MBLLEN: Low-light Image/Video Enhancement Using CNNs

Feifan Lv[1]
lvfeifan@buaa.edu.cn
Feng Lu[1,2]
lufeng@buaa.edu.cn
Jianhua Wu[3]
jianhua.wu@sg.panasonic.com
Chongsoon Lim[3]
chongsoon.lim@sg.panasonic.com

[1] State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, China
[2] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing, China
[3] Panasonic R&D Center Singapore, Singapore City, Singapore

Abstract

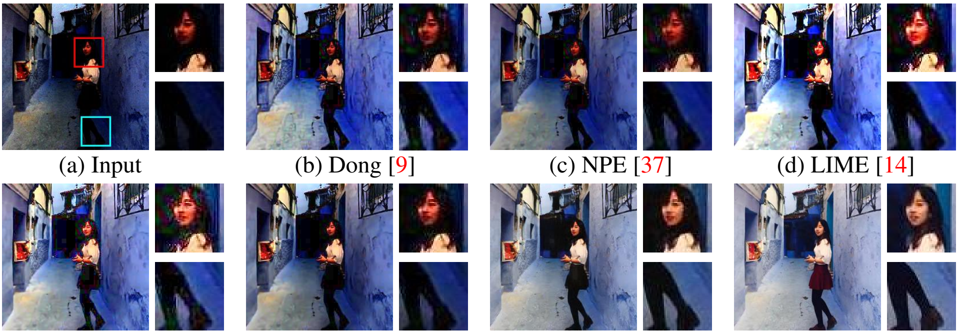
We present a deep learning based method for low-light image enhancement. This problem is challenging due to the difficulty in handling various factors simultaneously including brightness, contrast, artifacts and noise. To address this task, we propose the multi-branch low-light enhancement network (MBLLEN). The key idea is to extract rich features up to different levels, so that we can apply enhancement via multiple subnets and finally produce the output image via multi-branch fusion. In this manner, image quality is improved from different aspects. Through extensive experiments, our proposed MBLLEN is found to outperform the state-of-art techniques by a large margin. We additionally show that our method can be directly extended to handle low-light videos.

1 Introduction

Images and videos carry rich and detailed information of the real scenes. By capturing and processing the image and video data, intelligent systems can be developed for various tasks such as object detection, classification, segmentation, recognition, scene understanding and 3D reconstruction, and then used in many real applications, *e.g.*, automated driving, video surveillance and virtual/augmented reality.

However, real systems rely heavily on the quality of the input images/videos. In particular, they may perform well with high quality input data but perform badly otherwise. One typical case is to use the images captured in the poorly illuminated environment. When a camera cannot receive sufficient light during a capture, there will be information loss in the dark region and unexpected noise, as shown in Figure 1. Using such low quality images due to low light will certainly reduce the performance of most vision-based algorithms, and thus,

Corresponding Author: Feng Lu
This work is partially supported by NSFC under Grant 61602020 and Grant 61732016.



(a) Input (b) Dong [9] (c) NPE [37] (d) LIME [14]
(e) Ying [43] (f) BIMEF [41] (g) Ours (MBLLEN) (h) Ground Truth

Figure 1: The proposed MBLLEN can produce high quality images from low-light inputs. It also performs well in suppressing the noise and aritifacts in dark regions.

it is highly demanded by real applications to enhance the quality of the low-light images without requiring additional and expensive hardware.

Various researches have been done in the literature for low-light image enhancement. They typically focus on restoring the image brightness and contrast, and suppressing the unexpected visual effects like color distortion. Existing methods can be roughly divided into two categories, namely, the histogram equalization-based methods, and the Retinex theory-based methods. Algorithms in the former category optimize the pixel brightness based on the idea of histogram equalization, while methods in the latter category recover the illumination map of the scene and enhance different image regions accordingly.

Although remarkable progress has been made, there is still a lot room to improve. For instance, existing methods tend to rely on certain assumptions about the pixel statistics or visual mechanism, which, however, may not be applicable for certain real scenarios. Second, besides brightness/contrast optimization, other factors such as artifacts in the dark region and image noise due to low-light capture should be handled more carefully. Finally, developing effective techniques for low-light video enhancement requires additional efforts.

In this paper, we propose a novel method for low-light image enhancement by taking the success of the latest deep learning technology. At the core of our method is the proposed fully convolutional neural network, namely the multi-branch low-light enhancement network (MBLLEN). The MBLLEN consists of three types of modules, i.e., the feature extraction module (FEM), the enhancement module (EM) and the fusion module (FM). The idea is to learn to 1) extract rich features up to different levels via FEM, 2) enhance the multi-level features respectively via EM and 3) obtain the final output by multi-branch fusion via FM. In this manner, the MBLLEN is able to improve the image quality from different aspects and accomplish the low-light enhancement task to its full extent.

