Fusion-Correction Network for Single-Exposure Correction and Multi-Exposure Fusion

Jin Liang, Anran Zhang, Jun Xu, Hui Li, Xiantong Zhen

Abstract—The photographs captured by digital cameras usually suffer from over-exposure or under-exposure problems. The Single-Exposure Correction (SEC) and Multi-Exposure Fusion (MEF) are two widely studied image processing tasks for image exposure enhancement. However, current SEC and MEF methods ignore the internal correlation between SEC and MEF, and are proposed under distinct frameworks. What's more, most MEF methods usually fail at processing a sequence containing only under-exposed or over-exposed images. To alleviate these problems, in this paper, we develop an integrated framework to simultaneously tackle the SEC and MEF tasks. Built upon the Laplacian Pyramid (LP) decomposition, we propose a novel Fusion-Correction Network (FCNet) to fuse and correct an image sequence sequentially in a multi-level scheme. In each LP level, the image sequence is feed into a Fusion block and a Correction block for consecutive image fusion and exposure correction. The corrected image is upsampled and re-composed with the high-frequency detail components in next-level, producing the base sequence for the next-level blocks. Experiments on the benchmark dataset [2] demonstrate that our FCNet is effective on both the SEC and MEF tasks. The code will be publicly released.

Index Terms—Multi-exposure image fusion, single image exposure correction, deep learning, Laplacian Pyramid decomposition

I. INTRODUCTION

Exposure is an essential aspect to influence the brightness and visual quality of the captured image. During the capture process, the exposure may be improper due to inaccurate shutter speed, focal-aperture ratio, or ISO value [2]. Improper exposure is also prone to happen when the illumination of environment is very dark, bright, or in great/fast variation. It largely degrades the visual quality of the captured single-exposed image or multi-exposed image sequence. To address this issue, a number of methods have been proposed to tackle the task of single-exposure correction [1], [2], [7], [50], [35] or multi-exposure fusion [18], [33], [38], [39], [41].

This work was supported in part by The National Natural Science Foundation of China (No. 62002176 and 62176068). Corresponding author: Jun Xu (nankaimathxujun@gmail.com)

Jin Liang is with the School of Mathematics Science, Nankai University, Tianjin 300071, China. (E-mail: liangjinliang@mail.nankai.edu.cn)

Anran Zhang is with the Key Laboratory of Advanced Technology of Near Space Information System, Beihang University, Beijing, China. (E-mail: zhangranan@buaa.edu.cn)

Jun Xu is with the School of Statistics and Data Science, Nankai University, Tianjin 300071, China. (E-mail: nankaimathxujun@gmail.com)

Hui Li is with the Imaging Algorithm Research Department, vivo Mobile Communication Co. Ltd., Shenzhen 518101, China (e-mail: lhiui@vivo.com)

Xiantong Zhen is with University of Amsterdam and Inception Institute of Artificial Intelligence (IIAI) (e-mail: zhenxt@gmail.com).

Fig. 1: Enhanced images by SEC method [2], MEF method [38], and the proposed FCNet. On one hand, the SEC method [2] well corrects the over-exposed image (a) and outputs the image (c), but cannot directly fuse the image sequence (b). On the other, the MEF method [38] well fuses the image sequence (b) and outputs the image (d), but cannot correct the image (a). By integrating SEC and MEF in a unified framework, according to the input image (a) or sequence (b), our FCNet can output visual pleasing results (e) for SEC (a) (left side) and MEF (b) (right side).

Single-exposure correction (SEC) aims to generate a well-exposed and visual pleasing image from a single under-exposed or over-exposed photograph [1], [2], [7], [50]. At early stage, the under-exposure correction task, a.k.a., low-light enhancement, is mainly tackled under the Retinex framework [30]. Most of existing Retinex-based methods [15], [21], [32], [50] mainly decompose an under-exposed image into the reflectance and illumination components, and then enhance the illumination one by gamma correction techniques [56]. Though with promising performance, these Retinex-based methods likely fail at handling over-exposed images. Besides, the optimization process is usually time-consuming. Recent SEC methods [2], [7] employ deep neural networks to learn
exposure correction from pairs of over/under-exposed and properly-exposed images. However, these methods usually produce artifacts on challenging scenarios, which could not be properly presented in a single exposure image [41]. Generative SEC methods [28] can recover the details in over-exposed and under-exposed areas to some extent, but often produce unnatural image colors.

Multi-exposure fusion (MEF) methods [11], [18], [33], [39], [41] generate visually appealing images by exploiting the complementary information from a sequence of over-exposed and under-exposed images. As an alternative to High Dynamic Range (HDR) technology [41], [35], these MEF methods merge a series of images in low dynamic range with different exposure settings at the same scene into a high-quality image. These methods usually suffer from high computational complexity upon a long sequence of high-resolution images. Several MEF methods [22], [38] implement computation exhaustive operations, e.g., alignment or mask prediction, on a downsampled resolution to reduce the computational costs. However, most of existing MEF methods ignore the challenging scenarios with very dark or strong lighting conditions, in which under- or over-exposed images are prone to be captured. For these cases, performing fusion only is not enough to produce a well exposed image. Therefore, it is necessary for MEF methods to correct the exposure of arbitrary frames (include one) for image quality enhancement.

Current SEC and MEF methods are usually developed in different frameworks, which requires duplicated computational resource for both tasks in practical camera imaging pipelines. To exploit the inner correlation between SEC and MEF, it is necessary to perform exposure fusion and correction in a unified manner to achieve visually appealing results on under-exposed or over-exposed image sequences. To this end, in this paper, we develop an integrated framework to simultaneously tackle the SEC and MEF tasks. Specifically, we propose a Fusion-Correction Network (FCNet) to perform exposure fusion and enhancement alternatively. To obtain better results, we employ the Laplacian Pyramid decomposition [5] to divide each image of the sequence into a series of multi-level components, and implement image fusion and correction alternately at different levels for the final well-exposed image. As shown in Figure 1, the SEC method [2] cannot directly tackle MEF, while the MEF method [38] cannot directly tackle SEC. By incorporating the fusion and correction capability into a unified network, our FCNet well handles both the SEC and MEF tasks.

In summary, the contributions of our work are three-fold:
- We develop an integrated framework to simultaneously tackle single-exposure correction (SEC) and multi-exposure fusion (MEF). Our integrated framework can well process an image sequence of arbitrary length (including one) to output a well-exposed image.
- Under the Laplacian Pyramid decomposition framework, we propose a Fusion-Correction Network (FCNet) with efficient inference speed over the other competitors on handling long sequences of high-resolution images.
- Experiments show that our FCNet achieves comparable performance with existing methods on SEC, while better performance on MEF tasks, especially with over-exposed or under-exposed image sequences.

