IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 7, 2018, accepted December 21, 2018, date of publication January 10, 2019, date of current version February 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2891957

A Pipeline Neural Network for Low-Light Image Enhancement

YANHUI GUO, XUE KE, JIE MA, AND JUN ZHANG

Guangdong Provincial Key Laboratory of Digital Manufacturing Equipment, National Key Laboratory of Science and Technology on Multi-Spectral Information Processing, School of Automation, Guangdong HUST Industrial Technology Research Institute, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Jie Ma (majie@hust.edu.cn)

This work was supported in part by the Guangdong Innovative and Entrepreneurial Research Team Program under Grant 2014ZT05G304.

ABSTRACT Low-light image enhancement is an important challenge in computer vision. Most of the low-light images taken in low-light conditions usually look noisy and dark, which makes it more difficult for subsequent computer vision tasks. In this paper, inspired by multi-scale retinex, we present a low-light image enhancement pipeline network based on an end-to-end fully convolutional networks and discrete wavelet transformation (DWT). First, we show that multiscale retinex (MSR) can be considered as a convolutional neural network with Gaussian convolution kernel, and blending the result of DWT can improve the image produced by MSR. Second, we propose our pipeline neural network, consisting of denoising net and low-light image enhancement net, which learns a function from a pair of dark and bright images. Finally, we evaluate our method both in the synthetic dataset and public dataset. The experiments reveal that in comparison with other state-of-the-art methods, our methods achieve a better performance in the perspective of qualitative and quantitative analyses.

INDEX TERMS Convolutional neural network, low-light image enhancement, LLIE-Net.

I. INTRODUCTION

In reality, when we capture an image in a low light environment, the image quality would be strongly influenced by noise and low contrast, which makes it more difficult to deal with the following tasks such as image segmentation, object detection etc. At the present, digital video technology has been widely used in various fields, for example, safety monitoring of important places, traffic management, driving assistance and so on. Under the condition of good daytime illumination, the image quality can meet the application requirements, but when it comes to night, the image quality of low light image worsens, which brings a big challenge in digital image process.

Sorts of enhancement methods were proposed and they can be divided into three categories: methods based on histogram equalization algorithms (HE), Retinex theory, and using dehazing model. In this paper, we present a new method different from existing methods. The contribution of our work can be summed up as three aspects: First of all, we describe the relationship between the MSR and CNN. Secondly, we elaborate a pipeline network, which learn a function for denoising and low light enhancement. And finally, the performance evaluated on a number of low light natural images reveal that our method achieves better performance in comparison with other state-of-the-art approaches. Figure 1 gives some examples of our results with inputting the low light noise images.

Overall, the contribution of our work can be generalized to three aspects:(1) We explain the relationship between CNN and MSR and we find out blending the low-low component of the image DWT with the output of MSR can improve the result. (2) We consider the low light image enhancement as a machine learning problem and we elaborate our neural network (LLIE-net) for this task, furthermore, we present a pipeline neural network including denoising and enhancement. (3) And we measure our results quantitatively and qualitatively. The performance of our LLIE-net is better than the existing methods in both the synthetic low light images and real-world low light images.

II. RELATED WORK

Based on retinex theory [1], several methods were proposed. Single scale retinex (SSR) [2], multiscale scale retinex (MSR) [3] and multiscale retinex with color restoration (MSRCR) [4] enhance images in frequency domains. HE algorithms [5], [6] mainly focus on enhancing

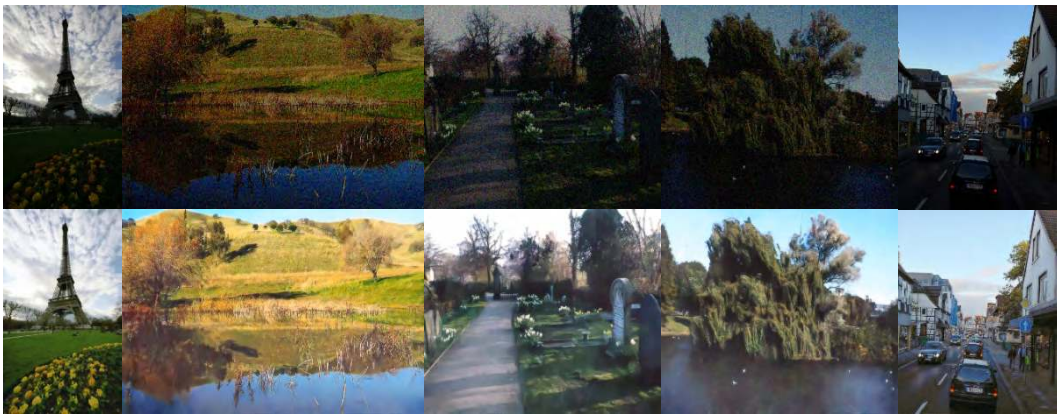

FIGURE 1. Some example results of our pipeline network.

image contrast. However, the details in dark regions are not enhanced appropriately [7]. Recently, some methods achieve state of-the-art enhancement performance, which are based on dehazing model. Dong *et al.* [8] found that low-light images are similar to hazy images after inverting. Similarly, Li *et al.* [9] and Zhang *et al.* [10] make use of dark channel prior [11] to enhance low light image. Nevertheless, these methods are lack of theoretical basis. Recently, some methods based on deep learning have been proposed. Chen *et al.* [34] build datasets of raw short-exposure low-light images, with corresponding long-exposure reference images and develop a network to learn the enhancement function. But the results perform well only on the constructed datasets.

