# **UnfoldIR: Rethinking Deep Unfolding Network in Illumination Degradation Image Restoration**

Chunming He<sup>1</sup>, Rihan Zhang<sup>1</sup>, Fengyang Xiao<sup>1</sup>, Chengyu Fang<sup>2</sup>, Longxiang Tang<sup>2</sup>, Yulun Zhang<sup>3</sup>, Sina Farsiu<sup>1,\*</sup>

<sup>1</sup>Duke University, <sup>3</sup>Tsinghua University, <sup>3</sup>Shanghai Jiao Tong University

#### **Abstract**

Deep unfolding networks (DUNs) are widely employed in illumination degradation image restoration (IDIR) to merge the interpretability of model-based approaches with the generalization of learning-based methods. However, the performance of DUN-based methods remains considerably inferior to that of state-of-the-art IDIR solvers. Our investigation indicates that this limitation does not stem from structural shortcomings of DUNs but rather from the limited exploration of the unfolding structure, particularly for (1) constructing task-specific restoration models, (2) integrating advanced network architectures, and (3) designing DUN-specific loss functions. To address these issues, we propose a novel DUN-based method, UnfoldIR, for IDIR tasks. UnfoldIR first introduces a new IDIR model with dedicated regularization terms for smoothing illumination and enhancing texture. We unfold the iterative optimized solution of this model into a multistage network, with each stage comprising a reflectance-assisted illumination correction (RAIC) module and an illumination-guided reflectance enhancement (IGRE) module. RAIC employs a visual state space (VSS) to extract non-local features, enforcing illumination smoothness, while IGRE introduces a frequency-aware VSS to globally align similar textures, enabling mildly degraded regions to guide the enhancement of details in more severely degraded areas. This suppresses noise while enhancing details. Furthermore, given the multistage structure, we propose an inter-stage information consistent loss to maintain network stability in the final stages. This loss contributes to structural preservation and sustains the model's performance even in unsupervised settings. Experiments verify our effectiveness across 5 IDIR tasks and 3 downstream problems. Besides, our analysis of the intrinsic mechanisms of DUNs provides valuable insights for future research. Code will be released.

# 1 Introduction

Illumination degradation image restoration (IDIR) [1–4], with representative tasks listed in Fig. 1, refers to a set of challenging restoration tasks in which images suffer from the adverse effects of degraded illumination, such as low contrast and non-uniform noise. By addressing illumination degradation, the restored data are expected to exhibit enhanced details and improved fidelity, thereby facilitating downstream tasks like nighttime object detection. To achieve this, substantial algorithms have been proposed [5–8], mainly including model-based and deep learning-based methods.

Model-based methods [5, 6] rely on manually designed restoration rules, which afford clear interpretability but suffer from limited generalization. In contrast, learning-based strategies [7, 8], which benefit from end-to-end training, have achieved significant success in IDIR tasks due to their improved generalization, although they remain less interpretable. To unify the merits of model-based and learning-based methods, deep unfolding networks (DUNs) have been proposed for IDIR. These

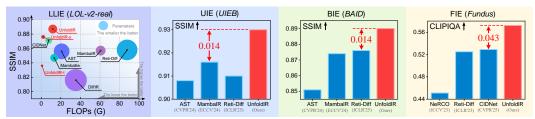


Figure 1: Results on IDIR tasks with commonly used datasets and metrics. In low-light image enhancement (LLIE), we compare our methods, UnfoldIR and its two lightweight variants (UnfoldIR-t and UnfoldIR-s), with cutting-edge methods. Our advancement is also proven by underwater image enhancement (UIE), backlit image enhancement (BIE), and fundus image enhancement (FIE).

methods [9, 10] unfold the iterative optimized solution of a task-specific model—typically one based on the Retinex model—into a multi-stage network, converting fixed parameters into learnable ones.

Notably, the performance of existing DUN-based methods is lower than that of state-of-the-art IDIR solvers, which challenges the applicability of DUN-based frameworks. This limitation does not reflect structural flaws in DUNs but rather indicates that the potential of the unfolding structure has not been fully realized—especially for (1) constructing restoration models with task-specific regularization terms during the modeling phase, (2) integrating advanced networks within the unfolding process, and (3) designing loss functions that align with DUN structures during training time.

To address these issues, we propose a novel DUN-based method, UnfoldIR, for IDIR tasks. First, UnfoldIR introduces a new IDIR model (IDIRM) based on the Retinex theory, which decomposes the input image into illumination and reflectance components. IDIRM incorporates dedicated regularization terms for the illumination component to enforce smoothness and for the reflectance component to suppress noise and enhance texture. We then use the proximal gradient algorithm to optimize the model and unfold the iterative solution into a multi-stage network architecture, UnfoldIR.

In UnfoldIR, each stage comprises two modules: the reflectance-assisted illumination correction (RAIC) module and the illumination-guided reflectance enhancement (IGRE) module, which iteratively optimize the illumination and reflectance components. The RAIC module employs a visual state space (VSS) to extract non-local features from the illumination component, effectively reducing color distortion and maintaining global color consistency. In addition to mitigating the inherent imaging noise in the reflectance, the IGRE module must also suppress noise introduced during the illumination recovery process. To achieve this, we introduce a frequency-aware VSS that globally aligns similar texture details using the estimated illumination as a condition. This mechanism allows lightly degraded regions to guide the enhancement of details in more severely degraded areas. Moreover, we integrate the VSS into a high-order ordinary differential equation framework—specifically, a second-order Runge-Kutta method—to accurately suppress noise and enhance image details.

Additionally, given our multi-stage structure, we propose an inter-stage information consistent (ISIC) loss. The ISIC loss is designed to maintain stability in the final stages so that small changes in illumination do not compromise essential reflectance details in the restored image, and vice versa. This contributes to structural preservation and distortion reduction. Moreover, since the ISIC loss serves as a framework-specific self-constraint, it can also be employed in unsupervised settings.

This paper also explores the intrinsic advantages of DUNs and investigates the deployment for image restoration in appendix A, aiming to guide the design of future DUN-based methods in this domain.

Our contributions are summarized as follows:

- (1) We propose UnfoldIR, a novel deep unfolding network designed for IDIR tasks, which incorporates a new IDIR model designed to enforce illumination smoothness, suppress noise, and enhance texture.
- (2) UnfoldIR introduces two new modules—RAIC and IGRE—specifically tailored to the illumination and reflectance components. The RAIC module enforces illumination smoothness, while the IGRE module enhances structural details and suppresses undesired noise in the reflectance component.
- (3) We propose a self-supervised ISIC loss for structural preservation and distortion elimination, maintaining the model's performance even in unsupervised settings.
- (4) Experiments validate our effectiveness across 5 IDIR tasks and 3 downstream problems. Besides, our analysis of DUN's internal mechanisms provides insights to guide future DUN-based methods.

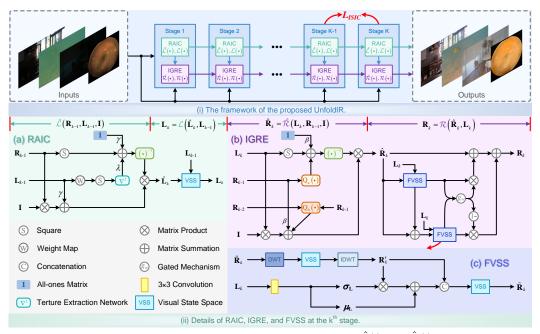


Figure 2: Framework of our UnfoldIR. The network connection in  $\hat{\mathcal{L}}(\cdot)$  and  $\hat{\mathcal{R}}(\cdot)$  are derived strictly based on mathematical principles, thus enhancing interpretability. For clarity, we replace certain redundant details with  $\mathbf{Q_a}$  and  $\mathbf{Q_b}$ , and present  $\hat{\mathcal{R}}(\cdot)$  according to Eq. (12).

### 2 Related Works

Illumination Degradation Image Restoration. IDIR has progressed from traditional methods to deep learning-based solutions [11–14]. Among them, Retinex theory-based methods play a significant role in decomposing inputs into reflectance and illumination components. Traditional ones typically introduce explicit Retinex priors to constrain Retinex maps [5, 15], offering interpretability but limited generalization. In contrast, learning-based methods, although lacking interpretability, achieve SOTA performance due to their powerful feature extractors. For example, DIE [16] integrated Retinex cues with a one-stage framework to address color distortion. Reti-diff [1] added estimated Retinex priors into a transformer for efficient inference. Besides, DUN, combining the merits of traditional and learning-based methods, has recently been introduced into IDIR, with the potential to achieve superior results. However, existing DUNs [9, 10] still underperform compared to current SOTAs.

