

ICLR: Inter-Chrominance and Luminance Interaction for Natural Color Restoration in Low-Light Image Enhancement

Xin Xu^{1,2}, Hao Liu¹, Wei Liu¹, Wei Wang^{1,2}, Jiayi Wu¹, Kui Jiang^{3*}

¹School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065 China

²Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System

³School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001 China
xuxin@wust.edu.cn, liuh85872@wust.edu.cn, liuwei@wust.edu.cn, wangwei8@wust.edu.cn, wuaddone@wust.edu.cn, jiangkui@hit.edu.cn

Abstract

Low-Light Image Enhancement (LLIE) task aims at improving contrast while restoring details and textures for images captured in low-light conditions. HVI color space has made significant progress in this task by enabling precise decoupling of chrominance and luminance. However, for the interaction of chrominance and luminance branches, substantial distributional differences between the two branches prevalent in natural images limit complementary feature extraction, and luminance errors are propagated to chrominance channels through the nonlinear parameter. Furthermore, for interaction between different chrominance branches, images with large homogeneous-color regions usually exhibit weak correlation between chrominance branches due to concentrated distributions. Traditional pixel-wise losses exploit strong inter-branch correlations for co-optimization, causing gradient conflicts in weakly correlated regions. Therefore, we propose an **Inter-Chrominance and Luminance inteRaction (ICLR)** framework including a Dual-stream Interaction Enhancement Module (DIEM) and a Covariance Correction Loss (CCL). The DIEM improves the extraction of complementary information from two dimensions, fusion and enhancement, respectively. The CCL utilizes luminance residual statistics to penalize chrominance errors and balances gradient conflicts by constraining chrominance branches covariance. Experimental results on multiple datasets show that the proposed ICLR framework outperforms state-of-the-art methods.

Introduction

Low-light image enhancement (LLIE) addresses critical challenges in computer vision. While recent RGB-based methods show progress, images in low-light conditions suffer from complex degradations. Simple RGB luminance adjustments often amplify noise and distort colors (Guo and Hu 2023). To mitigate this, researchers increasingly adopt color space decoupling—separating chrominance (color information) and luminance (brightness information) into distinct processing branches for independent optimization. Among decoupled spaces, the HVI color space (Yan et al. 2025) overcomes the pure-black-plane limitation of HSV space while maintaining the one-to-one RGB mapping. This

approach retains HSV’s decoupling advantages while enhancing interpretability of color transformations.

CIDNet provides a new paradigm for HVI-space optimization. It enables luminance-chrominance interaction via cross-attention mechanisms, resolving feature separation issues in conventional decoupling methods. Simultaneously, it applies consistent L1 constraints to both the HVI and RGB color spaces, mitigating the separation of optimization objectives across different color spaces. However, CIDNet overlooks inherent statistical distribution relationships that exist among HVI branches, constraining network optimization. Based on this, we identify two novel and representative observations in natural images, supported by statistical details in the **Appendix Section B**:

Observation 1: As illustrated in Figure 1(a), substantial distributional differences between luminance and chrominance cause oversmoothed cross-attention weights, limiting complementary feature extraction. Moreover, a luminance-dependent nonlinear parameter is coupled in chrominance branches for adjusting the color-dot density in dark regions, propagating luminance errors to chrominance channels.

Observation 2: As illustrated in Figure 1(b), images with large homogeneous-color regions usually exhibit weak correlation between chrominance branches due to concentrated distributions. However, traditional pixel-wise losses exploit strong inter-branch correlations for co-optimization, causing gradient conflicts (competing optimization directions) in weakly correlated regions. Consequently, conventional methods fail to optimize such areas effectively.

To address these challenges, we propose an **Inter-Chrominance and Luminance inteRaction (ICLR)** framework. ICLR leverages complementary information between chrominance (color data) and luminance (brightness data) branches to guide dual-branch optimization while calibrating joint statistical distributions of chrominance branches. Specifically, we improve the complementary information in two dimensions: enhancement and fusion. On the one hand, we design a Cross Dynamic Enhancement Module (CDEM) to leverage local and global contextual information to enhance complementary information. More importantly, we design a Multidimensional Attention-guided Fusion Module (MAFM) for chrominance branches, aiming to align chrominance and luminance branches from spatial, channel and pixel dimensions. Concurrently, we introduce a Covari-

*Corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

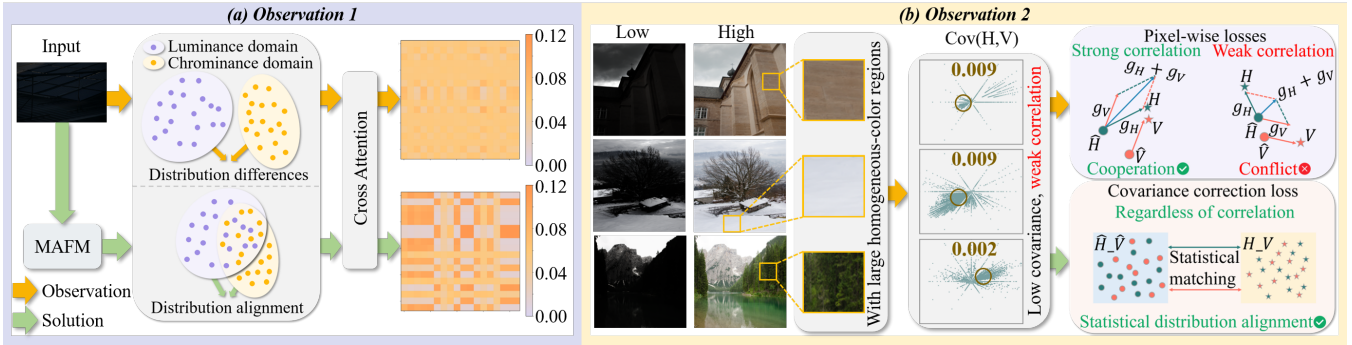


Figure 1: Our two key observations. **(a)** Substantial distributional differences between luminance and chrominance limit the complementary feature extraction. **(b)** In images with large homogeneous-color regions, chrominance branches typically exhibit weak correlation due to their relatively concentrated distributions, which leads to gradient conflicts.

