



# INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : [www.joiv.org/index.php/joiv](http://www.joiv.org/index.php/joiv)



## Ultra-wide-field Fundus Image Synthesis Using Various GAN Models

Ara Ko <sup>a,c</sup>, Jungwon Cho <sup>b,c,\*</sup>

<sup>a</sup> Major in Computer Education, Graduate School, Jeju National University, 102 Jejudaehak-ro, Jeju 63243, Republic of Korea

<sup>b</sup> Department of Computer Education, Jeju National University, 102 Jejudaehak-ro, Jeju 63243, Republic of Korea

<sup>c</sup> Intelligent Software Education Research Institute, Jeju National University, 102 Jejudaehak-ro, Jeju 63243, Republic of Korea

Corresponding author: \*jwcho@jejunu.ac.kr

**Abstract**—Many people lose sight due to diabetic retinopathy. The reason that diabetic retinopathy is dangerous is that it cannot return to its pre-onset state after the disease's onset. Most patients take fundus images that capture the retina, and the doctor uses the fundus images to determine the presence of disease. Existing fundus images could only identify a narrow range, making it difficult to diagnose the disease accurately. However, with technological advances, ultra-wide-field fundus images that allow the wider retina to be seen have emerged. However, in deep learning research, many studies use existing fundus images due to the lack of new data. In the case of new technologies such as ultra-wide-field fundus images, it was often difficult to obtain data, so deep learning research could not be done properly. In the case of ultra-wide-field fundus images, research was conducted using data from hundreds to ten thousand sheets, but compared to large-scale data sets, the deep learning performance is inevitably inferior compared to large-scale data sets. In this study, synthetic data were created using ultra-wide-field fundus images and various GAN models to solve this problem. As a result of the study, BEGAN was derived similarly to the real image in qualitative and quantitative evaluation. However, it fell into mode collapse and showed the same output even when a new input came in. Mode collapse in BEGAN could be appeared depending on the amount and size of data, so various studies using BEGAN are needed.

**Keywords**—GAN; generative adversarial networks; ultra-wide-field fundus image; diabetic retinopathy; deep learning.

Manuscript received 22 Oct. 2021; revised 24 Jan. 2022; accepted 7 Apr. 2022. Date of publication 30 Sep. 2022.  
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



### I. INTRODUCTION

Diabetic Retinopathy is adults' number one cause of blindness [1]. One of the diabetic complications, diabetic retinopathy, cannot be completely cured after onset. Therefore, early diagnosis through regular check-ups is essential for diabetic patients, and it is the best treatment method [2]. Due to the characteristics of these diseases, various disease diagnosis studies using deep learning are being conducted. The most representative example is a diagnostic study of diabetic retinopathy [3] using fundus images. However, just as deep learning technology advances, fundus imaging is also evolving. The fundus images used in [3] can be observed in the 30-50° range. However, for accurate disease diagnosis, an ultra-wide-field imaging technique capable of observing a 200° range has been developed [4] and is being used in many medical fields. If the imaging technology used in the medical field is changed, deep learning models using existing data cannot be used. Therefore, research using new data is conducted separately.

Research using ultra-wide-field fundus images was conducted in collaboration with a hospital, and a hospital retinal specialist or ophthalmologist did data labeling. In Toshihiko et al. [5] study, untreated proliferative diabetic meningitis was detected, and 378 images of 256\*192 size were used. To solve the lack of data, brightness adjustment, gamma correction, histogram equalization, noise addition, and inversion were used, and 18-fold augmentation was used for training.

In the study of Kangrok et al. [6], diabetic retinopathy was diagnosed by segmentation of ultra-wide-field fundus images using ENTDRS 7-field photographs. About 13,000 pieces of data were used, and data from Catholic Guangdong University were used. This study used segmentation to remove eyebrows and skin in ultra-wide-field fundus images. In the study of Toshihiko et al. [7], diabetic retinopathy was diagnosed using ultra-wide-field fundus images and Optical Coherence Tomography Angiography (OCTA) images. A total of 491 images were used, and each image was used to diagnose each stage of diabetic retinopathy. In the previous studies [5], [6], [7]), in order to increase the accuracy of disease diagnosis, data augmentation, segmentation

techniques, or other data are additionally used. This problem is a technique used because the data is small or biased, and if the number of data is sufficient, the existing CNN model can be used. However, the data used in each study is not disclosed to the outside due to privacy issues.

