DMFourLLIE: Dual-Stage and Multi-Branch Fourier Network for Low-Light Image Enhancement

Tongshun Zhang
tszhang23@mails.jlu.edu.cn
Jilin University
College of Computer
Science and Technology
Changchun, China

Pingping Liu*
liupp@jlu.edu.cn
Jilin University
College of Computer
Science and Technology
Changchun, China

Ming Zhao
mingzhao23@mails.jlu.edu.cn
Jilin University
College of Software
Changchun, China

Haotian Lv
lvht22@mails.jlu.edu.cn
Jilin University
College of Computer
Science and Technology
Changchun, China

Abstract

In the Fourier frequency domain, luminance information is primarily encoded in the amplitude component, while spatial structure information is significantly contained within the phase component. Existing low-light image enhancement techniques using Fourier transform have mainly focused on amplifying the amplitude component and simply replicating the phase component, an approach that often leads to color distortions and noise issues. In this paper, we propose a Dual-Stage Multi-Branch Fourier Low-Light Image Enhancement (DMFourLLIE) framework to address these limitations by emphasizing the phase component's role in preserving image structure and detail. The first stage integrates structural information from infrared images to enhance the phase component and employs a luminance-attention mechanism in the luminance-chrominance color space to precisely control amplitude enhancement. The second stage combines multi-scale and Fourier convolutional branches for robust image reconstruction, effectively recovering spatial structures and textures. This dual-branch joint optimization process ensures that complex image information is retained, overcoming the limitations of previous methods that neglected the interplay between amplitude and phase. Extensive experiments across multiple datasets demonstrate that DMFourLLIE outperforms current state-of-the-art methods in low-light image enhancement.

CCS Concepts

• **Computing methodologies → Computer vision; Image processing; Low level**.

Keywords

Low-light image enhancement, Fourier frequency information, Amplitude enhancement, Phase component
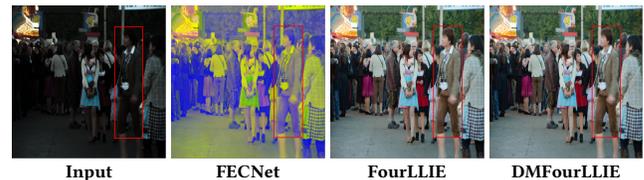
---

*Corresponding author.

1 Introduction

Input   **FECNet**   **FourLLIE**   **DMFourLLIE**

(a) First column is input, second to fourth columns represent the outputs of the respective methods in the first stage.

Ground Truth   **FECNet**   **FourLLIE**   **DMFourLLIE**
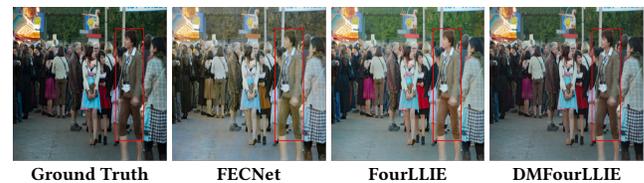
(b) First column is ground truth image, second to fourth columns represent the outputs of the respective methods in the second stage.

Figure 1: Comparison with State-of-the-Art Two-Stage Fourier-Based Methods. (a) displays initial results from FECNet [14] and FourLLIE [37], which exhibit color distortion and noise. Our method, in contrast, enhances brightness and preserves color fidelity, eliminating noise. (b) shows the superior noise suppression and color accuracy of our approach, particularly in highlighted areas.

Due to environmental constraints and hardware limitations, images often suffer from issues like diminished visibility, uneven exposure, loss of detail, and color inaccuracies. These challenges significantly hinder the performance of computer vision applications, affecting critical tasks such as autonomous driving [35], pedestrian recognition [56], and object detection [13]. The primary aim of low-light image enhancement (LLIE) techniques is to restore images to a state of normal lighting, preserving as much texture detail as possible, thereby enhancing the efficiency of subsequent vision processing tasks. With the advancement of neural networks, traditional non-learning methods are becoming less prevalent. Recent years have seen the emergence of innovative LLIE methods

MM '24, October 28–November 1, 2024, Melbourne, VIC, Australia.

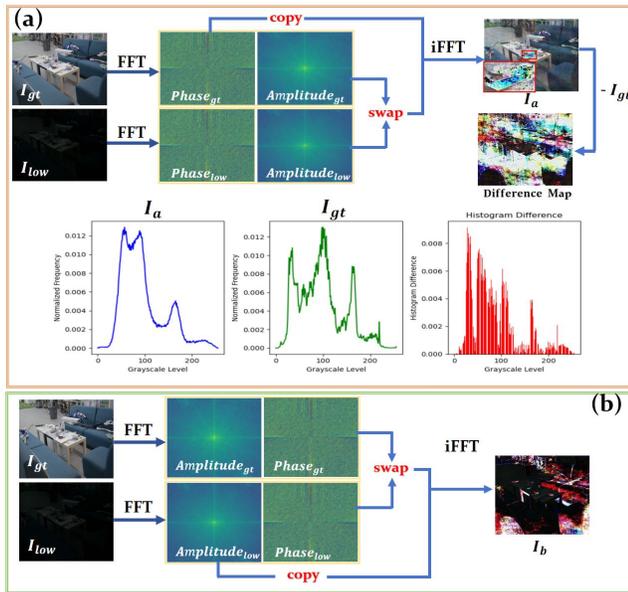Tongshun Zhang, Pingping Liu, Ming Zhao and Haotian Lv

Figure 2: Our Motivations. (a): Illustrates issues from replacing the amplitude component of a low-light image with that of a normally illuminated image, resulting in I_a. (b): Highlights the importance of the phase component, as shown by the distorted image I_b from replacing the phase component.

[9, 25, 27, 39, 53, 60, 61] in computer vision, yet a balance between method performance and computational demand remains elusive. High-performing models [44, 52] often come with substantial computational costs, whereas Fourier-based strategies [14, 37] present a cost-effective solution by leveraging Fourier frequency information for image brightening.

