

# Degradation-Aware Dynamic Fourier-Based Network for Spectral Compressive Imaging

Ping Xu , Lei Liu , Haifeng Zheng , Senior Member, IEEE, Xin Yuan , Senior Member, IEEE, Chen Xu , and Lingyun Xue 

**Abstract**—We consider the problem of hyperspectral image (HSI) reconstruction, which aims to recover 3D hyperspectral data from 2D compressive HSI measurements acquired by a coded aperture snapshot spectral imaging (CASSI) system. Existing deep learning methods have achieved acceptable results in HSI reconstruction. However, these methods did not consider the imaging system degradation pattern. In this article, based on observing the initialized HSIs obtained by shifting and splitting the measurements, we propose a dynamic Fourier network based on degradation learning, called the degradation-aware dynamic Fourier-based network (DADF-Net). We estimate the degradation feature maps from the degraded hyperspectral images to realize the linear transformation and dynamic processing of the features. In particular, we use the Fourier transform to extract the HSI non-local features. Extensive experimental results show that the proposed model outperforms state-of-the-art algorithms on simulation and real-world HSI datasets.

**Index Terms**—Deep learning, fourier transform, hyperspectral images, snapshot compressive imaging.

## I. INTRODUCTION

HYPERSPECTRAL images (HSIs) have been applied in various fields, such as remote sensing [1], [2], [3], object tracking [4], [5], and medical diagnosis [6], [7], due to their rich spectral information. Coded aperture snapshot spectral imaging (CASSI) is an emerging technique for spectral imaging that compresses multichannel HSIs into 2D measurements [8], [9], [10], [11], [12], [13]. After capturing a measurement using CASSI, it is crucial to reconstruct the desired HSI with high fidelity. The

Manuscript received 5 April 2023; revised 20 June 2023 and 1 August 2023; accepted 7 August 2023. Date of publication 11 August 2023; date of current version 2 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grants 62273125, 61971139, and 62271414, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR23F010001, and in part by the Research Center for Industries of the Future (RCIF) at Westlake University. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yuxin Peng. (*Corresponding authors: Ping Xu; Lingyun Xue.*)

Ping Xu, Lei Liu, Chen Xu, and Lingyun Xue are with the School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: xuping@hdu.edu.cn; liu\_lei@hdu.edu.cn; xuchen@hdu.edu.cn; xly@hdu.edu.cn).

Haifeng Zheng is with the Fujian Key Lab for Intelligent Processing and Wireless Transmission of Media Information, College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China (e-mail: zhenghf@fzu.edu.cn).

Xin Yuan is with the Research Center for Industries of the Future and School of Engineering, Westlake University, Hangzhou 310030, China (e-mail: xyuan@westlake.edu.cn).

The source code is available at: <https://github.com/CISMOLab/DADF-Net>  
Digital Object Identifier 10.1109/TMM.2023.3304450

model-based methods involve formulating an optimization problem for HSI reconstruction based on prior knowledge, such as total variation (TV) [14], [15], [16], sparsity [9], low-rank [17], and non-local similarity [18], [19]. This optimization problem is then mathematically solved to obtain the reconstructed HSI. However, these methods require manual parameter adjustments, leading to unpleasant reconstruction performance and long reconstruction times.

With the development of deep learning and its application in low-level vision tasks such as image restoration [20], [21] and image super-resolution [22], [23], [24], [25], [26], [27], [28], researchers have started utilizing deep learning-based methods [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39] for HSI reconstruction. For example, the plug-and-play (PnP) algorithm has been proposed, which incorporates a pre-trained denoising network into the model-based method to recover HSIs from CASSI measurements [39]. A series of deep unfolding methods have been proposed that utilize multi-stage networks that combine mathematical methods or prior knowledge to reconstruct HSIs [29], [30], [33], [34]. End-to-end methods employ a deep learning network to learn the mapping from measurements to target HSIs [31], [32], [35], [37], [38]. Most end-to-end methods [31], [35], [37], [38] utilize an encoder-decoder architecture, commonly in the form of a U-shaped network [40].

Previous deep learning-based methods for HSI reconstruction [31], [35], [37], [38] have achieved acceptable results. However, these methods do not consider the degradation patterns of the entire system. Since the degradation pattern is related to the imaging process, incorporating the degradation pattern into the network structure may enhance the reconstruction performance of the network. In addition, non-local features and long-range dependencies are crucial in HSI reconstruction. Transformer [41] and self-attention, known for their large receptive fields, have been successfully applied to capture non-local features and long-range dependencies. The self-attention mechanism was integrated into the bottleneck layer of a U-shaped network for extracting non-local features while reducing computational costs [31], [37]. Leveraging the power of the transformer, a reconstruction model that captures non-local features was developed, leading to impressive performance [35], [38]. Traditional small kernel convolution networks are limited in achieving a large receptive field. Although large convolution kernels can provide a larger receptive field, they pose challenges in training and network design [42]. In this article, we propose a degradation-aware dynamic Fourier-based network

dubbed DADF-Net to address these problems. First, we analyze the potential degradation pattern between the input of the network (initialized HSI) and the ground-truth HSI and then adopt the degradation pattern into the network architecture. We design an estimation block that generates a degradation feature map containing the degradation information of the input. This degradation feature map is utilized to generate the inverse factors required by the Trans-block and the filter weights for subsequent modules. Second, we design a locally enhanced Fourier block (LFB), which can extract local and non-local features simultaneously. Different from previous work, which uses a convolution network to extract features in the Fourier domain [43], the non-local branch of LFB relies on mapping the image convolution operation in the Fourier domain and directly learns to generate or dynamically generate filters to extract non-local features of input in the Fourier domain. Then these features are converted to the image domain by an inverse Fourier transform. The local branch relies on a convolution network to extract local features in the image domain. Furthermore, we introduce a multi-scale design to improve the performance of the entire network. The proposed method achieves state-of-the-art (SOTA) performance. The main contributions of this article are summarized as follows:

- We analyze the potential relationship between the network input and the ground-truth HSI and design an estimation block and a Trans-block based on this relationship. The estimation block generates a degradation feature map, which is used to generate the inverse factors required by the Trans-block and the dynamic filter weights for the subsequent modules in the network.
- We propose a locally enhanced Fourier block, which includes the local and non-local branches, to capture local and non-local features. Combining these non-local features with local features extracted by the local branch improves the overall performance of the network.
- We introduce a multi-scale design incorporating multi-scale feature fusion and multi-scale reconstruction loss constraints. This design aims to minimize information loss within the network and enhance the feature processing capability across different scales, leading to a reconstruction performance improvement.
- Our method DADF-Net achieves SOTA quantitative results and pleasant visual effects.

The rest of this article is organized as follows: Section II provides a survey of related work. Section III describes the CASSI System. Section IV provides a detailed description of the proposed method. Section V presents the experimental results, and Section VI concludes the article.

## II. RELATED WORK

### A. HSI Reconstruction

Traditional model-based methods rely on hand-crafted prior knowledge and mathematical optimization to achieve HSI reconstruction. The gradient projection for sparse reconstruction (GPSR) method was utilized for HSI reconstruction by incorporating the sparsity constraint [9]. The multiple snapshot strategy was employed to reduce the ill-posedness of HSI reconstruction

with a total variation (TV) constraint [14]. The two-step iterative shrinkage/thresholding (TwIST) algorithm [44] was employed to solve the HSI reconstruction problem using the TV prior constraint while incorporating the gray-scale image of the target scene to ensure spatial smoothness [15]. The generalized alternating projection (GAP) algorithm was utilized to solve the TV minimization problem for HSI reconstruction [16]. A dimension-discriminative low-rank tensor recovery model incorporating non-local similarity was proposed to restore HSIs [17]. Another approach involved building a joint model that integrates non-local self-similarity of HSIs and rank minimization, followed by utilizing an alternating minimization algorithm to solve it [18]. However, these model-based methods exhibit unsatisfactory performance and usually require long reconstruction times.

Deep learning-based methods have been applied in compressive HSI reconstruction. The convolutional autoencoder was used to reconstruct HSIs from measurements [45]. The pre-trained network was incorporated into the model-based method to improve reconstruction performance [39]. Several deep unfolding methods were developed using mathematical optimization [33], [46] and prior knowledge [29], [30]. The iterative shrinkage/thresholding (IST) algorithm was unfolded into a deep network for HSI reconstruction [33]. The GAP algorithm was unfolded into a deep network, where the denoiser was based on an autoencoder [46]. A deep network based on the Gaussian scale mixture model was proposed for HSI reconstruction [29]. The optimization-inspired network was introduced with data-driven priors [30]. The convolution and contextual Transformer (CCoT) block was incorporated into a network based on the GAP algorithm [47]. Several alternative methods were designed for end-to-end HSI reconstruction [31], [32], [35], [37], [38]. A self-attention block was embedded into the bottleneck of a U-shaped network, and progressive reconstruction was adopted [37]. Spatial-spectral self-attention was used to capture correlations across different dimensions [31]. A spectral-spatial attention mechanism was designed to extract spectral-spatial dependencies, and a dual domain loss was utilized to improve reconstruction performance [32]. Additionally, a spectral-wise multi-head self-attention approach was proposed to reduce computational complexity, and mask-guided attention was introduced to capture spatial correlations adaptively [35]. Another method employed spectra-aggregation hashing multi-head self-attention to capture self-similarity and reduce computational complexity while incorporating sparsity and coarse patch selection for high-fidelity HSI reconstruction [38]. However, these methods do not consider the degradation pattern between the input HSIs and the ground truth HSIs.

### B. Degradation Estimation

The degradation estimation is widely employed in blind image super-resolution. Based on the probabilistic framework, a degradation model was proposed to better capture the degradation factors and improve the restoration performance [22]. A flexible framework was developed to adapt to various degradations by utilizing learned representations of degradation [24].

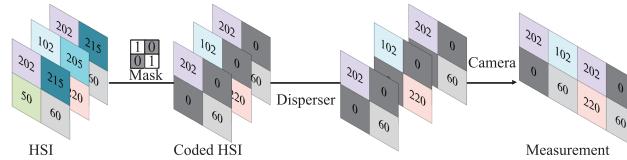


Fig. 1. Imaging pipeline of the CASSI system with a binary mask.

