Joint Image Super-resolution and Low-light Enhancement in the Dark

Feihu Zhou[1], Kan Chang[1✉], Mingyang Ling[2], Hengxin Li[3], and Shucheng Xia[1]

[1] School of Computer and Electronic Information,
Guangxi University, Nanning, China
flyhu2@126.com,kanchang@gxu.edu.cn,xsc@st.gxu.edu.cn,
[2] School of Electrical Engineering, Guangxi University, Nanning, China
lingmy@st.gxu.edu.cn
[3] Shenzhen Xiaomi Communications Co.,Ltd, China
lihengxin@xiaomi.com

Abstract. Existing super-resolution (SR) models are typically trained on datasets captured under normal-light conditions. However, when dealing with images captured under low-light conditions, the degradation becomes more complex, and the difference in lighting conditions often leads to low-quality results when using standard SR models. Due to error accumulation, simply cascading low-light enhancement (LE) and SR algorithms may not result in satisfactory results. Therefore, in this paper, we tackle this issue by jointly considering SR and LE. We first propose a new dataset called DarkSR, which contains low-resolution (LR) RAW and sRGB images captured under the low-light conditions, along with the corresponding high-resolution (HR) sRGB images captured under normal-light conditions. Noticing the linear relationship between pixel values and scene radiance, as well as the high bit depth of RAW images, and considering the presence of ISP pipeline-related information in sRGB images, we introduce JSLNet, a dual-input network that effectively explores the complementary information from the low-light LR RAW and sRGB images. Extensive experiments demonstrate that compared to other state-of-the-art (SOTA) methods, our method achieves the best quality of results, while maintaining a relatively low computational burden. The code and dataset are available at https://github.com/flyhu2/DarkSR

1 Introduction

Digital zoom can be achieved by using super-resolution (SR) methods, allowing long-distance objects in images taken with short focal length to be clearly seen. Nevertheless, in low-light conditions, long-distance objects captured by a short focal length camera suffer from more complex degradations, such as blurring, resolution reduction, low contrast, loss of detail, noise contamination, and color shifts. As most existing SR methods are trained on datasets under normal-light conditions, they cannot be directly applied in low-light conditions. Therefore,

(a) LR Bicubic x2 (b) L-Net→SR-F (c) P-LIE→SR-F (d) JSLNet(ours) (e) GT HR sRGB

Fig. 1: The visual comparison among different approaches. L-Net, P-LIE, SR-F are short for LEDNet [58], PairLLIE [15], and SRFormer [59], respectively.

simultaneously enhancing image illumination, suppressing noise, and improving spatial resolution becomes a challenging task.

To tackle this issue, a straightforward approach is to directly cascade a low-light enhancement (LE) model [27,55,57] with an SR model [12,23,59]. However, this may lead to error accumulation, as the artifacts produced by the LE model will be further propagated by the SR model. As an example, Fig. 1 (b) and (c) show the results of applying the SR model SRFormer [59] after the LE models LEDNet [58] and PairLLIE [15], respectively. We can see both LE-followed-by-SR approaches yield noticeable color distortions, and few details can be restored.

Recently, some researchers have jointly considered LE and other restoration tasks, such as deblurring [58], defogging [19] and deraining [25], so as to mitigate the issue of error accumulation. However, few studies have focused on the joint LE and SR problem. In this paper, we aim to address this issue by constructing a dataset containing low-light low-resolution (LR) images and their corresponding normal-light high-resolution (HR) ground-truth (GT), and proposing a joint SR and LE model that directly learns a sophisticated nonlinear mapping from the low-light LR space to the normal-light HR space in an end-to-end manner.

Note that although most LE and SR models are trained on sRGB images, it is sub-optimal to conduct joint SR and LE on this type of images. The main reason lies in that these images have undergone a series of processing steps in the camera's image signal processing (ISP) pipeline, including operations such as color demosaicing, denoising, color space conversion, color look-up table (LUT) application, tone mapping, gamma correction, and JPEG compression [20, 39]. On one hand, the ISP pipeline introduces nonlinear transformations like tone mapping and gamma correction, which break the linear relationship between the sRGB image and the corresponding scene radiance, and make sRGB images less suitable for tasks like LE and SR. On the other hand, the bit depth of sRGB images is limited to 8 bits, and the JPEG compression process introduces quantization noise, which further degrades image quality. As a result, sRGB images exhibit noticeable loss of information.

Fortunately, along with the compressed, low bit-depth sRGB images, RAW images are also available [10,11,18,32,49]. Since RAW images typically use 12 to 14 bits, they are able to capture more details than sRGB images. Moreover, the absence of post-processing through the ISP pipeline ensures that RAW images maintain a linear relationship with scene radiance. However, it is important to note that RAW images do not inherently contain knowledge about the ISP

pipeline. The specific system design and parameters of the ISP pipeline are often proprietary and closely guarded by camera manufacturers. Therefore, the real ISP pipeline is practically a "black box" and cannot be directly used for RAW image processing [26]. Although deep networks can be employed to learn the ISP pipeline [35], these models typically require separate training for different cameras, as the ISP pipeline varys significantly across different camera.

To tackle this problem, some methods [26, 38, 50] have proposed to utilize the sRGB image processed by the ISP pipeline as the reference image, so as to guide the conversion from the linear RAW space to the nonlinear sRGB space. However, to the best of our knowledge, there is no low-light image SR dataset containing both LR RAW and sRGB images. Inspired by the data-processing pipeline of real-world SR dataset [50] and the Day-to-Night image synthesis framework [31], we develop a pipeline to produce the low-light LR RAW and sRGB images from the given normal-light HR RAW image, and build an RAW SR dataset under low-light conditions called *DarkSR*.

Moreover, we propose a novel framework called joint super-resolution and low-light enhancement network (JSLNet). JSLNet accepts both the LR RAW and sRGB images as input. It first uses an initial enhancement sub-network to conduct feature extraction, and then employs a wavelet-domain enhancement sub-network (WEnet) to enhance the extracted features. The WEnet leverages a divide-and-conquer strategy to mitigate the difficulty of directly learning a nonlinear mapping. Specifically, the features are decomposed into low-frequency (LF) and high-frequency (HF) components, which are then enhanced separately. To facilitate an effective feature refinement, the frequency fusion module (FFM) within the WEnet fuses the frequency components of the RAW image and its corresponding sRGB counterpart. As shown in Fig. 1 (d), our joint model is able to produce visually appealing images with clear details.

