Depthwise Separable Convolution-Based Lightweight Network for Hyperspectral Image Denoising

Hui Luan, Yaoxin Tan, Jun Zhang, Mengling He, Zhaoming Wu

***Abstract*—In hyperspectral image denoising, the integration of a hypergraph convolutional neural network and a 3D convolutional neural network has demonstrated superior denoising performance. However, this hybrid approach suffers from high computational costs, which restricts its practical deployment in resource-constrained scenarios. To address this challenge, we propose a lightweight version of the hyperspectral image denoising network based on a hypergraph convolutional network (HGCDN) named Light-HGCDN, which maintains competitive denoising performance while significantly reducing computational complexity. The core innovation of Light-HGCDN lies in replacing the computationally expensive 3D module with an efficient multi-scale local feature extraction module based on depthwise separable convolution. Furthermore, to mitigate potential performance degradation resulting from model simplification, we introduce a multi-stage teacher-student framework, where the pre-trained HGCDN serves as the teacher model to guide Light-HGCDN through hierarchical knowledge distillation. Experimental results demonstrate that Light-HGCDN achieves a balance between denoising performance and computational efficiency, making it suitable for real-world applications. Additionally, we develop a web-based hyperspectral image denoising system, integrating Light-HGCDN into a Flask backend with an interactive frontend, further illustrating its practical applicability.**

***Index Terms*—Hyperspectral image denoising, Lightweight network, Depthwise separable convolution, Knowledge distillation, Teacher-student framework.**

I. Introduction

HYPERSPECTRAL imaging has emerged as a pivotal technology in various fields, including remote sensing, environmental monitoring, agriculture, and defense, owing to its ability to capture rich spectral-spatial information. Unlike conventional RGB images, hyperspectral images (HSIs) enable fine-grained material discrimination due to their contiguous spectral bands, making them indispensable for applications such as classification [1]–[3], and target recognition [4]–[6]. However, inherent limitations in imaging sensors and environmental interference often introduce complex noise patterns (e.g., Gaussian, stripe, and impulse noise) into acquired HSIs. Consequently, HSI denoising has become a critical preprocessing step to ensure reliable interpretation.

Traditional HSI denoising methods typically formulate the task as an optimization problem constrained by handcrafted priors that exploit intrinsic data properties. Widely adopted priors include total variation (TV) [7]–[9] for edge preservation, low-rank assumptions [10]–[12] for spectral correlation modeling, and sparse representations [13]–[15] for noise separation. Although these model-based approaches have achieved notable success, their performance is heavily dependent on manually designed priors that align with specific noise characteristics, requiring substantial domain expertise. Furthermore, the iterative optimization procedures employed in such methods incur high computational costs, and their specialization to particular noise types limits their adaptability to real-world scenarios with mixed or unknown noise distributions.

The advent of deep learning has revolutionized HSI denoising by enabling data-driven feature learning. Pioneering work by Chang et al. [16] introduced a multichannel 2D convolutional neural network (CNN) to leverage spatial correlations in HSIs. Subsequently, Yuan et al. [17] proposed HSID-CNN, which integrates multi-scale 2D-3D convolutional blocks to jointly capture spatial-spectral features. Despite these advances, early CNN-based methods struggled to model long-range dependencies, which are critical for suppressing structural noise. This limitation spurred interest in recurrent neural networks (RNNs). For instance, Wei et al. [18] developed QRNN3D, a quasi-recurrent architecture that enhances global spectral correlation modeling through 3D convolutions. Further innovations addressed spatial non-locality: Pang et al. [19] combined a Uformer block [20] with QRNN3D via bidirectional fusion, while Shi et al. [21] incorporated self-attention mechanisms to model inter-channel and inter-pixel relationships. In [22], Pan et al. incorporated the spectral attention calculation into QRNN to propose the spatial-spectral quasi-attention recurrent network (SQAD).

Recent advances in deep learning have demonstrated the superiority of transformer-based models in capturing long-range dependencies for HSI processing. Notably, Lai et al. [23] introduced a hybrid architecture integrating depthwise separable convolution with 3D self-attention to simultaneously model both local and global spectral-spatial relationships.

Manuscript received April 14, 2025; revised August 7, 2025.

This work was partly supported by the Jiangxi Provincial Natural Science Foundation (20232BAB201017, 20242BAB22013), the National Natural Science Foundation of China (62461043).

Hui Luan is a lecturer of College of Science, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, Jiangxi, China. (e-mail: nithui@163.com).

Yaoxin Tan is a postgraduate student of Jiangxi Province Key Laboratory of Smart Water Conservancy, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, Jiangxi, China. (e-mail: tyaoxin@163.com).

Jun Zhang is an associate professor of College of Science & Key Laboratory of Engineering Mathematics and Advanced Computing, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, Jiangxi, China. (corresponding author, e-mail: junzhang0805@126.com).

Mengling He is a postgraduate student of College of Science, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, Jiangxi, China (e-mail: heml0816@163.com).

Zhaoming Wu is an associate professor of Jiangxi Province Key Laboratory of Smart Water Conservancy, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, Jiangxi, China. (e-mail: zhaomingwu@cqu.edu.cn).

In a parallel development, Li et al. [24] developed a transformer-based framework employing non-local spatial self-attention and global spectral self-attention to effectively exploit HSI's inherent similarities. While 3D-CNNs demonstrate strong local feature extraction capabilities [25] and transformers excel at modeling long-range dependencies, researchers have increasingly focused on hybrid architectures that combine these complementary strengths. Dixit et al. [26] proposed an integrated framework merging both paradigms to achieve enhanced denoising performance. Li et al. [27] introduced a spectral-enhanced rectangle transformer (SERT), which employs rectangular self-attention for improved modeling of non-local similarities. A notable advancement in SERT is its incorporation of a memory unit that stores low-rank priors, effectively bridging model-based and learning-based denoising methodologies.

