Biomedical Signal Processing and Control 79 (2023) 104098

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Image-to-image translation with Generative Adversarial Networks via retinal masks for realistic Optical Coherence Tomography imaging of Diabetic Macular Edema disorders

Plácido L. Vidal, Joaquim de Moura *, Jorge Novo, Manuel G. Penedo, Marcos Ortega

Centro de investigación CITIC, Universidade da Coruña, Campus de Elviña, s/n, 15071 A Coruña, Spain
Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruña, Xubias de Arriba, 84, 15006 A Coruña, Spain

ARTICLE INFO

Keywords:
Optical coherence tomography
Generative adversarial network
Image-to-image translation
Diabetic macular edema
Synthetic data

ABSTRACT

One of the main issues with deep learning is the need of a significant number of samples. We intend to address this problem in the field of Optical Coherence Tomography (OCT), specifically in the context of Diabetic Macular Edema (DME). This pathology represents one of the main causes of blindness in developed countries and, due to the capturing difficulties and saturation of health services, the task of creating computer-aided diagnosis (CAD) systems is an arduous task. For this reason, we propose a solution to generate samples. Our strategy employs image-to-image Generative Adversarial Networks (GAN) to translate a binary mask into a realistic OCT image. Moreover, thanks to the clinical relationship between the retinal shape and the presence of DME fluid, we can generate both pathological and non-pathological samples by altering the binary mask morphology. To demonstrate the capabilities of our proposal, we test it against two classification strategies of the state-of-the-art. In the first one, we evaluate a system fully trained with generated images, obtaining 94.83% accuracy with respect to the state-of-the-art. In the second case, we tested it against a state-of-the-art expert model based on deep features, in which it also achieved successful results with a 98.23% of the accuracy of the original work. This way, our methodology proved to be useful in scenarios where data is scarce, and could be easily adapted to other imaging modalities and pathologies where key shape constraints in the image provide enough information to recreate realistic samples.

1. Introduction

In the current information era, and thanks to the advances in the information and machine learning technologies, we are able to solve problems more efficiently and, most importantly, automatically. One of the main sectors that greatly benefited from these new technologies is the field of bioinformatics, specially in computer-aided diagnostic issues. These systems confer the diagnosis process an independence from the human subjectivity, and make the results repeatable over the time (critical in the field for a proper follow-up of the patients). The main issue with these methodologies is the scarcity of available labeled data to train or adapt these methodologies, as well as problems with class imbalance. To solve this issue, methodologies have been developed for the two main components that comprise a machine learning system: the learning algorithm itself (or network architecture) and the dataset used to train it.

Regarding improvements to the learning algorithm, we can find methodologies that try to create systems able to learn and generalize from a greatly limited number of samples (also called “Few Shot learning” [1]). This is illustrated by even one of the most representative networks in the field of medical imaging, the U-Net [2]. This architecture was originally proposed as a network that, thanks to its skip-connections between encoder and decoder, is able to return a segmentation with a limited number of samples (as the decoder is able to salvage information that, otherwise, would be lost in the neck of the funnel).

The second approach consists in, instead of finding strategies that are able to extract information from a reduced number of samples, transforming the dataset so a machine learning strategy is able to better generalize from the limited number of patterns present, called “Data Augmentation” [3]. These strategies can be divided into three main categories: geometrical (random rotations, cropping, flipping, shearing, elastic transformation . . .), intensity based (contrast, brightness, noise, blur . . .) [4] and (more recently) synthetic, in an effort to generate

* Corresponding author.
E-mail addresses: placido.francisco.lizancos.vidal@udc.es (P.L. Vidal), joaquim.demoura@udc.es (J. de Moura), jnovo@udc.es (J. Novo), mgpenedo@udc.es (M.G. Penedo), mortega@udc.es (M. Ortega).

https://doi.org/10.1016/j.bspc.2022.104098
Received 1 March 2022; Received in revised form 27 June 2022; Accepted 11 August 2022
Available online 23 August 2022

completely new samples to balance and complement the dataset. An example of these methodologies is the approach of Zhao et al. [5], where they model the set of spatial and appearance transformations between images in the dataset to, then, synthesize new examples by sampling transformations and applying them to a single labeled example.

The issue of data scarcity is specially relevant in the ophthalmological domain as, despite the easy accessibility of the organs, the image modalities and precision required to obtain these images require the patient to maintain a firm stance, which often results in lesser quality images as the capturing time has to be reduced to attain a minimum usable product [6–8]. For this reason, methodologies have been presented in this field that address the problem of the lack of samples (and their quality) in both of the aforementioned approaches. For example, as a few-shot proposal, the work of Kim et al. [9] is centered in the problem of glaucoma diagnosis. This work presents a few-shot strategy based on a matching neural network architecture [10] that returns significantly better results with the same limited dataset than a conventional convolutional neural network and with similar accuracy to what a human expert would return. Another example of this paradigm is the work of Medela et al. [11], where they train a deep siamese neural network [12] in a domain with a large number of samples. Then, they use this siamese network to extract features from the limited number of new domain samples and train a shallow classifier; in this case, a Support Vector Machine (SVM).

On the other hand, as an example of works in the field of data augmentation, we can study the work of Burlina et al. [13], where they use progressively grown generative adversarial networks or GAN [14] to generate Age-Related macular degeneration eye fundus images. This work was tested with trained specialists, which could not distinguish with statistical significance synthetic images from real eye-fundus images. Another example is the work of Zhao et al. [15], where they generate a full eye fundus image from a retinal vasculature and random noise input as seeds. Finally, works like the ones proposed by Hu, Liang and Lu [16] by slicing samples into multiple iterations usable for training or Vidal et al. [17,18] by dividing a sample into thousands of overlapping windows in conjunction with a voting strategy, do not alter or generate new samples but rather further dissect existing samples to aid the model to extract more information from them. These two last approaches were also combined with a GAN in the work of Kugelman et al. [19], thus merging the two aforementioned paradigms into one single work.

In this work, we propose a fully automatic methodology to tackle this same problem of data scarcity in the field of DME, one of the main causes of blindness in developed countries and for which the availability of datasets is quite limited [20]. The retina is characterized for being the internal part of the human anatomy that is easiest to study, and its retinal vascularity and neural organization usually manifest symptoms that are just beginning to cause damage in other parts of the human body. This is the case of DME, where the fragile retinal microvasculature start to deteriorate and leak fluids in between the retinal layers, completely altering (and even destroying) its delicate morphology. If its not detected on time, it will lead to severe complications and the need for invasive treatments [21–23]. The main means to study this pathology is with the use of OCT imaging. This medical imaging modality allows for a non-invasive cross-sectional visualization of the retinal layers [24]. A comparison between a healthy retina and a pathological one with DME fluid accumulations can be seen in Fig. 1.

Thus, in this proposal, we exploit a GAN to generate new synthetic OCT samples of DME to complement the lack of real samples, as well as even to entirely replace a dataset (as we will show in subsequent sections). Recent works have applied GANs in the OCT domain with different strategies to produce synthetic samples. The basic idea for these works is to create new samples from a random noise seed, like the methodology proposed by Odaibo [25]. These works had the main issue that the resulting images were of low quality and the structures tended to be lost. That was improved by Zha et al. [26] by using SSIM or Structural Similarity loss to try to take into consideration structure features of the retina and by Zheng et al. [27], with a progressively grown GAN: a network trained initially with a low number of layers that was progressively grown and trained. This way, the gradients were not dispersed and could generate higher resolution images (256 × 256 pixels in that work). Another example of these methodologies is the work of He et al. [28], where they use a Least Squares GAN [29] to complement a classification problem. This raised the problem that the GAN generated samples are not labeled, so they used an alternative strategy: they assume the generated images do not belong to any of the considered categories, and use them more as a regularization strategy to increase the variance of the dataset. To all these labeling, stability and quality issues, we have to add the instability inherent to these learning strategies and the delicate balance needed to be achieved by the selection of hyperparameters. Some works in the field of machine learning try to solve these issues by offline learning strategies [30] that, during the learning process, vary their strategy to further optimize the learned strategies. For example, through the usage of bayesian-optimized strategies [31], others based on genetic algorithms [32], or even by the additional support of other networks, in special the based on Long-Term-Short Memory architectures [33]. This effect is also mitigated by the careful control of the batch size, and through the usage of normalization strategies as such is the batch normalization (which helps to smooth and stabilize the optimization process [34]). However, these issues mainly affect adversarial methodologies based on a randomized input strategy, as it relies heavily on the underlying context embedded in the hyperparameters. Strategies based on an input with domain information are able to maintain stability in the epochs thanks to the added independence from the learnt context. This is specially seen in architectures based on an image-to-image paradigm, as they precisely transform from one context to another, needing a reduced set of information to do so [35].

In our work, we present an alternative methodology by taking advantage on the retinal shape, limited by the Inner Limiting Membrane (ILM) and the Retinal Pigmented Epithelium (RPE) layers (presented in Fig. 2). Clinical studies have shown the direct correlation between the different deformations (and their absence) to the different types of fluid accumulations [36–38]. This means that, in theory, by only knowing the retinal region of interest limited by the ILM and RPE layers, we could infer the different fluid accumulations (as well as the lack of them) present in the retina.

This way, we use the aforementioned image-to-image translation strategy using as input a binary mask indicating a retinal region. This mask represents an easy way to obtain abstraction of the retinal layers that we can use to generate pathological or healthy retinas. This binary mask is enough to generate all the complex internal layers of the retina, artifacts and fluid present in the vitreous humor, choroidal structures and even decide what kind of pathological patterns will be generated. Moreover, given that this binary mask represents a high level abstraction of the structures in a retinal OCT image, with an strategy that can generate new binary retinal masks, we are able to generate a potentially unlimited number of completely unseen retinal OCT images. Finally, we evaluate our proposal with two strategies from the state-of-the-art that proved themselves in our clinical scenario: detection of DME in OCT images. One scenario where we train a proven convolutional network only with fully generated samples (and compare it to the same network with real samples from the state-of-the-art) and other where we test against a pretrained expert model from the state-of-the-art the performance of our synthetic samples as test subset compared to the real test subset.

In summary, the main contributions of our work are:

- Fully-automatic approach to generate synthetic retinal OCT images.
- Successfully able to create images with both pathological (DME) and healthy retinal patterns.

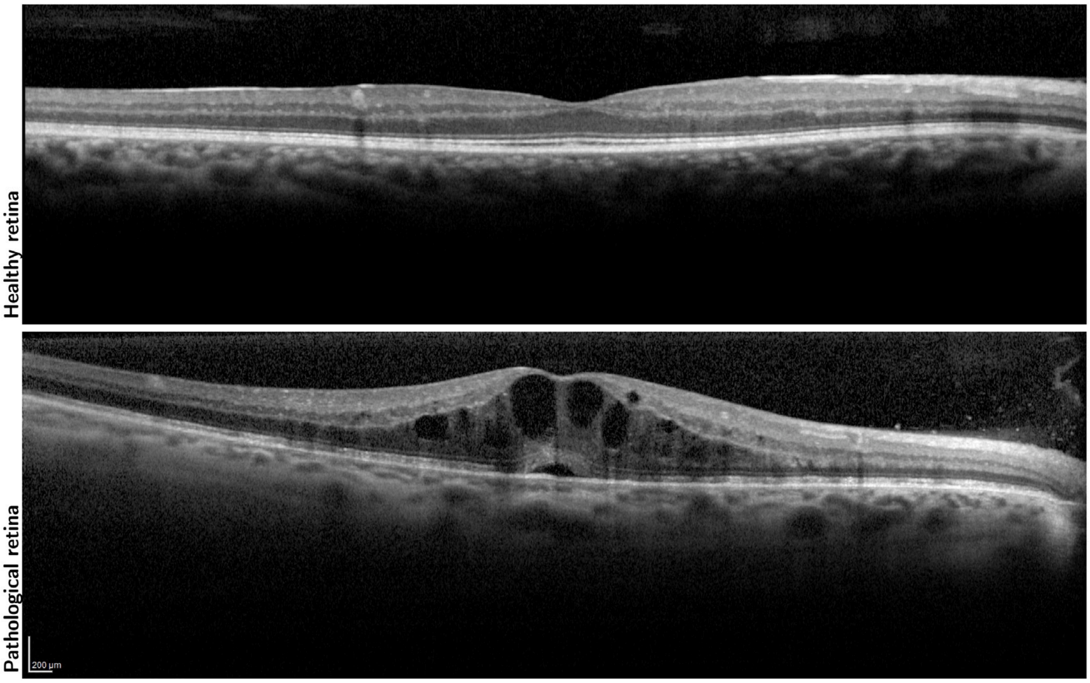

Fig. 1. Comparison between a healthy retina and one with multiple types of intraretinal DME fluid accumulations.

