Article

A Two-Stage GAN for High-Resolution Retinal Image Generation and Segmentation

Paolo Andreini [1], Giorgio Ciano [1,2,*], Simone Bonechi [1,3], Caterina Graziani [1], Veronica Lachi [1], Alessandro Mecocci [1], Andrea Sodi [4], Franco Scarselli [1] and Monica Bianchini [1]

1 Department of Information Engineering and Mathematics, University of Siena, 53100 Siena, Italy; paolo.andreini@unisi.it (P.A.); simone.bonechi@di.unipi.it (S.B.); caterina.graziani@student.unisi.it (C.G.); veronica.lachi@student.unisi.it (V.L.); mecocci@unisi.it (A.M.); franco@diism.unisi.it (F.S.); monica.bianchini@unisi.it (M.B.)
2 Department of Information Engineering, University of Florence, 50121 Florence, Italy
3 Department of Computer Science, University of Pisa, 56127 Pisa, Italy
4 S.O.D. Oculistica, Azienda Ospedaliero Universitaria Careggi, 50134 Firenze, Italy; andreasodi2@gmail.com
* Correspondence: giorgio.ciano@unifi.it

Abstract: In this paper, we use Generative Adversarial Networks (GANs) to synthesize high-quality retinal images along with the corresponding semantic label-maps, instead of real images during training of a segmentation network. Different from other previous proposals, we employ a two-step approach: first, a progressively growing GAN is trained to generate the semantic label-maps, which describes the blood vessel structure (i.e., the vasculature); second, an image-to-image translation approach is used to obtain realistic retinal images from the generated vasculature. The adoption of a two-stage process simplifies the generation task, so that the network training requires fewer images with consequent lower memory usage. Moreover, learning is effective, and with only a handful of training samples, our approach generates realistic high-resolution images, which can be successfully used to enlarge small available datasets. Comparable results were obtained by employing only synthetic images in place of real data during training. The practical viability of the proposed approach was demonstrated on two well-established benchmark sets for retinal vessel segmentation—both containing a very small number of training samples—obtaining better performance with respect to state-of-the-art techniques.

Keywords: deep learning; convolutional neural networks; semantic segmentation; generative adversarial networks; retinal images; image augmentation

1. Introduction

The retinal microvasculature is the only part of human circulation that can be directly and non-invasively visualized in vivo [1]. Hence, it can be easily acquired and analyzed by automatic tools. As a result, retinal fundus images have a multitude of applications, including biometric identification, computer-assisted laser surgery, and the diagnosis of several disorders [2,3]. One important processing step in such applications is the proper segmentation of retinal vessels. Image semantic segmentation aims to make dense predictions by inferring the object class for each pixel of an image and, indeed, the segmentation of digital retina images allows us to extract various quantitative vessel parameters and to obtain more objective and accurate medical diagnoses. In particular, the segmentation of retinal blood vessels can help the diagnosis, treatment, and monitoring of diseases such as diabetic retinopathy, hypertension, and arteriosclerosis [4,5].

Deep Neural Networks (DNNs) has become the standard approach in semantic segmentation [6–8] and in many other computer vision tasks [9–12]. DNN training, however, requires large sets of accurately labeled data, so the availability of annotated images is becoming increasingly critical. This is particularly true in medical applications, where data collection is often difficult and expensive. For this reason, generating synthetic data is of

great interest. Nevertheless, synthesizing high-resolution realistic medical images remains an open challenge. Most of the leading approaches for semantic segmentation, in fact, rely on thousands of supervised images, while supervised public datasets for retinal vessel segmentation are very small (most datasets contain fewer than 30 images).

To face the scarcity of data, we propose a new approach for the generation of retinal images along with the corresponding semantic label-maps. Specifically, we propose a novel generation procedure based on two distinct phases. In the first phase, a generative adversarial network (GAN) [13] generates the blood vessel structure (i.e., the vasculature). The GAN is trained to learn the typical semantic label-map distribution from a small set of training samples. To generate high-resolution label-maps, the Progressively Growing GAN (PGGAN) [14] approach has been employed. In a second, distinct phase, an image-to-image translation algorithm [15] is used to translate blood vessels structures into realistic retinal images (see Figure 1).
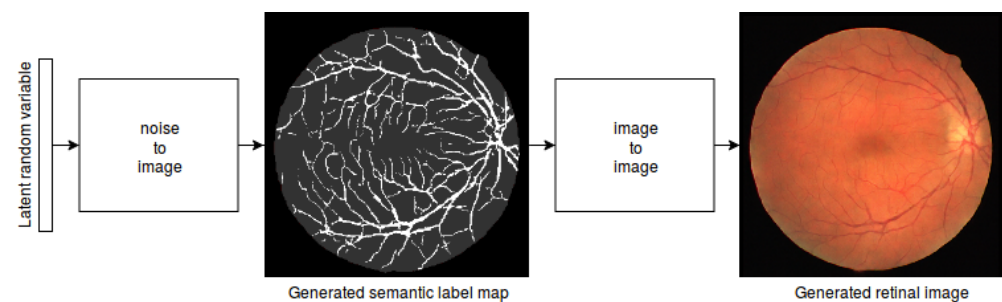
Figure 1. The proposed two-step image generation method.

The rationale behind this approach is that, in many applications, the semantic structure of an image can be learned regardless of its visual appearance. Once the semantic label-map has been generated, visual details can be incorporated using an image-to-image translation algorithm, thus obtaining realistic synthesized images. By separating the whole process into two stages, the generation task is simplified and the number of samples required for training is significantly reduced. Moreover, the training is very effective and we obtained retinal images with an unprecedented high resolution and quality, along with their semantic label-maps. It is worth noting that the proposed two-step approach also reduces the GPU memory requirements with respect to a single-step method. Finally, the generation of the label-maps, based on GANs, allows us to synthesize a virtually infinite number of different training samples with different vasculature.

To assess the usefulness and correctness of the proposed approach, the generation procedure has been applied on two public datasets (i.e., DRIVE [16] and CHASE_DB1 [17]). Moreover, the two-step generation procedure has been compared with a single-stage generation, in which label-maps and retinal images have been generated simultaneously in two different channels (see Figure 2). Indeed, in our experiments, the multi-stage approach allows us to significantly improve performance of vessels segmentation when used for data augmentation. In particular, the generated data have been used to train a Segmentation Multiscale Attention Network (SMANet) [18]. Comparable results have been obtained by training the SMANet on the generated images in place of real data. It is interesting to note that, if the network is pre-trained on the synthesized data and then fine-tuned on real images, the segmentation results obtained on the DRIVE dataset come very close to those obtained by the best state-of-the-art approach [19]. If the same approach is applied to the CHASE_DB1 benchmark, the results overcome (to the best of our knowledge) those obtained by any other previously proposed method.
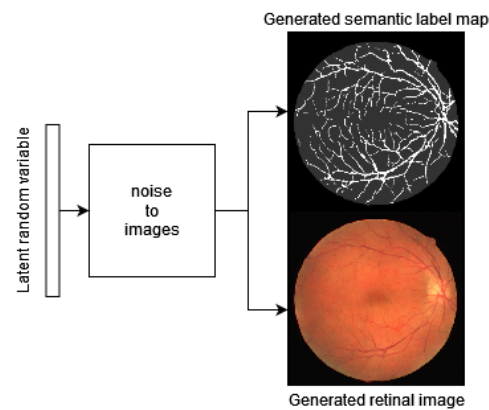
Figure 2. The single-stage image generation scheme.

