Spatial-frequency Dual-Domain Feature Fusion Network for Low-Light Remote Sensing Image Enhancement

Zishu Yao, Guodong Fan, Jinfu Fan, Min Gan, *Senior Member, IEEE*, and C. L. Philip Chen, *Fellow, IEEE*

***Abstract*—Low-light remote sensing images generally feature high resolution and high spatial complexity, with continuously distributed surface features in space. This continuity in scenes leads to extensive long-range correlations in spatial domains within remote sensing images. Convolutional Neural Networks, which rely on local correlations for long-distance modeling, struggle to establish long-range correlations in such images. On the other hand, transformer-based methods that focus on global information face high computational complexities when processing high-resolution remote sensing images. From another perspective, Fourier transform can compute global information without introducing a large number of parameters, enabling the network to more efficiently capture the overall image structure and establish long-range correlations. Therefore, we propose a Dual-Domain Feature Fusion Network (DFFN) for low-light remote sensing image enhancement. Specifically, this challenging task of low-light enhancement is divided into two more manageable sub-tasks: the first phase learns amplitude information to restore image brightness, and the second phase learns phase information to refine details. To facilitate information exchange between the two phases, we designed an information fusion affine block that combines data from different phases and scales. Additionally, we have constructed two dark light remote sensing datasets to address the current lack of datasets in dark light remote sensing image enhancement. Extensive evaluations show that our method outperforms existing state-of-the-art methods. The code is available at https://github.com/iijjlk/DFFN.**

I. Introduction

Remote sensing (RS) technology aims to capture clear ground scene information. Compared to ordinary images, remote sensing images display richer color information and broader spatial information. However, due to nighttime conditions and other inherent natural factors, images captured in low-light environments often suffer from color degradation and loss of detail. This inevitably affects many downstream tasks [1], [2] that rely on remote sensing images, such as disaster recovery, wildlife monitoring, and environmental protection. Therefore, the development of algorithms dedicated to enhancing low-light remote sensing images is crucial.

This work was supported in part by the National Natural Science Foundation of China under Grant 62073082, Grant 71701136, Grant U1813203, and Grant U1801262; in part by the Taishan Scholar Program of Shandong Province.

Zishu Yao, Guodong Fan, Jinfu Fan and Min Gan are with the College of Computer Science & Technology, Qingdao University, Qingdao 266071, China (e-mail: yaozishu@qdu.edu.cn; fgd96@outlook.com; aganmin@aliyun.com; fan_jinfu@163.com).

C. L. Philip Chen is with the College of Computer Science & Technology, Qingdao University, Qingdao 266071, China and the School of Computer Science & Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: philip.chen@ieee.org).

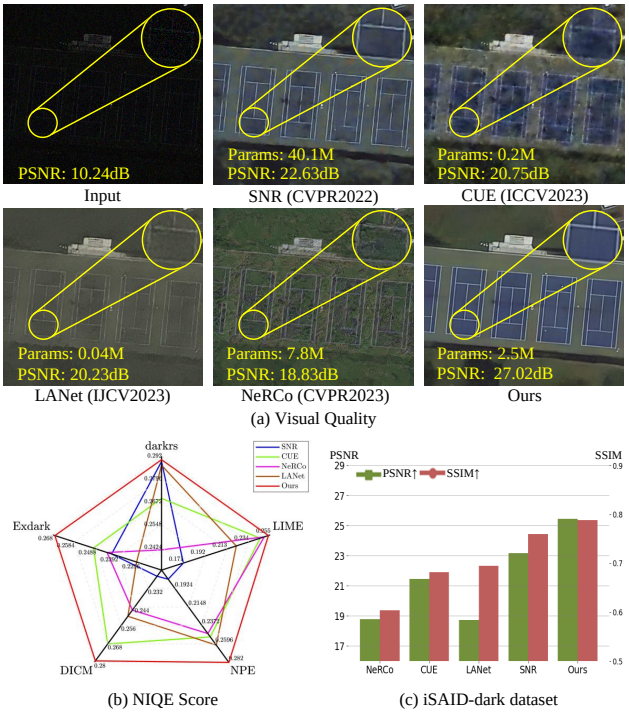


Fig. 1. Comparison between the latest state-of-the-art methods and our approach. As shown in the zoomed-in region, these methods exhibit color distortion and excessive noise, leading to reduced visual quality. In contrast, our method presents vibrant colors and sharp outlines. Transformer-based method SNR [13] achieves similar visual effects to ours but with 40.1M parameters. Additionally, in (b), we compute the NIQE [14] scores on a no-reference dataset and display them in inverse form. Furthermore, in (c), scores of SSIM [15] and PSNR [16] metrics on a reference dataset are shown. It can be easily observed that our method significantly outperforms others.

Compared to images in traditional Low-light image enhancement tasks [3]–[9], RS images have larger lengths and widths. Some surface features are often continuously distributed in space, such as forests covering several square kilometers or roads traversing the entire image. The continuity of such scenes results in remote sensing images exhibiting long-distance correlations in spatial aggregation. Convolutional Neural Networks (CNNs) [10]–[12] rely on local perception. For high-resolution images, the receptive field of a single convolutional operation struggles to capture larger spatial structures or global information in the image. Although deep convolutions can expand the receptive field, as the network depth increases, the loss of detailed information gradually occurs, making it challenging for models to establish long

correlations in RS images. Recently, a number of transformer methods [17]–[21], with Vision Transformer [22] as a prominent example, have successfully addressed the challenge of establishing long-range dependencies by implementation of self-attention mechanisms. However, the quadratic complexity problem remains difficult to solve. Additionally, these methods tend to overfit when applied to small datasets, making practical application challenging. Fig. 1 (a) presents visual results and number of parameters for several representative methods. It can be observed that SNR, built upon the transformer, achieves suboptimal visual results, yet concurrently bears an onerous computational burden.

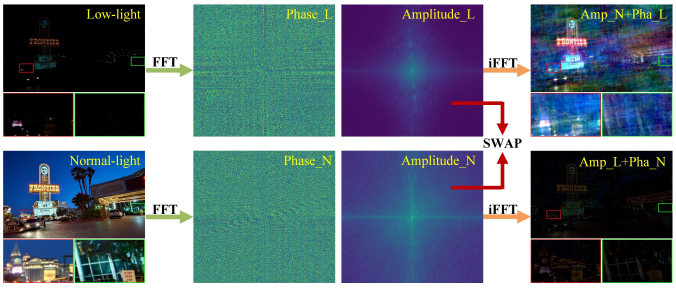


Fig. 2. Our Motivation. As observed in the figure, when swapping the amplitude and phase of low-light images with those of normal images, the overall ambient light of the image containing degraded amplitude darkens while maintaining clear details. On the other hand, in the image containing phase degradation, there is no change in brightness, but the image details vanish. This indicates that Fourier transform can decouple degradation information.

Most existing methodss [23]–[25] process images in the spatial domain based on a pixel perspective. In contrast, we shift our focus to the Fourier domain, motivated by two unique phenomena inherent in the Fourier domain: 1) Under Fourier transformation, global information of the image can be extracted without introducing a large number of parameters [26]. 2) It is well-known that Fourier phase preserves high-level semantics, while amplitude contains low-level features [27]. Fig. 2 displays the visualization image of exchanging the Fourier amplitude and phase between normal and low-light images. It can be observed that brightness degradation exists in the amplitude, while other information is preserved in the phase. This phenomenon suggests that the Fourier transform can decouple the degradation of low-light images, breaking down the complex coupling problem into two relatively easier-to-solve sub-problems. Existing methods have explored the frequency domain information, but still present many shortcomings. For example, Wang et al. [26] opts to simultaneously recover both amplitude and phase information without considering spatial information. Li et al. [28] only performs Fourier decomposition on images within the network, yet the loss function continues to make comparisons with the ground truth (GT) solely in the spatial domain, lacking effective supervision. This undoubtedly makes the recovery process more difficult. Therefore, the challenge of how to integrate frequency domain information with spatial features and efficiently apply this integration for the enhancement of remote sensing low-light images remains unresolved.

In this work, based on two unique phenomena of the Fourier transform, we propose a Dual-domain Feature Fusion Network (DFFN). Specifically, the DFFN consists of two-stage transformations to progressively restore high-quality images. Initially, the amplitude illumination phase learns the amplitude mapping from low-light to normal lighting conditions to brighten the image. Subsequently, the phase refinement phase enhances image details further. To integrate frequency domain and spatial features, we designed the Dual-Domain Amplitude Block (DDAB) for the amplitude illumination phase and the Dual-Domain Phase Block (DDPB) for the phase refinement phase. By splitting the Low-light Remote Image Enhancement task into two sub-problems, we reduce the complexity of degradation coupling.

Meanwhile, to improve information interaction capability between the two-stage networks, we designed an Information Fusion Affine Module (IFAM) that integrates rich feature information from different stages and scales, achieving adaptive learning through dynamic fusion with affine filtering. The obtained feature information is convolved with the decoder features of the phase refinement stage to enhance the network representation capability.

In addition, due to the lack of available image pairs, the development of existing remote sensing low-light enhancement tasks has been slow. In this paper, we constructed two large datasets, with the paired dataset iSAID-dark used for training and the real dataset darkrs used for testing.

In summary, our work contributes the following:

- We propose a Dual-domain Feature Fusion Network for low-light remote sensing image enhancement, which decomposes the degradation problem into two relatively easier-to-solve sub-problems through a two-stage network.
- We design an Information Affine Fusion Module to facilitate information interaction across scales, stages, and domains within the two-stage network. This module promotes information fusion between the two stages and enhances the representation of global context.
- We construct two low-light remote sensing datasets, including a real dataset captured by UAVs at night and a large-scale synthetic dataset. These datasets provide paired data for low-light remote sensing image enhancement and real low-light scenes for testing purposes.
- We conduct extensive experiments on multiple public benchmarks for comprehensive evaluation, demonstrating that our method performs well compared to state-of-the-art methods. Additionally, our approach shows excellent trade-offs between model complexity and performance.

