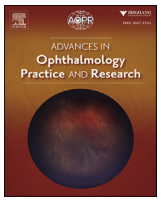
Advances in Ophthalmology Practice and Research 3 (2023) 192–198

Contents lists available at ScienceDirect

Advances in Ophthalmology Practice and Research

journal homepage: www.journals.elsevier.com/advances-in-ophthalmology-practice-and-research



Full Length Article

A deep learning model for generating fundus autofluorescence images from color fundus photography

Fan Song [a,b], Weiyi Zhang [a,b], Yingfeng Zheng [c], Danli Shi [a,b,*], Mingguang He [a,b]

[a] Experimental Ophthalmology, School of Optometry, The Hong Kong Polytechnic University, Hong Kong, China
[b] Research Centre for SHARP Vision, The Hong Kong Polytechnic University, Hong Kong, China
[c] State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangdong Provincial Clinical Research Center for Ocular Diseases, Guangzhou, China

ARTICLE INFO

Keywords:
Generative adversarial networks
Color fundus to fundus autofluorescence generation
Age-related macular degeneration
Deep learning

ABSTRACT

Background: Fundus Autofluorescence (FAF) is a valuable imaging technique used to assess metabolic alterations in the retinal pigment epithelium (RPE) associated with various age-related and disease-related changes. The practical uses of FAF are ever-growing. This study aimed to evaluate the effectiveness of a generative deep learning (DL) model in translating color fundus (CF) images into synthetic FAF images and explore its potential for enhancing screening of age-related macular degeneration (AMD).

Methods: A generative adversarial network (GAN) model was trained on pairs of CF and FAF images to generate synthetic FAF images. The quality of synthesized FAF images was assessed objectively by common generation metrics. Additionally, the clinical effectiveness of the generated FAF images in AMD classification was evaluated by measuring the area under the curve (AUC), using the LabelMe dataset.

Results: A total of 8410 FAF images from 2586 patients were analyzed. The synthesized FAF images exhibited an impressive objectively assessed quality, achieving a multi-scale structural similarity index (MS-SSIM) of 0.67. When evaluated on the LabelMe dataset, the combination of generated FAF images and CF images resulted in a noteworthy improvement in AMD classification accuracy, with the AUC increasing from 0.931 to 0.968.

Conclusions: This study presents the first attempt to use a generative deep learning model to create authentic and high-quality FAF images from CF images. The incorporation of the translated FAF images on top of CF images improved the accuracy of AMD classification. Overall, this study presents a promising approach to enhance large-scale AMD screening.

1. Introduction

Age-related macular degeneration (AMD) is a condition characterized by the gradual deterioration of central vision due to the damage occurring in the macular region of the retina.[1] It accounts for around 9% of global blindness cases, particularly among individuals aged 60 and above.[2] AMD can be classified as early AMD for individuals with medium drusen, intermediate AMD for those with large drusen or pigmentary abnormalities, and late AMD for individuals with neovascular abnormalities or with geographic atrophy.[3] Early detection plays a crucial role in preventing the progression of AMD.[4]

Fundus autofluorescence (FAF) is an essential imaging technique used to assess the function of the retinal pigment epithelium (RPE).[5] It relies on the intrinsic fluorescence emitted by lipofuscin granules, which are metabolic byproducts of RPE cells that can accumulate with age. Abnormal FAF patterns, characterized by localized areas of hyperfluorescence and/or hypofluorescence, are occasionally observed and are believed to be indicative of changes in retinal function. These patterns offers significant advantages in detecting, evaluating, and treating both atrophic AMD and wet AMD.[6–8] Despite the important clinical applications of FAF in specific diseases, its overall utilization in ophthalmology is relatively limited compared to color fundus (CF) photography. This is primarily due to limited awareness among some doctors regarding the potential of FAF and patients' hesitancy to undergo extensive testing, and the shortage of FAF imaging device in screening scenario. Consequently, synthesizing precise FAF images based on corresponding CF images

* Corresponding author. Research Centre for SHARP Vision, Experimental Ophthalmology, School of Optometry, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China.
E-mail address: danli.shi@polyu.edu.hk (D. Shi).

https://doi.org/10.1016/j.aopr.2023.11.001
Received 29 August 2023; Received in revised form 4 November 2023; Accepted 5 November 2023
Available online 10 November 2023

provides a practical solution to overcome those challenges.

The development of deep learning models has opened up the possibility of modality translation from CF images to FAF images. Image-to-image translation is an innovative technique that converts images from one style to another while preserving their essential features. Previous studies have been dedicated to the task of translating CF images to fundus fluorescein angiography (FFA) images using generative adversarial networks (GANs), and the results have demonstrated high-quality generated images.[9–11] In addition, a previous study utilized a geographic atrophy region-aware conditional GAN for generating trustworthy FAF images from en-face optical coherence tomography (OCT).[12] However, there are no prior arts addressing the translation of CF images to FAF images.

