Knowledge Distillation-Driven Quality Transfer with Importance-Guided Supervision for Fundus Image Enhancement

Qingshan Hou[1,†], Wenju Yang[2,†], Meng Wang[3], Peiyun Xue[1], Xuhui Huang[1(✉)], and Yangyang Liu[4,5(✉)]

[1] College of Electronic Information Engineering, Taiyuan University of Technology, Taiyuan 030600, Shanxi, China
[2] College of Software, Liaoning Technical University, Huludao, 125105, Liaoning, China
[3] Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[4] School of Computer Engineering, Jiangsu Second Normal University, Nanjing 211200, Jiangsu, China
[5] State Key Lab. of Novel Software Technology, Nanjing University, P.R. China
huangxuhui@tyut.edu.cn; liuyynjus@gmail.com

Abstract. Low-quality fundus images are pervasive in real-world clinical screening, severely compromising the reliability of diabetic retinopathy (DR) diagnosis and degrading the performance of automated grading systems. Existing methods either rely on scarce paired training data or treat enhancement as a uniform image-level operation, neglecting the spatially varying importance of clinically sensitive regions such as lesions and vessels. To address these limitations, we propose a semi-supervised dual framework that integrates unsupervised knowledge distillation-driven quality transfer with importance-guided supervised enhancement. The quality transfer module aligns the global quality distribution of low-quality images to the high-quality domain by exploiting the representational discrepancy between source and target CNN models, without requiring paired training data. The importance-guided module employs a pixel-level importance decoder to adaptively weight the supervision signal, prioritizing clinically critical regions during reconstruction. Extensive experiments on the EyeQ benchmark demonstrate that our method achieves state-of-the-art enhancement performance, outperforming all compared methods. Moreover, the enhanced images yield significant improvements in downstream DR grading, validating the clinical utility of our framework.

Keywords: Fundus Image Enhancement; Knowledge Distillation; Unsupervised Domain Adaptation; Diabetic Retinopathy Grading.

† Qingshan Hou and Wenju Yang contribute equally to this work.
✉ Corresponding authors

1 Introduction

Retinal fundus photography is the primary imaging modality for screening sight-threatening diseases including diabetic retinopathy (DR), glaucoma, and age-related macular degeneration [8,19]. However, as illustrated in Fig. 1, a substantial proportion of fundus images acquired in real-world clinical settings suffer from quality degradation caused by factors such as poor illumination, optical blur, and patient motion, obscuring pathological findings in clinically critical regions—including the optic disc, retinal vessels, and lesion areas—and severely degrading both diagnostic reliability and automated grading performance, motivating the development of robust enhancement methods.

Early deep learning-based fundus enhancement methods rely on paired high- and low-quality training images. However, collecting large-scale paired fundus images in clinical practice remains prohibitively difficult. Representative works such as Cofe-net [17] and I-SECRET [2] construct paired data through degradation simulation and achieve impressive results under controlled settings. To relax the paired data requirement, unpaired methods based on cycle-consistent adversarial training [21,16] have been explored for domain-level quality transfer. As illustrated in Fig. 1(b), while these methods avoid the need for paired supervision, the cycle-consistency constraint operates as a global image-level regularization and is known to introduce structural artifacts and color distortions in the reconstructed images, which can obscure fine retinal structures and compromise diagnostic reliability. More recently, structure-aware enhancement methods [15,11] incorporate image quality metrics or structural priors to guide the enhancement process. Despite their progress, these methods still treat quality enhancement as a spatially uniform operation: all pixels are supervised equally regardless of their clinical significance, neglecting the fact that regions such as the optic disc, retinal vessels, and DR lesions carry disproportionately higher diagnostic importance than background areas. To overcome these limitations, we

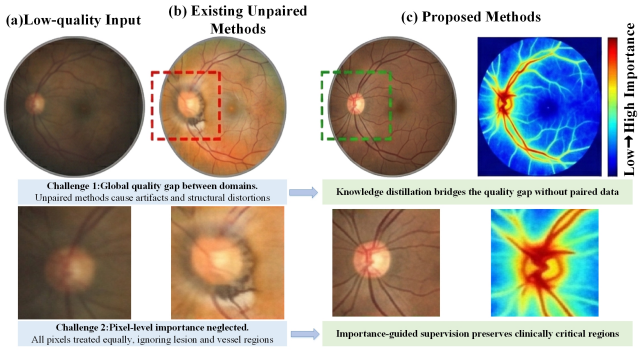

Fig. 1: (a) Real-world low-quality input. (b) Existing unpaired methods introduce artifacts and structural distortions. (c) Our method bridges the domain gap and prioritizes clinically critical regions.

