Cross-Scale Pansharpening via ScaleFormer and the PanScale Benchmark

Supplementary Material

8. Related Work

8.1. Pansharpening

Pan-sharpening methods can generally be categorized into two major paradigms: traditional techniques and deep learning-based approaches. Traditional methods are further divided into three main categories: component substitution (CS) [22, 28], multi-resolution analysis (MRA) [30, 55], and variational optimization (VO) [10, 23]. CS [12, 14] techniques enhance spatial resolution by substituting the spatial textures of the LRMS image with high-frequency details extracted from the PAN image. MRA [36, 40] approaches perform image information through multi-scale fusion, and VO [11, 33] methods generate the HRMS by optimizing the loss function. However, traditional methods typically rely on hand-crafted features and lack robust priors, often resulting in suboptimal results. Driven by the success of deep neural networks, deep learning-based methods have subsequently emerged as the predominant approach in the pan-sharpening community [27, 52, 58]. The pioneering work of PNN [31] initiated the application of convolutional neural networks to pan-sharpening. Following this, research efforts have increasingly focused on devising more sophisticated network architectures (e.g., MSDCNN [50], GPPNN [48]), incorporating auxiliary information (e.g., frequency analysis in SFINet[57] and MSDDN[16]), and exploring innovative learning paradigms [9, 45, 49]. Concurrently, pan-sharpening has witnessed a significant expansion in its richness and diversity. HFIN [38] introduces a hierarchical frequency integration network. Mamba-based methods [4, 34, 46] employ the state-space model to achieve effective pan-sharpening. Furthermore, ARConv [44] presents an adaptive rectangular convolution, enabling the network to learn convolution kernel dimensions tailored to objects of varying scales within remote sensing imagery.

8.2. Scale Generalization in Perceptual Tasks

Scale generalization presents a critical challenge for visual models in perceptual tasks, particularly when dealing with inputs of varying sizes and resolutions. The Flexivit [2] was one of the first attempts to address this issue, enabling the model to adapt to inputs with different patch sizes during training, ensuring stable performance across changing scales. Similarly, Resformer [39] enhances Vision Transformers' generalization ability in tasks with significant scale variations by employing a multi-resolution training strategy with a scale consistency loss function. Building on this, ViTAR [8] further improves scale generalization by incorporating dynamic resolution adjustments. Additionally, it introduces fuzzy positional encoding to prevent overfitting to a single resolution. The MSPE [29] method optimizes the patch embedding process with multiple variable-sized patch kernels, pushing the performance of Vision Transformers to new heights, especially in tasks involving multi-scale images, where it demonstrates a substantial improvement in model performance.

9. PanScale Dataset and Benchmark

9.1. Dataset Preprocessing

As is shown in Fig. 7, we first selected 16 geographically diverse regions across the globe, covering a broad spectrum of terrains such as urban, mountainous, oceanic, and desert. This diversity ensures that the resulting dataset effectively captures the challenges of pansharpening under varying environmental conditions, thereby facilitating robust evaluation of model generalization across different geographic contexts. Based on this selection, we sourced imagery from three widely used satellites: Jilin-1, Landsat-9, and Skysat. Taking Landsat imagery as an example, each raw satellite image can be represented as $\mathbf{I} \in \mathbb{R}^{H \times W \times B}$, where H,W, and B denote the height, width, and number of spectral bands. To remove cloud-contaminated pixels, we utilize the quality assessment band Q, which provides metadata indicating the presence of clouds and cirrus. Using this metadata, a binary cloud mask $Mask$ is generated, and the resulting cloud-free image is defined as:

$$\mathbf{I}' = \mathbf{I} \odot Mask \tag{12}$$

Building upon the cloud-free images, we collect all available Landsat satellite observations covering a selected geographic region within a defined time interval $[t_0, t_1]$, resulting in an image set:

$$\mathcal{I} = \{\mathbf{I'}_1, \mathbf{I'}_2, ..., \mathbf{I'}_N\} \tag{13}$$

For each spectral band $b \in [1, B]$, we perform a pixel-wise median compositing across all cloud-free images in the set $\mathcal{I}$ to generate a temporally stable and noise-reduced composite band:

$$\mathbf{I}_b^{mid}(i,j) = median(\{\mathbf{I}_{1,b}(i,j), \mathbf{I}_{2,b}(i,j), ..., \mathbf{I}_{N,b}(i,j)\}) \tag{14}$$

where (i, j) denotes the spatial pixel location with $i \in [1, H]$ and $j \in [1, W]$. By stacking the median composites of all spectral bands, we obtain the final multi-band synthesized image, denoted as $\mathbf{I}^{\text{med}}$. Leveraging the band extraction and export capabilities of Google Earth Engine, we further derive the corresponding panchromatic (PAN) images and

