Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images

Jianrui Cai, Shuhang Gu, and Lei Zhang, *Fellow, IEEE*

***Abstract*— Due to the poor lighting condition and limited dynamic range of digital imaging devices, the recorded images are often under-/over-exposed and with low contrast. Most of previous single image contrast enhancement (SICE) methods adjust the tone curve to correct the contrast of an input image. Those methods, however, often fail in revealing image details because of the limited information in a single image. On the other hand, the SICE task can be better accomplished if we can learn extra information from appropriately collected training data. In this paper, we propose to use the convolutional neural network (CNN) to train a SICE enhancer. One key issue is how to construct a training data set of low-contrast and high-contrast image pairs for end-to-end CNN learning. To this end, we build a large-scale multi-exposure image data set, which contains 589 elaborately selected high-resolution multi-exposure sequences with 4,413 images. Thirteen representative multi-exposure image fusion and stack-based high dynamic range imaging algorithms are employed to generate the contrast enhanced images for each sequence, and subjective experiments are conducted to screen the best quality one as the reference image of each scene. With the constructed data set, a CNN can be easily trained as the SICE enhancer to improve the contrast of an under-/over-exposure image. Experimental results demonstrate the advantages of our method over existing SICE methods with a significant margin.**

***Index Terms*— Single image contrast enhancement, multi-exposure image fusion, convolutional neural network.**

I. Introduction

REPRODUCING the natural scene with good contrast, vivid color and rich details is an essential goal of digital photography. The acquired images, however, are often under-exposed or over-exposed because of poor lighting conditions and the limited dynamic range of imaging device. The resulting low contrast and low quality images will not only degenerate the performance of many computer vision and image analysis algorithms, but also degrade the visual aesthetics of images [1]. Contrast enhancement is thus an important step to improve the quality of recorded images and make the image details more visible.

Manuscript received July 11, 2017; revised November 9, 2017 and December 28, 2017; accepted January 5, 2018. Date of publication January 15, 2018; date of current version February 6, 2018. This work was supported by Hong Kong RGC GRF under Grant PolyU 5313/13E. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiaochun Cao. *(Corresponding author: Lei Zhang.)*

The authors are with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong (e-mail: csjcai@comp.polyu.edu.hk; cssgu@comp.polyu.edu.hk; cslzhang@comp.polyu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2794218

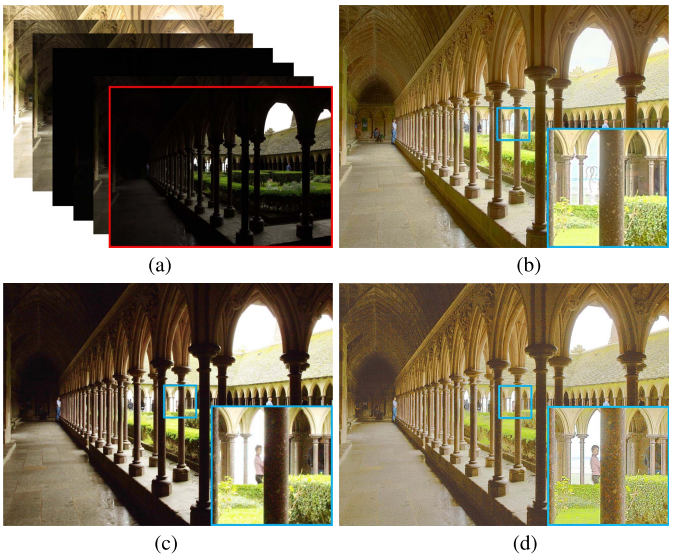

Fig. 1. (a) A sequence with under- and over-exposed images. (b) The enhanced image by a deghosting MEF method [2]. (c) The enhanced image by a SICE method [3]. (d) The enhanced image by our method. (Note that (c) and (d) are the enhanced results by a single under-exposed image.)

Traditional single image contrast enhancement (SICE) techniques include those histogram-based algorithms [4]–[6], which increase the contrast of an image by redistributing the luminous intensity on histogram, and Retinex based algorithms [7]–[9], which enhance the reflectance and illumination components of the image separately. These methods, however, are difficult to reproduce a high-quality image due to the complex natural scenes and the limited information in a single low-contrast image. Thanks to the development of imaging devices, we are able to capture a sequence of multi-exposure images in a short time to fulfil the dynamic range of a scene [10], [11]. With the sequence, multi-exposure image fusion (MEF) [2], [12], [13] and stack-based high dynamic range (HDR) imaging (with a following tone mapping operator) methods [14], [15] can be applied to blend the multiple images with different exposures into a perceptually more appealing image.

Generally speaking, MEF and stack-based HDR methods will produce images with better visual quality than those SICE methods since more information is available in the multi-exposure sequence. However, the acquisition of multi-exposure images will complicate the imaging process, and camera shake or moving objects will lead to unpleasant fusion artifacts such as the ghosting artifacts [16], [17]. Figure 1(a)

shows a sequence of under/over-exposure images. Figure 1(b) shows the output image by a state-of-the-art deghosting MEF algorithm [2], which merges the multi-exposure images into a high-visibility image. Figure 1(c) shows the result by a state-of-the-art SICE method [3], which only takes an under-exposure image as input (the image with red box shown at Figure 1(a)). One can see that the MEF method could recover more image details, which cannot be revealed by the SICE method; however, it generates some ghosting artifacts due to the displacement of different frames caused by the object motion (such as human movement and ripples). In contrast, the SICE method will not have such ghosting artifacts because it takes only one single exposure image as input. Due to the above reasons, SICE is more attractive and easier to implement in practice, yet it is much more challenging because of the limited information in a single image.

Considering that multi-image based MEF and single-image based SICE methods have their pros and cons, one interesting question is: can we develop a SICE method which can approximate the contrast enhancement performance of MEF methods while being free of the ghosting artifacts? In this work, we make the first attempt to address this challenging problem. Our idea is inspired by the success of discriminative learning methods [18]–[20], especially the deep convolutional neural network (CNN) methods [21], [22], in image restoration. Compared with generative models which use high-quality images to learn image priors, discriminative methods utilize a set of degraded and ground-truth image pairs to learn a model to enhance the given degraded image. As a powerful discriminative learning method, CNN has been successfully used in many low-level vision problems such as single image super-resolution [23] and image denoising [24], where a large amount of paired training samples can be generated or simulated. For example, one can down-sample a high-resolution image to generate a corresponding low-resolution image, and add noise to a clean image to generate a noisy observation of it. In the application of contrast enhancement, unfortunately, it is very hard to generate such paired images due to the lack of a simple model to approximate the low-contrast image generation process. To the best of our knowledge, by far there is no dataset of paired low-contrast and high-contrast images available for training a discriminative model for SICE.

One significant contribution of this work is that we build such a dataset of low-contrast and good-contrast image pairs, which makes the discriminative learning of SICE enhancers possible. The key idea is that we use state-of-the-art MEF and stack-based HDR methods to reconstruct the reference good-contrast image of a scene, while those under-exposure or over-exposure images of the scene can be naturally taken as the low-contrast counterparts. To build this dataset, we collect multi-exposure sequences from 2 categories of scenes (indoor and outdoor), and employ 13 recently developed MEF and HDR algorithms to generate the high-contrast images for each scene. Then, subjective experiments are conducted to select the best MEF/HDR result for each scene, and exclude those sequences which do not have satisfactory outputs (e.g, ghosting artifacts). Finally, the multi-exposure sequences of 589 scenes and their corresponding high-quality reference images are selected in the dataset. Each sequence has 3 to 18 low-contrast images with different exposure levels, and there are 4,413 low-contrast images in total. With the constructed dataset, we design a simple yet effective CNN to learn a SICE enhancer, which is able to automatically enhance the low-contrast images with different exposure levels. The learned CNN based SICE enhancer demonstrates clear advantages over existing SICE methods, outperforming them by a large margin. The contributions of our work are summarized as follows:

1) We build, for the first time to the best of our knowledge, a large-scale multi-exposure image dataset which contains low-contrast images with different exposure levels and their corresponding high-quality reference image. The constructed dataset makes end-to-end discriminative learning of high performance SICE methods possible. It also provides a platform to quantitatively evaluate, at least to some extent, the performance of different contrast enhancement algorithms.
2) With the constructed dataset, a well designed CNN is trained for SICE, which demonstrates significant advantages over existing SICE methods. Our work provides a new solution to high performance SICE.

II. Related Work

A. Single Image Contrast Enhancement

Single image contrast enhancement (SICE) aims to improve the visibility of the scene in a given single low-contrast image. It provides a way to enhance the low contrast photographs captured from a high dynamic range scene [25]. Many histogram and Retinex based SICE methods have been proposed in the past decades. Histogram-based methods [4], [5] have been widely used because of their simplicity in enhancing low-contrast images. Those methods attempt to redistribute the luminous intensity on histogram in a global or local manner. However, such simple redistribution operations may produce serious unrealistic effects in the enhanced images since they ignore image structural information [26]. To excavate the structural information from the low-contrast image, Retinex-based methods [7], [27] decompose the input image into albedo and illumination layers, and adopt different strategies to enhance the reflectance and illumination components. Most of the previous SICE methods are based on some assumptions on high-quality images, while they may not fully exploit the information in the input image. On the other hand, the enhancement capability of existing SICE methods is rather limited due to the limited information in a single low-contrast image [9]. Recently, methods [28], [31] have been proposed to train a CNN network to map the low dynamic range (LDR) images to HDR images. In [29], a CNN is trained to set the parameters of bilateral filters, which are then used to enhance an input image to a desired image edited by professional photographers. Since extra information can be learned from the external dataset, in this work, we will elaborately build a dataset to learn a powerful CNN-based SICE enhancer from multi-exposure images.

