DSTrans: Dual-Stream Transformer for Hyperspectral Image Restoration Supplementary Material

Dabing Yu, Qingwu Li, Xiaolin Wang, Zhiliang Zhang, Yixi Qian and Chang Xu
Hohai university
{yudadabing, zhangzl, 211620010037, xuchang}@hhu.edu.cn, {li_qingwu, xlwang1998}@163.com

1. Loss Function

We combine the L_1 loss and the spatial-spectral total variation (SSTV) loss [7] to optimize the parameters of DSTrans. L_1 loss computes the mean absolute error (MAE) between the restored images and the ground truth. It is beneficial to penalize pixel errors and ensure better convergence throughout the training process,

$$L_1(\Theta) = \frac{1}{N}\sum_{n=1}^{N}\|X^n - I^n\|, \tag{1}$$

where I^n and X^N are the n-th reconstructed result and ground truth. N denotes the number of images in one training batch, and Θ refers the parameters of DSTrans. SSTV loss is designed to smooth the reconstructed result in both spatial and spectral dimensions,

$$\begin{aligned} L_{SSTV}(\Theta) =& \frac{1}{N}\sum_{n=1}^{N}(\|\bigtriangledown_h I^n\|_1 + \|\bigtriangledown_w I^n\|_1 \\ &+ \|\bigtriangledown_c I^n\|_1), \end{aligned} \tag{2}$$

where $\bigtriangledown_h$, $\bigtriangledown_w$, and $\bigtriangledown_c$ present the horizontal, vertical, and spectral directions gradient of the reconstruction result, respectively. The final objective loss function for the our DSTrans is the sum of the L_1 loss and SSTV loss,

$$L(\Theta) = L_1 + L_{SSTV}. \tag{3}$$

The overall loss of our restoration task contains the loss of RGBI restoration task and HSI restoration task,

$$\begin{aligned} L_{Total}(\Theta) =& L^{HSI}(X^n_{HSI}, I^n_{HSI}) \\ &+ L^{RGBI}(X^n_{RGBI}, I^n_{RGBI}). \end{aligned} \tag{4}$$

2. HSI Super-Resolution Results

In this section, we evaluate our DSTrans both qualitatively and quantitatively on two benchmark datasets: CAVE dataset [20] and Harvard dataset [4]. These two benchmark datasets are shown in 2.1. The results of the proposed DSTrans are compared with state-of-the-art methods, *i.e.,* EDSR[13], RCAN[23], HAN[15], GDRNN[12], MCNet[10], ERCSR[11], SSPSR[7], HSISR[9]. In the HSI SR task, we focus on upscaling factors $\times 4$ and $\times 8$ for comparison.

2.1. Datasets

CAVE. This dataset contains 32 images with the spatial resolution of 512×512 and 31 bands collected by a tunable filter and a cooled CCD camera ranging from 400 nm to 700nm. We use 20 images for training and 12 images for testing in our experiment.
Harvard. This dataset is gathered by Nuance FX, CRI Inc., which includes 50 images totally with the cube size of $31 \times 1040 \times 1392$ from 420 nm to 720nm. We use 40 for training and 10 for testing.
DIV2K [1]. DIV2K is adopted for the auxiliary RGBI SR task. There are 1000 high-quality images with 2K resolution. We use 800 samples for the training set and remain 200 for the test set.

2.2. Experimental Results on Harvard Dataset

For qualitative analysis, the 21st, 13th, and 5th bands are selected as R-G-B channels to exhibit the visual results in Fig. 1. Since the PSNR values of the results are all above 38dB, it is hard to distinguish the difference between visual results.

3. HSI Denoising Results

In this section, we conduct HSI denoise experiments on both synthetic Gaussian and real-world noise HSI. For Gaussian denoising, the different level of Gaussian noise is added to ICVL dataset [3]. Our DSTrans learns a separate model for each noise level. Meanwhile, the auxiliary RGBI gaussian denoising task is performed on DIV2K by adding Gaussian noise. For real HSI denoising, we evaluate our DSTrans on a real HSI noise dataset HSIDwRD [22]. The