Overall, our contributions are threefold. 1) We propose a novel method for low-light image enhancement based on the deep neural networks. It improves both the objective and subjective image quality. 2) Our method also works well in terms of suppressing image noise and artifacts in the low light regions. 3) Our method can be directly extended to process low light videos by using the temporal information. These properties make our method superior to existing methods, and both quantitative and qualitative evaluations demonstrate that our method outperforms the state-of-the-arts by a large margin.

2 Related Work

This section briefly overviews existing techniques for low-light image/video enhancement.

Low-light image enhancement. Methods for low-light image enhancement can be mainly divided into two categories. The first category is built upon the well-known histogram equalization (HE) technique and also uses additional priors and constraints. In particular, BPDHE [15] tries to preserve image brightness dynamically; Arici *et al*. [2] propose to analyze and penalize the unnatural visual effects for better visual quality; DHECI [29] introduces and uses the differential gray-level histogram; CVC [5] uses the interpixel contextual information; LDR [26] focuses on the layered difference representation of 2D histogram to try to enlarge the gray-level differences between adjacent pixels.

The other category is based on the Retinex theory [22], which assumes that an image is composed of reflection and illumination. Typical methods, e.g., MSR [17] and SSR [18], try to recover and use the illumination map for low-light image enhancement. Recently, AMSR [24] proposes a weighting strategy based on SSR. NPE [37] balances the enhancement level and image naturalness to avoid over-enhancement. MF [11] processes the illumination map in a multi-scale fashion to improve the local contrast and maintain naturalness. SRIE [12] develops a weighted vibrational model for illumination map estimation. LIME [14] considers both the illumination map estimation and denoising. BIMEF [41, 42] proposes a dual-exposure fusion algorithm and Ying *et al*. [43] use the camera response model for further enhancement. In general, conventional low-light enhancement methods rely on certain statistical models and assumptions, which only partially explain the real world scenes.

Deep learning-based methods. Recently, deep learning has achieved great success in the field of low-level image processing. Powerful tools such as end-to-end networks and GANs [13] have been employed by various applications, including image super resolution [4, 23], image denoising [31, 44] and image-to-image translation [16, 45]. There are also methods proposed for low-light image enhancement. LLNet [28] uses the deep autoencoder for low-light image denoising. However, it does not take advantage of recent developments in deep learning. Other CNN-based methods like LLCNN [35] and [34] do not handle brightness/contrast enhancement and image denoising simultaneously.

Low-light video enhancement. There are few researches for low-light video enhancement. Some of them [27, 36] also follow the Retinex theory, while others use the gamma correction technique [19] or the similar framework for image de-haze [9, 30]. In order to suppress artifacts, similar patches from adjacent frames can be used [21]. Although these methods have achieved certain progress in low-light video enhancement, they still share some limitations, e.g., temporal information has not been well utilized to avoid flickering.

3 Methodology

The proposed method is introduced in this section with all the necessary details. Due to the complexity of the image content, it is often difficult for a simple network to achieve high quality image enhancement. Therefore, we design the MBLLEN in a multi-branch fashion. It decomposes the image enhancement problem into sub-problems related to different feature levels, which can be solved respectively to produce the final output via multi-branch fusion.

The input to the MBLLEN is a low-light color image and the output is an enhanced clean image of the same size. The overall network architecture and the data process flow is shown in Figure 2. The three modules, namely, FEM, EM and FM, are described later in detail.

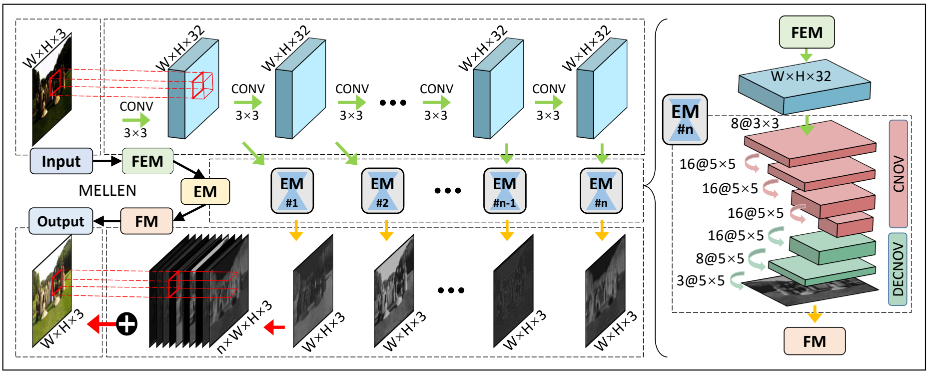


Figure 2: The proposed network with feature extraction module (FEM), enhancement module (EM) and fusion module (FM). The output image is produced via feature fusion.

