Unveiling Details in the Dark: Simultaneous Brightening and Zooming for Low-Light Image Enhancement

Ziyu Yue[1,2], **Jiaxin Gao**[1], **Zhixun Su**[1,2*]

[1]Dalian University of Technology
[2]Key Laboratory for Computational Mathematics and Data Intelligence of Liaoning Province
{11901015, zxsu}@mail.dlut.edu.cn, jiaxinn.gao@outlook.com

Abstract

Existing super-resolution methods exhibit limitations when applied to nighttime scenes, primarily due to their lack of adaptation to low-pair dynamic range and noise-heavy dark-light images. In response, this paper introduces an innovative customized framework to simultaneously **Br**ighten and **Zo**om in low-resolution images captured in low-light conditions, dubbed **BrZoNet**. The core method begins by feeding low-light, low-resolution images and their corresponding ground truths into the Retinex-induced siamese decoupling network. This process yields distinct reflectance maps and illuminance maps, guided by supervision from the ground truth's decomposition maps. Subsequently, these reflectance and illuminance maps transition into an intricate super-resolution sub-network. This sub-network employs a meticulously designed cross-layer content-aware interactor - Illumination-aware Interaction Unit (IaIU), elegantly endowed with a gating mechanism. The IaIU facilitates meaningful feature interaction between illuminance and reflectance features while effectively reducing unwanted noise. An intricate super-resolution cage is also constructed to comprehensively integrate information, ultimately resulting in the generation of high-resolution images featuring intricate details. Thorough and diverse experiments validate the superiority of the proposed BrZoNet, surpassing contemporary cutting-edge technologies by proficiently augmenting brightness and intricately recovering complex details, showcasing advancements of 7.1% in PSNR, 2.4% in SSIM, and an impressive 36.8% in LPIPS metrics.

Introduction

The task of low-light image processing has been a hot research topic in the field of computer vision (Sharma and Tan 2021; Jin, Yang, and Tan 2022; Jin et al. 2023; Tan et al. 2021; Xie et al. 2023). It has practical applications across diverse domains, encompassing nighttime photography, nighttime detection, and segmentation, as well as security surveillance and autonomous driving (Wu and Deng 2022; Deng et al. 2022). These tasks necessitate the acquisition of bright and detailed images or video frames to ensure effective visual perception and capture rich feature information, thereby enhancing high-level perceptual performance. Hence, enhancing the brightness of low-light images along with improving their resolution to capture more details holds substantial research significance for the aforementioned task.

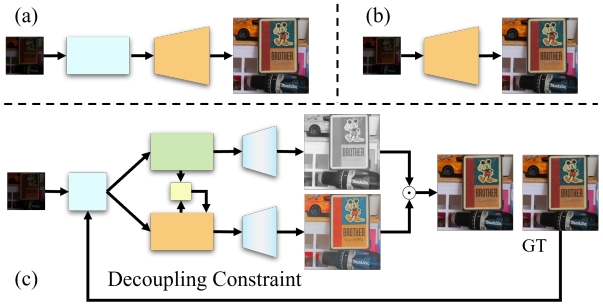

Figure 1: We explore three different approaches to address the challenging task of super-resolution in low-light scenes: (a) task cascade, (b) direct super-resolution, and (c) siamese decoupling and illumination-guided super-resolution.

To tackle this issue, two of the most intuitive approaches are considered: one involves cascading low-light enhancement with the super-resolution method, while the other entails training the existing super-resolution network model directly on the corresponding dataset. The corresponding schemes are shown in Figure 1(a) and Figure 1(b). We explored the effectiveness of these two implementation strategies by using the state-of-the-art super-resolution model (i.e., HAT (Chen et al. 2023)) as well as the low-light enhancement models (i.e., LLFormer (Wang et al. 2023) and SCI (Ma et al. 2022)) for direct and cascade training. The corresponding visual results are presented in Figure 2. Notably, we retrained both the cascaded and standalone super-resolution models using the low-light super-resolution dataset to ensure a fair comparison.

Nevertheless, both of the aforementioned approaches present sub-optimal solutions to this joint task. As evidenced in Figure 2, the results produced by cascaded models, namely SCI+HAT and LLFormer+HAT, are contingent upon the efficacy of the low-light enhancement model. However, this approach is hindered by prominent noise and a deficiency of detail. Direct utilization of existing super-resolution models produces artifacts and is not sufficiently adaptable to different levels of darkness. The root cause of these limitations lies in the inherent characteristics of existing super-resolution

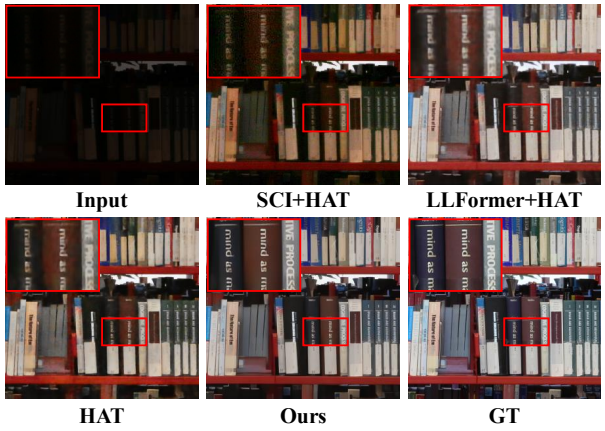

Figure 2: Visual results of various methods in low-light scenes. Cascade mode and direct super-resolution mode exhibit inferior performance. In contrast, the proposed approach achieves more natural and intricate texture details.

methods under normal lighting conditions, making it challenging to robustly learn rich features from extremely dark images to enhance super-resolution. However, in the cascade model, because of the inherent limitations of low illumination enhancement methods, the enhanced input propagated to the subsequent super-resolution network tends to suffer from color biases, noise, and even blurred details, thereby significantly impeding the overall performance of the super-resolution model. Conversely, network models without a specific design tailored to low-light images face considerable challenges when attempting to glean sufficient fine-grained information from such inputs, given their low pixel values and narrow dynamic ranges. Consequently, these models are prone to generating issues such as artifacts, noise, and color biases. Moreover, the actual images and video frames captured in real-world scenarios often exhibit varying degrees of darkness, thus further compounding the challenge of adapting to diverse levels of low-light conditions. This variability necessitates a flexible and adaptable approach to address the specific darkness level in a given scenario.