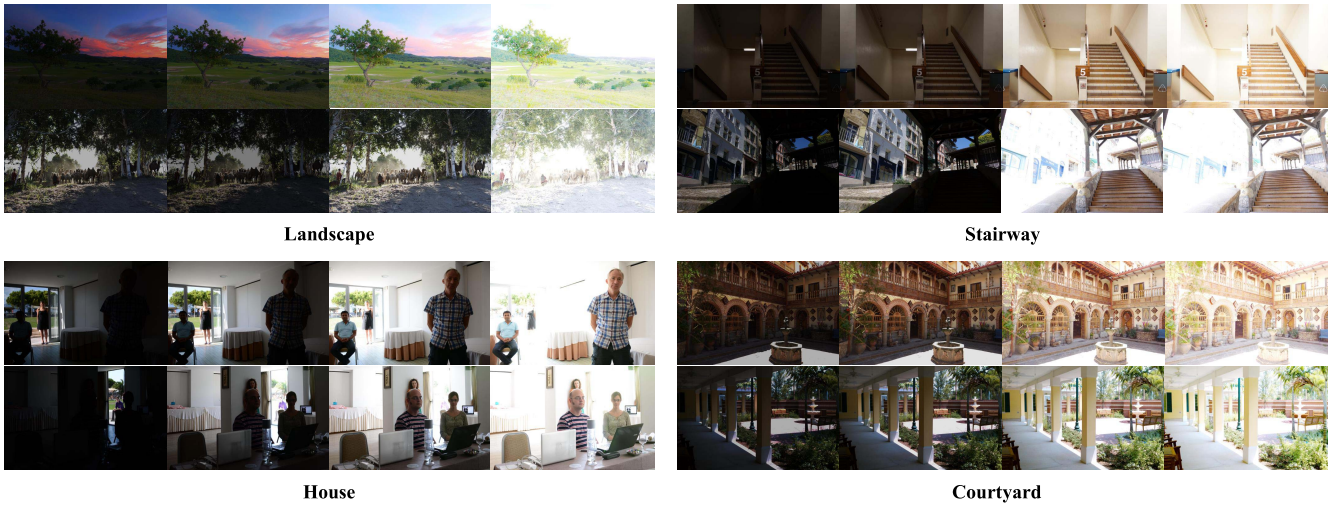

Fig. 2. Sample source image sequences with different exposure levels in our dataset.

B. MEF and Stack-Based HDR

Because of the limited dynamic range, traditional digital imaging systems may lose structural details when shooting a natural scene [30]. To address this issue, stack-based HDR methods [14], [15] propose to merge bracketed multiple exposure images into an HDR irradiance map, then employ a tone mapping operator to compress the dynamic range of HDR irradiance map so that the high-contrast image can be displayed on regular monitors. Different from the HDR approaches, MEF methods [2], [13] attempt to fuse the images directly in the non-liner brightness domain to reproduce a high-visibility image. Despite their successes on well-aligned image sequences, the existence of camera shake and object motion in many scenes often leads to ghosting artifacts in the final enhanced results, limiting the applications of MEF and HDR in practice. In the last decade, researchers have spent much efforts to design de-ghosting algorithms [2], [14], [32] and learning-based methods like [33] proposed to map the multi-exposure image sequences to an HDR image. However, it remains a challenging problem in MEF and HDR for dynamic scenes. In this work, we elaborately select the well-aligned image sequences to generate a good reference images by MEF and HDR reconstruction methods.

C. CNN for Image Restoration

CNN has demonstrated its effectiveness in image restoration and enhancement applications such as denoising [34], [35], super-resolution [21] and deblurring [36]. In those applications, pairs of degraded images and their high-quality counterparts can be easily generated. With those paired training data, CNN can be used to learn a mapping function between the degraded observations and their corresponding high-quality reference images. However, for the application of SICE, such rule-based, computer-generated training datasets are too ideal to be true for real-world low-contrast images, where the distribution of luminance is much more complex and varies with different scenes, cameras and camera settings. The lack of training data has hindered the adoption of CNN methods for end-to-end SICE enhancer learning. In this work, an elaborately designed dataset of low-/good-contrast image pairs is built, with which CNN can be easily adopted to learn powerful SICE enhancers.

III. Multi-Exposure Image Dataset and Reference Image Generation

As discussed in the previous sections, the lack of paired training data impedes the application of CNNs to SICE tasks. In order to make end-to-end learning of SICE enhancers possible, in this section we construct a dataset of multi-exposure image sequences as well as the reference good-contrast image for each sequence.

A. Objectives

There are two major objectives for our multi-exposure image dataset construction. First, the dataset should contain enough high resolution multi-exposure image sequences and cover a diversity of scenes. Second, for each sequence, a high-quality reference image should be generated so that image pairs can be constructed for end-to-end learning.

Some multi-exposure image sequences are available in literature [37], [38] and most of them were captured for the study of MEF and stack-based HDR methods. However, the total amount of such publicly available sequences is very limited, and many of them were taken under indoor environment. Neither the number of sequences nor the diversity of sequence exposure levels meets the requirement of real-world applications. To achieve the first objective, we collect a large number of sequences from both indoor and outdoor scenes, and make sure that the photographs in our dataset cover a broad range of scenes, subjects and lighting conditions. Some sample sequences of our multi-exposure image dataset are presented in Figure 2. Some sequence are from [37][1], while the others are collected by us.

The second objective is more challenging. Considering that MEF and stack-based HDR methods can reproduce an image with much higher contrast and visibility than a single exposure

[1] http://rit-mcsl.org/fairchild/HDR.html

image, we adopt the latest and state-of-the-art MEF and HDR techniques to construct the reference image. However, there is not an MEF or HDR algorithm which outperforms all the other methods for different scenarios, and most of those methods will generate certain ghosting artifacts for dynamic sequences with moving objects under uncontrolled environment. To address this issue, we adopt 13 state-of-the-art MEF and HDR algorithms to generate the high-contrast reference images of all scenes, and conduct subjective experiments to select the best output for each sequence. In addition, sequences which cannot produce satisfactory outputs will be excluded from the dataset. In this way, the reference images generated in our dataset will have higher quality than any individual existing MEF or HDR method.

B. Multi-Exposure Image Collection

To achieve the objectives mentioned above, we collect and select multi-exposure image sequences of relatively static scenes. The details of data collection and screening are described as follows.

1) Data Collection: To ensure that a robust and general SICE enhancer can be trained, the training data should be collected from representative real-world scenarios with commonly used imaging devices. In our dataset, the image sequences are taken by different cameras and from different scenes. Seven types of consumer grade cameras are used to collect the image sequences, including Sony α7RII, Sony NEX-5N, Canon EOS-5D Mark II, Canon EOS-750D, Nikon D810, Nikon D7100 and iPhone 6s.

Exposures of indoor and outdoor scenes will be very different even with the same Exposure Value (EV) setting. Since our goal is to learn a SICE enhancer for automatically correcting the exposure of a low-contrast image, the images we collected should cover most of the exposure levels we would see in our daily life. In this work, we collect image sequences from both indoor and outdoor scenes under certain EV settings. For the indoor scenes, we are able to set up a static environment and use a tripod to capture well-aligned image sequences. We collect 7 to 18 images for each indoor scene. The exposure levels are manually set based on the lighting ratio of the scene. For the uncontrolled outdoor environment, moving objects (e.g., cars, walking people, shaking trees) make the acquisition of well-aligned sequences very challenging. We use the continuous bracket mode to automatically shoot image sequences with shifted exposures. To ensure that the sequence can be well-aligned, for each outdoor scene, multiple sequences (with EV shifted by $\pm$ $\{0.5, 0.7, 1.0, 2.0, 3.0\}$) are collected, and each sequence contains 3 to 5 images. After collecting the source images, a further screening process is conducted to select desirable sequences for reference image generation.

2) Data Screening: In the data collection stage, we collected more than 10,000 image sequences with different exposure levels (including repeated sequences). However, many of them contain distorted images or moving objects. We therefore conduct an uphill screening process to refine the dataset. Sequences with significantly distorted images (e.g., motion blur, out of focus and visible sensor noise) or obvious moving objects are discarded (see Figure 3 for some examples). As for those repeated sequences, we manually choose the most well-aligned one. In the screening stage, more than 85% of the collected sequences are weeded out, and about 1,200 candidate sequences are remaining in the dataset. We then apply state-of-the-arts MEF and stack-based HDR algorithms to those sequences to generate reference images, and further screen the sequences based on the quality of reference images.

Fig. 3. Sample source image sequences excluded from the dataset. Blue and red arrows point out the moving objects in different frames, which will cause "ghosting" artifacts in the reference image generated by MEF or stack-based HDR algorithms.

C. Reference Image Generation

Having the candidate sequences, we propose to generate high-quality reference images with MEF and stack-based HDR methods. 13 state-of-the-art MEF and HDR algorithms are employed in this process, including 8 MEF methods: Mertens09 [13], Raman09 [39], Shen11 [40], Zhang12 [41], Li13 [12], Shen14 [42], Ma17 [2], Kou17 [43], and 5 stack-based HDR methods: Sen12 [14], Hu13 [32], Bruce14 [44], Oh15 [15], Photomatix [45]. The implementations of those algorithms are obtained from the original authors, except for Raman09 and Bruce14 whose implementations are from an HDR toolbox in Github [46]. To generate the faithful results of original scenes and for a fair comparison, we manually tune the tone-mapping operators for Sen12, Hu13 and Oh15 by Photomatix. As a result, there is one image for each scene by each method for subjective evaluation. Note that we also use Photomatix to generate the HDR irradiance map for Hu13, which outputs a stack of well-aligned low dynamic range images. As for Ranman09 and Bruce14, we adopt the tone-mapping operators as presented in the original papers.