ance Correction Loss (CCL) that utilizes luminance residual statistics to penalize chrominance errors, suppressing nonlinear parameter-induced diffusion errors. Meanwhile, this design balances gradient conflicts in weakly correlated distributions by constraining chrominance branches covariance. Our contributions can be summarized as follows:

- **Empirical Contribution.** We have two important observations based on the HVI: 1) Substantial distributional differences between luminance and chrominance limit complementary feature extraction, while luminance errors propagate to chrominance channels via the nonlinear parameter. 2) The chrominance branches exhibit weak correlation in images with large homogeneous-color regions, causing gradient conflicts in weakly correlated regions during optimization.
- **Framework Contribution.** We propose an Inter-Chrominance and Luminance inteRaction (ICLR) framework, which aims to leverage complementary information between chrominance and luminance branches to guide dual-branch optimization while calibrating joint statistical distributions of chrominance branches.
- **Technical Contribution.** We propose a Dual-stream Interactive Enhancement Module (DIEM) and a Covariance Correction Loss (CCL). The DIEM improves the complementary information from fusion and enhancement perspectives, by incorporating Multidimensional Attention-guided Fusion Module (MAFM) and Cross Dynamic Enhancement Module (CDEM). The CCL utilizes luminance residual statistics to penalize chrominance errors and balances gradient conflicts by constraining chrominance branches covariance.

Related Work

Low-Light Image Enhancement

Deep learning-based methods (Jiang et al. 2021b; Yang et al. 2021a; Jiang et al. 2022) have achieved remarkable success in LLIE tasks. RetinexNet (Wei et al. 2018) decomposes low-light images into illumination and reflectance components, performing both enhancement and denoising. Zero-DCE (Guo et al. 2020) formulates LLIE as the task of esti-

mating image-specific curves for a deep network, and introduces a zero-reference learning strategy for training. Additionally, SNRNet (Xu et al. 2022) employs a signal-to-noise ratio perceptual converter and convolutional modeling to dynamically enhance pixels through spatially varying operations. More recently, DRGN (Jiang et al. 2024) introduces a degradation-aware two-stage generation network that first learns intrinsic degradation to simulate the distortion of environmental illumination and then refines image content to restore details and colors. Despite these advancements, however, the aforementioned deep learning methods remain confined to the study of the RGB color space.

Color Space

RGB channels couple both chrominance and luminance, so simple luminance enhancement adjusts all channels together, misdirecting chrominance and causing color shifts. Therefore, the paradigm of decoupling chrominance and luminance by transforming the color space is introduced to LLIE tasks. For example, HSV (Chobola et al. 2024) and Ycbr color space (Brateanu et al. 2025; Guo and Hu 2023), which decouple chrominance and luminance, the conversion process with RGB color space results in multi-mapping and pure black planes, which leads to black artifacts in the image. To solve this problem, HVI color space (Yan et al. 2025) decouples luminance while introducing Horizontal/Vertical (HV) chrominance map as a plane to quantize the chrominance reflection map, which achieves one-to-one mapping to RGB color space and eliminates the pure black plane.

Methodology

Overall Architecture

As illustrated in Figure 2, our ICLR framework is a three-level U-Net architecture. Details of the framework can be found in **Appendix Section D.1**.

Dual-stream Interaction Enhancement Module

Substantial distributional differences between the two branches prevalent in natural images limit complementary

