SSP-Net: A Siamese-based Structure-Preserving Generative Adversarial Network for Unpaired Medical Image Enhancement

Guoxia Xu, *Member, IEEE,* Hao Wang, *Member, IEEE,* Marius Pedersen, *Member, IEEE,* Meng Zhao, *Member, IEEE,* Hu Zhu, *Member, IEEE,*

Abstract—Recently, unpaired medical image enhancement is one of the important topics in medical research. Although deep learning-based methods have achieved remarkable success in medical image enhancement, such methods face the challenge of low-quality training sets and the lack of a large amount of data for paired training data. In this paper, a dual input mechanism image enhancement method based on Siamese structure (SSP-Net) is proposed, which takes into account the structure of target highlight (texture enhancement) and background balance (consistent background contrast) from unpaired low-quality and high-quality medical images. Furthermore, the proposed method introduces the mechanism of the generative adversarial network to achieve structure-preserving enhancement by jointly iterating adversarial learning. Experiments comprehensively illustrate the performance in unpaired image enhancement of the proposed SSP-Net compared with other state-of-the-art techniques.

Index Terms—Siamese GAN ,Structural Preservation, Unpaired Medical Image Enhancement, Corneal Confocal Microscopy

1 Introduction

IMPROVING medical image quality has become a focus of research in recent years since it serves as a crucial foundation for clinical diagnosis and therapy. Because different medical image modalities are invariably impacted by various equipment, background, and other factors. The inability of low quality (LQ) images increases clinical observation uncertainty and presents significant challenges to ensuing image analysis and downstream applications [1]–[3]. Unpaired medical image enhancement [4]–[6] restored low-quality images to high-quality images with uniform illumination and clear structural details, which is crucial for many medical imaging applications. With the emergence of deep learning [7], [8], one of the most significant study areas in biomedical applications and digital health had emerged, including smart biomaterials [9], biomedical imaging [10], heartbeat or blood pressure measurements [11], providing valuable suggestions for better medical treatment and lifestyle. While major advancements in deep learning for medical and healthcare research had made, there are still gaps between the computer-aided treatment design and the needs of real-world healthcare.

- *Guoxia Xu, Hao Wang, and Marius Pedersen are with Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway. Meng Zhao is with the Engineering Research Center of Learning-Based Intelligent System (Ministry of Education), the Key Laboratory of Computer Vision and System (Ministry of Education), and the School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China . Hu Zhu is with Jiangsu Province Key Lab on Image Processing and Image Communication, Nanjing University of Posts and Telecommunications, Nanjing 210003, China. (Corresponding author: Hao Wang) (E-mail: hawa@ntnu.no)*

The performance of unpaired medical image quality enhancement technology has significantly increased with the advent of deep learning technology, thanks to paired training datasets. Lore et al. [12] proposed a method based on a deep autoencoder to identify features from low-light images and to enhance the image adaptively without excessive amplification. A clinically oriented fundus enhancement network (COFE-Net) was proposed by Shen et al. [13] to suppress global degradation factors based on the degradation model, which simultaneously preserved anatomical retinal structures and pathological characteristics for clinical observation and analysis. However, most of these fully supervised models require strict alignment of high/low-quality image pairs for training. And, the quality of the training set largely determines the performance of the model, which is very time-consuming and money-costly. These limitations greatly hinder the application of these models in clinical practice and usually unpaired dataset for medical imaging devices is easily obtained. Therefore, unsupervised methodology is a good trial to enhance image quality by learning knowledge from typical images and transferring it without the need for paired images. To address the problems of adversarial training instability and anatomical structure mismatch, Liu et al. [14] proposed a bidirectional multilayer contrast adaptive network (BMCAN) for unpaired cross-modal segmentation. With fewer training components and better feature representation, this method can effectively overcome the overfitting and domain shifting problems.

There have been an increasing number of studies on image quality improvement utilizing the unpaired learning approach in recent years thanks to the works of the genera-

tive adversarial network (GAN). [15] transformed the input image into the enhanced image based on a bidirectional generative adversarial network utilizing an image intensifier. Unpaired generative adversarial network model was proposed in [16] that aimed to solve the task with unpaired training data. It learned a map by converting images from the source domain to the target domain and combined with inverse mapping to enhance images by using cyclic consistency loss. A model for converting MR images to CT images was proposed in [17] using paired-unpaired unsupervised attention-guided generative adversarial networks (uagGANs). Based on paired dataset for pre-training and initialization, the uagGAN model was then retrained on unpaired datasets using a cascade method. In order to produce fine-structured images, pairwise pre-training was used to combine the Wasserstein GAN adversarial loss function with two new non-adversarial losses. In [18], they suggested a colorization network based on the CycleGAN model with a combination of the perceptual loss function and full variational loss function, in order to secure color medical images and enhance the quality of synthetic images while using unpaired training image data. The model in [19] also was constructed without low-light/normal light image pairs and can well handle various real-world test images. However, the biggest problem with these methods is that they focus on the global constraints of appearance and consistency and have poor performance in local detail learning.

Siamese networks [20] was originally proposed to deal with the classification problem, and it adopted the two-channel network with shared weights to measure the local distribution of the network. Not only is the classification label information considered, but also the local spatial distribution information between samples is achieved. It definitely helps for a small sample size task for the classification performance [21]. Li et al. [22] presented a discriminative self-attentive recurrent generative adversarial network, based on a recurrent GAN architecture, to address the super-resolution issue of actual face images. This network used unpaired samples to train both degraded and reconstructed networks. Contextual data was collected using a self-attentive method to reduce detail degradation. Bertinetto et al. [23] introduced a siamese network into a target tracking task, which greatly improved the sample limitations of online learning over traditional tracking methods. All these work well to demonstrate the strong feature generalization capacity of a siamese network structure, which can solve the situation with mismatched training data. Here, we tailor the same property to the medical image enhancement task. In this study, we employ a siamese structure for training unpaired data, allowing the network to learn important features from high-quality (HQ) images and preserving the structural details of low-quality (LQ) images, such as target highlight (texture enhancement) and background balance (consistent background contrast). To learn the texture-preserving representation and enhance the visual quality, we introduce two generators for learning a random pair (LQ and HQ). Our network structure is similar to the work in [24] which is used for face hallucination. Instead of relying on faulty paired labeling and the same identity, our random input pair aims to learn a common and salient distribution between LQ and HQ medical images, as these prerequisites are not necessary for our input pair. The HQ image can be treated as prior knowledge to guide the enhancement of the LQ image. Overall, our model combines the siamese structure with a GAN, using shared weights to produce high-quality enhancement results in an adversarial manner, which ensures the robustness of the model to texture blur, structure weakening, and background noise.

Our contributions can be summarized as follows:

- The mechanism of a Siamese-based structure with dual inputs, low-quality (LQ) images, and high-quality (HQ) images, allows the network to learn salient features from the HQ images while maintaining structural information in the LQ images.
- The proposed SSP-Net can generate high-quality enhanced results in an adversarial manner, which ensures the robustness of SSP-Net to blurred textures, weakened structure, and background noise.
- Sufficient experiments comprehensively illustrate the performance in image enhancement of the proposed SSP-Net compared with other state-of-the-art (SOTA) techniques both qualitatively and quantitatively. Moreover, the applicable verification in image segmentation also powerfully demonstrates the applicability of our SSP-Net.

The rest of the paper is organized as follows: Section II presents a review of image enhancement tasks. Section III shows the model of our SSP-Net framework and the specific structure of Siamese network. Section IV presents extensive experimental simulations, as well as a detailed discussion of the results. Finally, Section V draws conclusions.

