

Self-supervised Domain Adaptation for Breaking the Limits of Low-quality Fundus Image Quality Enhancement

Qingshan Hou^{1,2}, Peng Cao^{1,2*}, Jiaqi Wang^{1,2}, Xiaoli Liu³, Jinzhu Yang^{1,2} and Osmar R. Zaiane⁴

¹Computer Science and Engineering, Northeastern University, Shenyang, China

²Key Laboratory of Intelligent Computing in Medical Image of Ministry of Education, Northeastern University, Shenyang, China

³DAMO Academy, Alibaba Group, China

⁴Alberta Machine Intelligence Institute, University of Alberta, Edmonton, Alberta, Canada

houqingshancv@gmail.com, caopeng@mail.neu.edu.cn, wjq010222@gmail.com,
liuxiaoli.lxl@alibaba-inc.com yangjinzhu@cse.neu.edu.cn, zaiane@cs.ualberta.ca

Abstract

Retinal fundus images have been applied for the diagnosis and screening of eye diseases, such as Diabetic Retinopathy (DR) or Diabetic Macular Edema (DME). However, both low-quality fundus images and style inconsistency potentially increase uncertainty in the diagnosis of fundus disease and even lead to misdiagnosis by ophthalmologists. Most of the existing image enhancement methods mainly focus on improving the image quality by leveraging the guidance of high-quality images, which is difficult to be collected in medical applications. In this paper, we tackle image quality enhancement in a fully unsupervised setting, i.e., neither paired images nor high-quality images. To this end, we explore the potential of the self-supervised task for improving the quality of fundus images without the requirement of high-quality reference images. Specifically, we construct multiple patch-wise domains via an auxiliary pre-trained quality assessment network and a style clustering. To achieve robust low-quality image enhancement and address style inconsistency, we formulate two self-supervised domain adaptation tasks to disentangle the features of image content, low-quality factor and style information by exploring intrinsic supervision signals within the low-quality images. Extensive experiments are conducted on EyeQ and Messidor datasets, and results show that our DASQE method achieves new state-of-the-art performance when only low-quality images are available.

1 Introduction

Medical fundus images have been extensively used for clinical analysis of various ocular diseases [Teo *et al.*, 2021;

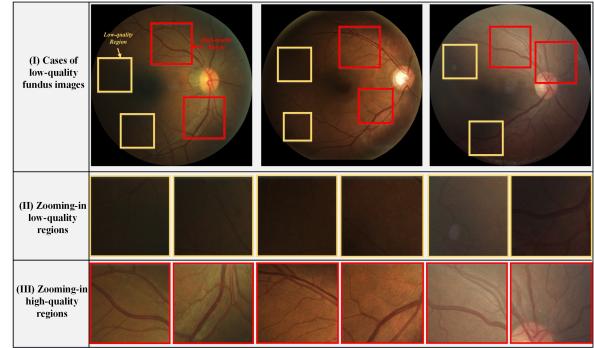


Figure 1: Some cases of low-quality fundus images, where the red box and the yellow box mark the different quality regions in the low-quality images, respectively. The main low-quality factors include uneven illumination, noticeable blur and artifacts.

Dai *et al.*, 2021; Li *et al.*, 2019]. However, the real clinical fundus datasets usually contain a large number of low-quality images. The quality of fundus images is critical to the diagnosis and screening of eye diseases. In contrast, the low-quality fundus images easily mislead the clinical diagnosis and lead to unsatisfactory results of downstream tasks like blood vessel segmentation. Existing deep learning methods [Shen *et al.*, 2020; Cheng *et al.*, 2021; Gui *et al.*, 2021] rely on a large amount of high-quality fundus images or paired images, which limits their practicality and generalization in clinical applications due to lack of availability of high-quality fundus images. To this end, this paper explores a new perspective: could a model achieve quality enhancement without requiring high-quality images?

In low-quality fundus images, we observe some interesting phenomena shown in Fig.1: not all the regions are low quality, as shown in Fig.1(I). Besides, there are significant differences among regions in low-quality images with respect to the image styles, as shown in Fig.1(III). Taking all of the above into consideration, the major challenges of the fundus image quality enhancement lie in: How to disentangle the low-quality factors and image content information for simul-

*Corresponding author

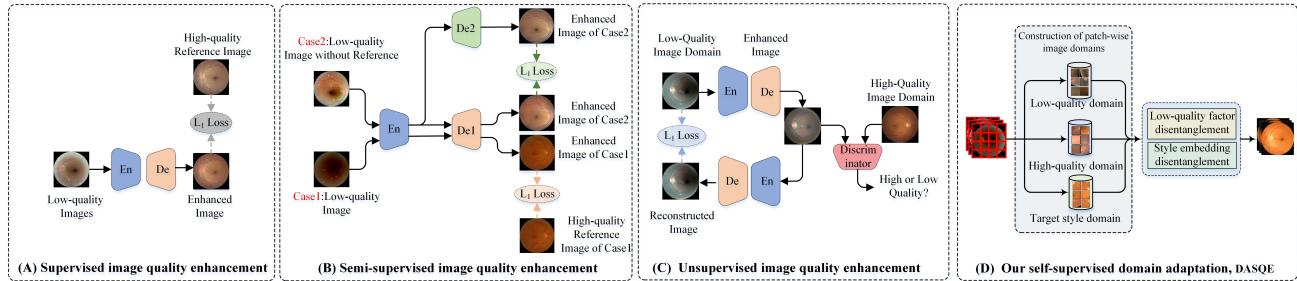


Figure 2: Comparison of different image quality enhancement schemes. Previous fundus image quality enhancement methods (A) require paired images of high-quality and low-quality as training data, which are often difficult to acquire. Although these semi-supervised and unsupervised methods (B) and (C) eliminate the need for paired images, they still need the guidance of high-quality images. (D) In contrast, our method is a truly unsupervised image quality enhancement, requiring neither paired fundus images nor high-quality images.

taneously 1) improving the quality and 2) unifying the style of images under the condition that only low-quality images are available. Specifically, inter-domain variations include two aspects: the variation between high-quality domain and low-quality domain, and the variation between source high-quality style domain and target high-quality style domain.

