SRENet: Saliency-Based Lighting Enhancement Network

Yuming Fang, *Senior Member, IEEE*, Chen Peng, Chenlei Lv, *Member, IEEE*, and Weisi Lin, *Fellow, IEEE*

***Abstract*—Lighting enhancement is a classical topic in low-level image processing. Existing studies mainly focus on global illumination optimization while overlooking local semantic objects, and this limits the performance of exposure compensation. In this paper, we introduce SRENet, a novel lighting enhancement network guided by saliency information. It adopts a two-step strategy of foreground-background separation optimization to achieve a balance between global and local illumination. In the first step, we extract salient regions and implement the local illumination enhancement that ensures the exposure quality of salient objects. Next, we utilize a fusion module to process global lighting optimization based on local enhanced results. With the two-step strategy, the proposed SRENet yield better lighting enhancement for local illumination while preserving the globally optimal results. Experimental results demonstrate that our method obtains more effective enhancement results for various tasks of exposure correction and lighting quality improvement. The source code and pre-trained models are available at https://github.com/PlanktonQAQ/SRENet**

***Index Terms*—Lighting enhancement, saliency extraction, low light image enhancement.**

I. Introduction

CAPTURING images with high-quality illumination is a very challenging task due to diverse environmental and technical constraints, including abnormal contrast, incorrect settings of photosensitive, darkness of shadow, direct highlight, *etc*. Such negative effects produce abnormal exposure regions which reduce the image quality. Some vision-based applications such as object tracking and face recognition are affected by the quality of exposure. Therefore, the low-light image enhancement (LLIE) is required to improve the quality of illumination. It can correct exposure while restoring the semantic details in images. The LLIE task can be regarded as a kind of low-level image processing.

Although there have been many solutions for LLIE, some limitations still exist. Traditional methods attempt to optimize the quality of illumination based on high dynamic range imaging (HDR) optimization [1] and histogram equalization (HE) [2]. However, such methods produce undesirable artifacts and image distortion with high probability. One improved solution relies on the retinex theory [3], which decomposes the low-light image into reflection and illumination. Such solution concentrates on lighting estimation that is useful for exposure correction. Unfortunately, it frequently introduces some unstable color distributions in real images. The reason is that the retinex-based solution assumes the input image is noise-free, which cannot be guaranteed in low-quality images.

Recently, deep learning-based methods [4], [5] have obtained significant progress in LLIE tasks. Such methods employ deep neural networks to learn the semantic-based mapping between low-light and normal regions in images, which can restore semantic details with high quality illumination. Benefited from the ability, the mentioned solutions can establish reliable lighting models and implement the related optimization tasks. Naturally, these solutions are sensitive to the relationship between semantic regions. Once there is a deviation in semantic analysis, their performance in light optimization is significantly reduced. On the contrary, if the method attempts to establish global lighting optimization, it produces negative impacts on the optimization for specific semantic objects. In addition, some methods [6] attempt to incorporate semantic priors into the low-light enhancement process. However, while SKF introduces semantic information in a more structured way, it still focuses on general semantic guidance rather than performing explicit, targeted optimization for specific object instances. In summary, it is a challenging issue for existing solutions to balance global and local illumination.

In this paper, we propose a novel LLIE method SRENet based on a separation optimization process with two steps. Firstly, we detect salient regions from images to implement saliency-based lighting enhancement. It ensures that the local illumination of the saliency objects can be accurately optimized. Compared to some semantics-based LLIE solutions, using salient regions as the foreground reduces the sensitivity to certain semantic objects, and this improves the generalization in practice. Next, we utilize a fusion module

Received 17 September 2024; revised 27 April 2025; accepted 15 June 2025. Date of publication 15 July 2025; date of current version 21 July 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62132006 and Grant 62311530101; and in part by the Natural Science Foundation of Jiangxi Province of China under Grant 20223AEI91002, Grant 20232BAB202001, and Grant 20224BAB212012. The associate editor coordinating the review of this article and approving it for publication was Dr. Junhui Hou. *(Corresponding author: Chenlei Lv.)*

Yuming Fang and Chen Peng are with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330032, China (e-mail: fa0001ng@e.ntu.edu.sg; ispengchen@outlook.com).

Chenlei Lv is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: chenleilv@mail.bnu.edu.cn).

Weisi Lin is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

Digital Object Identifier 10.1109/TIP.2025.3587588

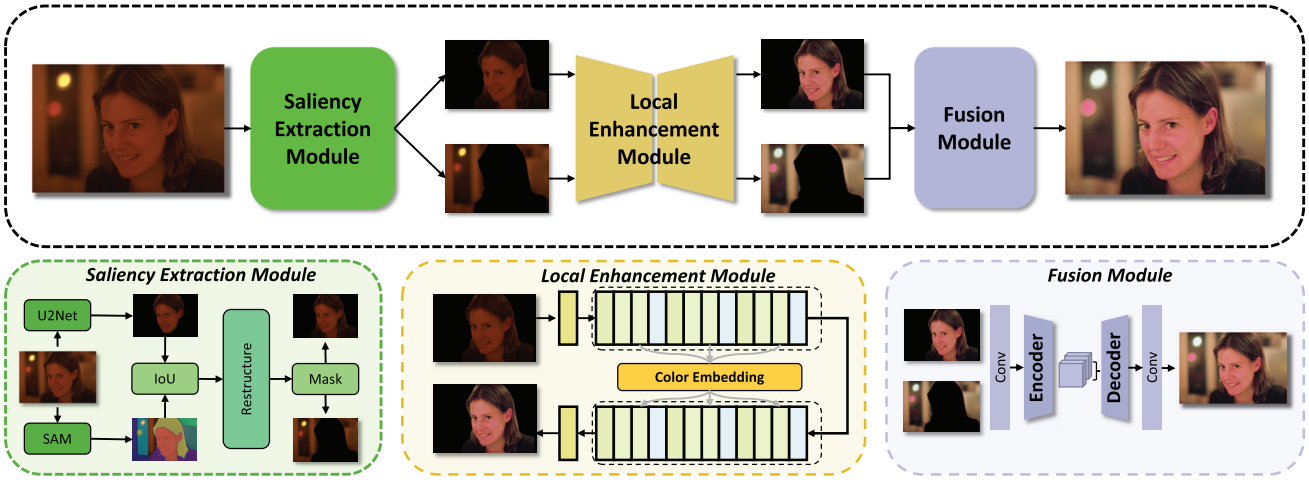


Fig. 1. The pipeline of SRENet. It contains saliency extraction module, local enhancement module, and fusion module.

