Seeing Beyond the Surface: Retinal Thickness Prediction from Color Fundus Photography for DME Management

Wenquan Cheng[1,2], Yihua Sun[1], Jinyuan Wang[3], Jia Guo[1], Zihan Li[1], Zhuhao Wang[1], Guochen Ning[3], Yingfeng Zheng[4], Hongen Liao[1], Tien Yin Wong[2,3,5], and Su Jeong Song[6]

[1] School of Biomedical Engineering, Tsinghua Medicine, Tsinghua University, Beijing, China

[2] Beijing Visual Science and Translational Eye Research Institute (BERI), Beijing Tsinghua Changgung Hospital Eye Center, Tsinghua Medicine, Tsinghua University, Beijing, China

[3] School of Clinical Medicine, Tsinghua Medicine, Tsinghua University, Beijing, China
ningguochen@tsinghua.edu.cn

[4] State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China

[5] Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore

[6] Department of Ophthalmology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
sjsong7@gmail.com

Abstract. Diabetic macular edema (DME) is a leading cause of severe vision loss in the working-age population. Optical coherence tomography (OCT) is the gold standard for DME management and primary care referrals, providing retinal thickness maps (RTMs) that quantify retinal pathologies. However, its limited accessibility in resource-constrained settings necessitates more efficient solutions. While color fundus photography (C-FP) is a cost-effective screening tool, its potential for quantitative thickness evaluation remains underexplored. In this paper, we propose a novel Global-to-Local conditional Diffusion model for Retinal Thickness prediction (GLD-RT), the first attempt to predict RTM solely from C-FP. Our framework predicts thickness distributions of macular region from 2D inputs through a diffusion process guided by hierarchical global-to-local retinal features. Experimental results demonstrate that GLD-RT accurately depicts both physiological and pathological retinal morphology, achieving superior performance in thickness quantification and enabling a more detailed examination of retinal structures. Furthermore, C-FP-generated RTMs exhibit promising utility in facilitating DME diagnosis. This approach transforms conventional fundus imaging into a comprehensive and cost-effective diagnostic tool for DME screening and monitoring in resource-limited settings, thereby holding significant clinical implications.

S. J. Song and G. Ning are the co-corresponding authors.

Keywords: Diabetic macular edema · Retinal thickness prediction · Color fundus photography · Conditional diffusion model

1 Introduction

Diabetic macular edema (DME) is a major cause of severe vision loss in working-age populations, imposing substantial global healthcare burdens [6]. While anti-Vascular Endothelial Growth Factor (anti-VEGF) therapy is the primary treatment, its efficacy heavily relies on regular retinal monitoring [27]. Optical coherence tomography (OCT) serves as the gold standard for retinal assessment [4]. It generates retinal thickness maps (RTMs) representing en-face thickness between the internal limiting membrane and Bruch's membrane. RTM provides detailed quantification and visualization of retinal structures, highlighting abnormalities with accurate location, contour, and volume.

However, OCT's accessibility is significantly constrained due to its high cost and operational complexity [22]. Real-world studies reveal suboptimal anti-VEGF treatment outcomes due to insufficient follow-up visits [9, 21]. Therefore, there is a pressing need for accessible monitoring solutions, which could also facilitate home monitoring and referral triage in resource-limited settings.

Color Fundus Photography (C-FP) is one of the most widely performed examinations in ophthalmology and is even feasible with smartphones [11]. However, as a 2D imaging technique, C-FP lacks depth resolution for quantitative retinal layer evaluation. Previous approaches relied on surrogate markers of retinal thickening (lipid deposits, laser scars, etc.) [28], demonstrating limited specificity and sensitivity for DME detection [29]. Arcadu et al. [3] developed a model to predict retinal thickening and thickness metrics from C-FP. However, the prediction accuracy failed to meet clinical standards, especially when image quality was poor.

