

Enhancing Images with Coupled Low-Resolution and Ultra-Dark Degradations: A Tri-level Learning Framework

Jiaxin Gao

Dalian University of Technology
Dalian, China
jiaxinn.gao@outlook.com

Yaohua Liu

Dalian University of Technology
Dalian, China
liuyaohua_918@163.com

ABSTRACT

Due to device constraints and lighting conditions, captured images frequently exhibit coupled low-resolution and ultra-dark degradations. Enhancing the visibility and resolution of ultra-dark images simultaneously is crucial for practical applications. Current approaches often address both tasks in isolation or through simplistic cascading strategies, while also relying heavily on empirical and manually designed composite loss constraints, which inevitably results in compromised training efficacy, increased artifacts, and diminished detail fidelity. To address these issues, we propose **TriCo**, the first to adopt a **Tri-level** learning framework that explicitly formulates the bidirectional **Co**operative relationship and devises algorithms to tackle coupled degradation factors. In the optimization across Upper (U)-Middle (M)-Lower (L) levels, we model the synergistic dependencies between illumination learning and super-resolution tasks within the M-L levels. Moving to the U-M levels, we introduce hyper-variables to automate the learning of beneficial constraints for both learning tasks, moving beyond the traditional trial-and-error pitfalls of the learning process. Algorithmically, we establish a Phased Gradient-Response (PGR) algorithm as our training mechanism, which facilitates a dynamic, inter-variable gradient feedback and ensures efficient and rapid convergence. Moreover, we merge inherent illumination priors with universal semantic model features to adaptively guide pixel-level high-frequency detail recovery. Extensive experimentation validates the framework's broad generalizability across challenging ultra-dark scenarios, outperforming current state-of-the-art methods across 4 real and synthetic benchmark datasets over 6 metrics (e.g., **5.8%↑** in PSNR and **26.6%↑** in LPIPS).

CCS CONCEPTS

• **Computing methodologies** → *Scene understanding*.

KEYWORDS

Nighttime vision, super-resolution, coupled degradations, bi-level

ACM Reference Format:

Jiaxin Gao and Yaohua Liu. 2024. Enhancing Images with Coupled Low-Resolution and Ultra-Dark Degradations: A Tri-level Learning Framework.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM '24, October 28-November 1, 2024, Melbourne, Australia

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0686-8/24/10

<https://doi.org/10.1145/3664647.3681682>

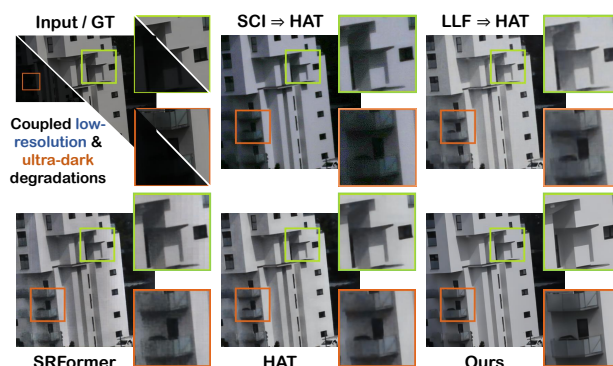


Figure 1: Visual comparison of advanced LLE [46, 61] and SR methods [4, 76] applied both independently and in a cascaded manner to inputs with coupled degradations.

In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24), October 28-November 1, 2024, Melbourne, Australia*
Proceedings of the 32nd ACM International Conference on Multimedia (MM'24), October 28-November 1, 2024, Melbourne, Australia. ACM, New York, NY, USA, 10 pages.
<https://doi.org/10.1145/3664647.3681682>

1 INTRODUCTION

Enhancing visibility and enlarging the resolution of ultra-dark images simultaneously is a daunting task with substantial real-world significance for fields such as intelligent surveillance and nocturnal autonomous driving [28, 44, 60, 62]. Due to inherent limitations in imaging devices and constraints posed by environmental lighting conditions, captured data frequently exhibits coupled degradations characterized by low resolution and extreme darkness [8, 34, 45, 67]. Imaging devices may struggle to capture clear details in poorly lit environments, resulting in images of low resolution; concurrently, insufficient ambient lighting exacerbates the darkness of the images, making content difficult to discern. This paper addresses the integrated challenge of enhancing brightness and increasing resolution in ultra-dark images plagued by these intertwined degradations.

Capturing images in ultra-low-light settings introduces a plethora of challenges that amplify the complexity of this joint task, including *uneven exposure resulting in highly irregular lighting, diminished contrast, color inaccuracies, and an overflow of artifacts*. Standard image Super-Resolution (SR) techniques [20, 31, 63, 65], crafted with modular techniques for image resolution enhancement under normal-lighting scenarios, cannot be straightforwardly adapted to enhance the luminance and resolution of images captured in low-light conditions. Indeed, the direct application of these techniques

would inevitably amplify hidden noise, blur, and artifacts present in darkness, leading to unnatural edges and textures and deviating from the primary goal of super-resolution. Contrarily, recent Low-Light Enhancement (LLE) methods [28, 40, 68], while capable of brightening, fall short in concurrently amplifying resolution and authentically enhancing high-frequency details. This prompts a further inquiry: *Can LLE and SR be effectively combined in a simple “A+B” cascaded format to achieve the desired outcomes?* Upon evaluation, we ascertain that this direct “A+B” does not address the entangled degradation factors at the data level, possibly akin to a “A×B” degraded form, with ongoing shortcomings in enhancing brightness and rendering texture details.

As illustrated in Fig. 1, we present a visual comparison of two cutting-edge LLE methods – SCI [46] and LLFormer [61] – alongside two normal-light SR techniques, HAT [4] and SRFormer [76]. A closer examination reveals that employing HAT and SRFormer independently falls short in recapturing fine details, producing blurred artifacts. Similarly, cascaded LLE⇒SR approaches (e.g., SCI⇒HAT, LLFormer⇒HAT) also fail to restore adequate brightness, exacerbating noise and structural distortions when magnified. In stark contrast, our proposed method generates natural and authentic exposure and color fidelity, alongside improved structural clarity and texture detail. In addition to the methods previously mentioned, a few recent studies have emerged focusing on super-resolution within low-light scenes [6, 19]. They tend to rely on simple brightness corrections and resolution scaling on synthetic datasets, leading to poor generalization in real-world scenarios. Thus, we summarize the two primary shortcomings limiting the efficacy of existing methods: (i) *The failure to recognize the intricacies of degradation-coupled data, which extends beyond a simplistic additive enhancement model. This overlooks the intrinsic bidirectional cooperation necessary for joint processing.* (ii) *A heavy reliance on empirical network design and manual aggregation of losses. Such methods disregard the guiding principles of physical image formation, and overlook the crucial role that suitably chosen loss constraints play in facilitating cooperative learning between intertwined tasks.*

Stemming from these insights, this paper seeks to explore a tri-level optimization perspective that formulates the cooperative relationships and devises a corresponding solution strategy. We propose TriCo, aiming to automate the optimization of these weighted constraints with hyper-variable and the coupling dependencies between two entangled tasks, striving to achieve a unified enhancement. Specifically, we initiate the process with an illumination interpolation mapping inspired by Retinex theory, yielding a brightened reflectance that serves as the foundation for subsequent feature-level super-resolution enhancement. We leverage universal foundational semantic model priors and illumination features under the self-regularized luminance constraint to provide dual guidance for the super-resolution process. On the algorithmic front, we have crafted a phased gradient-response algorithm as our training mechanism, meticulously designed to synergize the optimization of three key variables while offering dynamic gradient feedback throughout the training phase, thereby ensuring streamlined training efficiency and rapid convergence. In summary, our contributions are fourfold:

- We introduce a novel **Tri**-level optimization method that explicitly models the bidirectional **Co**operative relationship

between illumination learning and super-resolution, named **TriCo**. This approach effectively addresses the combined challenges of low-resolution and ultra-dark conditions by synergistically enhancing and enlarging degraded images.

