Dual-Path Learning based on Frequency Structural Decoupling and Regional-Aware Fusion for Low-Light Image Super-Resolution

Ji-Xuan He[1,†], Jia-Cheng Zhao[1,†], Feng-Qi Cui[2,*], Jinyang Huang[1,*], Yang Liu[3], Sirui Zhao[2], Meng Li[1], Zhi Liu[4]

[1]Hefei University of Technology, Hefei, China   [2]University of Science and Technology of China, Hefei, China
[3]Zhejiang University, Hangzhou, China   [4]The University of Electro-Communications, Tokyo, Japan

Abstract—Low-light image super-resolution (LLISR) is essential for restoring fine visual details and perceptual quality under insufficient illumination conditions with ubiquitous low-resolution devices. Although pioneer methods achieve high performance on single tasks, they solve both tasks in a serial manner, which inevitably leads to artifact amplification, texture suppression, and structural degradation. To address this, we propose Decoupling then Perceive (*DTP*), a novel frequency-aware framework that explicitly separates luminance and texture into semantically independent components, enabling specialized modeling and coherent reconstruction. Specifically, to adaptively separate the input into low-frequency luminance and high-frequency texture subspaces, we propose a Frequency-aware Structural Decoupling (FSD) mechanism, which lays a solid foundation for targeted representation learning and reconstruction. Based on the decoupled representation, a Semantics-specific Dual-path Representation (SDR) learning strategy that performs targeted enhancement and reconstruction for each frequency component is further designed, facilitating robust luminance adjustment and fine-grained texture recovery. To promote structural consistency and perceptual alignment in the reconstructed output, building upon this dual-path modeling, we further introduce a Cross-frequency Semantic Recomposition (CSR) module that selectively integrates the decoupled representations. Extensive experiments on the most widely used LLISR benchmarks demonstrate the superiority of our *DTP* framework, improving $+1.6\%$ PSNR, $+9.6\%$ SSIM, and -48% LPIPS compared to the most state-of-the-art (SOTA) algorithm. Codes are released at https://github.com/JXVision/DTP.

Index Terms—Low-light super-resolution; dual-path learning;

I. INTRODUCTION

Low-light image super-resolution (LLISR) is a critical task in computer vision, with applications in night surveillance, autonomous driving, and low-light photography [1]. Unlike traditional super-resolution, LLISR must simultaneously address severe luminance degradation and detail loss, making the interaction between luminance and texture a central challenge [2]. As shown in Fig. 1, early approaches often adopt a two-stage paradigm that sequentially performs low-light enhancement and resolution reconstruction [3]. Although achieving high accuracy, such decoupled designs inevitably suffer from error amplification and structural artifacts due to the error accumulation of the serial structure. To address this, recent methods move toward joint frameworks that integrate enhancement and reconstruction in a unified pipeline [4], or leverage frequency priors to suppress artifacts [5]. However,
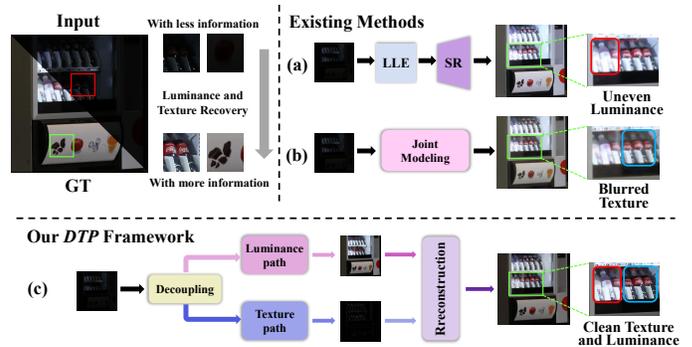
Fig. 1. Comparison of existing methods and our proposed approach. (a) and (b) represent sequential and joint spatial-domain LLISR methods, respectively, which suffer from uneven luminance and blurred textures due to early-stage semantic entanglement. (c) illustrates our frequency-aware disentanglement strategy, which explicitly separates and models luminance and texture components, enabling faithful recovery with both crisp luminance and fine textures.

these methods still struggle with semantic confusion between luminance and texture, especially under extreme lighting conditions [6], which highlights the need for deeper modeling mechanisms and more robust optimization strategies.

