

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

Paolo Andreini
University of Siena

Simone Bonechi
University of Siena

Franco Scarselli
University of Siena

Monica Bianchini
University of Siena

Alessandro Mecocci
University of Siena

Andrea Sodi
University of Florence

Abstract

In recent years, the use of deep learning is becoming increasingly popular in computer vision. However, the effective training of deep architectures usually relies on huge sets of annotated data. This is critical in the medical field where it is difficult and expensive to obtain annotated images. In this paper, we use Generative Adversarial Networks (GANs) for synthesizing high quality retinal images, along with the corresponding semantic label-maps, to be used instead of real images during the training process. Differently from other previous proposals, we suggest a two step approach: first, a progressively growing GAN is trained to generate the semantic label-maps, which describe the blood vessel structure (i.e. vasculature); second, an image-to-image translation approach is used to obtain realistic retinal images from the generated vasculature. By using only a handful of training samples, our approach generates realistic high resolution images, that can be effectively used to enlarge small available datasets. Comparable results have been obtained employing the generated images in place of real data during training. The practical viability of the proposed approach has been demonstrated by applying it on two well established benchmark sets for retinal vessel segmentation, both containing a very small number of training samples. Our method obtained better performances with respect to state-of-the-art techniques.

1. Introduction

The retinal microvasculature is the only part of the human circulation that can be directly and non-invasively visualized *in vivo* [50]. Hence, it can be easily acquired and analyzed by automatic tools. As a result, retinal fundus images have a multitude of applications, ranging from biometric identification, to computer-assisted laser surgery, to the diagnosis of several disorders [16]. One important processing step in such applications is the proper segmentation of

the retinal vessels. For this reason, we propose a new deep learning approach for retinal image generation and vessel segmentation. Image semantic segmentation aims at making dense predictions by inferring the object class for each pixel of an image. The segmentation of digital retina images allows to extract various quantitative vessel parameters, in order to obtain more objective and accurate medical diagnosis. In particular, the segmentation of retinal blood vessels, can help the diagnosis, treatment, and monitoring of diseases such as diabetic retinopathy, hypertension, and arteriosclerosis [6, 1].

It is widely recognized that Deep Neural Networks (DNNs) are becoming the standard approach in semantic segmentation [40, 7, 69] and in many other computer vision tasks [32, 9, 24]. DNN training, however, requires large sets of accurately labeled data, so that 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 a complex unsolved challenge. Most of the leading approaches for semantic segmentation rely on thousands of supervised images while supervised public datasets for retinal vessel segmentation are very small (most datasets contain less 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) [21] learns to generate 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 [30] (PGGAN) approach has been employed. In a second, and distinct phase, an image-to-image translation algorithm [62] is used to translate the blood vessels structures into realistic retinal images

(see Fig. 1).

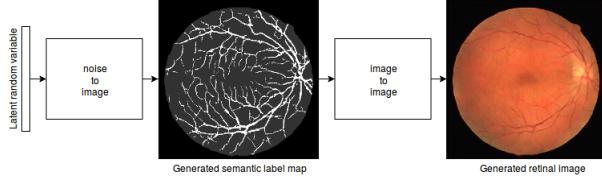


Figure 1: The proposed two-step image generation scheme.

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. Separating the whole process into two phases, we obtained retinal images with an unprecedented high resolution and quality, along with their semantic label-maps. Furthermore, the number of samples required for training is significantly reduced. It is worth noting that the proposed two-step approach reduces the GPU memory requirements w.r.t. a single step method. The generation of the label-maps through a GAN, allows 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 [61] and CHASE_DB1 [17]). The generated data have been used to train a Segmentation Multiscale Attention Network (SMANet) [5].

The SMANet is a deep convolutional neural network, originally developed for text segmentation in outdoor/indoor scenes. Its architecture is based on the Pyramid Scene Parsing Network [69], which is a popular semantic segmentation network. The SMANet employs a two-level convolutional decoder and a multi-scale attention mechanism to better detect small objects at different scales. This multi-scale detection capability is fundamental in case of retinal vessel segmentation, because the vessels show different characteristics depending on their diameter and spatial location. 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 [56]. 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.

The 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 experimental results. Finally, Section 5 draws conclusions and future perspectives.