3.1 Network Architecture

As shown in Figure 2, the proposed MBLLEN consists of three types of modules: the feature extraction module (FEM), the enhancement module (EM) and the fusion module (FM).

FEM. It is a single stream network with 10 convolutional layers, each of which uses kernels of size 3×3, stride of 1 and ReLU nonlinearity, and there is no pooling operation. The input to the first layer is the low-light color image. The output of each layer is both the input to the next layer and also the input to the corresponding subnet of EM.

EM. It contains multiple sub-nets, whose number equals to the number of layers in FEM. The input to a sub-net is the output of a certain layer in FEM, and the output is a color image with the same size of the original low-light image. Each sub-net has a symmetric structure to first apply convolutions and then deconvolutions. The first convolutional layer uses 8 kernels of size 3×3, stride 1 and ReLU nonlinearity. Then, there are three convolutional layers and three deconvolutional layers, using kernel size 5×5, stride 1 and ReLU nonlinearity, with kernel numbers of 16, 16, 16, 16, 8 and 3 respectively. Note that all the sub-nets are trained simultaneously but individually without sharing any learnt parameters.

FM. It accepts the outputs of all EM sub-nets to produce the finally enhanced image. We concatenate all the outputs from EM in the color channel dimension and use a 1×1 convolution kernel to merge them. This equals to the weighted sum with learnable weights.

Network for video. Our method can handle video enhancement after simple modification. 1) Let FEM perform 3D convolution instead of 2D convolution with 16 kernels of size $3 \times 3 \times 3$. The input of the first layer is the low-light color video which has 31 frames. The first three dimensions of the output from each layer are sent to EM, and the rest dimensions are used as the input to the next convolution. 2) EM is modified to perform 3D convolutions. 3) FM uses original low-light video as additional input for fusion.

3.2 Loss Function

In order to improve the image quality both qualitatively and quantitatively, using common error metrics such as MSE and MAE is shown to be insufficient. Therefore, we propose a novel loss function by further considering the structure information, context information and

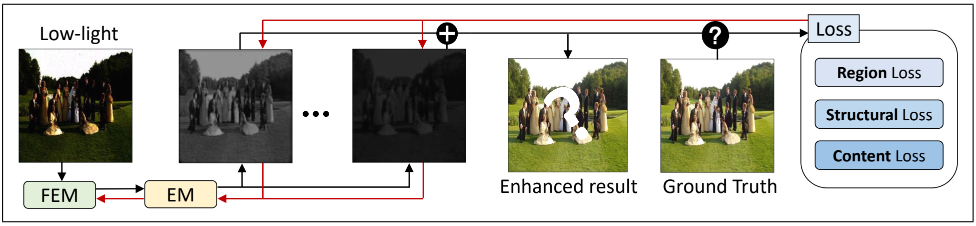


Figure 3: Data flow for training. The proposed loss function consists of three parts.

regional difference of the image, as shown in Figure 3. It is computed as:

$$Loss = L_{Str} + L_{VGG/i,j} + L_{Region}, \tag{1}$$

where details of the structure loss, context loss and region loss are given below.

Structure loss. This loss is designed to improve the visual quality of the output image. In particular, low-light capture usually causes structure distortion such as blur effect and artifacts, which is visually salient but cannot be well handled by MAE. Therefore, we introduce the structure loss to measure the difference between the enhanced image and the ground truth, so that to guide the learning process. In particular, we use the well-known image quality assessment algorithms SSIM [39] and MS-SSIM [38] to build our structure loss. A similar strategy has also been adopted in a recent method LLCNN [35].

We use a simplified form of SSIM computed for a pixel p by

$$L_{SSIM} = -\frac{1}{N}\sum_{p\in img}\frac{2\mu_x\mu_y + C_1}{\mu_x^2+\mu_y^2+C_1}\cdot\frac{2\sigma_{xy}+C_2}{\sigma_x^2+\sigma_y^2+C_2}, \tag{2}$$

where μ_x and μ_y are pixel value averages, σ_x^2 and σ_y^2 are variances, σ_{xy} is covariance, and C_1 and C_2 are constants to prevent the denominator to zero. Due to the page limit, the definition of MS-SSIM can be checked in [38]. The value ranges of SSIM and MS-SSIM are $(-1,1]$ and $[0,1]$, respectively. The final structure loss is defined as $L_{Str} = L_{SSIM} + L_{MS-SSIM}$.

