Degradation-invariant Enhancement of Fundus Images via Pyramid Constraint Network

Haofeng Liu[1,2], Heng Li[2,✉], Huazhu Fu[3], Ruoxiu Xiao[4], Yunshu Gao[1,2], Yan Hu[2,✉], and Jiang Liu[1,2,5]

[1] Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation, Southern University of Science and Technology, Shenzhen, China
[2] Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China
`lih3, huy3@sustech.edu.cn`
[3] IHPC, A*STAR, Singapore
[4] The School of Computer and Communication Engineering, University of Science and Technology Beijing, China
[5] Singapore Eye Research Institute, Singapore National Eye Centre, Singapore

Abstract. As an economical and efficient fundus imaging modality, retinal fundus images have been widely adopted in clinical fundus examination. Unfortunately, fundus images often suffer from quality degradation caused by imaging interferences, leading to misdiagnosis. Despite impressive enhancement performances that state-of-the-art methods have achieved, challenges remain in clinical scenarios. For boosting the clinical deployment of fundus image enhancement, this paper proposes the pyramid constraint to develop a degradation-invariant enhancement network (PCE-Net), which mitigates the demand for clinical data and stably enhances unknown data. Firstly, high-quality images are randomly degraded to form sequences of low-quality ones sharing the same content (SeqLCs). Then individual low-quality images are decomposed to Laplacian pyramid features (LPF) as the multi-level input for the enhancement. Subsequently, a feature pyramid constraint (FPC) for the sequence is introduced to enforce the PCE-Net to learn a degradation-invariant model. Extensive experiments have been conducted under the evaluation metrics of enhancement and segmentation. The effectiveness of the PCE-Net was demonstrated in comparison with state-of-the-art methods and the ablation study. The source code of this study is publicly available at https://github.com/HeverLaw/PCENet-Image-Enhancement.

Keywords: Fundus image, degradation-invariant enhancement, Laplacian pyramid feature, feature pyramid constraint.

1 Introduction

Due to the convenience, economy, and safety, fundus photography has widely served as a routine clinical examination. Fundus photography is taken by a specialized low-power microscope with an attached CCD (Charge Coupled Device)

camera designed to photograph the interior of the eye, including the retina, optic disc, macula, and posterior pole (i.e., the fundus) [20, 21]. Unfortunately, collecting fundus images under non-ideal light conditions is a frustrating experience. The causation of image quality degradation includes imperfections in the fundus camera optics, aberrations of the human eye, improper camera adjustment, flash lighting, or focusing during the exam. Therefore fundus observation is prone to be impacted by the degradation, resulting in the uncertainty of diagnosis.

Efforts have been long made to enhance the quality of fundus images [13, 15]. Contrast limited adaptive histogram equalization (CLAHE) and Fourier transforms have been collaborated to restore degraded fundus images [17]. Inheriting the guided image filtering (GIF), Cheng et al. [2] designed a structure-preserving guided retinal image filtering (SGRIF) to correct the fundus images from cataract patients. With filtering technique and root domain in frequency, Cao et al. [1] proposed an enhancement method for the retinal images. More recently, with the advance of deep learning techniques, fundus image enhancement has achieved state-of-the-art performances. Segmentation networks were introduced to preserve retinal vessels in enhancement [19]. Based on unpaired image translation [22, 18], importance-guided semi-supervised contrastive constraining (I-SECRET) [3] as well as structure and illumination constrained GAN (StillGAN) [16] were proposed to learn the enhancing mappings from unpaired data. Considering the gap between synthesized and clinical data, domain adaptation [12, 11] was leveraged to generalize the restoration model from synthesized data to real ones by accessing the clinical data.

Despite existing enhancement methods performed decently in the laboratory, challenges remain in the clinical scenarios. i) Due to the complexity of clinical imaging interference, it is impractical for the model to traverse every possible degradation in training. ii) Unpaired data tends to drive enhancement models to abandon some fundus characteristics (such as vessels and abnormality), while synthesized data probably result in performance-drop on real data. iii) As the foundation of fundus assessment, retinal structures need to be more efficiently preserved in the enhancement. iv) Access to clinical data is still necessary for domain adaptation to bridge the gap between synthesized and clinical domains.

