Band Representation-Based Semi-Supervised Low-Light Image Enhancement: Bridging the Gap Between Signal Fidelity and Perceptual Quality

Wenhan Yang, *Member, IEEE*, Shiqi Wang, *Member, IEEE*, Yuming Fang, *Senior Member, IEEE*, Yue Wang, and Jiaying Liu, *Senior Member, IEEE*

***Abstract*— It has been widely acknowledged that under-exposure causes a variety of visual quality degradation because of intensive noise, decreased visibility, biased color, *etc*. To alleviate these issues, a novel semi-supervised learning approach is proposed in this paper for low-light image enhancement. More specifically, we propose a deep recursive band network (DRBN) to recover a linear band representation of an enhanced normal-light image based on the guidance of the paired low/normal-light images. Such design philosophy enables the principled network to generate a quality improved one by reconstructing the given bands based upon another learnable linear transformation which is perceptually driven by an image quality assessment neural network. On one hand, the proposed network is delicately developed to obtain a variety of coarse-to-fine band representations, of which the estimations benefit each other in a recursive process mutually. On the other hand, the extracted band representation of the enhanced image in the recursive band learning stage of DRBN is capable of bridging the gap between the restoration knowledge of paired data and the perceptual quality preference to high-quality images. Subsequently, the band recomposition learns to recompose the band representation towards fitting perceptual regularization of high-quality images with the perceptual guidance. The proposed architecture can be flexibly trained with both paired and unpaired data. Extensive experiments demonstrate that our method produces better enhanced results with visually pleasing contrast and color distributions, as well as well-restored structural details.**

***Index Terms*— Low light, image enhancement, perceptual quality.**

Manuscript received July 9, 2020; revised January 15, 2021; accepted February 22, 2021. Date of publication March 3, 2021; date of current version March 9, 2021. This work was supported in part by the Science, Technology, and Innovation Commission of Shenzhen Municipality under Project JCYJ20180307123934031, in part by the National Natural Science Foundation of China under Grant 62022002 and Grant 61772043, in part by the Hong Kong ITF UICP under Grant UIM359 (CityU 9440203), in part by the Hong Kong Research Grants Council, Early Career Scheme (RGC ECS) under Grant 21211018, in part by the General Research Fund (GRF) under Grant 11203220, and in part by the Fundamental Research Funds for the Central Universities. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhenzhong Chen. *(Corresponding author: Shiqi Wang.)*

Wenhan Yang and Shiqi Wang are with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: wyang34@cityu.edu.hk; shiqwang@cityu.edu.hk).

Yuming Fang is with the School of Information Management, Jiangxi University of Finance and Economics, Nanchang 330013, China (e-mail: fa0001ng@e.ntu.edu.sg).

Yue Wang is with Bytedance (HK) Ltd., Hong Kong (e-mail: wangyue.v@bytedance.com).

Jiaying Liu is with the Wangxuan Institute of Computer Technology, Peking University, Beijing 100080, China (e-mail: liujiaying@pku.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2021.3062184, provided by the authors.

Digital Object Identifier 10.1109/TIP.2021.3062184

I. Introduction

In LOW-LIGHT acquisition environments, a variety of visual quality degradation, *i.e.* low contrast, low visibility, and intensive noises may appear in the acquired images. It is beneficial to apply more advanced capture devices and adopt specialized photographic techniques to ease these quality degradation issues. However, it is still quite challenging to completely avoid the presented noise, especially when insufficient light can reach camera sensors such that the background scene signals might be inevitably contaminated and even buried by the system noise. Though increasing the exposure time to suppress noise would be beneficial, blurriness is then introduced inevitably in this scenario. Therefore, it is highly desired to apply the software low-light enhancement solutions instead of resorting to advanced hardware. In particular, such methodologies enhance images acquired in the low-light environment back to a normal one, such that the contrast, visibility and noise are expected to be enhanced, improved and suppressed, respectively. The enhancement process not only brings in visual quality improvement, but also provides a promising start point for high-level visual understanding tasks (*e.g.,* recognition or object detection).

In general, it is very challenging to enhance low-light images to the perfect quality as the inherent noise is also prone to be amplified in this process. In the past decades, numerous efforts have been dedicated to addressing this issue. The most traditional histogram equalization (HE) methods [1], [2] enhance the low-light images via stretching the dynamic range of the images. However, noise is amplified and regions might be characterized with undesirable illumination. Retinex theory-based methods [3] decompose an image into two separate layers, *i.e.* reflectance and illumination layers, which are further processed for enhancement purpose. Moreover, a variety of filters [4]–[6] have been built to decompose the images, and different priors [7]–[11] have also been developed and enforced on the decomposed illumination and reflectance.

Recently, the strong representation capability of deep neural networks also inspires a series of deep-learning based approaches for low-light enhancement. These methods [12]–[16] learn to alleviate the related composite complex degradation relying on the sophisticated designed learning models as well as the learned mapping of the paired images which are well-prepared as low and normal-light. However, the existing loss functions are not specifically

designed according to the human visual perception and may cause the in-appropriate estimation of the intrinsic image structure, resulting in unsatisfactory visual quality, *e.g.,* residual noise and biased color distribution. Recently, EnlightenGAN [17] was developed without the constraint of paired supervision. The method only relies on a database containing low as well as normal light images that are not necessarily paired. EnlightenGAN demonstrates the feasibility to train a low-light enhancement model with unpaired data. However, without paired data as supervision, it is difficult to restore fine structural details reliably, and the enhanced results still contain the intensive noise.

In general, deep-learning based methods are classified into two categories, including *fully supervised methods* and *unsupervised methods*. In *fully supervised methods*, the enhancement models are trained with the ground truth guidance in terms of *paired supervision* which is assumed to be available to model the detailed signal structure in the training phase. Therefore, with the ground truth guidance, the networks are more capable of learning to suppress noise and impair local structural details. In *unsupervised methods*, the enhancement mapping is trained based on *unpaired low/normal-light image sets*, as it is effortless to collect such large-scale data with diversified content. As such, these methods can learn to recover the illumination, contrast, and color in an adaptive and economical manner.

In this work, we consider the strengths, weaknesses, and potentials of existing methods comprehensively and explore to construct a unified architecture with the merits of both *fully supervised* and *unsupervised* methods. More specifically, we develop a novel semi-supervised learning framework for low-light image enhancement. We propose a deep recursive band network (DRBN) relying on the band representation to link the signal-level fidelity constraint obtained from paired supervision and the nature scene statistics prior [18] based upon a quality assessment neural network, measuring the preference in terms of human visual perception. In particular, this is rooted in the widely accepted philosophy that the human visual system evolves with the perception of natural scenes, such that the network is learned based on a large collection of natural images with human annotated mean opinion scores (MOS). In the first stage of DRBN, a linear band is first recovered via training with paired low/normal-light images. The estimations of this representation are beneficial mutually in a recursive process. Subsequently, the extracted band representation of the enhanced image in the first stage of DRBN (recursive band learning) is capable of bridging the gap between the paired data knowledge and the perceptual quality provided by a quality assessment neural network. In the second stage of DRBN (band recomposition), the band representation is reconstructed towards fitting underlying visual characteristics of high-quality images with the perceptual guidance provided by a quality assessment neural network. The two-stage design facilitates to generate enhanced results with visually pleasing contrast, color distributions, and well-reconstructed structural details.

This work is the extension of our previous conference paper [19]. Compared to the conference version, this work further makes significant contributions. First, we introduce the Long Short Term Memory (LSTM) networks to further augment the feature representation capacity in a form of residual learning in each recurrence, which improves the performance with slightly increased complexity. Second, instead of applying quality-guided adversarial learning, we introduce an image quality assessment neural network to infer the nature scene statistics, implying the preference in terms of human visual perception. The image quality assessment network is trained with the whole Aesthetic Visual Analysis (AVA) [20] dataset, such that it can provide more comprehensive and accurate feedback to train the enhancement network, making the enhancement network produce visually satisfactory results. The contributions of this work are summarized as follows:

- To our best knowledge, this is the first attempt to enhance the quality of low-light images relying on a *semi-supervised learning* framework, where a deep recursive band representation is specifically designed to combine fully-supervised and unsupervised learning to integrate both superiority.
- The proposed framework is sophisticatedly designed to achieve a series of coarse-to-fine band representations. The estimations of these band representations are mutually beneficial through the end-to-end training in a recursive way, enabling the capabilities of noise removal and local structural details recovery.
- The perceptual characteristics are incorporated into deep band representations based upon a quality assessment neural network learned from an aesthetic analysis dataset with human labeled mean opinion score (MOS), which guides the recomposed results to move towards better visual quality. To our best knowledge, this is also the first trial to pursue perception-driven low-light image enhancement.

