SNOC: Subtle Nested Objective Configuration for Joint Ultra-Low-Light Enhancement and Super-Resolution

Jiaxin Gao
Department of Data Science and Artificial Intelligence
The Hong Kong Polytechnic University
Hong Kong, China
jiaxinn.gao@outlook.com

Yaohua Liu*
School of Computing and Data Science
The University of Hong Kong
Hong Kong, China
liuyaohua.918@gmail.com

Danchen Cui
College of Computer Science and Technology
Taiyuan University of Technology
Taiyuan, China
2023008025@link.tyut.edu.cn

Zhihui Zhao
College of Computer Science and Technology
Taiyuan University of Technology
Taiyuan, China
Shanxi Key Laboratory of Industrial Internet Security
Taiyuan, China
zhaozhihui@tyut.edu.cn

Abstract

Ultra-low-light image restoration remains challenging in part because static optimization objectives are often brittle across the diverse degradation patterns induced by extreme darkness. Existing joint enhancement and super-resolution paradigms predominantly rely on manually specified objective trade-offs, leading to scene-dependent failures such as color bias, exposure inconsistency, and artifact propagation. To address this, we present **SNOC** (**S**ubtle **N**ested **O**bjective **C**onfiguration), a unified framework that integrates a subtle rectification architecture with adaptive objective-portfolio configuration. Architecturally, SNOC employs three refinement middlewares, i.e., Semantic Illumination Cross Calibration (SICC), Exposure-aware Rectification Unit (EaRU), and Grid-aware Dynamic Up-sampler (GaDU), to coordinate illumination-aware, exposure-aware, and detail-oriented representations for faithful recovery. Beyond architectural design, SNOC introduces a compact *nested objective configuration* mechanism that adaptively updates learnable coefficients over a comprehensive objective portfolio, thereby aligning module-specific restoration targets with global image quality and avoiding labor-intensive manual objective tuning. Extensive experiments on real-world benchmarks and our newly synthesized LLLR-NLHR dataset verify that SNOC consistently achieves superior perceptual quality and reconstruction fidelity.

CCS Concepts

• **Computing methodologies → Reconstruction**.

Keywords

Ultra-low-light imaging, Joint enhancement and super-resolution, Nested learning, Adaptive objective configuration

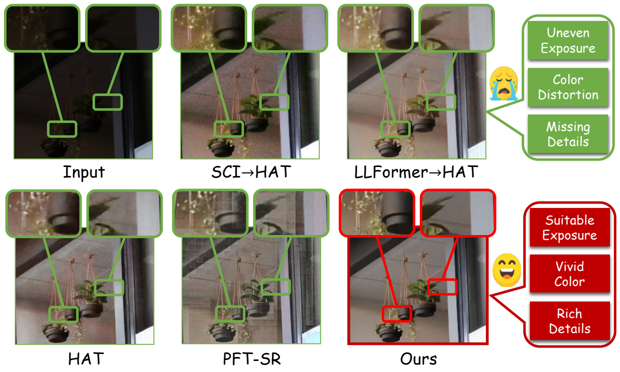

Figure 1: Comparison of joint ultra-low-light enhancement and super-resolution (×2). SCI [41] and LLFormer [55] denote LLE methods; HAT [5] and PFT-SR [39] denote NSR methods.

ACM Reference Format:
Jiaxin Gao, Yaohua Liu, Danchen Cui, and Zhihui Zhao. 2026. SNOC: Subtle Nested Objective Configuration for Joint Ultra-Low-Light Enhancement and Super-Resolution. In *International Conference on Multimedia Retrieval (ICMR '26), June 16–19, 2026, Amsterdam, Netherlands.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3805622.3810773

1 Introduction

Ultra-low-light image restoration remains challenging because extreme darkness severely degrades illumination, color fidelity, and structural details, while effective recovery is highly sensitive to objective coordination during training [7, 45, 77]. In real-world ultra-low-light scenes, non-uniform exposure, low signal-to-noise ratios, sensor noise, and detail attenuation vary substantially across images, making static objective designs difficult to generalize. Such scenarios frequently arise in practical applications such as nighttime surveillance [29, 56], autonomous driving in dark environments [50, 53], and detection and segmentation tasks [11, 25, 51]. Beyond visibility degradation, ultra-low-light images also suffer from significant resolution loss, making it more difficult to jointly recover natural brightness and high-frequency details [18, 24, 70].

---

[*] Corresponding author.

ICMR '26, Amsterdam, Netherlands

ACM ISBN 979-8-4007-2617-0/26/06
https://doi.org/10.1145/3805622.3810773

Most existing super-resolution (SR) methods assume well-lit inputs, while LLE methods mainly improve visibility without explicitly addressing resolution recovery [20, 21]. Consequently, directly applying normal-light SR (NSR) to ultra-low-light images often amplifies noise and artifacts, whereas cascaded LLE→NSR pipelines tend to cause color inconsistency, texture degradation, and pseudo-artifacts. Under ultra-low-light conditions, the challenge of joint enhancement and super-resolution lies in both *how* restoration features are coordinated and *how* restoration objectives are configured. Illumination-aware cues, exposure-sensitive responses, and high-frequency detail reconstruction interact strongly during ultra-low-light restoration, calling for dedicated architectural mechanisms for cross-representation guidance, exposure-aware modulation, and content-adaptive upsampling. Meanwhile, the relative importance of brightness enhancement, exposure consistency, detail preservation, and artifact suppression varies substantially across scenes, rendering fixed objective aggregation inadequate. This motivates a unified framework that jointly addresses subtle restoration architecture and adaptive objective-portfolio configuration.

Here we present **SNOC** (**S**ubtle **N**ested **O**bjective **C**onfiguration), a unified framework for joint ultra-low-light enhancement and super-resolution. As illustrated in Fig. 1, both standalone SR models and cascaded pipelines often struggle under extreme conditions where illumination correction and detail magnification interfere. SNOC addresses this through two complementary aspects. First, we employ a **subtle rectification architecture** featuring three refinement middlewares, i.e., **Semantic Illumination Cross Calibration (SICC)**, **Exposure-aware Rectification Unit (EaRU)**, and **Grid-aware Dynamic Up-sampler (GaDU)**. These modules provide fine-grained solutions to physical degradations, specifically rectifying color distortions via cross-covariance, alleviating regional exposure non-uniformity via piecewise-affine transformations, and suppressing upsampling artifacts through learned coordinate-shifting. Second, SNOC introduces a **nested objective configuration** strategy that adaptively navigates a comprehensive objective portfolio. This formulation turns objective coordination into a learnable component and reduces manual tuning efforts. The main contributions are summarized as follows:

- We present **SNOC**, a unified paradigm establishing **adaptive objective configuration** via compact nested objective learning, which navigates an objective portfolio to align restoration targets with quality and overcome.
- Our subtle architecture integrates SICC, EaRU, and GaDU to respectively resolve color distortions, regional exposure non-uniformity, and upsampling artifacts via cross-covariance, piecewise-affine, and coordinate-shifting modulations.
- Real-world evaluations and our LLLR-NLHR dataset (simulating RAW noise) verify SNOC's superior perceptual quality and reconstruction fidelity over state-of-the-arts.

2 Related Work

Ultra-Low-Light Image Restoration. Traditional research primarily addresses isolated degradations. Low-light image enhancement (LLE) focuses on visibility and plausible illumination via pixel-wise curve mappings or transformer-based modeling [23, 30, 37, 55, 69, 76]. Conversely, normal-light SR targets high-frequency detail recovery through residual learning, feedback networks, and advanced transformer interactions [4, 5, 8, 9, 64, 66, 72, 78]. Recent SR advances have expanded receptive fields [6, 39, 75] or utilized diffusion-based generative priors [12, 48, 57] to restore sophisticated textures. However, these methods frequently struggle with ultra-low-light joint restoration (LLISR) tasks due to the inherent interference between illumination correction and magnification-induced artifacts. Recent unified frameworks have addressed this through end-to-end mapping [52, 62, 65, 77], joint scaling strategies [1, 68], and conditional flows [71]. These approaches have demonstrated efficacy even in specialized domains such as image restoration [21, 49, 60, 61, 63], yet they do not explicitly address localized exposure variation and content-adaptive upsampling in a unified manner [19]. Such paradigms often lack the requisite sensitivity to rectify localized physical degradations, notably regional exposure non-uniformity and grid-based pseudo-artifacts. In contrast, SNOC introduces a *subtle* rectification scheme, instantiated by SICC, EaRU, and GaDU, to support cross-representation guidance, exposure-sensitive modulation, and content-adaptive upsampling in a unified framework.