Recent approaches have attempted to bridge this gap by extracting thickness information from infrared fundus photography (IR-FP). In clinical practice, C-FP is acquired during the initial examination, while RTM examination requires additional OCT acquisition with IR-FP as the localizer. The early treatment diabetic retinopathy study (ETDRS) grid segments the macular region for standardized thickness evaluation [10]. Holmberg et al. [15] developed DeepRT for RTM prediction from IR-FP, while Sun et al. [26] extended this approach by incorporating an unregistered C-FP to predict the ETDRS grid and more accurate RTM. Despite these efforts, the need for an additional device to capture IR-FP limits their clinical application, and the ETDRS grid provides only coarse measurements insufficient for detailed monitoring. In addition, IR-FP exhibits inherent limitations: hyperreflective artifacts, restricted illumination wavelength optimized for subretinal structures, and inadequate visualization of critical pathological markers including hard exudates and hemorrhages [1].

In contrast, C-FP offers superior spatial resolution, a broader spectral range, and a wider field of view. It captures critical global retinal structures, such as vascular networks and optic nerve heads, which are key pathological biomarkers.

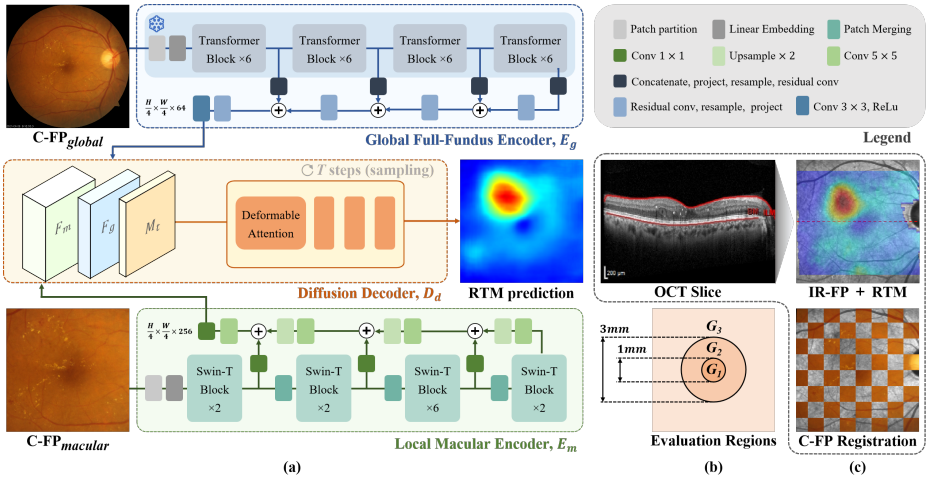

Fig. 1. (a) GLD-RT leverages a dual-stream diffusion framework to integrate global contextual features with fine-grained details. (b) Evaluation regions for RTM prediction. (c)IR-FP, the localizer in OCT imaging that generates RTM, is used to register C-FP$_{global}$ with RTM, generating C-FP$_{macular}$.

These structures are notably affected by pathologies like edema, which alters vessel characteristics (trajectory, branching, caliber, density) and impacts adjacent retinal layer integrity. Meanwhile, DME-induced retinal layer disorganization causes subtle texture and color variations in C-FP, which are challenging to identify even for experienced clinicians. These subtle manifestations of pathological structural modifications are particularly pronounced in the macular region. Therefore, accurate interpretation of these biomarkers facilitates more precise retinal thickness prediction, which requires both robust high-level semantic features and precise local macular details.

Firstly, to ensure spatial correspondence, we implement precise multi-modal registration between C-FP and RTM in the macular region(Fig. 1c) [24]. We present GLD-RT (Global-to-Local conditional Diffusion model for Retinal Thickness prediction), a novel framework that predicts accurate RTMs directly from C-FP. GLD-RT incorporates a hybrid CNN-transformer conditional diffusion architecture that processes both global context and local structural details through dual-stream feature extraction. We enable robust multi-scale feature fusion using RETFound [32], to effectively capture heterogeneous DME manifestations.

To the best of our knowledge, this study represents the first attempt to predict RTMs solely from C-FP, enabling accurate retinal thickness assessment from widely available C-FP. The proposed GLD-RT has significant clinical potential to transform DME management by facilitating effective and timely intervention, ultimately improving patients' visual outcomes.

2 Methodology

GLD-RT employs a decoupled encoder-decoder architecture (Fig. 1a). Parallel encoders extract high-level semantic representations and fine-grained local features respectively. A hierarchical diffusion decoder subsequently integrates these features to predict RTM with global-to-local anatomical consistency.