Furthermore, a novel image super-resolution framework was designed to estimate different blur kernels and real noise distributions and achieve high perception [48]. Similarly, the introduction of degradation information is beneficial for HSI reconstruction.

### III. CASSI SYSTEM

The imaging process of the CASSI system is illustrated in Fig. 1.  $\mathbf{X} \in \mathbb{R}^{H \times W \times N}$  represents a 3D HSI cube, where  $H$ ,  $W$  and  $N$  denote the height, width, and number of spectral bands, respectively. Firstly, the HSI cube  $\mathbf{X}$  is modulated by a coded aperture, which is a physical mask denoted as  $\mathbf{M} \in \mathbb{R}^{H \times W}$ . The resulting coded HSI cube after the modulation is represented as  $\mathbf{X}' \in \mathbb{R}^{H \times W \times N}$ . The coding process can be represented as follows:

$$\mathbf{X}'(:,:,n) = \mathbf{X}(:,:,n) \odot \mathbf{M}, n \in [1, \dots, N], \quad (1)$$

where  $\odot$  denotes the element-wise multiplication. To obtain the coded HSI, element-wise multiplication is performed between the HSI and the mask. After passing through the disperser,  $\mathbf{X}'' \in \mathbb{R}^{H \times (W+d(n-1)) \times N}$  is obtained by translating  $\mathbf{X}'$  along the  $y$ -axis, where  $d$  represents the shifting step for adjacent spectral channels. Finally, the measurement  $\mathbf{Y} \in \mathbb{R}^{H \times (W+d(n-1))}$  can be represented as follows:

$$\mathbf{Y} = \sum_{n=1}^N \mathbf{X}''(:,:,n) + \Theta, \quad (2)$$

where  $\Theta$  denotes the imaging noise.

### IV. METHOD

#### A. Degradation Analysis

The CASSI system with a binary mask (described in Fig. 1) is taken as an example. By performing shifting and splitting operations on  $\mathbf{Y}$ , the initialized HSI cube  $\mathbf{X}_{init} \in \mathbb{R}^{H \times W \times N}$  is obtained as the input of our proposed network. As shown in Fig. 2, each band in the initialized HSI can be considered as the corresponding band of the original HSI modulated by a scale factor and adjusted using a bias factor. This bias term is primarily contributed by the other (especially neighboring) bands in the HSI. Therefore, the degradation from the original HSI to the initialized HSI can be expressed as follows:

$$\mathbf{X}_{init} = \mathbf{A} \odot \mathbf{X} + \mathbf{B}, \quad (3)$$

where  $\mathbf{A}$  denotes the scale factor,  $\mathbf{B}$  denotes the bias factor, and  $\odot$  denotes the element-wise multiplication.

It is reasonable to utilize deep learning methods to generate the inverse scale factor  $\bar{\mathbf{A}}$  and the inverse bias factor  $\bar{\mathbf{B}}$  to reverse

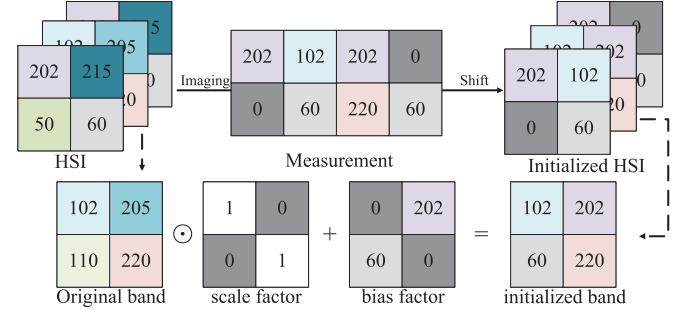


Fig. 2. Degradation analysis between the original and initial hyperspectral images, using a band as an example.

the process mentioned in (3). Consequently, the approximately reconstructed HSI can be represented as follows:

$$\bar{\mathbf{X}} = \bar{\mathbf{A}} \odot \mathbf{X}_{init} + \bar{\mathbf{B}}. \quad (4)$$

Based on this observation, we propose a degradation-aware dynamic Fourier-based network, which predicts the corresponding inverse factors from degraded HSIs and utilizes the predicted factors to process features.

#### B. Overall Architecture

As shown in Fig. 3(a), the proposed DADF-Net is a 3-level encoder-decoder network. Given an initialized HSI  $\mathbf{X}_{init} \in \mathbb{R}^{H \times W \times N}$ ,  $\mathbf{X}_{init}$  is concatenated with a mask  $\mathbf{M} \in \mathbb{R}^{H \times W \times N}$  (copy the physical mask  $N$  times). The concatenated result is then passed through a fusion layer ( $conv1 \times 1$ ) to generate  $\mathbf{X}_{in} \in \mathbb{R}^{H \times W \times N}$ . Firstly,  $\mathbf{X}_{in}$  is fed into an embedding layer ( $conv3 \times 3$ ) to obtain the feature map  $\mathbf{f}_{in} \in \mathbb{R}^{H \times W \times C}$ . Secondly,  $\mathbf{f}_{in}$  undergoes a series of blocks to generate hierarchical features. These blocks consist of  $N_1$  Trans-Fourier blocks, a down-sample layer,  $N_2$  Trans-Fourier blocks, another down-sample layer and  $N_3$  Trans-Fourier blocks. The down-sample layers use a strided  $conv4 \times 4$  to reduce the spatial size and double the number of channels. Next,  $\mathbf{f}_3 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 4C}$  goes through a series of blocks to get  $\mathbf{f}_{out} \in \mathbb{R}^{H \times W \times C}$ . These blocks consist of  $N_1$  Trans-Fourier blocks, an up-sample layer,  $N_2$  Trans-Fourier blocks, and another up-sample layer. In detail, the up-sample layer is a de-convolution layer with kernel size = 2. To alleviate information loss, skip connections are employed. Finally, a mapping layer ( $conv3 \times 3$ ) is used to map the output feature map  $\mathbf{f}_{out}$  to the residual HSI  $\mathbf{X}_{res} \in \mathbb{R}^{H \times W \times N}$ . This residual HSI is added to  $\mathbf{X}_{in}$  to reconstruct the final HSI denoted as:  $\bar{\mathbf{X}} = \mathbf{X}_{in} + \mathbf{X}_{res}$ .

By introducing a multi-stage design to enhance the reconstruction performance of DADF-Net, as shown in Fig. 4, we propose 2-stage (DADF-Plus-2) and 3-stage networks (DADF-Plus-3). Specifically, the reconstructed HSIs from the previous stage are used as input for the next stage. In addition, we employ a  $conv1 \times 1$  layer to integrate the hierarchical features from the decoder and bottleneck layer of the previous stage with the corresponding feature maps at the current stage (e.g.,  $\mathbf{F}_1$  and  $\mathbf{F}'_1$  in Fig. 4). This fusion mechanism helps to reduce information loss between stages.

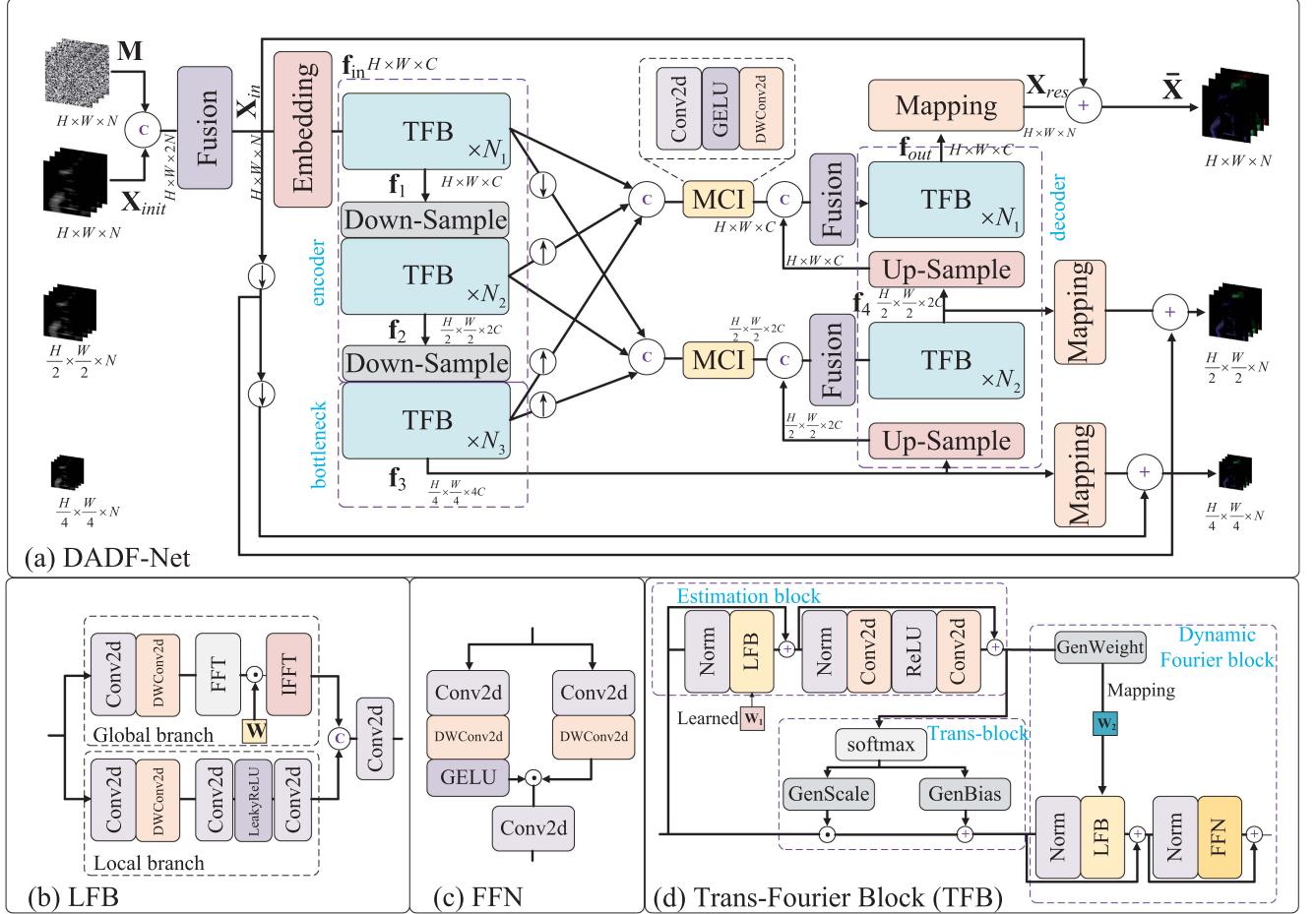


Fig. 3. Overall architecture of our proposed DADF-Net. (a) DADF-Net is a 3-level U-shaped network. (b) The locally enhanced Fourier block (LFB). (c) The Feed-Forward Network (FFN). (d) The Trans-Fourier block (TFB) consists of an estimation block, a Trans-block, and a dynamic Fourier block.