To the best of our knowledge, this is the first work to rethink the image quality enhancement problem from a self-supervised learning perspective, without the requirement of any high-quality image. Based on the design of the self-supervision, we propose two assumptions to formulate the self-supervised image quality enhancement: 1) the patches from the same position of the source and translated images should have consistent content. Hence, the low-quality image patches are assumed to be composed of low-quality factors and image content embedding. The procedure of image quality enhancement can be formulated as disentangling the quality factors from the image content. 2) The low-quality image patches consist of style embedding and image content embedding. To unify the styles, we aim to factorize style-related and content-related embeddings without any auxiliary supervisory signals. Specifically, we propose a high-quality unaware self-supervised domain adaptation framework, named DASQE. The proposed framework first detects the high-quality regions inside the low-quality images by a pre-trained quality assessment network. Given the patch-wise high-quality domain consisting of high-quality patch-wise images, we obtain multiple style domains by clustering and choose a target style domain consisting of high-quality patch-wise images with uniform illumination style. Then, we disentangle patch-wise images into content-related, quality-related and target style-related embedding by separate encoders, and further reconstruct them into original images by different image generators to reduce quality and style variations. The overall procedure involves the cycle consistency loss and the adversarial losses in the latent and image spaces to achieve a self-supervised quality enhancement without any explicit supervision of high-quality data.

In summary, our contributions can be summarized as follows. 1) A major limitation of most current low-quality fundus image enhancement methods is that they rely on the guidance of paired images or the presence of high-quality images. Hence, we propose a novel medical image quality enhance-

ment method to loosen the requirement of pair-wise training images and only require low-quality input images. To the best of our knowledge, this is the first attempt to apply a self-supervised reference-free method coupled with domain adaptation to the quality enhancement task. We tackle image quality enhancement in a fully unsupervised setting, i.e., neither paired images nor high-quality images are needed. It is a truly unsupervised fundus image quality enhancement. 2) To enhance low-quality fundus images while preserving pathological features and major retinal structures, we propose a representation decoupling strategy for disentangling content-related, style-related and quality-related embeddings. The low-quality fundus images are enhanced by recombining the learned content-related and target style-related embeddings. It is worth mentioning that our method not only improves the quality of fundus images by the proposed strategy, but also unifies the image style to eliminate style variations. 3) Our self-supervised version without the guidance of any high-quality data significantly outperforms the state-of-the-art methods by a considerable 28.47 PSNR / 91.4% SSIM and 29.36 PSNR / 89.7% SSIM on Messidor and EyeQ benchmark datasets, a 1.11/2.63 PSNR and 0.6/1.4% SSIM improvement compared to the previous best results. Moreover, the proposed method is easily expanded to the quality enhancement task of other medical images. We also further verify that our method is beneficial for a variety of fundus imaging analysis tasks, such as retinal vessel segmentation, lesion segmentation and disease grading.

2 Related Work

The quality enhancement methods for fundus images mainly contain traditional non-parametric methods based on hand-crafted priors and data-driven methods based on deep learning. For example, [Shome and Vadali, 2011] enhance the contrast of fundus images by the contrast limited adaptive histogram equalization method. [Tian and Cohen, 2017] design a global and local contrast enhancement method for the quality enhancement of non-uniform illumination images. [Zhou *et al.*, 2017] adjust the luminosity of fundus images based on the luminance gain matrix to achieve enhancement of fundus images with uneven illumination. However, these methods rely heavily on hand-crafted priors, which are hardly applicable to all cases of low-quality fundus image enhancement,