2. Related works

2.1. Synthetic Image Generation

Methods for generating images are by no means new, and can be classified into two main categories: model-based approaches and learning-based approaches. The most conventional approach is to formulate a model of the observed data and to render the images by a dedicated engine. This approach has been used, for example, to extend the available datasets of driving scenes in urban environments [51], [53] or for object detection [26], and text segmentation [4]. Also in the field of medical image analysis, synthetic image generation has been extensively employed. For example, realistic digital brain-phantom have been synthesized in [10] while, more recently, synthetic agar plate images have been generated for image segmentation [2]. The design of specialized engines for data generation requires an accurate model of the scene and a deep knowledge of the specific domain. For this reason, in recent years, the learning-based approach 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 model for data generation is the Generative Adversarial Network (GAN) [21]. 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 at distinguishing the fake samples from real ones, $\mathbf{x} \in p_{data}(x)$. The GAN training is formulated as a min-max game between G and D :

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_r(x)} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

One example of GANs used for data augmentation in the medical field has been given in [18] for the classification of liver lesions, and in [58] to generate synthetic abnormal MRI images containing by brain tumors.

2.2. Image-To-Image Translation

Recently, beside image generation, adversarial learning has been also extended to the image-to-image translation, whose goal is to translate an input image from one domain to another. Many computer vision tasks, such as image super-resolution [33], image inpainting [49], and style transfer [20] can be casted into the image-to-image translation framework. Both unsupervised [37], [38], [65], [71]

and supervised approaches can be used [28], [30], [8]. 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 a corresponding image in the target domain. As an example, Pix2Pix [28] consists of a conditional GAN that operates in a supervised way, and Pix2PixHD [30] 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 the seminal work [55], in which an anatomic model of the eye and of the surrounding face has been implemented for surgical simulations. More recently, in [15], a large dictionary of small image patches containing no vessels, has been used to model the retinal background and fovea. A parametric intensity model, whose parameters have been estimated from real images, is used to generate the optical disk. Complementary to [15], the contribution in [43] focuses on the generation of the vascular network, based on a parametric model whose parameters are learned from real vessel trees. Despite these methods provide reasonable results, they are complex and heavily depend on domain knowledge. To reduce the domain knowledge requirements, a completely learning-based approach has been proposed in [11], where an image-to-image translation model has been employed to transform existing vessel networks into realistic retinal images. The 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 performances of the segmentation module. In [68], 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. variability) of the synthesized images. The model still relies on pre-existing vessel-networks (obtained manually or by a suitable segmentation technique). In [12], the use of an adversarial auto-encoder for the task of retinal-vessel synthesis has been adopted to avoid the dependence of the model on the availability of preexistent vessel maps. Nevertheless, this approach allows to generate only low resolution images, and the performance in vessel segmentation by using the synthesized data, is far below the state-of-the-art. Higher resolution retinal images, along with their segmentation label-maps, have been generated in [3] with an approach based on a Progressively Growing GAN (PGGAN)[30]. The method allows to generate images up to a resolution of 512×512 pixels. A set of 5550 images, segmented by a pre-trained U-Net [52] 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 preexisting images (actually, 20 images with supervised segmentation maps). Both the retinal images and the corresponding semantic label-maps (the vasculature) are generated. Furthermore, we prove that combining real retinal images with synthesized ones during the training of a segmentation network, improves the final segmentation performance.

2.4. Retinal Vessel Segmentation

During the last decades, several approaches for retinal vessel segmentation have been proposed, both supervised and unsupervised. Unsupervised methods depend heavily on prior knowledge about the vessel structure. For example, Vessel Tracking Techniques define an initial set of seed points and, thereafter, by chaining pixels that minimize a given cost function, the vasculature is iteratively extracted [36][66]. In [27], retinal images are convolved with a 2D filter to produce a Gaussian intensity profile of the blood vessels, that are subsequently thresholded to give the vessels map. Adaptive thresholding has been used in [54] and [45]. An Active Contour Model, that combines intensity and local phase information, is used in [70]. Supervised methods are currently the leading techniques in semantic segmentation. In this framework, truth 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. A k-Nearest Neighbor classifier is used in [46], which adopts a pixel-wise feature vector, based on Gaussian functions and their derivatives. In [60], a Gaussian Mixture Model classifier is applied to the pixel intensities augmented by coefficients obtained through a 2D-Gabor Wavelet Transform, evaluated at multiple scales. In [42], a neural network is used to classify vectors comprising gray-levels and moment invariant features. A random forest classifier, applied on a 29-dimensional feature vector, has been used in [67]. Supervised methods are strongly affected by the feature engineering stage.

