Learning Semantic-Aware Knowledge Guidance for Low-Light Image Enhancement

Yuhui Wu[1], Chen Pan[1], Guoqing Wang[1]*, Yang Yang[1], Jiwei Wei[1], Chongyi Li[2], Heng Tao Shen[1]

[1]Center for Future Media, University of Electronic Science and Technology of China, China
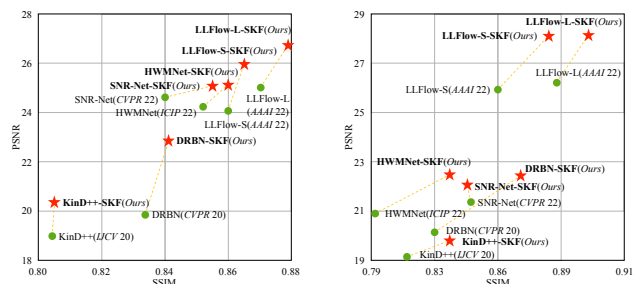[2]S-Lab, Nanyang Technological University, Singapore

wuyuhui132@gmail.com; panchen0103@163.com; gqwang0420@hotmail.com;

dlyyang@gmail.com; mathematic6@gmail.com; chongyi.li@ntu.edu.sg; shenhengtao@hotmail.com

Figure 1. **Motivation and superiority.** (a) The enhancement results (bottom row) without semantic priors show color deviations (*e.g.*, the black car turns gray). (b) Our SKF provides remarkable performance boost on LOL/LOL-v2 datasets in terms of PSNR/SSIM.

Abstract

Low-light image enhancement (LLIE) investigates how to improve illumination and produce normal-light images. The majority of existing methods improve low-light images via a global and uniform manner, without taking into account the semantic information of different regions. Without semantic priors, a network may easily deviate from a region's original color. To address this issue, we propose a novel semantic-aware knowledge-guided framework (SKF) that can assist a low-light enhancement model in learning rich and diverse priors encapsulated in a semantic segmentation model. We concentrate on incorporating semantic knowledge from three key aspects: a semantic-aware embedding module that wisely integrates semantic priors in feature representation space, a semantic-guided color histogram loss that preserves color consistency of various instances, and a semantic-guided adversarial loss that produces more natural textures by semantic priors. Our SKF is appealing in acting as a general framework in LLIE task. Extensive experiments show that models equipped with the SKF significantly outperform the baselines on multiple datasets and our SKF generalizes to different models and scenes well. The code is available at Semantic-Aware-Low-Light-Image-Enhancement

1. Introduction

In real world, low-light imaging is fairly common due to unavoidable environmental or technical constraints such as insufficient illumination and limited exposure time. Low-light images not only have poor visibility for human perception, but also are unsuitable for subsequent multimedia computing and downstream vision tasks designed for high-quality images [4, 9, 36]. Thus, low-light image enhancement (LLIE) is proposed to reveal buried details in low-light images and avoid degraded performance in subsequent vision tasks. Mainstream traditional methods for LLIE include Histogram Equalization-based methods [2] and Retinex model-based methods [18].

Recently, many deep learning-based LLIE methods have proposed, such as end-to-end frameworks [5,7,34,45,46,48] and Retinex-based frameworks [29, 41, 43, 44, 49, 53, 54]. Benefiting from their ability in modeling the mapping between the low-light and high-quality image, deep LLIE methods commonly achieve better results than traditional approaches. However, existing methods typically improve low-light images globally and uniformly, without taking into account the semantic information of different regions, which is crucial for enhancement. As shown in Fig. 1(a), a network that lacks the utilization of semantic priors can easily deviate from a region's original hue [22]. Furthermore, studies have demonstrated the significance of incorporating semantic priors into low-light enhancement. Fan *et al*. [8] utilize semantic map as prior and incorporated it into the feature representation space, thereby enhancing image quality. Rather than relying on optimizing intermediate fea-

tures, Zheng *et al.* [58] adopt a novel loss to guarantee the semantic consistency of the enhanced images. These methods successfully combine the semantic priors with LLIE task, demonstrating the superiority of semantic constraints and guidance. However, their methods fail to fully exploit the knowledge that semantic segmentation networks can provide, limiting the performance gain by semantic priors. Furthermore, the interaction between segmentation and enhancement is designed for specific methods, limiting the possibility of incorporating semantic guidance into LLIE task. Hence, we wonder two questions: *1. How can we obtain various and available semantic knowledge? 2. How does semantic knowledge contribute to image quality improvement in LLIE task?*

We attempt to answer the first question. First, a semantic segmentation network pre-trained on large-scale datasets is introduced as a semantic knowledge bank (SKB). The SKB can provide richer and more diverse semantic priors to improve the capability of enhancement networks. Second, according to previous works [8, 19, 58], the available priors provided by the SKB primarily consist of intermediate features and semantic maps. Once training a LLIE model, the SKB yields above semantic priors and guides the enhancement process. The priors can not only refine image features by employing techniques like affinity matrices, spatial feature transformations [40], and attention mechanisms, but also guide the design of objective functions by explicitly incorporating regional information into LLIE task [26].

Then we try to answer the second question. We design a series of novel methods to integrate semantic knowledge into LLIE task based on the above answers, formulating in a novel semantic-aware knowledge-guided framework (SKF). First, we use the High-Resolution Network [38] (HRNet) pre-trained on the PASCAL-Context dataset [35] as the previously mentioned SKB. In order to make use of intermediate features, we develop a semantic-aware embedding (SE) module. It computes the similarity between the reference and target features and employs cross-modal interactions between heterogeneous representations. As a result, we quantify the semantic awareness of image features as a form of attention and embed semantic consistency in enhancement network.

Second, some methods [20, 55] propose to optimize image enhancement using color histogram in order to preserve the color consistency of the image rather than simply enhancing the brightness globally. The color histogram, on the other hand, is still a global statistical feature that cannot guarantee local consistency. Hence, we propose a semantic-guided color histogram (SCH) loss to refine color consistency. Here, we intend to make use of local geometric information derived from the scene semantics and global color information derived from the content. In addition to guarantee original color of the enhanced image, it can also

add spatial information to the color histogram, performing a more nuanced color recovery.

Third, existing loss functions are not well aligned with human perception and fail to capture an image's intrinsic signal structure, resulting in unpleasing visual results. To improve visual quality, EnlightenGAN [16] employs global and local image-content consistency and randomly chooses the local patch. However, the discriminator do not know where the regions are likely to be 'fake'. Thus, we propose a semantic-guided adversarial (SA) loss. Specifically, the ability of the discriminator is improved by using segmentation map to determine the fake areas, which can improve the image quality further.