Context loss. Metrics such as MSE and SSIM only focus on low-level information in the image, while it is also necessary to use some kind of higher-level information to improve the visual quality. Therefore, we refer to the idea in SRGAN [23] and use similar strategies to guide the training of the network. The basic idea is to employ a content extractor. Then, if the enhanced image and the ground truth are similar, their corresponding outputs from the content extractor should also be similar.

A suitable content extractor can be a neural network trained on a large dataset. Because the VGG network [33] is shown to be well-structured and well-behaved, we choose the VGG network as the content extractor in our method. In particular, we define the context loss based on the output of the ReLU activation layers of the pre-trained VGG-19 network. To measure the difference between the representations of the enhanced image and the ground truth, we compute their sum of absolute differences. Finally, the context loss is defined as follows:

$$L_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}C_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\sum_{z=1}^{C_{i,j}}\|\phi_{i,j}(E)_{x,y,z} - \phi_{i,j}(G)_{x,y,z}\| \tag{3}$$

where E and G are the enhanced image and ground truth, and $W_{i,j}$, $H_{i,j}$ and $C_{i,j}$ describe the dimensions of the respective feature maps within the VGG network. Besides, $\phi_{i,j}$ indicates the feature map obtained by j-th convolution layer in i-th block in the VGG-19 Network.

Region loss. The above loss functions take the image as a whole. However, for our low-light enhancement task, we need to pay more attention to those low-light regions. As a result, we propose the region loss, which balances the degree of enhancement between low-light and other regions in the image.

In order to do so, we first propose a simple strategy to separate low-light regions from other parts of the image. By conducting preliminary experiments, we find that choosing the top 40% darkest pixels among all pixels gives a good approximation of the low-light regions. One can also propose more complex ways for dark region selection and in fact there are many in the literature. Finally, the region loss is defined as follows:

$$L_{Region} = w_L \cdot \frac{1}{m_L n_L} \sum_{i=1}^{n_L} \sum_{j=1}^{m_L} (\|E_L(i,j) - G_L(i,j)\|) + w_H \cdot \frac{1}{m_H n_H} \sum_{i=1}^{n_H} \sum_{j=1}^{m_H} (\|E_H(i,j) - G_H(i,j)\|), \quad (4)$$

where E_L and G_L are the low-light regions of the enhanced image and ground truth, and E_H and G_H are the rest parts of the images. In our case, we suggest $w_L = 4$ and $w_H = 1$.

3.3 Implementation Details

Our implementation is done with Keras [7] and Tensorflow [1]. The proposed MBLLEN can be quickly converged after being trained for 5000 mini-batches on a Titan-X GPU with a set of 16925 images from the PASCAL VOC dataset [10]. We use mini-batches of 24 patches of size $256 \times 256 \times 3$. The input image values should be scaled to $[0,1]$.

In terms of designing the context loss, we test each convolutional layer of VGG-19. From the fourth convolution block, the enhancement effect decreases slightly. On the other hand, with deeper layers, the feature map size decreases, which increases the computational efficiency. As a trade-off, we use the output of the fourth convolutional layer of the third block of VGG as the context loss extraction layer.

In the experiment, training is done using the ADAM optimizer [20] with a learning rate of $\alpha = 0.002$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$. We also use the learning rate decay strategy, which reduces the learning rate to 95% before the next epoch.

4 Experimental Evaluation

The proposed method is evaluated and compared with existing methods through extensive experiments. For comparison, we use the published codes of the existing methods.

Overall, we have done four major sets of experiments as follows. 1) We compare our method with a number of existing methods including those latest and most representative ones on the task of low-light image enhancement. 2) We show another set of comparisons on the task of low-light image enhancement in the presence of Poisson noise. 3) We show experimental results using real-world low-light images. 4) We conduct further experiments on low-light video enhancement.

4.1 Dataset and Metrics

Capturing real-world low-light images with ground truth is difficult. Therefore, following the previous research [28], we produce a large set of low-light images via synthesis based on the PASCAL VOC images dataset [10] for this research.