II. Related Work

A. Low-light image enhancement

In recent years, Low-light image enhancement has developed rapidly, which can mainly be classified into three categories: methods based on value mapping, model-based methods, and data-driven methods.

Value mapping-based methods enhance image exposure by altering the pixel distribution through mapping relationships. Notably, such methods include histogram equalization [29]–[31] and curve transformations, such as gamma transformation

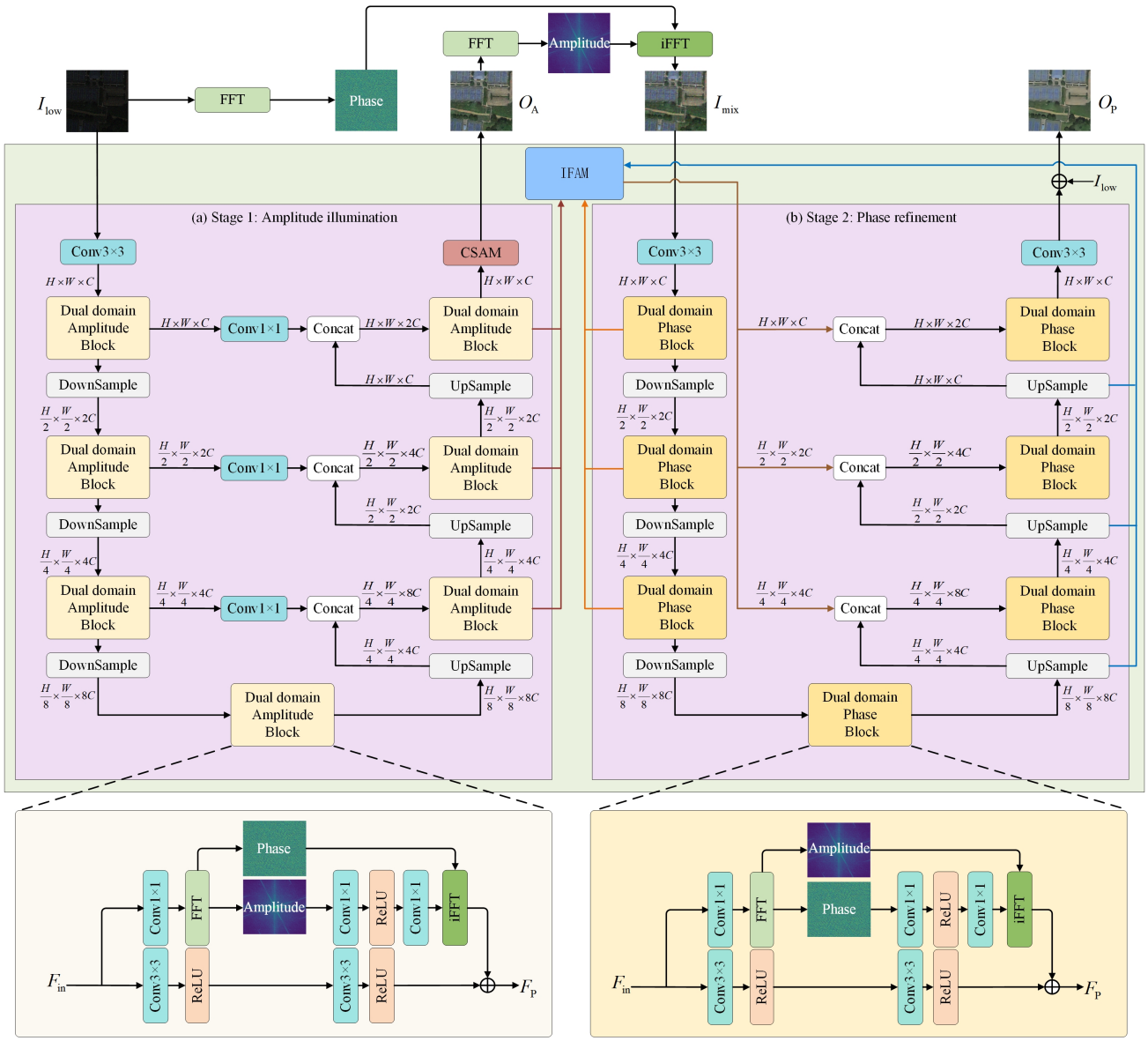


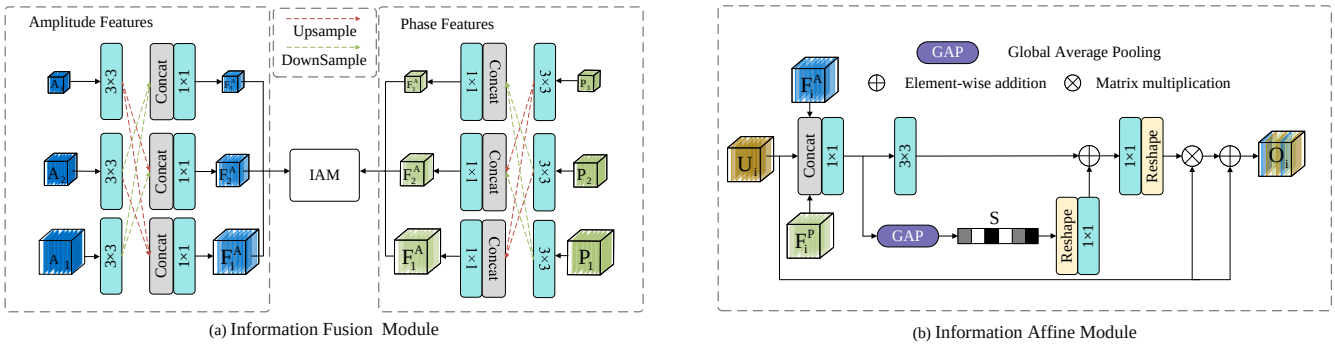
Fig. 3. Overall network structure diagram. The specific structures of DDAB and DDPB are displayed below the diagram. The low-light image I_{low} first undergoes an amplitude illumination stage to obtain a preliminary enhanced image O_A, followed by a phase refinement stage to obtain the final enhanced image O_P. In this process, the feature information from different stages is passed into IFAM for mapping to enhance the contextual representation capability of the model.

[32], [33]. However, due to the lack of consideration for pixel context information, these methods are prone to causing local illumination inconsistencies and color abnormalities.

Model-based methods [34]–[36], grounded in Retinex theory, view images as the product of reflectance and irradiance. By dividing low-light images by the irradiance image, the resulting reflectance is used as the final enhanced image. For example, Li et al. [37], building upon traditional Retinex theory and taking noise maps into account, successfully developed a noise-robust Retinex model. Guo et al. [38] estimated the illumination for each pixel using the maximum channel prior and added a prior structure on top of it to refine the illumination. The crux of such methods lies in designing well-crafted regularization functions. However, these regularization functions are based on prior assumptions, which, when not valid, greatly diminish the effectiveness of these methods.

Data-driven approaches [39]–[41] are based on large-scale datasets and enhance network performance through the design of various complex model structures. For instance, Ma et al. [42] proposed a context-sensitive decomposition network that focuses on the contextual dependencies in the feature space, estimating lighting and reflections at the feature level to ultimately achieve high-quality enhancement effects. Li et al. [43] designed a pixel-level curve estimation method and developed a series of loss functions that, while ensuring high-speed computation, yielded excellent results. Wang and colleagues [44] combined gamma correction with transformers, establishing dependencies for all pixels from local to global, efficiently recovering dark areas.

Although existing low-light image enhancement methods



(a) Information Fusion Module (b) Information Affine Module

Fig. 4. The specific structure of the proposed IFAM consists of two key components: the Information Fusion Module and the Information Mapping Module.

have achieved remarkable enhancement effects, due to the lack of corresponding paired datasets and the rarity of methods specifically designed for remote sensing images, current mainstream methods struggle to achieve ideal results in the task of enhancing dark light remote sensing images. Recently, Ling et al. [45] proposed an unpaired training method for low-light remote sensing image enhancement, which maximizes the mutual information between dark and normal light images in a generative adversarial network through self-similarity based contrastive learning. However, due to limited supervision, the final enhancement effect is not satisfactory. To facilitate the development of this task, we propose a large-scale paired dataset for dark light remote sensing image enhancement to provide effective supervision.

B. Fourier Transform in Neural Networks

Recently, due to the unique properties of images in the Fourier space, such as capturing global representations, Fourier frequency information [26], [28] has attracted widespread attention. A series of works have been based on Fourier domain transformations to enhance the generalization performance of networks. For instance, Mao et al. [46] designed a residual Fourier block to address issues like the potential neglect of low-frequency information and the difficulty in modeling long-distance information in Resblocks. Roman et al. [47] used fast Fourier convolution to design a large mask inpainting network, achieving impressive results even in challenging scenarios. Dario et al. [48] employed Fourier loss to aid networks in recovering high-frequency information in super-resolution images. While these works explore the application of Fourier information in various scenarios, few have considered the relationship between frequency domain information and remote sensing images. In contrast, our proposes DFFN effectively decouples dark light degradation, achieving excellent enhancement results.

III. Method

In this section, we first introduced the Fourier information transformation. Subsequently, we described the overall architecture of the DFFN and the basic units constituting the network. Finally, we elaborated on the IFAM and the loss functions used in the DFFN.