With the 1,200 sequences and 13 MEF/HDR algorithms, we generate $1{,}200 \times 13 = 15{,}600$ fusion results. We then invite 13 amateur photographers and 5 volunteers who do not have much photographing experience to perform a pairwise comparison among the 13 MEF/HDR results of each sequence. All of the 18 volunteers perform subjective evaluation under the same environment with a 4K Ultra HD LED Monitor. Besides, all volunteers do not have a bias on this task and they were given instructions before the experiments. As shown in Figure 4, a customized interface is adopted to render a pair of images simultaneously at their original resolutions but in random order. The subject can move the mouse on the image so that the zoom-in windows of the local region can be shown for better comparison. For each pair of images, the subject

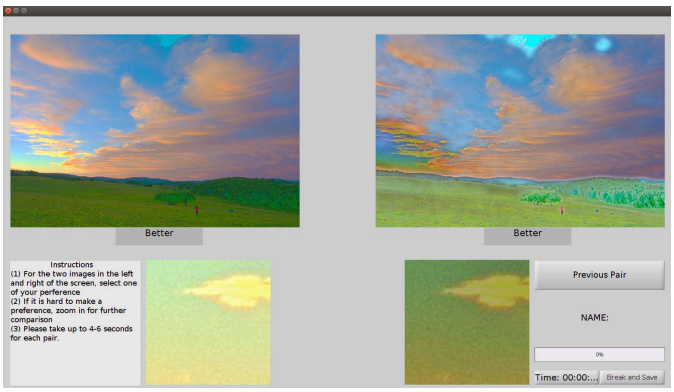

Fig. 4. User interface for subjective testing.

only need to indicate his/her preference to either the left image or the right image. Therefore, for each subject the best result of one scene will be selected after 12 pair comparisons. By majority voting, the 1,200 best reference images for the 1,200 sequences are selected after the subjective test.

However, for some challenging sequences which contain unaligned contents, even the best MEF/HDR results can have unsatisfactory quality. We then abandon those images from our dataset. Eventually, only 589 high-quality reference images and their corresponding sequences are remained in our dataset. Figure 5 shows the percentage of selected images for each MEF or HDR algorithm, as well as some examples of the reference image selection. As one popular image quality assessment (IQA) metric for MEF images, MEF-SSIM [16] is employed here to quantitatively compare the 13 selected MEF/HDR algorithms. Note that since the MEF-SSIM metric is designed to evaluate MEF algorithms on static scenes, it is not applicable to the evaluation of MEF/HDR methods on dynamic scenes. In addition, some of the MEF/HDR methods (such as [12], [13], [39], and [43]) are designed to process static scene sequences. When applied to dynamic scene sequences (even with small misaligned), these methods often generate ghosting artifacts in the final fusion results. Therefore, we only adopt MEF-SSIM to evaluate the MEF/HDR algorithms on 100 indoor static image sequences. The results are listed in Table I. One can see that the quality metric MEF-SSIM is in accordance with the proposed subjective testing to some extent. For example, [43] is one of the mostly selected MEF methods, while it also gets the highest score of quality metric [16].

In summary, our dataset includes 589 sequences from indoor and outdoor scenes, containing a total number of 4,413 multi-exposure images. Among them, 56 sequences are obtained from existing literature [37], and the remaining are collected by ourselves. The resolution of most images are between 3000×2000 and 6000×4000. To our best knowledge, it is the largest multi-exposure image dataset so far (dataset available at https://github.com/csjcai/SICE). Furthermore, the results of 13 MEF/HDR algorithms on each sequence, as well as the selected reference images, are also provided in the dataset. Our dataset has various potential applications, for example, MEF and HDR algorithm evaluation, image quality assessment metric design, and SICE enhancer training. Particularly, in this paper we utilize the constructed dataset to train a CNN based powerful SICE enhancer, as introduced in the next section.

TABLE I

THE QUALITY METRIC MEF-SSIM [16] OF THIRTEEN MEF AND HDR ALGORITHMS ON 100 INDOOR IMAGE SEQUENCES. THE QUALITY VALUE RANGES [0, 1] WITH A HIGHER VALUE INDICATING BETTER PERCEPTUAL QUALITY

Method	Average Score	Rank
Raman09 [39]	0.817	13
Shen14 [42]	0.839	12
Li13 [12]	0.925	5
Bruce14 [44]	0.873	10
Shen11 [40]	0.907	9
Zhang12 [41]	0.871	11
Mertens09 [13]	0.931	3
Photomatix [45]	0.919	7
Ma17 [2]	0.934	2
Kou17 [43]	0.937	1
Sen12 [14]	0.927	4
Oh15 [15]	0.921	6
Hu13 [32]	0.919	7

IV. CNN-BASED SICE LEARNING

With the constructed dataset, we can design a CNN based SICE enhancer to learn a mapping function between the low-contrast input image $\boldsymbol{I}(x, y) \in \Re^3$ and its corresponding reference image $\boldsymbol{I}_{ref}(x, y) \in \Re^3$. Intuitively, one can directly train a deep CNN $H(\boldsymbol{I}, W)$ with parameters W to achieve this goal, and Figure 6 shows such a network architecture with 15 layers. We train the direct network with Mean Squared Error (MSE) loss, ℓ_1-norm loss and Structural dissimilarity (DSSIM) [47] loss, respectively. The MSE loss function to be minimized is:

$$l_2(W) = \frac{1}{n}\sum_{i}^{n} \|\boldsymbol{I}_{ref}^{(i)} - H(\boldsymbol{I}^{(i)}, W)\|_F^2 \tag{1}$$

ℓ_1-norm loss function can be formulated as:

$$l_1(W) = \frac{1}{n}\sum_{i}^{n} \|\boldsymbol{I}_{ref}^{(i)} - H(\boldsymbol{I}^{(i)}, W)\|_1 \tag{2}$$

The DSSIM loss function, which is derived from structural similarity (SSIM) [48], can be formulated as:

$$DSSIM(W) = \frac{1}{n}\sum_{i}^{n}(1 - ssim(\boldsymbol{I}_{ref}^{(i)} - H(\boldsymbol{I}^{(i)}, W)))/2 \tag{3}$$

However, we experimentally found that the result obtained by directly training such a network in original image domain is not very satisfactory. In Figures 7(a) - 7(e), we show an example of the (cropped) original low-contrast image, the reference image (generated by MEF method [43]) and the enhanced result by the CNN with MSE loss, DSSIM loss and ℓ_1-norm loss, respectively. One can see that the enhancement results exhibit some color shift. This is probably because one stage CNN in original intensity may have difficulties in balancing the enhancement of smooth and texture components

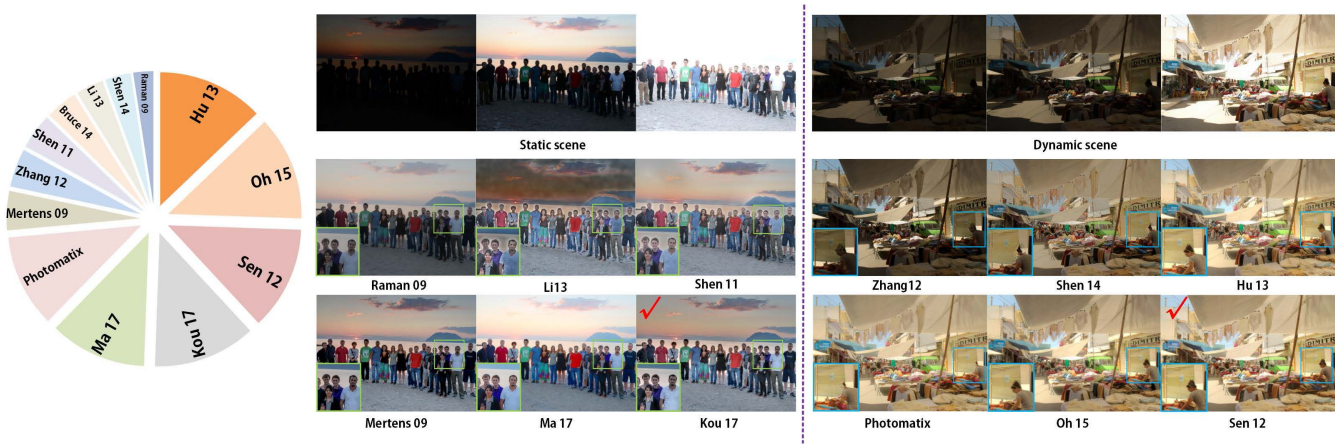

Fig. 5. Reference image generation. **Left:** Pie chart illuminates the percentage of images generated by different MEF and stack-based HDR algorithms. **Right:** Sample fusion results by different MEF and HDR algorithms. The one with red tick has the highest score in the subjective quality discrimination test, which is then selected as the reference image for that scene in our dataset.