To circumvent the above issues, a degradation-invariant fundus enhancement method, termed PCE-Net, is developed in this paper via pyramid constraint. SeqLCs are generated from identical images as the training data with content consistency. Then, the LPF is employed to preserve the retinal structures in the enhancement, and FPC is proposed to constrain the consistency of embeddings for enforcing the learning of a degradation-invariant model. The main contributions of this paper are summarised as follows:

- For boosting the enhancement of clinical fundus images, the PCE-Net is developed to learn the degradation-invariant enhancement model from simulated data.
- The SeqLC and FPC are married to enforce the training of a degradation-invariant model, and the LPF is employed to preserve retinal structures.
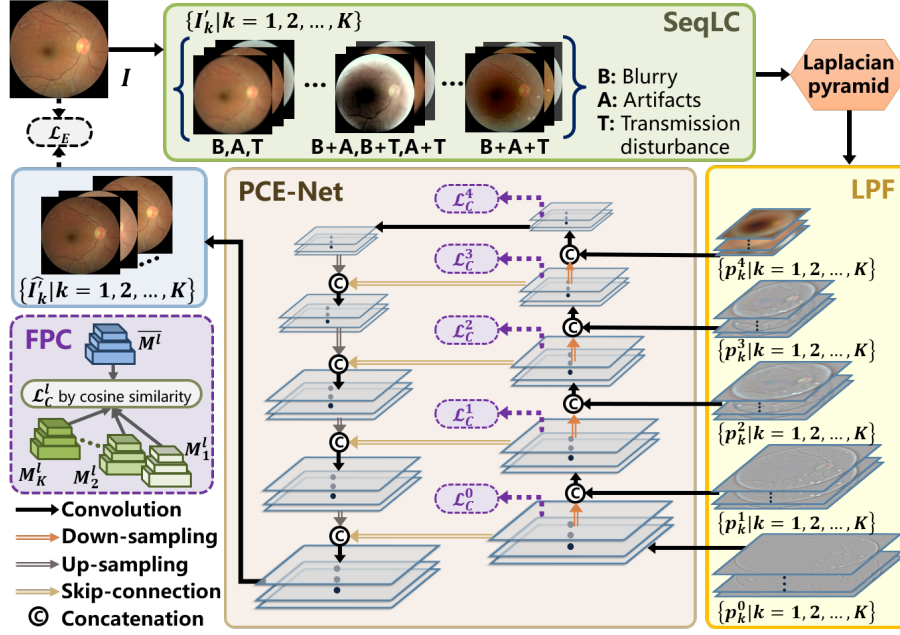
Fig. 1. Overview of PCE-Net. Given a high-quality image I, the SeqLC $\{I'_k|k = 1, 2, ..., K\}$ is randomly synthesized. Then I'_k is decomposed as the LPF $\{p^l_k|l = 0, 1, ..., L\}$ to boost structure-preserving in PCE-Net. Subsequently, the FPC is quantified by the consistency loss $\mathcal{L}^l_C$ at each encoding layer to enforce the model to learn degradation-invariant representations.

– Three evaluations are presented in the experiment to verify the effectiveness of the PCE-Net, and superior performance to state-of-the-art algorithms is achieved in the enhancement of clinical fundus images.

2   Methodology

As exhibited in Fig. 1, SeqLC, LPF, and FPC are implemented in the proposed PCE-Net. Specifically, the SeqLC is randomly degraded from an identical high-quality image for the training of the enhancement network. The LPF is adopted to forward multi-level inputs to boost structure-preserving of the enhancement network. The FPC leads the network to learn the degradation-invariant enhancement model.

2.1   Sequence of Low-quality Images with the Same Content

In clinics, fundus images often suffer from quality degradations caused by imaging interferences [19]. Unfortunately, the enormous cost of collecting high-low quality fundus image pairs and the limited volume of available data prevent

the development of enhancement models. To collect training data and boost the learning of degradation-invariant models, high-quality images are randomly degraded to generate low-quality ones with the same content. SeqLCs are thus acquired to import content consistency into the model training.

As demonstrated in the green block of Fig. 1, the SeqLC $\{I'_k | k = 1, 2, ..., K\}$ is generated from an identical high-quality fundus image I to enforce the learning of the degradation-invariant model using the content consistency. The three imaging interferences modeled in [19], including blurry, artifacts, and transmission disturbance, are randomly combined to degrade I. Accordingly, $\{I'_k | k = 1, 2, ..., K\}$ is formed by the low-quality images with various degradations but the same content.