In order to solve this problem, data synthesis research using the Generative Adversarial Network (GAN) [8] is in progress. Ju et al. [9] conducted a study to synthesize ultra-wide-field fundus images using existing fundus images, and Cycle GAN was used. For synthetic data validation, ResNetV2 was used to measure the accuracy, precision, recall, specificity, and F1-score of disease classification. Although three ophthalmologists participated in data labeling in this study, there are limitations in that the patients of fundus images and ultra-wide-field fundus images are not the same, and the pathology of each image is different.

In the case of ultra-wide-field fundus images, research is not active compared to fundus images due to the lack of open datasets. Existing fundus image synthesis studies were also conducted using image-to-image translation and blood vessel images. In the study of Costa et al. [10], the amount of data in which blood vessel images and fundus images exist in pairs was small, so U-Net was used to add data. Also, in Zhao et al. [11] and Iqbal et al., although vascular images were not used, a synthesis study using style transfer was conducted.

During the study on the synthesis of fundus images, Biswas et al. [13] synthesized fundus images using DCGAN. The results were measured using Structural-Similarity (SSIM) and made similar results to the real fundus images, as shown in Fig.1. In this study, the possibility of data synthesis using only fundus images was found. Therefore, instead of using a separate data set, images are synthesized using only ultra-wide-field fundus images. Since there is no representative GAN model, this study tries to find an appropriate model using various GAN models.

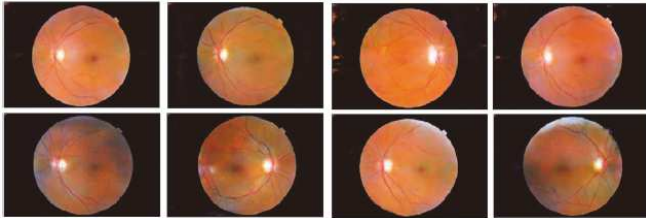


Fig. 1 Some result of Biswas paper (first row : k=5, second row : k=6)

## II. MATERIALS AND METHOD

### A. Hardware Specifications

This study utilized certain hardware specifications to support the research. All the items and hardware specifications are expected to run the software excellently.

TABLE I  
SPECIFICATIONS OF HARDWARE AND SOFTWARE ARE USED

Item	Specification
CPU	IntelCore®Processor i7-7700 3.60Ghz
SSD	ADATA SU650 (240GB)
RAM	16GB DDR4
GPU	NVIDIA Geforce GTX1060Ti
OS	Ubuntu 18.04 LTS

### B. Dataset

In this study, a total of 152 ultra-wide-field fundus images were collected through collaboration with the Department of Ophthalmology at Jeju National University Hospital. For the data, only the image was delivered after personal information was deleted, and the file name was 'grade (order number)'. Data grades are divided into five categories according to the severity of international clinical diabetic retinopathy: No-DR, Mild, Moderate, Severe, and PDR. In order to reduce data bias, each grade has 30 sheets, and only the Severe grade consists of 32 sheets.

The data was too small to be used for deep learning research, so the study was conducted by dividing five grades into two classes. The normal group was classified into No-DR, Mild grade, and the disease group was divided into Moderate, Severe, and PDR grades. The original data was 3900\*3072, which was reduced to 256\*192 for research. A sample of the data is equal to Fig. 2.

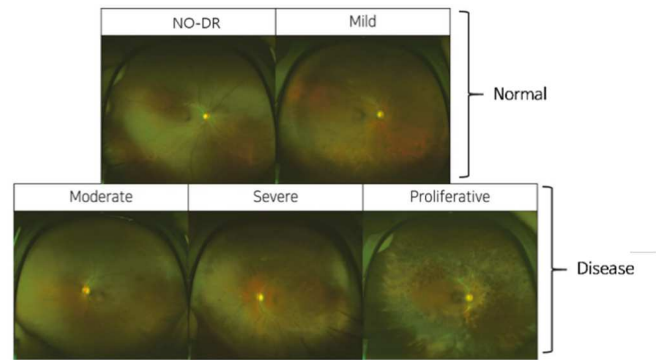


Fig. 2 Sample of Dataset

### C. DCGAN

Deep Convolutional GAN(DCGAN) [14] is a model which is a modified structure for the stabilization of learning. The structure was changed using Convolutional Neural Networks, which are often used in supervised learning. DCGAN used strided convolution in the generator, and up-sampling is possible in the discriminator. Instead of a fully connected layer, a convolutional layer was used, and batch normalization was used.

