Fairness-Aware vCDR-Controlled Generation for Glaucoma Diagnosis

Ziheng Wang[1]*, Shuran Yang[2]*, Wen Chen[3], Zhen Zhang[3], Mengyu Wang[4], Feixiang Zhou[5], Yu Tian[6], Meng Wang[7], Yitian Zhao[8], Yalin Zheng[5], and Yanda Meng[1](✉)

[1] Computer Science Department, University of Exeter, Exeter, UK
[2] School of Journalism and Communication, Xizang Minzu University, Xianyang, China
[3] Department of Ophthalmology, Xuzhou Central Hospital, Xuzhou, China
[4] Harvard Ophthalmology AI Lab, Harvard University, Cambridge, USA
[5] Eye and Vision Sciences Department, University of Liverpool, Liverpool, UK
[6] Department of Computer Science, University of Central Florida, Orlando, USA
[7] Centre for Innovation and Precision Eye Health & Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[8] Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo, China
Y.M.Meng@exeter.ac.uk

Abstract. Glaucoma is a leading cause of irreversible blindness, and early diagnosis is crucial for effective treatment. However, AI-assisted glaucoma diagnosis faces challenges in fairness and data scarcity, because AI model biases can lead to disparities across demographic groups. To address this, we propose GlaucoDiff, a diffusion-based generative model that synthesizes SLO images with precise control over the vertical cup-to-disc ratio. Unlike previous methods, GlaucoDiff enables bidirectional synthesis, generating both healthy and glaucomatous samples of varying severity, thus enhancing the dataset diversity. To ensure anatomical fidelity, GlaucoDiff leverages real fundus backgrounds while generating the optic nerve head regions. We also introduce a sample selection strategy that filters generated images based on the alignment agreement percentage, compared with target optic structures, ensuring the high-quality of the synthetic data. Experiments on two public ophthalmic datasets demonstrate that GlaucoDiff outperforms state-of-the-art approaches in both diagnosis and fairness measurement settings. Two independent ophthalmologists' evaluations confirm the clinical relevance of the generated images, highlighting GlaucoDiff's potential for improving AI-driven glaucoma diagnosis. Our code is available [2].

Keywords: Glaucoma Diagnosis · Image Synthesis · Fairness Learning.

[1] * Ziheng Wang and Shuran Yang contributed equally to this work.
[2] https://github.com/WANG-ZIHENG/GlaucoDiff

1 Introduction

Severe glaucoma can lead to partial vision loss or irreversible blindness, making early diagnosis crucial for timely treatment [1]. The scanning laser ophthalmoscope (SLO) is widely used in clinical screening for the detection of glaucoma because of its cost-effectiveness compared to Optical Coherence Tomography or other devices [2]. Previous studies [3,4] found that the incidence of glaucoma is associated with attributes such as race, gender, and age. Although AI-powered computer vision techniques have been widely utilized in glaucoma diagnosis across various imaging modalities [5], the potential biases in AI models, such as data imbalance, pose a major challenge. These biases can result in unfair suboptimal performance, particularly for the demographic minority groups [6,7], leading to disparities in diagnosis and treatment. Thus, improving the fairness of AI-assisted glaucoma diagnosis is an important and urgent task.

Fairness in AI-powered medical imaging is a critical ethical concern. Recent studies in ophthalmic imaging have increasingly focused on addressing demographic group fairness. For example, [6] improves both fairness and overall model performance by minimizing the Sinkhorn distance between the overall data distribution and the distributions of individual demographic groups. Another study [7] introduced an error-bound scaling approach that re-weighted the loss function based on the performance disparities among different identity groups, prioritizing the underrepresented groups. This approach ensures that the model does not solely optimize overall performance but also maintains equitable accuracy across various demographic groups. Differently, GlaucoDiff uses demographic identity information as textual prompts for a diffusion-based generative model. By synthesizing high-quality samples that are tailored to specific demographic groups, GlaucoDiff enhances model fairness while preserving diagnostic accuracy.

