Generation of synthetic intermediate slices in 3D OCT cubes for improving pathology detection and monitoring

Emilio López-Varela [a,b,*], Noelia Barreira [a,b], Nuria Olivier Pascual [c], Maria Rosa Arroyo Castillo [c], Manuel G. Penedo [a,b]

[a] VARPA Group, Biomedical Research Institute of A Coruña (INIBIC), University of A Coruña, A Coruña, Spain
[b] CITIC-Research Center of Information and Communication Technologies, University of A Coruña, A Coruña, Spain
[c] Servizo de Oftalmoloxía, Complejo Hospitalario Universitario de Ferrol (CHUF), Av. da Residencia S/N, Ferrol 15405, Spain

ARTICLE INFO

ABSTRACT

OCT is a non-invasive imaging technique commonly used to obtain 3D volumes of the ocular structure. These volumes allow the monitoring of ocular and systemic diseases through the observation of subtle changes in the different structures present in the eye. In order to observe these changes it is essential that the OCT volumes have a high resolution in all axes, but unfortunately there is an inverse relationship between the quality of the OCT images and the number of slices of the cube. This results in routine clinical examinations using cubes that generally contain high-resolution images with few slices. This lack of slices complicates the monitoring of changes in the retina hindering the diagnostic process and reducing the effectiveness of 3D visualizations. Therefore, increasing the cross-sectional resolution of OCT cubes would improve the visualization of these changes aiding the clinician in the diagnostic process. In this work we present a novel fully automatic methodology to perform the synthesis of intermediate slices of OCT image volumes in an unsupervised manner. To perform this synthesis, we propose a fully convolutional neural network architecture that uses information from two adjacent slices to generate the intermediate synthetic slice. We also propose a training methodology, where we use three adjacent slices to train the network by contrastive learning and image reconstruction. We test our methodology with three different types of OCT volumes commonly used in the clinical setting and validate the quality of the synthetic slices created with several medical experts and using an expert system.

1. Introduction

Medical imaging plays a fundamental role in the diagnosis and monitoring of a large number of diseases today, as it reveals the internal anatomy of the body and allows quantitative biomarkers to be obtained. In many cases, the quality of the images determines the quality of the medical diagnosis. Adequate visualization of the anatomy of the body, both structurally and functionally, is therefore indispensable for clinicians to be able to carry out their work efficiently. For this purpose, various imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT) or optical coherence tomography (OCT) are available Mahesh [1], Zhang et al. [2].

OCT is a non-invasive imaging technique with a resolution of microns that allows light-scattering biological tissues to be scanned in vivo. This technique is commonly used to obtain 3D volumes of the ocular structure where the different layers that compose the eye are represented Medeiros et al. [3]. This volume is used to diagnose and

monitor both ocular pathologies, such as central serous chorioretinopathy [4–6], diabetic macular edema [7] or age-related macular degeneration [8], and systemic diseases such as hypertension or multiple sclerosis (MS) [9–11], a chronic neurological disease of the central nervous system [12]. The progression of these pathologies produces different effects that have a very negative impact on the lives of the patients. For example, the progression of age-related macular degeneration produces an irreversible loss of visual capacity [13], while MS causes a progressive loss of mobility in general [14]. This makes early diagnosis and accurate monitoring of the progression of these diseases vitally important.

In order to make a good diagnosis of these diseases using OCT images, it is essential that the 3D OCT volumes have a good resolution in each axis. This factor helps the clinician to make more accurate and robust decisions, being able to observe more clearly the subtle changes that affect the diagnosis both at the slice level and in a 3D

reconstruction. The problem is that due to the OCT capture mechanism, there is an inverse relationship between image quality and the number of volume slices. Although there are several differences between the multiple acquisition devices, the general process is based on three steps. First, all information from a single one-dimensional scan of the tissue (A-scan) is acquired simultaneously by evaluating the frequency spectrum of the reflected light. Second, a two-dimensional image (B-scan) is obtained by laterally combining a series of A-scans. Finally, the device repeats this process by obtaining a series of B-scans or slices that are finally combined to form the final 3D volume [8,15]. If this process is performed too slowly, it leads to errors caused by the movement of the patient, the fluid or the structures of the eye. It is therefore not possible to increase the total acquisition time without increasing errors of this type. To increase the quality of the images (B-scans), the acquisition time of each slice must be increased, so that fewer slices can be taken with a greater separation between them or covering a smaller surface area. On the other hand, to increase the number of slices, the acquisition time for each slice must be reduced, thus reducing the quality and definition of each slice. In Fig. 1, macular OCT volumes with different number of slices obtained by various configurations are shown. It can be seen that in general, the more slices the worse the image quality, while the fewer slices the worse the cross-sectional resolution.

For all these reasons, volumes with different characteristics are used in the clinical setting depending on the purpose of the study. Many routine clinical studies are performed using volumes with high-resolution images but few slices, which make it very difficult to create a 3D visualization and are very inefficient in monitoring changes in the three-dimensional structure of the different layers and vessels. It is therefore of great interest to create a system capable of synthesizing the intermediate images between adjacent slices in order to increase the cross-sectional resolution of the volume, and thus create visualizations that allow clinicians to estimate pathological changes in a robust and effective way.

Intermediate slice synthesis has some similarity with frame interpolation in a video as both use a sequence of continuous images. Most differences between consecutive frames are caused by object or camera movements, so several works have used the optical flow field to synthesize intermediate frames [16,17]. In contrast, slices of a volume can have different content and complex distortions, which makes this task much more difficult to solve. Recently, some state of the art works have attempted to solve slice synthesis in different medical imaging modalities using neural networks such as generative adversarial networks (GANs). To the best of our knowledge, there are no OCT works on intermediate slice synthesis. The OCT papers most similar to this thematic are those about image generation through segmentation masks [18] or papers that focus on the conversion from one type of OCT image to another Gende et al. [19] and therefore involve style transfer. These works deal with 2D image generation and methodologically have nothing to do with the synthesis of an intermediate slice, so we cannot use them for comparison.

Although there are some works using MRI and CT image volumes in the synthesis area. These works attempt to transform low resolution image volumes to high resolution image volumes. In the work of Chen et al. [20] a multi-level densely connected super-resolution network was used to recover HR details from LR MRI images. In a similar way, Sánchez and Vilaplana [21] proposed to use a 3D GAN to exploit volumetric information to generate super-resolution brain MRI. In the work of You et al. [22] a semi-supervised deep learning approach (GAN) was used to accurately recover HR CT images. Also in CT, Kudo et al. [23] proposed a novel architecture based on conditional (GANs) with the goal of generating HR images of the different body parts. Finally, Peng et al. [24] proposed a two-stage framework to solve the slice synthesis task. It first employs 2D convolutional neural networks (CNNs) to enlarge sagittal and coronal images individually, and then fuse the enlarged images of two views to produce the final result. The main drawback of using this 3D neural network is the huge amount of network parameters and memory consumption. A major limitation of previous works is that the presented neural networks require a large amount of paired LR and HR image volumes. This complicates the task quite a lot as these volumes are usually not available in routine medical practice. Thus, it is essential to develop unsupervised optimization algorithms for medical slice synthesis. Because of this, Fang et al. [25] explored a self-supervised learning framework to train the slice synthesizer without the ground-truth data. For this purpose, the originals HR volumes were downsampled so that the LR versions of the volumes used as inputs were obtained. This type of methodology works fine when using imaging modalities such as MRI or CT where large HR volumes are involved. In contrast, as previously explained, in the OCT imaging modality there is a general inverse relationship between image quality and number of slices. Thus, volumes with fewer slices are those with generally higher image quality. This makes it impossible to adapt any of the methods used in other imaging modalities, as an unsupervised approach that extends the number of slices of the volume without ground-truth data is needed.

Given the significant gap that exists in the literature regarding the synthesis of intermediate slices in OCT image volumes, and taking into account the huge potential for improving the visualizations used by clinicians to diagnose and monitor different pathologies such as MS, in this work we propose a novel fully automatic methodology to perform the synthesis of intermediate slices of OCT image volumes in an unsupervised manner. This allows us to create quality image cubes with a small inter-slice distance, improving both slice visualization and 3D reconstruction. For this purpose, we propose a fully convolutional neural network architecture that uses the information from two slices to generate the intermediate synthetic slice. We also propose a training methodology that allows us to train the network to generate these intermediate slices (which do not exist) in an unsupervised way using contrastive learning and image reconstruction. We test our methodology with three different types of OCT volumes commonly used in the clinical setting and validate the quality of the synthetic slices created with several medical experts and using a expert system.

To the best of our knowledge, this is the only work that presents a fully automatic methodology specifically dedicated to increase the number of slices of OCT cubes in order to improve the different visualizations used by medical staff. Our methodology is robust, allowing better diagnosis and monitoring of patients with different diseases, such as, patients affected by multiple sclerosis.

This manuscript is organized as follows: "**Materials**" presents all the resources needed to fully reproduce our work. "**Methodology**" presents a detailed explanation of the proposed approach. "**Results and Discussion**" includes the results and their subsequent discussion, as well as the main challenges faced in this work. Finally, "**Conclusions**" presents the general conclusions of the study and possible lines of future work.

2. Materials

In this section, we describe the OCT dataset, as well as the software and hardware resources that were used in this work. Both resources are described in detail below.