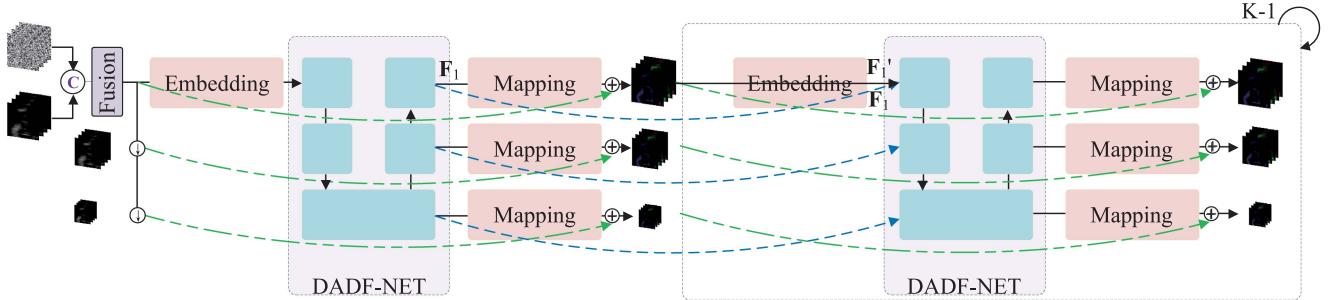


Fig. 4. Architecture of our proposed DADF-Net with multi-stage design.  $K$  denotes the number of stages.

### C. Locally Enhanced Fourier Block

Given the importance of non-local features in HSI reconstruction and the difficulty of designing networks with large convolution kernels, we extract the non-local features in the Fourier domain. Inspired by [49], [50], as depicted in Fig. 5, the convolution operation in the image domain can be equivalently represented as the element-wise product in the Fourier domain:

$$\mathcal{F}(\mathbf{X} * \mathbf{K}) = \mathcal{F}(\mathbf{X}) \odot \mathcal{F}(\mathbf{K}), \quad (5)$$

where  $\mathbf{X}$  denotes the input signal,  $\mathbf{K}$  denotes the convolution kernel,  $*$  denotes the convolution operation,  $\mathcal{F}$  denotes the Fourier transform and  $\odot$  denotes the element-wise product.

Based on (5), we can set  $\mathbf{W}$  as a learned filter to extract non-local features in the Fourier domain. Subsequently, the features in the Fourier domain are converted back to the image domain using the inverse Fourier transform. This process can be represented as follows:

$$\mathbf{f} = \mathcal{F}^{-1}(\mathbf{W} \odot \mathcal{F}(\mathbf{X})), \quad (6)$$

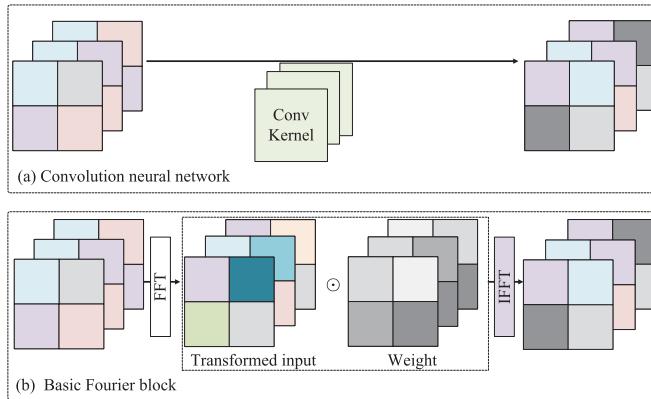


Fig. 5. (a) The process of feature extraction using a convolutional block. (b) The process of feature extraction using a basic Fourier block.

where  $\mathcal{F}^{-1}$  denotes the inverse Fourier transform,  $\mathbf{W}$  denotes the filter weight and  $\mathbf{f}$  denotes the output feature map.

As shown in Fig. 3(b), we propose a locally enhanced Fourier block (LFB) with a convolution operation and a two-branch design, based on the basic Fourier block. In the global branch, the input  $\mathbf{t} \in \mathbb{R}^{w \times h \times c}$  goes through a  $conv1 \times 1$  layer followed by a depth-wise  $conv3 \times 3$  layer. The result then passes through the Fourier block using the given filter weight  $\mathbf{W}$  to generate the non-local features. In the local branch, the input  $\mathbf{t}$  goes through a  $conv1 \times 1$  layer and a depth-wise  $conv3 \times 3$  layer. The result then undergoes a  $conv3 \times 3$ , an activation layer, and another  $conv3 \times 3$  layer to generate the local features. Finally, LFB concatenates these features and feeds the concatenated result into a fusion layer ( $conv1 \times 1$ ) to generate the locally enhanced feature  $\mathbf{T} \in \mathbb{R}^{w \times h \times c}$ .

#### D. Feed-Forward Network

For the Feed-Forward Network (FFN) shown in Fig. 3(c), inspired by [51], [52], we utilize a gating mechanism to control feature generation. The input  $\mathbf{G} \in \mathbb{R}^{w \times h \times c}$  passes through two parallel branches. One branch consists of a  $conv1 \times 1$  and a depth-wise  $conv3 \times 3$ , which encodes the features; the other branch consists of a  $conv1 \times 1$ , a depth-wise  $conv3 \times 3$ , and an activation layer  $GELU$ , which generates the gating signal. Element-wise product is performed on the results from these two branches. Finally, a convolution layer is used to generate the final features  $\hat{\mathbf{G}} \in \mathbb{R}^{w \times h \times c}$ .

#### E. Trans-Fourier Block

Inspired by (3) and recent image super-resolution [24], [53], as shown in Fig. 3(d), we propose a Trans-Fourier block (TFB), which consists of an estimation block, a Trans-block, and a dynamic Fourier block. The estimation block estimates the degradation feature map, which contains the degradation information of the input. The inverse factors required by the Trans-block and the filter weight of the subsequent module are generated using the degradation feature map.

Firstly, given an input feature  $\mathbf{F} \in \mathbb{R}^{w \times h \times c}$ , it passes through the estimation block to generate the degradation feature map

$\mathbf{K} \in \mathbb{R}^{w \times h \times c}$ . The estimation block consists of a layer normalization, a LFB, a layer normalization, a  $conv3 \times 3$ , an  $ReLU$ , and a  $conv3 \times 3$ . It is worth noting that the filter weight of the LFB within the estimation block is generated by learning.

Then,  $\mathbf{K}$  passes through a softmax layer and two parallel branches (the GenScale module and GenBias module) to generate the inverse scale factor  $\hat{\mathbf{A}} \in \mathbb{R}^{w \times h \times c}$  and the inverse bias factor  $\hat{\mathbf{B}} \in \mathbb{R}^{w \times h \times c}$ . The refined feature  $\mathbf{F}_{refine} \in \mathbb{R}^{w \times h \times c}$  can be represented as follows:

$$\mathbf{F}_{refine} = \hat{\mathbf{A}} \odot \mathbf{F} + \hat{\mathbf{B}}. \quad (7)$$

In detail, the GenScale and GenBias modules have the same structure, consisting of a  $conv3 \times 3$  layer, an activation layer, and a  $conv3 \times 3$  layer.

Moreover, to realize the adaptive processing of the input and improve the modeling ability of the network for feature,  $\mathbf{K}$  also passes through the GenWeight module to dynamically generate filter weight  $\mathbf{W}_2$  for the LFB within the dynamic Fourier block. Specifically, the GenWeight module consists of a  $conv3 \times 3$ , an activation layer, and a  $conv3 \times 3$ . The refined feature  $\mathbf{F}_{refine}$  is then normalized using layer normalization, and the normalized result is passed to the LFB. The LFB processes the feature using the dynamically generated filter weight  $\mathbf{W}_2$ . This process can be represented as follows:

$$\mathbf{F}'_{refine} = \mathbf{F}_{refine} + \text{LFB}(\text{LayerNorm}(\mathbf{F}_{refine})), \quad (8)$$

where LFB denotes the locally enhanced Fourier block, and LayerNorm denotes the layer normalization module.

Next,  $\mathbf{F}'_{refine} \in \mathbb{R}^{w \times h \times c}$  undergoes a layer normalization and a FFN to generate the final feature  $\mathbf{F}' \in \mathbb{R}^{w \times h \times c}$ . This process can be represented as follows:

$$\mathbf{F}' = \mathbf{F}'_{refine} + \text{FFN}(\text{LayerNorm}(\mathbf{F}'_{refine})), \quad (9)$$

where FFN denotes the Feed-Forward Network.

Overall, the LFB focuses on modeling non-local and local features, while the FFN controls the information within the feature.

#### F. Multi-Scale Design

We introduce the multi-scale design to enhance the reconstruction capability of DADF-Net. This design comprises two components: multi-scale context interaction and multi-scale loss constraints.

The features of each encoder layer and the bottleneck layer are resized to the same resolution scale through up-sampling or down-sampling. These features of different scales are then aggregated using the multi-scale context interaction (MCI) module to reduce the loss of information. The MCI module consists of a  $conv1 \times 1$  layer, a  $GELU$  activation layer, and a depth-wise  $conv3 \times 3$  layer.