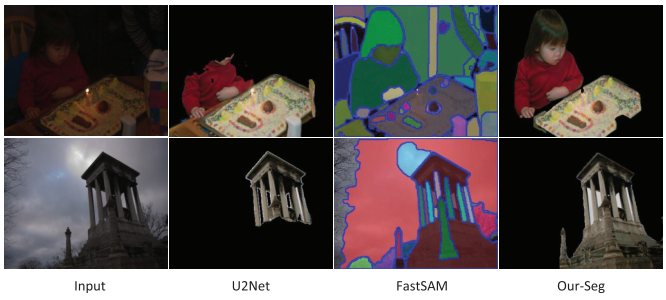


Fig. 2. Visualization of detected salient regions by the proposed module.

to process global lighting optimization based on optimized semantic regions. It keeps the semantic consistency between foreground and background, and balance between global and local illumination. The contributions of the proposed method can be summarized as follows.

- We provide a saliency extraction module to detect foreground semantic objects in images. It first extracts fragments with preliminary semantic relevance and then combines them to obtain salient regions with well-defined boundaries, significantly enhancing the quality of the salient regions.
- We develop a local enhancement module with an encoder-decoder network to optimize the illumination for salient regions and background separately. The foreground objects in salient regions can be significantly highlighted. The network shares weight parameters for the two parts and considers the color changing during the optimization, which provides more flexibility and stability.
- We design a fusion module to integrate optimized detection of salient regions and background. It keeps semantic consistency of exposures and color distributions while preserving fine-grained image details. The fusion module achieves balance between global and local illumination and successfully controls the distortions.

The pipeline of our method is shown in Fig. 2. The rest of the paper is organized as follows. In Sec. II, we discuss the related works about LLIE. In Sec. III, we introduce the implementation details of the saliency extraction module, local enhancement module, and fusion module. We show the performance of proposed method in Sec. IV and Sec. V provides the conclusion.

II. Related Work

According to related implementation details, LLIE methods can be divided into three categories: statistical-based, retinex-based, and data-driven LLIE methods.

Statistical-based LLIE methods utilize statistical analysis to guide lighting optimization. Representative solutions include HDR optimization and histogram equalization (HE). An image with HDR regions may experience tone reproduction issues with a high probability [1]. Some semantic details are lost due to the influence of highlights or shadows. HDR optimization solutions [7], [8], [9] attempt to solve the problem by low dynamic range (LDR) mapping, which ensures more details to be displayed. However, such solutions inevitably lead to a contrast loss. The HE-based methods [2], [10], [11] improve image contrast by modifying the dynamic range of the image. However, such methods suffer from the under-enhancement or over-enhancement without global illumination control. The mentioned works raise a crucial issue: how to establish balance between local and global illumination.

Retinex-based LLIE methods implement lighting optimization according to the retinex theory [3] that reveals the principles of the human visual perception for image illumination. To decompose the image into reflectance and illumination, the retinex-based solutions reconstruct the balance between local and global illumination. According to the retinex theory, some works [12], [13] implement dynamic range optimization to achieve good illumination for images. Following the development of deep learning, the improved solutions [14], [15], [16], [17], [18], [19] combine retinex-based analysis and deep neural works to obtain more adaptable light optimization models. Compared to the statistical-based methods, retinex-based methods improve the ability to balance global and local illumination. However, the performance of

such methods are affected by absence of explicit semantic specification, which leads to the loss of image details in local regions.

Data-driven LLIE methods directly construct lighting optimization models from prior images. Such methods can be divided into two technical routes: supervised and unsupervised schemes. The supervised LLIE methods [5], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30] learn the distributions of illumination based on clearly annotated images. The training images share identical semantic information with varying lighting conditions, which provide clear optimization target. Obviously, such approaches are limited by the training dataset. The unsupervised LLIE methods [4], [6], [31], [32], [33] enhance low-lighting images without explicit paired samples. Some objective rules are employed to establish loss functions, including color curve [4], visual quality [32], semantic-aware knowledge [6], etc. The advantage of such methods is the ability to train models on larger-scale datasets. However, the performance of such methods is limited by the accuracy of trained model and semantic sensitivity in practice. Additionally, other methods such as diffusion-based approaches [34], [35] have been proposed, which leverage diffusion models to enhance low-light images through a generative process.

Overall, it is important to establish a balance between global and local illumination while considering certain semantic details in LLIE task. Such requirements take challenge in terms of performance, robustness, and generalization. Using salient region to guide lighting is a promising solution [36]. The reason is that the salient region has relationship with semantic information but not correspond to specific semantic objects, which reduces the semantic sensitivity. The main drawback of the solution is that if salient region optimization is performed independently, the lighting adjustment for other regions are limited which breaks the balance between global and local illumination. We provide a practical solution to address the issue.

III. Proposed Method

The SRENet contains three main components, including saliency extraction module, local enhancement module, and fusion module. The saliency extraction module detects the accurate salient regions and separates an image into foreground and background. The local enhancement module optimizes the illumination of the two parts by an encoder-decoder network. Finally, the optimized parts are merged to obtain the final results by fusion module. The pipeline is shown in Fig. 1. In following parts, the implementation details are introduced.

A. Saliency Extraction Module

To implement accurate LLIE tasks, we extract salient regions that represent the optimization target for local illumination. The premise is that human subjective attention primarily focuses on salient regions in images. Enhancing these regions independently as the foreground improves human subjective perception. Additionally, salient regions do not require precise semantic annotation, which reduces semantic sensitivity in practice. We design the saliency extraction module to separate the image into two parts: salient regions as the foreground and other regions as the background. First, we implement the saliency detection by

$$S = U2Net(I_{in}), \tag{1}$$

where I_{in} is the input image, and S represents the mask of the salient region captured by U2Net [37]. However, the detected salient region cannot fully satisfy the requirements of the LLIE task due to the limited generalization capability, which often produces fragmented semantic objects with a high probability in low-light scenes. To address this problem, we employ the fast segment anything model (FSAM [38]) to assist in saliency detection. Trained on a large-scale dataset such as SA-1B, SAM is capable of performing segmentation tasks on arbitrary images without the need for task-specific fine-tuning. It supports zero-shot segmentation, maintaining strong performance even on previously unseen objects or scenes. Building on SAM, the FastSAM model is trained using only 2% of the SA-1B dataset released with SAM. While achieving comparable segmentation performance, FastSAM offers a 50-fold improvement in inference speed. FSAM can detect semantic parts of images with more accurate boundaries, represented as

$$\{O_i\}_{i=1}^{m} = FSAM(I_{in}), \tag{2}$$

where O_i represents a semantic part with index i, and the input image is divided into m semantic parts by FSAM. To refine the saliency detection, we compute the IoU scores (pixel-based area overlap ratio) between $\{O_i\}$ and S to improve the boundary accuracy of the identified salient regions. This process can be formulated as

$$S' = S \cup \{O_i\}_{s_i>0.2} - \{O_j\}_{s_j<0.05}, \tag{3}$$

where s_i represents the IoU score between O_i and S, and s_j represents the IoU score between O_j and S. According to this computation, the original salient region S is updated to the new region S' with semantically consistent boundaries. Some blurry pixels distributed along the boundaries can be effectively restored by subtracting the weakly overlapping regions $\{O_j\}$. Although we utilize semantic segmentation to improve the accuracy of boundaries, the proposed module does not rely on specific semantic annotations. By incorporating salient regions, semantic sensitivity can be controlled within an acceptable range. In Fig. 2, we show some saliency results produced by the proposed module.

