Structure and Illumination Constrained GAN for Medical Image Enhancement

Yuhui Ma, Jiang Liu, Yonghuai Liu, Huazhu Fu, Yan Hu, Jun Cheng, Hong Qi, Yufei Wu, Jiong Zhang, Yitian Zhao

***Abstract*— The development of medical imaging techniques has greatly supported clinical decision making. However, poor imaging quality, such as non-uniform illumination or imbalanced intensity, brings challenges for automated screening, analysis and diagnosis of diseases. Previously, bi-directional GANs (e.g., CycleGAN), have been proposed to improve the quality of input images without the requirement of paired images. However, these methods focus on global appearance, without imposing constraints on structure or illumination, which are essential features for medical image interpretation. In this paper, we propose a novel and versatile bi-directional GAN, named Structure and illumination constrained GAN (StillGAN), for medical image quality enhancement. Our StillGAN treats low- and high-quality images as two distinct domains, and introduces local structure and illumination constraints for learning both overall characteristics and local details. Extensive experiments on three medical image datasets (e.g., corneal confocal microscopy, retinal color fundus and endoscopy images) demonstrate that our method performs better than both conventional methods and other deep learning-based methods. In addition, we have investigated the impact of the proposed method on different medical image analysis and clinical tasks such as nerve segmentation, tortuosity grading, fovea localization and disease classification.**

***Index Terms*— Bi-directional GAN, Illumination regularization, structure loss, medical image enhancement.**

I. Introduction

Recently, the rapid development of medical imaging technology has brought about a revolution in the field of clinical medicine [1]. Medical images usually provide clinicians with a great deal of information related to biological or anatomical tissues; this plays a crucial role in effective diagnosis and treatment. However, whether acquired by the same or different devices, medical images tend to exhibit large variations in quality - exhibiting defects such as intensity inhomogeneity, low contrast, noticeable blur or noise, all of which can occur during the image acquisition process. Fig. 1 illustrates one low-quality and one high-quality examples, captured by confocal microscopy, color fundus cameras and endoscopy respectively. For the high-quality examples (the bottom row of Fig. 1), almost all details can be easily identified by clinicians. For the low-quality images (the top row of Fig. 1), however, it is difficult to observe with clarity the complete structure of corneal nerve fibers, blood vessels, digestive tract or other tissues and lesions of interest. By contrast with natural or scenery images, most medical images result from a specialized imaging process with unique degradation factors, which may lead to a variety of low-quality appearance artifacts and additional challenges to clinical applications. A screening study by Philip et al. [2] demonstrated that about 12% of fundus images from 5,575 consecutive patients were unreadable by ophthalmologists due to lack of adequate quality. Another study based on UK BioBank also showed that about 30% of retinal images were not of sufficiently high quality for accurate diagnosis [3]. In addition, these obstacles also impair the performance of many subsequent image analysis tasks, such as specific structure segmentation [4] and lesion detection [5], or other computer-aided diagnosis [6]. Consequently, fully automatic and reliable medical image enhancement techniques

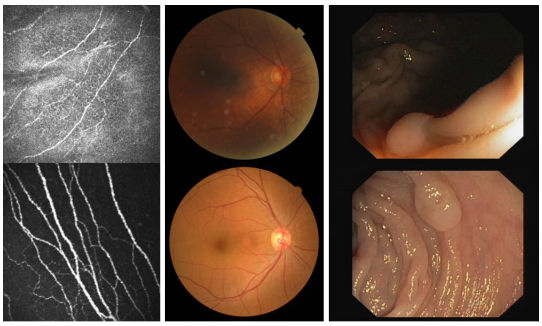

Fig. 1: Examples of different low-quality (top row) and high-quality (bottom row) medical images.

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China (LZ19F010001Y), in part by the Youth Innovation Promotion Association CAS (2021298), in part by the Key Research and Development Program of Zhejiang Province (2020C03036), in part by the Ningbo 2025 S&T Megaprojects (2019B10033 and 2019B1006). *Y. Ma and J. Liu are contributed equally to this work. (Corresponding author: Yitian Zhao, e-mail: yitian.zhao@nimte.ac.cn)*

Y. Ma, J. Zhang, J. Cheng and Y. Zhao are with Cixi Institute of Biomedical Engineering, Ningbo Institute of Material Technology and Engineering, Chinese Academy of Sciences, Ningbo, China. They are also with Zhejiang International Scientific and Technological Cooperative Base of Biomedical Materials and Technology, and Zhejiang Engineering Research Center for Biomedical Materials. Y. Ma is also with the University of Chinese Academy of Sciences, Beijing 100049, China. J. Liu and Y. Hu are with Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China; Y. Liu is with Department of Computer Science, Edge Hill University, UK; H. Fu is with the Inception Institute of Artificial Intelligence, UAE; H. Qi are with Department of Ophthalmology, Peking University Third Hospital, Beijing, China; Y. Wu is with Department of Ophthalmology, the Affiliated People's Hospital of Ningbo University, Ningbo, China;

have long been deemed worthwhile as the preceding step of clinical applications, as they are crucial for achieving high-quality images with comprehensive details and adequate contrast.

In recent decades, many conventional methods have been proposed for image enhancement. These include histogram equalization (HE) [7], dark channel prior (DCP) [8], filtering-based [9], [10] and Retinex-based methods [11]–[13]. However, they are usually sensitive to a few parameters [14], which are not sufficiently adaptive and usually require manual adjustment. Recently, due to the increase in the amount of data and the availability of computing capabilities, deep learning techniques have also revealed their superiority in low-level image processing and computer vision tasks, where image enhancement can be treated as a task of image-to-image translation. The most common deep learning-based methods are fully supervised learning methods [15], [16], which require aligned image pairs in the training phase. For medical images, it is however difficult to obtain such low/high-quality image pairs in real scenarios for training. Therefore, a few unsupervised learning frameworks have also been proposed [14], [17]–[20], but they are usually unstable, and sometimes amplify noise, or suffer from halo artefacts.

As a popular unpaired learning architecture of image-to-image translation, a Cycle-consistent Generative Adversarial Network (CycleGAN) [18] has the advantage of learning knowledge represented with typical images in one domain, and transferring it to the other domain, without the need for aligned image pairs. However, most existing bi-directional GANs are usually under-constrained. For example, CycleGAN focuses primarily on learning intra-domain global appearance and inter-domain cycle-consistency, and is thus often ineffective in capturing local details. In medical images, local details are particularly important for decision-making. A high-quality medical image usually should exhibit uniform illumination and clear structural details.

Taking all of the above into consideration, we propose a novel framework for medical image enhancement, called Structure and illumination constrained GAN (StillGAN). To this end, we develop two novel constraints - illumination regularization and structure loss, and incorporate them into the objective function of a bi-directional GAN, in order to obtain images with better illumination condition and structural details for clinical interpretation and subsequent analysis. Illumination regularization aims at improving illumination uniformity via minimizing the difference of illumination distribution in the enhanced images, while structure loss is introduced to preserve structural details as much as possible by reducing the dissimilarity in terms of structure between the low-quality image and its enhanced version. Compared with other deep learning approaches, the proposed StillGAN achieves overall better performance in various metrics for enhancing multi-modality images. The proposed method extends considerably our previous work published in MICCAI-2020 [21], which was verified only on two medical imaging modalities. In this work, medical image enhancement is regarded as a transformation task from a low-quality image domain to a high-quality image domain. Our contributions are summarized as follows:

- A novel bi-directional GAN called StillGAN has been proposed to improve the readability of poor quality medical images. The new model introduces illumination regularization and structure loss to improve illumination conditions and to preserve structural details, respectively.
- The proposed method has undergone rigorous quantitative and qualitative evaluation using three different medical image modalities - confocal microscopy, color fundus and endoscopy images. For each medical image modality, we adopt different image quality assessment approaches, according to their respective imaging characteristics and clinical interests. We have released the source code of our StillGAN and the corneal confocal microscopy dataset, CORN-2 [21] (containing both low- and high-quality image sets) online available to the public at https://imed.nimte.ac.cn/CORN.html

II. Related Works

Many methods have been proposed for a variety of image enhancement tasks. Examples of well-known global enhancement methods include histogram equalization (HE) [7] and contrast limited adaptive histogram equalization (CLAHE) [22]. They enhance images by stretching their dynamic ranges, and have been widely used in medical imaging community. Recently, some methods [23], [24] have produced high-quality results by applying dehazing methods [8], [25] to the inverted low-quality images. In addition, some filtering-based methods [9], [10] and Retinex-based methods [11]–[13] have been proposed to improve image quality either by filtering, or by decomposing the given image into illumination and reflectance components. However, these methods usually process foreground and background indiscriminately, and as a result sometimes amplify noise, or oversmooth regions close to flat, and in consequence struggle to preserve fine details.

