FEDD-NET: A FREQUENCY DIAGONAL FEATURE ENHANCED DUAL-BRANCH DIFFUSION NETWORK FOR LOW-LIGHT IMAGE ENHANCEMENT

Supplementary Material

A. Experiments

A1. The visualization results on LOL-v1 and VE-LOL-L test datasets

Fig. 1 presents a visual comparison of FEDD-Net with 5 state-of-the-art LLIEs. In the first group, the results from Diff-Retinex, SWANet, and HVI exhibit noticeable blur. The results from PairLIE, SWANet and URetinex-Net++ exhibit color distortion, *i.e*.the enhanced images of URetinex-Net++ exhibit a red color shift. In the second group, FEDD-Net demonstrates clear advantages in edge preservation, whereas PairLIE and URetinex-Net++ show blurring and artifacts. The zoomed-in views further highlight that diagonal feature enhancement allows FEDD-Net to maintain fine structural details, particularly in edges and subtle textures.

A2. Comparison results of FEDD-Net with other LLIEs on unpaired datasets

To evaluate the robustness of the FEDD-Net across different domains, this paper evaluates FEDD-Net on 5 unpaired datasets: LIME, DICM, MEF, NPE, and VV. The evaluation is conducted using the no-reference image quality assessment metric NIQE, with lower scores indicating better perceptual quality and naturalness. In Table 1, the FEDD-Net outperforms several SOTA LLIEs, including Diff-retinex, HVI and URetinex-Net++. Fig.2 shows the qualitative comparison on the 5 unpaired datasets. URetinex-Net++ fails to achieve sufficient brightness enhancement, while Diff-retinex and HVI produce reasonable brightness but suffer from blurriness and lack of detail. In contrast, FEDD-Net generates sharper textures and natural colors, further validating the robustness of our dual-diffusion modeling in capturing fine-grained global structures and local details.

B. Alation Study

B1. The impact of decomposition levels in DTCWT on the performance of FEDD-Net This paper explores the impact of the setting levels J of DTCWT decomposition in the feature extraction stage. The second row of Fig. 3 presents 3D visualizations of the pixel distributions of the enhanced results using DTCWT decomposition at levels $J = \{1, 3, 5\}$ in the HSV color space. It can be observed that a decomposition level of $J = 3$ yields a pixel distribution most similar to the ground truth across all three scales. As shown in the first row of Fig. 3, a lower decomposition level ($J = 1$) introduces noise, artifacts, and poor contrast, while a higher level ($J = 5$) leads to overexposure and excessive sharpening. Furthermore, increasing the decomposition level also results in higher computational cost. Therefore, $J = 3$ is selected as an optimal compromise between enhancement quality and computational efficiency.

This work was supported in part by the National Natural Science Foundation of China (Grant No. 62371389, 62031023, 62272383) and in part by the Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No. 23JP105).

Table 1. NIQE scores on LIME, DICM, MEF, NPE, and VV datasets. Lower NIQE scores indicate better perceptual quality. The best and second-best results are highlighted in **bold** and <u>underlined</u>. "AVG" denotes the average NIQE scores across the five datasets.

Methods	LIME	DICM	MEF	NPE	VV	AVG
RUSA	5.39	4.43	5.45	7.09	4.88	5.45
MIRNet	6.45	**4.04**	5.50	5.24	4.74	5.19
FECNet	6.04	4.18	4.71	<u>4.50</u>	3.75	4.64
Retinexformer	5.44	4.01	<u>3.73</u>	**3.89**	3.71	<u>4.16</u>
PairLIE	5.45	4.75	5.78	5.01	5.21	5.22
Diff-retinex	5.13	4.34	4.42	5.19	4.44	4.70
WF-Diff	4.55	4.97	4.71	5.01	5.35	4.92
SWANet	4.56	5.76	4.28	4.42	5.37	4.88
HVI	<u>3.61</u>	4.77	4.82	4.78	3.84	4.36
URetinex-Net++	4.42	5.35	5.89	5.11	<u>3.51</u>	4.86
(FEDD-Net)Ours	**3.44**	<u>4.14</u>	**3.65**	4.90	**3.32**	**3.87**

Table 2. Comparison of using fixed weight and learnable coefficient in the DFCA module. "*" denotes learnable weights. The best results are highlighted in **bold**.