B. Local Enhancement Module

Based on achieved salient regions and related background, we propose a local enhancement module featuring a two-branch architecture to optimize foreground and background separately. This design enhances the contrast of salient regions, aligning with human subjective perception. Inspired by the backbone structure described in [24], we introduce an encoder-decoder architecture with a color embedding module. To prevent semantic fragmentation from independent enhancement, the two branches share the same weights during training.

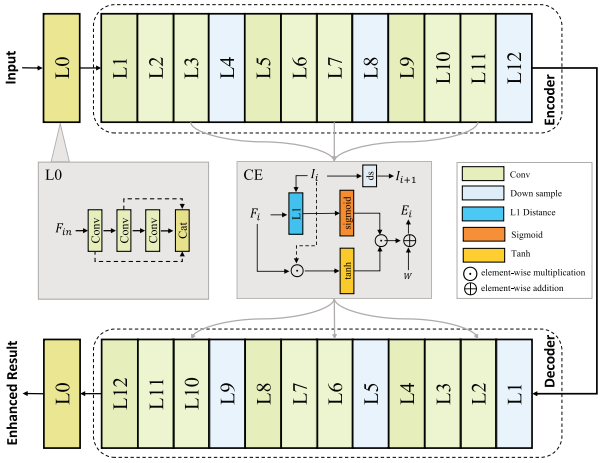


Fig. 3. Architecture of local illumination enhancement module.

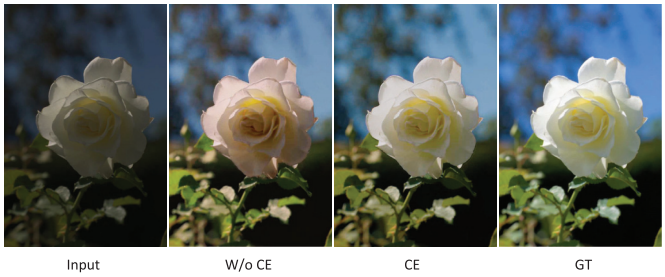


Fig. 4. Visualization of color embedding. Without color embedding, the white semantic objects take significant color shift. W/o CE: without color embedding; CE: with color embedding; GT: ground truth.

To enrich input features, we employ three convolutional blocks to preprocess the original image before encoding. Each block contains a Conv-Norm unit with a GELU (Gaussian Error Linear Unit) activation function. The encoding module consists of 12 convolutional layers aiming to reduce the spatial dimension of the feature map. At the same time, it increases the number of feature channels and facilities the aggregation of contextual features across multiple scales without sacrificing resolution. The first layer of this module comprises a convolutional block, while the second and third layers consist of two convolutional blocks with residual connections. The fourth layer performs down-sampling using convolutional blocks. Layers five to eight and nine to twelve share similar structures to the preceding four layers. The structure of the decoding module corresponds to the structure of the encoding module. During the encoding and decoding processes, down-sampling and up-sampling are achieved using convolutional blocks with a stride of 2. Following the decoding module, three convolutional blocks are included to connect features. Finally, convolutional blocks with a tanh activation function are employed to normalize the number of channels and control data overflow. In Fig. 3, we illustrate the architecture of local illumination enhancement.

The color consistency should be taken into consideration in lighting enhancement [24]. Due to the separate optimization used in our framework, the need for color consistency control becomes especially necessary. We employ a color embedding

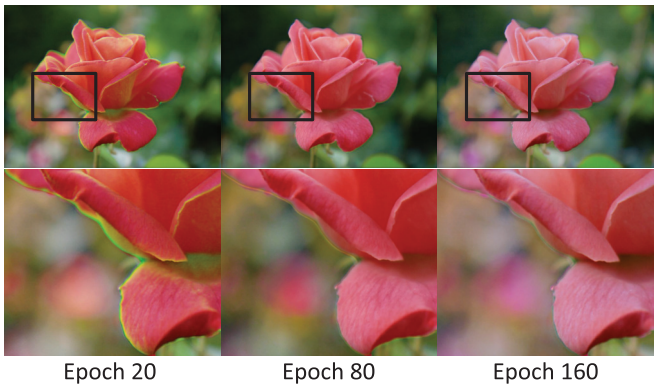


Fig. 5. An instance of fusion for boundary integration.

module to fit the requirement. It combines the original input with encoded features to compute the affinity matrix used for color distribution matching. Throughout the training process, it continuously learns the distribution of colors, ultimately achieving optimized color consistency. To capture the relationship between the input image and its encoded features, we first define two intermediate variables:

$$M = -\|F - I_{in}\|_1, P = F \cdot I_{in}, \tag{4}$$

where F denotes the feature map extracted from the encoder, I_{in} is the input image adjusted via convolution blocks to match the channel dimensions of F, M represents the Manhattan distance (L1 norm) between F and I_{in}, P represents the element-wise product (inner product) of F and I_{in}, indicating feature alignment or consistency. These two variables are then combined to compute the affinity matrix A, which characterizes both semantic and color correlations across different regions of the image:

$$A = 2 \times sigmoid(M) \odot tanh(P), \tag{5}$$

where $sigmoid(M)$ maps the similarity scores into the range $[0.0, 0.5]$, and multiplying by 2 constrains the final values within $[0.0, 1.0]$, preventing data overflow, $tanh(P)$ maps the product values to the range $[-1, 1]$, preserving both positive and negative correlation information, $\odot$ denotes element-wise multiplication, effectively merging the similarity and consistency measurements. The resulting affinity matrix A is then passed to the decoding module to guide color embedding. As shown in Fig. 4 in paper, the color embedding module enhances the color consistency between the output image and the ground truth, producing more visually coherent results.

C. Fusion Module

To obtain the final enhanced image with the balance between local and global illumination, the two parts of the image should be merged with global consistency which takes a significant challenge. Introducing strong global consistency constraint produces degradation of local contrast. On the contrary, weak global consistency leads to significant semantic fragmentation with non-natural boundaries. To address the issue, we propose a fusion module to concentrate salient

Fig. 6. Comparison of enhanced results by different methods. According to the labeled foreground and background regions, our method achieves better balance between exposure correction and contrast preservation.