Deep learning-based methods automatically learn from the input data an increasingly complex hierarchy of features, bypassing the need for problem specific knowledge. In retinal image segmentation a deep convolutional neural network (DCNN) is used in [35] where the training examples are subject to various preprocessing 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 [23]. In [29], it is employed a fully convolutional network [40] with an AlexNet [32] encoder. Fully convolutional networks have been used also in [13] and [14]. In [34] the task of segmentation is remolded into a problem of cross-modality data

transformation from retinal images to vessel map. A modified U-Net [52] is used in [64] which exploits a combination between a segment-level loss and a pixel-level loss, to deal with the unbalanced ratio between thick and thin vessels in fundus images. A Holistically-Nested Edge Detection (HED) network [63], originally designed for edge detection, followed by a conditional random field are employed for the retinal blood vessel segmentation in [19]. Deep supervision is incorporated in some intermediate layers of a VGG network [39] in [44] and [41]. In [47] a Fully Convolutional Neural Network uses a stationary wavelet transform pre-processing step to improve the network performance. Finally, in [56], a CNN is pre-train on image patches and then fine tuned at the image level.

3. Retinal Image Generation

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 different phases: the first one is related to the generation of semantic label-maps of the vessels, and the second to the synthesis of realistic images starting from those label-maps. The quality and usefulness of the generated images have been validated by the performance obtained on two public benchmark datasets using the synthesized images to train a segmentation network. In particular, Section 3.1 gives an overview of the approach used to generate the semantic label-maps. Section 3.2 describes the image-to-image translation algorithm which synthesizes retinal images from the semantic label-maps. Instead, Section 3.3 describes the semantic segmentation network used to segment the retinal vessels. Finally, some details about the training method are reported in Section 3.4.

3.1. Vasculature Generation

The generation of the vessel structure is based on the PGGAN approach, 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 risks related to the lack of an adequate descriptive power, due to the very limited number of available training samples, data augmentation has been applied. Specifically, the semantic label-maps have been slightly rotated ($\pm 15^\circ$) and flipped in different ways (horizontal, vertical and horizontal followed by vertical flips). The generation starts at low-resolution and then the resolution is progressively increased by adding new layers to the networks. The generator and the discriminator are symmetric and grow in sync. The transition from low-resolution image generation to high-resolution image generation follows the procedure described in [30] to avoid the problems related to a sudden transition. The training

starts with both the generator and the discriminator having a low spatial resolution (e.g. 4×4 pixels), then the resolution increases progressively until the final resolving power is reached. The Wasserstein loss, using a gradient penalty [22], has been used as loss function for the discriminator. The learning procedure is illustrated in Fig. 2.

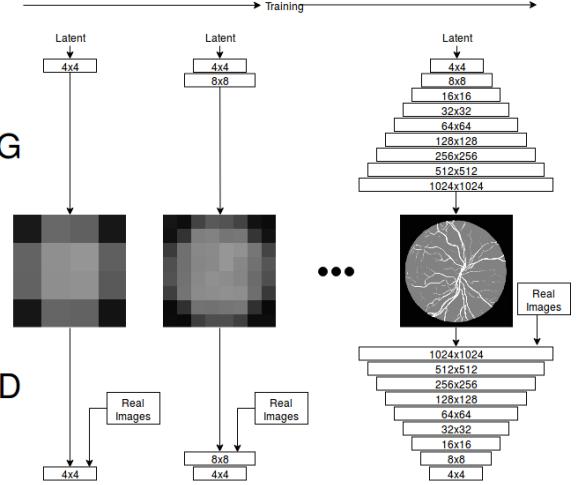


Figure 2: Training schema of the semantic label-maps generation.

It can be observed that the global structure of the vessel distribution is learned at the beginning of the training, whereas finer details are added as the resolution increases. The generation procedure allows to obtain a virtually infinite number of different vasculatures. To reduce the probability of introducing artifacts, a simple post-processing has been carried out. Specifically, a morphological opening [57] has been applied to the generated retinal fundus mask to improve its circularity. Small holes have been filled, and segments of small dimension have been removed from the generated vessel structure.

3.2. Translating Vessel Maps into Retinal Images

Once the vessel networks have been obtained, they must be transformed into realistic color retinal images. Our method is based on Pix2PixHD [30], a supervised image-to-image translation framework based on Pix2Pix [28]. In Pix2Pix, a conditional GAN learns to generate the output conditioned on the corresponding input image. The generator has an encoder-decoder structure, and takes as input the images belonging to a certain domain A and generates images in a different domain B . The discriminator observes couples of images, the image from A is provided as input along with the corresponding image of B (real or generated). The discriminator aims at distinguishing between real and fake (generated) couples. Pix2PixHD improves upon

Pix2Pix by introducing a coarse-to-fine generator composed of two subnetworks that operate at different resolutions. A multiscale discriminator is also employed, with an adversarial loss which incorporates a feature-matching loss for training stabilization. In our setup, the semantic label-maps, generated in the previous step, are given in input to the generator that is trained to generate realistic retinal images. Images have been resized to the nearest power-of-two resolution (i.e. the retinal images in the DRIVE dataset that have a resolution of 565×584 pixels have been resized to 512×512 pixels, whereas the CHASE dataset images that have a resolution of 999×960 pixels have been resized to 1024×1024 pixels).

