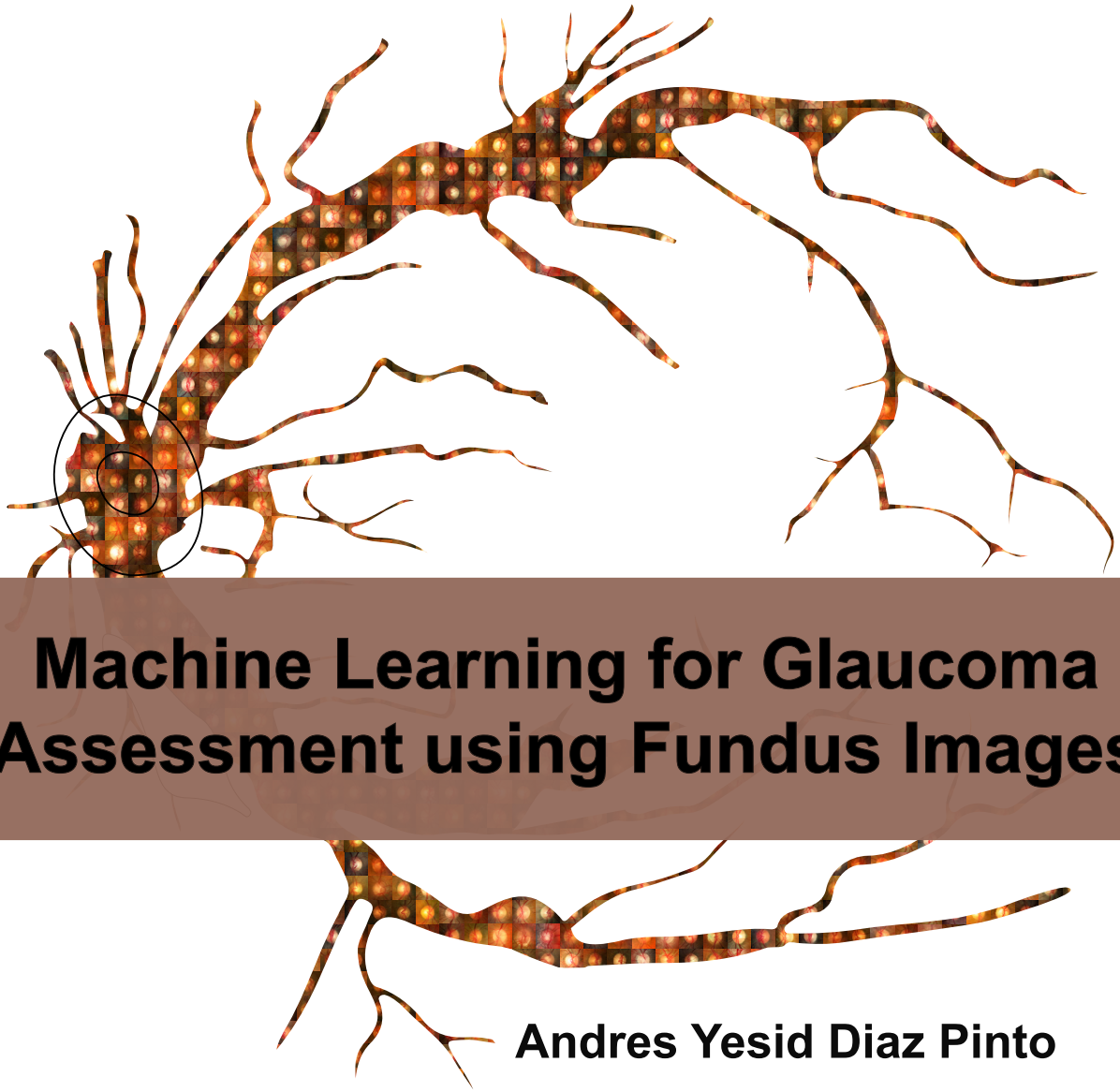


Machine Learning for Glaucoma Assessment using Fundus Images

Andres Yesid Diaz Pinto

PhD Thesis

June 2019

Department of Electronic Engineering

Supervisors

Dr Valery Naranjo Ornedo

Dr Sandra Morales Martinez

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Machine Learning for Glaucoma Assessment using Fundus Images

June 2019

Author: Andres Yesid Diaz Pinto

Supervisors: Dr Valery Naranjo Ornedo
Dr Sandra Morales Martínez

Departamento de Ingeniería Electrónica
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

The research derived from this doctoral thesis has been supported by the Generalitat Valenciana under the scholarship Santiago Grisolía [GRISOLIA/2015/027].

Acknowledgements

Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible to do without the emotional and intellectual support and guidance that I received from many people.

- Firstly, I would like to express my sincere gratitude to *Valery* and *Sandra*, my research supervisors, for their patient guidance and useful critiques of this research work. Many thanks for giving me the great opportunity to be part of this group.
- Besides my supervisors, I would like to thank Amparo Navea, the clinical expert who collaborated in the development of this thesis. Her contribution from the clinical point of view was invaluable.
- Quisiera también agradecer enormemente a toda mi familia. En especial a mi madre, mi padre, mis hermanas y los niños (Felipe, Jeronimo y Santiago). Quienes han sido el motor de mi pasión por seguir adelante y desarrollarme como persona. Quienes han estado muy pendientes de mi y de cada paso en este camino.
- I would also like to say a heartfelt thank you to Morenita, who has encouraged me during the hard times and have helped me each time I needed. Thank you Morenita, without you, this would not have been possible.
- Mis agradecimientos también para mi familia en España, Mari Carmen, Julio, Diego y Mireya. Muchas gracias por estar ahí.

- Quiero agradecer también a Xiomara por su incondicional ayuda durante toda mi estadía en Valencia. Muchas gracias por todo los momentos compartidos y por hacerme sentir en casa.
- I am also very grateful to all those at LME in Erlangen and CISTIB in Sheffield. Thanks for all you guys have done for me.
- Last but not the least, I thank my fellow labmates Jose, Felix, Adri, Fer, Reinier, Fran, Rober, Adri B. and Jorge. Without you guys, this experience would not have been the same. Many thanks!

Abstract

Fundus images are widely used by ophthalmologists to assess the retina and detect glaucoma, which is, according to studies from the World Health Organization (WHO), the second cause of blindness worldwide.

In this thesis, machine learning algorithms for automatic glaucoma assessment using fundus images are studied. First, two methods for automatic segmentation are proposed. The first method uses the Stochastic Watershed transformation to segment the optic cup and measures clinical features such as the Cup/Disc ratio and ISNT rule. The second method is a U-Net architecture focused on the optic disc and optic cup segmentation task.

Secondly, automated glaucoma assessment systems using convolutional neural networks (CNNs) are presented. In this approach, different ImageNet-trained models are fine-tuned and used as automatic glaucoma classifiers. These new techniques allow detecting glaucoma without previous segmentation or feature extraction. Moreover, it improves the performance of other state-of-art works.

Thirdly, given the difficulty of getting large amounts of glaucoma-labelled images, this thesis addresses the problem of retinal image synthesis. Two different architectures for image synthesis, the Variational Autoencoder (VAE) and Generative Adversarial Networks (GAN) architectures, were analysed. Using these models, synthetic images that were qualitative and quantitative analysed, reporting state-of-the-art performance, were generated.

Finally, an adversarial model is used to create an alternative automatic glaucoma assessment system. In this part, a semi-supervised learning algorithm was implemented to reach this goal.

Resumen

Las imágenes de fondo de ojo son muy utilizadas por los oftalmólogos para la evaluación de la retina y la detección de glaucoma. Esta patología es la segunda causa de ceguera en el mundo, según estudios de la Organización Mundial de la Salud (OMS).

En esta tesis doctoral, se estudian algoritmos de aprendizaje automático (machine learning) para la evaluación automática del glaucoma usando imágenes de fondo de ojo. En primer lugar, se proponen dos métodos para la segmentación automática. El primer método utiliza la transformación Watershed Estocástica para segmentar la copa óptica y posteriormente medir características clínicas como la relación Copa/Disco y la regla ISNT. El segundo método es una arquitectura U-Net que se usa específicamente para la segmentación del disco óptico y la copa óptica.

A continuación, se presentan sistemas automáticos de evaluación del glaucoma basados en redes neuronales convolucionales (CNN por sus siglas en inglés). En este enfoque se utilizan diferentes modelos entrenados en ImageNet como clasificadores automáticos de glaucoma, usando fine-tuning. Esta nueva técnica permite detectar el glaucoma sin segmentación previa o extracción de características. Además, este enfoque presenta una mejora considerable del rendimiento comparado con otros trabajos del estado del arte.

En tercer lugar, dada la dificultad de obtener grandes cantidades de imágenes etiquetadas (glaucoma/no glaucoma), esta tesis también aborda el problema de la síntesis de imágenes de la retina. En concreto se analizaron dos arqui-

tecturas diferentes para la síntesis de imágenes, las arquitecturas Variational Autoencoder (VAE) y la Generative Adversarial Networks (GAN). Con estas arquitecturas se generaron imágenes sintéticas que se analizaron cualitativa y cuantitativamente, obteniendo un rendimiento similar a otros trabajos en la literatura.

Finalmente, en esta tesis se plantea la utilización de un tipo de GAN (DCGAN) como alternativa a los sistemas automáticos de evaluación del glaucoma presentados anteriormente. Para alcanzar este objetivo se implementó un algoritmo de aprendizaje semi-supervisado.

Resum

Les imatges de fons d'ull són molt utilitzades pels oftalmòlegs per a l'avaluació de la retina i la detecció de glaucoma. Aquesta patologia és la segona causa de ceguesa al món, segons estudis de l'Organització Mundial de la Salut (OMS).

En aquesta tesi doctoral, s'estudien algoritmes d'aprenentatge automàtic (machine learning) per a l'avaluació automàtica del glaucoma usant imatges de fons d'ull. En primer lloc, es proposen dos mètodes per a la segmentació automàtica. El primer mètode utilitza la transformació Watershed Estocàstica per segmentar la copa òptica i després mesurar característiques clíniques com la relació Copa/Disc i la regla ISNT. El segon mètode és una arquitectura U-Net que s'usa específicament per a la segmentació del disc òptic i la copa òptica.

A continuació, es presenten sistemes automàtics d'avaluació del glaucoma basats en xarxes neuronals convolucionals (CNN per les sigles en anglès). En aquest enfocament s'utilitzen diferents models entrenats en ImageNet com classificadors automàtics de glaucoma, usant fine-tuning. Aquesta nova tècnica permet detectar el glaucoma sense segmentació prèvia o extracció de característiques. A més, aquest enfocament presenta una millora considerable del rendiment comparat amb altres treballs de l'estat de l'art.

En tercer lloc, donada la dificultat d'obtenir grans quantitats d'imatges etiquetades (glaucoma/no glaucoma), aquesta tesi també aborda el problema de la síntesi d'imatges de la retina. En concret es van analitzar dues arquitectures diferents per a la síntesi d'imatges, les arquitectures Variational Autoencoder

(VAE) i la Generative adversarial Networks (GAN). Amb aquestes arquitectures es van generar imatges sintètiques que es van analitzar qualitativament i quantitativament, obtenint un rendiment similar a altres treballs a la literatura.

Finalment, en aquesta tesi es planteja la utilització d'un tipus de GAN (DCGAN) com a alternativa als sistemes automàtics d'avaluació del glaucoma presentats anteriorment. Per assolir aquest objectiu es va implementar un algoritme d'aprenentatge semi-supervisat.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents the motivations behind this thesis, its objectives and its main contributions. Additionally, the thesis framework and the thesis outline are also presented

1.1 Motivation

Glaucoma is an irreversible neuro-degenerative eye disease that is considered one of the main reasons of visual disability in the world (World Health Organization 2004). According to the World Health Organization (WHO), glaucoma affects more than 65 million people around the globe (Bourne 2006). As it may be asymptomatic, early detection and treatment are important to prevent vision loss. This silent eye disease is mainly characterized by optic nerve fibre loss and that is given by the increased intraocular pressure (IOP) and/or loss of blood flow to the optic nerve. However, IOP measurement is found to be neither specific nor sensitive enough to be an effective glaucoma indicator since visual damage can be present without increased IOP. For that reason, a comprehensive glaucoma exam should also include an analysis of the retina using images and a field of vision test.

Fundus images are an imaging modality that is extensively used for the detection of glaucoma. Its simplicity, low cost of their acquisition process and non-invasive nature are the most important advantages. However, expert analysis of these images is expensive for large scale screening and the diagnostics vary significantly from one physician to another. Therefore, the development

of automatic glaucoma screening systems using fundus images are being actively explored for the scientific research community. These systems take as input a fundus image and apply a series of computations to provide feedback to ophthalmologists, including a labelling of the regions of interest such as the optic disc or give a probability indicating the level of risk of the patient from suffering glaucoma. These algorithms allow to improve accuracy in detecting the disease and reduce the intra-expert variability.

1.2 Objectives

The main goal of this thesis is the development of new methods that help ophthalmologists to automatically detect glaucoma by using retinal fundus images. In particular, the main objectives of this work are:

- To study the state-of-the-art of segmentation techniques and automatic glaucoma assessment algorithms using retinal fundus images.
- To propose new algorithms to segment and extract clinical features using retinal fundus images with the aim of automatically detecting glaucoma.
- Development and implementation of algorithms based on machine learning and/or deep learning that help ophthalmologists to detect glaucoma by using retinal fundus images.

1.3 Main Contributions

The main contribution of this thesis is the development of different automatic glaucoma algorithms based on classical and deep learning techniques. Each contribution will be described as follows:

Our first contribution is described in Section 4.1, and consists in an automatic segmentation algorithm. Using previous optic disc segmentation, this algorithm is based on the use of stochastic watershed transformation for optic cup detection in retinal fundus images. After that, clinical indicators in the optic nerve such as the Cup/Disc ratio (CDR), the area Cup/Disc ratio (ACDR) and the ISNT rule, that checks the disc rim thickness, are also analysed, which are the same clinical indicators used by experts to determine if a person has glaucoma. This contribution is based on the following publications:

- Diaz-Pinto A, Morales S, Naranjo V, Alcocer P, Lanzagorta A. Glaucoma Diagnosis by Means of Optic Cup Feature Analysis in Color Fundus Im-

ages. 24th European Signal Processing Conference (EUSIPCO). August 2016.

- Diaz-Pinto A, Morales S, Naranjo V, Alcocer P, Lanzagorta A. Diagnóstico Automático del Glaucoma a través de la Segmentación y Análisis de la Copa Óptica Usando Imágenes de Fondo de Ojo. XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica. pp 383-386. 2016

- Vesal S, Diaz-Pinto A, Ravikumar N, Ellman S, Davari A, Maier A. Semi-Automatic Algorithm for Breast MRI Lesion Segmentation Using Marker-Controlled Watershed Transformation. Nuclear Science Symposium and Medical Imaging Conference. October 2017.

- Diaz-Pinto A, Morales S, Naranjo V, Navea A. Computer-aided Glaucoma Diagnosis using Stochastic Watershed Transformation on Single Fundus Images. Accepted for publication at the Journal of Medical Imaging and Health Informatics.

After a brief introduction to the CNN architectures (Section 4.2.1), our second contribution is discussed in Section 4.2.2. We proposed a deep learning method for optic disc and optic cup segmentation. Using the U-Net network, we obtained competitive results that were used for the REFUGE challenge, hosted at the MICCAI 2018 conference in conjunction with OMIA workshop. (`https://refuge.grand-challenge.org/Home/`, Group name: Cvblab (Orlando, Fu, et al. 2019)).

Our third contribution, presented in Section 5.1, is a study of five different CNN architectures fine-tuned for the glaucoma assessment task using fundus images. The obtained results from this work show the robustness of the proposed method. Additionally, we present the biggest glaucoma-labelled database that can be found in the literature.

Details of this contribution were accepted for publication as a research paper for the BioMedical Engineering OnLine journal. This paper is titled "CNNs for Automatic Glaucoma Assessment using Fundus Images: An Extensive Validation".

In our fourth contribution, presented in Section 5.3, we used the variational autoencoder and a generative adversarial network to generate synthetic retinal images. We quantitatively compared synthetic and real images by using the portion of vessels, optic disc and background occupied by each type of image. For qualitative evaluation, we used a web application (`https:`

//cvblab.synology.me/ganval/) in which experts were able to label the images into real or fake. Therefore, we could know how similar are the synthetic and the real images.

This contribution is based on this paper:

- Diaz-Pinto A, Colomer A, Naranjo V, Morales S, Xu Y, Frangi A F. Retinal Image Synthesis for Glaucoma Assessment using DCGAN and VAE Models, 19th International Conference on Intelligent Data Engineering and Automated Learning. Nov 2018. pp 224-232

The last contribution of this thesis, presented in Section 5.3, is a study of an image synthesizer and a semi-supervised learning method for glaucoma assessment using cropped retinal fundus images. These two systems were trained on 86926 retinal images cropped around the optic disc using the Deep Convolutional Generative Adversarial Network (DCGAN). We qualitatively compared the synthetic images generated by other state-of-the-art work and the real images with our synthetic images by using t-SNE plots. Moreover, a quantitative evaluation was carried out by analyzing the structural properties of synthetic and real images. We also compared the performance of the proposed glaucoma classifier obtained from the semi-supervised learning method with the state-of-the-art algorithms.

An article with all the contributions made in this chapter was accepted for publication as a research paper for the IEEE Transactions on Medical Imaging journal. This article is titled "Retinal Image Synthesis and Semi-supervised Learning for Glaucoma Assessment".

1.4 Framework

This thesis stands within the framework of the research project "Fundus image processing for automatic screening of ophthalmological diseases - ACRIMA" (TIN2013-46751-R). The main goal of ACRIMA is to develop an automatic screening system for three of the most significant diseases related to permanent blindness: glaucoma, diabetic retinopathy and age-related macular degeneration. However, as previously mentioned, this thesis is focused on the development of automatic algorithms for glaucoma detection.

1.5 Outline

This thesis is divided into six chapters. In this chapter, we presented the motivations behind the research involved in this thesis, the main objectives, the main contributions, and the thesis framework.

Chapter 2 introduces the human retina anatomy, the types and the main characteristics of glaucoma and finally the most used imaging modalities. We also present the mainly used approaches for automated detection of glaucoma.

Chapter 3 presents all the public and private databases that were used to train, validate and test the methods proposed in this thesis.

Chapter 4 presents the theory related to the methods used for optic disc and optic cup segmentation. We present a method for optic cup segmentation and further glaucoma classification by using the stochastic watershed transformation. Additionally, we present a brief introduction to the CNNs and a new method based on deep learning to segment the optic disc and optic cup. Specifically, we used a modified U-Net network and contrast histogram equalization algorithm as a preprocessing technique to obtain the optic disc and optic cup masks. Finally, results obtained from the methods used for optic disc and optic cup segmentation are also presented. Advantages and disadvantages of each method are also discussed.

In Chapter 5 we study deep learning techniques for glaucoma classification and retinal images synthesis. First, we present the performance of five ImageNet-trained convolutional neural networks (CNNs) used as glaucoma classifiers. We analyse each CNN and its architecture and the images used for this task. Secondly, we present the theory behind the variational autoencoder (VAE) and a generative adversarial network (GAN) to be used as retinal images synthesizers. Finally, as our last contribution, we present a retinal image synthesizer and a semi-supervised learning algorithm for glaucoma assessment using cropped retinal images. We also present and discuss the results obtained from these methods in the last sections of this chapter.

Finally, we conclude the thesis presenting its main conclusions in Chapter 6. We also provide further research lines that could be derived from our contributions.

Chapter 2

Background

This chapter presents a brief explanation about anatomical aspects of the human retina, the main characteristics of the glaucoma disease, how it affects the human retina, the main imaging modalities to analyse the retina and the state-of-the-art algorithms developed for glaucoma assessment.

2.1 Anatomy of the Retina

The eye is a complex organ comprising a great number of structures that work together to provide the sense of sight. It is divided into two main parts, the anterior and the posterior part.

The anterior part is visible from outside and is composed by the cornea, the pupil, the iris and the lens. The posterior part is composed of the vitreous humour and a wall made up of three different layers of tissue, the fibrous (cornea and sclera), the vascular (blood vessels) and the inner layer called the retina (See Figure 2.1).

The retina is a thin layer of tissue that lines the back of the eye, it is approximately 0.2 mm thick, and has an area of approximately 1100 mm^2. The retina processes the information gathered by the photoreceptor cells (rods and cones) and sends this information to the brain via the optic nerve.

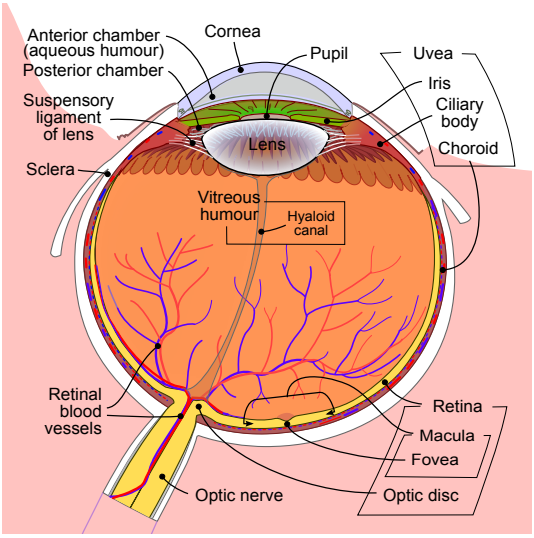


Figure 2.1: Schematic diagram of the eye. Anterior and posterior section of the eye (Orlando 2017; Rhcastilhos and Jmarchn 2007).

The retina is anatomically composed by three main structures: the optic disc, the vascular network, the macula and the fovea. The optic disc, also called the optic nerve, is the cable of nerve fibres that carries the electrical signals from the retina to the brain. Inside the optic disc is located the optic cup, which is characterised by the brightest area in a fundus image. The vascular network is composed by arteries and veins that supply of oxygen and food the retina. The macula is a yellow region which is specialized in the fine detail view located in the posterior part of the retina. The fovea, located in the centre of the macula, is a small pit (approximately 500 μm in diameter) that contains no rods and has the densest concentration of cones in the eye. (See Figure 2.2).

2.2 Types of Glaucoma

Glaucoma refers to a group of clinical diseases that share a common characteristic, consisting of a deepening or excavation of the optic nerve head. The main forms of glaucoma are open-angle, angle closure, and congenital glaucoma. The open-angle glaucoma is also called primary or chronic glaucoma because it is the most common type of glaucoma (at least 90% of all glaucoma cases). Although the majority of open-angle glaucoma is age-related and the risk of having glaucoma increases as we get older, this eye disease can affect

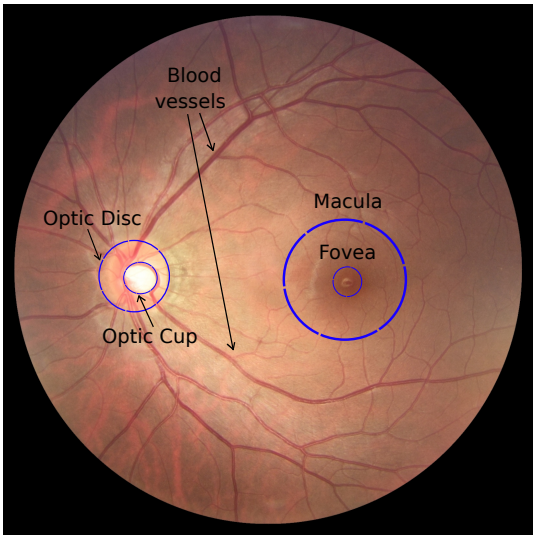


Figure 2.2: Main structures of the retina: blood vessels, optic disc, macula and fovea.

people of all ages, even newborns. Open-angle glaucoma is caused by the slow clogging of the drainage canals, resulting in increased eye pressure with symptoms and damage that are not noticed (Weinreb and Khaw 2004; Weinreb, Aung, and Medeiros 2014). "Open-angle" means that the angle where the iris meets the cornea is as wide and open as it should be (See Figure 2.3).

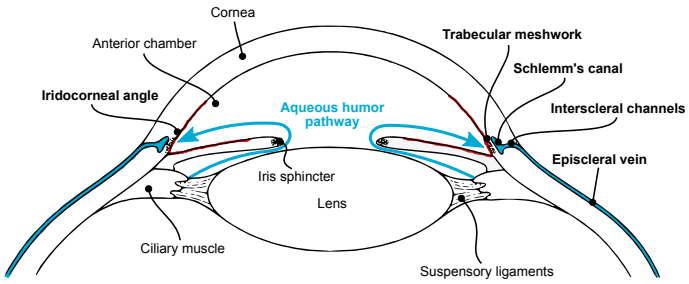


Figure 2.3: Schematic representation of the aqueous humor flow. The light-blue lines represent the aqueous humor pathways and the red lines are the iridocorneal angles.

The angle closure glaucoma is a less common form of glaucoma that is caused by blocked drainage canals, resulting in a sudden rise in intraocular pressure. It is a type of glaucoma that presents very noticeable symptoms and damage. It is also called acute glaucoma or narrow-angle glaucoma.

The congenital glaucoma occurs in babies when there is incorrect or incomplete development of the eye's drainage canals during the prenatal period. This is a rare condition that may be inherited. When it is uncomplicated, microsurgery can often correct the structural defects. Other cases are treated with medication and surgery.

All these forms make glaucoma the second leading cause of blindness around the globe after cataract. This eye disease could be treated by lowering the intraocular pressure (IOP), accomplished by daily eye drop administration, laser treatment to the eye, or ocular surgery (Quigley 2018).

People with glaucoma will slowly lose their peripheral (side) vision and may miss objects to the side. They seem to be looking through a tunnel. As an example, in Figure 2.4 is possible to see the differences between people with a normal vision and people with glaucoma.

(a) (b)

Figure 2.4: A scene as it might be viewed by a (a) person with normal vision (National Eye Institute, NIH 2012b) and a (b) person with glaucoma (National Eye Institute, NIH 2012a).

2.3 Imaging Technology

The most used imaging modalities to analyze the retina and make a glaucoma assessment are fundus photograph and optical coherence tomography (OCT). These two imaging modalities are briefly described in the following sections.

2.3.1 Fundus photographs

The fundus photograph or fundus image is acquired by a fundus camera (mydriatic or non-mydriatic) that has a digital camera. The digital camera operates in the same mode as a conventional camera using an image sensor(Patton et al. 2006) (See Figure 2.5).

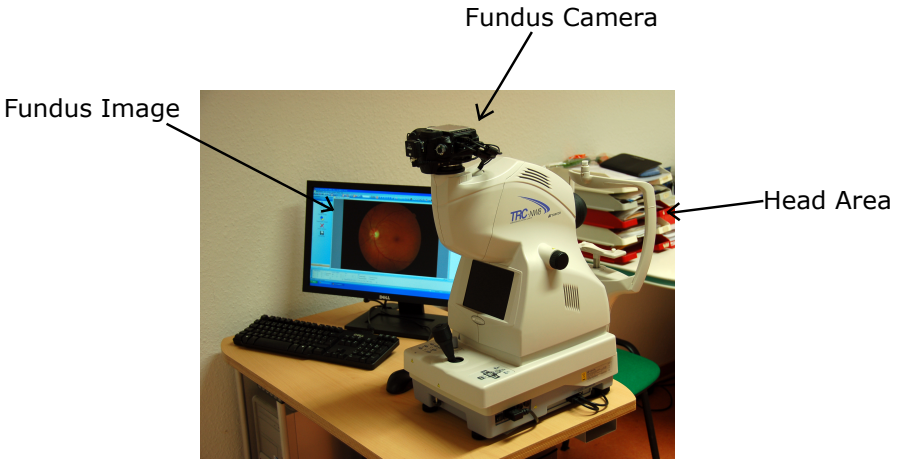


Figure 2.5: Fundus camera (Roletschek 2010)

A fundus photograph is used to record the appearance of a patient's retina. It allows to study the retina and to detect retinal diseases such as macular degeneration, diabetic retinopathy and glaucoma. A fundus photograph could also be used to evaluate irregularities in the fundus, to monitor the progression of a disease and to analyse the therapeutic outcome. The acquisition process is non-invasive and relatively easy to perform. It is the most economical imaging modality for diagnosing retinal diseases. An example of a fundus photograph is shown in Figure 2.2

In a fundus photograph the first plane represents the red pixel intensities, the second plane represents the green pixel intensities and the third plane represents the blue pixel intensities. Although many techniques use the grey-

scale image to extract features or assess images, the grey-scale image needs to be extracted from an RGB colour image.