In order to address the above challenges, we have devised a novel framework and corresponding training strategy, which leverages the principles of the Retinex theory within the context of decomposition space. The corresponding scheme is shown in Figure 1(c). Specifically, a decomposition space network is introduced, trained by siamese unsupervised learning, to decompose the low-illumination image into distinct reflection and illumination maps. These individual components, representing the inherent reflectance and illuminance characteristics, are then directed into a subsequent multi-scale contextual UNet for dedicated enhancement. To further improve the quality of the reflectance maps and adapt them to varying degrees of darkness, we have introduced a cross-layer aware interactor with gating mechanisms, which serves a dual purpose of implicit denoising and illuminating the reflectance maps. By employing gating mechanisms, we can selectively regulate the information flow, enabling the reflectance maps to better acclimate to diverse low-light conditions. Furthermore, the enriched multi-scale illumination and reflection features undergo a meticulous super-resolution fusion process, meticulously designed within an intricate fusion cage. This fusion facilitates comprehensive information integration across multiple scales and ultimately leads to the reconstruction of super-resolution images. The final results are obtained by a dot product, which effectively enhances both luminance and fine-grained details, thus achieving superior visual enhancement outcomes.

The main contributions of the paper are summarized:

- In order to solve the low-light super-resolution problem, **BrZoNet** is proposed from the perspective of decomposition space for simultaneously **Br**ightening and **Zo**oming low-light low-resolution images.
- To enable the interaction of information on different scales of illuminance and reflectance features, this paper proposes a cross-layer content-aware interactive component - Illumination-aware Interaction Unit (IaIU), which enhances the adaptability of features to different darkness levels and suppresses the noise implicitly.
- In order to enhance the detail of the final reconstruction results, this paper proposes an intricate super-resolution fusion cage - Multi-stream Super-resolution Cage (MSC), which emphasizes faithful texture details by fusing reflectance and illumination features at different scales.
- Thorough and comprehensive experimentation unequivocally establishes the superiority of the proposed BrZoNet over contemporary cutting-edge technologies, as it not only enhances brightness but also adeptly restores intricate details, achieving remarkable improvements of 7.1% in PSNR, 2.4% in SSIM, and an impressive 36.8% in LPIPS metrics.

Related Work

Image Super-resolution

In recent years, remarkable advancements have been achieved in image super-resolution algorithms, driven by the rapid development of deep learning (Zhou et al. 2023a; Sun, Pan, and Tang 2022). The utilization of convolutional neural network (Ledig et al. 2017; Zamir et al. 2020; Gao et al. 2023a) and generative adversarial network (Wang et al. 2018, 2021) in addressing image super-resolution challenges has resulted in outstanding performance across various datasets. The application of transformer-based networks (Liu et al. 2021c; Liang et al. 2021; Chen et al. 2023; Gao et al. 2023b) in the realm of image super-resolution has significantly contributed to advancements in model architecture, computational efficiency, and practical application, thereby fostering progress in the research and development of super-resolution tasks. These methods promote the development of super-resolution tasks in terms of model structure, computational consumption, and application to realistic scenarios. However, these methods are not specifically designed for nighttime scenes, and their direct use can result in insufficient brightening, artifacts, color deviation, and noise amplification.

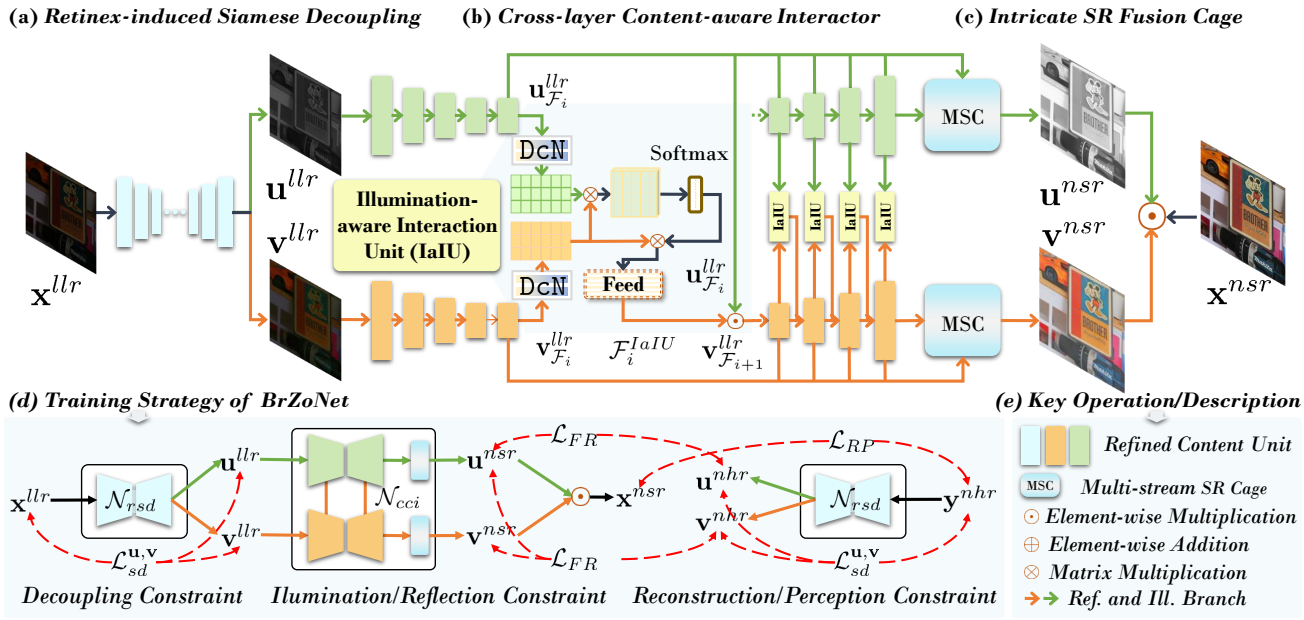

Figure 3: The pipeline of the BrZoNet. The overall network consists of three parts: Part (a) utilizes a siamese decoupling module to decompose the low-light low-resolution input. Part (b) constructs cross-layer content-aware interactor between the illumination and reflection branches, proposing an illumination-based perceptual-guided reflection for fine enhancement. Part (c) builds a multi-stream feature aggregation super-resolution module to improve high-frequency details and enlarge the resolution. (d) illustrates the proposed training strategy and the specific constraint losses (i.e., marked with red dashed arrows) introduced in each module, including decoupling constraint, illumination-reflection constraint, and reconstruction constraint, respectively.