propose a semi-supervised dual framework for fundus image enhancement that addresses both the global quality gap and the spatially heterogeneous importance of retinal structures. Inspired by unsupervised domain adaptation, our quality transfer module exploits the representational discrepancy between a source CNN pre-trained on high-quality images and a target CNN fine-tuned on low-quality images. By aligning the feature-space outputs of the generator with those of the source CNN, the framework bridges the quality domain gap without requiring any paired training data, while substantially reducing the structural artifacts characteristic of cycle-consistency-based approaches. Complementing this, an importance-guided supervised learning branch introduces a lightweight pixel-level importance decoder that dynamically assigns higher supervision weights to clinically critical regions during training on synthetically degraded image pairs. As visualized in Fig. 1(c), this spatially adaptive supervision mechanism enables the framework to recover fine structural details in high-importance regions that uniform supervision strategies consistently fail to preserve. Our main contributions are as follows:

1) A knowledge distillation-driven unsupervised quality transfer mechanism is proposed, which aligns the intermediate feature representations of the generator with those of a high-quality domain CNN. This design bridges the global quality gap without requiring paired training data, while mitigating the structural artifacts commonly introduced by conventional cycle-consistency-based methods.

2) To enable preservation of clinically critical structures — including the optic disc, retinal vessels, and DR lesions — a pixel-level importance-guided supervised learning branch is designed. A lightweight importance decoder is incorporated to adaptively modulate supervision intensity across spatial locations, ensuring that regions of high clinical relevance receive proportionally stronger guidance.

3) By effectively exploiting large-scale unpaired real clinical images alongside synthetic paired data, the proposed unified semi-supervised framework achieves state-of-the-art enhancement performance and demonstrates consistent improvements in downstream DR grading across standard benchmarks.

2 Methodology

Given unpaired low-quality images $\mathcal{X} : \{x_i\}_{i=1}^{N_x}$ and high-quality images $\mathcal{Y} : \{y_i\}_{i=1}^{N_y}$, we synthesize a degraded dataset $\hat{\mathcal{Y}} : \{\hat{y}_i\}_{i=1}^{N_y}$ following the degradation pipeline of Shen et al. [17] to construct supervised training pairs. The goal is to train a generator $g(\cdot)$ that maps $x \in \mathcal{X}$ to an enhanced image $\tilde{x}$ approximating the high-quality distribution. As illustrated in Fig. 2, the proposed framework adopts a semi-supervised design for two complementary reasons. 1) Unsupervised methods lack explicit structural supervision and tend to produce artifacts when enhancing fine retinal details. 2) Supervised methods require large-scale paired real clinical data, which is prohibitively expensive to collect. By jointly optimizing an unsupervised quality transfer branch over unpaired real images $\mathcal{X}$ and a supervised enhancement branch over synthetic pairs $(\hat{\mathcal{Y}}, \mathcal{Y})$, the framework leverages the complementary strengths of both paradigms: the unsupervised branch

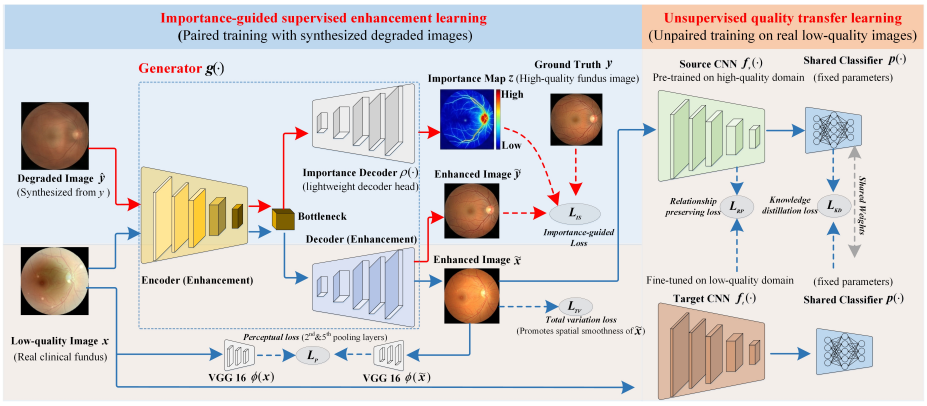

Fig. 2: The overall architecture of the proposed framework.

aligns global quality distributions across domains, while the supervised branch provides pixel-level structural fidelity guidance.