2.1. Dataset

In this work the dataset used is composed of a total of 42 OCT cubes obtained using two equivalent Spectralis R OCT capture devices from Heidelberg Engineering. Within these volumes there are three different types of OCT cubes extracted in the macular region and captured using different configurations commonly used in the clinical setting. The "Fast" type comprises 20 of the 42 cubes and has 25 image slices with a resolution of 496×512. The "Dense" type comprises 20 of the 42 cubes and has 49 image slices with a resolution of 496×512. Finally, the
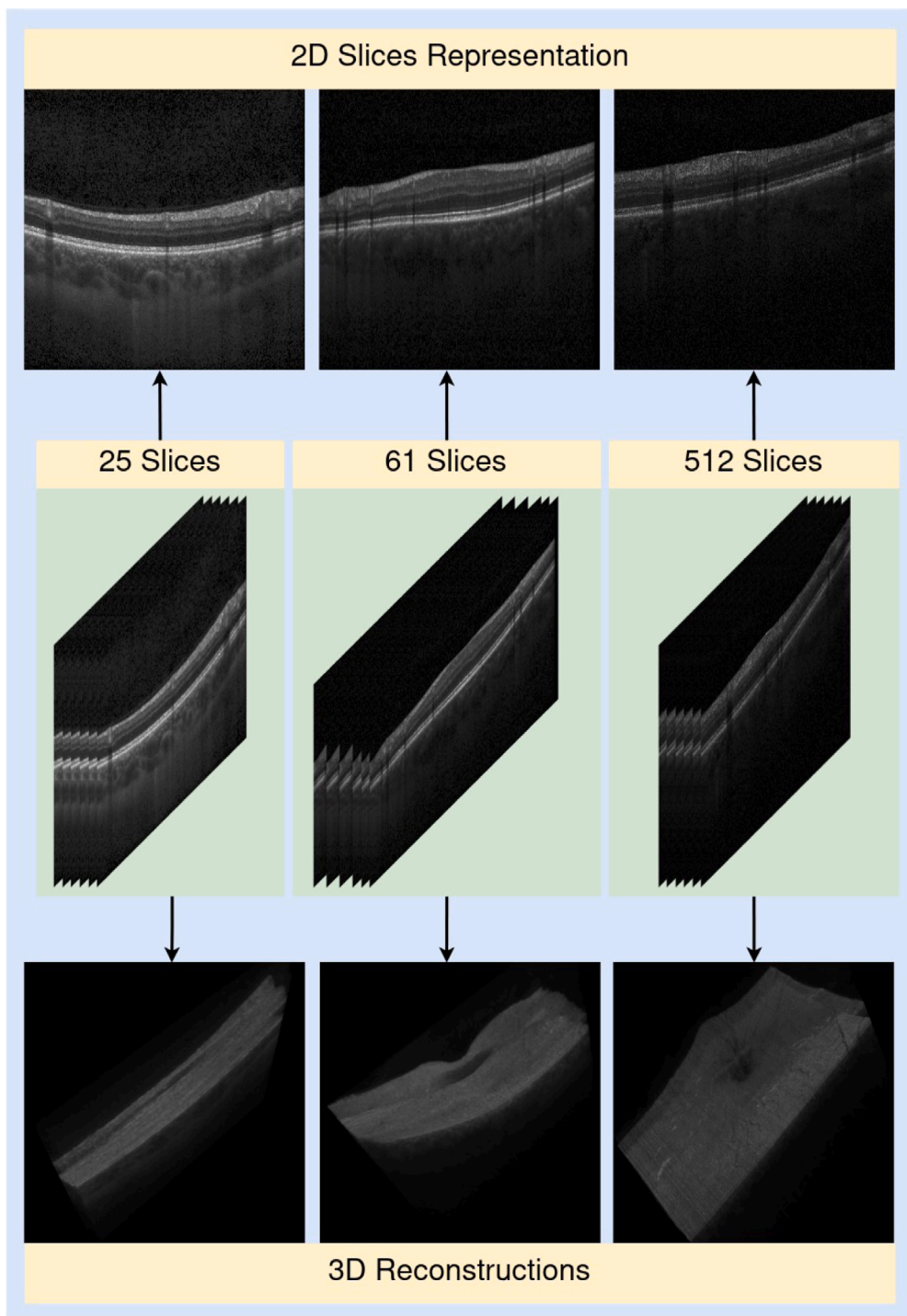
Fig. 1. Representative examples of different types of OCT cubes. The 2D representation of a single image and a 3D reconstruction of the volume are shown.

"Posterior Pole" type spans 2 of the 42 cubes and has 61 image slices with a resolution of 496×768. These three types of cubes are used to test whether our network is able to generate the intermediate synthetic slices in a robust and efficient way regardless of the distance between the two adjacent slices used. All the images were resized to 496×496. In Fig. 2, we can see an example of 3 consecutive slices for each type of OCT cube. This amount of volumes as a whole is comparable to other datasets in the public domain as [26], so it is significant enough to carry out the experiments. Furthermore, unlike other datasets, our OCT volumes have a large variability of resolutions and number of slices, which makes them perfect for testing our slice generation methodology.

2.2. Software and hardware resources

Regarding the software resources, we used Python 3.9.5 with Py-Torch 1.9.1 and cuda 11.2 [27]. OpenCV (4.5.3) [28] and Numpy (1.20.3) were used to perform image processing operations. In addition, to test and evaluate the usefulness of our proposal we used a 3D reconstruction algorithm [8] and a segmentation network trained to segment the choroid [29]. Regarding hardware resources, we include in Table 1 the full disclosure of the components, drivers and software that have been used throughout the development of this work and that may have influenced its computational performance.
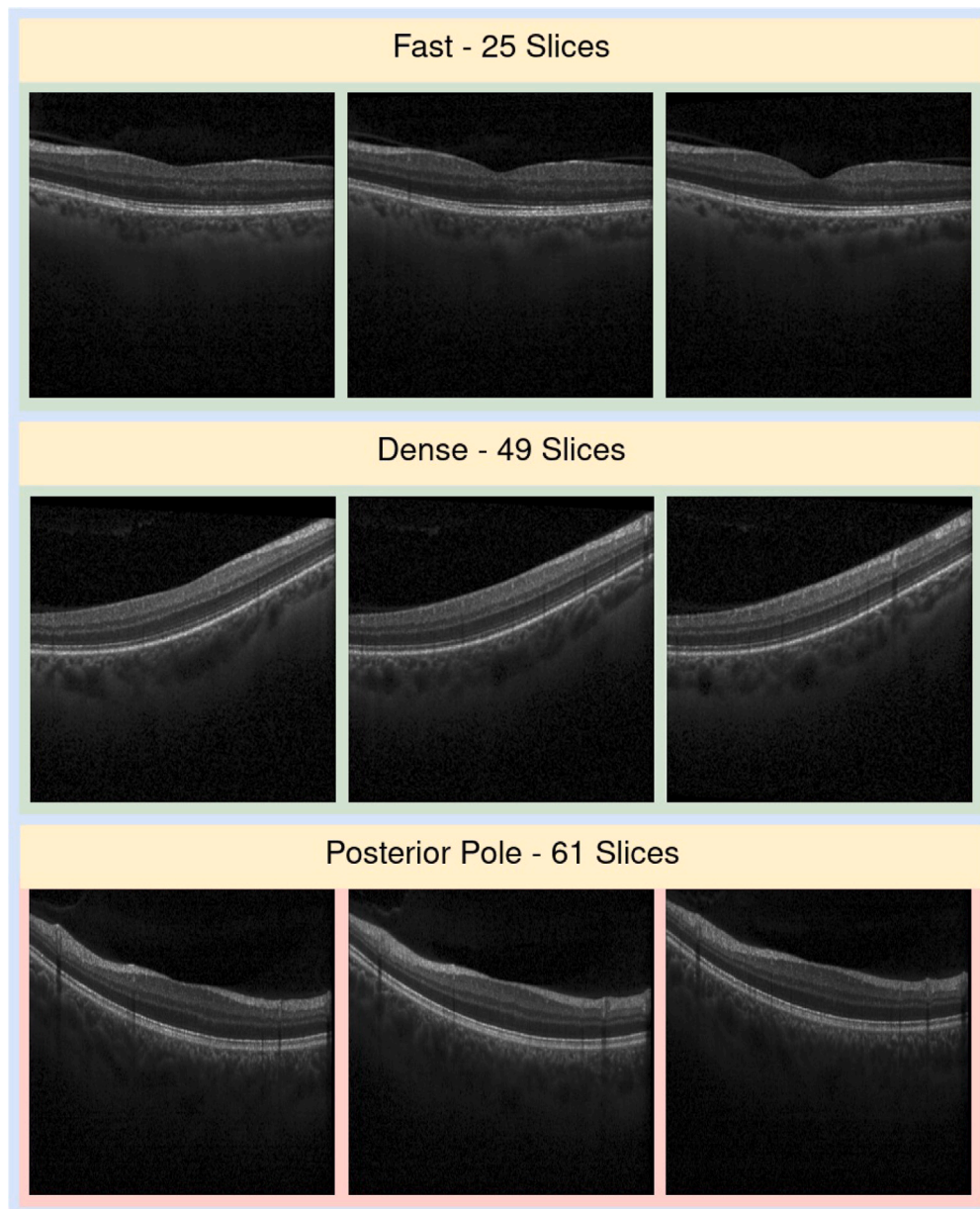
Fig. 2. Example of three consecutive slices for the cube types used.

Table 1
Specifications of the equipment that was used during this work.

Name	Description
OS	Ubuntu 20.04.1 LTS (Focal Fossa)
Kernel	Linux 5.4.0-81-generic
Architecture	x86-64
CPU	Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20 GHz
GPU	NVIDIA Corporation Tesla M60
Driver Version	460.91.03
CUDA Version	11.2

3. Methodology

We show below all the methodological details of our proposal. In order to generate image-realistic intermediate slices consistent with the continuity of the cube, we have developed a neural architecture that we use as a generator and a specific training methodology where we use two smaller networks as discriminators. This neural architecture is a derivation of the ResNet [30] adapted to perform slice synthesis and basically works as the generator of a GAN. Our training methodology is based on using three adjacent slices to generate intermediate slices with which to reconstruct the central slice. We use contrast learning [31] to force the reconstructed images and indirectly the intermediate images to have similar content to the original slices.

3.1. Network architecture

In this work we use three different neural networks consisting of a generator and two discriminators. In Fig. 3, a schematic showing the main parts of the three neural networks used can be seen.

