VISIONARY: Novel Spatial-Spectral Attention Mechanism for Hyperspectral Image Denoising

Aditya Dixit
IIT Indore, India
phd2201101009@iiti.ac.in

Nischit Hosamani
IIT Indore, India
cse200001054@iiti.ac.in

Puneet Gupta
IIT Indore, India
puneet@iiti.ac.in

Ankur Garg
SAC, Ahmedabad, India
agarg@sac.isro.gov.in

Qualitative denoising results of real data for various band combinations are presented in Fig. 1, where it is shown that our proposed method surpasses existing methods in terms of visual quality. We also perform experiments using the Urban [4] and RealHSI datasets [9]. The visual results is shown in Fig. 4. It's clear that VISIONARY delivers cleaner and sharper images.

For a thorough evaluation, we compared HSI classification accuracy before and after denoising using various methods Fig. 2 shows SVM classification results for the IP and PU datasets and Fig. 5 shows SVM classification results for the Urban dataset, while Table 1 provides the quantitative results of the same dataset. Post-denoising, our proposed method surpasses existing methods in terms of visual quality, indicating superior denoising performance. Additionally, Fig. 3 provides spectral profiles to assess further the potency of different denoising methods in the spectral domain, showing that our method outperforms all others in spectral performance and closely matches the ground truth.

1. Computational Efficiency

We give an overview of the computational efficiency of our proposed method in terms of GFLOPs, number of parameters, and inference/training time. The GFLOP for VISIONARY is 144.47, which is relatively lesser than the GFLOPS of state-of-the-art methods like SERT with values of 1018.9 and comparable to UNFOLD's value of 141.5. Additionally, the parameter count is relatively comparable but gives better results as a trade-off. Moreover, the average inference time of VISIONARY is 5.1 seconds, which is better than UNFOLD [9] and relatively comparable to SERT [17], which have average inference times of 5.8 and 3.2 seconds, respectively. The training time of VISIONARY is 15 hours.

References

[1] Yi Chang, Luxin Yan, and Sheng Zhong. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4260–4268, 2017. 2

[2] Aditya Dixit, Anup Kumar Gupta, Puneet Gupta, Saurabh Srivastava, and Ankur Garg. UNFOLD: 3D U-Net, 3D CNN and 3D Transformer based Hyperspectral Image Denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–10, 2023. 2

[3] Wei He, Quanming Yao, Chao Li, Naoto Yokoya, and Qibin Zhao. Non-local meets global: An integrated paradigm for hyperspectral denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6868–6877, 2019. 2

[4] Linda S Kalman and Edward M Bassett III. Classification and material identification in an urban environment using hydice hyperspectral data. In *Imaging Spectrometry III*, volume 3118, pages 57–68. SPIE, 1997. 1, 3

[5] Miaoyu Li, Ying Fu, and Yulun Zhang. Spatial-Spectral Transformer for hyperspectral image denoising. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 1368–1376, 2023. 2

[6] Miaoyu Li, Ji Liu, Ying Fu, Yulun Zhang, and Dejing Dou. Spectral Enhanced Rectangle Transformer for hyperspectral image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5805–5814, 2023. 2

[7] Kaixuan Wei, Ying Fu, and Hua Huang. 3-D quasi-recurrent neural network for hyperspectral image denoising. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):363–375, 2020. 2

[8] Qiangqiang Yuan, Qiang Zhang, Jie Li, Huanfeng Shen, and Liangpei Zhang. Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):1205–1218, 2018. 2

[9] Tao Zhang, Ying Fu, and Cheng Li. Hyperspectral image denoising with realistic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2248–2257, 2021. 1, 3