The main contributions of our work are as follows:

- We propose a semantic-aware knowledge-guided framework (SKF) to boost the performance of existing methods by jointly maintaining color consistency and improving image quality.

- We propose three key techniques to take full advantage of semantic priors provided by semantic knowledge bank (SKB): semantic-aware embedding (SE) module, semantic-guided color histogram (SCH) loss, and semantic-guided adversarial (SA) loss.

- We conduct experiments on LOL/LOL-v2 datasets and unpaired datasets. The experimental results demonstrate large performance improvements by our SKF, verifying its effectiveness in resolving the LLIE task.

2. Related Work

2.1. Low-light Image Enhancement

Traditional methods. Traditional methods for low-light enhancement include Histogram Equalization-based methods [2] and Retinex model-based methods [18]. The former improve low-light images by extending the dynamic range. The latter decompose a low-light image into reflection and illumination maps and the reflection component is treated as the enhanced image. Such model-based methods require explicit priors to fit data well, but designing proper priors for various scenes is difficult [44].

Learning-based methods. Recent deep learning-based methods show promising results [15, 29, 43, 44, 53, 54, 56]. We can further divide existing designs into Retinex-based methods and end-to-end methods. Retinex-based methods use deep network to decompose and enhance an image. Wei *et al.* proposed a two-stage Retinex-based method called Retinex-Net [43]. Inspired by Retinex-Net, Zhang *et al.* proposed two refined methods, called KinD [54] and KinD++ [53]. Recently, Wu *et al.* [44] proposed a novel deep unfolding Retinex-based network to further integrate the strengths of model-based and learning-based methods.

In comparison to Retinex-based method, end-to-end methods directly learning an enhanced result [5–7, 27, 32, 34, 37, 41, 45, 46, 51, 57, 59]. Lore *et al.* [30] made the
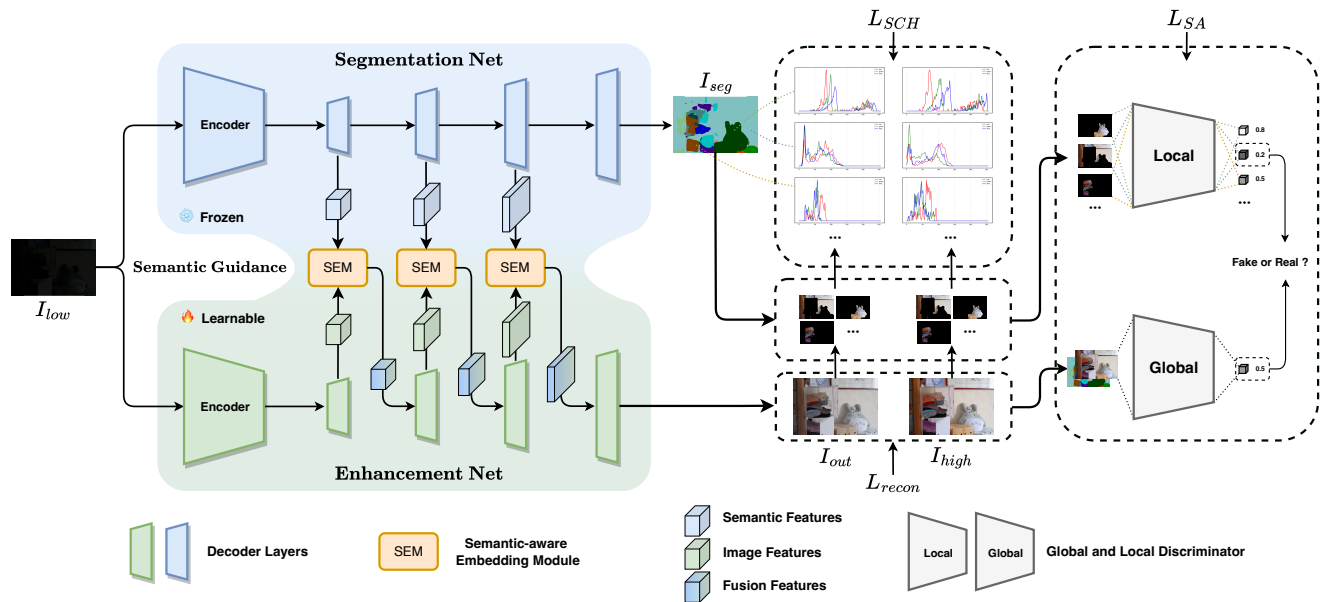
Figure 2. Overview of our Semantic-aware Knowledge-guided Framework (SKF). With a pre-trained Segmentation Net, our SKF utilizes semantic priors to improve the enhancement process in two aspects: **(a)** In feature-level, the multi-scale semantic-aware embedding modules enable cross-modal interactions between semantic features and image features in representation space. **(b)** In loss-level, the semantic segmentation result is introduced into the computation of color histogram loss and adversarial loss as a guidance.

first attempt by proposing a deep autoencoder named Low-Light Net (LLNet). Later on, various end-to-end methods are proposed. Physics-based concepts, *e.g.* Laplacian pyramid [27], local parametric filter [34], Lagrange multiplier [57], De-Bayer-Filter [5], normalization flow [41] and wavelet transform [7], are proposed to improve model interpretability and lead to visually pleasing results. In [16, 17, 48], adversarial learning is introduced to capture the visual properties. In [11], the light enhancement is creatively formulated as a task of image-specific curve estimation using zero-shot learning. In [20, 47, 55], 3D lookup table and color histogram are utilized to preserve the color consistency. However, existing designs focus on optimizing enhancement process, while ignoring the semantic information of different regions. In contrast, we design a SKF with three key techniques to explore the potential of semantic priors and thus produce visually pleasing enhanced results.

2.2. Semantic-Guided Methods

Recently, semantic-guided methods prove the reliability of semantic priors. These methods could be divided into two kinds: loss-level semantic-guided methods and feature-level semantic-guided methods.

Loss-level semantic-guided methods. In order to make use of semantic priors, some works focus on utilizing semantic-aware losses as extra objective functions of the original vision tasks. In image denoising [28], image super-resolution [1], low-light image enhancement [58], researchers directly utilized semantic segmentation loss as an extra constrain to guide the training process. Furthermore,

Liang *et al.* [26] better maintained the details of the images by using a semantic brightness consistency loss.