2.1 Unsupervised Quality Transfer

Inspired by prior works on unsupervised domain adaptation (UDA) [1,3], we first pre-train a source CNN model $f_s(\cdot)$ and a shared classifier $p(\cdot)$ on the high-quality source domain. With the shared classifier parameters fixed, we fine-tune $f_s(\cdot)$ on the low-quality target domain to obtain the target CNN model $f_t(\cdot)$. The weight parameters of $f_s(\cdot)$, $f_t(\cdot)$, and $p(\cdot)$ are then frozen. $f_s(\cdot)$ is pre-trained on high-quality images and encodes high-quality domain knowledge in its feature representations, while $f_t(\cdot)$ is fine-tuned on low-quality images and captures target-domain characteristics. If the generator $g(\cdot)$ can produce an enhanced image $\tilde{x}$ such that $f_s(\tilde{x})$ yields similar classifier outputs to $f_t(x)$—i.e. the enhanced image is recognised by the high-quality-domain model in the same way the low-quality image is recognised by the target-domain model—then the generator has successfully transferred quality-related knowledge from the source domain into the enhanced image, without requiring any paired supervision. To align the classifier's probability distributions $p(\cdot)$ over disease categories, the knowledge distillation loss is formulated as:

$$\mathcal{L}_{KD} = D_{\mathrm{KL}}\big(p(f_t(x)) \parallel p(f_s(\tilde{x}))\big), \tag{1}$$

where $p(f_t(x))$ is the classifier output for the low-quality image under the target model—representing what the target domain "perceives"—and $p(f_s(\tilde{x}))$ is the classifier output for the enhanced image under the source model—representing whether the enhanced image has acquired high-quality domain characteristics. Minimising $\mathcal{L}_{KD}$ drives the generator to produce images that are indistinguishable from the high-quality domain in terms of semantic feature distributions.

To preserve global quality features, motivated by style transfer [9], we employ a Gram-matrix-based channel-level relationship preserving loss following

Hou et al. [9], where channel-wise Gram matrices capture inter-channel co-activation patterns encoding global style and texture. We reshape feature maps as $f_s(\tilde{x}), f_t(x) \in \mathbb{R}^{D\times H\times W} \longrightarrow F_s, F_t \in \mathbb{R}^{D\times HW}$ and compute $G_s = F_s F_s^\top$, $G_t = F_t F_t^\top \in \mathbb{R}^{D\times D}$. Row-wise ℓ_2 normalisation is then applied to ensure equal channel contributions:

$$\bar{G}_s[i,:] = \frac{G_s[i,:]}{\|G_s[i,:]\|_2}, \qquad \bar{G}_t[i,:] = \frac{G_t[i,:]}{\|G_t[i,:]\|_2}, \tag{2}$$

where $[i,:]$ denotes the i-th row. The relationship preserving loss $\mathcal{L}_{RP} = \frac{1}{D}\|\bar{G}_s - \bar{G}_t\|_F^2$ is defined as the MSE between the normalised Gram matrices.

To make the enhanced fundus image visually realistic and to preserve lesion shape and structure from the original low-quality image, we adopt a perceptual loss following Cycle-Dehaze [5], a perceptual loss is constructed by fusing low-level and high-level features extracted from the 2nd and 5th pooling layers of VGG16 [18], formulated as $\mathcal{L}_P = \|\phi(x) - \phi(g(x))\|_2^2$, where $\phi(\cdot)$ denotes feature mappings corresponding to the selected pooling layers. To further suppress high-frequency noise along vessel boundaries while preserving the continuity of fundus structures, we incorporate a total variation (TV) loss:

$$\mathcal{L}_{TV} = \sum_{i=1}^{H}\sum_{j=1}^{W}\Big[(\tilde{x}_{i,j} - \tilde{x}_{i+1,j})^2 + (\tilde{x}_{i,j} - \tilde{x}_{i,j+1})^2\Big]^{\beta/2}, \tag{3}$$

where $\tilde{x}$ denotes the enhanced image, i and j are pixel coordinates, and β controls the degree of smoothness penalisation. Unlike isotropic smoothing, which uniformly blurs all regions, the TV loss selectively penalises abrupt local discontinuities, allowing the generator to maintain sharp vessel boundaries while suppressing noise in homogeneous background areas.