The rest of this paper is organized as follows. Section II briefly reviews the related work. Section III presents the proposed deep recursive band network for semi-supervised low-light image enhancement. Experimental results and concluding remarks are presented in Sections IV and V, respectively.

II. Related Works

A. Global Adjustment-Based Image Enhancement

The earliest low-light enhancement strategies adjust the illumination by stretching out the range, which could easily lead to over-exposure and under-exposure, such as HE based methods [1], [2]. Without the local adaptation, such a straightforward enhancement methodology results in undesirable illumination and intensive noise. Along this vein, several kinds of priors have also been adopted to constrain the equalization process, *e.g.* mean intensity preservation [21], white and black stretching [22] and noise robustness to improve the overall visual quality of the output image. In [23], [24], the HE is applied to the difference of pixels for fine-grained manipulation on the histograms. In [25], depth information is introduced as side information to adaptively guide the pixel value variation. In [26], [27], the imaging and visual perception models are used as the guidance information for the low-light image enhancement. Moreover, the visibility can also

be enhanced by adopting dehaze models to inverted low-light images [28], [29]. However, the off-line denoising [30] is thereby applied for noise removal which unfortunately leads to structural detail blurriness.

B. Retinex Model-Based Image Enhancement

In the Retinex-based methods [3], the design philosophy is to decompose to-be-enhanced images into illumination and reflectance layers which could be further adaptively adjusted. The joint noise suppression and illumination adjustment can be performed via various priors, *e.g.,* structure aware prior [8], weighted variation [9], and multiple derivatives of illumination [7]. There are also a series of Retinex models, *e.g.,* single-scale Retinex [4], multi-scale Retinex [5], naturalness Retinex [6], and robust Retinex [10], [11] developed towards this goal. Based on these models, impressive enhancement results in terms of thin noise removal and illumination adjustment have been achieved. In [31], the weight of each single-scale Retinex is adaptively computed based on the input image. Wang *et al.* [5] construct a bright-pass filter for Retinex decomposition. In [32], prior distributions of the reflectance and the illumination are modeled with a hierarchical Bayesian model.

Efforts have also been devoted to exploiting more effective priors [7]–[9] to regularize the enhancement of illumination and reflectance layers. These methods may ignore the constraints on the reflectance, such that the latent intensive noises in the low-light regions are usually amplified. Li *et al.* [10] proposed to extend the traditional Retinex model to a robust one with an explicit noise term. Ren *et al.* [11] also developed a sequential algorithm to address the inverse recovery problem. These methods are built with constraints obtained by hand-crafted design, which may not be adaptive enough to model the complicated signal characteristics of various kinds of images. Therefore, their results might inevitably involve noises, over-exposed and under-exposed local regions.

C. Deep Learning-Based Image Enhancement

Inspired by the impressive performance of deep neural networks in low-level vision tasks, a series of deep learning based low light enhancement models have been developed. Lore *et al.* [12] made the first attempt in this trail by developing a deep auto-encoder named Low-Light Net (LLNet) for contrast adjustment and noise removal. Subsequently, different networks [6], [15], [16], [33], [34] towards this goal have been proposed. In [33], [35], [36], the multi-scale features are injected into the multi-branch architecture to form better low-light enhancement results. Attempts have also been made to build paired low/normal-light datasets for the model training [6], [12], [16]. Diversified losses are utilized to help train the enhancement model, such as MSE [12], SSIM [6], and compound loss [16]. In [33], [34], [37], Retinex structure is incorporated into the design of effective deep neural networks, to equip with the advantages of Retinex-based methods, *i.e.* good signal structure, as well as deep learning-based methods, *i.e.* the general effective priors extracted from the large-scale training data. In [14], [48], the layer decomposition and separated processing are used to better model the image structure (low-frequency component) and texture detail (high-frequency component). In [50], dark regions are precisely detected by a new visual attention module and the brightness is adjusted to maintain the color, tone, and brightness of the input images and prevent normally illuminated areas from being saturated and distorted. The results and performance of these methods, trained on the paired dataset, could be largely dependent on the used training dataset. As the synthetic data could not fully account for the degradation in real application scenarios and practically acquired paired data cover limited scenes, there is still a large gap between the results of these methods and satisfactory quality.

In [17], [38], adversarial learning is utilized to capture the statistical characteristics of visual signals beyond the traditional metrics. In particular, for EnlightenGAN [17], Jiang *et al.* applied the unpaired learning to train a low-light enhancement model, which gets rid of paired dataset construction and addresses the domain shift problem between the limited training data and practical applications in the wild. In [49], the light enhancement problem is formulated as a task of image-specific curve estimation with a deep network. Pixel-wise and high-order curves are estimated for dynamic range adjustment of the input low-light image. In addition, efforts have also been devoted to deep-learning based image enhancement from raw images [13], as well as joint low-light image enhancement and high-level visual understanding such as face detection [39] and object detection [40].

In our work, we concentrate on perceptual quality improvement of low-light images in RGB format. Compared with previous relevant studies, we build a semi-supervised framework. In our framework, useful knowledge is creatively used to offer joint perceptual guidance via a quality assessment network. In particular, the quality assessment network is learned from a large collection of natural images with human annotated MOS values for both structural detail modeling and global illumination, color and contrast recovery.

III. DRBN for Semi-Supervised Low-Light Image Enhancement

A. Motivations

1) Recursive Band Learning: The training data in the paired form is capable of offering signal fidelity constraint with strong capability to recover the signal structure details. Herein, we adopt a recursive band learning to recover structural detail with the paired training. In addition to the enhanced image $\hat{y}$, a variety of band representations $\left\{\Delta\hat{y}_{s_1}^T, \Delta\hat{y}_{s_2}^T, \ldots, \Delta\hat{y}_{s_n}^T\right\}$ are also produced progressively from a low-light image x, where $\hat{y} = \sum_{i=1}^{n}\hat{y}_{s_i}^T$ and s_i indicates the order of the band representation. $\Delta\hat{y}_{s_i}^T$ is learned by fully-supervised learning on paired low/normal-light data, such that the high-order band $\hat{y}_{s_i}^T$ depends on the low-order one $\hat{y}_{s_{i-1}}^T$.

2) Facilitating Recursive Band Representation with Adversarial Learning: Relying on signal fidelity constraint in the first stage only cannot fully guarantee the visual quality. Therefore, in the second stage, the learned band representations are recomposed, leading to better visual quality from the perspective of visual perception,

$$\hat{y} = \sum_{i=1}^{n} w_i\left(x, \left\{\Delta\hat{y}_{s_1}^T, \Delta\hat{y}_{s_2}^T, \ldots, \Delta\hat{y}_{s_n}^T\right\}\right)\Delta\hat{y}_{s_i}^T(x), \quad (1)$$

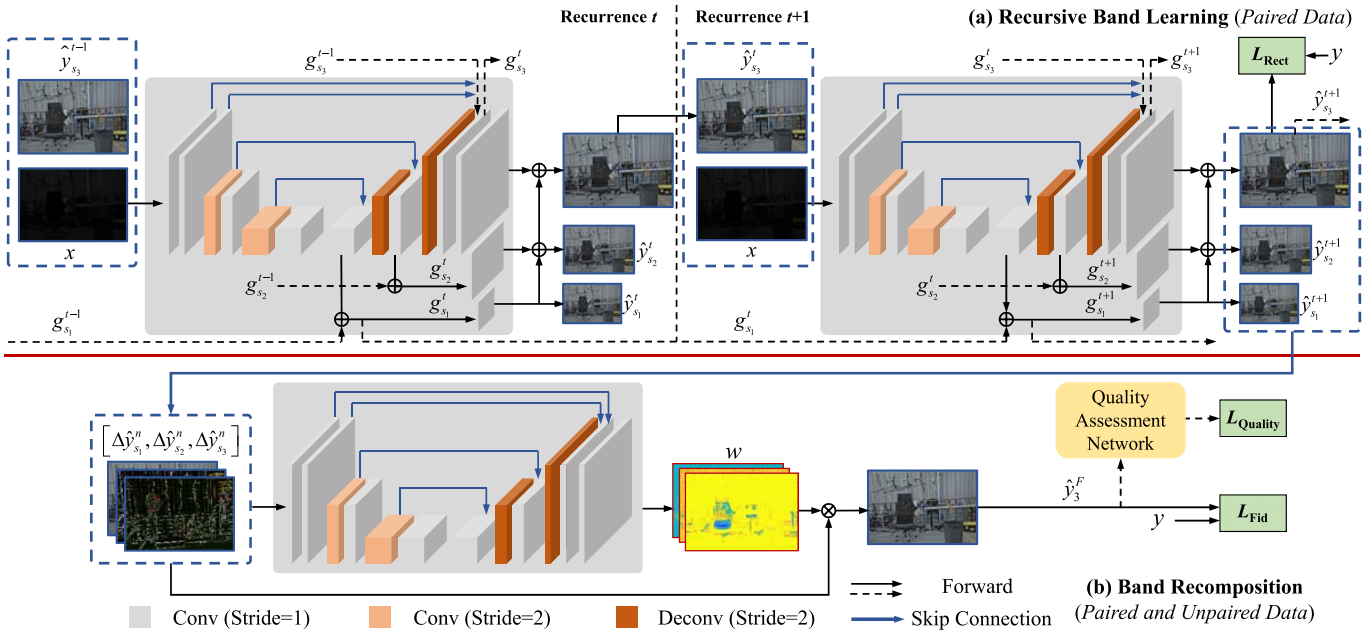


Fig. 1. Illustration of the proposed DRBN consisting of recursive band learning and band re-composition. In the first stage, a coarse-to-fine band representation can be obtained in a recursive process, such that different band signals are jointly inferred. The enhancement result from the last recurrence is further leveraged as the guidance of the next recurrence, making the later recurrence learn the residue in the feature and signal domains at different scales. In the second stage, the band representation improves perceptual quality based upon a neural network which is responsible for image quality assessment.

