. With ISBI2017 and ISIC2018 test datasets, they obtained IoU scores of 77.00% and 70.20%, respectively.

Most of the methods stated in Table 6.1 have utilized data augmentation/preprocessing techniques in the training phase while others applied postprocessing techniques (e.g., morphological operations) on the resulting masks. In turn, during the training of the proposed FCA-Net, we utilize the original images of ISBI2016, ISBI2017, and ISIC2018 datasets without applying any data augmentation technique. For a fair comparison with the state-of-the-art methods, we separately trained and tested the proposed model on the training and testing sets of the aforementioned datasets. The proposed cGAN model that includes FCA blocks and a multi-scale stage highlights the most important features (of a highly receptive field) and disregards the artifacts from images.

6.3 Methodology

6.3.1 The factorized channel attention block

Figure 6.2 presents the design of the proposed FCA block, which applies a weighted aggregation between the output features of two mechanisms: channel attention (upper branch) Fu et al. (2018b) and residual 1-D factorized convolutions (lower branch) Romera et al. (2018). The proposed FCA block increases the representational power of features computed by encoder layers in generator and discriminator networks.

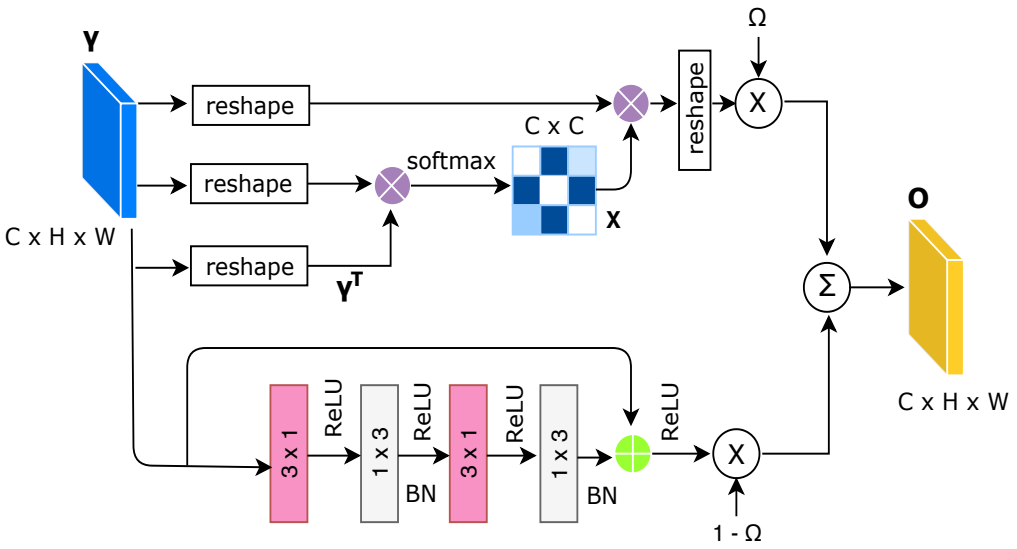


Figure 6.2: Proposed FCA block with integration of channel attention and residual 1-D factorized convolution.

Channel attention. This mechanism is intended to boost feature channels that have similar values in the same image positions. Assume that $\gamma \in R^{C\times H\times W}$ is the activation map (i.e. set of features) obtained by the original encoder layer. To calculate the channel attention map $X \in R^{C\times C}$, γ is firstly reshaped to $R^{C\times N}$ ($N = H \times C$), then multiplied by its transpose, and finally normalized with the softmax function:

$$x_{ji} = \frac{exp(\gamma_i.\gamma_j)}{\sum_{i=1}^{C} exp(\gamma_i.\gamma_j)} \tag{6.1}$$

where γ_i and γ_j are vectors of length N, containing the values of all map positions

in channels i and j, respectively, and $\gamma_i.\gamma_j$ represents their dot product. Hence, x_{ji} represents a normalized correlation degree between those two channels. The output of the channel attention branch, $O_1 \in \mathbb{R}^{C\times H\times W}$, can be expressed for each channel j as follows:

$$O_{1j} = \eta\sum_{i=1}^{C}(x_{ji}\gamma_i) + \gamma_j \tag{6.2}$$

where $\sum_{i=1}^{C}(x_{ji}\gamma_i)$ includes the feature values of all channels modulated by the correlation degree between each channel with respect to the j^{th} channel. Moreover, this summation is weighted by η, which is a learned weighting parameter and then added to the original activation map. In this way, channels that present more similarities increase their relevance in the output. This mechanism improves the segmentation accuracy because relevant patterns corresponding to skin lesion areas create high activation values in several feature channels, while other irrelevant patterns, like healthy skin areas or hairs, may have associated very few feature channels as they representatives.

Residual 1-D factorized convolutions. This mechanism is intended to boost feature values that are similar in different image positions. The core working is the same as a typical residual convolution, i.e., to learn the difference between input and output activation maps, but using factorized 1-D kernels instead of regular 2D kernels. We hypothesize that the output of this branch will tend to detect image areas that present similar skin patterns in neighboring image positions. This output does not entirely assume the role of the full spatial attention mechanism, where similar patterns in disconnected image areas can be enhanced. Nevertheless, in the skin lesion context, the residual convolution is enough if we assume that most of the lesions are contained in a contiguous image region.

Factorized kernels can effectively preserve the spatial information of regular 2D kernels and maintain the accuracy with significantly less computation.

Assume that $\mathbf{W} \in \mathbb{R}^{C\times d^h\times d^v\times F}$ are the weights of a typical 2D convolutional layer, where C is the number of input planes (channels), F is the number of output planes (i.e. feature maps) and $d^h \times d^v$ is the kernel size of each feature map (typically $d^h = d^v = d$). $\mathbf{f^i} \in \mathbb{R}^{d^h\times d^v}$ is the i^{th} kernel in the layer. As proposed in Romera et al.

(2018), $\mathbf{f^i}$ can be expressed as a linear combination of 1-D filters:

$$\mathbf{f^i} = \sum_{k=1}^{K} \sigma_k^i \bar{v}_k^i \left(\bar{h}_k^i\right)^T \tag{6.3}$$

where σ_k^i is a scalar weight, K is the rank of $\mathbf{f^i}$, $\bar{v}_k^i$ and $\left(\bar{h}_k^i\right)^T$ are vectors of length d. The i^{th} output of the decomposed layer, O_{2i} can be expressed as a function of its input γ, as follows:

$$O_{2i} = \varphi\left(b_i^h + \sum_{l=1}^{L} \bar{h}_{il}^T * \left[\varphi\left(b_l^v + \sum_{c=1}^{C} \bar{v}_{lc} * \gamma_c\right)\right]\right) + \gamma_i \tag{6.4}$$

where $\varphi(.)$ represents the non-linearity of the 1-D decomposed filters (where we used ReLU), b_i^h and b_l^v are the horizontal and vertical biases of each filter. The residual strategy of this branch combines the original features provided by the previous layer with a new set of feature maps with 1-D convolutional filters. 1-D convolutional filters have intrinsically less computational cost and less number of parameters than their 2D equivalent filters. Additionally, the 1-D combinations improve the compactness of the generator layers by minimizing redundancies in the features coming from the previous 2D convolution layers and theoretically improving the learning capacity. Besides, the residual 1-D kernel factorization is faster in terms of computation time than the normal non-bottleneck design He et al. (2016b).

To determine the final output of the FCA block, we aggregate the channel attention and the residual 1-D factorized convolutions outputs as follows:

$$O = (1 - \Omega) \times O_1 + \Omega \times O_2 \tag{6.5}$$

Here, Ω is the weighting factor. We checked the system performance for Ω values from 1.0 to 0.0, in steps of 0.1. We have found that $\Omega = 0.3$ provides the best results.

6.3.2 Network architecture

The proposed model comprises a generator and a discriminator network. The generator network includes an encoder section and a decoder section. As shown in Figure 6.3, both encoder and decoder sections include seven sequential layers (E_n refers to an encoder layer and D_n refers to a decoder layer).

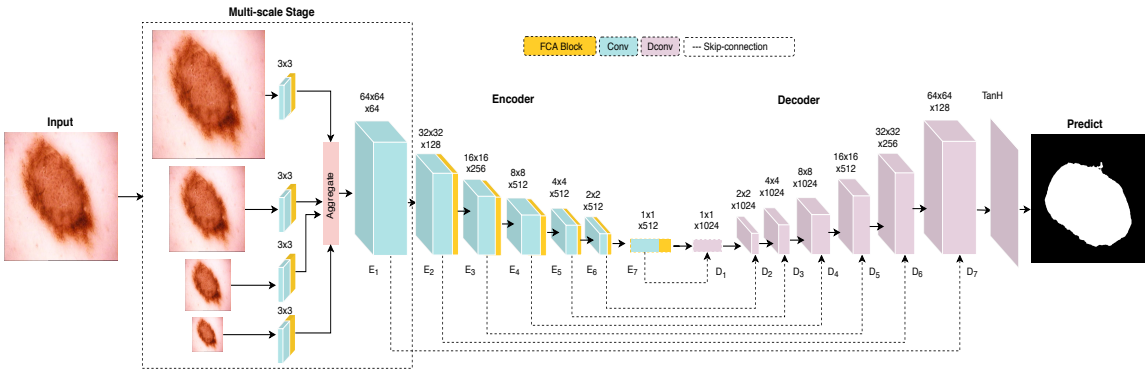


Figure 6.3: The architecture of the generator network.