Researches have validated the efficacy of Fourier domain-based low-level image enhancement techniques across various tasks, including enhancement under low-light conditions [14, 37], denoising [49, 62], and super-resolution [8]. In LLIE tasks, since luminance information is predominantly encoded in the amplitude component, common strategies [14, 37] involve amplifying the amplitude component and duplicating the phase component. However, the efficacy of such strategies warrants scrutiny. Through rigorous experimentation, we have arrived at the following two critical insights:

(1) As demonstrated in Fig. 2(a), directly replacing the amplitude component of a low-light image with that of a normally illuminated image results in the generated image I_a suffering from compromised brightness and texture, thereby introducing noise. Therefore, this indicates significant limitations in previous methods [14, 37], where they directly approximate the amplitude component between the low-light input and the ground truth image. This approach leads to the degraded image I_a being used as the learning target for the network, making the subsequent optimization challenging. We employ difference maps and histogram difference curves to visually demonstrate the disparity between I_a and I_{gt}. Furthermore, enhancing the amplitude component may result in overflow of inherently brighter regions in the original image, introducing distorted patches in the visual output (highlighted by the red boxes in I_a). As consistently demonstrated in our results shown in Fig. 1, although FECNet [14]

and FourLLIE [37] attempt to mitigate degradation by restoring the phase component and spatial information in the second stage, merely recovering these aspects does not fully address the distortions caused by the first stage.

(2) As depicted in Fig. 2(b), the distorted image I_b demonstrates that the phase component also contains a significant amount of image information, which likewise requires meticulous handling during the process of frequency domain information transfer. The phase component retains more detailed information of the image, sometimes even more critical than the amplitude component. Furthermore, swapping the phase components between two images while keeping their amplitude components unchanged results in the content of the images being exchanged along with the phase information, underscoring the significance of phase information for the structure and content of images.

Building upon the insights garnered, we propose a novel framework, DMFourLLIE (Dual-Stage Multi-Branch Fourier Low-Light Image Enhancement), which employs a two-stage design rooted in Fourier-based methods [14, 20, 37], as illustrated in Fig. 3. The inaugural stage, detailed in Fig. 4, is dedicated to Fourier reconstruction. Within the Fourier frequency domain, our aim is to concurrently augment the expressiveness of both amplitude and phase components. To bolster the structural integrity of the phase component in low-light images, we integrate infrared prior information. For the amplitude component, we employ a luminance attention map to transition the luminance space, thereby enhancing the accuracy of amplification across various brightness levels. Progressing to the second stage, depicted in Fig. 5, our focus shifts towards the reconstruction of spatial structures and textures. Here, we introduce a dual-pathway architecture that merges multi-scale spatial convolution branches with branches based on Fourier convolution [4]. This configuration is strategically designed to refine spatial structure representation while capturing intricate texture details.

Overall, the main contributions of this paper can be summarized as follows:

- In this paper, we introduce a framework named Dual-Stage Multi-Branch Fourier Low-Level Image Enhancement (DMFourLLIE). It addresses the limitations of existing methods by accurately enhancing frequency domain information through cross-modal enhancements within the Fourier domain. Additionally, it employs a dual-pathway architecture that combines multi-scale spatial convolution with Fourier convolution techniques to reconstruct fine-grained textures and spatial structures with high fidelity.

- To the best of our knowledge, DMFourLLIE is the first approach in the LLIE field to leverage the synergistic potential of infrared and luminance priors in the Fourier space. By using cross-modal infrared images to guide the structural information of the phase component and luminance attention maps for precise amplitude component enhancement, our method introduces a novel strategy for embedding complementary perceptual priors within the Fourier space.

- Extensive experiments validate the effectiveness of our hypotheses and the architectural advantages of DMFourLLIE. Our approach outperforms current state-of-the-art methods in the LLIE domain.
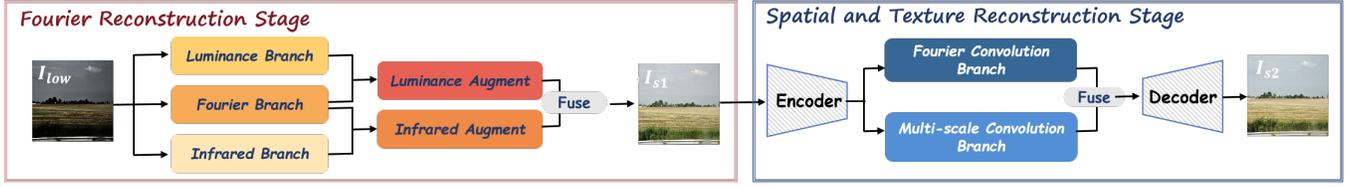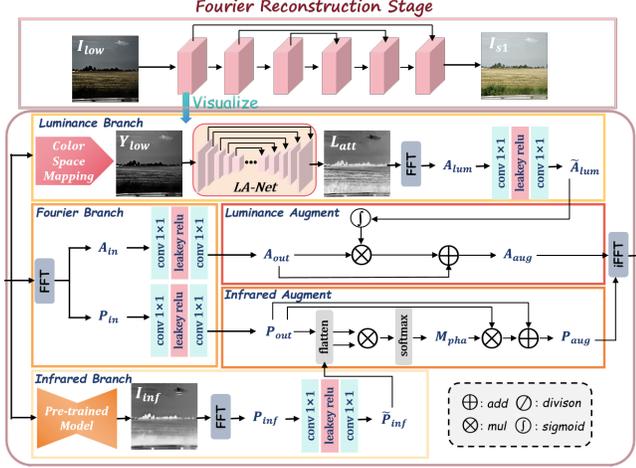
Figure 3: Overall framework of DMFourLLIE.

Figure 4: Structure of the Fourier Reconstruction Stage.

Figure 5: Structure of the Spatial and Texture Reconstruction Stage.

2 Related Work

Non-learning and Learning-based LLIE Methods. The spectrum of Low-Light Image Enhancement (LLIE) techniques is bifurcated into non-learning and learning-based methodologies. Traditional non-learning strategies encompass Histogram-based [1, 18, 31] and Retinex-based [6, 21, 43] methods. These approaches, however, fall short in accurately capturing scene dynamics and perceptual cues, leading to suboptimal brightness and color fidelity in real-world scenarios, along with exacerbated noise in extremely low-light conditions. The advent of deep learning catalyzed the development of a plethora of learning-based LLIE strategies [9, 25, 27, 53, 60]. Notably, [7] introduced an unsupervised model that derives adaptive priors from pairs of low-light images. Similarly, [42] unveiled a semantic-aware framework that imbues models with a rich tapestry of semantic knowledge. Furthermore, studies [14, 37] have validated the efficacy of Fourier space manipulation for enhancing illumination in low-light scenes, achieving nuanced improvements with minimal model complexity.