In summary, our contributions are fourfold: 1) We introduce a novel low-light SR dataset called *DarkSR*. This dataset comprises low-light LR RAW and sRGB images, as well as their corresponding normal-light HR sRGB images. This dataset aims to facilitate further research and development in the field of low-light image super-resolution. 2) We develop a unified framework named *JSLNet* that can simultaneously enhance low-light images and increase their spatial resolution. JSLNet takes both the LR RAW and sRGB images as input, and exploits the ISP-related information in the sRGB image and the rich information in the RAW image to achieve faithful image reconstruction. 3) To fully leverage the complementary properties between the input RAW and sRGB images, we design a frequency fusion module (FFM) that employs cross-attention to fuse the frequency components of the two input modalities. 4) Extensive experiments demonstrate the effectiveness of our JSLNet.

1.1 Image Super-resolution

In the past a few years, a large number of convolutional neural networks (CNN)-based SR methods have been proposed [14,22,24,44]. More recently, the attention

mechanism [40] has been widely used in low-level vision, and many attention-based SR methods have achieved high quality of image reconstruction. The representative methods include SwinIR [23], HAT [12] and SRFormer [59], etc. However, the performance of the above SR methods is limited by the compressed and low bit-depth sRGB input.

Some RAW image SR methods have been proposed. For example, Xing *et al*. [49] introduced a novel SR method named JDnDmSR, which jointly conducts demosaicing, denoising and SR; the residual in residual dense block (RRDB) [32] was introduced into TENet to boost the performance of joint demosaicing and SR. However, these RAW SR methods can not be well generalized to different types of cameras since RAW images do not contain ISP-pipeline-related information. To address this issue, dual-input models RawSR [50] and PRNet [26] that leverage RAW image for detail restoration and sRGB image for color correction were proposed. Inspired by these works, we also propose a dual-input model named JSLNet for joint SR and LE.

To train the SR models, the majority of SR methods have relied on simple bicubic downsampling to generate LR images, which significantly differs from real-world degradations. To obtain more complicated degradations, a few real-world SR datasets were present, such as ZoomSR [54], RealSR [8] and DRealSR [47]. These datasets are obtained by employing a digital single-lens reflex (DSLR) camera with varying focal lengths, which, however, present several drawbacks. Firstly, controlling the magnification factors through different focal lengths is challenging. Secondly, although image alignment has been conducted on image pairs, minor misalignments still exist. As a result, the SR models trained on these datasets are required to account for additional constraints [54], which largely limit their practical applications. In addition, these datasets are all collected under normal lighting conditions, which are not suitable for training low-light image SR models. To address these issues, in this paper, we propose a data-processing pipeline that synthesizes low-light LR RAW and sRGB images from given normal-light HR RAW images. This approach enables the creation of a training dataset with diverse degradations.

1.2 Low-light Enhancement

The CNN-based LE methods have achieved a great success recently. A notable trend in this domain is the application of Retinex theory [9, 15, 46, 55, 57] to decompose images into luminance and reflection maps. The luminance map controls the overall brightness, while the reflection map captures the shapes and details of objects. Similarly, other methods also use the divide-and-conquer strategy, effectively dividing the complex nonlinear mapping problem into multiple simpler sub-problems, which, in turn, leads to superior performance, such as [17, 48, 52]. Additionally, visual transformer has been successfully introduced to LE methods. For instance, Xu *et al*. [51] proposed a signal-to-noise ratio-aware transformer, and Cui *et al*. [13] proposed a lightweight transformer that optimises the ISP parameters. To mitigate the reliance on large training datasets, zero-reference methods have also been presented, such as [16] and [42].

The well-known dataset for training LE models is the LOL dataset [46], which is collected by changing the settings of ISO and exposure time in digital camera to capture low-light and normal-light sRGB image pairs. Similarly, Chen *et al*. [11] collected a RAW image LE dataset named SID by controlling the exposure time in digital cameras. However, it is important to note that RAW images do not encompass information related to the in-camera ISP pipeline. Due to varying camera styles across different brands and models, the nonlinear mapping from the RAW image space to the sRGB image space is essentially one-to-many mapping. Therefore, without prior knowledge of the specific camera ISP pipeline, the model trained on the SID dataset may not generalize well to other camera brands and types. To address this issue, we propose extracting features from both the RAW and sRGB inputs, thereby leveraging the ISP-pipeline-related information embedded in sRGB images to guide the conversion from the RAW color space to the sRGB color space of a specific camera type.

1.3 Joint Image Super-resolution and Low-light Enhancement

There have been only a few studies exploring joint SR and LE problem. Aakerberg *et al*. [4] introduced the first low-light image SR dataset, RELLISUR, which contains various types of low-light degradation. Similar to other real-world SR datasets [8, 47, 54], the RELLISUR dataset was collected by using a DSLR with various focal lengths, leading to a common issue of misalignment within the image pairs. Furthermore, the dataset only consists of sRGB images, which inherently lack sufficient bit depth to capture rich details and exhibit a non-linear relationship with scene radiance. Rasheed *et al*. [34] directly applied bicubic downsampling on the low-light dataset LOL [46] to obtain LR low-light images. On one hand, the simplicity of bicubic downsampling fails to represent the diverse degradations present in real-world scenarios. On the other hand, similar to the RELLISUR dataset, the LOL dataset only comprises sRGB images. Consequently, training a joint SR and LE model on the above two datasets may not yield optimal results.

To address these issues, we propose a new low-light image SR dataset called *DarkSR*. It contains both LR RAW and sRGB images with diverse types of degradations. Based on *DarkSR*, we also present a novel model called *JSLNet* to directly learn a sophisticated nonlinear mapping from the low-light LR RAW space to the normal-light HR sRGB space of a specific camera type.

2 DarkSR Dataset

We assume the well-exposed images are sharp and noise-free, which is usually valid as well-lit scenes can be captured with a short exposure and low ISO. Therefore, we begin with the well-exposed HR images $\boldsymbol{X}_{\text{ori}} \in \mathbb{R}^{4H \times 4W \times 1}$, and propose a pipeline shown in Fig. 2 to synthesize the LR RAW and sRGB images in the dark. Our pipeline mainly contains the procedures of lighting degradation, resolution degradation and ISP simulation, the details are described below.

