BEAM: Boosting Fundus Image Enhancement via Adapted Text-to-Image Models

Ziheng Wang[1(✉)],Pujin Cheng[2,3], and Xiaoying Tang[2]

[1] School of Artificial Intelligence,
The Chinese University of Hong Kong, Shenzhen, China
zihengwang3@link.cuhk.edu.cn
[2] Department of Electronic and Electrical Engineering,
Southern University of Science and Technology, Shenzhen, China
[3] Department of Electrical and Electronic Engineering,
The University of Hong Kong, Hong Kong SAR, China

Abstract. High-quality fundus images are crucial in the diagnosis of ophthalmic diseases. However, these images in real-world settings often suffer from degradation due to motion blur, illumination irregularities, and artifacts. Existing enhancement methods that rely on paired datasets or simplified degradation models have difficulty addressing the complex degradations commonly observed in clinical realities. We state that pre-trained large-scale text-to-image models contain rich image priors to enhance fundus images to high-quality ones, and we can take low-quality images directly as input with the skip-connection maintaining structure consistency, reducing the ambiguity brought from random noise sampling while simultaneously eliminating the need for additional controlling modules. We then fine-tune the pre-trained network in a single step with a small fraction of trainable parameters to adapt it to the fundus image enhancement task, and show the superiority of BEAM through extensive experiments over other state-of-the-art approaches. Moreover, we expanded our framework to an unpaired scheme and showcased its capacity to generate a realistic paired simulation dataset. The source code and dataset are available at https://github.com/wangzh1/BEAM.

Keywords: Fundus Image Enhancement · Text-to-Image Models · Diffusion Models

1 Introduction

The quality of fundus images is crucial for the accurate diagnosis and management of various ophthalmic conditions. The ability to identify subtle signs of disease greatly relies on clear and high-quality fundus images, in turn enabling early diagnosis and more effective treatment [6]. However, obtaining such images is often challenging due to factors that degrade image quality, such as glare from light sources, patient movement causing motion blur, poor illumination, and

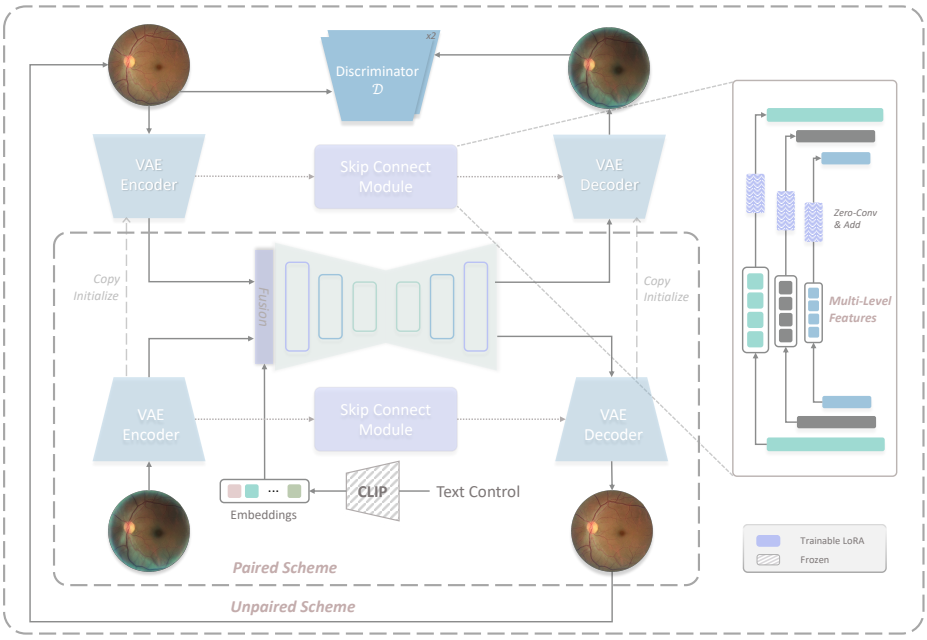

Fig. 1: **Pipeline of BEAM.** BEAM is a framework using a pre-trained latent diffusion model, VAE encoder-decoder, and text encoder for fundus image enhancement. It takes low-quality images as input, retrains with a skip connect module for detail retention, and supports unpaired training via introducing duplicated trainable VAE and adversarial training.

retinal artifacts. These degradations can obscure critical diagnostic information and hinder the performance of automated analysis systems.