This paper is organized as follows. In Section 2, the related literature is reviewed. Section 3 presents a description of the proposed approach. Section 4 shows and discusses the experimental results. Finally, Section 5 draws the conclusions and future perspectives.

2. Related Work

2.1. Synthetic Image Generation

Methods for generating images can be classified into two main categories: model-based and learning-based approaches. The most conventional procedure is to formulate a model of the observed data and to render the images using a dedicated engine. This approach has been used, for example, to extend the available datasets of driving scenes in urban environments [20,21] or for object detection [22]. In the field of medical image analysis, synthetic image generation has been extensively employed. For example, realistic digital brain-phantom has been synthesized in [23], while more recently, synthetic agar plate images have been generated for image segmentation [24,25]. The design of specialized engines for data generation requires an accurate model of the scene and deep knowledge of the specific domain. For this reason, in recent years, the learning-based approach has attracted increasing research resources. In this context, machine learning techniques are used to capture the intrinsic spatial variability of a set of training images, so that the specific domain model is acquired implicitly from the data. Once the probability distribution that underlies the set of real images has been learned, the system can be used to generate new images that are likely to mimic the original ones. One of the most successful machine learning models for data generation is the Generative Adversarial Network (GAN) [13]. A GAN is composed of two competing networks: a generator G and a discriminator D. G is trained to map a latent random variable $\mathbf{z} \in \mathbb{R}^Z$ into fake images $\tilde{\mathbf{x}} = G(\mathbf{z})$, whereas D aims to distinguish fake samples from real data. The GAN training is formulated as a min-max game between G and D. GANs strive to generate images that resemble real data. If the synthetic images are close enough to the real ones, they can be used to enlarge existing datasets for training machine learning models. For example, GANs have been used in [26] to augment data for a patch-based approach to OCT chorio-retinal boundary segmentation. In [27], synthetic chest X-ray (CXR) images are generated by developing an Auxiliary Classifier Generative Adversarial Network (ACGAN) model, called CovidGAN. Synthetic images produced by CovidGAN were used to improve the performance of a standard CNN for detecting COVID-19.

In [28], synthetic abnormal MR images containing brain tumors are generated. An image-to-image translation algorithm is employed to construct semantic label-maps of real MR brain images, distortions are introduced on the generated segmentation (i.e., tumors are shrunk or enlarged, or their position is changed), and then the segmentation is translated back to images. Indeed, manually introducing distortions on the generated label-maps is not trivial because they can alter the image semantic—for instance, in the case of retinal

image generation, enlarging or reducing blood vessels is not meaningful. We solve this issue directly by learning the semantic label-map distribution with a GAN.

The fundamental idea behind this work is the decomposition of the hard problem of image generation in multiple stages to simplify the generation process. In particular, we aim to generate the semantic structure of the scene in the initial phase, while all of the visual details are incorporated in the second phase. Our experiments demonstrate that this approach is preferable to a single-step generation procedure, resulting in higher-quality images that can be used for data augmentation.

2.2. Image-to-Image Translation

Recently, beside image generation, adversarial learning has also been extended to image-to-image translation, in which the goal is to translate an input image from one domain to another. Many computer vision tasks, such as image super-resolution [29], image inpainting [30], and style transfer [31] can be casted into the image-to-image translation framework. Both unsupervised [32–35] and supervised approaches [14,36,37] can be used. Supervised training uses a set of pairs of corresponding images $\{(s_i, t_i)\}$, where s_i is an image of the source domain and t_i is the corresponding image in the target domain. As an example, Pix2Pix [36] consists of a conditional GAN that operates in a supervised way, and Pix2PixHD [14] employs a coarse-to-fine generator and discriminator, along with a feature-matching loss-function, to translate images with higher resolution and quality.

2.3. Retinal Image Synthesis

One of the first applications of retinal image synthesis has been described in seminal work [38], in which an anatomic model of the eye and of the surrounding face has been implemented for surgical simulations. More recently, in [39], a large dictionary of small image patches containing no vessels, has been used to model the retinal background and fovea. A parametric intensity model, in which the parameters have been estimated from real images, is used to generate the optical disk. Complementary to [39], the contribution in [40] focuses on the generation of the vascular network, based on a parametric model, in which the parameters are learned from real vessel trees. Despite these methods providing reasonable results, they are complex and heavily depend on the domain knowledge. To reduce the knowledge requirements, a completely learning-based approach has been proposed in [41], where an image-to-image translation model has been employed to transform existing vessel networks into realistic retinal images. Vessel networks used for learning have been obtained using a suitable segmentation technique applied to a set of real retina images. However, the quality of the generated images heavily depends on the segmentation module performance. In [42], a generative adversarial approach, together with a style transfer algorithm, is used to reduce the need for annotated samples and to improve the representativeness (e.g., the variability) of synthesized images. The model still relies on pre-existing vessel networks (obtained manually or by a suitable segmentation technique). In [43], an adversarial auto-encoder for retinal vessel synthesis has been adopted to avoid the dependence of the model on the availability of pre-existent vessel maps. Nevertheless, this approach allows us to generate only low-resolution images, and the performance in vessel segmentation using synthesized data is far below that of the state-of-the-art. Higher-resolution retinal images, along with their segmentation label-maps, have been generated in [44], using Progressively Growing GANs (PGGANs) [14]. This method allows us to generate images up to a resolution of 512×512 pixels. A set of 5550 images segmented by a pre-trained U-Net [45] have been used during training. Unfortunately, the usefulness of the generation for image segmentation is not demonstrated.

The present paper improves previous approaches generating synthetic images up to a resolution of 1024×1024 pixels. The generation is based on a very small set of pre-existing images (actually, 20 images with supervised segmentation maps). Both the retinal images and the corresponding semantic label-maps (the vasculature) are generated. Fur-

thermore, we prove that combining real retinal images with synthesized ones for training a segmentation network improves the final segmentation performance.

2.4. Retinal Vessel Segmentation

During recent decades, several approaches for retinal vessel segmentation have been proposed, both supervised and unsupervised. Unsupervised methods depend heavily on prior knowledge on the vessel structure. For example, the so-called vessel tracking techniques define an initial set of seed points and, thereafter, by chaining pixels that minimize a given cost function, iteratively extract the vasculature [46,47]. In [48], retinal images are convolved with a 2D filter to produce a Gaussian intensity profile of the blood vessels, that is subsequently thresholded to give the vessel map. Adaptive thresholding has been used in [49] and in [50]. An active contour model that combines intensity and local phase information is used in [51]. In [52], a hybrid unsupervised approach was proposed. To obtain the vessel location map, the composition of two pre-processed images is fused with the enhanced image of B-COSFIRE filters followed by thresholding. Instead, an ensemble strategy automatically combining multiple segmentation results is presented in [53]. Moreover, since the retinal blood vessels' diameter significantly changes based on the distance from the optic disc, multi-scale approaches can be particularly effective for the vessel segmentation [54,55]. Supervised methods are currently the leading techniques in semantic segmentation. In this framework, true annotations are used to train a classifier aimed at distinguishing the vessels from the background. Various classification models have been employed for blood vessel segmentation based on a preliminary feature engineering stage [56–58], which, however, has a fundamental impact on performance.