In recent years, deep learning approaches have been widely used in computer vision, which has also enabled the rapid advancement of image enhancement. Most deep learning approaches are fully supervised, which attempt to learn a mapping between a low-quality image and its reference high-quality one. Lore *et al.* [15] utilized synthetically darkened and noise-added images to train a deep stacked-sparse denoising autoencoder, aiming at achieving both low-light enhancement and denoising. Tai *et al.* [26] proposed a persistent memory network, MemNet, for image restoration. Interestingly, a few works have also appeared that combine deep networks with Retinex theory. Inspired by multi-scale Retinex, Shen *et al.* [16] designed MSR-Net for low-light image enhancement. Wei *et al.* [27] proposed a two-stage framework, Retinex-Net, for low-light image enhancement.

Although achieving impressive results in image enhancement, fully-supervised learning methods have shortcomings. These methods require rigorously aligned low/high-quality image pairs for training, and their performance depends largely on the quality of the training set. For medical images in particular, such image pairs are usually not available, and synthetic image pairs cannot fully characterize low- and high-quality images in clinical scenarios: this is likely to lead to unexpected visual results such as color shift and intensive

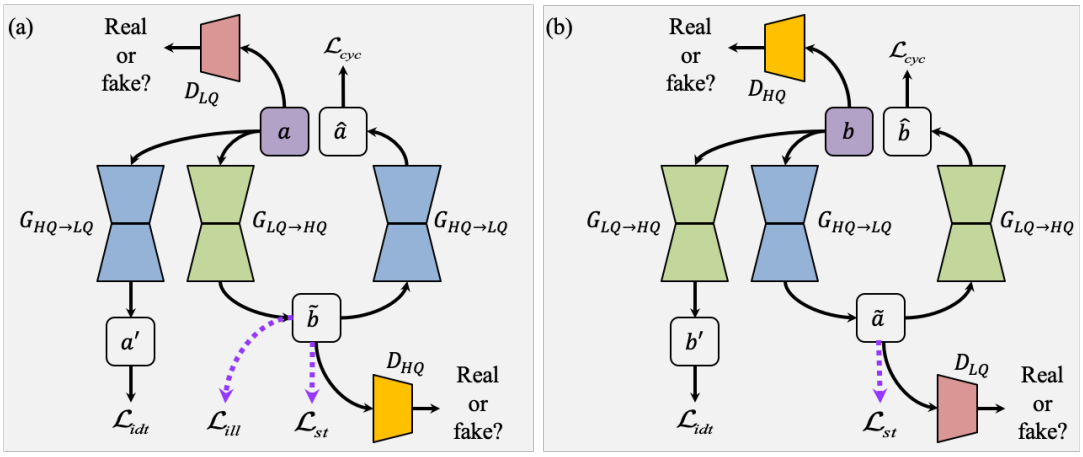

Fig. 2: The overall structure diagram of StillGAN. It comprises two generator (G)/discriminator (D) pairs ($G_{LQ\to HQ}/D_{HQ}$, $G_{HQ\to LQ}/D_{LQ}$) and two types of cycle consistency: (a) forward cycle consistency; (b) backward cycle consistency. Variables a and b represent real images from the low-quality (LQ) and high-quality image domain (HQ) respectively. Other variables are defined as follows: $\tilde{b} = G_{LQ\to HQ}(a), \hat{a} = G_{HQ\to LQ}(\tilde{b}), a' = G_{HQ\to LQ}(a); \tilde{a} = G_{HQ\to LQ}(b), \hat{b} = G_{LQ\to HQ}(\tilde{a}), b' = G_{LQ\to HQ}(b)$. $\mathcal{L}_{cyc}$ and $\mathcal{L}_{idt}$ represent the cycle consistency term and the identity mapping loss. The proposed illumination regularization and structure loss are represented as $\mathcal{L}_{ill}$ and $\mathcal{L}_{st}$, respectively.

noise. In consequence, unsupervised learning models like CycleGAN [18] were proposed recently. Most of these models attempt to learn knowledge represented with typical images in one domain and transfer it to the other without the requirement of paired images. Gatys *et al.* [17] proposed a neural transfer algorithm (NST) for unpaired image transformation. Zhang *et al.* [19] introduced a multi-style generative network (MSG-Net) to achieve real-time image style translation. Chen *et al.* [14] proposed a two-way GAN with several improvements for photograph enhancement. By contrast, Jiang *et al.* [20] proposed EnlightenGAN, a one-way GAN with a global-local discriminator structure, a self-regularized perceptual loss fusion and an attention mechanism, for low-light image enhancement. Nevertheless, compared with supervised learning, it is difficult for these unsupervised learning methods to precisely learn characterization of one domain and produce stable results in the other (i.e., amplifying noise or generating halo artefacts).

III. Proposed Method

In our work, we treat the medical image enhancement as a translation of general knowledge from low-quality (LQ) domain to high-quality (HQ) domain. Then we propose a novel unpaired learning framework, StillGAN, for medical image enhancement. It learns a suitable mapping from domain LQ to domain HQ without requiring paired images in the training phase, as shown in Fig. 2. StillGAN also introduces two new loss terms - illumination regularization and structure loss, which aim at achieving illumination uniformity and restoring structural details in the enhanced images.

A. Network Architecture

Our StillGAN adopts two generator/discriminator pairs ($G_{LQ\to HQ}/D_{HQ}$, $G_{HQ\to LQ}/D_{LQ}$), where $G_{LQ\to HQ}$ ($G_{HQ\to LQ}$) learns to translate an image from domain LQ (HQ) into domain HQ (LQ), and D_{LQ} (D_{HQ}) is trained to distinguish between real samples from domain LQ (HQ) and the generated images from domain HQ (LQ). Generators $G_{LQ\to HQ}$ and $G_{HQ\to LQ}$ adopt an encoder-decoder architecture, with residual blocks similar to ResU-Net [28]. The constructed generative network consists of eight encoder layers and the symmetric decoder layers with skip connections, as shown in Fig. 3. For each encoder layer, we use a residual block followed by a max pooling layer; while for each decoder layer, we adopt an upsampling layer using bilinear interpolation followed by a residual block with the same structure. The residual block takes the form of two stacked 3×3 Convolution-BatchNorm-LeakyReLU with shortcut connection between the input and output.

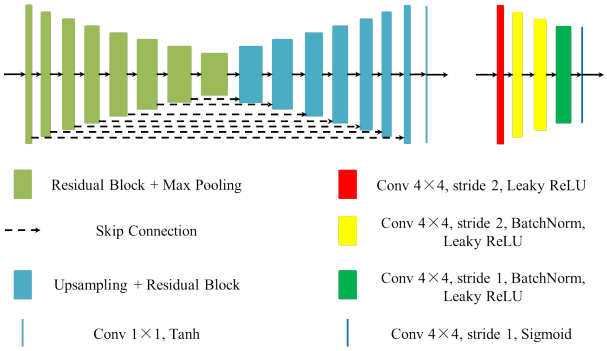

Fig. 3: The network structure of the generators (left) and discriminators (right). The generators adopt an encoder-decoder architecture with skip connections and several residual blocks, and the discriminators utilize PatchGAN [29] with five convolutional layers.

For both discriminators D_{LQ} and D_{HQ}, we utilize PatchGAN [29] for the classification of an image as real or fake based on image patches rather than the whole image:

this differentiates PatchGAN from traditional discriminators. PatchGAN contains five 4×4 convolutional layers, with a stride of 2 in the first three layers and a stride of 1 in the last two layers, as illustrated on the right hand side of Fig. 3. Leaky ReLU with a slope of 0.2 is applied in the first four layers. Batch normalization is applied in the middle three layers. For the above settings, we set the receptive field of PatchGAN, or the patch of the input image to be identified as 70×70, which makes PatchGAN more lightweight and faster than traditional discriminators, but still guides the generator to produce realistic results [30]. Finally, the Sigmoid activation function is adopted in the output layer to identify each patch of the input image. In consequence, each output pixel represents the probability that the corresponding 70×70 patch of the input image is from one real sample.

B. Objective Function

As a kind of bi-directional GAN framework, the basic objective function of StillGAN contains three terms, including adversarial loss, cycle consistency loss and identity mapping loss. In addition, StillGAN introduces two novel terms, illumination regularization and structure loss, to further constrain the bi-directional GAN framework in order to achieve illumination uniformity and preserve subtle structural details for medical image enhancement.

• **Transfer Loss** Transfer loss is defined as the basic objective function of StillGAN, represented by the summation of adversarial loss, cycle consistency loss and identity mapping loss. In StillGAN, the adversarial loss $\mathcal{L}_{adv}$ is applied to both the generator/discriminator pairs ($G_{LQ\to HQ}/D_{HQ}$, $G_{HQ\to LQ}/D_{LQ}$). It is defined as:

$$\begin{aligned}&\mathcal{L}_{adv}\left(G_{LQ\to HQ}, G_{HQ\to LQ}, D_{LQ}, D_{HQ}\right)\\&= \mathbb{E}_{b\in HQ}\left[logD_{HQ}(b)\right] + \mathbb{E}_{a\in LQ}\left[log\left(1 - D_{HQ}(G_{LQ\to HQ}(a))\right)\right]\\&+ \mathbb{E}_{a\in LQ}\left[logD_{LQ}(a)\right] + \mathbb{E}_{b\in HQ}\left[log\left(1 - D_{LQ}(G_{HQ\to LQ}(b))\right)\right],\end{aligned} \tag{1}$$

where $G_{LQ\to HQ}$ ($G_{HQ\to LQ}$) attempts to convert an image from domain *LQ* (*HQ*) into domain *HQ* (*LQ*), and D_{LQ} (D_{HQ}) tries to identify differences between real samples from domain *LQ* (*HQ*) and the generated images from domain *HQ* (*LQ*).

