Diversified and Structure-realistic Fundus Image Synthesis for Diabetic Retinopathy Lesion Segmentation

Xiaoyi Feng[1*][0009−0000−7279−1874], Minqing Zhang[1*][0000−0002−7214−0569], Mengxian He[1,2][0009−0005−7585−374X], Mengdi Gao[1][0009−0003−5291−8294], Hao Wei[1][0000−0002−5719−8826], and Wu Yuan[1✉][0000−0001−9405−519X]

[1] The Chinese University of Hong Kong, Sha Tin, New Territories, Hong Kong.
wyuan@cuhk.edu.hk
[2] Johns Hopkins University, Baltimore, Maryland

Abstract. Automated diabetic retinopathy (DR) lesion segmentation aids in improving the efficiency of DR detection. However, obtaining lesion annotations for model training heavily relies on domain expertise and is a labor-intensive process. In addition to classical methods for alleviating label scarcity issues, such as self-supervised and semi-supervised learning, with the rapid development of generative models, several studies have indicated that utilizing synthetic image-mask pairs as data augmentation is promising. Due to the insufficient labeled data available to train powerful generative models, however, the synthetic fundus data suffers from two drawbacks: 1) unrealistic anatomical structures, 2) limited lesion diversity. In this paper, we propose a novel framework to synthesize fundus with DR lesion masks under limited labels. To increase lesion variation, we designed a learnable module to generate anatomically plausible masks as the condition, rather than directly using lesion masks from the limited dataset. To reduce the difficulty of learning intricate structures, we avoid directly generating images solely from lesion mask conditions. Instead, we developed an inpainting strategy that enables the model to generate lesions only within the mask area based on easily accessible healthy fundus images. Subjective evaluations indicate that our approach can generate more realistic fundus images with lesions compared to other generative methods. The downstream lesion segmentation experiments demonstrate that our synthetic data resulted in the most improvement across multiple network architectures, surpassing state-of-the-art methods.

Keywords: DR lesion segmentation · Labeled data scarcity · Data augmentation · Image-mask pair synthesis · Diffusion model

1 Introduction

Diabetic retinopathy (DR) is one of the most widespread ophthalmic disorders, with global DR prevalence projected to reach approximately 160 million

* Xiaoyi Feng and Minqing Zhang equally contribute to this paper.

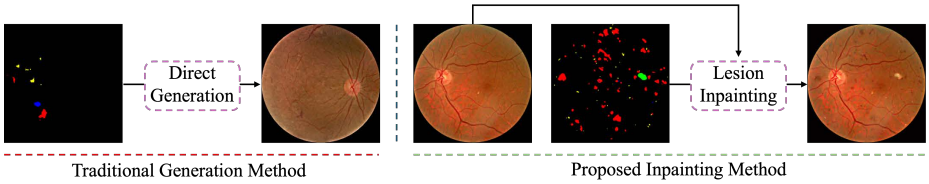

Fig. 1. Comparison of the pipeline for traditional conditional generative approaches and the proposed method.

by 2045 [14]. Since DR can lead to a heightened risk of irreversible vision impairment, early detection and appropriate treatments of DR hold great clinical value [10]. The screening of DR typically involves analyzing DR lesions using fundus images, including microaneurysms (MAs), hemorrhages (HEs), soft exudates (SEs), and hard exudates (EXs) [12]. Therefore, automatic DR lesion segmentation is an important computer-aided diagnosis application with the potential to enhance the efficiency and accuracy of clinical DR screening [5]. With the advancement of deep learning, many approaches based on deep neural networks [5, 6, 8, 9, 16] have been proposed to tackle the DR lesion segmentation task, but they suffer from insufficient segmentation accuracy. This is because training models with satisfactory performance demands abundant labeled data. However, the public availability of annotated datasets for DR lesions remains scarce, with even the largest benchmark FGADR [21] comprising only 1,741 annotated images. Such limited availability of annotated data poses considerable challenges for training well-performing segmentation methods.

