

STAR-Net: An Interpretable Tensor Representation Network for Hyperspectral Image Denoising

Xianchao Xiu

Department of Automation



International Conference on Optimization and Statistics with Applications, June 21-23, 2024

Joint work with [Jingjing Liu](#) (SHU), [Wanquan Liu](#) (SYSU) and others

Outline

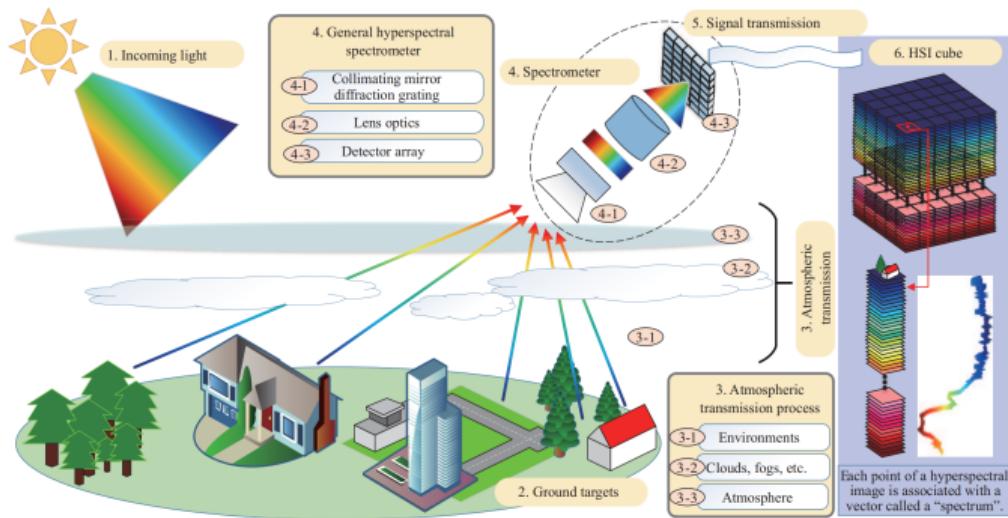
Introduction

Methodology

Experiment

Conclusion

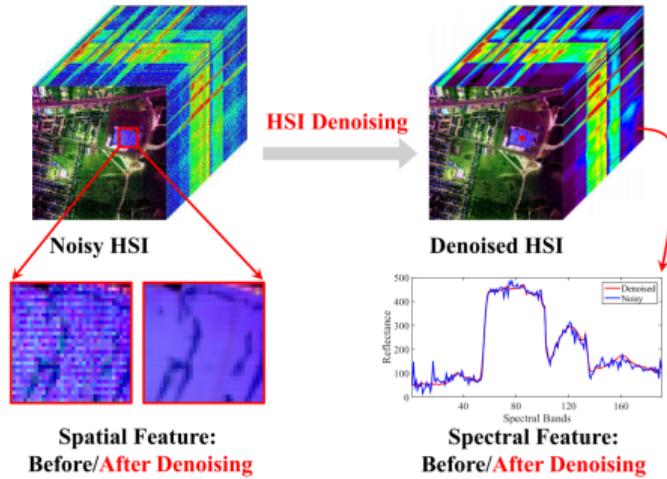
- ▶ Hyperspectral image (HSI): Rich spectral and spatial information



- ▶ Board applications: Agriculture, astronomy, geosciences, medicine
- ▶ Various tasks: Denoising, classification, detection, fusion, unmixing
- ▶ <https://github.com/xianchaoxiu/Hyperspectral-Imaging>

Denoising

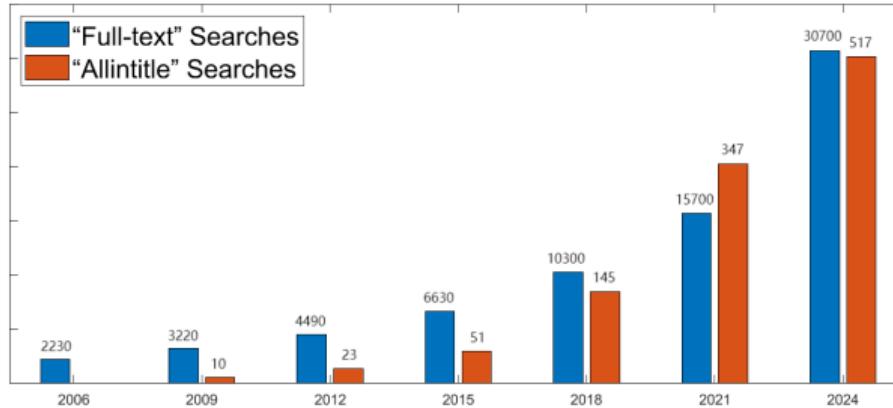
- ▶ HSI denoising/HSI restoration



- ▶ Taxonomy of existing methods
 - ▶ **Model-driven:** Non-local, TV, sparse coding, low-rank representation
 - ▶ **Data-driven:** CNN, hybrid networks, unsupervised networks
 - ▶ **Model-data-driven**

