Real-world low-light image enhancement *via* domain-gap aware framework and reverse domain-distance guided strategy

Yong Chen[1], Meiyong Huang[1]*, Huanlin Liu[2], Kaixin Shao[1] and Jinliang Zhang[1]

[1]The Key Laboratory of Industrial Internet of Things and Network Control, Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing, China, [2]School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China

Low-light image enhancement (LLIE) has high practical value and development potential in real scenarios. However, the current LLIE methods reveal inferior generalization competence to real-world low-light (LL) conditions of poor visibility. We can attribute this phenomenon to the severe domain bias between the synthetic LL domain and the real-world LL domain. In this article, we put forward the Domain-Gap Aware Framework, a novel two-stage framework for real-world LLIE, which is the pioneering work to introduce domain adaptation into the LLIE. To be more specific, in the first stage, to eliminate the domain bias lying between the existing synthetic LL domain and the real-world LL domain, this work leverages the source domain images *via* adversarial training. By doing so, we can align the distribution of the synthetic LL domain to the real-world LL domain. In the second stage, we put forward the Reverse Domain-Distance Guided (RDDG) strategy, which takes full advantage of the domain-distance map obtained in the first stage and guides the network to be more attentive to the regions that are not compliance with the distribution of the real world. This strategy makes the network robust for input LL images, some areas of which may have large relative domain distances to the real world. Numerous experiments have demonstrated the efficacy and generalization capacity of the proposed method. We sincerely hope this analysis can boost the development of low-light domain research in different fields.

KEYWORDS

real-world low-light image enhancement, domain-gap aware framework, domain adaptation, reverse domain-distance guided strategy, adversarial training

1 Introduction

Real-world LLIE aims to reconstruct normal light images from observations acquired under low-light conditions with low visibility and poor details. Numerous scientific deep-learning approaches [1–5] with the advantage of the powerful capability to learn features [6–10] have been extensively proposed. For the efficient LLIE task [11–13], they recover the visibility and precise details of low-illumination images by learning the relationship between LL and NL images. As the efficiency of deep learning methods is subordinate to the dataset, some methods collect bursts of images with multiple exposure levels captured in real scenarios for real-world LLIE applications [14, 15]. However, since the collections of large-scale paired datasets are incredibly laborious and expensive [16], the existing paired datasets

are usually of small scale, which may cause overfitting when training networks using them. Therefore, some methods have been put forward to enlarge the scale of datasets by synthesizing low-illumination images and forming paired datasets with normal illumination images [17, 18]. However, the synthetic LL images are usually not compliant with real-world distribution, leading to poor generalization capability to the real world for the LLIE methods trained on these datasets [19]. Specifically, the illumination level cannot be improved sufficiently to recover details, or the white balance cannot be maintained correctly. Therefore, it is a worthwhile but challenging task to generate enhancement results that match real-world distribution.

Unsupervised methods are of high practical value and development potential because they do not require paired datasets captured in the same static scenarios [20, 21]. They implement LLIE tasks by taking full advantage of unpaired real-world NL images and LL images. To realize the concept of aligning the distribution of enhanced NL images to the unpaired NL domain, existing methods usually adopt adversarial training directly for the enhanced results against the real-world NL images. Further, to ensure that all regions in the enhanced images are close to the real ones, EnlightenGAN [22] crops image patches randomly from the enhanced images and adopts adversarial training against the real NL image patches. However, these methods seldom notice that the severe domain gap may impede the enhancement performance but only focus on the enhancement procedure, which degrades the generalization performance of the networks trained on synthetic datasets. Moreover, randomly comparing image patches does not guarantee that all regions of the enhanced images match the real-world distribution.

Over recent years, researchers have extensively proposed ways to address the shortage of data with labels for training. For Domain adaptation (DA) methods, the labeled data enables adequate training in the source domain as well as performing new tasks on the unlabeled target domain with new distribution [23]. It greatly improves the effectiveness of methods on the target domain, which is appropriate for real-world LLIE tasks.

In this article, by comprehensively reviewing the potential and reaping the full benefits of alternative methods, we put forward a two-stage framework with the merit of both adversarial learning and domain adaptive methods. Specifically, we propose the Domain-Gap Aware Framework to implement real-world LLIE tasks, which addresses the issue that the input LL images deviate from the real-world distribution.

As shown in Figure 1, the noticeable domain gap between the real-world and the synthetic LL domain can be observed. Besides, different areas in a single low-light image may have different relative domain distances. We find that the domain gap severely degrades the generalization competency of the network to real-world low-light conditions. Therefore, unlike existing methods that ignore the domain bias for synthetic LL images, we propose the Domain-Gap Aware Framework. Specifically, in the first stage, we impose adversarial training on the Darkening Network to eliminate the severe domain gap and generate realistic pseudo-LL images. By doing so, we obtain pseudo-LL images that are consistent with real-world distribution, as well as domain-distance maps. In the second stage, we propose the Reverse Domain-Distance Guided strategy to capitalize fully on the domain-distance maps and mitigate the unrealistic areas of pseudo-LL images. In detail, we assign higher
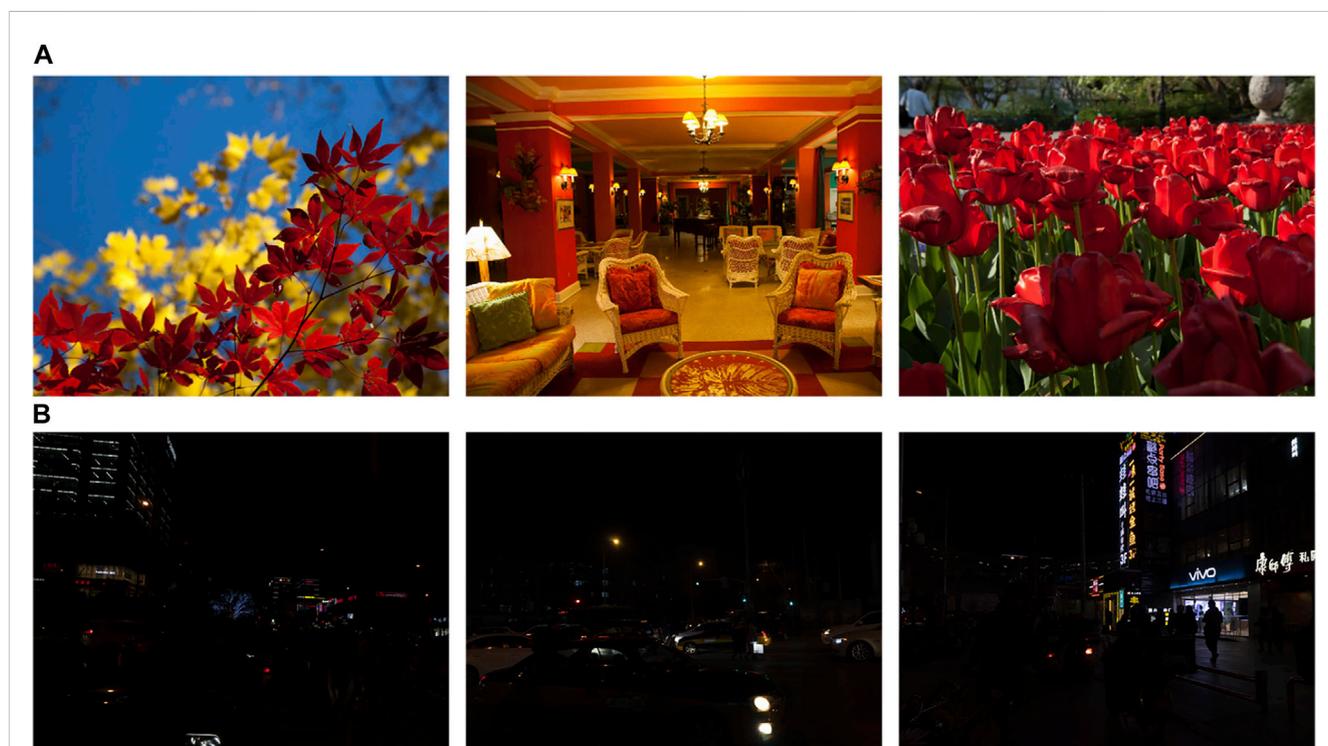
FIGURE 1
(A) The presentation of the existing synthetic LL dataset and **(B)** the real-world LL images. **(A,B)** shows the apparent domain gap in terms of illumination level and white balance lying between the real-world and the existing synthetic LL domain.

weights to the regions in the generated NL images that are relatively far from the real-world domain; while assigning smaller weights to the realistic regions in the training phase, thus mitigating the uncompetitive enhancement competence to real-world scenarios due to the unrealistic input LL patches. The proposed two-stage framework generalizes well to the real world with boosted illumination level and clearly reconstructed structural details, which can significantly facilitate subsequent computer vision tasks and systems [24] focusing on objects at nighttime.

