SCI-Net: Semantic-Guided Color and Intensity Decoupling Network for Low-Light Image Enhancement

Chuancheng Fu[1], Zhaokun He[1], Li Chen[1,2,*]
[1]School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China.
[2]Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan, China.
Email: {fcc13093896521@wust.edu.cn, hezhaokun0602@163.com, chenli@wust.edu.cn}

***Abstract*—Low-light image enhancement (LLIE) is vital for vision system reliability. However, most existing methods remain "semantically agnostic" pixel-wise mappings that ignore scene content, leading to blurred edges and color deviations. While foundation models like SAM offer powerful semantic extraction, the low-light-induced domain shifts impair the reliability of captured priors. To address this, we propose the Semantic-Guided Color and Intensity Decoupling Network (SCI-Net), focusing on two key challenges: (1) acquiring reliable semantic priors from degraded images, and (2) efficiently incorporating these features into image restoration process. For the first challenge, we design a Residual Fourier Guidance Module (RFGM) that generates high-quality pseudo-enhanced images via phase structural compensation and amplitude refinement, ensuring reliable inputs for semantic extraction. For the second challenge, we introduce the Cross-Modal Guided Attention (CMGA) module. By employing a cross-modal interaction with semantic features as the Key, CMGA precisely injects guidance into the decoupled HVI branches (HV and I), effectively leveraging the inherent color-intensity separation of the HVI color space. Extensive experiments on eight datasets demonstrate that SCI-Net outperforms existing state-of-the-art methods in contrast enhancement, detail recovery, and semantic consistency.**

***Index Terms*—Low-Light Image Enhancement, HVI Color Space, Semantic Guidance, SAM, Fourier Domain Enhancement**

I. Introduction

Low-light image enhancement (LLIE) is vital for vision system reliability. Historically, research has established robust frameworks, from Retinex-based models to deep-learning-based pixel mappings, achieving impressive results in brightness restoration. Despite progress in deep learning, most methods remain "semantically agnostic" pixel-wise mappings that ignore scene content. Lacking semantic awareness, they typically employ uniform enhancement strategies, failing to achieve content-adaptive restoration. Consequently, they struggle to balance structural details and color, leading to blurred edges and color deviations, as shown in Fig. 1 (a).

Foundation models like the Segment Anything Model (SAM) offer new tools for semantic extraction, yet their direct application to LLIE is limited. Extreme noise and low contrast induce a significant domain shift, severely impairing the ability to capture reliable scene semantics. Consequently, as shown in Fig. 1 (b), methods directly utilizing SAM, such as PPTformer [1], still suffer from chromatic aberrations and structural loss. Furthermore, existing works typically treat semantic features as static priors, lacking deep interaction with low-level features to achieve content-aware guidance.

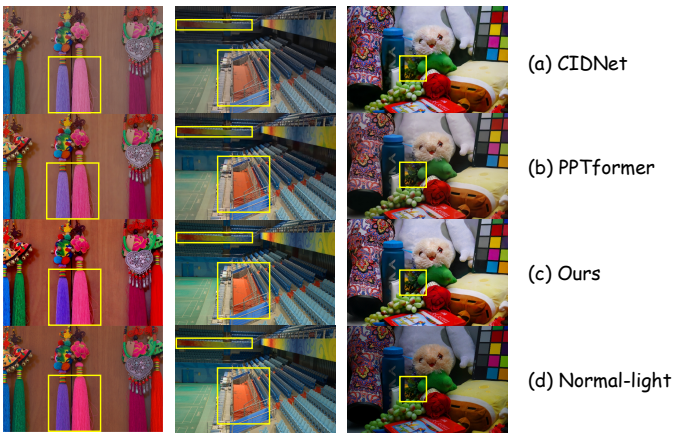

Fig. 1. Visual comparison of different methods against normal-light image: (a) CIDNet (semantic-agnostic): obvious artifacts and color deviation; (b) PPTformer (direct SAM input): reduced artifacts but detail loss; (c) Our SCI-Net (semantic-guided): clear textures, natural tones, close to (d) normal-light.

Meanwhile, the HVI color space provides an efficient decoupling paradigm. The I branch handles illumination, while HV branches focus on denoising and chromaticity correction. This structure serves as an ideal pivot for semantic integration, where multi-scale semantic features can offer targeted guidance for both I branch illuminance recovery and refined chromatic restoration within the HV branches. However, existing HVI-based methods rarely exploit the synergy between large-scale semantic models and this decoupling paradigm.

