Adversarial Synthesis of Retinal Images from Vessel Trees

Pedro Costa[1(✉)], Adrian Galdran[1], Maria Ines Meyer[1], Ana Maria Mendonça[1,2], and Aurélio Campilho[1,2]

[1] INESC TEC - Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal
{pvcosta,adrian.galdran,maria.i.meyer}@inesctec.pt
[2] Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
{amendon,campilho}@fe.up.pt

Abstract. Synthesizing images of the eye fundus is a challenging task that has been previously approached by formulating complex models of the anatomy of the eye. New images can then be generated by sampling a suitable parameter space. Here we propose a method that learns to synthesize eye fundus images directly from data. For that, we pair true eye fundus images with their respective vessel trees, by means of a vessel segmentation technique. These pairs are then used to learn a mapping from a binary vessel tree to a new retinal image. For this purpose, we use a recent image-to-image translation technique, based on the idea of adversarial learning. Experimental results show that the original and the generated images are visually different in terms of their global appearance, in spite of sharing the same vessel tree. Additionally, a quantitative quality analysis of the synthetic retinal images confirms that the produced images retain a high proportion of the true image set quality.

Keywords: Retinal image synthesis · Generative adversarial learning

1 Introduction

Modern machine learning methods require large amounts of training data. This data is rarely available in the field of medical image analysis, since obtaining clinical annotations is often a costly process. Therefore, the possibility of synthetically generating medical visual data is greatly appealing, and has been explored for years. However, the realistic generation of high-quality medical imagery still remains a complex unsolved challenge for current computer vision methods.

Early methods for medical image generation consisted of digital phantoms, following simplified mathematical models of human anatomy [2]. These models slowly evolved to more complex techniques, able to reliably model relevant aspects of the different acquisition devices. When combined with anatomical and physiological information arising from expert medical knowledge, realistic images can be produced [4]. These are useful to validate image analysis techniques, for medical training, therapy planning, and a wide range of applications [6,11].

F. Karray et al. (Eds.): ICIAR 2017, LNCS 10317, pp. 516–523, 2017.
DOI: 10.1007/978-3-319-59876-5_57

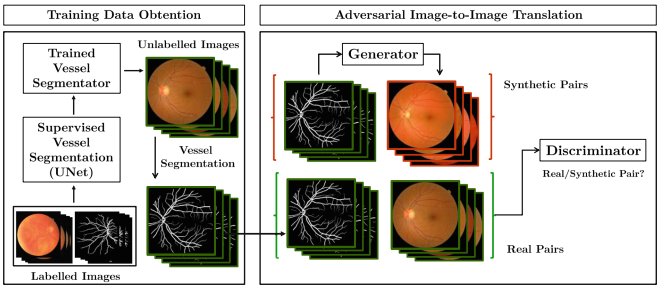

Fig. 1. Overview of the proposed retinal image generation method.

However, the traditional top-down approach of observing the available data and formulating mathematical models that explain it (*image simulation*) implies modeling complex natural laws by unavoidably simplifying assumptions. More recently, a new paradigm has arisen in the field of medical image generation, exploiting the bottom-up approach of directly learning from the data the relevant information. This is achieved with machine learning systems able to automatically learn the inner variability on a large training dataset [18]. Once trained, the same system can be sampled to output a new but plausible image (*image synthesis*).

In the general computer vision field, the synthesis of natural images has recently experimented a dramatic progress, based on the general idea of adversarial learning [5]. In this context, a generator component synthesizes images from random noise, and an auxiliary discriminator system trained on real data is assigned the task of discerning whether the generated data is real or not. In the training process, the generator is expected to learn to produce images that pose an increasingly more difficult classification problem for the discriminator.

Although adversarial techniques have achieved a great success in natural image generation, medical imaging applications are still incipient. This is partially due to the lack of large amounts of training data, and partially to the difficulty of finely controlling the output of the adversarial generator. In this work, we propose to apply the adversarial learning framework to retinal images. Notably, instead of generating images from scratch, we propose to generate new plausible images from binary retinal vessel trees. Therefore, the task of the generator remains achievable, as it only needs to learn how to generate part of the retinal content, such as the optical disk, or the background's texture (Fig. 1).

