Joint Low-light Enhancement and Super Resolution with Image Underexposure Level Guidance

Mingjie Xu[1]
xumingjie@buaa.edu.cn
Chaoqun Zhuang[1]
zhuangchaoqun@buaa.edu.cn
Feifan Lv[2]
lvfeifan@buaa.edu.cn
Feng Lu[1*]
lufeng@buaa.edu.cn

[1] State Key Laboratory of VR Technology and Systems, School of CSE, Beihang University, Beijing, China
[2] Huawei Consumer BG Consumer Cloud Service Dept, Nanjing, China

Abstract

Obtaining high-quality images with high resolution in poor illumination environments using a limited spatial resolution image sensor poses a significant challenge. Low-light Enhancement (LLE) and Super-Resolution (SR) are crucial technologies for overcoming this challenge. However, current approaches usually generate normal-light high-resolution images with non-uniform brightness and loss of details from low-light low-resolution images, and suffer from significant performance degradation in cross-dataset settings. To alleviate these problems, we propose a novel solution for low-light image super-resolution. For non-uniform brightness problem, we propose a Relative Underexposure Level Estimation Module (RUL-EM) that estimates the relative underexposure levels of input images to adjust the image brightness to a uniform level and avoid artifacts. For detail loss and cross-dataset problems, we introduce the Multi-Scale Sampling (MSS) strategy for sampling multi-scale patches. MSS involves randomly cropping low-light and low-resolution patches of different sizes and positions and resizing them to a given patch size. Combining RUL-EM with MSS can improve the model performance in detail restoration and generalization. Additionally, we also incorporate channel attention to enable the Joint LLE & SR Network (JLSN) to adaptively adjust the influence of estimated relative underexposure levels. Our proposed method can be applied to various backbone architectures. Experimental results show that our proposed method achieves state-of-the-art performance on the joint LLE & SR task in both within-dataset and cross-dataset settings. Our proposed solution can convert low-resolution low-light images into high-resolution images with satisfactory brightness, vivid colors, and more details.

1 Introduction

In the real world, various image degradation problems significantly reduce image quality [28, 29] and create inconveniences for many kinds of applications, such as autonomous

* Corresponding Author.

driving [48], medical imaging [36], *etc*. Among these problems, two main issues are low-light (LL) and low-resolution (LR). Low-light problems are mainly caused by poor lighting conditions or limited exposure time when capturing images, while low-resolution problems are mainly caused by limited spatial resolution image sensors and image transmission over the Internet [34]. In a common scenario, images captured in low-light environments using regular cameras usually suffer from both low-light and low-resolution problems [2, 34].

In fact, obtaining high-quality images with high resolution in poor illumination environments using a limited spatial resolution image sensor is a highly challenging task. Both Low-Light Enhancement (LLE) [27, 45, 46] and Super-Resolution (SR) [23, 42] are crucial technologies for tackling this task. However, when inputting LL & LR images, the output normal-light (NL) & high-resolution (HR) images of current approaches are facing the following problems: (1) **non-uniform brightness** when inputting images of varying brightness, (2) **loss of details** and (3) **significant performance degradation on unseen datasets**, *i.e.*, cross-dataset problem, which limits the generalization to the real world. These problems present a challenge in solving joint LLE & SR problem.

For the non-uniform brightness problem, it is crucial to incorporate the relative underexposure levels into the input to obtain uniform brightness from input images with varying brightness. Relative Underexposure Level (RUL) reflects the lightness of an image, ranging from $\{0,1,2,3,4,5,6\}$, with higher values denoting darker images and lower values indicating lighter images. Note that RUL is not a precise physical quantity. As such relative underexposure levels are difficult to obtain in the real world, we propose a Relative Underexposure Level Estimation Module (RUL-EM) to predict the accurate relative underexposure levels from LL & LR images, thereby enabling the model to adjust the image brightness to a uniform level from the input images of different brightness, and avoid artifacts. However, the RUL-EM alone cannot tackle the detail loss and the cross-dataset problems. We argue that these two problems are caused by the use of single-scale patches in the joint LLE & SR task, which is due to the fixed-size patch sampling strategy used for training (frequently used in low-level vision tasks). To overcome this, inspired by [50], we introduce the Multi-Scale Sampling (MSS) strategy [50] into our proposed method, which is used in other fields like image captioning [50] but rarely used in low-level vision. Different from [50], we randomly crop LL & LR patches of different sizes and positions, and resize them to a given patch size. Combining RUL-EM with MSS can improve detail restoration and generalization performance of the model, which is imperative for the joint LLE & SR task. Finally, we propose a Joint LLE & SR Network (JLSN) that incorporates channel attention [14] to enable the network to adaptively adjust the influence of the estimated relative underexposure levels, where various backbone architectures can be used. Based on these three components, we obtain a novel solution for low-light image super-resolution, which takes relative underexposure levels and multi-scale patches into consideration. Our proposed joint LLE & SR solution can alleviate the non-uniform brightness, the detail loss and the cross-dataset problems described above, as shown in Fig. 2, 3 and 5.

To summarize, our contributions are: 1) We propose a novel joint LLE & SR solution which can address the above-mentioned problems including non-uniform brightness, detail loss and cross-dataset problem. 2) We propose RUL-EM to accurately predict relative underexposure levels for adjustment of the image brightness to a uniform level and artifact avoidance, introduce MSS [50] to sample multi-scale patches of the scene and improve the ability of detail restoration and generalization with the help of RUL-EM, and propose JLSN that uses channel attention [14] to enable the network to adaptively adjust the influence of the estimated relative underexposure levels. 3) Both the quantitative and qualitative results show that our proposed method can achieve state-of-the-art (SOTA) performance in the joint LLE & SR task in both within-dataset and cross-dataset settings, including all evaluation metrics, and produce high resolution images with satisfactory brightness, vivid colors, and more details.

