Letter

D2Net: Deep Denoising Network in Frequency Domain for Hyperspectral Image

Erting Pan, Yong Ma, Xiaoguang Mei, Jun Huang, Fan Fan, and Jiayi Ma

Dear Editor,

Since the existing hyperspectral image denoising methods suffer from excessive or incomplete denoising, leading to information distortion and loss, this letter proposes a deep denoising network in the frequency domain, termed D2Net. Our motivation stems from the observation that images from different hyperspectral image (HSI) bands share the same structural and contextual features while the reflectance variations in the spectra are mainly fallen on the details and textures. We design the D2Net in three steps: 1) spatial decomposition, 2) spatial-spectral denoising, and 3) refined reconstruction. It achieves multi-scale feature learning without information loss by adopting the rigorous symmetric discrete wavelet transform (DWT) and inverse discrete wavelet transform (IDWT). In particular, the specific design for different frequency components ensures complete noise removal and preservation of fine details. Experiment results demonstrate that our D2Net can attain a promising denoising performance.

Introduction: Due to various unstable factors in the complex imaging chain, HSIs are always contaminated by noises, which will severely degrade the visual qualtiy and affect their further analysis and subsequent interpretations [1], [2]. Therefore, removing HSI noise is of utmost significance for HSI exploitation.

Traditional HSI denoising methods commonly employ a definite model such as low-rank matrix recovery [3], domain transform [4], sparse representation [5], [6], tensor decomposition [7], etc. Unfortunately, they suffer from manual parameterization for specific noise cases and befall complex optimization problems. With the benefit of solid nonlinear representative abilities and fewer priors, convolutional neural networks (CNN)-based techniques have achieved impressive success in image restoration [8]. However, every coin has two sides. Few priors imply a large number of parameters, resulting in huge redundant calculations, especially with high-dimensional HSIs. On the other hand, mainstream CNN-based denoising approaches typically follow either an encoder-decoder [9] or a high-resolution (single-scale) [10] feature processing architecture. The former achieves extensive contextual feature learning by upsampling and downsampling operations, but it loses fine spatial details, making it challenging to reconstruct clean HSIs accurately. The latter do not vary the spatial resolution and the limitation of the receptive field makes such networks incapable of encoding contextual information. Furthermore, due to a neglect of the intrinsic HSI spatial-spectral properties, these methods represent quite finite denoising performance. Therefore, investigating a tailor-made denoising approach is an urgent and challenging task, which requires efficiently removing HSI noises while carefully preserving the high-frequency details.

We have observed that capturing from the same scene, images in different bands show strong correlations. As evident in Fig. 1, they share the same structural and contextual features, which are also called low-frequency features. On the other hand, the reflectance difference between bands may induce by details, textures, and noises, which mainly belong to high-frequency features. Motivated by this, we argue that denoising HSIs in the frequency domain can achieve better results and propose D2Net (as shown in Fig. 2) to efficiently and precisely restore a clean HSI. First, we replace common up-sampling and down-sampling operations with DWT/IDWT, which is mathematically strict and symmetric, to provide rich frequency features for domain transformation and clean HSI reconstruction. Second, DWT/IDWT produces multi-scale features without any information loss, enabling us to design sub-branches for denoising in different frequency components. In particular, we propose a progressive spatial-spectral mixed convolution block (PMCB) to protect the effective transfer of high-frequency information. Third, we deploy a spatial-spectral consistency regularization block (SCRB) to explore its coherence further and finely reconstruct the clean HSI. The experimental results prove that the proposed method has a good trade-off between efficiency and denoising performance.

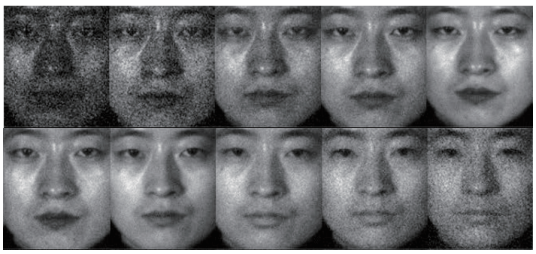

Fig. 1. Examples of an HSI [1] taken from the PolyU hyperspectral face database. Sample bands covering the visible range from 420 nm to 690 nm in 30 nm intervals.

