Binarized Neural Network for Multi-spectral Image Fusion

Supplementary Material

We organize the supplementary material as follows:
Sec. 1 introduces the binarization of activation and weights.
Sec. 2 provides the experimental details.
Sec. 3 presents additional experimental results.

1. Binarization of activation and weights

In binarization operation, the decomposed full-precision features X_f are binarized to a set $\mathbb{B} \in \{-1, +1\}$ through the Sign function. Specifically, the binarization of the activation can be formulated as follows:

$$X_b = \pi(X_f - \varepsilon) = \begin{cases} +1, & \text{if } X_f > \varepsilon \\ -1, & \text{if } X_f \leq \varepsilon \end{cases}, \tag{1}$$

where π is the mapping of Sign function and $\varepsilon \in \mathbb{R}^C$ is a learnable threshold applied to adjust the binarized range of the corresponding channel of X_f. Likewise, the binarization of the kernel weights is written as follows:

$$\begin{aligned} W_b^j &= s_j \cdot \pi(W_f^j) = \begin{cases} +s_j, & \text{if } W_f^j > 0 \\ -s_j, & \text{if } W_f^j \leq 0 \end{cases}, \\ s_j &= \frac{\left\| W_f^j \right\|_1}{C \times K \times K}, \quad j = 1, \cdots, C \ , \end{aligned} \tag{2}$$

where $W_f^j, W_b^j \in \mathbb{R}^{C \times K \times K}$ denotes the j-th kernel of the full-precision kernels W_f and the binarized kernels W_b, while s_j is a scaling factor used to narrow the difference between the full-precision and binary weights. $\|\cdot\|_1$ indicates the $\mathcal{L}_1$ norm.

2. Experimental Details

Dataset Simulation. We assess our proposed method using three widely recognized satellite datasets for the pansharpening task: WorldView-3 (WV3), GaoFen-2 (GF2), and QuickBird (QB). Notably, we leverage Wald's protocol [16] to simulate the training data. In this way, each satellite dataset provides a comprehensive collection of image pairs designated for training, validation, and testing. The training set includes up-sampled MS (denoted as LRMS), PAN, and GT images at a spatial resolution of 64×64, and originally observed MS images at 16×16. For the reduced-resolution testing set, LRMS, PAN, and GT images are at 256×256, with MS images at 64×64. While the full-resolution testing set comprises LRMS and PAN images at 512×512, and MS images at 128×128. Additional dataset details can be found in [2].

Binary Benchmark. To evaluate our BNNPan's performance, we compare with state-of-the-art binary methods. These include BNN [4], IRNet [11], ReActNet [9], BTM [5], E2FIF [12], FABNet [6], and BBCU [18].

Full-Precision Benchmark. We conduct a comprehensive comparison with state-of-the-art full-precision approaches, including PNN [10], MSDCNN [20], GPPNN [19], FusionNet [1], LAGConv [7], HFEAN [17], and HFIN [13], to further highlight the superiority of our model in terms of both accuracy and efficiency.

Metrics. For reduced-resolution assessments, we use the Spectral Angle Mapper (SAM) [21], the ERGAS [15], Q2n (Q8 for 8-band datasets and Q4 for 4-band datasets) [3], and the Peak Signal-to-Noise Ratio (PSNR). For full-resolution evaluations, we utilize three no-reference metrics: HQNR, the spectral distortion D_λ index, and the spatial distortion D_s index [14].

Experimental Settings. All deep learning models are implemented in PyTorch and trained on a single NVIDIA RTX 4090 GPU. The Adam optimizer [8] is used with beta values of (0.9, 0.999) and a weight decay of 1×10^{-6}. We set a minibatch size of 32, with an initial learning rate of 3×10^{-4}. The learning rate decays by a factor of 0.1 every 250 epochs, with training concluding after 1000 epochs.

3. Additional Experimental Results

3.1. Evaluation on Reduced-resolution Scene

We first evaluate the performance of our model on the reduced-resolution GF2 and QB datasets, as summarized in the left panels of Tabs. 1 and 2. It is evident that the proposed model consistently outperforms other binary methods across all metrics in the reduced-resolution setting. Notably, our BNNPan outshines the second-best FABNet by a remarkable 2 dB in PSNR on the GF2 dataset. On the QB dataset, BNNPan still excels in both spectral and spatial metrics. These findings robustly validate its superior spectral preservation and spatial reconstruction quality. Additionally, when compared to full-precision approaches, BNNPan surpasses the representative FusionNet on both the GF2 and QB datasets, achieving results comparable to those of resource-intensive full-precision methods. As illustrated in Figs. 1 and 2, the visual comparison further demonstrates that the images produced by our model showcase smaller aberrations against ground truth. This is corroborated by the mean absolute error (MAE) map between the fused image and the ground truth, where the magnified residual regions exhibit fewer bright spots. These visual observations further support the quantitative results in Tabs. 1 and 2.