Furthermore, denoting the enhancement results of $\{I'_k | k = 1, 2, ..., K\}$ as $\{\widehat{I'_k} | k = 1, 2, ..., K\}$, the enhancement loss $\mathcal{L}_E$ under the supervision from I is defined as:

$$\mathcal{L}_E = \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\left\|I - \widehat{I'_k}\right\|_1\right]. \tag{1}$$

2.2 Laplacian Pyramid Features

Laplacian Pyramid decomposes an image into a linear invertible image representation of multi-level frequency features and has been extensively used to boost the structure information in image generation [4], super-resolution translation [14], and semantic segmentation [6]. Considering retinal structures are essential in clinical fundus examination, Laplacian Pyramid is introduced to constrain retinal structures in the enhancement model.

Laplacian Pyramid is implemented with down-sampling and up-sampling operations. For an image I with the size of $s \times s$, the down-sampling operation $\delta(\cdot)$ is carried out by smoothing with a Gaussian filter and then down-sampling to $\frac{s}{2} \times \frac{s}{2}$. Subsequently, the up-sampling operation $\mu(\cdot)$ is performed by resizing the image back to $s \times s$. To acquire an L-level Laplacian pyramid feature (LPF) from an image I, a Gaussian pyramid $\mathcal{G}(I) = \{g^l | l = 0, 1, ..., L\}$ is firstly computed, where $g^0 = I$ and g^n denotes the output of I underwent n times $\delta(\cdot)$. Then the LPF $\mathcal{P}(I) = \{p^l | l = 0, 1, ..., L\}$ is achieved, where $p^l = g^l - \mu(g^{l+1})$ and $p^L = g^L$.

As shown in the yellow block of Fig. 1, an LPF with the L of 4 is extracted for retinal structure preservation in the proposed PCE-Net. LPFs $\{\mathcal{P}(I'_k) | k = 1, 2, ..., K\}$ are calculated from $\{I'_k | k = 1, 2, ..., K\}$, and then the LPFs are forwarded to the network by concatenating with the output of the corresponding layers. Denote f^l as the output from the l-th layer in PCE-Net, the output of the next layer is given by

$$f_k^{l+1} = \text{Conv}([p_k^{l+1}, f_k^l]), l = 0, 1..., L - 1, k = 1, 2, ..., K, \tag{2}$$

where $[\cdot]$ is the concatenation operation, $\text{Conv}(\cdot)$ refers to the convolution operation at each layer, and $f_k^0 = \text{Conv}(p_k^0)$.

2.3   Feature Pyramid Constraint

The underlying content consistency in SeqLC enables the learning of an enhancement model invariant to degradations. Concretely, for a degradation-invariant enhancement model, corresponding consistency should share in the features embedded from $\{I'_k | k = 1, 2, ..., K\}$. However, straightly constraining the numerical consistency of the feature maps in PCE-Net is too inflexible to converge at the optimal solution. Therefore, the FPC is proposed to boost the enhancement model training effectively.

The FPC replaces the numerical consistency constraint to learn the representations invariant to degradations, which leads to a degradation-invariant model. Concretely, spatial pyramid pooling (SPP) [7], a multi-scale pooling operation, is employed to quantify the FPC. Denote SPP as $\sigma(\cdot)$, where the pooling scales of 2×2, 4×4, and 8×8 are conducted. The feature map extracted by SPP is defined as $M_k^l = \sigma(f_k^l)$, where f_k^l denotes the output at the l-th layer in PCE-Net from I'_k. Subsequently, as exhibited by the purple blocks in Fig. 1, the feature maps extracted with SPP are adopted to calculate the consistency loss $\mathcal{L}_C^l$ at each encoding layer of PCE-Net using cosine similarity:

$$\mathcal{L}_C^l = \mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K}\left(1 - \frac{M_k^l \cdot \overline{M^l}}{\|M_k^l\|\|\overline{M^l}\|}\right)\right], \tag{3}$$

where $\overline{M^l} = \frac{1}{K}\sum_{k=1}^{K} M_k^l$. The FPC is hence given by $\mathcal{L}_C = \sum_{l=0}^{L} \mathcal{L}_C^l$.

Accordingly, the total loss function is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_E + \lambda_C \mathcal{L}_C, \tag{4}$$

where λ_C is the weight to balance the total loss and set to 0.1 in training.

3   Experiments

3.1   Implementation Details

To demonstrate the effectiveness of the proposed PCE-Net, a comparison and ablation study were implemented in three evaluations, including enhancement evaluations with and without reference, and vessel segmentation.

