Unmixing Diffusion for Self-Supervised Hyperspectral Image Denoising

Supplementary Material

Summary

As outlined in the main body of our manuscript, this supplementary material encompasses specific components: 1. Elaboration on the customized controlled reverse denoising diffusion model, detailing its methodology and implementation, 2. Details of the U-Net-like "hourglass" architecture embedded within the conditioning function Φ, 3. Comprehensive visual comparisons demonstrating various spectral unmixing strategies, and 4. Further visual comparisons, encompassing both simulated and real noise scenarios.

1. Details of Controlled Reverse Diffusion

Due to space limitations, the main body of our manuscript focuses solely on the essential procedures involved in the controlled reverse denoising diffusion model. In this section, we aim to delve deeper into the intricacies of the model by presenting additional comprehensive details:

Controlled Reverse Diffusion Process. A match at state $\mathcal{A}_t$ indicates that, given the specific noise schedule β, there exists at least one potential sample from the posterior at state $\mathcal{A}_t$ in the baseline unconditional generation process that closely approximates the provided input $\mathcal{A}'$. Consequently, a more precise image can be sampled at state $\mathcal{A}_0$ through an iterative reverse process denoted as $p(\mathcal{A}_0|\mathcal{A}_t)$ with condition $\mathcal{A}_c$. With the matched state t and trained $\epsilon_\theta(\cdot, t)$, reverse diffusion process [5] starting from $\mathcal{A}_t$ with noisy abundance $\mathcal{A}_t = \mathcal{A}'$, and the reverse process is updated as follows:

$$\mathcal{A}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathcal{A}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathcal{A}_t, t)\right) + \sqrt{1-\alpha_t}z_t, \quad (1)$$

where $z_t \sim \mathcal{N}(0,1), t \in [T]$. As [10, 12], we formulate the ancestral sampling process (1) as the discretization of reverse Stochastic Differential Equations (SDE).

$$d\mathcal{A} = \left[f(\mathcal{A},t) - g^2(t)\nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}(t))\right] dt + g(t)d\bar{\mathbf{w}}, \quad (2)$$

Recent work [10, 12] shows that as the total diffusion step "T" goes infinity and the forward series $\{\mathcal{A}_t\}_{t=1}^T$ becomes $\{\mathcal{A}(t)|t \in [0,1]\}$ indexed by continuous time variable, the diffusion process $\mathcal{A}(t)$ is actually the solution to an Itô SDE: $dX = f(\mathcal{A},t)dt + g(t)d\mathbf{w}$, where $\mathbf{w}$ represents the standard Wiener process. For example, the diffusion process with transition distribution $q(\mathcal{A}(t)|\mathcal{A}(t-1)) = \mathcal{N}(\mathcal{A}(t)|\sqrt{\alpha(t)}\mathcal{A}(t-1), (1-\alpha(t))I)$ corresponds to the SDE as follows

$$d\mathcal{A} = -\frac{1}{2}(1-\alpha(t))dt + \sqrt{1-\alpha(t)}d\mathbf{w}. \quad (3)$$

In this case, $f(\mathcal{A},t) = -\frac{1}{2}(1-\alpha(t))$ and $g(t) = \sqrt{1-\alpha(t)}$. Also, the reverse process is a solution to an SDE:

$$d\mathcal{A} = \left[f(\mathcal{A},t) - g^2(t)\nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}(t))\right] dt + g(t)d\bar{\mathbf{w}}, \quad (4)$$