While LLISR has achieved remarkable progress, existing methods still struggle in structurally complex regions under severe illumination degradation. A key limitation is their reliance on joint spatial-domain modeling and static fusion, which entangles luminance and texture features and leads to ambiguous semantics, artifact confusion, and degraded perceptual quality. At the heart of these limitations lie three fundamental challenges: **1) Semantic entanglement leads to structural ambiguity.** Luminance and texture are typically modeled together in a unified feature space, without explicit structural disentanglement. This early-stage coupling blurs semantic boundaries and makes it difficult for the model to distinguish luminance artifacts from true textures. As a result, structural distortion and perceptual inconsistency frequently occur in low-light scenarios. **2) Representation lacks semantic selectivity and subspace sensitivity.** Despite the inherently different perceptual roles and frequency characteristics of luminance and texture, most methods adopt flat or undifferentiated representations, *e.g.*, shared channels, uniform scales, or shallow stacks. This lack of subspace-aware

modeling weakens the system's ability to adaptively respond to different semantic targets, leading to feature interference, insufficient specialization, and unstable representation quality. **3) Static fusion fails to adapt to spatial heterogeneity.** Mainstream fusion strategies often employ fixed weights or naive concatenation, without accounting for regional variations in luminance and texture distribution. Such rigid mechanisms are unable to modulate the integration of heterogeneous features based on content relevance, resulting in semantic misalignment, redundancy, and local structural degradation.

To address the aforementioned challenges, we propose Decoupling to Perceive (*DTP*), a frequency-aware low-light super-resolution framework that explicitly decomposes, models, and fuses heterogeneous visual representations in a structurally guided manner. Specifically, to mitigate early-stage semantic entanglement, a Frequency-aware Structural Disentanglement (FSD) module is first introduced. By adaptively transforming input features, FSD separates perceptually distinct luminance and texture components into low- and high-frequency subspaces, reducing cross-domain interference. Based on the disentangled representation, we design a Semantics-specific Dual-path Representation (SDR) strategy that performs targeted modeling for each subspace. The low-frequency luminance branch enhances brightness perception through a bio-inspired mechanism, while the high-frequency texture branch reconstructs fine structural details using residual learning. This design improves robustness and task-specific expressiveness under low-light conditions. To further enhance integration, we propose a Cross-frequency Semantic Recomposition (CSR) module that adaptively fuses the decoupled features via spatial-channel attention, enabling structure-aware fusion and perceptual consistency. In summary, these modules synergistically form a unified pipeline that systematically addresses the core limitations of LLISR, which further achieves superior fidelity, detail recovery, and perceptual quality across diverse low-light scenarios.

Our main contributions are summarized as follows:

- To the best of our knowledge, this paper is the first attempt to introduce the principle of frequency-aware structural disentanglement into LLISR. By explicitly decoupling luminance and texture in the frequency domain and aligning modeling pathways with their semantic roles, we construct a unified *DTP* framework that systematically mitigates early-stage semantic entanglement and enhances perceptual consistency in complex low-light conditions.
- A novel frequency-aware disentanglement module FSD is first proposed by adopting an adaptive transformation scheme to separate luminance and texture into distinct frequency subspaces. This design enables early semantic decomposition and produces compact representations for targeted modeling and interference suppression.
- We further introduce a dual-path modeling and adaptive fusion strategy, consisting of SDR and CSR modules. SDR enhances low-frequency and high-frequency components through biologically inspired adjustment and residual refinement, while CSR employs spatial-channel attention to selectively integrate heterogeneous features, improving

regional generalization and structural fidelity.
- Extensive experiments on standard LLISR benchmarks validate the effectiveness of *DTP*. Our method consistently outperforms all state-of-the-art approaches across ×2 and ×4 scales, which achieves superior perceptual quality, detail restoration, and generalization robustness in diverse low-light scenarios.

II. METHODOLOGY

To address semantic ambiguity and structural degradation in LLISR, by introducing a structured pipeline of decomposition, specialization, and fusion, we propose a frequency-aware low-light super-resolution framework termed Decoupling then Perceive (*DTP*), whose overall workflow is illustrated in Fig. 2. Specifically, *DTP* first employs a Frequency-aware Structural Decoupling (FSD) module to adaptively separate low-light inputs into low-frequency luminance and high-frequency texture components, mitigating early-stage entanglement. On top of the decoupled representation, a Semantics-specific Dual-path Representation (SDR) strategy enhances brightness perception and reconstructs texture details via two dedicated branches. Finally, a Cross-frequency Semantic Recomposition (CSR) module selectively fuses the processed features through joint spatial and channel attention, promoting semantic alignment and perceptual consistency.

A. Frequency-aware Structural Decomposition

To alleviate semantic ambiguity caused by joint spatial modeling [7], we introduce the FSD module to establish a structurally stable representation.

