OPEN ACCESS
Journal of Innovative Optical Health Sciences
Vol. 17, No. 3 (2024) 2450003 (9 pages)

DOI: 10.1142/S1793545824500032

World Scientific
www.worldscientific.com

# Unified deep learning model for predicting fundus fluorescein angiography image from fundus structure image

Yiwei Chen *,†, Yi He *,†,§, Hong Ye *, Lina Xing *,†, Xin Zhang *,† and Guohua Shi *,†,‡

*Jiangsu Key Laboratory of Medical Optics
Suzhou Institute of Biomedical Engineering and Technology
Chinese Academy of Sciences, Suzhou 215163, P. R. China

†School of Biomedical Engineering (Suzhou),
Division of Life Sciences and Medicine
University of Science and Technology of China
Hefei 230026, P. R. China

‡Center for Excellence in Brain Science and Intelligence Technology,
Chinese Academy of Sciences, Shanghai 200031, P. R. China
§heyi@sibet.ac.cn

Received 19 October 2023
Accepted 19 February 2024
Published 22 March 2024

The prediction of fundus fluorescein angiography (FFA) images from fundus structural images is a cutting-edge research topic in ophthalmological image processing. Prediction comprises estimating FFA from fundus camera imaging, single-phase FFA from scanning laser ophthalmoscopy (SLO), and three-phase FFA also from SLO. Although many deep learning models are available, a single model can only perform one or two of these prediction tasks. To accomplish three prediction tasks using a unified method, we propose a unified deep learning model for predicting FFA images from fundus structure images using a supervised generative adversarial network. The three prediction tasks are processed as follows: data preparation, network training under FFA supervision, and FFA image prediction from fundus structure images on a test set. By comparing the FFA images predicted by our model, pix2pix, and CycleGAN, we demonstrate the remarkable progress achieved by our proposal. The high performance of our model is validated in terms of the peak signal-to-noise ratio, structural similarity index, and mean squared error.

*Keywords*: Fundus fluorescein angiography image; fundus structure image; image translation; unified deep learning model; generative adversarial networks.

§Corresponding author.

## 1. Introduction

Retinal circulation imaging plays a crucial role in both ophthalmic diagnosis and the study of eye diseases.[1,2] Widely used ocular angiography modalities include fundus fluorescein angiography (FFA),[3] indocyanine green angiography,[4] and optical coherence tomography angiography.[5] Indocyanine green angiography is mainly used to image choroidal blood flow, and optical coherence tomography angiography fails to suitably show static leakage. On the other hand, FFA is the most precise modality of retinal circulation imaging. However, it requires the administration of intravenous dyes, which may have adverse effects, including symptoms such as nausea, vomiting, syncope, shock, and, in the worst case, death. Owing to their weakened physical condition, the elderly population, which comprises most of the individuals undergoing FFA, is the most susceptible to such adverse effects.

Given the related risks, the frequency of FFA examinations should be reduced. Thus, screening diagnostic methods have recently been developed for predicting FFA images from conventional fundus structure images.[6–17] The core of this algorithm involves training a deep neural network to acquire the ability to transform input fundus structure images into FFA images. During the training process, real FFA images are used to supervise the network. Once the training is completed, the network can convert fundus structure images into pseudo-FFA images without the need for intravenous dyes, thereby reducing the frequency of intravenous dyes usage. These methods involve three tasks: predicting FFA images from fundus camera images,[6–14] single-phase FFA images from scanning laser ophthalmoscopy (SLO) images,[14] and three-phase FFA images from SLO images.[15–17] However, the existing models can only perform one or two of these prediction tasks. Hence, different approaches must be developed for different tasks, rendering the design, implementation, and debugging time-consuming.

To solve this problem, we propose a unified deep-learning model that takes a fundus structure image as the input and jointly performs the three prediction tasks, avoiding the construction of different approaches for these tasks. This model is based on image-to-image translation using conditional adversarial networks (pix2pix)[18] and self-attention.[19] To demonstrate the improvements of our model, we compared its FFA image predictions with the FFA images generated by pix2pix[18] and cycleGAN.[20] Specifically, we showed typical predicted FFA images and compared the methods with regard to the peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM),[21] and mean squared error (MSE).