III. THE PROPOSED MODEL

We elaborate that the multi-scale retinex as an image enhancement method can be regarded as a forward propagation in convolutional neural network [33] with different Gaussian convolution kernels. And we find that this method will lead to an effect of overexposure. In order to get more natural result, we fuse the image produced by convolution with the low-low component of the image DWT (LL), which represents the main content of the image, produced by discrete wavelet transform. Finally, we propose a full convolution network which can learn end-to-end function to enhance the low light image.

A. MULTI SCALE RETINEX IS CNN

As we all know, the dominant assumption of Retinex theory is that the image can be decomposed into reflection and illumination. Based on this theory, SSR [2] has been proposed, of which the general mathematical form is:

$$R(x,y) = \log(I(x,y)) - \log(I(x,y) * G(x,y)) \quad (1)$$

where the $G(x,y)$ represents the Gaussian kernel function, $I(x,y)$ is the input RGB image, $*$ is the convolution operation, and the $R(x,y)$ is the reflectance image. The $G(x,y)$ is given by:

$$G(x,y) = Kexp[-(x^2+y^2)/2\sigma^2] \quad (2)$$

where σ, the filter standard deviation, controls the amount of spatial detail which is retained, and K is a normalization factor.

We can represent the Equation (1) with following form:

$$R(x,y) = \phi(I(x,y)) + \psi(I(x,y)) \quad (3)$$

where the $\phi(\cdot)$ represents a logarithmic transformation for image $I(x,y)$ and $\psi(\cdot)$ is a non-linear transformation for image $I(x,y)$.

Further, multi-scale retinex (MSR) [3] is a weighted average of these different SSR outputs, which is given as:

$$R_{MSR} = \sum_{i=1}^{N} w_i R_i = \sum_{i=1}^{N} w_i[\phi_i(I(x,y)) + \psi_i(I(x,y))] \quad (4)$$

where R_i denotes the i^{th} component of the i^{th} scale, N represents the number of scales, w_i is the weight of the i^{th} scale, R_{MSR} represents the result of MSR. The Equation (3) can be extended to all spectral channels of the input image. (Generally, the number of spectral channels is 3 for RGB color space.)

We can use convolutional neural network to transform the image to a logarithmic domain and a higher mathematic space, which given as:

$$L^{k+1}(x) = h_k(\left(W^k * x\right) + b_k) \quad (5)$$

where h_k is the k^{th} layer activation function and the W^k, b^k are the parameters of k^{th} convolution layer. The $L^{k+1}(x)$ is the output of k^{th} layer.

So, we can set the $\phi(\cdot)$ and $\psi(\cdot)$ as following:

$$\phi(\cdot) = L_\phi^N\left(L_\phi^{N-1}\left(\cdots L_\phi^1(I)\right)\right) \quad (6)$$

$$\psi(\cdot) = L_\psi^M\left(L_\psi^{M-1}\left(\cdots L_\psi^1(I)\right)\right) \quad (7)$$

where the N and M represent the number of layers in ϕ network and ψ network. The Figure 2 shows the structure of MSR. We split the three channels of input image to forward respectively pass a series of Gaussian convolution

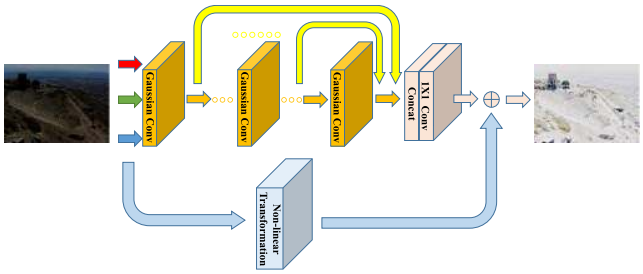

FIGURE 2. The structure of multi-scale retinex(MSR) regarded as a convolution neural network.

layers with different standard deviation. At last, the concatenation and 1x1 convolution layer represent the weighted w_i in Equation (4). And, finally, we add together the output of 1x1 convolution layer and the result of forwarding pass a non-linear layer.

B. WAVELET FOR LOW LIGHT IMAGE ENHANCEMENT

There are some problems in MSR method, for example, we find that an inappropriate scale would lead to the 'halo' effect near edges [4]. And except this, because of the difficulty of estimation for the illumination, the enhancement for the highlight and dark area is not good. So, we introduce wavelet transformation into MSR for reducing above adverse effect.

