uc3m | Universidad Carlos III de Madrid

Bachelor's degree in Data Science and Engineering
2021-2022

Bachelor Thesis

"DESIGN OF ALGORITHMS FOR THE CONSTRUCTION OF 3-DIMENSIONAL MODELS OF THE HUMAN CRYSTALLINE LENS"

Andrés Ruiz Calvo

Eduardo Martínez Enríquez
Fernando Díaz de María

Madrid, June 2022

ABSTRACT

Quantification of the anterior segment of the human eye (cornea, iris, and crystalline lens) is crucial for the diagnosis and treatment of common eye conditions such as presbyopia and cataracts. Quantification requires measurement of patient's eye and the construction of accurate 3-D models from these measurements. Measurements are usually obtained using a non-invasive high-resolution optical imaging technique known as Optical Coherence Tomography (OCT). From these measurements, one of the most relevant steps to obtain accurate 3-D models is the development of high-quality and reliable image segmentation algorithms.

In this bachelor's thesis we first describe state-of-the-art image segmentation algorithms of the anterior segment of the eye, which are based on traditional image processing techniques. Also, we present a discussion about their results and shortcomings.

Then, we present the main goal of this project: to improve the performance and address the shortcomings of the traditional image processing-based segmentation algorithms. To do so, an image segmentation model based on deep learning was created.

A comparison between both image segmentation models will be carried out. The results obtained show that model performance and reliability, key aspects for the quantification of anterior segment of the human eye, are superior using deep learning. The proposed deep learning model could have a strong positive impact on the quantification of the human eye, which in turn could help in the diagnosis and treatment of high-prevalence eye conditions.

Keywords: Image segmentation, Deep learning, Transfer learning, Optical Coherence Tomography (OCT), Anterior segment of the eye.

TABLE OF CONTENTS

TABLE OF FIGURES

1. MOTIVATION AND OBJECTIVES

1.1. Motivation

The anterior segment of the human eye is the front-most region of the eye and includes the cornea, the crystalline lens, and the iris.

The measurement of the geometrical shape (quantification) of the anterior segment's optical surfaces (cornea and crystalline lens) is crucial for the diagnosis and treatment of common (high prevalence) eye conditions such as presbyopia and cataracts.

Nevertheless, accurate quantification of the anterior segment is very challenging, requiring the measurement of the eye's patient and the construction of accurate 3-D models from these measurements. The construction of the 3-D models consists of four main processes:

1. Imaging the eye with high resolution and non-invasive techniques.
2. Correcting the distortions that appear in the images coming from the different imaging techniques.
3. Registering the information coming from different measurements in the same coordinate system.
4. Performing image segmentation of the surfaces of interest.

High-quality semantic segmentation models (a pixel-wise classification problem in which each of the pixels of the image must be classified according to the label it belongs to) are of utmost importance for obtaining accurate 3-D models of the human eye.

1.2. Objectives

The main goal of this project is the design and implementation of a semantic segmentation model based on deep learning that improves the performance and addresses the shortcomings of the traditional image processing-based segmentation approaches. Traditional approaches have been used for segmentation of eye's anterior segment images in the literature and implemented in current commercial clinical imaging systems.

In this work we firstly describe the segmentation approach based on traditional image processing techniques, providing:

1. A detailed explanation about how it was used to perform semantic segmentation over the regions of interest.
2. A discussion about the results obtained.
3. Shortcomings of the approach.

Then, we will turn to our own implementation of a semantic segmentation solution based on deep learning and compare the results with the traditional approach. To understand it, the following topics will be included:

1. An introduction to deep learning.
2. A detailed explanation about how it was used to perform semantic segmentation over the regions of interest.
3. A discussion about the results obtained.
4. A comparison between the traditional approach and deep learning, and if the shortcomings of the traditional approach are addressed.
5. Future lines of improvement.

Note that in this project we use anterior segment images taken with a non-invasive high-resolution optical imaging technique known as Optical Coherence Tomography (OCT).

2. INTRODUCTION

In this section we describe different concepts that will be used throughout this document. Specifically, we first describe the anatomy of the anterior segment of the human eye and the OCT technique used for image acquisition. Then, we introduce the concepts of deep learning and convolutional neural networks, which will be used in this project to approach the segmentation problem.

2.1. The anterior segment of the human eye

The regions of the anterior segment of the human eye, essential for understanding this project, are the following:

1. Cornea: External part of the eye that acts as the outermost lens, accounting for about two-thirds of the eye's total optical power. It is transparent and roughly spherical with a radius of curvature of 8 mm approximately [1]. When light strikes into the cornea, it bends or refracts the incoming light onto the lens. The cornea is composed of two main parts: anterior cornea and posterior cornea.

2. Iris: Colorful part of the eye that makes up the pupil and controls its size and, therefore, the amount of light that enters the eye (dilating or constricting the pupil to let in more or less amount of light, respectively).

3. Crystalline lens: Sophisticated optical structure that, along with the cornea, projects the images of the world onto the retina and changes its shape to focus near and far objects in what is known as accommodation, although this accommodation ability is progressively lost during life (presbyopia). Furthermore, the crystalline lens transparency is lost with age in what is known as a cataract, requiring its replacement by an artificial intraocular lens (IOL). It is composed of two main parts: anterior and posterior crystalline lens.

Figure 1 shows a schematic eye with the different anatomic parts, defining the anterior segment of the eye.

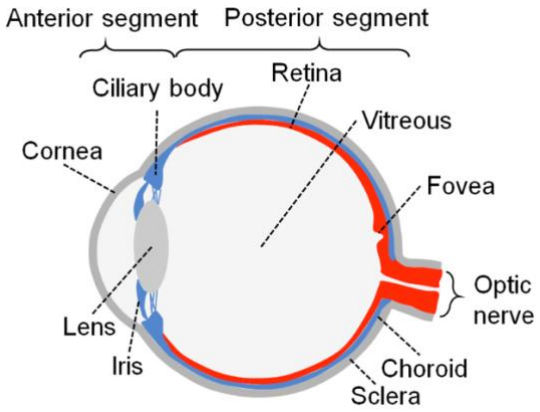


Fig. 1. Anatomy of the human eye. From [2].

2.2. Image acquisition: Optimal Coherence Tomography (OCT)

OCT imaging is based on the detection of reflected signals in the different internal layers of tissue. OCT is analogous to ultrasound imaging but using light instead of sound. For obtaining 3-D information, the position of the incident light is swept with a scanner.

The advantages of this technique compared to other image acquisition techniques are [3]:

- **High resolution:** OCT has a resolution of 5-10 μm. The ultrasound systems (US) can achieve resolutions of 150 μm, while computed tomography (CT) and Magnetic resonance imaging (MRI) acquire images with a resolution of 300 μm and 1000 μm, respectively. In this way, OCT is considered as a high-resolution imaging technique.
- **Non-invasive and contactless**: This property increases the safety of the patient, thus increasing its use and expanding its application in several medical areas.
- **High acquisition speed:** OCT does not have mechanical elements for scanning, allowing to obtain 3-D images with a speed of the order of tenths of a second.
- **Use of non-harmful radiation:** The OCT system operates in the visible region and infrared, being not harmful to the eye, unlike the CT system, which operates in the X-ray region. In addition, a very relevant characteristic of this technology is that it allows the real-time study of pathologies or samples of biological studies, granting considerable medical benefits.

Figure 2 shows a comparison between different imaging techniques as a function of their resolution (Y-axis) and their ability for tissue penetration (X-axis).

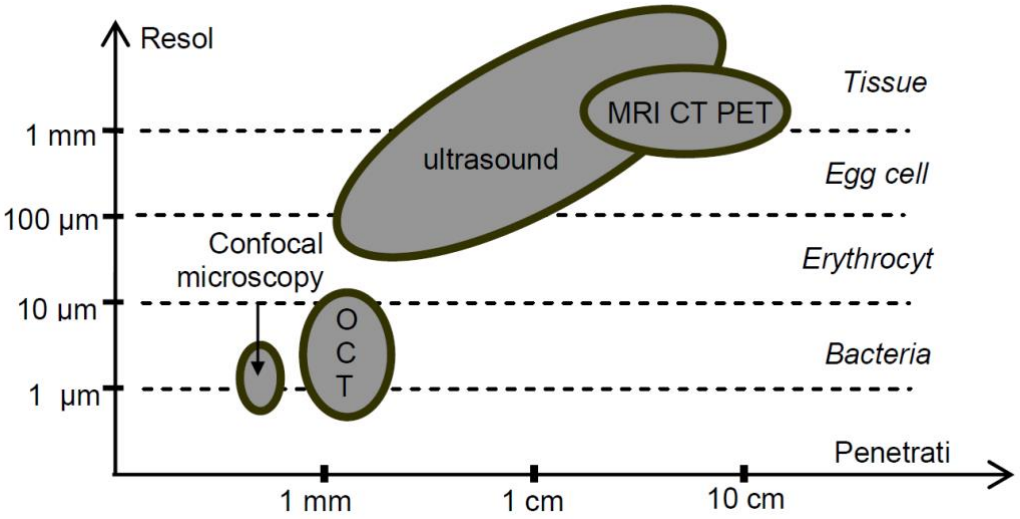


Fig. 2. Comparison between the different imaging techniques. From [3].

Thanks to the advantages offered by OCT compared to other techniques, its study and development has increased considerably in recent years, being very useful for various disciplines of medicine such as ophthalmology, oncology, and cardiology. OCT systems can be used both for the diagnosis of the disease and for its monitoring, as well as support to specialists in various surgeries. Figure 3 shows the very fast increase in the number of scientific publications related to OCT in recent years and the topic of these publications [4].

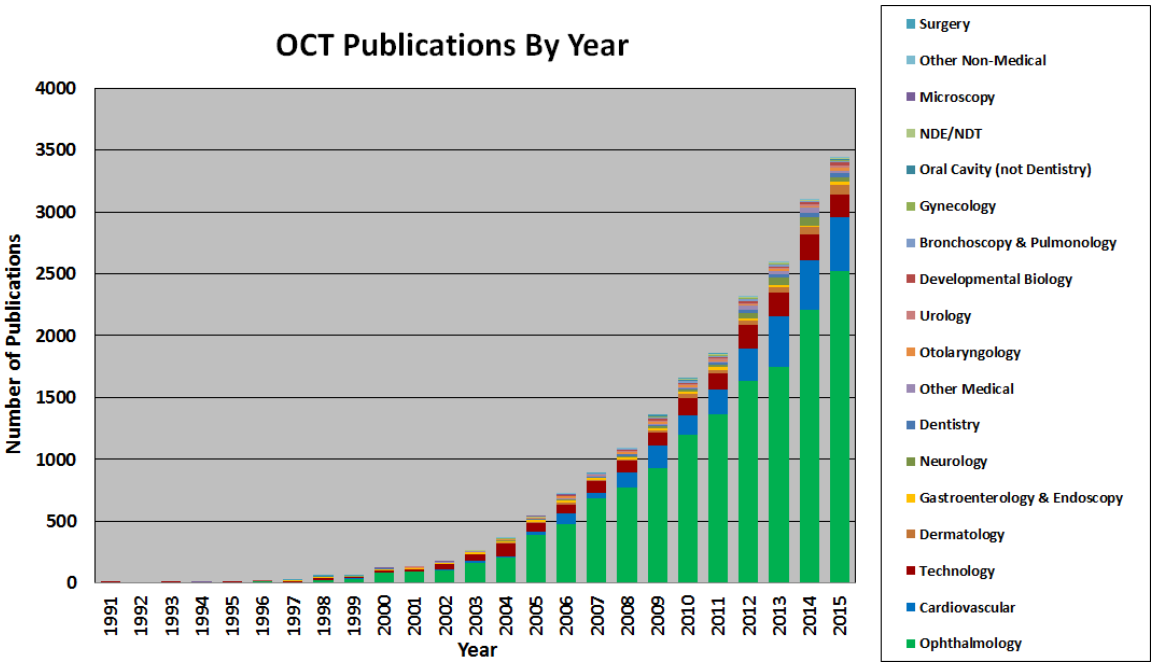


Fig. 3. Evolution on the number of publications on OCT. From [4].