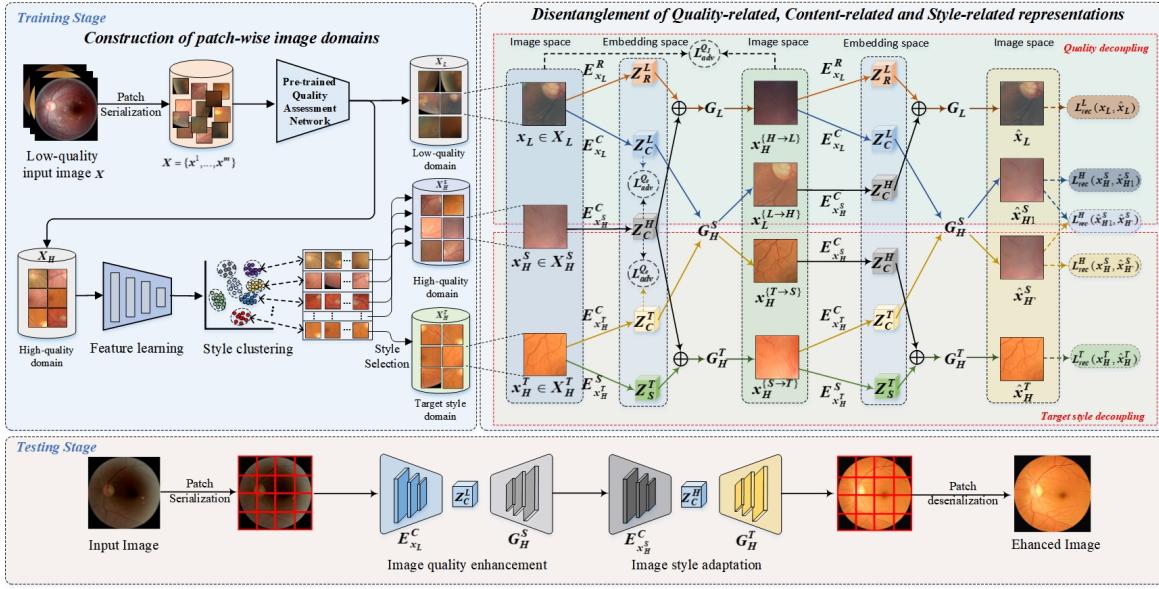


Figure 3: The overall architecture of the proposed framework. It involves two stages: **Construction of patch-wise image domains**: The low-quality input images X are first serialized as the patch-wise fundus images $\mathbf{X} = \{x^1, \dots, x^m\}$. With the help of the pre-trained quality assessment network, the patch-wise fundus images are divided into low-quality and high-quality domains, e.g. $\mathbf{X}_L, \mathbf{X}_H$. A style clustering is performed on \mathbf{X}_H , and a target style domain \mathbf{X}_H^T is determined. **Decoupling of multiple features**: We factorize content-related, quality-related and style-related for patch-wise domains through the supervisions derived from the image data itself to enhance the image quality and align style features. The framework involves content-related encoders $\{E_{x_L}^C, E_{x_H^S}^C, E_{x_H^T}^C\}$ for extracting latent content embedding, quality-related encoder $E_{x_L}^R$ for extracting low-quality factor embedding, target style-related encoder $E_{x_H^T}^S$ for extracting target style embedding, low-/high-quality image generators G_L and G_H^S for generating different quality images, and high-quality target style image generator G_H^T for generating target style images.

such as artifacts. In addition, they also do not provide any learnable parameters to prevent the over-enhancement of fundus images. Recently, deep learning has shown its advantages in a wide range of fields, such as vessel segmentation [Jin *et al.*, 2019], lesion detection [Wang *et al.*, 2018], disease classification [Cao *et al.*, 2022]. For the quality enhancement of fundus images, the most common quality enhancement methods [Wang *et al.*, 2021; Shen *et al.*, 2020] based on deep learning belong to the fully supervised learning scheme (Fig.2(A)), which requires supervision with high-quality reference images corresponding to the input during the training stage. However, obtaining paired fundus images is expensive and time-consuming, resulting in limited applications due to the requirement for pairwise training data. In addition, the supervised enhancement methods limit their generalization and practicality across different datasets owing to inter-domain variations between datasets. To loosen the limitations of paired fundus images, a semi-supervised quality enhancement method (Fig.2(B)) [Cheng *et al.*, 2021] is proposed, which requires only a portion of paired images during the training phase. Inspired by generative adversarial learning, some unsupervised methods (Fig.2(C)) [You *et al.*, 2019; Ma *et al.*, 2021] based on bi-directional GAN have been proposed to enhance low-quality fundus images, which eliminates the requirement for paired fundus images. However, such methods usually require learning knowledge representation in the high-quality domain, and transferring it to the low-

quality domain. Its heavy reliance on high-quality data proves to be the pain point of image quality enhancement methods.

Furthermore, most existing unsupervised methods are usually under-constrained, which may introduce undesirable artifacts or fail to preserve the fine retinal structures and pathological signatures in the real clinical fundus images, and the high-quality image domains are similarly uncommon. From the above analysis and comparison of related work, further exploration of unsupervised quality enhancement method without any guidance of high-quality images is significant and needed for adapting the extensive clinical applications.

3 Methodology

Our main goal is to develop a clinically oriented fundus image quality enhancement method, and uniform illumination style of fundus images. In this section, we first describe the scheme for the construction of the patch-wise domains. Subsequently, we introduce the specific details of multiple feature decoupling, mainly including quality feature decoupling and target style feature decoupling. The quality feature decoupling is designed to remove low-quality interference indicators from patch-wise low-quality images, whereas the style decoupling aims to eliminate style variations among patch-wise images. Finally, we provide the optimization process of the proposed method, the overview of which is illustrated in Fig.3.

3.1 Construction of patch-wise image domains

The goal of this part is to obtain the patch-wise low-/high-quality domain and select a high-quality target style domain which is beneficial to the ophthalmologist's diagnosis. Formally, the low-quality fundus images \mathbf{X} are serialized into a patch-wise image set $\mathbf{X} = \{x^1, \dots, x^m\}$ as shown in Fig.3. Then, \mathbf{X} are fed into a pre-trained quality assessment network to obtain the patch-wise low-quality image domain $\mathbf{X}_L = \{x_L^1, \dots, x_L^n\}$ and high-quality image domain $\mathbf{X}_H = \{x_H^1, \dots, x_H^{n'}\}$, where $m = n + n'$. For quality assessment network, a base DesNet121 model is trained based on the EyeQ dataset [Fu *et al.*, 2019] with high-/low-quality binary labels. For the patch-wise high-quality domains \mathbf{X}_H , there are several distinct illumination styles. This poses significant challenges to the enhancement of low-quality images. In order to align the style of patch-wise images, it is essential to construct a target style domain for the subsequent style decoupling. More specifically, our primary goal is to learn an embedding function $\text{De}(\cdot)$ mapping \mathbf{X}_H to embeddings $\mathbf{Z} = \{z_1, z_2, \dots, z_j\}$ in a D -dimension representation space. The loss \mathcal{L}_e of latent embedding learning is defined as:

$$\mathcal{L}_e = \|\text{De}(\text{En}(x_H^i)) - x_H^i\|_2^2, \quad (1)$$

where x_H^i indicates the patch-wise images from \mathbf{X}_H , $\text{De}(\cdot)$ and $\text{En}(\cdot)$ denote the encoder and decoder, respectively. Subsequently, we cluster the obtained embeddings $\mathbf{Z} = \text{En}(\mathbf{X}_H)$ to obtain the style domains $\{\mathbf{X}_{S1}, \mathbf{X}_{S2}, \dots, \mathbf{X}_{Sn}\}$. Finally, we select a style domain containing the highest number of patches as the target style domain \mathbf{X}_H^T , and the others are considered as source style domains \mathbf{X}_H^S .