Infrared and Visible Image Fusion. In computer vision, combining infrared and visible light imagery [3, 58] enhances visual representation in various tasks. However, the lack of annotated visible-infrared image pairs hinders the development of infrared-based deep learning models. This challenge has led to methods [16, 29] for training models in visible-to-infrared image translation. [15, 19] introduced a translation model focused on edge preservation to maintain spatial textures. Unlike traditional shared encoder designs, [59] explored a novel approach for feature extraction and fusion. Inspired by the unique properties of infrared imagery and
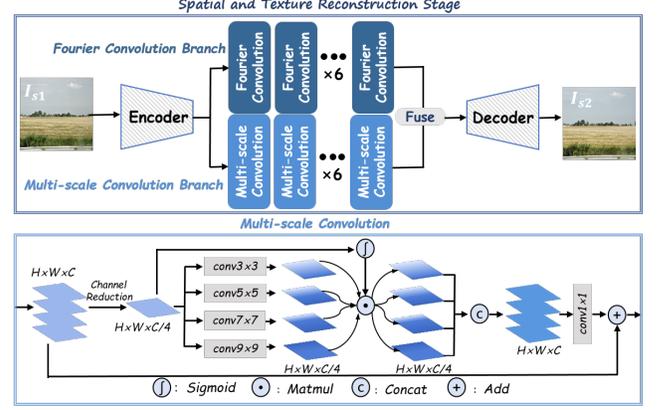
previous research, we investigate the application of infrared in the Fourier domain for illumination enhancement.

Fourier-based LLIE Methods. The exploration of Fourier frequency information for LLIE is still in its early stages. [8] investigated Fourier loss for super-resolution enhancement, while [62] used Fourier transform for image fusion to combine spatial-frequency signals. [49] tackled dehazing by exploring both frequency and spatial domains. Additionally, [14] highlighted the role of amplitude in encoding brightness and phase in capturing structural details, proposing a network for exposure correction. FourLLIE [37] showed that global feature extraction via Fourier analysis can be achieved without increasing model parameters. UHDFour [20] noted the similarity of amplitude components across image resolutions and proposed an illumination technique for ultra-high-definition content. Despite these advancements, the isolated manipulation of amplitude and phase components has limitations, highlighting the need for further exploration of Fourier frequency information.

3 Method

3.1 Fourier Frequency Information

Firstly, we provide a brief introduction to Fourier frequency information. Fourier frequency information refers to the representation of an input image x in the Fourier space X through a transform function $\mathcal{F}$. The input image x has a shape of $H \times W$, where H represents the height and W represents the width. $\mathcal{F}$ is expressed as:

$$\mathcal{F}(x)(u,v) = X(u,v) = \frac{1}{\sqrt{H \times W}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} x(h,w) e^{-j2\pi\left(\frac{h}{H}u + \frac{w}{W}v\right)},$$

(1)

where h, w are the coordinates in the spatial space and u, v are the coordinates in the Fourier space, j is the imaginary unit, the inverse

MM '24, October 28–November 1, 2024, Melbourne, VIC, Australia.

Tongshun Zhang, Pingping Liu, Ming Zhao and Haotian Lv

process $\mathcal{F}$ is denoted as $\mathcal{F}^{-1}$. Each complex component $X(u,v)$ can be represented by the amplitude component $\mathcal{A}(X(u,v))$ and phase component $\mathcal{P}(X(u,v))$. These two components are expressed as:

$$\mathcal{A}(X(u,v)) = \sqrt{R^2(X(u,v)) + I^2(X(u,v))}, \qquad (2)$$

$$\mathcal{P}(X(u,v)) = \arctan\left[\frac{I(X(u,v))}{R(X(u,v))}\right], \qquad (3)$$

where $R(x)$ and $I(x)$ represent the real and imaginary part of $X(u,v)$ respectively. In our method, the Fourier transform and inverse procedure is computed independently on each channel of feature maps.

3.2 Fourier Reconstruction Stage

As shown in Fig .4, the low-light input image I_{low} is fed into the Fourier reconstruction phase, resulting in the output of the first stage, I_{s1}. This stage consists of six identical components with skip connections, where the first component is visualized in detail. The LA-Net and pretrained models of Infrared Branch are only activated in the first component, with their outputs being passed and processed in the subsequent five components. The branches are described as follows:

Fourier Branch: The input image is first transformed into the Fourier space to obtain the amplitude component A_{in} and the phase component P_{in}. Once transformed into the Fourier space, the assumption of spatial invariance no longer holds [8]. Therefore, we apply a 1×1 convolutional layer and LeakyReLU activation function in this space to extract the amplitude component, to avoid information loss. The amplitude and phase components are each extracted through a 1×1 convolutional layer with LeakyReLU activation, resulting in A_{out} and P_{out}. Unlike existing methods [14, 37], which only enhance the amplitude component and adopt a strategy of replicating the phase component, our analysis in Fig .2 emphasizes the importance of phase information for the structure and content of images, indicating that merely replicating the phase component is not the best strategy.

Luminance Branch: To address the potential overflow issue in the Fourier domain space, we learn a luminance attention map through the luminance branch to precisely guide the enhancement level of the amplitude component in different areas, correctly enhancing underexposed areas while avoiding over-enhancement of correctly exposed areas. According to the study by Bread [10], among various color spaces, the luminance component (Y) in the YCbCr color space performs best and is least affected by noise. Inspired by this finding, we predict the luminance attention map L_{att} through LA-Net (Luminance Attention Net) based on the luminance component Y_{low} from the YCbCr color space. LA-Net follows the classic U-Net architecture with an encoder-decoder structure. This process and the learning target can be represented as:

$$L_{att} = LA\text{-}Net(Y_{low}), \qquad L_{att} = \frac{|Y_{low} - Y_{gt}|}{Y_{gt}}, \qquad (4)$$

where Y_{low} and Y_{gt} represent the luminance components (Y channel) extracted from the input low-light image and the ground truth image, respectively. In the predicted L_{att}, brighter areas correspond to darker areas in the input low-light image, and vice versa. Then, L_{att} is transformed into the Fourier space to obtain the amplitude

component A_{lum}, and features are extracted through two 1×1 convolutional layers with LeakyReLU activation, resulting in $\tilde{A}_{lum}$.

Luminance Augment: To address the issue of varying brightness across different regions and the potential overflow problem during the transfer of amplitude components in the Fourier space, we propose a Luminance Augment operation. This operation leverages L_{att} to make A_{out} pay more attention to darker areas while reducing focus on brighter areas. Specifically, $\tilde{A}_{lum}$ undergoes a sigmoid operation and is then element-wise multiplied by A_{out}, and through a residual connection, we obtain the enhanced amplitude component A_{aug}. The process can be described as follows:

$$A_{aug} = A_{out} \cdot sigmoid(\tilde{A}_{lum}) + A_{out}, \qquad (5)$$

through the aforementioned operation, A_{aug} effectively adjusts the attention towards enhancing different areas, thereby more accurately amplifying the amplitude component.