2.3. Introduction to deep learning

An algorithm could be defined as a series of steps, normally defined by humans, aiming to solve a specific problem. Algorithms are fed with some inputs and produce outputs that represent the solution to the original problem. If the problem is straightforward, then coming up with a series of steps that solves the problem feels natural. An example might be sorting a list of natural numbers.

However, there are some problems for which coming up with an algorithm is more complex, or even out of reach. To illustrate this, let's consider the following example. You are given an image with a bunch of people, and you are asked to recognize and place a bounding box on each of them. Coming up with a series of steps does not feel natural to us because the task of recognizing people in an image is done subconsciously by our brain.

For this type of problems, it would be nice to have a computer program or machine capable of "*learning*" by itself how to obtain the desired output from the given inputs. In his book *Machine Learning* [5], Tom Mitchell, computer scientist and long-term researcher, provided a definition of machine learning: "A computer program is said to

learn with experience E with respect to some class of task T and performance measure P, if its performance at task T, as measured by P, improves with E."

Let's say we want to solve some task T: recognizing people on an image. The given *inputs* are fed into the *model*, and together with the *model parameters*, which define how the model operates, computes the *model predictions*. The *model predictions* and the *labels* (given *outputs*, which in this case would be the input image with the people correctly recognized) are compared according to the performance measure P (which could be how many persons were correctly classified by the model predictions). Then, the *model parameters* are updated so that the *model predictions* match the *labels*. This iterative process of updating the *model parameters* is called *model training* and it is what Tom Mitchell was referring to when he said learning with experience E.

This high-level explanation about machine learning can also be used to illustrate deep learning, which is a subfield of machine learning. However, some minor modifications need to be made. In deep learning, the model is defined as the combination of the *model parameters* and the *architecture*. The performance measure P is called *loss function*, and it is used to evaluate model performance and to update the *model parameters* during training so that the predictions match the *labels*. The concept of a *metric* is like the one of *loss function*, except for the fact that the *metric* is not used to update the *model parameters* during *model training*; just for evaluating how good our *model* is.

2.4. Introduction to Convolutional Neural Networks (CNNs)

Deep learning has a wide range of applications, including natural language processing, computer vision, biology, robotics, and recommendation systems, among others. In this section, we will provide a high-level explanation of the Convolutional Neural Network (CNNs), which are the building block of any computer vision technique based on deep learning.

CNNs were introduced by Yann LeCun et al., in their paper called "Gradient-Based Learning Applied to Document Recognition" [6], and they were used to successfully recognize handwritten characters.

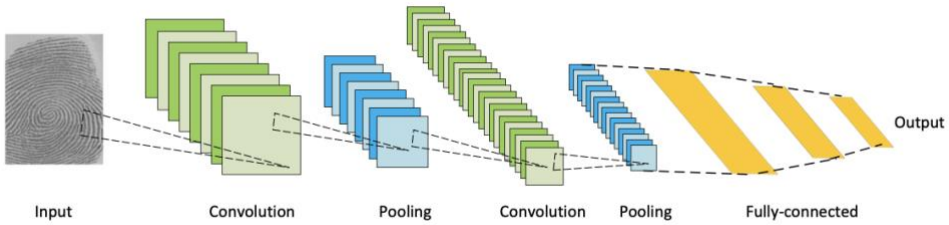


Fig. 4. Architecture of convolutional neural network. From [6].

As can be seen in Figure 4, the architecture of a convolutional neural network consists of a series of stacked up convolution layers (also known as kernels or feature maps), which are used to extract different features of the images. Each convolutional filter extracts a different feature; the feature extracted depends on the parameters of the filters, which are learnt during model training using a gradient-based optimization algorithm. The number of parameters used depend on the size of the input images, and the number and size of convolution layers, among other things. Since the number of parameters can be very high, pooling layers, whose goal is to shrink the input image and therefore reduce the number of parameters, are introduced.

The final fully connected layer is designed to output, for each input image, a vector of probabilities for each of the classes to be predicted. The structure of this layer varies a lot depending on the type of image segmentation problem.

2.5. Train, validation, and test sets

Neither a model can't be created without data, nor it can discover patterns and relationships that are not present in the data used to train the model. Therefore, the way in which we decide to split our dataset prior to the training phase can have a huge effect on the capabilities of our deep learning model.

It is important to keep in mind that one of the most important aspects of deep learning models is their ability to generalize well on unseen instances. If we were to use all our images to train the model and then evaluate model performance, we would not be capable

of assessing the quality of the model on data that it has not seen. To address this important issue, we divide our data into two sets: the training set and the validation set.

The training set is used for model training and the validation set is used to evaluate the performance of our trained model. This process is never done once; when creating a model, several choices need to be made (type of architecture, batch size, number of epochs, type of optimizer, learning rate, etc.) and we iterate until we find the combination of them that maximizes performance on the validation set.

However, if the validation set is not representative, our choice of hyperparameters will be biased, and our model will lose generalization power. To solve that, another data partition is required: the test set, which is not accessible until we have trained our model and decide to evaluate its final performance.

One of the biggest obstacles when it comes to training deep learning models is overfitting. Overfitting refers to the situation in which models have no predictive or scientific value because they perform extremely well on training data, but very bad on unseen data. Following the splitting strategy already mentioned helps us combat overfitting.

The size of the dataset, and all partitions being representative enough is also an important aspect to be considered. Regarding the size of the different sets, there is not a specific rule that states how much data the training, validation and test set needs to include. Different problems may require different approaches, and the decision often boils down to the designer of the machine learning model.

2.6. Transfer learning

Image segmentation models rely on convolutional neural networks (CNN), which use a stack of layers made up of convolutional filters that learn important features of the training images. Early layers are intended to learn general image features like edges and shapes, while final layers learn features specific to the task at hand.

Deep learning image segmentation models involve a huge number of parameters, often in the order of millions. When this is the case, training from scratch may not be the best

approach due to high computational and time requirements. In addition, there may exist other possible obstacles: model parameters may not be initialized properly, which may prevent the model from converging to a well-defined solution; or the dataset may be too small for the model to learn useful features and make accurate predictions.

A very common solution is to use a pretrained model. A pretrained model is a model whose parameters have been learned on another dataset. Since pretrained models are normally trained on huge and varied datasets, it is reasonable to assume that convolutional filters learned in the early layers can be reused.

The concept of pretrained models is related to that of transfer learning, which consists of using a pretrained model for a task different from what it was originally trained for.

The most common use of transfer learning in deep learning consists of the following steps:

1. Download the pretrained model and reuse its layers.
2. Change the head, which is the last layer of the model, to adapt for the specific task at hand.
3. Freeze the weights of the pretrained model so that they can't be updated during model training; otherwise, the information associated with the pretrained model will be lost.
4. Train the new head of the model. Once it has been trained, use the model to make predictions.

3. STATE OF THE ART

3.1. Anterior segment quantification

Numerous works have focused on the quantification of the anterior segment of the human eye and the study of the radius of curvature, thickness, topography, volume, and diameter of the cornea and the crystalline lens. This has motivated the development of devices capable of performing measurements with high resolution and speeds of acquisition. Thus, the geometry of the human eye has been extensively studied *ex vivo* and *in vivo* using different imaging techniques. For example, the crystalline lens has typically been evaluated using Purkinje images [7, 8], Scheimpflug camera [9, 10], ultrasound [11] or MRI [12, 13]. However, these imaging methods have disadvantages in terms of resolution and speed of acquisition in comparison with the OCT as mentioned in the previous sections. Therefore, in this project we used OCT imaging technique because it is a non-invasive technology, with high resolution and acquisition speed, and that is available in most of the ophthalmologic clinics in the world and implemented in several commercial systems.

3.2. 3-D models construction from OCT images

In 2009, the CSIC Institute of Optics in collaboration with the Nicolaus Copernicus University in Poland developed a spectral domain OCT for capturing images of the anterior segment of the eye [14]. Thanks to developed system and algorithms for the optical distortion correction [15] (due to the refraction of the rays in the different surfaces of the eye) and the fan distortion correction [16] (due to the OCT scanning architecture), the CSIC Institute of Optics was a pioneer in the accurate quantification of the different surfaces of the anterior segment of the eye using OCT, both in the cornea [16] and in the crystalline lens [17-19]. This has allowed the creation of personalized models of the anterior segment for a specific patient, very useful for the investigation of various pathologies such as myopia [20] or cataract surgery [21-23]. These models have been also applied for the discovery of new knowledge as, for example, the change in the topography [24] and in the full shape of the crystalline lens [25] with accommodation, useful for understanding presbyopia. The models have been validated with experimental measurements based on the wavefront measured in the retina [26, 27].

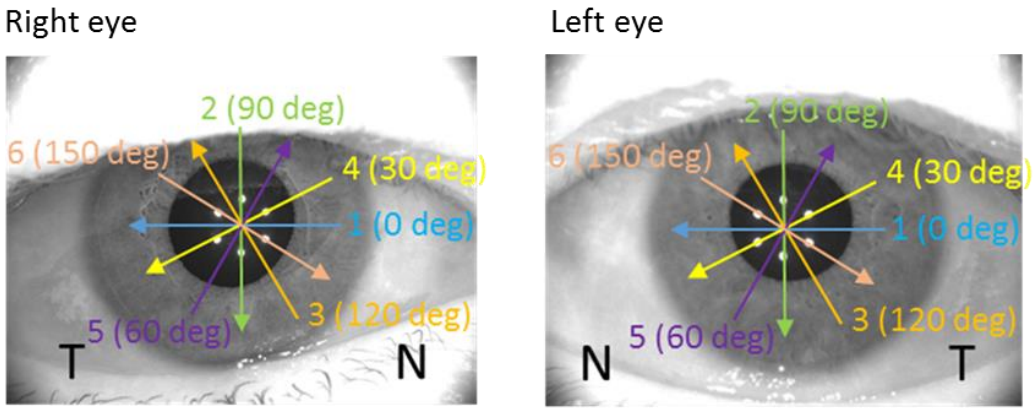
3.3. Anterior segment segmentation in OCT images

In the last years, deep learning has been extensively used to solve different segmentation problems [28, 29]. In the visual optics field, deep learning techniques have been lately used for segmentation of the different layers or the blood vessels of the retina [30-32]. Nevertheless, only a few works have approached the anterior segment images segmentation using deep learning techniques, and they have focused in the segmentation of the ciliary muscle [33]. Instead, classical image processing-based segmentation techniques have been extensively used for the segmentation of the different surfaces of the anterior segment of the eye [17, 20, 24, 34-36]. Nevertheless, there is room for improvement in the obtained results using classical techniques in terms of accuracy and robustness of the algorithms.

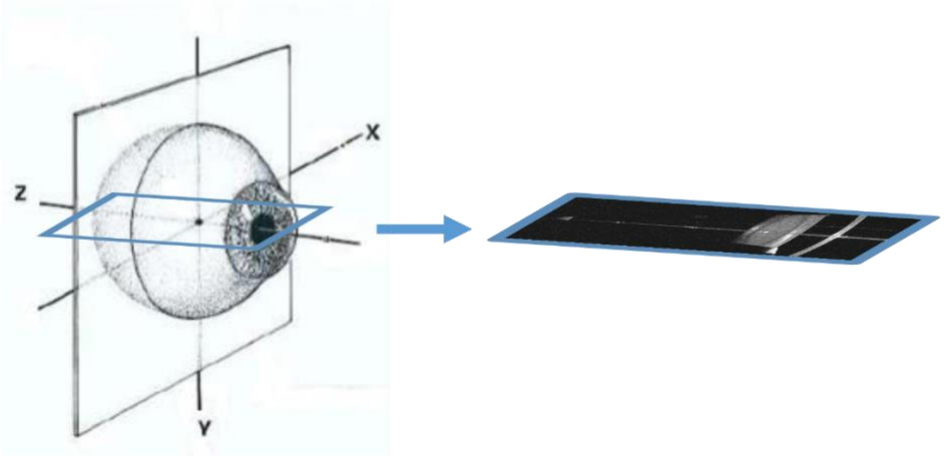
4. PROBLEM DEFINITION

4.1. Dataset

Our dataset consists of 588 images of size 903x703 pixels that were collected using a commercial OCT system developed by the company ZEISS (IOLMaster 700). The images belonged to 49 different subjects. Both eyes were measured for each subject, obtaining 6 different images for each eye. Each image has information about the structures (cornea, iris, etc.) within the eyes for a given scanning orientation corresponding to a given meridian, as shown in Figure 5 (a). Figure 5 (b) shows a representation of the 3-D information obtained for a given meridian (specifically the 0 degrees scanning).



