Integrating Semantic Segmentation and Retinex Model for Low Light Image Enhancement

Minhao Fan, Wenjing Wang
Peking University

Wenhan Yang
City University of Hong Kong

Jiaying Liu*
Peking University

ABSTRACT

Retinex model is widely adopted in various low-light image enhancement tasks. The basic idea of the Retinex theory is to decompose images into reflectance and illumination. The ill-posed decomposition is usually handled by hand-crafted constraints and priors. With the recently emerging deep-learning based approaches as tools, in this paper, we integrate the idea of Retinex decomposition and semantic information awareness. Based on the observation that various objects and backgrounds have different material, reflection and perspective attributes, regions of a single low-light image may require different adjustment and enhancement regarding contrast, illumination and noise. We propose an enhancement pipeline with three parts that effectively utilize the semantic layer information. Specifically, we extract the segmentation, reflectance as well as illumination layers, and concurrently enhance every separate region, *i.e.* sky, ground and objects for outdoor scenes. Extensive experiments on both synthetic data and real world images demonstrate the superiority of our method over current state-of-the-art low-light enhancement algorithms. Our code will be public available at: https://mm20-semanticreti.github.io/.

CCS CONCEPTS

• **Computing methodologies → Image manipulation**.

KEYWORDS

low light enhancement, image decomposition, semantic segmentation, image restoration, Retinex model

ACM Reference Format:
Minhao Fan, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2020. Integrating Semantic Segmentation and Retinex Model for Low Light Image Enhancement. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20), October 12–16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3394171.3413757

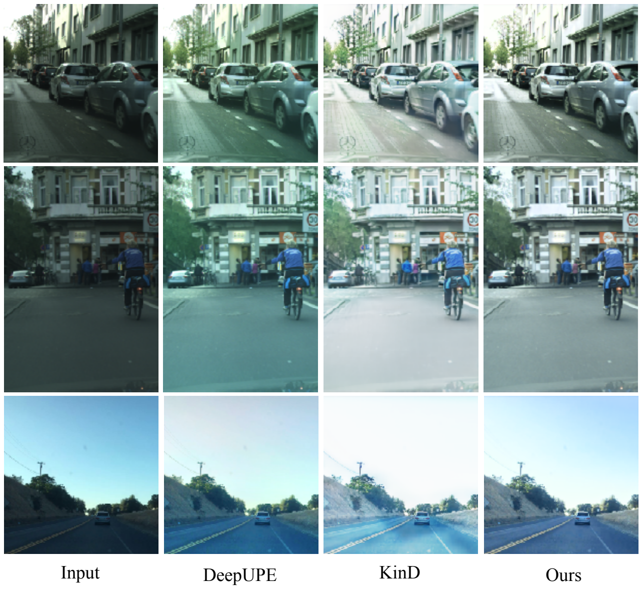


Figure 1: Our low-light enhancement method can reconstruct the visual quality of under-exposure images. Compared with existing methods DeepUPE [38] and KinD [51], with the help of the integrated semantic segmentation and the Retinex-based framework design, our result is of more natural color and illumination distribution.

1 INTRODUCTION

Images captured in low-light environment are degraded due to insufficient exposure and various sensor noises. In a dark scene, cameras fail to capture enough information to develop the details in images. Such images not only deteriorate visually, but also fail to be effectively processed by some machine vision systems, *e.g.* human and object recognition and detection in surveillance or auto-driving systems. As the capturing condition varies, there exist a large number of images with inconsistent degradation taken in such environments. Thus, there is a demand for enhancement methods to improve the quality of low-light images. However, the problem of enhancing underexposed images is challenging due to the large variance of content and illumination in such images. Such images suffer from insufficient visibility, low contrast, and usually variate levels of noise.

Many techniques have been developed to tackle the challenging problem of low-light enhancement. Early methods [1, 31] equalize the histogram of the image by brightening the dark regions and compressing bright pixels. Li *et al.* [23] proposed to apply dehazing methods on inverted low-light images to enhance visibility. Fu *et al.* [10] designed priors for light adjustment as well as noise suppression. Retinex theory based methods [15, 16, 18] decompose the signal into two components, *i.e.* reflectance and illumination. Specially designed priors (smoothness, structure invariance, *etc.*) are adopted subsequently to preserve details and suppress noise.

[*] Corresponding author. Email: liujiaying@pku.edu.cn.

MM '20, October 12–16, 2020, Seattle, WA, USA
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-7988-5/20/10...$15.00
https://doi.org/10.1145/3394171.3413757

Recently, deep learning based methods have been proposed for low-light enhancement [4, 14, 26, 38, 40, 43, 45]. Lore *et al.* [26] utilized stacked denoising auto-encoders to jointly learn low-light enhancement and noise reduction. SICE [4] decomposes low-light images into smooth and texture components and enhances them separately. Wei *et al.* [43] introduced a data-driven Retinex decomposition method, which integrates image decomposition and illumination mapping, and employed BM3D for noise reduction in the reflectance component. Wang *et al.* [38] proposed to learn an image-to-illumination mapping and design loss functions based on image priors. EnlightenGAN [14] utilizes gray-scale images as an attention map for self regularization with Generative Adversarial Network (GAN) to handle unpaired low-light image enhancement problem. However, existing models still suffer from regional degradation, *e.g.* overexposure, amplified noise, and color distortion, and fail to achieve satisfying visual quality in the whole image.

In this paper, we present a semantic layer-aware Retinex model to handle the above issues. We introduce a novel information extraction network, which learns to obtain the reflectance (R), illuminance (I) and semantic layers (S) of a low-light image. A subsequent enhancement network is proposed to improve the quality of the Retinex components R and S. Our model utilizes the structure information from S in a parallel enhancing architecture. To train and evaluate the proposed model, we build a dataset consisting of both synthetic and real images. The normal-light images in the dataset are collected from the Camvid [3] and Cityscapes [7], while the low-light images are synthesized by simulating the image capturing process with noise and illumination degradation to fit the distribution of real underexposed images. To thoroughly evaluate the proposed method, besides the synthetic images, we also collected a set of images captured in real low-light conditions. Experimental results show that our approach can successfully handle noise and color distortion, which outperforms multiple state-of-the-art approaches both quantitatively and qualitatively.

Our contributions are summarized as follows.