Feature-level semantic-guided methods. In comparison to loss-level semantic-guided methods, feature-level semantic-guided methods concentrate on extracting intermediate features from semantic segmentation network and introduce semantic priors in feature representation space to combine with image features. Similar works have been done on image restoration [23], image deraining [24], image super-resolution [40], low-light image enhancement [8], depth estimation [10, 19].

Existing semantic-guided methods are limited by the insufficient interaction between semantic priors and original tasks. Hence, we propose a semantic-aware framework to fully exploit semantic information both on loss-level and feature-level, including two semantic-guided losses and a semantic-aware embedding module. Specifically, comparing to semantic-guided methods in LLIE task [8, 26, 58], our SKF is appealing in acting as a general framework.

3. Method
3.1. Motivation and Overview

Illumination enhancement is the process of making an underexposed image look better by adjusting the lighting, eliminating noise, and restoring lost details. Semantic priors can provide a wealth of information for improving the enhancement performance. Specifically, semantic priors can help reformulate the existing LLIE methods as a region-aware enhancement framework. In particular, the novel model will blur noises on smooth regions in a simple way,

such as skies, whereas being careful on regions with rich details, such as indoor scene. Furthermore, combining with semantic prior, the color consistency of enhanced image will be carefully preserved. A network that lacks access to semantic priors can easily deviate from a region's original hue [22]. Existing low-light enhancement methods, however, ignore the importance of semantic information and thus have limited capability.

In this paper, we propose a novel SKF, jointly optimizing image features, maintaining regional color consistency and improving image quality. As shown in Fig. 2, semantic priors are provided by SKB and integrated into LLIE task by three key components: SE module, SCH loss and SA loss.

Problem definition of semantic-aware LLIE. Given a low-light image $I_l \in \mathbb{R}^{W \times H \times 3}$ with width W and height H. Combining with semantic segmentation, the LLIE process can be modeled as two functions, first:

$$M = \mathbf{F}_{segment}(I_l; \theta_s), \qquad (1)$$

where M is the semantic prior, including segmentation result and intermediate features with multi-scale dimensions. $\mathbf{F}_{segment}$ represents the pre-trained semantic segmentation network, acting as the SKB, and θ_s is frozen in training stage. Then M is used as input:

$$\widehat{I_h} = \mathbf{F}_{enhance}(I_l, M; \theta_e), \qquad (2)$$

where $\widehat{I_h} \in \mathbb{R}^{W \times H \times 3}$ is the enhanced result and $\mathbf{F}_{enhance}$ represents the enhancement network. During training stage, θ_e will be updated by minimizing the objective function with the guidance of M while θ_s is fixed:

$$\widehat{\theta_e} = argmin\mathcal{L}(\widehat{I_h}, I_h, M), \qquad (3)$$

where $I_h \in \mathbb{R}^{W \times H \times 3}$ is the ground truth, $\mathcal{L}(\widehat{I_h}, I_h, M)$ is the objective function of semantic-aware LLIE.

3.2. Semantic-Aware Embedding Module

When refining image features with the help of semantic priors, another challenge should be particularly considered is the discrepancy between the two sources. To alleviate this issue, we propose the SE module to refine the image feature maps, as illustrated in Fig. 3. The SE modules are like bridges between Segmentation Net and Enhancement Net (see Fig. 2), establishing connections between two heterogeneous tasks.

In our framework, we choose HRNet [38] as the SKB due to its exceptional performance and make some task-specific modifications. Besides the semantic map, we utilize output features before the representation head as multi-scale semantic priors. For further illustration, three SE modules are shown in Fig. 2, thus we take three semantic/image features $(F_s^b/F_i^b, b = 0, 1, 2)$ with three spatial resolutions $(H/2^{4-b}, W/2^{4-b})$, where H and W are the height and width of the input image. The SE module performs a pixel-wise interaction between F_s^b and F_i^b, and gives
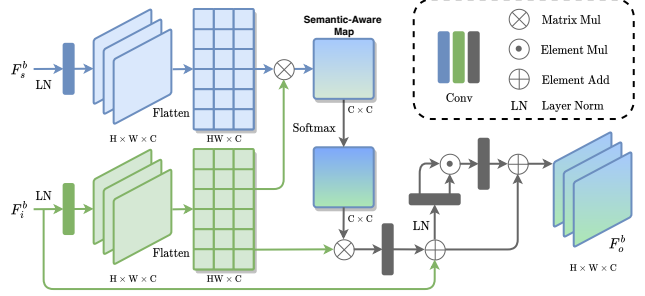

Figure 3. Architecture of the semantic-aware embedding (SE) module. At the bth decoder layer, SE module transforms the image feature map F_i^b with the semantic feature map F_s^b and produces the refined output feature F_o^b.

the final refined feature map F_o^b. Details of the learning process are provided below.

The SE module computes the semantic awareness of the image features through cross-modal similarity and produces a semantic-aware map. We first apply convolution layers to transform F_s^b and F_i^b to the same dimension. Next, inspired by Restormer [50], we adopt a transposed-attention mechanism to compute the attention map with a low computational cost. Hence, the semantic-aware attention map is described as follows:

$$A^b = Softmax\left(W_k(F_i^b) \times W_q(F_s^b)/\sqrt{C}\right), \qquad (4)$$

where $W_k(\cdot)$ and $W_q(\cdot)$ are convolution layers, LN is layer normalization, C is channel of features. Here, $A^b \in \mathbb{R}^{C \times C}$ indicates the semantic-aware attention map, which represents the interrelationship between F_i^b and F_s^b. Then we use A^b to fabricate image feature F_i^b as follows:

$$F_o^b = FN(W_v(F_i^b) \times A^b + F_i^b), \qquad (5)$$

where FN denotes feed-forward network, F_o^b is the final refined feature map of bth SE module and becomes the input of $(b+1)^{th}$ layer of the Enhancement Net decoder.

3.3. Semantic-Guided Color Histogram Loss

Color histogram carries crucial underlying image statistics and is profitable for learning color representations. DCC-Net [55] uses PCE modules with affinity matrix to match the color histogram and content in feature-level, therefore retaining color consistency of enhanced image. However, color histogram describes a global statistic information, differences in color characteristics between various instances are eliminated. Thus, we propose an intuitive way to achieve local color adjustment, *i.e.*, semantic-guided color histogram (SCH) loss, as shown in Fig. 2. It focuses on adjusting color histogram of each instance, thereby retaining more detailed color information.