(a)



(b)

Fig. 5. (a) Scanning pattern: 12 images are measured for each subject (6 images per eye). Each image corresponds to a different meridian (oriented scanning). N refers to Nasal (the nearest part to the nose) and

T refers to temporal (the farthest part to the nose). (b) Representation of the information obtained for the 0 degrees meridian. Using the 6 meridians per eye, 3-D information of the eye is obtained.

Figure 6 shows the typical OCT image obtained indicating the regions of interest for this project: the cornea (in green), the iris (in brown), the crystalline lens (in blue) and the retina (in red, not used in this project). Each of them was briefly explained in the introduction to the anatomy of the human eye.

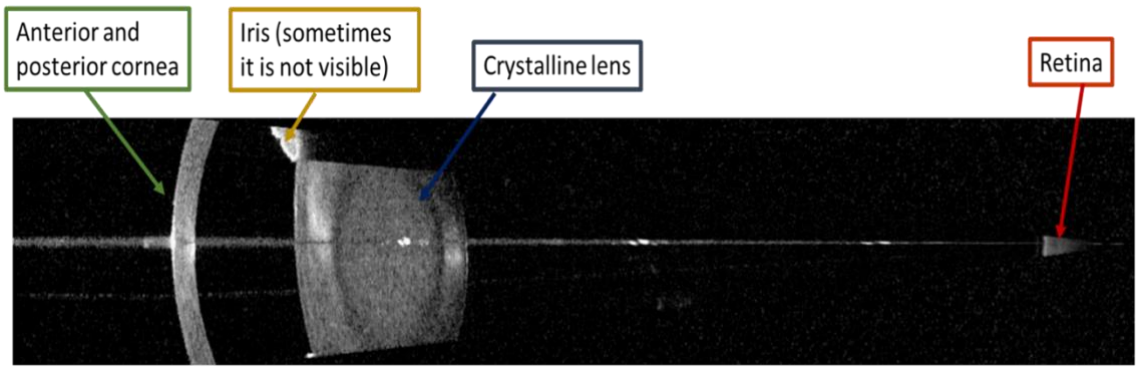


Fig. 6. Typical IOLMaster 700 OCT image including the anterior segment of the eye and the retina.

Note that throughout this document images will be presented rotated 90 degrees and cropped so that only the anterior segment is shown.

In our dataset, there are some images in which the iris is not visible. This is not a problem because we are not interested in segmenting it. Due to physical limitations of OCT, specifically that the light source used cannot penetrate the iris and track the whole crystalline lens, only its central region is visible.

Horizontal lines appearing in the middle of the images are noise; they occur whenever rays of light waves incident orthogonally to the cornea.

Even though it has been mentioned that there are several images of the same eye, none of them are the same, as there are small variations due to scanning angles and strong asymmetries of the human eye. Figure 7 shows three different scanning of the same eye and patient.

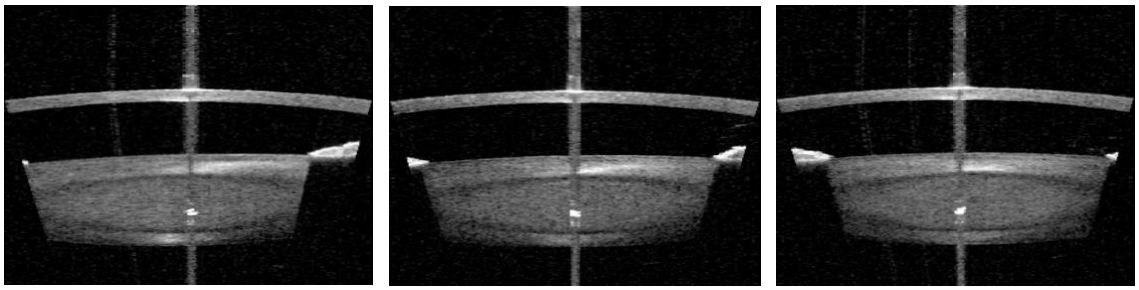
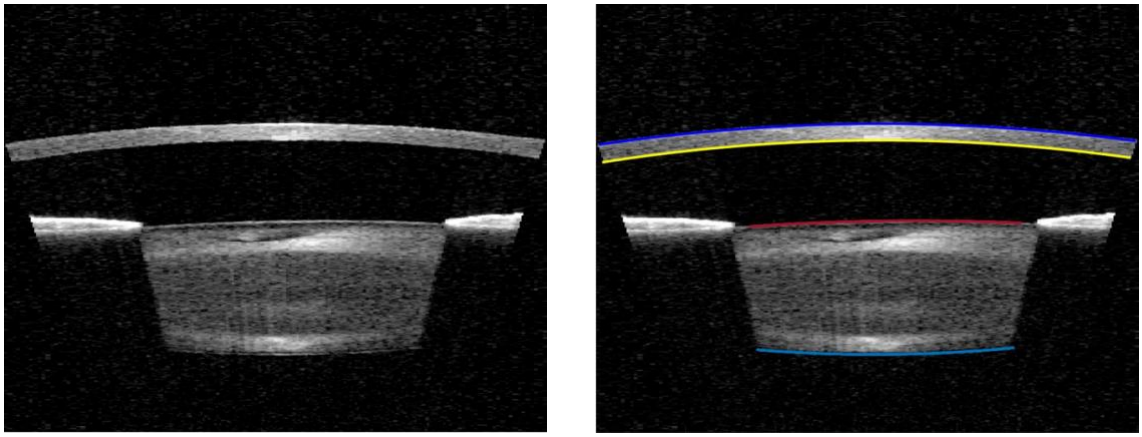


Fig. 7. Images of the same human eye corresponding to different meridians.

4.2. Problem definition

The problem that we face in this project consists in detecting the edges and label the anterior and posterior cornea and crystalline lens surfaces. This is a semantic segmentation problem because each of the pixels of the image will be classified into one of the following five classes: edge of the anterior cornea, edge of the posterior cornea, edge of the anterior crystalline lens, edge of the posterior crystalline lens, or background.

An example of a successful semantic segmentation model is shown in Figure 8. The model would take as input the image on the left (Figure 8 (a)) and produce as output the image on the right (Figure 8 (b)). In the output image, the edges of the four regions of interest are represented by using different colors and can be perfectly identified (in dark blue and yellow the anterior and posterior parts of the cornea, respectively; in red and light blue the anterior and posterior surfaces of the crystalline lens, respectively).



(a) (b)

Fig. 8. Example of input image (a). Example of a successful semantic segmentation model (b).

5. TRADITIONAL SEMANTIC SEGMENTATION APPROACH

Many traditional image segmentation algorithms have been developed in the literature using image processing tools such as thresholding [37] , edge detection [38], and k-means clustering [39], among many others.

In this section we describe a method based on traditional image processing tools that has been used to perform semantic segmentation of the different parts of the eye.

5.1. Segmentation algorithm

In general, the cornea and crystalline lens image processing-based segmentation algorithm includes the following steps:

1. Image cropping: The segmentation algorithm started cropping the image using a-priori knowledge about the typical eye anatomy; e.g., assuming that the cornea is always placed in the upper part of the image, or that the crystalline lens will always appear in a specific region of the image.
2. Edge detection: obtained using a Canny edge detector with a sigma of 3 and threshold of 0.11.
3. Vertical edge removal: Vertical edges that don't belong to the surfaces of interest were removed with a median filter of size 2x3. These edges appear mainly due to the strong reflection in the center of the image.
4. Thresholding: Uni-modal thresholding was performed to obtain a binary image with the masks where the information of interest should be located. The less noisy and blurry the image, the more effective thresholding will be. Because of that, before applying thresholding, a gaussian filter (used to smooth the image) with a sigma value of 5 was applied.
5. Finding masks/regions of interest: Morphological operations for region filling were applied. Then, some mask properties were calculated (i.e., centroid positions or areas) and a-priori knowledge about the measurements (i.e., relative position or relative area of crystalline lens and cornea) was used for labelling the regions that belonged to cornea and crystalline lens.

6. AND operation between obtained masks and edges was performed, so that only the edges in the regions of interest were saved.
7. Labeling: To assign a specific label to cornea and crystalline lens edges (i.e., anterior cornea, posterior cornea, anterior crystalline lens, or posterior crystalline lens) a-priori knowledge was used, assigning the label "anterior" to samples in the upper part of the cornea or crystalline lens masks and "posterior" to samples in the lower part.
8. Polynomial denoising: Finally, denoising was performed by recursively fitting a low order polynomial to each of the surfaces, removing samples that were far away from the best fitting polynomial of degree 2.

Figure 9 shows a flowchart of the segmentation process, illustrating the segmentation of the anterior part of the cornea.

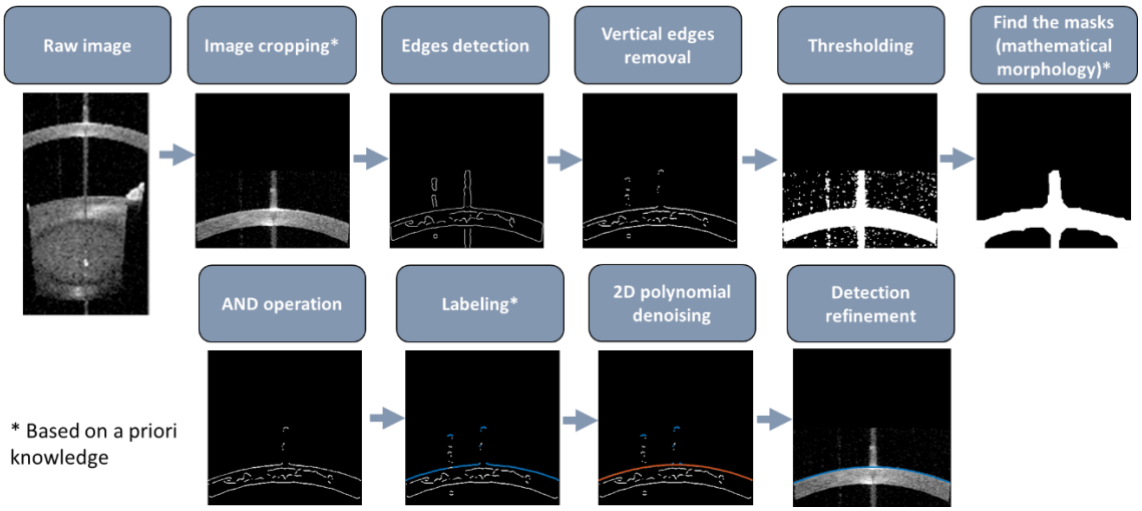
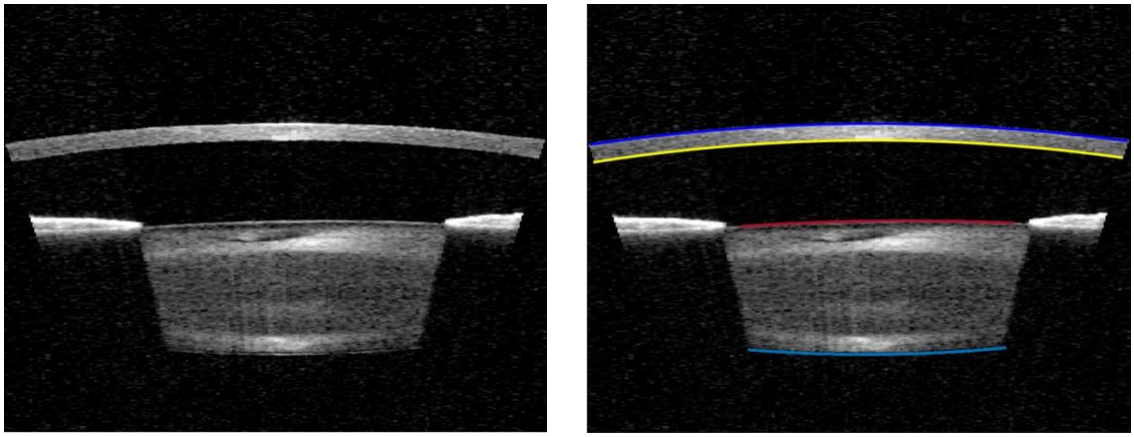


Fig. 9. Flowchart of the segmentation process. Anterior cornea example.