The rest of this paper is organized as follows. In §II, we briefly introduce the related works of our FCNet. In §III, we present our FCNet for SEC and MEF tasks. Extensive experiments are conducted in §IV to evaluate the performance of our FCNet, with in-depth analysis. §V concludes this work.

II. RELATED WORK

Single-exposure correction (SEC). The Retinex theory [30] is widely studied to tackle the under-exposure problem by decomposing an image into the illumination and reflection components and correcting the illumination [15], [21], [32], [39], [56]. Fu et al. [15] analysed the side effect of logarithmic transformation and designed a weighted variation model to refine the regularization terms. In [21], Guo et al. first estimated a coarse illumination map and refined it by optimization techniques. The work of [32] introduces a noise term [59], [58], [25] for robust enhancement. However, these methods have to solve complex optimization problems, which are often very time-consuming. Recently, deep learning methods [8], [20], [28] have achieved efficient performance on single image exposure correction. Chen et al. [8] learned to directly enhance the under-exposed raw sensor images. Guo et al. [20] proposed to estimate adjustment curves for under-exposure correction. EnlightenGAN [28] is an unsupervised GAN-based network trained by unpaired of under-exposed and normal images. Since focusing on the under-exposed images, these methods are unable to correct over-exposed images.

The work of [2] is among the first deep learning based method for both over- and under-exposure correction. Due to the limited exposure information in a single image, [2] is usually not robust enough upon challenging scenarios. In this paper, we propose an integrated framework to flexibly handle over- or under-exposure correction and produce perceptually appealing results for multi-exposure fusion.

Multi-exposure fusion (MEF) is a promising alternative for high dynamic range imaging [41], [48]. Early MEF methods mainly resort to pixel-wise operations. Mertens et al. [41] proposed a multi-scale fusion mechanism via Laplacian pyramid decomposition [5], which is boosted by Shen et al. [48] on computational efficiency. To well preserve details in both bright and dark regions within limited dynamic range, the methods of [3], [33], [46] separately process the base and detail components of the images decomposed by edge-preserving filtering [23], [49]. Gradient information is also exploited in [4], [19], [44] for detail-enhanced MEF. However, pixel-wise methods is not robust to the image sequences with misalignment problem [39]. For this, later methods [18], [31], [39] usually perform MEF in a patch-wise manner to optimize the MEF-SSIM metric [40]. But patch-based methods prone to produce over-smooth results upon dynamic image sequences.

Recently, deep MEF methods [38], [45], [55] achieve robust performance upon challenging scenarios. DeepFuse [45] converts the image sequence to YCbCr format and performs fusion on the Y channel. To achieve efficient MEF, Ma et al. [38] proposed to predict fusion maps in a downsampled resolution, and upsampled the maps to the original resolution for final fusion. MEF-GAN [55] achieves visually appealing
results using the generative adversarial networks [17]. Overall, these MEF methods could not correct improper-exposed images. In this paper, we explore an integrated framework to tackle both the SEC and MEF tasks.

Multi-task Learning. Multi-Task Learning (MTL) aims to tackle several correlated tasks simultaneously, by exploiting the similarity and considering the differences among these tasks. Over past decades, MTL has been studied in various areas, such as natural language processing [10], [37], speech recognition [12], [9] and video analysis [54], [52], [53]. Comparing with single task case, MTL methods often reduce the memory cost and increase the inference speed due to its inherent layer sharing [53], but can potentially improve the performance on all tasks by sharing useful information between correlated tasks [54]. One MTL method closely related to our work is the Deep Coupled Feedback Network (CF-Net) [13], which improves the image quality via joint multi-task learning to tackle several correlated tasks simultaneously, by exploiting the similarity and considering the differences among these tasks.

III. PROPOSED FUSION-CORRECTION NETWORK

Our Fusion-Correction Network (FCNet) mainly contains three parts: Laplacian Pyramid decomposition (§III-A), Fusion-Correction (FC) block (§III-B), and Base-Detail composition between two FC blocks (§III-C), as shown in Figure 2.

A. Laplacian Pyramid Decomposition

Our FCNet employs the Laplacian Pyramid (LP) decomposition [5] to perform exposure enhancement in a coarse-to-fine manner. Given a sequence of \( K \geq 1 \) images \( \{I_k \in \mathbb{R}^{h \times w}\}_{k=1}^{K} \) of arbitrary exposures, we first decompose each image \( I_k \) into \( n \) hierarchical layers, including \( n - 1 \) high-frequency detail components \( H_k = \{H_k^1, H_k^2, \ldots, H_k^{n-1}\} \) and the low-frequency base component \( L_k^n \) at the \( n \)-th layer. Here, \( H_k^i \) is of size \( \frac{1}{2}h \times \frac{1}{2}w \) while \( L_k^n \) is of size \( \frac{1}{2^n}h \times \frac{1}{2^n}w \). The final image is reconstructed from these components sequentially by the proposed Fusion-Correction block introduced as follows.

B. Fusion-Correction Block

To tackle SEC and MEF in an integrated framework, we propose a Fusion-Correction (FC) block to consecutively enhance the image sequence under the LP framework. Given the \( n \)-th base image sequence \( \{L_k^n\}_{k=1}^{K} \), our FC block first weighted sums the base sequence by a Fusion block and then enhances the fusion result by a UNet-like Correction block. The output image \( O^n \) will be upsampled and composed with the high-frequency components \( \{H_k^{n-1}\}_{k=1}^{K} \) at level \( n-1 \), to generate the base sequence \( \{L_k^{n-1}\}_{k=1}^{K} \) as the input of the next FC block. We implement FC block on the base component \( \{L_k^n\}_{k=1}^{K} \) at layer \( i\)-th and composition on \( O^i \) with \( \{H_k^{n-1}\}_{k=1}^{K} \) (\( i = n, n-1, \ldots, 1 \)) alternately to output the final well-exposed image \( O^1 \) (we do not perform composition when \( i = 1 \)).