**Deep Unfolding Network**. DUNs, such as DM-Fusion [17] and DeRUN [18], aim to unfold iterative optimization algorithms into deep networks, converting traditionally fixed parameters and operators into learnable ones. Several DUN-based methods have been introduced in IDIR. For instance, Uretinex-Net [9] unfolded the Retinex model into a multi-stage network, mitigating noise interference. CUE [10] leveraged a masked autoencoder-based loss function to supervise an unfolded network for learning customized priors. However, these methods still fall notably short of the existing SOTAs. Our analysis indicates that this limitation does not stem from inherent structural deficiencies in DUNs but rather from the insufficient exploration of the unfolding structures. To address this, we propose UnfoldIR, a novel DUN-based framework tailored for IDIR tasks. Specifically, UnfoldIR constructs a task-specific restoration model (IDIRM), incorporates advanced VSS structures to form RAIC and IGRE modules, and introduces a DUN-specific ISIC loss to maintain network stability. With these enhancements, our UnfoldIR achieves leading performance across five IDIR tasks.

# 3 Methodology

### 3.1 IDIR Model

According to Retinex theory, an illumination-degraded image  ${\bf I}$  can be decomposed into its reflectance image  ${\bf R}$  and illumination map  ${\bf L}$  using the Hadamard product  $\odot$ , formulated as

$$\mathbf{I} = \mathbf{R} \odot \mathbf{L} = (\mathbf{R}_{HQ} + \hat{\mathbf{R}}) \odot (\mathbf{L}_{HQ} + \hat{\mathbf{L}}), \tag{1}$$

where  $\mathbf{R}_{HQ}$  and  $\mathbf{L}_{HQ}$  denote the latent high-quality reflectance and illumination components, and  $\mathbf{R}$ and L represent perturbations corresponding to textural degradation and color distortion. To reduce these perturbations, we restore  $\mathbf{R}$  and  $\mathbf{L}$  by optimizing the following objective function:

$$L(\mathbf{R}, \mathbf{L}) = \frac{1}{2} \|\mathbf{I} - \mathbf{R} \odot \mathbf{L}\|_{2}^{2} + \beta \varphi(\mathbf{R}) + \gamma \phi(\mathbf{L}),$$
 (2)

where  $\|\cdot\|_2$  is a  $\ell_2$ -norm.  $\beta$  and  $\gamma$  are trade-off parameters. The regularization terms  $\varphi(\mathbf{R})$  and  $\varphi(\mathbf{L})$ will be implicitly learned by deep networks to effectively suppress the perturbations. In addition to these implicit terms, we propose incorporating explicit constraints into the reflectance  $\mathbf{R}$  to suppress imaging noise and enhance texture details, and into the illumination L to reduce color distortion based on their degradation characteristics. Consequently, the objective function becomes:

$$L(\mathbf{R}, \mathbf{L}) = \frac{1}{2} \|\mathbf{I} - \mathbf{R} \odot \mathbf{L}\|_{2}^{2} + \beta \varphi(\mathbf{R}) + \gamma \phi(\mathbf{L}) + \mu \mathcal{S}(\mathcal{A}(\mathbf{I}) - \mathcal{A}(\mathbf{R})) + \frac{\lambda}{2} \|\mathbf{w} \odot \nabla \mathbf{L}\|_{2}^{2}, \quad (3)$$

where  $\mu$  and  $\lambda$  are trade-off parameters. Inspired by the Perona-Malik algorithm [19], we introduce  $\mathcal{A}(\cdot)$  and use  $\mathcal{S}(\cdot)$ , an  $\ell_1$ -norm, to enforce texture distribution consistency between  $\mathbf{I}$  and  $\mathbf{R}$ . At pixel location i,  $\mathcal{A}(\mathbf{I}_i) = \frac{1}{\tilde{\eta}_i} \sum_{j \in \eta_i} c(|\nabla \mathbf{I}_{i,j}|) \nabla \mathbf{I}_{i,j}$ , where  $\eta_i$  is the set of neighboring pixels around pixel i, and  $\tilde{\eta}_i$  is the number of neighboring pixels, set to 9.  $|\cdot|$  means gradient magnitude and  $\nabla$  is the gradient operator, such as Sobel operator [20]. The diffusion function  $c(|\nabla \mathbf{I}_{i,j}|)$  is:  $c(|\nabla \mathbf{I}_{i,j}|)$  $\exp[-(\frac{|\nabla \mathbf{I}_{i,j}|}{c})^2]$ , where s is a constant controlling the sensitivity of the diffusion process and is set as a learnable parameter in UnfoldIR. A small  $|\nabla \mathbf{I}_{i,j}|$  indicates a relatively smooth region, thereby intensifying the diffusion process and suppressing noise. Conversely, a large one implies the presence of important texture details, weakening the diffusion effect and enhancing texture preservation.

For the illumination map L, we propose a local smoothing constraint to mitigate color distortion. w is a gradient-aware weighting matrix and  $\mathbf{w} = \frac{1}{\exp|\nabla \mathbf{L}|}$ . This weighting matrix effectively preserves significant structural information while preventing over-smoothing.

#### UnfoldIR 3.2

#### 3.2.1 Model Optimization

We utilize the proximal gradient algorithm [21] to optimize Eq. (3), progressively suppressing the perturbations  $\hat{\mathbf{R}}$  and  $\hat{\mathbf{L}}$ , and ultimately obtaining the optimal Retinex components  $\mathbf{R}^*$  and  $\mathbf{L}^*$ :

$$\{\mathbf{R}^*, \mathbf{L}^*\} = \arg\min_{\mathbf{R}, \mathbf{L}} L(\mathbf{R}, \mathbf{L}). \tag{4}$$

 $\{{\bf R}^*,{\bf L}^*\} = \arg\min_{{\bf R},{\bf L}} L({\bf R},{\bf L}). \tag{4}$  The optimization process involves alternating updates of  ${\bf L}$  and  ${\bf R}$  over iterations. In the following, we select the  $k^{th}$  iteration  $(1 \le k \le K)$  to present the alternative solution process.

**Optimizing**  $L_k$ . The optimization function is partitioned to update the illumination map  $L_k$ :

$$\mathbf{L}_{k} = \arg\min_{\mathbf{L}} L(\mathbf{R}_{k-1}, \mathbf{L}) = \arg\min_{\mathbf{L}} \frac{1}{2} \|\mathbf{I} - \mathbf{R}_{k-1} \odot \mathbf{L}\|_{2}^{2} + \gamma \phi(\mathbf{L}) + \frac{\lambda}{2} \|\mathbf{w}_{k} \odot \nabla \mathbf{L}\|_{2}^{2}.$$
 (5)

The solution comprises two terms, that is, the gradient descent term and the proximal term. By introducing an auxiliary variable  $\hat{\mathbf{L}}_k$ , the two terms can be formulated as

$$\hat{\mathbf{L}}_{k} = \frac{1}{2} \| \mathbf{I} - \mathbf{R}_{k-1} \odot \hat{\mathbf{L}} \|_{2}^{2} + \frac{\gamma}{2} \| \hat{\mathbf{L}} - \mathbf{L}_{k-1} \|_{2}^{2} + \frac{\lambda}{2} \| \mathbf{w}_{k} \odot \nabla \hat{\mathbf{L}} \|_{2}^{2},$$
(6)

$$\mathbf{L}_k = \operatorname{prox}_{\phi}(\mathbf{R}_{k-1}, \hat{\mathbf{L}}_k),\tag{7}$$

where  $L_0$  and  $R_0$  are initialized following Uretinex-Net [9] and  $w_k$  is constructed based on  $L_{k-1}$ . Eq. (6) can be solved directly by equating its derivative to zero, while Eq. (7) will be replaced by a deep network in Sec. 3.2.2. The closed-form solution of  $\hat{\mathbf{L}}_k$  is

deep network in Sec. 3.2.2. The closed-form solution of 
$$\hat{\mathbf{L}}_k$$
 is 
$$\hat{\mathbf{L}}_k = \left(\mathbf{R}_{k-1}^2 + \lambda \mathbf{w}_k^2 \nabla^2 + \gamma \mathbf{1}\right)^{-1} \left(\mathbf{R}_{k-1} \mathbf{I} + \gamma \mathbf{L}_{k-1}\right),$$
 where  $\mathbf{1}$  is an all-ones matrix. (8)