An overview of the proposed setup is given in Fig. 3.

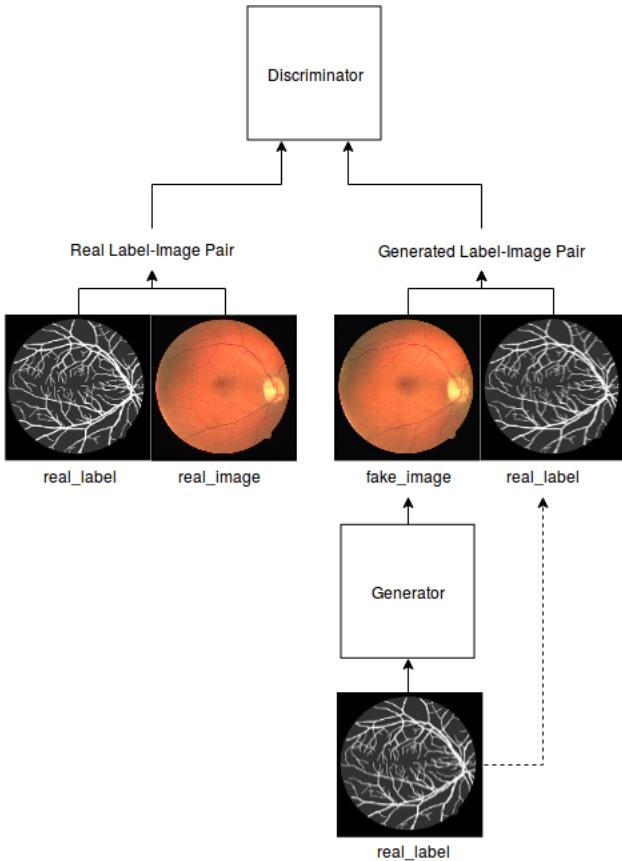


Figure 3: Scheme of the Pix2Pix training framework employed to translate label-maps into retinal images.

3.3. SMArt Architecture

The semantic segmentation network employed in this paper is a Segmentation Multiscale Attention Network (SMArt) [5]. The SMArt, originally proposed for scene

text segmentation, comprises three main components: a ResNet encoder, a multi-scale attention module, and a convolutional decoder (see Fig. 4).

The architecture is based on the PSPNet [69], a deep fully convolutional neural network with a ResNet [25] 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 convolution [48]). Moreover, in the PSPNet, a pyramid of pooling with different kernel size has been employed to gather context information. The pooled feature maps are then up-sampled at the same resolution of 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 SMArt, a multi scale 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.

3.4. Training Details

The SMArt, used in this work, is implemented in TensorFlow. Random crops of 281×281 pixels have been employed during training, whereas a sliding window of the same size has been used for the evaluation. The Adam optimizer [31], based on a learning rate of 10^{-4} and a minibatch of 17 examples, has been used to train the SMArt. All the experiments have been carried out in a Linux environment on a single NVIDIA Tesla V100 SXM2 with 32 GB RAM.

4. Experiments and Results

4.1. The benchmark datasets

- **DRIVE dataset** – The DRIVE dataset [61] includes 40 retinal-fundus images of size $584 \times 565 \times 3$ (20 images are for training and 20 for test). The images have been collected by a screening program for diabetic retinopathy in the Netherlands. Among the 40 photographs, 33 show no diabetic retinopathy, while 7 show mild early diabetic retinopathy. Segmentation ground-truth is provided both for training and 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 have adopted the same strategy of [64] and [34], selecting the first 20 images for training and the remaining 8 for test.

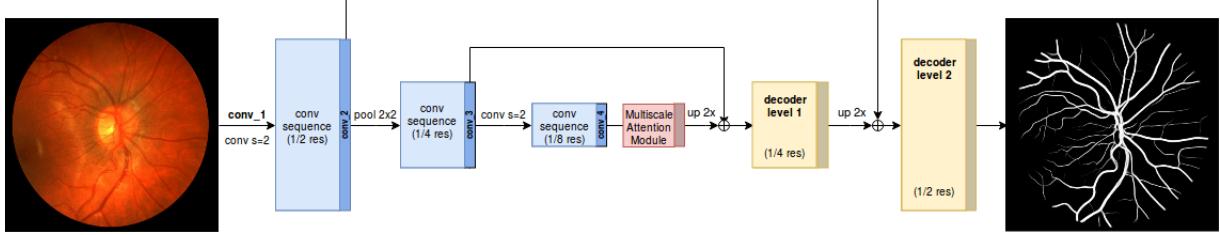


Figure 4: Scheme of the SMA-Net segmentation network.