Nested Optimization and Hierarchical Training Paradigms. The efficacy of joint restoration is heavily dictated by the balance of optimization objectives. Bi-level optimization has been extensively utilized to automate hyper-parameter tuning, meta-learning, and adversarial attack [15, 16, 33, 36, 38, 42, 46, 67]. Recent studies have also begun to revisit the interaction between architecture design and optimization strategy under broader nested learning perspectives [2]. Building on these foundations, hierarchical models have demonstrated remarkable potential in modeling complex multi-task synergies. In the context of image restoration and medical analysis, multi-level structures have been developed to synchronize illumination correction [17, 44], deformable registration [13, 35], and joint medical segmentation [14]. While these hierarchical approaches automate task-level coordination, they mainly emphasize variable-level or globally coupled optimization, with less focus on compact module-specific objective-portfolio adaptation for restoration. Our work advances this paradigm by introducing a compact *nested objective learning* mechanism. Instead of purely synergizing variables, SNOC focuses on the **adaptive configuration of a comprehensive objective portfolio**, ensuring that module-specific restoration targets are dynamically aligned with global image quality.

3 Subtle Rectification Architecture

3.1 Restoration Workflow

Given an ultra-low-light input $\mathbf{x}^{ll}$, SNOC aims to reconstruct a high-resolution image $\mathbf{y}^{ns}$ with natural illumination and faithful details, approximating the normal-light ground truth $\mathbf{y}^{nh}$. This task is formulated as a coupled restoration problem: exposure and color corrections are inherently amplified during upsampling, while noise suppression often erases high-frequency textures. SNOC departs from cascaded stages by learning an end-to-end mapping through a unified, rectification-oriented architecture.

As illustrated in Fig. 2(a), the framework facilitates the interaction between illumination-aware and structural-contextual representations. SNOC utilizes three *subtle rectification middlewares*

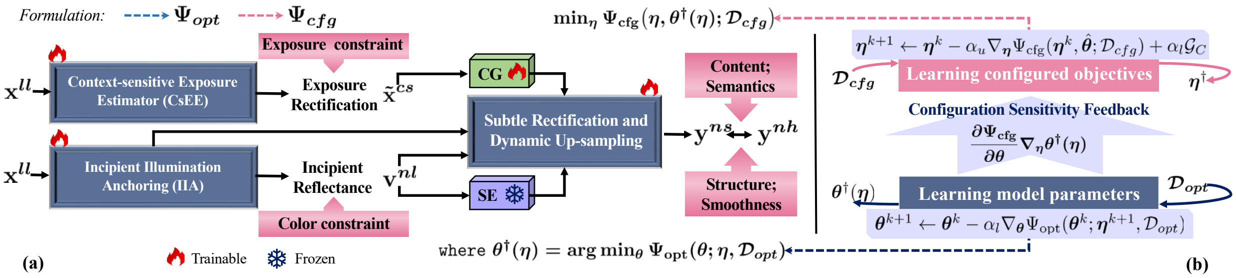

Figure 2: SNOC workflow: (a) *subtle* rectification scheme via SICC, EaRU, and GaDU; (b) nested objective-portfolio configuration.

to synchronize visibility recovery with resolution magnification. Crucially, we introduce a nested objective configuration mechanism to handle the intertwined optimization. Within this nested logic, an objective evaluator adaptively configures the weights of a comprehensive objective portfolio, while the network parameters are optimized under these scene-responsive constraints. This configuration-driven approach stabilizes the joint restoration trajectory and eliminates the reliance on trial-and-error objective tuning.

3.2 Subtle Rectification Architecture

As illustrated in Fig. 3, SNOC is built upon a suite of *rectification middlewares* designed to synchronize illumination recovery and resolution magnification within a unified mapping. Rather than treating these as independent processes, our design utilizes these components as neural mediators to resolve the physical interferences inherent in coupled restoration tasks.

Incipient Illumination Anchoring. To establish a physically consistent baseline [28], we utilize an anchoring component $\mathcal{M}_{I_2}$ based on [17] to capture scene-dependent illumination characteristics. It generates a dynamic-range-constrained illumination map $\mathbf{u}^{ll}$ through a learned mapping ψ_I:

$$\mathbf{u}^{ll} := \gamma \otimes \mathbf{x}^{ll} + (1 - \gamma) \otimes I, \tag{1}$$

where γ denotes the learned interpolation factors. The resulting reflectance representation, $\mathbf{v}^{nl} = \mathbf{x}^{ll} \oslash \mathbf{u}^{ll}$, provides a normalized feature space that suppresses the interference of extreme darkness and facilitates subsequent fine-grained rectification.

Semantic Illumination Cross Calibration. To calibrate illumination-aware representations, SNOC combines multi-scale illumination features from the decoder of $\mathcal{M}_{I_2}$, $[\tilde{u}_f^0, \cdots, \tilde{u}_f^J]$, with semantic features extracted from the reflectance representation $\mathbf{v}^{nl}$ by a frozen semantic extractor $\mathcal{M}_{SE}$ with parameters Θ_{SE}^* [59]:

$$\mathcal{M}_{SE}[\mathcal{V}_j^\uparrow](\mathbf{v}^{nl}|\Theta_{SE}^*) = [\tilde{s}_f^0, \cdots, \tilde{s}_f^j, \cdots, \tilde{s}_f^J], \tag{2}$$

where $\mathcal{V}_j^\uparrow$ extracts the j-th semantic feature and aligns it with $\tilde{u}_f^j$.

Inspired by recent self-attention architectures [31, 72], we then introduce Semantic Illumination Cross Calibration (SICC), a channel-wise attention module that calibrates reflectance features through cross-covariance interactions between illumination and semantic cues. At each scale j, the resulting channel responses are aggregated into a unified attention map ω_{IS}^j, which modulates reflectance features to produce the refined representation $\tilde{v}_f^j$:

$$\omega_{IS}^j = \mathsf{SM}(\omega_I^j \oplus \omega_S^j), \quad \tilde{v}_f^j = \mathsf{FFN}\Big(\mathsf{Conv}_{1\times1}\Big(\phi_R(v_b^j) \otimes \omega_{IS}^j\Big) \oplus v_f^j\Big), \tag{3}$$

where SM denotes the softmax function and $\phi_R(\cdot)$ is the reshape operator. SICC thereby enables context-aware calibration for subsequent high-fidelity reconstruction.

Context-sensitive Exposure Estimator. Due to uneven exposure in imaging equipment, overexposed or underexposed areas often occur in captured low-light images, meaning that some regions may be too bright while others are too dark [40]. To rectify the underexposed blocks hidden in the low-light image, we introduce the exposure-consistency prior loss as supervision[1], to generate the exposure guidance, i.e., a context-sensitive exposure mapping denoted as $\tilde{\mathbf{x}}^{cs} \in \mathbb{R}^{C\times H\times W}$, with C set to 64. This exposure-aware mapping is then input into the Coding Generator (CG) to produce a one-dimensional exposure code $\mathcal{V}^e \in \mathbb{R}^C$:

$$\tilde{\mathbf{x}}^{cs} = \mathcal{M}_{EE}(\mathbf{x}^{ll}|\Theta_{EE}), \ \mathcal{V}^e = \phi_{CG}(\tilde{\mathbf{x}}^{cs}), \tag{4}$$

where $\mathcal{M}_{EE}$ is a U-Net network mirroring the structure of $\mathcal{M}_{I_2}$, with parameters Θ_{EE}. ϕ_{CG}, as shown in Fig. 3, comprises a 1×1 convolution followed by global average pooling.

Exposure-aware Rectification Unit. EARU replaces handcrafted curves with a *learnable piecewise-affine* rectification: from an exposure code (see Fig. 4(a)), four small FC heads predict $(\tilde{\mu}_1, \tilde{\mu}_2, \tilde{l}, \tilde{\kappa})$. Given features $\tilde{v}_f^j$, we apply a self-correcting intensity mapping

$$\mathcal{G}(\tilde{v}_f^j) = \tilde{\mu}(\tilde{v}_f^j) \odot (\tilde{v}_f^j \ominus \tilde{l}) \oplus \tilde{\kappa}, \quad \tilde{\mu}(\tilde{v}_f^j) = \begin{cases} \tilde{\mu}_1, & \tilde{v}_f^j \le \tilde{l}, \\ \tilde{\mu}_2, & \tilde{v}_f^j > \tilde{l}, \end{cases} \tag{5}$$

where $\odot$ denotes element-wise multiplication. This compact parameterization effectively emulates non-linear exposure curves, enabling adaptive, region-wise rectification.

Grid-aware Dynamic Up-sampler. To suppress magnification-induced artifacts and preserve high-frequency structural integrity, we introduce **GaDU** as the terminal rectification middleware. Rather than utilizing static interpolation, GaDU implements a 'coordinate-aware warping mechanism' that adaptively refines the sampling process. As illustrated in Fig. 4(b), for an input feature map $V_f^{in} \in \mathbb{R}^{C\times H\times W}$, a linear rectification layer estimates the sampling offsets $\mathcal{O} \in \mathbb{R}^{2s^2\times H\times W}$, which capture localized geometric deformations. These offsets are integrated with the original sampling grid $\mathcal{CG}$ to produce the rectified coordinates $C_s = \mathcal{CG} + \mathcal{O}$. By applying a

---

[1] Please refer to the exposure-consistency loss in Eq. (8).