- The main network is a **generator** that uses two adjacent slices as input to generate the intermediate synthetic slice. In this network InstanceNorm layers are used to avoid the problems of using BatchNorm layers in the generators. These problems result from the interdependency introduced into the elements within the minibatches in batch normalization. This interdependency can
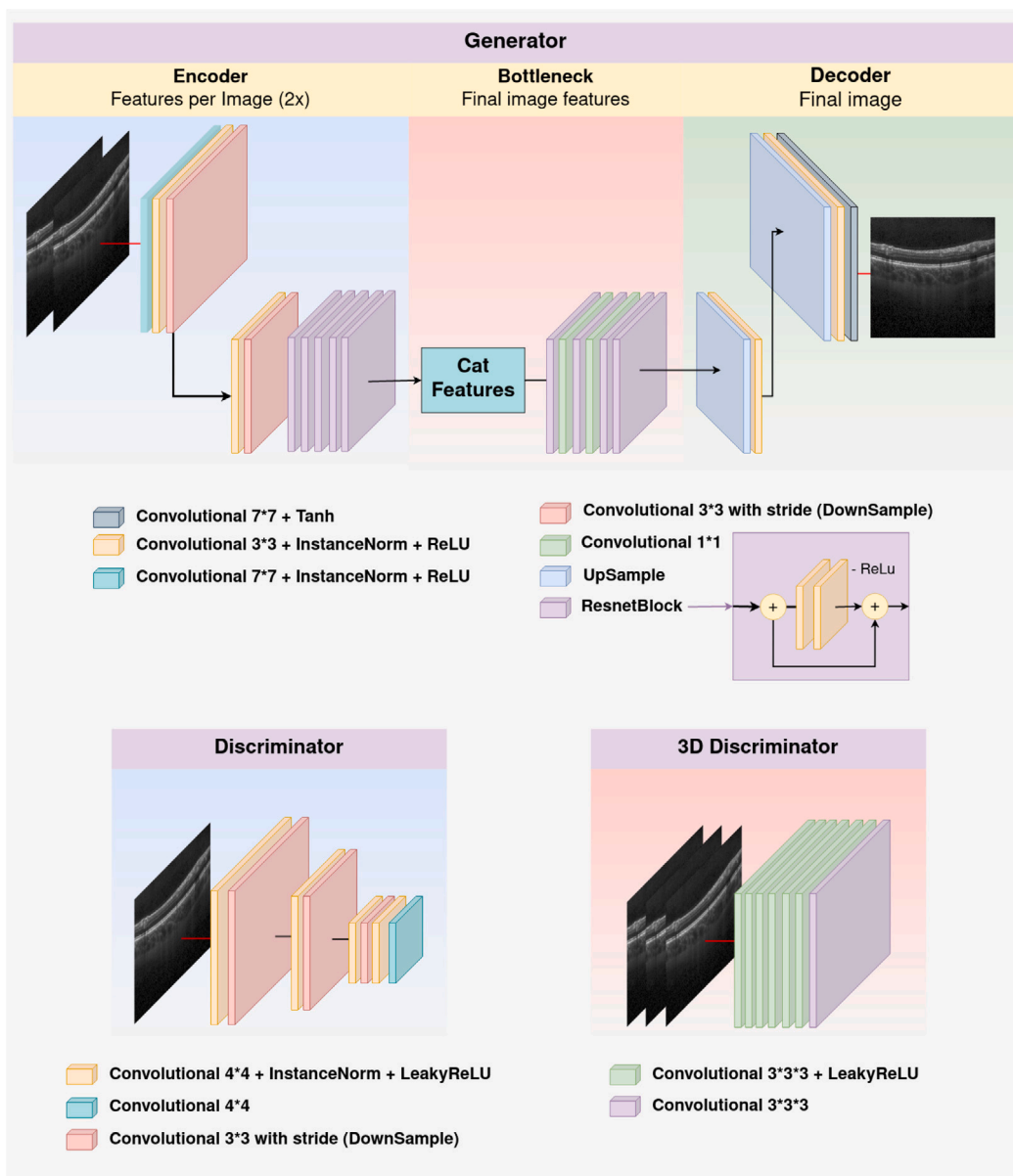
Fig. 3. Diagram showing the neural architectures of the generator and the two discriminators.

have a detrimental effect on performance, especially in scenarios where the minibatch size is small or the elements involved are correlated [32]. In our case, our network aims to generate the average slice using two closely adjacent slices so there is a clear correlation between these elements. In addition, a limited gpu memory forces us to use a small batch size, which also affects the performance of the BatchNorm layers. ReLu is used as a non-linear function, except for the last layer where a Tanh layer is used to create the final image. The network has three distinct parts that can be summarized as an encoder, a bottleneck and a decoder.

.   The **encoder** has a structure formed by several convolutional layers with kernels of size 3×3 and 7×7 and two convolutional layers with stride in charge of downsampling the feature maps. At the end we have 5 blocks of the ResNet with residual connections that help refine the feature maps that conform the image. The two images that conform the adjacent slices are passed through the encoder independently (as a batch of images), generating the feature maps for each slice separately.

.   The **bottleneck** is responsible for merging the sets of feature maps of the adjacent slices into a single set of feature maps characterizing the intermediate synthetic slice. To do this, the feature maps from both slices are concatenated and passed through a set of ResNet blocks and convolutional layers 1×1 responsible for doing the merging.

.   The **decoder** is formed by several convolutional and upsample layers and is in charge of creating the final synthetic image.

• The two **discriminators** act as support networks in the training process. We use a 2D discriminator that classifies individual images and a 3D discriminator that classifies a set of consecutive slices.

.   The **2D discriminator** is formed by a set of 3×3 and 4×4 convolutional layers where LeakyReLu is used as a nonlinear function.

.   The **3D discriminator** is formed by a set of $3 \times 3 \times 3$ convolutional layers where LeakyReLu is used as a nonlinear function.
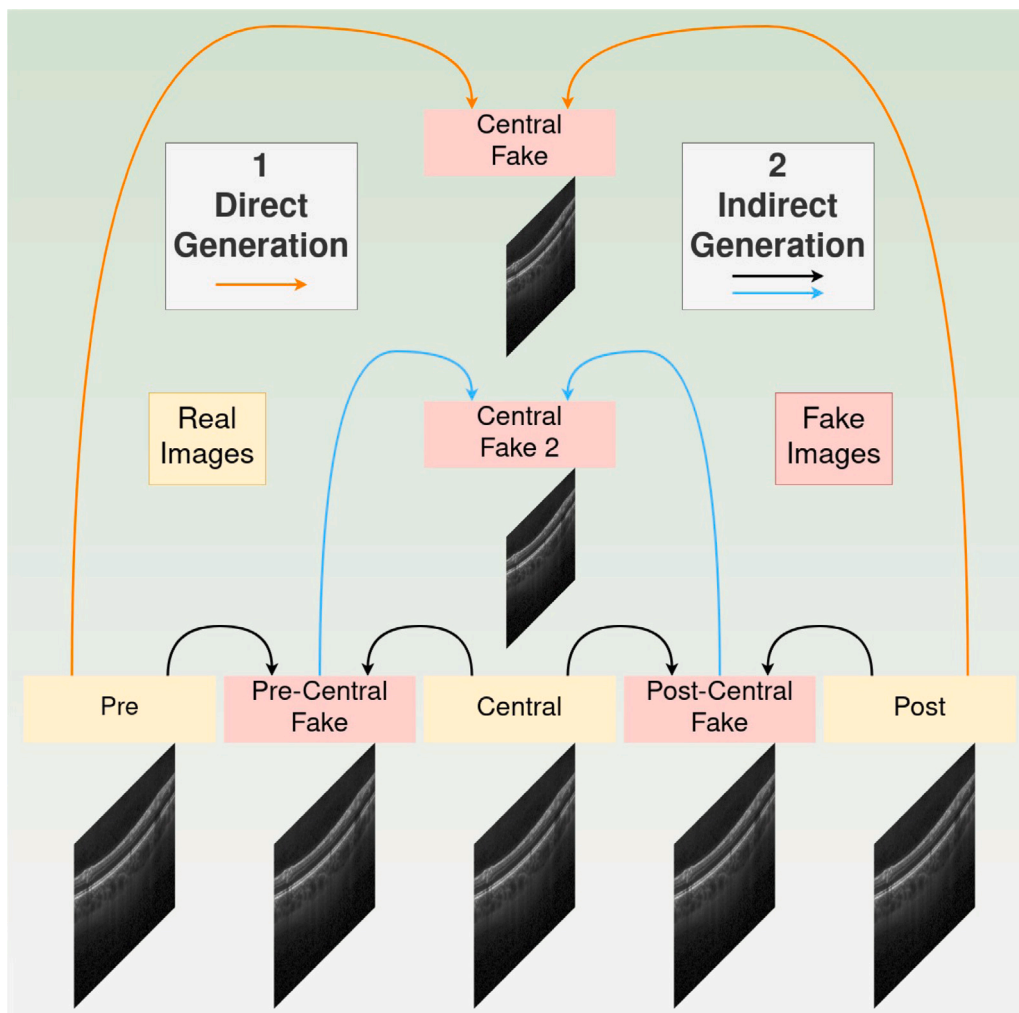
Fig. 4. Diagram showing how the central fake images are generated directly and indirectly.

3.2. Generator training

Unlike in a standard GAN training, we do not have a ground truth to compare the generated image with the real one using a similarity metric. This is because our aim is to generate a brand new synthetic intermediate slice from two adjacent slices. To achieve this objective, we modify the training process and, at each training step, we use three adjacent slices which we call `Pre`, `Central` and `Post`. The main idea is to use the central slice as a reference to make the similarity comparison. To do so, we generate two fake central slices that must be equivalent to the real `Central` slice, one directly and another indirectly. On the one hand, in a direct way we use as input for the generator the `Pre` and `Post` images to generate the intermediate `Central-Fake` slice. On the other hand, indirectly we use as input to the generator the adjacent slice pairs `Pre-Central` and `Central-Post` to generate intermediate slices which we can call `Pre-Central-Fake` and `Post-Central-Fake`. These two generated fake slices have the same distance between them as the `Pre-Central` and `Central-Post` slices so we can use them as input to the network in an equivalent way. By using these two generated fake slices as input we generate the intermediate slice between them which by distance has to be exactly the same as the `Central` slice and therefore we call `Central-Fake-2`. At the end of this process we have generated 4 synthetic slices in total; the two intermediate `Pre-Central-Fake` and `Post-Central-Fake` slices which are our final target and two slices equivalent to the real `Central` slice which are the `Central-Fake` and `Central-Fake-2` slices that we

use to calculate the loss based on the similarity metrics. Fig. 4 shows how each of these images are generated during the training.

By using the indirectly obtained `Central-Fake-2` we force the generator to generate the intermediate slices consistently even with subtle changes, since in order for `Central-Fake-2` to be similar to real `Central`, the `Pre-Central-Fake` and `Post-Central-Fake` intermediate slices used as input must have the appropriate changes. Yet our generator may be able to create these intermediate slices with strange appearances resulting in a `Central-Fake-2` slice similar to the real one. To avoid this situation, the `Central-Fake` slice forces the generator to produce slices similar in content to the real images. Moreover, the 2D discriminator forces the generator to make realistic images since it is trained to find the differences between real and generated images whereas the 3D discriminator, which uses a sequence of slices as input, compels the continuity of the sequence. This together with the similarity metrics prevents the generator from generating a realistic image that is not related to the adjacent slices. The exact methodology for training the GAN generator consists of three phases and is presented in Fig. 5.