Using generative models [8,9,10] for data augmentation in medical imaging has become widespread for various disease diagnosis tasks [11]. A major challenge, particularly in ophthalmology, is ensuring anatomical structure accuracy, especially when synthesizing SLO images with precise vascular structures. Common issues include disrupted vascular continuity, imbalanced branch-to-main vessel ratios, and missing peripapillary vessels [12]. Recent studies have leveraged the ControlNet-guided Stable Diffusion framework [10,14] to generate polyps from non-polyp frames (*e.g.* ControlPolypNet [15]) or to synthesize cardiac lesion MRI and lung nodule CT images from healthy samples (*e.g.* LeFusion [16]). These methods can only do unidirectional synthesis to generate diseased samples from healthy controls. FairDiff [13] employs diffusion models to generate cup-to-disc contours, which are used to synthesize SLO images for improving segmentation model performance. However, it fails to ensure accurate anatomical structures in the background and lacks control over vertical cup-to-disc ratio (vCDR) values in the synthesized images. Differently, GlaucoDiff offers greater flexibility, enabling bidirectional synthesis to controllably generate both healthy and glaucomatous samples with varying vCDR values. This enhances dataset diversity and allows for the simulation and visualization of disease prognosis and prediction. We generate the optic cup (OC) and optic disc (OD) regions via scal-

ing the OC shape to control vCDR and ensuring the generation of high-quality synthetic SLO images that preserve anatomical fidelity. Precise vCDR control, a critical biomarker for glaucoma diagnosis [17], significantly improves the clinical relevance and diagnostic applicability of synthetic data. OC and OD segmentation, widely used in glaucoma diagnosis [5,18] and commercial devices [19], facilitates seamless integration of GlaucoDiff into existing AI-based diagnostic frameworks, offering a valuable tool for clinical research and model training.

Our main contributions are: (1) We propose GlaucoDiff that generates OC and OD regions in SLO images with precise vCDR control, supporting bidirectional synthesis. (2) We introduce a sample selection strategy that automatically filters the generated images based on their alignment agreement with the target OC and OD shapes. (3) Experiments show that GlaucoDiff outperforms state-of-the-art methods in both classification accuracy and fairness metrics.

2 Method

2.1 Preliminaries

ControlNet [14] is a neural network architecture designed to enhance spatial control in image generation tasks for large pre-trained text-to-image diffusion models, such as Stable Diffusion [10]. It incorporates additional image conditions, like segmentation maps or human poses, into the generation process to provide precise spatial conditioning. The architecture consists of a locked pre-trained model, $\mathcal{F}(\cdot;\Theta)$, and a trainable copy with parameters Θ_c, connected through "zero convolution" layers, $\mathcal{Z}(\cdot;\cdot)$. This design enables the integration of spatial conditions while retaining the original capabilities of the pre-trained model. The output feature map of ControlNet, denoted as y_c, is expressed as:

$$y_c = \mathcal{F}(x;\Theta) + \mathcal{Z}\left(\mathcal{F}(x + \mathcal{Z}(c;\Theta_{z_1});\Theta_c);\Theta_{z_2}\right), \tag{1}$$

Here, the function $\mathcal{F}(x;\Theta)$ represents the pre-trained neural network blocks of the Stable Diffusion model, which process the input feature map x to generate a base output y. A zero convolution layer $\mathcal{Z}(c;\Theta_{z1})$ transforms the conditional input c into a feature map of the same dimensions as x. This transformed feature map is added to x, forming a combined feature map that is fed into the trainable copy $\mathcal{F}(\cdot;\Theta_c)$, producing a new feature map. The new feature map is further processed by another zero convolution layer $\mathcal{Z}(\cdot;\Theta_{z2})$ and subsequently added to the base output y, resulting in the final conditioned output feature map y_c. This process enables effective conditional control over image generation.

2.2 GlaucoDiff: vCDR-Controlled Glaucoma Generation

We propose GlaucoDiff, a generative model based on ControlNet that synthesizes OC and OD regions in fundus images with controllable vCDR. By overlaying OC and OD masks onto source images, it provides explicit spatial guidance. Unlike prior methods that regenerate entire SLO images from segmentation masks,

GlaucoDiff focuses on OC and OD regions and adjusts vCDR by scaling the OC. This enables high-quality, clinically relevant synthetic data for precise ophthalmic data augmentation. GlaucoDiff calculates the vCDR value following [20].

