SHARED BUT NOT SUBSTITUTABLE: FREQUENCY-AWARE COLOR–INTENSITY COLLABORATION FOR LOW-LIGHT IMAGE ENHANCEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

We propose SPARNet, a frequency-aware framework for low-light image enhancement that explicitly models shared and component-specific information between chromatic and intensity representations in the HVI color space. Motivated by the observation that low-frequency HV and I components exhibit increasing similarity at coarser scales while remaining weakly predictable from one another under linear mappings, SPARNet learns shared and private representations to exploit cross-component dependencies without compromising their complementary characteristics. Specifically, a three-level wavelet pyramid captures multi-scale low-frequency context, while shared–private latent fusion enables adaptive information exchange and preserves component-specific representations through explicit regularization. The fused low-frequency context further guides directional high-frequency subband correction, facilitating illumination restoration, detail recovery, and noise suppression. Experiments on paired and unpaired low-light benchmarks show that SPARNet achieves the best results in eight of nine metric–dataset combinations across the LOL benchmarks and the lowest average NIQE on five unpaired datasets. Ablation studies and representation analyses further demonstrate the effectiveness of shared–private frequency collaboration.

1 INTRODUCTION

Low-light images are degraded by signal attenuation, non-uniform exposure, sensor noise, and color shifts. Low-light image enhancement (LLIE) is therefore an ill-posed problem in which exposure correction, noise suppression, detail reconstruction, and color preservation are mutually constrained (Li et al., 2022). Traditional histogram equalization and Retinex methods rely on fixed mappings or idealized decompositions and often produce over-enhancement, halos, or amplified noise. Deep convolutional networks, Transformers, diffusion models, and signal-to-noise-ratio-aware approaches have improved restoration (Guo et al., 2020; Wang et al., 2023b; Yi et al., 2023; Xu et al., 2022). Nevertheless, most operate in sRGB, where coupled illumination and chromaticity make it difficult to improve brightness without distorting colors.

The HVI color space maps the periodic hue component of HSV onto a continuous HV plane and employs learnable intensity compression to suppress unstable responses in near-black regions, providing a representation for decoupled modeling of illumination and chromaticity (Yan et al., 2025). However, branch decoupling does not imply branch independence. After wavelet decomposition, the low-frequency HV and I components originate from the same scene and contain corresponding large-scale layouts and slowly varying degradation patterns. Meanwhile, they characterize chromatic distributions and intensity variations, respectively, and should not be assumed to be mutually substitutable. Completely isolating the branches discards useful cross-branch cues, whereas direct concatenation or unconstrained alignment may introduce redundancy and homogenize branch-specific information.

To examine this hypothesis, we conduct a controlled three-level HVI–wavelet analysis on 689 paired low-light and ground-truth images from the LOLv2-Real training set (Yang et al., 2020). As shown in Fig. 1, the paired linear centered kernel alignment (CKA) (Kornblith et al., 2019) between low-light HV and I components increases from 0.1550 at LL1 to 0.2633 at LL3. The conservative

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

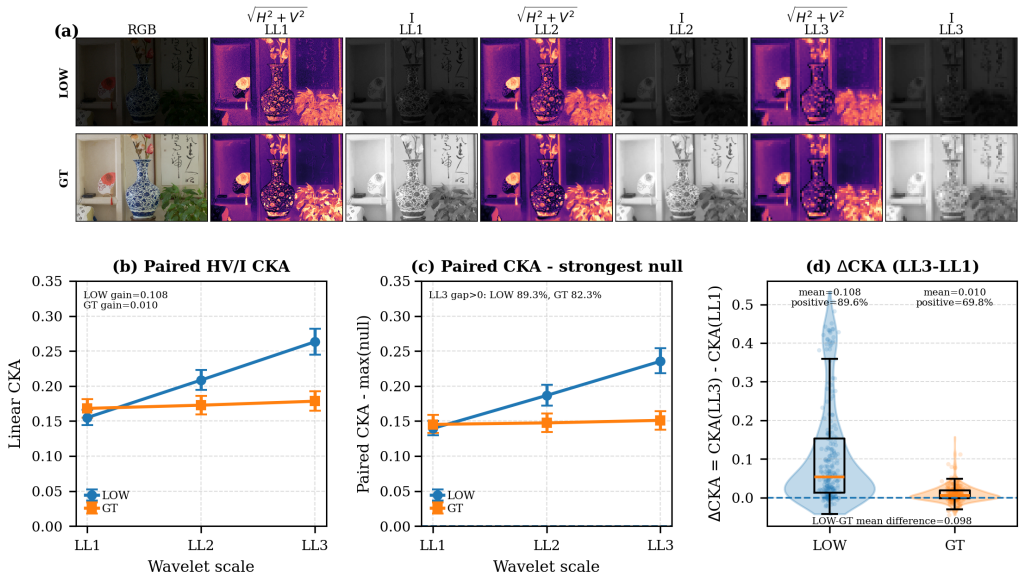

Figure 1: Multi-scale relationships between the HV and I components in HVI space. Paired CKA increases at coarser low-frequency scales and remains higher than spatial-shuffling and cross-image-mismatching controls, whereas fixed cross-image linear prediction remains limited.

margin over the stronger of two null controls increases from 0.1395 to 0.2354. The cross-scale CKA increment is 0.1083 for low-light images but only 0.0103 for ground-truth images. In contrast, five-fold ridge regression grouped by image yields a pooled out-of-fold R^2 no higher than 0.0598 across scales and prediction directions. Thus, the results support a bounded conclusion: exploitable spatial dependencies exist between low-frequency HV and I components, but a fixed linear mapping explains only a small fraction of the variation in one component from the other. We describe this relationship as *shared but not mutually substitutable*, rather than fully independent or interchangeable.

Motivated by this observation, we propose SPARNet (Shared–Private Adaptive Restoration Network). It builds a three-level wavelet pyramid within the HV and I branches to aggregate low-frequency illumination and scene structures. A Shared–Private Cross-Space Low-Frequency Latent Fusion module (CSLF) adaptively learns cross-branch shared factors and branch-specific factors through bidirectional interaction and complementary gating. Shared alignment, intra-branch decorrelation, and private-purity constraints preserve an explicit division of representation roles. Finally, fused low-frequency context and an uncertainty prior guide direction-specific correction of the HL, LH, and HH subbands.