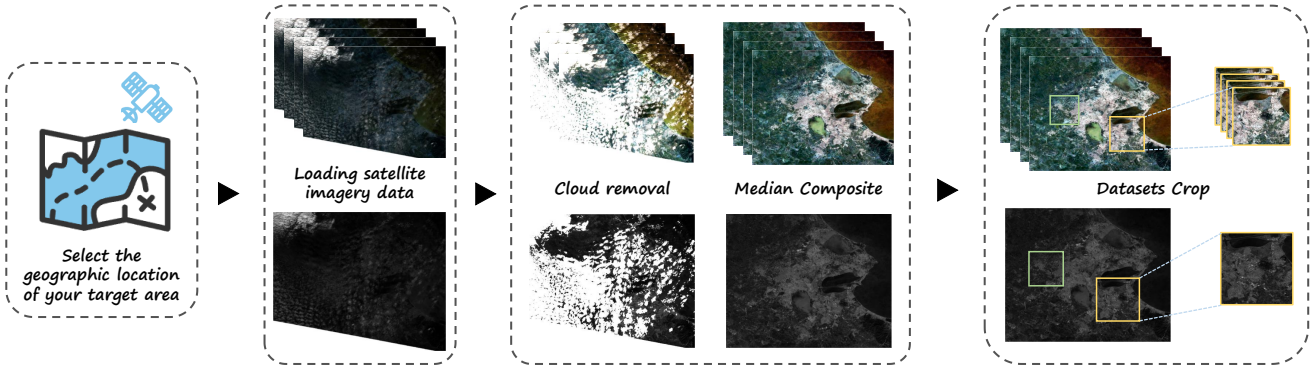

Figure 7. The construction pipeline of the PanScale dataset, which consists of three main stages: accessing satellite imagery via Google Earth Engine, data preprocessing, and data cropping.

low-resolution multi-spectral (LRMS) images for subsequent use in pansharpening.

9.2. Dataset Construction

Table 6 presents the detailed configuration of the dataset construction. The training sets for the three constituent sources—Jilin-1, Landsat-9, and Skysat—are designed with distinct upsampling ratios of 4×, 2×, and 2.5×, respectively. This variation in scaling allows the models to be exposed to diverse resolution levels during training.

For a comprehensive evaluation, we establish a multi-resolution testing framework to rigorously assess the cross-scale generalization ability of pansharpening models. Specifically, the reduced-resolution (RR) testing sets consist of images at resolutions of 1600×1600, 800×800, 400×400, and 200×200 pixels. Notably, for the Jilin dataset, the maximum available resolution is 800×800, and therefore, the 1600×1600 variant is not included. In addition to the RR evaluations, we provide full-resolution (FR) test sets at varying scales to further evaluate model robustness across real-world spatial variations.

9.3. Benchmark Design

To comprehensively evaluate diverse methods on the PanScale benchmark, we employ a suite of image quality assessment (IQA) metrics. Specifically, we select six prevalent reference-based metrics: Peak Signal-to-Noise Ratio (PSNR) [18], Structural Similarity Index (SSIM) [47], Spectral Angle Mapper (SAM) [51], Relative Dimensionless Global Error in Synthesis (ERGAS) [43], Spatial Correlation Coefficient (SCC) [19] and Quality index (Q) [42]. PSNR and SSIM primarily quantify spatial fidelity, with higher values indicating superior preservation of spatial details. Conversely, SAM and ERGAS focus on assessing spectral accuracy, where lower values denote diminished spectral distortion and enhanced consistency with the reference data. SCC assesses the spatial correlation between images, while Q provides a comprehensive measure of universal image quality for pansharpened MS images. Furthermore, to provide additional evaluation of model performance on PanScale, we also leverage three reference-free IQA metrics that do not rely on ground truth data. These include the spectral distortion index D_λ [20], the spatial distortion index D_S, and the Quality with No Reference (QNR) [1]. Specifically, D_λ and D_S quantify the extent of spectral and spatial distortion, respectively. QNR provides a holistic assessment of image quality without reference.

10. Additional experimental results

10.1. More Visual Comparison

Due to space constraints, we provide only the visualization results for the Jilin dataset in the main text. More visual comparison results for the three datasets are shown in Fig. 8, 9, and 10. The zoomed-in comparison images and corresponding MSE maps demonstrate that the proposed method excels in preserving fine details and spectral fidelity. These results highlight its strong adaptability to large-scale pansharpening tasks, showcasing its ability to handle high-resolution image fusion effectively.

10.2. Comprehensive Quantitative Comparison

Due to space limitations, we report only the average multi-scale performance on the reduced resolution datasets in Table 2 of the main text. For completeness and reproducibility, the full quantitative results across all resolutions and datasets are provided in the following tables: Table 7 (Jilin dataset at 200 and 400 resolutions), Table 8 (Jilin dataset at 800 resolution), Table 9 (Landsat dataset at 200 and 400 resolutions), Table 10 (Landsat dataset at 800 and 1600 resolutions), Table 11 (Skysat dataset at 200 and 400 resolutions), and Table 12 (Skysat dataset at 800 and 1600 resolutions). These results offer a more detailed perspective on the performance of each method across different scales. Notably,

Table 6. Detailed Information of the PanScale Datasets, where "samples" denotes the number of samples at each scale, "scale" represents the height or width of the square PAN image, and "res" indicates the image resolution used for dataset creation.