Inspired by [54], we use down-sampling to generate degraded HSIs at different scales from a degraded input  $\mathbf{X}_{in} \in \mathbb{R}^{H \times W \times N}$ . The outputs from the bottleneck layer and different decoder layers pass through their respective mapping layers to obtain residual HSIs at different scales. The residual HSIs at different scales are added to their corresponding HSIs at each scale to obtain the

reconstructed HSIs at different scales. The objective function of the network is as follows:

$$L_{MRMSE} = \sum_{b=1}^B \text{RMSE}(\hat{\mathbf{x}}_b, \mathbf{x}_b), \quad (10)$$

where RMSE denotes the Root Mean Square Error,  $\mathbf{x}_b$  and  $\hat{\mathbf{x}}_b$  represent the ground truth and reconstructed HSI at a specific scale, respectively.  $B$  is set to 3, and  $b$  represents a specific scale (where  $b = 1, 2, 3$  corresponds to  $\frac{1}{4}, \frac{1}{2}$ , and original resolution, respectively).

## V. EXPERIMENTS

### A. Experimental Settings

We validate the effectiveness of the proposed method on both simulation and real-world datasets. Following the setting of TSA-Net [31], we adopt 28 bands from 450 nm to 650 nm by spectral interpolation for HSIs of public dataset CAVE [55] and KAIST [45].

*Simulation HSI Datasets:* For the experiment on simulation dataset, we use CAVE [55] with a size of  $512 \times 512 \times 31$  and KAIST [45] with a size of  $2704 \times 3376 \times 31$ . Following TSA-Net [31], we adopt CAVE for training and 10 selected scenes from KAIST for testing.

*Real HSI Data:* For the experiment on real-world scenes, following TSA-Net [31], we adopt real compressive measurements collected by the CASSI system reported in [31] with a spatial size of  $660 \times 714$ .

*Evaluation Metrics:* To evaluate the performance of the proposed method and the other open-source state-of-the-art (SOTA) methods, we adopt peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [56], feature similarity (FSIM) [57], visual information fidelity (VIF) [58], spectral angle mapper (SAM), and log frequency distance (LFD) as metrics to compare different methods.

*Implementation Details:* We implemented our proposed method in PyTorch. The model is trained using the ADAM optimizer [59] with the Cosine Annealing scheme [60] for 300 epochs. The learning rate is initially set to  $6 \times 10^{-4}$ . The batch size is 2. Random flipping and rotation are used for data augmentation. The DADF-Net is trained on a single RTX 2080Ti GPU. The training takes 2.5 days, and the average testing time is 2.70 seconds.

### B. Ablation Study

In this section, we conduct ablation experiments on the public simulation HSI dataset to investigate the effectiveness of each proposed module.

- The baseline model (Case 1) is a U-shaped network using the basic Fourier block with inputting the initialized HSI  $\mathbf{H}$  and mask  $\mathbf{M}$ .
- Case 2 incorporates the locally enhanced Fourier block into the baseline model.
- Case 3 integrates the Trans-Fourier block into the baseline model.

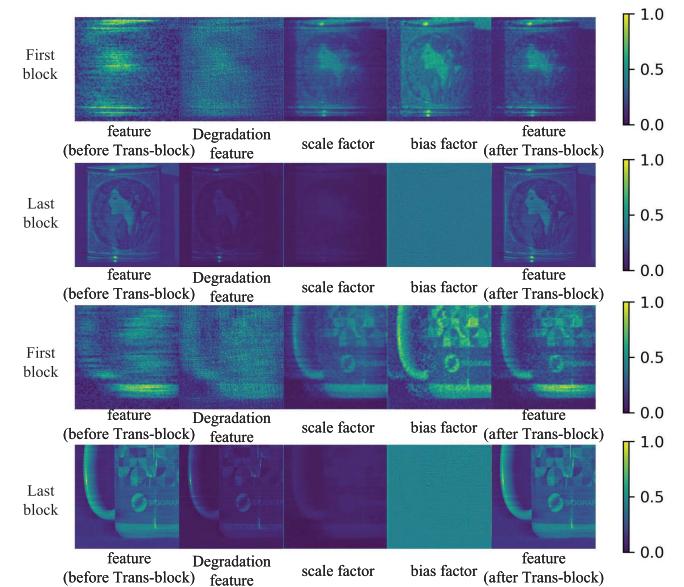


Fig. 6. Visualization of the feature (before Trans-block), the predicted degradation feature, the predicted inverse scale factor, the predicted inverse bias factor and the feature (after Trans-block).

- Case 4 introduces a multi-scale design to the baseline model.
- Case 5 integrates the locally enhanced Fourier block and the Trans-Fourier block into the baseline model.

The results are presented in Table I. The baseline model achieves 36.41 dB in PSNR. When the locally enhanced Fourier block is applied (Case 2), it improves the PSNR by 0.54 dB and the SSIM by 0.005 compared to the baseline model. Similarly, the Trans-Fourier block independently (Case 3) leads to an improvement of 0.46 dB in PSNR and a gain of 0.004 in SSIM, highlighting the effectiveness of the Trans-Fourier block. When the multi-scale design is applied independently (Case 4), there is only a slight improvement of 0.03 dB in PSNR compared to the baseline model. This might be because the Fourier blocks extract features within the frequency domain and are less sensitive to spatial information. However, when the locally enhanced Fourier block and the Trans-Fourier block are incorporated together (Case 5), there is an improvement of 0.66 dB in PSNR compared to the baseline model. Nevertheless, the gain is not significant compared to Case 2 (0.12 dB higher) and Case 3 (0.20 dB higher), suggesting the lack of effective constraints for feature modeling. Furthermore, incorporating the multi-scale design into Case 5 significantly improves performance. This is because the multi-scale design provides information interaction and loss constraints at multiple scales, thereby enhancing the ability of the network to extract features effectively. These techniques enable our approach to achieve desirable performance.

To further illustrate the advantages of the Trans-Fourier block, Fig. 6 visualizes the input features, predicted degradation features, predicted inverse scale factors, predicted inverse bias factors, and refined features in the first and last Trans-Fourier blocks of Case 3. It can be observed that the input features to the Trans-Fourier block are blurry. After the linear transformation

TABLE I  
ABLATION EXPERIMENTS ON SIMULATION DATASET

Method	baseline	Locally enhanced	Trans-Fourier block	Multi-Scale	PSNR	SSIM	SAM	GFLOPS	Params
Case 1	✓	✗	✗	✗	36.41	0.959	5.89	8.90	13.87
Case 2	✓	✓	✗	✗	36.95	0.964	5.33	22.15	14.95
Case 3	✓	✗	✓	✗	36.87	0.963	5.18	49.07	17.14
Case 4	✓	✗	✗	✓	36.44	0.958	5.98	9.60	13.89
Case 5	✓	✓	✓	✗	37.07	0.964	5.22	75.55	19.29
Ours	✓	✓	✓	✓	<b>37.66</b>	<b>0.967</b>	5.22	76.62	19.35

The quantitative results are the average PSNR, SSIM, and SAM of 10 simulation scenes. The computational complexity and number of parameters under different cases are reported.

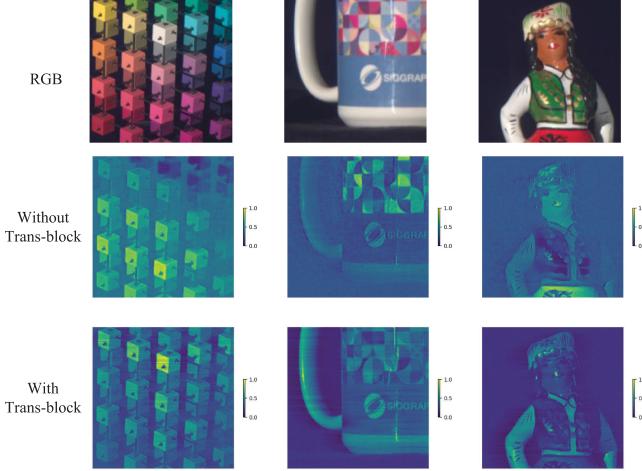


Fig. 7. Visualization of input features of the last Fourier block. The top row shows the RGB. The middle and bottom rows show the features without and with Trans-block, respectively. The introduction of the Trans-block enables input features to be clearer.

using the predicted factors in the trans-block of the TFB, the features become clearer, which benefits the subsequent modules in capturing the information within the features. Additionally, comparing the predicted factors in the first and last TFBs, it is evident that the last TFB captures less degradation information due to its clearer input features, resulting in weaker predicted factor strengths. Furthermore, Fig. 7 visualizes the inputs to the last feature processing module in Case 3 and Case 1 (baseline model). It can be seen that the introduction of the Trans-Fourier block provides clearer features to the subsequent submodules.

### C. Comparison Results

We compare our proposed method with the other SOTA methods, including two model-based methods (GAP-TV [16], De-SCI [18]), six CNN-based methods ( $\lambda$ -Net [37], TSA-Net [31], DGSMP [29], ADMM-Net [61], GAP-Net [46], HDNet [32]) and recent Transformer-based methods (GAP-CCoT [47], MST [35], CST [38], DAUHST [62]). To be fair, all comparative experiments are kept the same settings as DGSMP [29].

*Simulation HSI Reconstruction:* The PSNR, SSIM, FSIM, VIF, and SAM results of the different methods on the 10 scenes in the simulation dataset are listed in Table II. It can be observed that our basic DADF-Net achieves the PSNR of 37.66 dB, which is 1.58 dB higher than CST-L-Plus. The two-stage model (DADF-Plus-2) achieves PSNR of 37.99 dB, surpassing the

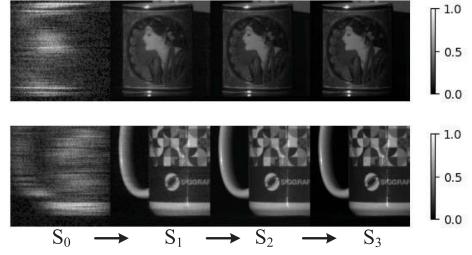


Fig. 8. Visualization of the results from each stage in the three-stage model (DADF-Plus-3).  $S_i$  denotes the result of  $i$ -th stage.  $S_0$  denotes the input of network.