Training. Fig. 1 illustrates the architecture of GlaucoDiff. During training, the Stable Diffusion decoder is unlocked to facilitate the generation of medical images. Instead of directly using OC and OD segmentation masks as condition images, we overlay these masks onto the source images during pre-processing to create Masked Images, which serve as the Condition Images for training. Demographic information and calculated vCDR are converted into sentence format, and these textual descriptions are used as prompts for the generative model.

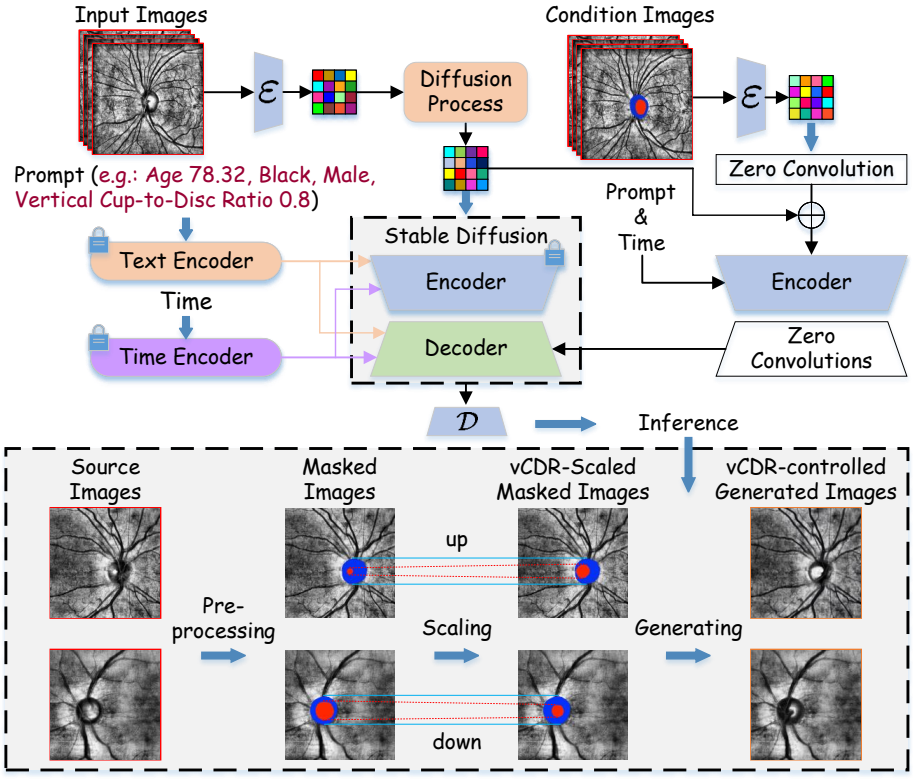

Fig. 1. Illustration of our proposed training framework, where the OC mask is scaled to the specified vCDR during inference to generate vCDR-controlled images.

Inference. During inference, the number of samples from minority groups is increased using the generated samples. We use the scaling method (shown in the lower part of Fig. 1) to construct vCDR-Scaled Masked Images, which are then used as condition inputs to the diffusion model. The vCDR value in

the text prompt is adjusted to match the specified ratio, corresponding to the condition images. This ensures that the generated images accurately represent different demographic groups while producing images with the desired vCDR ratio. To ensure accurate scaling, we maintain the center of the OC mask fixed while proportionally scaling the OC mask to the target vCDR ratio, ensuring that the shape of the OC mask remains unchanged. Although glaucomatous changes to the OC and OD are clinically irreversible, the ability to simulate the reduction of the OC is highly valuable for the purpose of data augmentation. This approach facilitates the expansion of healthy samples, particularly when they are underrepresented in the dataset, thus supporting more balanced and effective training of classification models to address the bias.