Recently, generative AI has gained significant popularity, especially diffusion models [7]. Empirical studies suggest the utilization of synthetically generated labeled data for data augmentation exhibits promising potential to alleviate the paucity of publicly available training data for segmentation task [2,3,13,17]. However, in the domain of DR lesion segmentation, leveraging generative models to produce diverse and anatomically plausible synthetic labeled data is hindered by the limited availability of public data, giving rise to two critical challenges: 1) As illustrated in fig. 1, during inference phase, conventional generative approaches directly leverage the limited lesion masks available in the training data to guide the image synthesis process. However, the scarcity of segmentation masks in publicly available datasets poses a considerable challenge for generating diverse synthetic data. 2) As shown in fig. 1 traditional generative methods directly employ DR lesion segmentation annotations as conditions, attempting to generate complete fundus images containing both lesion regions and complex physiological structures in a single step. However, this approach demands substantial training data, as training on limited public datasets can lead to a lack of realism in generated physiological structures outside lesion areas.

Therefore, we proposed a two-stage realistic DR fundus image synthesis method, which is capable of synthesizing diverse DR images with realistic phys-

iological structures. In stage I, to synthesize anatomically plausible DR lesion masks, our framework utilizes the anatomical structure information from healthy fundus images, including the region of interest (ROI), optic disc (OD), and vessel (VE), to guide the synthesis of the lesion masks. In stage II, to alleviate the challenge of limited training data, we propose a *Learnable Training Augmentation* (LTA) module. This module enables augmenting our training data in a learnable manner during the training of the second stage. Simultaneously, to mitigate the challenge of generating complex physiological structures in fundus images, as illustrated in fig. 1, we devise an inpainting approach. Our approach generates DR lesions within the masked regions of readily available healthy fundus images, thereby circumventing the need to synthesize complex physiological structures.

In this study, our contributions are as follows: 1)We highlight two key drawbacks in generating paired DR lesion segmentation data, stemming from limited training data: insufficient lesion diversity and difficulties in synthesizing complex physiological structures. 2) To address the paucity of training data and two stemming drawbacks, we propose a two-stage approach for DR image generation that can synthesize diverse images with realistic physiological structures. 3) We perform comprehensive experiments validating the realism of our generated images and their efficacy in boosting downstream segmentation performance. Compared to state-of-the-art approaches, our method achieves the most realistic generation and highest segmentation gains.

2 Method

As illustrated in fig. 2, our approach consists of two stages that enable the synthesis of lesions at specified locations on healthy fundus images. In stage I, leveraging structural information as conditional inputs, we employ a diffusion model to generate diverse and anatomically plausible lesion masks. In stage II, guided by these generated masks, we task the conditional diffusion model to focus on generating lesions within the designated areas, avoiding the need to generate complex physiological structures. Concurrently, we introduce the LTA module, which enables learnable augmentation of the training data used in stage II training.

2.1 Structure Guided Mask Synthesis

Inference. The distribution and structure of DR lesions on fundus images are closely related to their structures, namely the ROI, OD, and VE [18]. As illustrated in fig. 2, in the inference phase of the first stage, to obtain anatomically plausible lesion masks $\tilde{y}$, we utilize a conditional diffusion model, leveraging the structures m as to generate diverse and anatomical coherent lesion masks $\tilde{y}$ from Gaussian noise y_T.

Training. For the training of this conditional diffusion model, the forward process is the same as the unconditional diffusion model. Through the Markov

chain, gaussian noise is gradually added to the original image y_0 for a total of T timesteps, this process can be formulated as $q(y_{1:T}|y_0)$:

$$\begin{aligned} q(y_t|y_{t-1}) &= \mathcal{N}(y_t; \sqrt{1-\beta_t}y_{t-1}, \beta_t\mathbf{I}) \\ q(y_{1:T}|y_0) &= \prod_{t=1}^{T} q((y_t|y_{t-1}) \end{aligned} \tag{1}$$

Where β_t is the schedule-defined constant, t denotes the timestep and y_t indicates the noised image at timestep t. With eq. (2), the corresponding y_t at any timestep t can be effectively sampled. Where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t := \prod_{i=1}^{T} \alpha_i$

$$y_t = \sqrt{\bar{\alpha}}y_0 + \sqrt{1-\bar{\alpha}}\epsilon; \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{2}$$

