

EEMEFN: Low-Light Image Enhancement via Edge-Enhanced Multi-Exposure Fusion Network

Minfeng Zhu^{1,2*}, Pingbo Pan^{2,3†}, Wei Chen^{1‡}, Yi Yang²

¹State Key Lab of CAD&CG, Zhejiang University

²The ReLER Lab, University of Technology Sydney, ³Baidu Research

minfeng_zhu@zju.edu.cn, chenwei@cad.zju.edu.cn, pingbo.pan@student.uts.edu.au, yi.yang@uts.edu.au

Abstract

This work focuses on the extremely low-light image enhancement, which aims to improve image brightness and reveal hidden information in dark areas. Recently, image enhancement approaches have yielded impressive progress. However, existing methods still suffer from three main problems: (1) low-light images usually are high-contrast. Existing methods may fail to recover images details in extremely dark or bright areas; (2) current methods cannot precisely correct the color of low-light images; (3) when the object edges are unclear, the pixel-wise loss may treat pixels of different objects equally and produce blurry images. In this paper, we propose a two-stage method called Edge-Enhanced Multi-Exposure Fusion Network (EEMEFN) to enhance extremely low-light images. In the first stage, we employ a multi-exposure fusion module to address the high contrast and color bias issues. We synthesize a set of images with different exposure time from a single image and construct an accurate normal-light image by combining well-exposed areas under different illumination conditions. Thus, it can produce realistic initial images with correct color from extremely noisy and low-light images. Secondly, we introduce an edge enhancement module to refine the initial images with the help of the edge information. Therefore, our method can reconstruct high-quality images with sharp edges when minimizing the pixel-wise loss. Experiments on the See-in-the-Dark dataset indicate that our EEMEFN approach achieves state-of-the-art performance.

Introduction

High-quality images are critical to computer vision tasks, e.g., video surveillance, and object detection. However, images captured in extremely low-light conditions may lose information in the dark region and receive unexpected noise and color bias. The low-quality images may significantly reduce the performance of computer vision methods (Zhu et al. 2017; Long, Shelhamer, and Darrell 2015), which rely

heavily on the quality of input images. Therefore, low-light image enhancement has attracted growing attention in the computer vision field. This work focuses on solving the low-light image enhancement problem, which aims to reduce color bias and reveal hidden information from dark regions.

There have been numerous attempts in the literature to address the low-light image enhancement problem. Traditional techniques (Lee, Lee, and Kim 2013; Land 1977) enhance pixels with low values to obey a natural distribution. Deep learning-based methods design deep models to recover high-quality images (Chen et al. 2018; Lore, Akintayo, and Sarkar 2017). Though these methods achieve significant progress, enhancing extremely low-light images is still challenging due to three main problems. Firstly, when a low-light image is high-contrast with under-exposed or over-exposed regions, it is difficult to find a one-to-one mapping from raw sensor data to images. Existing methods fail to recover well-exposed image details from extremely dark or bright areas. The generated images are always blurry and noisy. Secondly, without well-exposed image information, existing models may suffer from color distortion due to the color bias between low-light images and ground truth images. Thirdly, pixel-wise loss tends to blur edges, destroy image details when object edges are unclear in low-light images. Pixel-wise loss weights all pixels equally regardless of their spatial distribution. For instance, l_1 or l_2 loss will be minimized by averaging nearby pixels or choosing the median of possible colors (Isola et al. 2017).

In this paper, we propose the Edge-Enhanced Multi-Exposure Fusion Network (EEMEFN) to enhance extremely low-light images and recover realistic image details. Our EEMEFN decomposes the image enhancement process into two stages. In the first stage, we propose a multi-exposure fusion (MEF) module to solve the high contrast and color bias issues. To recover a well-exposed image from a high-contrast low-light image, we need to assign different exposure time to different regions. For instance, bright areas require a shorter exposure time than dark areas. We generate a set of multi-exposure images with different exposure times from one single low-light image. Specifically, fusion blocks are proposed to combine well-exposed regions of the multi-exposure image set in a complementary way. Thus, the MEF

*This work was done when Minfeng Zhu was visiting the University of Technology Sydney.

†Part of this work was done when Pingbo Pan was a research intern with baidu research.

‡Corresponding author.

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

module can produce high-quality initial images and reason a color distribution closer to the ground truth. In the second stage, we introduce an edge enhancement (EE) module to address the edge issue. We predict edge information with fine structures from the initial image generated by the first stage. Then, the EE module enhances the initial image by taking advantage of global features from the image contents and local features from the edge information. We evaluate our proposed EEMEFN on the See-in-the-Dark dataset (Chen et al. 2018). Quantitatively, our method outperforms state-of-the-art methods. For instance, the Peak Signal-to-Noise Ratio (PSNR) of our approach on the Sony subset is 29.60 dB, compared to 28.88 dB by (Chen et al. 2018). Our approach improves the Structural SIMilarity (SSIM) on the Fuji subset from 0.680 to 0.723. Besides, qualitative results indicate that our EEMEFN achieves a more natural result with abundant textures and sharp edges.

