

Supplementary Materials of Deep Unfolding Network with Spatial-spectral Perception Enhanced for Pan-sharpening

Mengjiao Zhao^{1,*}
mj.zhao@zju.edu.cn

Mengting Ma^{2,*}
mtma@zju.edu.cn

Xiangdong Li¹
xiangdong.li@zju.edu.cn

Ao Gao¹
gaoao.olivia@zju.edu.cn

Siyang Song³
ss2796@cam.ac.uk

Wei Zhang^{1,4,†}
cstzhangwei@zju.edu.cn

¹ School of Software Technology, Zhejiang University, Hangzhou, China

² College of Computer Science and Technology, Zhejiang University, Hangzhou, China

³ School of Computing and Mathematical Sciences, University of Leicester, UK

⁴ Innovation Center of Yangtze River Delta, Zhejiang University, Jiaxing, China

1 More quantitative and qualitative results

In the main manuscript, due to space constraints, we only show quantitative results comparing our method with DL-based methods on the full-resolution WorldView-2 dataset. In this section, we present in Tab. 1 the quantitative results of our method compared to all other competing methods on the full-resolution Gaofen-2, WorldView-2, and Worldview-3 datasets. Additionally, in the main manuscript, we only provide qualitative results on Worldview-2. To further demonstrate the effectiveness of our method, we show visual results on the Gaofen-2 and WorldView-3 datasets in Fig. 1 and Fig. 2, respectively. From the results, it can be seen that our method produces visually pleasing outcomes.

2 Dataset Details

In this section, we introduce three satellite datasets used in the experimental section of the main manuscript, i.e., WorldView-2, WorldView-3, and GaoFen-2. The remote sensing images we use are collected by different satellite sensors, with the resolution of the captured PAN images being four times that of the corresponding LR-MS images. Due to the large size of remote sensing images, it is difficult to feed them into neural networks. Therefore,

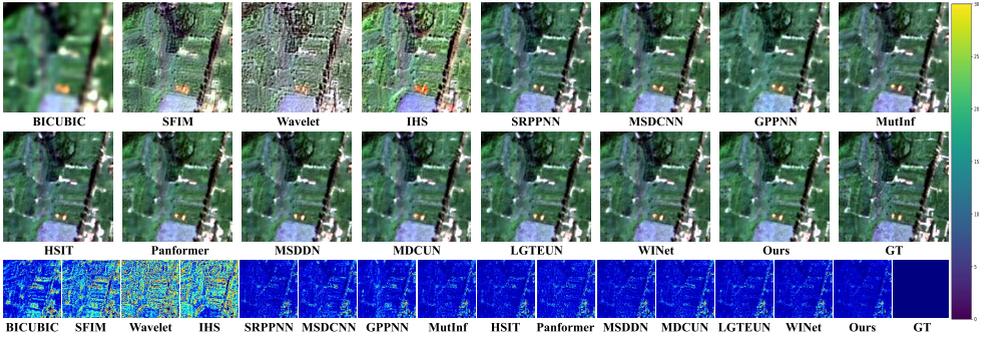


Figure 1: Visual comparison and absolute errors of our method versus other representative pan-sharpening methods on the GaoFen-2 dataset.

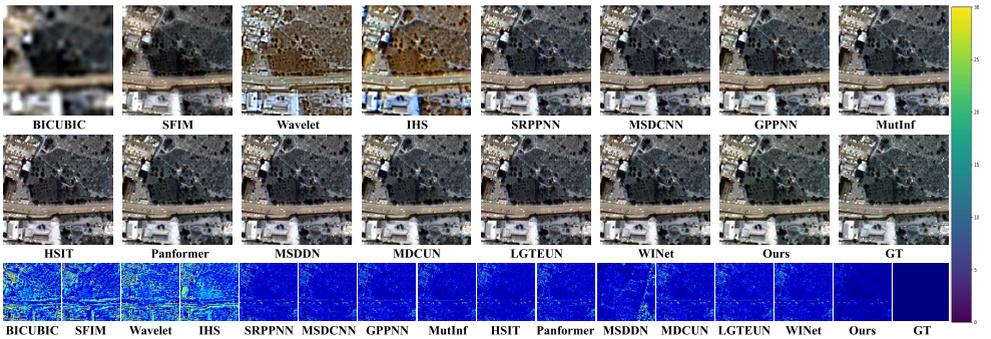


Figure 2: Visual comparison and absolute errors of our method versus other representative pan-sharpening methods on the WorldView-3 dataset.