- We propose a novel deep network that integrates semantic segmentation and Retinex model for low light image enhancement. With the power of semantic prior and signal structure guidance, our model can successfully handle the illumination distribution, moderate noise and color distortion simultaneously and provide the results with superior visual quality in the low-light enhancement.
- The semantic prior is used to guide the enhancement of both illumination and reflectance jointly via manipulating features with a spatial transform, which improves the restoration quality of regional restoration, *e.g.* noise suppression, color correction. Extensive experiments show the superiority of our method and each component's effectiveness.
- To facilitate the related researches, we build a novel low light image synthesis model and generate a dataset of 2458 underexposed images, each with a ground truth image retouched based on the Cityscapes and Camvid datasets. Exposure adjustment, noise generation and color distortion are taken into consideration during our synthesis process. Additionally, a set of 100 real low light images are collected (online resource and captured) for evaluating our method and other state-of-the-art low-light image enhancement approaches.

2 RELATED WORK

2.1 Conventional Methods

The earliest low-light enhancement methods make the adjustment of the illumination of the low-light image uniformly. Histogram equalization (HE) turn dim images to be visible by changing the dynamic range of the input image [31] by manipulating its histogram. However, This kind of operation, HE is naturally easy to cause over-exposure and under-exposure. Without the local adaptation, the enhancement results in intensive noise and undesirable illumination. Later methods constrain the equalization process with several kinds of priors, *e.g.* mean intensity preservation [13], noise robustness, white and black stretching [2], and a new distortion model [19], to improve overall visual quality of the adjusted image. To better make fine-grained manipulation on the histograms, in [20, 30], the histogram equalization is applied to the difference of pixels. Some methods make attempts to introduce side information, *e.g.* depth information [22], to guide the pixel value change adaptively. In [44, 47], the imaging and visual perception models are utilized to guide the low-light image enhancement, *e.g.* camera response model [47] to select the best exposure ratio and visual importance [44] to control the contrast gain. Some methods [23, 49] take the low-light enhancement as the coupled problem of dehazing and stretch the visibility of the image by performing dehazing methods to the inverted low-light image In these methods, some off-line denoising operations [8] are also needed to remove noise, which inevitably cause detail blurriness sometime. Furthermore, a physical explanation on their basic model is missing.

2.2 Retinex-Based Methods

Later on, Retinex models are injected into the low-light enhancement problem. Retinex-based methods [18] will separate the whole image signal into illumination and reflectance and then operate them adaptively. To make the enhanced illumination natural and suppress noise presented in the reflectance, various priors are enforced to guide the manipulation of these two layers, *e.g.* structure aware prior [12], weighted variation [11], and multiple derivatives of illumination [10]. Meanwhile, variants of Retinex models are proposed to balance the layer separation and manipulation for better low-light enhancement, *e.g.* single-scale Retinex [16], multi-scale Retinex [15], naturalness Retinex [39], and robust Retinex [24, 33]. In [21], the weight of each single-scale Retinex is adaptively computed based on the input image. Wang *et al.* [39] construct a bright-pass filter for Retinex decomposition, and try to preserve the naturalness while enhancing details in low-light images. In [37], prior distributions of the reflectance and the illumination as well as the parameters of the enhancement process are jointly modeled with a hierarchical Bayesian model. Some methods explore the proper domain to apply the reconstruction prior. In [9], a novel model without the logarithmic transform is built to well preserve edges. There are also methods focusing on exploiting more effective priors [10–12] to regularize the enhancement of illumination and reflectance layers. Fu *et al.* [10] propose an improved version by fusing different merits into a single one based on multiple derivatives of the

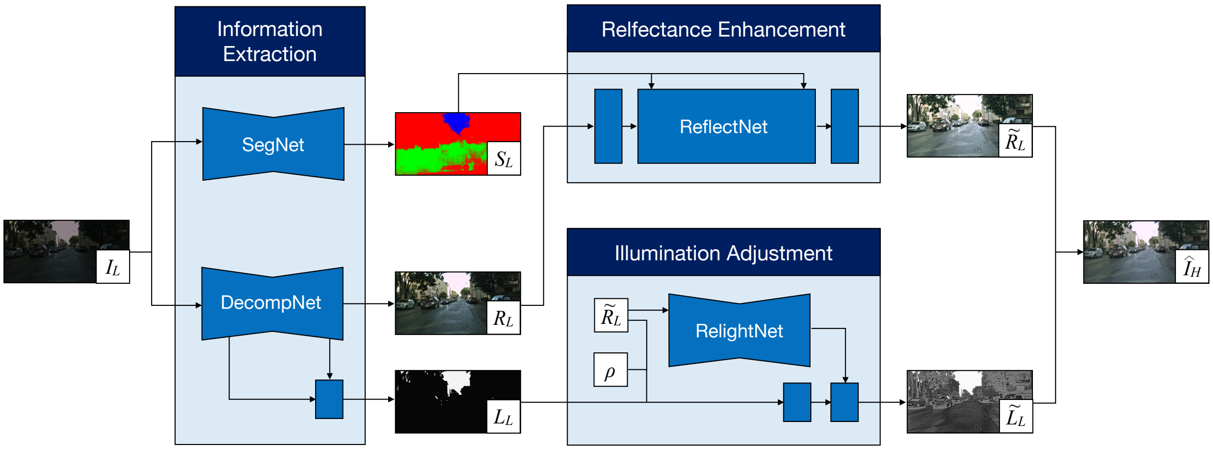


Figure 2: The architecture of the proposed semantic-aware Retinex-based low-light enhancement network, including three components: Information Extraction, Reflectance Restoration, and Illumination Adjustment. We first estimate semantic segmentation, reflectance, and illumination from the input underexposed image. Then, we enhance reflectance with the help of semantic information, and use the reconstructed reflectance to adjust the illumination. The final result is generated by fusing both reflectance and illumination.

estimated illumination. Guo *et al.* [12] proposed to refine an initial illumination map with a structure aware prior. In [11], a weighted variational model is proposed to impose better prior representation in the regularization terms. These methods consider less on the constraints on the reflectance, and the latent intensive noises in the low-light regions are usually amplified. Li *et al.* [24] proposed to extend the traditional Retinex model to a robust one with an explicit noise term, and made the first attempt to estimate a noise map out of that model via an alternating direction minimization algorithm. These methods obtain good results in illumination enhancement and light noise suppression. Nevertheless, they are built with only hand-crafted constraints. Therefore, these methods are not adaptive enough to capture the complex signal properties of the diverse kinds of nature images.