2.2 Importance-Guided Supervised Learning

To complement the unsupervised quality transfer branch and recover fine structural details beyond global domain alignment, we introduce an importance-guided supervised learning branch. This branch operates on synthetic paired data $(\hat{y}, y)$ and introduces a lightweight importance decoder $\rho(\cdot)$ that re-interprets pixel-level reconstruction difficulty as spatially adaptive supervision weights.

The importance decoder $\rho(\cdot)$ shares the encoder with the generator $g(\cdot)$ and consists of three convolutional layers with ReLU activations, followed by a sigmoid output layer that constrains the importance map to $z = \rho(\hat{y}) \in [0,1]^{H\times W}$. Higher values in z indicate regions where reconstruction is more uncertain and therefore require stronger regularisation of the supervision signal. A key design choice in our loss function is the use of $1/\exp(\alpha_i)$ as the per-pixel supervision weight, where α_i denotes the learned log-uncertainty of pixel i, following the heteroscedastic uncertainty formulation of Kendall and Gal [10]. This formulation should be interpreted as uncertainty-weighted supervision rather than importance-weighted supervision: when the decoder assigns a high α_i to a pixel,

it signals high reconstruction uncertainty at that location, and the loss function accordingly down-weights the MSE contribution of that pixel to prevent gradients from being dominated by inherently ambiguous regions. The regularisation term $\frac{1}{N}\sum_i \alpha_i$ penalises the decoder from trivially assigning infinite uncertainty to all pixels. In practice, the decoder learns to assign high uncertainty to background regions with little diagnostic value, and low uncertainty to clinically critical regions such as vessel boundaries and lesion areas, effectively concentrating supervision on diagnostically relevant structures:

$$\mathcal{L}_{IS} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\exp(\alpha_i)}\,\mathcal{L}_{\text{MSE}}(\tilde{y},\, y) + \frac{1}{N}\sum_{i=1}^{N}\alpha_i, \tag{4}$$

where N is the total number of pixels in the batch, $\mathcal{L}_{\text{MSE}}(\cdot,\cdot)$ is the batch MSE, and α_i is the learned log-uncertainty of the i-th pixel. The overall objective is $\mathcal{L} = \lambda_1\mathcal{L}_{KD} + \lambda_2\mathcal{L}_{RP} + \lambda_3\mathcal{L}_{P} + \lambda_4\mathcal{L}_{IS} + \lambda_5\mathcal{L}_{TV}$, where $\lambda_1, \ldots, \lambda_5$ are loss-balancing hyper-parameters.

3 Experiments

3.1 Datasets and Implementation Details

Datasets. The DDR dataset [13] serves as the high-quality source domain for pre-training $f_s(\cdot)$ and $p(\cdot)$, and EyePACS [6] as the low-quality target domain for fine-tuning $f_t(\cdot)$. For enhancement evaluation, we follow Shen et al. [17] on EyeQ [4]: "Good" images provide high-quality ground truth for supervised training pairs and PSNR/SSIM reference [7]; "Reject" images serve as real low-quality inputs.

Implementation Details. The proposed method is implemented in PyTorch and trained on 4 NVIDIA Quadro RTX 6000 GPUs. $f_s(\cdot)$ and $f_t(\cdot)$ are initialised from SHOT-IM [14] with a pre-trained ResNet-50 backbone. The generator $g(\cdot)$ follows the StillGAN architecture [16]. All images are resized to 512×512. The generator and importance decoder are trained end-to-end using the Adam optimiser with cosine annealing (initial $lr = 1{\times}10^{-4}$, $momentum = 0.9$, $batchsize = 8$) for 100 epochs, with early stopping (patience = 15) on validation PSNR. Loss weights are set to $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $\lambda_3 = 0.1$, $\lambda_4 = 1.0$, $\lambda_5 = 0.01$, determined via a one-at-a-time sensitivity analysis on the EyeQ validation split.

3.2 Comparison with State-of-the-Art Methods

We compare our method against seven state-of-the-art approaches: Cofe-net [17], I-SECRET [2], StillGAN [16], Diff-Retinex++ [20], PCE-Net [15], SCR-Net [11], and ARC-Net [12]. Table 1 reports PSNR and SSIM on the EyeQ test set, together with ΔPSNR and ΔSSIM gains relative to the unenhanced low-quality baseline. Our method achieves the highest PSNR of 27.86 dB and SSIM of 0.91, outperforming the closest competitor I-SECRET by +0.50 dB. Compared with

Diff-Retinex++, our method achieves a +1.83 dB gain. StillGAN shows a larger gap (+2.48 dB), confirming that global cycle-consistency constraints are insufficient for preserving fine retinal structures. As shown in Fig 3, zoomed regions

Table 1: Quantitative comparison on the EyeQ test set.

