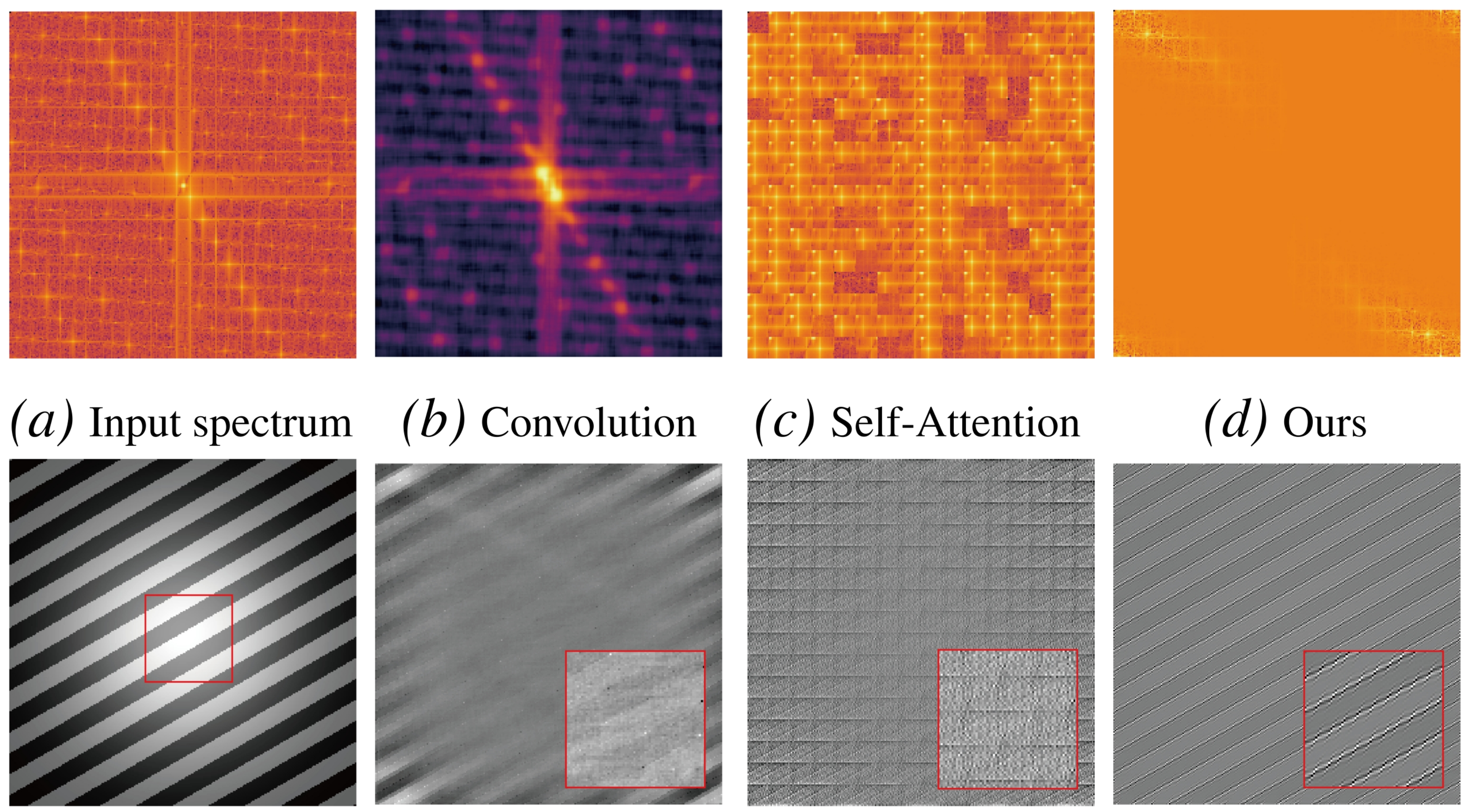
Butterworth as Attention: Anisotropic Spectral Gating for Pansharpening

Zhenggang Wang Wu Wang Huazhe Liang Tai-xiang Jiang

1. Background & Motivation



(a) Input spectrum (b) Convolution (c) Self-Attention (d) Ours

(e) Original image (f) Convolution (g) Self-Attention (h) Ours

Existing FFT-based methods naively transplant spatial operators (convolution or self-attention) into the frequency domain, ignoring its fundamental structure:

- Convolution assumes translation invariance → causes ringing artifacts.
- Self-attention splits the continuous spectrum into patches → destroys spectral continuity, introducing grid artifacts.