2.3 Learning-based Methods

Recently, deep-learning has brought in large changes to low-level image processing tasks and also brings in impressive performance gains for low-light image enhancement. Lore *et al.* [26] made the first efforts in creating a deep auto-encoder – Low-Light Net (LLNet) – to perform contrast enhancement and noise suppression jointly. Later on, various deep network-based methods are proposed with diversified kinds of network architectures and priors [4, 32, 35, 38, 43]. In [27, 35, 36], the multi-scale features are injected into the multi-branch architecture to form better low-light enhancement results. Some of these works [4, 26, 38] make efforts in building paired low/normal-light datasets for the model training. Diversified losses are utilized to help train the enhancement model, such as, MSE [26], SSIM loss [4], and compound loss [38]. In [35, 41, 43], Retinex structure is injected into the design of effective deep networks, to have both the advantages of Retinex-based methods, *i.e.* good signal structure, and deep learning-based methods, *i.e.* the general effective priors extracted from the large-scale training data. In [32], the layer decomposition and separative processing are used to better model structure and detail . In [14, 17], the adversarial learning is utilized to capture the visual properties beyond the traditional metrics. Especially for EnlightenGAN [14], 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 training data and practical testing applications. There are also works on deep-learning based image enhancement from raw images [6], or the joint task of low-light image enhancement and high-level computer vision tasks, such as face detection [48], object detection [25], *etc.* Compared to these works, our method integrates both Retinex-based layer separation and manipulation and semantic prior modeling jointly, which offer better low-light enhancement results due to the signal structure constraint and contextualized semantic awareness.

3 SEMANTIC-AWARE RETINEX-BASED LOW-LIGHT ENHANCEMENT

3.1 Motivation and Overall Architecture

The goal of illumination enhancement is to improve the visual quality of a given underexposed image comprehensively, requiring not only brightness distribution adjustment, but also noise suppression and detail correction. Semantic information can provide a wealth of information for low-light enhancement. For example, noises on smooth regions such as skies can be strongly blurred without hurting the subjective effect, while on regions with rich details such as street signs, denoising should be careful, or else details can be destroyed. However, existing low-light enhancement methods neglect the importance of semantic information, therefore are of limited capability.

In this paper, we propose a novel semantic-aware low-light enhancement network, which leverages semantic information for better scene understanding and intrinsic reflectance restoration. The overall architecture of the proposed model is illustrated in Fig. 2. Imitating Retinex decomposition, our network consists of three components: Information Extraction (Sec. 3.2), Reflectance Restoration (Sec. 3.3), and Illumination Adjustment (Sec. 3.4). In the following, we will provide the details of each part and how semantic awareness is introduced into the framework.

3.2 Information Extraction

The first step of low-light enhancement is information extraction, where three kinds of features are estimated: reflectance R, illumination I, semantic information S. The first two are obtained through a deep Retinex decomposition process, while the last one is obtained through a semantic segmentation network.

Image Decomposition. Inspired by RetinexNet [43] and KinD [51], we follow the Retinex theory and assume that the observed image I can be decomposed into two components, *i.e.* reflectance R and illumination L, where $I = R \cdot L$. Note that this problem is ill-posed, extra constraints required. Based on the observation that the same object looks different under different ambient light, we assume that a low-quality image (I_L) and its corresponding normal-light image (I_H) have consistent structures on their reflectance layers.

The decomposition network DecompNet is trained with pairs of low- and normal-light images. Mutual smoothness loss, reconstruction loss, and illumination smoothness loss are used to guide the network training. Please refer to [43] and [51] for the definition of each objective function.

Semantic Segmentation. In this step, we extract semantic information S_L from the input low-light image I_L by a SegNet. The estimated S_L is later used as guidance for the reflectance restoration stage. We focus on outdoor street scenes, which are common in autonomous driving and city surveillance. Streets contain various kinds of objects. However, to guide the low-light enhancement, fine-grained classification is not necessary. We simply split street scenes into three segments: sky, ground and foreground objects. These three regions are usually different in perspective and reflection attributes. The sky is usually smooth and often has a different light source than the objects on the ground. Compared with carriageways, other foreground objects are usually brighter and contain richer details. We will show later that by processing each semantic component separately, the enhancement result can be improved.

To extract semantic features, we adopt a light-weight U-Net, which is powerful enough to handle the three-category segmentation task. Both low- and normal-light images are used to train SegNet. Denote S_L and S_H as the segmentation estimation of I_L and I_H respectively, the basic segmentation objective function is

$$\mathcal{L}^H_{Seg} = \mathrm{CE}(S_H, S_{GT}), \tag{1}$$

$$\mathcal{L}^L_{Seg} = \mathrm{CE}(S_L, S_{GT}), \tag{2}$$

where CE denotes element-wise cross entropy loss, and S_{GT} is the ground-truth segmentation label.

Moreover, an L1 consistency constraint $\mathcal{L}_1$ is forced on S_H and S_L as follows,

$$\mathcal{L}_1 = ||S_H - S_L||_1. \tag{3}$$

Compared with the underexposed I_L, the objects in I_H are easier to detect, therefore SegNet usually performs better on I_H. With the consistency between S_H and S_L, the $\mathcal{L}_1$ can guide a more accurate estimation of S_L.

The final loss function for SegNet is:

$$\mathcal{L}_{Seg} = \lambda^H_{Seg}\mathcal{L}^H_{Seg} + \lambda^L_{Seg}\mathcal{L}^L_{Seg} + \lambda_1\mathcal{L}_1, \tag{4}$$

where λs are weights to balance the loss terms.

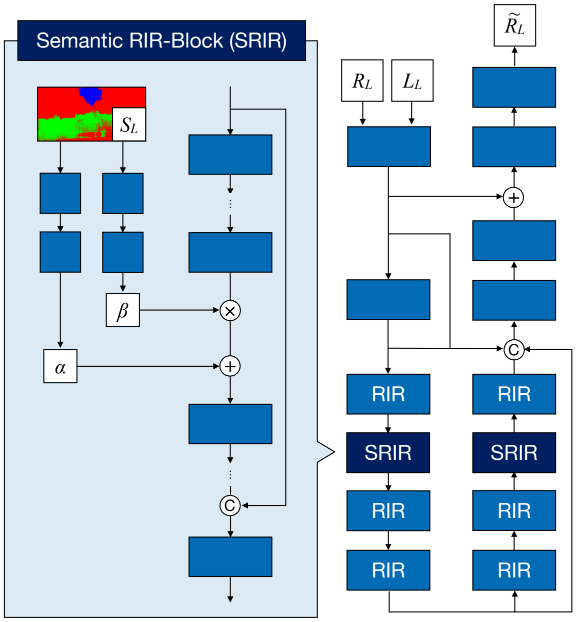


Figure 3: The architecture of our reflectance enhancement network. R_L, L_L and S_L denote the reflectance, illumination, and semantic information of the low-light image, respectively. The network yields the enhanced reflectance $\tilde{R}_L$.