Tensor

- ▶ The development of “tensor hyperspectral” on Google Scholar



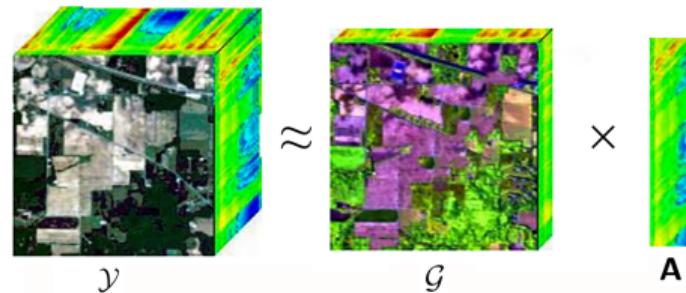
- ▶ What are the advantages of tensor modeling?
 - ▶ Excellent data representation
 - ▶ Various tensor decomposition
 - ▶ (possibly) Low computational complexity

Motivation

- ▶ General subspace tensor representation framework

$$(P) \quad \min_{\mathcal{G}, \mathbf{A}} \frac{1}{2} \|\mathcal{Y} - \mathcal{G} \times_3 \mathbf{A}\|_F^2 + \lambda \Omega(\mathcal{G})$$

s.t. $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$



- ▶ Recent surveys
 - ▶ Liu-Li-Wang-Tao-Du-Chanussot, SCIS, 2023
 - ▶ Wang-Hong-Han-Li-Yao-Gao, IEEE GRSM, 2023

Motivation

- Xiong-Zhou-Tao-Lu-Zhou-Qian, IEEE TIP, 2022

$$(P) \quad \min_{\mathcal{G}, \mathbf{A}} \quad \frac{1}{2} \|\mathcal{Y} - \mathcal{G} \times_3 \mathbf{A}\|_{\text{F}}^2 + \lambda \Omega(\mathcal{G})$$

s.t. $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$

↓

$$\min_{\mathcal{G}, \mathcal{B}_i, \mathbf{A}} \quad \frac{1}{2} \|\mathcal{Y} - \mathcal{G} \times_3 \mathbf{A}\|_{\text{F}}^2 + \lambda \sum_i (\phi(\mathcal{G}, \mathcal{B}_i) + \gamma_1 \|\mathcal{B}_i\|_1)$$

s.t. $\mathbf{A}^\top \mathbf{A} = \mathbf{I}$

- Multidimensional representation

$$\phi(\mathcal{G}, \mathcal{B}_i) = \frac{1}{2} \|\mathcal{R}_i \mathcal{G} - \mathcal{B}_i \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times_3 \mathbf{D}_3\|_{\text{F}}^2$$

How to characterize priors? How to develop algorithms?

Outline

Introduction

Methodology

Experiment

Conclusion

Model

- Sparse tensor aided representation (STAR)

$$\begin{aligned} (\text{P}) \quad & \min_{\mathcal{G}, \mathbf{A}} \quad \frac{1}{2} \|\mathcal{Y} - \mathcal{G} \times_3 \mathbf{A}\|_{\text{F}}^2 + \lambda \Omega(\mathcal{G}) \\ \text{s.t.} \quad & \mathbf{A}^\top \mathbf{A} = \mathbf{I} \end{aligned}$$

↓

$$\begin{aligned} (\text{STAR}) \quad & \min_{\mathcal{G}, \mathcal{B}_i, \mathbf{A}} \quad \frac{1}{2} \|\mathcal{Y} - \mathcal{G} \times_3 \mathbf{A}\|_{\text{F}}^2 + \lambda \sum_i (\phi(\mathcal{G}, \mathcal{B}_i) + \gamma_1 \|\mathcal{B}_i\|_1 + \gamma_2 \|\mathcal{B}_i\|_*) \\ \text{s.t.} \quad & \mathbf{A}^\top \mathbf{A} = \mathbf{I} \end{aligned}$$

↓

$$\begin{aligned} (\text{STAR-S}) \quad & \min_{\mathcal{G}, \mathcal{S}, \mathcal{B}_i, \mathbf{A}} \quad \frac{1}{2} \|\mathcal{Y} - \mathcal{G} \times_3 \mathbf{A} - \mathcal{S}\|_{\text{F}}^2 + \mu \|\mathcal{S}\|_1 \\ & + \lambda \sum_i (\phi(\mathcal{G}, \mathcal{B}_i) + \gamma_1 \|\mathcal{B}_i\|_1 + \gamma_2 \|\mathcal{B}_i\|_*) \\ \text{s.t.} \quad & \mathbf{A}^\top \mathbf{A} = \mathbf{I} \end{aligned}$$