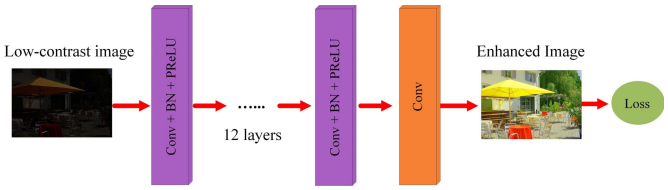

Fig. 6. The 15 layers direct CNN network architecture.

of an image. In Section V, we would take those 3 direct networks with ℓ_1, MSE and DSSIM loss function as the baseline methods, and provide both the visual and quantitative results of them for further comparisons.

According to the Retinex theory [49], the low-frequency information of an image represents the global naturalness, and the high-frequency information represents the local details. It is a common practice that we can separate the image low-frequency component and high-frequency component and process them individually. Several MEF/HDR algorithms, such as [50]–[53], have been proposed to decompose an image into high and low frequency components to preserve image details and colors in the brightest/darkest regions. In order to train a CNN which can not only enhance the luminance range of the low-contrast image but also reveal some missing details, it is important for the network to make an appropriate balance between high and low frequency components. Inspired by those previous works [50]–[53], we first decompose the low-contrast image and the reference image into a low-frequency luminance component $\boldsymbol{L}(x, y) \in \Re^3$ and a high-frequency detail component $\boldsymbol{R}(x, y) \in \Re^3$:

$$\boldsymbol{I}(x, y) = \boldsymbol{L}(x, y) + \boldsymbol{R}(x, y) \tag{4}$$

The decomposition is performed by applying the weighted least squares (WLS) method [54] to each channel.

After decomposition, a two-stage CNN scheme can be developed to learn the SICE enhancer. In the first stage, a luminance enhancer and a detail enhancer are trained in parallel to enhance the two components with different loss functions. In the second stage, the two enhanced components are merged as the input, and another CNN is trained to enhance the whole image to the desired reference. Figure 8 illustrates the training procedure of the proposed method.

A. Network Overview

The proposed CNN has 5 types of layers which are shown in Figure 8 with 5 different colors. i) Conv+PReLU: 64 filters of size 3×3, 5×5 and 9×9 with strides 1 and 2 are used to generate 64 feature maps, and PReLU (parametric rectified linear unit) [55] is utilized for the nonlinearity. ii) Deconv+PReLU: 64 filters of size 9×9, 5×5 and 3×3 with strides 2 and 1 are used to generate 64 feature maps, and PReLU is utilized as the activation function. iii) Conv+BN+PReLU: 64 filters of size 3×3 are used, and batch normalization [56] is added between convolution and PReLU. iv) Conv: 3 filters of size 1×1 are used to reconstruct the output. v) Skip connection: the add operation is used to connect the feature maps of two layers.

1) Stride Convolution and Deconvolution: The convolutional operations will reduce the size of feature maps. To ensure that the output image will have the same size as the input one, methods have been proposed to pad zeros before convolutions [21]. However, for the luminance enhancement network, we experimentally found that padding zeros would lead to artifacts around the boundary of the output image. Therefore, instead of padding zeros, we apply deconvolutions to keep the size of the output unchanged. The convolutional and deconvolutional strategy not only avoid artifacts in the boundary area, but also reduce computational burden with stride filters.

2) Parametric Rectified Linear Unit: In many CNN-based image restoration methods [21], [23], rectified linear unit (ReLU) is adopted as the activation function. However, since both the positive and negative coefficients contain important local structural information of the input image, simply setting the negative responses to zeros may not be a good choice. In this paper, we adopt the PReLU as the activation function, which could improve model fitting with nearly

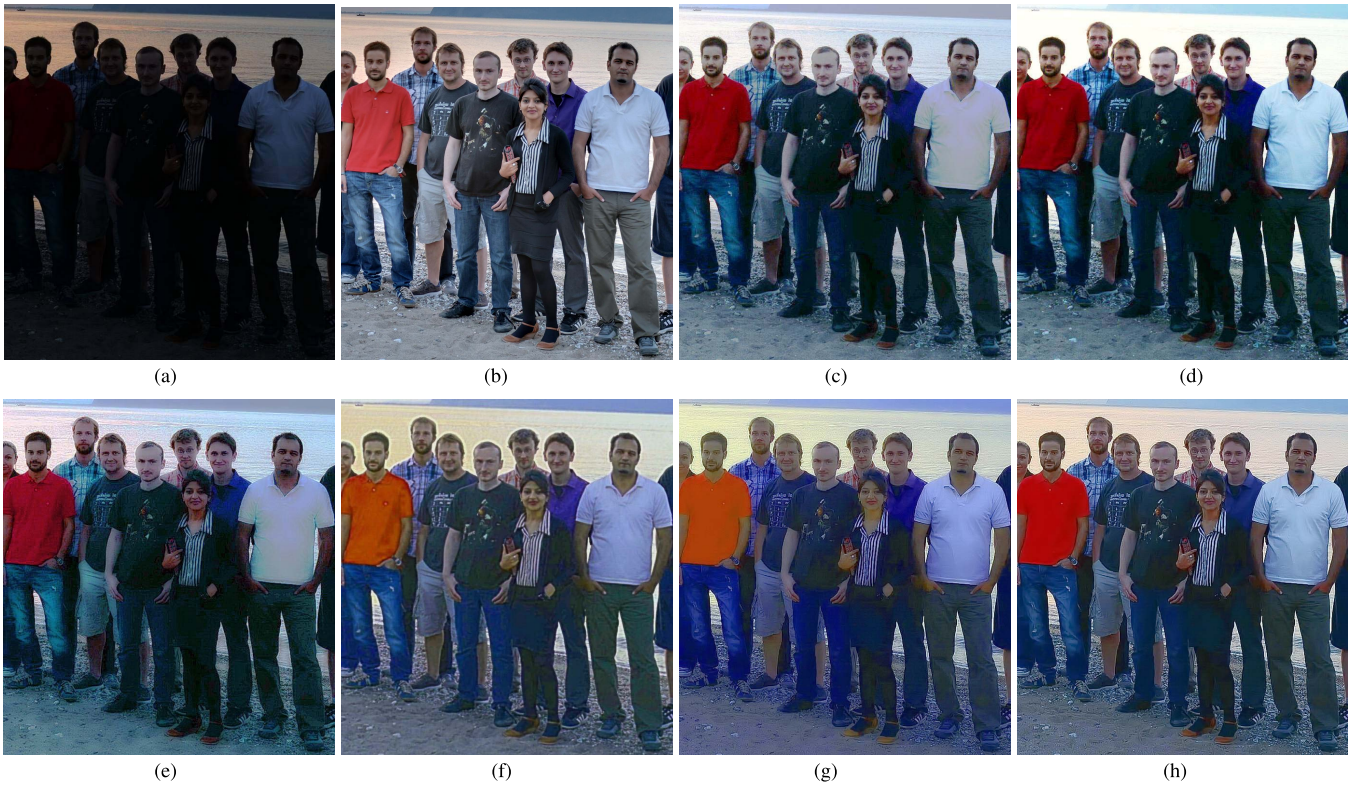

Fig. 7. The enhancement results of different networks. (a) Original. (b) Reference. (c) Direct network with MSE loss. (d) Direct network with DSSIM loss. (e) Direct network with ℓ_1 loss. (f) First stage of our network. (g) Jointly fine-tuning of first stage (DSSIM loss). (h) Second stage of our network.

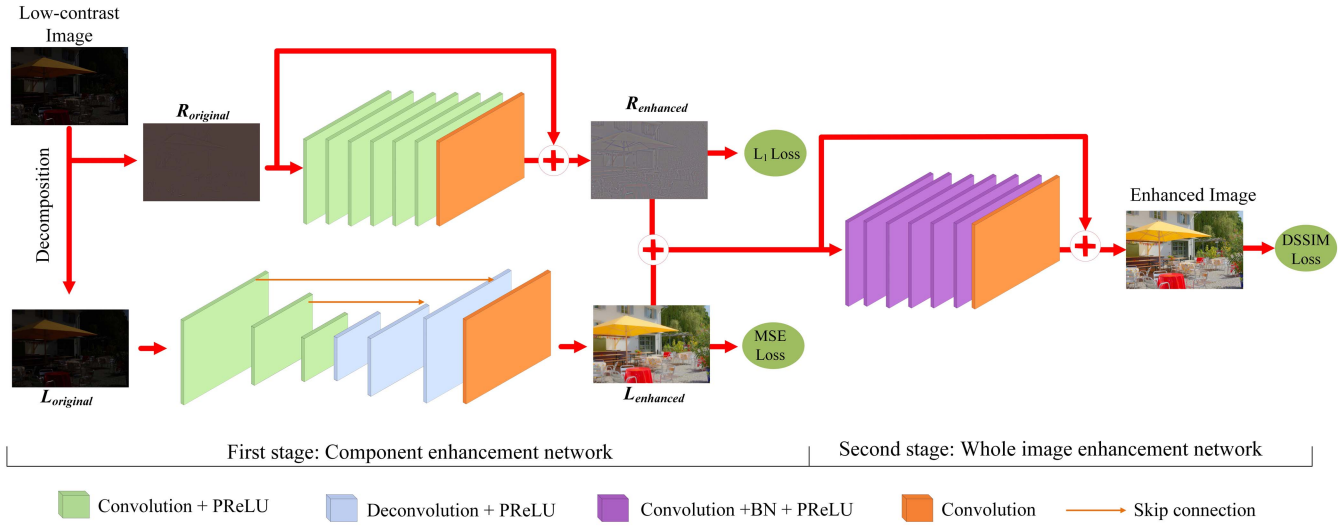

Fig. 8. The proposed CNN network architecture.

zero extra computational cost and little over-fitting risk [55]. Without ignoring negative coefficients, PReLU is able to generate high quality estimation with less filters. In Figure 9, we provide some visual examples to compare the two activation functions.

