

# Memory-Augmented Model-Driven Network for Pansharpening (Supplementary Material)

Keyu Yan<sup>1,2</sup>, Man Zhou<sup>1,2</sup>, Li Zhang<sup>1,2</sup>, and Chengjun Xie<sup>1</sup>

<sup>1</sup> Hefei Institute of Physical Science, Chinese Academy of Sciences, China

<sup>2</sup> University of Science and Technology of China, China

{keyu,manman,zanly20}@mail.ustc.edu.cn, cjxie@iim.ac.cn

The supplementary material is structured as follows. In Section 1, we give a detailed derivation process of the MAP model for Pan-sharpening. In Section 2, we show the qualitative visualization comparison over the additional WorldViewIII and GaoFen2 satellite datasets. Due to the page limits, it has not been shown in the main manuscript. The quantitative comparison over multiple satellite datasets are also reported in Section 2 and Section 3 aims to show the effect of different stage numbers.

## 1 MAP model for Pan-sharpening

In Pan-sharpening, we assume that LRMS image  $\mathbf{L}$  is obtained through performing the blurring kernel  $\mathbf{k}$  and down-sampling operator over the HRMS image  $\mathbf{H}$ , and thus the degradation model can be mathematically formulated as

$$\mathbf{L} = (\mathbf{H} \otimes \mathbf{k}) \downarrow_s + \mathbf{n}_s, \quad (1)$$

The spatial resolution ratio between  $\mathbf{H}$  and  $\mathbf{L}$  is  $r = M/m = N/n$ . The observation model in Eq. 1 can be equivalently reformulated as

$$\mathbf{L} = \mathbf{DKH} + \mathbf{n}_s, \quad (2)$$

where  $\mathbf{K}$  is the matrix form of kernel  $\mathbf{k}$ , and  $\mathbf{D}$  is the matrix form of down-sampling operator. Based on the observation model in Eq. 2, the distribution of  $\mathbf{L}$  is defined as

$$P(\mathbf{L}|\mathbf{H}) = \mathcal{N}(\mathbf{L}|\mathbf{DKH}, \sigma^2\mathbf{I}), \quad (3)$$

where  $\mathcal{N}(\mathbf{L}|\mathbf{DKH}, \sigma^2\mathbf{I})$  denotes the Gaussian distribution with mean  $\mathbf{DKH}$  and covariance matrix  $\sigma^2\mathbf{I}$ .

Pan-sharpening super-resolves the LRMS image using the PAN image, which is typically acquired in the same scene as the LRMS image, so the HRMS image and the PAN image share some global and local relevant features. We create a local implicit prior and a global implicit prior to capture both properties. The local implicit prior implicitly models the relationship between the HRMS image and the PAN image from a local perspective, and so can assist in capturing the local related information between the HRMS and the PAN image. The global implicit prior addresses the non-local auto-regression property between the two

images from a global perspective, allowing for effective use of the global correlation between the two images.

Specifically, we assume the local implicit prior distribution  $P_1(\mathbf{H}|\mathbf{P})$  and global implicit prior distribution  $P_2(\mathbf{H}|\mathbf{P})$  separately as follows:

$$P_1(\mathbf{H}|\mathbf{P}) \propto \exp\{-\eta\Omega_1(\mathbf{H}|\mathbf{P})\}, \quad (4)$$

$$P_2(\mathbf{H}|\mathbf{P}) \propto \exp\{-\lambda\Omega_{NL}(\mathbf{H}|\mathbf{P})\}, \quad (5)$$

where  $\Omega_1(\mathbf{H}|\mathbf{P})$  and  $\Omega_{NL}(\mathbf{H}|\mathbf{P})$  are two energy functions related to  $\mathbf{H}$  and  $\mathbf{P}$ ,  $\eta$  and  $\lambda$  are the weight parameters. For simplicity, we assume the distribution of  $\mathbf{H}$  is

$$P(\mathbf{H}|\mathbf{P}) \propto P_1(\mathbf{H}|\mathbf{P})P_2(\mathbf{H}|\mathbf{P}). \quad (6)$$