Fig. 7. Comparison of enhanced results by different methods proposed in recent years. Our method avoids color distortion in enhanced images and keeps exposure balance in complex environments.

regions and background with global consistency. It is designed based on the U-Net structure, which can be represented as

$$I_C = FNet(S'_e + B_e), \tag{6}$$

where S'_e and B_e denote the enhanced salient regions and background, the output I_c is computed by the concentration between S'_e and B_e with global consistency control. In Fig. 5, we show an instance of fusion process for boundary integration. With the iterative optimization, boundaries between foreground and background are eliminated to fit global consistency.

Loss function. The SRENet should consider the image detail preservation and boundary concentration during the fusion process. The primary purpose of the loss function is to establish global consistency constraint on the local enhanced results. It can be represented as

$$L_g = \lambda_1 L_P(I_E) + \lambda_2 L_1(I_E) + \lambda_3 L_B(I_C), \tag{7}$$

where I_E is the middle result that is enhanced by local enhancement module, the perceptual loss L_P is employed to evaluate the similarity between the generated image and the reference one. It emphasizes the semantic level of similarity rather than pixels. The $L1$ loss evaluates the absolute

difference according to predicted and target values of images. It is sensitive to singular values generated by noisy pixels and discrete distributions. The boundary loss L_B controls boundaries between salient regions and background in fusion process. During the training of SRENet, L_B estimates the difference in boundaries and guide the fusion module by gradient propagation. Firstly, it smooths the input image by Gaussian convolution kernel. Then, the boundaries are extracted by Laplace operator. Finally, the Charbonnier loss is employed to evaluate the loss for similarity measurement in boundaries. With the hyperparameters $(\lambda_1, \lambda_2, \lambda_3)$, the loss terms are combined to abtain the loss funtion L_g to balance local illumination enhancement and global consistency.

Training. The SRENet uses Adam optimizer to implement training. The related parameters are set to $\beta_1 = 0.99$ and $\beta_2 = 0.999$. The learning rate is set to 10^{-5}. To preserve the integrity of the images, we do not perform cropping for feature analysis. All training images are normalized to 400×320. The batch size is set to 2 for each iteration, and the epoch number is set to 320. We evaluate the performance with different tasks in the following part.

IV. Experiments

A. Datasets and Metrics

We evaluate the performance of SRENet in MIT-Adobe FiveK [39] and LOL dataset [17], which contain rich saliency information and various low-light scenes. Based on the MIT-Adobe FiveK, we select images edited by the expert C to be the reference data. Related original images are used to be the source samples. For exposure learning, we collect images with normal exposure from LOL dataset to be the reference ones. Some images with abnormal exposure are selected to be the source ones. In addition, we generate a synthetic image dataset to improve the ability for accurate light source perception. Based on the software "set-a-light-3D", we set precise photography parameters to render images with different light source intensities. Overall, the entire test dataset includes 3301 pairs of images, including source and reference ones that are regarded as the input samples and related ground truth. It includes training set (2547 pairs), validation set (643 pairs), and test set (111 pairs).

To provide quantitative analysis for different methods, we employ a set of image quality assessment (IQA) metrics, including full reference IQA and no-reference IQA. The full reference IQA methods require the reference image to be the baseline for quality evaluation. We employ the classical measurements PSNR and SSIM [41] to evaluate the performance of LLIE methods. For no-reference IQA, we select three mainstream metrics ILNIQE [42], DBCNN [43], and MUSIQ [44] to conduct the quantitative analysis. It should be noticed that each metric can only evaluate image quality from a specific perspective. In addition, we also establish a user study (45 participants, with 25 randomly selected images for each one to choose the best and worst images) as the subjective evaluation reference.

TABLE I

Quantitative Results of Different Methods on the Test Set. The Top Two Results Are Marked in Bold. U-Best and U-Worst are Percentage Results of the User Study

Methods	PSNR↑	SSIM↑	ILNIQE↓	DBCNN↑	MUSIQ↑	U-Best↑	U-Worst↓
RetinexNet [17]	11.51	0.577	26.502	48.586	58.930	——	——
KinD [18]	15.46	0.673	**22.414**	47.473	58.32	——	——
Zero-DCE [4]	15.17	0.661	22.431	44.830	54.706	——	——
MIRNet [20]	21.29	0.810	24.270	**54.387**	64.942	——	——
HWMNet [40]	21.25	0.814	23.400	49.680	62.638	——	——
IAT [25]	18.97	0.778	24.584	47.596	62.183	——	——
LLFormer [5]	**23.32**	**0.841**	22.790	49.603	63.729	**0.28**	**0.05**
Retinexformer [19]	21.56	0.793	25.921	49.563	63.551	0.08	0.23
SCI [32]	20.91	0.812	24.201	52.035	64.161	0.02	0.19
LLFIOW-SKF [6]	20.56	0.785	25.786	48.831	63.360	0.03	0.36
DCCNet [24]	17.58	0.742	22.539	51.870	**65.765**	0.24	0.11
SRENet	**21.95**	**0.821**	**21.987**	**52.091**	**65.884**	**0.33**	**0.02**

B. Comparisons

To provide a comprehensive evaluation of LLIE performance, we introduce a series of classical methods as the reference, including RetinexNet [17], KinD [18], Zero-DCE [4], MIRNet [20], HWMNet [40], IAT [25], LLFormer [5], Retinexformer [19], SCI [32], LLFIOW-SKF [6], and DCCNet [24]. The selected comparative approaches encompass most of the mainstream LLIE solutions in the past three years. Fig. 6 presents an instance enhanced by different methods. In general, traditional methods focus on global illumination optimization which results in a significant loss of contrast in the processed images. In contrast, the methods proposed in the past two years can achieve more natural results with the balance between contrast preserving and exposure optimization.

Building on the progress, our method achieves enhanced illumination for regions with significant exposure differences, benefiting from a separate optimization strategy. In Fig. 7, we compare the results of our method with several recent solutions. It exhibits a clear advantage of our method in achieving a balanced exposure between the foreground and background with complex lighting conditions. In Fig. 8, some images with more complex exposures and related RGB histograms are displayed. Our method achieves a better balance between exposure control and contrast preservation. Benefited from the separation optimization, our method is able to achieve better exposure results with adaptable color distributions. More results are shown in Fig. 9. The results labeled with "Our-Global" means that the entire images are regarded as the salient regions and processed by SRENet. Such conditions correspond to the case where the saliency detection fails.