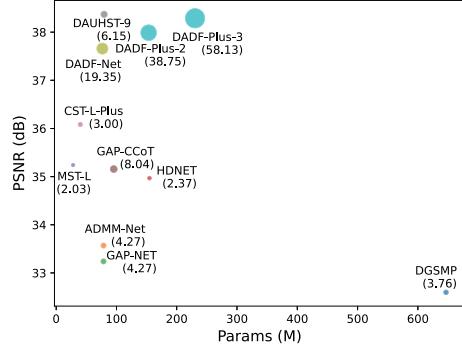


Fig. 9. PSNR-Params-GFLOPS comparisons with the other open-source SOTA methods for HSI reconstruction. The vertical axis is the peak signal-to-noise ratio (PSNR in dB performance), the horizontal axis is the GFLOPS (computational cost), and the circle radius is the number of parameters (memory cost).

basic DADF-Net by 0.33 dB. The three-stage model (DADF-Plus-3) reaches PSNR of 38.29 dB, comparable to DAUHST-9. Regarding FSIM, the basic DADF-Net surpasses CST-L-Plus by 0.004 and is on par with DAUHST-5. As for VIF, DADF-Net outperforms DAUHST-5 by 0.012. DADF-Plus-2 performs on par with DAUHST-9 in FSIM and outperforms DAUHST-9 by 0.009 in VIF. DADF-Plus-3 outperforms DAUHST-9 by 0.001 in FSIM and by 0.019 in VIF. We can observe that as the number of stages increases, the PSNR performance of our method gradually saturates. Fig. 8 illustrates the results of different stages in DADF-Plus-3. As the number of stages increases, the reconstruction quality is improved. It also indicates that the difficulty of obtaining degradation information increases, leading to reconstruction performance saturation of the proposed method. Table III reports the computation cost (giga floating-point operations per second, GFLOPS), memory cost (Params), and average reconstruction time for each method. To visually compare

TABLE II  
QUANTITATIVE RESULTS ON 10 SCENES IN SIMULATION

Method		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	avg	
GAP-TV [16]		PSNR SSIM FSIM VIF SAM	27.20 0.766 0.829 0.141 13.10	26.13 0.686 0.819 0.101 16.74	27.83 0.806 0.884 0.190 13.32	38.16 0.920 0.932 0.223 11.63	24.86 0.736 0.840 0.158 14.07	23.91 0.681 0.816 0.099 23.27	24.88 0.715 0.792 0.121 13.83	23.25 0.644 0.797 0.102 29.63	26.13 0.743 0.832 0.135 13.63	24.88 0.601 0.755 0.070 27.41	26.72 0.730 0.830 0.134 17.66
		PSNR SSIM FSIM VIF SAM	27.31 0.764 0.811 0.111 14.61	25.85 0.669 0.801 0.080 18.07	29.79 0.897 0.911 0.248 8.42	38.36 0.950 0.928 0.198 8.77	25.49 0.794 0.842 0.189 11.54	24.52 0.772 0.828 0.104 15.77	25.87 0.806 0.812 0.146 10.77	23.40 0.753 0.813 0.096 21.04	24.57 0.809 0.837 0.145 11.40	24.57 0.679 0.739 0.057 17.59	27.36 0.789 0.832 0.137 13.80
		PSNR SSIM FSIM VIF SAM	29.71 0.830 0.867 0.198 14.13	27.70 0.742 0.831 0.142 17.41	29.53 0.846 0.898 0.255 15.60	37.53 0.911 0.913 0.277 24.04	26.60 0.790 0.863 0.221 16.35	27.25 0.787 0.862 0.211 26.04	26.61 0.782 0.825 0.182 14.08	26.20 0.781 0.849 0.169 27.57	28.54 0.798 0.832 0.177 15.86	26.14 0.701 0.789 0.115 26.02	28.58 0.797 0.853 0.191 19.71
		PSNR SSIM FSIM VIF SAM	32.31 0.898 0.914 0.335 8.74	31.07 0.863 0.908 0.347 10.35	32.30 0.918 0.940 0.370 7.39	39.53 0.959 0.950 0.382 8.37	29.44 0.887 0.919 0.344 6.72	31.06 0.905 0.936 0.349 9.70	30.26 0.883 0.896 0.328 7.66	29.31 0.893 0.920 0.308 11.38	31.62 0.912 0.916 0.373 7.67	29.20 0.867 0.897 0.274 9.56	31.61 0.898 0.920 0.341 8.75
		PSNR SSIM FSIM VIF SAM	33.35 0.920 0.936 0.428 9.22	31.66 0.892 0.929 0.484 11.92	32.92 0.925 0.951 0.534 7.74	40.39 0.970 0.964 0.457 7.74	29.46 0.894 0.931 0.459 9.88	32.74 0.938 0.954 0.403 8.28	32.38 0.898 0.914 0.413 8.08	30.68 0.932 0.944 0.454 10.91	33.70 0.925 0.929 0.403 8.34	30.64 0.933 0.938 0.403 7.37	32.60 0.923 0.939 0.443 8.95
ADMM-Net [61]		PSNR SSIM FSIM VIF SAM	34.09 0.924 0.933 0.396 8.60	33.58 0.904 0.934 0.463 13.05	35.03 0.935 0.947 0.416 8.13	41.25 0.972 0.954 0.434 9.37	31.79 0.926 0.942 0.419 7.90	32.52 0.929 0.930 0.402 12.76	32.38 0.900 0.906 0.405 8.34	30.68 0.912 0.926 0.399 16.89	33.70 0.921 0.920 0.393 8.29	30.20 0.905 0.913 0.335 12.83	33.57 0.923 0.932 0.403 10.62
		PSNR SSIM FSIM VIF SAM	33.70 0.917 0.930 0.388 9.11	33.23 0.902 0.933 0.445 13.07	34.27 0.933 0.951 0.414 8.60	41.13 0.972 0.954 0.420 9.55	31.41 0.926 0.941 0.410 9.37	32.39 0.929 0.931 0.402 7.87	32.23 0.918 0.922 0.404 12.61	31.32 0.912 0.925 0.397 8.40	31.53 0.922 0.929 0.399 16.08	31.51 0.923 0.932 0.335 13.23	32.60 0.922 0.933 0.396 10.73
		PSNR SSIM FSIM VIF SAM	35.11 0.940 0.930 0.465 7.38	35.65 0.943 0.933 0.563 8.16	36.05 0.947 0.946 0.472 5.65	42.47 0.954 0.954 0.521 5.82	32.67 0.942 0.941 0.420 5.07	34.47 0.939 0.941 0.400 6.86	33.64 0.948 0.949 0.404 6.55	32.44 0.948 0.949 0.445 8.15	34.85 0.947 0.949 0.445 6.35	32.34 0.943 0.948 0.443 6.68	34.97 0.948 0.953 0.435 6.68
		PSNR SSIM FSIM VIF SAM	35.44 0.946 0.951 0.494 7.03	36.12 0.949 0.961 0.506 8.12	36.39 0.955 0.965 0.544 6.09	42.05 0.977 0.970 0.544 7.42	32.94 0.950 0.961 0.506 5.80	34.71 0.957 0.965 0.514 7.85	32.38 0.952 0.957 0.486 6.27	30.68 0.953 0.957 0.469 10.34	33.47 0.947 0.949 0.448 7.47	30.20 0.940 0.949 0.450 8.36	33.24 0.942 0.953 0.450 7.47
		PSNR SSIM FSIM VIF SAM	35.17 0.938 0.946 0.473 7.00	35.90 0.948 0.953 0.605 8.19	36.91 0.958 0.965 0.515 5.12	42.25 0.977 0.967 0.552 6.17	32.61 0.957 0.959 0.496 4.97	34.95 0.957 0.966 0.520 4.76	33.46 0.952 0.966 0.473 5.56	33.13 0.952 0.957 0.481 6.78	35.75 0.954 0.955 0.530 5.76	32.43 0.941 0.950 0.441 6.80	35.26 0.950 0.957 0.509 6.58
CST-L-Plus [38]		PSNR SSIM FSIM VIF SAM	35.86 0.954 0.959 0.529 6.40	36.80 0.957 0.969 0.617 6.47	38.23 0.982 0.972 0.544 5.78	42.56 0.958 0.966 0.575 4.69	33.16 0.958 0.966 0.546 5.92	35.74 0.958 0.964 0.542 5.92	34.84 0.958 0.965 0.526 5.44	34.26 0.958 0.965 0.526 6.71	36.41 0.958 0.965 0.541 5.46	33.00 0.952 0.961 0.542 5.71	36.08 0.961 0.965 0.542 5.71
		PSNR SSIM FSIM VIF SAM	35.89 0.948 0.950 0.488 6.14	36.68 0.949 0.963 0.576 7.12	37.97 0.954 0.967 0.546 4.34	44.58 0.972 0.972 0.575 5.87	34.12 0.957 0.957 0.542 4.37	35.44 0.957 0.957 0.542 3.62	34.74 0.957 0.957 0.542 4.39	33.61 0.957 0.957 0.542 7.84	37.38 0.960 0.960 0.545 4.90	33.03 0.948 0.957 0.545 5.60	36.35 0.957 0.960 0.540 6.00
		PSNR SSIM FSIM VIF SAM	36.56 0.953 0.956 0.488 5.80	37.92 0.960 0.963 0.576 6.76	39.36 0.968 0.969 0.567 5.36	44.97 0.972 0.972 0.589 5.36	34.82 0.966 0.966 0.585 3.96	36.22 0.972 0.972 0.576 6.23	35.99 0.962 0.962 0.563 4.39	34.24 0.962 0.962 0.553 7.84	38.50 0.961 0.961 0.557 4.65	33.63 0.952 0.953 0.547 5.46	37.22 0.961 0.964 0.547 5.54
		PSNR SSIM FSIM VIF SAM	36.88 0.959 0.962 0.555 5.43	38.51 0.965 0.965 0.655 6.40	40.55 0.971 0.971 0.567 3.45	45.29 0.986 0.986 0.589 5.57	35.33 0.967 0.967 0.585 3.62	36.59 0.974 0.974 0.585 3.65	36.78 0.974 0.974 0.587 3.65	34.71 0.974 0.974 0.587 4.39	38.67 0.974 0.974 0.587 7.84	34.24 0.974 0.974 0.587 4.90	37.75 0.967 0.969 0.577 5.43
		PSNR SSIM FSIM VIF SAM	37.22 0.962 0.965 0.575 5.20	39.02 0.969 0.977 0.683 6.04	41.09 0.975 0.975 0.611 3.45	46.35 0.987 0.987 0.605 3.57	35.81 0.974 0.974 0.612 3.62	37.12 0.974 0.974 0.612 4.39	35.06 0.974 0.974 0.587 3.45	39.98 0.974 0.974 0.587 4.72	34.56 0.974 0.974 0.587 6.20	38.37 0.971 0.971 0.587 4.09	37.66 0.971 0.971 0.587 5.10
DADF-Net		PSNR SSIM FSIM VIF SAM	36.95 0.959 0.962 0.575 5.58	38.90 0.971 0.977 0.681 5.62	39.47 0.987 0.987 0.611 3.68	45.39 0.987 0.987 0.620 3.45	35.22 0.974 0.974 0.605 3.57	36.56 0.974 0.974 0.585 3.45	36.17 0.974 0.974 0.585 3.17	34.77 0.974 0.974 0.585 4.72	39.51 0.974 0.974 0.585 6.20	34.06 0.974 0.974 0.585 4.65	37.99 0.971 0.971 0.585 5.22
		PSNR SSIM FSIM VIF SAM	37.39 0.963 0.967 0.589 5.22	39.66 0.974 0.976 0.691 5.27	40.26 0.987 0.987 0.571 3.46	45.53 0.987 0.987 0.620 4.44	35.46 0.974 0.974 0.601 3.45	36.75 0.974 0.974 0.585 3.17	35.06 0.974 0.974 0.585 4.72	39.98 0.974 0.974 0.585 6.20	34.56 0.974 0.974 0.585 4.09	38.37 0.971 0.971 0.585 5.10	37.66 0.971 0.971 0.585 4.71
		PSNR SSIM FSIM VIF SAM	37.46 0.965 0.968 0.599 5.18	39.86 0.974 0.978 0.615 5.10	41.03 0.987 0.989 0.654 3.08	45.98 0.987 0.989 0.620 4.46	35.53 0.974 0.974 0.601 3.59	37.02 0.974 0.974 0.585 3.52	36.76 0.974 0.974 0.585 4.43	34.78 0.974 0.974 0.585 6.13	40.07 0.974 0.974 0.585 4.14	34.39 0.974 0.974 0.585 5.38	38.29 0.972 0.972 0.585 4.67
		PSNR SSIM FSIM VIF SAM	37.47 0.968 0.971 0.603 5.19	39.87 0.975 0.978 0.618 5.12	41.07 0.987 0.989 0.642 3.46	45.54 0.987 0.989 0.620 4.44	35.53 0.974 0.974 0.601 3.45	37.02 0.974 0.974 0.585 3.17	36.76 0.974 0.974 0.585 4.72	34.78 0.974 0.974 0.585 6.20	40.07 0.974 0.974 0.585 4.09	34.39 0.974 0.974 0.585 5.38	38.29 0.972 0.972 0.585 4.67
		PSNR SSIM FSIM VIF SAM	37.48 0.968 0.971 0.603 5.19	39.87 0.975 0.978 0.618 5.12	41.07 0.987 0.989 0.642 3.46	45.54 0.987 0.989 0.620 4.44	35.53 0.974 0.974 0.601 3.45	37.02 0.974 0.974 0.585 3.17	36.76 0.974 0.974 0.585 4.72	34.78 0.974 0.974 0.585 6.20	40.07 0.974 0.974 0.585 4.09	34.39 0.974 0.974 0.585 5.38	38.29 0.972 0.972 0.585 4.67