The encoder. We use a multi-scale input strategy Van Noord and Postma (2017), where the input images are resized into three different scales with ratios of 1/8, 1/4 and 1/2 of the original size. In this way, the FCA-Net explicitly deals with variation in resolution, object size and image scale, by encouraging the development of filters which are scale-variant, while constructing a scale-invariant representation. This strategy helps to segment some small skin lesion pixels. After each scale, we add a convolution layer with 3×3 kernels along with the proposed FCA block to extract more rich features from a skin lesion. The sizes of the features that are fed into the aggregation module are $128 \times 128 \times 64$, $64 \times 64 \times 64$, $32 \times 32 \times 64$ and $16 \times 16 \times 64$, respectively. Before aggregation, lower-scale features are upsampled to the size of the feature vector extracted from the original image ($128 \times 128 \times 64$), and then inputted into the encoder layers. We added the proposed FCA block in all layers of the encoder part, and we did not add it to the decoder layers. We used batch normalization with $LeakyReLU$ (slope 0.2) after the first six layers of encoder and the $ReLU$ activation function after E_7. The size of all convolutional kernels is 4×4 with a stride of 2. The encoder can learn low-level features of the skin images, such as spatial information (e.g., edge, intensity, texture) throughout the training

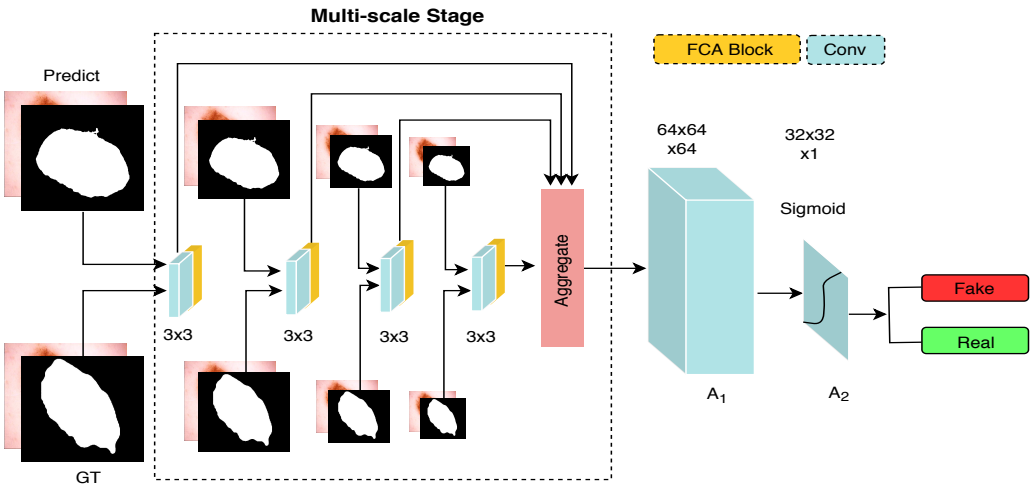


Figure 6.4: The architecture of the discriminator network.

process.

The decoder. To avoid overfitting, we used batch normalization and dropout (rate = 0.5) in D_1, D_2 and D_3. We added the $ReLU$ activation function after each layer of the decoder. The size of the deconvolutional kernels is 4×4 with a stride of 2. We also added skip connections between each convolutional layer to its corresponding deconvolutional layer. To outline the segmented skin lesion into a binary mask, we added *Tanh* after D_7. We used a threshold of 0.5 to convert the output of *Tanh* activation function to binary masks.

The discriminator network. In Figure 6.4, similar to the generator network, we apply a multi-scale stage with the FCA block to enhance the scale independence of the discriminator between the ground truth and the generated masks. The masks are also resized into three scales with ratios 1/8, 1/4 and 1/2 of the original image size. The extracted features are up-sampled, concatenated and inputted into two successive convolutional layers A_1 and A_2. We add $LeakyReLU$ activation function in A_1 and *sigmoid* function in A_2.

6.3.3 Loss function

To optimize the proposed segmentation model, we employ a loss function composed of three terms: Adversarial loss as Binary Cross Entropy (BCE), ℓ_{L1} loss and EPE proposed in Sarker et al. (2018). Assume x is the skin image containing a lesion, y

is the ground truth mask, $G(x,z)$ and $D(x,G(x,z))$ are the outputs of the generator and the discriminator, respectively. The loss function of the generator network G is as follows:

$$\begin{aligned}\ell_{Gen}(G,D) &= \mathbb{E}_{x,y,z}(-\log(D(x,G(x,z)))) \\ &+ \lambda\mathbb{E}_{x,y,z}(\ell_{L1}(y,G(x,z))) \\ &+ \beta\mathbb{E}_{x,y,z}(\ell_{EPE}(y,G(x,z)))\end{aligned} \tag{6.6}$$

where z is a random variable, and β and λ are empirical weighting factors. The ℓ_{L1} loss forces the model to suppress the outliers and artifacts and speed up the optimization process.

The EPE loss Baker et al. (2011) compares the magnitude and orientation of the edges of the predicted mask with its ground truth for preserving the boundaries of the segmented regions. The EPE can be defined as:

$$L_{epe} = \sqrt{(G(x,z)_x - y_x)^2 + (G(x,z)_y - y_y)^2} \tag{6.7}$$

where $(G(x,z)_x,\ G(x,z)_y)$ and $(y_x,\ y_y)$ are the first derivatives in x and y directions of $G(x,z)$ and y, respectively.
In the discriminator network, we only used the BCE loss, which is defined as:

$$\begin{aligned}\ell_{Dis}(G,D) &= \mathbb{E}_{x,y,z}(-\log(D(x,y))) \\ &+ \mathbb{E}_{x,y,z}(-\log(1-D(x,G(x,z))))\end{aligned} \tag{6.8}$$

The optimizer will fit D to maximize the loss values for ground truth masks (by minimizing $-\log(D(x,y))$) and minimize the loss values for generated masks (by minimizing $-\log(1-D(x,G(x,z)))$. The generator and discriminator networks are optimized concurrently, one optimization step for both networks at each iteration, where G tries to generate an accurate lesion segmentation mask and D learns how to discriminate between the synthetic and the real segmentation masks.

6.4 Experimental results and discussion

Datasets. To assess the efficacy of the proposed model, we use three skin lesion challenge datasets, which are publicly available: ISBI2016[1] Gutman et al. (2016), ISBI2017[2] Codella et al. (2018) and ISIC2018[3] Codella et al. (2019). The images of the datasets were acquired using different devices at several medical centers worldwide. In the ISBI2016 dataset, the training set has 900 annotated images while the testing set has 379 annotated images. The size of the images varies from 542×718 to 2848×4288 pixels. The ISBI2017 dataset has three sets: the training set (2000 images), validation set (150 images) and testing set (600 images). In the ISIC2018 dataset, the training set has 2594 images, the validation set has 100 images and the testing set contains 1000 images. Ground truth of mask images (provided for training and used internally for scoring validation and test phases) were generated using several techniques, but all data were reviewed and curated by practicing dermatologists with expertise in dermoscopy. Our model has been trained on randomly chosen 2546 skin lesion images from the ISIC 2018 dataset and validated on the rest 48 images. The model is then evaluated on the validation and testing sets. This process has only been used to tune the hyperparameters of the model. Note that the trained model is evaluated on the testing sets of ISBI2016 and ISBI2017 datasets. Our model is further assessed on the ISIC2018 validation and test sets.

Parameter settings. Each input image is resampled to 128×128 pixels and normalized for rescaling the pixel values between [0, 1], before feeding it into our network. The hyper-parameters of the model were empirically tuned. We used Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$, a learning rate of 0.0002 and a batch size of 2. The weighting factors of the ℓ_{L1} and EPE losses (λ and β) were set to 100 and 50, respectively. We have trained our model for 300 epochs, although the best results were obtained on 240 epochs.

Implementation details. The experiments were conducted on an NVIDIA GeForce GTX 1070 with 8 GB of video RAM. The operating system was Ubuntu

[1]https://challenge.kitware.com/challenge/560d7856cad3a57cfde481ba
[2]https://challenge.kitware.com/challenge/583f126bcad3a51cc66c8d9a
[3]https://challenge.kitware.com/challenge/5aab46f156357d5e82b00fe5

Table 6.2: Performance metrics of different configurations of the proposed method with ISBI2016 and ISBI2017 datasets.

Methods	ISBI2016					ISBI2017				
	ACC	*Dice*	*IoU*	*SEN*	*SPE*	*ACC*	*Dice*	*IoU*	*SEN*	*SPE*
BL	92.66	88.23	81.59	86.78	94.42	92.08	83.52	69.36	76.47	94.11
BL+CA	93.67	88.97	82.62	87.78	94.91	93.86	83.99	73.50	79.37	94.64
BL+FK	94.95	90.33	84.16	86.36	96.61	94.29	84.52	75.10	83.46	96.79
FCA-Net w/o MS	95.69	91.75	86.01	91.47	97.88	95.11	86.54	76.88	87.36	96.92
FCA-Net	**96.97**	**93.94**	**87.58**	**92.42**	**98.62**	**96.29**	**88.28**	**78.94**	**88.09**	**97.36**

16.04 using a 3.4 GHz Intel Core-i7 with 16 GB of RAM. The main required packages involve Python 3.6, CUDA 9.1, cuDNN 7.0 and PyTorch 0.4.1. The codes of the proposed model are publicly available at https://github.com/vivek231/Skin-Project.