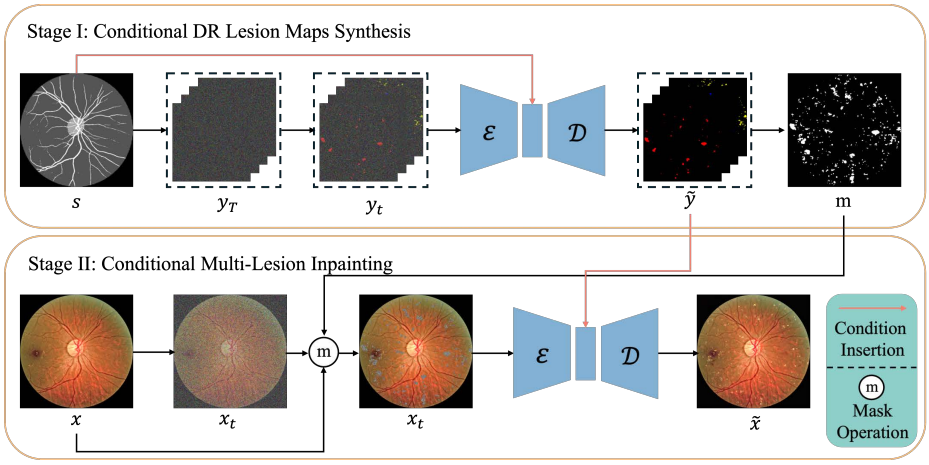

Fig. 2. An image to illustrate our two-stage method: Conditional lesion map generation and conditional image inpainting.

In the reverse process of training, sampled $y_t \in \mathbf{R}^{H\times W\times C}$ is fed into the network, while the structure $s \in \mathbf{R}^{H\times W}$ will be fed into the network as a condition. Therefore, the model is trained by minimizing the loss $\mathcal{L}_{condition}$ between the sampled noise ϵ and the noise $\epsilon_\theta(x_t, s)$ estimated by network ϵ_θ. Where the loss function $\mathcal{L}_{condition}$ is defined as follows:

$$\mathcal{L}_{condition} = E_{t,x,s}[\| \epsilon - \epsilon_\theta(x_t, s) \|] \tag{3}$$

2.2 Mask-guided Lesion Inpainting

Inference. In the inference phase of the second stage, to circumvent the need for the model to generate complex physiological structures, we employ the generated lesion masks as guidance, allowing the conditional diffusion model to focus

solely on generating lesions within the masked regions. Specifically, as shown in fig. 2, we first merge the generated four channel lesion masks $\tilde{y} \in \mathbf{R}^{H\times W\times 4}$ into a single binary mask $m \in \mathbf{R}^{H\times W}$. This mask m determines the regions that need to be modified in the healthy image x. With the well-trained inpainting model, we can generate the image with lesions $\tilde{x}$, where the region outside of m is the same as that of x and the region inside m contains corresponding lesions generated based on the four channel lesion maps. We follow [15] and use spatially adaptive normalization in the model architecture, This module can incorporate the masks as conditional inputs into the network decoder and guide the conditional generation.

Training. Obtaining a satisfactory inpainting model to inpaint lesions on the healthy image requires substantial data, yet public datasets for DR lesion segmentation are limited. Even the largest dataset, FGADR [21], contains only 1,741 images with lesion annotations. To enable the inpainting model to achieve effective performance with limited data, we devise the *Learnable Training Augmentation*, tailored to the characteristics of our inpainting approach, which can augment the for the training of inpainting model in a learnable manner.