Conversely, deep learning methods automatically learn an increasingly complex hierarchy of features from input data, bypassing the need for problem-specific knowledge and generally providing better results. Indeed, a deep convolutional neural network (DCNN) for retinal image segmentation has been used in [59], while the training examples are subjected to various forms of pre-processing and augmented based on geometric transformations and gamma corrections. A neural network that can be efficiently used in real-time on embedded systems is proposed in [60]. In [61], a fully convolutional network [6] was described, with an AlexNet [9] encoder. Fully convolutional networks have also been used in [62,63]. In [64], the segmentation task was remolded into a problem of cross-modality data transformation from retinal images to vessel maps. A modified U-Net [45] was used in [65] to exploit a combination between segment-level loss and pixel-level loss to deal with the unbalanced ratio between thick and thin vessels in fundus images. A Holistically Nested Edge Detection (HED) network [66]—originally designed for edge detection—followed by a conditional random field were employed for the retinal blood vessel segmentation in [67]. Deep supervision was incorporated in some intermediate layers of a VGG network [68] in [69,70]. In [71], a fully convolutional neural network used a stationary wavelet transform pre-processing step to improve the network performance. Finally, in [19], a CNN was pre-trained on image patches and then fine-tuned at the image level.

In this paper, we use the Segmentation Multiscale Attention Network (SMANet), which allows us to obtain excellent results, comparable with the state-of-the-art.

3. Materials and Methods

3.1. The Benchmark Datasets

- DRIVE dataset—The DRIVE dataset [16] includes 40 retinal fundus images of size $584 \times 565 \times 3$ (20 images for training and 20 for test). The images were collected by a screening program for diabetic retinopathy in the Netherlands. Among the 40 photographs, 33 showed no diabetic retinopathy, while 7 showed mild early diabetic retinopathy. The segmentation ground-truth was provided both for the training and the test sets.

- CHASE_DB1 dataset—The CHASE_DB1 dataset [17] is composed by 28 fundus images of size $960 \times 999 \times 3$, corresponding to the left and right eyes of 14 children. Each image is annotated by two independent human experts. An officially defined split between training and test is not provided for this dataset. In our experiments, we adopted the same strategy as [64,65], selecting the first 20 images for training and the remaining 8 for testing.

The main goal of this work is to generate realistic retinal images and the corresponding semantic segmentation masks by using a very small number of training samples. The proposed generation procedure is composed of two steps (see Figure 1): the first one involves the generation of semantic label-maps of the vessels while, during the second, the synthesis of realistic images based on label-maps is carried out. The quality of the generated images was validated by an expert and their usefulness was verified by the performance obtained on two public benchmark datasets, using the synthesized images to train a segmentation network. In particular, Section 3.2 gives an overview of the approach used to generate the semantic label-maps, while Section 3.3 describes the image-to-image translation algorithm that synthesizes retinal images from the semantic label-maps. Section 3.4 reports the semantic segmentation network used to segment retinal vessels. Finally, some details on the training method are collected in Section 3.5.

3.2. Vasculature Generation

The generation of the vessel structure is based on the use of PGGANs, which are capable of learning the distribution of the semantic label-maps. The label-maps are processed to encode both the retinal fundus and the vasculature (i.e., the vessel distribution). To reduce the risk related to the lack of an adequate descriptive power, due to the very limited number of available training samples, data augmentation was applied. Specifically, the semantic label-maps were slightly rotated ($\pm 15°$) and flipped in different ways (horizontal, vertical, and horizontal followed by vertical flips). The generation started at low resolution, and then, the resolution was progressively increased by adding new layers to the networks. The generator and the discriminator were symmetric and grew in sync. The transition from low-resolution image generation to high-resolution image generation followed the procedure described in [14], to avoid problems related to sudden transitions. The training started with both the generator and the discriminator having a spatial resolution of 4×4 pixels, progressively increasing until the final resolving power was reached. The Wasserstein loss, with a gradient penalty [72], was used as the loss function for the discriminator. The learning procedure is illustrated in Figure 3.
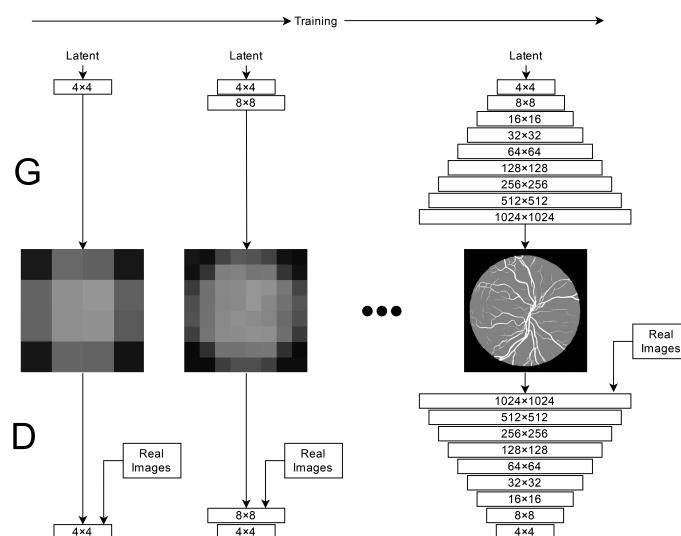

Figure 3. Training scheme for the generation of the semantic label-maps. The resolutions of the generator (G) and the discriminator (D) were progressively increased until the final resolving power was reached.

It can be observed that the global structure of the vessel distribution was learned at the beginning of the training, whereas finer details were added as the resolution increased. The generation procedure allows us to obtain a virtually infinite number of different vasculatures. To reduce the probability of introducing artifacts, a simple post-processing was carried out. Specifically, a morphological opening [73] was applied to the generated retinal fundus mask to improve its circularity. Small holes were filled, and segments of small dimension were removed from the generated vessel structure.

