FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Characterization of Retinal Fluid in OCT images

David Castanho Terroso

WORKING VERSION

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Mestrado em Engenharia Biomédica

Supervisor: Prof. Ana Maria Mendonça

Second Supervisor: Prof. Tânia Melo

June 24, 2025

Resumo

Abstract

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework to achieve a better and more sustainable future for all. It includes 17 goals to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis.

The specific Sustainable Development Goals mentioned have the following names:

SDG 7 Ensure access to affordable, reliable, sustainable and modern energy for all

SDG 8 Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

SGD	**Target**	**Contribution**	**Performance Indicators and Metrics**
7	7.1	Enhancing the efficiency of SCADA systems can help increase the reliability of solar energy production, facilitating universal access to clean energy.	Percentage of solar plants with improved ...
	7.2	Improving the management of solar plants helps enhance the efficiency and reliability of renewable energy ...	Increase in renewable energy share ...
8	8.1	Enhancing renewable energy infrastructure promotes resilience against climate-related hazards and supports sustainable energy sources.	Increase in resilience metrics ...

Acknowledgements

David Castanho Terroso

"Our greatest glory is not in never falling, but in rising every time we fall"

Confucius

Contents

List of Figures

List of Tables

List of Acronyms

AMD	Age-related macular degeneration
BM	Brunch’s membrane
CNN	Convolutional neural network
CT	Computed tomography
DME	Diabetic macular edema
GAN	Generative adversarial network
GT	Ground truth
HR	High-resolution
ILM	Internal limiting membrane
IRF	Intraretinal fluid
MAE	Mean absolute error
MRI	Magnetic resonance imaging
LR	Low-resolution
OCT	Optical coherence tomography
ONL	Outer nuclear layer
PED	Pigment epithelial detachment
PSNR	Peak signal-to-noise ratio
RPE	Retinal pigment epithelium
RVO	Retinal vein occlusion
SR	Super-resolution
SRF	Subretinal fluid
SSIM	Structural similarity index measure

Chapter 1

Introduction

The vision is the human's most important and complex sense, playing a critical role in our orientation in the world [1]. However, the health of the retina, an important part of the eye, can be compromised by multiple diseases, that lead to fluid accumulation in it. The characterization of the fluid present in the retina is important to assess the progression of diseases such as age-related macular degeneration (AMD), diabetic macular edema (DME), and macular edema secondary to retinal vein occlusion (RVO) [2].

AMD affects the macular region of the retina, leading, in later stages, to a significant and permanent loss of central visual acuity, which has a severe impact on the patient's quality of life. In patients with AMD, the formation of new blood vessels can occur, which leak fluid, lipids, and blood into the retina, resulting in the formation of retinal fluid [3]. It is one of the leading causes of visual impairment with an expected effect on 300 million people by 2040 [4].

In patients with diabetes mellitus, DME represents the most common cause of visual impairment, affecting approximately 150 million people worldwide, as of 2015. It is anticipated that this number will increase as the prevalence of diabetes in developed countries is growing [5]. The fluid accumulation is caused by a disruption of the blood-retinal barrier, which allows fluid to accumulate in the intraretinal layers of the macula, resulting in retinal thickening (edema) [6, 7].

Affecting 16 million people worldwide, RVO represents a significant cause of vision loss in older individuals. The occlusion of the retinal vein can result in swelling of the optic disc, which leads to a reduction in visual acuity [8].

The presence of intraretinal fluid (IRF) is a defining criterion of DME and RVO, while two in every three patients with AMD present this type of fluid. The majority of patients with AMD and 30% of the patients with DME and RVO have subretinal fluid (SRF). Pigment epithelial detachments (PED) occurs more frequently in patients with AMD [2].

Therefore, retinal fluids are important for the classification and progression of these diseases, and can be observed through retinal optical coherence tomography (OCT) [2]. OCT is a non-invasive imaging technique that analyzes the light behavior (such as its reflection, absorption, and time-of-flight) to estimate the spatial dimensions of the tissue's structure [9]. This allows for *in vivo* visualization of the individual retinal layers within the posterior segment of the eye. An

OCT is composed of multiple consecutive cross-sectional 2D images that, when stacked, form a volumetric representation of the posterior segment. Each of these two-dimensional images is referred to as B-scans and an example can be seen in Figure 1.1. The resolution is sufficiently high to assess the tissue integrity, the retinal layers, and the fluids present [10, 11]. There are multiple devices used for the acquisition of OCT volumes, resulting in different image attributes across the same technique, such as interslice distance, image quality, and appearance [2].

The classification of the fluid is dependent on its location within the retina. There are three different categories: IRF, which is situated in the inner and outer layers of the retina; SRF, positioned between the outer nuclear layer (ONL) and the retinal pigment epithelium (RPE); and PED, which appear beneath the RPE [2]. Figure 1.1 shows the characteristics and positions of these fluids on an OCT B-scan and Figure 1.2 exhibits the retinal layers in the OCT scan of a healthy patient.

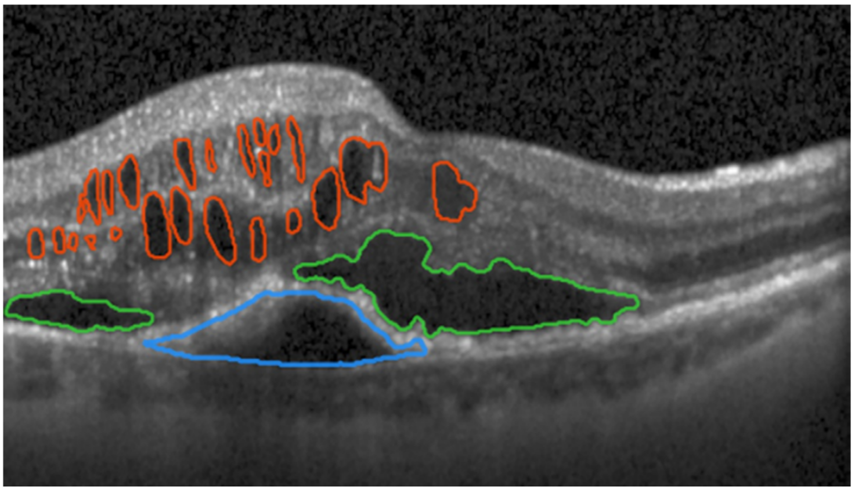

Figure 1.1: The three distinct fluid types on an OCT B-scan: IRF in red, SRF in green, and PED in blue [2].

By segmenting the fluids detected in the B-scans, their volume can be estimated and used as a progression marker of the mentioned retinal diseases. However, manual segmentation is laborious, expensive, and prone to bias, which motivates the search for automatic methods [11].

In OCT imaging, the precision of the estimated volume is not only dependent on the quality of the segmentation, but also on the interslice distance [13]. It is seen in other imaging techniques that the performance of the segmentation is improved when the neighboring slices are used as input. Consequently, the reduction in interslice distance and improvement of the resolution along this axis, betters the performance of models that include information from adjacent slices [14]. Given that the interslice space is reduced, the estimated segmented volume will also be closer to the real fluid volume.

Considering the previous statements, the dissertation general objective is to conduct an analysis of retinal OCT scans, classifying the retinal fluids in three distinct types (IRF, SRF, and PED) and quantifying their respective volumes. Another important objective is to increase the interslice resolution of the OCT volumes, with the aim of improving the fluid volume estimation. The

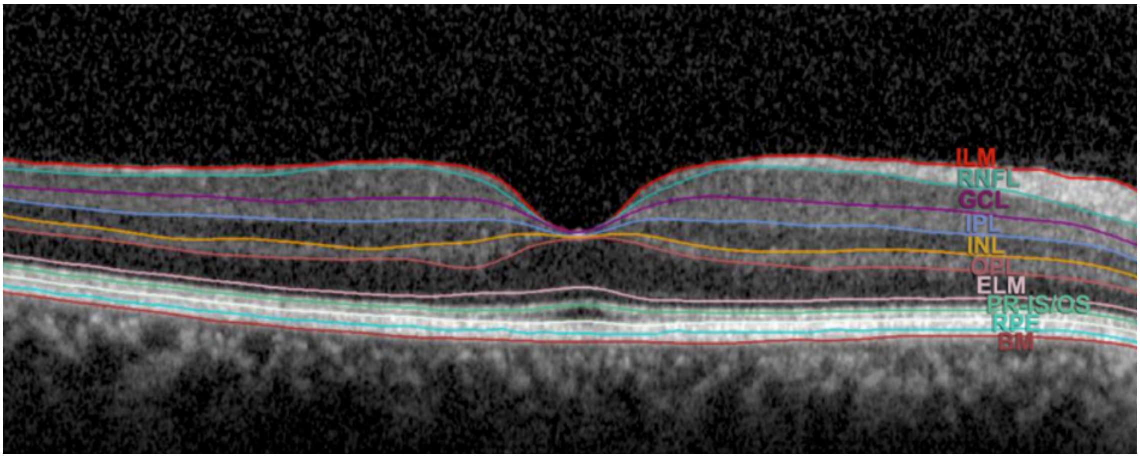

Figure 1.2: OCT scan of the retinal layers [12].
BM, Bruch's membrane; ELM, external limiting membrane; GCL, ganglion cell layer; ILM, internal limiting membrane; INL, inner nuclear layer; IPL, inner plexiform layer; ONL, outer nuclear layer; OPL, outer plexiform layer; PR-IS/OS, photoreceptor inner segment/outer segment; RNFL, retinal nerve fibre layer; RPE, retinal pigment epithelium.

specific objectives were determined as follows:

1. Develop different 2D deep learning models for multi-class segmentation of retinal fluids (IRF, SRF, and PED) in OCT volumes.