In order to achieve interconversion and reconstruction between the two domains via two generators, StillGAN contains both the forward and backward cycle consistency, as shown in Fig. 2. For the forward cycle consistency, each $a \in LQ$ is expected to be recovered as well as possible, which is denoted as $a \to \tilde{b} = G_{LQ\to HQ}(a) \to \hat{a} = G_{HQ\to LQ}(\tilde{b}) \approx a$. This holds for the backward cycle consistency as well: $b \to \tilde{a} = G_{HQ\to LQ}(b) \to \hat{b} = G_{LQ\to HQ}(\tilde{a}) \approx b$. Thus the cycle consistency loss $\mathcal{L}_{cyc}$ is defined as:

$$\begin{aligned}&\mathcal{L}_{cyc}\left(G_{LQ\to HQ}, G_{HQ\to LQ}\right)\\&= \mathbb{E}_{a\in LQ}\left[\|G_{HQ\to LQ}(G_{LQ\to HQ}(a)) - a\|_1\right]\\&+ \mathbb{E}_{b\in HQ}\left[\|G_{LQ\to HQ}(G_{HQ\to LQ}(b)) - b\|_1\right].\end{aligned} \tag{2}$$

In addition, two generators are regularized as an identity mapping separately when real samples from *LQ* (*HQ*) are applied to $G_{HQ\to LQ}$ ($G_{LQ\to HQ}$): $a' = G_{HQ\to LQ}(a) \approx a$ and $b' = G_{LQ\to HQ}(b) \approx b$. The identity mapping loss $\mathcal{L}_{idt}$ is thus defined as:

$$\begin{aligned}\mathcal{L}_{idt}\left(G_{LQ\to HQ}, G_{HQ\to LQ}\right) &= \mathbb{E}_{b\in HQ}\left[\|G_{LQ\to HQ}(b) - b\|_1\right]\\&+ \mathbb{E}_{a\in LQ}\left[\|G_{HQ\to LQ}(a) - a\|_1\right].\end{aligned} \tag{3}$$

Therefore, the transfer loss is finally defined as:

$$\begin{aligned}&\mathcal{L}_{transfer}\left(G_{LQ\to HQ}, G_{HQ\to LQ}, D_{LQ}, D_{HQ}\right)\\&= \mathcal{L}_{adv}\left(G_{LQ\to HQ}, G_{HQ\to LQ}, D_{LQ}, D_{HQ}\right)\\&+ \lambda_1\mathcal{L}_{cyc}\left(G_{LQ\to HQ}, G_{HQ\to LQ}\right) + \lambda_2\mathcal{L}_{idt}\left(G_{LQ\to HQ}, G_{HQ\to LQ}\right),\end{aligned} \tag{4}$$

where parameters λ_1 and λ_2 represent positive weighted coefficients of the cycle consistency loss and the identity mapping loss, respectively.

Although it is possible to achieve inter-domain image translation, this bi-directional GAN framework with the transfer loss only has two drawbacks when applied to medical images. Firstly, it is difficult to guarantee the generation of stable results due to its under-constraints in the adversarial training process. More specifically, the existing bi-directional GAN framework lacks adequate supervision information only based on the global adversarial loss and cycle consistency constraints. Secondly, for medical image enhancement, it is difficult to make sure that $G_{LQ\to HQ}$ and $G_{HQ\to LQ}$ capture important low-level features without extra detailed constraints being provided. On one hand, it is a challenge to remove excessively dark or bright regions so as to achieve a more uniform appearance consistent with human visual characteristics. On the other hand, subtle details of great significance to clinical analysis, such as the curvilinear structures of corneal nerve fibers or blood vessels, and the complete morphology of the digestive tract, might be blurred or even lost in the translated images. To address these drawbacks, we propose two novel terms - illumination regularization and structure loss (as shown in the purple arrows of Fig. 2), to guide the generator $G_{LQ\to HQ}$ in reaching a balance between illumination uniformity and structural restoration.

• **Illumination Regularization** The illumination regularization is proposed to improve overall illumination uniformity. It is realised as minimizing the illumination difference between local patches and the whole image. It represents a correcting factor that reflects the non-uniformity of illumination in the enhanced image, and can serve as prior knowledge of human vision. Calculation of the illumination correcting factor of a given image I is performed in the following steps:

1) Calculate the global average intensity of I;
2) Divide the image into $n \times m$ patches of the same size; then calculate the average intensity of each patch to obtain the illumination matrix D;
3) Subtract the average intensity of I from each element of D to form the illumination difference matrix E;
4) Rescale E into the illumination distribution matrix R of the same size as I via bicubic interpolation;
5) Calculate the average absolute value of elements in R.

A brief explanation of the above steps is given here. Global average intensity calculated in Step 1 represents the overall illumination level in the input image. Average intensity of each divided patch in Step 2 aims at achieving local illumination distribution in the input image. From Step 3 to Step 4, we obtain the illumination error distribution map of the input image. Note that both the matrices E and R represent the illumination distribution of the given image, and R is the rescaled version of E. Finally, we calculate the average illumination error in Step

5, in order to measure the illumination non-uniformity of the input image. The smaller the average illumination error, the more uniform the illumination of the given image. According to the above, the illumination regularization $\mathcal{L}_{ill}$ is defined as:

$$\mathcal{L}_{ill}\left(G_{LQ\to HQ}\right)=\mathbb{E}_{a\in LQ}\left[\mathbb{E}_{global}\left[\left|upsampling\left\{\mathbb{E}_{local}^{p\times p}\left[G_{LQ\to HQ}(a)\right]\right.\right.\right.\\ \left.\left.\left.-\mathbb{E}_{global}\left[G_{LQ\to HQ}(a)\right]\right\}\right|\right]\right], \tag{5}$$

where $\mathbb{E}_{global}\left[\cdot\right]$ denotes the global mean of the input image; $\mathbb{E}_{local}^{p\times p}\left[\cdot\right]$ aims at calculating the illumination matrix D based on each $p\times p$ patch divided in the input image; and $upsampling\left\{\cdot\right\}$ is intended to resize the illumination difference matrix E to the size of the original input image via bicubic interpolation. Note that the proposed illumination regularization is applied to $G_{LQ\to HQ}(a)$ only in that enhanced images generated from $G_{LQ\to HQ}$ should satisfy the constraints of illumination regularization. When generating low-quality images from high-quality ones, it is unnecessary to impose any constraint, as they may be caused by various unpredictable factors such as poor lighting or imaging noise.

• **Structure Loss** Although it is favorable to improving illumination uniformity, the using of illumination regularization alone might also lead to excessively low contrast, or even complete loss of vital details. The low-quality image and its enhanced version should exhibit similar structural features in spite of great differences in intensity and contrast distribution. Structural SIMilarity (SSIM) [31] provides a relatively appropriate measurement of this degree of similarity. Compared with mean squared error (MSE), SSIM can effectively characterize structural similarity between two images in three aspects: luminance, contrast and structure. Motivated by the structure comparison function in SSIM, we propose a kind of structure-aware prior - structure loss, based on the dissimilarity between the low-quality image and its enhanced version. Mathematically, it is formulated as:

$$\begin{aligned}&\mathcal{L}_{st}\left(G_{LQ\to HQ},G_{HQ\to LQ}\right)\\&=\mathbb{E}_{a\in LQ}\left[1-\frac{1}{M}\sum_{i=1}^{M}\frac{\sigma_{a_i,G_{LQ\to HQ}(a)_i}+c}{\sigma_{a_i}\sigma_{G_{LQ\to HQ}(a)_i}+c}\right]\\&+\mathbb{E}_{b\in HQ}\left[1-\frac{1}{M}\sum_{i=1}^{M}\frac{\sigma_{b_i,G_{HQ\to LQ}(b)_i}+c}{\sigma_{b_i}\sigma_{G_{HQ\to LQ}(b)_i}+c}\right],\end{aligned} \tag{6}$$

where a_i, b_i, $G_{LQ\to HQ}(a)_i$ and $G_{HQ\to LQ}(b)_i$ are the i-th local window in the images a and b and the corresponding generated images $G_{LQ\to HQ}(a)$, $G_{HQ\to LQ}(b)$ respectively; M is the number of local windows in each image; $\sigma_{a_i,G_{LQ\to HQ}(a)_i}$ and $\sigma_{b_i,G_{HQ\to LQ}(b)_i}$ are the covariance between a_i and $G_{LQ\to HQ}(a)_i$ and that between b_i and $G_{HQ\to LQ}(b)_i$ respectively; σ_{a_i}, σ_{b_i}, $\sigma_{G_{LQ\to HQ}(a)_i}$ and $\sigma_{G_{HQ\to LQ}(b)_i}$ are the standard deviations of a_i, b_i, $G_{LQ\to HQ}(a)_i$ and $G_{HQ\to LQ}(b)_i$ respectively; and c is a small positive constant used to avoid numerical instabilities.