| Methods | GaoFen-2 | | | WorldView-2 | | | WorldView-3 | | |
|-------------|------------------------|------------------|----------------|------------------------|------------------|----------------|------------------------|------------------|----------------|
| | $D_\lambda \downarrow$ | $D_s \downarrow$ | QNR \uparrow | $D_\lambda \downarrow$ | $D_s \downarrow$ | QNR \uparrow | $D_\lambda \downarrow$ | $D_s \downarrow$ | QNR \uparrow |
| SFIM | 0.0687 | 0.0624 | 0.8752 | 0.0737 | 0.0899 | 0.8439 | 0.0094 | 0.1061 | 0.8854 |
| BICUBIC | 0.0660 | 0.2144 | 0.7340 | 0.0628 | 0.1411 | 0.8058 | 0.0211 | 0.0626 | 0.9278 |
| Wavelet | 0.1310 | 0.0807 | 0.8041 | 0.0968 | 0.1020 | 0.8126 | 0.0552 | 0.133 | 0.8193 |
| IHS | 0.0782 | 0.0904 | 0.8405 | 0.0874 | 0.1187 | 0.8053 | 0.0176 | 0.1223 | 0.8621 |
| SRPPNN | 0.0663 | 0.2054 | 0.7413 | 0.0640 | 0.0858 | 0.8567 | 0.0396 | 0.0448 | 0.9176 |
| MSDCNN | 0.0715 | 0.2092 | 0.8630 | 0.0639 | 0.0783 | 0.8637 | 0.0242 | 0.0476 | 0.9292 |
| GPPNN | 0.0719 | 0.0734 | 0.7397 | 0.0670 | 0.0785 | 0.8607 | 0.0196 | 0.0484 | 0.9329 |
| MutInf | 0.0755 | 0.1762 | 0.7612 | 0.0638 | 0.0794 | 0.8644 | 0.0164 | 0.0420 | 0.9423 |
| HSIT | 0.0727 | 0.1637 | 0.7764 | 0.0640 | 0.0862 | 0.8609 | 0.0326 | 0.0453 | 0.9238 |
| Panformer | 0.0647 | 0.1996 | 0.7481 | 0.0627 | 0.0825 | 0.8609 | 0.0210 | 0.0444 | 0.9355 |
| MSDDN | 0.0693 | 0.1580 | 0.7840 | 0.0639 | 0.0758 | 0.8659 | 0.0170 | 0.0381 | 0.9457 |
| MDCUN | 0.0667 | 0.2334 | 0.7149 | 0.0661 | 0.0834 | 0.8568 | 0.0413 | 0.0345 | 0.9258 |
| LGTEUN | 0.0749 | 0.1468 | 0.7895 | 0.0645 | 0.0771 | 0.8644 | 0.0173 | 0.0328 | 0.9504 |
| WINet | 0.0681 | 0.2172 | 0.7283 | 0.0644 | 0.0761 | 0.8627 | 0.0190 | 0.0395 | 0.9422 |
| Ours | 0.0642 | 0.0605 | 0.8602 | 0.0622 | 0.0757 | 0.8652 | 0.0101 | 0.0280 | 0.9515 |

Table 1: Quantitative results of all competing methods on three full resolution datasets. The best and second best values are highlighted in **bold** and underline, respectively.

| Datasets | GaoFen-2 | WorldView-2 | WorldView-3 |
|------------------|---------------------------|---------------------------|---------------------------|
| bit depth | 11 | 11 | 11 |
| Training set | 1036 | 1012 | 910 |
| Test set | 136 | 145 | 144 |
| LR-MS image size | $32 \times 32 \times 4$ | $32 \times 32 \times 4$ | $32 \times 32 \times 4$ |
| PAN image size | $128 \times 128 \times 1$ | $128 \times 128 \times 1$ | $128 \times 128 \times 1$ |
| HR-MS image size | $128 \times 128 \times 4$ | $128 \times 128 \times 4$ | $128 \times 128 \times 4$ |

Table 2: Detailed information of the datasets used.

| Stage Number | Params(M) | Flops(G) | PSNR \uparrow | SSIM \uparrow | Q4 \uparrow | SAM \downarrow | ERGAS \downarrow |
|--------------|-----------|----------|-------------------------|------------------------|------------------------|------------------------|------------------------|
| $K = 1$ | 0.1417 | 0.9137 | 42.6860 | 0.9785 | 0.8415 | 0.0207 | 0.9243 |
| $K = 2$ | 0.1712 | 1.8275 | 42.7301 | 0.9787 | 0.8427 | 0.0206 | 0.9213 |
| $K = 3$ | 0.4249 | 2.7412 | 42.6785 | 0.9788 | 0.8426 | 0.0206 | 0.9191 |
| $K = 4$ | 0.5665 | 3.6549 | 42.6747 | 0.9781 | 0.8420 | 0.0208 | 0.9267 |

Table 3: Quantitative results of our method with different number of stages on WorldView-2.

we crop the LR-MS images and PAN images into small patches to form training and testing sets, similar to most pan-sharpening works [10, 2, 9]. The information about the testing and training sets for these three datasets is shown in Tab. 2.

3 Limitations

While the proposed method brings promising results, there are still some notable issues that require further research. Firstly, due to the specificity of different satellites, our method may not fully guarantee superior performance in all full-resolution scenarios, as shown in Tab. 1. Therefore, our method shows potential for performance improvement in full-resolution pan-sharpening scenarios. Additionally, our model involves a large number of floating-point operations. As shown in Tab. 3, the computational complexity of the model increases linearly with the number of stages. Therefore, further exploration of potential acceleration optimization strategies to improve model efficiency will make our proposed SSPEDUN more competitive.

4 Impact Statement

Our proposed SSPEDUN aims to advance the field of computer vision by providing an efficient and feasible approach for image fusion or restoration tasks, such as super-resolution, hyperspectral image reconstruction, and spectral compressed imaging. It has the potential to impact various industries, including agricultural development, environmental monitoring, and military applications, by enhancing the accuracy and efficiency of generated images.

References

- [1] Xuanhua He, Keyu Yan, Rui Li, Chengjun Xie, Jie Zhang, and Man Zhou. Pyramid dual domain injection network for pan-sharpening. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12908–12917, 2023.
- [2] Hebaixu Wang, Meiqi Gong, Xiaoguang Mei, Hao Zhang, and Jiayi Ma. Deep unfolded network with intrinsic supervision for pan-sharpening. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 5419–5426, 2024.
- [3] Zeyu Zhu, Xiangyong Cao, Man Zhou, Junhao Huang, and Deyu Meng. Probability-based global cross-modal upsampling for pansharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14039–14048, 2023.