Retinal Image Augmentation using Composed GANs

Manal Alghamdi
Department of Computer Science and Artificial Intelligence, University of Umm Al-Qura, Makkah, Saudi Arabia
maalghamdi@uqu.edu.sa (corresponding author)

Mohamed Abdel-Mottaleb
Department of Computer Science, Indiana University, Indiana, USA
mabdem@iu.edu

Received: 10 September 2024 | Revised: 1 October 2024 | Accepted: 12 October 2024

 DOI: https://doi.org/10.48084/etasr.8964

ABSTRACT

Medical image analysis faces a significant challenge in the scarcity of annotated data, which is crucial for developing generalizable Deep Learning (DL) models that require extensive training data. Consequently, the field of medical image generation has garnered substantial interest and potential for further exploration. Besides widely employed data augmentation techniques, such as rotation, reflection, and scaling, Generative Adversarial Networks (GANs) have demonstrated the ability to effectively leverage additional information from datasets by generating synthetic samples from real images. In the context of retinal image synthesis, an image-to-image translation approach is frequently adopted to generate retinal images from available vessel maps, which can be scarce and resource-intensive to obtain. Deviating from prior work reliant on pre-existing vessel maps, this study proposes a learning-based model that is independent of vessel maps, utilizing Progressive Growing GAN (PGGAN) to generate vascular networks from random noise. The visual and quantitative evaluations conducted suggest that the majority of the images generated by the proposed model are substantially distinct from the training set while maintaining a high proportion of true image quality, underscoring the model's potential as a powerful tool for data augmentation.

Keywords-retinal image synthesis; generative adversarial networks; image-to-image translation; medical image segmentation

I. INTRODUCTION

Retinal blood vessel image segmentation is extensively employed in medical and biological domains, underpinning applications, such as computer-assisted surgery, diabetic retinopathy diagnosis, foveal avascular region detection, vessel diameter measurement for hypertension assessment, fovea localization, and optic disk identification [1-4]. Typically, human experts provide annotations serving as the ground truth to validate the efficacy of retinal image analysis algorithms. However, manual segmentation of retinal images is a laborious and arduous task. Accordingly, computer scientists and physicians concur that automated segmentation is crucial for viable computer-assisted diagnostic systems [5]. Existing research demonstrates that vascular anatomy adheres to distinct patterns, leading to the development of early medical image segmentation techniques [5]. These include model-based methods, tracking-based strategies, matched filtering, and pattern recognition approaches. Subsequently, researchers have achieved improved outcomes through the utilization of advanced Machine Learning (ML) algorithms capable of automated feature extraction [6, 7].

In the past decade, Convolutional Neural Networks (CNNs) have gained widespread popularity within the field of computer vision and have been extensively adopted for the purpose of semantic segmentation tasks [8, 9]. Compared to traditional segmentation algorithms, CNNs have the capability to automatically extract features from a comprehensive training set, effectively capturing the high-level semantic information of the target image. This approach has demonstrated improved stability and generalizability across various applications. Consequently, the availability of accurately annotated data has become increasingly crucial, as conventional deep neural network training necessitates large and correctly annotated datasets. However, in the domain of medical image analysis, annotated images are often scarce and costly. As a result, the generation of synthetic images with a high degree of realism has garnered significant interest, and researchers are actively exploring this challenge [10].

The advent of GANs has enabled researchers to leverage representation learning-based techniques for medical image synthesis [11]. Specifically, researchers have extended adversarial networks to the domain of image-to-image

translation, which could be potentially utilized to generate retinal images from existing vessel maps [12-14]. However, a significant limitation of prior work is the reliance on available vessel networks, which may not be feasible for datasets lacking human-annotated vessel information. To address this challenge, a group of researchers proposed an adversarial autoencoder approach to generate vessel maps directly from random noise, enabling end-to-end synthesis of synthetic retinal images [15]. Nonetheless, their model exhibited a decline in performance when trained on a combination of real and synthetic data, suggesting that the generated images may be of suboptimal quality.