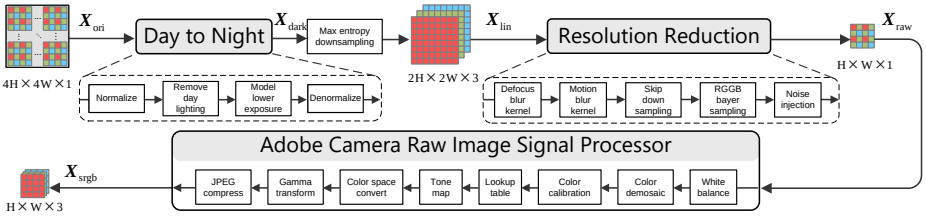

Fig. 2: The pipeline to generate *DarkSR* dataset. The input is a well-exposed HR RAW image $\boldsymbol{X}_{\text{ori}}$. Conducting lighting degradation and max entropy downsampling on $\boldsymbol{X}_{\text{ori}}$ leads to a low-light full color image $\boldsymbol{X}_{\text{lin}}$. Performing resolution degradation on $\boldsymbol{X}_{\text{lin}}$ results in a low-light LR RAW image $\boldsymbol{X}_{\text{raw}}$, and further rendering $\boldsymbol{X}_{\text{raw}}$ by a simulated ISP pipeline leads to a low-light LR sRGB image $\boldsymbol{X}_{\text{srgb}}$.

Lighting Degradation. As pixel values in RAW images are linear to scene radiance, it is reasonable to perform lighting degradation on well-exposed RAW images to synthesize low-light RAW images $\boldsymbol{X}_{\text{dark}} \in \mathbb{R}^{4H\times 4W\times 1}$. Therefore, we directly adopt the processing framework of Day-to-Night [31] to synthesize low-light RAW images. For the well-exposed HR RAW input, we first adjust their black and white levels, followed by normalization of the data. Afterwards, the daytime lighting is removed by white balance. Thereafter, exposure is reduced by multiplying the image by a randomly selected global scaling factor within the range of -1.8 to -2.2. For a detailed exploration of this technique, we refer the interested reader to [31]. However, the process of injecting noise is not included at this stage, as the noise characteristics might be compromised due to the subsequent steps of spatial resolution reduction.

Resolution Degradation. Performing resolution degradation on full-color HR images is a straightforward strategy. However, due to the usage of color filter array (CFA), there are two missing channels at each pixel of the recorded RAW image. Therefore, the full-color images are actually obtained by demosaicing the RAW images. This process inevitably introduces artifacts, and consequently, these images are not suitable for serving as GT images [21, 50]. Inspired by [21, 50], we define a 2×2 Bayer block on $\boldsymbol{X}_{\text{dark}}$ as a single virtual pixel, which contains the sampled values of red, green and blue. To compensate for the color bias caused by the center shift in each virtual pixels, we employ the maximum entropy downsampling method [21] to obtain a full-color linear image $\boldsymbol{X}_{\text{lin}} \in \mathbb{R}^{2H\times 2W\times 3}$. In order to simulate the blurring caused by capturing pictures at nighttime, both defocus blur and motion blur [50] are added to $\boldsymbol{X}_{\text{lin}}$. After that, we downsample the blurred image by a factor of 2. Finally, Bayer sampling and noise injection are applied to obtain the noisy low-light LR RAW image $\boldsymbol{X}_{\text{raw}} \in \mathbb{R}^{H\times W\times 1}$. In summary, the resolution degradation process is represented by:

$$\boldsymbol{X}_{\text{raw}} = f_{\text{Bayer}}(f_{\text{down}}(\boldsymbol{X}_{\text{lin}} * k_{\text{def}} * k_{\text{mot}})) + \mathbf{n} \tag{1}$$

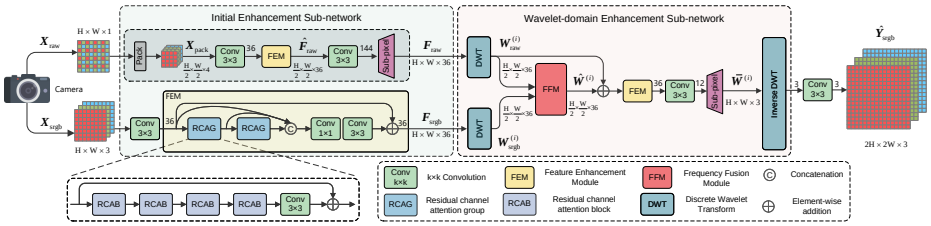

Fig. 3: Overview of JSLNet. It consists of two sub-networks. The first sub-network extracts features from both the LR RAW and sRGB images. In the second sub-network, the features are first decomposed into four frequency components by using DWT. Subsequently, FFM is employed to fuse the corresponding frequency components from the RAW and sRGB features. This fusion process aims to leverage the complementary information in both domains. The numbers on lines indicate the channel dimension.

where $f_{\text{Bayer}}(\cdot)$ is the RGGB Bayer sampling operation; $f_{\text{down}}(\cdot)$ stands for the downsampling operation; k_{def} and k_{mot} are the defocus and motion blur kernels, respectively; $*$ denotes the convolution operation; $\mathbf{n}$ is the heteroscedastic Gaussian noise [31], which resembles the noise in a low-quality short-exposure high-ISO nighttime image.

ISP Simulation. To generate low-light LR sRGB image $\boldsymbol{X}_{\text{srgb}} \in \mathbb{R}^{H\times W\times 3}$, we adopt the Adobe Camera Raw (ACR) ISP pipeline [1–3,5]. This pipeline includes various essential steps such as demosaicing, denoising, color space conversion, LUT application, tone mapping, gamma transformation, and JPEG compression to process the low-light LR raw image $\boldsymbol{X}_{\text{raw}}$. Note that we do not choose the commonly used DCRAW and Rawpy to simulate the ISP pipeline because the LUT and tone mapping employed by camera manufacturers are proprietary and not accessible through these tools. In contrast, the ACR ISP pipeline adequately simulates the comprehensive ISP steps implemented within real-world cameras. In addition to generating the low-light LR sRGB image, we also produce the normal-light HR sRGB image $\boldsymbol{Y}_{\text{srgb}} \in \mathbb{R}^{2H\times 2W\times 3}$ (i.e., the GT image) by sequentially applying maximum entropy downsampling, and ACR ISP pipeline to the well-exposed HR raw image $\boldsymbol{X}_{\text{ori}}$.

3 JSLNet for Joint SR and LE in the Dark

3.1 Network Architecture

As shown in Fig. 3, we propose a joint super-resolution and low-light enhancement network (JSNet), which takes both LR low-light RAW and sRGB images as inputs. This model consists of two sub-networks. The initial sub-network is responsible for extracting features from both two inputs separately, while the wavelet-domain enhancement sub-network (WEnet) uses discrete wavelet transform (DWT) to decompose the features and enhances the decomposed components in the frequency domain.