The semantic map is firstly used to divide enhanced result into image patches with different instance labels. Each patch includes a single instance with the same label. Hence, the process of producing patches are defined as follows:

$$P = \{P^0, P^1, \ldots, P^{class}\}, \quad P^c = I_{out} \odot I_{seg}^c, \quad (6)$$

where $\odot$ is the dot product, I_{out} denotes the enhanced result, I_{seg}^c denotes the cth channel of the one-hot semantic map, $P^c \in \mathbb{R}^{W \times H \times 3}$ denotes the cth image patch, P denotes the group of all the patches.

Due to the discrete nature of color histogram, we approximate the differentiable version inspired by Kernel Density Estimation [3] for model training. Considering the prediction error of the semantic result, pixels that are close to the boundary are not considered. We refine the patch group P to P' without edge pixels, mitigating effects of misclassification. In the case of R channel of the cth image patch $P^{c'}(R)$, the estimation process is defined as follows:

$$x_{ij}^h = x_j - \frac{i - 0.5}{255}, \quad x_{ij}^l = x_j - \frac{i + 0.5}{255}, \quad (7)$$

where x_j denotes j^{th} pixel in $P^{c'}(R)$, $i \in [0, 255]$ denotes pixel intensity. x_{ij}^h and x_{ij}^l represent higher anchor and lower anchor respectively, which are key variables to estimate histogram as follows:

$$H_i^c = \sum_j \left(Sigmoid(\alpha \cdot x_{ij}^h) - Sigmoid(\alpha \cdot x_{ij}^l) \right), \quad (8)$$

$$H^c = \{i, H_i^c\}_{i=0}^{255}, \quad (9)$$

where H^c denotes the differentiable histogram of $P^{c'}(R)$, H_i^c denotes estimated number of pixels with intensity value i. α is a scaling factor, we experimentally set it to 400 for better estimation. The difference between results of two $Sigmoid(\cdot)$ denotes the contribution of x_j to the number of pixels with intensity value i. Specifically, when x_j exactly equal to i, the difference is 1, *i.e.*, x_j adds 1 to H_i^c.

Finally, we apply l_1 loss to constrain the estimated differentiable histogram. Therefore, the SCH loss can be described as follows:

$$\mathcal{L}_{SCH} = \sum_c \| H^c(\hat{I}_h) - H^c(I_h) \|_1, \quad (10)$$

where $\hat{I}_h$ and I_h denote output and groundtruth respectively, $H^c(\cdot)$ denotes the process of histogram estimation.

3.4. Semantic-Guided Adversarial Loss

Global and local discriminator is used to encourage more realistic results in image inpainting tasks [14, 25]. EnlightenGAN [16] employs this idea as well, but the local patches are selected randomly instead of focusing on fake regions. Therefore we introduce semantic information to guide the discriminator to focus on intriguing regions. To achieve this, we further refine the global and local adversarial loss function respectively by the segmentation map I_{seg} and image patches P' mentioned in Sec. 3.3. Finally, we propose the semantic-guided adversarial (SA) loss.

For the local adversarial loss, we first use refined patch group P' as candidate fake patches of the output I_{out}. Then, we compare the discriminating result of image patches

among P', the worst patch is most likely to be 'fake' and can be chosen to update parameters of both discriminator and generator. Hence, the discriminator plausibly uses the semantic priors to find the target fake region $x_f \sim p_{fake}$ by itself. While the real patches $x_r \sim p_{real}$ are still randomly cropped from real images each time. The local adversarial loss function is defined as:

$$\mathcal{L}_{local} = \min_G \max_D \mathbb{E}_{x_r \sim p_{real}} MSE(D(x_r), 0) + \mathbb{E}_{x_f \sim p_{fake}} MSE(D(x_f), 1), \quad (11)$$

$$x_f = P^t, D(P^t) = \min(D(P^0), \ldots, D(P^{class})), \quad (12)$$

where $MSE(\cdot)$ denotes the mean squared error and P^t denotes the target fake patch.

For the global adversarial loss, we adopt a simple design to achieve semantic-aware guidance when discriminating a fake sample. We concatenate I_{out} and I'_{seg}, which is the output feature before $Softmax$, as a new x_f. The images x_r with real distribution are randomly sampled. Finally, the global adversarial loss function is defined as:

$$\mathcal{L}_{global} = \min_G \max_D \mathbb{E}_{x_r \sim p_{real}} MSE(D(x_r), 0) + \mathbb{E}_{x_f \sim p_{fake}} MSE(D(x_f, I'_{seg}), 1), \quad (13)$$

Therefore, the SA loss can be defined as:

$$\mathcal{L}_{SA} = \mathcal{L}_{global} + \mathcal{L}_{local}, \quad (14)$$

We define the original loss function of Enhancement Net as $\mathcal{L}_{recon}$, which may be l_1 loss, MSE loss, $SSIM$ loss, etc., or their combination according to original setting of each selected method. Thus, the overall loss function of our SKF can be formulated as follows:

$$\mathcal{L}_{all} = \mathcal{L}_{recon} + \lambda_{SCH}\mathcal{L}_{SCH} + \lambda_{SA}\mathcal{L}_{SA}, \quad (15)$$

where λs are weights to balance the loss terms.

4. Experiments

4.1. Experimental Settings

Datasets. We evaluate the proposed framework on several datasets from various scenes, including LOL [43], LOL-v2 [49], MEF [31], LIME [12], NPE [39] and DICM [21]. The LOL dataset [43] is a real captured dataset including 485 low/normal light image pairs for training and 15 pairs for testing. The LOL-v2 dataset [49] is the real part of LOL-v2, which is larger and more diverse than LOL, including 689 low/normal light pairs for training and 100 pairs for testing. The MEF (17 images), LIME (10 images), NPE (85 images) and DICM (64 images) are real captured datasets including unpaired images.

Metrics. To evaluate the performance of different LLIE methods with and without our SKF, we use both full-reference and non-reference image quality evaluation metrics. For LOL/LOL-v2 datasets, peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [42], learned perceptual image patch similarity (LPIPS) [52], natural image

Table 1. Quantitative comparison on the LOL [43] and LOL-v2 [49] datasets. ↑ (↓) denotes that, larger (smaller) values lead to better quality. + (-) denotes the improvement (reduction) of performance, corresponding to ↑ (↓). The bold denotes the best.