Other specific processes were performed to improve the segmentation, such as the use of a median denoising filtering for removing wrong detected-outlier edges (far away from the median value of the edges of a given surface) or removing posterior crystalline lens samples outside the pupil (i.e., those samples out of the range of anterior crystalline lens samples).

Figure 10 shows segmentation examples for the cornea (upper part of the figure) and crystalline lens (lower part of the figure). In Figure 10 (a), the input image is represented. In Figure 10 (b), the output image after applying the algorithm is visualized. As can be seen, both the anterior (blue) and posterior (yellow) cornea and the anterior (red) and posterior (blue) crystalline lens were successfully segmented. A more detailed discussion of the results obtained will be explained in later sections.



(a) (b)

Fig. 10. Input image (a). Example of image with the regions of interest segmented (b).

5.2. Discussion of the results

Although the described algorithm for semantic segmentation was capable of correctly identifying most of the anterior and posterior regions of the cornea and crystalline lens, there were some examples in which its performance was not satisfactory. In this section, we will be presenting and discussing some of these examples and try to infer some of the reasons for that wrong performance.

Overall, the results of the algorithm were satisfactory, as can be seen in the examples from Figure 11.

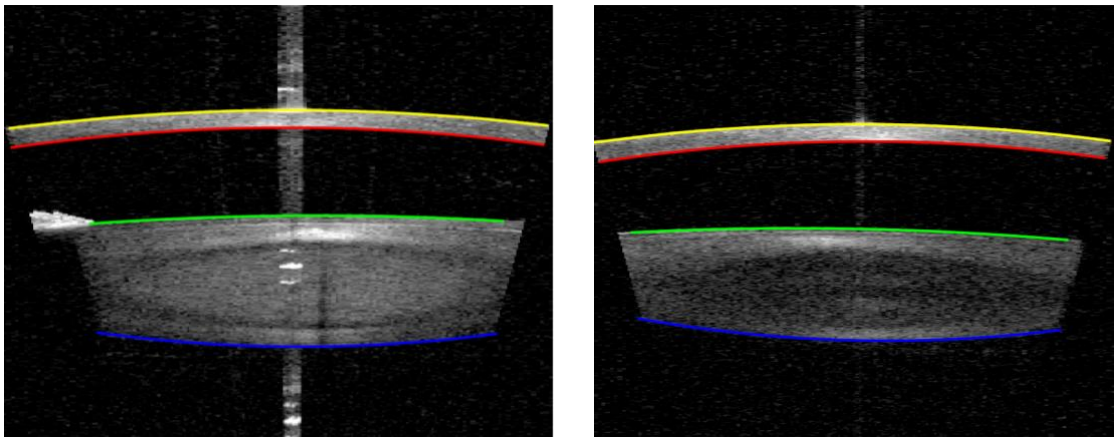
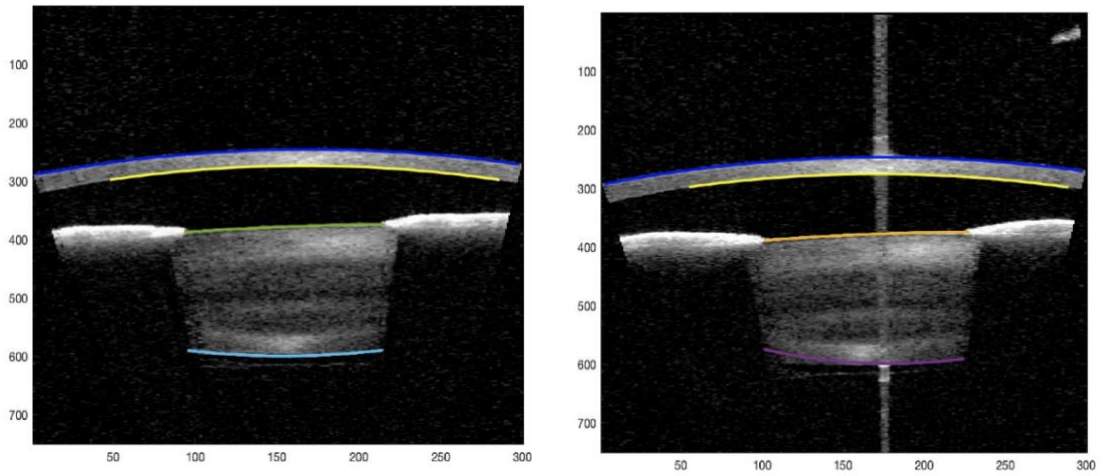


Fig. 11. Two images successfully segmented using the traditional image segmentation approach

However, there were approximately 30 images (5%) in which the segmentation results were wrong. For example, regarding the segmentation of the posterior cornea, there were 6 images in which the described algorithm could only partially capture it. Figure 12 shows two of them:



(a) (b)

Fig. 12. (a) Example where the posterior cornea was not segmented correctly. (b) Example where the posterior cornea and the crystalline lens were not properly segmented.

The reason for that is because the initial cropping assumes that the cornea is in the upper part of the image, specifically between the 150th and 300th rows. As can be seen from the vertical axis of the images, this assumption turns out to be wrong and that is why some parts of the regions of the cornea are not detected.

The need for “a-priori” knowledge is one of the main drawbacks of the image segmentation model being discussed. It is often the case that the future does not resemble the past and, if for some reason, the cornea of future images is not located between the 150th and 300th row, then the predictions will not be accurate. As previously discussed in the introduction, one of the desired features of the image segmentation model is reliability; and, in this situation, a-priori knowledge hampers reliability.

Regarding the quality of the predictions of the crystalline lens, the posterior crystalline lens is responsible for most of the errors. As can be seen in the examples of Figure 13, in some masks the predicted region for the posterior crystalline lens completely misses the point.

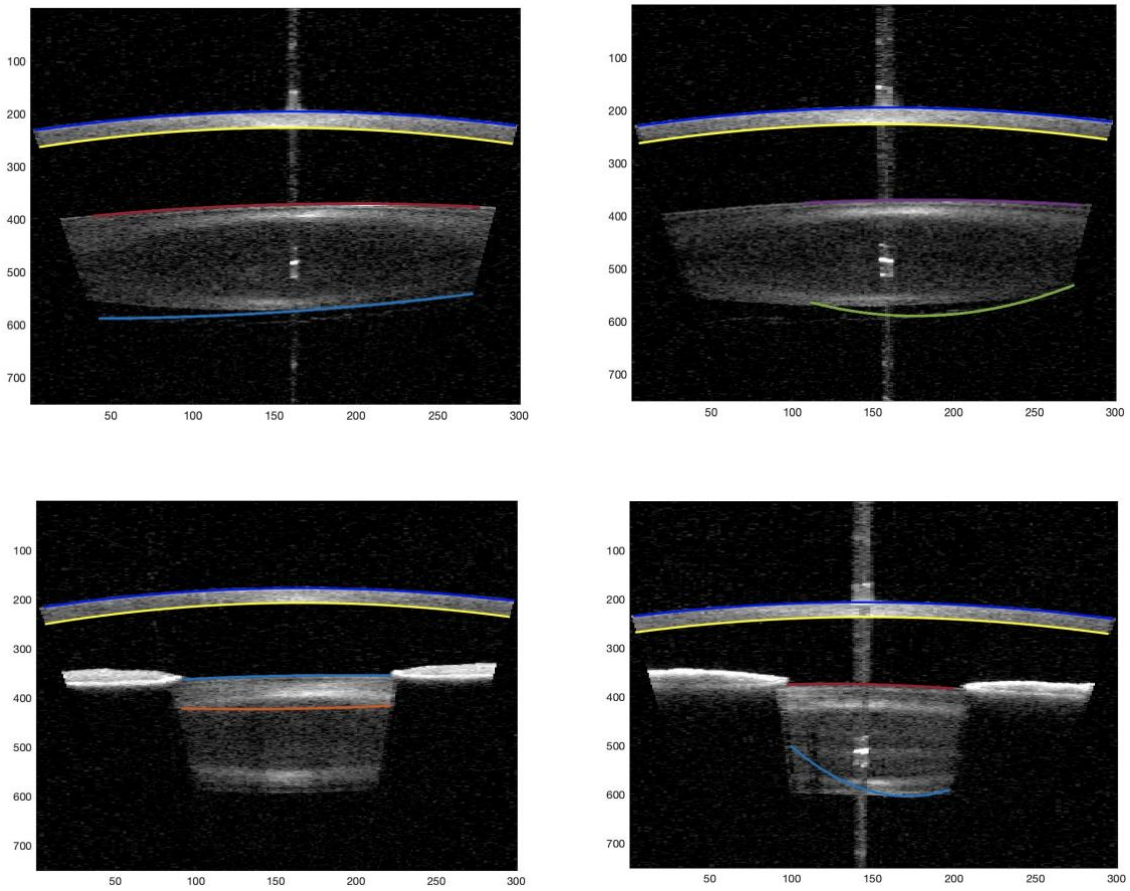


Fig. 13. Images with posterior crystalline lens incorrectly segmented.

It is worth mentioning that, out of the four regions to be segmented, the posterior crystalline lens is the most difficult to tackle. Across all images, the crystalline lens varies a lot more than cornea, so creating an algorithm with great generalization capabilities is more demanding.

In addition, both the anterior and posterior cornea are very similar, and detecting one of them could be useful for detecting the other; the anterior crystalline lens is in most of the cases surrounded by the iris, which could make its detection easier.

Finally, note that the Signal to Noise ratio in the posterior surface of the crystalline lens is lower because light must pass through the cornea, the anterior chamber (region between cornea and crystalline lens) and the crystalline lens to reach its posterior surface, losing part of its energy in the way.

5.3. Shortcomings

Though the algorithm already explained is capable of segmenting most of the regions of interest successfully, there are some shortcomings that need to be addressed.

We have illustrated that "incomplete" a-priori knowledge can have a negative impact on model predictions, thus resulting in less reliable segmentation models. It is important to appreciate that there are two more major factors affecting model performance:

1. The performance of most of the image processing techniques used depends on the parameters selected. Therefore, the choice of parameters is of utmost importance for obtaining good image segmentation models. Since one of the main objectives is to create a model capable of working with images not seen during the creation of the algorithm, it is important to make sure that the parameters selected are not biased towards the original image dataset. Otherwise, the generalization power of the model could be compromised.

2. Not only a well-defined procedure to select the right set of parameters is required, but also a means of testing the performance of our image segmentation model so that the algorithm is not biased towards our initial dataset. To do so, one common

approach is to obtain several partitions of our original dataset; one partition could be used to tune the model parameters, and another could be used to assess the model performance on unseen data.

However, neither of these two things were considered during the creation of the traditional semantic segmentation model.

In the following section, we will be presenting an image segmentation model based on deep learning that, among other things, aims to address these shortcomings.

6. SEMANTIC SEGMENTATION USING DEEP LEARNING

6.1. Data preprocessing and creation of the labels

The labelling process is central to the capability of generalization of our semantic segmentation deep learning model because it learns to make predictions that match the given labels.