2 Related Work

In recent decades, many excellent methods have been proposed for various image enhancement tasks. He et al. [4] proposed a simple but effective image prior-dark channel prior to removing haze from a single input image. Using this prior to the haze imaging model, researchers can directly estimate the thickness of the haze and recover a high-quality haze-free image. Similar with many conventional methods, however, this method necessitates extensive data training and manual parameter tweaking, making it challenging to apply in practical situations. [25] presented an explicit image filter, which is effective in detail enhancement and fog removal. Analogously, a 2D/3D symmetric filter was proposed in [26] solve the problem of automatic blood vessel detection, which had a strong response to vascular features under typical imaging conditions. However, most of these methods usually treat foreground and background indiscriminately, resulting in poor fidelity of image structure and loss of a lot of detailed information.

Actually, many research topics [27]–[29] have made great breakthroughs based on deep learning. By examining the characteristics of rain maps, Liu et al. [30] suggested a straightforward and efficient method for image rain removal based on unpaired learning. The semi-supervised learning component and the knowledge distillation component make up the majority of the algorithm. The estimation and image reconstruction of the rain maps are carried out by the semi-supervised portion using the established layer separation

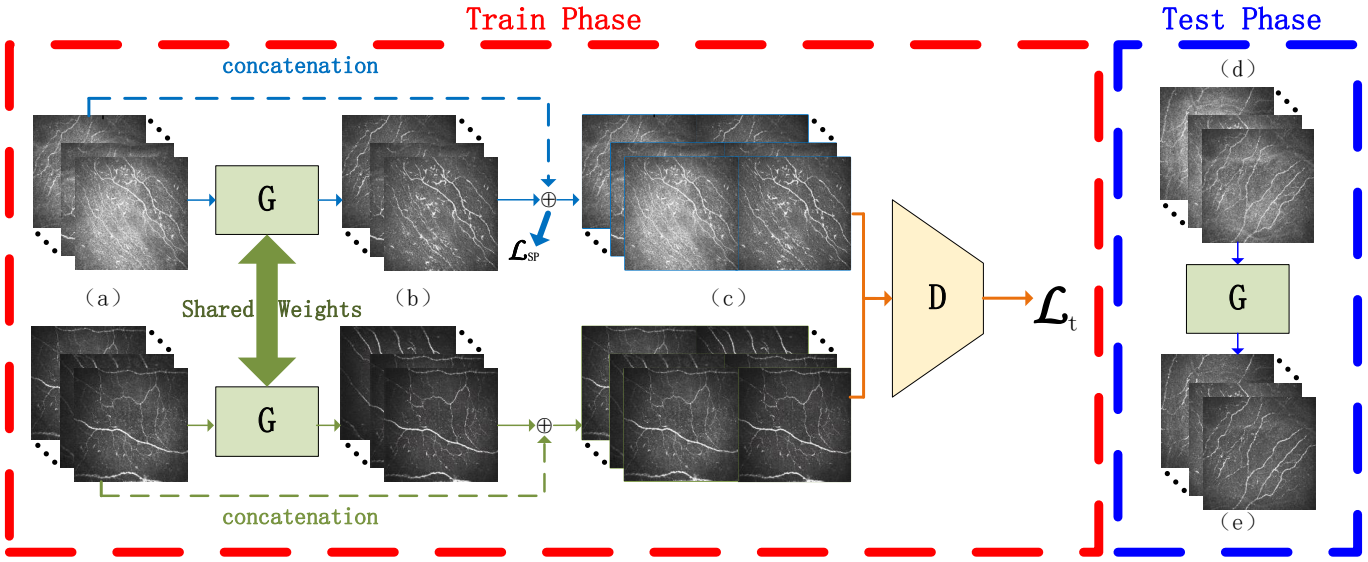

Fig. 1. The framework of the proposed SSP-Net has dual inputs (a),i.e., low-quality (LQ) images and high-quality (HQ) images, and adopts two generators with shared weights to jointly enhance the image quality (b) both in target highlight and background balance. In addition, (c) is the concatenation operation, (d) and (e) are the low-quality images and the enhanced results in the test phase, respectively.

principle. A rain direction regularizer is created to limit the rain estimation network in the semi-supervised learning phase. Noise2Noise (N2N) [31] was proposed to replace the clean target, which used independent noisy realization of the same image. Experimental results showed that networks trained in this manner can achieve the performance of those trained with clean targets. However, there is an obvious shortcoming to this approach: N2N training requires the availability of pairs of noisy images. Krull et al. [32] proposed a novel training scheme that overcame this limitation in the paper. The work in [33] presented a two-stage scheme, including self-supervised learning and knowledge distillation, to learn a blind image denoising network from an unpaired set of clean and noisy images. However, real noise can be more complex than pixel-independent heteroscedastic noise which they used in their experiments. Yang et al. [34] proposed a new deep learning-based method for micro-optical image enhancement that learns from both paired and unpaired data, while performing end-to-end contrast enhancement and noise reduction. Specifically, multi-level content loss is used on paired synthetic data to train enhancers for better detail recovery.

The adversarial learning mode of GAN [35] structure was applied to unpaired learning, among which CycleGAN [16] model brought about a great breakthrough. A framework was proposed to capture special characteristics of one image domain and figure out how these characteristics could be translated into the other image domain. Although achieving impressive results in image enhancement, the CycleGAN method is limited by the separated learning modules to realize the datasets. Moreover, the image quality is hard to keep only by restraining the background. Structure and illumination constrained GAN (StillGAN), a novel generalized bi-directional GAN that Ma et al. [36] devised, was intended to improve the quality of medical images. In order to learn global features and local details, StillGAN treats low-quality and high-quality images as two separate domains and introduces the local structure and lighting constraints. [37] presented a novel unsupervised shimmer image enhancement network,named LE-GAN. It was based on generative adversarial networks and trained with unpaired shimmer/normal light images, to address the issues of color bias and overexposure in shimmer image improvement. In order to solve the issues of noise and color bias and to boost the network's feature extraction, an illumination-aware attention module specifically was constructed. To address the overexposure issue, the author also put forth a novel identity-invariant loss, which enables the network to learn adaptive image quality improvement in low-light conditions. Furthermore, unpaired image enhancement technology is also applied to multi-mode image processing [38], [39]. [39] presented a novel method for cross-modality image synthesis by training the unpaired data, which enhanced the quality of the synthesized images with their generator. Although these techniques improve medical images somewhat, keeping texture detail and harmonizing the background pose difficulties.

3 SSP-Net Framework

In this paper, we propose a Siamese-based structure-preserving network, named SSP-Net, for corneal confocal microscopy image enhancement to handle the challenge of the deficiency of the paired (LQ) images and the HQ images. In this section, we introduce the details of our proposed SSP-Net, including an overview, of the siamese-based generative network structure and loss function.

3.1 Overview

In this section, we introduce the overall framework structure of SSP-Net. Then, the high and low-quality image contrast details and enhanced segmentation results are described.

The framework of the proposed SSP-Net is shown in Fig. 1, where the unpaired LQ images I^{T_1} and the HQ

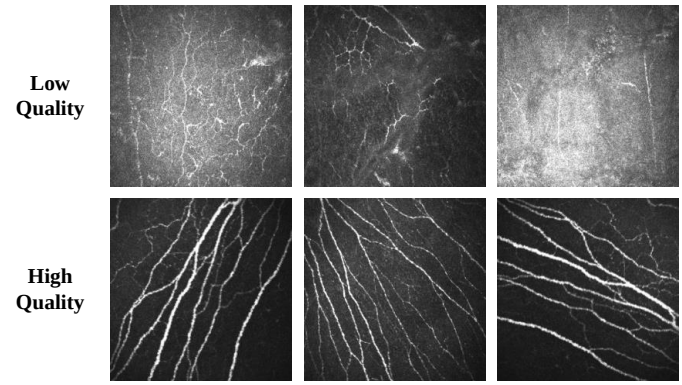

Fig. 2. Examples of unpaired LQ images and HQ images of corneal confocal microscopy, where the first and the second LQ images represent the challenges of heterogeneous illumination distribution and blurred texture detail, respectively. While the third LQ image is incorporated with the two aforementioned degradations, which is a more common phenomenon in LQ images.

images I^{T_2} are simultaneously used as input into a Siamese-based GAN and the corresponding enhanced outputs are discriminated by a discriminator with the original LQ and HQ input in an adversarial manner.