PSNR, SSIM, FSIM, VIF, AND SAM are reported. The higher the PSNR, SSIM, FSIM, AND VIF, the Better. The lower the SAM, the Better.

TABLE III  
GFLOPS-PARAMS-RECONSTRUCTION TIME (CPU WITH 16 GB) COMPARISONS WITH THE OPEN-SOURCE METHODS

Method	GAP-TV	DeSCI	$\lambda$ -Net	TSA-Net	DGSMP	ADMM-Net	GAP-NET	HDNET	MST-L	GAP-CCoT	CST-L-Plus	DAUHST-2	DAUHST-3	DAUHST-5	DAUHST-9	DADF-Net	DADE-Plus-2	DADE-Plus-3
GFLOPS	-	-	117.98	110.06	646.65	78.58	78.58	154.76	28.15	95.52	40.10	18.44	27.17	44.61	79.50	76.62	153.51	230.41
Params (M)	-	-	62.64	44.25	3.76	4.27	4.27	2.37	2.03	8.04	3.00	1.40	2.08	3.44	6.15	19.35	38.75	58.13
Time (s)	117.53	5602.93	0.48	1.48	8.12	1.45	1.32	1.73	2.08	2.69	2.51	1.21	1.74	2.88	30.44/7.32 (40GB)	2.70	9.05 (40GB)	11.23 (40GB)

The lower the better.

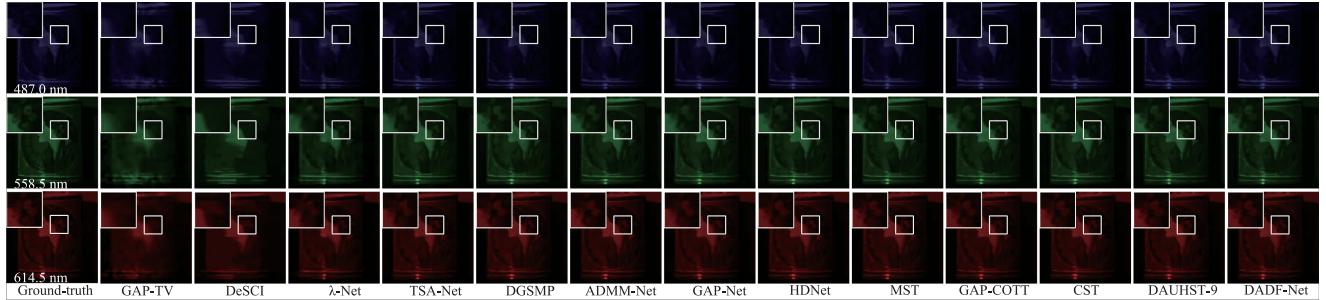


Fig. 10. Reconstructed simulation HSI comparisons of Scene 1 with 3 out of 28 spectral channels (wavelength: 487.0 nm, 558.5 nm and 614.5 nm), including the 11 SOTA methods and ours. Zoom in for a better view.

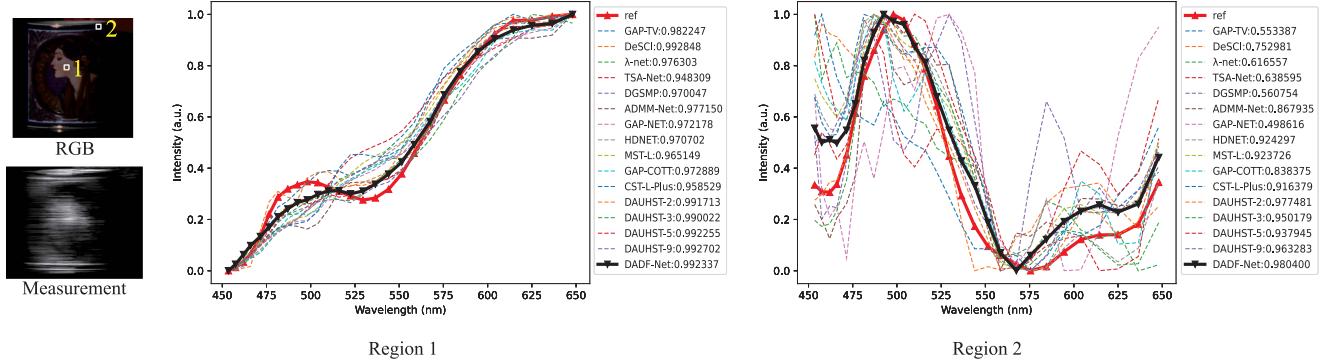


Fig. 11. Spectral curves of Scene 1 are corresponding to the selected white box of the RGB image, including the 11 SOTA methods and ours.

the relationship between computation cost (GFLOPS), memory cost (Params), and PSNR, Fig. 9 presents visualizations of the results obtained by different methods. When DADF-Net achieves the PSNR of 37.66 dB, the computational complexity is 76.62 GFLOPS, with 19.35 M parameters. The reconstruction time on Windows 11 with an Intel 10400F processor and 16 GB of memory is 2.70 seconds. The large number of parameters in DADF-Net is due to the need to store the weights of Fourier-domain filters. It is worth noting that under 16 GB memory, DAUHST-9 requires 30.44 seconds for reconstruction, indicating that DAUHST-9 may require a larger memory footprint. While the performance of the proposed method is improved when incorporating the multi-stage design, it is critical to consider factors such as computational complexity, memory cost, and reconstruction time. In this regard, the DADF-Net without the multi-stage design balances performance and cost.

Fig. 10 visualizes the reconstruction results of different methods for *Scene 1* in the simulation dataset across 3 out of 28 bands (487.0 nm, 558.5 nm, and 614.5 nm). The zoomed-in views allow for better observation. Model-based methods yield unpleasant artifacts, while previous deep-learning methods can result in

undesired effects. In contrast, DADF-Net provides a clearer reconstruction with more restored details. This advantage stems from the capability of DADF-Net to capture both local and non-local information while adaptively modeling features. The average spectral correlation curves of regions with a size of  $10 \times 10$  are shown in Fig. 11. Our method demonstrates the ability to provide more accurate or comparable spectral reconstructions. We also compare the performance of different methods in reconstructing frequency-domain information. Table IV reports LDF to evaluate the spectrum difference between the reconstructed and ground truth HSI. Fig. 12 visualizes the frequency spectrum of the reconstruction results obtained by each method. The results highlight the superiority of our proposed method in preserving frequency-domain information.