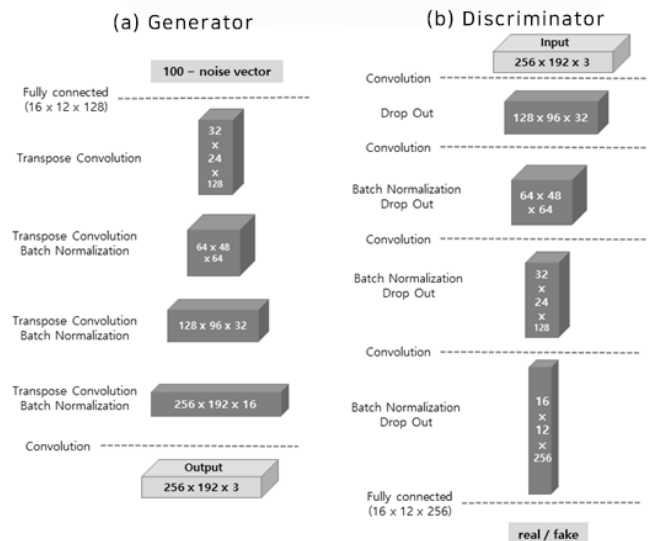


Fig. 3 Implemented DCGAN Architecture

In addition, a convolutional layer was used in the discriminator, and LeakyReLU was used as the activation function. The generator's structure was described as a deconvolutional layer, but it was a real transpose convolutional layer, and ReLU was used as the activation function. According to Yi et al. [15], among medical image synthesis studies, DCGAN was the most used among unconditional GANs. Therefore, DCGAN was also adopted in this study. The model implementation was based on the published paper, but because the data size was different, one layer was added, and the hyperparameter was maintained. The implementation model is shown in Fig. 3.

#### D. WGAN-GP

Wasserstein GAN(WGAN) [16] is a GAN model that uses the Wasserstein distance as a loss function. It points out the limitations of Total Variation (TV) distance, Kullback-Leibler (KL) divergence, and Jensen-Shannon (JS) divergence used in the existing GAN. When the real image's probability distribution and the generated image's probability distribution are parallel, if the TV distance does not match, 1 is always output. In the case of KL divergence, the value  $\infty$  is output, and in the case of JS divergence,  $\log 2$  is output. Since one value was output and learning did not proceed anymore, the Wasserstein distance was used. However, WGAN had a limitation because it was difficult to adjust the clipping parameter value. To overcome this limitation, the model produced is WGAN-GP [17], and by using gradient penalty instead of clipping parameter, the learning speed was increased while maintaining stability. The implemented WGAN-GP model is shown in Fig. 4. Up sampling and convolutional layer have been added to the generator, and other settings are the same as those implemented in the paper.

#### E. BEGAN

Boundary Equilibrium GAN(BEGAN)[18], compared to other GAN models, the structure of the model is simple, the learning speed is fast and stable learning is possible. Based on EBGAN, it uses an autoencoder structure. It uses Wasserstein distance as the loss function and is characterized by reducing the auto-encoder loss distribution of real and fake images. Also, by defining the diversity ratio, a hyperparameter was added to balance the generator and discriminator. The diversity ratio has a value between 0 and 1, and the diversity of fake images can be adjusted. When the value of the diversity ratio becomes small, the autoencoder discriminator concentrates on image restoration. Thus, the variety of images produced by the generator is reduced. Conversely, as the value increases, more focus is placed on the generator than on the discriminator. Thus, various images are created. In this study, the diversity ratio was conducted using three values of 0.3, 0.5, and 0.7.

In the decoder used in this study, up-sampling and two convolution arithmetic operations were added, and three convolution arithmetic operations were added in the encoder. The model structure can be found in Fig. 5. Also, unlike DCGAN, the Network structure was made simple, and batch normalization and Dropout were not used.