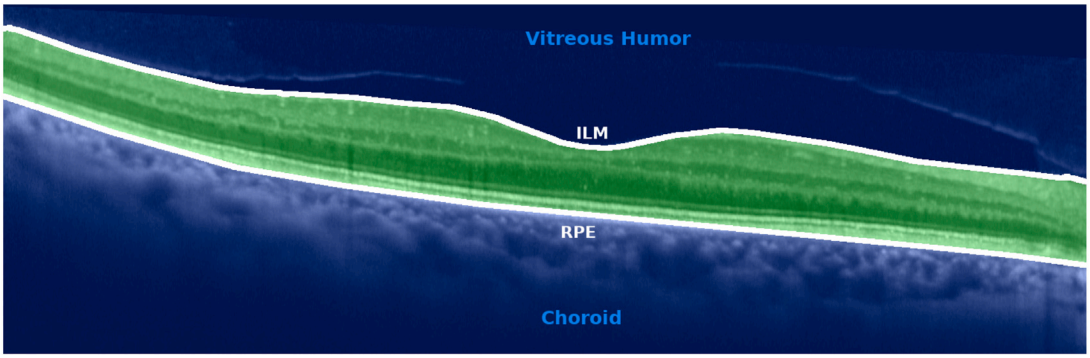

Fig. 2. Example of a retinal OCT image, with the retinal region marked in green, the Vitreous Humor/Choroid in blue and the limiting layers of the retina (ILM and RPE) in white.

- Tested with approaches from the state-of-the-art, obtaining comparable results to the metrics attained with real images.
- Generation using a context transformation paradigm, robust to the usual instability of the GANs based on randomized seeds.
- Able to generate high resolution fully synthetic images from a simple binary seed.

The present document is divided into main sections. Section 2: “Materials and methods”, presents all the resources needed to fully reproduce our work in detail, as well as a complete explanation of the algorithm and strategy followed in this work with the particular parameters for each experiment. Section 3: “Results” presents the outcomes of the training and comparisons with the state-of-the-art of our proposal. All these results are analyzed in Section 4: “Discussion”, where we comment different highlights of the results, weaknesses detected and their significance for the domain of the problem. Finally, Section 5: “Conclusions” includes a series of final notes drawn for this research and a commentary on future lines of work.

2. Material and methods

As previously indicated, in this section we will disclose all the necessary resources, procedure and different steps followed to recreate the experimentation and results of our work. This section is divided into three main subsections: Dataset (Section 2.1), where we explain the sources and characteristics of the used data in our experiment; Software resources (Section 2.2), where we present all the different libraries and software versions used in this work; Methodology (Section 2.3); where we explain all the steps of our experimentation; and, finally, Experiment configurations (Section 2.4), where we describe the precise configuration and parameters to allow the complete reproducibility of our work.

2.1. *Dataset*

The dataset that was used for this work was composed by 400 OCT images from a HRA+OCT SPECTRALIS© from Heidelberg Engineering, Inc. Originally, these images ranged in resolution from 512 × 156 pixels to 1535 × 497 pixels, ranging in aspect ratio from 1.03 to 5.75 pixels width in relation to height. All these images were captured from different patients, from both left and right eyes and with a wide range of configurations. All the images used in this work were resized to a resolution of 1024 × 1024 pixels.

The protocols followed during the development of this project were conducted in accordance with the Declaration of Helsinki, approved by the Ethics Committee of Investigation from A Coruña/Ferrol

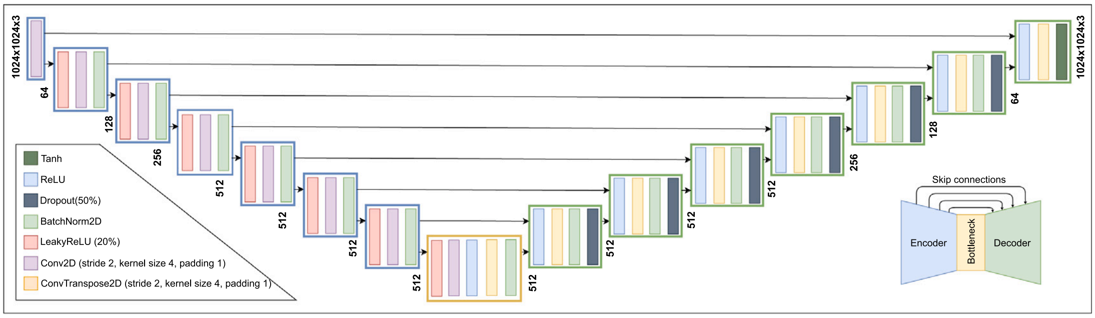

Fig. 3. Architecture of the U-Net 256 network.

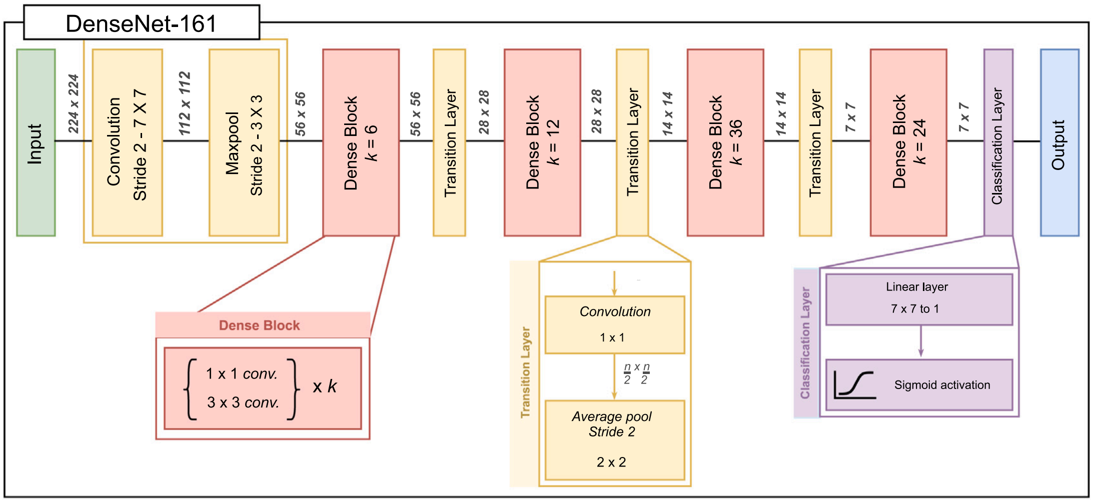

Fig. 4. Architecture of the DenseNet-161 network configuration.

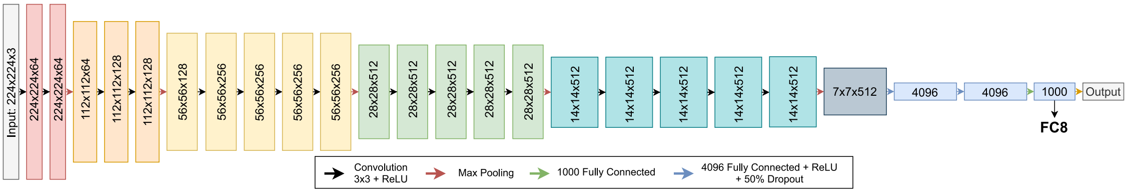

Fig. 5. Architecture of the VGG-19 network. Note the Fully Connected layer FC8 from where we will extract the features.

(2014/437). These images were labeled as normal (not presenting any fluid accumulations that could belong to DME) and DME (where there is a clear, beyond doubt, presence of intraretinal fluid belonging to DME).

2.2. *Software resources*

To develop this project, the generator is based on the pix2pix architecture, composed by an U-Net (Fig. 3) and a small encoder discriminator [10]. Additionally, we used an implementation of the DenseNet-161 [39]. Its precise architecture and configuration can be seen in Fig. 4. Finally, we also employed a pretrained version of a VGG-19 on the ImageNet dataset [40,41].

Regarding software resources, for the development of this project we used PyTorch 1.5.1 with torchvision 0.6.1 to train and test the DenseNet networks. The VGG-19 was tested using Matlab R2018a (9.4.0.813654) with the Deep Learning Toolbox and the Deep Learning Toolbox^{TM} Model for VGG-19 Network support package. The precise architecture and configuration of the used network can be seen in Fig. 5.

2.3. *Methodology*

In the present section, we will proceed to explain the proposed methodology as well as the approaches that were considered to demonstrate is capabilities against the state-of-the-art. In Fig. 6 we present the main generator stage, as well as the two complementary validation approaches where we test its usefulness and robustness.

In the first stage (Section 2.3.1), we will train a GAN architecture to generate realistic OCT retinal images from empty retinal binary masks. This generator will receive as input binary retinal masks and will infer the inner retinal structure to generate a realistic representation of a retina. We use an algorithm to deform binary masks so we can obtain new samples from knowledge already present in preexisting images. The dataset will be divided in this stage into training and test, so further stages of the methodology are not biased by additional knowledge gained by the generator.

In the second stage, we validate our generator with a proposal based on a screening problem. More precisely, in the identification of the samples belonging to retinas with DME fluid accumulations and to a normal retina. In this approach (Section 2.3.2), we compare the

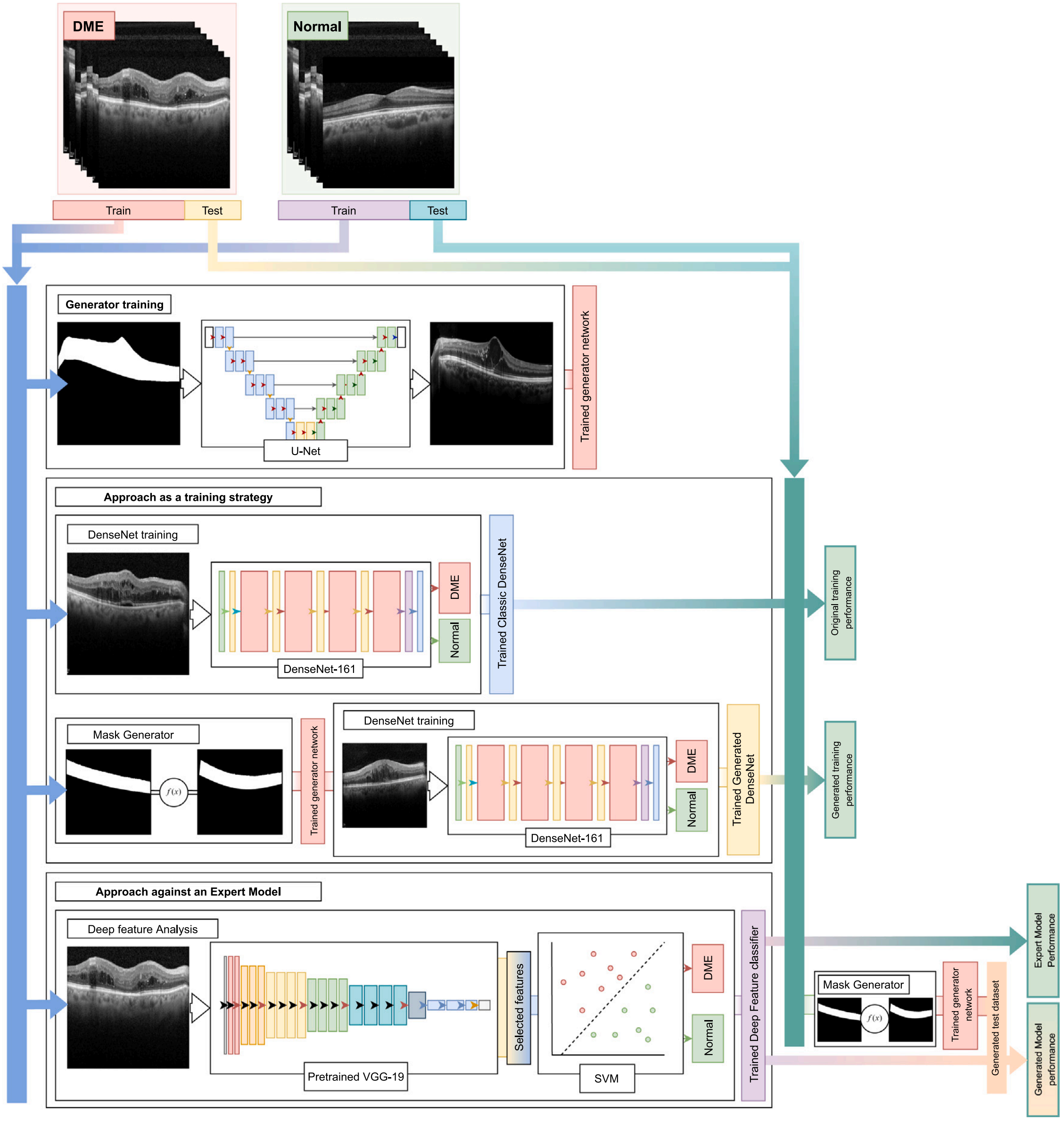

Fig. 6. Diagram of the methodology with the validation comparing it with the state-of-the-art.

performance of a representative densely connected network (proven in the state-of-the-art to be able to successfully discern healthy OCT images from pathological ones [42–46] trained with a real dataset and its performance trained with completely generated samples from our proposed methodology.