3.3. Translating Vessel Maps into Retinal Images

Once the vessel networks were obtained, they were transformed into realistic color retinal images. Our method is based on Pix2PixHD [14], a supervised image-to-image translation framework derived from Pix2Pix [36]. In Pix2Pix, a conditional GAN learns to generate the output conditioned on the corresponding input image. The generator has an encoder-decoder structure, takes in input images belonging to a certain domain *A*, and generates images in a different domain *B*. The discriminator observes pairs of images, and the image of *A* is provided as input along with the corresponding image of *B* (real or generated). The discriminator aims to distinguish between real and fake (generated) pairs. Pix2PixHD improves upon Pix2Pix by introducing a coarse-to-fine generator composed of two subnetworks that operate at different resolutions. A multiscale discriminator was also employed, with an adversarial loss that incorporates feature-matching loss for training stabilization. In our setup, the semantic label-maps, previously generated, were fed into the generator, which is trained to generate retinal images. An overview of the proposed setup is given in Figure 4.
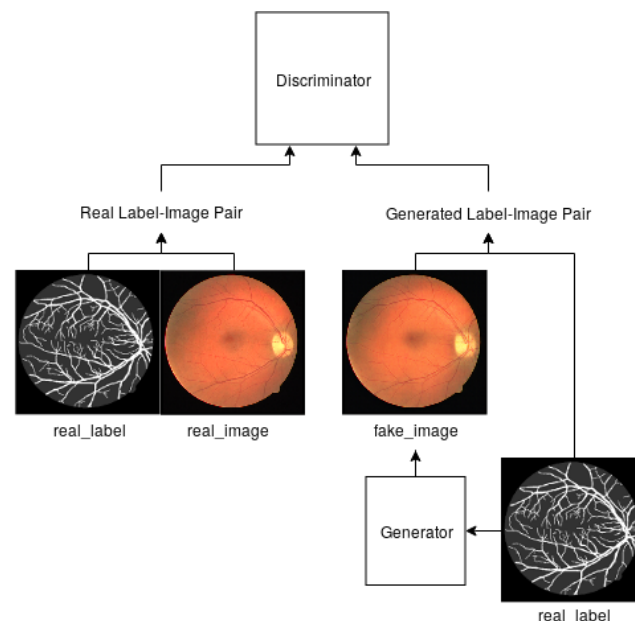

Figure 4. Scheme of the Pix2Pix training framework employed to translate label-maps into retinal images.

3.4. The SMANet Architecture

The semantic segmentation network employed in this paper is a Segmentation Multi-scale Attention Network (SMANet) [18]. The SMANet, originally proposed for scene text segmentation, comprises three main components: a ResNet encoder, a multi-scale attention module, and a convolutional decoder (see Figure 5).

The architecture is based on the PSPNet [8], a deep fully convolutional neural network with a ResNet [74] encoder. In the PSPNet, to enlarge the receptive field of the neural network, a set of standard convolutions of the ResNet backbone has been replaced with dilated convolutions (i.e., atrous convolutions [75]). Moreover, in the PSPNet, a pyramid of pooling layers, with different kernel size, has been employed to gather context information.

The pooled feature maps are then up-sampled at the same resolution as the ResNet output, concatenated, and fed into a convolutional layer to obtain an encoded representation. In the original PSPNet, this representation is followed by a final convolutional layer that reduces the feature maps to the number of classes. The desired per-pixel prediction is obtained directly up-sampling to the original image resolution. In the SMANet, a multiscale attention mechanism is adopted to focus on the relevant objects present in the image, while a two-level convolutional decoder is added to the architecture to better handle the presence of thin objects.
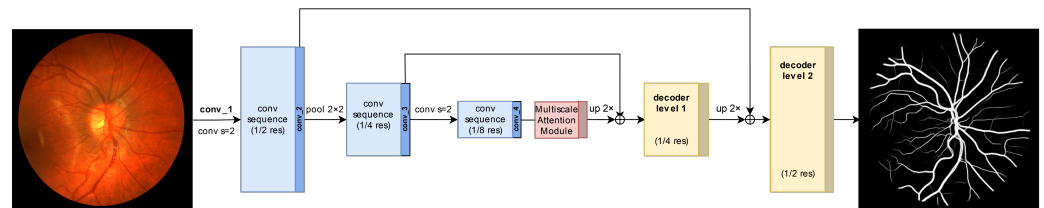
Figure 5. Scheme of the SMANet segmentation network.

3.5. Training Details

The SMANet, used in this work was implemented in TensorFlow. Random crops of 281×281 pixels were employed during training, whereas a sliding window of the same size was used for the evaluation. The Adam optimizer [76], based on a learning rate of 10^{-4} and a mini-batch of 17 examples, was used to train the SMANet. Early stops were employed using a validation set of three images, randomly extracted from the real data training set. Additionally, the PGGAN was realized in TensorFlow, while Pix2PixHD was implemented in PyTorch. The PGGAN architecture is similar to that proposed in [14], but to speed up the computation and to reduce overfitting, the maximum number of feature maps for each layer was fixed to 128. Moreover, since the aim of the generator is to produce a semantic label-map, the output image has only one channel, instead of three. The PGGAN and Pix2PixHD hyperparameters were tuned by visually inspecting the quality of the generated samples. The images were resized to the nearest power-of-two resolution (i.e., the retinal images in the DRIVE dataset, which have a resolution of 565×584 pixels, were resized to 512×512 pixels, whereas the CHASE images that have a resolution of 999×960 pixels were resized to 1024×1024 pixels).

All of the experiments were carried out in a Linux environment on a single NVIDIA Tesla V100 SXM2 with 32 GB RAM.

4. Results and Discussion

In this paper, we provide both qualitative and quantitative evaluations of the generated data. In particular, some qualitative results of the generated retinal images for the DRIVE and CHASE_DB1 dataset are given in Figures 6 and 7.

