From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement

Wenhan Yang[1], Shiqi Wang[1]*, Yuming Fang[2], Yue Wang[3], Jiaying Liu[4]
[1] City University of Hong Kong   [2] Jiangxi University of Finance and Economics
[3] ByteDance Technology Co., Ltd.   [4] Peking University

Abstract

Under-exposure introduces a series of visual degradation, i.e. decreased visibility, intensive noise, and biased color, etc. To address these problems, we propose a novel semi-supervised learning approach for low-light image enhancement. A deep recursive band network (DRBN) is proposed to recover a linear band representation of an enhanced normal-light image with paired low/normal-light images, and then obtain an improved one by recomposing the given bands via another learnable linear transformation based on a perceptual quality-driven adversarial learning with unpaired data. The architecture is powerful and flexible to have the merit of training with both paired and unpaired data. On one hand, the proposed network is well designed to extract a series of coarse-to-fine band representations, whose estimations are mutually beneficial in a recursive process. On the other hand, the extracted band representation of the enhanced image in the first stage of DRBN (recursive band learning) bridges the gap between the restoration knowledge of paired data and the perceptual quality preference to real high-quality images. Its second stage (band recomposition) learns to recompose the band representation towards fitting perceptual properties of high-quality images via adversarial learning. With the help of this two-stage design, our approach generates the enhanced results with well reconstructed details and visually promising contrast and color distributions. Extensive evaluations demonstrate the superiority of our DRBN.

1. Introduction

In low-light conditions, a series of visual degradations, *i.e.* low visibility, low contrast and intensive noises appear

(a) Input

(b) SICE [2]

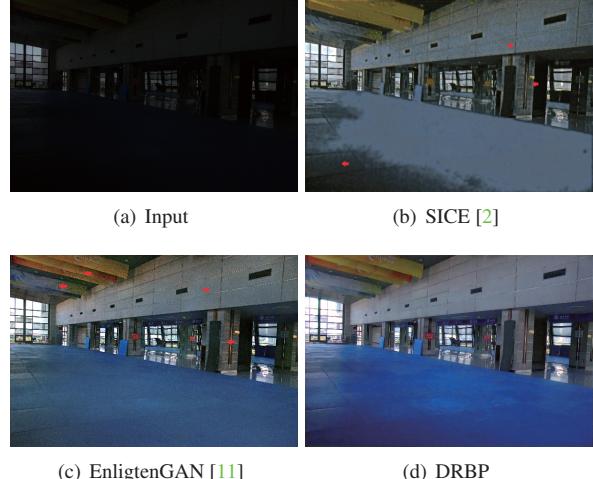(c) EnligtenGAN [11]

(d) DRBP

Figure 1. The visual results of different methods on a real low-light image in LOL [30]. SICE [2] is a state-of-the-art fully supervised method and EnlightenGAN [11] is an unsupervised method without paired supervision. Our DRBN well restores the global illumination, suppresses noise and preserves structural details.

in captured images. It will be beneficial to use more advanced shooting devices and specialized photographic techniques to alleviate some degradations. However, it is still difficult to totally avoid the presence of noise even with upgraded shooting devices. There is insufficient light reaching camera sensors, causing the scene signals buried by system noise. It would be helpful if longer exposure time is taken to suppress noise, which however introduces blurriness. Therefore, the low-light enhancement methods at the software end are expected. It aims to restore an image captured in the low-light condition to a normal one, where visibility, contrast, and noise are expected to be improved, stretched, and suppressed, respectively. The enhancement process leads to visual quality improvement and offers a good starting point for high-level computer vision tasks (*i.e.* object detection and recognition).

It is non-trivial to enhance low-light images, because the noise is usually easy to be amplified. In the past decades,

many researchers have tried to address this problem. Histogram equalization (HE) methods [23, 1] enhance the low-light images by stretching the dynamic range, providing undesirable illumination with unintentionally amplified intensive noise. Retinex theory-based methods [15] decompose and process two layers of an image, reflectance and illumination layers, respectively. Several filters [14, 12, 28] are designed for that decomposition. Many works [7, 9, 8, 18, 25] impose priors on the decomposed illumination and reflectance.

Recently, the prosperity of deep learning also inspires the development of new deep-learning based approaches for low-light enhancement. These methods [20, 4, 2, 24, 27] learn to remove the related composite degradation based on the well-prepared paired low/normal-light images and elaborately designed models. However, the existing loss functions are not well aligned to human perception and do not capture the intrinsic signal structure of an image, leading to unsatisfactory visual results, *e.g.* biased color distribution and residual noise. Recently, EnlightenGAN [11] is created without paired supervision, where the dataset with only low/normal-light images (unnecessarily paired) are required. This method proves the feasibility to learn with unpaired data for low-light enhancement. However, without the paired supervision, fine details cannot be restored, and the intensive noise is still present in enhanced results.

In general, the aforementioned popular deep-learning based methods can be divided into two categories. The *fully supervised methods* are trained *with paired supervision*, where the ground truth guidance is provided for detail signal modeling in the training phase. Therefore, the networks are more capable of suppressing noise and impairing details. In *unsupervised methods*, the paired low/normal-light images are unavailable. The knowledge on the enhancement mapping is extracted from *unpaired low/normal-light image sets*. It is easy to make this kind of dataset large-scale with diversified content, therefore the methods can learn to restore illumination, color, and contrast more adaptively.