2 Related Works

Image Super-Resolution. The Single Image Super-Resolution (SISR) task aims to increase the pixel density and enrich the details of an image. With the development of deep learning, SRCNN [6] first applied Convolutional Neural Network (CNN) to SISR. Since then, many works have made progress by stacking more convolutional layers and designing more complex network connections, like SRResNet [20], EDSR [24], RCAN [53], RDSR [17] and DCLS [26]. In order to solve the difficulty of signal-fidelity-oriented methods in reproducing image details, GAN [8] has a wide range of applications in the SISR task, including SRGAN [19], DBPN [12], ESRGAN [41], Real-ESRGAN [42] and LDL [22]. Recent SISR researches focus on Transformer [7, 25] architecture, such as SwinIR [23]. These methods perform well in recovering image details for normal-light low-resolution images. However, they cannot adjust image brightness and recover details well for low-light low-resolution images.

Low-light Image Enhancement (LLE). There exist various deep-learning-based LLE approaches, which are all able to achieve SOTA performance on popular benchmarks. These techniques typically rely on multi-level feature with multi-branch fusion [27], illumination cue [40], Retinex [18] theory [35, 44, 51, 52], pixel-wise and high-order curves [9, 21], multi-resolution features with spatial-wise and channel-wise attention [46], GAN [16], attention [30], self-calibrated illumination [31], Signal-to-Noise-Ratio prior [45], local-and-global components decoupling [5], local color distribution information [39] and over-and-under exposure consideration [33]. Although these approaches are able to tackle LLE problem well, their performance will degrade when applying them to joint LLE & SR tasks.

Joint LLE & SR. To the best of our knowledge, there are only a few methods that tackles joint LLE & SR task. Rasheed *et al*. proposed a joint LLE & SR network LSR using both the Lighten-Projection Unit and Darken-Projection Unit [34]. Aakerberg *et al*. also proposed a joint LLE & SR method named RELIEF [2]. Zhang *et al*. proposed a joint LLE & SR method for videos [47]. Guo *et al*. proposed a joint LLE & SR method for face images [10]. Moreover, RELLISUR dataset [1] is specifically designed for the image joint LLE & SR task. Such joint LLE & SR task is still open, which lacks research.

3 Methodology

In this section, we describe our proposed joint LLE & SR method. We begin by introducing the Relative Underexposure Level Estimation Module (RUL-EM) in Sec. 3.1, followed by the Multi-Scale Sampling (MSS) strategy in Sec. 3.2. Finally, we discuss the Joint LLE & SR Network (JLSN) in Sec. 3.3.

3.1 Relative Underexposure Level Estimation Module (RUL-EM)

Problem Analysis. Directly generating NL & HR images from LL & LR images can result in non-uniform brightness when inputting images with varying brightness. To tackle this problem, it is beneficial to use Relative Underexposure Levels (RUL) as input for the complex, degraded and coupled joint LLE and SR task. Relative underexposure levels are naturally and strongly correlated with brightness adjustment, which can help to enhance the

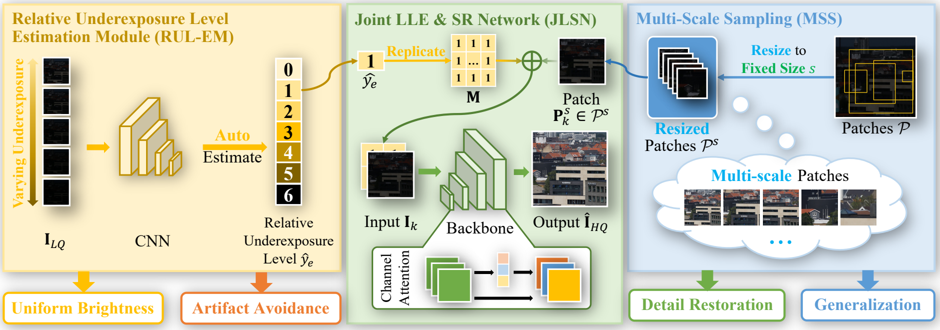


Figure 1: Overview of proposed relative underexposure level guided joint LLE & SR method.

brightness of low-light low-resolution images with varying brightness into a uniform level. Moreover, such relative underexposure levels can help avoid artifacts.

Obtaining relative underexposure levels for common RGB images distributed widely across the Internet is difficult. Thanks to the ground truth relative underexposure levels provided by the RELLISUR [1] dataset, we propose a Relative Underexposure Level Estimation Module (RUL-EM) to tackle such problem. Formally, this strategy can be expressed as $\hat{y}_e = \mathcal{E}(\mathbf{I}_{LQ})$, where $\hat{y}_e$ is the estimated relative underexposure level and $\mathcal{E}$ is the RUL-EM.

Implementation. The RELLISUR dataset [1] provides 7 relative underexposure levels, which are only accessible in the training set. The goal of RUL-EM is to accurately predict such relative underexposure level of an input image. To achieve this, we design RUL-EM as a classification module, which is based on ResNet-50 [13] with Squeeze-and-Excitation module [14], as shown in Fig. 1. The model is trained using widely-used Cross Entropy Loss, and used to guide the subsequent MSS (Sec. 3.2) and JLSN (Sec. 3.3).

Discussion. Intuitively, RUL-EM pre-determines a certain relative underexposure level for a low-light low-resolution image. Based on this, the model learns a restricted problem, which is easier for avoiding artifacts.

3.2 Multi-Scale Sampling (MSS)

Problem Analysis. Although RUL-EM can help tackle the non-uniform brightness problem and avoid artifacts, experimental results have shown that RUL-EM alone can not tackle the detail loss and the cross-dataset problems (see Fig. 5). With the guidance of relative underexposure levels estimated by RUL-EM, there needs a strategy for detail restoration. Inspired by [50], we argue that using multi-scale patches with RUL-EM, instead of single-scale patches, can help alleviate the detail loss problem, further improving the cross-dataset performance.