Finally, in the third stage (Section 2.3.3), we use another paradigm of machine learning (called deep feature analysis) from de Moura et al. [47]. Here, in the same medical screening context as in the previous stage, we use the generator to compare the separability the trained expert model that proved its capabilities in the state-of-the-art confers to a real test dataset with the separability it is able to confer to a set of completely generated images from our proposal. This way, we can study the variability and complexity of our methodology. In the first stage, we evaluate if our proposal can be used to learn the problem it itself tries to describe (that is, if the dataset has coherent information, as we test it with an strategy that has already proved itself to be able to with real images and achieve outstanding results). On the other hand, we also test if the variability of our proposal is similar to the original dataset, as the pretrained methodology from de Moura et al. should return similar error results as when using synthetic images. If the images are not representative, this expert model trained in the state-of-the-art should return significantly different metrics. Either better (thus, indicating that the images generated are easier to identify and, thus, not representative of the variability) or worse (and, thus, indicating that the images lack relevance to the issue as no patterns from the features learnt by the model were found in our proposal). Below, we will further explain each of the stages.

2.3.1. *Generator training*

The methodology followed to train the generator GAN is presented in Fig. 7. To train the generator, the main idea is to first train, at the same time, a discriminator network that will find the differences between the real images and the generated ones. Both networks (the generator and the discriminator) would enter a continuous competition, further improving the quality of the generated images. In our case, we used a basic encoder discriminator architecture, presented in Fig. 7, in the “Discriminator architecture”. The loop is repeated a fixed number of epochs, always following the same steps. First, we train the discriminator so it is able to learn new relevant features of the target domain to find if the image is real or generated. Then, we train the generator

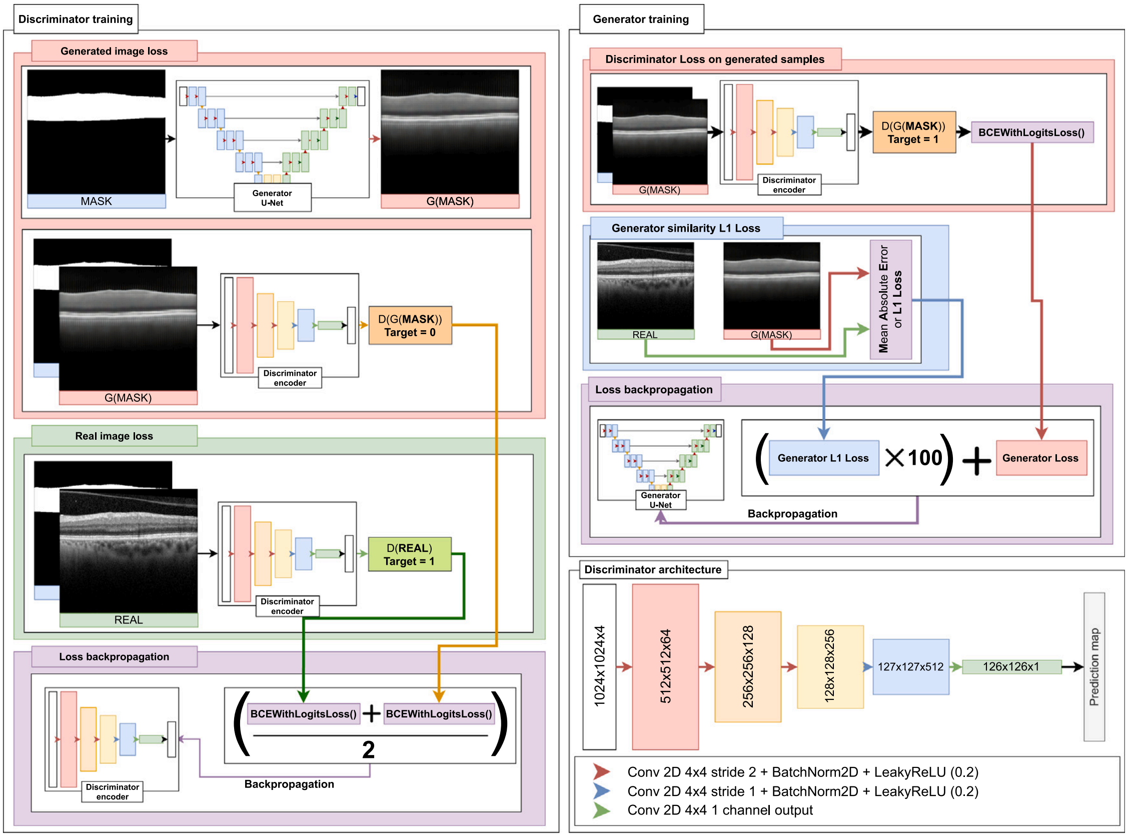

Fig. 7. Steps followed for each iteration to train the image-to-image network and Discriminator architecture.

so it is able to learn these new features from the discriminator. Both generator and discriminator have their own independent loss metric, depending on their point of view. The discriminator will use a binary loss, based on a classification problem where it combines the loss of the success (real image) and a failure (error made identifying the generated image). On the other hand, the loss of the discriminator is a bit more complex. This loss will combine a classification error similar to the one used by the discriminator and a similarity metric, that measures how much it the generated pixels have deviated from the original sample. These two losses: the binary classification loss and the similarity error loss are weighted and merged together.

To train the discriminator, we first use the generator to obtain what information of the domain have both abstracted (in the first iteration, this will be no information at all, only random noise, but will set a baseline for the generator to improve from). During this first generator step, the gradients of the generator are ignored, as we do not want to train it for now. Then, both input and output are fed to the discriminator (the mask and the generated image). They are fed by merging them as new channels in the image (what would be its third dimension) so the discriminator is not only assessing if its real or not, but also if it belongs to the input binary mask fed to the network. As what we are training is the discriminator, the loss will be calculated indicating the target of this image as a false (generated) image. Afterwards, we repeat this process, but this time with a real image and setting the target as real. The final backpropagated loss through the discriminator will be the average of both real and generated losses. When the discriminator iteration has ended, we perform the optimizer step, and the weights of the network are updated.

Then, after the discriminator training part of the iteration, we proceed to train the generator. The first step is similar: we calculate the loss on the generated samples. But, this time, instead of labeling the sample as generated like in the training of the discriminator, we label it as if it was real (albeit ignoring the gradients product of this forward pass through the discriminator, as we are not training it). This is because the loss of the generator (that is, what has it done wrong) is the distance between the discriminator considering the generated sample as real (what is needed to trick the discriminator) and the prediction it actually returned. Additionally, we use an additional loss component to force similarity between the original image and the generated one. Thus, the final loss is calculated as the sum of the generator loss and its similarity loss times a given weight to regulate the impact of the forced similarity. Finally, it is backpropagated through the generator U-Net, the optimizer step is performed and the weights of the generator are updated. Take into consideration that each network is using a different optimizer, as they are, effectively, two independent networks. This process is repeated through each iteration until the end of training a fixed number of epochs.

To generate the samples to use in subsequent steps of the methodology, we need to have the same level of complexity as in the original dataset, but different enough to prevent the generator to use memorized information instead of learnt one. Additionally, we need to know the features that characterize a pathological retina so we can generate samples that properly represent the biomarkers of a given label (so we can generate samples at will with the label we want). To achieve this, we use the original masks, but apply different transformations to obtain a warped version of them. This way, the model will not take advantage on features used during the learning process, but will maintain the pathological biomarkers so the label of the generated samples are known. This warped version is achieved by first applying Eq. (1) to each pixel in the image:

$$y(x) = abs(\sin(\frac{x \cdot \pi}{1024}) \cdot (-a)) + d \tag{1}$$

where a is the amplitude of the sinusoidal wave used to warp the retina and d the displacement of the wave in the y axis. As seen in Fig. 8, the generated retinal images preserve enough information from the original images, but also are warped enough so the deep neural network has to

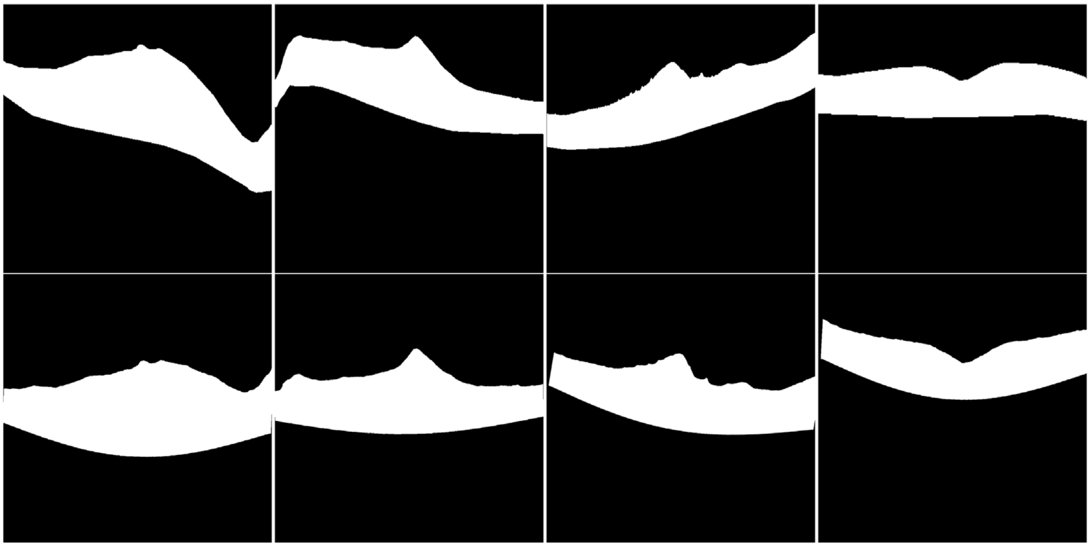

Fig. 8. Original retinal region of interest masks (1st row) and the corresponding warped masks (2nd row).

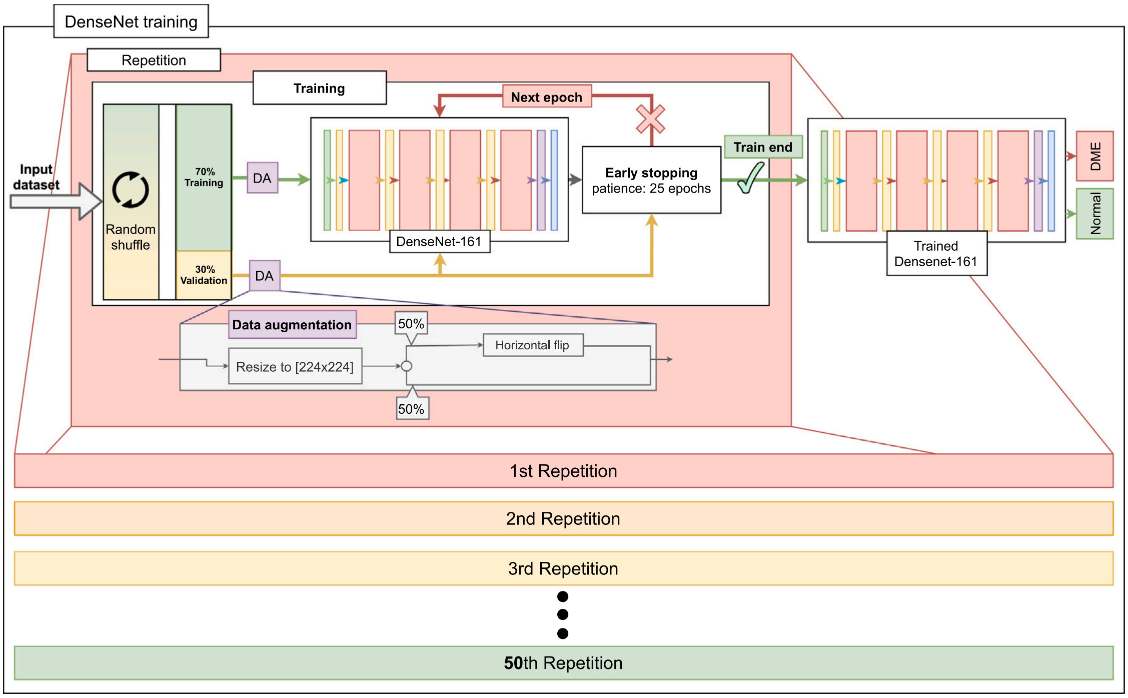

Fig. 9. General DenseNet training strategy used with both the generated and original datasets.

infer new information to generate new retinal regions. Additionally, a small binary closure morphological operation is applied to each mask to diminish the effect of the image rotation and rescaling in some images. It also helps to change the small morphological markers in the innermost retinal layer (like small vessels protruding from the top), helping to add variability to the final generation.

2.3.2. Approach as a training strategy

In this first evaluation stage, we will compare our work against a classification strategy from the state-of-the-art. The chosen architecture to this effect is the aforementioned DenseNet-161 as it has proven its reliability and success in the subject of a clinical screening for DME. As shown in Fig. 6. We will train the system with a real dataset and another using only generated samples. The strategy used to train both networks can be seen in Fig. 9.