	Jilin (4×)			Landsat (2×)			Skysat (2.5×)		
	Samples	**Scale**	**Res(m)**	**Samples**	**Scale**	**Res(m)**	**Samples**	**Scale**	**Res(m)**
Train	1055	200	2	2484	256	30	2370	200	2
RR Test	34	800,400 200	2	96	1600,800 400,200	30	88	1600,800 400,200	2
FR Test	121	2000,1600 1200,800	2	48	2000,1600 1200,800	30	34	2000,1600 1200,800	2

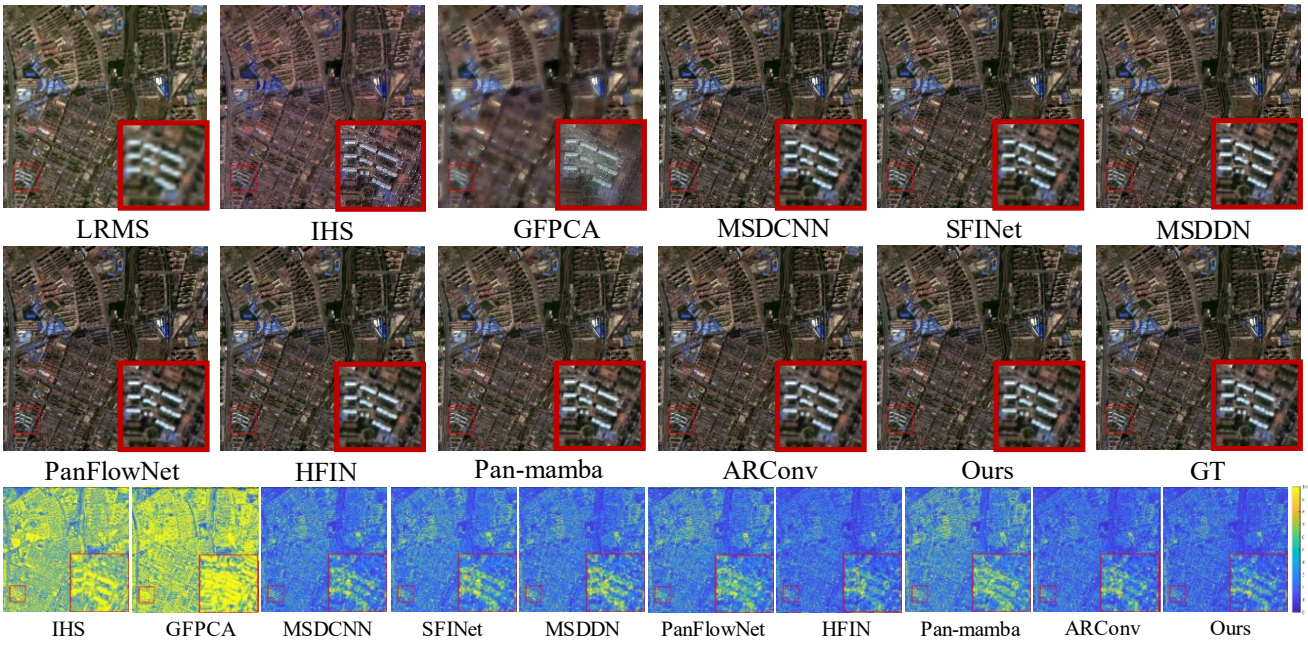

Figure 8. Comparative visual experiments of several methods on Jilin datasets.

our proposed method consistently outperforms existing approaches across nearly all metrics, demonstrating its superior cross-scale generalization capability.

Similarly, for the full-resolution experiments summarized in Table 3 of the main text, we report only the averaged results across multiple scales. To ensure completeness and facilitate further analysis, the detailed performance metrics for each resolution are presented in the following supplementary tables: Table 13 (Full Jilin dataset at 800, 1200, 1600, and 2000 resolutions), Table 14 (Full Landsat dataset at 800, 1200, 1600, and 2000 resolutions), and Table 15 (Full Skysat dataset at 800, 1200, 1600, and 2000 resolutions). These comprehensive results show that our method achieves performance on par with or superior to state-of-the-art techniques even under full-resolution settings, demonstrating strong generalization capability across scales in real-world scenarios.