6.4.1 Ablation study

We have run an ablation study to demonstrate the effect of each part of the proposed block. We firstly trained a baseline (BL) model without adding the channel attention or factorized convolution blocks. Then, we added the channel attention block to the encoder layers of the generator network (called the BL+CA model). Furthermore, the factorized kernel was also separately added to the encoding layers (called the BL+FK model). Finally, the proposed model (FCA-Net) is constructed by adding both CA and FK blocks with and without multi-scale. Note that all models used in this ablation study have been trained on the ISIC2018 dataset and tested on the ISBI2016 and ISBI2017 datasets. We did not consider the testing images of ISBI2016 and ISBI2017 that appear in the training set of the ISIC2018 dataset.

Table 6.2 presents the results of the ablation study. With the dataset of ISBI2016 dataset, the BL model yields a Dice and IoU scores of 88.23% and 81.59% respectively, while the BL+CA model provides a small improvement of 0.74% and 1.03% for Dice and IoU scores, respectively. This improvement is achieved because the CA block explicitly models inter-dependencies among channels. Besides, the BL+FK model gives a Dice of 90.33% and an IoU score of 84.16%, yielding better improvement than the BL+CA model. In turn, the proposed FCA-Net achieves an improvement of around 4.0% and 3.0% of Dice and IoU scores, respectively, with respect the BL model. Similarly, on ISBI2017 dataset, FCA-Net improved the

results comparing with other checked strategies and achieved a Dice of 88.28% and IoU scores of 78.94%.

To demonstrate the effectiveness of the multi-scale stage, Table 6.2 also provides the results of FCA-Net with the multi-scale stage and without employing the multi-scale stage (w/o MS). As shown, the multi-scale stage improves the segmentation performance of the proposed FCA-Net model with an increment of approximately 2% in Dice and IoU scores.

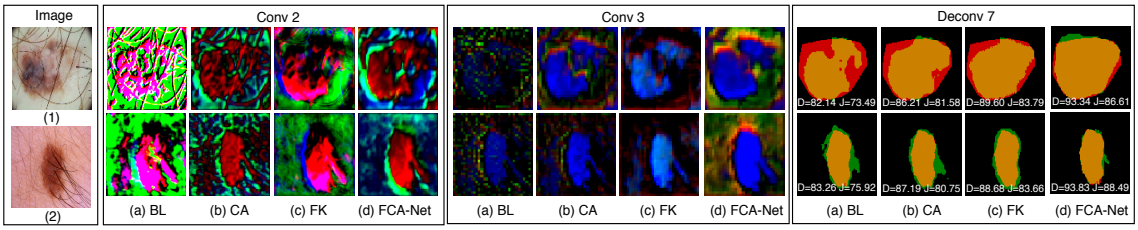


Figure 6.5: Visualization of two sample images and the corresponding activation maps generated by the second and third convolutional layers of the generator network in four variants w.r.t. the use of the FCA block: BL is our baseline cGAN; CA includes the channel attention branch; FK includes the residual 1-D factorized convolutions; FCA-Net is our fully-fledged network. The figure also shows the output (Deconv7) of the variants, graphically compared with the ground truth segmentation, color-coded as yellow: TP, green: FP, red: FN and black: TN, as well as the Dice and IoU indexes for each experiment.

In Fig. 6.5, we present a couple of difficult samples, jointly with visualizations of the activation maps created by the second (Conv2) and third (Conv3) encoder layers, as well as the output of the last decoder layer (Deconv7) graphically compared with the ground truth segmentation (more examples in Fig. 6.7). The layer outputs have been obtained with four variants of the FCA block (see figure caption).

The BL variant (basic cGAN) can distinguish lesion from non-lesion areas in low-level layers of the network, as shown with pink and green zones in Conv2 activation maps, but artifacts like hair and texture/color variability are interfering with the detection of the lesion area. Besides, for image 2 there is a false lesion detection at the left of the true lesion. Although next layers can get rid of such inaccuracies up to a certain degree, the final outputs show a high amount of false negatives (red pixels in output 1) or a fair amount of false positives (green pixels in output 2). Also, note that the output for image 1 comprises small holes (red spots inside the yellow area).

The CA variant obtains a more consistent identification of the target classes. In Conv2, almost all lesion region renders one single color (dark red), although their pixels show varying shading due to different activation degrees (especially for image 1). In Conv3, the lesion areas are more compact (shaded in dark blue), but for image 1 there is a visible break. Nevertheless, this groove will be filled in by further layers. Noticeably, the final output of CA for image 1 has reduced the number of FN with respect to the baseline output, although there are still too many red pixels. The output for image 2 has not trimmed the false (green) elongation at the right of the lesion, but it has trimmed the extra green pixels at its left boundary, so the performance metrics have been improved. Despite that these outputs still present some misled areas, the obtained degree of improvement empirically proves that the channel attention mechanism can significantly smooth the effect of artifacts since lesion features are consistently enhanced.

The FK variant obtains a much more compact coloring of the target areas, thanks to the local spatial coherence provided by the residual filters. Despite this good property, lesion regions in Conv2 maps contain different colors (red, pink, dark blue), which indicates that several feature channels are responsible for characterizing different areas of the lesions. In Conv3, however, the lesion features are more consistent, showing one single blue color. Segmentation outputs are better than outputs from BL and CA variants, which indicates that the residual branch is enhancing the compactness of the output activation maps.

Finally, the FCA-Net combines the good properties of CA and FK variants, since the activation maps from Conv2 and Conv3 tend to be both spatial and channel coherent within the lesion area. For the non-lesion area, however, in Conv3 there is a color gradation (halo) around the lesion, which may turn out into misleading boundary delineation, like the small green area on top of the lesion in the final segmentation of image 1. This inconvenience is largely compensated by the significant reduction of false negatives in output 1. At the same time, there is also a significant reduction of false positives in output 2. In summary, the combination of the two branches of the FCA block helps each other to provide the best results of the four

variants.

Table 6.3: Analysis of the effect of the FCA block on different segmentation models with ISBI2016 dataset.

Methods	Without FCA Block					With FCA Block				
	ACC	*Dice*	*IoU*	*SEN*	*SPE*	*ACC*	*Dice*	*IoU*	*SEN*	*SPE*
FCN8	90.10	80.60	72.52	81.11	97.17	91.02	83.37	74.42	85.15	98.39
UNet	89.23	79.51	71.66	80.26	96.87	91.22	82.10	73.92	84.92	97.41
SegNet	92.36	83.74	75.92	84.67	97.02	93.41	86.61	78.45	87.51	97.87
Link-Net	92.97	84.76	77.01	85.97	**97.26**	93.74	87.53	79.46	88.02	97.98
RefineNet	**93.03**	83.97	75.58	86.08	95.71	93.19	85.82	77.71	89.41	96.76
FCA-Net	92.66	**88.23**	**81.59**	**86.78**	94.42	**96.97**	**93.94**	**87.58**	**92.42**	**98.62**

Table 6.4: Analysis of the effect of the FCA block on different segmentation models with ISBI2017 dataset.

Methods	Without FCA Block					With FCA Block				
	ACC	*Dice*	*IoU*	*SEN*	*SPE*	*ACC*	*Dice*	*IoU*	*SEN*	*SPE*
FCN8	86.38	64.55	57.38	61.71	88.26	89.93	68.85	59.98	64.90	92.82
UNet	85.84	63.03	55.39	60.87	88.24	90.04	67.89	59.39	63.20	95.03
SegNet	88.76	72.45	64.30	71.52	92.44	91.66	75.24	66.06	74.68	96.02
LinkNet	89.62	75.97	69.37	73.81	94.35	92.57	79.59	71.20	77.77	96.01
RefineNet	88.28	74.05	63.64	73.88	**94.36**	91.52	75.31	67.82	74.92	95.57
FCA-Net	**92.08**	**83.52**	**76.36**	**84.47**	94.11	**96.29**	**88.28**	**78.94**	**88.09**	**97.36**

In our experiments, we added the proposed FCA block to different state-of-the-art image segmentation methods (FCN8 Long et al. (2015), UNet Ronneberger et al. (2015), SegNet Badrinarayanan et al. (2017), LinkNet Chaurasia and Culurciello (2017) and RefineNet Lin et al. (2017)) and evaluated them on ISBI2016 and ISBI2017 dataset. Tables 6.3 and 6.4 present the results of all models with and without the proposed FCA block. For a fair comparison, all models have the same configuration of the multi-scale input layer. As can be observed, all models give better results when adding the FCA block. Besides, our model gives the best results among all evaluated models.