To deal with the challenge of data scarcity in the medical domain, GANs can be used to derive additional information from the available training data by generating realistic outputs [16]. Unlike traditional data augmentation techniques, such as flipping, shifting, zooming, shearing, and scaling, GANs learn the underlying distribution from discrete training samples, and subsequently augment the training data with all potential variations after sufficient training. The adoption of GAN-based data augmentation has been found to enhance the model's generalizability and performance, and has been successfully employed in the generation of medical images, including CT and MRI scans [17-19]. Within the GAN framework, the generator (G) aims to fool the discriminator (D) by producing realistic images that are indistinguishable from the real ones, while the D is trained to identify the G-generated images as accurately as possible. This training process essentially represents a min-max game between the G and the D.

$$\min max\, V(D,G) = E_{x\sim P_{data}(x)}[logD(x)] + E_{z\sim P_z(z)}[\log\left(1 - D\big(G(z)\big)\right)] \quad (1)$$

However, the outputs of standard GANs often suffer from limited realism and mode collapse, with the model's training process being inherently unstable. Over time, researchers have proposed various GAN variants to enhance the synthetic image representations and stabilize the training process. Deep Convolutional GANs (DCGANs), demonstrate improved training stability by replacing fully connected layers with fully convolutional layers [20]. A cascading strategy was employed to integrate the GAN architecture, enabling the generation of images in a coarse-to-fine manner [21]. Similarly, a training method that progressively grows the G and D by symmetrically adding new layers to the model, with each layer contributing finer details to the generated images, was deployed [22].

In addition to generating images, researchers utilize adversarial networks to address the image-to-image translation challenge, aiming to transform images from one domain to another. A conditional GAN was developed that exerts increased control over the generated data by incorporating additional information into the G and D [16]. Although standard GANs for image generation learn the mapping from the random noise z to the output image y, conditional GANs learn the mapping from the input image x and the random noise z to the target image y'. The generalizability of conditional GANs was evaluated across various tasks and their applicability to numerous image-to-image translation problems, such as super-resolution, style transfer, image inpainting, and future state prediction, was demonstrated [23-27]. In the past five years, researchers have proposed numerous improvements to conditional GANs to enhance the accuracy and flexibility of the translation process. Specifically, CycleGAN enables mapping between different image domains without relying on paired training data, StarGAN applies the same model to translate images across multiple domains, and Pix2PixHD builds upon Pix2Pix to generate high-resolution images with realistic textures [17, 23, 28, 29].

In retinal image synthesis, conditional GANs have proven to be a useful technique for generating retinal images from existing vessel maps. A model was trained on pairs of vessel maps and the corresponding retinal images, enabling the model to learn the mapping between the two data modalities [12]. Building upon this, style transfer was incorporated into the GAN architecture, allowing the model to learn from a relatively small training set and increasing the variability of the generated outputs [13]. Furthermore, these models were improved upon by employing an adversarial auto-encoder to generate vessel maps from random noise, thus removing the dependency on pre-existing vessel maps [12, 15]. However, the performance of their model decreased when the training set included a mix of synthetic and real data, indicating the low quality of the generated images.

Semantic segmentation is a pixel-wise prediction task where each pixel is classified into a corresponding category. With the advancement of DL techniques, automated image segmentation has experienced substantial progress in recent years. Introducing a Fully Convolutional Network (FCN) architecture that removes the fully connected layers from traditional CNNs, enables predictions on images of arbitrary size [30]. Furthermore, a skip architecture was employed to combine low-level visual information with high-level semantic knowledge. Inspired by the FCN, the UNet architecture was developed, which has become the standard method for medical imaging segmentation [31]. To mitigate the vanishing gradient problem, the U-Net integrates feature maps across different image scales to facilitate signal propagation. Over the past two years, researchers have proposed numerous variations on the classic UNet, including Residual U-Net, Attention U-Net, R2U-Net, and IterNet, all of which improve model segmentation performance through techniques, such as residual learning, attention mechanisms, and recurrent network structures [32-35]. More recently, the concept of deformable convolution was incorporated into the U-Net architecture [36]. In Deformable U-Net, the sampling locations are adaptively trained to the scale and shape of vessels, enabling more precise segmentation.

In this work, the following contributions to fundus image vasculature segmentation are made:

- This study proposes employing a combination of two GANs to generate retinal fundus images without reliance on pre-existing vessel maps. Furthermore, it demonstrates that the utilization of synthetic data can enhance the performance of neural networks in retinal image segmentation tasks.

- Once trained, the model can generate vessel maps and the corresponding retinal images with 512×512 resolution.
- The proposed vessel-map agnostic fundus images retain a high proportion of the true images' quality, and the model is a powerful tool for data augmentation.