According to the unique characteristics of the inpainting method, for each (x, y, m) in Our training dataset $D = \{x^n, y^n, m^n\}$, the region of x to be learned is determined by m, where $m \in \mathbf{R}^{H\times W} = \text{fuse}(y \in \mathbf{R}^{H\times W\times 4})$. For each channel c of lesion map y, which is denoted as y_c can be viewed as a combination of j connected regions r, where $y_c = \{r_j | j = 1, 2, 3, ...J\}$. Therefore, an approach to achieve data augmentation is, while training with each data pair (x, y, m), we randomly drop a proportion p of connected regions in y. Thus, by modifying y, we can obtain new data pairs (x, y_{new}, m_{new}) for training. However, p as a hyperparameter, manually searching for the optimal value of it is not reasonable and time-consuming. Therefore, we follow the insight of Bayesian optimization [4], optimizing p every epoch based on the LPIPS loss calculated during validation. Thus, the value of p can be upgraded and the LTA module can enable learnable augmentation for our training data with the optimized p. The approach can be formulated as the following equation:

$$p^* = \underset{p\in(0.1,1)}{\arg\min}\, f(p) \tag{4}$$

Where p stands for the dropping ratio and f(r) means the LPIPS loss calculated in validation. The range of p is set to $(0.1, 1)$. In the forward process of training, we generate a noisy image $\tilde{x_t}$ based on the original image x_0, and then replace the non-lesion regions of $\tilde{x_t}$ with the original image x_0 according to the mask m, resulting in the final noised image x_t. This process can be formulated as follows:

$$\begin{aligned} \tilde{x_t} &= \sqrt{\bar{\alpha}}x_0 + \sqrt{1-\bar{\alpha}}\epsilon; \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\ x_t &= \tilde{x_t} \odot m + x_0 \odot (1-m) \end{aligned} \tag{5}$$

Where the ϵ denotes the added Gaussian noise. In the reverse process of inpainting training, the noised image x_t is input into the network encoder ϵ, while the

lesion mask $y \in \mathcal{R}^{H \times W \times 4}$ is inserted into the network decoder $\mathcal{D}$ as a condition. The estimated noise is indicated as $\mathcal{D}(\varepsilon(x_t), y)$. The loss function of the inpainting model is as follows:

$$\mathcal{L}_{Inpainting} = E_{t,x_0,m,y}[\| \epsilon - \mathcal{D}(\varepsilon(x_t), y) \|] \tag{6}$$

3 Experiments

3.1 Implementation Details

Table 1. Comparison of image quality for different data generative methods.

Methods	Objective Metrics	Subjective Metrics			
	LPIPS (↓)	Accuracy (↓)	Recall (↓)	Precision (↓)	F1-score (↓)
RetinaGAN [8]	0.594	0.800	0.625	1.000	0.769
SDM [15]	0.572	0.700	0.714	1.000	0.833
Ours	**0.523**	**0.550**	**0.545**	**0.600**	**0.571**

Datasets. Our experiments were conducted on two public DR lesion segmentation datasets, FGADR [21] and OIA-DDR [11]. The FGADR dataset consists of 101 healthy images and 1741 images with lesion labels. The OIA-DDR dataset includes data for DR classification and DR lesion segmentation, with 757 images containing DR lesion annotations. The structure of the training images is obtained by fine-tuned SAM. To evaluate our method's effectiveness, we divided the FGADR dataset into 975 training, 244 validation, and 522 test images. For OIA-DDR, we split it into 383 training, 149 validation, and 225 test images. All images were margin-cropped and resized to 1024×1024.

Inpainting Training Protocol. During the training phase of our generative model. For each dataset, we utilized the training set for training and the validation set to compute the LIPIS loss between the generated images and the original images for validation. We set the diffusion step to 1000, batch size to 4, learning rate to 0.0001, and employed AdamW as the optimizer. Each stage of training the generative model was conducted on NVIDIA RTX 4090 GPUs.

Segmentation Experimental Settings. In this work, we conduct a comparative evaluation of our proposed approach against various data augmentation techniques across two different segmentation networks. The comparative methods encompass the traditional augmentation technique CutMix [19], as well as three data generation approaches: RetinaGAN [8], SDM [15], and Polyp-DDPM [2]. Two segmentation networks are Dense UNet [1], a medical segmentation network designed based on CNN, Trans2U-Net [20], a network specifically tailored for Retinal Lesion segmentation.