As seen in this figure, we repeated the training to obtain robust and significant statistical values in the final metrics. For each epoch, we randomly shuffle the dataset (remember that a test dataset was previously separated from the original, common to all the steps in the methodology so we can perform a fair comparison). These samples are afterwards forwarded through the aforementioned DenseNet-161 architecture. Additionally, we use an external scheduler, which regulates the learning rate depending if the training has stagnated or not. This is done by checking if the validation loss is still decreasing after a given number of epochs or it does not reach a new local low. This approach was chosen instead of the linear descent used in the training of the generator as it allows to explore better the search space of the domain. In the generator, due to the nature of chasing each other of both networks, we could not guarantee that the evolution seen in them was not product of a circular chase that would lead to nowhere. In this case, as we are training only one network with a clear objective, there is a clear defined task to optimize. The same way, instead of choosing a given number of epochs to stop the training, we use an early stopping strategy. This strategy, same as the loss regulation strategy, will stop the training of the model if it the validation loss has stagnated even further. Additionally, every time a new best validation loss is reached, the model is chosen as the candidate, and every time the loss is readjusted, that same checkpoint is recovered to continue training in the most promising stage of the training process.

The second model, as mentioned, is trained the same way as was the previous one, with the difference that the masks used to train the system are generated following the steps presented in Section 2.3.1 with the train dataset images.

2.3.3. Approach against an expert model

In this second validation stage against the state-of-the-art, we use an expert model methodology based on deep features and test it with both images from the real test subset and also with images that were generated with our expert model. We used the proposal of de Moura et al. [47], as indicated in Fig. 6. This methodology takes advantage of the features learnt in the deep layers of the convolutional neural network in a complex domain to classify a simpler one. In this case, we will use an VGG-19 network architecture (as defined in Section 2.2) pre-trained in the ImageNet dataset. This network has learnt, in its internal layers, different texture and morphological features that it uses to classify more than 1000 different classes. What we do is extract the features learnt by the network in the fully convolutional layer 8 (FC8, as shown in Fig. 5), select the most relevant ones with a feature selection strategy and, then, use them to train a simpler classifier. This strategy allows to train a model with a low number of images, with a reduced use of resources and with greater abstraction capabilities thanks to using a network trained in a open-ended image domain.

After the expert model is trained as indicated by de Moura et al., we generate new test images using the strategy explained in Section 2.3.1 using as seed masks from the test dataset (further preventing any bias, since we are both generating the images and also using masks that were never seen by neither the generator or the expert model). Finally, we compare the real test results from this methodology of the state-of-the-art and the generated test dataset to obtain a performance evaluation of the model and analyze its behavior. If our system is able to obtain similar results to the expert model, it means that it has successfully learnt the retinal patterns, being indistinguishable from real retinal images for computer-aided diagnosis purposes.

2.4. Experiment configurations

In this section we will now present the precise configuration and parameters needed to fully replicate or methodology and experimentation performed in this work.

2.4.1. Mask generator

As mentioned in Section 2.3 this step, the dataset was divided into what would be used for training and testing of our methodology. The division chosen for our work is 70% for the training and 30% reserved for all the testing purposes. This last set of samples were not used in any step of the methodology but for statistical purposes. This way, from the 400 original images, 280 will belong to the training dataset, and 120 to the test dataset. All the images were randomly chosen from both positive and negative samples, maintaining the balance between the number of positive (120 + 60) and negative (120 + 60) OCT images in both sets.

To train both the generator and the discriminator we used as main loss the Binary Cross-Entropy (BCE) with a sigmoid layer, defined in Eq. (2):

$$\begin{aligned} &BCE(x,y) = mean(L), L = \left\{l_1, \dots, l_N\right\}^T, \\ &l_n = -w_n[y_n \cdot \log \sigma(x_n) + (1-y_n) \cdot \log(1-\sigma(x_n))] \end{aligned} \tag{2}$$

where N represents the batch size and w_n the weight (which, in our case, was not considered). Additionally, as mentioned, to train the generator, a similarity loss component is added to the loss value so we have a force that pressure the synthetic image to preserve information from the original sample (as well as aiding to converge faster). In this case, the L1 Loss is used to measure this distance between the real image and the generated image (called in Fig. 7 the "Generator similarity L1 Loss"). With this, we compare the mean absolute error (MAE) between each element of the real image and the generated image [35], defined in Eq. (3):

$$MAE(x,y) = mean(L), L = \left\{l_1, \dots, l_N\right\}^T, l_n = |x_n - y_n| \tag{3}$$

where N represents the batch size. This loss is summed with the generator loss, weighted so we can regulate establish the magnitude of the effect of this similarity force. In this experiment, was established at 100. For the training of both discriminator and generator, we used the Adam optimizer [48]. As both models are trained independently, each of the models uses its own independent optimizer.

The learning rate was set as a baseline of 0.0002, maintained static for 100 epochs. This way, we allow the generator to find a general abstraction of the images. Afterwards, the loss value is updated linearly following Eq. (4) during 500 additional epochs until zero:

$$Lr(n) = 1 - \max(0, n + n_f)/(n_d + 1) \tag{4}$$

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.

For the generation of the binary masks for posterior use of the generator, the values used in the formula presented in Section 2.3.1 are randomly set for each instance. The amplitude (a) is set as a random integer between 0 and 150. The displacement (d), as a random integer between −100 and 100. This ensures that the retina in all the warped masks maintains a valid shape and the conditions for a correctly captured retinal shape. Additionally, to add more variability, each mask is rotated a random angle between −15 and 15 degrees (a feasible rotation in this medical imaging domain). The kernel chosen to smooth the resulting masks to palliate the effects of the resizing and rotation was a disk kernel with 20 pixels of radius (so the effect on the image would be the same whatever the resulting rotation angle).

2.4.2. Approach as training strategy

First, the dataset was divided 70% into training and 30% into validation. To increase the efficiency of the methodology, all samples are resized to 224 × 224. Additionally, to increase the effective number of samples present in the dataset, we perform a data augmentation consisting in randomly flipping horizontally the samples with a 50% chance. This is done as the retina can present the same samples in both east and west directions, commonly due to belonging to the right or left eye. Thus, horizontally flipping the samples is actually showing new samples that could perfectly appear in real retinas.

To train both DenseNet strategies, we used as loss the BCE loss, defined in Eq. (5)

$$\begin{aligned} &BCE(x,y) = mean(L), L = \left\{l_1, \dots, l_N\right\}^T, \\ &l_n = -w_n[y_n \cdot \log(x_n) + (1-y_n) \cdot \log(1-(x_n))] \end{aligned} \tag{5}$$

where N represents the batch size. As optimizer, we are using the Adam optimizer with an initial learning rate of 0.01. This learning rate, unlike the linear approach considered in the generator, was changed dynamically depending on the improvement or stagnation of the model. The patience for this dynamic loss is set as 10 epochs to determine the stagnation of the validation value. Thus, if the validation loss has not improved in 10 epochs, the learning rate will be halved each time until the end of the training. In this case, the early stopping patience has been set at 25 epochs, so the training has enough epochs in the middle to further adjust the learning rate and allow the model to fall to a new

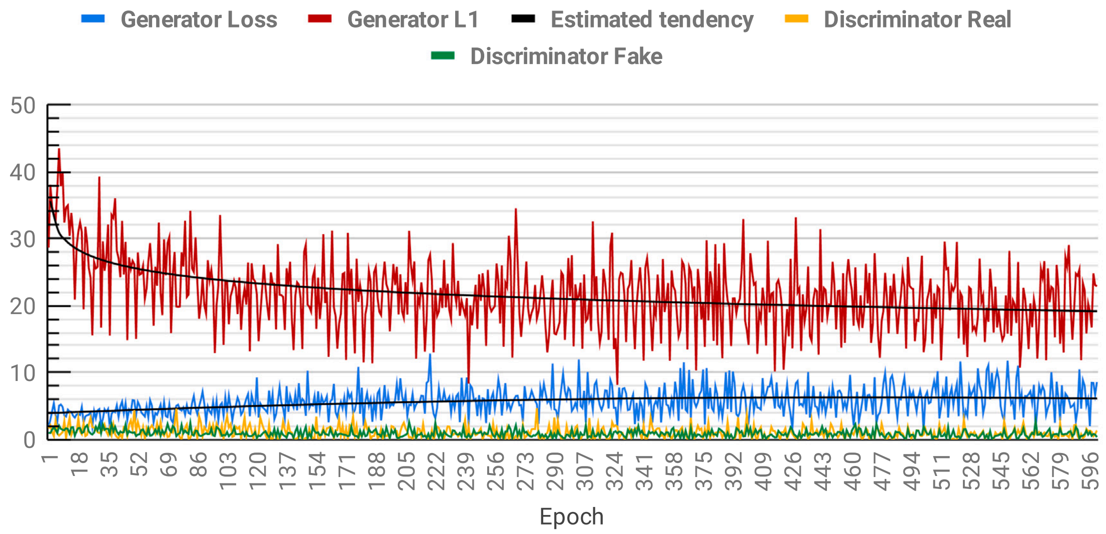

Fig. 10. Metrics during the training of the generator network.

lower point of the decision space with higher precision (preventing thus reaching a non-minimal stopping point). For this experiment, a batch size of 3 was chosen, as in preliminary tests it attained satisfactory results. The training process was repeated 50 times to ensure the robustness of the final metrics.

2.4.3. Approach against an expert model

In this case, as mentioned, we followed the proposal of de Moura et al. [47] to train the expert system from the state-of-the-art to which we will compare our proposal with. This way, we used the pretrained model, and only the feature selection of the original work was used. This way, we extracted the selected 532 features from the FC8 layer. These features were selected using the Relief-F algorithm [49] (53.20% of the features contained in the layer). In this layer most of the features have already been abstracted enough so a simple linear Support Vector Machine [50] using the Sequential minimal optimization [51] solver could obtain satisfactory results without a kernel trick, allowing for better generalization capabilities. So, to classify the final samples, we will first process it through the pretrained network. Then, we extract the selected features from the chosen layer. Finally, we classify these abstracted features by means of the SVM, obtaining the final result to evaluate. As we used the trained model of de Moura et al. [47] to allow full comparison, please refer to cited work for more information on the hyperparameters and configuration of the classic classifier.

3. Results

In this section we will proceed to present the results obtained in each of the experiments carried out in this work (both the proposed strategy and the two stages of comparison with the state-of-the-art works) using the configuration presented in Section 2.4. For the convenience of the reader, the same subsection structure was maintained, as per in Section 2.3.

In order to evaluate the performance of the models, we employed a combination the metrics shown in Eqs. (6) to (9), where TP are the True Positives, TN the True Negatives, FP the False Positives and FN the False Negatives. These metrics were chosen to ensure a complete analysis from different points of view of the results.

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$\textbf{Precision} = \frac{TP}{TP+FP},\quad \textbf{Recall} = \frac{TP}{TP+FN},\quad \textbf{Specificity} = \frac{TN}{FP+TN} \tag{7}$$

$$\textbf{F1Score} = \frac{2\times \text{Precision}\times \text{Recall}}{\text{Precision}+\text{Recall}} \tag{8}$$

$$\textbf{MCC} = \frac{TP\times TN - FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{9}$$

These measures are the *accuracy* (or percentage of correctly classified samples), *precision* (or percentage of positives returned by the system that are real from the total positives returned), *recall* (or percentage of true positives returned by the system from the total of positives available), *sensitivity* (or percentage of true negative samples over the total available detected by the system), F_1 *Score* (or the harmonic mean of the precision and recall), and *Matthews correlation coefficient* or MCC (the correlation between the true labels and the labels returned by the system). This last score is between −1 and 1, as it is based on the correlation coefficient of Pearson.

3.1. Evaluation of the generator training

In Fig. 10, we can see the results of the progressive training of the generator. As shown by the estimated tendency plotted over the “Generator L1” function plot, the similarity loss between the generated image and the seed stagnates early on, reaching no significant variability from epoch 290 onwards. Thus, at this point, the system has successfully replicated the retinal structure. As this epoch represents an stagnation in the structural similarity, the main driver of the generator will be the “Generator Loss”. This means that the last epochs the system would act as a normal GAN, just focusing in tricking the discriminator network, which shortly after it also stagnates into an stability region around the same amplitude loss spikes.

In that same figure, we can see how the inherent instability of the GANs can be seen, with a significant variability during the training of all the considered metrics. However, we have also added the estimated tendency to show how, despite the inherent variability of the GANs, our proposal is actually oscillating around a descending point of equilibrium. By using the retinal shape as seed for the generated image we are severely constraining the possible outcomes of the methodology. Thus, this local variability is caused by the changes in the generated textural patterns. However, fixed structures that are dependent on the retinal shape (such as the layers, fluid structures and other pathological artifacts) are a constant provided by the mask itself. Thus, as shown by this progression, the inherent stability of the GANs is actually severely reduced by our proposal. Moreover, this is precisely what allowed our proposal to be able to generate high resolution images

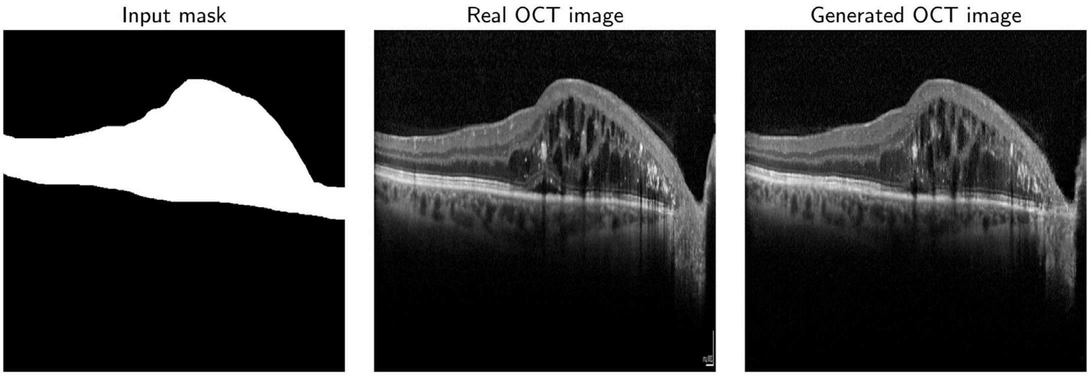

Fig. 11. Example of input, output and real image of the generator in training epoch 599.