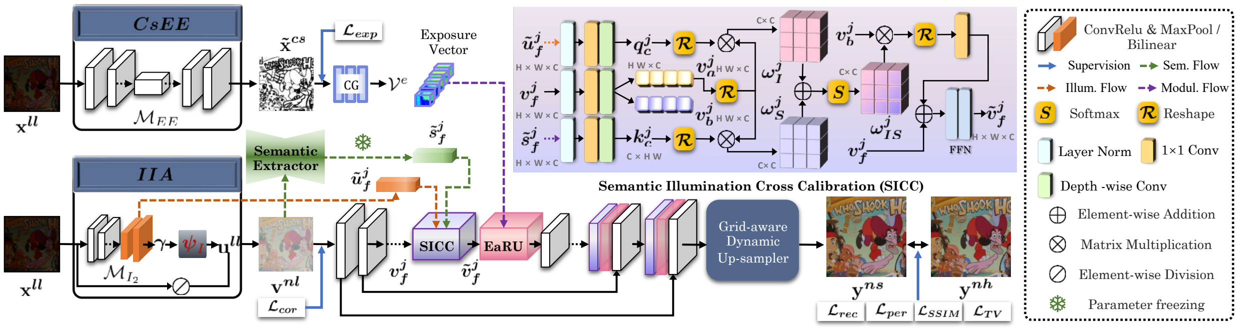

Figure 3: SNOC architecture: subtle rectification middlewares with adaptive nested objective-portfolio learning.

grid-sampling operation, GaDU generates the final normal-light, super-resolution output:

$$\mathbf{y}^{ns} = \texttt{GridSample}(V_f^{in}, \mathcal{C}_s), \tag{6}$$

where $\mathbf{y}^{ns} \in \mathbb{R}^{3\times sH\times sW}$. This learned coordinate-shifting ensures that pixel-wise intensities are sampled from the most informative regions, effectively rectifying grid-based pseudo-artifacts and achieving faithful structural recovery in a single feed-forward pass.

4 Nested Objective Configuration Strategy

This section introduces the *objective portfolio* of SNOC and describes how adaptive objective configuration is instantiated during training. Rather than relying on fixed objective aggregation, SNOC treats restoration targets as a configurable portfolio whose coefficients are updated jointly with network optimization.

4.1 Objective Portfolio Construction

As depicted in Fig. 3, SNOC constructs an objective portfolio containing six candidate objectives that cover color fidelity, exposure consistency, perceptual realism, structural preservation, and smoothness regularization. This portfolio acts as a dynamic constraint set for the rectification middlewares of SNOC, allowing different restoration aspects to be emphasized adaptively under diverse ultra-low-light conditions.

Self-regularized color loss. To ensure that the colors of generated reflection images and natural images exhibit statistically consistent distributions, we use $\mathcal{L}_{cor}$ by constraining the means and variances of both distributions:

$$\mathcal{L}_{cor}(\mathbf{v}^{nl}) = e^{|\bar{\mathbf{v}}_c^{nl} - \mu_c - \sigma_c|} - 1,\ \ c \in \{R, G, B\}. \tag{7}$$

Here, $(\boldsymbol{\mu}_c, \boldsymbol{\sigma}_c)$ denote the predefined channel-wise statistics of natural images estimated from ImageNet [10, 17], and $\bar{\mathbf{v}}_c^{nl}$ represents the channel-wise mean of $\mathbf{v}^{nl}$.

Exposure consistency loss. We use $\mathcal{L}_{exp}$ with the L_2 metric to measure the prediction error between the generated $\tilde{\mathbf{x}}^{cs}$ and the exposure prior map $\mathbf{x}^{exp}$:

$$\begin{aligned}&\mathcal{L}_{exp}(\tilde{\mathbf{x}}^{cs}) = \|\tilde{\mathbf{x}}^{cs} - \mathbf{x}^{exp}\|_2,\ \ \pi \in \{R, G, B\},\\ &\texttt{where } \mathbf{x}^{exp} = |\max_{\xi\in\pi}(\mathbf{x}^{ll}) - \max_{\xi\in\pi}(\mathbf{z})|/\max_{\xi\in\pi}(\mathbf{z}).\end{aligned} \tag{8}$$

Here, $\mathbf{z}$ denotes a clear image of the same size as $\mathbf{x}^{ll}$.

Algorithm 1 Configuration-Driven Training for SNOC

Require: Initial configuration coefficients $\boldsymbol{\eta}^0$, restoration parameters $\boldsymbol{\theta}^0$, learning rates α_u, α_l, finite-difference coefficient β, split ratio s, total iterations $\mathcal{K}$.

Ensure: Configured coefficients $\boldsymbol{\eta}$ and optimized parameters $\boldsymbol{\theta}$.

```
Define objective subsets:
    $\mathcal{T}_l = \{\mathcal{L}_{cor}, \mathcal{L}_{exp}, \mathcal{L}_{con}\}$
    $\mathcal{T}_u = \mathcal{T}_l \cup \{\mathcal{L}_{per}, \mathcal{L}_{SSIM}, \mathcal{L}_{TV}\}$
Partition the training set into $\mathcal{D}_{cfg}$ and $\mathcal{D}_{opt}$ with ratio $1 : s$
for $k = 0$ to $\mathcal{K} - 1$ do
    // Step 1: restoration response under current configuration
    $\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}^k - \alpha_l \nabla_{\boldsymbol{\theta}} \Psi_{\text{opt}}(\boldsymbol{\theta}^k; \boldsymbol{\eta}^k, \mathcal{D}_{opt})$
    // Step 2: estimate configuration sensitivity via finite difference
    $\boldsymbol{\theta}^+ \leftarrow \boldsymbol{\theta}^k + \beta \nabla_{\boldsymbol{\theta}} \Psi_{\text{opt}}(\hat{\boldsymbol{\theta}}; \boldsymbol{\eta}^k, \mathcal{D}_{opt})$
    $\boldsymbol{\theta}^- \leftarrow \boldsymbol{\theta}^k - \beta \nabla_{\boldsymbol{\theta}} \Psi_{\text{opt}}(\hat{\boldsymbol{\theta}}; \boldsymbol{\eta}^k, \mathcal{D}_{opt})$
    $\mathcal{G}_C \leftarrow \frac{1}{2\beta}\left[\nabla_{\boldsymbol{\eta}} \Psi_{\text{opt}}(\boldsymbol{\theta}^+; \boldsymbol{\eta}^k, \mathcal{D}_{opt}) - \nabla_{\boldsymbol{\eta}} \Psi_{\text{opt}}(\boldsymbol{\theta}^-; \boldsymbol{\eta}^k, \mathcal{D}_{opt})\right]$
    // Step 3: update configuration and restoration parameters
    $\boldsymbol{\eta}^{k+1} \leftarrow \boldsymbol{\eta}^k - \alpha_u \nabla_{\boldsymbol{\eta}} \Psi_{\text{cfg}}(\boldsymbol{\eta}^k, \hat{\boldsymbol{\theta}}; \mathcal{D}_{cfg}) + \alpha_l \mathcal{G}_C$
    $\boldsymbol{\theta}^{k+1} \leftarrow \boldsymbol{\theta}^k - \alpha_l \nabla_{\boldsymbol{\theta}} \Psi_{\text{opt}}(\boldsymbol{\theta}^k; \boldsymbol{\eta}^{k+1}, \mathcal{D}_{opt})$
end for
```

Content reconstruction loss. We adopt the widely used L_1 reconstruction objective:

$$\mathcal{L}_{con}(\mathbf{y}^{ns}, \mathbf{y}^{nh}) = \frac{1}{hwc}\sum\nolimits_{i,j,k} |\mathbf{y}^{ns}{}_{i,j,k} - \mathbf{y}^{nh}{}_{i,j,k}|, \tag{9}$$

where h, w, and c denote height, width, and the number of channels, respectively.

Perceptual, structural, and smoothness objectives. We use perceptual loss [26] to preserve semantic consistency, denoted by $\mathcal{L}_{per}(\mathbf{y}^{ns}, \mathbf{y}^{nh})$. We further employ the SSIM loss $\mathcal{L}_{SSIM}$ to maintain structural similarity between $\mathbf{y}^{ns}$ and $\mathbf{y}^{nh}$. To suppress noise and encourage local smoothness, we adopt the standard total variation objective [47]:

$$\mathcal{L}_{TV}(\mathbf{y}^{ns}) = \sum\nolimits_{\xi\in\pi} (|\nabla_h \mathbf{y}^{ns}_{\xi}| + |\nabla_v \mathbf{y}^{ns}_{\xi}|), \tag{10}$$

where $\pi = \{R, G, B\}$, and ∇_h and ∇_v are the horizontal and vertical gradient operators, respectively.