3.2 Multiple Feature Decoupling

To maintain the consistency of image-content and enforce the disentanglement of the low-quality factors and the target style during quality enhancement without any extra annotations, the multiple feature decoupling consists of two main aspects: 1) the purpose of the the quality decoupling is to eliminate interference factors from patch-wise low-quality images, and 2) as shown in Fig1 (III), we consider that the original patch style may be changed during the quality decoupling, and in light of the inconsistency of the high-quality patch styles obtained from low-quality images. The target style decoupling is introduced for aligning the style of patch-wise images.

We are given the unpaired high-quality patch-wise image $x_H^S \in \mathbf{X}_H^S$, low-quality patch-wise image $x_L \in \mathbf{X}_L$, and target style patch-wise image $x_H^T \in \mathbf{X}_H^T$. The encoders $E_{x_L}^C$, $E_{x_H^S}^C$, and $E_{x_H^T}^C$ are employed to extract the content embedding Z_C^L , Z_C^H , and Z_C^T of x_L , x_H^S , and x_H^T , respectively. For x_L and x_H^T , quality-related encoder $E_{x_L}^R$ and style-related encoder $E_{x_H^T}^S$ are designed to extract low-quality factor embedding Z_R^L and style-related embedding Z_S^T . For the quality decoupling, the process of feature extraction and image generation can be formulated as:

$$\begin{aligned} x_L^{\{L \rightarrow H\}} &= G_H^S(E_{x_L}^C(x_L)) \\ x_H^{\{H \rightarrow L\}} &= G_L(E_{x_L}^R(x_L) \oplus E_{x_H^S}^C(x_H^S)), \end{aligned} \quad (2)$$

where \oplus indicates the channel-wise concatenation. The learned Z_C^H and Z_C^L are recombined to generate low-quality image $x_H^{\{H \rightarrow L\}}$ via generator G_L . For high-quality images $x_L^{\{L \rightarrow H\}}$, G_H^S only utilizes the content embedding Z_C^L .

After decoupling the low-quality factors, there is a visible perceptual disparity among the high-quality patches, which suggests the presence of a style gap among the image patches. Hence, the target style decoupling is proposed for aligning image style. Analogous to the quality decoupling, generator G_H^S is used to transform a high-quality image x_H^T with the target style into a general high-quality image $x_H^{\{T \rightarrow S\}}$, while generator G_H^T jointly takes the $Z_S^T = E_{x_H^T}^S(x_H^T)$ and $Z_C^H = E_{x_H^T}^C(x_H^S)$ to generate a new style-changing image $x_H^{\{S \rightarrow T\}}$. We align different style domains by extracting the style embedding of target style domain, and re-feeding the content embeddings from source domains with the decoupled target style embedding to be reconstructed into the image space, thereby alleviating the domain shift problem.

Meanwhile, we define feature-level adversarial loss $\mathcal{L}_{adv}^{Q_e}$ to further encourage the content-related encoders $E_{x_L}^C$, $E_{x_H^S}^C$ and $E_{x_H^T}^C$ to pull their embedding Z_C^L , Z_C^H and Z_C^T together, so that inter-domain content variations are reduced.

$$\begin{aligned} \mathcal{L}_{adv}^{Q_e} = & \mathbb{E}_{x_L} \left[\frac{1}{2} \log D_C(Z_C^L) + \frac{1}{2} \log (1 - D_C(Z_C^L)) \right] \\ & + \mathbb{E}_{x_H^S} \left[\frac{1}{2} \log D_C(Z_C^H) + \frac{1}{2} \log (1 - D_C(Z_C^H)) \right] \\ & + \mathbb{E}_{x_H^T} \left[\frac{1}{2} \log D_C(Z_C^T) + \frac{1}{2} \log (1 - D_C(Z_C^T)) \right], \end{aligned} \quad (3)$$

where D_C denotes a content representation discriminator.

In addition to the adversarial loss on the embedding, we define an image-level adversarial loss $\mathcal{L}_{adv}^{Q_I}$ to constrain the generation of $x_L^{\{L \rightarrow H\}}$, $x_H^{\{H \rightarrow L\}}$, $x_H^{\{T \rightarrow S\}}$ and $x_H^{\{S \rightarrow T\}}$, which can force the content-related, quality-related and target style-related encoders to capture their respective embedding.

$$\begin{aligned} \mathcal{L}_{adv}^{Q_I} = & \mathbb{E}_{x_H^S} [\log D_I(x_H^S)] + \mathbb{E}_{x_L/x_H^T} [\log D_I(x^L/x_H^T)] \\ & + \mathbb{E}_{\{x_H^S, x_L/x_H^T\}} [\log (1 - D_I(x_H^{\{H \rightarrow L\}}/x_H^{\{S \rightarrow T\}}))] \\ & + \mathbb{E}_{\{x_L/x_H^T, x_H^S\}} [\log (1 - D_I(x_L^{\{L \rightarrow H\}}/x_L^{\{T \rightarrow S\}}))], \end{aligned} \quad (4)$$

where D_I denotes an image-level domain discriminator.

Besides the adversarial loss for constraining the image generation in the image space, to supervise the adversarial image generation and fully exploit the low-quality images for the quality enhancement model, self-supervised learning can be naturally harnessed for providing additional supervision. A re-feeding strategy is designed to re-disentangle and reconstruct new generated images for image-level self-supervision in the original image space and the disentangled feature space. Specifically, we reconstruct $x_L^{\{L \rightarrow H\}}$ and