The following are the key contributions to this article:

- We put forward the Domain-Gap Aware Framework to address the domain-gap issue and generate pseudo-low-light images consistent with real-world distribution, which is essential to attain models with high generalization capability for real-world LLIE.
- A Reverse Domain-Distance Guided strategy is proposed for real-world applications. The pixel-wise domain distance maps are taken full advantage of to further promote the robustness of the Enlightening Network. It is worth pointing out that this is the pioneering work to introduce DA to LLIE as far as we know.

The remainder of this paper is structured as follows. In Section 2, we present a brief review of some related works in the LLIE field. Section 3 introduces the proposed framework and strategy. Section 4 shows experimental results to demonstrate the effect of our method, and Section 5 gives a conclusion of the paper.

2 Related work

2.1 CNN-based approaches

CNN-based approaches have become a principal method in the LLIE field with their high efficiency in image analysis [25, 26]. They reconstruct the contrast and structural details of LL images by learning the mapping relationship between LL-NL images. Some methods have collected paired data in real scenarios [27]. However, it is difficult to construct large-scale paired datasets due to the required high cost and heavy workforce. Since the applications of deep learning methods are usually hampered by shortages of data in pairs for training [28], some methods have also made attempts to construct simulated datasets [17, 18, 29]. It is widely known that the data for training are essential for the networks' performance [30]. However, the synthetic dataset was generated under the assumption of simple degradation in terms of illumination level, noise, etc., which leads to the poor generalization of the trained networks to the real world and the side effect, e.g., color distortion and insufficiently improved illumination.

Real-world LLIE has attracted significant research due to its high practical value. Researchers have been extensively designing diverse architectures to achieve better generalization to the real world. In EnlightenGAN [22], the design philosophy is to address the domain gap issue by applying adversarial training using unpaired datasets. In addition, researchers have also made efforts to zero-shot LLIE. Zero-DCE [31] regards the LLIE as a task of curve estimation for image-wise dynamic range adjustment. However, it pays little attention to the domain gap between the to-be-enhanced LL images and the real-

world ones and only focuses on the enhanced NL images, which degrades the generalization performance of the networks trained on synthetic datasets.

2.2 Domain adaptation

Domain adaptation (DA) intends to enhance performance when confronting a new target domain despite domain bias [32]. It is beneficial to deal with data shortages for tasks that are difficult to obtain real data.

In this work, we concentrate on eliminating the domain gap to synthesize realistic LL images, which is a preparation phase for the enlightening stage. Inspired by relevant studies in super-resolution [33], we construct a Domain-Distance Aware framework to perform the real-world LLIE. We apply DA to improve the performance of LLIE on real data.

In the next sections, we introduce the proposed Domain-Distance Aware framework and Reverse Domain-Distance Guided strategy in detail.

3 Methods

3.1 Network architecture

Given two domains, which can be described as the LL domain and the NL domain, our goal is to learn an Enlightening Network to promote the visibility and reconstruct structural details of the images in the LL domain while generating enhanced NL estimations belonging to the real-world NL domain. To achieve this objective, we propose the Domain-Gap Aware Framework. We did not follow previous work that directly utilizes the existing synthetic low-illumination datasets to train the Enlightening Network. Instead, our framework takes the domain bias between x^g and x^r into full account. As shown in Figure 2, during the first phase, we train the Darkening Network using adversarial training, which generates pseudo-LL images belonging to the real-world LL domain as well as domain distance maps. Then, in the second stage, we put forward the Reverse Domain-Distance Guided strategy, which leverages the pseudo-LL-NL image pairs and domain distance maps to train the Enlightening Network.

In the next subsection, we first describe how to train the Darkening Network to generate LL-NL image pairs in line with real-world distributions. Then, we show the Reverse Domain-Distance Guided strategy.

3.1.1 Training of darkening network

The general procedure of synthesizing low-light images by existing methods is manually adjusting the illumination and adding noise [17, 18]. However, the illumination levels in the real world are diverse and may also vary spatially in a single image. Moreover, it is difficult to represent noise with simple and known distribution. In a word, the degradations assumed by existing methods are too simple to fully simulate the complex degradation in the real world, which unfortunately leads to domain bias lying between the synthesized LL images and the real-world ones. In contrast, our approach employs a deep

FIGURE 2
(A) The proposed Domain-Gap Aware Framework contains two stages, i.e., the Darkening stage and Enlightening stage. **(B)** Adversarial training during the training of Darkening Network.
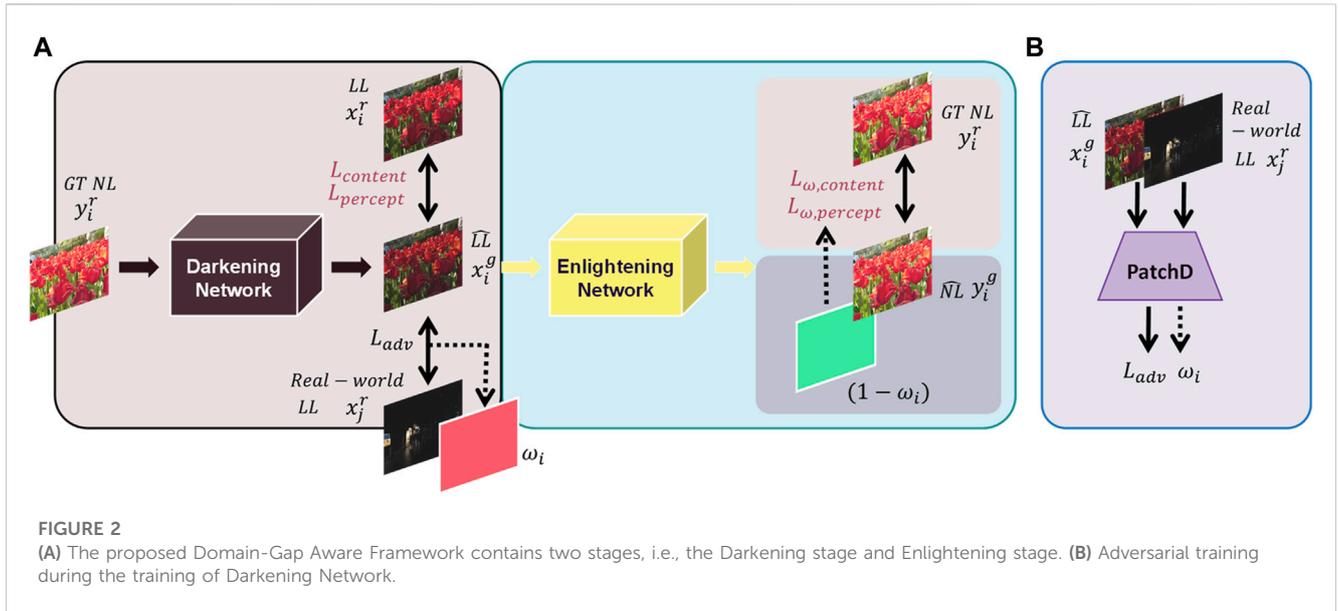
network (i.e., the Darkening Network) to learn the real-world degradation process. It works as the generator in the whole framework and extracts the features of NL images using eight blocks (each layer is convolved by a 3×3 kernel and activated by a ReLU activation in between).

