



Deep learning-driven automated quality assessment of ultra-widefield optical coherence tomography angiography images for diabetic retinopathy

Yixiao Jin^{1,2} · Fu Gui³ · Minghao Chen^{1,2} · Xiang Chen⁴ · Haoxuan Li⁵ · Jingfa Zhang^{1,2} 

Accepted: 23 March 2024
© The Author(s) 2024

Abstract

Image quality assessment (IQA) of fundus images constitutes a foundational step in automated disease analysis. This process is pivotal in supporting the automation of screening, diagnosis, follow-up, and related academic research for diabetic retinopathy (DR). This study introduced a deep learning-based approach for IQA of ultra-widefield optical coherence tomography angiography (UW-OCTA) images of patients with DR. Given the novelty of ultra-widefield technology, its limited prevalence, the high costs associated with equipment and operational training, and concerns regarding ethics and patient privacy, UW-OCTA datasets are notably scarce. To address this, we initially pre-train a vision transformer (ViT) model on a dataset comprising 6 mm × 6 mm OCTA images, enabling the model to acquire a fundamental understanding of OCTA image characteristics and quality indicators. Subsequent fine-tuning on 12 mm × 12 mm UW-OCTA images aims to enhance accuracy in quality assessment. This transfer learning strategy leverages the generic features learned during pre-training and adjusts the model to evaluate UW-OCTA image quality effectively. Experimental results demonstrate that our proposed method achieves superior performance compared to ResNet18, ResNet34, and ResNet50, with an AUC of 0.9026 and a Kappa value of 0.7310. Additionally, ablation studies, including the omission of pre-training on 6 mm × 6 mm OCTA images and the substitution of the backbone network with the ViT base version, resulted in varying degrees of decline in AUC and Kappa values, confirming the efficacy of each module within our methodology.

Keywords Deep learning · Image quality assessment (IQA) · Transfer learning · Diabetic retinopathy · Optical coherence tomography angiography (OCTA) · Ultra-widefield optical coherence tomography angiography (UW-OCTA)

Yixiao Jin and Fu Gui: Co-first authors contributed equally to this study.

✉ Haoxuan Li
2311519002@sus.edu.cn

✉ Jingfa Zhang
13917311571@139.com

¹ Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, 100 Hai Ning Road, Shanghai 200080, China

² National Clinical Research Center for Eye Diseases, Shanghai Clinical Research Center for Eye Diseases, Shanghai Key Clinical Specialty, Shanghai Key Laboratory of Ocular Fundus Diseases, Shanghai Engineering Center for Visual Science and Photomedicine, Shanghai Engineering Center for Precise Diagnosis and Treatment of Eye Diseases, Shanghai, China

³ Department of Ophthalmology, The Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China

1 Introduction

Diabetic retinopathy (DR) is a global leading cause of irreversible blindness, with the patients' number projected to increase from 103 million in 2020 to 161 million by 2040 [1]. Regular screening and timely treatment are essential for DR [2]. However, DR screening largely depends on the expertise of ophthalmologists, where a lack of sufficient training could lead to misdiagnoses and low accuracy. Moreover, the public health economic burden is substantial, particularly in resource-limited areas. Thus, the implementation of an efficient artificial intelligence (AI) system is invaluable for

⁴ Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

⁵ Sports Engineering, School Exercise and Health, Shanghai University of Sport, 650 Qingyuan Ring Road, Yangpu District, Shanghai 200438, China

aiding accurate DR diagnosis and alleviating the workload of ophthalmologists [3, 4].

Diagnosis of DR relies on imaging manifestations such as microaneurysms (MA), intraretinal microvascular abnormalities (IRMA), and neovascularization (NV) [5]. Ultra-widefield optical coherence tomography angiography (UW-OCTA) serves as a commonly used non-invasive technique that provides a three-dimensional, intuitive representation of pathological changes in all retinal layers [6]. UW-OCTA provides a broader range of peripheral retinal areas than traditional OCTA. UW-OCTA, generally capable of providing a field of view (FOV) up to 100 degrees or more, offers a significantly broader range of peripheral retinal areas compared to traditional OCTA, which typically provides a FOV of 30 to 50 degrees. This broader perspective is instrumental in the early detection of lesions, such as MA, facilitating timely treatment and intervention [7]. Such early diagnostic capability is crucial for preserving patients' vision, highlighting the significant advantage of UW-OCTA in managing and mitigating the progression of retinal diseases. Moreover, OCTA enables reproducible measurements of retinal pathological parameters and the evaluation of treatment efficacy and follow-up through quantifiable, intuitive, and repeatable values [8, 9]. However, the practicality of DR screening systems is hindered by low-quality fundus images due to certain problems, such as patient non-cooperation, operator skill, or equipment-related factors, which may affect the numerical values of OCTA-generated parameters [10, 11]. Such images, marred by significant artifacts and poorly lit areas, pose challenges to subsequent AI diagnosis and staging tasks, impacting model performance. Therefore, it is necessary to filter out poor quality images before conducting any DR analysis, such as lesion segmentation and DR grading.

The quality assessment for medical images is complex. Unlike natural images, the quality grading capability of medical images is not solely dependent on pixels, signal, noise, or distortion, but also on the specific visibility and interpretability of clinically relevant features. Even images with acceptable signal strength may still present challenges in assessing other OCTA image quality issues, such as off-centering, out of registration, signal loss, motion artifacts, and projection artifacts [11–13]. Image quality assessment (IQA) requires trained operators and interpreters with ophthalmic clinical knowledge, a significant challenge due to clinic staffing and training time constraints. Moreover, manual evaluation of each OCTA scan by human assessors is impractical, time-consuming, and tedious within the busy clinical workflow [14]. Additionally, subjective differences may arise even among experienced ophthalmologists. Furthermore, judgment of human assessors on whether the overall image quality is sufficient for disease detection or

needs further analysis is crucial for distinguishing medical image grading. For instance, despite acceptable overall image quality or satisfactory noise levels in non-vascular areas, if the vascular quality, in terms of contrast or continuity, is poor and insufficient for the clear identification of MA, such an image would be deemed of poor quality, failing to meet the clinical diagnostic requirements. Conversely, images where vascular imaging appears blurred or poor due to retinal disease states, such as edema or exudation, yet the lesion manifestations are recognizable by clinical physicians, are considered clinically usable. The judgment of human assessors in manually evaluating image quality forms the foundation of training algorithmic models that can automatically assess large image datasets with less human effort and lower costs. This is key for automated tasks like disease diagnosis, grading, and lesion segmentation.