To address these issues, traditional solutions often rely on operations in the transform domain. For instance, Cheng et al. [2] applied a structure-preserving guided filtering technique to deblur fundus images, while Cao et al. [1] removed low-frequency components in the root domain to obtain clearer results. However, these methods struggle with more complex degradations. Therefore, recent research has shifted towards learning-based enhancement approaches, focusing both on realistically simulating the degradation process to synthesize paired datasets and on effectively enhancing low-quality images. Shen et al. proposed an restoration network with an artificial degradation pipeline on frequency domain [22] to synthesize paired dataset, but it was not enough to simulate complex realistic degradation factors. Using this degradation pipeline, Liu et al. developed the Pyramid Constraint Network [14] to improve the representation of clinically significant features. Similarly, Li et al. introduced the Structure-Consistent Restoration Network [12] for cataract fundus images, grounding their approach in preserving high-frequency component consistency. Li et al. [13] and

Cheng et al. [4] introduced frequency self-supervision and importance-guided self-supervision with synthesized data supervision, respectively, to further improve the network's ability of domain adaption. However, these methods are all strongly biased by the artificial degradation simulation. Cheng et al. [3] utilized a diffusion model to learn the enhancement process, but it also needed a data-driven degradation model and was quantitatively inferior to other state-of-the-art (SOTA) methods in terms of low-level features. On the other hand, Zhao et al. [28] directly employed an unpaired training manner to enhance fundus images using Generative Adversarial Network and cycle consistency [29].

We realize that due to the significant intersections between fundus image enhancement and clinical knowledge, this challenge not only lies at a low-level processing layer, but also needs high-level features and semantic understandings. Recent development of text-to-image (T2I) models in the computer vision field [18, 20] offer a promising avenue for this task. Therefore, we propose to take advantage of their rich diffusion priors to better facilitate the transfer between low-quality and high-quality fundus images. In this paper, we dedicate to adapt pre-trained T2I models to enhance fundus images. Successful applications adapting T2I diffusion for downstream tasks [23, 24] introduced additional modules such as ControlNet and regularizers. While these techniques maintain the efficiency of the original latent representation, they complicate the process by duplicating network parameters, significantly increasing memory overhead. To further reduce computational cost, we replace input noise with image itself and utilize skip-connections, which have proven to be effective in preserving detailed features [19]. However, this shift requires retraining in the image space, which can be costly with hundreds or thousands of steps of fine-tuning. Leveraging recent advances in few-step T2I diffusion using distillation and adversarial training [21] and successful cases of one-step diffusion fine-tuning [23, 25], we also demonstrate the effectiveness of single-step adaptation in our task. Moreover, we show that our pipeline can be easily extended to an unpaired scheme by duplicating another pair of VAE and introduce adversarial training.

The main contribution of BEAM can be summarized as follows: (1) We present a simple but effective pipeline to adapt T2I models to the task of fundus image enhancement without changing the main structure or adding additional controlling modules. (2) We demonstrate that fine-tuning a very small fraction of the parameters in a pre-trained T2I model is sufficient to outperform existing methods. (3) Through unpaired training, we use BEAM to synthesize a realistic degraded fundus image dataset and make it public.

2 Method

In this section, we detail how we adapt a pre-trained text-to-image model for fundus image enhancement. An overview of the BEAM framework is presented in Fig. 1. By adapting the structure of the pre-trained text-to-image model, our method originally supports paired training, when high-quality and low-quality image pairs are available. Additionally, we extend the framework by introduc-

ing an duplicated VAE encoder-decoder pair and discriminators for adversarial training.

2.1 BEAM Pipeline

Our approach provides a simple yet efficient solution for adapting text-to-image models to enhance fundus images. We employ a distilled version of Stable Diffusion 2.1 [21] as the backbone, retaining its core components: a frozen CLIP [17] text encoder, a VAE encoder-decoder pair, and a latent diffusion network. To maintain structural integrity, such as the preservation of blood vessels, we introduce a skip-connect module that has been proved to be effective [19] to transfer features from the VAE encoder to the decoder.