Our objective was to create and evaluate a generative deep learning model capable of accurately translating color CF images into FAF images. The model was trained on a large-scale clinical dataset, and we assessed its effectiveness in improving AMD classification using the LabelMe dataset. The synthesized FAF images generated by our model hold promise as an alternative to traditional FAF in enhancing the efficiency of screening for AMD and other retinal disorders.

2. Methods

2.1. Participants

This retrospective study included a collection of 7049 color fundus (CF) images and 8410 fundus autofluorescence (FAF) images from 2586 patients who underwent regular clinical examinations at a tertiary hospital between 2016 and 2019. The CF images were obtained using Topcon TRC-50XF and Zeiss FF450 Plus cameras (Carl Zeiss, Inc., Jena, Germany), featuring resolutions ranging from 1110 × 1467 to 2600 × 3200. The FAF images were captured using Zeiss FF450 Plus cameras and Heidelberg Spectralis cameras (Heidelberg, Germany) with a resolution of 768 × 768. All patient information was meticulously anonymized and devoid of any identifying details.

To assess the clinical application of the generated FAF images, we utilized the LabelMe dataset (http://www.labelme.org, Guangzhou, China) to evaluate their potential in enhancing age-related macular degeneration (AMD) screening. Experienced ophthalmologists performed a widely used AMD classification system (Beckman clinical classification system), which ranges from 0 to 3, to assess the presence of AMD in each image.[3]

The study adhered to the principles of the Declaration of Helsinki. This retrospective study received ethical approval from the Institutional Review Board (No.2021KYPJ164-3), and the need for individual consent for the analysis was waived.

2.2. CF and FAF matching

We performed image matching between CF and FAF images obtained during the same eye examination session. To achieve precise pixel-level alignment, we utilized the Retina-based microvascular health assessment system (RMHAS)[13] to extract retinal vessels from CF images, and an in-house multimodality vessel segmentation model for extracting vessels from FAF images. Similar to our previous study,[14] we registered the FAF images and subsequently aligned them with their corresponding CF images. Key points were extracted from the vessel maps of both CF and FAF images using the AKAZE key point detector.[15] These key points were then utilized for feature matching, and Random Sample Consensus (RANSAC)[16] was applied to generate homography matrices and eliminate outliers. To ensure accurate registration, we imposed validity restrictions by constraining the rotation scale within the range of 0.8–1.3 and limiting the absolute value of the rotation radian to less than 2. We also discarded image pairs with inadequate registration performance, which was determined by a dice coefficient below 0.5, based on empirical observations from our dataset.

2.3. FAF image synthesis

Our model was trained using CF images as input and real FAF images as ground truth, with a split ratio of 8 : 1 : 1 for training, validation, and evaluation at patient-level. During training, we resized the images to a resolution of 512 × 512 and utilized pix2pixHD model,[17] which is a widely-used generative adversarial networks (GAN) model for image translation. This model employed a minimax game approach, where the generator G aimed to produce a convincing FAF image to deceive the discriminator D, while the discriminator D endeavored to differentiate between the generated image and the real one.[17] To generate high-resolution FAF images, we employed a multi-scale convolutional neural network as the discriminator.[17] This network partitioned the images into patches and independently assessed the fidelity of each patch independently. By considering the quality of individual patches, our approach aimed to produce coherent and detailed FAF images at a high resolution. Moreover, we introduced Gradient Variance Loss to improve the generation performance of high-frequency elements, such as retinal structure and lesions.[18] During training, we utilized a batch size of 4 and set the learning rate to 0.0002. To address the issue of overfitting, we implemented data augmentation strategies during training, such as randomly resized crops within a scale range of 0.5–2, applying random horizontal or vertical flipping. To optimize the model's performance, we conducted a total of 50 epochs for each training session. We implemented the image translation model using PyTorch and conducted the training process on an NVIDIA GeForce RTX 3090 GPU.

2.4. DL evaluation methodology

2.4.1. Objective evaluation

For the quantitative evaluation of the internal test set, we used common objective metrics: structural similarity measures (SSIM),[19] multi-scale structural similarity index (MSSSIM),[20] Fréchet inception distance (FID),[21] mean absolute error (MAE),[22] and peak signal-to-noise ratio (PSNR).[23] SSIM is a metric used to measure the structural similarity between two images. A value closer to 1 indicates a higher similarity between the images.[19] Unlike conventional SSIM, which focuses solely on the local structure at a specific scale, MS-SSIM integrates multiple scales through a sequence of image down-sampling and filtering operations. This enables MS-SSIM to encompass structural information across various levels of granularity.[20] FID is a metric used to assess the dissimilarity between generated and real images. It calculates the discrepancy in feature representations extracted from these images using a pre-trained inception network, where lower FID values indicate better similarity between the generated and real images.[21] MAE and PSNR are metrics used to assess image reconstruction quality. MAE calculates the average difference between corresponding pixels in the generated and real images, where lower values mean better reconstruction quality.[22] PSNR measures the quality of the reconstructed image by evaluating the ratio of the peak signal power to the noise power, where higher values correspond to superior image quality.[23]