Infrared Branch: The input I_{low} is processed by a pre-trained RGB-to-TIR image translation model [15, 19], which focuses on edge preservation. This model, trained on the VIPER[32], FLIR-ADAS, and STheReO[50] datasets, generates infrared images from low-light scenes, improving infrared image performance in dark conditions. This process retains spatial texture information and generates the corresponding infrared image I_{inf} from I_{low}. Infrared images, unaffected by illumination, color, and distortion, enhance object outlines. Since contour information primarily resides in the phase components of the Fourier domain, we extract the phase component P_{inf} of I_{inf} using FFT. Features are then extracted through two 1×1 convolutional layers with LeakyReLU activation, resulting in $\tilde{P}_{inf}$.

Infrared Augment: The phase component in the Fourier space is crucial for the structure and content of an image. Considering the rich structural information of infrared images and their high robustness in dark scene tasks, we are inspired by the transposed attention in Restormer [51] and design an Infrared Augment operation. This operation achieves phase attention maps M_{pha} through cross-modal fusion of phase component features of infrared and visible light images in the Fourier space. The phase attention map M_{pha} is described as follows:

$$M_{pha} = softmax\left(flatten(P_{out}) \times flatten(\tilde{P}_{inf})\right), \qquad (6)$$

then, P_{out} is multiplied by M_{pha} and added through a residual connection, resulting in the enhanced phase component P_{aug}. This process is as follows:

$$P_{aug} = P_{out} \times M_{pha} + P_{out}, \qquad (7)$$

to this end, we obtain the enhanced amplitude component A_{aug} and phase component P_{aug}, which are then transformed back into the spatial domain through the inverse Fourier transform to yield the output I_{s1}:

$$I_{s1} = \mathcal{F}^{-1}(A_{aug} \times \cos(P_{aug}) + A_{aug} \times \sin(P_{aug})). \qquad (8)$$

3.3 Spatial and Texture Reconstruction Stage

While the Fourier space conveys global information, it lacks the enhancement of spatial details and may introduce additional noise and feature loss during the Fourier transform and feature transfer processes. To address this, we design the Spatial and Texture

Figure 6: Visual comparison on LSRW-Huawei, LOL-v2-real, and LOL-v2-Synthesis datasets. DMFourLLIE effectively enhances visibility while preserving image texture and original colors without introducing additional noise.

Reconstruction Stage to further refine image details and texture structures. Specifically, as shown in Fig. 5, this stage consists of two branches: the Multi-scale Convolution Branch and the Fourier Convolution Branch. The first-stage output I_{s1} is initially downsampled by an encoder, then processed through dual-path branches composed of six Multi-scale Convolutions and Fourier Convolutions, respectively. Finally, features from both branches are concatenated and fed into a decoder, producing the final brightening result I_{s2}. These two branches work collaboratively to optimize pixel-level reconstruction, effectively recovering important details and structural patterns in the original image.

Multi-scale Convolution Branch: The structure of the Multi-scale Convolution is depicted below Fig. 5. In this branch, multiple convolutional layers with varying kernel sizes are utilized to capture a wide range of spatial features across multiple scales and enhance the overall representation of spatial structures. To reduce computational cost, we downscale the dimension from (C=64) to (C/4=16), and gradually increase the kernel size by 2 from 3×3 to 9×9. After applying a sigmoid operation to the reduced features, they are multiplied with the output features of different convolutional kernels, then concatenated, restoring the dimension to (C=64). The features fused with multi-scale spatial information are processed through a 1×1 convolutional layer and combined with the input via a residual connection.

Fourier Convolution Branch: Fast Fourier Convolution (FFC) [4] has been proven to handle high-resolution restoration cases with strong periodic texture, achieving robust enhancement at the resolution level. It consists of two branches: 1) a local branch that

uses regular convolution, and 2) a global branch that convolves features after a fast Fourier transform. The outputs of both branches are then merged to obtain a larger receptive field and local invariance during the restoration process [36]. However, such a powerful model is unable to learn a reasonable overall spatial structure [5]. Therefore, we pair it with the Multi-scale Convolution to form a dual-path structure, complementing it to achieve precise overall spatial structure through the final feature fusion.

3.4 Loss Function

To further improve image quality from both qualitative and quantitative aspects by considering perceptual information and regional differences, our loss function consists of four parts:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{s1} + \lambda_2 \mathcal{L}_{s2} + \lambda_3 \mathcal{L}_{per} + \lambda_4 \mathcal{L}_{lum}, \quad (9)$$

where λ denotes the loss weights, we empirically set $\lambda_1, \lambda_2, \lambda_3, \lambda_4 = [0.5, 1, 0.2, 0.1]$. $\mathcal{L}_{s1} = \|I_{s1} - GT\|_2$ represents the l_2 loss between the preliminary enhanced result I_{s1} and ground truth image GT. Similar to $\mathcal{L}_{s1}$, $\mathcal{L}_{s2} = \|I_{s2} - GT\|_2$ represents the loss between the final output I_{s2} and GT. $\mathcal{L}_{per}$ is the perceptual loss between the I_{s2} and GT, which constrains the features extracted from VGG [34] to obtain better visual results.

Additionally, to acquire the luminance attention map for correctly guiding enhancement of luminance in image regions by $LA\text{-}Net$, we adopt l_2 error to measure the prediction error:

$$\mathcal{L}_{lum} = \|LA\text{-}Net(I_{low}) - L_{att}\|_2, \quad (10)$$

where $LA\text{-}Net(I_{low})$ and L_{att} represent the predicted and expected luminance attention maps, respectively.