Fusion block. We perform image fusion before exposure correction to reduce the potential large memory consumption and computational costs, when dealing with a long image sequence. Besides, a good fusion result tends to have rich details and little noise [57], reducing the difficulty of exposure correction that is prone to bring displeasing artifacts [7]. Inspired by the acceleration schemes in [22], [16], [38], we design our fusion block under the “Downsample-Execute-Upsample” scheme [38]. Specifically, the Fusion block in the \( i\)-th \( (i = 1, \ldots, n) \) LP level first downsamples the base sequence \( \{L_k^n\}_{k=1}^{K} \) by a bilinear interpolation and predicts the weight

Fig. 2: Overview of the proposed FCNet for both single exposure correction and multi-exposure fusion. Given an input image sequence of arbitrary length \( I = \{I_k \in \mathbb{R}^{h \times w}\}_{k=1}^{K} \) \( (K \geq 1) \), we decompose the images into different Laplacian pyramid levels, and perform fusion and correction level-by-level in a coarse-to-fine manner. Our FCNet sequentially fuses and corrects the Laplacian pyramid images in each level, by a series of fusion and correction blocks. The fusion block, in a light-weight structure, produces weigh maps for weighted image summation. The correction block is in a UNet-like structure.
TABLE I: Detailed structure of the fusion block in our FCNet. \( m \) is the number of intermediate convolutional layers.

| Layer | Channel | Kernel Size | Stride | Padding | Dilation |
|-------|---------|-------------|--------|---------|----------|
| Conv\(v\) | \(3 \rightarrow 24\) | \(3 \times 3\) | 1 | 1 | 1 |
| Conv\(v\) \(\text{from 1 to } m\) | \(24 \rightarrow 24\) | \(3 \times 3\) | 1 | \(2'\) | \(2'\) |
| Conv\(v\) \(\text{+1}\) | \(24 \rightarrow 24\) | \(3 \times 3\) | 1 | 1 | 1 |
| Conv\(v\) \(\text{+2}\) | \(24 \rightarrow 3\) | \(1 \times 1\) | 1 | 0 | 1 |

maps \(\{W_k\}^K_{k=1}\) by several dilated convolutions \([60]\) at a downsampled size of \(\{L_k\}^K_{k=1}\). Then we upsample the weight maps \(\{W_k\}^K_{k=1}\) with the guidance of the input sequence \(\{L_k\}^K_{k=1}\) \([23]\) to fuse them at their original resolutions. The fused image \(F^i\) in the \(i\)-th layer is obtained as follows:

\[
F^i = \bigoplus_{k=1}^K g^i(W_k) \otimes L_k^i, \tag{1}
\]

where \(\otimes\) is the Hadamard product and \(g^i\) denotes the guided upsampling operation \([23]\). The dilated convolutions used in each Fusion block are shown in Table I. In each Fusion block, we use \(m\) intermediate convolutional layers between the first convolutional layer and last two convolutional layers. Here, we set \(m = 4, 3, 2, 1\) in corresponding LP levels. Thus, the size of Fusion block decreases gradually in different LP levels.

**Correction block.** Given the fused image \(F^i\), a Correction block at the \(i\)-th layer is applied on it to produce an exposure-enhanced image \(O^i\). Since the UNet \([47]\) proves to be effective in many low-level tasks \([14], [36], [26]\), our Correction block is also designed as a UNet-like network. To alleviate potential checkerboard artifacts \([28]\), we replace the standard deconvolutional layer at each upsampling stage of UNet by a bilinear upsampling layer with a \(3 \times 3\) convolutional layer. To improve its correction capability, we add a skip connection between the encoder and decoder of the UNet-like network. The Correction block in the 4-th LP layer is a 4-layer UNet-like network, which has 24 channels in the first convolutional layer of the encoder. As a standard UNet, the number of channels from layer to layer is doubled with halved image resolution. The rest Correction blocks are designed in the same manner. The second Correction block is also a 4-layer UNet-like network and has 16 channels in the first convolutional layer of the encoder. The third and the final Correction blocks are 3-layer UNet-like networks with 16 channels in the first convolutional layer of the encoder. Overall, the size of FC block gradually decreases with the progress of the image reconstruction. This enables our FCNet focus more on fusing and correcting the based LP levels of the input image or image sequence. The ablation studies in §IV-D validate the effectiveness of this structure design for our Fusion and Correction blocks.

**C. Base-Detail Composition**

As the image reconstruction process under the LP framework \([5]\), the image \(O^i\) output by the FC block at layer \(i\) \((i = n, ..., 2)\) will be upsampled and composed with the high-frequency components \(\{H_k^{i-1}\}\) in layer \(i - 1\). The resulting base sequence \(\{L_k^{i-1}\}\) is the input of the FC block at layer \(i - 1\). The composition process is formulated as:

\[
L_k^{i-1} = f^i(O^i) + H_k^{i-1}, k = 1, ..., K, \tag{2}
\]

where \(f^i\) denotes the upsampling operation implemented by a bilinear interpolation layer and a \(3 \times 3\) convolutional layer.

The intermediate images output in different stages of our FCNet are shown in Figure 3. One can see that, from an input base image sequence, our FCNet successively fuses the base image sequence in different levels and corrects the fused image subsequently with enhanced image quality.

**D. Loss Function**

To endow our FCNet with the capability to well handle the SEC and MEF tasks simultaneously, our FCNet is trained end-to-end by the following loss function:

\[
\mathcal{L} = \mathcal{L}_r + \mathcal{L}_{pr} + \lambda \mathcal{L}_{ps}, \tag{3}
\]

where \(\mathcal{L}_r\) is the reconstruction loss, \(\mathcal{L}_{pr}\) is the pyramid reconstruction loss, \(\mathcal{L}_{ps}\) is the pyramid spatial consistency loss, and \(\lambda\) is used to trade-off different terms.

**Reconstruction loss.** We use the \(L_1\) loss function to penalize the gap between the final output \(O^1\) and the corresponding ground truth \(G\) as follows:

\[
\mathcal{L}_r = \sum_{p=1}^{3hw} |O^1(p) - G(p)|. \tag{4}
\]

**Pyramid reconstruction loss.** As suggested by \([2]\), we decompose the ground-truth image \(G\) into a Gaussian pyramid...
and use each pyramid layer to supervise the corresponding output by the FC block at that layer, as follows:

\[ L_{pr} = \sum_{i=2}^{n} \sum_{p=1}^{3h_iw_i} \left( G^i(p) - O^i(p) \right), \quad (5) \]

where \( O^i \) is the output image of \( i \)-th FC block and \( G^i \) is the \( i \)-th pyramid layer decomposed from the ground truth image \( G \). \( h_i = \frac{1}{2^i} h \) and \( w_i = \frac{1}{2^i} w \) are the height and width of \( O^i \). The pyramid reconstruction loss \( L_{pr} \) supervises the intermediate output of each FC block, making our FCNet robust to large exposure deviation during the reconstruction. Our ablation study in §IV-D also validates the effectiveness of the pyramid reconstruction loss on exposure correction.

