Unveiling Advanced Frequency Disentanglement Paradigm for Low-Light Image Enhancement

Kun Zhou^{1,2*}, Xinyu Lin^{1,2*}, Wenbo Li³, Xiaogang Xu³ Yuanhao Cai⁴, Zhonghang Liu⁵, Xiaoguang Han¹, and Jiangbo Lu²

¹CUHK-Shenzhen, China ²SmartMore Corporation, China ³CUHK ⁴Johns Hopkins University, USA ⁵SMU, Singapore jiangbo.lu@gmail.com

Abstract. Previous low-light image enhancement (LLIE) approaches, while employing frequency decomposition techniques to address the intertwined challenges of low frequency (e.g., illumination recovery) and high frequency (e.g., noise reduction), primarily focused on the development of dedicated and complex networks to achieve improved performance. In contrast, we reveal that an advanced disentanglement paradigm is sufficient to consistently enhance state-of-the-art methods with minimal computational overhead. Leveraging the image Laplace decomposition scheme, we propose a novel low-frequency consistency method, facilitating improved frequency disentanglement optimization. Our method, seamlessly integrating with various models such as CNNs, Transformers, and flow-based and diffusion models, demonstrates remarkable adaptability. Noteworthy improvements are showcased across five popular benchmarks, with up to 7.68dB gains on PSNR achieved for six state-of-the-art models. Impressively, our approach maintains efficiency with only 88K extra parameters, setting a new standard in the challenging realm of lowlight image enhancement. https://github.com/redrock303/ADF-LLIE.

Keywords: Low-light Image Enhancement, Disentanglement Optimization, Frequency Consistency

1 Introduction

Low-light image enhancement (LLIE) endeavors to enhance the quality and exposure of photographs taken in dark environments. The state-of-the-art (SOTA) LLIE methods [3, 11, 15, 25, 45, 51, 55] achieve remarkable progress by leveraging advanced algorithms [13, 15, 45, 51] and various deep architectures [3, 55]. These models generally address the combined challenges of low-frequency adjustment and high-frequency restoration within a unified framework. However, these two tasks exhibit distinct characteristics and differ from each other in several aspects. For example, low-frequency adjustments can unintentionally amplify noise [50, 59], while high-frequency restoration may impact the recovery of illumination intensity. Relying on a single, unified model presents significant

^{*} Equal contribution



Fig. 1: (a) Illustration of predictions and corresponding frequency-disentangled components. 'SNR-De' denote the improved version of SNR [51] by our method. Visualization of the decoupled components shows that our method accurately recovers intensity in the low-frequency domain and effectively denoises in the high-frequency domain (see supplementary materials for more visual results). (b) Comparison between SOTA models and their improved versions on four representative benchmarks. It can be seen that our method significantly improves six representative SOTA models. \bullet and \star denote these baseline (SOTA) models and their enhanced version by our method, respectively.

challenges due to the complexities (combined low-frequency and high-frequency enhancement) involved in the optimization processes. Consequently, this routine may result in suboptimal outcomes, such as inaccurate illumination corrections or residual noise artifacts, as illustrated in Fig. 1(a).

Given the challenges associated with coupled frequency optimization, a pertinent question arises: How can we devise a generic frequency disentanglement paradigm capable of (1) seamlessly integrating with existing LLIE methods, (2) enhancing their frequency restoration capabilities, and (3) requiring a minimal additional model complexity (e.g., extra parameters, computational costs)? In response to this query, we introduce a generic frequency disentanglement optimization paradigm aimed at augmenting previous LLIE models.

Existing frequency decomposition learning schemes neglect the optimization interaction of these distinct frequency components and require addressing the combined degradations. In contrast, we take advantage of the image Laplace pyramid and decompose images into different frequency bands. More importantly, we present a novel low-frequency consistency constraint to decompose the complex combined restoration problem into two relatively simpler tasks, namely coarse low-frequency adjustment and frequency decoupled restoration.

In the coarse low-frequency adjustment, we present a novel Adaptive Convolutional Composition Aggregation module (ACCA), focusing on challenges at low frequency. In ACCA, we design a convolution-style spatial-channel attention mechanism to perform effective content aggregation, largely reducing the overall complexity. It outperforms Retinexformer [3] with a substantial 1dB margin in PSNR on the LOL-v2 [52], which is 5.5% model size of Retinexformer (1.6M).

In our frequency decoupled restoration, we offer the Laplace Decoupled Restoration Model (LDRM), which is mainly for high-frequency enhancement. By introducing a novel low-frequency consistency term, LDRM vastly reduces the complexity of optimization. As illustrated in Fig. 1 (b), our proposed approach reports remarkable improvement over SOTA image enhancement models, including MIR-Net [56], Restormer [55], LLFlow [45], SNR [51], Retinexformer [3] and Diff-L [15], all achieved with only a marginal increase in computational costs (*i.e.*, 88K parameters and 2.53GFLOPS for 256×256 inputs in Tab. 2). Significantly, our method notably enhances Restormer by an impressive 6.11dB, 7.68dB in PSNR on the SDSD-in [43] and SDSD-out [43] benchmarks, respectively. In Fig. 1(a) and Fig. 3, it is evident that our approach significantly improves the visual quality of the restored predictions for SOTA models, effectively addressing illumination degradation and mitigating artifacts.

In summary, our contributions are threefold:

Disentanglement learning. We introduce an effective decomposition learning framework, yielding a significant improvement. This strategy achieves separate optimization in the low-frequency and high-frequency domains while maintaining low-frequency consistency between the two stages.