Recent advances in deep learning, particularly through the integration of three-dimensional convolutional neural networks (3D-CNNs) with Transformer architectures [28], have demonstrated remarkable success in HSI denoising while effectively preserving spectral fidelity. Nevertheless, current state-of-the-art models typically employ computationally intensive architectures that hinder their deployment in resource-constrained edge computing environments, such as unmanned aerial vehicles (UAVs) and embedded systems [29]. A representative example is the Hypergraph Convolutional Denoising Network (HGCDN) [30], which combines 3D convolution with hypergraph convolution [31] to achieve superior denoising performance through group correlation analysis [32]. Although HGCDN demonstrates exceptional denoising capabilities, its high computational complexity poses a critical challenge for real-world applications.

To reconcile the competing demands of denoising performance and computational efficiency, we propose Light-HGCDN, a computationally efficient variant of HGCDN optimized for practical deployment. The complexity analysis in [30] identifies the 3D Local Feature Module (3D-LFM) as the primary computational bottleneck in the original HGCDN architecture. To address this, we redesign the 3D-LFM using depthwise separable convolution, a technique well-established for its ability to significantly reduce both parameter count and computational costs (measured in GFLOPs) while preserving multi-scale feature extraction capabilities. However, such architectural simplification typically compromises the model's representational capacity. To mitigate this, we introduce a Multi-stage Teacher-Student Framework (MTSF), leveraging hierarchical knowledge distillation to transfer high-order features from the pre-trained HGCDN (the teacher model) to Light-HGCDN (the student model). This strategy ensures that the lightweight model inherits the teacher's robust feature learning capacity while maintaining computational efficiency, effectively addressing the performance degradation inherent in model compression.

Experimental results on benchmark datasets demonstrate that Light-HGCDN achieves an optimal balance between denoising performance and computational efficiency. For example, the proposed model incurs only a marginal 0.19 dB PSNR reduction compared to HGCDN, while reducing computational costs by 44.4% in terms of GFLOPs. These results highlight Light-HGCDN's suitability for resource-constrained environments while maintaining the stringent requirements of hyperspectral data analysis. To further validate Light-HGCDN's practical utility, we develop a web-based hyperspectral image denoising system. The system architecture consists of an interactive frontend built with standard web technologies (HTML, CSS, and JavaScript) and a lightweight Flask backend hosting Light-HGCDN for server-side inference. This integrated framework demonstrates the model's real-time processing capabilities and provides an accessible interface for real-world applications, bridging the gap between theoretical research and practical implementation. In summary, the principal contributions of this work are threefold.

- Lightweight Architecture: We propose a novel multi-scale feature extraction module based on depthwise separable convolution, reducing HGCDN's computational cost by 44.4% in terms of GFLOPs while maintaining competitive denoising performance (in mixed noise case).
- Hierarchical Knowledge Distillation: We propose a MTSF that enables progressive knowledge transfer from the teacher model (HGCDN) to the student model (Light-HGCDN) model, thereby preserving crucial spectral-spatial features in the lightweight architecture.
- Practical Deployment: We demonstrate Light-HGCDN's edge computing readiness via a web-based implementation, confirming its efficacy in real-world scenarios.

The remainder of this paper is organized as follows. Section II presents the proposed method in detail. Section III encompasses some experiments aimed at showcasing the superiority of our method through quantitative and visual result comparisons. Section IV analyzes and discusses the complexity of the proposed approach. Finally, we conclude the paper in Section V.

II. The Proposed Approach

HGCDN adopts a U-shape architecture inspired by U-Net [33], comprising two 3D convolutional feature extraction layers, seven blocks for extracting local and high-order features (ELHF), and a reconstruction layer. Each ELHF block consists of three key components: Local Feature Extraction Module (LFM), Dimensional Transformation Module (DTM), and High-Order Feature Extraction Module (HOFM). Through comprehensive complexity analysis, the computational bottleneck of ELHF stems primarily from its 3D-LFM. To address this, we develop a lightweight version of 3D-LFM (Light-LFM) using depthwise separable convolution, resulting in the proposed Light-HGCDN architecture, as shown in Fig. 1. However, such architectural simplification may compromise the network's feature-learning capacity. To ensure optimal knowledge transfer from HGCDN to Light-HGCDN, we design a multi-stage teacher-student framework (MTSF) for the training phase. In addition, to avoid gradient explosion during the training process, we implement a hybrid distillation loss function that combines smooth loss functions L_1 and L_2. In summary, depthwise separable convolution improves the computational efficiency of the model, and knowledge distillation technology retains the higher-order feature learning ability of the teacher model, thus achieving a balance between noise reduction performance and efficiency.