Fig. 6.6 shows boxplots of Dice and IoU scores of the proposed model, as well as the FCN8, UNet, SegNet, LinkNet and RefineNet models, evaluated on ISBI2016 and ISBI2017 datasets. As shown in plot (a), the proposed model has the highest median Dice and the smallest range with few outliers. In plot (b), with the 379 images of the test set of ISBI2016, the proposed model produces 13 outliers with IoU scores while the LinkNet and SegNet models have 18 and 17 outliers, respectively.

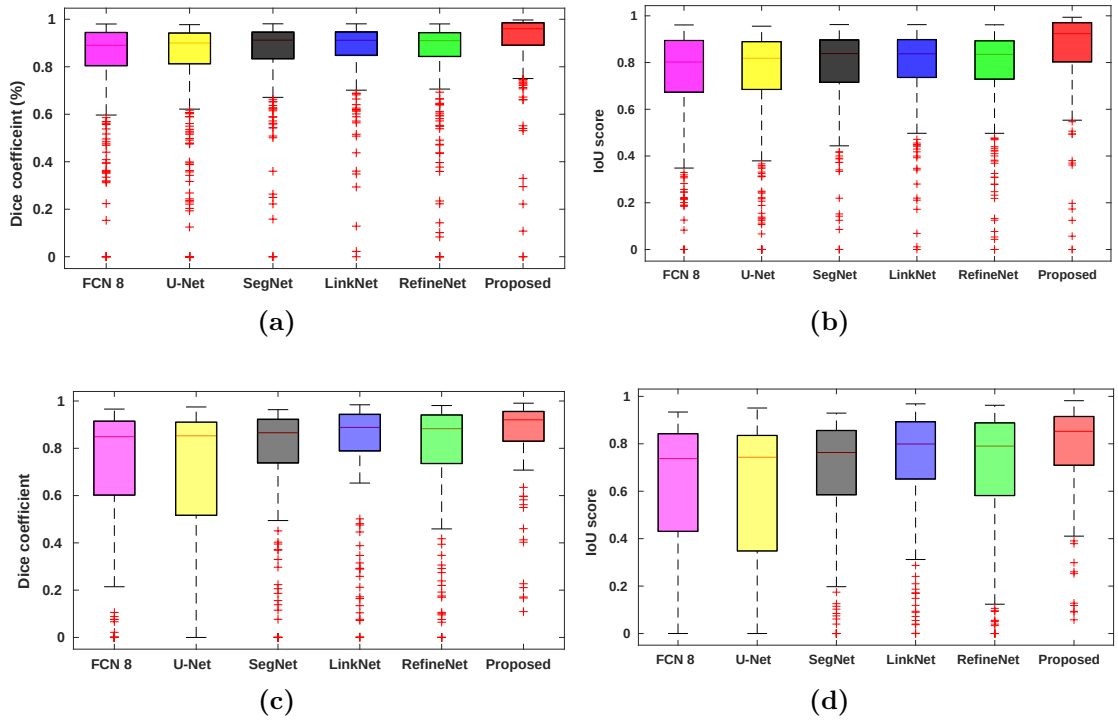


Figure 6.6: Boxplots of Dice and IoU scores for all test samples in ISBI2016 dataset in the upper row (plots a, b) and ISBI2017 dataset in the bottom row (plots c, d). Different color boxes indicate the score range of several methods, the red line inside each box represents the median value, box limits include interquartile ranges Q2 and Q3 (from 25% to 75% of samples), upper and lower whiskers are computed as 1.5 times the distance of upper and lower limits of the box, and all values outside the whiskers are considered as outliers, which are marked with the (+) symbol.

In turn, plot (d) shows that the FCN8 and UNet models have no outliers but the range of IoU metric is much bigger than the one from our model. With the 600 images of the test set of ISBI2017, our model has 10 outliers with IoU scores while the RefineNet model has 14 outliers.

6.4.2 Comparisons

In Table 6.5, the FCA-Net model is compared with 8 state-of-the-art skin lesion segmentation methods on ISBI2016 dataset. For fair comparisons, we have also trained and tested FCA-Net with the ISBI2016 training set and evaluate them with the ISBI2016 test set. As shown, FCA-Net gives a Dice score of 92.80% and an IoU score of 86.41%, which are the second best values in each metric. Li et al. (2019) also achieves a sensitivity score higher than our model but we obtain better specificity

and accuracy.

Table 6.5: Comparing the proposed model with 8 state-of-the-art methods on ISBI2016 dataset. Best results are marked in bold.

Methods	*ACC*	*Dice*	*IoU*	*SEN*	*SPE*
ExB[4]	95.30	91.00	84.30	91.00	96.50
Mirikharaji et al. (2018)	95.02	90.11	83.30	90.15	97.00
Li et al. (2019)	95.90	**93.10**	**87.00**	**95.10**	96.00
Bi et al. (2019)	95.80	91.70	85.90	93.10	96.00
Jahanifar et al. (2019)	94.30	90.70	83.80	90.10	96.60
CUMED Yu et al. (2017)	94.90	89.70	82.90	91.10	95.70
Rahman et al. (2016)	95.20	89.50	82.20	88.00	96.90
Yuan (2017)	95.50	91.20	84.70	91.80	96.60
FCA-Net	**95.93**	92.80	86.41	91.63	**97.07**

Table 6.6 shows a comparison between the results of our FCA-Net and 11 state-of-the-art skin lesion segmentation methods on ISBI2017 dataset. For fair comparisons, we have also trained FCA-Net with the ISBI2017 training set and evaluated it with the ISBI2017 test set. Although the IoU score of Li et al. (2018) is slightly better than our's, we have achieved the second best result. Moreover, we have achieved the best Dice and accuracy scores. Also, Sarker et al. (2018) equals the best Dice score and obtains the best specificity score, 1% higher than FCA-Net, but significantly lower sensitivity and 1% less accuracy than our model. In the IoU score, which is the most strict metric, it performed slightly lower than our's, so we conclude that our method can be ranked as the second best, closely followed by Sarker et al. (2018).

The performance of FCA-Net is also assessed on the validation set of ISIC2018. The segmented images are submitted to the *Leaderboards* platform, which calculates an IoU_{th} score of the provided results. Note that this IoU_{th} score is computed as follows Codella et al. (2019):

$$IoU_{th} = \begin{cases} IoU, & \text{if } IoU > 0.65 \\ 0, & otherwise. \end{cases} \tag{6.9}$$

On the validation set of the live leaderboard, our model has achieved an IoU score

Table 6.6: Comparing the proposed model with 11 the state-of-the-art methods on ISBI2017 datasets. Best results are marked in bold. Dashes (-) indicate that results are not reported in the cited papers.

Methods	*ACC*	*Dice*	*IoU*	*SEN*	*SPE*
Bisla et al. (2019)	-	-	77.00	-	-
Al-Masni et al. (2018)	94.03	87.08	77.11	85.40	96.69
Li et al. (2018)	94.30	87.40	**79.80**	87.90	95.30
Xue et al. (2018)	94.10	86.70	78.50	-	-
Yuan (2017)	93.40	84.90	76.50	82.50	97.50
Berseth (2017)	93.20	84.70	76.20	82.00	97.80
Jahanifar et al. (2019)	93.00	83.90	74.90	81.00	98.10
Venkatesh et al. (2018)	93.60	85.60	76.40	83.00	97.60
Galdran et al. (2017)	92.30	82.40	73.50	81.30	96.80
Sarker et al. (2018)	93.60	**87.80**	78.20	81.60	**98.30**
Vesal et al. (2018)	93.20	85.10	76.70	**93.00**	90.50
FCA-Net	**94.95**	**87.80**	78.65	87.91	97.05

of 77.2%. On the test set (with 1000 skin images), it has achieved an average Dice score of 85.8% and IoU score of 78.2%.

An additional comparison has been performed by comparing the FCA-Net model with FCN, UNet, SegNet, FrCN (reported in Al-masni et al. (2018)), GAN-FCN Bi et al. (2018), and Hardie et al. (2018). In the case of GAN-FCN model, GAN is used to derive additional training data from ISIC2018 dataset, and then this data is combined with the original training data to train an FCN model for skin lesion segmentation. The authors of Hardie et al. (2018) transform skin images to the RGB space and train a color classifier to discriminate between lesions and normal skin tissue based only on RGB color vectors. Then, they use a Gaussian mixture model (GMM) to model the probability density functions of skin lesions and a support vector machine regression algorithm to segment the images. As shown in Table 6.7, our model achieves an IoU score close to the best one provided by GAN-FCN. In turn, the UNet model yields the worst results with an IoU of 54%.

Fig. 6.7 presents qualitative results of skin lesion segmentation that include a variety of challenging conditions: hair presence, blurriness, illumination variations (intensity, chromaticity, fading, etc.), fuzzy or irregular borders, big and small lesions, and multi-color lesions. Each row of Fig. 6.7 includes a skin lesion image along with

Table 6.7: The performance of FCA-Net on the ISIC2018 validation dataset. The proposed model has been evaluated on skin lesion leaderboard (https://submission.challenge.isic-archive.com/)

Methods	**IoU**$_{th}$(%)
FCN	74.70
UNet	54.40
SegNet	69.50
FrCN	74.60
GAN-FCN Bi et al. (2018)	**77.80**
Hardie et al. (2018)	66.30
FCA-Net	77.20

its segmentation obtained with FCN8, UNet, SegNet, Refine-Net, LinkNet and the proposed model. To visualize the accuracy of each model, we compare the ground truth mask with the generated mask and use four different color codes to mark up the classification result for each pixel. Note that yellow refers to TP, red refers to FN, green refers to FP and black refers to TN. An ideal segmentation model will assign yellow to skin lesion pixels and black to the background pixels.