Fig. 2 shows an example of three unpaired sets of LQ and HQ images, where the first row is LQ and the second row is HQ. The first and second LQ images represented the challenges of uneven lighting distribution and blurred texture detail, respectively, while the third LQ image contained both degradations. In contrast to these LQ images, the second row of HQ images is visually superior, with uniform illumination and clear structural detail, which is critical for many medical imaging applications: segmentation, computer-aided diagnosis, etc. Therefore, recovering HQ images from LQ images is the focus of modern medical research. Fig. 3 shows the comparison between a group of LQ images and the enhanced images, where (a) is ground truth and (b) is the original image. The enhanced image is displayed in (e). We can clearly observe that the enhanced image has more uniform illumination and obtains clearer detailed structure information. (c) is the segmentation result of the original image. Compared with the segmentation result of the enhanced image (f), the quality of (f)is significantly improved, which ensures the performance of medical images in downstream tasks and improves the accuracy of medical treatment. (d) and (g) are the different results of (a) (c) and (a) (f) respectively, as we can see that (g) is significantly less than (d). Therefore, the research work on medical image enhancement is of great significance, and automated and reliable image enhancement methods need more attention.

3.2 Siamese Network Structure for Enhancement

In general, the siamese network structure guided by two input for self-supervised-like feature extraction leads to avoid the problem of complex concatenation and high coupling in training stage. Due to the different quality level, illumination and blur case, these problems would tend to accumulate throughout the entire unpaired images. This type of variability will bring negative influence to perceive separately LQ and HQ domain images. Thus, our SSP-Net adopts weight sharing generator G, which share the same structure and parameters. In the proposed network structure, we first design domain generators to determine the LQ and HQ images. The backbone of the generator is a U-Net-like structure, where there exist skip connections between the encoder and the decoder. Furthermore, the U-Net-like encoder-decoder module in the generator is not a symmetrical structure, where the backbone of the encoder module is a modified ResNet50 [40], where the fully-connected (FC) layer is removed and the max-pooling operation is replaced by a stride convolution for a more powerful texture preservation ability. Moreover, the decoder module is composed of five 3×3 convolution layers in case of local illumination variations. Since the difference between the feature of LQ and HQ images, the structure have high ability to extract the salient features by the weight sharing. The overall input images (LQ and HQ images) can be regarded as a pair of images with similar background information, but diverse appearance domains T_1 and T_2. As illustrated in Fig. 1, given a LQ image I^{T_1} and HQ images I^{T_2}, the primary goal of SSP-Net is to learn a map model as $T_1 \leftarrow T_2$, making the generated image appear similar to HQ image but retaining their salient information.

Besides, we introduce a discriminator D implemented by a PatchGAN [41] structure, which discriminates one image by its patch rather than the whole image, with a local illumination-sensitive constraint. The discriminator is composed of three convolution layers with the kernel size of 4×4, whose initialization follows a normal distribution with the parameter of $\sigma\,(0, 0.02)$. The shallow layer design lead to extract more discriminative information.

3.3 Loss Function

In this section, the loss function is introduced to ensure the enhanced images to obtain uniform structure-preserving texture details, which consists of three sub-parts, i.e., the adversarial loss and structure-preserving loss:

$$L_t = L_{GAN} + \alpha L_{SP}, \tag{1}$$

where L_{GAN} and L_{SP} correspond to the adversarial loss and the structure-preserving loss. α is the trade-off parameters that control the illumination and texture in the enhanced image.

Basic Loss. Considering that the proposed SSP-Net is a Siamese-based GAN, the adversarial loss is defined as the basic loss function to ensure the generating ability of the generator module in structural preservation and the discriminatory ability of the discriminator in detail fidelity. The basic GAN loss is formulated as follows:

$$\begin{aligned} L_{GAN} &= \min_G \max_D L(G, D) \\ &= \lambda_1 E_{I^{\hat{T}_1}}[\log D(I^{\hat{T}_1}) + \log[1 - D(I^{\hat{T}_1})]] \\ &+ \lambda_2 E_{I^{\hat{T}_2}}[\log D(I^{\hat{T}_2}) + \log[1 - D(I^{\hat{T}_2})]], \end{aligned} \tag{2}$$

where $I^{\hat{T}_1} = G(I^{T_1})$ is a loss sample of the generator, $I^{\hat{T}_2} = G(I^{T_2})$ is a loss sample of the Discriminator, λ_1 and λ_2 are the trade-off parameter to guarantee that the generator is able to retain key information in the original

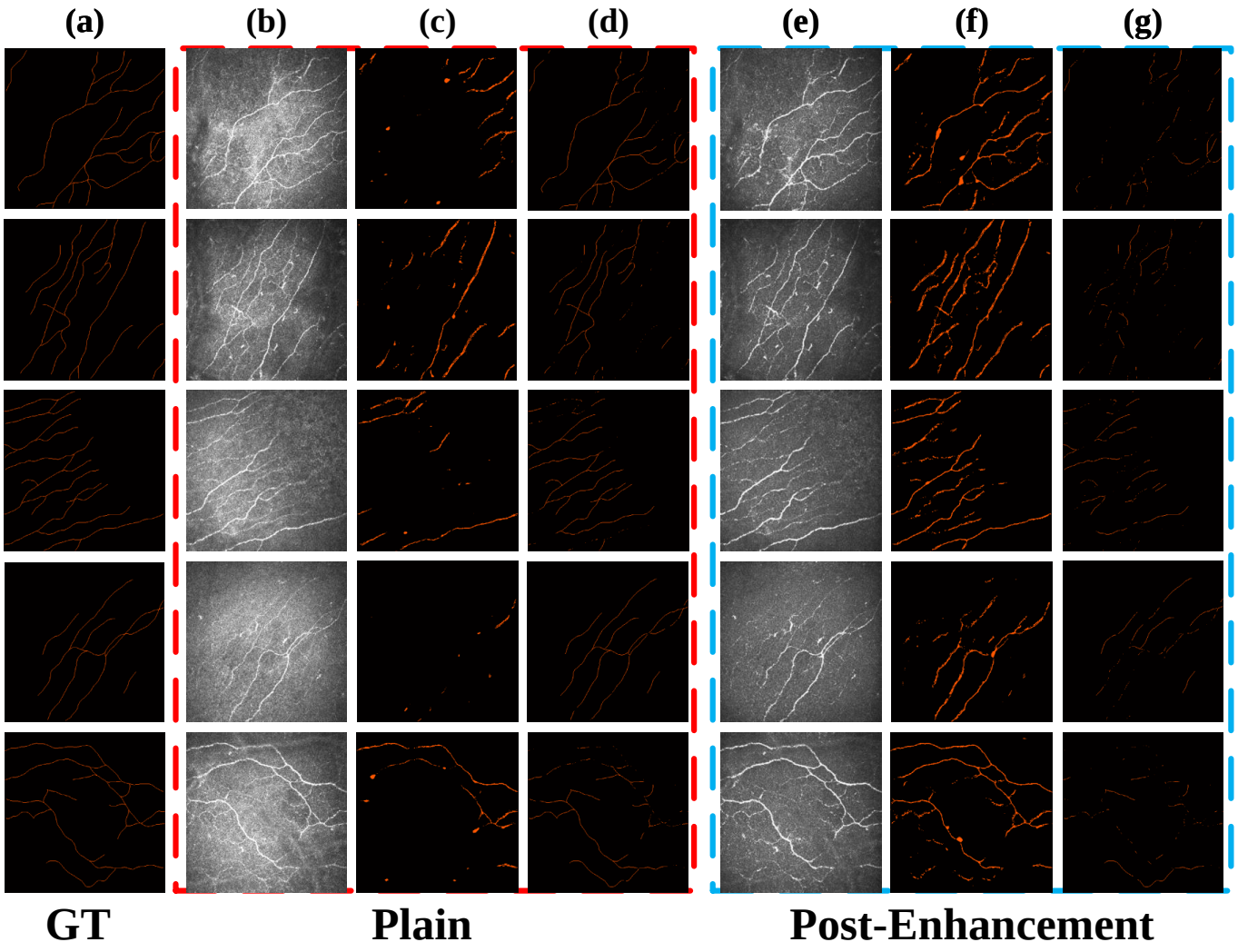

Fig. 3. Illustrations of the facilitation of the enhanced image for the segmentation task, where (a) is the ground truth labels, (b)-(d) are the original LQ images, the segmented results of LQ images, and the residual maps between ground truth labels and segmented results, (e)-(g) are those of the images enhanced by the proposed SSP-Net.

image while incorporating high-quality features, which is specially designed for the enhancement task under unpaired data.