Following the introduction of optical coherence tomography (OCT) equipment, the advent of the split-spectrum amplitude-decorrelation angiography (SSADA) in 2012 marked a significant milestone [15]. Optovue, Inc. swiftly integrated OCTA into their commercial SD-OCT platform as a research tool for the broader ophthalmic community [16]. Subsequently, OCTA technology matured and found its application in clinical practice [17]. UW-OCTA, a later development based on OCTA, is relatively new and has only begun to be utilized clinically in recent years, with its widespread adoption still emerging. Additionally, the high cost of ultra-widefield equipment and the significant operation and training expenses have limited its use, particularly in resource-constrained regions. The need for specialized operational skills and experience to acquire high-quality UW-OCTA images, coupled with the novelty of the technology, means that comprehensive training for relevant personnel may not yet be widespread, potentially affecting the efficiency and quality of data collection. The ethical and privacy standards required for collecting and sharing medical imaging data necessitate time to establish appropriate data sharing mechanisms for emerging imaging technologies. These factors contribute to the scarcity of UW-OCTA datasets compared to OCTA datasets.

To address this, we developed an algorithm that utilizes a standard 6 mm × 6 mm OCTA dataset for model pre-training, followed by fine-tuning with a 12 mm × 12 mm ultra-widefield OCTA dataset, ultimately applying it for the quality assessment of UW-OCTA images. Therefore, this research aims to develop a deep learning system (DLS) for the quality assessment of UW-OCTA images, which enhancing the accuracy and efficiency of IQA and improving the precision of human judgment in the screening, diagnosis, and monitoring of DR, leveraging advanced image analysis techniques.

2 Related work

2.1 The application of deep learning in ophthalmology

In recent years, the application of deep learning in ophthalmology has been increasingly prevalent [18]. The study by De Fauw et al. demonstrated significant advancements in the application of deep learning for the diagnosis and referral of retinal diseases. Their system, trained on OCT datasets, autonomously analyzed and diagnosed various retinal diseases, including age-related macular degeneration (AMD) and DR, with remarkable accuracy. The model is capable of prioritizing patients for referral based on the severity and urgency of their condition, performing comparably to or even surpassing human experts [19]. Dai and colleagues developed a DLS named DeepDR, trained on 466,247 fundus images from 121,342 diabetic patients, for real-time IQA, lesion detection, and grading. It can detect DR lesions such as MA, cotton wool spots, hard exudates, and hemorrhages [20]. In glaucoma, Berchuck et al. developed a DLS to improve the estimation of progression rates and predict future patterns of visual field loss [21]. Li and colleagues developed a convenient DLS based on a smartphone application to detect changes in the visual field for glaucoma [22]. Yoo et al. developed a method using fundus photographs to detect anterior chamber depth, a critical risk factor for angle closure glaucoma, thereby screening for the condition [23]. In AMD, Yim et al. used deep learning to predict the progression of the second eye in patients with wet AMD. The system can predict conversion to wet AMD within a clinically viable 6-month window, outperforming five out of six experts and showcasing the potential of using AI to predict disease progression [24]. Hwang and colleagues developed an AI-based system for diagnosing AMD based on OCT images, achieving detection accuracy comparable to ophthalmologist, and providing treatment recommendations on par with experts. Furthermore, an operational cloud computing website was developed based on this AI platform, allowing patients to upload OCT images to verify if they have AMD and require treatment. The use of AI-based cloud services represents a genuine solution for medical imaging diagnosis and telemedicine [25].

2.2 Transfer learning

Transfer learning is an effective strategy when the dataset for the target task is too small to train a model from scratch [26]. Transfer learning leverages the knowledge (features, weights, and biases) a model has learned from a large and comprehensive dataset to enhance its performance on another, often smaller dataset [27]. This approach has become increasingly popular in various domains, including natural language processing, computer vision, and medical imaging, due to its

ability to improve model performance with minimal computational resources and data requirements. In recent years, the development and improvement in transfer learning algorithms have been significant. For instance, in computer vision, pre-trained models like VGGNet, ResNet, and Inception have been widely adopted for tasks such as image classification and object detection by fine-tuning the models on specific datasets [28]. In natural language processing, models like bidirectional encoder representations from transformers (BERT) [29] and generative pre-trained transformer (GPT) have revolutionized the field by providing a robust foundation for tasks like text classification, sentiment analysis, and question-answering systems. The advancements in transfer learning algorithms have also made a substantial impact on medical imaging, where models pre-trained on general images are fine-tuned to detect and diagnose diseases from medical scans with high accuracy [30–32]. This approach has proved particularly beneficial in areas with limited labeled medical datasets.

2.3 ViT in image analysis

The ViT has emerged as a groundbreaking architecture in computer vision, marking a significant departure from the convolutional neural networks (CNNs) that have dominated the landscape for the past decade. Introduced by Dosovitskiy et al., ViT applies the transformer model, originally designed for natural language processing tasks, to image analysis by treating images as sequences of patches [33]. This approach allows ViT to capture global dependencies within an image, a feat that traditional CNNs achieve through extensive depth or complex architectures.

The benefits of ViT are manifold. Firstly, it demonstrates an exceptional ability to scale with increased data and computational resources, often surpassing the performance of state-of-the-art CNNs on benchmark datasets. Secondly, ViT offers a more flexible architecture that is inherently capable of handling various input sizes, making it adaptable to a wide range of vision tasks without significant modifications [34].