One of the main disadvantages of fundus photographs is that they are a 2D projection of a 3D object (retina). Representing the retina as a 2D object using a colour fundus image makes the automatic glaucoma assessment process particularly difficult. However, its very low cost in comparison with other imaging modalities, make them useful for screening a large population (Abràmoff and Niemeijer 2015).

2.3.2 Optical Coherence Tomography

Optical Coherence Tomography (OCT) has emerged in importance because of the wide variety of information it can provide, its high resolution, and the complex 3-dimensional (3D) data it is able to gather (Geitzenauer, Hitzenberger, and U. M. Schmidt-Erfurth 2011; Wojtkowski et al. 2005). It can be used to scan through the layers of structured transparent and translucent tissue samples such as the retina with very high axial (depth) resolution (3-20μm). In particular, it measures the echo time delay and magnitude of backscattered or -reflected light to build up two- and three-dimensional images with a resolution of $\sim$ 20 μm (transversal) $\times$ 5 μm (axial) for commercial systems, and of 2-3 μm (isotropic) for high-end research systems (Pircher, Hitzenberger, and U. Schmidt-Erfurth 2011).

OCT images use the backscattered light to depict variations in optical reflectance through the depth of the retina along a point, creating what is known as an A-scan (Drexler and Fujimoto 2008; Schuman 2008; Arevalo, Krivoy, and Fernandez 2009). These A-scans through the retina can be gathered linearly or in a different scan protocol, making one cross-sectional image, which is known as a B-scan or tomogram. A collection of parallel B-scans can be used to gather a 3D data set as it is shown in Figure 2.6.

Nowadays, there are two OCT technologies available. The traditional time domain optical coherence tomography (TD-OCT) and the spectral domain OCT (SD-OCT). The TD-OCT obtains the depth information of the retina after a longitudinal translation in time of a reference arm, meanwhile, the SD-OCT measures the detected backscattered light as a function of optical frequencies, allowing for imaging speeds 50 times faster than TD-OCT, and providing a greater number of images per unit area.

While SD-OCT is the current commercially available state-of-the-art, new OCT technologies that improve acquisition speed, axial and transverse res-

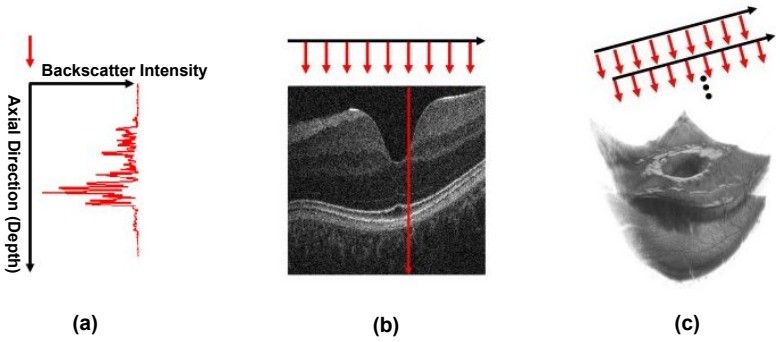


Figure 2.6: Types of OCT images. (a) One example of an A-scan, (b) One example of a B-scan and (c) One example of a C-scan.

olution of images are also under investigation. For example, technologies with much higher acquisition speeds and axial resolution are the swept-source OCT (SS-OCT), which is another Fourier domain technique for OCT and differs from SD-OCT in the method used to rapidly acquire data (Huber, Adler, and Fujimoto 2006; B. Liu and Brezinski 2007; Srinivasan et al. 2007; Potsaid et al. 2010; J. Zhang, Rao, and Z. Chen 2005; Choma, Hsu, and Izatt 2010), and the Polarization-sensitive OCT (PS-OCT) technology which can detect three parameters simultaneously: reflectivity, retardation and optic axis orientation (Drexler and Fujimoto 2008; Gabriele et al. 2011; Pircher, Götzinger, et al. 2004; Pircher, Hitzenberger, and U. Schmidt-Erfurth 2011).

Current OCT devices

The commercialization of OCT technology was critical for its clinical acceptance and utilization in ophthalmology because of its high cost associated with the development and engineering required to produce a clinical instrument from a research prototype. Fortunately, after the first commercial OCT instrument (Zeiss OCT) was introduced in 1996 with an axial image resolution of 10μm and a speed of 100 axial scans/s (A-scans), there have been significant advances in this technology.

Currently, available SD-OCT instruments are:

- Cirrus HD-OCT (Carl Zeiss Meditec, Inc, Dublin, California)

- Spectralis SD-OCT (Heidelberg Engineering GmbH, Heidelberg, Germany)
- RTVue (Optovue, Inc, Fremont, California)
- 3D-OCT 1000/200 (Topcon Medical Systems, Inc, Paramus, New Jersey)
- Biopitgen Envisu SD-OCT (Bioptigen, Inc, Research Triangle Park, North Carolina)
- SOCT Copernicus HR (Optopol Technology, SA, Zawiercie, Poland). This instrument is still awaiting FDA approval.

Each instrument has unique features and software that allow not only circular scans, line scans and macular cube-volume scans but can also include eye tracking, optic disc parameters, deviation map, enhanced depth imaging or anterior segment capabilities.

Although this image modality allows more accurate clinical measurements, it is prohibitively expensive for mass screening. For that reason, all the contributions of this thesis have been developed using fundus images, which are cheaper for the examination of a large population.

2.4 Retinal Image Analysis for Automatic Glaucoma Assessment

In this section, we summarize the most significant works in the literature for optic disc and optic cup segmentation, as well as automatic glaucoma diagnosis and retinal image synthesis methods.

2.4.1 Optic Disc and Optic Cup Segmentation

The detection process of glaucoma, using fundus images, usually involves the optic disc and optic cup segmentation, which helps ophthalmologists to register changes in the retina. For that reason, several approaches for localisation and segmentation of the optic disc and optic cup have been presented in the literature.

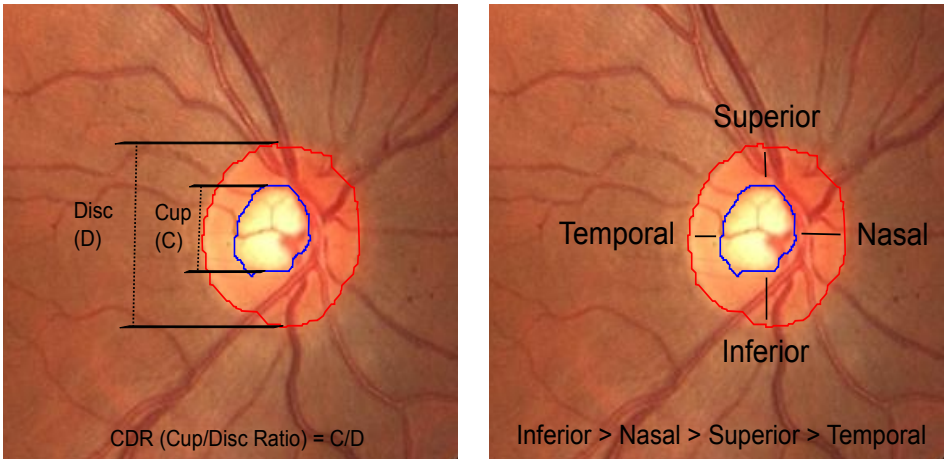
As it is mentioned in the survey made by (Mary, Rajsingh, and Naik 2016), the current works in the literature can be divided into two main subgroups: optic disc localisation and optic disc segmentation. Among all the optic disc localisation methods, the most significant works are based on the highest image

variation (Li and Chutatape 2001), optimum threshold (Lalonde, Beaulieu, and Gagnon 2001), mask generation (Gagnon et al. 2001), principal component analysis (PCA) (Morales, Naranjo, Angulo, et al. 2013; Li and Chutatape 2004), Hough transform (Chrástek et al. 2005) and line operator (Lu and Lim 2010; Lu 2011).

Regarding the works focused on optic disc segmentation, the more known are based on machine learning and traditional classifiers (Parfitt, Mikelberg, and Swindale 1995; Park and Yae 2002; Chan et al. 2002), geometric based active contours models (Lowell et al. 2004; Li and Chutatape 2004; Joshi, Sivaswamy, and Krishnadas 2011), level set methods (Wong et al. 2008), watershed transformation (Morales, Naranjo, Angulo, et al. 2013) and deep learning (Sevastopolsky 2017; Maninis et al. 2016).

Once the optic disc is segmented, there are different ways to measure the changes in it. A common way is a ratio that expresses the vertical diameter proportion of the optic disc and the cup (Cup/Disc ratio or CDR) (See Figure 2.7(a)). For normal discs, this proportion falls in the range of 0.3 to 0.5 and is higher than 0.5 for glaucomatous discs (Nath and Dandapat 2012). There exist other characteristics such as the ISNT rule and the area Cup/Disc ratio or ACDR. The ISNT rule in a healthy optic disc has a characteristic configuration of the neuro-retinal rim; which is thickest inferiorly, followed by superiorly, then nasally, then temporally (I >S >N >T) (Cobb 2010) (See Figure 2.7(b)). On the other hand, the ACDR is the ratio between the area occupied by the optic nerve and the area occupied by the cup. Therefore, CDR, ISNT rule and ACDR are quantitative characteristics commonly used for glaucoma assessment and can be gathered manually or by automatic algorithms.

In order to measure CDR, ACDR and ISNT rule, the first step is to segment the optic cup. However, given that the optic cup is actually an excavation, representing it as a 2D object using a colour fundus image makes the automatic cup segmentation process particularly difficult. Different approaches have been developed towards cup segmentation and glaucoma detection using colour fundus images. For instance, Wong et al. and Liu et al. presented a method to calculate the CDR after obtaining the optic cup and optic disc masks using level-set technique (Wong et al. 2008). Wong et al. tested their method on 104 images from the Singapore Malay Eye Study and found that their method produced results with a variation of up to 0.2 CDR units from the ground truth. Another approach for optic disc and optic cup segmentation is presented by Cheng et al. which developed a technique to measure the CDR based on superpixel classification (Cheng et al. 2013). They evaluated their method on 650 images achieving areas under the ROC curve of 0.800



(a) (b)

Figure 2.7: Illustration of how to measure CDR and ISNT rule. (a) CDR of a normal disc and (b) ISNT rule of a normal disc.

and 0.822 in two databases. There is also another method proposed by Joshi et al., which is based on anatomical evidence such as vessel bends at the cup boundary. They localised the optic cup using the vessel geometry and circular Hough transform obtaining a CDR error of 0.12 ± 0.10 (Joshi, Sivaswamy, and Krishnadas 2011). Different techniques have only been focused on optic disc and/or cup segmentation (Morales, Naranjo, Angulo, et al. 2013; Sivaswamy et al. 2014). In (Almazroa et al. 2015), a review of optic disc and cup segmentation methodologies is presented. This review provides a flowchart for each developed technique. It discusses the pros and cons of each optic disc and cup segmentation method.

2.4.2 Glaucoma Classification and Retinal Image Synthesis using Deep Learning

Important limitations of the methods that are based on handcrafted characteristics (CDR, Area Cup/Disc ratio (ACDR), vessel kinks and ISNT rule) is the significant disagreement in estimating them even between expert human graders. For that reason, new algorithms have been focused on automatic feature extraction such as the data-driven methods and convolutional neural networks (CNNs).

For instance, in the paper published by Bock, Meier, Nyúl, et al. 2010, they proposed a data-driven method. This method is not based on accurate measurements of geometric optic nerve head structures such as the CDR. Instead, they used the idea of "Eigenimages" to extract features that are later classified by a Support Vector Machine (SVM). They evaluated their algorithm on 575 images randomly selected from the Erlangen Glaucoma Registry (EGR), obtaining a competitive AUC of 0.88. However, the images used in their work are private and do not allow a direct and reliable comparison with other works.

CNNs emerged as a powerful tool for image classification and semantic segmentation since the ImageNet competition in 2012, in which the main goal was to estimate the content of natural images for the purpose of automatic annotation using a subset of the ImageNet dataset (Russakovsky et al. 2015). Their success came through the use of GPUs, rectifiers such as ReLU, data augmentation techniques and new regularization techniques such as Dropout (Srivastava et al. 2014). The main power of the CNN architectures relies on their ability to automatically extract highly discriminating features at multiple levels of abstraction.

CNNs are designed to process data that come in form of a raw format (e.g. images). However, training a CNN from scratch is not an easy task. They require a huge amount of labelled data -a requirement that is difficult to meet in the glaucoma assessment task- and high computational resources.

However, there are two alternatives to train a CNN from scratch that have been previously applied to several medical image classification tasks. The first alternative consists in fine-tuning a CNN that has been trained using a large labelled dataset from a different application (e.g., ImageNet). An example of this alternative is the work of Carneiro, Nascimento, and Bradley 2015, where they showed that CNN models that were pre-trained on natural images, such as the ImageNet, are useful in medical image applications, despite the significant differences in image appearance. The study made by H. Chen et al. 2015 demonstrated that the use of a fine-tuned pre-trained CNN for localizing standard planes in ultrasound images outperforms the state-of-the-art for the fetal abdominal standard plane (FASP). Another example is the study made by Tajbakhsh et al. 2016, in which they conducted a set of experiments for four medical imaging applications showing the use of pre-trained CNN performed as well as a CNN trained from scratch.

The second alternative consists in using an ImageNet-trained CNN as a feature extractor, where the CNN is applied to an input image and then features are extracted from a certain hidden layer of the network. Then, the extracted

features are used to train a new classifier such as Support Vector Machines (SVM), Decision Trees, K-nearest-neighbor or Naive Bayes classifier. For example, Bar et al. 2015 pre-trained CNNs that were used as a feature extractor for chest pathology identification. Another study made by Razavian et al. 2014 showed that using features extracted from the OverFeat network and feeding an SVM classifier, it is possible to obtain superior results compared to the highly tuned state-of-the-art systems.

For glaucoma assessment, there are also several works in the literature that employ CNNs. For instance, X. Chen et al. 2015 proposed and trained from scratch a CNN architecture that contains six layers: four convolutional layers and two fully-connected layers, to automatically classify glaucomatous fundus images. They performed the experiments on two private databases: ORIGA-(light) which contains 650 images and SCES which contains 1676 images, achieving an AUC of 0.831 and 0.887 respectively. For ORIGA database, they trained their CNN architecture by randomly selecting 99 images, and using the remaining 551 images for test. For SCES database, they used the 650 images from ORIGA database for training, and all the 1676 images of SCES database for test. The main disadvantage is the unbalanced data. The ORIGA database is comprised of 168 glaucomatous and 482 normal fundus images and the SCES database contains 1676 fundus images of which only 46 are glaucomatous. Another limitation of this work is that the obtained results are difficult to reproduce because ORIGA and SCES databases are not publicly available.

A study conducted by Alghamdi et al. 2016 makes use of eight databases (four public and four private databases) to detect optic disc abnormality. They developed a new approach using two CNNs: one CNN was trained to first classify the optic disc region and the other CNN to classify the optic disc region into normal, suspicious and abnormal classes. However, the four public databases (DRIVE, STARE, DIARETDB1 and MESSIDOR) used in the work of Alghamdi et al. cannot be used for glaucoma classification because they were taken for different purposes. This means those images do not have any glaucoma sign or do not have glaucoma annotations. The glaucoma labelled databases they used are private and, for that reason, it is difficult to reproduce the results presented in their work.

In the study made by Abbas 2017, he developed and implemented a system known as Glaucoma-Deep. This system consists of an unsupervised CNN architecture that automatically extracts features from the fundus images. Afterwards, it uses a deep-belief network (DBN) model to select the most discriminative features. In his work, Qaisar Abbas uses four databases to test his

method, three of them are public and one private. Although his work shows good results (Specificity: 0.9801 and Sensitivity: 0.8450), details of the CNN and architecture were not given.

It is worthy to mention the work made by Orlando, Prokofyeva, et al. 2017, in which they showed how two different CNNs, OverFeat and VGG-S, could be used as feature extractors. They also investigated how the performance of these networks behave when Contrast-limited adaptive histogram equalization (CLAHE) and vessels deletion are applied to the fundus images. In their work, they used Drishti-GS1 database to test the performance of the fine-tuned CNNs. They observed that OverFeat CNN performed better than VGG-S, obtaining an AUC of 0.7626 and 0.7180, respectively. The main limitation of this work is the small number of images (101 images) used to test the performance of the CNNs. However, their method achieved a competitive AUC score with respect to other existing strategies.

Although previous works in the literature present a great performance in glaucoma classification task, all of them have a huge problem when trying to generalise. The number of glaucoma-labelled images available in the literature is scarce. For these reasons, image synthesis and semi-supervised learning methods will be discussed below.

Retinal image synthesis has been a focus of the scientific community. For instance, Fiorini et al. 2014 used a system that generated the retinal background and the fovea, and another system to generate the optic disc by using a large dictionary of patches with no vessels that are later registered. After that, the authors developed a complementary work that is mainly focused on vessel generation (Bonaldi et al. 2016). Although their method allows the generation of high-quality and large resolution images, the process of concatenating the generation of the main parts of the images is a considerable computational complex algorithm that relies on how well the images are registered.

Another approach to retinal image synthesis is the one developed by Costa, Galdran, M. I. Meyer, Niemeijer, et al. 2017. In their work, they trained an adversarial method on vessel networks and their corresponding retinal fundus images. In other words, they learned a transformation between the vessel trees and the retinal fundus. The main limitation of their method is the dependency of an independent algorithm to segment the vessels.

In another paper, Costa et al. presented a method which improves their previous work. Instead of learning a transformation between the vessel trees and the corresponding retinal image, the authors used the original vessel trees to

train an autoencoder. Then, the synthetic vessel trees are used as input to the retinal image synthesizer (Costa, Galdran, M. I. Meyer, Abràmoff, et al. 2017).

Although the latter system proposed by Costa et al. is a substantial improvement in their previous work, both methods are dependent on how well the independent method extracts the vessels. The quality of the segmented vessel tree will affect the synthetic vessel trees and then, the final retinal image.

Chapter 3

Materials

In this chapter, a description of private and public databases used to train, validate and test the methods proposed in this thesis are presented in this chapter.

3.1 Private Databases

ACRIMA

ACRIMA database is composed of 705 fundus images (396 glaucomatous and 309 normal images) with resolution 2048×1536 px. They are part of the ACRIMA project and were obtained from glaucomatous and normal patients. All patients were selected by experts based on their criteria and clinical findings during the examination. Most of the fundus images from this database were taken from the left and right eye previously dilated and centred in the optic disc. They were captured using the Topcon TRC retinal camera and IMAGEnet® capture System. Images were taken with a field of view of 35º.

All images from ACRIMA database were annotated by two glaucoma experts at the Fundación Oftalmológica del Mediterráneo (FOM) with eight years of experience. No other clinical information was taken into account while providing labels for the images. This first version of ACRIMA database can only be used for classification tasks because optic disc and optic cup annotations are not provided.

Update: ACRIMA database is now publicly available with the publication of this article: **Diaz-Pinto A**, Morales S, Naranjo V, Köhler T, Mossi M J, Navea A. CNNs for Automatic Glaucoma Assessment using Fundus Images: An Extensive Validation. March 2019. BioMedical Engineering OnLine journal. doi: https://doi.org/10.1186/s12938-019-0649-y.

Autogla

Autogla is a private database provided by the FISABIO oftalmología médica (FOM). It is composed of 83 images with resolution 3216×2136 px, containing 50 glaucomatous and 33 normal images.

12Octubre

12Octubre is a private database provided by the 12 de Octubre Hospital in Madrid. It is composed of 53 images with resolution 768×576 px (Román Morán et al. 2014). Additionally to the original images, this database contains annotations for the optic disc and optic cup.

3.2 Public Databases

CHASEDB1

The CHASEDB1 database is a retinal vessel reference database acquired from multiethnic school children (Owen et al. 2011). This database is a part of the Child Heart and Health Study in England (CHASE), a cardiovascular health survey in 200 primary schools in London, Birmingham, and Leicester. The images were captured at 30° FOV with a resolution of 1280×960 px. Images of this database are characterized by having nonuniform background illumination and poor contrast of blood vessels as compared with the background. The 28 images that compose CHASEDB1 database are divided such that 20 images are included in the test set and 8 images comprise the training set (Ng et al. 2014). Database available at https://blogs.kingston.ac.uk/retinal/chasedb1/

DRIONS-DB

DRIONS stands for Digital Retinal Images for Optic Nerve Segmentation (Carmona et al. 2008). It is a public database that consists of 110 colour digital retinal images (600 x 400 px) whose optic disc is manually segmented by two specialists. The mean age of the patients was 53.0 years (S.D. 13.05), with 46.2% male and 53.8% female and all of them were Caucasian ethnicity. 23.1% patients had chronic simple glaucoma and 76.9% eye hypertension. Database available at http://www.ia.uned.es/~ejcarmona/DRIONS-DB.html

Drishti-GS1

Drishti-GS1 dataset consists of 101 retinal images (Sivaswamy et al. 2014). It is divided into 50 training and 51 testing images and all the images have been labelled by 4 clinical experts. The images were collected at Aravind eye hospital which correspond to patients between 40-80 years of age with a roughly equal number of males and females.

All images were taken centred on OD with a Field-Of-View (FOV) of 30º and of dimensions 2896×1944 px. Clinical experts with experience of 3,5,9 and 20 years created the ground truth for the 50 training images by using a dedicated marking tool that allows precise boundary marking. Additionally to the original images, this database contains the ground truth represented by a segmentation Soft Map, average OD and Cup boundaries, CDR Values, image-level decisions which is a binary image-level decision on whether each image is normal or glaucomatous and notching which is the decision on the occurrence of notching in the superior, inferior, nasal, and temporal sectors, assessed by a single expert. Database available at https://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php

DRIVE

DRIVE stands for Digital Retinal Images for Vessel Extraction (Staal et al. 2004). It was obtained from a diabetic retinopathy screening program in The Netherlands from subjects between 25-90 years of age. DRIVE databased is composed of 40 images in which 33 do not show any sign of diabetic retinopathy and 7 show signs of mild early diabetic retinopathy.

The images were acquired using a Canon CR5 non-mydriatic 3CCD camera with a 45-degree field of view (FOV). Each image was captured using 8 bits per colour plane at 768×584 px. The FOV of each image is circular with a

diameter of approximately 540 pixels. For this database, a mask image that delineates the FOV is provided.

DRIVE database was divided into a training and a test set, both containing 20 images. A single manual segmentation of the vasculature is available for the training set and two manual segmentations for the test set. Database available at https://www.isi.uu.nl/Research/Databases/DRIVE/

DR KAGGLE

The DR Kaggle database was presented for the Diabetic Retinopathy (DR) Detection in Kaggle (*Kaggle* 2015). It is a large set of high-resolution fundus images taken under a variety of imaging conditions and was rated by a clinician into 5 classes:

- No DR
- Mild
- Moderate
- Severe
- Proliferative DR

As in most real-world datasets, it is possible to encounter noise in both the images and labels. Artefacts, unfocused, underexposed, or overexposed images are founded in DR Kaggle database. It is composed of 88702 images divided into two subsets: 35.126 images in the training set and 53.576 test images. DR Kaggle database is available at https://www.kaggle.com/c/diabetic-retinopathy-detection/

e-ophtha

e-ophtha is a database of colour fundus images specially designed for scientific research in Diabetic Retinopathy (DR) containing 463 images (Decencière, Cazuguel, et al. 2013). This database is made of two sub databases named e-ophtha-MA (MicroAneurysms) with 148 images with microaneurysms or small haemorrhages and 233 images with no lesion, and e-ophtha-EX (EXudates) with 47 images with exudates and 35 images with no lesion. It was generated from the OPHDIAT® Tele-medical network for DR screening, in the framework of the ANR-TECSAN-TELEOPHTA project funded by the French

Research Agency (ANR). Database available at http://www.adcis.net/en/Download-Third-Party/E-Ophtha.html

HRF

HRF stands for High-Resolution Fundus (Köhler et al. 2013). This database contains 45 images which 15 images of healthy patients, 15 images of patients with diabetic retinopathy and 15 images of glaucomatous patients. They were captured by a Canon CR-1 fundus camera with a field of view of 45° with a resolution of 3504×2336 px. Binary gold standard vessel segmentation images generated by clinicians is available for each image. Database available at https://www5.cs.fau.de/research/data/fundus-images/

MESSIDOR

MESSIDOR stands for Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology (Decencière, X. Zhang, et al. 2014). It is a database composed of 1200 eye fundus images acquired by 3 ophthalmologic departments using a colour video 3CCD camera on a Topcon TRC NW6 non-mydriatic retinograph with a 45-degree field of view. The images were captured using 8 bits per colour plane at 1440×960, 2240×1488 or 2304×1536 px. From the 1200 images, 800 were acquired with pupil dilation and 400 without dilation. The Retinopathy grade and the risk of macular oedema were provided by the medical experts for each image. Database available at http://www.adcis.net/en/Download-Third-Party/Messidor.html

ONHSD

ONHSD stands for Optic Nerve Head Segmentation Dataset (Lowell et al. 2004). It is composed of 99 fundus images taken from 50 patients. Those patients are from various ethnic backgrounds (Asian 20%, Afro-Caribbean 16%, Caucasian 50%, Unknown 14%). The images of this database were acquired using a Canon CR6 45MNf fundus camera, with a field angle lens of 45° and resolution 640×480 px. In this database, annotations of the optic discs, made by clinical experts, are also available. Database can be downloaded from http://www.aldiri.info/Image%20Datasets/ONHSD.aspx

ORIGA-light

ORIGA-light contains 650 retinal images annotated by trained professionals from Singapore Eye Research Institute (Z. Zhang et al. 2010). It is composed of 168 glaucomatous images and 482 normal images with a resolution of 3072×2048 px. This database is widely used as a benchmark for automatic glaucoma classification methods. Database available at http://imed.nimte.ac.cn/en-imed-origa-650.html (Accessed: October 2017)

REFUGE

The REFUGE database is composed of 1200 retinal fundus images divided into three subsets: training (400 images), validation (400 images) and test (400 images) (REFUGE 2018). They were acquired with two different fundus cameras: Zeiss Visucam 500 (2124×2056 pixels) for the training set and Canon CR-2 (1634×1634 pixels) for the validation and test sets. All images are centred at the posterior pole with both macula and optic disc. Only the training set has glaucoma labels (40 glaucomatous images and 360 normal images).