Typically, current LLE or SR approaches sample fixed-size image patches for training [2, 21, 23, 45]. Such fixed-size patches are single-scale. For both single LLE or SR tasks, such single-scale patches contain enough information for color/illumination enhancement or detail restoration, respectively. However, single-scale patches are not enough for the joint LLE & SR task. If directly using single-scale patches for training the joint LLE & SR model, the model tends to generate over-smoothed and detail-lost images, as shown in Fig. 5. Although RUL-EM (Sec. 3.1) can help to avoid artifacts, using it alone cannot help the model for detail restoration. Moreover, using single-scale patches can cause the model to overfit as they do not contain enough information for the joint LLE & SR task, resulting in significant performance degradation on unseen datasets. The supplementary material provides evidence that even large fixed-size patches are unable to address these issues. To tackle these issues, based on RUL-EM, multi-scale patches need to be applied into the training procedure. Thus, we introduce the Multi-Scale Sampling (MSS) strategy [50] into our method for sampling multi-scale patches, which is used in other fields like image captioning [50] but rarely used in low-level vision.

Overview. Different from [50], in brief, the MSS strategy randomly crops LL & LR patches with **different sizes and positions** and resizes them to a given patch size, as shown in Fig. 1. With the help of both RUL-EM and MSS, the joint LLE and SR model can easily obtain multi-scale patches, improving its ability of detail restoration and generalization on unseen datasets, which is imperative for the joint LLE & SR task.

Procedure. Given the full low-light low-resolution image $\mathbf{I}_{LQ}$ with size $H \times W \times 3$, we randomly sample multiple patches $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots\}$ with different sizes $p \times p \times 3$ where $p \sim \mathcal{U}(p_{low}, p_{high})$ and different top-left positions $x \sim \mathcal{U}(0, H-p)$ and $y \sim \mathcal{U}(0, W-p)$, where $\mathcal{U}$ is the Uniform distribution. Then we resize all the patches $\mathbf{P}_k$ in $\mathcal{P}$ into the given size s, which outputs $\mathcal{P}^s = \{\mathbf{P}_1^s, \mathbf{P}_2^s, \ldots\}$ with size $s \times s \times 3$.

Discussion. MSS addresses the cross-dataset generalization problem by sampling patches with a wide range of scales, making the model learn various scales that may appear in unseen datasets during training. Without MSS, the model is unable to effectively learn the diverse scales present in unseen datasets since the training dataset only covers a limited range of scales across different input images.

3.3 Joint LLE & SR Network (JLSN)

Overview. In the joint LLE & SR procedure, the relative underexposure levels estimated by RUL-EM (Sec. 3.1) and the multi-scale patches sampled by MSS strategy (Sec. 3.2) are helpful for restoring NL & HR images $\hat{\mathbf{I}}_{HQ}$ from LL & LR images $\mathbf{I}_{LQ}$. Here we propose a Joint LLE & SR Network (JLSN) to restore NL & HR images with the guidance of such information, which is shown in Fig. 1. The JLSN consists of a Generator G and Discriminator D following the classic GAN [8] structure, and can use various backbone architectures.

Generator G. Our proposed method is in the form of add-on (plug-and-play), which constitutes one of its key advantages. We employ commonly-used LLE architecture MIRNet [46] (with upsampling module [37], a representative LLE backbone), SR architectures, including RRDB [41, 42] (a representative CNN-based backbone) and SwinIR [23] (a representative Transformer-based backbone), and a cascaded LLE & SR architecture MIRNet [46]+RRDB [41, 42] as G. To fully exploit the connection between the relative underexposure levels and the LL & LR image $\mathbf{I}_{LQ}$, we incorporate the Channel Attention (CA) structure [14] into RRDB, SwinIR and MIRNet+RRDB Module (MIRNet+Upsample originally contains CA), allowing the JLSN to adjust the influence of the estimated relative underexposure levels adaptively. We refer to the modified generator models as CA-RRDB, CA-SwinIR and MIRNet+CA-RRDB, respectively. Implementation details can be found in the supplementary material. Further experiments show that plugging our proposed method into various backbone can all achieve SOTA results and does not increase the inference time, as shown in Sec. 4.2 and the supplementary material, respectively.

Discriminator D. Here we use the U-Net Discriminator with spectral normalization following [42] to increase the adaptability of the network to real-world scenarios.

RUL-EM and MSS Information Combination. To make use of the information obtained by RUL-EM and MSS, we replicate the estimated relative underexposure level y_e into the relative underexposure map $\mathbf{M}$ with size $s \times s \times 1$. We then concatenate $\mathbf{M}$ and the multi-scale patch $\mathbf{P}_k$ to obtain the input $\mathbf{I}_k = \mathbf{P}_k \oplus \mathbf{M}, \mathbf{P}_k \in \mathcal{P}^s$ to G, where $\oplus$ indicates the concatenate operation. Note that this procedure is only processed in the training phase. To simplify the implementation, we randomly select one p and one (x,y) for each image in one batch. In the testing phase, we directly feed the full $\mathbf{I}_{LQ}$ concatenated with the estimated relative underexposure map (replicated using the estimated relative underexposure level $\hat{y}_e$) into the JLSN to generate the normal-light high-resolution image.

Loss Function. We use L_1 Pixel Loss $\mathcal{L}_{pix}$, Perceptual Loss $\mathcal{L}_{per}$ [41, 42] and Adversarial Loss $\mathcal{L}_{adv}$ [8] for training G, and the Adversarial Loss $\mathcal{L}_D$ [8] for training D, which is commonly used in both LLE methods and SR methods. The total loss function of the generator G is $\mathcal{L}_G = \mathcal{L}_{pix} + \lambda_{per}\mathcal{L}_{per} + \lambda_{adv}\mathcal{L}_{adv}$, where λ_{per} and λ_{adv} are the coefficients to balance different terms of $\mathcal{L}_G$, and we empirically set them as $\{1.0, 0.005\}$ respectively.