Parameter	α	β	φ	η	PSNR↑	SSIM↑	LPIPS↓	NIQE↓
A	1	1	1	1	19.36	0.499	0.455	5.036
B	*	1	1	1	19.39	0.770	0.403	5.045
C	1	*	1	1	21.33	0.698	0.364	5.851
D	1	1	*	1	22.09	0.802	0.326	4.235
E	1	1	1	*	23.58	0.806	0.321	**4.088**
F	*	*	1	1	21.47	0.780	0.356	4.316
G	1	1	*	*	20.97	0.793	0.473	4.124
H	1	*	1	*	23.84	0.815	0.261	4.655
I	*	1	*	1	23.65	0.864	0.194	4.342
FEDD-Net	*	*	*	*	**24.73**	**0.877**	**0.065**	4.294

B2. The impact of setting $\alpha, \beta, \eta, \varphi$ parameters in DFCA on FEDD-Net performance

Table 2 compares the performance of fixed-weight and data-driven learning strategies in the DFCA module. When fixed weights ($\alpha = \beta = \varphi = \eta = 1$) are applied, both PSNR (19.36dB) and SSIM (0.499) simultaneously drop to their

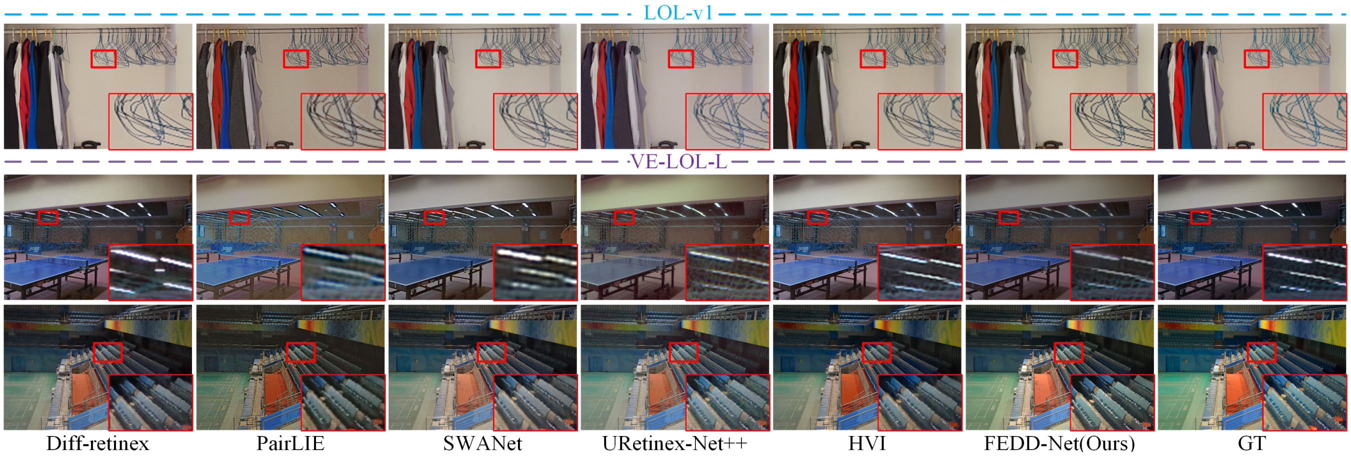

Fig. 1. Visual comparison of LLIEs on the LOLv1 and VE-LOL-L test datasets.

Fig. 2. Visual comparison of LLIE methods on the LIME, DICM, MEF, NPE, and VV datasets.