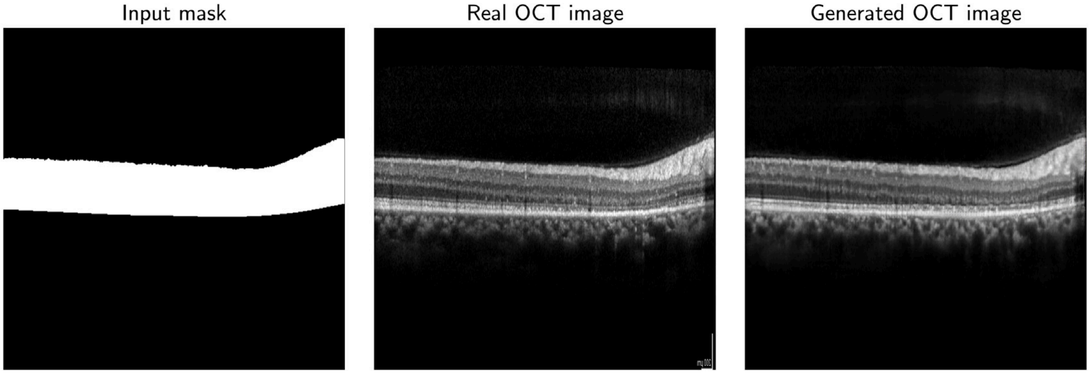

Fig. 12. Example of input, generated image and real image of the generator in training epoch 500.

while also maintaining this aforementioned robustness and stability. In the following sections we will analyze further which elements of the OCT images affect this variability, but mostly are those who their position is not determined by the retinal shape (such as elements from the graphical interface of the device).

This way, and by examining the resulting generated images in the last epochs (Fig. 11) where even textural patterns are replicated (as well as major structural artifacts), we decided to use an earlier epoch as the generator to be used in further steps of our proposal. As shown in Fig. 12, we chosen the generator model of epoch 500, as it maintains texture patterns while presenting minor variations that indicate the possibility of generalization of the model.

3.2. Evaluation of the approach as a training strategy

In Figs. 13 and 14 we present the training process of both the DenseNet trained with real samples and the one trained with generated samples. As can be seen in both graphs, both models have achieved satisfactory accuracy values. The main shown difference is that, in later epochs, the model trained with generated samples started to overfit the model to the training dataset. This is an indicator of what we will further comment in the discussion section. An early sign that, while our proposal retains most of the variability, is at the fine-grained level where the variability is mostly lost. That is, in fine texture patterns learnt in later stages of learning the model starts to be hindered by the lack of information while, in the real dataset, there is still an slight room for improvement (albeit negligible).

Nonetheless, in Table 1, the reader can see that the proposed methodology, using exclusively generated samples, was able to achieve 94.83% of the accuracy of the expert model.

Table 1
Mean accuracy and standard deviation for all the DenseNet training repetitions trained with both the real and the generated dataset and tested with real images.

	Mean train	Mean validation	Mean test
State-of-the-art	0.9574 ± 0.0417	0.9767 ± 0.0175	0.9315 ± 0.0203
Synthetic dataset	0.8956 ± 0.0521	0.9300 ± 0.0212	0.8833 ± 0.0328
Relative accuracy	93.55%	95.22%	94.83%

3.3. Evaluation of the approach against an expert model

In Fig. 15 we present a visualization of the main metrics that describe the behavior of our proposal in this state-of-the-art scenario. In this plot, we can see how the expert model was able to successfully separate both classes, achieving even better results than a standard Deep Learning approach presented in the previous section. The same way, when testing the model with images generated by our methodology, the expert model is able to obtain almost the same statistics as they were obtained with the real dataset.

This shows that our methodology is not merely recreating what it has seen before with deformed masks. The masks used as seed in this step were not seen before by the model. All the results obtained are purely unbiased and the generated masks are completely original, not giving clues to the methodology from the training dataset. Additionally, in the metrics presented in Fig. 15, we can see that the specificity of the synthetic dataset is on par with the generated dataset statistics; but the sensitivity of the results are slightly lower (a 96.49% of the sensitivity reached by the real test dataset).

As seen in Fig. 16, this sensitivity is result of two samples being misclassified as normal samples, while their mask was set as being pathological. We will further expand these ideas in the following discussion section, as well as possible causes and solutions. In Table 2 we include the comparison between the rest of metrics. As shown by

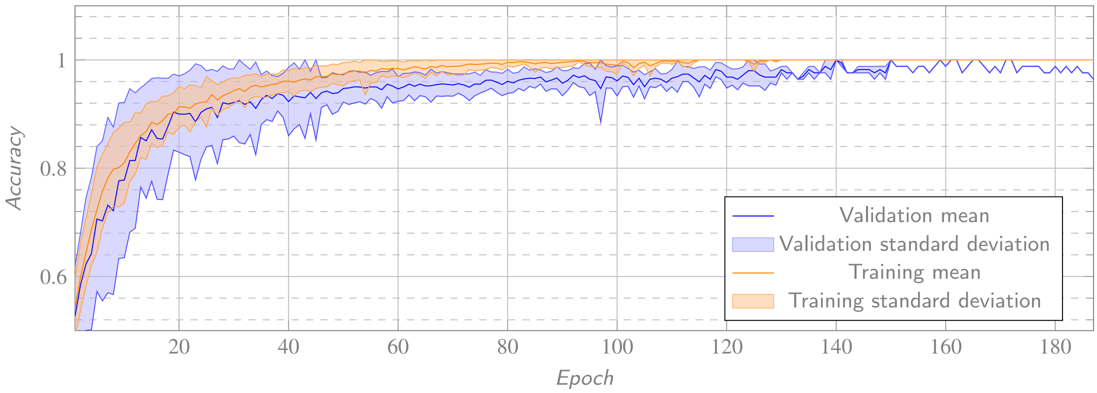

Fig. 13. Mean and standard deviation of the validation and training accuracy per epoch for the model from the state-of-the-art (control).

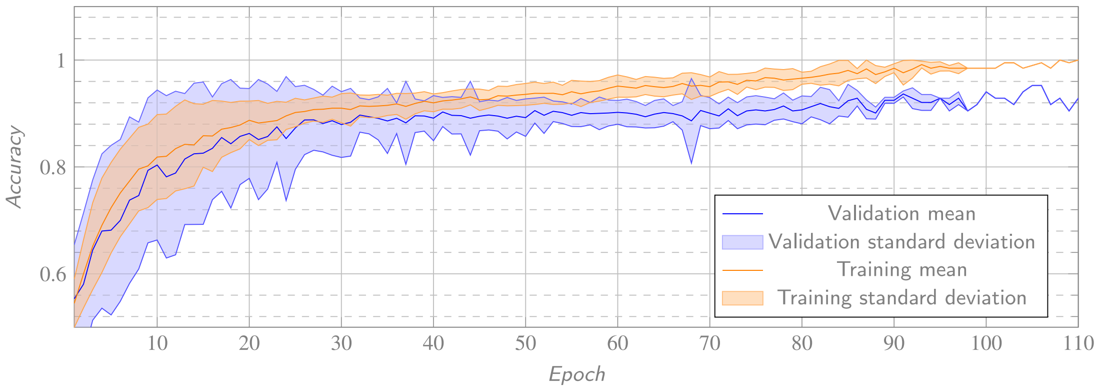

Fig. 14. Mean and standard deviation of the validation and training accuracy per epoch for the model trained with generated samples.

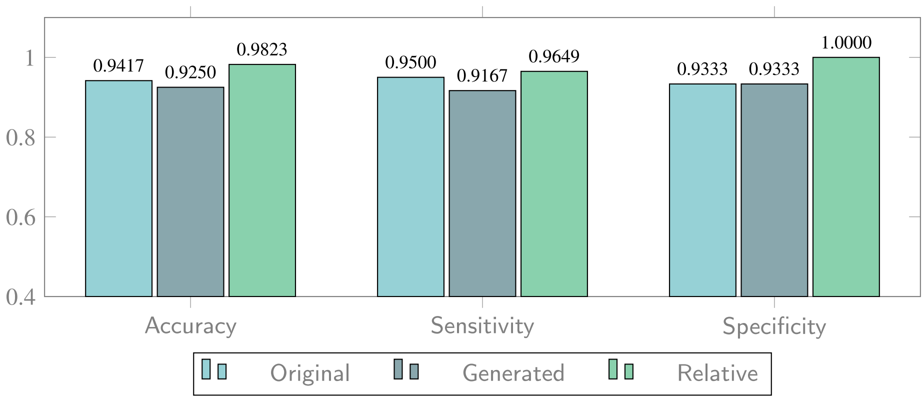

Fig. 15. Metrics obtained with the first state-of-the-art model using both the generated and original test datasets. The relative metrics (generated/original) are also presented.

de Moura *et al.*

Target		Predicted Normal	Predicted DME
Target	Normal	56	4
	DME	3	57

Synthetic test dataset

Target		Predicted Normal	Predicted DME
Target	Normal	56	4
	DME	5	55

Fig. 16. Confusion matrix for both the expert model of de Moura et al. and the test results using our generated dataset.

the confusion matrices, we see that the most significant discrepancy between the proposed approach and the original work from the state-of-the-art is on the recall. That is, the positive samples generated by our proposal are being misclassified slightly more than in the real scenario. However, we see that, overall, the results are comparable to the state-of-the-art, being still completely generated. The same way, and as presented in the specificity, our proposal is able to generate healthy samples indistinguishable from the real OCT images, as the state-of-the-art work performed the same with our synthetic images than the real ones.

Table 2
Comparison between the results of the expert model of de Moura et al. and the results when trained with our synthetic OCT dataset.

	Accuracy	Precision	Recall	Specificity	F1 Score	MCC
de Moura et al.	0.9417	0.9344	0.9500	0.9333	0.9421	0.8835
Synthetic dataset	0.9250	0.9322	0.9167	0.9333	0.9244	0.8501
Difference	0.0167	0.0022	0.0333	0.0000	0.0177	0.0334

4. Discussion

First of all, we would like to remark the results obtained in the comparison with the state-of-the-art. As we saw in the first comparison, in test, a system trained exclusively with images generated with our proposal is able to achieve a comparable result with the original work trained with real images. In fact, we see that we reach 94.83% of the accuracy of the methodology trained with completely real data. This slight reduction in test accuracy is possibly due to the fact that the variability of the generated images, in some situations, is diminished. While it is correct that the shape of the retina is representative of the present accumulations within it, it is also possible that different shapes of accumulations may result in a somewhat similar generated retina. This results in that, for the same mask, the generator network has found a consensus between the variability for the same retinal shape and the variability of the fluid patterns. For this reason, this slight reduction in variability due to the surjective (but not injective) nature of our proposal is reflected in a small reduction in the training contribution. However, this can be easily solved by adding information to the binary masks that would allow better determination of the inner fluid types. We have to think that right now it is inferring the entire internal structure based on a binary mask. With the mere addition of simple weak labeling we could get much more consistent images that would give even more competitive results.

On the other hand, when we compare our proposal to an expert system already trained and with features from the state-of-the-art work, we see that the variability provided by our dataset is indeed competitive and comparable to that of the original state-of-the-art work. Moreover, in the comparative metrics presented, we see that the main metric affected is recall. That is, of the total number of positives generated by our system, a negligible amount of them is ignored when compared to the state-of-the-art model. This precisely is due to what we explained before. This behavior is specially seen and justified in the presented specificity results of that same comparison with the state-of-the-art, which obtains an indistinguishable performance with respect to the state-of-the-art. Thus, we see how it is able to perfectly capture the shape of the retina and provide as much information as the real images. Moreover, it is able to generate pathological images with realistic positions of cysts even though there is no explicit information in the binary mask about their location, extent or type.