- We establish an Upper (U)-Middle (M)-Lower (L) level nested formulation, which in its M-L level, explicitly delineates the collaborative dependency of two entangled tasks. In the U-M level, we integrate hyper-variables to autonomously enforce positive constraint feedback, thus dismantling the reliance on manual trial-and-error intervention.
- We propose a Phased Gradient-Response (PGR) algorithm as the training mechanism, designed to synergistically optimize three variables while providing dynamic gradient feedback, thus achieving efficient training and rapid convergence.
- Extensive experimentation validates the framework’s broad generalizability and performance advantages across 4 real and synthetic benchmark datasets over 6 metrics (e.g., 5.8%↑ in PSNR and 26.6%↑ in LPIPS).

2 RELATED WORK

Low-light Image Enhancement/Image Super-Resolution. LLE’s goal is to make images engulfed in darkness visible [9, 10, 44]. Early works generally concentrated on leveraging handcrafted priors and empirical insights for LLE, such as Retinex model [21, 23, 27] for separate treatment of illumination and reflection. Recent advancements have been seen with models based on convolutional neural networks, addressing these fundamental challenges [2, 18, 40, 41, 46, 59]. Especially, the method [66] proposed constructing a flexible semantic-aware embedding module for the low-light image enhancement task to consider the semantic information of different regions and alleviate color deviation issues. The author proposed a Retinex-based single-stage network architecture called Retinexformer [2], which guides the non-local interaction modeling of regions under different lighting conditions through an illumination-guided transformer. Typically, these techniques always rely on manually selecting complex loss functions, and they cannot be directly applied to effectively address the super-resolution task of low-light images. Instead, we introduce a tri-level learning strategy to model this coupled multi-degradation task, as well as pinpoint beneficial constraint feedback, diverging from the reliance on empirical hyperparameter tuning.

Normal-light SR task generates high resolution images from low resolution inputs under standard lighting conditions [13, 13, 30, 51, 56]. Recently, a large number of methods based on convolutional neural networks have emerged to continuously refresh the performance [31, 58, 66]. With the growing popularity of transformer-based technologies, many leading-edge methods [4, 5, 32, 76] have been developed for super-resolution enhancement, including SwinIR [32], Restormer [72], the recently proposed SRFormer [76] and HAT [4]. Indeed, these methods tailored for normal lighting conditions fail to address the challenges posed by low-light environments, often resulting in undesirable outcomes such as artifact spreading and texture blurring. Drawing on this, we fuse semantic cues from universal models [25, 74] with illumination attributes for modulation, meticulously steering the detail restoration of reflectance features and maintaining color consistency.

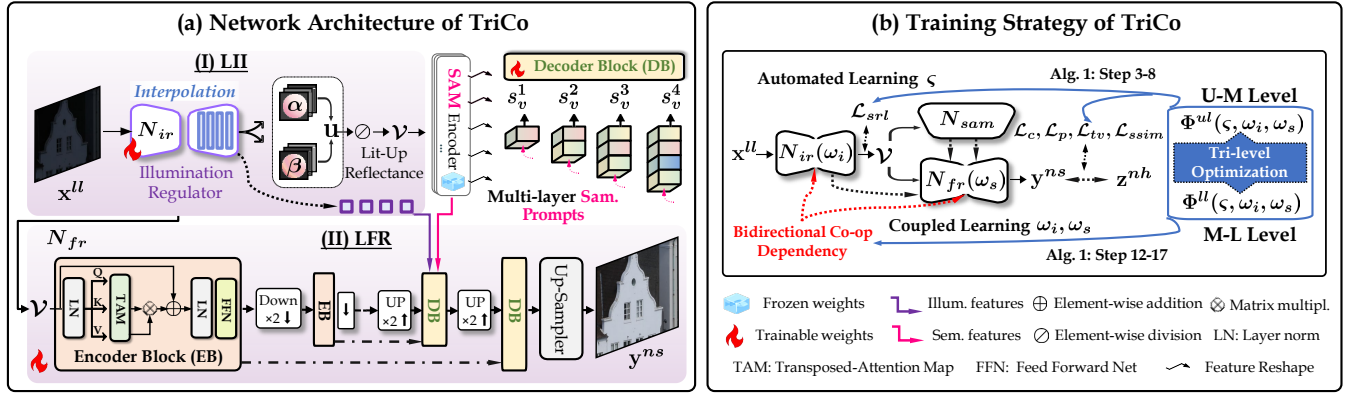


Figure 2: The overall TriCo framework. (a) initiates with an interpolation-based illumination regulator N_{ir} (parameterized by ω_i), producing a lit-up reflectance v . Then v feeds into a frozen SAM for multi-scale semantic prompts, while concurrently undergoing refinement for the LFR (N_{fr} parameterized by ω_s). (b) establishes tri-level learning paradigm with a phased gradient-response algorithm to foster a collaborative, automated, and efficient training process.

Low-light Image Super-Resolution. Recent forays into super-resolution focused on low-light imagery, have yet to yield satisfactory results [6, 16, 17, 19, 52, 57, 71]. For instance, Cheng *et al.* [6] proposed a light-guided and cross-fusion U-Net, featuring an intensity estimation unit, targets uneven-light image super-resolution. Yet, its sole reliance on pixel shuffling for resolution enlargement introduces notable color distortion and a lack of clarity in structure. A potential reason is that previous approaches did not account for the coupled collaboration between the two tasks, treating them in isolation. Hence, we employ hierarchical optimization to model and solve this multi-tiered coupled task.

Hierarchical Bi-level Optimization. Bi-level optimization is the hierarchical mathematical program, where the feasible region of upper-level task is restricted by the solution set mapping of lower-level task and the two task are mutually reinforced [26, 36]. Subsequently, the bi-level optimization framework has been widely applied across various application fields in machine learning and computer vision, e.g., hyper-parameter optimization [22, 39, 42, 43], multi-task and meta learning [38, 47, 55], neural architecture search [70, 77], and image processing and analysis [14, 15, 35, 37, 50]. Especially recently, the paper [50] proposed a triple-level optimization framework, which designs network architectures through cooperating optimization and auto-searching mechanisms to handle various video rain circumstances, significantly improving the fidelity and temporal consistency of rain removal. Similarly driven by this triple-level optimization, we explicitly consider the collaborative relationship between super-resolution and brightness adjustment tasks, constructing a novel perspective of automated learning with multiple constraints for modeling and solving.

3 METHODOLOGY

Our TriCo aims to transform extremely dim, low-resolution images x^{ll} into luminance-friendly, super-resolution counterparts y^{ns} . In the following, we firstly delve into the architecture of our model, followed by the details of the proposed learning strategy.

3.1 Illumination-Guided Integrated Network

Acknowledging the consensus that direct enlarging of dark images can result in the loss of details, noise amplification, and artifacts, we meticulously crafted an illumination-guided integrated network. This network does not merely cascade brightness adjustment with resolution enhancement modules; instead, it adopts a more profound integration approach to ensure a holistic improvement. As depicted in Fig. 2, our network adopts a top-down integration approach, seamlessly incorporating an illumination adjustment sub-network for initial brightness enhancement and a feature-level refinement sub-network for resolution upscaling. In this process, we introduce a key cross-attention-based transformer bridging component [16] that utilizes illumination and semantic priors as cues for guided refinement. This architecture is summarized into two phases: (i) the Learning Interpolated Illuminance (LII), focusing on adjusting the initial luminance, and (ii) the Learning Feature Refinement (LFR) for super-resolution, dedicated to enhancing and enlarging the image details at the feature level.