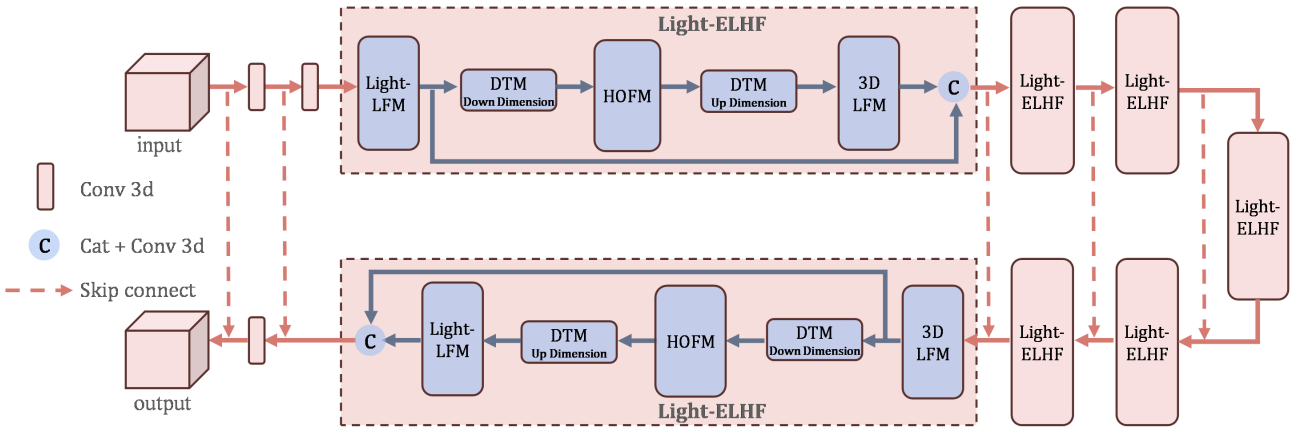


Fig. 1. Overall architecture of HGDCN.

A. *Lightweight local feature extraction module*

In HGCDN, the 3D-LFM was designed to extract local features from HSI, inspired by DenseNet [34] (Fig. 2). Although experimental results confirmed its superior feature extraction capabilities, its substantial computational complexity limited practical deployment in real-world applications. To address this issue, we propose an optimization scheme using depthwise separable convolution, which decomposes the 3D convolution into sequential 1D and 2D convolutional operations. We further incorporate a multi-scale convolutional mechanism to enhance feature representation. The optimized module structure is shown in Fig. 3.

Light-LFM employs efficient $1 \times 1 \times 1$ 3D convolution kernels at both ends for feature map transitions, while two intermediate convolutional layers handle core feature extraction. Although this dimensional reduction may weaken feature extraction capacity, our multi-scale strategy compensates by expanding the receptive field.

Through systematic evaluation, we identify that combining 3×3 and 5×5 convolution kernels achieves the best performance-efficiency trade-off, whereas 7×7 kernels incur impractical overhead. The optimized Light-LFM achieves remarkable parameter reduction. For single-channel operations, the parameter count decreases from 108 ($3 \times 3 \times 3 \times 4$) in 3D-LFM to 44 ($1 + (3 \times 3 + 3) + (5 \times 5 + 5) + 1$), a 60% reduction. This design preserves strong representation power while dramatically improving computational efficiency, enhancing practical utility without compromising performance.

B. *Multi-stage teacher-student framework based on feature transfer*

Given that network simplification significantly increases training difficulty, we employ the knowledge distillation technique to ensure that Light-HGCDN effectively preserves complex features. The traditional knowledge distillation technique [35] utilizes temperature parameters and soft labels to formulate distillation loss functions, facilitating knowledge transfer from complex teacher models to compact student models via a teacher-student framework. However, this approach is primarily limited to Softmax-based classification models, restricting its broader applicability. Recent advances have diversified distillation methodologies through various strategies, including feature transfer [36], and relationship transfer [37], though most applications focus on image classification. Recently, knowledge distillation techniques have been applied to image restoration. For example, Hong et al. [38] proposed an image dehazing method using knowledge distillation, while Li et al. designed a heterogeneous knowledge distillation strategy for image denoising [39].

Inspired by prior research, we propose a multi-stage high-order feature knowledge distillation strategy, allowing Light-HGCDN to acquire more comprehensive and precise high-order representations. As illustrated in Fig. 4, our approach leverages the pre-trained HGCDN as the teacher model and its robust high-order feature learning capabilities to progressively transfer knowledge to the student model (Light-HGCDN) through hierarchical feature distillation. The distillation adopts a three-stage progressive strategy:

- **Stage 1**: The knowledge distillation is performed using features from layers 3-5 in the teacher model's EHLF modules.
- **Stage 2**: The distillation scope expands to layers 3-7, promoting deeper knowledge transfer, which helps the student model learn complex high-order features.
- **Stage 3**: we optimize feature representation by distilling the input and output HOFM layers (i.e., 1, 2, 8, and 9).

This multi-stage progressive approach effectively mitigates training challenges from model compression, gradually increasing task complexity to overcome training bottlenecks while enriching feature representations.

For the loss function design, we propose a composite loss function to optimize model training. Specifically, Knowledge Distillation Loss (KDL), defined by the Smooth L_1 norm, prevents gradient explosion during the distillation process. Mathematically, KDL is expressed as

$$Loss_{KDL} = \sum_{i=1}^{n} Smooth_{L_1}(Teacher_i - Student_i), \quad (1)$$

where

$$Smooth_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1, \\ |x| - 0.5, & |x| \geq 1. \end{cases} \quad (2)$$

Here, $Teacher_i$ and $Student_i$ represent the output feature maps from the i-th layer of the teacher and student models, respectively.