Using the scaling method, GlaucoDiff efficiently generates synthetic SLO images with arbitrary vCDR values, allowing flexible creation of healthy or glaucomatous samples for diverse, tailored datasets.

Selecting Generated Samples. Given the focus of GlaucoDiff on the structural integrity of the OC and OD regions, we devised an automatic strategy for selecting high-quality generated samples, as illustrated in Fig. 2. The generated samples are processed through three pre-trained segmentation models, such as UNet [21], SAM [22], and TransUNet [23]. The predicted segmentation masks from these models are averaged and compared to the vCDR-Scaled Masked Images using Dice Similarity Coefficient (Dice) and the 95th percentile Hausdorff Distance (HD95) to ensure that the generated images maintain the desired morphological structure. To rank the samples, we sort them by Dice scores in descending order and HD95 values in ascending order. The ranks are summed for each sample, and we empirically select the top 50% of samples based on cross-

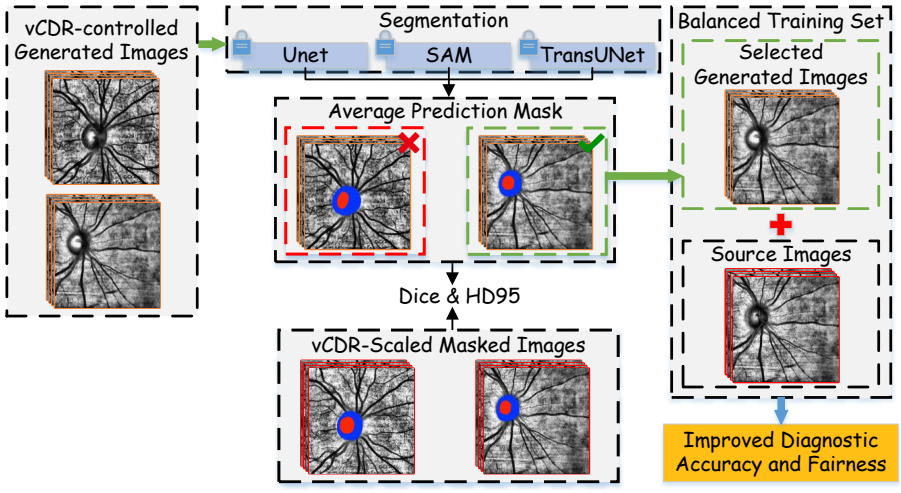

Fig. 2. Selection strategy for the generated samples based on Dice and HD95 metrics.

validation. This selection strategy effectively balances the quantity and quality of retained generated samples, ensuring their quality and contribution to downstream tasks. Based on the vCDR values of these generated samples, we label them as either *healthy* or *suspected glaucoma* cases following the labeling criteria established in [24]. Finally, the selected synthetic images are added to the training dataset. This augmentation enhances the model's overall classification and fairness performance by mitigating potential biases in data distribution.

3 Experiments

3.1 Datasets and Implementation Details

We used the FairSeg [7] and FairVLMed10k [6] datasets for our experiments, both containing SLO fundus images with demographic and clinical information. After removing noisy images (e.g., pure black ones), we retained 5,000 images in FairSeg, split into 3,500 for training, 500 for validation, and 1,000 for testing. For FairVLMed10k, 7,363 images were retained, with 5,266 for training, 692 for validation, and 1,405 for testing. Two independent ophthalmologists annotated FairSeg with *healthy* and *suspected glaucoma* labels, while a SAM model [22], pre-trained on the FairSeg [7] training set, was used to generate OC and OD masks in FairVLMed10k. Both datasets grouped samples by age (*young* < 65, *elderly* ≥ 65). We trained the GlaucoDiff for 10 epochs with a batch size of 16 and a learning rate of 5e-5. For the classification model, we used a pre-trained EfficientNet [25] and trained for 30 epochs with a batch size of 32 and an initial learning rate of 1e-5. All hyperparameters were empirically set via cross-validation experiments. As for the evaluation metrics, we adopted classification metrics, including ACC, AUC, Precision (Prec.), Recall, F1-Score (F1), Matthews Correlation Coefficient (MCC), and Quadratic Weighted Kappa (QWK). Also, we assess the model's fairness using the Demographic Parity Difference (DPD) and the Difference in Equalized Odds (DEOdds), followed by [6]. Lower values of DPD and DEOdds indicate better fairness performance. All results are reported as percentages.