Thus the overall objective function of the proposed StillGAN for medical image enhancement is defined as:

$$\begin{aligned}&\mathcal{L}\left(G_{LQ\to HQ},G_{HQ\to LQ},D_{LQ},D_{HQ}\right)\\&=\mathcal{L}_{transfer}\left(G_{LQ\to HQ},G_{HQ\to LQ},D_{LQ},D_{HQ}\right)\\&+\alpha\mathcal{L}_{ill}\left(G_{LQ\to HQ}\right)+\beta\mathcal{L}_{st}\left(G_{LQ\to HQ},G_{HQ\to LQ}\right),\end{aligned} \tag{7}$$

where α and β are the positive parameters controlling the weights of the illumination regularization and structure loss respectively.

IV. Experiments

A. Datasets

Three different medical imaging modalities, confocal microscopy, color fundus and endoscopy, were used to validate the proposed StillGAN method.

• **CORN-2 (CORneal Nerve Database)** The dataset was constructed for confocal image enhancement, which is based on a publicly-available corneal confocal microscopy (CCM) dataset [4]. The CORN-2 dataset contains a total of 688 confocal images of size 384×384 acquired using a Heidelberg Retina Tomograph equipped with a Rostock Cornea Module (HRT-III) microscope. Low quality in this confocal dataset manifests as low contrast, speckle noise and non-uniform intensity. Accordingly, one image expert and one clinician were invited to grade the confocal image quality based on [13], and they come to a consensus to divide the CORN-2 dataset into 340 low-quality images and 288 high-quality images for training, with the remaining 60 low-quality images reserved for testing. In addition, all visible nerve fibers in confocal images were manually annotated at centerline level.

• **Fundus Multi-disease Diagnosis (iSee) Dataset for enhancement** The iSee dataset [32] was collected by a local hospital for research on automated disease analysis and diagnosis in clinical applications. This dataset contains a total of 10000 color fundus images of size 1942×1940, and includes instances of some common eye diseases, such as age-related macular degeneration (AMD), pathological myopia (PM), glaucoma and diabetic retinopathy (DR). There are large variations of image quality in the iSee dataset, including examples of under/over exposure, blur/noise and artifacts. To evaluate the performance of enhancement approaches on color fundus images, our image experts and clinicians were also invited to select 1,520 color fundus samples from this dataset based on [6]: 733 low-quality images and 637 high-quality images for training, and the remaining 150 low-quality images for testing. Note that all the samples showing normal eyes and the various eye diseases were distributed uniformly in low- and high-quality image subsets. In addition, two experienced clinicians annotated foveal locations of these selected samples for quantitative assessment. Firstly, clinicians manually labeled the foveal centre point in each color fundus image. Then we generated a bounding box centered at the annotated point as the final ground truth of the foveal region. Following our experts' observation of these color fundus images and their suggestions, we set the size of the bounding box as 150×150.

• **EASE (Endoscopy Automated Scene Enhancement)** The EASE dataset is an endoscopy dataset collected from the public CVC-EndoSceneStill dataset [33] for endoluminal scene enhancement. Specular highlights and dark shadows also degrade the visual quality of these endoscopy images. Through careful selection based on [34], two clinicians selected 267 low-quality images and 123 high-quality endoscopy images as the training set and 70 low-quality images as the testing set. All the images in the EASE dataset have a size of 384×288.

B. Implementation Settings

The proposed StillGAN was implemented with PyTorch library, and the experiments were conducted on a single NVIDIA GPU (Tesla P40 with 24 GB). All training images were resized to 512×512, and a random flipping in the lateral or vertical direction was applied for data augmentation. For our StillGAN, we selected two patch sizes - 48×48 and 96×96 respectively, to obtain the illumination regularization and then the average of these two options was computed as the final one, and set local windows of 11×11 for the calculation of the structure loss. Adam optimization was applied to train the two adversarial pairs, with the initial learning rate of 0.0002 and a batch size of 1. The weighted parameters in the final objective function were experimentally set as: $\lambda_1 = 10$, $\lambda_2 = 5$, $\alpha = 1$, $\beta = 5$. Note that even though these hyperparameters need fine-tuning carefully, their settings do follow certain principles. The weighted coefficient λ_1 of the cycle consistency loss in fact controls the content consistency between a low-quality image and its high-quality version in one cycle mapping. Thus λ_1 should be large enough to ensure the correspondence before and after enhancement. The identity mapping loss enforces invariance of intra-domain translation. In particular, one real high-quality image should remain unchanged after enhancement. To this end, the weighted coefficient λ_2 should also hold a certain proportion. For illumination regularization, too large a value of α often leads to low contrast in the whole image or even loss of local details, while too small a value of α often makes it difficult to attain the expected uniform illumination. In general, the weighted parameter β of the structure loss should be large enough, but too large a value of β usually results in amplifying noise or producing artifacts. All above settings were consistent in applying each dataset.

In order to validate our proposed StillGAN, the following state-of-the-art approaches were selected for comparison on each dataset: three conventional methods, including contrast limited adaptive histogram equalization (CLAHE) [22], dark channel prior (DCP) [8], and low-light image enhancement (LIME) [35], and three deep learning methods, including neural style transfer (NST) [17], multi-style generative network (MSG-Net) [19], and EnlightenGAN [20]. The parameters in the conventional methods were set to the default values as in the corresponding articles. For each deep learning method, the same training datasets and data augmentation were adopted, with the hyperparameters tuned to achieve a relatively satisfied performance. Furthermore, we also conduct ablation studies on the illumination regularization and structure loss and see how they affect the performance of our proposed StillGAN. Finally, we investigate the clinical impact of image enhancement on three tasks: image reclassification, nerve fibre tortuosity grading and disease diagnosis.

C. Evaluation over Corneal Confocal Microscopy

Firstly, we validate the proposed StillGAN on the CORN-2 dataset. In addition to making visual comparisons, we also evaluate it in the following metric and task: by calculating signal-to-noise ratio (SNR) based on regions of nerve fibers and by comparing the performance of nerve fiber segmentation guided by enhancement.

TABLE I: SNR (unit: dB) of the original and enhanced corneal confocal microscopy images using different approaches. (S: structure loss; I: illumination regularization)

Methods	r=3	r=5	r=7
Original	17.47±1.09	17.61±1.14	17.65±1.18
CLAHE [22]	16.56±0.59	16.73±0.62	16.79±0.65
DCP [8]	14.59±1.03	14.88±1.13	14.99±1.21
LIME [35]	16.43±1.43	16.78±1.51	16.89±1.57
NST [17]	16.61±1.23	16.89±1.27	17.01±1.28
MSG-Net [19]	19.12±0.61	19.92±0.57	20.22±0.54
EnlightenGAN [20]	18.40±1.13	19.26±1.10	19.70±1.10
Baseline	19.55±0.85	20.14±0.84	20.41±0.87
Baseline + I	20.30±0.89	20.93±0.80	21.22±0.77
Baseline + S	20.11±1.05	20.75±0.97	21.04±0.93
StillGAN	**20.35±0.93**	**21.06±0.88**	**21.41±0.88**

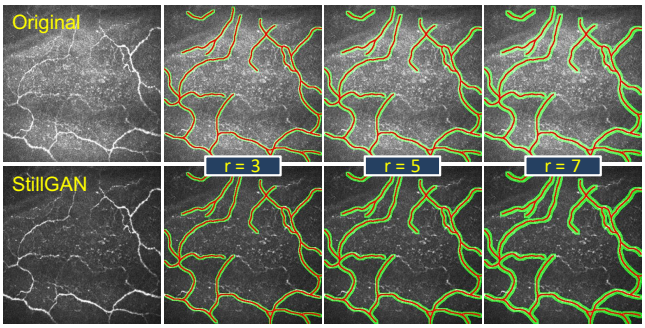

Fig. 4: An example to show the regions selected as background so as to calculate the SNR. The background (green color) was determined by a disk-shaped dilation operation on the manually traced fibers (red color) with a radius of 3, 5 and 7 pixels, respectively. Top row: an original image; Bottom row: the example enhanced by our StillGAN.