II. G ENERATING VESSEL MAPS AND RETINAL IMAGES

A. *Overview*

Retinal fundus images independent of pre-existing vascular networks were developed while maintaining relatively high resolution. This was achieved by introducing a model comprising two GANs: PGGAN for generating vessel maps from random noise inputs, and a Pix2PixHD for translating the generated vessel maps into retinal fundus images [22, 29]. PGGAN is renowned for its ability to generate high-resolution images, exhibiting high training stability and robustness to hyperparameter selection. In contrast, the conditional GAN requires additional information to guide the data generation process and has been utilized for translating from vessel maps to fundus images [16]. The proposed approach involves the application of the Pix2PixHD model, which employs a coarse-to-fine generator and a multi-scale discriminator architecture, making it well-suited for generating high-resolution images with finer details. The overview of the specific approach is depicted in Figure 1. A trained PGGAN can generate synthetic vessel maps by sampling from a latent space. In Pix2PixHD, the generator maps the input label to the corresponding retinal image and the discriminator learns to distinguish between synthetic and real pairs. Once trained, the model will infinitely generate new retinal images with the corresponding vessel maps.

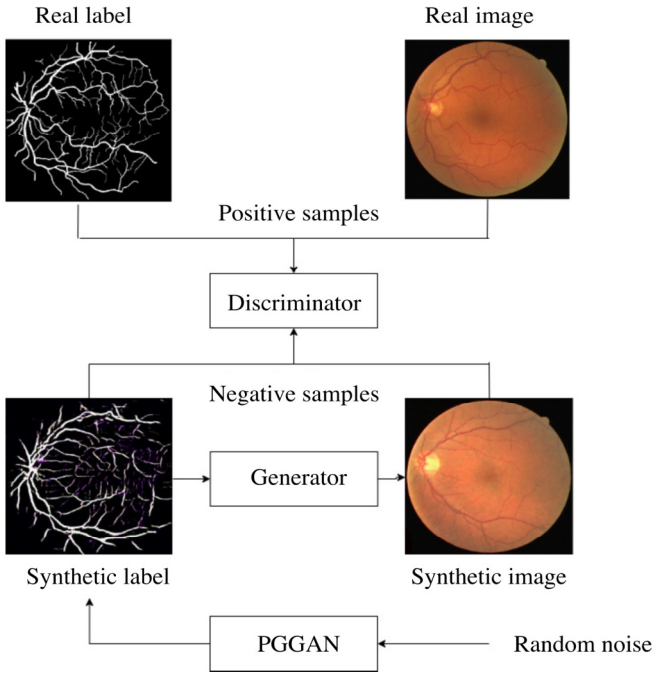

Fig. 1. Overview of the approach of the current study.

B. *Generation of Vessel Maps*

This stage seeks to generate vessel images from the latent space. PGGAN is trained in a gradual approach to provide enhanced training stability and variability [22]. The current research, initially trains PGGAN to produce realistic images at a 4×4 scale. It then iteratively adds pairs of convolutional layers and up-sampling/down-sampling layers to the generator and discriminator. The model can generate images at resolutions of 2n×2n, with the maximum resolution being 512×512. An overview of the PGGAN approach is depicted in Figure 2. First, the generator and discriminator start with a low resolution of 4×4 and grow synchronously with the training progresses. With more and more layers being added to the generator and discriminator, the model is able to generate images with higher resolution and better details.

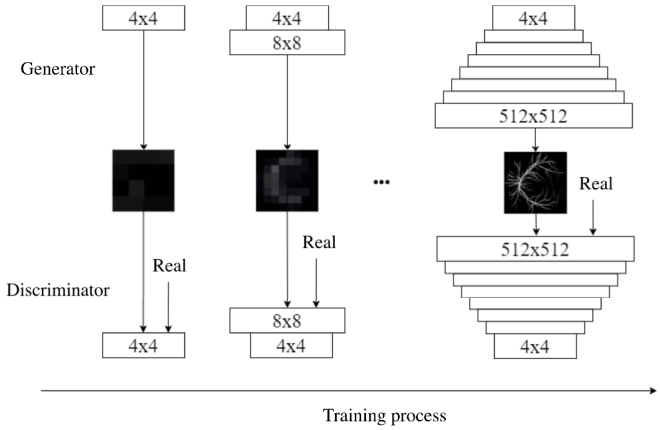

Fig. 2. The training process of PGGAN.

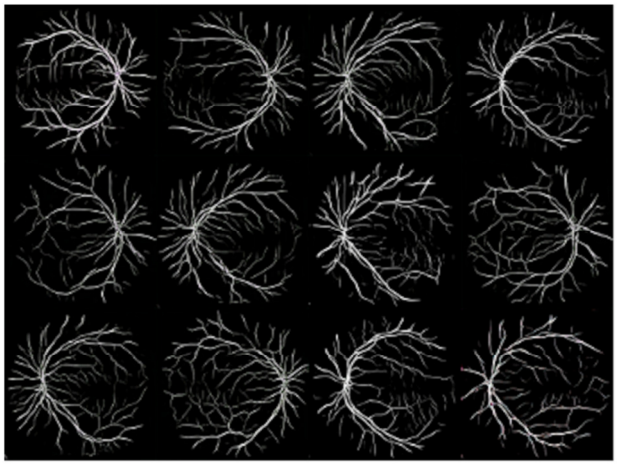

Fig. 3. Example of vessel maps generated by PGGAN. The resolution is set to 512×512, the nearest power-of-two resolution of DRIVE dataset.