Table 1. Comparison for the overall performance of GlaucoDiff with existing methods.

Dataset	Method	ACC	AUC	Prec.	Recall	F1	MCC	QWK
FairSeg	Baseline	83.38	92.03	83.59	82.95	83.14	66.54	66.33
	FairCLIP [6]	85.42	92.96	85.69	87.73	86.69	71.87	71.56
	FIN [26]	85.48	93.41	85.36	87.51	86.42	69.38	69.23
	FairDomain [27]	85.36	92.79	87.21	86.69	86.95	72.45	72.14
	FairVision [28]	84.89	92.72	81.93	87.46	84.60	67.34	66.01
	Ours	**87.31**	**95.24**	**88.62**	**93.85**	**91.13**	**74.76**	**74.19**
FairVL Med10k	Baseline	69.47	75.93	71.46	60.41	65.50	38.42	41.19
	FairCLIP [6]	74.03	82.78	83.41	61.25	70.63	53.29	53.36
	FIN [26]	74.86	80.43	75.86	65.49	70.29	49.98	48.61
	FairDomain [27]	73.91	80.16	81.47	62.83	70.95	50.35	48.42
	FairVision [28]	74.78	81.29	75.55	65.41	70.12	48.76	48.39
	Ours	**78.10**	**85.48**	**84.07**	**66.45**	**74.23**	**56.09**	**55.97**

3.2 Results

Classification Performance and Fairness. According to the results in Table 1, our proposed method outperforms state-of-the-art methods across all classification metrics. On the FairSeg dataset, our recall rate is at least 6.12% higher

Table 2. Comparison of fairness metrics across various attributes for GlaucoDiff and previous approaches, along with AUC for each group within the different attributes.

Dataset	Attribute	Method	DPD↓	DEOdds↓	Group-wise AUC↑		
					Asian	Black	White
		Baseline	27.55	44.74	96.60	94.86	91.17
		FairCLIP [6]	23.39	24.28	96.83	95.72	92.26
	Race	FIN [26]	23.46	25.97	97.56	95.59	92.41
		FairDomain [27]	26.71	29.62	96.59	93.41	92.17
		FairVision [28]	23.83	37.03	97.41	96.34	91.62
		Ours	**21.92**	**13.63**	**98.02**	**97.06**	**94.47**
					Female	Male	
		Baseline	6.36	14.17	91.19	92.36	
		FairCLIP [6]	1.76	6.49	91.91	93.32	
FairSeg	Gender	FIN [26]	1.15	3.52	92.08	93.14	
		FairDomain [27]	5.81	12.43	90.63	93.74	
		FairVision [28]	2.55	6.02	91.42	93.56	
		Ours	**0.89**	**1.97**	**94.57**	**95.12**	
					Young	Elderly	
		Baseline	9.60	10.41	91.73	93.19	
		FairCLIP [6]	4.81	3.48	93.04	92.71	
	Age	FIN [26]	3.04	7.51	93.69	93.09	
		FairDomain [27]	7.05	3.65	92.72	92.52	
		FairVision [28]	5.65	3.44	92.55	93.36	
		Ours	**2.36**	**2.46**	**94.26**	**95.93**	
					Asian	Black	White
		Baseline	14.90	14.38	75.53	72.26	76.46
		FairCLIP [6]	14.19	9.72	84.92	78.29	83.14
	Race	FIN [26]	7.79	13.45	85.78	79.04	80.18
		FairDomain [27]	7.94	3.51	80.75	77.91	80.63
		FairVision [28]	14.73	9.72	86.41	77.30	81.26
		Ours	**5.59**	**3.07**	**87.59**	**80.75**	**85.54**
					Female	Male	
		Baseline	7.69	13.92	74.51	77.78	
		FairCLIP [6]	2.17	5.94	80.82	84.90	
FairVL Med10k	Gender	FIN [26]	1.91	5.91	78.64	82.41	
		FairDomain [27]	5.22	9.96	77.08	83.79	
		FairVision [28]	2.16	4.63	79.76	82.73	
		Ours	**1.53**	**4.19**	**83.85**	**87.65**	
					Young	Elderly	
		Baseline	47.54	41.04	70.49	77.16	
		FairCLIP [6]	32.67	30.92	77.72	85.14	
	Age	FIN [26]	36.05	36.28	74.39	83.39	
		FairDomain [27]	31.41	24.71	74.33	82.42	
		FairVision [28]	34.09	31.15	75.64	84.01	
		Ours	**27.23**	**22.41**	**78.89**	**89.23**	