Method	LOL				LOL-v2				Param(M)
	PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓	
LIME [13] TIP'16	16.760	0.560	0.350	-	15.240	0.470	-	-	-
Zero-DCE [11] CVPR'20	14.861	0.562	0.335	7.767	18.059	0.580	0.313	8.058	0.33
EnlightGAN [16] TIP'21	17.483	0.652	0.322	4.684	18.640	0.677	0.309	5.089	8.64
ISSR [8] MM'20	18.846	0.788	0.243	5.249	16.994	0.798	0.206	5.179	12.12
MIRNet [51] PAMI'22	24.140	0.842	0.131	4.203	20.357	0.782	0.317	5.094	5.90
RetinexNet [43] BMVC'18	16.770	0.462	0.474	8.873	18.371	0.723	0.365	5.849	0.62
RetinexNet-SKF(Ours)	20.418 (+3.648)	0.711 (+0.249)	0.216 (+0.258)	4.211 (+4.662)	19.849 (+1.478)	0.719 (-0.004)	0.255 (+0.110)	4.233 (+1.616)	0.66
KinD [54] MM'19	20.870	0.799	0.207	5.189	17.544	0.669	0.375	6.849	8.03
KinD-SKF(Ours)	21.913 (+1.043)	0.835 (+0.036)	0.143 (+0.064)	5.031 (+0.158)	19.821 (+2.277)	0.833 (+0.164)	0.201 (+0.174)	4.778 (+2.071)	8.50
DRBN [48] CVPR'20	19.860	0.834	0.155	4.793	20.130	0.830	0.147	4.961	2.21
DRBN-SKF(Ours)	22.837 (+2.977)	0.841 (+0.007)	0.138 (+0.017)	4.464 (+0.329)	22.441 (+2.311)	0.871 (+0.041)	0.132 (+0.015)	4.460 (+0.501)	2.43
KinD++ [53] IJCV'20	18.970	0.804	0.175	4.760	19.087	0.817	0.180	5.086	9.63
KinD++-SKF(Ours)	20.363 (+1.393)	0.805 (+0.001)	0.201 (-0.026)	**4.142 (+0.618)**	19.779 (+0.692)	0.837 (+0.020)	0.178 (+0.002)	4.179 (+0.907)	10.21
HWMNet [7] ICIP'22	24.240	0.852	0.114	5.141	20.928	0.798	0.359	5.970	66.56
HWMNet-SKF(Ours)	25.086 (+0.846)	0.860 (+0.008)	0.108 (+0.006)	4.346 (+0.795)	22.490 (+1.562)	0.836 (+0.038)	0.175 (+0.184)	4.683 (+1.288)	69.98
SNR-LLIE-Net [46] CVPR'22	24.608	0.840	0.151	5.179	21.479	0.848	0.157	4.623	39.13
SNR-LLIE-Net-SKF(Ours)	25.031 (+0.552)	0.855 (+0.015)	0.113 (+0.038)	4.722 (+0.457)	21.927 (+0.448)	0.842 (-0.006)	0.160 (-0.003)	**3.963 (+0.660)**	39.44
LLFlow-S [41] AAAI'22	24.060	0.860	0.136	5.412	25.922	0.860	0.173	6.150	4.97
LLFlow-S-SKF(Ours)	25.942 (+1.882)	0.865 (+0.005)	0.125 (+0.011)	5.606 (-0.194)	28.107 (+2.185)	0.884 (+0.024)	0.133 (+0.040)	5.415 (+0.735)	5.26
LLFlow-L [41] AAAI'22	24.999	0.870	0.117	5.582	26.200	0.888	0.137	5.406	37.68
LLFlow-L-SKF(Ours)	**26.798 (+1.799)**	**0.879 (+0.009)**	**0.105 (+0.012)**	5.589 (-0.007)	**28.451 (+2.251)**	**0.905 (+0.017)**	**0.112 (+0.025)**	5.725 (-0.319)	39.91

quality evaluator (NIQE) [33] are employed. For the MEF, LIME, NPE and DICM datasets without paired data, only NIQE is used, as there is no ground-truth.

Compared Methods. To verify the effectiveness of our designs, we compare our method with a rich collection of SOTA methods for LLIE, including LIME [13], RetinexNet [43], KinD [54], DRBN [48], KinD++ [53], Zero-DCE [11], ISSR [8], EnlightGAN [16], MIRNet [51], HWMNet [7], SNR-LLIE-Net [46], LLFlow [41]. To demonstrate the superiority of our method faithfully, we reasonably select several methods as the baseline networks. Specifically, both the most representative methods including RetinexNet, KinD and KinD++, and three latest methods including HWMNet, SNR-LLIE-Net and LLFlow are selected. Thus, our methods are denoted as RetinexNet-SKF, KinD-SKF, DRBN-SKF, KinD++-SKF, HWMNet-SKF, SNR-LLIE-Net-SKF, LLFlow-S-SKF and LLFlow-L-SKF (small and large version of LLFlow respectively).

Implementation Details. We conduct our experiments on NVIDIA 3090 GPU and NVIDIA A100 GPU, which are based on the released code of the baseline networks with the same training settings. Specifically, only the last subnets of Retinex-SKF, KinD-SKF and KinD++-SKF are trained with SCH loss and SA loss, while the other subnets are trained with the original loss functions. Furthermore, we do not apply SA loss to LLFlow because there is no enhanced output in training stage. Additionally, SE modules are reasonably located in decoders of all the baseline networks.

4.2. Quantitative Evaluation

Quantitative results on LOL and LOL-v2 datasets. The evaluation results are shown in Table 1. We can observe that our SKF achieves consistent and significant performance gain over each baseline method. Specifically, our SKF provides an average improvement of 1.750 dB/1.611 dB on LOL/LOL-v2 datasets respectively and this is achieved by introducing the capability of suppressing noise and artifacts and preserving color consistency. Notably, our LLFlow-L-SKF earns PSNR values of 26.798 dB/28.451 dB on LOL/LOL-v2 datasets, establishing a new SOTA. Furthermore, SSIM values achieve similar performance as well. Our SKF yields better SSIM values by an average of 0.041/0.037 on LOL/LOL-v2 datasets, which illustrates that our SKF helps baseline methods restore the luminance and contrast and preserve the structural information with details. Besides, the substantial gain of LPIPS and NIQE provided by our SKF reasonably indicates that human intuition is more closely matched by introducing semantic priors from our designs.