Recent works have leveraged the ViT architecture for large-scale image recognition tasks, showcasing its potential as a foundation model in the realm of visual data. For instance, the application of ViT in models like BigGAN and DALL-E underscores its versatility and efficiency in generating high-fidelity images and understanding complex visual concepts [35, 36]. Furthermore, the integration of ViT into foundation models has set new benchmarks in tasks such as image classification, object detection, and semantic segmentation, highlighting its robustness and scalability.

Fig. 1 Overview of the methodology. Initially, the ViT model is initialized with ImageNet-derived weights, followed by the pre-training with the 6 mm × 6 mm OCTA images. Subsequently, the ViT model is finetuned on 12 mm × 12 mm UW-OCTA images and output the image quality levels

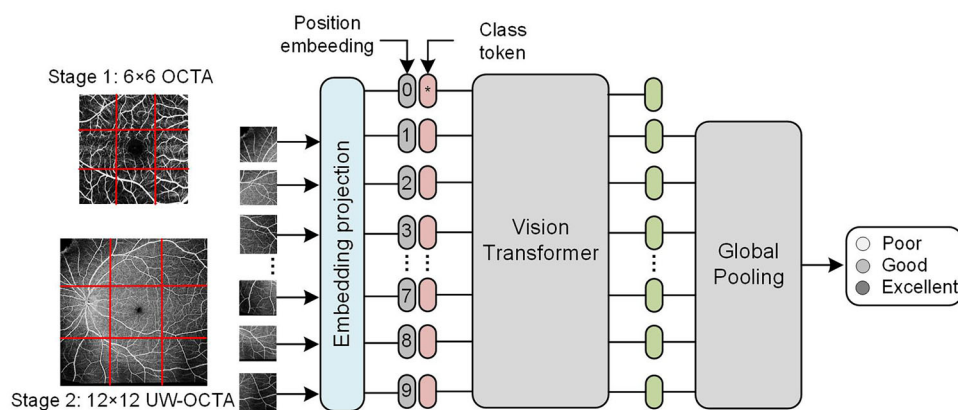
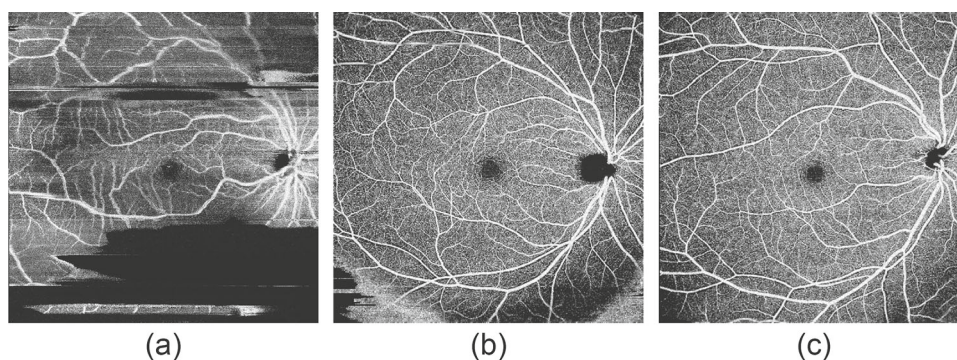


Fig. 2 Representative images for distinct quality labels:

a illustrates an image categorized under label 0, denoting poor quality. **b** showcases an image classified as label 1, signifying good quality. **c** displays an image attributed to label 2, indicative of excellent quality



3 Methods

3.1 Overview

Our methodology initiates with the pre-training phase of a ViT model on a dataset consisting of 6 mm × 6 mm OCTA images. This preliminary stage allows the model to acquire a foundational comprehension of OCTA image characteristics and quality indicators. Subsequently, we employ a fine-tuning phase on a higher field of view dataset of 12 mm × 12 mm UW-OCTA images, aimed at enhancing the model's accuracy in quality assessment. This transfer learning strategy leverages the generic features learned during pre-training and adapts the model to perform the specialized task of UW-OCTA image quality assessment. An illustrative overview of this methodology is presented in Fig. 1.

3.2 Data augmentation

In the data augmentation, we employed a series of transformations to enrich our dataset and enhance the robustness of our model against various imaging conditions. These transformations include random horizontal and vertical flips to simulate different orientations of the images, introducing variability in the dataset. Color jittering is also utilized to adjust the brightness, contrast, saturation, and hue of the images, further augmenting the diversity of the dataset. To

introduce a range of rotational perspectives, we implement random rotations with a degree range of -180 to 180. Subsequently, all images are normalized using mean values and standard deviations of ImageNet dataset, aligning with common practice and ensuring consistency in image input to the model. These data augmentation steps are instrumental in developing a model that is adaptable and performs consistently across a UW-OCTA image presentations.

3.3 Classification architecture

In selecting the architecture for our model, we considered the strengths and limitations of two prominent architectures: Residual Networks (ResNet) and ViT. ResNet is renowned for its deep architecture that effectively addresses the vanishing gradient problem using skip connections. These connections allow the network to learn identity functions, ensuring that deeper layers can at least perform as well as shallower ones, which prevents performance degradation with increased depth. The ability of ResNet to leverage deep convolutional layers makes it adept at capturing hierarchical features in images [28]. However, its reliance on convolutional operations can limit its ability to learn the global image features within an image, which may be crucial for understanding complex scenes or contextual information.

