

# A Reference-free Self-supervised Domain Adaptation Framework for Low-quality Fundus Image Enhancement

Qingshan Hou  
School of Computer Science and  
Engineering, Northeastern University  
Shenyang, Liaoning, China  
Key Laboratory of Intelligent  
Computing in Medical Image of  
Ministry of Education, Northeastern  
University  
Shenyang, Liaoning, China  
houqingshancv@gmail.com

Peng Cao\*  
School of Computer Science and  
Engineering, Northeastern University  
Shenyang, Liaoning, China  
Key Laboratory of Intelligent  
Computing in Medical Image of  
Ministry of Education, Northeastern  
University  
Shenyang, Liaoning, China  
caopeng@cse.neu.edu.cn

Jiaqi Wang  
School of Computer Science and  
Engineering, Northeastern University  
Shenyang, Liaoning, China  
Key Laboratory of Intelligent  
Computing in Medical Image of  
Ministry of Education, Northeastern  
University  
Shenyang, Liaoning, China  
wjq010222@gmail.com

Xiaoli Liu  
School of Computer Science and  
Engineering, Northeastern University  
Shenyang, Liaoning, China  
Key Laboratory of Intelligent  
Computing in Medical Image of  
Ministry of Education, Northeastern  
University  
Shenyang, Liaoning, China  
neuxiaoliliu@gmail.com

Jinzhong Yang\*  
School of Computer Science and  
Engineering, Northeastern University  
Shenyang, Liaoning, China  
Key Laboratory of Intelligent  
Computing in Medical Image of  
Ministry of Education, Northeastern  
University  
Shenyang, Liaoning, China  
yangjinzhong@cse.neu.edu.cn

Osmar R. Zaiane  
Alberta Machine Intelligence Institute,  
University of Alberta  
Edmonton, Canada  
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 fundus 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, and proposed a Domain Adaptation Self-supervised Quality Enhancement framework, named DASQE. Specifically, we construct multiple patch-wise domains via a well-designed rule-based quality assessment scheme and 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 factors and style information by exploring intrinsic supervision signals within the low-quality images. Extensive experiments are conducted on four benchmark datasets, and results show that our DASQE method achieves new state-of-the-art performance when only low-quality images are available.

## CCS CONCEPTS

• **Computing methodologies** → **Computer vision; Reconstruction; Biometrics.**

## KEYWORDS

reference-free, domain adaptation, self-supervised, fundus image quality enhancement

### ACM Reference Format:

Qingshan Hou, Peng Cao, Jiaqi Wang, Xiaoli Liu, Jinzhong Yang, and Osmar R. Zaiane. 2023. A Reference-free Self-supervised Domain Adaptation Framework for Low-quality Fundus Image Enhancement. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*, October 29–November 3, 2023, Ottawa, ON, Canada. ACM, Ottawa, ON, Canada, 11 pages. <https://doi.org/10.1145/3581783.3612049>

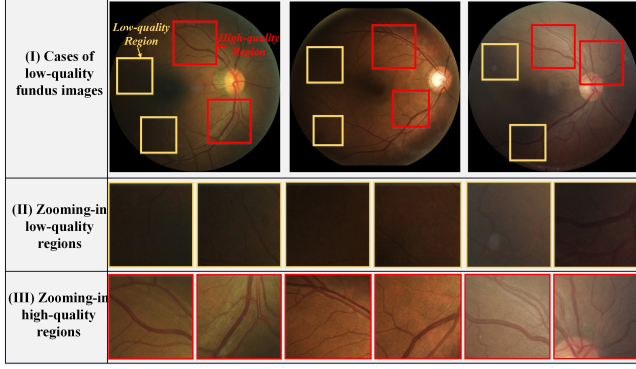
## 1 INTRODUCTION

Medical fundus images have been extensively used for clinical analysis of various ocular diseases [7, 26, 36, 50]. However, the real clinical fundus datasets usually contain a large number of low-quality images. The quality of fundus images is critical to

\*corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
MM '23, October 29–November 3, 2023, Ottawa, ON, Canada.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 979-8-4007-0108-5/23/10...\$15.00  
<https://doi.org/10.1145/3581783.3612049>



**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 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 vessel/lesion segmentation. Existing deep learning methods [5, 12, 34] 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 Figure 1: not all the regions are low quality, as shown in Figure 1 (I). Besides, there are significant differences among regions in low-quality images with respect to the image styles, as shown in Figure 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 simultaneously 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.

Our work rethinks 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 rule-based quality assessment scheme. Given the high-quality patch-wise 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 embedding 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 medical fundus image quality enhancement 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 medical fundus image quality enhancement task. We tackle fundus image quality enhancement in a fully unsupervised setting.

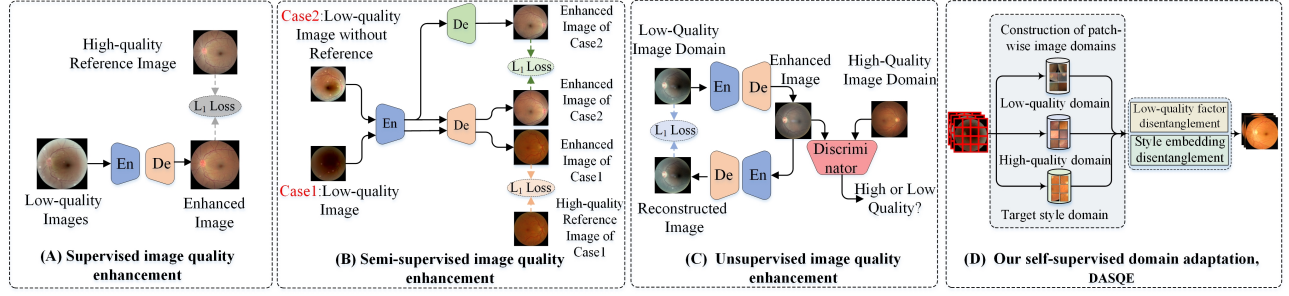
2) To enhance low-quality fundus images while preserving pathological features and major retinal structures, we propose a representation decoupling strategy for disentangling content-related, style-related and quality-related embeddings. The low-quality fundus images are enhanced by recombining the learned 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 domain adaptation framework without the guidance of any high-quality data significantly outperforms the state-of-the-art methods by a considerable 28.32 PSNR / 91.5% SSIM and 29.14 PSNR / 89.2% SSIM on EyeQ and Messidor benchmark datasets, a 0.96/2.41 PSNR and 0.7/1.8% 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 in the appendix section, such as retinal vessel segmentation, lesion segmentation and disease grading.