In summary, we make the following contributions:

- We propose a novel multi-exposure fusion module together with fusion blocks to combine generated images with different light conditions so that the high contrast and color bias issues can be addressed.
- We introduce an edge enhancement module to enhance images with sharp edges and fine structures.
- The experimental results demonstrate that the proposed approach achieves state-of-the-art performance. Furthermore, we conduct an ablation study to demonstrate the effectiveness of each module.

Related Work

Low-Light Image Enhancement

Low-light images captured in extremely low-light conditions will undoubtedly reduce the performance of computer vision algorithms (Ying et al. 2017). Thus, various low-light image enhancement approaches have been proposed to recover a high-quality image. Traditional approaches can be categorized into two main categories: histogram-based methods (Lee, Lee, and Kim 2013) and retinex-based methods (Land 1977). For example, histogram equalization (Cheng and Shi 2004) tries to map the histogram of the whole image as a simple mathematical distribution. However, these methods recover each pixel individually without considering surrounding pixels. Retinex-based methods (Guo 2016) first estimate an illumination map according to the retinex theory and then enhance each pixel using the well-constructed illumination map. Ying, Li, and Gao (2017) proposed to dynamically adjust exposure time based on a single image and fuse images using the illumination estimation techniques. However, it is difficult to estimate the illumination map of extremely low-light images with severe noise.

Recently, deep learning-based methods have achieved significant improvements in image enhancement, such as deblurring (Aittala and Durand 2018), denoising (Godard, Matzen, and Uyttendaele 2018) and low-light image enhancement (Lv et al. 2018). LLNet (Lore, Akintayo, and Sarkar 2017) consists of a contrast enhancement module

and a denoising module. LLCNN (Tao et al. 2017) applies a special-designed convolutional module to utilize multi-scale feature maps for image enhancement. Retinex-Net (Wei et al. 2018) consists of a Decom-Net for decomposition and an Enhance-Net for illumination adjustment. CAN (Chen, Xu, and Koltun 2017) approximates a variety of processing operators. Different from existing methods operate on sRGB images, Chen et al. (2018) employed a fully-convolutional network (Long, Shelhamer, and Darrell 2015) for extremely low-light image enhancement based on raw sensor data. Although these methods may produce a satisfying result sometimes, they may generate low-quality images due to high contrast and serious color bias of extremely low-light images. In addition to single-image enhancement, multiple-image methods can address these issues since more information is collected (Mildenhall et al. 2018; Hasinoff et al. 2016). For instance, (Godard, Matzen, and Uyttendaele 2018) introduced a parallel recurrent network to combine images. These methods focus on how to integrate the information of all frames in the burst. In this paper, we propose the multi-exposure fusion module, which generates images with different light-conditions from a single image and intelligently combines well-exposed information. The multi-exposure fusion network can reduce color bias and produce a high-quality image.

Edge Detection

Edge detection is one of the most fundamental computer vision tasks. Existing methods can be roughly categorized into three groups. The first one usually produces an edge map by designing various filters manually. For instance, (Canny 1986) introduced the Gaussian smoothing in the process of extracting the image gradient. The second category predicts edges using data-driven models according to human-designed features. Structured Edges (Dollár and Zitnick 2015) employs random decision forests to learn the structure of edge patches. Third, deep learning methods learn complex feature representations from raw data and have achieved considerable progress. HED (Xie and Tu 2015) is an end-to-end edge detection model that combines side outputs from multiple scales. Deepedge (Bertasiu, Shi, and Torresani 2015) averages the outputs of a classification branch and a regression branch to produce the final outputs. Deepcontour (Shen et al. 2015) divides edge data into subclasses and fits each subclass using different model parameters. RCF (Liu et al. 2017) uses rich features from all the convolution layers to perform an image-to-image prediction task in real-time. Liu et al. (2018) proposed the diverse deep supervision which minimizes different loss functions for high-level and low-level feature learning. Image enhancement methods with l_1 and mean squared error loss tend to blur sharp edges, destroy lines, and other fine image details. Given that the human visual system is highly sensitive to edges, preserving structure information is crucial to the performance of the image reconstruction task. Inspired by previous works, we propose the edge enhancement module to reconstruct a high-quality image with abundant textures and rich local structures using edge information for the low-light image enhancement task.