3.3 Reflectance Enhancement

The roughly decomposed reflectance of the low-light image R_L suffers from strong noises and color bias. Therefore, we use the reflectance enhancement procedure to reconstruct R_L with the help of semantic information.

The reflectance enhancement sub-network ReflectNet contains a stack of Residual In Residual (RIR) [5] blocks. We further add concatenation-based and addition-based skip connections between the front and back end of the network. The residual-based architecture provides ReflectNet with the strong capacity of pixel adjustment, which is beneficial for noise suppression and color correctness.

To introduce semantic information, we design a variant of Semantic Residual In Residual (SRIR) block, where semantic information adjusts the feature of R_L through a linear transformation. As shown in Fig. 3, in our proposed SRIR block, semantic information S_L is processed by several convolutional layers, after which we obtain α and β. The feature of R_L first multiplies with $beta$ and then adds with α.

ReflectNet is guided with the reflectance R_H extracted from the ground truth normal-light image. We adopt the combination of Mean Squared Error (MSE), Structure SIMilarity index (SSIM) [42] and Gradient loss (Grad) [51] as the final loss for this stage. Denote $\tilde{R}_L$ as the reconstruction result by ReflectNet, the loss function can be formulated as follows:

$$\mathcal{L}_{\mathrm{R}} = \mathrm{MSE}(\tilde{R}_L, R_H) + \lambda^S_R \mathrm{SSIM}(\tilde{R}_L, R_H) + \lambda^G_R \mathrm{Grad}(\tilde{R}_L, R_H), \tag{5}$$

where λs are weights to balance the loss terms.

3.4 Illumination Adjustment

The directly decomposed illumination of the low-light image L_L is usually buried in darkness. For the adjustment of L_L, we designed a RelightNet and use the restored reflectance $\tilde{R}_L$ as well as an enhancement ratio ρ to rebalance the lightness distribution.

First, we use a light-weighted U-Net to extract features from the reconstructed reflectance $\tilde{R}_L$. In this way, the semantic information S_L is indirectly introduced into the illumination adjustment process. The result of RelightNet $\tilde{L}_L$ is limited to range $[0, 1]$ through a Sigmoid layer. By combining $\tilde{L}_L$ and $\tilde{R}_L$ through element-wise multiplication, our low-light enhancement framework generates the final prediction $\hat{I}_H = \tilde{L}_L \times \tilde{R}_L$.

In our training dataset, the distribution of illumination spans a wide range, which is close to the real application scenario, but it can pose challenges to model training. To solve this problem, a ratio parameter ρ is introduced to guide the adjustment procedure. For training, ρ is defined by the ratio of the mean pixel value of the ground truth I_H to that of the input I_L, while in the testing phase, we use a fixed $\rho = 5.0$. The ratio can be also provided by users as an interface to control the degree of brightening.

We use the ground truth I_H to directly guide RelightNet, which introduce more context information for model training. For better utilization of the illumination adjustment ratio ρ, besides MSE, SSIM, and Gradient loss, we propose a ratio learning loss:

$$\mathcal{L}_\rho = |\rho I_L - \tilde{L}_L \times \tilde{R}_L|. \tag{6}$$

The final loss function for this stage can be formulated as follows:

$$\begin{aligned}\mathcal{L}_{\text{RE}} =& \text{MSE}(\hat{I}_H, I_H) + \lambda_L^S \text{SSIM}(\hat{I}_H, I_H) \\ &+ \lambda_L^G \text{Grad}(\hat{I}_H, I_H) + \lambda_L^\rho \mathcal{L}_\rho,\end{aligned} \tag{7}$$

where λs are weights to balance the loss terms.

4 EXPERIMENTS

In this section, we first provide the details of data preparation and implementation, then demonstrate the superiority of the proposed method through experimental comparisons. We also demonstrate the effectiveness of our designs through ablation studies.

4.1 Datasets

Data Collection. In order to guide our model to utilize the intrinsic semantic information, we synthesize paired low/normal-light images based on two semantic segmentation datasets, Cityscapes [7] and Camvid [3]. We gather a collection of 2,458 normal-light images from these datasets, and retouch them to have a more pleasant illumination distribution. During the collection, we remove those with severe motion-blur or lens impediments and crop broken image boundary if necessary. The slight color distortion of the Cityscapes dataset is also adjusted for a better quality of the ground truths. We use the selected normal-light images to synthesize the corresponding low-light images, and split 2,118 pairs for training and the other 340 pairs for evaluation.

We also collect 100 real low-light images for testing and comparison. Among them, 20 are dim images selected from Cityscapes, 30 are twilight images from BDD100k, and others are manually captured from mobile phones. Illumination and noise level are diversified among the images in both global and local distribution.

Low-Light Data Synthesis. Our comprehensive synthesis procedure includes illumination adjustment, slight color distortion as well as noise simulation, accommodating various degradations in the low-light environment. Our process is as follows: first, a normal-light image is mapped to range $[0, 1]$ using inverse camera function [34] and linearly scaled with a value δ randomly sampled from the range $[0.5, 1]$ to simulate raw sensor data generation. Note that this param only serves for noise generation and is not related to the final illumination distribution. Then, we introduce moderate color distortion by multiplying a parameter from the ranges of $[0.9, 1.1]$ with each channel of the image. Read and shot noise, of which standard deviation is sampled based on the work proposed by Mildenhall *et al.* [28] is then added to generate noisy signals under dim ambient light. Afterwards, the current noisy image is divided by δ while another factor indicating illumination scale β sampled from $[0.5, 1.1]$ is used to adjust the illumination linearly. We allow the upper bound to be above 1.0 so that the over-exposure condition is retained. Nevertheless, to avoid too many bright images appearing in the set, we adjust the distribution of the illumination, which ensures bright images count for no more than 3 percent of the whole dataset. Finally, another gamma correction is adopted, so that the output synthetic image is with the gamma factor sampled from $[0.77, 0.91]$.