## 2. Materials and Methods

### 2.1. *Deep learning model*

The architecture of the proposed deep learning model for FFA image prediction is shown in Fig. 1. A generator extracts features of the input fundus structure image and synthesizes an FFA image based on the extracted features. A discriminator distinguishes real FFA images from synthetic FFA images. To improve the prediction performance, we add a self-attention module[19] to the generator. The module architecture is shown in the bottom-left block of Fig. 1. To train the deep learning model, we utilize a loss function that combines the L1 and adversarial loss functions, which are determined by a least-squares generative adversarial network.[22]

### 2.2. *Datasets preparation*

We prepared three types of datasets for predicting FFA images from fundus camera images, single-phase FFA images from SLO images, and three-phase FFA images from SLO images. Because the input and output of our deep learning model were in the red–green–blue (RGB) format, being three-channel images, we prepared our datasets to conform to this format.

For the dataset used to predict FFA images from fundus camera images, the input was the latter, and the output was a single-phase FFA image. The fundus camera image was originally in the RGB format, and the single-phase FFA image, which was originally a grayscale image, was converted to RGB format by separately copying the grayscale information to the red, green, and blue channels.

For the dataset used to predict a single-phase FFA image from an SLO image, the input was the latter, and the output was the former. Both the SLO and single-phase FFA images were originally represented in grayscale and converted to RGB format by separately copying the grayscale information to the red, green, and blue channels.

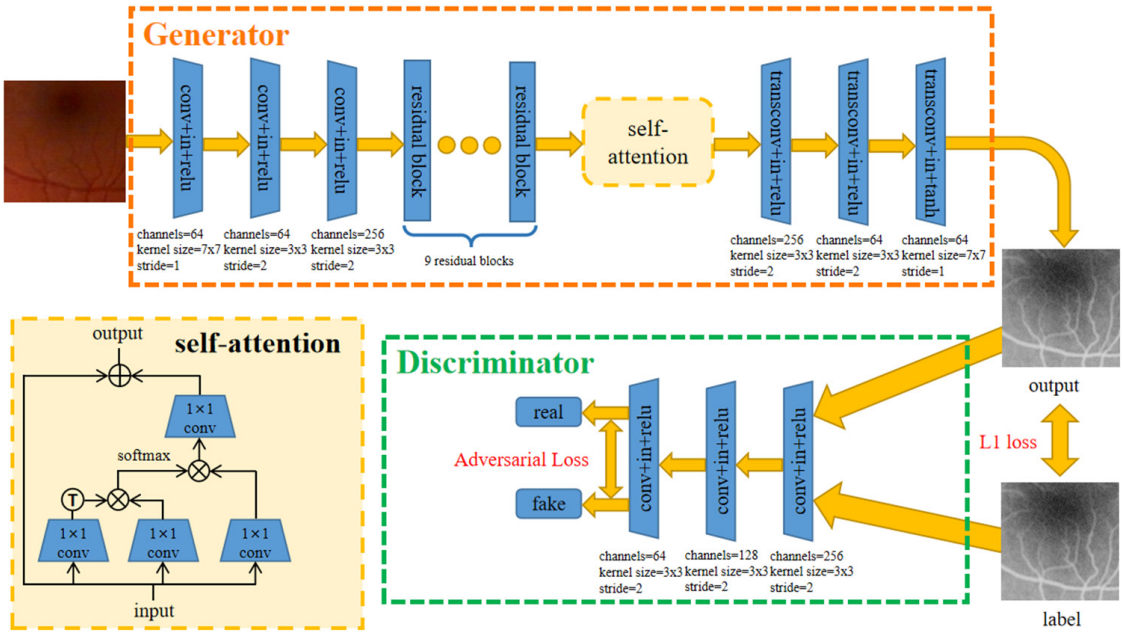


Fig. 1. Architecture of proposed deep learning model (conv = convolution, in = instance normalization, relu = rectified linear unit, and transconv = transposed convolution).

For the dataset used to predict a three-phase FFA image from an SLO image, the input was the latter, and the output was the former. The SLO image, which was originally in grayscale, was converted to RGB format by separately copying the grayscale information to the red, green, and blue channels. The three-phase FFA image, which was originally composed of three grayscale images, was converted to RGB format by copying the late, venous, and arterial phases to the red, green, and blue channels, respectively, as illustrated in Fig. 2.

