Hyperspectral Pansharpening via Diffusion Models with Iteratively Zero-Shot Guidance

Supplementary Material

This supplementary material provides more discussions about running time, the downsampling factor between the low-resolution hyperspectral image (LR-HSI) and the high-resolution hyperspectral image (HR-HSI), and the kernel size of the Gaussian blurring operation.

1. Running Time

In the main paper, we demonstrated that diffusion models (DMs) guided by zero-shot guidance achieve superior performance with fewer sampling steps. In this section, we further evaluate the running time of each component of the proposed method, which comprises two key parts: DMs with zero-shot guidance and neural spatial-spectral decomposition (NSSD). Additionally, we compare the proposed method with PLRDiff [30], a state-of-the-art pre-trained DMs-based approach for hyperspectral pansharpening.

As shown in Figure 1, NSSD accounts for the majority of the total computation time. This is due to the fact that NSSD is not pre-trained and requires iterative updates, making it more time-consuming compared to the DMs with zero-shot guidance. Although the total running time of the proposed method exceeds that of PLRDiff [30], it is important to note that PLRDiff achieves its speed under a low-value constraint of rank for spatial-spectral decomposition, i.e., r in (13) is set to 3. The low value of rank constraint limits the representation capability of hyperspectral data. In contrast, the iterative update process of DMs with zero-shot guidance and NSSD remains stable across different rank configurations. Figure 2 compares the time consumption of PLRDiff [30] with the proposed method. It is evident that as the rank increases, the time consumption of PLRDiff rises significantly. In comparison, the proposed method maintains a consistent running time across data with varying rank constraints, demonstrating its robustness and scalability.

2. Downsampling Factor

In the experimental section of the main text, the downsampling factor between LR-HSI and HR-HSI is set to 4, a common setting in hyperspectral pansharpening tasks. In this part, we further conduct experiments with a downsampling factor of 2. As shown in Table 1, the proposed method achieves the best performance across multiple metrics. Moreover, the top two rows of Figure 3 illustrate the visual comparisons on the Pavia dataset under this downsampling setting. The proposed method preserves more details compared to other approaches.

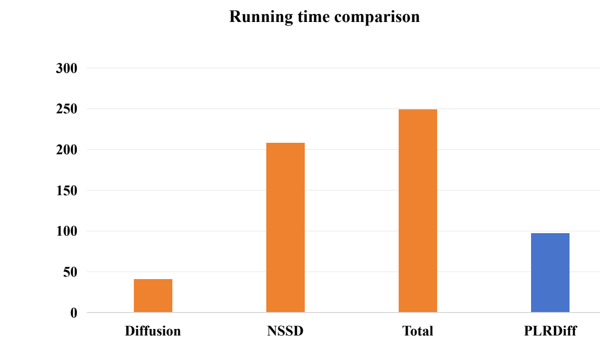

Figure 1. Running time (in seconds) of PLRDiff [30] and the proposed method.

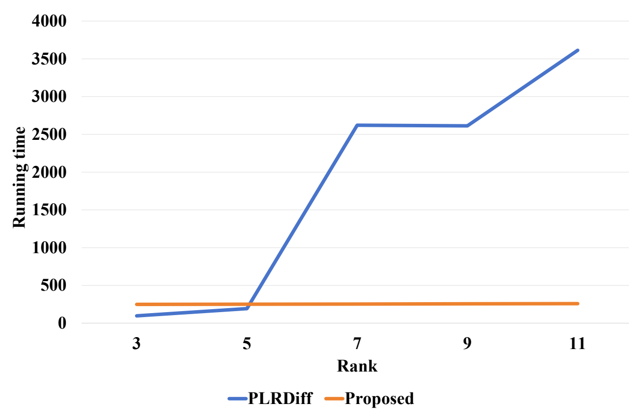

Figure 2. The time (in seconds) comparison of PLRDiff [30] and the proposed method with different rank sets.

Table 1. Quantitative results with downsampling factor (DF) as 2 on Pavia dataset. (Bold: best; Underline: second best)

DF	Method	PSNR↑	SSIM↑	ERGAS↓	SAM↓	RMSE↓
2	GSA [1]07'TGRS	31.721	0.887	7.731	5.162	7.014
	CNMF [47] 11'TGRS	31.967	0.909	7.486	4.433	6.621
	HySure [32] 14'TGRS	29.009	0.827	10.389	6.639	9.508
	ZSL [11]23'TPAMI	29.266	0.857	10.317	7.675	9.547
	PLRDiff [30] 24'IF	<u>32.978</u>	<u>0.926</u>	6.489	4.845	<u>6.314</u>
	HIR-Diff [29]24'CVPR	32.967	0.925	<u>6.381</u>	<u>4.139</u>	**6.245**
	Proposed	**33.125**	**0.962**	**6.369**	**3.815**	6.544
	Ideal value	+∞	**1**	**0**	**0**	**0**

3. Kernel Size

In this section, we analyze the performance of different methods in the benchmark under varying blurring kernel

