Cycle Structure and Illumination Constrained GAN for Medical Image Enhancement

***

***

Abstract. The non-uniform illumination or imbalanced intensity in medical images brings challenges for automated screening, examination and diagnosis of diseases. Previously, CycleGAN was proposed to transform input images into enhanced ones without paired images. However, it did not consider many local details of the structures, which are essential for medical images. In this paper, we propose a Cycle Structure and Illumination constrained GAN (CSI-GAN), for medical image enhancement. Inspired by CycleGAN based on the global constraints of the adversarial loss and cycle consistency, the proposed CSI-GAN treats low and high quality images as those in two domains and computes local structure and illumination constraints for learning both overall characteristics and local details. To evaluate the effectiveness of CSI-GAN, we have conducted experiments over two medical image datasets: corneal confocal microscopy (CCM) and endoscopic images. The experimental results show that our method yields better performance than both conventional methods and other deep learning based methods. As a complementary output, we will release the CCM dataset to the public.

Keywords: Illumination · structural · enhancement

1 Introduction

High-quality images with adequate contrast and details are crucial for many medical imaging applications: e.g., segmentation [1] and computer-aided diagnosis [2]. However, medical images acquired using the same or different sensors usually have a large variation in quality - intensity inhomogeneity, noticeable blur and poor contrast, that are often inherited from the image acquisition process. Fig. 1 (a) and (c) illustrate two examples of low-quality images captured by confocal microscopy and endoscopy - where it is difficult to observe the complete structure of corneal nerve fibers and digestive tract respectively due to the imperfect focus and poor light condition. These obstacles pose significant challenges to many subsequent image analysis tasks, such as curvilinear structure segmentation [3] and lesion detection [4]. As a consequence, fully automated and reliable image quality enhancement approaches have long been deemed desirable.

Many image enhancement methods have been proposed such as histogram equalization (HE) [5], dark channel prior [6] and guided filtering [7]. However, these conventional methods usually enhance images uniformly, irrespective of

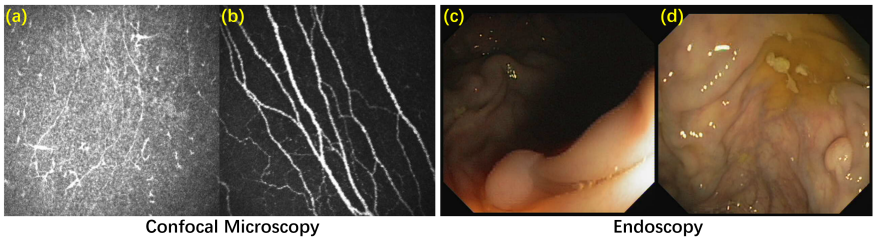

Fig. 1: Examples of low-/high-quality images captured by the same equipment. (a) and (b) corneal confocal microscopic images. (c) and (d) endoscopic images.

whether a given region is in the foreground or background [8, 9]. Deep learning has provided new avenues for image enhancement, nevertheless, a large portion of these methods [10–12] require aligned image pairs for training. In consequence, recently a few unpaired learning models have been proposed. Gatys et al. [13] proposed a neural transfer algorithm (NST) for unpaired image transformation. Zhang et al. [14] introduced a multi-style generative network (MSG-Net) to achieve real-time image style translation. Jiang et al. [15] proposed a EnlightenGAN with a global-local discriminator structure, a self-regularized perceptual loss fusion, and attention mechanism for low-light image enhancement. Nevertheless, one common limitation of these unpaired learning methods is that they often amplify noise in the dark background area and suffer from halo artifacts.

Cycle-consistent generative adversarial network (CycleGAN) [16] has an advantage of learning knowledge represented with typical images in one domain and transferring it to the other domain without paired images. However, CycleGAN mainly exploits global constraints on appearance and cycle-consistency, which is weak in learning local details. To address the weakness, two novel constraints including an illumination regularization and a structure loss are proposed in our new method, which we refer it to CSI-GAN for medical image enhancement. In our work, low- and high-quality images are treated as those in two different domains and high-quality images can be easily identified by clinicians, as shown in Fig. 1 (b) and (d). The main contributions of this paper are summarized as follows. (1) A novel CSI-GAN is proposed to improve low-quality medical images with better illumination conditions while well-preserving structure details. (2) The proposed method has undergone rigorous quantitative and qualitative evaluation using corneal confocal microscopy and endoscopic images in an unified manner. (3) As a complementary output, we release the CCM dataset (both poor and good quality image sets) online available to the public.