3.1.1.1 Losses

We employ multiple loss functions to train the Darkening Network. To ensure that the content of the pseudo-LL images is preserved consistently with the GT (Ground-Truth) LL images, we adopt content loss along with the perceptual loss to optimize the distance between them at the image level and the feature level, respectively. In detail, the content loss contains reconstruction loss, which is L1-norm and SSIM (Structural SIMilarity Index) [34] loss, which aims at measuring structural similarities between two images. The reason why we adopt L1-norm as our reconstruction loss is that it treats all errors equally so that the training can keep going even though the error is tiny. Perceptual loss is also widely used in the image reconstruction field, which measures the distance between features extracted *via* deep neural networks.

$$L_{Recons.} = \mathbf{E}_{y^r} \left\| x_i^r - DN\left(y_i^r\right) \right\|_1$$

$$SSIM\left(m, n\right) = \frac{\left(2\mu_m \mu_n + c_1\right)\left(2\sigma_{mn} + c_2\right)}{\left(\mu_m^2 + \mu_n^2 + c_1\right)\left(\sigma_m^2 + \sigma_n^2 + c_2\right)}$$

$$L_{SSIM} = 1 - SSIM\left(x_i^r, DN\left(y_i^r\right)\right)$$

$$L_{content} = L_{Recons.} + L_{SSIM}$$

$$L_{percept} = \mathbf{E}_{y^r} \left\| \Phi\left(x_i^r\right) - \Phi\left(DN\left(y_i^r\right)\right) \right\|_1$$

We show the adopted loss functions above, where $\Phi(\cdot)$ denotes the convolutional layers of the conv5_3 of VGG-16 [35], and SSIM$(\cdot, \cdot)$ means the SSIM score between two input images.

In addition to the above training, to address the domain gap issue and align the distribution of the pseudo-LL images to the real world, the pseudo-LL images are trained against the real-world LL images by adversarial training. Specifically, we adopt a similar strategy as DASR

[33], which uses a patch discriminator with four layers of fully convolutional layers to determine whether each image block matches the real-world distribution. This strategy facilitates pseudo-LL images to fit the real-world distribution.

$$L_{Adv.}^G = -\mathbf{E}_{y^r}\left[logD\left(DN\left(y^r\right)\right)\right]$$

$$L_{Adv.}^D = -\mathbf{E}_{y^r}\left[logD\left(x^r\right)\right] - \mathbf{E}_{y^r}\left[log\left(1 - D\left(DN\left(y^r\right)\right)\right)\right]$$

The loss functions are shown above, where $D(\cdot)$ denotes the patch discriminator.

3.1.2 Reverse domain-distance guided strategy

As shown in Figure 1 previously, each region in the generated LL image may distant diversely from the domain of the real world, i.e., some regions lie relatively close to the domain of the real world, while some regions are relatively far. Since the regions relatively far from the domain of the real world may degrade the enhancement competency of the network, we should endow different regions with diverse attention. We realize this concept by reversing the domain distance maps first and then applying them to eliminate the discrepancy between y_i^g and x_i^r. Thereby, we adaptively adjust the loss functions by assigning diverse weight parameters to these regions adaptively. We present the Reverse Domain-Distance Guided strategy in Figure 3.

3.1.2.1 Losses

We denote the supervised losses as follows, where ω_i denotes the domain distance map for x_i^g, which is attained by the patch Discriminator trained in the first stage.

$$L_{\omega,content} = -\mathbf{E}_{x_i^g, y_i^r} \left\| \left(1 - \omega_i\right) \odot \left(EN\left(x_i^g\right) - y_i^r\right) \right\|_1$$

$$L_{\omega,percept} = -\mathbf{E}_{x_i^g, y_i^r} \left\| \left(1 - \omega_i\right) \odot \left(\Phi\left(EN\left(x_i^g\right)\right) - \Phi\left(y_i^r\right)\right) \right\|_1$$

The trained patch Discriminator can differentiate between the pseudo-LL patches and those from the real-world domain. A smaller value in ω_i means a lower probability that the pseudo-LL patches
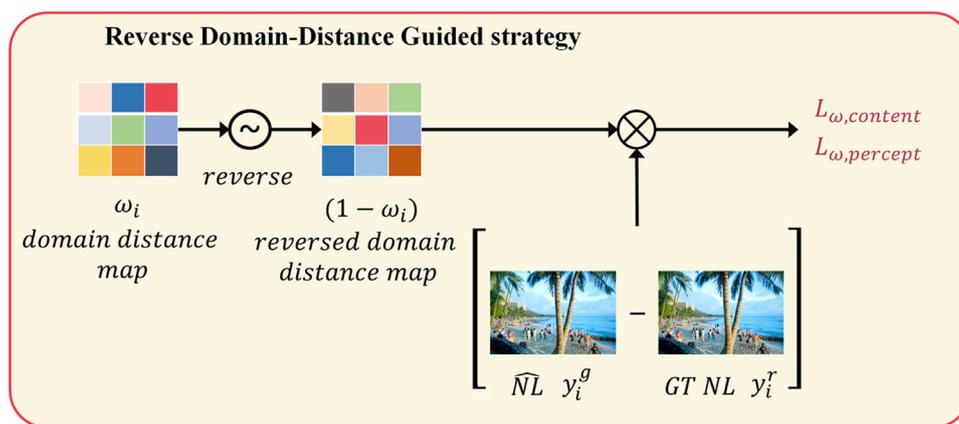
FIGURE 3
The proposed Reverse Domain-Distance Guided strategy. It facilitates the Enlightening Network to be more attentive to the less realistic regions of the input LL images.

belong to the real-world domain. It also indicates a higher value in the reverse of ω_i, i.e., $1- \omega_i$, and a larger domain distance from the pseudo-LL to the real-world domain. Therefore, we guide the network to be attentive to the enhanced outcomes of the input pseudo-LL patches relatively far from the real-world domain by endowing distance-related importance to different areas. The Reverse Domain-Distance Guided strategy makes full use of domain distance to remedy the unrealistic areas of pseudo-LL images and further improves the generalization to the real world.

To evaluate the proposed method, we describe experimental settings and results in detail in the next section.

4 Experiments

Since the similarity with the ground-truth NL images can reflect the enhanced result to a large extent, we adopt PSNR and SSIM [34] as reference metrics, two widely adopted quality metrics in the image restoration field. In addition, as our method adopts a generative-adversarial network, we also focus on perceptual quality. Therefore, we also adopt LPIPS (Learned Perceptual Image Patch Similarity [36] as the quality metric. Diverse ablation studies are carried out by us to figure out the effect of the proposed strategies in our framework. Then, to figure out our method's generalization competency, the real-world LL dataset is assigned as the testing set. Finally, we further make comparisons with other competing LLIE approaches by applying them to real-world LL datasets.

4.1 Training settings

Researchers have constructed MIT-Adobe FiveK dataset [37], which consists of 5,000 photos retouched by five experts, to adjust the global tone. It has been widely leveraged in the LLIE field. We applied GladNet [38] to the normal-light images retouched by MIT-Adobe FiveK dataset's Expert E to obtain synthetic LL images. We separate 4,000 paired NL-LL images from it to prepare for training

and 1,000 paired images to prepare for validation. Then we resize the images to 600×400 resolution. Besides, we adopt the DARK FACE dataset (4,000 for training, 1,000 for validation) [39], which consists of 6,000 images obtained under real-world nighttime conditions, as the real-world low-light references.

Let us now turn our attention to the main framework. The network is assigned random weights initially. The Adam method (with momentum and weight decay set to be 0.9 and 0.001, respectively) is adopted to update the network's parameters. Besides, the learning rate is assigned to be 0.0001 initially and then is halved every ten epochs. During the whole training procedure, we maintain a batch size of 16. We carry out all the evaluation experiments on the NVIDIA GeForce GTX3090 and NVIDIA GeForce GTX1080Ti with PyTorch.

4.2 Ablation studies

Before conducting comparison experiments with recently competing methods, we carried out a variety of ablation experiments to delve into diverse loss functions as well as the proposed framework.

4.2.1 Effect of loss functions
We carry out a variety of trying outs for diverse loss functions and figure out the quantitative outcomes on the widely adopted metrics, i.e., PSNR and SSIM, along with LPIPS. During the computation of LPIPS, we extract features of input images through AlexNet [40] to calculate the distance between them. A small LPIPS value means a high similarity. Table 1 displays the quantitative outcomes.

Firstly, let us analyze the effect of each loss function. We can find from the 3rd, 4th, 5th row that in comparison to being supervised by the reconstruction loss and adversarial loss, adding SSIM loss boosts the performance on PSNR, SSIM, and LPIPS metrics with 3.871dB, 0.0654dB, and 0.1177dB, respectively, and adding perceptual loss improves the performance on the three metrics with 3.19dB,

TABLE 1 Ablation study of diverse loss functions and corresponding image quality evaluations.