Viewed through the lens of SDE, the process of sampling from $p(\mathcal{A}(0))$ can be achieved through an appropriate discretization of Equation (4). Consequently, in the context of HSI denoising, we aim to leverage the learned distribution pertaining to $\mathcal{A}$ from a diffusion model. This model inherently encapsulates the prior information of the image, aiding in the restoration of denoised HSI from the observed noisy image and a clean approximation. Together with condition $\mathcal{A}_c$ and the estimated endmembers E_y from STU as conditioning variables, we can reformulate the reverse SDE concerning $\mathcal{A}$ as

$$d\mathcal{A} = \left[f(\mathcal{A},t) - g^2(t)\nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}(t)|\mathcal{A}_c, E_y)\right] dt + g(t)d\bar{\mathbf{w}}. \quad (5)$$

where $f(\mathcal{A},t) = -\frac{1}{2}(1-\alpha(t))$ and $g(t) = \sqrt{1-\alpha(t)}$, $\bar{\mathbf{w}}$ is the reverse of the standard Wiener process. The gradient $\nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}(t))$ is commonly referred to as the 'score function' of $\mathcal{A}(t)$. Using Bayes's rule, the score function can be separated into two parts

$$\begin{aligned}&\nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}(t)|\mathcal{A}_c, E_y)\\ =&\nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}(t)) + \nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}_c, E_y|\mathcal{A}(t)). \quad (6)\end{aligned}$$

The first part can be derived under the general unconditional framework. However, the second part is intractable, since only the relation between $\mathcal{A}_c$, $\mathcal{X}$ and $p(\mathcal{A}(t)|\mathcal{A}(0))$ are known. Following [3, 10], we approximate the second term as

$$\begin{aligned}&\nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}_c, E_y|\mathcal{A}(t))\\ =&\nabla_{\mathcal{A}(t)} \log \int p(\mathcal{A}_c, E_y|\mathcal{A}(0))p(\mathcal{A}(0)|\mathcal{A}(t))d\mathcal{A}(0)\\ \approx&\nabla_{\mathcal{A}(t)} \log p(\mathcal{A}_c, E_y|\hat{\mathcal{A}}_0), \quad (7)\end{aligned}$$

where $\hat{\mathcal{A}}_0$ is the expectation of $\mathcal{A}(0)|\mathcal{A}(t)$ by Tweedie's formula:

$$\begin{aligned}&\hat{\mathcal{A}}_0(\mathcal{A}(t)) = \mathbb{E}[\mathcal{A}(0)|\mathcal{A}(t)]\\ &= \frac{1}{\sqrt{\bar{\alpha}_t}}\left[\mathcal{A}(t) + (1-\bar{\alpha}_t)\nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}(t))\right]. \quad (8)\end{aligned}$$