Our contributions are threefold: (i) a controlled HVI–wavelet analysis using low-light/ground-truth comparisons, two null controls, and cross-image prediction, which identifies low-frequency dependencies alongside limited linear substitutability; (ii) a shared–private cross-space low-frequency fusion mechanism that exploits cross-branch commonality while preserving branch-specific features along a three-level LL pathway; and (iii) low-frequency-guided directional subband correction, whose complementary contribution with CSLF is examined on paired and unpaired LLIE benchmarks.

2 Related Work

Low-light image enhancement. Early LLIE methods relied on histogram adjustment, gamma correction, and Retinex decomposition. Deep methods subsequently introduced supervised CNNs, zero-reference curve estimation (Guo et al., 2020), Transformers (Wang et al., 2023b; Cai et al., 2023), diffusion models (Yi et al., 2023; Hou et al., 2023), and SNR-aware routing (Xu et al., 2022) to improve nonlinear restoration and long-range modeling. However, jointly controlling chromatic coupling, noise amplification, and unrealistic details in extremely dark regions remains challenging.

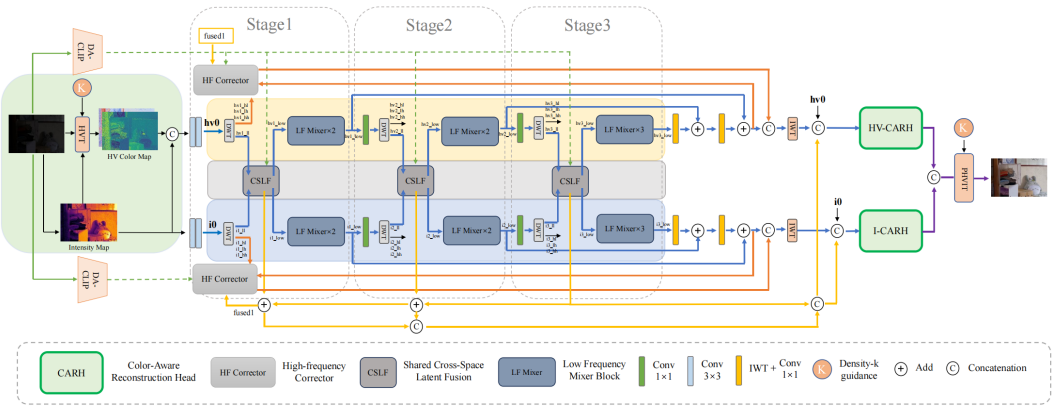

Figure 2: Overall architecture of SPARNet. A three-level LL pathway performs multi-scale low-frequency modeling; CSLF learns shared and branch-specific factors; the fused low-frequency representation guides directional subband correction.

HVI-based enhancement. HVI alleviates the red discontinuity and black-plane noise inherent in HSV through continuous HV coordinates and learnable intensity collapse. CIDNet established a dual-branch paradigm for separately modeling HV chromatic and I intensity components (Yan et al., 2025). Subsequent work extended HVI enhancement through two-stage Transformers, Mamba architectures, physical SNR priors, and differentiated branch processing (Li et al., 2026; Xu & Hao, 2026; Wang et al., 2026; He et al., 2025). DLFE-Net, HVI-FourNet, and HVI-FFNet further illustrate frequency-domain approaches to illumination correction and detail restoration (He et al., 2025; Zhang et al., 2025; Tang et al., 2026). These approaches do not explicitly distinguish shared from branch-specific factors within the low-frequency HV/I representations in the manner investigated here.

Frequency-domain and shared–private learning. Fourier-based approaches model global frequency characteristics (Wang et al., 2023a; Huang et al., 2022); wavelet transforms also retain scale, orientation, and spatial localization (Zou et al., 2024; Jiang et al., 2023; Mallat, 1989). Shared–private representation learning separates cross-domain consistency from domain-specific differences (Bousmalis et al., 2016; Hazarika et al., 2020). We apply this principle to the low-frequency pathways of an HVI wavelet pyramid. Individual latent channels are not assigned predetermined physical semantics; instead, their functional roles are learned using architecture and statistical regularization.

3 Method

3.1 Overview

Given a low-light image $x \in \mathbb{R}^{3\times H\times W}$, the HVI transform (Yan et al., 2025) decomposes it into chromatic and intensity components:

$$[x^{\mathrm{hv}}, x^{\mathrm{i}}] = \mathcal{T}_{\mathrm{hvi}}(x). \tag{1}$$

As shown in Fig. 2, the two components are projected into shallow features and processed by a three-level wavelet pyramid. At each scale, the LL subbands are modeled by CSLF. Deep low-frequency features propagate to shallower scales through top-down aggregation; the first-level directional subbands are corrected using fused low-frequency context. Two reconstruction heads predict HVI residuals; the final enhanced image is

$$\hat{x} = \mathcal{T}_{\mathrm{hvi}}^{-1}\left([x^{\mathrm{hv}}, x^{\mathrm{i}}] + [\Delta x^{\mathrm{hv}}, \Delta x^{\mathrm{i}}]\right). \tag{2}$$

A frozen DA-CLIP encoder (Luo et al., 2024) extracts image-content and degradation-condition representations, used only as auxiliary priors at different stages.