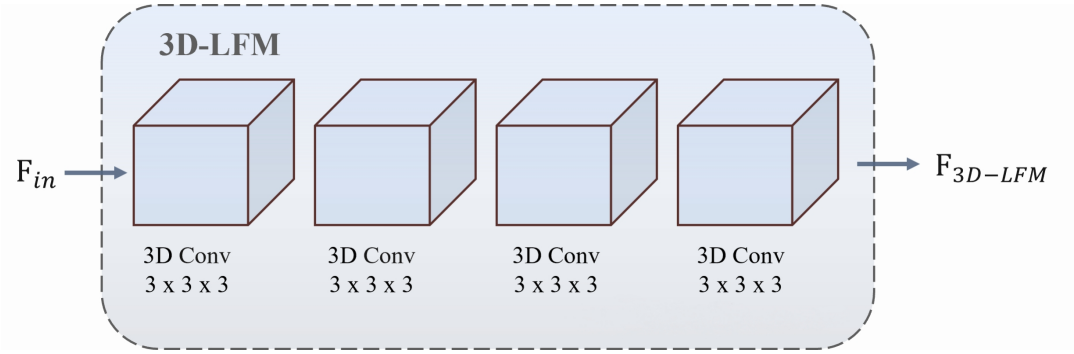


Fig. 2. 3D-LFM.

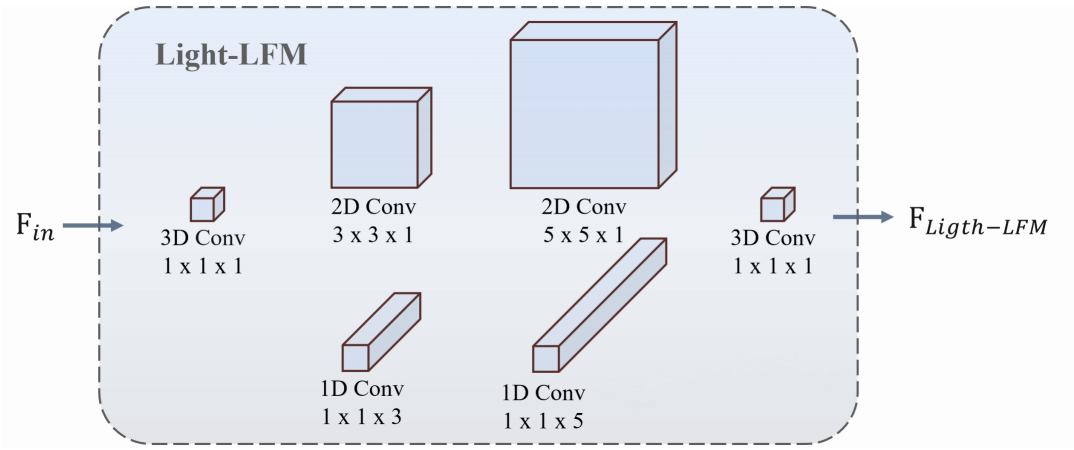


Fig. 3. Light-LFM.

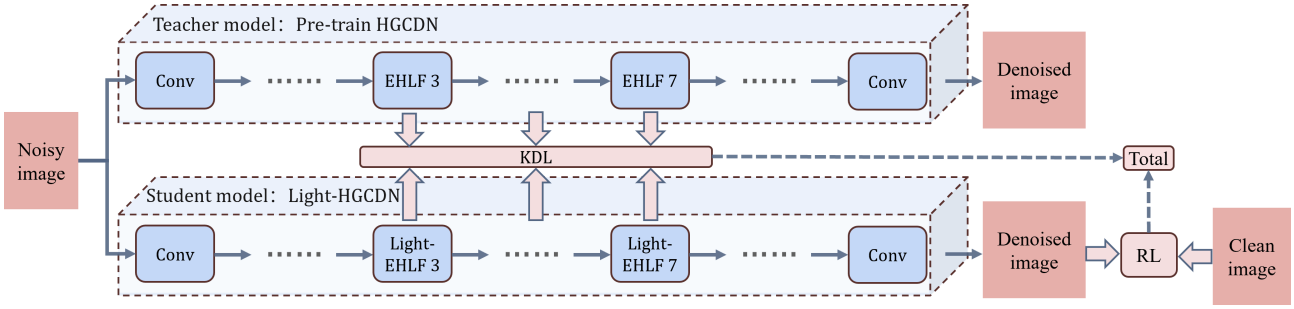


Fig. 4. MTSF.

To assess reconstruction quality, we employ an L_2-norm-based Reconstruction Loss (RL) to measure the discrepancy between the student model's denoised output and the ground truth image. The RL is defined as follows:

$$Loss_{RL} = \frac{1}{2}\|\mathrm{X} - Student_{output}\|_2^2, \tag{3}$$

where $Student_{output}$ indicates the denoised image from the student model, and X stands for the ground truth.

The total loss function is formulated as

$$Loss_{total} = Loss_{RL} + \omega Loss_{KDL}. \tag{4}$$

Here, ω serves as a balancing factor to harmonize the two loss items. During the first two training stages, we set $\omega = 0.2$ and reduce it to 0.1 in the final stage. This composite loss function enables Light-HGCDN to simultaneously maintain fundamental image reconstruction quality while effectively acquiring high-order feature representations. By balancing these objectives, we achieve significantly improved denoising performance.

In summary, Light-LFM employs the depthwise separable convolution to decompose 3D convolution into sequential 1D and 2D operations, achieving a 60% reduction in parameters compared to the original 3D-LFM while substantially decreasing the computational complexity. This is combined with a multi-stage progressive training strategy that gradually increases task complexity, addressing the training challenges inherent to lightweight models and enabling Light-HGCDN to acquire more comprehensive high-order feature representations. The training uses a composite loss function with RL and KDL terms. The RL term maintains output fidelity by minimizing pixel-wise errors between denoised and clean images, while the KDL term facilitates the transfer of high-order features from the teacher model, enabling the student model to retain both high-order features and local details. Through these coordinated designs, Light-HGCDN balances efficiency and performance, maintaining competitive denoising quality with significant computational savings.