Specifically, in the initial enhancement sub-network, similar to [10,11,26,49], the LR low-light RAW image $\boldsymbol{X}_{\text{raw}} \in \mathbb{R}^{H\times W\times 1}$ is rearranged to RGGB matrices $\boldsymbol{X}_{\text{pack}} \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times 4}$ according to the Bayer pattern. To effectively extract features, we build feature enhancement module (FEM), which consists of two residual channel attention groups (RCAGs) [56]. For the LR low-light sRGB image $\boldsymbol{X}_{\text{srgb}} \in \mathbb{R}^{H\times W\times 3}$, we obtain the features $\boldsymbol{F}_{\text{srgb}} \in \mathbb{R}^{H\times W\times C}$. On the other hand, for the packed RAW input, we further use a 3×3 convolution layer and a sub-pixel layer [37] to increase its spatial resolution, leading to features $\boldsymbol{F}_{\text{raw}} \in \mathbb{R}^{H\times W\times C}$. In this paper, we set $C = 36$.

Inspired by the widely used divide-and-conquer strategy [9, 15, 26, 46, 52, 55, 57], we apply discrete wavelet transform (DWT) in WEnet to decompose the extracted features, i.e., $\boldsymbol{F}_{\text{raw}}$ and $\boldsymbol{F}_{\text{srgb}}$, into different frequency components to ease the difficult of directly learning a sophisticated nonlinear mapping. The DWT is employed due to the following reasons: 1) DWT preserves both frequency and location information in multiple directions, which is beneficial to restoring fine details and sharp edges; 2) the LF component primarily captures global information such as illumination and shape of objects, while the HF components mainly represent details and edges, making DWT well-suited for our network; 3) the spatial resolution of each frequency component of DWT is reduced to half of the input feature, which is helpful for maintaining a relatively low computational burden. The decomposed frequency components of $\boldsymbol{F}_{\text{raw}}$ and $\boldsymbol{F}_{\text{srgb}}$ are denoted as $\boldsymbol{W}^{(i)}_{\text{raw}} \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times C}$ and $\boldsymbol{W}^{(i)}_{\text{srgb}} \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times C}$, respectively, where $i = 0$ represents the LF component and $i = 1, 2, 3$ represent the HF components with the directions of LH, HL, and HH, respectively.

Since co-located positions in corresponding frequency components of RAW and sRGB domains are spatially correlated, it is reasonable to fuse the corresponding frequency components, so that the complementary information present in both domains can be leveraged. To this end, the frequency fusion module (FFM) is developed, the details of which are described in Sec. 3.2. To enable residual learning, $\boldsymbol{W}^{(i)}_{\text{raw}}$ is directly added to the output of FFM. Afterwards, FEM is employed to further refine the fused features. To reconstruct the normal-light HR full-color images, a 3×3 convolution layer followed by a sub-pixel layer is used on each frequency component to obtain $\bar{\boldsymbol{W}}^{(i)} \in \mathbb{R}^{H\times W\times 3}$. Next, we perform the inverse discrete wavelet transform (IDWT) on the reconstructed LF and HF components. To further reduce artifacts, we apply an additional 3×3 convolution layer and obtain the final normal-light HR output $\hat{\boldsymbol{Y}}_{\text{srgb}} \in \mathbb{R}^{2H\times 2W\times 3}$.

3.2 Frequency Fusion Module

The structure of FFM is shown in Fig. 4, where the corresponding frequency components from both the RAW and sRGB domains are fused. Although sRGB images contain information related to the ISP pipeline, they suffer from information loss due to limited bit-depth and JPEG compression. In contrast, RAW images preserve more details, and their pixel values are linearly related to scene

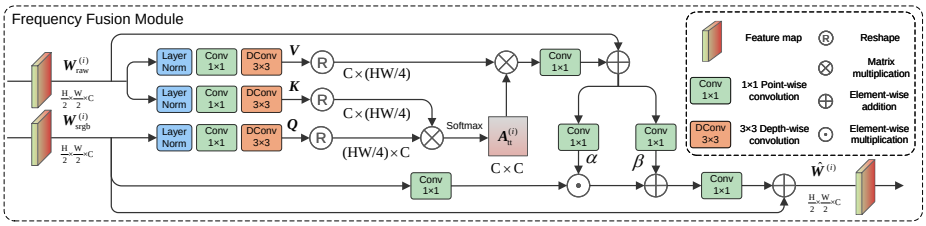

Fig. 4: The structure of frequency fusion module (FFM). Cross attention is applied to the frequency components from the RAW domain, and the resulting output is used to learn the scale and shift parameters of an affine transformation. This transformation is then used to modulate the frequency components from the sRGB domain.

radiance. Therefore, in FFM, we try to compensate for the lost details in the sRGB domain by taking advantage of RAW images.

To effectively and efficiently fuse the two type of features, affine transformation [33, 43] could be a possible solution. However, the scaling parameter $\boldsymbol{\alpha}$ and shift parameter $\boldsymbol{\beta}$ of affine transformation are usually obtained by using traditional convolution layers, which have a limited receptive field and fail to capture long-range dependencies. To address this issue, we take advantage of the cross-attention mechanism [40] to calculate the scaling and shift parameters. Since computing the attention map in the spatial dimension is computationally intensive, inspired by [53], we resort to computing cross-covariance across feature channels to obtain attention map.

Specifically, we generate query matrix $\boldsymbol{Q}$ from the layer-normalized feature $\boldsymbol{W}^{(i)}_{\text{srgb}}$, and correspondingly, we construct the key matrix $\boldsymbol{K}$ and the value matrix $\boldsymbol{V}$ from the layer-normalized feature $\boldsymbol{W}^{(i)}_{\text{raw}}$. Subsequently, these matrices are reshaped before undergoing a matrix multiplication operation between $\boldsymbol{K}$ and $\boldsymbol{Q}$. Following this, a Softmax function is applied to the output to derive the cross-attention map $\boldsymbol{A}^{(i)}_{\text{tt}} \in \mathbb{R}^{C\times C}$. By doing so, we can fully explore the global interactions between the frequency components extracted from the RAW and sRGB domain. The calculation of $\boldsymbol{A}^{(i)}_{\text{tt}}$ can be formulated as:

$$\boldsymbol{A}^{(i)}_{\text{tt}} = \text{Softmax}(\boldsymbol{K}\cdot\boldsymbol{Q}/\tau) \tag{2}$$

where τ is a learnable temperature parameter.

After applying cross attention on $\boldsymbol{V}$, information from $\boldsymbol{W}^{(i)}_{\text{srgb}}$ has been incorporated into the frequency components from the RAW domain. Therefore, the fused result is suitable for learning the scale parameter $\boldsymbol{\alpha}$ and the shift parameter $\boldsymbol{\beta}$. Subsequently, these parameters are utilized to modulate and compensate the sRGB frequency components, thereby contributing significantly to the fidelity and accuracy of the representation.