ViT utilizes the transformative capabilities of transformers, a paradigm initially conceived for natural language

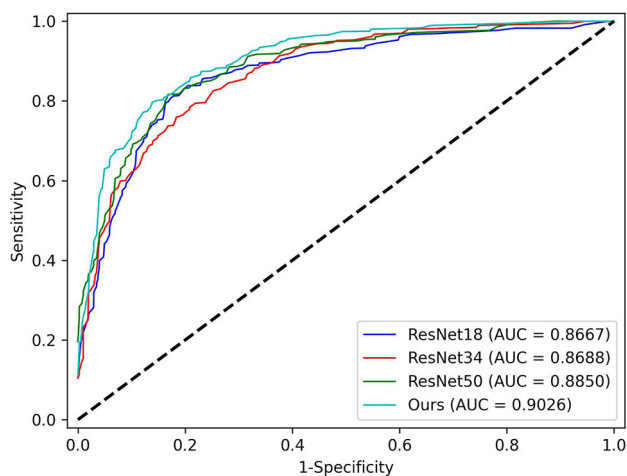


Fig. 3 ROC curves for various methods applied to image quality assessment of UW-OCTA images, including ResNet-18, ResNet-34, ResNet-50, and our model

Table 1 Performance comparison of image quality assessment in the test dataset

Methods	AUC	Kappa
ResNet18	0.8667	0.6624
ResNet34	0.8688	0.6668
ResNet50	0.8850	0.7048
Ours	0.9026	0.7310

processing, and adeptly repurposes them within the visual domain. By segmenting an image into discrete patches and subsequently processing these patches as sequential entities analogous to words in textual analysis, ViT introduces a novel methodology for image interpretation. This segmentation and sequential processing facilitate the capacity to assimilate global features dispersed throughout the entirety of the image, thereby rendering ViT exceptionally proficient for tasks necessitating a comprehensive understanding of the image context. Central to ViT's architecture is the incorporation of an attention mechanism, which strategically allocates focus to the most salient segments of the input. Through this innovative adaptation of transformers to the visual sphere, ViT emerges as a potent tool, offering nuanced insights and enhanced analytical capabilities for image-based assessments [37, 38].

Considering these aspects, we chose the ViT as our classification model due to its superior capability in capturing global image contexts and features, which is critical for assessing the quality of OCTA images that have diverse and complex retinal structures.

3.4 Training strategy

Transfer learning represents a formidable strategy within the domain of machine learning, wherein a model devised for a primary task is repurposed as the foundational model for a secondary task. This methodology proves exceptionally advantageous in contexts where the dataset pertinent to the target task is relatively diminutive, yet related, more extensive datasets exist for the initial task. Motivated by this paradigm, we employ transfer learning to surmount the challenge posed by the limited availability of $12 \text{ mm} \times 12 \text{ mm}$ UW-OCTA images. Despite the scarcity of datasets for $12 \text{ mm} \times 12 \text{ mm}$ UW-OCTA images, there exists a relatively ample collection of traditional $6 \text{ mm} \times 6 \text{ mm}$ OCTA images. Consequently, our approach entails initially pre-training our model on the abundant $6 \text{ mm} \times 6 \text{ mm}$ OCTA images, utilizing ImageNet weights for model initialization to harness features learned from a broad spectrum of natural images. This pre-training phase equips the model with the capability to discern general features and patterns pertinent to OCTA images. The subsequent phase involves fine-tuning the model with the rarer $12 \text{ mm} \times 12 \text{ mm}$ UW-OCTA images, with a specific focus on enhancing the proficiency of this model in assessing the quality of OCTA images across a wider field of view.

By implementing this two-step training regimen, we effectively utilize OCTA data across different fields of view, thereby augmenting efficacy of this model in appraising the quality of UW-OCTA images. This methodological framework not only optimizes the utilization of available data but also significantly enhances the precision of quality assessments for UW-OCTA images.

4 Experiments

4.1 Dataset

This study utilized a dataset from the Diabetic Retinopathy Analysis Challenge (DRAC2022) website [39], captured using the VG200D ultra-wide swept source OCTA (UW SS-OCTA) device, manufactured by SVision Imaging, Ltd. This dataset encompasses a total of 1103 images, segmented into two subsets: 665 images designated for training and 438 for testing purposes. Each image within the training subset was annotated with a corresponding label, delineating the image quality into one of three categorically distinct levels: label 0 denotes poor quality, label 1 signifies good quality, and label 2 is indicative of excellent quality. The representative images are shown in Fig. 2.

During the training, we divide 20% of the images from the original training set to be used as the validation set, while the