Recons.	SSIM	Adv	Percept	PSNR↑	SSIM↑	LPIPS↓
√	×	×	×	**20.107**	0.7704	0.1023
√	√	×	×	20.028	**0.7772**	**0.0960**
√	×	√	×	15.761	0.6995	0.2218
√	√	√	×	19.632	0.7649	0.1041
√	×	√	√	18.951	0.7190	0.1880
√	√	√	√	20.079	0.7718	0.1035

Recons., Adv., and Percept. indicate reconstruction loss function, adversarial loss, and perceptual loss, respectively. Scores marked in bold indicate the highest scores on the corresponding metric.

0.0195dB, 0.0338dB, respectively. It indicates the effectiveness of both SSIM loss and perceptual loss in reconstructing texture and details of contents.

Secondly, we can find from the 1st and 2nd row that the best performance is achieved under the settings of using merely content loss, which includes reconstruction loss and SSIM loss. Note that content loss aims at reducing the distance between input images. Therefore, training with them equals supervised learning, which easily achieves better performance in comparison to adversarial training. Nevertheless, our method, which contains adversarial learning for fitting with real-world low-light image distributions, achieves similar quantitative results with supervised learning. In specific, our method performs second best on PSNR and SSIM and obtains rank third on LPIPS with a difference of only 0.028dB, 0.0054dB, and 0.0075 dB with the rank first scores.

As shown qualitatively in Figure 4, our method generates images with sufficiently low exposure levels and correct white balance. Therefore, we can conclude that the LL images generated by our method keep contents consistent with LL images from the existing dataset but closer to the ones captured under poor light conditions. We finally chose $\omega_{col} = 1$, $\omega_{ssim} = 1$, $\omega_{per} = 0.02$, and $\omega_{adv} = 0.02$ for the weight parameters of each loss function.

4.2.2 Effect of the darkening network
4.2.2.1 Comparisons of LL images

We adopt a generative adversarial network to generate pseudo-LL images so that they are close to LL images from the existing dataset in terms of contents while in compliance with the distribution of real-world LL images. Figure 4 presents the contrast between the MIT-Adobe FiveK dataset [37] and pseudo-LL images synthesized by our method. We can find in the 2nd row of the panel (A) and (B) in Figure 4 that the white balance of several images in the existing low-light dataset is going wrong, where white areas of the original NL images appear orange in the existing low-light dataset. This may lead to the color shift in the enhanced images enhanced, which is further proven in Figure 5. Besides, the 2nd row of Figure 4B suggests that the illumination level is not low enough to simulate night lighting conditions. In contrast, the proposed Darkening Network maintains the correct white balance and decreases the illumination level sufficiently in LL versions, as displayed in the 3rd row of the panel (A) and (B) in Figure 4, which facilitates the lightening network to generalize better to the real-world low-light condition.

4.2.2.2 Comparisons of enhancement results

Furthermore, the effect of the proposed Darkening Network is investigated in this subsection. We train the Darkening Network both using pseudo-LL-NL pairs $\{x^g, y^r\}$ and existing paired dataset $\{x^r, y^r\}$, and compare the outcomes in terms of quality and quantity. Figure 5 shows qualitative comparisons. As shown in Figure 5A, we can clearly see that the enhanced outcomes of the existing LL dataset suffer from the color shift. This is because the input LL images are of imperfect white balance, as shown in Figure 4 previously. Besides, as shown in Figure 5C, it easily appears over-exposure, which hinders some regions (e.g., regions in the dark color such as hair, ribbon, skin, and so on.) from retaining semantic darkness, unfortunately. This is due to the insufficient illumination level in existing LL images. Moreover, the enhanced results of the backlit image (in the 5th row of Figure 5C) suffer from artifacts severely. In contrast, as shown in Figures 5B, D, the enhanced results of pseudo-LL images are of correct white balance and appropriate exposure level with good preservation of semantic information, as well as much fewer artifacts introduced to backlit images. Therefore, we confirm that the Enlightening Network can produce superior enhancement results collaborated with the Darkening Network, which fully reflects the effect of the Darkening Network.

Next, we display quantitative comparison results in Table 2.