Algorithm

- ▶ Alternating direction method of multipliers (ADMM)

$$\begin{aligned} \min_{\mathcal{G}, \mathcal{B}_i, \mathcal{L}_i, \mathbf{A}} \quad & \frac{1}{2} \|\mathcal{Y} - \mathcal{G} \times_3 \mathbf{A}\|_{\text{F}}^2 + \lambda \sum_i (\phi(\mathcal{G}, \mathcal{B}_i) + \gamma_1 \|\mathcal{B}_i\|_1 + \gamma_2 \|\mathcal{L}_i\|_*) \\ \text{s.t.} \quad & \mathbf{A}^\top \mathbf{A} = \mathbf{I}, \quad \mathcal{L}_i = \mathcal{B}_i \end{aligned}$$

↓

$$\begin{aligned} L_\beta(\mathcal{G}, \mathcal{B}_i, \mathcal{L}_i, \mathbf{A}, \mathcal{P}_i) = & \frac{1}{2} \|\mathcal{Y} - \mathcal{G} \times_3 \mathbf{A}\|_{\text{F}}^2 + \langle \mathcal{P}_i, \mathcal{L}_i - \mathcal{B}_i \rangle \\ & + \lambda \sum_i (\phi(\mathcal{G}, \mathcal{B}_i) + \gamma_1 \|\mathcal{B}_i\|_1 + \gamma_2 \|\mathcal{L}_i\|_*) + \frac{\beta}{2} \|\mathcal{L}_i - \mathcal{B}_i\|_{\text{F}}^2 \end{aligned}$$

- ▶ From iterative optimization to deep unrolling
 - ▶ Gregor-LeCun, ICML, 2010
 - ▶ Yang-Sun-Li-Xu, NIPS, 2016
 - ▶ Monga-Li-Eldar, IEEE SPM, 2021
 - ▶ Elad-Kawar-Vaksman, SIIMS, 2023

STAR-Net

- ▶ Update \mathcal{G} -block

$$\begin{aligned}\mathcal{G}^{k+1} = & \arg \min_{\mathcal{G}} \frac{1}{2} \|\mathcal{Y} - \mathcal{G} \times_3 \mathbf{A}^k\|_{\text{F}}^2 \\ & + \frac{\lambda}{2} \sum_i \|\mathcal{R}_i \mathcal{G} - \mathcal{B}_i^k \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times_3 \mathbf{D}_3\|_{\text{F}}^2 \\ & \Downarrow\end{aligned}$$

$$\begin{aligned}\mathcal{G}^{k+1} = & (\mathcal{I} + \lambda \sum_i \mathcal{R}_i^\top \mathcal{R}_i)^{-1} \left(\lambda \sum_i \mathcal{R}_i^\top \mathcal{B}_i^k \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times_3 \mathbf{D}_3 + \mathcal{Y} \times_3 \mathbf{A}^{k\top} \right) \\ & \Downarrow\end{aligned}$$

$$\mathcal{G}^{k+1} = \mathcal{E}_1 * \mathcal{E}_2$$

↓

$$\mathcal{G}^{k+1} = \text{LargNet}(\mathcal{E}_1, \mathcal{E}_2)$$

STAR-Net

- ▶ Update \mathcal{B}_i -block

$$\mathcal{B}_i^{k+1} = \arg \min_{\mathcal{B}_i} \frac{\lambda}{2} \|\mathcal{R}_i \mathcal{G}^{k+1} - \mathcal{B}_i \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times_3 \mathbf{D}_3\|_F^2$$

$$+ \frac{\beta}{2} \|\mathcal{L}_i^k - \mathcal{B}_i + \mathcal{P}_i^k / \beta\|_F^2 + \lambda \gamma_1 \|\mathcal{B}_i\|_1$$

↓

$$\mathcal{B}_i^{k+1} = \arg \min_{\mathcal{B}_i} \frac{1}{2} \|(\beta \mathcal{I} + \lambda \mathcal{I} \times_1 \mathbf{D}_1 \times_2 \mathbf{D}_2 \times_3 \mathbf{D}_3) \mathcal{B}_i\|_F^2$$

$$- (\lambda \mathcal{R}_i \mathcal{G}^{k+1} + \beta \mathcal{L}_i^k + \mathcal{P}_i^k)\|_F^2 + \lambda \gamma_1 \|\mathcal{B}_i\|_1$$

↓

$$\mathcal{B}_i^{k+1} = \text{Shrink}(\mathcal{F}_i, \lambda \gamma_1 / \nu)$$

↓

$$\mathcal{B}_i^{k+1} = \text{sgn}(\mathcal{F}_i) \text{ReLU}(|\mathcal{F}_i| - \lambda \gamma_1 / \nu)$$

STAR-Net

- ▶ Update \mathbf{A} -block

$$\mathbf{A}^{k+1} = \arg \min_{\mathbf{A}^\top \mathbf{A} = \mathbf{I}} \frac{1}{2} \|\mathcal{Y} - \mathcal{G}^{k+1} \times_3 \mathbf{A}\|_F^2$$