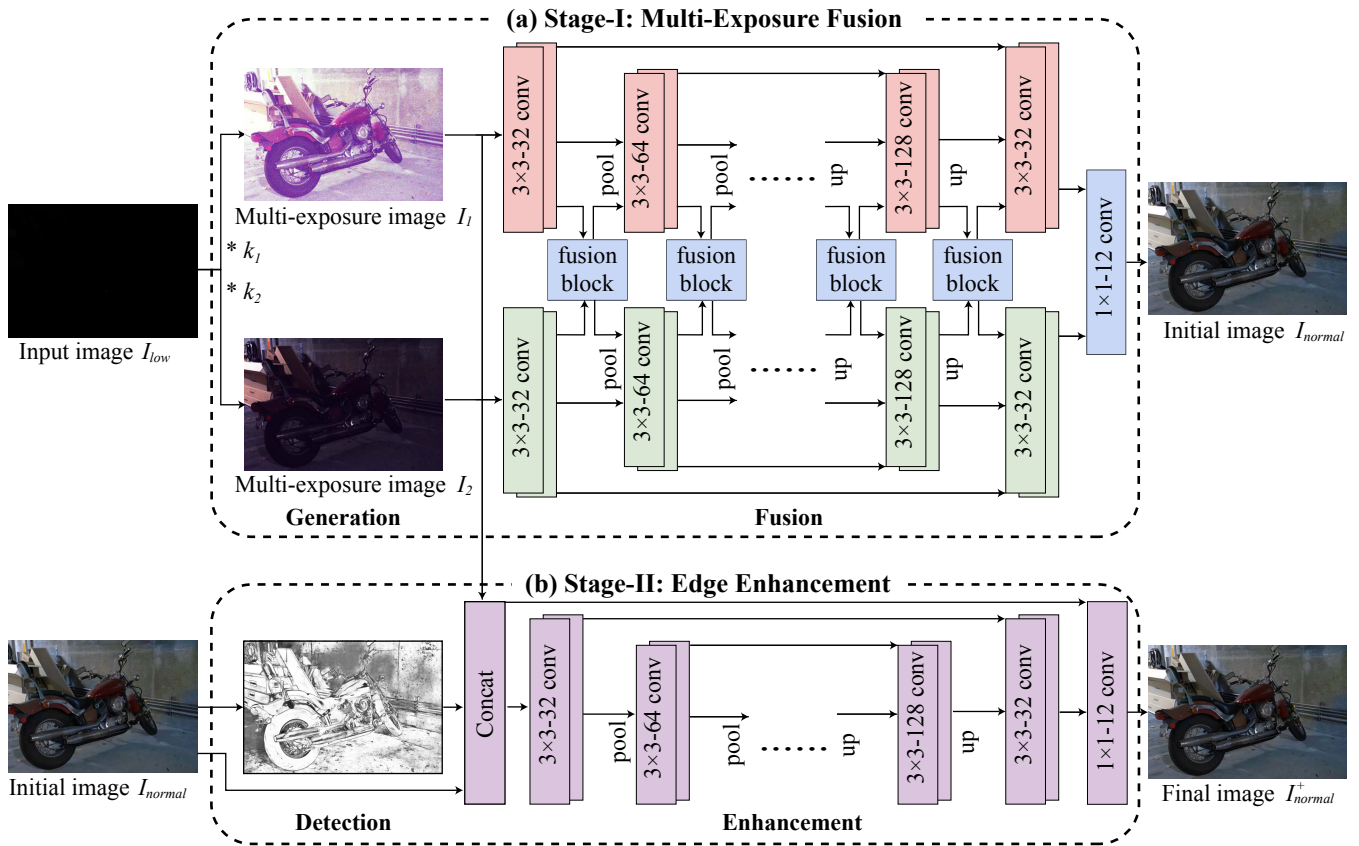


Figure 1: Demonstration of our framework for low-light image enhancement. The proposed EEMEFN consists of two stages: (a) multi-exposure fusion and (b) edge enhancement. The multi-exposure fusion module first generates several images in different light conditions and then fuses images into one high-quality initial image. The edge enhancement module obtains an edge map from the initial image and combines edge information to yield the final enhanced image.

Method

The Edge-Enhanced Multi-Exposure Fusion Network (EEMEFN) aims to recover well-exposed image details, reduce noise and color bias, and maintain sharp edges for extremely low-light image enhancement. As illustrated in Figure 1, our EEMEFN approach consists of two components: multi-exposure fusion and edge enhancement.

Stage-I: Multi-Exposure Fusion

The multi-exposure fusion (MEF) module is introduced to fuse images with different exposure ratios into one high-quality initial image. As shown in Figure 1 (a), the MEF module includes two main steps: generation and fusion. In the generation step, we produce a set of multi-exposure images in different light conditions by scaling the pixel value with different amplification ratio. The synthesized images may be well-exposed in the regions where the original image is under-exposed. In the fusion step, we fuse partial information from synthesized images into an initial image that holds well-exposed information. Specially, we propose fusion blocks to share the image features obtained by different convolution layers to fully exploit valuable information from the synthesized images.

Generation. Given an raw image $I_{low} \in \mathbb{R}^{H \times W \times 1}$ and a set of exposure ratios $\{k_1, k_2, \dots, k_N\}$, we can generate a set of multi-exposure images $I = \{I_1, I_2, \dots, I_N\}$. The i -th image is defined as follows:

$$I_i = Clip(I_{low} * k_i), \quad (1)$$

where the $Clip(x) = \min(x, 1)$ function performs per-pixel clipping of the image. We denote k^* as the ratio of exposure times between the low-light image and the reference image. Note that Chen et al. (2018) only generated one image by specifying the exposure ratio as k^* . Considering the information redundancy in the multi-exposure images, feeding all images into our model may not improve the performance but increase the computational cost.

Fusion. In this step, the MEF module combines well-exposed regions from the generated multi-exposure images $I = \{I_1, I_2, \dots, I_N\}$ to obtain an initial image I_{normal} :

$$I_{normal} = MEN(I_0, I_1, \dots, I_N). \quad (2)$$

Figure 1 (a) illustrates the architecture of our MEF module, which is fed with images in different light conditions. Our architecture can easily scale to multiple images. We take two images as an example to demonstrate our framework.

Firstly, each image is processed by a U-net branch with the same architecture (Ronneberger, Fischer, and Brox 2015). We add skip connections in the U-net to aid the reconstruction of details at different scales.