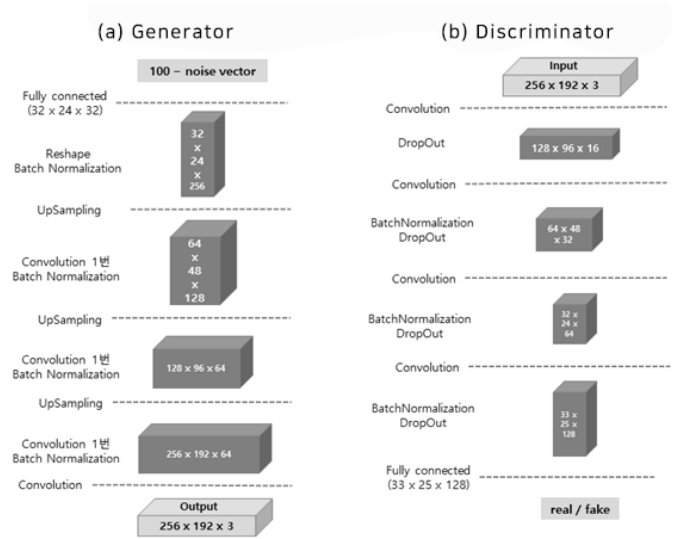


Fig. 4 Implemented WGAN-GP Architecture

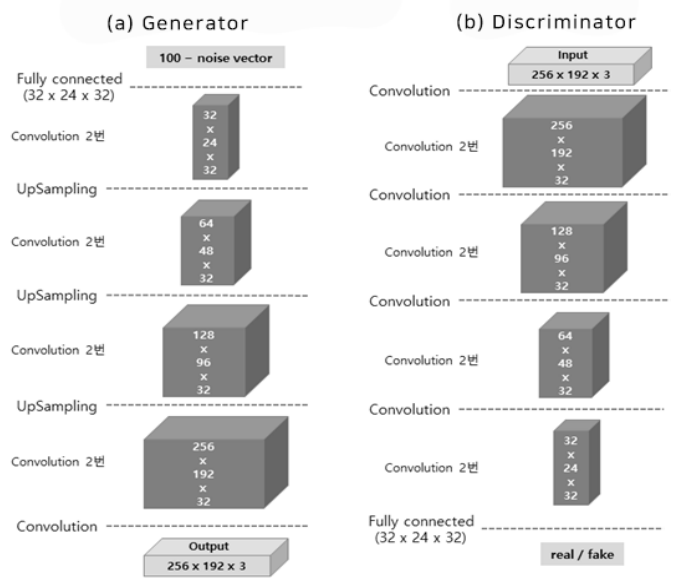


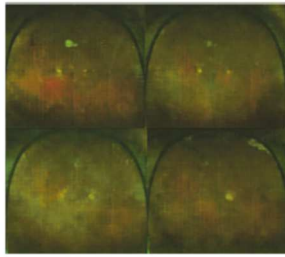
Fig. 5 Implemented BEGAN Architecture

### III. RESULT AND DISCUSSION

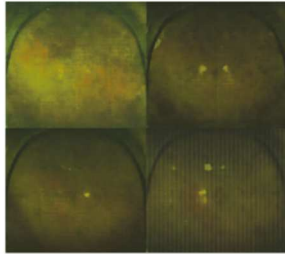
Using 152 ultra-wide-field fundus images, images were synthesized using three GAN models. DCGAN, WGAN-GP, and BEGAN models were used, and the results of this study are as follows. The evaluation of the generated image was divided into qualitative evaluation and quantitative evaluation. For quantitative evaluation, it was measured using the frequently used FID.

#### A. Synthesis Images

The image, which is the learning result, was selected based on 20,000 epochs, and presented by dividing the case using normal data and disease data. The synthesis result of DCGAN can be confirmed in Fig. 6, the result of WGAN-GP in Fig. 7, and the result of BEGAN in Fig. 8.

