EFINet: Restoration for Low-Light Images via Enhancement-Fusion Iterative Network

Chunxiao Liu, Fanding Wu, and Xun Wang, *Member, IEEE*

***Abstract*— The lighting environment in the real world is so complex that most existing low-light image restoration methods suffer from color cast and local over-exposure. In order to solve these problems, this paper proposes the enhancement-fusion iterative network (EFINet) for low-light image enhancement. Within each iteration of EFINet, a stretching coefficient estimation network based enhancement module is designed to adjust the input image pixel-wisely to obtain the initial enhancement result with the estimated coefficient maps. Then, an encoder-decoder based fusion network is devised to extract the deep features and combine the well-exposed local areas in both the input image and the initially enhanced image, to obtain a visually pleasing, high-quality image enhancement result. The coefficient estimation network and the fusion network are weight shared among all iterations. What's more, most of the low-light image datasets are generated through illumination reduction in a global way. To better simulate the diverse illumination distribution in the real world, we put forward a new low-light image synthesis method to produce the low-light images with non-uniform illumination for the network training purpose. After conducting extensive experiments on both synthetic and real-world low-light images, the results verify the superiority of our algorithm over the state-of-the-art (SOTA) methods, especially in balancing the brightness difference and preventing over-enhancement.**

***Index Terms*— Non-uniform illumination, low-light image enhancement, light-weight CNNs.**

I. Introduction

NIGHTTIME accounts for at least half of the whole day, but the poor lighting situations at night will result in dark images. In addition, extreme weathers such as cloudy or rainy days as well as indoor environments with dim light sources can also lead to the imaging results with weak lighting effects. Low-light images are characterized by low visibility, color degradation and rich noise, which are unpleasant to human eyes. The useful information in the dark areas can be hidden to the human investigators at the same time. More importantly, unclear contents in the low-light images are big obstacles to the high-level downstream computer vision tasks conducted by the machine. Sufficiently and uniformly illuminated images tend to be vivid in color as well as clear in details, which not only can give the human eye a good visual experience, but also can be beneficial to the subsequent tasks such as nighttime monitoring, as a result the low-light image enhancement technique is particularly essential.

Many methods have been put forward for the low-light image enhancement task, but they all have limitations either in efficiency or in performance. Traditional image processing-based methods generally fall into histogram equalization (HE) [5], [6] or Retinex [7] based approaches. To some extent, they solve part of the issues about image contrast and brightness. In comparison, deep learning-based methods become more popular recently. Retinex based CNN models [1], [4], [8] decompose the low-light image into the illumination and reflectance maps, which are then enhanced and combined to reconstruct the enhancement result. Fusion based CNN models [2], [9] obtain the enhancement result by merging the enhanced features up to different levels or the candidate images with different exposure levels. The unsupervised CNN models [10]–[12] enhance the low-light images under the constraints of the prior loss functions or the unpaired image dataset. For above deep learning based methods, on the one hand, it is challenging to recover both under- and over-exposed areas simultaneously, and their results are prone to be over-enhanced. Fig. 1 shows the results of different low-light image enhancement methods. Evidently, color distortion, over- or under-enhancement are ubiquitous in Fig. 1(b)-(e). On the other hand, the training dataset used by them only contains the low-light images generated through global illumination reduction, thus this monotonous data synthesis approach easily leads to poor model generalization performance.

Considering the limitations of existing methods, we propose an enhancement and fusion based network architecture with iterative optimization strategy in a relatively simplified and lightweight form, which is named as EFINet. Each iteration of EFINet includes two stages, i.e., image enhancement and image fusion. In the first stage, the lightweight stretching coefficient estimation network-based enhancement module brightens or darkens the under- or over-exposed areas rapidly by analyzing the illumination distribution of the input images, so as to generate the initial enhancement result. This process improves the global visibility and contrast of the input images, but the result is not satisfactory enough. Then, in the second stage, to refine the local illumination, we enter both the input image and the initial enhancement result to the Fusion Network. The Fusion Network can extract image features adaptively and combine the well-lit areas of the two images to get the final enhancement result for the current

Manuscript received 17 January 2022; revised 24 April 2022 and 29 June 2022; accepted 21 July 2022. Date of publication 3 August 2022; date of current version 6 December 2022. This work was supported by the National Natural Science Foundation of China under Grant 61976188. This article was recommended by Associate Editor M. Teutsch. *(Corresponding author: Chunxiao Liu.)*

The authors are with the School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China (e-mail: cxliu@zjgsu.edu.cn; kyrie111219@gmail.com; xw@zjgsu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSVT.2022.3195996.

Digital Object Identifier 10.1109/TCSVT.2022.3195996

iteration, which is regarded as the input of the next iteration. Through iterative optimization, the proposed method avoids the quick emergence of noise, and restore the structure and color information in conjunction with our loss functions. It is worth mentioning that our model with only 129k parameters is more lightweight than that of current learning-based methods, which makes a better trade-off between model effectiveness and computational requirements.

Moreover, we further propose a new low-light image synthesis method to generate a low-light image dataset with non-uniform illumination based on the principle of locally random brightness assignment. Our synthesis method can embody the middle-level visual perception well based on superpixel segmentation, which reflects certain semantic information while maximizing the randomness of the illumination distributions. Previous datasets do not consider non-uniform illumination simulation on a single image, therefore the models trained with them have difficulties in adapting to extremely complex lighting conditions. By comparison, our dataset facilitates the network to recover the image colors and details while improving the image visibility. In short, the contributions of our work can be summarized as follows:

- We propose a novel network architecture called the enhancement-fusion iterative network (EFINet) for the low-light image enhancement task. It can well balance the non-uniform illumination on the image, and restore the image color and details.
- We construct a new low-light image synthesis method to generate a low-light image dataset with non-uniform illumination, which provides powerful data support for our model training process as well as further research works on the low-light image enhancement field.
- We carry out comprehensive experiments on our dataset and network architecture, which proves that our method is superior to the most advanced methods both qualitatively and quantitatively.

II. Related Work

A. Traditional Image Processing Based Methods

Before the appearance of data-driven methods, the representative methods in the past can mainly be divided into two categories. The first one is devised based on image histograms [13]. HE [5], [6] attempts to map the image histograms to a uniform distribution so as to improve the image contrast. CLAHE [14] divides the image into multiple regions and applies local adaptive histogram equalization under the condition of contrast limitation, which can effectively suppress noise amplification while increasing the image contrast. On the basis of depth information, DGACE [15] enhances the image by changing pixel values adaptively according to a 2-dimensional histogram. Generally speaking, these methods are not flexible enough for local image adjustment, thus serious noise and unnatural illumination still exist.

The Retinex theory [16] lays a foundation for the second kind of methods [17]–[21], whose core is to decompose the image into such two parts as illumination and reflectance, and

Fig. 1. Low-light image enhancement example. (a) Low-light input image. (b)-(e) Enhancement results of SOTA methods (RetinexNet [1], MBLLEN [2], GLADNet [3], KinD [4]). (f) Our result.

enhance them respectively to reconstruct the normal-exposed image. Variant Retinex based methods like SSR [22] and MSR [17] just treat the reflectance map as the final enhancement result. LIME [19] treats the maximum of three channel of pixels as the illumination and develops a structure-aware smoothing model to achieve illumination consistency. Nevertheless, LIME cannot recover extremely dark images. MF [18] adjusts the initially estimated illumination and fuses its multiple derivation results, which sometimes makes the local areas with dense texture look unrealistic. Since Retinex is actually an ill-posed inverse process, the introduction of reasonable prior information is crucial. However, due to the limitations of human experience, above methods fail in adapting to various circumstances. In addition, because of the non-linearity of color perception and the complexity of the image data, they may deviate from the original color information in the low-light images.

Additionally, ENR [23] and SepDehaze [24] creatively apply the defogging method to the inverted input images, and can sometimes get the enhancement results with better visibility. However, ENR adopts BM3D [25] as the post-processing step, which is likely to cause blur.