Continuing with the generated images, we would like to highlight about the images generated by our system is its surprising capability to recreate the same noise distribution present in this medical imaging modality. In Fig. 17, we can observe how in both the original image and the image generated by the system there is (especially in the vitreous humor) a grainy pattern of almost constant intensity. This noise pattern is a distinctive feature of the capture device, being considerably different in other OCT devices (depending on the configuration, reconstruction strategy to generate the OCT images and characteristics of the tissues and fluids of the examined patient). Moreover, as mentioned a before, we can see that the system was able to completely recreate the inner structure of the retinal layers using only shape information provided by the binary mask. In fact, we see how the foveal region (represented as a depression in the center of the retinal region) has been correctly reproduced. There, we can clearly recognize how the Henle's fibers layer (axons of neurons that expand from the fovea) are correctly detected and generated by the proposed methodology.

In the same way, we see how the vascular structure that nourishes the retinal tissues (the choroid) has also been perfectly represented in the inferior section of the retina. Moreover, for both cases (both inner retinal and choroid regions) present similar high-level patterns, but low-level texture and artifacts have been realistically generated. That is, the network has managed to infer the features that depend solely on the retinal morphology, but has generated and correctly filled those that do not alter the retinal shape (and, thus, cannot be predicted). Thanks to this, we can see how this methodology is not only applicable to the field studied in this work, but could be easily transferred to other pathologies and imaging modalities.

On the other hand, one of the main aspects to note about the training of the generator is its evolution along the different training epochs. In Figs. 18 & 19 we present the evolution of two images belonging to the normal and DME class along different epochs. As can be seen in these images, the generator quickly learnt to complete the retinal structure from the binary mask. Nonetheless, the finer grain of the texture is only seen in latter epochs, as in earlier epochs we see repetitive artificial patterns (such as the white hyperreflective spots randomly distributed in the epoch 100 for the DME class and on its noise) and textures that lack the richness and coarseness of a living tissue, such as seen in the DME image, epoch 300. As saw in Fig. 10, the model reached an stability point where the generator and the discriminator achieved their maximum potential, and from this point they were pursuing each other by changing minor details in the images (and even overfitting). The real images and the generated OCT images presented features that could not be determined by the shape of the retina alone. For example, the hyperreflective dense spots (shown as white structures inside the retinal layers) in the generated retinas were situated in the same positions as in the original retinas. They could have been placed in any other place without interfering with the retinal shape as much as any other fluid accumulation or pathology would. This is a clear sign that the network was learning to draw retinas using the retinal shape as reference for a memorized pattern, and not as a seed for the generator. This is precisely what we mentioned in the results section. Our proposal is able to successfully train the methodology independently of the inherent instability common to the GANs thanks to the chosen input. As the generated masks contain enough information to generate most of the structures, the real variability appears when considering the internal texture patterns. As mentioned, these patterns oscillate around a given tendency, but still learning the general layer structure that define the retina. For this reason, when these main structures are achieved, the system enters in a pendulum-like phase, continuously fitting to texture patterns with no real improvement for the overall model.

This is also seen by the scale that indicates the relationship between the size of the structures in the retina (in microns) and the pixels shown in the OCT image. As seen in Figs. 18 & 19, the generator seems to continuously try to both recreate and dissimulate this scale during the training of the model. This is normal, as it represent a very representative (and easy to find) feature of the real retinal images, so its easy to view why the discriminator forced the generator to sometimes go back and try to generate this ruler. Nonetheless, the generator does not have any indication on where this scale can appear, as the images are horizontally flipped the ruler moves from right to left and the input binary mask only contains information about the retina. While this does not impact the final quality of the image, it could easily be removed by removing this rule from all the training images, so the discriminator of the GAN does not center its attention on it and force the generator to recreate this part of the images.

Additionally, as the images are resized, the scale is also deformed and pixelated (as it is only a few pixels wide). We can see how, in the healthy images for example, said scale is generated in epoch 100, but cleverly dissimulated in posterior epochs with the noise (as well as what happened in epoch 500). It is clear that, at this point, the discriminator was taking advantage on the knowledge that the

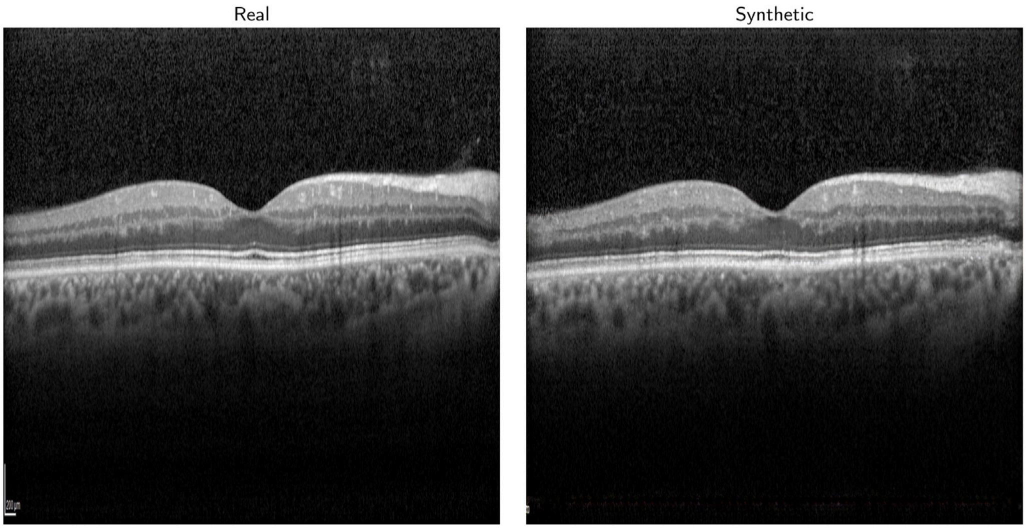

Fig. 17. Example of a real image and a generated test image for a healthy OCT example.

Table 3
Summary of the contributions and the analysis of the results.

Advantages of our proposal	The proposal is able to robustly generate synthetic images. When validated as a means to train a classifier strategy from the state-of-the-art, the model was able to successfully perform as well as when trained with the original dataset. The validation against an expert system from the state-of-the-art demonstrated that the variability and information of our generated images are on-par with the real data. The retinal mask thickness and shape are representative of the pathology, and allow to properly generate both pathological and normal OCT images. The proposed input is a robust descriptor of the problem, providing stability to the training. This allows the generation of high-resolution images with a reduced computational cost.
Weaknesses	User interface elements from the device are misplaced as no reference is given in the input. In earlier epochs, the textures present severe smoothing of the textures and coarse in later epochs. Early-stopping strategies should be considered to prevent this phenomenon. The model relies on the binary mask input to infer the shape. Thus, if the input mask is not realistic, the errors are propagated to the final generation.

ruler was possibly most of the time in the lowest part of the image, so the fake samples would be the only ones with that scale in the vitreous humor. Thus, the discriminator hid this artifact progressively by merging it with the fundus noise pattern. This also resulted in some artificial-looking artifacts, such as the ones present in epoch 300 of Fig. 19. Nonetheless, is clear how epoch 500 presents the most balanced version of the presented images, as it is the one that presents a finer grain of detail in the retinal tissues while also hiding the artifacts product of the scale generation attempts. The noise pattern is correctly simulated without sings of artificial repetitive patterns and the retinal histological structures perfectly match the real counterparts.

However, an interesting drawback of our proposal is that, while the usage of the retinal shape to generate the images allows for more stability, it also causes the GAN results to be more dependent on them. Thus, if the binary mask is not representative of a valid retina, neither the generated result. An example of this event is presented in Fig. 20, where the borders of some of the retinas were cropped. This issue is caused by the warping strategy, as it is the one that rotates the image in such ways that, as shown in said figure in the presented masks, leaves black background at the sides of the retinas. Such structure is not present in the real-life retinas, so these generated samples (despite being perfectly realistic in all other aspects) would not pass the analysis of a human expert. Nonetheless, this issue is related to the mask warping algorithm, and not to the generator itself. Moreover, it shows the robustness of our proposal, as the generator correctly completed the retinal structure and vitreous humor noise patterns even when presented with a faulty retinal mask. It shows a high level of reliance on the shape of the binary mask, but this is precisely what gives it the ability to generate both normal and DME labeled images as mentioned in previous sections of this work.

We can see that this is not an impediment for the correct performing of our proposal, as both strategies presented in this work to test the validity of our performance attained more than satisfactory results. Moreover, we were able to find the factor that was causing the lesser sensitivity of the model when tested against an expert model: there are a few configurations that can occur with the warping strategy that deform the retina in such ways that its shape losses al the clinical features that characterized its internal retinal fluids. That is, the warping strategy can, with some lesser retinal fluid accumulations, correct the retinal shape so it now represents an almost normal retina. In Fig. 21, we can see the example of a warped pathological retina that was transformed into an almost nearly healthy retina, creating what we could effectively call "antagonistic" samples. The system correctly generated an small trace of pathological tissues in some of the most external layers of the retina due to the sensible thickness of the retinal layers, but the new shape mostly represented a normal retina.

This way, we can see that most of the weaknesses of our proposal rely on the simple warping strategy used, and that a slightly perfected algorithm can greatly improve the already satisfactory results. Changing this warping algorithm for, for example, deformable models (or by slightly extending the retinal borders to the border of the image) will solve this issue.

An example of the success of our proposal can be seen in Fig. 22. In this image, a retina that presents a pathological formation is shown. More precisely, the ILM layer has detached and is floating in the vitreous humor. This is a real clinical scenario that was present in some of the training samples of the generator. This scenario is characterized

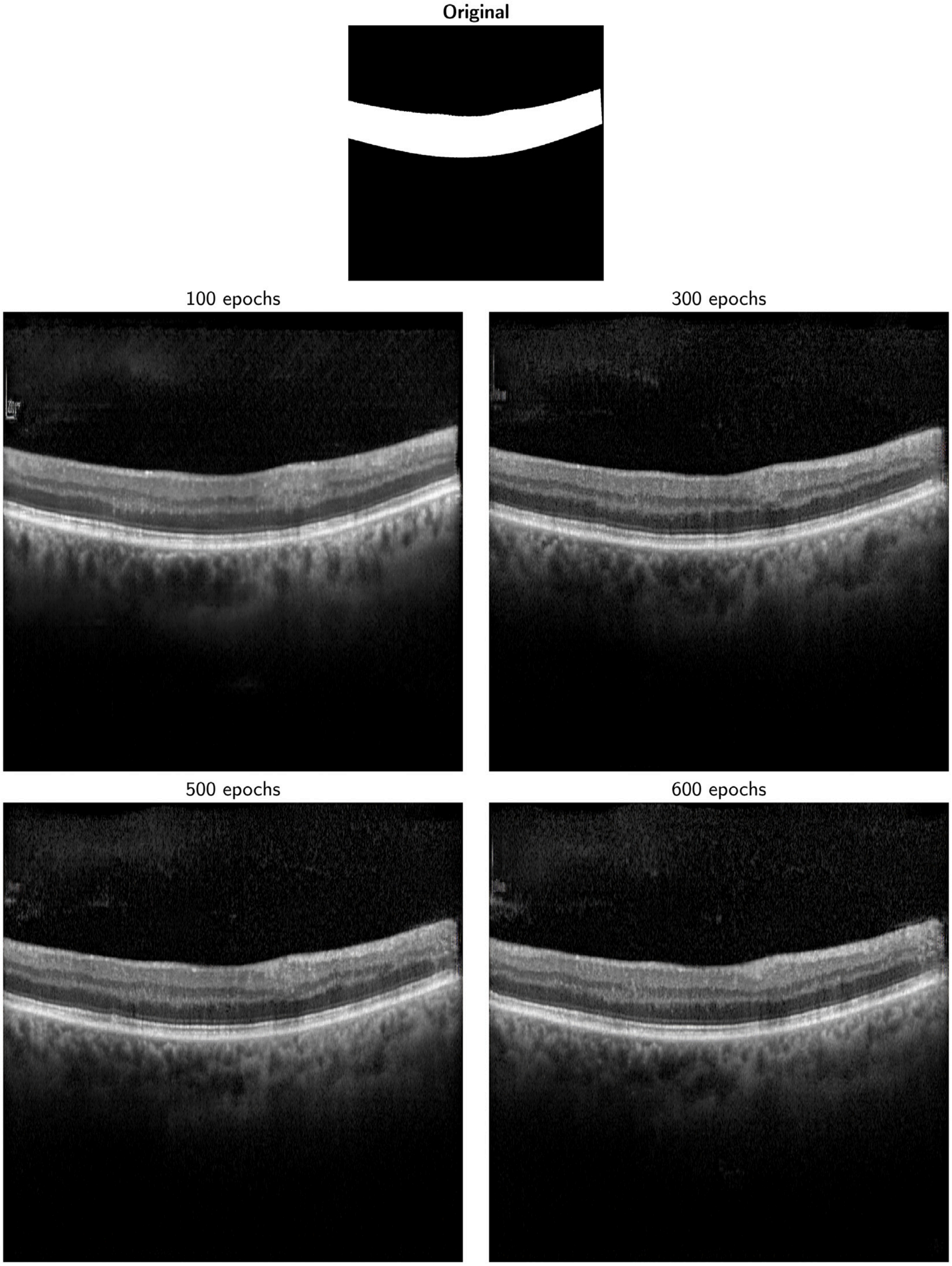

Fig. 18. Example of generated images with the seed mask along different epochs for a normal retina.

by its tendency to generate a traction in the innermost retinal tissues, warping the shape of the retina. What we are seeing in this image is how the generator, when presented with a case where the retinal shape is warped in a way that could be an indicator of this pathological scenario, it successfully recreates a realistic representation of it. (See Table 3.)

