2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871771

Fundus GAN - GAN-based Fundus Image Synthesis for Training Retinal Image Classifiers

[1] Dereje Shenkut and Vijayakumar Bhagavatula[1]

Abstract— **Two major challenges in applying deep learning to develop a computer-aided diagnosis of fundus images are the lack of enough labeled data and legal issues with patient privacy. Various efforts are being made to increase the amount of data either by augmenting training images or by synthesizing realistic-looking fundus images. However, augmentation is limited by the amount of available data and it does not address the patient privacy concern. In this paper, we propose a Generative Adversarial Network-based (GAN-based) fundus image synthesis method (Fundus GAN) that generates synthetic training images to solve the above problems. Fundus GAN is an improved way of generating retinal images by following a two-step generation process which involves first training a segmentation network to extract the vessel tree followed by vessel tree to fundus image-to-image translation using unsupervised generative attention networks. Our results show that the proposed Fundus GAN outperforms state of the art methods in different evaluation metrics. Our results also validate that generated retinal images can be used to train retinal image classifiers for eye diseases diagnosis.**

Clinical relevance— **Our proposed method Fundus GAN helps in solving the shortage of patient privacy-preserving training data in developing algorithms for automating image-based eye disease diagnosis. The proposed two-step GAN-based image synthesis can be used to improve the classification accuracy of retinal image classifiers without compromising the privacy of the patient.**

I. INTRODUCTION

The main motivation of this research is to contribute to solving the problem of the shortage of ophthalmologists in Africa by implementing a robust image classifier for the automatic diagnosis of eye diseases that can reduce the burden on the ophthalmologists by providing initial screening. In Africa, it is known that there is a very low medical doctor to population ratio. When it comes to eye doctors, this ratio is even lower with 2.24 ophthalmologists per million people in sub-Saharan Africa (SSA) [12]. This problem can be mitigated by leveraging the latest machine learning (ML) advancements to automate initial eye screening and reduce the need for medical experts. However, getting annotated data to train ML systems is challenging due to the high cost of labeling data and the issue of patient privacy. Our proposed method focuses on solving the shortage of training data by generating synthetic retinal images that can be used to train ML systems. Also, since the synthetic retinal images are artificially created, they do not come from any particular patients, thus allaying patient privacy concerns.

The imaging process whereby a two-dimensional representation of a three-dimensional retina is obtained by using reflected light is called fundus imaging [2]. The image obtained from this process is known as fundus image. In this paper, we use the word fundus image, retinal image and retinal fundus image synonymously. In retinal image analysis applications, the availability of high-quality fundus images is becoming increasingly critical [1] to build practical image-based eye disease classifiers. However, annotated retinal images are not available in sufficient amount. As a result, generating synthetic fundus images has drawn attention in recent years and some works have been done in the area of synthesizing realistic-looking medical image data for retinal image diagnosis.

Before the introduction of General Adversarial Networks (GANs) [11], synthesizing realistic retinal images was attempted by using a complex mathematical model of the eye anatomy [3], [15]. [4] used a pair of retinal fundus images with vessel tree segmentation to synthesize color retinal images. The vessel tree was obtained using UNet [7] and then this vessel tree was used to learn a mapping to retinal image using image-to-image translation [16]. In their follow-up work, [6] used an adversarial autoencoder to synthesize the retinal vessel network and used this vessel map to generate retinal image using GAN. Similarly, [14] trained DCGAN [13] to generate vascular structure from noise and generated the corresponding fundus image using cGAN [16]. In another work, [3] implemented a method of generating fundus images from binary segmentation masks and then [17] presented a synthetic retinal image dataset generated using a variant of a gated recurrent unit [18]. Likewise, [5] proposed MI-GAN to generate synthetic medical images and their segmented masks.

However, the retinal images generated lack enough blood micro-vessel details. The results obtained also suffer from problems including having inconsistent diameter and geometry of optical disk and macula, abnormal interruption, and lack of enough details to distinguish between veins and arteries [6]. These are critical features of the retina in a clinical context if synthetic images are going to be used in practice. The synthesized images in most of the methods were evaluated using segmentation performance [3] and traditional GAN metrics only. To address these challenges, we design Fundus GAN, a two-stage Generative Adversarial Network (GAN) for high-quality retinal image synthesis. Inspired by the effectiveness of two-stage training for gen-

[1]Dereje Shenkut is a graduate student of Electrical and Computer Engineering, Carnegie Mellon University Africa, Kigali, Rwanda dshenkut@andrew.cmu.edu