Most of the images were manually labelled by a human supervisor. Nevertheless, in some cases the labels were obtained using the image processing-based segmentation algorithm described in the previous section. This was done only in the cases where the algorithm segmentation results were considered subjectively "perfect", to avoid the very time-consuming task of manual labelling. The consequences of this labelling process will be discussed in the following sections.

Labels as in Figure 14 (a) are not adequate for semantic segmentation problems because every pixel of the image needs to be classified according to the class it belongs to. Pixels that do not belong to any of the four regions of interest should be classified as background and represented using a different color (black), as shown in Figure 14 (b).

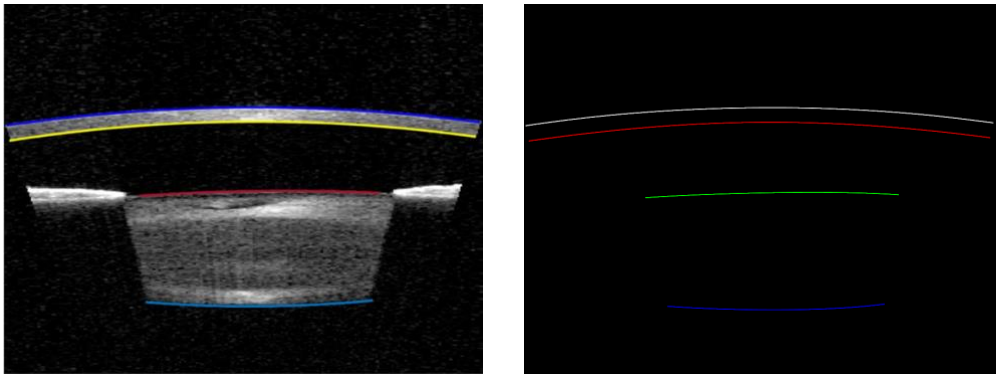


Fig. 14 (a) Initial labels. (b) New labels adopted for semantic segmentation: white for the anterior cornea; red for the posterior cornea; green for the anterior crystalline lens; blue for the posterior crystalline lens; black for the background.

As shown in Figure 14 (b), the number of pixels of the background class, which we are not interested in, is much greater than the number of pixels of the classes to be predicted. This undesirable situation in which there is an unequal class distribution in the training

set is known as class imbalance. Class imbalance is an important concern to this project, and it will be addressed in the following sections.

6.2. Selection of train, validation, and test set

Based on the assumption that the dataset containing OCT images of human eye of the same individual are more similar that images belonging to different ones, our image dataset was divided according to the individual the images belong to.

Since there are 49 different individuals, our training set included all images belonging to 35 individuals, the validation set included all images of 5 individuals, and the test set included the remaining 9 individuals. In terms of percentages, our image dataset is divided as follows: 70% of images to the training set, 10% to the validation set and 20% to the test set.

Another splitting strategy could have been to randomly assign images to the different sets. Nevertheless, we are not sure if OCT images of the human eye of the same individual are more similar than images belonging to different individuals. Therefore, we decided to play safe and split the data based on individuals. However, the exercise of testing and discussing how different splitting strategies might affect the overall performance of the model is very interesting and a possible future line of improvement.

6.3. Obstacles encountered during model training

Semantic segmentation is a not a trivial problem. Some of the challenges that need to be consider are the number of training instances needed to successfully train the model, the generalization capability of the trained models and the impact that incorrect predictions could have in the real world, among others [40].

One of the most common obstacles during model training in semantic segmentation is class imbalance. It refers to the situation in which the number of instances of the specific class to be predicted is very small compared to the remaining ones. It is a very common pitfall in fields such as medical and satellite imagery, where objects of interest represent only a small fraction of the overall image.

This situation is a great challenge because the objective during model training is to update the model parameters so that the loss function is minimized. If a specific class clearly dominates over the rest, the model will probably learn to label all pixels as the majority class because that may be one of the easiest forms of minimizing the loss function.

Class imbalance has been one of the greatest challenges of this project because, as have been explained in the labelling section, the regions of interest of the cornea and crystalline lens are very small compared to the background of the image. After training different models using these labels, all predictions obtained consisted of a black image, which is the majority class.

Without a doubt, this was an undesirable situation because our model was not learning what it was supposed to. In the context of image segmentation models, a wide range of techniques to deal with class imbalance have been proposed [41].

There exist numerous methods in the literature to alleviate the problem of imbalanced (imaging) datasets:

1. Synthetic augmentation, which consists of creating partial or completely artificial new images using Generative Adversarial Networks (GANs) [42].
2. Data augmentation [43], which consists of creating new images derived from the originals by applying transformations like rotation, translation, or cropping, to count a few.
3. Specific loss functions designed for heavily class imbalance, which work by giving more weight to mistakes on rare classes.

None of these methods were implemented in our solution because of the following reasons:

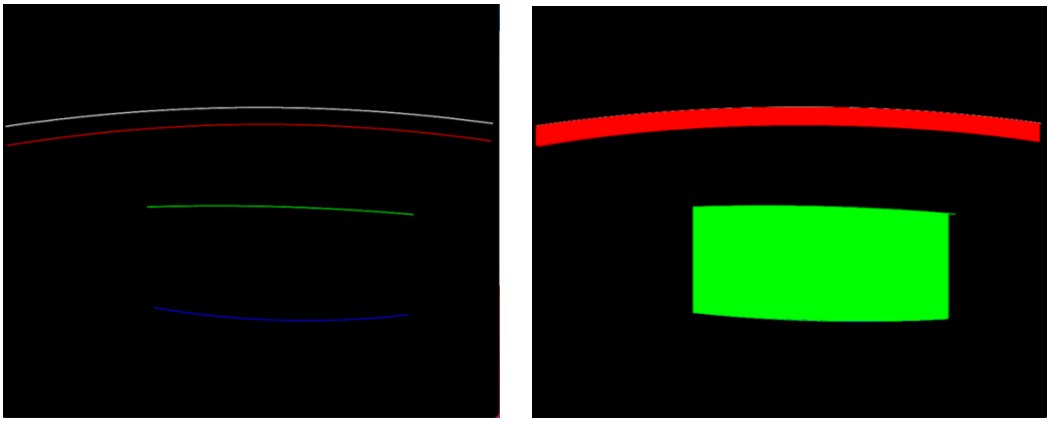
1. Synthetic augmentation was not considered because creating a new deep learning model to generate new images was out of the scope of this project.

2. Regarding data augmentation, we did not find any specific transformation that could alleviate the problem.

3. Some models were trained using a loss function specifically designed to work with imbalanced datasets, such as Focal Loss [44], but the issue of obtaining a black image as model predictions remained. We hypothesize that it may be due to the acute class imbalance. It is also important to mention that every dataset has its own particularities, and there is not a "one-size fits all" solution.

In the following paragraphs, we explain how we solved the problem of class imbalance.

Ideally, we would like to have more pixels of the classes we are interested in. The good news is that thanks to the topology of the original ground truth mask, we can, in fact, increase the number of pixels of interest, and therefore tackle the imbalance problem, without changing the actual regions of interest. Since we are only interested in the edges of the cornea and crystalline lens, the two ground truths shown in Figure 15 are equivalent for the model prediction purposes.



(a) (b)

Fig. 15. (a) Initial labels. (b) New labels designed to deal with class imbalance.

The pixels added in the new ground truth mask, represented in Figure 15 (b), are irrelevant to the original semantic segmentation task because information about the edges of interest is still contained in the new ground truth masks, but helpful because they help fight class imbalance.

The process for creating the new ground truth masks was straightforward. We used a straight line to connect the anterior and posterior cornea, and the same process to connect

the anterior and posterior crystalline lens. After that, we colored all pixels delimited by the new cornea and crystalline regions. Note that with the new masks we are now predicting three classes instead of five. Nevertheless, the former and the new masks are interchangeable, because one can go from one representation (mask) to the other without information loses.

Following the previous reasoning, we fed the deep learning model with the new masks, and then once the predictions were obtained, they were converted back to the original ground truth masks.

After training different models, this idea turned out to be right. The persistent problem of obtaining as output prediction a black image because of the class imbalance disappeared, and the deep learning model was able to identify the areas of interest.

6.4. Architecture

There exist a wide range of architectures in the field of image segmentation. The state-of-the-art semantic segmentation models are usually based on the Fully Convolutional Network (FCN) [45]. FCN are very similar to CNNs, but instead of having at the latter stages a fully connected layer, they only have convolutional layers. This makes possible to use FCN to perform semantic segmentation.

Some of the most common architectures for semantic segmentation are FCN with an encoder-decoder structure. An example might be U-Net [28] or Feature Pyramid Network (FPN) [46].

As has been already explained, the number of model parameters to be trained depends, among other things, on the dimensions of the images. With the original size images, we did not have enough computational resources to train a model using U-Net as architecture (the images had to be down sampled by a factor of 4 for U-Net to run, which could result in loss of information). In addition, U-Net requires that the width and height of input images are divisible by 32, which could affect the aspect ratio of the original images.

One advantage of FPN is that there are no requirements about the size of the original images. Another is that we did not have to substantially reduce the size of the input images to train models using FPN.

These reasons, together with the fact that the model predictions using FPN were satisfactory, justify the choice of FPN.

6.5. Pretraining and fine-tuning

As already explained in the previous section, FPN follows an encoder-decoder structure. Since we are interested in using a pretrained model, we need to select a pretrained encoder and decoder. ResNet50 [47] pretrained using ImageNet [48] was selected. ImageNet is an image database which contains millions of images. It is widely used in computer vision and deep learning research.

Since we are solving a semantic segmentation problem, the new head of the model needs to output a three-dimensional mask whose size is equal to the input images. For each pixel of the image, our model will output a three-dimensional vector containing the probability of that pixel belonging to each of the three classes to be predicted. The final predicted class will correspond to that with higher probability.

The whole model consisted of 26,116,291 parameters, but since we are freezing all layers except the head of the model, we were left with only 387 parameters to tune. Provided that the features of the pretrained model are useful for our problem, the drastic reduction of the number of model parameters to learn is an important advantage that considerably reduces the computational load and execution time during model training.

6.6.Metric

Metrics play a fundamental role when it comes to evaluating the performance of deep learning models. They are useful for quantifying what we mean by success and failure. When it comes to choosing the right type of metric for our problem, some aspects need to be taken into consideration.

First, it is often the case that not all the predictions errors count the same. For example, in medical diagnosis, the error of diagnosing a person with cancer when in fact it does not have cancer is more tolerable than diagnosing a person as healthy when it has cancer. In this example, the first error should count less than the second one.

Second, the nature of our problem and the potential particularities of the dataset. To illustrate this, let's consider the metric of accuracy, which is defined as the number of pixels that are correctly classified.

If there are 100 pixels and 90 of them are correctly classified, then accuracy is 90%. This metric is intuitive and straightforward. However, accuracy is not a suitable metric for class imbalance problems like ours. If we were only interested in detecting the minority class but because of skewness of the classes the model only predicts the majority class, then the accuracy would be very high, but, for our purposes, that model would be useless. This shows another important feature of metrics: what we understand by success and failure needs to be aligned with the definition of success and failure of the metric.

Two additional metrics for semantic segmentation are Intersection over Union (IoU), also known as Jaccard index, and F1 score.

IoU outputs values between 1 (ground truth and predicted mask perfectly overlap) and 0 (there is no overlap between the ground truth and predicted mask) and is computed as the area of overlap divided by the sum of the areas between the two masks:

$$IoU = J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Dice score also takes values between 0 and 1 but weighs more true positives.

$$Dice\ Score = \frac{2|A \cap B|}{|A|+|B|}. \quad (2)$$

However, we did not find any advantages in weighting true positive values, so Dice score was discarded.

We decided to evaluate our model using IoU because of the following reasons:

1. It is suitable for working with class imbalance, as opposed to accuracy.
2. It is aligned with our goals because it reflects what we mean by success and failure.
3. It is interpretable and easy to understand.