B. Recent Deep Learning Based Methods

As deep learning techniques have made great progress in recent years, they are widely used in various computer vision tasks, such as image denoising, dehazing, and

Fig. 2. Four representative image pairs of our dataset. First row: the original normal-light images. Second row: the degraded low-light images with non-uniform illumination.

de-raining, etc. Recent deep learning based low-light image enhancement methods [26]–[29] have gradually become the mainstream due to their superior effect and generalization performance. MBLLEN [2] enhances and fuses the features extracted from different layers to obtain the enhancement result. With the global illumination estimation module and detail reconstruction module, GLADNet [3] and PRIEN [30] are enabled to maintain image details and improve image visibility. STANet [31] and ELLIE [32] specifically exploit structure features to strengthen perceptual quality. Zhao *et al.* [33] use invertible neural networks to learn image feature transformation. STAR [34] provides a real-time image enhancement algorithm by a lightweight transformer. RetinexNet [1] introduces a Retinex model based image decomposition network. Similarly, the Retinex-based approaches, such as KinD [4], DeepUPE [35], SICE [36], CSDNet [37], SCU [38], RetinexDIP [39], all decompose the input image into the illumination and reflectance components first, which are then enhanced separately to get the final enhancement result. These methods belong to the supervised learning based approaches. Two of the most critical factors influencing the enhancement performance are the quality and quantity of the paired training image samples. In other words, the training dataset plays a decisive role for the model performance.

Aside from the above methods, the unsupervised learning-based approaches [11], [40]–[42] become increasingly popular because they get rid of the dependence on paired training samples. What is worth mentioning is EnlightenGAN [10], which constrains the unpaired training with global-local discriminators, the self-regularized perceptual loss, as well as the attention mechanism. Another representative unsupervised method is a lightweight network named Zero-DCE [12], which estimates a series of parameters for the adjustment curves to enhance the dynamic range of the input image. Even though these methods do not need paired training image samples and narrow the domain gap between training samples and real-world images to some degree, their results often suffer from detail loss and color degradation.

III. Synthetic Dataset Generation

To overcome the limitations of existing methods, we first propose a data synthesis method to generate the low-light images with non-uniform illumination based on the locally random brightness assignment strategy. Our motivation is to degrade the normal-light images to simulate the low-light images with complex illumination distribution so as to build a large-scale paired image dataset. Our low-light image synthesis method can embody the middle-level visual perception well, which reflects the semantic information to some extent based on the superpixel segmentation and merging while maximizing the randomness of the illumination distributions. Fig. 2 shows a few representative examples generated. The detailed steps of our low-light image dataset generation process are described in the following subsections.

A. Normal-Light Image Selection

Our dataset is synthesized on the basis of MIT-Adobe FiveK [43]. MIT-Adobe FiveK contains 5000 images with diverse scenes and subjects, which will be beneficial to the robustness of our model. Although MIT-Adobe FiveK is an excellent dataset for learning photography adjustments, some images inside fail to meet the requirements of the normal-light ground truths for low-light image enhancement. As a consequence, we put forward three rules for the selection of normal-light images, i.e. (1) The overall exposure of the image should be full and uniform; (2) The details of the image must be clear without any blurring effects; (3) The image needs to be clean without obvious noise. According to the above conditions, we obtain 1364 qualified high-resolution images.

Since the original size of the images in MIT-Adobe FiveK is generally too large to train the network model, we first resize the images with different scaling factors and then perform non-overlapping cropping operations. In this way, on the one hand, the network can be exposed to image content of different scales in the training phase, which is conducive to increasing the generalization performance of the network. On the other hand, this is also an efficient means for data augmentation.

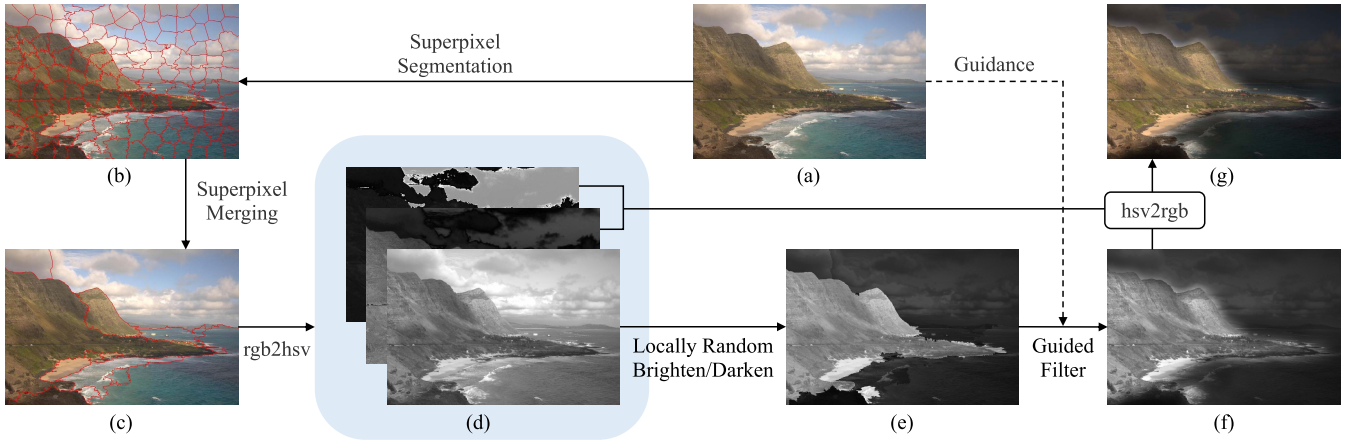

Fig. 3. Synthesizing process of low-light images with non-uniform illumination. (a) Original image. (b) Superpixel segmentation result. (c) Superpixel merging result. (d) H, S, V channels (from top to bottom). (e) Non-uniform illumination map. (f) Guided filtering result. (g) Synthesized low-light images with non-uniform illumination.

B. Low-Light Image Synthesis

Our purpose of synthesizing low-light images with non-uniform illumination is to better simulate the random lighting environment in the real world. The images in the real-world paired image dataset, like the LOL dataset [1], are captured mainly by changing the exposure time or ISO. When it comes to the outdoor scene, object movement or camera shake always makes it challenging to align image pairs, which results in ghosting artifacts and blur in the enhancement results. For the existing synthetic datasets, most of them only make changes with the global brightness reduction, which leads lead to poor performance of the network for the low-light images with local exposure adjustment.

On the basis of above considerations, we propose a method to generate the low-light images with non-uniform illumination, as shown in Fig. 3. Compared to seeing pixels as the basic unit of image adjustment, superpixels can better reflect the semantic information of the image such as similar textures. First of all, as Fig. 3(a)(b) shows, we perform superpixel segmentation [44] operation on the cropped normal-light image and partition it into several superpixels, which fit image edges and structures. Next, we merge the adjacent superpixels with similar color means to embody the middle-level visual information further according to a threshold $\tau = 50$. Fig. 3(c) shows the superpixel merging result. It can be seen that the merged superpixels express part of semantic information in the image, like the silhouette of the mountain, which benefits our network model for learning to recognize the objects with different appearance.

With the merged superpixels, we randomly brighten or darken each local area in the image, which helps the network learning to restore the under- and over-exposed areas at the same time. We first transform the image from RGB into HSV color space, and separate the brightness channel V. Based on statistical analysis of the images adjusted through different parameters, we have determined the limits of brightening and darkening operations to ensure that image content will not be lost after the illumination adjustment operation. Our local illumination adjustment for the merged superpixels can be expressed as follows:

$$V'_{local} = \begin{cases} (V_{local})^{\gamma_1}, & if\ op = Brighten \\ r \cdot (V_{local})^{\gamma_2}, & if\ op = Darken \end{cases} \tag{1}$$

where V_{local} denotes the V channel of a superpixel in the original normal-light image Y. As for the brightening operation, we perform gamma correction with the factor $\gamma_1 \in [0.55, 0.8]$ on the V channel of current superpixel. As for the darkening operation, we first perform gamma correction with the factor $\gamma_2 \in [1.5, 1.6]$ and then apply the linear adjustment with the local multiplier $r \in [0.3, 0.55]$ on the V channel of current superpixel. Besides, for those pixels with lower brightness value which may be located on the objects with darker color in the normal-light images, we do not change their color characteristics and turn them to be brighter, to avoid deviating from the general color distribution of real-world dark images. Accordingly, we measure the color and brightness of superpixels to decide whether to brighten or darken them. For a superpixel with an average brightness value less than 64, only the darkening function can be operated, while the local multiplier r is constrained in $[0.4, 0.55]$ correspondingly. It is worth mentioning that r, γ_1 and γ_2 all follow a random uniform distribution. In order to make the brightness transitions among the superpixels in the generated image more smooth and natural, we take the original image Y as the guidance and perform guided filtering [45] on V'_{local}. Fig. 3(g) shows a sample of the generated low-light images with non-uniform illumination. At last, we must complement that our generated dataset also contains the low-light images with relatively uniform illumination, which makes our network model deal with the low-light images with both uniform and non-uniform illumination distributions at the same time.