For all examples, our model achieves the best Dice and IoU scores in all images, as shown in the scores printed on each output image. Moreover, it produces a tiny amount of red or green pixels, which are usually distributed around the border of the skin lesion, while the other methods have a higher number of FP and FN.

The sample in the first row has a small lesion with thick hair in the background. The UNet model failed to properly segment by over-extending the lesion area, while the rest of the models give accurate segmentation and high Dice and IoU results. The sample of the second row has low contrast with bad illumination conditions, along with a framing effect due to the optic lenses. In this case, FCN8, UNet, SegNet, RefineNet, and LinkNet give a low Dice and IoU scores ($\leq$ 88%), while our method outputs a decent result. In the third row, the pixels inside the lesion region have inhomogenous colors, and some of them are similar to the ones of the background. The UNet and LinkNet method give bad segmentation results while the proposed model achieves the most accurate fit of the lesion, with very high Dice and IoU scores. In the fourth row, the lesion region is not homogeneous and has low contrast. With this case, our model achieves a Dice of 87.34% and an IoU of 77.53%, which correspond to the best matching of the lesion. The last row has a small skin lesion

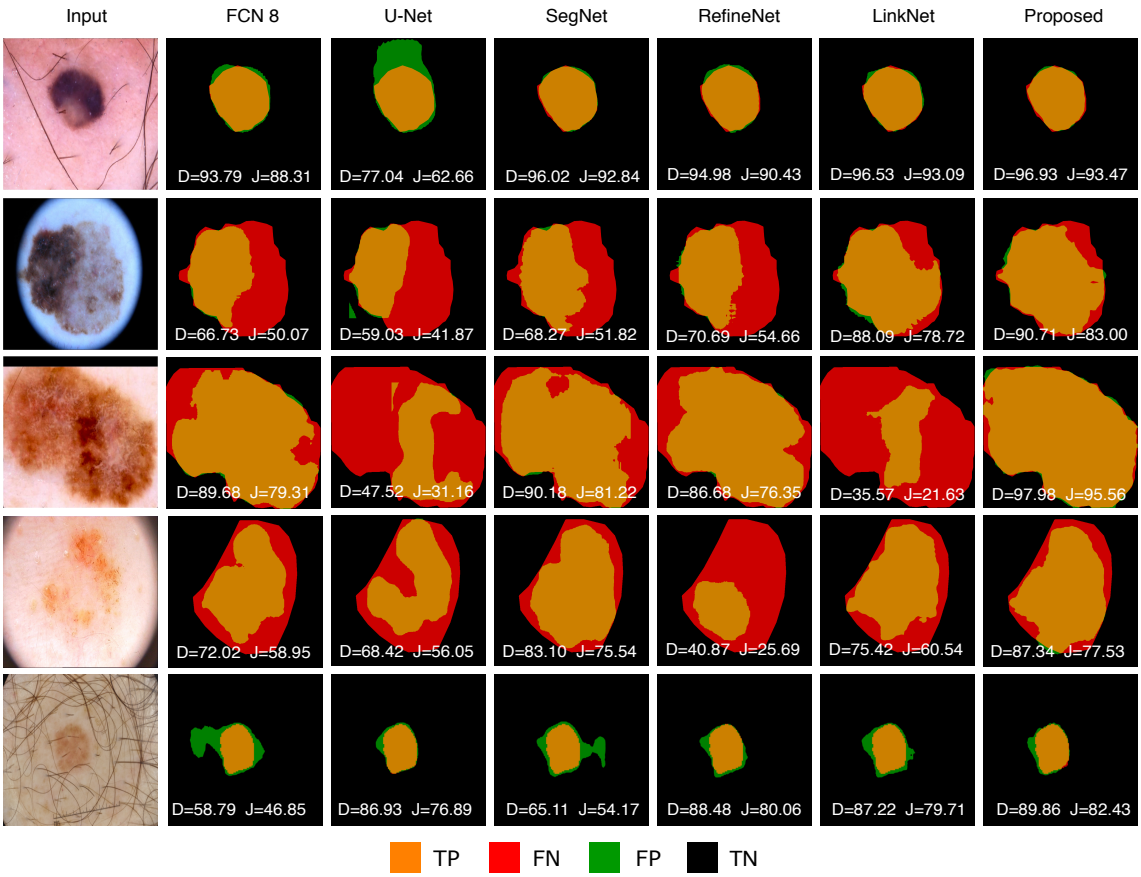


Figure 6.7: Skin lesion segmentation using the FCN8, UNet, SegNet, RefineNet, LinkNet and FCA-Net models. Note that D and J represent the Dice and IoU scores, respectively. Further visualization for the segmentation results of the proposed method can be found at https://youtu.be/GeUM8FglhFA.

with very dense hair in the background. For this sample, the proposed model obtains promising segmentation results compared to the other models.

6.4.3 Limitations

Although the proposed FCA-Net outperforms several deep learning-based models (FCN8, UNet, SegNet, RefineNet and LinkNet), it may produce inaccurate results with some cases as shown in Figure 6.8. As we can see, it is difficult to segment such images manually. The skin image of the first row has fuzzy boundaries, low contrast and intensity inhomogeneity. With this case, our model achieves a Dice of 84.21% and an IoU score of 78.56%. The other five tested models provide less accurate segmentation results, while the FCA-Net almost fit the lesion area. The second sample has two distinct color shades in the lesion. All models have failed to

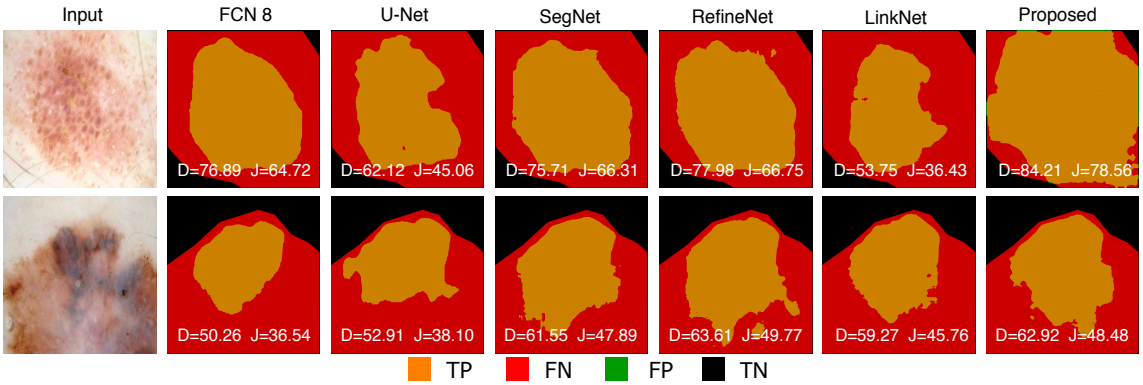


Figure 6.8: Examples of inaccurately segmented lesions with the proposed FCA-Net model and compared with other baseline segmentation models. Note that D and J represent the Dice and IoU scores, respectively.

accurately segment the lesion because of this dual shading, selecting the stronger one as the lesion and misclassifying the weaker one as background. It leads to IoU scores less than 50% in all segmentations, although FCA-Net has achieved the second best result.

6.5 Conclusion

In this chapter, we have proposed an accurate skin lesion segmentation model based on a generative adversarial network with the proposed FCA block that integrates a channel attention mechanism with residual 1-D factorized kernels convolutions. The FCA block noticeably improves the performance of our cGAN model as well as other well-known architectures (FCN8, UNet, etc.). We have run several qualitative and quantitative experiments that show how both channel attention and 1-D residual convolution mechanisms contribute to the segmentation improvement. Those mechanisms boost similar features across all channels and in neighboring regions of the encoder activation maps. As we expected, those similar features tend to correspond more robustly with relevant patterns of lesion/non-lesion areas.

Our model is fully automated and fully self-contained, in the sense that we did not use any data augmentation techniques in the training phase or pre-processing steps. The efficacy of the proposed model is assessed on three publicly available skin lesion segmentation challenge datasets: ISBI2016, ISBI2017, and ISIC2018. Our

model outperforms several state-of-the-art methods, such as FCN8, UNet, SegNet, ExB, CUMED, MResNet-Seg, and FrCN, in terms of Dice and IoU metrics.

CHAPTER 7

Concluding remarks

This final chapter presents the most important contributions and main conclusions of this dissertation, emphasizing their significance. Likewise, the chapter also includes approaches for future work.

7.1 Thesis highlights

The main contribution of this thesis is the method for segmenting different types of lesions in several modalities of medical images, which is based on conditional cGAN. This type of network is able to learn the intrinsic features of the lesions from a relatively low number (hundreds) of training samples, and then generate the corresponding image mask that selects the pixels belonging to the ill tissue. Numerous experiments performed in the thesis reveal that the proposed method segments very efficiently either breast tumors in mammograms and ultrasound

images, skin lesions in dermoscopic images or optic disc in fundus images. Numerical and visual comparisons indicate that the proposed method is better than other state-of-the-art segmentation methods in the majority of situations.