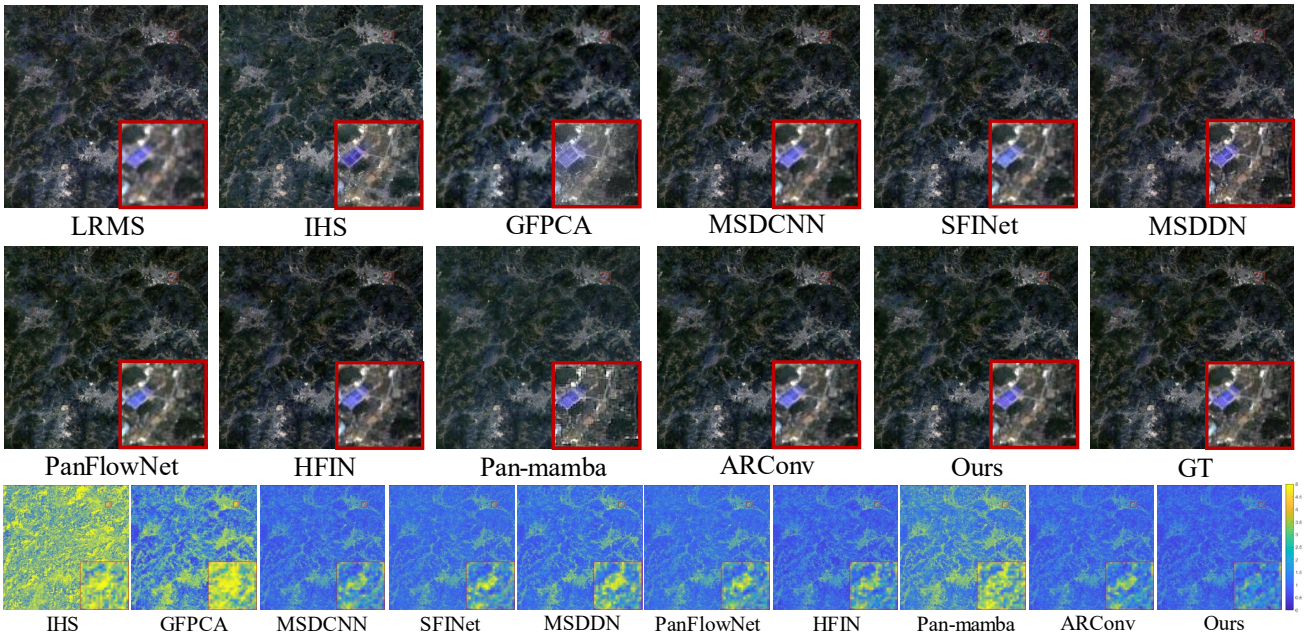

Figure 9. Comparative visual experiments of several methods on Landsat datasets.

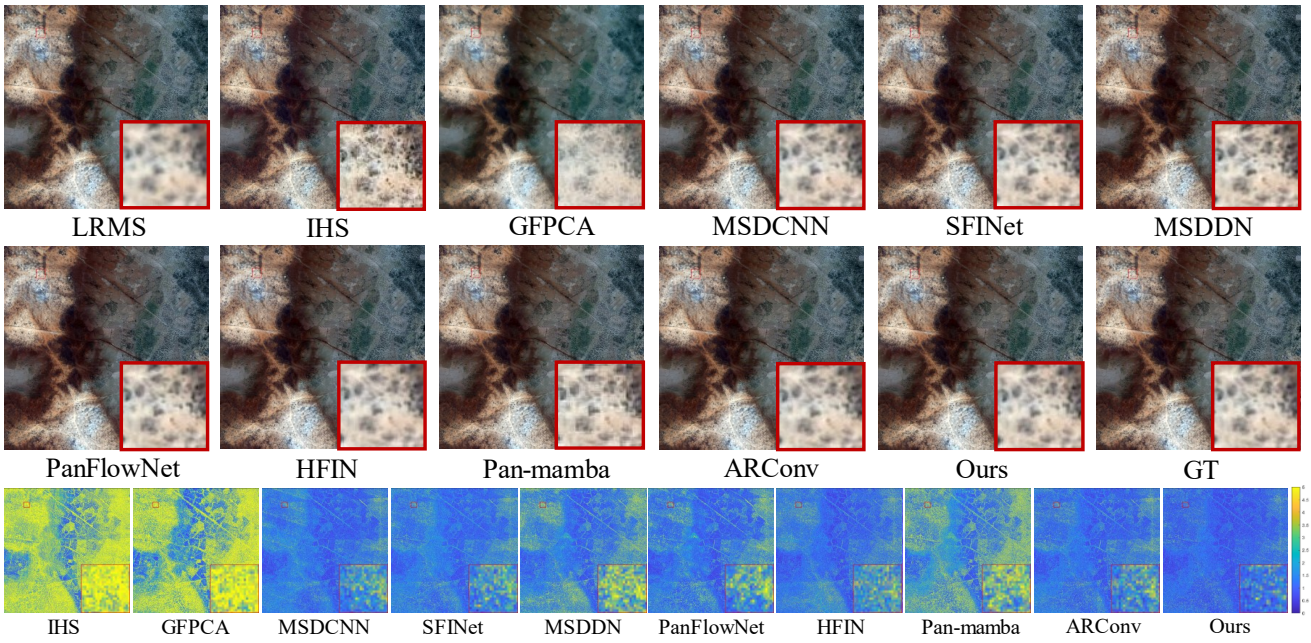

Figure 10. Comparative visual experiments of several methods on Skysat datasets.