Structure-preserving Loss. The structure-preserving loss has two parts, including illumination and texture information. The uniformity of the background illumination is a fundamental and essential requirement for high-quality images causing that variable illumination can be challenging for subsequent treatment and physician diagnosis. Therefore, a structure-preserving loss function is used to regularize the illumination of the entire image from [36]. The formulation of illumination uniformity loss is shown as follows:

$$L_{IU} = E_{I^{\hat{T}_1}}\left[E_g\left[\nabla\left\{E_l^P[G(I^{T_1})] - E_g[G(I^{T_1})]\right\}\right]\right], \quad (3)$$

where E_g and E_l^P denote the global and the local average illumination with the patch size of $p \times p$, respectively. $\nabla(\cdot)$ is a upsampling operator implemented by bicubic interpolation to address the scale discrepancy. Illumination uniformity loss focuses on balancing the illumination of the whole image, which may result in some unremarkable textures being homogenised into the surrounding background.

Therefore, a texture-aware loss function should be designed to ensure the enhanced result of highlighting textural details while maintaining uniform illumination. It is common sense to apply an MSE-based formulation to ensure the similarity of the pixel-level textural distribution between the source image and the enhanced image. However, the textural information from low-quality images can sometimes be blurred and low-contrast, which leads to poor performance in textural fidelity by the contrast information. We follow the setting in [36] with the following mathematical definition:

$$\begin{aligned} L_{TE} = {} & E_{I^{T_1}}\left[1 - \frac{1}{k}\sum_{l=1}^{K}\frac{2\Theta_{[I_l^{T_1}]}\cdot\Theta_{[G(I^{T_1})_l]} + C_2}{\left[\Theta_{[I_l^{T_1}]}\right]^2 + \left[\Theta_{[G(I^{T_1})_l]}\right]^2 + C_2}\right] \\ & + E_{I^{T_2}}\left[1 - \frac{1}{k}\sum_{l=1}^{K}\frac{2\Theta_{[I_l^{T_2}]}\cdot\Theta_{[G(I^{T_2})_l]} + C_2}{\left[\Theta_{[I_l^{T_2}]}\right]^2 + \left[\Theta_{[G(I^{T_2})_l]}\right]^2 + C_2}\right], \end{aligned} \quad (4)$$

where $I_l^{T_1}$, $I_l^{T_2}$, $G(I^{T_1})_l$, and $G(I^{T_2})_l$ denote the i^{th} local

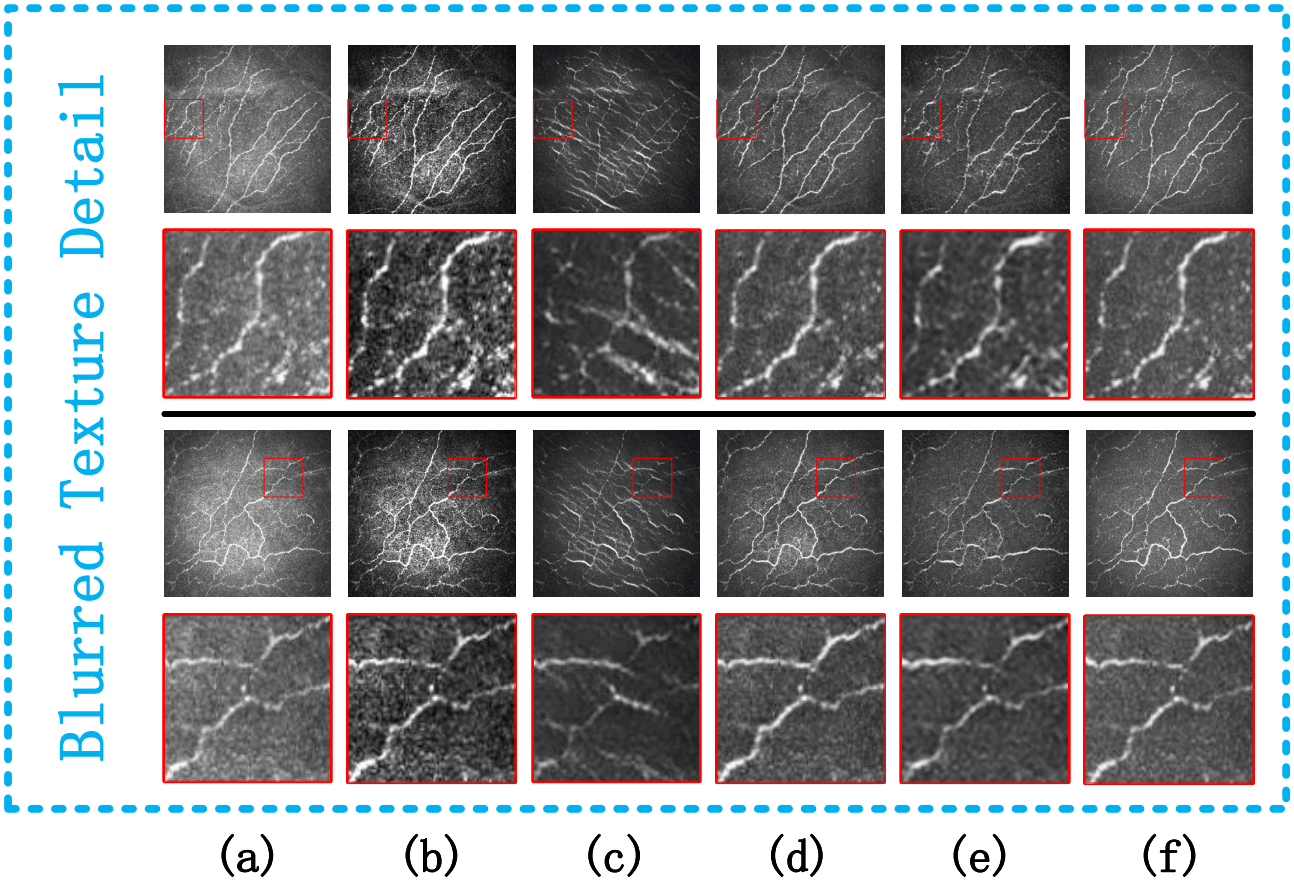

Fig. 4. Visual comparison of the enhanced results of the original image from the CORN-2 dataset with blurred texture detail, where the second row is the zoom-in operation from the first row. The top two groups of the original image are the examples with blurred texture detail, while the rests are the examples of those with heterogeneous illumination distribution. Among them, (a) is the original image, (b) - (e) is the contrast method, including the traditional method DCP (b), two deep learning-based methods MSG-Net (c) and EnlightenGAN (d), and the unpaired learning construction method StillGAN (e). The last (f) is our SSP-net.

TABLE 1
Average values and standard deviations of four selected no-reference metrics for the test set on CORN-2, where origin represents the low quality images in the test set and the best values are marked with bold.