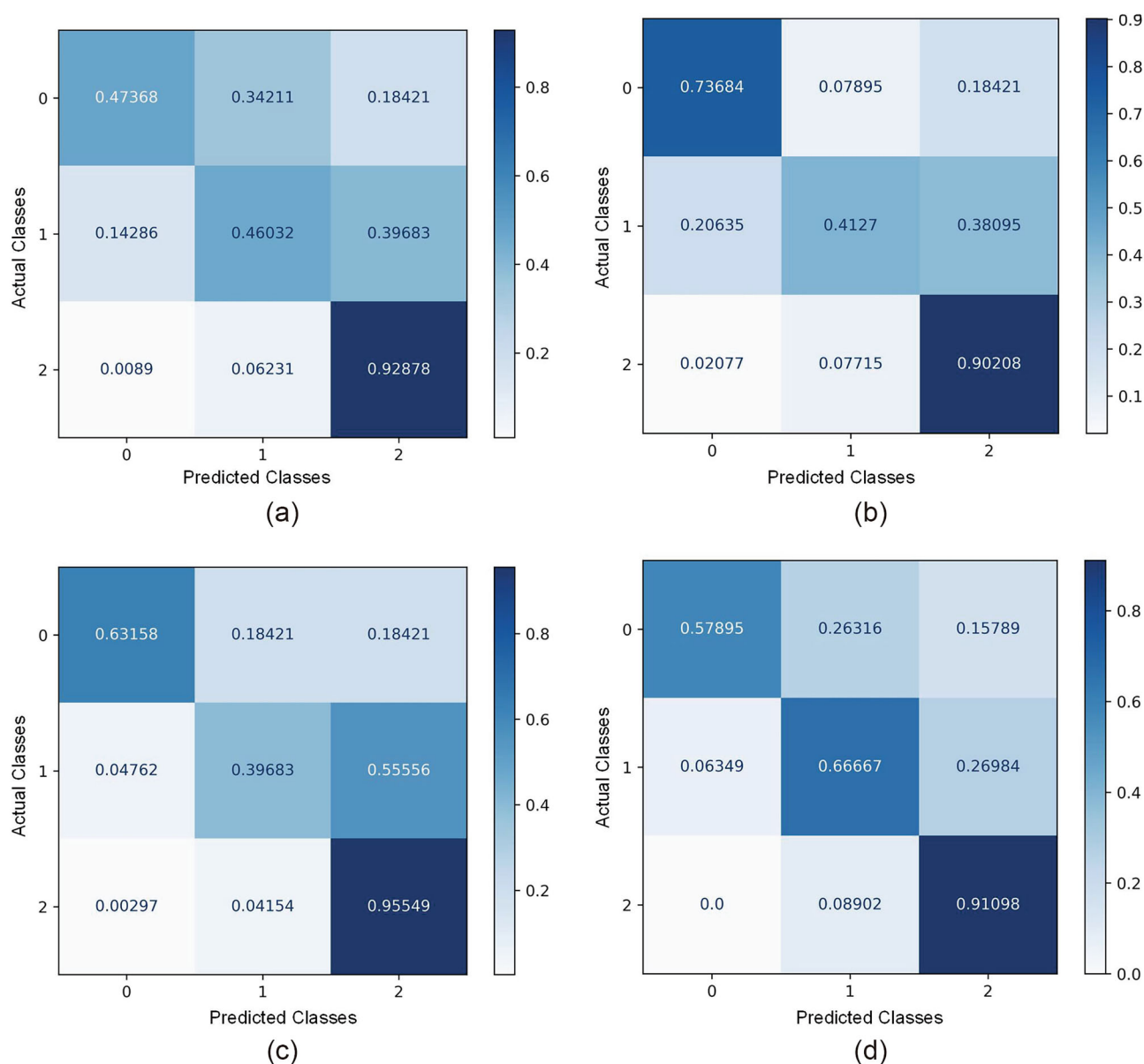


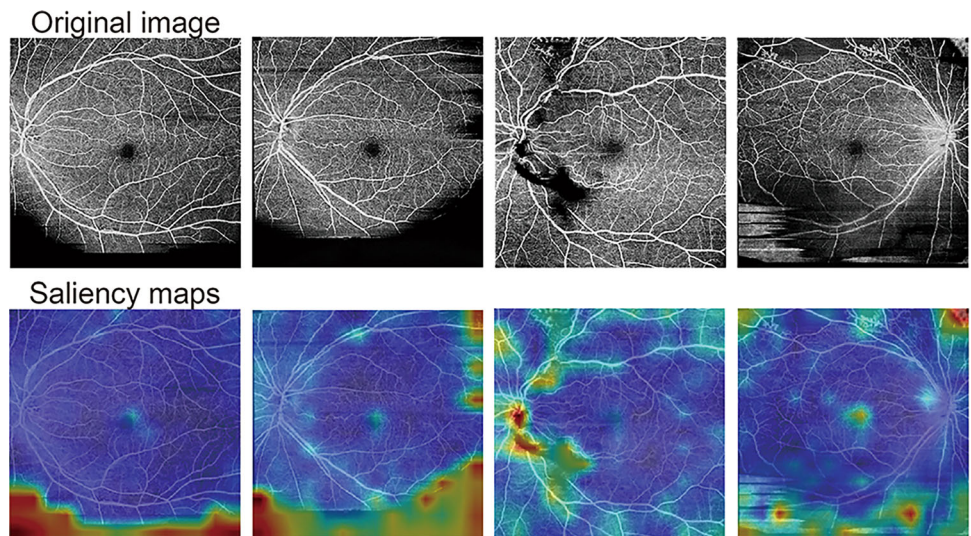
Fig. 4 Confusion matrix (CM) of each method evaluated in the context of UW-OCTA image quality assessment. **a** delineates the CM for the original ResNet18 model. **b** depicts the CM for ResNet34 model, illustrating its performance metrics. **c** displays the CM for the ResNet50

model. **d** elucidates the CM for our proposed method, showcasing enhanced IQA precision

remaining images serve as the training set. The model performance is evaluated on the test set. For the pre-training images of $6 \text{ mm} \times 6 \text{ mm}$ images, we collected a total of 278 images from the Shanghai General Hospital, using an UW SS-OCTA device, manufactured by SVision Imaging, Ltd. The inclusion criteria were patients diagnosed with diabetes mellitus, possessing OCTA images, regardless of the imaging quality. The exclusion criteria included patients who declined to participate in the study or were non-cooperative during the

examination process. This research adhered to the Declaration of Helsinki's principles and underwent ethical review by the committee of Shanghai General Hospital, affiliated with Shanghai Jiao Tong University School of Medicine (ethical approval number: 2023–263). These images are divided into training set, validation set, and test set with a ratio of 6:2:2. All images were labeled aligned with the standards set by the Diabetic Retinopathy Analysis Challenge (DRAC2022) [39]. An image is considered “poor quality” if it is not sufficient for analysis, with high level of artifacts and blurred

Fig. 5 Representative examples of the saliency maps and corresponding original UW-OCTA images. Heatmaps highlight the areas that contribute to IQA of this model, with red color indicating high contribution



vascular details, and there are 10 poor quality images (label 0), 26 good quality images (label 1), and 242 excellent quality images (label 2).

4.2 Implemental details

In this investigation, the ViT large variant was selected, with the patch size configured to 16. To accommodate input requirements of the network, images were resized to dimensions of 224×224 pixels. The optimization of the network was facilitated through the employment of the Adam optimizer, with an initial learning rate meticulously set at 0.005. To further refine the training process, a multi-step learning rate adjustment strategy was implemented, characterized by predetermined milestones at the 20th and 40th epochs, accompanied by a gamma adjustment factor of 0.1. The training regimen spanned 50 epochs, maintaining a batch size of 4, and utilized the cross-entropy loss function as the criterion for network optimization. Model performance was rigorously evaluated against the validation set upon the completion of each epoch. The epoch demonstrating optimal performance on the validation set was subsequently designated as the final model configuration. This model was then applied to assess performance metrics on the test set, ensuring a comprehensive evaluation of its diagnostic capabilities.