Unlike conventional diffusion methods that initialize with random noise, our pipeline directly uses the degraded image as input. This eliminates the need for additional modules such as ControlNet [26], significantly reduces the number of trainable parameters. Inspired by the success of Low-Rank Adaptation (LoRA) [9] in fine-tuning large text-to-image diffusion models for various downstream tasks [16, 23, 24], we also adopt LoRA to accelerate the training process. Furthermore, since the original backbone was trained using adversarial diffusion distillation with very few steps and has proven effective with a single step [21], we implement a single-step training approach, which also boosts the training efficiency. Together, these make it possible for BEAM to outperform SOTA methods with only 3M trainable parameters.

2.2 Paired Training

In the paired training scheme, BEAM is trained for an enhancement mapping $\mathcal{T} : \mathcal{X} \to \mathcal{Y}$ using pairs of high-quality and synthetically degraded fundus images.

Loss Functions We use the Learned Perceptual Image Patch Similarity (LPIPS) loss [27] to measure perceptual differences between the enhanced and high-quality images. To preserve fine details, such as the sharpness of blood vessels, we introduce a high-frequency loss by applying a high-pass filter $\mathcal{H}$ to both the enhanced image $\hat{\mathbf{x}}$ and the ground truth $\mathbf{y}$, minimizing their L-1 difference:

$$\mathcal{L}_{\text{high-freq}} = \|\mathcal{H}(\hat{\mathbf{x}}) - \mathcal{H}(\mathbf{y})\|_1 . \tag{1}$$

To prevent over-enhancement and ensure stability when processing already high-quality images, we include an identity loss to penalize deviations from the input when enhancement is unnecessary:

$$\mathcal{L}_{\text{idt}} = \mathbb{E}_{\mathbf{y}\sim\mathcal{Y}} \left[\|\mathcal{T}(\mathbf{y}) - \mathbf{y}\|_1\right], \tag{2}$$

where $\mathcal{T}$ denotes the enhancement function. The total loss for paired training is a weighted combination of these terms:

$$\mathcal{L}_{\text{paired}} = \lambda_{\text{LPIPS}}\mathcal{L}_{\text{LPIPS}} + \lambda_{\text{high-freq}}\mathcal{L}_{\text{high-freq}} + \lambda_{\text{idt}}\mathcal{L}_{\text{idt}}. \tag{3}$$

2.3 Unpaired Training

Our unpaired training includes two domain translations $\mathcal{T}(\mathbf{x}, t_{\mathcal{Y}}) : \mathcal{X} \to \mathcal{Y}$ and $\mathcal{T}(\mathbf{y}, t_{\mathcal{X}}) : \mathcal{Y} \to \mathcal{X}$, where $t_{(.)}$ denotes the text description for the corresponding target domain.

Cycle Consistency Loss We utilize a pixel-wise loss $\mathcal{L}_{\text{rec}}$ combining L-1 norm and LPIPS [27] between input image $\mathbf{x}$ and reconstructed $\hat{\mathbf{x}}$ from $\mathbf{y}$:

$$\mathcal{L}_{\text{rec}}(\mathbf{x}, \hat{\mathbf{x}}) = \lambda_1 \left\|\mathbf{x} - \hat{\mathbf{x}}\right\|_1 + \lambda_{\text{LPIPS}} \mathcal{L}_{\text{LPIPS}}(\mathbf{x}, \hat{\mathbf{x}}). \tag{4}$$

The cycle consistency loss can be described as:

$$\begin{aligned} \mathcal{L}_{\text{cycle}} = \mathbb{E}_{\mathbf{x}\sim\mathcal{X}} \left[\mathcal{L}_{\text{rec}} \left(\mathcal{T}\left(\mathcal{T}\left(\mathbf{x}, t_{\mathcal{Y}}\right), t_{\mathcal{X}}\right), \mathbf{x}\right)\right] \\ + \mathbb{E}_{\mathbf{y}\sim\mathcal{Y}} \left[\mathcal{L}_{\text{rec}} \left(\mathcal{T}\left(\mathcal{T}\left(\mathbf{y}, t_{\mathcal{X}}\right), t_{\mathcal{Y}}\right), \mathbf{y}\right)\right]. \end{aligned} \tag{5}$$

