2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

ASP-LED: Learning Ambiguity-Aware Structural Priors for Joint Low-Light Enhancement and Deblurring

Jing Ye[1], Yang Liu[1], Congjing Yu[1], Changzhen Qiu[1], Zhiyong Zhang[1*]

***Abstract*— Low-light enhancement and deblurring is vital for high-level vision-related nighttime tasks. Most existing cascade and joint enhancement methods may provide undesirable results, suffering from severe artifacts, deteriorating blur, and unclear details. In this paper, we propose a novel ambiguity-aware network (ASP-LED) with structural priors, including high-frequency and edge, to enable effective image representation learning for joint low-light enhancement and deblurring. Specifically, we employ a Transformer backbone to explore the global clues of the image. To compensate for the inadequate local detail optimization, we propose a multi-patch perception pyramid block that models the correlation between different size patches and ambiguity, and identifies non-uniform deblurring spatial features, facilitating the reconstruction of potential high-frequency and edge information. Furthermore, a prior-guided reconstruction block based on the parallel attention mechanism is present to adaptively correct global image with statistical features, which helps guide the model to refine sharp texture and structure. Extensive experiments performed on simulated and real-world datasets demonstrate the efficacy of our proposed method in restoring low-light blurry images with increased visual perception compared to state-of-the-art methods.**

I. INTRODUCTION

Acquiring images with good visibility is crucial in fields where intelligent robots are required to interact with environments[1], [2]. However, complex nighttime conditions usually lead to severe degradation of imaging quality. Thus, low-light enhancement and deblurring are vital for vision-related nighttime tasks.

Most existing methods attempt to address the co-existence of low light and blur degradation by cascading low-light enhancement [3], [4], and deblurring techniques [5], [6], [7]. One paradigm is to perform low-light enhancement followed by deblurring. However, most low-light enhancement methods may overexpose the saturated areas of the night blurry images, resulting in worsening blur (see Fig. 1(b)). In addition, low-light enhancement risks masking deblurring information clues by smoothing the image during the denoising process, the cascade method fails to achieve the desired deblurring effect as shown in Fig. 1(c). Another strategy is to deblurring and then conduct low-light enhancement. Previous deblurring methods are trained only on daytime scene datasets, which poorly perceive motion information and features in low illumination regions of blurry images. It is non-trivial to restore sharp low-light images directly using existing deblurring methods. While the low-light enhancement performance is highly dependent on the quality of the deblurred image, the pixel errors introduced in the deblurring stage propagate to the low-light enhancement stage, which affects the final enhancement effect. As shown in Fig. 1(d), the cascade work MIMO[5] → Zero-DCE [4] yields suboptimal results. The above cascade methods treat low-light enhancement and deblurring as independent tasks, ignoring the correlation between low light and motion blur degradation. The forceful combined methods fail to handle degradation with co-existing low light and blur.

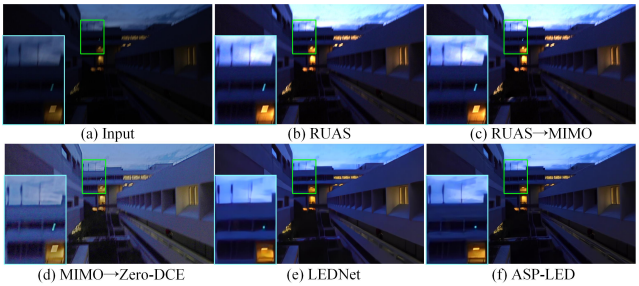

Fig. 1. A visual comparison of the proposed ASP-LED with the representative cascade methods RUAS [3] → MIMO [5], MIMO[5] → Zero-DCE [4], and the latest joint method LEDNet [8].

Considering the cooperation of low-light enhancement and deblurring mechanism, LEDNet [8] firstly customizes the joint task and proposes to solve the illumination enhancement and deblurring problems in a unified framework. However, the approach fails to recover sharp details in some cases due to the lack of guidance from prior information. As shown in Fig. 1(e), although the quality of night blurry images is improved, the texture and structure of the enhanced results are still blurred.

This paper aims to improve the visibility of low-light blurry images and restore sharp texture detail simultaneously. Our method is based on two insights. First, blurring low-light images can be viewed as a low-pass filtering process. The blurry image's information ($e.g.$, texture and edge) is distorted or lost. Image deblurring is to recover this information. As shown in Fig. 2, we observe that the high-frequency and edge information significantly differs between the low-light blurry image and the normal-light sharp image. It implies that the accurate high-frequency and edge features tend to represent the information of the normal-light sharp image. Therefore, we introduce high-frequency and edge priors to guide the model in learning the recovery of distinct texture and structure. Second, knowing the non-uniform ambiguity helps to eliminate them appropriately. However, ambiguity is difficult to predict because it depends on the spatial characteristics of target motion and depth variation. As described in [9], the ambiguity decreases as the distance

[1]School of Electronics and Communication Engineering, Sun Yat-Sen University, Shenzhen, China. (Corresponding author* Zhiyong Zhang: zhangzhy99@mail.sysu.edu.cn)