where $w_i(\cdot)$ is the learned weighting factors in the recomposition process, which aims to recompose band signals of an enhanced image. In this manner, well-reconstructed images with better perceptual quality can be obtained which are characterized with abundant structural details, little noise, superior contrast, illumination, and color distributions.

B. Deep Recursive Band Network

1) Architecture: As illustrated in Fig. 1, there are two modules in our framework, including recursive band learning (learning from paired data), and band recomposition (learning from unpaired data). In the recursive band learning, DRBN is built to generate an image with normal-light with the guidance of the low-light input recursively. The output of the previous recurrence leads to the intermediate estimation which can be regarded as the guidance input of the next recurrence, such that all band estimations are tied together and jointly optimized. The DRBN employs the residual learning in both feature and signal domains, such that the latter recurrence only accounts for estimating the residual features and images to obtain better estimations. In this manner, structural details are better modeled with the later recurrence with the suppression of noise. In each recurrence, a variety of coarse-to-fine band representations are obtained and subsequently merged into the enhancement results. Such band representation is effective in terms of the fidelity priors learned from the paired images and the natural scene statistics from high-quality images. In the band re-composition, we feed-forward the band representation into another network to generate a set of transformation coefficients, which manipulate and fuse these bands linearly. An image quality assessment network determining the perceptual quality of a given image, is utilized to provide the feedback to the enhancement network as guidance. A set of natural images with human annotated MOS values are used to train the quality assessment network which aims at extracting the prior natural scenes statistics and inferring the preference in terms of human visual perception. In this way, the overall promising enhancement results from the perspective of both signal fidelity and perceptual quality can be obtained.

2) Recursive Band Learning: The intrinsic mapping that could be learned from paired data learning is leveraged to recover each band of an enhanced image. The band learning networks (BLN) based upon a series of U-Net like deep neural networks are established, as shown in Fig. 1. In each BLN, the input formed by the concatenation of x and the enhanced image of the last recurrence $\hat{y}_{s_3}^{t-1}$ is projected into the feature space. Subsequently, the features are transformed with several convolutional layers. In intermediate layers, we first down-sample the spatial resolutions of features and then up-sample them via stride convolutions and deconvolutions. Skip connections are used to link the features with the same spatial resolution from shallow layers to deep layers, making the local information that is contained in the features generated by shallow layers reach the output feature. Each BLN generates three features at the scales $s_1 = 1/4$, $s_2 = 1/2$ and $s_3 = 1$.

For convenience, the first recurrence of the recursive learning is given by,

$$\begin{aligned}
\left[g_{s_1}^1, g_{s_2}^1, g_{s_3}^1\right] &= G_{\text{BLN_G}}^1(x), \\
\left[\beta_{s_i}^1, h_{s_i}^1\right] &= G_{\text{LSTM}}^i\left(g_{s_i}^1, h_{s_i}^0\right), i = 1, 2, 3, \\
\widetilde{g}_{s_i}^t &= g_{s_i}^t + \beta_{s_i}^1, i = 1, 2, 3, \\
\hat{y}_{s_1}^1 &= G_{\text{R_}s_1}^1\left(\widetilde{g}_{s_1}^1\right), \\
\hat{y}_{s_2}^1 &= G_{\text{R_}s_2}^1\left(\widetilde{g}_{s_2}^1\right) + G_{\text{U}}\left(\hat{y}_{s_1}^1\right), \\
\hat{y}_{s_3}^1 &= G_{\text{R_}s_3}^1\left(\widetilde{g}_{s_3}^1\right) + G_{\text{U}}\left(\hat{y}_{s_2}^1\right),
\end{aligned} \tag{2}$$

where $g^1_{s_1}, g^1_{s_2}, g^1_{s_3}$ denote features extracted from x at their corresponding scales. Moreover, $\widetilde{g}^1_{s_1}, \widetilde{g}^1_{s_2}, \widetilde{g}^1_{s_3}$ are features augmented by LSTMs. $G^1_{\text{BLN\_G}}(\cdot)$ denotes the feature extraction process, and $G^1_{\text{R\_}s_1}(\cdot)$, $G^1_{\text{R\_}s_2}(\cdot)$ and $G^1_{\text{R\_}s_3}(\cdot)$ indicate projection of the augmented features back to the image domain at the corresponding scales. Herein, $G_{\text{U}}(\cdot)$ is the up-sampling process. As such, the image is first reconstructed at the roughest scale s_1, then at finer scales the residual signals are predicted to compose the whole.

Our convolution LSTM unit $G^i_{\text{LSTM}}(\cdot)$ consists of an input gate $\mathbf{j}^i_t$, a forget gate $\mathbf{f}^i_t$, an output gate $\mathbf{o}^i_t$ as well as a cell state $\mathbf{C}^i_t$. Given the input $\mathbf{X}^i_t = g^t_{s_i}$ and the cell state as well as the output feature $\left[\mathbf{C}^i_{t-1}, \mathbf{H}^i_{t-1}\right] = h^{t-1}_{s_i}$ in Eqn. (2) and (3), the interaction between states and gates along the temporal dimension t is defined as follows:

$$\begin{aligned}
\mathbf{j}^i_t &= \sigma\left(\mathbf{W}^i_{xj} * \mathbf{X}^i_t + \mathbf{W}^i_{hj} * \mathbf{H}^i_{t-1} + \mathbf{b}^i_j\right),\\
\mathbf{f}^i_t &= \sigma\left(\mathbf{W}^i_{xf} * \mathbf{X}^i_t + \mathbf{W}^i_{hf} * \mathbf{H}^i_{t-1} + \mathbf{b}^i_f\right),\\
\mathbf{C}^i_t &= \mathbf{f}^i_t \odot \mathbf{C}^i_{t-1} + \mathbf{j}^i_t \odot \tanh\left(\mathbf{W}^i_{xc} * \mathbf{X}^i_t + \mathbf{W}^i_{hc} * \mathbf{H}^i_{t-1} + \mathbf{b}^i_c\right),\\
\mathbf{o}^i_t &= \sigma\left(\mathbf{W}^i_{xo} * \mathbf{X}^i_t + \mathbf{W}^i_{ho} * \mathbf{H}^i_{t-1} + \mathbf{b}^i_o\right),\\
\mathbf{H}^i_t &= \mathbf{o}^i_t \odot \tanh\left(\mathbf{C}^i_t\right),
\end{aligned}$$

where $\mathbf{W}$ and $\mathbf{b}$ are learnable weights and biases in LSTM, respectively. The operators $*$ and $\odot$ denote the convolution and element-wise multiplication . At the output end, we have: $\beta^t_{s_i} = \mathbf{H}^i_t$ and $h^t_{s_i} = \left[\mathbf{C}^i_t, \mathbf{H}^i_t\right]$.