$x_H^{\{H \rightarrow L\}}$ as \hat{x}_L and \hat{x}_{H1}^S in the original image space via the encoders $\{E_{x_L}^C, E_{x_L}^R, E_{x_H^S}^C\}$ as well as the generators $\{G_L, G_H^S\}$ according to the same procedure as Eq.2. Hence, the quality-related self-supervision loss \mathcal{L}_Q for supervision on the reconstructed images can be defined as:

$$\begin{aligned}\mathcal{L}_Q &= \mathcal{L}_{rec}^L(x_L, \hat{x}_L) + \mathcal{L}_{rec}^H(x_H^S, \hat{x}_{H1}^S) \\ &= \mathbb{E}_{x_L} [\|\hat{x}_L - x_L\|_1] + \mathbb{E}_{x_H^S} [\|\hat{x}_{H1}^S - x_H^S\|_1],\end{aligned}\quad (5)$$

where $\|\cdot\|_1$ indicates the L_1 norm that is widely adopted in self-supervised learning for preserving the consistency of the reconstructed images. Similarly, due to the lack of reference images corresponding to the inputs, we reconstruct $x_H^{\{T \rightarrow S\}}$ and $x_H^{\{S \rightarrow T\}}$ into the space of the original images \hat{x}_H^T and \hat{x}_H^S . An identical style-related self-supervised learning loss $\mathcal{L}_T = \mathcal{L}_{rec}^T(x_H^T, \hat{x}_H^T) + \mathcal{L}_{rec}^H(x_H^S, \hat{x}_H^S)$ is proposed to align style while maintaining pathological and structural features.

In summary, through the self-supervised domain adaptation strategy, we can decouple quality-related, target style-related and content-related embedding, and further recombine content-related and style-related embedding so as to unify the style of the patch-wise fundus image while removing low-quality interference factors. As a result, the proposed method enables the improvement of fundus image quality without any high-quality reference images.

3.3 Jointly Optimizing

The final self-supervised loss \mathcal{L}_s is designed to:

$$\mathcal{L}_s = \mathcal{L}_Q + \mathcal{L}_T + \mathcal{L}_C,\quad (6)$$

where $\mathcal{L}_C = \mathbb{E}_{x_H^S} [\|\hat{x}_{H1}^S - \hat{x}_{H'}^S\|_1]$ to constrain \hat{x}_{H1}^S and $\hat{x}_{H'}^S$ to be similar, which establishes an intrinsic link between the quality-related and style-related disentanglement. Then the overall objective function \mathcal{L} of DASQE can be formulated as:

$$\mathcal{L} = \mathcal{L}_s + \lambda_1 \mathcal{L}_{adv}^{Q_e} + \lambda_2 \mathcal{L}_{adv}^{Q_I},\quad (7)$$

where $\lambda_1 = \lambda_2 = 0.1$ are regularization parameters to balance the losses \mathcal{L}_s , $\mathcal{L}_{adv}^{Q_e}$, and $\mathcal{L}_{adv}^{Q_I}$.

4 Experiments

In this section, we conduct experiments on two benchmarks including EyeQ [Fu *et al.*, 2019] and Messidor datasets [Deencière *et al.*, 2014] for evaluating the performance of quality enhancement of the proposed method. Moreover, we further investigate the effects of the enhanced images on extensive medical image analysis tasks, such as lesion segmentation, vessel segmentation, and disease grading.

4.1 Datasets and Evaluation Metrics

EyeQ dataset is a large-scale public benchmark for fundus image quality assessment, which consists of 28,792 fundus images with different quality labels, including high-quality, usable-quality and low-quality. It is worth mentioning that we re-label the images of "Usable-quality" with "low-quality" for pre-training the binary quality assessment network.

Messidor dataset includes 1,200 fundus images from three

Methods	PSNR	SSIM	Number of low-quality images	Number of reference images	Types
LIME(2016)	13.54	0.868	/	/	Traditional
Fu et al.(2014)	9.76	0.564	/	/	Traditional
He et al.(2010)	15.56	0.759	/	/	Traditional
cofe-net(2020)	20.51	0.885	7196	7196	Supervised
cGAN(2017)	26.35	0.894	7196	7196	Supervised
I-SECRET(2021)	27.36	0.908	12543	7196	Semi-Supervised
CutGAN(2020)	22.76	0.872	7196	7196	Unsupervised
Cycle-CBAM(2019)	21.56	0.843	7196	7196	Unsupervised
StillGAN(2021)	25.38	0.894	7196	7196	Unsupervised
DASQE	28.47	0.914	7196	/	Unsupervised

Table 1: The comparison between our method with the SOTA methods for low-quality fundus image enhancement on EyeQ dataset.

medical institutions, and the DR grading annotations are provided to measure the severity of diabetic retinopathy.

Evaluation Metrics. We use peak signal-to-noise ratio (PSNR) and structural index measure (SSIM) [Hore and Ziou, 2010] to evaluate the quality enhancement performance. In addition, it should be noted that for the quantitative metrics PSNR and SSIM, paired images are required. Due to the nonexistence of low-quality images in the relevant datasets, we follow the degradation pipeline proposed by [Shen *et al.*, 2020] to synthesize the corresponding low-quality dataset to measure the performance of image enhancement. In addition, in order to investigate the influence of higher quality images enhanced by our method on DR grading task, we also further train the disease grading network on the messdior dataset. The accuracy and quadratic weighted kappa are chosen to evaluate the grading performance.

4.2 Implementation

In this paper, we implement the proposed method using PyTorch with 4 NVIDIA Quadro RTX 6000 GPUs. The architecture of encoder and decoder is similar to the one in [Liu *et al.*, 2017], and the weights are not shared across encoders. Due to the large-sized and diversity of the original images, all images are resized to 512×512 , and the size of patch-wise fundus images is set to 128×128 . During the training, we apply Adam optimizer to update the parameters with *learning rate*=0.0001, *momentum*=0.9, *batch size*=32. The maximum number of training iterations is set to 300K. The weights of the proposed method are initialized from a Gaussian distribution $\mathcal{N}(0, 0.02)$.