Pyramid spatial consistency loss. To improve our FCNet robustness upon challenging images with large homogeneous regions, here we propose a pyramid spatial consistency loss \( L_{ps} \) to preserve the difference of neighboring regions between the image output and its corresponding ground-truth. Inspired by [20], we divide the image \( O^i \) (output by the FC block at layer \( i \)) and the \( i \)-th pyramid ground-truth layer \( G^i \) into two sets of \( M \) corresponding regions, and penalty the discrepancy between the image \( O^i \) and the layer \( G^i \) on the average difference between the \( j \)-th (\( j = 1, ..., M \) region and its neighboring regions. Specifically, the \( L_{ps} \) is defined as:

\[ L_{ps} = \sum_{i=1}^{n} \left( G^i_h - O^i_h \right)^2 + \sum_{j=1}^{M} \sum_{h \in \Omega(j)} \frac{1}{M} \left( \left| G^i_h - O^i_h \right| - \left| G^i_j - O^i_j \right| \right)^2, \quad (6) \]

where \( M \) is the number of regions, \( \Omega(j) \) denotes the regions centered at region \( j \), \( O^i_j \) is the region \( j \) of \( i \)-th layer image restored by our FCNet, \( G^i_j \) is the region \( j \) of \( i \)-th pyramid layer decomposed from the ground-truth image \( G \).

E. Implementation Details

In our FCNet, we set \( n = 4 \) in the Laplacian Pyramid decomposition. The parameters of our FCNet are initialized by Kaiming-initialization [24] and optimized by Adam [29] with default parameters. Our FCNet is trained in a total of 150 epochs. The learning rate is initialized as \( 10^{-4} \) and decayed by a factor of 0.8 for every 50 epochs. In training, we randomly select 1 \( \sim \) 10 image(s) from each multi-exposure image sequence. The batch size is 1, since we cannot fix the number of images beforehand. We set \( \lambda = 4000 \) to trade-off different loss terms. Our FCNet is implemented in PyTorch [43] and trained on a Titan RTX GPU with 24Gb memory.

IV. Experiment

A. Dataset and Metric

Dataset. We use the dataset in [2], which contains 24,330 images. As far as we know, it is the largest dataset with multi-exposed image sequences. The images in [2] are rendered from MIT-Adobe FiveK dataset [6] with 5 digital Exposure Values (EVs, i.e., -1.5, -1, +0, +1, +1.5) estimated by Adobe Camera Raw SDK. This dataset is divided into a training set with 17,675 images, a validation set with 750 images, and a test set with 5,905 images. As suggested in [16], we choose the images retouched by “Expert C” in [6] as the ground-truths during the training and test stages.

Metric. We evaluate different methods by employing the widely used Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [51] metrics.

B. Comparison on Multi-Exposure Fusion

Experimental setting. We evaluate our FCNet on three fusion tasks: Under-Exposure Fusion (Under-EF) on three images rendered with EVs of +0, -1, and -1.5, Over-Exposure Fusion (Over-EF) on three images rendered with EVs of +0, +1, and +1.5, and Multi-Exposure Fusion (MEF) on all five images rendered with EVs of -1, -1.5, +0, +1, and +1.5.

Comparison methods. We compare our FCNet on the three fusion tasks with three MEF methods: Mertens09 [41], MEF-GAN [55], MEF-Net [38], and two single-exposure correction (SEC) methods: Zero-DCE [20], and MSEC [2]. We evaluate these methods by their official implementations with default settings, except that we use the provided results by MSEC [2].
Fig. 5: Qualitative comparison on an over-exposed image sequence (with EVs of +0, +1, +1.5) by different MEF methods (Mertens09 [41], MEF-GAN [55], MEF-Net [38]) and SEC methods (Zero-DCE [20], MSEC [2]).

Since MEF-GAN only fuses two images, we run it for different two-exposure combinations and provide the results with the best visual quality. To perform Zero-DCE and MSEC on MEF, we directly apply them on each image of a multi-exposure sequence, and average the exposure-corrected images.

Quantitative results on PSNR and SSIM by different methods are listed in Table II. One can see that our FCNet achieves higher PSNR and SSIM results than the other methods [38], [41], [55] on the Over-EF and MEF tasks, and comparable results on the Under-EF task. The MEF methods [38], [41], [55] can hardly achieve promising fusion results on the Under-EF or Over-EF task, which may attributed to the lack of correction capability. By simply averaging the corrected images, MSEC [2] achieves competing PSNR and SSIM results, demonstrating the effectiveness of exposure correction on multi-exposure fusion tasks. By integrating fusion and correction into a single framework, our FCNet is able to achieve robust results upon all three fusion tasks.

Visual quality. The results of MEF with all EVs by different methods are shown in Figure 4. One can see that, when compared to the ground-truth, our FCNet well preserves the color and contrast information. The Mertens09, MEF-Net, or MEF-GAN prone to generate images with a certain contrast or color bias. The result of Zero-DCE is over-exposed, while the result of MSEC presents clear color shift.

We also show the results on Over-EF in Figure 5. We observe that the result of our FCNet is closer to the ground-truth than those of other methods. Due to lack of correction
capability, Mertens09 and MEF-Net produce over-exposed images. Performing Over-EF for MSEC by simple averaging brings clear color bias, e.g., the ship. As a low-light enhancer, Zero-DCE cannot obtain well-exposed results upon over-exposed images. MEF-GAN suffers from clear color deviation on the Over-EF task, since it is trained on pairs of under- and over-exposed images. Besides, as shown in Figure 6, on the Under-EF task, the image contrast and color saturation of the image produced by our FCNet are closest to that of ground-truth over the other competitors.

C. Comparison on Single Exposure Correction

Comparison methods. Our FCNet is able to tackle single-exposure correction (SEC). Here we compare with 5 state-of-the-art SEC methods: WVM [15], LIME [21], EnlightenGAN [28], Zero-DCE [20], and MSEC [2].

Quantitative comparisons. In Table III, we observe that our FCNet arrives at comparable PSNR and SSIM results with MSEC [2] and much better results than the others. Our FCNet is slightly better than the MSEC on the over-exposure correction tasks, while is slightly inferior to the MSEC on under-exposure correction. These results validate the flexibility and effectiveness of our FCNet on SEC.