2.4.2. AMD classification

We performed a comparative study to evaluate the effectiveness of generated FAF images in the classification of AMD on the LabelMe dataset. It's worth noting that LabelMe does not inherently contain ground truth data for FAF images. In this context, we utilized the model developed on our internal dataset to generate synthetic FAF images within the LabelMe dataset. The AMD severity scale was classified using the Swin-transformer[24] model under consistent experimental conditions. Swin-transformer is able to capture vital long-range dependencies, which is essential for intricate fundus image analysis. Its layered attention mechanism enables multi-scale feature extraction, enhancing sensitivity to diverse pathological features. We considered CF images alone and CF images combined with the generated FAF images for classification, with AMD severity classes ranging from 0 to 3 (0 = no AMD, 1 = early or

intermediate AMD, 2 = late dry AMD, 3 = wet AMD).[3] The LabelMe dataset was utilized for the experiment and was divided into three sets: training (60%), validation (20%), and testing (20%). The images were resized to a resolution of 512 × 512 and underwent data augmentation during training, including random horizontal flips and rotations between −30° to +30°. We employed the Adam optimizer with a small learning rate of 1e-5 and a modest batch size of 4. The training process consisted of 30 epochs, and for evaluation, we selected the models that achieved the highest area under the curve (AUC) value on the validation set.

Our DL algorithm was implemented using PyTorch, and training of the models was performed on an NVIDIA GeForce RTX 3090 GPU.

3. Results

After excluding 2895 CF images and 1620 FAF images for reasons such as being off-centered from the macula, unsuccessful CF-FAF pairwise matching, or poor image quality, a total of 6790 pairs of CF-FAF images obtained from 1850 participants were used for model development. The participants had a median age of 47.99 years (interquartile range: 16.52), and 1024 (55.4%) of them were male. The study flow chart is depicted in Fig. 1.

3.1. Generation of FAF images

Samples of the real and generated images from both the internal and external test sets are shown in Fig. 2. In the internal and external test sets, a portion of the generated FAF images exhibited poor quality attributed to the following factors: synthetic FAF images from blurry CF images might fail to accurately depict lesions. The algorithm may overlook lesions, such as subtle drusen, that were not visibly evident on the CF images. Furthermore, hyperautofluorescent lesions presented a greater challenge in accurate representation on translated FAF images compared to hypoautofluorescent lesions, as illustrated in Supplementary Fig. 1. While this study primarily focused on AMD, Supplementary Fig. 2 provides generation samples for other retinal diseases, such as papilledema, central serous chorioretinopathy, diabetic retinopathy, retinopathy post-laser treatment and branch retinal vein occlusion to further exemplifying the model's generation performance. Additionally, we have provided examples of hyperautofluorescent and hypoautofluorescent lesions in Supplementary Fig. 3.

3.2. Objective evaluation

Pairwise comparison between original and generated FAF was conducted on the internal test set, the SSIM, MS-SSIM, FID, MAE, and PSNR were 0.51, 0.67, 53.73, 127.16, 19.22. The synthesized image quality is considered better when the PSNR, SSIM, and MS-SSIM values are higher, and the FID and MAE values are lower. The results above are demonstrated in Table 1.

3.3. AMD classification evaluation

The incorporation of generated FAF images resulted in a significant improvement in the quantitative outcomes of AMD classification on the LabelMe dataset, as clearly illustrated in Table 2 and Fig. 3. The inclusion of generated FAF images alongside CF images notably enhanced the accuracy of AMD classification, particularly in the 0 and 1 categories. Detailed characteristics of the LabelMe dataset are illustrated in Table 3.

4. Discussion

To the best of our knowledge, this study represents the first attempt to develop and evaluate a generative deep learning model for synthesizing realistic FAF images from CF images. The integration of the generated FAF images with CF images led to a significant enhancement in the accuracy of AMD classification. Furthermore, our study introduced a compelling alternative for FAF and presented a promising approach to improve AMD screening on a larger scale.

The translation of CF to FAF images offers several benefits. FAF has expanded its clinical utility, proving valuable in diagnosing and managing a diverse range of chorioretinal conditions, encompassing AMD, retinal drug toxicities, central serous chorioretinopathy, and inherited retinal degenerations, such as Stargardt disease and retinitis pigmentosa (RP).[25] By adeptly mapping the metabolic activity of the retina and effectively highlighting areas with lipofuscin accumulation, FAF enhances its diagnostic efficacy and clinical relevance. The qualitative and quantitative analysis of FAF has shown its value in detecting clinically significant AMD-associated lesions, such as reticular pseudodrusen, regressed drusen, and smaller geographic atrophy regions, offering essential diagnostic and prognostic insights for medical decision-making.[6,26–28] However, many patients are hesitant to undergo extensive auxiliary examinations, and primary care hospitals may lack