4 Experiments

4.1 Setup

Dataset. We use the RELLISUR ($\mathcal{D}_R$) dataset [1] for within-dataset evaluation, which provides real-world LL & LR images of 7 different relative underexposure levels and NL & HR images with $\times 1$, $\times 2$ or $\times 4$ magnifications, which can be used to model real-world physical down-sampling and noise. There are 3610, 215, and 425 images for training, validation, and testing, respectively. Moreover, to evaluate the cross-dataset performance, we also use LOL ($\mathcal{D}_L$) [44], LSRW ($\mathcal{D}_S$) [11] and DIV2K ($\mathcal{D}_D$) [3, 38] test sets, which provide 15, 50 and 100 images, respectively. This forms 3 cross-dataset tasks: $\mathcal{D}_R \to \mathcal{D}_L$, $\mathcal{D}_R \to \mathcal{D}_S$ and $\mathcal{D}_R \to \mathcal{D}_D$, respectively. Since LOL [44] and LSRW [11] are designed for LLE task, they only contain NL & LR ground truth images. Since DIV2K [3, 38] is designed for SR task, we perform Gamma correction with $\gamma \sim \mathcal{U}(2, 3.5)$ on the input LR images following [27].

Comparison Methods. Our task is joint LLE and SR. So it is necessary to compare our proposed method with SOTA LLE and SR methods. We conduct a comparison with SOTA and latest methods on single LLE or SR tasks on RELLISUR dataset. Results are shown in the supplementary material. Based on these results, we select the LLE methods SNR-Aware [45], MBLLEN [27] and MIRNet [46], as well as the SR methods SRResNet [20], EDSR [24], Real-ESRGAN [42], SwinIR [23] and LDL [22] which have the best performances, for further experiments and comparison. Moreover, we compare our proposed method with SOTA joint LLE & SR methods LSR [34] and RELIEF [2]. We compare our proposed method with 6 strategy designs: (I) Apply LLE methods to obtain NL & LR images from LL & LR images, followed by SR methods to obtain the desired NL & HR images, (II) Reversed Sequential Process of Type I, (III) Cascaded LLE Network and SR Network, (IV) LLE Network + Upsampling Module [37], (V) SR Network, and (VI) Joint LLE & SR Network. Implementation details will be discussed in the supplementary material.

Our proposed method is in the form of add-on (plug-and-play). To demonstrate the SOTA performance of our proposed method with different backbones, we use 4 versions of our proposed method with different backbones for comparison, including CA-RRDB, CA-SwinIR, MIRNet+CA-RRDB and MIRNet+Upsample, as stated in Sec. 3.3.

Evaluation Metrics. We use Peak Signal to Noise Ratio (PSNR) [15] and Structural Similarity (SSIM) [43] to evaluate the similarity between a HR image and its generated SR counterpart (higher is better), while we use Learned Perceptual Image Patch Similarity (LPIPS) [49] and Natural Image Quality Evaluator (NIQE) [32] to evaluate human perceptual quality of