↓

$$\mathbf{A}^{k+1} = \mathbf{U} \mathbf{V}^\top$$

↓

$$\mathbf{A}^{k+1} = \text{LargNet}(\mathbf{U}, \mathbf{V}^\top)$$

- ▶ Update \mathcal{P}_i -block

$$\mathcal{P}_i^{k+1} = \mathcal{P}_i^k + \beta(\mathcal{L}_i^{k+1} - \mathcal{B}_i^{k+1})$$

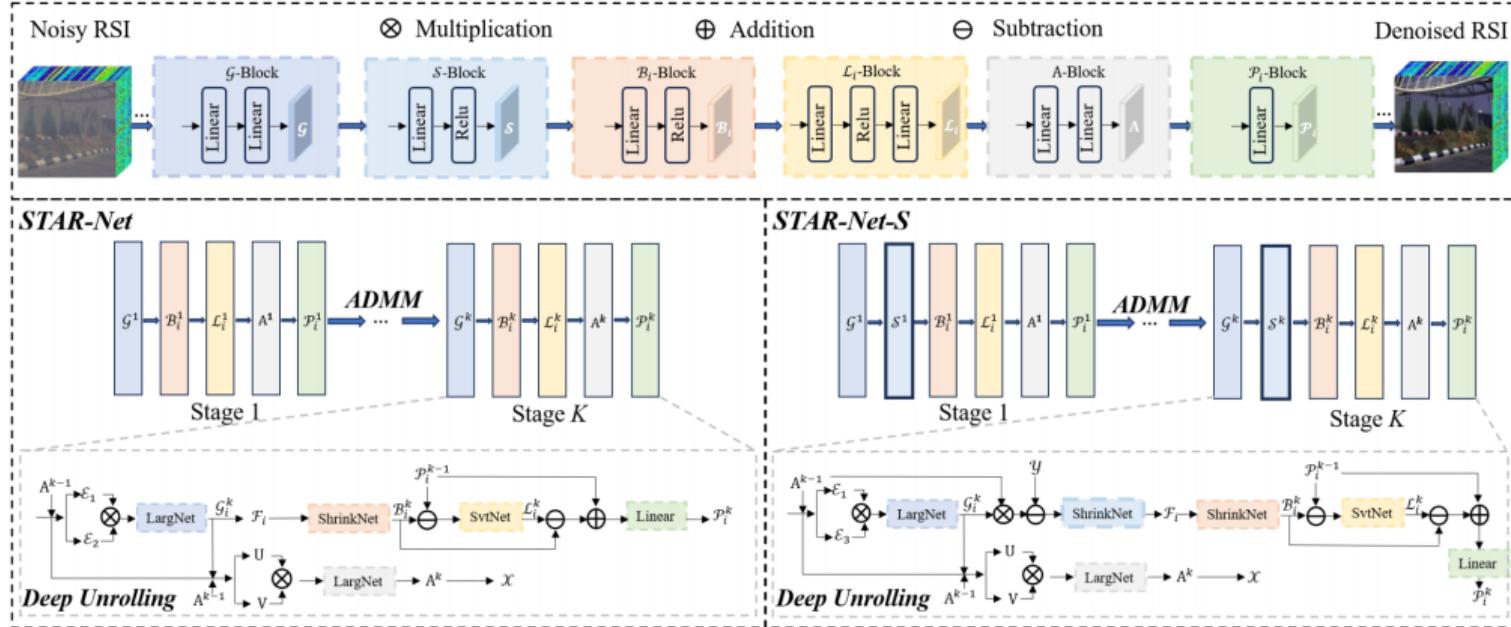
↓

$$\mathcal{P}_i^{k+1} = \text{Linear}(\Theta_i)$$

STAR-Net

- ▶ **Input:** Noisy HSI \mathcal{Y} , parameters $\lambda, \beta, \gamma_1, \gamma_2, \nu$
- ▶ **Initialize:** $(\mathcal{G}^0, \mathcal{B}_i^0, \mathcal{L}_i^0, \mathbf{A}^0, \mathcal{P}_i^0)$
- ▶ **While** $k = 1, \dots, K$ **do**
 - ▶ Update \mathcal{G} -block by
$$\mathcal{G}^{k+1} = \text{LargNet}(\mathcal{E}_1, \mathcal{E}_2)$$
 - ▶ Update \mathcal{B}_i -block by
$$\mathcal{B}_i^{k+1} = \text{ShrinkNet}(\mathcal{F}_i, \lambda\gamma_1/\nu)$$
 - ▶ Update \mathcal{L}_i -block by
$$\mathcal{L}_i^{k+1} = \text{SvtNet}(\mathcal{B}_i^{k+1} - \mathcal{P}_i^k/\beta, \lambda\gamma_2/\beta)$$
 - ▶ Update \mathbf{A} -block by
$$\mathbf{A}^{k+1} = \text{LargNet}(\mathbf{U}, \mathbf{V}^\top)$$
 - ▶ Update \mathcal{P}_i -block by
$$\mathcal{P}_i^{k+1} = \text{Linear}(\Theta_i)$$
- ▶ **Output:** Denoised HSI $\mathcal{X} = \mathcal{G}^{k+1} \times_3 \mathbf{A}^{k+1}$