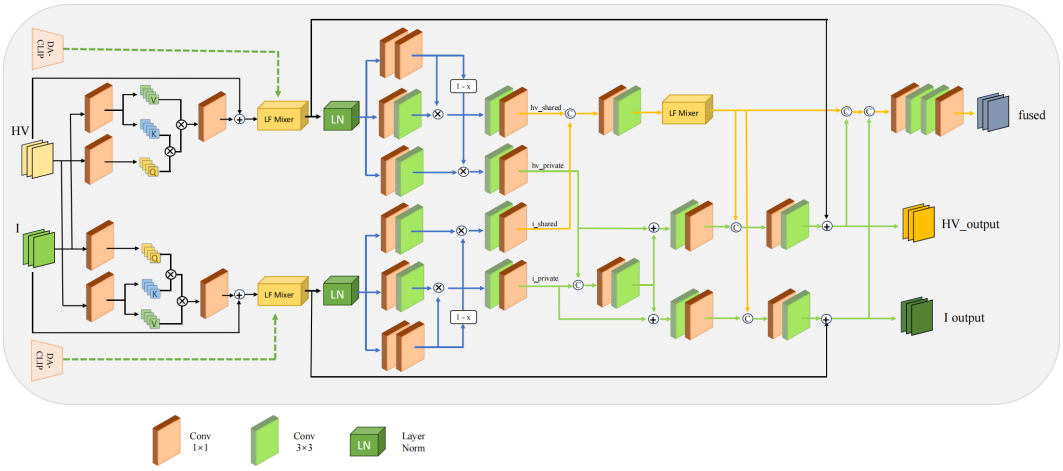

Figure 3: Shared–Private Cross-Space Low-Frequency Latent Fusion (CSLF). The module exchanges HV/I low-frequency context, separates shared and private features, and reinjects a fused shared latent alongside branch-specific representations.

3.2 Cross-Branch Wavelet Pyramid

Let $\mathcal{R} = \{\mathrm{hv}, \mathrm{i}\}$ denote the two HVI branches, let $\bar{r}$ denote the complementary branch to $r \in \mathcal{R}$, and let $F^r_{\ell-1}$ be the input features at stage ℓ. Haar discrete wavelet transform (Mallat, 1989) decomposes the features into a low-frequency subband and three directional high-frequency subbands:

$$\left(L^r_\ell, H^{r,\mathrm{hl}}_\ell, H^{r,\mathrm{lh}}_\ell, H^{r,\mathrm{hh}}_\ell\right) = \mathrm{DWT}\left(F^r_{\ell-1}\right), \quad \ell = 1, 2, 3. \tag{3}$$

Only enhanced LL features are propagated to the next stage, successively reducing spatial resolution to $H/2$, $H/4$, and $H/8$. This decomposition increases receptive field while concentrating slowly varying illumination and large-scale structure into low-frequency representations. Deeper high-frequency subbands are not directly used in output reconstruction: synthesizing fine details at highly reduced resolutions may introduce structures unsupported by the input. Instead, the first-level directional subbands, closest to the output resolution, are preserved and corrected before inverse wavelet reconstruction.

3.3 Shared–Private Cross-Space Low-Frequency Fusion

As illustrated in Fig. 3, CSLF performs bidirectional cross-branch interaction at each stage:

$$C^r_\ell = L^r_\ell + \mathcal{A}^r_\ell\left(\mathrm{LN}(L^r_\ell), \mathrm{LN}(L^{\bar{r}}_\ell)\right), \quad r \in \mathcal{R}, \tag{4}$$

where $\mathcal{A}^r_\ell$ denotes transformer-based cross-branch interaction. A low-frequency mixer $\mathcal{M}^r_\ell$ gives $\bar{C}^r_\ell = \mathcal{M}^r_\ell(C^r_\ell)$. Complementary gating then assigns mixed features to shared and private pathways:

$$g^r_\ell = \sigma\left(\mathcal{G}^r_\ell(\bar{C}^r_\ell)\right), \tag{5}$$

$$S^r_\ell = \mathcal{P}^r_{S,\ell}(\bar{C}^r_\ell) \odot g^r_\ell, \qquad P^r_\ell = \mathcal{P}^r_{P,\ell}(\bar{C}^r_\ell) \odot (1 - g^r_\ell). \tag{6}$$

Here $\mathcal{G}^r_\ell$ predicts the gate, $\mathcal{P}^r_{S,\ell}$ and $\mathcal{P}^r_{P,\ell}$ are shared/private projections, and $\odot$ is elementwise multiplication.

The shared HV/I features are aggregated into a compact token-based latent Z_ℓ. Conditioned on image-content and degradation prior T_ℓ, this latent is reinjected together with the corresponding private representation:

$$\widetilde{L}^r_\ell = \bar{C}^r_\ell + \mathcal{R}^r_{\mathrm{inj},\ell}\left(Z_\ell + \mathcal{A}^r_{T,\ell}(Z_\ell, T_\ell), P^r_\ell\right). \tag{7}$$

The reinjection lets the branches exchange shared context without replacing their private representations.

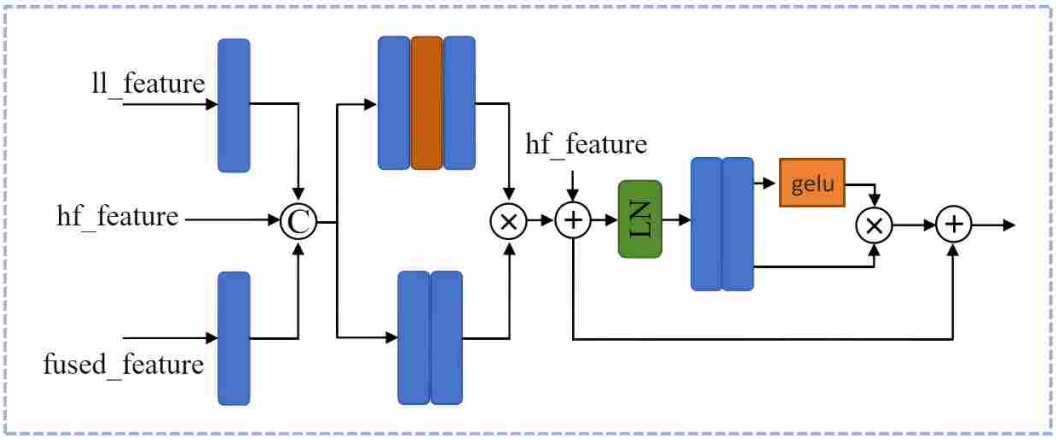

Figure 4: Low-frequency-guided directional subband correction. Branch-specific LL features and fused low-frequency context jointly guide gated correction of the HL, LH, and HH subbands.