	Origin	DCP [4]	MSG-Net [42]	EnlightenGAN [43]	StillGAN [36]	SSP-Net
Entropy ↑	4.7433±0.0925	5.5652±0.0411	6.5454±0.0432	6.6772±0.2914	6.5488±0.0837	**6.6951±0.1723**
AvG ↑	5.1374±1.3022	7.3361±1.4815	6.5834±0.9673	7.3547±1.0276	5.9066±0.5373	**8.5481±1.1588**
NIQE ↓	6.9277±0.4526	4.8018±0.4499	6.5568±0.2993	4.2323±0.3274	8.2876±0.4323	**3.9778±0.3629**
PIQE ↓	9.7944±1.8574	11.4941±2.4798	7.3329±2.1681	6.8681±1.6882	21.8277±3.5749	**5.3141±1.2229**

window in LQ $\rightarrow I^{T_1}$, HQ $\rightarrow I^{T_2}$, and their corresponding enhanced result. $\Theta_{[I_l^{T_1}]}$, and $\Theta_{[G(I^{T_1})_l]}$ are the covariance of the set of $I_l^{T_1}$ and $G(I^{T_1})_l$, respectively. C_2 is a positive constant to avoid the numerical error and E_m, E_n is the energy function.

4 Experiment

In this section, we compare the enhancement performance on CORN-2 dataset [36] with four SOTA techniques. Moreover, the image segmentation task is evaluated on the five compared enhancement algorithms to further explore the latent applicable potential.

4.1 Dataset and Comparison Methods

Dataset: A confocal microscopy-based dataset, namely CORN-2 [36], is selected within the total of 688 confocal images equipped with manual labels at the size of 384×384, where 628 images are divided into the training set with 340 low-quality medical images and 288 high-quality medical images, while the remaining 60 low-quality images are reserved for testing.

Comparison Methods. In comparison with the proposed SSP-Net, which consist of one traditional method, i.e., DCP [4], and three convolution-based approaches, i.e., MSG-Net [42], EnlightenGAN [43], and StillGAN [36], whose configuration are consistent with the public available implementation without any amendments. Furthermore, the

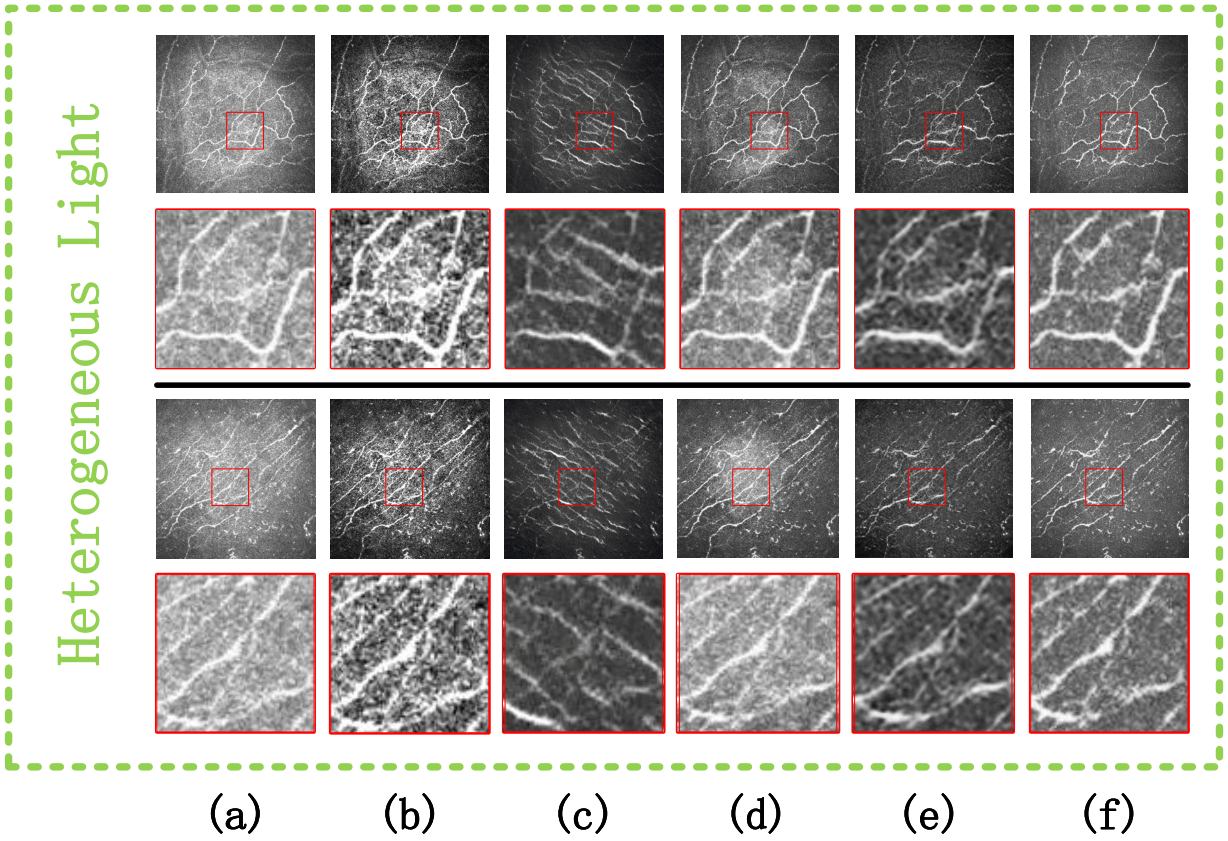

Fig. 5. Visual comparison of the enhanced results of the original image from the CORN-2 dataset with heterogeneous illumination distribution, where the second row is the zoom-in operation from the first row. The top two groups of the original image are the examples with blurred texture detail, while the rests are the examples of those with heterogeneous illumination distribution. Among them, (a) is the original image, (b) - (e) is the contrast method, including the traditional method DCP (b), two deep learning-based methods MSG-Net (c) and EnlightenGAN (d), and the unpaired learning construction method StillGAN (e). The last (f) is our SSP-net.

aforementioned learning-based methods are all retrained in the same dataset with the proposed SSP-Net.

4.2 Experiment Settings

All the training images are initially resized by 384×384 with a series of data augmentation strategies [36], e.g., random flipping, and color space transformation. Adam optimizer [44] is applied to train the network with momentum terms $(0.5, 0.9)$ for generator and $(0.9, 0.999)$ for discriminator, whose learning rate is initialized as 3×10^{-5} and linearly warmed up to 1×10^{-4}. When reaching 500 iterations, the learning rate will be halved every 15 epochs. The batch size and the related parameters λ_1, λ_2, α, β are set as 4, 0.6, 0.4, and 5.1. All the experiments are implemented in PyTorch [45] on two RTX2080TI GPUs.

4.2.1 Evaluation Metrics

Four no-reference evaluation metrics are provided to comprehensively illustrate the superiority of our proposed SSP-Net, which are entropy, average gradient (AvG) [46], natural image quality evaluator (NIQE) [47], and perception-based image quality evaluator (PIQE) [48]. Entropy and AvG measure the amount of information and detail textures contained in the enhanced images, respectively, where a larger value indicates a better performance. Nevertheless, NIQE and PIQE correspond to the natural quality and perception-based feature quality modeled by the multivariate Gaussian model, where a lower score means better reconstructed quality. We employ the area under the ROC curve (AUC) [49], accuracy (ACC) [49], sensitivity (SEN) [50], specificity (SPE) [50], false discovery rate (FDR) [51], dice coefficient (Dice) [52], G-Mean score (G-Mean) [53], and Kappa coefficient [54] as the evaluation metrics for segmentation task. Higher value denotes better performance.

4.3 Results and Analysis

In this section, the five aforementioned techniques are applied to CORN-2 dataset to qualitatively and quantitatively evaluate the enhancement performance.