2 Proposed Method

2.1 Overall Architecture

Different from conditional adversarial network (cGAN) [17], CycleGAN [16] learns a suitable translation function between the source domain A and the

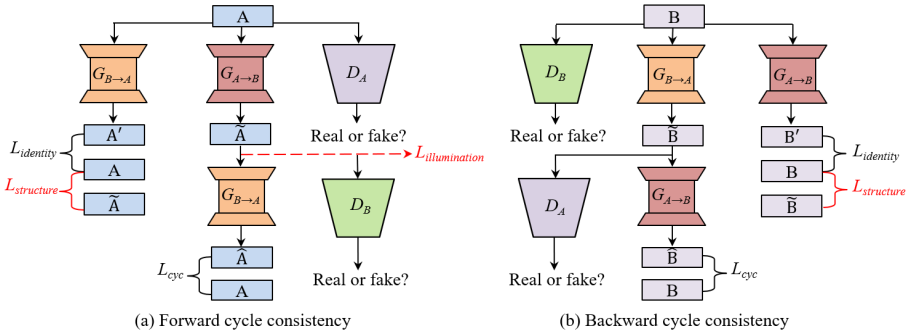

Fig. 2: CSI-GAN structure diagram. It comprises of two generator/discriminator pairs ($G_{A\to B}/D_B$, $G_{B\to A}/D_A$) and two types of cycle consistency (forward/backward cycle consistency). A and B refer to low-quality and high-quality image domain, respectively. L_{cyc} and $L_{identity}$ represent the cycle consistency term and the identity mapping loss. The proposed illumination regularization and structure loss are represented as $L_{illumination}$ and $L_{structure}$, respectively.

target domain B without the paired images in the training. In this paper, we assume that A and B are image domains with low and high quality images, respectively. CycleGAN adopts two generator/discriminator pairs ($G_{A\to B}/D_B$, $G_{B\to A}/D_A$), where $G_{A\to B}$ ($G_{B\to A}$) learns to translate an image from domain A (B) into domain B (A), and D_A (D_B) is trained to distinguish between real samples from domain A (B) and the translated images from domain B (A).

In order to prevent two generators from contradicting each other, the whole framework contains both forward and backward cycle consistency, as shown in Fig. 2. Each $a \in A$ is expected to be reconstructed as much as possible in forward cycle, which is represented as $a \to G_{A\to B}(a) \to G_{B\to A}(G_{A\to B}(a)) \approx a$. This holds for backward cycle as well: $b \to G_{B\to A}(b) \to G_{A\to B}(G_{B\to A}(b)) \approx b$. In addition, two generators are regularized as an identity mapping separately when real samples from A (B) are applied to $G_{B\to A}$ ($G_{A\to B}$), i.e., $G_{B\to A}(a) \approx a$ and $G_{A\to B}(b) \approx b$. The objective function of CycleGAN is defined as:

$$\begin{aligned} & L\left(G_{A\to B}, G_{B\to A}, D_A, D_B\right) \\ &= L_{GAN}\left(G_{A\to B}, D_B, A, B\right) + L_{GAN}\left(G_{B\to A}, D_A, B, A\right) \\ &+ \lambda_1 L_{cyc}(G_{A\to B}, G_{B\to A}, A, B) + \lambda_2 L_{identity}\left(G_{A\to B}, G_{B\to A}, A, B\right), \end{aligned} \tag{1}$$

where L_{GAN} denotes the adversarial loss; L_{cyc} and $L_{identity}$ represent the cycle consistency term and the identity mapping loss with weighted coefficients λ_1 and λ_2 respectively.

In training, the generators try to minimize the objective function against the discriminators that try to maximize it, which can be formulated as:

$$G^*_{A\to B}, G^*_{B\to A} = \arg \min_{G_{A\to B}, G_{B\to A}} \max_{D_A, D_B} L\left(G_{A\to B}, G_{B\to A}, D_A, D_B\right). \tag{2}$$

2.2 Objective Function

Although achieving somewhat success in inter-domain image translation, CycleGAN has two obvious shortcomings when applied to medical images: (i) CycleGAN often produces unstable results due to GAN's characteristic of high freedom. In other words, the existing CycleGAN architecture lacks adequate supervision based only on the global adversarial loss and cycle consistency constraints; and (ii) For medical image enhancement, it is difficult to make sure that $G_{A\to B}$ and $G_{B\to A}$ focus on important features without extra constraints provided. On one hand, it is a challenge to remove too dark or bright regions and achieve more uniform appearance. On the other hand, subtle vital details such as curvilinear structures of corneal nerve fibers and complete morphology of digestive tract might be blurred or even lost in the translated images. To address these shortcomings, we propose to formulate and incorporate two new terms - illumination regularization and structure loss (as shown in red in Fig. 2), to guide the generator to improve illumination uniformity and structural details in the enhanced images.