For the comparative experiments setting in table 2, each network's baseline and CutMix [19] methods are trained using the original training set. For the three generative approaches and the proposed approach, we generate the same

number of data as the training sets: 975 images for FGADR and 383 images for OIA-DDR. We then combine the generated data with the original training set to form new training sets for these methods, and the segmentation models are trained on these new training sets.

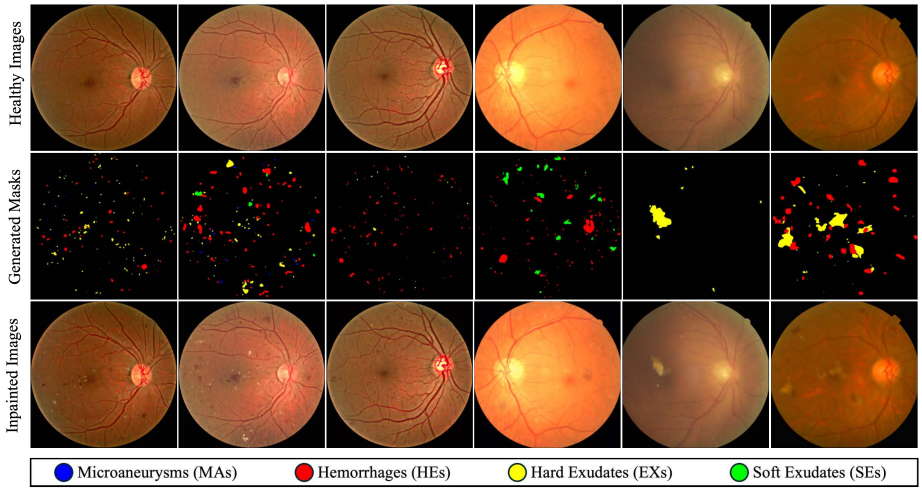

Fig. 3. Display of generated fundus images with lesions, healthy images in the first row, generated lesion maps in the second row, and the results of inpainting in the third row. The legend at the bottom indicates the mask color for four types of retinal lesions.

3.2 Image Quality Evaluation

To validate the realism of the images generated by our method, we evaluate it against two other well-performing methods using both objective and subjective metrics. As shown in table 1, for objective evaluation, we employ the LPIPS as the evaluation metrics, and our method outperforms the other two methods. For the subjective evaluation, we randomly selected 40 images for each method, with 20 real images and 20 generated images. Professional ophthalmologists are tasked to discriminate which image is generated. table 1 presents the Accuracy, Precision, Recall, and F1 scores for this subjective assessment. The results demonstrate that the synthetic images produced by our approach exhibit the highest degree of realism, most closely resembling the real data distribution, thereby posing significant challenges for discrimination.

3.3 Qualitative Analysis

To further illustrate the image quality of our proposed method, we present visualization results in fig. 3, where the left three columns and right three columns

Table 2. Comparison of different data augmentation methods for enhancing three lesion segmentation baselines.