Secondly, we propose fusion blocks to combine the image features obtained by different branches to fully exploit valuable information in a complementary way (see Figure 2). The fusion block is build up a permutation-invariant technique (Aittala and Durand 2018) with more aggregation operations between features. Therefore, the MEF module can recover accurate image details from dark regions and push the color distribution closer to the ground truth. Each fusion block takes N image features $f_1, f_2, \dots, f_N \in \mathbb{R}^{c \times hw}$ from N branches as inputs and performs max and average operations to extract valuable information.

$$f_{max} = \max(f_1, f_2, \dots, f_N), \quad (3)$$

$$f_{avg} = (f_1 + f_2 + \dots + f_N)/N. \quad (4)$$

Then, we transform features f_{max} and f_{avg} into the input feature space and seed them back to each branch.

$$f = W * [f_{max}, f_{avg}], f \in \mathbb{R}^{c \times hw}, W \in \mathbb{R}^{c \times 2c}, \quad (5)$$

where $[\cdot, \cdot]$ denotes the concatenation operation, f is the output feature, and W is the learned weight matrix.

Finally, the last features of all branches are concatenated together and fed into a 1×1 Conv layer to produce the desired output by jointly learning from all branches.

The loss function is defined as l_1 loss between the output I_{normal} of our MEF module and the ground truth:

$$loss_{Fusion} = \|I_{normal} - I_{gt}\|_1. \quad (6)$$

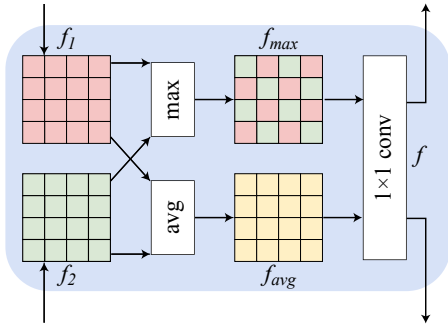


Figure 2: The architecture of the proposed fusion block.

Stage-II: Edge Enhancement

The edge enhancement (EE) module aims to enhance the initial image generated by the MEF module. As shown in Figure 1 (b), the EE module consists of two main steps: detection and enhancement. In the detection step, we firstly produce an edge map from the initial image instead of the input image. The input image and the multi-exposure images are extremely noisy, which makes it challenging to extract edges. The MEF module effectively removes noise and generates clear normal-light images with accurate edge maps. In the enhancement step, the EE module enhances

the initial image by leveraging the edge information. With the help of the predicted edge map, the EE module can generate more smooth object surfaces with consistent color, and recover abundant textures and sharp edges.

Detection. In this step, we employ the edge detection network (Liu et al. 2017) to predict edges $E \in \mathbb{R}^{H \times W \times 1}$ of I_{normal} , $E = Detection(I_{normal})$. Then the edge information is utilized to guide the reconstruction of high-quality images. The edge detection network consists of five stages, each of which makes use of all the activation of convolution layers to perform the pixel-wise prediction (E_1, E_2, E_3, E_4, E_5). Finally, a fusion layer is employed to combine CNN features of all the stages carefully. The edge detection network can obtain an accurate edge map E .

Considering the fact that the distribution of edge/non-edge pixels is heavily unbalanced, we employ two class-balancing weights α and β to offset such unbalance. The edge loss between predicted edge map E_i and ground truth $E_{gt} = (e_j, j = 1, \dots, |E_{gt}|), e_j = \{0, 1\}$ is defined as a weighted cross entropy loss with respect to the pixel label:

$$l_{edge}(E_i, E_{gt}) = -\alpha \sum_{j \in E_{gt}^+} \log \Pr(e_j = 1|i) - \beta \sum_{j \in E_{gt}^-} \log(1 - \Pr(e_j = 1|i)), \quad (7)$$

$$\alpha = \frac{|E_{gt}^-|}{|E_{gt}^+| + |E_{gt}^-|}, \beta = \frac{|E_{gt}^+|}{|E_{gt}^+| + |E_{gt}^-|}, \quad (8)$$

where $|E_{gt}^+|$ and $|E_{gt}^-|$ denote the size of the edge and non-edge ground truth label sets, $e_j = 1$ represents an edge point at pixel j and $\Pr(e_j = 1|i)$ is the predicted value of pixel j at stage i . The loss function is computed by aggregating the loss function from different stages and the fusion layer:

$$loss_{Detection} = \sum_{i=1}^5 l_{edge}(E_i, E_{gt}) + l_{edge}(E, E_{gt}). \quad (9)$$

To train the edge detection network, we generate a set of input-output pairs. The input is the initial images generated by our MEF module. We use the Canny edge detector (Canny 1986) to compute the output as the corresponding edge map of the ground truth image. Compared with Canny, our EE module can extract useful edges from blurry initial images and achieves better performance.