Low-light Image Enhancement

The integration of Retinex theory (Rahman, Jobson, and Woodell 2004) into deep learning represents a predominant approach in addressing the majority of current methods (Wei et al. 2018; Wang et al. 2019; Zhang et al. 2021; Ma et al. 2023; Gao et al. 2023a; Liu et al. 2023; Li et al. 2024). Further, Retinex theory is also combined with neural architecture search (Liu, Simonyan, and Yang 2018; Liu et al. 2021b,a) and unrolling modes (Wu et al. 2022) to address low-light enhancement tasks. Besides supervised methods, there are semi-supervised and unsupervised approaches (Jiang et al. 2021; Guo et al. 2020; Ma et al. 2022; Liu et al. 2022) for low-light enhancement. Drawing inspiration from the Retinex theory principles, this paper delves into the learning process of super-resolution tasks in low-light scenarios.

Methodology

The problem of super-resolution in low-light scenarios is a complex and challenging task within a practical application context, and as such, it has not received sufficient attention for an extended period. It aims to reconstruct a super-resolution image with normal illumination $\mathbf{x}^{nsr}$ from a low-resolution image captured under low-lighting conditions $\mathbf{x}^{llr}$. In the following, we propose a specialized methodology that encompasses three essential processing steps: siamese decoupling, cross-layer interaction, and fusion reconstruction. The overall pipeline of the proposed method is illustrated in Figure 3. Firstly, leveraging the principles of Retinex theory, we establish a Retinex-induced siamese decoupling module $\mathcal{N}_{rsd}$ to derive the low-resolution illumination mapping $\mathbf{u}^{llr}$ (or $\mathbf{u}^{nhr}$) and reflection mapping $\mathbf{v}^{llr}$ (or $\mathbf{v}^{nhr}$) in terms of $\mathbf{x}^{llr}$ and its corresponding normal-light high-resolution image $\mathbf{y}^{nhr}$. This decoupling process can be formulated as

$$\mathbf{x}^{llr}/\mathbf{y}^{nhr} \xrightarrow{\mathcal{N}_{rsd}} \{\mathbf{u}^{llr}/\mathbf{u}^{nhr}, \mathbf{v}^{llr}/\mathbf{v}^{nhr}\}. \tag{1}$$

Building upon this foundation, we further devise a cross-layer content-aware interactor $\mathcal{N}_{cci}$ that facilitates feature exchange between two mapping branches. Finally, through the meticulous integration of a intricate super-resolution fusion cage and the retinex-based element-wise multiplication, we obtain the enhanced super-resolution image $\mathbf{x}^{nsr}$ with faithfully restored lighting conditions. The workflow can be succinctly formalized as follows:

$$\{\mathbf{u}^{llr}, \mathbf{v}^{llr}\} \xrightarrow{\mathcal{N}_{cci}} \{\mathbf{u}^{nsr}, \mathbf{v}^{nsr}\} \xrightarrow{\odot} \mathbf{x}^{nsr}. \tag{2}$$

Retinex-induced Siamese Decoupling

We posit that the decomposition mechanism guided by Retinex theory can be effectively extended across different resolutions. Thus, for a given data pair $(\mathbf{x}^{llr}, \mathbf{y}^{nhr})$, where $\mathbf{x}^{llr}$ represents the low-resolution input captured under weak lighting conditions and $\mathbf{y}^{nhr}$ corresponds to the high-resolution ground truth with optimal illumination, we jointly input them into a shared-parameter representation space. Specifically, we construct a parallel Retinex-induced siamese decoupling network $\mathcal{N}_{rsd}$ parameterized by θ_{rsd},

which is an improved version of the Unet-type architecture. It is worth noting that, in order to enhance the representational capacity of features, we introduce refined content reconstruction units as fundamental building blocks at each layer, as depicted in Figure 4. The entire module is learned by incorporating a decoupling constraint[1], yielding the low-resolution illumination map and reflection map. The entire process can be formalized as follows:

$$\begin{aligned}\{\mathbf{u}^{llr}, \mathbf{v}^{llr}\} &= \mathcal{N}_{rsd}(\mathbf{x}^{llr}; \theta_{rsd}),\ \mathbf{x}^{llr} = \mathbf{u}^{llr} \odot \mathbf{v}^{llr} \\ \{\mathbf{u}^{nhr}, \mathbf{v}^{nhr}\} &= \mathcal{N}_{rsd}(\mathbf{y}^{nhr}; \theta_{rsd}),\ \mathbf{y}^{nhr} = \mathbf{u}^{nhr} \odot \mathbf{v}^{nhr}.\end{aligned} \tag{3}$$

Cross-layer Content-aware Interactor

Building upon the aforementioned framework, the separated illumination mapping and reflection mapping $\{\mathbf{u}^{llr}, \mathbf{v}^{llr}\}$ are fed into two parallel Unet-style network, designed for fine-grained feature enhancement and interaction. This module is denoted as $\mathcal{N}_{cci}$ parameterized by θ_{cci}, which is formulated as $\{\mathbf{u}^{nsr}, \mathbf{v}^{nsr}\} = \mathcal{N}_{cci}(\mathbf{u}^{llr}, \mathbf{v}^{llr}; \theta_{cci})$. To elaborate further, we first extract features at n different scales from the decoder of the illumination sub-network denoted as $\{\mathbf{u}^{llr}_{\mathcal{F}_i}\}_{i=1}^{n}$. Subsequently, utilizing the designed Illumination-aware Interaction Unit (IaIU; ψ^{IaIU}), the preceding features are employed as guidance masks to fuse with the features $\{\mathbf{v}^{llr}_{\mathcal{F}_i}\}_{i=1}^{n}$ from the reflection sub-network at each layer, ultimately obtaining the guided reflection map $\mathbf{v}^{nsr}$:

$$\mathbf{v}^{llr}_{\mathcal{F}_{i+1}} = \psi^{IaIU}(\mathbf{u}^{llr}_{\mathcal{F}_i}, \mathbf{v}^{llr}_{\mathcal{F}_i}),\ i = 1, \cdots, n. \tag{4}$$