To address these challenges, we propose the **S**emantic-Guided **C**olor and **I**ntensity Decoupling Network (SCI-Net), delivering superior enhancement performance, as shown in Fig. 1 (c). SCI-Net establishes a closed loop from feature rectification to deep-fusion guidance. Specifically, the Residual Fourier Guidance Module (RFGM) generates pseudo-enhanced images via phase

* Correspoding Author

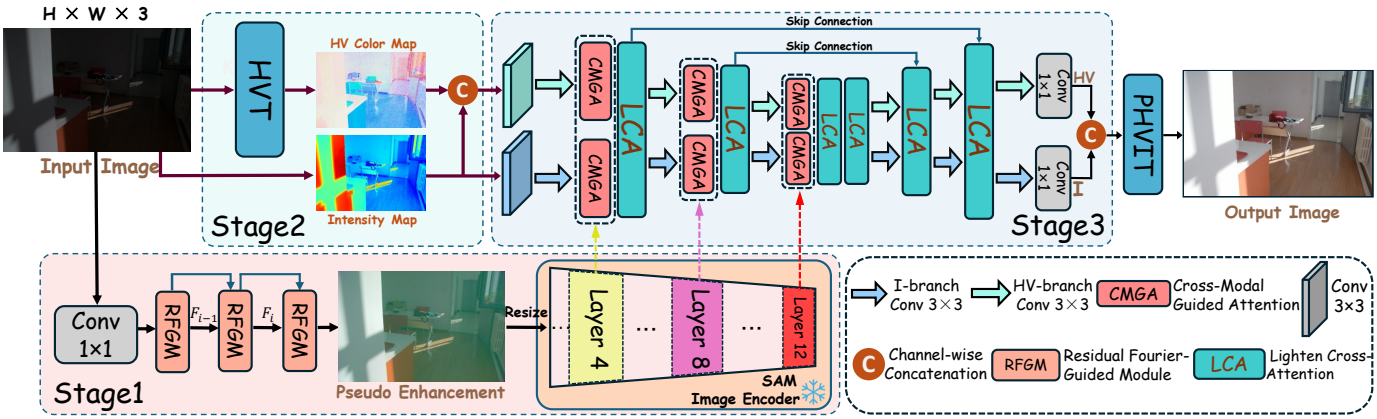

Fig. 2. The overview of the proposed SCI-Net. **Stage 1:** The input image is processed via RFGM, followed by a frozen SAM encoder to extract multi-scale semantic priors. **Stage 2:** The HVI Color Transformation (HVIT) decouples the input into HV and I maps. **Stage 3:** A dual-branch U-Net utilizes CMGA and LCA modules to perform semantic-guided enhancement. Finally, the PHVIT module reconstructs the enhanced HVI maps to generate the final sRGB image.

structure compensation and amplitude optimization guided by phase differences. These images serve as reliable inputs for SAM, enabling accurate semantic extraction. Subsequently, the Cross-Modal Guidance Attention (CMGA) module utilizes semantic features as the Key to inject multi-scale semantic information into the decoupled HV and I branches, imposing explicit, content-aware constraints throughout the enhancement process.

The main contributions of our work are as follows:

- A novel SCI-Net is proposed to synergize SAM's multi-scale priors with HVI space, effectively suppressing semantic-inconsistency artifacts and color distortion.
- The RFGM module is introduced to leverage the structural robustness of phase information, providing reliable SAM inputs via phase structure compensation and phase-difference-guided amplitude optimization.
- We design the CMGA module, which utilizes semantic features as the Key to provide precise semantic guidance for the decoupled branches.

II. Methods

A. Network Architecture

Accurate scene structure recovery under low-light requires precise content awareness, yet the domain shift in low-light images renders direct semantic extraction from foundation models unreliable. Therefore, our SCI-Net focuses on two key challenges: capturing reliable semantic priors from degraded images and integrating them into the image restoration.

As illustrated in Fig. 2, the entire network pipeline is divided into three consecutive stages. **Stage 1: Multi-Level Semantic Features Extraction.** The RFGM derives a structure-preserving pseudo-enhanced image $\mathbf{I}_P$ from the input image $\mathbf{I}_L$ via phase structure compensation and phase-difference guided magnitude optimization. Subsequently, a frozen SAM encoder extracts hierarchical high-fidelity semantic features $\{\mathbf{F}_S^L, \mathbf{F}_S^M, \mathbf{F}_S^H\}$ from $\mathbf{I}_P$. **Stage 2: HVI Color Transformation.** The HVI Color Transformation (HVIT) module decouples the raw image into HV and I components, initializing the dual branches. **Stage 3: Image Enhancement**. Within the dual-branch U-Net, the CMGA module employs a cross-modal mechanism, utilizing semantic features as the Key to inject multi-scale semantic priors into the HV and I branches for guided enhancement. Meanwhile, the LCA module facilitates periodic interaction between the two branches. Finally, the results are reconstructed via PHVIT to produce the final image $\mathbf{I}_R$. Details are provided below.