2. Compare the best performing 2D model with a previously implemented 2.5D model.

3. Evaluate the performance of the best segmentation model and estimate the volume of each fluid using the masks predicted by it.

4. Use of a generative model for synthesizing intermediate slices in OCT volumes, generating one or more slices between two real slices to improve the interslice resolution of the volume, while assessing the quality of these generated images.

5. Investigate the impact of intermediate slices synthesis on the fluid volume estimation by the segmentation models.

Apart from the "Introduction", the dissertation is composed of the following chapters: "Literature Review", "Methods", "Results", "Discussion", and "Conclusion". In the "Literature Review" chapter, an analysis is performed on the latest papers in the field of retinal fluid segmentation using 2D deep learning networks, as well as the latest publications on interslice resolution enhancement. The "Methods" chapter details the selection of the dataset for the experiments performed during the dissertation, alongside with an insightful description of the experiments. In the "Results" chapter, the results from each experiment are shared, showing the performance of each model in their respective task, while the "Discussion" chapter explains the performance differences between experiments, comparing them to the literature. Finally, the "Conclusion" shows the main findings from the experiments performed and suggests directions of further research.

Chapter 2

Literature Review

For this dissertation, research was conducted to find the most recent trends in 2D fluid segmentation of OCT volumes using deep learning and in the use of generative models in the intermediate slice synthesis.

2.1 Fluid Segmentation

In the fluid segmentation state-of-the-art research, articles were retrieved using the methodology of a systematic review. The next subsection details the retrieval process and the criteria for inclusion and exclusion of the articles. "2.1.2 Literature Review" shows the trends on the methodologies utilized for fluid segmentation.

2.1.1 Search Strategy

The search query was defined as: ""OCT" AND "segmentation" AND ("deep learning" OR "CNN" OR "neural network")". Using the query, papers were retrieved from four different databases: 398 articles from PubMed, 105 from IEEE, 125 from ScienceDirect, and 80 from ACM.

In the process of collecting the papers, those published over the previous five years and regarding 2D or 2.5D fluid segmentation in OCT volumes were included. Additionally, conferences proceedings, articles not written in English, and articles for which the full text was not accessible were excluded.

A total of 708 articles were initially identified, of which 133 were duplicates. Afterwards, 575 articles were subjected to screening, based on their titles and abstracts. These articles were analyzed in accordance with the inclusion and exclusion criteria, resulting in the removal of 499 papers. Of the remaining 76 articles for the full-text screening, 20 met the established criteria. These final articles represent the state-of-the art in 2D deep learning fluid segmentation in OCT volumes included in this dissertation.

2.1.2 Literature Review

The selected papers can be divided into two broad groups, according to the type of segmentation: binary segmentation [15, 16, 17, 18, 19, 20], where the fluid is classified in one whole class, and multi-class [21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34], where the segmented fluid is classified in two or more classes (namely IRF, SRF, and PED). We have also considered other criteria to group the papers, such as the segmentation architecture, and the use of retinal delimitation, as shown in Figure 2.1.

In binary segmentation, the approaches to the segmentation problem are simpler, but include both convolutional neural network (CNN) [16, 17, 18, 19, 20] and transformer solutions [15]. The CNN solutions differ among them, depending on the modules that constitute each network, but all are inspired by the U-Net [35]. In Figure 2.2, an instance of a CNN used for binary fluid segmentation is shown.

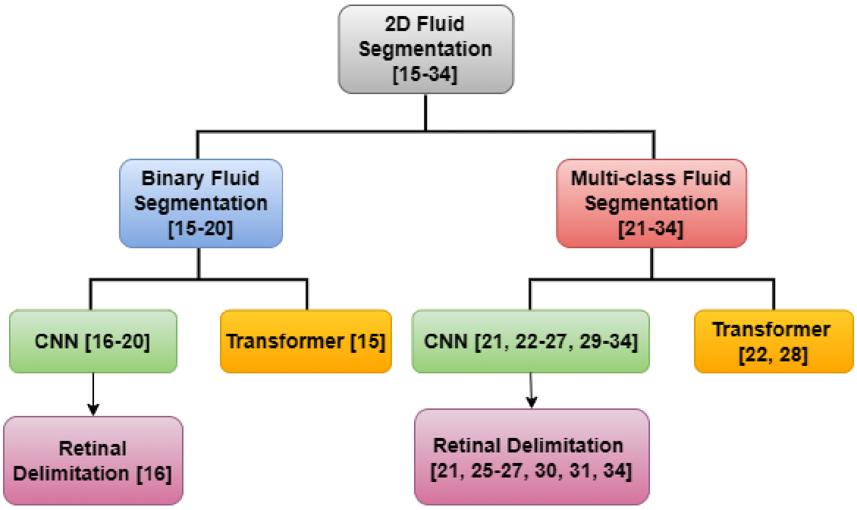

Figure 2.1: Grouping of the articles included in the literature review.

Pawan et al. [16] is the only paper in binary segmentation that restricts the input of the segmentation CNN to the content within the retinal layer. This approach is frequently observed in the papers focused on multi-class segmentation. In this article, this is achieved by performing a retinal layer segmentation and assigning all the values outside the boundaries to zero. The result of this operation is an input for the segmentation CNN. The removal of irrelevant information surrounding the retina simplifies the learning process and improves the model's focus on essential information [34].

In the framework proposed by Liu et al. [17], the slice's fluid mask and distance map are generated. The distance map consists of the predicted distance of each pixel to the background or retinal tissue, with only the values above a specified distance threshold being kept. This is achieved through the use of a double-branched network, where the encoder is the same, while the decoders vary. One encoder is responsible for generating the fluid segmentation map, while

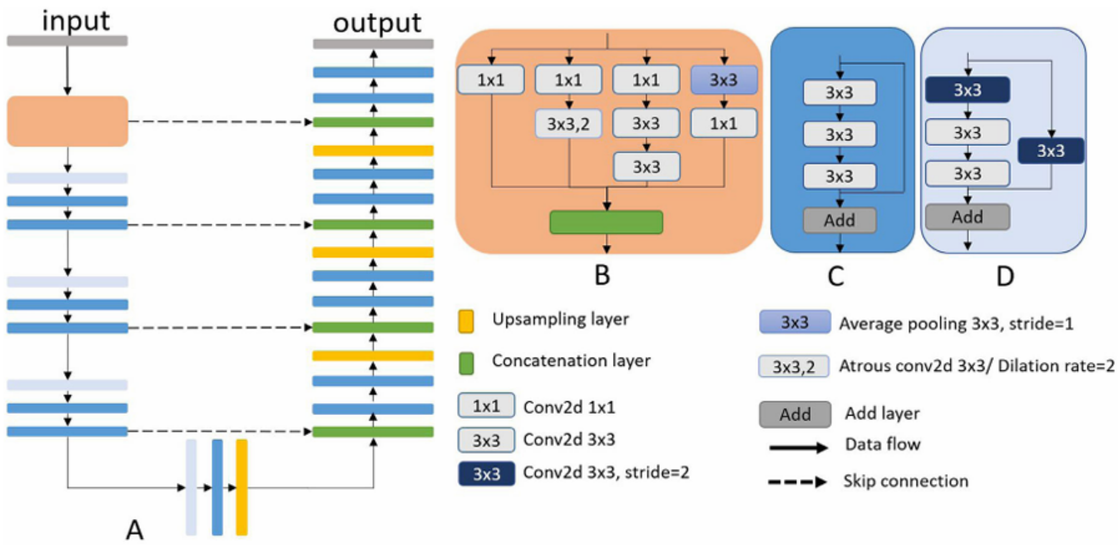

Figure 2.2: Example of a CNN architecture used in fluid binary segmentation. Image A depicts the neural network architecture. B shows the used multi-scale block, while C and D exhibit the residual convolutional blocks [18].

the other predicts the distance map. The intersection between these outputs forms the final segmentation. This approach mitigates the issue of inappropriate merging of small and proximate fluid regions, as the distance branch is better than the fluid segmentation network in discerning the boundaries that delineate fluid regions.