The remaining of this work is organized as follows: we first describe a recent generative adversarial framework [7] that can be employed on pairs of vessel trees and retinal images to learn how to map the former to the latter. Then, we briefly review U-Net, a Deep Convolutional Neural Network designed for image segmentation, which allows us to generate pairs of retinal images and corresponding binary vessel trees. This model provides us with a dataset of vessel trees and corresponding retinal images that we then use to train an adversarial model, producing new good-quality retinal images out of a new vessel tree. Finally, the quality of the generated images is evaluated qualitatively and quantitatively, and a description of potential future research directions is presented.

2 Adversarial Retinal Image Synthesis

2.1 Adversarial Translation from Vessel Trees to Retinal Images

Image-to-image translation is a relatively recent computer vision task in which the goal is to learn a mapping G, called *Generator*, from an image x into another representation y [7]. Once the model has been trained, it is able to predict the most likely representation $G(x_{new})$ for a previously unseen image x_{new}.

However, for many problems a single input image can correspond to many different correct representations. If we consider the mapping G between a retinal vessel tree v and a corresponding retinal fundus image r, variations in color or illumination may produce many acceptable retinal images that correspond to the same vessel tree, i.e. $G(v) = \{r_1, r_2, \ldots, r_n\}$. Directly related to this is the choice of the objective function to be minimized while learning G, which turns out to be critical. Training a model to naively minimize the $L2$ distance between $G(v_i)$ and r_i for a collection of training pairs given by $\{(r_1, v_1), \ldots, (r_n, v_n)\}$ is known to produce low-quality results with lack of detail [12], due to the model selecting an average of many equally valid representations.

Instead of explicitly defining a particular loss function for each task, it is possible to employ Generative Adversarial Networks to implicitly build a more appropriate loss [7]. In this case, the learning process attempts to maximize the misclassification error of a neural network (called *Discriminator*, D) that is trained jointly with G, but with the goal of discriminating between real and generated images. This way, not only G but also the loss are progressively learned from examples, and adapt to each other: while G tries to generate increasingly more plausible representations $G(v_i)$ that can deceive D, D becomes better at its task, thereby improving the ability of G to generate high-quality samples. Specifically, the adversarial loss is defined by:

$$\mathcal{L}_{adv}(G, D) = \mathbb{E}_{v,r\sim p_{data}(v,r)}[logD(v, r)] + \mathbb{E}_{v\sim p_{data}(v)}[log(1 - D(v, G(v)))], \quad (1)$$

where $\mathbb{E}_{v,r\sim p_{data}}$ represents the expectation of the log-likelihood of the pair (v, r) being sampled from the underlying probability distribution of real pairs $p_{data}(v, r)$, while $p_{data}(v)$ corresponds to the distribution of real vessel trees. An overview of this process is shown in Fig. 2.

To generate realistic retinal images from binary vessel trees, we follow recent ideas from [7,15], which propose to combine the adversarial loss with a global $L1$ loss to produce sharper results. Thus, the loss function to minimize becomes:

$$\mathcal{L}(G, D) = \mathcal{L}_{adv}(G, D) + \lambda \mathbb{E}_{v,r\sim p_{data}(v,r)} \left(||r - G(v)||_1 \right), \quad (2)$$

where λ balances the contribution of the two losses. The goal of the learning process is thus to find an equilibrium of this expression. The discriminator D attempts to maximize Eq. (2) by classifying each $N \times N$ patch of a retinal image, deciding if it comes from a real or synthetic image, while the generator aims at minimizing it. The $L1$ loss controls low-frequency information in images generated by G in order to produce globally consistent results, while the adversarial loss promotes sharp results. Once G is trained, it is able to produce a realistic retinal image r from a new binary vessel tree v.