Enhancement. The enhancement step adopts an U-Net architecture, which takes multi-exposure images $I = \{I_1, I_2, \dots, I_N\}$, initial image I_{normal} and edge map E as inputs and integrates these images to produce the final enhanced image I_{normal}^+ .

$$I_{normal}^+ = Enhancement(I, I_{normal}, E). \quad (10)$$

The Enhancer takes advantage of global pixel information from images and local edge information from the edge map. The loss function of the enhancement step is defined as:

$$loss_{Enhancement} = \|I_{normal}^+ - I^{gt}\|_1. \quad (11)$$

We also evaluated other auxiliary loss, such as edge-preserving loss and perceptual loss. However, edge-preserving loss significantly reduces performance and the perceptual loss cannot improve performance.

Experiments

In this section, to demonstrate the capability of our proposed method, we evaluate the EEMEFN model quantitatively and qualitatively on the See-in-the-Dark dataset (Chen et al. 2018) for extremely low-light image enhancement. We implemented our method based on the Tensorflow framework and the Paddlepaddle framework.

Dataset. The See-in-the-Dark dataset consists of two image sets: Sony set and Fuji set. The Sony set captured by Sony α 7S II includes 2697 raw short-exposure images and 231 long-exposure images. The Fuji set captured by Fujifilm X-T2 contains 2397 raw short-exposure images and 193 long-exposure images. The resolution is 4240×2832 for the Sony Set and 6000×4000 for the Fuji set. The exposure time of the raw images was set between 1/30 and 1/10 seconds. The corresponding long-exposure images (ground truth) were captured with 100 to 300 times longer.

Implementation Details. We train EEMEFN for 5000 epochs using ADAM (Kinga and Adam 2015) optimizer with an initial learning rate of 10^{-4} , which is decreased to $5 * 10^{-5}$ after 2500 epochs and 10^{-5} after 3500 epochs. Following Chen et al. (2018), we preprocess all raw images by subtracting the black level. The MEF module generates two images with exposure ratios $\{1, k^*/2\}$. Our model takes these two images and their corresponding long-exposure image as an input-output pair to train the model. In experiments, we pretrain the edge detection network using long-exposure images as inputs and fine-tune the network using images generated by the MEF module. With this training strategy, our model converges much faster than the one started from scratch and achieves better performance.

Quantitative Evaluation

We compare our approach with the state-of-the-art methods, including CAN (Chen, Xu, and Koltun 2017) and the U-net used by Chen et al. (2018). We also introduce a baseline that has the same U-net and is fed with concatenated multi-exposure images. The results reported in this paper are based on Tensorflow because most of the comparisons (Chen, Xu, and Koltun 2017; Chen et al. 2018) used this framework.

Table 1 reports quantitative results for low-light image enhancement. It can be seen that the baseline model outperforms CAN and Chen et al., which demonstrates the effectiveness of using multi-exposure low-light images. Besides, thanks to the fusion block, better performance can be achieved by MEF method when compared with the baseline model. In addition, our EEMEFN method is 0.54 dB and 0.43 dB higher than the baseline on the Sony and Fuji set with respect to PSNR. It suggests that exchanging partial information and combining edge information enable our EEMEFN model to make full use of global features from the raw images and local features from the edge information. As this table shows, our approach achieves state-of-the-art performance over other methods on both Sony and Fuji set. For the comparison with (Chen et al. 2018), our EEMEFN improves PSNR from 28.88 dB to 29.60 dB on the Sony set and from 26.61 dB to 27.38 dB on the Fuji set. The SSIM is improved from 0.787 to 0.795 on the Sony set and from

0.680 to 0.723 on the Fuji set. We employ the LPIPS metric (Zhang et al. 2018) to measure the perceptual distance. Compared with (Chen et al. 2018), we reduce the LPIPS distance from 0.476 to 0.458 on the Sony set and from 0.586 to 0.547 on the Fuji set. Overall, the experimental results demonstrate that our proposed EEMEFN achieves the best performance in terms of PSNR and SSIM for low-light image enhancement.

Table 1: Quantitative evaluation of low-light image enhancement algorithms in terms of PSNR/SSIM. The best results are highlighted in bold.

Model	Sony set	Fuji set
CAN	27.40 / 0.792	25.71 / 0.710
Chen et al.	28.88 / 0.787	26.61 / 0.680
baseline	29.06 / 0.787	26.95 / 0.717
MEF	29.43 / 0.791	27.21 / 0.719
EEMEFN	29.60 / 0.795	27.38 / 0.723

Qualitative Evaluation

Figure 3 shows some representative results for visual comparison. We show the processed input image, the ground truth, the results by Chen et al. (2018) and EEMEFN, and the error map between the enhanced image and the ground truth. Note that the input image is processed by the traditional pipeline and linearly scaled to match the reference image for better visual representation. We employ the error map to visualize pixel-level l_1 distance in RGB space. The red and green rectangles represent where the corresponding zoomed sub-images were taken from.

As we can see, the input image captured in extremely low-light condition suffers from information loss, high contrast, and color bias. Although the model by Chen et al. (2018) handles the noise effectively, the output image is blurry. The severely distorted content, e.g., colors and edges, can not be well restored. For instance, the general edges of the train are not clear enough and the green color standing for virescence is missing in Figure 3(c). When the edges of different objects are unclear, it is challenging to reduce such distortion. In contrast, the proposed EEMEFN method reconstructs high-quality images and preserves the local structures by leveraging edge information. Figure 3(f) shows that our approach produces error maps with small values. The error maps demonstrate that our EEMEFN model produces images with higher quality and smaller color distortion. By checking the details from the green rectangle of Figure 3 (d) and (f), it is clear that our method recovers more correct colors and pushes the color distribution closer to the ground truth.