Finally, the output feature $\hat{\boldsymbol{W}}^{(i)}$ is obtained by:

$$\hat{\boldsymbol{W}}^{(i)} = \text{Conv1}(\boldsymbol{\alpha}\odot\text{Conv1}(\boldsymbol{W}^{(i)}_{\text{srgb}}) + \boldsymbol{\beta}) + \boldsymbol{W}^{(i)}_{\text{srgb}} \tag{3}$$

where $\odot$ denotes element-wise multiplication, and Conv1$(\cdot)$ denotes the function of 1×1 convolution layer.

4 Experiments

4.1 Dataset and Experimental Settings

The well-known MIT-Adobe FiveK dataset [7], which comprises 5000 RAW images, is applied to generate our dataset DarkSR. We manually remove the underexposed and over-exposed images. As a result, we retain 1548 well-exposed RAW images, of which 1395 images are used for training, while the remaining images are reserved for testing. The dataset contains 20 camera models from 4 camera brands, and the training and testing sets are synthesized by the method described in Fig. 2. The radius of the defocus blur is randomly sampled from $[3, 5]$, and the size of the motion blur kernel is chosen from three odd numbers: $\{3, 5, 7\}$. The short-exposure high-ISO noise applied in Day-to-Night model [31] is used, where the ISO range is randomly sampled from $[400, 1600]$ in our dataset. For more details of DarkSR, please refer to the supplementary materials.

When training JSLNet, we use the $\mathcal{L}_1$ loss function, which measures the pixel-by-pixel difference between the reconstructed output $\hat{\boldsymbol{Y}}_{\text{srgb}}$ of JSLNet and the GT image $\boldsymbol{Y}_{\text{srgb}}$. We randomly crop one 256×256 patch from each RAW/sRGB image, and then feed these patches into JSLNet for training. No data augmentation is used. The batch size is set as 6, and the training lasts 500 epochs. The AdamW optimizer with parameters $\beta_1 = 0.9$, $\beta_2 = 0.99$ is used, with the weight decay $\omega = 0.05$. A warm-up strategy is used, and the learning rate is first linearly increased to 1.5×10^{-4} and then decayed to 6×10^{-6} by using the cosine annealing strategy. Our JSLNet is implemented with the Pytorch framework, and all experiments are performed on a single NVIDIA GeForce RTX 3090 GPU.

4.2 Evaluation on the DarkSR Dataset

We evaluate the proposed JSLNet quantitatively and qualitatively on our DarkSR dataset. Since the joint SR and LE tasks have not been widely studied and there are no publicly available methods for comparison, we compare our method with the representative LE and SR methods in this paper. For a fair comparison, all compared methods have been re-trained on our DarkSR dataset in an end-to-end manner. Specifically, the compared methods can be divided into three categories:

1) RAW-input methods. For RAW input, the representative LE algorithms SID [11] and DNF [18] followed by the SR model SRFormer (abbreviated as SR-F) [59] are compared. In addition, the joint demosaicing and SR methods JDnDmSR [49] and TENet [32] are also evaluated.

2) sRGB-input methods. For sRGB input, the Retinex-based LE methods, including RetinexFormer (abbreviated as Re-F) [9], PairLLIE (abbreviated as P-LIE) [15], and CUE [57], and the jointly LE and deblurring approach LED-Net (abbreviated as L-Net) [58] are cascaded with SR-F [59]. Moreover, we directly train the SR methods HAT [12] and SR-F [59] by using the sRGB image pairs from the DarkSR dataset.

Table 1: Quantitative comparison on the DarkSR dataset. The best/second best results are marked in red/underlined. Due to varying image resolutions, fixed-size patches with a resolution of 256×256 are used for measuring FLOPs and runtime.

Methods	Input Type	CPSNR ↑	SSIM ↑	ΔE ↓	Para.(M)	Flops(G)	RT.(ms)
SID [11]→SR-F [59]	RAW	22.751	0.7560	11.2756	9.996	149.35	204.00
DNF [18]→SR-F [59]	RAW	23.551	0.7713	10.7789	5.069	162.25	212.55
JDnDmSR [49]	RAW	24.664	0.7896	8.8616	6.479	426.41	65.76
TENet [32]	RAW	24.641	0.7972	8.8594	20.219	827.94	98.67
L-Net [58]→SR-F [59]	sRGB	25.844	0.7743	9.1125	9.644	184.49	213.73
P-LIE [15]→SR-F [59]	sRGB	26.091	0.7890	8.3904	2.577	168.27	204.56
CUE [57]→SR-F [59]	sRGB	26.859	0.7962	7.7538	2.492	152.23	235.95
Re-F [9]→SR-F [59]	sRGB	27.904	0.8156	7.0833	3.841	162.94	217.26
HAT [12]	sRGB	27.530	0.8110	7.1930	9.476	613.24	972.23
SRFormer [59]	sRGB	27.602	0.8109	7.2104	2.235	145.92	411.73
RawSR [50]	Dual-input	28.017	0.8171	6.8842	5.163	165.49	15.13
PRNet [26]	Dual-input	28.226	0.8205	6.8260	1.965	69.65	22.14
JSLNet(ours)	Dual-input	28.430	0.8229	6.7715	1.555	36.84	21.69

3) Dual-input methods. Two dual-input SR methods RawSR [50] and PRNet [26] are re-trained on the DarkSR dataset.

The results are shown in Tab. 1. Note that for all types of methods, a single model is used for the entire testing set. The color peak signal-to-noise ratio (CPSNR) [28], structural similarity (SSIM) [45], and the CIELAB color space difference (also known as ΔE [36]) are employed to evaluate the reconstruction accuracy. The number of parameters, Flops and running time are used to verify the efficiency of the compared methods.

As shown in the Tab. 1, the RAW-input methods exhibit the poorest performance due to the absence of ISP pipeline-related information. The sRGB-input methods yield better results compared to the RAW-input methods, but they are still inferior to the Dual-input methods. This suggests that their performance is primarily constrained by the compressed, low bit-depth sRGB input. Furthermore, the dual-input methods outperform the other two types of methods, indicating the necessity of using dual inputs. Among these methods, our JSNLet approach achieves the best results while maintaining a relatively low complexity burden. This can be attributed to the following reasons: 1) the divide-and-conquer strategy reduces the difficulty of directly learning a complex nonlinear mapping; 2) due to the usage of DWT, both localization and frequency information are well represented, leading to well preservation of fine details; 3) the DWT also reduces the spatial resolution of features, resulting in a lower computational burden; 4) through the FFM, our approach fully explores the global interactions between RAW and sRGB wavelet features, thus leveraging the complementary information from both inputs and contributing to high performance.