Quantitative results on MEF, LIME, NPE and DICM datasets. The evaluation results on the MEF, LIME, NPE and DICM datasets are described in Table 2. In general, each method with SKF obtain better NIQE results than baseline on all six datasets except three worse cases of DRBN-SKF and HWMNet-SKF. The RetinexNet-SKF performs the best with NIQE of 3.632 on MEF dataset, while the KinD++-SKF achieves the best performance on other five datasets. Overall, it is notable that our SKF yields an
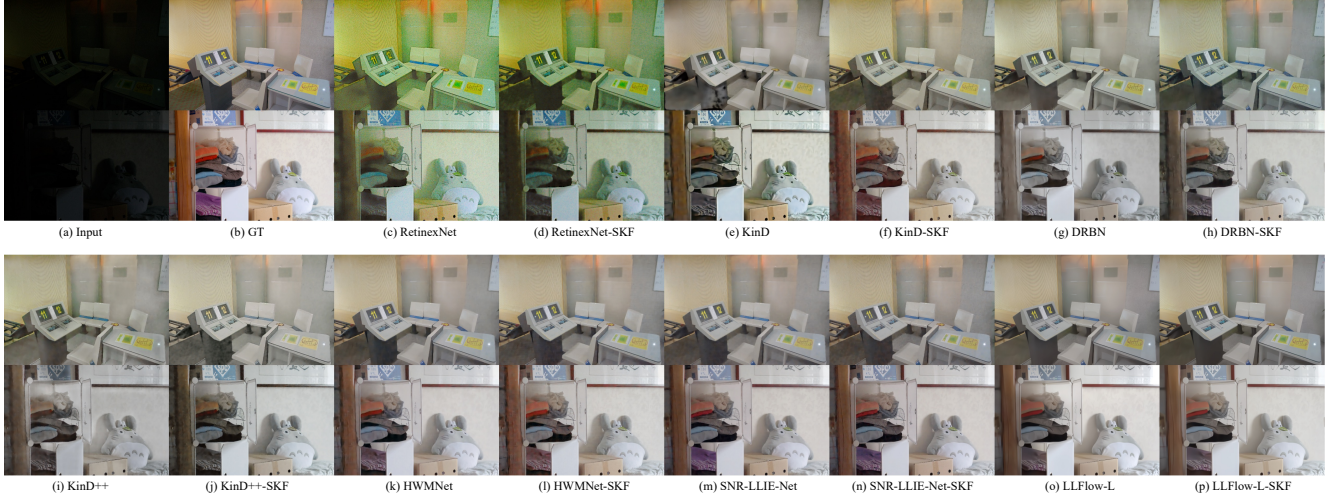
Figure 4. Visual comparison of baseline methods with and without SKF on LOL dataset. Our SKF enables baseline methods produce images with less noise, more color information and realistic details.

Figure 5. Visual comparison of baseline methods with and without SKF on LIME dataset.

Table 2. Quantitative comparison on the LOL [43], LOL-v2 [49], MEF [31], LIME [12], NPE [39] and DICM [21] datasets in terms of NIQE, where smaller values lead to better quality.

Method	LOL	LOL-v2	MEF	LIME	NPE	DICM
Input	6.7488	6.7911	4.2650	4.4380	4.3124	4.2550
RetinexNet [43] BMVC'18	6.8731	5.8488	4.1490	4.4200	4.5008	4.5912
RetinexNet-SKF(Ours)	4.2118	4.2331	**3.6321**	4.0779	4.0152	3.6945
KinD [54] MM'19	5.1891	6.8490	4.1344	4.6418	4.6896	3.9371
KinD-SKF(Ours)	5.0306	4.7783	3.9460	4.3607	3.8721	3.7909
DRBN [48] CVPR'20	4.7930	4.9612	4.0956	4.4019	3.9205	4.0433
DRBN-SKF(Ours)	4.4636	4.4599	4.0894	4.3392	4.0192	3.8541
KinD++ [53] IJCV'20	4.7602	5.0856	3.7498	4.3756	3.9848	3.7076
KinD++-SKF(Ours)	**4.1415**	**4.1785**	3.7645	**3.9892**	**3.8201**	**3.5382**
HWMNet [7] ICIP'22	5.1407	5.9702	4.2175	4.3549	4.0683	3.9196
HWMNet-SKF(Ours)	4.3460	4.6826	4.0312	4.3699	3.9942	4.0760

average gain of 0.519 on NIQE across all the methods and datasets. The better NIQE shows that the methods with our SKF can produce images with more natural textures and become more effective for restoring low-light images.

4.3. Qualitative Evaluation

The qualitative evaluations on LOL and LIME datasets are shown in Figs. 4 and 5 respectively. As indicated by Fig. 4, our SKF can improve the enhancement capability of baseline methods and generate images with more pleasing perceptual quality. Specifically, the results of RetinexNet are unreal due to the obvious color gap and serious noise, which can be mitigated by our SKF. Compared to results of KinD and KinD++, KinD-SKF and KinD++-SKF resolve the issue of inconsistent lighting and strange white artifacts. For other results, more consistent color and natu-

ral details recovery for desk, wall and clothes are achieved by our SKF.

We further exhibit the visual enhancement results on the LIME dataset in Fig. 5. It can be observed that the methods with our SKF suppress the unnatural halo around the lamp and restore naturalistic color and details. Hence, methods with our SKF yield more visually pleasing results as compared to baselines, supporting our method's excellent performance in quantitative evaluation. More visualization results are provided in the supplementary material.

4.4. Ablation Study

We conduct ablation studies on LOL dataset to prove the effectiveness of our SKF from various aspects.

SCH loss, SA loss and SE module. As shown in Table 3, we conduct experiments of KinD++-SKF, DRBN-SKF and HWMNet-SKF. The addition of SCH loss and SE module improve the PSNR by an average of 0.243 dB and 0.841 dB over the baseline respectively. Simultaneously applying SCH loss and SE module further improves the baseline method by yielding an average gain of 1.741 dB over the baseline. This verifies that more beneficial semantic-aware priors are integrated into enhancement process. Despite that adding SA loss causes minor drops in some full-reference metrics, average gain of 0.292 with NIQE are obtained across all the cases. Therefore, the baseline methods are refined through semantic-aware knowledge by each component, and the total framework leads to a significant performance boost. Additionally, results in Fig. 6 demonstrate that model with SCH loss and SE module can pre-

(a) Baseline   (b) W/ SCH loss   (c) W/ SE   (d) W/SCH loss & SE   (e) W/SCH loss & SA loss & SE   (f) GT

Figure 6. Visual comparison of DRBN-SKF for investigating the contribution of key techniques of our SKF.

Table 3. Ablation study of KinD++-SKF, DRBN-SKF and HWMNet-SKF for investigating the contribution of key techniques of our SKF.