5. Conclusions

In this work, to the best of our knowledge, we propose a novel fully-automatic methodology for the generation of realistic labeled OCT images with and without DME fluid accumulations as a solution to the data scarcity problem present in this medical imaging domain. Our proposal represents the first methodology that solves three common problems of the state-of-the-art: the inconsistency in the generations (and lack of control over the result), the impossibility to know the predefined classes of the generated samples (needed to train most computer-aided diagnosis systems) in the classification of DME retinal OCT images, and the inherent instability of the training of GAN-based methodologies when managing images from a considerable resolution.

We base our work in discoveries from the biomedical domain, as the shape of the retinal layers was demonstrated in clinical studies that directly correlates with the different types and distribution of fluid accumulations in the retinal layers. Our methodology uses binary masks

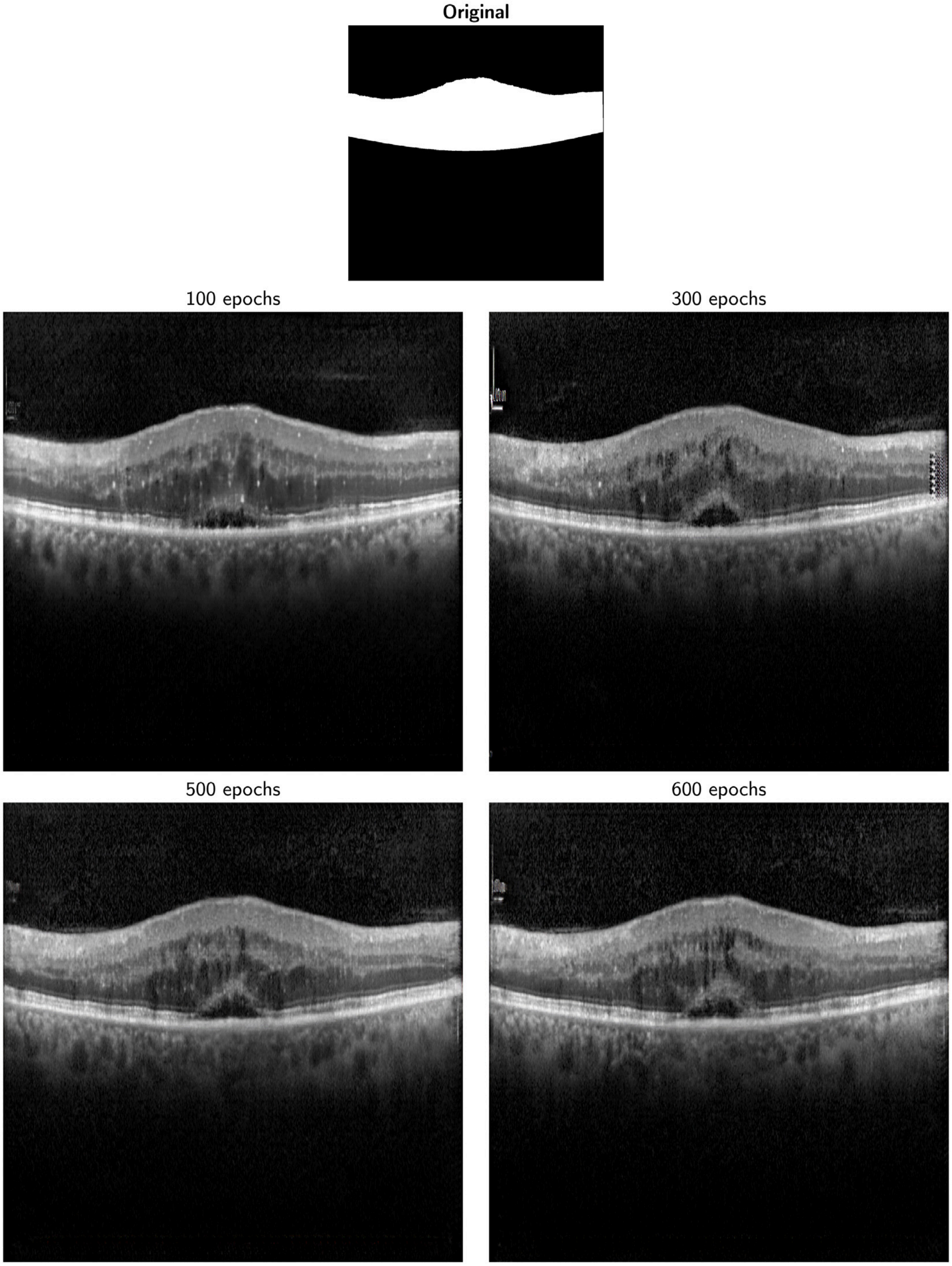

Fig. 19. Example of generated images with the seed mask along different epochs for a retina with DME.

of the retinal morphology as seed to generate a final reconstruction of the retinal layers. This mask gives the generator information about the limiting membranes of the retina, as well as about the distribution and shape of the internal vitreous humor and the external choroidal tissues. Additionally, implicitly, it provides information about what DME fluid accumulations should be present in the synthetic generated image. Thus, we can decide if the generated image will be from a normal retina or from a DME-afflicted patient.

As we are using an image-to-image architecture, the GAN will always try to adjust the generated pattern to the shape of the "proposed" retina; ensuring so that every generated image will be coherent and realistic (as long as the input binary mask is). Thanks to this mask, we are able to generate higher resolution images that even what we could have achieved by using progressively grown GANs: the binary mask is aiding to prevent gradient dispersion in the first learning stages (what would slow down significantly – or even invalidate – the training of the generator) by giving the learning process a solid base to build the retinal structure from. Moreover, thanks to using a binary retinal mask generator strategy, we can potentially create as much new OCT images as desired.

We also tested our proposal with two methodologies from the state-of-the-art. We demonstrated that the proposed methodology is able to produce a satisfactory trained screening model that can successfully

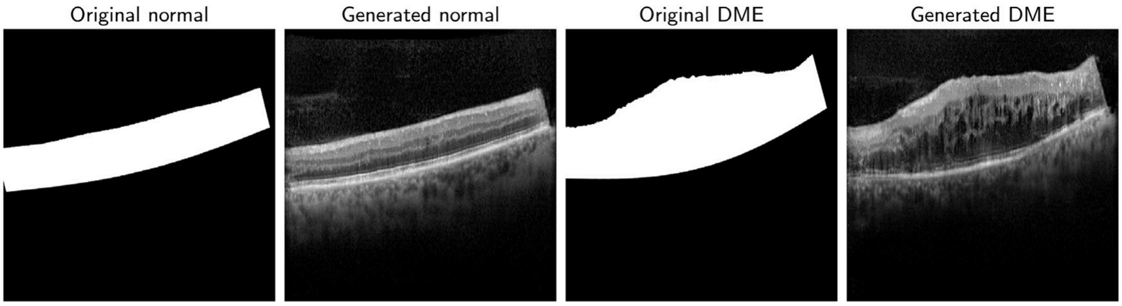

Fig. 20. Examples of images with incomplete masks that result in cropped retinas.

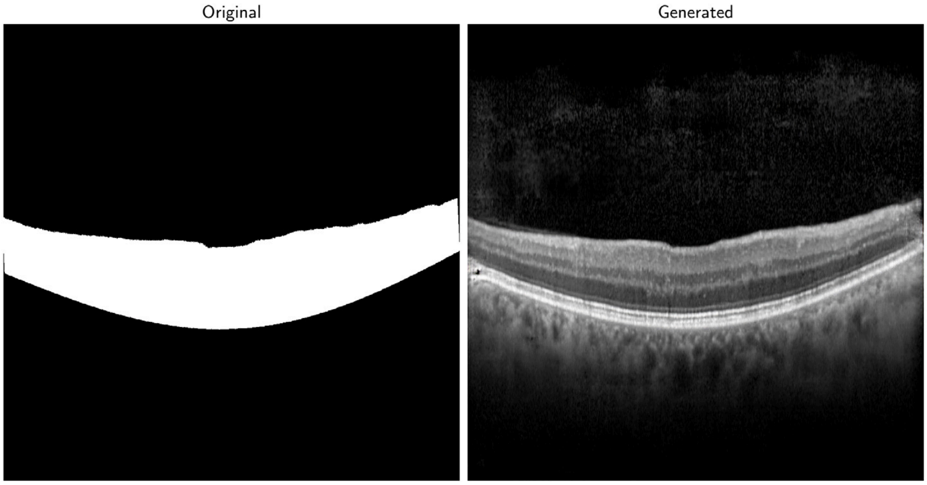

Fig. 21. Example of DME-labeled images that were generated as normal retinal OCT images.

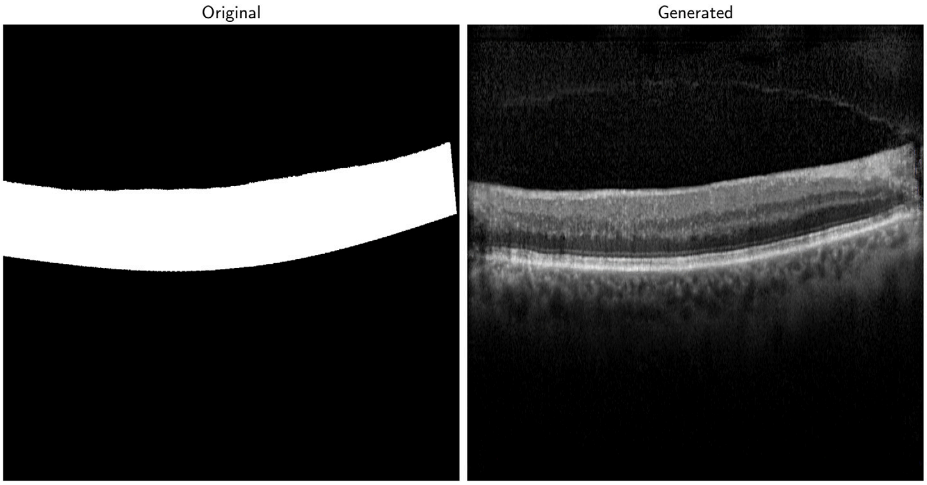

Fig. 22. Curious example on how the generator inferred a detachment of the ILM due to the shape of the generated mask.

separate normal OCT images from DME trained with a completely synthetic dataset. Moreover, we tested our proposal with an expert model, achieving nearly-identical results than a real test dataset. We were able to witness how the methodology even generated extraretinal structures (such as tractions caused by detachment of the ILM layer) without explicit reference to it in the images, only inferring the pattern from what could have caused a deformation of the retina in the provided binary seed. This binary information used as input also helped to reduce the instability of the model during training. The mask provides with a defined constraint for the models to follow, significantly limiting the degrees of freedom the model can reach. This feature is seen in elements of the image that are not constraint to the binary mask (such as the scale indicator of the device) that is placed in different locations.

As future work, we plan to further study the retinal shape features, so we can create a better strategy to generate both pathological and non-pathological retinas without the issues presented before. Additionally, using a loss strategy that combines the L1 norm we used in our work with another structural (and more coarse grain) loss function could also solve (and significantly speed up) the balance between overfitting and smoothing that we seen along the training of the generator. Also, a further study is needed on the reason behind the random positioning of the scale in the generated OCT images. It seems that, without a reference point in the images like it has with the rest of the retina, the network is positioning it in seemingly random positions. This way, as future work, this and other patterns generated by the GAN could be further analyzed by means of explainable artificial intelligence strategies. Finally, thanks to the potential of our proposal to be applied to other imaging modalities and pathologies, it would be interesting to study its implementation in them. Moreover, in the same way that we take advantage of the shape characteristics of the retina to choose at will the type of retina generated, similar applications could be studied in these other modalities (or even discover new biomarkers and shape relationships of relevant structures in the human body by reverse engineering the generator).

Funding

This research was funded by Instituto de Salud Carlos III, Government of Spain, DTS18/00136 research project; Ministerio de Ciencia e Innovación y Universidades, Government of Spain, RTI2018-095894-B-I00 research project, Ayudas para la formación de profesorado universitario (FPU), grant ref. FPU18/02271; Ministerio de Ciencia e Innovación, Government of Spain through the research project with reference PID2019-108435RB-I00; Consellería de Cultura, Educación e Universidade, Xunta de Galicia, Spain, Grupos de Referencia Competitiva, grant ref. ED431C 2020/24 and through the postdoctoral grant contract ref. ED481B 2021/059; Axencia Galega de Innovación (GAIN), Xunta de Galicia, grant ref. IN845D 2020/38; CITIC, Centro de Investigación de Galicia ref. ED431G 2019/01, receives financial support from Consellería de Educación, Universidade e Formación Profesional, Xunta de Galicia, through the ERDF (80%) and Secretaría Xeral de Universidades (20%).