B. Multi-Level Semantic Features Extraction

Due to noise and extreme low-light conditions, pre-trained SAM often struggles to extract reliable semantics directly. Therefore, we design the Residual Fourier Guidance Module (RFGM), which leverages the structural robustness of Fourier phase components to generate pseudo-enhanced images $\mathbf{I}_P$. The core of RFGM comprises two guidance mechanisms:

Phase Structure Guidance: As shown in Fig. 3, input features are decomposed into amplitude $\mathbf{A}_{i-1}$ and phase $\mathbf{P}_{i-1}$ via FFT. $\mathbf{P}_{i-1}$ is preliminarily denoised to $\mathbf{P}_i$. To extract the most reliable structural prior, we compute the channel-wise cosine similarity $\mathrm{MS}(\cdot,\cdot)$ between $\mathbf{P}_{i-1}$ and $\mathbf{P}_i$ to obtain a similarity matrix $\mathbf{M} \in \mathbb{R}^{C\times C}$. From $\mathbf{M}$, we identify the Top-1 index to locate the channel with the most stable structural information between the two phases, since higher similarity indicates greater resilience to low-light noise. By leveraging this most reliable channel as a structural reference, it is processed via convolution and Sigmoid activation to adaptively modulate $\mathbf{P}_i$, yielding the final phase $\mathbf{P}_{final}$:

$$\begin{aligned} \mathbf{M} &= \mathrm{MS}(\mathbf{P}_{i-1}, \mathbf{P}_i) \\ \mathrm{Index} &= \text{Top-1}(\mathbf{M}) \\ \mathbf{P}_{final} &= \mathbf{P}_i \odot \mathrm{Sigmoid}(\mathrm{Conv}(\mathbf{P}_{i-1}[\mathrm{Index}])) \end{aligned} \tag{1}$$

Phase Difference Guided Magnitude Refinement: Simultaneously, $\mathbf{A}_{i-1}$ is spatially processed to obtain the initially enhanced magnitude $\mathbf{A}_i$. To keep brightness restoration consistent with the structure and avoid unnatural results, we calculate the phase difference $\Delta\mathbf{P} = \mathbf{P}_{final} - \mathbf{P}_{i-1}$. Then, $\mathbf{A}_i$ and $\Delta\mathbf{P}$ are

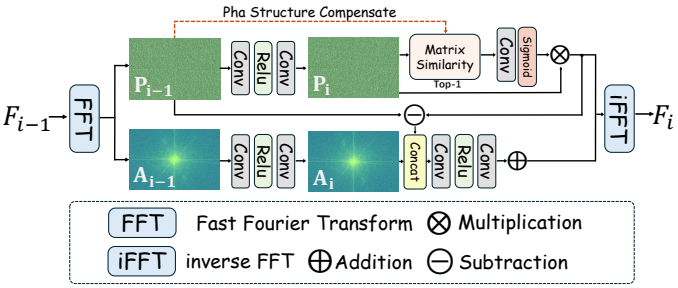

Fig. 3. Residual Fourier Guided Module (RFGM). RFGM transforms the input feature F_{i-1} into the enhanced feature F_i via phase structure guidance and phase-difference guided magnitude refinement.

concatenated along the channel dimension. Convolutional layers extract the brightness correction under structural constraints, which is integrated via a residual connection to produce the final magnitude $\mathbf{A}_{final}$:

$$\mathbf{A}_{final} = \mathbf{A}_i + \text{Conv}(\text{Concat}(\mathbf{A}_i, \Delta\mathbf{P})) \tag{2}$$

In implementation, three RFGM modules are cascaded to progressively refine the frequency features and output the pseudo-enhanced image $\mathbf{I}_P$. This image is subsequently fed into a pre-trained SAM image encoder for semantic parsing.

Finally, to efficiently harness SAM's general semantic priors, we adopt a parameter-freezing extraction strategy. Specifically, we use the smallest and most efficient SAM-B (ViT-Base) variant of the Segment Anything Model, which contains a total of 12 transformer layers in its image encoder. By removing absolute position encodings to accommodate flexible input resolutions and utilizing a hook mechanism, multi-scale semantic features $\mathbf{F}_S^l$ are extracted from various layers of the image encoder to provide fine-grained guidance:

$$\mathbf{F}_S^l = \text{SAM}(\mathbf{I}_P)\big|_{\text{Layer } k} \tag{3}$$

where $(l, k) \in \{(L, 4), (M, 8), (H, 12)\}$.