Given a low-light input $O^{llr} \in \mathbb{R}^{H \times W \times 3}$, we first apply a learnable wavelet transform to generate frequency subbands $\{x^{LL}, x^{LH}, x^{HL}, x^{HH}\} = \mathcal{W}_\theta(O^{llr})$, where x^{LL} encodes global luminance and others capture high-frequency textures. To enhance representation consistency, we utilize a learnable vector α (subject to $||\alpha||_1 = 1$) to reweigh the stacked subbands f via $\hat{f} = \alpha^\top \cdot f$. Furthermore, we constrain the luminance component f^{LL} using a KL divergence loss $\mathcal{L}_{KL} = D_{KL}(P_{LL}||\mathcal{N}(\mu_0, \sigma_0^2))$ to ensure compactness. Finally, the fused feature $\hat{f}$ is explicitly separated into independent branches $\{I^{llr}, T^{llr}\}$, which contain structure-preserving low-frequency luminance and fine-grained high-frequency textures, respectively, laying a foundation for targeted processing.

B. Semantics-specific Dual-path Representation

To fully leverage the separated frequency-aware features in the FSD module, we propose a semantically specific dual-path representation (SDR) module. This module introduces two structurally independent branches that model luminance and texture, respectively, allowing each branch to focus on its respective modality. Specifically, the luminance branch focuses on enforcing global illumination consistency, while the texture branch strives to recover fine-grained structure and suppress noise artifacts. SDR effectively avoids interference between luminance and texture representations. This design enables frequency-specific enhancement while suppressing
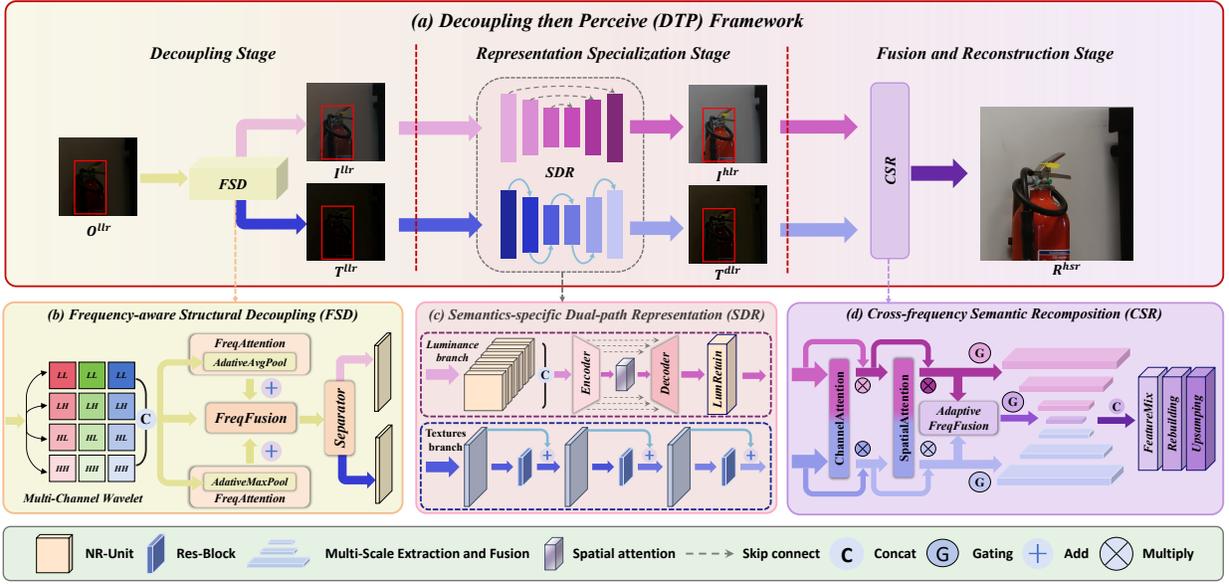
Fig. 2. An overview of the proposed *DTP* (a) The overall processing flow of *DTP*. (b) The pipeline of Decoupling Stage (FSD) in *DTP*. (c) The flow of Representation Specialization Stage (SDR). (d) The sketch of Fusion and Reconstruction Stage (CSR).

cross-domain interference, ensuring robust and interpretable representation learning in low-light conditions.