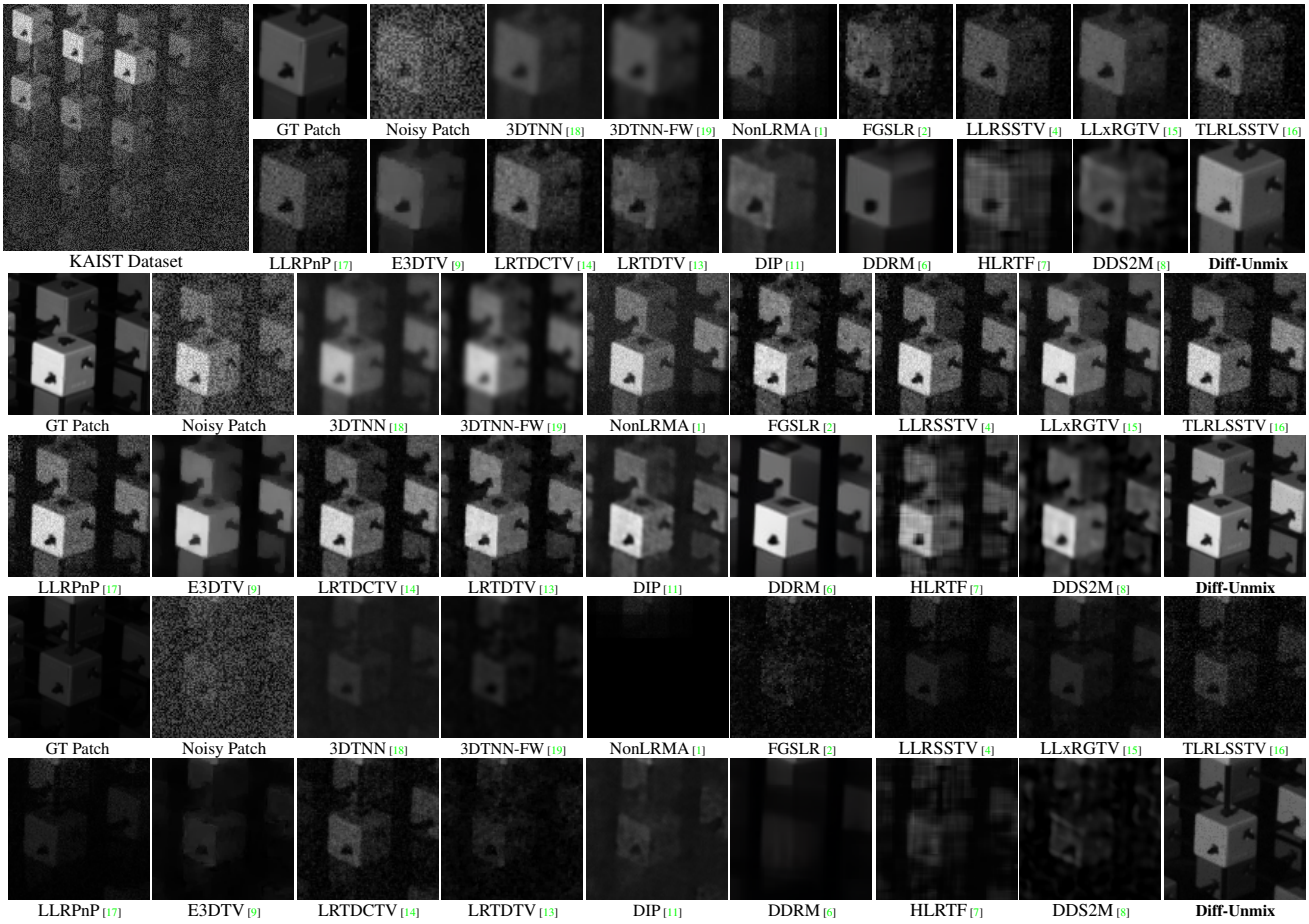

Figure 1. Visual comparison of HSI denoising methods on *KAIST* dataset.

The term $\log p(\mathcal{A}_c, E_y|\hat{\mathcal{A}}_0)$ is much more available since $\hat{\mathcal{A}}_0$ can be seen as an approximation to $\mathcal{A}$, by using computational relaxation [10], it can be relaxed and formulated as,

$$\begin{aligned}\log p(\mathcal{A}_c, E_y|\hat{\mathcal{A}}_0) &= \log p(\mathcal{A}_c, E_y|\hat{\mathcal{A}}_0) \\ &\approx -\gamma\|\mathcal{A}_c - \hat{\mathcal{A}}_0 \times_3 E_y\|_F, \end{aligned} \tag{9}$$

where γ is trade-off parameter.

Then, we discretize the reverse SDE (5) using the form of ancestral sampling process (1):

$$\begin{aligned}\mathcal{A}_{t-1} &= \frac{1}{\sqrt{\alpha_t}}\left(\mathcal{A}_t + (1-\alpha_t)\nabla_{\mathcal{A}(t)} \log p_t(\mathcal{A}(t)|\mathcal{A}_c, E_y)\right) \\ &\approx \frac{1}{\sqrt{\alpha_t}}\left(\mathcal{A}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathcal{A}_t, t)\right) + \sqrt{1-\alpha_t}z_t \\ &\quad - \eta\nabla_{\mathcal{A}_t}\|\mathcal{A}_c - \hat{\mathcal{A}}_0 \times_3 E_y\|_F, \end{aligned} \tag{10}$$

where $\eta = \frac{1-\alpha_t}{\sqrt{\alpha_t}}\gamma$. At time t, we can see that the sampling consists of two parts. The first part is equal to sampling from parameterized $p(\mathcal{A}_{t-1}|\mathcal{A}_t)$ with fixed variance $\sqrt{1-\alpha_t}$. The second part pushes the sample towards the consistent form with constraint on abundance. See supplementary materials for the detailed inference of (5) and (10).

