SSMP-Net: Spatial-Spectral Prior-Aware Unfolding Network for Pan-sharpening Supplementary Materials

Shijie Fang, Hongping Gan*,

School of Software, Northwestern Polytechnical University, China
fangshijie@mail.nwpu.edu.cn, ganhongping@nwpu.edu.cn

Overview

In this supplementary material, we first provide more hyperparameter settings for our proposed SSMP-Net in Section 1 to facilitate reproduction. We then provide more details and comparisons in Section 2, including qualitative comparisons with more models in Section 2.1 and more quantitative comparisons in Section 2.2. Finally, we provide an ablation analysis of number of convolutional hidden layers (mid), stage (N), and hyperparameters of the loss function (λ_{loss}) in Section 3.

1 More Model Reproducible Details

We use SSUN-Net with 5 reconstruction stages ($N = 5$) and 32 hidden layers ($mid = 32$) that obtained by ablation of Section 3 as the default model, it upsamples the LRMS using the bicubic interpolation to initialize $\mathbf{H}^{(0)}$. In addition, the auxiliary variables ($\mathbf{d_e}^{(0)}$, $\mathbf{d_a}^{(0)}$, $\mathbf{b_e}^{(0)}$, $\mathbf{b_a}^{(0)}$) are initialized to zero matrices of the same size as $\mathbf{H}^{(0)}$. Finally, we initialize the learnable parameters $\{\lambda_a^{(k)}, \lambda_e^{(k)}\}_{k=1}^{N}$ to 0.001 and $\{u^{(k)}, v^{(k)}, \alpha^{(k)}\}_{k=1}^{N}$ to 0.5.

All experiments are conducted on a Windows 11 system with an NVIDIA GeForce RTX 3090 GPU and a 4.10 GHz Intel i5-10600KF CPU. The SSUN-Net training employs a batch size of 4, and uses an ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, epsilon = 10^{-8}, decay = 0. The learning rate is adjusted using the cosine annealing strategy over 1000 training cycles, transitioning from 5×10^{-4} to 5×10^{-8} every 50 cycles, with the global random number seed set to 123. Furthermore, a comparison of learnable internal parameters in the SSUN-Net is presented before and after training in the Tab. 1.

Based on this experimental configuration, the default SSUN-Net inference time consumption for a multispectral image with a band number of 4 and a spatial size of 128×128 is approximately 30 ms.

The supplementary material includes the source code, and detailed experimental settings necessary to replicate the findings outlined in this paper. Furthermore, both the source code and pre-trained models will be made publicly available to enhance accessibility and facilitate reproducibility.

*Corresponding Author.

		u	v	λ_e	λ_a	α
INIT		0.50000	0.50000	0.00010	0.00010	0.50000
		0.50000	0.50000	0.00010	0.00010	0.50000
		0.50000	0.50000	0.00010	0.00010	0.50000
		0.50000	0.50000	0.00010	0.00010	0.50000
		0.50000	0.50000	0.00010	0.00010	0.50000
GF-2	1^{th} stage	0.35079	0.49849	0.03487	0.04178	0.27577
	2^{th} stage	1.07813	0.42333	0.02501	0.07627	0.22244
	3^{th} stage	0.52679	0.39109	0.03464	0.07852	0.20853
	4^{th} stage	0.68389	0.41166	0.00468	0.03032	0.21179
	5^{th} stage	0.98291	0.47788	0.02121	0.00211	0.23048
WV-II	1^{th} stage	0.40815	0.47499	0.07199	0.02251	0.42138
	2^{th} stage	0.60423	0.46966	0.00918	0.08669	0.42775
	3^{th} stage	0.76919	0.44415	0.20064	0.01772	0.44900
	4^{th} stage	0.79330	0.50096	0.09342	0.10913	0.47564
	5^{th} stage	0.56806	0.47466	0.14257	0.00726	0.42644
WV-III	1^{th} stage	0.25150	0.60125	0.07080	0.01571	0.60898
	2^{th} stage	0.65381	0.65611	0.25261	0.05625	0.48862
	3^{th} stage	0.84260	0.57591	0.15098	0.01661	0.53334
	4^{th} stage	0.48746	0.61369	0.04278	0.01519	0.51497
	5^{th} stage	0.37152	0.57717	0.31080	0.00121	0.38110

Table 1: Initialization of learnable internal parameters for SSUN-Net.