4.2. Experimental Results

In this paper, we provide both a qualitative and a quantitative evaluation of the generated data. In particular, the quantitative analysis consists in evaluating the usefulness of the generated images for training a semantic segmentation network. This approach is similar to [59] and it is based on the assumption that the performances of a deep learning architecture can be directly related with the quality and variety of GAN generated images. Some qualitative results of the generated retinal images for the DRIVE and CHASE_DB1 dataset are given in Figs. 5–6.

In Fig. 7, a zoom on a random patch of a high resolution generated image shows that the image-to-image translation allows to effectively translate the generated vessel structures in retinal images maintaining the semantic information provided by the semantic label map.

It must be noted that, even if the major part of the generated samples accurately resembles real retinal fundus images, few examples are evidently suboptimal (see Fig. 8 that shows disconnected vessels and an unrealistic optical disc).

In Table 1, the proposed method for retinal image generation is compared with other learning-based approaches found in literature.

Methods	Gen. Vessels	Max Res.	Samples
Costa et al. [11]	No	512 × 512	614
Zhao et al. [68]	No	2048 × 2048	10-20
Costa et al. [12]	Yes	256 × 256	634
Beers et al. [3]	Yes	512 × 512	5550
Our	Yes	1024 × 1024	20

Table 1: Comparison among different generation approaches.

The generation procedure described in Section 3 has been employed to generate 10000 synthetic retinal images for both the DRIVE and the CHASE_DB1 datasets. To eval-

uate the usefulness of the generated data for semantic segmentation, we employed the following experimental set up:

- SYNTH – the semantic segmentation network is trained by using only the 10000 generated synthetic images.
- REAL – only real data are used to train the semantic segmentation network.
- SYNTH + REAL – synthetic data are used to pre-train the semantic segmentation network and real data are employed for fine-tuning.

In Table 2 and Table 3, the results of the vessel segmentation for the DRIVE and CHASE_DB1 datasets, obtained using the previously described approaches, are respectively reported.

Methods	AUC	Acc
SYNTH	98.5 %	97.9 %
REAL	98.48 %	96.87 %
SYNTH + REAL	98.65 %	96.9 %

Table 2: Evaluation of the use of generated data on the DRIVE dataset.

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

Table 3: Evaluation of the use of generated data on the CHASE_DB1 dataset.

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 the generated images effectively capture the training image distribution, so that they can be used to adequately

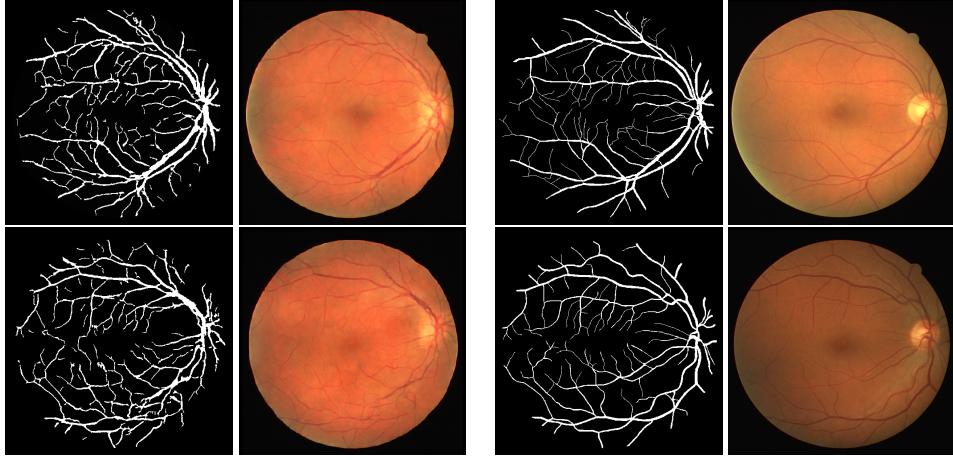


Figure 5: Examples of generated (left) and corresponding real images (right) for DRIVE.



Figure 6: Examples of generated (left) and corresponding real images (right) for CHASE_DB1.

train a deep neural network. Moreover, if fine-tuning with real data is applied after a pre-training with synthetic data only, the results further improve w.r.t. 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. A comparison with other state-of-the-art techniques applied to the two benchmarks, is reported in Table 4 and 5.