Visual quality. As shown in Figure 7, similar to MSEC [2], our FCNet achieves visually favorable results on correcting over-exposed images. By contrast, the methods of LIME and WVM produce over-bright images, while EnlightenGAN and Zero-DCE have obvious flaws such as whitening and artifacts. For under-exposed correction (low-light enhancement), as shown in Figure 8, MSEC, LIME and WVM lead to some over-exposure artifacts (e.g., the top of the image), while EnlightenGAN and Zero-DCE generate unsatisfactory visual results in terms of both image contrast and naturalness. On the contrary, our FCNet generates proper results with reasonable image contrast and vivid color. This demonstrates that our FCNet is also very effective on the SEC task.

D. Ablation Study

We now conduct detailed examinations of our FCNet on the SEC and MEF tasks to assess: 1) the effectiveness of gradually decreasing the size of the proposed FC block during the image reconstruction progress; 2) how the decomposition depth in LP affects the performance of our FCNet; 3) the necessity of combining Fusion and Correction blocks to process arbitrary number of input frames; 4) the order of Fusion and Correction blocks; 5) the influence of loss functions to our FCNet.

1) The effectiveness of gradually decreasing the size of FC block during the progress of the image reconstruction. As most of the illumination and color information of the image sequence is mainly extracted at the bottom LP levels, it is efficient to use larger networks to fuse and correct the bottom levels with lower resolutions and smaller networks to process...
the top LP levels with higher resolutions [34], as introduced in §III-B. To validate the effectiveness of this design for our FCNet, we compare three different network design schemes, which are described as follows.

**Model 1.** As the basis for comparison, this model applies the smallest Fusion block and Correction block in each LP level. The parameter $m$ of each fusion block is set as 1, and each correction block is a 3-layer UNet-like network with 16 channels in the first convolutional layer of the encoder. We denote this model as “Small→Small”.

**Model 2.** In this model, the size of FC block gradually increases with the progress of the image reconstruction. For different LP levels from bottom to top, the parameter $m$ of four Fusion blocks are set as 1,2,3,4, respectively. The first and second Correction blocks are a 3-layer UNet-like network, with 16 channels in the first convolutional layer of the encoder. The third Correction block is a 4-layer UNet-like network and has 16 channels in the first convolutional layer of the encoder. The final (4-th) Correction block is a 4-layer UNet-like network, but with 24 channels in the first convolutional layer of the encoder. We denote this model as “Small→Large”.

**Model 3.** Contrary to the Model 1, this model applies the largest Fusion block and Correction block to each LP level. The parameter $m$ of each Fusion block is set as 4. All the Correction blocks are 4-layers UNet-like network, with 24 channels in the first convolutional layer of the encoder. We denote this model as “Large→Large”.

**Model 4.** This is our FCNet, in which the FC blocks are introduced in §III-B. We denote this model as “Large→Small”.

As shown in Table IV, our FCNet (Model 4, “Large→Small”) with decreasing amounts of model size during the image reconstruction progress obtains better PSNR and SSIM results on SEC and MEF, similar results on Under-EF, while comparable results on Over-EF, when compared with the Model 3 which employs largest Fusion block and Correction block in each LP level. At the same time, the Model 1 (“Small→Small”) with smallest Fusion block and Correction block has fewest learnable parameters and FLOPs, but cannot well handle the tasks. It is worth noting that the FLOPs of our FCNet and Model 1 are similar, but our FCNet achieves better performance than Model 1 on both SEC and MEF tasks. Thus it is reasonable and effective to allocate more parameters and computational costs to the bottom LP levels, which contain more visual attributes than the top LP levels under the LP decomposition framework.

2) **How the depth of Laplacian Pyramid decomposition affects the performance of our FCNet on SEC and MEF?**

For this problem, we train our FCNet by varying the depth of
Fig. 8: Qualitative comparison on low-light enhancement task by different single exposure correction methods.

Fig. 9: Ablation study of the contribution of each loss (pyramid reconstruction loss $L_{pr}$ and pyramid space consistency loss $L_{ps}$) with EVs of -1.5, -1, +0, +1, +1.5. Also, in order to prove the necessity of using loss function to constrain the output of each layer of model, we compared the effect of loss functions which only applying in the last layer output of the model (reconstruction loss $L_r$ and space consistency loss $L_s$). Red boxes indicate the obvious differences and amplified details.

TABLE IV: PSNR and SSIM results of our FCNet with difference designs for the FC block on the Single-Exposure Correction (SEC) and Multi-Exposure Fusion (MEF) tasks. The best results are highlighted in bold. We also compare the number of learnable parameters (Params.) and FLOPs upon processing five $512 \times 512$ multi-exposed images, in our FCNet.

| Model   | Task       | SEC PSNR | SEC SSIM | Under-EF PSNR | Under-EF SSIM | Over-EF PSNR | Over-EF SSIM | MEF PSNR | MEF SSIM | Network Complexity |
|---------|------------|----------|----------|--------------|--------------|-------------|-------------|----------|----------|-------------------|
|         |            | PSNR     | SSIM     | PSNR         | SSIM         | PSNR        | SSIM        | PSNR     | SSIM     | Params. (×M) FLOPs (×G) |
| Small→Small | SEC       | 18.04    | 0.7780   | 18.65        | 0.7649       | 19.01       | 0.7816      | 19.75    | 0.8228   | 0.530 12.16 |
| Small→Large  | SEC       | 18.56    | 0.7816   | 18.86        | 0.7608       | 19.27       | 0.7878      | 20.11    | 0.8379   | 2.055 34.18 |
| Large→Large   | SEC       | 19.27    | 0.8025   | 20.51        | 0.8271       | 20.38       | 0.8411      | 20.42    | 0.8444   | 4.904 39.78 |
| Large→Small   | SEC       | 19.71    | 0.8054   | 20.14        | 0.8160       | 20.41       | 0.8393      | 20.81    | 0.8465   | 2.055 12.90 |

LP decomposition as $n = 1, 2, 3, 4$. For a fair comparison, we adjust the number of filters in convolutional layers to guarantee similar parameter amounts in different networks. As shown in Table V, the performance of these variants varies on different exposure fusion (EF) tasks, but drops clearly on SEC task as the LP depth $n$ decreases. To trade-off the performance of our FCNet on correction and fusion tasks, we set the depth $n = 4$ in the LP decomposition of our FCNet.