To comprehensively demonstrate our method to restore overexposed images, we also selected the two SOTA methods for a detailed comparison. As shown in Fig. 11, DCCNet tends to increase image brightness while neglecting color information, and LLFormer similarly results in a warmer tone. In contrast, our method better preserves the color information and overall contrast of the images. To provide compelling quantitative analysis, we reports results of IQA metrics and related user study report in Table I and Fig. 10. The IQA metrics show that our method achieves more stable quality based on various measurements. For the user study, we designed a simple and intuitive web-based interface that displays the enhancement results from different LLIE methods side by side. Participants were asked to select the image they considered the "best" and the one they considered the "worst" among

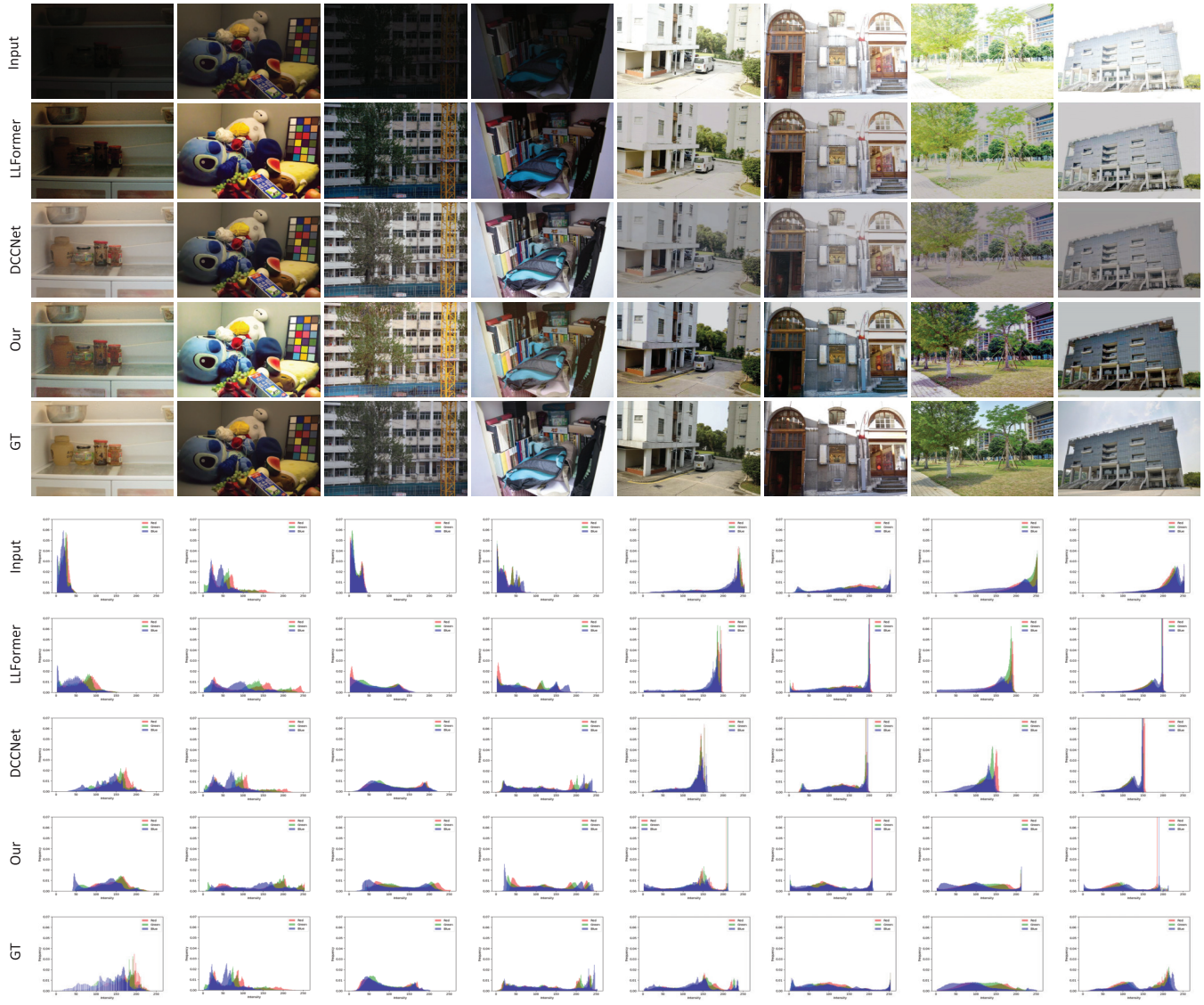


Fig. 8. Comparison of enhanced results by different methods proposed in recent years. Our method can handle images with underexposure and overexposure. According to the histograms, our method achieves uniform distributions of RGB channels while maintaining the contrast with high quality.

all displayed results for each scene. To ensure diversity and reduce bias, we randomly shuffled the display order of the methods across scenes and collected responses from more than 100 participants with varied backgrounds. Through statistical analysis of the collected selections, we aim to determine which LLIE method is most preferred by human observers. The results indicate that our approach was selected as "best" with the highest proportion and as "worst" with the lowest proportion. Related results are shown in Fig. 10.

Due to the benefits of the color embedding module, our method can better maintain the color consistency of the images. As shown in Fig. 12, we selected the two methods with the best perception based on the user study as the References. We calculate the color differences between the ground truth and enhanced results, and generate related heatmaps. In Fig. 12, such heatmaps show that our method is more consistent with the natural color distribution according to the ground truth. Furthermore, our method also has better color consistency for the salient objects.

C. Applications

Low-light Object Detection. The LLIE not only meets human subjective visual perception and provides better visualization effects, but also plays an important role in some practical vision applications. In this part, we report some quantitative results for object detection tasks in low-light environments. The test dataset is collected from ExDark dataset [45], which includes 5890 training images, 737 validation images, and 736 test images. The YOLO-V5 [46] is used as the test classifier. We use different LLIE methods to preprocess low-light images and compare the object