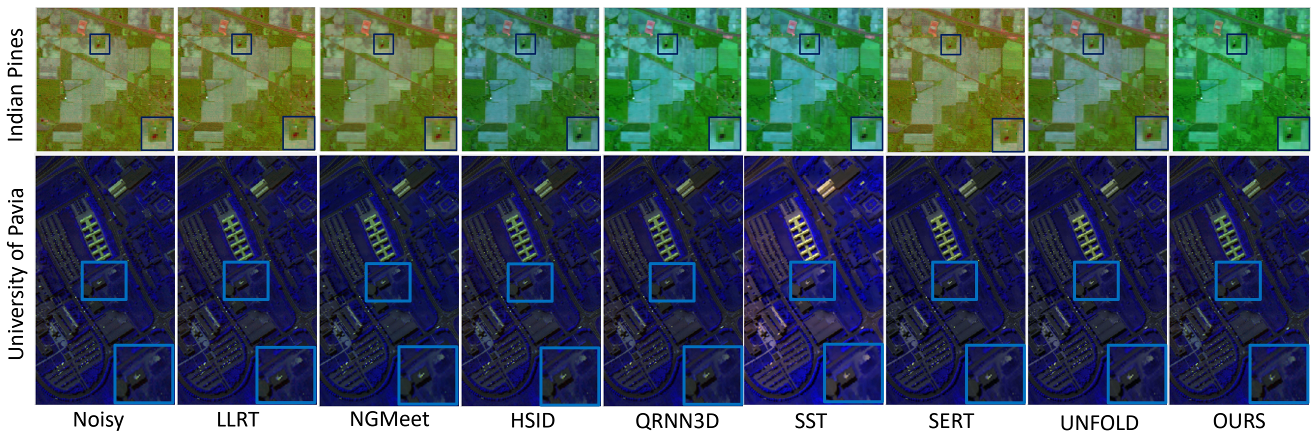

Figure 1. Real denoised results on IP and PU datasets for bands (3, 103, 203) and (5, 13, 95), respectively.

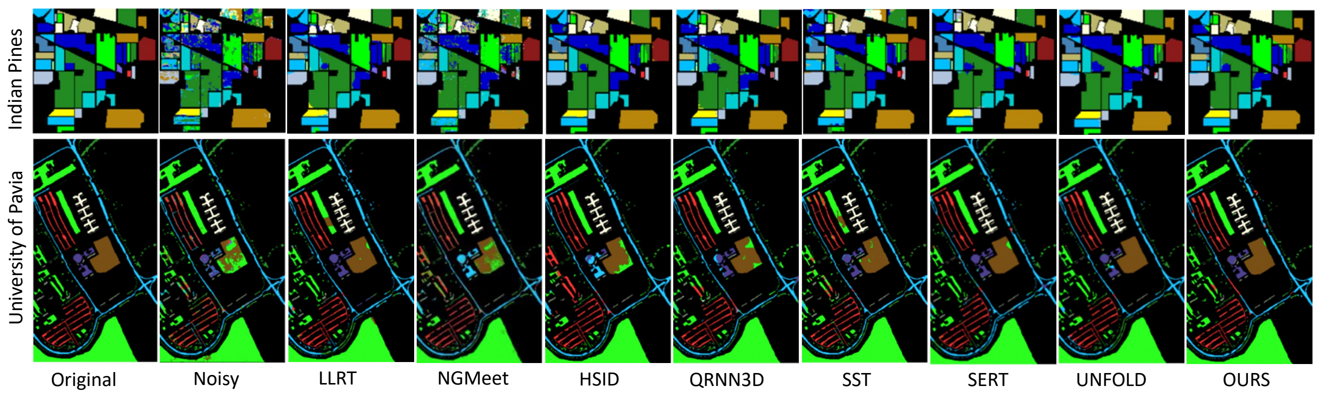

Figure 2. Classification outcomes for IP and PU datasets.