2 More Comparisons with SOTA Methods

2.1 More Quantitative Comparison In this subsection, we compare the proposed SSUN-Net with various representative Pan-sharpening methods including traditional approaches (GSA (Aiazzi, Baronti, and Selva 2007), SFIM (Liu 2000)), pure deep learning methods (PDDN (He et al. 2023a), PANF (Zhou, Liu, and Wang 2022), INNF (Zhou et al. 2022), MSDDN (He et al. 2023b),WINET (Zhang et al. 2024),RFCO (Qu et al. 2024),HFEAN (Wang et al. 2023), SFINet++ (Zhou et al. 2024), HFIN (Tan et al. 2024)), and deep unfolding methods (LGTEUN (Li et al. 2023a), MDCUN (Yang et al. 2022), GPPNN (Xu et al. 2021), NLUNET (Li et al. 2023b)). We set all comparison methods to their optimal configurations proposed in their paper. It is important to note that we have also included the comparative performance of representative different configurations of SSUN-Net. For instance, SSUN-Net (32-9) signifies a configuration with 32 hidden layers ($mid = 32$) and 9 iterations of expansion ($N = 9$) in SSUN-Net.

We highlight the results of the top three simulated data

Table 2: More quantitative results comparing SSUN-Net with other methods on simulated data. The symbol ↑ or ↓ is used to indicate that a higher or lower value corresponds to a better result.

		WorldView II					WorldView III					GaoFen 2					Flops (G)	Params (M)
		ERGAS↓	SSIM↑	PSNR↑	SCC↑	SAM↓	ERGAS↓	SSIM↑	PSNR↑	SCC↑	SAM↓	ERGAS↓	SSIM↑	PSNR↑	SCC↑	SAM↓		
GSA	TGRS'07	9.1864	0.5335	21.833	0.0619	0.2310	4.6064	0.9266	36.980	0.4552	0.0873	1.7981	0.8855	36.501	0.0128	0.0105	-	-
SFIM	IJRS'07	8.7638	0.5483	22.152	0.0670	0.2243	4.9651	0.9147	35.846	0.4534	0.0897	1.5923	0.8964	37.665	0.0170	0.0212	-	-
SFINet++	TPAMI'24	0.9538	0.9675	41.587	0.5147	0.0236	3.0217	0.9261	30.767	0.8333	0.0720	0.4376	0.9906	49.358	0.5897	0.0087	1.3112	0.0848
HFIN	CVPR'24	0.9906	0.9694	41.903	0.5697	0.0226	3.1523	0.9192	30.481	0.8156	0.0766	0.4907	0.9891	48.311	0.5583	0.0090	1.0104	0.0773
WINET	TGRS'24	0.9024	0.9714	42.048	0.5753	0.0226	3.1753	0.9187	30.327	0.8205	0.0802	0.4472	0.9904	49.107	0.5630	0.0090	1.9597	0.3336
PDDN	ICCV'23	0.9889	0.9702	41.374	0.5819	0.0240	3.3839	0.9088	29.844	0.8095	0.0852	0.5098	0.9892	48.947	0.5440	0.0102	0.1284	0.0395
INNF	AAAI'22	0.9176	0.9706	41.895	0.5756	0.0222	3.1000	0.9216	30.542	0.8250	0.0745	0.4831	0.9894	48.520	0.5664	0.0095	1.2201	0.0613
RFCO	TGRS'24	1.0992	0.9663	40.357	0.5285	0.0241	3.3407	0.9153	29.953	0.8134	0.0891	0.7378	0.9839	44.439	0.4492	0.0139	4.0509	5.2089
PANF	ICME'22	0.8943	0.9703	42.103	0.5638	0.0217	3.0638	0.9230	30.654	0.8231	0.0812	0.4690	0.9894	48.742	0.5377	0.0094	2.9426	1.5242
HFEAN	MM'23	0.9538	0.9675	41.587	0.5147	0.0236	3.4045	0.9086	29.716	0.7963	0.0842	0.4574	0.9009	44.910	0.5343	0.0092	24.892	0.5439
MSDDN	TGRS'23	1.0416	0.9635	40.885	0.5407	0.0256	3.5076	0.9023	29.424	0.8053	0.0831	0.6740	0.9807	45.686	0.4881	0.0133	0.0900	0.9605
DISP	AAAI'24	0.8759	0.9720	42.253	0.5837	0.0215	3.0096	0.9267	30.835	0.8315	0.0720	0.4493	0.9904	49.090	0.5679	0.0097	55.334	3.0872
NLUNET	TGRS'23	1.0034	0.9644	41.064	0.5413	0.0246	3.2981	0.9140	29.972	0.8162	0.0811	0.4976	0.9885	48.229	0.5458	0.0099	4.6099	0.3062
LGTEUN	IJCAI'23	0.8968	0.9734	42.677	0.5832	0.0211	3.0151	0.9246	30.788	0.8291	0.0723	0.4798	0.9894	48.529	0.5728	0.0097	3.2113	0.3004
MDCUN	CVPR'22	1.0060	0.9635	41.130	0.5274	0.0249	3.3531	0.9111	29.834	0.8142	0.0828	0.4830	0.9890	48.414	0.5461	0.0098	118.30	0.1538
GPPNN	CVPR'21	0.9248	0.9702	41.838	0.5807	0.0222	3.0923	0.9226	30.577	0.8290	0.0738	0.4812	0.9893	48.496	0.5596	0.0096	4.1901	0.3594
SSMP-Net (16-13)	Ours	0.8312	0.9743	42.661	0.5958	0.0202	2.9056	0.9292	31.109	0.8362	0.0693	0.4870	0.9904	49.001	0.5600	0.0098	3.8862	0.5422
SSMP-Net (32-9)	Ours	0.8483	0.9740	42.560	0.5911	0.0206	2.8891	0.9293	31.152	0.8364	0.0691	0.4290	0.9911	49.301	0.5803	0.0087	4.7905	0.5272
SSMP-Net (32-5)	Ours default	0.8349	0.9740	42.705	0.5942	0.0202	2.9054	0.9292	31.116	0.8361	0.0690	0.4286	0.9913	49.481	0.5996	0.0087	2.6698	0.2934