Initially, the model learns the global distribution of the vascular structure. As training progresses, the generator and discriminator develop synchronously, incorporating finer details into the generated images. A specific scheme is followed to avoid abrupt transitions when new layers are added

to the *G* and *D* [22]. Additionally, minibatch standard deviation is employed to enhance variation, and the model is trained with a batch size of four. The Wasserstein loss function is utilized during training [37]. Once the training is complete, the model can generate a virtually unlimited number of vessel maps exhibiting diverse vascular structures, and some of these generated samples are presented in Figure 3.

C. Translation from Vessel Maps to Retinal Images

After generating vessel maps, they are converted into corresponding retinal images exhibiting realistic colors and structures. This objective is achieved by leveraging Pix2PixHD, an enhanced version of Pix2Pix [23, 29]. Within the supervised image-to-image translation framework, the model is trained in pairs of samples to learn the mapping from the source domain to the target domain. In this study, vessel maps are fed into both the *G* and the *D*. The *D* is tasked with distinguishing between a genuine retinal image and a synthetic one generated from the input vessel map.

The present study employs the established Pix2PixHD architecture: The *G* is partitioned into two sub-networks that operate at different resolutions, and multi-scale discriminators with varying receptive fields are utilized to consolidate global and local information [29]. Additionally, feature-matching loss is incorporated into the objective function to stabilize the training process and enhance the model's performance. The output resolution is set to 512×512 to ensure that the generated images match the input resolution. Following the training phase, the model can transform the vessel maps produced in the preceding step into realistic retinal images, featuring enhanced details and photorealism.

III. EXPERIMENTS

A. Implementation

The experiments are conducted on an NVIDIA GeForce RTX 2080 GPU and the implementations are mainly based on the Keras library using a Tensorflow backend.

B. The Data

The DRIVE dataset contains twenty 586×565 training samples and twenty 586×565 testing samples, along with annotations from human experts as ground truth [38]. The initial UNet is trained on the DRIVE dataset, which is later applied to obtain vessel segmentation from the Messidor-1 dataset [39].

The Messidor-1 dataset consists of 1200 retinal images with varying degrees of diabetic retinopathy, enabling the evaluation of the model's cross-domain robustness. The U-Net trained on the DRIVE dataset is applied to the Messidor-1 dataset to generate vessel maps from the existing retinal images. However, the DRIVE dataset contains only seven retinal images with mild diabetic retinopathy. To ensure a fair comparison, images with level three retinopathy from the Messidor-1 dataset are excluded, resulting in a sample size of 952. Furthermore, the remaining vessel maps and annotations from the DRIVE training set are utilized to train the GANs.

All the images are resized using the nearest to 512×512 neighbor interpolation before training, and the model's outputs are also constrained to 512×512 resolution.

C. Segmenting Synthetic Images

The quantitative assessment of the generated images poses a significant challenge. Generative models should be evaluated based on their intended application, while it is not advisable to extrapolate from one evaluation criterion to another [40]. In this study, the model functions as a tool for data augmentation, hence the evaluation should focus on the performance of subsequent segmentation networks. Accordingly, a U-Net is trained on three distinct training sets:

- A real dataset containing 20 images from the DRIVE's training set.
- A synthetic dataset containing 20 images generated from the proposed GANs.
- A combined dataset containing all the images in the previous two datasets.

The current study subsequently assesses the segmentation network's performance on the DRIVE test dataset.

D. Different Backbones

In addition to the standard U-Net structure, to test model generalizability, the evaluation of the performance of different U-Net backbones was completed:

- Residual U-Net is built on the U-Net architecture with two main differences: Plain neural units are replaced with residual units, and unnecessary cropping operations are removed from the network [32].
- Attention U-Net introduces the attention gate mechanism from natural language processing, which emphasizes salient information and suppresses irrelevant information [33].
- In the Bi-directional ConvLSTM U-Net with Densely connected convolutions (BCDU-Net), the authors combine the advantages of bi-directional ConvLSTM with those of U-Net [41]. The densely connected convolutions in the encoding path facilitate feature propagation and reuse.