To compare the different retinal blood vessel segmentation methods is somehow difficult in datasets for which an explicit train-test split is not given (e.g CHASE_DB1), because the split may differ from one paper to another. For instance, [64] and [34] use the same split employed in this paper, while [44], [56] and [47] use a 4-fold cross-validation strategy (in [47] each fold included 3 images of one eye and 4 images of the other), whereas [35] only considers patches

Methods	AUC	Acc
Jiang et al.[29]	96.80 %	95.93 %
Li et al. [34]	97.38 %	95.27 %
Dasgupta and Singh [13]	97.44 %	95.33 %
Yan et al. [64]	97.52 %	95.42 %
Mo and Zhang [44]	97.82 %	95.21 %
Liskowski and Krawiec [35]	97.90 %	95.35 %
Feng et al. [14]	97.92 %	95.60 %
Oliveira et al. [47]	98.21 %	95.76 %
Sekou et al. [56]	98.74 %	96.90 %
Our	98.65 %	96.90 %

Table 4: A comparison of vessel segmentation results on the DRIVE dataset.

that are fully inside the field of view. However, our method demonstrates improved (or at least state-of-the-art) performances on both the DRIVE and CHASE_DB1 datasets.

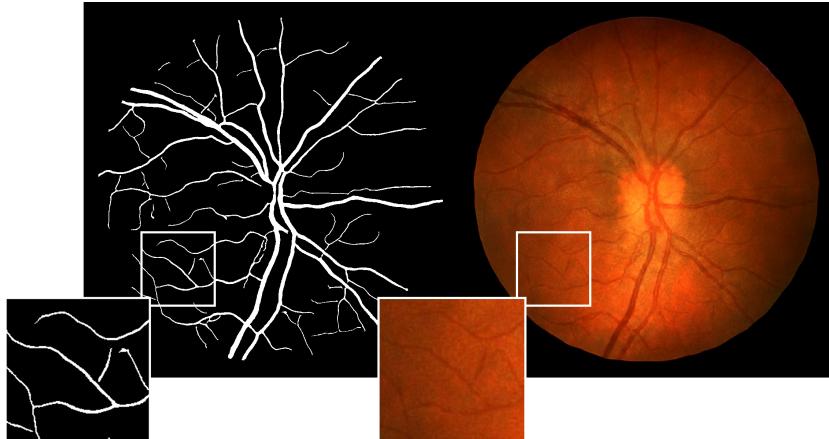


Figure 7: An example of a generated image with resolution 1024×1024 , and of the corresponding label map, for the CHASE_DB1 dataset.

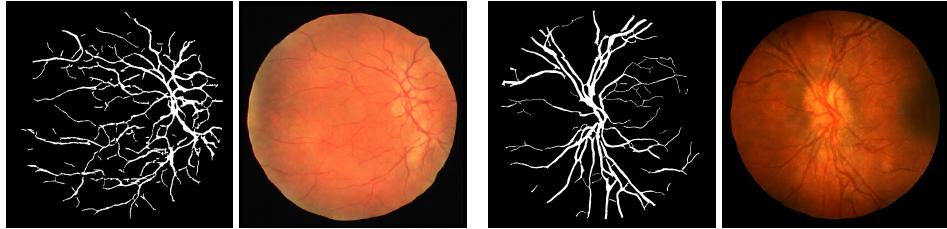


Figure 8: Examples of DRIVE and CHASE generated images with unrealistic optical disc and vasculature (left and right respectively).

Methods	AUC	Acc
Jiang et al. [29]	95.80 %	95.91 %
Li et al. [34]	97.16 %	95.81 %
Yan et al. [64]	97.81 %	96.10 %
Liskowski and Krawiec [35]	98.45 %	95.77 %
Mo and Zhang [44]	98.12 %	95.99 %
Oliveira et al. [47]	98.55 %	96.53 %
Sekou et al. [56]	98.78 %	97.37 %
Our	99.16 %	97.72 %

Table 5: A comparison of vessel segmentation results on the CHASE dataset.

5. Conclusions and future perspectives

In this paper, we have proposed a two stage procedure to generate synthetic retinal images. During the first stage, the semantic label masks, which correspond to the retinal vessels, are generated by a Progressively Growing GAN. Then

an image-to-image translation approach is employed to obtain the retinal images from the label masks. The proposed approach allows to generate images with unprecedented high resolution and realism. The experiments demonstrate the usefulness of the synthetic images, that can be effectively used to train a deep segmentation network. Moreover, if a fine-tuning based on real images is applied after a preliminary learning phase based only on synthetic images, the performances of the segmentation network further improve, reaching or outperforming the state-of-the-art methods. It is worth noting that the proposed framework for image generation is general and not limited to retinal image generation. It is a matter of further investigation the possibility of extending the proposed two-phase generation procedure to different domains.