C. Cross-Modal Guided Attention

Unlike conventional methods that rely on simple concatenation for auxiliary information, the proposed CMGA module is embedded within the U-Net encoder to facilitate scene-aware interaction between branch features (HV or I branch) and multi-scale semantic priors $\mathbf{F}_S^{\{L,M,H\}}$. For notation simplicity, we denote the branch feature as $\mathbf{x}$ and the semantic prior at the corresponding scale as $\mathbf{y}$.

The core philosophy of CMGA is to utilize branch feature $\mathbf{x}$ to provide Query ($\mathbf{Q}_x$) and Value ($\mathbf{V}_x$), while attending to the Key ($\mathbf{K}_y$) that is synthesized from the interaction between semantic feature $\mathbf{y}$ and branch feature $\mathbf{x}$. By performing local spatial operations on the concatenated $[\mathbf{x}, \mathbf{y}]$, the semantic information within $\mathbf{K}_y$ is dynamically aligned with the visual context, ensuring precise semantic prior injection:

$$\mathbf{Q}_x = \text{Proj}_Q(\mathbf{x}), \quad \mathbf{V}_x = \text{Proj}_V(\mathbf{x}) \tag{4}$$

$$\mathbf{K}_y = \text{Proj}_K(\text{Conv}(\text{Concat}(\mathbf{x}, \mathbf{y}))) \tag{5}$$

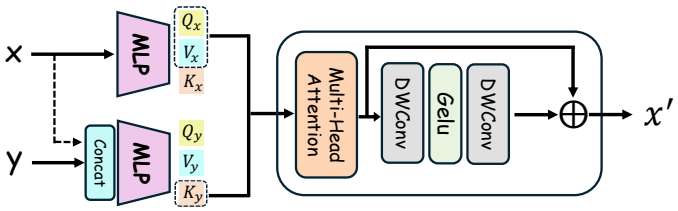

Fig. 4. Cross-Modal Guided Attention (CMGA). CMGA generates $\mathbf{K}_y$ by fusing branch feature $\mathbf{x}$ and semantic prior $\mathbf{y}$ via concatenation and convolution, while deriving $\mathbf{Q}_x$ and $\mathbf{V}_x$ from $\mathbf{x}$. After multi-head attention and DWConv, it outputs the branch feature $\mathbf{x}'$ injected with semantic information.

Specifically, CMGA computes the cross-modal attention map via the interaction of $\mathbf{Q}_x$ and $\mathbf{K}_y$, using a learnable factor ζ to adjust the attention sharpness. To maintain spatial consistency, a positional embedding (Pos) is parallelly extracted from $\mathbf{V}_x$ via depth-wise convolutions. This design compensates for the permutation-invariant nature of self-attention by re-injecting local structural priors into the enhanced features. The refined feature $\mathbf{x}'$ is formulated as:

$$\mathbf{x}' = \text{Softmax}(\mathbf{Q}_x\mathbf{K}_y^T \cdot \zeta)\mathbf{V}_x + \text{Pos}(\mathbf{V}_x) \tag{6}$$

Consequently, the semantically-infused branch feature $\mathbf{x}'$ effectively enhances restoration accuracy: the I branch leverages semantic boundary priors to align illumination with structures, while the HV branch utilizes category-specific priors to guide color correction and denoising, ensuring semantic consistency in the final output.

D. Loss Function

To integrate the advantages of both RGB and HVI spaces, we convert the normal-light image to HVI space to constrain the HVI output, while transforming the output back to RGB for constraint against the normal-light image. We define a universal loss $\mathcal{L}_j$ to supervise both color spaces, where $j \in \{\text{RGB}, \text{HVI}\}$:

$$\mathcal{L}_j = \lambda_{L1}\mathcal{L}_{L1} + \lambda_s\mathcal{L}_{ssim} + \lambda_e\mathcal{L}_{edge} + \lambda_p\mathcal{L}_{perc} \tag{7}$$

$$\mathcal{L}_{total} = \mathcal{L}_{RGB} + \lambda_{HVI}\mathcal{L}_{HVI} \tag{8}$$

In our experiments, λ_{HVI} is set to 1, and other hyperparameter configurations are detailed in the experimental section.