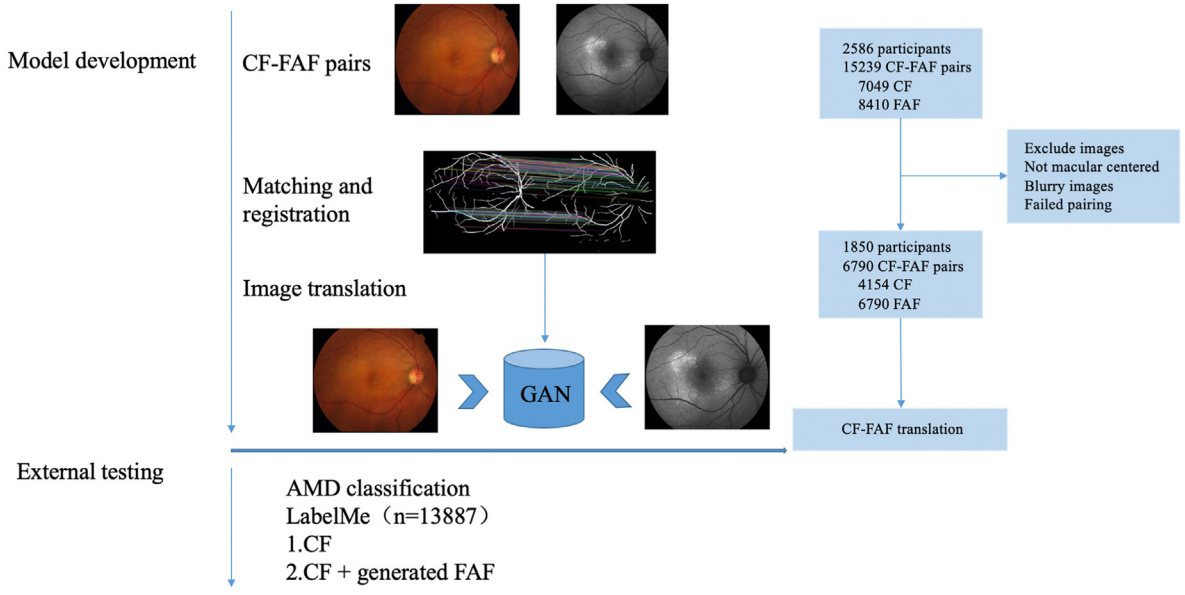


Fig. 1. Flow chart of the study. CF = color fundus, FAF = fundus autofluorescence, AMD = age-related macular degeneration, GAN = generative adversarial network, n = number of images.