To address the problems due to the inherent sources of variability and uncertainty present in all types of medical images, the thesis also proposes several extensions to enhance the core cGAN, being the most noticeable the CAW block and the FCA block, which has improved significantly ($> 7\%$ in Jaccard index) the segmentation performance for breast tumors in ultrasound images and skin lesions in dermoscopic images, respectively. Other variants include different loss functions, which has lead to slight improvements.

Besides, the thesis also proposes several CNNs to classify the segmented regions into different types of findings, for example, the shape of a breast tumor (irregular, lobular, oval and round), its molecular subtype (Luminal A/B, Her-2 and Basal-like) and if it is benign or malignant. Although the results are not as outstanding as the segmentation ones (around 80% of classification accuracy compared to more than 90% in segmentation accuracy), they are still better than other state-of-the-art methods in these fields.

In summary, according to the obtained results it can be assessed that CNNs in general and cGANs in particular are highly reliable for the tasks of medical image segmentation and classification on different types of image modalities and different organs. One can expect that this statement will hold for other modalities and organs than the ones experimented in this thesis.

7.2 Future research lines

The work explained in this thesis and the achieved results permit to foresee future research lines. Some of them are studies that were omitted from the thesis due to time restrictions, and others are new predictions originated from issues that emerged during the outcome of the investigation.

Deep learning-based medical image analysis is always a challenging task that

requires a large amount of data and huge computational power. Designing a lightweight network with less number of parameters and the same degree of accuracy of our current model is an exciting direction for our research.

On the other hand, the medical image datasets to train the model are usually small in the number of samples. This situation reduces the generalization capacity of the model, which means that it may fail to correctly process test examples that exhibit small variations in the expected patterns. For example, the same body part of the same patient can be interpreted differently by the model when gathered with different image-acquisition machines. Another source of variability is to try to handle medical images from a different population of patients than the one used as a reference set. Our final target will be to design a robust deep learning architecture able to overcome such variability, thus truly increasing the efficacy of the CAD systems. Moreover, our future model should be suitable for being trained and process different modalities of medical images.

References

Abdel-Nasser, M., Melendez, J., Moreno, A., Omer, O.A., Puig, D., 2017. Breast tumor classification in ultrasound images using texture analysis and super-resolution methods. Engineering Applications of Artificial Intelligence 59, 84–92.

Akram, F., Kim, J., Lee, C., Choi, K.N., 2015. Segmentation of regions of interest using active contours with SPF function. Comp. Math. Methods in Medicine 2015, 710326:1–710326:14. doi:10.1155/2015/710326.

Al-antari, M.A., Al-masni, M.A., Choi, M.T., Han, S.M., Kim, T.S., 2018. A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. International Journal of Medical Informatics 117, 44–54.

Al-Bander, B., Al-Nuaimy, W., Williams, B.M., Zheng, Y., 2018. Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc. Biomedical Signal Processing and Control 40, 91–101.

Al-masni, M., Al-antari, M., Rivera, P., Valarezo, E., Gi, G., Kim, T., Park, H.,

Kim, T., 2018. Automatic skin lesion boundary segmentation using deep learning convolutional networks with weighted cross entropy .

Al-Masni, M.A., Al-antari, M.A., Choi, M.T., Han, S.M., Kim, T.S., 2018. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Computer methods and programs in biomedicine 162, 221–231.

Almazroa, A., Burman, R., Raahemifar, K., Lakshminarayanan, V., 2015. Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey. Journal of ophthalmology 2015.

Argenziano, G., Soyer, H.P., Chimenti, S., Talamini, R., Corona, R., Sera, F., Binder, M., Cerroni, L., De Rosa, G., Ferrara, G., et al., 2003. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the internet. Journal of the American Academy of Dermatology 48, 679–693.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 2481–2495.

Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M.J., Szeliski, R., 2011. A database and evaluation methodology for optical flow. IJCV 92, 1–31.

Berseth, M., 2017. Isic 2017-skin lesion analysis towards melanoma detection. arXiv preprint arXiv:1703.00523 .

Bi, L., Feng, D., Kim, J., 2018. Improving automatic skin lesion segmentation using adversarial learning based data argumentation. arXiv preprint arXiv:1807.08392 .

Bi, L., Kim, J., Ahn, E., Feng, D., 2017. Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv preprint arXiv:1703.04197 .

Bi, L., Kim, J., Ahn, E., Kumar, A., Feng, D., Fulham, M., 2019. Step-wise integration of deep class-specific learning for dermoscopic image segmentation. Pattern Recognition 85, 78–89.

Bisla, D., Choromanska, A., Stein, J.A., Polsky, D., Berman, R., 2019. Skin lesion segmentation and classification with deep learning system. arXiv preprint arXiv:1902.06061 .

Bissoto, A., Perez, F., Ribeiro, V., Fornaciali, M., Avila, S., Valle, E., 2018. Deep-learning ensembles for skin-lesion segmentation, analysis, classification: Recod titans at isic challenge 2018. arXiv preprint arXiv:1808.08480 .

Boisserie-Lacroix, M., Hurtevent-Labrot, G., Ferron, S., Lippa, N., Bonnefoi, H., Mac Grogan, G., 2013. Correlation between imaging and molecular classification of breast cancers. Diagnostic and interventional imaging 94, 1069–1080.

Bovis, K., Singh, S., 2000. Detection of masses in mammograms using texture features, in: Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, IEEE. pp. 267–270.

Cardoso, J.S., Domingues, I., Oliveira, H.P., 2015. Closed shortest path in the original coordinates with an application to breast cancer. International Journal of Pattern Recognition and Artificial Intelligence 29.

Cardoso, J.S., Marques, N., Dhungel, N., Carneiro, G., Bradley, A., 2017. Mass segmentation in mammograms: A cross-sensor comparison of deep and tailored features, in: Proceeding of the IEEE International Conference on Image Processing (ICIP), pp. 1737–1741.

Chaurasia, A., Culurciello, E., 2017. Linknet: Exploiting encoder representations for efficient semantic segmentation, in: 2017 IEEE Visual Communications and Image Processing (VCIP), IEEE. pp. 1–4.

Cheng, H., Cai, X., Chen, X., Hu, L., Lou, X., 2003. Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recognition 36, 2967 – 2991. URL: http://www.sciencedirect.com/science/article/pii/S0031320303001924, doi:https://doi.org/10.1016/S0031-3203(03)00192-4.

Cho, N., 2016. Molecular subtypes and imaging phenotypes of breast cancer. Ultrasonography 35, 281.

Chollet, F., et al., 2015. Keras: Deep learning library for theano and tensorflow.(2015). There is no corresponding record for this reference .

Chrástek, R., Wolf, M., Donath, K., Niemann, H., Paulus, D., Hothorn, T., Lausen, B., Lämmer, R., Mardin, C.Y., Michelson, G., 2005. Automated segmentation of the optic nerve head for diagnosis of glaucoma. Medical Image Analysis 9, 297–314.

Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., Halpern, A., 2019. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). `arXiv:1902.03368`.

Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., et al., 2018. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 168–172.

Day, G.R., Barbour, R.H., 2000. Automated melanoma diagnosis: where are we at? Skin Research and Technology 6, 1–5.

Dev, J., Dash, S.K., Dash, S., Swain, M., 2012. A classification technique for microarray gene expression data using pso-flann. International Journal on Computer Science and Engineering 4, 1534.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 .

Dhungel, N., Carneiro, G., Bradley, A., 2015a. Tree re-weighted belief propagation using deep learning potentials for mass segmentation from mammograms, in: 12th

IEEE International Symposium on Biomedical Imaging, ISBI 2015, Brooklyn, NY, USA, April 16-19, 2015, pp. 760–763. doi:10.1109/ISBI.2015.7163983.

Dhungel, N., Carneiro, G., Bradley, A.P., 2015b. Deep learning and structured prediction for the segmentation of mass in mammograms, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 605–612.

Dhungel, N., Carneiro, G., Bradley, A.P., 2017. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Medical image analysis 37, 114–128.

Döhler, F., Mormann, F., Weber, B., Elger, C.E., Lehnertz, K., 2008. A cellular neural network based method for classification of magnetic resonance images: towards an automated detection of hippocampal sclerosis. Journal of neuroscience methods 170, 324–331.

Du, X., Zhang, W., Zhang, H., Chen, J., Zhang, Y., Warrington, J.C., Brahm, G., Li, S., 2018. Deep regression segmentation for cardiac bi-ventricle mr images. IEEE Access 6, 3828–3838.

Erkol, B., Moss, R.H., Joe Stanley, R., Stoecker, W.V., Hvatum, E., 2005. Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. Skin Research and Technology 11, 17–26.

Falck, A.K., Fernö, M., Bendahl, P.O., Rydén, L., 2013. St gallen molecular subtypes in primary breast cancer and matched lymph node metastases-aspects on distribution and prognosis for patients with luminal a tumours: results from a prospective randomised trial. BMC cancer 13, 558.