Low-light image synthesis. Low-light images differ from common images due to two most salient features: low brightness and the presence of noise. For the former feature, we apply a random gamma adjustment to each channel of the common images to produce the low-light images, which is similar to [28]. This process can be expressed as $I_{out} = A \times I_{in}^{\gamma}$, where A is a constant determined by the maximum pixel intensity in the image and γ obeys a uniform distribution $U(2, 3.5)$. As for the noise, although many previous methods ignore it, we still take it into account. In particular, we add Poisson noise with peak value = 200 to the low-light image. Finally, we select 16925 images in the VOC dataset to synthesize the training set, 56 images for the validation set, and 144 images for the test set.

Low-light video synthesis. We chose e-Lab Video Data Set (e-VDS) [8] to synthesize low-light videos. We cut the original videos into video clips ($31 \times 255 \times 255 \times 3$) to build a dataset of around 20000 samples, 95% of which form the training set and the rest for test.

Performance metrics. To evaluate the performance of different methods from different aspects and in a more fair way, we use a variety of different metrics, including PSNR, SSIM [39], Average Brightness(AB) [6], Visual Information Fidelity(VIF) [32], Lightness order error(LOE) as suggested in [41] and TMQI [40]. Note that in the tables below, red, green and blue colors indicate the best, sub-optimal and third best results, respectively.

4.2 Low-light Image without Additional Noise

We conduct experiments using the synthetic dataset. Results are compared between our method and other 10 latest methods, as shown in Table 1. Our method outperforms all the other methods in all cases and is far ahead of the second (green) and third best (blue).

	PSNR	SSIM [39]	VIF [32]	LOE	TMQI [40]	AB [6]
Input	12.80	0.43	0.38	606.85	0.79	-54.62
SRIE [12]	15.84	0.59	0.43	788.53	0.82	-37.02
BPDHE [15]	15.01	0.59	0.39	**607.43**	0.81	-37.38
LIME [14]	15.16	0.60	0.44	1215.58	0.82	10.90
MF [11]	18.48	0.67	0.45	882.24	0.84	-23.24
Dong [9]	17.80	0.64	0.37	1398.35	0.82	-16.43
NPE [37]	17.65	0.68	0.43	1051.15	0.84	-25.28
DHECI [29]	18.18	0.68	0.43	**606.98**	**0.87**	**-1.21**
WAHE [2]	17.64	0.67	**0.48**	648.29	0.84	-30.28
Ying [43]	**19.66**	**0.73**	**0.47**	892.56	**0.86**	**2.57**
BIMEF [41]	**19.80**	**0.74**	**0.48**	675.15	0.85	-20.22
Ours	**26.56**	**0.89**	**0.55**	**478.02**	**0.91**	**-0.96**

Table 1: Comparison of low-light image (without additional noise) enhancement.

Representative results are visually shown in Figure 4. By checking the details, it is clear that our method achieves better visual effects, including good brightness/contrast and less artifacts. It is highly encouraged to zoom in to compare the details.

In order to emphasize our advantage in detail recovery besides the brightness recovery, we scale the image brightness of all methods according to the ground truth so that they have the exactly correct maximum and minimum values. Then, we compare the results in Table 2. Due to space limit, we only select those best methods from Table 1. The results show that our method still outperforms all other methods by a large margin.

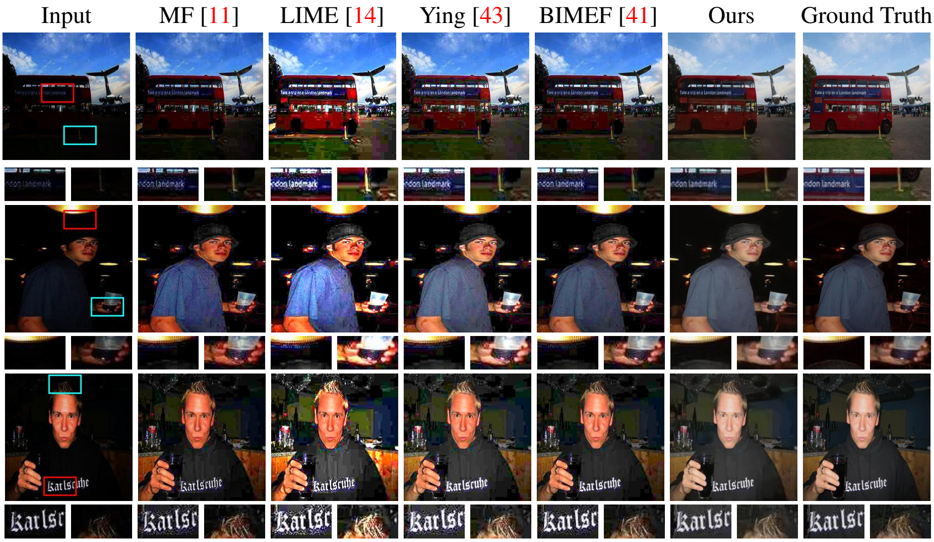


Figure 4: Comparison of low-light(no additional noise) images.