Table 7. Multi-resolution results on the Jilin datasets (Part 1), with numbers indicating the size of the PAN image. The best are highlighted in **bold**.

Method	Jilin-200						Jilin-400					
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑
GS	25.7681	0.7128	0.0573	5.3592	0.8592	0.6002	25.6914	0.7099	0.0570	5.5151	0.8683	0.5805
IHS	25.9327	0.7081	0.0590	5.6410	0.8488	0.5930	25.8671	0.7063	0.0588	5.7120	0.8618	0.5754
GFPCA	25.6482	0.7106	0.0761	5.1018	0.8730	0.4579	25.8554	0.7252	0.0733	4.9484	0.8882	0.4469
MSDCNN	36.5984	0.9555	0.0277	1.5336	0.9885	0.9146	37.0869	0.9588	0.0265	1.4275	0.9907	0.9056
SFINet	36.2407	0.9508	0.0306	1.6494	0.9872	0.9017	36.6916	0.9532	0.0301	1.5592	0.9895	0.8915
MSDDN	37.5748	0.9653	0.0242	1.3600	0.9910	0.9289	37.7031	0.9667	0.0244	1.3366	0.9922	0.9174
PanFlowNet	36.6522	0.9571	0.0277	1.5303	0.9886	0.9163	36.8303	0.9583	0.0272	1.4832	0.9903	0.9038
HFIN	38.0366	0.9686	0.0228	1.3103	0.9919	0.9343	38.2695	0.9693	0.0224	1.2464	0.9932	0.9223
Pan-mamba	36.1465	0.9517	0.0278	1.6057	0.9873	0.9092	36.3710	0.9534	0.0280	1.5434	0.9891	0.8985
ARConv	38.1115	0.9682	0.0233	1.2904	0.9919	0.9338	38.5962	0.9701	0.0225	1.2113	0.9934	0.9246
Ours	**39.3300**	**0.9759**	**0.0202**	**1.1152**	**0.9941**	**0.9449**	**39.6334**	**0.9762**	**0.0197**	**1.0728**	**0.9950**	**0.9336**

Table 8. Multi-resolution results on the Jilin datasets (Part 2), with numbers indicating the size of the PAN image. The best are highlighted in **bold**.

Method	Jilin-800					
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑
GS	24.8563	0.6939	0.0605	5.9530	0.8700	0.5739
IHS	24.9852	0.6926	0.0611	6.0682	0.8659	0.5726
GFPCA	25.3367	0.7198	0.0738	5.0877	0.8974	0.4479
MSDCNN	36.0279	0.9578	0.0263	1.5284	0.9907	0.9074
SFINet	35.9211	0.9528	0.0295	1.6301	0.9901	0.8942
MSDDN	35.5442	0.9617	0.0266	1.6012	0.9903	0.9122
PanFlowNet	35.5154	0.9543	0.0278	1.6527	0.9897	0.9001
HFIN	37.6967	0.9714	0.0219	1.2696	0.9940	0.9285
Pan-mamba	34.1367	0.9388	0.0311	1.9399	0.9858	0.8822
ARConv	37.9942	0.9708	0.0221	1.2368	0.9940	0.9288
Ours	**38.9161**	**0.9762**	**0.0196**	**1.1093**	**0.9953**	**0.9369**

Table 9. Multi-resolution results on the Landsat datasets (Part 1), with numbers indicating the size of the PAN image. The best are highlighted in **bold**.

Method	Landsat-200						Landsat-400					
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑
GS	36.6649	0.9049	0.0413	4.1725	0.8227	0.5872	36.1124	0.8994	0.0481	4.0663	0.8504	0.5853
IHS	37.3551	0.9001	0.0364	4.1847	0.7969	0.5566	36.7972	0.8952	0.0398	4.0051	0.8263	0.5547
GFPCA	37.1018	0.8968	0.0432	3.7622	0.8255	0.4707	36.9330	0.8975	0.0431	3.3999	0.8967	0.4744
MSDCNN	38.8386	0.9590	0.0407	2.8789	0.8242	0.6667	38.6542	0.9597	0.0396	2.7907	0.8442	0.6678
SFINet	38.7072	0.9523	0.0486	3.3100	0.8428	0.6666	39.1371	0.9569	0.0484	2.7937	0.8649	0.6683
MSDDN	37.8541	0.9360	0.0609	3.2539	0.8212	0.6442	38.3405	0.9391	0.0524	2.7208	0.8213	0.6433
PanFlowNet	37.9274	0.9343	0.0721	3.8958	0.8263	0.6571	38.3768	0.9423	0.0627	3.3714	0.8470	0.6592
HFIN	39.9710	0.9635	0.0345	2.8424	0.8768	0.7165	40.6129	0.9698	0.0337	2.2585	0.9031	0.7208
Pan-mamba	39.505	0.9553	0.0530	3.0258	0.8701	0.7093	38.7678	0.9471	0.0487	2.6194	0.8512	0.6703
ARConv	39.6267	0.9624	0.0365	2.8184	0.8411	0.6770	39.8234	0.9645	0.0368	2.4587	0.8713	0.6812
Ours	**40.6063**	**0.9677**	**0.0341**	**2.6561**	**0.8869**	**0.7317**	**41.3715**	**0.9725**	**0.0333**	**2.1051**	**0.9081**	**0.7367**