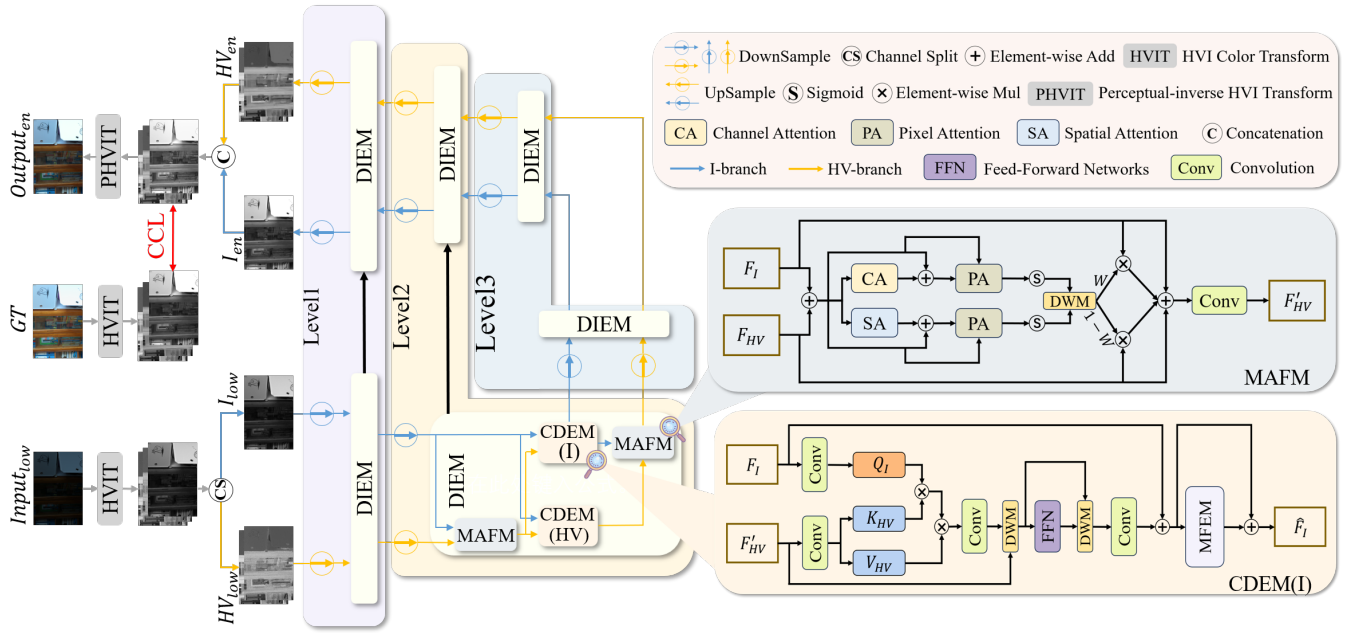


Figure 2: The overall architecture of the proposed ICLR framework. ICLR is a three-level U-Net architecture integrated with Dual-stream Interaction Enhancement Modules (DIEM) and a Covariance Correction Loss (CCL). The DIEM module mainly consists of Multidimensional Attention-guided Fusion Modules (MAFM) and Cross Dynamic Enhancement Modules (CDEM).

feature extraction. An introduction to shortcomings of traditional cross-attention can be found in **Appendix Section D.2**. Inspired by the progressive coupling mechanism (Jiang et al. 2021a), we propose the DIEM to improve the cross-attention mechanism from dimensions of enhancement and fusion (Dang et al. 2024c, 2023), respectively. The core idea of DIEM is to adaptively fuse (Chen, He, and Lu 2024) the luminance and chrominance branches to increase their distributional similarity and control the fusion scale for effective complementary feature extraction, while leveraging both local and global contextual information through dedicated enhancement modules (Shen et al. 2024) to further enhance complementary information. Details are shown in Figure 2, and here we take the enhancement of luminance branch as an example in order to clearly express the flow of the module.

Dimensions of Fusion: Multidimensional Attention-guided Fusion Module. There are substantial distributional differences between the chrominance and luminance branches. Inspired by adaptive fusion strategies in multi-model networks (Jiang et al. 2019), we hope to effectively model the cross-subspace relationship between the two branches from different fusion dimensions to achieve adaptive distribution alignment (Dang et al. 2024a,b), thereby improving the extraction of complementary information. The details of MAFM are shown in Figure 2, given the input features F_I and F_{HV} of the input luminance branch and chrominance branch, the initial feature fusion is first realized to obtain the base fused features F_{init} . Then, channel attention is used to adjust weights of each channel to emphasize complementary features that are more discriminative at the global level, so that channels that are more sensitive to com-

plementary information can dominate in subsequent fusion. Spatial attention is used in parallel to compute the mean and maximum of each spatial location, capturing the saliency of different spatial locations and ensuring that complementary information on local details is not weakened. And these two attention maps are combined with the initial fused features, allowing the attention map to reinforce the complementary information while still retaining the semantics of the original features. Then W_c and W_s are obtained by pixel attention by combining the original features and the two attention weights, respectively, such that pixel attention assigns an attention weight to each pixel location. It combines complementary information of chrominance and luminance at the pixel level, allowing the model to finely tune the attention at each pixel location, thus ensuring that the chrominance and luminance information is fully complementary at each spatial location. Finally, we fuse the attention weights W_c and W_s to obtain the final refined attention weight W by dynamic weighting to ensure that the complementary features in both channel and spatial dimensions are fully integrated. We adaptively adjust the contribution ratio of chrominance to luminance by adjusting the initial features with learned attention weights W to ensure the coordination of chrominance and luminance features and finally obtain the fused feature F'_{HV} . Figure 3 compares luminance and chrominance distributions before and after MAFM, demonstrating MAFM’s capability to achieve adaptive distribution alignment. This process can be formulated as follows:

$$F_{init} = F_I + F_{HV}, \quad (1)$$

$$W_c = \sigma(PA(CA(F_{init}) + F_{init}, F_{init})), \quad (2)$$

$$W_s = \sigma(PA(SA(F_{init}) + F_{init}, F_{init})), \quad (3)$$

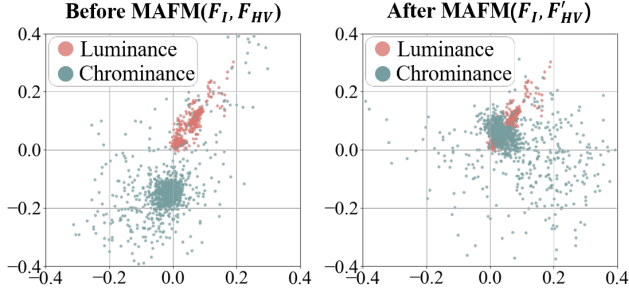


Figure 3: Comparison of luminance and chrominance branches feature distributions before and after MAFM.

$$W = \varphi \cdot W_c + \omega \cdot W_s, \quad (4)$$

$$F'_{HV} = F_{init} + W \cdot F_I + (1 - W) \cdot F_{HV}, \quad (5)$$

where σ denotes the sigmoid operation, $CA(\cdot)$, $SA(\cdot)$ and $PA(\cdot)$ denote channel attention, spatial attention and pixel attention. Details can be found in **Appendix Section D.3**. φ and ω represent learnable parameters.