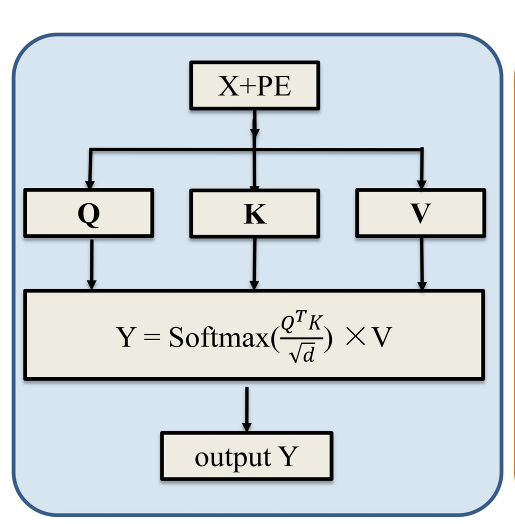
2. Butterworth as Attention

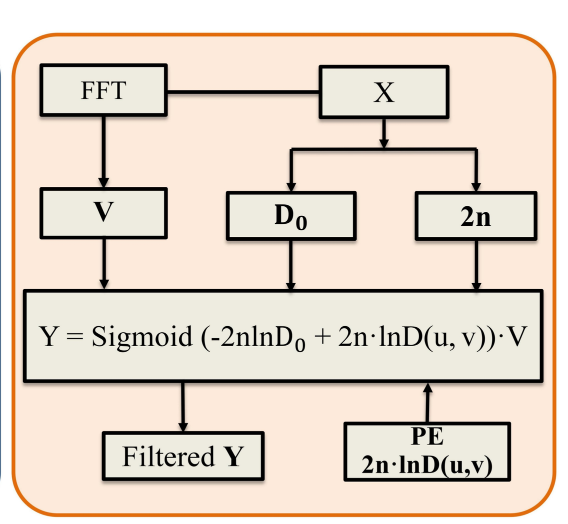
- **Reformulating the Butterworth Filter.** The classical Butterworth low-pass filter is defined by its transfer function in the frequency domain:

$$H(u,v) = \frac{1}{1+\left(\frac{D(u,v)}{D_0}\right)^{2n}},$$

Applying the identity x = eln x and rearranging reveals a Sigmoid structure:

$$H(u,v) = \sigma\left(-2n \cdot \ln D_0 + 2n \cdot \ln D(u,v)\right).$$





Query → D_0 (cutoff frequency)

Key → 2n (filter order)