Table 10. Multi-resolution results on the Landsat datasets (Part 2), with numbers indicating the size of the PAN image. The best are highlighted in **bold**.

Method	Landsat-800						Landsat-1600					
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑
GS	34.9034	0.8949	0.0517	4.6372	0.8873	0.5911	32.6318	0.8817	0.0633	5.5056	0.9064	0.5857
IHS	35.6661	0.8915	0.0407	4.2507	0.8640	0.5616	33.6923	0.8791	0.0461	4.8532	0.8755	0.5486
GFPCA	36.1215	0.8963	0.0413	3.5643	0.9074	0.4790	34.8823	0.8946	0.0418	3.6901	0.9320	0.4748
MSDCNN	38.7256	0.9600	0.0388	2.6008	0.8973	0.6809	39.0508	0.9582	0.0430	2.3577	0.9553	0.6792
SFINet	39.1135	0.9575	0.0461	2.5864	0.9167	0.6804	38.3526	0.9600	0.0389	2.5399	0.9314	0.6803
MSDDN	37.7455	0.9364	0.0506	2.7781	0.8704	0.6480	36.8196	0.9269	0.0515	2.8563	0.9072	0.6329
PanFlowNet	38.2522	0.9438	0.0573	3.1225	0.8994	0.6735	38.3067	0.9527	0.0475	2.6904	0.9453	0.6777
HFIN	40.2656	0.9664	0.0343	2.1820	0.9429	0.7249	39.9946	0.9665	0.0349	2.0090	0.9701	0.7243
Pan-mamba	35.7809	0.9131	0.0579	3.2307	0.8508	0.6291	32.8549	0.8668	0.0744	4.1172	0.8376	0.5661
ARConv	39.7381	0.9642	0.0342	2.3293	0.9185	0.6943	39.4700	0.9642	0.0339	2.1698	0.9511	0.6939
Ours	**41.1326**	**0.9719**	**0.0327**	**1.9870**	**0.9471**	**0.7438**	**41.0335**	**0.9722**	**0.0317**	**1.7934**	**0.9735**	**0.7458**

Table 11. Multi-resolution results on the Skysat datasets (Part 1), with numbers indicating the size of the PAN image. The best are highlighted in **bold**.

Method	Skysat-200						Skysat-400					
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑
GS	39.4693	0.9354	0.0326	2.2604	0.9121	0.6343	39.2596	0.9346	0.0310	2.3096	0.9357	0.6478
IHS	40.1768	0.9354	**0.0264**	2.0909	0.8934	0.6259	39.7388	0.9326	**0.0260**	2.1468	0.9206	0.6362
GFPCA	38.0562	0.9289	0.0454	2.4658	0.8896	0.4720	37.8788	0.9275	0.0441	2.4482	0.9213	0.4867
MSDCNN	43.0041	0.9758	0.0406	2.8975	0.8886	0.7373	43.1001	0.9788	0.0399	2.2283	0.9259	0.7450
SFINet	41.6073	0.9417	0.1010	3.4414	0.8496	0.6763	41.3735	0.9418	0.0939	3.1682	0.8911	0.6829
MSDDN	42.3969	0.9462	0.0825	2.9244	0.8947	0.7115	41.7617	0.9489	0.0697	2.7420	0.9243	0.7096
PanFlowNet	41.0920	0.9484	0.0757	3.2361	0.8467	0.6674	40.9916	0.9508	0.0667	2.9674	0.9009	0.6703
HFIN	43.7949	0.9533	0.0749	2.6743	0.9073	0.7637	43.9910	0.9669	0.0561	2.2725	0.9443	0.7758
Pan-mamba	43.2983	0.9516	0.0908	2.8129	0.8955	0.7546	42.7551	0.9590	0.0679	2.5895	0.9298	0.7461
ARConv	43.5032	0.9785	0.0423	2.1815	0.8934	0.7475	43.4617	0.9795	0.0406	2.0500	0.9344	0.7596
Ours	**44.9275**	**0.9826**	0.0362	**1.7945**	**0.9339**	**0.7837**	**44.6718**	**0.9824**	0.0364	**1.7266**	**0.9519**	**0.7874**

Table 12. Multi-resolution results on the Skysat datasets (Part 2), with numbers indicating the size of the PAN image. The best are highlighted in **bold**.