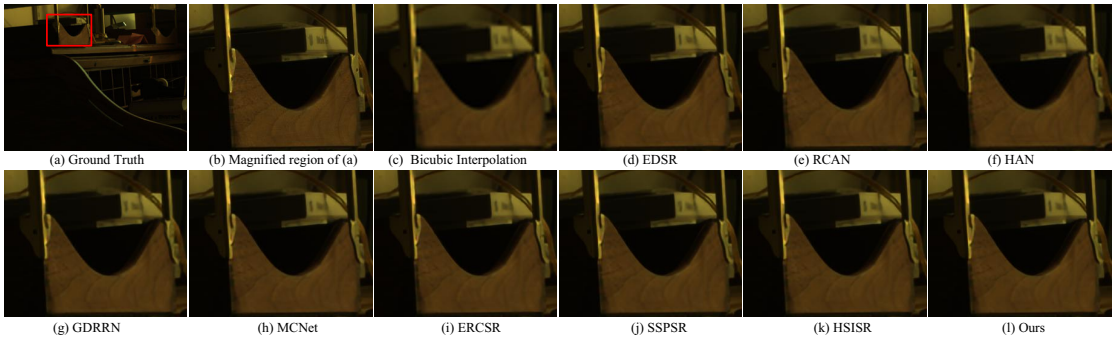


Figure 1. Visual comparison for HSI SR on one representative test image *imgd2_nl* from Harvard dataset with spectral bands 21-13-5 as R-G-B with the scale factor 4.

auxiliary RGBI denoise task is performed on RENOIR [2]. These benchmark datasets are shown in 3.1.

3.1. Datasets

ICVL. ICVL dataset contains 201 images with the spatial resolution of 1392×1300 over 31 spectral bands. We randomly select 100 images as training set and 50 images for testing.

HSIDwRD. HSIDwRD dataset is the first real-world dataset for training and testing HSI denoising model. There are 62 real-world HSIs collected by the SOC710-VP hyperspectral camera, of which each is with a size of $696 \times 520 \times 34$. The paired high-quality inference image and noise image are captured by adjusting aperture, focus, and exposure time. We select paired noisy and reference images of 45 scenes to form the training set and the rest 17 scenes are chosen for the testing set.

RENOIR [2]. RENOIR is a dataset for real noise images denoising tasks, which contains 40 scenes with the spatial resolution of 3684×2760 collected by the Canon S90, 40 scenes collected at 5202×3465 spatial resolution by the Canon Rebel T3i, and 40 scenes collected at 4208×3120 by the Xiaomi Mi3. RENOIR is adopted for the auxiliary RGBI real-world denoise task. We select image pairs of 30 scenes to form the testing set and the rest scenes are used for training.

3.2. Experimental Results on ICVL Dataset

To demonstrate the equally superior performance on HSI denoise, we compare our DSTrans with six state-of-the-art HSI denoising algorithms, including three traditional methods, BM4D [14], KBR [18], WLRTR [5], and NGmeet [6], and three recently developed deep learning methods, including HSID-CNN [21], QRNN3D [17] and DPPR [8].

In order to visually demonstrate the superiority of our method, we output the qualitative results and perform a local magnification comparison. Considering the test images have a substantial number of bands, we randomly selected one of them for comparison. The visual results are presented in Fig. 2. "Noisy" is obtained by adding the additive Gaussian white noise with noise levels of 50.

3.3. Experimental Results on HSIDwRD Dataset

We also provide more visual results on the HSIDwRD dataset. For visualization,we compare our DSTrans with five state-of-the-art HSI denoising algorithms, including BM4D [14], ITSReg [19], LRTDTV [16], QRNN3D [17] and DPPR [8]. More specifically, we sample the 15th band of the qualitative results and perform a local magnification comparison. The results of all methods on two different images are provided in Fig. 3 and Fig. 4. It is clear that our method eliminates the real-world noise and generates the clearest text signal than other methods.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1122–1131, 2017.

