

You Do Not Need Additional Priors or Regularizers in Retinex-based Low-light Image Enhancement

Huiyuan Fu¹, Wenkai Zheng¹, Xiangyu Meng¹, Xin Wang², Chuanming Wang¹, Huadong Ma¹

¹Beijing University of Posts and Telecommunications, ²Stony Brook University

{fhy, ciki, clearlove7, wcm, mhd}@bupt.edu.cn, x.wang@stonybrook.edu

Abstract

Images captured in low-light conditions often suffer from significant quality degradation. Recent works have built a large variety of deep Retinex-based networks to enhance low-light images. The Retinex-based methods require decomposing the image into reflectance and illumination components, which is a highly ill-posed problem and there is no available ground truth. Previous works addressed this problem by imposing some additional priors or regularizers. However, finding an effective prior or regularizer that can be applied in various scenes is challenging, and the performance of the model suffers from too many additional constraints. We propose a contrastive learning method and a self-knowledge distillation method for Retinex decomposition that allow training our Retinex-based model without elaborate hand-crafted regularization functions. Rather than estimating reflectance and illuminance images and representing the final images as their element-wise products as in previous works, our regularizer-free Retinex decomposition and synthesis network (RFR) extracts reflectance and illuminance features and synthesizes them end-to-end. In addition, we propose a loss function for contrastive learning and a progressive learning strategy for self-knowledge distillation. Extensive experimental results demonstrate that our proposed methods can achieve superior performance compared with state-of-the-art approaches.

1. Introduction

High quality images are highly desirable for many computer vision and machine learning applications. In a low-light environment, images often suffer from visual degradations such as poor visibility and low contrast. The darker the area in an image, the more the information is lost, and the harder it is to recover the image with a good quality. Many computer vision systems may malfunction or fail completely in low-light conditions. If dark regions can be removed or alleviated after the low-light image enhancement, images can

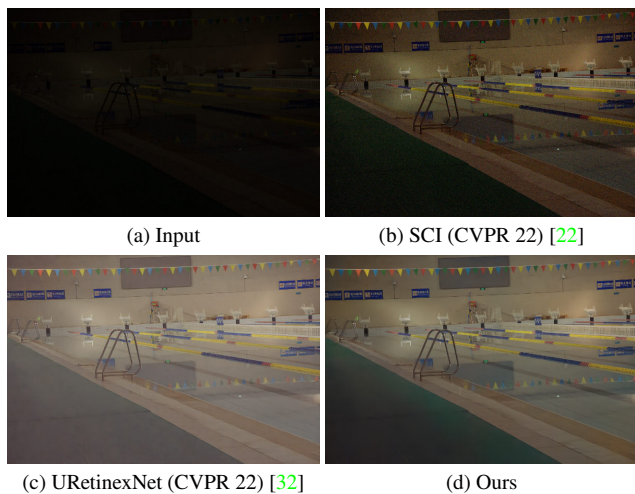


Figure 1. Visual comparison on a real-captured low-light image using state-of-the-art approaches.

become more clear. Thus low-light image enhancement can potentially benefit many computer vision applications.

Traditional methods for low-light image enhancement can be divided into two categories, one based on histogram equalization [1, 2, 7, 12, 25] and the other based on Retinex theory [4, 6, 9, 10, 13, 28]. However, relying on the hand-crafted features, they are not robust against the degraded visibility and unexpected noise.

Recently, deep learning based approaches for low-light image enhancement have received extensive attention. Instead of using prior handcrafted features, these methods, such as directly learning enhanced results in an end-to-end network and deep Retinex-based methods, can automatically learn the features with deep neural networks [5, 8, 16, 17, 19, 20, 22, 26, 27, 29, 31–35, 37–39] for low-light image enhancement. A specific survey can be found in [14].

Since Retinex theory [11] models the color perception of human vision on natural scenes, deep Retinex-based methods have better enhancement performance and generalization in most cases than directly learning enhanced results in an end-to-end network. However, Retinex decomposition

is a highly ill-posed problem and there is no well-defined ground truth for the reflectance and illumination of real images. Previous works [26, 31–33, 37–39] addressed this problem by introducing some additional priors or regularizers, but these manually well-designed regularization functions are difficult to apply to all scenes. The joint optimization with too many constraints also leads to the absence of adaptivity and efficiency. Therefore previous Retinex-based methods often generate unnatural images.

To tackle these challenging issues, we propose a contrastive learning method and a self-knowledge distillation method for Retinex decomposition. With these methods, our proposed regularizer-free Retinex decomposition and synthesis network (RFR) can be trained without additional priors or regularizers. Considering the effect of image degradations on the Retinex decomposition, instead of estimating reflectance and illumination images and representing the final image as their element-wise product, our model extracts reflectance and illumination features and synthesizes them through a neural network. Rather than using multiple networks to perform decomposition, enhancement and denoising separately for low-light image enhancement, this method can be learned through an end-to-end network. Moreover, a novel loss function for contrastive learning is introduced, where some low-quality negative samples are taken into account. Then, a progressive learning strategy for self-knowledge distillation is presented, which allows the student to learn the knowledge of the teacher more effectively. We conduct extensive experiments to demonstrate that our proposed methods are beneficial and reasonable. Fig. 1 shows the visual comparisons of images recovered from the low-light one using different approaches.

In summary, the contributions of our work are as follows:

- We propose a contrastive learning method and a self-knowledge distillation method that enable the network to extract high-quality reflectance and illumination without additional priors or regularizers.
- We propose a regularizer-free Retinex decomposition and synthesis network (RFR) for low-light image enhancement that extract reflectance and illumination features and synthesizes them end-to-end to suppress the effect of image degradations on the Retinex theory.
- We propose a loss function named Weighted Normalized Temperature-Scaled Cross-Entropy Loss for contrastive learning and a progressive learning strategy for self-knowledge distillation.
- Comprehensive experiments on different datasets show the superiority of our proposed methods compared with state-of-the-art approaches.