The ψ^{IaIU} is illustrated in the middle of Figure 3. Specifically, the illumination features $\mathbf{u}^{llr}_{\mathcal{F}_i}$ and reflection features $\mathbf{v}^{llr}_{\mathcal{F}_i}$ obtained from the encoder are inputted separately into the multi-dconv block with norm layers, denoted as $\texttt{DcN}$. After that, the dimensions are reshaped to obtain the illumination-aware guidance map G^{IaIU}, which represents the relationship between the illumination and reflection features, expressed as

$$G^{IaIU} = \texttt{Softmax}\big(\texttt{DcN}(\mathbf{u}^{llr}_{\mathcal{F}_i}) \otimes \texttt{DcN}(\mathbf{v}^{llr}_{\mathcal{F}_i})\big), \tag{5}$$

where $\otimes$ denotes the matrix multiplication operation. Using this guidance map as an attention map, we apply softmax normalization and modulate the transformed reflection features. At the same time, we introduce a feed forward block with stack of gated conv. layers, denoted as $\tilde{\Psi}_{Feed}$. This stack performs self-coordinated transformation and produces a reflection map with the same dimensions as the input, i.e.,

$$\mathcal{F}_i^{IaIU} = \tilde{\Psi}_{Feed}\big(\texttt{DcN}(\mathbf{v}^{llr}_{\mathcal{F}_i}) \otimes G^{IaIU}\big). \tag{6}$$

It's important to note that the illumination map contains more structural detail information. Thus, it can serve as a mask to guide the enhancement of the reflection map. Moreover, inspired by the Retinex theory, we perform element-wise multiplication between the reflection map and the illumination map in the feature level, formulated as $\mathbf{v}^{llr}_{\mathcal{F}_{i+1}} = \texttt{Up}_{\uparrow}(\mathbf{u}^{llr}_{\mathcal{F}_i} \odot \mathcal{F}_i^{IaIU})$, where $\texttt{Up}_{\uparrow}$ denotes the ConvTranspose layer for upsampling. The purpose of this step is to obtain the guided reflection feature $\mathbf{v}^{llr}_{\mathcal{F}_{i+1}}$ that are more consistent in terms of contextual content.

[1]Please refer to the self-regularized decoupling loss in Eq. (8).

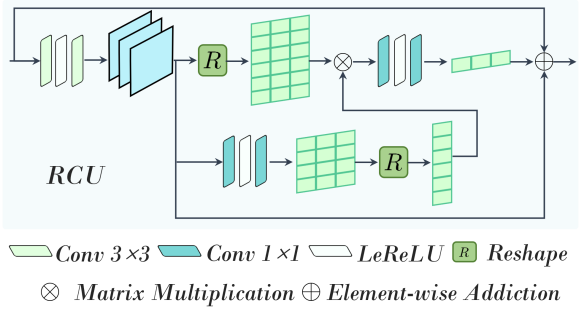

Figure 4: Illustrations of the basic RCU module.

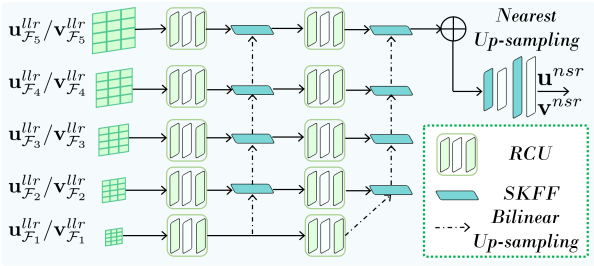

Figure 5: Architectural details of the MSC module.

Intricate Super-resolution Fusion Cage

We designed the Multi-stream Super-resolution Cage (MSC) module for the illumination and reflection branches to perform feature aggregation and amplification across multiple scales. As shown in Figure 5, for each specific indexed scale layer of the illumination and reflection features, they undergo sequential refined content unit (RCU) and selective attention mechanism (SKFF) layers. The SKFF layer is inspired by the setting of method (Zamir et al. 2020), enabling feature selection and aggregation interaction between the two scales to enhance representational capacity. By applying two sets of the same operation forms, along with residual connections, we obtain the amplified illumination and reflection maps through convolution and upsampling operations at the desired magnification ratio.

Loss Function

The following is the loss function used in this paper:

$$\mathcal{L}_{total} = \lambda_{SD} * \mathcal{L}_{SD} + \lambda_{FR} * \mathcal{L}_{FR} + \lambda_{RP} * \mathcal{L}_{RP}, \tag{7}$$

where $\mathcal{L}_{SD}$ and $\mathcal{L}_{FR}$ and $\mathcal{L}_{RP}$ are self-regularized decoupling loss, and fusion resolution loss and reconstruction perception loss respectively. λ_{SD}, λ_{FR} and λ_{RP} are the corresponding loss weights.

Self-regularized Decoupling Loss. Within the decomposition space, following the principles of Retinex, we impose constraints on the input data pairs separately, ensuring that the decomposed illumination and reflection components satisfy fundamental imaging rules. The constructed decoupling