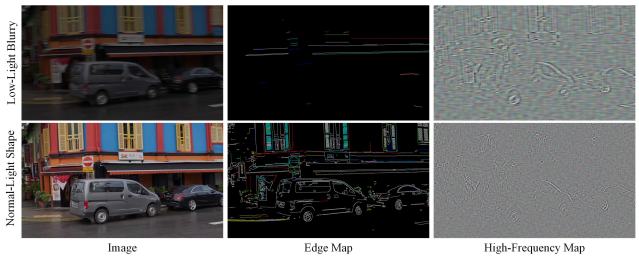

Fig. 2. A visual comparison of edge and high-frequency maps of the low-light blurry image, and normal-light sharp image.

between the target and the camera increases in static scenes, and ambiguity increases as the target moves faster in dynamic scenes. We note that motion and depth information in the spatial domain can provide some valuable clues for ambiguity estimation, which inspires us to aggregate multi-grained spatial features to remove non-uniform blur. Based on these insights, in this paper, we propose a novel ambiguity-aware network (ASP-LED) with structural priors, including high-frequency and edge, to enable effective image representation learning for joint low-light enhancement and deblurring. Our contributions are summarized as follows:

- We propose a novel ambiguity-aware framework with structural priors for joint low-light enhancement and deblurring, where multiple statistical information is introduced to restore mixed degraded images.
- We present a multi-patch perception pyramid block to model the correlation between different size patches and ambiguity, and the captured expressive multi-grained deblurring spatial features are used to remove non-uniform blur of different intensities.
- To efficiently exploit statistical features to recover sharp textures and structures cooperatively, we propose a prior-guided reconstruction block based on the parallel attention mechanism to adaptively correct global image features with high-frequency and edge information.
- Extensive experiments demonstrate that the proposed method outperforms SOTA methods on both simulated and real-world low-light blurry datasets.

II. Related Works

A. Low-Light Image Enhancement

Traditional low-light enhancement methods include histogram equalization [10], [11], gamma correction [12], and Retinex methods [13], [14], [15]. However, these methods tend to produce artifacts and color distortions when dealing with complex real-world images. Recently, learning-based image enhancement methods have achieved impressive results [16], [17], [3], [18], [19], [20], [21]. Lore *et al.* [22] proposed a deep autoencoder (LLNet) to illuminate and denoise low-light images adaptively. Zhang *et al.* [23] proposed a network (KinD) consisting of three subnets to achieve low-light image enhancement collaboratively. Zero-DCE [4] and Zero-DCE++ [24] transform low-light enhancement into a high-order curve estimation task. The above methods are excellent for enhancing low-light images, but they ignore the spatial degradation of motion blur. In contrast, we improve image visibility while accounting for non-uniform blur.

B. Image Deblurring

Traditional deblurring methods [25], [26], [27], [28], [29] based on hand-crafted priors and assumptions for image deblurring. However, their generality and expressiveness are somewhat limited. Lately, learning-based methods have dominated with the advent of paired deblurring datasets. [5], [30], [31] use multi-scale structures to expand the perceptual field to remove stronger blur. To avoid information loss, the multi-patch strategy [32], [33], [34] is proposed to deal with non-uniform blur in a coarse-to-fine manner. Meanwhile, some effective priors have been introduced into deep networks to solve the deblurring problem [35], [36], [37], [38], [39], [40]. Existing deblurring methods focus on removing blur from daytime images rather than nighttime scenes, and they cannot handle degradation where low light and blur co-exist. In this work, we propose a multi-patch perception pyramid block to perceive non-uniform blur in low-light images.

C. Joint Low-light Enhancement and Deblurring

Previous studies have shown that a unified framework is much better than simply combining networks from different domains[41], [42]. Zhou *et al.* [8] proposed a network called LEDNet, which includes an encoder for illuminate enhancement and a decoder for deblurring, using a joint optimized scheme to generate normal-light sharp images. However, the method achieves unsatisfied enhancement for low-light images with strong blur, because it does not take into account the guidance of the prior information in tackling low light and blur degradation. In comparison, we introduce structural priors guidance within a unified task model to achieve better blur removal and enhance image illumination.

III. Proposed Method

A. Overview

The overall framework of our proposed ASP-LED in Fig. 3, the low-light blurry image $I_{in} \in \mathbb{R}^{H\times W\times 3}$ is first through an encoder consisting of convolutional layer and pyramid pooling module (PPM) [43] to extract shallow image features. Then, they are fed into the SNR Transformer [16] global information enhancement branch and structure prior learning branch in the feature reconstruction subnet. Finally, the image and statistical features aggregated by prior-guided reconstruction block (PR) are used to reconstruct the normal-light sharp image $I_{out} \in \mathbb{R}^{H\times W\times 3}$ in the decoder.