Labels in the training set were obtained from the health records. This means, labels were not obtained only based on fundus image, but also the OCT and visual field. Regarding the optic disc and optic cup annotations, they were obtained by seven independent clinical experts from Zhongshan Ophthalmic Center, Sun Yat-sen University in China. The final reference standard was made by merging all the seven annotations into one. REFUGE database was made publicly available for the OMIA workshop at MICCAI2018. Database available at https://refuge.grand-challenge.org/Home/

RIM-ONE

RIM-ONE is a set of retinal image databases for optic nerve evaluation and glaucoma detection (Medina-Mesa et al. 2015). There are several versions of this database: RIM-ONE v1 and RIM-ONE v3 that are composed of 455 and 159 images, respectively. These databases were created by ophthalmologists from the Department of Ophthalmology at the Universitario de Canarias Hospital in Spain. The RIM-ONE v1 dataset was labelled according to a binary classification (healthy vs. glaucomatous), and RIM-ONE v3 was labelled as a 3-class classification problem (healthy, suspicious and glaucoma). Additionally to the images, this database contains annotations for the optic disc and optic cup. Database available at http://medimrg.webs.ull.es/research/retinal-imaging/rim-one/

SASTRA

SASTRA database is composed of 40 fundus images. Eighteen images have a resolution of 768×576 px and the other 22 have a resolution of 1504×1000 px. SASTRA database was used to validate a semi-automated method for glaucoma detection using the CDR and the ISNT rule (Narasimhan et al. 2012). Database obtained by emailing corresponding author knr@ece.sastra.edu.

sjchoi86-HRF

sjchoi86-HRF database (Abbas 2017; sjchoi86 2017) is composed of 601 fundus images divided into 4 subsets: normal (300 images), glaucoma (101 images), cataract (100 images) and retina disease (100 images). The images were captured using 8 bits per colour plane at 2592×1728, 2464×1632 or 1848×1224 px. Database available at `https://github.com/sjchoi86/retina_dataset/tree/master/dataset` (Accessed: July 2017)

STARE

STARE stands for STructured Analysis of the Retina (Hoover, Kouznetsova, and Goldbaum 2000). It is a project conceived and initiated in 1975 by Michael Goldbaum, M.D., at the University of California, San Diego. From this project, they created the STARE database composed of 400 images of 700×605 px designed for blood vessel and OD segmentation. Database available at `http://cecas.clemson.edu/~ahoover/stare/`

All the previously mentioned databases are shown alphabetically ordered in Table 3.1. For each database, the number of normal and glaucomatous images and what databases contain optic disc and optic cup annotations are presented.

Experts of the FISABIO Oftalmología Médica (FOM) analysed the images of the DRIVE database for glaucoma through visual inspection, and based on their analysis they were labelled as Glaucoma or Normal. They also segmented the optic disc and optic cup manually for the images in the HRF, RIM-ONE v1 and Autogla databases.

Table 3.1: Databases used to train, validate and test the proposed methods. GT stands for Ground Truth

Database	Glaucoma	Normal	OD/Cup GT	Total
ACRIMA	396	309	No	705
Autogla	50	33	Both	83
CHASEDB1 (Owen et al. 2011)	-	-	No	28
DRIONS-DB (Carmona et al. 2008)	-	-	No	110
Drishti-GS1 (Sivaswamy et al. 2014)	70	31	Both (train set)	101
DRIVE (Staal et al. 2004)	20	20	Both	40
DR KAGGLE (*Kaggle* 2015)	-	-	No	88702
e-ophtha (Decencière, Cazuguel, et al. 2013)	-	-	No	463
HRF (Köhler et al. 2013)	27	18	Both	45
MESSIDOR (Decencière, X. Zhang, et al. 2014)	-	-	No	1200
ONHSD (Lowell et al. 2004)	-	-	Only OD	99
ORIGA-light (Z. Zhang et al. 2010)	168	482	No	650
REFUGE	40	360	Both (train set)	1200
RIM-ONE v1 (Medina-Mesa et al. 2015)	194	261	Both	455
RIM-ONE v3 (Medina-Mesa et al. 2015)	74	85	Both	159
SASTRA (Narasimhan et al. 2012)	-	-	No	40
sjchoi86-HRF (Abbas 2017)	101	300	No	601
STARE (Hoover, Kouznetsova, and Goldbaum 2000)	-	-	Only OD	400
12Octubre (Román Morán et al. 2014)	29	24	Both	53
	1169	1923	-	**95134**

Chapter 4

Optic Disc and Optic Cup Segmentation

The first step in most of the classical approaches for glaucoma assessment using retinal fundus images is the segmentation of the optic disc and optic cup. The goal of this step is to determine, as accurate as possible, the masks for these regions of interest. These masks are used to measure clinical features such as the Cup/Disc ratio and the ISNT rule. This chapter presents two approaches for addressing this step, one based on stochastic watershed transformation for optic cup segmentation and further glaucoma diagnosis and another that makes use of a convolutional neural network for segmenting the optic disc and optic cup.

4.1 Stochastic-watershed-based approach

This section is mainly focused on using the Stochastic Watershed transformation for the cup segmentation and further glaucoma diagnosis by measuring the CDR, ACDR and the ISNT rule. First, the optic disc mask for each image was obtained automatically by using the method proposed in (Morales, Naranjo, Angulo, et al. 2013). Secondly, images are resized and cropped around the optic disc. Thirdly, grey-scale transformation and vessel removal are necessary steps previous to finally apply the Stochastic Watershed transformation.

The flowchart of the proposed method is presented in Figure 4.1 and described below.

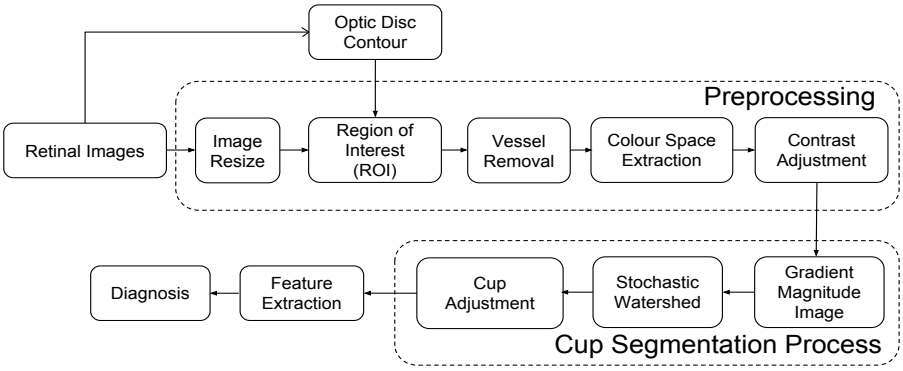


Figure 4.1: Complete block diagram of the proposed algorithm

Preprocessing

Image resize

Image resize is the first part of the preprocessing block. Due to the fact that the images under study belong to different databases, the image size varies. This block resizes the images to a standardized size in order to obtain comparable results between the databases. In this work, we used the image size 768x576 px because is the lowest resolution of the images available for this method.

To resize the images, the length of the horizontal diameter of the fundus was used as reference (X. Zhang et al. 2012). Bicubic interpolation was used for resizing, in which the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighbourhood.

ROI localization

After image resizing, the next step is to crop the images strategically. The proposed method crops the original image using a bounding box of 1.2 times the optic disc radius. An example of a cropped image is shown in Figure 4.2.

Figure 4.2: Cropped image using as reference the optic disc mask.

Vessel removal

The glaucoma disease is mainly characterized by the loss of the optic fibre nerves and astrocytes. However, to the best of our knowledge, the vessels are not significantly affected by this disease as it is the optic nerve. For that reason, an important preprocessing step is to remove the vessels which do not provide extra information for glaucoma diagnosis and can make the cup segmentation more difficult.

A method based on mathematical morphology, curvature evaluation and the k-means clustering algorithm was used to remove the vessels (Morales, Naranjo, Navea, et al. 2014). Afterwards, an iterative inpainting technique (Criminisi, Pérez, and Toyama 2004), used in photo restoration and video processing, replaces the pixels of the vessel mask using exemplar-based synthesis. An example of the vessel removal result is shown in Figure 4.3.

Figure 4.3: Example of the vessel removal result. (a) Cropped and resized image, (b) Vessel mask and (c) Inpainted image.

Component selection

After images are resized, cropped and without vessels, the components of the colour spaces CMYK (Cyan, Magenta, Yellow, and Key (black) components), Principal Component Analysis (PCA) (Pearson 1901), YIQ, Lab, Luv (Mustroph 2014) and RGB were analysed. The reason for this analysis is to check the performance of each component in cup segmentation and choose the one that provides optimal results.

From the analysis made by (Diaz-Pinto et al. 2016), it was found that the components with better performance for cup segmentation were Cyan plus Key component in CMYK, "Q" in colour space YIQ, "u" component in Luv colour space, first component in PCA and "a" component in Lab colour space. Additionally to their study, in this section the RGB colour space was also analysed. Performance results of each of these components are presented in Section 4.3.

Examples of these components are shown in Figure 4.4. In this figure, it is also possible to see that the cup is usually darker or brighter than the other parts of the image, which is essential for the proposed algorithm.

Contrast adjustment

It was observed that the nonuniform-contrast image through the available databases affects the classification performance. For this reason, a contrast adjustment is applied to the grey-scale image obtained after component selection. After obtained the grey-scale image using the colour space model, the intensity values in that image are linearly mapped into another grey-scale im-

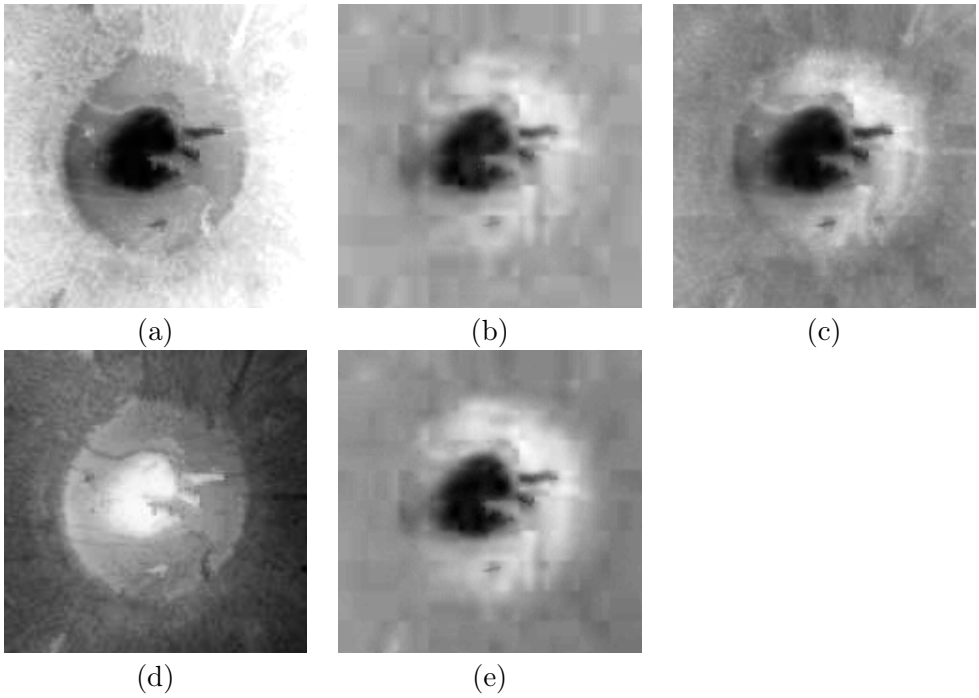


Figure 4.4: Colour space components used in the proposed method. (a) **C**+**K** components in CMYK space, (b) **Q** in YIQ space, (c) **u** in Luv space,(d) First PCA component and (e) **a** component in Lab space.

age such that 1% of pixels is saturated at low and high intensities, with the aim of obtaining a higher difference between the intensity values in the cup and the rest of the image. An example of this adjustment can be seen in Figure 4.5(d)

Cup segmentation

The cup segmentation block is mainly divided into 2 parts: the Stochastic Watershed transformation and the Cup adjustment block.

Stochastic Watershed

Fundamentally, the stochastic watershed transformation is the neuralgic step in this method. It is an improved variant of the watershed transformation. Based on the gradient image, the watershed transformation is a segmentation technique in which the minimum pixel values of the image represent the object of interest and the maximum pixel values represent the separation boundaries between objects (Beucher and F. Meyer 1992).

One problem with the watershed technique is the over-segmentation, which is caused by the existence of numerous local minima in the image due to the presence of noise. This problem is solved with the marker-controlled watershed, establishing the image minima artificially and defining a marker per minimum. Internal and external markers are needed. The internal markers determine the object of interest (optic cup) and the external markers are used to limit the segmented area (optic disc). Although in the marker-controlled watershed there exist several strategies to choose the most appropriate internal markers for each case, the following strategies are the most used: internal markers are selected based on a manual or automatic seed, uniformly selected in the image (stratified markers) or following a non-uniform distribution that is restricted to areas that accomplished a specific condition (regionalized random markers). Independently on the chosen strategy, a problem of the marker-controlled watershed is the under-segmentation, which means part of the area of interest is not detected because the existence of not enough markers representing the entire area.

Both problems, over- and under-segmentation, are solved with the Stochastic Watershed transformation. In this transformation, a given number M of marker-controlled watershed realizations are performed selecting N regionalized random markers (or pseudo-random markers) in each realization. The idea is to estimate a probability density function (*pdf*) for the contours of the

image, which filter out non-significant border fluctuations. The probability density function is computed by Parzen window method (Duda, Hart, and Stork 1995) as follows:

$$pdf(\mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M}(WS_i(\mathbf{x}) * G(\mathbf{x}; s)) \tag{4.1}$$

where $G(\mathbf{x}; s)$ represents a Gaussian function of variance s^2 and mean $\mu(\mu = 0)$, M represents the number of marker-controlled watershed realizations with N regionalized random markers and $WS_i = WS(\varrho)_{f_{mrk_i}}$ the ith output watershed image, being ϱ the gradient image. Afterwards, it is necessary to perform a last marker-controlled watershed transformation on the *pdf*, which defines the resulting mask by joining all the watershed regions.

In this work, the gradient contour of the optic disc (See Figure 4.5(e)) was used as the external marker. It is worthy to mention that the optic disc masks used for this process were obtained from the algorithm proposed in (Morales, Naranjo, Angulo, et al. 2013).

The used regionalized random markers are random markers whose distribution is restricted to low-intensity or high-intensity areas depending on the colour space used (See Figure 4.5(f)). Starting with an initial marker or seed which is defined by the darkest or the brightest pixel that will belong to the optic cup, the regionalized markers or internal markers are chosen by following a Poisson distribution with variance σ^2 (Angulo and Jeulin 2007). The stochastic watershed transformation makes M marker-controlled watershed realizations with N pseudorandom markers to finally obtain a *pdf* (See Figure 4.5(g)). We say pseudorandom, instead of random markers, because they are limited by the region of interest.

In this part, we carefully set the variance of the Poisson function that generates the pseudorandom markers, the number of pseudorandom markers and the number of realizations. Given the small intensity variance between the pixels in the cup, the variance σ^2 was set experimentally in a small value ($\sigma^2 = 0.0003$). The algorithm was programmed to generate a minimum number of pseudorandom markers (N) of 100. This means if the initial variance is not enough to generate 100 pseudorandom markers, the variance values increases until there are at least the minimum number of markers. Another important factor is the number of realizations (M). It was observed that 5 realizations are a good trade-off between a suitable *pdf* for segmentation and a relatively low computational cost. For that reason, this factor was set at 5.

Afterwards, an additional last marker-controlled watershed transformation on the *pdf*, combined with the image gradient, was performed, which defined the resulting mask by joining all the watershed regions (Angulo and Jeulin 2007).

$$\varrho = (1 - \lambda)\varrho + \lambda pdf; \text{ where } \lambda = 0.5. \tag{4.2}$$

In Figure 4.5, it is possible to see the optic cup segmentation process using the stochastic watershed transformation.

Cup adjustment

The cup adjustment block improves the measure of the Cup/Disc ratio and ISNT rule. In this block, the vertical radius of the mask obtained by the watershed transformation is calculated, then a circle is generated with the same vertical radius of the watershed mask and placed in the centre of the optic disc. The main motivation for this adjustment came from the fact that the temporal part of the optic cup is usually hidden by the vessels (Nath and Dandapat 2012). The reason for placing this circle in the centre of the optic disc is because the optic cup is generally located in the centre of the optic disc.

In Figure 4.5(j) an example of the cup adjustment result is shown. The black line represents the segmentation using the proposed method based on Stochastic Watershed, the white line represents the adjustment made to the segmented cup and the yellow and blue lines identify the ground truth of the optic disc and cup, respectively.

Glaucoma diagnosis

At this point, the cup segmentation process using Stochastic Watershed has been shown. Next step is to use the obtained optic cup mask to measure the CDR, ACDR and the ISNT rule and make a glaucoma assessment.

ACDR is defined as the ratio between the area occupied by the cup and the area occupied by the optic disc. CDR is the ratio of the vertical diameter of the cup and the vertical diameter of the optic disc (Cobb 2010). The proposed algorithm measures the CDR by computing the diameter of the optic disc and the cup, as the mean of the two highest vertical diameters because of the irregular shape of the optic disc and the cup.

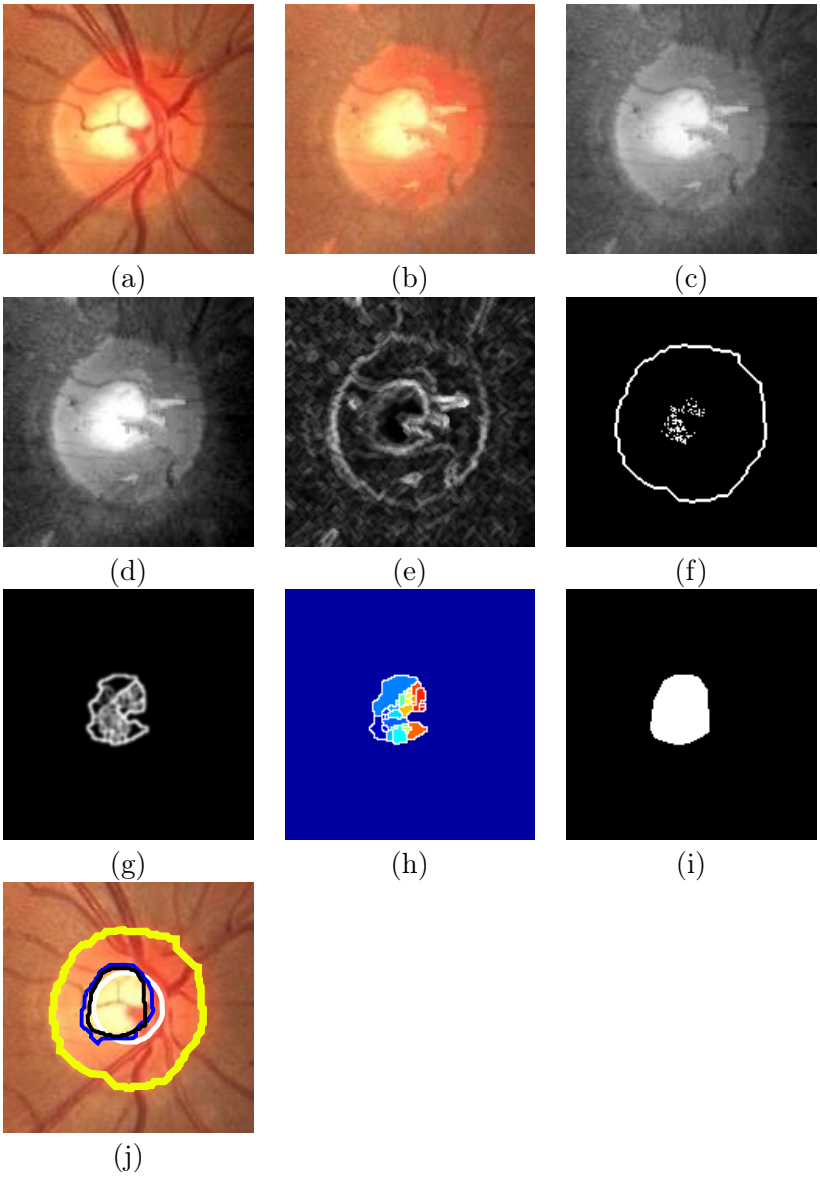


Figure 4.5: Process of the Stochastic Watershed transformation: (a) Resized and cropped image (ROI), (b) Image inpainting, (c) Grey-scale image (first PCA component), (d) Contrast adjust, (e) Gradient image, (f) Pseudo-random markers, (g) *pdf* of image contours, (h) Watershed regions, (i) Final segmentation and (j) Cup adjustment (in white).

Optimum thresholds for glaucoma diagnosis were stablished to **0.50** for the CDR and **0.30** for ACDR (Nath and Dandapat 2012).

Regarding the ISNT rule, the horizontal and vertical thickness of the neuro-retinal rim were measured as it was done with the CDR (by mean of the two highest values). The horizontal thickness represents the temporal part (T) plus the nasal part (N) of the neuro-retinal rim: HorDiam $= T + N$, and the vertical thickness represents the inferior part (I) plus the superior part (S) of the neuro-retinal rim: VerDiam $= I + S$. If the horizontal thickness is smaller than the vertical thickness, the optic disc in the image follows the ISNT rule, and then, classified as "Normal", otherwise it is classified as "Glaucomatous".

4.2 U-Net-based approach

4.2.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) were first introduced by (LeCun, Y 1989) and are biological-inspired (Hubel and Wiesel 1968) variants of Multilayer Perceptrons. Since then, they have been used in computer vision and artificial intelligence. However, their relevance was not discovered until the ImageNet competition in 2012 (http://image-net.org/challenges/LSVRC/).

CNNs are a specialized type of neural network to process data in a raw format or grid-like topology (colour image, 2D audio spectrograms, etc). They extract a set of highly discriminative features at multiple levels of abstraction. The first layers of the architecture extract edges at particular orientations and locations in the image. The middle layers detect structures composed by particular arrangements of edges and the last layers detect more complex structures that correspond to parts of familiar objects, or objects that are combinations of these parts (LeCun, Bengio, and G. Hinton 2015).

A typical CNN architecture consists of a sequence of layers, and every layer transforms its corresponding input into an output through a differentiable function. In this way, the CNN transforms the original image layer by layer from the original pixel values to the final class scores.

Next, a brief description of the main types of layers to build CNN architectures will be presented.

Convolutional layer

The convolutional layer is the main layer of a CNN. It does most of the feature extraction in a CNN by using a set of learnable filters. These filters, composed by a set of values, also called weights, are used to convolve across the width and height of the input volume and compute dot products between the weights and the input at any position (Goodfellow, Bengio, and Courville 2016). The result of this operation is a 2-dimensional feature map for each filter used (See Figure 4.6).

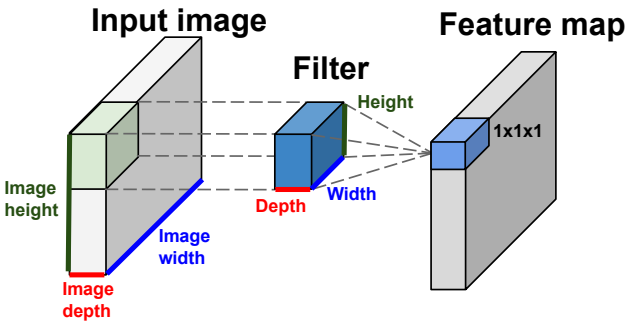


Figure 4.6: Feature map generated by one filter when applied to an input image.

Activation layer

After the feature map computation carried out by the convolutional layer, an activation function is applied. The main objective of using the activation layer is to introduce non-linearities, which allows representing almost any arbitrary complex function that maps inputs to outputs in a CNN. Although there exist several activation functions, the most commonly used are the Sigmoid or Logistic, Tanh and Rectified Linear Unit (ReLU).

The Sigmoid function is a curve of form $\sigma(x) = \frac{1}{1+e^{-x}}$. Its range is between 0 and 1 as it is shown in Figure 4.7

The Tanh function is the mathematical formula: $\sigma(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$. Its output is zero-centred and its range is in between -1 to 1, as it is shown in Figure 4.8

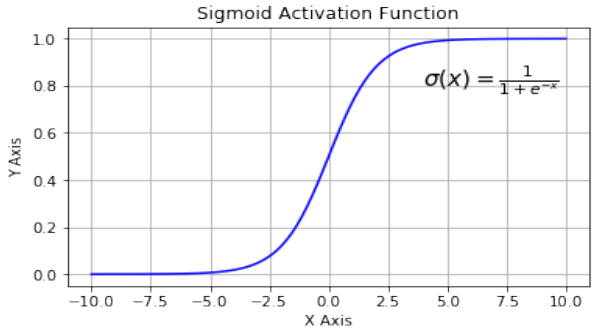


Figure 4.7: Sigmoid or logistic activation function (Sharma 2017)

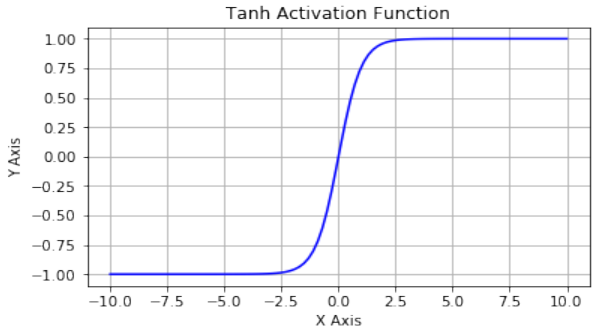


Figure 4.8: Tanh activation function (Sharma 2017).

The ReLU function is the most popular of all these functions because of its simplicity and improvement in convergence. It is defined as $\sigma(x) = max(0, x)$ and its curve can be seen in Figure 4.9

Pooling layer

The aim of using pooling layers in a CNN is to reduce the spatial size of the feature maps. This layer reduces the number of parameters and the computation in the network, and hence, controls overfitting. The most common form is the MAX pooling with filters of size 2×2 applied with a stride of 2. In Figure 4.10 a representation of the MAX pooling form is shown.

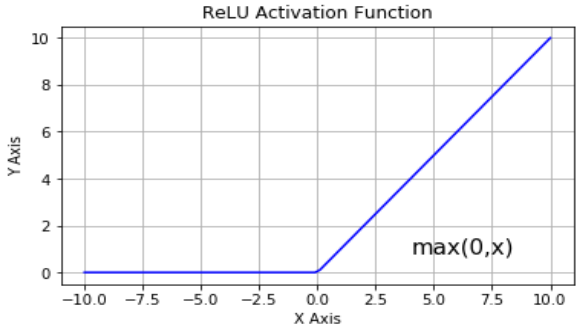


Figure 4.9: ReLU activation function (Sharma 2017).

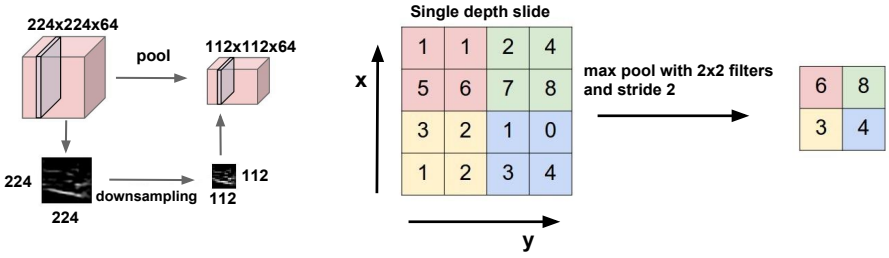


Figure 4.10: Example of the pooling layer (Johnson and Karpathy 2017).

Fully-connected layer

Finally, the fully-connected layer is the one that takes an input volume of activations and outputs an N-dimensional vector, where N is the number of classes. Its outputs can be computed with simple matrix multiplication.

Although there are more types of layers such as normalisation layers, locally-connected layer, etc, the previously described layers are the typical layers that composed a CNN. In Figure 4.11 it is possible to see a typical architecture of a CNN using the convolutional, activation, pooling and fully-connected layers.

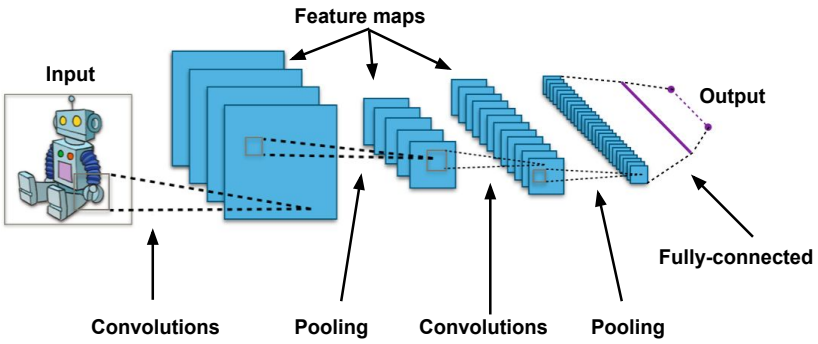


Figure 4.11: Example of a CNN using the main layers above described (convolutional, activation, pooling and fully-connected layer) (Aphex34 2015).