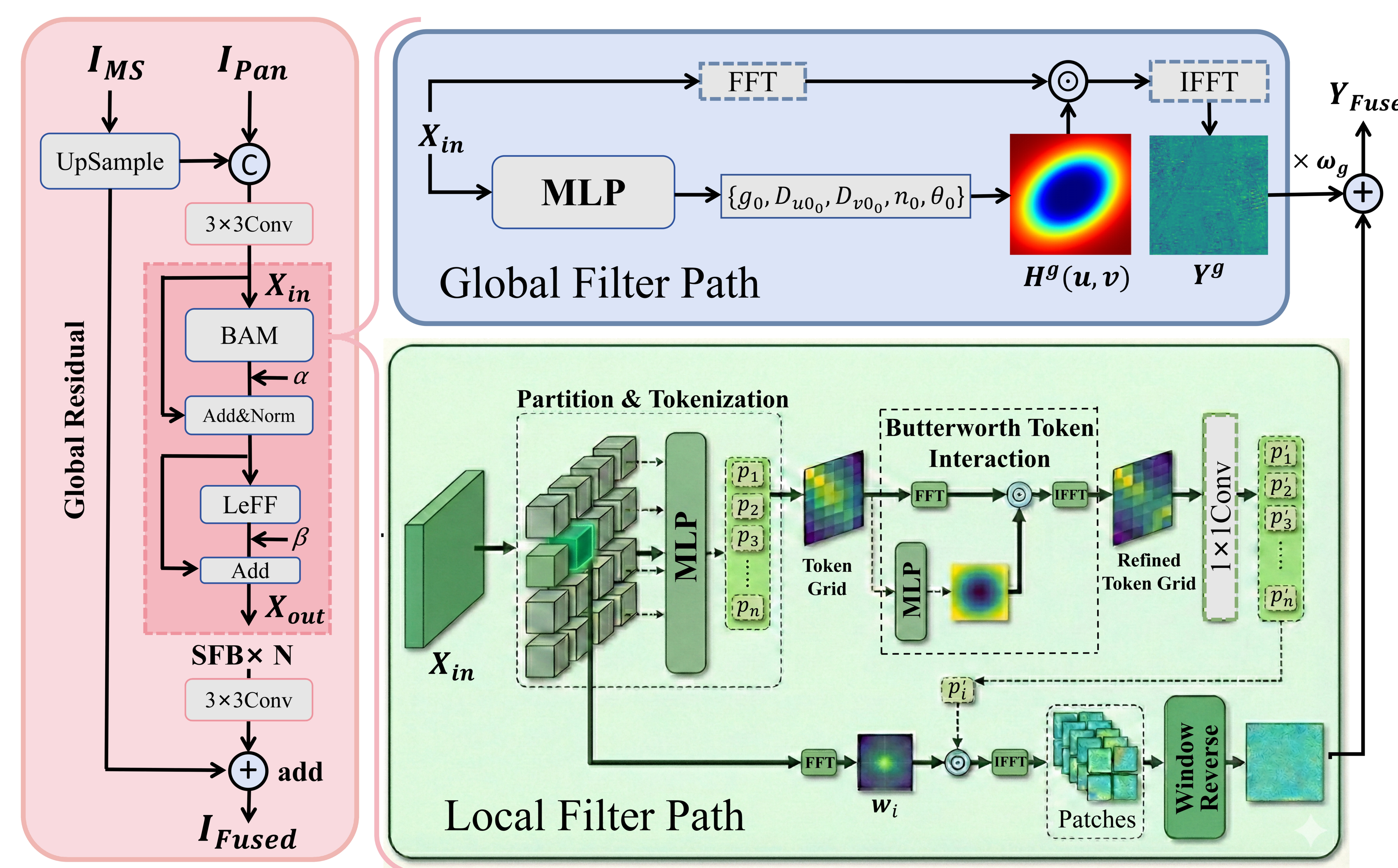
Value → Spectrum X (FFT)

Positional Encoding → 2n·ln D(u,v)

- **From Isotropic → Anisotropic.** The standard Butterworth filter has a circular (isotropic) passband — treating all orientations equally. We generalize via a 2D rotation:

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \cdot \ \mathcal{R}(u',v') = \sqrt{\left(\frac{u'}{D_{u0}}\right)^2 + \left(\frac{v'}{D_{v0}}\right)^2}.$$

3.Method: Anisotropic Butterworth Fusion Network



Overall pipeline

- LMS + PAN → concatenate → 3x3 Conv → N stacked SFBs → HRMS
- Each SFB = BAM (replaces self-attention) + LeFF (locally-enhanced FFN)

Dual-branch BAM

- Global path： One anisotropic Butterworth mask over the full feature map — captures holistic structure
- Local path： Patch-level filtering guided by BTI — models regional directional textures via O(N log N) token interaction

Outputs fused via learnable per-channel weight wg

Key advantages

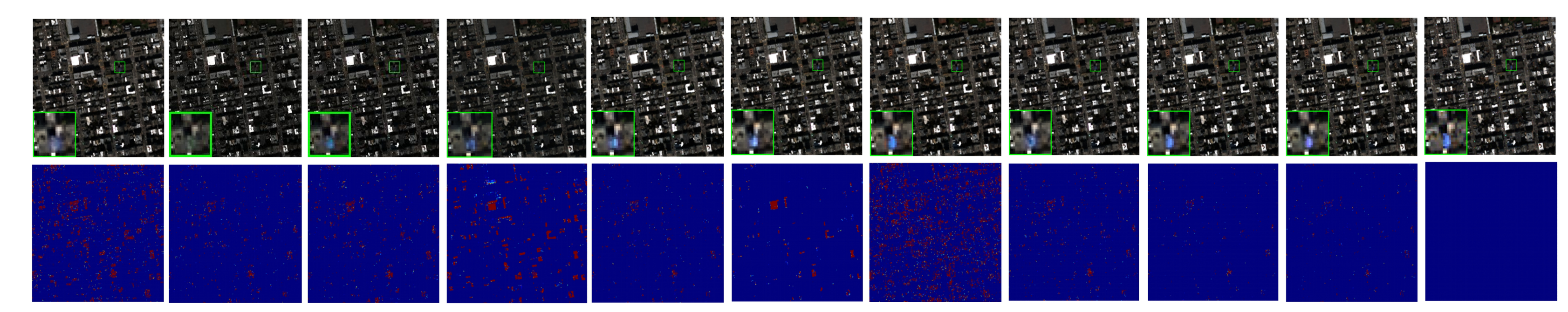
- Global receptive field — no local window limitation
- Direction-aware filtering — adapts to dominant texture orientation
- O(N log N) complexity — cheaper than standard self-attention O(N²)