3.4 Low-Frequency-Guided Subband Correction

High-frequency subbands contain both genuine edges and textures and sensor noise. Uniform enhancement may therefore conflict with noise suppression. As in Fig. 4, we independently learn a gated residual correction for every first-level directional subband $b \in \{\mathrm{hl}, \mathrm{lh}, \mathrm{hh}\}$:

$$\widehat{H}_1^{r,b} = H_1^{r,b} + G_1^{r,b} \odot \Phi_1^{r,b}\left(H_1^{r,b}, \widetilde{L}_1^{r}, F_1^{\mathrm{fuse}}\right). \tag{8}$$

$G_1^{r,b}$ is an adaptive gate, $\Phi_1^{r,b}$ is a direction-specific correction operator, $\widetilde{L}_1^r$ provides branch-specific low-frequency features, and F_1^{fuse} provides fused cross-branch context. The fused context supplies exposure and structural information, while the branch-specific feature preserves chromatic or intensity constraints. Corrected HL, LH, and HH subbands are combined with the enhanced LL features by inverse wavelet transform.

3.5 Training Objectives

The model is optimized with reconstruction losses in sRGB and HVI space, $\mathcal{L}_{\mathrm{rgb}}$ and $\mathcal{L}_{\mathrm{hvi}}$. To define the three shared–private regularizers, let $\mathcal{N}(X)$ denote the channel-wise centered, Frobenius-normalized matrix formed by flattening the spatial dimensions of X. Shared alignment promotes agreement between shared pathways, orthogonality suppresses within-branch shared/private leakage, and private purity discourages private features from encoding the complementary branch:

$$\mathcal{L}_{\mathrm{align}} = \frac{1}{3}\sum_{\ell=1}^{3} \left\|\mathcal{N}(S_\ell^{\mathrm{hv}}) - \mathcal{N}(S_\ell^{\mathrm{i}})\right\|_F^2, \tag{9}$$

$$\mathcal{L}_{\mathrm{orth}} = \frac{1}{6}\sum_{\ell=1}^{3}\sum_{r\in\mathcal{R}} \left\|\mathcal{N}(S_\ell^{r})\mathcal{N}(P_\ell^{r})^\top\right\|_F^2, \tag{10}$$

$$\mathcal{L}_{\mathrm{pur}} = \frac{1}{6}\sum_{\ell=1}^{3}\sum_{r\in\mathcal{R}} \left\|\mathcal{N}(P_\ell^{r})\mathcal{N}(L_\ell^{\bar{r}})^\top\right\|_F^2. \tag{11}$$

The total objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{rgb}} + \lambda_{\mathrm{hvi}}\mathcal{L}_{\mathrm{hvi}} + \lambda_{\mathrm{align}}\mathcal{L}_{\mathrm{align}} + \lambda_{\mathrm{orth}}\mathcal{L}_{\mathrm{orth}} + \lambda_{\mathrm{pur}}\mathcal{L}_{\mathrm{pur}}. \tag{12}$$

The regularizers constrain statistical relationships among representations rather than imposing predetermined physical semantics on particular latent channels.

Table 1: Paired comparison on LOLv1, LOLv2-Real, and LOLv2-Synthetic. Higher PSNR/SSIM and lower LPIPS are better. A dash indicates a value not reported in the source.

	LOLv1			LOLv2-Real			LOLv2-Synthetic		
Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RUAS (2021)	18.635	.516	.256	19.042	.509	.296	17.778	.693	.342
MIRNet (2023)	24.140	.842	.131	20.357	.782	.317	21.940	.846	–
LLFlow (2022)	25.132	.872	.117	26.200	.888	.137	24.807	.919	.067
DDNet (2024)	25.530	.823	.182	28.374	.861	.165	27.938	.927	.069
LLFormer (2023)	25.758	.823	.167	26.197	.819	.209	28.006	.927	.061
SNR-Aware (2022)	26.716	.851	.152	27.209	.871	.157	27.787	.941	.054
Diff-Retinex (2023)	27.091	.873	.087	25.049	.859	.152	25.127	.911	.087
RetinexFormer (2023)	27.167	.848	.125	27.667	.855	.165	28.965	.937	.059
LYT-Net (2025)	26.572	.833	.129	28.352	.877	.128	26.305	.913	.095
RetinexMamba (2025)	26.343	.838	.136	27.765	.864	.169	29.400	.942	.055
GSAD (2023)	27.640	.875	.091	28.350	.891	**.096**	28.325	.943	.049
Wave-Mamba (2024)	28.418	.890	.087	28.542	.889	.105	29.630	.949	.046
CIDNet (2025)	28.141	.889	.099	28.139	.892	.101	29.566	.950	.044
SPARNet	**28.668**	**.891**	**.057**	**28.602**	**.893**	.110	**30.442**	**.953**	**.036**

Table 2: Paired comparison on Sony-Total-Dark (SID-derived).

Method	PSNR↑	SSIM↑
RetinexNet (2018)	15.695	.395
Zero-DCE (2020)	14.087	.090
URetinexNet (2022)	15.519	.323
LLFlow (2022)	16.226	.367
LEDNet (2022)	20.830	.648
CIDNet (2025)	22.904	.676
SPARNet	**23.030**	**.688**

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

We conduct paired evaluations on LOLv1 (Wei et al., 2018), LOLv2-Real and LOLv2-Synthetic (Yang et al., 2020), and Sony-Total-Dark derived from SID (Chen et al., 2018). Unpaired generalization is evaluated on DICM (Lee et al., 2013), LIME (Guo et al., 2017), MEF (Ma et al., 2015), NPE (Wang et al., 2013), and VV (Vonikakis et al., 2018). We report PSNR, SSIM, and LPIPS on paired data and NIQE on unpaired data.

SPARNet is implemented in PyTorch 2.6.0 and trained on a single NVIDIA RTX 4090. Its three stages use 32, 64, and 128 channels and 2, 4, and 8 attention heads, respectively. Training images from LOLv1/LOLv2-Real, LOLv2-Synthetic, and Sony-Total-Dark are cropped into 400×400, 384×384, and 256×256 patches. LOL models use batch size 2 for 1,000 epochs; Sony-Total-Dark uses batch size 4 for 800 epochs. Adam is initialized with a learning rate of 10^{-4}.