B. Component Enhancement Network

1) Luminance Enhancement Network: In order to increase the contrast of luminance map $\boldsymbol{L}$ and enforce spatial smoothness on it, we train a luminance enhancer to learn a mapping function between the luminance component $\boldsymbol{L}_{original}$ of the input low-contrast image and the luminance component $\boldsymbol{L}_{ref}$ of the reference image. Since the luminance component represents the global naturalness, the local receptive fields of the network is set larger to connect with more pixels in the original image. In order to increase the receptive fields while preventing the loss of information caused by the stride convolution operation, we adopt U-net [57] as the luminance

(a) (b) (c) (d)

Fig. 9. The enhancement results by using ReLU and PReLU. (a) Original. (b) PReLU with 64 filters. (c) ReLU with 64 filters. (d) ReLU with 128 filters.

TABLE II
LUMINANCE ENHANCEMENT NETWORK ARCHITECTURE

Layer	Activation size
Input	$129 \times 129 \times 3$
$9 \times 9 \times 64$ conv, stride 2	$61 \times 61 \times 64$
$5 \times 5 \times 64$ conv, stride 2	$29 \times 29 \times 64$
$3 \times 3 \times 64$ conv, stride 1	$27 \times 27 \times 64$
$3 \times 3 \times 64$ deconv, stride 1	$29 \times 29 \times 64$
$5 \times 5 \times 64$ deconv, stride 2	$61 \times 61 \times 64$
Skip connection	$61 \times 61 \times 64$
$9 \times 9 \times 3$ deconv, stride 2	$129 \times 129 \times 3$
Skip connection	$129 \times 129 \times 3$
$1 \times 1 \times 3$ conv, stride 1	$129 \times 129 \times 3$

TABLE III
DETAIL ENHANCEMENT NETWORK ARCHITECTURE

Layer	Activation size
Input	$129 \times 129 \times 3$
$3 \times 3 \times 64$ conv, stride 1, pad 1	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1	$129 \times 129 \times 64$
$1 \times 1 \times 3$ conv, stride 1	$129 \times 129 \times 3$
Residual sum	$129 \times 129 \times 3$

TABLE IV
WHOLE IMAGE ENHANCEMENT NETWORK ARCHITECTURE

Layer	Activation size
Input	$129 \times 129 \times 3$
$3 \times 3 \times 64$ conv, stride 1, pad 1, BN	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1, BN	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1, BN	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1, BN	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1, BN	$129 \times 129 \times 64$
$3 \times 3 \times 64$ conv, stride 1, pad 1, BN	$129 \times 129 \times 64$
$1 \times 1 \times 3$ conv, stride 1	$129 \times 129 \times 3$
Residual sum	$129 \times 129 \times 3$

enhancement network. The details of the proposed luminance enhancement network are summarized in Table II.

The MSE is adopted as the loss function, where $F_L(\cdot, \Theta)$ denotes the luminance CNN mapping function with parameters Θ:

$$l(\Theta) = \frac{1}{n}\sum_{i}^{n} \|\boldsymbol{L}_{ref}^{(i)} - F_L(\boldsymbol{L}_{original}^{(i)}, \Theta)\|_F^2 \tag{5}$$

2) Detail Enhancement Network: A CNN mapping function F_R with parameters Ω is trained to enhance the detail component $\boldsymbol{R}_{original}$ of the input low-contrast image to the detail component $\boldsymbol{R}_{ref}$ of the reference image. Inspired by [21], [35], and [58], we adopt residual learning strategy for our detail enhancement network. According to [58], this structure guarantees that the input information can be propagated through all parameter layers, which helps to train the network. And we experimentally found that residual learning can not only result in faster but also more stable convergence than the direct mapping network. The details of the proposed detail enhancement network are summarized in Table III.

Considering that the high-frequency detail component will generally follow Laplacian distribution and contain some noise and outliers, we use the ℓ_1-norm loss function:

$$l(\Omega) = \frac{1}{n}\sum_{i}^{n} |\boldsymbol{R}_{ref}^{(i)} - F_R(\boldsymbol{R}_{original}^{(i)}, \Omega)|_1 \tag{6}$$

C. Whole Image Enhancement Network

By using the luminance and detail enhancement CNNs, we are able to enhance the luminance range and recover more details of the original low-contrast image. However, we cannot ensure the overall visual quality of the enhanced image because the two CNNs are trained separately on luminance and detail components. Moreover, since the source image contain both dark region and brightness region, using only the component network is not powerful enough to model the mapping function from low-contrast image to high quality image, which may cause color shifts in the final image. Therefore, we merge the two enhanced components into one image, and introduce another CNN to further refine it to the desired reference image. The whole image enhancement network architecture is same as the detail enhancement network, except that the batch normalization (BN) operation is used here. The details of the proposed whole image enhancement network are summarized in Table IV.

To promote the perceptual quality of the final outputs, the perceptually-motivated DSSIM measure is employed as

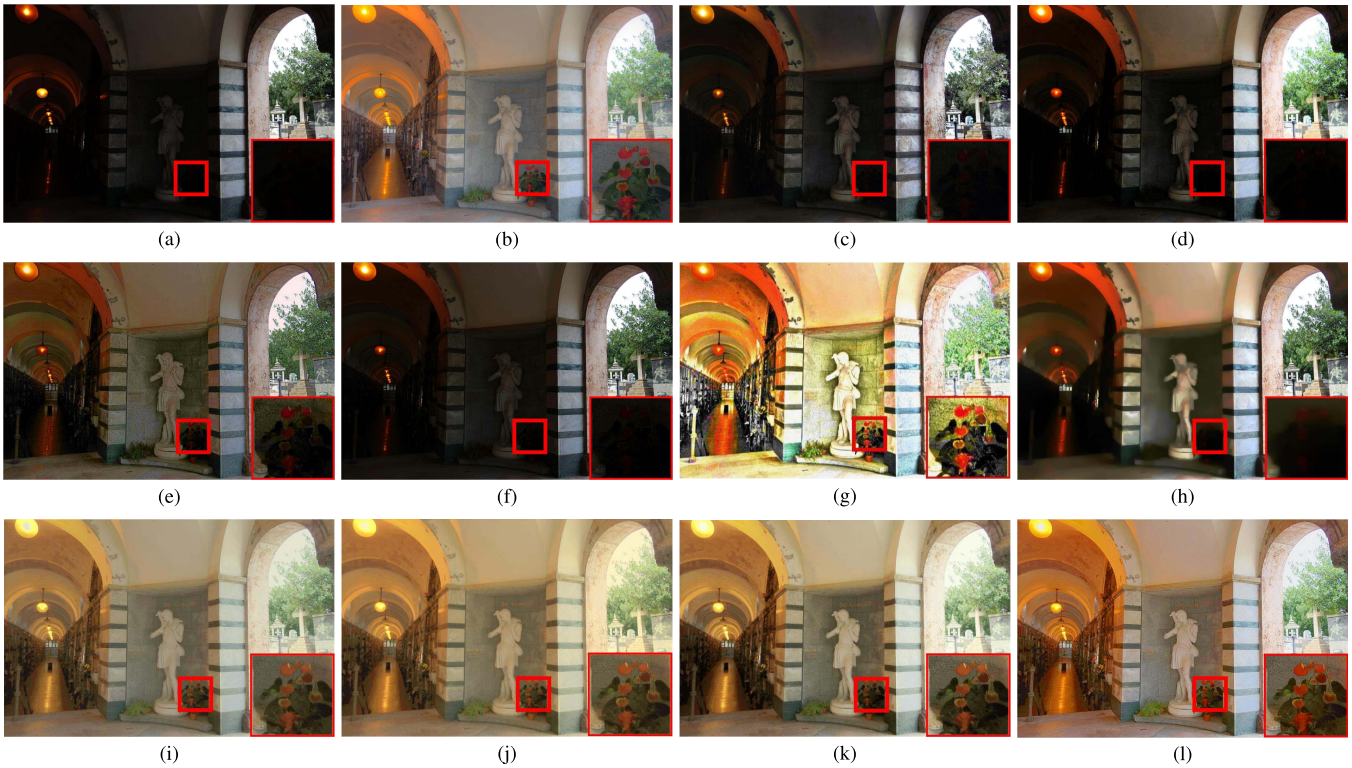

Fig. 10. Single image contrast enhancement results on an under-exposed image by different methods (PSNR/FSIM). (a) Original. (b) Reference. (c) CVC [5] (13.01/0.8775). (d) AGCWD [6] (13.55/0.8863). (e) NEP [8] (16.44/0.8913). (f) SRIE [3] (14.88/0.8902). (g) LIME [1] (13.66/0.8372). (h) Li [59] (15.13/0.8897). (i) DN (MSE) (19.58/0.9140). (j) DN (ℓ_1) (18.94/0.9037). (k) DN (DSSIM) (18.66/0.9206). (l) Ours (20.27/0.9379).