1) Luminance Enhancement via Bio-inspired Activation: In the luminance pathway, we propose a biologically-inspired enhancement mechanism that mimics the adaptive response of human photoreceptors towards varying illumination. Given the luminance component $\mathcal{I}^{llr}$, a nonlinear transformation modeled on the Naka–Rushton equation is applied to enhance the quality of dark regions while avoiding overexposure. The enhanced luminance is computed as:

$$\mathcal{I}^{hlr} = \frac{(\mathcal{I}^{llr})^\gamma}{(\mathcal{I}^{llr})^\gamma + \sigma^\gamma + \beta} \cdot \frac{\mu_{\text{in}}}{\mu_{\text{out}} + \epsilon}, \quad (1)$$

where γ, σ, and β control the contrast sensitivity, perceptual threshold, and saturation suppression, respectively. μ_{in} and μ_{out} denote the mean intensities before and after activation, respectively, and ϵ is a small constant for numerical stability.

This transformation adaptively scales brightness while maintaining structural contrast, guided by local photometric statistics. Crucially, due to the prior frequency-domain decoupling in FSD, this operation exclusively enhances luminance instead of affecting texture details, thereby preserving semantic integrity and avoiding undesired signal interference.

2) Texture Restoration via Residual-based Hierarchical Denoising: To deal with the high-frequency texture component $\mathcal{T}^{llr}$, which contains rich structural cues but is vulnerable to noise under extreme low-light degradation, we thus introduce a hierarchical denoising framework based on residual learning. The denoising proceeds iteratively, which can be expressed as:

$$\mathcal{T}_i = \mathcal{T}_{i-1} + \mathcal{R}_i(\mathcal{T}_{i-1}), \quad i = 1, \ldots, L, \quad (2)$$

where we set $\mathcal{T}_0 = \mathcal{T}^{llr}$, and each $\mathcal{R}_i(\cdot)$ denotes a residual unit that selectively removes structured noise while preserving fine-grained details.

This progressive refinement scheme enables stable texture enhancement, balancing high-frequency detail preservation and denoising robustness. After L stages, the final representation $\mathcal{T}^{dlr} = \mathcal{T}_L$ serves as a clean and expressive high-frequency feature fed for subsequent semantic fusion.

C. Cross-frequency Semantic Recomposition

To integrate the enhanced luminance and restored texture features into a unified high-fidelity representation [8], we design a CSR module. Instead of naive concatenation, CSR leverages dual-branch attention to perform content-aware, frequency-sensitive feature fusion. By adaptively calibrating the fusion process according to semantic importance, this design ensures perceptual consistency and structural alignment.

Given the outputs from the previous stage, *i.e.*, enhanced luminance $\mathcal{I}^{hlr}$ and denoised texture $\mathcal{T}^{dlr}$, we first apply two parallel attention pathways to assess semantic contributions across spatial and channel dimensions. In particular, spatial attention $\phi_{\text{SA}}(\cdot)$ and channel attention $\phi_{\text{CA}}(\cdot)$ are computed to capture complementary fusion cues. Their contributions are then balanced by a learnable gating function $\mathcal{G}_{\text{att}} \in [0, 1]$, which produces a soft selection mask based on input dynamics and can be represented as:

$$\mathcal{F}^{csr} = \mathcal{G}_{\text{att}} \cdot \phi_{CA}(\mathcal{I}^{hlr}, \mathcal{T}^{dlr}) + (1 - \mathcal{G}_{\text{att}}) \cdot \phi_{SA}(\mathcal{I}^{hlr}, \mathcal{T}^{dlr}). \quad (3)$$

To further mitigate information degradation during fusion, we design a residual refinement path from $\mathcal{T}^{dlr}$, reinforcing high-frequency structure retention. The final high-resolution output $\mathcal{R}^{hsr}$ is reconstructed through a lightweight decoder consisting of feature mixing, structural refinement, and resolution upsampling, which can be denoted as:

$$\mathcal{R}^{hsr} = \text{Upsample}(\text{Rebuild}(\mathcal{F}^{csr} + \mathcal{T}^{dlr})). \quad (4)$$
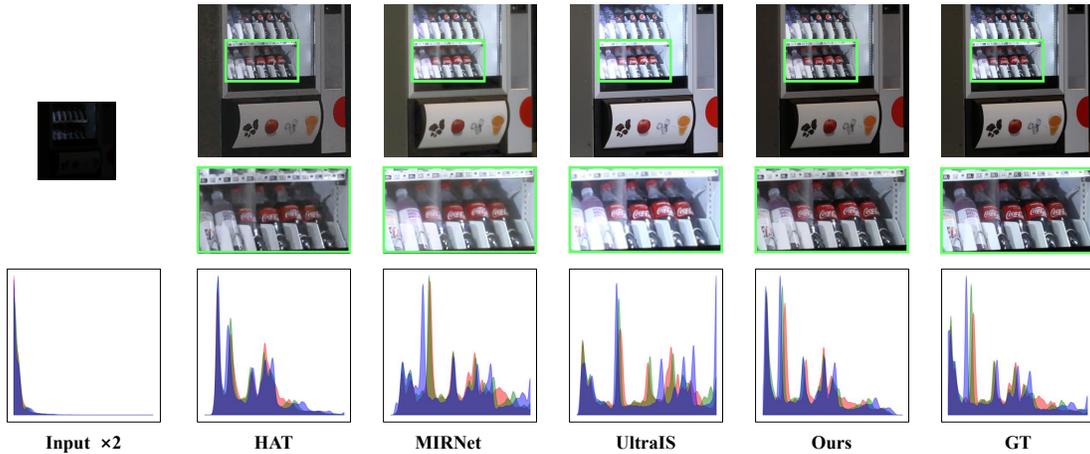
Fig. 3. Qualitative comparisons of ×2 tasks on RELLISUR. The top row shows the full restored images, the middle displays zoomed-in patches for local structure inspection, and the bottom presents RGB histogram distributions.
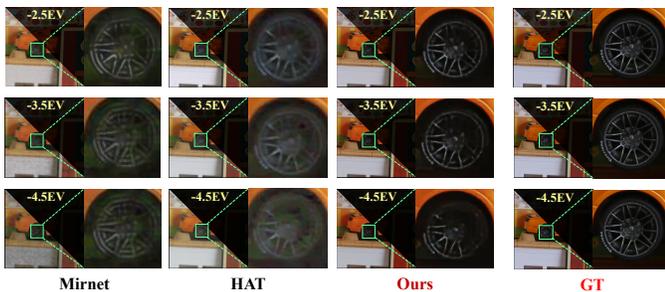

Fig. 4. Visual comparison under extreme low-light conditions (−2.5 EV to −4.5 EV) on the RELLISUR dataset.

By explicitly aligning cross-frequency representations and dynamically adjusting fusion priorities, CSR effectively reconciles global luminance coherence with local texture fidelity. This ensures that the reconstructed image maintains consistent structure and perceptual realism under diverse low-light degradation patterns.

III. EXPERIMENTS

A. Experiment Settings

1) Datasets: We evaluate our proposed method on the RELLISUR dataset [1], a recently curated benchmark for low-light image super-resolution. This dataset provides paired low-resolution dark-light and high-resolution normal-light images at three scaling levels: ×1, ×2, and ×4. Specifically, it contains 3,610 training pairs and 425 test pairs, with the test set encompassing diverse illumination conditions and varying degrees of darkness. To evaluate the robustness of our method under different upscaling demands, we conduct experiments on both the ×2 and ×4 subsets.

2) Validation Metrics: To comprehensively evaluate the quality of image restoration, we adopt three widely used metrics, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS), each capturing different aspects of fidelity and perceptual quality. These metrics jointly evaluate pixel accuracy,

perceptual structure fidelity, and semantic perceptual similarity, offering a comprehensive assessment of restoration quality under challenging low-light conditions.

3) Implementation Details: The specific implementation details and training configurations are provided in the Supplementary Material.

B. Comparison with State-of-the-Artst Methods

We compare our proposed *DTP* framework against a broad range of state-of-the-art methods, not only including general-purpose image restoration models (*e.g.*, PAN, MIRNet, Restormer), super-resolution models (*e.g.*, SwinIR, HAT, OmniSR), but also containing recent LLISR methods tailored to the dual-degradation setting (*e.g.*, PDHAT+, CollaBA, TriCo). All baselines are retrained under unified protocols on the RELLISUR dataset to ensure a fair comparison.

General image restoration methods, while effective in isolated tasks, often struggle with the compound degradation in LLISR due to the lack of semantic decoupling. Their tendency to process brightness and texture jointly in the spatial domain leads to over-smoothed structures and amplified artifacts under extreme lighting conditions. In contrast, our *DTP* framework explicitly separates these factors in the frequency domain, providing more stable and structurally coherent representations. Super-resolution approaches are powerful in capturing long-range dependencies and restoring high-frequency details. However, their performance deteriorates under illumination degradation, largely due to their reliance on spatial coupling and the absence of dedicated luminance modeling. Our approach complements frequency-aware decomposition with perceptually driven brightness enhancement, enabling more robust and balanced recovery. Although SOTA LLISR-specific methods attempt to jointly address illumination and resolution to achieve high performance, they rely on complex fusion strategies and lack structured disentanglement of heterogeneous components that inevitably increase computational cost without solving the underlying semantic interference. Instead of using complex structures, our method adopts a streamlined pipeline, *i.e.*,