4.2 QUANTITATIVE COMPARISONS

Table 1 shows that SPARNet obtains the highest values in eight of the nine reported metric–dataset combinations over three LOL benchmarks. Relative to Wave-Mamba, its LOLv1 PSNR improves by 0.250 dB and LPIPS decreases from 0.087 to 0.057. On LOLv2-Synthetic, it obtains 30.442 dB PSNR, 0.953 SSIM, and 0.036 LPIPS. On LOLv2-Real, it has the highest PSNR and SSIM among the tabulated methods, although its LPIPS of 0.110 leaves room for perceptual improvement.

Table 3: NIQE on five unpaired low-light datasets; lower is better.

Method	DICM	LIME	MEF	NPE	VV	Avg.
RUAS (2021)	6.62	5.39	5.41	6.46	4.94	5.76
LLFlow (2022)	3.72	4.12	3.69	4.06	**3.05**	3.73
Retinexformer (2023)	3.85	4.31	3.67	3.95	3.09	3.77
QuadPrior (2024)	4.82	5.22	5.11	4.78	3.99	4.78
GSAD (2023)	4.12	4.32	3.71	3.95	6.23	4.47
Wave-Mamba (2024)	3.76	4.19	3.82	**3.75**	3.09	3.72
CIDNet (2025)	3.79	4.13	3.56	3.87	3.21	3.71
SPARNet	**3.66**	**3.69**	**3.31**	3.80	3.21	**3.53**

Table 4: Component ablation on LOLv1. CSLF denotes shared–private cross-space low-frequency latent fusion; SP Reg. denotes shared–private regularization.

ID	LF fusion	SP Reg.	Subband corr.	PSNR↑	SSIM↑	LPIPS↓
M0	Concat.	–	–	27.863	.882	.069
M1	CSLF	Yes	–	28.228	.884	.065
M2	Concat.	–	Yes	28.090	.889	.064
M3	CSLF	Yes	Yes	**28.668**	**.891**	**.057**

On Sony-Total-Dark, SPARNet yields 23.030 dB PSNR and 0.688 SSIM (Table 2), improving on the listed CIDNet values by 0.126 dB and 0.012. Robustness under severe signal-dependent noise remains a limitation.

On the five unpaired datasets, Table 3 reports an average NIQE of 3.53, the lowest among the methods tabulated. SPARNet obtains the lowest NIQE on DICM, LIME, and MEF; it is second on NPE and not the best on VV.

4.3 Qualitative Comparisons

Figure 5 shows paired LOL examples and five unpaired real-world datasets. In the LOLv1 swimming-pool scene, SPARNet separates the strokes of LED digits while suppressing chromatic noise in dark windows. In the LOLv2-Real gymnasium example, it restores thin railings and repetitive seat textures with fewer color artifacts and less local blurring. In the LOLv2-Synthetic outdoor example, it maintains the brightness of the white boat hull while preserving railings, structural boundaries, and color transitions. In the unpaired examples, enhanced shadow regions, highlight retention, and edge detail illustrate the intended effects of low-frequency collaboration and directional subband correction. These are qualitative observations, not additional metric claims.

5 Ablation Studies and Representation Analysis

5.1 Component Analysis

Table 4 evaluates CSLF and directional subband correction on LOLv1. Relative to the simple low-frequency concatenation baseline M0, the full M3 configuration improves PSNR by 0.805 dB and SSIM by 0.009, and reduces LPIPS from 0.069 to 0.057. Adding subband correction to M1 or