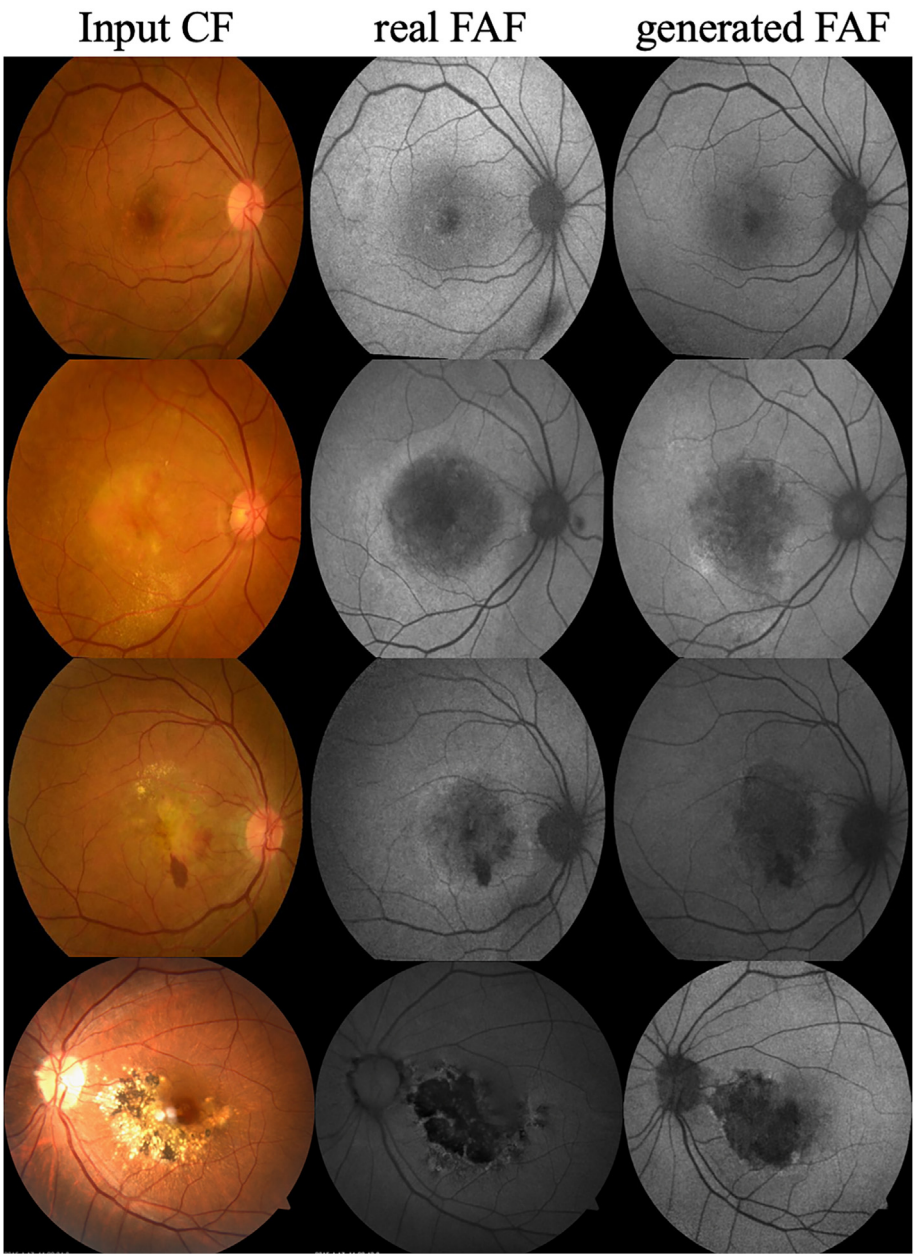


Fig. 2. Samples of real and generated fundus autofluorescence (FAF) images.1st row, early age-related macular degeneration (AMD), 2nd row, wet AMD, 3rd row, wet AMD, 4th row, long-standing wet AMD. 1–3 rows: internal test set, color fundus (CF) were registered with FAF, rotation was applied during this process. 4th row: external test set.

Table 1
Objective evaluation between original and generated fundus autofluorescence (FAF) images.

	MAE	PSNR	SSIM	MS-SSIM	FID
Generated FAF	127.16	19.22	0.51	0.67	53.73

MAE = mean absolute error, PSNR = peak signal-to-noise ratio, SSIM = structural similarity measures, MS-SSIM = multi-scale structural similarity index, FID = Fréchet inception distance.

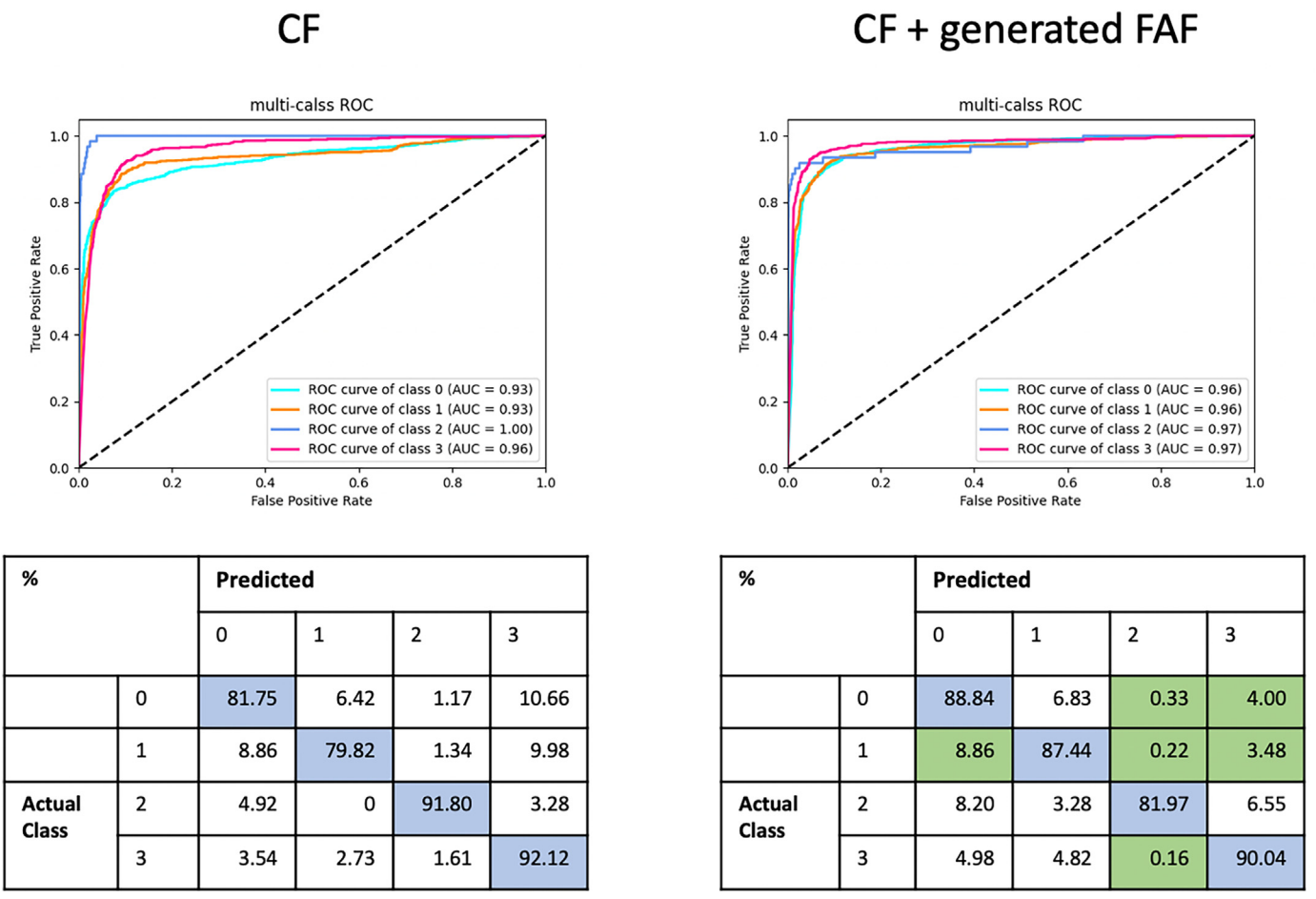
Table 2
Age-related macular degeneration (AMD) classification with color fundus (CF), CF + generated fundus autofluorescence (FAF) images.