A. Fourier Frequency Information

In this subsection, we will briefly introduce Fourier frequency information. Specifically, consider an image $I \in \mathbb{R}^{H\times W\times 1}$, whose Fourier transformation function $\mathcal{F}$ can be represented as:

$$\mathcal{F}(X)(i,j) = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} X(h,w) e^{-j2\pi(\frac{h}{H}i + \frac{w}{W}j)}, \quad (1)$$

where k represents the imaginary unit, i and j represent the horizontal and vertical coordinates respectively, h and w represent spatial coordinates, and the inverse process of $\mathcal{F}$ is denoted as $\mathcal{F}^{-1}$.

Correspondingly, the amplitude $\mathcal{A}(I)$ and phase $\mathcal{P}(I)$ information can be defined as:

$$\begin{cases} \mathcal{A}(I)(i,j) = \sqrt{R^2(X)(i,j) + I^2(X)(i,j)}, \\ \mathcal{P}(I)(i,j) = \arctan\left[\dfrac{I(X)(i,j)}{R(X)(i,j)}\right], \end{cases} \quad (2)$$

where $I(X)(i,j)$ and $R(X)(i,j)$ denote the imaginary and real component of $F(X)(i,j)$. For feature maps or color images, Fourier calculations are computed separately for each channel. The $R(X)(i,j)$ and $I(X)(i,j)$ can also be obtained by:

$$\begin{cases} R(X)(i,j) = \mathcal{A}(X)(i,j) \times \cos(\mathcal{P}(X)(i,j)), \\ I(X)(i,j) = \mathcal{A}(X)(i,j) \times \sin(\mathcal{P}(X)(i,j)). \end{cases} \quad (3)$$

B. Network Architecture

The structure of DFFN is shown in Fig. 3. Specifically, in the amplitude illumination stage, image brightness is restored by adjusting the amplitude, while in the phase refinement stage, phase information is learned to refine fine details. Since the two stages have different subtasks, their input information and supervision targets should also differ. The low-light image I_{low} as the input for the first stage. The supervision target is defined as $\mathcal{F}^{-1}(\mathcal{A}(I_{gt}), \mathcal{P}(I_{low}))$. To prevent the first stage network from affecting the phase information, the output O_A is not directly used as the input for the second stage. Instead, let $I_{mix} = \mathcal{F}^{-1}(\mathcal{A}(O_A), \mathcal{P}(I_{low}))$, and the supervision target is I_{gt}.

Although our primary motivation lies within the Fourier domain, the utilization of spatial branches is necessary. This

is because spatial branches and Fourier branches complement each other. Spatial branches, employing convolution operations, effectively model structural dependencies in the spatial domain. Fourier branches can engage in global information, facilitating the unraveling of energy and degradation. Therefore, for more efficient information recovery, we introduce DDAB and DDPB.

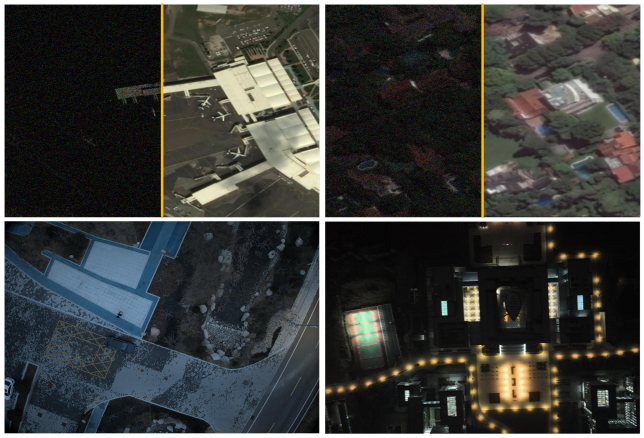


Fig. 5. Samples from the proposed iSAID-dark(Up) and darkrs(Down) dataset.

1) Dual-Domain Amplitude Block: As shown in the bottom-left of Fig. 3, The DDAB consists of two parallel branches: spatial domain branch and frequency domain branch. The spatial domain branch has two 3×3 Conv layers to capture spatial information. In the frequency domain branch, the feature stream F_{in} first undergoes 1×1 Conv to refine the features. Then, FFT is applied to obtain the amplitude component $\mathcal{A}(F_{in})$ and the phase component $\mathcal{P}(F_{in})$. The amplitude component $\mathcal{A}(F_{in})$ is passed through a Conv layer with two 1×1 kernels (to avoid damaging the frequency domain structure, only 1×1 Conv is used here). Meanwhile, the corresponding phase information $\mathcal{P}(F_{in})$ is preserved through a separate branch. Subsequently, we use the iFFT to map $\mathcal{A}(F_{in})$ and $\mathcal{P}(F_{in})$ back to the image space to obtain frequency domain features. Finally, we integrate the spatial and frequency domain features through residual connections. The process can be described as follows:

$$\begin{cases} F_{sa} = \sigma_{3\times3}\left[\sigma_{3\times3}\left(F_{in}\right)\right], \\ \mathcal{A}(F_{in}), \mathcal{P}(F_{in}) = \mathcal{F}[\Theta_{1\times1}\left(F_{in}\right)], \\ F_A' = \mathcal{F}^{-1}\left\langle \Theta_{1\times1}\left\{\sigma_{1\times1}\left[\mathcal{A}(F_{in})\right]\right\}, \mathcal{P}(F_{in})\right\rangle, \\ F_A = F_{sa} + F_A', \end{cases} \tag{4}$$

where $\sigma(\cdot)$ represents Conv and ReLU operation, $\mho(\cdot)$ represents concatenation operation, and $\Theta(\cdot)$ represents Conv operation.

2) Dual-Domain Phase Block: Compared to DDAB, DDPB simply swaps the operations of amplitude and phase. It can be described as:

$$F_P^{'} = \mathcal{F}^{-1}\left\langle \Theta_{1\times1}\left\{\sigma_{1\times1}\left[\mathcal{P}(F_{in})\right]\right\}, \mathcal{A}(F_{in})\right\rangle. \tag{5}$$

In summary, the amplitude illumination stage utilizes DDAB as the basic unit and employs an encoder-decoder structure to extract multi-scale feature information. Similarly, the phase refinement stage adopts DDPB as the basic unit. Skip connections are established between the corresponding encoder and decoder at each scale of the amplitude illumination stage. In contrast, the phase refinement stage employs IFAM to map features from the previous stage and the current encoder features more efficiently for fusing information between the two stages. Furthermore, to optimize the cross-stage network, a cross-stage feature fusion and supervised attention module (CSAM) [49] is introduced at the end of the amplitude illumination stage.

Next, we will provide a detailed explanation of the proposed IFAM module and Loss function.

C. Information Fusion Affine Module

Multi-stage networks [4], [26], [48] enhance performance by decomposing complex problems into simpler sub-problems and solving them separately. In recent years, various works have achieved significant results using this approach. However, improving the ability of cross-stage information interaction remains a major challenge for multi-stage networks. As the depth of the network increases, shallow features tend to be lost gradually. Simply passing the output from the previous stage to the next stage may result in the loss of a significant amount of potential features. Another mainstream approach is to transfer information between stages of the same size, but this lacks interaction between features of different scales.

Therefore, our designed IFAM first aggregates feature information from different stages and scales. Subsequently, it performs adaptive fusion on the aggregated features to enhance global contextual representation. Specifically, IFAM (as shown in Fig. 4) comprises two sub-modules:

1) Information Fusion Module: IAM integrates all scale information from the first-stage decoder and second-stage encoder, and merges information from different scales. As shown in Fig. 4 (a), the decoder features $\{A_i\}_{i=1}^{3}$ first undergo a 3×3 Conv, followed by combining different scale features through upsample and downsample operations. Then, 1×1 Conv is used to merge these features with different receptive fields, obtaining rich contextual information. The encoder features $\{P_i\}_{i=1}^{3}$ undergo similar operations. Subsequently, the fused features are passed into IAM.

2) Information Affine Module: As shown in Fig. 4 (b), taking the feature information from the i-th layer as an example, three features of the same size from different encoders and decoders are first fused through a 1×1 Conv. Subsequently, they pass through separate 3×3 Conv to obtain spatial context information, and through a pooling layer and 1×1 Conv to obtain channel context information. Finally, the two streams of information are added element-wise to fully utilize their information.

$$\begin{cases} F_{fuse} = \Theta\left[\mho(A_i, P_i, U_i)\right], \\ F_{fuse}^{'} = \Theta(F_{fuse}) + \text{Pool}\left\{\text{Reshape}\left[\Theta(F_{fuse})\right]\right\}, \end{cases} \tag{6}$$

where U_i represents the decoder features from the second stage.

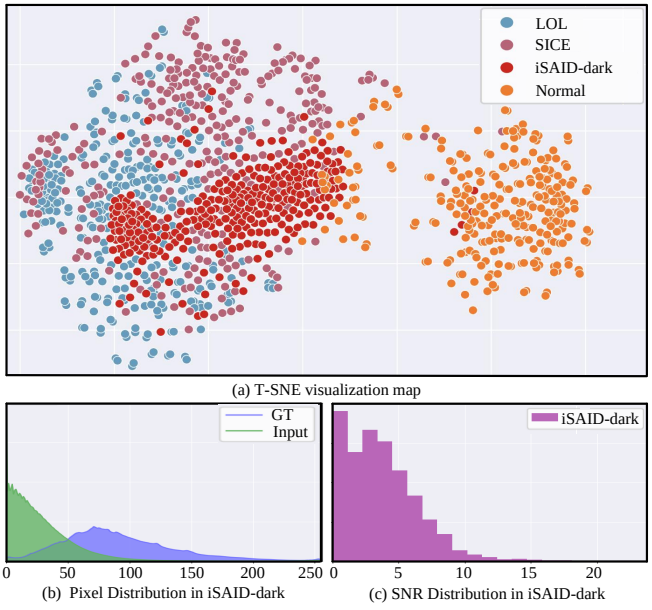
TABLE I
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE REFERENCE DATASET, WITH THE BEST RESULTS INDICATED IN BOLD AND THE SECOND-BEST RESULTS UNDERLINED.