Type	Method	RELLISUR ($\mathcal{D}_R$) ×2 PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓	RELLISUR ($\mathcal{D}_R$) ×4 PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓	$\mathcal{D}_R \to \mathcal{D}_L$ ×4 NIQE ↓	$\mathcal{D}_R \to \mathcal{D}_S$ ×4 NIQE ↓	$\mathcal{D}_R \to \mathcal{D}_D$ ×4 PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓
I	SNR-Aware [■]→SRResNet [■]	20.19	0.75	0.32	7.46	19.49	0.75	0.46	9.29	7.50	7.59	14.02	0.41	0.56	7.41
	MIRNet [■]→SRResNet [■]	21.46	0.75	0.47	8.79	20.49	0.76	0.59	10.59	8.78	8.72	15.70	0.56	0.60	7.80
	MBLLEN [■]→SRResNet [■]	18.88	0.74	0.46	8.92	18.23	0.74	0.57	10.43	8.53	8.25	14.00	0.51	0.65	8.03
	SNR-Aware [■]→EDSR [■]	20.23	0.75	0.33	7.67	19.15	0.75	0.48	9.06	7.11	7.32	13.91	0.42	0.57	7.11
	MIRNet [■]→EDSR [■]	21.45	0.75	0.46	8.96	20.08	0.75	0.59	9.88	8.34	8.33	15.53	0.55	0.60	7.71
	MBLLEN [■]→EDSR [■]	18.88	0.75	0.46	8.95	17.79	0.74	0.58	9.82	8.39	7.94	13.87	0.51	0.66	7.99
	SNR-Aware [■]→Real-ESRGAN [■]	20.02	0.75	0.30	6.91	19.32	0.74	0.41	7.40	5.13	5.20	13.38	0.37	0.52	5.69
	MIRNet [■]→Real-ESRGAN [■]	21.50	0.74	0.41	7.65	20.48	0.75	0.51	8.85	6.66	6.49	15.71	0.54	0.52	5.99
	MBLLEN [■]→Real-ESRGAN [■]	18.78	0.74	0.42	7.90	18.12	0.73	0.51	8.67	6.49	5.90	13.63	0.49	0.59	6.52
	SNR-Aware [■]→SwinIR [■]	20.31	0.75	0.33	7.46	19.46	0.75	0.48	9.32	7.43	7.41	13.91	0.40	0.58	7.03
	MIRNet [■]→SwinIR [■]	21.44	0.75	0.46	8.94	20.48	0.76	0.59	10.37	8.88	8.74	15.88	0.55	0.61	7.66
	MBLLEN [■]→SwinIR [■]	18.83	0.75	0.46	8.91	18.21	0.74	0.57	9.94	8.44	7.82	13.93	0.51	0.66	7.55
	SNR-Aware [■]→LDL [■]	19.99	0.75	0.30	6.78	19.35	0.74	0.41	7.53	4.91	4.99	13.56	0.38	0.51	5.24
	MIRNet [■]→LDL [■]	21.43	0.74	0.41	7.37	20.40	0.75	0.51	8.84	6.94	6.43	15.51	0.54	0.52	5.83
	MBLLEN [■]→LDL [■]	18.80	0.74	0.42	7.73	18.11	0.74	0.51	8.71	6.66	5.85	13.66	0.49	0.59	6.20
II	SRResNet [■]→SNR-Aware [■]	20.00	0.75	0.40	7.53	18.52	0.72	0.57	8.29	6.74	6.53	14.39	0.48	0.49	5.94
	SRResNet [■]→MIRNet [■]	21.32	0.74	0.49	8.17	20.01	0.74	0.62	9.23	6.96	7.31	15.31	0.56	0.54	6.68
	SRResNet [■]→MBLLEN [■]	18.60	0.73	0.50	8.40	17.74	0.72	0.65	9.61	6.78	7.17	13.72	0.53	0.57	6.65
	EDSR [■]→SNR-Aware [■]	19.14	0.73	0.45	7.18	18.40	0.74	0.58	9.11	5.86	6.18	14.05	0.48	0.53	5.37
	EDSR [■]→MIRNet [■]	20.18	0.73	0.52	8.28	19.64	0.74	0.60	9.83	7.16	7.32	14.90	0.55	0.57	6.93
	EDSR [■]→MBLLEN [■]	17.84	0.71	0.55	8.54	17.44	0.73	0.65	10.17	6.92	7.09	13.23	0.50	0.61	6.34
	Real-ESRGAN [■]→SNR-Aware [■]	20.00	0.76	0.35	7.73	18.34	0.71	0.46	7.29	4.79	5.23	14.44	0.42	0.50	5.23
	Real-ESRGAN [■]→MIRNet [■]	21.11	0.75	0.37	7.26	19.59	0.72	0.47	7.27	5.95	6.12	15.28	0.50	0.54	6.12
	Real-ESRGAN [■]→MBLLEN [■]	18.65	0.74	0.41	8.11	17.61	0.71	0.50	8.57	6.04	6.23	14.08	0.49	0.57	6.04
	SwinIR [■]→SNR-Aware [■]	18.00	0.72	0.46	7.64	18.55	0.73	0.62	8.79	6.33	6.51	14.50	0.48	0.51	5.58
	SwinIR [■]→MIRNet [■]	19.35	0.72	0.51	8.10	19.45	0.74	0.66	9.14	7.29	7.40	15.48	0.56	0.56	6.69
	SwinIR [■]→MBLLEN [■]	17.42	0.71	0.54	8.67	17.67	0.72	0.68	9.49	7.06	7.34	13.89	0.53	0.58	6.69
	LDL [■]→SNR-Aware [■]	19.87	0.75	0.35	7.77	18.84	0.74	0.45	7.79	5.12	5.40	14.93	0.46	0.51	5.24
	LDL [■]→MIRNet [■]	20.59	0.73	0.38	6.96	20.01	0.74	0.47	7.46	6.13	6.25	15.56	0.53	0.55	6.13
	LDL [■]→MBLLEN [■]	18.23	0.73	0.42	8.00	17.73	0.73	0.50	8.85	6.24	6.40	14.40	0.49	0.59	6.04
III	SNR-Aware [■]+RRDB [■]	20.76	0.77	0.35	8.21	19.22	0.76	0.45	7.98	6.80	6.36	15.44	0.47	0.53	6.86
	MIRNet [■]+RRDB [■]	22.24	0.79	0.26	7.69	20.93	0.78	0.38	8.35	6.63	6.59	15.08	0.51	0.52	6.73
	MBLLEN [■]+RRDB [■]	19.00	0.76	0.29	7.57	18.36	0.75	0.41	8.62	6.60	6.58	15.77	0.49	0.53	6.86
IV	SNR-Aware [■]+Upsample	20.98	0.77	0.34	7.73	20.06	0.76	0.49	9.00	7.64	7.42	13.68	0.22	0.72	10.06
	MIRNet [■]+Upsample	22.57	0.78	0.37	8.07	20.81	0.78	0.52	9.11	7.04	6.88	17.03	0.50	0.52	5.83
	MBLLEN [■]+Upsample	19.24	0.77	0.38	8.46	18.39	0.76	0.53	10.26	7.69	7.94	16.48	0.52	0.57	6.95
V	SRResNet [■]	18.94	0.72	0.45	7.09	18.15	0.74	0.56	7.42	5.78	5.78	15.39	0.52	0.54	5.17
	EDSR [■]	19.03	0.74	0.45	8.02	18.13	0.74	0.60	9.53	9.16	7.99	7.00	0.43	0.78	8.24
	Real-ESRGAN [■]	19.30	0.77	0.28	7.45	18.59	0.76	0.40	8.32	6.52	6.40	15.61	0.47	0.53	6.58
	SwinIR [■]	19.49	0.74	0.49	9.71	18.44	0.74	0.63	9.71	8.28	8.01	15.23	0.55	0.57	6.65
	SwinIR [■]+GAN [■]	19.13	0.75	0.30	7.91	18.58	0.75	0.41	8.35	6.47	6.49	16.58	0.56	0.51	6.74
	LDL [■]	18.98	0.76	0.28	6.94	18.28	0.74	0.40	7.67	5.06	5.09	13.73	0.43	0.52	5.71
VI	LSR [■]	19.50	0.74	0.47	8.06	18.65	0.75	0.61	8.91	7.13	6.98	16.49	0.51	0.64	5.95
	LSR [■]+GAN [■]	18.70	0.74	0.29	7.60	17.90	0.74	0.41	8.16	6.55	6.30	16.49	0.54	0.54	6.27
	RELIEF [■]	21.18	0.75	0.46	8.63	20.10	0.76	0.59	9.85	8.11	8.31	16.48	0.58	0.63	7.47
	LCUN [■]	22.39	0.78	0.34	6.81	21.23	0.77	0.50	8.75	6.61	7.58	15.90	0.49	0.57	6.94
	Ours (CA-RRDB)	22.55	0.75	0.32	5.39	21.52	0.77	0.39	6.77	5.90	4.91	17.29	0.58	0.34	4.73
	Ours (CA-SwinIR)	22.74	0.78	0.25	6.32	21.30	0.77	0.37	6.41	4.31	4.03	17.33	0.61	0.32	4.29
	Ours (MIRNet+CA-RRDB)	23.06	0.79	0.26	6.58	21.56	0.78	0.38	7.21	4.18	4.30	17.25	0.59	0.36	4.35
	Ours (MIRNet+Upsample)	23.08	0.79	0.27	6.79	21.60	0.78	0.38	6.77	4.75	4.49	17.48	0.59	0.35	4.70

Table 1: Quantitative results of SOTA methods. ↑ indicates larger is better, ↓ indicates lower is better. **Red**, **blue** and **green** denotes the first, second and third best results, respectively.

an image (lower is better). Note that PSNR, SSIM and LPIPS are full-reference metrics, whereas NIQE is a no-reference metric.