Comprehensively considering the strengths, weaknesses, and potentials of existing methods, in this paper, we explore to construct a unified architecture to have two kinds of merits. Specifically, a novel semi-supervised learning framework is constructed for low-light image enhancement. A deep recursive band network (DRBN) is proposed to offer a band representation to connect the signal fidelity prior obtained from paired supervision and the perceptual visual quality prior extracted by an unpaired high-quality dataset, where images are selected by mean opinion scores. In the first stage of DRBN, the network is trained on the paired low/normal-light images. A linear band representation is first recovered via training with paired low/normal-light images, whose estimations are mutually beneficial in a recursive process. After that, the extracted band repre-

sentation of the enhanced image in the first stage of DRB-N (**recursive band learning**) is used to bridge the gap between the restoration knowledge of paired data and the perceptual quality provided by a high-quality image dataset. Then, in the second stage of DRBN (**band recomposition**), it learns to recompose the band representation towards fitting visual properties of high-quality images via adversarial learning. With the help of this two-stage design, our approach generates the enhanced results with well reconstructed details, and visually promising contrast and color distributions. Our contributions are summarized as follows:

- To the best of our knowledge, we make the first attempt to propose a *semi-supervised learning* framework for low-light image enhancement, where a deep recursive band representation is designed to connect fully-supervised and un-supervised frameworks to integrate their superiorities.

- The proposed framework is well designed to extract 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, capable of removing noise and correcting details.

- The deep band representations are recomposed under the perceptual guidance of a quality-guided adversarial learning. The "real images" for the discriminator are selected perceptually based on mean opinion score (MOS). As far as we know, this is also the first trial in low-light image enhancement tasks.

2. Related Work

The earliest low-light enhancement methods adjust the illumination uniformly, which easily causes over-exposure and under-exposure, such as Histogram equalization (HE) [23, 1]. Without local adaptation, the enhancement leads to undesirable illumination and intensive noise. Some methods [17, 35] enhance the visibility by applying dehazing methods to the inverted low-light images. In these methods, the off-line denoising operation [5] is applied to suppress noise, which sometimes also leads to detail blurriness.

Later on, Retinex-based methods [15] perform the joint illumination adjustment and noise suppression via decomposing the image into illumination and reflectance layers and adjusting them adaptively. Various priors, *e.g.* structure aware prior [9], weighted variation [8], and multiple derivatives of illumination [7] are utilized to guide manipulation of these two layers. Variants of Retinex models, *e.g.* single-scale Retinex [14], multi-scale Retinex [12], naturalness Retinex [28], and robust Retinex [18, 25] are developed to facilitate low-light image enhancement. These
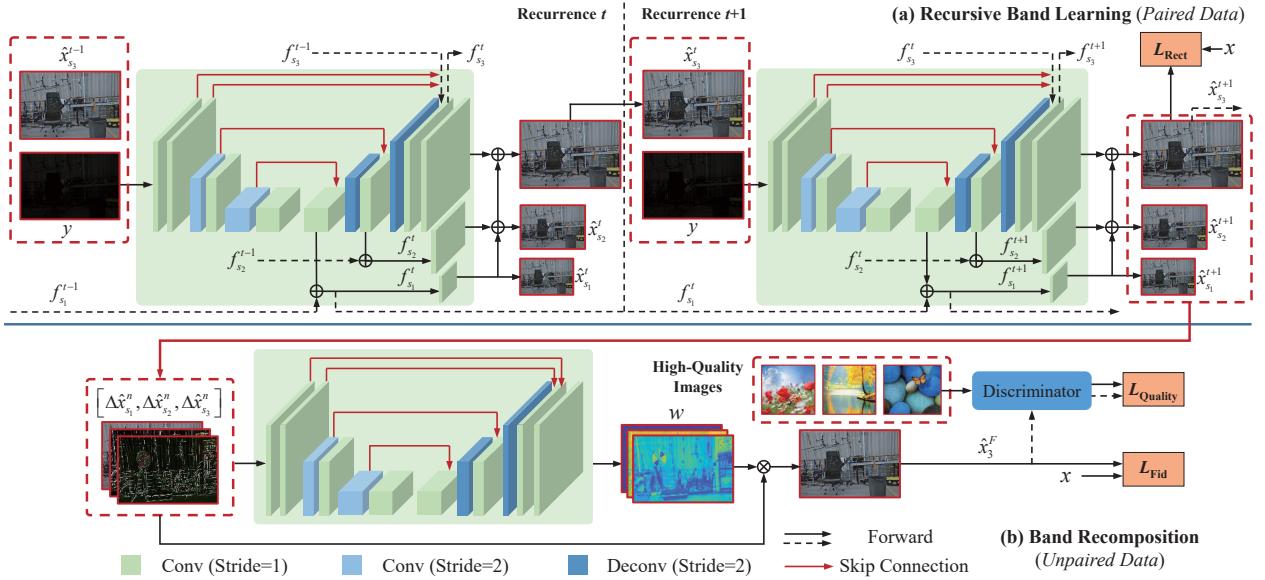
Figure 2. The framework of the proposed Deep Recursive Band Network (DRBN), which consists of two stages: recursive band learning and band recomposition. (1) In the first stage, a coarse-to-fine band representation is learned and different band signals are inferred jointly in a recursive process. The enhanced result from the last recurrence is used as the guidance of the next recurrence and the later recurrence is only responsible to learn the residue in the feature and image domains at different scales. (2) In the second stage, the band representation is recomposed to improve perceptual quality of the enhanced low-light image via a perceptual quality-guided adversarial learning.

methods show impressive results in illumination adjustment and small noise removal. However, with only hand-crafted constraints, these methods are not adaptive enough, and their results present intensive noises and sometimes under-exposed and over-exposed local details.