than the best-performing method (*e.g.* FairCLIP [6]), a significant improvement in the medical field as it directly reduces the risk of missed diagnoses, which is critical for patient care. On both the FairSeg and FairVLMed10k datasets, GlaucoDiff leads by at least 4.18% and 3.28%, respectively, in the F1 score compared with FairDomain [27]. The results in Table 2 show that GlaucoDiff enhances the classification AUC performance across all groups for each attribute and provides the most fair classification performance in terms of the DPD and DEOdds metrics. The improvements in MCC and QWK (Table 1), combined with consistent performance across imbalanced subgroups (Table 2), demonstrate GlaucoDiff's effectiveness and stability in medical image classification tasks. The *Baseline*, as shown in Tables 1 and 2, is the model that ablates all generated data and uses only the classification model to classify the source data. GlaucoDiff significantly outperforms the Baseline in both classification performance and fairness metrics, demonstrating the effectiveness of its high-quality synthetic samples for the glaucoma diagnosis task. Additionally, while not included in the tables for brevity, the Selecting Generated Samples strategy improves AUC and F1 by nearly 2% on both datasets, underscoring its contribution to model performance.

Evaluation of Generated Data. The quality of the generated images was evaluated by two ophthalmologists with 100 randomly selected synthetic images, vCDR-undisclosed. Over 90% of generated images had vCDR consistent with text prompts, as judged by experts. The combination of background and vCDR-controlled OC and OD regions ensures clinical relevance, reflecting glaucomatous and healthy characteristics. As shown in Fig. 3, GlaucoDiff generates

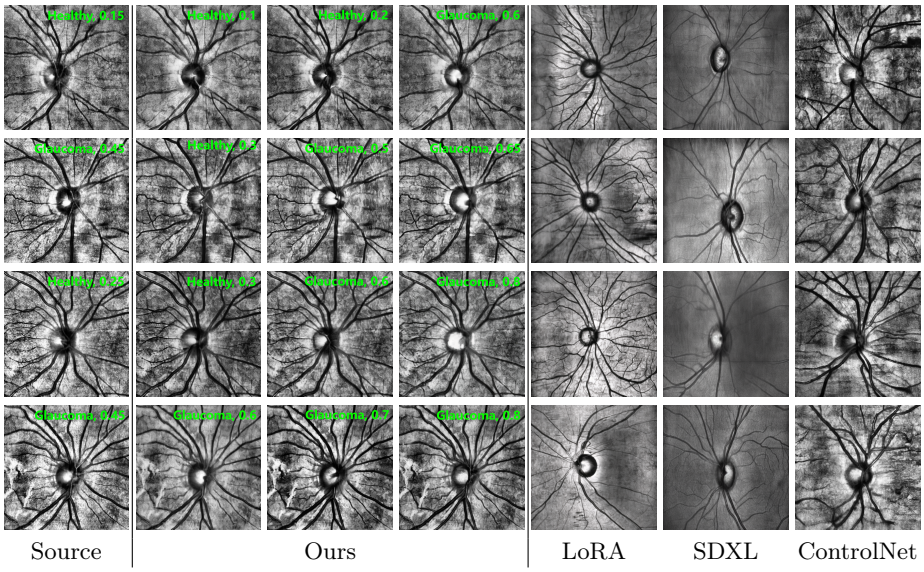

Fig. 3. Synthetic images generated by LoRA [29], SDXL [30], ControlNet [14], and Ours. Green labels indicate glaucoma status and vCDR values.