**Optimizing**  $\mathbf{R}_k$ . The optimization function of  $\mathbf{R}_k$  is formulated as:

$$\mathbf{R}_{k} = \arg\min_{\mathbf{R}} L(\mathbf{R}, \mathbf{L}_{k}) = \arg\min_{\mathbf{R}} \frac{1}{2} \|\mathbf{I} - \mathbf{R} \odot \mathbf{L}_{k}\|_{2}^{2} + \beta \varphi(\mathbf{R}) + \mu \mathcal{S}(\mathcal{A}(\mathbf{I}) - \mathcal{A}(\mathbf{R})). \tag{9}$$

Same as the optimization rule for  $L_k$ , the gradient descent term and the proximal term are defined as:

$$\hat{\mathbf{R}}_k = \frac{1}{2} \|\mathbf{I} - \hat{\mathbf{R}} \odot \mathbf{L}_k\|_2^2 + \frac{\beta}{2} \|\hat{\mathbf{R}} - \mathbf{R}_{k-1}\|_2^2 + \mu \mathcal{S}(\mathcal{A}(\mathbf{I}) - \mathcal{A}(\hat{\mathbf{R}})), \tag{10}$$

$$\mathbf{R}_k = \operatorname{prox}_{\mathcal{O}}(\hat{\mathbf{R}}_k, \mathbf{L}_k). \tag{11}$$

The closed-form solution of  $\hat{\mathbf{R}}_k$  can be acquired similarly:

$$\hat{\mathbf{R}}_k = (\mathbf{L}_k^2 + \beta \mathbf{1} + \mathbf{Q_a})^{-1} (\mathbf{L}_k \mathbf{I} + \beta \mathbf{R}_{k-1} + \mathbf{Q_b}), \tag{12}$$

where  $\mathbf{Q_a} = \frac{\mu L_S}{\tilde{\eta}_i^2} \sum_{j \in \eta_i} c(|\nabla \mathbf{R}_{k-1}|)^2 \nabla^2$ ,  $\mathbf{Q_b} = \frac{\mu L_S}{\tilde{\eta}_i^2} \sum_{j \in \eta_i} c(|\nabla \mathbf{R}_{k-1}|) c(|\nabla \mathbf{R}_{k-2}|) \nabla^2 + \frac{\mu}{\tilde{\eta}_i} \sum_{j \in \eta_i} c(|\nabla \mathbf{R}_{k-1}|) \nabla \mathcal{S}_L(\frac{1}{\tilde{\eta}_i} \sum_{j \in \eta_i} c(|\nabla \mathbf{I}|) \nabla \mathbf{I} - c(|\nabla \mathbf{R}_{k-2}|) \nabla \mathbf{R}_{k-2})$ . Redundant subscripts are omitted.  $L_S$  is the Lipschitz constant.  $\mathcal{S}_L(\cdot)$  is the Lipschitz continuous gradient function of  $\mathcal{S}(\cdot)$ .

#### 3.2.2 Deep Unfolding Mechanism

We unfold the iterative solutions into a multi-stage network, UnfoldIR, with each step corresponding to a stage. As shown in Fig. 2, each stage has two modules: the reflectance-assisted illumination correction (RAIC) and illumination-guided reflectance enhancement (IGRE) modules.

**RAIC**. RAIC, derived from Eqs. (7) and (8), uses  $\hat{\mathcal{L}}(\cdot)$  and  $\mathcal{L}(\cdot)$  to compute the optimized result  $\hat{\mathbf{L}}$ and the refined illumination map L, respectively. Given  $\mathbf{R}_{k-1}$  and  $\mathbf{L}_{k-1}$ , we define  $\hat{\mathbf{L}}_k$  as follows:

$$\hat{\mathbf{L}}_k = \hat{\mathcal{L}}(\mathbf{R}_{k-1}, \mathbf{L}_{k-1}, \mathbf{I}) = \left(\mathbf{R}_{k-1}^2 + \lambda \mathbf{w}_k^2 \nabla^2 + \gamma \mathbf{1}\right)^{-1} \left(\mathbf{R}_{k-1} \mathbf{I} + \gamma \mathbf{L}_{k-1}\right), \tag{13}$$
 Eq. (13) retains the same formulation as Eq. (8), with originally fixed parameters to be learnable.

To refine the estimated illumination map  $\hat{\mathbf{L}}_k$ , we introduce a visual state space (VSS) module [22], denoted as  $VSS(\cdot)$ , to extract non-local features. This enables the network to adjust global illumination and promotes uniform lighting across the image. The design is particularly effective under backlit conditions, where the illumination often exhibits region-level heterogeneity. Under this formulation, the refined illumination map  $L_k$  is given by:

$$\mathbf{L}_k = \mathcal{L}(\hat{\mathbf{L}}_k, \mathbf{L}_{k-1}) = VSS(\hat{\mathbf{L}}_k, \mathbf{L}_{k-1}). \tag{14}$$

**IGRE**. in IGRE, the calculation of  $\hat{\mathbf{R}}_k$  relies on  $\hat{\mathcal{B}}(\cdot)$ , similar to Eq. (12) but with fixed parameters, including  $S_L(\cdot)$ , made learnable. Given  $\mathbf{L}_k$  and  $\mathbf{R}_{k-1}$ ,  $\hat{\mathbf{R}}_k$  can be calculated as

$$\hat{\mathbf{R}}_{k} = \hat{\mathcal{R}}(\mathbf{L}_{k}, \mathbf{R}_{k-1}, \mathbf{R}_{k-2}, \mathbf{I}) = (\mathbf{L}_{k}^{2} + \beta \mathbf{1} + \mathbf{Q}_{\mathbf{a}})^{-1} (\mathbf{L}_{k} \mathbf{I} + \beta \mathbf{R}_{k-1} + \mathbf{Q}_{\mathbf{b}}).$$
(15)

Refining the reflectance component  $\hat{\mathbf{R}}_k$  is inherently more complex. In addition to mitigating intrinsic imaging noise and recovering weakened texture details—both common in illumination-degraded scenarios—the refinement process must also account for interference introduced by the optimized illumination component L. To address these issues, we begin by unrolling Eq. (1), resulting in:

$$\mathbf{I} = \mathbf{R}_{HQ} \odot \mathbf{L}_{HQ} + \mathbf{R}_{HQ} \odot \hat{\mathbf{L}} + \hat{\mathbf{R}} \odot (\mathbf{L}_{HQ} + \hat{\mathbf{L}}). \tag{16}$$

In the third term of Eq. (16), the optimization of L—that is, the light-up process—aims to suppress the perturbation L, but it can also amplify noise hidden in the dark scenes. Therefore, the optimization of R must concurrently address noise suppression during the illumination recovery phase.

To this end, we introduce a frequency-aware VSS (FVSS) module,  $FVSS(\cdot)$ , which conditions on the estimated illumination map  $\mathbf{L}_k$ . The FVSS module is designed to globally align similar texture patterns and enables lightly degraded regions to guide the enhancement of details in more severely degraded areas under illumination-aware guidance. The preliminary refined result  $\hat{\mathbf{R}}_k$  is defined as

$$\widetilde{\mathbf{R}}_k = FVSS(\widehat{\mathbf{R}}_k, \mathbf{L}_k) = VSS(conca(\mathbf{R}'_k, \boldsymbol{\sigma_L} \mathbf{R}'_k + \boldsymbol{\mu_L})), \tag{17}$$

$$\mathbf{R}_{k}' = IDWT(VSS(DWT(\hat{\mathbf{R}}_{k}))), \boldsymbol{\sigma}_{\mathbf{L}} = conv3_{\boldsymbol{\sigma}}(\mathbf{L}_{k}), \boldsymbol{\mu}_{\mathbf{L}} = conv3_{\boldsymbol{\mu}}(\mathbf{L}_{k}),$$
(18)

where  $conca(\cdot)$  is concate.  $DWT(\cdot)$  and  $IDWT(\cdot)$  means discrete wavelet transform and its inverse.  $conv3(\cdot)$  is a  $3\times3$  convolution. In the preliminary refinement, the wavelet transform decomposes the reflectance component into frequency bands, which are then processed by a VSS module for restoration within each band. The reconstructed reflectance  $\mathbf{R}'_k$ , along with its illumination-aware modulation, is then passed through another VSS for enhanced texture alignment and restoration.

