DDA: A DUAL-DOMAIN ATTENTION PLUG-AND-PLAY PRIOR FOR PANSHARPENING

Wen-Jie Shu[1], Zi-En Zhang[1] *
[1]University of Electronic Science and Technology of China
wenjieshu2003@gmail.com, zienzhang@std.uestc.edu.cn

ABSTRACT

Pansharpening is an image processing technique that enhances spatial resolution of multispectral images by fusing them with higher-resolution panchromatic images, becoming increasingly critical for remote sensing and geospatial analysis applications. Despite advancements, current deep learning algorithms for pansharpening face limitations: lack of global information extraction in the spatial domain and insufficient interaction across spectral channels. To tackle these challenges, we propose DDA, a dual-domain attention plug-and-play prior, integrating transformer attention with Convolutional Neural Networks, to facilitate better spatial and spectral detail integration. The code is available at https://github.com/zennnnnnnnnnnn/DDANet.

1 INTRODUCTION AND RELATED WORK

Pansharpening merges spatial enhancement with spectral preservation. It combines image fusion and hyperspectral super-resolution2, aiming to retain spectral integrity while enhancing spatial detail.

Deep learning has significantly impacted pansharpening, offering robust solutions for remote sensing data. Traditional CNN-based methods Vivone (2019) Choi et al. (2010) Vivone et al. (2018) have achieved great performance due to the superior capabilities of DL in feature extraction and nonlinear fitting. The pioneering PNN, with its convolutional architecture, set the stage for subsequent innovations. Models such as PanNetYang et al. (2017), DiCNNHe et al. (2019), and FusionNetDeng et al. (2020) have built upon this foundation, each enhancing the deep learning framework's application in pansharpening. Transformer models in pansharpening address CNN limitations but require extensive computational resources and large datasets, challenging remote sensing's data-scarce domain. Addressing these shortcomings, our research introduces a dual-domain plug-and-play module that synergizes spatial and spectral attention mechanismsVaswani et al. (2017)Dosovitskiy et al. (2020). Preceding transformer modules, due to their architecture, incurred substantial computational costsLiang et al. (2021). Our module combines CNNs and attention blocks for efficiency, capturing global context. This dual-domain approach balances spatial clarity and material characterization.

2 METHOD

The proposed Dual-Domain Attention (DDA) module, a plug-and-play component for CNNs, comprises the High-Resolution Spectral Attention (HRSA) and High-Resolution Spatial Attention (HRSpA) branches. Motivated by CvTWu et al. (2021), we incorporate convolution into the transformer, which can reduce parameters and improve computational speed. The Dual-Domain Attention (DDA) module for CNNs features two branches: High-Resolution Spectral Attention (HRSA) and High-Resolution Spatial Attention (HRSpA). HRSA processes input (C×H×W) through convolution, halving channel dimensions, and uses softmax to generate attention weights, applied to doubled-channel input, creating a C×1×1 spectral attention map1. HRSpA, also starting with C×H×W input, produces a spatial attention map after realignment and sigmoid function application2. The final output combines both attention maps with the input, through element-wise multiplication and summation3.

$$SA(X) = \sigma(\text{fc}(\text{Softmax}(\text{fc}(X)))) \tag{1}$$

*Corresponding author

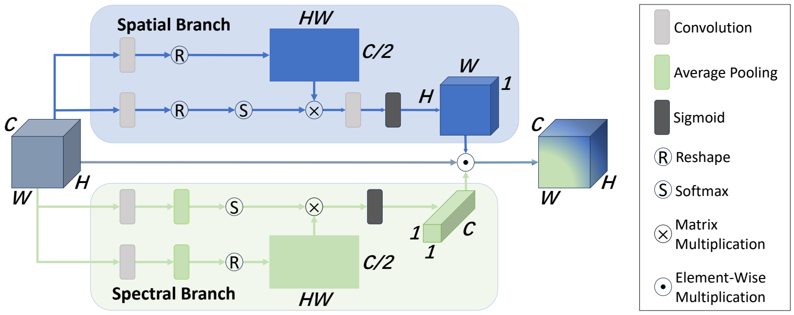

Figure 1: Overall Structure of the Proposed Method. The DDA module extracts spectral and spatial attention features separately and multiplies them with the input data.

$$SPA(X) = \sigma(\text{Reshape}(\text{Softmax}(\text{GAP}(\text{fs}(X))) \times \text{fs}(X))) \tag{2}$$

$$Y = X \times SA(X) \times SPA(X) \tag{3}$$

Here, σ denotes the sigmoid function, Softmax denotes the softmax function, GAP represents global average pooling, and $\times$ is element-wise multiplication. fc, fs represent convolution operation for spatial and spectral branch.