FIGURE 3. Left: Original Image. Middle: Low Light Image. Right: Results of DWT.

Fourier transform is a good tool for analyzing images in the domain of frequency, however, the spectrum of an image loses lots of information, which makes it difficult to apply convolutional neural network and restored well. Another frequently used method called wavelet transformation can make it easier to solve the problem, which can split the frequency of an image. As we all know, the low frequency part consists the majority of information of an image. And there are mainly noise and drastic brightness changes in the high frequency part. We can retain the low frequency part to get an enhanced image. As shown in Figure 3, the image can be transformed into four sub-images by discrete wavelet transformation [12] using Haar basis function. The four sub-images correspond to the approximation sub-band LL, horizontal detail LH, vertical detail HL and diagonal detail HH, respectively. We find that the LL image has been brighten a little and the noise mainly exist in high frequency area. So, inspired with this, we use MSR in LL image and then resize the result to the original shape. There is still a problem existing, which lead to a color distortion using MSR. To solve this problem, Jobson *et al.* [4] proposed MSRCR to complete the algorithm with a color restoration step. Their first move is to compute the chromaticity coordinates:

$$I_i'(x,y) = \frac{I_i(x,y)}{\sum_{j=1}^{S} I_j(x,y)} \tag{8}$$

For the i^{th} color channel, where S is the number of spectral channels, and the restored color MSR is given by:

$$R_{MSRCR_i} = C_i(x,y)R_{MSR_i}(x,y) \tag{9}$$

$$C_i(x,y) = f\left(I_i'(x,y)\right) = \beta\log[\alpha I_i'(x,y)] \tag{10}$$

To simplify, we change the Equation (10):

$$R_{MSRCR_i} = \mathrm{T}(I_i(x,y))R_{MSR_i}(x,y) \tag{11}$$

where $\mathrm{T}(\cdot)$ is a serial of transformations for original image $I(x,y)$, however, we find Equation (9) can be approximately equal to the Equation (12) (simply shown in Figure 4 (a)):

$$R_{MSRCR_i} = \lambda \cdot resize(DWT^{LL}(I_i(x,y))) + (1-\lambda)R_{MSR_i}(x,y) \tag{12}$$

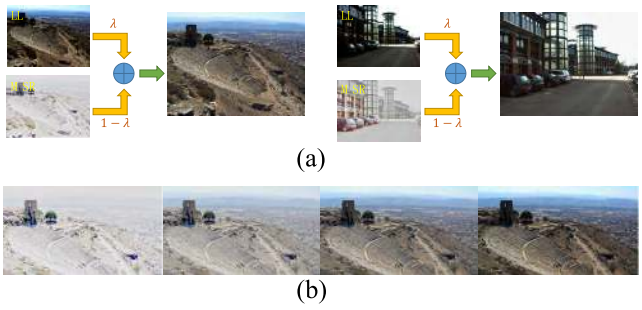

FIGURE 4. (a): Sketch map for Equation (12). (b): The results of different λ.

where the *DWT* is discrete wavelet transformation using Haar basis function, and the *resize* operation is to reshape the LL image into the original $I(x,y)$'s shape because of the *DWT* cause the image to be shrunken twice the size. Absolutely, the Equation (9) and Equation (12) are not equivalent in mathematical form. But we can get results with more natural color than MSR. As shown in Figure 4 (b), we show the enhanced images using different blending parameters λ. The blending results can reduce the color distortion. However, with the λ increasing, the result will be dark and blurred because of the effect of LL image produced by biquadratic interpolation. We should choose an appropriate parameter to get a natural image.