Table 3: More quantitative results comparing SSUN-Net with other methods on real data. The symbol ↑ or ↓ is used to indicate that a higher or lower value corresponds to a better result.

Metrics	Pure Deep Learning Methods									Deep Unfolding Methods					
	SFINet++	HFIN	WINET	PDDN	INNF	RFCO	PANF	HFEAN	MSDDN	DISP	NLUNET	LGTEUN	MDCUN	GPPNN	**SSMP-Net**
$D_\lambda \downarrow$	0.0784	0.0984	0.1187	0.0798	0.0708	0.1056	0.0751	0.0607	0.0882	0.0807	0.0763	0.1134	0.0765	0.0912	**0.0658**
$D_s \downarrow$	0.0886	0.0849	0.0827	0.1063	0.1076	0.1376	0.1027	0.1801	0.0807	0.0822	0.1074	0.1716	0.0846	0.0948	**0.0804**
$QNR \uparrow$	0.8400	0.8251	0.8084	0.8223	0.8292	0.7714	0.8299	0.7701	0.8382	0.8437	0.8245	0.7345	0.8454	0.8226	**0.8591**

tests in Tab. 2, color-coded in red, blue, and green for the best, second-best, and third-best performances, respectively. Our proposed SSUN-Net demonstrates distinct advantages in performance across the three datasets, consistently outperforming other methods to varying degrees. Specifically, compared with SOTA DUN-based methods DISP, NLUNET, LGTEUN, MDCUN and GPPNN SSUN-Net reduces ERGAS↓ by 0.0207 (4.8%), 0.0690 (16.1%), 0.0512 (11.9%), 0.0543 (12.6%), and 0.0526 (12.3%) on the GaoFen2 dataset; 0.1042 (3.6%), 0.3927 (13.5%), 0.1097 (3.8%), 0.4477 (15.5%), and 0.1869 (6.4%) on the WorldView III dataset; and 0.0410 (4.9%), 0.1684 (20.2%), 0.0619 (7.4%), 0.1710 (20.5%), and 0.0899 (10.8%) on the WorldView II dataset , respectively. Furthermore, we present supplementary results from testing on real data in Tab. 3, where our default SSUN-Net configuration continues to exhibit superior performance, particularly in the QNR performance index and maintaining a balanced hardware burden. In addition, our computational cost is also the lowest compared to these DUN-based methods.