B. Multi-Patch Perception Pyramid Block

In a blurry image, the ambiguity of neighboring pixels is similar, the difference in ambiguity is more significant for pixels that are farther away. To accurately predict the non-uniform blur, we propose using multi-grained spatial information to identify the blur by constructing the parallel perceptual pyramid. Although this strategy has been explored in previous deblurring methods [33], [34], [40], these methods usually ignore the negative effects caused by non-uniform blur, such as the variation of ambiguity. Unlike these works, we further consider the correlation between different size patches and ambiguity to model the features

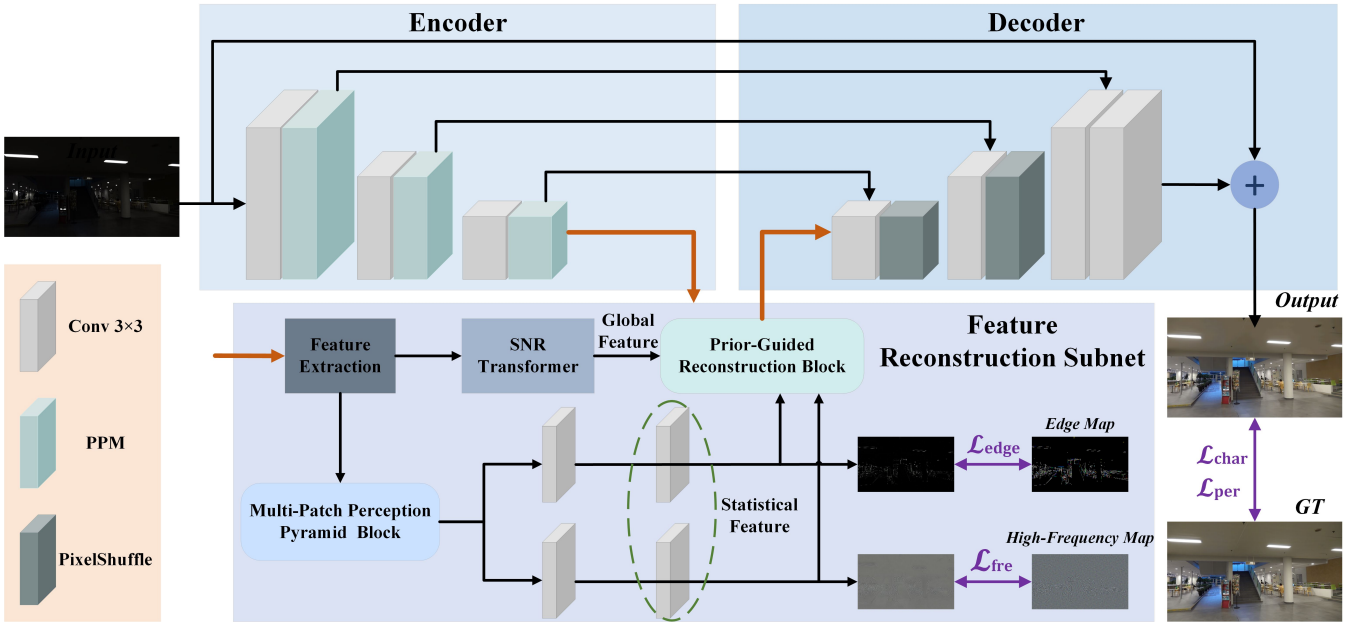

Fig. 3. Overview of the proposed ASP-LED. Given a low-light blurry image, our method first extracts shallow image features with an encoder and then sends them to the feature reconstruction subnet, where a multi-patch perception pyramid block (MPP) uses multi-grained spatial features to remove non-uniform blur and output potential high-frequency and edge maps. Furthermore, a prior-guided reconstruction block (PR) adaptively fuses the global features captured by the Transformer and statistical features. Finally, the decoder refines and restores the normal-light sharp image.

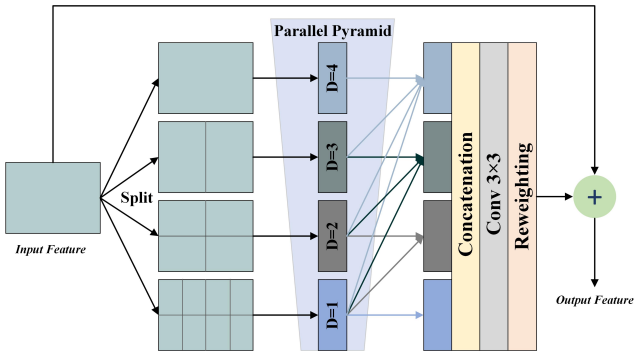

Fig. 4. Illustration of the proposed multi-patch perception pyramid block.

used to remove non-uniform blur. We first construct a multi-patch perception pyramid for depth image features and then capture multi-grained features more beneficial for deblurring within patches of different sizes using appropriate perceptual fields. Finally, the features of multiple layers are adaptively reaggregated using a spatial reweighting mechanism.

The specific structure of the multi-patch perception pyramid block is shown in Fig. 4. We first perform the split operation $S^T(\cdot)$ on the depth image features $\mathcal{F}_{in}$, and then feed them to a set of parallel dilated convolutions to identify the non-uniform blur. In particular, similar deblurring features are captured with convolutions of smaller receptive fields within the fine-grained patches, and discrepant deblurring features are extracted with convolutions of larger receptive fields within the coarse-grained patches. This process can be defined as follows:

$$\mathcal{F}_1^T = Conv_3^D\left(S^T\left(\mathcal{F}_{in}\right)\right), \tag{1}$$

where $Conv_3^D(\cdot)$ denotes an dilated convolution with kernel 3×3 and D is the dilation rate. T represents the parallel pyramid hierarchy. Subsequently, we gradually aggregate multiple layers of deblurring features within the pyramid using a top-to-down fashion, *i.e.*, we aggregate deblurring features from fine-grained blocks to other features from coarse-grained blocks. This ensures that features with more similar ambiguity are not ignored, as they will be aggregated with features of larger perceptual fields. On the other hand, the blur of different intensities will not be incorrectly estimated because their features are aggregated from the fine-grained patches. The aggregated features $\mathcal{F}_2^T$ can be expressed as:

$$\mathcal{F}_2^T = \sum_{T=1}^{T} \mathcal{F}_1^T. \tag{2}$$

Next, features of the T layers are concatenated along the channels as non-uniform deblurring features $\mathcal{F}_2$, and the contributions of the features are redistributed through a spatial reweighting mechanism. The final output deblurred features $\mathcal{F}_{out}$ are calculated as:

$$\mathcal{F}_{out} = \mathcal{F}_2 \odot \xi\left(\varepsilon\left(\mathcal{F}_2\right)\right) + \mathcal{F}_{in}, \tag{3}$$

where $\varepsilon(\cdot)$ represents the two sets of repeated operations consisting of deep convolution and RELU activation function [44]. $\xi(\cdot)$ is the sigmoid activation function. This recombination process learns to pay adaptive attention to the correlation between deblurring features extracted at different grains. By assigning different weights to the spatial context information, better deblurred features are output.

C. Prior-Guided Reconstruction Block

We reconstructed the potential high-frequency and edge maps based on the deblurred features. However, simply concatenating image, high-frequency, and edge features to guide low-light blurry image enhancement is not helping

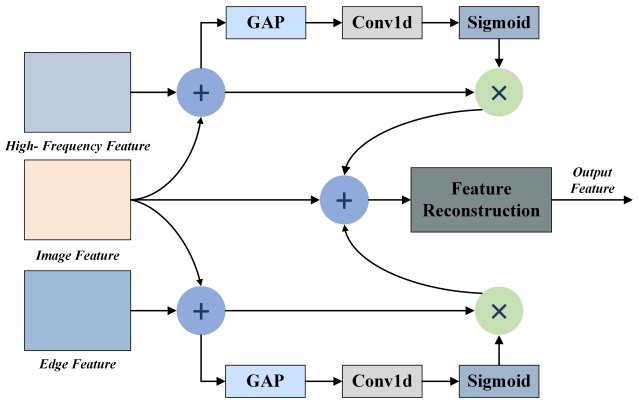

Fig. 5. The proposed architecture of the prior-guided reconstruction block.

the effective use of structural priors, because these three types of features are independent of each other. To make the high-frequency and edge information emphasize texture and structure refinement, respectively. In this paper, we propose a prior-guided reconstruction block as shown in Fig. 5, which exploits high-frequency and edge maps to refine the missing structural and textural information in the original image by adaptively computing the spatial attention weights.

Specifically, high-frequency features $\mathcal{F}_h$ and edge features $\mathcal{F}_e$ are first summed with image features $\mathcal{F}_i$ at the pixel level separately and then fed into a parallel attention mechanism built based on the ECA block [45]. The obtained features after channel reallocation can be expressed as:

$$\begin{aligned} \mathcal{F}_i^h &= \left(\mathcal{F}_i + \mathcal{F}_h\right) \odot \xi\left(Conv_1\left(G\left(\mathcal{F}_i + \mathcal{F}_h\right)\right)\right), \\ \mathcal{F}_i^e &= \left(\mathcal{F}_i + \mathcal{F}_e\right) \odot \xi\left(Conv_1\left(G\left(\mathcal{F}_i + \mathcal{F}_e\right)\right)\right), \end{aligned} \tag{4}$$

where $G(\cdot)$ represents global average pooling. $Conv_1(\cdot)$ is a 1×1 convolutional layer. Specifically, suppose the image feature map superimposed on the high-frequency feature map results in larger information response values at some locations. In that case, these regions will be given greater weights to enhance texture detail. Similar to the high-frequency map, the purpose of edge features is to refine the main structure of the image. Finally, the adaptive fusion of the deblurred deep image features is performed by combining ten residual blocks $\mathbb{R}$ for feature reconstruction, which can be written as:

$$F = \mathbb{R}\left(\mathcal{F}_i + \mathcal{F}_i^h + \mathcal{F}_i^e\right). \tag{5}$$

Overall, the parallel attention guides the cooperation between statistical features to correct the global image features, thus the output shape features compensate for the insufficient global clues.

D. Loss Functions

In this work, we integrate two reconstruction losses and two structural prior losses for end-to-end training of the network.

1) Charbonnier Loss: To enjoy better robustness and performance, we introduce Charbonnier loss [46] to supervise the network training. $\mathcal{L}_{char}$ is expressed as:

$$\mathcal{L}_{char} = \sqrt{\left\|I_{out} - I_{gt}\right\|_2 + \epsilon^2}, \tag{6}$$

where I_{gt} denotes the ground truth. The regularization term ϵ is set as 10^{-3} in all experiments.

2) Perceptual Loss: To improve the fidelity, we measure the distance between features based on the pre-trained VGG-19 model. $\mathcal{L}_{per}$ is written as:

$$\mathcal{L}_{per} = \frac{1}{C_n H_n W_n}\left\|\phi_n\left(I_{out}\right) - \phi_n\left(I_{gt}\right)\right\|_2^2, \tag{7}$$

where $\phi_n(\cdot)$ is the feature map of the n-th convolutional layer in VGG-19 model. C_n, H_n, and W_n denote the dimensions of the corresponding feature maps, respectively.