Dimensions of Enhancement: Cross Dynamic Enhancement Module. To further enhance the representation of complementary information captured by the cross-attention mechanism, we design a novel dynamic weighting structure and a multi-branch convolutional structure from the enhancement point of view (Hu et al. 2023), the details of which are shown in Figure 2. Given the input luminance feature F_I and the fused chrominance feature F'_{HV} , firstly, the underlying complementary information Z_I is captured through the cross-attention mechanism, and then the dynamic weighting mechanism is utilized to regulate the priority enhancement of the complementary information. On the one hand, the structured cues captured in the luminance branch are forced to directly intervene in the reconstruction of chrominance features to uplift the complementary information in the region of strong cross-subspace correlation into the dominant restoration signals, while on the other hand, the original luminance features can be effectively retained to strengthen the expression of complementary information. Next, the nonlinear mapping of feed-forward neural network is utilized to fully explore the correlation between luminance and chrominance in high-dimensional space, and the reconstruction of the feature flow is further adaptively adjusted to obtain the feature \hat{Z}_I through the dynamic weighting mechanism. Finally, the MFEM utilizes the multi-branch convolutional structure to extract features in different directions, scales, and semantic spaces in parallel, thereby enhancing the ability to express complementary information, to obtain the final enhanced feature \hat{F}_I . This process can be formulated as follows:

$$Z_I = \text{CrossAttention}(F_I, F'_{HV}), \quad (6)$$

$$\hat{Z}_I = \lambda \cdot \text{FFN}(\alpha \cdot Z_I + \beta \cdot F'_{HV}) + \mu \cdot Z_I, \quad (7)$$

$$\hat{F}_I = \mathcal{F}_{MFEM}(F_I + \hat{Z}_I) + (F_I + \hat{Z}_I), \quad (8)$$

where $\text{CrossAttention}(\cdot)$ denotes the cross-attention mechanism, $\text{FFN}(\cdot)$ stands for feed-forward neural network in standard Transformer. $\mathcal{F}_{MFEM}(\cdot)$ represents the multi-branch feature enhancement module. α , β , λ and μ represent the learnable parameters.

Covariance Correction Loss (CCL)

Error Correction for Luminance Guidance. In the HVI, the nonlinear parameter C_k is used to adjust the color point density of the low-luminance color plane thus allowing it to be coupled in chrominance branches. This parameter can be considered as a function of the luminance $C_k(I)$, where $I \in [0, 1]$ and $\frac{\pi \hat{I}}{2} \in [0, \frac{\pi}{2}]$. So the function $C_k(I)$ is a monotonically increasing function on the range $[0, 1]$. For luminance error $\Delta I = \hat{I} - I$, when the luminance error ΔI is increasing, the error in the nonlinear parameter ΔC_k is also increased. Since the nonlinear parameters are coupled in the calculation of the chrominance branch, the corresponding chrominance channels errors ΔH and ΔV also increase. Therefore, the luminance error is diffused to the chrominance channels through the nonlinear parameter C_k , which makes the prediction of the chrominance branch heavily shifted. To address this problem, we propose to guide the correction of chrominance branches errors through luminance, and more specifically, we use the mean and variance of the luminance error as adaptive weights to adjust the optimization of the chrominance branches, and aggravate the penalty on the chrominance branches errors when the luminance error is large in order to mitigate the effect of the luminance error diffused by the nonlinear parameter C_k :

$$L_Z = \frac{1}{N} \sum_{i=1}^N (\hat{Z}_i - Z_i)^2, \quad Z \in \{I, H, V\} \quad (9)$$

$$W_H = 1.0 + \frac{1}{N} \sum_{i=1}^N |\hat{I}_i - I_i|, \quad (10)$$

$$W_V = 1.0 + \sqrt{\frac{1}{N} \sum_{i=1}^N (|\hat{I}_i - I_i| - \frac{1}{N} \sum_{i=1}^N |\hat{I}_i - I_i|)^2}, \quad (11)$$

$$L_{I-HV} = W_H L_H + W_V L_V, \quad (12)$$

where $N = B \times H \times W$ represents the total number of pixels per batch, L_H , L_V and L_I denote the independent losses of the three channels of HVI, respectively. W_H and W_V stand for the adaptive weight of the chrominance branches penalty strength. L_{I-HV} represents the luminance-guided error correction loss function.

Covariance Statistical Constraints. Conventional loss minimize the discrepancy between the predicted image and the ground truth by pixel-wise constraints, and these constraints have a tendency to be present in the HVI: stronger linear correlation between chrominance branches H and V correlates with higher prediction quality, while weaker correlation leads to degradation. This phenomenon arises because when H and V are strongly linearly correlated, their gradients update directions align, enabling simultaneous optimization of both components through a unified parameter

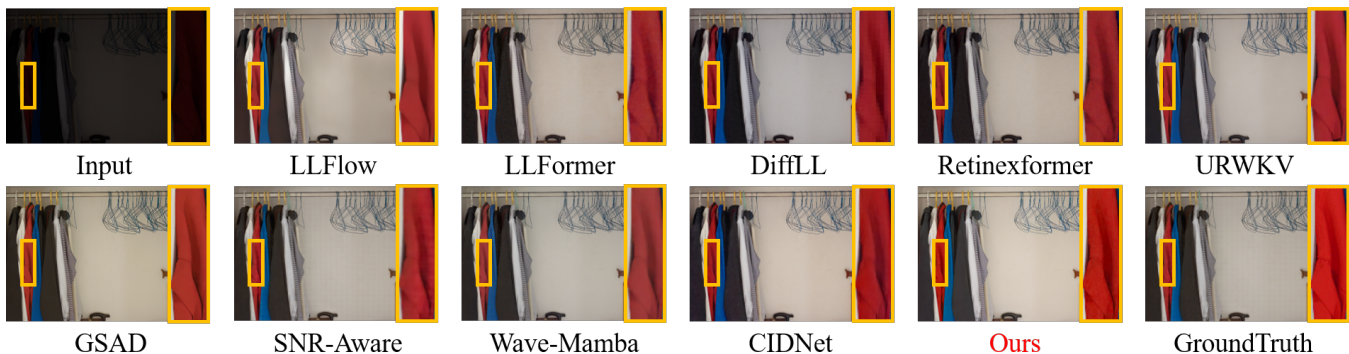


Figure 4: The qualitative comparison of the methods we compared in terms of color restoration. Zoom in for the best view.