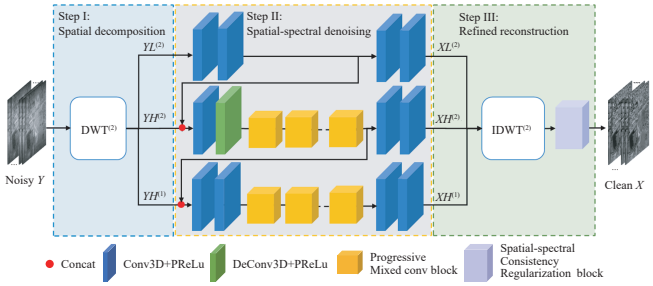

Fig. 2. The proposed D2Net framework for HSI denoising.

Methodology: The goal of the HSI denoising task is to recover a clean HSI X from a noise-contaminated HSI Y, where $X, Y \in \mathbb{R}^{(C\times H\times W)}$, and C presents the number of spectral bands, H and W describe their spatial scale.

1) Spatial decomposition: As aforementioned, mainstream CNN-based HSI denoising methods easily fail in recovering spatial and spectral details. On the contrary, denoising in the frequency domain utilizing strictly symmetric DWT/IDWT enables multi-scale feature learning without information loss, alleviating this problem. In this paper, we utilize DWT in the second level with the *Haar* wavelet kernel to spatially decompose the noisy HSI Y before denoising so that the network can perform customized and accurate denoising for different frequency components. It can be formulated as

$$YH^{(1)},\ YH^{(2)},\ YL^{(2)} = \mathbf{DWT}(\mathbf{DWT}(Y)) \tag{1}$$

where $YL^{(2)}$ is in size of $(B, 1, C, H/2, W/2)$, $YH^{(2)}$ and $YH^{(1)}$ are in size of $(B, 3, C, H/4, W/4)$, B is the batch size, 1 and 3 indicate the corresponding number of wavelet components. Intuitively, the employed DWT can derive wavelet subbands in multi-resolution, allowing multi-scale feature learning and benefiting the non-local similarities exploration in HSIs. Besides, DWT/IDWT would not affect the end-to-end training of our network, making the proposed

Corresponding author: Xiaoguang Mei.

Citation: E. T. Pan, Y. Ma, X. G. Mei, J. Huang, F. Fan, and J. Y. Ma, "D2Net: Deep denoising network in frequency domain for hyperspectral image," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 3, pp. 813–815, Mar. 2023.

The authors are with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: panerting@whu.edu.cn; mayong@whu.edu.cn; meixiaoguang@gmail.com; junhwong@whu.edu.cn; fanfan@whu.edu.cn; jyma2010@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2022.106019

D2net simple and effective.

2) Spatial-spectral denoising: In view of noise distribution in different frequency components and the intrinsic spatial and spectral characteristics of HSIs, we design three subbranches to fulfil spatial-spectral denoising. After DWT, the low-frequency components of HSIs retain more structural information and are less damaged by noises, while the high-frequency details and textures are unfortunately drowned in noises. Accordingly, we infer that it is easier to remove noise in low-frequency, for which we deploy a simple multi-layer network consisting of Conv3D+PReLU. Conversely, for the other two high-frequency subbranches in different resolutions, we design the PMCB to exploit inherent spatial-spectral properties and finely discriminate the detailed texture and noise of each band. Noting that the $YH^{(1)}$ and $YH^{(2)}$ have similar information patterns in high-frequency but are in different resolutions, we utilize a DeConv3D+PReLU block to spatially amplify the $YH^{(2)}$ in twice. It empowers recursively sharing subsequent parameters of spectral-spatial denoising in PMCB. Concretely, inspired by the superiority of the information multi-distillation block (IMDB) proposed in [11], we follow its idea of information distillation and design the PMCB as illustrated in Fig. 3 (a). Here, instead of the general convolution calculation in IMDB, we develop a mixed convolution designed to fulfil comprehensive and efficient spatial-spectral feature learning. The calculation in PMCB can be formulated as

$$\begin{aligned} f &= \mathbf{Concat}(\mathbf{ConvA}(g), \mathbf{ConvB}(g'), \mathbf{ConvC}(g''), \mathbf{ConvD}(g''')) \\ \hat{f} &= \mathbf{CA}(f) + g \end{aligned} \tag{2}$$

where g, g', g'' and g''' indicate the input features of each convolution as labeled in Fig. 3 (a). Concretely, ConvA with $3\times3\times3$ kernel and ConvD with $1\times1\times1$ kernel are intuitive 3D convolution. Pseudo-3D convolution ConvB in the kernel size of $3\times3\times1$ is employed for the spatial domain, and ConvC in the kernel size of $1\times1\times3$ is specific for the spectral domain. Such 3D convolutions allow exploration of inherent spatial-spectral correlations with different kernel size and the pseudo-3D convolutions purely concentrate on extracting features in the specific domain. It also significantly decreases the number of parameters and construct a more lightweight progressive feature learning architecture. After that, a feature channel attention (CA) block is adding to explore the cross-channel features further and extract global information. Besides, we stack multiple PMCBs to yield the greatest returns and boost the denoising performance.