6.8. Hyperparameters

6.8.1. Batch size

The batch size is the number of instances used in every iteration to update the weights of the model. It can influence training time, because a powerful GPU can accelerate training, and model's performance, since the number of instances may affect the computation of the gradients during backpropagation, which in turn influences how the model parameters are updated.

We used a batch of size 8 for model training and of size 4 for model validation. Since some of the models considered were computationally expensive and we did not have access to a high-performance GPU and RAM, we were not able to work with a bigger batch size. Smaller batches were also tried, but there were no noticeable improvements.

6.8.2. Optimizer

Optimizers are methods that define how the model parameters are updated during training. The optimizer selected was Adam [49], which is currently recommended as the default optimizer to use for computer vision tasks.

6.8.3. Learning rate

The learning rate is one of the most important hyperparameters. It controls how much model parameters are updated at each iteration during model training. If the learning rate is too small, model parameters will be updated very slowly, and training may get stuck. In addition, a huge number of epochs may be required. On the other hand, if the learning

rate is too high, the difference between model parameters in subsequent iterations can be quite drastic, which can make the model diverge or find sub-optimal solutions.

The optimal learning rate depends on many factors such as the optimizer being used or batch size, among others. A good practice for finding the optimal learning rate is to consider a set of initial values using a logarithmic scale that range from very low to very high. For this specific project, we found that the best results were achieved using a learning rate between $3 \cdot 10^{-4}$ and $7 \cdot 10^{-4}$. Learning rates of the order of 10^{-3} or 10^{-5} were also tested but their performance was significantly worse (the model was not predicting all labels, or the model parameters were updated very slowly, among others).

6.8.4. Number of epochs

The number of epochs denotes the number of times that the training set has been used for model training. We trained for 5 epochs, but it just needed fewer epochs to yield acceptable results. A higher number of epochs was tried, but it did not translate into better performance.

6.8.5. Loss function

In the introduction, we defined learning as the iterative process of updating the model parameters so that the model predictions match the true labels. From a mathematical standpoint, making model predictions match the true labels is equivalent to minimizing a loss function. Loss functions also help evaluate the performance of deep learning models.

Different aspects should be taken into consideration when selecting a loss function: the type of problem we face, the particularities of our dataset (whether classes are imbalanced), and what we understand by good performance, among other things. Failing to address any of these concerns could prevent the model from succeeding. It is no wonder, then, that loss functions are essential for learning.

For an image segmentation problem, there exist a wide range of loss functions [50]. Some of the most used loss functions are: Jaccard loss, Focal loss, and Dice loss.

Using the selected batch size, optimizer, learning rate, and number of epochs already discussed, all previous loss functions were tried. Across all images of the validation set, the following results in terms of IoU were obtained:

- Jaccard loss: the mean IoU of the cornea across all epochs was 94,10%, while the mean IoU of the crystalline lens was 96,46%.
- Dice loss: the mean IoU of the cornea across all epochs was 93,53%, while the mean IoU of the crystalline lens was 96,66%.
- Focal loss: the mean IoU of the cornea across all epochs was 95,66%, while the mean IoU of the crystalline lens was 96,86%.

Even though the results obtained using different loss function were very similar, Focal loss, which was designed to work better with highly imbalanced datasets, was slightly better and, therefore, selected as loss function.

7. EXPERIMENTAL RESULTS

7.1. Model predictions

In this section, we analyze some of the predictions carried out by the deep learning model. It is important to note that the images in Figure 16 belong to the test set, which means that they were not seen during model training. Therefore, they illustrate the prediction capabilities of our semantic segmentation deep learning model.

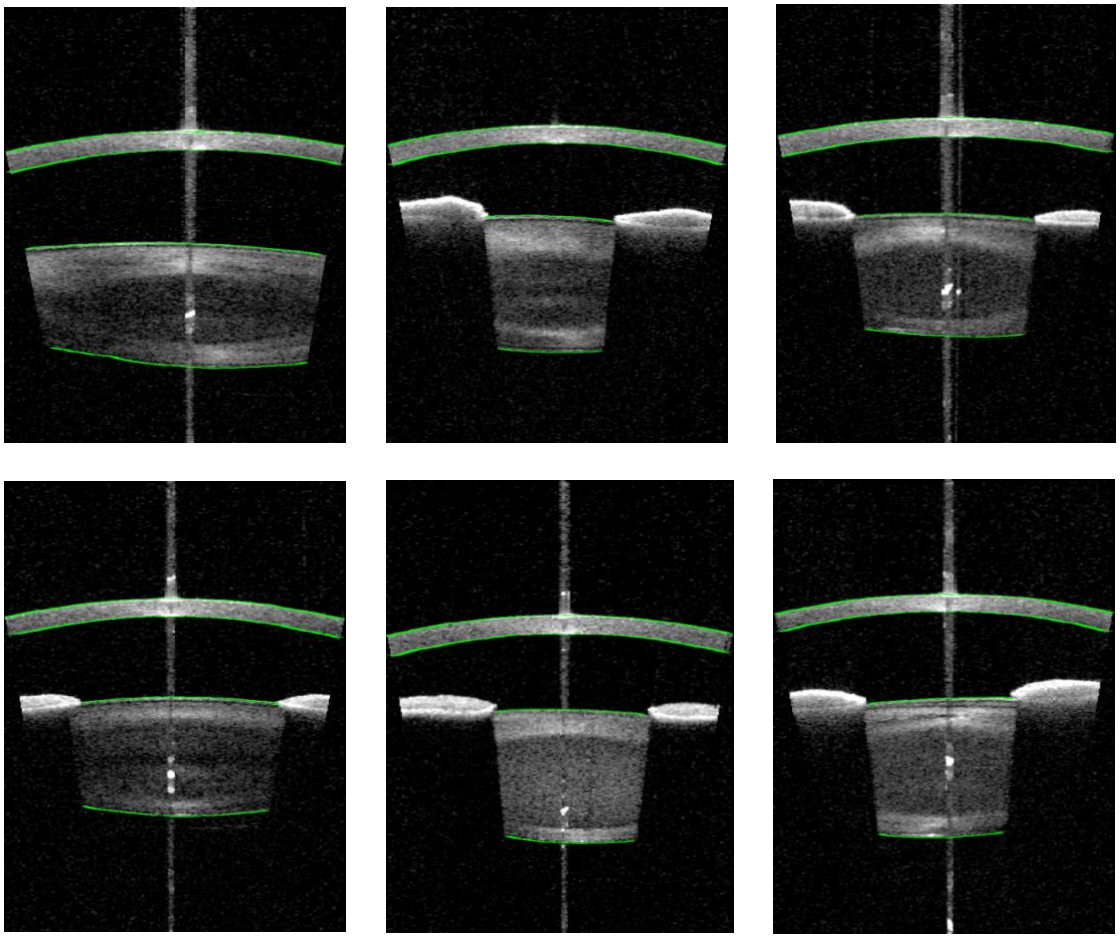


Fig. 16. Predictions of the semantic segmentation based on deep learning.

As can be seen in Figure 16, areas of interest of the cornea and crystalline lens were successfully segmented across a wide range of different images.

One desired feature of our segmentation model is to segment images as fast as possible. Our semantic segmentation deep learning model takes approximately 1.3 seconds to segment new images.

7.2. Deep learning vs traditional approach

The aim of this section is to show some of the images that were wrongly segmented by the described traditional approach, and visually compare them with the predictions made by the deep learning model. This comparison included images belonging to the training and test set. In addition, IoU will be used to compute the model performance of both algorithms on the test set.

Figure 17 shows some of the segmentation results obtained by the traditional approach (left column) and by the deep learning model (right column) on test images.

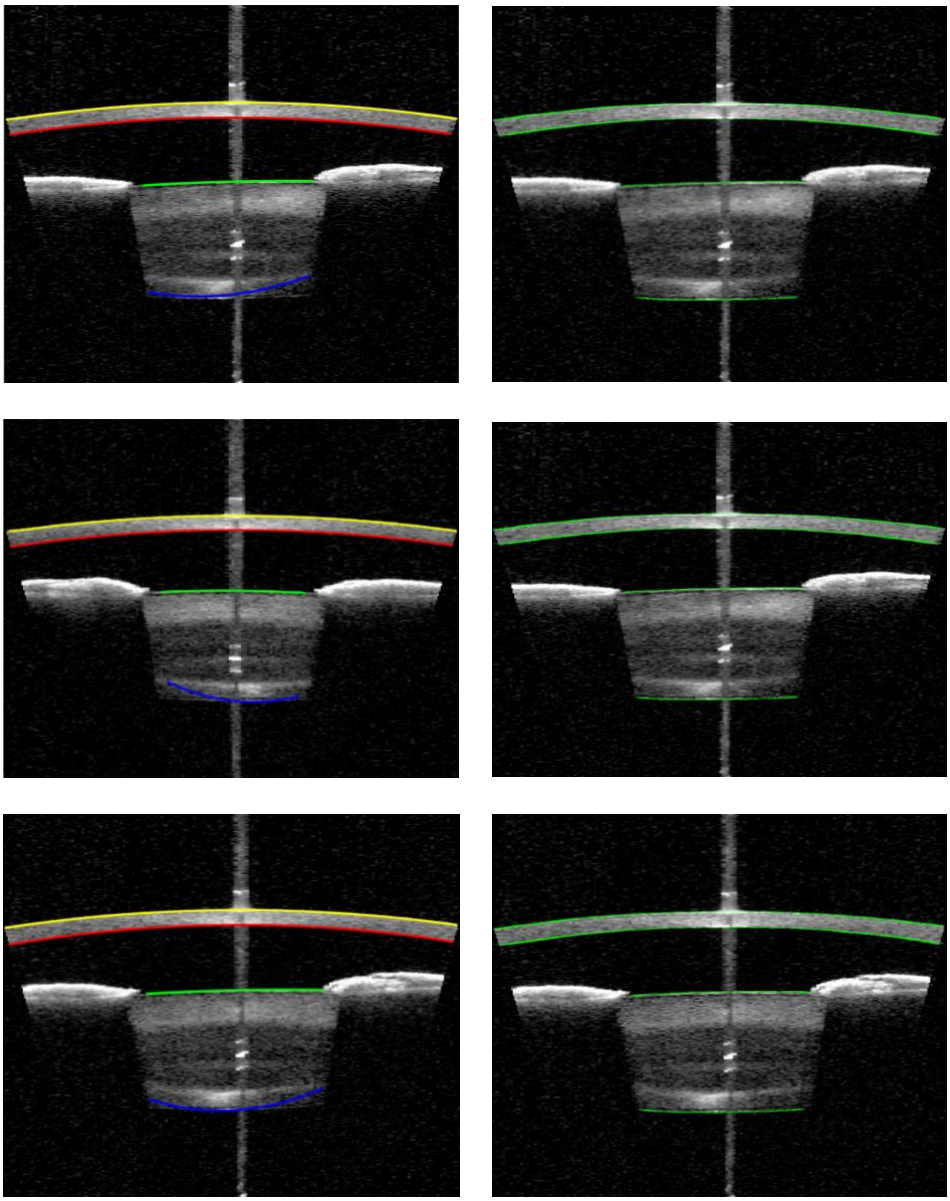


Fig. 17. Predictions made by the traditional approach (left column). Predictions obtained using deep learning (right column). Images included in the test set.

While the described traditional algorithm failed to segment the posterior crystalline lens in the images shown, the deep learning model was capable of segmenting it correctly. This seems to make the deep learning model more suitable and reliable for semantic segmentation.

For each image of the test set, IoU between the ground truth and predicted mask made by the traditional approach was computed. As already explained in previous sections, the ground truth masks used for model training were three-dimensional (background, cornea, and crystalline lens), so results below are given using these masks. Across all images of the test set, the following results were obtained:

1. The mean IoU of the cornea was 92,86% with a standard deviation of 1%.
2. The worst IoU of the cornea was 83,2%.
3. The mean IoU of the crystalline lens was 95,19% with a standard deviation of 2,5%.
4. The worst IoU of the crystalline lens was 83,60%.
5. The mean IoU of the background class was 98,58% with a standard deviation of 0,001%.
6. The mean IoU of all classes was 95,54%.