	PSNR	SSIM [39]	VIF [32]	LOE	TMQI [40]	AB [6]
WAHE [2]	17.43	0.65	**0.48**	**648.30**	0.84	-30.70
MF [11]	**18.64**	0.67	0.45	882.23	0.84	-22.74
DHECI [29]	18.34	0.68	0.43	**607.01**	**0.87**	**-2.09**
Ying [43]	**19.93**	**0.73**	**0.47**	892.53	**0.86**	**2.22**
BIMEF [41]	16.75	**0.71**	0.46	674.53	0.83	-34.44
Ours	**26.65**	**0.89**	**0.55**	**477.95**	**0.90**	**-1.30**

Table 2: Comparison of different methods after brightness scale according to ground truth.

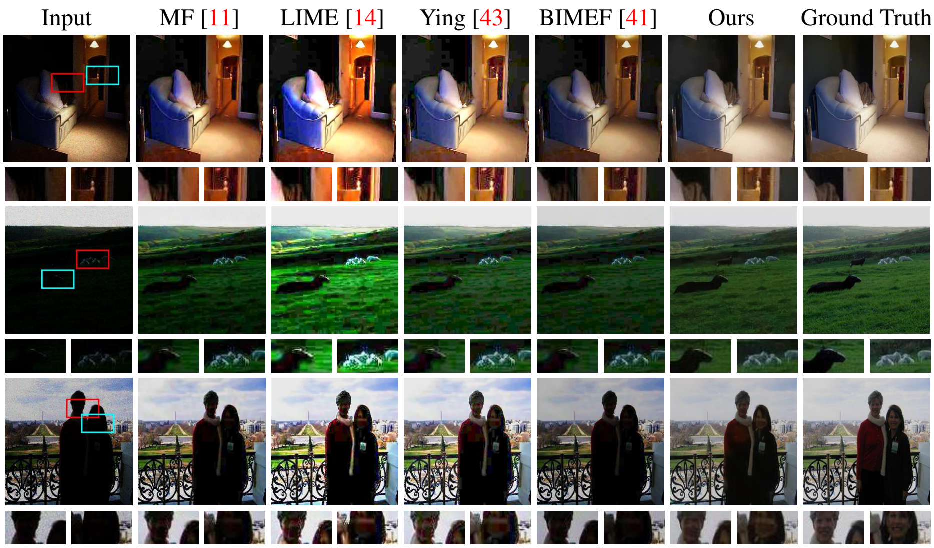


Figure 5: Enhancement comparison of low-light(with additional Poisson noise) images.

4.3 Low-light Image with Additional Poisson Noise

We test our method on low-light images with additional Poisson noise. For comparison, we choose the best comparison methods from Table 1, and use them along with the IVST [3] denoising method to produce the final comparison results. A variety of quality metrics are used for evaluation in the cases of original output (left number) and scaled output (right number) as in the last experiments, as shown in Table 3.

	PSNR	SSIM [39]	VIF [32]	LOE	TMQI [40]	AB [6]
WAHE [2]	17.91/17.37	0.62/0.59	**0.40**/**0.40**	771.34/771.33	0.83/0.82	-26.41/-30.04
MF [11]	**19.37**/**19.66**	0.67/0.67	0.39/0.38	896.67/896.46	0.84/**0.84**	-13.77/-16.88
DHECI [29]	18.03/18.71	0.67/0.68	0.36/0.36	**687.60**/**687.61**	**0.86**/**0.86**	**3.75**/ **0.16**
Ying [43]	18.61/**19.69**	**0.70**/**0.71**	**0.40**/**0.39**	928.13/927.83	**0.86**/**0.86**	**10.99**/ **6.99**
BIMEF [41]	**20.27**/17.56	**0.73**/**0.70**	**0.41**/**0.39**	**725.72**/**725.61**	**0.85**/0.83	-11.58/-31.13
Ours	**25.97**/**26.39**	**0.87**/**0.87**	**0.49**/**0.49**	**573.14**/**573.14**	**0.90**/**0.89**	**1.45**/ **-1.57**

Table 3: Comparison on low-light (with additional Poisson noise) images enhancement. As described earlier, the numbers on the right indicate the brightness-scaled results.