• **Illumination Regularization** The proposed illumination regularization is a constraint aimed at improving overall illumination. It represents a correcting factor that reflects the non-uniformity of illumination in the enhanced image. The illumination correcting factor of the given image I is based on the following steps: (1) Calculate the average intensity of I; (2) Divide the image into $n \times m$ patches with a certain size, then calculate the average intensity of each patch and form the luminance matrix D; (3) Subtract the average intensity of I from each element of matrix D to obtain the luminance difference matrix E; (4) Resize the matrix E into the luminance distribution matrix R of the same size as I using bicubic interpolation; and (5) Calculate the average absolute value of elements in the matrix R. The illumination uniformity constraint is defined as:

$$\begin{aligned} L_{\text{illumination}}\left(G_{A\to B}\right) = E_{a\in A}\left[E_{global}\left[\left|upsampling\left\{E_{local}^{p\times p}\left[G_{A\to B}(a)\right]\right.\right.\right. \\ \left.\left.\left. -E_{global}\left[G_{A\to B}(a)\right]\right\}\right|\right]\right], \end{aligned} \tag{3}$$

where $E_{global}[\cdot]$ denotes the global mean of the whole input image; $E_{local}^{p\times p}[\cdot]$ is intended to calculate the luminance matrix D based on each $p \times p$ patch divided in the input image; and $upsampling\{\cdot\}$ aims at rescaling the input image to the size of the original image with bicubic interpolation.

•**Structure Loss** In addition to the illumination constraint, the low-quality image and the corresponding high-quality image should share similar structural features despite vast differences in terms of intensity and contrast distribution. Structural similarity (SSIM) [18] provides relatively reasonable measurement for this task. Compared with mean squared error (MSE), SSIM effectively characterizes the similarity of image structures from three aspects: luminance, contrast and structure. Inspired from the structure comparison function in SSIM, we propose a structure loss based on the dissimilarity between the low-quality image and the corresponding high-quality image. Mathematically, it is formulated as:

$$L_{\text{structure}}(G, X) = E_{x\in X}\left[1 - \frac{1}{M}\sum_{j=1}^{M}\frac{\sigma_{x_j, G(x)_j} + c}{\sigma_{x_j}\sigma_{G(x)_j} + c}\right], \tag{4}$$

where x_j and $G(x)_j$ are the j-th local window in the image x and the corresponding generated image $G(x)$ respectively; M is the number of local windows in each of the two images; $\sigma_{x_j,G(x)_j}$ is the covariance of x_j and $G(x)_j$; σ_{x_j} and $\sigma_{G(x)_j}$ are the standard deviations of x_j and $G(x)_j$ respectively; and c is a small positive constant. Thus the objective function of the proposed CSI-GAN for image enhancement is computed as:

$$\begin{aligned} & L\left(G_{A\to B}, G_{B\to A}, D_A, D_B\right) \\ &= L_{GAN}\left(G_{A\to B}, D_B, A, B\right) + L_{GAN}\left(G_{B\to A}, D_A, B, A\right) \\ &+ \lambda_1 L_{cyc}(G_{A\to B}, G_{B\to A}, A, B) + \lambda_2 L_{identity}\left(G_{A\to B}, G_{B\to A}, A, B\right) \\ &+ \gamma L_{\text{illumination}}\left(G_{A\to B}\right) + \alpha L_{\text{structure}}(G_{A\to B}, A) + \beta L_{\text{structure}}(G_{B\to A}, B), \end{aligned} \tag{5}$$

where γ, α and β are the parameters to control the weights of each part.

3 Experimental Results

To evaluate the performance of the proposed method, we applied it on two types of medical images, including the corneal confocal microscopy and the endoscopy. A clinical expert was invited to select low and high quality images of each dataset, primarily according to illumination uniformity, contrast and edge sharpness.