3) The necessity of combining Fusion block and Correction block to process input image sequence with arbitrary length. To study this problem, we develop two variants of our FCNet by only using Fusion blocks (“Fusion Only”) or only using Correction blocks (“Correction Only”). The comparison results shown in Table VI demonstrate that, the variant “Fusion Only” well handles the MEF task but fails on the SEC task. The variant “Correction Only” can handle both the SEC and MEF tasks. Note that for MEF, it performs correction image-by-image and simply averages the corrected images. By combining Fusion and Correction blocks, our FCNet obtains robust PSNR/SSIM results on both SEC and MEF.

4) The order between Fusion and Correction blocks. In our FCNet, we implement the Fusion block before the Correction block in each level of LP decomposition. If we change the order of these two blocks in all 4 LP levels, the resulting computational costs and memory consumption is prohibitive on our RTX GPU with 24G memory. Thus, here we only...
TABLE V: PSNR and SSIM results of our FCNet with difference LP depth $n$ on the Single-Exposure Correction (SEC) and Multi-Exposure Fusion (MEF) tasks. For MEF, we also include Under-EF and Over-EF settings described in Table II. The best results are shown in bold.

| $n$ | SEC | Under-EF | Over-EF | MEF |
|-----|-----|----------|--------|-----|
|     | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| 1   | 17.76 | 0.793 | 19.55 | 0.792 | 20.08 | 0.827 | 21.54 | 0.827 |
| 2   | 17.74 | 0.755 | 19.12 | 0.778 | 21.53 | 0.843 | 21.56 | 0.829 |
| 3   | 18.41 | 0.771 | 20.27 | 0.792 | 21.05 | 0.836 | 21.93 | 0.837 |
| 4   | 19.71 | 0.805 | 20.14 | 0.816 | 20.41 | 0.839 | 20.81 | 0.847 |

TABLE VI: Results of our FCNet with only fusion or correction blocks on Single-Exposure Correction (SEC) and Multi-Exposure Fusion (MEF) tasks. Note that the “Fusion Only” model does not change the input image on SEC.

| Model | Task | SEC | PSNR | SSIM | MEF | PSNR | SSIM |
|-------|------|-----|------|------|-----|------|------|
| Fusion Only | Correction Only | 16.40 | 0.7399 | 21.67 | 0.8364 | 19.69 | 0.8202 | 20.05 | 0.8386 |
| Our FCNet | Correction Only | 19.71 | 0.8054 | 20.81 | 0.8465 |

TABLE VII: Results of our FCNet with different orders of fusion and correction blocks on Single-Exposure Correction (SEC) and Multi-Exposure Fusion (MEF) tasks.

| Model | Task | SEC | PSNR | SSIM | MEF | PSNR | SSIM |
|-------|------|-----|------|------|-----|------|------|
| Correction→Fusion | Fusion→Correction | 19.31 | 0.794 | 20.71 | 0.838 | 19.71 | 0.805 | 20.81 | 0.847 |

TABLE VII: Comparison of model complexity and running time by different methods on the MEF task. We compare the FLOPs ($\times$G) and running time ($\times$second) by our FCNet and different methods in processing five multi-exposed images in different resolutions. * on MEF-GAN means that, since MEF-GAN can only tackle two-exposure fusion task, we compute its FLOPs and running time when fusing two images.

| Model | Task | 256 × 256 | 512 × 512 | 1024 × 1024 |
|-------|------|----------|----------|-------------|
| Params ($\times$M) | FLOPs Time | FLOPs Time | FLOPs Time |
| MEMT09 | 0.448 | 61.87 | 0.950 | 247.49 | 2.348 | 989.95 | 8.474 |
| MEF-GAN* | 0.026 | 8.62 | 0.008 | 34.48 | 0.030 | 137.90 | 0.121 |
| MEF-Net | 0.079 | 26.02 | 0.007 | 104.09 | 0.009 | 416.57 | 0.014 |
| Zero-DCE | 2.055 | 3.22 | 0.014 | 12.90 | 0.016 | 51.58 | 0.033 |
| FCNet (Ours) | 0.816 | 0.792 | 0.837 |

TABLE IX: Results of our FCNet with different loss functions on Single-Exposure Correction (SEC) and Multi-Exposure Fusion (MEF). The best results are shown in bold.

| Loss | Task | SEC | PSNR | SSIM | MEF | PSNR | SSIM |
|------|------|-----|------|------|-----|------|------|
| $L_r$ | Correction→Fusion | 19.14 | 0.7848 | 20.14 | 0.8208 |
| $L_r + L_{pr}$ | Correction→Fusion | 19.27 | 0.8042 | 20.13 | 0.8420 |
| $L_r + L_{pr} + L_s$ | Correction→Fusion | 18.73 | 0.7822 | 19.21 | 0.8100 |
| $L_r + L_{pr} + L_{ps}$ | Correction→Fusion | 19.71 | 0.8054 | 20.81 | 0.8465 |

TABLE VIII: Comparison of model complexity and running speed on different loss functions. In this paper, we proposed a Fusion-Correction Network (FCNet) to simultaneously tackle the Single-Exposure Correction (SEC) and Multi-Exposure Fusion (MEF) tasks in an integrated framework. In our FCNet, we implemented Laplacian Pyramid (LP) decomposition to exploit multi-scale context information of natural images. In each LP level, we processed the base image sequence by our Fusion and Correction blocks sequentially. The Fusion is implemented before correction to reduce the computation costs. The Fusion block is a light-weight network for efficiency consideration, while the Correction block is a UNet-like network to enhance the exposure of the fused image. The corrected image is

change the order in the first FC Block of our FCNet for comparisons. As shown in Table VII, the results of “first Correction then Fusion” (Correction→Fusion) are slightly inferior to our original FCNet (Fusion→Correction). Since it is prone to introduce unpleasing artifacts when correcting challenging over- or under-exposed images, applying fusion first in our FCNet would generate a nearly well exposed image and thereby reduce the correction difficulty. In addition, compared with correcting a single fusion image, correcting a multi-exposure image sequence before fusion needs more computational costs and larger GPU memory. Therefore, we finally adopt the idea of “first Fusion then Correction”. 5) The contribution of each term loss to our FCNet on SEC and MEF. To study the effect of different reconstruction and spatial loss functions to our FCNet on constraining the output of each LP level, we evaluate our FCNet by employing these loss functions only on the final output image. We denote $L_r$ as the reconstruction loss function in Eq. (3), and denote the $L_s$ as the spatial loss function only on the final output. The $L_r$ can be viewed as a general form of the $L_{ps}$ in Eq. (6) with $i = 1$. As shown in Figure 9, when our FCNet is trained only with $L_r$, it would produce some unnatural color spots. Our FCNet trained with the reconstruction loss $L_r$ and pyramid loss $L_{pr}$ alleviates this problem, but produces results with low contrasts. Besides, our FCNet trained with both $L_s$, $L_r$ and $L_{pr}$ produces brighter results. By combining the reconstruction loss $L_r$, pyramid reconstruction loss $L_{pr}$, and our pyramid spatial consistency loss $L_{ps}$, our FCNet excludes the influence of artifacts to produce well-exposed images. These results show that our loss function plays an important role in our FCNet to produce visually pleasing results. Table IX also shows that our FCNet trained with $L_r$, $L_{pr}$, and $L_{ps}$ attains the best results.