Methods	LOL-v2-Real			LOL-v2-Syn			LSRW-Huawei			#Param	#FLOPs
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	(M)	(G)
NPE [38] (TIP, 13)	17.33	0.4642	0.2359	16.60	0.7781	0.1079	17.08	0.3905	0.2303	-	-
LIME [11] (TIP, 16)	15.24	0.4190	0.2203	16.88	0.7578	0.1041	17.00	0.3816	0.2069	-	-
SRIE [6] (ICCV, 16)	14.15	0.5524	0.2160	14.50	0.6640	0.1484	13.42	0.4282	0.2166	-	-
Kind [57] (MM, 19)	20.01	0.8412	0.0813	22.62	0.9041	0.0515	16.58	0.5690	0.2259	8.02	34.99
MIRNet [52] (ECCV, 20)	22.11	0.7942	0.1448	22.52	0.8997	0.0568	19.98	0.6085	0.2154	31.79	785.1
Kind++ [55] (IJCV, 21)	20.59	0.8294	0.0875	21.17	0.8814	0.0678	15.43	0.5695	0.2366	8.27	-
SGM [47] (TIP, 21)	20.06	0.8158	0.0727	22.05	0.9089	0.4841	18.85	0.5991	0.2492	2.31	-
FECNet [14] (ECCV, 22)	20.67	0.7952	0.0995	22.57	0.8938	0.0699	21.09	0.6119	0.2341	0.15	5.82
HDMNet [22] (MM, 22)	18.55	0.7132	0.1717	20.54	0.8539	0.0690	20.81	0.6071	0.2375	2.32	-
SNR-Aware [44] (CVPR, 22)	21.48	0.8478	0.0740	24.13	0.9269	0.0318	20.67	0.5911	0.1923	39.12	26.35
Bread [10] (IJCV, 23)	20.83	0.8217	0.0949	17.63	0.8376	0.0681	19.20	0.6179	0.2203	2.12	1.54
FourLLIE [37] (MM, 23)	22.34	0.8403	0.0573	24.65	0.9192	0.0389	21.11	0.6256	0.1825	0.12	4.07
SNR-SKF [42] (CVPR, 23)	20.66	0.8128	0.0757	17.21	0.7738	0.0731	16.21	0.5560	0.1615	39.44	27.88
UHDFour [20] (ICLR, 23)	19.42	0.7896	0.1151	23.64	0.8998	0.0341	19.39	0.6006	0.2466	17.54	4.78
Retinexformer [2] (ICCV, 23)	22.79	0.8397	0.0724	25.67	0.9295	0.0273	21.23	0.6309	0.1699	1.61	15.57
DMFourLLIE (Ours)	22.64	0.8589	0.0520	25.83	0.9314	0.0234	21.47	0.6331	0.1781	0.41	1.69

Table 1: Quantitative comparison on LOL-v2-Real [47], LOL-v2-Syn [47], and LSRW-Huawei [12]. The best results are marked in red, the second-best are in blue.

Methods	LIME	VV	DICM	NPE	MEF	AVG
Kind	4.772	3.835	3.614	4.175	4.819	4.194
MIRNet	6.453	4.735	4.042	5.235	5.504	5.101
SGM	5.451	4.884	4.733	5.208	5.754	5.279
FECNet	6.041	3.346	4.139	4.500	4.707	4.336
HDMNet	6.403	4.462	4.773	5.108	5.993	5.056
Bread	4.717	3.304	4.179	4.160	5.369	4.194
Retinexformer	3.441	3.706	4.008	3.893	3.727	3.755
FourLLIE	4.402	3.168	3.374	3.909	4.362	3.907
DMFourLLIE	3.233	3.298	3.613	3.564	3.565	3.455

Table 2: NIQE scores on LIME, VV, DICM, NPE, and MEF datasets. The top results are highlighted in red and the second-best in blue. "AVG" denotes the average NIQE scores across these five datasets. All evaluated methods have been pre-trained on the LSRW-Huawei dataset.

Settings	PSNR↑	SSIM↑	LPIPS↓
w/o Fourier Reconstruction Stage	20.92	0.8435	0.0473
w/o Spatial and Texture Reconstruction Stage	20.88	0.8337	0.0596
w/o Luminance Branch	22.51	0.8504	0.0524
w/o Fourier Branch	20.85	0.8391	0.0579
w/o Infrared Branch	22.63	0.8476	0.0673
w/o Luminance Augment	20.84	0.8382	0.0532
w/o Infrared Augment	20.63	0.8317	0.0567
w/o Fourier Convolution Branch	22.46	0.8492	0.0576
w/o Multi-scale Convolution Branch	21.60	0.8404	0.0732
DMFourLLIE	22.64	0.8585	0.0520

Table 3: Ablation Studies on DMFourLLE with the LOL-v2-Real Dataset. The term 'w/o' denotes the absence of a specific component. 'w/o Fourier Branch' refers to a configuration that enhances only the amplitude component, similar to previous Fourier-based illumination networks, without involving the phase component. 'w/o Fourier Convolution Branch' and 'w/o Multi-scale Convolution Branch' represent scenarios where traditional convolutional layers replace these specialized modules.

4 Experiments

4.1 Datasets and Experimental Setting

DMFourLLIE is trained on three commonly used LLIE datasets: LOL-V2-Real [47], LOL-v2-synthesis [47] and LSRW-Huawei [12]. LOL-v2-Real is captured under varying exposure times and ISO in real scenes, including 689 low/normal light image pairs for training and 100 pairs for testing. LOL-v2-synthesis is synthesized from original images by analyzing the luminance channel distribution of low/normal light images. It contains 900 low light/normal light image pairs for training and 100 pairs for testing. Compared with the original LOL dataset [41], LOL-v2 is larger and more diverse, providing more convincing evaluation of performance. LSRW-Huawei was captured in real scenes using different devices, containing 3150 training image pairs and 20 test image pairs. Additionally, we also evaluated DMFourLLIE on five unpaired datasets: DICM [17] (64 images), LIME [11] (10 images), MEF [26] (17 images), NPE [38] (85 images) and VV(24 images).

DMFourLLIE adopts an end-to-end training strategy to jointly optimize the network parameters of both stages. We implement DMFourLLIE in Pytorch. During training, images are randomly cropped to 256×256 and augmented with random flipping. An ADAM optimizer with an initial learning rate of 4.0×10^{-4} is used to optimize the network, along with a multi-step scheduler. The batch size is set to 8 and the total number of training iterations is set to 1.0×10^5. Training takes approximately 18 hours on an NVIDIA 3090 GPU.

4.2 Comparison with Current Methods

In this paper, our DMFourLLIE is compared to thirteen state-of-the-art LLIE methods, including traditional approaches LIME [11],

DMFourLLIE: Dual-Stage and Multi-Branch Fourier Network for Low-Light Image Enhancement

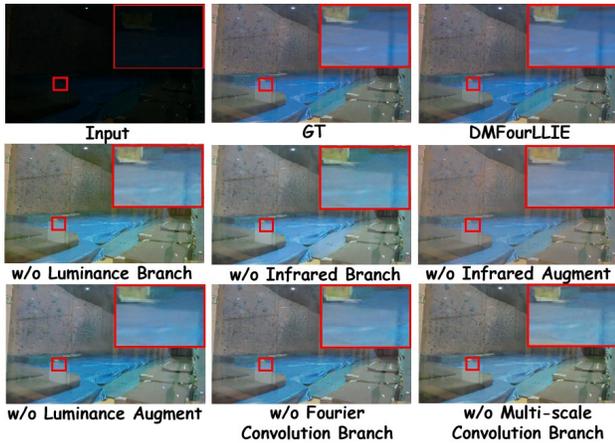MM '24, October 28–November 1, 2024, Melbourne, VIC, Australia.

Figure 7: Visualizing Component Ablation in DMFourLLIE. Visual comparisons show noticeable differences against the ground truth when specific components of DMFourLLIE are removed. Each omitted component causes varying levels of color distortion and noise, highlighting their importance for high-fidelity image enhancement.