Table 1. Datasets used in the experiments

Evaluation	Training set	Test set	Metrics
Full reference enhancement	Good subset in EyeQ: 16,817 high-quality fundus images	FIQ: 196 low-high quality fundus image pairs	SSIM, PSNR
Non-reference enhancement		Usable & Reject subset in EyeQ test dataset: 6,435 mediocre- and 5,540 low-quality fundus images	FIQA, WFQA
Segmentation		FIQ	IoU, DSC

As reported in Table 1, the public dataset of EyeQ [1] and a private fundus dataset FIQ were used in the experiments. Six state-of-the-art enhancement methods, including Cao et al. [1], pix2pix [9], CycleGAN [22], I-SECRET [3], CofeNet [19], Li et al. [12] were compared with PCE-Net. The 'Good' subset in EyeQ was employed as the high-quality images to construct paired data for the model training. High-low quality image pairs are contained FIQ, serving as the test data for the full reference evaluation of enhancement and segmentation. The non-reference evaluation was tested on the 'Usable' and 'Reject' subsets in EyeQ. The enhancement performance was quantified by structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) [8] in the full reference evaluation, while by fundus image quality assessment (FIQA) [3] and weighted FIQA (WFQA) in the non-reference one. FIQA was defined as the ratio of the test images predicted as 'Good' by MCF-Net [5], and WFQA was calculated by assigning the weights of {2, 1, 0} to the images predicted as {'Good', 'Usable', 'Reject'}. The segmentation was measured by the metrics of intersection over union (IoU) and Dice score (DSC) [10]. The segmentation results were obtained from enhanced images and the reference using a U-Net trained on the DRIVE dataset.

In the experiments, a U-Net was designed to construct PCE-Net illustrated in Fig. 1. Note that low-quality images are individually degraded from high-quality ones without the SeqLC. The model trained 200 epochs with Adam optimizer, which the learning rate was 0.001 for the first 150 epochs and decayed gradually to 0 for the following 50 epochs. The batch size was 16, and instance normalization was applied. The input image was resized to 256×256. The same training strategy and augmentation data were applied to all algorithms.

3.2   Comparison and Ablation Study

Comparison of enhancement and segmentation with the state-of-the-art algorithms is visualized in Fig. 2. And quantitative evaluations of the comparison and ablation study are summarized in Table 2.

Comparison As illuminated in Fig. 2 (a), as a result of the quality degradation, it is difficult to observe the fundus, and the segmentation is also severely impacted. Based on image filtering, Cao et al. [1] improved the image clarity in Fig. 2 (b), but the enhancement result was apparently different from the common ones. Due to the grievous interferences, the detailed structures are abandoned in the enhancement by pix2pix [9]. Thanks to training on unpaired images, CycleGAN [22] and I-SECRET [3] are more convenient to deploy in clinical scenarios. However, their preservation of retinal structures is limited by the unpaired data, such that artifacts appear in the enhancement and segmentation result of Fig. 2 (d) and (e). Using vessel annotations as the structure guidance, promises CofeNet [19] remarkable preservation for fundus vessels, but leads to neglect of the unannotated structures. To suppress the gap between the synthesized and