adjustment, and when H and V are weakly linearly correlated, their gradients require updates in divergent or even opposing directions, leading to conflicts in parameter optimization and cancellation effects that undermine model convergence. To mitigate gradient competition caused by weak linear correlation between chrominance branches H and V , we shifts the focus from a deterministic pixel-wise comparison to a statistical perspective, emphasizing the learning of distributions rather than individual pixel values (Dang et al. 2025a,b). The core idea is to optimize the overall pixel distribution by means of a loss function corresponding to the computation of the fitted covariance, which leads to the optimization of the overall pixel distribution and gets rid of the reliance on the pixel-wise linear correlation between H and V . More specifically, the formula for calculating the covariance between H and V can be expressed as:

$$Cov(H, V) = E(HV) - E(H)E(V), \quad (13)$$

where $Cov(\cdot)$ denotes the covariance function, $E(\cdot)$ denotes the mean value function.

The independent channel losses for H and V in Eq.9, L_H and L_V have already constrained the individual means $E(H)$ and $E(V)$. To further enforce covariance-based constraints on the global statistical distribution of H and V pixels, we must additionally constrain the joint mean term $E(HV)$, and this constraint can be formulated as follows:

$$L_{HV} = \frac{1}{B} \sum_{i=1}^B \left(\frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \hat{H}_i \hat{V}_j - \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W H_i V_j \right)^2, \quad (14)$$

where B , H and W denote batch size, height and width, respectively, L_{HV} represents the joint mean-constrained loss.

Finally, we simple combine L_I , L_{I-HV} and L_{HV} these three losses as the final objective CCL.

Experiments

Datasets and metrics

We train and evaluate our model on the LLIE benchmark dataset LOL, which consists of two versions: LOLv1 (Wei et al. 2018) and LOLv2 (Yang et al. 2021b). LOLv2 is divided into two subsets: LOLv2-real and LOLv2-synthetic. The training and testing splits are 485:15 for LOLv1,

689:100 for LOLv2-real and 900:100 for LOLv2-synthetic. In addition, we test on five unpaired datasets, DICM (Lee, Lee, and Kim 2013), LIME (Guo, Li, and Ling 2016), MEF (Ma, Zeng, and Wang 2015), NPE (Wang et al. 2013) and VV (Vonikakis, Kouskouridas, and Gasteratos 2018), to validate generalization performance. For paired datasets, we adopt Peak Signal-to-Noise Ratio (PSNR) (Wang et al. 2004), Structure Similarity Measure (SSIM) (Wang et al. 2004), and Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al. 2018) as evaluation metrics. For unpaired datasets, we adopt Natural Image Quality Evaluator (NIQE) (Mittal, Soundararajan, and Bovik 2012) as evaluation metric. Among them, PSNR and SSIM are better when higher, while LPIPS and NIQE are better when lower. Experiments with more datasets can be found in **Appendix Section E.1**, which further validate the performance of our method on large-scale datasets.

Experimental settings

We set the patch size to 256×256 , the batch size to 8, and train the model with the Adam optimizer for a total of 1000 epochs on a single NVIDIA 3090 GPU by using PyTorch. The learning rate is initially set to 1×10^{-4} and then steadily decreased to 1×10^{-7} by the cosine annealing scheme during the training process. We train on the LOLv1 dataset and test on the LOLv1, LOLv2-real and LOLv2-synthetic datasets to demonstrate the generalization performance of our method.

Comparisons with State-of-the-Art Methods

We compare our method with state-of-the-art (SOTA) methods for LLIE, including LLFlow (Wang et al. 2022), SNR-Aware (Xu et al. 2022), Retinexformer (Cai et al. 2023), LLFormer (Wang et al. 2023), GSAD (Hou et al. 2023), DiffLL (Jiang et al. 2023), Wave-Mamba (Zou et al. 2024), URWKV (Xu et al. 2025), and CIDNet (Yan et al. 2025).

Comparison on LOLv1 and LOLv2 Datasets. Table 1, Figure 4 and Figure 7 list quantitative and qualitative comparisons with SOTA methods on LOLv1 and LOLv2. For quantitative comparison, our method gains 0.87 dB, 0.88 dB and 0.20 dB on LOLv1, LOLv2-real and LOLv2-syn datasets, respectively. This indicates that our method has better generalization performance, mainly because we en-

Methods	References	LOLv1			LOLv2-Real			LOLv2-Syn			Params(M)
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	
LLFlow	AAAI'22	24.046	0.860	0.136	26.428	0.903	0.096	19.219	0.824	0.214	37.68
SNR-Aware	CVPR'22	24.609	0.842	0.105	30.923	0.893	0.092	16.251	0.767	0.192	39.13
Retinexformer	ICCV'23	25.155	0.845	0.085	28.983	0.882	0.068	16.190	0.771	0.189	1.53
LLFormer	AAAI'23	23.649	0.816	0.169	27.749	0.860	0.143	17.163	0.784	0.244	24.55
GSAD	NeurIPS'23	27.595	0.875	0.091	28.818	0.895	0.095	19.784	0.806	0.225	17.36
DiffLL	TOG'23	24.634	0.798	0.201	27.781	0.845	0.180	19.269	0.766	0.279	22.07
Wave-Mamba	MM'24	25.847	0.858	0.143	30.714	0.905	0.103	19.700	0.819	0.233	1.51
URWKV	CVPR'25	26.513	0.869	0.107	31.413	0.906	0.073	20.518	0.810	0.214	2.25
CIDNet	CVPR'25	27.732	0.870	0.117	31.436	0.896	0.101	20.375	0.816	0.243	1.88
Ours	-	28.603	0.885	0.068	32.320	0.907	0.054	20.719	0.837	0.217	4.27

Table 1: Quantitative results on LOLv1, LOLv2-Real and LOLv2-Syn. The best performance is marked in bold. To compare the generalization performance of methods, We train on the LOLv1 and test on LOLv1, LOLv2-Real and LOLv2-Syn.