Scale	*Metrics*	EDSR	D-DBPN	ESRGAN	RDN	RCAN	SRFBN	PAN	MSRResNet	MIRNet	SwinIR	Restormer	SRFormer	HAT	Ours
×2	PSNR↑	18.38	18.70	18.08	18.79	19.76	18.42	18.78	18.15	21.05	18.38	21.21	19.55	20.21	**22.79**
	SSIM↑	0.679	0.682	0.655	0.701	0.712	0.662	0.693	0.677	0.720	0.640	0.727	0.704	0.719	**0.745**
	LPIPS↓	0.466	0.460	0.300	0.455	0.426	0.510	0.450	0.451	0.436	0.577	0.385	0.469	0.454	**0.243**
	RMSE↓	0.125	0.120	0.135	0.120	0.110	0.125	0.119	0.128	0.095	0.125	0.095	0.110	0.103	**0.078**
	FSIM↑	0.851	0.862	0.873	0.874	0.881	0.847	0.867	0.848	0.889	0.845	0.892	0.877	0.882	**0.902**
	SRE↑	55.62	55.80	55.52	55.86	56.37	55.67	55.85	55.52	57.06	55.64	57.09	56.23	56.58	**57.90**
×4	PSNR↑	17.69	17.96	17.18	18.21	19.07	17.67	18.10	17.59	19.78	17.53	20.29	18.72	19.75	**21.41**
	SSIM↑	0.679	0.674	0.647	0.701	0.712	0.665	0.700	0.684	0.704	0.663	0.720	0.705	0.715	**0.726**
	LPIPS↓	0.623	0.575	0.471	0.584	0.550	0.640	0.559	0.581	0.599	0.688	0.492	0.613	0.561	**0.383**
	RMSE↓	0.135	0.132	0.149	0.128	0.119	0.136	0.129	0.137	0.109	0.139	0.106	0.121	0.110	**0.090**
	FSIM↑	0.832	0.848	0.858	0.866	0.874	0.836	0.859	0.841	0.878	0.840	0.885	0.869	0.873	**0.886**
	SRE↑	58.51	58.64	58.30	58.77	59.21	58.51	58.71	58.45	59.63	58.43	59.82	59.02	59.57	**60.39**

Table 1: Quantitative comparison of ×2 and ×4 tasks on the *RELLISUR* dataset. Best results are bolded. Six reference indicators, including PSNR, SSIM, LPIPS, RMSE, FSIMC and SRE are quantitatively analyzed.

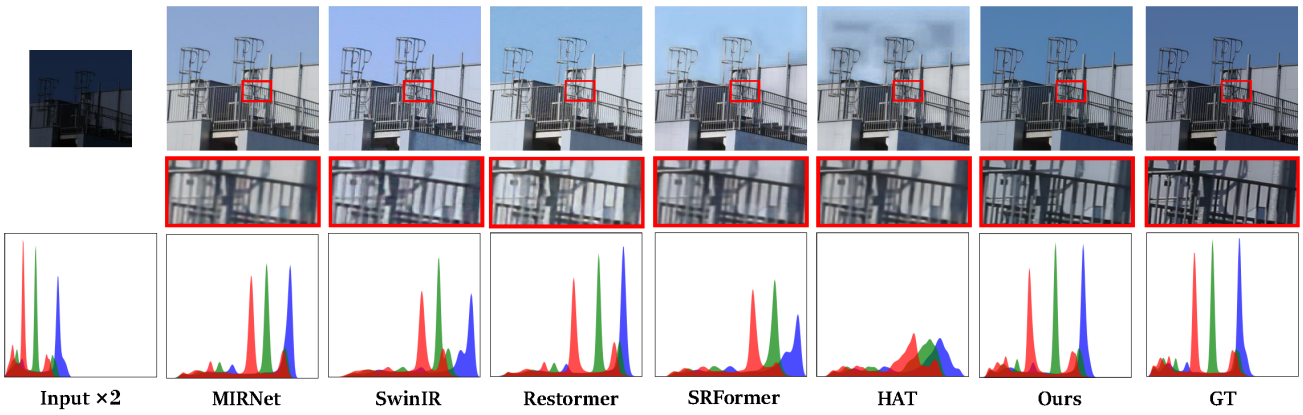

Figure 6: Qualitative comparisons of ×2 tasks on *RELLISUR* dataset. The top is the full result images, the middle are local zoom images of the red boxes, and the bottom is the statistical distributions of the RGB channels.

constraints can be represented as follows:

$$\mathcal{L}_{SD}^{\mathbf{u},\mathbf{v}} = \sum_{p\in\{llr,nhr\}} ||\mathbf{u}^p \otimes \mathbf{v}^p - \mathbf{x}^p||_1 + SATV(\mathbf{u}^p, \mathbf{v}^p), \quad (8)$$

where $SATV(\cdot,\cdot)$ denotes the structure-aware total variation loss (Wei et al. 2018).

Fusion Resolution Loss. After merging the two branches and performing super-resolution, we apply constraints separately to the resulting illumination and reflection components to ensure their consistency with the corresponding ground truth illumination and reflection content. Additionally, we introduce an illumination smoothness constraint to guarantee structural smoothness. This loss term can be represented as follows:

$$\mathcal{L}_{FR} = \sum_{q\in\{\mathbf{u},\mathbf{v}\}} ||\mathbf{q}^{nsr} - \mathbf{q}^{nhr}||_1 + SATV(\mathbf{u}^{nsr}, \mathbf{v}^{nsr}). \quad (9)$$

Reconstruction and Perception Loss. In the concluding stage of the network reconstruction, we apply the following loss terms to enforce content consistency and perceptual content consistency, expressed as

$$\mathcal{L}_{RP} = ||\mathbf{y}^{nsr} - \mathbf{y}^{nhr}||_1 + ||\psi(\mathbf{y}^{nsr}) - \psi(\mathbf{y}^{nhr})||_{Perc}, \quad (10)$$

where ψ denotes the pretrained VGG-19 network with specific network layers. By imposing content consistency and perceptual consistency losses, the network is encouraged to generate visually consistent and more realistic enhancement results.

Experiments

Experimental Setting

We use the widely recognized dataset RELLISUR[2] to train and evaluate the proposed method. The dataset comprises paired images, including low-resolution dark light/high-resolution normal light images at ×1, ×2, and ×4 resolutions. The training set consists of 3610 pairs, while the test