In summary, our proposed model can recover colors by fully exploiting valuable information from multi-exposure images and preserve sharper edges such as the structure of objects and the boundary between objects.

Ablation Studies

For a comprehensive understanding of our model, we conduct ablation experiments to demonstrate the improvements

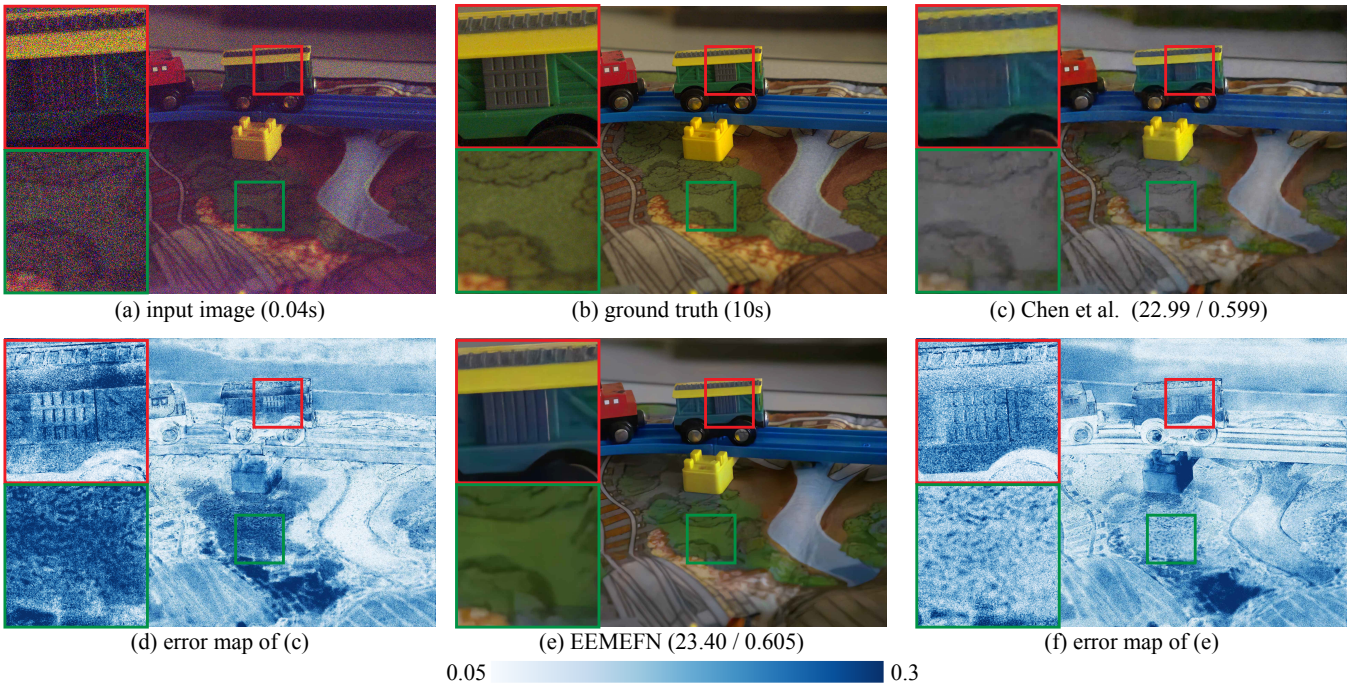


Figure 3: Qualitative results for extremely low-light image enhancement by U-net and our EEMEFN. The numbers in parentheses represent the PSNR and SSIM values.

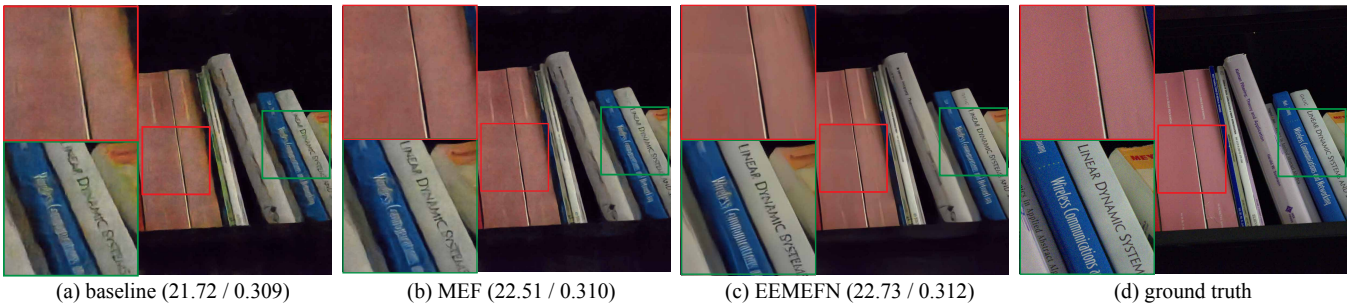


Figure 4: Qualitative comparison of baseline, MEF and EEMEFN.

obtained by each component.