To further improve noise suppression and texture enhancement, we integrate the FVSS module into a high-order ordinary differential equation (ODE) framework, specifically a second-order Runge-Kutta (RK2) method. Compared with the traditional residual network structure—essentially a first-order Euler discretization of an ODE with non-negligible truncation error [23]—the RK2 framework provides more accurate numerical solutions, which better accommodate the fine-grained requirements of noise removal and detail enhancement. To increase flexibility, we incorporate a learnable weighted

Table 1: Results on the LLIE task. The best two results are in **red** and **blue** fonts, respectively.

Methods	Sources	Effic	ciency		LOL				LOL-v	2-real			LOL-v2-s	ynthetic	:
Methous	Sources	Para. ↓	FLOPs ↓	PSNR ↑	SSIM ↑	FID↓	$BIQE \downarrow$	PSNR ↑	SSIM ↑	FID↓	$\text{BIQE} \downarrow$	PSNR ↑	SSIM ↑	FID↓	BIQE ↓
URetinex [9]	CVPR22	0.36	233.09	21.33	0.835	85.59	30.37	20.44	0.806	76.74	28.85	24.73	0.897	33.25	33.46
UFormer [24]	CVPR22	5.20	10.68	16.36	0.771	166.69	41.06	18.82	0.771	164.41	40.36	19.66	0.871	58.69	39.75
Restormer [25]	CVPR22	26.13	144.25	22.43	0.823	78.75	33.18	19.94	0.827	114.35	37.27	21.41	0.830	46.89	35.06
SNR-Net [26]	CVPR22	4.01	26.35	24.61	0.842	66.47	28.73	21.48	0.849	68.56	28.83	24.14	0.928	30.52	33.47
SMG [27]	CVPR23	14.02	17.55	24.82	0.838	69.47	30.15	22.62	0.857	71.76	30.32	25.62	0.905	23.36	29.35
Diff-Retinex [8]	ICCV23	56.88	198.16	21.98	0.852	51.33	19.62	20.17	0.826	46.67	24.18	24.30	0.921	28.74	26.35
MRQ [28]	ICCV23	8.45	20.66	25.24	0.855	53.32	22.73	22.37	0.854	68.89	33.61	25.54	0.940	20.86	25.09
IAGC [29]	ICCV23	_	_	24.53	0.842	59.73	25.50	22.20	0.863	70.34	31.70	25.58	0.941	21.38	30.32
DiffIR [30]	ICCV23	27.80	35.32	23.15	0.828	70.13	26.38	21.15	0.816	72.33	29.15	24.76	0.921	28.87	27.74
CUE [10]	ICCV23	0.25	157.32	21.86	0.841	69.83	27.15	21.19	0.829	67.05	28.83	24.41	0.917	31.33	33.83
GSAD [31]	NIPS23	17.17	670.33	23.23	0.852	51.64	19.96	20.19	0.847	46.77	28.85	24.22	0.927	19.24	25.76
AST [32]	CVPR24	19.90	13.25	21.09	0.858	87.67	21.23	21.68	0.856	91.81	25.17	22.25	0.927	37.19	28.78
RetiMamba [33]	ArXiv	3.59	37.98	24.03	0.827	75.33	16.28	22.45	0.844	56.96	21.76	25.89	0.934	20.17	16.29
MambaIR [34]	ECCV24	4.30	60.66	22.23	0.863	63.39	20.17	21.15	0.857	56.09	24.46	25.75	0.937	19.75	20.37
Mamballie [35]	NIPS24	2.28	20.85	23.24	0.861	_	_	22.95	0.847	_	_	25.87	0.940	_	_
Reti-Diff [1]	ICLR25	26.11	87.63	25.35	0.866	49.14	17.75	22.97	0.858	43.18	23.66	27.53	0.951	13.26	15.77
CIDNet [36]	CVPR25	1.88	7.57	23.50	0.900	46.69	14.77	24.11	0.871	48.04	18.45	25.71	0.942	18.60	15.87
UnfoldIR-t	Ours	0.09	0.86	21.08	0.858	66.82	28.73	20.73	0.836	66.04	24.19	24.49	0.920	29.11	28.83
UnfoldIR-s	Ours	0.35	2.00	22.57	0.897	50.37	15.50	21.64	0.876	44.65	18.12	24.92	0.952	20.05	15.66
UnfoldIR	Ours	3.45	11.83	24.41	0.911	43.08	13.21	22.99	0.887	34.86	17.84	27.55	0.959	18.07	15.08
		Pho To				Marin .		Altho -		Altre .		- State	700 PM	-	the Control
										3/2					
									2 3		2		2		2
	1				1		7				A	75		TH	1
				THE PARTY NAMED IN				- PES	1	J. C.	See 1		area or	1	Parks a
Input	Input EnGAN RetiM		Iamba	Mamb	aLLIE	R	eti-Diff	(	CIDNet		Ours	(	Ground	Truth	

Figure 3: Visual results on the low-light image enhancement task.

gating mechanism  $g_w$  into the RK2 solver. The final refined reflectance  $\mathbf{R}_k$  is computed as:

$$\mathbf{R}_k = \mathcal{R}(\hat{\mathbf{R}}_k, \mathbf{L}_k) = \hat{\mathbf{R}}_k + g_w \widetilde{\mathbf{R}}_k + (1 - g_w) \widetilde{\mathbf{R}}_k^1, \tag{19}$$

$$g_w = S(\boldsymbol{\sigma}_q conv3(\widetilde{\mathbf{R}}_k, \widetilde{\mathbf{R}}_k^1) + \boldsymbol{\mu}_q), \ \widetilde{\mathbf{R}}_k = FVSS(\widehat{\mathbf{R}}_k, \mathbf{L}_k), \ \widetilde{\mathbf{R}}_k^1 = FVSS(\widehat{\mathbf{R}}_k + \widetilde{\mathbf{R}}_k, \mathbf{L}_k), \ (20)$$

where  $S(\cdot)$  is the softmax operator.  $\sigma_g$  and  $\mu_g$  are two learnable parameters in  $g_w$ . As the stages progress, UnfoldIR jointly promotes illumination smoothness, noise removal, and texture enhancement, while mitigating the intrinsic conflict between the recovery of illumination and reflectance.

ISIC Loss. Leveraging the multi-stage nature of DUNs, we design an inter-stage information consistency (ISIC) loss,  $L_{ISIC}$ , to enhance stability during the final stages of restoration. This loss ensures that small variations in illumination do not compromise essential reflectance details in the restored image, and vice versa. Hence,  $L_{ISIC}$  contributes to reducing color distortion in the illumination map and preserving structural details in the reflectance component.  $L_{ISIC}$  is defined as:

$$L_{ISIC} = \|\mathbf{R}_k \odot \mathbf{L}_{k-1} - \mathbf{R}_k \odot \mathbf{L}_k\|_2 + \|\nabla(\mathbf{R}_{k-1} \odot \mathbf{L}_k) - \nabla(\mathbf{R}_k \odot \mathbf{L}_k)\|_1. \tag{21}$$

Eq. (21) encourages consistency in the restored image rather than in the individual Retinex components, relaxing the constraint and allowing small changes that do not degrade essential information. Notably,  $L_{ISIC}$  is a DUN-specific constraint, which enables broader applicability of UnfoldIR in the unsupervised settings. To ensure the model remains responsive to stage-specific dynamics, we apply  $L_{ISIC}$  only during the final two stages. Except  $L_{ISIC}$ , all other loss functions follow Uretinex [9].

### 4 Experiments

**Experimental setup**. Our UnfoldIR is implemented in PyTorch on RTX4090 GPUs and is optimized by Adam with momentum terms (0.9, 0.999). Random rotation and flips are used for augmentation. The stage number K is set as 3. Other parameters inherited from traditional methods are optimized in a learnable manner. For efficiency, different stages in UnfoldIR share the same weights. For fairness, We abandon GT-mean and all compared results are obtained using the official codes.

### 4.1 Compative Evaluation

**Low-light image enhancement.** Following Reti-Diff [1], we report results on LOL-v1 [47], LOL-v2-real [48], and LOL-v2-syn [48] with four metrics: PSNR, SSIM, FID [49], and BIQE [50]. Higher PSNR and SSIM values, and lower FID and BIQE scores, indicate better performance. we compare UnfoldIR with state-of-the-art techniques on  $256 \times 256$  resolution inputs. The quantitative results are presented in Table 1, where our method achieves top performance across all datasets while maintaining competitive efficiency. Besides, we introduce two lightweight variants, UnfoldIR-t and UnfoldIR-s, which outperform existing lightweight models, highlighting the flexibility and scalability of UnfoldIR.