2. Methodology

2.1. Motivation

Following the physically explicable Retinex theory, deep Retinex-based methods have better enhancement performance and generalization in most cases than directly learning enhanced results in end-to-end networks. Retinex theory states that an image \mathbf{I} can be decomposed into reflectance \mathbf{R} and illumination \mathbf{L} , and expressed as their element-wise product as:

$$\mathbf{I} = \mathbf{R} \circ \mathbf{L} \quad (1)$$

Retinex-based methods usually require decomposing the observed image into its reflectance and illumination components. This problem is highly ill-posed and there is no ground truth for reflectance and illumination of real images. Previous works addressed the problem by introducing some additional priors or regularizers, which imposed some elaborately hand-crafted constraints to train their decomposition networks. The objective function of their Retinex decomposition can be summarized as:

$$\min_{\hat{\mathbf{R}}, \hat{\mathbf{L}}} \|\mathbf{I} - \hat{\mathbf{R}} \circ \hat{\mathbf{L}}\|_F^2 + \Phi(\hat{\mathbf{R}}) + \Psi(\hat{\mathbf{L}}) \quad (2)$$

where $\hat{\mathbf{R}}, \hat{\mathbf{L}}$ are the estimated reflectance and illuminance, $\|\cdot\|_F$ represents Frobenius norm, and $\Phi(\hat{\mathbf{R}}), \Psi(\hat{\mathbf{L}})$ denote the regularizers of reflectance and illuminance respectively.

However, it is challenging to find an effective prior or regularizer that can be applied in various scenes. Inaccurate priors or regularizers may lead to artifacts and color deviations in the enhanced results. In addition, the joint optimization with too many additional constraints may lead to possible suboptimal solutions to balance the mismatched goals of them. Therefore, these solutions often suffer from the absence of adaptivity and efficiency and require introducing some tricks during the model training (e.g., adding pure black or pure white images to the training data). We believe that they are not optimal and not elegant. In this work, we would like to use only the most fundamental assumptions in the Retinex theory as the objective function for Retinex decomposition without introducing additional priors or regularizers, as follows:

$$\min_{\hat{\mathbf{R}}_l, \hat{\mathbf{L}}_l, \hat{\mathbf{R}}_n, \hat{\mathbf{L}}_n} \sum_{i \in \{l, n\}} \|\mathbf{I}_i - \hat{\mathbf{R}}_i \circ \hat{\mathbf{L}}_i\|_F^2 + \|\hat{\mathbf{R}}_l - \hat{\mathbf{R}}_n\|_F^2 \quad (3)$$

where \mathbf{I}_l and \mathbf{I}_n denote the low-light image and its corresponding normal-light image respectively, $\hat{\mathbf{R}}_l$ and $\hat{\mathbf{R}}_n$ denote the estimated reflectance of \mathbf{I}_l and \mathbf{I}_n respectively, and $\hat{\mathbf{L}}_l$ and $\hat{\mathbf{L}}_n$ denote the estimated illumination of \mathbf{I}_l and \mathbf{I}_n respectively. $\|\mathbf{I} - \hat{\mathbf{R}} \circ \hat{\mathbf{L}}\|_F^2$ is derived from Eq. (1). And $\|\hat{\mathbf{R}}_l - \hat{\mathbf{R}}_n\|_F^2$ indicates that the reflectance of the low-light and normal-light images of the same scene is identical, which is obvious.

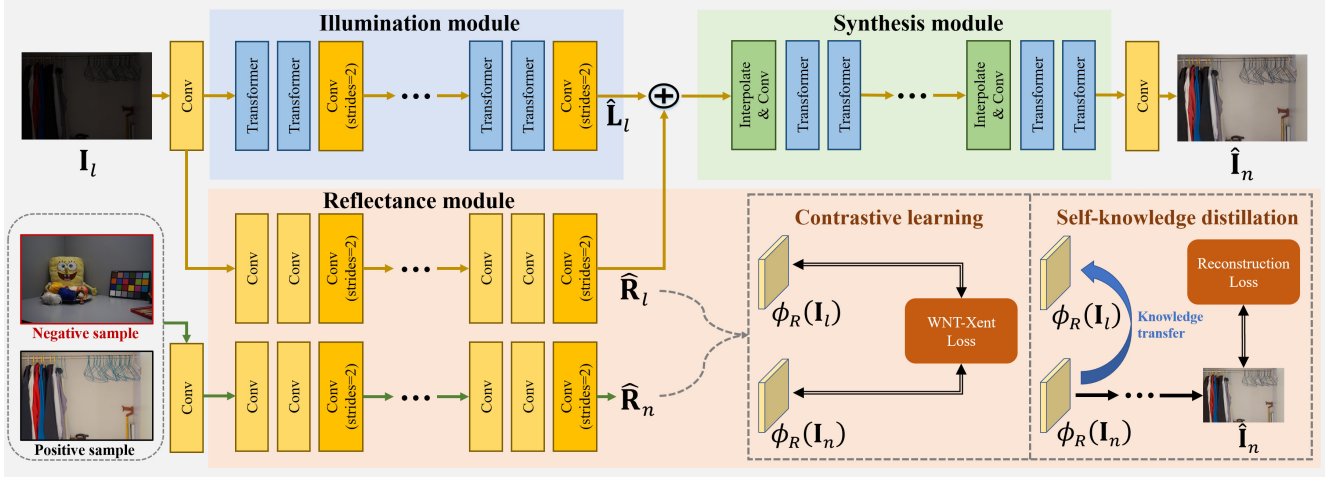


Figure 2. The framework of our RFR, which consists of the reflectance module, the illumination module and the synthesis module. The reflectance module trained using our proposed contrastive learning method or self-knowledge distillation method extracts reflectance features, the illumination module extracts illumination features, and the synthesis module synthesizes them to generate the enhanced images.