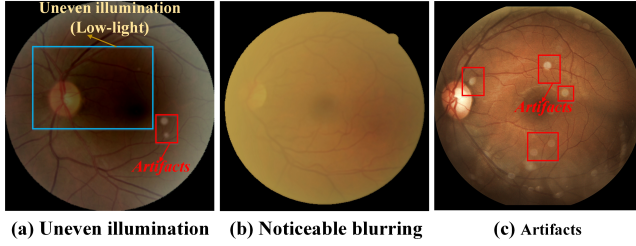
## 2 RELATED WORK

### 2.1 Quality enhancement challenges for natural and medical fundus images

The data-driven deep learning methods have been widespread used in natural image enhancement and other fields [27, 28, 38, 39, 44, 45] due to their impressive feature extraction and representation abilities. However, natural image enhancement methods usually focus only on a specific aspect, such as low-light image enhancement [13, 30, 42, 43], dehazing and deraining [3, 4, 23], and deblurring [6, 24]. Unlike the focus of natural image enhancement tasks,



**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.**



**Figure 3: The low-quality factors in fundus images include: (a) uneven illumination, (b) noticeable blurring, and (c) artifacts.**

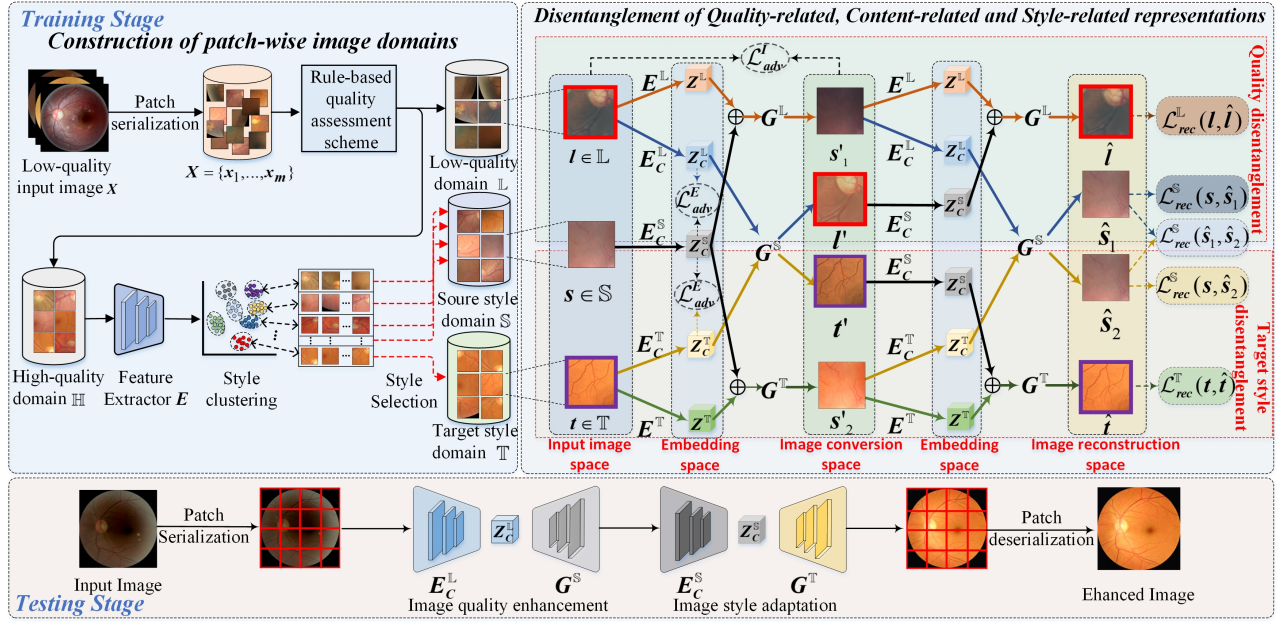
medical image enhancement aims to obtain a clear and accurate representation of internal anatomy or physiological processes from medical imaging modalities for the purposes of clinical diagnosis, treatment planning, and monitoring of disease progression. Therefore, enhancing low-quality fundus images is more challenging than natural images. The challenges mainly lie in: (1) Low-quality factors (e.g. artifacts, noticeable blurring and uneven illumination, as shown in Figure 3.) on fundus images are more diverse compared to natural images. (2) The local lesion regions are critical for clinical decisions. However, the lesion-related regions on the fundus image are small and complex, which require that clear lesion boundaries and fundus structures to be recognized. Hence, the natural image enhancement methods (e.g. low-light image enhancement, deblurring and so on.) are not adequate for low-quality fundus image enhancement.

## 2.2 Technical progress for fundus image quality enhancement

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 et al. [35] enhance the contrast of fundus images by the contrast limited adaptive histogram equalization method. Tian et al. [37] design a global and local contrast enhancement method for the quality enhancement of non-uniform illumination images. Zhou et al. [48] 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, 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 [22], lesion detection [40], disease classification [2]. For the quality enhancement of fundus images, the most common quality enhancement methods [34, 41] based on deep learning belong to the fully supervised learning scheme (Figure 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 (Figure 2 (B)) [5] is proposed, which requires only a portion of paired images during the training phase. Inspired by generative adversarial learning, some unsupervised methods (Figure 2 (C)) [31, 46] 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.



**Figure 4: The overall architecture of the proposed framework.** It involves two stages: 1) **Construction of patch-wise image domains:** The low-quality input images  $X$  are first serialized as the patch-wise fundus images  $X = \{x_1, \dots, x_m\}$ . With the help of the rule-based quality assessment scheme, the patch-wise fundus images are divided into low-quality and high-quality patch domains, e.g.  $\mathbb{L}, \mathbb{H}$ . Then, a style clustering is performed on  $\mathbb{H}$ , and a Source/Target style patch domain  $\mathbb{S}/\mathbb{T}$  is determined. 2) **Disengagement of multiple features:** We factorize content-related, quality-related and style-related for patch-wise domains through the supervision derived from the image data itself to enhance the image quality and align style features. The framework involves content-related encoders  $\{E_C^L, E_C^S, E_C^T\}$  for extracting latent content embeddings  $\{Z_C^L, Z_C^S, Z_C^T\}$  of  $\{l; s; t\}$ , quality-related encoder  $E^L$  for extracting low-quality factor embedding  $Z^L$  of  $l$ , target style-related encoder  $E^T$  for extracting target style embedding of  $Z^T$  of  $t$ , generators  $G^L, G^S, G^T$  for generating low-quality patch, source and target style patches.