Table 2: Results on the UIE task. Table 3: Results on the BIE task. Table 4: Results on the FIE task.

Methods	Sources	es $PSNR \uparrow SSIM \uparrow UCIQE \uparrow UIQM \uparrow$			Methods Sources $BAID$ PSNR $\uparrow$ SSIM $\uparrow$ LPIPS $\downarrow$ FID						Methods	Resources	BIQE ↓	Fundus ↓ CLIPIQA ↑	↑ FID↓	
S-uwnet [37]	AAAI21	18.28	0.855	0.544	2.942	EnGAN [7]	TIP21	17.96	0.819	0.182	43.55	SNR-Net [26]	CVPR22	6.144	0.557	79.284
PUIE [38]	ECCV22		0.882	0.566	3.021	URetinex [9]	CVPR22		0.845	0.206	42.26	URetinex [9]		12.158		33.347
U-shape [39]	TIP23	22.91	0.905	0.592	2.896	CLIP-LIT [3]	ICCV23		0.853	0.159	37.30	SCI [43]		23.527		85.175
PUGAN [40]	TIP23	23.05	0.897	0.608	2.902	Diff-Retinex [8]			0.861	0.160	38.07	MIRNetV2 [44]				47.607
ADP [41]	IJCV23	22.90	0.892	0.621	3.005	DiffIR [30]	ICCV23		0.835	0.175	40.35	FourLLE [45]	MM23	7.741	0.508	28.736
NU2Net [2]	AAAI23	22.38	0.903	0.587	2.936		CVPR24		0.851	0.156	32.47	CUE [10]	ICCV23	11.721		111.336
	CVPR24		0.908	0.602	2.981		ECCV24		0.874	0.153		NeRCO [46]		17.256		95.241
MambaIR [34]			0.916	0.617	2.991	RAVE [42]	ECCV24		0.872	0.096	64.89	Reti-Diff [1]	ICLR25	10.788		27.637
Reti-Diff [1]	ICLR25	24.12	0.910	0.631	3.088	Reti-Diff [1]	ICLR25	23.19	0.876	0.147	27.47	CIDNet [36]	CVPR25	10.663		41.089
UnfoldIR	Ours	24.16	0.930	0.651	3.248	UnfoldIR	Ours	24.83	0.890	0.091	34.64	UnfoldIR	Ours	6.719	0.572	27.398
											100					
Input		EnG	AN		S-uwnet	PU	TE C		NU2Net		Pai	ti-Diff	Ours		Ground	Truth
Input						esults on t								í.	Glound	Hum
				-		The state of the s				15				15	Page 1	
1117					1											

GAN URetinex CLIP-LIT RAVE Reti-Diff Our Figure 5: Visual results on the backlit image enhancement task.

Ground Truth

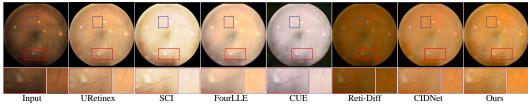


Figure 6: Visual results on the fundus image enhancement task.

Visualizations are shown in Fig. 3, where our method demonstrates superiority in producing visually coherent restorations with accurately corrected illumination and enhanced textures.

**Underwater image enhancement.** We evaluate our method on *UIEB* [51] with two metrics, UCIQE [52] and UIQM [53], where higher values indicate better results. The quantitative results are shown in Table 2, where our method achieves a leading place. Visualizations in Fig. 4 verify our effectiveness in correcting color aberrations and enhancing fine texture in underwater scenes.

**Backlit image enhancement**. Following CLIP-LIT [3], we train our network using *BAID* dataset [54] and evaluate with PSNR, SSIM, LPIPS [55], and FID [49]. The results in Table 3 show that our method consistently outperforms existing approaches. Furthermore, visualizations in Fig. 5 demonstrate our effectiveness in detail reconstruction and color correction under challenging backlit conditions.

**Fundus image enhancement**. Same as Reti-Diff [1], we test performance on the *Fundus* dataset by employing models pretrained on *LOL-v2-syn* and evaluate performance on BIQE, CLIPIQA [56], and FID. A larger CLIPIQA indicates a better performance. As depicted in Table 4 and Fig. 6, our proposed UnfoldIR consistently outperforms existing methods, both qualitatively and quantitatively.

**Real-world illumination degradation image restoration**. We select four real-world datasets: *DICM* [57], *LIME* [58], *MEF* [59], and *VV* [60]. Following [61], we employ a model pretrained on *LOL-v2-syn* for inference and evaluate with PI [62] and NIQE [63], where lower values indicate better results. As shown in Table 5, our method achieves a leading place across all datasets.

#### 4.2 Ablation Study and Further Analysis

Effect of basic network components in UnfoldIR. We conduct experiments on Table 6 to verify the effectiveness of the core components in UnfoldIR, including RAIC, IGRE, and  $L_{ISIC}$ . We first replace the VSS module, *i.e.*,  $\mathcal{L}$ , with three alternatives:  $\mathcal{L}_1$ ,  $\mathcal{L}_2$ , and  $\mathcal{L}_3$ .  $\mathcal{L}_1$  is a transformer block from Reti-Diff [1] with comparable parameters;  $\mathcal{L}_2$  and  $\mathcal{L}_3$  correspond to more advanced VSS modules taken from RetiMamba [33] and Mamballie [35]. As shown in Table 6, VSS, compared to them, can better balance performance and efficiency. Then we assess the effect of individual

Table 5: Results on the real-world IDIR task.

Methods	Sources	D	ICM	L	IME	N	1EF		VV
Methods	Sources	PI↓	$NIQE \downarrow$	PI↓	NIQE↓	PI↓	NIQE↓	PI↓	NIQE ↓
EnGAN [7]	TIP21	4.173	4.064	3.669	4.593	4.015	4.705	3.386	4.047
KinD++ [64]	IJCV21	3.835	3.898	3.785	4.908	4.016	4.557	3.773	3.822
SNR-Net [26]	CVPR22	3.585	4.715	3.753	5.937	3.677	6.449	3.503	9.506
DCC-Net [65]	CVPR22	3.630	3.709	3.312	4.425	3.424	4.598	3.615	3.286
UHDFor [66]	ICLR23	3.684	4.575	4.124	4.430	3.813	4.231	3.319	4.330
PairLIE [67]	CVPR23	3.685	4.034	3.387	4.587	4.133	4.065	3.334	3.574
GDP [68]	CVPR23	3.552	4.358	4.115	4.891	3.694	4.609	3.431	4.683
Reti-Diff [1]	ICLR25	2.976	3.523	3.111	4.128	2.876	3.554	2.651	2.540
CIDNet [36]	CVPR25	3.045	3.796	3.146	4.132	2.683	3.568	2.826	3.218
UnfoldIR	Ours	2.952	3.381	3.085	4.099	2.722	3.387	2.553	2.306

Table 6: Ablation study in the LLIE task.

Datasets	Metrics		Effect of RAIC				ct of IGR			of $L_{ISIC}$	UnfoldIR
	Wictires	$\mathcal{L}_1(\bullet) \to \mathcal{L}(\bullet)$	$\mathcal{L}_2(\bullet) \to \mathcal{L}(\bullet)$	$\mathcal{L}_3(\bullet) \to \mathcal{L}(\bullet)$	w/o RK2	w/o $g_w$	w/o $L_k$	$VSS \to FVSS$	w/o $L_{ISIC}$	$L^1_{ISIC}$ $L^2_{ISIC}$	(Ours)
L-v2-s	PSNR ↑ SSIM ↑	26.85 0.943	27.51 0.956	<b>27.58</b> 0.958	27.31 0.954	27.23 0.955	26.93 0.946	26.74 0.942	27.37 0.943	27.52 27.50 0.954 0.952	
L-v2-r	PSNR ↑ SSIM ↑	22.11 0.882	22.87 <b>0.892</b>	22.96 0.888	22.78 0.883	22.87 0.885	22.06 0.876	22.06 0.879	22.26 0.874	22.82 22.88 0.884 0.881	

Table 7: Performance of UnfoldIR with different restora-Table 8: Experiments on the unsupervised tion models and stage numbers.