1) Evaluation in SNR: For quantitative assessment of confocal image quality, we first calculated signal-to-noise ratio (SNR) based on manual annotations of nerve fibers, which is calculated as: $SNR = 10\log_{10}\left(\max(I_s)^2/\sigma_b^2\right)$, where $\max(I_s)$ denotes the maximum intensity of signal regions I_s (centerline-level regions of the manually traced nerve fibers) in the image, and σ_b is the standard deviation of the background regions. In our experiments, we defined the regions (except signal regions) after a disk-shaped dilation operation on signal regions with a radius (r) of 3, 5 and 7 pixels, respectively as background regions. Fig. 4 shows one example with signal regions (marked in red) and three kinds of background regions (marked in green). The SNR results of different enhancement methods are shown in Table I. As illustrated in the table, our StillGAN achieves the highest SNR when compared with all the selected state-of-the-art methods. It indicates that the proposed StillGAN is more successful in eliminating uneven intensity in background regions and highlighting signal regions. Furthermore, its enhanced results demonstrate a huge advantage over the original images by an improvement in SNR of 2.88 dB, 3.45 dB and 3.76 dB for $r = 3$, 5, and 7, respectively. The significant improvement in SNR is confirmed

TABLE II: Segmentation performance of the original and enhanced corneal confocal microscopy images using different approaches. (S: structure loss; I: illumination regularization)

Methods	AUC	ACC	SEN	G-mean	Kappa	Dice
Original	0.735±0.107	0.969±0.013	0.421±0.186	0.628±0.159	0.528±0.182	0.541±0.182
CLAHE [22]	0.777±0.087	0.970±0.010	0.488±0.160	0.685±0.122	0.570±0.139	0.584±0.139
DCP [8]	0.899±0.034	0.964±0.007	0.708±0.084	0.830±0.050	0.615±0.093	0.633±0.093
LIME [35]	0.895±0.033	0.960±0.009	0.698±0.080	0.823±0.048	0.585±0.102	0.606±0.102
NST [17]	0.777±0.080	0.958±0.016	0.490±0.148	0.686±0.108	0.494±0.167	0.515±0.162
MSG-Net [19]	0.754±0.086	0.964±0.009	0.441±0.167	0.647±0.133	0.495±0.160	0.512±0.160
EnlightenGAN [20]	0.853±0.037	0.960±0.010	0.671±0.072	0.807±0.046	0.580±0.104	0.601±0.103
Baseline	0.900±0.052	0.971±0.006	0.748±0.112	0.854±0.069	0.673±0.113	0.688±0.112
Baseline + I	0.918±0.042	**0.977±0.006**	0.776±0.098	0.873±0.060	0.735±0.084	0.747±0.084
Baseline + S	0.908±0.054	0.971±0.007	0.769±0.114	0.865±0.073	0.680±0.118	0.695±0.117
StillGAN	**0.922±0.041**	**0.977±0.006**	**0.788±0.096**	**0.879±0.058**	**0.736±0.090**	**0.748±0.090**

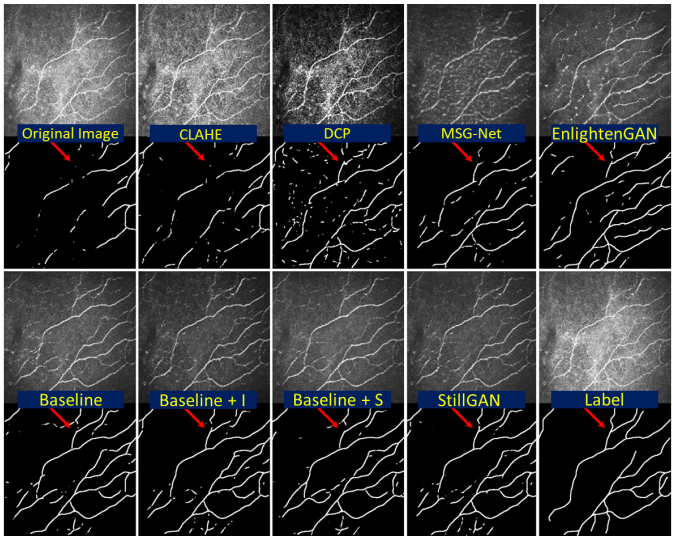

Fig. 5: An example of corneal confocal microscopy and its enhancement using different approaches (the top and third row), and their guided nerve fiber segmentation results via CS-Net (the second and bottom row).

by the statistical analysis (all $p < 0.05$).

2) Evaluation in nerve fiber segmentation: In order to confirm the impact of image enhancement on subsequent analysis tasks, we further performed corneal nerve fiber segmentation and compared segmentation results guided by enhancement with that of the original images. To this end, we employed a pre-trained CS-Net [4], which had been trained on high-quality corneal confocal microscopy images with manually traced nerve fibers, for corneal nerve fiber segmentation in the low-quality and the enhanced images via enhancement approaches. For assessment of the segmentation performance, we calculated the following metrics between the predicted centerlines and ground truth: area under the ROC curve (AUC), accuracy (ACC), sensitivity (SEN), G-mean score [36], Kappa score, and Dice coefficient (Dice). Note, a three-pixel tolerance region around the manually-traced nerves is considered as true positive [37] for the calculation of these metrics.

The top row and the third row of Fig. 5 show an original image and its the enhancement results using different methods, while the second row and the bottom row depict enhancement-guided fiber segmentation results obtained using CS-Net. It can be seen that more completed fibers have been identified in the sample enhanced by our StillGAN, whose location is indicated by the red arrows, since the contrast between the nerve fibers and the background regions has been significantly improved, and more uniform responses in both the regions have been achieved. With the guidance of our StillGAN, CS-Net is more sensitive in detecting small fibers with low contrast. This finding is also confirmed by the segmentation results in Table II: our StillGAN achieves the best segmentation performance and outperforms the state-of-the-art EnlightenGAN and the baseline framework by 17.44% and 5.35% in SEN, 26.90% and 9.36% in Kappa, 8.92% and 2.93% in G-mean, 24.46% and 8.72% in Dice respectively. Paired t-tests were conducted on AUC, and all $p < 0.05$ demonstrate that our method can significantly improve nerve fiber segmentation performance, especially in reducing missing rate, which is more useful for monitoring and diagnosing nerve-related diseases.

TABLE III: High-quality score and fovea localization performance (mean ± standard deviation) on the original and enhanced color fundus images via different enhancement approaches. (S: structure loss; I: illumination regularization)

Methods	s_{HQ}	d
Original	0.0940±0.0760	129.00±358.19
CLAHE [22]	0.1906±0.1668	76.16±249.69
DCP [8]	0.1156±0.0802	339.54±542.34
LIME [35]	0.1140±0.0855	95.57±273.09
NST [17]	0.0978±0.1165	200.31±426.20
MSG-Net [19]	0.0709±0.0477	115.37±292.47
EnlightenGAN [20]	0.0920±0.0535	191.91±389.62
Baseline	0.2426±0.1767	74.38±239.29
Baseline + I	0.3164±0.2428	64.30±198.01
Baseline + S	0.3103±0.2610	67.13±227.17
StillGAN	**0.3487±0.2437**	**62.87±205.59**

D. Evaluation over Color Fundus Images

Two different experiments have been conducted, so as to verify the effectiveness of our StillGAN on color fundus images.

1) Evaluation in retinal image quality assessment score: We adopted a state-of-the-art classification network called

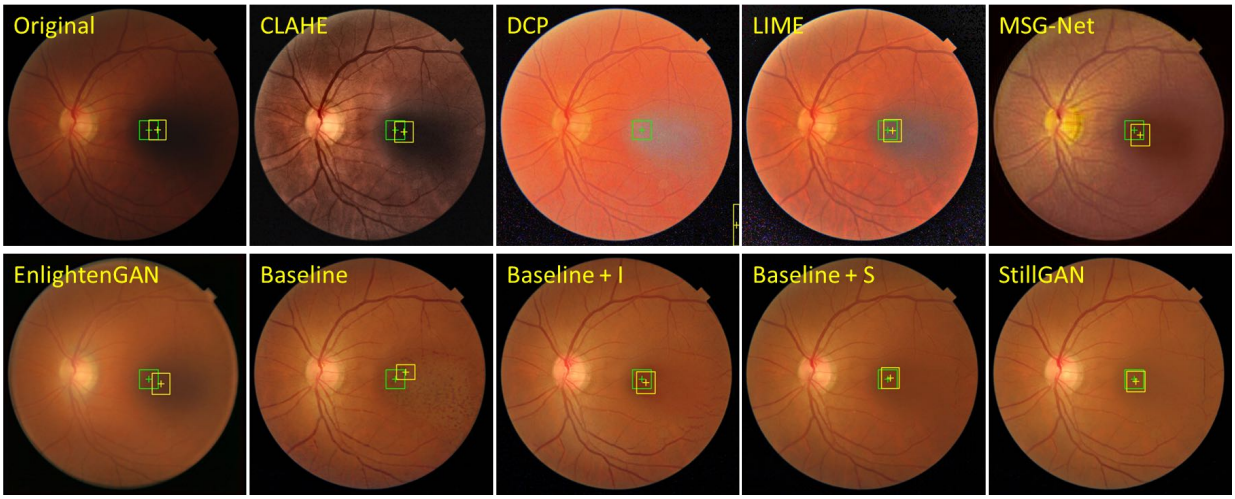

Fig. 6: An example of color fundus enhanced using different approaches, and their guided fovea localization results (including the bounding box and its center). Ground truth: green box and cross; Prediction: yellow box and cross.