3) Refined reconstruction: After obtaining the denoising components $XH^{(1)}$, $XH^{(2)}$ and $XL^{(2)}$, we utilize IDWT to recomposite the denoising HSI into the original resolution. Furthermore, to preserve spatial-spectral consistency and promote the refine reconstruction, a regularization block is organized to model the spatial-spectral correlations in the recovered scale. Combining ideas of dense connection and residual learning, as shown in Fig. 3 (b), SCRB utilizes dense shortcuts to merge features from different layers and a global residual shortcut to enhance the denoising performance. The pseudo-3D convolution mentioned before is deployed in SCRB to ensure efficient extraction of spatial-spectral features with less computational overhead. Besides, the final convolution layer acts like a regularization operator, which utilizes the $(1,1,1)$ kernel to further integrate features from previous layers and regularize the final output. The refined reconstruction can be formulated as

$$X = \mathbf{SCRB}(\mathbf{IDWT}(\mathbf{IDWT}(XL^{(2)}, [XH^{(2)}, XH^{(1)}]))). \tag{3}$$

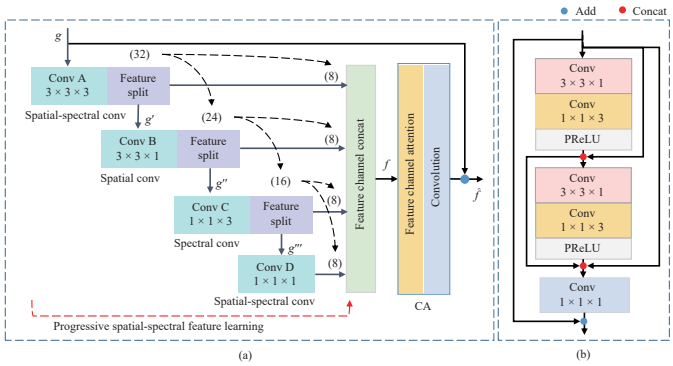

Fig. 3. Illustration of two main blocks in detail: (a) progressive spatial-spectral mixed convolution block; (b) spatial-spectral consistency regularization block.

Experiments: We organize training and evaluation of the proposed D2Net via mimicking synthetic Gaussian noise in different intensities and typically mixed noise cases on the ICVL hyperspectral dataset (http://icvl.cs.bgu.ac.il/hyperspectral/). To fairly assessment, we chose a series of advanced denoising methods for comparison, including BM4D [4], TDL [7], GLF [3], HSID-CNN [10], QRNN3D [9] and DSWN [12]. We employ five metrics for HSI denoising performance evaluation, inclusive of peak signal-to-noise ratio (PSNR), structure similarity (SSIM) and feature similarity (FSIM) for spatial-based image quality measurement, spectral angle mapper (SAM) for spectral fidelity evaluation, and time-consuming during testing per HSI.

1) Quantitative comparisons: As listed in Table 1, where Cases 1, 2, 3 are: Gaussian noise intensity of $\sigma = 70$, blind σ, mixed noise of Gaussian noise and deadline noise, D2Net represents superior in the majority of quantitative metrics, especially the spatial quality metrics. It indicates the superior flexibility of our proposed D2Net. However, it falls slightly (0.007) on the SAM metric behind the QRNN3D in the complex noise case. It might be caused by the fact that the DWT/IDWT are conducted only in the spatial domain while QRNN3D emphasis spectral features by recurrent. On the other hand, specific designs in D2Net like PMCB and SCRB have greatly compensated for the gap in spatial-spectral fidelity, bringing considerable gains involving nearly 1dB gains in PSNR and average 0.005 incomes in SSIM and FSIM, but a bit of sacrifice in speed (average 0.04 s slower than QRNN3D).

Table 1. Quantitative Results on the ICVL Dataset. The best and the second results are shown in red and blue, respectively.