Methods	FGADR [21]					OIA-DDR [11]				
	MA	HE	EX	SE	Mean	MA	HE	EX	SE	Mean
Dense UNet [1]	25.9	52.4	52.9	32.0	40.8	25.1	45.3	53.3	19.0	35.7
+Cutmix [19]	25.2	55.4	58.8	33.3	43.2 (+2.4)	24.3	46.0	58.0	19.1	36.8 (+1.1)
+Polyp-DDPM [2]	28.5	57.0	57.3	32.2	43.8 (+3.0)	27.4	48.4	56.2	18.3	37.6 (+1.9)
+SDM [15]	28.3	58.5	60.3	35.5	45.7 (+4.9)	27.1	49.8	58.3	19.5	38.7 (+3.0)
+RetinaGAN [8]	26.5	57.8	59.1	33.2	44.2 (+3.4)	26.3	48.2	58.4	19.1	38.0 (+2.3)
+Ours (w/o LTA)	27.9	59.2	60.2	34.8	45.5 (+4.7)	28.4	51.6	61.3	21.1	40.6 (+4.9)
+Ours	**29.0**	**61.1**	**61.8**	**37.9**	**47.5** (+**6.7**)	**30.6**	**53.9**	**63.9**	**23.7**	**43.0** (+**7.3**)
Trans2U-Net [20]	32.7	66.0	61.4	35.8	49.0	32.6	53.7	54.7	41.3	45.6
+Cutmix [19]	32.4	66.8	63.1	36.3	49.7 (+0.7)	32.8	54.9	56.2	41.6	46.4 (+0.8)
+Polyp-DDPM [2]	33.2	67.9	64.3	36.6	50.5 (+1.5)	33.4	56.0	55.8	41.8	46.8 (+1.2)
+SDM [15]	33.4	68.5	65.3	37.4	51.2 (+2.2)	33.1	56.9	57.3	42.6	47.5 (+1.9)
+RetinaGAN [8]	31.9	67.2	64.8	36.5	50.1 (+1.1)	32.7	55.8	58.4	42.3	47.3 (+1.7)
+Ours (w/o LTA)	32.7	69.4	65.8	38.0	52.0 (+3.0)	34.6	58.2	59.7	43.9	49.1 (+3.5)
+Ours	**33.9**	**72.1**	**67.1**	**39.6**	**53.2** (+**4.2**)	**36.2**	**60.4**	**61.1**	**45.8**	**50.9** (+**5.3**)

indicate the images produced by our models trained on FGADR and OIA-DDR datasets, respectively. As shown in fig. 3, our method generates diverse and anatomic plausible masks. Subsequently, our method inpaint lesions within masked regions while preserving the complex retinal physiological structure. Moreover, the distinction between lesion classes is noticeable.

3.4 Downstream Lesion Segmentation Task

The Dice scores of the downstream segmentation result are presented in table 2. Our proposed method consistently outperforms other data augmentation methods in enhancing the segmentation performance across two networks. Specifically, on the FGADR dataset, our method improved the performance of Dense UNet by 6.7%, surpassing the second-best method, SDM, by 1.8%. Moreover, our method enhances Trans2U-Net gain a 4.2% improvement, surpassing the second-best method, SDM, by 2.0%. On the OIA-DDR dataset, our method resulted in a 7.3% performance gain for Dense UNet, surpassing the second-best method, SDM, by 4.3%. Furthermore, our method contributed to a 5.3% enhancement for Trans2U-Net, surpassing the second-best method, SDM, by 3.4%. As illustrated in table 2, we conducted ablation experiments to evaluate the effectiveness of the LTA module. Our experimental results demonstrate that the LTA module consistently improves the performance of segmentation models on both of the two datasets.

4 Conclusion

In this work, we underscore two critical limitations in generating paired DR lesion segmentation data when faced with scarce training data and stemming

drawbacks: The lack of lesion variability and the challenges in synthesizing intricate physiological structures. To mitigate the limitations of data scarcity in DR lesion segmentation, we proposed a novel pipeline that can transform healthy fundus images into realistic lesion-containing images under limited public data. The extensive experiment results on two DR lesion segmentation datasets and evaluation of the image quality of different generative methods demonstrate that our method outperforms others in terms of both image quality and boosting the performance of segmentation models.

5 Acknowledgments

This work was supported in part by the Research Grants Council (RGC) of Hong Kong SAR (GRF14203821, GRF14216222), the Innovation and Technology Fund (ITF) of Hong Kong SAR (ITS/240/21, ITS/252/23), and the Science, Technology, and Innovation Commission (STIC) of Shenzhen Municipality (SGDX20220530111005039), and the Strategic Seed Funding for Collaborative Research Scheme of CUHK (SSFCRS 3133341).