4.3 Comparison with State-of-the-Art Methods

In this part, we provide a comprehensive comparison of the proposed method with the traditional and deep learning-based methods. The traditional image enhancement methods include: LIME [Guo *et al.*, 2016], distribution fitting [Fu *et al.*, 2014], and latent structure-drive [He *et al.*, 2010]. Deep learning-based approaches include: cofe-Net [Shen *et al.*, 2020], I-SECRET [Cheng *et al.*, 2021], cGAN [Isola *et al.*, 2017], CutGAN [Park *et al.*, 2020], CycleCAN [You *et al.*, 2019], and StillGAN [Ma *et al.*, 2021]. Experimental results are reported in Table 1 where the best results are boldfaced. The proposed method, DASQE, surpasses all the compared methods significantly, in terms of PSNR and SSIM without

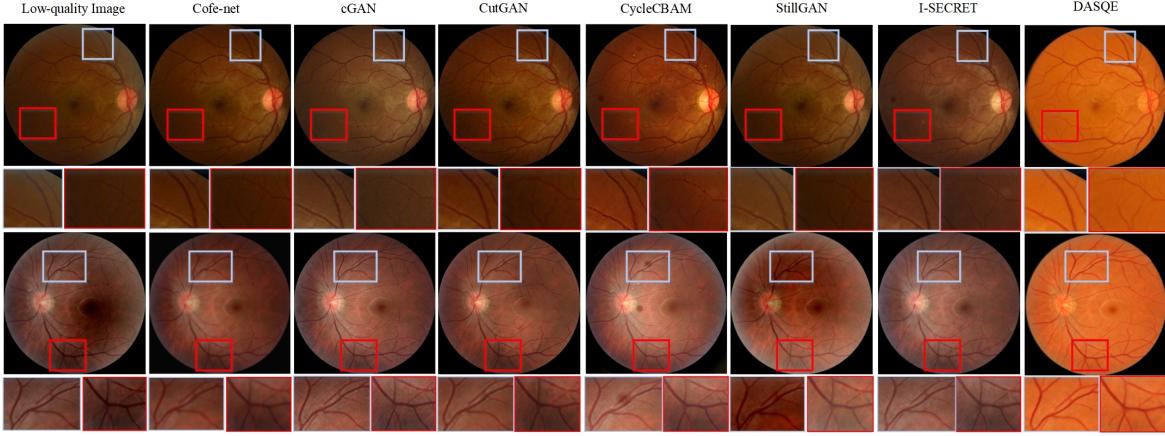


Figure 4: Visual comparisons on the low-quality image enhancement between the proposed and other deep learning methods.

Methods	PSNR	SSIM	Methods	PSNR	SSIM
LIME	12.36	0.852	cGAN	25.83	0.867
Fu et al.	11.32	0.573	CutGAN	24.37	0.883
He et al.	16.23	0.764	Cycle-CBAM	23.56	0.851
cofe-net	21.97	0.852	StillGAN	26.32	0.871
I-SECRET	26.73	0.874	DASQE	29.36	0.897

Table 2: The comparison between our method with the SOTA methods for low-quality image enhancement on the Messidor dataset.

the guidance of any high-quality reference images. Although these comparable deep learning methods obtain better results than the traditional methods, the fact that they require pairwise high-quality images largely reduces their practicality for the problem that we aim to solve. Notably, no high-quality images are available for our self-supervised learning method, and the supervision is derived from the image data itself, we also further visualize some cases of low-quality image enhancement for a range of comparable methods in Fig.4. Although all methods exhibit decent ability of enhancing low-quality fundus images, unsupervised methods based on domain transfer suffer from apparent domain shifts. For example, the enhanced images obtained by Cycle-CBAM introduce undesired artifacts, which might mislead ophthalmologists to diagnose fundus disease. For StillGAN, the blurring of low-quality images is not well improved. In addition, due to the fact that the effect of style shifts on image enhancement is not considered, the comparable methods show limited improvement for uneven illumination. Our method presents remarkable performance on low-quality fundus image enhancement, and alleviates the domain shift problem, resulting in better visual perception. In order to prove the robustness of the proposed method for low-quality image enhancement on other datasets, we also report the experimental results of the proposed method on the Messidor dataset in Table 2. From the results, we can observe that our method again achieves competitive performance against the other methods.

4.4 Ablation Study

In this section, we perform a series of ablation experiments to analyze the effect of patch-wise image domains and multiple

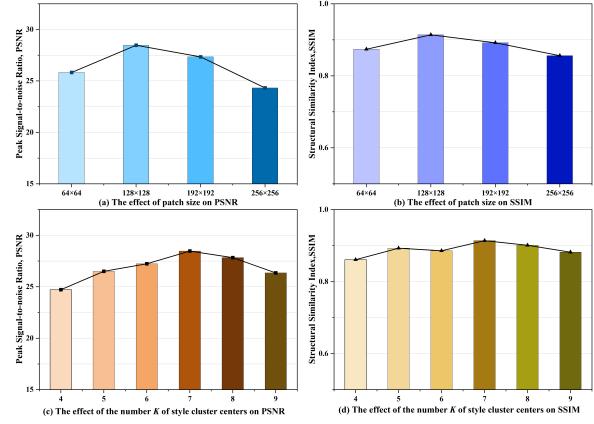


Figure 5: Comparison of low-quality fundus image enhancement results based on the different patch sizes (a-b) and different number K of style cluster centers (c-d).

feature decoupling for image quality enhancement. During the construction of the patch-wise image domain, there are two key hyper-parameters, including the size of the patch and the number of style cluster centers. We first explore the effect of the sizes of patches on quality enhancement. Based on the EyeQ dataset, we report the quality enhancement performance at different patch sizes in Fig. 5(a-b). It can be found that the best performance metrics for quality enhancement are achieved when the patch size is 128×128 . Furthermore, both oversized and undersized patches result in degraded performance of image enhancement. The reason is that the number of low-quality regions is affected by the size of the patch-wise images. When the size of patches changes from 128×128 to 64×64 , the number of high-quality patch-wise images increases. This may lead to more incorrect quality assessment results for pre-trained quality assessment networks. In contrast, when the size of patches changes from 128×128 to 256×256 , the performance of quality enhancement degrades with the reduction in the number of high-quality patches.