From LII to LFR. LII performs the initial mapping from “low-light” to “normal-light”. Based on Retinex theory, the normalized illumination map satisfies the inequality within the dynamic range: $0 \leq x^{ll} \leq u \leq I$. Thus, LII is designed to construct an interpolation mapping to estimate the illumination map u . Finally, the initial reflectance map v is obtained by applying element-wise division to u . The initial reflectance map v is then input to the LFR sub-network (i.e., N_{fr}) for fine-grained feature modulation, ensuring that the upsampling process generates more high-frequency details:

$$\begin{cases} u = \alpha \cdot x^{ll} + \beta \cdot I, \{\alpha, \beta\} = N_{ir}(\omega_i; x^{ll}), \\ v = x^{ll} \oslash u, y^{ns} = N_{fr}(\omega_s; v), \end{cases} \quad (1)$$

where the interpolation factors α and β are generated by the underlying Unet-style illumination regulator N_{ir} and satisfy the constraints within the unit interval, with their sum equaling 1. ω_i and ω_s are network parameters of N_{ir} and N_{fr} , respectively. Finally, we introduce a dynamic grid up-sampling module [11] to enlarge the image dimensions. The uniqueness of LII lies in its reliance solely

on a single luminance loss for unsupervised learning, eliminating cumbersome training with multiple stages and losses.

Also, we feed \mathbf{v} into a pre-trained large-scale base semantic model (i.e., MobileSAM [25, 74]) N_{sam} to extract multi-layer semantic features. We note that the multi-scale illuminance features and semantic features can serve as expert cues containing degradation priors (i.e., exposure and color information of different local areas). Therefore, we use the cross-attention-based transformer to modulate reflectance features layer-by-layer within the LFR sub-network's decoder, guiding the generation of high-frequency texture details. This module is designed based on the fact that illumination, as an intrinsic attribute of local material properties, is independent of resolution changes. Thus, we introduce general semantic features (reflecting local region attributes) and illumination features (reflecting regional illumination properties) for modulation.

In the Decoder Block (DB), each layer's reflectance feature in the decoder undergoes layer normalization, 1×1 convolution, and 3×3 depth-wise convolution, generating semantic query, reflection key, and reflection value projections. Next, the semantic query and reflection key perform matrix multiplication to generate the semantic attention map, which is then normalized and used for adaptively updating the reflectance feature. This involves element-wise multiplication and convolution, with a learnable scaling factor adjusting the operations. The updated reflectance feature is enhanced by a feed-forward network (FFN) for content reconstruction [24]. Similarly, illumination features from the LII decoder are combined with semantically guided reflectance features and processed through illumination attention. The resulting illumination query, reflection key, and reflection value projections are dynamically enhanced. Finally, the enhanced features are processed again by the FFN, yielding the doubly modulated reflection feature that integrates both semantic and illumination information for improved low-light image enhancement.

3.2 Tri-level Optimization Formulation

Bidirectional Co-op Dependency. Existing methods often focus on a single task, either brightness adjustment or resolution enhancement, seldom considering the interdependent coupling between the two. Recognizing that LII and LFR, can mutually reinforce each other—where precise luminance improvement by LII can facilitate better super-resolution outcomes in LFR, and conversely, the detailed enhancement by LFR can enhance the illumination learning in LII—we model these consecutive learning tasks as a hierarchical optimization problem, formalized as follows:

$$\begin{cases} \min_{\omega_i} \Phi^{ul}(\omega_i, \omega_s^*; \{\mathcal{D}_{ul}\}), \text{ s.t., } \omega_s^* \in \mathcal{P}_I(\omega_i), \\ \mathcal{P}_I(\omega_i) := \arg \min_{\omega_s} \Phi^{ll}(\omega_i, \omega_s; \{\mathcal{D}_{ll}\}), \end{cases} \quad (2)$$

where $\mathcal{P}_I(\cdot)$ denotes the solution set, with \mathcal{D}_{ll} and \mathcal{D}_{ul} representing the lower and upperlevel datasets, respectively.

The hierarchical formulation explicitly delineates the collaborative training modality between N_{lr} and N_{fr} . This collaboration is heavily contingent upon the judicious selection of lower and upper level objectives, ensuring that the sub-networks can reciprocally foster enhancement and positive feedback. This raises a pivotal question: *How can we automate the assignment of high hyperparameters that significantly foster positive influences on the learning*

Algorithm 1 Optimization strategy for TriCo.

Require: Initialize $\omega := \{\omega_i, \omega_s\}$, with ζ as a unit vector. Learning rate: $\gamma_u, \gamma_l, \mathbf{o}_u$ and \mathbf{o}_l ; Total iterations \mathcal{K} . Split $\{\mathcal{D}\} := \{\mathcal{D}_{ul}\} \cup \{\mathcal{D}_{ll}\}$ with partition ratio s . Set candidate loss space \mathcal{T} .

Ensure: The optimal parameters ζ, ω .

```

1: %% S1: Automated learning for  $\zeta$ .
2: while not converged do
3:   %% Upper-level variable update:
4:    $\hat{\omega} \leftarrow \omega - \gamma_l \frac{\partial \Phi^{ll}(\zeta, \omega)}{\partial \omega}, \omega^\pm \leftarrow \omega \pm \lambda \frac{\partial \Phi^{ll}(\zeta, \hat{\omega})}{\partial \omega}$ 
5:    $\mathcal{A}_\omega \leftarrow \frac{1}{2\lambda} \left( \frac{\partial \Phi^{ll}(\zeta, \omega^+)}{\partial \zeta} - \frac{\partial \Phi^{ll}(\zeta, \omega^-)}{\partial \zeta} \right)$ 
6:    $\zeta \leftarrow \zeta - \gamma_u \frac{\partial \Phi^{ul}(\zeta, \hat{\omega})}{\partial \zeta} + \gamma_l \mathcal{A}_\omega$ 
7:   %% Middle-level variable update:
8:    $\omega \leftarrow \omega - \gamma_l \frac{\partial \Phi^{ll}(\zeta, \omega)}{\partial \omega}$ 
9: end while
10: %% S2: Optimization for  $\{\omega_i, \omega_s\}$  with frozen  $\zeta$ .
11: while not converged do
12:   %% Middle-level variable update:
13:    $\hat{\omega}_s \leftarrow \omega_s - \mathbf{o}_l \frac{\partial \Phi^{ll}(\omega_i, \omega_s)}{\partial \omega_s}, \omega_s^\pm \leftarrow \omega_s \pm \lambda \frac{\partial \Phi^{ll}(\omega_i, \hat{\omega}_s)}{\partial \omega_s}$ 
14:    $\mathcal{B}_{\omega_s} \leftarrow \frac{1}{2\lambda} \left( \frac{\partial \Phi^{ll}(\omega_i, \omega_s^+)}{\partial \omega_i} - \frac{\partial \Phi^{ll}(\omega_i, \omega_s^-)}{\partial \omega_i} \right)$ 
15:    $\omega_i \leftarrow \omega_i - \mathbf{o}_u \frac{\partial \Phi^{ul}(\omega_i, \hat{\omega})}{\partial \omega_i} + \mathbf{o}_l \mathcal{B}_{\omega_s}$ 
16:   %% Lower-level variable update:
17:    $\omega_s \leftarrow \omega_s - \mathbf{o}_l \frac{\partial \Phi^{ll}(\omega_i, \omega_s)}{\partial \omega_s}$ 
18: end while
```

tasks? Delving deeper into this inquiry, we transcend the confines of the hierarchical optimization framework and venture into an expanded horizon—establishing a nested optimization problem that encompasses both lower and upper levels, aimed at the autonomous learning of beneficial constraints for two learning tasks.