Metrics	Loss Settings					
	w/o $\mathcal{L}_{s1}$	w/o $\mathcal{L}_{s2}$	w/o $\mathcal{L}_{lum}$	w/o $\mathcal{L}_{per}$	DMFourLLIE	$+\mathcal{L}_{amp}$
PSNR↑	21.61	20.41	21.98	21.31	**22.64**	22.33
SSIM↑	0.8480	0.8419	0.8531	0.8319	**0.8589**	0.8403

Table 4: Ablation experiments on the loss function.

NPE [38] and SRIE [6], Fourier-based methods FourLLIE [37], FEC-Net [14] and UHDFour [20]. Deep learning-based methods involved in the comparison are Kind [57], Kind++ [55], MIRNet [52], SGM [47], HDMNet [22], SNR-Aware [44], Bread [10], SNR-SKF [42] and Retinexformer [2]. All deep learning models are trained on the same datasets using their original public codes, ensuring a fair evaluation.

Quantitative results on LOL-v2-Real, LOL-v2-synthesis and LSRW-Huawei datasets. In this work, we use Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [40] and Learned Perceptual Image Patch Similarity (LPIPS) [54] as our evaluation metrics. Generally, higher PSNR and SSIM as well as lower LPIPS indicate higher similarity between two images.

As shown in Tab. 1, we evaluate our approach using the LOL-v2-Real, LOL-v2-synthesis, and LSRW-Huawei datasets. Compared to the current state-of-the-art methods, it is evident that DMFourLLIE achieves nearly the best performance across the PSNR, SSIM, and LPIPS metrics. Furthermore, we also have advantages in terms of model parameters and computational efficiency.

Visual Quality Comparisons. Due to space limitations, we select the most recent and superior methods FECNet [14], SNR-SKF [42], FourLLIE [37], UHDFour [20] and Retinexformer [2] for an intuitive comparison with our DMFourLLIE. Fig. 6 shows the comparison results on LSRW-Huawei, LOL-v2-real, and LOL-v2-Synthesis datasets. In LSRW-Huawei, FourLLIE causes over-brightening leading to image overexposure, while SNR-SKF and Retinexformer do not achieve fully brightened visualization results.

	FourLLIE	FourLLIE$_{s1}$ + DMFourLLIE$_{s2}$	DMFourLLIE
PSNR ↑	24.65	24.94 (↑ 0.29)	**25.83**
SSIM ↑	0.9192	0.9269 (↑ 0.0077)	**0.9314**
LPIPS ↓	0.0389	0.0235 (↓ 0.0154)	**0.0234**

Table 5: Objective validation of effectiveness by replacing the second stage of FourLLIE with our DMFourLLIE.

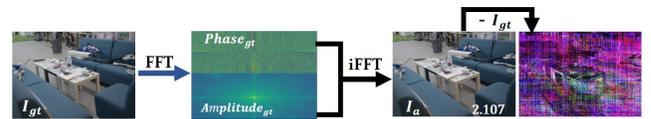**Figure 8: Stage-wise performance comparison.**

Figure 9: Fourier transform loss. I_a number is average error with GT.

UHDFour exhibits an issue of insufficient brightness. FECNet, even with better brightening effects, lacks the clarity in motorcycle detail and overall image color (especially the floor) compared to our results, which are closer to ground truth (GT). In LOL-v2-real, FourLLIE and Retinexformer evidently introduce unavoidable blotchiness, greatly affecting the visualization results. FECNet and SNR-SKF show noticeable noise issues, whereas our method produces the clearest and most natural brightening results. In LOL-v2-synthesis, FECNet, SNR-SKF, FourLLIE, and Retinexformer suffer from insufficient brightness or overexposure issues, as well as some degree of color deviation. UHDFour introduces noise leading to blurry images, while our method achieves visually similar results to the GT.

Quantitative results on LIME, VV, DICM, NPE and MEF datasets. We measure the no-reference Natural Image Quality Evaluator (NIQE [28]) scores on five non-paired datasets. Lower NIQE scores indicate images with higher naturalness quality. Tab. 2 shows the NIQE evaluation results. It can be seen that our DMFourLLIE outperforms most of the existing LLIE methods.

4.3 Effectiveness of the second stage

We elaborate on the effectiveness of the second stage from three perspectives: (1) As shown qualitatively in Fig.8 and quantitatively through mean square error metrics, the second stage demonstrates its effectiveness and outperforms other methods; (2) By replacing the second stage of FourLLIE with the second stage of our DM-FourLLIE, as reflected in Tab.5, all performance metrics indicate improvements compared to the FourLLIE baseline; (3) Due to the inevitable information loss introduced during Fourier transform and its inverse in the first stage (refer to Fig.9, owing to discretization), and after multiple transformations and network inferences, the refinement provided by the second stage is essential.

MM '24, October 28–November 1, 2024, Melbourne, VIC, Australia.

Tongshun Zhang, Pingping Liu, Ming Zhao and Haotian Lv

	Infrared Image (StawGAN [33])	Infrared Image (PearlGAN [24])	Depth Map [46]	Edge Map [45]	DMFourLLIE
PSNR ↑	25.17	25.32	24.97	25.15	25.83
SSIM ↑	0.9287	0.9307	0.9256	0.9275	0.9317
LPIPS ↓	0.0265	0.0225	0.0238	0.0245	0.0234

Table 6: The impact of infrared image , depth map , and edge map on model performance was investigated. Comparative analysis revealed that infrared images achieve superior performance.

Metrics	FourLLIE	UHDFour	DMFourLLIE
UICM ↑	0.9133	0.7464	**0.9426**
NIQE ↓	3.143	5.385	**3.047**

Table 7: Visualization quality comparison on Dark Face.

4.4 Ablation Studies

Ablation experiments on components: In Tab. 3 and Fig. 7, we conducted quantitative and qualitative ablation studies on each component of DMTourLLIE to assess their effectiveness. By controlling a single variable, we verified the effectiveness of each component and the overall network design.

Ablation experiments on loss function: To further analyze the impact of loss functions on model performance, we conducted an ablation study. As shown in Tab. 4, following the principle of controlling a single variable, we observed the contributions of different loss function configurations to model performance. The results indicate that among all compared configurations, DMFourLLIE achieved the best performance.