Resorting to generative adversarial networks (GANs), Wu et al. [20] make images from different vendors, visually similar to the images of a singular, specific vendor. Subsequently, a U-Net, which has extensively been trained on images from the specific vendor, is used for segmentation. This approach is intended to reduce the burden of learning the segmentation on multiple vendors by ensuring that all volumes are similar to one in which the segmentation model performs well. Similarly, the multi-class segmentation framework proposed by Li et al. [23] was designed based on the same idea.

CNNs inspired by the U-Net can also be combined with transformers in the context of image segmentation. While CNNs capture the information from local receptive fields, visual transformers integrate features from global receptive fields. Despite being more prevalent in multi-class segmentation frameworks, in this paper by Quek et al. [15], the visual transformers are located between the encoder and decoder paths, thus incorporating features from both receptive fields in the encoding branch.

The majority of the papers included in this review perform multi-class segmentation models, therefore presenting more diverse implementations. While all these articles segment two or more fluids, Hassan et al. [27] and Padilla-Pantoja et al. [32] also segment other biomarkers. Similarly to binary segmentation, the multi-class segmentation papers can also be divided according to the presence [22, 28] or absence [21, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, 34] of transformers in the segmentation network. All the papers that have transformers in their framework, combine them with CNNs.

Similar to what was developed in Quek et al. [15], Liu et al. [22] have integrated transformers in the bottleneck section of a segmentation network inspired by the U-Net. Liu et al. [22] utilize two networks for the segmentation: one for coarse segmentation and other for the refinement of the results from the first. Both networks are similar to the U-Net, but in the refine branch, a transformer is included. Its purpose is to provide features from global fields, compensating for the deep features that are used as input in this branch. In contrast, Zhang et al. [28] replaced the CNN encoder with a transformer encoder, exploiting its modeling capacity with self-attention.

The limitation of the input to the region within the retinal layer, ignoring what is outside of it, is seen in many of the multi-class papers [21, 25, 26, 27, 30, 31, 34], similarly to what was done in Pawan et al. [16]. There are various approaches for this delimitation, with some using CNNs trained for the segmentation of the retinal layers or the retina [31, 34], and others using algorithms leveraging on the noticeable transition between the retinal layers and its background [16, 21, 25, 26, 27, 30].

The retinal delimitation is conducted as a separate process from the fluid segmentation. In [21, 25, 26, 27, 30, 31], the retinal layer is segmented prior to the fluid segmentation, conditioning the input of the fluid segmentation network and simplifying the learning process. However, the retinal delimitation can also limit the final segmentation by intersecting the network's output, limiting the segmentation results to the boundaries of the retinal layer, as observed in Mantel et al. [34].

The fluid segmentation network input is conditioned in multiple ways. In Xing et al. [30], the image is cropped to fit its region of interest. [21, 26, 31] combined the B-scan with the retinal delimitation result, either through concatenation or along another channel. In [16, 25, 27] the information outside the retinal layer is set to zero and ignored.

Contrasting with the work of Liu et al. [17] who used a CNN to output a distance map (relative to the background or the retinal tissue), Tang et al. [31], and Rahil et al. [21], inspired by the work of Lu et al. [26], calculate a relative distance map to combine with the input slice in a CNN. Starting with the retinal delimitation, the relative distance to the internal limiting membrane (ILM) is calculated for each pixel located between the ILM and the Bruch's membrane (BM) (see Figure 1.2). This map provides information about the relative position of each pixel to the ILM, influencing their classification. An example of such framework can be seen in Figure 2.3.

Regarding the segmentation CNNs adopted by the analyzed papers, most are directly inspired by the U-Net, to which changes are done, when considering the objectives of each study. Examples of such changes are the introduction of blocks (such as residual [22, 25, 27, 28, 32, 34]), and modules (like atrous sampling pyramid pooling [25, 27, 29, 33]), which makes the network distinctive. However, some papers use other variations of the U-Net that are also popular: the Deeplab [36] in Li et al. [23] and Hassan et al. [27], and the VGG [37] in Hassan et al. [25] and Padilla-Pantoja et al. [32].

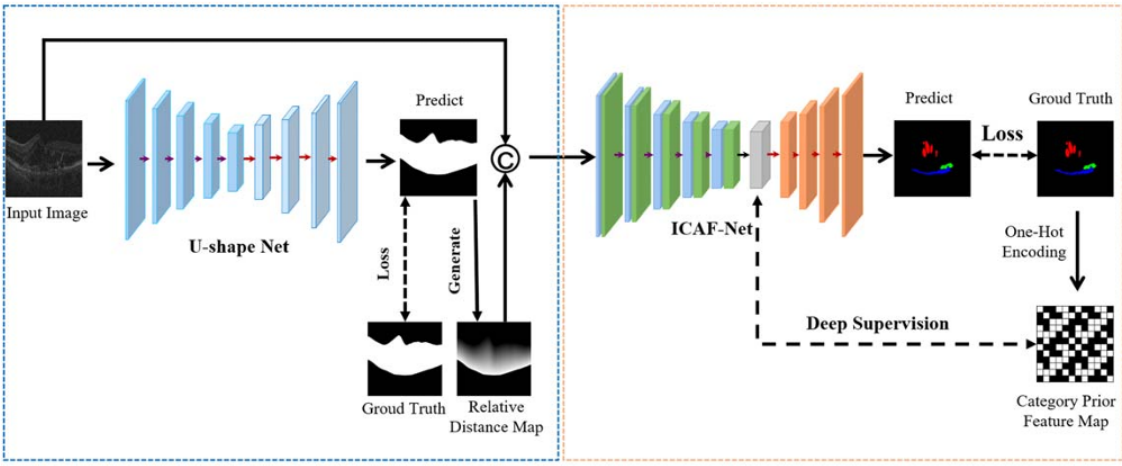

Figure 2.3: Example of a framework that includes delimitation of the retinal layer and a relative distance map (left side). The generated map is included in the segmentation network (denominated ICAF-Net, by the authors) [31].

2.2 Intermediate Slice Synthesis

For many years, there have been attempts to improve the resolution of OCT exams using computational methods, a process called super-resolution (SR). In 3D applications, such as magnetic resonance imaging (MRI), computed tomography (CT), and OCT, SR can be done intra-slice, which improves the resolution of each slice in the volume along one plane, or inter-slice, bettering the resolution of the volume along one axis, by generating one or more slices between a pair of original ones. Some frameworks may contain both approaches [38].

The use of GANs to generate slices between other known slices is commonly used technique in MRI and CT, but with few examples in OCT [38]. The systematic literature review performed by Ibrahim et al. [39], which analyzed the latest trends in the use of generative models in medical data, only presents one example of GAN for inter-slice resolution improvement in OCT volumes [13]. In this imaging technique, the use of GANs is mainly done for the generation of OCT images and conversion between different vendors [39].

In the following subsection, the state-of-the-art architectures used for the improvement of inter-slice resolution are presented.

2.2.1 Architectures

Given the lack of examples in OCT imaging, it was considered appropriate to study works from other imaging techniques, given that the working principle is the same across them. The selected papers can be classified into three distinct categories: inter-slice SR, which leverages information from adjacent slices to generate one or more intermediate slices [13, 40, 41, 42]; intra-slice SR combined with inter-slice SR, which improves the resolution of the slices from orthogonal planes and combines them with the results of inter-slice SR [43, 44, 45, 46]; and SR applied directly in 3D volumes, utilizing three-dimensional convolutions in the generation process, which incorporates the information along all the axes from multiple slices simultaneously [47, 48, 49, 50].

In López-Varela et al. [13], present an inter-slice SR framework based on a GAN (inspired by the ResNet) for the generation of three B-scan slices between two known slices. The GAN training process, as illustrated in Figure 2.4, begins with the generation of an intermediate slice (Central Fake) located between two original B-scans (Pre and Post), which are separated by another original one (Central). The Central B-scan will serve as the ground truth (GT) and will be used for the assessment of image quality generated by the network. Subsequently, the network generates other two slices: one between the Pre and Central slices, designated as Pre-Central fake, and another between the Central and Post slices, named Post-Central fake. As the mentioned generations lack a corresponding GT, the network performance is regulated by using these two new synthetic slices to generate an additional Central Fake (Central Fake 2), which is then compared to the true Central. Consequently, if the generation of Pre- and Post-Central fakes are inadequate, the Central Fake 2 will also be of poor quality, resulting in a higher loss value. During the inference process, one slice is synthesized for every two known B-scans, reducing the inter-slice distance to half of the original value.

The importance of this study comes not only from it being the only study in OCT but also from the approach selected, which is similar to the foundation of the frameworks implemented in other papers.

In a more straightforward approach, Nishimoto et al. [42] utilize a baseline U-Net that uses two spaced slices as input to generate the slices between them. This methodology was tested in the generation of three, four, and five intermediate slices, and obtained better outcomes than those generated through linear interpolation.