Tri-level Constraint Modeling. Evidently, to automate the determination of constraints that significantly influence the learning tasks through weight allocation, we introduce a novel concept, the hyper-variable ζ . This hyper-variable, along with the two preceding variables ω_i and ω_s , forms a new set of constraint relationships, thereby constituting a hierarchical optimization problem based on three variables, as shown below:

$$\begin{cases} \min_{\zeta} \Phi^{ul}(\zeta, \omega_i^*, \omega_s^*; \{\mathcal{D}_{ul}\}), \text{ s.t., } (\omega_i^*, \omega_s^*) \in \mathcal{P}_I(\zeta), \\ \mathcal{P}_I(\zeta) := \arg \min_{\omega_i, \omega_s} \Phi^{ll}(\zeta, \omega_i, \omega_s; \{\mathcal{D}_{ll}\}), \end{cases} \quad (3)$$

where ω_i^* and ω_s^* represent the best-reponse for a given ζ . These three variables are highly interdependent and dynamically influence each other throughout the training process. This modeling approach offers significant advantages: firstly, it explicitly defines the mathematical relationships between multiple variables, allowing for dynamic feedback during training, thereby enhancing training efficiency. Secondly, it automates the determination of constraints' positive feedback, overcoming the reliance on manual hyper-parameter tuning based on empirical knowledge, thereby reducing the need for extensive manual intervention.

3.3 Algorithmic Procedure

Moving forward, we devise a Phased Gradient-Response (PGR) algorithm that iterates from the upper to the lower layers, serving as the training strategy. Specifically, we define a comprehensive function as the weighted sum of multiple losses related to the hyper-variables ζ , addressing various specific attributes (e.g., brightness, color, exposure, smoothness, and content) pertinent to multi degradation restoration tasks. The total loss is defined as follows:

$$\mathcal{L}_{total}(\zeta, \omega_i, \omega_s; \{\mathcal{D}\}) = \sum_{u=1}^N \zeta_u \cdot \mathcal{L}_u(\zeta, \omega_i, \omega_s), \quad \mathcal{L}_u \in \mathcal{T}, \quad (4)$$

where the hyper-variable is denoted as $\zeta := \{\zeta_u\}_{u=1}^N \in \mathbb{R}^N$. \mathcal{T} represents the loss selection set. Please refer to Sec. 3.4 for \mathcal{T} . To prevent ambiguous solutions during training and safeguard against overfitting, we introduce a regularization constraint term for ζ , thereby reformulating the total loss as: $\mathcal{L}_{total}(\zeta, \omega_i, \omega_s; \{\mathcal{D}\}) = \sum_{u=1}^N \frac{1}{2\zeta_u} \cdot \mathcal{L}_u(\zeta, \omega_i, \omega_s) + \ln(1 + \zeta_u^2)$. The training set is divided into proportions denoted by η , thus the upper and lower levels are abstractly defined as: $\Phi^{ul} := \mathcal{L}_{total}(\zeta, \omega_i, \omega_s; \mathcal{D}_{ul})$, $\Phi^{ll} := \mathcal{L}_{total}(\zeta, \omega_i, \omega_s; \mathcal{D}_{ll})$. Next, we decompose the problem into two stages of hierarchical optimization to solve the tri-level coupled problem step by step.

Optimization Algorithm. Following the first-order gradient algorithm based on hierarchical optimization [36], we compute the composite upper gradients based on the best-response from the lower optimization. We first calculate the upper-level gradient:

$$\nabla_{\zeta} \Phi^{ul}(\zeta, \omega) = \frac{\partial \Phi^{ul}(\zeta, \omega^*(\zeta))}{\partial \zeta} + \frac{\partial \Phi^{ul}(\zeta, \omega^*(\zeta))}{\partial \omega} \nabla_{\zeta} \omega^*(\zeta). \quad (5)$$

For simplicity, we define the lower-level variables as $\omega := \{\omega_i, \omega_s\}$. The second term, the coupled gradient, is denoted as \mathcal{A}_{ω} . Subsequently, based on a single-step gradient descent to approximate the best-response, we calculate the finite difference approximation [33] for the coupled gradient \mathcal{A}_{ω} as $\mathcal{A}_{\omega} = \frac{1}{2\lambda} (\frac{\partial \Phi^{ul}(\zeta, \omega^+)}{\partial \zeta} - \frac{\partial \Phi^{ul}(\zeta, \omega^-)}{\partial \zeta})$,

where $\omega^{\pm} \leftarrow \omega \pm \lambda \frac{\partial \Phi^{ul}(\zeta, \omega)}{\partial \omega}$, and λ denotes a constant learning rate. For the second phase, a similar derivation to the first phase is implemented. Given the optimal hyper-variable ζ^* obtained from the first stage, we compute the upper-level gradient with respect to

the variable ω_i : $\nabla_{\omega_i} \Phi^{ul}(\omega_i, \omega_s) = \frac{\partial \Phi_{\zeta^*}^{ul}(\omega_i, \omega_s^*(\omega_i))}{\partial \omega_i} + \mathcal{B}_{\omega_s}$, where $\mathcal{B}_{\omega_s} = \frac{\partial \Phi_{\zeta^*}^{ul}(\omega_i, \omega_s^*(\omega_i))}{\partial \omega_s} \nabla_{\omega_i} \omega_s^*(\omega_i)$. Ultimately, the optimization process across both stages is amalgamated to form our training strategy, which is summarized in Alg. 1.

3.4 Loss Objectives

As illustrated in Fig. 2, we propose a set of five specific loss objectives constituting a candidate space that encapsulates the model's constraints on brightness, color, exposure, smoothness, and content attributes, denoted as the set \mathcal{T} , as follows:

- To ensure that the generated reflectance aligns with the luminance attributes of large-scale natural ImageNet dataset [12] in a consistent distribution, we introduce the self-regularized luminance loss: \mathcal{L}_{srl} :

$$\mathcal{L}_{srl}(\mathbf{v}) = e^{|\bar{v}_c - \mu_c - \sigma_c|} - 1, \quad c \in \{R, G, B\}, \quad (6)$$

Table 1: Quantitative comparison among cascaded LLE methods and SR method (i.e., LLE \Rightarrow HAT) on RELISUR dataset and LOL-v2 dataset. The best three results are bolded.

Method	RELISUR @ $\times 2$ / @ $\times 4$		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
HAT [4]	20.213 / 19.751	0.719 / 0.715	0.454 / 0.561
ZeroDCE [29] \Rightarrow HAT [4]	12.927 / 12.524	0.354 / 0.321	0.698 / 0.739
SCI [46] \Rightarrow HAT	14.963 / 14.776	0.439 / 0.452	0.591 / 0.697
LLFormer [61] \Rightarrow HAT [4]	21.218 / 20.135	0.720 / 0.718	0.455 / 0.575
Retinexformer [2] \Rightarrow HAT [4]	21.326 / 20.142	0.725 / 0.722	0.421 / 0.544
Ours	22.456 / 21.056	0.744 / 0.731	0.304 / 0.432
Method	LOL-v2 @ $\times 2$ / @ $\times 4$		
	BRISQUE \downarrow	NIQE \downarrow	MetalQA \uparrow
HAT [4]	48.978 / 66.220	9.325 / 10.381	0.387 / 0.371
Retinexformer [2] \Rightarrow HAT [4]	49.510 / 58.280	8.026 / 9.561	0.360 / 0.362
Ours	44.191 / 56.475	7.910 / 8.821	0.368 / 0.391

where \bar{v}_c signifies the operation of computing the mean across channels. Channel means and standard deviations are $\mu_c = [0.485, 0.456, 0.406]$ and $\sigma_c = [0.229, 0.224, 0.225]$.