IV. Proposed Method

To achieve the effect of progressive low-light image enhancement optimization, we design a novel two-stage iterative network architecture named EFINet, which is composed

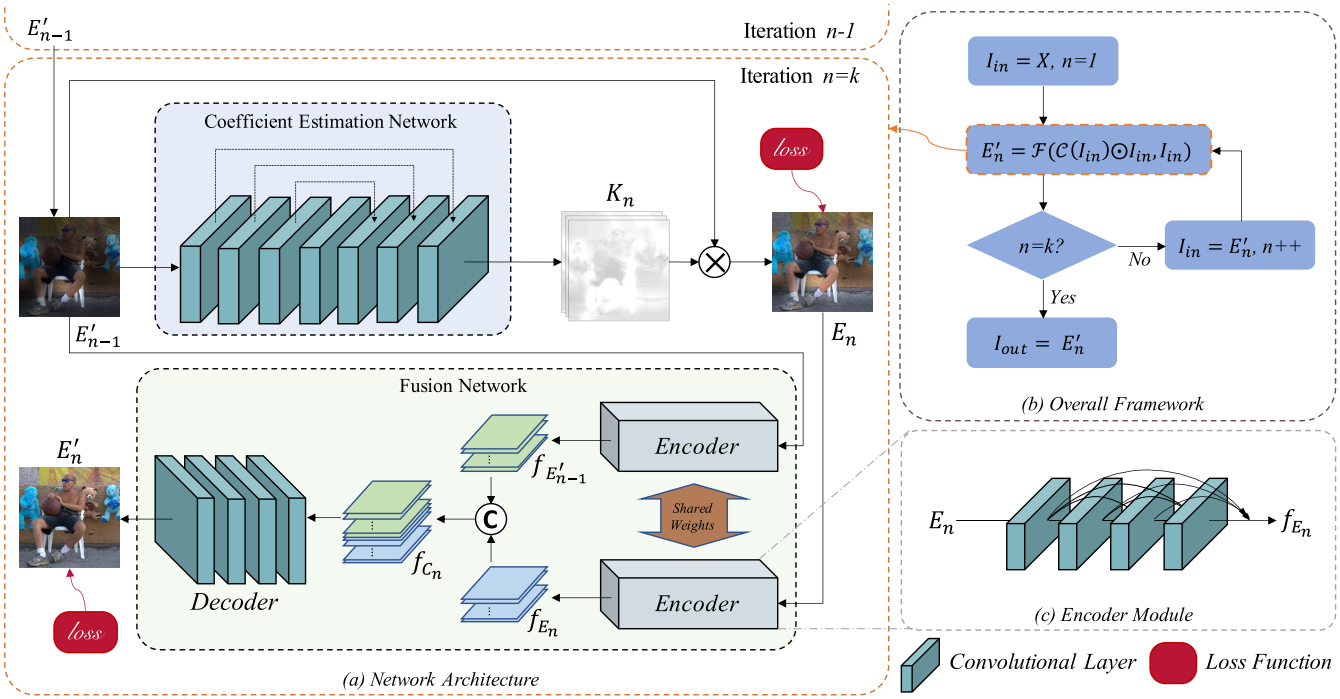

Fig. 4. The architecture of our enhancement-fusion iterative network (EFINet). (a) Our EFINet consists of two sequential stages, i.e. the Coefficient Estimation Network and the Fusion Network. The figure shows the process of the n-th iteration. (b) The overall framework of EFINet, where I_{in} represents the input of EFINet, n is the current number of iterations, I_{out} represents the output of EFINet, and k is the total number of iterations. (c) The encoder module.

of stretching coefficient estimation based enhancement and fusion based refinement. Below is the detailed description of our EFINet and its two main components, i.e., the Coefficient Estimation Network and the Fusion Network.

A. Network Architecture

Fig. 4(a) shows the network architecture of EFINet. As the initial input, the low-light image X is first put into the Coefficient Estimation Network $\mathcal{C}$ in the first iteration, which estimates a 3-channel stretching coefficient map K_1 corresponding to the RGB channels of X pixel by pixel. K_1 embodies the illumination distribution and color information of X, and adaptively reflects different enhancement strengths for different areas. Subsequently, K_1 and X are combined by multiplying the corresponding pixel values to obtain the initial enhancement result E_1. Optimization through back-propagation under the constraints of normal-light ground truths enables the Coefficient Estimation Network to automatically identify the pixels or regions in the input image to be brightened or darkened and predict their stretching degrees at the same time. We believe that E_1 has a preliminary enhancement effect on the illumination of the image.

Furthermore, in order to combine the useful information and better parts from both the input image X and the initial enhancement result E_1 to get the better result, we design an image fusion network in the second stage. After E_1 is generated as the initial enhancement result, it enters the Fusion Network $\mathcal{F}$ together with X. The weight-shared encoder module extracts deep features f_X and f_{E_1} from the two images respectively. We combine these two feature maps with the concatenation operator, thus the decoder module can merge them more flexibly. Eventually, the decoder reconstructs the feature maps f_{C_1} to E'_1, which can be treated as the final enhancement result of the first iteration.

Similar to the first iteration, E'_1 will be put into the network as the input image of the next iteration to obtain the enhancement result E'_2. By analogy, we obtain the final enhancement result E'_n after n iterations through above iterative way. In each iteration, the Fusion Network integrates the input and output images of the previous coefficient estimation based initial enhancement stage many times, which is beneficial to get the final result without losing the information in the previous iteration. The overall framework of our method can be seen in Fig. 4(b) and expressed as follows:

$$E'_1 = \mathcal{F}(\mathcal{C}(X) \odot X,\ X) \tag{2}$$

$$E'_2 = \mathcal{F}(\mathcal{C}(E'_1) \odot E'_1,\ E'_1) \tag{3}$$

$$\vdots$$

$$E'_n = \mathcal{F}(\mathcal{C}(E'_{n-1}) \odot E'_{n-1},\ E'_{n-1}) \tag{4}$$

where X is the initial input low-light image, $\mathcal{C}$ and $\mathcal{F}$ represent the Coefficient Estimation Network and the Fusion Network, $\odot$ is the matrix dot product operator, E'_n is the final enhancement result after n iterations. Through experiments and comparisons, we finally set n to 3 in this work.

Fig. 5(a)-(c) show the stretching coefficient maps of an image in three rounds of iterative enhancement. As can be seen by observing, the coefficient map mainly emphasizes the overall structure boundary and weakens the texture characteristics on the input image. Compared with Fig. 5(a), the smoother version Fig. 5(c) can take into accounts the overall information of the image, which indicates that the

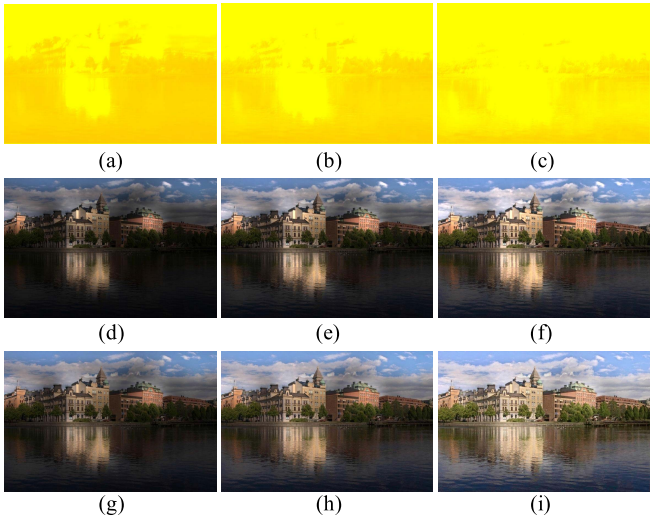

Fig. 5. Intermediate and final results of EFINet. (a)-(c) The stretching coefficient maps. (d)-(f) The initial enhancement results. (g)-(i) The final enhancement results. Left Column: the first iteration. Middle Column: the second iteration. Right Column: the third iteration.

network can correct the illumination of low-light images from coarse to fine sequentially. Fig. 5(d)-(f) and Fig. 5(g)-(i) are the initial and final enhancement results of each iteration. In comparison, the initial enhancement and fusion part also plays the roles of coarse and fine enhancement respectively. From the perspective of human vision, the illumination in the results of the latter iterations are more uniform and pleasant than those in the previous iterations, which embodies the main purpose of our network to enhance the image round by round.

1) Coefficient Estimation Network: In order to learn the brightness mapping between the input low-light image and the initial enhancement result, we use a lightweight network with only seven convolutional layers to estimate the stretching coefficient map. Skip connections [46] are added between the first and seventh layers, the second and sixth layers, and the third and fifth layers of the network, which strengthens the information transfer and improve the running stability of the network model. The activation function we used in the first six layers is ReLU.