In addition, Retinex theory is an assumption for an ideal situation, but low-light images in practice tend to have significant noise and loss of details due to insufficient lighting. If the effects of these image degradations are not taken into account, it will result in inaccurate estimates of reflectance and illumination. Although some Retinex-based works perform processes such as image denoising, it is difficult to completely eliminate the effects of image degradation. Simply representing an enhanced image by the element-wise multiplication of reflectance and illumination according to Eq. (1) would amplify the deviation of the estimated reflectance and illumination, as shown in Eq. (4). The noise, loss of details in the reflectance image, and blurring of the illumination image would all be amplified in the final image, resulting in poor visual effects.

$$\begin{aligned}\hat{\mathbf{I}}_n &= \hat{\mathbf{R}}_l \circ \hat{\mathbf{L}}_n = (\mathbf{R}_l + \varepsilon_1) \circ (\mathbf{L}_n + \varepsilon_2) \\ &= \mathbf{I}_n + \underbrace{\mathbf{R}_l \circ \varepsilon_2 + \mathbf{L}_n \circ \varepsilon_1 + \varepsilon_1 \circ \varepsilon_2}_{\text{deviation}}\end{aligned}\quad (4)$$

where $\varepsilon_1, \varepsilon_2$ represent the deviations of the estimated reflectance and illumination, respectively.

In order to reduce the effect of deviations in reflectance and illumination on the final result, instead of estimating the reflectance and illumination images and representing the enhanced image as their element-wise product as in previous works, we estimate the reflectance and illumination features and synthesize them into the final image by neural network.

2.2. Retinex Decomposition with Contrastive Learning

To improve the performance and generalization of Retinex-based models, we use only the most fundamental

assumptions in the Retinex theory without introducing additional priors or regularizers as in Eq. (3), and generate the enhanced images by synthesizing the reflectance and illumination feature maps by neural network. The objective function of our Retinex decomposition is as follows:

$$\min_{\phi_S, \phi_R, \phi_L} \|\mathbf{I}_n - \phi_S(\phi_R(\mathbf{I}_l), \phi_L(\mathbf{I}_l))\|_F^2 + \|\phi_R(\mathbf{I}_n) - \phi_R(\mathbf{I}_l)\|_F^2 \quad (5)$$

where ϕ_R, ϕ_L denote the modules used to estimate reflectance and illumination respectively, and their outputs are reflectance and illumination features. And ϕ_S denotes the module used to synthesize reflectance and illumination.

However, this objective function cannot be directly used to train the model because the reflectance module of the network only needs to output the same reflectance for all images to make $\|\phi_R(\mathbf{I}_n) - \phi_R(\mathbf{I}_l)\|_F^2 = 0$. In order to extract high-quality reflectance, we propose a contrastive learning based Retinex decomposition method without introducing additional priors or regularizers.

Specifically, the estimated reflectance should be luminance-independent (i.e., $\min \|\phi_R(\mathbf{I}_n) - \phi_R(\mathbf{I}_l)\|_F^2$) and meaningful (i.e., it can help the subsequent optimization of $\min \|\mathbf{I}_n - \phi_S(\phi_R(\mathbf{I}_l), \phi_L(\mathbf{I}_l))\|_F^2$). For a low-light image, we take the corresponding normal-light image as a positive sample, and the rest of the normal-light images in the minibatch as negative samples, and then extract the luminance-independent features through contrastive learning. The network extracts similar features from positive samples, and completely different features from negative samples. An image and its positive sample, i.e. a low-light image and its corresponding normal-light image, will extract the same luminance-independent features in perfect accordance with the assumption of reflectance consistency in Retinex theory (i.e. $\min \|\mathbf{R}_l - \mathbf{R}_n\|_F^2$). Due to contrastive

learning, the reflectance module no longer outputs the same features for all input images and the estimated reflectance will include the spatial features of the image.

We also propose a loss function for contrastive learning named Weighted Normalized Temperature-Scaled Cross-Entropy Loss (WNT-Xent Loss). We notice that most of the low-light image datasets contain a lot of similar or identical scene images, and it is unreasonable to treat these images as negative samples. We therefore improve the commonly used NT-Xent loss [3] in contrastive learning by weighting the negative samples using the similarity of the original normal-light images, and our WNT-Xent loss is shown below:

$$\mathcal{L}_{\text{WNT-Xent}} = -\log \frac{\exp(\|\phi_R(\mathbf{I}_{I_i}) - \phi_R(\mathbf{I}_{n_i})\|_F^2/\tau)}{\sum_{k=1}^B (1 - \omega_{i,k}) \exp(\|\phi_R(\mathbf{I}_{I_i}) - \phi_R(\mathbf{I}_{n_k})\|_F^2/\tau)} \quad (6)$$

where $\omega_{i,k} \in [0, 1]$ represents the structural similarity [30] of normal-light images \mathbf{I}_{n_i} and \mathbf{I}_{n_k} , B is the batch size and τ denotes a temperature parameter.

2.3. Retinex Decomposition with Self-knowledge Distillation

The reflectance extracted by the model trained directly with Eq. (3) is luminance-independent, but it is meaningless for the subsequent enhancement. Although our contrastive learning approach can solve this problem, its performance is affected by the quality of the training set. In order for the estimated reflectance to help the subsequent optimization of $\min \|\mathbf{I}_n - \phi_S(\phi_R(\mathbf{I}_I), \phi_L(\mathbf{I}_I))\|_F^2$, we also propose another Retinex decomposition method based on self-knowledge distillation. Specifically, we use a branch of the network to estimate the reflectance of the low-light image (student branch) and a branch with the same architecture to estimate the reflectance of the corresponding normal-light image (teacher branch). The reflectance features extracted by the teacher branch are used to guide the training of the student branch.