Efficient coarse adjustment. Our ACCA introduces adaptive spatial-channel aggregation, resulting in SOTA low-frequency adjustment results. Remarkably, it achieves competitive results while being parameter-efficient (Tab. 4).

Universal improvements over SOTA models. Leveraging our disentanglement learning framework and lightweight coarse adjustment, we significantly improve prevailing LLIE models both quantitatively and qualitatively. Meanwhile, it allows effortless deployment of advanced low-frequency or high-frequency enhancement models as validated in Sec. 4.

2 Related Works

Traditional LLIE methods. Traditional low-light image enhancement methods have been extensively explored to improve the visibility and quality of images captured under challenging lighting conditions. Histogram equalization [34] has been widely employed to enhance contrast, but it tends to amplify noise and lacks local adaptability. On the other hand, Retinex-based methods [7,16] have aimed to separate reflectance and illumination components, but they often struggle with low-light scenarios, leading to over-enhancement and unrealistic results. Contrast-limited Adaptive Histogram Equalization (CLAHE) [35] attempts to address local contrast enhancement, but it may introduce halos around sharp edges, causing distortions and unnatural-looking enhancements.

Deep LLIE models. Deep learning-based methods [10, 14, 18, 19, 29–31, 36, 38, 40, 47, 48, 53] have gained significant attention for low-light image enhancement, aiming to tackle the challenges of both lighting improvement and denoising through a unified framework. Retinex-Net [47] combines the benefits of traditional Retinex algorithms with convolutional neural networks (CNNs). It utilizes a multi-scale architecture to decompose the input low-light image into illumination and reflectance components. MIRNet [56] utilizes a multi-scale architecture to capture and distill information at different levels. On the other hand, transformer-based architectures [6, 19, 33, 40, 51, 55, 58, 60] have dominated the LLIE field with impressive performance. For example, Star [58] uses a lightweight transformer to learn coarse image structure information for image enhancement. More recently, Retinexformer [3] presents a one-stage Retinex-based transformer

for enhancing low-light images. These transformer-based models are generally good at long-range correlation modeling but require high computational costs. **Decomposition-based LLIE.** There are two primary categories of decomposed LLIE methods [8,12,25,25,50]. First, frequency-based image restoration has garnered significant attention, particularly within the LLIE field. Xu *et al.* [50] propose a two-stage enhancement pipeline for low-light image enhancement, where they introduce a unified framework to optimize the main backbone module across both stages. Secondly, Retinex-based approaches represent low-light and normal-light images using two key components: reflectance and illumination layers. These methods estimate the reflectance as the predicted enhancement result. Hao *et al.* [12] present a semi-decoupled learning method to restore both layers using different optimization terms. RFR [8] introduces a self-knowledge distillation method for Retinex decomposition based on contrastive learning.

In summary, lots of frequency-based decomposition models have explored various techniques for LLIE. Though they explicitly decompose images into different frequency bands, they *do not decouple the optimization process* of low-frequency and high-frequency components. In contrast, our decoupled frequency restoration framework, featuring a new low-frequency consistent loss, establishes a universal decoupled enhancement paradigm that enhances various baselines.

3 Method

As mentioned previously, LLIE aims to address issues of coupled low-frequency (e.g., illumination recovery) and high-frequency (e.g., noise reduction) enhancement. Different from existing LLIE models utilizing a unified solution to handle diverse (combined) conditions, our research delves into the potential advantages of a decomposition learning strategy for image enhancement.

Based on the observation that enhancement in the low-frequency domain reports a greater impact and thus encourages lightweight structures to favor coarse learning, we design a lightweight coarse adjustment mainly for low-frequency degradation (Sec. 3.2) and a consistent coarse-to-fine (C2F) restoration (Sec. 3.3) principally for high-frequency issues, as illustrated in Fig. 2. Initially, we present the Adaptive Convolutional Composition Aggregation module (ACCA) to provide coarse adjustments, primarily focusing on illumination recovery. Following this, we utilize the Laplace Decoupled Restoration Model (LDRM) to separate the coarse results and original inputs into multi-scale high and low-frequency maps. It's worth noting that LDRM seamlessly integrates with state-of-the-art LLIE models with minimal redesign. Lastly, our proposed low-frequency consistent supervision ensures efficient decomposition learning.

3.1 Overview

Our goal is to recover a high-quality output image I_c from a low-light noisy image I. In general, we disentangle the prevailing coupled optimization into coarse phase and coarse-to-fine phase. During the coarse adjustment, we employ



Fig. 2: Overview of our proposed frequency disentanglement learning framework. It consists of two phases: (1) coarse phase: ACCA conducts coarse adjustment to initially enhance the input image I and produce the preliminary result I_l , and (2) coarse-to-fine phase: LDRM integrates Laplace representations $(M, M_l \text{ from the } I, I_l)$ for subsequent fine-grained restoration. Additionally, a low-frequency consistent loss L_i (Eq. 5) between the two phases is introduced to achieve effective disentanglement optimization.

the ACCA module to mainly recover the low-frequency information in the image I, obtaining the roughly enhanced results I_l that usually have better brightness:

$$I_l = f_l(I, \theta_l), \tag{1}$$

where θ_l is the learnable parameters of the ACCA module.