Finally, the HSI is reconstructed through unmixing reconstruction, achieved by mixing the diffusion generative adjusted abundance map with the spectral endmembers,

$$\mathcal{X}_{\text{diff}} = \hat{\mathcal{A}}_0 \times_3 E_y. \tag{11}$$

2. U-Net-like "hourglass" architecture within Φ

In the proposed Diff-Unmix framework, Φ functions as the conditioning mechanism, introducing conditions $(\mathcal{A}_c, E_y)$ into the denoising diffusion model's sampling process. To expedite this sampling process, we adopt a U-Net-inspired "hourglass" architecture [11]. Please refer to Figure 2 for a comprehensive representation of the network architecture used to train Φ.

3. Comparison on Spectral Unmixing Strategy

In evaluating the effectiveness of the STU network, we conduct a visual analysis of the decomposition process. It's important to recognize that spectral unmixing inherently lacks an exact optimal solution, presenting an ill-posed problem. A

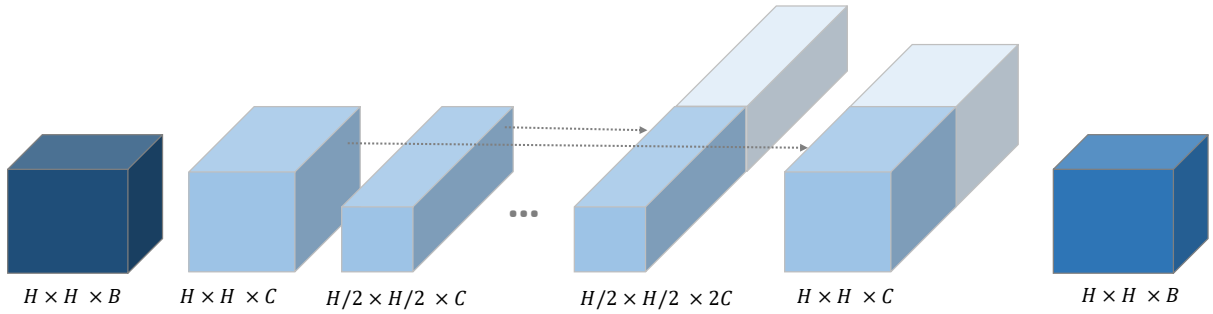

Figure 2. The network architecture within conditioning function Φ.

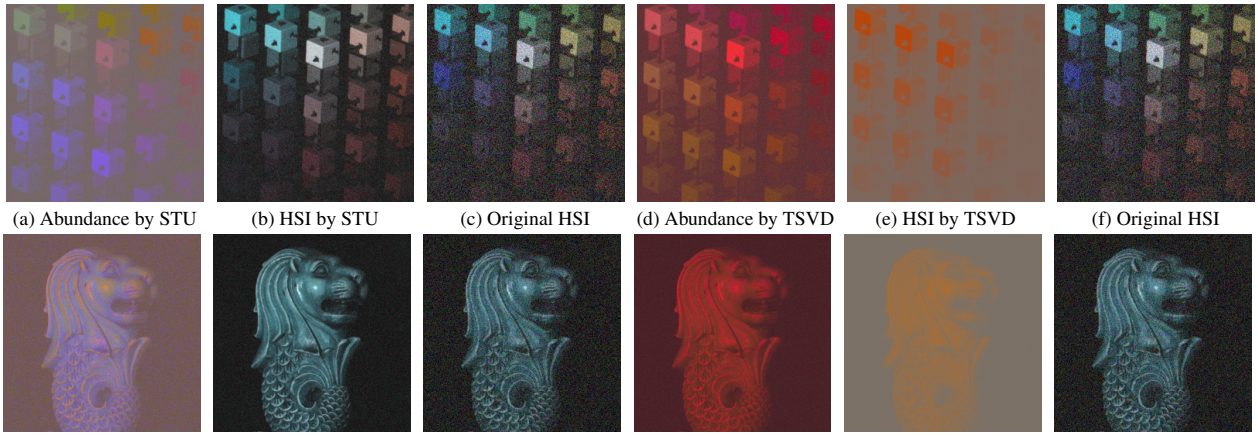

Figure 3. Comparison of different spectral unmixing strategies: STU and tensor SVD (TSVD).