E. Comparison on Model Complexity and Speed

In Table VIII, we compare the number of parameters and FLOPs by different SEC and MEF methods, on five multi-exposed images with the resolution of $256 \times 256$, $512 \times 512$, or $1024 \times 1024$. We also compare the running time of our FCNet with the other methods on the MEF task, on five multi-exposed images with the resolution of $256 \times 256$, $512 \times 512$, or $1024 \times 1024$. The running times of different methods are evaluated on a computer with an Intel i9 10920X CPU and a Titan RTX GPU. We observe that our FCNet is faster than MEF-GAN [55] and MEF-Net [38], but slower than Zero-DCE [20]. By employing the LP decomposition, our FCNet assigns minor computational costs on high-resolution top LP levels, and thus its running time is relatively robust to the increasing of image resolutions over the other methods.

V. CONCLUSION

In this paper, we proposed a Fusion-Correction Network (FCNet) to simultaneously tackle the Single-Exposure Correction (SEC) and Multi-Exposure Fusion (MEF) tasks in an integrated framework. In our FCNet, we implemented Laplacian Pyramid (LP) decomposition to exploit multi-scale context information of natural images. In each LP level, we processed the base image sequence by our Fusion and Correction blocks sequentially. The Fusion is implemented before correction to reduce the computation costs. The Fusion block is a light-weight network for efficiency consideration, while the Correction block is a UNet-like network to enhance the exposure of the fused image. The corrected image is
upsampled and composed with the high-frequency components in the next LP level. By integrating of fusion and correction blocks, our FCNet is feasible to process an image sequence of arbitrary length (including one). Experimental results on the benchmark dataset [2] demonstrated that, our FCNet not only achieves competitive or even better performance on both MEF and SEC tasks when compared to the corresponding state-of-the-arts methods, but also well resolves the Over-EF and Under-EF tasks that previous MEF methods usually fail at.

REFERENCES

[1] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. IEEE Transactions on Consumer Electronics, 53(2):593–600, 2007.

[2] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In IEEE Conf. Comput. Vis. Pattern Recogn., pages 9157–9167, 2021. 1, 2, 4, 5, 6, 7, 8, 9, 11

[3] Durga Prasad Bavirisetti, Gang Xiao, Junhao Zhao, Ravindra Dhuli, and Gang Liu. Multi-scale guided image and video fusion: A fast and efficient approach. Circuits, Systems, and Signal Processing, 38(12):5576–5605, 2019.

[4] Marcelo Bertalmio and Stacey Levine. Variational approach for the fusion of exposure bracketed pairs. IEEE Trans. Image Process., 22(2):712–723, 2012.

[5] Peter J Bart and Edward H Adelson. The laplacian pyramid as a compact image code. In Readings in Computer Vision, pages 671–679. Elsevier, 1987.

[6] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédéric Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In IEEE Conf. Comput. Vis. Pattern Recogn., pages 1–10. 2011.

[7] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. IEEE Trans. Image Process., 27(4):2049–2062, 2018.

[8] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In IEEE Conf. Comput. Vis. Pattern Recogn., pages 3291–3300, 2018.

[9] Zhuo Chen, Shijin Watanabe, Hakan Erdogan, and John R Hershey. Speech enhancement and recognition using multi-task learning of long short-term memory recurrent neural networks. In Sixteenth Annual Conference of the International Speech Communication Association, 2015.

[10] Roman Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Int. Conf. Mach. Learn., pages 160–167, 2008.

[11] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In ACM SIGGRAPH 2008 Classes, pages 1–10. 2008.

[12] Li Deng, Geoffrey Hinton, and Brian Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In IEEE Int. Conf. Acoust. Speech & Signal Process., pages 8599–8603. IEEE, 2013.

[13] Xin Deng, Yutong Zhang, Mai Xu, Shuhang Gu, and Yiping Duan. Deep coupled feedback network for joint exposure fusion and image super-resolution. IEEE Trans. Image Process., 30:3098–3112, 2021.

[14] Gabriel Eilertsen, Joel Kromander, Georgy Denes, Rafal K Mantiuk, and Jonas Unger.Hdr image reconstruction from a single exposure using deep cnns. ACM Trans. Graph., 36(6):1–15, 2017.

[15] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In IEEE Conf. Comput. Vis. Pattern Recogn., pages 2792–2790, 2016. 1, 2, 7, 8, 9

[16] Michael Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédéric Durand. Deep bilateral learning for real-time image enhancement. ACM Trans. Graph., 36(4):1–12, 2017.

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Adv. Neural Inform. Process. Syst., 27, 2014.

[18] A Arshed Goshatsky. Fusion of multi-exposure images. Image and Vision Computing, 23(6):611–618, 2005.

[19] Bo Gu, Wujing Li, Jiatao Wang, Minyun Zhu, and Minhui Wang. Gradient field multi-exposure images fusion for high dynamic range image visualization. Journal of Visual Communication and Image Representation, 23(4):604–610, 2012.

[20] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In IEEE Conf. Comput. Vis. Pattern Recogn., pages 1780–1789, 2020. 2, 5, 6, 7, 8, 9, 10

[21] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. IEEE Trans. Image Process., 26(2):982–993, 2016.

[22] Michel Gharbi, Jitendra Malik. Fusion of multi-exposure images. In IEEE Conf. Comput. Vis. Pattern Recogn., pages 0–0, 2019.

[23] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell., 35(6):1397–1409, 2012.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Int. Conf. Comput. Vis., pages 1026–1034, 2015.

[25] Yingkun Hou, Jun Xu, Mingxia Liu, Guanghui Liu, Li Liu, Fan Zhu, and Ling Shao. Nllb: A blind pixel-level non-local method for real-world image denoising. IEEE Trans. Image Process., 29:5121–5135, 2020.