Implementation Details. We use PyTorch to conduct all the experiments on NVIDIA GPUs. All proposed models are trained for 200,000 iterations with a batch size of 24. For both G and D, we use the Adam optimizer with $lr = 10^{-4}$, $\beta_1 = 0.9$ and $\beta_2 = 0.99$, and set the weight decay as 0. The LR patch size s is set to 64 for all the experiments. We set the dimension of the feature of both G and D as 64. For CA-RRDB, we use 7 blocks, and the channel growth of the dense block is set to 32. For other architectures, we use their original settings. We use the same setting of Perceptual Loss as Real-ESRGAN [42]. For MSS, we set p_{low} as 64 and p_{high} as 512. For RUL-EM, the model is pretrained before training other modules. The input size is 625 as it is not cropped, and the model is trained for 200 epochs with the SGD optimizer with $lr = 0.002$, $\gamma = 0.9$, and we set the weight decay as 5×10^{-4}.

4.2 Comparison with SOTA Methods

To show the superior performance of our proposed method on joint LLE & SR task, we conduct comparison experiments with SOTA methods on RELLISUR dataset [■]. Quantitative and qualitative results are shown in Tab. 1 and Fig. 2, respectively. In Tab. 1, our proposed joint LLE & SR methods can outperform all types of SOTA methods on all evaluation met-

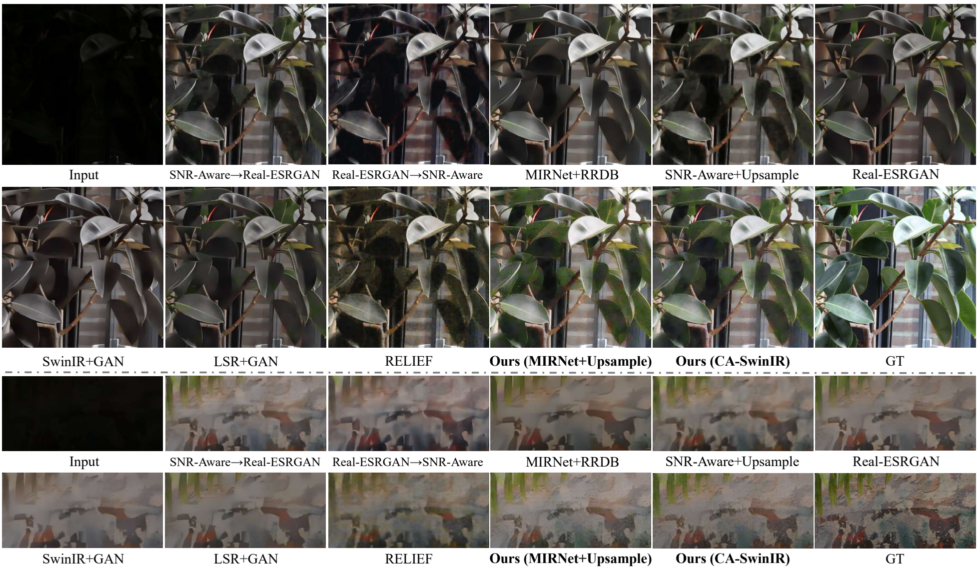


Figure 2: Qualitative results on RELLISUR dataset (x4). Zoom for best view. Our proposed methods tend to generate **vivid colors and more details**.

rics, indicating that our proposed methods can generate NL & HR images from LL & LR images with superior quality. Furthermore, in Fig. 2, our proposed methods produce more vivid colors, realistic textures and fewer artifacts than other methods, closely approximating the ground truth images.

4.3 Real-World Results

The above experiments focused on evaluating joint LLE & SR methods all on a within-dataset setting. However, some methods tend to overfit on one single dataset. As it is important to apply joint LLE & SR methods in the real world, it is necessary to evaluate them in the real-world setting (cross-dataset setting). In this section, we evaluate joint LLE & SR methods on LOL [44], LSRW [11] and DIV2K [3, 38] datasets. Note that we directly use the model trained on the RELLISUR dataset for evaluation without re-training. Qualitative results are shown in Fig. 3. Due to the space limitation, we only present some examples from LOL dataset, while additional examples are available in the supplementary material. The results indicate that other methods tend to generate over-smoothed images with more artifacts and detail loss. This is due to overfitting on the training dataset. Thanks to the proposed RUL-EM with MSS, our proposed methods tend to produce clearer images with more details and vivid colors, demonstrating their superior performance in solving such a challenging joint LLE & SR problem in the cross-dataset setting. Moreover, the quantitative results in the 6 rightmost columns of Tab. 1 show that in cross-dataset setting, our proposed methods also achieve the best performance (as there is no HR ground truth HR images available in LOL and LSRW datasets, we only report NIQE results, which is a no-reference metric).

4.4 Understanding of Our Proposed Method

Effectiveness of RUL-EM. During the testing phase, relative underexposure levels cannot be obtained. Therefore, accurately estimating relative underexposure levels is required. When testing on the RELLISUR test split, our RUL-EM can achieve 77.8% accuracy. Meanwhile, we draw the confusion matrix (Fig. 4(a)). As the figure shows, even though RUL-EM cannot