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

References

1. Cai, S., Tian, Y., Lui, H., Zeng, H., Wu, Y., Chen, G.: Dense-unet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quantitative imaging in medicine and surgery **10**(6), 1275 (2020)
2. Dorjsembe, Z., Pao, H.K., Xiao, F.: Polyp-ddpm: Diffusion-based semantic polyp synthesis for enhanced segmentation. arXiv preprint arXiv:2402.04031 (2024)
3. Du, Y., Jiang, Y., Tan, S., Wu, X., Dou, Q., Li, Z., Li, G., Wan, X.: Arsdm: colonoscopy images synthesis with adaptive refinement semantic diffusion models. In: International conference on medical image computing and computer-assisted intervention. pp. 339–349. Springer (2023)
4. Frazier, P.I.: A tutorial on bayesian optimization. arXiv preprint arXiv:1807.02811 (2018)
5. Guo, T., Yang, J., Yu, Q.: Diabetic retinopathy lesion segmentation using deep multi-scale framework. Biomedical Signal Processing and Control **88**, 105050 (2024)
6. He, A., Wang, K., Li, T., Bo, W., Kang, H., Fu, H.: Progressive multiscale consistent network for multiclass fundus lesion segmentation. IEEE transactions on medical imaging **41**(11), 3146–3157 (2022)
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
8. Hou, B.: High-fidelity diabetic retina fundus image synthesis from freestyle lesion maps. Biomedical Optics Express **14**(2), 533–549 (2023)
9. Huang, S., Li, J., Xiao, Y., Shen, N., Xu, T.: Rtnet: relation transformer network for diabetic retinopathy multi-lesion segmentation. IEEE Transactions on Medical Imaging **41**(6), 1596–1607 (2022)

10. Lawrenson, J., Bourmpaki, E., Bunce, C., Stratton, I., Gardner, P., Anderson, J., Group, E.S.: Trends in diabetic retinopathy screening attendance and associations with vision impairment attributable to diabetes in a large nationwide cohort. Diabetic Medicine **38**(4), e14425 (2021)
11. Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., Kang, H.: Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. Information Sciences **501**, 511 – 522 (2019). https://doi.org/https://doi.org/10.1016/j.ins.2019.06.011, http://www.sciencedirect.com/science/article/pii/S0020025519305377
12. Salamat, N., Missen, M.M.S., Rashid, A.: Diabetic retinopathy techniques in retinal images: A review. Artificial intelligence in medicine **97**, 168–188 (2019)
13. Shrivastava, A., Fletcher, P.T.: Nasdm: Nuclei-aware semantic histopathology image generation using diffusion models. arXiv preprint arXiv:2303.11477 (2023)
14. Teo, Z.L., Tham, Y.C., Yu, M., Chee, M.L., Rim, T.H., Cheung, N., Bikbov, M.M., Wang, Y.X., Tang, Y., Lu, Y., et al.: Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. Ophthalmology **128**(11), 1580–1591 (2021)
15. Wang, W., Bao, J., Zhou, W., Chen, D., Chen, D., Yuan, L., Li, H.: Semantic image synthesis via diffusion models. arXiv preprint arXiv:2207.00050 (2022)
16. Yin, M., Soomro, T.A., Jandan, F.A., Fatihi, A., Ubaid, F.B., Irfan, M., Afifi, A.J., Rahman, S., Telenyk, S., Nowakowski, G.: Dual-branch u-net architecture for retinal lesions segmentation on fundus image. IEEE Access **11**, 130451–130465 (2023)
17. Yu, X., Li, G., Lou, W., Liu, S., Wan, X., Chen, Y., Li, H.: Diffusion-based data augmentation for nuclei image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 592–602. Springer (2023)
18. Yu, Z., Yan, R., Yu, Y., Ma, X., Liu, X., Liu, J., Ren, Q., Lu, Y.: Multiple lesions insertion: boosting diabetic retinopathy screening through poisson editing. Biomedical Optics Express **12**(5), 2773–2789 (2021)
19. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6023–6032 (2019)
20. Zhang, L., Fang, Z., Li, T., Xiao, Y., Zhou, J.T., Yang, F.: Retinal multi-lesion segmentation by reinforcing single-lesion guidance with multi-view learning. Biomedical Signal Processing and Control **86**, 105349 (2023)
21. Zhou, Y., Wang, B., Huang, L., Cui, S., Shao, L.: A benchmark for studying diabetic retinopathy: segmentation, grading, and transferability. IEEE Transactions on Medical Imaging **40**(3), 818–828 (2020)