2) Fusion Network: With the purpose of combining good local areas and generate better results, we design the Fusion Network with an encoder module, a fusion layer, and a decoder module. Inspired by [47], dense connections [48] are added between each convolutional layer in the encoder. As shown in Fig. 4(c), the outputs of previous layers are sent to the subsequent layers, which helps to preserve the deep features obtained from different layers. In the fusion layer, the strategy we adopt is to concatenate two sets of feature maps. The decoder transforms the features into images through four convolutional layers in the end.

B. Loss Function

The network is trained end-to-end under the constraints of the loss function. It is worth noting that we only apply the loss function to the result of the last iteration. With the expectations of restoring the color of the image, maintaining the details, and improving the visibility, our loss function can be defined as follows:

$$\mathcal{L} = \lambda_r \mathcal{L}_r + \lambda_p \mathcal{L}_p + \lambda_s \mathcal{L}_s + \lambda_c \mathcal{L}_c, \tag{5}$$

where λ_r, λ_p, λ_s and λ_c are the balance coefficients of image reconstruction loss $\mathcal{L}_r$, perceptual loss $\mathcal{L}_p$, structure similarity loss $\mathcal{L}_s$ and color angle loss $\mathcal{L}_c$, respectively.

1) Reconstruction Loss: We apply the basic loss function L_1 to measure the reconstruction error of the final enhancement result E'_n during the training process. Formally, $\mathcal{L}_r$ can be expressed as:

$$\mathcal{L}_r(E'_n, Y) = |E'_n - Y|_1, \tag{6}$$

where Y is the expected normal-light ground truth corresponding to E'_n.

2) Perceptual Loss: To make the enhancement result with better visual perception and lower perceptual error, we extract the features of E'_n and Y by the VGG Network [49], and wish their feature differences as smaller as possible. We define the $\mathcal{L}_p$ based on the pre-trained VGG-16 network as:

$$\mathcal{L}_p(E'_n, Y) = \frac{\sum_{x=1}^{w_{ij}} \sum_{y=1}^{h_{ij}} \sum_{z=1}^{c_{ij}} |\phi_{ij}(E'_n)_{xyz} - \phi_{ij}(Y)_{xyz}|_2}{w_{ij} h_{ij} c_{ij}} \tag{7}$$

where w_{ij}, h_{ij} and c_{ij} describe the dimensions of each feature map in the VGG-16 network respectively. Additionally, ϕ_{ij} represents the feature map obtained by the j-th convolutional layer in the i-th block of the VGG-16 network.

3) Structural Loss: The images captured under poor lighting conditions usually present visually significant structural distortion, such as blurring and artifacts, thus we use SSIM [50] to evaluate reconstructed image quality. More specifically, the value of SSIM ranges from 0 to 1, where the smaller value represents the worse similarity. Consequently, the structural loss L_s can be expressed as:

$$\mathcal{L}_S(E'_n, Y) = 1 - SSIM(E'_n, Y) \tag{8}$$

$$SSIM(S, T) = \frac{2\mu_S \mu_T + c_1}{\mu_S^2 + \mu_T^2 + c_1} \cdot \frac{2\sigma_{ST} + c_2}{\sigma_S^2 + \sigma_T^2 + c_2} \tag{9}$$

where S, T represent the images to be compared. μ_S and μ_T are the mean pixel values of the two images. σ_S^2 and σ_T^2 are the variances of the two images, and σ_{ST} is their covariance. c_1 and c_2 are two smaller constants that prevent the denominator from being zero.

4) Color Angle Loss: Compared with the normal-light images, low-light images generally suffer from color degradation. Therefore, we formulate $\mathcal{L}_c$ to constrain the colors of E_n and E'_n to be consistent with the normal-light ground truth Y as much as possible. The definition of $\mathcal{L}_c$ is shown as the following formula:

$$\mathcal{L}_c = \mathcal{L}_c(E_n, Y) + \mathcal{L}_c(E'_n, Y) \tag{10}$$

$$\mathcal{L}_c(S, T) = \sum_{p=1}^{N} \boldsymbol{CA}(S_p, T_p) \tag{11}$$

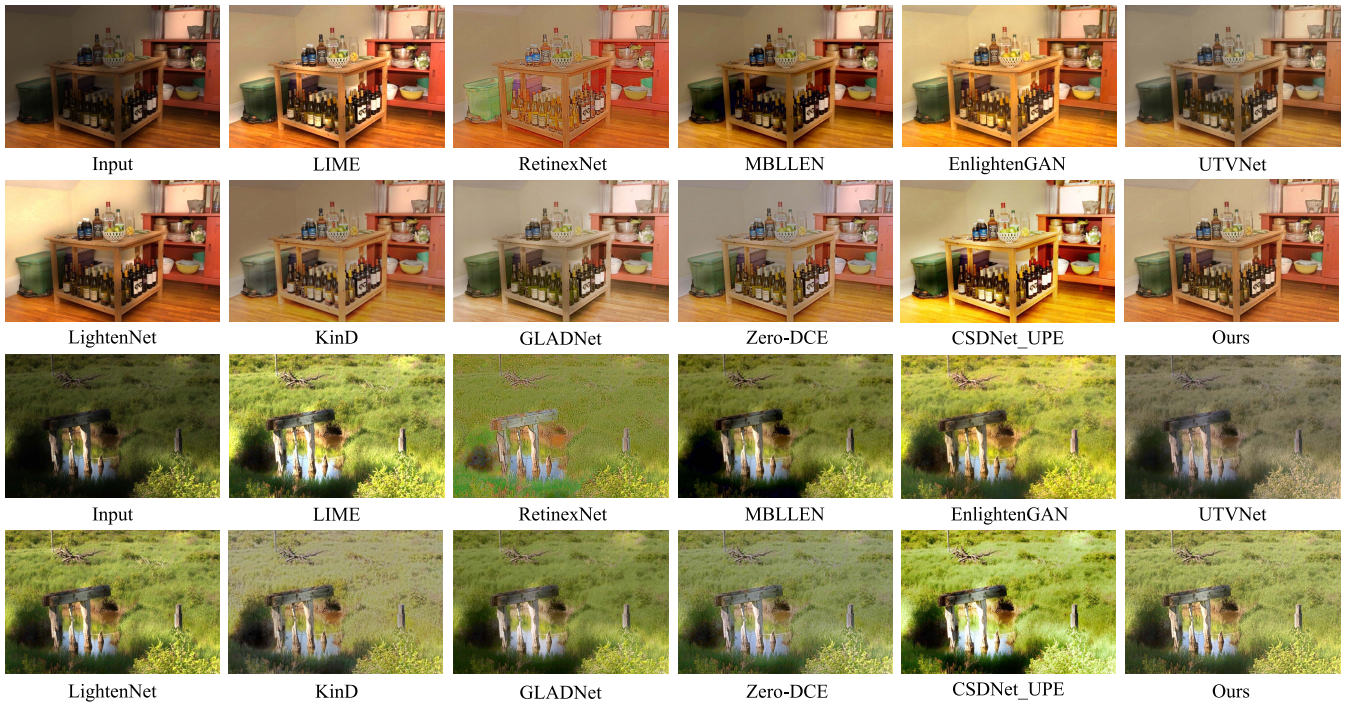

Fig. 6. Visual comparison of different methods on synthetic low-light images.