3) High-frequency Loss: By using Discrete Cosine Transform (DCT) [47] to decouple the frequency domain information of the normal-light sharp image, the high-frequency map is extracted as the corresponding ground truth I_{gt}^{fre}. We use the L1 distance to measure the similarity of the two high-frequency maps as follows:

$$\mathcal{L}_{fre} = \left\|I_{out}^{fre} - I_{gt}^{fre}\right\|, \tag{8}$$

where I_{out}^{fre} is the predicted high-frequency map of the feature reconstruction subnet.

4) Edge Loss: We extract the edge map of the normal-light sharp image by the Canny operator [48] as the corresponding ground truth I_{gt}^{edge}. This intermediate supervised process is expressed as follows:

$$\mathcal{L}_{edge} = \left\|I_{out}^{edge} - I_{gt}^{edge}\right\|, \tag{9}$$

where I_{out}^{edge} is the predicted edge map of the feature reconstruction subnet.

The total training loss function is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{char} + \lambda_1 \mathcal{L}_{per} + \lambda_2\left(\mathcal{L}_{fre} + \mathcal{L}_{edge}\right), \tag{10}$$

where λ_1 and λ_2 are constants that are used to balance the importance between the loss functions.

IV. Experiments

A. Evaluation Datasets and Metrics

We use the LOL-Blur dataset [8] to train our model and other baselines, and test the generalization of the method with a real-world low-light blurry dataset, Real-LOL-Blur [8]. The LOL-Blur dataset contains 10,200 pairs and 1,800 pairs of 1120×640 images for training and testing, while Real-LOL-Blur contains 240 low-light blurry images captured in the wild and 160 night blurry images from the RealBlur dataset [49]. For the LOL-Blur dataset, we use PSNR, SSIM [50] and LPIPS [51] to evaluate the enhancement results. For the Real-LOL-Blur, we chose MUSIQ [52], NRQM [53] and NIQE [54] as our perceptual metrics.

B. Experimental Settings

The proposed model is implemented on the PyTorch framework. The model parameters are randomly initialized and trained on an NVIDIA RTX 3090 GPU. We randomly clip 256×256 patches, flip and rotate $90°$, $180°$, $270°$ for data augmentation. The mini-batch size is 8. We train our network using the Cosine Annealing learning scheduler [55] and Adam [56] optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$. The initial learning rate is set to 10^{-4}, and the total number of iterations of the network is 500K. In all experiments, we set $T = 4$, $\lambda_1 = 0.01$, and $\lambda_2 = 0.001$.

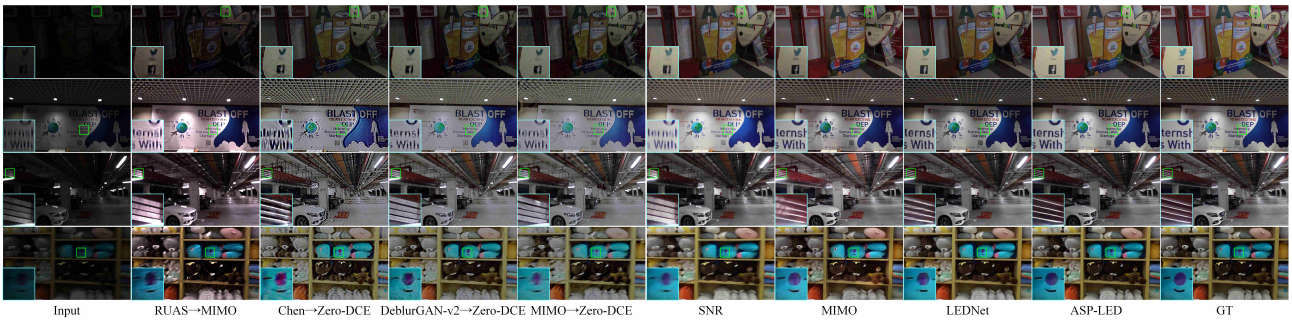

Fig. 6. Visual comparisons of the proposed method with other enhancement methods on the LOL-Blur dataset.

TABLE I

QUANTITATIVE COMPARISON ON THE LOL-BLUR DATASET.

Method	PSNR↑	SSIM↑	LPIPS↓
Zero-DCE[4] → MIMO[5]	17.68	0.542	0.510
RUAS[3] → MIMO[5]	17.81	0.569	0.523
Chen[6] → Zero-DCE[4]	17.02	0.502	0.516
DeblurGAN-v2[7] → Zero-DCE[4]	18.33	0.589	0.476
MIMO[5] → Zero-DCE[4]	17.52	0.570	0.498
KinD++[18]	21.26	0.753	0.359
DRBN[17]	21.78	0.768	0.325
SNR[16]	22.45	0.770	0.389
DeblurGAN-v2[7]	22.30	0.745	0.356
DMPHN[33]	22.20	0.817	0.301
MIMO[5]	22.41	0.835	0.262
LEDNet[8]	25.74	0.850	0.224
ASP-LED	**26.73**	**0.866**	**0.199**

C. Low-Light enhancement and Deblurring

We compare the proposed model with twelve representative baselines. These include two cascade low-light enhancement and deblurring methods Zero-DCE[4] → MIMO[5], RUAS[3] → MIMO[5], three cascade deblurring and low-light enhancement methods Chen[6] → Zero-DCE[4], DeblurGAN-v2[7] → Zero-DCE[4], MIMO[5] → Zero-DCE[4], three low-light enhancement models KinD++[18], DRBN[17], SNR[16], three deblurring models DeblurGAN-v2[7], DMPHN[33], MIMO[5], and the only advanced joint low-light enhancement and deblurring method LEDNet[8]. To compare the model performance fairly, we retrained the low-light enhancement and deblurring networks on the LOL-Blur dataset.