Subsequently, at the t-th recurrence, only the learned residual features and images are guided by previously estimated results. x and previously estimation $\hat{y}^{t-1}_{s_3}$ are concatenated and regarded as the input as follows:

$$\begin{aligned}
\left[\Delta g^t_{s_1}, \Delta g^t_{s_2}, \Delta g^t_{s_3}\right] &= G^t_{\text{BLN_G}}\left(x, \hat{y}^{t-1}_{s_3}\right),\\
g^t_{s_i} &= \Delta g^t_{s_i} + g^{t-1}_{s_i}, i = 1, 2, 3,\\
\left[\beta^t_{s_i}, h^t_{s_i}\right] &= G^i_{\text{LSTM}}\left(g^t_{s_i}, h^{t-1}_{s_i}\right), i = 1, 2, 3,\\
\widetilde{g}^t_{s_i} &= \beta^1_{s_i} + g^t_{s_i}, i = 1, 2, 3,\\
\hat{y}^t_{s_1} &= G^t_{\text{R_}s_1}\left(\widetilde{g}^t_{s_1}\right),\\
\hat{y}^t_{s_2} &= G^t_{\text{R_}s_2}\left(\widetilde{g}^t_{s_2}\right) + G_{\text{U}}\left(\hat{y}^t_{s_1}\right),\\
\hat{x}^t_{s_3} &= G^t_{\text{R_}s_3}\left(\widetilde{g}^t_{s_3}\right) + G_{\text{U}}\left(\hat{y}^t_{s_2}\right).
\end{aligned} \tag{3}$$

As such, all band features are closely connected, forming a joint optimization of all bands. At the final recurrence T (set to 4 in our work), the reconstruction loss is given by,

$$\begin{aligned}
L_{\text{Rect}} = -\Big(&\phi\left(\hat{y}^T_{s_3}, y\right) + \lambda_1 \phi\left(\hat{y}^T_{s_2}, F_D\left(y, s_2\right)\right)\\
&+ \lambda_2 \phi\left(\hat{y}^T_{s_1}, G_D\left(y, s_1\right)\right)\Big),
\end{aligned} \tag{4}$$

where $G_D(\cdot)$ is the down-sampling process given the scaling factor s_i. $\phi(\cdot)$ is the calculation of quality in terms of structural similarity index (SSIM) [41]. Moreover, λ_1 and λ_2 denote the weighting factors.

Overall, there are three main characteristics for our recursive band learning that benefit the performance of the enhancement task,

- The low-order band in the current recurrence is influenced by the high-order one based on the inference from the last recurrence. As such, the relationship between the low and high-order bands is bi-directional, since the high-order bands can offer useful guidance for low-order bands recovery.
- The recursive estimation makes different bands learn to adjust their estimations with the guidance of previous estimations of all bands.
- The recursive learning boosts the modeling capacities. The later recurrence only is expected to recover the residue signals with the guidance of the estimation from previous recurrences. Hence, accurate estimations can be acquired with the attention paid to fine structural details.

3) Band Recomposition: Based upon the paired data learning, the band recovery process is performed by converting from the low-light images to the normal-light images, which reliably restores structural details and suppresses noise. However, it has been widely acknowledged that the signal fidelity cannot fully account for the characteristics of human visual perception, especially due to the widely acknowledged gap in terms of global statistics such as lighting and color distribution. Therefore, our model learns to recompose the restored band signals with the perceptual guidance via learning based perceptual quality assessment model trained with an aesthetic quality assessment image dataset. The images in the aesthetic visual analysis dataset AVA [20] are labeled with MOS values which characterize prior statistics of natural scenes that effectively measure the preference in terms of human visual perception. We utilize another U-like network to model the recomposition process $G_{\text{RC}}(\cdot)$ to produce the coefficients used for recomposing the band signals as follows,

$$\begin{aligned}
\{w_1, w_2, w_3\} &= G_{\text{RC}}\left(\left\{\Delta\hat{y}^T_{s_1}, \Delta\hat{y}^T_{s_2}, \Delta\hat{y}^T_{s_3}\right\}\right),\\
\hat{y}^F_3 &= \sum_{i=1}^{3} w_i \Delta\hat{y}^T_{s_i},\\
\Delta\hat{y}^T_{s_i} &= \hat{y}^T_{s_i} - G_{\text{U}}\left(\hat{y}^T_{s_{i-1}}\right),\ i = 2, 3,\\
\Delta\hat{y}^T_{s_1} &= \hat{y}^T_{s_1},
\end{aligned} \tag{5}$$

where $\hat{y}^F_3$ is trained based on the following three loss functions,

$$L_{\text{Detail}} = -\phi\left(\hat{y}^F_3 - y\right), \tag{6}$$

$$L_{\text{Percept}} = \left\| G_{\text{P}}(\hat{y}^F_3) - G_{\text{P}}(y)\right\|^2_2, \tag{7}$$

$$L_{\text{Quality}} = \left\| D\left(\hat{y}^F_3\right) - l_r\right\|^2_2, \tag{8}$$

Herein, $D(\cdot)$ is a neural network-based quality predictor [42] trained with AVA dataset [20], which effectively measures the probability that $\hat{y}^F_3$ is perceptually preferred. l_r is set to 10 (10 representing the best quality). $G_{\text{P}}(\cdot)$ is the process to extract deep features from a pretrained VGG network.

In this stage, the whole loss function is given by,

$$L_{\text{SBR}} = L_{\text{Percept}} + \lambda_3 L_{\text{Detail}} + \lambda_4 L_{\text{Quality}}, \tag{9}$$

where λ_3 and λ_4 denote weighting factors.

TABLE I

QUANTITATIVE PERFORMANCE COMPARISONS ON IMAGES FROM *LOL-Syn* DATASET. EG DENOTES ENLIGHTENGAN

Metric	BIMEF [43]	BPDHE [21]	CRM [44]	DHECE [24]	Dong *et al.* [45]	EFF [26]	CLAHE [46]	LIME [8]	MF [7]
PSNR	17.20	12.49	18.91	17.75	16.90	17.20	15.37	16.88	17.50
SSIM	0.7127	0.5290	0.7864	0.7800	0.7487	0.7127	0.5527	0.7762	0.7514
SSIM-GC	0.7844	0.6922	0.8138	0.8189	0.7915	0.7844	0.5561	0.7834	0.8261
Metric	MR [5]	JED [11]	SRIE [9]	RRM [10]	DRD [34]	DeepUPE [16]	SICE [14]	EG [17]	Ours
PSNR	14.98	17.48	14.50	17.15	17.13	15.08	18.50	16.57	**23.22**
SSIM	0.6248	0.7444	0.6163	0.7277	0.7978	0.6225	0.7631	0.7338	**0.9275**
SSIM-GC	0.6367	0.7965	0.7544	0.7757	0.8156	0.7446	0.7885	0.7849	**0.9368**

TABLE II

QUANTITATIVE PERFORMANCE COMPARISONS ON IMAGES FROM *LOL-Real* DATASET. EG DENOTES ENLIGHTENGAN

Metric	BIMEF [43]	BPDHE [21]	CRM [44]	DHECE [24]	Dong *et al.* [45]	EFF [26]	CLAHE [46]	LIME [8]	MF [7]
PSNR	17.85	13.84	19.65	14.64	17.26	17.85	13.13	15.24	18.73
SSIM	0.6526	0.4254	0.6623	0.4450	0.5270	0.6526	0.3709	0.4702	0.5590
SSIM-GC	0.7231	0.5936	0.6968	0.4521	0.5715	0.7231	0.3947	0.4905	0.5765
Metric	MR [5]	JED [11]	RRM [10]	SRIE [9]	DRD [34]	DeepUPE [16]	SICE [14]	EG [17]	Ours
PSNR	11.67	17.33	17.34	17.34	15.47	13.27	19.40	18.23	**20.29**
SSIM	0.4269	0.6654	0.6859	0.6859	0.5672	0.4521	0.6906	0.6165	**0.8310**
SSIM-GC	0.5158	0.7236	0.7459	0.7075	0.7476	0.7051	0.7250	0.6452	**0.8479**

4) Summarization: In the DRBN, the band representation learning is first performed. Then, we make each band signal learn to be recovered based on the guidance of the paired dataset, to ensure both signal fidelity and structural detail recovery. Subsequently, we perform band recomposition to further improve the visual quality of enhanced images based upon the perceptual guidance provided by a neural network-based quality assessment model trained with aesthetic quality assessment dataset, where a large number of natural images with MOS labels serve as the guidance information for extracting natural scene statistics.

IV. EXPERIMENTAL RESULTS

In this section, the performance of the proposed scheme is validated through both quantitative and qualitative evaluations. Moreover, the results and analyses based on ablation studies are also shown in an effort to provide more useful evidence regarding the effectiveness of the proposed scheme.