4.3 Evaluation metrics

Within the scope of our methodological approach, two critical evaluation metrics were employed to assess model performance: the area under the receiver operating characteristic curve (AUC) and the quadratic weighted Kappa (QWK). The AUC metric serves as a comprehensive measure of the model to discriminate between classes across all possible thresholds. It is calculated as the area under the curve

plotted with the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. Mathematically, the AUC can be expressed as:

$$AUC = \int_{x=0}^1 TPR(x) dFPR(x) \tag{1}$$

where TPR is the true positive rate and FPR is the false positive rate.

On the other hand, QWK is a more sophisticated statistical measure that evaluates the agreement between two raters who each classify N items into K mutually exclusive categories. Unlike simple agreement measures, the QWK accounts for the possibility of agreement occurring by chance and introduces a weighting scheme to penalize disagreements proportionally to the squared distance between categories. The QWK is calculated using the formula:

$$QWK = 1 - \frac{\sum_{i,j} w_{ij} O_{ij}}{\sum_{i,j} w_{ij} E_{ij}} \tag{2}$$

where O_{ij} is the observed count of items in category i predicted to be in category j , E_{ij} is the expected count of items in category i predicted to be in category j under the assumption of chance agreement, and W_{ij} is the weight assigned to the disagreement between categories i and j , typically calculated as $(i - j)^2 / (K - 1)^2$.

These metrics, AUC and QWK, collectively provide a robust framework for evaluating the performance of predictive models, offering insights into both the discriminative power of the model and the consistency of its predictions with respect to a standard or another rater, respectively.

4.4 Quantitative results

To substantiate the efficacy of our proposed methodology in the domain of image quality assessment for UW-OCTA images, we conducted a comparative analysis against established benchmarks, including ResNet18, ResNet34, and ResNet50. The quantitative outcomes of this evaluation are delineated in Table 1. It is evident from the analysis that our approach outperforms the comparative models in terms of the AUC and Kappa metrics, registering improvements of 1.76% and 2.62% over the second-best performing method for AUC and Kappa, respectively. The Receiver Operating Characteristic (ROC) curves, illustrating the diagnostic ability of our method alongside the baseline models, are presented in Fig. 3. Additionally, the classification accuracy and misclassification patterns are encapsulated within the confusion matrix (CM), depicted in Fig. 4.

4.5 Explainability analysis

To better understand how this DLS performs quality assessment of UW-OCTA images, we conducted a heatmap analysis to gain insight into regions of the retinal fundus image that may affect DLS predictions. Based on the technique proposed by Chefer et al., we employ Layer-wise Relevance Propagation (LRP)-based correlations to compute scores for each attention head within every layer of the transformer model [40]. The method combines these scores across the attention graph, using both relevance and gradient data to progressively eliminate negative impacts. This process leads to a visualization that is specific to each class for self-attention models, offering a fresh perspective on the model's interpretability and reliability. Figure 5 shows some representative examples of original UW-OCTA images and the corresponding heatmap maps. In these images, the red regions represent areas of high contribution. The visualization results suggest that our DLS can discriminate image quality based on signal deficiencies and artifacts.

4.6 Ablation study

An ablation study was conducted to ascertain the contribution of individual components within our proposed methodology. Initially, the pre-training phase on 6 mm × 6 mm OCTA images was omitted, restricting the model training exclusively to 12 mm × 12 mm UW-OCTA images. This modification led to a decrement of 1.11% in the AUC and 2.39% in the Kappa metric, underscoring the significance of the pre-training step. Subsequently, the architectural foundation was altered by substituting the original network with the ViT basic model. This adjustment resulted in a reduction of 2.42% in AUC and 2.27% in Kappa, as detailed in Table 2.

Table 2 Ablation analysis of the proposed methodology for image quality assessment utilizing UW-OCTA images

Method	AUPR	Kappa
w/o pre-training	0.8915	0.7071
w/ ViT-B	0.8964	0.7083
Ours	0.9026	0.7310

w/o, without; w/, with; ViT-B, vision transformer base

These findings unequivocally demonstrate that each component integrated into our framework plays a pivotal role in enhancing the overall performance of image quality assessment, thereby validating the efficacy of our comprehensive approach.

5 Conclusion

This study introduces a robust DLS that significantly advances the automated IQA of UW-OCTA images, particularly for DR patients. The complexity is inherent in medical image quality assessment, where evaluation criteria extend beyond mere pixel quality to encompass the visibility and interpretability of clinically relevant features. The manual evaluation of each fundus scan, especially in clinics lacking experienced personnel, is both inefficient and impractical. Our methodology, leveraging a ViT model pre-trained on standard 6 mm × 6 mm OCTA images and fine-tuned on 12 mm × 12 mm UW-OCTA scans, addresses these challenges by enhancing the accuracy and efficiency of IQA processes.

Our approach, utilizing transfer learning and data augmentation strategies, effectively navigates the limitations imposed by the scarcity of UW-OCTA datasets—a scarcity driven by the novelty of ultra-widefield technology, associated high costs, and ethical considerations. The experimental results, showcasing superior performance over conventional models with an AUC of 0.9026 and a Kappa value of 0.7310 (in Table 1 and Fig. 3), alongside ablation studies, underscore the critical importance of each component in our framework.

Recently, some DLS have been developed using OCTA images, including the diagnosis of diabetic macular edema and choroidal neovascularization, disease progression and vision prediction in DR patients, retinal vessel segmentation, and retinal layering [41–46]. These advancements underscore the potential of deep learning to reduce the interpretative costs associated with fundus image diseases. Consequently, the necessity for IQA to pre-emptively filter out unusable images for enhanced accuracy is evident. However, the manual filtration of poor-quality images currently demands significant human, material, and financial

resources, with research on UW-OCTA remaining scarce. Therefore, our DLS holds the potential for integration with other systems to further disease detection. A significant future application of our DLS is its embedded installation in UW-OCTA machines, enabling operators to be notified and immediately reacquire images when the device classifies an image as of poor quality. This integration would substantially alleviate the manual burden of image quality control and efficiently provide higher-quality images for further analysis, marking a significant stride toward automating and enhancing the precision of medical imaging in the diagnosis and management of retinal diseases.