Architecture



Outline

Introduction

Methodology

Experiment

Conclusion

Setup

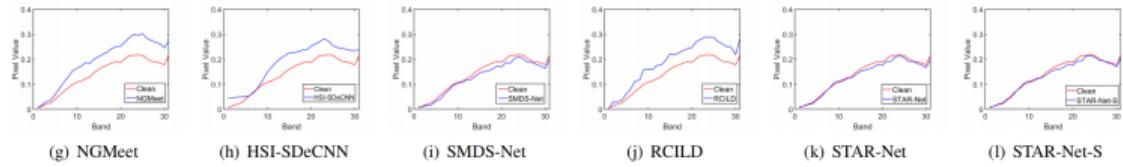
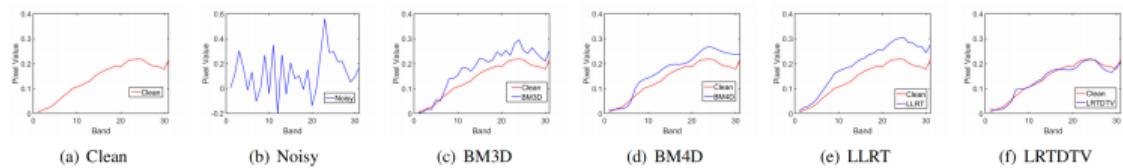
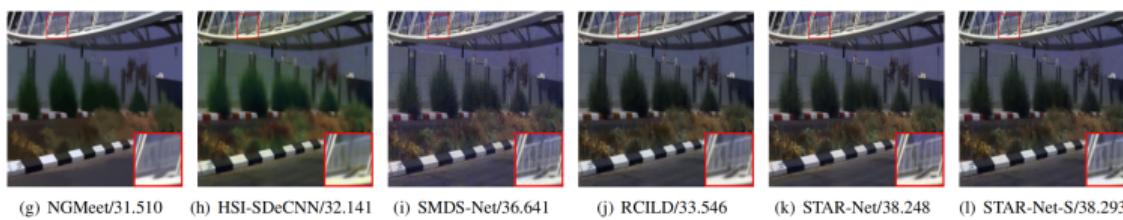
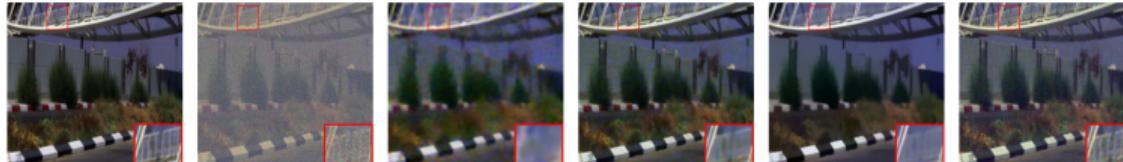
- ▶ Compared methods
 - ▶ BM3D: Dabov-Katkovnik-Egiazarian, IEEE TIP, 2007
 - ▶ BM4D: Maggioni-Katkovnik-Egiazarian-Foi, IEEE TIP, 2012
 - ▶ LLRT: Chang-Yan-Zhong, CVPR, 2017
 - ▶ LRTDTV: Wang-Chen-Han-He, RS, 2017
 - ▶ NGMeet: He-Yao-Li-Yokoya-Zhao-Zhang-Zhang, IEEE TPAMI, 2022
 - ▶ HSI-SDeCNN: Maffei-Haut-Paoletti-Plaza, IEEE TGRS, 2020
 - ▶ SMDS-Net: Xiong-Zhou-Tao-Lu-Zhou-Qian, IEEE TIP, 2022
 - ▶ RCILD: Peng-Wang-Cao-Zhao-Yao-Zhang-Meng, IEEE TGRS, 2024
- ▶ Implementation details
 - ▶ Training loss: $L = \|\text{STAR-Net}(\mathcal{Y}) - \mathcal{X}\|_F^2$
 - ▶ Training dataset: 100 HSIs from ICVL, data augmentation
 - ▶ Testing dataset: ICVL, Washington DC Mall
 - ▶ Hyperparameters: Learning rate=0.005, batch=2, epoch=300, K=6, dictionary=[9, 9, 9]
 - ▶ Training parameters: $\lambda, \beta, \mu, \gamma_1, \gamma_2, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$