Exposure Ratio. The choice of exposure ratios is essential to our MEF module. Firstly, we evaluate the impact of the exposure ratios. As shown in Table 2, we can achieve higher performance when more multi-exposure images are generated. Though the image generated by exposure ratio $k = 1$ achieves the worst performance, this image can provide more under-exposed information. The exposure ratio set, which includes $k = 1$, leads to better results. We find that both MEF and EEMEFN models with exposure ratios $\{1, k^*/2\}$ or $\{1, k^*/2, k^*\}$ can achieve the best performance. However, taking three images into consideration is a waste of computation resources and cannot improve the performance significantly. Therefore, we choose the exposure ratios as $\{1, k^*/2\}$ during the rest of our experiments.

Fusion Block. Secondly, we study the effectiveness of the fusion block and its variant. To this end, we perform the

following experiments:

- (1) baseline, which employs the U-net architecture. We concatenate multi-exposure images before feeding them into the network.
- (2) basic MEF, which represents a basic multi-exposure fusion module without fusion blocks, such that the information of two U-net branches are fused at the last layer.
- (3) MEF (max), which employs fusion blocks with the max operation.
- (4) MEF (avg), which employs fusion blocks with the average operation.
- (5) MEF (max + avg), which employs the max and average operations in the fusion block.

Table 3 shows the evaluation results of the fusion block and its variant. Note that all models take two multi-exposure images with exposure ratios $\{1, k^*/2\}$ as input. We can see that basic MEF achieves a slightly better performance

Table 2: Ablation study on the choice of exposure ratios. The best results in terms of PSNR/SSIM are highlighted in bold.

Exposure ratio	Model	Sony set	Fuji set
$\{1\}$	U-net	22.46 / 0.732	20.51 / 0.654
$\{k^*/2\}$	U-net	28.90 / 0.785	26.61 / 0.711
$\{k^*\}$	U-net	28.88 / 0.786	26.66 / 0.712
$\{1, k^*/2\}$	MEF	29.43 / 0.791	27.21 / 0.719
$\{1, k^*\}$	MEF	29.41 / 0.792	27.05 / 0.718
$\{k^*/2, k^*\}$	MEF	29.27 / 0.791	27.04 / 0.715
$\{1, k^*/2, k^*\}$	MEF	29.41 / 0.792	27.22 / 0.720
$\{1, k^*/2\}$	EEMEFN	29.60 / 0.795	27.38 / 0.723
$\{1, k^*\}$	EEMEFN	29.58 / 0.795	27.29 / 0.723
$\{k^*/2, k^*\}$	EEMEFN	29.44 / 0.795	27.24 / 0.719
$\{1, k^*/2, k^*\}$	EEMEFN	29.56 / 0.796	27.36 / 0.723

than baseline because basic MEF processes each image individually which can increase the width (number of channels) of image features. Furthermore, MEF (max) and MEF (avg) perform better than basic MEN, because they transmit information between two branches to make full use of the partial information. By combining two aggregation operations, MEF (max + avg) can further improve performance.

Table 3: Ablation study on the design of fusion block. The best performance is highlighted in bold.

Model	Sony set	Fuji set
baseline	29.06 / 0.787	26.95 / 0.717
basic MEF	29.18 / 0.788	26.97 / 0.717
MEF (max)	29.28 / 0.790	27.15 / 0.718
MEF (avg)	29.30 / 0.791	27.16 / 0.718
MEF (max + avg)	29.43 / 0.791	27.21 / 0.719

Edge Enhancement. Thirdly, we study the network design of the edge enhancement module, e.g. the importance of edge map and the initial image. We perform the following experiments:

- (1) MEF, which is the first stage of our EEMEFN model.
- (2) MEF + edge loss, which is a MEF module followed by an edge-preserving loss and a l_1 loss:

$$loss = \|I_{normal} - I^{gt}\|_1 + \lambda * l_{edge}(Detection(I_{normal}), E_{gt}).$$

- (3) EEMEFN ($E + I_{normal}$), which only takes as input edge information and the initial image.
- (4) EEMEFN ($E + I$), which only takes as input edge information and multi-exposure images.
- (5) EEMEFN ($E + I + I_{normal}$), which takes edge information, multi-exposure images and the initial image as input.

Table 4 shows the evaluation results of EE module in terms of PSNR/SSIM. We can see that considering edge information is able to improve performance further. Enhancing the initial image with edge information outperforms an edge-aware smoothing loss, which demonstrates that the edge enhancement module can extract more information

Table 4: Ablation study on the design of edge enhancement network. The best performance is highlighted in bold.

Model	Sony set	Fuji set
MEF	29.43 / 0.791	27.21 / 0.719
MEF + edge loss	28.44 / 0.766	26.96 / 0.689
EEMEFN ($E + I_{normal}$)	29.56 / 0.795	27.32 / 0.722
EEMEFN ($E + I$)	29.15 / 0.790	27.03 / 0.716
EEMEFN ($E + I + I_{normal}$)	29.60 / 0.796	27.38 / 0.723

using edge maps. The comparison between EEMEFN models shows that the initial image is a very critical element which has a significant impact on performance. By feeding the network with multi-exposure images, the network can extract more information with the help of edge maps and initial images. The improvement achieved by our proposed EEMEFN demonstrates the effectiveness of edge enhancement module, which enhances the initial images according to the edge information.