	F1-score	Sensitivity	Precision	Specificity	Accuracy	AUC
Only CF	0.839	0.837	0.851	0.932	0.837	0.931
CF + FAF	0.885	0.885	0.885	0.940	0.885	0.967

LabelMe dataset: 13887 images (train:validation:test = 6 : 2 : 2). The classification metrics results in this table are generated using the test set of the LabelMe dataset. AMD classification (0 = no AMD, 1 = early or intermediate AMD, 2 = late dry AMD, 3 = wet AMD). AUC = the area under the curve.

the necessary diagnostic equipment. This technology can help reduce costs for patients and decrease the likelihood of misdiagnosis by healthcare professionals under limited conditions. Therefore, CF-FAF conversion holds promise as a clinically valuable technique.

Our study employed the widely-used GAN model, specifically the pix2pixHD[17] model, for the translation of FAF images from CF images. Unlike traditional GAN methods, pix2pixHD stands out due to its coarse to fine-grained generator, which enables the preservation of



%		Predicted			
		0	1	2	3
Actual Class	0	81.75	6.42	1.17	10.66
	1	8.86	79.82	1.34	9.98
	2	4.92	0	91.80	3.28
	3	3.54	2.73	1.61	92.12

%		Predicted			
		0	1	2	3
Actual Class	0	88.84	6.83	0.33	4.00
	1	8.86	87.44	0.22	3.48
	2	8.20	3.28	81.97	6.55
	3	4.98	4.82	0.16	90.04

Fig. 3. The additive value of generated fundus autofluorescence (FAF) atop of color fundus image (CF) for age-related macular degeneration (AMD) classification (0 = no AMD, 1 = early or intermediate AMD, 2 = late dry AMD, 3 = wet AMD). 1st row represents the receiver operating characteristic (ROC) curves for two models. 2nd row represents confusion matrixes. The color blue indicates correct predictions, while the color green indicates a reduced error rate with the addition of the generated FAF.

Table 3
LabelMe dataset characteristics.

	no AMD	early or intermediate AMD	late dry AMD	wet AMD	Total
Train	3600 (43.21%)	2673 (32.08%)	182 (2.19%)	1876 (22.52%)	8331
Validation	1200 (43.15%)	891 (32.04%)	61 (2.19%)	629 (22.62%)	2781
Test	1200 (43.24%)	892 (32.14%)	61 (2.20%)	622 (22.42%)	2775
Total	6000 (43.20%)	4456 (32.09%)	304 (2.19%)	3127 (22.52%)	13887

AMD = age-related macular degeneration.

high-resolution elements, such as retinal structures and lesions, with remarkable accuracy. Both the local and global parts of the generator are structured with a convolutional front end, a series of residual blocks, and a transposed convolutional back end. Moreover, the model incorporated a multi-scale convolutional neural network[17] as the discriminator to distinguish the FAF images generated by the generator from real images. Besides, we introduced Gradient Variance Loss to improve the generation performance of high-frequency elements, including retinal structure and lesions.[18] Furthermore, data augmentation strategies were implemented during training to address overfitting issues. These methods collectively contributed to achieving remarkable performance in generating high-resolution and authentic FAF images.

To ensure the algorithm's applicability across clinical images, we conducted evaluations on internal datasets. The synthesized images underwent quantitative evaluation and demonstrated favorable image quality compared to real FAF images, as evidenced by metrics such as the MS-SSIM score of 0.67, the SSIM score of 0.51, and the PSNR score of 19.22, which served as indicators of the quality of the generated FAF images. As this study represents the first attempt to synthesize FAF images from CF images, we compared the corresponding scores with those assessing the synthetic fundus fluorescence angiography images translated from CF images using a pix2pix model in a previous study, where SSIM and PSNR achieved 0.78 and 18.68, respectively.[29] Compared to the previous study, our PSNR score was higher, ensuring the high quality of generated FAF images. Although the SSIM score was not as satisfactory, our study introduced the MS-SSIM to comprehensively assess the similarity of images, and the result indicated that the generated FAF images were reliable when compared to real ones. In terms of classifying AMD on the LabelMe dataset, we observed improved results by incorporating the generated FAF images. Our approach achieved high accuracy in AMD classification, particularly when adding the synthesized FAF images. Furthermore, our algorithm was evaluated for enhancing AMD diagnosis, highlighting the potential of synthesized FAF images as an additional tool for AMD screening. Moreover, the algorithm demonstrated compelling performance in translating FAF images for a range of retinal diseases beyond AMD. Evaluation on both internal and external datasets revealed promising generation samples, including central serous chorioretinopathy, diabetic retinopathy, and branch retinal vein occlusion, as depicted in Supplementary Fig. 2. Additionally, FAF has