Fang, Y., Xie, J., Dai, G., Wang, M., Zhu, F., Xu, T., Wong, E., 2015. 3d deep shape descriptor, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2319–2328.

Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018a. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. arXiv preprint arXiv:1801.00926 .

Fu, J., Liu, J., Tian, H., Fang, Z., Lu, H., 2018b. Dual attention network for scene segmentation. arXiv preprint arXiv:1809.02983 .

Fumero, F., Alayón, S., Sanchez, J., Sigut, J., Gonzalez-Hernandez, M., 2011. Rim-one: An open retinal image database for optic nerve evaluation, in: Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on, IEEE. pp. 1–6.

Galdran, A., Alvarez-Gila, A., Meyer, M.I., Saratxaga, C.L., Araújo, T., Garrote, E., Aresta, G., Costa, P., Mendonça, A.M., Campilho, A., 2017. Data-driven color augmentation techniques for deep skin image analysis. arXiv preprint arXiv:1703.03702 .

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: Advances in neural information processing systems, pp. 2672–2680.

Gutman, D., Codella, N.C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A., 2016. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1605.01397 .

Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., Zwiggelaar, R., 2018. Deep learning in mammography and breast histology, an overview and future trends. Medical image analysis 47, 45–67.

Harbeck, N., Thomssen, C., Gnant, M., 2013. St. gallen 2013: brief preliminary summary of the consensus discussion. Breast care 8, 102–109.

Hardie, R.C., Ali, R., De Silva, M.S., Kebede, T.M., 2018. Skin lesion segmentation

and classification for isic 2018 using traditional classifiers with hand-crafted features. arXiv preprint arXiv:1807.07001 .

Hazarika, M., Mahanta, L.B., 2018. A new breast border extraction and contrast enhancement technique with digital mammogram images for improved detection of breast cancer. Asian Pacific journal of cancer prevention: APJCP 19, 2141.

He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.

He, K., Zhang, X., Ren, S., Sun, J., 2016b. Deep residual learning for image recognition, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

He, X., Zhang, H., Landis, M., Sharma, M., Warrington, J., Li, S., 2017. Unsupervised boundary delineation of spinal neural foramina using a multi-feature and adaptive spectral segmentation. Medical image analysis 36, 22–40.

Hesamian, M.H., Jia, W., He, X., Kennedy, P., 2019. Deep learning techniques for medical image segmentation: Achievements and challenges. Journal of digital imaging , 1–15.

Hofer, C., Kwitt, R., Niethammer, M., Uhl, A., 2017. Deep learning with topological signatures, in: Advances in Neural Information Processing Systems, pp. 1634–1644.

Howell, A., Anderson, A.S., Clarke, R.B., Duffy, S.W., Evans, D.G., Garcia-Closas, M., Gescher, A.J., Key, T.J., Saxton, J.M., Harvie, M.N., 2014. Risk determination and prevention of breast cancer. Breast Cancer Research 16, 446.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141.

Hu, Y., Guo, Y., Wang, Y., Yu, J., Zhou, S., Chang, C., 2019. Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. Medical physics 46, 215–228.

Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. arXiv preprint .

Izadi, S., Mirikharaji, Z., Kawahara, J., Hamarneh, G., 2018. Generative adversarial networks to segment skin lesions, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 881–884.

Jahanifar, M., Tajeddin, N.Z., Asl, B.M., Gooya, A., 2019. Supervised saliency map driven segmentation of lesions in dermoscopic images. IEEE journal of biomedical and health informatics 23, 509–518.

Jeleń, Ł., Fevens, T., Krzyżak, A., 2008. Classification of breast cancer malignancy using cytological images of fine needle aspiration biopsies. International Journal of Applied Mathematics and Computer Science 18, 75–83.

Jiao, Z., Gao, X., Wang, Y., Li, J., 2018. A parasitic metric learning net for breast mass classification based on mammography. Pattern Recognition 75, 292–301.

Kim, S.T., Lee, H., Kim, H.G., Ro, Y.M., 2018. ICADx: Interpretable computer aided diagnosis of breast masses, in: Proceedings of the SPIE - Medical Imaging 2018: Computer-Aided Diagnosis.

Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J., 2017. Learning to discover cross-domain relations with generative adversarial networks, in: Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp. 1857–1865.

Kisilev, P., Sason, E., Barkan, E., Hashoul, S., 2016. Medical image description using multi-task-loss cnn, in: Deep Learning and Data Labeling for Medical Applications. Springer, pp. 121–129.

Kisilev, P., Walach, E., Hashoul, S.Y., Barkan, E., Ophir, B., Alpert, S., 2015. Semantic description of medical image findings: structured learning approach, in: Proceedings of the British Machine Vision Conference (BMVC), pp. 171.1–171.11.

Kozegar, E., Soryani, M., Minaei, B., Domingues, I., et al., . Assessment of a novel mass detection algorithm in mammograms .

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, pp. 1097–1105.

Kshema, George, M.J., Dhas, D.A.S., 2017. Preprocessing filters for mammogram images: A review, in: 2017 Conference on Emerging Devices and Smart Systems (ICEDSS), pp. 1–7. doi:10.1109/ICEDSS.2017.8073694.

Kurnianggoro, L., Jo, K.H., et al., 2018. A survey of 2d shape representation: Methods, evaluations, and future research directions. Neurocomputing 300, 1–16.

Lankton, S., Tannenbaum, A., 2008. Localizing region-based active contours. IEEE transactions on image processing 17, 2029–2039.

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Benbrahim-Tallaa, L., Bouvard, V., Bianchini, F., Straif, K., 2015. Breast-cancer screening-viewpoint of the IARC working group. New England Journal of Medicine 372, 2353–2358.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature 521, 436.

Lee, C.Y., Chen, G.L., Zhang, Z.X., Chou, Y.H., Hsu, C.C., 2018. Is intensity inhomogeneity correction useful for classification of breast cancer in sonograms using deep neural network? Journal of healthcare engineering 2018.

Li, H., He, X., Zhou, F., Yu, Z., Ni, D., Chen, S., Wang, T., Lei, B., 2019. Dense deconvolutional network for skin lesion segmentation. IEEE journal of biomedical and health informatics 23, 527–537.

Li, X., Yu, L., Chen, H., Fu, C.W., Heng, P.A., 2018. Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model. arXiv preprint arXiv:1808.03887 .

Lin, G., Milan, A., Shen, C., Reid, I., 2017. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1925–1934.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Medical image analysis 42, 60–88.

Liu, S., Wu, X.D., Xu, W.J., Lin, Q., Liu, X.J., Li, Y., 2016a. Is there a correlation between the presence of a spiculated mass on mammogram and luminal a subtype breast cancer? Korean journal of radiology 17, 846–852.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016b. Ssd: Single shot multibox detector, in: European conference on computer vision, Springer. pp. 21–37.

Liu, Z., Zhuo, C., Xu, X., 2018. Efficient segmentation method using quantised and non-linear cenn for breast tumour classification. Electronics Letters .

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.

Lowell, J., Hunter, A., Steel, D., Basu, A., Ryder, R., Fletcher, E., Kennedy, L., 2004. Optic nerve head segmentation. IEEE Transactions on medical Imaging 23, 256–264.

Luciano, L., Hamza, A.B., 2018. Deep learning with geodesic moments for 3d shape classification. Pattern Recognition Letters 105, 182–190.

MacGillivray, T., Trucco, E., Cameron, J., Dhillon, B., Houston, J., Van Beek, E., 2014. Retinal imaging as a source of biomarkers for diagnosis, characterization

and prognosis of chronic illness or long-term conditions. The British journal of radiology 87, 20130832.

Maier, A., Syben, C., Lasser, T., Riess, C., 2019. A gentle introduction to deep learning in medical image processing. Zeitschrift für Medizinische Physik 29, 86–101.

Maninis, K.K., Pont-Tuset, J., Arbeláez, P., Van Gool, L., 2016. Deep retinal image understanding, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 140–148.

Masoumi, M., Hamza, A.B., 2017. Spectral shape classification: A deep learning approach. Journal of Visual Communication and Image Representation 43, 198–211.

Materka, A., Strzelecki, M., et al., 1998. Texture analysis methods–a review. Technical university of lodz, institute of electronics, COST B11 report, Brussels , 9–11.

Matos, C.E.F., Souza, J.C., Diniz, J.O.B., Junior, G.B., de Paiva, A.C., de Almeida, J.D.S., da Rocha, S.V., Silva, A.C., 2018. Diagnosis of breast tissue in mammography images based local feature descriptors. Multimedia Tools and Applications , 1–26.

Mirikharaji, Z., Izadi, S., Kawahara, J., Hamarneh, G., 2018. Deep auto-context fully convolutional neural network for skin lesion segmentation, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 877–880.

Perou, C.M., Sørlie, T., Eisen, M.B., Van De Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al., 2000. Molecular portraits of human breast tumours. nature 406, 747.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. Pointnet: Deep learning on point

sets for 3d classification and segmentation. Proc. Computer Vision and Pattern Recognition (CVPR), IEEE 1, 4.

Rahman, M., Alpaslan, N., Bhattacharya, P., 2016. Developing a retrieval based diagnostic aid for automated melanoma recognition of dermoscopic images, in: 2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), IEEE. pp. 1–7.