### 3 METHODOLOGY

In this section, we present a clinically oriented fundus image quality enhancement method DASQE to address the low-quality image enhancement problem. Specifically, our framework is designed and built to address two challenges: How to construct the patch-wise domains? How to design auxiliary supervision to improve the representation disentanglement for better generalization ability of DASQE? An overview of the method is illustrated in Figure 4. A summary of key notations used in this paper is reported in Table 1.

#### 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 contains the highest number of patches. Formally, the low-quality fundus images  $X$  are serialized into a patch-wise image set  $X = \{x_1, \dots, x_m\}$  as shown in Figure 4. Then, according to the rule-based quality assessment scheme,  $X$  are divided into the low-quality patch domain  $\mathbb{L} = \{l_1, \dots, l_n\}$  and high-quality patch domain  $\mathbb{H} = \{h_1, \dots, h_{n'}\}$ , where  $m = n + n'$ .

More specifically, the peak signal-to-noise ratio (PSNR) is generally higher between the patches with the similar quality compared to the patches with different qualities. Consequently, the proposed

rule-based patch quality assessment scheme involves the following main steps:

##### *Initializing the high-/low-quality quality domains.*

Utilizing an auto-encoder to encode the patches into embeddings for the following cluster, which aims to group the patches into a high-quality domain  $\mathbb{H}$  and a low-quality domain  $\mathbb{L}$ .

##### *Updating the high-/low-quality domains.*

1) The average PSNR  $AVG_{PSNR}$  between patches within  $\mathbb{H}$  is calculated as following:

$$AVG_{PSNR} = \frac{2 \sum_{i=1}^{n'} \sum_{j=i+1}^{n'} P(h_i, h_j)}{n' * (n' - 1)} \quad (1)$$

where  $P(\cdot)$  indicates a function of calculating the PSNR between a pair of patch  $h_i$  and patch  $h_j$ , and  $n'$  denotes the number of patches in  $\mathbb{H}$ .

2) Randomly select patch pairs  $h_i$  and  $h_j$  from  $\mathbb{H}$  and calculate their PSNR  $p(h_i, h_j)$ . If the value of  $p(h_i, h_j)$  is lower than  $AVG_{PSNR}$ , we calculate the average PSNR between  $h_i$  and the patches in  $\mathbb{H}$ , and the one between  $h_j$  and the patches in  $\mathbb{H}$ , respectively. Choose the patch with the larger average PSNR as the reference patch.



**Table 1: Important notations in this study**

Notation	Description
$\mathbb{L}; l$	Low-quality patch domain; Low-quality patch;
$l'$	Conversion of low-quality patch to high-quality patch;
$\hat{l}$	Reconstruction of low-quality patch
$\mathbb{S}$	High-quality Source style patch domain;
$s$	High-quality Source style patch
$s'_1$	Conversion of high-quality source style patch to low-quality patch;
$\hat{s}_1$	Reconstruction of high-quality source style patch during the quality disentanglement
$s'_2$	Conversion of high-quality source style patch to high-quality target style patch;
$\hat{s}_2$	Reconstruction of high-quality source style patch during the style disentanglement
$\mathbb{T}$	High-quality Target style patch domain;
$t$	High-quality Target style patch;
$t'$	Conversion of high-quality target style patch to high-quality source style patch;
$\hat{t}$	Reconstruction of high-quality target style patch
$E^{\mathbb{L}}$	Quality-related encoder for extracting low-quality factor embedding of $l$
$E^{\mathbb{T}}$	Target Style-related encoder for extracting target style embedding of $t$
$E_C^{\mathbb{L}}; E_C^{\mathbb{S}}; E_C^{\mathbb{T}}$	Content-related encoders for extracting latent content embeddings of $l; s; t$
$Z^{\mathbb{L}}$	Low-quality factor embedding of $l$
$Z^{\mathbb{T}}$	Target style embedding of $t$
$Z_C^{\mathbb{L}}; Z_C^{\mathbb{S}}; Z_C^{\mathbb{T}}$	Latent Content embeddings of $l; s; t$
$G^{\mathbb{L}}$	Low-quality patch generator
$G^{\mathbb{S}}$	High-quality Source style patch generator
$G^{\mathbb{T}}$	High-quality Target style patch generator

3) In order to more accurately remove low-quality patches from the high-quality domain, repeat step 2) to obtain multiple reference patches.

4) Calculate the average PSNR between each remaining patch in  $\mathbb{H}$  and the chosen reference patches, respectively. Remain the patch in  $\mathbb{H}$  if its average PSNR is higher than  $AVG_{PSNR}$ , otherwise move it to  $\mathbb{L}$ .

For the patch-wise high-quality domains  $\mathbb{H}$ , 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 disentanglement. More specifically, our primary goal is to learn an embedding function  $De(\cdot)$  mapping  $\mathbb{H}$  to embeddings  $Z = \{z_1, z_2, \dots, z_{n'}\}$  in a  $D$ -dimension representation space. The loss  $\mathcal{L}_e$  of latent embedding learning is defined as:

$$\mathcal{L}_e = \mathbb{E}_h [\|De(En(h)) - h\|_1], \quad (2)$$

where  $h$  indicates the patch-wise images from  $\mathbb{H}$ ,  $En(\cdot)$  and  $De(\cdot)$  denote the encoder and decoder, respectively. Subsequently, we cluster the obtained embeddings  $Z = En(\mathbb{H})$  to obtain the style domains  $\{\mathbb{H}_1, \mathbb{H}_2, \dots, \mathbb{H}_k\}$ . Finally, we select a domain containing the highest number of patches as the Target style domain  $\mathbb{T}$ , and the others are considered as Source style domains  $\mathbb{S}$ .