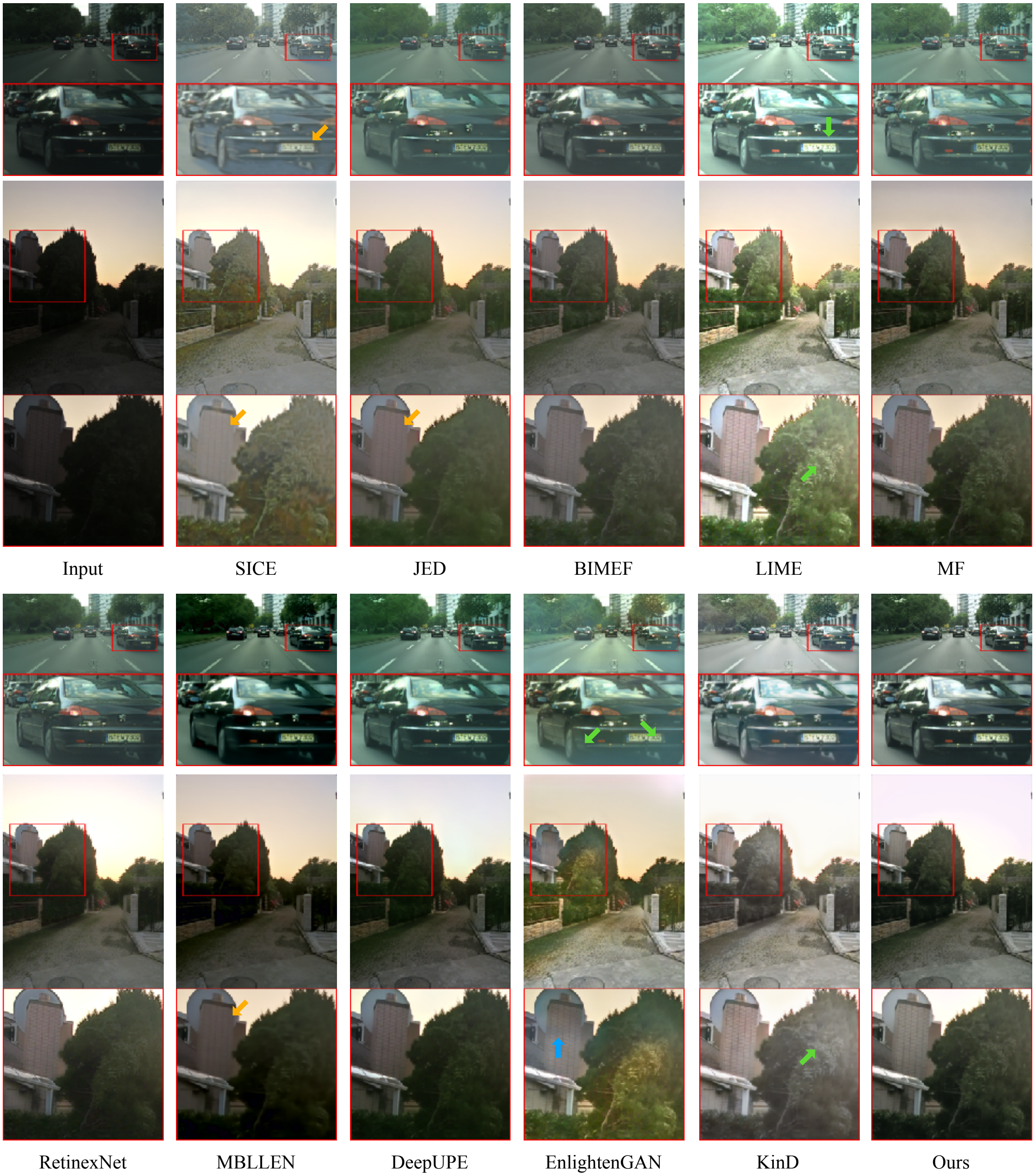
4.2 Implementation Details

Our pipeline consists of three major parts, *i.e.* Information Extraction, Reflectance Enhancement, and Illumination Adjustment. The training process is also split into four stages, corresponding to the training of DecompNet, SegNet, ReflectNet, and RelightNet. First, we pre-train SegNet with a batch size of 16 and a patch size of 128×128. The DecompNet is trained with a batch size of 10. For ReflectNet, the batch size is set to 10 and the patch size is set to 48×48. In the final illumination adjustment stage, the patch size is again set to 128×128 in order to avoid unstable training due to the highly variational light condition. We train the DecompNet for 2000 epochs and the other stages for 500 epochs with the learning rate initialized as 1×10^{-3} for DecompNet and 10^{-4} for other sub-networks. Adam is used for optimization.

Please refer to the supplementary material for the detailed architecture of each sub-network and other implementation details.

4.3 Evaluation on Synthesis Data

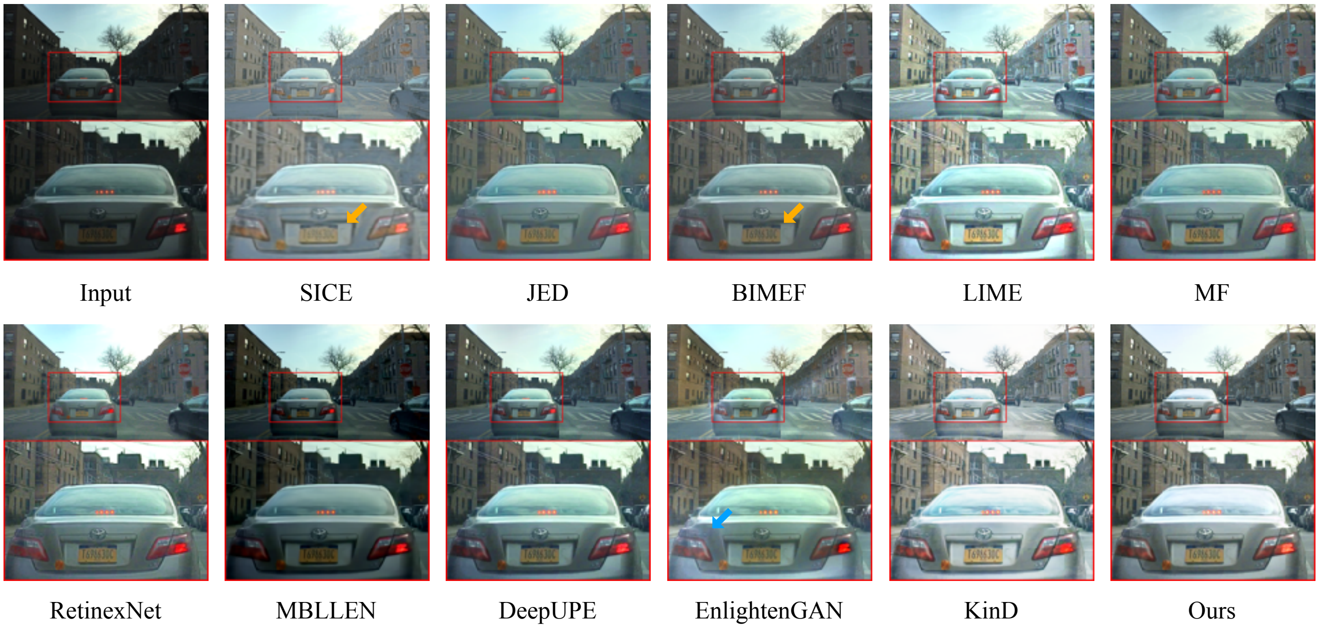
We first evaluate the performance on our synthesized data. We adopt three metrics to evaluate our method. Full reference metrics Peak Signal to Noise Ratio (PSNR) and Structure SIMilarity index (SSIM) [42] are used to measure signal and structure fidelity, while the Natural Image Quality Evaluator (NIQE) [29] is chosen as the blind image quality assessment for the naturalness of the enhanced images. We compare our methods with several state-of-the-art methods including LIME [12], BIMEF [46], MF [10], JED [33], SICE [4], RetinexNet [43], MBLLEN [27], EnlightenGAN [14], DeepUPE [38], and KinD [51]. Among them, deep-based RetinexNet and KinD are retrained on our dataset. As reported in Table 1, our method outperforms all the state-of-the-art methods on all of the metrics, demonstrating the effectiveness of our designs.