First the generator makes its forward pass and gives rise to the four fake slices. Then, secondly, the discriminators are trained and the gradients of the generator are ignored by doing a detach, since we do not want to train it for now. Regarding the 2D discriminator, the generated images are labeled as false and the real images as true. In the 3D discriminator, one of the randomly chosen fake sequences (e.g. `Pre-real`, `Central-Fake`, `Post-Real`) is labeled as false and the real sequence as true. The optimizer step is performed and the
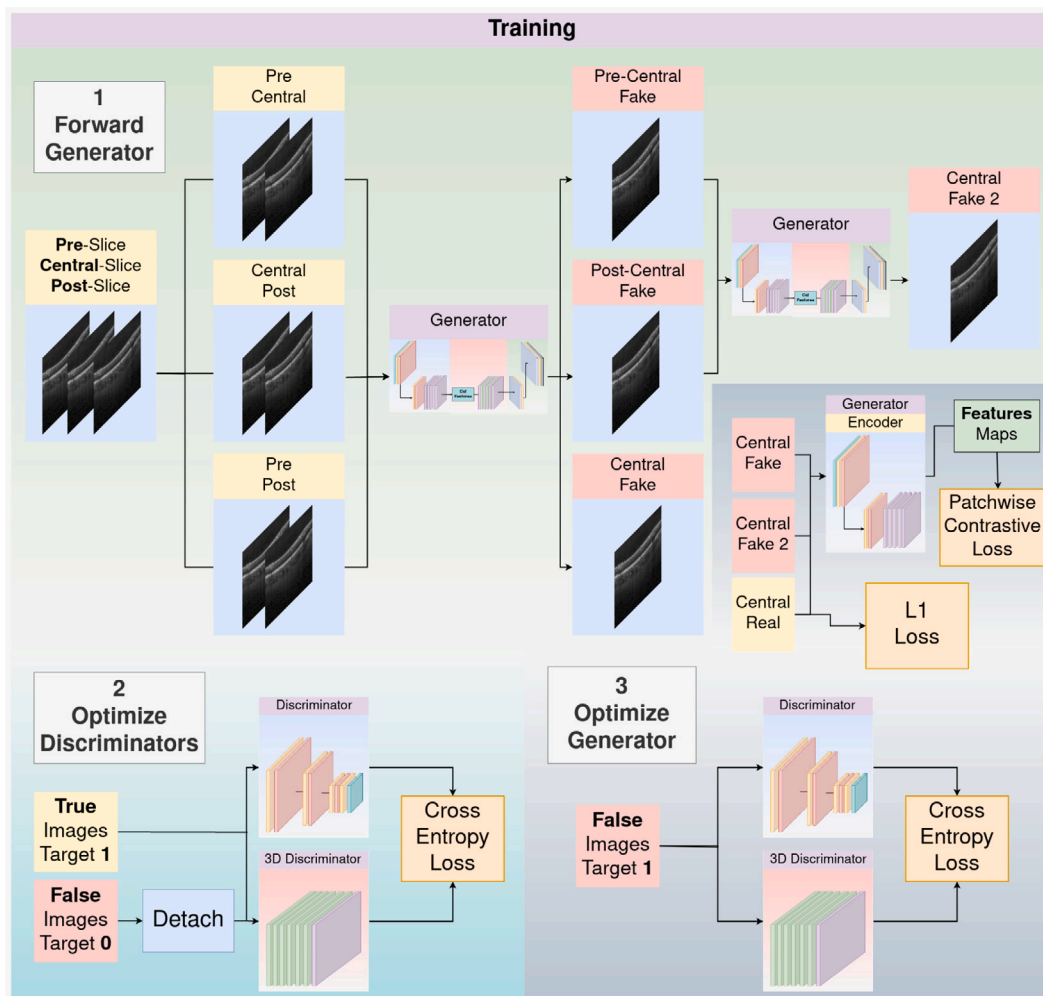
Fig. 5. Diagram showing the three main phases of the training process.

weights of both networks are updated. Finally, we proceed to train the generator. This training consists of two steps. A first step, similar to the discriminator training, where we label the generated images and fake sequences as true. We make the forward pass through the two discriminators and train the generator (ignoring the gradients generated by the forward passes through the discriminators, since we are not training them). We thus force our generator to try to fool the discriminators. The second step consists of two additional loss components that force the similarity between the fake `Central` images and the real one and, indirectly, also help to make the content of the intermediate slices realistic. The L1 loss is calculated in a direct way, while to calculate the Patchwise Contrastive loss [31] it is necessary to use the encoder of the generator to extract the multi-layer feature maps of the three images (real `Central`, `Central-Fake`, `Central-Fake-2`). Comparing these maps gives a more realistic similarity measure of the difference between the images. We compare the feature maps of the fake images with the real one so that patches with the same location in both images are forced to be similar, while patches in different locations are forced to be different, basically applying contrast learning with positive and negative examples. The final loss is calculated as the sum of the individual losses multiplied by a weight that determines their impact. This process is repeated at each iteration until the end of the training a fixed number of epochs.

3.3. Training details

In this section we present all the configuration details and all the parameters needed to fully replicate the methodology and experimentation performed in this work.

The dataset was divided into a training and a test set. The split chosen for our work is 85% of the cubes for training and 15% of the cubes for testing. This 85% reserved for training is composed of 36 OCT cubes where 18 are of Fast type and 18 are of Dense type. We do not use Posterior Pole type OCT cubes during training because we want to test if the network is able to generate the slices robustly even when using adjacent slices with different separation distances as input.

Each network uses a different optimizer, since they are effectively three independent networks. The Adam optimizer was used as the optimizer for the three networks. In the case of the generator, the optimizer was set with a weight decay of 0.001 to stabilize the training and prevent the generator from changing its predictions abruptly. The learning rate was set as a base at 0.00001 and kept static for 250 epochs. In this way, we allowed the generator to find the general abstraction of the images. After this, the value of the learning rate is updated linearly following Eq. (1) for an additional 100 epochs until it reaches zero.

$$Lr(n) = \frac{1 - max(0, n + n_f)}{n_d + 1} \tag{1}$$

where n is the epoch number, n_f represents the number of epochs without linear decrease in the learning rate and n_d the number of epochs that the learning rate will be decreasing. A total of three different losses were used to carry out the training. To train both the discriminator and the generator, the Cross Entropy defined in Eq. (2) was used as the adversarial loss.

$$CE = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c}) \tag{2}$$

where M is the number of classes, y is a binary indicator and p is the predicted probability observation o is of class c. In addition, to train the generator we used two losses that provide the measure of similarity of the images and force the synthetic images to preserve the information of the real ones. The first one used is the L1 loss with which we compare the mean absolute error (MAE) between the elements of the real image and the generated one defined in Eq. (3). The second one is the Patchwise Contrastive loss defined in Eq. (4) with which we compare the feature maps of the real image and the generated ones extracted by the encoder of our generator. In this case we also make a contrast comparison by forcing the patches located in the same location to be the same while the patches located in different locations of the image will be different. In our case we use the feature maps of layers 0,4,8,12,16 to make the comparison. Full details of this loss can be found in the original paper [31].

$$MAE(x, y) = mean(L), L = \{l_l, \ldots, l_N\}^T, l_n = |x_n - y_n| \tag{3}$$

$$L_{PatchNCE}(G, H, X) = E_x \sim X, \sum_{l=1}^{L} \sum_{s=1}^{S_t} l(\hat{z}_l^s, z_l^s, z_l^{S/s}) \tag{4}$$

where l specifies the particular layer that is being referred and L refers to the total number of layers. Moreover, for each layer, there is a specific number of spatial locations S_l.

3.4. Evaluation strategy

To evaluate the effectiveness of our methodology we used the 15% of OCT cubes that we separated for the test. These 6 OCT cubes consist of two Fast, two Dense and two Posterior pole cubes. This last type was never used during training so it is useful to test the robustness of our slice synthesis methodology. To see the effect that our proposal has on the visualization, we apply the network to all adjacent slices of the OCT cubes so we double the number of slices in each cube. We further apply the network to this new cube, quadrupling the number of slices with respect to the original size (minus one slice each time which is doubled by the edge effect). This results in Fast cubes going from 25 to 97 slices, Dense cubes going from 49 to 194 slices, and Posterior pole cubes going from 61 to 241 slices. Apart from observing the improvement in cross-sectional resolution, we use a 3D reconstruction of the cubes to observe the effect that our proposal has on their visualization. We compare the results obtained with our methodology with those obtained using a spline interpolation of order 3. We selected this type of interpolation as baseline for several reasons. OCT volumes are composed of images containing 3D structures, such as retinal layers or vessels, which are in different planes. This makes it very difficult to perform an accurate and effective interpolation, as most interpolation methods produce intermediate values that do not correspond to any of the layers present in the OCT. Previous tests with these interpolation methods produced disastrous results, which led us to use spline interpolation. By performing a smoother interpolation, this technique incurs much smaller errors, which makes it the most suitable option to be used as a baseline in our problem.

To check that the generated slices are realistic we follow two independent approaches. On the one hand, we randomly selected 25 fake images generated by our network and 25 real images. Three different medical experts labeled the 50 images using a scale from 1

to 5 where 1 meant that the image is very fake, 3 meant that they do not know how to identify it as either fake or true and 5 meant that the image is very real. We used these data to make comparisons in the response distributions of the medical personnel to check for significant differences between the generated images and the real ones. On the other hand, we tested the cubes with synthetic slices against an expert model. This model was trained to robustly segment the choroid in OCT cubes and if the generated images are realistic it should be able to segment the choroid in them equivalently to how it does in real images. To test this an expert labeled the choroid in 65 generated images and 65 real images. Two metrics were calculated to evaluate the segmentations, the Dice score (DSC) and pixel accuracy (PA) shown in Eq. (5), where X represents the prediction region and Y is the ground truth region.

$$DSC = \frac{2 \cdot |X \cap Y|}{|X| + |Y|} \qquad PA = \frac{|X \cap Y|}{|Y|} \tag{5}$$

Finally, to evaluate and compare the 3D visualizations created using the images generated using our network, the images generated with the spline, and the original images, we have calculated two different metrics. On the one hand, we used the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) algorithm. [33]. BRISQUE is an image quality evaluator which, unlike other measures such as Peak Signal to Noise Ratio, does not require a reference image to compare. Instead, it returns a score indicative of the perceptual quality of the processed image, with lower BRISQUE scores indicating a higher image quality. In this line, BRISQUE has been previously used to assess the quality of medical images with favorable results [34]. On the other hand, we use a smoothness metric. This metric is calculated taking into account that we are trying to improve the resolution and therefore the continuity of the different regions and structures that make up the OCT volume. For this we calculate the autocorrelation of the imaging volume with respect to the volume shifted by one pixel. A higher autocorrelation implies a greater smoothness in the volume and therefore an improvement in the visualization.

4. Results and discussion

In this section, we show all the results that were acquired throughout the evaluation of the proposed slice synthesis methodology. First, we show the results obtained at the 2D image level, where we can appreciate the generated images as opposed to the real images. We also show here the effect that this slice increase has on the cross-sectional resolution of the cube at the image level. Secondly, we show the effect that our methodology has on the visualization of the 3D reconstruction. We also show and compare the segmentation results obtained by the expert system and show the 3D reconstruction obtained using the segmented choroid.

4.1. Evaluation of the generated images

The results obtained by our slice generation methodology are shown below. In Fig. 6, two graphs can be seen which show the evolution of the set of losses used during the training. We can see how the losses of the discriminator and the generator converge quite stably to the central zone around 0.5 as is usual in training with Gans. On the other hand, the reconstruction losses that force the images to have certain structural characteristics converge to a lower value as the generated images are improved until they stabilize around 270 epochs. This improvement in the images can be seen in the examples of images synthesized throughout the training. The final model was chosen once the stabilization of the structural losses had occurred from epoch 300 onwards. The model of epoch 316 was selected as it produced the set of structural losses with the lowest value.