In Table 3, our method almost achieves all the best results under all quality metrics. The only case it ranks second is on the brightness-scaled result under the AB metrics. However, note that for brightness-scaling, we need to provide all the methods with the ground truth brightness, and such results are just for reference but not for a fair comparison. Visual demonstration is given in Figure 5.

4.4 Real-world Image

Besides the above synthetic dataset, our method also performs well on the natural low-light images and outperforms existing ones. Due to page limit, comparison on one representative example is shown in figure 6, along with additional low-light image enhancement results by our method. More results and comparisons, including those for previous sections, are included in the **supplementary files**, which fully demonstrate the effectiveness of our method.

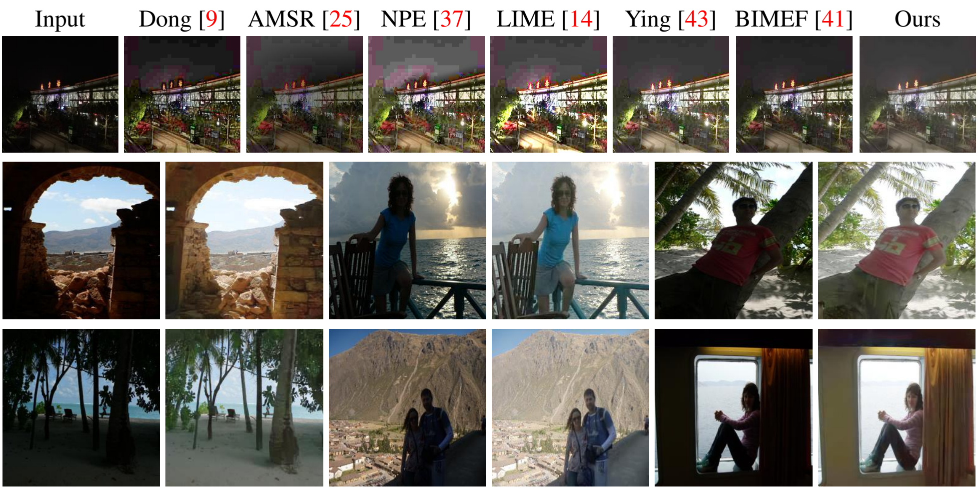


Figure 6: Some real-world results. Image in the first row is captured in a railway station. The rest images are from the Vassilios Vonikakis dataset downloaded from the Internet.

4.5 Low-light Video

Existing methods typically process videos in a frame-by-frame manner. On the other hand, our modified method can process low light videos more efficiently via 3D convolution.

	LIME [14]	Ying [43]	BIMEF [41]	MBLLEN	MBLLVEN
PSNR	14.26	**22.36**	**19.80**	19.71	**24.98**
SSIM [39]	0.59	**0.78**	0.76	**0.88**	**0.83**
AB(Var) [6]	**0.015**	0.025	0.044	**0.010**	**0.006**

Table 4: Quantitative evaluation on low-light video enhancements.

Comparison of our MBLLEN, its video version (MBLLVEN) and three other methods are shown in Table 4. We also introduce the AB(var) metric to measure the difference of the average brightness variance between the enhanced video and the ground truth. This metric reflects whether the video has unexpected brightness changes or flickers, and the proposed MBLLVEN achieves the best performance in preserving the inter-frame consistency. The enhanced videos are provided in the **supplementary files** for intuitive comparison.

5 Conclusion

This paper proposes a novel CNN-based method for low-light enhancement. Existing methods usually rely on certain assumptions and often ignore additional factors such as image noise. To solve those challenges, we aim at training a powerful and flexible network to address this task more effectively. Our network consists of three modules, namely the FEM, EM and FM. It is designed to be able to extract rich features from different layers in FEM, and enhance them via different sub-nets in EM. By fusing the multi-branch outputs via FM, it produces high quality results and outperforms the state-of-the-arts by a large margin. The network can also be modified to handle low-light videos effectively.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[2] Tarik Arici, Salih Dikbas, and Yucel Altunbasak. A histogram modification framework and its application for image contrast enhancement. *IEEE Transactions on image processing*, 18(9):1921–1935, 2009.

[3] Lucio Azzari and Alessandro Foi. Variance stabilization for noisy+ estimate combination in iterative poisson denoising. *IEEE signal processing letters*, 23(8):1086–1090, 2016.

[4] Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[5] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12):3431–3441, 2011.

[6] ZhiYu Chen, Besma R Abidi, David L Page, and Mongi A Abidi. Gray-level grouping (glg): an automatic method for optimized image contrast enhancement-part i: the basic method. *IEEE transactions on image processing*, 15(8):2290–2302, 2006.