Methods	Venue	LOL			iSAID-dark			iSAID-dark (retrain)		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Zero-DCE++ [43]	TPAMI'22	12.17	0.511	0.258	12.07	0.040	0.818	8.46	0.055	0.950
SCI [50]	CVPR'22	14.78	0.525	0.238	12.14	0.194	0.901	13.08	0.205	0.745
SNR [13]	CVPR'22	21.22	**0.834**	0.097	13.31	0.246	0.722	23.17	0.760	0.179
LLFormer [51]	AAAI'23	23.25	0.819	0.105	16.66	0.376	0.658	23.11	0.725	0.221
CUE [52]	ICCV'23	21.68	0.769	0.151	15.70	0.343	0.717	21.47	0.682	0.263
FourLLIE [26]	ACMM'23	20.02	0.818	0.088	14.49	0.350	0.579	19.78	0.643	0.309
LANet [53]	IJCV'23	21.74	0.816	0.101	11.88	0.089	0.804	18.72	0.695	0.231
NeRCo [54]	CVPR'23	22.94	0.783	0.102	**16.73**	**0.513**	**0.381**	18.79	0.604	0.326
Ours	-	**23.31**	0.832	**0.087**	14.99	0.282	0.718	**25.38**	**0.773**	**0.165**

Subsequently, weight filters $F_{sv} \in \mathbb{R}^{(k\times k\times c)\times h\times w}$ are generated based on 1×1 Conv, and reshaped into independent pixel kernels $F_{sv}^{'} = \{W_{i,j} \mid i \in [1,h], j \in [1,w]\}$, where $W_{i,j} \in \mathbb{R}^{k^2\times c}$. In this way, each position is adjusted to the optimum with adaptive weights. The filtering result can be represented as:

$$O_i = U_i \otimes F_{sv}^{'} + U_i, \tag{7}$$

where $\otimes$ indicates the convolution operation. Through the above process, IFAM generates adaptive weight filtering guided by multi-stage, multi-scale, and cross-domain interactive features, enabling the generated features to adapt more flexibly to image content.



(a) T-SNE visualization map

(b) Pixel Distribution in iSAID-dark

(c) SNR Distribution in iSAID-dark

Fig. 6. Distribution analysis of the iSAID-dark dataset. (a) T-SNE map indicates that the proposed dataset exhibits a broader data range. (b) demonstrates the brightness range of normal-light (blue) and low-light images (green). (c) The whole dataset covers a wide range of SNR distribution ranging from 1 to 25 and the SNR values center at the range of [0,10], indicating the challenging noise levels.

D. Loss Function

Due to our DFFN having two outputs O_A and O_P, the loss functions L_A and L_P for the two stages are as follows:

$$\begin{cases} L_A = \left\|O_A - \mathcal{F}^{-1}(\mathcal{A}(I_{gt}), \mathcal{P}(I_{low})\right\|_1 + \alpha \left\|\mathcal{A}(O_A) - \mathcal{A}(I_{gt})\right\|_1, \\ L_P = \left\|O_P - I_{gt}\right\|_1 + \beta \left\|\mathcal{A}(O_P) - \mathcal{A}(I_{gt})\right\|_1, \end{cases} \tag{8}$$

where $\|\cdot\|_1$ presents the Mean Absolute Error(MAE), α and β are the weight factor, and we empirically set them to 0.05. Therefore, the overall loss is :

$$L = L_A + L_P. \tag{9}$$

IV. DATASET

Due to the particularity of remote sensing images, collecting paired normal/low-light image sets is almost an impossible task. In this work, inspired by the significant achievements of previous synthetic datasets in various low-quality image enhancement tasks, we introduced a synthesis mechanism to generate the corresponding low-light remote sensing dataset. Additionally, to validate the reliability of the proposed DDFN and iSAID-dark in real low-light environments, we collected real nighttime remote sensing images by capturing different scenes with drones at night, forming a test dataset named dark remote sensing (darkrs), which comprises 86 test images. The two proposed datasets are shown in Fig. 5.

A. Data Collection

The iSAID [55] dataset has collected a large number of high-resolution aerial images from various scenes including rural and urban areas. Therefore, we selected 751 images from the iSAID test set as our benchmark data. Considering the significant differences in size and shape of remote sensing images, and the difficulty of training high-resolution images in mainstream networks, and in order to increase the diversity of scenes included in the dataset, we performed multiple random crops on these 751 images. The cropped images are resized to 500x500. Subsequently, we divided them into 3755 training image pairs and 66 validation image pairs.

B. Data Generation

1) Low-light Simulation: Inspired by Zero-DCE++ [43], which treats enhancing illumination as an estimation task of

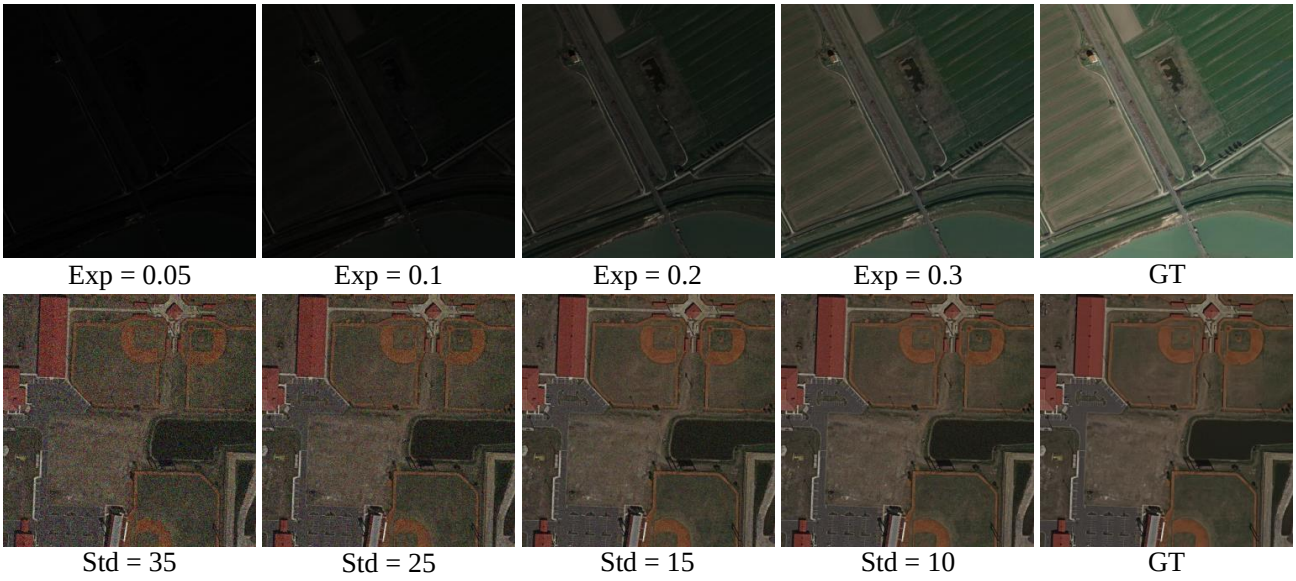


Fig. 7. The proposed iSAID-dark dataset contains low-light images with varying exposure levels and noise levels.

TABLE II
COMPARISON WITH STATE-OF-THE-ART METHODS ON THE NO-REFERENCE DATASET, WITH THE BEST RESULTS INDICATED IN BOLD AND THE SECOND-BEST RESULTS UNDERLINED.

Methods	Venue	Exdark	DICM	NPE	LIME	darks	Ave.	Params (M)
Zero-DCE++ [43]	TPAMI'22	3.992	3.924	3.489	4.188	3.760	3.870	0.0106
SCI [50]	CVPR'22	3.968	3.657	3.912	4.165	5.450	4.230	**0.0004**
SNR [13]	CVPR'22	4.125	4.458	5.503	5.844	3.456	4.677	40.0842
LLFormer [51]	AAAI'23	3.960	3.890	3.629	4.175	4.502	4.031	24.5183
CUE [52]	ICCV'23	3.989	3.722	3.977	4.068	3.708	3.892	0.2407
FourLLIE [26]	ACMM'23	3.941	3.685	3.550	4.056	4.119	3.870	1.4855
LANet [53]	IJCV'23	4.322	3.992	3.828	4.493	3.478	4.022	0.0422
NeRCo [54]	CVPR'23	4.092	4.052	4.048	4.007	4.140	4.067	7.8414
Ours	-	**3.568**	**3.402**	**3.182**	**3.940**	**3.447**	**3.507**	2.5765

specific curves for images and achieves excellent results, we adopted an inverse transformation of Zero-DCE++ to simulate low-light images by adjusting the reverse curve. This allows us to generate low-light images with adjustable darkness levels. Specifically, we set a random darkness range to replace the fixed exposure value of brightness loss in Zero-DCE++. By using different exposure levels, the trained reverse Zero-DCE++ can generate low-light images with varying darkness levels.

2) Noise Simulation: To simulate more realistic low-light images, we also added Gaussian noise of varying degrees to the generated images, all of which followed a normal distribution. Additionally, for images with higher levels of darkness, the added noise was more intense.

C. Data Distribution

Fig. 6 (a) displays the T-SNE visualization results of our proposed iSAID-dark dataset compared to other datasets, demonstrating that iSAID-dark encompasses common features of LOL and SICE, indicating a broader data range in our dataset. Fig. 6 (b) illustrates the brightness distribution of our proposed dataset. We also plotted the SNR distribution of the dataset in Fig. 6 (c) to showcase the noise level of the images. The SNR distribution indicates that our dataset possesses a wide and challenging SNR range.

D. Discussion

Our dataset simulates real low-light environments, encompassing the majority of common remote sensing scenes. Additionally, our randomly set dark intensity range and noise cover most common challenging low-light scenarios. When given different exposure strengths, the reverse Zero-DCE++ can generate low-light images with different brightness levels. Fig. 7 illustrates images with different levels of low-light and noise.