In order to allow the teacher branch to extract the reflectance information instead of outputting the same feature map for all inputs, we use the reflectance estimated by the teacher branch to generate the final image and use the reconstruction loss of the global network to guide the training of the teacher branch. The objective function of the knowledge distillation process is shown below:

$$\mathcal{L}_{\text{kd}} = \lambda (\|\phi_R(\mathbf{I}_I) - \phi_R(\mathbf{I}_n)\|_F^2/\tau) + \mathcal{L}_{\text{rec}} \quad (7)$$

where λ is a trade-off parameter and \mathcal{L}_{rec} is the reconstruction loss defined in Eq. (9).

However, we find it difficult to distill information from the teacher branch directly to the student branch. [23] shows that the performance of student networks decreases when there is a large gap between students and teachers. In our

knowledge distillation, the severe image degradations of the low-light images lead to a very large gap between the reflectance features extracted through the teacher branch and through the student branch. This makes the student branch difficult to converge during the training.

To bridge the gap, we propose a progressive learning strategy. Specifically, in the process of self-knowledge distillation, we mix the low-light image with the corresponding normal-light image as input to the student branch, and we gradually increase the proportion of low-light images thus increasing the difficulty for the student branch to learn the knowledge of the teacher branch. The input can be expressed as:

$$\mathbf{I}_{\text{input}} = \alpha \mathbf{I}_n + (1 - \alpha) \mathbf{I}_I, \quad \alpha = \frac{1 + \cos\left(\frac{T_{\text{cur}}}{T_{\text{total}}} \pi\right)}{2} \quad (8)$$

where T_{cur} represents the current epoch of training, and T_{total} represents the total epoch number.

It is worth mentioning that our progressive learning strategy can not only significantly improve the performance of our self-knowledge distillation, but also be used in our contrastive learning based Retinex decomposition method, as it can be considered as a data augmentation method that increases the number of positive and negative samples. In addition, our strategy allows the model to be trained with images of different brightness, further improving the generalization of the model. As a result, the images generated by our model are rarely underexposed or overexposed even in cross-dataset evaluation.

2.4. Overall Architecture

Based on the above approaches, we propose an end-to-end regularizer-free Retinex decomposition and synthesis network (RFR) for low-light image enhancement. We use RFR_{CL} and RFR_{SD} to denote RFR trained with our proposed contrastive learning and RFR trained with our proposed self-knowledge distillation, respectively. As shown in Fig. 2, our proposed RFR consists of three parts: the reflectance module, the illumination module, and the synthesis module.

Reflectance module The reflectance module consists of two branches with the same structure but without shared parameters. Each branch contains four stages, and each stage contains two convolution layers and a down-sampling layer. In the training phase of Retinex decomposition, the two branches receive low-light images and normal images respectively to extract reflectance features. In other phases, only the branch for low-light images is used to estimate the reflectance.

Illumination module Since the global illumination information is critical for the low-light image enhancement task, we introduce transformer layers to work with the convolution layers to capture longer dependencies without los-

ing local dependencies. The illumination module is a hierarchical structure containing four stages, and each stage contains two transformer layers and a convolution layer. The transformer layers extract illumination features, which are then down-sampled by the convolution layer.

Synthesis module The synthesis module also has four stages, and each stage contains an up-sampling layer and two transformer layers. We use a combination of nearest neighbor interpolation and a convolutional layer instead of a deconvolution layer for up-sampling, because we find that the mechanism of splitting patches in the transformer layer exacerbates the checkerboard artifacts of deconvolution. In addition, a skip connection is used for the corresponding stage of the illumination module and the synthesis module.

We fine-tune the entire network after training the reflectance module with our methods. In the fine-tuning phase and testing phase, only the low-light images are input into the model and sent to the illumination module and the branch for low-light images of the reflectance module for the extraction of the corresponding features, respectively. The following is the expression of our reconstruction loss in the fine-tuning phase.

$$\mathcal{L}_{\text{rec}} = (1 - \lambda) \sqrt{\|\mathbf{I}_n - \hat{\mathbf{I}}_n\|_F^2 + \epsilon^2} + \lambda \mathcal{L}_{\text{SSIM}}(\mathbf{I}_n, \hat{\mathbf{I}}_n) \quad (9)$$

where the weighting parameter λ is set to 0.8, the constant ϵ is set to 0.001, and $\mathcal{L}_{\text{SSIM}}$ represents the structural similarity loss [30].

3. Experiments

3.1. Implementation Details

We train our model on LOL dataset [31] containing 500 pairs of real-captured images. We augment the data using rotation and horizontal flipping and optimize the network by AdamW optimizer [18] with the momentum terms of (0.9, 0.999). We set the learning rate to 0.0001 in the contrastive learning and self-knowledge distillation phase and 0.00005 in the fine-tuning phase, and we use the cosine decay strategy to decrease it. We train RFR for 300 epochs in all phases on a Tesla V100 GPU.

3.2. Quantitative Evaluation

To fully evaluate the proposed approaches, we test our methods on the images from various scenes, including paired and unpaired datasets.

Evaluation on LOL. We use the test dataset in LOL dataset [31] for quantitative evaluations, since it has images of different light illuminations including particularly dark images, which most methods cannot recover very well. We compare the quality of images enhanced using RFR with state-of-the-art methods.

Table 1. Quantitative results of different methods on LOL dataset. **Retinex** indicates whether the method is based on Retinex theory.