In addition, IoU between the ground truth and predicted mask made by the deep learning model was also computed. Across all images of the test set, the following results were obtained:

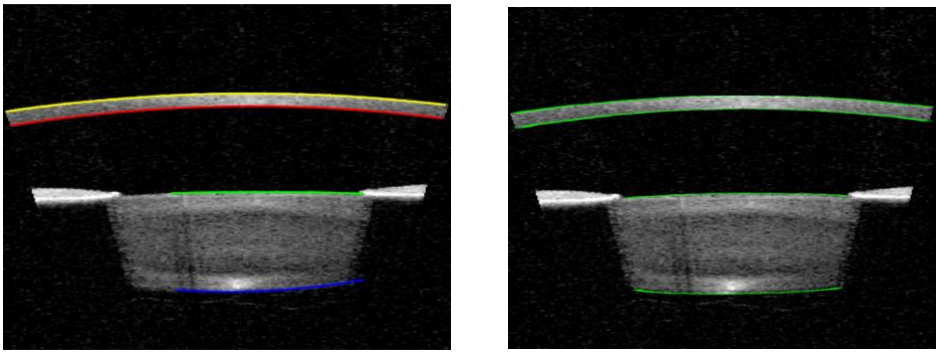
1. The mean IoU of the cornea was 94,50% with a standard deviation of 1%.
2. The worst IoU of the cornea was 89,00%.
3. The mean IoU of the crystalline lens was 96,36% with a standard deviation of 1%.
4. The worst IoU of the crystalline lens was 86,40%.
5. The mean IoU of the background class was 98,78% with a standard deviation of 0,001%.
6. The mean IoU of all classes was 96,54%.

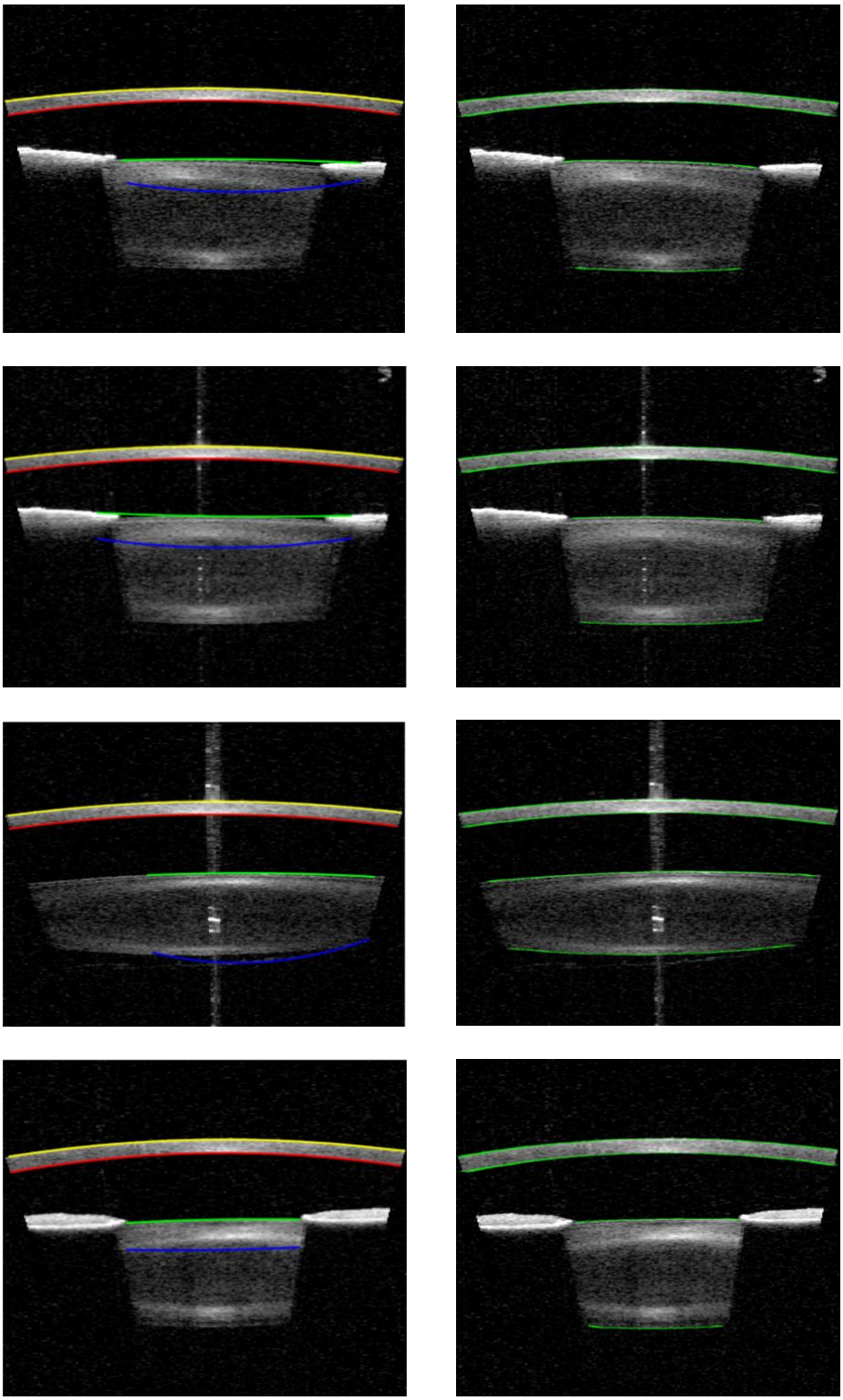
As can be seen from the previous results, the cornea and crystalline lens were better segmented using the deep learning approach. Furthermore, the standard deviation across

images in the crystalline lens segmentation was lower with the deep learning approach (1%) than with the traditional approach (2,5 %), and the IoU of the worst cases were higher with the deep learning approach. These results show the reliability and robustness of the deep learning algorithm.

It is important to note that some of the ground truth masks used throughout this project were obtained using the described traditional semantic segmentation approach. This means that the comparative results between the deep learning model and the traditional approach in terms of IoU would be even more favorable to the deep learning approach if all ground truth masks had been manually labelled.

Figure 18 shows segmentation results obtained by the traditional approach (left column) and by the deep learning model (right column) on the training set of images. It is important to note that we are interested in a deep learning model that can generalize well on unseen images, so comparing the performance using images from the training set is not ideal. Our deep learning model could be overfitting these images, so this comparison is just for illustrative purposes. Seeing how good model performance on the test set is and the similarity of all images, we hypothesize that we if the following images were to be included in the test set, the results would be similar.





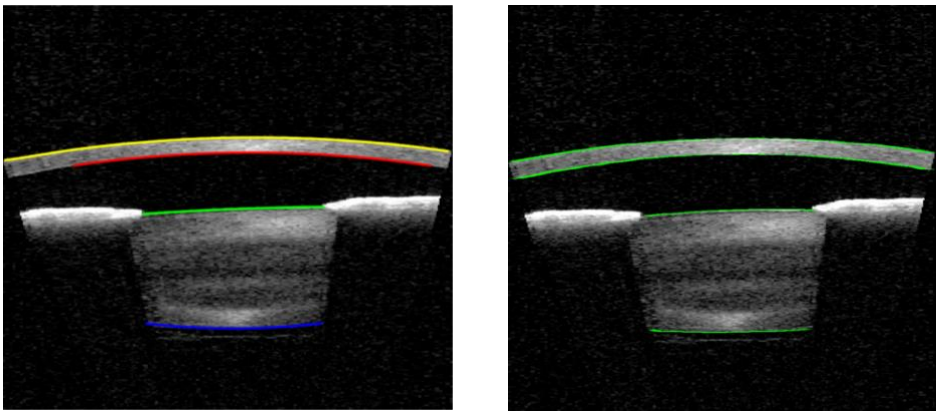


Fig. 18. Predictions of the traditional approach (left column). Predictions obtained using deep learning (right column). Images included in the train set.

As can be seen from Figure 18, one of the biggest advantages of the implemented deep learning model is reliability. As opposed to the described traditional approach, the predictions made by the deep learning are always satisfactory (we did not find any wrongly segmented example), which could have a positive impact for the quantification of the human eye.

8. CONCLUSIONS

8.1. Conclusions

The goal of this project was to create a computer vision solution that could improve the performance and address the shortcomings of a semantic segmentation model based on traditional image processing techniques. To do so, we created a semantic segmentation model based on deep learning that consisted of:

- Architecture: Feature Pyramid Network (FPN) with ResNet50 as an encoder-decoder pretrained on ImageNet.
- Loss function: Focal loss.
- Optimizer: Adam.
- Learning rate: 0.003.
- Metric: Intersection over Union.
- Training batch size of 8.
- Validation batch size of 4.

Our proposed semantic segmentation deep learning model achieved several accomplishments:

1. Improved the performance of the traditional semantic segmentation approach.
2. Successfully segmented images that could not be segmented using the traditional model.

These two accomplishments imply that the images are segmented in a reliable an automatic way.

The main obstacle that we encountered in the design of the algorithm was class imbalance, and it was solved by creating new ground truth masks. Furthermore, the proposed semantic segmentation approach took approximately 1.3 seconds to segment new images. Therefore, it could be easily implemented in a commercial system to be used in a clinical setting. These features could have a huge positive impact in the quantification of the human eye, which in turn could help in the diagnosis and treatment of common eye conditions such as presbyopia, and cataracts.

8.2. Future research lines

- As already explained, some of the ground truth masks were obtained with the traditional image processing algorithm. Labelling all images manually would help at improving model performance, reliability, and to compare both algorithms in a fairer way.

- Since a reliable and automatic means of segmenting the images was required, deploying our model in cloud-based environments (AWS, Microsoft Azure, or Google Cloud Platform) could offer some advantages. For example, the possibility of training our model with more capable computational resources.

- It would be interesting to train the algorithm with images from other commercial or custom-designed OCT systems (that have the same structure but can be significantly different in terms of quality and visual aspect) and evaluate its performance and generalization ability in this framework.

- As the results of the project are of general interest for the visual science community, we are planning the publication of the results in a scientific journal.

9. REGULATORY FRAMEWORK

The dataset used for this project consisted of images of the human eye belonging to different patients. Medical data is considered by the GDPR [51], which is Europe's new data privacy and security law, as sensitive data. Data must be treated in accordance with the GDPR. To comply with it, the following aspects need to be considered:

- Confidentiality of medical data.
- Informed consent of every patient.
- Treatment of health data according to some of the following principles:
    - Transparency.
    - Data minimization.
    - Accuracy.
    - Integrity and confidentiality.
- Provide to the patient information about the protection of medical data.
- Risk analysis of the possible contingencies of the treatments carried out.
- Impact assessment and implementation of security measures.

The measurements were taken for a study that met the tenets of the Declaration of Helsinki. Ethical approval was granted by the Ethics Committee of Consejo Superior de Investigaciones Científicas (CSIC). Written informed consent was obtained from the patients after detailed explanation of the procedure.

The intellectual property of this project belongs to the author.

10. SOCIO-ECONOMIC ENVIRONMENT

10.1. Budget

To estimate the budget for this project, human cost together with hardware and software requirements will be considered.

It is estimated that the salary of the author of this project could be approximately 15€/h (like a junior machine learning engineer). Likewise, the salary of the tutor of this project could be 35€/h (like a senior engineer). Considering 250 hours as the time it took for the author to complete this project, and 45 hours by the tutor in supervision tasks and meetings, the total human cost associated to this project would be 5,325€ (3,750€/author, 1,575€/tutor).

The main hardware used for the code and the document of this project was an Apple MacBook Pro, valued at 1,800€, with an amortized cost of 600€ per year. Considering that it took 4 months to complete this project, the associated cost would be 200€.