In Fig. 7 the responses of the two medical experts individually and the response as a whole are represented in boxplot form. We can see
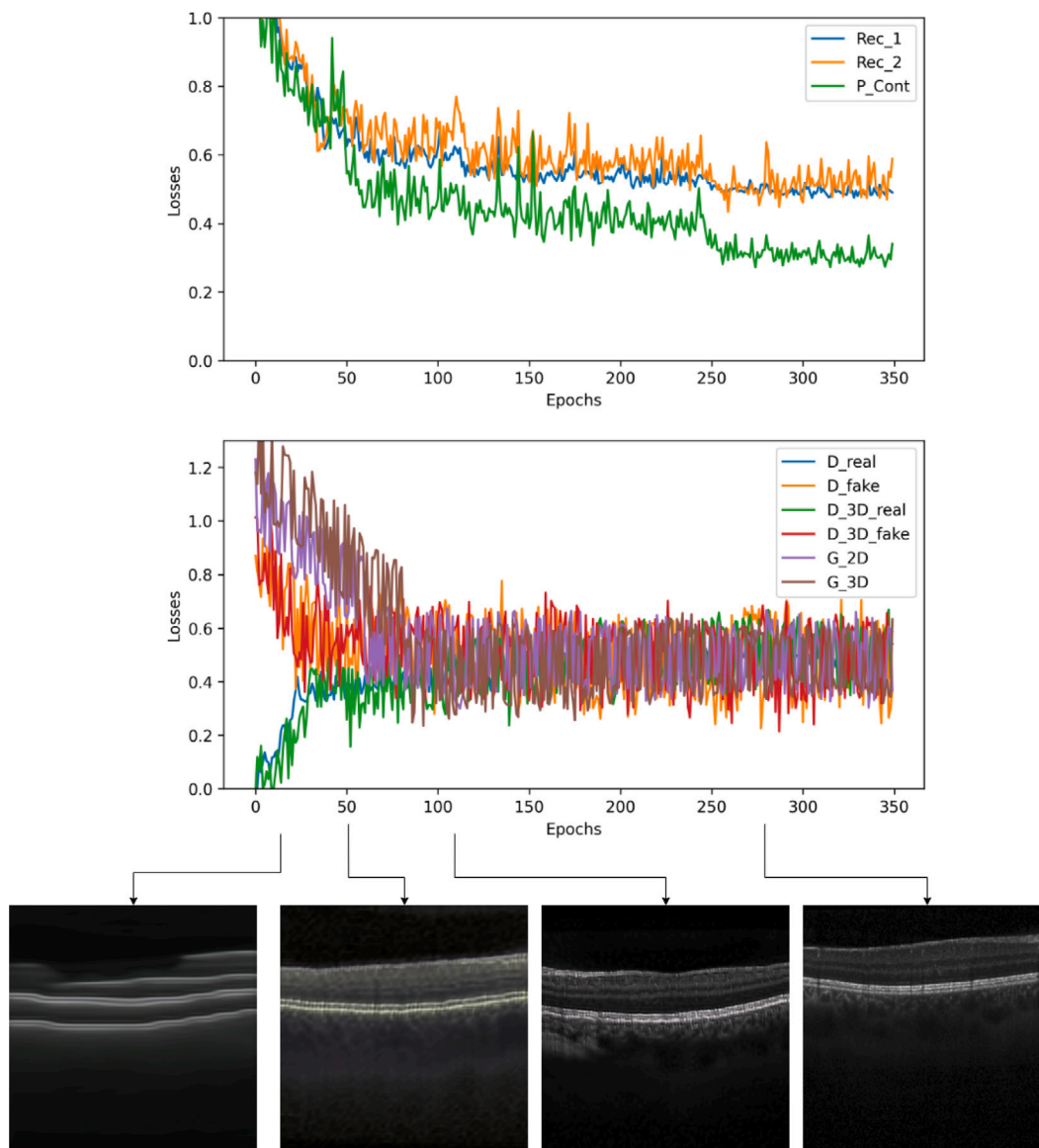
Fig. 6. Evolution of the different losses during training and examples of images generated in each iteration.

how the responses of the experts are consistent, labeling both true and generated images as true. Roughly speaking, we can see how the distribution in labeling is very similar between true and false images. This implies that medical experts are not able to distinguish generated images from real images, so our network is able to generate realistic images that do not have large defects or extraneous artifacts unbecoming of the imaging modality. The overall mode of the distribution is 3 while the mean is 3.37. Separated by expert the mean is 3.44 for expert 1 and 3.3 for expert 2. We can observe how there is a general tendency to mark the images as real possibly due to the fact that there are no strange artifacts in the generated images. The experts present a similar and conservative behavior and tend to see the images as real and only in a few of them they suspect that there is something strange.In Fig. 8, a several examples of real images and several examples of images generated by our network can be seen. In general the generated images are realistic and indistinguishable from the real ones.

We can see that in general there is an extreme consensus among experts when assessing whether an image is real or fake. Table 2 compares the expert consensus prediction of whether an image is real or generated with respect to the target and also shows the number of images where there is a dissenting opinion. We consider the experts

Table 2

On the left side, confusion matrix based on expert consensus classification of images as real or fake. On the right side, table showing the number of images with discrepancies.

		Predicted		Discrepancy	
		Real	Fake	Real	Fake
Target	Real	19	1	0	2
	Fake	19	1		

to be in consensus when all indicate the same option using 3 as the threshold. For example, the real image consensus is reached when one expert votes more than 3 on that image and the other vote 3 or higher.

The experts label as real the same number of real images as generated images, which implies that the generated images are hardly distinguishable from the real ones. There is a discrepancy with respect to two images where one expert labeled the image as real and another as fake. Finally, there is one generated images that all experts labeled as fake. Looking at any of these images, one can see some defect that may be the cause that inclined the medical experts to label this image as fake. In Fig. 9, a series of three real adjacent slices, the
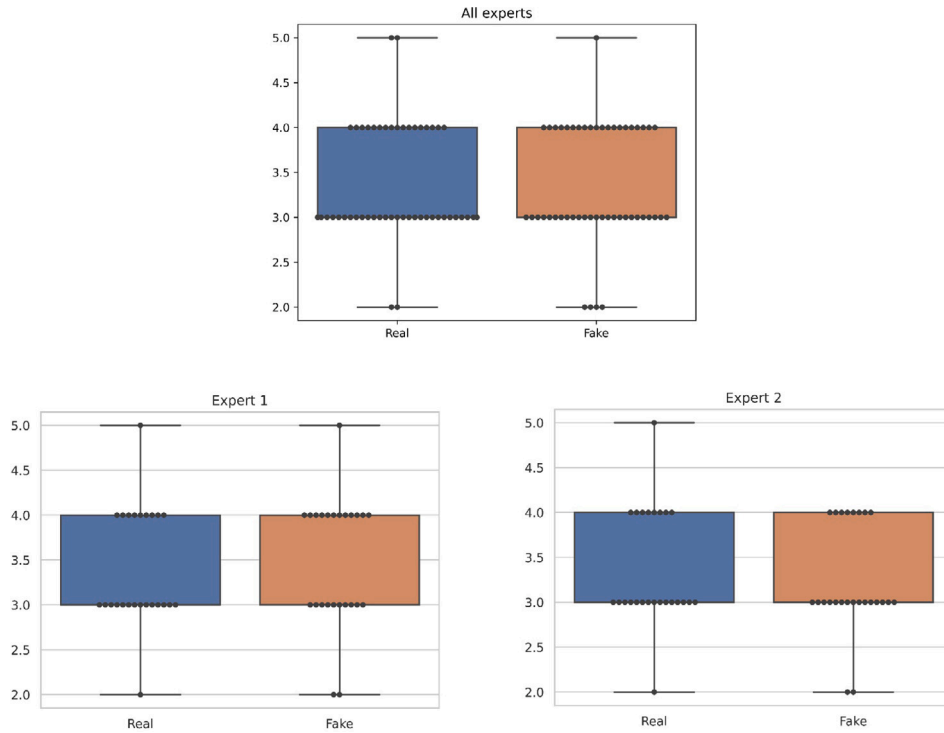
Fig. 7. Boxplot showing the distribution of the responses of the two medical experts classifying each image as fake or real. The blue left column indicates real images while the orange right column indicates false images. The scale corresponds to 1 very fake and 5 very real.
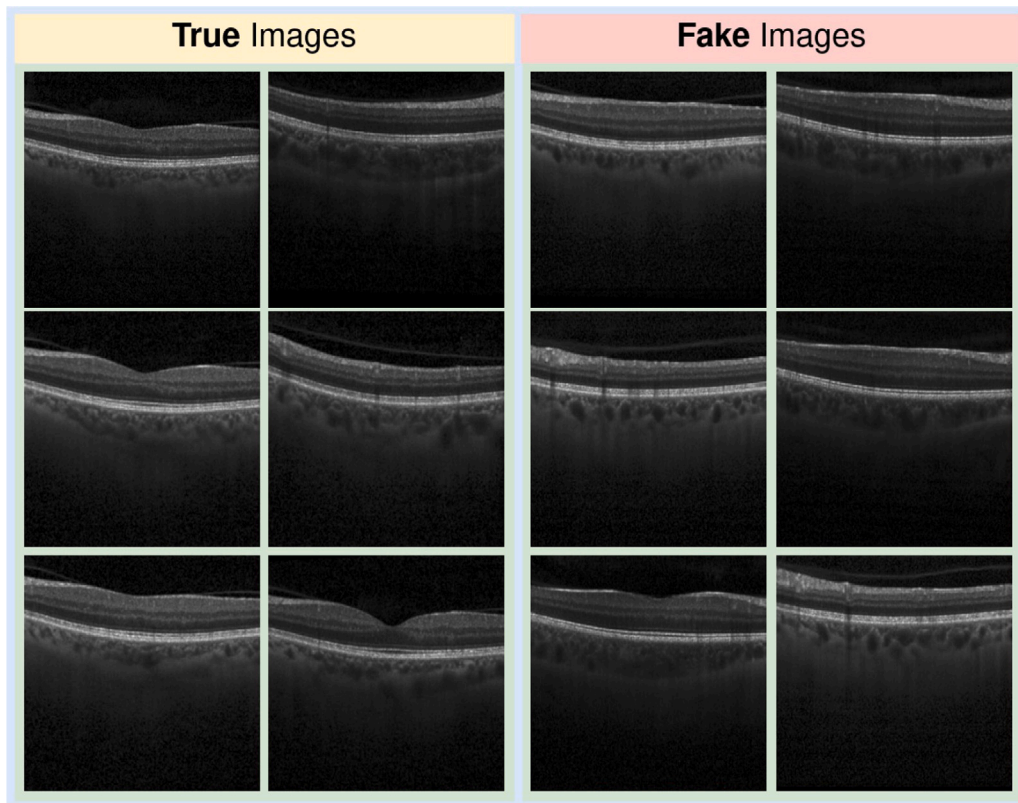