TABLE I
COMPARISON OF OUR *DTP* WITH SOTA METHODS ON RELLISUR DATASET AT ×2 AND ×4 SCALES. (**BOLD**: BEST, <u>UNDERLINE</u>: SECOND BEST)

Method	Years	RELLISUR ×2			RELLISUR ×4		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SwinIR [9]	ICCV'21	$18.38_{\uparrow 4.77}$	$0.640_{\uparrow 0.176}$	$0.577_{\downarrow 0.453}$	$17.53_{\uparrow 4.33}$	$0.663_{\uparrow 0.151}$	$0.688_{\downarrow 0.435}$
Restormer [10]	CVPR'22	$21.21_{\uparrow 1.94}$	$0.727_{\uparrow 0.089}$	$0.385_{\downarrow 0.261}$	$20.29_{\uparrow 1.57}$	$0.720_{\uparrow 0.094}$	$0.492_{\downarrow 0.239}$
OmniSR [11]	CVPR'23	$20.87_{\uparrow 2.28}$	$0.740_{\uparrow 0.076}$	$0.470_{\downarrow 0.346}$	$19.89_{\uparrow 1.97}$	$0.723_{\uparrow 0.091}$	$0.461_{\downarrow 0.208}$
HAT [12]	CVPR'23	$20.21_{\uparrow 2.94}$	$0.719_{\uparrow 0.097}$	$0.454_{\downarrow 0.330}$	$19.75_{\uparrow 2.11}$	$0.715_{\uparrow 0.099}$	$0.561_{\downarrow 0.308}$
PDHAT+ [6]	T-MM'24	-	-	-	$21.00_{\uparrow 0.86}$	<u>0.753</u>$_{\uparrow 0.061}$	$0.452_{\downarrow 0.159}$
CollaBA [13]	PR'24	$21.62_{\uparrow 1.53}$	<u>0.787</u>$_{\uparrow 0.029}$	$0.247_{\downarrow 0.123}$	$20.42_{\uparrow 1.44}$	$0.734_{\uparrow 0.080}$	<u>0.371</u>$_{\downarrow 0.118}$
TriCo [14]	ACM MM'24	$22.45_{\uparrow 0.70}$	$0.744_{\uparrow 0.072}$	$0.304_{\downarrow 0.180}$	$21.05_{\uparrow 0.81}$	$0.731_{\uparrow 0.083}$	$0.432_{\downarrow 0.179}$
UltraIS [15]	T-NNLS'24	$22.26_{\uparrow 0.89}$	$0.743_{\uparrow 0.073}$	$0.254_{\downarrow 0.130}$	$21.03_{\uparrow 0.83}$	$0.726_{\uparrow 0.088}$	$0.371_{\downarrow 0.118}$
MSIRNet [2]	IF'24	$21.33_{\uparrow 1.82}$	$0.750_{\uparrow 0.066}$	$0.365_{\downarrow 0.241}$	$20.32_{\uparrow 1.54}$	$0.735_{\uparrow 0.079}$	$0.437_{\downarrow 0.184}$
BrZoNet [4]	AAAI'24	<u>22.79</u>$_{\uparrow 0.36}$	$0.745_{\uparrow 0.071}$	<u>0.243</u>$_{\downarrow 0.119}$	<u>21.41</u>$_{\uparrow 0.45}$	$0.726_{\uparrow 0.088}$	$0.383_{\downarrow 0.130}$
DTP (Ours)	-	**23.15**	**0.816**	**0.124**	**21.86**	**0.814**	**0.253**

TABLE II
ABLATION STUDY OF FSD, SDR, AND CSR MODULES IN *DTP* ACROSS ×2 AND ×4 UPSCALING FACTORS.