DAUHST and our method are both based on degradation-aware for HSI reconstruction and superior to the other SOTA methods on performance. In the initialization stage of HSI reconstruction, DAUHST predicts the auxiliary variables needed in the reconstruction process according to the degraded HSI. It ignores the variation of degradation information between stages in the HSI reconstruction process. Unlike DAUHST, our proposed method dynamically generates the desired degradation

TABLE IV  
LOG FREQUENCY DISTANCE (LFD) ON 10 SCENES IN SIMULATION

Method	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	avg
GAP-TV	15.91	16.19	15.83	13.74	16.45	16.67	16.47	16.84	16.2	16.45	16.07
DeSCI	15.88	16.22	15.32	13.35	16.30	16.52	16.23	16.80	15.67	16.52	15.88
$\lambda$ -Net	15.33	15.80	15.38	13.81	16.04	15.89	16.06	16.15	15.63	16.16	15.62
TSA-Net	14.73	15.03	14.77	13.44	15.39	15.02	15.26	15.45	15.00	15.45	14.95
DGSMP	14.54	14.89	14.91	13.56	15.40	14.69	15.12	15.02	15.06	14.96	14.81
ADMM-Net	14.34	14.44	14.15	13.09	14.84	14.70	14.80	15.15	14.49	15.14	14.51
GAP-NET	14.43	14.53	14.35	13.16	14.93	14.74	14.85	15.21	14.56	15.23	14.60
HDNET	14.12	13.97	13.99	12.59	14.67	14.27	14.51	14.74	14.31	14.75	14.19
MST-L	14.05	13.85	13.96	12.95	14.61	14.22	14.39	14.63	14.26	14.66	14.16
GAP-CCoT	14.13	13.95	13.89	13.09	14.69	14.17	14.62	14.61	14.18	14.75	14.21
CST-L-Plus	13.97	13.70	13.49	12.65	14.59	13.98	14.22	14.33	13.97	14.62	13.95
DAUHST-2	13.92	13.72	13.46	12.06	14.33	14.03	14.25	14.47	13.64	14.59	13.85
DAUHST-3	13.77	13.44	13.13	11.99	14.16	13.87	13.97	14.32	13.38	14.46	13.65
DAUHST-5	13.70	13.30	12.84	11.83	14.05	13.77	13.78	14.20	13.35	14.32	13.51
DAUHST-9	13.62	13.18	12.71	11.52	13.93	13.65	13.59	14.12	13.01	14.25	13.36
DADF-Net	13.69	13.21	13.14	11.77	14.08	13.79	13.92	14.27	13.17	14.33	13.54
DADF-Plus-2	13.59	13.04	12.96	11.82	14.04	13.75	13.83	14.20	13.14	14.37	13.47
DADF-Plus-3	13.57	13.00	12.74	11.65	14.03	13.69	13.77	14.19	12.99	14.30	13.39

The lower the better.

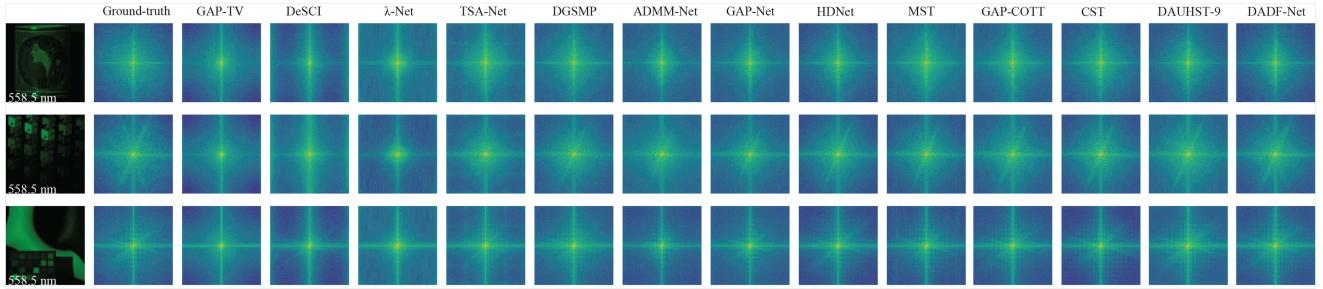


Fig. 12. Frequency spectrum visualization of Scene 1, 2, and 3, including the 11 SOTA methods and ours.

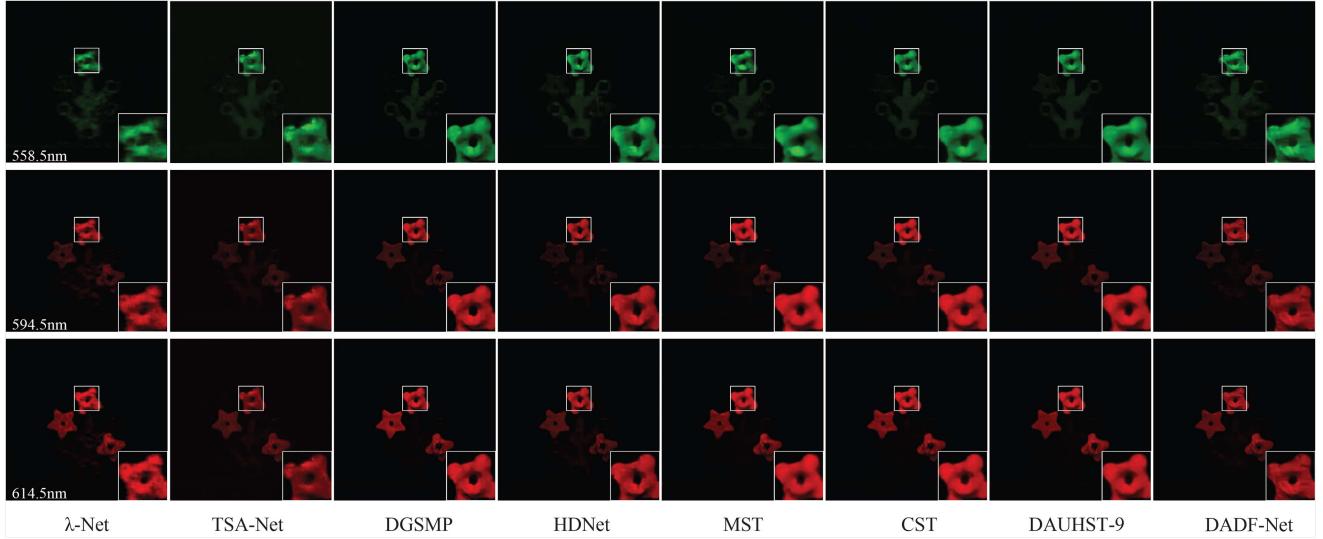


Fig. 13. Reconstructed real HSI comparisons of Scene 1 with 3 out of 28 spectral channels (wavelength: 558.5 nm, 594.5 nm and 614.5 nm), including the 6 SOTA methods and ours. Zoom in for a better view.

factors based on the current features during reconstruction. It increases the flexibility of the network, leading to improved network performance. The performance of our method is higher than DAUHST-2 and DAUHST-3 and comparable to DAUHST-5, and our method outperforms DAUHST-5 in SAM. The computational cost and reconstruction time of DADF-Net are

comparable to DAUHST-5. However, the performance of our approach is slightly weaker than that of DAUHST-9. This may be because DAUHST is designed as a multi-stage network, where each stage consists of a linear projection module and a U-shaped denoising network, and the weights are not shared between phases. It can iteratively optimize the intermediate

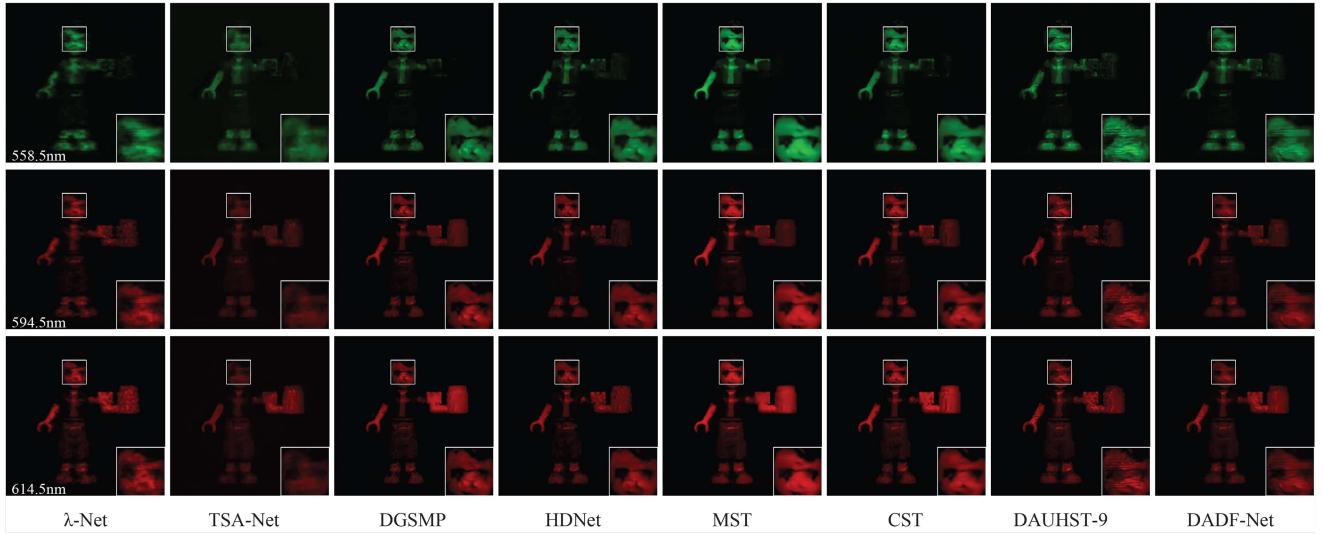


Fig. 14. Reconstructed real HSI comparisons of Scene 2 with 3 out of 28 spectral channels (wavelength: 558.5 nm, 594.5 nm and 614.5 nm), including the 6 SOTA methods and ours. Zoom in for a better view.

reconstruction results at each stage, resulting in a pleasant performance. When introducing a multi-stage design to enhance DADF-Net, the three-stage DADF-Net (DADF-Plus-3) performance can be comparable to that of DAUHST-9. However, basic DADF-Net achieves a pleasant trade-off between performance and computational cost.

**Real HSI Reconstruction:** Following the TSA-Net [31], we train our proposed model on CAVE [55] and KAIST [45] with 11-bit shot noise injected. Measurements obtained with a real SCI system are used for testing. Figs. 13 and 14 visualize the reconstructed result of real HSIs data of *Scene 1* and *Scene 2* with 3 out of 28 bands (558.5 nm, 594.5 nm, and 614.5 nm) using the 7 SOTA methods and our proposed method. The HSIs reconstructed by our method can provide better details and pleasant visualization results. This demonstrates the effectiveness of our approach in real imaging systems.