4.2 Configuration-Driven Training

To overcome the brittleness of manual objective tuning, SNOC adopts a *configuration-driven training* paradigm that couples adaptive objective configuration with restoration optimization. Instead

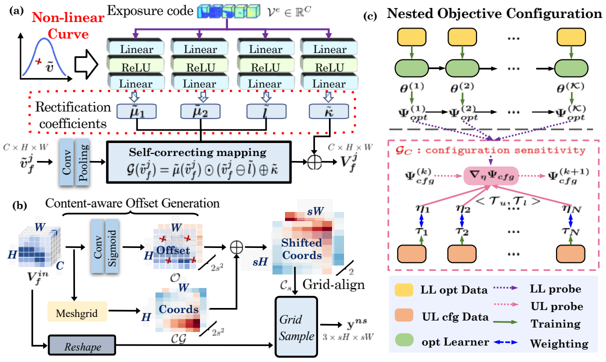

Figure 4: (a) EaRU, (b) GaDU, and (c) SNOC with training flow.

of prescribing fixed objective trade-offs, SNOC treats the coefficient vector $\boldsymbol{\eta}$ as a configurable controller over the objective portfolio, while the network parameters $\boldsymbol{\theta}$ evolve as the restoration response under the configured objectives.

Formally, let $\boldsymbol{\theta}^\dagger(\boldsymbol{\eta})$ denote the parameter response induced by configuration $\boldsymbol{\eta}$ on the optimization set $\mathcal{D}_{opt}$:

$$\boldsymbol{\theta}^\dagger(\boldsymbol{\eta}) = \arg\min_{\boldsymbol{\theta}} \Psi_{\text{opt}}(\boldsymbol{\theta}; \boldsymbol{\eta}, \mathcal{D}_{opt}). \tag{11}$$

The configuration itself is then optimized on a held-out set $\mathcal{D}_{cfg}$ by evaluating the restoration quality achieved under $\boldsymbol{\theta}^\dagger(\boldsymbol{\eta})$:

$$\min_{\boldsymbol{\eta}} \Psi_{\text{cfg}}\big(\boldsymbol{\eta}, \boldsymbol{\theta}^\dagger(\boldsymbol{\eta}); \mathcal{D}_{cfg}\big). \tag{12}$$

In this way, $\boldsymbol{\eta}$ acts as a dynamic regulator over the objective portfolio, while $\boldsymbol{\theta}^\dagger(\boldsymbol{\eta})$ feeds back the restoration response to guide configuration. We instantiate the configuration and optimization objectives from the portfolio $\mathcal{T}$ as

$$\Psi_{\text{cfg}} = \sum_{u=1}^{N} \eta_u \, \mathcal{L}_{\tau_u}\big(\boldsymbol{\eta}, \boldsymbol{\theta}^\dagger(\boldsymbol{\eta}); \mathcal{D}_{cfg}\big), \quad \mathcal{L}_{\tau_u} \in \mathcal{T}_u, \tag{13}$$

$$\Psi_{\text{opt}} = \sum_{l=1}^{M} \eta_l \, \mathcal{L}_{\tau_l}\big(\boldsymbol{\eta}, \boldsymbol{\theta}; \mathcal{D}_{opt}\big), \quad \mathcal{L}_{\tau_l} \in \mathcal{T}_l, \tag{14}$$

where $\boldsymbol{\eta} := \{\boldsymbol{\eta_u}\}_{\boldsymbol{u=1}}^{\boldsymbol{N}} \in \mathbb{R}^{\boldsymbol{N}}$ denotes learnable configuration coefficients, and we set $N = 6$, $M = 3$. Here, $\mathcal{T}_u$ and $\mathcal{T}_l$ denote the candidate objective subsets for configuration and optimization, respectively. During training, the dataset is partitioned into $\mathcal{D}_{cfg}$ and $\mathcal{D}_{opt}$ with an allocation ratio of $1:s$. This split allows the configuration stage to update coefficients using held-out feedback, thereby reducing the risk of overfitting to the optimization trajectory.

To improve generalization, we incorporate a Self-placed regularizer (SpR), $\mathcal{R}(\eta_u) = \ln(1+\eta_u^2)$, into the configuration objective:

$$\Psi_{\text{cfg}} = \sum_{u=1}^{N} \left(\frac{1}{2\eta_u^2} \mathcal{L}_{\tau_u}\big(\boldsymbol{\eta}, \boldsymbol{\theta}^\dagger(\boldsymbol{\eta}); \mathcal{D}_{cfg}\big) + \mathcal{R}(\eta_u) \right). \tag{15}$$

This regularization discourages degenerate coefficient configurations and stabilizes objective allocation across scenes. To capture how the restoration response changes with objective configuration, we further estimate a *configuration sensitivity* term: $\boldsymbol{\mathcal{G}_C} = \frac{\partial \Psi_{\text{cfg}}}{\partial \boldsymbol{\theta}} \nabla_{\boldsymbol{\eta}} \boldsymbol{\theta}^\dagger(\boldsymbol{\eta})$. In practice, $\boldsymbol{\mathcal{G}_C}$ is approximated by a first-order finite-difference scheme [32, 34], where β is a constant coefficient and $\boldsymbol{\theta}^\pm$ are the perturbed parameters used in the approximation. The configuration-driven procedure is summarized in Alg. 1.

5 Experimental Results and Analysis

5.1 Data Preparation

Real-world Datasets. We evaluate SNOC on both real-world and synthesized ultra-low-light datasets. For real-world evaluation, we use RELLISUR[2] and DarkFace[3]. Following the standard protocol, we use the 3,610 paired samples in RELLISUR for training and 425 paired samples for testing under both ×*2* and ×*4* scaling tasks. To further assess generalization to real nighttime scenes, we additionally select 100 images from DarkFace for testing.

Synthesized LLLR-NLHR Dataset. To complement real-world evaluation, we further construct a synthesized **LLLR-NLHR** dataset from high-quality normal-light images in DIV2K[4] and Flickr2K[5]. Specifically, we first generate low-light images through randomized brightness attenuation [40]: $\mathbf{x}_{out}^c = (\alpha_1 \cdot \mathbf{x}_{in}^c)^{\alpha_2} \cdot \alpha_3, \quad c \in \{R, G, B\}$, where $\alpha_1 \sim \mathcal{U}(0.7, 1)$ and $\alpha_3 \sim \mathcal{U}(0.5, 1)$ control linear transformations, and $\alpha_2 \sim \mathcal{U}(1.5, 4.5)$ is the gamma coefficient. We then simulate realistic low-light imaging by transforming RGB images into the RAW domain, injecting mixed shot/read noise, and converting them back to RGB:

$$\begin{aligned} \mathbf{x}_{raw} &= \hbar_w^{-1} \cdot (\hbar_g^{-1} \cdot (\hbar_c^{-1} \cdot (\hbar_t^{-1} \cdot (\mathbf{x}_{in})))), \\ \mathbf{x}_{noise} &= \hbar_t \cdot (\hbar_c \cdot (\hbar_g \cdot (\hbar_w \cdot (\Re_\delta(\mathbf{x}_{raw}))))), \end{aligned} \tag{16}$$

where $\hbar_t$, $\hbar_c$, $\hbar_g$, and $\hbar_w$ denote tone mapping, color correction, gamma processing, and white balance functions, respectively, together with their inverse operations, and $\Re_\delta$ denotes the realistic mixed-noise transformation [22, 73]. This process yields 7,100 synthesized pairs in total, evenly split into 3,550 pairs for ×*2* and 3,550 pairs for ×*4* scaling, of which 3,350 pairs are used for training and the remaining 200 for testing.

5.2 Experimental Settings

We follow the Alg. 1 for training. Specifically, we initialize model parameters $\boldsymbol{\theta}$ randomly and set configuration coefficients $\boldsymbol{\eta}$ as unit vectors. We set the total number of training iterations (i.e., $\mathcal{K}$) to 150,000 and employ an Adam-type optimizer with betas values set to $[0.9, 0.999]$. The initial learning rates for $\boldsymbol{\alpha_u}$ and $\boldsymbol{\alpha_l}$ are set to $1e^{-4}$ and $2e^{-4}$, respectively. We partition the dataset $\mathcal{D}$ into $\boldsymbol{\mathcal{D}_{cfg}}$ and $\boldsymbol{\mathcal{D}_{opt}}$ using an allocation ratio of $1 : \boldsymbol{s} = 1 : 7$, identified as the optimal choice. The experiments were conducted using PyTorch version 2.0.1 with CUDA 11.7 support, utilizing a single NVIDIA RTX A6000 GPU and 48GB RAM. For the evaluation, we employ six widely-used full-reference metrics, i.e., PSNR [3], SSIM [58], LPIPS [74], Root Mean Square Error (RMSE), Signal to Reconstruction Error ratio (SRE) [27] and MSE. We also introduce two no-reference metrics (i.e., NIQE [43] and LOE [54]) for unpaired evaluation.