MCF-Net [6], which was proposed for retinal image quality assessment. We employed the pre-trained MCF-Net to predict the high-quality score (denoted as s_{HQ}) of those low-quality images and their enhanced ones using different approaches. The high-quality score mainly measures the overall perceptual quality of color fundus images in terms of different color-spaces. It is apparent that s_{HQ} of the original low-quality image is usually lower; the higher the metric s_{HQ} of the enhanced image, the better the performance of the enhancement approach. Table III provides s_{HQ} of the original images and their enhanced results using different methods. The proposed StillGAN has achieved the highest high-quality score among these competing methods. Compared with the original images, our method increases the high-quality score by over 2.7 times, which demonstrates that our StillGAN can significantly improve the overall visual perception of the color fundus images.

2) Evaluation in fovea localization: We conducted fovea localization of the enhanced fundus images to verify the localization performance gains brought about by the proposed method and the others. We utilized a pre-trained fovea localization framework based on Faster R-CNN [38], which had been trained on high-quality color fundus images with manual fovea localization, for fully automatic fovea localization on the low-quality images, with and without application of image enhancement approaches. To measure the precision of fovea localization, we used the Euclidean distance (denoted as d) between the predicted box center and the box center of the ground truth following [38] as the fovea localization error.

Fig. 6 shows the fovea localization results achieved by different enhancement approaches. For their comparison, the predicted foveal region and its center are marked in the yellow box and cross respectively, while the ground truths are marked in the green box and cross instead. It can be seen that the original image of the example in Fig. 6 exhibits poor exposure around the foveal region, which leads to imprecise localization. It is worth noting that fovea localization based on the result of DCP is entirely wrong, though it shows the overall homogeneous appearance. By visual inspection, we found that it amplifies the noise and even produces some color distortion in the image, especially in the foveal region. This is because its dehazing method changes the characteristics of the foveal region and makes the enhanced image far different from those in the real high-quality domain. EnlightenGAN tends to oversmooth the image including the foveal region, leading to certain localization deviations. Although our baseline framework is able to achieve the comparatively better overall visual effects, it still produces some artifacts around the foveal region which reduces its localization accuracy. In sharp contrast, for our StillGAN, these artifacts are almost entirely removed, and higher fovea localization accuracy is achieved. The fovea localization error is also presented in Table III. We can see that enhanced images using our StillGAN have yielded the smallest error in fovea localization.

E. Evaluation over Endoscopy

Finally, the proposed StillGAN was verified over the EASE dataset. For quantitative evaluation, we adopted three no-reference image quality assessment metrics: Natural Image Quality Evaluator (NIQE) [39], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [40] and Perception based Image Quality Evaluator (PIQE) [41]. Note that the lower the score achieved using these no-reference assessments, the better the endoscopy image quality.

Table IV shows the results of endoscopy images enhanced using different approaches. When the proposed StillGAN is compared with the competing methods - it achieves the best performance in NIQE and PIQE, and similar performance to EnlightenGAN in BRISQUE, where the former is only 0.17 lower than the latter. The statistical analysis also indicates that differences found were not statistically significant between EnlightenGAN and StillGAN in terms of BRISQUE ($p = 0.85 > 0.05$).

The top row of Fig. 7 illustrates the enhancement results achieved by two conventional (CLAHE and DCP) and one deep learning-based (EnlightenGAN) enhancement method,

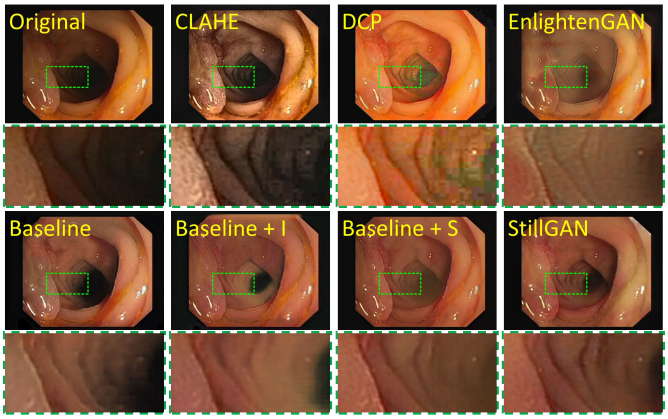

Fig. 7: An example of endoscopy enhancement using different approaches.

TABLE IV: No-reference assessment results (mean ± standard deviation) of different enhancement approaches. (S: structure loss; I: illumination regularization)

Methods	NIQE	BRISQUE	PIQE
Original	4.40±0.67	36.70±5.42	37.27±11.50
CLAHE [22]	4.52±0.89	30.76±3.88	29.52±4.89
DCP [8]	4.13±0.64	35.45±5.59	35.09±6.37
LIME [35]	4.27±0.69	30.74±4.89	34.95±7.20
NST [17]	9.42±1.96	30.28±3.64	25.48±6.17
MSG-Net [19]	7.26±0.55	56.72±2.29	93.43±11.96
EnlightenGAN [20]	4.38±0.88	**24.35±4.18**	33.07±6.47
Baseline	4.38±0.62	27.81±7.70	29.54±5.28
Baseline + I	4.33±0.61	27.70±6.71	25.25±9.35
Baseline + S	4.27±0.60	26.00±6.59	28.12±6.01
StillGAN	**3.84±0.64**	24.52±5.38	**23.48±6.42**

respectively. CLAHE demonstrates limited improvement in dark regions. It can be seen that DCP improves the overall illumination conditions of the image, but also amplifies noise in extremely dark regions, and even leads to some color distortions. This might be because our endoscopy image does not meet the assumption of the reverse dehazing method. EnlightenGAN generates universally over-smoothed images with many details blurred. In contrast, the proposed StillGAN produces visually satisfactory results with both more uniform illumination and clearly perceivable structural details, especially in poorly-illuminated regions. These results show that the proposed StillGAN is powerful in enhancing images with uniform illumination conditions and preserving local details.

F. Ablation studies

In this paper, the proposed StillGAN incorporates two novel terms - illumination regularization and structure loss, into our bi-directional GAN framework for medical image enhancement. In order to investigate their contributions, we carry out the following ablation studies on the baseline bi-directional GAN framework in conjunction with different combinations of these two terms.

1) Illumination regularization.: To discuss the effectiveness of the proposed illumination regularization, we compared the performance of the baseline method and that with illumination regularization only over the three medical imaging modalities. The experimental results in Fig. 5-Fig. 7 and Table I-Table IV show that the illumination regularization brings significant improvements of overall illumination uniformity to the baseline GAN framework. In particular for some degradation factors, such as intensity inhomogeneity or speckle noise in confocal microscopy, and uneven exposure or other light disturbance in color fundus photography, the illumination regularization usually works well. For corneal confocal microscopy, we found that our bi-directional GAN framework using the illumination regularization had higher SNR with different background regions. This indicates that non-uniform intensity and noise in the background regions could be further suppressed by introducing the illumination constraint. In addition, better nerve fiber segmentation could be achieved via enhancement using the illumination term, indirectly confirming that the illumination regularization is conducive to eliminating the influence of non-uniform illumination on the nerve fiber segmentation task. For color fundus images, the illumination regularization also helped our baseline method to improve both its high-quality score and fovea localization. Especially for those samples with under-exposure or slight over-exposure, it could lead to great improvement in overall visual quality that is well aligned to human perception, resulting in a better high-quality score. Some fundus degradation factors, such as light transmission disturbance [42] and absence of exposure, could impair observation and fovea localization. The illumination regularization is an appropriate way for overcoming degradation factors to a large extent, thereby improving fovea localization. However, it alone can also over-smooth regions of interest or even blur vital details: as we can see in Fig. 7, that the severely dark region in the endoscopy image becomes brighter but also appears blurry or even loss of its texture.

2) Structure loss.: Furthermore, we also verified the impact of the proposed structure loss on the enhancement performance of our bi-directional GAN framework. By contrast with the illumination regularization, the structure loss attempts to mine and retain structural information from the original images. In corneal confocal microscopy images, the topology of nerve fibers is the most important structural information. As shown in Tables I and II, the application of the structure loss also resulted in a slightly higher SNR and better nerve fiber segmentation compared to the baseline method. This demonstrates that this structure-aware prior could assist the bi-directional GAN framework in focusing on and highlighting the structural details, leading to improvement of contrast between signal and background regions. For color fundus images, the structure loss could spotlight structural features of some important retinal biomarkers, such as the fovea, optic disc and vessels. Thus both the high-quality score and fovea localization are achieved by the structure constraint. In addition, it also guides the bi-directional GAN framework in producing clearer digestive tract imagery and thus improves endoscopy quality. Even though the proposed structure loss would help to reduce the risk of missing structural details, it may be sensitive to noise or other interferences. For example in Fig. 5, the corneal confocal microscopy image enhanced by Baseline + S seems inadequate

TABLE V: Nerve fiber tortuosity classification results before and after enhancement using our StillGAN (w/o enhancement).