The work by Xia et al. [40] demonstrates the enhancement of inter-slice resolution in MRI, through the utilization of multiple networks and a multi-scale discriminator that considers both the image from a large and a small field of view images. Additionally, the framework also estimates the depth and flow of each pair of images, generating an independent image from the GAN. This subsequently facilitates the generation of the intermediate slice.

Similarly, Wu et al. [41] improved the inter-slice resolution by training a single GAN to generate bi-directional spatial transformations instead of producing fake images. The advantage of this process is that it allows the same transformations to be applied to the segmentation masks from the surrounding slices, generating fake masks for the fake slices. As in Xia et al. [40], the discriminator also judges the generated images in both a larger and smaller field of view, but combines this with an object classifier that checks if the structures present in the neighboring slices appear in the fake one.

As an example of the use of intra-slice SR to improve inter-slice resolution, Zhang et al. [43] implemented two GANs that increase the resolution in the slices of the two planes with the lowest resolution of the volume (sagittal and coronal). In these planes, the dimensions of each slice are increased in the direction of the axis of lowest inter-slice resolution. Simultaneously, in the plane of highest resolution (axial), an architecture similar to López-Varela et al. [13] is used

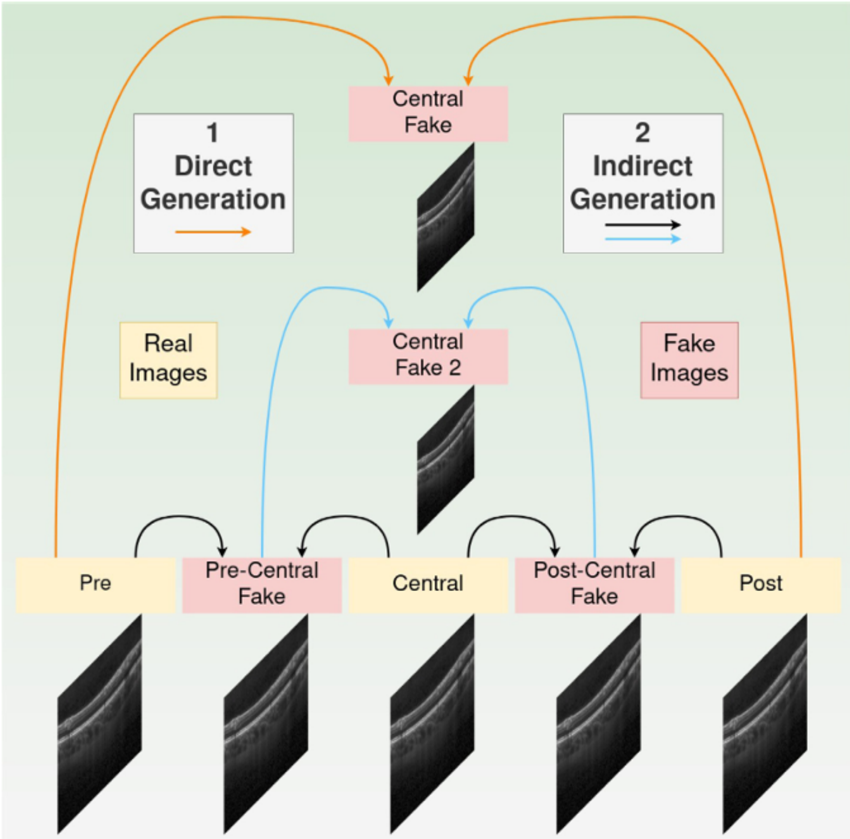

Figure 2.4: López-Varela et al. [13] training process.

in the generation of intermediate fake slices between each pair of known slices. The results from these three networks are then compared between each other, promoting the accurate and consistent generation of data among the networks.

A similar approach was done by Fang et al. [44], in which three networks (one for each axis) are trained to generate intermediate slices along one axis with a lower inter-slice resolution. However, during the unsupervised phase, upon each increase in resolution (a process that occurs twice), the information generated by the networks is compared between each other and a loss value that quantifies the performance is calculated, as illustrated in Figure 2.5.

Similarly, Nimitha and Ameer [45] use GANs to improve intra-slice resolution and a CNN to improve inter-slice resolution. The method starts by increasing the resolution of low-resolution (LR) slices, making them high-resolution (HR). Then an intermediate HR slice is generated between every set of two slices, using a CNN.

The same approach was also used by Georgescu et al. [46] to enhance the intra- and inter-slice resolution in CT and MRI scans. Two independently trained CNNs were used in LR volumes. One CNN was tasked with generating HR slices from the LR slices. Concurrently, the other CNN was utilized to reduce the distance between slices by increasing the resolution of the images from the orthogonal plane. By inferring an image with increased resolution along this plane, new intermediate slices were generated, improving the inter-slice resolution along the low resolution

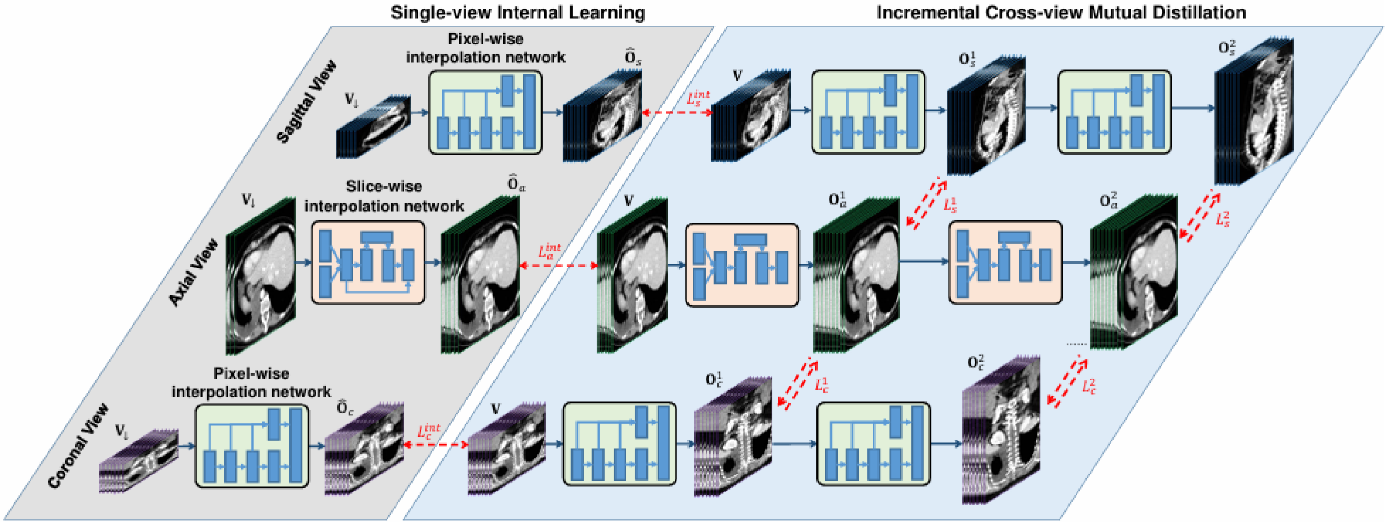

Figure 2.5: Pipeline of the methodology utilized by Fang *et al*., 2022 [44].

axis.

The methodologies utilizing 3D GANs are similar between each other, as they all apply networks based on the GANs implemented in 2D images. As this method already considers the information across all the axes simultaneously, there is no need to use multiple networks for each, as seen in some of the previous approaches. Therefore, the differences between papers mainly originate from the medical imaging technique to which it is applied, the modules that constitute the 3D GANs used, and the datasets used for evaluation [47, 48, 49, 50].

Chapter 3

Methods

The Methods section starts with an overview of the dataset selected for the fluid segmentation and intermediate slice synthesis tasks, while regarding the requirements needed for the training of each model and the reasoning behind the selection. Afterwards, it provides an explanation of the experiments that will be done during the dissertation, regarding fluid segmentation, inter-slice generation, and fluid volume estimation, while explaining the methodologies that will be implemented.

3.1 Dataset

The application of deep learning to fluid segmentation in OCT volumes requires a large number of images annotated with the three retinal fluids for the training process. The manual segmentation of large amounts of B-scans is a laborious process, which results in a shortage of publicly available annotated OCT datasets. Consequently, the majority of these datasets contain a limited quantity of images.

The dataset selected for this dissertation is the RETOUCH dataset [51]. This dataset consists of 112 OCT volumes, obtained with four different devices: 38 from the Cirrus HD-OCT (Zeiss Meditec), 38 from the Spectralis (Heidelberg Engineering), and 36 from the T-1000/T-2000 (Topcon). The 112 volumes are split into training (70 volumes) and testing (42 volumes). Only those in the training set have annotations of the retinal fluids (IRF, SRF, and PED). For the training and testing of the segmentation models, only the annotated volumes will be used.