Synthetic

► Quantitative comparisons on ICSVL

σ	Index	Noisy	Model-based methods					Deep learning-based methods				
			BM3D	BM4D	LLRT	LRTDTV	NGMeet	HSI-SDeCNN	SMDS-Net	RCILD	STAR-Net	STAR-Net-S
10	PSNR \uparrow	29.018	37.310	42.987	39.810	43.882	42.383	41.519	46.371	42.458	47.286	47.345
	SSIM \uparrow	0.521	0.924	0.973	0.962	0.979	0.966	0.969	0.985	0.987	0.988	0.989
	SAM \downarrow	0.229	0.121	0.080	0.045	0.077	0.074	0.075	0.028	0.044	0.025	0.025
30	PSNR \uparrow	21.591	32.582	37.630	34.250	38.245	36.791	36.840	42.337	38.514	42.435	42.500
	SSIM \uparrow	0.146	0.846	0.930	0.921	0.877	0.915	0.926	0.972	0.971	0.972	0.972
	SAM \downarrow	0.535	0.208	0.142	0.084	0.149	0.139	0.124	0.040	0.067	0.039	0.038
50	PSNR \uparrow	18.402	29.982	35.242	32.067	33.618	34.399	34.342	37.481	35.838	39.853	39.963
	SSIM \uparrow	0.042	0.790	0.888	0.899	0.862	0.887	0.893	0.907	0.951	0.956	0.956
	SAM \downarrow	0.779	0.264	0.190	0.107	0.195	0.177	0.134	0.066	0.092	0.050	0.047
70	PSNR \uparrow	18.126	28.654	33.586	30.746	30.565	32.389	32.794	37.197	33.980	37.342	38.237
	SSIM \uparrow	0.038	0.742	0.844	0.852	0.762	0.858	0.855	0.923	0.930	0.943	0.943
	SAM \downarrow	0.897	0.310	0.231	0.214	0.304	0.217	0.186	0.066	0.132	0.058	0.055
Ave.	PSNR \uparrow	21.784	32.132	37.361	34.218	36.578	36.490	36.374	40.463	37.212	41.729	42.011
	SSIM \uparrow	0.187	0.826	0.909	0.909	0.870	0.907	0.911	0.943	0.953	0.965	0.965
	SAM \downarrow	0.610	0.226	0.161	0.113	0.181	0.152	0.130	0.050	0.077	0.043	0.041

Synthetic

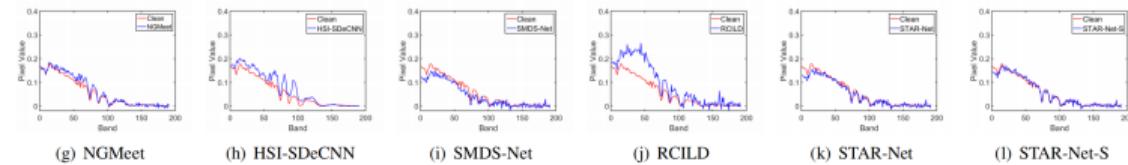
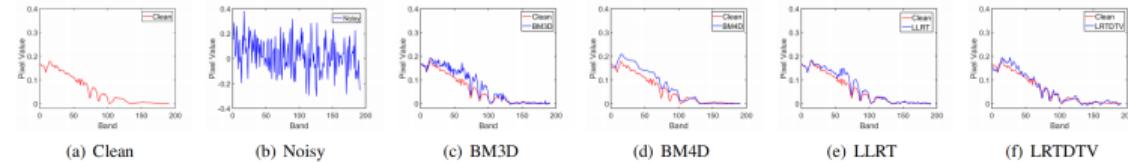
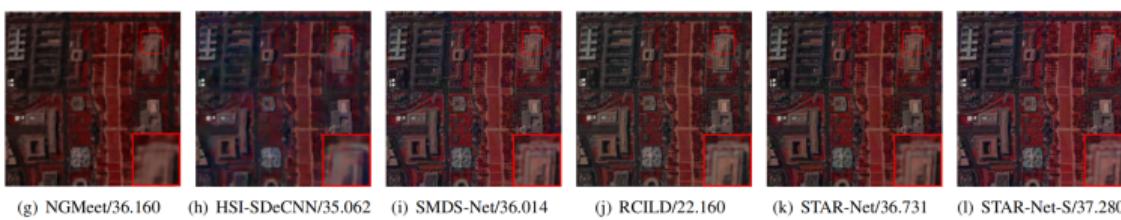