2.1 Multi-Modal Fundus Image Registration

Due to OCT's limited en-face resolution, we use IR-FP as an intermediary for pixel-wise registration between the entire C-FP (C-FP$_{global}$) and RTM ground truth $M \in \mathbb{R}^{H \times W}$. We extract retinal vessels as modality-agnostic features to align images from two modalities. We use a pre-trained Unet [18] to segment vessels from C-FP$_{global}$ and IR-FP. Vessel mask keypoints are detected by AKAZE [2], and homography matrices are computed with RANSAC [8] to eliminate outliers. The filtered key points are matched with the k-nearest neighbor (k-NN) matching strategy. C-FP$global$ is nonrigidly registered to IR-FP and cropped to obtain C-FP$macular$, retaining only the macular region that corresponds pixel-wise to IR-FP.

2.2 Global-to-Local Feature Extraction

Local Macular Encoder. The local macular encoder (Fig.1, E_m) extracts fine-grained features $F_m \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 256}$ from C-FP$_{macular}$. E_m uses the swin-transformer [19] to capture hierarchical representations through progressive patch merging, followed by the feature pyramid network [17] to integrate features across four resolution levels.

Global Full-Fundus Encoder. To compensate for peripheral context loss in C-FP$_{macular}$, we incorporate C-FP$_{global}$ to capture comprehensive retinal structure. The proposed global full-fundus encoder (Fig.1, E_g) leverages RETFound [32], a SOTA retinal foundation model, to encode robust anatomical and pathological retinal patterns. E_g employs RETFound's ViT [7] encoder with frozen pre-trained weights to extract rich semantic context from C-FP$_{global}$. An adapter aggregates high-dimensional tokens from four encoding stages into a global feature map $F_g \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 64}$.

2.3 Conditional diffusion decoder

The decoder implements a conditional diffusion framework [31] with deformable attention modules. To expedite sampling while maintaining high fidelity, we adopt the denoising diffusion implicit model (DDIM) [14], which employs a non-Markovian, deterministic sampling process that requires significantly fewer steps. During training, scheduled Gaussian noise is applied to M following a noise scheduler [20]. Concatenated features $[F_m, F_g]$ from dual encoders condition the

diffusion process. The diffusion decoder(Fig. 1, E_d) learns to predict M from M_t, the noisy RTM at timestep t. At inference, GLD-RT reverses the diffusion process to generate RTM predictions. The model is optimized using mean squared error:

$$\mathcal{L} = \mathbb{E}_{M,t,F_m,F_g}\left[\|M - D_d(M_t, t, F_m, F_g)\|_2^2\right] \quad (1)$$

3 Experiments

3.1 Experimental Setup

Dataset. The GLD-RT development utilizes 2,918 data triplets, each consisting of corresponding OCT, IR-FP, and C-FP images, from 1,418 DME patients undergoing anti-VEGF therapy at Kangbuk Samsung Hospital (IRB: KBSMC 2022-12-016-004). The cohort presents mean retinal thickness of $296.99 \pm 30.67\,\mu$m and central macular thickness of $292.46 \pm 75.00\,\mu$m, indicating substantial macular edema prevalence. IR-FP and 31 OCT B-scans are acquired via Heidelberg devices, which automatically generate membrane segmentations. Experienced ophthalmologists exclude scans with poor fixation or segmentation errors. For DME diagnosis, we utilize the Mobile Brazilian Retinal Dataset (mBRSET) [30], a collection of C-FP captured with portable cameras. It comprises 5,164 images from 1,291 diabetics in Itabuna, Bahia, Brazil. All images are annotated with DME diagnosis and image quality labels.

Data Pre-processing. IR-FP is centrally cropped to 544×544, corresponding to the OCT scanning area. C-FP$_{global}$ is downsampled from 3608×3608 to 544×544 to match IR-FP. RTM is computed from 31 B-scan lines and linearly interpolated to match the resolution of IR-FP. RTM undergoes sequential smoothing to preserve structure while reducing artifacts: Gaussian filtering ($\sigma = 3$) followed by non-local means denoising (patch size: 5×5, search window: 6×6, $h = 0.1$). For DME diagnosis, C-FP is cropped to 800×800 according to fovea location and downsampled to 544×544.