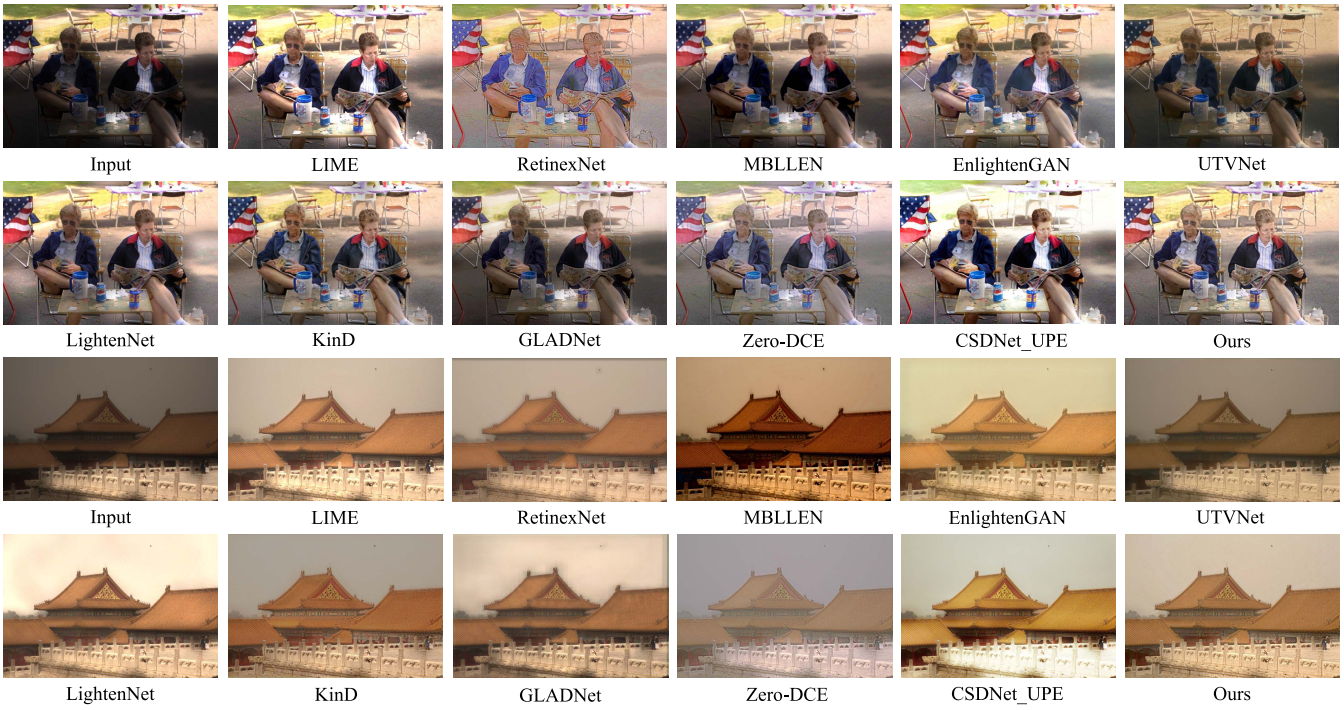

Fig. 7. Visual comparison of different methods on synthetic low-light images.

where S_p and T_p denote the RGB color vectors of pixel p in images S and T respectively, $\boldsymbol{CA}(\cdot,\cdot)$ is an operator used to calculate the color angle between two 3-dimensional color vectors.

C. Implementation Details

We use PyTorch [51] to implement our EFINet on a PC with NVIDIA RTX 2080Ti GPU and Intel Core i7-9700 3.00GHz CPU. The network is trained for 500 epochs with the batch-size of 8. For data augmentation, we resize the image to be 1/3, 1/5, 1/7 of the original size, clip them into 256 $\times$ 256 non-overlapping patches, and randomly mirror or rotate all image patches. Ultimately, we get 15096 images for training, 1837 images for validation and testing. The network is optimized by using Adam optimizer [52], and the hyperparameters of optimizer are set as $\beta_1 = 0.9$,

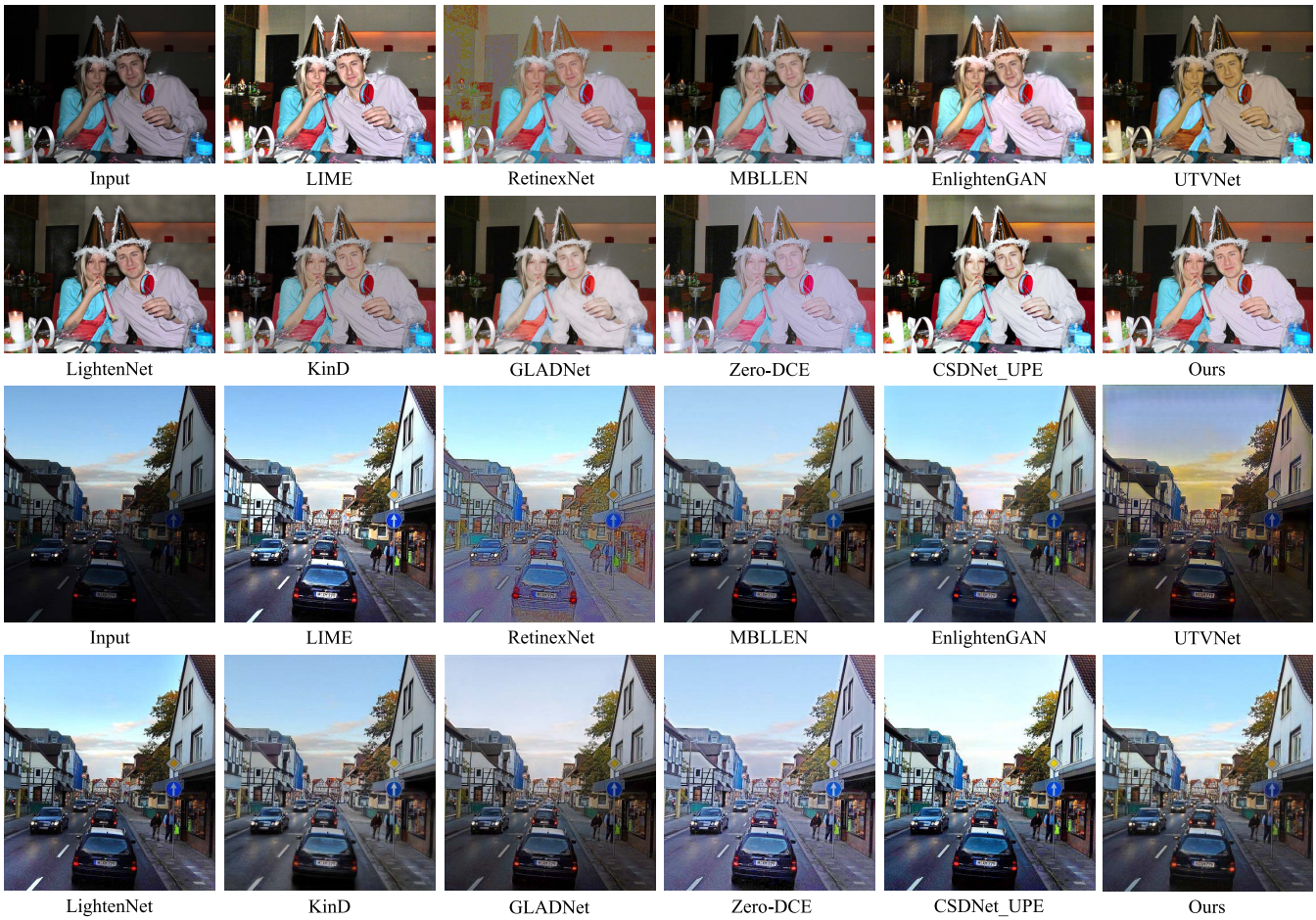

Fig. 8. Visual comparison of different methods on real-world low-light images.

$\beta_2 = 0.999$ and $\varepsilon = 10e-8$. λ_r, λ_p, λ_s and λ_c are set to 1, 1, 1, and 3 respectively. We fix the learning rate at $10e-4$ during the training phase.

V. Experimental Results

To evaluate the performance of our proposed method in enhancing low-light images, we qualitatively and quantitatively compare our method with eleven methods with available codes, including NPE [53], LIME [19], RetinexNet [1], MBLLEN [2], EnlightenGAN [10], UTVNet [29], LightenNet [8], KinD [4], GLADNet [3], Zero-DCE [12], CSDNet_UPE [37]. To make it fair, we re-train other methods on our synthetic dataset based on the source code and the hyper-parameter settings provided by the authors. Furthermore, we would like to verify the generalization performance of different network architectures trained on their original training dataset, by comparing the performances of the authors provided network models on untouched low-light images, which come from five real-world captured image dataset.

A. Qualitative Evaluation

We conduct extensive qualitative evaluations on both synthetic and real-world low-light images, comparing our visual effects with those of other eleven SOTA methods. To make it fair, we re-train other deep learning-based methods on our synthetic dataset based on the source code and the hyper-parameter settings provided by authors. Fig. 6 and Fig. 7 show the results of different methods tested on our synthetic dataset. Compared with existing methods, ours can better balance the illumination among brighter and darker areas, so that our results hold more uniform illumination while maintaining the image contrast. Especially for the transition between the regions with different illumination, our method can produce more natural and smoother results with better visual perception quality.

Fig. 8 and Fig. 9 show the enhancement results for the real-world low-light images by different methods. It is easy to see that, for the images with wide illuminartion distribution, the results of LIME and CSDNet_UPE are tend to be locally over-enhanced. Meanwhile, GLADNet and Zero-DCE can hardly restore the color of the image, and there is a general tendency to lose color authenticity in their results. To some extent, MBLLEN performs well for the noise removal, but it sacrifices the texture details of the image in some cases. In comparison with other methods, our method can recover more details in the foreground and background. Our results present good visibility while restoring the color, and hold closer color distribution to the real-world normal-light images.

Fig. 9. Visual comparison of different methods on real-world low-light images.