References

- [1] M. D. Abràmoff, M. K. Garvin, and M. Sonka. Retinal imaging and image analysis. *IEEE reviews in biomedical engineering*, 2013.

- ical engineering*, 3:169–208, 2010.
- [2] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, and F. Scarselli. A deep learning approach to bacterial colony segmentation. In *International Conference on Artificial Neural Networks*, pages 522–533. Springer, 2018.
- [3] A. Beers, J. M. Brown, K. Chang, J. P. Campbell, S. Ostmo, M. F. Chiang, and J. Kalpathy-Cramer. High-resolution medical image synthesis using progressively grown generative adversarial networks. *CoRR*, abs/1805.03144, 2018.
- [4] S. Bonechi, P. Andreini, M. Bianchini, and F. Scarselli. Coco_ts dataset: Pixel-level annotations based on weak supervision for scene text segmentation. *CoRR*, abs/1904.00818, 2019.
- [5] S. Bonechi, P. Andreini, M. Bianchini, and F. Scarselli. Weak supervision for generating pixel-level annotations in scene text segmentation. 2019. Available at http://www3.diism.unisi.it/priv/papers/papers_doc/84.pdf.
- [6] B. Bowling. *Kanski’s clinical ophthalmology: a systematic approach*. Saunders Ltd, 2015.
- [7] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.
- [8] Q. Chen and V. Koltun. Photographic image synthesis with cascaded refinement networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1520–1529, 2017.
- [9] G. Cheron, I. Laptev, and C. Schmid. P-cnn: Pose-based cnn features for action recognition. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- [10] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans. Design and construction of a realistic digital brain phantom. *IEEE transactions on medical imaging*, 17(3):463–468, 1998.
- [11] P. Costa, A. Galdran, M. Meyer, M. Abramoff, M. Niemeijer, A. Mendona, and A. Campilho. Towards adversarial retinal image synthesis, 01 2017.
- [12] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abramoff, A. M. Mendona, and A. Campilho. End-to-end adversarial retinal image synthesis. *IEEE Transactions on Medical Imaging*, 37(3):781–791, March 2018.
- [13] A. Dasgupta and S. Singh. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 248–251, April 2017.
- [14] Z. Feng, J. Yang, and L. Yao. Patch-based fully convolutional neural network with skip connections for retinal blood vessel segmentation. *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1742–1746, 2017.
- [15] S. Fiorini, M. D. Biasi, L. Ballerini, E. Trucco, and A. Ruggeri. Automatic Generation of Synthetic Retinal Fundus Images. In A. Giachetti, editor, *Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference*. The Eurographics Association, 2014.
- [16] M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. Rudnicka, C. Owen, and S. Barman. Blood vessel segmentation methodologies in retinal images a survey. *Computer Methods and Programs in Biomedicine*, 108(1):407 – 433, 2012.
- [17] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. Barman. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59:2538–2548, 2012.
- [18] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan. Synthetic data augmentation using gan for improved liver lesion classification. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 289–293, 2018.
- [19] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 698–701, 2016.
- [20] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015.
- [21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [22] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, pages 5769–5779, USA, 2017. Curran Associates Inc.

- [23] M. Hajabdollahi, R. Esfandiarpoor, K. Najarian, N. Karimi, S. Samavi, and S. Reza-Soroushmeh. Low complexity convolutional neural network for vessel segmentation in portable retinal diagnostic devices. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2785–2789. IEEE, 2018.
- [24] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [25] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [26] T. Hodan, V. Vineet, R. Gal, E. Shalev, J. Hanzelka, T. Connell, P. Urbina, S. N. Sinha, and B. Guenter. Photorealistic image synthesis for object instance detection. *arXiv preprint arXiv:1902.03334*, 2019.
- [27] A. D. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210, March 2000.
- [28] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.
- [29] Z. Jiang, H. Zhang, Y. Wang, and S.-B. Ko. Retinal blood vessel segmentation using fully convolutional network with transfer learning. *Computerized Medical Imaging and Graphics*, 68:1 – 15, 2018.
- [30] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *CoRR*, abs/1710.10196, 2018.
- [31] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [32] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [33] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, 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*, pages 4681–4690, 2017.
- [34] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang. A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Transactions on Medical Imaging*, 35(1):109–118, Jan 2016.
- [35] P. Liskowski and K. Krawiec. Segmenting retinal blood vessels with deep neural networks. *IEEE Transactions on Medical Imaging*, 35(11):2369–2380, Nov 2016.
- [36] I. Liu and Y. Sun. Recursive tracking of vascular networks in angiograms based on the detection-deletion scheme. *IEEE Transactions on Medical Imaging*, 12(2):334–341, June 1993.
- [37] M.-Y. Liu, T. Breuel, and J. Kautz. Unsupervised image-to-image translation networks. *ArXiv*, abs/1703.00848, 2017.
- [38] M.-Y. Liu and O. Tuzel. Coupled generative adversarial networks. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 469–477. Curran Associates, Inc., 2016.
- [39] S. Liu and W. Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 730–734, Nov 2015.
- [40] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [41] K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. V. Gool. Deep retinal image understanding. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2016.
- [42] D. Marin, A. Aquino, M. E. Gegundez-Arias, and J. M. Bravo. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Transactions on Medical Imaging*, 30(1):146–158, Jan 2011.
- [43] E. Menti, L. Bonaldi, L. Ballerini, A. Ruggeri, and E. Trucco. Automatic generation of synthetic retinal fundus images: Vascular network. In S. A. Tsaftaris, A. Gooya, A. F. Frangi, and J. L. Prince, editors, *Simulation and Synthesis in Medical Imaging*, pages 167–176, Cham, 2016. Springer International Publishing.
- [44] J. Mo and L. Zhang. Multi-level deep supervised networks for retinal vessel segmentation. *International Journal of Computer Assisted Radiology and Surgery*, 12(12):2181–2193, Dec 2017.
- [45] L. C. Neto, G. L. Ramalho, J. F. R. Neto, R. M. Veras, and F. N. Medeiros. An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images. *Expert Systems with Applications*, 78:182 – 192, 2017.
- [46] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff. Comparative study of retinal ves-