Input SICE JED BIMEF LIME MF

RetinexNet MBLLEN DeepUPE EnlightenGAN KinD Ours

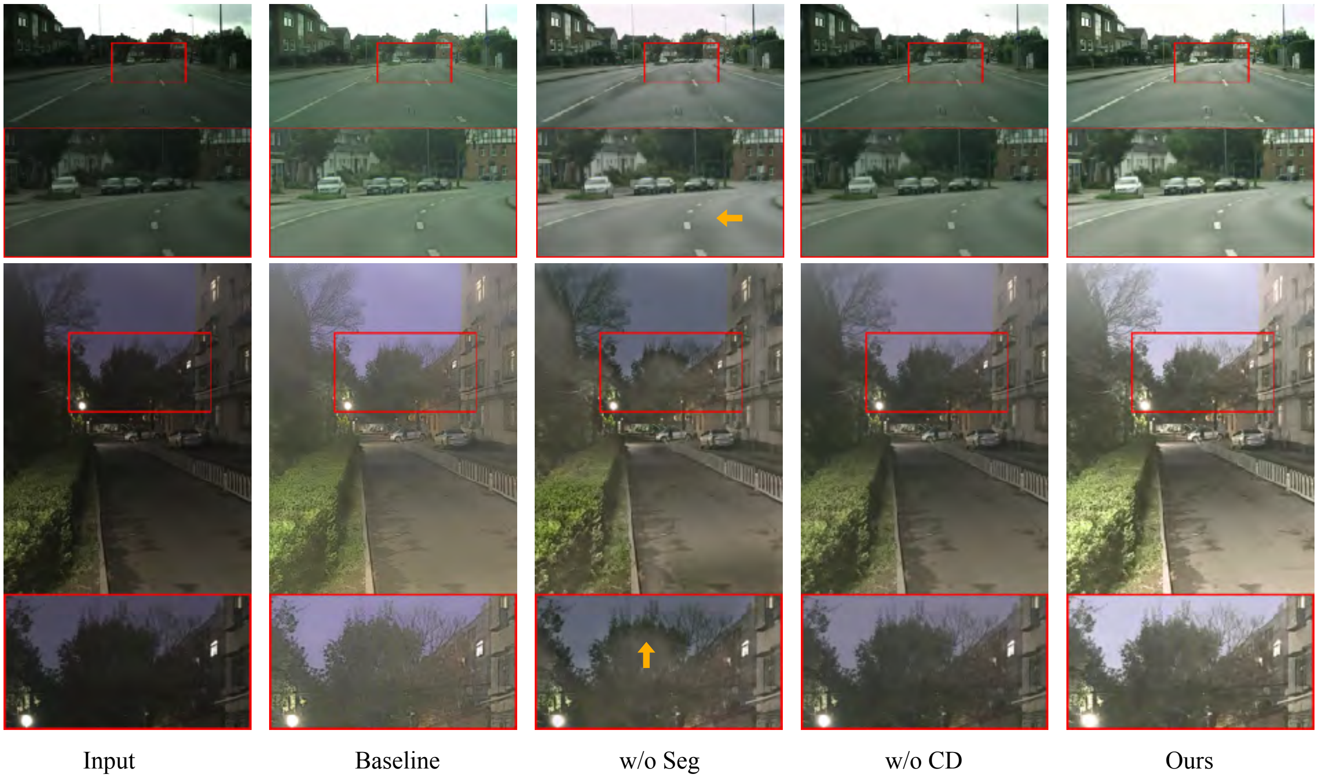
Figure 4: Comparison results of low-light enhancement. Detail blurriness, over-exposure, and weird artifacts are pointed by yellow, green, and blue arrows, respectively.



Input SICE JED BIMEF LIME MF

RetinexNet MBLLEN DeepUPE EnlightenGAN KinD Ours

Figure 5: Comparison results of low-light enhancement. Detail blurriness and weird artifacts are pointed by yellow and blue arrows, respectively.



Input Baseline w/o Seg w/o CD Ours

Figure 6: Ablation studies of the proposed method. Weird black edges and shadows are pointed by yellow arrows. CD denotes the color distortion in the training data synthesis stage.

Table 1: Quantitative results on our synthetic dataset.

Methods	PSNR↑	SSIM↑	NIQE↓
LIME [12]	12.304	0.508	5.385
BIMEF [46]	22.642	0.762	4.952
MF [10]	20.115	0.651	5.458
JED [33]	21.604	0.798	5.149
SICE [4]	16.227	0.783	5.254
RetinexNet [43]	23.763	0.722	4.960
MBLLEN [27]	17.228	0.744	3.571
DeepUPE [38]	22.503	0.710	4.946
EnlightenGAN [14]	16.953	0.731	4.881
KinD [51]	22.842	0.921	3.180
Ours	**28.816**	**0.951**	**3.048**

Table 2: Quantitative results on real-captured dataset.