III. Experiments

A. Experimental Settings

Datasets and Metrics. We evaluate our model on LOLv1 [2] (485 training/15 testing pairs), as well as the real (689 training/100 testing pairs) and synthetic (900 training/100 testing pairs) subsets of LOLv2. Additionally, five unpaired datasets are employed: DICM [3], LIME [4], MEF [5], NPE [6], and VV [7] (containing 64, 10, 17, 85, and 24 images respectively). Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [8], and Learned Perceptual Image Patch Similarity (LPIPS) [9] are used for LOLv1 and LOLv2, while

TABLE I

QUANTITATIVE COMPARISON OF DIFFERENT METHODS ON LOLV1 AND LOLV2 DATASETS. THE BEST AND SECOND-BEST RESULTS ARE HIGHLIGHTED IN **BOLD** AND UNDERLINED, RESPECTIVELY. SYMBOLS "↑" (OR "↓") INDICATE THAT HIGHER (OR LOWER) VALUES ARE BETTER.

Methods	LOLv1			LOLv2-Real			LOLv2-Synthetic			Params	FLOPs
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	(M)	(G)
RetinexNet [2]	18.915	0.427	0.470	16.097	0.401	0.543	17.137	0.762	0.255	0.84	584.47
EnlightenGAN [11]	20.001	0.691	0.317	18.233	0.617	0.309	16.572	0.734	0.220	8.64	61.01
KinD [12]	23.018	0.843	0.156	17.544	0.669	0.375	18.320	0.796	0.252	8.02	34.99
ZeroDCE [13]	21.880	0.640	0.335	16.059	0.580	0.313	17.712	0.815	0.169	0.075	4.83
LLFlow [14]	21.149	0.854	0.119	17.433	0.832	0.176	17.119	0.812	0.224	17.42	358.4
SNR-Aware [15]	23.931	0.846	0.152	21.480	0.849	0.163	24.140	0.928	0.056	4.01	26.35
LYT-Net [16]	21.607	0.813	0.134	21.092	0.834	0.163	22.045	0.895	0.090	0.045	3.49
GSAD [17]	22.021	0.850	0.137	20.153	0.846	0.115	24.472	0.930	0.051	17.36	442.02
PairLIE [18]	23.526	0.755	0.248	19.885	0.778	0.317	19.074	0.800	0.230	0.33	20.81
MambaLLIE [19]	22.801	0.832	0.090	21.853	0.828	0.167	25.874	0.940	0.047	20.85	2.28
LightenDiff [20]	23.620	0.829	0.180	22.878	0.855	0.166	21.582	0.869	0.153	-	-
RetinexMamba [21]	24.003	0.827	0.147	22.452	0.845	0.174	25.883	0.935	0.055	4.59	42.82
CWNet [22]	23.601	0.850	0.123	21.647	0.860	0.135	25.501	0.936	0.048	1.23	11.3
CIDNet [23]	23.809	0.857	0.086	23.904	0.865	0.121	25.129	0.938	0.045	1.88	7.57
Ours	**24.039**	**0.860**	**0.085**	**24.205**	**0.872**	**0.114**	**25.910**	**0.939**	**0.044**	88.25	190.19