Fig. 8. Several examples of real slices and slices generated by the network.
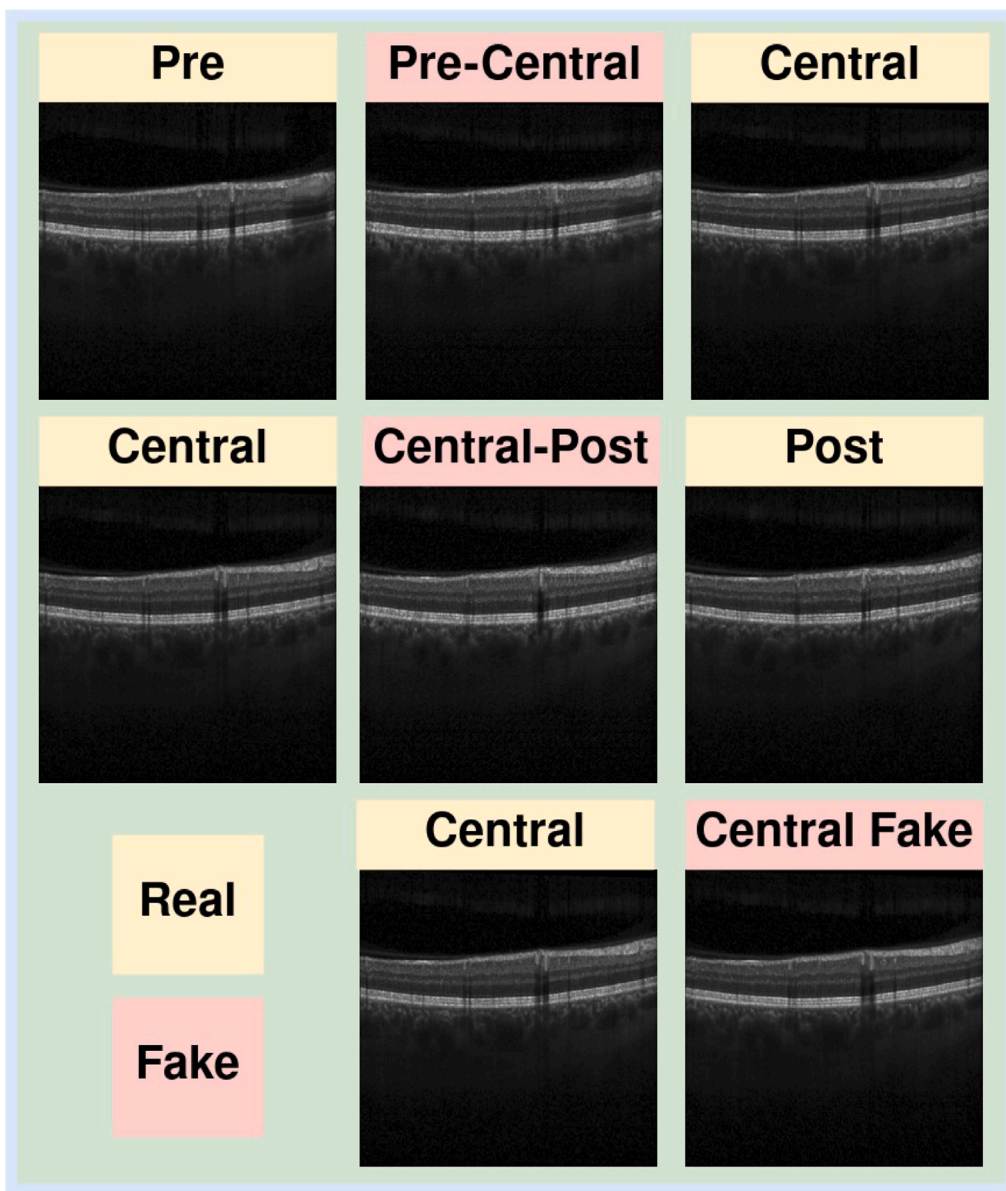
Fig. 9. Example showing the three real adjacent slices (`Pre`, `Central` and `Post`), the two generated intermediates and the fake central one reconstructed using the two intermediates generated during training.

two intermediate slices generated by our network and the fake central slice reconstructed in the training process are shown. The `Central-Post` slice was labeled as false by the two medical experts. We can see that in this slice there is a strange noise pattern at the bottom of the image which gives away its fakeness. In Fig. 10, a zoom of this image can be seen with increased brightness showing the extraneous noise in greater detail. This noise pattern arises when trying to interpolate the noise at the bottom of the images which may not have coherence. Even so, the generated image is accurate and realistic presenting the necessary changes in the different structures that exist in the eye to show continuity with respect to its adjacent slices. We can also observe how the reconstructed central slice is very similar to the real central slice.

Below we show the results obtained by our evaluation approach using our expert system. In Table 3, you can see the values obtained for DSC and PA using real images and using fake generated images created by our network. We can see that, although the values of both metrics are minimally lower, there are no significant differences between them.

Table 3
Segmentation metrics obtained by the expert system using the real images and using the fake images.

	Real	Fake
DSC	89.5	89.3
PA	89.8	89.5

This indicates that for our expert system the fake images generated by our network are equivalent to the real images and the choroid is continuous between slices.

Finally, in order to observe the effect on the improvement of cross-sectional resolution, three cross-sectional slices at different heights of an OCT cube are presented in Fig. 11. A representative image of the OCT cube and three cross-sectional images obtained using the original cube, the cube with twice the number of generated slices and the cube with four times the number of generated slices are presented. It can be
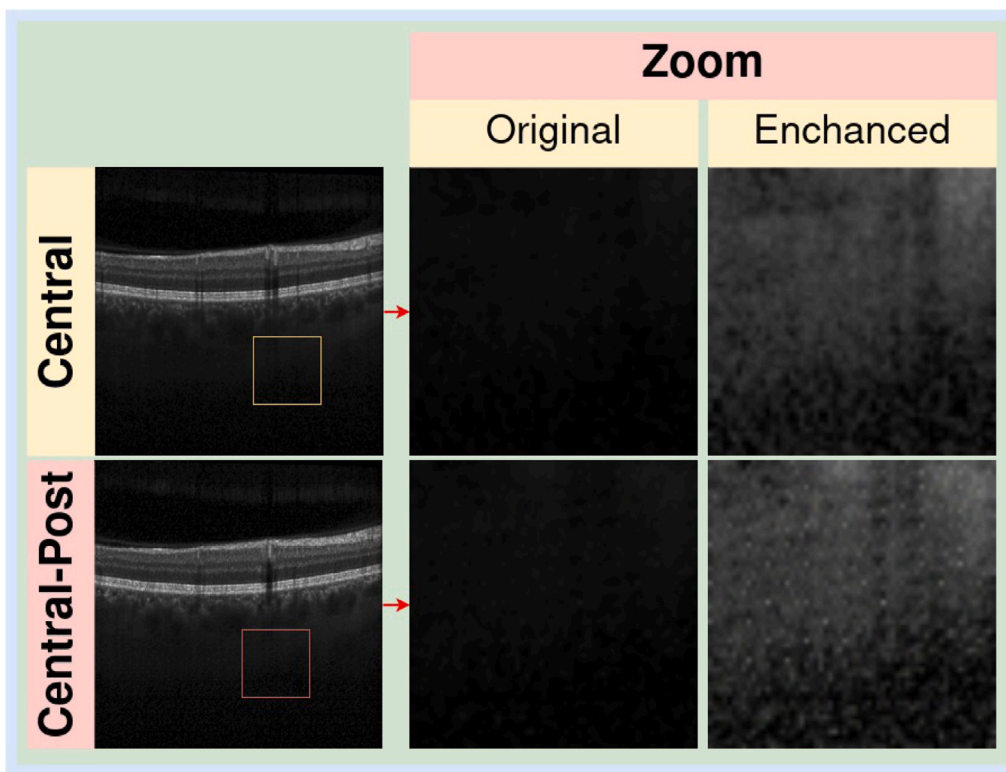
Fig. 10. Comparison of the real `Central` image and the fake `Central-Post` image labeled by the two experts as fake. A strange noise at the bottom of the image can be seen with the zoom.

Table 4
Average BRISQUE score obtained for the 3D visualizations created using the original volumes, the ones generated with the baseline and the ones generated with our network.

	Fast	Dense	Posterio pole
Original Size	42.1	39.2	37.2
Spline x2	41.8	42.4	36.5
Spline x4	42.3	43.3	36.8
Network x2	41.3	37.0	35.4
Network x4	39.7	36.1	33.1
Network x8	42.8	39.1	37.1

Table 5
Average autocorrelation value obtained for the 3D visualizations created using the original volumes, the ones generated with the baseline and the ones generated with our network.

	Fast	Dense	Posterio pole
Original Size	0.76	0.79	0.82
Spline x2	0.77	0.80	0.84
Spline x4	0.79	0.81	0.85
Network x2	0.78	0.83	0.83
Network x4	0.83	0.85	0.86
Network x8	0.82	0.85	0.84

seen how the increase in the number of slices improves the resolution in the different heights of the cube allowing a more precise observation of the changes produced in the vessels and the different retinal layers.

4.2. Effect on 3D visualization

In this section we show the effects of our slice synthesis methodology on 3D reconstructions and visualizations of OCT cubes. We show below the results obtained for the BRISQUE and autocorrelation metrics using the original volumes, the volumes expanded using the baseline and the volumes expanded using our network. In Table 4 and in Table 5, the scores obtained for these two metrics are shown in each case. A low BRISQUE score and a high autocorrelation indicate a better quality and smoothness in the final 3D visualization. It can be observed how both metrics improve in the three types of cubes tested when using our network. This improvement increases when multiplying by 4 the existing slices, being very noticeable the improvement in posterior pole volume type. The results show that the deterioration in the quality of the images starts after multiplying the number of slices by 8, which indicates that this is the maximum limit of slice enlargement, since from this point onwards our network starts to introduce artifacts that worsen the quality of the reconstructions generated.

In Fig. 12, a Fast type OCT cube, a Dense type OCT cube and a Posterior Pole OCT cube are shown. It can be clearly observed the improvement against the display using the original size, where the different textures that compose the cube are almost not visible. Our methodology allows us to obtain a clear representation in contrast to other interpolation methodologies such as the spline, that increases the artifacts of the cube. In addition, Fig. 13 shows the same 3D visualization with a zoom of the OCT Posterior pole cube. This type of cube was not used in the training and has a different slice spacing than the Fast and Dense cubes used. Our methodology is able to generate these slices improving the visualization of the 3D cube. In the visualization we can observe a clear change in the smoothness of the texture and the definition of important structures for diagnosis such as vessels. Our methodology is therefore robust being able to generate appropriate images even when using slice distances different from those originally used as input.