1) Evaluation on LOL-Blur Dataset: Table I reports the quantitative testing results of the compared methods on the LOL-Blur dataset. Our method achieves the best scores in terms of all three metrics, PSNR, SSIM, and LPIPS, which implies that the proposed method has the best ability to handle the co-existing degradation of low light and blur. During regularizing the structural priors, our method adaptively exploits spatial information from different grains to robustly learn deblurring features. Moreover, with the effective guidance of high-frequency and edge information, the proposed model performs better in low-light blurry image enhancement, with PSNR, SSIM, and LPIPS outperforming the second-best method by 0.99dB, 0.016, and 0.025, respectively. Fig. 6 shows some visual examples with different illumination levels and scenes (*e.g.*, saturated areas). Our method successfully removes blur, including saturated areas, and produces images with more proper illumination and sharper content.

TABLE II

QUANTITATIVE COMPARISON ON THE REAL-LOL-BLUR DATASET.

Method	MUSIQ↑	NRQM↑	NIQE↓
Zero-DCE[4] → MIMO[5]	39.36	5.206	4.459
RUAS[3] → MIMO[5]	34.39	3.322	6.812
Chen[6] → Zero-DCE[4]	45.79	**5.844**	5.043
MIMO[5] → Zero-DCE[4]	28.36	3.697	6.892
KinD++[18]	31.74	3.854	7.299
DRBN[17]	31.27	4.019	7.129
SNR[16]	34.58	4.662	5.310
DMPHN[33]	35.08	4.470	5.190
MIMO[5]	35.37	5.140	4.851
LEDNet[8]	39.11	5.643	4.764
ASP-LED	**46.95**	5.583	**4.442**

Fig. 7. Visual comparisons on the LOL-Blur dataset for low-light images with different ambiguity.

To demonstrate that our method can handle low-light images with the blur of different intensities, we visualize the enhancement results of low-light images with varying degrees of blur in Fig. 7. It can be observed that the results generated by other methods exist visible artifacts, blurred details, and ringing effect. The proposed model is better at restoring normal-light images with more sharp texture.

2) Evaluation on Real-World Dataset: Table II reports the evaluation results of MUSIQ, NRQM, and NIQE metrics. Our model achieves the best MUSIQ and NIQE scores, and NRQM obtains competitive results with the best scores. This indicates that our method has the best perceptual results in color contrast and sharpness, and the image quality is generally superior to other methods. Fig. 8 shows visual com-

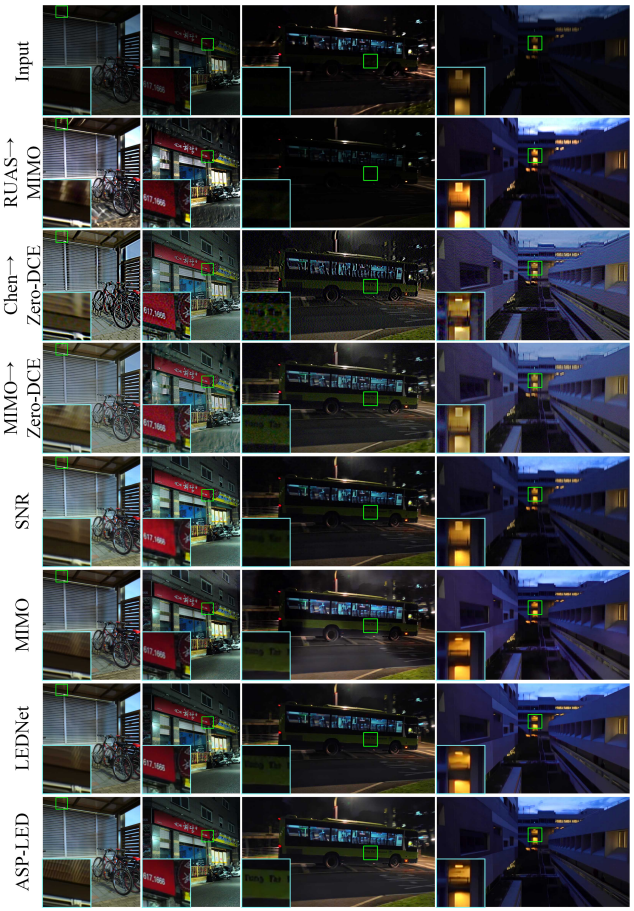

Fig. 8. Visual comparisons of the proposed method with other enhancement methods on the Real-LOL-Blur dataset.

TABLE III

ABLATION STUDIES OF STRUCTURAL PRIOR LOSS FUNCTIONS.