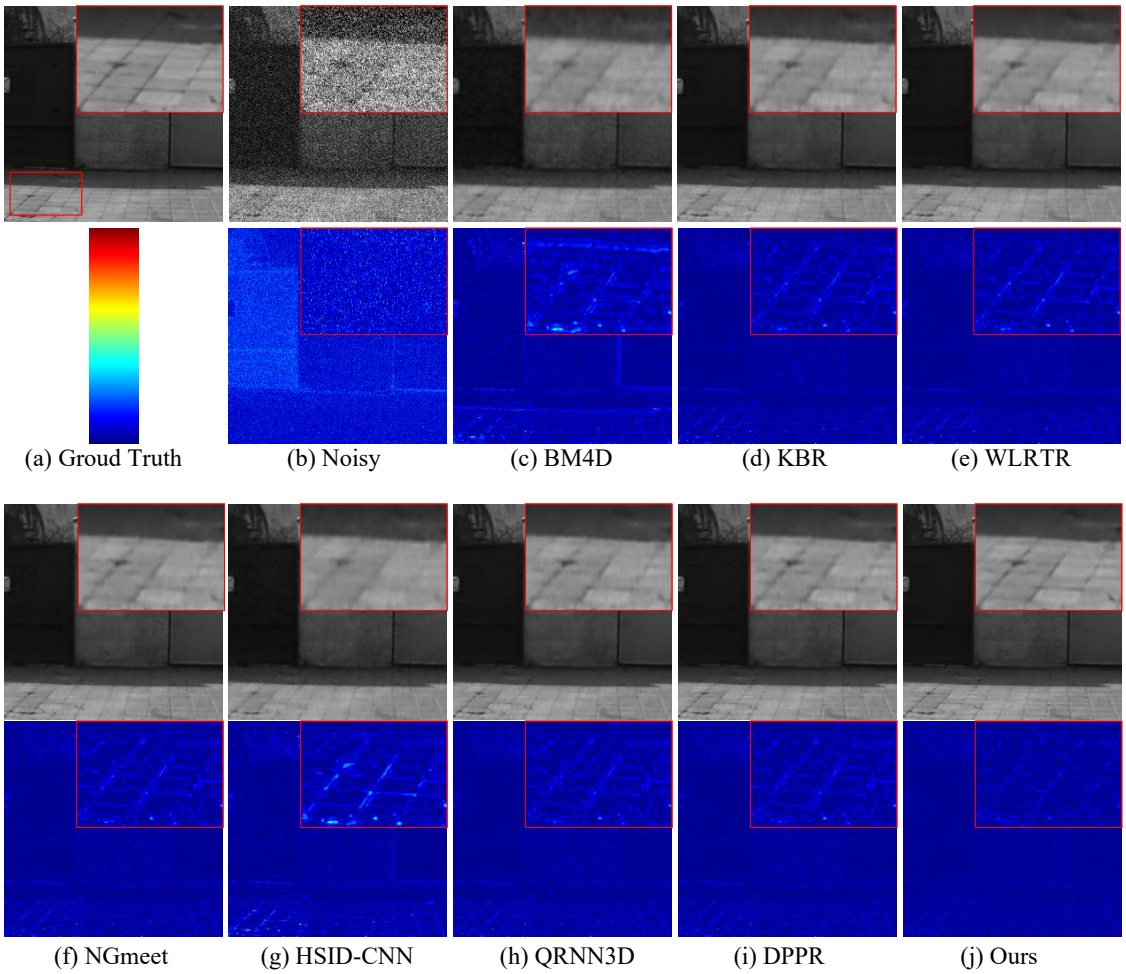
[2] Josue Anaya and Adrian Barbu. Renoir–a dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation*, 51:144–154, 2018.

[3] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016.

[4] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *CVPR 2011*, pages 193–200. IEEE, 2011.

[5] Yi Chang, Luxin Yan, Xi-Le Zhao, Houzhang Fang, Zhijun Zhang, and Sheng Zhong. Weighted low-rank tensor recovery for hyperspectral image restoration. *IEEE transactions on cybernetics*, 50(11):4558–4572, 2020.

[6] Wei He, Quanming Yao, Chao Li, Naoto Yokoya, Qibin Zhao, Hongyan Zhang, and Liangpei Zhang. Non-local meets global: An integrated paradigm for hyperspectral image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.



(a) Groud Truth (b) Noisy (c) BM4D (d) KBR (e) WLRTR

(f) NGmeet (g) HSID-CNN (h) QRNN3D (i) DPPR (j) Ours

Figure 2. Denoising results and error maps of the representative image *eve_0331-1549* under Gaussian noise with spectral bands 15.

[7] Junjun Jiang, He Sun, Xianming Liu, and Jiayi Ma. Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Transactions on Computational Imaging*, 6:1082–1096, 2020.

[8] Zeqiang Lai, Kaixuan Wei, and Ying Fu. Deep plug-and-play prior for hyperspectral image restoration. *Neurocomputing*, 2022.