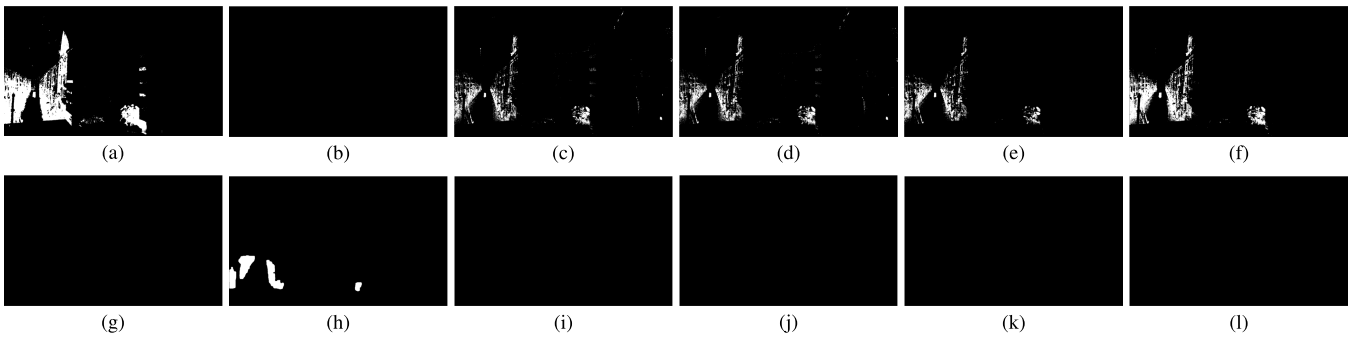

Fig. 11. Binary mask visualization of Figure 10 (Intensity less than 5 would be assigned 1, otherwise 0). (a) Original. (b) Reference. (c) CVC. (d) AGCWD. (e) NEP. (f) SRIE. (g) LIME. (h) Li. (i) DN (MSE). (j) DN (ℓ_1). (k) DN (DSSIM). (l) Ours.

the loss function, where $\boldsymbol{I}_{input}$ and $\boldsymbol{I}_{ref}$ represent the input image and its corresponding reference image, respectively, and $F(\cdot, \Psi)$ is the CNN mapping function with parameters Ψ:

$$DSSIM(\Psi) = \frac{1}{n}\sum_{i}^{n}(1 - ssim(\boldsymbol{I}_{ref}^{(i)} - F(\boldsymbol{I}_{input}^{(i)}, \Psi)))/2 \quad (7)$$

For those two stages, we first train them separately. After the first stage network is trained, we fix the learned weights (Θ and Ω) and train the second stage network to learn the weights Ψ. Having finished the training of the two networks, we remove the loss functions used in the first stage, and jointly fine-tune the whole system with the DSSIM loss as the loss function. In other words, the pre-trained two networks are used as initialization to fine-tune the whole network. This is an end-to-end training scheme.

Figures 7(f) - 7(h) show the enhanced image in the first stage, jointly fine-tuning of first stage with DSSIM loss and the final output, respectively. Although the enhancement result by jointly training the CNNs in the first stage looks sharper than those direct networks with MSE, DSSIM and ℓ_1 loss, it still exhibits some color distortions and detail artifacts. One can see that a two-stage training strategy yields better visual quality than a direct one-stage CNN, and it can correct such color shift and balance the enhancement of smooth and texture components of an image.

V. Experimental Results

We evaluate the proposed CNN enhancer on the built multi-exposure image dataset as well as images outside the dataset. We first present the experimental settings, and then present the

Fig. 12. Single image contrast enhancement results on an over-exposed image by different methods (PSNR/FSIM). (a) Original. (b) Reference. (c) CVC [5] (14.66/0.8344). (d) AGCWD [6] (16.03/0.8843). (e) NEP [8] (15.94/0.8561). (f) SRIE [3] (17.01/0.9013). (g) LIME [1] (15.99/0.8577). (h) Li [59] (15.76/0.8547). (i) DN (MSE) (19.12/0.9117). (j) DN (ℓ_1) (17.16/0.9068). (k) DN (DSSIM) (18.64/0.9296). (l) Ours (20.14/0.9316).

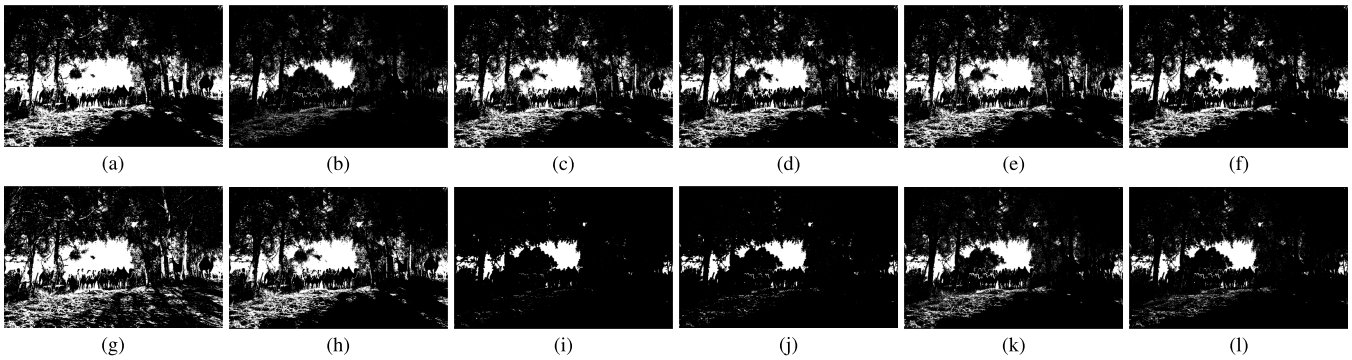

Fig. 13. Binary mask visualization of Figure 12 (Intensity higher than 225 would be assigned 1, otherwise 0). (a) Original. (b) Reference. (c) CVC. (d) AGCWD. (e) NEP. (f) SRIE. (g) LIME. (h) Li. (i) DN (MSE). (j) DN (ℓ_1). (k) DN (DSSIM). (l) Ours.

comparison results with state-of-the-art SICE methods, as well as MEF and stack-based HDR methods. We end up with a discussion on failure case.

A. Experimental Setting

We split all the 589 sequences randomly into training, validation, and test sets with a ratio of 7:1:2. All the three sets are guaranteed to contain images from indoor and outdoor scenes, which contain images with different exposure levels. Note that to further demonstrate the robustness of our method, we also conduct experiments on images outside our dataset, specifically, images from [14].[2] 720, 128 patches of size 129×129 are cropped from the training images, and stochastic gradient descent (SGD) with a batch size of 80 patches is used in training. We implement our model using the TensorFlow package. The momentum parameter and weight decay parameter are set to 0.9 and 0.0001, respectively. The method described in [55] is employed to initialize the weights, and the learning rate is initially set to 0.1 with a decaying factor of 10 for every 30 epochs. Our training process takes about 1 hours for one epoch with a Nvidia Titan X GPU. All the experiments are carried out on a PC with Intel(R) Core(TM) i7-5820K CPU 3.30GHz and 64G memory.

[2]https://images.google.com/.

B. Comparisons With SICE Methods

We compare the proposed CNN-based SICE enhancer with 6 state-of-the-art and representative SICE methods, including

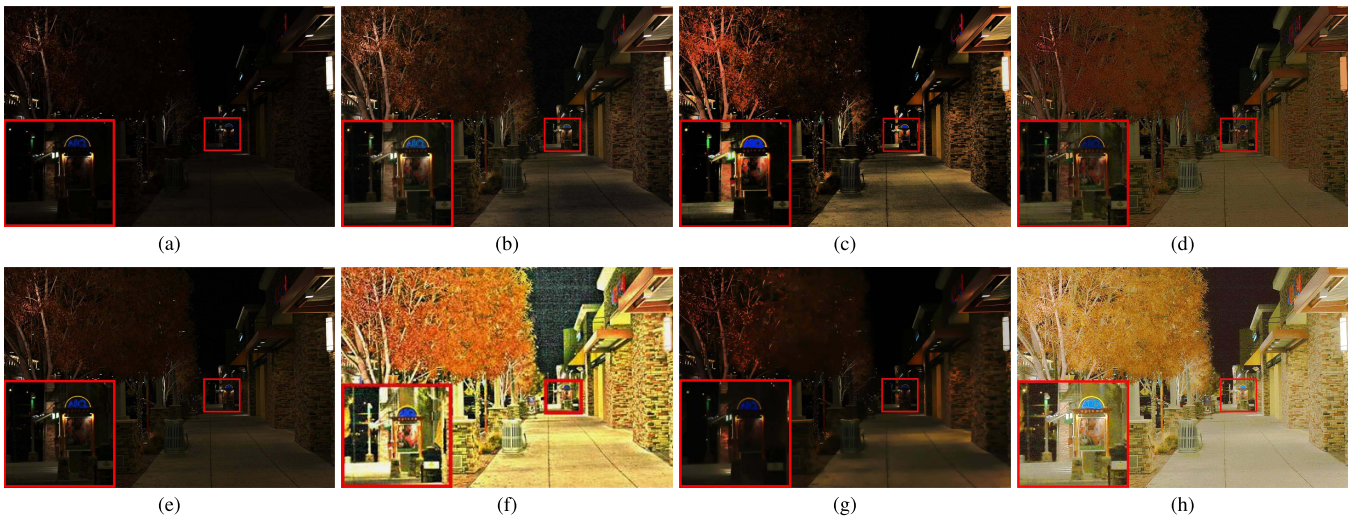

Fig. 14. Single image contrast enhancement results on an under-exposed image from [14]. (a) Original. (b) CVC [5]. (c) AGCWD [6]. (d) NEP [8]. (e) SRIE [3]. (f) LIME [1]. (g) Li [59]. (h) Ours.