Datasets Metrics Restoration models Stage numbers UnfoldIR CM1 CM2 CM3 CM4 CM5 K=2 K=4 K=5 IDIRM,K=3	NeRCO [46] CLIP-LIT [3] LightenDiff [69] UnfoldIR ICCV23 ICCV23 ECCV24 Ours
$L-v2-s$   PSNR $\uparrow$   26.32 27.13 27.08 26.73 27.25   26.89 28.32 <b>28.50</b>   27.55   SSIM $\uparrow$   0.938 0.946 0.944 0.949 0.952   0.949 0.968 <b>0.972</b>   0.959	19.14 16.18 20.44 <b>20.47</b> 0.743 0.792 0.843 <b>0.852</b>
$L$ - $v$ 2- $r$   PSNR $\uparrow$   21.23 21.94 22.12 22.30 22.17   22.57 23.86 <b>24.15</b>   22.99   SSIM $\uparrow$   0.860 0.873 0.872 0.870 0.882   0.881 0.893 <b>0.896</b>   0.887	19.23 17.06 <b>19.32</b> 19.28 0.671 0.589 0.684 <b>0.686</b>

components within IGRE, including RK2, the gated mechanism  $g_w$ , the illumination-aware texture enhancement mechanism, and the proposed FVSS module. Additionally, we evaluate the contribution of our  $L_{ISIC}$  by (1) removing it entirely; (2) conducting consistency constraints directly to the reflectance and illumination components instead of the restored image  $(L_{ISIC}^1)$ ; and (3) enforcing consistency across all stages rather than only the final ones  $((L_{ISIC}^2))$ . These results highlight the importance of the placement and formulation of ISIC loss.

Other configurations in UnfoldIR. We explore how the restoration model and stage number influence the performance. As shown in Table 7, we first conduct breakdown ablations of our IDIRM model by removing two proposed explicit constraints (CM1), removing only the explicit restriction of  $\mathcal{R}$  (CM2) or  $\mathcal{L}$  (CM3). CM4 and CM5 correspond to the restoration model used in URetinex [9] and CUE [10], further validating the effectiveness of our proposed restoration model. Besides, we explore the influence of stage number and discover that increasing the stage number can improve performance. We discover that UnfoldIR achieves cutting-edge performance when K reaches 3, thus setting K=3.

**Potential applications of UnfoldIR.** We further explore the potential of the UnfoldIR framework, including its adaptability to different supervision paradigms, its positive impact on existing methods, and its benefits for downstream tasks. First, as shown in Table 8, we extend UnfoldIR to the unsupervised setting following LightenDiff [69]. Notably, even when supervised solely by the proposed  $L_{\rm ISIC}$  loss, our method achieves performance comparable to existing SOTA unsupervised methods, highlighting the effect of  $L_{\rm ISIC}$  as a framework-specific loss function. Next, as shown in Table 9a, we examine how UnfoldIR can enhance existing restoration methods by (1) directly using UnfoldIR with our pretrain model as a refiner method by inputting the enhanced result of existing methods into UnfoldIR("-R"), and (2) integrating our UnfoldIR after the existing methods and then end-to-end training the combined network("+"). We observe performance gains from both settings. Finally, in Table 9b, we explore how to better facilitate the performance of downstream tasks. "Comb." refers to combining enhanced outputs from multiple UnfoldIR stages as a form of enhancement-based data augmentation, thus improving downstream tasks. When further combining this strategy with the bi-level optimization (BLO) framework, we can observe further performance gains.

# 4.3 User Study and Downstream Tasks

**User Study**. We conduct a user study to evaluate the visual quality of IDIR methods, including LLIE (L-v1 and L-v2), UIE (UIEB), and BIE (BAID). In this study, 29 participants rated enhanced images on a scale of 1 (worst) to 5 (best) based on four criteria: (1) the presence of underexposed or overexposed regions; (2) the degree of color distortion; (3) the occurrence of unwanted noise or

Table 9: Potential applications of UnfoldIR. In (a), "-R" and "+" means refining the method with UnfoldIR and integrating UnfoldIR with the method for end-to-end training. In (b), "Comb." means aggregating the outputs of each stage in UnfoldIR, while BLO is short for bi-level optimization [4].

(a) Enhancing restoration performance. (b) I

(	b)	Promoting	downstream	tasks
---	----	-----------	------------	-------

Datasets	Metrics	Reti-Diff	Reti-Diff-R	Reti-Diff+	CIDNet	CIDNet-R	CIDNet+	Tasks	Metrics	UnfoldIR	+ Comb.	+ BLO	+ Comb. & BLO
L-v2-s	PSNR ↑ SSIM ↑	27.53	27.55	28.82	25.71	25.86	26.55	Detection	mAP↑	78.9	79.3	79.8	80.3
L-V2-3	SSIM ↑	0.951	0.958	0.973	0.942	0.946	0.958	Seman. Seg	. mIoU ↑	62.5	63.2	63.3	64.3
L-v2-r	PSNR ↑ SSIM ↑	22.97	23.08	23.67	24.11	24.10	24.37	Concealed	$F_{\beta} \uparrow$	0.716	0.724	0.726	0.731
L-V2-1	SSIM ↑	0.858	0.862	0.872	0.871	0.875	0.883	Object Seg.	$E_{\phi} \uparrow$	0.875	0.882	0.882	0.886

Table 10: User study.

Table 11: Low-light image detection on *ExDark*.

Methods	L-v1	L-v2	UIEB	BAID	Methods (AP)	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Motor	People	Table	Mean
Uretinex	3.69	3.63		3.13	Baseline	74.7	64.9	70.7	84.2	79.7	47.3	58.6	67.1	64.1	66.2	73.9	45.7	66.4
Restormer	3.17	3.20	_	3.08	RetinexNet	72.8	66.4	67.3	87.5	80.6	52.8	60.0	67.8	68.5	69.3	71.3	46.2	67.5
SNR	3.73	3.88	_	3.25	KinD	73.2	67.1	64.6	86.8	79.5	58.7	63.4	67.5	67.4	62.3	75.5	51.4	68.1
CUE	3.30	3.57	_	_	MIRNet	74.9	69.7	68.3	89.7	77.6	57.8	56.9	66.4	69.7	64.6	74.6	53.4	68.6
AST	3.58	3.65	3.82	3.37	RUAS	75.7	71.2	73.5	90.7	80.1	59.3	67.0	66.3	68.3	66.9	72.6	50.6	70.2
MambaIR	3.70	3.77	4.03	3.45	SCI	73.4	68.0	69.5	86.2	74.5	63.1	59.5	61.0	67.3	63.9	73.2	47.3	67.2
Reti-Diff	3.89	4.13	4.18	3.60	SNR-Net	78.3	74.2	74.5	89.6	<b>82.7</b>	66.8	66.3	62.5	74.7	63.1	73.3	57.2	71.9
CIDNet	3.67	3.82	_	_	Reti-Diff	82.0	<b>77.9</b>	<b>76.4</b>	92.2	83.3	69.6	67.4	74.4	75.5	74.3	78.3	<b>57.9</b>	75.8
Ours	4.12	4.17	4.34	3.75	Ours	87.5	81.0	78.2	86.4	74.6	76.4	80.1	80.8	83.4	84.1	69.1	65.0	<b>78.9</b>

Table 12: Low-light semantic segmentation, where images are Table 13: Low-light concealed obdarkened by [70]. ject segmentation.

Methods (IoU)	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Dog	Horse	People	Mean	Methods	$M\downarrow$	$F_{\beta} \uparrow$	$E_{\phi} \uparrow$	$S_{\alpha} \uparrow$
Baseline	43.5	36.3	48.6	70.5	67.3	46.6	11.2	42.4	56.7	57.8	48.1	Baseline	0.049	0.631	0.818	0.762
RetinexNet	48.6	41.7	51.7	77.6	68.3	52.7	15.8	46.3	60.2	62.3	52.5	RetinexNet	0.041	0.663	0.847	0.789
KinD	51.3	40.2	53.2	76.8	69.4	50.8	14.6	47.3	60.3	60.9	52.5	KinD	0.038	0.670	0.855	0.793
MIRNet	50.3	42.9	47.4	73.6	62.7	50.4	15.8	46.3	61.0	63.3	51.4	MIRNet	0.036	0.701	0.860	0.800
RUAS	53.0	37.3	50.4	71.3	72.3	47.6	15.9	50.8	63.6	60.8	52.3	RUAS	0.037	0.707	0.866	0.805
SCI	54.5	46.3	57.2	78.4	73.3	49.1	22.8	49.0	62.1	66.9	56.0	SNR-Net	0.035	0.708	0.857	0.807
SNR-Net	57.7	48.6	59.5	81.3	74.8	50.2	24.4	50.7	64.3	68.7	58.0	SCI	0.036	0.702	0.867	0.802
Reti-Diff	59.8	51.5	62.1	85.5	<b>76.6</b>	<b>57.7</b>	28.9	56.3	66.2	73.4	61.8	Reti-Diff	0.034	0.708	0.867	0.809
Ours	60.2	51.8	61.3	84.7	<b>78.5</b>	<b>58.8</b>	30.2	57.5	66.8	75.2	62.5	Ours	0.033	0.716	0.875	0.807

artifacts; and (4) the preservation of essential structural details. Each low-light image was displayed alongside its enhanced version with the method's name concealed. As shown in Table 10, our method outperforms across all four datasets, verifying its effectiveness in producing visually pleasing results.