After unifying the input and output image formats for the proposed deep learning model, we described the sources and divided the datasets. The publicly available Isfahan MISP dataset,[23] which contains 58 fundus camera/FFA image pairs of 720 $\times$ 576 pixels, was used to predict FFA images from the corresponding fundus camera images. Of the image pairs, 50 were used as the training set, 4 were used as the validation set, and the remaining 4 were used as the test set. For prediction of a single-phase FFA image from an SLO image, a nonpublicly

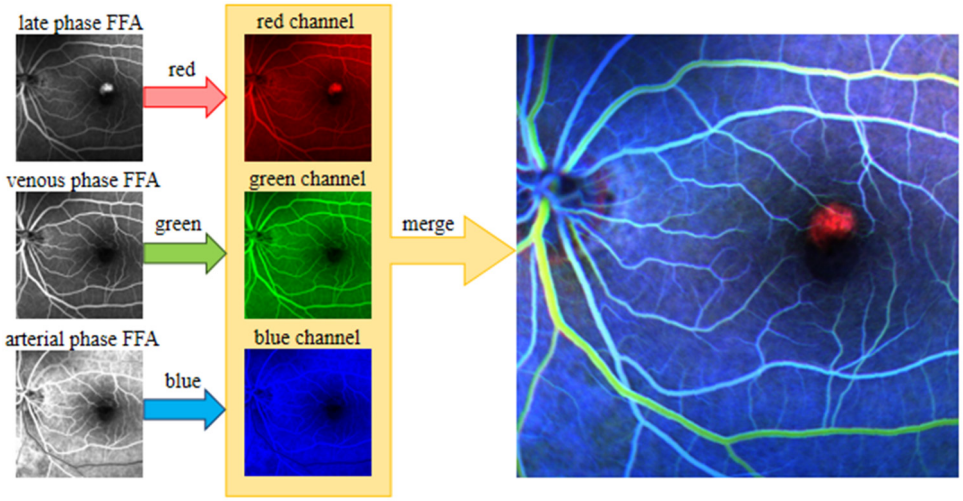


Fig. 2. Three-phase FFA image originally represented by three grayscale images and conversion into RGB image by copying the late, venous, and arterial phases to the red, green, and blue channels, respectively.

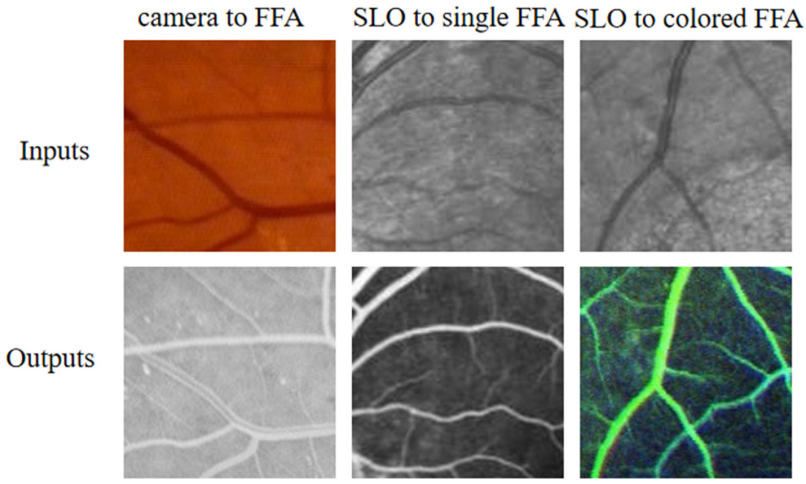


Fig. 3. Examples of manually registered and cropped image pairs for predicting FFA images.

available dataset was used. This dataset contained 97 image pairs of $512 \times 512$ pixels captured using the Spectralis HRA imaging platform (Heidelberg Engineering, Heidelberg, Germany) at the Third People's Hospital of Changzhou (Jiangsu, China), with 75, 11, and 11 image pairs being used for the training, validation, and test sets, respectively. For prediction of a three-phase FFA image from an SLO image, another nonpublicly available dataset containing 66 image groups of $768 \times 768$ pixels was used. The dataset was also collected using the Spectralis HRA (Heidelberg Engineering) at the Third People's Hospital of Changzhou (Jiangsu, China). Three-phase FFA images were obtained during the arterial (11–15 s), venous (16–20 s), and late (5–6 min) phases. The 51 collected image pairs were used for the training set, and the remaining 2 and 13 image pairs were used for the validation and test sets, respectively.