## VI. CONCLUSION

In this article, we proposed a degradation-aware dynamic Fourier-based network, named DADF-Net, for high-fidelity HSI reconstruction. Inspired by the degradation pattern inherent in imaging systems, we incorporated the potential relationship between the network input and the ground-truth HSI into the network architecture. We introduced the locally enhanced Fourier block to model both non-local and local features of the HSI effectively. Furthermore, we employed a multi-scale design to improve the overall reconstruction performance. Thanks to these techniques, our method achieved state-of-the-art (SOTA) results on both simulation and real datasets.

## REFERENCES

- [1] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2007.
- [2] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [3] Y. Yuan, X. Zheng, and X. Lu, “Hyperspectral image superresolution by transfer learning,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.
- [4] M. H. Kim et al., “3D imaging spectroscopy for measuring hyperspectral patterns on solid objects,” *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–11, 2012.
- [5] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, “Face recognition in hyperspectral images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1552–1560, Dec. 2003.
- [6] A. Bjorgan and L. L. Randeberg, “Towards real-time medical diagnostics using hyperspectral imaging technology,” in *Proc. Eur. Conf. Biomed. Opt.*, 2015, Art. no. 953712.
- [7] Z. Meng et al., “Snapshot multispectral endomicroscopy,” *Opt. Lett.*, vol. 45, no. 14, pp. 3897–3900, 2020.
- [8] X. Yuan, D. J. Brady, and A. K. Katsaggelos, “Snapshot compressive imaging: Theory, algorithms, and applications,” *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 65–88, Mar. 2021.
- [9] A. Wagadarikar, R. John, R. Willett, and D. Brady, “Single disperser design for coded aperture snapshot spectral imaging,” *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, 2008.
- [10] H. Du, X. Tong, X. Cao, and S. Lin, “A prism-based system for multispectral video acquisition,” in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 175–182.
- [11] P. Llull et al., “Coded aperture compressive temporal imaging,” *Opt. Exp.*, vol. 21, no. 9, pp. 10526–10545, 2013.
- [12] A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, “Video rate spectral imaging using a coded aperture snapshot spectral imager,” *Opt. Exp.*, vol. 17, no. 8, pp. 6368–6388, 2009.
- [13] X. Cao et al., “Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world,” *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 95–108, Sep. 2016.
- [14] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, “Multiframe image estimation for coded aperture snapshot spectral imagers,” *Appl. Opt.*, vol. 49, no. 36, pp. 6824–6833, 2010.
- [15] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, “Dual-camera design for coded aperture snapshot spectral imaging,” *Appl. Opt.*, vol. 54, no. 4, pp. 848–858, 2015.
- [16] X. Yuan, “Generalized alternating projection based total variation minimization for compressive sensing,” in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 2539–2543.
- [17] S. Zhang, L. Wang, Y. Fu, X. Zhong, and H. Huang, “Computational hyperspectral imaging based on dimension-discriminative low-rank tensor recovery,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 10183–10192.

- [18] Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai, "Rank minimization for snapshot compressive imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 2990–3006, Dec. 2019.
- [19] L. Wang, Z. Xiong, G. Shi, F. Wu, and W. Zeng, "Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 10, pp. 2104–2111, Oct. 2017.
- [20] S. W. Zamir et al., "Learning enriched features for real image restoration and enhancement," in *Proc. 16th Eur. Conf. Comput. Vis.*, 2020, pp. 492–511.
- [21] S. W. Zamir et al., "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14821–14831.
- [22] Z. Luo, Y. Huang, S. Li, L. Wang, and T. Tan, "Learning the degradation distribution for blind image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 6063–6072.
- [23] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ersgan: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1905–1914.
- [24] L. Wang et al., "Unsupervised degradation representation learning for blind super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 10581–10590.
- [25] Y. Yan et al., "Fine-grained attention and feature-sharing generative adversarial networks for single image super-resolution," *IEEE Trans. Multimedia*, vol. 24, pp. 1473–1487, 2022.
- [26] Y. Liu et al., "Iterative network for image super-resolution," *IEEE Trans. Multimedia*, vol. 24, pp. 2259–2272, 2022.
- [27] K. Park, J. W. Soh, and N. I. Cho, "A dynamic residual self-attention network for lightweight single image super-resolution," *IEEE Trans. Multimedia*, vol. 25, pp. 907–918, 2023.
- [28] G. Liu, H. Yue, J. Wu, and J. Yang, "Intra-inter view interaction network for light field image super-resolution," *IEEE Trans. Multimedia*, vol. 25, pp. 256–266, 2023.
- [29] T. Huang, W. Dong, X. Yuan, J. Wu, and G. Shi, "Deep Gaussian scale mixture prior for spectral compressive imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 16216–16225.
- [30] L. Wang, C. Sun, M. Zhang, Y. Fu, and H. Huang, "DNU: Deep non-local unrolling for computational spectral imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1661–1671.
- [31] Z. Meng, J. Ma, and X. Yuan, "End-to-end low cost compressive spectral imaging with spatial-spectral self-attention," in *Proc. 16th Eur. Conf. Comput. Vis.*, 2020, pp. 187–204.
- [32] X. Hu et al., "Hdnet: High-resolution dual-domain learning for spectral compressive imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 17542–17551.
- [33] X. Zhang, Y. Zhang, R. Xiong, Q. Sun, and J. Zhang, "HerosNet: Hyper-spectral explicable reconstruction and optimal sampling deep network for snapshot compressive imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 17532–17541.
- [34] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8032–8041.
- [35] Y. Cai et al., "Mask-guided spectral-wise transformer for efficient hyper-spectral image reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 17502–17511.
- [36] Z. Meng, Z. Yu, K. Xu, and X. Yuan, "Self-supervised neural networks for spectral snapshot compressive imaging," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2622–2631.
- [37] X. Miao, X. Yuan, Y. Pu, and V. Athitsos, "l-net: Reconstruct hyperspectral images from a snapshot measurement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 4059–4069.
- [38] Y. Cai et al., "Coarse-to-fine sparse transformer for hyperspectral image reconstruction," in *Proc. 17th Eur. Conf. Comput. Vis.*, 2022, pp. 686–704.
- [39] X. Yuan, Y. Liu, J. Suo, and Q. Dai, "Plug-and-play algorithms for large-scale snapshot compressive imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1447–1457.
- [40] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.
- [41] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 6000–6010.
- [42] X. Ding, X. Zhang, J. Han, and G. Ding, "Scaling up your kernels to 31x31: Revisiting large kernel design in CNNs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 11963–11975.
- [43] L. Chi, B. Jiang, and Y. Mu, "Fast fourier convolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 4479–4488.
- [44] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.
- [45] I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, and M. H. Kim, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–13, 2017.
- [46] Z. Meng, S. Jalali, and X. Yuan, "Gap-net for snapshot compressive imaging," 2020, *arXiv:2012.08364*.
- [47] L. Wang, Z. Wu, Y. Zhong, and X. Yuan, "Snapshot spectral compressive imaging reconstruction using convolution and contextual transformer," *Photon. Res.*, vol. 10, no. 8, pp. 1848–1858, 2022.
- [48] X. Ji et al., "Real-world super-resolution via kernel estimation and noise injection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 466–467.
- [49] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, 2015, vol. 28, pp. 2449–2457.
- [50] Y. Rao, W. Zhao, Z. Zhu, J. Lu, and J. Zhou, "Global filter networks for image classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 980–993.
- [51] S. W. Zamir et al., "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5728–5739.
- [52] J. Yu et al., "Free-form image inpainting with gated convolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 4471–4480.
- [53] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 606–615.
- [54] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 4641–4650.
- [55] J.-I. Park, M.-H. Lee, M. D. Grossberg, and S. K. Nayar, "Multispectral imaging using multiplexed illumination," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
- [56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [57] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [58] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.
- [60] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," in *Proc. Int. Conf. Learn. Representations*, 2016, pp. 1–15.
- [61] J. Ma, X.-Y. Liu, Z. Shou, and X. Yuan, "Deep tensor admm-net for snapshot compressive imaging," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 10223–10232.
- [62] Y. Cai et al., "Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 37749–37761.

**Ping Xu** received the Ph.D. degree in control theory and control engineering from Zhejiang university, Hangzhou, China, in 2006. He is currently an Associate Professor of Hangzhou Dianzi University, Hangzhou, China. From 2018 to 2019, he was a Visiting Scholar with Oklahoma State University, Stillwater, OK, USA. His research interests include computational imaging, image processing, and deep learning.

**Lei Liu** is currently working toward the M.S. degree with the School of Automation, Hangzhou Dianzi University, Hangzhou, China. His research interests include deep learning, computational imaging, and image restoration.

**Haifeng Zheng** (Senior Member, IEEE) received the B.Eng. and M.S. degrees in communication engineering from Fuzhou University, Fuzhou, China, in 2000 and 2003, respectively, and the Ph.D. degree in communication and information system from Shanghai Jiao Tong University, Shanghai, China, in 2014. He is currently a Professor with the College of Physics and Information Engineering, Fuzhou University. From 2015 to 2016, he was a Visiting Scholar with the State University of New York, Buffalo, NY, USA. His research interests include tensor theory, machine learning and its applications in image processing.

**Xin Yuan** (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from Xidian University, Xi'an, China, in 2007 and 2009, respectively, and the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2012. From 2012 to 2015, he was a Postdoctoral Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA, and a video analysis and coding Lead Researcher with Bell Labs, Murray Hill, NJ, USA, from 2015 to 2021. In 2021, he joined Westlake University, Hangzhou, China. His research interests include computational imaging, signal processing, and machine learning.

**Chen Xu** received the B.S. degree in optoelectronics and the Ph.D. degree in optical engineering from the Beijing Institute of Technology, Beijing, China, in 2012 and 2018, respectively. He is currently with Hangzhou Dianzi University, Hangzhou, China. His research interests include optical system design, freeform optics, and imaging systems.

**Lingyun Xue** received the B.Eng. degree in electronic precision machinery from Hangzhou Dianzi University, Hangzhou, China, in 1989, the M.Eng. degree in computer science and technology and the Ph.D. degree in biomedical engineering from Zhejiang University, Zhejiang, China, in 1999 and 2008, respectively. She is currently a Professor with Hangzhou Dianzi University. Her research interests include pattern recognition and intelligent system.