Our method almost outperforms other algorithms, verifying the effectiveness of multi-scale priors.

2.2 More Qualitative Comparison In this subsection, we enhance the qualitative visual comparisons by comparing the reconstructed images of our SSUN-Net and several state-of-the-art Pan-sharpening methods. Initially, we conduct a qualitative visualization comparison by extracting the gradient of the reconstructed image and the ground truth image on WorldView II and WorldView III, where structural information is prominent, as depicted in Fig. 2 and Fig. 4. Detailed annotations are added in the locally enlarged areas, revealing that our SSUN-Net effectively captures finer texture features due to our gradient-based prior settings and multiscale calculations.

Furthermore, we present a comparison of the reconstruction results of the RGB band of GaoFen 2 and WorldView II, along with the MSE residual of the infrared band, showcased in Fig. 5, Fig. 1, and Fig. 3. Our proposed SSUN-Net demonstrates superior performance in mitigating the distortion of subtle textures and spectra compared to other methods, preserving more detailed information, particularly evident in the enlarged area of Fig. 1 and Fig. 3. In terms of MSE residuals, our reconstruction results closely align with the ground truth.

3 More Ablation Experiments.

Table 4: Comparative results of ablation for different numbers of iteration stages and hidden layer.

mid	stage (N)	GaoFen 2			WorldView II		WorldView III		Params (M)	Flops (G)
		SSIM	PSNR	QNR	SSIM	PSNR	SSIM	PSNR		
16	3	0.9903	48.899	0.8157	0.9719	42.200	0.9238	30.731	0.1256	0.904
16	5	0.9905	48.973	0.7741	0.9726	42.321	0.9265	30.939	0.2089	1.501
16	7	0.9911	49.224	0.8081	0.9732	42.405	0.9273	30.991	0.2922	2.097
16	9	0.9906	49.063	0.8673	0.9734	42.439	0.9276	31.015	0.3755	2.693
16	11	0.9905	49.014	0.7885	0.9739	42.528	0.9279	31.040	0.4589	3.290
16	13	0.9904	49.001	0.8499	0.9743	42.661	0.9292	31.109	0.5422	3.886
32	3	0.9906	48.891	0.8472	0.9729	42.371	0.9269	30.945	0.1765	1.609
32	5	**0.9913**	**49.481**	**0.8591**	0.9740	**42.705**	0.9292	31.117	0.2934	2.670
32	7	0.9911	49.120	0.8034	0.9738	42.519	0.9286	31.071	0.4103	3.730
32	9	0.9911	49.301	0.8510	0.9740	42.560	0.9293	**31.152**	0.5272	4.791
32	11	0.9911	49.271	0.8562	0.9737	42.511	0.9292	31.137	0.6441	5.851
32	13	0.9911	49.261	0.7644	**0.9742**	42.593	**0.9296**	31.111	0.7609	6.911

3.1 The Number of Convolutional Hidden Layers (mid) and Stage (N) We investigate the impact of varying quantities on different N of SSUN-Net and analyze the influence

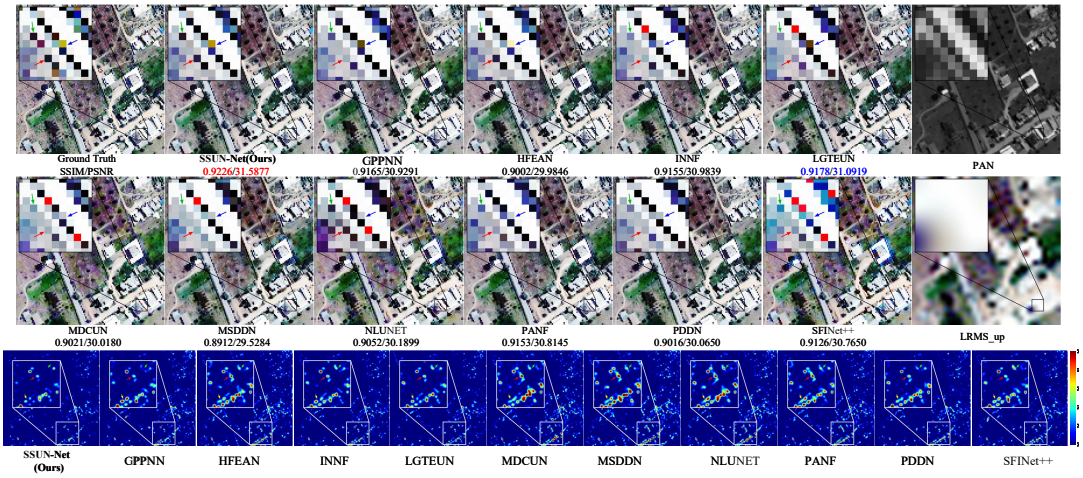

Figure 1: Visual comparison of SSUN-Net with other methods on simulated data from the WorldView III. The arrows point to details that are zoomed for better comparison.