III. Numerical Experiments

In this section, we present experiments designed to validate the proposed method, including simulated HSI denoising tests and ablation studies.

A. Experimental settings

1) Datasets: For synthetic HSI denoising evaluation, we use the ICVL dataset, which serves as a benchmark standard in hyperspectral imaging research. This dataset contains 201 high-quality HSIs, each with dimensions of 1392×1300 pixels across 31 spectral bands covering the 400-700 nm wavelength range. Following [18], we divide the dataset into three subsets: 100 images for model training, with the remainder allocated for validation and testing. To optimize neural network training, we implement comprehensive preprocessing procedures. For synthetic data generation, each training image is cropped into 1024×1024-pixel spatial patches. We further enhance data diversity through random rotational augmentation. During evaluation, we extract the central 512×512 region from all test images to ensure standardized assessment.

2) Comparison methods: To evaluate Gaussian noise removal performance, we compare our method with five state-of-the-art denoising approaches: LLRT [40], NGmeet [41], QRNN3D [18], Trq3d [19], and MAN [42]. In the complex noise case, the comparison methods are LRTDTV [8], LRTV [9], QRNN3D, Trq3d, and SERT [27].

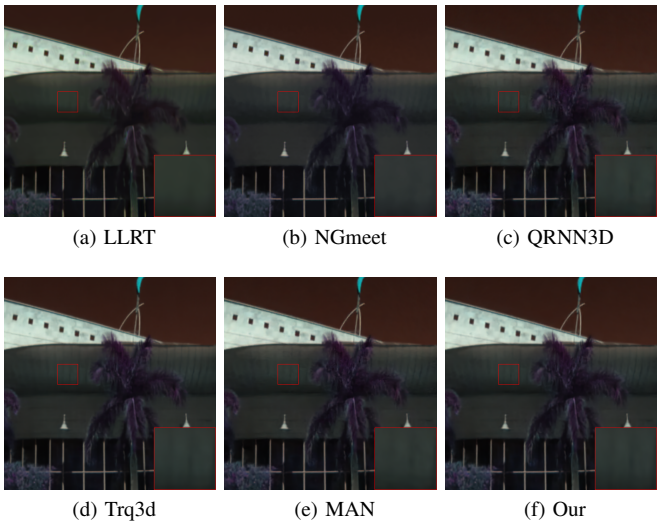
3) Evaluation metrics: Four standard metrics are employed to assess model performance in our experiments, including peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [43], spectral angle mapper (SAM) [44], and erreur relative globale adimensionnelle de synthse (ERGAS) [45].

4) Experimental scenarios: To comprehensively evaluate the denoising capability of Light-HGCDN, we conduct two synthetic noise experiments: (1) Gaussian noise ($\sigma = 50$) and (2) complex noise (comprising non-i.i.d. Gaussian, stripe, deadline, and impulse noise). Additionally, to validate the effectiveness of MTSF, we perform ablation experiments. By comparing model performance across training stages, we quantitatively analyze the contribution of the progressive training strategy.

5) Implementation details: All experiments are implemented using the PyTorch framework, with neural network model training and testing carried out on two NVIDIA GeForce RTX 3080 GPUs. For practical deployment, Light-HGCDN runs on a computer equipped with an Intel Core i5 processor and 32GB RAM. The web-based application utilizes a Flask backend and features an interactive frontend developed with HTML, CSS, and JavaScript.

B. Synthetic experiments

To demonstrate the superior denoising capability of Light-HGCDN, we conduct experiments under two noise scenarios: Gaussian noise scenario and complex noise scenario. Quantitative results are presented in Tables I and II. And visual comparisons are shown in Figs. 5 and 6, which consist of bands 15, 27 and 31. In the Gaussian noise scenario, Light-HGCDN without knowledge distillation achieves best PSNR and SSIM, as well as second-best SAM. Although NGmeet shows superiority on SAM and ERGAS, its results exhibit noticeable oversmoothing artifacts. These findings



(a) LLRT (b) NGmeet (c) QRNN3D

(d) Trq3d (e) MAN (f) Our

Fig. 5. Visual comparison of the Gaussian noise case.

demonstrate that our lightweight architecture alone retains excellent denoising capability. In the complex noise scenario, the Light-HGCDN optimized by knowledge distillation shows significant denoising advantages, outperforming all comparison methods in all evaluation metrics. This confirms Light-HGCDN's practical utility for real-world applications.

TABLE I Quantitative comparison of Light-HGCDN (not distilled) and competing methods in Gaussian noise scenario ($\sigma = 50$).

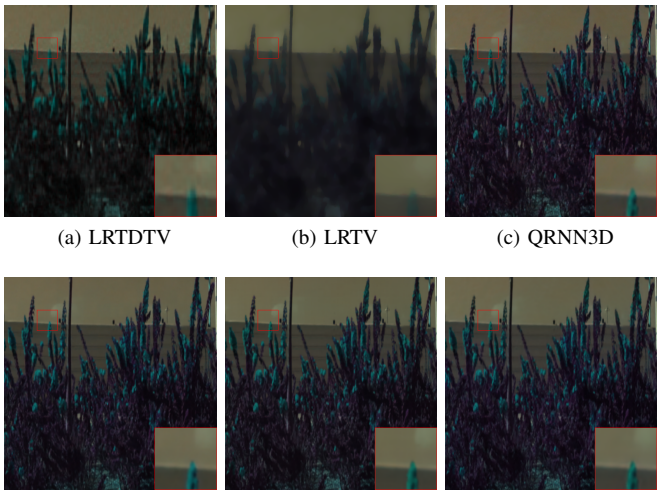
Metric	LLRT	NGmeet	QRNN3D	Trq3d	MAN	Light-HGCDN
PSNR	38.83	40.35	40.30	40.21	40.51	**40.67**
SSIM	0.942	0.954	0.955	0.957	0.956	**0.957**
SAM	0.0735	**0.0587**	0.0711	0.0636	0.0607	0.0607
ERGAS	70.61	**45.25**	49.90	56.46	46.21	55.13

TABLE II Quantitative comparison of Light-HGCDN (distilled) and competing methods in complex noise case.