### 3.2 Multiple Feature Disentanglement

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 disentanglement consists of two main aspects: 1) the purpose of the quality disentanglement is to eliminate interference factors from low-quality patches, and 2) as shown in Fig1 (III), we consider that the original patch style may be changed during the quality disentanglement, and in light of the inconsistency of the high-quality patch styles obtained from low-quality images. The target style disentanglement is introduced for aligning the style of patches.

We are given the unpaired low-quality patch  $l \in \mathbb{L}$ , high-quality source style patch  $s \in \mathbb{S}$ , and high-quality target style patch  $t \in \mathbb{T}$ . The encoders  $E_C^{\mathbb{L}}$ ,  $E_C^{\mathbb{S}}$ , and  $E_C^{\mathbb{T}}$  are employed to extract the content embedding  $Z_C^{\mathbb{L}}$ ,  $Z_C^{\mathbb{S}}$ , and  $Z_C^{\mathbb{T}}$  of  $l$ ,  $s$ , and  $t$ , respectively. For  $l$  and  $t$ , quality-related encoder  $E^{\mathbb{L}}$  and style-related encoder  $E^{\mathbb{T}}$  are designed to extract low-quality factor embedding  $Z^{\mathbb{L}}$  and style-related embedding  $Z^{\mathbb{T}}$ . For the quality disentanglement, the process of feature extraction and image generation can be formulated as:

$$l' = G^{\mathbb{S}} \left( E_C^{\mathbb{L}}(l) \right); \quad s'_1 = G^{\mathbb{L}} \left( E^{\mathbb{L}}(l) \oplus E_C^{\mathbb{S}}(s) \right), \quad (3)$$

where  $\oplus$  indicates the channel-wise concatenation. The learned  $Z_C^{\mathbb{S}} = E_C^{\mathbb{S}}(s)$  and  $Z^{\mathbb{L}} = E^{\mathbb{L}}(l)$  are recombined to generate low-quality patch  $s'_1$  via generator  $G^{\mathbb{L}}$ . For  $l'$ ,  $G^{\mathbb{S}}$  only utilizes the content embedding  $Z_C^{\mathbb{L}} = E_C^{\mathbb{L}}(l)$  of  $l$ .

After disentangling 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 patches. Hence, the target style disentangling is proposed for aligning patch style. Analogous to the quality disentangling, generator  $G^{\mathbb{S}}$  is used to convert a high-quality patch  $t$  with the target style into a high-quality patch  $t'$  with the source style, while generator  $G^{\mathbb{T}}$  jointly takes the  $Z^{\mathbb{T}} = E^{\mathbb{T}}(t)$  and  $Z_C^{\mathbb{S}} = E_C^{\mathbb{S}}(s)$  to generate a new style-changing patch  $s'_2$ . We align different style domains by extracting the style embedding  $Z^{\mathbb{T}}$  of the target style domain  $\mathbb{T}$ , and re-feeding the content embeddings  $Z_C^{\mathbb{S}}$  from source domain  $\mathbb{S}$  with the disentangled target style embedding  $Z^{\mathbb{T}}$  to be reconstructed into the image conversion space, thereby alleviating the domain shift problem.

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

$$\begin{aligned} \mathcal{L}_{adv}^E = & \mathbb{E}_l \left[ \frac{1}{2} \log D_C \left( Z_C^{\mathbb{L}} \right) + \frac{1}{2} \log \left( 1 - D_C \left( Z_C^{\mathbb{L}} \right) \right) \right] \\ & + \mathbb{E}_s \left[ \frac{1}{2} \log D_C \left( Z_C^{\mathbb{S}} \right) + \frac{1}{2} \log \left( 1 - D_C \left( Z_C^{\mathbb{S}} \right) \right) \right] \\ & + \mathbb{E}_t \left[ \frac{1}{2} \log D_C \left( Z_C^{\mathbb{T}} \right) + \frac{1}{2} \log \left( 1 - D_C \left( Z_C^{\mathbb{T}} \right) \right) \right], \end{aligned} \quad (4)$$

where  $D_C$  denotes a feature-level content representation discriminator.

In addition to the adversarial loss on the embedding, we define an image-level adversarial loss  $\mathcal{L}_{adv}^I$  to constrain the generation of  $s'_1$ ,  $l'$ ,  $t'$  and  $s'_2$ , which can force the content-related, quality-related and target style-related encoders to capture their respective embedding.

$$\begin{aligned}\mathcal{L}_{adv}^I = & \mathbb{E}_s [\log D_I(s)] + \mathbb{E}_l [\log D_I(l)] + \mathbb{E}_t [\log D_I(t)] \\ & + \mathbb{E}_{\{s,l\}} [\log (1 - D_I(s'_1))] + \mathbb{E}_{\{s,t\}} [\log (1 - D_I(s'_2))] \quad (5) \\ & + \mathbb{E}_{\{l,s\}} [\log (1 - D_I(l'))] + \mathbb{E}_{\{t,s\}} [\log (1 - D_I(t'))],\end{aligned}$$

where  $D_I$  denotes an image-level domain discriminator.

Besides the adversarial loss for constraining the patch generation in the image conversion space, to supervise the adversarial patch 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 newly generated patches for image-level self-supervision in the original input image space and the image reconstruction space. Specifically, we reconstruct  $l'$  and  $s'_1$  as  $\hat{l}$  and  $\hat{s}_1$  in the original image space via the encoders  $\{E^L, E^C, E^S\}$  as well as the generators  $\{G^L, G^S\}$  according to the same procedure as Eq.3. Hence, the quality-related self-supervision loss  $\mathcal{L}_Q$  for supervision on the reconstructed patches can be defined as:

$$\begin{aligned}\mathcal{L}_Q = & \mathcal{L}_{rec}^L(l, \hat{l}) + \mathcal{L}_{rec}^S(s, \hat{s}_1) \quad (6) \\ = & \mathbb{E}_l [\|\hat{l} - l\|_1] + \mathbb{E}_s [\|\hat{s}_1 - s\|_1],\end{aligned}$$

where  $\|\cdot\|_1$  indicates the L<sub>1</sub> norm that is widely adopted in self-supervised learning for preserving the consistency of the reconstructed patches. Similarly, due to the lack of reference patches corresponding to the inputs, we reconstruct  $t'$  and  $s'_2$  into the patches  $\hat{t}$  and  $\hat{s}_2$ . An identical style-related self-supervised learning loss  $\mathcal{L}_T = \mathcal{L}_{rec}^T(t, \hat{t}) + \mathcal{L}_{rec}^S(s, \hat{s}_2)$  is proposed to align style while maintaining pathological and structural features.