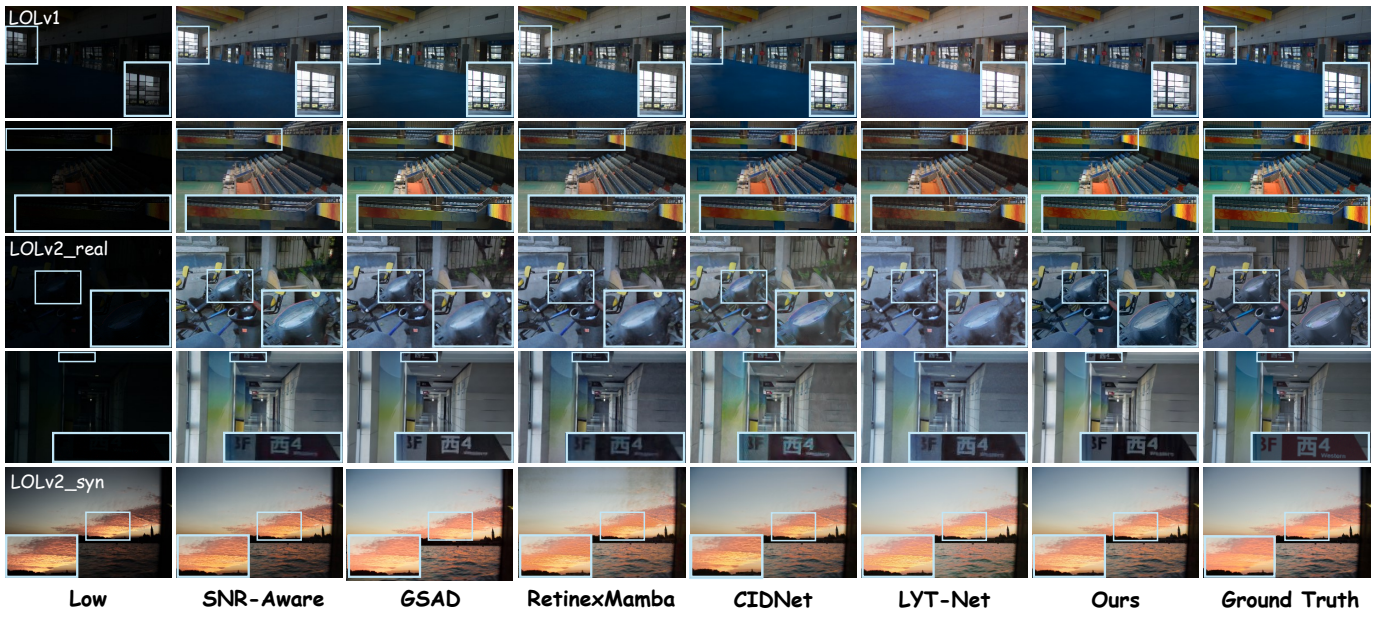

Fig. 5. Visual comparisons of enhancement results by using different methods on the LOLv1, LOLv2-real, and LOLv2-synthetic datasets.

Natural Image Quality Evaluator (NIQE) [10] is adopted for unpaired datasets.

Implementation Details. Our U-net-based SCI-Net is trained on a single RTX 2070 GPU with a patch size of 256×256 and a batch size of 2 for 1000 epochs. For the optimizer, we use the Adam optimizer with an initial learning rate of 1×10^{-4}, which is gradually reduced to 1×10^{-7} via a cosine annealing scheme. For the loss function weights, the coefficients for SSIM loss, edge loss [24], and perceptual loss [25] are set to 0.50, 50.00, and 0.01, respectively.

B. Comparisons with States of the Art

Results on LOL Datasets. As shown in Table I, SCI-Net outperforms SOTA methods across all metrics. The high PSNR stems from RFGM's frequency-domain denoising, while superior SSIM and LPIPS result from the CMGA, which injects semantic features into the HV and I branches to maintain structural consistency. Visual comparisons in Fig. 5 further validate these quantitative gains. Unlike the blurring or color distortion prevalent in existing methods, our approach effectively suppresses local overexposure in highlights while restoring faithful colors and intricate textures. In summary, SCI-Net effectively balances artifact removal and detail reconstruction, maintaining natural illumination and structural integrity.

Results on Unpaired Datasets. We further evaluate SCI-Net on five unpaired datasets using the NIQE metric. As shown in Table II, SCI-Net achieves the best performance on DICM, MEF, and NPE, demonstrating superior naturalness and realism.

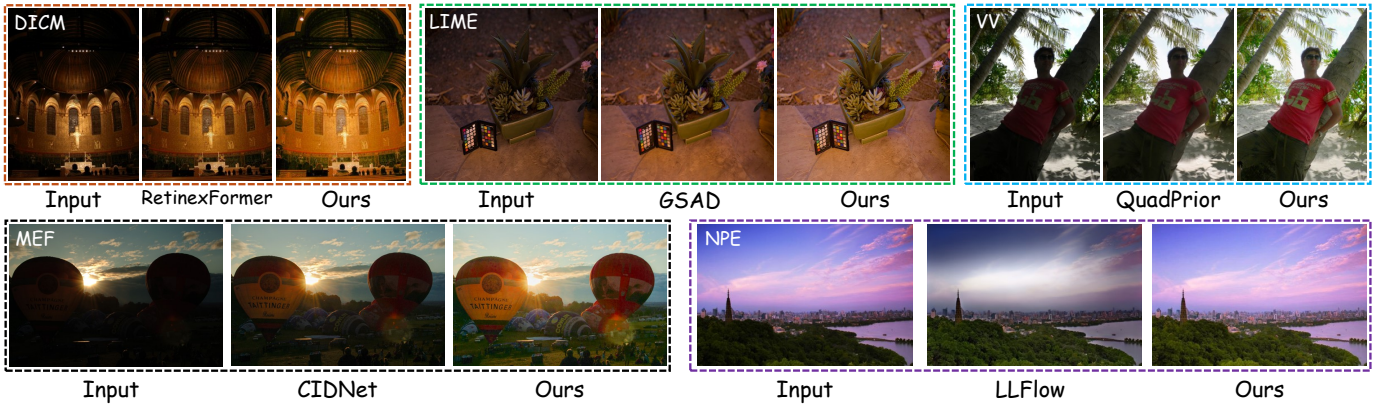

Fig. 6. Visual comparison across five unpaired datasets. Following the evaluation protocol of RetinexFormer [26], we select a representative image from each dataset to demonstrate the qualitative advantages of our method over existing approaches.