In recent years, deep-learning based low-light image enhancement brings in impressive performance gains. Lore *et al.* [20] made the first attempt by proposing a deep auto-encoder named Low-Light Net (LLNet) for contrast enhancement and noise removal. Later on, various methods based on different network designs [26, 3, 24, 27, 30] are proposed. These methods are trained on the paired dataset, and their enhancement performance is largely dependent on the dataset. Because the synthetic data cannot fully characterize degradation in real scenarios and real captured paired data includes limited kinds of scenarios, the results of these methods are still imperfect, especially failing to handle intensive noise. There are also works on deep-learning based image enhancement from raw images [4], or the joint task of low-light image enhancement and high-level computer vision tasks, such as face detection [34], object detection [19], *etc*. In our work, we focus on perceptual quality improvement of low-light images in RGB format. Different from previous related studies, we develop a semi-supervised framework, where the useful knowledge from the paired and unpaired datasets is used jointly to provide perceptual guidance for both detail signal modeling and global illumination, color and contrast recovery.

3. Deep Recursive Band Network for Semi-Supervised Low-Light Enhancement

In this section, we illustrate our motivations at first. After that, we convert the related methodologies into a deep recursive band network (DRBN) for low-light image enhancement.

3.1. Motivation: Bridging the Gap Between Signal Fidelity and Perceptual Quality

Recursive Band Learning. The paired training data provides strong signal fidelity constraint to correct detailed signals. Therefore, we first perform a recursive band learning to restore detail signals based on the guidance of the paired data. In this process, besides the enhanced image $\hat{x}$, a series of band representations $\left\{ \Delta\hat{x}_{s_1}^T, \Delta\hat{x}_{s_2}^T, ..., \Delta\hat{x}_{s_n}^T \right\}$ are also generated progressively from y, where $\hat{x} = \sum_{i=1}^{n} \hat{x}_{s_i}^T$. s_i denotes the order of the band representation. $\Delta\hat{x}_{s_i}^T$ is learned by fully-supervised learning on paired low/normal-light data. The high-order band $\hat{x}_{s_i}^T$ depends on the low-order one $\hat{x}_{s_{i-1}}^T$.

Connecting Recursive Band Representation and Adversarial Learning. However, with the signal fidelity constraint in the first stage, the good visual quality cannot naturally achieved. Therefore, inspired by the recent image enhancement methods based on the unpaired dataset, in the second stage, we recompose the learned band representations in the first stage to obtain the results with better per-

ceptual quality from the view of human perception as follows,

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

where $w_i(\cdot)$ is the learned weighting parameter for recomposition. It recomposes band signals of an enhanced image, which has been almost noise-free with well-reconstructed details, to a new one with more superior illumination, contrast, and color distributions, namely better perceptual quality from the perspective of human vision.

3.2. Deep Recursive Band Network

Architecture Overview. As shown in Fig. 2, with the above-mentioned motivations in mind, our network consists of two parts: recursive band learning (relying on paired data), and band recomposition (relying on unpaired data). In the recursive band learning part, DRBN is constructed to recover a normal-light image based on the low-light input in a recursive way. The intermediate estimation, the output of the previous recurrence, is used as the guidance input of the next recurrence, which connects all band estimations together in a joint estimation. Our DRBN adopts the residual learning in both feature and image domains. That is to say, the latter recurrence only serves to estimate the residual features and images to acquire better estimations. Therefore, the later recurrence is more capable to model structural details and suppress noise. In each recurrence, a series of coarse-to-fine band representations are extracted and then merged into the enhancement results. This band representation provides an effective tool to combine the power of both the enhancement knowledge learned from the paired images and the data prior of high-quality images. In the band recomposition part, this band representation is feed-forwarded into another network to produce a set of transform coefficients that manipulate and fuse these bands linearly. An adversarial loss (named as quality loss), judging whether an image is perceptually high-quality or not, is used as a constraint. A set of high-quality images, selected by mean opinion scores, are used to serve as the priors of human vision perception. In this way, the overall good results from the perspective of both signal fidelity and human perceptual quality are achieved.

Recursive Band Learning. We first aim to fully exploit the power of paired data learning to recover each band signal of an enhanced image. A series of U-Net like deep networks, called band learning networks (BLN), are built as shown in Fig. 2. Each BLN projects the input, *i.e.* the concatenation of y and the enhanced result of the last recurrence $\hat{x}_{s_3}^{t-1}$, into the feature space and then transforms the features by several convolutional layers. In intermediate layers, the spatial resolutions of features are first down-sampled and then up-sampled via stride convolutions and deconvolutions. There

are skip connections (denoted by red) to connect the features with the same spatial resolution from shallow layers to deep ones, which helps local information contained in the features generated by shallow layers to reach the output. Each BLN produces three features at the scales $s_1 = 1/4$, $s_2 = 1/2$ and $s_3 = 1$, respectively.