4.2.2 U-Net architecture

The U-Net is a convolutional network architecture for fast and precise segmentation of images. It was introduced by (Ronneberger, Fischer, and Brox 2015). The network is based on the fully convolutional network and its architecture was modified and extended to work with fewer training images and to yield more precise segmentation.

The U-Net network consists of a contracting path and an expansive path, which gives it the u-shaped architecture. The contracting path is a typical convolutional network that consists of repeated application of convolutions, each followed by a rectified linear unit (ReLU) and a max pooling operation. During the contraction, the spatial information is reduced while feature information is increased. The expansive pathway combines the feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the contracting path.

The U-Net architecture (Ronneberger, Fischer, and Brox 2015; Sevastopolsky 2017) used in this section can be seen in Figure 4.12.

This network was originally introduced as a fully-convolutional neural network in which feature maps are depth-concatenated to layers that are upsampled from the bottleneck layer. In simple terms, this network receives as input an RGB image and outputs a probability map.

As a pre-processing technique, we used Contrast Limited Adaptive Histogram Equalization (CLAHE) (Pizer, Johnston, et al. 1990; Pizer, Amburn, et al.

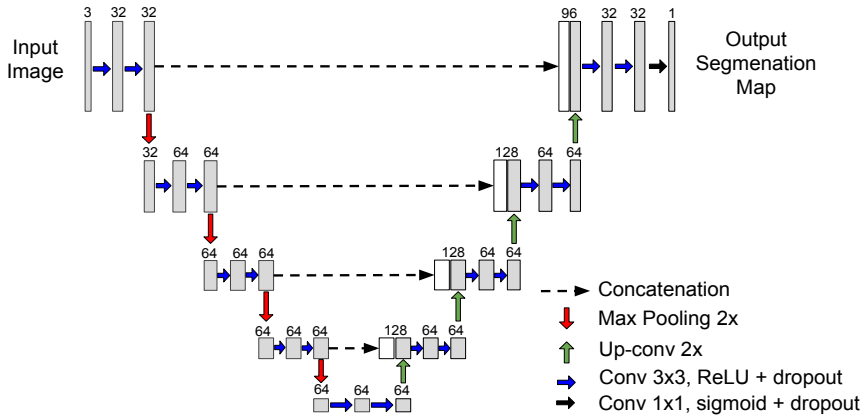


Figure 4.12: U-Net architecture used for Optic disc and Optic Cup segmentation. The number on top of the box represents the number of channels.

1987) for both optic disc and optic cup segmentation. This technique improves the local contrast and enhances the definitions of edges in each region of an image. After that, images were resized to 256×256 px.

A pipeline of our approach for optic disc and optic cup segmentation is shown in Figure 4.13.

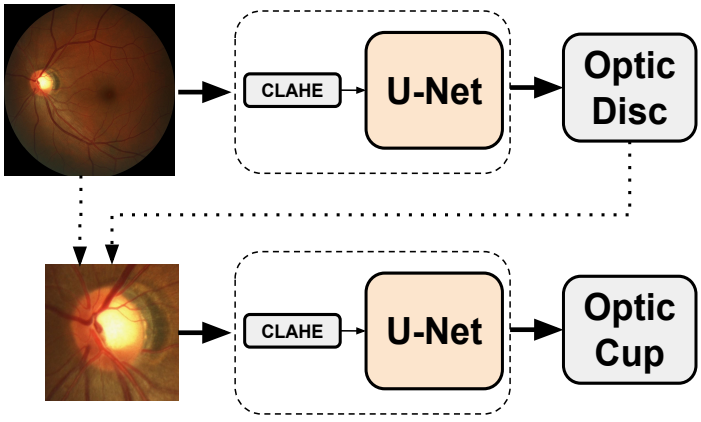


Figure 4.13: Schema used for Optic disc and Optic Cup segmentation

As it is possible to see in Figure 4.13, we first segment the optic disc to obtain the optic cup mask. This means a two-stage process: we first segment the optic disc and used the mask to crop the image and segment the optic cup.

We also used data augmentation for both optic disc and optic cup segmentation. To do that, we employed the Keras library (Chollet et al. 2015) to transform images and masks simultaneously and in the same fashion. We applied vertical and horizontal flipping, a rotation range of 50 degrees, applied both width and height shift of 0.15 and finally randomly zoomed the images by a range between 0.7 and 1.3. All these transformations were applied for each batch during the training stage.

This method was used for the REFUGE challenge, hosted at the MICCAI 2018 conference in conjunction with OMIA workshop[1]. The goals of this challenge were to evaluate and compare automated algorithms for glaucoma detection and obtain, for the optic disc and optic cup segmentation task, the highest value of Dice index on the validation and test set of the REFUGE database. The dataset is composed of 1200 retinal fundus images: 400 images for each set (training, validation and test). These images were annotated by seven independent glaucoma experts from the Zhongshan Ophthalmic Center, Sun Yat-sen University, China.

4.3 Results

Results for the Stochastic-watershed-based approach

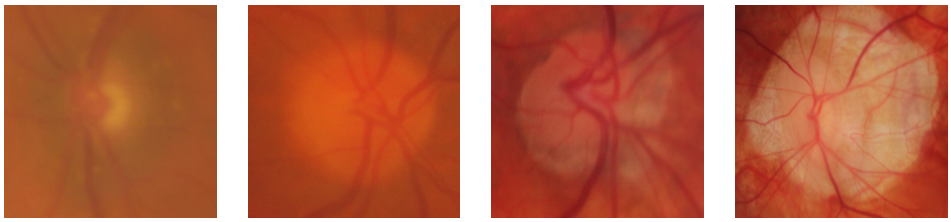
Images belonging to the 12Octubre (Román Morán et al. 2014), DRIVE (Staal et al. 2004), Autogla, HRF (Köhler et al. 2013), RIM-ONE v1 (Medina-Mesa et al. 2015) and the training set of the Drishti-GS1 database were used to validate the stochastic watershed method presented in Section 4.1. They were carefully analysed by a clinical expert and excluded images that had at least one of the following issues:

- Low resolution or low image quality
- Image with doubtful diagnosis

After that, 54 images were excluded from the RIM-ONE database. Examples of the excluded images are shown in Figure 4.14. Therefore, a total of 672

[1] https://refuge.grand-challenge.org/home/ (Group name: Cvblab)

images were used for the experiments. 308 were related to glaucomatous and 364 were related to healthy images.



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

Figure 4.14: Different samples of the excluded images. (a-c) Images with low resolution and (d) Image with doubtful diagnosis.

Cup segmentation

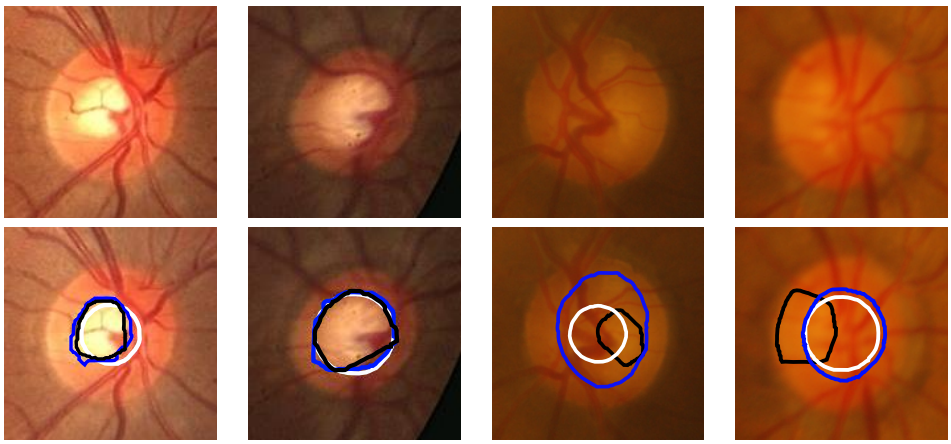
In order to test the performance of the cup segmentation, the Jaccard and Dice indexes were calculated for every image, and the mean and standard deviation were obtained for each database. After that, the weighted average (according to the number of images of each database) and the standard deviation of the Jaccard and Dice indexes were calculated for all databases. This analysis was performed for all the six colour spaces (CMYK, PCA, YIQ, Lab, Luv and RGB).

As it can be seen in Table 4.1, HRF and RIM-ONE databases have better cup segmentation results when YIQ colour space is used. Given that these databases have the majority of the images (446/723 images), the weighted average is higher when YIQ colour space is applied. However, the results obtained when using the CMYK colour space, have also a considerable performance in cup segmentation.

In most cases, using the Stochastic Watershed transformation a proper contour of the optic cup is obtained. Examples of these results can be observed in Figure 4.15(a-b). However, wrong contours are obtained when the optic cup has not its characteristic pallor. For that reason, a weakness of this method lies in the pallor absence of the optic cup. Which makes the optic cup not discernible from other parts of the optic disc. Examples of this problem can be seen in Figure 4.15(c-d).

Table 4.1: Results of the cup segmentation using five colour spaces.

Database	Index	CMYK	YIQ	Luv	Lab	PCA	RGB
12Octubre	Jaccard	$0,553 \pm 0,185$	$0,565 \pm 0,215$	$0,526 \pm 0,260$	$0,551 \pm 0,238$	$0,558 \pm 0,214$	$\mathbf{0,575 \pm 0.192}$
	Dice	$0,692 \pm 0,173$	$0,695 \pm 0,204$	$0,643 \pm 0,281$	$0,674 \pm 0,250$	$0,689 \pm 0,202$	$\mathbf{0,709 \pm 0.178}$
DRIVE	Jaccard	$\mathbf{0,567 \pm 0,198}$	$0,481 \pm 0,227$	$0,415 \pm 0,247$	$0,461 \pm 0,224$	$0,471 \pm 0,211$	$0,495 \pm 0.200$
	Dice	$\mathbf{0,700 \pm 0,191}$	$0,615 \pm 0,233$	$0,541 \pm 0,269$	$0,596 \pm 0,239$	$0,611 \pm 0,209$	$0,638 \pm 0.187$
Drishti-GS1	Jaccard	$0,656 \pm 0.191$	$0,661 \pm 0.215$	$0,347 \pm 0.350$	$0,492 \pm 0.328$	$\mathbf{0,668 \pm 0.200}$	$0,664 \pm 0.185$
	Dice	$0,775 \pm 0.160$	$0,771 \pm 0.190$	$0,416 \pm 0.391$	$0,583 \pm 0.356$	$\mathbf{0,781 \pm 0.175}$	$0,780 \pm 0.161$
Autogla	Jaccard	$0,568 \pm 0,197$	$0,551 \pm 0,210$	$0,512 \pm 0,270$	$0,527 \pm 0,240$	$0,560 \pm 0,209$	$\mathbf{0,572 \pm 0.201}$
	Dice	$0,702 \pm 0,177$	$0,683 \pm 0,207$	$0,626 \pm 0,296$	$0,653 \pm 0,245$	$0,693 \pm 0,194$	$\mathbf{0,706 \pm 0.180}$
HRF	Jaccard	$0,607 \pm 0,185$	$\mathbf{0,627 \pm 0,150}$	$0,545 \pm 0,250$	$0,592 \pm 0,221$	$0,575 \pm 0,212$	$0,614 \pm 0.179$
	Dice	$0,738 \pm 0,161$	$\mathbf{0,759 \pm 0,131}$	$0,663 \pm 0,269$	$0,712 \pm 0,234$	$0,704 \pm 0,199$	$0,744 \pm 0.161$
RIM-ONE	Jaccard	$0,523 \pm 0,190$	$\mathbf{0,540 \pm 0,178}$	$0,329 \pm 0,287$	$0,464 \pm 0,253$	$0,491 \pm 0,209$	$0,515 \pm 0.198$
	Dice	$0,665 \pm 0,175$	$\mathbf{0,683 \pm 0,169}$	$0,423 \pm 0,336$	$0,586 \pm 0,281$	$0,630 \pm 0,202$	$0,656 \pm 0.185$
All databases	Jaccard	$0,559 \pm 0,203$	$\mathbf{0,565 \pm 0,218}$	$0,546 \pm 0,259$	$0,468 \pm 0,233$	$0,546 \pm 0,227$	0.537 ± 0.205
	Dice	$0,696 \pm 0,187$	$\mathbf{0,702 \pm 0,221}$	$0,679 \pm 0,281$	$0,590 \pm 0,226$	$0,682 \pm 0,215$	0.674 ± 0.190



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

Figure 4.15: Results of best and worst cases of the optic cup segmentation (upper images are the original images cropped around the optic disc). The blue line represents the annotation made by the experts, the black line represents the contour obtained from the Stochastic Watershed and the white line represents the cup adjustment.

In order to compare the proposed method with other state-of-the-art methods, the overlapping error E was computed. This evaluation metric is defined in equation 4.3.

$$E = 1 - \frac{Area(S \cap G)}{Area(S \cup G)} \tag{4.3}$$

where S and G denote the segmented and the ground truth cup respectively.

Table 4.2 shows the percentage of images per interval of overlapping error and the average μ_E using the annotated images as ground truth. The results are shown in detail for each database, and for all the images combined when using the YIQ colour space.

Table 4.2: Percentage of images per overlapping error E interval and the average μ_E for cup segmentation purpose.

	# **images**	$E <= 0,1$	$E <= 0,2$	$E <= 0,3$	$E <= 0,4$	$E <= 0,5$	μ_E
12Octubre	53	1,89 %	15,09 %	39,62 %	62,26 %	69,81 %	38,60 %
DRIVE	40	0 %	12,50 %	27,50 %	52,50 %	75 %	41,40 %
Drishti-GS1	101	**6** %	**34** %	**64** %	**68** %	**86** %	**30,85** %
Autogla	83	0 %	9,64 %	37,35 %	56,63 %	71,08 %	40,42 %
HRF	45	0 %	17,78 %	40 %	71,11 %	86,67 %	35,09 %
RIM-ONE	401	0,25 %	5,74 %	22,44 %	42,14 %	61,60 %	44,72 %
All databases	723	**1.11** %	**11.94** %	**32.59** %	**51.27** %	**69.00** %	**41.06** %

It can be seen from the table that the percentage of images for the Drishti-GS1 database has better relative results compare to the other databases. However, the RIM-ONE database has more impact on the final result because of the number of images.

For the comparison, the results presented in (Cheng et al. 2013) were used. Results in that article were obtained from experiments using cup segmentation algorithms such as the threshold method (Joshi, Sivaswamy, and Krishnadas 2011), R-bend (Bock, Meier, Michelson, et al. 2007), ASM method (Yin et al. 2012) and regression method (Y. Xu et al. 2011).

Table 4.3 gives a comparative analysis of the performance of the proposed method against some existing methods of cup segmentation. It is possible to see from the table that the proposed algorithm is among the methods with the best performance. In addition, this method is evaluated with a greater number of images.

Table 4.3: Comparison between existing methods and the proposed method using the overlapping error E in cup segmentation. Other works results shown in this table were taken from (Cheng et al. 2013)

	$E <= 0,1$	$E <= 0,2$	$E <= 0,3$	$E <= 0,4$	$E <= 0,5$	μ_E	**Images used**
Thresholding (Joshi, Sivaswamy, and Krishnadas 2011)	0 %	3 %	15 %	31 %	47 %	53,50 %	138
R-bend (Bock, Meier, Michelson, et al. 2007)	0 %	4 %	28 %	56 %	77 %	39,50 %	200
Proposed method	**1** %	**10** %	**30** %	**50** %	**69** %	**41.06** %	**723**
ASM (Yin et al. 2012)	3 %	25 %	51 %	76 %	88 %	31,30 %	325
Regression (Y. Xu et al. 2011) (Z. Zhang et al. 2010)	6 %	29 %	62 %	81 %	95 %	28,40 %	650
Superpixel (Cheng et al. 2013)	8 %	42 %	75 %	90 %	96 %	24,10 %	650

Unlike existing works, the proposed method uses databases from different hospitals and research centres, which implies different image resolution and image quality, thus it demonstrates the robustness of the method in case of variability conditions. For instance, Regression and Superpixel methods were evaluated using fewer images and from only one private database.

Another comparison with works that only use the Drishti-GS1 database was made. For this comparison the F-score, which is defined as $F = 2P \times R/(P+R)$ was calculated, where P is precision and R is the recall. For instance, Sedai et al. (S Sedai and P Roy and D Mahapatra and R Garnavi 2016) obtained an F-score of 0.86 using only 50 of the 101 images in the Drishti-GS1 database and Chakravarty et al. (Chakravarty and Sivaswamy 2014) obtained an F-score of 0.81 using the whole Drishti-GS1 database. As in the latter, in this work, all the images in the Drishti-GS1 were used; obtaining an F-score of 0.77. It shows that the proposed algorithm has a competitive performance for optic cup segmentation.

Glaucoma diagnosis

After the cup is segmented, the CDR, ACDR and ISNT rule was measured for all the images. Table 4.4 shows the specificity (Sp) and sensitivity (Se) obtained for each colour space that was used in the cup segmentation process.

It can be observed from Table 4.1 that YIQ colour space presents better performance overall the other colour spaces in cup segmentation. It is also possible to see from Table 4.4, that measuring the CDR using the colour space YIQ, the glaucoma diagnosis is more reliable than measuring the ACDR, the ISNT rule or combining all of them. For that reason, the next results were obtained using the YIQ colour space and measuring the CDR.

Table 4.4: Results of the Glaucoma diagnosis using the CDR, ACDR, ISNT and the combination of all of them (Combined). In this table, Sp stands for Specificity and Se stands for Sensitivity

	CMYK		**YIQ**		**Luv**		**Lab**		**PCA**		**RGB**	
	Sp	**Se**	**Sp**	**Se**	**Sp**	**Se**	**Sp**	**Se**	**Sp**	**Se**	**Sp**	**Se**
CDR	0,574	0,697	**0,675**	**0,674**	0,650	0,731	0,832	0,563	0,487	0,716	0.545	0.716
ACDR	0,601	0,633	**0,715**	**0,604**	0,688	0,673	0,849	0,509	0,517	0,663	0.574	0.655
ISNT	0,495	0,570	**0,431**	**0,568**	0,422	0,561	0,337	0,609	0,523	0,544	0.499	0.511
Combined	0,545	0,702	**0,730**	**0,602**	0,685	0,635	0,373	0,760	0,376	0,778	0.513	0.742

In Figure 4.16, the ROC curves for each database can be observed. In these figures, blue lines represent the ROC curves for the ground truth and the red lines represent the obtained results from the proposed method measuring the CDR and using the YIQ colour space. It was observed from the experiments that the ROC plot for the Drishti-GS1 database has the best relative result compared to the other databases.

After the ROC curves for each database were generated, a weighted average ROC curve was computed using the results of all databases. In Figure 4.17(a) it is possible to see a comparison between the ground truth and the results of the proposed method and in Figure 4.17(b), a comparison between the superpixel results, using SiMES database (Cheng et al. 2013), and the results of the proposed method. It is important to remark that the superpixel method and the proposed method were tested with different databases. It can be seen that although the experiments were made with a wide variety of images, the proposed method performs properly and its results could be of interest in a clinical setting.

Computation time

Our whole method was implemented in MatLab on a 3.40 GHz quad-core CPU running Windows 8.1. Once the Stochastic Watershed parameters were tuned, an average of 0.508 s per image was needed to segment, adjust the optic cup contour and measured the CDR, ACDR and ISNT rule. This time clearly highlights a strength of this method, its computational efficiency.

Results for the U-Net-based approach

For the REFUGE challenge, we implemented and trained an U-Net architecture for the optic disc and optic cup segmentation task. This U-Net was trained with the same hyper-parameters used in (Sevastopolsky 2017): Stochastic Gradient Descent (SGD) as optimizer, a batch size of 1, a learning rate of $1e4$ and a momentum of 0.95. As mentioned in Section 4.2.2, the main and only goal for the optic disc and optic cup Segmentation task was to obtain the highest value of Dice index on the validation (400 images) and test set (400 images) of the REFUGE database.

For training the U-Net network, not only the training set (400 images) of the REFUGE database (REFUGE 2018) was used, but also the DRIONS-DB (110 images) (Carmona et al. 2008), Drishti-GS1 (101 images) (Sivaswamy et

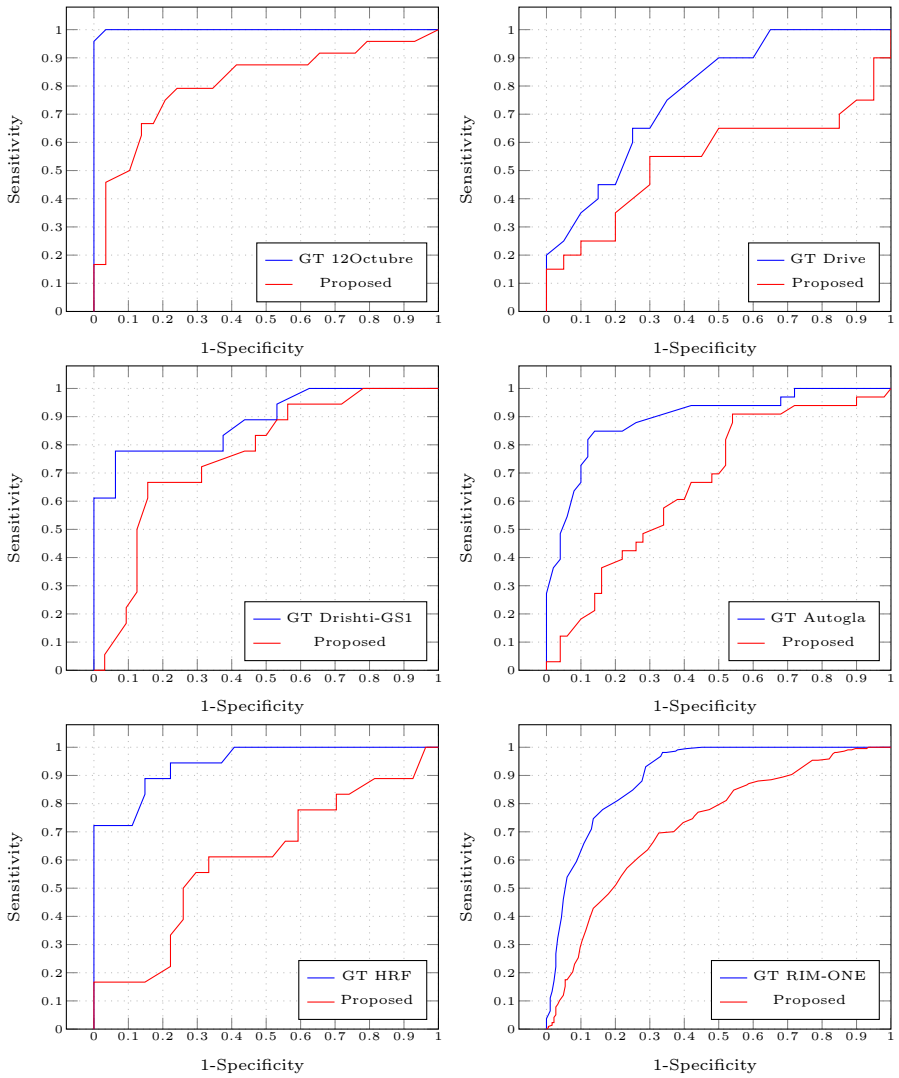


Figure 4.16: Receiver operating characteristic (ROC) curves for the automatic glaucoma diagnosis for each database using the ground truth (GT) images and the results of the proposed method using the YIQ colour space and the CDR.

al. 2014) and RIM-ONE v3 (159 images) (Medina-Mesa et al. 2015), giving place to a total of 770 images.

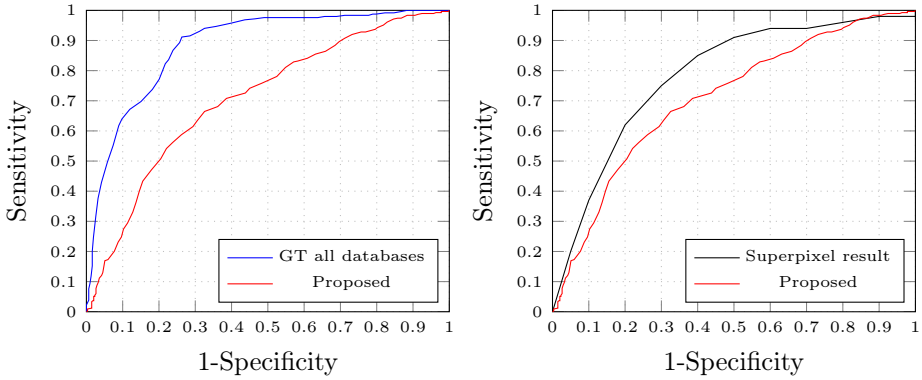


Figure 4.17: A comparison ROC plots between the weighted average ROC curves for all databases using the ground truth (GT) images and the proposed method (left) and the Superpixel method and the proposed method (right).

Using this method, a Dice index of 0.91 and 0.78 were obtained for the optic disc and optic cup segmentation, respectively, in the validation set and a Dice index of 0.91 and 0.77 for the optic disc and optic cup segmentation in the test set. Examples of the optic disc and optic cup segmentation results using the U-Net are shown in Figure 4.18

It is worthy to mention that the open source Deep Learning library Keras (Chollet et al. 2015) and NVIDIA Titan Xp GPU were used to carry out the experiments.

4.4 Discussion

In this chapter, we have analysed the capability of two different approaches for image segmentation. In section 4.1, we have presented an automatic algorithm based on the Stochastic Watershed transformation for optic cup segmentation. Training and testing were performed on 672 fundus images, obtaining a Dice index of $0,70$. After cup segmentation, further feature extraction and glaucoma diagnosis were also performed. This method is a classical approach for classification, in which handcrafted features such as the CDR, ACDR and ISNT rule were used to classify normal and glaucomatous images.

Although a fair comparison with state-of-the-art methods is not always possible as some databases are not public, the obtained results were compared

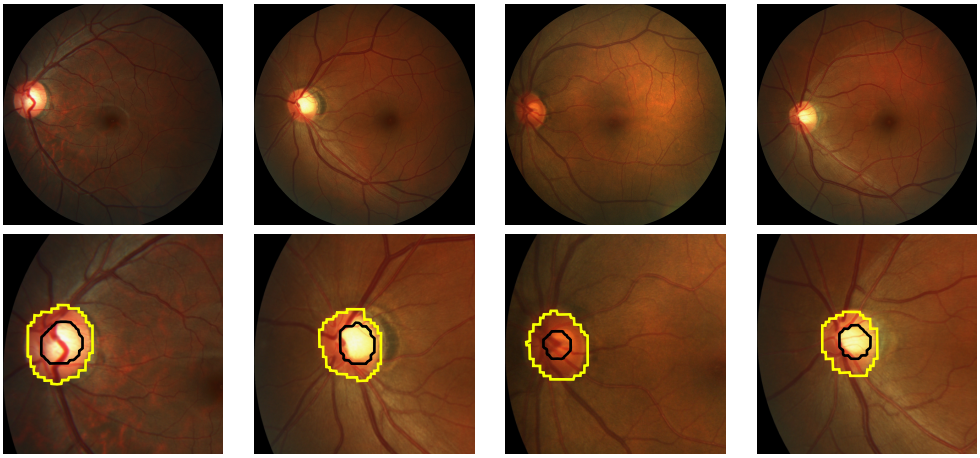


Figure 4.18: Examples of the optic disc and optic cup segmentation obtained from the U-Net. First row are the original images. Optic disc and optic cup are represented by yellow and black lines, respectively. For visibility purposes, images in the second row were intentionally cropped.

against some methods showing competitive sensitivity (0.675) and specificity (0.674) when using the YIQ colour space, the cup adjustment technique and measuring the CDR. However, the main disadvantage of this method is its high dependence on the selected features. For that reason, we proposed a second method based on convolutional neural networks.