[1]Vijayakumar Bhagavatula is Professor of Electrical and Computer Engineering, Carnegie Mellon University, PA 15213, USA kumar@ece.cmu.edu

erating high-quality fundus images for segmentation tasks, we adopt the technique of two-stage training for generating fundus images for training image classifiers. Specifically, we utilize a patch-level modified version of UNet (PUNet-33) based on [7] for generating the vessel tree. The vessel tree is then converted to a fundus image using image-to-image translation. In recent image-to-image translation tasks, unsupervised generative attention networks with adaptive layer-instance normalization technique [8] have exhibited superior results in different domains including fundus and angiography generation. We adopt this technique for the challenge of vessel tree to fundus image translation. The segmentation model is compared qualitatively and quantitatively to different methods. We also validated the resulting synthesized images by using them as training data for a retinal image-based eye diseases classifier. This shows that Fundus GAN can be potentially used to synthesize retinal images to be used in increasing the number of retinal images available for training eye diseases classifiers. The main contributions of this paper are as follows:

1) Patch-level deep layered UNet (PUNet-33) segmentation network for segmenting retinal images.
2) Based on unsupervised generative attention networks with adaptive layer-instance normalization for image-to-image translation, we propose a vessel tree to fundus image translation.
3) With PUNet-33 followed by vessel tree to fundus image translation, we build a two-stage retinal image generation network for generating retinal images with different medical conditions, healthy or with different pathological cases.

The rest of the paper is organized as follows. Section II describes our materials and methods. In Section III and Section IV, our experimentation and results are discussed, while section V concludes the paper.

II. MATERIALS AND METHODS

A. Dataset

The dataset includes three groups for training and testing segmentation, image-to-image translation, and classification models. For training and validating PUNet-33, four publicly available datasets (details provided in Table I), DRIVE [19], CHASE_DB1 [22], STARE [23], and HRF [21] are preprocessed to get images that are cropped into smaller image patches with a stride. The second stage of Fundus GAN, image-to-image translation, is also trained and validated using those datasets with their vessel structure.

B. Dataset Pre-processing (Image Patches)

In order to train and validate the model with multiple images, overlapping image patches with a stride of s and image patch resolution of *patch dim* are used from CHASE-DB1 [22], DRIVE [19], STARE [23], and HRF [21] datasets. As a result, the total images used are

$$n_{patches} = n\left(\frac{width - patch\ dim}{s} + 1\right)\left(\frac{height - patch\ dim}{s} + 1\right) \quad (1)$$

TABLE I

DATASET: NUMBER OF IMAGE PATCHES FOR TRAINING, VALIDATING, AND TESTING PUNET-33. EXCEPT FOR HRF, ALL THE OTHER DATASETS' IMAGES ARE CROPPED WITH A STRIDE OF 32 AND A PATCH DIMENSION OF 128X128. DUE TO ITS VERY HIGH RESOLUTION, HRF IS CROPPED WITH A STRIDE OF 128 AND PATCH DIMENSION 512.

Dataset	Original Image Dim.	Image Patch Dim.	Stride Size	No. of Image Patches
DRIVE	584x565	128	32	8400
STARE	700×605	128	32	5400
CHASE_DB1	999×960	128	32	21168
HRF	3504x2336	512	128	16200

Where n is the number of images in the dataset and *width* and *height* correspond to the resolution of the image. For instance, for 40 images in the DRIVE dataset with a stride of 32 and patch dimension of 128, we end up having 8400 image patches. We cropped 28 CHASE_DB1, 20 STARE, and 45 HRF images in this way. Table I contains the details about all the datasets and the training image patches used.

C. Two Step Image generation

The architecture of Fundus GAN, the two-stage image generation is shown in Fig. 1. A patch level UNet based image segmentation is used to capture enough pathological case details from the micro-vessels of the fundus image. The vessel trees obtained from PUNet-33 and their corresponding fundus images are then fed to the stage of an unsupervised image-to-image translation network. In [8], adding attention module and using adaptive layer instance normalization function has shown to be critical to generate realistic images in different domains. Inspired by this, we adopted this unsupervised image-to-image translation with attention module and adaptive layer instance normalization technique for the vessel to fundus image translation as part of the second stage of Fundus GAN. The details of the architectures used in the two stages are discussed in the following sub-sections.