Method	PSNR↑	SSIM↑	ΔPSNR↑	ΔSSIM↑
Low-quality (baseline)	18.23	0.75	—	—
Cofe-net [17]	20.51	0.88	+2.28	+0.13
I-SECRET [2]	27.36	0.90	+9.13	+0.15
StillGAN [16]	25.38	0.89	+7.15	+0.14
Diff-Retinex++ [20]	26.03	0.84	+7.80	+0.09
PCE-Net [15]	22.44	0.87	+4.21	+0.12
SCR-Net [11]	26.37	0.84	+8.14	+0.09
ARC-Net [12]	21.24	0.80	+3.01	+0.05
Ours	**27.86**	**0.91**	**+9.63**	**+0.16**

†Baseline PSNR/SSIM computed on "Reject"-quality EyeQ images against "Good"-quality references.

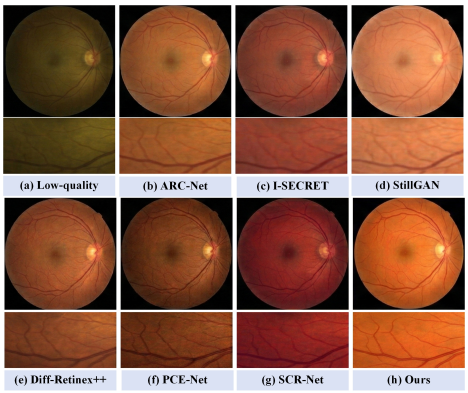

Fig. 3: Comparison of the image enhancement on the EyeQ dataset.

highlight fine capillaries for each method. Regarding global quality, our method achieves the most natural illumination and colour correction, closely resembling "Good"-quality EyeQ references, while StillGAN produces noticeable colour cast artefacts. Regarding structural fidelity, I-SECRET produces blurred capillary boundaries, ARC-Net fails to recover small vessel details, and SCR-Net introduces ringing artefacts along vessel edges. In contrast, our method preserves sharp vessel boundaries and recovers the fine capillary network, which we attribute to the spatially adaptive supervision provided by the importance-guided branch: the importance decoder assigns high supervision weight precisely to these vessel regions.

3.3 Ablation Study

Figure 4 presents qualitative and quantitative ablation results on the EyeQ test set. For each ablation variant, we crop a consistent region enclosing the optic disc and a major vessel branch for structural comparison. Variant (b), removing both $\mathcal{L}_{KD}$ and $\mathcal{L}_{IS}$, exhibits pronounced domain shift with an unnatural warm colour cast, confirming knowledge distillation as the primary mechanism for global quality alignment. Variant (c), $\mathcal{L}_{RP}$ and $\mathcal{L}_{IS}$, produces less structured vessel appearance, suggesting channel-level relationship preservation contributes to natural colour rendition. Variant (d), removing $\mathcal{L}_{P}$ and $\mathcal{L}_{IS}$, reveals broken vessel edges and loss of high-frequency textural detail, confirming perceptual loss is indispensable for structural fidelity. Variant (e), retaining all unsupervised

losses but omitting $\mathcal{L}_{IS}$, achieves acceptable global quality yet still exhibits subtle capillary blurring. The full model (f) recovers the sharpest vessel boundaries, clearest optic disc rim, and most natural colour tone, with the +2.62 dB PSNR gain over variant (e) directly confirming that the importance-guided branch targets fine structural details the unsupervised branch alone cannot recover. To

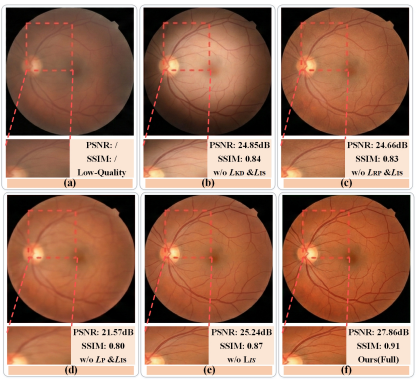

Fig. 4: Qualitative and quantitative ablation study on the EyeQ dataset.