In Figure 8, a zoom on a random patch of a high-resolution generated image shows that the image-to-image translation allows us to effectively translate the generated vessel structures in retinal images by maintaining the semantic information provided by the semantic label-map. It is worth noting that, although most of the generated samples closely resemble real retinal fundus images, few examples are clearly sub-optimal (see Figure 9, which shows disconnected vessels and an unrealistic optical disc).
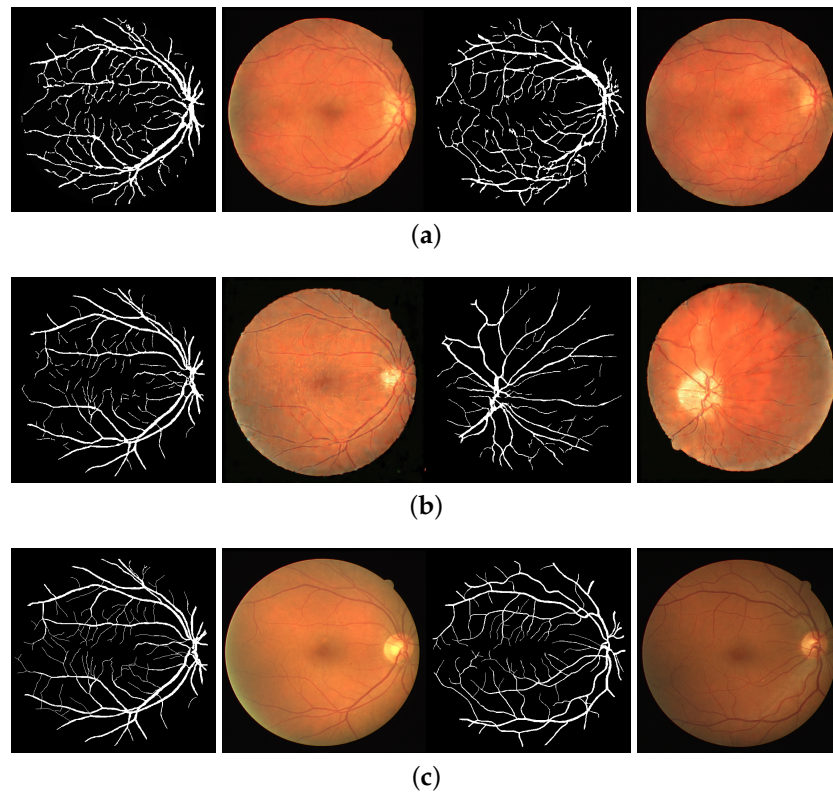
Figure 6. Examples of real and generated DRIVE images. (**a**) Generated DRIVE images with our two-step method; (**b**) Generated DRIVE images with the single-step method; (**c**) Real DRIVE images.
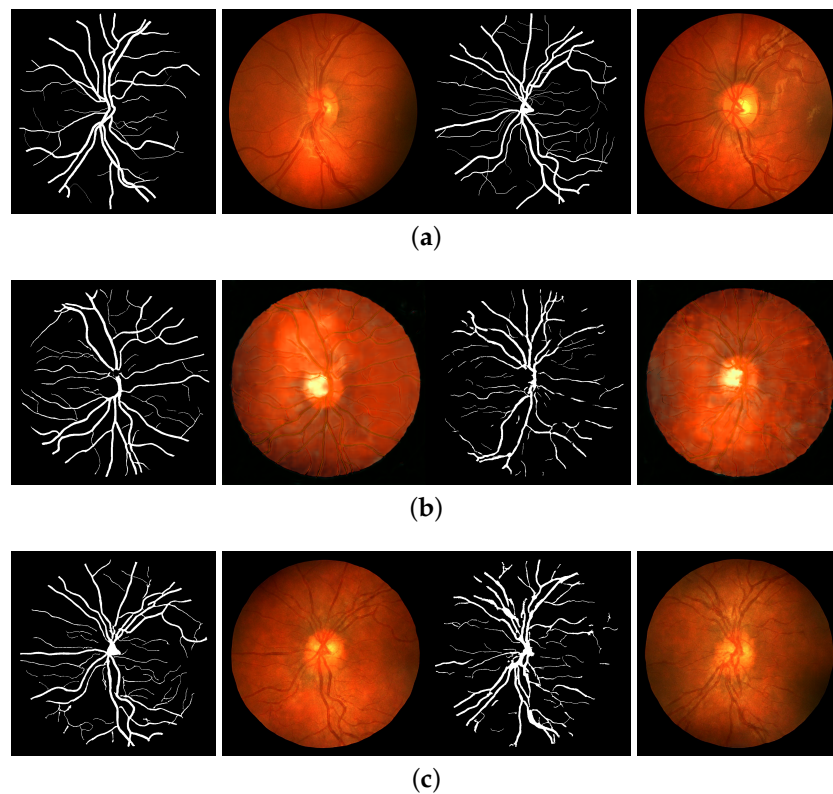

Figure 7. Examples of real and generated CHASE_DB1 images. (**a**) Generated CHASE_DB1 images with our two-step method; (**b**) Generated CHASE_DB1 images with the single-step method; (**c**) Real CHASE_DB1 images.
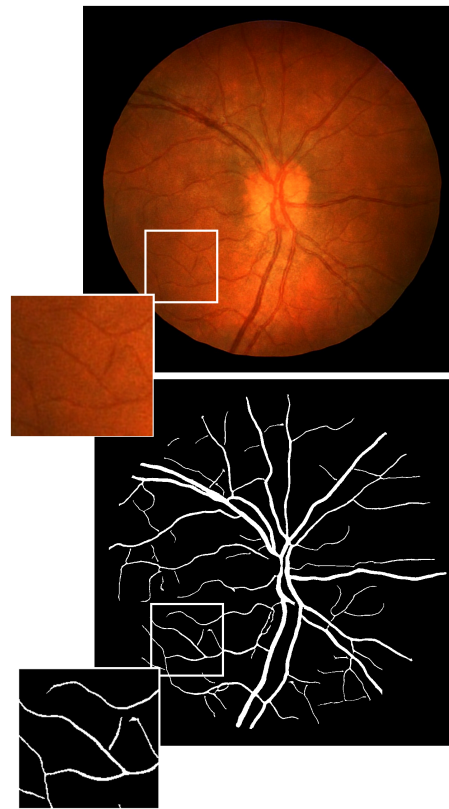
Figure 8. Example of a generated image (resolution 1024 × 1024) with the corresponding label-map from the CHASE_DB1 dataset.

To further validate the quality of the generation process, a sub-sample of 100 synthetically generated retinal images were examined by an expert ophthalmologist. The evaluation showed that 35% of the images are of medium-high quality. The remaining 65% is visually appealing but contains small details that reveal an unnatural anatomy, such as an optical disc with feathered edges—which actually occur only in the case of specific diseases— or blood vessels that pass too close to the macula—while usually, except in the case of malformations, the macular region is avascular or at least paucivascular.
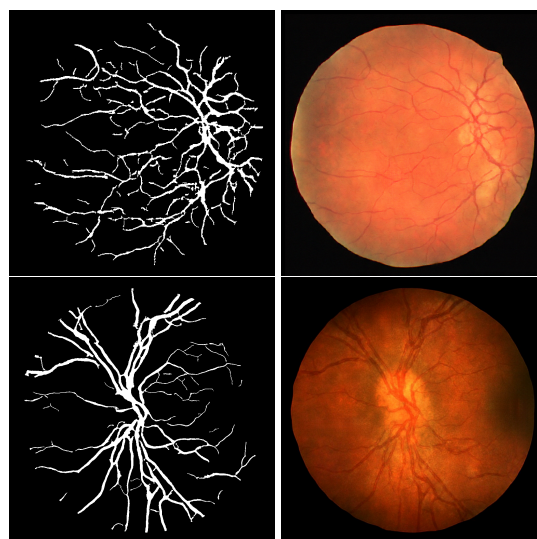
Figure 9. Examples of generated images with an unrealistic optical disc and vasculature from DRIVE (**top**) and CHASE _DB1 (**bottom**).

Table 1 compares the characteristics of the proposed method with respect to other learning-based approaches for retinal image generation found in the literature.

It can be observed that our approach is able to synthesize higher-resolution images, with less training samples, with respect to methods that generate both the image and the corresponding segmentation. Moreover, for such methods, the usefulness of the inclusion of synthetic images in semantic segmentation was not assessed. Instead, in this paper, we demonstrate that synthetic images can be effectively used for data augmentation, which indirectly guarantees the high quality of the generated data.

Table 1. Comparison with other generation approaches.