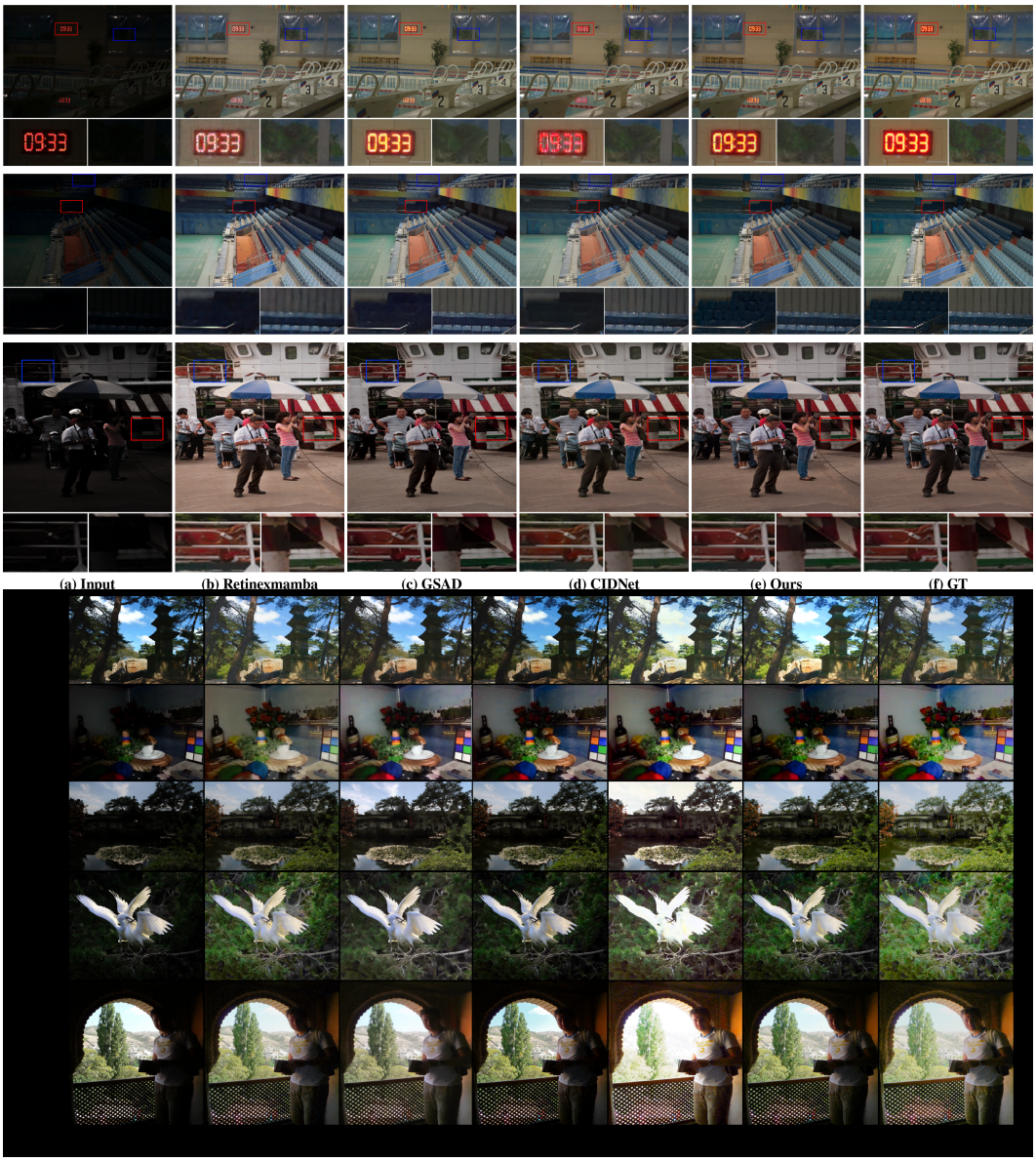

Figure 5: Qualitative comparison on three paired LOL benchmarks (top) and five unpaired datasets (bottom). Red and blue boxes mark chromatic noise, thin structures, repetitive textures, and highlight regions. The "Ours" column denotes SPARNet.

introducing CSLF into M2 yields further improvements, consistent with complementary contributions of the two components.

5.2 Effect of Shared–Private Regularization

To isolate regularization, we retain CSLF on LOLv2-Real while setting $\lambda_{\text{align}} = \lambda_{\text{orth}} = \lambda_{\text{pur}} = 0$. Restoring the terms improves PSNR and SSIM by 0.473 dB and 0.0082 and decreases LPIPS by 0.0179 (Table 5). We also report image-level linear CKA averaged across three stages: A_s is cross-branch shared alignment, L_s is within-branch shared/private leakage, and $G_s = A_s - L_s$ is the separation margin.

The full model increases A_s, reduces L_s from 0.589 to 0.051, and changes G_s from -0.157 to 0.506. Figure 6 illustrates the corresponding stage-wise CKA matrices on a shared color scale. These

Table 5: Shared–private regularization on LOLv2-Real. A_s: shared alignment; L_s: leakage; $G_s = A_s - L_s$: separation margin.

Variant	SP Reg.	PSNR↑	SSIM↑	LPIPS↓	A_s ↑	L_s ↓	G_s ↑
w/o SP Reg.	No	28.129	.8848	.1279	.432	.589	-.157
Full	Yes	**28.602**	**.8930**	**.1100**	**.558**	**.051**	**.506**

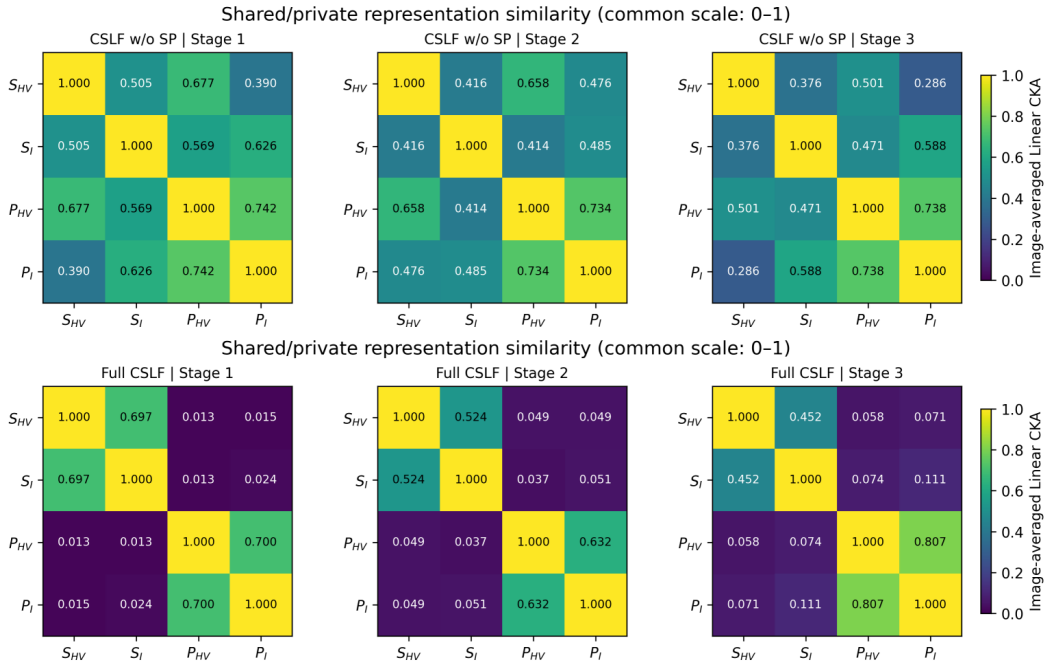

Figure 6: Stage-wise linear CKA matrices without (top) and with (bottom) shared–private regularization. The two matrices use the same color scale.

measurements are consistent with clearer cross-space alignment in shared pathways and stronger separation of shared and private features.

6 Conclusion

We proposed SPARNet to address the correlated yet weakly linearly substitutable characteristics of low-frequency HV and I components in HVI space. Its three-level wavelet pyramid aggregates multi-scale low-frequency context, CSLF models cross-branch dependencies through shared–private latent collaboration, and low-frequency-guided directional correction coordinates detail restoration with noise suppression. Experiments on paired and unpaired benchmarks, component ablations, and post-training CKA analyses support the proposed design. Remaining limitations include perceptual distance on LOLv2-Real and restoration under extremely dark, heavily noisy conditions.

References

Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Yuanhao Cai, Hao Bian, Jing Lin, Hao Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12504–12513, 2023.

C. Chen, Q. Chen, J. Xu, and V. Koltun. Learning to see in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3291–3300, 2018.

Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1780–1789, 2020.

X. Guo, Y. Li, and H. Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017.

Devamanyu Hazarika, Roger Zimmermann, and Soujanya Poria. Misa: Modality-invariant and -specific representations for multimodal sentiment analysis. In *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 1122–1131, 2020.

Z. He, X. Yuan, G. Hao, and W. Wang. Dlfe-net: Preserving details and removing noise using hvi color space for low-light image enhancement. *Sensors*, 25(17):5353, 2025.