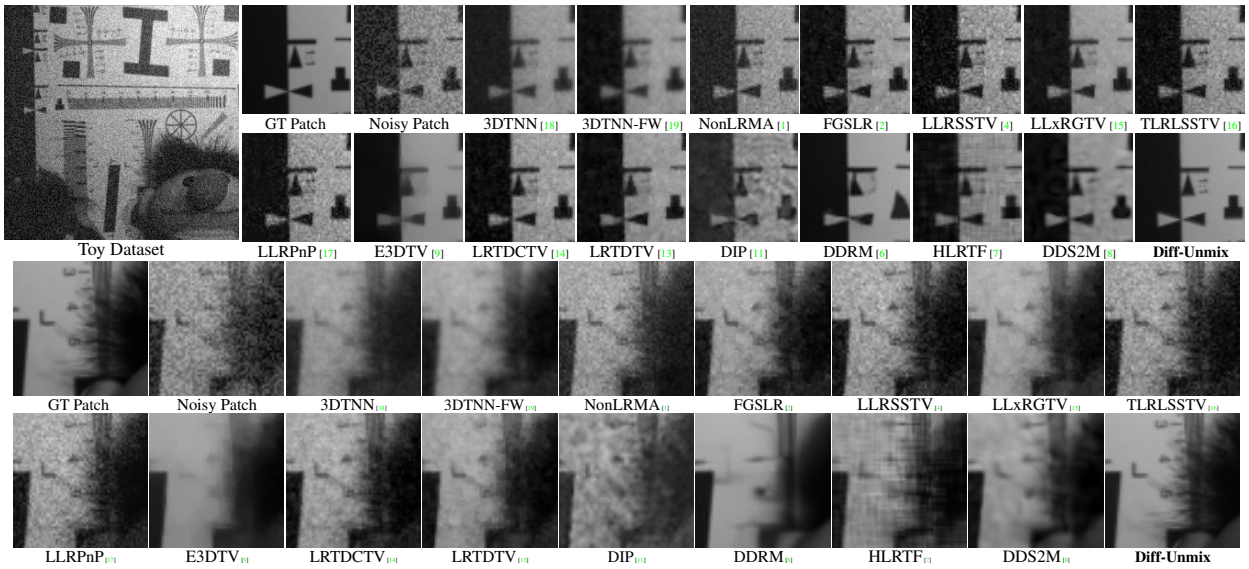

Figure 4. Visual comparison of HSI denoising methods on *Toy* dataset.

critical consideration involves ensuring consistent endmember information across varying noise levels. For comparison, we utilize Singular Value Decomposition (SVD) and DNN techniques for unmixing, as illustrated in Figure 3. The reconstructed HSI resulting from the unmixing reconstruction, achieved by blending the tensor SVD decomposed abundance map with spectral endmembers, notably differs from the original reference HSI. Conversely, STU demonstrates improved HSI reconstruction through its unmixing process, effectively blending the abundance map and the spectral