[2]https://vap.aau.dk/rellisur/

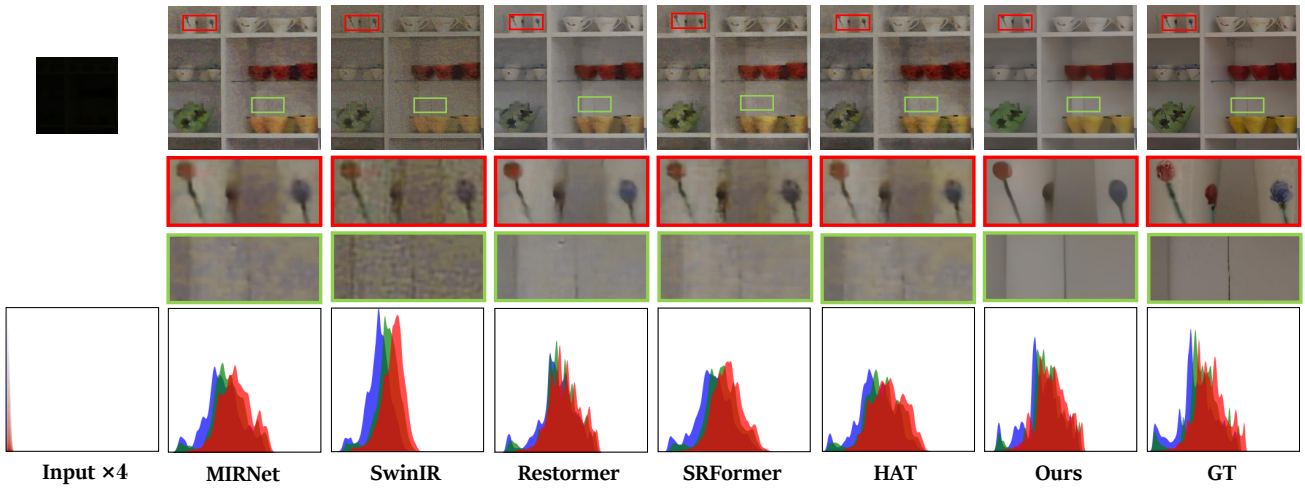

Figure 7: Qualitative comparisons of $\times\mathbf{4}$ tasks on *RELLISUR* dataset. The top is the full result images, the middle two rows are local zoom images of the red and green boxes, and the bottom is the statistical distribution of the RGB channels.

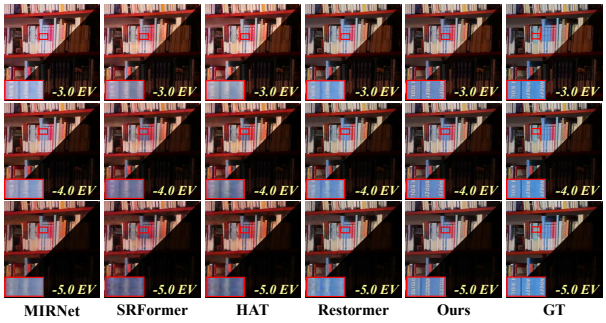

Figure 8: Enhancement results comparison for input images under different low-light levels (i.e., -3.0EV, -4.0EV and -5.0EV). Across all varying low-light levels, the proposed method excels in both detail and luminance restoration, effectively suppressing color distortion.

set includes 425 pairs of images with varying darkness levels at each resolution. We experimented with data at $\times\mathbf{2}$ and $\times\mathbf{4}$ resolution. To augment our dataset, we utilize three data augmentation techniques, namely random cropping, random rotation, and random flipping (Liu et al. 2020). The experiments were conducted using the PyTorch 2.0.1 framework on a single NVIDIA GeForce GTX 2080Ti GPU, and the optimizer used was AdamW with 15W iterations. Dynamic patch size and batch size were employed during training. The initial learning rate was set to 2×10^{-3}, and we opted for the CosineAnnealingRestartCyclicLR as the learning rate tuning method.

Experimental Evaluation

To fully verify the effectiveness of our method, we compare 13 state-of-the-art normal light super-resolution methods, including MSRFBN (Li et al. 2019), D-DBPN (Haris, Shakhnarovich, and Ukita 2018), RDN (Zhang et al. 2018c), EDSR (Lim et al. 2017), ESRGAN (Wang et al. 2018), RCAN (Zhang et al. 2018b), PAN (Zhao et al. 2020), ESRGAN (Wang et al. 2018), SwinIR (Liang et al. 2021), MIRNet (Zamir et al. 2020), Restormer (Zamir et al. 2022), SRFormer (Zhou et al. 2023b) and HAT (Chen et al. 2023). All compared methods are retrained on the RELLISUR dataset according to the official parameters. And we choose six evaluation metrics: *Peak Signal to Noise Ratio (PSNR)* (Chan and Whiteman 1983) and *Structural Similarity Index Measure (SSIM)* (Wang et al. 2004), *Learned Perceptual Image Patch Similarity (LPIPS)* (Zhang et al. 2018a), *Root Mean Square Error (RMSE)* (Ferrari et al. 2018), *Feature-based Similarity Index (FSIMC)* (Zhang et al. 2011) and *Signal to Reconstruction Error Ratio (SRE)* (Lanaras et al. 2018).

Quantitative Evaluation. As shown in Table 1, our BrZoNet outperforms other state-of-the-art techniques, ranking first in all six evaluation metrics. This indicates that the results recovered by the proposed method consistently exhibit superior performance in terms of illumination, color, and texture details compared to alternative approaches. Particularly for the $\times\mathbf{2}$ upscaling task, PSNR and RMSE show improvements of 7.4% and 17.8%, respectively. For the $\times\mathbf{4}$ upscaling task, the improvements are 5.5% for PSNR and 15.1% for RMSE.

Qualitative Evaluation. The qualitative comparison visualization results are presented in Figure 6 and Figure 7. In comparison to other state-of-the-art methods, the proposed method excels not only enhances illumination but also exhibits superior restoration of texture details. It effectively suppresses the generation of artifacts and noise issues. Furthermore, the statistical distribution of RGB values indicates that, our method produces results with color distributions closer to those of the reference image, as compared to the contrastive methods.