We performed preprocessing before using the datasets. First, we manually registered the paired images via translation and rotation, which is essential for supervised learning with generative adversarial networks. The training, validation, and test sets were augmented to 2000, 160, and 100 image pairs, respectively, by applying random cropping. Examples of the manually registered and cropped image pairs are shown in Fig. 3.

### 2.3. *Model training and testing*

The proposed model was trained using image pairs with a batch size of 64 over 200 epochs. During optimization, the generator and discriminator were alternately optimized. The generator was optimized with the L1 loss and its adversarial loss function, whereas the discriminator was only optimized with its adversarial loss function. The Adam optimizer was adopted with an initial learning rate of 0.0002, following a linear decay policy of multiplying by 0.1 every 50 epochs. The loss of our model in the training process is showed in Fig. 4. From our perspective, we hold the view that the similarity metric between the synthesized FFA and real FFA images is an important basis for assessing the accuracy and clinical relevance of the generated FFA images. To ensure the accuracy and clinical relevance of the synthesized FFA images, we trained our proposed model a lot of times and chose the one with the highest similarity metric. Once the model is trained, the synthesized FFA images are stable. The trained model was used to predict FFA images from the corresponding fundus structure images.

## 3. Results

Image processing was performed using a computer running 64-bit Python on the Linux Ubuntu operating system equipped with an Intel Xeon E5-2620 processor (2.10 GHz), 32 GB of memory, and an NVIDIA GeForce GTX 1080Ti graphics card. It worthy to note that the typical computation time of processing one image group is 0.89s, so our technique is potential for real-time application in clinical settings, which is much faster traditional FFA imaging techniques that takes about half an hour.

Figure 5 shows three typical examples for a qualitative comparison of our model with pix2pix[18]

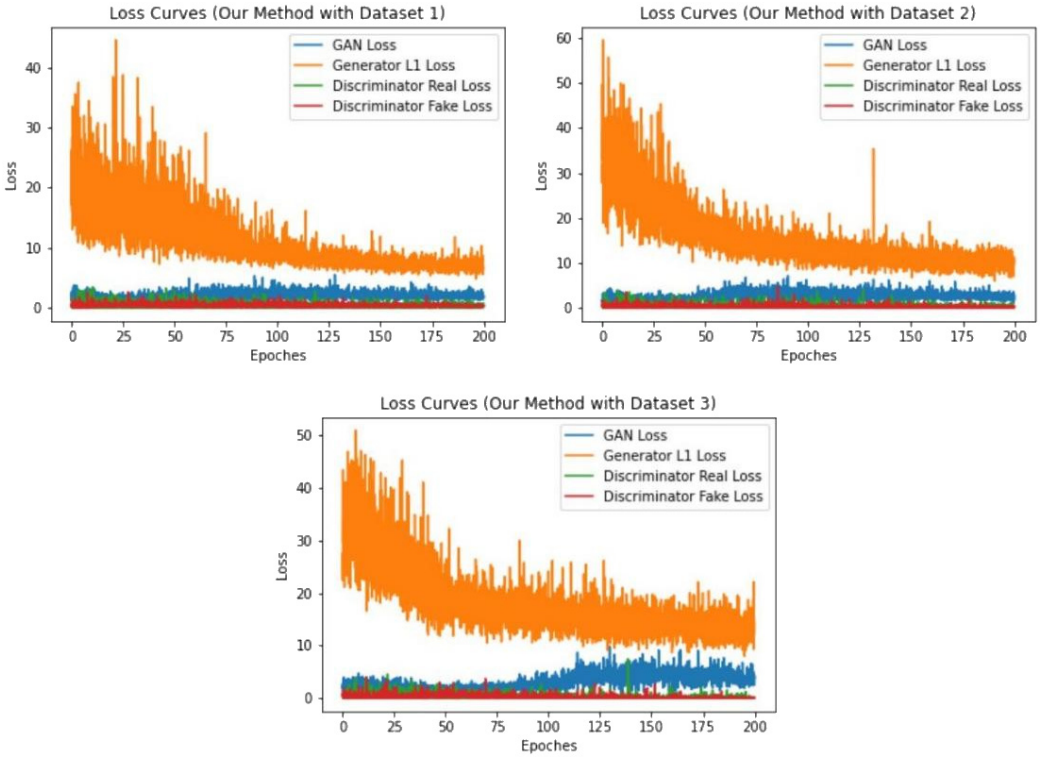


Fig. 4. The loss of our model in the training process (Dataset 1: fundus camera images to FFA images; Dataset 2: SLO images to single-phase FFA images; Dataset 3: SLO images to three-phase FFA images).