expanded its clinical utilization, becoming an indispensable tool for assessing macular dystrophies and a variety of inherited retinal disorders, such as RP, Stargardt disease, and Bestrophinopathies.[30] Further research is necessary to improve the algorithm's performance across the aforementioned diseases.

There are some limitations to consider. Firstly, the utilization of FAF images with a limited field of view of 55° may have resulted in the omission of certain peripheral manifestations of retinal diseases. Future research should focus on generating wide-field CF-based FAF images to enhance the comprehensiveness of this approach. Secondly, the impact of blurry CF images on the generation of FAF images should be considered. Blurry CF images may lead to blurry generated FAF images and potentially incorrect associations with AMD lesions. To mitigate this issue, patients with sub-optimal CF image quality should be excluded from the training data, and real-FAF examinations are recommended for accurate assessment. Thirdly, it is important to note that small lesions, such as subtle drusen, may not be prominent on CF images and could be missed by the algorithm. Finally, accurately representing hyperautofluorescent lesions on translated FAF images proved to be more challenging compared to hypoautofluorescent lesions. This limitation could potentially hinder the timely reflection of AMD progression, as the presence of hyperautofluorescent lesions at the margin of geographic atrophy suggests the involvement of larger lesions.[25,31] Future studies are warranted for improving the generation of small lesions and hyperautofluorescent lesions on the translated FAF images and the generated FAF should be used in combination with CF.

5. Conclusions

Our generative deep learning model demonstrates a feasible method of translating CF images into realistic FAF images for the first time. The ability to generate synthetic FAF images that closely resemble real ones holds promise for various applications, such as enhancing large-scale AMD screening and assisting in medical decision-making processes.

Study approval

The study adhered to the principles of the Declaration of Helsinki. This retrospective study received ethical approval from the Institutional Review Board, and the need for individual consent for the analysis was waived.

Author contributions

Conception and design of study: DS; Data collection: YZ, MH; Analysis and interpretation of results: WZ, FS, DS; Drafting the manuscript: FS; All authors reviewed the results and approved the final version of the manuscript.

Funding

This research received support from the Global STEM Professorship Scheme (P0046113). The sponsor or funding organization did not participate in the design or implementation of this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Thanks to all the peer reviewers for their opinions and suggestions.

Abbreviations

FAF	Fundus autofluorescence
RPE	Retinal pigment epithelium
DL	Deep learning
CF	Color fundus
AMD	Age-related macular degeneration
GAN	Generative adversarial network
AUC	Area under the curve
MS-SSIM	Multi-scale structural similarity index
RMHAS	Retina-based microvascular health assessment system
RANSAC	Random sample consensus
SSIM	Structural similarity measures
FID	Fréchet inception distance
MAE	Mean absolute error
PSNR	Peak signal-to-noise ratio

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aopr.2023.11.001.