Jiachuan Hou, Zhe Zhu, Junhui Hou, Haoning Liu, Huan Zeng, and Hong Yuan. Global structure-aware diffusion process for low-light image enhancement. In *Advances in Neural Information Processing Systems*, volume 36, pp. 79734–79747, 2023.

J. Huang, Y. Liu, F. Zhao, K. Yan, J. Zhang, Y. Huang, M. Zhou, and Z. Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *Computer Vision – ECCV 2022*, pp. 163–180. Springer, 2022.

H. Jiang, A. Luo, H. Fan, S. Han, and S. Liu. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics*, 42(6):238:1–238:14, 2023.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey E. Hinton. Similarity of neural network representations revisited. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 3519–3529. PMLR, 2019.

C. Lee, C. Lee, and C.-S. Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE Transactions on Image Processing*, 22(12):5372–5384, 2013.

Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9396–9416, 2022.

Y. Li, L. Luo, and H. Li. Hviformer: A dual-stage low-light image enhancement method based on hvi representation. *Applied Sciences*, 16(5):2450, 2026.

Z. Luo, F. K. Gustafsson, Z. Zhao, J. Sjölund, and T. B. Schön. Controlling vision-language models for multi-task image restoration. In *International Conference on Learning Representations*, 2024.

K. Ma, K. Zeng, and Z. Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015.

Stéphane Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

L. Tang, C. Wang, C. Yang, L. Mei, and M. Shao. Hvi-ffnet: Low-light enhancement via hvi color space and wavelet-adapted frequency domain fusion. In *Fifth International Conference on Algorithms, High Performance Computing, and Artificial Intelligence (AHPCAI 2025)*, volume 14142, pp. 141421A. SPIE, 2026.

V. Vonikakis, R. Kouskouridas, and A. Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77(8):9211–9231, 2018.

C. Wang, H. Wu, and Z. Jin. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 7459–7469, 2023a.

S. Wang, J. Zheng, H.-M. Hu, and B. Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.

Tao Wang, Kai Zhang, Tianfan Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 2654–2662, 2023b.

Z. Wang, Z. Shang, L. Zhang, H. Ghayvat, S. He, and X. Ning. Fusing physical priors and visual mamba: An snr-aware framework for low-light enhancement in hvi space. *Information Fusion*, 134:104350, 2026.

C. Wei, W. Wang, W. Yang, and J. Liu. Deep retinex decomposition for low-light enhancement. In *Proceedings of the British Machine Vision Conference*, pp. 155, 2018.

Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17714–17724, 2022.

Z. Xu and S. Hao. Mamba-transformer for low-light image enhancement in hvi color space. *Multimedia Systems*, 32:7, 2026.

Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kaiming Shi, Peng Wu, Weisheng Dong, Jing Sun, and Yanning Zhang. Hvi: A new color space for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5678–5687, 2025.

Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3063–3072, 2020.

Xunpeng Yi, Han Xu, Hao Zhang, Lin Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12302–12311, 2023.

Y. Zhang, H. Zheng, X. Xu, and H. Zhu. Hvi-based spatial–frequency-domain multi-scale fusion for low-light image enhancement. *Applied Sciences*, 15(19):10376, 2025.

W. Zou, H. Gao, W. Yang, and T. Liu. Wave-mamba: Wavelet state space model for ultra-high-definition low-light image enhancement. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 1534–1543, 2024.

Table B.1: Core architecture configuration.

Stage	1	2	3
Channels	32	64	128
Attention heads	2	4	8
LL resolution	$H/2$	$H/4$	$H/8$

TECHNICAL SUPPLEMENTARY MATERIAL

This appendix provides implementation, optimization, loss-weight, evaluation, and ablation details that complement the main text. It is intended to make the reported experiments easier to reproduce without duplicating the method derivations and benchmark discussions already presented above.

A SCOPE AND RELATION TO THE MAIN TEXT

The main text defines the SPARNet architecture, the three-level HVI wavelet pathway, the Shared–Private Cross-Space Low-Frequency Latent Fusion module (CSLF), the low-frequency-guided directional subband correction module, and the complete optimization objective. This appendix specifies the final dataset-dependent training configurations and clarifies the evaluation and ablation protocols.

Following the anonymized submission format, this appendix contains no author identities, institutional information, or non-anonymous repository links. Additional implementation, optimization, evaluation, and reproducibility details are provided below.

B ARCHITECTURE AND COMPUTING ENVIRONMENT

All experiments are implemented in PyTorch 2.6.0 and trained on a single NVIDIA GeForce RTX 4090 GPU. SPARNet uses three wavelet stages. The stage channel dimensions are 32, 64, and 128, and the corresponding cross-branch attention modules use 2, 4, and 8 attention heads, respectively. Haar discrete wavelet transform and inverse wavelet transform are used for multi-scale decomposition and reconstruction. The DA-CLIP encoder is frozen and is used only to provide auxiliary image-content and degradation-condition priors; its parameters are not updated during SPARNet training.

C TRAINING PROTOCOL

C.1 COMMON OPTIMIZATION SETTINGS

Training uses paired low-light and normal-light images as supervision. Unless a dataset-specific value is listed in Table C.1, the following settings are shared across datasets:

- Optimizer: Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.999$.
- Initial learning rate: 1×10^{-4}.
- Learning-rate schedule: a three-epoch warm-up followed by cosine annealing to a minimum learning rate of 1×10^{-7}.
- Data augmentation: random horizontal and vertical flipping after random patch cropping.
- Gradient clipping: maximum gradient norm of 1.0.
- Checkpoint interval: one checkpoint every 10 epochs.
- Validation metrics for paired data: PSNR, SSIM, and LPIPS.

Table C.1: Dataset-specific training configurations. The Sony-Total-Dark setting uses a larger batch and fewer epochs than the LOL settings.

Dataset	Patch size	Batch size	Epochs	Initial LR	Minimum LR
LOLv1	400×400	2	1000	1×10^{-4}	1×10^{-7}
LOLv2-Real	400×400	2	1000	1×10^{-4}	1×10^{-7}
LOLv2-Synthetic	384×384	2	1000	1×10^{-4}	1×10^{-7}
Sony-Total-Dark	256×256	4	800	1×10^{-4}	1×10^{-7}

Table D.1: Dataset-specific reconstruction and shared–private regularization weights.