We can clearly find that training with pseudo-LL-NL pairs $\{x^g, y^r\}$ achieves better scores on PSNR, SSIM, and LPIPS, which exceeds the GT LL-NL pairs $\{x^r, y^r\}$ to a great extent. More specifically, training with $\{x^g, y^r\}$ achieves considerably higher scores on PSNR and SSIM than training with $\{x^r, y^r\}$, both with and without the Reverse Domain-Distance Guided strategy. More specifically, as shown in column2 and 4, training with $\{x^g, y^r\}$ exceeds training with $\{x^r, y^r\}$ by 16.252 dB (= 35.276–18.751 dB) on PSNR and by 0.1022 dB (= 0.9696–0.8674 dB) on SSIM. It demonstrates that the synthesized pseudo-LL-NL pairs are more suitable for real-world LLIE than GT LL-NL pairs from the existing dataset. The reason is that the proposed Darkening Network aims to address the domain gap *via* adversarial training so that the pseudo-LL images match the real-world distribution in terms of exposure level, hue, noise, and so on. However, the existing procedure of synthesizing low-light images assumes simple degradations from NL images, which is far from the complex degradations of the real world. Therefore, we confirm that by adequately taking advantage of target domain data during the training process, the proposed Darkening Network makes a significant contribution to the improvement of enhancement quality.
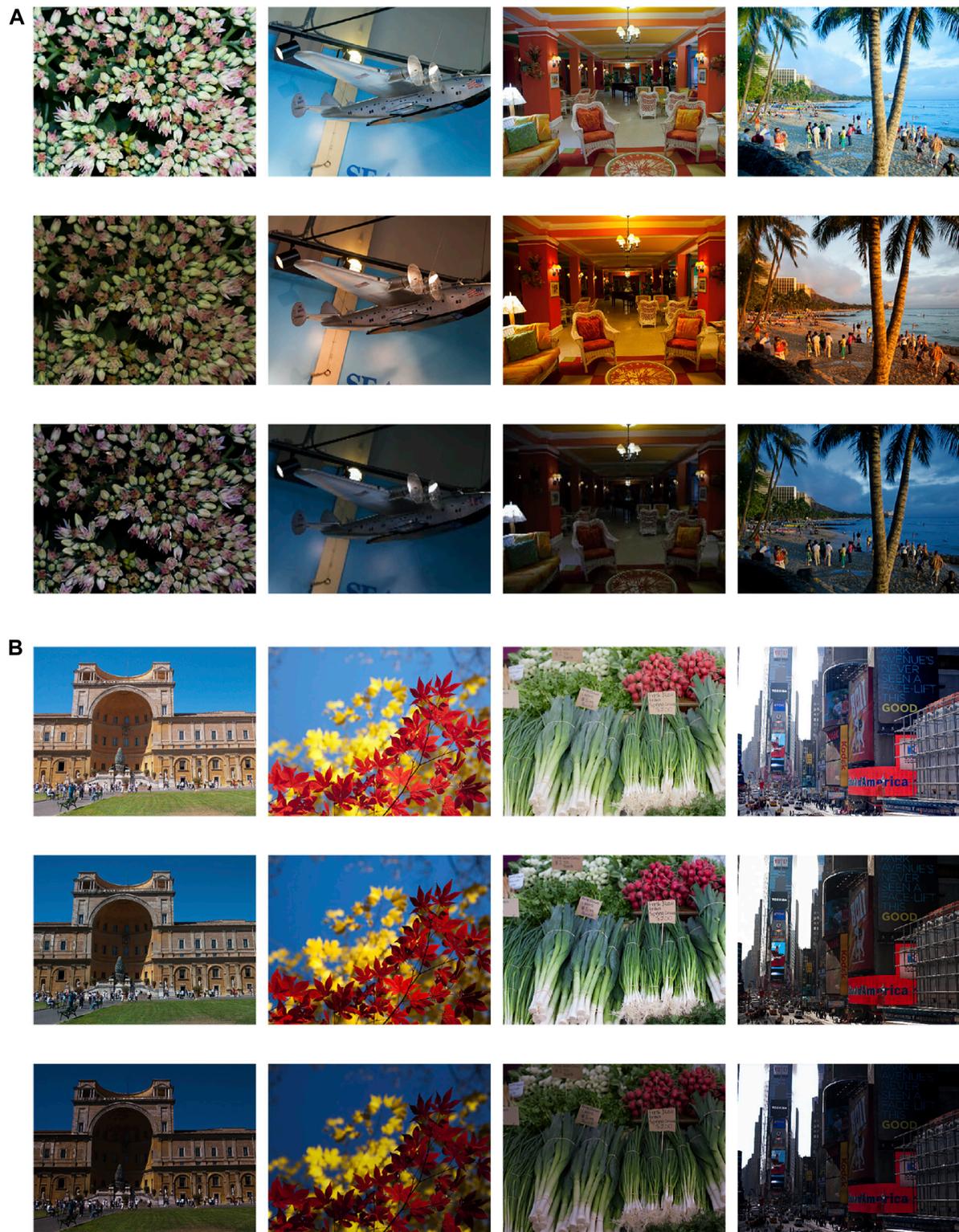
FIGURE 4
The comparison of LL images from the existing dataset and pseudo-LL ones synthesized by our method. **(A)** The comparison of white balance of LL-images; **(B)** The comparison of illumination level. The 1st, 2nd, and 3rd in both **(A)** and **(B)** show original normal-light images, LL images from the existing dataset, and pseudo-LL ones generated by our method, respectively. The images in the 3rd row of each panel are in line with the distribution of the real world in terms of illumination level and white balance.
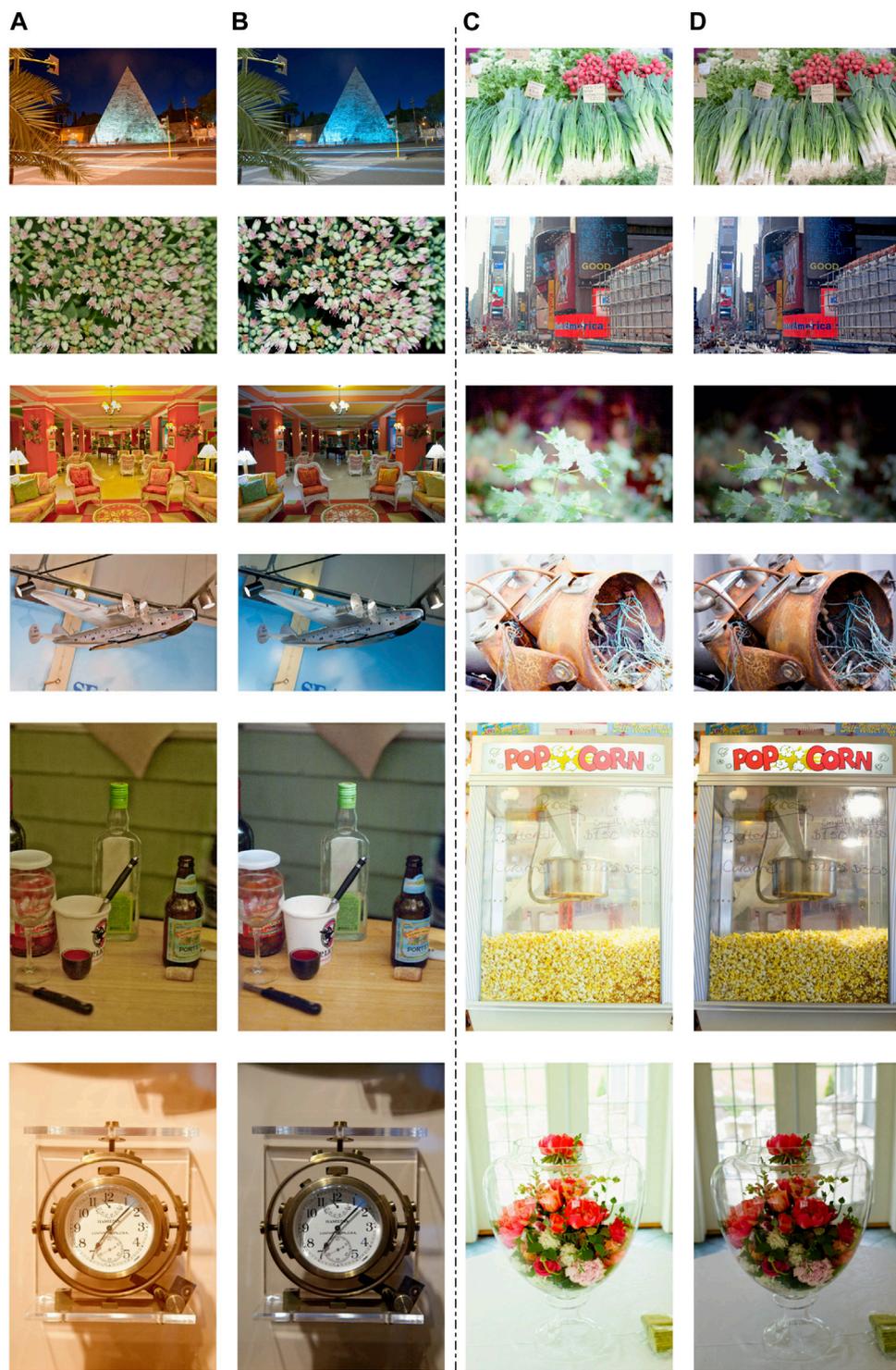
FIGURE 5
Qualitative comparison of the enhancement results of Darkening Network trained with GT LL-NL pairs $\{x^r, y^r\}$ and pseudo-LL-NL pairs $\{x^g, y^r\}$. **(A,C)** shows the enhanced results by GT LL-NL pairs $\{x^r, y^r\}$, **(B,D)** shows the enhanced results by pseudo-LL-NL pairs $\{x^g, y^r\}$. The proposed Darkening Network performs better by retaining semantically dark regions and correct white balance.

TABLE 2 Ablation studies of Darkening Network and Reverse Domain-Distance Guided strategy with the settings of both pseudo-LL-NL pairs and GT LL-NL pairs from the existing dataset.

	$\{x^r, y^r\}$		$\{x^g, y^r\}$	
Reverse domain-distance guided strategy	√	×	√	×
PSNR↑	18.796	18.751	**37.257**	35.276
SSIM↑	0.8618	0.8674	**0.9731**	0.9696
LPIPS↓	0.1764	0.1674	0.1587	**0.1547**

The scores marked in bold indicate the highest scores on the corresponding metric.
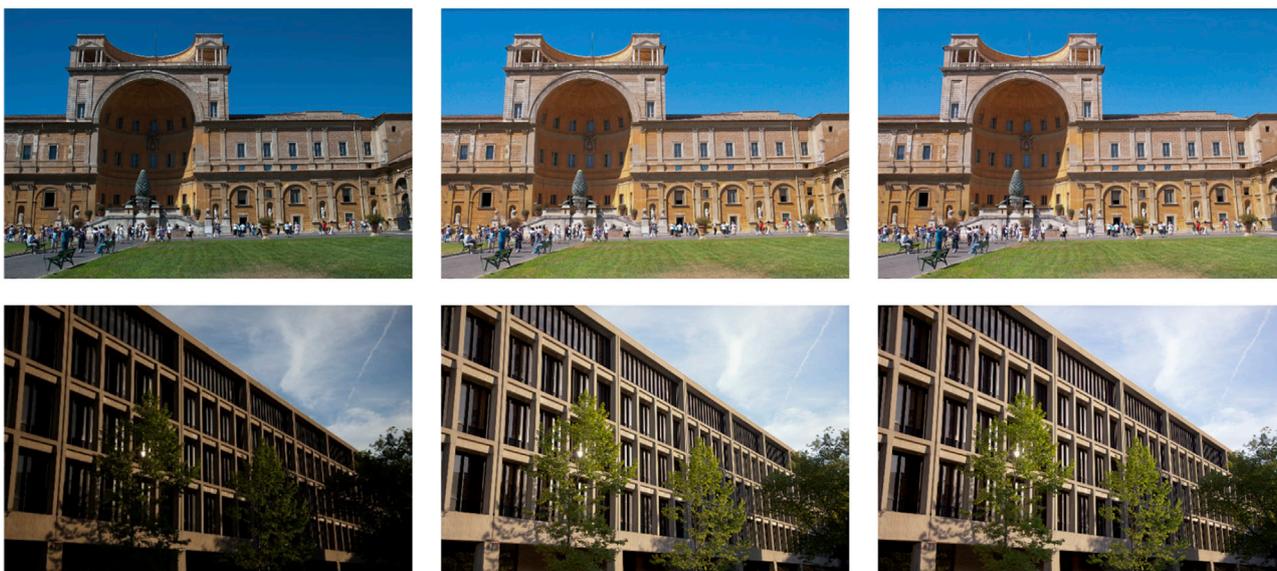
FIGURE 6
Qualitative comparison of the enhancement results without and with Reverse Domain-Distance Guided strategy. The texture and details are better reconstructed with the propsed strategy.