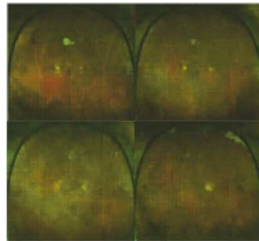


(a) Sample of synthesis image using Normal data

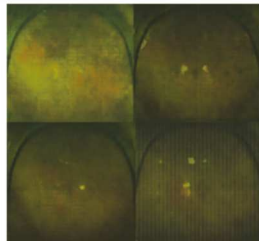


(b) Sample of synthesis image using Disease data

Fig. 6 Result image using DCGAN

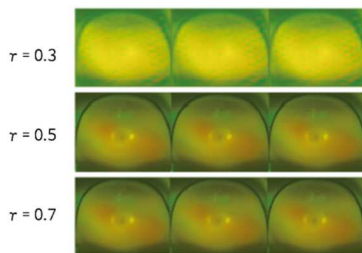


(a) Sample of synthesis image using Normal data

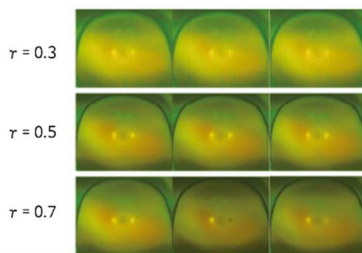


(b) Sample of synthesis image using Disease data

Fig. 7 Result image using WGAN-GP



(a) Sample of synthesis image using Normal data



(b) Sample of synthesis image using Disease data

Fig. 8 Result image using BEGAN

## B. Qualitative Evaluation

Using the resulting images, it was evaluated whether there were elements to be included in the fundus images. The overall outline, the optic disk, the macula, and blood vessels should be observed, and the pathology should be judged. Table 2 shows the results of evaluating fundus images for each model using sample data.

TABLE II  
EVALUATION OF SYNTHESIS UWF IMAGES ELEMENTS BY MODEL

Element	DCGAN	WGAN-GP	BEGAN (0.7)
Contour	○	○	○
Optic disk	△	△	△
Macula	×	△	○
Vessel	×	×	×
Distinguishing the disease	×	×	△

In the case of DCGAN, it was possible to check the outline, and the optic disk could be identified in some images. However, the macula and blood vessels could not be observed, and the data could not be distinguished according to the presence or absence of disease. In the case of WGAN-GP, the macula could be observed in some data and the optic disk. However, it was not possible to distinguish between blood vessels and the presence or absence of disease.

In the case of BEGAN, not only the outline but also the macula could be observed. Optic disks could be identified in some images, but two optic disks were observed in other images. Among the three models, when only the BEGAN model was used, the results differed when normal, and disease data were used. When the components are discriminated, it can be seen that among the three models, when BEGAN is used, the image most similar to the real one is generated.

## C. Quantitative Evaluation

The generated image evaluation scales using GAN include Inception Score (IS)[19], and Fréchet Inception Distance(FID)[20]. Both IS, and FID is methods that use Inception models pre-trained with ImageNet data. Since IS uses a generated image, it has a disadvantage in not reflecting the probability distribution of a real image. Therefore, in this study, quantitative evaluation is carried out using FID. FID measures the score using the mean and covariance values using the probability vectors obtained by putting both the real image and the generated image into the Inception model. The smaller the value, the more similar it can be evaluated to the real image. The FID score for each model can be seen in Table 3.

TABLE III  
FID SCORE BY MODEL

Model	Normal	Disease
Real-Real	160.8	189.1
DCGAN	322.1	310.7
WGAN-GP	323.3	307.2
BEGAN(0.3)	469.1	<b>251.3</b>
BEGAN(0.5)	268.7	278.3
BEGAN(0.7)	<b>237.8</b>	252.0

The Inception model using ImageNet data is used when using the FID score. However, since ImageNet data and ultra-wide-field fundus images are different, the FID of Real-Real data is compared based on the measured value. Looking at the FID value, it can be determined that the case with a small value is similar to the real data. Both the Normal and Disease data had the smallest values when the BEGAN model was used. When the Normal data was used, the diversity ratio value was 0.7, and in the case of Disease data, it was the lowest when it was 0.3. When the FID scores were compared, it was found that the BEGAN model had the smallest score among the three models, making it a suitable model for image synthesis.

#### IV. CONCLUSIONS

We have studied to generate new data using ultra-wide-field fundus images. Since there were no related existing studies, three GAN models, DCGAN / WGAN-GP / BEGAN, were selected, and the results of each study were compared. In the qualitative evaluation, the BEGAN (0.7) model was most similar to the real, and in the quantitative evaluation, the BEGAN (0.3) and BEGAN (0.7) generated data similar to the actual data according to the type of data. When synthesizing data using ultra-wide-field fundus images, BEGAN was the most suitable model among the three models. This study was conducted based on 20,000 epochs, but a similar image was generated when it exceeded 60,000 epochs. In Fig. 9, blood vessels, optic disk, and macula were all observed, and an image similar to the real one was created.