Subsequently, we proceed to the coarse-to-fine phase. In this stage, with our primary emphasis on the high-frequency domain, we propose a Laplace decoupled representation to separate low-frequency and high-frequency maps from both the original input I and the coarsely recovered image I_l :

$$M = \operatorname{Lap}(I), M_l = \operatorname{Lap}(I_l), \tag{2}$$

where $M = [m^1, m^2, \dots, m^K]$, $M_l = [m_l^1, m_l^2, \dots, m_l^K]$ refers to the decomposed maps of I, I_l , and K is the pyramid level. The coarsest feature map m^K, m_l^K generally contains the most low-frequency information of the images I and I_l , while the other feature maps $\{m^1, \dots, m^{K-1}, m_l^1, \dots, m_l^{K-1}\}$ capture the higherfrequency nuances. Following an upsampling operation on all the K - 1 lowresolution maps, we stack all of these maps alone with the channel dimension:

$$M_s = \operatorname{Stack}[m^1, \operatorname{Up}(m^2), \cdots, \operatorname{Up}(m^K), m_l^1, \operatorname{Up}(m_l^2), \cdots, \operatorname{Up}(m_l^K)], \quad (3)$$

where "Up" means bilinear up-sampling to up-scale all the low-resolution maps to the same size with the first map m^1 . Finally, Our proposed LDRM takes the stacked maps M_s as input and restores the enhanced Laplace maps $M_d = [m_d^1, m_d^2, \cdots, m_d^K]$:

$$M_d = f_d(M_s, \theta_d),\tag{4}$$

where θ_d is the learnable parameter of LDRM. To build effective supervision, we first use the reconstruction loss L_r to guide the multi-scale predictions M_d closer to corresponding decomposed maps of ground truth M_{gt} obtained by the same Laplace decouple pipeline. In addition, we also introduce a low-frequency consistent supervision L_i to effectively decompose the coarse low-frequency adjustment and high-frequency refine learning. The overall loss of LDRM is:

$$L_{total} = L_r + \alpha * L_i, L_r = \sum_{i=1}^{K} (||m_d^i - m_{gt}^i||_1^1), L_i = ||m_d^K - m_l^K||_1^1, \quad (5)$$

where α is the scale factor to balance these two loss terms, and $|| \cdot ||_1^1$ refers to ℓ_1 -norm. In the next section, we will explain ACCA for the coarse learning phase and Laplace decomposition schemes for coarse-to-fine restoration, respectively.

3.2 ACCA for Coarse Adjustment

Framework overview. Inspired by [6], we utilize a hybrid dual-branch neural network aimed at regressing both global and local low-frequency adjustment parameters. As shown in Fig. 2, our proposed Adaptive Convolutional Composition Aggregation (ACCA) is composed of two branches. In the local branch, we first employ a convolution layer to get a corresponding deep image feature of the input image. Subsequently, two parallel Window-based Convolutional Composition Attention blocks (W-CCA) are developed to regress two local transformation factor maps $\{A_l, B_l\}$, where $A_l \in R^{\{H \times W \times 3\}}$ is a scale amplification map and $B_l \in R^{\{H \times W \times 3\}}$ is a linear compensation map. The result of local parameters adjustment is calculated as:

$$I_l^{local} = A_l \odot I \oplus B_l, \tag{6}$$

where \odot refers to element-wise multiplication and \oplus means the element-wise addition. Simultaneously, we also adopt a global ISP (Image Signal Processing) branch¹ that employs a transformer-based framework, achieving a global enhancement for the locally adjusted result:

$$I_l = (A_q \otimes I_l^{local})^{B_g},\tag{7}$$

where A_g, B_g are the predicted colour matrix and gamma adjustment parameters and \otimes is the matrix multiplication.

W-CCA. Exiting studies [41, 42] have proved the importance of both spatial and channel aggregation for some low-level tasks, including denoising, super-resolution, deblurring, colorization, etc. However, performing spatial-channel fusion encounters challenges due to the high computational demands. To tackle this issue, we introduce a novel window-based convolutional composition method for efficient local enhancement.

¹ The structure of our global branch is illustrated in our supplementary material.

Given a 2D image feature F, we first employ a stridden convolutional layer to split F into several non-overlapped image patches: P = Conv(F, s, s).

Conv(·) denotes a grouped convolution with group size and stride size s. As a result, F is partitioned into $N = H \times W/s^2$ patches $P = [p_1, p_2, \cdots, p_N]$. For n-th divided image patch $p_n \in R^{\{s \times s \times C\}}$, where C represents the channel dimension and s denotes the window size. Then we perform Omni aggregation using a 3D similarity map $o_n \in R^{\{s \times s \times C\}}$:

$$\hat{p}_n = o_n \odot p_n. \tag{8}$$

Instead of directly regressing the corresponding 3D Omni similarity map o_n , we employ the tensor composition technique [21, 22, 49] to produce o_n by three regressed 1D tensors:

$$o_n = f_n^h \odot \ f_n^w \odot f_n^c. \tag{9}$$

The terms f_n^h and f_n^w represent two separable tensors in 2D space, specifically referring to the height and width dimensions. The operation \odot denotes the composition of these tensors. Additionally, the term f_n^c represents a 1D tensor that is used to establish correlations along the channel dimension. Towards a lower computational cost and pixel-adaptive aggregation, as depicted in Fig. 2, triple convolutions with a stride of *s* are utilized to regress the three separable kernels:

$$f_n^{\{h,w,c\}} = \text{Conv}_{\{h,w,c\}}(p_n, s, s).$$
(10)

To summarize, we derive 3D Omni similarity maps by Eq. 9 and perform Omni similarity aggregation by Eq. 8.