**Low-light Object Detection**. Enhanced images are expected to facilitate subsequent tasks. We first evaluate the impact on low-light object detection to test this hypothesis. Following the protocol in [1], all compared methods are assessed on *ExDark* [71] using YOLO. "Baseline" corresponds to the original low-quality images without enhancement. As shown in Table 11, our UnfoldIR outperforms competing methods, confirming its effectiveness in enhancing high-level vision performance.

Low-light Image Segmentation. We also performed segmentation tasks by retraining segmentation models for each enhancement method, following that employed for detection. (1) For semantic segmentation, following [1], we apply image darkening to samples from VOC [72] and utilize Mask2Former [73] to segment the enhanced images, evaluating the performance using Intersection over Union (IoU). As shown in Table 12, our approach achieves a leading place across most classes. (2) Besides, we explore concealed object segmentation, a challenging task aimed at delineating objects with visual similarity to their backgrounds. This evaluation was conducted on COD10K [74]. We similarly apply image darkening and employ RUN [75] to segment the enhanced images. Performance was assessed using four metrics: mean absolute error (M), adaptive F-measure  $(F_{\beta})$ , mean E-measure  $(E_{\phi})$ , and structure measure  $(S_{\alpha})$ . As reported in Table 13, our method outperforms existing methods.

## 5 Conclusions

In this paper, we propose a novel DUN-based method, UnfoldIR, for IDIR tasks. UnfoldIR introduces a new IDIR model with specifically designed regularization terms for smoothing illumination and enhancing texture. By unfolding into a multistage network, we get RAIC and IGRE modules in each stage. RAIC employs VSS to attract non-local features for color correction, while IGRE introduces a frequency-aware VSS to globally align similar textures, enhancing details. We also propose an ISIC loss to maintain network stability in the final stages. Abundant experiments verify our effectiveness.

### References

- [1] Chengyu Fang, Yulun Zhang, Tian Ye, Kai Li, Longxiang Tang, Zhenhua Guo, Xiu Li, and Sina Farsiu. Reti-diff: Illumination degradation image restoration with retinex-based latent diffusion model. *ICLR*, 2025. 1, 3, 6, 7, 8, 9
- [2] Chunle Guo, Ruiqi Wu, Xin Jin, Linghao Han, Weidong Zhang, Zhi Chai, and Chongyi Li. Underwater ranker: Learn which is better and how to be better. In *AAAI*, volume 37, pages 702–709, 2023. 7
- [3] Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *ICCV*, pages 8094–8103, 2023. 7, 8
- [4] Chunming He, Kai Li, Guoxia Xu, Jiangpeng Yan, Longxiang Tang, Yulun Zhang, Yaowei Wang, and Xiu Li. Hqg-net: Unpaired medical image enhancement with high-quality guidance. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 1, 9
- [5] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, pages 2782–2790, 2016. 1, 3
- [6] Neng-Tsann Ueng and Louis L Scharf. The gamma transform: A local time-frequency analysis method. In *ACSSC*, volume 2, pages 920–924. IEEE, 1995. 1
- [7] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021. 1, 7, 8
- [8] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *ICCV*, pages 12302–12311, 2023. 1, 6, 7
- [9] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinexnet: Retinex-based deep unfolding network for low-light image enhancement. In *CVPR*, pages 5901–5910, 2022. 2, 3, 4, 6, 7, 8
- [10] Naishan Zheng, Man Zhou, Yanmeng Dong, Xiangyu Rui, Jie Huang, Chongyi Li, and Feng Zhao. Empowering low-light image enhancer through customized learnable priors. In *ICCV*, pages 12559–12569, 2023. 2, 3, 6, 7, 8
- [11] Heng-Da Cheng and XJ Shi. A simple and effective histogram equalization approach to image enhancement. *Digital signal processing*, 14(2):158–170, 2004. 3
- [12] Chunming He, Xiaobo Wang, Lizhen Deng, and Guoxia Xu. Image threshold segmentation based on glle histogram. In *CPSCom*, pages 410–415. IEEE, 2019.
- [13] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE transactions on image processing*, 22(3):1032–1041, 2012.
- [14] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In CVPR, pages 10561–10570, 2021. 3
- [15] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 3
- [16] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In CVPR, pages 6849–6857, 2019.

- [17] Guoxia Xu, Chunming He, Hao Wang, Hu Zhu, and Weiping Ding. Dm-fusion: Deep model-driven network for heterogeneous image fusion. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 3
- [18] Chunming He, Kai Li, Guoxia Xu, Yulun Zhang, Runze Hu, Zhenhua Guo, and Xiu Li. Degradation-resistant unfolding network for heterogeneous image fusion. In *ICCV*, pages 12611–12621, 2023. 3, 16
- [19] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639, 1990. 4
- [20] Irwin Sobel, Gary Feldman, et al. A 3x3 isotropic gradient operator for image processing. Stanford Artificial Project, 1968:271–272, 1968. 4
- [21] Chengyu Fang, Chunming He, Fengyang Xiao, Yulun Zhang, Longxiang Tang, Yuelin Zhang, Kai Li, and Xiu Li. Real-world image dehazing with coherence-based label generator and cooperative unfolding network. *NeurIPS*, 2024. 4
- [22] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. Vmamba: Visual state space model. *NeurIPS*, 37:103031–103063, 2024. 5
- [23] Ee Weinan. A proposal on machine learning via dynamical systems. *Commun. Math. Stat.*, 5(1):1–11, 2017. 5
- [24] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In CVPR, pages 17683– 17693, 2022. 6
- [25] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In CVPR, pages 5728–5739, 2022. 6
- [26] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In CVPR, pages 17714–17724, 2022. 6, 7, 8
- [27] Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. Low-light image enhancement via structure modeling and guidance. In *CVPR*, pages 9893–9903, 2023. 6
- [28] Yunlong Liu, Tao Huang, Weisheng Dong, Fangfang Wu, Xin Li, and Guangming Shi. Low-light image enhancement with multi-stage residue quantization and brightness-aware attention. In *ICCV*, pages 12140–12149, 2023. 6
- [29] Yinglong Wang, Zhen Liu, Jianzhuang Liu, Songcen Xu, and Shuaicheng Liu. Low-light image enhancement with illumination-aware gamma correction and complete image modelling network. In *ICCV*, pages 13128–13137, 2023. 6
- [30] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *ICCV*, 2023. 6, 7
- [31] HOU Jinhui, Zhiyu Zhu, Junhui Hou, LIU Hui, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. In *NeurIPS*, 2023. 6
- [32] Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, and Jufeng Yang. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2952–2963, June 2024. 6, 7
- [33] Jiesong Bai, Yuhao Yin, Qiyuan He, Yuanxian Li, and Xiaofeng Zhang. Retinexmamba: Retinex-based mamba for low-light image enhancement. *arXiv preprint arXiv:2405.03349*, 2024. 6, 7
- [34] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In ECCV, 2024. 6, 7