For the convenience, we illustrate the first recurrence of the recursive learning:

$$\begin{aligned} \left[f_{s_1}^1, f_{s_2}^1, f_{s_3}^1 \right] &= F_{\text{BLN\_F}}^1 (y), \\ \hat{x}_{s_1}^1 &= F_{\text{R\_}s_1}^1 \left(f_{s_1}^1 \right), \\ \hat{x}_{s_2}^1 &= F_{\text{R\_}s_2}^1 \left(f_{s_2}^1 \right) + F_{\text{U}} \left(\hat{x}_{s_1}^1 \right), \\ \hat{x}_{s_3}^1 &= F_{\text{R\_}s_3}^1 \left(f_{s_3}^1 \right) + F_{\text{U}} \left(\hat{x}_{s_2}^1 \right), \end{aligned} \quad (2)$$

where $f_{s_1}^1, f_{s_2}^1, f_{s_3}^1$ are features extracted from y at their corresponding scales, respectively; $F_{\text{BLN\_F}}^1(\cdot)$ is the related process; $F_{\text{R\_}s_1}^1(\cdot)$, $F_{\text{R\_}s_2}^1(\cdot)$ and $F_{\text{R\_}s_3}^1(\cdot)$ are the processes of projecting the features back to the image domains at the corresponding scales; $F_{\text{U}}(\cdot)$ is the up-sampling process. The image is firstly reconstructed at the roughest scale s_1. Then, at fine scales, the residual signals are predicted to be parts of the whole result.

After that, at the t-th recurrence, only the residual features and images are learned with the guidance of previous estimated results. The concatenation of y and previously estimated result $\hat{x}_{s_3}^{t-1}$ is regarded as the input:

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

This formulation connects all band features closely, which forms a joint optimization of all bands. At the final recurrence T (set to 4 in our work), the reconstruction loss is applied as follows,

$$\begin{aligned} L_{\text{Rect}} = - \big(&\phi \left(\hat{x}_{s_3}^T, x \right) + \lambda_1 \phi \left(\hat{x}_{s_2}^T, F_D \left(x, s_2 \right) \right) \\ &+ \lambda_2 \phi \left(\hat{x}_{s_1}^T, F_D \left(x, s_1 \right) \right) \big), \end{aligned} \quad (4)$$

where $F_D(\cdot)$ is the down-sampling process, given the scaling factor s_i; $\phi(\cdot)$ calculates SSIM values of the input images; λ_1 and λ_2 are weighting parameters.

In this way, our recursive band learning has the following advantages:

- The high-order band inferred from the last recurrence will have influence on the inference of the low-order one in this recurrence. Therefore, the connection between the low and high-order bands are bi-directional, and the high-order bands also provide useful guidance to recover low-order bands.

- The recursive estimation enables different bands to learn to correct their estimations based on previous estimations of all bands.

- The recursive learning enhances the modeling capacities. The later recurrence only needs to recover the residue signals, guided by the estimation from previous recurrences. Therefore, accurate estimations can be obtained and fine details are paid attention to.

Band Recomposition. With the power of paired data, the band recovery process from the low-light images to the normal-light ones can be well learned, with well restored details and suppressed noise. Because the signal fidelity is not alway well aligned to human visual perception, especially for some global properties of images, such as light, color distribution, *etc*. Therefore, we further make our model learn to recompose the restored band signals with the perceptual guidance of a high-quality image dataset via perceptual quality-guided adversarial learning. The high-quality images selected from aesthetic visual analysis dataset [21] based on the MOS values are used to represent the prior knowledge of human perception. Another U-like network is also utilized to model the recomposition process $F_{\text{RC}}(\cdot)$ to generate the coefficients to recompose the band signals as follows,

$$\{w_1, w_2, w_3\} = F_{\text{RC}}\left(\left\{\Delta \hat{x}_{s_1}^T, \Delta \hat{x}_{s_2}^T, \Delta \hat{x}_{s_3}^T\right\}\right),$$

$$\hat{x}_3^F = \sum_{i=1}^{3} w_i \Delta \hat{x}_{s_i}^T, \qquad (5)$$

$$\Delta \hat{x}_{s_i}^T = \hat{x}_{s_i}^T - F_{\text{U}}\left(\hat{x}_{s_{i-1}}^T\right), \; i = 2, 3,$$

$$\Delta \hat{x}_{s_1}^T = \hat{x}_{s_1}^T,$$

where $\hat{x}_3^F$ is trained with the following three losses:

$$L_{\text{Detail}} = -\phi\left(\hat{x}_3^F - x\right), \qquad (6)$$

$$L_{\text{Percept}} = \left\|F_{\text{P}}(\hat{x}_3^F) - F_{\text{P}}(x)\right\|_2^2, \qquad (7)$$

$$L_{\text{Quality}} = -\log D\left(\hat{x}_3^F\right), \qquad (8)$$

where D is the discriminator measuring the probability that $\hat{x}_3^F$ is preferred by human vision. $F_{\text{P}}(\cdot)$ is the process to extract deep features from a pretrained VGG network.