Complexity analysis of W-CCA. In this part, we will show the significant reduction in computational resource consumption achieved by W-CCA. As can be seen in the last paragraphs, we employ a stridden convolution layer to split an input image feature F into $M_h \times M_w$ image patches². To further reduce the parameter count, we implement a group-wise convolution with a group size of s and a kernel size of $s \times s$. The total computational complexity of a single W-CCA is determined by the cumulative sum of operations in Eq. 8, Eq. 9, Eq. 10 and the first stridden convolution:

$$\mathcal{O}(W\text{-}CCA) = 4HWC + 2HWC^2/s . \tag{11}$$

A more detailed complexity analysis is reported in supplementary material.

From the Eq. 11, we can see the complexity of our proposed W-CCA is *linearly* increased with the image resolution H and W.

3.3 LDRM for Consistent C2F Restoration

In the Laplace Decoupled Restoration Model (LDRM), we re-design SOTA existing LLIE models by integrating Laplace representation (also known as Laplace

² $M_h = H/s, M_w = W/s, H \times W$ is the feature resolution.

pyramid [1,2,23,24]) for subsequent effective fine-grained enhancement. In general, we only modify the first and last convolutional layers to employ our Laplace pyramid representation, dissecting a given image into multi-scale high-frequency components and a low-frequency part.

To begin with, we first utilize a Laplace pyramid decomposition layer to obtain multi-scale image maps from both input image I and the coarse enhanced result I_l . Taking the input image I for illustration, we follow [2] and obtain the Gaussian pyramid:

$$I_G^k = \begin{cases} g * I ; & \text{if } k = 1\\ g * \text{resize}(I_G^k, \downarrow 2); & \text{if } k > 1 \end{cases}$$
(12)

where $g(\cdot)$ refers to a $s \times s$ Gaussian kernel, * means the convolutional operation and 'resize' is a bilinear interpolation. Then, We derive a Laplacian pyramid by computing the difference between two neighboring Gaussian maps:

$$m^{k} = \begin{cases} I - I_{G}^{k}; & \text{if } k = 1\\ \text{resize}(I_{G}^{k-1}, \downarrow 2) - I_{G}^{k}; & \text{if } K > k > 1\\ \text{resize}(I_{G}^{k-1}, \downarrow 2); & \text{if } k = K \end{cases}$$
(13)

As shown in Eq. 13, the high-frequency components are produced by the DoG (Difference of Gaussians).

After that, all these decomposed Laplace maps are stacked and fed into our LDRM to generate the corresponding restored Laplace maps. The final reconstructed output is obtained by inverse Laplace transformation:

$$\hat{m}^{k} = \begin{cases} m^{K}; & \text{if } k = K\\ m^{k} + \text{resize}(m^{k+1}, \uparrow 2); & \text{otherwise} \end{cases}$$
(14)

where \hat{m}^1 is the final reconstructed image, and the inverse Laplace decouple is performed in reverse order, from \hat{m}^K to \hat{m}^1 . Note that we set K = 4 in our work, for it outperforms other K settings.

4 Experiments

4.1 Settings

Datasets. We evaluate our approach on several datasets for low-light image enhancement, including LOL-v2 [52], SID [4], SDSD [43] (both indoor and outdoor parts), and SMID [5]. LOL-v2 dataset: we specifically focus on its real subset, which presents challenging low-light conditions and noise degradations. This subset comprises 689 training samples and 100 testing samples. SID dataset: we utilize 2099 images for training, while the remaining 598 image pairs are reserved for evaluation. SDSD-in/SDSD-out datasets: these datasets consist of static indoor and outdoor scenes captured in various environments. SMID dataset: it contains

LOL-v2 SID SDSD-in SDSD-out SMID Method PSNR/SSIM PSNR/SSIM PSNR/SSIM PSNR/SSIM PSNR/SSIM MIR-Net [56] 20.02/0.820 20.84/0.60524.38/0.864 27.13/0.837 25.66/0.762 MIR-Net-De 24.19/0.882 24.18/0.68228.14/0.90029.45/0.88429.27/0.822Improve. +0.06+0.07+0.036+0.047+0.072-3.76-1.16/-4.17-3.34Restormer [55] 19.94/0.82722.27/0.649 25.67/0.82724.79/0.802 26.97/0.758 24.56/0.893 Restormer-De 24.76/0.694 31.78/0.918 32.47/0.904 29.11 /0.816 $-4.62/\pm0.066$ Improve. $-2.49/\pm0.043$ +6.11/+0.09 $-7.68/\pm0.102$ $\pm 2.14 / \pm 0.058$ LLFlow [45] 26.20/0.88821.72/0.61826.51/0.88326.02/0.85927.84/0.803 LLFlow-De 28.90/0.908 24.64/0.68231.60/0.917 34.58/0.916 29.37/0.820 Improve. -1.70/+0.020 $-2.92/\pm0.064$ $\pm 5.09/\pm 0.034$ $-8.83/\pm0.057$ +1.53/+0.017SNR [51] 21.48/0.849 22.87/0.625 29.44/0.89428.66/0.866 28.49/0.805 SNR-De 24.00/0.87223.55/0.66730.31/0.901 31.98/0.89730.48/0.822Improve. +0.023+0.68 +0.042+0.87+0.00+0.03+1.99/+0.01722.80/0.840 24.44/0.68029.77/0.89629.49/0.87729.15/0.815Retinexformer [3] 24.21/0.881Retinexformer-De 24.64/0.69430.54/0.90933.16/0.90530.85 /0.828 1.41/+0.04-0.20/+0.014-0.77+0.013+0.028+1.70/'+0.013Improve. -3.67Diff-L [15] 18.95/0.72221.45/0.57123.93/0.83624.19/0.83227.57/0.783 Diff-L-De 23.93/0.85323.48/0.67528.73/0.867 28.33/0.888 28.88/0.817 Improve. -4.98/+0.0812.03/+0.104-4.80/ +0.03-4.14/+0.056+1.31/+0.034

Table 1: Quantitative results (PSNR (dB)/SSIM) on five challenging LLIE benchmarks. '-De' refers to the enhanced version of our proposed approach and 'Improve.' refers to the improvements attained by our method compared to the six baselines.

static scenes captured in outdoor environments. We follow previous SOTA methods and train/evaluate various baseline models and their counterparts (with our disentanglement optimization) on the training part of each dataset, *ensuring the fairness comparison and evaluation*.