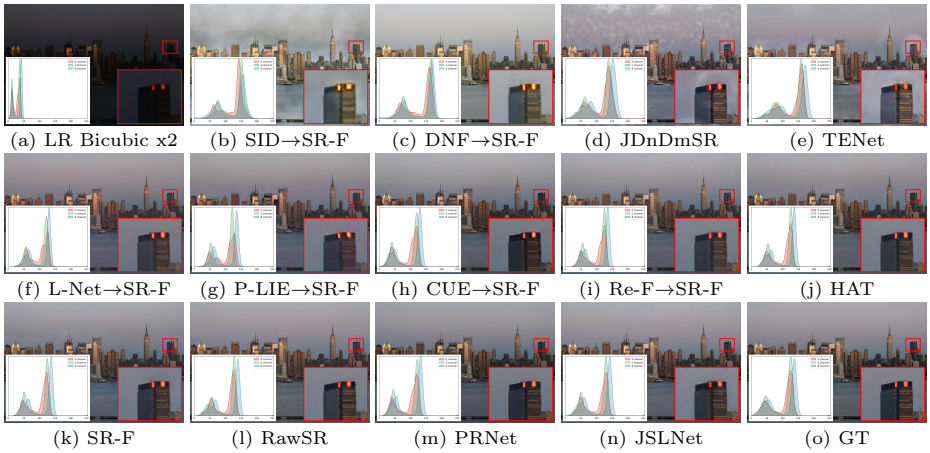

Fig. 5: Visual comparisons of different methods. Local details and color histograms are displayed in the lower-right and lower-left positions, respectively. Zoom for best view.

Fig. 5 shows the visual comparison among different methods. It can be observed that the RAW-input methods have a significant color distortion, demonstrating the difficulty of achieving an accurate mapping from the RAW domain to the sRGB domain of a specific camera type. Meanwhile, due to information loss in sRGB images, the sRGB-input methods fail to recover rich details. Benefiting from the complementary information of the RAW and sRGB inputs, the dual-input methods achieve higher quality of results. Compared to RawSR [50] and PRNet [26], our JSLNet recovers more details hidden in the dark region.

4.3 Ablation Study

To demonstrate the effectiveness of the key modules of JSLNet, we conduct ablation experiments on the DarkSR dataset, and the results are shown in Tab. 2. For a fair comparison, all variants have been adjusted to a similar model size. The variant $\mathbb{N}_1$ removes DWT/IDWT; variant $\mathbb{N}_2$ removes the sRGB input, while $\mathbb{N}_3$ discards the RAW input; in variant $\mathbb{N}_4$, the FFM is replaced by directly concatenating RAW and sRGB features. Moreover, variants $\mathbb{N}_5$ and $\mathbb{N}_6$ are evaluated to further investigate the effectiveness of the structure in FFM. In $\mathbb{N}_5$, the cross attention is removed, and the scale and shift parameters of affine transformation are learned by the concatenated frequency components from both domains; in $\mathbb{N}_6$, the affine transformation is removed, and the frequency components from both domains are directly fused by the cross attention mechanism.

As shown in Tab. 2, variant $\mathbb{N}_1$ exhibits a noticeable performance drop, highlighting the essential role of DWT in establishing the complex mapping relationship. The performance of variants $\mathbb{N}_2$ and $\mathbb{N}_3$ significantly degrades, confirming the effectiveness of the dual-input structure. In particular, variant $\mathbb{N}_2$ is notably inferior to $\mathbb{N}_3$, indicating the importance of the sRGB input for accurate

Table 2: Ablation study on the DarkSR dataset. The best results are marked in red. CAM is short for cross attention mechanism, and AF is short for affine transformation.

Metrics	$\mathbb{N}_1$ w/o DWT	$\mathbb{N}_2$ w/o sRGB	$\mathbb{N}_3$ w/o RAW	$\mathbb{N}_4$ w/o FFM	$\mathbb{N}_5$ w/o CAM	$\mathbb{N}_6$ w/o AF	JSLNet
Para.(M)	1.582	1.581	1.558	1.596	1.523	1.524	1.555
CPSNR(dB) ↑	27.710	25.446	27.666	27.934	27.895	27.929	28.430
SSIM ↑	0.8074	0.8021	0.8074	0.8100	0.8084	0.8148	0.8229
ΔE ↓	7.1667	8.7139	7.2443	6.9904	7.0408	7.1377	6.7715

Table 3: Quantitative comparison on the realistic testing dataset. The best results are marked in red, and the second best results are underlined.

Methods	Input Type	NIQE ↓	PIQE ↓	PI ↓	BRISQUE ↓
SID [11]→SR-F [59]	RAW	7.67	75.15	7.48	64.07
DNF [18]→SR-F [59]	RAW	6.68	63.40	6.86	54.28
JDnDmSR [49]	RAW	6.32	63.25	6.56	51.38
TENet [32]	RAW	6.43	61.52	6.62	57.31
L-Net [58]→SR-F [59]	sRGB	6.72	64.58	6.49	50.75
P-LIE [15]→SR-F [59]	sRGB	6.45	63.34	6.38	46.74
CUE [57]→SR-F [59]	sRGB	<u>6.04</u>	61.50	6.09	44.02
Re-F [9]→SR-F [59]	sRGB	6.48	60.71	6.45	47.75
HAT [12]	sRGB	6.43	61.71	6.36	48.16
SR-F [59]	sRGB	6.32	<u>60.21</u>	<u>6.29</u>	45.60
RawSR [50]	Dual-input	6.23	61.39	6.38	<u>42.97</u>
PRNet [26]	Dual-input	6.45	60.86	6.39	46.27
JSLNet(ours)	Dual-input	5.84	59.48	6.09	39.19

color correction. Variant $\mathbb{N}_4$ demonstrates that directly concatenating the two frequency components from different domains fails to effectively leverage the complementary information. Variants $\mathbb{N}_5$ and $\mathbb{N}_6$ demonstrate that removing either cross attention or affine transformation results in poorer performance.

4.4 Evaluation on Realistic Low-light LR Images

We collected a realistic testing dataset to assess the generalization ability of our method. The dataset consists of 18 low-light RAW and sRGB images captured using a *Xiaomi 14 Ultra* in a light box. The images have exposure biases ranging from -1.7 to -2.2 and ISO levels from 200 to 1600.