In the second approach, presented in subsection 4.2.2, we proposed an U-Net to segment the optic disc and optic cup. Different to the previous method, we were focused on segmentation only, not glaucoma assessment. It is because this method was presented for the REFUGE challenge. In this challenge, we were asked to develop a method for optic disc and optic cup segmentation that presents a high Dice index on the validation and test set of the REFUGE database. A Dice index of 0.91 and 0.78 were obtained for the optic disc and optic cup, respectively. The results obtained from this approach demonstrate the potential of this system to be used in a clinical setting to help ophthalmologists in the glaucoma detection task.

Although a rigorous comparison between these two approaches is interesting, it has been demonstrated in other works in the literature that segmentation algorithms based on Deep Learning techniques significantly outperform algorithms based on classical approaches (Sevastopolsky 2017; Ronneberger, Fischer, and

Brox 2015). For that reason, the comparison between these two methods will not significantly contribute to this thesis.

Even though Deep Learning approaches seem more interesting for segmentation tasks, classical approaches, such as the Stochastic Watershed transformation, are still of great interest. They could be used in cases where the number of images is very limited and/or Deep Learning techniques do not perform well. Additionally to this, classical approaches could be used in segmentation tools to obtain semi-automatic ground-truth, in which experts can verify and correct if necessary. This could significantly reduce the time and cost of manual segmentation.

Nevertheless, we show in Figure 4.19 some of the best and the worst examples of cup segmentation when using the Stochastic Watershed transformation and the U-Net. We can see from this figure that although both systems perform well, the U-Net makes fewer mistakes than the Stochastic Watershed approach. It is because the U-Net has the ability to learn more discriminative features than only the intensity.

In summary, we have presented two approaches for optic cup segmentation. The first one uses the classical approach of machine learning and a second approach uses a deep learning technique. In the first approach, we employed an external algorithm to segment the optic disc. In the second approach, however, we segment the optic disc and the optic cup by using the same architecture. Another difference is that cup adjustment was only applied for the Stochastic Watershed approach.

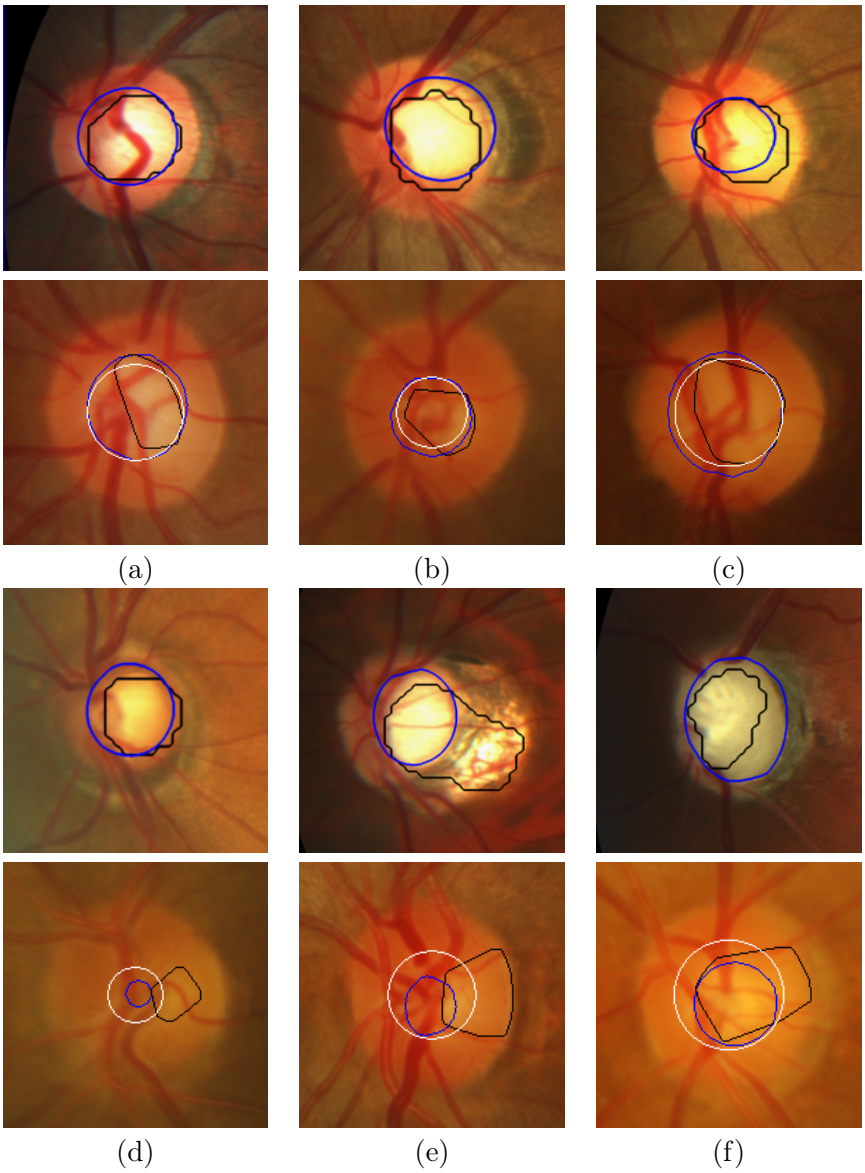


Figure 4.19: Examples of the optic cup segmentation performed by the U-Net (upper rows) and the Stochastic Watershed transformation (lower rows). Upper row (a-d) U-Net best cases and upper row (e-f) U-Net worst cases. Lower row (a-c) Stochastic Watershed best cases and lower row (d-f) Stochastic Watershed worst cases. Black lines represent the predicted cup, blue lines represent the ground-truth and white lines represent the cup adjustment.

Chapter 5

Glaucoma Classification and Retinal Image Synthesis using Deep Learning

Current algorithms for automatic glaucoma assessment using fundus images usually rely on handcrafted features based on segmentation, which are affected by the performance of the chosen segmentation method and the extracted features. To avoid this problem, we address the glaucoma classification task skipping the segmentation step. In this chapter, we present the latest convolutional neural network architectures fine-tuned for the automatic glaucoma assessment task. Additionally, retinal image synthesizer and semi-supervised learning methods for automatic glaucoma classification are presented.

5.1 Convolutional Neural Networks (CNNs) for Glaucoma Assessment

CNNs have been applied to a wide variety of applications in image segmentation and classification. In this section, we show how CNNs, that were used to classify natural non-medical images, can be applied for retinal image classification using ImageNet-trained CNN architectures.

ImageNet-trained CNN architectures

In this subsection, the CNN architectures used for glaucoma assessment are described. They are the state-of-the-art in image classification, object detection and localization. Moreover, their pre-trained versions are publicly available to use for applications that differ from the initial task.

VGG16 and VGG19

The architecture of the VGG16 and VGG19 networks are based on the same model and characterized by their simplicity (Simonyan and Zisserman 2014). They used 3x3 convolutional filters stacked on top of each other in increasing depth and reducing the volume size by max pooling. In the end, three fully-connected (FC) layers, two of 4096 and one of 1000 neurons (this one represents the class scores) are followed by a softmax classifier (See Figure 5.1). This architecture was the basis of their ImageNet 2014 submission, with which they secured the first and the second places in the localization and classification tracks.

The main advantages of this architecture in comparison with previous architectures such as AlexNet (Krizhevsky, Sutskever, and G. E. Hinton 2012), is that it has fewer parameters to optimize and has more non-linear rectification layers, which makes the decision function more discriminative.

We used the CNN weights, available in Keras Applications, to classify glaucomatous images: `https://keras.io/applications/#available-models`.

GoogLeNet

This CNN was first introduced by (Szegedy, W. Liu, et al. 2015). With this architecture, the authors made the submission to the ImageNet 2014 and secured the first place in the object detection track. Based on the Inception module, they developed an architecture significantly more complex and deeper than all previous CNN architectures. The main idea of the Inception module is the concatenation of filters of different sizes and dimensions into a single new filter, acting as a "multi-level feature extractor" by computing 1x1, 3x3, and 5x5 convolutions within the same module of the network. As it is shown in Figure 5.2.

The initial architecture of this CNN was called GoogLeNet or Inceptionv1. However, after some improvements to the architecture, it has been called Incep-

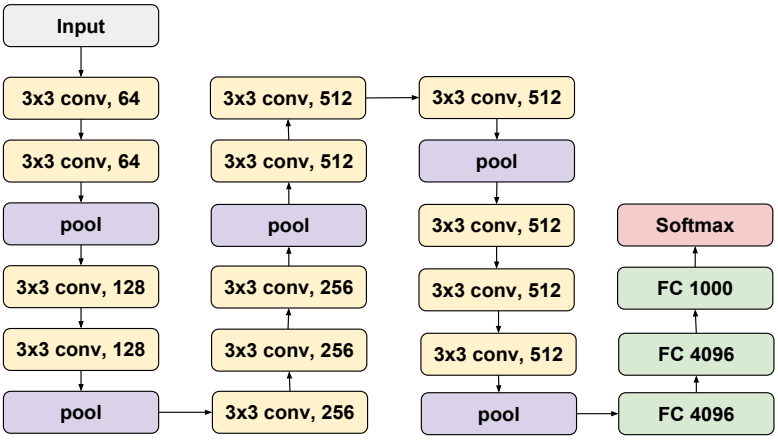


Figure 5.1: Illustration of the VGG16 architecture with its 16 weighted layers (13 convolution layers and 3 Fully Connected Layers, represented by yellow and green colours, repectively). FC stands for Fully Connected and is followed by the number of neurons. In the yellow blocks, the first part represents the convolution block dimension (i.e. 3x3 conv), followed by the number of filters (64, 128, 256 and 512).

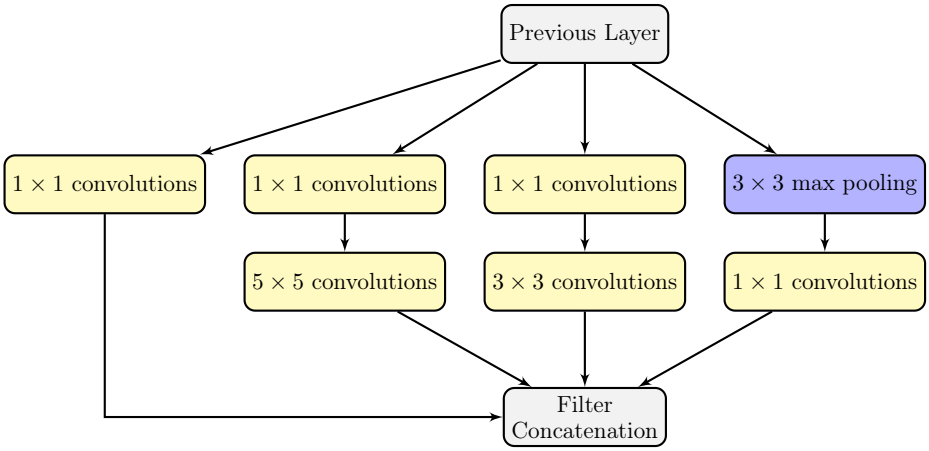


Figure 5.2: Illustration of the inception module proposed by (Szegedy, W. Liu, et al. 2015).

tionvN where N refers to the version number put out by Google. For instance, the **Inceptionv3** architecture, included in the Keras core, was presented by

(Szegedy, Vanhoucke, et al. 2016), where they proposed updates to the inception module to further boost classification accuracy. For that reason, we used the pre-trained architecture Inceptionv3 for the glaucoma classification task.

It is worthy to mention that, because of the weights of the last Inception version (Inceptionv4) are not in Keras, that Inception version was not used in this work.

Microsoft ResNet

Microsoft ResNet is the CNN architecture proposed by the Microsoft Research Asia team (MSRA) to the ImageNet competition in 2015. With this architecture, the MSRA not only won the competition, but also set new records in classification, detection, and localization. Unlike traditional sequential network architectures such as AlexNet (Krizhevsky, Sutskever, and G. E. Hinton 2012) and VGG, ResNet is an "exotic architecture" that relies on residual blocks (See Figure 5.3). With these blocks, they solved the degradation problem, a problem that appears when the network depth increases, and is not caused by overfitting.

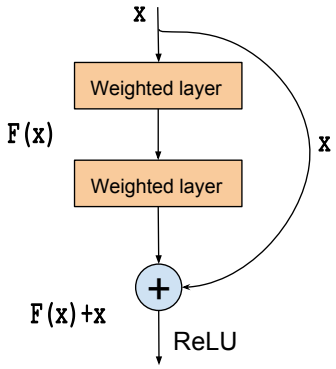


Figure 5.3: Illustration of a Residual block proposed by (He et al. 2016), where x is the input and $F(x) + x$ the output of the residual block, before the rectified linear unit (ReLU).

The residual block is used to construct the network itself. It is a collection of convolutions and pooling layers that leads to a macro-architecture. The description of this block and the whole architecture was published by (He et al. 2016). In their work, they demonstrated that extremely deep networks

can be trained using standard Stochastic Gradient Descent (SGD) optimizer through the use of residual blocks.

Given the obtained performance by (He et al. 2016) and the fact that the weights of this network are available in Keras, the ResNet50 architecture is also fine-tuned to the glaucoma assessment task. The obtained results will be described in the results section.

Xception

Xception stands for Extreme Inception and is the name of the architecture proposed by (Chollet 2016). It is an extension of the Inception architecture which replaces the standard Inception modules with depthwise separable convolutions. This type of convolution is also called "separable convolution" in frameworks such as TensorFlow and Keras. It consists in a depthwise convolution (spatial convolution performed independently over each channel), followed by a pointwise convolution, i.e. a 1x1 convolution (Chollet 2016). In Figure 5.4(a) it is possible to see an illustration of the simplified version of the Inception module and in Figure 5.4(b) an illustration of the extreme Inception module.

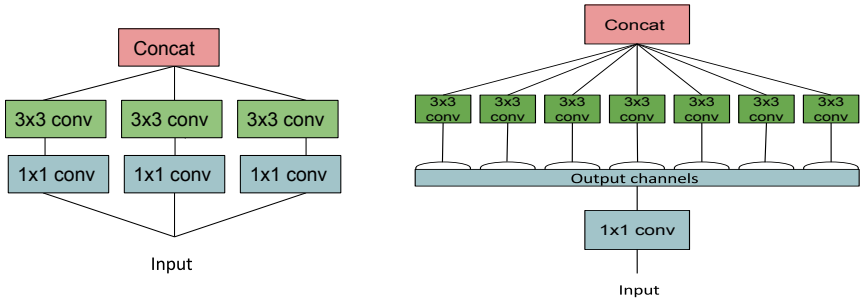


Figure 5.4: Illustration of the Inception modules. (a) A simplified Inception module and (b) an "extreme" version of the Inception module (Xception) with one spatial convolution per output channel. Concat and conv stand for concatenation and convolutions, respectively. The numbers 1×1 and 3×3 are the convolution block dimensions.

The fundamental hypothesis behind this architecture is that the cross-channel correlations and spatial correlations are sufficiently decoupled to be mapped separately.

As the other previously mentioned CNN architectures, the weights of the ImageNet-trained Xception architecture are provided as part of the Keras Applications module. For that reason, and given the results presented in the work made by (Chollet 2016), this architecture was fine-tuned to the glaucoma assessment task and its performance results are also presented in the results section.

Fine-Tuning

CNN architectures can be either trained from scratch (with random initialization) or fine-tuned from models that were trained on a very large database (e.g. ImageNet, which contains 1.2 million images with 1000 categories (Deng et al. 2009)).

Fine-tuning a CNN is a procedure that is based on transfer learning (Bengio 2012). In order to optimise the network and minimize the error in a more specific task, fine-tuning consists of two parts: a) the weight initialization of the convolutional layers using the weights of the pre-trained CNN with the same architecture and b) the replacement of the classification function or the number of nodes in the last fully connected layer. All the weights in the layers are initialized except for the last fully connected layer whose nodes depend on the number of classes of the new task. Using fine-tuning technique, the initialized parameters, or "knowledge", of the pre-trained networks from natural non-medical images is transferred to the glaucoma classification task that uses fundus images.

After the weight initialization, the CNN can be fine-tuned in a layer-wise manner. It means, start fine-tuning only the last layer, then include layers until fine-tuned all the layers in the network. Depending on the number of fine-tuned layers, this procedure can be called "shallow tuning" if only the last few layers are fine-tuned and "deep tuning" if all the layers are fine-tuned (Tajbakhsh et al. 2016).

In the literature, other applications have successfully used shallow- and deep-tuned CNNs, obtaining promising results in different computer vision tasks (Tajbakhsh et al. 2016; H. Chen et al. 2015). In this work, both techniques, shallow- and deep-tuning, were explored and evaluated. Deep-tuning turned out to be the best technique when fine-tuning ImageNet-trained CNN for glaucoma assessment.

For all the experiments executed in this work, the open source Deep Learning library Keras (Chollet et al. 2015) and NVIDIA GeForce GTX 1080 graphic

cards were used. Keras library is a simple way to use, implement and fine-tune CNNs architectures built on top of Theano (Al-Rfou et al. 2016) or TensorFlow (Abadi et al. 2015).

Preprocessing

The fundus images used for the fine-tuning process were cropped around the optic disc with a size of 1.5 times the optic disc radius, which turned out to be the more efficient way to obtain better results when using CNN for glaucoma assessment task. Cropping the images around the optic disc has also a clinical reason, glaucoma disease affects mainly the optic disc and its surroundings.

Subtracting the vessels and doing CLAHE enhancement do not provide any improvement to the classification results (Orlando, Prokofyeva, et al. 2017). An example of the images used for the fine-tuning and testing time is showed in Figure 5.5(b).

(a) (b)

Figure 5.5: Digital fundus images. (a) Original fundus image and (b) Cropped image used for fine-tuning and testing.

Cross Validation

Cross-validation (CV) (Stone 1974; Geisser 1975) is one of the most commonly used methods for the performance evaluation of a predictive model. Basically, based on data splitting, part of the data is used for training and the rest of the data is used for test. A version of CV is the k-fold CV. In this version, the data is randomly partitioned into k equal-size subsets. $k-1$ subsets are used

for training and the remaining subset is used for test. The average prediction error of each split is obtained to estimate the model performance (Y. Zhang and Yang 2015). More information about the k-fold CV can be found on the survey made by (Arlot and Celisse 2010).

Following the procedure described by (Hastie, Tibshirani, and Friedman 2009), the performance evaluation of all the pre-trained models used in this subsection were obtained using the k-fold CV with $k = 10$.

5.2 Ensemble Setting with CNNs

Several papers have shown that an ensemble method can deliver outstanding performance in reducing the testing error. For instance, (Krizhevsky, Sutskever, and G. E. Hinton 2012) showed that their ensemble model with 5 convolutional nets achieved a top-1 error rate of 38.1% on the ImageNet 2012 classification benchmark, compared to the top-1 error rate of 40.7% given by the single model. Another example is the work made by (Zeiler and Fergus 2013) in which they showed that by the ensemble of 6 convolutional nets, they reduced the top-1 error from 40.5% to 36.0%.

In this section, we proposed a system based on an ensemble of four CNNs for automatic glaucoma assessment using single retinal fundus images. A schema of the proposed method can be seen in Figure 5.6

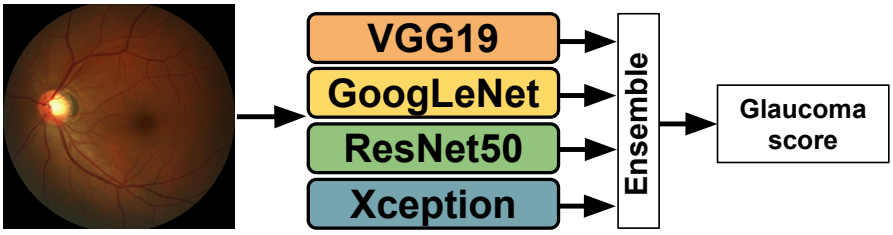


Figure 5.6: Schema of the ensemble setting used for glaucoma classification

We used the VGG19, GoogLeNet (InceptionV3), ResNet50 and Xception networks for this task. Each network was fine-tuned independently and then combined in an average ensemble setting to finally obtain a glaucoma score for each image.

This method was also used for the REFUGE challenge (See subsection 4.2.2), hosted at the MICCAI 2018 conference in conjunction with OMIA workshop https://refuge.grand-challenge.org/home/. (Group name: Cvblab). Additionally to the evaluation and comparison of optic disc/cup segmentation algorithms, this challenge also evaluates the performance of automated algorithms for glaucoma detection.

5.3 Retinal Image Synthesis and Semi-supervised Learning

Retinal Image Synthesis using VAE and DCGAN

The limited number of glaucoma-labelled images available in the literature is a huge problem when trying to generalise. Therefore, retinal image synthesizer has been a focus of the scientific community, as it was mentioned in Subsection 2.4.2.

In this subsection, the Variational Autoencoder (VAE) and the Deep Convolutional Generative Adversarial Network (DCGAN) are trained to synthesize cropped retinal images. Further image quality evaluation is also presented.

Variational Autoencoder

The Variational Autoencoder (VAE) is composed by two neural networks: the approximate inference network (or encoder), that maps a training example to a latent (hidden) space, and the decoder network that maps from the latent space to a synthetic sample. In this work, we used the architecture proposed in (D P Kingma and Welling 2013), in which the prior over the latent space is a centred isotropic multivariate Gaussian, and the encoder and decoder are fully-connected neural networks with a single hidden layer.

During training or learning phase, the encoder obtains the latent variables z from the input data and the decoder draws those variables to generate a sample. After that, during the generation phase, VAE draws samples from the latent space that run through the decoder to finally obtain a synthetic sample. The VAE architecture can be seen in Figure 5.7(a).

Generative Adversarial Network

Generative Adversarial Networks, or GAN, are deep neural net architectures comprised of two nets. One is called the generator and the other (the adversary) is called the discriminator.

A class of CNN called Deep Convolutional Generative Adversarial Networks (DCGAN) that are based on the adversarial strategy was used for this work. This architecture was a major improvement on the first GAN, generating better quality images and more stability during the training stage. Following the guidelines to construct the generator and discriminator, described in the paper written by (Radford, Metz, and Chintala 2015), we implemented and trained them on cropped retinal images using the original discriminator and generator cost functions.

In the same way, as in the VAE approach, synthetic image generation using the DCGAN mainly consists of two phases: a learning phase and a generation phase. For the training phase, the generator draws samples from an N-dimension normal distribution (latent variables) that run through the generator to obtain a synthetic sample and the discriminator attempts to distinguish between images drawn from the generator and images from the training set. A figure of a DCGAN architecture can be seen in Figure 5.7(b).

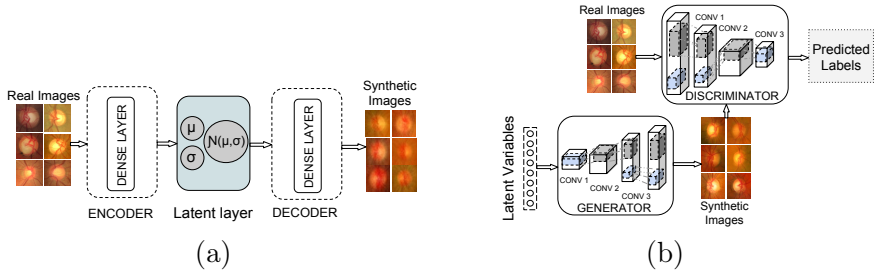


Figure 5.7: Schema of the VAE and DCGAN architecture. (a) VAE architecture and (b) DCGAN architecture.

In the following subsection, we focused on the development of an image synthesizer and, at the same time, a semi-supervised learning method for glaucoma assessment using the DCGAN architecture. Different to this subsection, the method presented in the next subsection employed a lot more images (86926 cropped retinal images) and not only trained an image synthesizer, but also a glaucoma classifier.

Semi-supervised Learning and Retinal Image Synthesis using DCGAN

Semi-supervised learning is an area in machine learning and a special form of classification in which a large amount of unlabelled data, along with the labelled data, are used to build better classifiers. Other names for this technique are "learning from labelled and unlabelled data" or "learning from partially labelled/classified data" (Zhu 2005).

Semi-supervised learning has been of great interest both in theory and in practice because it requires less human effort and gives higher accuracy. Given the scarce number of glaucoma-labelled images, this technique can significantly help the development of automatic glaucoma assessment systems using retinal images. For that reason, we decided to use the power of the DCGAN to develop a semi-supervised learning method for training a glaucoma classifier and at the same time an image synthesizer. In that way, we train a glaucoma classifier making use of a small set of glaucoma-labelled images, a set of unlabelled images and the synthetic images generated by the DCGAN to train a glaucoma classifier. This setting learns not only from the small set of glaucoma-labelled images but also from the unlabelled and synthetic images.

As in Section 5.3, we used the Deep Convolutional Generative Adversarial Networks (DCGAN) as image synthesizer and semi-supervised learning method. A schema of the trained architecture can be seen in Figure 5.8.

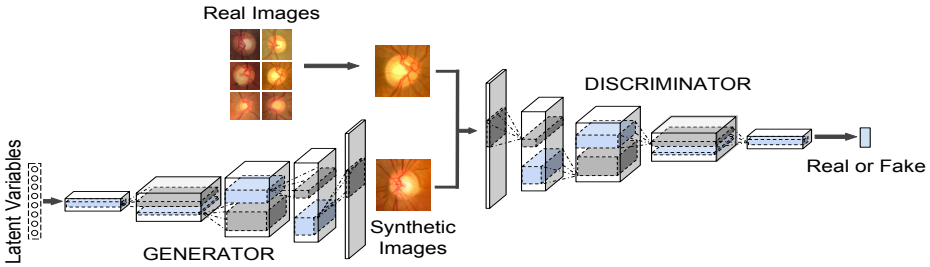


Figure 5.8: Schema of the DCGAN architecture. The generator takes as input a vector of latent variables (N-dimension normal distribution) to synthesize retinal images while the discriminator tries to predict whether the input is a real or a generated image.

Model Architecture and Hyperparameters

The DCGAN architecture has several improvements on the vanilla GAN. Among them are the replacement of all pooling layers with strided convolutions in the discriminator and fractional-strided convolutions in the generator, the use of batch normalization (batchnorm) in both the generator and the discriminator, the replacement of fully connected hidden layers with the average pooling at the end, the use of ReLU activation in the generator for all layers except for the output and the use of LeakyReLU activation for all layers in the discriminator. We followed the guidelines to construct the generator and discriminator described in the paper written by Radford et al. (Radford, Metz, and Chintala 2015).

The architecture of the image synthesis and semi-supervised learning method differs only on the last output layer (Fully connected layer) of the discriminator: one neuron for image synthesis (Synthetic or Real, **FC-1**) and three neurons for semi-supervised learning method (Normal, Glaucoma and Synthetic class **FC-3**. See Figure 5.9). The architecture details are presented in Table 5.1.

Table 5.1: The discriminator and generator CNNs used for retinal image synthesis. conv stands for convolution, upconv stands for upconvolution, FC stands for Fully Connected and batchnorm stands for batch normalization.

Discriminator *D*	**Generator *G***
Input 128×128 Color image	Input $\in \mathbb{R}^{100}$
5×5 conv, LeakyReLU (alpha 0.2), stride 2, Dropout 0.4	FC 32×32×256, ReLU, batchnorm
5×5 conv, LeakyReLU (alpha 0.2), stride 2, Dropout 0.4	UpSampling2D size 2 5×5 upconv, ReLU, stride 1, batchnorm
5×5 conv, LeakyReLU (alpha 0.2), stride 1, Dropout 0.4	UpSampling2D size 2 5×5 upconv, ReLU, stride 1, batchnorm
FC-1 output layer, sigmoid activation (Output for **DCGAN**)	5×5 upconv, ReLU, stride 1, batchnorm
FC-3 output layer, softmax activation (Output for **SS-DCGAN**)	5×5 upconv, Tanh, stride 1

Regarding image resolution, we modified the architecture to handle 128×128 px, which is closer to the average resolution of original fundus images cropped around the optic disc. No pre-processing was applied to the training images, no data augmentation was used and class weights for the Glaucoma, Normal and Not-labelled images were set to train the semi-supervised learning method.