SCH loss	SA loss	SE module	KinD++-SKF				DRBN-SKF				HWMNet-SKF			
			PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓
			18.970	0.804	0.175	4.760	19.860	0.834	0.155	4.793	24.240	0.852	0.114	5.141
✓			19.170	0.806	**0.170**	4.759	20.040	0.835	0.154	4.793	24.590	0.859	0.112	5.023
	✓		19.385	0.800	0.189	4.392	20.070	0.834	0.149	4.701	24.305	0.853	0.111	4.712
		✓	19.781	0.808	0.181	4.712	21.334	0.837	0.143	4.678	24.477	0.859	0.111	4.988
✓		✓	**20.620**	**0.815**	0.176	4.536	22.550	0.836	0.150	4.581	**25.123**	0.860	0.111	4.711
✓	✓	✓	20.363	0.805	0.201	**4.142**	**22.837**	**0.841**	**0.138**	**4.464**	25.086	**0.860**	**0.108**	**4.346**

Table 4. Ablation study of HWMNet-SKF for investigating the effect of semantic priors in the loss function.

$\mathcal{L}_{SCH}$		$\mathcal{L}_{SA}$			PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓
w/o S	w/ S	w/o SA	w/o S	w/ S				
					24.240	0.852	0.114	5.141
✓		✓			24.477	0.859	0.111	4.988
✓			✓		24.568	0.857	0.113	4.613
✓				✓	24.668	0.859	0.111	4.567
	✓	✓			**25.123**	**0.860**	**0.108**	4.711
	✓		✓		25.040	0.859	0.111	4.546
	✓			✓	25.086	**0.860**	**0.108**	**4.311**

Table 5. Ablation study for investigating whether the performance improvement comes from semantic priors or more parameters.

Method		PSNR ↑	SSIM ↑	LPIPS ↓	Param(M)
HWMNet	Baseline	24.240	0.852	0.114	66.56
	Large	24.445	0.853	0.115	69.99
	w/ SKF	**25.086**	**0.860**	**0.108**	69.98
LLFlow-S	Baseline	24.060	0.860	0.136	4.97
	Large	24.167	0.858	0.137	5.38
	w/ SKF	**25.942**	**0.865**	**0.125**	5.26
LLFlow-L	Baseline	24.999	0.870	0.117	37.68
	Large	25.292	0.873	0.113	40.55
	w/ SKF	**26.798**	**0.879**	**0.105**	39.91

serve color consistency and details and the SA loss reduces fake regions by producing more natural textures.

Semantic-guided losses. Table 4 lists the results of different settings of losses. The w/o S and w/ S denote calculating the global histogram and our semantic-guided histogram respectively. For SA loss, the w/o SA, w/o S and w/ S denote without SA loss, classic global and local adversarial loss like EnlightGAN [16] and our SA loss. First, the HWMNet-SKF with SCH loss presents better performance, achieving average margin of 0.512 dB improvement on PSNR, indicating the significant capability of SCH loss in preserving color consistency. Furthermore, the average gain of 0.271 on NIQE by adding classic adversarial loss can be attributed to the capability of discriminator to improve visual quality. Finally, our SA loss provides favorable gain of 0.411 on NIQE over the baseline and faithfully demonstrates that the semantic priors help find out fake regions and thus produce more natural images.

Superiority of semantic priors. We choose HWMNet-SKF, LLFlow-S-SKF and LLFlow-L-SKF to investigate whether the improvement of performance benefits from semantic priors provided by our SKF or the more parameters of our SE modules. As shown in Table 5, Baseline, Large and w/ SKF denote the original model, original model with more layers or channels and original model with our SKF. Our methods achieve significant improvement by average margin of 1.272 dB on PSNR comparing to large versions with similar number of parameters. Hence, we prove the superiority of the semantic priors instead of extra parameters.

5. Conclusion

This work has proposed a novel framework for semantic-aware image enhancement, named SKF. The SKF incorporates semantic priors into Enhancement Net to preserve color consistency and visual details by SE module, SCH loss and SA loss. SE module allows image features to perceive rich and spatial information by semantic feature representations. SCH loss offers effective semantic-aware regional constrain for preserving color consistency. SA loss combines global and local adversarial loss and semantic priors to seek target fake region and generates natural results. Extensive experiments show that our SKF achieves superior performance in the case of all six baseline methods, and the LLFlow-L-SKF outperforms all the competitors. However, the improvement is limited when dealing with unknown category, inducing more possibility when improving the capability of identifying unknown instance by SKB. Furthermore, we will also explore the potential of our SKF in other low-level vision tasks.

References

[1] Andreas Aakerberg, Anders S Johansen, Kamal Nasrollahi, and Thomas B Moeslund. Semantic segmentation guided real-world super-resolution. In *WACV*, pages 449–458, 2022. 3

[2] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *TCE*, 53:593–600, 2007. 1, 2

[3] Mor Avi-Aharon, Assaf Arbelle, and Tammy Riklin Raviv. DeepHist: Differentiable joint and color histogram layers for image-to-image translation. *arXiv preprint arXiv:2005.03995*, 2020. 5

[4] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *CVPR*, pages 4690–4699, 2019. 1

[5] Xingbo Dong, Wanyan Xu, Zhihui Miao, Lan Ma, Chao Zhang, Jiewen Yang, Zhe Jin, Andrew Beng Jin Teoh, and Jiajun Shen. Abandoning the bayer-filter to see in the dark. In *CVPR*, pages 17431–17440, 2022. 1, 2, 3

[6] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *CVPR*, pages 5759–5768, 2022. 2

[7] Chi-Mao Fan, Tsung-Jung Liu, and Kuan-Hsien Liu. Half wavelet attention on M-Net+ for low-light image enhancement. *arXiv preprint arXiv:2203.01296*, 2022. 1, 2, 3, 6, 7

[8] Minhao Fan, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Integrating semantic segmentation and Retinex model for low-light image enhancement. In *ACMMM*, pages 2317–2325, 2020. 1, 2, 3, 6

[9] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *CVPR*, pages 3146–3154, 2019. 1

[10] Vitor Guizilini, Rui Hou, Jie Li, Rares Ambrus, and Adrien Gaidon. Semantically-guided representation learning for self-supervised monocular depth. In *ICLR*, 2019. 3

[11] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, pages 1780–1789, 2020. 3, 6

[12] Xiaojie Guo. LIME: A method for low-light image enhancement. In *ACMMM*, pages 87–91, 2016. 5, 7

[13] Xiaojie Guo, Yu Li, and Haibin Ling. LIME: Low-light image enhancement via illumination map estimation. *TIP*, 26:982–993, 2016. 6

[14] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Trans. Graph.*, 36:1–14, 2017. 5