- We employ the standard content reconstruction loss between y^{ns} and z^{nh} utilizing the L_1 norm:

$$\mathcal{L}_c(y^{ns}, z^{nh}) = \frac{1}{hwc} \sum_{i,j,k} |y_{i,j,k}^{ns} - z_{i,j,k}^{nh}|, \quad (7)$$

where h, w, c are the height, width, and channel count.

- We also introduce the semantic perceptual loss function to maintain semantic congruence between y^{ns} and z^{nh} :

$$\mathcal{L}_p(y^{ns}, z^{nh}) = \|\text{VGG19}_j(y^{ns}) - \text{VGG19}_j(z^{nh})\|_1, \quad (8)$$

where j indicates the j -th feature extraction layer, which includes layers from conv1, \dots , conv5.

- We employ the SSIM loss \mathcal{L}_{ssim} to maintain the structural similarity between y^{ns} and y^{nh} .
- We incorporate a total variation metric [53] to reduce noise and enhance image smoothness:

$$\mathcal{L}_{tv}(y^{ns}) = \sum_{\xi \in \pi} (|\nabla_h y_{\xi}^{ns}| + |\nabla_v y_{\xi}^{ns}|), \quad (9)$$

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

4 EXPERIMENTS

4.1 Experimental Settings

Datasets and Metrics. We evaluated the benchmark performance of all compared methods across four datasets: 1) RELISUR [1]¹, 2) Dark-Zurich [54], 3) LOL-v2 [69], and 4) Cityscapes [7]. In the evaluation phase, we employ three full-reference metrics to assess the performance, namely Peak Signal-to-Noise Ratio (PSNR) [3], Structural Similarity Index (SSIM) [64], Learned Perceptual Image Patch Similarity (LPIPS) [75]. Additionally, we introduce three non-reference assessments, namely Natural Image Quality Evaluator (NIQE) [49], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [48] and MetalQA [78], to evaluate non-paired metrics.

¹<https://vap.aau.dk/rellisur/>

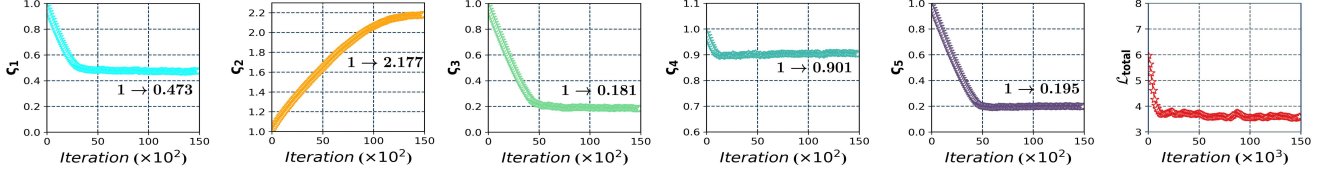


Figure 3: Illustrating the convergence curve of the hyper-variable $\{\zeta_u\}_{u=1}^5$ and the total loss \mathcal{L}_{total} , based on Alg. 1.

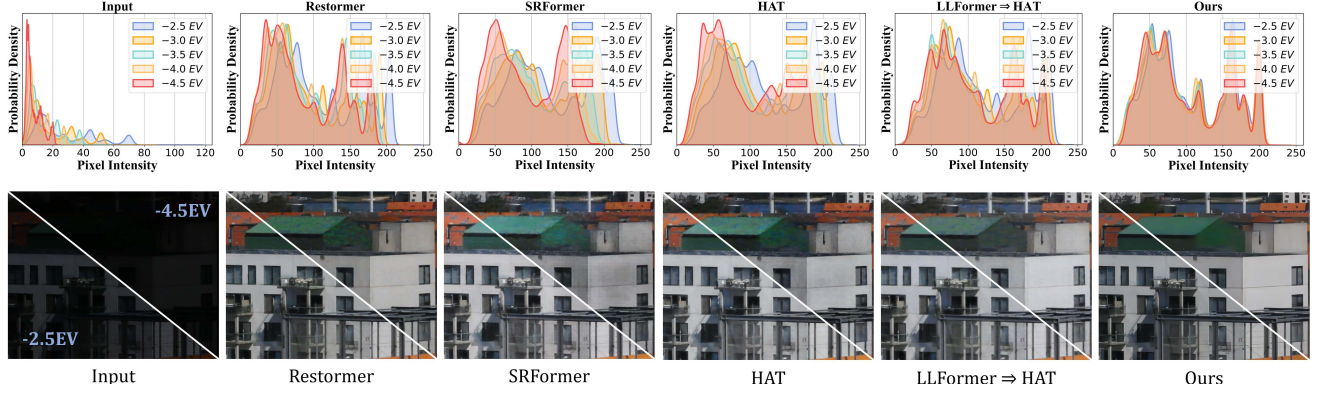


Figure 4: Illustrating (Above) the probability density histogram trends and (Below) enhancement outcomes for the same sample (i.e., 00018) under five different levels of darkness. Note that from -2.5EV to -4.5EV indicates increasing darkness.

Table 2: Quantitative comparison on RELISUR dataset for $\times 2 / \times 4$ tasks. The best three results are bolded.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	#Param. (M) \downarrow	#FLOPs (G) \downarrow	#Infer. (s) \downarrow	#FPS \uparrow
MIRNet [73]	21.052 / 19.784	0.720 / 0.704	0.436 / 0.599	31.814	196.563	0.036	27.24
SwinIR [32]	18.383 / 17.531	0.640 / 0.663	0.577 / 0.688	11.756	57.381	0.116	8.64
LCUN [6]	18.911 / 18.463	0.684 / 0.657	0.531 / 0.644	--	--	--	--
Restormer [72]	21.217 / 20.290	0.727 / 0.720	0.385 / 0.492	26.126	35.375	0.033	26.87
SRFormer [76]	19.554 / 18.792	0.704 / 0.705	0.469 / 0.613	10.162	81.797	0.218	3.01
HAT [4]	20.213 / 19.751	0.719 / 0.715	0.454 / 0.561	9.473	58.990	0.184	5.31
Ours	22.456 / 21.056	0.744 / 0.731	0.304 / 0.432	1.421	20.774	0.024	41.67

Implementation Details. We adhere to the tri-level learning strategy as outlined in Alg. 1 for our network training, with the total number of iterations set to 150,000. We utilize the Adam optimizer with beta values configured at [0.9, 0.999]. The initial learning rates for the upper and lower layers of the two stages are set to $\gamma_u = 1e-4$, $\gamma_l = 2e-4$, $\mathbf{o}_u = 1e-4$ and $\mathbf{o}_l = 1e-4$, respectively. A cosine annealing restart strategy is implemented for cyclic learning rate scheduling. The dataset $\{\mathcal{D}\}$ is partitioned into $\{\mathcal{D}_{ul}\} \cup \{\mathcal{D}_{ll}\}$ and at a distribution ratio of 1 : 5. Experiments are conducted using PyTorch version 2.0.1, which supports CUDA 11.7, on a single NVIDIA RTX A6000 GPU with 48GB of RAM.

Compared Methods. To substantiate the efficacy of our proposed methodology, we conduct a comprehensive comparison with a diverse array of SOTA methods in LLE and SR. Specifically, we meticulously benchmark against 4 emblematic LLE techniques, namely ZeroDCE [29], SCI [46], and LLFormer [61], Retinexformer

[2], alongside 6 SR methods, which include 5 under normal lighting conditions—SwinIR [32], MIRNet [73], Restormer [72], SRFormer [76], and HAT [4]—and one dedicated to low-light scenarios, LCUN [6]. Notably, the enhancement results on the RELISUR dataset for the sole low-light SR method, LCUN, were furnished by the authors themselves. To ensure a fair comparison, we retrain the publicly available codes of all competing methods on the training set of the RELISUR dataset.