both glaucomatous and healthy samples, showcasing its flexibility in expanding datasets while outperforming other methods in image quality. GlaucoDiff also provides valuable visual support for glaucoma prognosis by generating synthetic samples with progressively increasing vCDR (Fig. 3, last row), simulating clearly disease progression. Compared to the images generated by ControlNet[14], which ablates GlaucoDiff's improvements (*e.g.*, background, vCDR control, and generated sample selection), using only OC and OD masks as conditions, our generated images exhibit superior clinical relevance and anatomical accuracy.

4 Conlusion

We introduce GlaucoDiff, a generative model that addresses data imbalance in fundus SLO-based glaucoma diagnosis through precise, vCDR-controlled image synthesis with preserved anatomical structure. Leveraging mature clinical OC and OD segmentation tools and the ControlNet-guided Stable Diffusion framework, GlaucoDiff is reproducible and suitable for real-world deployment. To ensure data quality, we adopt a sample selection strategy based on OC and OD alignment. Compared to existing methods, GlaucoDiff offers greater flexibility and produces clinically relevant images, as validated by expert assessments. Extensive experiments demonstrate consistent improvements in classification accuracy and fairness across overall and demographic-specific evaluations.

Acknowledgments. This work is supported by Y. Meng's The Royal Society Fund (IEC\NSFC\242172), United Kingdom.

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

References

1. Neto, A., Camera, J., Oliveira, S., Cláudia, A. & Cunha, A. Optic disc and cup segmentations for glaucoma assessment using cup-to-disc ratio. *Procedia Computer Science*. **196** pp. 485-492 (2022)
2. Chopra, R., Wagner, S. & Keane, P. Optical coherence tomography in the 2020s—outside the eye clinic. *Eye*. **35**, 236-243 (2021)
3. Tham, Y., Li, X., Wong, T., Quigley, H., Aung, T. & Cheng, C. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. **121**, 2081-2090 (2014)
4. Halawa, O., Jin, Q., Pasquale, L., Kang, J., Lorch, A., Sobrin, L., Miller, J., Li, Y., Eslami, M., Wang, M. & Others Race and ethnicity differences in disease severity and visual field progression among glaucoma patients. *American Journal Of Ophthalmology*. **242** pp. 69-76 (2022)
5. Meng, Y., Zhang, H., Zhao, Y., Gao, D., Hamill, B., Patri, G., Peto, T., Madhusudhan, S. & Zheng, Y. Dual consistency enabled weakly and semi-supervised optic disc and cup segmentation with dual adaptive graph convolutional networks. *IEEE Transactions On Medical Imaging*. **42**, 416-429 (2022)