Table 1. Quantitative comparison of our BNNPan with representative full-precision and binary methods on the GaoFen-2 dataset.

Category	Method	Reduced resolution				Full resolution		
		SAM(± std)	ERGAS(± std)	Q2n(± std)	PSNR(± std)	D_λ(± std)	D_s(± std)	HQNR(± std)
Full	PNN [10]	1.0477 ± 0.2264	1.0572 ± 0.2355	0.9604 ± 0.0100	39.0712 ± 2.2927	0.0317 ± 0.0286	0.0943 ± 0.0224	0.8771 ± 0.0363
	MSDCNN [20]	1.0472 ± 0.2210	1.0413 ± 0.2309	0.9612 ± 0.0108	39.2216 ± 2.2275	0.0243 ± 0.0133	0.0730 ± 0.0093	0.9044 ± 0.0126
	GPPNN [19]	0.7972 ± 0.1605	0.7107 ± 0.1296	0.9791 ± 0.0080	42.4459 ± 1.7997	0.0229 ± 0.0119	0.0665 ± 0.0091	0.9122 ± 0.0139
	FusionNet [1]	0.9735 ± 0.2117	0.9878 ± 0.2222	0.9641 ± 0.0093	39.6386 ± 2.2701	0.0350 ± 0.0124	0.1013 ± 0.0134	0.8673 ± 0.0179
	LAGConv [7]	0.7859 ± 0.1478	0.6869 ± 0.1125	0.9804 ± 0.0085	42.7348 ± 1.4469	0.0284 ± 0.0130	0.0792 ± 0.0136	0.8947 ± 0.0200
	HFEAN [17]	0.7424 ± 0.1474	0.6574 ± 0.1174	0.9818 ± 0.0073	43.0458 ± 1.6412	0.0235 ± 0.0129	0.0456 ± 0.0123	0.9319 ± 0.0165
	HFIN [13]	0.8427 ± 0.1475	0.7347 ± 0.1260	0.9774 ± 0.0113	42.1889 ± 1.7517	0.0272 ± 0.0197	0.0620 ± 0.0093	0.9124 ± 0.0176
Binary	BNN [4]	1.1014 ± 0.2171	1.1144 ± 0.2249	0.9552 ± 0.0137	38.4899 ± 2.1089	0.0295 ± 0.0197	0.0779 ± 0.0134	0.8947 ± 0.0141
	IRNet [11]	1.5226 ± 0.2913	1.6186 ± 0.2968	0.9055 ± 0.0235	35.2960 ± 1.9704	0.0323 ± 0.0281	0.0495 ± 0.0141	0.9196 ± 0.0224
	ReActNet [9]	1.0605 ± 0.2261	1.0885 ± 0.2513	0.9577 ± 0.0128	38.7711 ± 2.2746	0.0314 ± 0.0135	0.0936 ± 0.0128	0.8779 ± 0.0159
	BTM [5]	1.0068 ± 0.2023	0.9620 ± 0.1901	0.9649 ± 0.0114	39.8011 ± 2.0545	0.0292 ± 0.0122	0.0890 ± 0.0124	0.8844 ± 0.0144
	E2FIF [12]	1.2212 ± 0.2045	1.1387 ± 0.2051	0.9496 ± 0.0157	38.2484 ± 1.7581	0.0392 ± 0.0232	0.0997 ± 0.0143	0.8649 ± 0.0200
	FABNet [6]	0.9805 ± 0.2154	0.9623 ± 0.2126	0.9654 ± 0.0115	39.8453 ± 2.2207	**0.0255 ± 0.0146**	0.0805 ± 0.0155	0.8960 ± 0.0180
	BBCU [18]	1.0167 ± 0.2018	0.9710 ± 0.1938	0.9650 ± 0.0109	39.7648 ± 2.0492	0.0288 ± 0.0155	0.0843 ± 0.0123	0.8892 ± 0.0132
	Ours	**0.8334 ± 0.1656**	**0.7564 ± 0.1288**	**0.9767 ± 0.0088**	**41.8447 ± 1.6214**	0.0259 ± 0.0131	**0.0494 ± 0.0134**	**0.9259 ± 0.0190**

Table 2. Quantitative comparison of our BNNPan with representative full-precision and binary methods on the QuickBird dataset.