Method	Retinex	PSNR↑	SSIM↑	LPIPS↓
LLNet [17]	×	17.96	0.713	0.360
RetinexNet [31]	✓	16.77	0.559	0.474
KinD [38]	✓	20.87	0.802	0.170
RRDNet [39]	✓	11.33	0.534	0.365
Zero-DCE [5]	×	14.86	0.589	0.335
EnlightenGAN [8]	×	17.48	0.677	0.322
KinD++ [37]	✓	21.30	0.823	0.160
RUAS [16]	✓	18.23	0.717	0.354
LLFlow [29]	×	25.19	0.925	0.113
SCI [22]	✓	14.78	0.646	0.339
SNR-Aware [34]	×	24.61	0.842	0.151
URetinexNet [32]	✓	21.33	0.835	0.122
RFR_{CL}	✓	26.33	0.928	0.109
RFR_{SD}	✓	26.41	0.928	0.107

Table 2. Quantitative results of cross-dataset evaluation on VE-LOL dataset.

Method	PSNR↑	SSIM↑	LPIPS↓
LLNet [17] (PR17)	17.57	0.739	0.402
RetinexNet [31] (BMVC18)	14.68	0.525	0.642
KinD [38] (ACMMM19)	18.42	0.766	0.288
RRDNet [39] (ICME20)	14.31	0.620	0.283
Zero-DCE [5] (CVPR20)	21.12	0.771	0.248
EnlightenGAN [8] (TIP21)	20.43	0.792	0.242
KinD++ [37] (IJCV21)	17.63	0.799	0.226
RUAS [16] (CVPR21)	14.36	0.671	0.337
LLFlow [29] (AAAI22)	23.85	0.899	0.146
SCI [22] (CVPR22)	16.37	0.711	0.270
SNR-Aware [34] (CVPR22)	23.52	0.874	0.216
URetinexNet [32] (CVPR22)	19.67	0.880	0.148
RFR_{CL}	26.22	0.918	0.144
RFR_{SD}	25.97	0.920	0.140

We adopt three well-known objective evaluation metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [30] and Learned Perceptual Image Patch Similarity (LPIPS) [36]. PSNR is the ratio between the maximum possible power of the normal light image and the power of the enhanced image and measures the fidelity between them. SSIM is a perception-based model that considers more about the structural message. LPIPS is a metric designed for human perception. So we can measure the quality of images from many perspectives, including structure, details and visual effects. Table 1 shows that our methods outperform all the other methods in all metrics.

Evaluation on VE-LOL. We use a cross-dataset evaluation to further evaluate the performance and generalization of our methods, in which we first train our model on the LOL training set and then evaluate it on the VE-LOL [15] test set

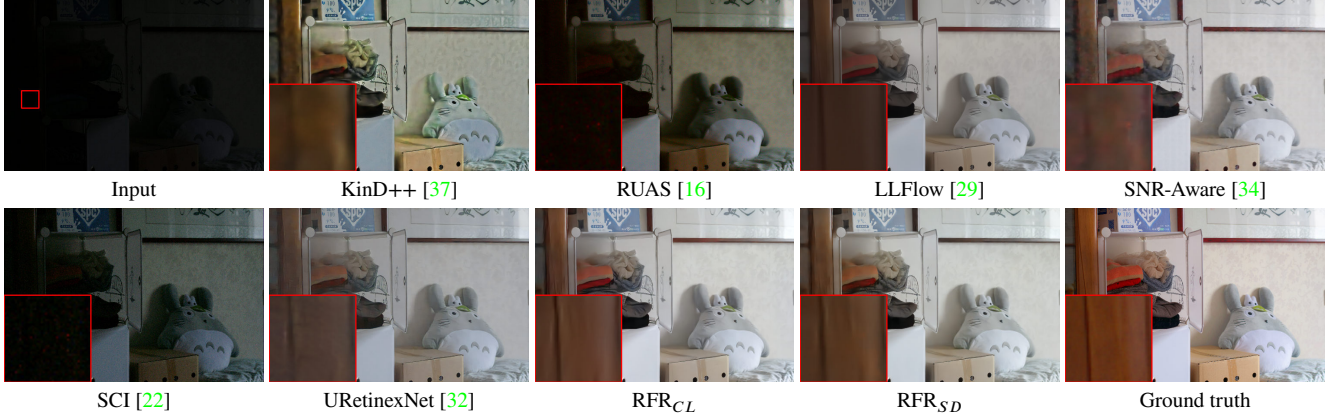


Figure 3. Visual comparison with state-of-the-art low-light image enhancement methods on LOL dataset.

Table 3. Quantitative results of cross-dataset evaluation on some unpaired datasets using NIQE metric. Smaller values represent better quality.

	DICM	LIME	MEF	NPE	All
LLFlow [29]	3.86	3.97	3.92	4.20	3.91
SCI [22]	4.10	<u>3.68</u>	3.63	4.47	4.01
SNR-Aware [34]	4.62	5.51	4.14	4.36	4.60
URetinexNet [32]	4.15	3.86	3.79	4.69	4.11
RFR_{CL}	<u>3.80</u>	3.65	<u>3.75</u>	4.07	3.79
RFR_{SD}	3.75	3.81	3.92	<u>4.13</u>	<u>3.81</u>

which contains 100 pairs of real-captured images and 100 pairs of synthetic images.

The evaluation results are shown in Table 2. We can see that our RFR outperforms other models in all metrics, indicating that it has excellent performance as well as good generalization capabilities.

Evaluation on unpaired datasets. We also perform a cross-dataset evaluation on some unpaired datasets including DICM [12], LIME [6], MEF [21] and NPE [28].

Since there is no ground truth of normal-light images in the unpaired datasets, we cannot evaluate the enhanced results by the metrics PSNR, SSIM, LPIPS. Therefore, we use the commonly used non-reference metric Natural Image Quality Evaluator (NIQE) [24] to evaluate the enhanced images. NIQE is based on the construction of a “quality aware” collection of statistical features based on a simple and successful space domain natural scene statistic model. We compare our methods with the latest state-of-the-art methods, which are all trained on the LOL training set. The evaluation results are shown in Table 3, which shows that our methods have excellent generalization capabilities and the enhanced images are visually more natural.