6. Luo, Y., Shi, M., Khan, M., Afzal, M., Huang, H., Yuan, S., Tian, Y., Song, L., Kouhana, A., Elze, T. & Others Fairclip: Harnessing fairness in vision-language learning. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*. pp. 12289-12301 (2024)
7. Tian, Y., Shi, M., Luo, Y., Kouhana, A., Elze, T. & Wang, M. FairSeg: A Large-Scale Medical Image Segmentation Dataset for Fairness Learning Using Segment Anything Model with Fair Error-Bound Scaling. *The Twelfth International Conference On Learning Representations*.
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. Generative adversarial nets. *Advances In Neural Information Processing Systems*. **27** (2014)
9. Kingma, D. Auto-encoding variational bayes. *ArXiv Preprint ArXiv:1312.6114*. (2013)
10. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. High-resolution image synthesis with latent diffusion models. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*. pp. 10684-10695 (2022)
11. Kazerouni, A., Aghdam, E., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I. & Merhof, D. Diffusion models in medical imaging: A comprehensive survey. *Medical Image Analysis*. **88** pp. 102846 (2023)
12. Guo, J., Pang, Z., Yang, F., Shen, J. & Zhang, J. Study on the method of fundus image generation based on improved GAN. *Mathematical Problems In Engineering*. **2020**, 6309596 (2020)
13. Li, W., Xu, H., Zhang, G., Gao, H., Gao, M., Wang, M. & Zhao, H. Fairdiff: Fair segmentation with point-image diffusion. *International Conference On Medical Image Computing And Computer-Assisted Intervention*. pp. 617-628 (2024)
14. Zhang, L., Rao, A. & Agrawala, M. Adding conditional control to text-to-image diffusion models. *Proceedings Of The IEEE/CVF International Conference On Computer Vision*. pp. 3836-3847 (2023)
15. Sharma, V., Kumar, A., Jha, D., Bhuyan, M., Das, P. & Bagci, U. ControlPolypNet: Towards Controlled Colon Polyp Synthesis for Improved Polyp Segmentation. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*. pp. 2325-2334 (2024)
16. Zhang, H., Yang, J., Wan, S. & Fua, P. Lefusion: Synthesizing myocardial pathology on cardiac mri via lesion-focus diffusion models. *ArXiv Preprint ArXiv:2403.14066*. (2024)
17. Haider, A., Arsalan, M., Park, C., Sultan, H. & Park, K. Exploring deep feature-blending capabilities to assist glaucoma screening. *Applied Soft Computing*. **133** pp. 109918 (2023)
18. Veena, H., Muruganandham, A. & Kumaran, T. A novel optic disc and optic cup segmentation technique to diagnose glaucoma using deep learning convolutional neural network over retinal fundus images. *Journal Of King Saud University-Computer And Information Sciences*. **34**, 6187-6198 (2022)
19. Marques, R., De Jesus, D., Barbosa-Breda, J., Van Eijgen, J., Stalmans, I., Walsum, T., Klein, S., Vaz, P. & Brea, L. Automatic segmentation of the optic nerve head region in optical coherence tomography: A methodological review. *Computer Methods And Programs In Biomedicine*. **220** pp. 106801 (2022)
20. Fu, H., Cheng, J., Xu, Y., Wong, D., Liu, J. & Cao, X. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Transactions On Medical Imaging*. **37**, 1597-1605 (2018)

21. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *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 (2015)
22. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A., Lo, W. & Others Segment anything. *Proceedings Of The IEEE/CVF International Conference On Computer Vision*. pp. 4015-4026 (2023)
23. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. & Zhou, Y. Transunet: Transformers make strong encoders for medical image segmentation. *ArXiv Preprint ArXiv:2102.04306*. (2021)
24. Soorya, M., Issac, A. & Dutta, M. An automated and robust image processing algorithm for glaucoma diagnosis from fundus images using novel blood vessel tracking and bend point detection. *International Journal Of Medical Informatics*. **110** pp. 52-70 (2018)
25. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference On Machine Learning*. pp. 6105-6114 (2019)
26. Luo, Y., Tian, Y., Shi, M., Pasquale, L., Shen, L., Zebardast, N., Elze, T. & Wang, M. Harvard glaucoma fairness: a retinal nerve disease dataset for fairness learning and fair identity normalization. *IEEE Transactions On Medical Imaging*. (2024)
27. Tian, Y., Wen, C., Shi, M., Afzal, M., Huang, H., Khan, M., Luo, Y., Fang, Y. & Wang, M. Fairdomain: Achieving fairness in cross-domain medical image segmentation and classification. *European Conference On Computer Vision*. pp. 251-271 (2024)
28. Luo, Y., Tian, Y., Shi, M., Elze, T. & Wang, M. Eye fairness: A large-scale 3d imaging dataset for equitable eye diseases screening and fair identity scaling. *ArXiv Preprint ArXiv:2310.02492*. (2023)
29. Hu, E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. & Chen, W. Lora: Low-rank adaptation of large language models. *ArXiv Preprint ArXiv:2106.09685*. (2021)
30. Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J. & Rombach, R. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *ArXiv Preprint ArXiv:2307.01952*. (2023)