Affected by various light sources, nighttime images usually show the characteristics of large illumination difference between bright and dark areas. As shown in Fig. 10, in the extremely dark areas, it is difficult for RetinexNet and LightenNet to recover the original pixel information, and unexpected noises appear in their results. In the results of UTVNet and Zero-DCE, the hues of the bright regions are excessively changed due to over-enhancement. Compared with other methods, ours better restores the visibility in the dark areas while avoiding local over-enhancement.

B. Quantitative Evaluation

We quantitatively evaluate the performances of all the methods with four commonly used metrics, i.e., NIQE [54], Angle Error (AE), PSNR and SSIM [50]. PSNR is the ratio of the maximum possible power between the normal-light image and the enhanced image, which is used to measure the fidelity of the low-light image enhancement results. SSIM treats image degradation as perceptual changes in structural information, and incorporates the differences in brightness and contrast at the same time. AE measures the similarity between two images in the RGB color space. NIQE is the only non-reference metric here. It summarizes the visual quality of images based on naturalness and focuses more on visual quality evaluation. Smaller NIQE and AE scores as well as larger PSNR and SSIM scores indicate better image quality.

1) Evaluation on Synthetic Dataset: Except for LIME and NPE which do not belong to the learning based approaches, we train the other methods on our training dataset using the authors provided parameters, and test them on our test dataset with 1837 images. Additionally, since Zero-DCE is a completely unsupervised method that does not require paired training image samples, we use only the low-light parts of our synthetic dataset for its training purpose.

TABLE I

QUANTITATIVE RESULTS ON TEST IMAGES FROM OUR SYNTHETIC LOW-LIGHT IMAGE DATASET. THE SYMBOL "↓" MEANS THAT LOWER INDICATORS CORRESPOND TO BETTER RESULTS, WHILE THE SYMBOL "↑" IS THE OPPOSITE

Methods	PSNR↑	SSIM↑	NIQE↓	AE↓
NPE [53]	17.19	0.8161	5.13	4.31
LIME [19]	17.79	0.8640	5.12	3.09
RetinexNet [1]	17.97	0.8385	5.51	2.36
MBLLEN [2]	19.01	0.7959	5.15	18.49
EnlightenGAN [10]	21.92	0.8989	5.07	1.83
UTVNet [29]	18.66	0.8778	5.14	6.60
LightenNet [8]	18.62	0.8681	5.32	1.80
GLADNet [3]	19.88	0.8955	5.10	3.50
KinD [4]	21.35	0.8982	5.26	1.64
Zero-DCE [12]	17.91	0.8115	5.79	11.61
CSDNet_UPE [37]	21.68	0.8853	4.96	2.41
Ours	**22.70**	**0.9077**	**4.93**	**1.53**

Table I shows the quantitative evaluation results of our model and the other eleven methods on our synthetic dataset. Our model achieves the best level in PSNR, SSIM, AE and NIQE, which means that our enhancement results reach a minimum bias from the real-world normal-light images, and our method gains the best enhancement effects.

2) Evaluation on Real-World Datasets: To compare different enhancement methods on real-world low-light images without corresponding normal-light images in DICM [13] (44 dark images selected from 69 images), LIME [19] (10 images), Fusion [25] (18 images), MEF [55] (79 images) and VV[1] (24 images).

[1]https://sites.google.com/site/vonikakis/datasets

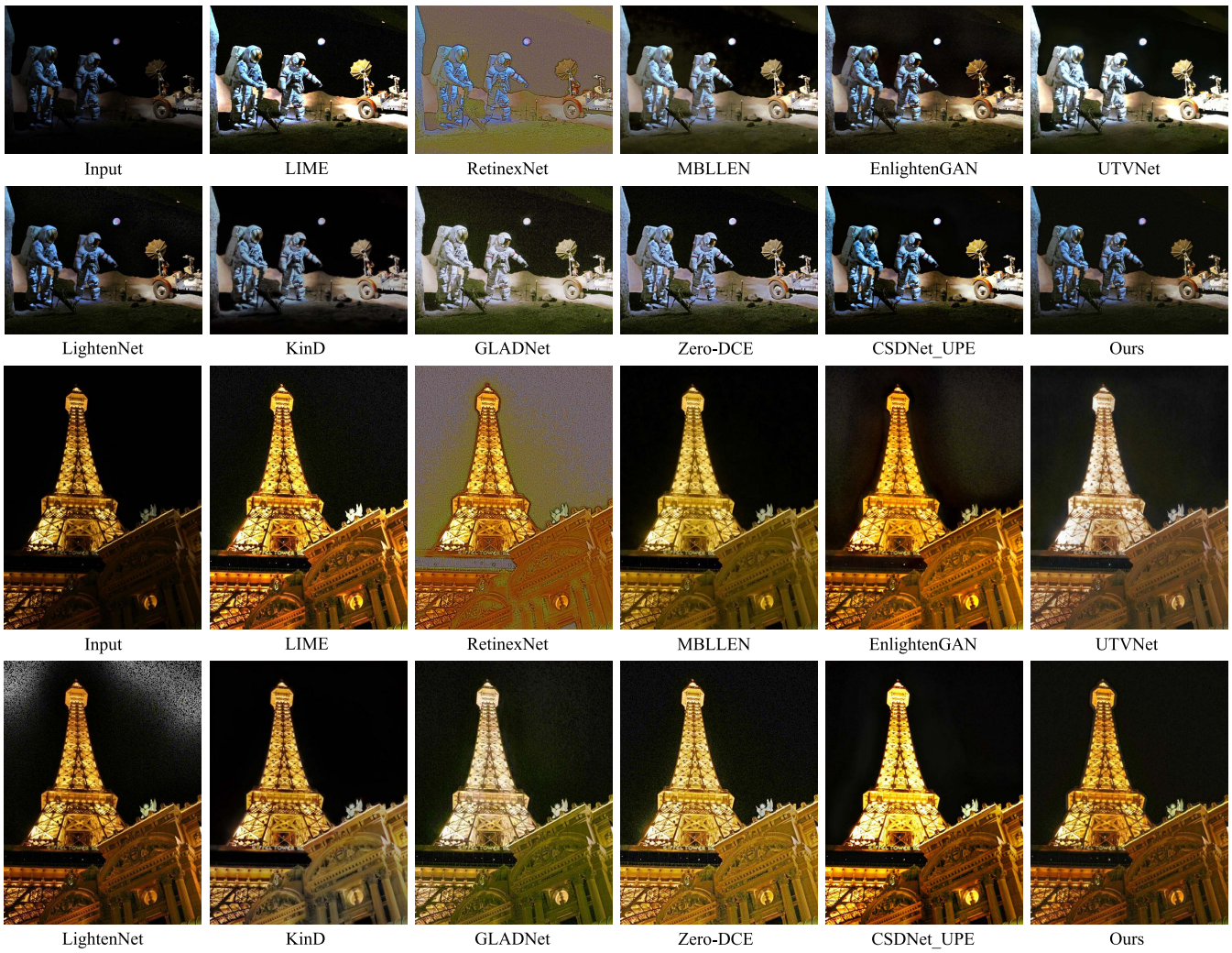

Fig. 10. Visual comparison of different methods on real-world nighttime images.

TABLE II

QUANTITATIVE RESULTS ON TEST IMAGES FROM 5 REAL-WORLD LOW-LIGHT IMAGE DATASETS. THE SYMBOL "↓" MEANS THAT LOWER INDICATORS CORRESPOND TO BETTER RESULTS, WHILE THE SYMBOL "↑" IS THE OPPOSITE

	NIQE↓					BTMQI↓					NIQMC↑				
Methods	DICM	LIME	Fusion	MEF	VV	DICM	LIME	Fusion	MEF	VV	DICM	LIME	Fusion	MEF	VV
NPE [53]	2.79	3.84	3.22	4.41	2.52	4.72	3.64	4.70	4.74	2.88	5.09	4.61	5.08	4.88	5.27
LIME [19]	2.69	3.50	3.15	4.28	2.39	4.48	3.83	4.21	5.14	3.19	4.49	5.03	5.06	5.05	5.20
RetinexNet [1]	4.12	5.07	4.18	5.23	2.67	3.71	3.72	5.05	4.73	3.22	4.54	4.23	4.68	4.04	4.62
MBLLEN [2]	2.79	3.84	3.07	3.62	2.38	4.05	3.64	4.57	4.81	4.27	5.31	5.13	5.07	5.20	5.41
EnlightenGAN [10]	2.78	3.48	2.99	3.59	3.38	4.26	3.18	5.86	**3.10**	**3.12**	5.25	5.08	4.78	4.92	5.43
UTVNet [29]	2.72	3.73	3.63	3.71	2.44	4.03	3.66	4.80	4.63	4.46	4.95	4.93	4.86	4.79	5.04
LightenNet [8]	2.54	3.49	3.04	3.81	2.28	4.42	3.76	4.22	4.65	4.27	5.10	4.83	5.18	5.03	5.30
GLADNet [3]	3.14	3.93	2.98	3.66	2.23	**3.61**	3.42	4.35	3.85	4.06	5.27	5.16	5.12	5.02	5.34
KinD [4]	2.86	3.63	2.74	3.72	**2.16**	4.02	3.21	4.09	4.91	3.43	5.15	4.94	5.22	4.83	5.42
Zero-DCE [12]	2.84	3.43	3.26	4.23	2.58	4.54	3.57	4.55	5.25	3.15	5.21	5.05	5.21	4.74	5.35
CSDNet_UPE [37]	3.45	**3.34**	3.33	4.14	3.22	4.96	3.93	5.08	5.36	4.35	5.28	**5.29**	5.21	**5.22**	5.33
Ours	**2.40**	3.39	**2.69**	**3.55**	2.22	4.25	**3.16**	**4.04**	5.07	3.65	**5.34**	4.98	**5.29**	4.95	**5.48**