Implementation Details. We split the dataset at the patient level into 2043 training triplets (993 patients), 292 validation triplets (142 patients), and 583 testing triplets (283 patients). We split the mBRSET into 2409 training images (671 patients), 345 validation images (98 patients), and 688 testing images (186 patients). For GLD-RT, Adam optimizer [16] is employed with $(\beta_1, \beta_2) = (0.9, 0.999)$, cosine decay scheduling after a 16-epoch warm-up (ramping from 1×10^{-8} to 6×10^{-5}), and fixed weight decay of 1×10^{-2} over 300 epochs. In the diffusion process, the timesteps for training and inference are set to 20 and 5. For DME diagnosis, Adam optimizer is employed with a learning rate of 2×10^{-5}, $(\beta_1, \beta_2) = (0.9, 0.999)$, weight decay of 1×10^{-2}, and ExponentialLR scheduling (with $\gamma = 0.999$) for 100 epochs. All experiments were implemented on NVIDIA GeForce RTX 3090 with data augmentation via random flipping and rotation.

Table 1. Quantitative comparison of different methods for RTM prediction, with MAE (μm) and PSNR (dB). * indicates GLD-RT outperforms baselines with p-values$<$0.01.

Inputs	Methods	RTM		G_1		G_2		G_3	
		MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑
IR-FP	UNet [23]	27.12*	27.53*	52.33*	23.14*	32.51*	26.98*	25.46*	27.63*
	DeepRT [15]	25.75*	28.33*	49.62*	23.49*	29.82*	27.64*	23.87*	27.95*
Both	M^{2}FRT [26]	23.88*	29.08*	40.13*	27.17*	26.74*	29.97	21.69*	29.61
Registered C-FP	UNet [23]	26.75*	27.86*	46.47*	25.58*	30.82*	27.95*	25.05*	27.84*
	DeepRT [15]	25.15*	28.71*	43.01*	25.73*	29.63*	28.39*	23.76*	28.05*
	Swin-Unetr [12]	24.66*	28.63*	42.14*	26.94*	28.36*	28.58*	22.62*	28.13*
	Trans-UNet [5]	23.89*	28.87*	41.65*	27.62*	28.03*	28.95*	22.47*	28.70*
	M^{2}FRT [26]	23.21*	29.34*	36.25*	28.31*	26.12*	30.11	21.43*	29.72
	GLD-RT	**20.08**	**30.91**	**31.97**	**30.18**	**22.85**	**31.16**	**19.70**	**30.25**

Performance Metrics. For RTM prediction, we evaluated model performance using mean absolute error (MAE) and peak signal-to-noise ratio (PSNR) across evaluation regions (G_1, G_2, G_3) illustrated in Fig. 1b. These regions correspond to clinically relevant areas of the ETDRS grid. The Wilcoxon signed-rank test is employed to compare the performance of GLD-RT with the baselines. For DME diagnosis, model performance is quantified using accuracy, recall, F1-score, and area under the receiver operating characteristic curve (AUC).

3.2 Quantitative and Qualitative Evaluations on RTM Predictions

We evaluated our GLD-RT against SOTA methods, including medical image dense prediction models (UNet [23], Trans-UNet [5], Swin-Unetr [12]) and specialized RTM prediction frameworks (DeepRT [15], M^{2}FRT [26]), as shown in Table 1. Notably, the SOTA M^{2}FRT [26] achieved significant improvements by incorporating unregistered C-FP, demonstrating that C-FP offers extra thickness information over IR-FP.

Our cross-modal registration enhanced spatial coherence for fine-detail representation learning. Substituting IR-FP with registered C-FP in baseline models (UNet, DeepRT, M^{2}FRT) consistently improved model performance. This improvement was most evident in the challenging and clinically important foveal region (G1), where MAE peaks. The fovea is critical for high-acuity central vision, where minor thickness variations can substantially impact visual function. The improved foveal prediction could facilitate more accurate treatment response assessment, optimizing the balance between intervention strategy and visual outcome.