- sel segmentation methods on a new publicly available database. In *Advances in neural information processing systems*, volume 5370, pages 2672–2680, 2004.
- [47] A. Oliveira, S. Pereira, and C. A. Silva. Retinal vessel segmentation based on fully convolutional neural networks. *Expert Systems with Applications*, 112:229 – 242, 2018.
- [48] G. Papandreou, I. Kokkinos, and P.-A. Savalle. Untangling local and global deformations in deep convolutional networks for image classification and sliding window detection. *arXiv preprint arXiv:1412.0296*, 2014.
- [49] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2536–2544, 2016.
- [50] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yugesan, and I. J. Constable. Retinal image analysis: Concepts, applications and potential. *Progress in Retinal and Eye Research*, 25(1):99 – 127, 2006.
- [51] S. R. Richter, V. Vineet, S. Roth, and V. Koltun. Playing for data: Ground truth from computer games. In *European Conference on Computer Vision*, pages 102–118. Springer, 2016.
- [52] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.
- [53] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez. 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*, pages 3234–3243, 2016.
- [54] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi. Iterative vessel segmentation of fundus images. *IEEE Transactions on Biomedical Engineering*, 62(7):1738–1749, July 2015.
- [55] M. Sagar, D. P. Bullivant, G. D. Mallinson, and P. J. Hunter. A virtual environment and model of the eye for surgical simulation. In *SIGGRAPH*, 1994.
- [56] T. B. Sekou, M. Hidane, J. Olivier, and H. Cardot. From patch to image segmentation using fully convolutional networks - application to retinal images. *ArXiv*, abs/1904.03892, 2019.
- [57] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA, 1983.
- [58] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *SASHIMI@MICCAI*, 2018.
- [59] K. Shmelkov, C. Schmid, and K. Alahari. How good is my gan? In *ECCV*, 2018.
- [60] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, 25(9):1214–1222, Sep. 2006.
- [61] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23:501–509, 2004.
- [62] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8798–8807, 2018.
- [63] S. Xie and Z. Tu. Holistically-nested edge detection. *Int. J. Comput. Vision*, 125(1-3):3–18, Dec. 2017.
- [64] Z. Yan, X. Yang, and K. T. T. Cheng. Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 65:1912–1923, 2018.
- [65] Z. Yi, H. Zhang, P. Tan, and M. Gong. Dualgan: Unsupervised dual learning for image-to-image translation. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2868–2876, 2017.
- [66] Y. Yin, M. Adel, and S. Bourennane. Retinal vessel segmentation using a probabilistic tracking method. *Pattern Recognition*, 45(4):1235 – 1244, 2012.
- [67] J. Zhang, Y. Chen, E. Bekkers, M. Wang, B. Dashtbozorg, and B. M. ter Haar Romeny. Retinal vessel delineation using a brain-inspired wavelet transform and random forest. *Pattern Recognition*, 69:107 – 123, 2017.
- [68] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng. Synthesizing retinal and neuronal images with generative adversarial nets. *Medical Image Analysis*, 49:14 – 26, 2018.
- [69] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6230–6239, 2017.
- [70] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. *IEEE Transactions on Medical Imaging*, 34(9):1797–1807, Sep. 2015.

- [71] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017.