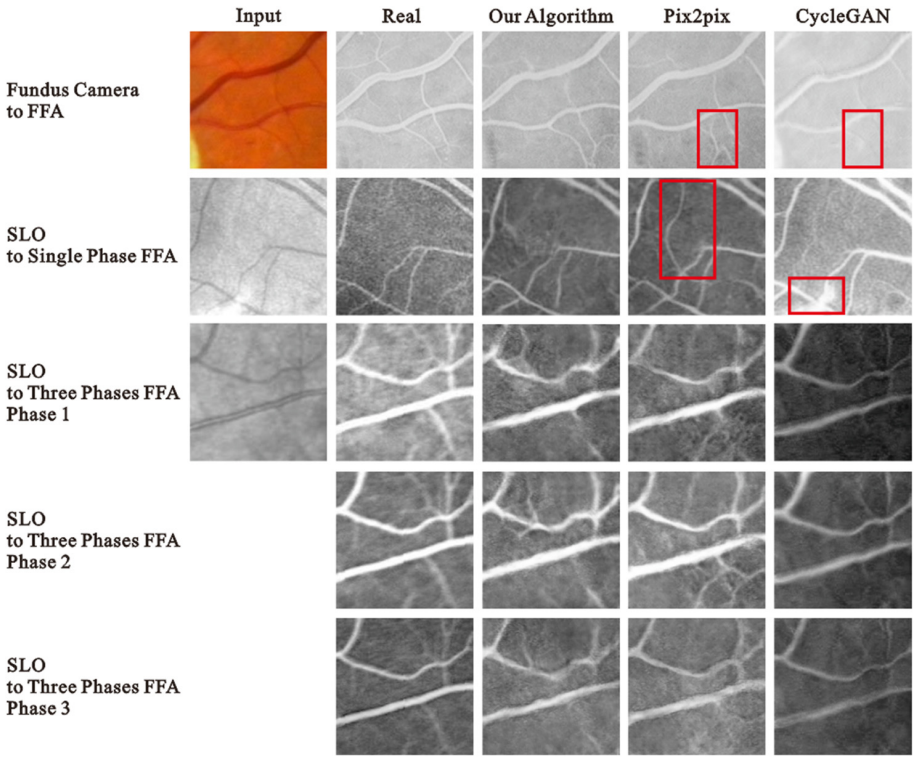


Fig. 5. Three examples of qualitative comparison between proposed model and state-of-the-art methods (pix2pix[18] and cycle-GAN[20]). The red boxes enclose prediction artifacts.

Table 1. Similarity metrics between real and predicted FFA images.

| | | | | |
|---|---|---|---|---|
| Fundus camera image to FFA image | Metric | Our model | pix2pix[18] | cycleGAN[20] |
| | SSIM | 0.6237 | 0.6069 | 0.5662 |
| | PSNR | 20.3488 | 19.6295 | 19.2347 |
| | MSE | 838.9363 | 1033.7321 | 1363.5402 |
| SLO image to single-phase FFA image | Metric | Our model | pix2pix[18] | cycleGAN[20] |
| | SSIM | 0.5102 | 0.5042 | 0.3494 |
| | PSNR | 16.6799 | 16.1639 | 12.7712 |
| | MSE | 1963.4722 | 2120.7682 | 4304.5082 |
| SLO image to three-phase FFA image | Metric | Our model | pix2pix[18] | cycleGAN[20] |
| | SSIM | 0.4198 | 0.4123 | 0.3633 |
| | PSNR | 14.7525 | 14.3044 | 12.9954 |
| | MSE | 3054.3708 | 3377.9785 | 4945.2119 |

and cycleGAN.[20] The first example is a fundus camera image converted into an FFA image. The second example is conversion of an SLO image into a single-phase FFA image, and the third example is conversion of an SLO image into a three-phase FFA image. For each example, the input, real FFA, and predicted images are shown. The predictions were generated by our model, pix2pix,[18] and cycleGAN.[20] The FFA image predicted by our model had a higher degree of similarity to the real image than the predictions of pix2pix[18] and cycleGAN.[20] For the fundus-camera-to-FFA and SLO-to-single-phase-FFA image conversions, the images from pix2pix[18] and cycleGAN[20] exhibited prediction artifacts (red boxes). For the conversion from the SLO image to the three-phase FFA image, most images exhibited prediction artifacts owing to fluctuations in the real three-phase FFA images and the limited size of our datasets. In this case, our model had similar performance to pix2pix[18] and outperformed cycleGAN.[20]