V. EXPERIMENTS

In this section, we first describe the detailed setup of the experiments, including the datasets used, evaluation metrics, comparison methods, and implementation details. Subsequently, we quantitatively and qualitatively evaluate the proposed DFFN against state-of-the-art methods on the proposed low-light remote sensing dataset and real-world low-light dataset. Additionally, we provide extensive ablation studies to validate the effectiveness of the proposed DFFN architecture and its components.

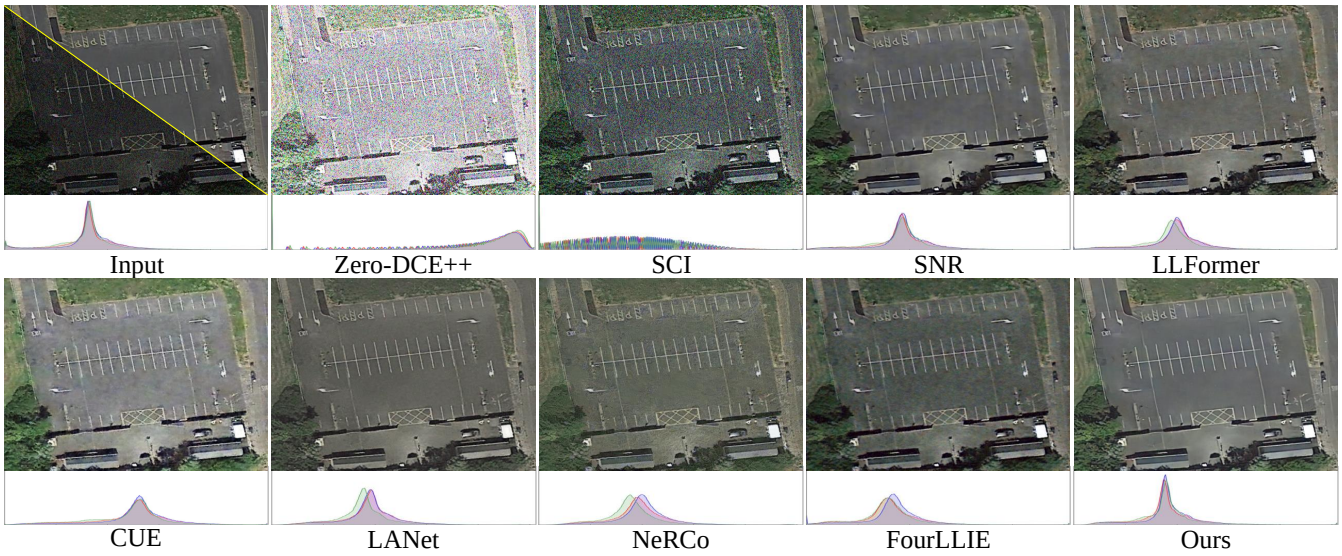


Fig. 8. The visualization results on the iSAID-dark dataset. We present the histogram of color distribution for the images. The histograms placed in Input/GT represent the color distribution of the GT. It can be observed that our method's histogram is closer to the GT histogram.

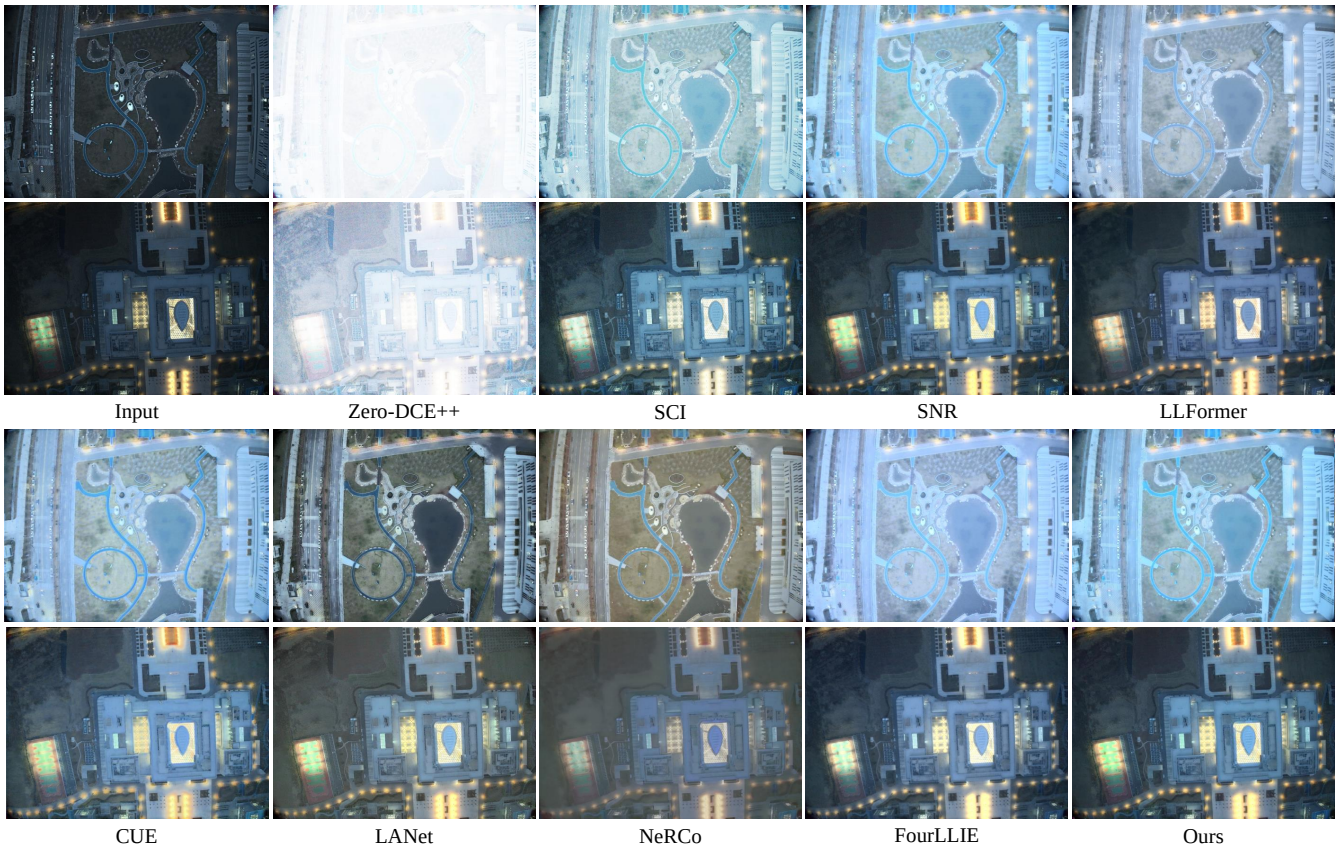


Fig. 9. The visualization results on the darkrs dataset.

A. Experiment Details

1) Datasets and Evaluation Metrics: We first validate the performance of the proposed method on the iSAID-dark and dark-rs datasets. Additionally, to assess the generalization ability of our model to low-light images in other scenarios, we further evaluate the performance of the proposed method on five widely used datasets, including a reference-based dataset (LOL [56]) and three no-reference datasets (DICM, Exdark, and NPE).

We evaluate the visual quality of reference-based datasets using PSNR [16], SSIM [15], and LPIPS [57] metrics. For no-reference images, we assess their naturalness using the NIQE [14] metric. Higher PSNR and SSIM scores indicate better

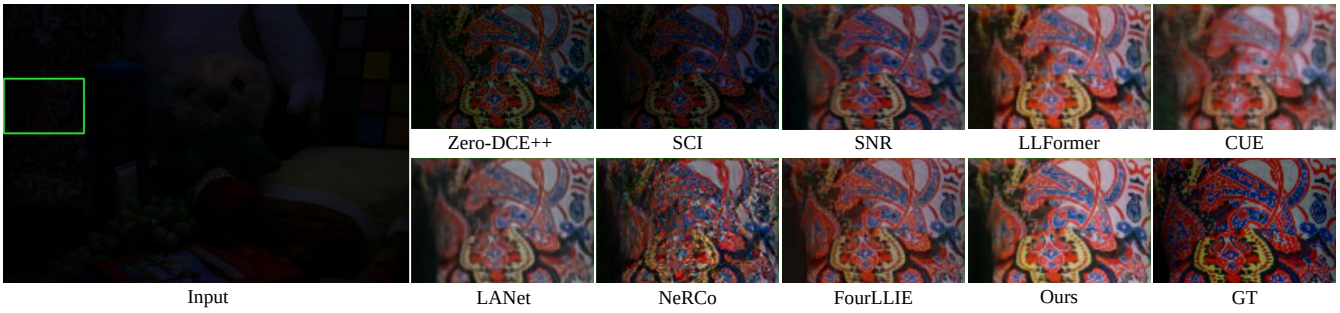


Fig. 10. The visualization results on the LOL dataset. The selected area is magnified to show details.