Synthetic

► Quantitative comparisons on WDC

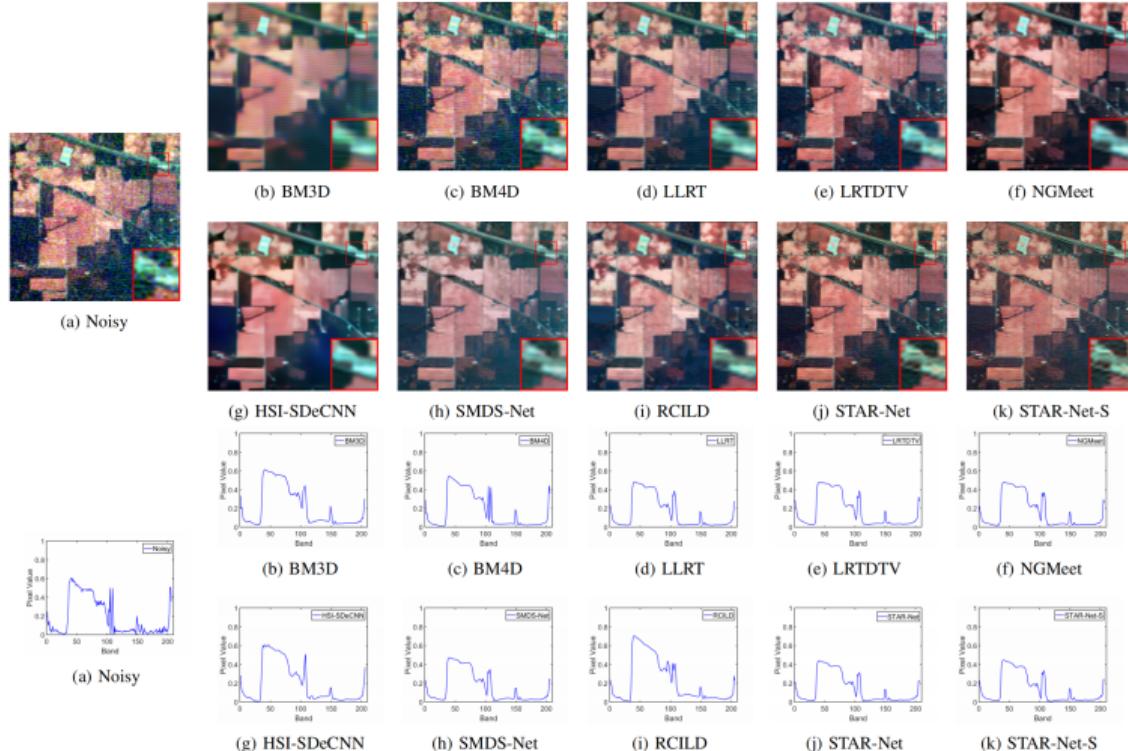
σ	Index	Noisy	Model-based methods					Deep learning-based methods				
			BM3D	BM4D	LLRT	LRTDTV	NGMeet	HSI-SDeCNN	SMDS-Net	RCILD	STAR-Net	STAR-Net-S
10	PSNR \uparrow	28.561	36.142	38.435	39.540	40.887	42.712	40.756	43.616	28.968	46.173	46.410
	SSIM \uparrow	0.522	0.923	0.946	0.966	0.909	0.978	0.937	0.961	0.972	0.980	0.982
	SAM \downarrow	0.738	0.338	0.271	0.265	0.236	0.200	0.304	0.066	0.215	0.043	0.042
30	PSNR \uparrow	20.897	30.869	32.793	32.681	38.887	37.391	36.710	39.498	22.509	39.924	39.950
	SSIM \uparrow	0.150	0.784	0.834	0.858	0.888	0.854	0.832	0.916	0.882	0.923	0.928
	SAM \downarrow	1.020	0.519	0.445	0.387	0.335	0.260	0.399	0.085	0.243	0.083	0.083
50	PSNR \uparrow	17.778	28.882	30.901	30.470	35.250	36.160	35.062	36.014	22.160	36.731	37.280
	SSIM \uparrow	0.066	0.702	0.770	0.789	0.813	0.793	0.771	0.834	0.872	0.852	0.871
	SAM \downarrow	1.164	0.603	0.528	0.433	0.504	0.340	0.446	0.124	0.263	0.122	0.121
70	PSNR \uparrow	16.966	27.694	30.140	29.159	34.198	35.110	32.891	35.315	21.262	35.964	36.286
	SSIM \uparrow	0.051	0.657	0.740	0.761	0.776	0.761	0.602	0.811	0.857	0.831	0.843
	SAM \downarrow	1.205	0.652	0.560	0.433	0.547	0.500	0.980	0.135	0.290	0.130	0.129
Ave.	PSNR \uparrow	21.051	30.897	33.067	32.962	37.305	37.843	36.355	38.611	23.725	39.698	39.982
	SSIM \uparrow	0.197	0.766	0.823	0.844	0.847	0.846	0.786	0.881	0.896	0.897	0.906
	SAM \downarrow	1.032	0.528	0.451	0.380	0.406	0.325	0.532	0.102	0.253	0.095	0.094