The present study provides real training sets from the DRIVE dataset and synthetic training sets from the prior stages for the segmentation models. The quality of the synthetic images is assessed through the model's precision and Area Under Curve for the Receiver Operating Characteristic (ROCAUC) metrics.

IV. RESULTS

In this study, the generated images are assessed both qualitatively and quantitatively. For qualitative visual assessments, some of the model outputs are portrayed in Figure 4. The first row of Figure 4 presents original retinal images from the Messidor-1 dataset. The second and third rows show the obtained segmentation map through U-Net and the synthetic retinal images, respectively. The generated retinal images can maintain the vascular structure based on a given vessel map. However, failure cases in the last two columns

indicate that the translation process may not work as expected due to the limitations of the segmentation network. A retinal image could be poorly segmented due to the missing details of vessel maps or advanced diabetic retinopathy. In such cases, the model cannot produce a realistic synthetic image since the input vessel map contains insufficient or misleading information. To evaluate the model's ability to generalize, Pix2PixHD was employed to produce multiple retinal images from a single vessel map, and the outcomes are displayed in Figure 5. The generated images exhibit diverse coloration, illumination, and textural characteristics, indicating that the model does not merely memorize the training data but can create images with specific variations based on high-level features.

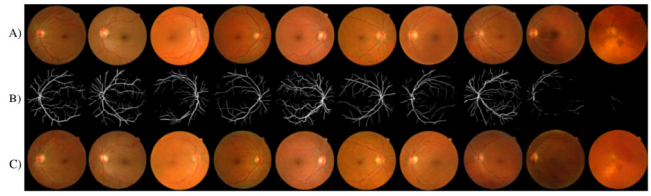

Fig. 4. Example of vessel maps generated by PGGAN. The resolution is set to 512×512, the nearest power-of-two resolution of DRIVE dataset. Comparison between true retinal images and synthetic retinal images based on the same vessel map. Rows from top to bottom are: a) True retinal images from Messidor-1, b) Obtained vessel maps after a segmentation network is applied to row A, c) Retinal images generated by the model.

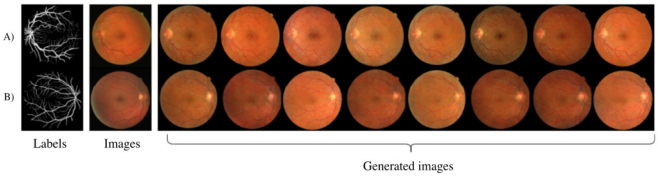

Fig. 5. Rows A and B are two examples demonstrating how Pix2PixHD generate retinal images from different vessel maps. Vessel maps and the corresponding retinal images in DRIVE's test set are shown in the first and the second columns respectively. The rest of the columns illustrate images generated by Pix2PixHD from the corresponding vessel maps.

The evaluation metric encompasses both accuracy and the ROCAUC analysis. Given that the performance of semantic segmentation is impacted by data augmentation, the findings for scenarios involving and excluding data augmentation are presented in Table I and Table II, respectively.

TABLE I. PERFORMANCE OF U-NET TRAINED ON DIFFERENT TRAINING SET WITHOUT DATA AUGMENTATION

Approach	Accuracy	ROC/AUC
Real Data (20)	0.9650	0.9705
Synthetic Data (20)	0.9370	0.8787
Real+Synthetic (20+20)	0.9652	0.9714

The findings in Tables I and II suggest that integrating conventional data augmentation techniques with the proposed GAN-based approach can further enhance the performance of the segmentation network. To assess the effectiveness of the proposed method, the segmentation capabilities of additional network architectures are evaluated, including the Residual U-Net, Attention U-Net, and BCDU-Net, as presented in Table III and Figure 6 [32, 33, 41]. These experiments were conducted using the same data augmentation strategies as in Table I.

TABLE II. PERFORMANCE OF U-NET TRAINED ON DIFFERENT TRAINING SET WITH DATA AUGMENTATION

Approach	Accuracy	ROC/AUC
Real Data (20)	0.9651	0.9670
Synthetic Data (20)	0.9353	0.9086
Real+Synthetic (20+20)	0.9652	0.9747

Figure 6 depicts Receiver Operating Characteristic (ROC) curves for models trained using synthetic images. These curves are significantly higher than the random guess baseline, suggesting that the generated samples maintain a substantial proportion of the true image quality. Notably, the two ROC curves for the BCDU-Net model are closely aligned, indicating that the model can achieve a comparable level of performance using solely synthetic data. However, across different experimental settings, a gap in performance was observed between models trained on synthetic images and those trained on real data. This discrepancy may be attributed to visual artifacts introduced by GANs, such as local inconsistencies and vessel interruptions.