To confirm the high performance of our model, we quantitatively compared the FFA images predicted by our model, pix2pix,[18] and cycleGAN[20] with regard to the PSNR, SSIM,[21] and MSE, as shown in Table 1. The real FFA images were the ground truth. Our model achieved a higher mean PSNR and SSIM[21] and smaller mean MSE than the other methods. This suggests that the FFA images predicted by our model have a higher overall similarity to the real FFA images than those predicted by pix2pix[18] and cycleGAN.[20]

## 4. Discussion

We propose a method for representing the three phases of an FFA image as a single RGB image and introduce a unified deep learning model to predict FFA images from fundus structural images. The RGB representation may be used in other applications. For instance, this representation can be used for leakage detection. As shown in Fig. 6, leakage often appears as red shades because it mainly appears in the late-phase FFA image, which is encoded in the red channel. In addition, the RGB representation can be used in image classification for retinal disease diagnosis. Because common image classification methods often support only one input image, the RGB format is a suitable representation choice.

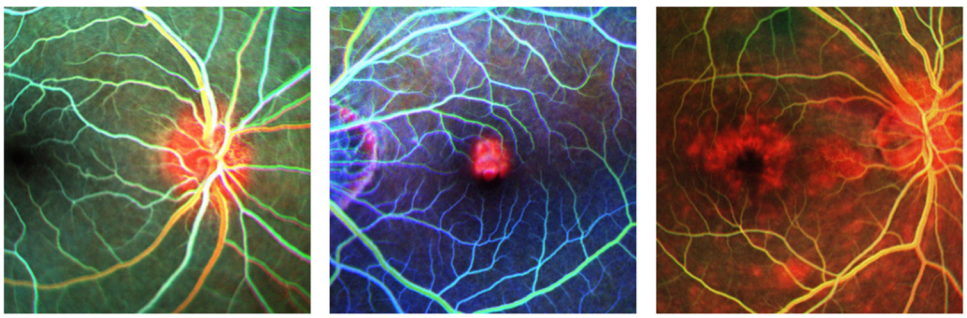

Fig. 6. Leakage images in RGB representation.

The ophthalmology community considers FFA imaging as the gold standard for diagnosing retinal vascular diseases. However, this technique requires injecting intravenous dyes into the human body, which can lead to certain levels of toxic side effects and, in severe cases, even pose a threat to life. As the elderly population is the main target group for FFA examinations, they are more prone to experiencing adverse reactions due to their physical condition. An important clinical significance of this technology lies in its ability to provide FFA-style images of fundus vasculature without the need for intravenous dyes injection in elderly individuals. Therefore, this technique can serve as a noninvasive diagnostic tool for retinal diseases in the elderly.

As a potential way to integrate this technology into existing diagnostic workflows, we can first use this technology for population screening. When the screening results indicate a potential disease condition, further confirmation and precise diagnosis can then be conducted using traditional FFA techniques.

At present, our dataset may not be sufficient to fully investigate the limitations of our model in predicting FFA images for specific disease states or patient demographics. However, we acknowledge the possibility of potential limitations in certain scenarios, and the investigation of these potential limitations can be future research direction.

In order to address the scalability and adaptability of our proposed model to accommodate future technological developments and evolving clinical needs, several key considerations are taken into account. First, our model is designed with a modular and flexible architecture, allowing for easy integration of new techniques and algorithms as they emerge. This will ensure that the model can be updated and adapted to incorporate the latest advancements in artificial intelligence and deep learning. Second, establishing a robust and scalable infrastructure is crucial. This includes ensuring sufficient computational resources and storage capacity, as well as implementing distributed computing frameworks to handle large-scale data processing. The use of cloud-based solutions can also provide scalability and flexibility by allowing the model to scale up or down based on demand. Third, continuous evaluation and validation of the model's performance and accuracy are necessary. Regular monitoring and feedback collection from clinicians and domain experts will enable timely adjustments and improvements to accommodate evolving clinical needs.