Table 2: We present comprehensive details regarding model complexities, including parameters, GFLOPS, and inference time. Consistent with [3], we assess GFLOPS and inference time using an input size of 256×256 , utilizing an NVIDIA-RTX 3090 GPU. It's worth highlighting that our framework (denoted as 'w/ ours') entails only 0.2%-5.5% additional parameters for enhancing the state-of-the-art LLIE models.

Methods	MIR-Net	Restormer	LLFlow	SNR	Retinexformer	Diff-L	w/ ours
Param.(M)	31.79	26.13	37.68	39.12	1.61	22.08	+0.088
GFLOPS	785	144.25	287	26.35	15.57	88.92	+2.53
$\mathbf{Speed}(\mathbf{s})$	0.205	0.104	0.267	0.039	0.079	0.227	+0.008

Evaluation metrics. Following the evaluation of most LLIE approaches [3, 51, 55, 56], we assess the effectiveness of our method using the $PSNR(\uparrow)$ (Peak Signal-to-Noise Ratio) and SSIM [46](\uparrow) (Structural Similarity Index Measure) metrics. Higher values indicate better results.

Implementation details. Our framework is implemented in PyTorch [32]. We use common data augmentation techniques, including random cropping, vertical/horizontal flipping, and rotation. The Adam [20] optimizer and cosine annealing scheme [28] dynamically adjust the learning rate from 5×10^{-4} to 0. We initially train the ACCA module for 100 epochs on a single NVIDIA RTX 3090 GPU. Then, we freeze the ACCA weights and train the LDRM for 40K itera-

tions on two NVIDIA RTX 3090 GPUs. The batch sizes are set to 1 (ACCA) and 32 (LDRM), respectively. All original baseline models and their corresponding LDRMs are *trained from scratch*. To achieve decoupled high-frequency restoration, we propose a novel low-frequency consistent loss in our coarse-to-fine enhancement, as illustrated in Eq. 5. Following IAT [6], the ACCA module is trained using L1 loss and VGG perceptual loss [17].

4.2 Improvement over SOTA LLIE Models

We conduct extensive experiments to assess our frequency-based disentanglement optimization. We integrate our learning strategy on six SOTA LLIE models: MIR-Net [56], Restormer [55], LLFlow [45], SNR [51], Retinexformer [3] and Diff-L [15], which serve as our baseline models. The modified versions of these models, equipped with our strategy, are denoted as MIR-Net-De, Restormer-De, LLFlow-De, SNR-De, Retinexformer-De, and Diff-L-De, respectively.

Quantitative comparison. Tab. 1 presents a quantitative comparison between the six baseline models and their enhanced counterparts using our disentanglement learning approach. It clearly demonstrates that our algorithm significantly improves the performance of all baseline models across all six benchmarks. Notably, our method achieves a remarkable enhancement of up to 7.68dB (31% improvement) for Restormer [55] on the SDSD-out benchmark. When evaluated on the same dataset, our approach showcases a substantial improvement, surpassing Restormer [55] by over 0.1 in terms of SSIM. This improvement highlights the exceptional restoration of our proposed LDRM in the high-frequency domain. Moreover, our decomposition representation leads to universal improvements for all popular frameworks, including CNNs, Transformers, flow-based, and diffusion-based methods. Even MIR-Net [55], an early SOTA LLIE model, achieves top-ranked results when enhanced with our algorithm.

Furthermore, we report the model complexity of these six LLIE models and the extra cost introduced by our approach. The results in Tab. 2 indicate that the re-designed models require only an additional 0.2% to 5% in terms of model parameters and FLOPS, demonstrating the efficiency of our method.

Table 3: Quantitative comparison (PSNR (dB)/SSIM) between our proposed frequency disentanglement scheme and segmentation-guided LLIE strategy [54]. * means segmentation information is required. Note that '+' ('-') highlighted in red (green) denotes performance improvement (reduction).

LOL-v2	SNR	SNR-SKF*	SNR-De	LLFlow	$LLFlow-SKF^*$	LLFlow-De
$PSNR \uparrow$	21.48	21.93(+0.45)	24.00(+2.52)	26.20	28.45(+2.25)	28.90(+2.70)
SSIM \uparrow	0.849	0.845(-0.006)	0.872(+0.023)	0.888	0.905(+0.017)	0.908(+0.020)
Param.(M)	39.12	39.44(+0.32)	39.21(+0.088)	37.68	39.91(+2.23)	37.77(+0.088)

Comparison with segmentation guided LLIE. In previous arts, Wu *et al.* [54] explored utilizing semantic prior for guidance LLIE (named as "SKF"),

Table 4: Quantitative comparison between ACCA and other SOTA LLIE models on LOL-v2 [52] benchmark. The best and second-best results are highlighted in red and blue, respectively. We test the inference speed (FPS: Frames Per Second) with an input size of 256×256 on an NVIDIA RTX 3090 GPU.