Moreover, to further validate the limitations in the design of the first-stage loss functions in FECNet and FourLLIE, we present the results of adding $\mathcal{L}_{amp}$ in the last column of Tab. 4 (following the design of FECNet and FourLLIE, where $\mathcal{L}_{amp}$ is defined as the difference between the enhanced amplitude component A_{aug} and the ground truth amplitude component A_{gt}). We found that the performance actually decreases after adding $\mathcal{L}_{amp}$, which again supports our motivational hypothesis that the strategy of solely boosting the amplitude component leads to using degraded images as the learning target for the network.

4.5 Impact of different pre-trained model

We examined the effects of different models on infrared information extraction, and the impact of incorporating depth and edge maps. In Tab.6, we provide a comprehensive analysis of the impact of different models on infrared information extraction, including the use of depth maps and edge maps. It is well-known that infrared images have significant advantages in extremely low-light conditions. In contrast, depth maps and edge maps are often susceptible to interference from factors such as lighting and color noise. This validates our motivation for using infrared images to enhance the structural integrity of the phase component. The infrared extraction model employed by DMFourLLIE achieves optimal overall performance.

4.6 Object Detection and Visualization

On the Dark Face dataset [48], we conducted object detection experiments under low-light conditions to evaluate the preprocessing

Figure 10: Detection comparison results on the Dark Face dataset.

effects of the latest Fourier-based methods on advanced visual understanding tasks. The Dark Face dataset consists of 6000 low-light images with real-world annotations. For testing, we randomly selected 200 images from the dataset and validated using the official YOLOv5 model pre-trained on the COCO [23] dataset. Fig. 10 presents the visual comparison results, where our method outperforms other methods significantly in terms of detection accuracy and recall rate. Notably, in the results of the second row, DMFourLLIE successfully detects the backpack, while in the fourth row, the detection rate of bicycles is significantly higher than other methods, with other methods showing instances of false detections.

Furthermore, through visual analysis, it is evident that DMFourLLIE surpasses other methods in terms of brightness enhancement, naturalness, and clarity. This is particularly noticeable when observing the faces in the second and fifth rows. In contrast, both FourLLIE and UHDFour methods exhibit blurriness and noise issues. Therefore, as shown in Tab. 7, we also compared image quality metrics (NIQE [28] and UICM [30]) on the Dark Face dataset to further support the advantages of our method.

5 Conclusion

In this paper, we introduce the Dual-Stage Multi-Branch Fourier Low-Light Image Enhancement (DMFourLLIE) framework, a novel structure designed to enhance the expressiveness of Fourier frequency domain information and improve the spatial and textural details of images. The first stage of our framework utilizes a distinctive multi-branch structure that integrates infrared and brightness priors, thereby refining the expressiveness and accuracy of frequency domain information. This integration significantly bolsters the phase component's capacity to preserve image structure while enabling precise amplification of the amplitude component, tailored to regional luminance variations. In the second stage, DMFourLLIE employs a multi-scale spatial perception module in conjunction with the innovative application of fast Fourier convolution. This dual-branch approach markedly enhances the representation of spatial structures and the delineation of subtle texture details in the enhanced images, leading to superior image quality.

DMFourLLIE: Dual-Stage and Multi-Branch Fourier Network for Low-Light Image Enhancement

MM '24, October 28–November 1, 2024, Melbourne, VIC, Australia.

References

[1] Tarik Arici, Salih Dikbas, and Yucel Altunbasak. 2009. A histogram modification framework and its application for image contrast enhancement. *IEEE TIP* 18, 9 (2009), 1921–1935.

[2] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. 2023. Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In *ICCV*.

[3] Bing Cao, Yiming Sun, Pengfei Zhu, and Qinghua Hu. 2023. Multi-Modal Gated Mixture of Local-to-Global Experts for Dynamic Image Fusion. In *ICCV*. 23555–23564.

[4] Lu Chi, Borui Jiang, and Yadong Mu. 2020. Fast fourier convolution. *Advances in Neural Information Processing Systems* 33 (2020), 4479–4488.

[5] Qiaole Dong, Chenjie Cao, and Yanwei Fu. 2022. Incremental transformer structure enhanced image inpainting with masking positional encoding. In *CVPR*. 11358–11368.

[6] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. 2016. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*. 2782–2790.

[7] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. 2023. Learning a Simple Low-Light Image Enhancer From Paired Low-Light Instances. In *CVPR*. 22252–22261.

[8] Dario Fuoli, Luc Van Gool, and Radu Timofte. 2021. Fourier space losses for efficient perceptual image super-resolution. In *ICCV*. 2360–2369.

[9] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*. 1780–1789.

[10] Xiaojie Guo and Qiming Hu. 2023. Low-light image enhancement via breaking down the darkness. *IJCV* 131, 1 (2023), 48–66.

[11] Xiaojie Guo, Yu Li, and Haibin Ling. 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE TIP* 26, 2 (2016), 982–993.

[12] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. 2023. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation* 90 (2023), 103712.

[13] Khurram Azeem Hashmi, Goutham Kallempudi, Didier Stricker, and Muhammad Zeshan Afzal. 2023. FeatEnHancer: Enhancing Hierarchical Features for Object Detection and Beyond Under Low-Light Vision. In *CVPR*. 6725–6735.

[14] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. 2022. Deep fourier-based exposure correction network with spatial-frequency interaction. In *ECCV*. Springer, 163–180.

[15] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. 2018. Multimodal Unsupervised Image-to-image Translation. In *ECCV*.

[16] Vladimir V Kniaz, Vladimir A Knyaz, Jiri Hladuvka, Walter G Kropatsch, and Vladimir Mizginov. 2018. Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset. In *ECCV*. 0–0.

[17] Chulwoo Lee, Chul Lee, and Chang-Su Kim. 2012. Contrast enhancement based on layered difference representation. In *ICIP*. IEEE, 965–968.

[18] Chulwoo Lee, Chul Lee, and Chang-Su Kim. 2013. Contrast enhancement based on layered difference representation of 2D histograms. *IEEE TIP* 22, 12 (2013), 5372–5384.

[19] Dong-Guw Lee, Myung-Hwan Jeon, Younggun Cho, and Ayoung Kim. 2023. Edge-guided multi-domain rgb-to-tir image translation for training vision tasks with challenging labels. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 8291–8298.

[20] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. 2023. EmbeddingFourier for Ultra-High-Definition Low-Light Image Enhancement. In *ICLR*.

[21] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. 2018. Structure-revealing low-light image enhancement via robust retinex model. *IEEE TIP* 27, 6 (2018), 2828–2841.

[22] Yudong Liang, Bin Wang, Wenqi Ren, Jiaying Liu, Wenjian Wang, and Wangmeng Zuo. 2022. Learning hierarchical dynamics with spatial adjacency for image enhancement. In *ACM MM*. 2767–2776.

[23] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *ECCV*. Springer, 740–755.