Category	Method	Reduced resolution				Full resolution		
		SAM(± std)	ERGAS(± std)	Q2n(± std)	PSNR(± std)	D_λ(± std)	D_s(± std)	HQNR(± std)
Full	PNN [10]	5.2054 ± 0.9625	4.4722 ± 0.3734	0.9180 ± 0.0938	36.9343 ± 2.5364	0.0577 ± 0.0110	0.0624 ± 0.0239	0.8837 ± 0.0304
	MSDCNN [20]	5.1397 ± 0.9604	4.3581 ± 0.3647	0.9210 ± 0.0926	37.1482 ± 2.5378	0.0602 ± 0.0150	0.0667 ± 0.0289	0.8774 ± 0.0388
	GPPNN [19]	4.4577 ± 0.8060	3.6656 ± 0.3354	0.9360 ± 0.0860	38.5663 ± 2.4525	0.0519 ± 0.0159	0.0366 ± 0.0080	0.9134 ± 0.0190
	FusionNet [1]	4.9226 ± 0.9077	4.1594 ± 0.3212	0.9252 ± 0.0902	37.5317 ± 2.5184	0.0572 ± 0.0182	0.0522 ± 0.0088	0.8936 ± 0.0213
	LAGConv [7]	4.5473 ± 0.8296	3.8259 ± 0.4196	0.9335 ± 0.0878	38.1813 ± 2.4563	0.0859 ± 0.0237	0.0676 ± 0.0136	0.8522 ± 0.0178
	HFEAN [17]	4.5747 ± 0.8335	3.8058 ± 0.3582	0.9353 ± 0.0827	38.2314 ± 2.4929	0.1034 ± 0.0264	0.1059 ± 0.0242	0.8014 ± 0.0248
	HFIN [13]	4.5416 ± 0.8051	3.8131 ± 0.3217	0.9344 ± 0.0845	38.2465 ± 2.4028	0.0666 ± 0.0252	0.0784 ± 0.0185	0.8600 ± 0.0183
Binary	BNN [4]	5.3635 ± 0.9800	4.8212 ± 0.4260	0.9067 ± 0.1078	36.3151 ± 2.5303	0.0764 ± 0.0172	0.1130 ± 0.0227	0.8195 ± 0.0344
	IRNet [11]	6.4167 ± 1.3354	6.5089 ± 0.6417	0.8609 ± 0.1049	33.7823 ± 3.1119	0.0745 ± 0.0178	0.1053 ± 0.0175	0.8283 ± 0.0304
	ReActNet [9]	5.1745 ± 0.9511	4.5900 ± 0.3928	0.9136 ± 0.1024	36.7038 ± 2.5174	0.0965 ± 0.0176	0.0884 ± 0.0286	0.8240 ± 0.0379
	BTM [5]	5.0541 ± 0.9373	4.3799 ± 0.3382	0.9191 ± 0.1001	37.0801 ± 2.4970	0.0818 ± 0.0193	0.0902 ± 0.0248	0.8358 ± 0.0382
	E2FIF [12]	5.5124 ± 0.9356	4.7499 ± 0.4039	0.9083 ± 0.1008	36.4312 ± 2.3098	0.0907 ± 0.0221	0.1197 ± 0.0314	0.8011 ± 0.0458
	FABNet [6]	5.0681 ± 0.9586	4.2521 ± 0.3350	0.9225 ± 0.0933	37.3476 ± 2.6041	**0.0738 ± 0.0145**	0.0878 ± 0.0280	0.8452 ± 0.0375
	BBCU [18]	4.9388 ± 0.9053	4.2079 ± 0.3273	0.9231 ± 0.0933	37.4267 ± 2.5812	0.0752 ± 0.0168	0.0892 ± 0.0244	0.8427 ± 0.0363
	Ours	**4.8260 ± 0.8860**	**3.9713 ± 0.3620**	**0.9294 ± 0.0899**	**37.9446 ± 2.4301**	0.0761 ± 0.0207	**0.0661 ± 0.0136**	**0.8627 ± 0.0188**