3.3. Qualitative Evaluation

We conduct mass qualitative evaluations on many paired and unpaired datasets. From the results on LOL dataset

shown in Fig. 3, RUAS and SCI cannot restore the brightness of extreme dark parts of the image, KinD++, SNR-Aware and URetinexNet have obvious artifacts or noise, and LLFlow does not recover a satisfactory overall brightness leading to obvious color deviations. In contrast, our methods achieve superior visual performance on both global illumination and details. Fig. 4 shows the visual results of the cross-dataset evaluation on different unpaired datasets. For the first three low-light images with different brightness, most of the low-light enhancement methods are overexposed and all except our model have serious color deviation. For the last image, only our methods recover the structure of the cave and do not introduce strange artifacts. Comparatively, our methods largely achieve a significant lead in the qualitative evaluation. Compared to the previous Retinex-based method, our methods generate more natural and realistic images without visible noise, nonrealistic color and texture distortions.

We also present the visual results of Retinex decomposition to evaluate the effectiveness of our proposed methods, since quantitative evaluation of the reflectance and illumination components is very difficult in the lack of ground truths for them. However, the previous Retinex-based methods estimate reflectance and illumination images while our methods estimate reflectance and illumination features, and we cannot directly compare the feature maps with the images qualitatively. Therefore, we propose a method to generate approximate illumination images of our illumination features and approximate reflectance images of our reflectance features. Specifically, for a pure white image \mathbf{I}_w , its reflectance image \mathbf{R}_w and illumination image \mathbf{L}_w are both pure white images. The element-wise product of any image and \mathbf{R}_w (or \mathbf{L}_w) is the image itself. According to Eq. (1), an image can be expressed as the element-wise product of reflectance and illumination. Therefore, a reflectance image can be expressed as the element-wise product of itself and \mathbf{L}_w and an illumination image can be expressed as the element-wise product of itself and \mathbf{R}_w . Our synthesis

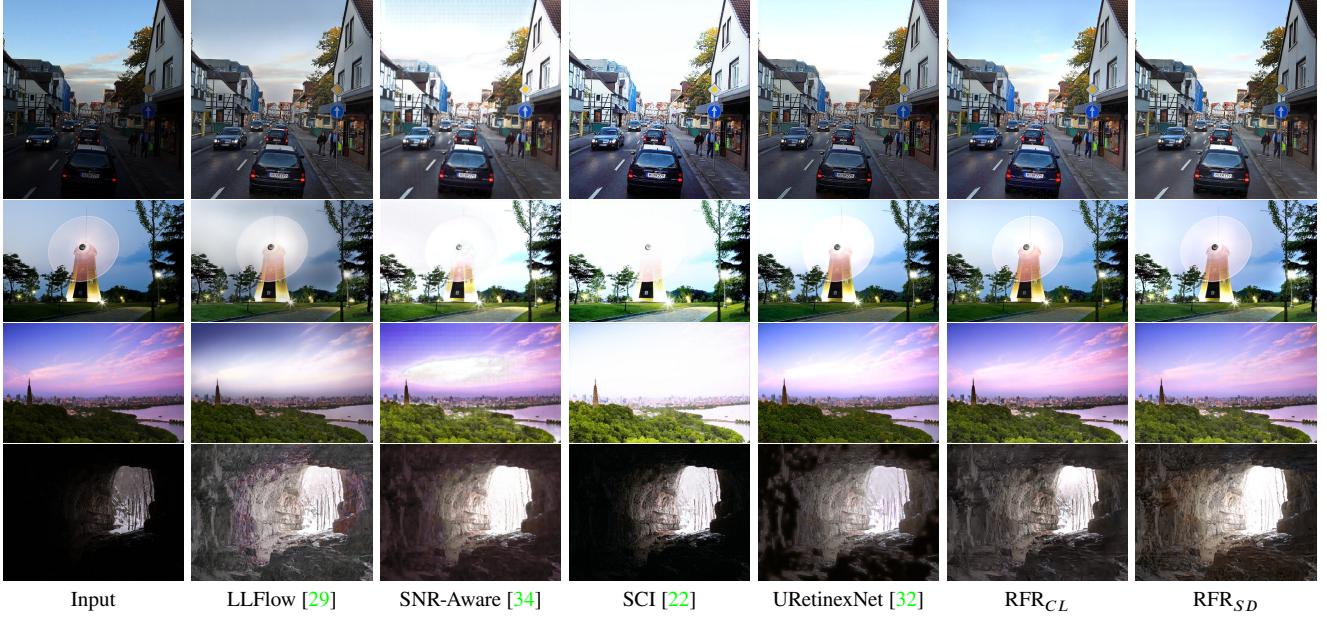


Figure 4. Visual results of cross-dataset evaluation on unpaired datasets. These 4 images are from LIME, DICM, NPE and MEF respectively.