Despite these advances, current methods fail to fully leverage the rich information in C-FP. Our GLD-RT significantly outperformed SOTA methods across all retinal regions. As shown in Fig. 2, GLD-RT robustly detected thickness variations of diverse pathological lesions, providing accurate representations of subtle and irregular retinal anatomy. The results demonstrate that the diffusion process with global-to-local conditions effectively captures the underlying structure of

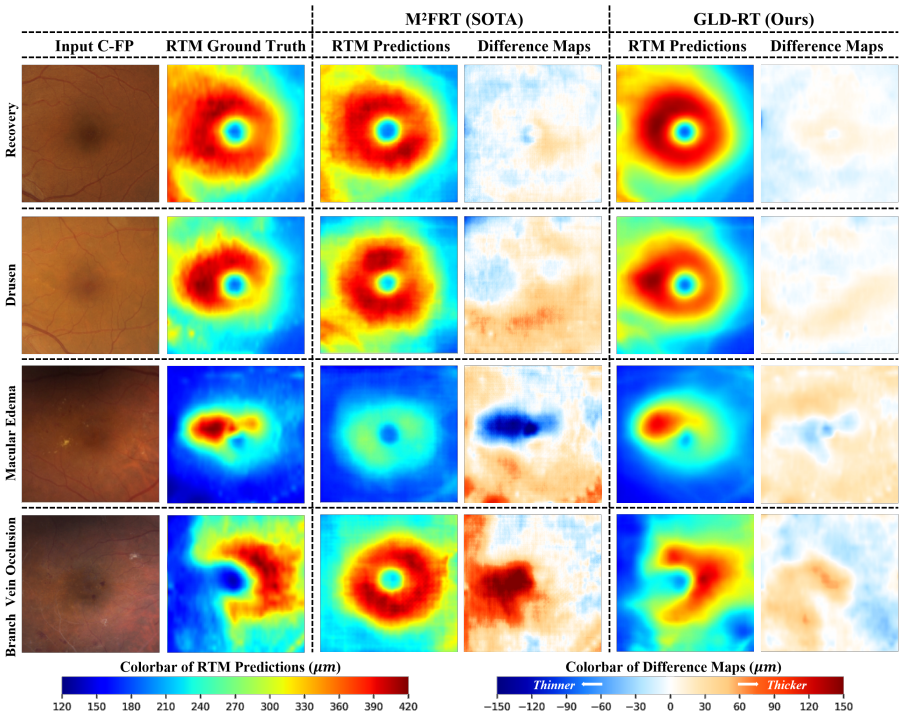

Fig. 2. Qualitative results of our model and M^{2}FRT. GLD-RT robustly detects diverse pathological lesions and accurately depicts subtle and irregular retinal anatomy. Difference maps (prediction minus ground truth) further underscore these gains.

retinal layers. By generating precise and comprehensive RTM from widely available C-FP, GLD-RT provides an efficient diagnostic tool for timely intervention, informed treatment decisions, and enhanced prognostication in retinal care.

3.3 Ablation Study

We conducted comprehensive ablation experiments to evaluate the contribution of each component (Table 2). Our backbone, consisting of E_m and D_{nd} (decoder without diffusion), outperformed M^{2}FRT, demonstrating the effectiveness of our enhanced local feature extraction in macular regions.

The proposed D_d remarkably improved model performance, particularly in the central macula (G_1). This region exhibits a complex layered structure with DME-induced pathological alterations. Qualitative analysis shows that the diffusion process effectively captures macular structure with fine details (Fig. 3a).

The proposed E_g significantly enhanced model performance, with domain-specific pre-training for retinal images playing a crucial role in this improvement. RETFound pre-trained on fundus photography achieved overall improvement

Table 2. Ablation study of RTM prediction, with MAE (μm) and PSNR (dB). * indicates GLD-RT outperforms baselines with p-values<0.01.