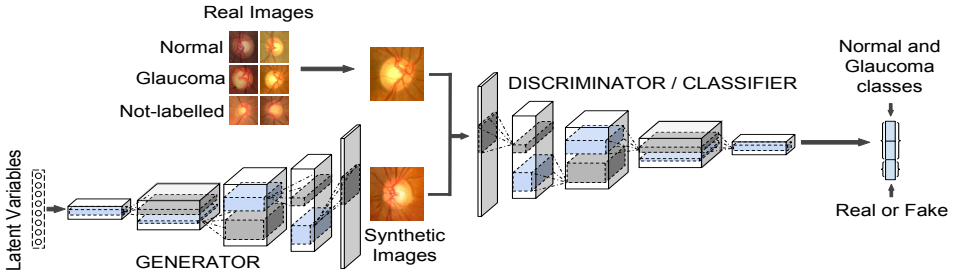


Figure 5.9: Schema of the DCGAN architecture used as a Semi-supervised learning method. The DCGAN discriminator is converted into a 3-class classifier (Normal, Glaucoma and Real/Fake class).

Although research in adversarial models continues to improve, stability on training these models is still a challenging task. For that reason, we followed the recommendations given in (Chintala et al. 2016) to reach stability on training the DCGAN and the semi-supervised learning method (SS-DCGAN). Recommendations such as normalizing the input images between -1 and 1, using Stochastic Gradient Descent (SGD) optimizer for the discriminator and ADAM optimizer for the generator, using a Gaussian distribution for the latent space and mini-batches containing only all real images or all generated images were used for training the models in this work.

Model Losses

As in a vanilla GAN, the DCGAN model emulates a competition in which the Generator G attempts to produce realistic images, while the Discriminator D classifies between images from the training set with their corresponding labels and images produced by the generator. The main goal of the DCGAN model is to maximise the miss-classification error of the Discriminator while the generator produces more realistic images trying to fool the discriminator. This competition is also called a two-player minimax game and it can be described as follows:

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (5.1)$$

where $\mathbb{E}_{x\sim p_{data}(x)}$ is the expectation over the training data and $\mathbb{E}_{z\sim p_z(z)}$ is the expectation over the data produced by the generator. $D(x)$ represents the probability that x came from the training data rather than the data produced

by the generator and $G(z)$ represents the probability of z being produced by the generator. Therefore, the system is trained to minimize $\log(1 - D(G(z)))$ and maximise $\log(D(x))$ (Goodfellow, Pouget-Abadie, et al. 2014).

However, regarding the semi-supervised learning method using the DCGAN architecture, instead of binary classification, the discriminator is transformed into a K-class classifier (Salimans et al. 2016; Dai et al. 2017). Therefore, the semi-supervised setting loss function is composed of two parts; the supervised and the unsupervised loss function (Diederik P Kingma et al. 2014):

$$L = L_{supervised} + L_{unsupervised}, \tag{5.2}$$

where the supervised loss is defined by the cross-entropy loss function as in a supervised learning setting with K classes:

$$L_{supervised} = -\mathbb{E}_{x,y\sim p_{data}(x,y)} \log(p_{model}(y|x, y < K + 1)), \tag{5.3}$$

and the unsupervised loss function is, in fact, the standard GAN minimax game:

$$L_{unsupervised} = -\left\{\mathbb{E}_{x\sim p_{data}(x)} \log D(x) + \mathbb{E}_{z\sim p_z(z)} \log(1 - D(G(z)))\right\} \tag{5.4}$$

where $D(x) = 1 - p_{model}(y = K + 1|x)$, being $p_{model}(y = K + 1|x)$ the model predictive distribution and K the number of real classes.

In other words, the unsupervised loss function is computed to differentiate real training images and fake images and the supervised loss computes the individual real classes probabilities. In this work, these classes are Glaucoma and Normal.

5.4 Results

Convolutional Neural Networks (CNNs) for Glaucoma Assessment

All the CNN architectures described in Section 5.1 were fine-tuned to the glaucoma assessment task using their ImageNet-trained versions available in the Keras core. To properly fine-tune these architectures, the last fully connected layer was changed for a global average pooling layer (GlobalAveragePooling2D) followed by a fully connected layer of two nodes (class scores, glaucoma and healthy) and a softmax classifier. With this change to the original architec-

tures, all the models were trained using the Stochastic Gradient Descent (SGD) optimizer, with a batch size of 8 a fixed learning rate of $1e^{-4}$ and a momentum of 0.9.

A total of 1707 images belonging to the ACRIMA, HRF (Köhler et al. 2013), Drishti-GS1 (Sivaswamy et al. 2014), RIM-ONE v1 (Medina-Mesa et al. 2015), and sjchoi86-HRF (sjchoi86 2017) databases were used (788 glaucomatous images and 919 normal images). It is important to remark that images from all databases were cropped around the optic disc. Only the images in the RIM-ONE database came originally cropped around the optic disc.

Although we got a considerable number of images to fine-tune the networks, images were randomly rotated, zoomed and horizontal and vertical flipped to augment the number of images; which avoid overfitting and increase the performance of the models. The images were also rescaled to the default input size of each CNN architecture (224×224 for VGG16, VGG19 and ResNet50 and 299×299 for Inceptionv3 and Xception).

The results presented in this section were obtained for 200 epochs, which turned out to be the optimum number of epochs to get the best performance for the fine-tuning process.

As a first experiment, only the last weighted layer of the CNN architectures was fine-tuned, keeping the other layers in a "not trainable" mode. Afterwards, the number of weighted layers that were fine-tuned was increased until updating all the layers in the CNN. After these experiments, it was observed that the best results were obtained when all the layers were fine-tuned.

For the model evaluation, 10-fold cross validation was performed. Therefore, 10 values of area under the curve (AUC), accuracy, specificity, sensitivity and F-score were obtained. Afterwards, the average and standard deviation of these values were calculated for each CNN architecture. Results for each fine-tuned model are presented in Table 5.2.

Table 5.2: Results for each model doing deep tuning and 10-fold cross validation

ModelName	AUC	Accuracy	Specificity	Sensitivity	Fscore
VGG16 (Simonyan and Zisserman 2014)	0.9632 (0.0149)	0.8948 (0.0253)	0.8816 (0.0612)	0.9057 (0.0331)	0.9005 (0.0231)
VGG19 (Chatfield et al. 2014)	0.9686 (0.0158)	0.9069 (0.0318)	0.8846 (0.0362)	0.9240 (0.0434)	0.9125 (0.0312)
InceptionV3 (Szegedy, W. Liu, et al. 2015)	0.9653 (0.0135)	0.9000 (0.0201)	0.8752 (0.0358)	0.9216 (0.0311)	0.9056 (0.0236)
ResNet50 (He et al. 2016)	0.9614 (0.0171)	0.9029 (0.0249)	0.8943 (0.0350)	0.9105 (0.0282)	0.9076 (0.0251)
Xception (Chollet 2016)	**0.9605 (0.0170)**	**0.8977 (0.0264)**	**0.8580 (0.0398)**	**0.9346 (0.0247)**	**0.9051 (0.0274)**

Additionally to the measurements reported in Table 5.2, the ROC curves for the average specificity and sensitivity, obtained by performing 10-folds cross validation, were plotted in Figure 5.10.

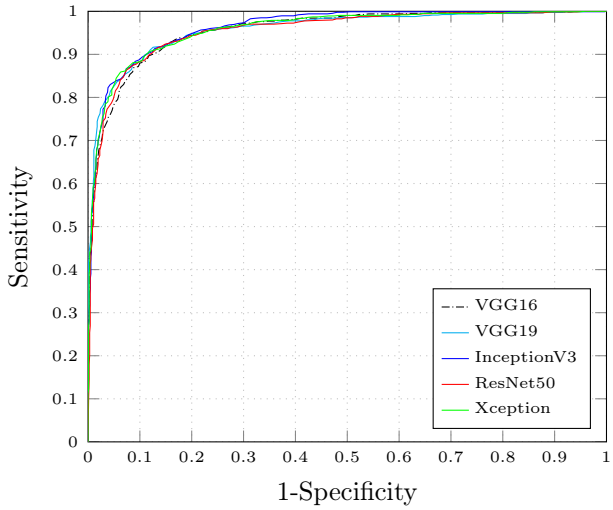


Figure 5.10: Average ROC curves for each fine-tuned CNN architecture

It is possible to see that all the proposed fine-tuned models have really good performance for the glaucoma assessment task. As the performance of the models is comparable, characteristics of the CNN such as the number of parameters can be used to determine which model is better than the others.

In Table 5.3, the number of parameters and the obtained AUC of each CNN architecture are shown. It is possible to see that, although VGG16 and VGG19 present higher AUC than the Xception architecture, they have much more parameters to fine-tune, which requires more computation power and resources. Therefore, the Xception architecture presents a better trade-off between the number of parameters and obtained AUC than the other architectures.

Given the changes in illumination and high variability among databases, the glaucoma assessment using fundus images is not an easy task. A developed method that properly classifies images from a certain database/s does not necessarily perform well when it is applied to images that come from a different database. A critical experiment that evaluates the performance of a glaucoma classifier consists of using images that come from a different sensor or database. For that reason, five different experiments using the Xception

Table 5.3: Number of parameters and obtained AUC for each architecture.

Model Name	# parameters (in millions)	AUC
VGG16	138	0.9632 (0.0149)
VGG19	144	0.9686 (0.0158)
InceptionV3	23	0.9653 (0.0135)
ResNet50	25	0.9614 (0.0171)
Xception	**22**	**0.9605 (0.0170)**

architecture and all public glaucoma-labelled databases (HRF, Drishti-GS1, RIM, sjchoi86-HRF and ACRIMA) were done. First, the Xception architecture was fine-tuned using all the databases except the images that belong to the database to be tested. Secondly, the trained model is tested on the desired database. This approach is repeated to test HRF, Drishti-GS1, RIM, sjchoi86-HRF and ACRIMA database. The results obtained from these experiments are presented in Figure 5.11 and Table 5.4, in which is possible to see that although the Xception architecture was fine-tuned without using images from these databases, its performance is promising.

Table 5.4: Obtained results for HRF, Drishti-GS1, RIM, sjchoi86-HRF and ACRIMA databases, using Xception architecture represented in AUC, Accuracy, sensitivity and specificity.

Database	AUC	Accuracy	Sensitivity	Specificity	# images
HRF	0.8354	0.8000	0.8333	0.7778	45
Drishti-GS1	0.8041	0.7525	0.7419	0.7143	101
RIM-ONE	0.8575	0.7121	0.7931	0.7990	455
sjchoi86-HRF	0.7739	0.7082	0.7033	0.7030	401
ACRIMA	0.7678	0.7021	0.6893	0.7020	705

It can be seen from Table 5.4 how robust is the developed system in terms of AUC, accuracy, sensitivity and specificity when evaluating images from databases that differ to the ones used during the fine-tuning process.

Thanks to the public availability of the Drishti-GS1 database, a comparison with other state-of-the-art algorithms that used this database is possible. For instance, in the work developed by (Chakravarty and Sivaswamy 2016), they

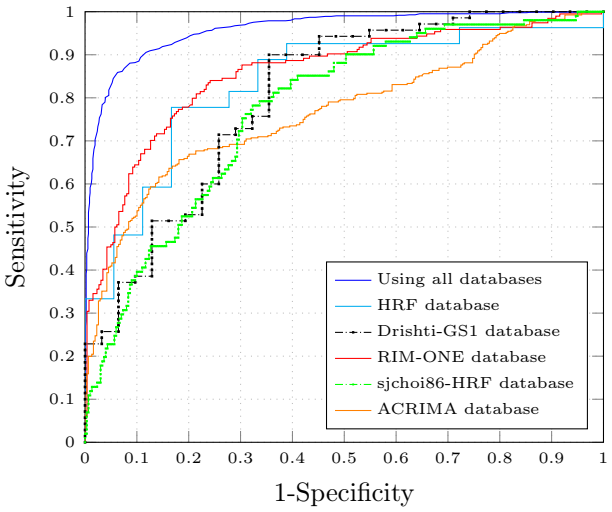


Figure 5.11: ROC curves for the Xception architecture in different experiments. When using HRF, Drishti-GS1, RIM, sjchoi86-HRF and ACRIMA database only as test set and using as test set a percentage of all the HRF, Drishti-GS1, RIM-ONE, sjchoi86-HRF and ACRIMA databases combined (Using all databases).

obtained an AUC of 0.78 when tested their method on this database. Or the work presented by (Orlando, Prokofyeva, et al. 2017), in which they obtained an AUC of 0.76 using pre-trained CNNs applied to the Drishti-GS1 database. It can be seen from Table 5.4 that the proposed method outperformed the existing works. Moreover, it must be taken into account that (Chakravarty and Sivaswamy 2016) and (Orlando, Prokofyeva, et al. 2017) evaluated their methods using the same database for training and test different to the experiments we performed in which Drishti-GS1 database is only used for test.

Although it is not possible to do a fair comparison with other CNN-based methods, as the databases used in other works are private and complete details about the implemented algorithms are not given, the results presented in (X. Chen et al. 2015) are outperformed by the proposed method. In their work, they obtained a maximum AUC of 0.887, which is lower than the AUC we obtained (0.9605).

Ensemble Setting with CNNs

For the ensemble setting (using VGG19, InceptionV3, ResNet50 and Xception networks), we employed a total of 2140 images from five public databases. These images belongs to the HRF (30 images) (Köhler et al. 2013), Drishti-GS1 (101 images) (Sivaswamy et al. 2014), RIM-ONE v3 (159 images) (Medina-Mesa et al. 2015), ORIGA-light (650 images) (Z. Zhang et al. 2010), and REFUGE (1200 images) (REFUGE 2018) databases.

As mentioned in Section 5.2, the ensemble setting was implemented for the REFUGE challenge. In this challenge, the main goal in the classification task was to obtain the maximum AUC on the validation (400 images) and test set (400 images) of the REFUGE database. Therefore, we used 1340 for training and 800 images for validation and test, obtaining an AUC of 0.94 for the validation set and 0.88 for the test set.

Retinal Image Synthesis and Semi-supervised Learning

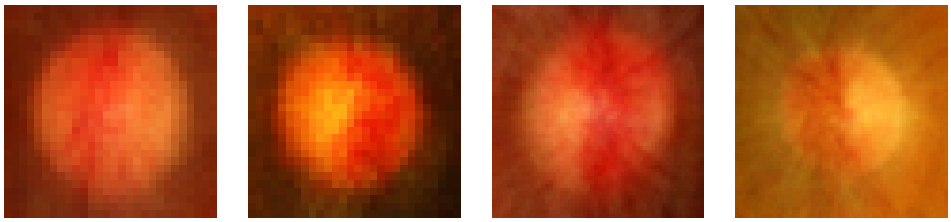
Our main goal in this section is to show the methods used for image synthesis and semi-supervised learning. In the first part (subsection "Retinal Image Synthesis using VAE and DCGAN"), we present the performance obtained from the VAE and DCGAN architectures, when trained on 2357 glaucoma-labelled images, for the images synthesis task. After that (subsection "Semi-supervised Learning and Retinal Image Synthesis using DCGAN"), we present the performance results of the DCGAN and the SS-DCGAN when trained on 86926 images. This second part is focused on both, image synthesis and the semi-supervised learning method.

Retinal Image Synthesis using VAE and DCGAN

In this section, we will present the obtained results of using the VAE and DCGAN architectures for retinal image synthesis. These architectures were trained on 2357 cropped retinal images from five public glaucoma-labelled databases: HRF (Köhler et al. 2013) (45 images), Drishti-GS1 (Sivaswamy et al. 2014) (101 images), ORIGA-light (Z. Zhang et al. 2010) (650 images) RIM-ONE (Medina-Mesa et al. 2015) (455 images) and sjchoi86-HRF (sjchoi86 2017) database (401 images) and our private database, ACRIMA (705 images). All images were manually cropped around the optic disc by an expert, with the exception of RIM-ONE images that are already cropped in the original database.

In order to keep a trade-off between performance and system complexity, the images were automatically re-scaled to the following resolutions: 28×28 pix, 56×56 pix, 112×112 pix and 224×224 pix. For each image size, we tested a range of N-dimensional latent spaces from 32 to 100 latent variables. Each latent space was explored in order to check that the systems do not memorise the training database and, at the same time, it generates plausible retinal images. To do that, we used spherical interpolation to evaluate intermediate latent representation points (White 2016).

For training the VAE model, we ran several tests and found out that the best results are obtained when using a 100-dimension latent space and image resolution of 28×28 and 56×56 pix. Running for 500 epochs and a small batch size of 64, we obtained the synthetic images presented in Figure 5.12



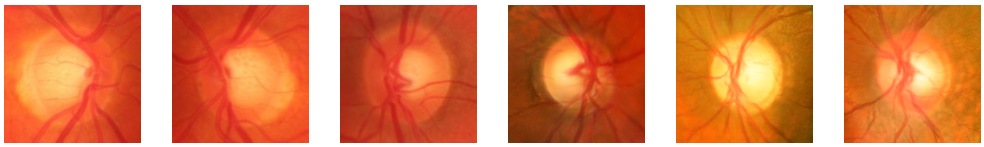
(a) (b) (c) (d)

Figure 5.12: Examples of images synthesised by the VAE architecture. (a-b) Images of 28×28 pix and (c-d) images of 56×56 pix.

Although the texture of the synthetic images obtained from VAE is similar to the real images, they are blurry and do not have sharp vessels as in real fundus images. For that reason, we only trained on the resolution 28×28 and 56×56 pix.

Regarding the DCGAN architecture, we found that realistic images were obtained when using an image size of 224×224 pix, a small batch size of 32 and 35000 steps. Examples of them are shown in Figure 5.13(d-f).

The main advantage of using the DCGAN architecture is that synthetic images are sharper than the ones synthesised by the VAE approach. We can see from Figure 5.13(d-f) well-defined optic disc shapes, how blood vessels converge into the optic disc and the right and left eye symmetry. For this reason, we continued with the evaluation of only the images synthesised by the DCGAN architecture and discarded those obtained by the VAE architecture.



(a) (b) (c) (d) (e) (f)

Figure 5.13: Examples of real images and examples of synthetic images generated by the DCGAN architecture. (a - c) Real images and (d - f) Synthetic images.

Experimental Evaluation

Although a great effort to develop objective metrics that correlate with perceived quality measurement has been made in recent years, it is still a challenging task. In the case of quality evaluation of synthetic images, it should be specific for each application (Theis, Oord, and Bethge 2016). For that reason, we created a database composed of 200 images: 100 synthetic images and 100 real images (randomly selected from the training set) to perform a qualitative and quantitative evaluation of our methods.

Qualitative evaluation of the database was carried out by ten experts with 3 to 10 years of experience. They evaluated the anatomical consistency and plausibility of the synthetic images. For this purpose, we developed a web application (`https://cvblab.synology.me/ganval/`) that randomly shows the 200 images (one each time) to be evaluated by the expert. For each expert, we calculated the Cohen's kappa coefficient using the ground-truth labels (Fake - Real) and the labels given by each expert. The Cohen's kappa coefficient ranges from -1 to $+1$, where 0 represents the amount of agreement that can be expected from random chance, and 1 represents a perfect agreement between the ground-truth and the expert. The obtained results are presented in Figure 5.14.

It can be seen from Figure 5.14 that although the Cohen's Kappa coefficient is high for two experts with high expertise, most of them were fooled when evaluating synthetic images.

Quantitative evaluation was carried out by measuring and further comparison of the average proportion of pixels belonging to the vessel and optic disc structures between the synthetic and real images. To obtain these percentages, optic disc masks were manually segmented by clinical experts and the vessel masks were automatically segmented using the method proposed in (Morales, Naranjo, Navea, et al. 2014). Table 5.5 shows the obtained results.

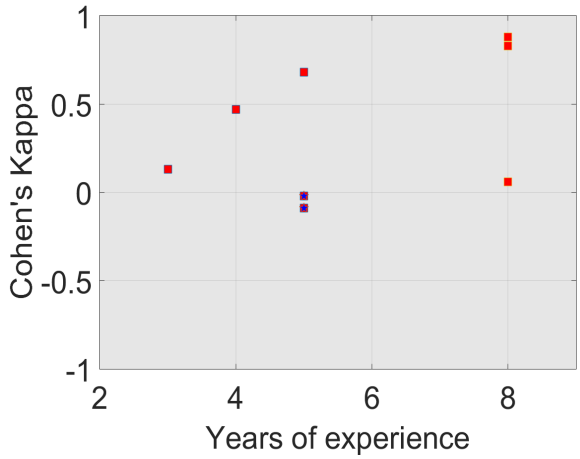


Figure 5.14: Qualitative evaluation using Cohen's Kappa coefficient and years of experience. Blue stars mean two squares are in the same point.

Table 5.5: Mean and standard deviation of pixel proportion occupied by the vessels, optic disc and background on the evaluation database.

	Synthetic Images	**Real Images**
Vessel proportion	0.1431 ± 0.0306	0.1519 ± 0.0306
Optic Disc proportion	0.1776 ± 0.0339	0.2456 ± 0.0722
Background	0.6792 ± 0.0428	0.6025 ± 0.0795

It is possible to observe from Table 5.5 that the mean proportions between synthetic and real images are very similar. The small difference between the mean proportion of the synthetic and real optic discs depends on the normal variation of the optic disc size among real fundus images.

To support the quantitative evaluation and to analyse the similarity between synthetic and real images, we also obtained the averaged 2D-histogram of real and synthetic images. These 2D-histograms were constructed using the Red and Green channels normalized by the luminance (See Figure 5.15).

Moreover, we calculated the mean-squared error between the averaged 2D-histograms and the chromaticity diagram of each of the 200 images of the database. The obtained results are presented in Table 5.6.

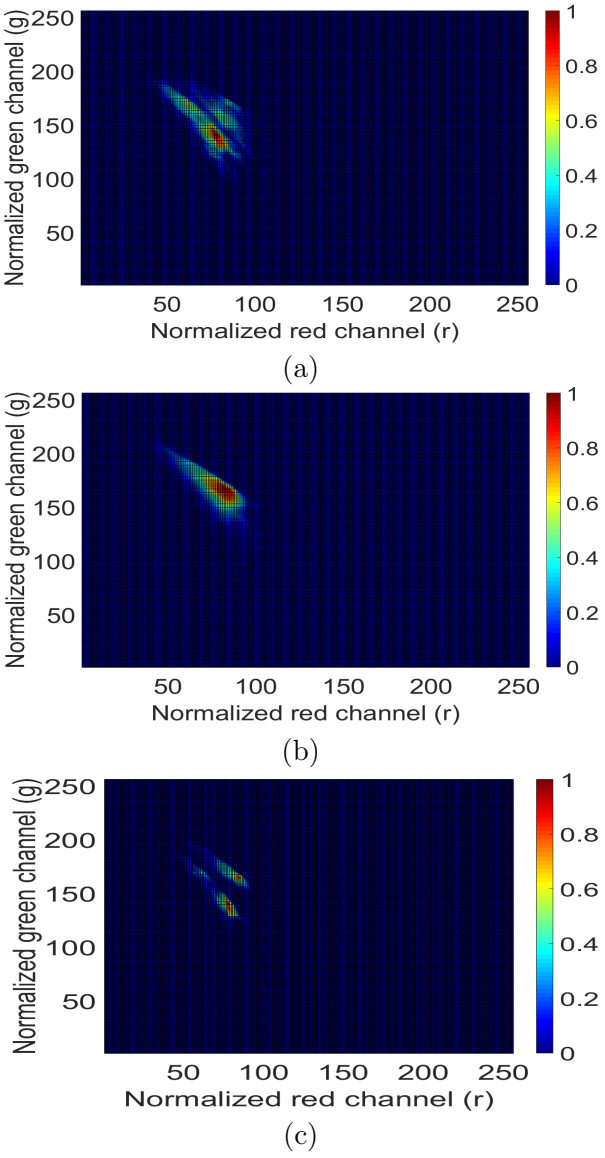


Figure 5.15: Averaged 2D-histograms of the synthetic and real images. (a) Averaged 2D-histogram of real images, (b) Averaged 2D-histogram of synthetic images and (c) Mean-squared error between synthetic and real 2D-histogram.

Table 5.6: Average and standard deviation of the mean-squared error between the averaged 2D-histograms and all images.

Averaged 2D-histogram	Real Images	Synthetic Images
Real	$0.0028 \pm 3.25 \times 10^{-4}$	$0.0036 \pm 5.43 \times 10^{-4}$
Synthetic	$\mathbf{0.0031 \pm 4.61 \times 10^{-4}}$	$0.0022 \pm 5.62 \times 10^{-4}$

The obtained results of this evaluation show that the mean-squared error between synthetic and real images is smaller than the resulting standard deviation among real images (3.25×10^{-4}).

An additional experiment to further analyse the latent space and cup size of the images was performed. We automatically measured the Cup/Disc ratio (CDR) to 1500 synthetic images using the method proposed by (Fu et al. 2018). From these samples, we obtained 743 glaucomatous images when setting the CDR threshold to 0.6 and 344 glaucomatous images when setting the CDR threshold to 0.7. This shows how this method could be used to assign glaucoma labels to synthetic images by using the CDR value.

Semi-supervised Learning and Retinal Image Synthesis using DCGAN

For this method, we used all the publicly available databases in the literature. A total of 86926 images belonging to the ORIGA-light (Z. Zhang et al. 2010), Drishti-GS1 (Sivaswamy et al. 2014), RIM-ONE (Medina-Mesa et al. 2015), sjchoi86-HRF (sjchoi86 2017), HRF (Köhler et al. 2013), ACRIMA, DRIVE (Staal et al. 2004), MESSIDOR (Decencière, X. Zhang, et al. 2014), DR KAGGLE (*Kaggle* 2015), STARE (Hoover, Kouznetsova, and Goldbaum 2000), e-ophtha (Decencière, Cazuguel, et al. 2013), ONHSD (Lowell et al. 2004), CHASEDB1 (Owen et al. 2011), DRIONS-DB (Carmona et al. 2008), and SASTRA (Narasimhan et al. 2012) databases were used to carry out the experiments.

All images were automatically cropped around the optic disc by using the method proposed in (P. Xu et al. 2017), with the exception of RIM-ONE where images are already cropped (See Figure 5.16). In order to fully covered the optic disc, we used a bounding box with ten more pixels around it. After cropping the images, they were revised by an expert in order to remove images with no optic disc or with the presence of huge artefacts. For that reason, we used fewer images of the DR KAGGLE (82447 instead of the 88702 images),

SASTRA (34 instead of 40 images), DRIONS-DB (105 instead of 110 images), ONHSD (89 instead of 99 images) and e-ophtha (431 instead of 463 images).

In this section, and as in previous sections in which deep learning techniques were employed, we used the open source Deep Learning library Keras (Chollet et al. 2015) and NVIDIA Titan Xp GPU.

In order to test the performance of the semi-supervised method, images with glaucoma and normal labels were divided into train and test using a typical division: 70% for training and 30% for test (287 glaucomatous and 420 normal images). In that way, the performance of the resulting discriminator/classifier can be evaluated using the test set.

As already mentioned in subsection 5.3, we trained an image synthesizer and a semi-supervised learning method. These models were trained on 86926 cropped retinal images from fourteen different databases. In the process of training these models, we tested a range of N-dimensional latent spaces from 32 to 100 latent variables. Each latent space was explored in order to check that the systems do not memorise the training database and, at the same time, it generates plausible retinal images. To accomplish this goal, we used spherical interpolation (White 2016) to evaluate intermediate latent representation points as it was done in (Costa, Galdran, M. I. Meyer, Abràmoff, et al. 2017). Regarding the image size, all the images were rescaled into 128×128 px because this size represents the nearest power of two to the averaged height and width of a retinal image cropped around the optic disc. We utilised a power of two image size to optimise speed and computational performance.

Retinal Image Synthesis

In this subsection, information about training stage (i.e. number of epochs) and a description of the obtained results (i.e. image resolution, image quality) using the DCGAN and SS-DCGAN will be presented.