4.2.3 Effect of reverse domain-distance guided strategy

To verify the effectiveness of the Reverse Domain-Distance Guided strategy, we conduct ablation studies with the settings of training with $\{x^g, y^r\}$, and $\{x^r, y^r\}$. Table 2 and Figure 6 show quantitative and qualitative results, respectively. For convenient comparison, the quantitative outcomes are displayed in Table 2. We can easily discover that adopting the Reverse Domain-Distance Guided strategy improves the PSNR and SSIM with a certain magnitude of 1.981 dB (=37.257–35.276 dB) on PSNR and 0.0035 dB (= 0.9731–0.9696 dB) on SSIM when training with $\{x^g, y^r\}$. The reason is that domain distances between pseudo-LL images x^g and real-world LL images x^r are taken full advantage of at the enhancement stage. Specifically, the Enlightening Network is driven to emphasize the regions that are not in compliance with the real world by allocating greater weight to them during the training process. Therefore, it is easy to understand that the collocation of the Reverse Domain-Distance Guided strategy and the proposed Darkening network is beneficial to the reconstruction of texture and details with the pseudo-LL-NL pairs $\{x^g, y^r\}$.

Let us investigate the effect of the Reverse Domain-Distance Guided strategy. We can find from Figure 6 that those semantically dark regions maintain their semantic darkness during the improvement of illumination level without under-exposure for other regions, which facilitates the images to appear more realistic. Therefore, the proposed Reverse Domain-Distance Guided strategy is of significance for the generalization of LLIE to the real world.

4.3 Evaluations of generalization on the real-world dataset

The Exclusively Dark dataset [41] is proposed to facilitate better detection under poor visibility conditions for nighttime systems and applications. It contains a total of 7,363 images of 12 specified object categories. Some images were sub-sampled from existing large-scale datasets, including Microsoft COCO, PASCAL VOC, and ImageNet. We carry out evaluations for the generalization capacity on the Exclusively Dark dataset and DARK FACE dataset. Figure 7 shows a

FIGURE 7
Evaluation of generalization capability on real-world datasets, i.e., **(A)** Exclusive Dark dataset and **(B)** DARK FACE dataset. In both **(A,B)** panels, the 1st row indicates input LL images, and the 2nd row shows enhanced results by the proposed framework. Our method has superior generalization capability to extremely low-illumination real-world conditions.
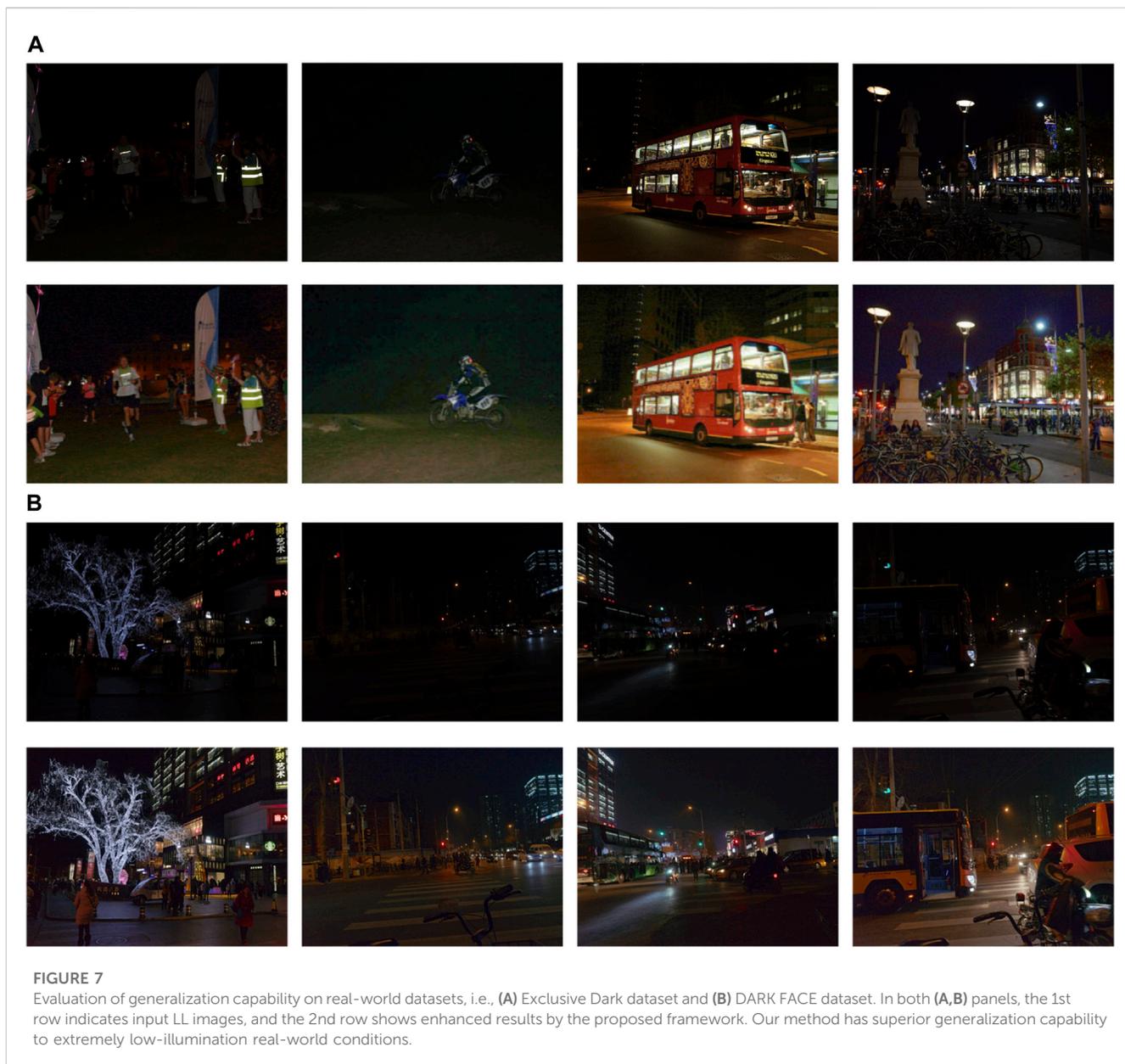
visual representation of enhancement results. Numerous outcomes demonstrate that the proposed approach can greatly boost the visibility of objects under extremely poor or significant variational illumination conditions. Therefore, we can confirm that the proposed approach can generalize well to extremely dark light conditions in the real world. Furthermore, our method can potentially facilitate subsequent computer vision systems for night vision surveillance since the performance of object-focused works usually drops when the given images are degraded [42, 43].