## 5. Conclusions

We introduce a unified deep learning model for predicting FFA images from fundus structure images. This model is based on supervised learning and combines the pix2pix[18] architecture and a self-attention module.[19] The high performance of our model is confirmed qualitatively and quantitatively through comparisons with the state-of-the-art pix2pix[18] and cycleGAN[20] methods. The proposed model can predict FFA images from fundus structure images well in a unified architecture. In addition, it may be useful for ophthalmic examinations with the generated FFA, especially for screening and diagnosis, while avoiding frequent intravenous dye injections.

## Acknowledgments

The authors thank the anonymous reviewers for their valuable suggestions. This work was supported in part by the Gusu Innovation and Entrepreneurship Leading Talents in Suzhou City, grant numbers ZXL2021425 and ZXL2022476; Doctor of Innovation and Entrepreneurship Program in Jiangsu Province, grant number JSSCBS20211440; Jiangsu Province Key R&D Program, grant number BE2019682; Natural Science Foundation of Jiangsu Province, grant number BK20200214; National Key R&D Program of China, grant number 2017YFB0403701; National Natural Science Foundation of China, grant numbers 61605210, 61675226, and 62075235; Youth Innovation Promotion Association of Chinese Academy of Sciences, grant number 2019320; Frontier Science Research Project of the Chinese Academy of Sciences, grant number QYZDB-SSW-JSC03; and Strategic Priority Research Program of the Chinese Academy of Sciences, grant number XDB02060000.

## Conflict of Interest

The authors declare that there is no conflict of interest relevant to this paper.

## ORCID

Yiwei Chen https://orcid.org/0000-0002-3638-2309

## References

1. S. S. Hayreh, "Ocular vascular occlusive disorders: Natural history of visual outcome," *Prog. Retin. Eye Res.* **41**, 1–25 (2014).