[15] Kui Jiang, Zhongyuan Wang, Zheng Wang, Chen Chen, Peng Yi, Tao Lu, and Chia-Wen Lin. Degrade is upgrade: Learning degradation for low-light image enhancement. In *AAAI*, volume 36, pages 1078–1086, 2022. 2

[16] 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. *TIP*, 30:2340–2349, 2021. 2, 3, 5, 6, 8

[17] Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: when layer decomposition meets light-effects suppression. *arXiv preprint arXiv:2207.10564*, 2022. 3

[18] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround Retinex. *TIP*, 6:451–462, 1997. 1, 2

[19] Hyunyoung Jung, Eunhyeok Park, and Sungjoo Yoo. Fine-grained semantics-aware representation enhancement for self-supervised monocular depth estimation. In *ICCV*, pages 12642–12652, 2021. 2, 3

[20] Bomi Kim, Sunhyeok Lee, Nahyun Kim, Donggon Jang, and Dae-Shik Kim. Learning color representations for low-light image enhancement. In *WACV*, pages 1455–1463, 2022. 2, 3

[21] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2D histograms. *TIP*, 22:5372–5384, 2013. 5, 7

[22] Chongyi Li, Chunle Guo, Ling-Hao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *TPAMI*, 2021. 1, 4

[23] Xiaoming Li, Chaofeng Chen, Shangchen Zhou, Xianhui Lin, Wangmeng Zuo, and Lei Zhang. Blind face restoration via deep multi-scale component dictionaries. In *ECCV*, pages 399–415. Springer, 2020. 3

[24] Yi Li, Yi Chang, Changfeng Yu, and Luxin Yan. Close the loop: A unified bottom-up and top-down paradigm for joint image deraining and segmentation. In *AAAI*, 2022. 3

[25] Yijun Li, Sifei Liu, Jimei Yang, and Ming-Hsuan Yang. Generative face completion. In *CVPR*, pages 3911–3919, 2017. 5

[26] Dong Liang, Ling Li, Mingqiang Wei, Shuo Yang, Liyan Zhang, Wenhan Yang, Yun Du, and Huiyu Zhou. Semantically contrastive learning for low-light image enhancement. In *AAAI*, pages 1555–1563, 2022. 2, 3

[27] Seokjae Lim and Wonjun Kim. DSLR: Deep stacked Laplacian restorer for low-light image enhancement. *TMM*, 23:4272–4284, 2020. 2, 3

[28] Ding Liu, Bihan Wen, Xianming Liu, Zhangyang Wang, and Thomas S. Huang. When image denoising meets high-level vision tasks: A deep learning approach. In *IJCAI*, 2018. 3

[29] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *CVPR*, pages 10561–10570, 2021. 1, 2

[30] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. LL-Net: A deep autoencoder approach to natural low-light image enhancement. *PR*, 61:650–662, 2017. 2

[31] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *TIP*, 24:3345–3356, 2015. 5, 7

[32] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *CVPR*, pages 5637–5646, 2022. 2

[33] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a 'completely blind' image quality analyzer. *SPL*, 20:209–212, 2012. 6

[34] Sean Moran, Pierre Marza, Steven McDonagh, Sarah Parisot, and Gregory Slabaugh. DeepLPF: Deep local parametric filters for image enhancement. In *CVPR*, pages 12826–12835, 2020. 1, 2, 3

[35] Roozbeh Mottaghi, Xianjie Chen, Xiaobai Liu, Nam-Gyu Cho, Seong-Whan Lee, Sanja Fidler, Raquel Urtasun, and Alan Yuille. The role of context for object detection and semantic segmentation in the wild. In *CVPR*, pages 891–898, 2014. 2

[36] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *NIPS*, 28, 2015. 1

[37] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. MAXIM: Multi-axis mlp for image processing. In *CVPR*, pages 5769–5780, 2022. 2

[38] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *TPAMI*, 43:3349–3364, 2020. 2, 4

[39] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *TIP*, 22:3538–3548, 2013. 5, 7

[40] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *CVPR*, pages 606–615, 2018. 2, 3

[41] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *AAAI*, volume 36, pages 2604–2612, 2022. 1, 2, 3, 6

[42] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *TIP*, 13:600–612, 2004. 5

[43] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018. 1, 2, 5, 6, 7

[44] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement. In *CVPR*, pages 5901–5910, 2022. 1, 2

[45] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, pages 2281–2290, 2020. 1, 2

[46] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. SNR-aware low-light image enhancement. In *CVPR*, pages 17714–17724, 2022. 1, 2, 6

[47] Canqian Yang, Meiguang Jin, Xu Jia, Yi Xu, and Ying Chen. AdaInt: Learning Adaptive Intervals for 3D Lookup Tables on Real-time Image Enhancement. In *CVPR*, pages 17522–17531, 2022. 3

[48] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-

[49] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep Retinex network for robust low-light image enhancement. *TIP*, 30:2072–2086, 2021. 1, 5, 6, 7

[50] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pages 5728–5739, 2022. 4

[51] Syed Waqas Zamir, Aditya Arora, Salman Hameed Khan, Hayat Munawar, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning Enriched Features for Fast Image Restoration and Enhancement. *TPAMI*, 2022. 2, 6

[52] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 5

[53] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *IJCV*, 129:1013–1037, 2021. 1, 2, 6, 7

[54] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACMMM*, pages 1632–1640, 2019. 1, 2, 6, 7

[55] Zhao Zhang, Huan Zheng, Richang Hong, Mingliang Xu, Shuicheng Yan, and Meng Wang. Deep color consistent network for low-light image enhancement. In *CVPR*, pages 1899–1908, 2022. 2, 3, 4

[56] Zunjin Zhao, Bangshu Xiong, Lei Wang, Qiaofeng Ou, Lei Yu, and Fa Kuang. RetinexDIP: A unified deep framework for low-light image enhancement. *TCSVT*, 32:1076–1088, 2021. 2

[57] Chuanjun Zheng, Daming Shi, and Wentian Shi. Adaptive unfolding total variation network for low-light image enhancement. In *ICCV*, pages 4439–4448, 2021. 2, 3

[58] Shen Zheng and Gaurav Gupta. Semantic-guided zero-shot learning for low-light image/video enhancement. In *WACV*, pages 581–590, 2022. 2, 3

[59] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *AAAI*, volume 34, pages 13106–13113, 2020. 2

supervised approach for low-light image enhancement. In *CVPR*, pages 3063–3072, 2020. 1, 3, 6, 7