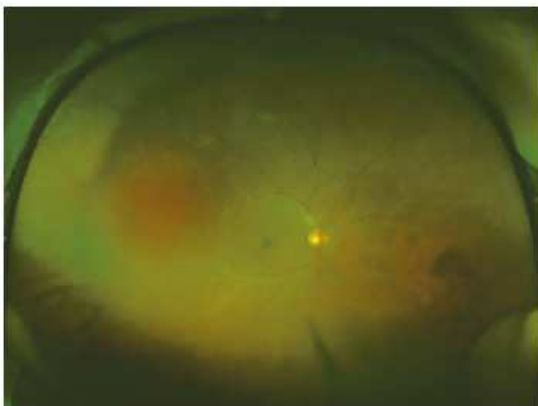


Fig. 9 Result synthesis image using BEGAN (60,000 epochs)

However, the same result was shown even if the input was changed due to mode collapse. There was no exact cause for the mode collapse, but according to the BEGAN v2[21] study, data quantity and data size can have an effect. Since the amount of data used in this study was small and the data size was larger than that of the existing GAN study, it could be inferred that mode collapse occurred. Therefore, it is necessary to study generative models suitable for medical data.

#### ACKNOWLEDGMENT

This work was supported by the 2022 education, research, and student guidance grant funded by Jeju National University. This article is excerpted from the presented at Master dissertation of the first author.

- [1] Health Chosun. (2017) 3 major blindness diseases You should not rest assured that you are young. [Online]. Available: [https://health.chosun.com/site/data/html\\_dir/2017/02/01/2017020101123.html](https://health.chosun.com/site/data/html_dir/2017/02/01/2017020101123.html). [Accessed: Aug. 15, 2022]
- [2] Kong Ophthalmology, "Special Feature: Diabetes mellitus and eye diseases -Diabetic retinopathy easily understood through cartoons," *A Monthly Diabetes Magazine*, vol.2011, no.2, pp.12-17, 2011.
- [3] V. Gulshan *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," (in eng), *Jama*, vol. 316, no. 22, pp. 2402-2410, Dec 13 2016, doi: 10.1001/jama.2016.17216.
- [4] K. Ghasemi Falavarjani, I. Tsui, and S. R. Sadda, "Ultra-wide-field imaging in diabetic retinopathy," (in eng), *Vision Res*, vol. 139, pp. 187-190, Oct 2017, doi: 10.1016/j.visres.2017.02.009.
- [5] T. Nagasawa *et al.*, "Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naïve proliferative diabetic retinopathy," (in eng), *Int Ophthalmol*, vol. 39, no. 10, pp. 2153-2159, Oct 2019, doi: 10.1007/s10792-019-01074-z.
- [6] K. Oh, H. M. Kang, D. Leem, H. Lee, K. Y. Seo, and S. Yoon, "Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images," (in eng), *Sci Rep*, vol. 11, no. 1, p. 1897, Jan 21 2021, doi: 10.1038/s41598-021-81539-3.
- [7] T. Nagasawa *et al.*, "Accuracy of Diabetic Retinopathy Staging with a Deep Convolutional Neural Network Using Ultra-Wide-Field Fundus Ophthalmoscopy and Optical Coherence Tomography Angiography," (in eng), *J Ophthalmol*, vol. 2021, p. 6651175, 2021, doi: 10.1155/2021/6651175.
- [8] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [9] L. Ju *et al.*, "Bridge the domain gap between ultra-wide-field and traditional fundus images via adversarial domain adaptation," *arXiv preprint arXiv:2003.10042*, 2020.
- [10] P. Costa *et al.*, "End-to-end adversarial retinal image synthesis," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 781-791, 2017.
- [11] 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, 2018.
- [12] 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.
- [13] S. Biswas, J. Rohdin, and M. Drahanický, "Synthetic retinal images from unconditional GANs," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019: IEEE, pp. 2736-2739.
- [14] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [15] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical image analysis*, vol. 58, p. 101552, 2019.
- [16] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, 2017: PMLR, pp. 214-223.
- [17] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.
- [18] D. Berthelot, T. Schumm, and L. Metz, "Began: Boundary equilibrium generative adversarial networks," *arXiv preprint arXiv:1703.10717*, 2017.
- [19] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, 2016.
- [20] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, vol. 30, 2017.
- [21] C.-C. Chang, C. H. Lin, C.-R. Lee, D.-C. Juan, W. Wei, and H.-T. Chen, "Escaping from collapsing modes in a constrained space," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 204-219.