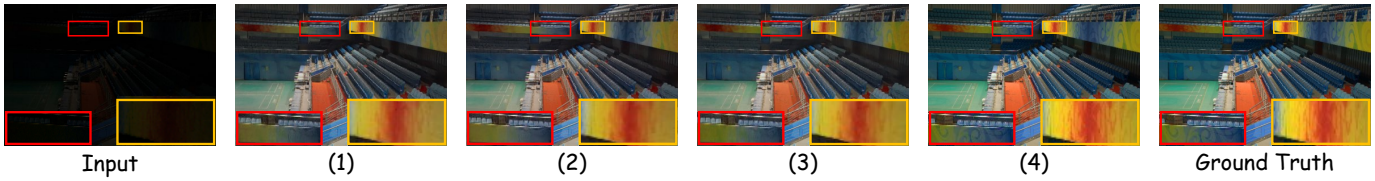

Fig. 7. Qualitative comparison of enhancement results across four ablation studies on the LOLv1 dataset. The labels (1) ~ (4) are in direct correspondence with the experimental configurations detailed in Table III.

TABLE II
QUANTITATIVE COMPARISON OF DIFFERENT METHODS ON DICM, LIME, MEF, NPE, AND VV DATASETS FOR NIQE. THE BEST AND SECOND-BEST RESULTS ARE HIGHLIGHTED IN **BOLD** AND UNDERLINED, RESPECTIVELY. A SMALLER NIQE VALUE INDICATES BETTER PERCEPTUAL QUALITY.

Methods	DICM	LIME	MEF	NPE	VV
ZeroDCE [13]	4.58	5.82	4.93	4.53	4.81
LLFlow [14]	4.06	4.59	4.70	4.67	4.04
SNR-Aware [15]	4.71	5.74	4.18	4.32	3.87
GSAD [17]	4.03	4.21	3.67	4.17	**3.11**
QuadPrior [27]	4.82	5.22	5.11	4.78	3.99
CIDNet [23]	3.55	**3.85**	3.46	3.82	3.24
Ours	**3.29**	3.97	**3.32**	**3.81**	3.19

TABLE III
QUANTITATIVE ABLATION RESULTS FOR (A) RFGM AND (B) CMGA ON LOLV1 DATASET. "✓/×" DENOTES THE MODULE STATUS. NOTE: WITHOUT RFGM, SEMANTICS ARE DERIVED FROM RAW IMAGES; WITHOUT CMGA, FEATURES ARE SIMPLY CONCATENATED. THE BEST RESULTS ARE IN **BOLD**.

	(a)	(b)	PSNR ↑	SSIM ↑	LPIPS ↓
(1)	×	×	23.757	0.850	0.107
(2)	✓	×	24.006	0.859	0.092
(3)	×	✓	24.001	0.855	0.096
(4)	✓	✓	**24.039**	**0.860**	**0.085**

As visually shown in Fig. 6, our method achieves smoother luminance transitions. Specifically, in the first row, our method exhibits more natural illumination and sharp edges. In the second row, our approach preserves fine cloud textures, avoiding the detail loss or overexposure common in other methods. Overall, SCI-Net exhibits strong robustness in complex real-world scenes with uneven illumination.

C. Ablation Study

Effectiveness of Key Components. We conduct four controlled experiments on LOLv1 to verify each module's contribution. As shown in Table III, compared with the baseline (1), individual integrations of RFGM (2) and CMGA (3) both yield noticeable performance gains. Specifically, (2) demonstrates that RFGM-based pseudo-enhancement helps the network capture clearer semantic features, while (3) suggests that CMGA effectively guides the image restoration with semantic priors. The best results from (4) further validate our design logic: RFGM provides a robust foundation for semantic extraction, enabling CMGA to inject reliable priors to maintain structural integrity and visual naturalness. This synergy is particularly evidenced by the improved LPIPS. Visual comparisons in Fig. 7 show that our full model (4) surpasses (1) ~ (3) in color recovery and noise suppression, achieving more natural exposure.