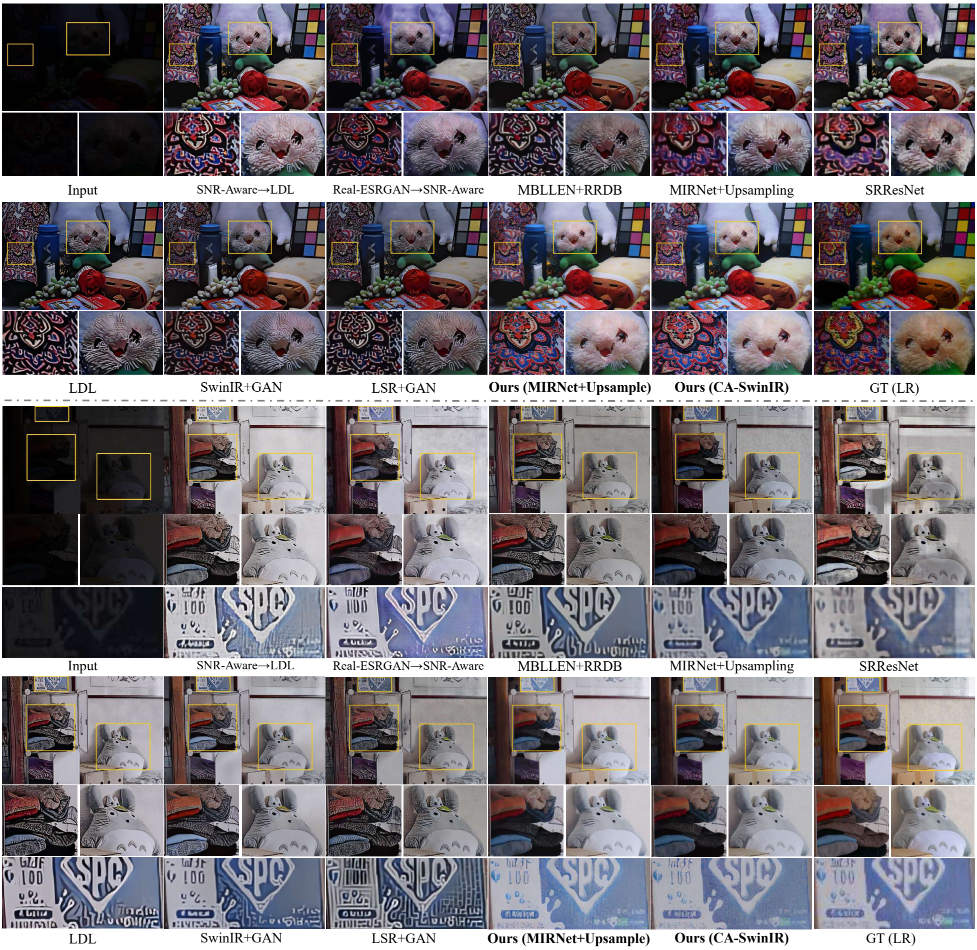
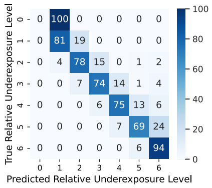


Figure 3: Qualitative results on LOL dataset (x4). Zoom for best view. Our proposed methods tend to generate **vivid colors, more details and fewer artifacts** on unseen datasets.

always accurately predict the relative underexposure levels, it can predict them similar to ground truth for most of the time. So RUL-EM can be used in subsequent procedures.

Ablation Study. To further show the efficacy of our proposed method and the necessity of all its components, we conduct an ablation study using different combinations of components. All the experiments are conducted using the CA-RRDB backbone, which is a common and simple-structured CNN in LLE or SR and can represent prevalent cases. Quantitative and qualitative results are shown in Fig. 4(b) and Fig. 5 respectively. With the help of the RUL-EM, the quality of output $\hat{\mathbf{I}}_{HQ}$ significantly increases, as indicated by the increased PSNR and SSIM values, and the joint LLE & SR network tends to adjust the input low-light images to uniform brightness. Additionally, RUL-EM can help the network to avoid some artifacts. However, using RUL-EM alone cannot tackle the detail loss problem. Based on RUL-EM, if MSS is combined into the network, an image with uniform brightness with more details and fewer artifacts can be obtained than RUL-EM alone. Moreover, using RUL-EM with Channel Attention (CA) can further improve the PSNR, SSIM and LPIPS values, and make the network product more vivid colors and fewer artifacts. This is mainly because CA can

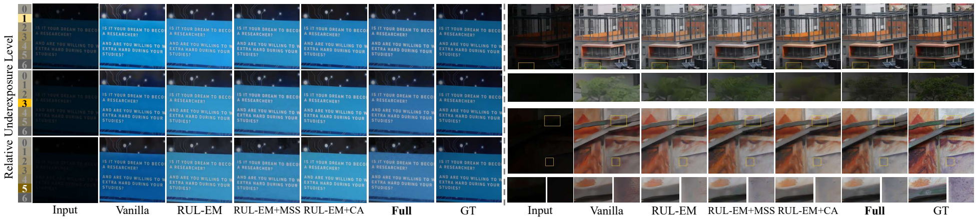


(a)

Strategy	PSNR↑	SSIM↑	LPIPS↓	NIQE↓
Vanilla	18.59	0.76	0.40	8.32
RUL-EM	20.97	0.78	0.40	8.18
RUL-EM + MSS	20.76	0.76	0.42	5.85
RUL-EM + CA	21.66	0.79	0.39	8.62
RUL-EM + MSS + CA	21.52	0.77	0.39	6.77

(b)

Figure 4: (a) Confusion matrix of RUL-EM on the test split of RELLISUR dataset, percentage is shown. (b) Ablation study on different strategies on joint LLE & ×4 SR task. ↑ indicates larger is better, ↓ indicates lower is better.



(a) Comparisons of different strategies on **adjustment brightness to a uniform level** (b) Comparisons of different strategies on **detail recovery** and **artifact avoidance**

Figure 5: Qualitative results of ablation study (x4). Zoom for best view.

help the network better handle the relationship between the relative underexposure levels and the input. Finally, using both RUL-EM and MSS with CA can obtain the most details and the colors closest to the ground truth with the most uniform brightness. Note that MSS may cause a slight decrease in PSNR and SSIM values. This is because details and PSNR/SSIM values are a pair of trade-off as noted in [19]. Higher PSNR/SSIM values tend to be related to over-smoothing problems, while more details may cause lower PSNR/SSIM values [19]. It can be concluded from the ablation study that all the components are necessary for the joint LLE & SR task.