TABLE V

AVERAGE PSNR (DB), FSIM INDICES AND RUNNING TIME FOR DIFFERENT SICE AND BASELINE CNN METHODS. (**DN** MEANS DIRECT NETWORK)

	Under-exposure		Over-exposure		Time
	PSNR	FSIM	PSNR	FSIM	(sec.)
CVC [5]	13.47	0.8901	15.11	0.8448	4.64
AGCWD [6]	13.96	0.8996	15.24	0.8601	18.75
NEP [8]	17.21	0.9013	15.62	0.8971	689
SRIE [3]	16.53	0.8978	17.03	0.9209	1e3
LIME [1]	17.68	0.9042	15.79	0.8712	246
Li [59]	15.79	0.8966	15.38	0.8743	1e3
DN (MSE)	19.43	0.9171	20.01	0.9269	23.69
DN (ℓ_1)	18.76	0.9108	19.96	0.9237	23.69
DN (DSSIM)	18.47	0.9284	19.60	0.9304	23.69
Proposed	**19.77**	**0.9347**	**20.21**	**0.9354**	26.47

TABLE VI

THE NUMBER OF FLOPS (FLOATING-POINT OPEARATIONS) OF THE PROPOSED NETWORKS. (M: MEGABYTE)

Network	**FLOPs**
Luminance enhancement network	57 MFLOPS
Detail enhancement network	443 MFLOPS
Whole Image enhancement network	444 MFLOPS
Total	944 MFLOPS

histogram-based methods (CVC [5] and AGCWD [6]), Retinex-based methods (NEP [8], SRIE [3] and LIME [1]) and Li's method [59]. The codes of [1], [3], [8], and [59] are from the original authors, and [5], [6] are from a contrast enhancement toolbox.[3] To verify the effectiveness of our two-stage network, we provide both the visual and quantitative comparisons between the proposed method and 3 baseline direct networks (with ℓ_1, MSE and DSSIM loss). Note that for all the methods, we set the different model parameters for under-exposed and over-exposed images.

1) Comparison on Images From Our Dataset: Figure 10 shows the results on an under-exposure image. The reference image provided in our dataset is also shown. Since the input image contains both bright and dark areas, the histogram-based methods show limited capacity in enhancing image details. Retinex-based methods extract information locally, and the results by NEP, LIME and Li's methods improve the overall visibility of the scene. For those baseline CNN methods, they can produce acceptable visual quality compared to the reference. However, these methods tend to generate unnatural enhancement results, lose many details of the scene, and exhibits some color distortions. Compared with these conventional methods and the baseline CNN methods, the results by our CNN based enhancer have balanced contrast. The image details in most regions are revealed. Figure 12 shows the results on an over-exposure image. Our CNN based enhancer recovers vivid color as well as more details of the scene.

To find out whether regions are saturated or not, the saturation binary mask visualization is also provided. We first convert the RGB image into gray, then threshold the intensity for classifying under-/over-saturation regions. For under-saturation images, intensity less than 5 is assigned as true, otherwise false, as shown in Figure 11. For over-saturation images, intensity higher than 225 is assigned as true, otherwise false, as shown in Figure 13. From Figures 11 and 13, one can see that the CNN methods (both the direct network and two-stage network) can recover almost the same details as the reference images, which demonstrates the advantages of our established dataset and proposed learning-based method.

By using the reference images provided in our dataset as enhancement "groundtruth", we are able to quantitatively evaluate different methods in terms of PSNR and FSIM [60] indices. The results are summarized in Table V. It is not a surprise that our CNN based enhancer has much higher PSNR and FSIM indices than other methods since it learns additional information from external training data. Nonetheless, this

[3]https://github.com/yunfuliu/pixkit.

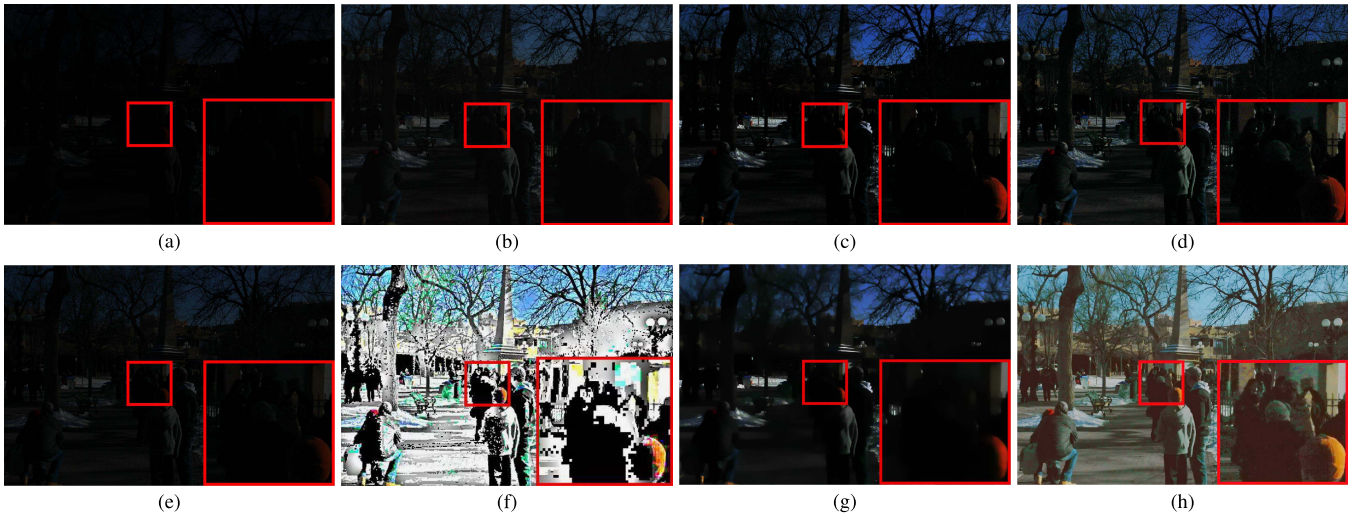

Fig. 15. Single image contrast enhancement results on another under-exposed image from [14]. (a) Original. (b) CVC [5]. (c) AGCWD [6]. (d) NEP [8]. (e) SRIE [3]. (f) LIME [1]. (g) Li [59]. (h) Ours.

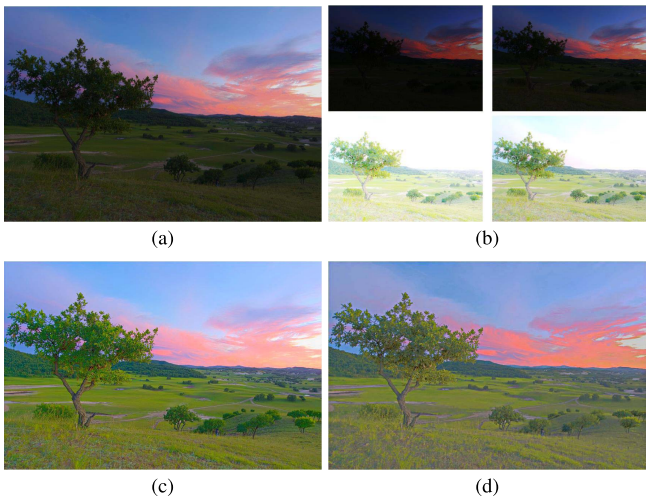

Fig. 16. Comparison between MEF and our method on a static scene. (a) The scene content. (b) Image sequence. (c) Fusion result by [43]. (d) Ours.

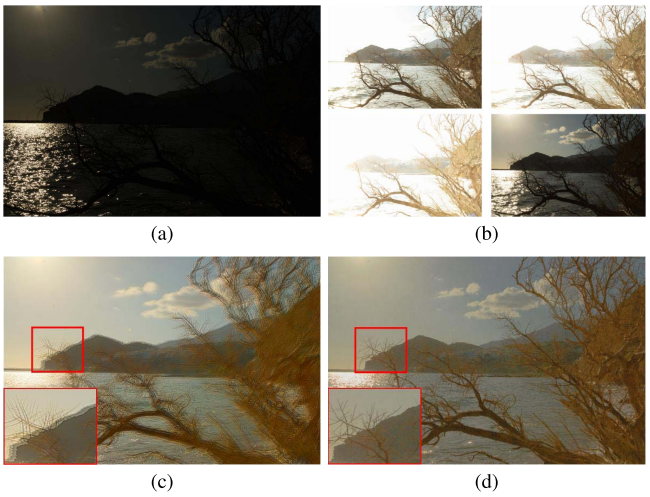

Fig. 17. Comparison between MEF and our method on dynamic scene. (a) The scene content. (b) Image sequence. (c) Fusion result by [43]. (d) Ours.

validates our original motivation to develop a SICE method which could approximate the performance of MEF methods but using only a single image as input. The running time by different methods is also listed in Table V. The number of FLOPs (FLoating-point OPerations) of our algorithms to process an image of resolution $129 \times 129 \times 3$ are also provided, as shown in Table VI. One can see that our method (run on CPU) is comparable to histogram-based methods on speed, and runs much faster than others.

2) Visual Comparison on Images Outside our Dataset: To further demonstrate the robustness of our method, we evaluate our method on images outside our dataset. Considering that most of the existing SICE methods are designed to enhance low light images, we primarily conduct experiments on low light images for a fair comparison. Figures 14 and 15 show the results on two images. One can see that although methods such as NEP, LIME, Li and the baseline CNN methods can reveal the detail from the dark region to some extent, the results enhanced by those methods are not pleasant enough (with obvious artifacts, unnaturalness and color distortions). Our two steam and two stage CNN not only reveals the structural information from the original image, but also presents a more natural result. Since reference images are not available for those images outside our dataset, quantitative measures such as PSNR and FSIM cannot be computed.