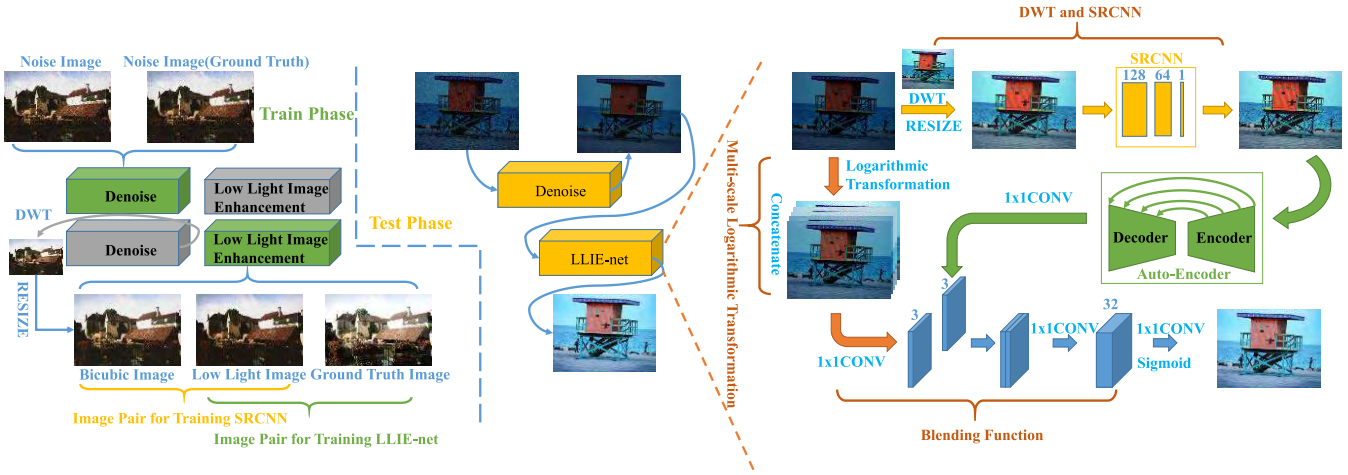

FIGURE 5. Left: The training and test process of our pipeline network. Right: The detail of LLIE-net.

C. OUR LOW LIGHT IMAGE ENHANCEMENT MODEL

In the previous section, we present the fact that MSR can be considered as a forward network and the wavelet transformation can improve the result of MSR. Inspired by the experiment combining the two approaches, we proposed an end-to-end fully convolution network to enhance a low light image. Our method differs entirely from existing approaches, which takes low light image enhancement as a supervised learning problem. Profiting from the fitting ability of CNN, we have no more need to adjust some parameters in traditional methods and the all needed parameters would be learnt in training phase. More detail about training dataset will be shown in **Sec. IV**.

Our mode consists of two tasks: denoising and low light image enhancement. We cascade the two network in test phase and fix one to train another one in training phase.

As shown in the left of Figure 5, for the first part, we train a denoising network using the method of Noise2Noise presents by Lehtinen *et al.* [15]. For simplicity, we only train a denoising network using Gaussian noise, which is commonest noise in the simulation. The structure of denoising network is SRRes-Net [14]. We express the empirical risk minimization task mathematically:

$$\underset{\theta}{\operatorname{argmax}} \sum_{i} L(f_\theta(\hat{x}_i), \hat{y}_i) \tag{13}$$

where both the inputs $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$ and the targets $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ are drawn from a corrupted distribution (not necessarily the same), conditioned on the underlying, unobserved clean target $\boldsymbol{y}_{\boldsymbol{i}}$ such that $\mathbb{E}\{\hat{\boldsymbol{y}}_{\boldsymbol{i}}|\hat{\boldsymbol{x}}_{\boldsymbol{i}}\} = \boldsymbol{y}_{\boldsymbol{i}}$. For the second part, we elaborate a structure for the low light image enhancement, which we call LLIE-net. The detail about LLIE-net can be seen in right of Figure 5 and our model consists of four components: **DWT and SRCNN**, **Multi-scale Logarithmic Transformation**, **Auto-Encoder net**, and **Blending Function**. In **Sec III.B**, we find that fusing the result of MSR with LL image produced by DWT will achieve a better performance in practice. Compared to MSR, our model attempt to use multi-scale logarithmic transformation, because we want to blend more diverse priori images and it's similar to non-linear transformation in MSR. The Auto-Encoder net plays an analogical role with Gaussian convolution in MSR and the Blending Function can learn best parameters blending with the results of another component to get an image with more natural color.

1) DWT AND SRCNN

We suppose this sub-function $\mathcal{H}_1$. The input of this function is the low-light image denoised X and the output is a the same size image $X_{\mathcal{H}_1}$. Firstly, we acquire the LL image by using DWT, which is $X^{LL} = DWT^{LL}(X)$. This operation can reduce the noise of image and bright the image a little. On the account of the shape of X^{LL} is half as X's shape, we should upsampling the image. But if we use bicubic interpolation, the result would become blur. It's difficult for the network to learn the task of super-resolution reconstruction and low image enhancement at the same time in training. We separate the task of super-resolution reconstruction out, and build a small network SRCNN [13] to get a clearer image for next function. Our SRCNN is the basic structure including only three convolution layers: 128,64, and 1 kernels respectively.

2) MULTI-SCALE LOGARITHMIC TRANSFORMATION

Multi-scale logarithmic transformation $\mathcal{H}_2$ takes the original low-light image X as input and computes the same size output $X_{\mathcal{H}_2}$. The dark image is enhanced by several difference logarithmic transformation. The formula is as follows:

$$S_j = c \cdot \log_{v_j+1}(1 + v_j \cdot r) \tag{14}$$

where $r \in [0, 1]$, c usually is 1, and $j = [1, 2, \ldots, n]$, S_j represents the output of j^{th} scale on the logarithmic base $v_j + 1$, and the number of logarithmic transformation function n. In our model, we set $n = 4$. and $v = \{10, 50, 100, 200\}$. Next, we concatenate these 3D tensors $\mathcal{M}_j$ (width $\times$ height $\times$ 3 channels) to a larger 3D tensor $\mathcal{M}$ (width $\times$ height $\times$ 3n

channels). We make the tensor $\mathcal{M}$ go through a Relu layers, the output of which is shown in Figure 6.