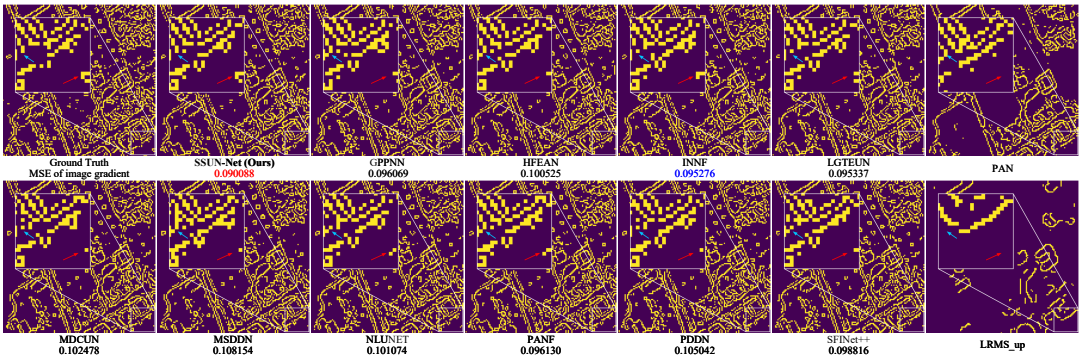

Figure 2: Gradient map visualization of reconstructed images and ground truth images on simulated data from the WorldView III dataset. The arrows point to details that are zoomed for better comparison.

of mid. The experimental results, presented in Tab. 4, indicate that enhancing the number of iterations stage and hidden layer can lead to improved performance; however, this enhancement comes with the trade-off of increased model parameters and hardware requirements. Consequently, we establish the default configuration with 5 stages and a hidden layer of 32 to strike a balance between model intricacy and performance.

Table 5: Comparative results of ablation for different Hyperparameters of Loss Function.

λ_{loss}	10	0.1	0.05	0.01	0
SSIM	0.9735	0.9740	0.9739	0.9735	0.9728
PSNR	42.673	42.705	42.695	42.687	42.630

3.2 The Hyperparameters of Loss Function (λ_{loss}) We investigated the effect of the hyperparameter λ_{loss} from Eq. (29) in the main text on the structural loss trade-off. Experimental results presented in Tab. 5, indicate that structural loss encourages SSUN-Net to preserve edge and texture consistency during training, thereby enhancing performance. Furthermore, variations in parameter settings can introduce fluctuations in model training. Our default setting of 0.1 yielded the most consistent optimal performance in our experiments.

4 Broader Impact

Pan-sharpening overcomes the limitations of sensor hardware by fusing panchromatic and multispectral images to obtain higher resolution multispectral images. Extracting and integrating spatial and spectral features from panchromatic and multispectral images is an important area of research interest. Our proposed SSUN-Net demonstrates superior computational cost, accuracy, and transparency compared to existing state-of-the-art (SOTA) methods. We will continue to focus on the potential of model interpretability and performance.

5 Limitations

Our work may have two potential limitations. First, our model can benefit from incorporating more prior knowledge,