1) Segmentation Network: The segmentation architecture (PUNet-33) is based on the original UNet [25]. UNet is shown to be effective in biomedical image segmentation even with a low number of images. PUNet-33, similar to the original UNet is built from contracting and expanding blocks. However, PUNet-33 is patch level and has greater depth with 33 convolutional neural networks (CNN) layers. The first half processes the image patches sequentially. After passing through the downsampling path consisting of contracting blocks and the corresponding symmetrical expanding blocks in the upsampling path, each image patch is then positioned in its corresponding location to form the full label of the input image as shown in Fig.1.

2) Unsupervised Image-to-Image Translation Network: Based on the success of U-GAT-IT [8] in generating images in different domains, the second stage of Fundus GAN incorporates an unsupervised method of translating vessel

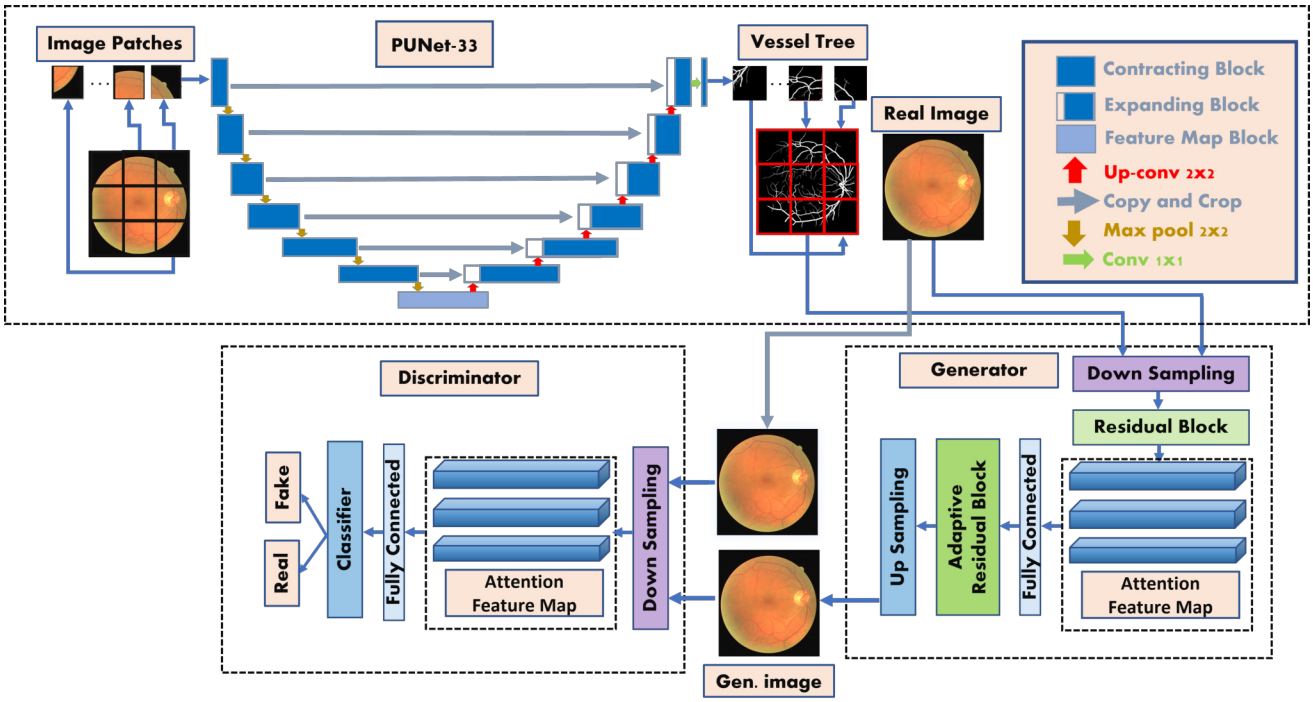

Fig. 1. Fundus GAN: the first stage is the patch level segmentation network—PUNet33 which segments the fundus image patches sequentially and produces their corresponding vessel trees. The vessel trees obtained from PUNet-33 and their fundus image pair are then fed to an unsupervised image-to-image translation network to produce the synthetic image. The synthetic image (**Gen. image**) is then used for training eye diseases classifier.

tree to fundus image with attention module and learnable layer normalization function. The generator has a CNN-based downsampling and residual blocks that compose the encoder part. After passing through the attention module the features are then decoded using adaptive residual and upsampling CNN layers. The discriminator employs two scales of PatchGAN [16] which classify image patches as real or fake. As shown in Fig 1, both the generator and discriminator include attention maps that help to focus on semantically important regions. In addition, adding an adaptive layer instance normalization function enables learning how to control the amount of change in the image features from the dataset.