Methods	RTM		G_1		G_2		G_3	
	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑
E_m, D_{nd}	22.68*	29.40*	35.09*	28.84*	24.62*	30.51*	20.50*	29.72*
E_m, D_d	21.22*	29.53*	32.78*	29.36	23.82*	30.60	20.18*	29.75*
$E_g(ImageNet), E_m, D_{nd}$	22.16*	29.48*	34.36*	29.09*	24.33*	30.55*	20.25*	29.84
$E_g(RETFound), E_m, D_{nd}$	21.28*	29.57*	33.89*	29.17*	23.98*	30.72	20.08	29.92
$E_g(ImageNet), E_m, D_d$	20.99*	29.43*	33.03*	29.39	24.15*	30.61	20.07	29.74*
$E_g(RETFound), E_m, D_d$	**20.08**	**30.91**	**31.97**	**30.18**	**22.85**	**31.16**	**19.70**	**30.25**

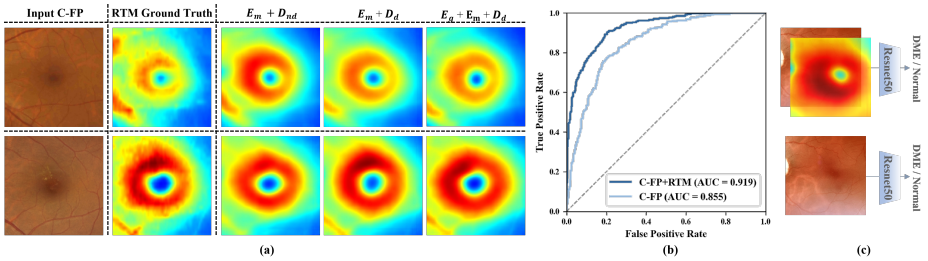

Fig. 3. (a) Qualitative results of ablation study on GLD-RT. The proposed D_d significantly enhanced fine macular structural details, while E_g further enforced global anatomical consistency. (b) Receiver operating characteristic curve of DME diagnosis. (c)Experiment design of DME diagnosis.

over ImageNet initialization. Global-to-local feature conditioning outperformed standard decoding approaches. The iterative diffusion process successfully fuses high-level contextual information with local embeddings. Improved anatomical consistency in predictions validates the enhanced multi-scale structural information from C-FP (Fig. 3a).

3.4 RTMs Assistance in DME Diagnosis

To mitigate ambiguous foveal location in portable imaging, we segmented the foveal region in C-FP using histogram equalization, thresholding, and morphological erosion [25]. We subsequently generated corresponding RTMs of the macular region and concatenated the C-FP and RTM as dual input. We employed ResNet50 [13] (without pre-trained weights) for binary classification (Fig. 3c).

The dual-input approach consistently outperformed the CFP-only baseline(Fig. 3b), with significant improvements in accuracy (93.44% vs. 92.21%), recall (79.49% vs. 64.41%), F1-score (85.90% vs. 75.84%), and AUC (Fig. 3b). These results demonstrate that C-FP-generated RTMs provide complementary diagnostic insights for DME diagnosis, aligning with clinical protocols requiring both OCT and C-FP examination. Our findings underscore the potential for effective portable retinal imaging systems in resource-constrained healthcare environments.

4 Conclusion and Dicussion

This study presents a novel diffusion-based paradigm for RTM prediction from C-FP in DME patients. The architecture employs parallel encoders to capture high-level semantic representations and fine-grained local features, integrated by a hierarchical diffusion decoder that ensures global-to-local anatomical consistency. Our approach demonstrates superior accuracy and robust generalizability. Our validation in DME diagnosis further highlights the practicality of C-FP-generated RTMs in resource-constrained clinical settings. Our method offers portable, cost-effective OCT alternatives, enabling more frequent retinal assessments and personalized treatment strategies for DME. Future directions include: prospective clinical validation and integration into existing healthcare workflows; incorporating higher-resolution OCT for detailed RTM prediction.