Method	Venue	$PSNR \uparrow$	SSIM \uparrow	Params. \downarrow	$\mathrm{FPS}\uparrow$
ZeroDCE [11]	CVPR 2020	17.63	0.617	0.079 M	62.50
Star [58]	ICCV 2021	18.26	0.546	$0.027 \mathrm{M}$	58.82
IAT [6]	BMVC 2022	23.50	0.824	$0.091 \mathrm{M}$	45.45
PairLIE [9]	CVPR 2023	18.80	0.721	$0.342 \mathrm{M}$	55.56
ACCA (Ours)	ECCV 2024	23.80	0.829	0.088M	125
MIR-Net [56]	ECCV 2020	20.02	0.820	$31.79 \mathrm{M}$	4.87
MIR-Net-v2 [57]	T-PAMI 2022	21.10	0.821	$5.86 \mathrm{M}$	15.6
Restormer [55]	CVPR 2022	19.94	0.827	26.13M	9.62
SNR [51]	CVPR 2022	21.48	0.849	$39.12 \mathrm{M}$	83.33
Retinexformer [3]	ICCV 2023	22.80	0.840	$1.61 \mathrm{M}$	27.03
LLformer [44]	AAAI 2023	20.69	0.759	$24.55 \mathrm{M}$	8.85
Diff-L [15]	Sig. Asia 2023	18.96	0.723	22.08 M	4.41

Table 5: Improvements over Restormer [55] when integrating existing fast LLIE models and our proposed ACCA into our disentanglement learning framework.

Method	Restormer	w/ ZeroDCE	w/ Star	w/ PairLIE	w/IAT	w/ ACCA
	(Baseline)	& L_i	& L_i	& L_i	& L_i	& L_i
PSNR (dB)	19.94	21.53	22.40	22.36	23.97	24.56
PSNR Gain	-	+1.59	+2.46	+2.42	+4.03	+4.62
SSIM	0.827	0.861	0.866	0.860	0.887	0.893
SSIM Gain	-	+0.034	+0.039	+0.033	+0.060	+0.066

leading to improved performance with SOTA LLIE methods. Different from their works, we present a decoupled optimization scheme to enhance existing approaches. The detailed comparisons between SKF and ours are presented in Table 3. It is clear that our proposed method attains superior improvements than SKF, with fewer extra model complexities. It is noted that our method does not require segmentation information.

Effectiveness of ACCA. In this study, we propose an efficient Adaptive Convolutional Composition Aggregation (ACCA) for coarse adjustment. To assess the effectiveness of ACCA, we quantitatively compare it with state-of-the-art LLIE models (lightweight models: ZeroDCE [11], Star [37], IAT [6], PairLIE [9]; large models: MIR-Net [56], MIR-Net-v2 [57], Restormer [55], SNR [51], Retinexformer [3]), LLformer [44] and Diff-L [15] on the LOL-V2. From the results in Tab. 4, it is evident that ACCA surpasses both lightweight and parameter-



Fig. 3: Qualitative evaluation on several benchmarks. Our integration provides accurate outcomes for both high-frequency (clearer image detail restoration) and low-frequency (more accurate illumination recovery) areas.



Fig. 4: (a) Visual validation of our disentanglement learning. (b-c) Impact of adopting different settings in our method: (b.1-2) different pyramid levels (K) in Laplace representation; (c.1-2) different α values to conduct low-frequency consistent supervision.

intensity LLIE methods in terms of PSNR and achieves competitive SSIM value compared with larger LLIE models, highlighting its superior performance.

Integration of other coarse models. We integrate lightweight LLIE methods—Star, ZeroDCE, PairLIE, and IAT—into our disentanglement learning framework to replace ACCA. The results in Tab. 5 show that these models significantly enhance the SOTA LLIE method, Restormer. Notably, IAT [6] improves Restormer by 4.03dB in PSNR and 0.06 in SSIM, demonstrating the effectiveness of our disentanglement strategy. This experiment underscores our framework's flexibility, allowing seamless integration of existing coarse adjustment algorithms.

Qualitative comparison. The qualitative comparisons are depicted in Fig. 3. It is evident that the SOTA models, primarily providing uniformly coupled optimization solutions, struggle to restore illumination and eliminate noise effectively. In contrast, with the integration of our strategy, they are capable of yielding more accurate outcomes with improved image details and fewer artifacts. More visual results are shown in our supplementary materials.

AFD-LLIE 13

Model	Restormer	ACCA	L_i	PSNR/SSIM	Param./FLOPS
Baseline	1			19.94/0.827	26.13M/144.25G
Ours w/o L_i	1	1		20.21/0.837	$26.22 \mathrm{M}/146.78 \mathrm{G}$
Ours w/ L_i	1	1	1	24.56 / 0.893	$26.22 \mathrm{M} / 146.78 \mathrm{G}$

Table 6: Effectiveness of our ACCA and L_i .

4.3 Ablation Studies

To verify the effectiveness of our proposed disentanglement learning scheme and ACCA, we conduct a comprehensive ablation analysis on the LOL-v2 dataset. Note that Restormer [55] is chosen as our baseline model.