A. Experimental Setting

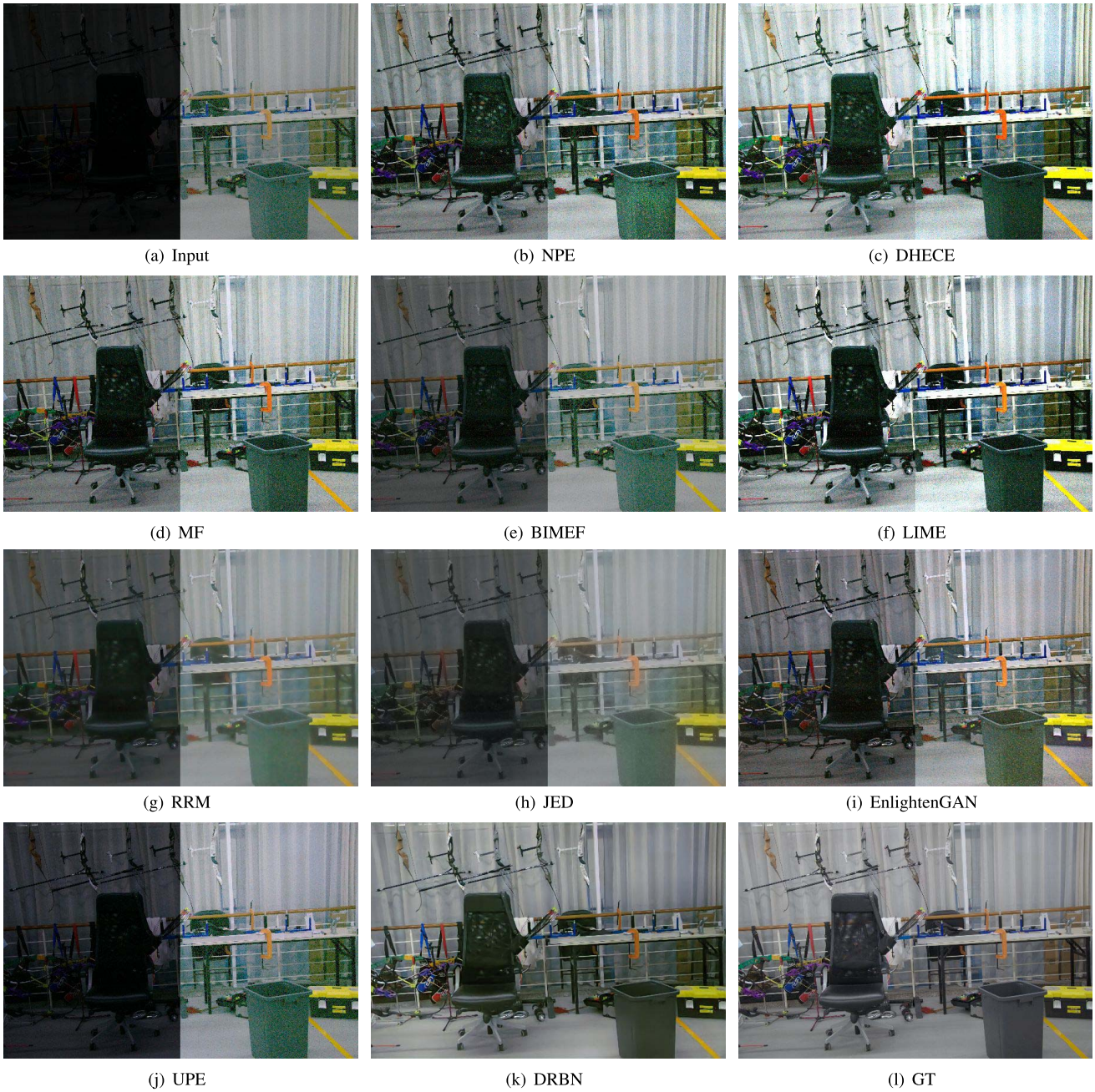
To comprehensively evaluate the proposed method, images of diverse content are adopted. In particular, the LOL dataset with actually acquired low and normal light images [34] are used for objective and subjective evaluations. The compared methods include Bio-Inspired Multi-Exposure Fusion (BIMEF) [43], Brightness Preserving Dynamic Histogram Equalization (BPDHE) [21], Camera Response Model (CRM) [44], Differential value Histogram Equalization Contrast Enhacement (DHECE) [24], Dong *et al.*'s method [45], Exposure Fusion Framework (EFF) [15], Contrast Limited Adaptive Histogram Equalization (CLAHE) [46], Low-Light Image Enhancement via Illumination Map Estimation (LIME) [8], Multiple Fusion (MF) [7], Multiscale Retinex (MR) [26] Joint Enhancement and Denoising Method (JED) [11], Refined Retinex Model (RRM) [10], Simultaneous Reflectance and Illumination Estimation (SRIE) [9], Deep Retinex Decomposition (DRD) [34], Deep Underexposed Photo Enhancement (DeepUPE) [16], Single Image Contrast Enhancer (SICE) [14], and EnlightenGAN [17]. In addition, we also provide the results of NPE [6] and DICM [23] for visual quality comparisons.

B. Implementation Details

There are two stages in training the network. In the first stage, we optimize the deep neural network with the Adam optimizer [47] and the learning rate is set to 0.0001. The cropped image size and batch size are set to be 256×256 and 4, respectively. Regarding the second stage training, the generator and discriminator are also trained with the Adam optimizer. The learning rates of the generator and discriminator are set to 0.0001 and 1e-6, respectively. We set the cropping image size and batch size for training to 320×320 and 1, respectively. The parameters of λ_1, λ_2, λ_3, and λ_4 are set to 0.1, 0.1, 0.01 and 1, respectively. Moreover, the LOL dataset [34] is split for training and testing. In the first stage, 300 epochs are allowed and by contrast 30 epochs are allowed in the second stage. The first module (BLN) is fixed in the second stage training. After 200 epochs in the first stage training, the learning rate drops by 0.5. The hyper-parameters of Adam optimizer are set with $\beta_1 = 0.9$, $\beta 2 = 0.999$ and $\epsilon = 1e-8$.

C. Evaluation Criteria

First, quantitative evaluations are performed for performance comparisons based on three objective evaluation metrics, including Peak Signal-to-Noise Ratio (PSNR), SSIM [41], and SSIM based on the corresponding Gamma corrected results (SSIM-GC). PSNR implies the *fidelity* of between normal light image and the enhanced image, and SSIM emphasizes more on image structures by taking the majority of quality degradation as a perceived change in *structural direction*. In the low-light enhancement task, the average illumination level is difficult to be predicted. Therefore, the degradation of details might not be well captured by PSNR and SSIM, motivating the introduction of SSIM-GC. In this metric, we first correct the global illumination via the Gamma transformation and then calculate the corresponding SSIM index to inspect the detail restoration capacity of different methods.



(a) Input (b) NPE (c) DHECE

(d) MF (e) BIMEF (f) LIME

(g) RRM (h) JED (i) EnlightenGAN

(j) UPE (k) DRBN (l) GT

Fig. 2. Visual quality comparisons of different methods. Left part: the original results. Right part: the results corrected by Gamma transformation for better visibility.

D. Quantitative Evaluations

In Table II, it is shown that our method is superior in the capacity of both structure recovery and illumination restoration based on the adopted metrics. Better results in terms of SSIM-GC also demonstrate our superiority when the global illumination is deducted. Moreover, the EnligtenGAN, SICE and CRM can also obtain superior PSNR values, indicating that they have already well restored the global illumination. The SSIM and SSIM-GC results of other methods imply their limitations on repairing structural details and stretching contrast, which has also been clearly confirmed in the latter qualitative evaluations.

E. Qualitative Evaluations

Extensive qualitative evaluations are conducted in Figs. 2 and 3. The results demonstrate that, compared to the state-of-the-art methods, our DRBN obtains superior performance in terms of both qualitative and quantitative evaluations. Generally speaking, most previous methods fail to well restore global structures and illumination. DHECE, NPE, EnlightenGAN, and LIME well estimate the global illumination. However, the burred noise is unintentionally amplified in their results. RRM, JED, MF, and UPE suffer from poor visibility and under-exposure. In their Gamma corrected results (right parts), we observe noise in the results

(a) Input (b) NPE (c) DHECE

(d) MF (e) BIMEF (f) LIME

(g) RRM (h) JED (i) EnlightenGAN

(j) UPE (k) DRBN (l) GT

Fig. 3. Visual quality comparisons of different methods. Left part: the original results. Right part: the results corrected by Gamma transformation for better visibility.

of JED and RRM. However, for these two methods the quality in terms of contrast is not satisfactory. Comparatively, our method obtains appealing visual quality, visually pleasing color distribution and illumination, as well as sharp and clean details.

F. Ablation Study

An ablation study is further conducted based on our two-stage design in Figs. 4 and 5. From Fig. 4, comparing the results before and after the band recomposition, more colorful results and better boosted contrast can be obtained. Benefiting from the quality prior for guidance, the quality as well as overall visibility even outperform the ground-truths (bottom rows). From Fig. 5, it is observed that our method is better at handling under-exposed and over-exposed regions, such as the light spot in Fig. 5 (a), the screen border in Fig. 5 (b), the face areas in Fig. 5 (c), and the sky region in Fig. 5 (d).