Reliability of Semantic Extraction via RFGM. Fig. 8 reveals the foundational role of RFGM in supporting semantic extraction under extreme low light. In the original inputs, SAM suffers from perceptual failure due to noise and low contrast. As shown in Fig. 8 (a), SAM fails to capture object contours in the dark interior. Similarly, it misses the shadowed regions beneath the table in Fig. 8 (b). Conversely, with RFGM-generated pseudo-enhanced images, SAM obtains clear and

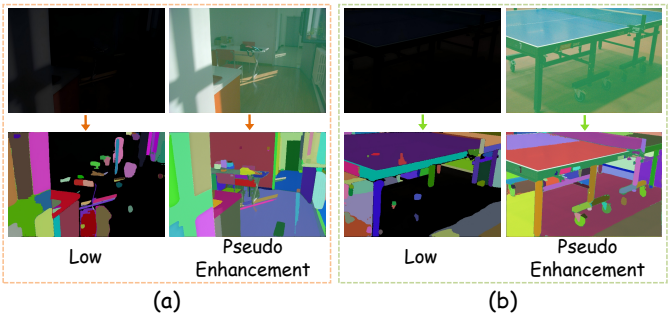

Fig. 8. Semantic extraction comparison. (a) and (b) both present the original low-light image, pseudo-enhanced image, and their corresponding semantic masks. The results indicate that using the pseudo-enhanced images generated by RFGM as input enables SAM to extract more reliable semantic priors.

distinct structural cues for accurate layout segmentation. This demonstrates that RFGM is a critical prerequisite for reliable semantic priors and high-quality enhancement.

IV. Conclusions

This paper presents SCI-Net, a novel framework leveraging multi-scale semantic priors in the HVI color space to effectively mitigate semantic artifacts and color distortions in low-light enhancement. To ensure reliable semantic extraction under degradation, the Residual Fourier Guidance Module (RFGM) leverages phase-structural constancy to generate pseudo-enhanced images for SAM. To precisely incorporate these priors, the Cross-Modal Guided Attention (CMGA) module strategically fuses hierarchical features into decoupled branches, enforcing content-aware constraints from global layout to fine details. Extensive experiments demonstrate that SCI-Net consistently outperforms state-of-the-art methods. Future research will explore efficient semantic fusion mechanisms to minimize computational overhead, facilitating broader deployment in real-time applications.

Acknowledgment

This work was supported by National Natural Science Foundation of China (62271359).

References

[1] C. Wang, J. Pan, L. Wang, and W. Wang, "Intra and inter parser-prompted transformers for effective image restoration," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 7, 2025, pp. 7609–7618.

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

[3] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE transactions on image processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

[4] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.

[5] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.

[6] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE transactions on image processing*, vol. 22, no. 9, pp. 3538–3548, 2013.

[7] V. Vonikakis, R. Kouskouridas, and A. Gasteratos, "On the evaluation of illumination compensation algorithms," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9211–9231, 2018.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[9] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[10] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.

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

[12] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 1632–1640.

[13] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1780–1789.

[14] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot, "Low-light image enhancement with normalizing flow," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 3, 2022, pp. 2604–2612.

[15] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17 714–17 724.

[16] A. Brateanu, R. Balmez, A. Avram, C. Orhei, and C. Ancuti, "Lyt-net: Lightweight yuv transformer-based network for low-light image enhancement," *IEEE Signal Processing Letters*, 2025.

[17] J. Hou, Z. Zhu, J. Hou, H. Liu, H. Zeng, and H. Yuan, "Global structure-aware diffusion process for low-light image enhancement," *Advances in Neural Information Processing Systems*, vol. 36, pp. 79 734–79 747, 2023.

[18] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K.-K. Ma, "Learning a simple low-light image enhancer from paired low-light instances," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 22 252–22 261.

[19] J. Weng, Z. Yan, Y. Tai, J. Qian, J. Yang, and J. Li, "Mamballie: Implicit retinex-aware low light enhancement with global-then-local state space," *Advances in Neural Information Processing Systems*, vol. 37, pp. 27 440–27 462, 2024.

[20] H. Jiang, A. Luo, X. Liu, S. Han, and S. Liu, "Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models," in *European Conference on Computer Vision*. Springer, 2024, pp. 161–179.

[21] J. Bai, Y. Yin, Q. He, Y. Li, and X. Zhang, "Retinexmamba: Retinex-based mamba for low-light image enhancement," in *International Conference on Neural Information Processing*. Springer, 2024, pp. 427–442.

[22] T. Zhang, P. Liu, Y. Lu, M. Cai, Z. Zhang, Z. Zhang, and Q. Zhou, "Cwnet: Causal wavelet network for low-light image enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 8789–8799.

[23] Q. Yan, Y. Feng, C. Zhang, G. Pang, K. Shi, P. Wu, W. Dong, J. Sun, and Y. Zhang, "Hvi: A new color space for low-light image enhancement," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 5678–5687.

[24] G. Seif and D. Androutsos, "Edge-based loss function for single image super-resolution," in *2018 IEEE International conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 1468–1472.

[25] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[26] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 12 504–12 513.

[27] W. Wang, H. Yang, J. Fu, and J. Liu, "Zero-reference low-light enhancement via physical quadruple priors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 26 057–26 066.