The broader implications of our research extend well into the field of medical imaging, offering a scalable and efficient solution for the automated quality assessment of fundus images. This advancement not only facilitates early detection and intervention in diabetic retinopathy but also potentially improves patient outcomes by ensuring high-quality image analysis for accurate diagnosis and grading. As the field of ophthalmology continues to develop DLS for various OCTA-based diagnostics, our study contributes significantly to reducing the manual burden of image quality control and enhancing the reliability of disease detection and progression monitoring through improved image quality assessment.

Author contributions YXJ and FG wrote the main manuscript text; YXJ, MHC, and JFZ collected OCTA images from the hospital; YXJ, XC, and HXL designed the algorithmic framework and developed the network; YXJ prepared figure 1 and tables; XC prepared figures 2-5. JFZ and FG revised and polished the manuscript. All authors reviewed the manuscript.

Data availability Data and material are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Teo, Z.L., Tham, Y.C., Yu, M., et al.: Global Prevalence of diabetic retinopathy and projection of burden through 2045: Systematic review and meta-analysis. *Ophthalmology* **128**, 1580–1591 (2021)
2. Lee, R., Wong, T.Y., Sabanayagam, C.: Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. *Eye Vis (Lond)* **2**, 17 (2015)
3. Wong, T.Y., Bressler, N.M.: Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *JAMA* **316**, 2366–2367 (2016)
4. Wong Tien, Y., Sabanayagam, C.: Strategies to tackle the global burden of diabetic retinopathy: from epidemiology to artificial intelligence. *Ophthalmologica* **243**, 9–20 (2019)
5. Cheung, N., Mitchell, P., Wong, T.Y.: Diabetic retinopathy. *Lancet* **376**, 124–136 (2010)
6. Niederleithner, M., Sisternes, Ld., Stino, H., et al.: Ultra-widefield OCT angiography. *IEEE Transact. Med. Imag.* **42**, 1009–1020 (2023)
7. Cui, Y., Zhu, Y., Wang, J.C., et al.: Comparison of widefield swept-source optical coherence tomography angiography with ultra-widefield colour fundus photography and fluorescein angiography for detection of lesions in diabetic retinopathy. *Br. J. Ophthalmol.* **105**, 577–581 (2021)
8. Fenner, B.J., Tan, G.S.W., Tan, A.C.S., Yeo, I.Y.S., Wong, T.Y., Cheung, G.C.M.: Identification of imaging features that determine quality and repeatability of retinal capillary plexus density measurements in OCT angiography. *Br. J. Ophthalmol.* **102**, 509–514 (2018)
9. Spaide, R.F., Fujimoto, J.G., Waheed, N.K., Sadda, S.R., Staurenghi, G.: Optical coherence tomography angiography. *Prog. Retin. Eye Res.* **64**, 1–55 (2018)
10. Ghasemi Falavarjani, K., Al-Sheikh, M., Akil, H., Sadda, S.R.: Image artefacts in swept-source optical coherence tomography angiography. *Br. J. Ophthalmol.* **101**, 564–568 (2017)
11. Spaide, R.F., Fujimoto, J.G., Waheed, N.K.: Image artifacts in optical coherence tomography angiography. *Retina* **35**, 2163–2180 (2015)
12. Ran, A.R., Shi, J., Ngai, A., et al.: Artificial intelligence deep learning algorithm for discriminating ungradable optical coherence tomography three-dimensional volumetric optic disc scans. *Neurophotonics* **6**, 041110 (2019)
13. Camino, A., Zhang, M., Gao, S.S., et al.: Evaluation of artifact reduction in optical coherence tomography angiography with real-time tracking and motion correction technology. *Biomed. Opt. Express* **7**, 3905–3915 (2016)
14. Wang, S., Jin, K., Lu, H., Cheng, C., Ye, J., Qian, D.: Human visual system-based fundus image quality assessment of portable fundus camera photographs. *IEEE Trans. Med. Imaging* **35**, 1046–1055 (2016)
15. Jia, Y., Tan, O., Tokayer, J., et al.: Split-spectrum amplitude-decorrelation angiography with optical coherence tomography. *Opt. Express* **20**, 4710–4725 (2012)
16. Kraus, M.F., Liu, J.J., Schottenhamml, J., et al.: Quantitative 3D-OCT motion correction with tilt and illumination correction, robust similarity measure and regularization. *Biomed. Opt. Express* **5**, 2591–2613 (2014)
17. Gao, S.S., Jia, Y., Zhang, M., et al.: Optical Coherence tomography angiography. *Invest. Ophthalmol. Vis. Sci.* **57**, 27–36 (2016)
18. Sheng, B., Guan, Z., Lim, L.L., et al.: Large language models for diabetes care: potentials and prospects. *Sci. Bull. (Beijing)* **69**, 583–588 (2024)
19. De Fauw, J., Ledsam, J.R., Romera-Paredes, B., et al.: Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350 (2018)