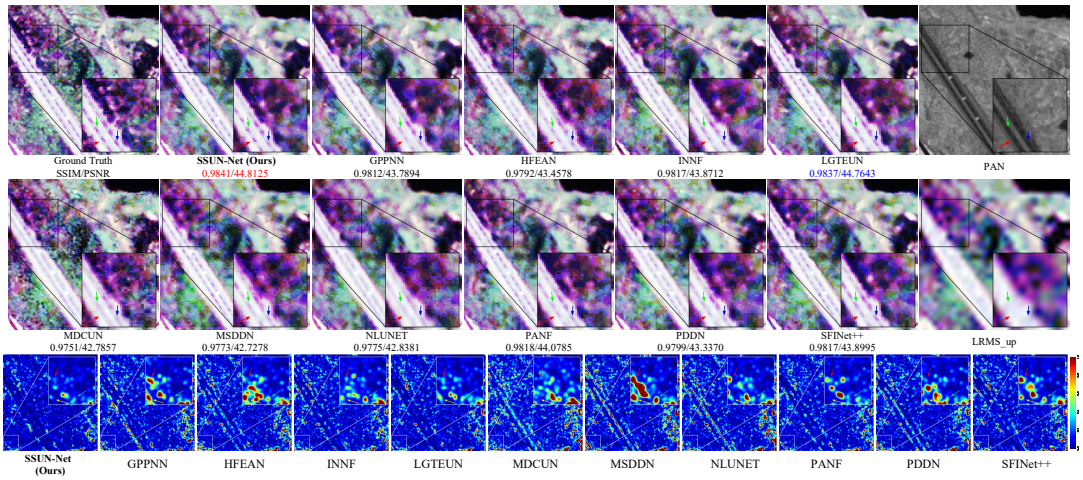

Figure 3: Visual comparison of SSUN-Net with other methods on simulated data from the WorldView II. The arrows point to details that are zoomed for better comparison.

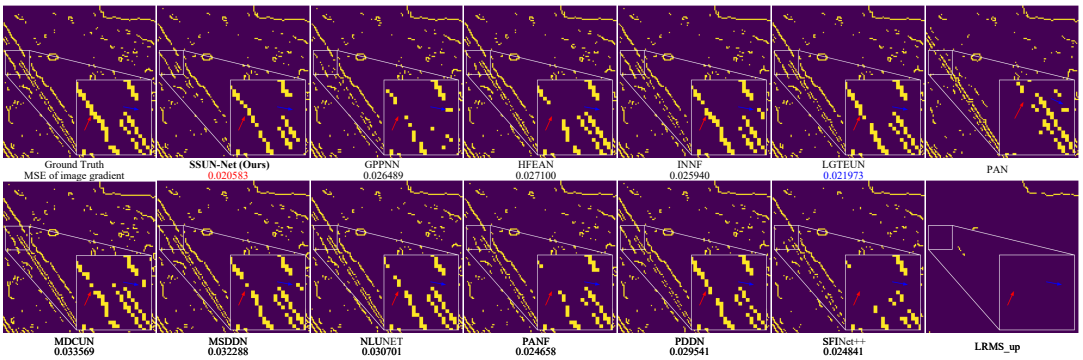

Figure 4: Gradient map visualization of reconstructed images and ground truth images on simulated data from the WorldView II dataset. The arrows point to details that are zoomed for better comparison.

such as low-rank prior (Liu, Xiao, and Li 2018) or spatial-spectral sparse prior (Zhu and Bamler 2013), to strengthen the connection between deep learning and VO techniques, thereby improving network performance. Secondly, we can make more attempts in areas such as hyperspectral fusion and remote sensing image super-resolution.

References

Aiazzi, B.; Baronti, S.; and Selva, M. 2007. Improving Component Substitution Pansharpening Through Multivariate Regression of MS +Pan Data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10): 3230–3239.

He, X.; Yan, K.; Li, R.; Xie, C.; Zhang, J.; and Zhou, M. 2023a. Pyramid Dual Domain Injection Network for Pan-sharpening. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 12862–12871.

He, X.; Yan, K.; Zhang, J.; Li, R.; Xie, C.; Zhou, M.; and Hong, D. 2023b. Multiscale Dual-Domain Guidance Network for Pan-Sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–13.

Li, M.; Liu, Y.; Xiao, T.; Huang, Y.; and Yang, G. 2023a. Local-Global Transformer Enhanced Unfolding Network for Pan-sharpening. In Elkind, E., ed., *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, 1071–1079. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Li, X.; Li, Y.; Shi, G.; Zhang, L.; Li, W.; and Lei, D. 2023b. Pansharpening Method Based on Deep Nonlocal Unfolding. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–11.

Liu, J. G. 2000. Smoothing Filter-based Intensity Modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*, 21(18): 3461–3472.