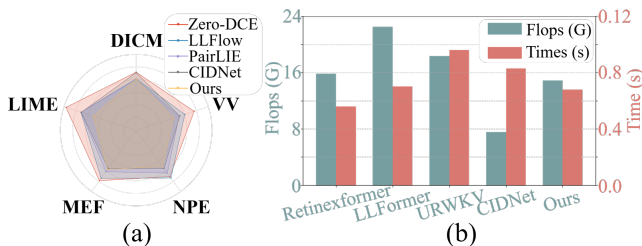


Figure 5: (a) Comparison of NIQE \downarrow on five unpaired datasets, where lower values indicate better performance. (b) Comparison of Flops and Inference Time.

hanced the expression of complementary information between the chrominance and luminance branches through the dual-stream interaction enhancement module, enabling the capture of multi-scale features more effectively for better generalization and adaptability across datasets. For qualitative comparison, our method is able to restore more faithful colors than other methods for the purple and orange sweaters in the first row and the red sign in the second row. Moreover, the red clothing in Figure 4 demonstrates that our method can restore a more vibrant red color compared to other methods. The advantage of our method in color restoration is attributed to the CCL function specifically designed for chrominance branches. More visual comparison experiments can be found in **Appendix Section E.3**.

Comparison on unpaired datasets. Figure 5(a) shows a quantitative comparison of the NIQE metric on five unpaired datasets. It can be observed that our method achieves the lowest values on all four datasets except for a slightly higher value than CIDNet on the MEF dataset, demonstrating the superior generalization capability of our method.

Comparison on computational overhead. Table 1 and Figure 5(b) show comparisons of parameters, flops and inference time. It can be observed that although our method has higher parameters than URWKV, we achieve advantages in both flops and inference time. This indicates that our method can achieve a noticeable improvement in restoration quality at the cost of small computational overhead.

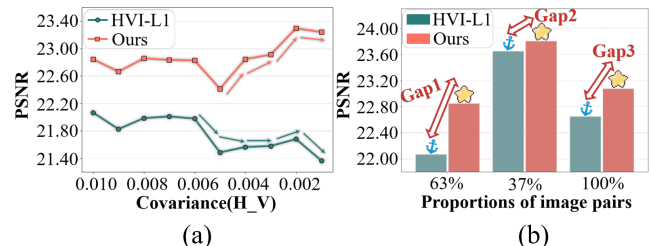


Figure 6: Evaluation and comparison of our method and conventional pixel-wise loss under different covariance conditions, using PSNR to evaluate performance.

Comparison on a New Synthetic Dataset. To further demonstrate the superiority of our method over pixel-wise losses approaches in resolving gradient conflicts, we merged the test sets from LOLv1, LOLv2-real and LOLv2-synthetic into a new dataset containing a total of 215 low-light image pairs. The new synthetic dataset consists of 215 images, of which **136 images (63%)** have a covariance coefficient less than or equal to 0.01, while **79 images (37%)** have a covariance coefficient greater than 0.01. As shown in Figure 6(a), for image pairs with covariances less than or equal to 0.01, we divided the covariance range into more detailed categories and evaluated the PSNR. It can be clearly observed that the pixel-wise loss shows a significant downward trend in PSNR as the covariance decreases. In contrast, our method not only outperforms the pixel-wise loss under all covariance conditions. Moreover, this advantage becomes even more pronounced when the covariance is relatively low. To illustrate the specific mechanism of our method for low-covariance images, we evaluated images with proportions of 63% ($Cov \leq 0.01$), 37% ($Cov > 0.01$) and 100% (All), as shown in Figure 6(b). Through quantitative analysis, 90% of the overall performance improvement (Gap3) comes from improvements in the portion with a covariance less than or equal to 0.01 (Gap1), while the remaining 10% comes from improvements in the portion greater than 0.01 (Gap2). This result shows that our method can effectively solve the gradient conflict problem caused by pixel-wise loss under low-covariance conditions.