In summary, through the self-supervised domain adaptation strategy, we can disentangle quality-related, target style-related and content-related embeddings, and further recombine content-related and style-related embeddings 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 Joint Training

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

$$\mathcal{L}_s = \mathcal{L}_Q + \mathcal{L}_T + \mathcal{L}_C, \quad (7)$$

where  $\mathcal{L}_C = \mathbb{E}_s [\|\hat{s}_1 - \hat{s}_2\|_1]$  is to constrain  $\hat{s}_1$  and  $\hat{s}_2$  to be similar, for establishing an intrinsic link between the quality-related and style-related disentanglement. Then the overall objective function  $\mathcal{L}$  of DASQE can be formulated as:

$$\mathcal{L} = \mathcal{L}_s + \lambda_1 \mathcal{L}_{adv}^E + \lambda_2 \mathcal{L}_{adv}^I, \quad (8)$$

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

## 4 EXPERIMENTS

In this section, we conduct experiments on four benchmarks including EyeQ [9], Messidor [8], DDR [25], and DRIVE [33] datasets 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, e.g. *high-/usable- and low-quality*.

**Messidor dataset** includes 1,200 fundus images from three medical institutions, and the DR grading annotations are provided to measure the severity of diabetic retinopathy.

**DDR dataset** includes 13,673 retinal fundus images for Diabetic Retinopathy diagnosis. Additionally, 775 images of the DDR are annotated with pixel-wise lesion segmentation masks.

**DRIVE dataset** contains 40 retinal fundus images with pixel-level vessel segmentation annotations.

**Evaluation Metrics.** We employ two types of metrics for DASQE and the comparable methods, e.g. full-reference and non-reference evaluation metrics. **For full-reference metrics**, peak signal-to-noise ratio (PSNR) and structural index measure (SSIM) [17] are adopted as the quantitative metrics to measure the performance of image enhancement. It should be noted that for the 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 [34] to synthesize the corresponding low-quality dataset. **For non-reference metrics**, we use fundus image quality assessment score (FIQA) [5], which indicates the ratio of the total number of images predicted as *high-quality* by a pre-trained quality assessment network [9] to the total number of the entire dataset. Furthermore, the typical non-reference metrics inception scores (IS) [1] and natural image quality evaluator (NIQE) [14, 47] are also utilized to assess the quality of image enhancement. Higher IS and lower NIQE metrics correspond to more satisfying quality enhancement results.

Additionally, in the appendix section, in order to investigate the influence of higher quality images enhanced by our method on DR grading, and lesion/vessel segmentation tasks, we also further train the DesnetNet-121 [18], UNet3+ [19], and CE-Net [11] on the messidor, DDR, and DRIVE datasets. The accuracy and quadratic weighted kappa [49] are chosen to evaluate the grading performance. The intersection over union (IoU) and Dice score (DSC)[21] are chosen to measure the segmentation 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 [29], and the weights are not shared across encoders. Due to the large-sized and diversity of the original images, all images are resized to  $512 \times 512$ , and the size of patch-wise fundus images is set to  $128 \times 128$ . 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 [15], distribution fitting [10], and latent structure-drive [16]. Deep learning-based approaches include: cofe-Net [34], I-SECRET [5], cGAN [20], CutGAN [32], CycleCAN [46], and StillGAN [31]. Experimental results are reported in Table 2 where the best results are boldfaced. The proposed method, DASQE, surpasses all the

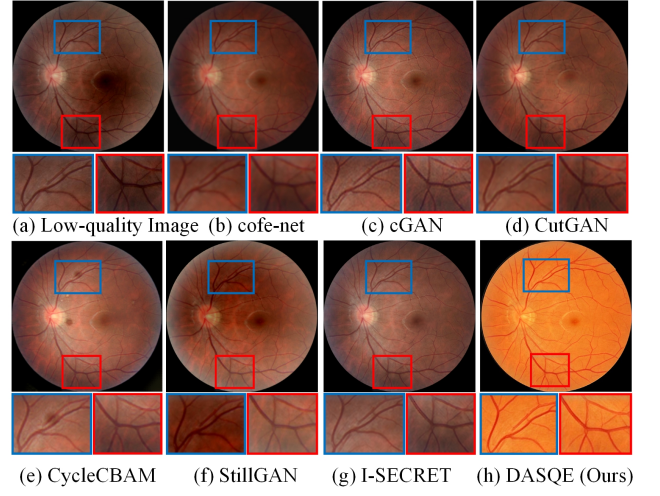
**Table 2: The comparison between our method with the SOTA methods for fundus image enhancement on EyeQ dataset.**

Methods	Types	Full-reference		Non-reference		
		PSNR $\uparrow$	SSIM $\uparrow$	FIQA $\uparrow$	IS $\uparrow$	NIQE $\downarrow$
LIME [15]	Traditional	13.54	0.868	0.346	1.28	6.63
Fu et al. [10]	Traditional	9.76	0.564	0.235	1.17	7.24
He et al. [16]	Traditional	15.56	0.759	0.368	1.33	6.52
cofe-net [34]	Supervised	20.51	0.885	0.482	1.54	5.31
cGAN [20]	Supervised	26.35	0.894	0.634	1.63	4.73
I-SECRET [5]	Semi-supervised	27.36	0.908	0.664	1.71	4.56
CutGAN [32]	Unsupervised	22.76	0.872	0.576	1.62	5.14
Cycle-CBAM [46]	Unsupervised	21.56	0.843	0.534	1.52	5.26
StillGAN [31]	Unsupervised	25.38	0.896	0.619	1.67	4.84
DASQE (Ours)	Unsupervised	<b>28.32</b>	<b>0.915</b>	<b>0.683</b>	<b>1.76</b>	<b>4.13</b>

compared methods significantly, in terms of full-reference and non-reference metrics without 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 Figure 5. 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 3. 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 feature disentanglement for image quality enhancement. During the