The whole loss function in this stage is as follows,

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

where λ_3 and λ_4 are weighting parameters.

Summarization. In our DRBN, we first perform band representation learning. Each band signal is learned to be recovered based on the guidance of paired dataset. This stage ensures the signal fidelity and detail recovery. After, band recomposition is performed to improve the visual quality of the enhanced images with the perceptual guidance of the unpaired dataset, where high-quality images serve as the prior knowledge of human vision perception.

4. Experiments

This section illustrates the experimental evaluation. More results and analysis (including some ablation studies) are provided in the supplementary material.

Experimental Setting. To fully evaluate the proposed method, we test our method on images from various scenes. The LOL real captured low/normal light images [30] are used for objective and subjective evaluations, since it is a real captured dataset including highly degraded images which most methods cannot achieve promising results. Beside, the results of NPE [28] and DICM [16] are also provided in the supplementary material. The compared methods include Bio-Inspired Multi-Exposure Fusion (BIMEF) [31], Brightness Preserving Dynamic Histogram Equalization (BPDHE) [10], Camera Response Model (CRM) [33], Differential value Histogram Equalization Contrast Enhacement (DHECE) [22], Dong [6], Exposure Fusion Framework (EFF) [32], Contrast Limited Adaptive Histogram Equalization (CLAHE) [36], Low-Light Image Enhancement via Illumination Map Estimation (LIME) [9], Multiple Fusion (MF) [7], Multiscale Retinex (MR) [13] Joint Enhancement and Denoising Method (JED) [25], Refined Retinex Model (RRM) [18], Simultaneous Reflectance and Illumination Estimation (SRIE) [8], Deep Retinex Decomposition (DRD) [30], Deep Underexposed Photo Enhancement (DeepUPE) [27], Single Image Contrast Enhancer (SICE) [2], EnlightenGAN [11].

Implementation Detail. The network is trained via two stages. In the first stage, the network is optimized by ADAM optimizer. The learning rate is set to 0.0001. The crop image size and batch size for training are set to 256 and 4, respectively. In the second stage, the generator and discriminator are trained by ADAM optimizer. The learning rate of the generator is set to 0.0001 and that of the discriminator is set 1e-6. The crop image size and batch size for training are set to 320 and 1, respectively. λ_1, λ_2, λ_3, and λ_4 are set to 0.1, 0.1, 0.01 and 1, respectively.

We partition LOL dataset [30] for training and testing. The training dataset includes 689 paired images with the index [1, 689] while the testing set includes 100 paired images with the index [690, 789]. In the first stage, 300 epochs are allowed while in the second stage, only 30 epochs are allowed. In the second stage training, the first module (BLN) is fixed. After 200 epochs in the first stage training, the learning rate drops by 0.5. The hyper-parameter of Adam optimizer is set as: $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1e - 8$.

Evaluation Criteria. We perform quantitative evaluations to compare the performance of different methods. We

Table 1. Quantitative results on real test images in *LOL-Real* dataset. EG denotes EnlightenGAN.

Metric	BIMEF [31]	BPDHE [10]	CRM [33]	DHECE [22]	Dong [6]	EFF [32]	CLAHE [36]	LIME [9]	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 [13]	JED [25]	RRM [18]	SRIE [8]	DRD [30]	DeepUPE [27]	SICE [2]	EG [11]	DRBN
PSNR	11.67	17.33	17.34	17.34	15.47	13.27	19.40	18.23	**20.13**
SSIM	0.4269	0.6654	0.6859	0.6859	0.5672	0.4521	0.6906	0.6165	**0.8295**
SSIM-GC	0.5158	0.7236	0.7459	0.7075	0.7476	0.7051	0.7250	0.6452	**0.8492**

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

(e) BIMEF   (f) LIME   (g) RRM   (h) JED

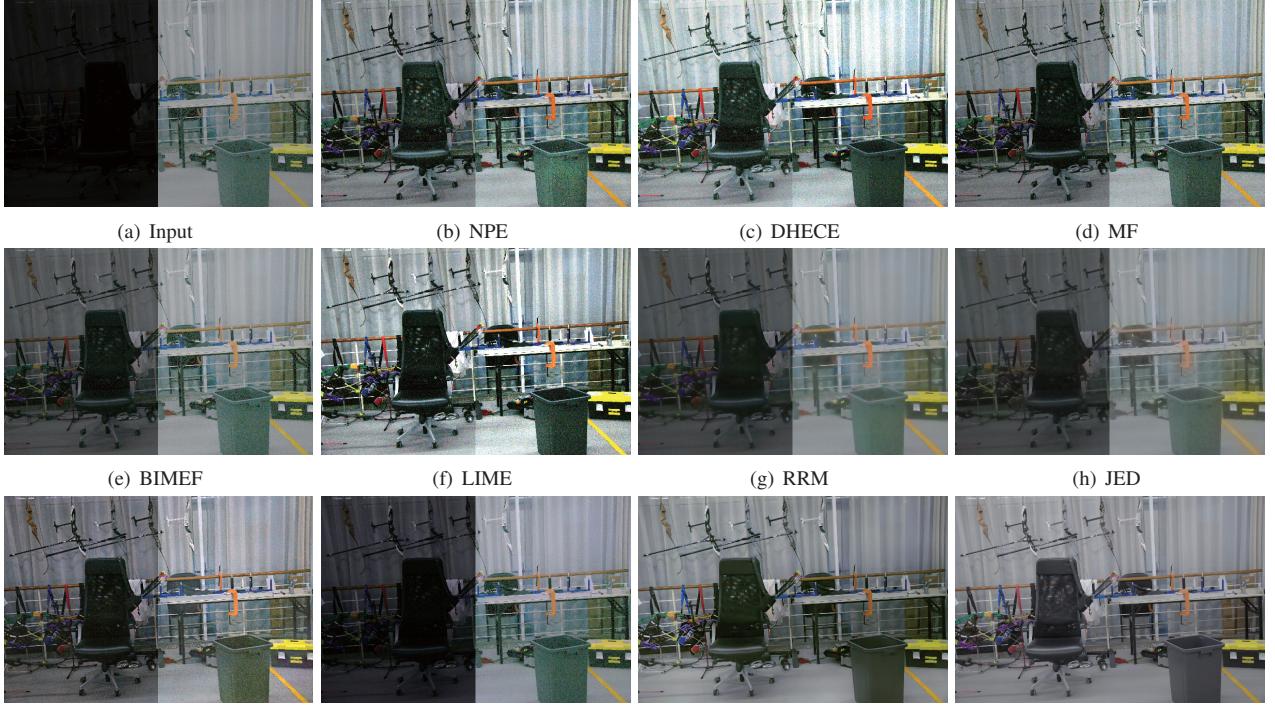(i) EnlightenGAN   (j) UPE   (k) DRBN   (l) GT