All the models trained by using the DarkSR dataset are directly tested on this realistic testing dataset, and the results are provided in Tab. 3. Since there is no GT normal-light HR image in the real-world testing dataset, four no-reference

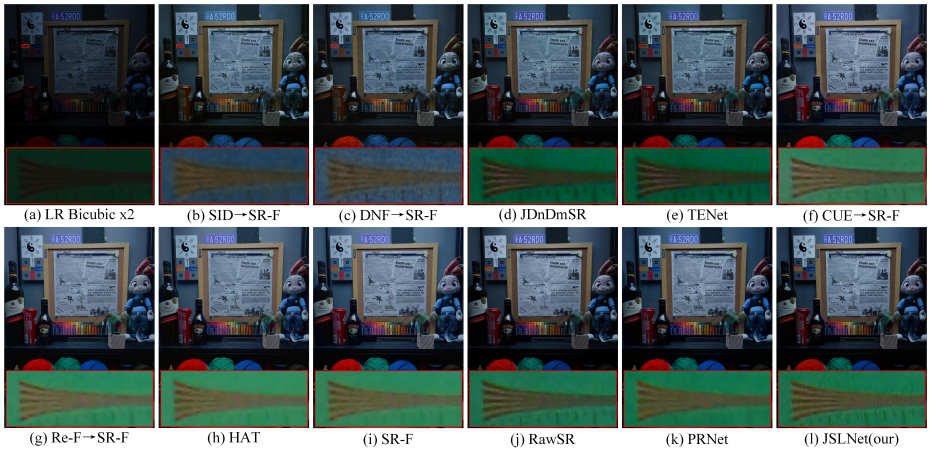

Fig. 6: Visual comparison of different methods on an unseen camera (*Xiaomi 14 Utral*).

image quality assessment metrics, including NIQE [30], PIQE [41], PI [6], and BRISQUE [29], are used to evaluate the performance of different methods. It can be found that our JSLNet still achieves the best results in all metrics.

The visual comparison among different methods are given in Fig. 6. Among the RAW-input methods, Fig. 6 (b) and (c) have obvious color shift, while (d) and (e) produce darker results. The sRGB-input methods, i.e., Fig. 6 (f)-(i), produce overly smooth content. In contrast, among the three dual-input methods, our JSLNet not only effectively restores fine details and sharp edges but also accurately reproduces vibrant colors. These results indicate that our model can generalize well to other unseen camera types.

5 Conclusion

To address the joint SR and LE problem, we introduce the DarkSR dataset, which includes low-light LR RAW and sRGB images, as well as normal-light HR sRGB images of the scenes. This dataset accounts for complex lighting and resolution degradation in RAW images and utilizes the ACR ISP pipeline to generate the corresponding sRGB images. Furthermore, we propose JSLNet, a lightweight yet effective model that takes both the low-light LR RAW and sRGB images as input. Experimental results on the DarkSR dataset show that JSLNet outperforms SOTA methods, and the results on the realistic testing dataset demonstrate the generalization ability of JSLNet. Future work will explore the potential of using realistic low-light LR images directly in training, potentially through unsupervised learning methods.

Acknowledgments. This work was supported by NSFC under Grant 62171145, Guangxi Key R&D Program under Grant AB23075106, and Guangxi Key Laboratory of Multimedia Communications and Network Technology.

References

1. Adobe. Color and Camera RAW, https://helpx.adobe.com/ca/photoshop-elements/using/color-camera-raw.html. Accessed: 2024-05-01.
2. Adobe. DNG SDK, https://helpx.adobe.com/cn/security/products/dng-sdk.html. Accessed: 2024-09-01.
3. Adobe. DNG_Spec_1_7_1_0, https://helpx.adobe.com/content/dam/help/en/photoshop/pdf/DNG_Spec_1_7_1_0.pdf. Accessed: 2024-09-01.
4. Aakerberg, A., Nasrollahi, K., Moeslund, T.B.: RELLISUR: A Real Low-Light Image Super-Resolution Dataset. In: Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). pp. 1–12. Online (2021)
5. Afifi, M., Derpanis, K.G., Ommer, B., Brown, M.S.: Learning Multi-Scale Photo Exposure Correction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9157–9167. Online (2021)
6. Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: The 2018 PIRM Challenge on Perceptual Image Super-Resolution. In: Proceedings of the European Conference on Computer Vision Workshops (ECCVW). pp. 315–333. Munich, Germany (2018)
7. 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 Conference on Computer Vision and Pattern Recognition (CVPR). pp. 97–104. Colorado Springs, USA (2011)
8. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward Real-World Single Image Super-Resolution: A New Benchmark and A New Model. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 3086–3095. Seoul, Korea (2019)
9. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: RetinexFormer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 12504–12513. Paris, France (2023)
10. Chang, K., Li, H., Tan, Y., Ding, P.L.K., Li, B.: A Two-Stage Convolutional Neural Network for Joint Demosaicking and Super-Resolution. IEEE Transactions on Circuits and Systems for Video Technology **32**(7), 4238–4254 (2022)
11. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to See in the Dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3291–3300. Salt Lake City, USA (2018)
12. Chen, X., Wang, X., Zhou, J., Qiao, Y., Dong, C.: Activating More Pixels in Image Super-Resolution Transformer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22367–22377. Vancouver, Canada (2023)
13. Cui, Z., Li, K., Gu, L., Su, S., Gao, P., Jiang, Z., Qiao, Y., Harada, T.: You Only Need 90K Parameters to Adapt Light: a Light Weight Transformer for Image Enhancement and Exposure Correction. In: Proceedings of the British Machine Vision Conference (BMVC). pp. 1–25. London, UK (2022)
14. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a Deep Convolutional Network for Image Super-Resolution. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 184–199. Zurich, Switzerland (2014)
15. Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.K.: Learning a Simple Low-Light Image Enhancer From Paired Low-Light Instances. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22252–22261. Vancouver, Canada (2023)