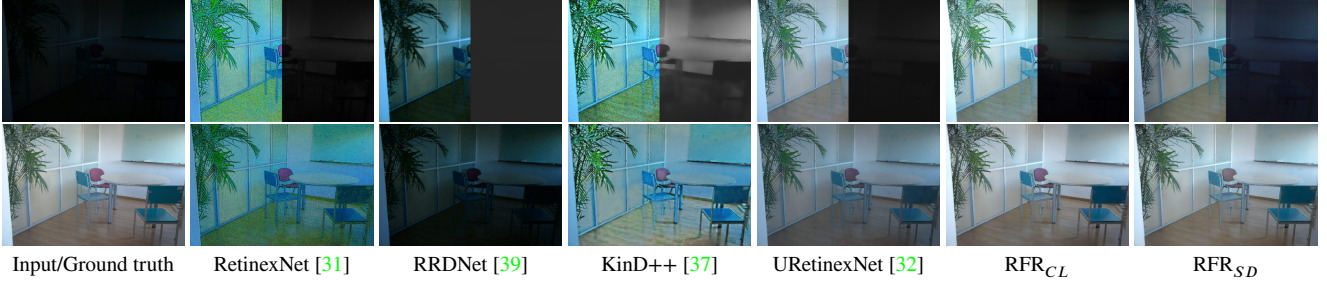


Figure 5. Comparisons of illumination and reflectance components estimated by different Retinex decomposition methods on VE-LOL dataset. The first column is the input low-light image and the ground truth. The left part of the images in the rest of the first row is the reflectance, and the right part is the illumination. The images in the rest of the second row are the corresponding enhanced results.

module can be approximated as the element-wise product of illumination and reflectance, so we can get an approximate reflectance or illumination image by replacing the input of the reflectance module or illumination module with a pure white image, as shown below:

$$\begin{aligned}\phi_S(\phi_R(\mathbf{I}_l), \phi_L(\mathbf{I}_w)) &\approx \mathbf{R}_l \circ \mathbf{L}_w = \mathbf{R}_l, \\ \phi_S(\phi_R(\mathbf{I}_w), \phi_L(\mathbf{I}_l)) &\approx \mathbf{R}_w \circ \mathbf{L}_l = \mathbf{L}_l\end{aligned}\quad (10)$$

As shown in Fig. 5, we present the reflectance, illumination and enhanced results of our methods and other Retinex-based methods. It should be noted that we only compare the methods that can output reflectance and illumination. It can be seen that RRDNet cannot extract good reflectance in the extremely dark environment, and the reflectance images estimated by RetinexNet and KinD++ both contain severe noise. URetinexNet and our methods can estimate high quality reflectance, but enhanced results of URetinexNet are obtained from Eq. (1) while our results are generated by the synthesis module. So our methods do not magnify the

defects of the reflectance and illumination, and can recover more satisfactory brightness without introducing artifacts.

3.4. User Study

We conduct a user study to compare the subjective visual quality of RFR and other Retinex-based methods. We randomly select 30 images from the LOL [31], VE-LOL [15], MEF [21], DICM [12], NPE [28] and LIME [6] test sets, 5 from each. The images are then enhanced using six Retinex-based methods (RUAS [16], KinD++ [37], SCI [22], URetinexNet [32], RFR_{CL} and RFR_{SD}). We ask 20 people to rate each image’s six enhancements. Specifically, we provide the original low-light images and present 6 kinds of enhancement results in random order for participants to rate their image quality from good to bad. The human subjects are instructed to consider 1) whether the images contain over or under-exposure artifacts 2) whether the images contain visible noise; and 3) whether the images show nonrealistic color or texture distortions. We calculate the average rank-

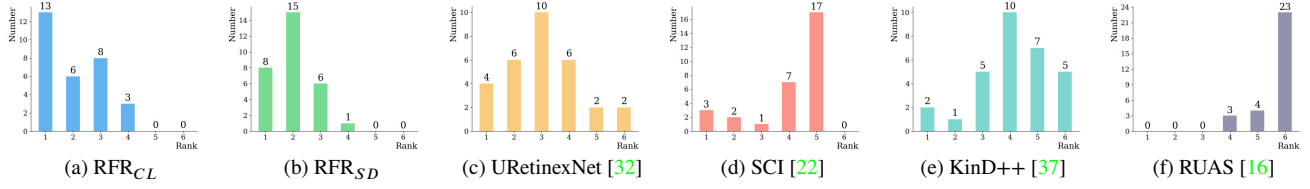


Figure 6. The results of six methods in the user study. In each histogram, the x-axis denotes the ranking index (1~6, 1 represents the highest), and the y-axis denotes the number of images in each ranking index.

Table 4. Quantitative results of the ablation study. **WX** denotes our WNT-Xent loss and **PL** denotes our progressive learning strategy.

Method	WX	PL	PSNR↑	SSIM↑	LPIPS↓
w/o Retinex	×	×	22.48	0.871	0.239
RFR _{CL}	×	×	25.33	0.922	0.122
RFR _{SD}	×	×	23.21	0.913	0.143
RFR _{CL}	✓	×	26.00	0.925	0.116
RFR_{CL}	✓	✓	26.33	0.928	0.109
RFR_{SD}	×	✓	26.41	0.928	0.107

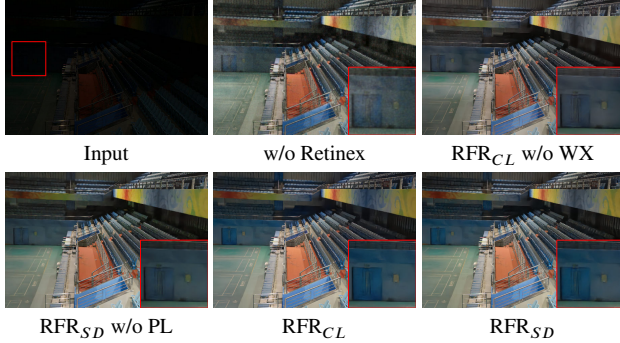


Figure 7. Visual results of the ablation study. **WX** denotes our WNT-Xent loss and **PL** denotes our progressive learning strategy.

ing of each method on each image and rank the results. As a result, each method is assigned with rank 1-6 on that image. The final results are shown in Fig. 6. RFR_{CL} has 13 results ranked 1, 6 results ranked 2, 8 results ranked 3, 3 results ranked 4, and no results ranked 5 or 6 out of the 30 images evaluated. RFR_{SD} has 8 results ranked 1, 15 results ranked 2, 6 results ranked 3, 1 result ranked 4, and no results ranked 5 or 6. When the histograms are compared, it is clear that our methods produce better results for human subjective evaluations across all datasets.