Methods	Gen. Vessels	Max Res.	Samples
[41]	No	512×512	614
[42]	No	2048×2048	10–20
[43]	Yes	256×256	634
[44]	Yes	512×512	5550
Our	Yes	1024×1024	20

Indeed, the quantitative analysis consists of assessing the usefulness of the generated images for training a semantic segmentation network. This approach, similar to [77], is based on the assumption that the performance of a deep learning architecture can be directly related with the quality and variety of GAN-generated images. The generation procedure described in Section 3 was employed to generate 10,000 synthetic retinal images for both the DRIVE and the CHASE_DB1 datasets; the samples were generated in a single run without any selection strategy. To evaluate the usefulness of the generated data for semantic segmentation, we employed the following experimental setup:

- SYNTH—the segmentation network was trained using only the 10,000 generated synthetic images;
- REAL—only real data were used to train the semantic segmentation network;
- SYNTH + REAL—synthetic data were used to pre-train the semantic segmentation network and real data were employed for fine-tuning.

Tables 2 and 3 report the results of the vessel segmentation for the DRIVE and CHASE_DB1 datasets, respectively.

Table 2. Segmentation performance using the generated and real images from the DRIVE dataset.

Methods	AUC	Acc
SYNTH	98.5%	96.88%
REAL	98.48%	96.87%
SYNTH + REAL	98.65%	96.9%

Table 3. Segmentation performance using the generated and real images from the CHASE_DB1 dataset.

Methods	AUC	Acc
SYNTH	98.64%	97.49%
REAL	98.82%	97.5%
SYNTH + REAL	99.16%	97.72%

It can be observed that the semantic segmentation network, trained on synthetic data, produces results very similar to those obtained by training on real data. This demonstrates that synthetic images effectively capture the training image distribution, so that they can be used to adequately train a deep neural network. Moreover, if fine-tuning with real data is applied after pre-training with synthetic data only, the results further improve with respect

to the use of real data only. This fact indicates that the generated data can be effectively used to enlarge small training sets, such as DRIVE and CHASE_DB1. Specifically, the AUC is improved by 0.17% and 0.34% on the DRIVE and CHASE_DB1 datasets, respectively.

Another set of experiments was designed to compare the proposed two-stage generation procedure with a traditional single-step approach. In particular, in the one-stage method, the label-maps and the retinal images were generated simultaneously. The results of the single-step approach on the DRIVE and CHASE_DB datasets are shown in Tables 4 and 5.

Table 4. Segmentation performance, using the single-step method, on the DRIVE dataset.

Methods	AUC	Acc
SYNTH	93.49%	91.01%
REAL	98.48%	96.87%
SYNTH + REAL	98.57%	96.88%

Table 5. Segmentation performance, using the single-step method, on the CHASE_DB1 dataset.

Methods	AUC	Acc
SYNTH	66.96%	92.62%
REAL	98.82%	97.5%
SYNTH + REAL	98.87%	97.65%

Tables 6 and 7 allows us to quickly visualize the differences between the two methods. In can be observed that better results are obtained in all the setups by employing the two-stage generation approach. In particular, if only synthetic data are used, the AUC increases by 5.01% (31.68%) with the two-stage method in the DRIVE (CHASE_DB1) dataset. As expected, the difference between the two methods is smaller if fine-tuning on real data is applied. Finally, we observe that the gap increases with higher image resolution. In the CHASE_DB1 dataset, in which the images have twice the resolution of the DRIVE dataset, the one-step generated images cannot be effectively used as data augmentation.

Table 6. A comparison of the vessel segmentation results on the DRIVE dataset between the one-step and the two-step methods.

Methods	AUC	Acc
One-Step (S)	93.49%	91.01%
Two-Step (S)	98.5%	96.88%
One-Step (S + R)	98.57%	96.88%
Two-Step (S + R)	98.65%	96.90%

Table 7. A comparison of the vessel segmentation results on the CHASE_DB1 dataset between the one-step and the two-step methods.

Methods	AUC	Acc
One-Step (S)	66.96%	92.62%
Two-Step (S)	98.64%	97.49%
One-Step (S + R)	98.87%	97.65%
Two-Step (S + R)	99.16%	97.72%

In the end, Tables 8 and 9 compare the proposed approach with other state-of-the-art techniques.

Table 8. A comparison with the state-of-the-art vessel segmentation methods on the DRIVE dataset.

Methods	AUC	Acc
[61]	96.80%	95.93%
[64]	97.38%	95.27%
[62]	97.44%	95.33%
[65]	97.52%	95.42%
[69]	97.82%	95.21%
[59]	97.90%	95.35%
[63]	97.92%	95.60%
[71]	98.21%	95.76%
[19]	98.74%	96.90%
Our	98.65%	96.90%

Table 9. A comparison with the state-of-the-art vessel segmentation on the CHASE_DB1 dataset.

Methods	AUC	Acc
[61]	95.80%	95.91%
[64]	97.16%	95.81%
[65]	97.81%	96.10%
[59]	98.45%	95.77%
[69]	98.12%	95.99%
[71]	98.55%	96.53%
[19]	98.78%	97.37%
Our	99.16%	97.72%

The results show that the proposed approach reaches the state-of-the-art on the DRIVE dataset, where it is only outperfomed by [19], when the AUC measure is used and outperforms all of the other methods on the CHASE_DB1 dataset. It is worth remembering that the experimental setups adopted by the previous approaches are varied and that a perfect comparison was impossible. For example, CHASE_DB1 does not provide an explicit training/test split, and in [64,65], the same split as in this paper was employed, while in [19,69,71] a fourfold cross-validation strategy was applied (in [71], where each fold included three images of one eye and four images of the other). Moreover, in [59], only patches that were fully inside the field of view were considered. However, even with those inevitable experimental limits, the results of Tables 8 and 9 suggest that the proposed method is promising and is at least as good as the best state-of-the-art techniques.

5. Conclusions

In this paper, we proposed a two-stage procedure to generate synthetic retinal images. During the first stage, the semantic label masks, which correspond to the retinal vessels, were generated by a Progressively Growing GAN. Then, an image-to-image translation approach was employed to obtain the retinal images from the label masks. The proposed approach allowed us to generate images with unprecedented high resolution and realism. The reported experiments demonstrate the usefulness of synthetic images, which can be effectively used to train a deep segmentation network. Moreover, if fine-tuning based on real images is applied, after a preliminary learning phase based only on synthetic images, the performance of the segmentation network further improves, reaching the performance of or outperforming the best methods. It is worth noting that the proposed framework for image generation is general and not limited to retinal image generation. The possibility of extending the proposed two-phase generation procedure to different domains is a matter of further investigations.

Author Contributions: Conceptualization, P.A.; data curation, P.A. and S.B.; funding acquisition, A.M., A.S., F.S. and M.B.; investigation, P.A., G.C. and S.B.; methodology, P.A. and S.B.; project administration, A.M., A.S., F.S. and M.B.; resources, A.M., A.S., F.S. and M.B.; software, P.A., G.C. and

S.B.; supervision, P.A., F.S. and M.B.; validation, P.A., G.C., S.B., C.G. and V.L.; visualization, P.A., G.C. and S.B.; writing—original draft, P.A., G.C. and S.B.; writing—review and editing, P.A., G.C., C.G., V.L., A.M., F.S. and M.B. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Patient consent was waived due the anonymous nature of the data.