As shown in Table II, we calculate the NIQE, BTMQI [56], NIQMC [57] scores of different methods on each dataset. NIQMC evaluates image quality by measuring the local detail and global histogram of the images, which particularly prefers the image with higher contrast. As for BTMQI, it assesses image quality by considering the average intensity, contrast, and structure information of the enhanced images. For each method, we use the authors provided

Fig. 11. Enhancement results of our EFINet trained on different datasets.

pre-trained models or default parameters for the testing process.

For NIQE, our method achieves the best on DICM, Fusion, and MEF datasets. It also ranks second on LIME and VV datasets, and the gap with the first-ranked method does not exceed 0.1. As for BTMQI, our method achieves the best results on LIME and Fusion. Finally, for NIQMC, our method performs best on DICM, Fusion, and VV. Overall, for each real-world dataset, our method always performs best on some evaluation metrics. The evaluation results show that our enhancement results hold better visual quality, which also proves that our method has better generalization performance and stability than other methods on untouched images.

C. Comparison With Other Datasets

To demonstrate the advantages of our synthetic dataset, we additionally train our model with two commonly used low-light image datasets, i.e., LOL [1] and MBLLEN [2] datasets. The training set of the LOL dataset contains 485 real-world image pairs and 1000 synthetic image pairs, and the training set of the MBLLEN dataset contains 16925 pairs of images generated by applying the random gamma correction on the PASCAL VOC dataset [58]. Since the number of images in each dataset varies, we keep all hyper-parameters consistent except for the number of epochs, and train our network to convergence separately. We choose part of the images in the real-world dataset TM-DIED[2] as the test dataset. Compared with the other two datasets, the normal-light ground truths in our dataset are more saturated and vivid in color, and richer in details. It can be seen from Fig. 11 that the model trained on the MBLLEN dataset can easily lead to color cast and over-smoothing phenomena. The generalization performance of the model trained on the LOL dataset is not good enough, and the color blocks caused by erroneous enhancement are commonly found in the results, especially in the darker areas. In contrast, our synthetic dataset takes into account the recovery of both under- and over-exposed areas and better describes the complex illumination distribution in the real worlds, which makes the enhancement results aesthetically pleasing while strengthening visibility.

[2]https://www.flickr.com/photos/73847677@N02/sets/72157718844828948/

D. Ablation Studies

1) Number of Iterations: We conduct comparative experiments on our synthetic dataset to figure out the influence of iteration times on the model performance. We keep fixed hyper-parameters and datasets to compare the impact of iteration times on PSNR, SSIM, and running time. In the training phase, we start with one iteration on our training dataset, gradually increase the number of iterations, and test them on our test dataset.

Table III shows our experimental results. On the one hand, for the network models with less than three iterations, their enhancement results are less capable of recovering image quality and maintaining image structure compared with the models with more iterations. On the other hand, for the network model with more than three iterations, it is difficult to further improve the performance of the network by increasing the iteration times, which also heavily increases the training

TABLE III

PSNR AND SSIM INDICATORS OF OUR EFINET MODELS WITH DIFFERENT ITERATIONS TESTED ON OUR SYNTHETIC DATASET

Iterations	1	2	3	4
PSNR↑	21.12	22.04	**22.70**	22.68
SSIM↑	0.8999	0.9008	**0.9077**	0.9076
Running Time	0.0017	0.0027	0.0046	0.0065
Training Time	67h	98h	129h	156h

TABLE IV

NIQE OF OUR EFINET MODELS TRAINING TRAINED ON THE DATASETS WITH DIFFERENT FILTERING KERNEL SIZES ON 50 REAL-WORLD LOW-LIGHT IMAGES. MIXING MEANS THAT THE KERNEL SIZE OF THE FILTER IS RANDOMLY SELECTED BETWEEN 32 AND 64

Kernel Size	**No Filtering**	**2×2**	**8×8**
NIQE ↓	2.963	3.109	2.908
Kernel Size	**32×32**	**64×64**	**Mixing**
NIQE ↓	2.861	2.860	**2.857**

and running time. Therefore, we set the number of iterations to 3 in the final.

2) Data Authenticity: We aim to improve the learning capabilities and the enhancement performance of the network model by simulating more realistic low-light images. Consequently, we compare the impact of different training datasets on the enhancement results, which includes the dataset with overall darkened images and the low-light images with non-uniform illumination. For low-light images with non-uniform illumination, we further compare the images processed by the guided filter with different kernel sizes.

As shown in Fig. 12, it looks obviously different when the transition between the bright and dark areas in the images are processed by the guided filter with different kernel sizes. The illumination boundaries obtained with smaller kernel sizes appear abrupt, while those obtained with larger kernel sizes are smoother. We attempt to simulate the real-world lighting environment with the above diverse data for the model training purpose. Fifty real-world dark images are selected for the model testing purpose, and the specific NIQE scores are shown in Table IV. Evidently, the model trained on the dataset with diverse kernel sizes for the guided filter adapts better to the real-world low-light images.

E. User Study

Furthermore, we also conduct a user study with 30 participants to test the human visual preference for the enhancement results from different methods. Twenty real-world low-light images are randomly selected and then enhanced by our method and other eleven representative methods. In the beginning, for each original low-light image, participants receive twelve enhanced versions presented in a random order. Then, we ask participants to subjectively choose the one with the best comprehensive quality based on the aspects of brightness, color, and visibility, etc. For the sake of comparison, we count the number of votes received for each method in the end. Fig.13 shows the user study results. Intuitively, our method receives the most votes, which proves that our results are more visually favorable.

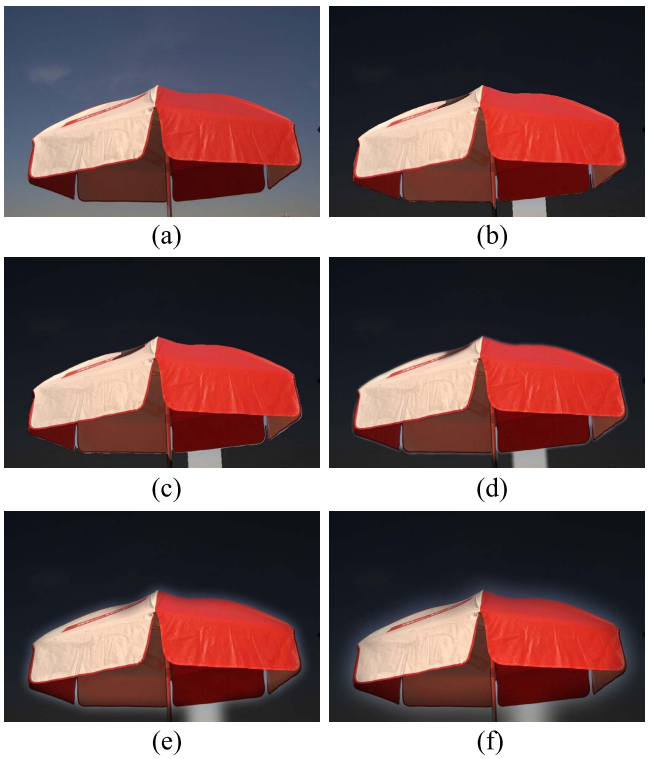

Fig. 12. Influences of different kernel sizes of the guided filter on the synthesized low-light images. (a) Original normal-light image. (b) Low-light image generated without the guided filtering process. (c)-(f) Low-light images generated with the guided filtering process, whose kernel sizes are 2 × 2, 8 × 8, 32 × 32 and 64 × 64 respectively.