**Figure 5: Visual comparisons on the image quality enhancement between DASQE and other deep learning methods.**

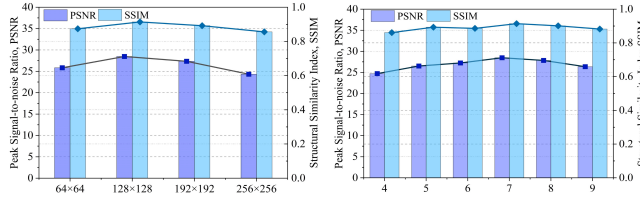
**Table 3: The comparison between our method with the SOTA methods for image quality enhancement on the Messidor.**

Methods	Full-reference		Non-reference		
	PSNR $\uparrow$	SSIM $\uparrow$	FIQA $\uparrow$	IS $\uparrow$	NIQE $\downarrow$
LIME [15]	12.36	0.852	0.338	1.25	6.82
Fu et al. [10]	11.32	0.573	0.263	1.21	6.87
He et al. [16]	16.23	0.764	0.374	1.42	6.14
cofe-net [34]	21.97	0.852	0.554	1.56	5.24
cGAN [20]	25.83	0.867	0.625	1.68	4.81
I-SECRET [5]	26.73	0.874	0.658	1.70	4.62
CutGAN [32]	24.37	0.883	0.617	1.64	4.93
Cycle-CBAM [46]	23.56	0.851	0.583	1.62	5.18
StillGAN [31]	26.32	0.871	0.649	1.72	4.75
DASQE (Ours)	<b>29.14</b>	<b>0.892</b>	<b>0.694</b>	<b>1.78</b>	<b>3.94</b>

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 and considering the sizes of lesions and artifacts, we report the quality enhancement performance at different patch sizes in Figure 6-left. It can be found that the best performance metrics for quality enhancement are achieved when the patch size is  $128 \times 128$ . Furthermore, both oversized and undersized patches result in degraded performance of image enhancement. The reason is that the number of high-/low-quality patches is affected by the size of the patches. When the size of patches changes from  $128 \times 128$  to  $64 \times 64$ , the number of high-quality patch-wise images increases. This may lead to more incorrect quality assessment results for the rule-based quality assessment framework. In contrast, when the size of patches changes

from  $128 \times 128$  to  $256 \times 256$ , the performance of DASQE degrades with the reduction in the number of high-quality patches.

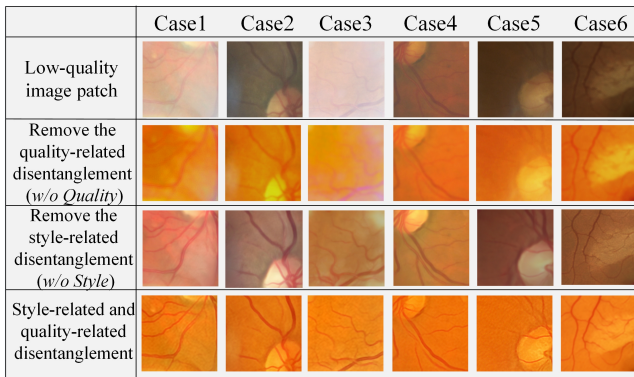


**Figure 6: Comparison of low-quality fundus image enhancement results based on the different patch sizes (left) and different number  $K$  of style cluster centers (right).**

In addition to the patch size, we also explore the intrinsic relationship between the number of style cluster centers and low-quality image enhancement as shown in Fig 6-right. We observe that the proposed method achieves the best performance when the number of cluster centers is set to 7. More specifically, 1) when the number of the style cluster centers is set from [4, 6], the clustering happens to overlap, which leads to inconsistent styles of patches in the target style domain. 2) when the number of style cluster centers exceeds 7, there are prominent outlier points in the clustering results, which increases the complexity of optimizing for style disentanglement.

**Table 4: Quantitative results for image enhancement on EyeQ.**

Methods	Disentanglement		Full-reference		Non-reference		
	Quality	Style	PSNR $\uparrow$	SSIM $\uparrow$	FIQA $\uparrow$	IS $\uparrow$	NIQE $\downarrow$
DASQE	✗	✓	16.32	0.768	0.372	1.37	6.45
DASQE	✓	✗	19.24	0.832	0.475	1.51	5.42
DASQE	✓	✓	<b>28.32</b>	<b>0.915</b>	<b>0.683</b>	<b>1.76</b>	<b>4.13</b>



**Figure 7: The effect of the different disentanglement branches for the patch-wise image quality enhancement.**

To investigate the effectiveness of multiple feature disengagement for image quality enhancement, we compare the proposed method with its two variants, respectively. Quantitative results are shown in Table 4. When the quality disengagement 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 disengagement branch and keep the quality disengagement branch, the PSNR also drops to 19.24 on EyeQ dataset. Our results suggest that the simultaneously disengagement of both quality and style features is critical for improving the quality of the fundus image. To demonstrate the effectiveness of the different disengagement branches more intuitively, we also further visualize some patch-wise fundus images in Figure 7.

As shown in the 2nd row in Figure 7, 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 noticeable blurring of patches is also not eliminated. From the 3rd row in Figure 7, we can observe that although *w/o Style* is able to transform general knowledge from low-quality patch domains to high-quality patch domains, there exist a large style gap for the enhanced patch-wise images, which is also detrimental to the quality enhancement of fundus images.

## 5 CONCLUSION

In this work, we propose a reference-free domain adaptation method to extend a self-supervised learning framework to the quality enhancement task. Our approach is the first to address the medical fundus image quality enhancement problem from a complete self-supervised learning perspective. 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 fundus 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 fundus 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), the Science Project of Liaoning Province (2021-MS-105) and the 111 Project (B16009).