---

[1] https://github.com/HzFu/EyeQ
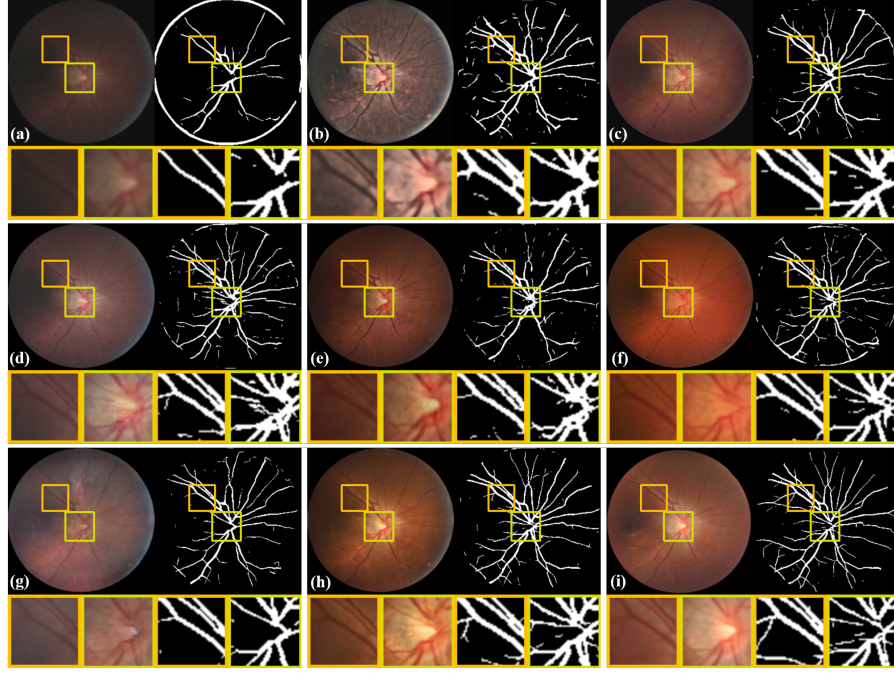
Fig. 2. Visual comparison on the fundus image with full reference. PCE-Net outstandingly enhances the clinical degraded fundus image and outperforms the state-of-the-art methods. (a) Low-quality real image. (b) Cao et al. [1]. (c) pix2pix [9]. (d) CycleGAN [22]. (e) I-SECRET [3]. (f) CofeNet [19]. (g) Li et al. [12]. (h) PCE-Net (ours). (i) Reference.

real data, Li et al. [12] introduces domain adaptation to generalize the enhancement model from synthesized images to real ones. Despite the advance in the performance of Fig. 2 (g), access to real data is required in the model training. Though only learning from synthesized data, superior performance on clinical image enhancement is presented by PCE-Net in Fig. 2 (h). An enhanced image with prominent structure visualization is provided by PCE-Net, and the corresponding segmentation result is highly consistent with that of the high-quality reference in Fig. 2 (i).

Table 2 summarizes the quantitative performances in the comparison. As the image clarity is improved by Cao et al. [1], decent segmentation is achieved. However, the appearance different from common images leads to inferior performance in the enhancement evaluation. Although the image-to-image translation methods of pix2pix [9] and CycleGAN [22] present reasonable performances in the enhancement evaluation with reference, the mediocre results in the non-reference evaluation and segmentation demonstrate their limitations on enhancing retinal structures. Because an importance map and contrastive constraint are introduced in I-SECRET [3] to boost structure preservation and local consistency,

Table 2. Comparisons and ablation study of PCE-Net on evaluations of full-reference, non-reference enhancement, and segmentation. Clear and degraded denote the high-quality images and the low-quality ones before enhancement. N.A. denotes the results were not available. PCE-Net achieves superior performance in the enhancement and segmentation.

	Full reference		Non-reference		Segmentation	
	SSIM	PSNR	FIQA	WFQA	IoU	DSC
Clear	N.A.	N.A.	0.95	1.95	N.A.	N.A.
Degraded	0.766	16.55	0.09	0.67	0.350	0.518
Cao et al. [1]	0.741	20.17	0.27	0.68	0.431	0.602
pix2pix [9]	0.846	21.96	0.18	0.87	0.420	0.592
CycleGAN [22]	0.844	21.36	0.25	0.90	0.420	0.591
I-SECRET [3]	0.841	21.85	0.43	1.18	0.423	0.593
CofeNet [19]	0.857	22.41	0.45	1.26	0.427	0.599
Li et al. [12]	0.839	21.88	0.37	1.15	0.435	0.606
PCE-Net w/o SeqLC, LPF, FPC	0.849	21.46	0.11	0.73	0.400	0.572
PCE-Net w/o LPF, FPC	0.854	22.35	0.24	0.99	0.424	0.596
PCE-Net w/o FPC	0.860	22.55	0.50	1.32	0.447	0.618
PCE-Net (ours)	**0.871**	**23.09**	**0.55**	**1.38**	**0.464**	**0.634**

advantages in the enhancement and segmentation are exhibited. CofeNet [19] and Li et al. [12] respectively employ vessel segmentation and edge detection as structure guidance, achieving reasonable segmentation performance. Nevertheless, as a restoration method of cataract fundus images, the performance of Li et al. [12] has been impacted in the enhancement of degraded images.

Owing to the degradation-invariant representation and multi-level structure preservation, PCE-Net achieves outstanding quantification results in each evaluation compared to the state-of-the-art methods.