	ACC	PRE	F1-score
Level 1	95.33%/89.67%	93.94%/61.11%	81.58%/68.04%
Level 2	90.67%/77.67%	87.01%/61.84%	82.72%/58.39%
Level 3	91.00%/83.00%	76.92%/57.97%	78.74%/61.07%
Level 4	93.67%/87.00%	86.40%/85.15%	91.91%/81.52%
Average	92.50%/83.91%	85.70%/69.48%	85.10%/68.81%

in the elimination of non-uniform background intensity.

The above ablation studies show that both the illumination regularization and structure loss have their own advantages and drawbacks. The former focuses on overall illumination uniformity at the risk of oversmoothing or blurring, while the latter tends to preserve some vital structural details rather than eliminate those low-quality factors. Thus, a combination of both the terms could reach a balance between illumination uniformity and structural preservation to avoid either blurring or other excessive degradation.

G. Clinical impact of enhancement

To further evaluate the clinical impact of image enhancement, we carry out three experiments on image reclassification, objective nerve fibre tortuosity grading and subjective disease diagnosis as follows.

1) Impact on image reclassification: In order to further validate the clinical impact of our method, we conducted a simple experiment of image quality re-classification for each dataset. We asked the same clinicians to re-classify all the images including the low-quality images from testing subsets (60 corneal confocal images, 150 retinal color fundus images, and 70 endoscopy images) after enhancement. In order to avoid bias from the experts, we did not disclose that these images had already been enhanced. As expected, 42 out of 60 corneal confocal images, 145 out of 150 retinal color fundus images, and 58 out of 70 endoscopy images have been identified as high-quality ones by the same experts under the same assessment protocol. These results show that the proposed method have successfully improved the quality of most images from the clinical point of view.

2) Impact on nerve fiber tortuosity grading: Previous studies have shown that corneal nerve tortuosity is related to hypertensive retinopathy [43], dry eye disease [44] or diabetic neuropathy [45], so the tortuosity level grading is of great importance in clinical practice. We employed a state-of-the-art tortuosity grading method [13], to estimate the nerve fiber tortuosity levels of 300 confocal images from an in-house dataset, with and without applying our enhancement method. These images were categorized into four groups based on fiber tortuosity levels by two experts based on a previously published protocol [46], and these labels were used as ground truth for objective tortuosity level evaluation. Finally, these images consist of 43, 85, 62 and 110 images at tortuosity levels 1 to 4 respectively.

Table V shows the nerve fiber tortuosity classification results. It demonstrates that our StillGAN promotes the performance of nerve tortuosity analysis, especially the average accuracy, precision and F1-score of four tortuosity levels have increased by 10.24%, 23.34% and 23.67% respectively after having applied our StillGAN on the original images. These objective results show that the quality improvement of confocal images can promote the nerve fibre tortuosity grading, which further confirms the clinical values of our StillGAN.

TABLE VI: Diagnosis results on the original and enhanced color fundus images via our StillGAN.

	ACC	PRE	SEN	SPE	F1-score
Original	75.00%	71.93%	82.00%	68.00%	76.64%
StillGAN	**81.50%**	**80.00%**	**84.00%**	**79.00%**	**81.95%**

3) Impact on disease diagnosis: In order to verify the usefulness in clinical decision-making, we invited an ophthalmologist to diagnose diabetic retinopathy from images with and without enhancement. To this end, we constructed a new dataset, and it includes 200 low-quality color fundus images from 100 healthy eyes and 100 eyes with diabetic retinopathy. All the 200 low-quality images were selected from the 'usable' grade of Eye-Quality (EyeQ) dataset [6], with their pathology condition provided, i.e., with or without diabetic retinopathy. Then these images were enhanced using our StillGAN. The ophthalmologist was invited to complete the diagnostic task over the original images first, and to review the enhanced ones two days later in order to avoid subjective factors.

The diagnostic performances on the original and enhanced images are shown in Table VI. It can be seen that our StillGAN improves the diagnosis performance of the ophthalmologist by 8.67% in ACC, 11.22% in PRE, 2.44% in SEN, 16.18% in specificity (SPE) and 6.93% in F1-score respectively. These results clearly verify the effectiveness of image enhancement using the proposed StillGAN in clinical practice.

V. Discussion and Conclusion

As a pre-processing step of automatic analysis and diagnosis, medical image enhancement is crucial to produce high-quality versions of captured images for the tasks. However, it is still challenging to obtain high-quality images due to diversity in illumination conditions across different medical imaging devices. Low-quality images not only inhibit clinicians from observation of important tissues or lesions, but also degrade the performance of subsequent automatic analysis methods.

A. Limitations

We further analyze the unsatisfactory enhancement cases. Fig. 8 illustrates three examples from different datasets. For the corneal confocal microscopy image, some regions with non-uniform intensity, e.g., corneal scar, as the red arrow indicated in Fig. 8 (a), appear as nerve-like structure after image enhancement (Fig. 8 (b)). Such structure may falsely be recognized as nerve fibers by computer or even clinicians. This implies that our structure loss needs to be further improved - for those objects with similar structures, it is difficult for the loss term to distinguish between our concerning biological tissues and low-quality factors. Faculae may exist in some

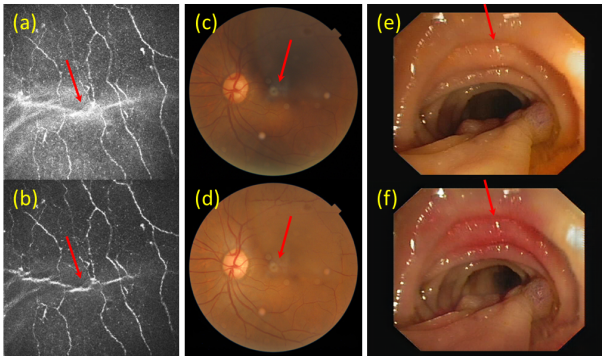

Fig. 8: Three typical cases with unsatisfactory enhancement results using our StillGAN: low-quality images (top row) and the corresponding enhanced ones (bottom row). From the left to the right: corneal confocal microscopy, color fundus and endoscopy, respectively.

color fundus images due to different imaging conditions, as shown by the red arrow of Fig. 8 (c). Unfortunately, it is hard for our StillGAN to remove this artifact, and may further lead to a lesion-like appearance. Thus, the structure loss needs to be improved for more adaptability. As the case of Fig. 8 (e-f), part of the intestinal mucosa appears artificially red after having been enhanced by our StillGAN, which might mislead clinicians to diagnose it as hemorrhage or inflammation. This is partly caused by color diversity of those training samples, which might bring a risk of color transfer to the training of our StillGAN. We would consider introducing color consistency constraints to alleviate such color transfer in future.

From the above unsatisfactory enhancement cases, we can see that there is a risk of incorrect translation (e.g., change color or create lesion-like artifacts) for image enhancement using most GANs including our StillGAN. That is because our model may not capture enough heterogeneity from different diseases or conditions during the training, and it may be encountered in clinical practice in both the low- and high-quality domains. This may be mitigated when a training set contains sufficient healthy and unhealthy samples in both the low- and high-quality domains, the generators could then learn to distinguish more accurately between imaging quality factors and disease conditions, and thus finally achieve reliable translation.

The complicated procedure of hyperparameter adjustment is another limitation of our method. Apart from the weighted coefficients of the loss terms, the improper setting of the patch size for the proposed illumination regularization may lead to a risk of altering the image. Structural details would be partly or even completely lost with a small patch size. Especially in the case where the patch size is 1, all the pixels of the generated high-quality image will tend to have the same global average intensity value. On the contrary, if the patch size is too large, the illumination term will play a limited role in improving the overall illumination uniformity. In the extreme case where the patch size is the same as the image size, the calculated illumination term will be zero and will have no impact on improving overall illumination uniformity during training.

B. Conclusion

In this paper, we have proposed an unpaired learning framework called StillGAN for medical image enhancement, where low- and high-quality images are treated as being in two different domains. The primary advantage of our StillGAN is that it learns to migrate the characteristics of high-quality images into low-quality ones via unpaired training, and thus has an advantage of easy implementation. Furthermore, by incorporating constraints on illumination and structure, overall illumination uniformity and well-restored structural details could be achieved in the enhanced images. Experimental results demonstrate that the performances of nerve fiber segmentation, nerve tortuosity grading, fovea localization, and disease diagnosis could be improved via our StillGAN.