[26] Xiaodan Hu, Mohamed A Naief, Alexander Wong, Mark Lamm, and Paul Fieguth. Runet: A robust unet architecture for image super-resolution. In IEEE Conf. Comput. Vis. Pattern Recogn., pages 0–0, 2019.

[27] Yuanning Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. Exposure: A white-box photo post-processing framework. ACM Trans. Graph., 37(2):1–17, 2018.

[28] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. IEEE Trans. Image Process., 30:2340–2349, 2021.

[29] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. Proc. Int. Conf. Learn. Represent., abs/1412.6980, 2015.

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

[31] Hui Li, Kede Ma, Hongwei Yang, and Lei Zhang. Fast multi-scale structural patch decomposition for multi-exposure image fusion. IEEE Trans. Image Process., 29:5805–5816, 2020.

[32] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. IEEE Trans. Image Process., 27(6):2828–2841, 2018.

[33] Shutoo Li, Xiaodong Kang, and Jianwen Hu. Image fusion with guided filtering. IEEE Trans. Image Process., 27(7):2864–2875, 2013.

[34] Jie Liang, Hui Zeng, and Lei Zhang. High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network. In IEEE Conf. Comput. Vis. Pattern Recogn., pages 9392–9400, 2021.

[35] Zhetong Liang, Jun Xu, David Zhang, Zisheng Cao, and Lei Zhang. A hybrid 11-10 layer decomposition model for tone mapping. In IEEE Conf. Comput. Vis. Pattern Recogn., June 2018.

[36] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Chia-Bin Huang. Single-image hdr reconstruction by learning to reverse the camera pipeline. In IEEE Conf. Comput. Vis. Pattern Recogn., pages 1651–1660, 2020.

[37] Minh-Thang Luong, Quoc V Le, Ilya Sutskever, Oriol Vinyals, and Lukasz Kaiser. Multi-task sequence to sequence learning. arXiv preprint arXiv:1511.06144, 2015.

[38] Kede Ma, Zhengfang Duanmu, Hanwei Zhu, Yuming Fang, and Zhou Wang. Structured multi-exposure image fusion: A coupled feedback network for joint exposure fusion and image super-resolution. IEEE Trans. Image Process., 29:2808–2819, 2019.

[39] Kede Ma, Hui Li, Hongwei Yang, Zhou Wang, Deyu Meng, and Ling Shao. Robust multi-exposure image fusion: a structural patch decomposition approach. IEEE Trans. Image Process., 26(5):2519–2532, 2017.

[40] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. IEEE Trans. Image Process., 24(11):3345–3356, 2015.

[41] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. In Computer Graphics Forum, volume 28, pages 161–171. Wiley Online Library, 2009.

[42] Jongchan Park, Joong-Young Lee, Donggeun Yoo, and In So Kweon. Distort-and-recover: Color enhancement using deep reinforcement learn-
[43] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In Adv. Neural Inform. Process. Syst., 2019. 5

[44] Sujoy Paul, Ioana S Severcenco, and Panagiotis Agathoklis. Multi-exposure and multi-focus image fusion in gradient domain. Journal of Circuits, Systems and Computers, 25(10):1650123, 2016. 2

[45] K Ram Prabahakar, V Sai Srikar, and R Venkatesh Babu. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In Int. Conf. Comput. Vis., pages 4714–4722, 2017. 2

[46] Shanmuganathan Raman and Subhasis Chaudhuri. Bilateral filter based compositing for variable exposure photography. In Eurographics, pages 1–4, 2009. 2

[47] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 234–241. Springer, 2015. 4

[48] Jianbing Shen, Ying Zhao, Shuicheng Yan, Xuelong Li, et al. Exposure fusion using boosting laplacian pyramid. IEEE Trans. Cybern., 44(9):1579–1590, 2014. 4

[49] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In Int. Conf. Comput. Vis., pages 839–846. IEEE, 1998. 2

[50] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans. Image Process., 22(9):3538–3548, 2013. 1, 2

[51] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process., 13(4):600–612, 2004. 5, 7

[52] Ziqin Wang, Jun Xu, Li Liu, Fan Zhu, and Sounthirm Chintala. Pytorch: An imperative style, high-performance deep learning library. In Adv. Neural Inform. Process. Syst., 2019. 5

[53] Anran Zhang received the B.S. degrees from Jilin University, Jilin, China in 2014, and M.E. degrees from Inner Mongolia University, Inner Mongolia, China in 2017. She is now a Ph.D student with the Key Laboratory of Advanced Technology of Near Space Information System, in the School of Electronics and Information Engineering, Beihang University, Beijing. She was a research intern with Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, UAE. Her research interests include machine learning and computer vision.

[54] Jun Xu received his B.Sc. and M.Sc. degrees from School of Mathematics Science, Nankai University, Tianjin, China, in 2011 and 2014, respectively, and Ph.D. degree from Department of Computing, Hong Kong Polytechnic University, in 2018. He worked as a Research Scientist at IIAI, Abu Dhabi, UAE. He is currently a Lecturer with School of Statistics and Data Science, Nankai University. More information can be found at https://csjunxu.github.io/.

[55] Hui Li received his Ph.D. degree in computer science from The Hong Kong Polytechnic University, Hong Kong, in 2019. Currently, he is working as a Senior Imaging Algorithm Engineer in vivo Mobile Communication Co. Ltd., Shenzhen, China. Prior to that, he worked as a Senior Engineer in Huawei 2012 Labs, Shenzhen, China. His research interests include image processing, computational photography, and computer vision.

[56] Xiayong Zhan received the B.S. and M.E. degrees from Lanzhou University, Lanzhou, China in 2007 and 2010, respectively, and the Ph.D. degree from the Department of Electronic and Electrical Engineering, The University of Sheffield, U.K., in 2013. He is currently with University of Amsterdam and Inception Institute of Artificial Intelligence (IIAI). His research interests include Machine Learning and Computer Vision.

[57] Jun Xu, Lei Zhang, Wangmeng Zuo, David Zhang, and Xiangchu Feng. Patch group based nonlocal self-similarity prior learning for image denoising. In Int. Conf. Comput. Vis., 2015. 2

[58] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. Proc. Int. Conf. Learn. Represent., page 1–13, 2016. 4

[59] Jun Xu, Yingkun Hou, Dongwei Ren, Li Liu, Fan Zhu, Mengyang Yu, Lei Zhang, and David Zhang. A trilateral weighted sparse coding scheme for real-world image denoising. In Eur. Conf. Comput. Vis., September 2018. 3

[60] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. In Int. Conf. Comput. Vis., Oct 2017. 2