Comparisons with State-of-the-Arts. We selected six state-of-the-art super-resolution methods: Restormer [72], SRFormer [78], HAT [5], RGT [6], PFT-SR [39] and HiT-SR [75]. To validate that straightforward cascaded LLE and normal-light SR methods are not efficient for addressing this joint task, we meticulously selected three different LLE methods, including RUAS [37] and SCI [41], and the supervised LLFormer [55].

---

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

[3] https://flyywh.github.io/CVPRW2019LowLight/

[4] https://data.vision.ee.ethz.ch/cvl/DIV2K/

[5] https://www.kaggle.com/datasets/hliang001/flickr2k

Table 1: Quantitative comparison on *RELLISUR* dataset for @×2 and @×4 tasks. Pink, orange, and yellow denote the top three results. † signifies training using the 1x dataset for LLE. ‡ indicates training using the 2x or 4x dataset for normal-light SR.

Method	*RELLISUR @ ×2*						*RELLISUR @ ×4*					
	❶PSNR↑	❷SSIM↑	❸LPIPS↓	❹RSME↓	❻SRE↑	❼MSE↓	❶PSNR↑	❷SSIM↑	❸LPIPS↓	❹RSME↓	❻SRE↑	❼MSE↓
SwinIR (ICCV'21)	18.383	0.640	0.577	0.125	55.647	96.993	17.531	0.663	0.688	0.139	58.432	98.923
Restormer (CVPR'22)	**21.217**	**0.727**	**0.385**	**0.095**	57.091	**85.628**	**20.290**	**0.720**	**0.492**	0.106	**59.822**	**87.977**
SRFormer (CVPR'23)	19.554	0.704	0.469	0.110	56.239	91.461	18.792	0.705	0.613	0.121	59.029	93.448
HAT (CVPR'23)	20.213	0.719	**0.454**	0.103	56.588	88.687	19.751	0.715	0.561	0.110	59.573	**90.574**
RGT (ICLR'24)	18.599	0.670	0.547	0.124	55.775	95.923	19.153	0.716	**0.538**	**0.083**	59.237	93.349
HiT-SR (ECCV'24)	16.291	0.624	0.589	0.159	54.638	105.534	17.647	0.683	0.643	0.136	58.539	99.430
PFT-SR (CVPR'25)	17.901	0.649	0.547	0.133	56.613	99.193	18.533	0.713	0.550	0.125	58.941	95.321
‡RUAS→†SwinIR	17.154	0.617	0.558	0.148	55.102	95.223	17.077	0.662	0.656	0.146	58.242	99.819
‡SwinIR→†RUAS	18.660	0.690	0.462	0.121	55.838	98.558	17.609	0.683	0.625	0.137	58.546	99.125
†SCI→‡HAT	14.963	0.439	0.591	0.200	**59.375**	104.586	14.776	0.439	0.591	0.200	57.100	104.057
‡LLFormer→†HAT	**21.218**	**0.720**	0.455	**0.093**	**57.114**	**87.645**	**20.135**	**0.718**	0.575	**0.105**	**59.765**	91.918
SNOC (-)	**22.426**	**0.742**	**0.310**	**0.081**	**57.712**	**77.247**	**20.906**	**0.724**	**0.407**	**0.095**	**60.134**	**86.954**

Table 2: Quantitative comparison on the *Synthesized LLLR-NLHR* dataset for @×2 and @×4 tasks. Pink, orange, and yellow denote the top three results. All methods are retrained on the synthesized dataset.

Method	*Synthesized LLLR-NLHR @ ×2*						*Synthesized LLLR-NLHR @ ×4*					
	❶PSNR↑	❷SSIM↑	❸LPIPS↓	❹RMSE↓	❻SRE↑	❼MSE↓	❶PSNR↑	❷SSIM↑	❸LPIPS↓	❹RMSE↓	❻SRE↑	❼MSE↓
SwinIR (ICCV'21)	20.440	0.745	0.326	0.099	59.084	93.045	19.636	0.638	0.508	0.108	58.652	93.610
Restormer (CVPR'22)	**21.217**	**0.750**	**0.313**	**0.094**	**59.329**	**91.019**	**20.410**	**0.689**	**0.451**	**0.093**	**59.034**	**90.768**
SRFormer (CVPR'23)	20.767	**0.749**	**0.319**	0.095	59.243	92.095	20.104	**0.646**	0.492	0.103	58.886	91.639
HAT (CVPR'23)	**21.012**	0.744	0.329	**0.092**	**59.469**	**91.168**	**20.266**	0.640	0.510	**0.101**	**58.986**	**91.291**
HiT-SR (ECCV'24)	20.009	0.733	0.347	0.105	58.873	95.176	19.399	0.634	0.488	0.112	58.549	95.164
PFT-SR (CVPR'25)	19.816	0.723	0.363	0.108	58.770	97.637	20.130	0.660	**0.454**	0.103	58.898	92.406
SNOC (-)	**21.738**	**0.753**	**0.271**	**0.087**	**59.754**	**89.458**	**20.941**	**0.704**	**0.426**	**0.088**	**59.108**	**88.261**

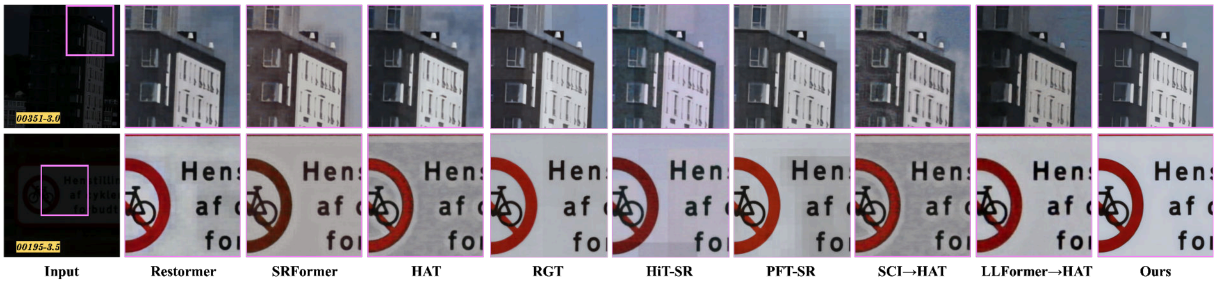

Figure 5: Qualitative assessments conducted on ultra-low-light *RELLISUR* samples, with a scale factor of ×4.

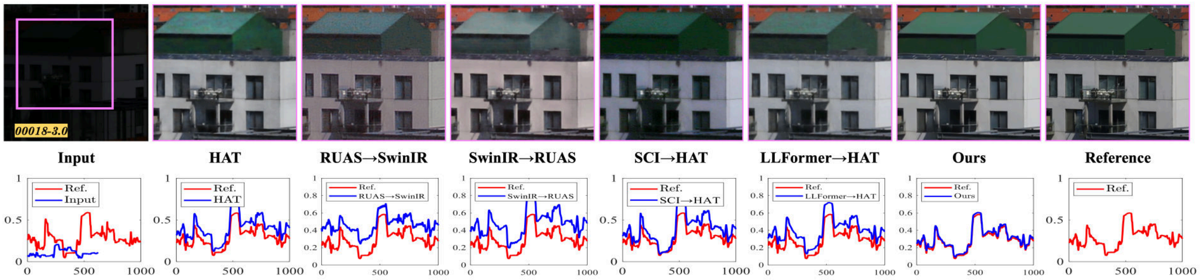

Figure 6: Qualitative evaluations of simultaneous brightening and zooming in low-light images on *RELLISUR* with ×2 scaling.

5.3 Quantitative and Qualitative Evaluation

5.3.1 ***Quantitative Evaluation.*** Tab. 1 presents the quantitative results of the joint task of simultaneous brightness enhancement and magnification on the real *RELLISUR* dataset. Of particular significance, the RMSE metric is highly sensitive to outliers; however, our method outperforms the second-place by 12.9% in RMSE score, highlighting the robustness of our approach across different lighting conditions. In Fig. 12, the violin plot provides a more comprehensive representation of the distribution of four indicators on the RELLISUR dataset. Tab. 2 reports the quantitative comparison on the synthesized *LLLR-NLHR* dataset, indicating that the proposed framework is effective not only on real ultra-low-light scenes but also under controlled synthetic degradations.

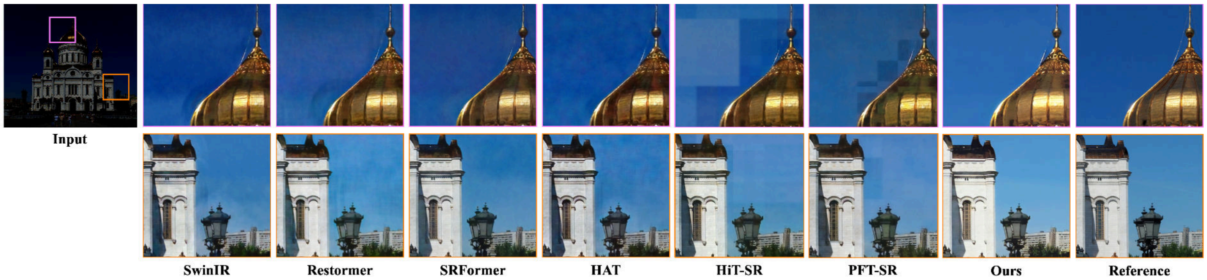

Figure 7: Qualitative evaluation on *Synthesized LLLR-NLHR* dataset with ×2 scaling.