4.3.1 Qualitative Analysis

The visual contrast performance between our SSP-Net and several state-of-the-art enhancement methods is shown in Fig. 4 and Fig. 5. Due to the heterogeneous illumination distribution in the backdrop of the original image, we can see that the methods DCP (b) and EnlightenGAN (d) provide an unequal distribution of brightness over the entire image. This issue, which causes the structural information to deteriorate in zoomed-in photos of the DCP, is very severe and makes medical diagnosis extremely difficult. The improved image produced by the MSG-Net (c) approach

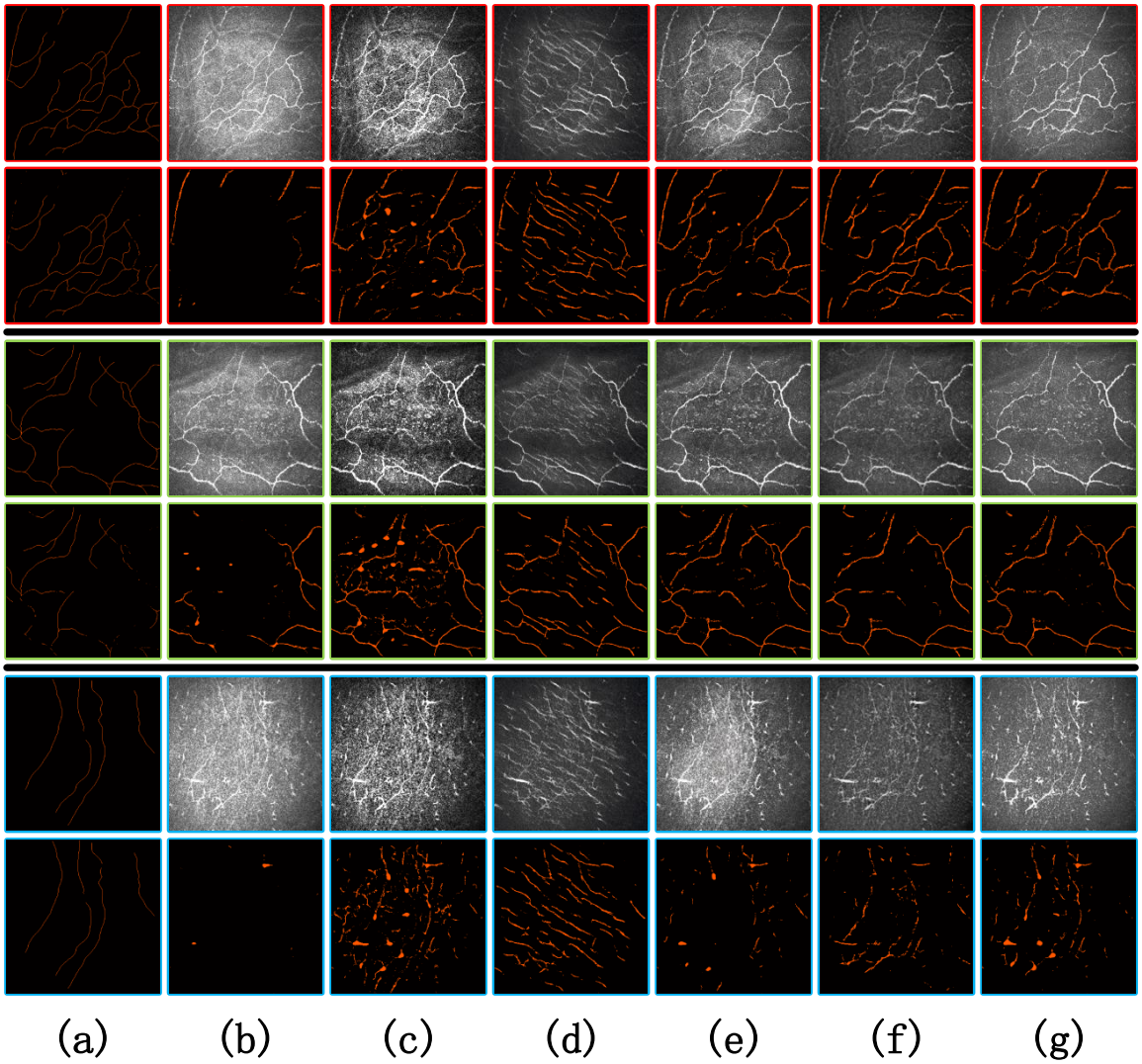

Fig. 6. Comparison of three groups of enhanced image segmentation results. In the first column, the first image of each group represents ground truth while the second image is the difference between the ground truth and the original image. (b) - (g) in the first row of each group are the original image (b), enhanced images by DCP(c), MSG-Net(d), EnlightenGAN(e), StillGAN(f) and our SSP-Net(g) respectively. (b) - (g) of the second row of each group are the segmentation results of the enhanced image corresponding to the first row.

displays visible distortion because of the weak texture structure. Although the StillGAN (e) technique nearly duplicated the results of our SSP-Net in terms of the issue of consistent backdrop, its foreground was still not sufficiently apparent, leading to the loss of some texture features. In contrast, our approach produces the best visual result, which not only keeps the backdrop color constant and makes the boosted image brightness uniform but also draws attention to the texture structure of target. For monitoring and diagnosis in medicine, it is extremely important.

4.3.2 Quantitative Analysis

As shown in Table 2, compared with other enhancement methods, our SSP-NET model achieved the best results on all metrics, especially AvG and PIQE, which made breakthrough progress. The optimal AvG and Entropy result demonstrates how well our structure-maintaining network achieves the goal of picture enhancement, extracts the intricate details of the original, and highlights the performance, making the enhanced image very faithful to the texture structure. Our values of NIQE and PIQE are relatively low compared to the StillGAN approach with non-paired learning model, which guarantees the SSP-Net reconstruction performance.

4.3.3 Ablation Study

One benefit of our approach is the proposed loss function. In order to confirm the benefits of the combination, we choose four elements: L_{GAN}, $L_{GAN} + L_{IU}$, $L_{GAN} + L_{TE}$, and $L_{GAN} + L_{SP}$ (i.e., our method). Table. 3 lists the metrics we chose for medical image quality enhancement. As can be observed in the table, our lossy combination technique performs better in comparison experiment, which supports the combination in Equ. (1)- (4). The improvement of the suggested partial loss combination is further demonstrated by the fact that the mean values of $L_{GAN} + L_{IU}$, $L_{GAN} + L_{TE}$, and $L_{GAN} + L_{SP}$ are higher than L_{GAN}.

TABLE 2
Average values and standard deviations for the segmentation results from the aforementioned enhanced images of eight selected metrics, where origin represents the segmentation results of the low quality images and the best values are marked with bold.

	Origin	DCP [4]	MSG-Net [42]	EnlightenGAN [43]	StillGAN [36]	SSP-Net
AUC ↑	0.4944 ± 0.0333	0.5144 ± 0.0269	0.5072 ± 0.0215	0.5192 ± 0.0267	0.5044 ± 0.0281	**0.5673 ± 0.0259**
ACC ↑	0.9533 ± 0.0499	0.9554 ± 0.0104	0.9491 ± 0.0051	0.9598 ± 0.0113	0.9623 ± 0.0096	**0.9749 ± 0.0097**
SEN ↑	0.4213 ± 0.1284	0.7089 ± 0.0874	0.4411 ± 0.1673	0.6713 ± 0.0726	0.7853 ± 0.0967	**0.8214 ± 0.0645**
SPE ↑	0.9575 ± 0.0511	0.9628 ± 0.0099	0.9563 ± 0.0046	0.9672 ± 0.0108	0.9698 ± 0.0092	**0.9724 ± 0.0093**
FDR ↑	0.9736 ± 0.0202	0.9745 ± 0.0109	0.9729 ± 0.0108	0.9743 ± 0.0116	0.9756 ± 0.0116	**0.9794 ± 0.0123**
Dice ↑	0.5414 ± 0.0201	0.6335 ± 0.0162	0.5125 ± 0.0163	0.6018 ± 0.0165	0.6937 ± 0.0159	**0.7736 ± 0.0158**
G-mean ↑	0.6283 ± 0.1377	0.8301 ± 0.0591	0.6477 ± 0.0408	0.7493 ± 0.0791	0.8207 ± 0.0586	**0.8847 ± 0.0640**
Kappa ↑	0.5283 ± 0.0177	0.6152 ± 0.0133	0.4953 ± 0.0131	0.5801 ± 0.0143	0.6733 ± 0.0136	**0.7254 ± 0.0138**

TABLE 3
Ablation study with different combinations of loss functions and the best results are bolded in red.