Adversarial Loss We employ adversarial losses [7], with discriminators $\mathcal{D}_{\mathcal{X}}$ and $\mathcal{D}_{\mathcal{Y}}$ using CLIP backbone [11]. The adversarial loss for domain mapping function $\mathcal{X} \to \mathcal{Y}$ can then be defined as:

$$\mathcal{L}_{\text{adv}}^{\mathcal{X}\to\mathcal{Y}} = \mathbb{E}_{\mathbf{y}\sim\mathcal{Y}} \left[\log \mathcal{D}_{\mathcal{Y}}(\mathbf{y})\right] + \mathbb{E}_{\mathbf{x}\sim\mathcal{X}} \left[\log\left(1 - \mathcal{D}_{\mathcal{Y}}\left(\mathcal{T}\left(\mathbf{x}, t_{\mathcal{Y}}\right)\right)\right)\right]. \tag{6}$$

Identity Loss To ensure that the model does not over-enhance or distort images unnecessarily, we incorporate an identity loss:

$$\mathcal{L}_{\text{idt}} = \mathbb{E}_y[\mathcal{L}_{\text{rec}}(\mathcal{T}(\mathbf{y}, t_{\mathcal{X}}), \mathbf{y})] + \mathbb{E}_x[\mathcal{L}_{\text{rec}}(\mathcal{T}(\mathbf{x}, t_{\mathcal{Y}}), \mathbf{x})]. \tag{7}$$

Total Loss The total loss function for BEAM is a combination of the losses with corresponding coefficient λ:

$$\mathcal{L}_{\text{unpaired}} = \mathcal{L}_{\text{cycle}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{idt}} \mathcal{L}_{\text{idt}}. \tag{8}$$

3 Experiments

3.1 Training BEAM

Dataset. We utilize the EyeQ dataset [6], which comprises 28,792 fundus images categorized into three quality grades: Good, Usable, and Reject. For paired training, we select all images labeled as "Good" to serve as our high-quality dataset and generate a corresponding low-quality image by applying synthetic degradations, including Gaussian blur, additive noise, and contrast reduction, based on the degradation model [22]. These degradation parameters are randomly sampled to mimic diverse real-world degradation scenarios. The dataset is divided into training and test sets following the original EyeQ split.

Table 1: **Comparison with SOTA supervised methods on paired dataset.** Note that BEAM-B and BEAM-S both have ∼1B parameters in total.

Method	Trainable Params ↓	Full-Reference PSNR ↑	SSIM ↑	VSD ↑	DRA ↑	Non-Reference FIQA ↑
Original	–	–	–	–	0.7236	0.1502
Degraded [22]	–	19.37	0.7794	0.7189	0.5841	–
pix2pix [10]	211M	24.28	0.7619	0.6521	0.6146	0.5093
SCRNet [12]	341M	27.84	0.8594	0.7244	0.6465	0.7364
GFENet [13]	341M	28.79	0.8759	0.7367	0.6849	0.7499
I-SECRET [4]	151M	28.22	0.8692	0.7351	0.6823	0.8025
PCENet [14]	102M	27.43	0.8471	0.7367	0.6245	0.7163
BEAM-S (ours)	**3M**	29.58	0.8946	**0.7824**	0.7068	0.7910
BEAM-B (ours)	133M	**30.08**	**0.8982**	**0.7824**	**0.7104**	**0.8255**

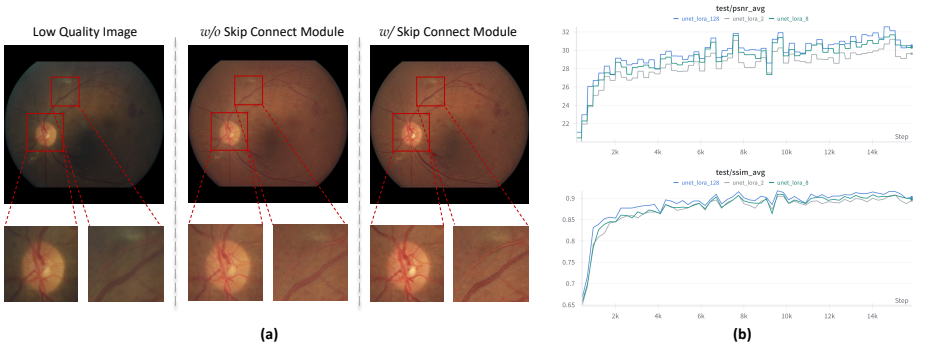

Fig. 2: **(a)** Performance comparison of intermediate training result with and without skip connect module. The skip connect module keeps more structure details and improve the sharpness of the enhanced vessels. **(b)** Quantitative comparison between different sizes of BEAM.