References

1. Mitchell P, Liew G, Gopinath B, et al. Age-related macular degeneration. *Lancet*. 2018;392(10153):1147–1159. https://doi.org/10.1016/S0140-6736(18)31550-2.
2. Wong WL, Su X, Li X, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Global Health*. 2014;2(2):e106–e116. https://doi.org/10.1016/S2214-109X(13)70145-1.
3. Ferris FL, Wilkinson CP, Bird A, et al. Clinical classification of age-related macular degeneration. *Ophthalmology*. 2013;120(4):844–851. https://doi.org/10.1016/j.ophtha.2012.10.036.
4. Thomas CJ, Mirza RG, Gill MK. Age-related macular degeneration. *Med Clin North Am*. 2021;105(3):473–491. https://doi.org/10.1016/j.mcna.2021.01.003.
5. Tsang SH, Sharma T. Fundus autofluorescence. In: Tsang SH, Sharma T, eds. *Atlas Of Inherited Retinal Diseases. Vol 1085. Advances in Experimental Medicine and Biology*. Springer International Publishing; 2018:15–16. https://doi.org/10.1007/978-3-319-95046-4_4.
6. Ly A, Nivison-Smith L, Assaad N, et al. Fundus autofluorescence in age-related macular degeneration. *Optom Vis Sci*. 2017;94(2):246–259. https://doi.org/10.1097/OPX.0000000000000997.
7. Midena E, Vujosevic S, Convento E, et al. Microperimetry and fundus autofluorescence in patients with early age-related macular degeneration. *Br J Ophthalmol*. 2007;91(11):1499–1503. https://doi.org/10.1136/bjo.2007.119685.
8. Bindewald A, Bird AC, Dandekar SS, et al. Classification of fundus autofluorescence patterns in early age-related macular disease. *Investig Opthalmol Vis Sci*. 2005;46(9):3309. https://doi.org/10.1167/iovs.04-0430.
9. Tavakkoli A, Kamran SA, Hossain KF, et al. A novel deep learning conditional generative adversarial network for producing angiography images from retinal fundus photographs. *Sci Rep*. 2020;10(1):21580. https://doi.org/10.1038/s41598-020-78696-2.
10. Shi D, Zhang W, He S, et al. Translation of color fundus photography into fluorescein angiography using deep learning for enhanced diabetic retinopathy screening. *Ophthalmol Sci*. 2023;3(4):100401. https://doi.org/10.1016/j.xops.2023.100401.
11. You A, Kim JK, Ryu IH, et al. Application of generative adversarial networks (GAN) for ophthalmology image domains: a survey. *Eye Vis*. 2022;9(1):6. https://doi.org/10.1186/s40662-022-00277-3.
12. Wu M, Cai X, Chen Q, et al. Geographic atrophy segmentation in SD-OCT images using synthesized fundus autofluorescence imaging. *Comput Methods Progr Biomed*. 2019;182:105101. https://doi.org/10.1016/j.cmpb.2019.105101.
13. Shi D, Lin Z, Wang W, et al. A deep learning system for fully automated retinal vessel measurement in high throughput image analysis. *Front Cardiovasc Med*. 2022;9:823436. https://doi.org/10.3389/fcvm.2022.823436.
14. Shi D, He S, Yang J, et al. One-shot retinal artery and vein segmentation via cross-modality pretraining. *Ophthalmol Sci*. 2024;4(2):100363. https://doi.org/10.1016/j.xops.2023.100363.
15. Alcantarilla PF, Nuevo J, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: *British Machine Vision Conference 2013*. British Machine Vision Association; 2013.
16. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM*. 1981;24(6):381–395. https://doi.org/10.1145/358669.358692.
17. Wang T.C., Liu M.Y., Zhu J.Y., et al. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. 2017. Published online. doi:10.48550/ARXIV.1711.11585.
18. Abrahamyan L, Truong AM, Philips W, et al. *Gradient Variance Loss for Structure-Enhanced Image Super-resolution*. 2022. https://doi.org/10.48550/ARXIV.2202.00997. Published online.

19. Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600–612. https://doi.org/10.1109/TIP.2003.819861.
20. Abdel-Salam Nasr M, AlRahmawy MF, Tolba AS. Multi-scale structural similarity index for motion detection. *J King Saud Univ - Comput Inf Sci*. 2017;29(3):399–409. https://doi.org/10.1016/j.jksuci.2016.02.004.
21. Heusel M, Ramsauer H, Unterthiner T, et al. *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*. 2017. https://doi.org/10.48550/ARXIV.1706.08500. Published online.
22. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res*. 2005;30(1):79–82.
23. Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM. In: *2010 20th International Conference on Pattern Recognition*. 2010:2366–2369. https://doi.org/10.1109/ICPR.2010.579.
24. Liu Z, Lin Y, Cao Y, et al. *Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows*. 2021. https://doi.org/10.48550/ARXIV.2103.14030. Published online.
25. Pole C, Ameri H. Fundus autofluorescence and clinical applications. *J Ophthalmic Vis Res*. 2021;29. https://doi.org/10.18502/jovr.v16i3.9439. Published online July.
26. Khanifar AA, Lederer DE, Ghodasra JH, et al. Comparison of color fundus photographs and fundus autofluorescence images in measuring geographic atrophy area. *Retina*. 2012;32(9):1884–1891. https://doi.org/10.1097/IAE.0b013e3182509778.
27. Keenan TDL, Chen Q, Peng Y, et al. Deep learning automated detection of reticular pseudodrusen from fundus autofluorescence images or color fundus photographs in AREDS2. *Ophthalmology*. 2020;127(12):1674–1687. https://doi.org/10.1016/j.ophtha.2020.05.036.
28. Miere A, Capuano V, Kessler A, et al. Deep learning-based classification of retinal atrophy using fundus autofluorescence imaging. *Comput Biol Med*. 2021;130:104198. https://doi.org/10.1016/j.compbiomed.2020.104198.
29. Li W, Kong W, Chen Y, et al. *Generating Fundus Fluorescence Angiography Images from Structure Fundus Images Using Generative Adversarial Networks*. ArXiv Prepr ArXiv200610216; 2020. Published online.
30. Pichi F, Abboud EB, Ghazi NG, et al. Fundus autofluorescence imaging in hereditary retinal diseases. *Acta Ophthalmol*. 2018;96(5):e549–e561. https://doi.org/10.1111/aos.13602.
31. Holz FG, Bindewald-Wittich A, Fleckenstein M, et al. Progression of geographic atrophy and impact of fundus autofluorescence patterns in age-related macular degeneration. *Am J Ophthalmol*. 2007;143(3;463:472.e2. https://doi.org/10.1016/j.ajo.2006.11.041.