16. Guo, C. and Li, C. and Guo, J. and Loy, C. and Hou, J. and Kwong, S. and Cong, R.:Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1780–1789. Seattle, WA, USA (2020)
17. Han, F., Chang, K., Li, G., Ling, M., Huang, M., Gao, Z.: Illumination-aware divide-and-conquer network for improperly-exposed image enhancement. Neural Networks **180**, 106733 (2024)
18. Jin, X., Han, L.H., Li, Z., Guo, C.L., Chai, Z., Li, C.: DNF: Decouple and Feedback Network for Seeing in the Dark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 18135–18144. Vancouver, Canada (2023)
19. Jin, Y., Lin, B., Yan, W., Yuan, Y., Ye, W., Tan, R.T.: Enhancing Visibility in Nighttime Haze Images Using Guided APSF and Gradient Adaptive Convolution. In: Proceedings of the ACM International Conference on Multimedia (ACM MM). pp. 2446–2457 (2023)
20. Karaimer, H.C., Brown, M.S.: A Software Platform for Manipulating the Camera Imaging Pipeline. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 429–444. Amsterdam, Netherlands (2016)
21. Khashabi, D., Nowozin, S., Jancsary, J., Fitzgibbon, A.W.: Joint Demosaicing and Denoising via Learned Nonparametric Random Fields. IEEE Transactions on Image Processing **23**(12), 4968–4981 (2014)
22. Kim, J., Lee, J.K., Lee, K.M.: Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1646–1654. Las Vegas, USA (2016)
23. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: SwinIR: Image Restoration Using Swin Transformer. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 1833–1844. Montreal, Canada (2021)
24. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced Deep Residual Networks for Single Image Super-Resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 136–144. Honolulu, USA (2017)
25. Lin, X., Yue, J., Ding, S., Ren, C., Guo, C.L., Li, C.: Unlocking Low-Light-Rainy Image Restoration by Pairwise Degradation Feature Vector Guidance. arXiv preprint arXiv:2305.03997 (2023)
26. Ling, M., Chang, K., Huang, M., Li, H., Dang, S., Li, B.: PRNet: Pyramid Restoration Network for RAW Image Super-Resolution. IEEE Transactions on Computational Imaging **10**, 479–495 (2024)
27. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5637–5646. New Orleans, USA (2022)
28. Menon, D., Calvagno, G.: Color image demosaicking: An overview. Signal Processing: Image Communication **26**(8-9), 518–533 (2011)
29. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-Reference Image Quality Assessment in the Spatial Domain. IEEE Transactions on Image Processing **21**(12), 4695–4708 (2012)
30. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "Completely Blind" Image Quality Analyzer. IEEE Signal Processing Letters **20**(3), 209–212 (2012)
31. Punnappurath, A., Abuolaim, A., Abdelhamed, A., Levinshtein, A., Brown, M.S.: Day-to-Night Image Synthesis for Training Nighttime Neural ISPs. In: Pro-

ceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10769–10778. New Orleans, USA (2022)
32. Qian, G., Wang, Y., Gu, J., Dong, C., Heidrich, W., Ghanem, B., Ren, J.S.: Rethinking Learning-based Demosaicing, Denoising, and Super-Resolution Pipeline. In: IEEE International Conference on Computational Photography (ICCP). pp. 1–12 (2022)
33. Qin, Q., Chang, K., Huang, M., Li, G.: DENet: Detection-Driven Enhancement Network for Object Detection Under Adverse Weather Conditions. In: Proceedings of the Asian Conference on Computer Vision (ACCV). pp. 2813–2829. Macau, China (2022)
34. Rasheed, M.T., Shi, D.: LSR: Lightening super-resolution deep network for low-light image enhancement. Neurocomputing **505**, 263–275 (2022)
35. Schwartz, E., Giryes, R., Bronstein, A.M.: DeepISP: Toward Learning an End-to-End Image Processing Pipeline. IEEE Transactions on Image Processing **28**(2), 912–923 (2018)
36. Sharma, G., Bala, R.: Digital Color Imaging Handbook. CRC press (2017)
37. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1874–1883. Las Vegas, USA (2016)
38. Song, B., Zhou, J., Chen, X., Zhang, S.: Real-Scene Reflection Removal With RAW-RGB Image Pairs. IEEE Transactions on Circuits and Systems for Video Technology **33**(8), 3759–3773 (2023)
39. Tang, Y., Chang, K., Huang, M., Li, B.: BMISP: Bidirectional mapping of image signal processing pipeline. Signal Processing **212**, 109135 (2023)
40. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention Is All You Need pp. 6000–6010 (2017)
41. Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S.: Blind image quality evaluation using perception based features. In: National Conference on Communications. pp. 1–6. Mumbai, India (2015)
42. Wang, W., Yang, H., Fu, J., Liu, J.: Zero-Reference Low-Light Enhancement via Physical Quadruple Priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 26057–26066. Seattle , USA (2024)
43. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 606–615. Salt Lake City, USA (2018)
44. Wang, Z., Chen, J., Hoi, S.C.: Deep Learning for Image Super-Resolution: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **43**(10), 3365–3387 (2020)
45. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
46. Wei, C., Wang, W., Yang, W., Liu, J.: Deep Retinex Decomposition for Low-Light Enhancement. In: Proceedings of the British Machine Vision Conference (BMVC). pp. 155–167. Newcastle, UK (2018)
47. Wei, P., Xie, Z., Lu, H., Zhan, Z., Ye, Q., Zuo, W., Lin, L.: Component Divide-and-Conquer for Real-World Image Super-Resolution. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 101–117. Online (2020)

48. Wei, X., Chang, K., Li, G., Huang, M., Qin, Q.: DLEN: Deep Laplacian Enhancement Networks for Low-Light Images. In: IEEE International Conference on Image Processing (ICIP). pp. 2120–2124. Kuala Lumpur, Malaysia (2023)
49. Xing, W., Egiazarian, K.: End-to-End Learning for Joint Image Demosaicing, Denoising and Super-Resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3507–3516. Nashville, USA (2021)
50. Xu, X., Ma, Y., Sun, W., Yang, M.H.: Exploiting Raw Images for Real-Scene Super-Resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(4), 1905–1921 (2022)
51. Xu, X., Wang, R., Fu, C.W., Jia, J.: SNR-Aware Low-Light Image Enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17714–17724. New Orleans, USA (2022)
52. Yang, K.F., Cheng, C., Zhao, S.X., Yan, H.M., Zhang, X.S., Li, Y.J.: Learning to Adapt to Light. International Journal of Computer Vision **131**(4), 1022–1041 (2023)
53. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient Transformer for High-Resolution Image Restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5728–5739. New Orleans, USA (2022)
54. Zhang, X., Chen, Q., Ng, R., Koltun, V.: Zoom to Learn, Learn to Zoom. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3762–3770. Long Beach, USA (2019)
55. Zhang, Y., Zhang, J., Guo, X.: Kindling the Darkness: A Practical Low-light Image Enhancer. In: Proceedings of the ACM International Conference on Multimedia (ACM MM). pp. 1632–1640. Nice France (2019)
56. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 286–301. Munich, Germany (2018)
57. Zheng, N., Zhou, M., Dong, Y., Rui, X., Huang, J., Li, C., Zhao, F.: Empowering Low-light Image Enhancer through Customized Learnable Priors. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 12559–12569. Paris, France (2023)
58. Zhou, S., Li, C., Loy, C.C.: LEDNet: Joint Low-light Enhancement and Deblurring in the Dark. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 573–589. Tel-Aviv, Israel (2022)
59. Zhou, Y., Li, Z., Guo, C.L., Bai, S., Cheng, M.M., Hou, Q.: SRFormer: Permuted Self-Attention for Single Image Super-Resolution. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 12780–12791. Paris, France (2023)