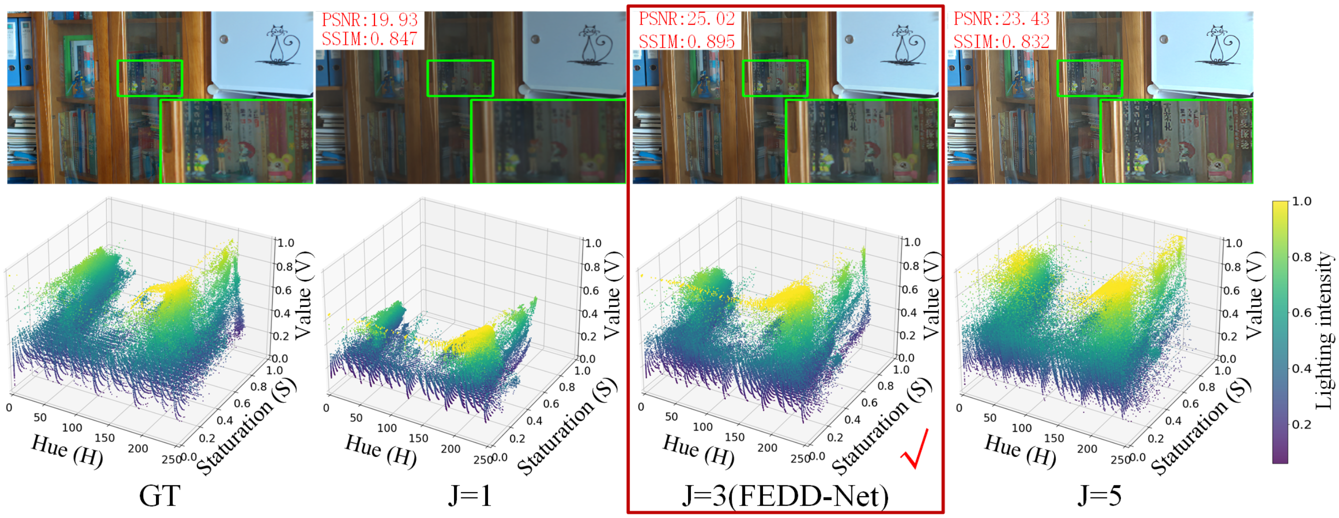

Fig. 3. The LLIE results of FEDD-Net when setting different levels of DTCWT-decomposition and the corresponding frequency visualization in HSV color space.

lowest points. Once any learnable parameter is introduced, PSNR achieves varying degrees of improvement. Partially learnable configurations improve performance, yet color bias and detail loss remain. By contrast, setting all four coefficients as learnable and optimizing them end-to-end enables the model to achieve optimal results, surpassing any fixed or partially learnable setup. This confirms that data-driven, dynamic adaptation is essential for DFCA to capture high-frequency diagonal features. Specifically, α and β adaptively balance cross-frequency aggregated and original diagonal features, suppressing artifacts while preserving details. φ and η regulate the fusion of residual-enhanced and FFN outputs, providing robust representations under complex lighting conditions.

B2. The impact of loss functions on FEDD-Net performance

Table 3. Ablation experiment on the loss function on the VE-LOL-L dataset. The best results are highlighted in bold.

Loss function	PSNR↑	SSIM↑	LPIPS↓	NIQE↓	FID↓
w/o L_{fre}	19.35	0.705	0.298	4.802	47.213
w/o $L_{content}$	23.51	0.843	0.258	4.558	48.124
w/o L_{diff}	20.89	0.766	0.319	4.423	45.359
All	**24.73**	**0.877**	**0.065**	**4.294**	**45.492**

As shown in Table 3, each loss term contributes complementarily to the overall performance. Removing the frequency loss (L_{fre}) results in the largest drop, with PSNR decreasing by 5.38 dB and SSIM by 0.172, underscoring its role in preserving structural fidelity. Excluding the content loss ($L_{content}$) leads to significant deterioration in all perceptual metrics (LPIPS, NIQE, FID), highlighting its importance for visual realism. Omitting the diffusion loss (L_{diff}) causes noticeable perceptual degradation (LPIPS = 0.319) alongside reduced PSNR and SSIM, confirming its necessity for accurate probabilistic modeling. Incorporating all losses yields the best performance, demonstrating that joint supervision of L_{fre}, $L_{content}$, and L_{diff} terms enables a balanced enhancement of both objective quality and perceptual fidelity.