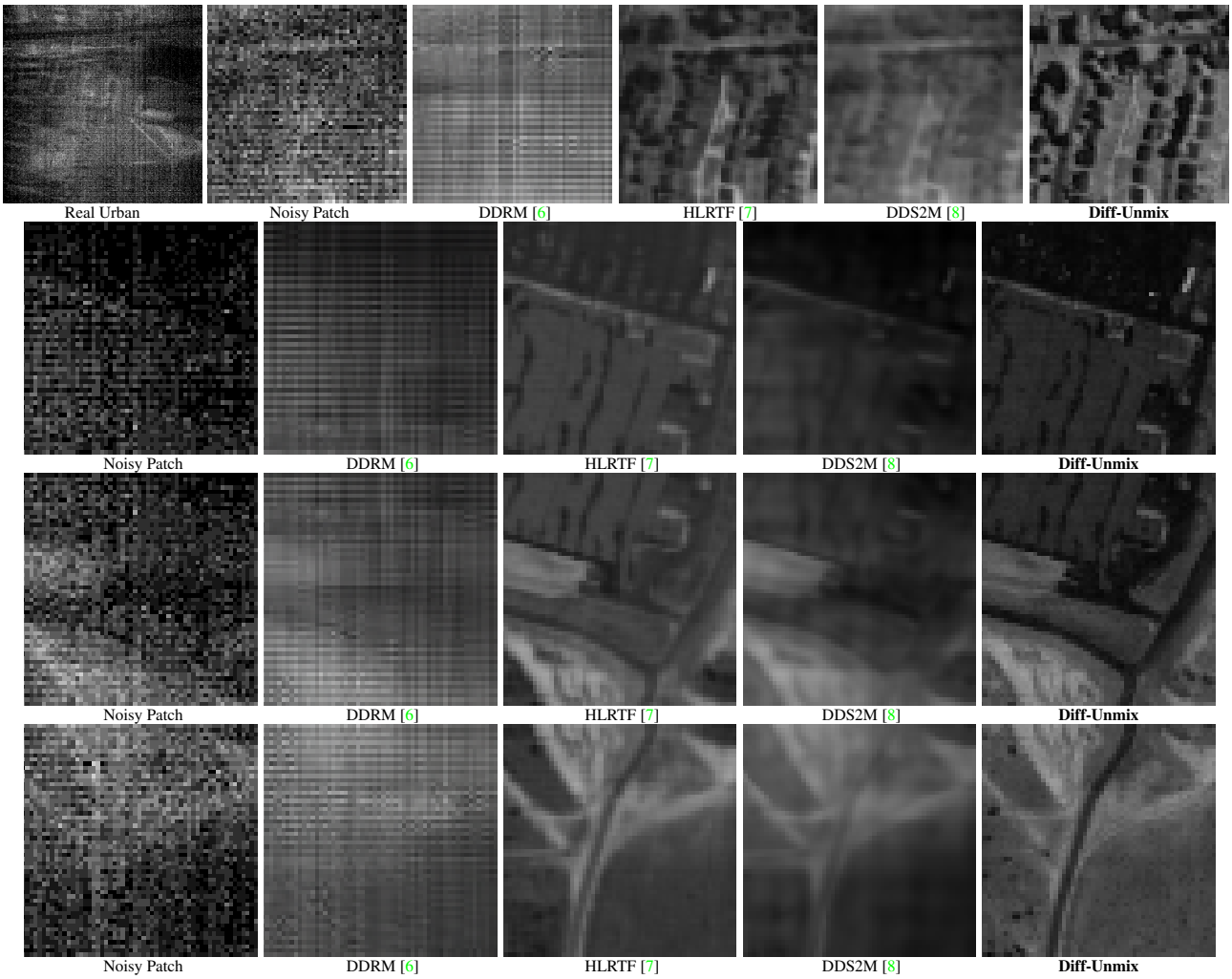

Figure 5. Results on Urban with real-world noise including stripes, deadlines, atmospheric interference, water absorption, and other unidentified sources. One can see that our Diff-Unmix adeptly mitigates mixed noise, ensuring the retention of fine-grained details.

endmembers it decomposed.

4. Additional Visual Comparison

Figures 1, 4, and 5 offer additional visual comparisons between the proposed Diff-Unmix and state-of-the-art (SOTA) methods. Evidently, Diff-Unmix excels in challenging scenarios, such as patches with severe noise or specific conditions like the lower-right context, adeptly reconstructing intricate image details and textures—areas where SOTA methods falter. However, akin to numerous applications of DDPM, Diff-Unmix does possess the propensity to generate extraneous details (over-enhancement) owing to its generative nature. For instance, observe the cube artifact within the Toy dataset depicted in the bottom of Figure 1.