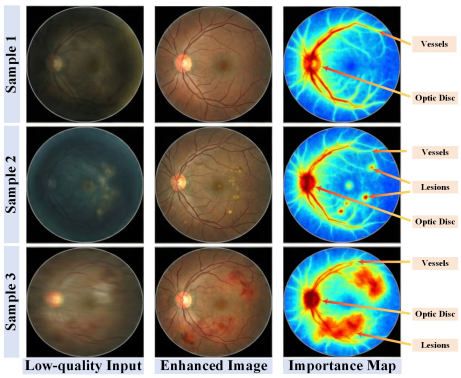

Fig. 5: Visualization of importance maps for three representative samples.

further validate that the importance decoder $\rho(\cdot)$ learns clinically meaningful spatial priors rather than arbitrary weighting patterns, Fig. 5 visualises importance maps for three representative samples across different DR severity levels. Across all samples, the highest weights consistently concentrate on two structure categories. The optic disc receives the highest importance in all three samples, as expected given its critical role in downstream retinal analysis. Major vessel branches also receive consistently high weights, reflecting that the decoder treats elongated vascular structures as high-priority reconstruction targets. Most notably, in Samples 2 and 3 the decoder additionally identifies lesion regions—corresponding to pathological findings such as microaneurysms and hard exudates—as high-importance areas without any explicit lesion annotation. This emergent localisation arises because pathological regions are inherently harder to reconstruct from degraded inputs, and the uncertainty-weighted loss naturally emphasises them—a form of weakly supervised lesion attention. In contrast, background regions consistently receive the lowest weights, confirming that supervision capacity is concentrated on diagnostically relevant structures, directly explaining the +2.62 dB gain from adding $\mathcal{L}_{IS}$ in the ablation study.

4 Conclusion

We presented a semi-supervised dual framework for image enhancement that addresses two key limitations: reliance on paired training data and neglect of spa-

tially varying pixel importance. The unsupervised quality transfer branch aligns global quality distributions without paired data, while the importance-guided supervised branch concentrates supervision on clinically critical regions without explicit anatomical annotation. Experiments on EyeQ demonstrate state-of-the-art enhancement performance and the largest downstream DR grading improvement among all compared methods, validating that structural fidelity gains translate into clinically meaningful improvements. Ablation studies confirm the importance-guided branch contributes the single largest performance gain, and importance map visualisations reveal that the decoder identifies lesion regions as high-priority targets without any lesion-level supervision—an emergent weakly supervised behaviour requiring no additional annotation.

Acknowledgments. This research was supported by the Open Project of the State Key Lab. of Novel Software Technology at Nanjing University(No. KFKT2026B63).

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

References

1. Chen, B., Zhang, X., Shen, C., Li, Q., Song, Z.: Couda: Continual unsupervised domain adaptation for industrial fault diagnosis under dynamic working conditions. IEEE Transactions on Industrial Informatics (2025)
2. 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)
3. Cui, C., Liu, Z., Gong, S., Zhu, L., Zhang, C., Liu, H.: When adversarial training meets prompt tuning: Adversarial dual prompt tuning for unsupervised domain adaptation. IEEE Transactions on Image Processing (2025)
4. 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)
5. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2414–2423 (2016)
6. Graham, B.: Kaggle diabetic retinopathy detection competition report. University of Warwick **22**(9) (2015)
7. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)
8. Hou, Q., Zhou, Y., Goh, J.H.L., Zou, K., Yew, S.M.E., Srinivasan, S., Wang, M., Lo, T.W.S., Lei, X., Wagner, S.K., et al.: Can a natural image-based foundation model outperform a retina-specific model in detecting ocular and systemic diseases? Ophthalmology Science **6**(1), 100923 (2026)
9. Hou, Y., Zheng, L.: Visualizing adapted knowledge in domain transfer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13824–13833 (2021)

10. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? Advances in neural information processing systems **30** (2017)
11. 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. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 487–496. Springer (2022)
12. 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 **41**(7), 1699–1710 (2022)
13. 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)
14. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: International conference on machine learning. pp. 6028–6039. PMLR (2020)
15. 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)
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 **40**(12), 3955–3967 (2021)
17. 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)
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
19. Wang, M., Lin, T., Lin, A., Yu, K., Peng, Y., Wang, L., Chen, C., Zou, K., Liang, H., Chen, M., et al.: Enhancing diagnostic accuracy in rare and common fundus diseases with a knowledge-rich vision-language model. Nature communications **16**(1), 5528 (2025)
20. Yi, X., Xu, H., Zhang, H., Tang, L., Ma, J.: Diff-retinex++: Retinex-driven reinforced diffusion model for low-light image enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence (2025)
21. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)