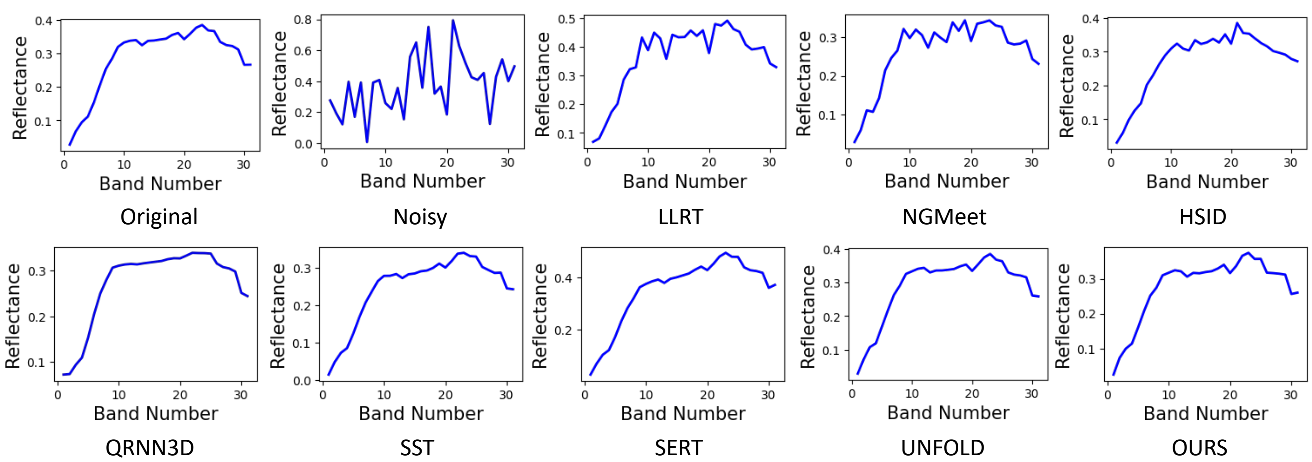

Figure 3. Spectral curves of the ICVL dataset at coordinates (110, 110) under Gaussian noise ($\sigma = 50$).

Table 1. Classification outcomes on Urban Dataset Pre- and Post-HSI Denoising

Dataset	**Index**	Noisy	[1]	[3]	[8]	[7]	[5]	[6]	[2]	**OURS**
Urban	OA	86.11	91.27	92.33	92.76	92.99	94.17	94.91	95.13	**96.02**
	Kappa	0.8385	0.8847	0.8909	0.9088	0.9086	0.9189	0.9267	0.9405	**0.9528**

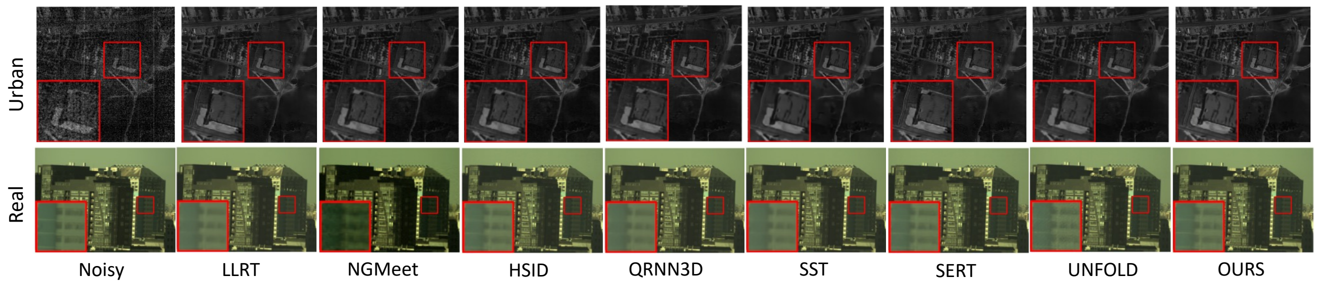

Figure 4. Visual outcomes for real-world noise reduction on the Urban [4] and Real [9] Datasets.

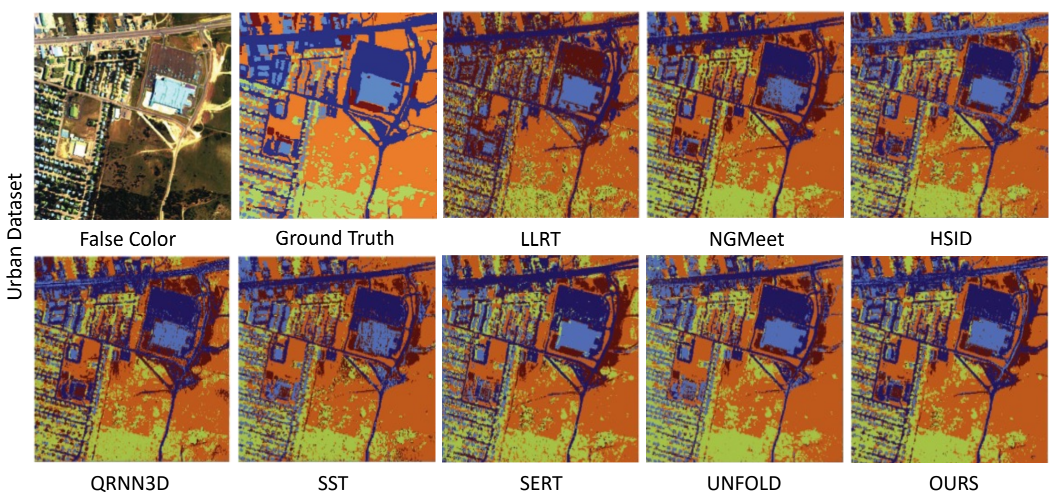

Figure 5. Classification outcomes for Urban [4] dataset.