Methods	NIQE↓	Methods	NIQE↓
Input	3.787	RetinexNet	3.249
LIME	3.123	MBLLEN	4.111
BIMEF	3.230	EnlightenGAN	3.153
MF	3.238	DeepUPE	3.226
JED	3.988	KinD	2.957
SICE	5.833	Ours	**2.886**

Table 3: The setting of our ablation study.

Methods	RIR	Semantic	Color Distortion
Baseline	×	×	×
w/o Seg	✓	×	✓
w/o CD	✓	✓	×
Ours	✓	✓	✓

Table 4: Ablation study of our proposed network. PSNR, SSIM, and NIQE are results on the synthetic data, while NIQE_R denotes results on the real-captured data.

Methods	PSNR↑	SSIM↑	NIQE↓	NIQE_R↓
Baseline	23.763	0.718	4.960	3.249
w/o Seg	28.915	0.952	3.299	3.159
w/o CD	**31.082**	**0.956**	3.113	3.182
Ours	28.816	0.951	**3.048**	**2.886**

4.4 Evaluation on Real-Captured Data

As there is no normal-light ground truth in our real-captured data, we only use the no-reference metric NIQE [29]. Deep-based methods compared in this section are all not retrained on our synthetic dataset, as training data preparation belongs to parts of their methods and contributions.

The quantitative results are shown in Table 2, where our method performs the best. It verifies the superiority of our method in generating more natural images over the other state-of-the-art methods.

Qualitative comparison results are shown in Fig. 4 and Fig. 5. SICE, BIMEF and MF fades the color, while SICE also generates black edges. JED focues on noise removal, however it can blur the details. LIME improves the overall brightness but overexposes certain regions. MBLLEN generates relatively natural illumination distribution, while its results are still not bright enough. RetinexNet causes severe color and contrast distortion, leading to unnatural results. DeepUPE presents a pleasant contrast while there are details missing in the underexposed parts. EnlightenGAN generates regional color artifacts that somehow degenerate visual quality. KinD fails to predict appropriate illumination for the enhancement results, which overexposes brighten parts and causes wired white artifacts. Moreover, most of the methods produce color distortion. Comparatively, the color distribution of our result is natural, which echoes our superior performance in the quantitative evaluation.

4.5 Ablation Study

We conduct ablation studies to demonstrate the design of our framework, specifically, the effectiveness of network architecture, applying color distortion in the data synthesis process, and introducing semantic information. The setting can be found in Table 3. For the baseline model, we use the architecture of RetinexNet [50]. RetinexNet also follows the pipeline of Retinex decomposition but has different network architecture choices.

Table 4 presents quantitative ablation study results on both synthetic and real sets. Qualitative results are shown Fig. 6. Compared with the baseline, our new network architectures (w/o Seg) largely improves the overall performance, with PSNR improving 7.3 and SSIM improving 0.238. However, as shown in Fig. 6, the result suffers from severe black edge artifacts, and the illumination is naturally distributed. The injection of semantic information (w/o CD) further improves the PSNR by 2.2 and NIQE by 0.2. More importantly, as shown in Fig. 6, the weird black edges disappeared and the result is more natural. This is because semantic segmentation guides the ReflectNet and RelightNet to understand the context of each region in the image. Finally, with the improved data synthesis design, our full model has the best visual quality as shown in Fig. 6. Although the full model performs slightly worse on synthetic data for PSNR and SSIM, the NIQE qualitative performance on both real and synthetic data improves. This is because the network has learned to remove color distortion, therefore generating results of higher visual quality.

5 CONCLUSION

In this paper, we propose a semantic aware Retinex-based model to address the low-light image enhancement problem. a novel joint decomposition and semantic segmentation method is introduced to extract a single image, which learns to decompose an image into reflectance, illumination, and semantic information. With the help of semantic prior and signal structure guidance, our model is successful to handle the illumination distribution, moderate noise and color distortion simultaneously and obtain the superior results in the low-light enhancement. Extensive experiments demonstrate the superiority of our method and the effectiveness of its each component.

ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China under Grand No.2018AAA0102700, National Natural Science Foundation of China under contract No.61772043, and Beijing Natural Science Foundation under contract No.L182002.

REFERENCES

[1] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. Akber Dewan, and O. Chae. 2007. A Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Trans. on Consumer Electronics* 53, 2 (May 2007), 593–600.

[2] T. Arici, S. Dikbas, and Y. Altunbasak. 2009. A histogram modification framework and its application for image contrast enhancement. *IEEE Trans. on Image Processing* 18, 9 (2009), 1921–1935.

[3] G. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla. 2008. Segmentation and Recognition Using Structure from Motion Point Clouds. In *Proc. IEEE European Conf. Computer Vision.* 44–57.

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

[5] J. Cai, W. Zuo, and L. Zhang. 2019. Extreme Channel Prior Embedded Network for Dynamic Scene Deblurring. *arXiv* abs/1903.00763 (2019). arXiv:1903.00763

[6] C. Chen, Q. Chen, J. Xu, and V. Koltun. 2018. Learning to See in the Dark. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition.*

[7] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition.*

[8] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. 2007. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Trans. on Image Processing* 16, 8 (Aug 2007), 2080–2095.

[9] X. Fu, Y. Sun, M. LiWang, Y. Huang, X. Zhang, and X. Ding. 2014. A novel retinex based approach for image enhancement with illumination adjustment. In *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing.* 1190–1194.

[10] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley. 2016. A fusion-based enhancing method for weakly illuminated images. *Signal Processing* 129 (2016), 82 – 96.

[11] X. Fu, D. Zeng, Y. Huang, X. P. Zhang, and X. Ding. 2016. A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition.* 2782–2790.

[12] X. Guo, Y. Li, and H. Ling. 2017. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Trans. on Image Processing* 26, 2 (Feb 2017), 982–993.

[13] H. Ibrahim and N. S. Pik Kong. 2007. Brightness Preserving Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Trans. on Consumer Electronics* 53, 4 (Nov 2007), 1752–1758.

[14] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang. 2019. EnlightenGAN: Deep Light Enhancement without Paired Supervision. *arXiv e-prints* (June 2019), arXiv:1906.06972. arXiv:1906.06972

[15] D. Jobson, Z. Rahman, and G. Woodell. 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. on Image Processing* 6, 7 (Jul 1997), 965–976.

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

[17] G. Kim, D. Kwon, and J. Kwon. 2019. Low-Lightgan: Low-Light Enhancement Via Advanced Generative Adversarial Network With Task-Driven Training. In *Proc. IEEE Int'l Conf. Image Processing.* 2811–2815.