[9] Ke Li, Dengxin Dai, Ender Konukoglu, and Luc Van Gool. Hyperspectral image super-resolution with spectral mixup and heterogeneous datasets. *arXiv preprint arXiv:2101.07589*, 2021.

[10] Qiang Li, Qi Wang, and Xuelong Li. Mixed 2d/3d convolutional network for hyperspectral image super-resolution. *Remote Sensing*, 12(10):1660, 2020.

[11] Qiang Li, Qi Wang, and Xuelong Li. Exploring the relationship between 2d/3d convolution for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 59(10):8693–8703, 2021.

[12] Yong Li, Lei Zhang, Chen Dingl, Wei Wei, and Yanning Zhang. Single hyperspectral image super-resolution with grouped deep recursive residual network. In *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, pages 1–4. IEEE, 2018.

[13] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[14] Matteo Maggioni, Vladimir Katkovnik, Karen Egiazarian, and Alessandro Foi. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE transactions on image processing*, 22(1):119–133, 2012.

[15] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *European conference on computer vision*, pages 191–207. Springer, 2020.

[16] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4):1227–1243, 2017.

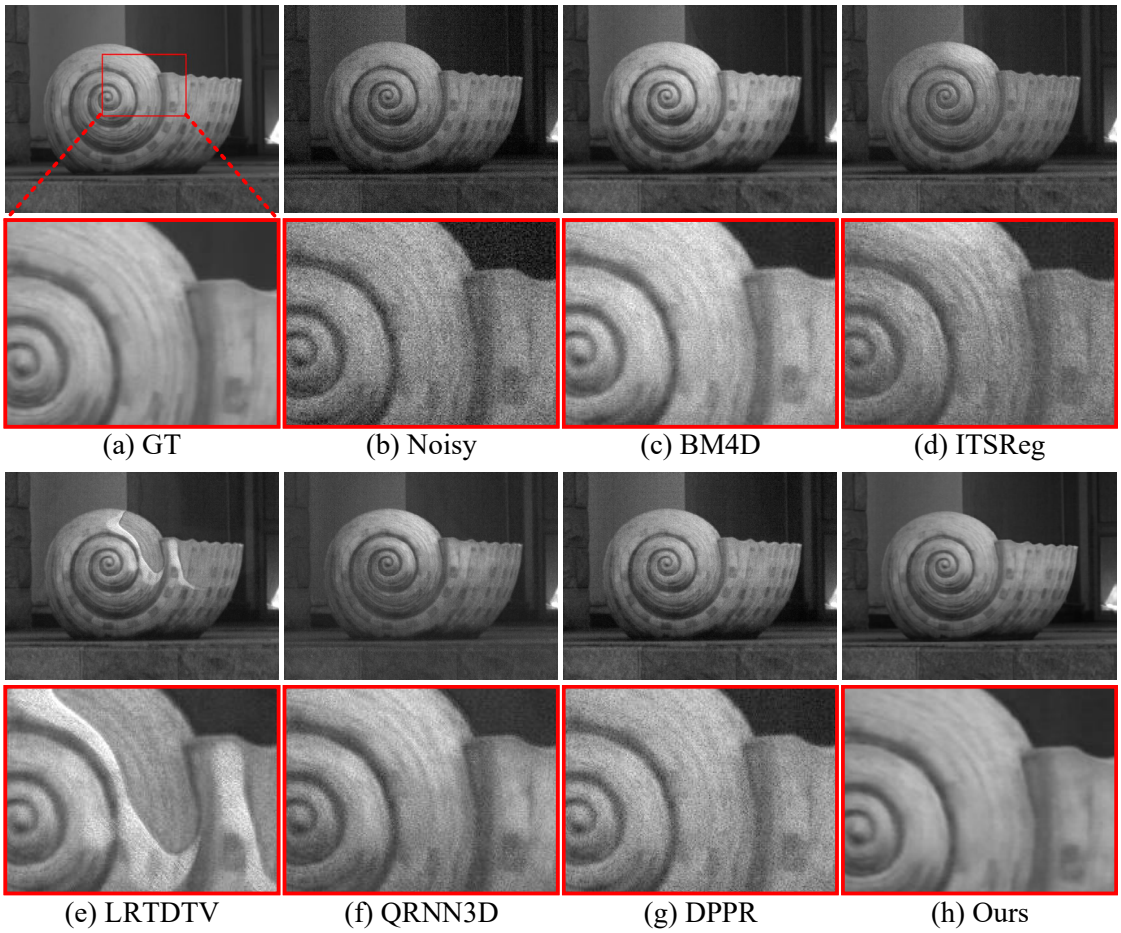


Figure 3. Denoising results of *image 50* under real-world noise with spectral bands 15.