In addition to the patch size, we also explore the intrinsic relationship between the number of style clusters and low-

Variants	EyeQ		Messidor	
	PSNR	SSIM	PSNR	SSIM
Remove the quality-related disentanglement	16.32	0.768	17.63	0.758
Remove the style-related disentanglement	19.24	0.832	18.52	0.817
DASQE	28.47	0.914	29.36	0.897

Table 3: Quantitative results for low-quality image enhancement.

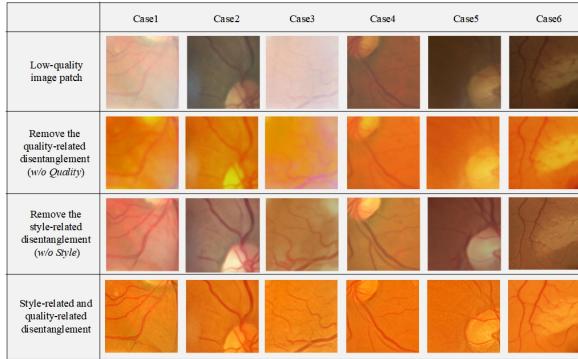


Figure 6: The effect of the different disentanglement branches for the patch-wise image quality enhancement.

quality image enhancement as shown in Fig 5(c-d). We observe that the proposed method achieves the best performance when the number of clusters is set to 7. More specifically, 1) when the number of style clusters is set from [4, 6], the clustering happens to overlap, which leads to inconsistent styles of patch-level images in the target style domain. 2) when the number of style clusters exceeds 7, there are prominent outlier points in the clustering results, which increases the complexity of optimizing for style disentanglement.

To investigate the effectiveness of multiple feature decoupling for image quality enhancement, we compare the proposed method with its two variants, respectively. Quantitative results are shown in Table 3. When the quality decoupling branch is removed, the performance metrics PSNR/SSIM for the quality enhancements decrease to 16.32/0.768 on the EyeQ dataset. This reveals that the method fails to effectively improve the quality of the fundus image. When we remove the style decoupling branch and keep the quality decoupling branch, the PSNR also drops to 19.24 on EyeQ dataset. Our results suggest that the simultaneously decoupling of both quality and style features is critical for improving the quality of the fundus image. To demonstrate the effectiveness of the different decoupling branches more intuitively, we also further visualize some patch-wise fundus images in Fig.6. As shown in the 2nd row in Fig.6, only the style-related disentanglement branch is incapable of enhancing the quality of the images. For example, the retinal vessels of patch-wise images do not appear clearer by introducing style disentanglement. The low-quality factors of images are also not eliminated. From the 3rd row in Fig.6, we can observe that although *w/o Style* is able to transform general knowledge from patch-wise low-quality domains to patch-wise high-quality

Methods	ACC	Kappa	Methods	ACC	Kappa
w/o enhancement	0.681	0.657	CutGAN	0.717	0.694
cofe-net	0.704	0.687	Cycle-CBAM	0.701	0.682
I-SECRET	0.735	0.718	StiLiGAN	0.724	0.706
cGAN	0.731	0.712	DASQE	0.752	0.734

Table 4: The DR grading performance on the Messidor dataset.

domains, there exist a large style gap for the enhanced patch-wise images, which is also detrimental to the quality enhancement of image-wise images.

4.5 Application

To investigate the effects of the DASQE on downstream medical image analysis tasks, we present a series of visual comparison results in Fig. 7. The experimental results show that the proposed method enables considerable performance improvements in the various fundus image analysis tasks, including vessel/lesion segmentation and disease grading.

Vessel segmentation: For computer-aided diagnosis of ophthalmic diseases, the precise segmentation of retinal vessels from fundus images is an important prerequisite. Such as, in *proliferative diabetic retinopathy* (PDR), the number of neovascularisations is a key indicator to reflect its severity. For the vessel segmentation of fundus images, we test CE-Net [Gu *et al.*, 2019] pre-trained on DRIVE dataset [Staal *et al.*, 2004] for vessel segmentation on the enhanced images of proposed and comparable methods. As shown in the 2nd column of Fig.7, we visualize some cases of vessel segmentation. As can be seen that CE-Net can capture more fine-grained vessel segmentation results of the enhanced fundus images obtained by our proposed method DASQE.

Lesion segmentation: In *non-proliferative diabetic retinopathy* (NPDR), the number of retinal hemorrhages (HEs) is an important indicator of disease severity. To investigate the effect of the proposed method on the segmentation of small lesions (HEs), we test UNet3+ [Huang *et al.*, 2020] pre-trained on the IDRiD dataset [Porwal *et al.*, 2018] for HEs segmentation on enhanced images. The segmentation results of HEs are shown in the 3rd column of Fig.7. Our method can produce enhanced images with more clear structures of clinical lesions, which can be precisely identified by ophthalmologists or automated diagnostic systems.

Disease grading: To verify the influence of the quality improvement by our method on the disease grading, we further conduct the comparison among the improved-quality images by competing methods using a DenseNet-121 [Huang *et al.*, 2017] as the grading model. The DR grading results are shown in Table 4. We observe that our method is the most beneficial for the disease grading of the classification model.

5 Conclusion

The enhancement methods of low-quality fundus images have long been considered to be a critical step before the diagnosis of eye disease. In this work, we propose a novel reference-free method to extend a self-supervised learning framework to the quality enhancement task. Our approach is the first to address the image quality enhancement problem from a complete self-supervised learning perspective.

	RGB fundus Image	Vessel segmentation	HEs segmentation
Low-quality Image			
Cofe-net			
cGAN			
CutGAN			
CycleCBAM			
StillGAN			
I-SECRET			
DASQE			

Figure 7: Visualization of Vessel and Lesion(HEs) segmentation on low-quality images and enhanced images.