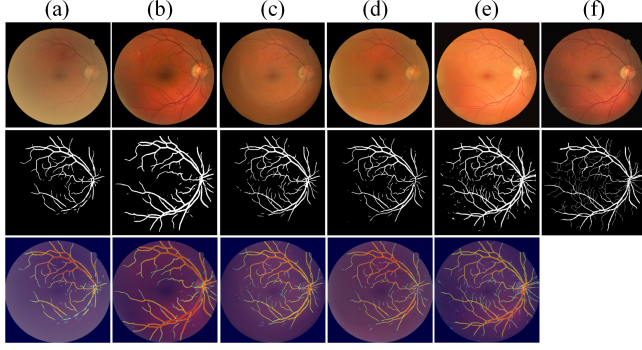
## REFERENCES

- [1] Shane Barratt and Rishi Sharma. 2018. A note on the inception score. *arXiv preprint arXiv:1801.01973* (2018).
- [2] Peng Cao, Qingshan Hou, Ruoxian Song, Haonan Wang, and Osmar Zaiane. 2022. Collaborative learning of weakly-supervised domain adaptation for diabetic retinopathy grading on retinal images. *Computers in Biology and Medicine* 144 (2022), 105341.
- [3] Chenghao Chen and Hao Li. 2021. Robust representation learning with feedback for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 7742–7751.
- [4] Xiang Chen, Zhentao Fan, Pengpeng Li, Longgang Dai, Caihua Kong, Zhuoran Zheng, Yufeng Huang, and Yufeng Li. 2022. Unpaired Deep Image Dehazing Using Contrastive Disentanglement Learning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII*. Springer, 632–648.
- [5] Pujin Cheng, Li Lin, Yijin Huang, Junyan Lyu, and Xiaoying Tang. 2021. I-secret: Importance-guided fundus image enhancement via semi-supervised contrastive constraining. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 87–96.
- [6] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. 2021. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4641–4650.
- [7] Ling Dai, Liang Wu, Huating Li, Chun Cai, Qiang Wu, Hongyu Kong, Ruhan Liu, Xiangning Wang, Xuhong Hou, Yuexing Liu, et al. 2021. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature communications* 12, 1 (2021), 1–11.
- [8] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, et al. 2014. Feedback on a publicly distributed image database: the Messidor database. *Image Analysis & Stereology* 33, 3 (2014), 231–234.
- [9] Huazhu Fu, Boyang Wang, Jianbing Shen, Shanshan Cui, Yanwu Xu, Jiang Liu, and Ling Shao. 2019. Evaluation of retinal image quality assessment networks in different color-spaces. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 48–56.
- [10] Xueyang Fu, Peixian Zhuang, Yue Huang, Yinghao Liao, Xiao-Ping Zhang, and Xinghao Ding. 2014. A retinex-based enhancing approach for single underwater image. In *2014 IEEE international conference on image processing (ICIP)*. IEEE, 4572–4576.
- [11] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. 2019. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging* 38, 10 (2019), 2281–2292.
- [12] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. 2021. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering* (2021).
- [13] Chunle Guo, Chongyi Li, Jichang Guo, Chen Chang Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1780–1789.
- [14] Jifeng Guo, Zhiqi Pang, Fan Yang, Jiayou Shen, and Jian Zhang. 2020. Study on the method of fundus image generation based on improved GAN. *Mathematical Problems in Engineering* 2020 (2020), 1–13.
- [15] Xiaojie Guo, Yu Li, and Haibin Ling. 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing* 26, 2 (2016), 982–993.
- [16] Kaiming He, Jian Sun, and Xiaoou Tang. 2010. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence* 33, 12 (2010), 2341–2353.
- [17] Alain Hore and Djemel Ziou. 2010. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition*. IEEE, 2366–2369.
- [18] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.
- [19] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. 2020. 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)*. IEEE, 1055–1059.
- [20] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.
- [21] Shruti Jadon. 2020. A survey of loss functions for semantic segmentation. In *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*. IEEE, 1–7.
- [22] Qiangguo Jin, Zhaopeng Meng, Tuan D Pham, Qi Chen, Leyi Wei, and Ran Su. 2019. DUNet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems* 178 (2019), 149–162.
- [23] Guisik Kim and Junseok Kwon. 2022. Self-Parameter Distillation Dehazing. *IEEE Transactions on Image Processing* (2022).
- [24] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. 2019. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF international conference on computer vision*. 8878–8887.
- [25] Tao Li, Yingqi Gao, Kai Wang, Song Guo, Hanruo Liu, and Hong Kang. 2019. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Information Sciences* 501 (2019), 511–522.
- [26] Xiaomeng Li, Xiaowei Hu, Lequan Yu, Lei Zhu, Chi-Wing Fu, and Pheng-Ann Heng. 2019. CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. *IEEE transactions on medical imaging* 39, 5 (2019), 1483–1493.
- [27] Ke Liang, Yue Liu, Sihang Zhou, Wenxuan Tu, Yi Wen, Xihong Yang, Xiangjun Dong, and Xinwang Liu. 2023. Knowledge Graph Contrastive Learning Based on Relation-Symmetrical Structure. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [28] Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, Xinwang Liu, and Fuchun Sun. 2022. Reasoning over Different Types of Knowledge Graphs: Static, Temporal and Multi-Modal. *arXiv preprint arXiv:2212.05767* (2022).
- [29] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. 2017. Unsupervised image-to-image translation networks. *Advances in neural information processing systems* 30 (2017).
- [30] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. 2022. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5637–5646.
- [31] Yuhui Ma, Jiang Liu, Yonghuai Liu, Huazhu Fu, Yan Hu, Jun Cheng, Hong Qi, Yufei Wu, Jiong Zhang, and Yitian Zhao. 2021. Structure and illumination constrained GAN for medical image enhancement. *IEEE Transactions on Medical Imaging* 40, 12 (2021), 3955–3967.
- [32] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. 2020. Contrastive learning for unpaired image-to-image translation. In *European conference on computer vision*. Springer, 319–345.
- [33] Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabudhe, and Fabrice Meriaudeau. 2018. Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. *Data* 3, 3 (2018), 25.
- [34] Ziyi Shen, Huazhu Fu, Jianbing Shen, and Ling Shao. 2020. Modeling and enhancing low-quality retinal fundus images. *IEEE transactions on medical imaging* 40, 3 (2020), 996–1006.
- [35] Saikat Kumar Shome and Siva Ram Krishna Vadali. 2011. Enhancement of diabetic retinopathy imagery using contrast limited adaptive histogram equalization. *International Journal of Computer Science and Information Technologies* 2, 6 (2011), 2694–2699.
- [36] Zhen Ling Teo, Yih-Chung Tham, Marco Yu, Miao Li Chee, Tyler Hyungtaek Rim, Ning Cheung, Mukharram M Bikbov, Ya Xing Wang, Yating Tang, Yi Lu, et al. 2021. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology* 128, 11 (2021), 1580–1591.
- [37] Qi-Chong Tian and Laurent D Cohen. 2017. Global and local contrast adaptive enhancement for non-uniform illumination color images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 3023–3030.
- [38] Xinhang Wan, Jiyuan Liu, Weixuan Liang, Xinwang Liu, Yi Wen, and En Zhu. 2022. Continual Multi-View Clustering. In *Proceedings of the 30th ACM International Conference on Multimedia (Lisboa, Portugal) (MM '22)*. Association for Computing Machinery, New York, NY, USA, 3676–3684. <https://doi.org/10.1145/3503161.3547864>
- [39] Xinhang Wan, Xinwang Liu, Jiyuan Liu, Siwei Wang, Yi Wen, Weixuan Liang, En Zhu, Zhe Liu, and Lu Zhou. 2023. Auto-weighted Multi-view Clustering for Large-scale Data. *arXiv:2303.01983 [cs.LG]*
- [40] Renzhen Wang, Benzhi Chen, Deyu Meng, and Lisheng Wang. 2018. Weakly supervised lesion detection from fundus images. *IEEE transactions on medical imaging* 38, 6 (2018), 1501–1512.
- [41] Xiaofei Wang, Mai Xu, Jicong Zhang, Lai Jiang, Liu Li, Mengxian He, Ningli Wang, Hanruo Liu, and Zulin Wang. 2021. Joint Learning of Multi-Level Tasks for Diabetic Retinopathy Grading on Low-Resolution Fundus Images. *IEEE Journal of Biomedical and Health Informatics* 26, 5 (2021), 2216–2227.
- [42] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. 2022. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 2604–2612.
- [43] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. 2022. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5901–5910.
- [44] Xihong Yang, Xiaochang Hu, Sihang Zhou, Xinwang Liu, and En Zhu. 2022. Interpolation-based contrastive learning for few-label semi-supervised learning. *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- [45] Xihong Yang, Yue Liu, Sihang Zhou, Siwei Wang, Wenxuan Tu, Qun Zheng, Xinwang Liu, Liming Fang, and En Zhu. 2023. Cluster-guided Contrastive Graph Clustering Network. In *Proceedings of the AAAI conference on artificial intelligence*,