From the 70 volumes, 24 were obtained with the Cirrus, 24 volumes were acquired with the Spectralis, and 22 were obtained with the two Topcon devices. The number of B-scans per volume, the dimensions of the B-scans, and the axial resolutions vary according to the device utilized to obtain the OCT. The volumes acquired using the Cirrus have 128 B-scans, while those obtained with Spectralis have 49 B-scans. The volumes acquired using the T-2000 have 128 B-scans and those obtained with the T-1000 have 64 B-scans. Of the 22 Topcon volumes, 20 were obtained using the T-2000 and 2 were acquired using the T-1000. In total, 6838 B-scans will be used on the train and test of the segmentation model.

When compared with other renown OCT datasets annotated with retinal fluid, such as the Duke dataset [52], the two datasets from the University of Minnesota [53, 54], and the Lu et al. [26] dataset, the RETOUCH presents a significantly larger quantity of annotated volumes. It also shows more variety since the volumes were obtained using four different devices instead of including volumes from just one device, as done in the mentioned datasets.

Table 3.1: Volumes, B-scans per volume, the total number of B-scans, and macular diseases in each dataset

	DUKE2015 [52]	UMN2017 [53]	UMN2018 [54]	LU2019 [26]	RETOUCH [51]
Volumes	10	24	29	528	70[a]
B-scans/Volume	11	25	25	Variable	128 (Cirrus and Topcon[b]), 64 (Topcon[b]), 49 (Spectralis)
B-scans	110	600	725	750	6838
Device	Spectralis	Spectralis	Spectralis	Spectralis	Cirrus, Topcon and Spectralis
Disease	DME	AMD	DME	DME	AMD and RVO

[a] 24 volumes from Cirrus, 22 volumes from Topcon, and 24 volumes from Spectralis.
[b] Volumes from Topcon were obtained using two different devices, resulting in volumes from the same vendor with different dimensions. Two of the training volumes have 64 slices.

For these reasons, the RETOUCH dataset is regarded as a diverse and large dataset, widely used in the literature that aims to perform fluid segmentation using deep learning, as done in [21, 22, 23, 25, 26, 28, 30, 31]. These aspects motivated the selection of the RETOUCH as the dataset that will be used for implementing the models for fluid segmentation in OCT volumes in this dissertation.

In intermediate slice synthesis, the 112 OCT volumes that constitute the RETOUCH dataset will be used for the training and evaluation of the models. The volumes that do not have segmentation masks can also be included since these masks are not necessary in the intermediate slice generation task.

The consistent number of slices per volume and large quantity of OCT volumes make the RETOUCH dataset suitable for the training and evaluation of the models developed to generate intermediate slices.

3.2 Experiments

In this subsection, the experiments that will be conducted during the dissertation are explained in depth. The subsection begins with a description of the data split, followed by the experiments in fluid segmentation and in fluid volume estimation.

3.2.1 Cross-validation

To promote consistency across all experiments, the conditions will be held identical. The train-test split will follow a 5-fold split. This split will be applied to the volumes used in fluid segmentation and will be common to all fluid segmentation experiments. Similarly, the 5-fold split will be applied to the volumes used in intermediate slice synthesis and the resulting split will be the same for all the experiments in slice generation.

The 5-fold split will be applied to each vendor (Cirrus, Spectralis, and Topcon) to ensure consistent vendor representation across all folds. The split will be applied to the volumes and not to the slices since the slices of the same volumes must be kept together to train the 2.5D segmentation model (3.2.2 Experiment 4).

In the beginning, each model will be trained using three folds. The fourth fold will be utilized for evaluating the effect of the changes done to the hyperparameters, and, when the best performing hyperparameters are found, the fifth fold will be dedicated to test the model. Afterwards, the training will be conducted using four folds and the testing will be done using one fold. Table 3.2 illustrates a potential 5-fold split in the RETOUCH dataset, for the training of the fluid segmentation models, in which 70 OCT volumes will be used. In the intermediate slice synthesis, the process will be the same but more volumes will be included.

Table 3.2: Number of OCT volumes per vendor in each fold, considering 5-fold validation.

	1st	**2nd**	**3rd**	**4th**	**5th**
Cirrus	5	5	5	5	4
Spectralis	5	5	5	5	4
Topcon	$4^a + 1^b$	$4^a + 1^b$	4^a	4^a	4^a

Volumes marked with ***a*** consist of 128 B-scans.

Volumes marked with ***b*** consist of 64 B-scans.

3.2.2 Fluid Segmentation

The initial experiments of this dissertation will focus on training networks on the fluid segmentation task. The goal of these experiments is to determine which segmentation network performs the best in the considered task, which will be later required for the fluid volume estimation.

The training will be performed on patches, following the same implementation as the one in Tennakoon et al. [55]. The extraction of patches aims at prioritizing the B-scan information relevant for the segmentation. To achieve this, the patches are not distributed uniformly. Instead, in each image, 12 patches are extracted from a random location inside the region of interest (ROI) and 2 patches are extracted from a random location outside the ROI. The image's ROI is the part of the image where the entropy is above a determined threshold or where there retinal fluid is present. The patches are then augmented by rotation and flipping.

In the first three experiments, the U-Net [35] will perform multi-class segmentation of the fluid regions in each B-scan. The U-Net is distinguished by its encoder-decoder structure, which resembles the letter U (see Figure 3.1). In the encoder path, two 3x3 unpadded convolutions are applied to the input image, with each being followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with a stride of 2, downsampling the image. In each downsampling step, the number of channels is doubled. In the expanding path, a 2x2 up-convolution is used, resulting in the halving of the number of channels. The result is then concatenated with the cropped feature map from the respective contracting path. A 1x1 convolution is applied to the final layer.

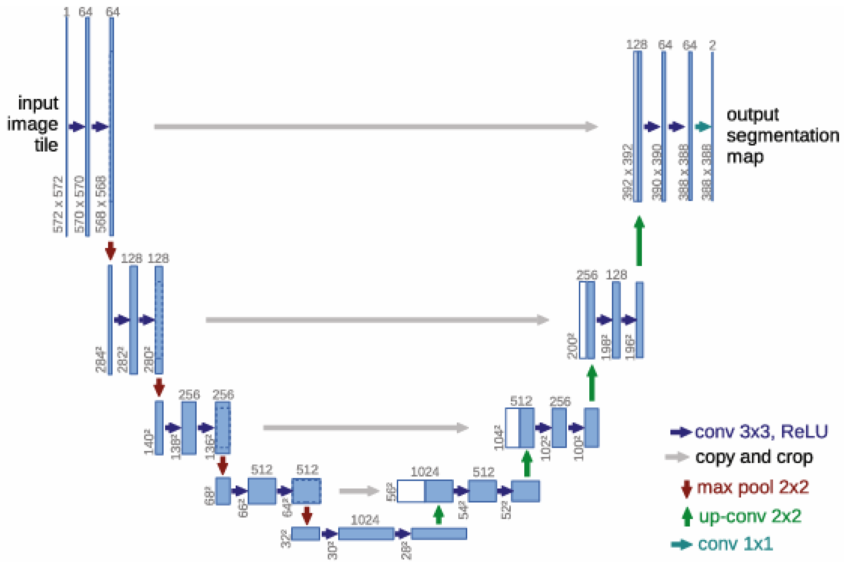

Figure 3.1: U-Net architecture [35]

The evaluation of all networks will be conducted using the Dice coefficient. The Dice coefficient is a commonly used metric for evaluating the similarity between two sets. In this context, it will be used for assessing the similarity between the segmentation mask generated by the segmentation network and the GT. The equation that describes the Dice coefficient can be seen in Equation 3.1, where A is a set that represents the GT binary mask of one fluid and B is another set that represents the predicted binary mask of the same fluid [56]. Considering a_i and b_i the binary pixels, the Dice coefficient can be rewritten as shown in Equation 3.2. The network that performs the best will be selected to estimate the fluid volumes in the fluid volume estimation experiments.

$$\text{Dice}(A,B) = \frac{2|A \cap B|}{|A| + |B|} \tag{3.1}$$

$$\text{Dice}(A,B) = \frac{2\sum_i a_i b_i}{\sum_i a_i + \sum_i b_i} \tag{3.2}$$

Experiment 1 In the first experiment, the base U-Net model, with no changes to the input, will be trained to perform 2D multi-class segmentation of the retinal fluids in OCT volumes. While this experiment is expected to perform worse than the following, it represents the baseline experiment, allowing for a comparison with the other fluid segmentation models that will be implemented.

Experiment 2 Similarly, in the second experiment, three U-Nets will be used for the segmentation task, one U-Net for each fluid type: one U-Net for the segmentation of IRF, one for SRF, and one for PED. This approach aligns with the methodology described in Rahil et al. [21] and Padilla-Pantoja et al. [32] and allows each network to specialize in the semantic segmentation of its fluid, improving the performance.