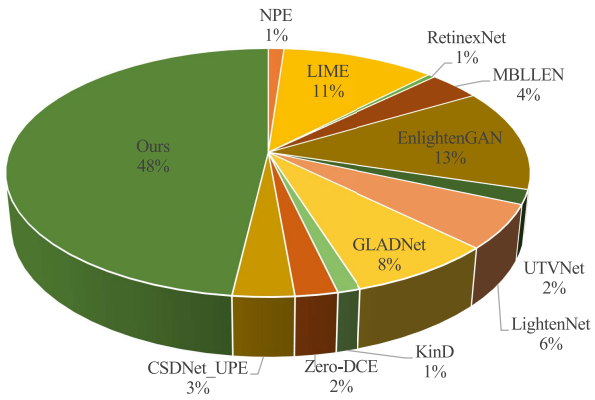

Fig. 13. User study results.

VI. CONCLUSION

In this paper, we propose a new method to degrade the normal-light images to be the low-light images with non-uniform illumination, which simulates the complex illumination distribution in the real-worlds through a locally random

brightness assignment strategy and embodies the preservation of mid-level visual perception through superpixel segmentation and merging operations. And, we design a weight-shared iterative network architecture named EFINet for the low-light image enhancement task. Within each iteration of EFINet, the Coefficient Estimation Network first estimates a set of stretching coefficient maps adaptively, which are used to enhance the low-light input image initially. Then, the Fusion Network further combines the well-lit local areas of the input image and the initial enhancement result, to generate more satisfying results. Our proposed method can recover image details and color information from the low-light images to generate the enhancement results with uniform and natural illumination. Extensive experiments show that our method outperforms the SOTA methods both in quantitative metrics and in human visual perception.

Considering the extremely dark scenes and more complex lighting conditions that are pretty common at night, we will try to make further efforts to enhance and restore the nighttime images in our future work. Furthermore, given that current evaluation metrics for the image enhancement task can hardly be consistent with the visual perception of human eyes, we are also interested in designing reasonable and widely used metrics for image quality evaluation.

REFERENCES

[1] W. Chen, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2018, pp. 1–12.

[2] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using CNNs," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2018, pp. 1–13.

[3] W. Wang, C. Wei, W. Yang, and J. Liu, "GLADNet: Low-light enhancement network with global awareness," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 751–755.

[4] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia (MM)*, 2019, pp. 1632–1640.

[5] E. D. Pisano *et al.*, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *J. Digit. Imag.*, vol. 11, no. 4, pp. 193–200, 1998.

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

[7] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.

[8] C. Li, J. Guo, F. Porikli, and Y. Pang, "LightenNet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Recognit. Lett.*, vol. 104, pp. 15–22, Mar. 2018.

[9] M. Zhu, P. Pan, W. Chen, and Y. Yang, "EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2020, vol. 34, no. 7, pp. 13106–13113.

[10] Y. Jiang, X. Gong, D. Liu, Y. Cheng, and C. Fang, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.

[11] Z. Zhu, Y. Meng, D. Kong, X. Zhang, Y. Guo, and Y. Zhao, "To see in the dark: N2DGAN for background modeling in nighttime scene," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 492–502, Feb. 2021.

[12] C. Guo *et al.*, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1780–1789.

[13] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5372–5384, Dec. 2013.

[14] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Müller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Visualizat. Biomed. Comput.*, 1990, pp. 337–345.

[15] J.-T. Lee, C. Lee, J.-Y. Sim, and C.-S. Kim, "Depth-guided adaptive contrast enhancement using 2D histograms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4527–4531.

[16] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, 1977.

[17] D. J. Jobson, Z.-U. 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, Jul. 1997.

[18] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Process.*, vol. 129, pp. 82–96, Dec. 2016.

[19] 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, Feb. 2017.

[20] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2782–2790.

[21] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.

[22] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[23] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3730–3734.

[24] X. Zhang, P. Shen, L. Luo, L. Zhang, and J. Song, "Enhancement and noise reduction of very low light level images," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 2034–2037.

[25] K. D. Abov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," *Proc. SPIE*, vol. 6064, pp. 354–365, Feb. 2006.

[26] Y. Atoum, M. Ye, L. Ren, Y. Tai, and X. Liu, "Color-wise attention network for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 506–507.

[27] K. Xu, X. Yang, B. Yin, and R. W. H. Lau, "Learning to restore low-light images via decomposition-and-enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2281–2290.

[28] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality," *IEEE Trans. Image Process.*, vol. 30, pp. 3461–3473, 2021.

[29] C. Zheng, D. Shi, and W. Shi, "Adaptive unfolding total variation network for low-light image enhancement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4439–4448.

[30] J. Li, X. Feng, and Z. Hua, "Low-light image enhancement via progressive-recursive network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 11, pp. 4227–4240, Nov. 2021.

[31] K. Xu, H. Chen, C. Xu, Y. Jin, and C. Zhu, "Structure-texture aware network for low-light image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Jan. 7, 2022, doi: 10.1109/TCSVT.2022.3141578.

[32] S. K. Dhara and D. Sen, "Exposedness-based noise-suppressing low-light image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 3438–3451, Jun. 2022, doi: 10.1109/TCSVT.2021.3113559.

[33] L. Zhao, S.-P. Lu, T. Chen, Z. Yang, and A. Shamir, "Deep symmetric network for underexposed image enhancement with recurrent attentional learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 12075–12084.

[34] Z. Zhang, Y. Jiang, J. Jiang, X. Wang, P. Luo, and J. Gu, "STAR: A structure-aware lightweight transformer for real-time image enhancement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4106–4115.

[35] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6849–6857.

[36] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.
[37] L. Ma, R. Liu, J. Zhang, X. Fan, and Z. Luo, "Learning deep context-sensitive decomposition for low-light image enhancement," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 30, 2021, doi: 10.1109/TNNLS.2021.3071245.
[38] L. Chen, L. Guo, D. Cheng, and Q. Kou, "Structure-preserving and color-restoring up-sampling for single low-light image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 1889–1902, Apr. 2022.
[39] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang, "RetinexDIP: A unified deep framework for low-light image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1076–1088, Mar. 2022.
[40] R. Zhang, L. Guo, S. Huang, and B. Wen, "ReLLIE: Deep reinforcement learning for customized low-light image enhancement," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 2429–2437.
[41] S. Zheng and G. Gupta, "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 581–590.
[42] L. Fu, H. Yu, F. Juefei-Xu, J. Li, Q. Guo, and S. Wang, "Let there be light: Improved traffic surveillance via detail preserving night-to-day transfer," *IEEE Trans. Circuits Syst. Video Technol.*, early access, May 19, 2021, doi: 10.1109/TCSVT.2021.3081999.
[43] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. CVPR*, Jun. 2011, pp. 97–104.
[44] J. Wu, C. Liu, and B. Li, "Texture-aware and structure-preserving superpixel segmentation," *Comput. Graph.*, vol. 94, pp. 152–163, Feb. 2021.
[45] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[47] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2614–2623, May 2019.
[48] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[50] 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, Apr. 2004.
[51] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 8026–8037.
[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[53] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.
[54] 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, 2012.
[55] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
[56] K. Gu *et al.*, "Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432–443, Mar. 2016.
[57] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017.
[58] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Sep. 2009.

Chunxiao Liu received the Ph.D. degree in mathematics from the State Key Laboratory of CAD and CG, Zhejiang University, Hangzhou, China, in 2019. He is currently an Associate Professor and a Master Supervisor in computer science and technology with the School of Computer Science and Information Engineering, Zhejiang Gongshang University. His current research interests include image and video processing, computer vision, computer graphics, machine learning, and intelligent systems.

Fanding Wu received the B.S. degree in information management and information systems from the Southwest University of Science and Technology, Mianyang, China, in 2019. He is currently pursuing the M.E. degree in computer science with the School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, China. His main research interests include computer graphics and deep learning-based computer vision applications.

Xun Wang (Member, IEEE) is currently a Professor and a Ph.D. Supervisor in computer science and technology and the Dean of the School of Computer Science and Information Engineering, Zhejiang Gongshang University, Hangzhou, China. He is the Head of the first-rate disciplines of computer science and technology, Zhejiang. He is also the Director of the Zhejiang Engineering Laboratory for Visual Media Big Data Technology. He was selected for the National Millions of Talents Program. In recent years, he has authored more than 80 articles in journals and conferences. He holds nine authorized invention patents. His current research interests include multimedia processing, computer vision, machine learning, computer graphics, and intelligent systems. He is a member of ACM and a Distinguished Member of CCF. He has five provincial and ministerial level scientific and technological progress awards.