[7] François Chollet et al. Keras. https://github.com/keras-team/keras, 2015.

[8] Eugenio Culurciello and Alfredo Canziani. e-Lab video data set. https://engineering.purdue.edu/elab/eVDS/, 2017.

[9] Xuan Dong, Guan Wang, Yi Pang, Weixin Li, Jiangtao Wen, Wei Meng, and Yao Lu. Fast efficient algorithm for enhancement of low lighting video. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1–6. IEEE, 2011.

[10] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[11] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82–96, 2016.

[12] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2782–2790, 2016.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[14] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017.

[15] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007.

[16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint*, 2017.

[17] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997.

[18] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997.

[19] Minjae Kim, Dubok Park, David K Han, and Hanseok Ko. A novel approach for denoising and enhancement of extremely low-light video. *IEEE Transactions on Consumer Electronics*, 61(1):72–80, 2015.

[20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[21] Seungyong Ko, Soohwan Yu, Wonseok Kang, Chanyong Park, Sangkeun Lee, and Joonki Paik. Artifact-free low-light video enhancement using temporal similarity and guide map. *IEEE Transactions on Industrial Electronics*, 64(8):6392–6401, 2017.

[22] Edwin H Land. The retinex theory of color vision. *Scientific American*, 237(6):108–129, 1977.

[23] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint*, 2016.

[24] Chang-Hsing Lee, Jau-Ling Shih, Cheng-Chang Lien, and Chin-Chuan Han. Adaptive multiscale retinex for image contrast enhancement. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on*, pages 43–50. IEEE, 2013.

[25] Chang-Hsing Lee, Jau-Ling Shih, Cheng-Chang Lien, and Chin-Chuan Han. Adaptive multiscale retinex for image contrast enhancement. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on*, pages 43–50. IEEE, 2013.

[26] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22 (12):5372–5384, 2013.

[27] Huijie Liu, Xiankun Sun, Hua Han, and Wei Cao. Low-light video image enhancement based on multiscale retinex-like algorithm. In *Control and Decision Conference (CCDC), 2016 Chinese*, pages 3712–3715. IEEE, 2016.

[28] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. L lnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017.

[29] Keita Nakai, Yoshikatsu Hoshi, and Akira Taguchi. Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms. In *Intelligent Signal Processing and Communications Systems (ISPACS), 2013 International Symposium on*, pages 445–449. IEEE, 2013.

[30] Jianhua Pang, Sheng Zhang, and Wencang Bai. A novel framework for enhancement of the low lighting video. In *Computers and Communications (ISCC), 2017 IEEE Symposium on*, pages 1366–1371. IEEE, 2017.

[31] Tal Remez, Or Litany, Raja Giryes, and Alex M Bronstein. Deep class-aware image denoising. In *Sampling Theory and Applications (SampTA), 2017 International Conference on*, pages 138–142. IEEE, 2017.

[32] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE Transactions on image processing*, 15(2):430–444, 2006.

[33] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[34] Li Tao, Chuang Zhu, Jiawen Song, Tao Lu, Huizhu Jia, and Xiaodong Xie. Low-light image enhancement using cnn and bright channel prior. In *Image Processing (ICIP), 2017 IEEE International Conference on*, pages 3215–3219. IEEE, 2017.

[35] Li Tao, Chuang Zhu, Guoqing Xiang, Yuan Li, Huizhu Jia, and Xiaodong Xie. Llcnn: A convolutional neural network for low-light image enhancement. In *Visual Communications and Image Processing (VCIP), 2017 IEEE*, pages 1–4. IEEE, 2017.

[36] Dongsheng Wang, Xin Niu, and Yong Dou. A piecewise-based contrast enhancement framework for low lighting video. In *Security, Pattern Analysis, and Cybernetics (SPAC), 2014 International Conference on*, pages 235–240. IEEE, 2014.

[37] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.

[38] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, volume 2, pages 1398–1402. Ieee, 2003.

[39] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[40] Hojatollah Yeganeh and Zhou Wang. Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing*, 22(2):657–667, 2013.

[41] Zhenqiang Ying, Ge Li, and Wen Gao. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv preprint arXiv:1711.00591*, 2017.

[42] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new image contrast enhancement algorithm using exposure fusion framework. In *International Conference on Computer Analysis of Images and Patterns*, pages 36–46. Springer, 2017.

[43] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new low-light image enhancement algorithm using camera response model. *manuscript submitted for publication*, 2017.

[44] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[45] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.