CRediT authorship contribution statement

Plácido L. Vidal: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Joaquim de Moura:** Conceptualization, Validation, Investigation, Data curation, Writing – review & editing, Supervision, Project administration. **Jorge Novo:** Validation, Investigation, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Manuel G. Penedo:** Supervision, Project administration, Funding acquisition. **Marcos Ortega:** Validation, Investigation, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

[1] J. Kotia, A. Kotwal, R. Bharti, R. Mangrulkar, Few shot learning for medical imaging, in: Studies in Computational Intelligence, Springer International Publishing, 2020, pp. 107–132, http://dx.doi.org/10.1007/978-3-030-50641-4_7.

[2] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Lecture Notes in Computer Science, Springer International Publishing, 2015, pp. 234–241, http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[3] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) http://dx.doi.org/10.1186/s40537-019-0197-0.

[4] Z. Hussain, F. Gimenez, D. Yi, D. Rubin, Differential data augmentation techniques for medical imaging classification tasks, AMIA Annu. Symp. Proc. 2017 (2018) 979–984, URL: https://pubmed.ncbi.nlm.nih.gov/29854165.

[5] A. Zhao, G. Balakrishnan, F. Durand, J.V. Guttag, A.V. Dalca, Data augmentation using learned transformations for one-shot medical image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019.

[6] M. Al-Sheikh, K.G. Falavarjani, H. Akil, S.R. Sadda, Impact of image quality on OCT angiography based quantitative measurements, Int. J. Retin. Vitr. 3 (1) (2017) http://dx.doi.org/10.1186/s40942-017-0068-9.

[7] G.M. Somfai, H.M. Salinas, C.A. Puliafito, D.C. Fernández, Evaluation of potential image acquisition pitfalls during optical coherence tomography and their influence on retinal image segmentation, J. Biomed. Opt. 12 (4) (2007) 041209, http://dx.doi.org/10.1117/1.2774827.

[8] M. Balasubramanian, C. Bowd, G. Vizzeri, R.N. Weinreb, L.M. Zangwill, Effect of image quality on tissue thickness measurements obtained with spectral domain-optical coherence tomography, Opt. Express 17 (5) (2009) 4019, http://dx.doi.org/10.1364/oe.17.004019.

[9] M. Kim, J. Zuallaert, W. De Neve, Few-shot learning using a small-sized dataset of high-resolution fundus images for glaucoma diagnosis, in: Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care, MMHealth '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 89—92, http://dx.doi.org/10.1145/3132635.3132650.

[10] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, D. Wierstra, Matching networks for one shot learning, 2017, arXiv:1606.04080.

[11] A. Medela, A. Picon, C.L. Saratxaga, O. Belar, V. Cabezón, R. Cicchi, R. Bilbao, B. Glover, Few shot learning in histopathological images:Reducing the need of labeled data on biological datasets, in: 2019 IEEE 16th International Symposium on Biomedical Imaging, ISBI 2019, 2019, pp. 1860–1864, http://dx.doi.org/10.1109/ISBI.2019.8759182.

[12] G. Koch, R. Zemel, R. Salakhutdinov, Siamese neural networks for one-shot image recognition, in: ICML Deep Learning Workshop, vol. 2, Lille, 2015, pp. 1–8.

[13] P.M. Burlina, N. Joshi, K.D. Pacheco, T.Y.A. Liu, N.M. Bressler, Assessment of deep generative models for high-resolution synthetic retinal image generation of age-related macular degeneration, JAMA Ophthalmol. 137 (3) (2019) 258–264, http://dx.doi.org/10.1001/jamaophthalmol.2018.6156.

[14] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of GANs for improved quality, stability, and variation, 2018, arXiv:1710.10196.

[15] H. Zhao, H. Li, S. Maurer-Stroh, L. Cheng, Synthesizing retinal and neuronal images with generative adversarial nets, Med. Image Anal. 49 (2018) 14–26, http://dx.doi.org/10.1016/j.media.2018.07.001, URL: https://www.sciencedirect.com/science/article/pii/S1361841518304596.

[16] L. Hu, H. Liang, L. Lu, Splicing learning: A novel few-shot learning approach, Inform. Sci. 552 (2021) 17–28, http://dx.doi.org/10.1016/j.ins.2020.11.028.

[17] P.L. Vidal, J. de Moura, J. Novo, M.G. Penedo, M. Ortega, Intraretinal fluid identification via enhanced maps using optical coherence tomography images, Biomed. Opt. Express 9 (10) (2018) 4730, http://dx.doi.org/10.1364/boe.9.004730.

[18] P.L. Vidal, J. de Moura, J. Novo, M. Ortega, Cystoid fluid color map generation in optical coherence tomography images using a densely connected convolutional neural network, in: 2019 International Joint Conference on Neural Networks, IJCNN, IEEE, 2019, pp. 1–8, http://dx.doi.org/10.1109/ijcnn.2019.8852208.

[19] J. Kugelman, D. Alonso-Caneiro, S.A. Read, S.J. Vincent, F.K. Chen, M.J. Collins, Data augmentation for patch-based OCT chorio-retinal segmentation using generative adversarial networks, Neural Comput. Appl. (2021) http://dx.doi.org/10.1007/s00521-021-05826-w.

[20] 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, http://dx.doi.org/10.3390/app10217718.

[21] R.E. Radi, E.T. Damanhouri, M.I. Siddiqui, Diabetic retinopathy causes, symptoms, and complications: A review, Int. J. Med. Dev. Countries 5 (1) (2021) 390–394, http://dx.doi.org/10.24911/IJMDC.51-1607272221, arXiv:http://www.ijmdc.com/?mno=22881.

[22] P.L. Vidal, J. de Moura, J. Novo, M.G. Penedo, M. Ortega, Intraretinal fluid map generation in optical coherence tomography images, in: Diabetes and Retinopathy, Elsevier, 2020, pp. 19–43, http://dx.doi.org/10.1016/b978-0-12-817438-8.00002-x.

[23] J. de Moura, P.L. Vidal, J. Novo, J. Rouco, M.G. Penedo, M. Ortega, Intraretinal fluid pattern characterization in optical coherence tomography images, Sensors 20 (7) (2020) 2004, http://dx.doi.org/10.3390/s20072004.

[24] D. Huang, E. Swanson, C. Lin, J. Schuman, W. Stinson, W. Chang, M. Hee, T. Flotte, K. Gregory, C. Puliafito, et al., Optical coherence tomography, Science 254 (5035) (1991) 1178–1181, http://dx.doi.org/10.1126/science.1957169.

[25] S.G. Odaibo, Generative adversarial networks synthesize realistic OCT images of the retina, 2019, CoRR, arXiv:1902.06676.

[26] X. Zha, F. Shi, Y. Ma, W. Zhu, X. Chen, Generation of retinal OCT images with diseases based on cGAN, in: E.D. Angelini, B.A. Landman (Eds.), Medical Imaging 2019: Image Processing, vol. 10949, International Society for Optics and Photonics, SPIE, 2019, pp. 544–549, http://dx.doi.org/10.1117/12.2510967.

[27] C. Zheng, X. Xie, K. Zhou, B. Chen, J. Chen, H. Ye, W. Li, T. Qiao, S. Gao, J. Yang, J. Liu, Assessment of generative adversarial networks model for synthetic optical coherence tomography images of retinal disorders, Transl. Vis. Sci. Technol. 9 (2) (2020) 29, http://dx.doi.org/10.1167/tvst.9.2.29.

[28] X. He, L. Fang, H. Rabbani, X. Chen, Z. Liu, Retinal optical coherence tomography image classification with label smoothing generative adversarial network, Neurocomputing 405 (2020) 37–47, http://dx.doi.org/10.1016/j.neucom.2020.04.044.

[29] X. Mao, Q. Li, H. Xie, R.Y.K. Lau, Z. Wang, S.P. Smolley, Least squares generative adversarial networks, 2017, arXiv:1611.04076.

[30] T.L. Paine, C. Paduraru, A. Michi, C. Gulcehre, K. Zolna, A. Novikov, Z. Wang, N. de Freitas, Hyperparameter selection for offline reinforcement learning, 2020, http://dx.doi.org/10.48550/ARXIV.2007.09055, URL: arXiv:2007.09055.

[31] M. Kaselimi, N. Doulamis, A. Doulamis, A. Voulodimos, E. Protopapadakis, Bayesian-optimized bidirectional LSTM regression model for non-intrusive load monitoring, in: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2019, http://dx.doi.org/10.1109/icassp.2019.8683110.

[32] F. Alarsan, M. Younes, Best selection of generative adversarial networks hyperparameters using genetic algorithm, 2020, http://dx.doi.org/10.21203/rs.3.rs-95571/v1.

[33] W. Li, W.W.Y. Ng, T. Wang, M. Pelillo, S. Kwong, HELP: An LSTM-based approach to hyperparameter exploration in neural network learning, Neurocomputing 442 (2021) 161–172, http://dx.doi.org/10.1016/j.neucom.2020.12.133.

[34] S. Santurkar, D. Tsipras, A. Ilyas, A. Madry, How does batch normalization help optimization? 2018, http://dx.doi.org/10.48550/ARXIV.1805.11604, URL: arXiv:1805.11604.

[35] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE Computer Society, 2017, pp. 5967–5976.

[36] B.Y. Kim, S.D. Smith, P.K. Kaiser, Optical coherence tomographic patterns of diabetic macular edema, Am. J. Ophthalmol. 142 (3) (2006) 405–412.e1, http://dx.doi.org/10.1016/j.ajo.2006.04.023.

[37] C.-S. Yang, C.-Y. Cheng, F.-L. Lee, W.-M. Hsu, J.-H. Liu, Quantitative assessment of retinal thickness in diabetic patients with and without clinically significant macular edema using optical coherence tomography, Acta Ophthalmol. Scand. 79 (3) (2001) 266–270, http://dx.doi.org/10.1034/j.1600-0420.2001.790311.x.

[38] T. Murakami, N. Yoshimura, Structural changes in individual retinal layers in diabetic macular edema, J. Diabetes Res. 2013 (2013) 1–11, http://dx.doi.org/10.1155/2013/920713.

[39] G. Huang, Z. Liu, K.Q. Weinberger, Densely connected convolutional networks, 2016, CoRR, arXiv:1608.06993.

[40] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252, http://dx.doi.org/10.1007/s11263-015-0816-y.

[41] L. Fei-Fei, J. Deng, O. Russakovsky, A. Berg, K. Li, Imagenet image database, 2021, URL: http://image-net.org/. (Accessed 10 March 2021).

[42] T. Nazir, M. Nawaz, J. Rashid, R. Mahum, M. Masood, A. Mehmood, F. Ali, J. Kim, H.-Y. Kwon, A. Hussain, Detection of diabetic eye disease from retinal images using a deep learning based CenterNet model, Sensors 21 (16) (2021) http://dx.doi.org/10.3390/s21165283, URL: https://www.mdpi.com/1424-8220/21/16/5283.

[43] S. Kaymak, A. Serener, Automated age-related macular degeneration and diabetic macular edema detection on OCT images using deep learning, in: 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing, ICCP, IEEE, 2018, http://dx.doi.org/10.1109/iccp.2018.8516635.

[44] K.T. Islam, S. Wijewickrema, S. O'Leary, Identifying diabetic retinopathy from OCT images using deep transfer learning with artificial neural networks, in: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems, CBMS, 2019, pp. 281–286, http://dx.doi.org/10.1109/CBMS.2019.00066.

[45] F. Fuentes-Hurtado, S. Morales, J.M. Mossi, V. Naranjo, V. Fedulov, D. Woldbye, K. Klemp, M. Torm, M. Larsen, Deep-learning-based classification of rat OCT images after intravitreal injection of ET-1 for glaucoma understanding, in: H. Yin, D. Camacho, P. Novais, A.J. Tallón-Ballesteros (Eds.), Intelligent Data Engineering and Automated Learning, IDEAL 2018, Springer International Publishing, Cham, 2018, pp. 27–34.

[46] X. Li, L. Shen, M. Shen, C.S. Qiu, Integrating handcrafted and deep features for optical coherence tomography based retinal disease classification, IEEE Access 7 (2019) 33771–33777, http://dx.doi.org/10.1109/ACCESS.2019.2891975.

[47] J. de Moura, J. Novo, M. Ortega, Deep feature analysis in a transfer learning-based approach for the automatic identification of diabetic macular edema, in: 2019 International Joint Conference on Neural Networks, IJCNN, IEEE, 2019, pp. 1–8, http://dx.doi.org/10.1109/ijcnn.2019.8852196.

[48] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015, URL: http://arxiv.org/abs/1412.6980.

[49] M. Robnik-Šikonja, I. Kononenko, Mach. Learn. 53 (1/2) (2003) 23–69, http://dx.doi.org/10.1023/a:1025667309714.

[50] V. Vapnik, S. Kotz, Estimation of Dependences Based on Empirical Data, vol. 41, 2006, p. 324, http://dx.doi.org/10.2307/2988246,

[51] J. Platt, Sequential minimal optimization: A fast algorithm for training support vector machines, Adv. Kernel Methods-Support Vector Learn. 208 (1998).