Dataset	w_1	w_{ssim}	w_{edge}	w_{perc}	λ_{hvi}	λ_{align}	λ_{orth}	λ_{pur}
LOLv1	1.00	0.60	0.00	0.010	1.00	0.10	0.05	0.05
LOLv2-Real	1.00	0.20	0.00	0.012	1.00	0.08	0.05	0.05
LOLv2-Synthetic	1.00	0.40	0.00	0.012	1.00	0.06	0.03	0.03
Sony-Total-Dark	1.25	0.60	0.00	0.000	0.35	0.06	0.03	0.03

D Objective Functions and Loss Weights

The reconstruction loss in each representation space $s \in \{\text{rgb}, \text{hvi}\}$ is implemented as

$$\begin{aligned}\mathcal{L}_s = w_1 \mathcal{L}_1^s + w_{\text{ssim}} \mathcal{L}_{\text{ssim}}^s \\ + w_{\text{edge}} \mathcal{L}_{\text{edge}}^s + w_{\text{perc}} \mathcal{L}_{\text{perc}}^s.\end{aligned} \tag{D.1}$$

The same four internal weights are used for the RGB-space and HVI-space reconstruction terms within a given dataset configuration. The complete objective is

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\text{rgb}} + \lambda_{\text{hvi}} \mathcal{L}_{\text{hvi}} + \lambda_{\text{align}} \mathcal{L}_{\text{align}} \\ + \lambda_{\text{orth}} \mathcal{L}_{\text{orth}} + \lambda_{\text{pur}} \mathcal{L}_{\text{pur}}.\end{aligned} \tag{D.2}$$

Here, $\mathcal{L}_{\text{align}}$ aligns the shared HV/I representations, $\mathcal{L}_{\text{orth}}$ reduces shared–private leakage within each branch, and $\mathcal{L}_{\text{pur}}$ discourages a private representation from encoding the complementary branch. Table D.1 lists all final weights used for the reported models.

The edge-loss coefficient is zero in all final configurations. The perceptual term is used for the LOL datasets and disabled for Sony-Total-Dark. No additional loss term is introduced at inference time.

E Evaluation and Reporting Protocol

E.1 Paired Benchmarks

PSNR, SSIM, and LPIPS are used for LOLv1, LOLv2-Real, LOLv2-Synthetic, and Sony-Total-Dark. Higher PSNR and SSIM indicate better fidelity and structural similarity, whereas lower LPIPS indicates better perceptual similarity. During training, checkpoints are saved every 10 epochs and the three paired metrics are computed on the validation set at the corresponding evaluation points.

E.2 Unpaired Benchmarks

Generalization is evaluated on DICM, LIME, MEF, NPE, and VV. NIQE is the primary no-reference metric reported in the main text; lower values indicate better perceptual naturalness under the NIQE criterion. BRISQUE is computed as an auxiliary no-reference indicator in the experiment pipeline but is not used to determine the main-text ranking.

E.3 Statistical Scope

The restoration tables in the main text report one point estimate for each method–dataset pair. They do not claim repeated-run means, standard deviations, confidence intervals, or statistical significance for the end-to-end restoration benchmarks. The configurations in Tables C.1 and D.1 are the final configurations used for the reported models; no systematic hyperparameter-search range is claimed.

F Reproducing the Ablation Studies

F.1 Component Ablation

The four variants in the main text are defined as follows:

- **M0**: three-stage wavelet pyramid with simple low-frequency concatenation; no shared–private regularization and no directional subband correction.
- **M1**: replace low-frequency concatenation with CSLF and enable the three shared–private regularizers; directional subband correction remains disabled.
- **M2**: retain simple low-frequency concatenation and enable directional subband correction; shared–private regularization is not used.
- **M3**: full SPARNet with CSLF, shared–private regularization, and directional subband correction.

All variants should use the same dataset split, optimizer, learning-rate schedule, crop size, augmentation, checkpoint interval, and evaluation implementation as the full model on the corresponding benchmark.

F.2 Shared–Private Regularization Ablation

To obtain the "w/o SP Reg." variant, the complete CSLF architecture is retained while

$$\lambda_{\text{align}} = \lambda_{\text{orth}} = \lambda_{\text{pur}} = 0. \tag{F.1}$$

All other architecture and training settings remain unchanged. The representation analysis uses image-level linear CKA for the stage-wise shared and private features. The reported shared-alignment score A_s and shared–private leakage score L_s are averaged across the three stages, and the separation margin is

$$G_s = A_s - L_s. \tag{F.2}$$

The same CKA color scale is used for all matrices so that the variants can be compared directly.

F.3 HVI–Wavelet Motivation Analysis

The controlled motivation experiment uses 689 paired low-light/ground-truth images from the LOLv2-Real training set. Each image is transformed into HVI space and decomposed through three Haar wavelet levels. Linear CKA is computed between the paired HV and I low-frequency components at LL1, LL2, and LL3. Two null controls are evaluated: spatial shuffling and cross-image mismatching. Cross-component predictability is additionally assessed by five-fold ridge regression with folds grouped by image, preventing samples from the same image from appearing in both training and validation folds. The analysis is intended to test whether low-frequency HV and I representations contain exploitable shared structure while remaining weakly substitutable under a fixed linear mapping.

G Reproducibility Boundaries

The information above supports reproduction of the model architecture, final optimizer settings, dataset-dependent crop sizes, training lengths, loss weights, evaluation metrics, checkpoint frequency, and ablation definitions. The submitted materials do not claim bitwise-identical reproduction across machines. In particular, the current experiment record does not specify a universal training random seed, and the main restoration experiments are not reported as averages over multiple independent training runs. These items should therefore be represented conservatively in the reproducibility checklist.

H Checklist-to-Material Mapping

The method and objective definitions are given in the main text. Architecture and hardware details are provided in Appendix B; optimization and dataset-specific training settings in Appendix C; complete

loss weights in Appendix D; evaluation details in Appendix E; ablation definitions in Appendix F; and statistical limitations in Appendix G.