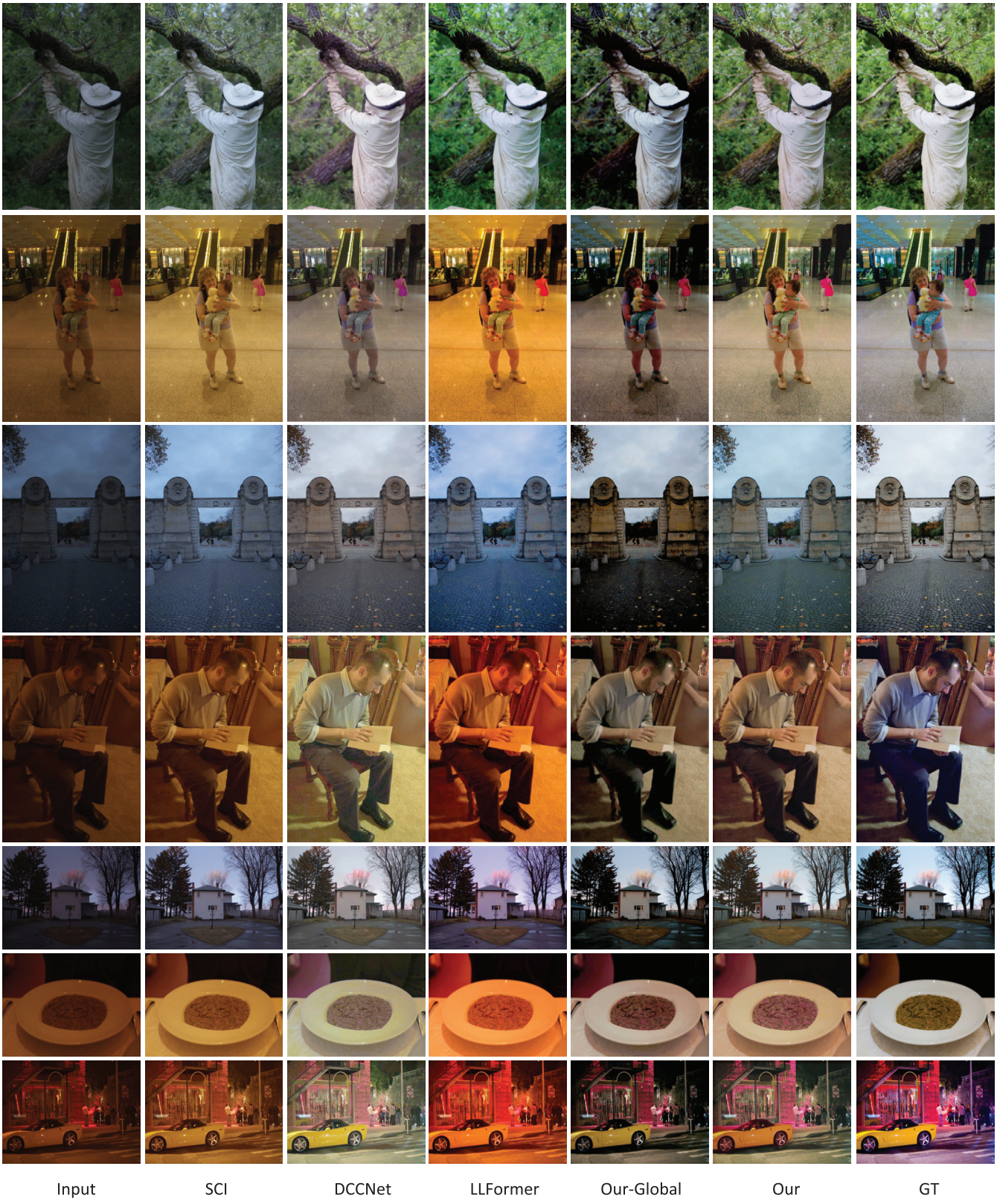


Fig. 9. Comparison of enhanced results by different methods on MIT-Adobe FiveK images.

detection performance. The quantitative results are shown in Table II. Compared to other LLIE methods, SRENet achieve performance improvement across more categories. To visually demonstrate the effect of LLIE methods on object detection

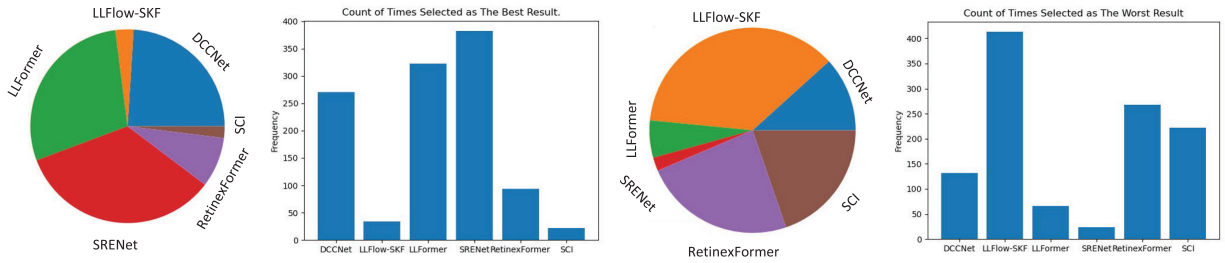


Fig. 10. Quantitative analysis of the user study. Our method achieves the most selections of "best" and the fewest "worst" selections.

TABLE II

COMPARISON OF THE PRE-PROCESSING EFFECTS BY DIFFERENT LLIE METHODS ON HIGH-LEVEL VISION UNDERSTANDING. THE TOP TWO RESULTS ARE MARKED IN BOLD

Methods	All	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Motor	People	Table
RetinexNet [17]	0.586	0.650	0.549	0.705	0.551	0.494	0.656	0.5	0.706	0.649	0.577	0.627	0.371
KinD [18]	0.675	0.803	**0.727**	0728	0.654	0.645	0.717	0.556	0.705	0.635	0.627	**0.704**	0.413
Zero-DCE [4]	0.65	0.757	0.679	0.759	0.652	0.643	0.751	0.518	0.704	0.599	0.546	0.678	0.39
MIRNet [20]	0.606	0.701	0.561	0.525	0.598	0.687	0.563	**0.588**	0.645	0.576	0.686	0.645	**0.495**
HWMNet [40]	0.635	0.734	0.577	0.545	**0.69**	**0.694**	0.613	0.57	0.667	0.567	**0.789**	0.673	**0.498**
IAT [25]	0.652	0.794	0.674	0.753	0.631	0.635	0.713	0.532	0.698	0.653	0.655	**0.701**	0.382
LLFormer [5]	0.605	0.679	0.531	0.513	0.609	0.691	0.621	0.55	0.63	0.544	**0.769**	0.642	0.483
Retinexformer [19]	0.667	**0.814**	0.697	0.693	0.683	0.69	0.733	0.543	**0.734**	0.64	0.699	0.691	0.377
SCI [32]	**0.678**	0.805	0.716	**0.799**	**0.691**	0.677	**0.753**	0.571	0.709	**0.696**	0.718	0.696	0.382
LLFIOW-SKF [6]	0.652	0.766	0.695	0.748	0.666	**0.697**	0.662	0.494	0.681	**0.689**	0.717	0.649	0.325
DCCNet [24]	0.626	0.75	0.703	0.715	0.635	0.639	0.672	0.557	0.702	0.523	0.579	0.664	0.374
SRENet	**0.679**	**0.81**	**0.737**	**0.769**	0.663	0.683	**0.888**	**0.573**	**0.718**	0.602	0.582	0.667	0.455