Method	Skysat-800						Skysat-1600					
	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑	PSNR↑	SSIM↑	SAM↓	ERGAS↓	SCC ↑	Q ↑
GS	38.9661	0.9302	0.0322	2.4344	0.9474	0.6367	38.5282	0.9264	0.0328	2.4648	0.9555	0.6294
IHS	39.3872	0.9280	**0.0275**	2.2406	0.9361	0.6247	39.0572	0.9255	**0.0269**	2.2241	0.9468	0.6217
GFPCA	37.4581	0.9282	0.0457	2.5427	0.9345	0.4751	37.2721	0.9272	0.0446	2.5326	0.9442	0.4755
MSDCNN	43.0467	0.9792	0.0391	2.1924	0.9347	0.7424	40.9681	0.9444	0.0851	3.0750	0.9211	0.6798
SFINet	41.3829	0.9452	0.0813	2.9911	0.9136	0.6787	42.6192	0.9786	0.0391	2.3726	0.9423	0.7419
MSDDN	40.7615	0.9475	0.0652	2.7640	0.9327	0.6900	39.2365	0.9397	0.0691	2.9878	0.9323	0.6684
PanFlowNet	41.1307	0.9558	0.0609	2.7860	0.9215	0.6733	40.8734	0.9570	0.0598	2.7015	0.9393	0.6801
HFIN	44.2075	0.9714	0.0508	2.0866	0.9581	0.7725	43.8435	0.9716	0.0516	2.0618	0.9680	0.7709
Pan-mamba	40.9513	0.9527	0.0626	2.6565	0.9320	0.7017	38.5358	0.9340	0.0689	3.0077	0.9245	0.6491
ARConv	43.4413	0.9803	0.0398	2.0106	0.9470	0.7594	43.2087	0.9806	0.0391	2.0094	0.9529	0.7591
Ours	**44.7328**	**0.9830**	0.0356	**1.6606**	**0.9613**	**0.7819**	**44.6416**	**0.9829**	0.0347	**1.7355**	**0.9686**	**0.7811**

Table 13. Multi-resolution results on the Full-Jilin datasets, with numbers indicating the size of the PAN image. The best are highlighted in **bold**, and the second are underlined.

Method	Full-Jilin-800			Full-Jilin-1200			Full-Jilin-1600			Full-Jilin-2000		
	$D_\lambda \downarrow$	$D_S \downarrow$	QNR ↑	$D_\lambda \downarrow$	$D_S \downarrow$	QNR ↑	$D_\lambda \downarrow$	$D_S \downarrow$	QNR ↑	$D_\lambda \downarrow$	$D_S \downarrow$	QNR ↑
GS	0.0795	0.1936	0.7433	0.0791	0.2012	0.7362	0.0814	0.2055	0.7304	0.0846	0.2081	0.7253
IHS	0.0955	0.2060	0.7191	0.0961	0.2121	0.7126	0.0972	0.2150	0.7092	0.0997	0.2165	0.7056
GFPCA	0.0464	0.1001	0.8584	0.0420	0.0980	0.8641	0.0405	0.0980	0.8654	0.0407	0.0979	0.8654
MSDCNN	0.0537	0.0499	0.8995	0.0460	0.0416	0.9144	0.0455	0.0394	0.9170	0.0454	0.0371	0.9194
SFINet	0.0596	0.0515	0.8925	0.0519	0.0430	0.9075	0.0517	0.0415	0.9092	0.0519	0.0395	0.9109
MSDDN	0.0481	0.0521	0.9046	0.0398	0.0509	0.9118	**0.0387**	0.0521	0.9116	0.0394	0.0533	0.9106
PanFlowNet	0.0502	**0.0438**	0.9044	0.0455	**0.0384**	0.9179	0.0441	**0.0364**	0.9201	0.0439	**0.0340**	**0.9196**
HFIN	0.0479	0.0522	0.9030	0.0414	0.0470	0.9138	0.0403	0.0453	0.9164	0.0398	0.0437	0.9182
Pan-mamba	0.0516	0.0501	0.9013	0.0446	0.0427	0.9148	0.0447	0.0402	0.9171	0.0450	0.0373	0.9195
ARConv	0.0468	0.0656	0.8907	0.0407	0.0629	0.8989	0.0398	0.0619	0.9007	0.0396	0.0611	0.9016
Ours	**0.0454**	0.0476	**0.9077**	**0.0394**	0.0428	**0.9181**	0.0394	0.0405	**0.9177**	**0.0393**	0.0363	0.9189

Table 14. Multi-resolution results on the Full-Landsat datasets, with numbers indicating the size of the PAN image. The best are highlighted in **bold**, and the second are underlined.