D. Multi-Class Classification

For eye disease classification, Inception model-v3 [24] is used with ImageNet weights pre-loaded. This enables transfer learning that provides image classification into multiclass classification with three classes as given in HRF [21] dataset. The Inception model has two main components, one for the feature extraction and the other for the sorting. In our method, both the feature extraction and the sorting layer components are enabled as that resulted in the best performance. The final layer of the Inception model is set up to enable computing loss for each class in multi-class classification.

E. Evaluation

1) Evaluation Metrics: Our architecture is evaluated using segmentation metrics including F1-score, sensitivity, specificity, and pixel accuracy for the segmented images by PUNet-33. The performance improvement of the classification task gained by the generated images is evaluated with multi-class classification accuracy.

2) Experiments: First, different modifications of UNet are investigated to obtain the best performing segmentation model. Adding two contracting blocks and two expanding blocks (10 CNN layers) to the original UNet and using small patch images (PUNet-33) showed the best result. Hence, PUNet-33 is selected for final experimentation. We trained this architecture on four datasets, namely, HRF [21], DRIVE [19], CHASE_DB1 [22] and STARE [23]. For training and validation, we used 4200, 15120, 4320 and 7920 image patches from 20 DRIVE, 20 CHASE_DB1, 16 STARE and 22 HRF images respectively. For testing, image patches of 23, 20, 8, and 4 images of HRF, DRIVE, CHASE_DB1, and STARE test images are used. All image patches are resized to 128x128 and are converted to grayscale images. The best result for PUNet-33 is obtained with batch size of 128 and 2000 epochs with learning rate 2×10^{-4}.

The second stage of vessel tree to fundus image translation is trained on pairs of vessel tree and fundus images of the four datasets. Since the HRF dataset is the only dataset with distinct pathological cases which contains images of three different groups in healthy, diabetic retinopathy, and glaucoma conditions, it is used as a final testing set of Fundus GAN to generate the final results to be tested on classification tasks. As suggested in the work U-GAT-IT [8], this step is trained with a learning rate of 1×10^{-4} for 100 epochs. Different augmentation techniques, namely, random

horizontal flip, resize, and random crop are used.

Finally, to test the benefit gained from the use of generated images in the classification task, the classifier is trained on the HRF image dataset using the Inception-v3 model. HRF contains 45 images in which 15 are healthy, 15 are for diabetic retinopathy, and 15 for glaucoma. The dataset is split into 60%/40% train/test split per class. In the first classification experiment, a classifier is trained on all real training images from the three classes. Then, we continued our experiments by incrementally adding from 25% up to 100% of generated images into our dataset. Thus, we created a total of five dataset groups as 100% real, 75%/25% real/generated, 50%/50% real/generated, 25%/75% real/generated splits, and 100% generated images. All of the dataset groups are then tested on the real images.

III. RESULTS

A. Segmentation and image synthesis

Table II shows the performance of the segmentation model on three datasets. We compared our segmentation network to some of the best performing models in the area of retinal image segmentation including UNet [25], M-GAN [5] and RV-GAN [20]. For all the methods including ours, the results shown are the average of the metrics over all test images from each dataset as our goal mainly focuses on improving segmentation in general. As illustrated in the table, our model outperforms the other models in specificity and accuracy while RV-GAN gives the highest F1-score. Fig 2 shows example segmented image results on DRIVE dataset. PUNet-33 segments the image to greater detail, whereas RV-GAN gives a good result in terms of capturing the global structure of the vessel. For our task of image generation, we want the segmentation network to capture enough details.

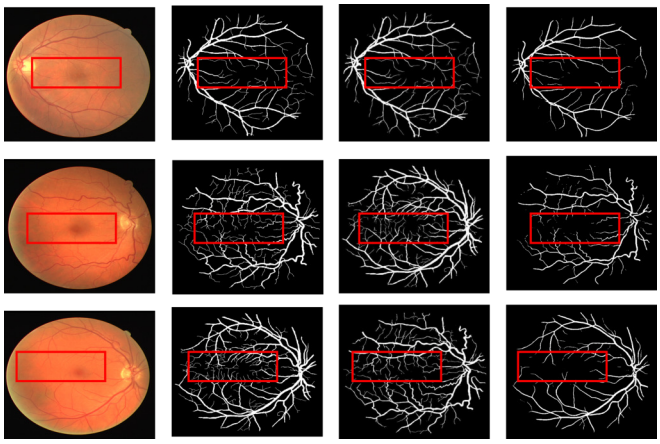

Fig. 2. Segmentation results. The first column is the input image, the second one is the ground label. The third one represents our result from PUNet-33 and the fourth column is the segmentation result from RVGAN [20] (Multi-scale GAN based segmentation)—the state of the art on DRIVE segmentation. Being trained with patch images, PUNet-33 segments micro-vessels, micro-branches and connections in detail as compared to RVGAN as shown inside the red boxes. Where as, RVGAN captures the general structure such as macro-branches in higher confidence.