Case	Method	PSNR↑	SSIM↑	FSIM↑	SAM↓	Time (s)
1	Noisy HSI	11.23	0.023	0.398	1.027	–
	BM4D [4]	33.71	0.854	0.903	0.182	305.19
	TDL [7]	36.92	0.910	0.945	0.099	47.32
	GLF [3]	37.01	0.883	0.954	0.145	537.02
	HSID-CNN [10]	36.42	0.923	0.948	0.099	5.71
	QRNN3D [9]	38.30	0.938	0.951	0.094	0.86
	DSWN [12]	37.34	0.946	0.953	0.105	1.01
	D2Net	39.86	0.951	0.957	0.087	0.91
2	Noisy HSI	17.58	0.121	0.598	0.776	–
	BM4D [4]	37.66	0.917	0.943	0.128	307.63
	TDL [7]	40.44	0.948	0.968	0.069	50.52
	GLF [3]	41.16	0.943	0.974	0.099	454.17
	HSID-CNN [10]	39.02	0.950	0.968	0.080	5.82
	QRNN3D [9]	41.65	0.965	0.972	0.076	0.89
	DSWN [12]	40.77	0.968	0.974	0.095	1.03
	D2Net	42.03	0.971	0.979	0.072	0.92
3	Noisy HSI	23.00	0.354	0.678	0.778	–
	BM4D [4]	28.56	0.539	0.898	0.338	311.25
	TDL [7]	34.12	0.906	0.901	0.110	41.96
	GLF [3]	38.68	0.963	0.974	0.076	351.82
	HSID-CNN [10]	36.76	0.933	0.964	0.088	5.86
	QRNN3D [9]	39.32	0.955	0.972	0.054	0.84
	DSWN [12]	38.17	0.945	0.942	0.159	1.06
	D2Net	40.43	0.990	0.982	0.061	0.91

2) Spatial quality comparisons: Fig. 4 illustrates denoising results of some representative scenes. From the specifics aspect, taking the second-row results in Fig. 4 as an example, traditional method like TDL has removed some noise but still obtain poor results; QRNN3D shows a cleaner result but loses some details due to excessive denoising; DSWN retains more high-frequency detail information, but fails to eliminate high-frequency noise. In contrast, our proposed D2Net obtains the best denoising results with its model superior, achieving high fidelity recovery and showing results much clearer with fewer artifacts and sharp edges. These results suggest that the proposed

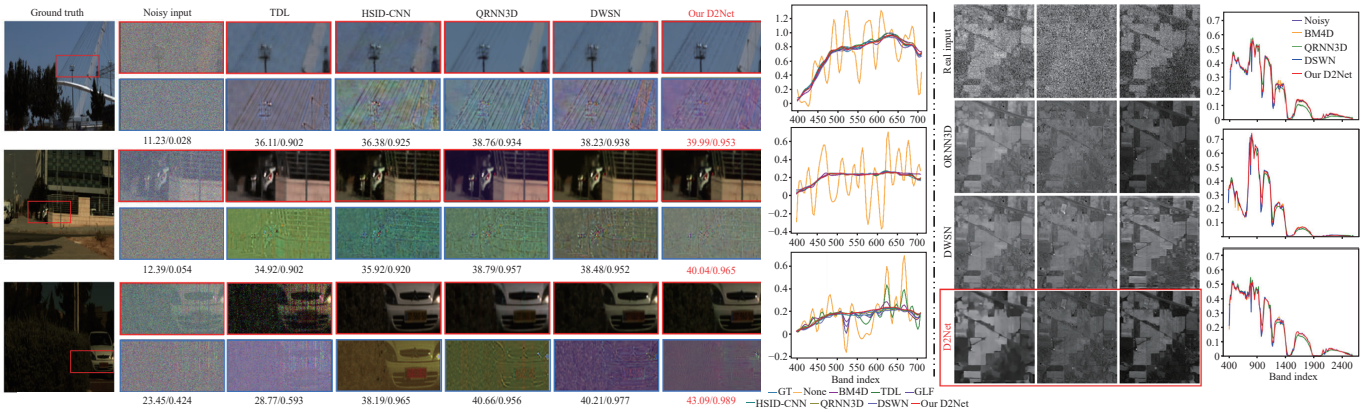

Fig. 4. Illustration of comparison experiments on synthetic noises and real HSIs. Bounded by the black dotted line, the left side shows denoising results (involving zoom-in region and residual noise map) on the ICVL dataset and recovered spectral curves of the sampled pixel, where three rows from top to bottom are in three noise cases; the right side presents denoising results on the Indian Pines dataset.

D2Net has a more robust capability to remove HSI noises.

3) Spectral fidelity comparisons: To verify the superiority of our method in spectral fidelity, we select one typical pixel in each noise case and draw the recovered spectral curves of some representative comparing methods in Fig. 4 . Apparently, compared to the other results, the recovered spectra curves of our D2Net are much closer to the reference. It indicates that our method accurately eliminates the negative effect of noise in the spectral domain and further confirms the advantage of our D2Net in maintaining high spectral fidelity.