G. Visualization of Recursive Band Learning and Band Recomposition

The learned band representations and the weighting masks for band recomposition are visualized in Figs. 6 and 7, respectively. It is obvious that the proposed RBL effectively extracts a series of coarse-to-fine layered representations. Subsequently, the band signals could be adaptively reconstructed with band recomposition. By comparing the weighting maps from different bands, it can also be observed that the

Fig. 4. Ablation study for the two-stage design. The top row: low-light images. The second row: the results without perceptual guidance. The third row: the results with perceptual guidance. The bottom row: normal light images.

higher-order weighting maps are sparser with more focus on edges and structures.

H. Ablation Study for Recursive Band Learning

We perform the ablation studies for Recursive Band Learning quantitatively. In particular, five alternative versions are compared.

- *Version 1* (*RBL-v1*) does not include the recursive processing (only once prediction) and multi-scale loss constraint, by replacing Eqn. (4) as follows,

$$L_{\text{Rect}} = -\phi\left(\hat{x}_{s_1}^T, x\right). \tag{10}$$

- *Version 2* (*RBL-v2*) additionally includes the multi-scale loss constraint.
- *Version 3* (*RBL-v3*) includes the recursive processing but without the feature bypass connection and the recursive input, by replacing Eqn. (3) as follows,

$$\begin{aligned}\left[\Delta f_{s_1}^t,\ \Delta f_{s_2}^t, \Delta f_{s_3}^t\right] &= F_{\text{BLN\_F}}^t\left(\hat{x}_{s_1}^{t-1}\right),\\ f_{s_i}^t &= \Delta f_{s_i}^t, i = 1, 2, 3,\\ \hat{x}_{s_3}^t &= F_{\text{R\_}s_3}^t\left(f_{s_3}^t\right),\\ \hat{x}_{s_2}^t &= F_{\text{R\_}s_2}^t\left(f_{s_2}^t\right) + F_{\text{U}}\left(\hat{x}_{s_3}^t\right),\\ \hat{x}_{s_1}^t &= F_{\text{R\_}s_1}^t\left(f_{s_1}^t\right) + F_{\text{U}}\left(\hat{x}_{s_2}^t\right).\end{aligned} \tag{11}$$

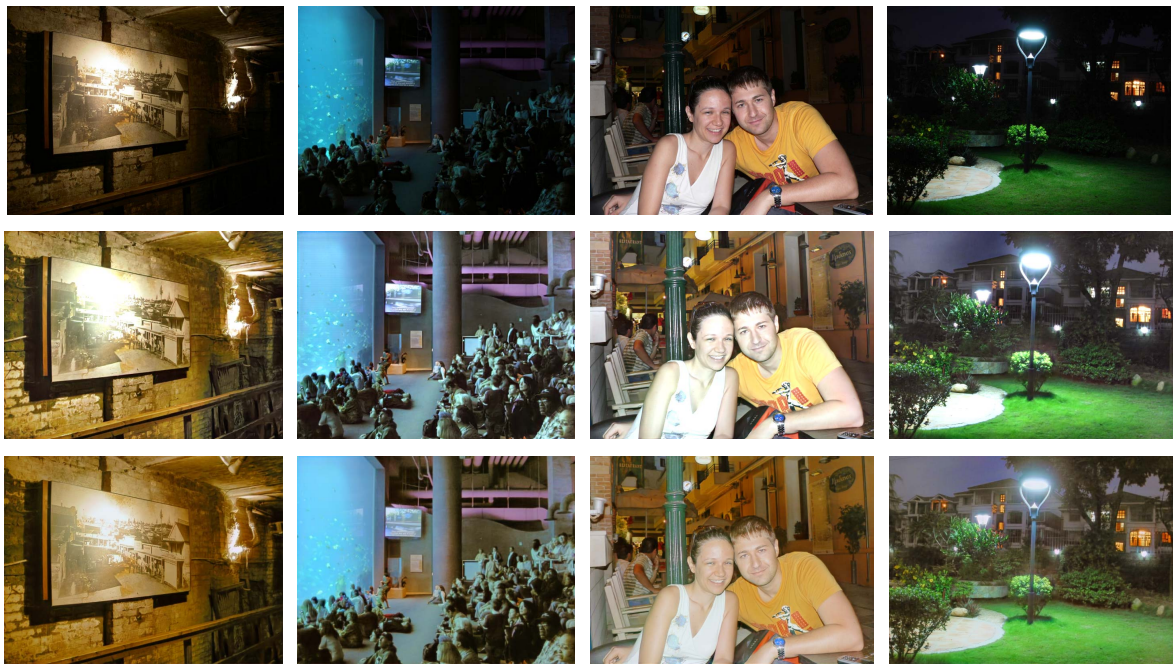
- *Version 4* (*RBL-v4*) includes the recursive processing and the recursive input without the feature bypass connection, by replacing Eqn. (3) as follows,

$$\begin{aligned}\left[\Delta f_{s_1}^t,\ \Delta f_{s_2}^t, \Delta f_{s_3}^t\right] &= F_{\text{BLN\_F}}^t\left(y, \hat{x}_{s_1}^{t-1}\right),\\ f_{s_i}^t &= \Delta f_{s_i}^t, i = 1, 2, 3,\\ \hat{x}_{s_3}^t &= F_{\text{R\_}s_3}^t\left(f_{s_3}^t\right),\\ \hat{x}_{s_2}^t &= F_{\text{R\_}s_2}^t\left(f_{s_2}^t\right) + F_{\text{U}}\left(\hat{x}_{s_3}^t\right),\\ \hat{x}_{s_1}^t &= F_{\text{R\_}s_1}^t\left(f_{s_1}^t\right) + F_{\text{U}}\left(\hat{x}_{s_2}^t\right).\end{aligned} \tag{12}$$

- *Version 5* (*RBL-v5*) includes the recursive processing, recursive input and feature bypass connection without introducing the feature augmentation by LSTMs:

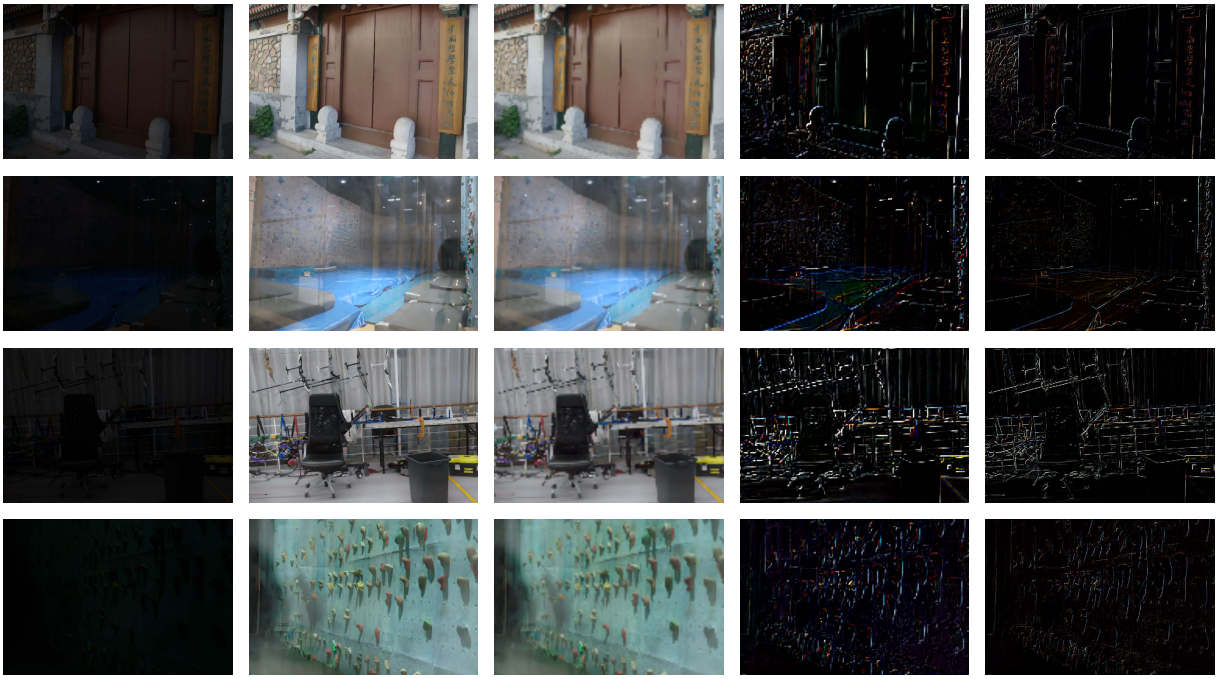
$$\begin{aligned}\left[\Delta f_{s_1}^t, \Delta f_{s_2}^t, \Delta f_{s_3}^t\right] &= F_{\text{BLN\_F}}^t\left(y, \hat{x}_{s_3}^{t-1}\right),\\ f_{s_i}^t &= \Delta f_{s_i}^t + f_{s_i}^{t-1}, i = 1, 2, 3,\\ \hat{x}_{s_1}^t &= F_{\text{R\_}s_1}^t\left(f_{s_1}^t\right),\\ \hat{x}_{s_2}^t &= F_{\text{R\_}s_2}^t\left(f_{s_2}^t\right) + F_{\text{U}}\left(\hat{x}_{s_1}^t\right),\\ \hat{x}_{s_3}^t &= F_{\text{R\_}s_3}^t\left(f_{s_3}^t\right) + F_{\text{U}}\left(\hat{x}_{s_2}^t\right).\end{aligned} \tag{13}$$

- *Version 6* (*RBL-v6*) is the full version and includes the recursive processing, recursive input, feature bypass connection and the feature augmentation by LSTMs.