TABLE II

PERFORMANCE COMPARISON BETWEEN PUNET-33, M-GAN [5] AND RVGAN [20]. THE RESULT SHOWN IS THE AVERAGE OF EACH METRICS ON DRIVE, CHASE_DB1, HRF AND STARE DATASET

Metrics	Method			
	UNet [25]	*M-GAN [5]*	PUNet-33 (Ours)	*RV-GAN [20]*
F1-Score	0.774	0.803	0.828	**0.842**
Sensitivity	0.733	**0.821**	0.785	0.791
Specificity	0.985	0.987	**0.993**	0.989
Accuracy	0.961	0.980	**0.985**	0.959

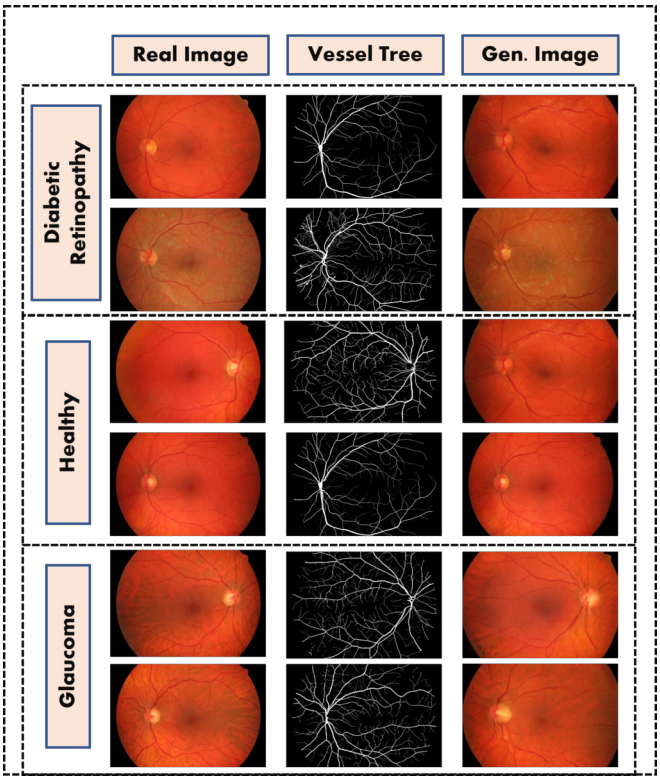

Fig. 3. Sample generated images. PUNet-33 allows feeding important pathological features which helps in translating the vessel tree to fundus image in a class specfic manner. We generated images with three classes healthy, glaucoma and diabetic retinopathy condition. For instance, the generated fundus images with glaucoma condition contains compression of blood vessels around the optic disc.

The extra details gained by using PUNet-33 allow capturing important pathological features which feed distinctive features to the image-to-image translation model. This improves the identification of different kinds of eye diseases when the generated image is used for classification. Sample generated images are shown in Fig. 3. The generated images contain features that are crucial in identifying pathological cases, namely, diabetic retinopathy, glaucoma, and healthy.

B. Disease classification

The results from our experiments on five dataset groups are presented in Table III. The table shows the best value of the validation loss (Val loss) and validation accuracy (Val acc) of each dataset group. The dataset with both synthetic

TABLE III

COMPARISON OF THE PERFORMANCE OF GENERATED IMAGES AND REAL IMAGES AS A TRAINING DATASET. VAL LOSS, VALIDATION LOSS;VAL ACC, VALIDATION ACCURACY

Dataset	Classification metrics (best values)	
	Val loss	*Val acc*
100% Real images	0.822	0.833
75% Real + 25% Gen. images	0.785	0.854
50% Real + 50% Gen. images	**0.765**	**0.872**
25% Real + 75% Gen. images	0.791	0.825
100% Gen. images	0.805	0.777

and real funds images in a 1:1 mixture outperforms all the other groups. Therefore, adding synthetic training data to real image dataset helps to improve the accuracy of retinal image classifier up to some threshold. The accuracy decreases as the synthetic data starts to dominate the dataset. However, with only generated images an accuracy of 77.7% is achieved, which shows that a classifier can be trained only on synthetic images in cases where no real labeled images are available.