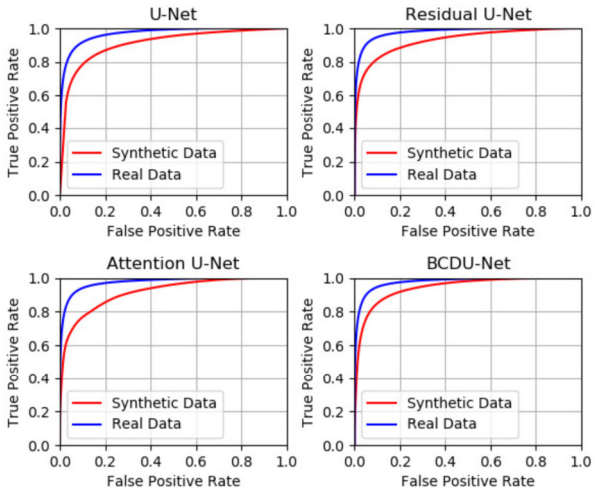

Fig. 6. ROC curves for models trained with 20 real images (blue) or 20 synthetic images (red). True Positive means that a vessel pixel is correctly classified as a vessel, and False Positive represents a background pixel being misclassified as a vessel.

The findings in Table III substantiate that combining synthetic and real-world data during training yields a consistent boost in performance, indicating the generative model's ability to produce high-quality images with subtle variations, rather than merely memorizing the training set. By augmenting the dataset with the proposed technique, improvements in segmentation across multiple algorithms are observed.

Despite the positive outcomes obtained from the proposed model, certain limitations can be identified, including:

1. The generated images sometimes suffer from artefacts like local inconsistencies and vessel interruptions, which can

negatively impact the quality of the synthetic images, and thus reduce their usefulness in segmentation tasks.

2. Although the model attempts to generate synthetic images independent of pre-existing vessel maps, the quality of the generated images still heavily depends on the input vessel maps. Poorly segmented or inaccurate vessel maps, particularly in cases of severe pathologies, such as high-level diabetic retinopathy, can lead to unrealistic synthetic retinal images.
3. The approach involves training two separate GANs, PGGAN and Pix2PixHD, which adds computational overhead.

TABLE III. PERFORMANCE OF RESIDUAL U-NET, ATTENTION U-NET AND BCDU-NET TRAINED ON DIFFERENT TRAINING SET

Residual U-Net	**Accuracy**	**ROC/AUC**
Real Data (20)	0.9661	0.9739
Synthetic Data (20)	0.9499	0.9336
Real+Synthetic (20+20)	0.9670	0.9795
Attention U-Net	**Accuracy**	**ROC/AUC**
Real Data (20)	0.9653	0.9706
Synthetic Data (20)	0.9432	0.9137
Real+Synthetic (20+20)	0.9654	0.9774
BCDU-Net	**Accuracy**	**ROC/AUC**
Real Data (20)	0.9652	0.9727
Synthetic Data (20)	0.9418	0.9414
Real+Synthetic (20+20)	0.9671	0.9767

V. CONCLUSIONS AND FUTURE WORK

This study proposes a data-driven approach to synthesize retinal images and their corresponding vessel maps. The model consists of two Generative Adversarial Networks (GANs): a Progressive Growing GA (PGGAN) for vessel map generation, and a Pix2PixHD for image-to-image translation. Once trained, the model can generate realistic retinal images without the need for pre-existing vessel maps. Researchers can simply sample from a predefined Gaussian distribution as the input to the model. This is significant because, in many real-world medical scenarios, obtaining annotated data is both time-consuming and expensive. The ability of the model to generate synthetic retinal images without the need for these annotations makes it a practical and valuable tool, particularly in settings where annotated medical images are scarce. The previous experimental results demonstrate that the model could also serve as a powerful tool for data augmentation in medical image segmentation tasks. By generating additional training data encompassing all potential variances, the model provides a more extensive training set for the segmentation network, which can ultimately improve the network's performance and generalizability for unseen cases.

The natural progression of the current research is to investigate more sophisticated segmentation models and address the segmentation limitations observed in the prior experiments. This study is presently examining whether employing alternative GAN architectures would impact model performance. The former's aim is to apply the proposed technique to additional datasets and domains in the future.

Within the context of retinal image augmentation using composed GANs, there are several promising directions for future research that could significantly enhance and build upon the existing work. These include:

- Improving GAN Architectures with more advanced GAN variants, such as StyleGAN or attention GANs [42].
- Exploring other medical imaging modalities, such as MRI, CT, or ultrasound [43].
- Combining the GANs with more sophisticated augmentation strategies [44].
- Exploring different learning techniques, such as semi-supervised or unsupervised learning approaches [45].
- Improving the Generalization of Synthetic Data [46].