Synthetic



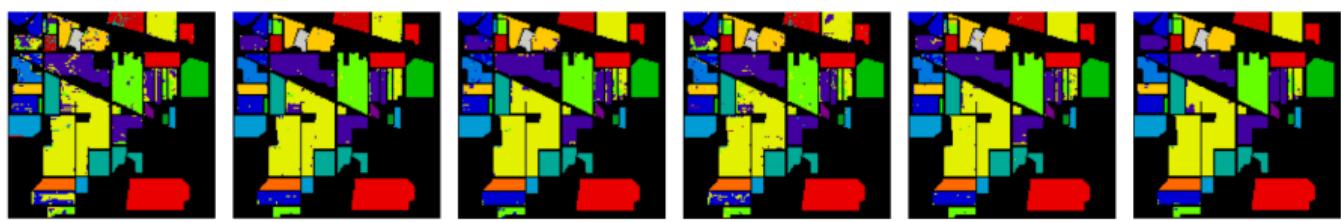
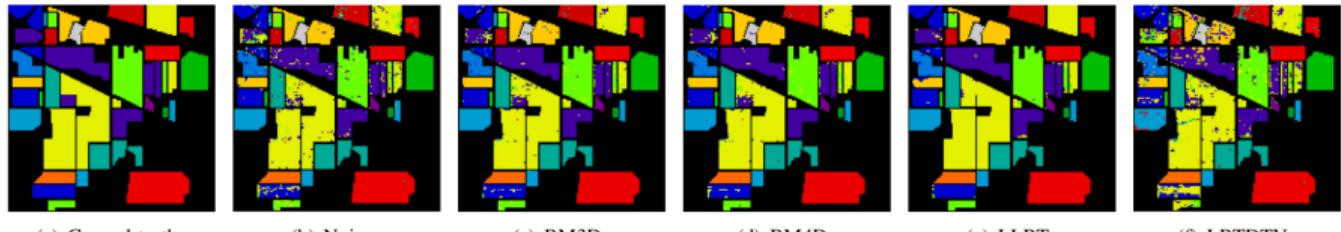
Real-world

► Case study on Indian Pines



Real-world

► Classification



Discussion

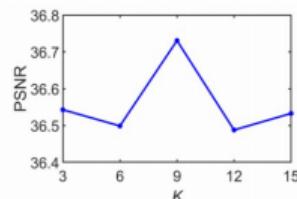
▶ Number of parameters

Methods	HSI-SDeCNN	SMDS-Net	RCILD	STAR-Net	STAR-Net-S
#Parameters	1,892,100	5,103	2,892,288	27,702	28,487

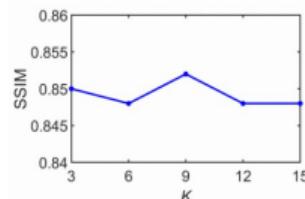
▶ Average runtime

Methods	HSI-SDeCNN	SMDS-Net	RCILD	STAR-Net	STAR-Net-S
Time	11.027	293.606	30.706	233.106	238.366

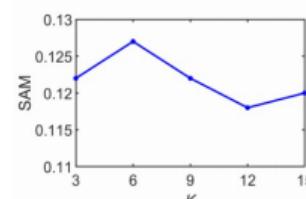
▶ Impact of unrolling iterations



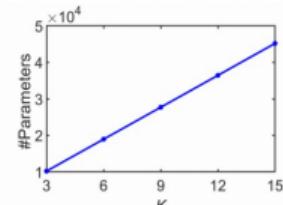
(a) PSNR



(b) SSIM



(c) SAM



(d) #Parameters

Outline

Introduction

Methodology

Experiment

Conclusion

Conclusion

- ▶ What have we done?
 - ▶ How to characterize priors? ⇒ Tensor + non-local self-similarity
 - ▶ How to develop algorithms? ⇒ ADMM + deep unrolling

Dataset	Index	Noisy	HSI-SDeCNN	SMDS-Net	RCILD	STAR-Net	STAR-Net-S
ICVL	PSNR ↑	18.402	34.342	37.481	35.838	39.853	39.963
	SSIM ↑	0.042	0.893	0.907	0.951	0.956	0.956
	SAM ↓	0.779	0.134	0.066	0.092	0.050	0.047
Deep unrolling		-	✗	✓	✗	✓	✓
Non-local self-similarity		-	✗	✗	✗	✓	✓

Thank you for your attention!

xcxiu@shu.edu.cn

References

- ▶ W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, Non-Local Meets Global: An Integrated Paradigm for Hyperspectral Denoising, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6861-6870, 2019
- ▶ V. Monga, Y. Li, Y. C. Eldar, Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing, IEEE Signal Processing Magazine, 38: 18-44, 2021
- ▶ F. Xiong, J. Zhou, S. Tao, J. Lu, J. Zhou, Y. Qian, SMDS-Net: Model Guided Spectral-Spatial network for Hyperspectral Image Denoising, IEEE Transactions on Image Processing, 31: 5469-5483, 2022
- ▶ L. Zhuang, M. K. Ng, FastHyMix: Fast and Parameter-Free Hyperspectral Image Mixed Noise Removal, IEEE Transactions on Neural Networks and Learning Systems, 34: 4702-4716, 2023
- ▶ J. Peng, H. Wang, X. Cao, Q. Zhao, J. Yao, H. Zhang, D. Meng, Learnable Representative Coefficient Image Denoiser for Hyperspectral Image, IEEE Transactions on Geoscience and Remote Sensing, 62: 1-16, 2024