Implementation Details. We utilize PyTorch-Lightning [5] as the framework for training BEAM. We use the AdamW optimizer with a base learning rate of 1.5×10^{-5} and a weight decay of 1×10^{-2}. All training processes are conducted on NVIDIA A40 GPU(s). For the paired scheme, the model is trained for 30 epochs with a batch size of 4 on 4 GPUs within 15 hours. For the unpaired scheme, the model is trained for 5 epochs with a batch size of 1 on a single GPU within 19 hours. We use LoRA rank 2 for training BEAM-S, while rank 128 for BEAM-B. For both model, we set the LoRA rank of VAE to 4. For more details, please refer to our code release.

3.2 Evaluation

We first conduct ablation experiments with different sizes of BEAM to show the power of the image priors embedded in the pre-trained text-to-image model.

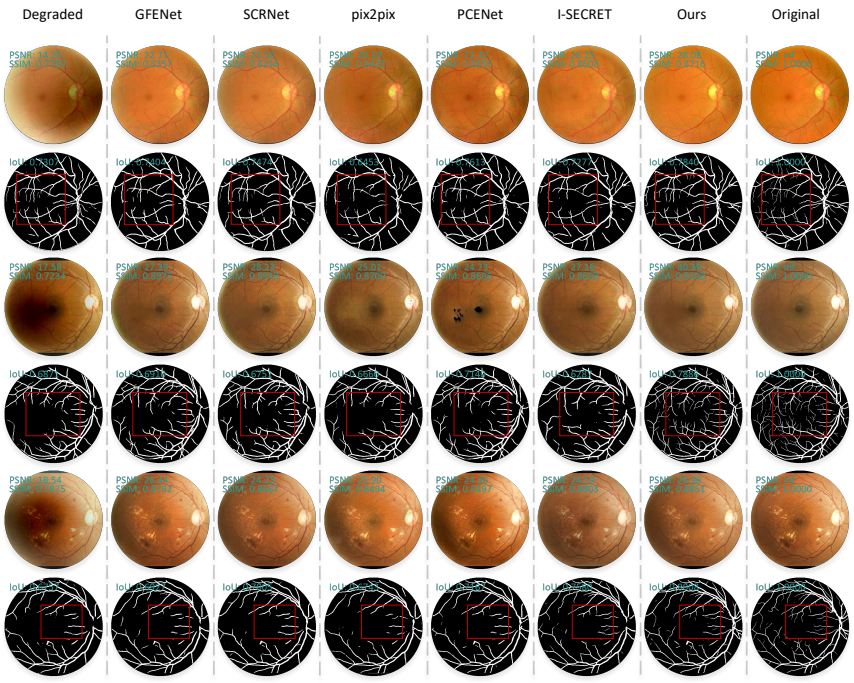

Fig. 3: Comparison of retinal fundus images before and after enhancement by different methods: GFENet [13], SCRNet [12], pix2pix [10], PCENet [14], and I-SECRET [4].

Specifically, we train BEAM with LoRA ranks of UNet ranging from 2 to 128. Selected performance curves are as shown in Fig. 2(b). The results demonstrate that fine-tuning only a minimal number of parameters enables us to effectively adapt a pre-trained text-to-image model for our task of fundus image enhancement. Furthermore, in Fig. 2(a), we highlight the effectiveness of the skip connect module in preserving detailed structures.