Ablation study The ablation study is also provided in Table 2 to verify the effectiveness of the proposed modules. The SeqLC not only prepares sufficient training data for the enhancement model, but also imports content consistency for enforcing the degradation invariance. By decomposing the images into multi-level inputs to the network, the LPF boosts the structure preservation in the enhancement and thus performs remarkably in the enhancement and segmentation. Based on the underlying consistency from the SeqLC, the model in training is constrained by the FPC to learn degradation-invariant representations. Consequently, a prominent enhancement model invariant to various degradations is acquired with the proposed modules.

4   Conclusion

To address the impact of quality degradation on clinical fundus examination, a fundus image enhancement network via pyramid constraint, called PCE-Net, is proposed to construct the degradation-invariant model from generated data.

The SeqLCs are first formed by low-quality images degraded from identical high-quality ones for training the enhancement model. Then the degraded images are decomposed into the LPFs to boost the retinal structure preservation of the enhancement model. Furthermore, FPC is proposed to enforce the learning of degradation-invariant representations based on the underlying content consistency in the SeqLCs. The qualitative and quantitative experiments in enhancement and vessel segmentation demonstrate that PCE-Net presents superior performance compared to state-of-the-art methods. And the ablation study verifies the effectiveness of the proposed modules.

Acknowledgment

References

1. Cao, L., Li, H., Zhang, Y.: Retinal image enhancement using low-pass filtering and α-rooting. Signal Processing **170**, 107445 (2020)
2. Cheng, J., et al.: 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., Lyu, J., Tang, X.: I-secret: Importance-guided fundus image enhancement via semi-supervised contrastive constraining. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 87–96. Springer (2021)
4. Denton, E., Chintala, S., Szlam, A., Fergus, R.: Deep generative image models using a laplacian pyramid of adversarial networks. Advances in Neural Information Processing Systems **2015**, 1486–1494 (2015)
5. 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: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 48–56. Springer (2019)
6. Ghiasi, G., Fowlkes, C.C.: Laplacian pyramid reconstruction and refinement for semantic segmentation. In: European conference on computer vision. pp. 519–534. Springer (2016)
7. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(9), 1904–1916 (2015)
8. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)
9. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134 (2017)

10. Jadon, S.: A survey of loss functions for semantic segmentation. In: 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). pp. 1–7. IEEE (2020)
11. Li, H., Liu, H., Hu, Y., Fu, H., Zhao, Y., Miao, H., Liu, J.: An annotation-free restoration network for cataractous fundus images. IEEE Transactions on Medical Imaging (2022)
12. Li, H., Liu, H., Hu, Y., Higashita, R., Zhao, Y., Qi, H., Liu, J.: Restoration of cataract fundus images via unsupervised domain adaptation. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 516–520. IEEE (2021)
13. Li, T., Bo, W., Hu, C., Kang, H., Liu, H., Wang, K., Fu, H.: Applications of deep learning in fundus images: A review. Medical Image Analysis p. 101971 (2021)
14. Liang, J., Zeng, H., Zhang, L.: High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9392–9400 (2021)
15. Liu, H., Li, H., Ou, M., Zhao, Y., Qi, H., Hu, Y., Liu, J.: Domain generalization in restoration of cataract fundus images via high-frequency components. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2022)
16. Ma, Y., Liu, J., Liu, Y., Fu, H., Hu, Y., Cheng, J., Qi, H., Wu, Y., Zhang, J., Zhao, Y.: Structure and illumination constrained gan for medical image enhancement. IEEE Transactions on Medical Imaging (2021)
17. Mitra, A., Roy, S., Roy, S., Setua, S.K.: Enhancement and restoration of non-uniform illuminated fundus image of retina obtained through thin layer of cataract. Computer methods and programs in biomedicine **156**, 169–178 (2018)
18. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: European Conference on Computer Vision. pp. 319–345. Springer (2020)
19. 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 (2020)
20. Zhang, X.Q., Hu, Y., Xiao, Z.J., Fang, J.S., Higashita, R., Liu, J.: Machine learning for cataract classification/grading on ophthalmic imaging modalities: A survey. Machine Intelligence Research **19**(3), 184–208 (2022)
21. Zhang, X., Xiao, Z., Higashita, R., Hu, Y., Chen, W., Yuan, J., Liu, J.: Adaptive feature squeeze network for nuclear cataract classification in as-oct image. Journal of Biomedical Informatics **128**, 104037 (2022)
22. Zhu, J.Y., et al.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)