Therefore, the posterior of  $\mathbf{H}$  given  $\mathbf{L}$  and  $\mathbf{P}$  can be computed by the Bayes formula:

$$P(\mathbf{H}|\mathbf{L}, \mathbf{P}) = \frac{P(\mathbf{L}|\mathbf{H})P(\mathbf{H}|\mathbf{P})}{P(\mathbf{L}|\mathbf{P})}, \quad (7)$$

where  $P(\mathbf{L}|\mathbf{P})$  is the marginal distribution of  $\mathbf{L}$  which is not related with  $\mathbf{H}$ . By using the maximum a posterior (MAP) principle,  $\mathbf{H}$  can be obtained by maximizing the log-posterior  $\log P(\mathbf{H}|\mathbf{L}, \mathbf{P})$ , which is equivalent to the following optimization problem:

$$\max_{\mathbf{H}} \log P(\mathbf{L}|\mathbf{H}) + \log P_1(\mathbf{H}|\mathbf{P}) + \log P_2(\mathbf{H}|\mathbf{P}). \quad (8)$$

Further, Eq. 8 can be reformulated as

$$\max_{\mathbf{H}} \frac{1}{2} \|\mathbf{L} - \mathbf{DKH}\|_2^2 + \eta\Omega_1(\mathbf{H}|\mathbf{P}) + \lambda\Omega_{NL}(\mathbf{H}|\mathbf{P}). \quad (9)$$

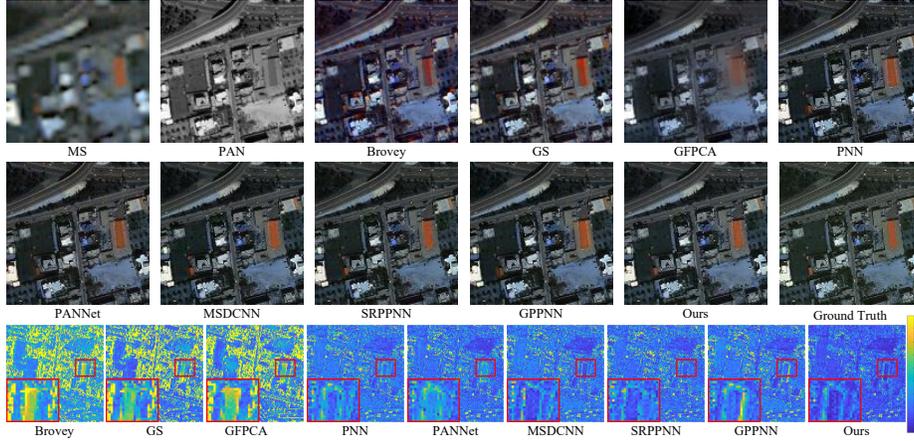
Eq. 9 is our final model that we develop the optimization algorithm to solve.

## 2 Qualitative and quantitative results

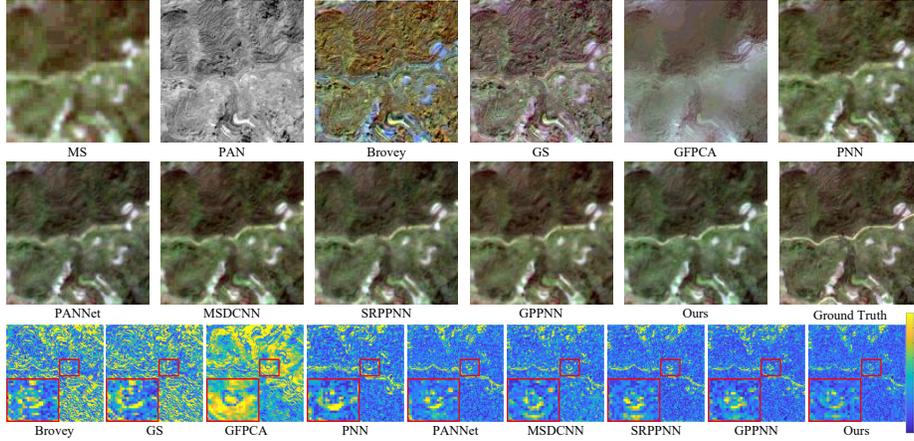
The visual comparison between our method and several state-of-the-art Pan-sharpening methods is shown in Fig. 1 and Fig. 2 on the representative samples of WorldViewIII and GaoFen2 datasets. Images in the last row are the MSE residues between the pan-sharpened results and the ground truth. Compared with other competing methods, our model has minor spatial and spectral distortions. It can be easily concluded from the observation of MSE maps. As for the MSE residues, it's noticed that our proposed method is closest to the ground truth than other comparison methods. Therefore, it can be affirmed that our method achieves the best performance than other competitive algorithms.