3 EXPERIMENT

Table 1: Quantitative results on 20 reduced-resolution samples of WV3. (red: best)

Method	**Reduced-Resolution**				**Params**
	PSNR	Q8	SAM	ERGAS	
PanNet	37.381±2.643	0.901±0.092	3.624±0.695	2.641±0.605	0.60MB
PanNet+DDA	38.014±2.541	0.908±0.092	3.328±0.622	2.440±0.614	0.61MB
MSDCNN	37.152±2.576	0.900±0.090	3.707±0.758	2.719±0.640	0.87MB
MSDCNN+DDA	37.371±2.713	0.903±0.090	3.580±0.668	2.666±0.677	0.88MB
FusionNet	37.647±2.601	0.903±0.091	3.388±0.657	2.544±0.615	0.58MB
FusionNet+DDA	37.834±2.564	0.906±0.090	3.317±0.643	2.480±0.632	0.60MB
LAGNet	38.584±2.519	0.916±0.087	3.129±0.642	2.297±0.593	0.58MB
LAGNet+DDA	38.666±2.637	0.918±0.086	3.085±0.576	2.261±0.565	0.59MB
Ideal value	$+\infty$	1	0	0	

We inserted our DDA module into recent SOTA works, including PanNetYang et al. (2017), MSDCNNYuan et al. (2018), FusionNetDeng et al. (2020), LAGNetJin et al. (2022). We insert DDA into the adjacent layers of the network backbone, such as in PanNet, where we insert it between two ResBlocks. The test results are presented in Table 1. The insertion of the DDA module resulted in a significant improvement across all quality metrics for the integrated networks. Due to space constraints, we have placed the details of the experimental setup and visual results in the appendix.

4 CONCLUSION

In summary, our study advances pansharpening by overcoming limitations in information extraction and spectral interaction in deep learning. We introduced DDA, a novel and efficient module that synergistically combines the global attention capabilities of transformers with the structural advantages of CNNs. This dual-domain module has exhibited exceptional performance in enhancing both the spatial and spectral quality of pansharpened images, as verified through comprehensive evaluations. Our open-source code contributes significantly to remote sensing, setting a benchmark in image fusion and hyperspectral super-resolution.

URM Statement

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

References

Jaewan Choi, Kiyun Yu, and Yongil Kim. A new adaptive component-substitution-based satellite image fusion by using partial replacement. *IEEE transactions on geoscience and remote sensing*, 49(1):295–309, 2010.

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.

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. doi: 10.1109/MGRS.2022.3187652.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Andrea Garzelli and Filippo Nencini. Hypercomplex quality assessment of multi/hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 6(4):662–665, 2009. doi: 10.1109/LGRS.2009.2022650.

Lin He, Yizhou Rao, Jun Li, Jocelyn Chanussot, Antonio Plaza, Jiawei Zhu, and Bo Li. Pansharpening via detail injection based convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(4):1188–1204, 2019.

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*, volume 36, pp. 1113–1121, 2022.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1833–1844, 2021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Gemine Vivone. Robust band-dependent spatial-detail approaches for panchromatic sharpening. *IEEE transactions on Geoscience and Remote Sensing*, 57(9), 2019.

Gemine Vivone, Rocco Restaino, and Jocelyn Chanussot. Full scale regression-based injection coefficients for panchromatic sharpening. *IEEE Transactions on Image Processing*, 27(7):3418–3431, 2018.

Lucien Wald. Data fusion. definitions and architectures - fusion of images of different spatial resolutions. 2010.

Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. Cvt: Introducing convolutions to vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 22–31, 2021.

Junfeng Yang, Xueyang Fu, Yuwen Hu, Yue Huang, Xinghao Ding, and John Paisley. Pannet: A deep network architecture for pan-sharpening. In *Proceedings of the IEEE international conference on computer vision*, pp. 5449–5457, 2017.

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.

A APPENDIX

A.1 SCHEMATIC DIAGRAM

Figure 2: schematic diagram of hyperspectral pansharpening

A.2 EXPERIMENT SETTINGS

Datasets. We conducted experiments on the WV3 dataset using data exclusively sourced from PanCollectionDeng et al. (2022), which comprises a total of 10,000 samples. Each sample is composed of a set of PAN/LRMS/GT images.The testing dataset is categorized into two types: a reduced-resolution dataset(8x256x256) and a full-resolution dataset(8x512x512). We tested on the reduced-resolution dataset.

Evaluation Metrics. For the evaluation of the reduced-resolution dataset, we employ four metrics: Peak Signal-to-Noise Ratio (PSNR), Quality Index (Q8)Garzelli & Nencini (2009), Spectral Angle Mapper (SAM), and the Error Relative to Global Amplitude of the Spectra (ERGAS)Wald (2010).

Through experimentation, our network demonstrated its suitability for integration into the deep intermediate region where spectral dual-domain information of PAN and MS is highly fused. Experimental results can be referenced on our GitHub repository.

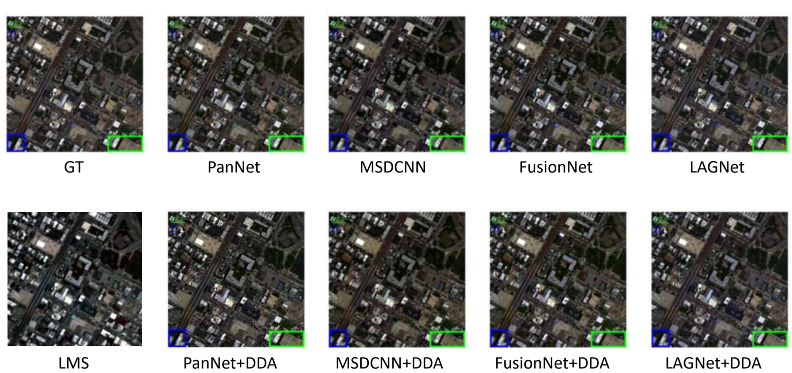

Figure 3: Visual results on the reduced-resolution sample from WV3 dataset