$$X_{\mathcal{H}_2} = \max(0, [\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4] * W_{log} + b_{log}) \quad (15)$$

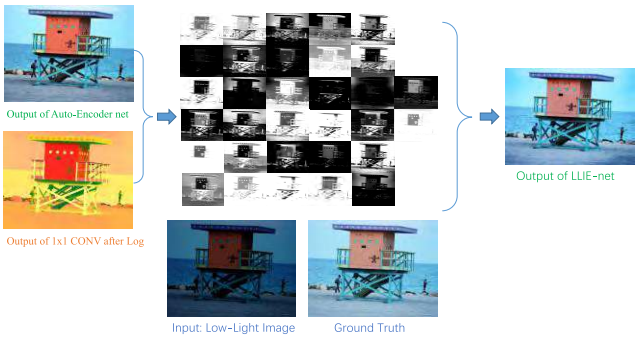

FIGURE 6. The detail of our blending function.

where the $*$ is a convolution operator, W_{log} is convolution kernel that shrinks 3n channels to 3 channels, which plays a role in computing a weighted enhancement image by using logarithmic transformation.

3) AUTO-ENCODER NET

Auto-Encoder
net takes the input $X_{\mathcal{H}_1}$, and we suppose this network $\mathcal{H}_3$. We design the structure referring to the MSR, and we establish an Auto-Encoder model [16], [31], [32], which is similar to the mu1lti-convolutional layers of MSR combing the shallow features with deep features. We denotes the output of $\mathcal{H}_3$ as $X_{\mathcal{H}_3}$.

4) BLENDING FUNCTION

As previously mentioned, the blending parameters can be learned in CNN network. Using 1x1 convolution kernel, we can compute the outputs weighted the value of input channels. As shown in Figure 6, we get an output of Auto-Encoder net and another one of 1x1 CONV. And we concatenate them to pass through the next 1x1 CONV. After that, we acquire features with 32 channels. Finally, we blend the features weighted to correct the color of image, the method of which is similar to MSRCR. Formally, we suppose this function $\mathcal{H}_4$, and the output is $X_{\mathcal{H}_4}$:

$$X^o_{\mathcal{H}_4} = \max(0, \left[X_{\mathcal{H}_2}, X_{\mathcal{H}_3}\right] * W^o_{blending} + b'_{blending}) \quad (16)$$

$$X_{\mathcal{H}_4} = \text{sigmoid}(X^o_{\mathcal{H}_4} * W_{blending} + b_{blending}) \quad (17)$$

where the $X^o_{\mathcal{H}_4}$ is the result of weighted image computed by $X_{\mathcal{H}_2}$ and $X_{\mathcal{H}_3}$, which is a tensor with 32 channels. Next, $W_{blending}$ is formed from three kernel parameters, which are used to compute the R, G and B channel respectively for final output from the tensor $X^o_{\mathcal{H}_4}$. And finally, we map the output into [0,1] using the activation function of $sigmoid\ (x) = \frac{1}{1+e^{-x}}$.

5) LOSS FUNCTION

The goal of our model is to learn a deep network to make the output $\mathcal{H}(p)$ and the label $Y(p)$ as close as possible using ℓ_1 loss, which is proved to be better than ℓ_2 loss for image restoration with neural networks [17]:

$$\mathcal{L}^{\ell_1}_{loss} = \frac{1}{N} \sum_{p \in P} |\mathcal{H}(p) - Y(p)| \quad (18)$$

where p is the pixel of P, which is the patch cropped from the image. $\mathcal{H}(p)$ and $Y(p)$ are the values of the pixels in the processed patch and the ground truth respectively.

IV. EXPERIMENTS

We elaborately establish dataset an image dataset and train a pipeline network for denoising and enhancing the low light images by using Keras [24].

In order to evaluate our method, we compare our method with six recent state-of-the-art image enhancement methods in our synthetic test data and public dataset. Finally, we display the results of our pipeline network with the input of low light noise images in reality.

A. IMAGE DATASET GENERATION

In order to learn the parameters of LLIE-net, we construct a new image dataset, which contains a number of normal light (NL) and low-light (LL) natural images. The all images selected are real-world scenes. We have collected the normal light images, some of which are from Google search, UCID [18] and BSD [19] dataset. And many of these images are strong motion blur, out of focus blur, low contrast, underexposure or overexposure and substantial sensor noise are deleted. Finally, we acquire 800 images. For every images, we generate 10 low-light images by two steps. Firstly, we scale the V (Value) component with a random factor between 0.5 and 1, by transforming the image to HSV space. And then we use gamma transform with parameters ranging between 1 and 3 to darken the image further [33]. Then, we obtain a dataset with 8,000 pairs of NL/LL images, some examples of which are shown in Figure 7. We randomly select 7,000 images in the dataset to generate one million 64 × 64 NL/LL patches for training. The remaining 1000 images are used to evaluate our network in test phase. In addition, to guarantee evaluating our approach objectively and impartially, we test our model using the real-world low-light images form the public MEF dataset [20], [21], DICM dataset [22] and LIME dataset [23].