Impacts of our designs. Compared to existing unified frameworks in SOTA LLIE methods, we propose a universal frequency-based disentanglement model and an efficient ACCA module for consistent coarse-to-fine enhancement. To evaluate our designs, we follow a systematic procedure, as shown in Tab. 6. First, we train a baseline model with the original unified framework. Next, we introduce the ACCA and LDRM modules, ablating the low-frequency consistent constraint between them. Finally, we train our full model with the proposed low-frequency consistency constraint. The results in Tab. 6 and Fig. 4(a) show that both the ACCA module and the disentanglement algorithm enhance performance. However, our full framework without the low-frequency consistent loss L_i only improves by 0.27dB over the baseline, highlighting the challenges of unified enhancement. Additionally, our designs require minimal computational demand while significantly outperforming the baseline method.

Pyramid levels (K) in eq. 13. In LDRM, we introduce an image Laplace representation method to decompose images into high- and low-frequency components, using four pyramid levels by default. To explore the impact of different pyramid levels, we train additional models with 3, 5, and 6 levels. The results in Fig. 4 (b.1-2) show that our default setting provides the best performance.

Why not end-to-end? As previously discussed, prior approaches may encounter challenges in optimizing low-frequency and high-frequency components simultaneously, possibly resulting in suboptimal outcomes. To validate our claim, we conduct experiments using six baselines, training their enhanced versions with two different optimization strategies: (1) an end-to-end training scheme (where the ACCA's weights are not frozen) and (2) freezing the weights of the ACCA (our default decoupled optimization). From Tab. 7, we can see that our decoupled optimization scheme consistently performs better than end-to-end learning.

Table 7: Quantitative comparison between two different optimization schemes.

Baseline	Retinexformer	Restormer	MIRNet	SNR	LLFlow	Diff-L
End-to-end	24.13/0.878	24.27/0.885	24.05/0.880	23.82/0.864	27.94/0.896	23.90/0.847
Ours	24.21/0.881	24.56/0.893	24.19/0.882	24.00/0.872	28.90/0.908	23.93/0.853

Different α values in Eq. 5. For decoupled frequency optimization, we introduce a novel low-frequency consistent supervision. By default, we set the value of α to 1.0, which balances both the reconstruction and low-frequency consistent loss. In this experiment, we further explore the impact of different α values on the training of our LDRM. The quantitative results shown in Fig. 4(c.1-2) demonstrate that an appropriate value of α leads to improved performance.

4.4 Limitation and future work.

While our frequency disentanglement optimization demonstrates significant enhancement over current state-of-the-art LLIE methods, it is confined to generating high-quality details that may be severely corrupted in the inputs. In the future, we intend to explore the incorporation of generative methods (such as Stable Diffusion models) into our framework.

5 Conclusion

In this study, we propose a novel framework for disentanglement learning in low-light image enhancement. Our approach involves explicitly decoupling the challenging LLIE problem into coarse low-frequency adjustment and fine high-frequency restoration tasks. To accomplish this, we introduce ACCA, a lightweight coarse adjustment module, that utilizes an efficient convolution composition module to achieve adaptive spatial-channel aggregation. Additionally, we develop LDRM by incorporating minimal re-designed SOTA LLIE methods. To ensure effective frequency decomposition, we introduce a low-frequency consistent loss for decoupled high-frequency restoration. Through comprehensive experiments, we demonstrate that our method significantly improves existing LLIE methods with minimal additional computational costs (88K learnable parameters, 2.53GFLOPS for 256×256 images).

Acknowledgments. This work is partially supported by Shenzhen Science and Technology Program KQTD20210811090149095 and also the Pearl River Talent Recruitment Program 2019QN01X226. The work was supported in part by the Basic Research Project No. HZQB-KCZYZ-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone, Guangdong Provincial Outstanding Youth Fund (No. 2023B1515020055), the National Key R&D Program of China with grant No. 2018YFB1800800, by Shenzhen Outstanding Talents Training Fund 202002, by Guangdong Research Projects No. 2017ZT07X152 and No. 2019CX01X104, by Key Area R&D Program of Guangdong Province (Grant No. 2018B030338001) by the Guangdong Provincial Key Laboratory of Future Networks of Intelligence (Grant No. 2022B1212010001), and by Shenzhen Key Laboratory of Big Data and Artificial Intelligence (Grant No. ZDSYS201707251409055). It is also partly supported by NSFC-61931024, NSFC-62172348, and Shenzhen Science and Technology Program No. JCYJ20220530143604010.

References

- 1. Adelson, E.H., Anderson, C.H., Bergen, J.R., Burt, P.J., Ogden, J.M.: Pyramid methods in image processing. RCA engineer (1984)
- 2. Burt, P.J., Adelson, E.H.: The laplacian pyramid as a compact image code. In: Readings in computer vision, pp. 671–679. Elsevier (1987)
- 3. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: ICCV (2023)
- Chen, C., Chen, Q., Do, M.N., Koltun, V.: Seeing motion in the dark. In: ICCV (2019)
- Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: CVPR (2018)
- Cui, Z., Li, K., Gu, L., Su, S., Gao, P., Jiang, Z., Qiao, Y., Harada, T.: You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. In: BMVC (2022)
- Feng, S., Bo, L.: Low-light color image enhancement based on retinex. In: International Conference on Automation, Control and Robotics Engineering (CACRE) (2020)
- Fu, H., Zheng, W., Meng, X., Wang, X., Wang, C., Ma, H.: You do not need additional priors or regularizers in retinex-based low-light image enhancement. In: CVPR (2023)
- Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.K.: Learning a simple lowlight image enhancer from paired low-light instances. In: CVPR (2023)
- 10. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. ACM TOG (2017)
- 11. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: CVPR (2020)
- 12. Hao, S., Han, X., Guo, Y., Xu, X., Wang, M.: Low-light image enhancement with semi-decoupled decomposition. IEEE TMM (2020)
- 13. Hou, J., Zhu, Z., Hou, J., Liu, H., Zeng, H., Yuan, H.: Global structure-aware diffusion process for low-light image enhancement. NeurIPS (2023)
- 14. Hu, Y., He, H., Xu, C., Wang, B., Lin, S.: Exposure: A white-box photo postprocessing framework. ACM TOG (2018)
- 15. Jiang, H., Luo, A., Fan, H., Han, S., Liu, S.: Low-light image enhancement with wavelet-based diffusion models. ACM TOG (2023)
- 16. Jobson, D.J., Rahman, Z.u., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE TIP (1997)
- 17. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV (2016)
- Kim, H.U., Koh, Y.J., Kim, C.S.: Global and local enhancement networks for paired and unpaired image enhancement. In: ECCV (2020)
- Kim, H., Choi, S.M., Kim, C.S., Koh, Y.J.: Representative color transform for image enhancement. In: ICCV (2021)
- Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
- Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM review (2009)
- 22. Kuleshov, V., Chaganty, A., Liang, P.: Tensor factorization via matrix factorization. In: Artificial Intelligence and Statistics (2015)