We show the quantitative experiment comparisons of several measurement metrics, including PSNR, SSIM, SAM, ERGAS, SCC, Q-index, the three non-reference metrics of  $D_\lambda$ ,  $D_S$ , QNR in Table 1, Table 2 and Table 3 between our

predictions and that of the baseline method. The best results are highlighted by **bold**. It can be clearly seen that our method performs the best compared with other state-of-the-art methods in all the indexes, indicating the superiority of



**Fig. 1.** Visual comparisons of the fused HRMS image for all the methods on one WorldView-III dataset. Images in the last row visualize the MSE between the pan-sharpened results and the ground truth.



**Fig. 2.** Visual comparisons of the fused HRMS image for all the methods on one GaoFen2 dataset. Images in the last row visualize the MSE between the pan-sharpened results and the ground truth.

### 3 Qualitative results of different stages

To further undertake the effect of stage number  $K$ , we present the representative sample generated by different model variants with stage number from  $K = 1$  to  $K = 6$  in Fig. 3. From Fig. 3, we can also observe that the model with

**Table 1.** Quantitative comparison of the SOTA methods on the WorldView-II dataset. Best results are highlighted by **bold**.  $\uparrow$  indicates that the larger the value, the better the performance, and  $\downarrow$  indicates that the smaller the value, the better the performance.

Methods	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$	ERGAS $\downarrow$	SCC $\uparrow$	Q $\uparrow$	$D_\lambda$ $\downarrow$	$D_S$ $\downarrow$	QNR $\uparrow$
SFIM	34.1297	0.8975	0.0439	2.3449	0.9079	0.6064	0.0915	0.1277	0.7942
GS	<b>35.6376</b>	0.9176	0.0423	1.8774	0.9225	0.6307	0.0607	0.1285	0.8195
Brovey	35.8646	0.9216	0.0403	1.8238	0.8913	0.6163	0.0770	0.1360	0.7977
IHS	35.2962	0.9027	0.0461	2.0278	0.8534	0.5704	0.0774	0.1578	0.7770
GFPCA	34.5580	0.9038	0.0488	2.1411	0.8924	0.4665	0.1016	0.1656	0.7508
PNN	40.7550	0.9624	0.0259	1.0646	0.9677	0.7426	0.065	0.1186	0.8250
PANNet	40.8176	0.9626	0.0257	1.0557	0.968	0.7437	0.0645	0.1189	0.8252
MSDCNN	41.3355	0.9664	0.0242	0.9940	0.9721	0.7577	0.0635	0.1172	0.8276
SRPPNN	41.4538	0.9679	0.0233	0.9899	0.9729	0.7691	0.0637	0.1164	0.8281
GPPNN	41.1622	0.9684	0.0244	1.0315	0.9722	0.7627	0.0642	0.1163	0.8278
Ours	<b>41.8577</b>	<b>0.9697</b>	<b>0.0229</b>	<b>0.9420</b>	<b>0.9745</b>	<b>0.7740</b>	<b>0.0629</b>	<b>0.1154</b>	<b>0.8299</b>

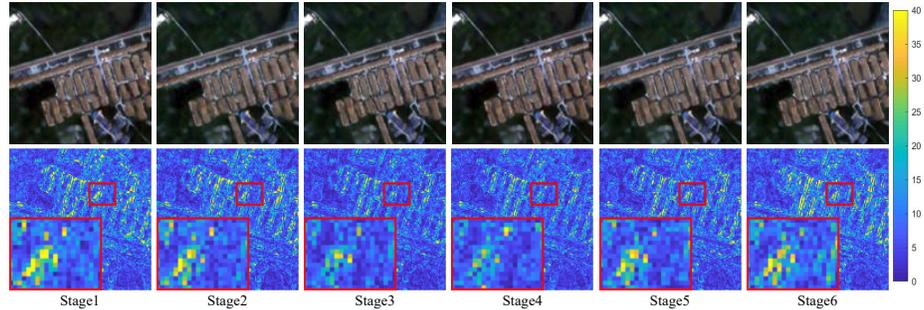
**Table 2.** Quantitative comparison of the SOTA methods on the WorldView-III dataset. Best results are highlighted by **bold**.  $\uparrow$  indicates that the larger the value, the better the performance, and  $\downarrow$  indicates that the smaller the value, the better the performance.