[18] E. Land. 1977. The retinex theory of color vision. *Sci. Amer* (1977), 108–128.

[19] C. Lee, J. Kim, C. Lee, and C. Kim. 2014. Optimized Brightness Compensation and Contrast Enhancement for Transmissive Liquid Crystal Displays. *IEEE Trans. on Circuits and Systems for Video Technology* 24, 4 (April 2014), 576–590.

[20] C. Lee, C. Lee, and C. Kim. 2013. Contrast Enhancement Based on Layered Difference Representation of 2D Histograms. *IEEE Trans. on Image Processing* 22, 12 (Dec 2013), 5372–5384.

[21] C. Lee, J. Shih, C. Lien, and C. Han. 2013. Adaptive multiscale retinex for image contrast enhancement. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on.* IEEE, 43–50.

[22] J. Lee, C. Lee, J. Sim, and C. Kim. 2014. Depth-guided adaptive contrast enhancement using 2D histograms. In *Proc. IEEE Int'l Conf. Image Processing.* 4527–4531.

[23] L. Li, R. Wang, W. Wang, and W. Gao. 2015. A low-light image enhancement method for both denoising and contrast enlarging. In *Proc. IEEE Int'l Conf. Image Processing.* 3730–3734.

[24] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo. 2018. Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model. *IEEE Trans. on Image Processing* 27, 6 (June 2018), 2828–2841.

[25] Y. Loh and C. Chan. 2019. Getting to know low-light images with the Exclusively Dark dataset. *Computer Vision and Image Understanding* 178 (2019), 30 – 42.

[26] K. Lore, A. Akintayo, and S. Sarkar. 2017. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition* 61 (2017), 650 – 662.

[27] F. Lv, F. Lu, J. Wu, and C. Lim. 2018. MBLLEN: Low-light Image/Video Enhancement Using CNNs. In *British Machine Vision Conference.*

[28] B. Mildenhall, J. Barron, J. Chen, D. Sharlet, R. Ng, and R. Carroll. 2018. Burst Denoising With Kernel Prediction Networks. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition.* 2502–2510.

[29] A. Mittal, R. Soundararajan, and A. Bovik. 2013. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* 20, 3 (2013), 209–212.

[30] K. Nakai, Y. Hoshi, and A. Taguchi. 2013. Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms. In *International Symposium on Intelligent Signal Processing and Communications Systems.* 445–449.

[31] S. Pizer, R. Johnston, J. Ericksen, B. Yankaskas, and K. Muller. 1990. Contrast-limited adaptive histogram equalization: speed and effectiveness. In *Proc. of Conf. on Visualization in Biomedical Computing.* 337–345.

[32] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M. Yang. 2019. Low-Light Image Enhancement via a Deep Hybrid Network. *IEEE Trans. on Image Processing* 28, 9 (Sep. 2019), 4364–4375.

[33] X. Ren, M. Li, W. Cheng, and J. Liu. 2018. Joint Enhancement and Denoising Method via Sequential Decomposition. In *IEEE Int'l Symposium on Circuits and Systems.* 1–5.

[34] Y. Ren, Z. Ying, T. Li, and G. Li. 2019. LECARM: Low-Light Image Enhancement Using the Camera Response Model. *IEEE Trans. Circuits Syst. Video Techn.* 29, 4 (2019), 968–981.

[35] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, and J. Ma. 2017. MSR-net:Low-light Image Enhancement Using Deep Convolutional Network. *ArXiv e-prints* (November 2017). arXiv:cs.CV/1711.02488

[36] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie. 2017. LLCNN: A convolutional neural network for low-light image enhancement. In *Proc. IEEE Visual Communication and Image Processing.* 1–4.

[37] L. Wang, L. Xiao, H. Liu, and Z. Wei. 2014. Variational Bayesian Method for Retinex. *IEEE Trans. on Image Processing* 23, 8 (Aug 2014), 3381–3396.

[38] R. Wang, Q. Zhang, C. Fu, X. Shen, W. Zheng, and J. Jia. 2019. Underexposed Photo Enhancement Using Deep Illumination Estimation. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition.*

[39] S. Wang, J. Zheng, H. M. Hu, and B. Li. 2013. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Trans. on Image Processing* 22, 9 (Sept 2013), 3538–3548.

[40] W. Wang, C. Wei, W. Yang, and J. Liu. 2018. GLADNet: Low-Light Enhancement Network with Global Awareness. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference.* IEEE, 751–755.

[41] Y. Wang, Y. Cao, Z. Zha, J. Zhang, Z. Xiong, W. Zhang, and F. Wu. 2019. Progressive Retinex: Mutually Reinforced Illumination-Noise Perception Network for Low Light Image Enhancement. In *ACM Trans. Multimedia.*

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

[43] C. Wei, W. Wang, W. Yang, and J. Liu. 2018. Deep Retinex Decomposition for Low-Light Enhancement. In *British Machine Vision Conference.*

[44] X. Wu, X. Liu, K. Hiramatsu, and K. Kashino. 2017. Contrast-accumulated histogram equalization for image enhancement. In *Proc. IEEE Int'l Conf. Image Processing.* 3190–3194.

[45] W. Yang, S. Wang, Y. Fang, Y. Wu, and J. Liu. 2020. From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition.*

[46] Z. Ying, G. Li, and W. Gao. 2017. A Bio-Inspired Multi-Exposure Fusion Framework for Low-light Image Enhancement. *ArXiv e-prints* (November 2017). arXiv:cs.CV/1711.00591

[47] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang. 2017. A New Image Contrast Enhancement Algorithm Using Exposure Fusion Framework. In *Proc. Int'l Conf. Computer Analysis of Images and Patterns.* 36–46.

[48] Y. Yuan, W. Yang, W. Ren, J. Liu, W. Scheirer, and Z. Wang. 2019. UG^{2+} Track 2: A Collective Benchmark Effort for Evaluating and Advancing Image Understanding in Poor Visibility Environments. *arXiv e-prints* (Apr 2019), arXiv:1904.04474.

[49] X. Zhang, P. Shen, L. Luo, L. Zhang, and J. Song. 2012. Enhancement and noise reduction of very low light level images. In *Proc. IEEE Int'l Conf. Pattern Recognition.* 2034–2037.

[50] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. 2018. Residual Dense Network for Image Super-Resolution. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition.*

[51] Y. Zhang, J. Zhang, and X. Guo. 2019. Kindling the Darkness: A Practical Low-light Image Enhancer. In *Proc. ACM Int'l Conf. Multimedia.*