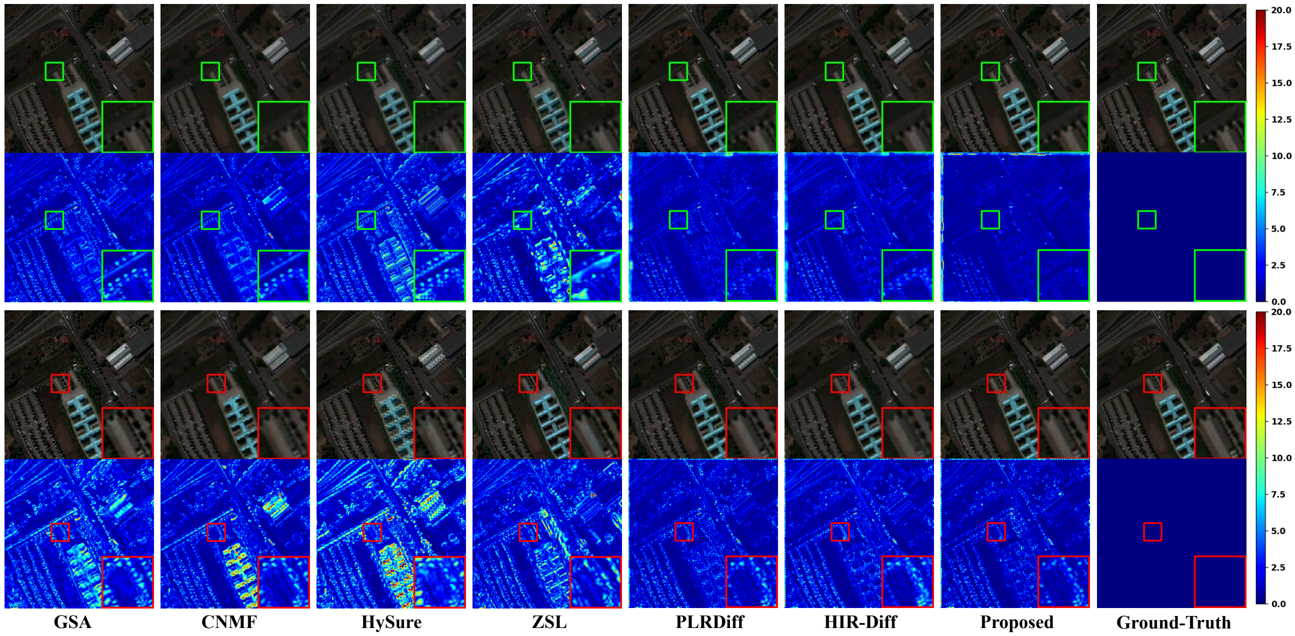

Figure 3. Visual comparisons on Pavia dataset. The upper two rows display the predicted HR-HSIs of different methods when the downsampling factor is 2. The bottom two rows show the predicted HR-HSIs of different methods when kernel size is 3×3.

sizes, i.e., 3×3 and 7×7. Table 2 presents the quantitative results for these kernel sizes. It is evident that the results vary with different kernel sizes. Notably, the proposed method consistently achieves the best hyperspectral pansharpening results across all conditions. The bottom rows of Figure 3 illustrate the visual comparisons, further demonstrating the superior performance of the proposed approach.

Table 2. Quantitative results for different kernel size (KS) on Pavia dataset. (Bold: best; Underline: second best)

KS	Method	PSNR↑	SSIM↑	ERGAS↓	SAM↓	RMSE↓
3	GSA [1]07'TGRS	27.899	0.801	6.030	8.179	10.769
	CNMF [47] 11'TGRS	27.462	0.838	6.729	6.099	11.236
	HySure [32] 14'TGRS	25.375	0.737	8.047	10.200	14.691
	ZSL [11]23'TPAMI	28.422	0.823	5.708	8.750	10.681
	PLRDiff [30] 24'IF	32.150	0.901	3.621	5.527	6.718
	HIR-Diff [29]24'CVPR	31.318	0.873	3.892	4.882	7.333
	Proposed	**32.655**	**0.914**	**3.390**	**4.749**	**6.387**
KS	**Method**	**PSNR↑**	**SSIM↑**	**ERGAS↓**	**SAM↓**	**RMSE↓**
7	GSA [1]07'TGRS	28.644	0.818	5.546	7.301	9.998
	CNMF [47] 11'TGRS	27.557	0.823	6.433	6.164	11.081
	HySure [32] 14'TGRS	26.323	0.762	7.339	9.675	13.330
	ZSL [11]23'TPAMI	28.707	0.826	5.426	9.046	10.447
	PLRDiff [30] 24'IF	31.816	0.901	3.710	5.439	7.177
	HIR-Diff [29]24'CVPR	30.790	0.867	4.098	4.884	7.966
	Proposed	**31.922**	**0.925**	**3.624**	**4.435**	**7.146**
	Ideal value	**+∞**	**1**	**0**	**0**	**0**