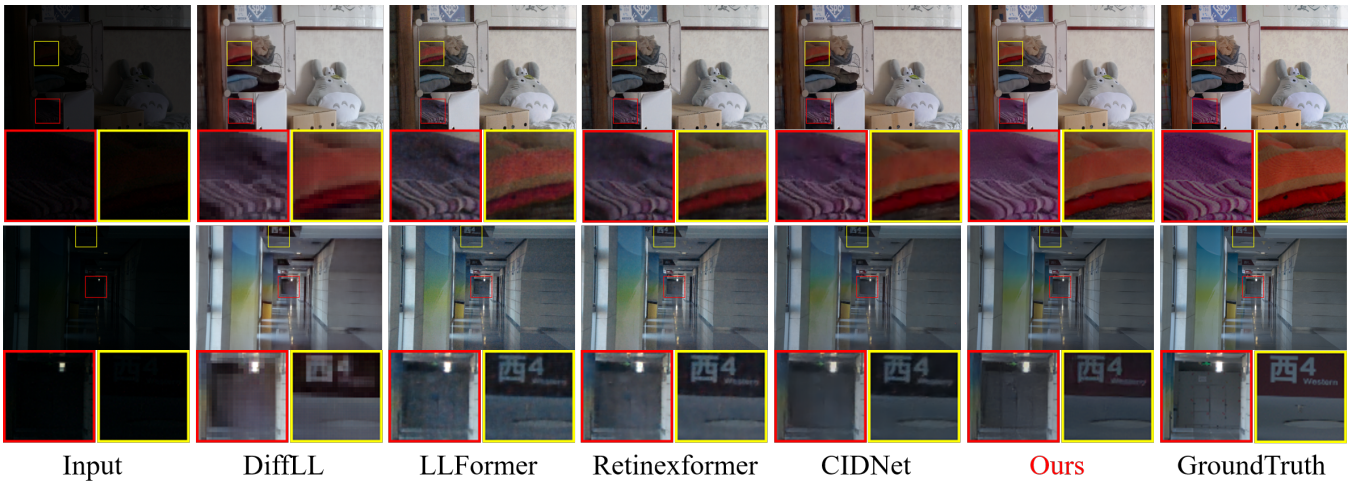


Figure 7: Visual comparisons of the enhanced results by different methods on LOLv1 and LOLv2. Zoom in for the best view.

Sets	Components			Metrics		
	MAFM	CDEM	CCL	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Ω_1	✓	✓	L1	28.202	0.879	0.073
Ω_2	✓	✓	L2	28.275	0.880	0.073
Ω_3	✓	TCA	✓	27.998	0.883	0.074
Ω_4	✗	✓	✓	28.028	0.880	0.080
Ω_5	✓	✓	✓	28.603	0.885	0.068

Table 2: Ablation Studies on MAFM, CDEM and CCL.

Ablation Study

In this section, we conduct ablation studies on the main components of our ICLR framework using the LOLv1 dataset.

Effectiveness of MAFM. As shown in Table 2, settings for Ω_4 and Ω_5 indicate that the PSNR without MAFM decreased by 0.575 dB. Combined with the distribution comparison before and after MAFM in Figure 3, this indicates that MAFM aligns chrominance and luminance branches to improve complementary feature extraction.

Effectiveness of CDEM. We compare the advantages of CDEM over Traditional Cross Attention (TCA) in two experimental settings, Ω_3 and Ω_5 . The PSNR decreased by 0.605 dB when CDEM is replaced by TCA, demonstrating that CDEM more effectively leverages local and global contextual information to enhance complementary features.

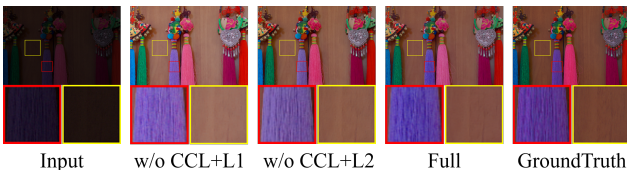


Figure 8: Visual comparison of CCL with conventional pixel-wise losses. Please zoom in for the best view.

Effectiveness of CCL. To compare CCL with conventional pixel-wise losses for color restoration, We replace

CCL with L1 and L2 in settings Ω_1 and Ω_2 . Compared with setting Ω_5 , the PSNR decreased by 0.401 dB and 0.328 dB. Combined with the visual comparison in Figure 8, this demonstrates that CCL can more effectively optimize the chrominance branches through covariance constraints, leading to a more faithful restoration of color information.

Effectiveness of the number of MAFMs. As shown in Table 3. We further conduct ablation studies on the two MAFMs before and after the DIEM. Compared to having both MAFM(1) (before) and MAFM(2) (after), having only MAFM(1) or only MAFM(2) resulted in PSNR decreases of 0.231 dB and 0.268 dB, respectively. This indicates that both MAFMs play an important role in aligning the chrominance and luminance branches.

Sets	Components		Metrics		
	MAFM(1)	MAFM(2)	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Ω_1	✗	✗	28.028	0.880	0.080
Ω_2	✗	✓	28.372	0.883	0.073
Ω_3	✓	✗	28.335	0.882	0.072
Ω_4	✓	✓	28.603	0.885	0.068

Table 3: Ablation Studies on the number of MAFMs.

Conclusion

In this paper, in order to solve the problems of interaction between chrominance and luminance branches, as well as within chrominance branches for natural images, we propose an Inter-Chrominance and Luminance inteRaction (ICLR) framework. Specifically, we improve the complementary feature extraction in the interaction of luminance and chrominance branches through a Multidimensional Attention-guided Fusion Module (MAFM) and a Cross Dynamic Enhancement Module (CDEM) in terms of the dimensions of fusion and enhancement, respectively. In addition, we design a Covariance Correction Loss (CCL) to optimize the interaction between chrominance branches from a statistical distribution perspective.

Acknowledgements

This research was financially supported by the National Natural Science Foundation of China (62501189, 62376201), Hubei Provincial Science & Technology Talent Enterprise Services Program (2025DJB059), Hubei Provincial Special Fund for Central Guided Local S&T Development (2025CSA017), and the Natural Science Foundation of Heilongjiang Province of China for Excellent Youth Project (YQ2024F006).