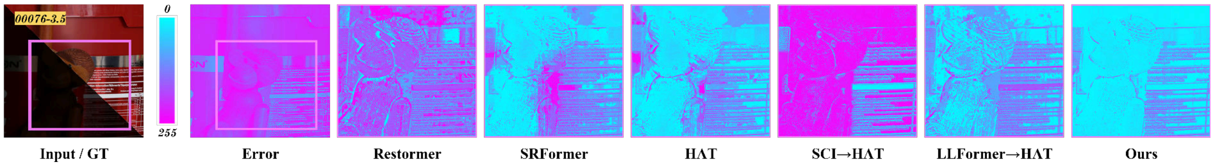

Figure 8: Error visualization for ×2 scaling. The transition from blue to purple indicates increasing pixel intensity error.

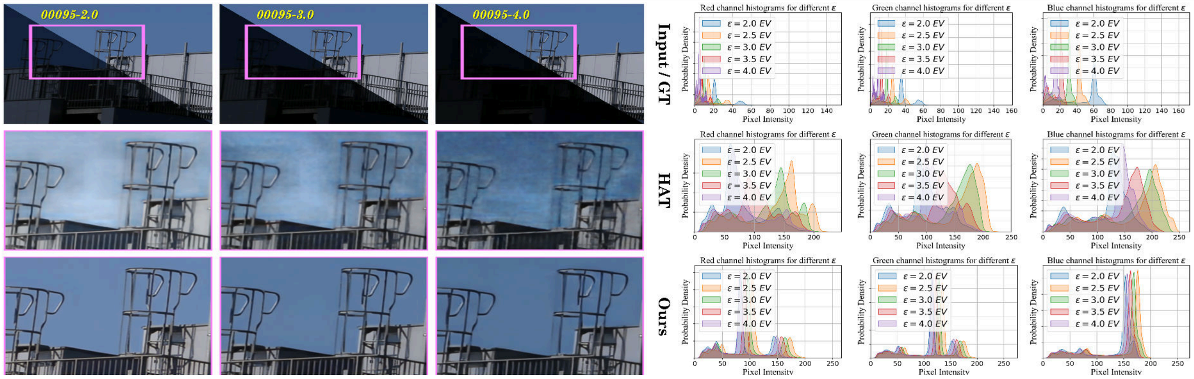

Figure 9: Qualitative evaluations captured under five distinct ultra-low-light conditions. Note that the images are arranged from left to right, becoming progressively darker with different darkness levels ϵ. The last three columns display the probability density curves for the *Red, Green, and Blue* color channels concerning various darkness levels.

Figure 10: Qualitative assessments for real-world nighttime images, on *DarkFace* samples for ×2 scaling.

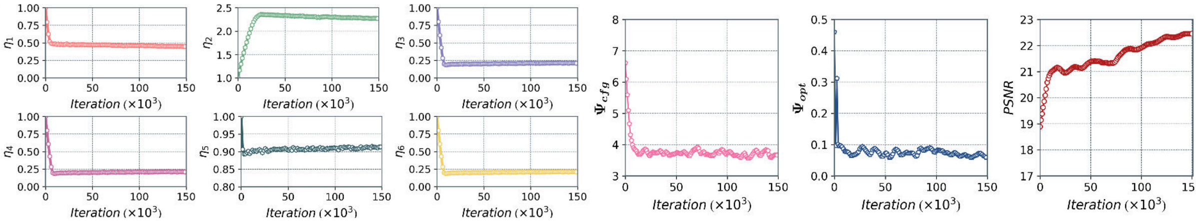

Figure 11: Configuration dynamics of SNOC (starting from unit initialization), including the evolution of configuration-values, objective energies, and PSNR over 150,000 training iterations.

5.3.2 ***Qualitative Evaluation*.** Fig. 6 displays the simultaneous brightness adjustment and magnification results for the 2x task on *RELLISUR*. Fig. 8 displays the error visualization of a 2x scaling. Our method achieves a more blue-shifted error map, indicating greater pixel-by-pixel consistency with the reference image. Fig. 5 presents the enhancement results for the ×*4* task. Our approach generates

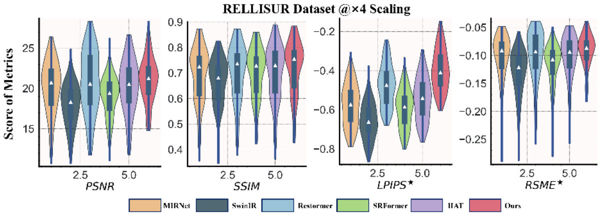

Figure 12: Quantitative assessment with the violin plot displaying the distribution of four indicators.

Table 3: Computational efficiency on images of size 128×128**.**

Metric	SwinIR	Restormer	SRFormer	HAT	RGT	HiT-SR	Ours
Params (MB)↓	11.683	26.126	10.162	9.473	10.051	**0.772**	6.356
FLOPs (G)↓	453.12	35.375	81.797	58.990	77.919	**18.290**	20.817
Inference Time (s)↓	3.879	**0.033**	0.218	0.184	0.615	0.044	0.040

Parameters of the pre-trained segmentation network are not included in our method.

images with pleasant brightness and preserved texture details, presenting a sense of realism. Fig. 7 shows that SNOC reconstructs cleaner structures and sharper local details on the synthesized *LLLR-NLHR* dataset than competing methods.

5.3.3 *Evaluation on Real Nighttime Scenarios*

Fig. 10 showcases the enhancement results of images on a real nighttime image dataset-DarkFace. Observing the highlighted regions in the pink boxes, it can be seen that our approach excels in maintaining the natural authenticity of nighttime scenes, reducing color deviations.

5.3.4 *Robustness across Diverse Darkness Levels*

Fig. 9 compares restoration performance under different input exposure levels. The right three columns show the RGB probability distributions under five representative darkness factors. SNOC maintains robust restoration quality across different degradation levels, and the corresponding curves indicate that the proposed configuration-driven strategy does not overfit to a single exposure regime but remains effective under varying low-light severities.

5.3.5 *Analysis of Configuration Dynamics*

Following Alg. 1, we analyze the evolution of configuration coefficients and objective values during training. As shown in Fig. 11, the final coefficient vector converges to [0.426, 2.253, 0.125, 0.061, 0.916, 0.216], indicating that the configured objective portfolio is not uniformly weighted. In particular, perceptual consistency receives the largest coefficient (i.e., $\eta_2 = 2.253$), followed by the content objective (i.e., $\eta_5 = 0.916$), which suggests that perceptual realism and content fidelity are the most influential factors in driving performance improvements.

5.3.6 *Analysis of Model Efficiency*

We provid the *Parameters*, *FLOPs* and *Inferences* for several recent methods in Tab. 3. These metrics were evaluated on the *RELLISUR* dataset using images with dimensions of 128 × 128 on a single NVIDIA 2080Ti GPU. It is worth noting that our proposed method achieves a favorable tradeoff between performance and computational efficiency.

5.4 Ablation Study

5.4.1 *Effectiveness of IIA and SICC*

As indicated in Tab. 4, deactivating both IIA and SICC modules led to performance degradation. Comparison of experiments No. 3 or 4 with No. 6 reveals that removing any branch from SICC (or segregating the semantic or illumination flow) leads to a decline in performance. Furthermore, the removal of color constraints in the initial stage of RiI_2 is critical for enhancing final performance, as demonstrated by comparing No. 5 with No. 6. Fig. 13 visualizes the feature maps at different decoding scales, showing that SICC produces more discriminative and consistent rectified representations.

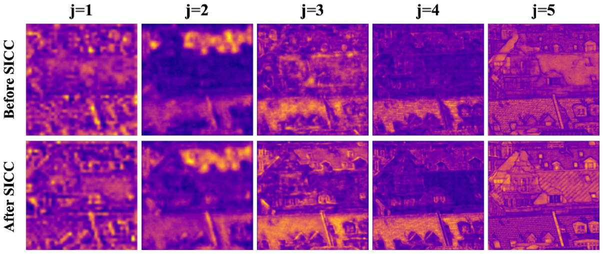

Figure 13: Ablation of feature visualization for SICC, on five different scale decoding layers, displayed at an equal size.

Table 4: Ablation studies of the IIA and SICC modules.