Method	PSNR↑	SSIM↑	LPIPS↓
$\mathcal{L}_{char}$ + $\mathcal{L}_{per}$	25.56	0.855	0.211
$\mathcal{L}_{char}$ + $\mathcal{L}_{per}$ + $\mathcal{L}_{fre}$	26.09	0.860	0.203
$\mathcal{L}_{char}$ + $\mathcal{L}_{per}$ + $\mathcal{L}_{edge}$	26.33	0.863	0.202
ASP-LED	**26.73**	**0.866**	**0.199**

parisons of low-light blurry images captured in various real-world scenes. These results also manifest that our method is effective in recovering normal-light sharp images with better human perception.

D. Ablation Study

To demonstrate the effectiveness of the proposed method, including structural priors and network structures, we perform a series of ablation experiments on the LOL-Blur dataset.

1) Structural Prior Study: Table III investigates the contributions of the high-frequency and edge loss terms in Eq. 10. The addition of the high-frequency or edge loss terms to $\mathcal{L}_{char}$ and $\mathcal{L}_{per}$ both enhances the model performance. Besides, by introducing two prior loss terms simultaneously, the quantitative metrics of our method achieve further improvement. With the joint regularization of $\mathcal{L}_{fre}$ and $\mathcal{L}_{edge}$, ASP-LED enjoys the ability to model sharper texture and structure.

TABLE IV

ABLATION STUDIES OF THE MULTI-PATCH PERCEPTION PYRAMID BLOCK.

Method	PSNR↑	SSIM↑	LPIPS↓
w/o MPP	25.76	0.853	0.217
Multi-Patch with Standard Convolution	25.87	0.859	0.204
Multi-Patch with Dilation Convolution	26.07	0.861	0.203
Multi-Patch with Aggregation	26.29	0.864	0.200
ASP-LED	**26.73**	**0.866**	**0.199**

TABLE V

ABLATION STUDIES OF THE PRIOR-GUIDED RECONSTRUCTION BLOCK.

Method	PSNR↑	SSIM↑	LPIPS↓
Concatenation Directly	26.08	0.860	0.205
Concatenation with an Attention	25.90	0.859	0.206
Cascade Attention (High-Fre → Edge)	26.33	0.861	0.201
Cascade Attention (Edge → High-Fre)	26.08	0.859	0.206
ASP-LED	**26.73**	**0.866**	**0.199**

2) Network Design Study: To verify the effectiveness of the proposed multi-patch perception pyramid block, we investigated the four baselines listed in Table IV. The results broadly validate our design choices. It shows that removing the MPP block degrades the model performance, which illustrates the necessity of effectively exploiting the multi-grained spatial features for deblurring (by comparing the result in row 1st with row 5th). In addition, by extracting deblurring features using different perceptual fields in parallel pyramids or aggregating deblurring features from top to bottom, low-light enhancement performance can be further improved (compare the results in row 2nd or 4th with row 3rd). Compared to the above baselines, ASP-LED achieves better enhancement performance by applying spatial adaptive weighted fusion of multi-grained deblurring features at the tail of the MPP block (see the last row).

To explore the effective way of fusing image, high-frequency, and edge information, we designed the four baselines listed in Table V for comparison. We observe that compared to Concatenation Directly, only Cascade Attention (High-Fr → Edge) achieves a slight performance gain, and the performance of the other two fusion methods does not improve. However, the proposed prior-guided reconstruction block has significant advantages in all metrics, indicating that ASP-LED can successfully incorporate the learned high-frequency and edge information into the enhancement process to recover sharper textures and structures.

V. CONCLUSION

In this paper, we propose a novel ambiguity-aware framework ASP-LED guided by structural priors for jointly addressing low light and blur co-exist degradation. Using the perceived multi-grained deblurring spatial features, the multi-patch perception pyramid block is able to remove non-uniform blur and facilitate the reconstruction of potential high-frequency and edge information. Then, with the proposed prior-guided reconstruction block, global image features are adaptively corrected based on statistical information to improve texture and structure reconstruction. Experiments on low-light blurry datasets show that our method outperforms state-of-the-art methods with the ability to restore visually better normal-light sharp images, and could lay a sound basis for future vision-related intelligent tasks.

REFERENCES

[1] J. Wen, J. Cui, Z. Zhao, R. Yan, Z. Gao, L. Dou, and B. M. Chen, "Syreanet: A physically guided underwater image enhancement framework integrating synthetic and real images," *arXiv preprint arXiv:2302.08269*, 2023.

[2] Q. Wen, Y. Wu, and Q. Chen, "Video waterdrop removal via spatio-temporal fusion in driving scenes," *arXiv preprint arXiv:2302.05916*, 2023.

[3] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 561–10 570.

[4] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1780–1789.

[5] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 4641–4650.

[6] L. Chen, J. Zhang, S. Lin, F. Fang, and J. S. Ren, "Blind deblurring for saturated images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6308–6316.

[7] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 8878–8887.

[8] S. Zhou, C. Li, and C. Change Loy, "Lednet: Joint low-light enhancement and deblurring in the dark," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI*. Springer, 2022, pp. 573–589.

[9] S. Zhou, J. Zhang, W. Zuo, H. Xie, J. Pan, and J. S. Ren, "Davanet: Stereo deblurring with view aggregation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 996–11 005.