5 Conclusion

In this paper, we propose a novel solution for the joint LLE and SR task. We propose a Relative Underexposure Level Estimation Module (RUL-EM) to accurately estimate relative underexposure levels for adjusting the image brightness to a uniform level and artifact avoidance. Furthermore, we introduce the efficient Multi-Scale Sampling (MSS) strategy [50] that enables the network to sample multi-scale patches of one scene. Cooperating RUL-EM and MSS can improve the detail restoration and generalization performance. Lastly, we design a Joint LLE & SR Network (JLSR), incorporating Channel Attention (CA) into the various architectures to adaptively adjust the influence of the estimated relative underexposure levels. Experimental results show that our proposed method achieves the SOTA performance on RELLISUR, LOL, and LSRW datasets with the highest quality, vivid colors, more details, and fewer artifacts in both within-dataset and cross-dataset settings. This paper can help for more robust applications of computer vision techniques in extreme environments.

References

[1] Andreas Aakerberg, Kamal Nasrollahi, and Thomas B Moeslund. Rellisur: A real low-light image super-resolution dataset. In *Thirty-fifth Conference on Neural Information*

Processing Systems Datasets and Benchmarks Track (Round 2), 2021.

[2] Andreas Aakerberg, Kamal Nasrollahi, and Thomas B. Moeslund. Relief: Joint low-light image enhancement and super-resolution with transformers. In *SCIA 2023 conference proceedings*, United States, October 2022. Springer Publishing Company.

[3] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[4] Deqiang Cheng, Liangliang Chen, Chen Lv, Lin Guo, and Qiqi Kou. Light-guided and cross-fusion u-net for anti-illumination image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(12):8436–8449, 2022.

[5] Z Cui, K Li, L Gu, S Su, P Gao, Z Jiang, Y Qiao, and T Harada. You only need 90k parameters to adapt light: A light weight transformer for image enhancement and exposure correction. In *33rd British Machine Vision Conference*, pages 21–24, 2022.

[6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

[7] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[9] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[10] Kehua Guo, Min Hu, Sheng Ren, Fangfang Li, Jian Zhang, Haifu Guo, and Xiaoyan Kui. Deep illumination-enhanced face super-resolution network for low-light images. *ACM Trans. Multimedia Comput. Commun. Appl.*, 18(3), mar 2022. ISSN 1551-6857. doi: 10.1145/3495258. URL https://doi.org/10.1145/3495258.

[11] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90:103712, 2023.

[12] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[14] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[15] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 2008.

[16] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021.

[17] Xiangtao Kong, Xina Liu, Jinjin Gu, Yu Qiao, and Chao Dong. Reflash dropout in image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6002–6012, 2022.

[18] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971.

[19] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[20] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[21] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4225–4238, 2021.

[22] Jie Liang, Hui Zeng, and Lei Zhang. Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5657–5666, 2022.

[23] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021.

[24] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[25] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[26] Ziwei Luo, Haibin Huang, Lei Yu, Youwei Li, Haoqiang Fan, and Shuaicheng Liu. Deep constrained least squares for blind image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17642–17652, 2022.

[27] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, page 4, 2018.

[28] Feifan Lv, Yinqiang Zheng, Bohan Zhang, and Feng Lu. Turn a silicon camera into an ingaas camera. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5980–5988, 2019. doi: 10.1109/CVPR.2019.00614.

[29] Feifan Lv, Yinqiang Zheng, Yicheng Li, and Feng Lu. An integrated enhancement solution for 24-hour colorful imaging. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07):11725–11732, Apr. 2020. doi: 10.1609/aaai.v34i07.6843. URL https://ojs.aaai.org/index.php/AAAI/article/view/6843.

[30] Feifan Lv, Yu Li, and Feng Lu. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129 (7):2175–2193, 2021.

[31] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5637–5646, June 2022.

[32] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[33] Hue Nguyen, Diep Tran, Khoi Nguyen, and Rang Nguyen. Psenet: Progressive self-enhancement network for unsupervised extreme-light image enhancement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1756–1765, 2023.

[34] Muhammad Tahir Rasheed and Daming Shi. Lsr: Lightening super-resolution deep network for low-light image enhancement. *Neurocomputing*, 505:263–275, 2022.

[35] Xutong Ren, Wenhan Yang, Wen-Huang Cheng, and Jiaying Liu. Lr3m: Robust low-light enhancement via low-rank regularized retinex model. *IEEE Transactions on Image Processing*, 29:5862–5876, 2020.

[36] Jianrun Shang, Xue Zhang, Guisheng Zhang, Wenhao Song, Jinyong Chen, Qilei Li, and Mingliang Gao. Gated multi-attention feedback network for medical image super-resolution. *Electronics*, 11(21):3554, 2022.

[37] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 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*, pages 1874–1883, 2016.

[38] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, Lei Zhang, Bee Lim, et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[39] Haoyuan Wang, Ke Xu, and Rynson WH Lau. Local color distributions prior for image enhancement. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII*, pages 343–359. Springer, 2022.

[40] Wenjing Wang, Chen Wei, Wenhan Yang, and Jiaying Liu. Gladnet: Low-light enhancement network with global awareness. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 751–755. IEEE, 2018.

[41] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

[42] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021.

[43] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *TIP*, 2004.

[44] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

[45] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17714–17724, 2022.

[46] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *European Conference on Computer Vision*, pages 492–511. Springer, 2020.

[47] Huan Zhang, Yihao Cao, Jianghui Cai, Xingjuan Cai, and Wensheng Zhang. Dual feature enhanced video super-resolution network based on low-light scenarios. *Signal Processing: Image Communication*, page 116984, 2023.

[48] Qiang Zhang, Xiaojian Hu, Yutao Yue, Yanbiao Gu, and Yizhou Sun. Multi-object detection at night for traffic investigations based on improved ssd framework. *Heliyon*, 8(11):e11570, 2022.

[49] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

[50] Xueting Zhang, Qi Wang, Shangdong Chen, and Xuelong Li. Multi-scale cropping mechanism for remote sensing image captioning. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 10039–10042, 2019. doi: 10.1109/IGARSS.2019.8900503.

[51] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019.

[52] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021.

[53] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.