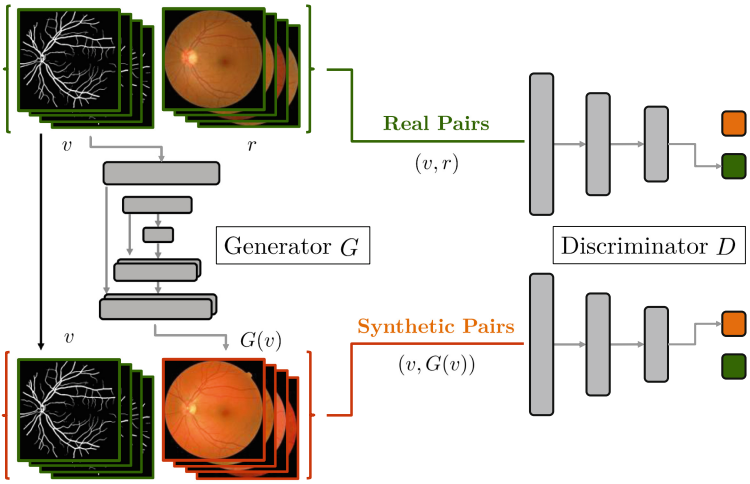

Fig. 2. Overview of the generative model mapping vessel trees to retinal images.

2.2 Obtaining Training Data

The model described above requires training data in the form of pairs of binary retinal vessel trees and corresponding retinal images. Since such a large scale manually annotated database is not available, we apply a state-of-the-art retinal vessel segmentation algorithm to obtain enough data for the model to learn the mapping from vessel trees to retinal images. There exist a large number of methods capable of providing reliable retinal vessel segmentations. Here we employ a supervised method based on Convolutional Neural Networks (CNNs), namely the U-Net architecture, first proposed in [13] for the segmentation of biomedical images. This technique is an extension of the idea of Fully-Convolutional Network (FCNs), introduced in [14], adapted to be trained with a low number of images and produce more precise segmentations.

The architecture of the U-Net consists of a contracting and an expanding part. The first half of the network follows a typical CNN architecture, with stacked convolutional layers of stride two and Rectified Linear Unit (ReLU) activations. The second part of the architecture is an expanding path, symmetric to the contracting path. The output feature map of the last layer of the contracting path is upsampled so that it has the same dimension of the second last layer. The result is concatenated with the feature map of the corresponding layer in the contracting path, and this new feature map undergoes convolution and activation. This is repeated until the expanding path layers reach the same dimensions as the first layer of the network.

The final layer is a convolution followed by a sigmoid activation in order to map each feature vector into vessel/non-vessel classes. The concatenation operation allows for very precise spatial localization, while preserving the coarse-level features learned during the contracting path. A representation of this architecture as used in the present work is represented in Fig. 3.

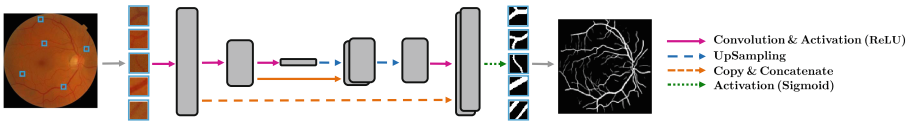

Fig. 3. Overview of the U-Net architecture. Each box corresponds to a multi-channel feature map.

2.3 Implementation

For the purpose of retinal vessel segmentation, the DRIVE database [16] was used to train the method described in the previous Section. Images and ground truth annotations were divided into overlapping patches of 64×64 pixels and fed randomly to the U-Net, with 10% of the patches used for validation. The network was trained with the Adam optimizer [8] and a binary crossentropy loss function.

Retinal vessel segmentation using the U-Net was evaluated on DRIVE's test set, achieving a 0.9755 AUC, aligned with state-of-the-art results [10]. The optimal binarization threshold maximizing the Youden index [19] was selected. Messidor [3] images were cropped, in order to only display the field of view, and downscaled to 512×512. Then, the segmentation method was applied to these images. Messidor contains 1200 images annotated with the corresponding diabetic retinopathy grade, and displays more color and texture variability than DRIVE's 20 training images. Due to the U-Net being trained and tested in different databases, some of the produced segmentations were not entirely correct. This may be related to DRIVE only containing 7 examples of images with signs of mild diabetic retinopathy (grade 1). For this reason, we retained only pairs of images and vessel trees in which the corresponding image had grade 0, 1, and 2.