4.4 Comparative experiments with state-of-the-arts

Let us conduct comparative experiments with recent competing LLIE approaches on the DARK FACE dataset [39]. Figure 8 displays

the qualitative contrastive results of different competing approaches, including EnlightenGAN [22], DSLR [44], DRBN [45], TBEFN [46], RRDNet [47], and MBLLEN [48]. Qualitative results show that all the methods can effectively enhance the LL images captured under severely real-world low-light nighttime environments in terms of illumination level. However, some methods introduce side effects. Specifically, it can be clearly found that the overall hue of the image is distorted in EnlightenGAN. Besides, DRBN, TBEFN, and MBLLEN introduce distinct artifacts to local areas. It can be concluded that DSLR, RRDNet, and our method attain the top three best performances. Let us further investigate their differences in detail. It can be clearly found that DSLR and RRDNet tend to generate blur artifacts, i.e., structural details cannot be clearly reconstructed. Besides, RRDNet cannot sufficiently improve the illumination level. In contrast, our method improves visibility without introducing blurriness and shows a better reconstruction
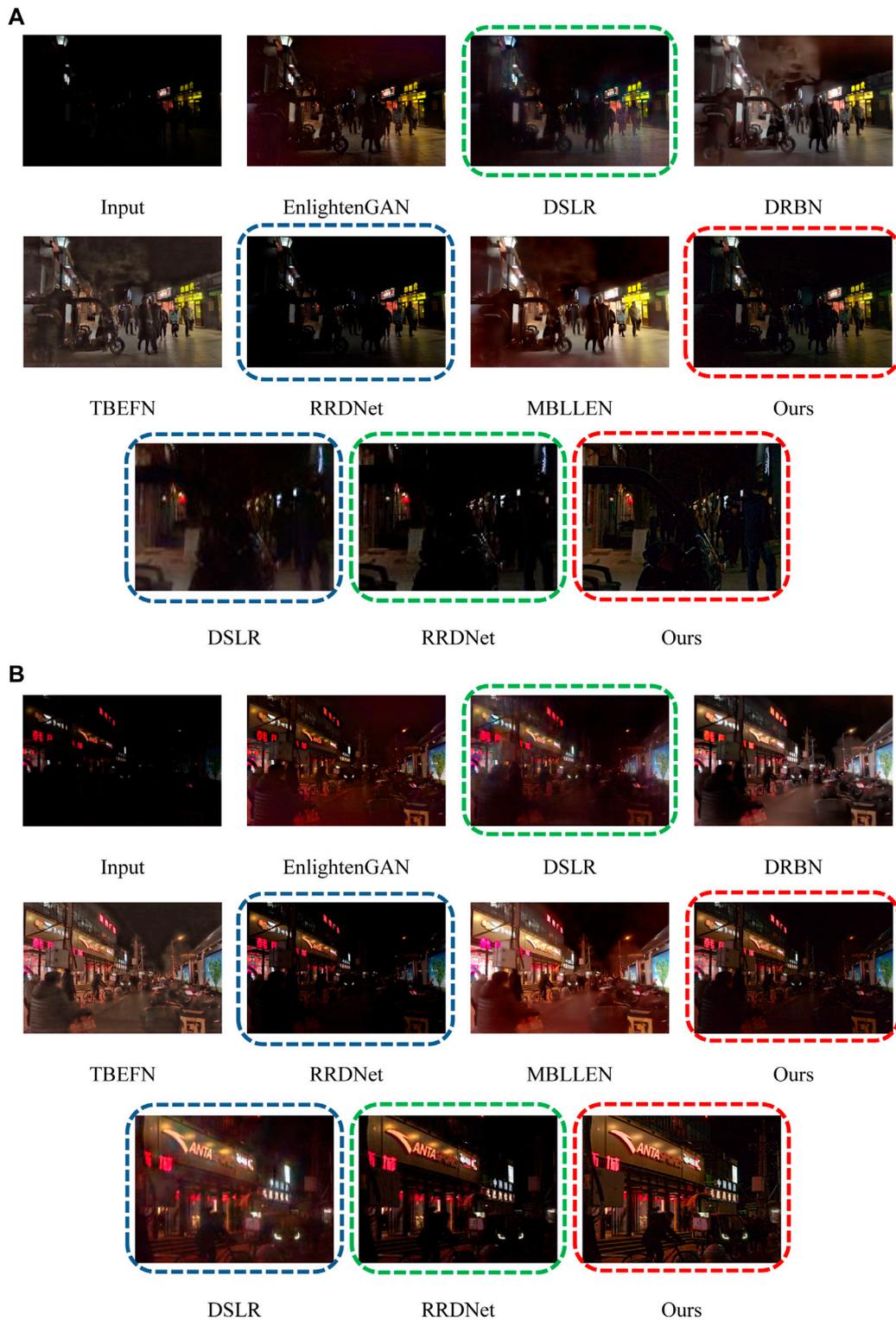
FIGURE 8
Vivid qualitative enhancement outcomes of recently competing network structures and our framework on the DARK FACE dataset. We present the results of SOTA methods on two specified images. We further compare the three most competing methods marked with boxes by zooming in on the local area of their enhanced results. Our method achieves the more superior enhancement results for real-world LL images than SOTA methods in respect of structural details and visibility.

of details, as shown in the zoomed-in comparison results. Therefore, we can confirm that our approach works most effectively relative to other recently competing LLIE methodologies.

Finally, we give a conclusion in Section 5.

5 Conclusion

This paper introduces domain adaptation to the LLIE field. Unlike previous methods that directly adopt existing synthetic low-light datasets, we propose the Domain-Gap Aware Framework, which addresses the dilemma of domain-gap lying between pseudo-LL and real-world LL domain. To eliminate the domain gap, we employ adversarial training to the Darkening Network in the first stage and obtain domain distance maps. In the second stage, we put forward a Reverse Domain-Distance Guided (RDDG) strategy, which further drives the enhancement network to focus on the regions that are not consistent with real-world distribution. In the second stage, we put forward a Reverse Domain-Distance Guided (RDDG) strategy, which further guides the Enlightening network to be attentive to the regions that are not consistent with real-world distribution. We objectively validate the effect of our framework on real-world LL datasets and conduct comparative experiments with other methods. Prominent experimental outcomes present that our framework outperforms other competing network structures.

In our future endeavors, we will explore more contributory approaches for the LLIE field. In addition, we will introduce LLIE methods to subsequent computer vision tasks and systems for diverse applications, such as driving assistant systems and nighttime surveillance.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://flyywh.github.io/CVPRW2019LowLight/, https://github.com/weichen582/GLADNet.

Author contributions

YC and MH contributed to putting forward core conceptions and design of the study; YC provided the computing platform; HL, KS, and JZ organized the database; YC and MH performed the statistical analysis and wrote the first draft of the manuscript; YC contributed to manuscript revision, read, and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Afifi M, Derpanis KG, Bjorn O, Michael SB. Learning multi-scale photo exposure correction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 20 2021 to June 25 2021; Nashville, TN, United States. IEEE (2021). p. 9157–67.

2. Gao X, Lu W, Zha L, Hui Z, Qi T, Jiang J. Quality elevation technique for UHD video and its VLSI solution. *J Chongqing Univ Posts Telecomm: Nat Sci Ed* (2020) 32(5): 681–97. doi:10.3979/j.issn.1673-825X.2020.05.001

3. Setyaningsih E, Wardoyo R, Sari AK. Securing color image transmission using compression-encryption model with dynamic key generator and efficient symmetric key distribution. *Digital Commun Networks* (2020) 6(4):486–503. doi:10.1016/j.dcan.2020.02.001

4. Xu X, Liu H, Li Y, Zhou Y. Image deblurring with blur kernel estimation in RGB channels. *J Chongqing Univ Posts Telecomm: Nat Sci Ed* (2018) 30:216–21. doi:10.3979/j.issn.1673-825X.2018.02.009

5. Zhu Z, Wei H, Hu G, Li Y, Qi G, Mazur N. A novel fast single image dehazing algorithm based on artificial multiexposure image fusion. *IEEE Trans Instrum Meas* (2020) 70:1–23. doi:10.1109/TIM.2020.3024335

6. Kyle M, Zhai X, Yu W. Image analysis and machine learning-based malaria assessment system. *Digital Commun Networks* (2022) 8(2):132–42. doi:10.1016/j.dcan.2021.07.011

7. Zhang G, Jian W, Yi Y. Traffic sign recognition based on ensemble convolutional neural network. *J Chongqing Univ Posts Telecomm: Nat Sci Ed* (2019) 31(4):571–7. doi:10.3979/j.issn.1673-825X.2019.04.019

8. Zheng M, Qi G, Zhu Z, Li Y, Wei H, Liu Y. Image dehazing by an artificial image fusion method based on adaptive structure decomposition. *IEEE Sens J* (2020) 20(14): 8062–72. doi:10.1109/JSEN.2020.2981719

9. Liu Y, Wang L, Cheng J, Li C, Chen X. Multi-focus image fusion: A survey of the state of the art. *Inf Fusion* (2020) 64:71–91. doi:10.1016/j.inffus.2020.06.013

10. Zhu Z, He X, Qi G, Li Y, Cong B, Liu Y. Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI. *Inf Fusion* (2023) 91:376–87. doi:10.1016/j.inffus.2022.10.022

11. Yang S, Zhou D, Cao J, LightingNet GY. LightingNet: An integrated learning method for low-light image enhancement. *IEEE Trans Comput Imaging* (2023) 9:29–42. doi:10.1109/TCI.2023.3240087

12. Fan S, Liang W, Ding D, Yu H. Lacn: A lightweight attention-guided ConvNeXt network for low-light image enhancement. *Eng Appl Artif Intel* (2023) 117:105632. doi:10.1016/j.engappai.2022.105632

13. Cotogni M, Cusano C. TreEnhance: A tree search method for low-light image enhancement. *Pattern Recognit* (2023) 136:109249. doi:10.1016/j.patcog.2022.109249

14. Cai J, Gu S, Zhang J. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Trans Image Process* (2018) 27(4):2049–62. doi:10.1109/TIP.2018.2794218

15. Chen C, Chen Q, Xu J, Koltun V. Learning to see in the dark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR); June 18 2018 to June 22 2018; Salt Lake City, UT, United States. IEEE (2018). p. 3291–300.