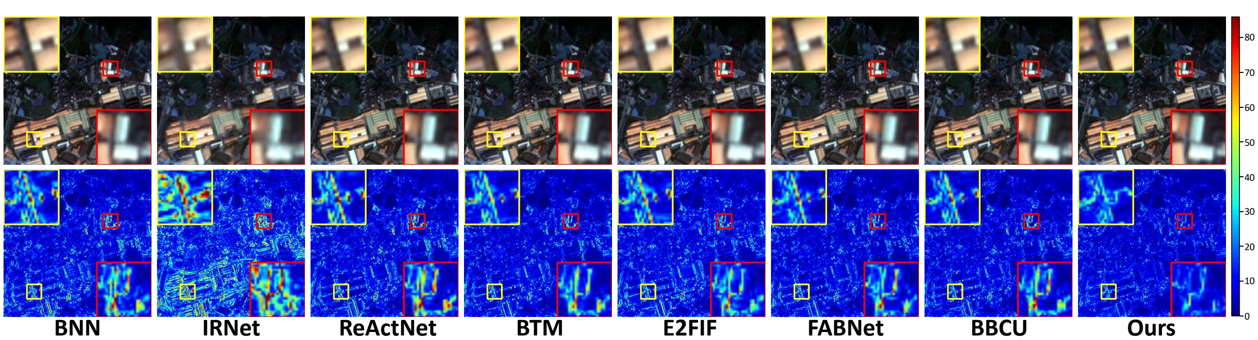

Figure 1. Qualitative comparison of our BNNPan with representative binary models on a reduced-resolution GF2 sample.

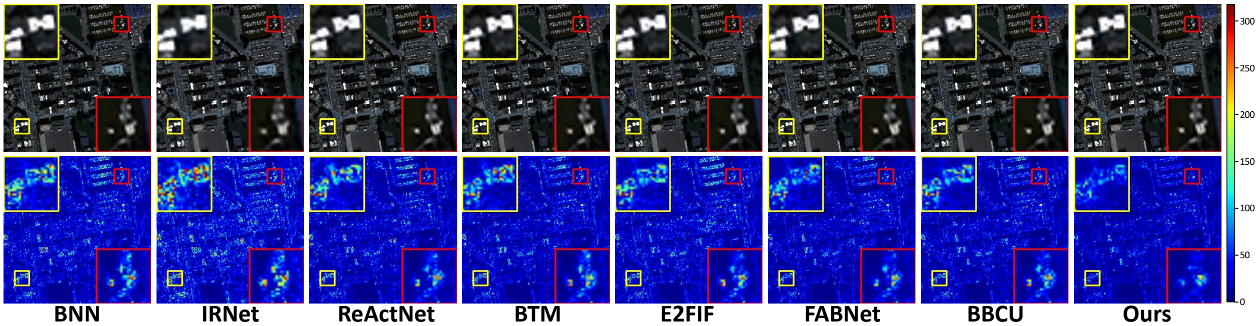

Figure 2. Qualitative comparison of our BNNPan with representative binary models on a reduced-resolution QB sample.

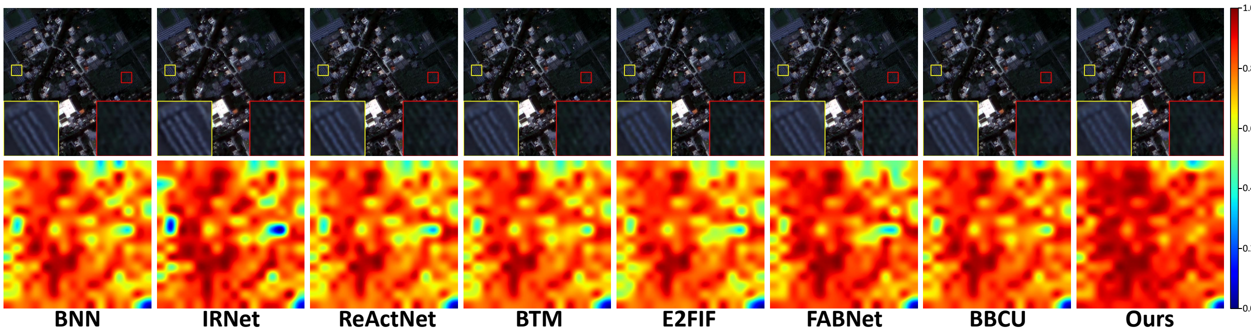

Figure 3. Qualitative comparison of our BNNPan with representative binary models on a full-resolution GF2 sample.