[24] Fuya Luo, Yunhan Li, Guang Zeng, Peng Peng, Gang Wang, and Yongjie Li. 2022. Thermal infrared image colorization for nighttime driving scenes with top-down guided attention. *IEEE Transactions on Intelligent Transportation Systems* 23, 9 (2022), 15808–15823.

[25] Feifan Lv, Bo Liu, and Feng Lu. 2020. Fast enhancement for non-uniform illumination images using light-weight CNNs. In *ACM MM*. 1450–1458.

[26] Kede Ma, Kai Zeng, and Zhou Wang. 2015. Perceptual quality assessment for multi-exposure image fusion. *IEEE TIP* 24, 11 (2015), 3345–3356.

[27] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. 2022. Toward fast, flexible, and robust low-light image enhancement. In *CVPR*. 5637–5646.

[28] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. 2013. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Processing Letters* 20, 3 (2013), 209–212. https://doi.org/10.1109/LSP.2012.2227726

[29] Adam Nyberg, Abdelrahman Eldesokey, David Bergstrom, and David Gustafsson. 2018. Unpaired thermal to visible spectrum transfer using adversarial training. In *ECCV*. 0–0.

[30] Karen Panetta, Chen Gao, and Sos Agaian. 2016. Human-Visual-System-Inspired Underwater Image Quality Measures. *IEEE Journal of Oceanic Engineering* 41, 3 (2016), 541–551. https://doi.org/10.1109/JOE.2015.2469915

[31] Gyu-Hee Park, Hwa-Hyun Cho, and Myung-Ryul Choi. 2008. A contrast enhancement method using dynamic range separate histogram equalization. *IEEE Transactions on Consumer Electronics* 54, 4 (2008), 1981–1987.

[32] Stephan R. Richter, Zeeshan Hayder, and Vladlen Koltun. 2017. Playing for Benchmarks. In *ICCV*.

[33] Luigi Sigillo, Eleonora Grassucci, and Danilo Comminiello. 2023. StawGAN: Structural-aware generative adversarial networks for infrared image translation. In *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 1–5.

[34] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[35] Tao Sun, Mattia Segu, Janis Postels, Yuxuan Wang, Luc Van Gool, Bernt Schiele, Federico Tombari, and Fisher Yu. 2022. SHIFT: a synthetic driving dataset for continuous multi-task domain adaptation. In *CVPR*. 21371–21382.

[36] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. 2022. Resolution-robust large mask inpainting with fourier convolutions. In *WACV*. 2149–2159.

[37] Chenxi Wang, Hongujun Wu, and Jin Zhi. 2023. FourLLIE: Boosting Low-Light Image Enhancement by Fourier Frequency Information. In *ACM MM*.

[38] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE TIP* 22, 9 (2013), 3538–3548.

[39] Yufei Wang, Yi Yu, Wenhan Yang, Lanqing Guo, Lap-Pui Chau, Alex C Kot, and Bihan Wen. 2023. Exposurediffusion: Learning to expose for low-light image enhancement. In *ICCV*. 12438–12448.

[40] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP* 13, 4 (2004), 600–612.

[41] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2018. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560* (2018).

[42] Yuhui Wu, Chen Pan, Guoqing Wang, Yang Yang, Jiwei Wei, Chongyi Li, and Heng Tao Shen. 2023. Learning Semantic-Aware Knowledge Guidance for Low-Light Image Enhancement. In *CVPR*. 1662–1671.

[43] Jun Xu, Yingkun Hou, Dongwei Ren, Li Liu, Fan Zhu, Mengyang Yu, Haoqian Wang, and Ling Shao. 2020. Star: A structure and texture aware retinex model. *IEEE TIP* 29 (2020), 5022–5037.

[44] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. 2022. SNR-aware low-light image enhancement. In *CVPR*. 17714–17724.

[45] Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. 2023. Low-light image enhancement via structure modeling and guidance. In *CVPR*. 9893–9903.

[46] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. 2024. Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. In *CVPR*.

[47] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. 2021. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE TIP* 30 (2021), 2072–2086.

[48] Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, et al. 2020. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE TIP* 29 (2020), 5737–5752.

[49] Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. 2022. Frequency and spatial dual guidance for image dehazing. In *ECCV*. Springer, 181–198.

[50] Seungsang Yun, Minwoo Jung, Jeongyun Kim, Sangwoo Jung, Younghun Cho, Myung-Hwan Jeon, Giseop Kim, and Ayoung Kim. 2022. STheReO: Stereo Thermal Dataset for Research in Odometry and Mapping. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 3857–3864. https://doi.org/10.1109/IROS47612.2022.9981857

[51] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*. 5728–5739.

[52] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2020. Learning enriched features for real image restoration and enhancement. In *ECCV*. Springer, 492–511.

[53] Fan Zhang, Yu Li, Shaodi You, and Ying Fu. 2021. Learning temporal consistency for low light video enhancement from single images. In *CVPR*. 4967–4976.

[54] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*. 586–595.

MM '24, October 28–November 1, 2024, Melbourne, VIC, Australia.

Tongshun Zhang, Pingping Liu, Ming Zhao and Haotian Lv

[55] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. 2021. Beyond brightening low-light images. *IJCV* 129 (2021), 1013–1037.

[56] Yukang Zhang and Hanzi Wang. 2023. Diverse Embedding Expansion Network and Low-Light Cross-Modality Benchmark for Visible-Infrared Person Re-identification. In *CVPR*. 2153–2162.

[57] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. 2019. Kindling the darkness: A practical low-light image enhancer. In *ACM MM*. 1632–1640.

[58] Wenda Zhao, Shigeng Xie, Fan Zhao, You He, and Huchuan Lu. 2023. MetaFusion: Infrared and Visible Image Fusion via Meta-Feature Embedding From Object Detection. In *CVPR*. 13955–13965.

[59] Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Shuang Xu, Zudi Lin, Radu Timofte, and Luc Van Gool. 2023. Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion. In *CVPR*. 5906–5916.

[60] Chuanjun Zheng, Daming Shi, and Wentian Shi. 2021. Adaptive unfolding total variation network for low-light image enhancement. In *ICCV*. 4439–4448.

[61] Naishan Zheng, Man Zhou, Yanmeng Dong, Xiangyu Rui, Jie Huang, Chongyi Li, and Feng Zhao. 2023. Empowering Low-Light Image Enhancer through Customized Learnable Priors. In *ICCV*. 12559–12569.

[62] Man Zhou, Jie Huang, Keyu Yan, Hu Yu, Xueyang Fu, Aiping Liu, Xian Wei, and Feng Zhao. 2022. Spatial-frequency domain information integration for pan-sharpening. In *ECCV*. Springer, 274–291.