[17] 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.

[18] Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu. Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8):1888–1902, 2018.

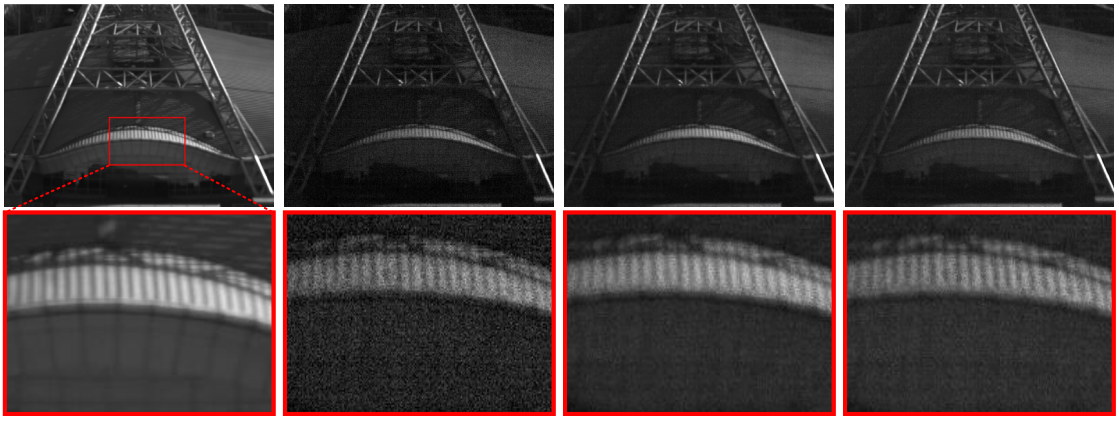
[19] Qi Xie, Qian Zhao, Deyu Meng, Zongben Xu, Shuhang Gu, Wangmeng Zuo, and Lei Zhang. Multispectral images denoising by intrinsic tensor sparsity regularization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1692–1700, 2016.

[20] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: post-capture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010.

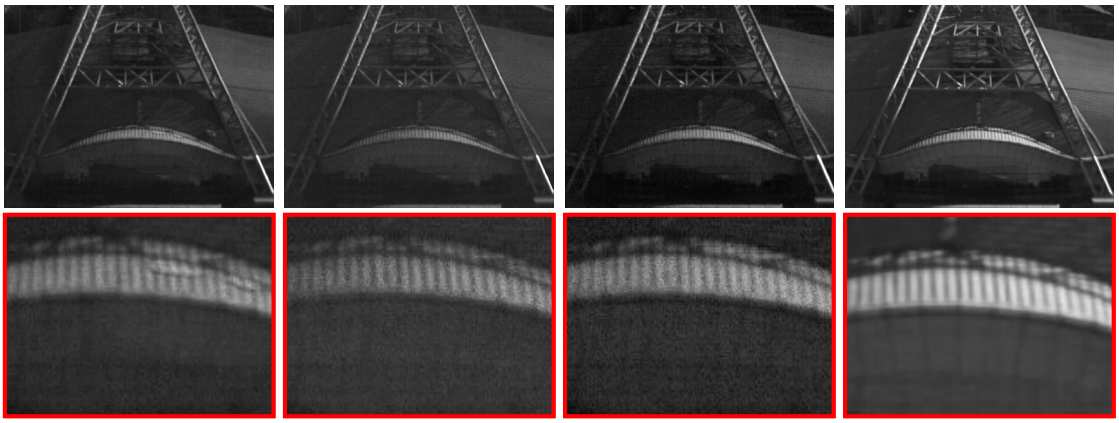
[21] Qiangqiang Yuan, Qiang Zhang, Jie Li, Huanfeng Shen, and Liangpei Zhang. Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):1205–1218, 2018.

[22] 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.

[23] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.



(a) GT (b) Noisy (c) BM4D (d) ITSReg



(e) LRTDTV (f) QRNN3D (g) DPPR (h) Ours

Figure 4. Denoising results of *image 47* under real-world noise with spectral bands 15.