Experiment 3 The subsequent experiment will be a repetition of the two previous experiments with an additional input. In papers such as [21, 26, 31], a relative distance map was used as an additional input channel to the network (see Figure 2.3). This requires the initial delineation of the retina between the ILM and BM for which the retinal layer segmentation model developed by Melo et al. [57] will be utilized. Afterwards, the relative distance map will be extracted, considering only the mentioned layers, for each B-scan that compose the dataset. By providing information on the fluid location in the retina, this additional input significantly improved the performance of the segmentation models in [21, 26, 31].

Experiment 4 The final experiment in the fluid segmentation task is inspired by the framework of Tennakoon et al. [55]. In this paper, 2.5D multi-class fluid segmentation is performed and patches of three consecutive slices are considered in the training phase. In order to keep the conditions equal across the experiments, the preprocessing and post-processing used in the original paper will not be performed. The network utilized in this framework, shown in Figure 3.2, is similar to a U-Net, with the inclusion of batch normalization to improve training efficiency and dropout to prevent overfitting. With the inclusion of multiple slices in the input, it is expected an improvement on the segmentation.

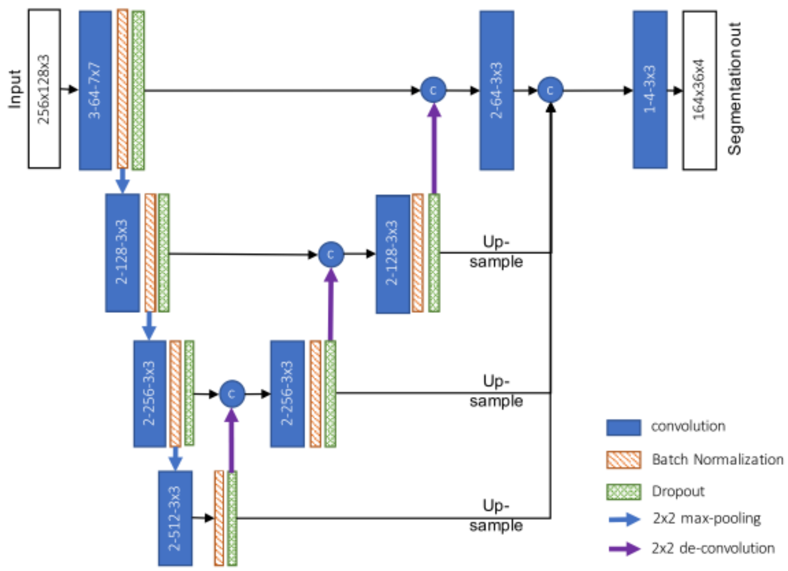

Figure 3.2: The segmentation network in [55].

3.2.3 Intermediate Slice Synthesis

The objective of the subsequent experiments is to improve the resolution between slices, thus approximating the fluid volume that will be estimated to the true value.

The intermediate slices will be generated using the RETOUCH dataset. In this experiment, subvolumes that consist of overlapping triplets of consecutive slices, sampled with a step size of 1, will be used. The first and the last slice of these triplets will be used for the generation of the middle slice. Consequently, it is possible to evaluate the generated slice in comparison to the original one, as done in other examples of the literature. For each volume, the number of potential subsets is then determined to be $n-2$, where n represents the number of slices within the same volume.

As proposed in the fluid segmentation experiments, the split between the train, test, and validation sets will remain consistent. In the slice generation experiments, since the entire slice is necessary for generation and not just the fraction that contains fluid, the B-scans will not be patched.

The generation of slices can be evaluated in specific metrics, as well as through qualitative assessment. To assess the efficacy of the generation model, the model utilized for fluid segmentation could be used for the estimation of the fluid's area in the generated image and to compare the resulting mask with the original image's mask. This comparison can be conducted using the Dice coefficient [13] (see Equation 3.2). However, this metric is insufficient for evaluating the generation performance, as it requires comparisons that encompass the entire slice and not just the fluid. Examples of such metrics include the mean absolute error (MAE) [13, 41, 50], the peak signal-to-noise ratio (PSNR) [38, 40, 43, 44, 45, 47, 48, 49, 50], and the structural similarity index measure (SSIM) [38, 43, 44, 45, 47, 48, 49, 50].

Experiment 5 In the first experiment focused on intermediate slice synthesis, a GAN will be used. The underlying principle of a GAN, originally proposed by Goodfellow et al. [58], is based on a competitive game between two networks. The generator network starts with the first and last slice of a subvolume, which is composed of three consecutive B-scans from an OCT scan, and aims to generate the intermediate slice. In contrast, the discriminator network is trained to distinguish between the generated and real slices. When the discriminator correctly labels generated slices as fake, the generator will be penalized, motivating it to fool the discriminator and consequently improving its generation, resulting in outputs more similar to the real inputs. However, the discriminator network loss also penalizes misclassifications, dependent on the probability of the prediction. As a result, as the generator improves, so does the discriminator [59]. The overall framework for GANs is illustrated in Figure 3.3.

Experiment 6 As in the previous experiment, the intermediate slice will be generated using the first and last slices of a subvolume that consists of three consecutive B-scans from an OCT scan. However, in this experiment, inspired by the work of Nishimoto et al. [42], the intermediate slice will be generated using a U-Net. While the U-Net is more commonly applied in segmentation, as

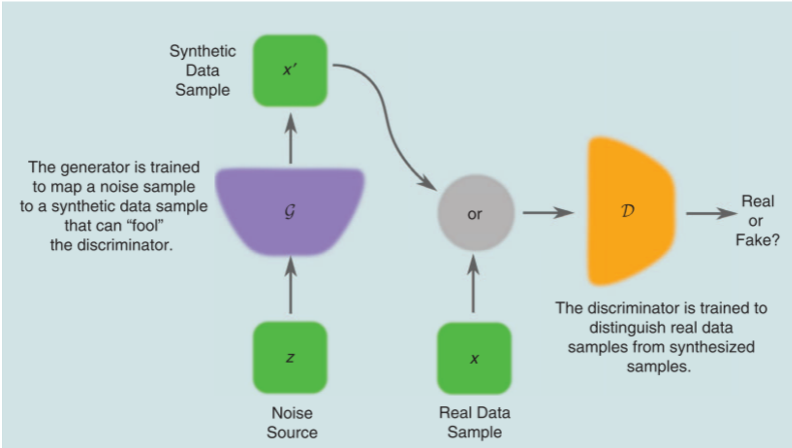

Figure 3.3: Example of a GAN framework, where $\mathscr{D}$ is the discriminator and $\mathscr{G}$ is the generator [60].

seen in the reviewed literature, Nishimoto et al. [42] apply it to generate the intermediate slices of a subvolume. The U-Net receives the edge slices as input and forces the output of the intermediate ones. In the paper [42], this was tested for three, four, and five slices. However, in this experiment it will be utilized to generate a single intermediate slice.

3.2.4 Fluid Volume Estimation

The estimation of fluid volume will be done once the optimal segmentation and intermediate slice generation models have been selected. Afterwards, the intermediate slice generation model will synthesize slices in the OCT volumes. The segmentation model will then be applied to both the unaltered volumes and the volumes with generated slices.

The area of each fluid in each OCT scan will be estimated considering the resolution of each OCT scan, which varies according to the device utilized to obtain the OCT volume. Afterwards, the area will be multiplied by the axial distance (half the axial distance to the previous slice plus half the axial distance to the following slice) to obtain the volume of fluid per slice. In the first and last slice of an OCT volume, the area will be multiplied by half of the axial distance (half the axial distance to the neighboring slice). The total volume of fluid in an OCT scan can be estimated by summing the fluid volumes of all individual B-scans. This allows the volume estimation of IRF, SRF, and PED, as well as the overall fluid volume in the OCT scan.

Experiment 7 In this experiment, the fluid volumes will be calculated for the OCT scans without the generated slices. The best segmentation model will segment the fluid in three classes and the volume will be estimated for each class as described. The results from this experiment will allow the comparison with the values obtained in the following experiment, where slice generation will be used.

Experiment 8 This experiment will consist of the fluid volume estimation in OCT scans with generated images. Similar to the previous experiment, the best segmentation model will predict the fluid masks and the fluid volume will be estimated. The values obtained will be compared with those calculated in the previous experiment. It is expected that the volumes will not vary more than one order of magnitude between experiments. In case a significant difference is observed, the generated images and their respective masks must be analyzed.

References

[1] F. Hutmacher, "Why Is There So Much More Research on Vision Than on Any Other Sensory Modality?" *Frontiers in Psychology*, vol. 10, 2019, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2019.02246. [Online]. Available: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.02246.

[2] H. Bogunović, W.-D. Vogl, S. M. Waldstein, and U. Schmidt-Erfurth, "Chapter 14 - OCT fluid detection and quantification," in E. Trucco, T. MacGillivray, and Y. Xu, Eds. Academic Press, 2019, pp. 273–298, ISBN: 978-0-08-102816-2. DOI: https://doi.org/10.1016/B978-0-08-102816-2.00015-0. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081028162000150.