Figure 3. The visual results of different methods. Left part: the original results. Right part: the results corrected by Gamma transformation for better visibility.

adopt three objective evaluation metrics: Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity (SSIM) [29], and SSIM calculated based on the Gamma corrected results, which is called SSIM-GC. PSNR is the ratio between the maximum possible power of the normal light image and the power of the enhanced image and measures the *fidelity* of between them. SSIM considers more on image structures, takes the image degradation as perceived change in *structural information* and incorporates luminance masking and contrast masking terms into the metric. In the low-light enhancement task, the average luminance level is hard to be predicted. Therefore, the detail fidelity might be not well captured by PSNR and SSIM. Therefore, SSIM-GC is introduced, where a global illumination is corrected first via the Gamma transformation and then SSIM values are calculated.

Quantitative Evaluation. We compare different methods

quantitatively in Table 1. Our method achieves much better results in both PSNR, SSIM, and SSIM-GC. The results show our superiority in illumination restoration and structure recovery. The better result in SSIM-GC also shows our superiority when the global illumination is deducted. The SICE, EnligtenGAN and CRM also achieve superior PSNR values, which means that they well restore the global illumination. The generally low SSIM and SSIM-GC results of previous methods demonstrate their limitations in restoring structural details and stretching contrast, which are also confirmed in the latter qualitative evaluations.

Qualitative Evaluation. We also conduct extensive qualitative evaluations in Fig. 3 and 4. The results shows that, our DRBN achieves much superior quantitative and qualitative results to previous methods. In general, most of previous methods fail to well restore global illumination and structures. DHECE, LIME, NPE, and EnlightenGAN