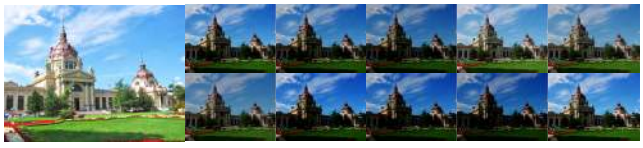

FIGURE 7. NL/LL example on the dataset. The left is NL image and the right are LL images.

TABLE 1. Quantitative results using SSIM/PSNR (dB)/ILNIQE on synthetic images.

Method	1st row	2nd row	3rd row	4th row	1000test images
GT	1/-/19.93	1/-/26.17	1/-/19.60	1/-/15.98	1/-/20.72
Input	0.28/8.69/26.62	0.33/11.59/39.69	0.31/10.88/31.50	0.47/8.73/20.21	0.55/12.38/24.42
LIME	0.71/15.76/47.89	0.73/18.51/37.10	0.66/16.50/31.11	0.79/17.76/29.49	0.76/15.66/30.81
MSRCR	0.88/16.88/25.37	0.62/10.91/32.72	0.77/12.39/20.15	0.89/17.15/19.77	0.79/13.51/22.44
Dong	0.58/12.57/35.44	0.71/18.59/31.84	0.58/15.08/29.62	0.75/14.42/26.12	0.75/17.17/28.52
MF	0.61/13.03/30.00	0.63/15.16/27.25	0.54/13.91/28.9	0.78/13.62/22.6	0.80/17.23/24.47
NPE	0.57/12.37/29.19	0.76/20.78/31.91	0.59/14.85/26.84	0.78/15.53/23.27	0.82/19.22/24.51
SRIE	0.49/11.64/28.49	0.52/14.93/30.27	0.42/12.79/29.75	0.65/11.90/20.51	0.72/16.02/23.37
Ours	0.92/21.22/22.07	0.94/28.48/26.47	0.94/28.82/18.44	0.92/25.03/18.75	**0.91/23.68/21.64**

FIGURE 8. Results using different methods on synthetic test images generated by Ground Truth. (a) GT. (b) Input. (c) LIME. (d) MSRCR. (e) Dong. (f) MF. (g) NPE. (h) SRIE. (i) Ours.

B. TRAINING STEP

The left of Figure 5 shows the step in training phase. We should train two sub-networks: denoising-net and LLIE-net. We would fix one and train another one, finally connecting them in series. In LLIE-net, we should train SRCNN firstly, and train Auto-Encoder net later. We set the depth of Auto-Encoder net to $D = 8$ and use Adam with a learning rate 10^{-3}. We train about 10K iterations with batch size 64.

C. EXPERIMENTS ON SYNTHETIC DATA

In Figure 8, we show visual compassion for four synthetic low light images. For the sake of fairness, we only use LLIE-net in our pipeline network without denoising. As we can see, the color of images produced by LIME [23] and MSRCR [25] is not natural. The method proposed by Dong *et al.* [8] always generates unexpected black edge. Multi-scale fusion model (MF) [26], naturalness preserved enhancement algorithm (NPE) [27] and SRIE [28] all tend to be dark in some areas of images. Visually, ours achieves the best performance.

To measure the results quantitatively, we use structural similarity index (SSIM) [7] and peak signal to noise ratio (PSNR) for evaluation, because the ground truth is known for the synthetic test image. And we use integrated local natural image quality evaluator (ILNIQE) [29] to assess the natural preservation, which is a state-of-the-art no-reference image quality assessment. As shown in Table 1, higher SSIM and PSNR mean that the enhanced image is closer to the ground truth, while a lower ILNIQE usually means a higher image quality. We have bolded the best results. Our method achieves the higher SSIM and PSNR, and lower ILNIQE average than other methods for 1000 test images.

D. EXPERIMENTS ON REAL DATA

Figure 9 gives the results of three real-world low light images. As shown in every red rectangle, our method LLIE-net always achieves better performance in dark regions. Besides the ILNIQE to evaluate the image quality, we evaluate the detail enhancement by using the statistical naturalness measure (SNM) [30]. A higher SNM score indicates that the enhanced image appears more natural. As shown in **Table 2**, for the real-world low light images from public datasets, LLIE-net can also obtain lower ILNIQE and higher SNM.

To our knowledge, we are the first to come up with this pipeline network jointing denoising and low light enhancement. As shown in Figure 10, we display the results of our pipeline network, with inputting the noise low-light images.