Paired Evaluation For full-reference evaluation, we conduct quantitative analysis on the test set mentioned in Sec. 3.1, against SOTA approaches requiring paired datasets (e.g., pix2pix [10], SCRNet [12], GFENet [13], I-SECRET [4], PCENet [14]). We employ PSNR and SSIM image quality metrics [8]. Furthermore, we segment images with a segmentation network [15] and use Vessel Segmentation Dice (VSD) to assess the enhancement of the structural vascular details. To quantify clinically meaningful enhancement effects, we measure Diabetic Retinopathy detection Accuracy (DRA) using the DR detection method stated in [3]. Fig. 3 shows sample comparisons between other SOTA methods. Our

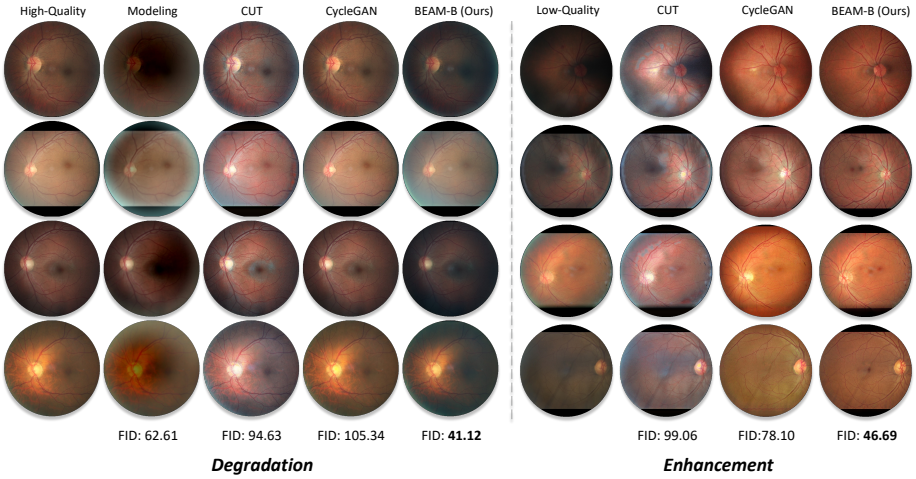

Fig. 4: **Evaluation of unpaired training.** Left: Comparison of degradation results; Right: Comparison of enhancement results.

BEAM is particularly superior to other methods in terms of structural details and achieves better metric results.

For non-reference evaluation, we use images labeled as "Usable" from the EyeQ test set and apply MCF-Net [6] to evaluate quality. The proportion of enhanced images assessed as "Good" was defined as the Fundus Image Quality Assessment Score (FIQA).

Unpaired Evaluation Generating realistic low-quality images through degradation is also crucial for creating paired datasets to train robust enhancement models. So, we evaluate BEAM on both degradation and enhancement tasks. For a fair comparison, we select unpaired domain translation methods, CycleGAN [29] and CUT [30]. Fig.4 demonstrates BEAMs superiority in both tasks. Due to complex degradation, CUT struggles to deblur low-quality images, while CycleGAN fails to preserve vessel structure consistency. In contrast, BEAM effectively removes low-level degradation while retaining vessel details. We also compare BEAM with a mathematical modeling method [22] in degradation. Qualitative results reveal that BEAM generates the most realistic outcomes in both degradation and enhancement, while achieving the lowest FID score.

4 Conclusion

This paper introduces BEAM, a framework adapting pre-trained text-to-image models for fundus image enhancement. The proposed BEAM directly takes image as input without introducing random noise or adding controlling branches, thus allowing for fine-tuning in a single step with minimal parameters while

outperforming SOTA methods. Furthermore, through extending BEAM to unpaired training, it performs realistic transformation between high-quality and low-quality fundus images, also beneficial for creating paired datasets.

Acknowledgments. This work was supported in part by the HPC Platform of ShanghaiTech University.