3.1 Experimental Settings

The proposed CSI-GAN was implemented in Python with PyTorch library. The experiments were carried out on a single NVIDIA GPU (GeForce GTX 1080, 8 GB). All training images were resized to 384×384. In our experiments, we selected two kinds of patch size - 48×48 and 96×96 to calculate illumination regularization and set local windows of 11×11 to calculate the structure loss. The Adam optimization with initial learning rate of 0.0002 and batch size of 1 were applied to the two adversarial pairs. The weighted parameters in the final loss function were set as: $\lambda_1 = 10$, $\lambda_2 = 5$, $\gamma = 1$, $\alpha = \beta = 5$.

3.2 Evaluation on Endoscopic Images

A total of 397 low-quality and 123 high-quality endoscopic images with the resolution of 384×288 pixels were selected. In our experiments, the low-quality images were randomly and equally divided into training and testing sets.

Five state-of-the-art approaches were selected for comparison: CLAHE [19], DCP [6], NST [13], MSG-Net [11], and EnlightenGAN [15]. An ablation study has also been conducted to justify the effectiveness of the illumination regularization and the structure loss. Due to page limitation, Fig. 3 only illustrates the enhancement results by one conventional (DCP) and one deep learning based (EnlightenGAN) enhancement method, respectively. It can be seen that DCP improves the overall brightness of the image, but it also amplifies the noise in severely dark regions and even produces some color distortion. This might be attributable to that our endoscopic images are not compatible with the haze

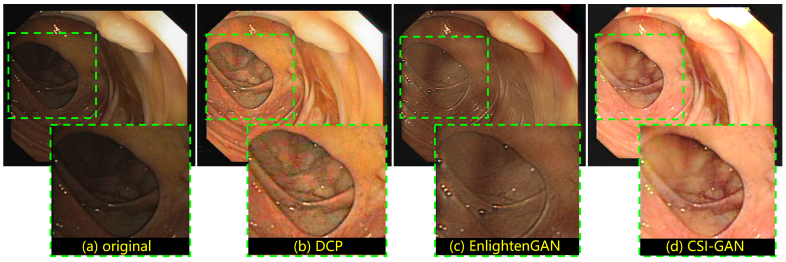

Fig. 3: Example result of different methods in endoscopic image enhancement.

Table 1: No-reference assessment results (mean ± standard deviation) of different enhancement methods.

Methods	NIQE	BRISQUE	PIQE
Original	4.40±0.67	36.70±5.42	37.27±11.50
CLAHE [19]	4.52±0.89	30.76±3.88	29.52±4.89
DCP [6]	4.13±0.64	35.45±5.59	35.09±6.37
NST [13]	9.42±1.96	30.28±3.64	25.48±6.17
MSG-Net [14]	7.26±0.55	56.72±2.29	93.43±11.96
EnlightenGAN [15]	4.38±0.88	**24.35 ± 4.18**	33.07±6.47
CycleGAN [16]	4.38±0.62	27.81±7.70	29.54±5.28
CycleGAN+I	4.33±0.61	27.70±6.71	25.25±9.35
CycleGAN+S	4.27±0.60	26.00±6.59	28.12±6.01
CSI-GAN	**3.84 ± 0.64**	24.52±5.38	**23.48 ± 6.42**

imaging model. The result of EnlightenGAN is universally over-smoothed with many details blurred. In contrast, the proposed CSI-GAN generates a visually pleasing result with more uniform illumination and visible structural details, especially in the region with poor lighting condition [see supplementary material for more examples]. For quantitative evaluation, we adopted three no-reference assessments: Natural Image Quality Evaluator (NIQE) [20], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [21] and Perception based Image Quality Evaluator (PIQE) [22]. Table 1 shows the results of the enhanced endoscopic images using different methods. Significant margins have been obtained when the proposed CSI-GAN is compared with other state-of-art methods - it achieves the highest scores in terms of NIQE and PIQE, and similar score to EnlightenGAN in BRISQUE, where ours is only 0.17 lower than EnlightenGAN. Overall, the proposed method has demonstrated its superiority in both visual comparison and quantitative evaluation over other competing methods.

Ablation study: Table 1 implies that our CSI-GAN is superior when compared with the results that CycleGAN enables with illumination (CycleGAN+I) or structure loss (CycleGAN+S) alone.