2. C. J. Flaxel, R. A. Adelman, S. T. Bailey, A. Fawzi, J. I. Lim, G. A. Vemulakonda, G. S. Ying, "Retinal and ophthalmic artery occlusions preferred practice pattern®," *Ophthalmology* **127**, 259–287 (2020).
3. R. F. Spaide, J. M. Klancnik, M. J. Cooney, "Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography," *JAMA Ophthalmol.* **133**, 45–50 (2015).
4. S. L. Owens, "Indocyanine green angiography," *Br. J. Ophthalmol.* **80**, 263 (1996).
5. Y. Chen, Y. J. Hong, S. Makita, Y. Yasuno, "Eye-motion-corrected optical coherence tomography angiography using Lissajous scanning," *Biomed. Opt. Express* **9**, 1111–1129 (2018).
6. S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. Zuckerbrod, S. A. Baker, K. M. Sanders, "Fundus2Angio: A conditional GANarchitecture for generating fluorescein angiography images from retinal fundus photography," *Proc. Int. Symp. Visual Computing*, San Diego, CA, USA, 5–7 November 2020, pp. 125–138, Springer, Cham, Switzerland (2020).
7. S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, "Attention2angiogan: Synthesizing fluorescein angiography from retinal fundus images using generative adversarial networks," *Proc. the 2020 25th Int. Conf. Pattern Recognition (ICPR)*, Milan, Italy, 10–15 January 2021, pp. 9122–9129.
8. Á. S. Hervella, J. Rouco, J. Novo, M. Ortega, "Retinal image understanding emerges from self-supervised multimodal reconstruction," *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention*, Granada, Spain, 16–20 September 2018, pp. 321–328, Springer, Cham, Switzerland (2018).
9. A. Tavakkoli, S. A. Kamran, K. F. Hossain, S. L. Zuckerbrod, "A novel deep learning conditional generative adversarial network for producing angiography images from retinal fundus photographs," *Sci. Rep.* **10**, 21580 (2020).
10. F. Schiffers, Z. Yu, S. Arguin, A. Maier and Q. Ren, "Synthetic fundus fluorescein angiography using deep neural networks," *Bildverarbeitung für die Medizin 2018*, pp. 234–238, Springer Vieweg, Berlin, Germany, 2018.
11. K. Li, L. Yu, S. Wang, P. A. Heng, "Unsupervised retina image synthesis via disentangled representation learning," *Proc. Int. Workshop on Simulation and Synthesis in Medical Imaging*, Shenzhen, China, 13 October 2019, pp. 32–41, Springer, Cham, Switzerland (2019).
12. Z. Cai, J. Xin, J. Wu, S. Liu, W. Zuo, N. Zheng, "Triple multi-scale adversarial learning with self-attention and quality loss for unpaired fundus fluorescein angiography synthesis," *Proc. 2020 42nd Annual Int. Conf. IEEE Engineering in Medicine Biology Society (EMBC)*, Montreal, QC, Canada, 20–24 July 2020, pp. 1592–1595.
13. Á. S. Hervella, J. Rouco, J. Novo, M. Ortega, "Deep multimodal reconstruction of retinal images using paired or unpaired data," *Proc. 2019 Int. Joint Conf. Neural Networks (IJCNN)*, Budapest, Hungary, 14–19 July 2019, pp. 1–8.
14. W. Li, W. Kong, Y. Chen, J. Wang, Y. He, G. Shi, G. Deng, "Generating fundus fluorescence angiography images from structure fundus images using generative adversarial networks," preprint (2020), arXiv:2006.10216.
15. W. Li, Y. He, W. Kong, J. Wang, G. Deng, Y. Chen, G. Shi, "SequenceGAN: Generating fundus fluorescence angiography sequences from structure fundus image," *Proc. Int. Workshop on Simulation and Synthesis in Medical Imaging, Strasbourg*, France, 27 September 2021, pp. 110–120, Springer, Cham, Switzerland (2021).
16. Y. Chen, Y. He, W. Li, J. Wang, P. Li, L. Xing, X. Zhang, G. Shi, "Series–parallel generative adversarial network architecture for translating from fundus structure image to fluorescence angiography," *Appl. Sci.* **12**(20), 10673 (2022).
17. P. Li, Y. He, P. Wang, J. Wang, G. Shi, Y. Chen, "Synthesizing multi-frame high-resolution fluorescein angiography images from retinal fundus images using generative adversarial networks," *BioMed. Eng. OnLine* **22**(1), 1–15 (2023).
18. P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 13–18 June 2010, pp. 1125–1134.
19. X. Wang, R. Girshick, A. Gupta, K. He, "Non-local neural networks," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 7794–7803, https://scholar.google.com/scholar?hl=zh-CN&as_sdt=0%2C5&q=X.+Wang%2C+R.+Girshick%2C+A.+Gupta%2C+K.+He%2C+%5CNon-local+neural+networks%2C%22+Proc.+IEEE+Conf.+Computer+Vision+and+Pattern+Recognition%2C+pp.+7794%E2%80%937803.&btnG=.
20. J. Y. Zhu, T. Park, P. Isola, A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proc. IEEE Int. Conf. Computer Vision*, Venice, Italy, 22–29 October 2017, pp. 2223–2232.
21. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**, 600–612 (2004).
22. X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, S. Paul Smolley, "Least squares generative adversarial

networks," *Proc. IEEE Int. Conf. on Computer Vision*, pp. 2794–2802, https://scholar.google.com/scholar?hl=zh-CN&as˙sdt=0%2C5&q=X.+Mao%2C+Q.+Li%2C+H.+Xie%2C+R.+Y.+Lau%2C+Z.+Wang%2C+S.+Paul+Smolley%2C+%5CLeast+squares+generative+adversarial+networks%2C%22+Proc.+IEEE+Int.+Conf.+on+Computer+Vision%2C+pp.+2794%E2%80%932802.&btnG=.

23. S. Hajeb Mohammad Alipour, H. Rabbani, M. R. Akhlaghi, Diabetic retinopathy grading by digital curvelet transform, *Comput. Math. Methods Med.* (2012), https://scholar.google.com/scholar?hl=zh-CN&as˙sdt=0%2C5&q=+S.+Hajeb+Mohammad+Alipour%2C+H.+Rabbani%2C+M.+R.+Akhlaghi%2C+Diabetic+retinopathy+grading+by+digital+curvelet+transform%2C+Comput.+Math.+Methods+Med.+%282012%29.&btnG=.