No.	IIA	SICC	PSNR↑	SSIM↑	LPIPS↓
1	✗	✗	21.423	0.719	0.403
2	✓	✗	$21.998_{\uparrow 0.575}$	$0.721_{\uparrow 0.002}$	$0.358_{\downarrow 0.045}$
3	✓	Only $\tilde{s}_f^j$	$22.155_{\uparrow 0.732}$	$0.724_{\uparrow 0.005}$	$0.331_{\downarrow 0.072}$
4	✓	Only $\tilde{u}_f^j$	$22.161_{\uparrow 0.738}$	$0.730_{\uparrow 0.011}$	$0.334_{\downarrow 0.069}$
5	w/o supervision	✓	$22.120_{\uparrow 0.697}$	$0.725_{\uparrow 0.006}$	$0.347_{\downarrow 0.056}$
6	✓	✓	$\mathbf{22.426}_{\uparrow \mathbf{1.003}}$	$\mathbf{0.742}_{\uparrow \mathbf{0.023}}$	$\mathbf{0.310}_{\downarrow \mathbf{0.093}}$

Table 5: Ablation studies of CsEE, EaRU and GaDU modules.

No.	CsEE	EaRU	GaDU	PSNR↑	SSIM↑
1	✗	✗	✓	22.047	0.729
2	✓	Concatenation	✓	$22.125_{\uparrow 0.078}$	$0.731_{\uparrow 0.002}$
3	w/o supervision	✓	✓	$22.210_{\uparrow 0.163}$	$0.737_{\uparrow 0.008}$
4	✓	✓	Bilinear	$22.148_{\uparrow 0.101}$	$0.735_{\uparrow 0.006}$
5	✓	✓	✓	$\mathbf{22.426}_{\uparrow \mathbf{0.379}}$	$\mathbf{0.742}_{\uparrow \mathbf{0.013}}$

5.4.2 *Effectiveness of CsEE, EaRU, and GaDU*

We further analyze the effectiveness of the exposure-aware branch and dynamic upsampling modules in Tab. 5. Removing the entire exposure-aware branch, including CsEE and EaRU, causes a clear PSNR drop of 0.379 dB (No. 1), which verifies that exposure-guided priors are critical for handling regional non-uniformity. Replacing EaRU with simple concatenation (No. 2) also degrades performance. Substituting GaDU with bilinear upsampling (No. 4) leads to inferior results, confirming the importance of grid-aware sampling for suppressing magnification-induced artifacts.

6 Conclusion

We presented **SNOC** (**S**ubtle **N**ested **O**bjective **C**onfiguration), a unified framework for joint ultra-low-light enhancement and super-resolution. By coupling a subtle rectification architecture with nested objective configuration over a module-specific objective portfolio, SNOC enables adaptive restoration target allocation and achieves strong perceptual quality and reconstruction fidelity under severe ultra-low-light conditions. Future work will focus on more efficient configuration mechanisms and lightweight rectification architectures for practical edge-device applications.

References

[1] Andreas Aakerberg, Kamal Nasrollahi, and Thomas B Moeslund. 2023. RELIEF: Joint low-light image enhancement and super-resolution with transformers. In *Scandinavian Conference on Image Analysis*. 157–173.

[2] Ali Behrouz, Meisam Razaviyayn, Peilin Zhong, and Vahab Mirrokni. 2025. Nested Learning: The Illusion of Deep Learning Architectures. In *NeurIPS*.

[3] Luen C Chan and Peter Whiteman. 1983. Hardware-constrained hybrid coding of video imagery. *IEEE Trans. Aerospace Electron. Systems* 1 (1983), 71–84.

[4] Chu Chen, Kangning Cui, Pasquale Cascarano, Wei Tang, Elena Loli Piccolomini, and Raymond H Chan. 2025. Blind Restoration of High-Resolution Ultrasound Video. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 77–87.

[5] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. 2023. Activating More Pixels in Image Super-Resolution Transformer. In *CVPR*. 22367–22377.

[6] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, and Xiaokang Yang. 2024. Recursive Generalization Transformer for Image Super-Resolution. In *ICLR*.

[7] Runmin Cong, Kaisheng Pang, Feng Li, Hua Li, Huihui Bai, Sam Kwong, and Wei Zhang. 2026. DiffLLFace: Learning Alternate Illumination-Diffusion Adaptation for Low-Light Face Super-Resolution and Beyond. *IEEE Transactions on Image Processing* 35 (2026), 2499–2512.

[8] Kangning Cui, Ruoning Li, Sam L Polk, Yinyi Lin, Hongsheng Zhang, James M Murphy, Robert J Plemmons, and Raymond H Chan. 2024. Superpixel-based and spatially regularized diffusion learning for unsupervised hyperspectral image clustering. *IEEE Transactions on Geoscience and Remote Sensing* 62 (2024), 1–18.

[9] Kangning Cui, Rongkun Zhu, Manqi Wang, Wei Tang, Gregory D Larsen, Victor P Pauca, Sarra Alqahtani, Fan Yang, David Segurado, David A Lutz, et al. 2025. Detection and Geographic Localization of Natural Objects in the Wild: A Case Study on Palms. In *IJCAI*. 9601–9609.

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *CVPR*. 248–255.

[11] Wang Di, Liu Jinyuan, Risheng Liu, and Fan Xin. 2023. An interactively reinforced paradigm for joint infrared-visible image fusion and saliency object detection. *Information Fusion* 98, 101828.

[12] Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, and Changqing Zou. 2025. Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution. In *CVPR*. 23174–23184.

[13] Xin Fan, Zi Li, Ziyang Li, Xiaolin Wang, Risheng Liu, Zhongxuan Luo, and Hao Huang. 2023. Automated learning for deformable medical image registration by jointly optimizing network architectures and objective functions. *IEEE Transactions on Image Processing* 32 (2023), 2800–2812.

[14] Xin Fan, Xiaolin Wang, Jiaxin Gao, Jia Wang, Zhongxuan Luo, and Risheng Liu. 2024. Bi-level Learning of Task-Specific Decoders for Joint Registration and One-Shot Medical Image Segmentation. In *CVPR*. 11574–11583.

[15] Luca Franceschi, Michele Donini, Massimiliano Pontil, and Paolo Pontil. 2018. Bilevel programming for hyperparameter optimization and meta-learning. In *ICML*. 1568–1577.

[16] Jiaxin Gao, Xiaokun Liu, Risheng Liu, and Xin Fan. 2023. Learning adaptive hyper-guidance via proxy-based bilevel optimization for image enhancement. *The Visual Computer* 39, 4 (2023), 1471–1484.

[17] Jiaxin Gao and Yaohua Liu. 2024. Enhancing Images with Coupled Low-Resolution and Ultra-Dark Degradations: A Tri-level Learning Framework. In *ACM MM*. 8642–8651.

[18] Jiaxin Gao, Yaohua Liu, Ziyu Yue, Xin Fan, and Risheng Liu. 2024. Collaborative brightening and amplification of low-light imagery via bi-level adversarial learning. *Pattern Recognition* 154 (2024), 110558.

[19] Jiaxin Gao, Ziyu Yue, Yaohua Liu, Sihan Xie, Xin Fan, and Risheng Liu. 2024. A dual-stream-modulated learning framework for illuminating and super-resolving ultra-dark images. *IEEE transactions on neural networks and learning systems* 36, 4 (2024), 7500–7513.

[20] Chunle Guo, Chongyi Li, Jichang Guo, Runmin Cong, Huazhu Fu, and Ping Han. 2018. Hierarchical features driven residual learning for depth map super-resolution. *IEEE Transactions on Image Processing* 28, 5 (2018), 2545–2557.

[21] Kehua Guo, Min Hu, Sheng Ren, Fangfang Li, Jian Zhang, Haifu Guo, and Xiaoyan Kui. 2022. Deep illumination-enhanced face super-resolution network for low-light images. *ACM Transactions on Multimedia Computing, Communications, and Applications* 18, 3 (2022), 1–19.

[22] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. 2019. Toward convolutional blind denoising of real photographs. In *CVPR*. 1712–1722.

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

[24] Xianghao Jiao, Yaohua Liu, Jiaxin Gao, Xinyuan Chu, Xin Fan, and Risheng Liu. 2023. PEARL: Preprocessing Enhanced Adversarial Robust Learning of Image Deraining for Semantic Segmentation. In *ACM MM*. 8185–8194.

[25] Xianghao Jiao, Di Wang, Jiawei Liang, Jianjie Huang, Wei Wang, and Xiaochun Cao. 2025. Object-Level Backdoor Attacks in RGB-T Semantic Segmentation with Cross-Modality Trigger Optimization. In *IJCAI*. 1269–1277.

[26] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*. 694–711.

[27] Charis Lanaras, José Bioucas-Dias, Silvano Galliani, Emmanuel Baltsavias, and Konrad Schindler. 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing* 146 (2018), 305–319.

[28] Edwin H Land. 1977. The retinex theory of color vision. *Scientific american* 237, 6 (1977), 108–129.

[29] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. 2021. Low-light image and video enhancement using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence* 44, 12 (2021), 9396–9416.

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

[31] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In *CVPR*. 1833–1844.

[32] Hanxiao Liu, Karen Simonyan, and Yiming Yang. 2019. DARTS: Differentiable Architecture Search. In *ICLR*.