16. Li C, Guo C, Han L, Jiang J, Cheng M-M, Gu J, et al. Low-light image and video enhancement using deep learning: A survey. *IEEE Trans Pattern Anal Mach Intell* (2021) 44(12):9396–416. doi:10.1109/TPAMI.2021.3126387

17. Wang L-W, Liu Z-S, Siu W-C, Lun DP. Lightening network for low-light image enhancement. *IEEE Trans Image Process* (2020) 29:7984–96. doi:10.1109/TIP.2020.3008396

18. Lv F, Li Y, Lu F. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *Int J Comput Vis* (2021) 129(7):2175–93. doi:10.1007/s11263-021-01466-8

19. Shi Y, Wu X, Zhu M. *Low-light image enhancement algorithm based on retinex and generative adversarial network* (2019). *arXiv preprint arXiv:1906.06027.*

20. Zhu J, Meng L, Wu W, Choi D, Ni J. Generative adversarial network-based atmospheric scattering model for image dehazing. *Digital Commun Networks* (2021) 7(2):178–86. doi:10.1016/j.dcan.2020.08.003

21. Jin S, Qi N, Zhu Q, Ouyang H. Progressive GAN-based transfer network for low-light image enhancement. In: Proceeding of the 28th International Conference on Multimedia Modeling (MMM); June 6 2022 to June 10 2022; PHU QUOC, Vietnam. IEEE (2022). p. 292–304.

22. Jiang Y, Gong X, Liu D, Cheng Y, Fang C, Shen X, et al. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans Image Process* (2021) 30:2340–9. doi:10.1109/tip.2021.3051462

23. Wang M, Deng W. Deep visual domain adaptation: A survey. *Neurocomputing* (2018) 312:135–53. doi:10.1016/j.neucom.2018.05.083

24. Leng J, Liu Y, Wang Z, Hu H, CrossNet GX. CrossNet: Detecting objects as crosses. *IEEE Trans Multimedia* (2021) 24:861–75. doi:10.1109/TMM.2021.3060278

25. Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang M-H, et al. Learning enriched features for real image restoration and enhancement. In: Proceeding of the European Conference on Computer Vision (ECCV); February 7 2020 to February 20 2020; Hilton Midtown, New York, United States (2020). p. 492–511.

26. Xu K, Yang X, Yin B, Lau RW. Learning to restore low-light images via decomposition-and-enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR); June 13 2020 to June 19 2020; Seattle, United States. IEEE (2020). p. 2281–90.

27. Wei C, Wang W, Yang W, Liu J. *Deep retinex decomposition for low-light enhancement* (2018). *arXiv preprint arXiv:1808.04560.*

28. Wang Y, Liu D, Jeon H, Chu Z, Matson ET. End-to-end learning approach for autonomous driving: A convolutional neural network model. In: Proceedings of the 15th International Conference on Agents and Artificial Intelligence; February 22 2023 to February 24 2023; Lisbon, Portugal. IEEE (2019). p. 833–9.

29. Liu D, Cui Y, Cao Z, Chen Y. A large-scale simulation dataset: Boost the detection accuracy for special weather conditions. In: Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN); July 19 2020 to July 24 2020; Glasgow, United Kingdom. IEEE (2020). p. 1–8.

30. Wang Q, Fang Y, Ravula A, Feng F, Quan X, WebFormer LD. The web-page transformer for structure information extraction. In: Proceedings of the ACM Web Conference 2022; April 25 2022 to April 29 2022; Virtual Event, Lyon France. IEEE (2022). p. 3124–33.

31. Guo C, Li C, Guo J, Loy CC, Hou J, Kwong S, et al. Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 13 2020 to June 19 2020; Seattle, United States. IEEE (2020). p. 1780–9.

32. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* (2010) 22(10):1345–59. doi:10.1109/TKDE.2009.191

33. Wei Y, Gu S, Li Y, Timofte R, Jin L, Song H. Unsupervised real-world image super resolution via domain-distance aware training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 20 2021 to June 25 2021; Nashville, TN, United States. IEEE (2021). p. 13385–94.

34. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* (2004) 13(4):600–12. doi:10.1109/TIP.2003.819861

35. Simonyan K, Zisserman A. *Very deep convolutional networks for large-scale image recognition* (2014). *arXiv preprint arXiv:1409.1556.*

36. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 18 2018 to Jun 22 2018; Salt Lake City, UT, United States. IEEE (2018). p. 586–95.

37. Bychkovsky V, Paris S, Chan E, Durand F. Learning photographic global tonal adjustment with a database of input/output image pairs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); June 21 2011 to June 23 2011; Seattle, United States. IEEE (2011). p. 97–104.

38. Wang W, Wei C, Yang W, Liu J. Gladnet: Low-light enhancement network with global awareness. In: Proceedings of the 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018); May 15 2018 to May 19 2018; Xi'an, China. IEEE (2018). p. 751–5.

39. Yuan Y, Yang W, Ren W, Liu J, Scheirer WJ, Wang Z. *UG²⁺ track 2: A collective benchmark effort for evaluating and advancing image understanding in poor visibility environments* (2019). *arXiv preprint arXiv:1904.04474.* doi:10.1109/TIP.2020.2981922

40. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM* (2017) 60(6):84–90. doi:10.1145/3065386

41. Loh YP, Chan CS. Getting to know low-light images with the exclusively dark dataset. *Comput Vis Image Underst* (2019) 178:30–42. doi:10.1016/j.cviu.2018.10.010

42. Leng J, Wang Y. RCNet: Recurrent collaboration network guided by facial priors for face super-resolution. In: Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME); July 18 2022 to July 22 2022; Taipei, Taiwan, China. IEEE (2022). p. 1–6.

43. Leng J, Mo M, Zhou Y, Gao C, Li W, Gao X, et al. Pareto refocusing for drone-view object detection. *IEEE Trans Circuits Syst Video Techn* (2022). doi:10.1109/TCSVT.2022.3210207

44. Lim S, Kim W. Dslr: Deep stacked laplacian restorer for low-light image enhancement. *IEEE Trans Multimedia* (2020) 23:4272–84. doi:10.1109/TMM.2020.3039361

45. Yang W, Wang S, Fang Y, Wang Y, Liu J. Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. *IEEE Trans Image Process* (2021) 30:3461–73. doi:10.1109/TIP.2021.3062184

46. Lu K, Zhang L. Tbefn: A two-branch exposure-fusion network for low-light image enhancement. *IEEE Trans Multimedia* (2021) 23:4093–105. doi:10.1109/TMM.2020.3037526

47. Zhu A, Zhang L, Shen Y, Ma Y, Zhao S, Zhou Y. Zero-shot restoration of underexposed images via robust retinex decomposition. In: Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME); July 6 2020 to July 10 2020; London, United Kingdom. IEEE (2020). p. 1–6.

48. Lv F, Lu F, Wu J, Lim C. Mbllen: Low-light image/video enhancement using CNNs. In: Proceedings of the British Machine Vision Conference (BMVC); November 21 2022 to November 24 2022; London, United Kingdom (2018). p. 4.