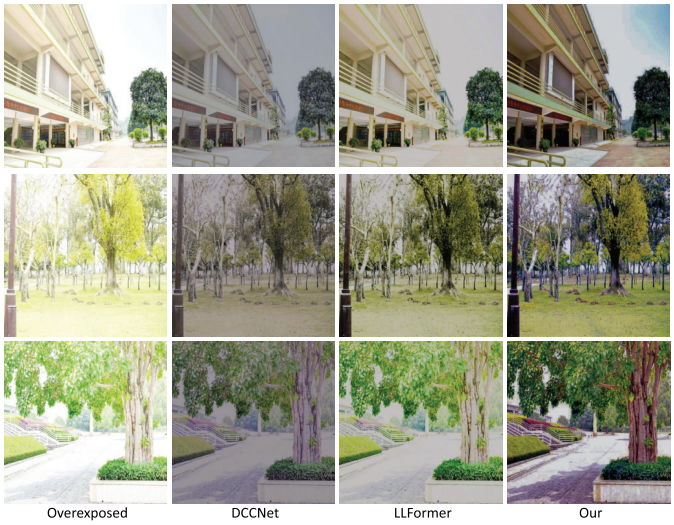


Fig. 11. Comparison of enhanced results by different methods on overexposed images. Our method achieves more natural illumination.

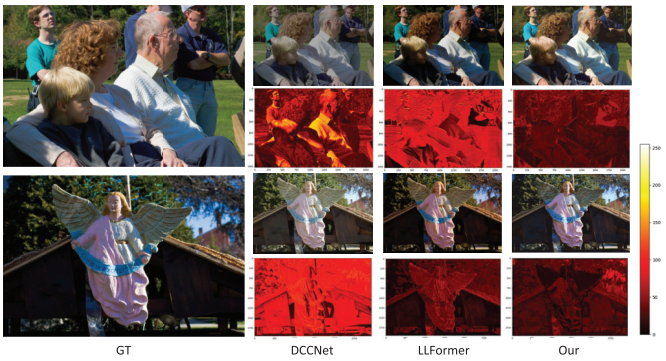


Fig. 12. Comparison of enhanced results by different methods with heatmaps. Our method achieves images closer to the ground truth.

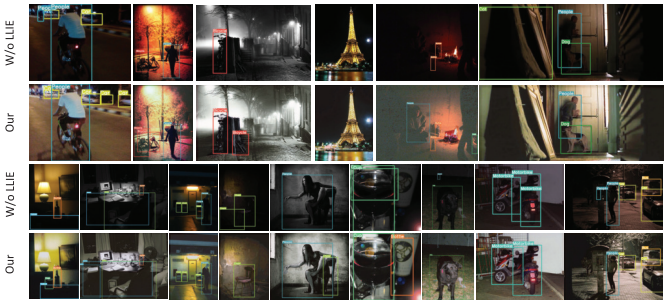


Fig. 13. Comparison of object detection without and with our method in low-light images.

tasks, we provide some examples in Fig. 13. Our method improves the accuracy for semantic detection. Some semantic objects influenced by low-exposure conditions can be revealed.

Color Transfer. The quality of image color distribution has close relationship with the lighting condition. Naturally, LLIE method can improve the performance in color-based enhancement applications such as color transfer and colorization. For the color transfer, the low-quality exposed areas affect color correspondences and degrade the transfer performance. With the improvement of LLIE, such limitation can be controlled. In Fig. 14, we compare some color transfer results for LOL images without and with SRENet improvement. Regardless

Fig. 14. Color transfer results without and with LLIE methods. The basic color transfer method is selected to WCT2 [47].

Fig. 15. The LLIE results by SRENet with different salient regions.

Fig. 16. The LLIE results by SRENet with global enhancement and local enhancement.

of whether the low-light areas appear in source images or reference ones, our method significantly improve the quality of color distribution. The contrast and tone in images are enhanced.

D. Ablation Study

To illustrate the functionality of different modules in SRENet, we conduct the ablation study to reveal their individual contributions. The saliency extraction module improves the accuracy of salient regions in the image with semantic consistency. The traditional saliency detection method break the semantic consistency with high probability. In Fig. 15, we compare the enhanced results by SRENet with traditional saliency detection and saliency extraction module. The instances show that the saliency extraction module extracts more accurate foreground information and output lighting optimization with better semantic consistency. In contrast, the traditional saliency detection methods produce unnatural breaks of color distribution.

Fig. 17. The LLIE results by SRENet without and with fusion processing.

Fig. 18. Negative examples by SRENet.

The local enhancement module is established according to the separation optimization strategy. A fundamental doubt is whether the strategy can achieve better results than solutions with global optimization. To validate the effectiveness of our strategy, we compare the SRENet with global enhancement (the entire image is regarded as the foreground) and local enhancement for LLIE tasks. In Fig. 16, some enhanced results are shown. Although the same structure is used by the global enhancement and local one, the latter is still able to achieve better results that are benefited from the enhancement module. More results have been shown in Fig. 9. Due to the selection of separate optimization, the importance of fusion module is self-evident. The key issue is that the boundary can be well processed by fusion processing. In Fig. 17, we compare enhanced images by SRENet without and with fusion processing. It can be seen that the fusion module effectively avoids artifact edges even for regions with complex boundary information.

Limitations. It should be acknowledged that our method relies on accuracy of salient regions. The performance of our method will degrade to global optimization (Fig. 16) when the image does not have obvious saliency objects as foreground, which can be regarded as a limitation. There is a special phenomenon where ghosting occurs at the edges of salient regions when the foreground appears in a backlit environment with extreme contrast difference. The reason is that in the backlit environment, boundaries of salient regions have abnormal high dynamic ranges in the illumination distribution, which makes it difficult for the fusion network to

achieve seamless blending results. In Figure 18, we show some negative examples. The color consistency is disrupted with color bleeding and ghosting effects when the boundaries of salient regions are significantly affected by back-lighting. Although there have some unsatisfactory areas in enhanced images, the optimization effect of SRENet on the foreground remains highly noticeable.

V. Conclusion

In this paper, we have proposed an LLIE method SRENet that utilizes a separation optimization strategy to balance local and global illuminations in images. It divides an image into foreground and background according to the salient regions. Based on the two parts, the SRENet implements local illumination enhancement to optimize the image separately. Benefited from the fusion module, the two enhanced parts can be concentrated with global consistency. The proposed solution successfully improves the contrast of salient regions while enhancing the quality of lighting and color distribution; for exposure correction, the SRENet overcomes the damage caused by the global exposure scheme to local semantic details. The resultant improvement is meaningful for many computer vision tasks.

References

[1] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, 2002, pp. 661–670.

[2] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3431–3441, Dec. 2011.

[3] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, Dec. 1977.

[4] C. Guo et al., "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1780–1789.

[5] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 3, 2023, pp. 2654–2662.