Realistic images from the DCGAN model were obtained after 15 Epochs. However, for the SS-DCGAN model 150 Epochs were needed to obtain good enough synthetic images. To train the SS-DCGAN algorithm, we weighted the classes during the training stage because there is less number of glaucomatous images in the training set than normal and/or images without labels.

Experimental Evaluation

To evaluate this method, we created a database composed of 400 images: 100 synthetic images from the DCGAN, 100 synthetic images from the SS-DCGAN, 100 images from a state-of-the-art method (Costa's method (Costa, Galdran, M. I. Meyer, Niemeijer, et al. 2017)), and 100 real images (randomly selected from the training set with the exception of ORIGA-light database that will be used for qualitative evaluation) to perform a qualitative and quantitative evaluation. The synthetic images used for this evaluation were generated after training the DCGAN for 15 Epochs and the semi-supervised learning method for 150 Epochs.

As previously mentioned, synthetic images evaluation should be specific for each application. For that reason, we believe a good way to compare synthetic and real retinal images is by comparing the features extracted by a CNN trained to classify retinal images. Therefore, we fine-tuned the ResNet50 architecture (He et al. 2016) on the ORIGA-light database as a glaucoma classifier. Once this network was fine-tuned, we took 100 features for each image using a fully connected layer with 100 neurons on the top model, in which each neuron's output represents one feature.

After obtaining the 100 features for each image, we qualitatively show with t-SNE (Maaten and G. Hinton 2008) the feature differences between real images and synthetic images generated by the DCGAN, the semi-supervised learning method and the Costa's method (Costa, Galdran, M. I. Meyer, Niemeijer, et al. 2017).

It is important to highlight that Costa's method was originally presented to synthesize images with a wider field of view and fewer images. For that reason, we retrained their method on the 86926 cropped retinal images with a resolution of 128×128 px. Examples of images used for this comparison are shown in Figure 5.16.

For the quantitative evaluation, we analyzed two important features in cropped retinal images: the anatomic characteristics such as vessels and the optic disc and the colour properties of the images. In particular, optic disc masks were manually segmented by clinical experts and the vessel masks were automatically segmented using the method proposed in (Morales, Naranjo, Navea, et al. 2014).

The colour properties of the images were analyzed by the averaged chromaticity diagram (Colomer, Naranjo, and Angulo 2017). We additionally measured

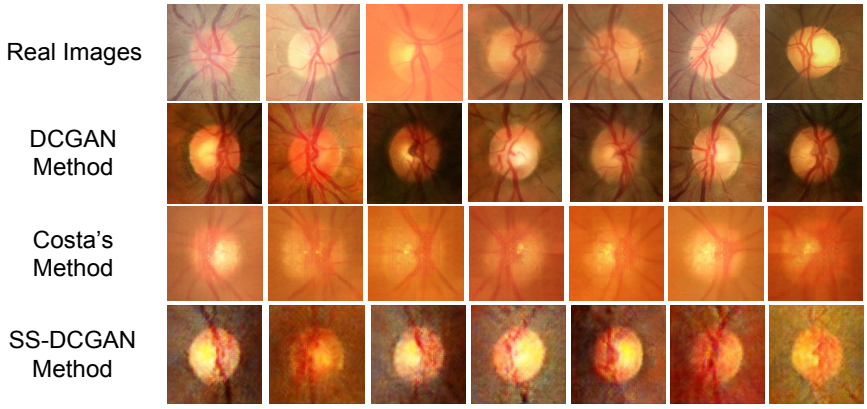


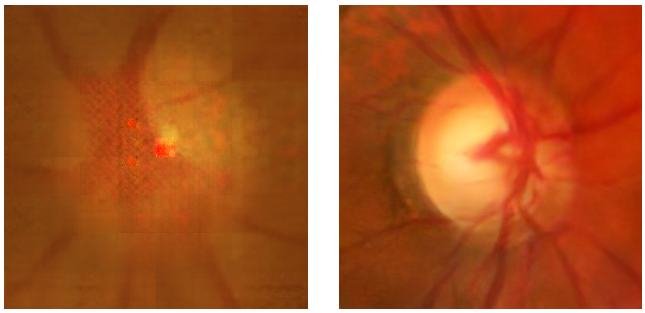
Figure 5.16: Examples of real images (first row), synthetic images generated by the DCGAN method (second row), synthetic images generated by Costa's method (third row) and synthetic images generated by the Semi-supervised DCGAN (fourth row).

the mean-squared error between the averaged histograms and the individual chromaticity diagram of each image. Results of this evaluation are presented in the next section.

In order to compare qualitatively and quantitatively the synthetic images obtained from our method, we first trained the method proposed by (Costa, Galdran, M. I. Meyer, Abràmoff, et al. 2017), following all the recommendations given in their paper, to generate cropped retinal images with the same size (128×128 px). See Figure 5.16.

As it can be seen from the Figure 5.16, synthetic images obtained from the DCGAN model are sharper, they present well-defined optic disc shapes, how the blood vessels clearly converge into the optic disc and right/left eye symmetry is evidenced in the resulting images. From this comparison, we found out that synthetic images from the Costa's algorithm have artefacts inside the optic disc as it is shown in Figure 5.17

These observations can be also qualitatively evaluated making use of the t-SNE plots. See Figure 5.18. From Figure 5.18 it is possible to see that features (extracted by the t-SNE algorithm) of the synthetic images generated by the DCGAN architecture are closer to the real images than the other methods and the features of images generated by Costa's method are closer to the real images than the SS-DCGAN method.



(a) (b)

Figure 5.17: Image sample generated by the Costa's (a) and the DCGAN (b) methods. Artifacts inside the optic disc are visible on the image generated by the Costa's method.

(a) (b) (c)

Figure 5.18: t-SNE plots of features associated to the different types of synthetic images. Yellow and blue dots indicate real and synthetic features respectively. Features of synthetic images using (a) DCGAN method, (b) Costa's method and (c) Semi-supervised DCGAN).

The quality of the images generated by the Semi-supervised DCGAN was expected to be low because this method is not only synthesizing images but also training a glaucoma classifier using labelled and not-labelled glaucomatous images. It was empirically demonstrated in (Dai et al. 2017) that a good semi-supervised learning method and a good generator cannot be obtained at the same time.

Given the results obtained from the qualitative evaluation, we decided to evaluate quantitatively the real images, the images generated by the Costa's method and the DCGAN. The semi-supervised DCGAN model will be further evaluated on glaucoma classification due to its low performance in synthesizing images.

As previously mentioned, the quantitative evaluation was carried out by measuring the average proportion of pixels belonging to the vessel and optic disc structures. Table 5.7 shows the obtained results.

Table 5.7: Mean and standard deviation of pixel proportion occupied by the vessels, optic disc, and background on the evaluation database.

	Real Images	**DCGAN method**	**Costa's method**
Vessel proportion	0.1519 ± 0.0306	0.1431 ± 0.0306	0.1026 ± 0.0195
Optic Disc proportion	0.2456 ± 0.0722	0.1776 ± 0.0339	0.1851 ± 0.0396
Background	0.6025 ± 0.0795	0.6792 ± 0.0428	0.7122 ± 0.0437

It is possible to observe from Table 5.7 that the mean proportions between synthetic and real images are very similar. The small difference between the mean proportion of the synthetic and real optic discs depends on the normal variation of the optic disc size among real fundus images. Moreover, the vessel proportion obtained from the Costa's images (0.1026) is significantly less than the averaged vessel proportion obtained from the real images (0.1519). This result could also be seen from Figure 5.17, in which vessels of the Costa's images are not as sharp as in the real or DCGAN images.

In order to evaluate the colour properties of the synthetic and real images, we also obtained the averaged 2D-histogram of real and synthetic images generated by our method and Costa's method. These 2D-histograms were constructed using the red and green channels normalized by the luminance (See Figure 5.19). It can be seen in Figure 5.19 that the shape of the histogram obtained from the DCGAN images (Figure 5.19(b)) is more similar to the shape of the histogram obtained from the real images (Figure 5.19(a)) than the shape of

the histogram obtained from the images generated by Costa's method (Figure 5.19(c)). This means that the colour properties of the images generated by the DCGAN method are closer to the properties of real retinal images.

Additionally, we calculated the mean-squared error between the averaged 2D-histograms and the chromaticity diagram of each of the 300 images of the database (100 Real images, 100 synthetic images using the DCGAN and 100 images using Costa's method). In other words, we compute for example the mean-squared error between the averaged 2D-histogram of real images and each image synthesized by the DCGAN and the Costa's method. The obtained results are presented in Table 5.8.

Table 5.8: Average and standard deviation of the mean-squared error between the averaged 2D-histograms of our method, Costa's method, and all images.

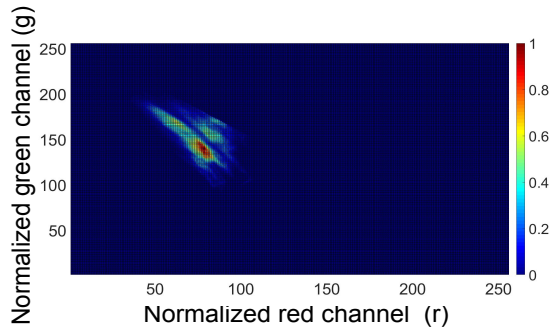
Averaged 2D-histogram	Real Images	DCGAN method	Costa's method
Real	0.0028 ± 0.000325	0.0036 ± 0.000543	0.0013 ± 0.000262
DCGAN method	0.0031 ± 0.000461	0.0022 ± 0.000562	0.0016 ± 0.000439
Costa's method	0.0031 ± 0.000126	0.0035 ± 0.000178	0.0010 ± 0.000163

In Table 5.8 it is possible to see that images generated by the DCGAN method are more heterogeneous among them than the images generated by the Costa's method. This is represented by the mean error distance between the averaged 2-D histogram and each image generated by the Costa's method (0.0013).

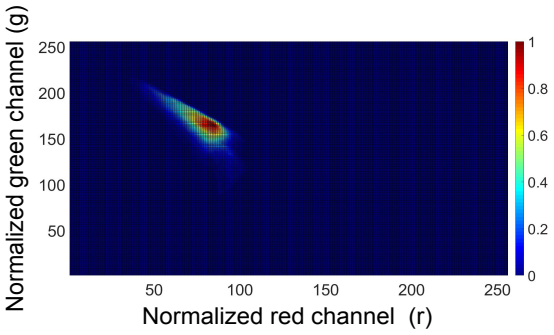
Glaucoma Diagnosis

In the qualitative evaluation we showed that although the SS-DCGAN system does not generate synthetic images better than the state-of-the-art methods, the resulting discriminator/classifier of the SS-DCGAN could be used as a glaucoma classifier. This classifier is the result of using glaucoma, normal and not-labelled images for training.

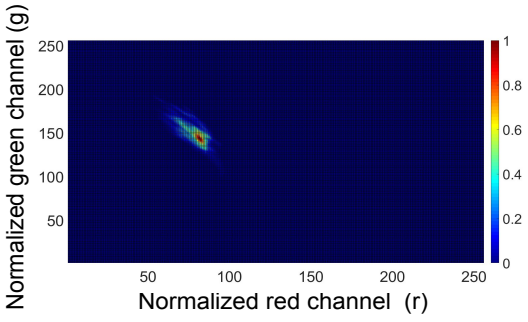
We computed the ROC curve, AUC, specificity, sensitivity and F-score to evaluate the performance of the proposed glaucoma classifier on the test set (287 glaucomatous and 420 normal images). Moreover, the obtained results were compared with other works in the literature such as the CNNs proposed by (X. Chen et al. 2015) and (Alghamdi et al. 2016). These networks were trained from scratch and tested on the same dataset used to evaluate our method. Additionally, we fine-tuned the ResNet50 architecture using the Im-



(a)



(b)



(c)

Figure 5.19: Averaged 2D-histograms of synthetic and real images. (a) Averaged 2D-histogram of real images, (b) Averaged 2D-histogram of synthetic images generated by the DCGAN model and (c) Averaged 2D-histogram of synthetic images produced by Costa's method. X-axis represents the normalized red channel and the Y-axis represents the normalized green channel.

ageNet weights. The obtained results from those models and our method are presented in Figure 5.20 and Table 5.9.

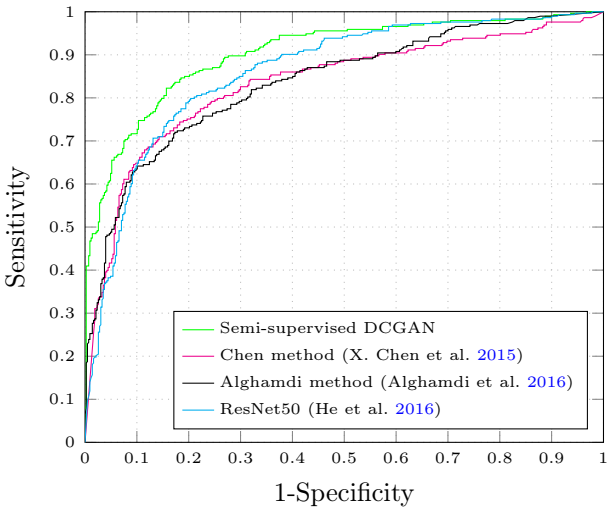


Figure 5.20: ROC curve for the glaucoma classifier trained by the Semi-supervised learning method.

Table 5.9: Comparison results of the proposed glaucoma classifier.

Model	Specificity	Sensitivity	AUC	F-score
Chen (X. Chen et al. 2015)	0.7440	0.8150	0.8330	0.8188
Alghamdi (Alghamdi et al. 2016)	0.6894	0.8384	0.8365	0.8174
ResNet50 (He et al. 2016)	0.8055	0.7775	0.8607	0.8137
SS-DCGAN	0.7986	0.8290	**0.9017**	0.8429

It is possible to see, from the Figure 5.20 and Table 5.9, that although the obtained results using the ResNet50 model, Chen's and Alghamdi's methods present a high AUC, the proposed glaucoma classifier outperforms them.

It is important to highlight that the architecture of the discriminator/classifier in the SS-DCGAN model is less complex than most of the works in the literature (4 layers). For instance, the CNN proposed by Chen is composed of 6 layers, the CNN proposed by Alghamdi is composed of 10 layers, and the ResNet50 architecture is composed of 50 layers. This improvement is given by the images without label and the synthetic images used to train the semi-supervised DCGAN. It was empirically demonstrated in (Dai et al. 2017) that

generative adversarial networks used as semi-supervised learning method boost the task performance because it uses the synthetic images generated while training the discriminator/classifier.

5.5 Discussion

In sections 5.1 and 5.2, we have proposed two methods for automatic glaucoma assessment using pretrained ImageNet CNN architectures. In the first method (Section 5.1), we used images cropped around the optic disc to fine-tune the VGG16, VGG19, InceptionV3, ResNet50, and Xception architectures. Using 10-fold cross-validation, we compared the performance between all these architectures. Although they have similar performance, the Xception architecture presents a better trade-off between the number of parameters and accuracy in glaucoma detection (AUC = 0.9605). Additionally, we carried out more experiments in order to analyse the performance of the Xception architecture on unseen images. From these experiments, we could see how this network outperformed the state-of-the-art CNNs for glaucoma assessment on a cross-testing setting.

The second glaucoma classifier (Section 5.2) is an average ensemble setting that uses four CNN architectures: VGG19, InceptionV3, ResNet50, and Xception. This configuration reduces the error rate by using the combination of output predictions of each CNN. Because this method was used to participate in the REFUGE challenge, we trained this algorithm on complete retinal images. Validation and test were performed on the validation and test set of the REFUGE database, obtaining an AUC of 0.94 and 0.88 for the validation and test set, respectively.

To summarise, we employed the fine-tuning technique on CNNs that were initially trained on natural (non-medical) images to classify glaucomatous images. Furthermore, we compared the performance of five CNNs by using cross-testing validation. Finally, we presented an average setting that was employed to compete in the REFUGE challenge.

In Subsections 5.3 and 5.3, two approaches for retinal image synthesis and semi-supervised learning have been introduced. In Subsection 5.3, a comparison between two image synthesizers, VAE and DCGAN, have been presented. Training these two models, we generated synthetic retinal images that were qualitatively and quantitatively evaluated. From this comparison, we saw that the DCGAN performed better than the VAE architecture in synthesizing retinal images. Additionally, we measured the CDR to 1500 synthetic images

which were used to put a label to each image. However, as CDR depends on how well the optic disc/cup is segmented, we decided to go further with retinal image synthesizers to produce glaucoma-labelled images by using a semi-supervised learning approach.

As our main goal was to implement a robust image synthesizer, capable of generating glaucoma-labelled images, we first collected and used, to the best of our knowledge, all the publicly available images (86926) to train again the DCGAN architecture as image synthesizer and a semi-supervised learning method based on the DCGAN (SS-DCGAN). Using the SS-DCGAN, we improved the state-of-the-art methods in retinal image synthesis and in automatic glaucoma classification (AUC = 0.9017) as shown in Section 5.3. The SS-DCGAN method is able to synthesise retinal image without previous vessel segmentation, which is an improvement to the methods proposed in the literature.

In summary, we showed how the DCGAN architecture outperformed the VAE architecture and the state-of-the-art methods in retinal image synthesis. Moreover, the proposed semi-supervised learning method (SS-DCGAN) is able to generate an unlimited number of glaucoma-labelled images that can be used to train a glaucoma classifier.

Chapter 6

Conclusions

In this thesis we have presented several contributions for automatic glaucoma assessment and optic disc and optic cup segmentation algorithms using retinal fundus images. As it was indicated in Chapter 2, the used of these types of images to assess glaucoma is highly valuable due to the low cost and non-invasive nature of this imaging modality.

In Chapter 4 we have presented a novel optic cup segmentation method based on the stochastic watershed transformation. We demonstrated that using the YIQ colour space and measuring the CDR, we obtained a sensitivity and a specificity comparable with state-of-the-art algorithms in the literature. This algorithm was validated on 723 colour fundus images from six different databases with high variability. However, because this method is highly dependent on the intensity to detect the cup, in Subsection 4.2.2 we presented how the U-Net performs better for optic disc and optic cup segmentation which has the ability to learn more discriminative features than only the intensity. Using only publicly available databases, we trained, validated and tested a two-stage segmentation method, which also allows us to participate in the REFUGE challenge hosted at the MICCAI2018 conference in conjunction with OMIA workshop.

Regarding automatic glaucoma classification task, we also proposed and improved the state-of-the-art results by using convolutional neural networks (CNNs)

pre-trained on the ImageNet database. In Chapter 5, we showed how five different CNNs could be used for automatic glaucoma assessment when small datasets are available. We fine-tuned these CNNs on 1707 images and augmented images. Additionally, we also developed a method using four CNNs on an average ensemble setting. This method was also used to participate in the same challenge previously mentioned in Subsection 4.2.2 (REFUGE challenge).

Given the fact glaucoma-labelled images are often scarce or costly to obtain and the performance of a glaucoma assessment system is affected by the number of labelled images used during the training stage, we presented two image synthesizers based on the VAE and DCGAN architecture (Subsection 5.3). These two models were trained on cropped retinal images from one private and five public databases (2357 retinal images). They showed an improvement in the state-of-the-art algorithms because the models we presented do not need the vessel masks to synthesise images. Ten clinical experts qualitatively evaluated the results of this method by randomly presenting 100 real images and 100 synthetic images. With the results of this evaluation, we demonstrated that most of the clinical experts were fooled when evaluating synthetic images. Furthermore, we measured the CDR and put a label according to its CDR value to 1500 synthetic images. However, as CDR depends on how well the optic disc/cup is segmented, we decided to go further with retinal image synthesizers.

Therefore, an algorithm able to generate and, at the same time, label images was proposed in Subsection 5.3. In this Subsection, we presented a semi-supervised learning method for automatic glaucoma assessment. This method used the DCGAN model and was trained on a small glaucoma-labelled and large unlabelled database. To train this model, we used 86926 retinal images, in which only 2357 images have glaucoma labels. Qualitative and quantitative evaluation were carried out on the obtained synthetic images, showing an improvement in quality when compared with the current works in the literature. Moreover, glaucoma labels to synthetic data were be assigned without using the CDR.

As future works, several lines of research could be carried out from the contributions made in this thesis. It was presented in this thesis that the use of generative adversarial networks, as semi-supervised learning methods, have great potential for training image synthesizers and glaucoma classifiers. For that reason, the use of more advanced generative adversarial networks such as image-to-image translation methods, for retinal image synthesis and semi-supervised learning can be considered as a future line of research.

Given the performance presented by the CNNs, another future line of research could be the implementation and analysis of deeper and new convolutional neural networks such as ResNet152 and InceptionResNetV2. A comparison between these networks trained with synthetic images against networks trained with real images could also be considered as a further line of research. In this way, we could explore all the potential that CNNs can give to the glaucoma classification task. Finally, we think a further line of research can be the combination of fundus images with clinical history and/or other image technology such as optical coherence tomography (OCT) to obtain more accurate screening results. As it can be used to measure the thickness of layers in the retina.

Merits

Journal articles

Orlando J I, Fu H, Barbossa Breda J, van Keer K, Bathula D R, **Diaz-Pinto, A** et al. REFUGE Challenge: A Unified Framework for Evaluating Automated Methods for Glaucoma Assessment from Fundus Photographs. Under review for the Medical Image Analysis journal.

Diaz-Pinto A, Colomer A, Naranjo V, Morales S, Xu Y, Frangi A F. Retinal Image Synthesis and Semi-supervised Learning for Glaucoma Assessment. March 2019. IEEE Transactions on Medical Imaging. doi: https://doi.org/10.1109/TMI.2019.2903434

Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi J M, Navea A. CNNs for Automatic Glaucoma Assessment using Fundus Images: An Extensive Validation. March 2019. BioMedical Engineering OnLine. doi: https://doi.org/10.1186/s12938-019-0649-y

Diaz-Pinto A, Morales S, Naranjo V, Navea A. Computer-aided Glaucoma Diagnosis using Stochastic Watershed Transformation on Single Fundus Images. August 2019. Journal of Medical Imaging and Health Informatics. doi: https://doi.org/10.1166/jmihi.2019.2721

International conferences

Diaz-Pinto A, Colomer A, Naranjo V, Morales S, Xu Y, Frangi A F. Retinal Image Synthesis for Glaucoma Assessment using DCGAN and VAE Models, 19th International Conference on Intelligent Data Engineering and Automated Learning. November 2018. pp 224-232. doi: https://doi.org/10.1007/978-3-030-03493-1_24

Vesal S, **Diaz-Pinto A**, Ravikumar N, Ellman S, Davari A, Maier A. Semi-Automatic Algorithm for Breast MRI Lesion Segmentation Using Marker-Controlled Watershed Transformation. Nuclear Science Symposium and Medical Imaging Conference. October 2017.

Diaz-Pinto A, Morales S, Naranjo V, Alcocer P, Lanzagorta A. Glaucoma Diagnosis by Means of Optic Cup Feature Analysis in Color Fundus Images. 24th European Signal Processing Conference (EUSIPCO). August 2016.

National conferences

Diaz-Pinto A, Morales S, Naranjo V, Alcocer P, Lanzagorta A. Diagnóstico Automático del Glaucoma a través de la Segmentación y Análisis de la Copa Óptica Usando Imágenes de Fondo de Ojo. XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica. pp 383-386. 2016

Bibliography

Abadi, M et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. https://www.tensorflow.org/. Accessed 2017-05-25 (cit. on p. 67).

Abbas, Q (2017). "Glaucoma-Deep: Detection of Glaucoma Eye Disease on Retinal Fundus Images using Deep Learning". In: *International Journal of Advanced Computer Science and Applications (IJACSA)* 8.6, pp. 41–45. DOI: 10.14569/IJACSA.2017.080606 (cit. on pp. 24, 33, 34).

Abràmoff, M D and M Niemeijer (2015). "Mass Screening of Diabetic Retinopathy Using Automated Methods". In: *Teleophthalmology in Preventive Medicine*. Ed. by Georg Michelson. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 41–50. ISBN: 978-3-662-44975-2. DOI: 10.1007/978-3-662-44975-2_4 (cit. on p. 18).

Alghamdi, H S et al. (2016). "Automatic Optic Disc Abnormality Detection in Fundus Images: A Deep Learning Approach". In: *OMIA3 (MICCAI 2016)*, pp. 17–24. DOI: 10.17077/omia.1042 (cit. on pp. 24, 90, 92).

Almazroa, A et al. (2015). "Optic Disc and Optic Cup Segmentation Methodologies for Glaucoma Image Detection: A Survey". In: *Hindawi Publishing Corporation. Journal of Ophthalmology*. 2015, p. 28. DOI: 10.1155/2015/180972 (cit. on p. 22).

Angulo, J and D Jeulin (2007). "Stochastic watershed segmentation". In: *Proceedings of the 8th International Symposium on Mathematical Morphology*. 8, pp. 265–276 (cit. on pp. 41, 42).

Aphex34 (2015). *Wikimedia Commons: Typical cnn*. https://upload.wikimedia.org/wikipedia/commons/6/63/Typical_cnn.png. Accessed: 2018-12-18 (cit. on p. 48).

Arevalo, J F, D Krivoy, and C F Fernandez (2009). "How Does Optical Coherence Tomography Work? Basic Principles". In: *Retinal Angiography and Optical Coherence Tomography*. Ed. by J. Fernando Arevalo. New York, NY: Springer New York, pp. 217–222. ISBN: 978-0-387-68987-6. DOI: 10.1007/978-0-387-68987-6_10 (cit. on p. 18).

Arlot, S and A Celisse (2010). "A survey of cross-validation procedures for model selection". In: *Statistics Surveys*, pp. 40–79. DOI: 10.1214/09-SS054 (cit. on p. 68).

Bar, Y et al. (2015). "Deep learning with non-medical training used for chest pathology identification". In: *Proc. SPIE 9414, Medical Imaging 2015: Computer-Aided Diagnosis*, pp. 94140V-1 –94140V-7. DOI: 10.1117/12.2083124 (cit. on p. 24).

Bengio, Y (2012). "Deep Learning of Representations for Unsupervised and Transfer Learning". In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. Vol. 27. Proceedings of Machine Learning Research. PMLR, pp. 17–36 (cit. on p. 66).

Beucher, S and F Meyer (1992). "The morphological approach to segmentation: The watershed transformation". In: *Mathematical Morphology in Image Processing*. Ed. by E.R. Dougherty. New York: Marcel Dekker. Chap. 12 (cit. on p. 40).

Bock, R, J Meier, G Michelson, et al. (2007). "Classifying Glaucoma with Image-Based Features from Fundus Photographs". In: *Pattern Recognition: 29th DAGM Symposium, Proceedings*. 29, pp. 355–364. DOI: 10.1007/978-3-540-74936-3_36 (cit. on p. 53).

Bock, R, J Meier, L G Nyúl, et al. (2010). "Glaucoma risk index: Automated glaucoma detection from color fundus images". In: *Medical Image Analysis.* 14, pp. 471–481. DOI: 10.1016/j.media.2009.12.006 (cit. on p. 23).

Bonaldi, L et al. (2016). "Automatic Generation of Synthetic Retinal Fundus Images: Vascular Network". In: *Procedia Computer Science* 90.Supplement C, pp. 54–60. DOI: 10.1016/j.procs.2016.07.010 (cit. on p. 25).

Bourne, R R A (2006). "Worldwide glaucoma through the looking glass". In: *British Journal of Ophthalmology.* 90, pp. 253–254. DOI: 10.1136/bjo.2005.083527 (cit. on p. 7).

Carmona, E J et al. (2008). "Identification of the optic nerve head with genetic algorithms". In: *Artificial Intelligence in Medicine* 43.3, pp. 243–259. ISSN: 0933-3657. DOI: 10.1016/j.artmed.2008.04.005 (cit. on pp. 29, 34, 55, 84).

Carneiro, G, J Nascimento, and A P Bradley (2015). "Unregistered Multiview Mammogram Analysis with Pre-trained Deep Learning Models". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III.* Ed. by Nassir Navab et al. Cham: Springer International Publishing, pp. 652–660. ISBN: 978-3-319-24574-4. DOI: 10.1007/978-3-319-24574-4_78 (cit. on p. 23).