4. Experiments & Results

- **Achieve SOTA on all three datasets,**outperform competitors in three metrics, using only 0.26M parameters and 12.64G FLOPs. Report inference latency and peak GPU memory on RTX 5060 Ti.

Method	WV3 Reduce			QB Reduce			GF2 Reduce			WV2 Zero-Shot			Params (M)↓	FLOPs (G)↓
	PSNR↑	ERGAS↓	SAM↓	PSNR↑	ERGAS↓	SAM↓	PSNR↑	ERGAS↓	SAM↓	PSNR↑	SAM↓	ERGAS↓		
SFIIN (Zhou et al., 2022)	34.29	3.01	4.31	36.00	4.74	5.49	47.32	1.03	1.21	28.04	6.01	4.86	0.09	5.37
TDNet (Zhang et al., 2023c)	36.03	2.54	3.64	37.17	4.15	4.58	47.27	1.02	1.23	26.68	6.04	9.23	0.49	20.00
BIMPan (Hou et al., 2023)	36.95	2.30	3.44	36.14	4.67	4.77	45.92	1.17	1.35	**29.58**	5.43	**4.03**	2.34	29.45
DISPnet (Wang et al., 2024a)	36.80	2.33	3.42	36.83	4.33	4.47	48.76	0.86	1.02	28.46	5.81	4.61	0.33	25.60
DCINN (Wang et al., 2024b)	37.22	2.24	3.30	38.43	3.57	4.22	49.28	0.81	0.98	27.93	5.82	4.94	0.49	13.62
WFAnet (Huang et al., 2025)	36.60	2.39	3.46	38.48	3.54	4.63	48.56	0.88	1.05	28.08	6.06	4.87	0.53	14.33
ARConv (Wang et al., 2025)	36.58	2.38	3.41	37.67	3.90	4.99	45.45	1.23	1.43	28.61	5.58	4.51	4.79	61.57
BNNPan (Hou et al., 2025)	36.00	2.54	3.63	37.66	3.90	5.07	40.77	1.09	1.26	27.43	6.47	5.16	0.25	15.28
Ours Only Global	36.71	2.36	3.40	38.36	3.59	4.19	48.78	0.85	1.00	--	--	--	**0.07**	**3.58**
Ours Full	**37.65**	**2.13**	**3.17**	**38.83**	**3.41**	**4.06**	**50.30**	**0.72**	**0.86**	29.05	**5.37**	4.29	0.26	12.64

- **Produces sharper edges and cleaner textures.** Residual maps show the lowest error intensity among all compared methods.



SDFI TDNet DISPNet BIMPan DCINN WFAnet ARConv BNNPan Ours Full Only Global GT

- The BAM backbone achieves 81.31% Top-1 accuracy with fewer parameters than most baselines in classification

Method	Params (M)↓	FLOPs (G)↓	Acc (%)↑
ViT-Base (Dosovitskiy et al., 2021)	5.36	**0.35**	69.00
ConvNeXt-Tiny (Liu et al., 2022)	**5.17**	0.75	79.44
GFNet (Rao et al., 2021)	6.26	0.80	80.30
DINO-V3 (Siméoni et al., 2025)	6.52	0.44	72.63
VMamba-V2 (Liu et al., 2024)	10.42	0.66	72.28
UniRepLK (Ding et al., 2024)	6.12	0.62	75.11
MobileNet-V4 (Qin et al., 2024)	6.26	0.76	81.25
TransNext (Shi, 2024)	6.59	1.01	80.79
Ours	5.54	0.70	**81.31**

OPEN TO OPPORTUNITIES

Zhenggang Wang

Seeking PhD / RA · Computer Vision

✉ 42311051@smail.swufe.edu.cn

wangzg06100470@gmail.com

https://Gang-ww.github.io