Liu, P.; Xiao, L.; and Li, T. 2018. A Variational Pan-Sharpening Method Based on Spatial Fractional-Order Geometry and Spectral-Spatial Low-Rank Priors. *IEEE Transactions on Geoscience and Remote Sensing*, 56(3): 1788–1802.

Qu, J.; Liu, X.; Dong, W.; Liu, Y.; Zhang, T.; Xu, Y.; and Li, Y. 2024. Progressive Multi-Iteration Registration-Fusion

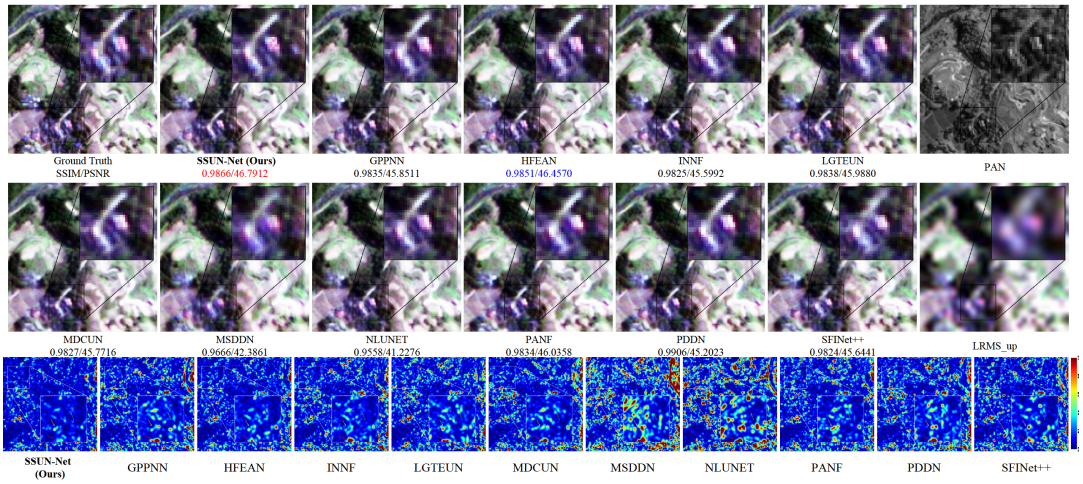

Figure 5: Visual comparison of SSUN-Net with other methods on simulated data from the GaoFen2. The arrows point to details that are zoomed for better comparison.

Co-Optimization Network for Unregistered Hyperspectral Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–14.

Tan, J.; Huang, J.; Zheng, N.; Zhou, M.; Yan, K.; Hong, D.; and Zhao, F. 2024. Revisiting Spatial-Frequency Information Integration from a Hierarchical Perspective for Panchromatic and Multi-Spectral Image Fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 25922–25931.

Wang, Y.; Lin, Y.; Meng, G.; Fu, Z.; Dong, Y.; Fan, L.; Yu, H.; Ding, X.; and Huang, Y. 2023. Learning High-frequency Feature Enhancement and Alignment for Pan-sharpening. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, 358–367. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701085.

Xu, S.; Zhang, J.; Zhao, Z.; Sun, K.; Liu, J.; and Zhang, C. 2021. Deep Gradient Projection Networks for Pan-sharpening. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1366–1375.

Yang, G.; Zhou, M.; Yan, K.; Liu, A.; Fu, X.; and Wang, F. 2022. Memory-augmented Deep Conditional Unfolding Network for Pansharpening. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1778–1787.

Zhang, J.; He, X.; Yan, K. R.; Cao, K.; Li, R.; Xie, C.; Zhou, M.; and Hong, D. 2024. Pan-Sharpening With Wavelet-Enhanced High-Frequency Information. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–14.

Zhou, H.; Liu, Q.; and Wang, Y. 2022. PanFormer: A Transformer Based Model for Pan-Sharpening. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.

Zhou, M.; Huang, J.; Fang, Y.; Fu, X.; and Liu, A. 2022. Pan-Sharpening with Customized Transformer and Invertible Neural Network. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(3): 3553–3561.

Zhou, M.; Huang, J.; Yan, K.; Hong, D.; Jia, X.; Chanussot, J.; and Li, C. 2024. A General Spatial-Frequency Learning Framework for Multimodal Image Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–18.

Zhu, X. X.; and Bamler, R. 2013. A Sparse Image Fusion Algorithm With Application to Pan-Sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5): 2827–2836.