One of the most important biomarkers when following the evolution of MS is the thickness of the choroid. Therefore, being able to detect changes in the thickness of this layer in an appropriate manner is essential. In relation to this, Fig. 14 shows the effect of our approach in the 3D reconstruction given a segmentation mask. In this case the choroid segmentation mask is shown in a green-red color gradient
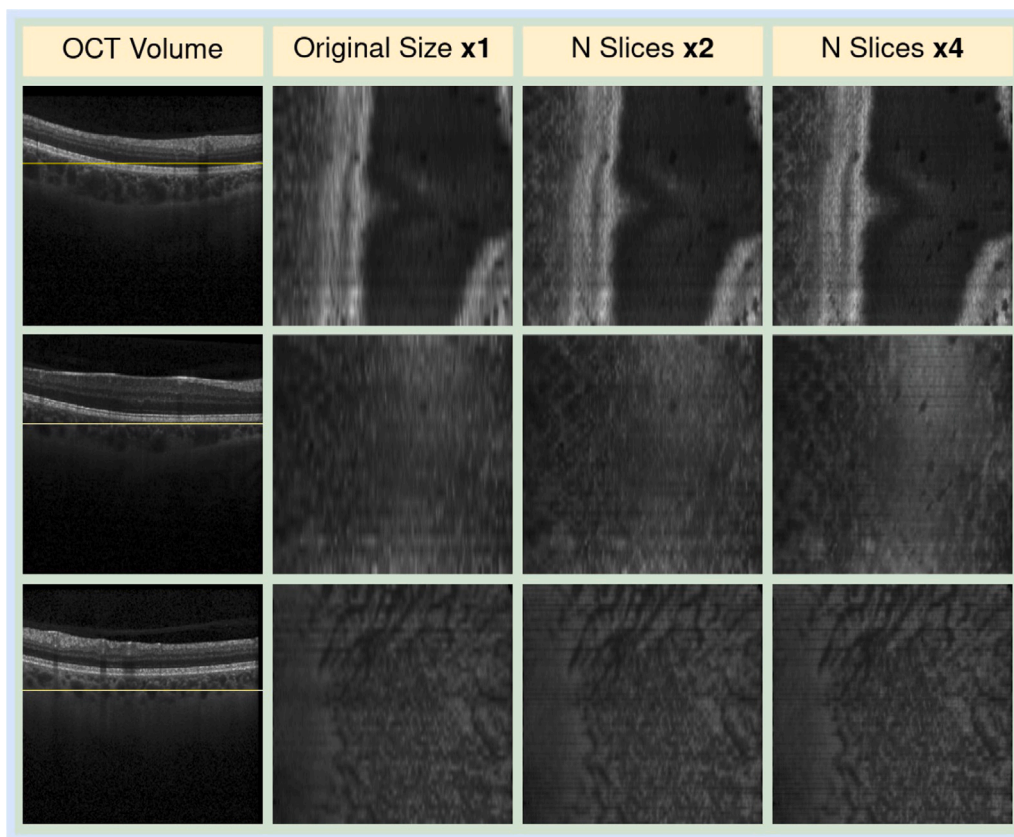
Fig. 11. In the first column there is a representative image of the OCT cube, the remaining three columns contain the cross-sectional image of the original size cube, the same image but doubling the number of slices with the net and the same image quadrupling the number of slices.

indicating as green the areas where the choroid is less thick and as red the areas where the choroid is thicker. It can be seen how the expansion of the cube using our methodology helps to better visualize the changes in thickness of this layer, allowing for more accurate detection of thickness in specific areas. While this methodology was designed to solve the problem of the low cross-sectional resolution of many OCT cubes, it is applicable to other medical imaging modalities that present similar problems. In addition, we have tested our methodology focusing on the monitoring of MS, but it is extendable and applicable to the monitoring of many other ocular diseases such as age-related macular degeneration and systemic diseases such as hypertension.

5. Conclusions

In this work, we have described how having a good quality and good resolution medical image plays a fundamental role in diagnosing a large number of diseases such as multiple sclerosis. Knowing the limitations that exist in various types of OCT cubes in the current clinic, we have proposed a fully automatic methodology that allows us to double the number of slices in the different types of cubes by increasing their resolution.

We have based our work on a generator network trained using three adjacent slices. This training allows us to recreate slices that do not exist in the cube through indirect reconstruction of the central slice, adversarial learning and contrast learning. This training methodology has proven to be robust, allowing to perform the training in an unsupervised way and allowing to generate OCT slices using as input slices with distances between them not seen in the training.

In this sense, the slices generated by our proposal were tested from two different approaches. On the one hand, three medical experts evaluated the realism of the slices in a blind test by comparing them with real images. On the other hand, we tested our images against an expert system based on a network trained to segment the choroid. We verified that the network is able to segment the choroid in the network-generated images in an equivalent way as it does in real images. We also tested the effect that our methodology causes in the cross-sectional image, in a 3D reconstruction and in a 3D reconstruction where we use a semantic segmentation mask. All tests allow us to conclude that our network is able to generate realistic and continuous intermediate slices with the cube effectively increasing the cross-sectional resolution which improves the efficiency of medical diagnosis.

As future work we intend to extend the use of this network to other imaging modalities. Furthermore, although our dataset has a considerable size and variability, we consider it of vital importance to test our methodology with other datasets in order to test the overall robustness of the synthesis method. In addition, to get the most out of this methodology it is necessary to integrate it as a previous step to other methodologies that perform segmentations, feature extractions or classifications, since our methodology allows to obtain a greater number of reference points when calculating biomarkers such as the thickness of retinal layers, which should provide robustness to the diagnostic process. In the case of sclerosis monitoring, an efficient segmentation methodology needs to be designed to segment all relevant structures in any type of OCT image, such as the multiple retinal layers, the choroid or the choroidal vessels. The integration of such methodologies with our synthesis methodology has a lot of potential and opens a wide horizon for future work.
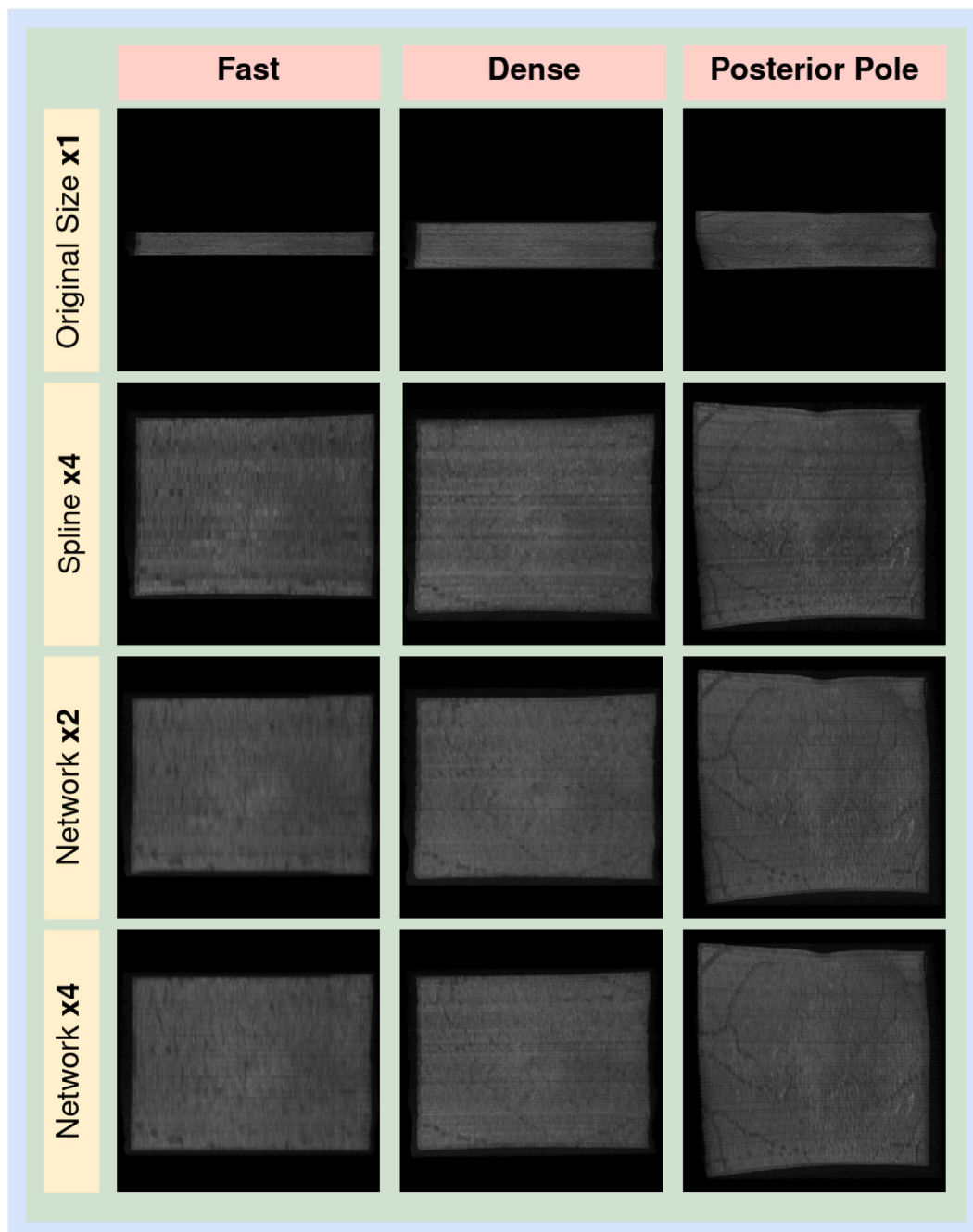
Fig. 12. 3D reconstruction of a Fast, Dense and Posterior Pole type cubes. From top to bottom: the original cube, a zoom using a spline, the duplicate slice by 2 and the duplicate by 4.
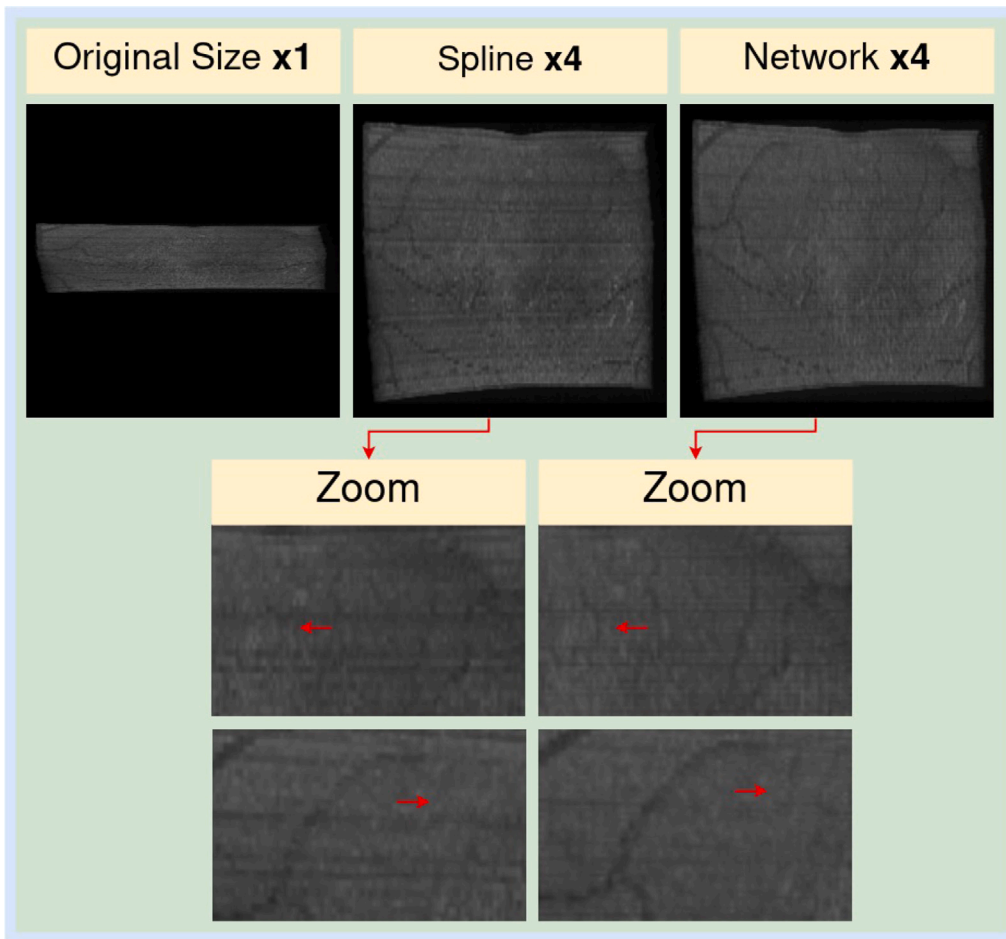
Fig. 13. 3D visualization of a posterior pole cube type. The red arrow indicates an example of the improvement in the definition of a vessel.
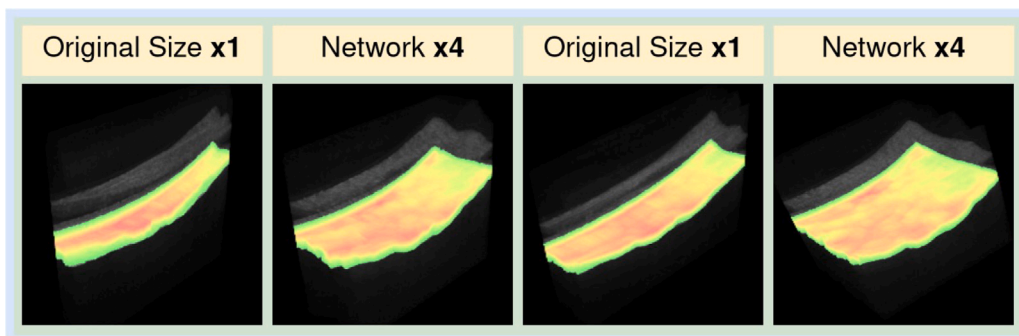