- 16 K. Zhou et al.
- 23. Li, Z., Shu, H., Zheng, C.: Multi-scale single image dehazing using laplacian and gaussian pyramids. IEEE TIP (2021)
- 24. Liang, J., Zeng, H., Zhang, L.: High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network. In: CVPR (2021)
- Liang, X., Chen, X., Ren, K., Miao, X., Chen, Z., Jin, Y.: Low-light image enhancement via adaptive frequency decomposition network. Scientific Reports (2023)
- 26. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: ICCV (2021)
- 27. Loh, Y.P., Chan, C.S.: Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding (2019)
- Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. In: ICLR (2017)
- Moran, S., Marza, P., McDonagh, S., Parisot, S., Slabaugh, G.: Deeplpf: Deep local parametric filters for image enhancement. In: CVPR (2020)
- Moran, S., McDonagh, S., Slabaugh, Gregorym, M., Ren, L., Tai, Y., Liu, X.: Curl: Neural curve layers for global image enhancement. In: International Conference on Pattern Recognition (ICPR) (2021)
- 31. Park, J., Lee, J.Y., Yoo, D., Kweon, I.S.: Distort-and-recover: Color enhancement using deep reinforcement learning. In: CVPR (2018)
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, highperformance deep learning library. NeurIPS (2019)
- Peng, L., Zhu, C., Bian, L.: U-shape transformer for underwater image enhancement. In: ECCV (2021)
- Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. Computer vision, graphics, and image processing (1987)
- 35. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. Journal of VLSI signal processing systems for signal, image and video technology (2004)
- Risheng, L., Long, M., Jiaao, Z., Xin, F., Zhongxuan, L.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: CVPR (2021)
- 37. Song, Y., Qian, H., Du, X.: Starenhancer: Learning real-time and style-aware image enhancement. In: ICCV (2021)
- Souibgui, M.A., Biswas, S., Jemni, S.K., Kessentini, Y., Fornés, A., Lladós, J., Pal, U.: Docentr: An end-to-end document image enhancement transformer. In: International Conference on Pattern Recognition (ICPR) (2022)
- Vonikakis, V., Kouskouridas, R., Gasteratos, A.: On the evaluation of illumination compensation algorithms. Multimedia Tools and Applications 77, 9211–9231 (2018)
- Wang, C., Pan, J., Wu, X.M.: Structural prior guided generative adversarial transformers for low-light image enhancement. arXiv preprint arXiv:2207.07828 (2022)
- Wang, H., Chen, X., Ni, B., Liu, Y., Liu, J.: Omni aggregation networks for lightweight image super-resolution. In: CVPR (2023)
- 42. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: CVPR (2020)
- 43. Wang, R., Xu, X., Fu, C.W., Lu, J., Yu, B., Jia, J.: Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In: ICCV (2021)

- Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., Lu, T.: Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In: AAAI (2023)
- 45. Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.P., Kot, A.: Low-light image enhancement with normalizing flow. In: AAAI (2022)
- 46. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE TIP (2004)
- 47. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: BMVC (2018)
- 48. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: BMVC (2018)
- Wu, X., Shi, B., Dong, Y., Huang, C., Chawla, N.: Neural tensor factorization for temporal interaction learning. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (2018)
- 50. Xu, K., Yang, X., Yin, B., Lau, R.W.: Learning to restore low-light images via decomposition-and-enhancement. In: CVPR (2020)
- 51. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: CVPR (2022)
- Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: CVPR (2020)
- 53. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE TIP (2021)
- Yuhui, W., Chen, P., Guoqing, W., Yang, Y., Jiwei, W., Chongyi, L., Shen, H.T.: Learning semantic-aware knowledge guidance for low-light image enhancement. In: CVPR (2023)
- 55. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR (2022)
- Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: ECCV (2020)
- Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for fast image restoration and enhancement. IEEE TPAMI (2022)
- 58. Zhang, Z., Jiang, Y., Jiang, J., Wang, X., Luo, P., Gu, J.: Star: A structure-aware lightweight transformer for real-time image enhancement. In: ICCV (2021)
- 59. Zheng, C., Shi, D., Shi, W.: Adaptive unfolding total variation network for low-light image enhancement. In: ICCV (2021)
- Zhou, K., Liu, K., Li, W., Han, X., Lu, J.: Mutual guidance and residual integration for image enhancement. arXiv preprint arXiv:2211.13919 (2022)