Metric	LRTDTV	LRTV	QRNN3D	Trq3d	SERT	Light-HGCDN
PSNR	34.46	31.31	39.33	40.16	39.34	**40.34**
SSIM	0.906	0.862	0.945	0.958	0.950	**0.962**
SAM	0.1058	0.2213	0.0841	0.0654	0.0649	**0.0599**
ERGAS	109.13	178.40	50.77	43.73	48.59	**42.79**

C. Ablation studies

To verify the effectiveness of MTSF, we conduct comprehensive testing of each training stage under mixed noise conditions, with results presented in Table III. Compared with the model without knowledge distillation (None-KD), single-stage distillation yields significant improvements across all metrics (PSNR, SSIM, and ERGAS), confirming KD's ability to enhance Light-HGCDN's capacity for complex noise modeling. Compared with single-stage KD, MTSF obtains better denoising performance in the second stage and significantly improves the model performance after three stages of KD. During the experiments, empirical observations reveal performance degradation beyond epoch 230, indicating overfitting. Therefore, we implement early stopping, selecting



(a) LRTDTV (b) LRTV (c) QRNN3D

(d) Trq3d (e) SERT (f) Our

Fig. 6. Visual comparison of the complex noise case.

the epoch 230 model as final. These findings collectively demonstrate that our multi-stage high-order feature transfer strategy enables richer feature representation learning while mitigating training instability.

TABLE III Ablation experiments about MTSF in mixed noise scenario.

Metric	HGCDN (baseline)	None-KD	Single Stage	MTSF-Stage 1	MTSF-Stage 2	MTSF-Stage 3
PSNR	40.53	38.74	39.97	39.76	40.15	40.34
(Δ)↓	-	(1.79)	(0.56)	(0.77)	(0.38)	**(0.19)**
SSIM	0.964	0.957	0.958	0.957	0.960	0.962
(Δ)↓	-	(0.007)	(0.006)	(0.007)	(0.004)	**(0.002)**
SAM	0.0562	0.0574	0.0631	0.0663	0.0646	0.0599
(Δ)↓	-	**(0.0012)**	(0.0069)	(0.0101)	(0.0084)	(0.0037)
ERGAS	41.20	57.71	44.26	45.87	44.47	42.79
(Δ)↓	-	(16.51)	(3.06)	(4.67)	(3.27)	**(1.59)**

D. Application example

We present a web-based implementation of Light-HGCDN to validate its real-world applicability. The backend utilizes Python's lightweight Flask framework to integrate our pre-trained model via RESTful APIs for efficient denoising. The frontend, implemented with standard web technologies (HTML, CSS, and JavaScript), employs a hierarchical design that optimizes usability while ensuring system maintainability and scalability.

The system architecture consists of three tiers. The first tier contains the authentication interface for login verification. The second tier hosts the home page with navigation controls, serving as the primary gateway to system functionalities. The third tier provides specialized functional modules accessible via the navigation menu. As shown in Fig. 7, the core image processing module enables HSI uploading, backend denoising algorithm interaction, and intuitive result visualization. This implementation demonstrates Light-HGCDN's deployment feasibility in production environments while proving its practical value for HSI denoising applications.

IV. Analysis and Discussion

Table IV presents the computational requirements of all deep learning-based denoising methods in our synthetic experiments, evaluating both parameter count and computational complexity. Compared with the original HGCDN, Light-HGCDN reduces the number of parameters by 0.53 M and computational complexity by 25.4 GFLOPs, representing an approximately 50% reduction in computational requirements. Although the lightweight model exhibits a slight denoising performance decline, the substantial complexity reduction demonstrates the efficacy of our proposed Light-HGCDN and knowledge distillation strategy. Moreover, Light-HGCDN maintains superior denoising performance relative to both Trq3d and QRNN3D while requiring lower computational complexity. This result further validates the importance of high-order correlation in HSI denoising, indicating Light-HGCDN's ability to efficiently leverage high-order features with minimal overhead.

TABLE IV Complexity comparison of Light-HGCDN and other deep learning methods.

Metric	QRNN3D	Trq3d	MAN	SERT	HGCDN	Light-HGCDN
Params (M)	0.86	0.68	0.50	1.91	2.59	2.06
GFLOPs	39.30	33.38	19.06	15.16	57.18	31.78
MACs	19.65	16.69	9.53	7.58	28.59	15.89

V. Conclusion

To address the high computational complexity of HGCDN, we proposed a lightweight multi-scale local feature extraction module via depthwise separable convolution, which achieved significant model compression. This design reduced parameters by 20.5% and computational costs by 44.4% (in FLOPs) compared to the original HGCDN. To preserve denoising performance, we designed a multi-stage teacher-student architecture leveraging knowledge distillation. This hierarchical training strategy enabled progressive knowledge transfer, which helped the lightweight model overcome optimization bottlenecks while maximizing the feature representation capacity of the lightweight 3D-LFM. Particularly, we utilized a hybrid distillation loss function to stabilize training by mitigating abrupt loss fluctuations. Experimental results in synthetic noise tests and ablation studies demonstrated that our approach maintains competitive denoising performance while substantially reducing computational costs. Additionally, we developed a web-based application to validate Light-HGCDN's practicality in real-world scenarios. The system features an interactive interface for HSI uploads, with backend denoising and visualization, confirming Light-HGCDN's operational feasibility for real-world HSI processing.