[6] Y. Wu et al., "Learning semantic-aware knowledge guidance for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 1662–1671.

[7] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do hdr displays support ldr content? A psychophysical evaluation," *ACM TOG*, vol. 26, no. 3, pp. 1–7, 2007.

[8] P. Didyk, R. Mantiuk, M. Hein, and H. P. Seidel, "Enhancement of bright video features for HDR displays," *Comput. Graph. Forum*, vol. 27, no. 4, pp. 1265–1274, Jun. 2008.

[9] Z. Li, Z. Wei, C. Wen, and J. Zheng, "Detail-enhanced multi-scale exposure fusion," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1243–1252, Mar. 2017.

[10] C. Lee, C. Lee, Y.-Y. Lee, and C.-S. Kim, "Power-constrained contrast enhancement for emissive displays based on histogram equalization," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 80–93, Jan. 2012.

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

[12] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[13] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[14] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1013–1037, Apr. 2021.

[15] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10561–10570.

[16] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5901–5910.

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

[18] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia (ACM MM)*, Oct. 2019, pp. 1632–1640.

[19] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," 2023, *arXiv:2303.06705*.

[20] S. W. Zamir et al., "Learning enriched features for real image restoration and enhancement," in *Proc. Eur. Conf. Comput. Vis.*, vol. 12370, 2020, pp. 492–511.

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

[22] S. W. Zamir et al., "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 14821–14831.

[23] J. Li, X. Feng, and Z. Hua, "Low-light image enhancement via progressive-recursive network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 11, pp. 4227–4240, Nov. 2021.

[24] Z. Zhang, H. Zheng, R. Hong, M. Xu, S. Yan, and M. Wang, "Deep color consistent network for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1899–1908.

[25] Z. Cui et al., "You only need 90K parameters to adapt light: A light weight transformer for image enhancement and exposure correction," in *Proc. BMVC*, Jan. 2022, pp. 1–25.

[26] H. Luo, B. Chen, L. Zhu, P. Chen, and S. Wang, "RCNet: Deep recurrent collaborative network for multi-view low-light image enhancement," *IEEE Trans. Multimedia*, vol. 27, pp. 2001–2014, 2025.

[27] L. Zhu, W. Yang, B. Chen, H. Zhu, X. Meng, and S. Wang, "Temporally consistent enhancement of low-light videos via spatial–temporal compatible learning," *Int. J. Comput. Vis.*, vol. 132, no. 10, pp. 4703–4723, Oct. 2024.

[28] L. Zhu, W. Yang, B. Chen, F. Lu, and S. Wang, "Enlightening low-light images with dynamic guidance for context enrichment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 8, pp. 5068–5079, Aug. 2022.

[29] Y. Feng et al., "DiffLight: Integrating content and detail for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2024, pp. 6143–6152.

[30] Z. Zhu, J. Hou, H. Liu, H. Zeng, and J. Hou, "Learning efficient and effective trajectories for differential equation-based image restoration," 2024, *arXiv:2410.04811*.

[31] A. Zhu, L. Zhang, Y. Shen, Y. Ma, S. Zhao, and Y. Zhou, "Zero-shot restoration of underexposed images via robust Retinex decomposition," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2020, pp. 1–6.

[32] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5637–5646.

[33] L. Zhu et al., "Unrolled decomposed unpaired learning for controllable low-light video enhancement," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2024, pp. 329–347.

[34] J. Hou, Z. Zhu, J. Hou, H. Liu, H. Zeng, and H. Yuan, "Global structure-aware diffusion process for low-light image enhancement," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2023, pp. 79734–79747.

[35] H. Jiang, A. Luo, X. Liu, S. Han, and S. Liu, "LightenDiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models," in *Proc. Eur. Conf. Comput. Vis.*, Nov. 2024, pp. 161–179.

[36] Y. Li et al., "Saliency guided naturalness enhancement in color images," *Optik*, vol. 127, no. 3, pp. 1326–1334, Feb. 2016.

[37] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107404.

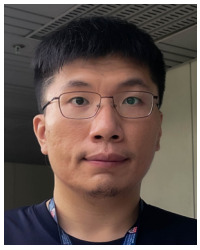
[38] X. Zhao et al., "Fast segment anything," 2023, *arXiv:2306.12156*.

[39] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 97–104.

[40] C.-M. Fan, T.-J. Liu, and K.-H. Liu, "Half wavelet attention on M-Net+ for low-light image enhancement," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 3878–3882.
[41] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.
[42] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
[43] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 36–47, Jan. 2020.
[44] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "MUSIQ: Multi-scale image quality transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 5148–5157.
[45] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark dataset," *Comput. Vis. Image Understand.*, vol. 178, pp. 30–42, Jan. 2019.
[46] R. Couturier, H. N. Noura, O. Salman, and A. Sider, "A deep learning object detection method for an efficient clusters initialization," 2021, *arXiv:2104.13634*.
[47] J. Yoo, Y. Uh, S. Chun, B. Kang, and J.-W. Ha, "Photorealistic style transfer via wavelet transforms," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9036–9045.
[48] W. Ren et al., "Low-light image enhancement via a deep hybrid network," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4364–4375, Sep. 2019.
[49] Y. Jiang et al., "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.
[50] X. Ren, W. Yang, W.-H. Cheng, and J. Liu, "LR3M: Robust low-light enhancement via low-rank regularized retinex model," *IEEE Trans. Image Process.*, vol. 29, pp. 5862–5876, 2020.
[51] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *Proc. ICCV*, 2023, pp. 12504–12513.

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

Chen Peng received the B.E. degree from the School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China, in 2022, where he is currently pursuing the M.E. degree. His research interests include image processing and computer vision.



Chenlei Lv (Member, IEEE) received the Ph.D. degree from the College of Information Science and Technology, Beijing Normal University (BNU). He was a Research Fellow with the School of Computer Science and Engineering, Nanyang Technological University (NTU). He is currently an Assistant Professor with the College of Computer Science and Software Engineering (CSSE), Visual Computing Research Center (VCC), Shenzhen University (SZU). His research interests include computer graphics, 3D vision, differential geometry, and machine learning. He has served as the ICME2023 Area Chair and the CVM2023 Session Chair. The personal page of the link is: https://aliexken.github.io/

Weisi Lin (Fellow, IEEE) received the Ph.D. degree from King's College London, U.K. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University. His research interests include image processing, perceptual signal modeling, video compression, and multimedia communication, in which he has published over 200 journal articles, over 230 conference papers, filed seven patents, and authored two books. He is a fellow of IET and an Honorary Fellow of Singapore Institute of Engineering Technologists. He was the Technical Program Chair of IEEE ICME 2013, PCM 2012, QoMEX 2014, and IEEE VCIP 2017.