References

[1] Yongyong Chen, Yanwen Guo, Yongli Wang, Dong Wang, Chong Peng, and Guoping He. Denoising of hyperspectral images using nonconvex low rank matrix approximation. *IEEE TGRS*, 55(9):5366–5380, 2017. 2, 3

[2] Yong Chen, Ting-Zhu Huang, Wei He, Xi-Le Zhao, Hongyan Zhang, and Jinshan Zeng. Hyperspectral image denoising using factor group sparsity-regularized nonconvex low-rank approximation. *IEEE TGRS*, 60:1–16, 2021. 2, 3

[3] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *ICLR*, 2023. 1

[4] Wei He, Hongyan Zhang, Huanfeng Shen, and Liangpei Zhang. Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation. *IEEE J-STARS*, 11(3):713–729, 2018. 2, 3

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020. 1

[6] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *NeurIPS*, 35:23593–23606, 2022. 2, 3, 4

[7] Yisi Luo, Xi-Le Zhao, Deyu Meng, and Tai-Xiang Jiang.

Hlrtf: Hierarchical low-rank tensor factorization for inverse problems in multi-dimensional imaging. In *CVPR*, pages 19303–19312, 2022. 2, 3, 4

[8] Yuchun Miao, Lefei Zhang, Liangpei Zhang, and Dacheng Tao. Dds2m: Self-supervised denoising diffusion spatio-spectral model for hyperspectral image restoration. In *ICCV*, pages 12086–12096, 2023. 2, 3, 4

[9] Jiangjun Peng, Qi Xie, Qian Zhao, Yao Wang, Leung Yee, and Deyu Meng. Enhanced 3dtv regularization and its applications on hsi denoising and compressed sensing. *IEEE TIP*, 29:7889–7903, 2020. 2, 3

[10] Xiangyu Rui, Xiangyong Cao, Zeyu Zhu, Zongsheng Yue, and Deyu Meng. Unsupervised pansharpening via low-rank diffusion model. *arXiv preprint arXiv:2305.10925*, 2023. 1, 2

[11] Oleksii Sidorov and Jon Yngve Hardeberg. Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution. In *ICCV*, pages 0–0, 2019. 2, 3

[12] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021. 1

[13] Yao Wang, Jiangjun Peng, Qian Zhao, Yee Leung, Xi-Le Zhao, and Deyu Meng. Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE J-STARS*, 11(4):1227–1243, 2017. 2, 3

[14] Haijin Zeng, Shaoguang Huang, Yongyong Chen, Hiep Luong, and Wilfried Philips. All of low-rank and sparse: A recast total variation approach to hyperspectral denoising. *IEEE J-STARS*, 2023. 2, 3

[15] Haijin Zeng and Xiaozhen Xie. Hyperspectral image denoising via global spatial-spectral total variation regularized nonconvex local low-rank tensor approximation. *Signal Processing*, 178:107805, 2021. 2, 3

[16] Haijin Zeng, Xiaozhen Xie, Haojie Cui, Hanping Yin, and Jifeng Ning. Hyperspectral image restoration via global l 1-2 spatial–spectral total variation regularized local low-rank tensor recovery. *IEEE TGRS*, 59(4):3309–3325, 2020. 2, 3

[17] Haijin Zeng, Xiaozhen Xie, Wenfeng Kong, Shuang Cui, and Jifeng Ning. Hyperspectral image denoising via combined non-local self-similarity and local low-rank regularization. *IEEE Access*, 8:50190–50208, 2020. 2, 3

[18] Yu-Bang Zheng, Ting-Zhu Huang, Xi-Le Zhao, Tai-Xiang Jiang, Teng-Yu Ji, and Tian-Hui Ma. Tensor n-tubal rank and its convex relaxation for low-rank tensor recovery. *Information Sciences*, 532:170–189, 2020. 2, 3

[19] Yu-Bang Zheng, Ting-Zhu Huang, Xi-Le Zhao, Tai-Xiang Jiang, Tian-Hui Ma, and Teng-Yu Ji. Mixed noise removal in hyperspectral image via low-fibered-rank regularization. *IEEE TGRS*, 58(1):734–749, 2019. 2, 3