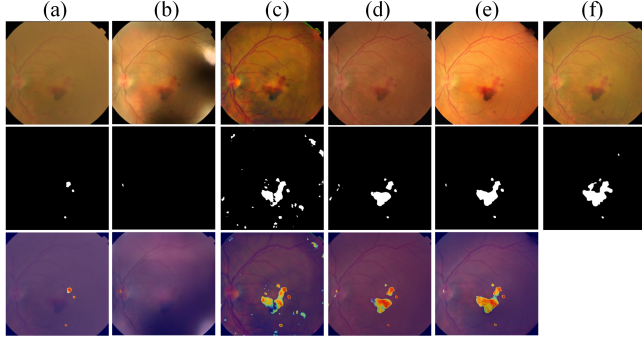
- Vol. 37. 10834–10842.
- [46] Qijing You, Cheng Wan, Jing Sun, Jianxin Shen, Hui Ye, and Qiuli Yu. 2019. Fundus image enhancement method based on CycleGAN. In *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 4500–4503.
- [47] He Zhao, Bingyu Yang, Lvchen Cao, and Huiqi Li. 2019. Data-driven enhancement of blurry retinal images via generative adversarial networks. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I* 22. Springer, 75–83.
- [48] Mei Zhou, Kai Jin, Shaoze Wang, Juan Ye, and Dahong Qian. 2017. Color retinal image enhancement based on luminosity and contrast adjustment. *IEEE Transactions on Biomedical engineering* 65, 3 (2017), 521–527.
- [49] Yi Zhou, Boyang Wang, Lei Huang, Shanshan Cui, and Ling Shao. 2020. A benchmark for studying diabetic retinopathy: segmentation, grading, and transferability. *IEEE Transactions on Medical Imaging* 40, 3 (2020), 818–828.
- [50] Yuqian Zhou, Hanchao Yu, and Humphrey Shi. 2021. Study group learning: Improving retinal vessel segmentation trained with noisy labels. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I* 24. Springer, 57–67.

## A APPENDICES

### A.1 Clinical Image Applications and Analysis



**Figure 8: Visualization of vessel segmentation (row 2) and heatmap (row 3) on low-quality images and enhanced images. (a) Low-quality image, (b) coe-net, (c) StillGAN, (d) I-SECRET, (e) DASQE, and (f) Ground Truth.**



**Figure 9: Visualization of lesion (HEs) segmentation (row 2) and heatmap (row 3) on low-quality images and enhanced images. (a) Low-quality image, (b) coe-net, (c) StillGAN, (d) I-SECRET, (e) DASQE, and (f) Ground Truth.**

**Table 5: DR grading, lesion/vessel segmentation evaluation.**

Methods	Messidor		DDR		DRIVE	
	DR Grading		Lesion Seg.		Vessel Seg.	
	Acc	Kappa	IoU	DSC	IoU	DSC
Low-quality image	0.681	0.657	0.413	0.539	0.483	0.678
cofe-net [34]	0.704	0.687	0.486	0.605	0.547	0.741
cGAN [20]	0.731	0.712	0.519	0.652	0.568	0.762
I-SECRET [5]	0.735	0.718	0.512	0.648	0.573	0.764
CutGAN [32]	0.717	0.694	0.497	0.625	0.558	0.746
Cycle-CBAM [46]	0.701	0.682	0.493	0.617	0.553	0.747
StillGAN [31]	0.724	0.706	0.504	0.634	0.562	0.758
DASQE (Ours)	<b>0.748</b>	<b>0.736</b>	<b>0.537</b>	<b>0.664</b>	<b>0.586</b>	<b>0.782</b>

To investigate the effects of the DASQE on downstream medical image analysis tasks, we present a series of visual comparison results in Figure 8 & 9 and Table 5, including vessel/lesion segmentation and DR grading tasks.

**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 pre-trained on DRIVE dataset for vessel segmentation on the enhanced images of proposed and comparable methods. As shown in Figure 8, 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. Moreover, we also report quantitative vessel segmentation results in Table 5.

**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+ pre-trained on the DDR dataset for HEs segmentation on enhanced images. The segmentation results of HEs are shown in Figure 9 and Table 5. 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 enhanced quality images by competing methods using a DenseNet-121 as the grading model. The DR grading results are shown in Table 5. We observe that our method is the most beneficial for the disease grading of the classification model.