4) Denoising on real HSIs: We also conduct denoising experiments on the Indian Pines dataset to verify the effectiveness and flexibility of our model. As Fig. 4 shown, our method still shows stable performance in dealing with varying noisy levels, confirming its generalization ability in a real-world scenario. We also draw the spectral curves of three typical pixels recovered by some leading methods in Fig. 4. It is quite clear that the spectral curves recovered by our proposed method are smoother after denoising, and their overall shapes are more consistent with the original data.

5) Ablation studies: To verify the effectiveness of three blocks in the D2Net, we conduct corresponding ablation studies and list the results in Table 2. Compared the first row with the second, replaced down-sampling/up-sampling convolution (D/U) with DWT/IDWT, denoising performance of the backbone model has improved without extra model parameters burden, except for some decreases in SAM due to neglecting the spectral properties. Similarly, the effectiveness of PMCB and SCRB also have been verified.

Conclusion: This letter insightfully combines a multi-branch network with DWT/IDWT and proposes D2Net to achieve fine recovery of noisy HSIs. On the one hand, DWT/IDWT employed in this work, which is mathematically strict symmetric, can support multi-scale feature decomposition and recomposition without information loss. On the other hand, HSI noise distribution varies in different frequency components, which inspired us to design tailor-made sub-branches and develop PMCBs to achieve accurate detachment of noise and texture details for high-frequency sub-branches. In addition, we deploy an SCRB in refined reconstruction to further explore its coherence in original resolution and enhance the spatial and spectral fidelity of the recovered HSIs. Experiments demonstrate the superiority of our D2Net. Such an idea also can be flexibly transferred or promoted to other vision tasks for future insightful research, like HSI reconstruction, super-resolution, and abnormal detection.

Table 2. Ablations on Remove Synthetic Gaussian Noise With $\sigma = 50$ on ICVl Dataset. Our D2Net is indicated by boldface.

No.	Components			Metrics			
	DWT/IDWT	PMCB	SCRB	PSNR ↑	SSIM↑	SAM ↓	Params (#)
1	D/U			36.84	0.967	0.095	0.40M
2	✓			37.61	0.970	0.113	0.40M
3	✓	CB		38.04	0.971	0.091	0.84M
4	✓	✓		38.77	0.978	0.086	0.98M
5	✓	✓	✓	**39.44**	**0.980**	**0.079**	**0.99M**

Acknowledgments: This work was supported by the National Natural Science Foundation of China (61903279).

References

[1] W. Cho, J. Jang, A. Koschan, M. A. Abidi, and J. Paik, "Hyperspectral face recognition using improved inter-channel alignment based on qualitative prediction models," *Optics Express*, vol. 24, no. 24, pp. 27637–27662, 2016.

[2] C. Wu, B. Du, and L. Zhang, " Hyperspectral anomalous change detection based on joint sparse representation," *ISPRS J. Photogrammetry and Remote Sensing*, vol. 146, pp. 137–150, 2018.

[3] L. Zhuang and J. M. Bioucas-Dias, " Hyperspectral image denoising based on global and non-local low-rank factorizations," in *Proc. IEEE Int. Conf. Image Processing*, 2017, pp. 1900–1904.

[4] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Processing*, vol. 22, no. 1, pp. 119–133, 2013.

[5] L. Geng, Z. Ji, Y. Yuan, and Y. Yin, " Fractional-order sparse representation for image denoising," *IEEE/CAA J. Autom. Sinica* , vol. 5, no. 2, pp. 555–563, 2018.

[6] Q. Xie, Q. Zhao, D. Meng, Z. Xu, S. Gu, W. Zuo, and L. Zhang, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, 2016, pp. 1692–1700.

[7] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2014, pp. 2949–2956.

[8] Q. Lian, W. Yan, X. Zhang, and S. Chen, "Single image rain removal using image decomposition and a dense network," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 6, pp. 1428–1437, 2019.

[9] K. Wei, Y. Fu, and H. Huang, "3-D quasi-recurrent neural network for hyperspectral image denoising," *IEEE Trans. Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 363–375, 2021.

[10] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network," *IEEE Trans. Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1205–1218, 2019.

[11] Z. Hui, X. Gao, Y. Yang, and X. Wang, " Lightweight image super-resolution with information multi-distillation network," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 2024–2032.

[12] W. Liu, Q. Yan, and Y. Zhao, "Densely self-guided wavelet network for image denoising," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops*, 2020, pp. 1742–1750.