Figure 4. The visual results of different methods. Left part: the original results. Right part: the results corrected by Gamma transformation for better visibility.
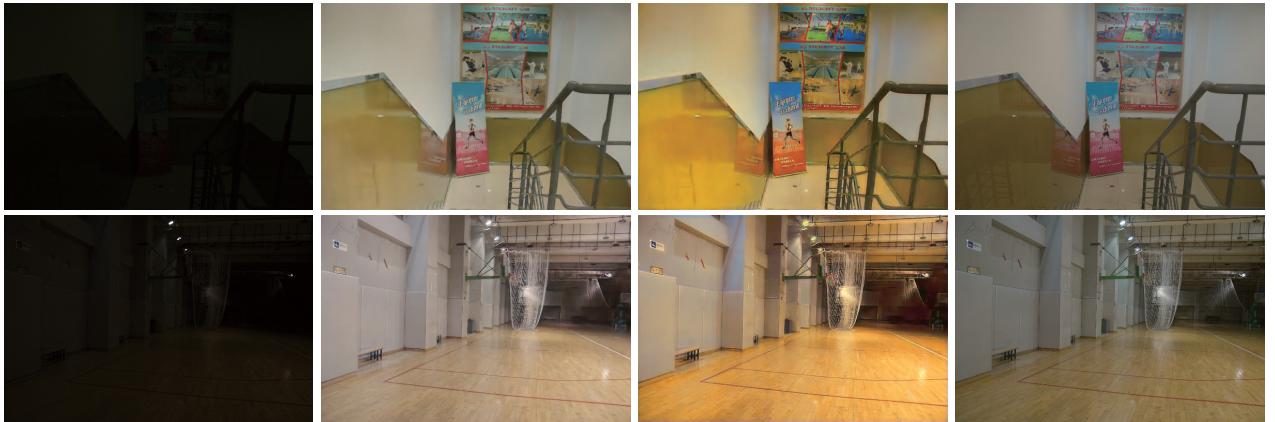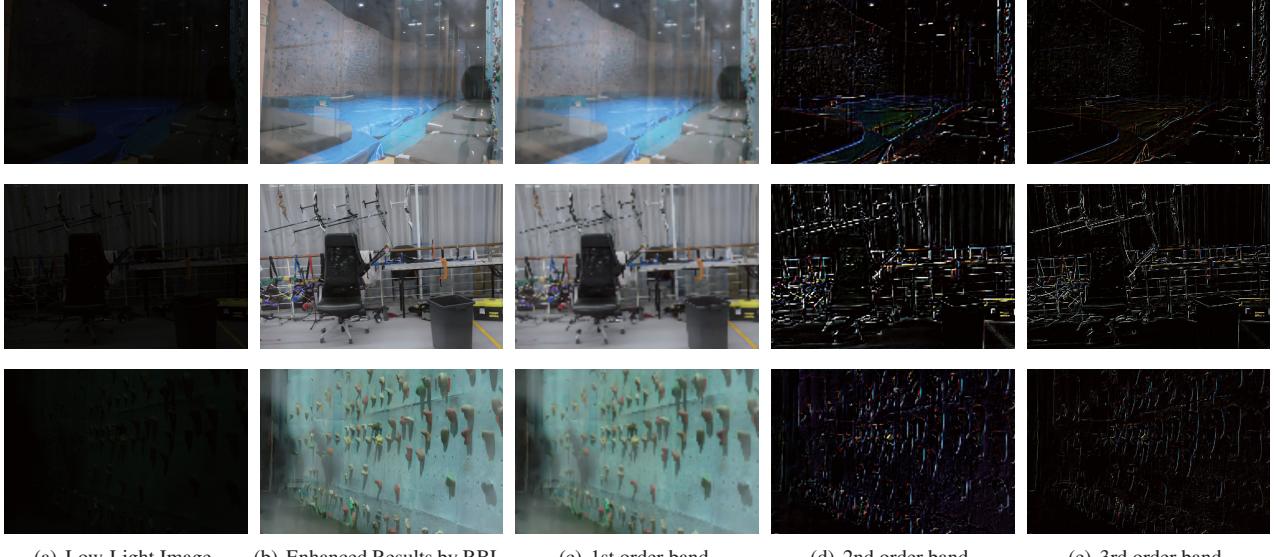

Figure 5. The first to fourth panels: low-light image, the result without perceptual guidance, the result with perceptual guidance, normal light image.

well restore the global illumination of the results. However, in their results, the burred noise is amplified, which heavily impairs the local details. RRM, MF, JED, and UPE suffer form under-exposure and poor visibility. In their Gamma corrected results (right parts), noise is observed in the results of RRM and JED. However, contrast of these two methods is not promising. Comparatively, our method achieves very good perceptual visual quality, with good illumination, color distribution, as well as clean and sharp
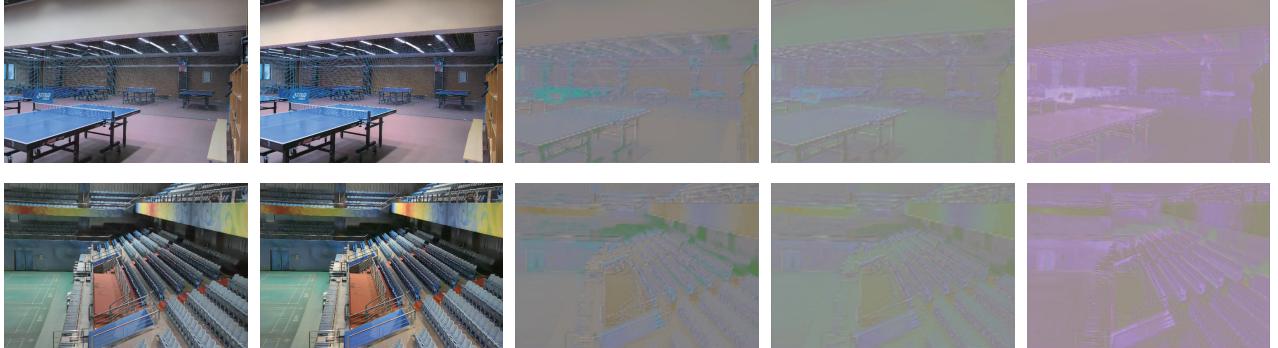
details.

Ablation Study for Two-Stage Design. We perform an ablation study of our two-stage design in Fig. 5. Comparing the results before and after the band recomposition, it is observed that, the results become more colorful and the contrast of the images is further boosted. Benefiting from the guided knowledge from a high-quality dataset, overall visibility and quality of these results even outperform the ground truths at the bottom rows.

3069

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

Figure 6. The visualization results of the learned bands by our DRBN. RBL denotes the recursive band learning.

| (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 |

Figure 7. The visualization results of the learned weighting maps for band recomposition. RBL and BR denote the recursive band learning and band recomposition, respectively.

5. Visualization of Recursive Band Learning and Band Recomposition

We visualize the learned band representations and the weighting masks for band recomposition in Fig. 6 and 7, respectively. It is observed that, our RBL effectively extracts a series of course-to-fine layered representations. After that, band recomposition reconstructs the band signals adaptively. Comparing the weighting maps for different bands, it is demonstrated that, the higher-order weighting maps are more sparse and focus more on edges and structures.