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

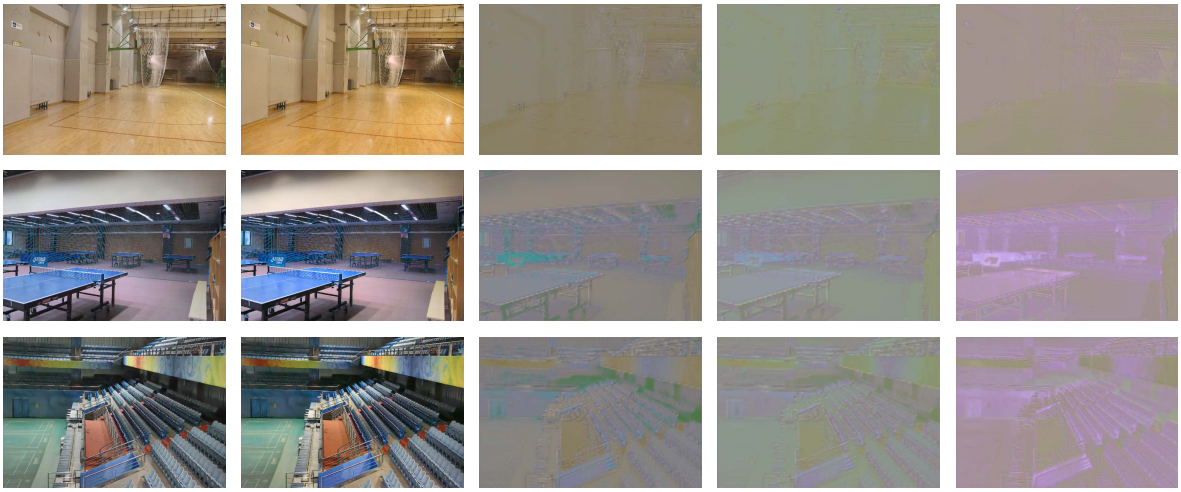
Fig. 5. Ablation study for the two-stage design. The top row: low-light images. The second row: the results without perceptual guidance. The bottom row: the results with perceptual guidance.



(a) Low-Light Image (b) Enhanced Results by RBL (c) 1st order band (d) 2nd order band (e) 3rd order band

Fig. 6. Visualization of the learned bands by our DRBN. RBL denotes the recursive band learning.

To better reflect the average performance of the six versions, their average performance is calculated in the last 50 epochs (250-300 epochs). The results are showed in Table III. Comparing RBL-V1 to RBL-V2, it is clearly observed that the multi-scale loss results in the performance gains in PSNR and SSIM. Comparing RBL-V3 and RBL-V4 to RBL-V2, it is observed that, the recursive structure might not achieve an improved performance, since RBL-V3 and RBL-V4 lead to performance drops in terms of PSNR and SSIM, respectively. With both the bypass connection and recursive input, RBL-V5 acquires performance gains in terms of both PSNR and SSIM. With the additional feature augmentation by LSTMs, RBL-V6 achieves further gains and the best results in PSNR and SSIM, which demonstrates the effectiveness and rationality of our RBL.



(a) Enhanced Results by RBL (b) Enhanced Results by BR (c) 1st order weighting map (d) 2nd order weighting map (e) 3rd order weighting map

Fig. 7. Visualization of the learned weighting maps for band recomposition. RBL and BR denote the recursive band learning and band recomposition, respectively.

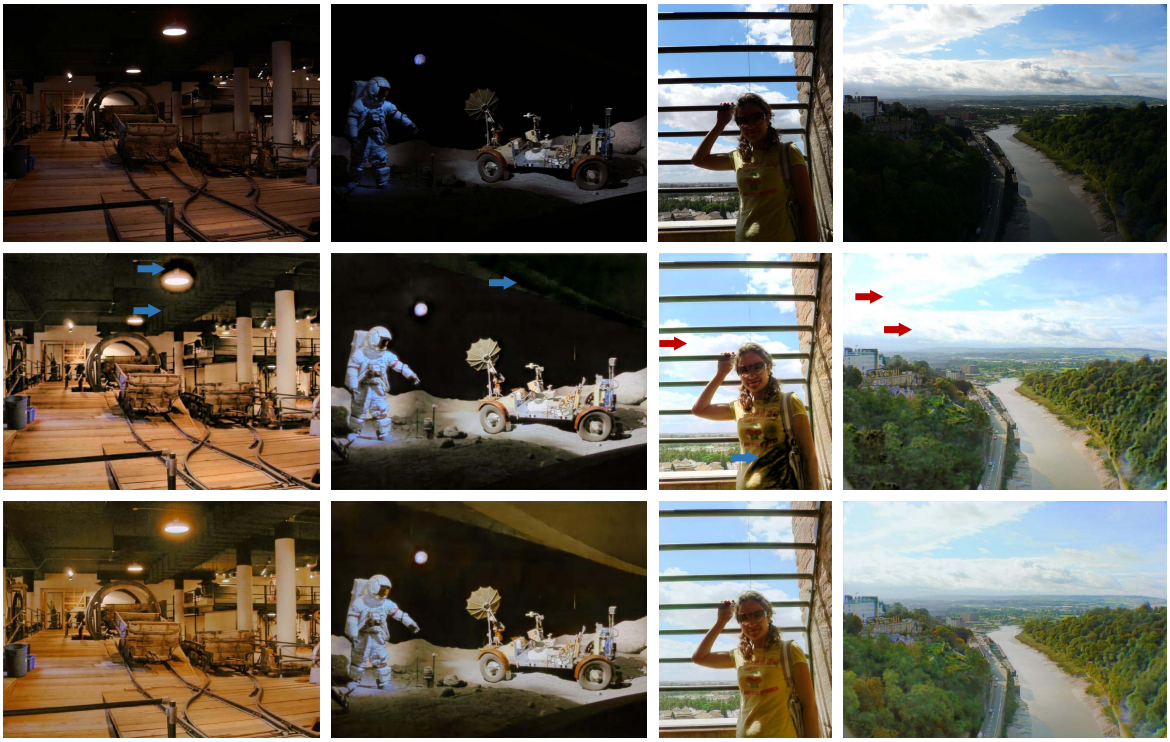


Fig. 8. Visual quality comparisons between quality-aware adversarial learning [19] and the proposed quality-guidance with assessment network. First row: low-light images. Middle row: images obtained by quality-aware adversarial learning. Bottom row: images obtained by quality-guidance via assessment network. Blue arrows: under-exposed details. Red arrows: over-exposed regions.

TABLE III

THE ABLATION STUDY FOR RECURSIVE BAND LEARNING (RBL)

Metrics	RBL-v1	RBL-v2	RBL-v3	RBL-v4	RBL-v5	RBL-v6
PSNR	18.52	19.07	19.33	18.51	19.71	**19.92**
SSIM	0.8638	0.8679	0.8413	0.8744	0.8853	**0.8880**

TABLE IV

THE ABLATION STUDY OF DARK AND BRIGHT CHANNEL PRIOR

Metrics	DRBN-	DRBN
PSNR	20.13	**20.29**
SSIM	0.8295	**0.8310**

I. Quantitative Evaluation Between DRBN [19] and DRBN in this Work

We also compare the performance of preliminary conference version [19] and the proposed scheme on *LOL-Real* in Table IV. It is obvious that the performance has been improved with very close computational complexity.

J. Visual Quality Comparison Between Quality-Aware Adversarial Learning in [19] and the Proposed Quality-Guidance with Assessment Network

We also compare the visual quality between quality-aware adversarial learning [19] and quality-guidance with assessment network. As shown in Fig. 8, it is observed that

compared to quality-aware adversarial learning, the result of the enhancement network with quality-guidance via the assessment network is more visually promising, containing less under-exposed regions (blue arrows) and over-exposed details (red arrows).