Fig. 14. 3D reconstruction showing the choroid on a green-red color scale based on its thickness.

References

[1] M. Mahesh, The essential physics of medical imaging, Med. Phys. 40 (7) (2013) 077301.

[2] K. Zhang, H. Hu, K. Philbrick, G.M. Conte, J.D. Sobek, P. Rouzrokh, B.J. Erickson, SOUP-GAN: Super-resolution MRI using generative adversarial networks, Tomography 8 (2) (2022) 905–919.

[3] F.A. Medeiros, L.M. Zangwill, C. Bowd, R.M. Vessani, R. Susanna Jr., R.N. Weinreb, Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography, Am. J. Ophthalmol. 139 (1) (2005) 44–55.

[4] E. López-Varela, J. Novo, J.I. Fernández-Vigo, F.J. Moreno-Morillo, M. Ortega, Unsupervised deformable image registration in a landmark scarcity scenario: Choroid OCTA, in: International Conference on Image Analysis and Processing, Springer, 2022, pp. 89–99.

[5] J.I. Fernández-Vigo, F.J. Moreno-Morillo, M. Ortega-Hortas, E. López-Varela, J. Novo-Bujan, B. Burgos-Blasco, L. López-Guajardo, J. García-Feijóo, J. Donate-López, Early changes in choriocapillaris flow voids as an efficacy biomarker of photodynamic therapy in central serous chorioretinopathy, Photodiagnosis Photodyn. Ther. 38 (2022) 102862.

[6] E. López-Varela, J. de Moura, J. Novo, J.I. Fernández-Vigo, F.J. Moreno-Morillo, M. Ortega, Fully automatic segmentation and monitoring of choriocapillaris flow voids in OCTA images, Comput. Med. Imaging Graph. 104 (2023) 102172.

[7] P.L. Vidal, J. de Moura, M. Díaz, J. Novo, M. Ortega, Diabetic macular edema characterization and visualization using optical coherence tomography images, Appl. Sci. 10 (21) (2020) 7718.

[8] E. López-Varela, P.L. Vidal, N.O. Pascual, J. Novo, M. Ortega, Fully-automatic 3D intuitive visualization of age-related macular degeneration fluid accumulations in OCT cubes, J Digit. Imaging (2022) 1–12.

[9] E. Garcia-Martin, M. Ortiz, L. Boquete, E.M. Sánchez-Morla, R. Barea, C. Cavaliere, E. Vilades, E. Orduna, M.J. Rodrigo, Early diagnosis of multiple sclerosis by OCT analysis using Cohen's d method and a neural network as classifier, Comput. Biol. Med. 129 (2021) 104165.

[10] F. Di Staso, M. Ciancaglini, S. Abdolrahimzadeh, F. D'Apolito, G. Scuderi, Optical coherence tomography of choroid in common neurological diseases, In Vivo 33 (5) (2019) 1403–1409.

[11] E. Garcia-Martin, L. Jarauta, E. Vilades, J.R. Ara, J. Martin, V. Polo, J.M. Larrosa, L.E. Pablo, M. Satue, Ability of swept-source optical coherence tomography to detect retinal and choroidal changes in patients with multiple sclerosis, J. Ophthalmol. 2018 (2018).

[12] A.G. Merino, J.A. Calizo, O.F. Fernández, L.L. Pascual, E.M. Torres, A.R.-A. Zarrantz, Consenso para el tratamiento de la esclerosis múltiple 2016. Sociedad Española de Neurología, Neurología 32 (2) (2017) 113–119.

[13] J.B. Hassell, E.L. Lamoureux, J.E. Keeffe, Impact of age related macular degeneration on quality of life, Br. J. Ophthalmol. 90 (5) (2006) 593–596.

[14] A. Sicras-Mainar, E. Ruíz-Beato, R. Navarro-Artieda, J. Maurino, Impact on healthcare resource utilization of multiple sclerosis in Spain, BMC Health Serv. Res. 17 (1) (2017) 1–7.

[15] D. Huang, E.A. Swanson, C.P. Lin, J.S. Schuman, W.G. Stinson, W. Chang, M.R. Hee, T. Flotte, K. Gregory, C.A. Puliafito, et al., Optical coherence tomography, Science 254 (5035) (1991) 1178–1181.

[16] M. Choi, S. Lee, H. Kim, K.M. Lee, Motion-aware dynamic architecture for efficient frame interpolation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 13839–13848.

[17] H. Li, Y. Yuan, Q. Wang, Video frame interpolation via residue refinement, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2020, pp. 2613–2617.

[18] P.L. Vidal, J. de Moura, J. Novo, M.G. Penedo, M. Ortega, Image-to-image translation with generative adversarial networks via retinal masks for realistic optical coherence tomography imaging of diabetic macular edema disorders, Biomed. Signal Process. Control 79 (2023) 104098.

[19] M. Gende, J. de Moura, J. Novo, M. Ortega, High/low quality style transfer for mutual conversion of OCT images using contrastive unpaired translation generative adversarial networks, in: Image Analysis and Processing–ICIAP 2022: 21st International Conference, Lecce, Italy, May 23–27, 2022, Proceedings, Part I, Springer, 2022, pp. 210–220.

[20] Y. Chen, F. Shi, A.G. Christodoulou, Y. Xie, Z. Zhou, D. Li, Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 91–99.

[21] I. Sánchez, V. Vilaplana, Brain MRI super-resolution using 3D generative adversarial networks, 2018, arXiv preprint arXiv:1812.11440.

[22] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong, et al., CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE), IEEE Trans. Med. Imaging 39 (1) (2019) 188–203.

[23] A. Kudo, Y. Kitamura, Y. Li, S. Iizuka, E. Simo-Serra, Virtual thin slice: 3D conditional GAN-based super-resolution for ct slice interval, in: International Workshop on Machine Learning for Medical Image Reconstruction, Springer, 2019, pp. 91–100.

[24] C. Peng, W.-A. Lin, H. Liao, R. Chellappa, S.K. Zhou, SAINT: spatially aware interpolation network for medical slice synthesis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7750–7759.

[25] C. Fang, L. Wang, D. Zhang, J. Xu, Y. Yuan, 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, 2022, pp. 20677–20686.

[26] M. Melinščak, M. Radmilović, Z. Vatavuk, S. Lončarić, Annotated retinal optical coherence tomography images (AROI) database for joint retinal layer and fluid segmentation, Automatika 62 (3–4) (2021) 375–385.

[27] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035, URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[28] Itseez, Open source computer vision library, 2015, https://github.com/itseez/opencv.

[29] E. López-Varela, N. Barreira, N.O. Pascual, E.G. Ben, S.R. Cid, M.G. Penedo, Fully automatic segmentation of the choroid in non-EDI OCT images of patients with multiple sclerosis, Procedia Comput. Sci. 207 (2022) 726–735, http://dx.doi.org/10.1016/j.procs.2022.09.128, URL https://www.sciencedirect.com/science/article/pii/S1877050922010092. Knowledge-Based and Intelligent Information and Engineering Systems: Proceedings of the 26th International Conference KES2022.

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

[31] T. Park, A.A. Efros, R. Zhang, J.-Y. Zhu, Contrastive learning for unpaired image-to-image translation, in: European Conference on Computer Vision, 2020.

[32] S. Singh, S. Krishnan, Filter response normalization layer: Eliminating batch dependence in the training of deep neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11237–11246.

[33] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.

[34] Z. Zhang, G. Dai, X. Liang, S. Yu, L. Li, Y. Xie, Can signal-to-noise ratio perform as a baseline indicator for medical image quality assessment, IEEE Access 6 (2018) 11534–11543.