The different software tools and their corresponding price used throughout the creation of this project are listed below:

- Python: an open-source and high-level programming language released under the Python Software Foundation License.
- PyTorch: an open-source machine learning framework developed by Facebook Inc.
- NumPy: an open-source package for scientific computation with Python.
- Pandas: an open-source data analysis built on top of the Python programming language.
- Google Colab Pro: product from Google Research designed to allow anybody to write and execute code through the browser using a cloud environment. The cost of Google Colab Pro was 11,99€ per month. The total cost associated to Google Colab Pro was 23,98€ (used for two months).
- Google Drive: cloud-storage service.
- Microsoft Word.

- MATLAB: free for students using UC3M license.

The total cost associated to this project would be of 5,500€.

10.2. Socio-economic impact

Image segmentation is critical for obtaining accurate 3-D human eye models. In addition, the segmentation process must be reliable and automatic to avoid repeated measurements in clinic or the need of human resources to perform the task, with the associated cost.

Obtaining accurate 3-D models of the anterior segment of the eye and their quantification is crucial to improve solutions to cataract and presbyopia, which would have a positive impact on society. To date, 28 million cataract procedures are performed yearly worldwide, being the most frequent surgical procedure in most hospitals of the world. Furthermore, cataract is responsible for 51% of the world blindness and it is considered by the World Health Organization as a priority eye disease. Presbyopia becomes noticeable at approximately age 45 y/o and occurs in nearly 100% of the population. Therefore, presbyopia compromises vision at a relatively young age, affecting productivity and quality of life.

The results of this project could be of interest for several companies that use OCT for anterior segment quantification, and then apply the quantification results for estimating the power of the IOL to be implanted in a cataract surgery. Examples of these companies are ZEISS (Germany, with its system, IOLMaster 700) or Heidelberg Engineering (Germany, with its system, ANTERION OCT). In a similar context, related patents have been previously licensed to leading companies in vision as Alcon research (USA, licensed in 2020) or Abbot Medical Optics (USA, licensed in 2018).

BIBLIOGRAPHY

[1] D. A. Atchison, "Optics of The Human Eye," 2000.

[2] K. Chuang, M. A. Fields, and L. V. Del Priore, " Potential of Gene Editing and Induced Pluripotent Stem Cells (iPSCs) in Treatment of Retinal Diseases " *The Yale Journal of Biology and Medicine,* vol. 90, pp. 635 - 642, 2017.

[3] S. Chang, Y. Mao, C. Flueraru, and S. Sherif, *Optical coherence tomography: technology and applications* vol. 7156: SPIE, 2009.

[4] J. Fujimoto and D. Huang, "Foreword: 25 Years of Optical Coherence Tomography," *Investigative Ophthalmology & Visual Science,* vol. 57, pp. OCTi-OCTii, 2016.

[5] T. Mitchell, *Machine Learning* 1ed. New York: McGraw-Hill, 1997.

[6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE,* vol. 86, pp. 2278-2324, 1998.

[7] P. Rosales, M. Dubbelman, S. Marcos, and R. van der Heijde, "Crystalline lens radii of curvature from Purkinje and Scheimpflug imaging," *J Vis,* vol. 6, pp. 1057-67, 2006.

[8] P. Rosales and S. Marcos, "Phakometry and lens tilt and decentration using a custom-developed Purkinje imaging apparatus: validation and measurements," *J Opt Soc Am A Opt Image Sci Vis,* vol. 23, pp. 509-20, Mar 2006.

[9] C. A. Cook, J. F. Koretz, A. Pfahnl, J. Hyun, and P. L. Kaufman, "Aging of the human crystalline lens and anterior segment," *Vision Res,* vol. 34, pp. 2945-54, Nov 1994.

[10] M. Dubbelman, G. L. Van der Heijde, and H. A. Weeber, "Change in shape of the aging human crystalline lens with accommodation," *Vision Res,* vol. 45, pp. 117-32, Jan 2005.

[11] V. Ramasubramanian and A. Glasser, "Objective measurement of accommodative biometric changes using ultrasound biomicroscopy," *J Cataract Refract Surg,* vol. 41, pp. 511-26, Mar 2015.

[12] D. A. Atchison, E. L. Markwell, S. Kasthurirangan, J. M. Pope, G. Smith, and P. G. Swann, "Age-related changes in optical and biometric characteristics of emmetropic eyes," *J Vis,* vol. 8, pp. 29 1-20, 2008.

[13] S. Kasthurirangan, E. L. Markwell, D. A. Atchison, and J. M. Pope, "MRI study of the changes in crystalline lens shape with accommodation and aging in humans," *J Vis,* vol. 11, 2011.

[14] I. Grulkowski, M. Gora, M. Szkulmowski, I. Gorczynska, D. Szlag, S. Marcos, *et al.,* "Anterior segment imaging with Spectral OCT system using a high-speed CMOS camera," *Opt Express,* vol. 17, pp. 4842-58, Mar 16 2009.

[15] S. Ortiz, D. Siedlecki, I. Grulkowski, L. Remon, D. Pascual, M. Wojtkowski, *et al.,* "Optical distortion correction in optical coherence tomography for quantitative ocular anterior segment by three-dimensional imaging," *Opt Express,* vol. 18, pp. 2782-96, Feb 1 2010.

[16] S. Ortiz, D. Siedlecki, L. Remon, and S. Marcos, "Optical coherence tomography for quantitative surface topography," *Appl Opt,* vol. 48, pp. 6708-15, Dec 10 2009.

[17] E. Martinez-Enriquez, M. Sun, M. Velasco-Ocana, J. Birkenfeld, P. Perez-Merino, and S. Marcos, "Optical Coherence Tomography Based Estimates of Crystalline Lens Volume, Equatorial Diameter, and Plane Position," *Invest Ophthalmol Vis Sci,* vol. 57, pp. OCT600-10, Jul 1 2016.

[18] S. Ortiz, P. Perez-Merino, E. Gambra, A. de Castro, and S. Marcos, "In vivo human crystalline lens topography," *Biomed Opt Express,* vol. 3, pp. 2471-88, Oct 1 2012.
[19] E. Martinez-Enriquez, A. de Castro, and S. Marcos, "Eigenlenses: a new model for full crystalline lens shape representation and its applications," *Biomed Opt Express,* vol. 11, pp. 5633-5649, Oct 1 2020.
[20] G. Muralidharan, E. Martínez-Enríquez, J. Birkenfeld, M. Velasco-Ocana, P. Pérez-Merino, and S. Marcos, "Morphological changes of human crystalline lens in myopia," *Biomedical Optics Express,* vol. 10, pp. 6084-6095, 2019/12/01 2019.
[21] E. Martinez-Enriquez, P. Perez-Merino, S. Duran-Poveda, I. Jimenez-Alfaro, and S. Marcos, "Estimation of intraocular lens position from full crystalline lens geometry: towards a new generation of intraocular lens power calculation formulas," *Sci Rep,* vol. 8, p. 9829, Jun 29 2018.
[22] J. Gao, M. Liu, Y. Zou, M. Mao, T. Shen, C. Zhang, *et al.*, "Long non-coding RNA growth arrest-specific transcript 5 is involved in ovarian cancer cell apoptosis through the mitochondria-mediated apoptosis pathway," *Oncol Rep,* Oct 1 2015.
[23] S. Marcos, E. Martinez-Enriquez, M. Vinas, A. de Castro, C. Dorronsoro, S. P. Bang, *et al.*, "Simulating Outcomes of Cataract Surgery: Important Advances in Ophthalmology," *Annu Rev Biomed Eng,* vol. 23, pp. 277-306, Jul 13 2021.
[24] P. Pérez-Merino, M. Velasco-Ocana, E. Martinez-Enriquez, and S. Marcos, "OCT-based crystalline lens topography in accommodating eyes," *Biomedical Optics Express,* vol. 6, pp. 5039-5054, 2015/12/01 2015.
[25] E. Martinez-Enriquez, P. Perez-Merino, M. Velasco-Ocana, and S. Marcos, "OCT-based full crystalline lens shape change during accommodation in vivo," *Biomed Opt Express,* vol. 8, pp. 918-933, Feb 01 2017.
[26] V. Akondi, P. Perez-Merino, E. Martinez-Enriquez, C. Dorronsoro, N. Alejandre, I. Jimenez-Alfaro, *et al.*, "Evaluation of the True Wavefront Aberrations in Eyes Implanted With a Rotationally Asymmetric Multifocal Intraocular Lens," *J Refract Surg,* vol. 33, pp. 257-265, Apr 01 2017.
[27] M. Sun, P. Pérez-Merino, E. Martinez-Enriquez, M. Velasco-Ocana, and S. Marcos, "Full 3-D OCT-based pseudophakic custom computer eye model," *Biomedical Optics Express,* vol. 7, pp. 1074-1088, 2016/03/01 2016.
[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Cham, 2015, pp. 234-241.
[29] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," Cham, 2018, pp. 3-11.
[30] Z. Mishra, A. Ganegoda, J. Selicha, Z. Wang, S. R. Sadda, and Z. Hu, "Automated Retinal Layer Segmentation Using Graph-based Algorithm Incorporating Deep-learning-derived Information," *Scientific Reports,* vol. 10, p. 9541, 2020/06/12 2020.
[31] M. Melinščak, P. Prentasic, and S. Lončarić, "Retinal Vessel Segmentation using Deep Neural Networks," in *VISAPP*, 2015.
[32] Q. Li, S. Li, Z. He, H. Guan, R. Chen, Y. Xu, *et al.*, "DeepRetina: Layer Segmentation of Retina in OCT Images Using Deep Learning," *Translational Vision Science & Technology,* vol. 9, pp. 61-61, 2020.
[33] I. Cabeza-Gil, M. Ruggeri, Y.-C. Chang, B. Calvo, and F. Manns, "Automated segmentation of the ciliary muscle in OCT images using fully convolutional networks," *Biomedical Optics Express,* vol. 13, pp. 2810-2823, 2022/05/01 2022.

[34] E. Martinez-Enriquez, A. de Castro, A. Mohamed, N. G. Sravani, M. Ruggeri, F. Manns, *et al.*, "Age-Related Changes to the Three-Dimensional Full Shape of the Isolated Human Crystalline Lens," *Invest Ophthalmol Vis Sci,* vol. 61, p. 11, Apr 9 2020.

[35] E. Martinez-Enriquez, P. Perez-Merino, M. Sun, S. Durán-Poveda, I. Jiménez-Alfaro, and S. Marcos, "Full shape crystalline lens quantification from 3-D OCT images and its application to predict the post-operative IOL position," *Investigative Ophthalmology & Visual Science,* vol. 58, pp. 1803-1803, 2017.

[36] P. Pérez-Merino, M. Velasco-Ocana, E. Martinez-Enriquez, L. Revuelta, S. A. McFadden, and S. Marcos, "Three-dimensional OCT based guinea pig eye model: relating morphology and optics," *Biomedical Optics Express,* vol. 8, pp. 2173-2184, 2017/04/01 2017.

[37] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 9, pp. 62-66, 1979.

[38] R. Sun, T. Lei, Q. Chen, Z. Wang, X. Du, W. Zhao, *et al.*, "Survey of Image Edge Detection," in *Frontiers in Signal Processing*, 2022.

[39] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm," *Procedia Computer Science,* vol. 54, pp. 764-771, 2015.

[40] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International Journal of Multimedia Information Retrieval,* vol. 7, pp. 87-93, 2017.

[41] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data,* vol. 6, pp. 1-54, 2019.

[42] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *CoRR,* vol. abs/1511.06434, 2016.

[43] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data,* vol. 6, pp. 1-48, 2019.

[44] T.-Y. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *2017 IEEE International Conference on Computer Vision (ICCV),* pp. 2999-3007, 2017.

[45] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, pp. 640-651, 2017.

[46] T.-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 936-944, 2017.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 770-778, 2016.

[48] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition,* pp. 248-255, 2009.

[49] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR,* vol. abs/1412.6980, 2015.

[50] S. Jadon, "A survey of loss functions for semantic segmentation," *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB),* pp. 1-7, 2020.

[51] J. F. A. Murphy, "The General Data Protection Regulation (GDPR)," *Irish medical journal*, vol. 111 5, p. 747, 2018.