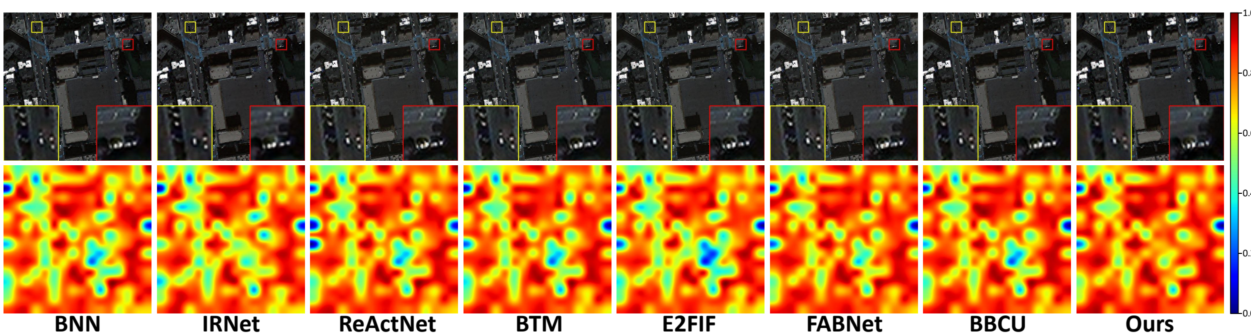

Figure 4. Qualitative comparison of our BNNPan with representative binary models on a full-resolution QB sample.

3.2. Evaluation on Full-resolution Scene

For the full-resolution evaluation, our model consistently delivers competitive performance across both the GF2 and QB datasets. As shown in the right panels of Tabs. 1 and 2, BNNPan leads in most metrics, yielding the highest HQNR scores across both datasets. While FABNet exhibits a marginal advantage in D_λ, our model achieves the optimal performance across all other evaluation metrics. Notably, BNNPan outperforms almost all full-precision methods in HQNR by a substantial margin on the GF2 dataset, whereas it surpasses the majority of full-precision approaches on the QB dataset. These results highlight BNNPan's exceptional fusion capability and competitive edge over both binary and full-precision models. Figs. 3 and 4 present the visual comparisons of various binary models on full-resolution GF2 and QB samples, respectively. It can be seen that our method produces images with rich details and appealing spectra, as shown in the first row of Figs. 3 and 4. Furthermore, the second row of Figs. 3 and 4 visualizes the corresponding HQNR maps, where the outcome of our method exhibits fewer bright spots, indicating higher values. These observations further substantiate the quantitative results shown in the right panels of Tabs. 1 and 2. Overall, these superior full-resolution results, encompassing both quantitative and qualitative aspects, robustly validate the effectiveness and adaptability of our model in real-world high-resolution scenarios.

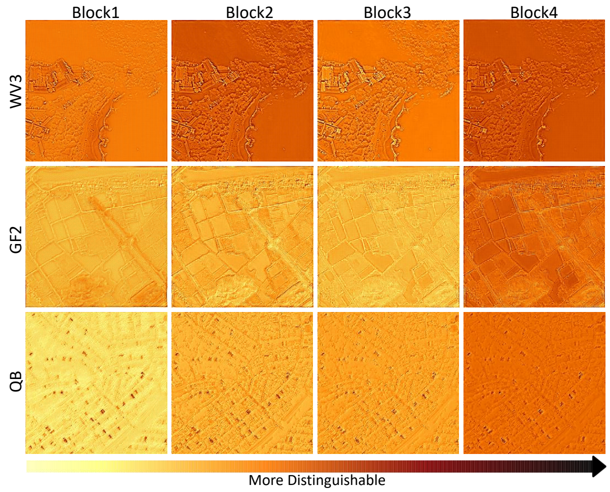

Figure 5. Feature visualization of different PIBF modules.

3.3. Feature Visualization

We further visualize the feature maps across different PIBF modules using samples from three datasets. As depicted in Fig. 5, these visualizations reveal a gradual improvement in clarity and detail information as the number of PIBF modules increases, emphasizing the model's representational capacity.