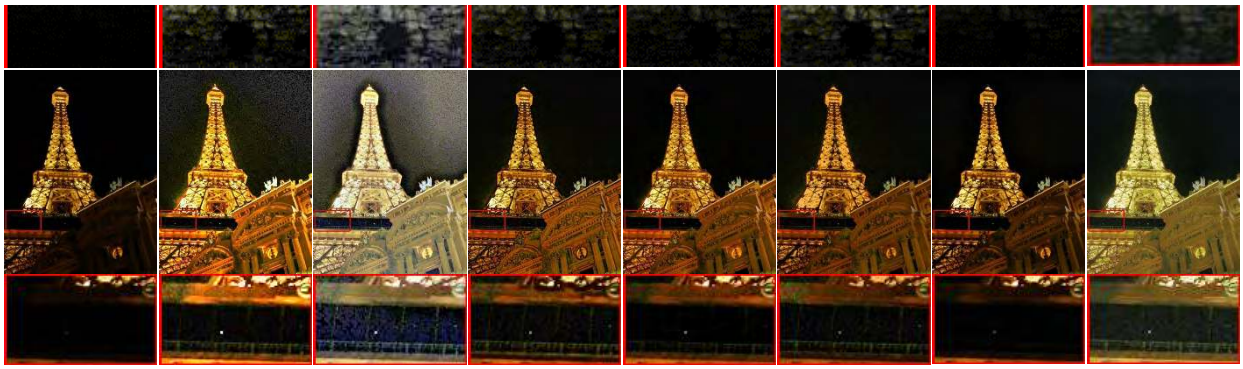

FIGURE 9. The results of real-world images using different methods and zoomed in region. (a) Input. (b) LIME. (c) MSRCR. (d) Dong. (e) MF. (f) NPE. (g) SRIE. (h) Ours.

TABLE 2. Quantitative results using ILNIQE/SNM on real-world low light images.

Dataset	Input	LIME	MSRCR	Dong	MF	NPE	SRIE	Ours
1st row	30.27/0.00	25.34/0.25	**19.47**/0.12	23.83/0.75	22.84/0.60	23.99/0.73	25.58/0.28	21.15/**0.93**
2nd row	22.01/0.00	21.48/0.01	16.79/0.16	18.92/0.15	17.76/0.17	17.11/0.24	20.64/0.10	**16.57/0.70**
3rd row	47.78/0.00	43.23/0.08	23.47/**0.69**	36.58/0.04	39.32/0.03	36.69/0.03	36.61/0.01	**22.84**/0.22
64 test images	29.36/0.00	28.77/0.27	23.61/0.31	27.33/0.41	26.22/0.46	25.91/0.51	26.17/0.35	**23.19/0.53**

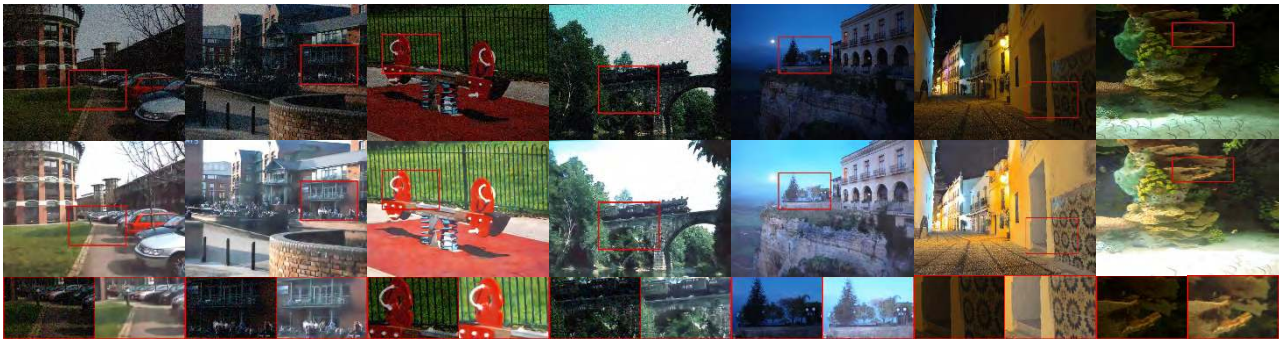

FIGURE 10. First row: the low light image with noise. Second row: the results of our pipeline network.

V. CONCLUTION

In this paper, we present a novel pipeline network for low-light image enhancement. We illustrate that multi-scale Retinex is equivalent to a CNN with different Gaussian convolution kernels. Furthermore, we find blending the low-low component of the image DWT with MSR can improve the results produced only by MSR. Inspired with this experiment, we propose LLIE-net, which is an end-to-end neural network without any artificial parameters in prediction phase. We compare our LLIE-net with existing state-of-the-art approaches from the qualitative and quantitative perspective respectively. Experiments reveal that our method achieves the better performance. Considering the fact that dealing with real-world low light images sometimes causes noise, we add a denoising-net and construct a pipeline network. Our method not only performs well in synthetic images but also in the real-world image sets. In **Table 2**, the performance of our method is not better than MSR in some images in Figure 9. It is because that the measure methods are subjective evaluation indexes, which would bring experimental error, and we test our method in a bigger image set consisting of the public real-world low light images. There are some methods to improve our model, such as training in a bigger dataset, adding hidden layers and so on, which can reduce the variance of our model.