Visual Comparison. Finally, we illustrate a visual comparison among baseline, MEF, EEMEFN and the ground truth in Figure 4. We can see that the baseline method suffers from color bias, e.g., the color of the books. Besides, the book titles and the general edges are not clear enough (Figure 4(a)). MEF utilizes fusion blocks to reduce color bias by fusing high-quality image features among multi-exposure images (Figure 4(b)). However, the result of MEF still suffers from blurry vision, because MEF may average nearby pixels of different objects when the object edge is unknown. The EEMEFN model can reconstruct high-quality images with sharp edges and smooth surfaces under the guidance of the edge map (Figure 4(c)). In summary, the experimental results demonstrate the effectiveness of MEF and EE, which leads to consistent improvement in recovering details, reducing color bias, and enhancing edges.

Conclusions

In this work, we propose a novel edge-enhanced multi-exposure fusion network for low-light image enhancement. The multi-exposure fusion module recovers well-exposed image details and decreases noise variance and color bias by generating and fusing multi-exposure images with different illumination conditions. Also, we introduce an edge enhancement module that yields high-quality images by merging low-light images and the edge information. Our experimental results have shown that our model outperforms against the state-of-the-art approaches in terms of PSNR and SSIM. The proposed method generates high-quality images with abundant textures and sharp edges. In the future, we plan to develop a more powerful architecture for real-time processing and apply the model to other enhancement tasks (e.g., low-light video enhancement).

Acknowledgment

This research has been supported by National Key Research and Development Program (2018YFB0904503) and National Natural Science Foundation of China (61772456).

References

- Aittala, M., and Durand, F. 2018. Burst image deblurring using permutation invariant convolutional neural networks. In *Proceedings of the European Conference on Computer Vision*, 731–747.
- Bertasius, G.; Shi, J.; and Torresani, L. 2015. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4380–4389.
- Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6):679–698.
- Chen, C.; Chen, Q.; Xu, J.; and Koltun, V. 2018. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3291–3300.
- Chen, Q.; Xu, J.; and Koltun, V. 2017. Fast image processing with fully-convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 9, 2516–2525.
- Cheng, H., and Shi, X. 2004. A simple and effective histogram equalization approach to image enhancement. *Digital signal processing* 14(2):158–170.
- Dollár, P., and Zitnick, C. L. 2015. Fast edge detection using structured forests. *IEEE transactions on pattern analysis and machine intelligence* 37(8):1558–1570.
- Godard, C.; Matzen, K.; and Uyttendaele, M. 2018. Deep burst denoising. In *Proceedings of the European Conference on Computer Vision*, 538–554.
- Guo, X. 2016. Lime: a method for low-light image enhancement. In *Proceedings of the ACM on Multimedia Conference*, 87–91. ACM.
- Hasinoff, S. W.; Sharlet, D.; Geiss, R.; Adams, A.; Barron, J. T.; Kainz, F.; Chen, J.; and Levoy, M. 2016. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics* 35(6):192.
- Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Kinga, D., and Adam, J. B. 2015. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations*, volume 5.
- Land, E. H. 1977. The retinex theory of color vision. *Scientific American* 237(6):108–129.
- Lee, C.; Lee, C.; and Kim, C.-S. 2013. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE Transactions on Image Processing* 22(12):5372–5384.
- Liu, Y.; Cheng, M.-M.; Hu, X.; Wang, K.; and Bai, X. 2017. Richer convolutional features for edge detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5872–5881. IEEE.
- Liu, Y.; Cheng, M.-M.; Bian, J.; Zhang, L.; Jiang, P.-T.; and Cao, Y. 2018. Semantic edge detection with diverse deep supervision. *arXiv preprint arXiv:1804.02864*.
- Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.
- Lore, K. G.; Akintayo, A.; and Sarkar, S. 2017. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition* 61:650–662.
- Lv, F.; Lu, F.; Wu, J.; and Lim, C. 2018. Mblen: Low-light image/video enhancement using cnns. In *British Machine Vision Conference*, 220.
- Mildenhall, B.; Barron, J. T.; Chen, J.; Sharlet, D.; Ng, R.; and Carroll, R. 2018. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2502–2510.
- Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.
- Shen, W.; Wang, X.; Wang, Y.; Bai, X.; and Zhang, Z. 2015. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3982–3991.
- Tao, L.; Zhu, C.; Xiang, G.; Li, Y.; Jia, H.; and Xie, X. 2017. Llcn: A convolutional neural network for low-light image enhancement. In *Visual Communications and Image Processing*, 1–4. IEEE.
- Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*.
- Xie, S., and Tu, Z. 2015. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, 1395–1403.
- Ying, Z.; Li, G.; Ren, Y.; Wang, R.; and Wang, W. 2017. A new low-light image enhancement algorithm using camera response model. In *Proceedings of the IEEE International Conference on Computer Vision*, 3015–3022.
- Ying, Z.; Li, G.; and Gao, W. 2017. A Bio-Inspired Multi-Exposure Fusion Framework for Low-light Image Enhancement. *arXiv preprint*.
- Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 586–595.
- Zhu, X.; Wang, Y.; Dai, J.; Yuan, L.; and Wei, Y. 2017. Flow-guided feature aggregation for video object detection. In *Proceedings of the IEEE International Conference on Computer Vision*.