4.2 Mechanism Evaluation

Algorithm Analysis. Following Alg. 1, we undertake a tri-level automated training regimen to sequentially optimize the three variables. Fig. 3 illustrates the convergence trend of the outermost variable and the overall loss function throughout the iterations. During the S1 cycle in Alg. 1 (refer to steps 1 to 8), the upper-level variable $\{\zeta_u\}_{u=1}^5$ evolves from an initial unit vector to eventually

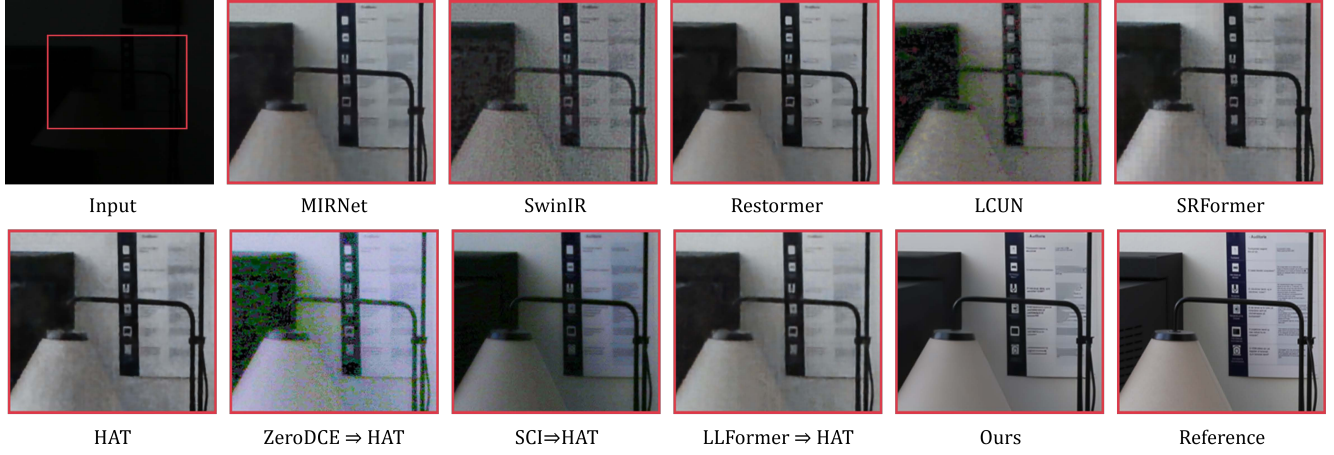


Figure 5: Visual assessments on RELLISUR examples for a $\times 2$ magnification task.

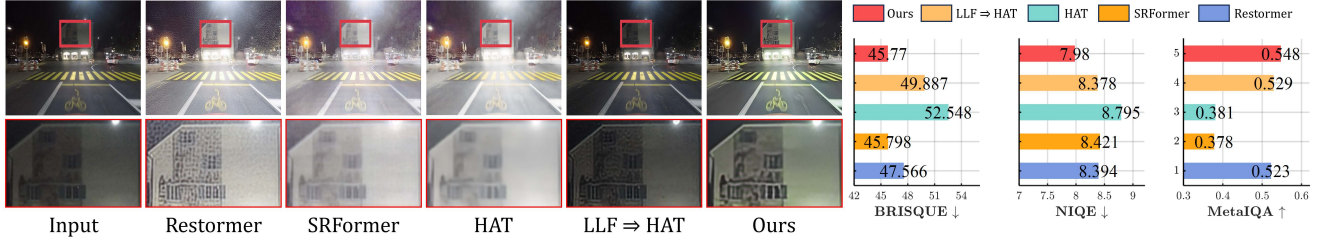


Figure 6: Visual comparisons (Left) and quantitative results on three metrics (Right) for enhancing brightness and enlarging low-light images on real nighttime Dark-Zurich samples. Note here, LLFormer⇒HAT is abbreviated as LLF⇒HAT.

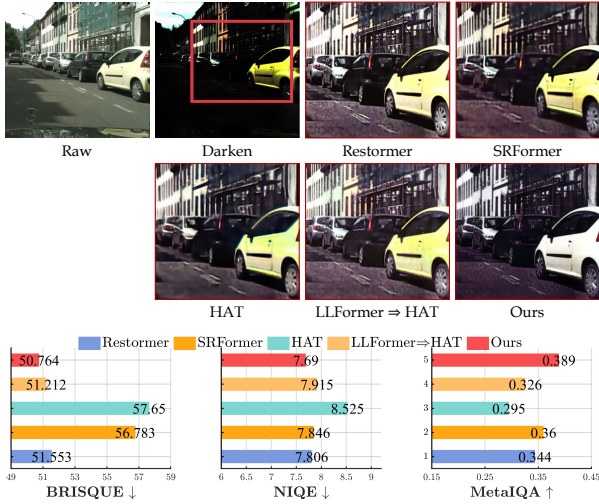


Figure 7: Qualitative and quantitative evaluations of various methods on the darken Cityscapes dataset.

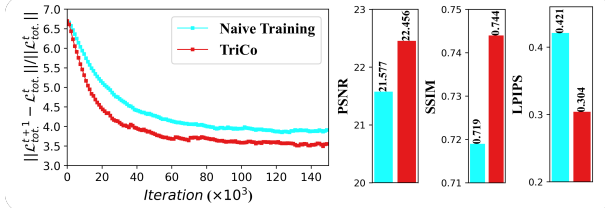
converge to [0.473, 2.177, 0.181, 0.901, 0.195]. This convergence elucidates that the constraints positively influencing the learning task,

as hypothesized by our algorithm, are indeed effective. It is possible to autonomously identify which constraints significantly foster a positive impetus for the learning task through adaptive weight allocation. Notably, the top three constraints—perceptual constraint, smoothness constraint, and reconstruction constraint—play a pivotal role in augmenting performance, underscoring the efficacy of our proposed approach in leveraging these constraints for enhanced learning outcomes. To evaluate model efficiency, we present the parameters, FLOPs, inference time, and FPS of compared SOTA methods in Tab. 2. The evaluations are performed on a single 2080 Ti GPU using images of size 128×128 . Excluding the parameters of the frozen SAM model, our network has a parameter count of less than 1.4MB. In conclusion, our network achieves a favorable balance between performance and efficiency.

Robustness Verification. Fig. 4 conducts a robustness analysis across varying levels of darkness. Five distinct levels of low exposure are generated by adjusting exposure time, resulting in corresponding dark images (e.g., -2.5EV, -3.0EV, -3.5EV, -4.0EV, and -4.5EV). Our method maintains consistent enhancement across various levels of darkness. This is visually corroborated by the probability density histograms, which demonstrate a uniform consistency distribution across the five different levels of darkness, highlighting the high robustness of our model to inputs under varying levels of darkness.

Table 3: Ablation studies of the Alg. 1 (in terms of w/ or w/o S1 and S2) and DB on RELLISUR dataset.