Chakravarty, A and J Sivaswamy (2014). "Coupled Sparse Dictionary for Depth-Based Cup Segmentation from Single Color Fundus Image". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part I.* Ed. by Polina Golland et al. Cham: Springer International Publishing, pp. 747–754. ISBN: 978-3-319-10404-1. DOI: 10.1007/978-3-319-10404-1_93 (cit. on p. 54).

— (2016). "Glaucoma classification with a fusion of segmentation and image-based features". In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 689–692. DOI: 10.1109/ISBI.2016.7493360 (cit. on pp. 77, 78).

Chan, K et al. (2002). "Comparison of machine learning and traditional classifiers in glaucoma diagnosis". In: *IEEE Transactions on Biomedical En-*

gineering 49.9, pp. 963–974. ISSN: 0018-9294. DOI: 10.1109/TBME.2002.802012 (cit. on p. 21).

Chatfield, K et al. (2014). "Return of the Devil in the Details: Delving Deep into Convolutional Nets". In: *ArXiv e-prints* abs/1405.3531 (cit. on p. 75).

Chen, H et al. (2015). "Standard Plane Localization in Fetal Ultrasound via Domain Transferred Deep Neural Networks". In: *IEEE Journal of Biomedical and Health Informatics* 19.5, pp. 1627–1636. DOI: 10.1109/JBHI.2015.2425041 (cit. on pp. 23, 66).

Chen, X et al. (2015). "Glaucoma detection based on deep convolutional neural network". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 715–718. DOI: 10.1109/EMBC.2015.7318462 (cit. on pp. 24, 78, 90, 92).

Cheng, J et al. (2013). "Superpixel Classification Based Optic Disc and Optic Cup Segmentation for Glaucoma Screening". In: *IEEE Transactions on Medical Imaging*. 32, pp. 1019–1032. DOI: 10.1109/TMI.2013.2247770 (cit. on pp. 21, 53, 55).

Chintala, S et al. (2016). *How to Train a GAN? Tips and tricks to make GANs work*. https://github.com/soumith/ganhacks (cit. on p. 73).

Chollet, F et al. (2015). *Keras*. https://github.com/fchollet/keras. Accessed: 2017-07-21 (cit. on pp. 50, 57, 66, 85).

Chollet, F (2016). "Xception: Deep Learning with Depthwise Separable Convolutions". In: *ArXiv e-prints* abs/1610.02357 (cit. on pp. 65, 66, 75).

Choma, M A, K Hsu, and J A Izatt (2010). "Swept source optical coherence tomography using an all-fiber 1300-nm ring laser source." In: *Journal of Biomedical Optics* 10.4, p. 44009. DOI: 10.1117/1.1961474 (cit. on p. 19).

Chrástek, R et al. (2005). "Automated segmentation of the optic nerve head for diagnosis of glaucoma". In: *Medical Image Analysis* 9.4. Functional Imaging and Modeling of the Heart - FIMH03, pp. 297–314. ISSN: 1361-8415. DOI: 10.1016/j.media.2004.12.004 (cit. on p. 21).

Cobb, C (2010). *optic-disc.org, an online resource for ophthalmologists, physicians, medical students and optometrists. Glaucoma Evaluation Tutorial.* http://www.optic-disc.org/tutorials/glaucoma_evaluation_basics/page13.html. Accessed: 2016-01-27 (cit. on pp. 21, 42).

Colomer, A, V Naranjo, and J Angulo (2017). "Colour normalization of fundus images based on geometric transformations applied to their chromatic histogram". In: *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3135–3139. DOI: 10.1109/ICIP.2017.8296860 (cit. on p. 86).

Costa, P, A Galdran, M I Meyer, M D Abràmoff, et al. (2017). "Towards Adversarial Retinal Image Synthesis". In: *arXiv: 1701.08974*. arXiv: 1701.08974 [cs.CV] (cit. on pp. 26, 85, 87).

Costa, P, A Galdran, M I Meyer, M Niemeijer, et al. (2017). "End-to-end Adversarial Retinal Image Synthesis". In: *IEEE Transactions on Medical Imaging* PP.99, pp. 1–1. ISSN: 0278-0062. DOI: 10.1109/TMI.2017.2759102 (cit. on pp. 25, 86).

Criminisi, A, P Pérez, and K Toyama (2004). "Region filling and object removal by exemplar-based image inpainting". In: *IEEE Transactions on Image Processing.* 13, pp. 1200–1212. DOI: 10.1109/TIP.2004.833105 (cit. on p. 37).

Dai, Z et al. (2017). "Good Semi-supervised Learning That Requires a Bad GAN". In: *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., pp. 6510–6520 (cit. on pp. 74, 89, 92).

Decencière, E, G Cazuguel, et al. (2013). "TeleOphta: Machine learning and image processing methods for teleophthalmology". In: *IRBM* 34.2, pp. 196–203. DOI: 10.1016/j.irbm.2013.01.010 (cit. on pp. 30, 34, 84).

Decencière, E, X Zhang, et al. (2014). "Feedback on a publicly distributed database: the Messidor database". In: *Image Analysis & Stereology* 33.3, pp. 231–234. DOI: 10.5566/ias.1155 (cit. on pp. 31, 34, 84).

Deng, J et al. (2009). "Imagenet: A large-scale hierarchical image database". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255 (cit. on p. 66).

Diaz-Pinto, A et al. (2016). "Glaucoma Diagnosis by Means of Optic Cup Feature Analysis in Color Fundus Images". In: *24th European Signal Processing Conference (EUSIPCO)*. Vol. 24, pp. 2055–2059. DOI: 10.1109/EUSIPCO.2016.7760610 (cit. on p. 38).

Drexler, W and J G Fujimoto (2008). "State-of-the-art retinal optical coherence tomography". In: *Progress in Retinal and Eye Research* 27.1, pp. 45–88. ISSN: 1350-9462. DOI: 10.1016/j.preteyeres.2007.07.005 (cit. on pp. 18, 19).

Duda, R O, P E Hart, and D G Stork (1995). "Pattern Classification and Scene Analysis (2nd ed.)" In: John Wiley & Sons (cit. on p. 41).

Fiorini, S et al. (2014). "Automatic Generation of Synthetic Retinal Fundus Images". In: *Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference*. The Eurographics Association. ISBN: 978-3-905674-72-9. DOI: 10.2312/stag.20141238 (cit. on p. 25).

Fu, H et al. (2018). "Joint Optic Disc and Cup Segmentation Based on Multi-label Deep Network and Polar Transformation". In: *IEEE Transactions on Medical Imaging* PP.99, pp. 1–1. ISSN: 0278-0062. DOI: 10.1109/TMI.2018.2791488 (cit. on p. 84).

Gabriele, M L et al. (2011). "Optical coherence tomography: History, current status, and laboratory work." In: *Iovs* 52, pp. 2425–2436. DOI: 10.1167/iovs.10-6312 (cit. on p. 19).

Gagnon, L et al. (2001). "Procedure to detect anatomical structures in optical fundus images". In: vol. 4322, pp. 4322–8. DOI: 10.1117/12.430999 (cit. on p. 21).

Geisser, S (1975). "The Predictive Sample Reuse Method with Applications". In: *Journal of the American Statistical Association* 70.350, pp. 320–328 (cit. on p. 67).

Geitzenauer, W, C K Hitzenberger, and U M Schmidt-Erfurth (2011). "Retinal optical coherence tomography: past, present and future perspectives". In: *British Journal of Ophthalmology* 95.2, pp. 171–177. DOI: 10.1136/bjo.2010.182170 (cit. on p. 18).

Goodfellow, I, Y Bengio, and A Courville (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press (cit. on p. 45).

Goodfellow, I, J Pouget-Abadie, et al. (2014). "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc., pp. 2672–2680 (cit. on p. 74).

Hastie, T, R Tibshirani, and J Friedman (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Second Edition*. Springer-Verlag New York. DOI: 10.1007/978-0-387-84858-7 (cit. on p. 68).

He, K et al. (2016). "Deep Residual Learning for Image Recognition". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (cit. on pp. 64, 65, 75, 86, 92).

Hoover, A D, V Kouznetsova, and M Goldbaum (2000). "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response". In: *IEEE Transactions on Medical Imaging* 19.3, pp. 203–210. ISSN: 0278-0062. DOI: 10.1109/42.845178 (cit. on pp. 33, 34, 84).

Hubel, D H and T N Wiesel (1968). "Receptive fields and functional architecture of monkey striate cortex". In: *The Journal of Physiology* 195.1, pp. 215–243. ISSN: 1469-7793. DOI: 10.1113/jphysiol.1968.sp008455 (cit. on p. 44).

Huber, R, D C Adler, and J G Fujimoto (2006). "Buffered Fourier domain mode locking (FDML): Unidirectional swept laser sources for OCT imaging at 370,000 lines per second". In: *Optics Letters* 31.20, pp. 2975–2977. DOI: 10.1364/OL.31.002975 (cit. on p. 19).

Johnson, J and A Karpathy (2017). *Stanford CS class CS231n*. http://cs231n.github.io/. Accessed: 2017-07-21 (cit. on p. 47).

Joshi, G D, J Sivaswamy, and S R Krishnadas (2011). "Optic Disk and Cup Segmentation From Monocular Color Retinal Images for Glaucoma Assessment". In: *IEEE Transactions on Medical Imaging*. 30, pp. 1192–1205. DOI: 10.1109/TMI.2011.2106509 (cit. on pp. 21, 22, 53).

Kaggle (2015). https://www.kaggle.com/c/diabetic-retinopathy-detection/data. Accessed: 2018-02-05 (cit. on pp. 30, 34, 84).

Kingma, D P and M Welling (2013). "Auto-Encoding Variational Bayes". In: *arXiv: 1312.6114*. arXiv: 1312.6114 [stat.ML] (cit. on p. 69).

Kingma, Diederik P et al. (2014). "Semi-supervised Learning with Deep Generative Models". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z Ghahramani et al. Curran Associates, Inc., pp. 3581–3589 (cit. on p. 74).

Köhler, T et al. (2013). "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation." In: *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. Pp. 95–100. DOI: 10.1109/CBMS.2013.6627771 (cit. on pp. 31, 34, 50, 75, 79, 84).

Krizhevsky, A, I Sutskever, and G E Hinton (2012). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., pp. 1097–1105 (cit. on pp. 62, 64, 68).

Lalonde, M, M Beaulieu, and L Gagnon (2001). "Fast and robust optic disc detection using pyramidal decomposition and Hausdorff-based template matching". In: *IEEE Transactions on Medical Imaging* 20.11, pp. 1193–1200. ISSN: 0278-0062. DOI: 10.1109/42.963823 (cit. on p. 21).

LeCun, Y (1989). *Generalization and network design strategies*. Ed. by R. Pfeifer et al. http://yann.lecun.com/exdb/publis/pdf/lecun-89.pdf. Technical Report CRG-TR-89-4, University of Toronto (cit. on p. 44).

LeCun, Y, Y Bengio, and G Hinton (2015). "Deep learning". In: *Nature* 521.7553, pp. 436–444. DOI: 10.1038/nature14539 (cit. on p. 44).

Li, H and O Chutatape (2001). "Automatic location of optic disk in retinal images". In: *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*. Vol. 2, 837–840 vol.2. DOI: 10.1109/ICIP.2001.958624 (cit. on p. 21).

— (2004). "Automated feature extraction in color retinal images by a model based approach". In: *IEEE Transactions on Biomedical Engineering* 51.2, pp. 246–254. ISSN: 0018-9294. DOI: 10.1109/TBME.2003.820400 (cit. on p. 21).

Liu, B and M E Brezinski (2007). "Theoretical and practical considerations on detection performance of time domain, Fourier domain, and swept source optical coherence tomography." In: *Journal of Biomedical Optics* 12.August, p. 044007. DOI: 10.1117/1.2753410 (cit. on p. 19).

Lowell, J et al. (2004). "Optic nerve head segmentation". In: *IEEE Transactions on Medical Imaging* 23.2, pp. 256–264. ISSN: 0278-0062. DOI: 10.1109/TMI.2003.823261 (cit. on pp. 21, 31, 34, 84).

Lu, S (2011). "Accurate and Efficient Optic Disc Detection and Segmentation by a Circular Transformation". In: *IEEE Transactions on Medical Imaging* 30.12, pp. 2126–2133. ISSN: 0278-0062. DOI: 10.1109/TMI.2011.2164261 (cit. on p. 21).

Lu, S and J H Lim (2010). "Automatic macula detection from retinal images by a line operator". In: *2010 IEEE International Conference on Image Processing*, pp. 4073–4076. DOI: 10.1109/ICIP.2010.5649080 (cit. on p. 21).

Maaten, L van der and G Hinton (2008). "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9, pp. 2579–2605 (cit. on p. 86).

Maninis, Kevis-Kokitsi et al. (2016). "Deep Retinal Image Understanding". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham: Springer International Publishing, pp. 140–148. ISBN: 978-3-319-46723-8 (cit. on p. 21).

Mary, M C V S, E B Rajsingh, and G R Naik (2016). "Retinal Fundus Image Analysis for Diagnosis of Glaucoma: A Comprehensive Survey". In: *IEEE Access* 4, pp. 4327–4354. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2016.2596761 (cit. on p. 20).

Medina-Mesa, E et al. (2015). "Estimating the Amount of Hemoglobin in the Neuroretinal Rim Using Color Images and OCT." In: *Current Eye Re-*

search. 41.6, pp. 798–805. DOI: 10.3109/02713683.2015.1062112 (cit. on pp. 32, 34, 50, 56, 75, 79, 84).

Morales, S, V Naranjo, J Angulo, et al. (2013). "Automatic Detection of Optic Disc Based on PCA and Mathematical Morphology". In: *IEEE Transactions on Medical Imaging*. 32, pp. 786–796. DOI: 10.1109/TMI.2013.2238244 (cit. on pp. 21, 22, 35, 41).

Morales, S, V Naranjo, A Navea, et al. (2014). "Computer-Aided Diagnosis Software for Hypertensive Risk Determination Through Fundus Image Processing". In: *IEEE Journal of Biomedical and Health Informatics* 18.6, pp. 1757–1763. ISSN: 2168-2194. DOI: 10.1109/JBHI.2014.2337960 (cit. on pp. 37, 81, 86).

Mustroph, H (2014). "Dyes, General Survey". In: *Ullmanns Encyclopedia of Industrial Chemistry*. Wiley-VCH Verlag GmbH & Co. DOI: 10.1002/14356007.a09_073.pub2 (cit. on p. 38).

Narasimhan, K et al. (2012). "Glaucoma Detection From Fundus Image Using Opencv". In: *Research Journal of Applied Sciences, Engineering and Technology* 4.24, pp. 5459–5463 (cit. on pp. 33, 34, 84).

Nath, M K and S Dandapat (2012). "Techniques of Glaucoma Detection From Color Fundus Images: A Review". In: *International Journal of Image, Graphics and Signal Processing*. 4, pp. 44–51. DOI: 10.5815/ijigsp.2012.09.07 (cit. on pp. 21, 42, 44).

National Eye Institute, NIH (2012a). *Wikimedia Commons: Glaucomatous vision*. https://upload.wikimedia.org/wikipedia/commons/4/4a/Eye_disease_simulation%2C_glaucoma.jpg. Accessed: 2018-09-25 (cit. on p. 16).

— (2012b). *Wikimedia Commons: Normal vision*. https://upload.wikimedia.org/wikipedia/commons/c/cc/Eye_disease_simulation%2C_normal_vision.jpg. Accessed: 2018-09-25 (cit. on p. 16).

Ng, E Y K et al. (2014). "Image Analysis and Modeling in Ophthalmology". In: Boca Raton: CRC Press (cit. on p. 28).

Orlando, J I (2017). "Machine learning for ophthalmic screening and diagnostics from fundus images". PhD thesis. Universidad Nacional del Centro de la Provincia de Buenos Aires (cit. on p. 14).

Orlando, J I, H Fu, et al. (2019). "REFUGE Challenge: A Unified Framework for Evaluating Automated Methods for Glaucoma Assessment from Fundus Photographs". In: *IEEE Transactions on Medical Imaging* Under review (cit. on p. 9).

Orlando, J I, E Prokofyeva, et al. (2017). "Convolutional neural network transfer for automated glaucoma identification". In: *SPIE Proceedings*. Vol. 10160, pp. 10160–10. DOI: 10.1117/12.2255740 (cit. on pp. 25, 67, 78).

Owen, C G et al. (2011). "Retinal Arteriolar Tortuosity and Cardiovascular Risk Factors in a Multi-Ethnic Population Study of 10-Year-Old Children; the Child Heart and Health Study in England (CHASE)". In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 31.8, pp. 1933–1938. ISSN: 1079-5642. DOI: 10.1161/ATVBAHA.111.225219 (cit. on pp. 28, 34, 84).

Parfitt, C M, F S Mikelberg, and N V Swindale (1995). "The detection of glaucoma using an artificial neural network". In: *Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society*. Vol. 1, 847–848 vol.1. DOI: 10.1109/IEMBS.1995.575393 (cit. on p. 21).

Park, J-M and H-D Yae (2002). "Analysis of active feature selection in optic nerve data using labeled fuzzy C-means clustering". In: *2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No.02CH37291)*. Vol. 2, 1580–1585 vol.2. DOI: 10.1109/FUZZ.2002.1006742 (cit. on p. 21).

Patton, N et al. (2006). "Retinal image analysis: Concepts, applications and potential". In: *Progress in Retinal and Eye Research* 25.1, pp. 99–127. ISSN: 1350-9462. DOI: 10.1016/j.preteyeres.2005.07.001 (cit. on p. 17).

Pearson, K (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space". In: *Philosophical Magazine*. 2, pp. 559–572. DOI: 10.1080/14786440109462720 (cit. on p. 38).

Pircher, M, E Götzinger, et al. (2004). "Imaging of polarization properties of human retina in vivo with phase resolved transversal PS-OCT". In: *Opt. Express* 12.24, pp. 5940–5951 (cit. on p. 19).

Pircher, M, C K Hitzenberger, and U Schmidt-Erfurth (2011). "Polarization sensitive optical coherence tomography in the human eye". In: *Progress in Retinal and Eye Research* 30.6, pp. 431–451 (cit. on pp. 18, 19).

Pizer, S M, E P Amburn, et al. (1987). "Adaptive histogram equalization and its variations". In: *Computer Vision, Graphics, and Image Processing* 39.3, pp. 355–368. ISSN: 0734-189X. DOI: 10.1016/S0734-189X(87)80186-X (cit. on p. 48).

Pizer, S M, R E Johnston, et al. (1990). "Contrast-limited adaptive histogram equalization: speed and effectiveness". In: *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, pp. 337–345. DOI: 10.1109/VBC.1990.109340 (cit. on p. 48).

Potsaid, B et al. (2010). "Ultrahigh speed 1050nm swept source / Fourier domain OCT retinal and anterior segment imaging at 100,000 to 400,000 axial scans per second". In: *Optics Express* 18.19, p. 20029. DOI: 10.1364/OE.18.020029 (cit. on p. 19).

Quigley, H A (2018). "Use of Animal Models and Techniques in Glaucoma Research: Introduction". In: *Glaucoma: Methods and Protocols*. Ed. by Tatjana C. Jakobs. New York, NY: Springer New York, pp. 1–10. ISBN: 978-1-4939-7407-8. DOI: 10.1007/978-1-4939-7407-8_1 (cit. on p. 16).

Radford, A, L Metz, and S Chintala (2015). "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks". In: *arXiv: 1511.06434*. arXiv: 1511.06434 [cs.LG] (cit. on pp. 70, 72).

Razavian, A S et al. (2014). "CNN Features off-the-shelf: an Astounding Baseline for Recognition". In: *arXiv e-prints* abs/1403.6382 (cit. on p. 24).

REFUGE, Retinal Fundus Glaucoma Challenge - (2018). *REFUGE challenge. 5th MICCAI Workshop on Ophthalmic Medical Image Analysis (OMIA)*. https://refuge.grand-challenge.org/Download/. Accessed: 2018-06-20 (cit. on pp. 32, 55, 79).

Al-Rfou, R et al. (2016). "Theano: A Python framework for fast computation of mathematical expressions". In: *arXiv e-prints* abs/1605.02688 (cit. on p. 67).

Rhcastilhos and Jmarchn (2007). *Schematic diagram of the human eye*. https://commons.wikimedia.org/wiki/File:Schematic_diagram_of_the_human_eye_en.svg. Accessed: 2018-09-24 (cit. on p. 14).

Roletschek, R (2010). *Wikimedia Commons: Fundus Camera*. https://upload.wikimedia.org/wikipedia/commons/7/73/2010-12-07-funduskamera-by-RalfR-02.jpg. Accessed: 2018-09-25 (cit. on p. 17).

Román Morán, R et al. (2014). "Color Analysis in Retinography: Glaucoma Image Detection". In: *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*. Vol. 41. IFMBE Proceedings. Springer International Publishing, pp. 325–329. DOI: 10.1007/978-3-319-00846-2_81 (cit. on pp. 28, 34, 50).

Ronneberger, O, P Fischer, and T Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et al. Cham: Springer International Publishing, pp. 234–241. ISBN: 978-3-319-24574-4 (cit. on pp. 48, 58).

Russakovsky, O et al. (2015). "ImageNet Large Scale Visual Recognition Challenge". In: *International Journal of Computer Vision (IJCV)* 115.3, pp. 211–252. DOI: 10.1007/s11263-015-0816-y (cit. on p. 23).

S Sedai and P Roy and D Mahapatra and R Garnavi (2016). "Segmentation of Optic Disc and Optic Cup in Retinal Fundus Images Using Coupled Shape Regression". In: *Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop*. 1, pp. 1–8. DOI: 10.17077/omia.1040 (cit. on p. 54).

Salimans, T et al. (2016). "Improved Techniques for Training GANs". In: *ArXiv e-prints*. arXiv: 1606.03498 (cit. on p. 74).

Schuman, J S (2008). "Spectral domain optical coherence tomography for glaucoma (an AOS thesis)." In: *Transactions of the American Ophthalmological Society* 106, pp. 426–58 (cit. on p. 18).

Sevastopolsky, A (2017). "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network". In: *Pattern Recognition and Image Analysis* 27.3, pp. 618–624. DOI: 10.1134/S1054661817030269 (cit. on pp. 21, 48, 55, 58).

Sharma, A (2017). *Understanding Activation Functions in Deep Learning*. https://www.learnopencv.com/understanding-activation-functions-in-deep-learning/. Accessed: 2018-12-18 (cit. on pp. 46, 47).

Simonyan, K and A Zisserman (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *ArXiv e-prints* abs/1409.1556 (cit. on pp. 62, 75).

Sivaswamy, J et al. (2014). "Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation." In: *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. Pp. 53–56. DOI: 10.1109/ISBI.2014.6867807 (cit. on pp. 22, 29, 34, 55, 75, 79, 84).

sjchoi86 (2017). *sjchoi86-HRF Database*. https://github.com/sjchoi86/retina_dataset/tree/master/dataset. Accessed: 2017-07-02 (cit. on pp. 33, 75, 79, 84).

Srinivasan, V J et al. (2007). "High-speed, high-resolution optical coherence tomography retinal imaging with a frequency-swept laser at 850 nm". In: *Optics Letters* 32.4, p. 361. DOI: 10.1364/OL.32.000361 (cit. on p. 19).

Srivastava, N et al. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15, pp. 1929–1958 (cit. on p. 23).

Staal, J et al. (2004). "Ridge based vessel segmentation in color images of the retina". In: *IEEE Transactions on Medical Imaging* 23.4, pp. 501–509 (cit. on pp. 29, 34, 50, 84).

Stone, M (1974). "Cross-Validatory Choice and Assessment of Statistical Predictions". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 36.2, pp. 111–147 (cit. on p. 67).

Szegedy, C, W Liu, et al. (2015). "Going deeper with convolutions". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594 (cit. on pp. 62, 63, 75).

Szegedy, C, V Vanhoucke, et al. (2016). "Rethinking the Inception Architecture for Computer Vision". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826 (cit. on p. 64).

Tajbakhsh, N et al. (2016). "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" In: *IEEE Transactions on Medical Imaging* 35.5, pp. 1299–1312. DOI: 10.1109/TMI.2016.2535302 (cit. on pp. 23, 66).

Theis, L, A van den Oord, and M Bethge (2016). "A note on the evaluation of generative models". In: *International Conference on Learning Representations* (cit. on p. 81).

Weinreb, Robert N., Tin Aung, and Felipe A. Medeiros (2014). "The Pathophysiology and Treatment of Glaucoma: A Review". In: *JAMA* 311.18, pp. 1901–1911. ISSN: 0098-7484. DOI: 10.1001/jama.2014.3192 (cit. on p. 15).

Weinreb, Robert N. and Peng T Khaw (2004). "Primary open-angle glaucoma". In: *The Lancet* 363.9422, pp. 1711–1720. ISSN: 0098-7484. DOI: 10.1016/S0140-6736(04)16257-0 (cit. on p. 15).

White, T (2016). "Sampling Generative Networks". In: *arXiv: 1609.04468*. arXiv: 1609.04468 (cit. on pp. 80, 85).

Wojtkowski, M et al. (2005). "Three-dimensional Retinal Imaging with High-Speed Ultrahigh-Resolution Optical Coherence Tomography". In: *Ophthalmology* 112.10, pp. 1734–1746. ISSN: 0161-6420. DOI: 10.1016/j.ophtha.2005.05.023 (cit. on p. 18).

Wong, D W K et al. (2008). "Level-set based automatic cup-to-disc ratio determination using retinal fundus images in ARGALI". In: *30th Annual International IEEE EMBS Conference*. Vol. 30, pp. 2266–2269. DOI: 10.1109/IEMBS.2008.4649648 (cit. on p. 21).

World Health Organization (2004). *Bulletin of the World Health Organization, Volume 82, Number 11*. http://www.who.int/bulletin/volumes/82/11/en/infocus.pdf?ua=1. Accessed 2016-05-05 (cit. on p. 7).

Xu, P et al. (2017). "Optic Disc Detection via Deep Learning in Fundus Images". In: *Fetal, Infant and Ophthalmic Medical Image Analysis*. Cham: Springer International Publishing, pp. 134–141. ISBN: 978-3-319-67561-9 (cit. on p. 84).

Xu, Y et al. (2011). "Sliding window and regression based cup detection in digital fundus images for glaucoma diagnosis". In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. 6893, pp. 1–8. DOI: 10.1007/978-3-642-23626-6 (cit. on p. 53).

Yin, F et al. (2012). "Automated segmentation of optic disc and optic cup in fundus images for glaucoma diagnosis". In: *25th International Symposium on Computer-Based Medical Systems (CBMS)*. 25, pp. 1–6 (cit. on p. 53).

Zeiler, M D and R Fergus (2013). "Visualizing and Understanding Convolutional Networks". In: *ArXiv e-prints*. arXiv: 1311.2901 [cs.CV] (cit. on p. 68).

Zhang, J, B Rao, and Z Chen (2005). "Swept Source Based Fourier Domain Functional Optical Coherence Tomography". In: *Proceedings of the 27th Annual Conference. IEEE Engineering in Medicine and Biology*, pp. 7230–7233. DOI: 10.1109/IEMBS.2005.1616179 (cit. on p. 19).

Zhang, X et al. (2012). "Spatial normalization of eye fundus images". In: *ISBI 2012: 9th IEEE International Symposium on Biomedical Imaging*. Poster (cit. on p. 36).

Zhang, Y and Y Yang (2015). "Cross-validation for selecting a model selection procedure". In: *Journal of Econometrics* 187.1, pp. 95–112 (cit. on p. 68).

Zhang, Z et al. (2010). "ORIGA-light: An online retinal fundus image database for glaucoma analysis and research". In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 3065–3068. DOI: 10.1109/IEMBS.2010.5626137 (cit. on pp. 32, 34, 53, 79, 84).

Zhu, X (2005). *Semi-Supervised Learning Literature Survey*. Tech. rep. 1530. Computer Sciences, University of Wisconsin-Madison (cit. on p. 71).