References

[1] P. Burai, B. Deák, O. Valkó, and T. Tomor, "Classification of Herbaceous Vegetation Using Airborne Hyperspectral Imagery," *Remote Sensing*, vol. 7, no. 2, pp. 2046-2066, 2015.

[2] S. G. Azar, S. Meshgini, T. Y. Rezaii, and S. Beheshti, "Hyperspectral Image Classification Based on Sparse Modeling of Spectral Blocks," *Neurocomputing*, vol. 407, pp. 12-23, 2020.

[3] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "SpectralFormer: Rethinking Hyperspectral Image Classification with Transformers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, p. 5518615, 2021.

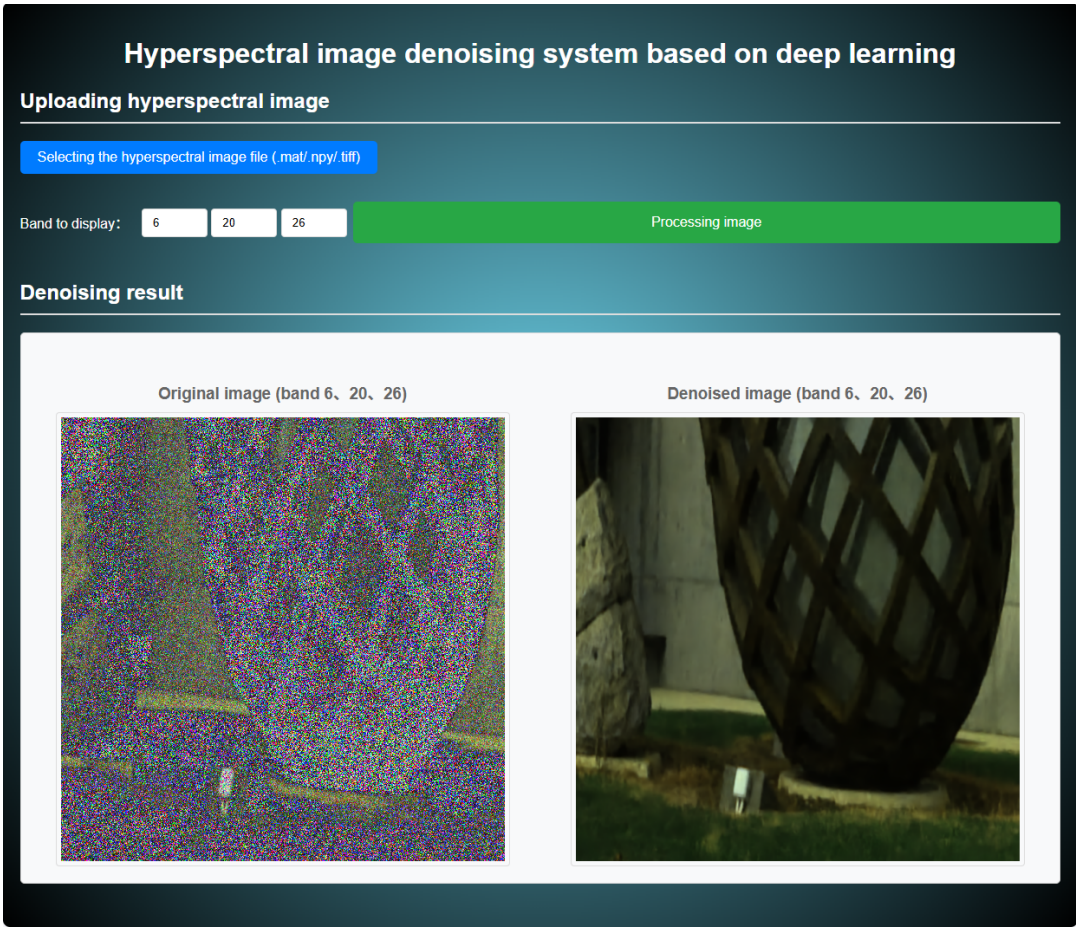


Fig. 7. System page.

[4] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral Face Recognition with Spatiospectral Information Fusion and PLS Regression," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 1127-1137, 2015.

[5] Y. Shi, L. Han, W. Huang, S. Chang, Y. Dong, D. Dancey, and L. Han, "A Biologically Interpretable Two-Stage Deep Neural Network (Bit-Dnn) for Vegetation Recognition from Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, p. 4401320, 2021.

[6] C. Chen and B. Wu. "An Improved YOLOv8 Algorithm for Detecting Remote Sensing Target Images," *IAENG International Journal of Applied Mathematics*, vol. 55, no. 5, pp. 1294-1303, 2025.

[7] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral Image Denoising Employing a Spectral-Spatial Adaptive Total Variation Model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3660-3677, 2012.

[8] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral Image Restoration Via Total Variation Regularized Low-Rank Tensor Decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1227-1243, 2017.

[9] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-Variation-Regularized Low-Rank Matrix Factorization for Hyperspectral Image Restoration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, p. 7167714, 2015.

[10] K. Wei and Y. Fu, "Low-Rank Bayesian Tensor Factorization for Hyperspectral Image Denoising," *Neurocomputing*, vol. 331, pp. 412-423, 2019.