Config.	Alg. 1 (w/o S1)	Alg. 1 (w/o S2)	w/ DB	PSNR↑	SSIM↑
0	✓	✓	✓	22.456	0.744
1	✓		✓	21.997 _{↓0.459}	0.727 _{↓0.017}
2		✓	✓	22.243 _{↓0.213}	0.735 _{↓0.009}
3	✓	✓		21.959 _{↓0.497}	0.726 _{↓0.018}

**Figure 8: Comparison analysis of the naive training strategy and our TriCo strategy.**

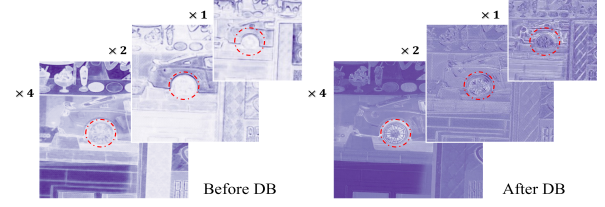
4.3 Comparisons with State-of-the-Art

Evaluation on RELLISUR. Tab. 2 presents the quantitative results for low-light super-resolution tasks at $\times 2$ and $\times 4$ scales on the RELLISUR dataset. While cascading strategies prove effective, the improvement is not drastic. Relative to the second-best method, our approach achieves significant enhancements across all metrics (e.g., a 5.8% increase in PSNR, a 2.3% boost in SSIM, and a 21.0% leap in LPIPS). The substantial improvements in LPIPS underscore our method’s capability to refine textures and robustly adapt to various extreme low-light conditions. Fig. 5 showcases a visual comparison on RELLISUR for simultaneous brightness adjustment and $\times 2$ upscaling. The majority of the compared methods suffer from significant noise and blur issues, especially observable in SwinIR, LCUN, and SRFormer. Some cascaded approaches exhibit severe color bias and insufficient brightness enhancement. In contrast, our method is capable of producing images with vivid luminance and excellent restoration of high-frequency structural details. Signal plots intuitively confirm the consistency between our method and the reference images at the pixel intensity level.

Evaluation on Real Nighttime Scenarios. As illustrated in Fig. 6, we assess the generalization performance of the entire benchmark suite under the real-world challenging scenario: dimly lit urban streetscapes at night, based on Dark-Zurich dataset. Fig. 7 showcases a comparison of enhancement methods applied to the Darken Cityscapes Dataset. Compared to other methods that either exhibit noticeable color bias or suffer from large areas of under-exposure and overexposure, it can be seen that our method has generated clearer structural details.

4.4 Ablation Analyses

Effectiveness of DB. When the DB module is removed, as seen in Config.3 of Tab. 3, there is a noticeable performance drop — approximately 2.2% in PSNR and 2.4% in SSIM — compared to the optimal model, Config.0. Fig. 9 presents the ablation results with the feature visualizations facilitated by DB. We visualized the

**Figure 9: Illustrating of intermediate layer feature visualization for the DB module.****Table 4: Ablation of loss functions on RELLISUR dataset.**

No. → Loss function ablation	PSNR↑ (trained during S1 cycle)
❶: w/o \mathcal{L}_p loss ❷: w/o \mathcal{L}_{ssim} loss	❶: 19.987 _{↓1.055} ❷: 20.906 _{↓0.136}
❸: w/o \mathcal{L}_{tv} loss ❹: w/ Full loss	❸: 21.036 _{↓0.006} ❹: 21.042

features before and after the DB process in the last three decoding layers. Upon comparison, the features prior to DB appear more sparse and scattered, with a distinct lack of textural detail (see the discernible regions within the dashed circles: the car wheels). Post-DB features, however, exhibit a more abstract and semantically rich visual representation. This indicates that DB fosters a greater focus on capturing higher-level semantics.

Analysis of the Solution Algorithm. We conduct the ablation study to quantify the impact of the proposed algorithm components, with comparative results detailed in Tab. 3, from Config.0 to Config.3. Omitting the S1 strategy alone leads to a performance degradation of approximately 2% in PSNR and 2.2% in SSIM compared to the best-performing model, Config.0. Similarly, removing only the S2 strategy results in a reduction of about 0.9% in PSNR and 1.2% in SSIM relative to Config.0. This delineation underscores the critical importance of synergistically integrating S1 and S2 strategies to achieve the superior performance set forth by Config.0.

Ablation of the Loss Functions. Tab. 4 reports the ablation analysis results of three empirical loss functions. Please note that the four sets of different experiments in the table were conducted with only the S1 training cycle (i.e., 6000 iterations) for a fair comparison. The PSNR improvements associated with each loss highlight their importance, which correlates with the coefficients of the upper-level hyper-variables automatically learned by our algorithm (see Fig. 3 in Sec. 4.2).

5 CONCLUSION AND REMARKS

This investigation delves into the intricate realm of brightening and magnifying ultra-dark images, a pursuit fraught with practical complexities due to the dual dilemmas of low resolution and profound darkness. Our tailored TriCo, adopts a tri-level learning strategy that intertwines the tasks of illumination enhancement and super-resolution. By fostering the collaborative learning, TriCo effectively negates the historical deficiencies of isolated or simplistic task handling. TriCo’s strategic innovation can be applied to other high-level semantic tasks in more adverse lighting scenarios.