References

[1] Liang-Jian Deng, Gemine Vivone, Cheng Jin, and Jocelyn Chanussot. Detail injection-based deep convolutional neural networks for pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 59(8):6995–7010, 2020. 1, 2

[2] Liang-Jian Deng, Gemine Vivone, Mercedes E Paoletti, Giuseppe Scarpa, Jiang He, Yongjun Zhang, Jocelyn Chanussot, and Antonio Plaza. Machine learning in pansharpening: A benchmark, from shallow to deep networks. *IEEE Geoscience and Remote Sensing Magazine*, 10(3): 279–315, 2022. 1

[3] Andrea Garzelli and Filippo Nencini. Hypercomplex quality assessment of multi/hyperspectral images. *IEEE Geosci. Remote Sens. Lett.*, 6(4):662–665, 2009. 1

[4] Itay Hubara, Matthieu Courbariaux, Daniel Soudry, Ran El-Yaniv, and Yoshua Bengio. Binarized neural networks. *Advances in neural information processing systems*, 29, 2016. 1, 2

[5] Xinrui Jiang, Nannan Wang, Jingwei Xin, Keyu Li, Xi Yang, and Xinbo Gao. Training binary neural network without batch normalization for image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1700–1707, 2021. 1, 2

[6] Xinrui Jiang, Nannan Wang, Jingwei Xin, Keyu Li, Xi Yang, Jie Li, Xiaoyu Wang, and Xinbo Gao. Fabnet: Frequency-aware binarized network for single image super-resolution. *IEEE Transactions on Image Processing*, 32:6234–6247, 2023. 1, 2

[7] Zi-Rong Jin, Tian-Jing Zhang, Tai-Xiang Jiang, Gemine Vivone, and Liang-Jian Deng. Lagconv: Local-context adaptive convolution kernels with global harmonic bias for pansharpening. In *Proceedings of the AAAI conference on artificial intelligence*, pages 1113–1121, 2022. 1, 2

[8] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. 1

[9] Zechun Liu, Zhiqiang Shen, Marios Savvides, and Kwang-Ting Cheng. Reactnet: Towards precise binary neural network with generalized activation functions. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*, pages 143–159. Springer, 2020. 1, 2

[10] Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva, and Giuseppe Scarpa. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7):594, 2016. 1, 2

[11] Haotong Qin, Ruihao Gong, Xianglong Liu, Mingzhu Shen, Ziran Wei, Fengwei Yu, and Jingkuan Song. Forward and backward information retention for accurate binary neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2250–2259, 2020. 1, 2

[12] Chongxing Song, Zhiqiang Lang, Wei Wei, and Lei Zhang. E2fif: Push the limit of binarized deep imagery super-resolution using end-to-end full-precision information flow. *IEEE Transactions on Image Processing*, 2023. 1, 2

[13] Jiangtong Tan, Jie Huang, Naishan Zheng, Man Zhou, Keyu Yan, Danfeng Hong, and Feng Zhao. 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*, pages 25922–25931, 2024. 1, 2

[14] Gemine Vivone, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio A Licciardi, Rocco Restaino, and Lucien Wald. A critical comparison among pansharpening algorithms. *IEEE Trans. Geosci. Remote Sens.*, 53(5):2565–2586, 2014. 1

[15] Lucien Wald. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002. 1

[16] Lucien Wald, Thierry Ranchin, and Marc Mangolini. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric engineering and remote sensing*, 63(6):691–699, 1997. 1

[17] Yingying Wang, Yunlong Lin, Ge Meng, Zhenqi Fu, Yuhang Dong, Linyu Fan, Hedeng Yu, Xinghao Ding, and Yue Huang. Learning high-frequency feature enhancement and alignment for pan-sharpening. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 358–367, 2023. 1, 2

[18] Bin Xia, Yulun Zhang, Yitong Wang, Yapeng Tian, Wenming Yang, Radu Timofte, and Luc Van Gool. Basic binary convolution unit for binarized image restoration network. In *The Eleventh International Conference on Learning Representations*, 2023. 1, 2

[19] Shuang Xu, Jiangshe Zhang, Zixiang Zhao, Kai Sun, Junmin Liu, and Chunxia Zhang. Deep gradient projection networks

for pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1366–1375, 2021. 1, 2

[20] Qiangqiang Yuan, Yancong Wei, Xiangchao Meng, Huanfeng Shen, and Liangpei Zhang. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):978–989, 2018. 1, 2

[21] Roberta H Yuhas, Alexander FH Goetz, and Joe W Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop*, 1992. 1