References

- Brateanu, A.; Balmez, R.; Avram, A.; Orhei, C.; and Ancuti, C. 2025. Lyt-net: Lightweight yuv transformer-based network for low-light image enhancement. *IEEE Signal Processing Letters*.
- Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; and Zhang, Y. 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.
- Chen, Z.; He, Z.; and Lu, Z.-M. 2024. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE transactions on image processing*, 33: 1002–1015.
- Chobola, T.; Liu, Y.; Zhang, H.; Schnabel, J. A.; and Peng, T. 2024. Fast context-based low-light image enhancement via neural implicit representations. In *European Conference on Computer Vision*, 413–430. Springer.
- Dang, J.; Chen, L.; Wu, J.; Lin, R.; Wang, B.; Wang, Y.; Wang, L.; Zhu, N.; and Wang, T. 2025a. Diff-LMM: Diffusion Teacher-Guided Spatio-Temporal Perception for Video Large Multimodal Models. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, 873–881.
- Dang, J.; Deng, S.; Chang, H.; Wang, T.; Wang, B.; Wang, S.; Zhu, N.; Niu, G.; Zhao, J.; and Liu, J. 2025b. Hallucination Reduction in Video-Language Models via Hierarchical Multimodal Consistency. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, 9167–9175.
- Dang, J.; Zheng, H.; Lai, J.; Yan, X.; and Guo, Y. 2023. Efficient and robust video object segmentation through isogenous memory sampling and frame relation mining. *IEEE Transactions on Image Processing*, 32: 3924–3938.
- Dang, J.; Zheng, H.; Wang, B.; Wang, L.; and Guo, Y. 2024a. Temporo-spatial parallel sparse memory networks for efficient video object segmentation. *IEEE Transactions on Intelligent Transportation Systems*.
- Dang, J.; Zheng, H.; Xu, X.; Wang, L.; and Guo, Y. 2024b. Beyond appearance: Multi-frame spatio-temporal context memory networks for efficient and robust video object segmentation. *IEEE Transactions on Image Processing*.
- Dang, J.; Zheng, H.; Xu, X.; Wang, L.; Hu, Q.; and Guo, Y. 2024c. Adaptive sparse memory networks for efficient and robust video object segmentation. *IEEE Transactions on Neural Networks and Learning Systems*.
- Guo, C.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1780–1789.
- Guo, X.; and Hu, Q. 2023. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, 131(1): 48–66.
- Guo, X.; Li, Y.; and Ling, H. 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2): 982–993.
- Hou, J.; Zhu, Z.; Hou, J.; Liu, H.; Zeng, H.; and Yuan, H. 2023. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, 36: 79734–79747.
- Hu, M.; Jiang, K.; Wang, Z.; Bai, X.; and Hu, R. 2023. CycMuNet+: Cycle-projected mutual learning for spatial-temporal video super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11): 13376–13392.
- Jiang, H.; Luo, A.; Fan, H.; Han, S.; and Liu, S. 2023. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (TOG)*, 42(6): 1–14.
- Jiang, K.; Wang, Q.; An, Z.; Wang, Z.; Zhang, C.; and Lin, C.-W. 2024. Mutual retinex: Combining transformer and cnn for image enhancement. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(3): 2240–2252.
- Jiang, K.; Wang, Z.; Wang, Z.; Chen, C.; Yi, P.; Lu, T.; and Lin, C.-W. 2022. Degrade is upgrade: Learning degradation for low-light image enhancement. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 1078–1086.
- Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Wang, Z.; Wang, X.; Jiang, J.; and Lin, C.-W. 2021a. Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining. *IEEE Transactions on Image Processing*, 30: 7404–7418.
- Jiang, K.; Wang, Z.; Yi, P.; Wang, G.; Gu, K.; and Jiang, J. 2019. ATMFN: Adaptive-threshold-based multi-model fusion network for compressed face hallucination. *IEEE Transactions on Multimedia*, 22(10): 2734–2747.
- Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; and Enlightengan, Z. W. 2021b. Deep light enhancement without paired supervision., 2021, 30. DOI: <https://doi.org/10.1109/TIP.2340-2349>.
- Lee, C.; Lee, C.; and Kim, C.-S. 2013. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE transactions on image processing*, 22(12): 5372–5384.
- Ma, K.; Zeng, K.; and Wang, Z. 2015. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11): 3345–3356.
- Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3): 209–212.

- Shen, J.; Chen, Y.; Liu, Y.; Zuo, X.; Fan, H.; and Yang, W. 2024. ICAFusion: Iterative cross-attention guided feature fusion for multispectral object detection. *Pattern Recognition*, 145: 109913.
- Vonikakis, V.; Kouskouridas, R.; and Gasteratos, A. 2018. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77(8): 9211–9231.
- Wang, S.; Zheng, J.; Hu, H.-M.; and Li, B. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9): 3538–3548.
- Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; and Lu, T. 2023. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 2654–2662.
- Wang, Y.; Wan, R.; Yang, W.; Li, H.; Chau, L.-P.; and Kot, A. 2022. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 2604–2612.
- Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.
- Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*.
- Xu, R.; Niu, Y.; Li, Y.; Xu, H.; Liu, W.; and Chen, Y. 2025. URWKV: Unified RWKV Model with Multi-state Perspective for Low-light Image Restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 21267–21276.
- Xu, X.; Wang, R.; Fu, C.-W.; and Jia, J. 2022. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17714–17724.
- Yan, Q.; Feng, Y.; Zhang, C.; Pang, G.; Shi, K.; Wu, P.; Dong, W.; Sun, J.; and Zhang, Y. 2025. Hvi: A new color space for low-light image enhancement. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 5678–5687.
- Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; and Liu, J. 2021a. Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. *IEEE Transactions on Image Processing*, 30: 3461–3473.
- Yang, W.; Wang, W.; Huang, H.; Wang, S.; and Liu, J. 2021b. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30: 2072–2086.
- Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.
- Zou, W.; Gao, H.; Yang, W.; and Liu, T. 2024. Wave-Mamba: Wavelet State Space Model for Ultra-High-Definition Low-Light Image Enhancement. In *ACM Multimedia 2024*.