[11] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral Image Restoration Using Low-Rank Matrix Recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729-4743, 2013.

[12] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal Low-Rank Regularized Tensor Decomposition for Hyperspectral Image Denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 5174-5189, 2019.

[13] Y. Qian and M. Ye, "Hyperspectral Imagery Restoration Using Nonlocal Spectral-Spatial Structured Sparse Representation With Noise Estimation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol, 6, no. 2, pp. 499-515, 2012.

[14] Y.-Q. Zhao and J. Yang, "Hyperspectral Image Denoising via Sparse Representation and Low-Rank Constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 296-308, 2014.

[15] T. Lu, S. Li, L. Fang, Y. Ma, and J. A. Benediktsson, "Spectral-Spatial Adaptive Sparse Representation for Hyperspectral Image Denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 373-385, 2015.

[16] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-Denet: Hyperspectral Image Restoration via Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 667-682, 2018.

[17] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral Image Denoising Employing a Spatial-Spectral Deep Residual Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1205-1218, 2019.

[18] K. Wei, Y. Fu, and H. Huang, "3-D Quasi-Recurrent Neural Network for Hyperspectral Image Denoising," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 363-375, 2020.

[19] L. Pang, W. Gu, and X. Cao, "TRQ3DNet: A 3D Quasi-Recurrent and Transformer Based Network for Hyperspectral Image Denoising," *Remote Sensing*, vol. 14, no. 18, p. 4598, 2022.

[20] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A General U-Shaped Transformer for Image Restoration," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17683-17693.

[21] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang, "Hyperspectral Image Denoising Using a 3-D Attention Denoising Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10348-10363, 2021.

[22] E. Pan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "SQAD: Spatial-Spectral Quasi-Attention Recurrent Network for Hyperspectral Image Denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, p. 5524814, 2022.

[23] Z. Lai, C. Yan, and Y. Fu, "Hybrid Spectral Denoising Transformer with Guided Attention," in *Proc. IEEE/CVF International Conference on Computer Vision*, 2023, pp. 13065-13075.

[24] M. Li, Y. Fu, and Y. Zhang, "Spatial-Spectral Transformer for Hyperspectral Image Denoising," in *Proc. AAAI Conference on Artificial Intelligence*, vol. 37, no. 1, 2023, pp.1368-1376.
[25] Q. Zhang, Y. Zheng, Q. Yuan, M. Song, H. Yu, and Y. Xiao, "Hyperspectral Image Denoising: From Model-Driven, Data-Driven, to Model-Data-Driven," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 10, pp. 13143-13163, 2024.
[26] A. Dixit, A. K. Gupta, P. Gupta, S. Srivastava, and A. Garg, "UNFOLD: 3-D U-Net, 3-D CNN, and 3-D Transformer-Based Hyperspectral Image Denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, p. 5529710, 2023.
[27] M. Li, J. Liu, Y. Fu, Y. Zhang, and D. Dou, "Spectral Enhanced Rectangle Transformer for Hyperspectral Image Denoising," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5805-5814.
[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All You Need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5999-6009.
[29] S. Wang, "A Face Recognition Method based on Lightweight Neural Network and Multi Hash Recognition Degree Weighting," *IAENG International Journal of Applied Mathematics*, vol. 54, no. 3, pp.581-586, 2024.
[30] J. Zhang, Y. Tan, and X. Wei, "Exploring High-Order Correlation For Hyperspectral Image Denoising with Hypergraph Convolutional Network," *Signal Processing*, vol. 227, p. 10971, 2025.
[31] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph Neural Networks," in *Proc. AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, 2019, pp. 3558-3565.
[32] Y. Chen, W. He, X. -L. Zhao, T. -Z. Huang, J. Zeng, and H. Lin, "Exploring Nonlocal Group Sparsity Under Transform Learning for Hyperspectral Image Denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, p. 5537518, 2022.
[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234-241.
[34] G. Huang, Z. Liu, L. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700-4708.
[35] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," arXiv preprint arXiv:1503.02531, 2015.
[36] S. Zagoruyko and N. Komodakis, "Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer," arXiv preprint arXiv:1612.03928, 2016.
[37] J. Yim, D. Joo, J. Bae, and J. Kim, "A Gift From Knowledge Distillation: Fast Optimization, Network Minimization and Transfer Learning," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4133-4141.
[38] M. Hong, Y. Xie, C. Li, and Y. Qu, "Distilling Image Dehazing with Heterogeneous Task Imitation," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. pp. 3462-3471.
[39] J. Li, H. Yang, Q. Yi, F. Fang, G. Gao, T. Zeng, and G. Zhang, "Multiple Degradation and Reconstruction Network for Single Image Denoising via Knowledge Distillation," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 558-567.
[40] Y. Chang, L. Yan, and S. Zhong, "Hyper-Laplacian Regularized Unidirectional Low-Rank Tensor Recovery for Multispectral Image Fenoising," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4260-4268.
[41] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-Local Meets Global: An Integrated Paradigm for Hyperspectral Denoising," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6868-6877.
[42] Z. Lai and Y. Fu, "Mixed Attention Network for Hyperspectral Image Denoising," arXiv preprint arXiv:2301.11525, 2023.
[43] 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.
[44] R. H. Yuhas, A. F. H Goetz, and J. W. Boardman, "Discrimination Among Semi-Arid Landscape Endmembers Using the Spectral Angle Mapper (SAM) Algorithm," in *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop*, 1992.
[45] L. Wald, "Quality of High Resolution Synthesised Images: Is There a Simple Criterion?," in *Proc. Third Conference "Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images"*, 2000, pp. 99-103.