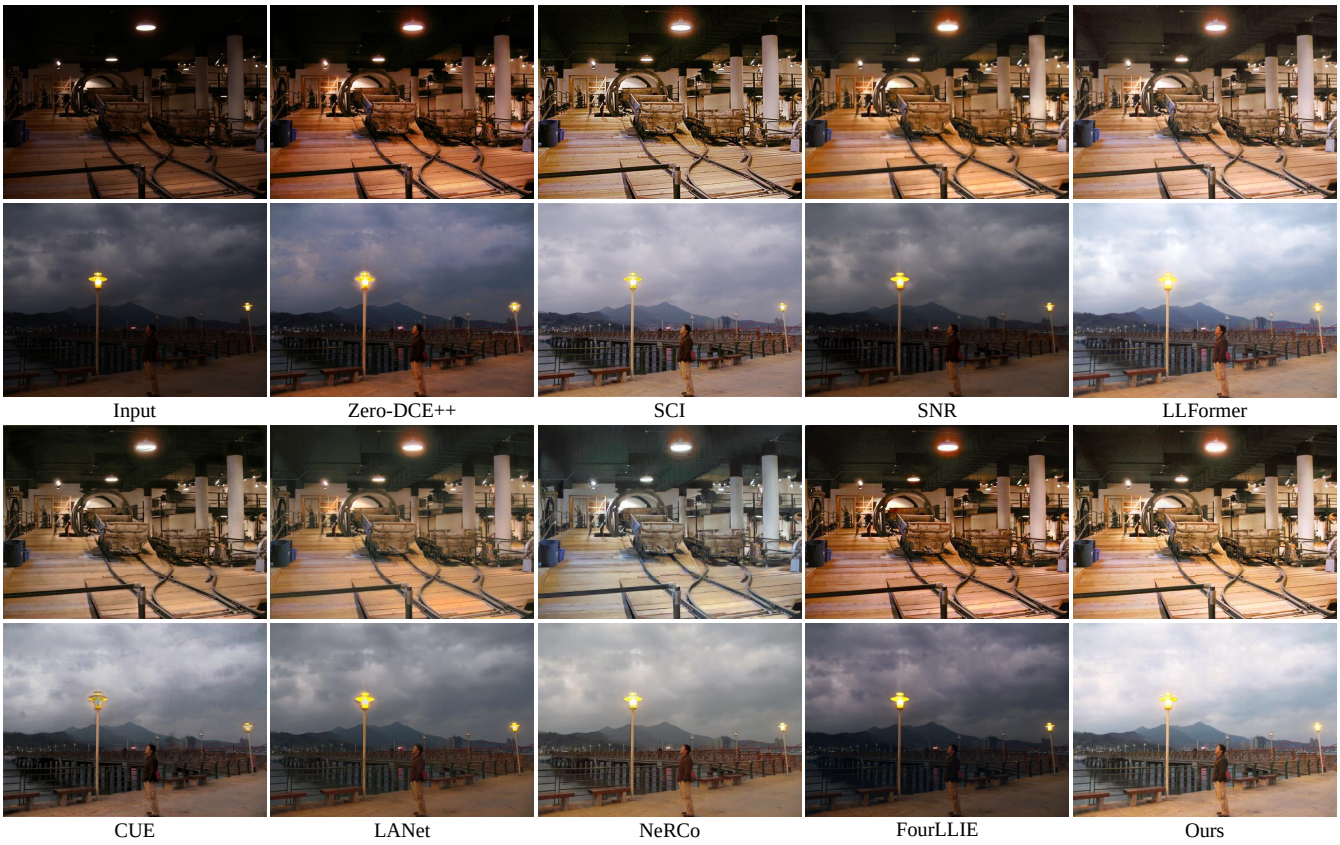


Fig. 11. The visualization results on the DICM dataset (top) and the NPE dataset (bottom).

quality, while lower NIQE and LPIPS scores are desirable.

2) Compared Methods: To evaluate the performance of the proposed method, we compare DFFN with current state-of-the-art methods, including a curve-based method (Zero-DCE++ [43]), two unsupervised learning-based methods (SCI [50], NeRCo [54]), and five supervised learning-based methods (LLFormer [51], SNR [13], CUE [52], FourLLIE [26], and LANet [53]). In the experiments, we retrain and test these methods using the weights and publicly available code provided by these approaches (for the proposed iSAID-dark dataset) for fair comparison.

3) Implementation Details: Our model is trained on the LOL dataset and the iSAID-dark dataset separately. To facilitate better convergence of the model, we set the initial learning rate to 0.0002, and then linearly decrease it by half every 100 epochs, with a total of 500 epochs for training. We use the ADAM optimizer for model optimization. All experiments are conducted on a Windows computer equipped with an NVIDIA RTX 3090 (24GB RAM) for training and testing. Additionally, we set the batch size to 12, randomly crop each image into 256×256 patches, and augment these patches using data augmentation techniques such as horizontal and vertical flips.

B. Comparison With Stat-of-the-Art Methods

1) Quantitative Results: We first quantitatively compare DFFN with state-of-the-art methods. Table I presents the PSNR, SSIM, and LPIPS results on reference-based datasets. From Table I, it can be seen that our method achieved the best or second-best results on both reference datasets. It is worth

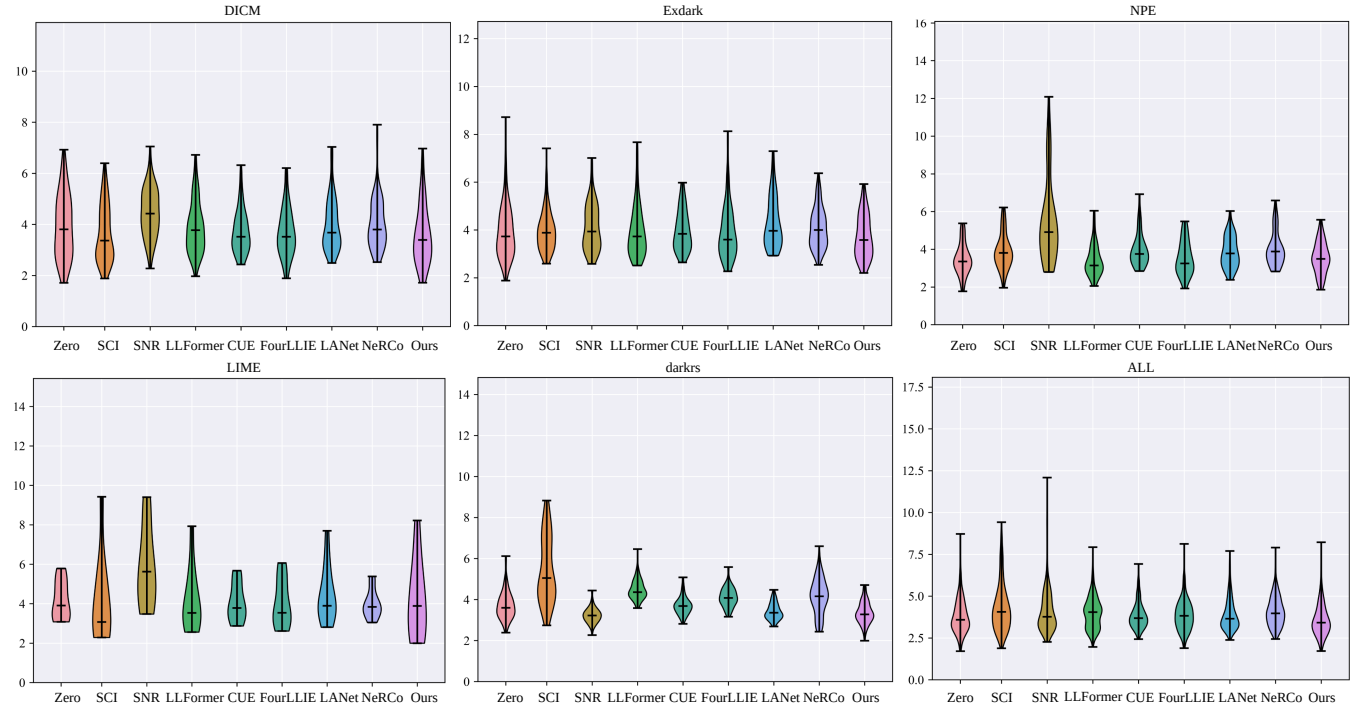


Fig. 12. On no-reference image datasets, the numerical distribution of NIQE test results varies across different enhancement methods. The evaluation value of the proposed method is relatively concentrated, with low scores across multiple images, and better performance in low-light enhancement scenarios.

noting that the results shown for iSAID-dark are obtained by testing with weights trained on LOL, indicating that models trained on existing datasets may not be suitable for low-light remote sensing image enhancement.

Table II presents the NIQE results on the no-reference dataset. It can be observed from Table II that our method achieves the best on all datasets, which fully demonstrates the superiority of DFFN. We also present the parameter count of the models in Table II. The results indicate that although our method cannot compete with no-reference methods in terms of complexity, the performance of no-reference methods on low-light remote sensing datasets is extremely unsatisfactory (for example, SSIM of Zero-DCE++ on iSAID-dark is only 0.055). In contrast, our method achieves a remarkable balance between model complexity and performance.

2) Qualitative Results: Fig. 8 illustrates the enhancement results of our method compared to other state-of-the-art methods on the iSAID-dark dataset. Additionally, histograms of color distributions for these images are plotted, where the histograms shown under Input/GT represent the GT histograms. Due to the significant Gaussian noise added to the iSAID-dark dataset, no-reference methods such as Zero-DCE++ and SCI, which were not designed to address denoising issues, still exhibit substantial noise in the recovered images. Consequently, these two methods show low SSIM and PSNR values as indicated in Table I. Furthermore, the restoration results of other methods either suffer from excessive color distortion or exhibit blurred details due to noise. In contrast, our method achieves the best visual results, while the histograms of color distributions are closer to the GT.

Fig. 9 depicts remote sensing images captured by drones in real nighttime environments. The dataset contains numerous areas with uneven exposure, providing a better balance for evaluating the model's ability to enhance low-light remote sensing images in real-world scenarios. It can be observed that Zero-DCE++ exhibits overexposure, while LANet, and NeRCo show limited brightening effects, resulting in overall low contrast in the images. SCI, LLFormer and FourLLIE demonstrate good enhancement in low-light areas, but some regions suffer from color distortion, and FourLLIE blurs image details. Our method, effectively enhances image brightness while better preserving image color.