Rangayyan, R.M., Banik, S., Desautels, J.L., 2010. Computer-aided detection of architectural distortion in prior mammograms of interval cancer. Journal of Digital Imaging 23, 611–631.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in neural information processing systems, pp. 91–99.

Rodrigues, P.S., 2017. Breast ultrasound image. Mendeley Data doi:`http://dx.doi.org/10.17632/wmy84gzngw.1`.

Romera, E., Alvarez, J.M., Bergasa, L.M., Arroyo, R., 2018. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. IEEE Transactions on Intelligent Transportation Systems 19, 263–272.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

Rouhi, R., Jafari, M., Kasaei, S., Keshavarzian, P., 2015. Benign and malignant breast tumors classification based on region growing and cnn segmentation. Expert Systems with Applications 42, 990–1002.

Sarker, M.M.K., Rashwan, H.A., Akram, F., Banu, S.F., Saleh, A., Singh, V.K., Chowdhury, F.U.H., Abdulwahab, S., Romani, S., Radeva, P., Puig, D., 2018. SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks, in: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, pp. 21–29. URL: https://doi.org/10.1007/978-3-030-00934-2_3, doi:10.1007/978-3-030-00934-2_3.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural networks 61, 85–117.

Shankaranarayana, S.M., Ram, K., Mitra, K., Sivaprakasam, M., 2017. Joint optic disc and cup segmentation using fully convolutional and adversarial networks, in: Fetal, Infant and Ophthalmic Medical Image Analysis. Springer, pp. 168–176.

Silveira, M., Nascimento, J.C., Marques, J.S., Marçal, A.R., Mendonça, T., Yamauchi, S., Maeda, J., Rozeira, J., 2009. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. IEEE Journal of Selected Topics in Signal Processing 3, 35–45.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Singh, V.K., Rashwan, H., Akram, F., Pandey, N., Sarker, M., Kamal, M., Saleh, A., Abdulwahab, S., Maaroof, N., Romani, S., et al., 2018a. Retinal optic disc segmentation using conditional generative adversarial network. arXiv preprint arXiv:1806.03905 .

Singh, V.K., Rashwan, H.A., Romani, S., Akram, F., Pandey, N., Sarker, M.M.K., Saleh, A., Arenas, M., Arquez, M., Puig, D., Torrents-Barrena, J., 2020. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. Expert Systems with Applications 139, 112855. URL: http://www.sciencedirect.com/science/article/pii/S0957417419305573, doi:https://doi.org/10.1016/j.eswa.2019.112855.

Singh, V.K., Romani, S., Rashwan, H.A., Akram, F., Pandey, N., Sarker, M.M.K., Abdulwahab, S., Torrents-Barrena, J., Saleh, A., Arquez, M., Arenas, M., Puig, D., 2018b. Conditional generative adversarial and convolutional networks for x-ray breast mass segmentation and shape classification, in: Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II, pp. 833–840. URL: https://doi.org/10.1007/978-3-030-00934-2_92, doi:10.1007/978-3-030-00934-2_92.

Singh, V.K., Romani, S., Torrents-Barrena, J., Akram, F., Pandey, N., Sarker, M.M.K., Saleh, A., Arenas, M., Arquez, M., Puig, D., 2017. Classification of breast cancer molecular subtypes from their micro-texture in mammograms using a vggnet-based convolutional neural network, in: Recent Advances in Artificial Intelligence Research and Development: Proceedings of the 20th International Conference of the Catalan Association for Artificial Intelligence, Deltebre, Terres de L'Ebre, Spain, October 25-27, 2017, IOS Press. p. 76.

Sivaswamy, J., Krishnadas, S., Chakravarty, A., Joshi, G., Tabish, A.S., et al., 2015. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomedical Imaging Data Papers 2, 1004.

Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L., 2016. Breast cancer histopathological image classification using convolutional neural networks, in: 2016 international joint conference on neural networks (IJCNN), IEEE. pp. 2560–2567.

Stewart, B., Wild, C.P., et al., 2014. World cancer report 2014 .

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging 35, 1299–1312.

Tamaki, K., Ishida, T., Miyashita, M., Amari, M., Ohuchi, N., Tamaki, N., Sasano, H., 2011. Correlation between mammographic findings and corresponding

histopathology: potential predictors for biological characteristics of breast diseases. Cancer science 102, 2179–2185.

Tang, J., Rangayyan, R.M., Xu, J., El Naqa, I., Yang, Y., 2009. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. IEEE Transactions on Information Technology in Biomedicine 13, 236–251.

Torrents-Barrena, J., Valls, A., Radeva, P., Arenas, M., Puig, D., 2015a. Automatic recognition of molecular subtypes of breast cancer in x-ray images using segmentation-based fractal texture analysis., in: CCIA, pp. 247–256.

Torrents-Barrena, J., Valls, Puig, D.A.M., Radeva, P., 2015b. Assessment of a multidiscrim-inant supervised classifier driven by textural features to distinguish molecular subtypes of breast cancer., in: CARS, pp. S31–S32.

Van Noord, N., Postma, E., 2017. Learning scale-variant and scale-invariant features for deep image classification. Pattern Recognition 61, 583–592.

Venkatesh, G., Naresh, Y., Little, S., OConnor, N.E., 2018. A deep residual architecture for skin lesion segmentation, in: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis. Springer, pp. 277–284.

Vesal, S., Ravikumar, N., Maier, A., 2018. Skinnet: A deep learning framework for skin lesion segmentation. arXiv preprint arXiv:1806.09522 .

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al., 2004. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13, 600–612.

Welfer, D., Scharcanski, J., Kitamura, C.M., Dal Pizzol, M.M., Ludwig, L.W., Marinho, D.R., 2010. Segmentation of the optic disk in color eye fundus images using an adaptive morphological approach. Computers in Biology and Medicine 40, 124–137.

Wong, A., Scharcanski, J., Fieguth, P., 2011. Automatic skin lesion segmentation via iterative stochastic region merging. IEEE Transactions on Information Technology in Biomedicine 15, 929–936.

Wong, M.T., He, X., Nguyen, H., Yeh, W.C., 2012. Mass classification in digitized mammograms using texture features and artificial neural network, in: International Conference on Neural Information Processing, Springer. pp. 151–158.

Xiong, W., Wu, L., Alleva, F., Droppo, J., Huang, X., Stolcke, A., 2018. The microsoft 2017 conversational speech recognition system, in: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE. pp. 5934–5938.

Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., Carson, P.L., 2019. Medical breast ultrasound image segmentation by machine learning. Ultrasonics 91, 1–9.

Xue, Y., Xu, T., Huang, X., 2018. Adversarial learning with multi-scale loss for skin lesion segmentation, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 859–863.

Yang, D., Xu, D., Zhou, S.K., Georgescu, B., Chen, M., Grbic, S., Metaxas, D., Comaniciu, D., 2017. Automatic liver segmentation using an adversarial image-to-image network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 507–515.

Yang, W., Zhang, S., Chen, Y., Li, W., Chen, Y., 2008. Measuring shape complexity of breast lesions on ultrasound images, in: Medical Imaging 2008: Ultrasonic Imaging and Signal Processing, International Society for Optics and Photonics. p. 69200J.

Yassin, N.I., Omran, S., El Houby, E.M., Allam, H., 2018. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. Computer methods and programs in biomedicine 156, 25–45.

Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 .

Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A., 2017. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE transactions on medical imaging 36, 994–1004.

Yuan, Y., 2017. Automatic skin lesion segmentation with fully convolutional-deconvolutional networks. arXiv preprint arXiv:1703.05165 .

Zhou, H., Li, X., Schaefer, G., Celebi, M.E., Miller, P., 2013. Mean shift based gradient vector flow for image segmentation. Computer Vision and Image Understanding 117, 1004–1016.

Zhu, W., Xiang, X., Tran, T.D., Hager, G.D., Xie, X., 2018. Adversarial deep structured nets for mass segmentation from mammograms, in: Proceedings of the IEEE 15^{th} International Symposium on Biomedical Imaging (ISBI), pp. 847–850.

Zilly, J.G., Buhmann, J.M., Mahapatra, D., 2015. Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images, in: International Workshop on Machine Learning in Medical Imaging, Springer. pp. 136–143.

UNIVERSITAT ROVIRA I VIRGILI
SEGMENTATION AND CLASSIFICATION OF MULTIMODAL MEDIAL IMAGES BASED ON GENERATIVE ADVERSARIAL LEARNING AND CONVOLUTIONAL NEURAL NETWORKS
Vivek Kumar Singh

UNIVERSITAT ROVIRA I VIRGILI
SEGMENTATION AND CLASSIFICATION OF MULTIMODAL MEDIAL IMAGES BASED ON GENERATIVE ADVERSARIAL LEARNING AND CONVOLUTIONAL NEURAL NETWORKS
Vivek Kumar Singh

UNIVERSITAT ROVIRA I VIRGILI
SEGMENTATION AND CLASSIFICATION OF MULTIMODAL MEDIAL IMAGES BASED ON GENERATIVE ADVERSARIAL LEARNING AND CONVOLUTIONAL NEURAL NETWORKS
Vivek Kumar Singh