REFERENCES

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

[2] 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.

[3] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proc. Int. Conf. Image Process.*, Sep. 1996, pp. 1003–1006.

[4] 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.

[5] 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.
[6] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.
[7] 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.
[8] X. Dong *et al.*, "Fast efficient algorithm for enhancement of low lighting video," in *Proc. IEEE Int. Conf. Multimedia Expo IEEE Comput. Soc.*, Jul. 2011, pp. 1–6.
[9] 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.*, Sep. 2015, pp. 3730–3734.
[10] X. Zhang, P. Shen, L. Luo, and L. Zhang, and J. Song, "Enhancement and noise reduction of very low light level images," in *Proc. Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 2034–2037.
[11] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
[12] H. Demirel and G. Anbarjafari, "Image resolution enhancement by using discrete and stationary wavelet decomposition," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1458–1460, May 2011.
[13] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[14] C. Ledig *et al.* (Aug. Sep. 2016). "Photo-realistic single image super-resolution using a generative adversarial network." [Online]. Available: https://arxiv.org/abs/1609.04802
[15] J. Lehtinen *et al.* (2018). "Noise2Noise: Learning image restoration without clean data." [Online]. Available: https://arxiv.org/abs/1803.04189
[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
[17] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.
[18] G. Schaefer and M. Stich, "UCID: An uncompressed color image database," *Proc. SPIE*, vol. 5307, pp. 472–481, Dec. 2003.
[19] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
[20] 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.
[21] K. Zeng, K. Ma, R. Hassen, and Z. Wang, "Perceptual evaluation of multi-exposure image fusion algorithms," in *Proc. 6th Int. Workshop IEEE Qual. Multimedia Exper. (QoMEX)*, Sep. 2014, pp. 7–12.
[22] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation," in *Proc. 19th IEEE Int. Conf. Image Process. (ICIP)*, Sep./Oct. 2012, pp. 965–968.
[23] 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.
[24] F. Chollet. (2015). *Keras, GitHub Repository*. [Online]. Available: https://github.com/fchollet/keras
[25] A. B. Petro, C. Sbert, and J.-M. Morel, "Multiscale retinex," *Image Process. Line*, pp. 71–88, Oct. 2014, doi: 10.5201/ipol.2014.97.
[26] 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.
[27] 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.
[28] 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.*, Jun. 2016, pp. 2782–2790.
[29] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
[30] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.
[31] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. ACM 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.
[32] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1017–1027, Apr. 2016.
[33] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma. (2017). "MSR-net: Low-light image enhancement using deep convolutional network." [Online]. Available: https://arxiv.org/abs/1711.02488
[34] C. Chen, Q. Chen, J. Xu, and V. Koltun. (2018). "Learning to see in the dark." [Online]. Available: https://arxiv.org/abs/1805.01934

YANHUI GUO received the B.S. degree from the Wuhan University of Technology, in 2017. He is currently pursuing the master's degree with the Department of Intelligent Science and Technology, Huazhong University of Science and Technology, China.

He has been a Researcher with the AILab, Tencent and NetEase Game. His research interests include the computer vision, robot control, deep learning, and artificial intelligence.

XUE KE was born in Huanggang, Hubei, China, in 1995. He received the B.S degree from Huazhong Agricultural University, in 2017. He is currently pursuing the master's degree with the Department of Intelligent Science and Technology, Huazhong University of Science and Technology, China.

His research interests include image process, robotic control, and objection detection.

JIE MA received the B.S. degree from the Wuhan University of Science and Technology, in 1996, and the M.S. and Ph.D. degrees from the Department of Intelligent Science and Technology, Huazhong University of Science and Technology, China, in 2000 and 2005, respectively.

Since 2006, he has been a Professor with the Department of Intelligent Science and Technology, Huazhong University of Science and Technology. His research interests include image process, pattern recognition, and 3D simulation.

He has presided over and completed more than 30 projects, including the National Natural Science Foundation of China, the Pre-research Foundation for General Assembly, the Pre-research Program for National Defense Weapons and Equipment, and National Defense 973 and 863. He is a member of the Chinese Society of Astronautics.

JUN ZHANG was born in 1966. He received the B.S. degree from Shanghai Jiaotong University, in 1986, and the M.S. and Ph.D. degrees from the Department of Intelligent Science and Technology, Huazhong University of Science and Technology, China, in 1999 and 2006, respectively, where he is currently an Assistant Professor with the School of Automation. His research interests include machine vision, machine learning, information retrieval, and data mining.

He has published 56 papers, including 12 articles of the Science Citation Index and 24 articles of the Engineering Index. He has participated in nine national projects and two provincial level projects and has hosted five transverse projects.