	Entropy ↑	AvG ↑	NIQE ↓	PIQE ↓
L_{GAN}	6.5643±0.2019	6.7084±1.8462	6.2351±0.7265	7.0518±1.4356
$L_{GAN} + L_{IU}$	6.6297±0.1645	7.8463±1.6752	4.9216±0.5738	6.1458±1.3567
$L_{GAN} + L_{TE}$	6.6518±0.1864	8.1941±1.1579	4.2387±0.4532	5.6039±1.3051
Our	**6.6951±0.1723**	**8.5481±1.1588**	**3.9778±0.3629**	**5.3141±1.2229**

4.4 Application Verification

To further explore the application of the medical image enhancement methods, a segmentation task is assigned to segment the enhanced images with a pre-trained segmentation network, called CS-Net [55], which was trained on the high-quality confocal database with manually labeled fibers. In this section, the superiority of the proposed SSP-Net will be illustrated subjectively and objectively.

4.4.1 Qualitative Analysis

Fig.6 shows three groups comparison of image segmentation results after enhancement. The groups are separated by thick solid lines, and each group images are selected by a different color frame to show the distinction. Among them, the first image in the first column of each group is the ground truth, and the second image is the difference between the ground truth and the original image segmentation result. (b) - (g) in the first row of each group is the original image (b), the enhanced image (c) - (f) of four comparison methods and the enhanced image (g) of the SSP-Net method in this paper. (b) - (g) of the second row of each group are the segmentation results of the enhanced image corresponding to the first row.

Overall, the segmentation results of all enhanced images are better than those of the original image. The segmentation result of the MSG-Net (d) method is the worst and has a lot of distortion because the enhanced image is distorted. Due to the problem of uneven brightness after enhancement, the segmentation results of the DCP (c) enhancement method have a lot of prominence. In contrast, the segmentation of EnlightenGAN (e) and StillGAN (f) enhanced images is relatively good, but there is still a problem that the details cannot be segmented due to insufficient structural fidelity, which is more obvious in the first group of images. The segmentation result of our SSP-Net method is the most similar to the ground truth and the best among all methods. All the details in the original image are basically retained and the texture structure is enhanced. which greatly promotes the performance of downstream tasks of medical images.

4.4.2 Quantitative Analysis

Eight metrics are selected to numerically make a comparison among the five enhancement solutions, which are area under the ROC curve (AUC), accuracy (ACC) [49], sensitivity (SEN), specificity (SPE) [50], false discovery rate (FDR) [51], Dice similarity coefficient (Dice), G-mean [56], and Kappa [54], where a higher score means a more accurate segmentation result.

For the enhanced image segmentation results, we can see the mean and standard deviation of the 8 indicators as shown in Table 2. Our SSP-Net achieves the best segmentation results on all metrics, especially AUC, Dice, and G-mean. Compared with EnlightenGAN and StillGAN methods, we can see that our SSP-Net outperforms the two methods by 5% and 6% in AUC, 17%, and 8% in Dice, and 14% and 8% in G-mean respectively. The significant performance of nerve fiber segmentation is more beneficial to the observation and diagnosis of neurological diseases.

5 Conclusion

The Siamese-based structure-preserving generative adversarial network is proposed in the paper for improving unpaired medical images. Low-quality (LQ) images and high-quality (HQ) images are used as dual inputs to the Siamese structure, allowing the network to learn salient features from the HQ images while maintaining the structural information in the LQ images. The method generates high-quality enhancement results in an adversarial manner, ensuring the robustness of the algorithm against blurred textures, weakened structures, and background noise. Experimental results on the simulated dataset demonstrate

that our method outperformed state-of-the-art methods in terms of quality and quantity. In future work, we will consider further style generative adversarial network as the network architecture to improve performance.

REFERENCES

[1] V. R. P. Borges, M. C. F. d. Oliveira, T. G. Silva, A. A. H. Vieira, and B. Hamann, "Region growing for segmenting green microalgae images," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 1, pp. 257–270, 2018.

[2] A. Dehzangi, K. Paliwal, J. Lyons, A. Sharma, and A. Sattar, "A segmentation-based method to extract structural and evolutionary features for protein fold recognition," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 11, no. 3, pp. 510–519, 2014.

[3] X. Zhou, X. Xu, W. Liang, Z. Zeng, and Z. Yan, "Deep learning enhanced multi-target detection for end-edge-cloud surveillance in smart iot," *IEEE Internet of Things Journal*, 2021.

[4] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[5] T. Dehghani, M. Naghibzadeh, and J. Sadri, "Enhancement of protein β -sheet topology prediction using maximum weight disjoint path cover," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 6, pp. 1936–1947, 2019.

[6] Y. Feng, F. Yang, X. Zhou, Y. Guo, F. Tang, F. Ren, J. Guo, and S. Ji, "A deep learning approach for targeted contrast-enhanced ultrasound based prostate cancer detection," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 6, pp. 1794–1801, 2019.

[7] X. Zhou, X. Yang, J. Ma, I. Kevin, and K. Wang, "Energy efficient smart routing based on link correlation mining for wireless edge computing in iot," *IEEE Internet of Things Journal*, 2021.

[8] W. Liang, Y. Hu, X. Zhou, Y. Pan, I. Kevin, and K. Wang, "Variational few-shot learning for microservice-oriented intrusion detection in distributed industrial iot," *IEEE Transactions on Industrial Informatics*, 2021.

[9] D. G. Anderson, J. A. Burdick, and R. Langer, "Smart biomaterials," *Science*, vol. 305, no. 5692, pp. 1923–1924, 2004.

[10] Q. Ren, Y. Wang, Y. Li, and S. Qi, "Sophisticated electromagnetic forward scattering solver via deep learning," 2021.

[11] M. Panwar, A. Gautam, D. Biswas, and A. Acharyya, "Pp-net: A deep learning framework for ppg-based blood pressure and heart rate estimation," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10 000–10 011, 2020.

[12] 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.

[13] Z. Shen, H. Fu, J. Shen, and L. Shao, "Modeling and enhancing low-quality retinal fundus images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 996–1006, 2021.

[14] H. Liu, Y. Zhuang, E. Song, X. Xu, and C.-C. Hung, "A bidirectional multilayer contrastive adaptation network with anatomical structure preservation for unpaired cross-modality medical image segmentation," *Computers in Biology and Medicine*, vol. 149, p. 105964, 2022.

[15] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6306–6314.

[16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[17] A. Abu-Srhan, I. Almallahi, M. A. Abushariah, W. Mahafza, and O. S. Al-Kadi, "Paired-unpaired unsupervised attention guided gan with transfer learning for bidirectional brain mr-ct synthesis," *Computers in Biology and Medicine*, vol. 136, p. 104763, 2021.

[18] Y. Liang, D. Lee, Y. Li, and B.-S. Shin, "Unpaired medical image colorization using generative adversarial network," *Multimedia Tools and Applications*, vol. 81, no. 19, pp. 26 669–26 683, 2022.

[19] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.