IV. CONCLUSIONS

In this paper, we proposed a two-stage GAN-based retinal image synthesis method to solve the problem of the shortage of training fundus images and the concern of privacy in retinal image data. We implemented a patch-level segmentation network followed by unsupervised image-to-image translation to synthesize retinal images from three different diseases group. Our segmentation network outperforms state-of-the-art architectures in pixel accuracy and specificity. The results from our image synthesis model improved the classifier's accuracy when mixed with real image dataset to train a classifier. In addition, we showed that eye-diseases classifiers can be trained only on synthetic retinal images.

REFERENCES

[1] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abr'amoff, A. M. Mendonc a, and A. Campilho, "End-to-end adversarial retinal image synthesis," IEEE Transactions on Medical Imaging, vol. 37, no. 3, pp. 781–791, 2018.

[2] Lim, G., Bellemo, V., Xie, Y. et al. Different fundus imaging modalities and technical factors in AI screening for diabetic retinopathy: a review. Eye and Vis 7, 21 (2020). https://doi.org/10.1186/s40662-020-00182-7

[3] V. Bellemo, P. Burlina, L. Yong, T. Y. Wong, and D. S. W. Ting, "Gen- erative adversarial networks (gans) for retinal fundus image synthesis," in Computer Vision – ACCV 2018 Workshops, G. Carneiro and S. You, Eds. Cham: Springer International Publishing, 2019, pp. 289–302.

[4] P. Costa, A. Galdran, M. I. Meyer, M. D. Abr'amoff, M. Niemeijer, A. M. Mendonc a, and A. Campilho, "Towards adversarial retinal image synthesis," 2017.

[5] T. Iqbal and H. Ali, "Generative adversarial network for medical images (mi-gan)," Journal of medical systems, vol. 42, no. 11, pp. 1–11, 2018.

[6] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abr'amoff, A. M. Mendonc a, and A. Campilho, "End-to-end adversarial retinal image synthesis," IEEE transactions on medical imaging, vol. 37, no. 3, pp. 781–791, 2017.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440.

[8] J. Kim, M. Kim, H. Kang, and K. Lee, "U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation," 2020.

[9] Y. Niu, L. Gu, Y. Zhao, and F. Lu, "Explainable diabetic retinopathy detection and retinal image generation," 2021.

[10] Y.-C. Liu, H.-H. Yang, C.-H. H. Yang, J.-H. Huang, M. Tian, H. Morikawa, Y.-C. J. Tsai, and J. Tegner, "Synthesizing new retinal symptom images by multiple generative models," 2019.

[11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[12] I. A. for the Prevention of Blindness (IAPB), "Addressing the eye health workforce crisis in sub-saharan africa: Business as usual is not an option*," 2014.

[13] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2016.

[14] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," 2018.

[15] E. Menti, L. Bonaldi, L. Ballerini, A. Ruggeri, and E. Trucco, "Auto- matic generation of synthetic retinal fundus images: Vascular network," in Simulation and Synthesis in Medical Imaging, S. A. Tsaftaris, A. Gooya, A. F. Frangi, and J. L. Prince, Eds. Cham: Springer International Publishing, 2016, pp. 167–176.

[16] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2018.

[17] H. Zhao, H. Li, S. Maurer-Stroh, Y. Guo, Q. Deng, and L. Cheng, "Supervised segmentation of un-annotated retinal fundus images by synthesis," IEEE transactions on medical imaging, vol. 38, no. 1, pp. 46–56, 2018.

[18] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014.

[19] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," in IEEE Transactions on Medical Imaging, vol. 23, no. 4, pp. 501-509, April 2004, doi: 10.1109/TMI.2004.825627.

[20] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, K. M. Sanders, and S. A. Baker, "Rv-gan: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network," 2021.

[21] B. Attila, B. Rudiger, M. Andreas, H. Joachim, and M. Georg, "Robust vessel segmentation in fundus images," 2015.

[22] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based ap- proach applied to retinal blood vessel segmentation," IEEE Transactions on Biomedical Engineering, vol. 59, no. 9, pp. 2538–2548, 2012.

[23] A. D. Hoover, V. Kouznetsova and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," in IEEE Transactions on Medical Imaging, vol. 19, no. 3, pp. 203-210, March 2000, doi: 10.1109/42.845178.

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015