3.3 Evaluations on Corneal Confocal Microscopy

We further validated our image enhancement method over a publicly available CCM dataset [3]. A clinical expert was invited and selected 340 low and 288 high

Table 2: SNR and segmentation performance of the original and enhanced CCM images using different methods. (S: structure loss; I: illumination regularization)

	SNR			**Segmentation**			
Methods	r=3	r=5	r=7	ACC	SEN	Kappa	Dice
Original	17.472	17.611	17.650	0.969	0.421	0.528	0.541
CLAHE [19]	16.560	16.733	16.793	0.970	0.488	0.570	0.584
DCP [6]	14.587	14.879	14.986	0.964	0.708	0.615	0.633
NST [13]	16.606	16.887	17.006	0.958	0.490	0.494	0.515
MSG-Net [14]	19.122	19.915	20.217	0.964	0.441	0.495	0.512
EnlightenGAN [15]	18.407	19.257	19.699	0.960	0.671	0.580	0.601
CycleGAN [16]	19.557	20.138	20.409	0.971	0.748	0.673	0.688
CycleGAN + I	20.297	20.932	21.217	**0.977**	0.776	0.735	0.747
CycleGAN + S	20.105	20.747	21.042	0.971	0.769	0.680	0.695
CSI-GAN	**20.352**	**21.057**	**21.413**	**0.977**	**0.788**	**0.736**	**0.748**

quality images in this dataset, respectively, for training and 60 low-quality images for test. Note, nerve fibers in all these CCM images were manually annotated at centerline pixel level. We have released these image quality-annotations to the public, includes both low and high quality image sets [download link removed for review]. Similarly, we also compare our method with the other five methods and conduct an ablation study.

Evaluation by SNR: in order to comprehensively measure the image quality improvement of our proposed CSI-GAN, we first calculated signal-to-noise ratio (SNR) based on manually traced nerve fibers: $SNR = 10\log_{10}\left(\max(I_s)^2/\sigma_b^2\right)$, where $\max(\cdot)$ represents the maximum intensity of signal regions (manually traced nerve fibers) in the image, and σ_b is the standard deviation of the background region. In our experiments, we defined the regions after a disk-shaped dilation operation on the manually traced fibers with a radius (r) of 3, 5, and 7 pixels, respectively as the background region. The quantitative results of different enhancement approaches are shown in Table 2. The proposed method has achieved the best performance when compared with all the competing methods. It exhibits a large advantage against the original images by an increase in SNR of about 2.88 dB, 3.45 dB, and 3.76 dB for r=3, r=5, and r=7, respectively.

Evaluation by nerve fiber segmentation: we further performed corneal nerve fiber segmentation of the enhanced images to confirm the relative benefits of the proposed method and the others. To this end, we employed a pre-trained corneal nerve fiber segmentation network, CS-Net [3], for fully automatic segmentation of corneal nerves on the low-quality images, with and without application of image enhancement methods. Then we computed *sensitivity* (SEN), *Accuracy* (ACC), Kappa score and Dice coefficient between the predicted centerlines and ground truth ones. The top row of Fig. 4 demonstrates the visual enhancement results by different methods, while its bottom row depicts the enhancement guided fiber segmentation results obtained by CS-Net. We observed that more completed fibers have been identified in our enhanced images, where

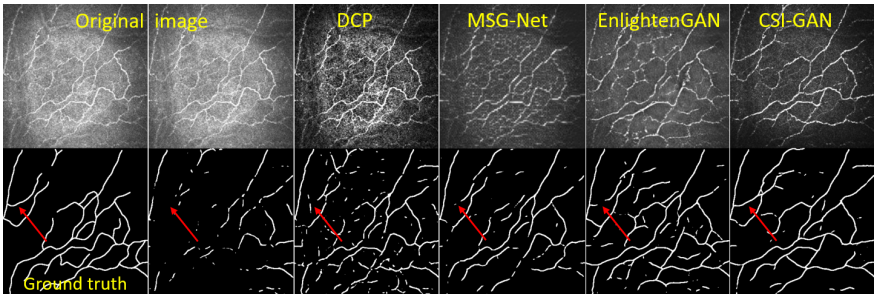

Fig. 4: An example of CCM image enhancement by different methods, and their guided nerve fiber tracing results using CS-Net.

indicated by the red arrows, since more uniform responses in both low and high intensity regions of the original image has been achieved by CSI-GAN. In consequence, CS-Net is able to provide relatively more sensitive segmentation on small fibers. This finding is also evidenced by the segmentation performance presented in Table 2. Our CSI-GAN method achieves the best performance and helps to improve the state-of-the-art EnlightenGAN and the baseline CycleGAN by 17.44% and 5.35% in SEN, 26.90% and 9.36% in Kappa, and 24.46% and 8.72% in Dice respectively, demonstrating that our method can effectively promote nerve fiber segmentation performance, especially in reducing missing rate, which is more useful for monitoring and diagnosing nerve-related diseases.