Scale	Modules			Metric		
	FSD	SDR	CSR	PSNR↑	SSIM↑	LPIPS↓
	×	×	×	18.51	0.765	0.387
	✓	×	×	$22.36_{\uparrow 3.85}$	$0.803_{\uparrow 0.038}$	$0.291_{\downarrow 0.096}$
	×	✓	×	$20.27_{\uparrow 1.76}$	$0.782_{\uparrow 0.017}$	$0.362_{\downarrow 0.025}$
×2	×	×	✓	$21.79_{\uparrow 3.28}$	$0.791_{\uparrow 0.026}$	$0.313_{\downarrow 0.074}$
	✓	✓	×	$22.75_{\uparrow 4.24}$	$0.811_{\uparrow 0.046}$	$0.203_{\downarrow 0.184}$
	✓	×	✓	$22.73_{\uparrow 4.22}$	$0.813_{\uparrow 0.046}$	$0.185_{\downarrow 0.202}$
	×	✓	✓	$22.13_{\uparrow 3.62}$	$0.809_{\uparrow 0.046}$	$0.239_{\downarrow 0.148}$
	✓	✓	✓	$23.15_{\uparrow 4.64}$	$0.817_{\uparrow 0.052}$	$0.124_{\downarrow 0.263}$
	×	×	×	17.72	0.733	0.479
	✓	×	×	$21.11_{\uparrow 3.39}$	$0.801_{\uparrow 0.068}$	$0.393_{\downarrow 0.086}$
	×	✓	×	$19.41_{\uparrow 1.69}$	$0.763_{\uparrow 0.030}$	$0.420_{\downarrow 0.059}$
×4	×	×	✓	$20.03_{\uparrow 2.31}$	$0.787_{\uparrow 0.054}$	$0.407_{\downarrow 0.072}$
	✓	✓	×	$21.47_{\uparrow 3.75}$	$0.809_{\uparrow 0.076}$	$0.326_{\downarrow 0.153}$
	✓	×	✓	$21.33_{\uparrow 3.61}$	$0.811_{\uparrow 0.078}$	$0.309_{\downarrow 0.170}$
	×	✓	✓	$20.79_{\uparrow 3.07}$	$0.806_{\uparrow 0.073}$	$0.377_{\downarrow 0.102}$
	✓	✓	✓	$21.86_{\uparrow 4.14}$	$0.814_{\uparrow 0.081}$	$0.253_{\downarrow 0.226}$

decoupling, specialization, and recomposition that promotes semantic clarity, localized adaptation, and perceptual alignment.

Overall, the results confirm that *DTP* 's structured modeling offers consistent gains across both fidelity and perceptual quality, especially in challenging low-light environments where conventional designs falter.

C. Visualization

1) Qualitative Visualization: To further validate the perceptual effectiveness of our method under severe low-light degradation, we present qualitative comparisons on the REL-LISUR dataset in Fig. 3. Each sample displays the restored full image (top), a zoomed-in region for local structure inspection (middle), and RGB histogram distributions (bottom). Compared to existing methods, our approach demonstrates superior restoration fidelity both visually and statistically. Specifically, general SR and LLISR methods often struggle to jointly recover global luminance and fine textures, resulting in either over-smoothed appearance or noisy over-enhancement. In contrast, our model produces visually coherent outputs with clearly restored structural boundaries and sharp local textures. The highlighted patch shows that bottle contours and label text are significantly more legible in our result, indicating better preservation of high-frequency details. Moreover, the RGB histogram of our reconstruction closely aligns with that of the ground truth, capturing both intensity spread and color consistency. This suggests that our frequency-aware decoupling and cross-semantic fusion contribute not only to spatial structure recovery but also to perceptual alignment across illumination and chrominance domains. The visualization clearly evidences our method's advantage in addressing dual degradation by the proposed semantically disentangled method and the structurally consistent manner.

2) Robustness under Extreme Low-Light Conditions: To evaluate the robustness of our method across varying levels of illumination degradation, we conduct a stress test on the RELLISUR dataset under extreme low-light settings ranging from −2.5 EV to −4.5 EV, as shown in Fig. 4. As the exposure drops, baseline methods progressively lose the ability to recover wheel contours and boundary edges, exhibiting either oversaturation or structure collapse. In contrast, our method retains the geometric layout and spoke sharpness even under −4.5 EV, demonstrating superior resilience to noise, blur, and luminance suppression. The zoom-in patches clearly show that our model avoids both over-smoothing and over-enhancement, instead producing coherent reconstructions with natural brightness and preserved detail. This robustness stems from our frequency-aware decoupling design and semantics-specific enhancement strategy, which isolate and reconstruct luminance and texture components under tailored constraints. The results highlight our model's ability to adapt to severe degradations while ensuring fine-grained detail recovery and perceptual realism, outperforming conventional approaches

across the entire exposure range.

D. Ablation Study

1) Effectiveness of Each Module: To evaluate the impact of each proposed component and their combined effects, we perform ablation studies by testing each module individually and in various combinations. As shown in Tab. II, all modules contribute positively to model performance, with the full *DTP* framework achieving the best results across scales and metrics. We first analyze the idividual contribution of each module. When a single module is activated, the FSD, SDR, CSR modules all yield a significant performance boost over the baseline. Specifically, FSD alone improves the PSNR from 18.51 dB to 22.36 dB, hightlighting its effectiveness in decoupling entangled semantics at the input stage. Next, we explore the synergistic effects of combining two modules. The combination of FSD and SDR further enhances the PSNR to 22.75 dB, which is a notable improvement over either module individually, validating their complementary roles in semantic confusion and disentangled representation. Similarly, pairing FSD with CSR achieves a PSNR of 22.75 dB, while SDR with CSR results in a PSNR of 22.13 dB. These results indicate that our modules are designed to work together, with each component reinforcing the others' strengths. The performance consistently improves as more modules are integrated, which effectively validates our design principle, progressive decoupling, specialization, and recomposition, further enabling more robust and semantically aligned restoration under low-light degradation.