[10] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1752–1758, 2007.

[11] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE transactions on consumer electronics*, vol. 53, no. 2, pp. 593–600, 2007.

[12] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE transactions on image processing*, vol. 22, no. 3, pp. 1032–1041, 2012.

[13] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.

[14] S. Park, S. Yu, B. Moon, S. Ko, and J. Paik, "Low-light image enhancement using variational optimization-based retinex model," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 2, pp. 178–184, 2017.

[15] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.

[16] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 714–17 724.

[17] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality," *IEEE Transactions on Image Processing*, vol. 30, pp. 3461–3473, 2021.

[18] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *International Journal of Computer Vision*, vol. 129, pp. 1013–1037, 2021.

[19] D. Zhang, J. Zhou, W. Zhang, Z. Lin, J. Yao, K. Polat, F. Alenezi, and A. Alhudhaif, "Rex-net: A reflectance-guided underwater image enhancement network for extreme scenarios," *Expert Systems with Applications*, p. 120842, 2023.

[20] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE transactions on image processing*, vol. 30, pp. 2340–2349, 2021.

[21] Y. Fu, Y. Hong, L. Chen, and S. You, "Le-gan: Unsupervised low-light image enhancement network using attention module and identity invariant loss," *Knowledge-Based Systems*, vol. 240, p. 108010, 2022.

[22] K. G. Lore, A. Akintayo, and S. Sarkar, "Llnet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognition*, vol. 61, pp. 650–662, 2017.

[23] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 1632–1640.

[24] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4225–4238, 2021.

[25] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *Acm transactions on graphics (tog)*, vol. 27, no. 3, pp. 1–10, 2008.

[26] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-laplacian priors," *Advances in neural information processing systems*, vol. 22, 2009.

[27] X. Yu, F. Xu, S. Zhang, and L. Zhang, "Efficient patch-wise non-uniform deblurring for a single image," *IEEE Transactions on Multimedia*, vol. 16, no. 6, pp. 1510–1524, 2014.

[28] L. Pan, R. Hartley, M. Liu, and Y. Dai, "Phase-only image based kernel estimation for single image blind deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6034–6043.

[29] Z. Hu, S. Cho, J. Wang, and M.-H. Yang, "Deblurring low-light images with light streaks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3382–3389.

[30] S. Nah, T. Hyun Kim, and K. Mu Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3883–3891.

[31] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8174–8182.

[32] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14 821–14 831.

[33] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5978–5986.

[34] X. Hu, W. Ren, K. Yu, K. Zhang, X. Cao, W. Liu, and B. Menze, "Pyramid architecture search for real-time image deblurring," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4298–4307.

[35] Z. Shen, W. Wang, X. Lu, J. Shen, H. Ling, T. Xu, and L. Shao, "Human-aware motion deblurring," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5572–5581.

[36] Y. Bai, H. Jia, M. Jiang, X. Liu, X. Xie, and W. Gao, "Single-image blind deblurring using multi-scale latent structure prior," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2033–2045, 2019.

[37] Y. Yuan, W. Su, and D. Ma, "Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3555–3564.

[38] S. Zheng, Z. Zhu, J. Cheng, Y. Guo, and Y. Zhao, "Edge heuristic gan for non-uniform blind deblurring," *IEEE Signal Processing Letters*, vol. 26, no. 10, pp. 1546–1550, 2019.

[39] Z. Fu, Y. Zheng, T. Ma, H. Ye, J. Yang, and L. He, "Edge-aware deep image deblurring," *Neurocomputing*, vol. 502, pp. 37–47, 2022.

[40] Y. Liu, F. Fang, T. Wang, J. Li, Y. Sheng, and G. Zhang, "Multi-scale grid network for image deblurring with high-frequency guidance," *IEEE Transactions on Multimedia*, vol. 24, pp. 2890–2901, 2021.

[41] S. Nah, S. Baik, S. Hong, G. Moon, S. Son, R. Timofte, and K. Mu Lee, "Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[42] Y. Zhao, Y. Xu, Q. Yan, D. Yang, X. Wang, and L.-M. Po, "D2hnet: Joint denoising and deblurring with hierarchical network for robust

night image restoration," in *European Conference on Computer Vision*. Springer, 2022, pp. 91–110.
[43] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.
[44] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," *Advances in neural information processing systems*, vol. 30, 2017.
[45] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 534–11 542.
[46] D. Triantafyllidou, S. Moran, S. McDonagh, S. Parisot, and G. Slabaugh, "Low light video enhancement using synthetic data produced with an intermediate domain mapping," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*. Springer, 2020, pp. 103–119.
[47] K. R. Rao and P. Yip, *Discrete cosine transform: algorithms, advantages, applications*. Academic press, 2014.
[48] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.
[49] J. Rim, H. Lee, J. Won, and S. Cho, "Real-world blur dataset for learning and benchmarking deblurring algorithms," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*. Springer, 2020, pp. 184–201.
[50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
[51] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.
[52] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "Musiq: Multi-scale image quality transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5148–5157.
[53] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Computer Vision and Image Understanding*, vol. 158, pp. 1–16, 2017.
[54] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.
[55] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.
[56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.