Methods	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$	ERGAS $\downarrow$	SCC $\uparrow$	Q $\uparrow$	$D_\lambda$ $\downarrow$	$D_S$ $\downarrow$	QNR $\uparrow$
SFIM	21.8212	0.5457	0.1208	8.973	0.6952	0.4531	0.0448	0.1265	0.8347
GS	22.5608	0.547	0.1217	8.2433	0.7131	0.4411	0.035	0.2011	0.7695
Brovey	22.506	0.5466	0.1159	8.2331	0.7033	0.4394	0.0481	0.2006	0.7603
IHS	22.5579	0.5354	0.1266	8.3616	0.6994	0.4301	0.0356	0.2073	0.7634
GFPCA	22.3400	0.4826	0.1294	8.3964	0.6987	0.3115	0.0528	0.1214	0.8325
PNN	29.9418	0.9121	0.0824	3.3206	0.954	0.8679	0.046	0.0933	0.8654
PANNet	29.684	0.9072	0.0851	3.4263	0.9512	0.8631	0.0474	0.0942	0.8634
MSDCNN	30.3038	0.9184	0.0782	3.1884	0.9577	0.8763	0.0432	<b>0.0877</b>	<b>0.8732</b>
SRPPNN	30.4346	0.9202	0.0770	3.1553	0.9581	0.8776	<b>0.0414</b>	0.0909	0.8719
GPPNN	30.1785	0.9175	0.0776	3.2593	0.9569	0.8739	0.0438	0.0936	0.8671
Ours	<b>30.5451</b>	<b>0.9214</b>	<b>0.0769</b>	<b>3.1032</b>	<b>0.9598</b>	<b>0.8804</b>	0.0435	0.0911	0.8698

**Table 3.** Quantitative comparison of the SOTA methods on the GaoFen2 dataset. Best results are highlighted by **bold**.  $\uparrow$  indicates that the larger the value, the better the performance, and  $\downarrow$  indicates that the smaller the value, the better the performance.

Methods	PSNR $\uparrow$	SSIM $\uparrow$	SAM $\downarrow$	ERGAS $\downarrow$	SCC $\uparrow$	Q $\uparrow$	$D_\lambda$ $\downarrow$	$D_S$ $\downarrow$	QNR $\uparrow$
SFIM	36.9060	0.8882	0.0318	1.7398	0.8128	0.4349	0.0691	0.1312	0.8109
GS	37.2260	0.9034	0.0309	1.6736	0.7851	0.4211	0.0397	0.1214	0.8445
Brovey	37.7974	0.9026	0.0218	1.3720	0.6446	0.3857	0.0905	0.1443	0.7790
IHS	38.1754	0.9100	0.0243	1.5336	0.6738	0.3682	0.0418	0.1345	0.8301
GFPCA	37.9443	0.9204	0.0314	1.5604	0.8032	0.3236	0.0898	0.1815	0.7445
PNN	43.1208	0.9704	0.0172	0.8528	0.9400	0.7390	0.0387	0.1162	0.8494
PANNet	43.0659	0.9685	0.0178	0.8577	0.9402	0.7309	0.0369	0.1219	0.8455
MSDCNN	45.6874	0.9827	0.0135	0.6389	0.9526	0.7759	0.0368	0.1112	0.8560
SRPPNN	47.1998	0.9877	0.0106	0.5586	0.9564	0.7900	0.0364	0.1087	0.8588
GPPNN	44.2145	0.9815	0.0137	0.7361	0.9510	0.7721	0.0362	0.1078	0.8612
Ours	<b>47.2668</b>	<b>0.9890</b>	<b>0.0102</b>	<b>0.5472</b>	<b>0.9597</b>	<b>0.7973</b>	<b>0.0332</b>	<b>0.1110</b>	<b>0.8594</b>

stage number  $K = 4$  obtains the best visual effect, which is consistent with the quantitative results.



**Fig. 3.** Visual comparison of different stages.