2) Effectiveness of Sub-Components in SDR Module: To explore the core of the SDR module, we further validate the effectiveness of the BioLuma and HicDeno components within the SDR module. The specific quantitative results and analysis are presented in the Supplementary Material.

IV. Conclusion

To tackle the dual challenges of luminance degradation and texture loss in low-light image super-resolution, we design a frequency-aware restoration framework *DTP*. Unlike conventional approaches that entangle semantic components in spatial-domain modeling, *DTP* explicitly separates luminance and texture into semantically independent frequency subspaces through a frequency-aware structural decoupling process. Based on this disentangled representation, *DTP* employs a semantics-specific dual-path mechanism to perform dedicated luminance enhancement and texture reconstruction, followed by a cross-frequency semantic recomposition module that adaptively fuses heterogeneous cues with spatial and channel attention to ensure perceptual alignment and structural fidelity, respectively. Totally, these modules form a unified framework that systematically mitigates semantic confusion and enhances generalization under diverse and extremely low-light conditions. Extensive experiments on public benchmarks validate the superiority of *DTP* on accuracy and recognition rate in terms of both distortion-based and perceptual quality metrics.

References

[1] Andreas Aakerberg, Kamal Nasrollahi, and Thomas B Moeslund, "Rellisur: A real low-light image super-resolution dataset," in *Thirty-fifth Conference on Neural Information Processing Systems-NeurIPS 2021*, 2021.

[2] Jing Ye, Shenghao Liu, Changzhen Qiu, and Zhiyong Zhang, "Learning multi-granularity semantic interactive representation for joint low-light image enhancement and super-resolution," *Information Fusion*, vol. 110, pp. 102467, 2024.

[3] Chongyi Li, Chunle Guo, and Chen Change Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 8, pp. 4225–4238, 2021.

[4] Ziyu Yue, Jiaxin Gao, and Zhixun Su, "Unveiling details in the dark: Simultaneous brightening and zooming for low-light image enhancement," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, vol. 38, pp. 6899–6907.

[5] Guanqun Liu, Xin Wang, Lei Wang, Daren Zha, Lin Zhao, Zhe Kong, and Peng Qi, "A noise-aware framework for blind image super-resolution," in *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2022, pp. 01–06.

[6] Yuezhou Li, Rui Xu, Yuzhen Niu, Wenzhong Guo, and Tiesong Zhao, "Perceptual decoupling with heterogeneous auxiliary tasks for joint low-light image enhancement and deblurring," *IEEE Transactions on Multimedia*, vol. 26, pp. 6663–6675, 2024.

[7] Feng-Qi Cui, Anyang Tong, Jinyang Huang, Jie Zhang, Dan Guo, Zhi Liu, and Meng Wang, "Learning from heterogeneity: Generalizing dynamic facial expression recognition via distributionally robust optimization," in *Proceedings of the 33rd ACM international conference on multimedia*, 2025, pp. 5587–5596.

[8] Jinyang Huang, Yuanhao Feng, Feng-Qi Cui, Xiang Zhang, Zhi Liu, Xin Liu, Jianchun Liu, Fusang Zhang, and Meng Li, "Identifying who you are no matter what you write through abstracting handwriting style," *IEEE Transactions on Dependable and Secure Computing*, pp. 1–15, 2026.

[9] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte, "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1833–1844.

[10] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5728–5739.

[11] Hang Wang, Xuanhong Chen, Bingbing Ni, Yutian Liu, and Jinfan Liu, "Omni aggregation networks for lightweight image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22378–22387.

[12] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong, "Activating more pixels in image super-resolution transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 22367–22377.

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

[14] Jiaxin Gao and Yaohua Liu, "Enhancing images with coupled low-resolution and ultra-dark degradations: A tri-level learning framework," in *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 8642–8651.

[15] Jiaxin Gao, Ziyu Yue, Yaohua Liu, Sihan Xie, Xin Fan, and Risheng Liu, "A dual-stream-modulated learning framework for illuminating and super-resolving ultra-dark images," *IEEE transactions on neural networks and learning systems*, 2024.