Acknowledgments. We thank Prof. Dawei Li, Prof. Yih Chung Tham and Prof. Yaxing Wang for providing constructive suggestions. This study was funded by National Key R&D Program (No. 2022YFC2502800), National Natural Science Fund of China (No. 82388101) and Beijing Natural Science Foundation (No. IS23096).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ajaz, A., Kumar, D.K.: Infrared retinal images for flashless detection of macular edema. Scientific Reports **10**(1), 14384 (2020)
2. Alcantarilla, P.F., Solutions, T.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. IEEE Trans. Patt. Anal. Mach. Intell **34**(7), 1281–1298 (2011)
3. Arcadu, F., et al.: Deep learning predicts oct measures of diabetic macular thickening from color fundus photographs. Investigative Opthalmology and Visual Science **60**, 852 (2019)
4. Baskin, D.E.: Optical coherence tomography in diabetic macular edema. Current opinion in ophthalmology **21**, 172–177 (2010)
5. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
6. Ciulla, T.A., Pollack, J.S., Williams, D.F.: Visual acuity outcomes and anti-vegf therapy intensity in diabetic macular oedema: a real-world analysis of 28 658 patient eyes. British Journal of Ophthalmology **105**, 216–221 (2021)
7. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
8. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)

9. Glassman, A., et al.: Five-year outcomes after initial aflibercept, bevacizumab, or ranibizumab treatment for diabetic macular edema (protocol t extension study). Ophthalmology **127**, 1201–1210 (2020)
10. Group, E.R.: Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified airlie house classification: Etdrs report number 10. Ophthalmology **98**, 786–806 (1991)
11. Haddock, L.J., Kim, D.Y., Mukai, S.: Simple, inexpensive technique for high-quality smartphone fundus photography in human and animal eyes. Journal of ophthalmology **2013**, 518479 (2013)
12. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
14. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
15. Holmberg, O., et al.: Self-supervised retinal thickness prediction enables deep learning from unlabelled data to boost classification of diabetic retinopathy. Nature Machine Intelligence **2**, 719–726 (2020)
16. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
17. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
18. Liu, W., Yang, H., Tian, T., Cao, Z., Pan, X., Xu, W., Jin, Y., Gao, F.: Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation. IEEE journal of biomedical and health informatics **26**(9), 4623–4634 (2022)
19. Liu, Z., et al.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
20. Lugmayr, A., et al.: Repaint: Inpainting using denoising diffusion probabilistic models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11461–11471 (2022)
21. Mehta, H., et al.: Real-world outcomes in patients with neovascular age-related macular degeneration treated with intravitreal vascular endothelial growth factor inhibitors. Progress in Retinal and Eye Research **65**, 127–146 (2018)
22. Organization, W.H., et al.: Monitoring progress and acceleration plan for ncds, including oral health and integrated eye care, in the who south-east asia region. Tech. rep., World Health Organization. Regional Office for South-East Asia (2022)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
24. Shi, D., Zhang, W., He, S., Chen, Y., Song, F., Liu, S., Wang, R., Zheng, Y., He, M.: Translation of color fundus photography into fluorescein angiography using deep learning for enhanced diabetic retinopathy screening. Ophthalmology science **3**(4), 100401 (2023)

25. Sigut, J., Nuñez, O., Fumero, F., Alayon, S., Diaz-Aleman, T.: Fovea localization in retinal images using spatial color histograms. Multimedia Tools and Applications **83**(6), 17753–17771 (2024)
26. Sun, Y., et al.: Retinal thickness prediction from multi-modal fundus photography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 585–595. Springer (2023)
27. Virgili, G., Parravano, M., Menchini, F., Evans, J.R.: Anti-vascular endothelial growth factor for diabetic macular oedema. Cochrane Database of Systematic Reviews (2014)
28. Wang, Y.T., Tadarati, M., Wolfson, Y., Bressler, S.B., Bressler, N.M.: Comparison of prevalence of diabetic macular edema based on monocular fundus photography vs optical coherence tomography. JAMA ophthalmology **134**, 222–228 (2016)
29. Wong, R.L., Tsang, C., Wong, D.S., McGhee, S., Lam, C., Lian, J., Lee, J.W., Lai, J.S., Chong, V., Wong, I.Y.: Are we making good use of our public resources? the false-positive rate of screening by fundus photography for diabetic macular oedema. Hong Kong Medical Journal **23**, 356 (2017)
30. Wu, C., et al.: Mbrset: A portable retina fundus photos benchmark dataset for clinical and demographic prediction. medRxiv pp. 2024–07 (2024)
31. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3836–3847 (2023)
32. Zhou, Y., et al.: A foundation model for generalizable disease detection from retinal images. Nature **622**, 156–163 (2023)