References

1. Cao, L., Li, H., Zhang, Y.: Retinal image enhancement using low-pass filtering and -rooting. Signal Processing **170**, 107445 (2020)
2. Cheng, J., Li, Z., Gu, Z., Fu, H., Wong, D.W.K., Liu, J.: Structure-preserving guided retinal image filtering and its application for optic disk analysis. IEEE Transactions on Medical Imaging **37**(11), 2536–2546 (2018)
3. Cheng, P., Lin, L., Huang, Y., He, H., Luo, W., Tang, X.: Learning Enhancement From Degradation: A Diffusion Model For Fundus Image Enhancement (Mar 2023). https://doi.org/10.48550/arXiv.2303.04603
4. Cheng, P., Lin, L., Huang, Y., Lyu, J., Tang, X.: I-SECRET: Importance-Guided Fundus Image Enhancement via Semi-supervised Contrastive Constraining. In: De Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, vol. 12908, pp. 87–96. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-87237-3_9
5. Falcon, W., The PyTorch Lightning team: PyTorch Lightning (Mar 2019). https://doi.org/10.5281/zenodo.3828935, https://github.com/Lightning-AI/lightning
6. Fu, H., Wang, B., Shen, J., Cui, S., Xu, Y., Liu, J., Shao, L.: Evaluation of retinal image quality assessment networks in different color-spaces. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. pp. 48–56. Springer International Publishing, Cham (2019)
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) Advances in Neural Information Processing Systems. vol. 27. Curran Associates, Inc. (2014)
8. Hore, A., Ziou, D.: Image Quality Metrics: PSNR vs. SSIM. In: 2010 20th International Conference on Pattern Recognition. pp. 2366–2369. IEEE, Istanbul, Turkey (Aug 2010). https://doi.org/10.1109/ICPR.2010.579
9. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-Rank Adaptation of Large Language Models (Oct 2021). https://doi.org/10.48550/arXiv.2106.09685
10. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on (2017)
11. Kumari, N., Zhang, R., Shechtman, E., Zhu, J.Y.: Ensembling off-the-shelf models for gan training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2022)

12. Li, H., Liu, H., Fu, H., Shu, H., Zhao, Y., Luo, X., Hu, Y., Liu, J.: Structure-consistent restoration network for cataract fundus image enhancement (2022), https://arxiv.org/abs/2206.04684
13. Li, H., Liu, H., Fu, H., Xu, Y., Shu, H., Niu, K., Hu, Y., Liu, J.: A generic fundus image enhancement network boosted by frequency self-supervised representation learning. arXiv preprint arXiv:2309.00885 (2023)
14. Liu, H., Li, H., Fu, H., Xiao, R., Gao, Y., Hu, Y., Liu, J.: Degradation-invariant enhancement of fundus images via pyramid constraint network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 507–516. Springer (2022)
15. 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 (Sep 2022). https://doi.org/10.1109/JBHI.2022.3188710
16. Parmar, G., Park, T., Narasimhan, S., Zhu, J.Y.: One-Step Image Translation with Text-to-Image Models (Mar 2024). https://doi.org/10.48550/arXiv.2403.12036
17. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision (2021), https://arxiv.org/abs/2103.00020
18. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models (2021)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015), https://arxiv.org/abs/1505.04597
20. Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S.K.S., Ayan, B.K., Mahdavi, S.S., Lopes, R.G., Salimans, T., Ho, J., Fleet, D.J., Norouzi, M.: Photorealistic text-to-image diffusion models with deep language understanding (2022), https://arxiv.org/abs/2205.11487
21. Sauer, A., Lorenz, D., Blattmann, A., Rombach, R.: Adversarial Diffusion Distillation. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) Computer Vision – ECCV 2024, vol. 15144, pp. 87–103. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-73016-0_6
22. Shen, Z., Fu, H., Shen, J., Shao, L.: Modeling and enhancing low-quality retinal fundus images. IEEE Transactions on Medical Imaging **40**(3), 996–1006 (2021). https://doi.org/10.1109/TMI.2020.3043495
23. Wu, R., Sun, L., Ma, Z., Zhang, L.: One-Step Effective Diffusion Network for Real-World Image Super-Resolution (Oct 2024). https://doi.org/10.48550/arXiv.2406.08177
24. Wu, R., Yang, T., Sun, L., Zhang, Z., Li, S., Zhang, L.: SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution (Jun 2024). https://doi.org/10.48550/arXiv.2311.16518
25. Yin, T., Gharbi, M., Zhang, R., Shechtman, E., Durand, F., Freeman, W.T., Park, T.: One-step diffusion with distribution matching distillation. In: CVPR (2024)
26. Zhang, L., Rao, A., Agrawala, M.: Adding Conditional Control to Text-to-Image Diffusion Models (Nov 2023). https://doi.org/10.48550/arXiv.2302.05543
27. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018)
28. Zhao, H., Yang, B., Cao, L., Li, H.: Data-Driven Enhancement of Blurry Retinal Images via Generative Adversarial Networks. In: Shen, D., Liu, T., Peters,

T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, vol. 11764, pp. 75–83. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_9

29. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. CoRR **abs/1703.10593** (2017), http://arxiv.org/abs/1703.10593
30. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks (2020), https://arxiv.org/abs/1703.10593