6. Conclusion

In this paper, we aim to create a novel semi-supervised learning method utilizing the knowledge of synthetic paired low/normal-light images and unpaired high-quality data for low-light image enhancement. To this end, we create a two-stage network which restores the signal based on fidelity first and then further enhances the results to improve overall visual quality. The DRBN recovers a linear band representation of an enhanced normal-light image with paired low/normal-light images, and then obtain an improved one by recomposing the given bands via another learnable linear transformation, which is trained with a perceptual quality-driven adversarial loss with unpaired high-quality data. The two-stage design makes our approach generate the enhanced results with well reconstructed details and visually promising contrast and color distributions. Both qualitative and quantitative evaluations demonstrate the advantages of the proposed method.

References

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

[2] J. Cai, S. Gu, and L. Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Trans. on Image Processing*, 27(4):2049–2062, April 2018. 1, 2, 5, 6

[3] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Trans. on Image Processing*, 27(4):2049–2062, April 2018. 3

[4] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 2018. 2, 3

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans. on Image Processing*, 16(8):2080–2095, Aug 2007. 2

[6] Xuan Dong, Guan Wang, Yi Pang, Weixin Li, Jiangtao Wen, Wei Meng, and Yao Lu. Fast efficient algorithm for enhancement of low lighting video. In *Proc. IEEE Int'l Conf. Multimedia and Expo*, pages 1–6, 2011. 5, 6

[7] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82 – 96, 2016. 2, 5, 6

[8] 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 Int'l Conf. Computer Vision and Pattern Recognition*, pages 2782–2790, June 2016. 2, 5, 6

[9] X. Guo, Y. Li, and H. Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Trans. on Image Processing*, 26(2):982–993, Feb 2017. 2, 5, 6

[10] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007. 5, 6

[11] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972*, 2019. 1, 2, 5, 6

[12] 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. on Image Processing*, 6(7):965–976, Jul 1997. 2

[13] 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. on Image Processing*, 6(7):965–976, July 1997. 5, 6

[14] D. J. Jobson, Z. Rahman, and G. A. Woodell. Properties and performance of a center/surround retinex. *IEEE Trans. on Image Processing*, 6(3):451–462, Mar 1997. 2

[15] Edwin H. Land. The retinex theory of color vision. *Sci. Amer*, pages 108–128, 1977. 2

[16] C. Lee, C. Lee, and C. S. Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE Trans. on Image Processing*, 22(12):5372–5384, Dec 2013. 5

[17] 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'l Conf. Image Processing*, pages 3730–3734, Sept 2015. 2

[18] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. on Image Processing*, 27(6):2828–2841, June 2018. 2, 5, 6

[19] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30 – 42, 2019. 3

[20] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Ll-net: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650 – 662, 2017. 2, 3

[21] N. Murray, L. Marchesotti, and F. Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pages 2408–2415, June 2012. 5

[22] Keita Nakai, Yoshikatsu Hoshi, and Akira Taguchi. Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms. In *International Symposium on Intelligent Signal Processing and Communications Systems*, pages 445–449, 2013. 5, 6

[23] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller. Contrast-limited adaptive histogram equalization: speed and effectiveness. In *Proceedings of Conference on Visualization in Biomedical Computing*, pages 337–345, May 1990. 2

[24] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M. Yang. Low-light image enhancement via a deep hybrid network. *IEEE Trans. on Image Processing*, 28(9):4364–4375, Sep. 2019. 2, 3

[25] X. Ren, M. Li, W. Cheng, and J. Liu. Joint enhancement and denoising method via sequential decomposition. In *IEEE Int'l Symposium on Circuits and Systems*, pages 1–5, May 2018. 2, 5, 6

[26] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma. MSR-net:Low-light Image Enhancement Using Deep Convolutional Network. *ArXiv e-prints*, November 2017. 3

[27] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2019. 2, 3, 5, 6

[28] S. Wang, J. Zheng, H. M. Hu, and B. Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. on Image Processing*, 22(9):3538–3548, Sept 2013. 2, 5

[29] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. on Image Processing*, 13(4):600–612, April 2004. 6

[30] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, Sept 2018. 1, 3, 5, 6

[31] Z. Ying, G. Li, and W. Gao. A Bio-Inspired Multi-Exposure Fusion Framework for Low-light Image Enhancement. *ArXiv e-prints*, November 2017. 5, 6

[32] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new image contrast enhancement algorithm using exposure fusion framework. In *International Conference on Computer Analysis of Images and Patterns*, pages 36–46. Springer, 2017. 5, 6

[33] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new low-light image enhancement algorithm using camera response model. In *Proc. IEEE Int'l Conf. Computer Vision*, Oct 2017. 5, 6

[34] Ye Yuan, Wenhan Yang, Wenqi Ren, Jiaying Liu, Walter J. Scheirer, and Zhangyang Wang. UG^{2+} Track 2: A Collective Benchmark Effort for Evaluating and Advancing Image Understanding in Poor Visibility Environments. *arXiv e-prints*, page arXiv:1904.04474, Apr 2019. 3

[35] 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'l Conf. Pattern Recognition*, pages 2034–2037, Nov 2012. 2

[36] Karel Zuiderveld. Graphics gems iv. chapter Contrast Limited Adaptive Histogram Equalization, pages 474–485. Academic Press Professional, Inc., San Diego, CA, USA, 1994. 5, 6