Fig. 10 displays the enhancement results on the LOL dataset. We have selected certain regions within the images for zooming in to present the image details more clearly. From the image, it can be seen that Zero-DCE++, SCI, and SNR only achieved limited enhancement. The results of CUE and LANet are too blurred in details. NeRCo exhibited some unexplained shadows. Although FourLLIE and LLFormer also achieved decent enhancement effects, compared to them, our method shows more prior colors and clearer details.

Fig. 11 illustrates the enhancement results on the DICM dataset and the NPE dataset. Zero-DCE++, SNR, and FourLLIE produce overly dark enhancement results, with little difference compared to the degraded images, especially in the images of the second row. SCI, CUE, LANet, and NeRCo introduce some color deviations and inadequate brightness restoration in the enhanced images. In comparison, our method effectively suppresses noise while restoring image brightness, presenting the most natural images.

Meanwhile, Fig. 12 displays the NIQE distributions of the comparison methods across five test datasets. It is evident

TABLE III
ABLATION STUDY OF THE TWO-STAGE ARCHITECTURE DESIGN INCLUDES EXPERIMENTS WHERE 'FREQUENCY' INDICATES WHETHER FREQUENCY DOMAIN IS USED, 'STAGE2' DENOTES WHETHER THE SECOND STAGE IS UTILIZED, AND 'SWAP' INDICATES WHETHER THE PHASE INFORMATION OF THE OUTPUT O_1 FROM THE FIRST STAGE IS REPLACED.

Model	Frequency	Stage2	Swap	PSNR↑	SSIM↑	LPIPS↓
M_a	✗	✗	✗	21.58	0.734	0.198
M_b	✓	✗	✗	22.59	0.750	0.178
M_c	✓	✓	✗	24.52	0.771	0.167
M_d	✓	✓	✓	**24.93**	**0.772**	**0.166**

TABLE IV
ABLATION STUDY ABOUT THE COMBINATION OF SPATIAL-FREQUENCY DOMAIN FEATURES.

Model	M_1	M_2	M_3	M_4	Ours
PSNR↑	21.55	24.88	24.88	24.93	**25.38**
SSIM↑	0.655	0.771	0.772	0.772	**0.773**
LPIPS↓	0.289	**0.162**	0.171	0.166	0.165

that DFFN exhibits a more concentrated data distribution, which strongly indicates the superior generalization capability of our model across various scenarios. Additionally, when all test set data is combined and plotted together (denoted as ALL), DFFN's data appears more centralized. This further demonstrates the exceptional stability of our model.

TABLE V
THE ABLATION STUDY FOR USING TRANSFORMER BLOCKS, 'ONE_BLOCK' INDICATES THAT EACH BLOCK USES ONLY ONE SELF-ATTENTION BLOCK AND ONE FFN, WHILE 'STACK_BLOCK' SIGNIFIES THAT MULTIPLE SELF-ATTENTION BLOCKS AND FFNS ARE STACKED WITHIN EACH TRANSFORMER BLOCK.

Model	Attention	Head	PSNR↑	SSIM↑	LPIPS↓	Params (M)
One_block	[1,1,1,1]	[2,2,2,2]	23.53	0.687	0.260	6.3139
Stack_block	[2,4,6,8]	[2,2,4,4]	23.02	0.726	0.203	26.9895
Ours	-	-	**25.38**	**0.788**	**0.146**	**2.5765**

C. Ablation Study

In this subsection, we conducted extensive ablation experiments to validate the effectiveness of the proposed modules, the two-stage network, and the fusion of spatial and frequency domains.

1) Contribution of Two-stage Architecture: In this subsection, we validate the effectiveness of the proposed two-stage network architecture. All experiments are conducted based on the foundational model without employing the IFAM modules. M_a represents the use of spatial information only in a single stage, M_b denotes the simultaneous utilization of spatial and frequency domain information in the single stage, and M_c employs the two-stage architecture, but with O_1 as the input for the second stage, M_d is identical to our training architecture

As shown in Table III, the performance of M_a is the poorest, indicating that solely recovering spatial information in a single stage may yield unsatisfactory results. Compared to M_b, M_c and M_d exhibit significantly improved performance, implying that decoupling and separately learning degraded information are crucial for information recovery. Moreover, M_d achieves a gain of 0.21dB in PSNR over M_c, indicating that preserving phase consistency is more critical for learning in the second stage.

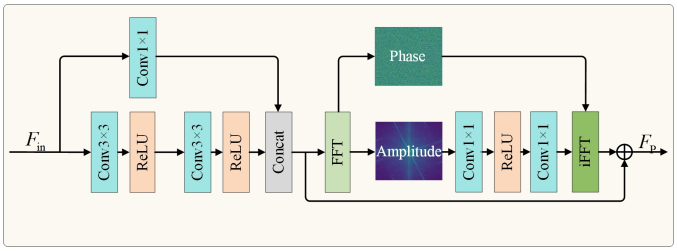


Fig. 13. The modified DDAB first learns spatial domain information in a ser ial manner, followed by learning frequency domain information.

2) Contribution of Dual-Domain Fusion: In this subsection, we discuss the effectiveness of integrating spatial and frequency domain information and the effectiveness of different integration methods, while also exploring the effectiveness of the proposed IFAM. M_1 denotes the use of spatial information only in a two-level network. M_2 and M_3 adopt a sequential method to learn spatial and frequency domain information, respectively. Specifically, M_2 first learns spatial information (as shown in Fig. 13), while M_3 first learns frequency domain information. M_4 has the same training architecture as ours, but none of the four configurations use IFAM.

As shown in Table IV, the model M_1 utilizing only spatial information performs poorly, indicating the effectiveness of integrating spatial-frequency dual-domain information. Compared to M_2 and M_3 ,M_4, which learns in parallel, achieves the best performance in most metrics. This suggests that compared to the parallel approach, serial learning is more prone to losing information from one domain while learning information from another. By employing parallel computation, this information loss is effectively mitigated.

Moreover, when IFAM is utilized, there is a significant improvement in PSNR, indicating that IFAM successfully aggregates features from different sources and enhances the global context representation of the network.

3) Compared with transformer blocks: Finally, we evaluated the effectiveness and efficiency of the proposed DFFN compared to transformers. Specifically, we replaced the frequency domain branches in DDAB and DDPB with the most advanced transformer blocks [58] currently available, and only used spatial domain information for supervision, with all other operations being the same as in DFFN.

Initially, for fairness, we used only one Self-Attention block and one FFN operation per block, setting heads to 2. Although transformers are designed to capture long-range dependencies, shallow transformer blocks might struggle to learn rich feature representations and perform less effectively than frequency domain networks when dealing with large images. As shown in Table V, transformer blocks with only a single operation were not very effective and introduced a large number of parameters.

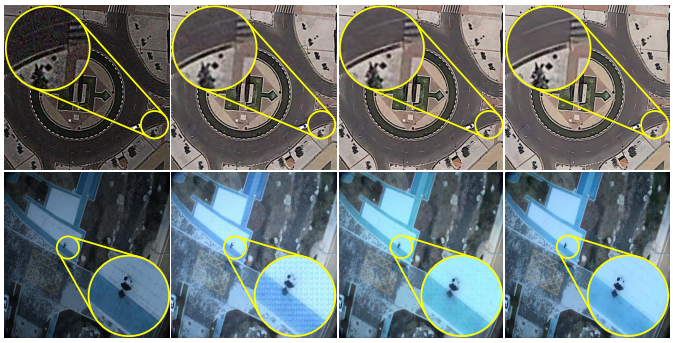


Fig. 14. The visualization results compared with transformer blocks, 'One_block' indicates that each block uses only one self-attention block and one FFN, while 'Stack_block' signifies that multiple self-attention blocks and FFNs are stacked within each transformer block.

Subsequently, we followed the mainstream setup, stacking [2,4,6,8] self-attention blocks and FFN operations in each stage, with heads set to [2,2,4,4]. However, this setup increased the parameter count to nearly 27M, and the final enhancement effect was still slightly inferior to our method. In contrast, our frequency domain strategy only had 2.57M parameters but achieved the best results. Additionally, as shown in Fig. 14, the visualization demonstrates that our method achieves the best visual effects.

VI. Conclusion

In this paper, we propose a two-stage DFFN for enhancing low-light remote sensing images. The DFFN combines the advantages of spatial and frequency domains, consisting of two sub-networks: the amplitude illumination stage and the phase refinement stage. To better integrate dual-domain information, we design the corresponding DDAB & DDRB. Due to the lack of information interaction in the two-stage network, we further design IFAM to interactively fuse features of different domains, stages, and sizes in the two stages, enhancing the model's contextual representation capability. Furthermore, we introduce the iSAID-dark and darkrs datasets for low-light remote sensing image enhancement.

Extensive quantitative and qualitative experiments demonstrate that the proposed method achieves outstanding performance in low-light remote sensing image enhancement, surpassing existing state-of-the-art methods on other datasets as well.

References

[1] J. Zhang, J. Lei, W. Xie, Y. Li, G. Yang, and X. Jia, "Guided hybrid quantization for object detection in remote sensing imagery via one-to-one self-teaching," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

[2] Y.-f. Zhang, J. Zheng, L. Li, N. Liu, W. Jia, X. Fan, C. Xu, and X. He, "Rethinking feature aggregation for deep rgb-d salient object detection," *Neurocomputing*, vol. 423, pp. 463–473, 2021.

[3] L. Ma, R. Liu, Y. Wang, X. Fan, and Z. Luo, "Low-light image enhancement via self-reinforced retinex projection model," *IEEE Transactions on Multimedia*, vol. 25, pp. 3573–3586, 2023.

[4] J. Li, X. Feng, and Z. Hua, "Low-light image enhancement via progressive-recursive network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 11, pp. 4227–4240, 2021.