Method	Full-Landsat-800			Full-Landsat-1200			Full-Landsat-1600			Full-Landsat-2000		
	$D_\lambda \downarrow$	$D_S \downarrow$	QNR ↑	$D_\lambda \downarrow$	$D_S \downarrow$	QNR ↑	$D_\lambda \downarrow$	$D_S \downarrow$	QNR ↑	$D_\lambda \downarrow$	$D_S \downarrow$	QNR ↑
GS	0.0519	0.1145	0.8405	0.0589	0.1155	0.8334	0.0567	0.1076	0.8425	0.0598	0.1064	0.8406
IHS	0.0862	0.1235	0.8012	0.0972	0.1286	0.7883	0.0982	0.1205	0.7945	0.1028	0.1205	0.7905
GFPCA	0.0731	0.0577	0.8737	0.0623	0.0543	0.8870	0.0627	0.0525	0.8883	0.0615	0.0510	0.8910
MSDCNN	0.0523	0.0378	0.9135	0.0476	0.0316	0.9234	0.0485	0.0307	0.9234	0.0468	0.0303	0.9254
SFINet	0.0493	0.0405	0.9137	0.0410	0.0331	0.9278	0.0404	0.0314	0.9299	0.0392	0.0317	0.9307
MSDDN	0.0402	0.0246	0.9372	0.0311	0.0224	0.9477	0.0317	0.0229	0.9466	0.0312	0.0246	0.9454
PanFlowNet	0.0570	0.0388	0.9077	0.0449	0.0379	0.9194	0.0458	0.0385	0.9178	0.0449	0.0384	0.9187
HFIN	0.0287	**0.0200**	0.9523	**0.0228**	**0.0201**	0.9543	0.0229	**0.0188**	**0.9590**	0.0217	0.0198	0.9591
Pan-mamba	0.0295	0.0525	0.9196	0.0297	0.0675	0.9049	0.0312	0.0816	0.8898	0.0330	0.0928	0.8773
ARConv	0.0278	0.0231	0.9501	0.0260	0.0200	0.9548	0.0251	0.0191	0.9566	0.0239	**0.0183**	0.9585
Ours	**0.0272**	0.0271	**0.9537**	0.0240	0.0225	**0.9563**	**0.0217**	0.0209	0.9580	**0.0206**	0.0206	**0.9594**

Table 15. Multi-resolution results on the Full-Skysat datasets, with numbers indicating the size of the PAN image. The best are highlighted in **bold**, and the second are underlined.

Method	Full-Skysat-800			Full-Skysat-1200			Full-Skysat-1600			Full-Skysat-2000		
	$D_\lambda \downarrow$	$D_S \downarrow$	QNR $\uparrow$	$D_\lambda \downarrow$	$D_S \downarrow$	QNR $\uparrow$	$D_\lambda \downarrow$	$D_S \downarrow$	QNR $\uparrow$	$D_\lambda \downarrow$	$D_S \downarrow$	QNR $\uparrow$
GS	0.0533	0.1033	0.8499	0.0657	0.1163	0.8273	0.0653	0.1129	0.8304	0.0649	0.1123	0.8311
IHS	0.0937	0.1282	0.7916	0.0973	0.1343	0.7834	0.0993	0.1333	0.7822	0.0994	0.1314	0.7836
GFPCA	0.0427	0.0633	0.8974	0.0445	0.0633	0.8956	0.0428	0.0620	0.8983	0.0432	0.0610	0.8989
MSDCNN	0.0449	**0.0262**	0.9344	0.0407	0.0282	0.9422	0.0422	**0.0187**	0.9391	0.0398	**0.0155**	**0.9455**
SFINet	0.1357	0.0331	0.8373	0.1304	0.0343	0.8412	0.1363	0.0333	0.8362	0.1361	0.0327	0.8368
MSDDN	0.0424	0.0386	0.9217	0.0388	0.0382	0.9250	0.0426	0.0355	0.9294	0.0408	0.0327	0.9293
PanFlowNet	0.0883	0.0376	0.8787	0.0761	0.0339	0.8935	0.0738	0.0312	0.8981	0.0726	0.0301	0.9003
HFIN	0.0623	0.0634	0.8796	0.0583	0.0573	0.8888	0.0584	0.0588	0.8872	0.0572	0.0565	0.8904
Pan-mamba	0.0579	0.0383	0.9071	0.0546	0.0330	0.9152	0.0515	0.0284	0.9224	0.0491	0.0248	0.9279
ARConv	**0.0405**	0.0416	0.9201	**0.0381**	0.0402	0.9237	**0.0380**	0.0415	0.9225	0.0380	0.0404	0.9236
Ours	0.0451	0.0292	**0.9365**	0.0423	**0.0248**	**0.9425**	0.0407	0.0269	**0.9426**	**0.0339**	0.0268	0.9416