Data Availability Statement: https://drive.grand-challenge.org/DRIVE/ (accessed on 1 December 2021), https://blogs.kingston.ac.uk/retinal/chasedb1/ (accessed on 1 December 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Patton, N.; Aslam, T.M.; MacGillivray, T.; Deary, I.J.; Dhillon, B.; Eikelboom, R.H.; Yogesan, K.; Constable, I.J. Retinal image analysis: Concepts, applications and potential. *Prog. Retin. Eye Res.* **2006**, *25*, 99–127. [CrossRef] [PubMed]
2. Fraz, M.; Remagnino, P.; Hoppe, A.; Uyyanonvara, B.; Rudnicka, A.; Owen, C.; Barman, S. Blood vessel segmentation methodologies in retinal images—A survey. *Comput. Methods Programs Biomed.* **2012**, *108*, 407–433. [CrossRef]
3. Patil, D.D.; Manza, R.R. Design new algorithm for early detection of primary open angle glaucoma using retinal Optic Cup to Disc Ratio. In Proceedings of the 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 3–5 March 2016; pp. 148–151.
4. Bowling, B. *Kanski's Clinical Ophthalmology: A Systematic Approach*; Saunders Ltd.: Philadelphia, PA, USA, 2015.
5. Abràmoff, M.D.; Garvin, M.K.; Sonka, M. Retinal imaging and image analysis. *IEEE Rev. Biomed. Eng.* **2010**, *3*, 169–208. [CrossRef] [PubMed]
6. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
7. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
8. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
9. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
10. Cheron, G.; Laptev, I.; Schmid, C. P-CNN: Pose-Based CNN Features for Action Recognition. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
11. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
12. Huynh, T.C. Vision-based autonomous bolt-looseness detection method for splice connections: Design, lab-scale evaluation, and field application. *Autom. Constr.* **2021**, *124*, 103591. [CrossRef]
13. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.
14. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *arXiv* **2018**, arXiv:1710.10196.
15. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8798–8807.
16. Staal, J.; Abràmoff, M.D.; Niemeijer, M.; Viergever, M.A.; van Ginneken, B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imaging* **2004**, *23*, 501–509. [CrossRef]
17. Fraz, M.M.; Remagnino, P.; Hoppe, A.; Uyyanonvara, B.; Rudnicka, A.R.; Owen, C.G.; Barman, S. An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2538–2548. [CrossRef]
18. Bonechi, S.; Bianchini, M.; Scarselli, F.; Andreini, P. Weak supervision for generating pixel–level annotations in scene text segmentation. *Pattern Recognit. Lett.* **2020**, *138*, 1–7. [CrossRef]
19. Sekou, T.B.; Hidane, M.; Olivier, J.; Cardot, H. From Patch to Image Segmentation using Fully Convolutional Networks—Application to Retinal Images. *arXiv* **2019**, arXiv:1904.03892.
20. Richter, S.R.; Vineet, V.; Roth, S.; Koltun, V. Playing for data: Ground truth from computer games. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 102–118.
21. Ros, G.; Sellart, L.; Materzynska, J.; Vazquez, D.; Lopez, A.M. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3234–3243.

22.   Hodan, T.; Vineet, V.; Gal, R.; Shalev, E.; Hanzelka, J.; Connell, T.; Urbina, P.; Sinha, S.N.; Guenter, B. Photorealistic Image Synthesis for Object Instance Detection. *arXiv* **2019**, arXiv:1902.03334.

23.   Collins, D.L.; Zijdenbos, A.P.; Kollokian, V.; Sled, J.G.; Kabani, N.J.; Holmes, C.J.; Evans, A.C. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imaging* **1998**, *17*, 463–468. [CrossRef]

24.   Andreini, P.; Bonechi, S.; Bianchini, M.; Mecocci, A.; Scarselli, F. A Deep Learning Approach to Bacterial Colony Segmentation. In Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; pp. 522–533.

25.   Andreini, P.; Bonechi, S.; Bianchini, M.; Mecocci, A.; Scarselli, F. Image generation by GAN and style transfer for agar plate image segmentation. *Comput. Methods Programs Biomed.* **2020**, *184*, 105268. [CrossRef] [PubMed]

26.   Kugelman, J.; Alonso-Caneiro, D.; Read, S.A.; Vincent, S.J.; Chen, F.K.; Collins, M.J. Data augmentation for patch-based OCT chorio-retinal segmentation using generative adversarial networks. *Neural Comput. Appl.* **2021**, *33*, 7393–7408. [CrossRef]

27.   Waheed, A.; Goyal, M.; Gupta, D.; Khanna, A.; Al-Turjman, F.; Pinheiro, P.R. Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access* **2020**, *8*, 91916–91923. [CrossRef] [PubMed]

28.   Shin, H.C.; Tenenholtz, N.A.; Rogers, J.K.; Schwarz, C.G.; Senjem, M.L.; Gunter, J.L.; Andriole, K.P.; Michalski, M. Medical Image Synthesis for Data Augmentation and Anonymization using Generative Adversarial Networks. In Proceedings of the International Workshop on Simulation and Synthesis in Medical Imaging, Granada, Spain, 16 September 2018.

29.   Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.

30.   Pathak, D.; Krähenbühl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context Encoders: Feature Learning by Inpainting. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.

31.   Gatys, L.A.; Ecker, A.S.; Bethge, M. A neural algorithm of artistic style. *arXiv* **2015**, arXiv:1508.06576.

32.   Liu, M.Y.; Breuel, T.; Kautz, J. Unsupervised Image-to-Image Translation Networks. *arXiv* **2017**, arXiv:1703.00848.

33.   Liu, M.Y.; Tuzel, O. Coupled Generative Adversarial Networks. In *Advances in Neural Information Processing Systems 29*; Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2016; pp. 469–477.

34.   Yi, Z.; Zhang, H.; Tan, P.; Gong, M. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2868–2876.

35.   Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2242–2251.

36.   Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5967–5976.

37.   Chen, Q.; Koltun, V. Photographic Image Synthesis with Cascaded Refinement Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 1520–1529.

38.   Sagar, M.; Bullivant, D.P.; Mallinson, G.D.; Hunter, P.J. A virtual environment and model of the eye for surgical simulation. In Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), Orlando, FL, USA, 24–29 July 1994.

39.   Fiorini, S.; Biasi, M.D.; Ballerini, L.; Trucco, E.; Ruggeri, A. Automatic Generation of Synthetic Retinal Fundus Images. In *Smart Tools and Apps for Graphics—Eurographics Italian Chapter Conference*; Giachetti, A., Ed.; The Eurographics Association: Cagliari, Italy, 2014. [CrossRef]

40.   Menti, E.; Bonaldi, L.; Ballerini, L.; Ruggeri, A.; Trucco, E. Automatic Generation of Synthetic Retinal Fundus Images: Vascular Network. In *Simulation and Synthesis in Medical Imaging*; Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 167–176.

41.   Costa, P.; Galdran, A.; Meyer, M.I.; Abramoff, M.D.; Niemeijer, M.; Mendonça, A.M.; Campilho, A. Towards adversarial retinal image synthesis. *arXiv* **2017**, arXiv:1701.08974.