[5] Y. Zhang, X. Di, B. Zhang, R. Ji, and C. Wang, "Better than reference in low-light image enhancement: conditional re-enhancement network," *IEEE Transactions on Image Processing*, vol. 31, pp. 759–772, 2021.

[6] Y. Huang, X. Tu, G. Fu, T. Liu, B. Liu, M. Yang, and Z. Feng, "Low-light image enhancement by learning contrastive representations in spatial and frequency domains," in *2023 IEEE International Conference on Multimedia and Expo*, 2023, pp. 1307–1312.

[7] G.-D. Fan, B. Fan, M. Gan, G.-Y. Chen, and C. P. Chen, "Multiscale low-light image enhancement network with illumination constraint," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7403–7417, 2022.

[8] L. Ma, R. Liu, J. Zhang, X. Fan, and Z. Luo, "Learning deep context-sensitive decomposition for low-light image enhancement," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5666–5680, 2021.

[9] Z. Li, Y. Wang, and J. Zhang, "Low-light image enhancement with knowledge distillation," *Neurocomputing*, vol. 518, pp. 332–343, 2023.

[10] G. Fan, M. Gan, B. Fan, and C. L. P. Chen, "Multiscale cross-connected dehazing network with scene depth fusion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 2, pp. 1598–1612, 2024.

[11] J. Zhou, Q. Liu, Q. Jiang, W. Ren, K.-M. Lam, and W. Zhang, "Underwater camera: Improving visual perception via adaptive dark pixel prior and color correction," *International Journal of Computer Vision*, pp. 1–19, 2023.

[12] S. Yuan, L. Li, H. Chen, and X. Li, "Surface defect detection of highly reflective leather based on dual-mask-guided deep-learning model," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–13, 2023.

[13] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 714–17 724.

[14] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[16] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[17] J.-N. Su, M. Gan, G.-Y. Chen, J.-L. Yin, and C. L. P. Chen, "Global learnable attention for single image super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 8453–8465, 2023.

[18] S. Zhang, J. Zhang, X. Wang, J. Wang, and Z. Wu, "Els2t: Efficient lightweight spectral–spatial transformer for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.

[19] J. Feng, Q. Wang, G. Zhang, X. Jia, and J. Yin, "Cat: Center attention transformer with stratified spatial-spectral token for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–1, 2024.

[20] C. Zhao, B. Qin, S. Feng, W. Zhu, W. Sun, W. Li, and X. Jia, "Hyperspectral image classification with multi-attention transformer and adaptive superpixel segmentation-based active learning," *IEEE Transactions on Image Processing*, vol. 32, pp. 3606–3621, 2023.

[21] J.-N. Su, M. Gan, G.-Y. Chen, W. Guo, and C. L. P. Chen, "High-similarity-pass attention for single image super-resolution," *IEEE Transactions on Image Processing*, vol. 33, pp. 610–624, 2024.

[22] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[23] W. Dong, Y. Yang, J. Qu, Y. Li, Y. Yang, and X. Jia, "Feature pyramid fusion network for hyperspectral pansharpening," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2023.

[24] J. Zhou, B. Li, D. Zhang, J. Yuan, W. Zhang, Z. Cai, and J. Shi, "Ugif-net: An efficient fully guided information flow network for underwater image enhancement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–17, 2023.

[25] L. Ma, D. Jin, N. An, J. Liu, X. Fan, Z. Luo, and R. Liu, "Bilevel fast scene adaptation for low-light image enhancement," *International Journal of Computer Vision*, pp. 1–19, 2023.

[26] C. Wang, H. Wu, and Z. Jin, "Fourllie: Boosting low-light image enhancement by fourier frequency information," in *Proceedings of the*

31st ACM International Conference on Multimedia, 2023, pp. 7459–7469.

[27] A. Oppenheim, J. Lim, G. Kopec, and S. Pohlig, "Phase in speech and pictures," in *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. IEEE, 1979, pp. 632–637.

[28] C. Li, C.-L. Guo, M. Zhou, Z. Liang, S. Zhou, R. Feng, and C. C. Loy, "Embedding fourier for ultra-high-definition low-light image enhancement," *arXiv preprint arXiv:2302.11831*, 2023.

[29] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 1921–1935, 2009.

[30] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3431–3441, 2011.

[31] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1752–1758, 2007.

[32] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1032–1041, 2012.

[33] H. Singh, A. Kumar, L. Balyan, and G. K. Singh, "A novel optimally gamma corrected intensity span maximization approach for dark image enhancement," in *2017 22nd International Conference on Digital Signal Processing*. IEEE, 2017, pp. 1–5.

[34] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2782–2790.

[35] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.

[36] C. Xu, H. Fu, L. Ma, W. Jia, C. Zhang, F. Xia, X. Ai, B. Li, and W. Zhang, "Seeing text in the dark: Algorithm and benchmark," *arXiv preprint arXiv:2404.08965*, 2024.

[37] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.

[38] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2016.

[39] H. Shen, Z.-Q. Zhao, Y. Zhang, and Z. Zhang, "Mutual information-driven triple interaction network for efficient image dehazing," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 7–16.

[40] J. Liang, Y. Xu, Y. Quan, B. Shi, and H. Ji, "Self-supervised low-light image enhancement using discrepant untrained network priors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7332–7345, 2022.

[41] Y. Luo, B. You, G. Yue, and J. Ling, "Pseudo-supervised low-light image enhancement with mutual learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 1, pp. 85–96, 2024.

[42] L. Ma, R. Liu, J. Zhang, X. Fan, and Z. Luo, "Learning deep context-sensitive decomposition for low-light image enhancement." *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5666–5680, 2022.

[43] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4225–4238, 2022.

[44] Y. Wang, Z. Liu, J. Liu, S. Xu, and S. Liu, "Low-light image enhancement with illumination-aware gamma correction and complete image modelling network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 13 128–13 137.

[45] L. Xing, H. Qu, S. Xu, and Y. Tian, "Clegan: Toward low-light image enhancement for uavs via self-similarity exploitation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[46] X. Mao, Y. Liu, W. Shen, Q. Li, and Y. Wang, "Deep residual fourier transformation for single image deblurring," *arXiv preprint arXiv:2111.11745*, vol. 2, no. 3, p. 5, 2021.

[47] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, and V. Lempitsky, "Resolution-robust large mask inpainting with fourier convolutions," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2149–2159.

[48] D. Fuoli, L. Van Gool, and R. Timofte, "Fourier space losses for efficient perceptual image super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2360–2369.

[49] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 821–14 831.

[50] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5637–5646.

[51] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 3, 2023, pp. 2654–2662.

[52] N. Zheng, M. Zhou, Y. Dong, X. Rui, J. Huang, C. Li, and F. Zhao, "Empowering low-light image enhancer through customized learnable priors," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 12 559–12 569.

[53] K.-F. Yang, C. Cheng, S.-X. Zhao, H.-M. Yan, X.-S. Zhang, and Y.-J. Li, "Learning to adapt to light," *International Journal of Computer Vision*, vol. 131, no. 4, pp. 1022–1041, 2023.

[54] S. Yang, M. Ding, Y. Wu, Z. Li, and J. Zhang, "Implicit neural representation for cooperative low-light image enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 12 918–12 927.

[55] S. Waqas Zamir, A. Arora, A. Gupta, S. Khan, G. Sun, F. Shahbaz Khan, F. Zhu, L. Shao, G.-S. Xia, and X. Bai, "isaid: A large-scale dataset for instance segmentation in aerial images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 28–37.

[56] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *British Machine Vision Conference*, 2018.

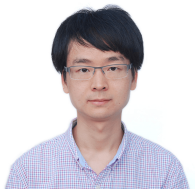
[57] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and pattern recognition*, 2018, pp. 586–595.

[58] X. Chen, H. Li, M. Li, and J. Pan, "Learning a sparse transformer network for effective image deraining," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2023, pp. 5896–5905.

Zishu Yao received the B.S. degree from Shandong University of Finance and Economics, China, in 2021, where he is currently pursuing the M.S. degree with the Qingdao University. His research interests are in image processing, machine learning, and computer vision.

Guodong Fan received the M. Eng. degree from Shandong Technology and Business University, China, in 2021, where he is currently pursuing the Ph.D. degree with the Qingdao University. His research interests are in image processing, machine learning, and computer vision.

Jinfu Fan the Ph.D. degree in control theory and control engineering from Tongji University, Shanghai. He is currently an associate professor with the College of Computer Science and Technology, Qingdao University, China. His current research interests include machine learning and computer vision, image processing, graph network, especially in learning from partial label data.



Min Gan (Senior Member, IEEE) received the B. S. degree in Computer Science and Engineering from Hubei University of Technology, Wuhan, China, in 2004, and the Ph.D. degree in Control Science and Engineering from Central South University, Changsha, China, in 2010. His current research interests include machine learning, inverse problems, and image processing.

C. L. Philip Chen (Fellow, IEEE) is the Chair Professor and Dean of the College of Computer Science and Engineering, South China University of Technology. Being a Program Evaluator of the Accreditation Board of Engineering and Technology Education(ABET) in the U.S., for computer engineering, electrical engineering, and software engineering programs, he successfully architects the University of Macau's Engineering and Computer Science programs receiving accreditations from Washington/Seoul Accord through Hong Kong Institute of Engineers(HKIE), of which is considered as his utmost contribution in engineering/computer Science education for Macau as the former Dean of the Faculty of Science and Technology. Prof. Chen is a Fellow of IEEE, AAAS, IAPR, CAA, and HKIE; a member of Academia Europaea(AE), European Academy of Sciences and Arts(EASA), and International Academy of Systems and Cybernetics Science(IASCYS). He received IEEE Norbert Wiener Award in 2018 for his contribution in systems and cybernetics, and machine learnings. He is also a Highly Cited Researcher by Clarivate Analytics in 2018 and 2021.