C. Comparisons With MEF Method

In this sub-section, we compare our single-image based SICE method with a state-of-the-art multi-image based MEF method [43] on a static scene and a dynamic scene, respectively. Figure 16 shows the results on a static scene. One can see that the MEF method [43] successfully delivers high quality estimation by extracting informative regions from multi-exposure images. Using only one observation as input, our CNN enhancer loses some color information, but its visual quality is still comparable to the result by MEF. Figure 17 shows the results on a dynamic scene, which demonstrate the advantage of SICE method over MEF. Our method produces similar contrast and structural details to the MEF method; however, our SICE method is free of ghosting artifacts, which

(a) (b) (c)

Fig. 18. Failure case. (a) Original. (b) MEF. (c) Ours.

are highly visible in the MEF result. It should be noted that the ghosting artifacts produced by the MEF method [43] could be significantly reduced by using ghost removal algorithms such as those in [53] and [32].

D. Failure Case

Our CNN based SICE method learns a complex nonlinear mapping function to map a low-contrast (either under-exposure or over-exposure) region to a good contrast region. Guided by the reference images generated by MEF/HDR methods, our method is trained to be able to reveal more details from a single low-contrast image than traditional SICE methods, which has been validated in our experiments presented in previous sections. However, it is also found that our method may fail to recover the details for large and severely over-exposed regions. Figure 18 shows an example. One can see that the missing color and structures in the color chart are not recovered, while the details can be seen in the reference image generated by MEF/HDR methods. The reason for the failure may be that the over-exposure is too severe (in terms of both level and area) so that there is little information the CNN can use to synthesize the missing details in the neighborhood.

VI. Conclusion and Future Work

We built a multi-exposure image dataset, which has 589 image sequences and 4,413 high-resolution images of different exposures. For each sequence, a corresponding high quality reference image was generated by using 13 MEF and stack-based HDR algorithms. Subjective tests are also conducted to screen the best quality one as the reference image of each scene. The availability of low-contrast images and their high-quality reference images in our dataset allows the end-to-end learning of high performance SICE methods. As a demonstration, we developed a simple yet powerful CNN-based SICE enhancer, which is capable of adaptively generating high quality enhancement result for a single over-exposed or under-exposed input image. Our experimental results showed that the developed SICE enhancer significantly outperforms state-of-the-art SICE methods, and even outperforms MEF and stack-based HDR methods for dynamic scenes.

Video enhancement is another important application. To apply the proposed methods to videos, we could consider enlarging our dataset and learning an LSTM (long short-term memory) based CNN enhancer to convert the conventional videos to HDR videos. This will be one of our future works.

Acknowledgement

We gratefully acknowledge the support from NVIDIA Corporation for providing us the Titan X GPU used in this research. We also like to thanks the anonymous reviewers for constructive comments.

References

[1] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[2] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, "Robust multi-exposure image fusion: A structural patch decomposition approach," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2519–2532, May 2017.

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

[4] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1921–1935, Sep. 2009.

[5] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3431–3441, Dec. 2011.

[6] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.

[7] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "A multiscale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[8] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.

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

[10] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH Classes*, 2008, p. 31.

[11] S. K. Nayar and V. Branzoi, "Adaptive dynamic range imaging: Optical control of pixel exposures over space and time," in *Proc. ICCV*, 2003, pp. 1168–1175.

[12] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[13] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2009.

[14] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based hdr reconstruction of dynamic scenes," *ACM Trans. Graph.*, vol. 31, no. 6, p. 203, 2012.

[15] T. H. Oh, J. Y. Lee, Y. W. Tai, and I. S. Kweon, "Robust high dynamic range imaging by rank minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1219–1232, Jun. 2015.

[16] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.

[17] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3086–3097, Oct. 2015.

[18] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.

[19] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 97–104.

[20] S. B. Kang, A. Kapoor, and D. Lischinski, "Personalization of image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 1799–1806.

[21] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1646–1654.

[22] L. Xu, J. Ren, Q. Yan, R. Liao, and J. Jia, "Deep edge-aware filters," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 1669–1678.

[23] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[24] V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 769–776.

[25] D. Guo, Y. Cheng, S. Zhuo, and T. Sim, "Correcting over-exposure in photographs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 515–521.

[26] L. Yuan and J. Sun, "Automatic exposure correction of consumer photographs," in *Proc. Comput. Vis.–ECCV*, 2012, pp. 771–785.
[27] Z. Li, J. Zheng, Z. Zhu, W. Yao, and S. Wu, "Weighted guided image filtering," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 120–129, Jan. 2015.
[28] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep CNNS," *ACM Trans. Graph.*, vol. 36, no. 6, p. 178, 2017.
[29] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Trans. Graph.*, vol. 36, no. 4, p. 118, 2017.
[30] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. San Mateo, CA, USA: Morgan Kaufmann, 2010.
[31] Y. Endo, Y. Kanamori, and J. Mitani, "Deep reverse tone mapping," *ACM Trans. Graph.*, vol. 36, no. 6, Nov. 2017, Art. no. 177.
[32] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: How to deal with saturation?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1163–1170.
[33] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, 2017, Art. no. 144.
[34] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 341–349.
[35] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[36] J. Zhang, J. Pan, W.-S. Lai, R. Lau, and M.-H. Yang, "Learning fully convolutional networks for iterative non-blind deconvolution," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6969–6977.
[37] H. Nemoto, P. Korshunov, P. Hanhart, and T. Ebrahimi, "Visual attention in LDR and HDR images," in *Proc. 9th Int. Workshop Video Process. Quality Metrics Consumer Electron. (VPQM)*, 2015.
[38] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video,' *ACM Trans. Graph.*, vol. 22, no. 3, pp. 319–325, 2003.
[39] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proc. Eurograph. (Short Papers)*, 2009, pp. 1–4.
[40] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3634–3646, Dec. 2011.
[41] W. Zhang and W.-K. Cham, "Gradient-directed multiexposure composition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2318–2323, Apr. 2012.
[42] J. Shen, Y. Zhao, S. Yan, and X. Li, "Exposure fusion using boosting Laplacian pyramid," *IEEE Trans. Cybern.*, vol. 44, no. 9, pp. 1579–1590, Sep. 2014.
[43] F. Kou, Z. Li, C. Wen, and W. Chen, "Multi-scale exposure fusion via gradient domain guided image filtering," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 1105–1110.
[44] N. D. B. Bruce, "Expoblend: Information preserving exposure blending based on normalized log-domain entropy," *Comput. Graph.*, vol. 39, pp. 12–23, Apr. 2014.
[45] Photomatrix. (2015). *Commercially-Available HDR Processing Software*. [Online]. Available: http://www.hdrsoft.com/
[46] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*. Boca Raton, FL, USA: CRC Press, 2011.
[47] A. Loza, L. Mihaylova, N. Canagarajah, and D. Bull, "Structural similarity-based object tracking in video sequences," in *Proc. IEEE 9th Int. Conf. Inf. Fusion*, Jul. 2006, pp. 1–6.
[48] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[49] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61 no. 1, pp. 1–11, 1971.
[50] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 257–266, 2002.
[51] Z. Li, Z. Wei, C. Wen, and J. Zheng, "Detail-enhanced multi-scale exposure fusion," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1243–1252, Mar. 2017.
[52] Z. G. Li, J. H. Zheng, and S. Rahardja, "Detail-enhanced exposure fusion," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4672–4676, Nov. 2012.
[53] Z. Li, J. Zheng, Z. Zhu, and S. Wu, "Selectively detail-enhanced fusion of differently exposed images with moving objects," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4372–4382, Oct. 2014.
[54] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, p. 67, 2008.
[55] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 1026–1034.
[56] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 448–456.
[57] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, 2015, pp. 234–241.
[58] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[59] Y. Li, F. Guo, R. T. Tan, and M. S. Brown, "A contrast enhancement framework with JPEG artifacts suppression," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 174–188.
[60] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

Jianrui Cai received the B.E. and M.E. degrees from the College of Computer Science and Electronic Engineering, Hunan University, China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Department of Computing, The Hong Kong Polytechnic University. His research interests include image restoration and image enhancement.

Shuhang Gu received the B.E. degree from the School of Astronautics, Beijing University of Aeronautics and Astronautics, China, in 2010, the M.E. degree from the Institute of Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, China, in 2013, and Ph.D. degree from the Department of Computing, The Hong Kong Polytechnic University, in 2017. He currently holds a post-doctoral position at ETH Zurich, Switzerland. His research interests include image restoration, sparse, and low rank models.

Lei Zhang (M'04–SM'14–F'18) received the B.Sc. degree from the Shenyang Institute of Aeronautical Engineering, Shenyang, China, in 1995, and the M.Sc. and Ph.D. degrees in control theory and engineering from Northwestern Polytechnical University, Xian, China, in 1998 and 2001, respectively. From 2001 to 2002, he was a Research Associate with the Department of Computing, The Hong Kong Polytechnic University. From 2003 to 2006, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, McMaster University, Canada. In 2006, he joined the Department of Computing, The Hong Kong Polytechnic University, as an Assistant Professor, where he has been a Chair Professor since 2017. His research interests include computer vision, pattern recognition, image and video analysis, and biometrics. He has published over 200 papers in those areas. As of 2017, his publications have been cited over 28,000 times in the literature. He was a Clarivate Analytics Highly Cited Researche from 2015 to 2017. He is an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, the *SIAM Journal of Imaging Sciences and Image and Vision Computing*.