42.   Zhao, H.; Li, H.; Maurer-Stroh, S.; Cheng, L. Synthesizing retinal and neuronal images with generative adversarial nets. *Med. Image Anal.* **2018**, *49*, 14–26. [CrossRef] [PubMed]

43.   Costa, P.; Galdran, A.; Meyer, M.I.; Niemeijer, M.; Abràmoff, M.; Mendonça, A.M.; Campilho, A. End-to-End Adversarial Retinal Image Synthesis. *IEEE Trans. Med. Imaging* **2018**, *37*, 781–791. [CrossRef] [PubMed]

44.   Beers, A.; Brown, J.M.; Chang, K.; Campbell, J.P.; Ostmo, S.; Chiang, M.F.; Kalpathy-Cramer, J. High-resolution medical image synthesis using progressively grown generative adversarial networks. *arXiv* **2018**, arXiv:1805.03144.

45.   Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015.

46.   Liu, I.; Sun, Y. Recursive tracking of vascular networks in angiograms based on the detection-deletion scheme. *IEEE Trans. Med. Imaging* **1993**, *12*, 334–341. [CrossRef]

47.   Yin, Y.; Adel, M.; Bourennane, S. Retinal vessel segmentation using a probabilistic tracking method. *Pattern Recognit.* **2012**, *45*, 1235–1244. [CrossRef]

48.   Hoover, A.D.; Kouznetsova, V.; Goldbaum, M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. Med. Imaging* **2000**, *19*, 203–210. [CrossRef] [PubMed]

49. Roychowdhury, S.; Koozekanani, D.D.; Parhi, K.K. Iterative Vessel Segmentation of Fundus Images. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1738–1749. [CrossRef]

50. Neto, L.C.; Ramalho, G.L.; Neto, J.F.R.; Veras, R.M.; Medeiros, F.N. An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images. *Expert Syst. Appl.* **2017**, *78*, 182–192. [CrossRef]

51. Zhao, Y.; Rada, L.; Chen, K.; Harding, S.P.; Zheng, Y. Automated Vessel Segmentation Using Infinite Perimeter Active Contour Model with Hybrid Region Information with Application to Retinal Images. *IEEE Trans. Med. Imaging* **2015**, *34*, 1797–1807. [CrossRef]

52. Khan, K.B.; Siddique, M.S.; Ahmad, M.; Mazzara, M. A hybrid unsupervised approach for retinal vessel segmentation. *BioMed Res. Int.* **2020**, *2020*, 8365783. [CrossRef]

53. Liu, B.; Gu, L.; Lu, F. Unsupervised ensemble strategy for retinal vessel segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019; pp. 111–119.

54. Khawaja, A.; Khan, T.M.; Khan, M.A.; Nawaz, S.J. A multi-scale directional line detector for retinal vessel segmentation. *Sensors* **2019**, *19*, 4949. [CrossRef]

55. Shah, S.A.A.; Shahzad, A.; Khan, M.A.; Lu, C.K.; Tang, T.B. Unsupervised method for retinal vessel segmentation based on gabor wavelet and multiscale line detector. *IEEE Access* **2019**, *7*, 167221–167228. [CrossRef]

56. Niemeijer, M.; Staal, J.; van Ginneken, B.; Loog, M.; Abramoff, M.D. Comparative study of retinal vessel segmentation methods on a new publicly available database. In Proceedings of the Medical Imaging 2004: Image Processing, San Diego, CA, USA, 16–19 February 2004; Volume 5370, pp. 2672–2680. [CrossRef]

57. Soares, J.V.B.; Leandro, J.J.G.; Cesar, R.M.; Jelinek, H.F.; Cree, M.J. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans. Med. Imaging* **2006**, *25*, 1214–1222. [CrossRef] [PubMed]

58. Toptaş, B.; Hanbay, D. Retinal blood vessel segmentation using pixel-based feature vector. *Biomed. Signal Process. Control* **2021**, *70*, 103053. [CrossRef]

59. Liskowski, P.; Krawiec, K. Segmenting Retinal Blood Vessels with Deep Neural Networks. *IEEE Trans. Med. Imaging* **2016**, *35*, 2369–2380. [CrossRef] [PubMed]

60. Hajabdollahi, M.; Esfandiarpoor, R.; Najarian, K.; Karimi, N.; Samavi, S.; Reza-Soroushmeh, S. Low complexity convolutional neural network for vessel segmentation in portable retinal diagnostic devices. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 2785–2789.

61. Jiang, Z.; Zhang, H.; Wang, Y.; Ko, S.B. Retinal blood vessel segmentation using fully convolutional network with transfer learning. *Comput. Med. Imaging Graph.* **2018**, *68*, 1–15. [CrossRef] [PubMed]

62. Dasgupta, A.; Singh, S. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; pp. 248–251. [CrossRef]

63. Feng, Z.; Yang, J.; Yao, L. Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1742–1746.

64. Li, Q.; Feng, B.; Xie, L.; Liang, P.; Zhang, H.; Wang, T. A Cross-Modality Learning Approach for Vessel Segmentation in Retinal Images. *IEEE Trans. Med. Imaging* **2016**, *35*, 109–118. [CrossRef] [PubMed]

65. Yan, Z.; Yang, X.; Cheng, K.T.T. Joint Segment-Level and Pixel-Wise Losses for Deep Learning Based Retinal Vessel Segmentation. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 1912–1923. [CrossRef] [PubMed]

66. Xie, S.; Tu, Z. Holistically-Nested Edge Detection. *Int. J. Comput. Vis.* **2017**, *125*, 3–18. [CrossRef]

67. Fu, H.; Xu, Y.; Wong, D.W.K.; Liu, J. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016; pp. 698–701.

68. Liu, S.; Deng, W. Very deep convolutional neural network based image classification using small training sample size. In Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 3–6 November 2015; pp. 730–734. [CrossRef]

69. Mo, J.; Zhang, L. Multi-level deep supervised networks for retinal vessel segmentation. *Int. J. Comput. Assist. Radiol. Surg.* **2017**, *12*, 2181–2193. [CrossRef] [PubMed]

70. Maninis, K.; Pont-Tuset, J.; Arbeláez, P.; Gool, L.V. Deep Retinal Image Understanding. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Athens, Greece, 17–21 October 2016.

71. Oliveira, A.; Pereira, S.; Silva, C.A. Retinal vessel segmentation based on Fully Convolutional Neural Networks. *Expert Syst. Appl.* **2018**, *112*, 229–242. [CrossRef]

72. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved Training of Wasserstein GANs. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017; pp. 5769–5779.

73. Serra, J. *Image Analysis and Mathematical Morphology*; Academic Press, Inc.: Orlando, FL, USA, 1983.

74. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

75. Papandreou, G.; Kokkinos, I.; Savalle, P.A. Untangling local and global deformations in deep convolutional networks for image classification and sliding window detection. *arXiv* **2014**, arXiv:1412.0296.
76. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
77. Shmelkov, K.; Schmid, C.; Alahari, K. How Good Is My GAN? In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.