Ablation study: Table 2 demonstrates our CSI-GAN yields significant improvements in either SNR or nerve fibre segmentation when compared with their performance using illumination (CycleGAN+I) or structure loss (CycleGAN+S) alone. It further verifies that by adding illumination and structure losses, our method achieves high visual image quality and well-restored structural details.

4 Conclusion

Image enhancement is helpful for improving visual quality and automatic analysis of medical images. However, it is still challenging due to different illumination conditions and diversity in quality of different medical imaging devices. This paper has proposed an unpaired learning architecture called CSI-GAN for medical image enhancement, where the low and high quality images are treated as those in two different domains. The primary advantage of this method is that it learns to migrate the features inside the high-quality images into low-quality images without paired images for training and thus has an advantage of easy implementation. Furthermore, by adding illumination regularization and structure loss, the overall illumination smoothness and well-restored structural details in the enhanced images are achieved. Compared with other traditional or deep learning-based methods, our method obtains better overall performance in enhancing images with different modalities in different metrics. In the future, we would consider further verifying our method on other medical image modalities and applying the enhanced images in the clinical settings for disease diagnosis.

References

1. Lai, M.: Deep learning for medical image segmentation. arXiv preprint arXiv:1505.02000 (2015)
2. Fu, H., Wang, B., Shen, J., Cui, S., Xu, Y., Liu, J., Shao, L.: Evaluation of retinal image quality assessment networks in different color-spaces. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2019) 48–56
3. Mou, L., Zhao, Y., Chen, L., Cheng, J., Gu, Z., Hao, H., Qi, H., Zheng, Y., Frangi, A., Liu, J.: Cs-net: Channel and spatial attention network for curvilinear structure segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2019) 721–730
4. Zhao, Y., Zheng, Y., Zhao, Y., Liu, Y., Chen, Z., Liu, P., Liu, J.: Uniqueness-driven saliency analysis forautomated lesion detection with applications to retinal diseases. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer (2018) 109–118
5. Abdullah-Al-Wadud, M., Kabir, M., Dewan, M., Chae, O.: A dynamic histogram equalization for image contrast enhancement. IEEE Transactions on Consumer Electronics **53**(2) (2007) 593–600
6. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(12) (2010) 2341–2353
7. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(6) (2012) 1397–1409
8. Chen, Y.S., Wang, Y.C., Kao, M.H., Chuang, Y.Y.: 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) 6306–6314
9. Guo, X.: Lime: a method for low-light image enhancement. In: Proceedings of the 24th ACM international conference on Multimedia. (2016) 87–91
10. Lv, F., Lu, F., Wu, J., Lim, C.: Mbllen: Low-light image/video enhancement using cnns. In: BMVC. (2018) 220
11. Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., Ma, J.: Msr-net: Low-light image enhancement using deep convolutional network. arXiv preprint arXiv:1711.02488 (2017)
12. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition **61** (2017) 650–662
13. Gatys, L., Ecker, A., Bethge, M.: A neural algorithm of artistic style. Journal of Vision **16**(12) (2016) 326
14. Zhang, H., Dana, K.: Multi-style generative network for real-time transfer. arXiv preprint arXiv:1703.06953 (2017)
15. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. arXiv preprint arXiv:1906.06972 (2019)
16. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. (2017) 2223–2232
17. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017) 1125–1134

18. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4) (2004) 600–612
19. Zuiderveld, K.: Contrast limited adaptive histogram equalization. In: Graphics gems IV, Academic Press Professional, Inc. (1994) 474–485
20. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. IEEE Signal Processing Letters **20**(3) (2012) 209–212
21. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing **21**(12) (2012) 4695–4708
22. Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S.: Blind image quality evaluation using perception based features. In: 2015 Twenty First National Conference on Communications (NCC), IEEE (2015) 1–6