Most existing bi-directional GANs such as CycleGAN primarily focus on learning intra-domain global appearance and inter-domain cycle-consistency, and are thus ineffective in capturing local details. In medical images, local details are particularly important for clinical interpretation. While the bi-directional GAN is usually under-constrained, in this paper, two novel constraints including illumination regularization and structure loss are developed and incorporated into its objective function, in order to obtain better illumination condition and clearer structural details for clinical interpretation and subsequent analysis. The former aims at improving illumination uniformity via minimizing the difference of illumination distribution in the enhanced images, while the latter is introduced to preserve structural details as much as possible by reducing the dissimilarity in terms of structure between the low-quality and enhanced images. Compared with other state-of-the-art methods, the StillGAN achieves overall better performance in various metrics for enhancing multi-modality images.

In clinical practice, we often cannot tell whether images show disease or not. It is difficult for clinicians to describe the appearance and identify the location of lesions, or even to judge whether a sample is normal or pathological from a low-quality medical image. Improvement of medical image quality and contrast is crucial to improve the interpretation of clinicians about the appearance of biological tissues, and thus the accuracy of decision making, which is of great clinical concern for diagnosis and therapy planning. The purpose of image enhancement is to help clinicians to more easily identify diseases from images. By visual inspection, many lesions existing in the most low-quality images could be easily spotted after enhancement by our StillGAN. With cycle consistency and identity constraints, the generators are well positioned to acquire the knowledge necessary for translating the input image to an output one, while maintaining the overall appearance of the images before and after translation. In addition, our structure loss further constrains the appearance. Last but not least, a training set with a certain heterogeneity from different disease or conditions that may be encountered in clinical practice in both the low-quality and high-quality domains could help to reduce the risk of changing lesions or generating lesion-like artifacts. In the end, clinicians can more easily judge from an enhanced image whether the sample is normal or pathological, and examine the appearance of and localize

lesions if pathological. In the future, we would consider further adapting our StillGAN to other medical imaging modalities, and apply the resulting enhanced images in the real-world clinical scenarios to assist in disease diagnosis.

REFERENCES

[1] K. Doi, "Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology," *Physics in Medicine and Biology*, vol. 51, no. 13, pp. R5–R27, Jun 2006.

[2] S. Philip, L. M. Cowie, and J. A. Olson, "The impact of the health technology board for scotland's grading model on referrals to ophthalmology services," *Br. J. Ophthalmol.*, vol. 89, no. 7, pp. 891–896, 2005.

[3] T. J. MacGillivray, J. R. Cameron, Q. Zhang, A. El-Medany, C. Mulholland, Z. Sheng *et al.*, "Suitability of uk biobank retinal images for automatic analysis of morphometric properties of the vasculature," *PLoS One*, vol. 10, no. 5, pp. 1–10, 05 2015.

[4] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao *et al.*, "Cs-net: Channel and spatial attention network for curvilinear structure segmentation," in *MICCAI*, Cham, 2019, pp. 721–730.

[5] Y. Zhao, Y. Zheng, Y. Zhao, Y. Liu, Z. Chen, P. Liu *et al.*, "Uniqueness-driven saliency analysis for automated lesion detection with applications to retinal diseases," in *MICCAI*, 2018, pp. 109–118.

[6] H. Fu, B. Wang, J. Shen, S. Cui, Y. Xu, J. Liu *et al.*, "Evaluation of retinal image quality assessment networks in different color-spaces," in *MICCAI*, 2019, pp. 48–56.

[7] M. Abdullah-Al-Wadud, M. Kabir, M. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, 2007.

[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, 2011.

[9] ——, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013.

[10] Y. Zhao, Y. Zheng, Y. Liu, Y. Zhao, L. Luo, S. Yang *et al.*, "Automatic 2-d/3-d vessel enhancement in multiple modality images using a weighted symmetry filter," *IEEE Trans. Med. Imaging*, vol. 37, no. 2, pp. 438–450, 2018.

[11] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, 1997.

[12] A. M. Gonzales and A. M. Grigoryan, "Fast retinex for color image enhancement: methods and algorithms," in *Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2015*, vol. 9411, International Society for Optics and Photonics. SPIE, 2015, pp. 129–140.

[13] Y. Zhao, J. Zhang, E. Pereira, Y. Zheng, P. Su, J. Xie *et al.*, "Automated tortuosity analysis of nerve fibers in corneal confocal microscopy," *IEEE Trans. Med. Imaging*, vol. 39, no. 9, pp. 2725–2737, 2020.

[14] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans," in *CVPR*, June 2018.

[15] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

[16] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "Msr-net: Low-light image enhancement using deep convolutional network," *arXiv preprint arXiv:1711.02488*, 2017.

[17] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *J. Vision*, vol. 16, no. 12, p. 326, 2016.

[18] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, Oct 2017, pp. 2223–2232.

[19] H. Zhang and K. Dana, "Multi-style generative network for real-time transfer," in *ECCV Workshops*, September 2018.

[20] Y. Jiang, X. Gong, D. Liu, Y. Cheng, F. Chen, X. Shen *et al.*, "Enlightengan: Deep light enhancement without paired supervision," *arXiv preprint arXiv:1906.06972*, 2019.

[21] Y. Ma, Y. Liu, J. Cheng, Y. Zheng, M. Ghahremani, H. Chen *et al.*, "Cycle structure and illumination constrained gan for medical image enhancement," in *MICCAI*, Cham, 2020, pp. 667–677.

[22] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV*, 1994, pp. 474–485.

[23] X. Zhang, P. Shen, L. Luo, L. Zhang, and J. Song, "Enhancement and noise reduction of very low light level images," in *ICPR*, November 2012, pp. 2034–2037.

[24] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *ICIP*, September 2015, pp. 3730–3734.

[25] M. Sulami, I. Geltzer, R. Fattal, and M. Werman, "Automatic recovery of the atmospheric light in hazy images," in *ICCP*, 2014, pp. 1–11.

[26] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *ICCV*, October 2017, pp. 4539–4547.

[27] W. Chen, W. Wenjing, Y. Wenhan, and L. Jiaying, "Deep retinex decomposition for low-light enhancement," *CoRR*, vol. abs/1808.04560, 2018.

[28] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, 2018.

[29] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *ECCV*, Cham, 2016, pp. 702–716.

[30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017, pp. 1125–1134.

[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[32] Y. Yan, M. Tan, Y. Xu, J. Cao, M. Ng, H. Min *et al.*, "Oversampling for imbalanced data via optimal transport," in *AAAI*, vol. 33, no. 01, 2019, pp. 5605–5612.

[33] D. Vázquez *et al.*, "A benchmark for endoluminal scene segmentation of colonoscopy images," *J. Healthc. Eng.*, vol. 2017, 2017.

[34] J. Bernal, J. Sánchez, and F. Vilariño, "Impact of image preprocessing methods on polyp localization in colonoscopy frames," in *EMBC*, 2013, pp. 7350–7354.

[35] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, 2017.

[36] J. Ri, G. Tian, Y. Liu, W. Xu, and J. Lou, "Extreme learning machine with hybrid cost function of g-mean and probability for imbalance learning," *Int. J. Mach. Learn. Cybern.*, 2020.

[37] P. Guimarães, J. Wigdahl, and A. Ruggeri, "A fast and efficient technique for the automatic tracing of corneal nerves in confocal microscopy," *Translational Vision Science & Technology*, vol. 5, no. 5, 09 2016.

[38] J. Wu, J. Wang, J. Xu, Y. Wang, K. Wang, Z. Shang *et al.*, "Fovea localization in fundus photographs by faster r-cnn with physiological prior," in *OMIA*, Cham, 2019, pp. 156–164.

[39] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013.

[40] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.

[41] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *NCC*, 2015, pp. 1–6.

[42] Z. Shen, H. Fu, J. Shen, and L. Shao, "Modeling and enhancing low-quality retinal fundus images," *IEEE Trans. Med. Imaging*, 2021.

[43] E. Grisan, M. Foracchia, and A. Ruggeri, "A novel method for the automatic grading of retinal vessel tortuosity," *IEEE Trans. Med. Imaging*, vol. 27, no. 3, pp. 310–319, 2008.

[44] A. Labbé, Q. Liang, Z. Wang, Y. Zhang, L. Xu, C. Baudouin, and X. Sun, "Corneal nerve structure and function in patients with non-sjögren dry eye: Clinical correlations," *Investigative Ophthalmology & Visual Science*, vol. 54, no. 8, pp. 5144–5150, 08 2013.

[45] K. Edwards, N. Pritchard, D. Vagenas, A. Russell, R. A. Malik, and N. Efron, "Standardizing corneal nerve fibre length for nerve tortuosity increases its association with measures of diabetic neuropathy," *Diabet. Med.*, vol. 31, no. 10, pp. 1205–1209, 2014.

[46] P. Su, Y. Zhao, T. Chen, J. Xie, Y. Zhao, H. Qi *et al.*, "Exploiting reliability-guided aggregation for the assessment of curvilinear structure tortuosity," in *MICCAI*, Cham, 2019, pp. 12–20.