- [35] Jiangwei Weng, Zhiqiang Yan, Ying Tai, Jianjun Qian, Jian Yang, and Jun Li. Mamballie: Implicit retinex-aware low light enhancement with global-then-local state space. *NeurIPS*, 2024. 6, 7
- [36] Qingsen Yan, Yixu Feng, Cheng Zhang, Pei Wang, Peng Wu, Wei Dong, Jinqiu Sun, and Yanning Zhang. You only need one color space: An efficient network for low-light image enhancement. *CVPR*, 2025. 6, 7, 8
- [37] Ankita Naik, Apurva Swarnakar, and Kartik Mittal. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In AAAI, volume 35, pages 15853–15854, 2021. 7
- [38] Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In ECCV, pages 465–482. Springer, 2022. 7
- [39] Lintao Peng, Chunli Zhu, and Liheng Bian. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 2023. 7
- [40] Runmin Cong, Wenyu Yang, Wei Zhang, Chongyi Li, Chun-Le Guo, Qingming Huang, and Sam Kwong. Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators. *IEEE Transactions on Image Processing*, 2023. 7
- [41] Jingchun Zhou, Qian Liu, Qiuping Jiang, Wenqi Ren, Kin-Man Lam, and Weishi Zhang. Underwater camera: Improving visual perception via adaptive dark pixel prior and color correction. *International Journal of Computer Vision*, pages 1–19, 2023. 7
- [42] Tatiana Gaintseva, Martin Benning, and Gregory Slabaugh. Rave: Residual vector embedding for clip-guided backlit image enhancement. In *ECCV*, pages 412–428. Springer, 2024. 7
- [43] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *CVPR*, pages 5637–5646, 2022. 7
- [44] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. *TPAMI*, 45(2):1934–1948, 2022. 7
- [45] Chenxi Wang, Hongjun Wu, and Zhi Jin. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *ACM MM*, pages 7459–7469, 2023. 7
- [46] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *ICCV*, pages 12918–12927, 2023. 7, 8
- [47] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 6
- [48] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. 6
- [49] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *NeurIPS*, 30, 2017. 6, 7
- [50] Anush Krishna Moorthy and Alan Conrad Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal processing letters*, 17(5):513–516, 2010. 6
- [51] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2019. 7
- [52] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015. 7

- [53] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2015. 7
- [54] Xiaoqian Lv, Shengping Zhang, Qinglin Liu, Haozhe Xie, Bineng Zhong, and Huiyu Zhou. Backlitnet: A dataset and network for backlit image enhancement. *Computer Vision and Image Understanding*, 218:103403, 2022. 7
- [55] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In CVPR, pages 586–595, 2018.
- [56] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In AAAI, volume 37, pages 2555–2563, 2023.
- [57] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE Trans. Image Process.*, 22(12):5372–5384, 2013. 7
- [58] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.*, 26(2):982–993, 2016. 7
- [59] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.*, 22(9):3538–3548, 2013. 7
- [60] Chunming He, Yuqi Shen, Chengyu Fang, Fengyang Xiao, Longxiang Tang, Yulun Zhang, Wangmeng Zuo, Zhenhua Guo, and Xiu Li. Diffusion models in low-level vision: A survey. arXiv preprint arXiv:2406.11138, 2024. 7
- [61] Yixu Feng, Cheng Zhang, Pei Wang, Peng Wu, Qingsen Yan, and Yanning Zhang. You only need one color space: An efficient network for low-light image enhancement. *arXiv* preprint *arXiv*:2402.05809, 2024. 7
- [62] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *ECCV*, pages 0–0, 2018. 7
- [63] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Lett.*, 20(3):209–212, 2012. 7
- [64] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *Int. J. Comput. Vision*, 129:1013–1037, 2021. 8
- [65] Zhao Zhang, Huan Zheng, Richang Hong, Mingliang Xu, Shuicheng Yan, and Meng Wang. Deep color consistent network for low-light image enhancement. In CVPR, pages 1899–1908, 2022. 8
- [66] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embeddingfourier for ultra-high-definition low-light image enhancement. In *ICLR*, 2023. 8
- [67] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In CVPR, pages 22252–22261, 2023. 8
- [68] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In CVPR, pages 9935–9946, 2023. 8
- [69] Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *European Conference on Computer Vision*, pages 161–179. Springer, 2024. 8
- [70] Fan Zhang, Yu Li, Shaodi You, and Ying Fu. Learning temporal consistency for low light video enhancement from single images. In CVPR, pages 4967–4976, 2021.

- [71] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. 9
- [72] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010. 9
- [73] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022. 9
- [74] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):6024–6042, 2021. 9
- [75] Chunming He, Rihan Zhang, Fengyang Xiao, Chenyu Fang, Longxiang Tang, Yulun Zhang, Linghe Kong, Deng-Ping Fan, Kai Li, and Sina Farsiu. Run: Reversible unfolding network for concealed object segmentation. *ICML*, 2025. 9, 16
- [76] Zhuoyuan Wu, Jian Zhang, and Chong Mou. Dense deep unfolding network with 3d-cnn prior for snapshot compressive imaging. *arXiv preprint arXiv:2109.06548*, 2021. 16
- [77] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In CVPR, pages 17399–17410, 2022. 16

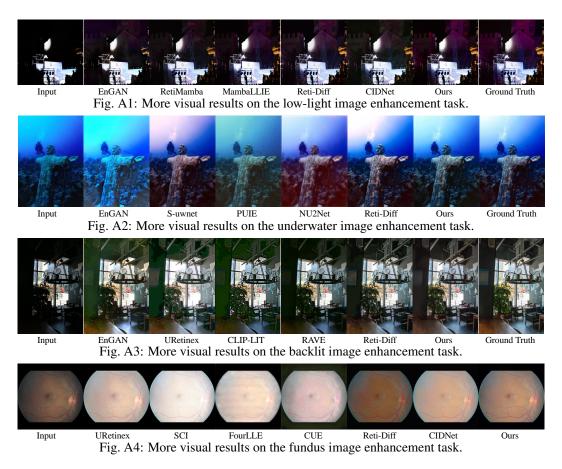
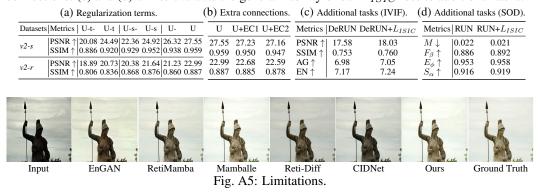


Table A1: Exploration of the intrinsic advantages of DUNs and investigation of the deployment for image restoration. where "U" denotes our UnfoldIR and the suffix "-" indicates constructing the network based on the basic restoration model, *i.e.*, CM1 in Table 7. In (b), EC refers to extra connections. (c) and (d) aim to evaluate the generalizability of our  $L_{ISIC}$  loss to additional tasks.



# **A** Discussions

In addition to proposing a powerful DUN-based method, this paper thoroughly explores the intrinsic advantages of DUNs and investigates their deployment for image restoration. From our analysis, we derive several significant conclusions:

First, as shown in Table A1a, the inclusion of explicit regularization terms is particularly beneficial for DUN-based frameworks, especially lightweight variants such as UnfoldIR-t (U-t) and UnfoldIR-s (U-s). These terms introduce explicit, task-specific priors, reducing the need for networks to implicitly learn complex priors. Consequently, this strategy decreases the number of required parameters and mitigates limitations faced by lightweight networks in modeling sophisticated priors.

Second, additional network connections—referred to here as extra connections (EC)—which have been commonly employed in previous DUN-based image restoration methods, such as EC1 [76] and EC2 [77], do not always yield performance improvements. As illustrated in Table A1b, these connections can sometimes degrade performance for two primary reasons: (1) DUNs inherently possess structured connections that rigorously derive from mathematical principles; (2) hence, adding arbitrary connections that are typically beneficial in purely learning-based, "black-box" connections can disrupt the mathematically principled structure of DUNs and impair their performance.

Third, as demonstrated in Tables A1c and A1d, our proposed  $L_{\rm ISIC}$  loss—a self-consistency supervision strategy that leverages the unique multi-stage structure of DUNs—shows strong generalizability to other image processing tasks. For instance, the DeRUN framework [18] in infrared and visible image fusion (IVIF), evaluated by additional metrics such as average gradient (AG) and entropy (EN), clearly benefits from  $L_{\rm ISIC}$  (where higher AG and EN scores indicate better performance). Similarly, RUN [75] in salient object detection (SOD) also demonstrates improved performance.

Finally, as presented in Table 9b, DUNs effectively enhance downstream tasks by leveraging outputs from multiple unfolding stages. These outputs serve as enhancement-based data augmentation inputs for downstream algorithms, thereby improving their overall performance.

## **B** Limitations and Future Work

As illustrated in Fig. A5, our method fails to recover certain subtle texture details. This limitation is common among existing approaches and is likely since such fine details, when obscured in dark regions, may be misinterpreted as artifacts or degradation and consequently removed during the restoration process.

This issue poses a challenge for preserving semantic consistency across image components and may lead to visually unnatural results. To address this, future work will explore the integration of high-level vision tasks to better capture and interpret the semantic context of image components. Additionally, we plan to enhance our DUN-based framework by incorporating generative techniques, such as diffusion models, to improve its capacity for producing results with higher perceptual fidelity.