To demonstrate that our proposed method can adaptively enhance and super-resolve low-light low-resolution images of different darkness levels, we show the comparison results in Figure 8. Compared to these above methods, the proposed

Scale	Method			Metric		
	IllNet	IaIU	MSC	PSNR↑	SSIM↑	LPIPS↓
×2	✗	✗	✗	21.91	0.725	0.270
	✓	✗	✗	$22.16_{\uparrow 0.25}$	$0.734_{\uparrow 0.009}$	$0.261_{\downarrow 0.009}$
	✓	✓	✗	$22.31_{\uparrow 0.40}$	$0.733_{\uparrow 0.008}$	$0.255_{\downarrow 0.015}$
	✓	✓	✓	$\mathbf{22.79}_{\uparrow 0.88}$	$\mathbf{0.745}_{\uparrow 0.020}$	$\mathbf{0.243}_{\downarrow 0.027}$
×4	✗	✗	✗	20.78	0.720	0.387
	✓	✗	✗	$20.90_{\uparrow 0.12}$	$0.723_{\uparrow 0.003}$	$0.392_{\downarrow 0.005}$
	✓	✓	✗	$21.16_{\uparrow 0.38}$	$0.724_{\uparrow 0.004}$	$0.389_{\downarrow 0.002}$
	✓	✓	✓	$\mathbf{21.42}_{\uparrow 0.64}$	$\mathbf{0.726}_{\uparrow 0.006}$	$\mathbf{0.383}_{\downarrow 0.004}$

Table 2: Ablation study regarding the proposed network components (i.e., IllNet, IaIU and MSC) on ×2 and ×4 task. Each numerical subscript in the bottom right corner is marked, indicating the difference from the baseline method.

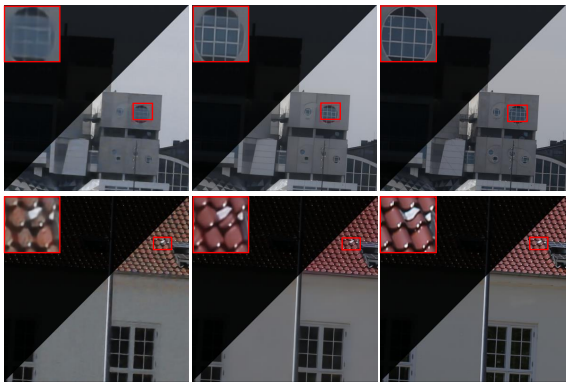

Figure 9: Illustrating ablation analysis of the perceptual loss.

method accurately recover the brightness and detail information of images with different levels of darkness, effectively avoiding color deviation. This confirms the remarkable adaptability of our method to scenes with varying levels of darkness and challenging lighting conditions.

Ablation Study

Effects of Decomposition Space. To validate the effectiveness of addressing low-light image super-resolution from the perspective of decomposition space, we conducted ablation experiments by training models with the retained illuminance sub-network (i.e., IllNet) and models without the illuminance sub-network. The results in the second and sixth rows of Table 2 demonstrate a significant performance improvement after incorporating the IllNet, thereby confirming the efficacy from the perspective of decomposition space.

Effects of IaIU. To demonstrate the effectiveness of the IaIU, we excluded the MSC while maintaining the IllNet. The results in the third and seventh rows of Table 2 show that the cross-layer content-aware interactor with IaIU improves model performance compared to results without its inclusion.

Effects of MSC. To confirm the effectiveness of the MSC, we compare the quantitative results of the complete network with those from the third and seventh rows in Table 2. The inclusion of the super-resolution fusion cage with MSC improves model performance, leading to more detailed reconstruction results.

Method	PSNR↑	SSIM↑	LPIPS↓
RSD w/o $\mathbf{y}^{nhr}$ (×2)	22.46	0.738	0.252
Ours (×2)	$\mathbf{22.79}_{\uparrow 0.33}$	$\mathbf{0.745}_{\uparrow 0.007}$	$\mathbf{0.243}_{\downarrow 0.009}$
RSD w/o $\mathbf{y}^{nhr}$ (×4)	21.25	0.724	0.378
Ours (×4)	$\mathbf{21.42}_{\uparrow 0.17}$	$\mathbf{0.726}_{\uparrow 0.012}$	$\mathbf{0.383}_{\downarrow 0.005}$

Table 3: Ablation study regarding the siamese decoupling for $\mathbf{y}^{nhr}$ on ×2 and ×4 task. The numbers annotated at the bottom right corner indicate the differences.

$\mathcal{L}_{SD}$ (w/o SATV)	$\mathcal{L}_{FR}$ (w/o SATV)	$\mathcal{L}_{RP}$	PSNR↑
0.5	0.5	0.5	22.33
1	0.1	0.1	22.57
0.8	0.5	0.5	22.74
1	0.5	0.5	**22.79**
w/o skip connection			22.17

Table 4: Analysis of loss weights and skip connections (Note that the SATV loss is by default set with a weight of 1).

Effects of Siamese Decoupling. To showcase the enhanced performance of the decomposition subnetwork with the siamese decoupling training strategy in our method, we excluded the decomposition of the normal light high-resolution images (i.e., RSD w/o $\mathbf{y}^{nhr}$) and the corresponding loss function during training. The resulting quantitative improvements are evident in Table 3, affirming the effectiveness of the siamese decoupling strategy.

Effects of Loss Weights and Perceptual Loss. To validate the impact of perceptual loss, we present the results of a qualitative comparison in Figure 9. It is evident that the perceptual loss effectively enhances detail recovery and mitigates color deviation. Table 4 shows the impact of different loss weights and the skip connections of Unet on the final performance. It can be observed that the decoupling loss accounts for the largest proportion, and removing the skip connections results in a 0.62dB decrease in PSNR performance.

Conclusion

This study introduces the BrZoNet framework to address limitations in existing super-resolution methods for nighttime scenes. It enhances adaptability to low-pair dynamic range and noise-laden dark-light images by combining siamese decomposition and a super-resolution network, featuring the IaIU for effective feature interaction and noise reduction. BrZoNet achieves high-resolution images through comprehensive information integration in a super-resolution cage. Extensive experiments demonstrate its superiority over state-of-the-art techniques, with significant improvements in brightness and detail recovery.

Acknowledgements

This paper was supported by the National Key R&D Program of China (No. 2018AAA0100300) and National Natural Science Foundation of China (No. 61976041).

Ziyu Yue and Jiaxin Gao equally contributed to this work. Professor Zhixun Su is the corresponding author of this paper (The corresponding author is marked with $*$), thank him for his guidance.

References

Chan, L. C.; and Whiteman, P. 1983. Hardware-constrained hybrid coding of video imagery. *IEEE Transactions on Aerospace and Electronic Systems*, (1): 71–84.

Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; and Dong, C. 2023. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22367–22377.

Deng, X.; Wang, P.; Lian, X.; and Newsam, S. 2022. NightLab: A dual-level architecture with hardness detection for segmentation at night. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16938–16948.