The final dataset collected for training our adversarial model consisted of 946 Messidor image pairs. This dataset was further randomly divided into training (614 pairs), validation (155 pairs) and test (177 pairs) sets. Regarding image resolution, the original model in [7] used pairs of 256×256 images, with a U-Net-like generator G. We modified the architecture to handle 512×512 pairs, which is closer to the resolution of DRIVE images. For that, we added one layer to the contracting part and another to the expanding part of G. The discriminator D classifies 16×16 overlapping patches of size 63×63. The implementation was developed in Python using Keras[1] [1]. The learning process starts by training D with real (v, r) and generated pairs $(v, G(v))$. Then, G is trained with real (v, r) pairs. This process was repeated iteratively until the losses of D and G stabilized.

3 Experimental Evaluation

For subjective evaluation of the images generated by our model, we show in Fig. 4 some visual results. The first row depicts a random sample of vessel trees extracted from the held-out test set, which was not used during training. The

[1] Code to reproduce our results is available at https://github.com/costapt/vess2ret.

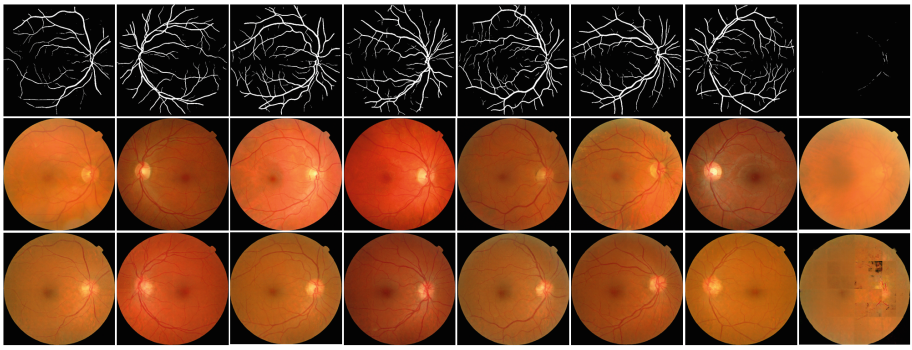

Fig. 4. Results of our model. First row: Vessel trees not used during training. Second row: True retinal images corresponding to the above vessel trees. Third row: Corresponding retinal images generated by our model. All images have 512×512 resolution.

second row shows the real images from which those vessel trees were segmented with the method outlined in Sect. 2.2, and the bottom row shows the synthetic retinal images produced by the proposed technique. We see that the original and the generated images share some global geometric characteristics. This is natural, since they approximately share the same vascular structure. However, the synthetic images have markedly different high-level visual features, such as the color and tone of the image, or the illumination. This information was extracted by our model from the training set, and effectively applied to the input vessel trees in order to produce realistic retinal images.

The last column in Fig. 4 shows a failure case of the proposed technique. Therein, the segmentation technique described in Sect. 2.2 failed to produce a meaningful vessel network out of the original image. This is probably due to the high degree of defocus that the input image had. In this situation, the binary vessel tree supplied to the generator contained too few information, and it reacted by creating spurious artifacts and chromatic noise in the synthetic image. Fortunately, the amount of cases in which this happened was relatively low: from our test set of 177 images, 7 were found to suffer from artifacts.

Regarding objective image quality verification, this is a hard challenge when no reference is available. In addition, for generative models it has been recently observed that specialized evaluation should be performed for each problem [17]. In our case, to achieve a meaningful objective quantitative evaluation of the quality of the generated images, we apply the no-reference retinal image quality assessment technique proposed in [9]. This score, denoted Q_v, is derived by calculating a local degree of vesselness around each pixel, computing a local estimate of anisotropy on regions that are good candidates for containing vessels, and averaging the results, see [9] for the technical details. The results of computing the Q_v metric on both sets of real and synthetic images are shown in Table 1.

The first two columns on Table 1 show the mean and standard deviation of the Q_v scores computed from the original and synthetic images. We can see that

Table 1. Result of computing the Q_v quality measure on real/synthetic images.