REFERENCES

- [1] Andreas Aakerberg, Kamal Nasrollahi, and Thomas Moeslund. 2021. RELLISUR: A Real Low-Light Image Super-Resolution Dataset. In *Neural Information Processing Systems Track on Datasets and Benchmarks*, J. Vanschoren and S. Yeung (Eds.), Vol. 1.
- [2] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. 2023. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12504–12513.
- [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] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. 2023. Activating More Pixels in Image Super-Resolution Transformer. In *CVPR*. 22367–22377.
- [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] Deqiang Cheng, Liangliang Chen, Chen Lv, Lin Guo, and Qiqi Kou. 2022. Light-Guided and Cross-Fusion U-Net for Anti-Illumination Image Super-Resolution. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 12 (2022), 8436–8449.
- [7] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Endzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *CVPR*.
- [8] Xiaohan Cui, Long Ma, Tengyu Ma, Jinyuan Liu, Xin Fan, and Risheng Liu. 2024. Trash to treasure: Low-light object detection via decomposition-and-aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 1417–1425.
- [9] Ziteng Cui, Lin Gu, Xiao Sun, Xianzheng Ma, Yu Qiao, and Tatsuya Harada. 2024. Aleth-nerf: Illumination adaptive nerf with concealing field assumption. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 1435–1444.
- [10] Ziteng Cui, Guo-Jun Qi, Lin Gu, Shaodi You, Zenghui Zhang, and Tatsuya Harada. 2021. Multitask aet with orthogonal tangent regularity for dark object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*. 2553–2562.
- [11] Yutong Dai, Hao Lu, and Chunhua Shen. 2021. Learning Affinity-Aware Upsampling for Deep Image Matting. In *CVPR*. 6841–6850.
- [12] 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*. 48–55.
- [13] Jiangxin Dong, Haoran Bai, Jinhui Tang, and Jinshan Pan. 2023. Deep Unpaired Blind Image Super-Resolution Using Self-supervised Learning and Exemplar Distillation. *International Journal of Computer Vision* (2023), 1–14.
- [14] 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.
- [15] 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.
- [16] Jiaxin Gao, Ziyu Yue, Yaohua Liu, Sihan Xie, Xin Fan, and Risheng Liu. 2023. Diving into darkness: A dual-modulated framework for high-fidelity super-resolution in ultra-dark environments. *arXiv preprint arXiv:2309.05267* (2023).
- [17] 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* (2024).
- [18] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Computer Vision and Pattern Recognition*. 1780–1789.
- [19] 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.
- [20] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. 2018. Deep back-projection networks for super-resolution. In *CVPR*. 1664–1673.
- [21] Kui Jiang, Qiong Wang, Zhaoyi An, Zheng Wang, Cong Zhang, and Chia-Wen Lin. 2024. Mutual retinex: Combining transformer and cnn for image enhancement. *IEEE Transactions on Emerging Topics in Computational Intelligence* (2024).
- [22] Dian Jin, Long Ma, Risheng Liu, and Xin Fan. 2021. Bridging the Gap between Low-Light Scenes: Bilevel Learning for Fast Adaptation. In *ACM Multimedia Conference '21*. ACM, 2401–2409.
- [23] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing* 6, 7 (1997), 965–976.
- [24] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. 2022. Transformers in vision: A survey. *ACM Computing Surveys (CSUR)* 54, 10s (2022), 1–41.
- [25] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. *arXiv:2304.02643* (2023).
- [26] Thomas Kleinert, Martine Labbé, Ivana Ljubić, and Martin Schmidt. 2021. A survey on mixed-integer programming techniques in bilevel optimization. *EURO Journal on Computational Optimization* 9 (2021), 100007.
- [27] Edwin H Land and John J McCann. 1971. Lightness and retinex theory. *Josa* 61, 1 (1971), 1–11.
- [28] 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.
- [29] 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.
- [30] Xiang Li, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. 2023. Dlganet: light-weight dynamic local and global self-attention networks for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12792–12801.
- [31] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. 2019. Feedback network for image super-resolution. In *CVPR*. 3867–3876.
- [32] 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.
- [33] Hanxiao Liu, Karen Simonyan, and Yiming Yang. 2019. DARTS: Differentiable Architecture Search. In *ICLR*.
- [34] Jinyuan Liu, Guanyao Wu, Junsheng Luan, Zhiying Jiang, Risheng Liu, and Xin Fan. 2023. HoLoCo: Holistic and local contrastive learning network for multi-exposure image fusion. *Information Fusion* 95 (2023), 237–249.
- [35] Risheng Liu, Jiaxin Gao, Xuan Liu, and Xin Fan. 2024. Learning With Constraint Learning: New Perspective, Solution Strategy and Various Applications. *IEEE Trans. Pattern Anal. Mach. Intell.* 46, 7 (2024), 5026–5043.
- [36] 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.
- [37] 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 Trans. Pattern Anal. Mach. Intell.* 44, 11 (2022), 7688–7704.
- [38] Risheng Liu, Yaohua Liu, Wei Yao, Shangzhi Zeng, and Jin Zhang. 2023. Averaged method of multipliers for bi-level optimization without lower-level strong convexity. In *International Conference on Machine Learning*. PMLR, 21839–21866.
- [39] Risheng Liu, Yaohua Liu, Shangzhi Zeng, and Jin Zhang. 2021. Towards gradient-based bilevel optimization with non-convex followers and beyond. *Advances in Neural Information Processing Systems* 34 (2021), 8662–8675.
- [40] 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.
- [41] Xiaofeng Liu, Jiaxin Gao, Xin Fan, and Risheng Liu. 2023. NAI₂: Learning Noise-Aware Illumination-Interpolator for Unsupervised Low-Light Image Enhancement. *arXiv preprint arXiv:2305.10223* (2023).
- [42] 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. *arXiv preprint arXiv:2406.02064* (2024).
- [43] Yaohua Liu and Risheng Liu. 2021. Boml: A Modularized Bilevel Optimization Library In Python For Meta Learning. In *2021 IEEE International Conference on Multimedia & Expo Workshops, ICME Workshops, Shenzhen, China, July 5-9, 2021*. IEEE, 1–2.
- [44] Rundong Luo, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2023. Similarity min-max: Zero-shot day-night domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 8104–8114.
- [45] Long Ma, Dian Jin, Nan An, Jinyuan Liu, Xin Fan, Zhongxuan Luo, and Risheng Liu. 2023. Bilevel fast scene adaptation for low-light image enhancement. *International Journal of Computer Vision* (2023), 1–19.
- [46] 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.
- [47] 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 *International Conference on Learning Representations ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.
- [48] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing* 21, 12 (2012), 4695–4708.
- [49] 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.
- [50] 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.

- [51] Jinshan Pan, Yang Liu, Deqing Sun, Jimmy Ren, Ming-Ming Cheng, Jian Yang, and Jinhui Tang. 2020. Image formation model guided deep image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 11807–11814.
- [52] Muhammad Tahir Rasheed and Daming Shi. 2022. LSR: Lightening super-resolution deep network for low-light image enhancement. *Neurocomputing* 505 (2022), 263–275.
- [53] 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.
- [54] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. 2020. Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 6 (2020), 3139–3153.
- [55] Amirreza Shaban, Ching-An Cheng, Nathan Hatch, and Byron Boots. 2019. Truncated Back-propagation for Bilevel Optimization. In *AISTATS 2019 (Proceedings of Machine Learning Research, Vol. 89)*. PMLR, 1723–1732.
- [56] Long Sun, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. 2023. Spatially-adaptive feature modulation for efficient image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 13190–13199.
- [57] Chenyang Wang, Junjun Jiang, Kui Jiang, and Xianming Liu. 2024. Low-Light Face Super-resolution via Illumination, Structure, and Texture Associated Representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 5318–5326.
- [58] Cong Wang, Jinshan Pan, Wei Wang, Jiangxin Dong, Mengzhu Wang, Yakun Ju, and Junyang Chen. 2023. PromptRestorer: A Prompting Image Restoration Method with Degradation Perception. In *NeurIPS*, Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (Eds.).
- [59] Cong Wang, Jinshan Pan, Wei Wang, Gang Fu, Siyuan Liang, Mengzhu Wang, Xiao-Ming Wu, and Jun Liu. 2024. Correlation Matching Transformation Transformers for UHD Image Restoration. In *AAAI*, Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (Eds.). AAAI Press, 5336–5344.
- [60] 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.
- [61] 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*. 2654–2662.
- [62] Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2021. Hla-face: Joint high-low adaptation for low light face detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16195–16204.
- [63] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *ECCV workshops*. 0–0.
- [64] 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.
- [65] 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.
- [66] Yuhui Wu, Chen Pan, Guoqing Wang, Yang Yang, Jiwei Wei, Chongyi Li, and Heng Tao Shen. 2023. Learning Semantic-Aware Knowledge Guidance for Low-Light Image Enhancement. In *CVPR*. 1662–1671.
- [67] Chengpei Xu, Hao Fu, Long Ma, Wenjing Jia, Chengqi Zhang, Feng Xia, Xiaoyu Ai, Binghao Li, and Wenjie Zhang. 2024. Seeing text in the dark: Algorithm and benchmark. *arXiv preprint arXiv:2404.08965* (2024).
- [68] 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.
- [69] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. 2021. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing* 30 (2021), 2072–2086.
- [70] Peng Ye, Tong He, Baopu Li, Tao Chen, Lei Bai, and Wanli Ouyang. 2023. β -DARTS++: Bi-level Regularization for Proxy-robust Differentiable Architecture Search. *CoRR* abs/2301.06393 (2023).
- [71] Ziyu Yue, Jiaxin Gao, and Zhixun Su. 2024. Unveiling Details in the Dark: Simultaneous Brightening and Zooming for Low-Light Image Enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 6899–6907.
- [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] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2020. Learning enriched features for real image restoration and enhancement. In *ECCV*. 492–511.
- [74] Chaoning Zhang, Dongshen Han, Yu Qiao, Jung Uk Kim, Sung-Ho Bae, Seungkyu Lee, and Choong Seon Hong. 2023. Faster Segment Anything: Towards Lightweight SAM for Mobile Applications. *arXiv preprint arXiv:2306.14289* (2023).
- [75] 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*.
- [76] 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*.
- [77] Guijing Zhu, Long Ma, Xin Fan, and Risheng Liu. 2022. Hierarchical Bilevel Learning with Architecture and Loss Search for Hadamard-based Image Restoration. In *IJCAI 2022*. ijcai.org, 1757–1764.
- [78] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. 2020. MetaQA: Deep Meta-Learning for No-Reference Image Quality Assessment. In *Computer Vision and Pattern Recognition*. 14143–14152.