Adopting these future research avenues could enhance the clinical utility of the work, leading to improved quality of synthetically generated medical images, and bolstering downstream applications, like disease identification, diagnosis, and treatment planning.

REFERENCES

[1] R. Ramesh and S. Sathiamoorthy, "A Deep Learning Grading Classification of Diabetic Retinopathy on Retinal Fundus Images with Bio-inspired Optimization," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11248–11252, Aug. 2023, https://doi.org/10.48084/etasr.6033.

[2] A. Haddouche, M. Adel, M. Rasigni, J. Conrath, and S. Bourennane, "Detection of the foveal avascular zone on retinal angiograms using Markov random fields," *Digit. Signal Process.*, vol. 20, no. 1, pp. 149–154, Jan. 2010, https://doi.org/10.1016/j.dsp.2009.06.005.

[3] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, and R. L. Kennedy, "Measurement of retinal vessel widths from fundus images based on 2-D modeling," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1196–1204, Jul. 2004, https://doi.org/10.1109/TMI.2004.830524.

[4] H. Li and O. Chutatape, "Automated feature extraction in color retinal images by a model based approach," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 2, pp. 246–254, Oct. 2004, https://doi.org/10.1109/TBME.2003.820400.

[5] M. M. Fraz *et al.*, "Blood vessel segmentation methodologies in retinal images – A survey," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 407–433, Oct. 2012, https://doi.org/10.1016/j.cmpb.2012.03.009.

[6] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised Feature Learning for Curvilinear Structure Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Berlin, Heidelberg, Germany, 2013, pp. 526–533, https://doi.org/10.1007/978-3-642-40811-3_66.

[7] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *British Journal of Ophthalmology*, vol. 83, no. 8, pp. 902–910, Aug. 1999, https://doi.org/10.1136/bjo.83.8.902.

[8] J. Son, S. J. Park, and K.-H. Jung, "Towards Accurate Segmentation of Retinal Vessels and the Optic Disc in Fundoscopic Images with Generative Adversarial Networks," *Journal of Digital Imaging*, vol. 32, no. 3, pp. 499–512, Jun. 2019, https://doi.org/10.1007/s10278-018-0126-3.

[9] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep Retinal Image Understanding," presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, Oct. 2016, vol. 9901, pp. 140–148, https://doi.org/10.1007/978-3-319-46723-8_17.

[10] C. Chen, J. H. Chuah, R. Ali, and Y. Wang, "Retinal Vessel Segmentation Using Deep Learning: A Review," *IEEE Access*, vol. 9, pp. 111985–112004, Aug. 2021, https://doi.org/10.1109/ACCESS.2021.3102176.

[11] I. Goodfellow *et al.*, "Generative Adversarial Nets," in *Proceedings of 27th Annual Conference on Neural Information Processing Systems*, Montreal, Quebec, Canada, Dec. 2014, pp. 2672–2680.

[12] P. Costa *et al.*, "Towards Adversarial Retinal Image Synthesis." Jan. 31, 2017, https://doi.org/10.48550/arXiv.1701.08974.

[13] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Medical Image Analysis*, vol. 49, pp. 14–26, Oct. 2018, https://doi.org/10.1016/j.media.2018.07.001.

[14] J. J. Jeong, A. Tariq, T. Adejumo, H. Trivedi, J. W. Gichoya, and I. Banerjee, "Systematic Review of Generative Adversarial Networks (GANs) for Medical Image Classification and Segmentation," *Journal of Digital Imaging*, vol. 35, no. 2, pp. 137–152, Apr. 2022, https://doi.org/10.1007/s10278-021-00556-w.

[15] P. Costa *et al.*, "End-to-End Adversarial Retinal Image Synthesis," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 781–791, Mar. 2018, https://doi.org/10.1109/TMI.2017.2759102.

[16] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets." arXiv, Nov. 06, 2014, https://doi.org/10.48550/arXiv.1411.1784.

[17] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, Jul. 2017, pp. 2242–2251, https://doi.org/10.1109/ICCV.2017.244.

[18] H.-C. Shin *et al.*, "Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks," in *Proceedings of Simulation and Synthesis in Medical Imaging*, Granada, Spain, Sep. 2018, pp. 1–11, https://doi.org/10.1007/978-3-030-00536-8_1.

[19] C. Bowles *et al.*, "GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks." arXiv, Oct. 25, 2018, https://doi.org/10.48550/arXiv.1810.10863.