Ferrari, V.; Hebert, M.; Sminchisescu, C.; and Weiss, Y. 2018. *Computer Vision–ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part V*, volume 11209. Springer.

Gao, J.; Liu, X.; Liu, R.; and Fan, X. 2023a. Learning adaptive hyper-guidance via proxy-based bilevel optimization for image enhancement. *The Visual Computer*, 39(4): 1471–1484.

Gao, J.; Yue, Z.; Liu, Y.; Xie, S.; Fan, X.; and Liu, R. 2023b. Diving into Darkness: A Dual-Modulated Framework for High-Fidelity Super-Resolution in Ultra-Dark Environments. *arXiv preprint arXiv:2309.05267*.

Guo, C.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1780–1789.

Haris, M.; Shakhnarovich, G.; and Ukita, N. 2018. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1664–1673.

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

Jin, Y.; Lin, B.; Yan, W.; Yuan, Y.; Ye, W.; and Tan, R. T. 2023. Enhancing visibility in nighttime haze images using guided apsf and gradient adaptive convolution. In *Proceedings of the 31st ACM International Conference on Multimedia*, 2446–2457.

Jin, Y.; Yang, W.; and Tan, R. T. 2022. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *European Conference on Computer Vision*, 404–421.

Lanaras, C.; Bioucas-Dias, J.; Galliani, S.; Baltsavias, E.; and Schindler, K. 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146: 305–319.

Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681–4690.

Li, Y.; Xu, R.; Niu, Y.; Guo, W.; and Zhao, T. 2024. Perceptual Decoupling with Heterogeneous Auxiliary Tasks for Joint Low-Light Image Enhancement and Deblurring. *IEEE Transactions on Multimedia*.

Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; and Wu, W. 2019. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3867–3876.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1833–1844.

Lim, B.; Son, S.; Kim, H.; Nah, S.; and Mu Lee, K. 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 136–144.

Liu, H.; Simonyan, K.; and Yang, Y. 2018. Darts: Differentiable architecture search. *arXiv preprint arXiv:1806.09055*.

Liu, M.; Yan, P.; Lian, C.; and Cao, X. 2020. *Machine Learning in Medical Imaging: 11th International Workshop, MLMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings*, volume 12436. Springer Nature.

Liu, R.; Gao, J.; Liu, X.; and Fan, X. 2023. Learning with Constraint Learning: New Perspective, Solution Strategy and Various Applications. *arXiv preprint arXiv:2307.15257*.

Liu, R.; Gao, J.; Zhang, J.; Meng, D.; and Lin, Z. 2021a. Investigating bi-level optimization for learning and vision from a unified perspective: A survey and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 10045–10067.

Liu, R.; Ma, L.; Ma, T.; Fan, X.; and Luo, Z. 2022. Learning with nested scene modeling and cooperative architecture search for low-light vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5): 5953–5969.

Liu, R.; Ma, L.; Zhang, J.; Fan, X.; and Luo, Z. 2021b. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10561–10570.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021c. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022.

Ma, L.; Jin, D.; An, N.; Liu, J.; Fan, X.; Luo, Z.; and Liu, R. 2023. Bilevel fast scene adaptation for low-light image enhancement. *International Journal of Computer Vision*, 1–19.

Ma, L.; Ma, T.; Liu, R.; Fan, X.; and Luo, Z. 2022. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5637–5646.

Rahman, Z.-u.; Jobson, D. J.; and Woodell, G. A. 2004. Retinex processing for automatic image enhancement. *Journal of Electronic imaging*, 13(1): 100–110.

Sharma, A.; and Tan, R. T. 2021. Nighttime visibility enhancement by increasing the dynamic range and suppression of light effects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11977–11986.

Sun, L.; Pan, J.; and Tang, J. 2022. Shufflemixer: An efficient convnet for image super-resolution. *Advances in Neural Information Processing Systems*, 35: 17314–17326.

Tan, X.; Xu, K.; Cao, Y.; Zhang, Y.; Ma, L.; and Lau, R. W. 2021. Night-time scene parsing with a large real dataset. *IEEE Transactions on Image Processing*, 30: 9085–9098.

Wang, R.; Zhang, Q.; Fu, C.-W.; Shen, X.; Zheng, W.-S.; and Jia, J. 2019. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6849–6857.

Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; and Lu, T. 2023. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 2654–2662.

Wang, X.; Xie, L.; Dong, C.; and Shan, Y. 2021. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1905–1914.

Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; and Change Loy, C. 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, 0–0.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

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

Wu, A.; and Deng, C. 2022. Single-domain generalized object detection in urban scene via cyclic-disentangled self-distillation. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 847–856.

Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; and Jiang, J. 2022. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5901–5910.

Xie, Z.; Wang, S.; Xu, K.; Zhang, Z.; Tan, X.; Xie, Y.; and Ma, L. 2023. Boosting Night-time Scene Parsing with Learnable Frequency. *IEEE Transactions on Image Processing*.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2020. Learning enriched features for real image restoration and enhancement. In *European Conference on Computer Vision*, 492–511. Springer.

Zhang, L.; Zhang, L.; Mou, X.; and Zhang, D. 2011. FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8): 2378–2386.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018a. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.

Zhang, Y.; Guo, X.; Ma, J.; Liu, W.; and Zhang, J. 2021. Beyond brightening low-light images. *International Journal of Computer Vision*, 129: 1013–1037.

Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; and Fu, Y. 2018b. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, 286–301.

Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; and Fu, Y. 2018c. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2472–2481.

Zhao, H.; Kong, X.; He, J.; Qiao, Y.; and Dong, C. 2020. Efficient image super-resolution using pixel attention. In *European Conference on Computer Vision*, 56–72. Springer.

Zhou, M.; Yan, K.; Pan, J.; Ren, W.; Xie, Q.; and Cao, X. 2023a. Memory-augmented deep unfolding network for guided image super-resolution. *International Journal of Computer Vision*, 131(1): 215–242.

Zhou, Y.; Li, Z.; Guo, C.-L.; Bai, S.; Cheng, M.-M.; and Hou, Q. 2023b. SRFormer: Permuted Self-Attention for Single Image Super-Resolution. *arXiv preprint arXiv:2303.09735*.