[3] L. S. Lim, P. Mitchell, J. M. Seddon, F. G. Holz, and T. Y. Wong, "Age-related macular degeneration," *The Lancet*, vol. 379, pp. 1728–1738, 9827 2012, ISSN: 0140-6736. DOI: https://doi.org/10.1016/S0140-6736(12)60282-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140673612602827.

[4] P. Mitchell, G. Liew, B. Gopinath, and T. Y. Wong, "Age-related macular degeneration," *The Lancet*, vol. 392, pp. 1147–1159, 10153 Sep. 2018, ISSN: 0140-6736. DOI: 10.1016/S0140-6736(18)31550-2. [Online]. Available: https://doi.org/10.1016/S0140-6736(18)31550-2.

[5] O. Musat et al., "Diabetic Macular Edema," *Rom J Ophthalmol*, vol. 59, pp. 133–136, 3 Jul. 2015.

[6] N. Bhagat, R. A. Grigorian, A. Tutela, and M. A. Zarbin, "Diabetic Macular Edema: Pathogenesis and Treatment," *Survey of Ophthalmology*, vol. 54, pp. 1–32, 1 2009, ISSN: 0039-6257. DOI: https://doi.org/10.1016/j.survophthal.2008.10.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0039625708001811.

[7] F. Bandello, R. Lattanzio, I. Zucchiatti, A. Arrigo, M. Battista, and M. V. Cicinelli, "Diabetic Macular Edema," in M. Attilio, L. Rosangela, Z. I. B. Francesco, and Zarbin, Eds. Springer International Publishing, 2019, pp. 97–183, ISBN: 978-3-319-96157-6. DOI: 10.1007/978-3-319-96157-6_3. [Online]. Available: https://doi.org/10.1007/978-3-319-96157-6_3.

[8] T. Y. Wong and I. U. Scott, "Retinal-Vein Occlusion," *New England Journal of Medicine*, vol. 363, pp. 2135–2144, 22 2010. DOI: 10.1056/NEJMcp1003934. [Online]. Available: https://www.nejm.org/doi/full/10.1056/NEJMcp1003934.

[9] D. Huang et al., "Optical coherence tomography," *Science*, vol. 254, pp. 1178–1181, 5035 Nov. 1991.

[10] W. Drexler and J. G. Fujimoto, "State-of-the-art retinal optical coherence tomography," *Progress in Retinal and Eye Research*, vol. 27, pp. 45–88, 1 2008, ISSN: 1350-9462. DOI: https://doi.org/10.1016/j.preteyeres.2007.07.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350946207000444.

[11] I. A. Viedma, D. Alonso-Caneiro, S. A. Read, and M. J. Collins, "Deep learning in retinal optical coherence tomography (OCT): A comprehensive survey," *Neurocomputing*, vol. 507, pp. 247–264, 2022, ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2022.08.021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523122200995X.

[12] M. Almonte, P. Capellàn, T. Yap, and M. F. Cordeiro, "Retinal correlates of psychiatric disorders," *Therapeutic Advances in Chronic Disease*, vol. 11, p. 204 062 232 090 521, Dec. 2020. DOI: 10.1177/2040622320905215.

[13] E. López-Varela, N. Barreira, N. O. Pascual, M. R. A. Castillo, and M. G. Penedo, "Generation of synthetic intermediate slices in 3D OCT cubes for improving pathology detection and monitoring," *Computers in Biology and Medicine*, vol. 163, p. 107 214, 2023, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2023.107214. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523006790.

[14] E. Selvi, M. Özdemir, and M. A. Selver, "Performance Analysis of Distance Transform Based Inter-Slice Similarity Information on Segmentation of Medical Image Series," *Mathematical and Computational Applications*, vol. 18, pp. 511–520, 3 2013, ISSN: 2297-8747. DOI: 10.3390/mca18030511. [Online]. Available: https://www.mdpi.com/2297-8747/18/3/511.

[15] T. C. Quek et al., "Predictive, preventive, and personalized management of retinal fluid via computer-aided detection app for optical coherence tomography scans," *EPMA Journal*, vol. 13, pp. 547–560, 4 2022, ISSN: 1878-5085. DOI: 10.1007/s13167-022-00301-5. [Online]. Available: https://doi.org/10.1007/s13167-022-00301-5.

[16] S. J. Pawan et al., "Capsule Network–based architectures for the segmentation of sub-retinal serous fluid in optical coherence tomography images of central serous chorioretinopathy," *Medical & Biological Engineering & Computing*, vol. 59, pp. 1245–1259, 6 2021, ISSN: 1741-0444. DOI: 10.1007/s11517-021-02364-4. [Online]. Available: https://doi.org/10.1007/s11517-021-02364-4.

[17] X. Liu, S. Wang, Y. Zhang, D. Liu, and W. Hu, "Automatic fluid segmentation in retinal optical coherence tomography images using attention based deep learning," *Neurocomputing*, vol. 452, pp. 576–591, 2021, ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2020.07.143. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231220319135.

[18] Y. Guo, T. T. Hormel, H. Xiong, J. Wang, T. S. Hwang, and Y. Jia, "Automated Segmentation of Retinal Fluid Volumes From Structural and Angiographic Optical Coherence Tomography Using Deep Learning," *Translational Vision Science & Technology*, vol. 9, p. 54, 2 Oct. 2020, ISSN: 2164-2591. DOI: 10.1167/tvst.9.2.54. [Online]. Available: https://doi.org/10.1167/tvst.9.2.54.

[19] Z. Wang et al., "Automated segmentation of macular edema for the diagnosis of ocular disease using deep learning method," *Scientific Reports*, vol. 11, p. 13 392, 1 2021, ISSN: 2045-2322. DOI: 10.1038/s41598-021-92458-8. [Online]. Available: https://doi.org/10.1038/s41598-021-92458-8.

[20] Y. Wu et al., "Training Deep Learning Models to Work on Multiple Devices by Cross-Domain Learning with No Additional Annotations," *Ophthalmology*, vol. 130, pp. 213–222, 2 2023, ISSN: 0161-6420. DOI: https://doi.org/10.1016/j.ophtha.2022.09.014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0161642022007497.

[21] M. Rahil, B. N. Anoop, G. N. Girish, A. R. Kothari, S. G. Koolagudi, and J. Rajan, "A Deep Ensemble Learning-Based CNN Architecture for Multiclass Retinal Fluid Segmentation in OCT Images," *IEEE Access*, vol. 11, pp. 17 241–17 251, 2023, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2023.3244922.

[22] H. Liu et al., "Semantic uncertainty Guided Cross-Transformer for enhanced macular edema segmentation in OCT images," *Computers in Biology and Medicine*, vol. 174, p. 108 458, 2024, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2024.108458. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482524005420.

[23] X. Li, S. Niu, X. Gao, X. Zhou, J. Dong, and H. Zhao, "Self-training adversarial learning for cross-domain retinal OCT fluid segmentation," *Computers in Biology and Medicine*, vol. 155, p. 106 650, 2023, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2023.106650. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523001154.

[24] K. Gao et al., "Double-branched and area-constraint fully convolutional networks for automated serous retinal detachment segmentation in SD-OCT images," *Computer Methods and Programs in Biomedicine*, vol. 176, pp. 69–80, 2019, ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2019.04.027. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260718318327.

[25] B. Hassan et al., "Deep learning based joint segmentation and characterization of multi-class retinal fluid lesions on OCT scans for clinical use in anti-VEGF therapy," *Computers in Biology and Medicine*, vol. 136, p. 104 727, 2021, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2021.104727. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521005217.

[26] D. Lu et al., "Deep-learning based multiclass retinal fluid segmentation and detection in optical coherence tomography images using a fully convolutional neural network," *Medical Image Analysis*, vol. 54, pp. 100–110, 2019, ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2019.02.011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841519300167.

[27] B. Hassan, S. Qin, T. Hassan, R. Ahmed, and N. Werghi, "Joint Segmentation and Quantification of Chorioretinal Biomarkers in Optical Coherence Tomography Scans: A Deep Learning Approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–17, 2021. DOI: 10.1109/TIM.2021.3077988.

[28] H. Zhang, J. Yang, C. Zheng, S. Zhao, and A. Zhang, "Annotation-efficient learning for OCT segmentation," *Biomed. Opt. Express*, vol. 14, pp. 3294–3307, 7 Jul. 2023. DOI: 10.1364/BOE.486276. [Online]. Available: https://opg.optica.org/boe/abstract.cfm?URI=boe-14-7-3294.

[29] L. B. Sappa et al., "RetFluidNet: Retinal Fluid Segmentation for SD-OCT Images Using Convolutional Neural Network," *Journal of Digital Imaging*, vol. 34, pp. 691–704, 3 2021, ISSN: 1618-727X. DOI: 10.1007/s10278-021-00459-w. [Online]. Available: https://doi.org/10.1007/s10278-021-00459-w.