[20] Y. Li, X. Tian, X. Shen, and D. Tao, "Classification and representation joint learning via deep networks." in *International Joint Conference on Artificial Intelligence*, vol. 2017, 2017, p. 67.

[21] G. Koch, R. Zemel, R. Salakhutdinov *et al.*, "Siamese neural networks for one-shot image recognition," in *International Conference on Machine Learning*, vol. 2. Lille, 2015.

[22] X. Li, N. Dong, J. Huang, L. Zhuo, and J. Li, "A discriminative self-attention cycle gan for face super-resolution and recognition," *IET Image Processing*, vol. 15, no. 11, pp. 2614–2628, 2021.

[23] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *European conference on computer vision*. Springer, 2016, pp. 850–865.

[24] C.-C. Hsu, C.-W. Lin, W.-T. Su, and G. Cheung, "Sigan: Siamese generative adversarial network for identity-preserving face hallucination," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 6225–6236, 2019.

[25] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2012.

[26] Y. Zhao, Y. Zheng, Y. Liu, Y. Zhao, L. Luo, S. Yang, T. Na, Y. Wang, and J. Liu, "Automatic 2-d/3-d vessel enhancement in multiple modality images using a weighted symmetry filter," *IEEE transactions on medical imaging*, vol. 37, no. 2, pp. 438–450, 2017.

[27] X. Zhou, W. Liang, I. Kevin, K. Wang, and L. T. Yang, "Deep correlation mining based on hierarchical hybrid networks for heterogeneous big data recommendations," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 1, pp. 171–178, 2020.

[28] X. Zhou, X. Xu, W. Liang, Z. Zeng, S. Shimizu, L. T. Yang, and Q. Jin, "Intelligent small object detection based on digital twinning for smart manufacturing in industrial cps," *IEEE Transactions on Industrial Informatics*, 2021.

[29] X. Zhou, W. Liang, W. Li, K. Yan, S. Shimizu, I. Kevin, and K. Wang, "Hierarchical adversarial attacks against graph-neural-network-based iot network intrusion detection system," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9310–9319, 2021.

[30] Y. Liu, Z. Yue, J. Pan, and Z. Su, "Unpaired learning for deep image deraining with rain direction regularizer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4753–4761.

[31] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[32] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2129–2137.

[33] N. Moran, D. Schmidt, Y. Zhong, and P. Coady, "Noisier2noise: Learning to denoise from unpaired noisy data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 064–12 072.

[34] Q. Yang, Y. Wu, D. Cao, M. Luo, and T. Wei, "A lowlight image enhancement method learning from both paired and unpaired data by adversarial training," *Neurocomputing*, vol. 433, pp. 83–95, 2021.

[35] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[36] Y. Ma, J. Liu, Y. Liu, H. Fu, Y. Hu, J. Cheng, H. Qi, Y. Wu, J. Zhang, and Y. Zhao, "Structure and illumination constrained gan for medical image enhancement," *IEEE Transactions on Medical Imaging*, 2021.

[37] Y. Fu, Y. Hong, L. Chen, and S. You, "Le-gan: unsupervised low-light image enhancement network using attention module and identity invariant loss," *Knowledge-Based Systems*, vol. 240, p. 108010, 2022.

[38] N. Arya and S. Saha, "Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1–1, 2020.

[39] L. Xiang, Y. Li, W. Lin, Q. Wang, and D. Shen, "Unpaired deep cross-modality synthesis with fast training," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 155–164.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for

image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[41] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
[42] H. Zhang and K. Dana, "Multi-style generative network for real-time transfer," in *Proceedings of the European Conference on Computer Vision Workshops*, 2018, pp. 0–0.
[43] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.
[44] R. Kumudham, G. D. Bharathi, and R. Priya, "An effective identification of microaneurysm in retinal eye using convnet with adam optimizer," in *2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, 2021, pp. 1–5.
[45] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.
[46] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," *Aquatic Procedia*, vol. 4, pp. 133–142, 2015.
[47] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.
[48] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *2015 Twenty First National Conference on Communications*. IEEE, 2015, pp. 1–6.
[49] J. Huang and C. X. Ling, "Using auc and accuracy in evaluating learning algorithms," *IEEE Transactions on knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.
[50] T. S. Genders, S. Spronk, T. Stijnen, E. W. Steyerberg, E. Lesaffre, and M. M. Hunink, "Methods for calculating sensitivity and specificity of clustered data: a tutorial," *Radiology*, vol. 265, no. 3, pp. 910–916, 2012.
[51] J. R. Chumbley and K. J. Friston, "False discovery rate revisited: Fdr and topological inference using gaussian random fields," *Neuroimage*, vol. 44, no. 1, pp. 62–70, 2009.
[52] D. Kang, S. Park, and J. Paik, "Sdban: Salient object detection using bilateral attention network with dice coefficient loss," *IEEE Access*, vol. 8, pp. 104 357–104 370, 2020.
[53] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro, and A. P. Braga, "Learning from imbalanced data sets with weighted cross-entropy function," *Neural processing letters*, vol. 50, no. 2, pp. 1937–1949, 2019.
[54] H. C. Kraemer, "Kappa coefficient," *Wiley StatsRef: statistics reference online*, pp. 1–4, 2014.
[55] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, and J. Liu, "Cs-net: channel and spatial attention network for curvilinear structure segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 721–730.
[56] J.-H. Ri, G. Tian, Y. Liu, W.-h. Xu, and J.-g. Lou, "Extreme learning machine with hybrid cost function of g-mean and probability for imbalance learning," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 9, pp. 2007–2020, 2020.

Guoxia Xu received the B.S. degree in information and computer science from Yancheng Teachers University, Jiangsu Yancheng, China in 2015, and the M.S. degree in computer science and technology from Hohai University, Nanjing, China in 2018. Now, he is pursing his Ph.D. degree in Department of Computer Science, Norwegian University of Science and Technology, Gjovik Norway. His research interest includes pattern recognition, image processing, and computer vision.

Hao Wang is an associate professor in the Department of Computer Science in Norwegian University of Science and Technology, Norway. He received his Ph.D. degree and B.Eng. degree, both in computer science and engineering, from South China University of Technology in 2006 and 2000, respectively. His research interests include big data analytics, industrial internet of things, high performance computing, and safety-critical systems. He has published 140+ papers in reputable international journals and conferences. He served as a TPC co-chair for IEEE CPSCom 2020, IEEE CIT 2017, ES 2017, and DataCom 2015, a senior TPC member for CIKM 2019, and reviewers/TPC members for many journals and conferences. He is the Chair for Sub-TC on Healthcare in IEEE IES Technical Committee on Industrial Informatics.

Marius Pedersen received the B.Sc. degree in computer engineering and the M.Sc. degree in media technology from the Gjøvik University College, and the Ph.D. degree in color imaging from the University of Oslo, in 2011. He is currently a Professor with the Norwegian University of Science and Technology, Gjøvik, Norway, and also the Head of the Norwegian Colour and Visual Computing Laboratory. His research interest includes print and image quality.

Meng Zhao received the B.S. degree in automation from Tianjin University, China, in 2010, and the M.S and Ph.D. degrees in control science and engineering from Tianjin University, China, in 2016. She completed a postdoctoral fellowship at NTNU, Norway, in 2020, where she successfully finished a research stay which was supported by the European Research Consortium for Informatics and Mathematics (ERCIM) 'Alain Bensoussan Fellowship Programme'. She is currently an associate professor with the School of Computer Science and Engineering, Tianjin University of Technology, China. Her research interests include medical image processing, medical/biomedical engineering, and machine learning/deep learning in medical informatics.

Hu Zhu received the B.S. degree in mathematics and applied mathematics from Huaibei Coal Industry Teachers College, Huaibei, China, in 2007, and the M.S. and Ph.D. degrees in computational mathematics and pattern recognition and intelligent systems from Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2013, respectively. In 2013, he joined the Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include pattern recognition, image processing, and computer vision.