[20] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." arXiv, Jan. 07, 2016, https://doi.org/10.48550/arXiv.1511.06434.

[21] E. L. Denton, S. Chintala, arthur szlam, and R. Fergus, "Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks," in *Proceedings of 28th Annual Conference on Neural Information Processing Systems*, Montreal, Canada, 2015.

[22] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation." arXiv, Feb. 26, 2018, https://doi.org/10.48550/arXiv.1710.10196.

[23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks." arXiv, Nov. 26, 2018, https://doi.org/10.48550/arXiv.1611.07004.

[24] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, Jul. 2017, pp. 105–114, https://doi.org/10.1109/CVPR.2017.19.

[25] C. Li and M. Wand, "Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks." arXiv, Apr. 15, 2016, https://doi.org/10.48550/arXiv.1604.04382.

[26] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, Jun. 2016, pp. 2536–2544, https://doi.org/10.1109/CVPR.2016.278.

[27] Y. Zhou and T. L. Berg, "Learning Temporal Transformations from Time-Lapse Videos," in *Computer Vision – ECCV 2016*, Amsterdam, Netherlands, Oct. 2016, pp. 262–277, https://doi.org/10.1007/978-3-319-46484-8_16.

[28] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation," in *Proccedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, USA, Jun. 2018, pp. 8789–8797, https://doi.org/10.1109/CVPR.2018.00916.

[29] 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." arXiv, Aug. 20, 2018, https://doi.org/10.48550/arXiv.1711.11585.

[30] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation." arXiv, Mar. 08, 2015, https://doi.org/10.48550/arXiv.1411.4038.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv, May 18, 2015, https://doi.org/10.48550/arXiv.1505.04597.

[32] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-Net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, Feb. 2018, https://doi.org/10.1109/LGRS.2018.2802944.

[33] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas." arXiv, May 20, 2018, https://doi.org/10.48550/arXiv.1804.03999.

[34] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation." arXiv, May 29, 2018, https://doi.org/10.48550/arXiv.1802.06955.

[35] L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki, "IterNet: Retinal Image Segmentation Utilizing Structural Redundancy in Vessel Networks," in *Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision*, Snowmass, Colorado, USA, Mar. 2020, pp. 3645–3654, https://doi.org/10.1109/WACV45572.2020.9093621.

[36] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149–162, Aug. 2019, https://doi.org/10.1016/j.knosys.2019.04.025.

[37] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN." arXiv, Dec. 06, 2017, https://doi.org/10.48550/arXiv.1701.07875.

[38] J. Xu, H. Ishikawa, G. Wollstein, and J. S. Schuman, "Retinal vessel segmentation on SLO image," in *Proceedings of 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, Canada, Dec. 2008, pp. 2258–2261, https://doi.org/10.1109/IEMBS.2008.4649646.

[39] E. Decencière *et al.*, "Feedback on a Publicly Distributed Image Database: The Messidor Database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014, https://doi.org/10.5566/ias.1155.

[40] L. Theis, A. van den Oord, and M. Bethge, "A note on the evaluation of generative models." arXiv, Apr. 24, 2016, https://doi.org/10.48550/arXiv.1511.01844.

[41] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-Directional ConvLSTM U-Net with Densley Connected Convolutions." arXiv, Aug. 31, 2019, https://doi.org/10.48550/arXiv.1909.00166.

[42] T. Karras *et al.*, "Alias-Free Generative Adversarial Networks," in *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 852–863.

[43] Y. Skandarani, P.-M. Jodoin, and A. Lalande, "GANs for Medical Image Synthesis: An Empirical Study," *Journal of Imaging*, vol. 9, no. 3, Mar. 2023, Art. no. 69, https://doi.org/10.3390/jimaging9030069.

[44] Y. Chen *et al.*, "Generative Adversarial Networks in Medical Image augmentation: A review," *Computers in Biology and Medicine*, vol. 144, Dec. 2022, Art. no. 105382, https://doi.org/10.1016/j.compbiomed.2022.105382.

[45] X. Feng, J. Lin, C.-M. Feng, and G. Lu, "GAN inversion-based semi-supervised learning for medical image segmentation," *Biomedical Signal Processing and Control*, vol. 88, no. 7, Feb. 2024, Art. no. 105536, https://doi.org/10.1016/j.bspc.2023.105536.

[46] H. Guan and M. Liu, "Domain Adaptation for Medical Image Analysis: A Survey," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1173–1185, Mar. 2022, https://doi.org/10.1109/TBME.2021.3117407.