[30] G. Xing et al., "Multi-Scale Pathological Fluid Segmentation in OCT With a Novel Curvature Loss in Convolutional Neural Network," *IEEE Transactions on Medical Imaging*, vol. 41, pp. 1547–1559, 6 2022. DOI: 10.1109/TMI.2022.3142048.

[31] W. Tang et al., "Multi-class retinal fluid joint segmentation based on cascaded convolutional neural networks," *Physics in Medicine & Biology*, vol. 67, p. 125 018, 12 Jun. 2022. DOI: 10.1088/1361-6560/ac7378. [Online]. Available: https://dx.doi.org/10.1088/1361-6560/ac7378.

[32] F. D. Padilla-Pantoja, Y. D. Sanchez, B. A. Quijano-Nieto, O. J. Perdomo, and F. A. Gonzalez, "Etiology of Macular Edema Defined by Deep Learning in Optical Coherence Tomography Scans," *Translational Vision Science & Technology*, vol. 11, p. 29, 9 Sep. 2022, ISSN: 2164-2591. DOI: 10.1167/tvst.11.9.29. [Online]. Available: https://doi.org/10.1167/tvst.11.9.29.

[33] J. Hu, Y. Chen, and Z. Yi, "Automated segmentation of macular edema in OCT using deep neural networks," *Medical Image Analysis*, vol. 55, pp. 216–227, 2019, ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2019.05.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841519300386.

[34] I. Mantel et al., "Automated Quantification of Pathological Fluids in Neovascular Age-Related Macular Degeneration, and Its Repeatability Using Deep Learning," *Translational Vision Science & Technology*, vol. 10, p. 17, 4 Apr. 2021, ISSN: 2164-2591. DOI: 10.1167/tvst.10.4.17. [Online]. Available: https://doi.org/10.1167/tvst.10.4.17.

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Springer International Publishing, 2015, pp. 234–241, ISBN: 978-3-319-24574-4.

[36] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, pp. 834–848, 4 2018. DOI: 10.1109/TPAMI.2017.2699184.

[37] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv e-prints*, arXiv:1409.1556, Sep. 2014. DOI: 10.48550/arXiv.1409.1556.

[38] C. You et al., "CT Super-Resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE)," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 188–203, 1 2020. DOI: 10.1109/TMI.2019.2922960.

[39] M. Ibrahim et al., "Generative AI for Synthetic Data Across Multiple Medical Modalities: A Systematic Review of Recent Developments and Challenges," *arXiv e-prints*, arXiv:2407.00116, Jun. 2024. DOI: 10.48550/arXiv.2407.00116.

[40] Y. Xia, N. Ravikumar, J. P. Greenwood, S. Neubauer, S. E. Petersen, and A. F. Frangi, "Super-Resolution of Cardiac MR Cine Imaging using Conditional GANs and Unsupervised Transfer Learning," *Medical Image Analysis*, vol. 71, p. 102 037, 2021, ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2021.102037. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841521000839.

[41] Z. Wu, J. Wei, J. Wang, and R. Li, "Slice imputation: Multiple intermediate slices interpolation for anisotropic 3D medical image segmentation," *Computers in Biology and Medicine*, vol. 147, p. 105 667, 2022, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2022.105667. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522004565.

[42] S. Nishimoto, K. Kawai, K. Nakajima, H. Ishise, and M. Kakibuchi, "Generating intermediate slices with U-nets in craniofacial CT images," *medRxiv*, p. 2024.05.08.24307089, Jan. 2024. DOI: 10.1101/2024.05.08.24307089. [Online]. Available: http://medrxiv.org/content/early/2024/05/09/2024.05.08.24307089.abstract.

[43] H. Zhang, X. Yang, Y. Cui, Q. Wang, J. Zhao, and D. Li, "A novel GAN-based three-axis mutually supervised super-resolution reconstruction method for rectal cancer MR image," *Computer Methods and Programs in Biomedicine*, vol. 257, p. 108 426, 2024, ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2024.108426. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016926072400419X.

[44] C. Fang, L. Wang, D. Zhang, J. Xu, Y. Yuan, and J. Han, "Incremental Cross-View Mutual Distillation for Self-Supervised Medical CT Synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 20 677–20 686.

[45] U. Nimitha and P. Ameer, "MRI super-resolution using similarity distance and multi-scale receptive field based feature fusion GAN and pre-trained slice interpolation network," *Magnetic Resonance Imaging*, vol. 110, pp. 195–209, 2024, ISSN: 0730-725X. DOI: https://doi.org/10.1016/j.mri.2024.04.021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0730725X24001346.

[46] M.-I. Georgescu, R. T. Ionescu, and N. Verga, "Convolutional Neural Networks With Intermediate Loss for 3D Super-Resolution of CT and MRI Scans," *IEEE Access*, vol. 8, pp. 49 112–49 124, 2020. DOI: 10.1109/ACCESS.2020.2980266.

[47] C. Yuhua, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, "Efficient and Accurate MRI Super-Resolution Using a Generative Adversarial Network and 3D Multi-level Densely Connected Network," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, A. F. Frangi, S. Julia A., C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds., Springer International Publishing, 2018, pp. 91–99, ISBN: 978-3-030-00928-1.

[48] I. Sanchez and V. Vilaplana, "Brain MRI super-resolution using 3D generative adversarial networks," *arXiv e-prints*, arXiv:1812.11440, Dec. 2018. DOI: 10.48550/arXiv.1812.11440.

[49] A. Kudo, Y. Kitamura, Y. Li, S. Iizuka, and E. Simo-Serra, "Virtual Thin Slice: 3D Conditional GAN-based Super-Resolution for CT Slice Interval," in *Machine Learning for Medical Image Reconstruction*, Andreas, R. Daniel, Y. J. C. K. Florian, and Maier, Eds., Springer International Publishing, 2019, pp. 91–100, ISBN: 978-3-030-33843-5.

[50] K. Zhang et al., "SOUP-GAN: Super-Resolution MRI Using Generative Adversarial Networks," *Tomography*, vol. 8, pp. 905–919, 2 2022, ISSN: 2379-139X. DOI: 10.3390/tomography8020073. [Online]. Available: https://www.mdpi.com/2379-139X/8/2/73.

[51] H. Bogunović et al., "RETOUCH: The Retinal OCT Fluid Detection and Segmentation Benchmark and Challenge," *IEEE Transactions on Medical Imaging*, vol. 38, pp. 1858–1874, 8 2019. DOI: 10.1109/TMI.2019.2901398.

[52] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomed. Opt. Express*, vol. 6, pp. 1172–1194, 4 Apr. 2015. DOI: 10.1364/BOE.6.001172. [Online]. Available: https://opg.optica.org/boe/abstract.cfm?URI=boe-6-4-1172.

[53] A. Rashno et al., "Fully-automated segmentation of fluid regions in exudative age-related macular degeneration subjects: Kernel graph cut in neutrosophic domain," *PLOS ONE*, vol. 12, pp. 1–26, 10 Dec. 2017. DOI: 10.1371/journal.pone.0186949. [Online]. Available: https://doi.org/10.1371/journal.pone.0186949.

[54] A. Rashno et al., "Fully Automated Segmentation of Fluid/Cyst Regions in Optical Coherence Tomography Images With Diabetic Macular Edema Using Neutrosophic Sets and Graph Algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 65, pp. 989–1001, 5 2018. DOI: 10.1109/TBME.2017.2734058.

[55] R. Tennakoon, A. K. Gostar, R. Hoseinnezhad, and A. Bab-Hadiashar, "Retinal fluid segmentation in OCT images using adversarial loss based convolutional neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 1436–1440. DOI: 10.1109/ISBI.2018.8363842.

[56] R. R. Shamir, Y. Duchin, J. Kim, G. Sapiro, and N. Harel, "Continuous Dice Coefficient: a Method for Evaluating Probabilistic Segmentations," *arXiv e-prints*, arXiv:1906.11031, Jun. 2019. DOI: 10.48550/arXiv.1906.11031.

[57] T. Melo, Â. Carneiro, A. Campilho, and A. M. Mendonça, "Retinal layer and fluid segmentation in optical coherence tomography images using a hierarchical framework," *Journal of Medical Imaging*, vol. 10, p. 14 006, 1 2023. DOI: 10.1117/1.JMI.10.1.014006. [Online]. Available: https://doi.org/10.1117/1.JMI.10.1.014006.

[58] I. Goodfellow et al., "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., vol. 27, Curran Associates, Inc., 2014. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

[59] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, pp. 139–144, 11 Oct. 2020, ISSN: 0001-0782. DOI: 10.1145/3422622. [Online]. Available: https://doi.org/10.1145/3422622.

[60] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, vol. 35, pp. 53–65, 1 2018. DOI: 10.1109/MSP.2017.2765202.

Appendix A

Lorem Ipsum