3.5. Ablation Study

We conduct comprehensive ablation studies to validate each component of our proposed RFR. The quantitative results are shown in Table 4 and the visual results are shown in Fig. 7.

To further validate the effectiveness of our approach, we try to use our synthesis module to improve previous Retinex-based methods. Specifically, for those methods that can output reflectance and illumination, we use our synthesis module to generate the final image with their estimated values.

Table 5. Quantitative results of different Retinex-based methods with our synthesis module. **SM** denotes our synthesis module.

Method	PSNR↑	SSIM↑	LPIPS↓
RetinexNet [31]	16.77	0.559	0.474
RetinexNet w/ SM	20.30	0.891	0.135
KinD [38]	20.87	0.802	0.170
KinD w/ SM	21.02	0.892	0.131
RRDNet [39]	11.33	0.534	0.365
RRDNet w/ SM	16.80	0.842	0.144
KinD++ [37]	21.30	0.823	0.160
KinD++ w/ SM	22.62	0.905	0.114
URetinexNet [32]	21.33	0.835	0.122
URetinexNet w/ SM	23.18	0.909	0.116

The quantitative results on LOL dataset [31] are shown in Table 5, and it is clear that our synthesis module can significantly improve the performance of the Retinex-based methods. In addition, the quality of reflectance and illumination is crucial to the results. The higher the quality of reflectance and illumination estimated by the previous Retinex-based methods, the higher the quality of the enhanced images generated by our synthesis module.

4. Conclusion

In this paper, we propose a contrastive learning method and a self-knowledge distillation method that allow training Retinex-based models without additional priors or regularizers. Instead of estimating reflectance and illumination images and representing the final image as their element-wise product, we propose an end-to-end Retinex-based network that extracts reflectance and illumination features and synthesizes them by a synthesis module. In addition, we propose a loss function for contrastive learning and a progressive learning strategy for self-knowledge distillation. A large number of experiments and comparisons prove the effectiveness of our proposed methods.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 62272059, the Funds for Creative Research Groups of China under Grant 61921003, the Beijing Nova Program under Grant Z201100006820124, and the 111 Project (B18008).

References

- [1] Tarik Arici, Salih Dikbas, and Yucel Altunbasak. A histogram modification framework and its application for image contrast enhancement. *IEEE Transactions on image processing*, 18(9):1921–1935, 2009. 1
- [2] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12):3431–3441, 2011. 1
- [3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 4
- [4] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2782–2790, 2016. 1
- [5] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020. 1, 5
- [6] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 1, 6, 7
- [7] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(4):1752–1758, 2007. 1
- [8] 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. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021. 1, 5
- [9] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 1
- [10] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997. 1
- [11] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 1
- [12] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013. 1, 6, 7
- [13] Chang-Hsing Lee, Jau-Ling Shih, Cheng-Chang Lien, and Chin-Chuan Han. Adaptive multiscale retinex for image contrast enhancement. In *2013 International Conference on Signal-Image Technology & Internet-Based Systems*, pages 43–50. IEEE, 2013. 1
- [14] 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. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (01):1–1, 2021. 1
- [15] Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision*, 129(4):1153–1184, 2021. 5, 7
- [16] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10561–10570, 2021. 1, 5, 6, 7, 8
- [17] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Ll-net: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 1, 5
- [18] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5
- [19] Kun Lu and Lihong Zhang. Tbefn: A two-branch exposure-fusion network for low-light image enhancement. *IEEE Transactions on Multimedia*, 23:4093–4105, 2020. 1
- [20] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, page 4, 2018. 1
- [21] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015. 6, 7
- [22] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5637–5646, 2022. 1, 5, 6, 7, 8
- [23] Seyed Iman Mirzadeh, Mehrdad Farajtabar, Ang Li, Nir Levine, Akihiro Matsukawa, and Hassan Ghasemzadeh. Improved knowledge distillation via teacher assistant. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 5191–5198, 2020. 4
- [24] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 6
- [25] Keita Nakai, Yoshikatsu Hoshi, and Akira Taguchi. Color image contrast enhancement method based on differential intensity/saturation gray-levels histograms. In *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, pages 445–449. IEEE, 2013. 1
- [26] Xutong Ren, Wenhan Yang, Wen-Huang Cheng, and Jiaying Liu. Lr3m: Robust low-light enhancement via low-rank regularized retinex model. *IEEE Transactions on Image Processing*, 29:5862–5876, 2020. 1, 2
- [27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1
- [28] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013. 1, 6, 7

- [29] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2604–2612, 2022. [1](#), [5](#), [6](#), [7](#)
- [30] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. [4](#), [5](#)
- [31] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*. British Machine Vision Association, 2018. [1](#), [2](#), [5](#), [7](#), [8](#)
- [32] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5901–5910, 2022. [1](#), [2](#), [5](#), [6](#), [7](#), [8](#)
- [33] Jun Xu, Yingkun Hou, Dongwei Ren, Li Liu, Fan Zhu, Mengyang Yu, Haoqian Wang, and Ling Shao. Star: A structure and texture aware retinex model. *IEEE Transactions on Image Processing*, 29:5022–5037, 2020. [1](#), [2](#)
- [34] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17714–17724, 2022. [1](#), [5](#), [6](#), [7](#)
- [35] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3063–3072, 2020. [1](#)
- [36] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. [5](#)
- [37] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021. [1](#), [2](#), [5](#), [6](#), [7](#), [8](#)
- [38] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. [1](#), [2](#), [5](#), [8](#)
- [39] Anqi Zhu, Lin Zhang, Ying Shen, Yong Ma, Shengjie Zhao, and Yicong Zhou. Zero-shot restoration of underexposed images via robust retinex decomposition. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020. [1](#), [2](#), [5](#), [7](#), [8](#)