[33] Risheng Liu, Jiaxin Gao, Xuan Liu, and Xin Fan. 2024. Learning with constraint learning: New perspective, solution strategy and various applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 7 (2024), 5026–5043.

[34] Risheng Liu, Jiaxin Gao, Jin Zhang, Deyu Meng, and Zhouchen Lin. 2021. 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 (2021), 10045–10067.

[35] Risheng Liu, Zi Li, Xin Fan, Chenying Zhao, Hao Huang, and Zhongxuan Luo. 2022. Learning Deformable Image Registration From Optimization: Perspective, Modules, Bilevel Training and Beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 11 (2022), 7688–7704.

[36] Risheng Liu, Yaohua Liu, Shangzhi Zeng, and Jin Zhang. 2026. Augmenting iterative trajectory for bilevel optimization: Methodology, analysis and extensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 48, 4 (2026), 4239–4256.

[37] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *CVPR*. 10561–10570.

[38] Yaohua Liu, Jiaxin Gao, Xuan Liu, Xianghao Jiao, Xin Fan, and Risheng Liu. 2024. Advancing generalized transfer attack with initialization derived bilevel optimization and dynamic sequence truncation. In *IJCAI*. 1137–1145.

[39] Wei Long, Xingyu Zhou, Leheng Zhang, and Shuhang Gu. 2025. Progressive Focused Transformer for Single Image Super-Resolution. In *CVPR*. 2279–2288.

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

[41] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. 2022. Toward fast, flexible, and robust low-light image enhancement. In *CVPR*. 5637–5646.

[42] Matthew MacKay, Paul Vicol, Jonathan Lorraine, David Duvenaud, and Roger B. Grosse. 2019. Self-Tuning Networks: Bilevel Optimization of Hyperparameters using Structured Best-Response Functions. In *ICLR*.

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

[44] Pan Mu, Zhu Liu, Yaohua Liu, Risheng Liu, and Xin Fan. 2021. Triple-level model inferred collaborative network architecture for video deraining. *IEEE Transactions on Image Processing* 31 (2021), 239–250.

[45] Deepak Rai and Shyam Singh Rajput. 2023. Low-light robust face image super-resolution via neuro-fuzzy inferencing-based locality constrained representation. *IEEE Transactions on Instrumentation and Measurement* 72 (2023), 1–11.

[46] Aravind Rajeswaran, Chelsea Finn, Sham M. Kakade, and Sergey Levine. 2019. Meta-Learning with Implicit Gradients. In *NeurIPS*, Vol. 32.

[47] Leonid I Rudin, Stanley Osher, and Emad Fatemi. 1992. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena* 60, 1-4 (1992), 259–268.

[48] Lingchen Sun, Rongyuan Wu, Zhiyuan Ma, Shuaizheng Liu, Qiaosi Yi, and Lei Zhang. 2025. Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach. In *CVPR*. 2333–2343.

[49] Chenyang Wang, Junjun Jiang, Kui Jiang, and Xianming Liu. 2024. Low-light face super-resolution via illumination, structure, and texture associated representation. In *AAAI*, Vol. 38. 5318–5326.

[50] Di Wang, Xianghao Jiao, Jinyuan Liu, and Xin Fan. 2025. Robust One-Stop Multi-Modality Image Registration-Fusion-Segmentation Framework Against Misalignments and Adversarial Attacks. *IEEE Transactions on Multimedia* 27 (2025), 4531–4543.

[51] Di Wang, Jinyuan Liu, Long Ma, Risheng Liu, and Xin Fan. 2024. Improving Misaligned Multi-modality Image Fusion with One-stage Progressive Dense Registration. *IEEE Transactions on Circuits and Systems for Video Technology* 34, 11 (2024), 10944–10958.

[52] Di Wang, Long Ma, Risheng Liu, and Xin Fan. 2022. Semantic-aware Texture-Structure Feature Collaboration for Underwater Image Enhancement. In *IEEE International Conference on Robotics and Automation ICRA*. 4592–4598.

[53] Hai Wang, Yanyan Chen, Yingfeng Cai, Long Chen, Yicheng Li, Miguel Angel Sotelo, and Zhixiong Li. 2022. SFNet-N: An improved SFNet algorithm for semantic segmentation of low-light autonomous driving road scenes. *IEEE Transactions on Intelligent Transportation Systems* 23, 11 (2022), 21405–21417.

[54] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing* 22, 9 (2013), 3538–3548.

[55] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. 2023. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *AAAI*, Vol. 37. 2654–2662.

[56] Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2021. Hla-face: Joint high-low adaptation for low light face detection. In *CVPR*. 16195–16204.

[57] Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C Kot, and Bihan Wen. 2024. Sinsr: diffusion-based image super-resolution in a single step. In *CVPR*. 25796–25805.

[58] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612.

[59] Zhihao Wang, Jian Chen, and Steven CH Hoi. 2020. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 10 (2020), 3365–3387.

[60] Zhijie Wang, Aiping Liu, Jinbao Wei, Qingguo Xie, Kongqiao Wang, and Xun Chen. 2024. DDC-Net: Dual-Domain Cascaded Network with POCS Prior for Fast MRI Reconstruction. *IEEE Sensors Journal* (2024).

[61] Zhijie Wang, Jinbao Wei, Gang Yang, Aiping Liu, Wei Wei, Bensheng Qiu, and Xun Chen. 2025. Dual-Domain Self-Consistency-Enhanced Deep Unfolding Network for accelerated MRI reconstruction. *Computer Methods and Programs in Biomedicine* (2025), 108995.

[62] Jinbao Wei, Yuhang Chen, Zhijie Wang, Gang Yang, Shimin Tao, Jian Gao, Aiping Liu, and Xun Chen. 2025. Rethinking Diffusion Bridge Model with Dual Alignments for Medical Image Synthesis. In *ACM MM*. 1052–1061.

[63] Jinbao Wei, Zhijie Wang, Kongqiao Wang, Li Guo, Xueyang Fu, Ji Liu, and Xun Chen. 2023. Accurate MRI Reconstruction via Multi-Domain Recurrent Networks. In *IJCAI*. 1524–1532.

[64] Jinbao Wei, Gang Yang, Zhijie Wang, Yu Liu, Aiping Liu, and Xun Chen. 2024. Misalignment-Resistant Deep Unfolding Network for multi-modal MRI super-resolution and reconstruction. *Knowledge-Based Systems* 296 (2024), 111866.

[65] Jinbao Wei, Gang Yang, Zhijie Wang, Shimin Tao, Aiping Liu, and Xun Chen. 2025. Degradation-Aware Prompted Transformer for Unified Medical Image Restoration. *IEEE Transactions on Image Processing* 34 (2025), 8583–8598.

[66] Jinbao Wei, Gang Yang, Wei Wei, Aiping Liu, and Xun Chen. 2025. Multi-contrast MRI arbitrary-scale super-resolution via dynamic implicit network. *IEEE Transactions on Circuits and Systems for Video Technology* (2025).

[67] Sihan Xie, Peiming Li, Jiaxin Gao, Ziyu Yue, Xin Fan, and Risheng Liu. 2024. Breaking the water dilemma: Transmission-guided bilevel adaptive learning for underwater imagery. *Neurocomputing* 596 (2024), 127909.

[68] Mingjie Xu, Chaoqun Zhuang, Feifan Lv, Feng Lu, and HCBC Cloud. 2023. Joint Low-light Enhancement and Super Resolution with Image Underexposure Level Guidance.. In *BMVC*. 46–52.

[69] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. 2020. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *CVPR*. 3063–3072.

[70] Ziyu Yue, Jiaxin Gao, and Zhixun Su. 2024. Unveiling details in the dark: Simultaneous brightening and zooming for low-light image enhancement. In *AAAI*, Vol. 38. 6899–6907.

[71] Ziyu Yue, Jiaxin Gao, Sihan Xie, Yang Liu, and Zhixun Su. 2024. LoLiSRFlow: Joint single image low-light enhancement and super-resolution via cross-scale transformer-based conditional flow. *arXiv preprint arXiv:2402.18871* (2024).

[72] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*. 5728–5739.

[73] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing* 26, 7 (2017), 3142–3155.

[74] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *CVPR*.

[75] Xiang Zhang, Yulun Zhang, and Fisher Yu. 2024. HiT-SR: Hierarchical transformer for efficient image super-resolution. In *ECCV*. 483–500.

[76] Yu Zhang, Xiaoguang Di, Bin Zhang, Ruihang Ji, and Chunhui Wang. 2021. Better than reference in low-light image enhancement: conditional re-enhancement network. *IEEE Transactions on Image Processing* 31 (2021), 759–772.

[77] Feihu Zhou, Kan Chang, Mingyang Ling, Hengxin Li, and Shucheng Xia. 2024. Joint Image Super-Resolution and Low-Light Enhancement in the Dark. In *ACCV*. 1332–1349.

[78] Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. 2023. SRFormer: Permuted Self-Attention for Single Image Super-Resolution. In *CVPR*.