V. Conclusion

In this paper, we have proposed a novel semi-supervised learning based low-light image enhancement method which absorbs the advantages of both synthetic paired low/normal-light enhancement for fidelity recovery and unpaired high-quality data for quality enhancement. To this end, we develop a two-stage network that first restores the signal based on fidelity and further enhances the quality to improve overall viewing experience. The two-stage design enables to obtain enhanced results with well-reconstructed structural details and visually pleasing contrast and color distributions. Qualitative and quantitative validations have provided useful evidence regarding the superior performance of the proposed approach.

References

[1] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Müller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Visualizat. Biomed. Comput.*, May 1990, pp. 337–345.

[2] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, May 2007.

[3] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–128, Dec. 1977.

[4] D. J. Jobson, Z. 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.

[5] D. J. Jobson, Z. 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.

[6] 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.

[7] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Process.*, vol. 129, pp. 82–96, Dec. 2016.

[8] 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.

[9] 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. (CVPR)*, Jun. 2016, pp. 2782–2790.

[10] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust Retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.

[11] X. Ren, M. Li, W.-H. Cheng, and J. Liu, "Joint enhancement and denoising method via sequential decomposition," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2018, pp. 1–5.

[12] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.

[13] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.

[14] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.

[15] W. Ren *et al.*, "Low-light image enhancement via a deep hybrid network," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4364–4375, Sep. 2019.

[16] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6849–6857.

[17] Y. Jiang *et al.*, "EnlightenGAN: Deep light enhancement without paired supervision," 2019, *arXiv:1906.06972*. [Online]. Available: http://arxiv.org/abs/1906.06972

[18] Z. Wang and A. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[19] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3063–3072.

[20] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2408–2415.

[21] H. Ibrahim and N. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.

[22] 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.

[23] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5372–5384, Dec. 2013.

[24] K. Nakai, Y. Hoshi, and A. Taguchi, "Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Nov. 2013, pp. 445–449.

[25] J.-T. Lee, C. Lee, J.-Y. Sim, and C.-S. Kim, "Depth-guided adaptive contrast enhancement using 2D histograms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4527–4531.

[26] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new image contrast enhancement algorithm using exposure fusion framework," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, Oct. 2017, pp. 36–46.

[27] X. Wu, X. Liu, K. Hiramatsu, and K. Kashino, "Contrast-accumulated histogram equalization for image enhancement," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3190–3194.

[28] L. Li, R. Wang, W. Wang, and W. Gao, "A low-light image enhancement method for both denoising and contrast enlarging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 3730–3734.

[29] X. Zhang, P. Shen, L. Luo, L. Zhang, and J. Song, "Enhancement and noise reduction of very low light level images," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 2034–2037.

[30] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[31] C.-H. Lee, J.-L. Shih, C.-C. Lien, and C.-C. Han, "Adaptive multiscale Retinex for image contrast enhancement," in *Proc. Int. Conf. Signal-Image Technol. Internet-Based Syst.*, Dec. 2013, pp. 43–50.

[32] L. Wang, L. Xiao, H. Liu, and Z. Wei, "Variational Bayesian method for Retinex," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3381–3396, Aug. 2014.

[33] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma, "MSR-Net: Low-light image enhancement using deep convolutional network," 2017, *arXiv:1711.02488*. [Online]. Available: http://arxiv.org/abs/1711.02488

[34] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2018, pp. 1–12.

[35] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie, "LLCNN: A convolutional neural network for low-light image enhancement," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[36] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using CNNs," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2018, pp. 1–13.

[37] Y. Wang *et al.*, "Progressive Retinex: Mutually reinforced illumination-noise perception network for low light image enhancement," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2015–2023.

[38] G. Kim, D. Kwon, and J. Kwon, "Low-lightgan: Low-light enhancement via advanced generative adversarial network with task-driven training," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2811–2815.

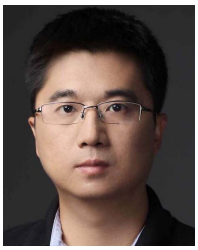
[39] W. Yang *et al.*, "Advancing image understanding in poor visibility environments: A collective benchmark study," *IEEE Trans. Image Process.*, vol. 29, pp. 5737–5752, 2020.

[40] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Comput. Vis. Image Understand.*, vol. 178, pp. 30–42, Jan. 2019.
[41] 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.
[42] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.
[43] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," 2017, *arXiv:1711.00591*. [Online]. Available: http://arxiv.org/abs/1711.00591
[44] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new low-light image enhancement algorithm using camera response model," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 3015–3022.
[45] X. Dong *et al.*, "Fast efficient algorithm for enhancement of low lighting video," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2011, pp. 1–6.
[46] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*. San Diego, CA, USA: Academic, 1994, pp. 474–485.
[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.
[48] K. Xu, X. Yang, B. Yin, and R. W. H. Lau, "Learning to restore low-light images via decomposition- and-enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 2281–2290.
[49] C. Guo *et al.*, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 1777–1786.
[50] D. Kwon, G. Kim, and J. Kwon, "DALE: Dark region-aware low-light image enhancement," in *Proc. Brit. Mach. Vis. Conf.*, 2020, pp. 1–12.

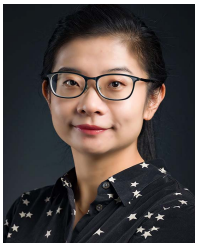
Wenhan Yang (Member, IEEE) received the B.S. and Ph.D. degrees (Hons.) in computer science from Peking University, Beijing, China, in 2012 and 2018. He is currently a Postdoctoral Research Fellow with the Department of Computer Science, City University of Hong Kong. He has authored over 80 technical articles in refereed journals and proceedings, and holds nine granted patents. His current research interests include image/video processing/restoration, bad weather restoration, human–machine collaborative coding. He received the IEEE ICME-2020 Best Paper Award, the IEEE CVPR-2018 UG2 Challenge First Runner-up Award, and the CSIG Best Doctoral Dissertation Award, in 2019. He served as the Area Chair for IEEE ICME-2021 and the Organizer for IEEE CVPR-2019/2020/2021 UG2+ Challenge and Workshop.

Shiqi Wang (Member, IEEE) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008 and the Ph.D. degree in computer application technology from Peking University in 2014. He was a Postdoctoral Fellow with the University of Waterloo, Canada, and the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. He has proposed over 40 technical proposals to ISO/MPEG, ITU-T, and AVS standards. He has authored/coauthored more than 200 refereed journal/conference articles. His research interests include video compression, quality assessment, and analysis. He received the Best Paper Award from IEEE ICME 2019, VCIP 2019, IEEE Multimedia 2018, and PCM 2017. He is the coauthor of an article that received the Best Student Paper Award in the IEEE ICIP 2018.

Yuming Fang (Senior Member, IEEE) received the B.E. degree from Sichuan University, Chengdu, China, the M.S. degree from the Beijing University of Technology, Beijing, China, and the Ph.D. degree from Nanyang Technological University, Singapore. He is currently a Professor with the School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include visual quality assessment, computer vision, and 3D image/video processing. He serves as an Associate Editor for IEEE ACCESS. He is on the Editorial Board of *Signal Processing: Image Communication*.



Yue Wang received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2006, and the Ph.D. degree from the Graduate University of the Chinese Academy of Sciences, Beijing, in 2012. He was a Postdoctoral Researcher with the Department of Computer Science, Peking University, Beijing, from 2015 and 2017. He has been the Director of the Multimedia Foundation Team with ByteDance Ltd., since 2018. His current research interests include image and video coding, processing, and transmission technology.



Jiaying Liu (Senior Member, IEEE) received the Ph.D. degree (Hons.) in computer science from Peking University, Beijing China, in 2010. She was a Visiting Scholar with the University of Southern California, Los Angeles, CA, USA, from 2007 to 2008. She was a Visiting Researcher with the Microsoft Research Asia in 2015 supported by the Star Track Young Faculties Award. She is currently an Associate Professor and a Peking University Boya Young Fellow with the Wangxuan Institute of Computer Technology, Peking University. She has authored over 100 technical articles in refereed journals and proceedings, and holds 50 granted patents. Her current research interests include multimedia signal processing, compression, and computer vision. She is a Senior Member of CSIG and CCF. She has served as a member for the Multimedia Systems and Applications Technical Committee (MSA TC), and the Visual Signal Processing and Communications Technical Committee (VSPC TC) in IEEE Circuits and Systems Society. She received the IEEE ICME-2020 Best Paper Award and the IEEE MMSP-2015 Top10% Paper Award. She has also served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUIT SYSTEM FOR VIDEO TECHNOLOGY, and *Journal of Visual Communication and Image Representation* (Elsevier), the Technical Program Chair of IEEE ICME-2021/ACM ICMR-2021, the Publicity Chair of IEEE ICME-2020/ICIP-2019, and the Area Chair of CVPR-2021/ECCV-2020/ICCV-2019. She was the APSIPA Distinguished Lecturer from 2016 to 2017.