By combining the strengths of self-supervised learning and domain adaptation, the proposed approach not only significantly improves the state-of-the-art results in image quality enhancement on multiple benchmark datasets but also effectively aligns the illumination style of the fundus images. Herein, we focus on this more challenging and more widely applicable approach. The results show that our DASQE framework can produce high-quality images of better generalizability and robustness compared to state-of-the-art quality

improvement methods on the various image analysis tasks. We strongly believe that the presented line of research is worth pursuing further.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No.62076059), and the Science Project of Liaoning province (2021-MS-105).

References

- [Cao *et al.*, 2022] Peng Cao, Qingshan Hou, Ruoxian Song, Haonan Wang, and Osmar Zaiane. Collaborative learning of weakly-supervised domain adaptation for diabetic retinopathy grading on retinal images. *Computers in Biology and Medicine*, 144:105341, 2022.
- [Cheng *et al.*, 2021] Pujin Cheng, Li Lin, Yijin Huang, Junyan Lyu, and Xiaoying Tang. I-secret: Importance-guided fundus image enhancement via semi-supervised contrastive constraining. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 87–96. Springer, 2021.
- [Dai *et al.*, 2021] Ling Dai, Liang Wu, Huating Li, Chun Cai, Qiang Wu, Hongyu Kong, Ruhan Liu, Xiangning Wang, Xuhong Hou, Yuexing Liu, et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature communications*, 12(1):1–11, 2021.
- [Decencière *et al.*, 2014] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordóñez, Pascale Massin, Ali Erginay, et al. Feedback on a publicly distributed image database: the messidor database. *Image Analysis & Stereology*, 33(3):231–234, 2014.
- [Fu *et al.*, 2014] Xueyang Fu, Peixian Zhuang, Yue Huang, Yinghao Liao, Xiao-Ping Zhang, and Xinghao Ding. A retinex-based enhancing approach for single underwater image. In *2014 IEEE international conference on image processing (ICIP)*, pages 4572–4576. IEEE, 2014.
- [Fu *et al.*, 2019] Huazhu Fu, Boyang Wang, Jianbing Shen, Shanshan Cui, Yanwu Xu, Jiang Liu, and Ling Shao. Evaluation of retinal image quality assessment networks in different color-spaces. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 48–56. Springer, 2019.
- [Gu *et al.*, 2019] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging*, 38(10):2281–2292, 2019.
- [Gui *et al.*, 2021] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [Guo *et al.*, 2016] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination

- map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016.
- [He *et al.*, 2010] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.
- [Hore and Ziou, 2010] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010.
- [Huang *et al.*, 2017] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [Huang *et al.*, 2020] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1055–1059. IEEE, 2020.
- [Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
- [Jin *et al.*, 2019] Qiangguo Jin, Zhaopeng Meng, Tuan D Pham, Qi Chen, Leyi Wei, and Ran Su. Dunet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, 178:149–162, 2019.
- [Li *et al.*, 2019] Xiaomeng Li, Xiaowei Hu, Lequan Yu, Lei Zhu, Chi-Wing Fu, and Pheng-Ann Heng. Canet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. *IEEE transactions on medical imaging*, 39(5):1483–1493, 2019.
- [Liu *et al.*, 2017] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30, 2017.
- [Ma *et al.*, 2021] Yuhui Ma, Jiang Liu, Yonghuai Liu, Huazhu Fu, Yan Hu, Jun Cheng, Hong Qi, Yufei Wu, Jiong Zhang, and Yitian Zhao. Structure and illumination constrained gan for medical image enhancement. *IEEE Transactions on Medical Imaging*, 40(12):3955–3967, 2021.
- [Park *et al.*, 2020] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European conference on computer vision*, pages 319–345. Springer, 2020.
- [Porwal *et al.*, 2018] Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, and Fabrice Meriaudeau. Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research. *Data*, 3(3):25, 2018.
- [Shen *et al.*, 2020] Ziyi Shen, Huazhu Fu, Jianbing Shen, and Ling Shao. Modeling and enhancing low-quality retinal fundus images. *IEEE transactions on medical imaging*, 40(3):996–1006, 2020.
- [Shome and Vadali, 2011] Saikat Kumar Shome and Siva Ram Krishna Vadali. Enhancement of diabetic retinopathy imagery using contrast limited adaptive histogram equalization. *International Journal of Computer Science and Information Technologies*, 2(6):2694–2699, 2011.
- [Staal *et al.*, 2004] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.
- [Teo *et al.*, 2021] Zhen Ling Teo, Yih-Chung Tham, Marco Yu, Miao Li Chee, Tyler Hyunghae Rim, Ning Cheung, Mukharram M Bikbov, Ya Xing Wang, Yating Tang, Yi Lu, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology*, 128(11):1580–1591, 2021.
- [Tian and Cohen, 2017] Qi-Chong Tian and Laurent D Cohen. Global and local contrast adaptive enhancement for non-uniform illumination color images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 3023–3030, 2017.
- [Wang *et al.*, 2018] Renzhen Wang, Benzhi Chen, Deyu Meng, and Lisheng Wang. Weakly supervised lesion detection from fundus images. *IEEE transactions on medical imaging*, 38(6):1501–1512, 2018.
- [Wang *et al.*, 2021] Xiaofei Wang, Mai Xu, Jicong Zhang, Lai Jiang, Liu Li, Mengxian He, Ningli Wang, Hanruo Liu, and Zulin Wang. Joint learning of multi-level tasks for diabetic retinopathy grading on low-resolution fundus images. *IEEE Journal of Biomedical and Health Informatics*, 26(5):2216–2227, 2021.
- [You *et al.*, 2019] Qijing You, Cheng Wan, Jing Sun, Jianxin Shen, Hui Ye, and Qiuli Yu. Fundus image enhancement method based on cyclegan. In *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 4500–4503. IEEE, 2019.
- [Zhou *et al.*, 2017] Mei Zhou, Kai Jin, Shaoze Wang, Juan Ye, and Dahong Qian. Color retinal image enhancement based on luminosity and contrast adjustment. *IEEE Transactions on Biomedical engineering*, 65(3):521–527, 2017.