20. Dai, L., Wu, L., Li, H., et al.: A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat. Commun.* **12**, 3242 (2021)
21. Berchuck, S.I., Mukherjee, S., Medeiros, F.A.: Estimating rates of progression and predicting future visual fields in glaucoma using a deep variational autoencoder. *Sci. Rep.* **9**, 18113 (2019)
22. Li, F., Song, D., Chen, H., et al.: Development and clinical deployment of a smartphone-based visual field deep learning system for glaucoma detection. *NPJ Digit. Med.* **3**, 123 (2020)
23. Yoo, T.K., Ryu, I.H., Kim, J.K., Lee, I.S., Kim, H.K.: A deep learning approach for detection of shallow anterior chamber depth based on the hidden features of fundus photographs. *Comput. Methods Progr. Biomed.* **219**, 106735 (2022)
24. Yim, J., Chopra, R., Spitz, T., et al.: Predicting conversion to wet age-related macular degeneration using deep learning. *Nat. Med.* **26**, 892–899 (2020)
25. Hwang, D.K., Hsu, C.C., Chang, K.J., et al.: Artificial intelligence-based decision-making for age-related macular degeneration. *Theranostics* **9**, 232–245 (2019)
26. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *ArXiv* 2014;abs/1411.1792.
27. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**, 1345–1359 (2010)
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual learning for image recognition. *IEEE Conf. Comput. Vis. Pattern Recogn. (CVPR)* **2016**, 770–778 (2016)
29. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *North American Chapter of the Association for Computational Linguistics*; 2019.
30. Le, D., Alam, M., Yao, C.K., et al.: Transfer learning for automated OCTA detection of diabetic retinopathy. *Translat. Vis. Sci. Technol.* **9**, 35–35 (2020)
31. Almasi, R., Vafaei, A., Kazeminasab, E., Rabbani, H.: Automatic detection of microaneurysms in optical coherence tomography images of retina using convolutional neural networks and transfer learning. *Sci. Rep.* **12**, 13975 (2022)
32. Rakocz, N., Chiang, J.N., Nittala, M.G., et al.: Automated identification of clinical features from sparsely annotated 3-dimensional medical imaging. *NPJ Digit Med* **4**, 44 (2021)
33. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv* 2020;abs/2010.11929.
34. Khan, S.H., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M.: Transformers in vision: a survey. *ACM Comput. Surv. (CSUR)* **54**, 1–41 (2021)
35. Brock A, Donahue J, Simonyan K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *ArXiv* 2018;abs/1809.11096.
36. Ramesh A, Pavlov M, Goh G, et al. Zero-Shot Text-to-Image Generation. *ArXiv* 2021;abs/2102.12092.
37. Dosovitskiy AB, Lucas; Kolesnikov, Alexander; Weissenborn, Dirk; Zhai, Xiaohua; Unterthiner, Thomas; Dehghani, Mostafa; Minderer, Matthias; Heigold, Georg; Gelly, Sylvain; Uszkoreit, Jakob; Houthsby, Neil. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint* 2020; [arXiv:2010.11929](https://arxiv.org/abs/2010.11929).
38. Zhuang Liu HM, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie. A ConvNet for the 2020s. *arXiv preprint* 2022.
39. Qian, B., Chen, H., Wang, X., et al.: A public benchmark for diabetic retinopathy analysis on ultra-wide optical coherence tomography angiography images. *Patterns* **2024**, 100929 (2022)
40. Chefer, H., Gur, S., Wolf, L.: Transformer Interpretability beyond attention visualization. *IEEE/CVF Conf. Comput. Vis. Pattern Recogn. (CVPR)* **2021**, 782–791 (2021)
41. Kreitner L, Paetzold JC, Rauch N, et al. Synthetic optical coherence tomography angiographs for detailed retinal vessel segmentation without human annotations. *IEEE Trans Med Imaging* 2024;Pp.
42. Hormel, T.T., Hwang, T.S., Bailey, S.T., Wilson, D.J., Huang, D., Jia, Y.: Artificial intelligence in OCT angiography. *Prog. Retin. Eye Res.* **85**, 100965 (2021)
43. Xu, X., Yang, P., Wang, H., et al.: AV-casNet: fully automatic arteriole-venule segmentation and differentiation in OCT angiography. *IEEE Trans. Med. Imaging* **42**, 481–492 (2023)
44. Liu, Y., Carass, A., Zuo, L., et al.: Disentangled representation learning for OCTA vessel segmentation with limited training data. *IEEE Trans. Med. Imaging* **41**, 3686–3698 (2022)
45. Ma, Y., Hao, H., Xie, J., et al.: ROSE: a retinal OCT-angiography vessel segmentation dataset and new model. *IEEE Trans. Med. Imaging* **40**, 928–939 (2021)
46. Yang, D., Tang, Z., Ran, A., et al.: Assessment of parafoveal diabetic macular ischemia on optical coherence tomography angiography images to predict diabetic retinal disease progression and visual acuity deterioration. *JAMA Ophthalmol.* **141**, 641–649 (2023)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Yixiao Jin received her B.S. degree in medicine from Xi'an Jiaotong University, Shaanxi, China. She is currently pursuing an M.S. degree in ophthalmology with Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. Her research interests include diabetic retinopathy and fundus image processing.



Fu Gui is a professor in the Department of Ophthalmology, The Second Affiliated Hospital, Jiangxi Medical College Nanchang University, Nanchang, China. Her academic and research focus on the pathogenesis and therapeutic strategies for vascular endothelial injury in diabetic retinopathy.



Minghao Chen received his B.S. degree in medicine from Wenzhou Medical University, Zhejiang, China. He is currently pursuing a Ph.D. degree in ophthalmology with Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. His research interests include diabetic retinopathy and age-related macular degeneration.



Jingfa Zhang is the Deputy Director of Shanghai Eye Research Institute, Shanghai, China, and the Associate Chief Physician, Principal Investigator and Doctoral Supervisor in the Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. His academic and research endeavors are primarily concentrated on the pathogenesis and therapeutic strategies for diabetic retinopathy and age-related macular degeneration. He has also delved into the innovative application of artificial intelligence in image analysis techniques for retina diseases.



Xiang Chen completed her undergraduate studies at Hunan University, located in Changsha, China, in the year 2024. Presently, she is in the process of earning her Ph.D. degree in computer science from Shanghai Jiao Tong University. Her academic and research pursuits are centered around key areas in the field of computer science, specifically focusing on natural language processing, computer vision, and time series forecasting.



Haoxuan Li received her MSc in digital media from Zhengzhou University of Light Industry, and she is currently pursuing her PhD in Sports Engineering at Shanghai Sport University. Her research interests include artificial intelligence, sports engineering, and digital media.