	Mean Q_v score	Std. dev	Avg. *per-image* variation
Real images	0.1234	0.0207	100%
Synthetic images	0.1040	0.0131	87.55%

the mean Q_v score obtained for the synthetic images was relatively close to the score computed from the dataset of true images. Furthermore, since from each vessel tree we have the corresponding true and synthetic images available, we can perform a *per-image* analysis of the results of the computation of the Q_v measure. For that, we considered the quality of the true retinal fundus images to be 100%, and for each synthetic image we computed the percentage of quality variation observed. Results of this analysis are shown in the third column, where we see that, on average, 87.55% of the true images quality was preserved. A more detailed analysis revealed that, from the 177 test binary vessel trees, the corresponding synthetically generated images achieved a better Q_v scores than the true images in 30 cases.

4 Conclusions and Future Work

The above results demonstrate the feasibility of learning to synthesize new retinal images from a dataset of pairs of retinal vessel trees and corresponding retinal images, applying current generative adversarial models. In addition, the dimension of the produced images was 512×512, which is greater than commonly generated images on general computer vision problems. We believe that achieving this resolution was only possible due to the constrained class of images in which the method was applied: contrarily to generic natural images, retinal images show a repetitive geometry, where high-level structures such as the field of view, the optical disc, or the macula, are usually present in the image, and act as a guide for the model to learn how to produce new texture and background intensities.

The main limitation of the presented method is its dependence on a pre-existing vessel tree in order to generate a new image. Furthermore, if the vessel tree comes from the application of a segmentation technique to the original image, the potential weaknesses of the segmentation algorithm will be inherited by the synthesized image. We are currently working on overcoming these challenges.

Acknowledgments. This work is funded by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme, by the FCT - Fundação para a Ciência e a Tecnologia within project CMUP-ERI/TIC/0028/2014 and by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement within the project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016".

References

1. Chollet, F.: Keras (2015). https://github.com/fchollet/keras
2. 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 **17**(3), 463–468 (1998)
3. Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., Charton, B., Klein, J.C.: Feedback on a publicly distributed database: the Messidor database. Image Anal. Stereol. **33**(3), 231–234 (2014)
4. Fiorini, S., Ballerini, L., Trucco, E., Ruggeri, A.: Automatic generation of synthetic retinal fundus images. In: Reyes-Aldasoro, C.C., Slabaugh, G. (eds.) Medical Image Understanding and Analysis 2014, pp. 7–12. BMVA Press (2014)
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
6. Hodneland, E., Hanson, E., Munthe-Kaas, A.Z., Lundervold, A., Nordbotten, J.M.: Physical models for simulation and reconstruction of human tissue deformation fields in dynamic MRI. IEEE Trans. Bio-Med. Eng. **63**(10), 2200–2210 (2016)
7. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks, November 2016. arXiv.org, arXiv: 1611.07004
8. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations, pp. 1–13 (2014)
9. Köhler, T., Budai, A., Kraus, M.F., Odstrcilik, J., Michelson, G., Hornegger, J.: Automatic no-reference quality assessment for retinal fundus images using vessel segmentation. In: Proceedings of the 26th IEEE CMBS, pp. 95–100, June 2013
10. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. IEEE Trans. Med. Imaging **35**(11), 2369–2380 (2016)
11. Liu, X., Liu, H., Hao, A., Zhao, Q.: Simulation of blood vessels for surgery simulators. In: 2010 International Conference on Machine Vision and Human-Machine Interface, pp. 377–380, April 2010
12. Lotter, W., Kreiman, G., Cox, D.: Unsupervised learning of visual structure using predictive generative networks. arxiv preprint (2015). arXiv:1511.06380
13. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_28
14. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE CVPR, pp. 3431–3440 (2015)
15. Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., Webb, R.: Learning from simulated and unsupervised images through adversarial training. arxiv preprint (2016). arXiv:1612.07828
16. Staal, J.J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B.: Ridge based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging **23**(4), 501–509 (2004)
17. Theis, L., Oord, A.v.d., Bethge, M.: A note on the evaluation of generative models. In: International Conference on Learning Representations (2016)
18. Tulder, G., Bruijne, M.: Why does synthesized data improve multi-sequence classification? In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 531–538. Springer, Cham (2015). doi:10.1007/978-3-319-24553-9_65
19. Youden, W.J.: Index for rating diagnostic tests. Cancer **3**(1), 32–35 (1950)