Local-Global Transformer Enhanced Unfolding Network for Pan-sharpening

Mingsong Li[1], Yikun Liu[1], Tao Xiao[1], Yuwen Huang[2], Gongping Yang[1]

[1]School of Software, Shandong University, Jinan, China; [2]School of Computer, Heze University, Heze, China

IJCAI/2023 MACAO

Summary

- **Task description:** *Pan-sharpening* aims to increase the spatial resolution of the low-resolution multispectral (LrMS) image with the guidance of the corresponding panchromatic (PAN) image.
  - **Input:** The coupled LrMS image and PAN image.
  - **Output:** A high-resolution multispectral (HrMS) image.
- **Existing two-fold deficiency:**
  - Model Interpretability
  - Local and Global Dependencies
- **Our main contributions:**
  - We customize a transformer module LGT as an image denoiser to efficiently model local and global dependencies at the same time and sufficiently mine the potential of the proposed unfolding pan-sharpening framework.
  - We develop an interpretable transformer-based deep unfolding network, LGTEUN.
  - To the best of our knowledge, LGTEUN is the first transformer-based deep unfolding network for the MS pan-sharpening, and LGT is also the first transformer module to perform spatial and frequency dual-domain learning.
  - **Code:** https://github.com/ lms-07/LGTEUN

Proposed Method

- **Model formulation and optimization**

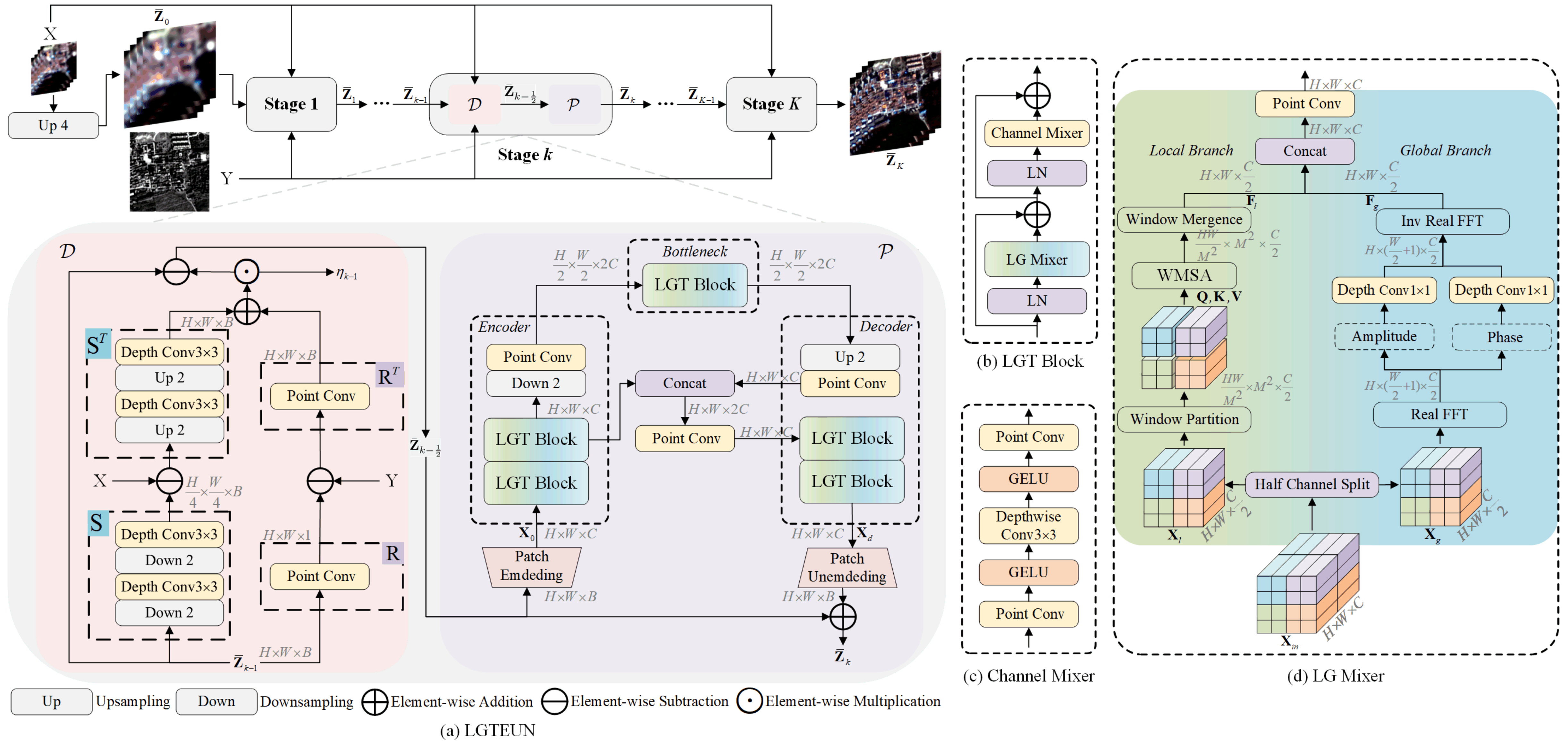
The degradation process of the HrMS image **Z**:

$$\mathbf{X} = \mathbf{S}\mathbf{Z} + \mathbf{N}_x,\ \ \mathbf{Y} = \mathbf{Z}\mathbf{R} + \mathbf{N}_y,$$

The energy function under MAP framework:

$$\bar{\mathbf{Z}} = \underset{\mathbf{Z}}{argmin}\frac{1}{2}\| \mathbf{X} - \mathbf{S}\mathbf{Z} \|^2 + \frac{1}{2}\| \mathbf{Y} - \mathbf{Z}\mathbf{R} \|^2 + \lambda J(\mathbf{Z}),$$

Fig: Illustration of the proposed LGTEUN.



Employing proximal gradient descent (PGD) algorithm:

$$\bar{\mathbf{Z}}_k = \underset{\mathbf{Z}}{argmin}\frac{1}{2}\| \mathbf{Z} - (\bar{\mathbf{Z}}_{k-1} - \eta\nabla_f(\bar{\mathbf{Z}}_{k-1}) \|^2 + \lambda J(\mathbf{Z}),$$

The data terms oriented differentiable operator:

$$\nabla_f(\bar{\mathbf{Z}}_{k-1}) = \mathbf{S}^T(\mathbf{S}\bar{\mathbf{Z}}_{k-1} - \mathbf{X}) + (\bar{\mathbf{Z}}_{k-1}\mathbf{R} - \mathbf{Y})\mathbf{R}^T.$$

The data subproblem and the prior subproblem:

$$\bar{\mathbf{Z}}_{k-\frac{1}{2}} = \bar{\mathbf{Z}}_{k-1} - \eta\nabla(\bar{\mathbf{Z}}_{k-1}),$$

$$\bar{\mathbf{Z}}_k = prox_{\eta,J}(\bar{\mathbf{Z}}_{k-\frac{1}{2}}),$$

- **Deep unfolding network**

Data module: $\bar{\mathbf{Z}}_{k-\frac{1}{2}} = \mathcal{D}(\bar{\mathbf{Z}}_{k-1}, \mathbf{X}, \mathbf{Y}, \eta_{k-1}).$

Prior module: $\bar{\mathbf{Z}}_k = \mathcal{P}(\bar{\mathbf{Z}}_{k-\frac{1}{2}}).$

- **Local-Global Transformer: LGT**
- **Local branch** calculates local window based self-attention in spatial domain.

$$\mathbf{F}_a^i = Softmax(\frac{\mathbf{Q}^i\mathbf{K}^{iT}}{\sqrt{d}} + \mathbf{P}^i)\mathbf{V}^i,\ \ i = 1, ..., h,$$

- **Global branch** extracts global contextual feature representation in frequency domain.

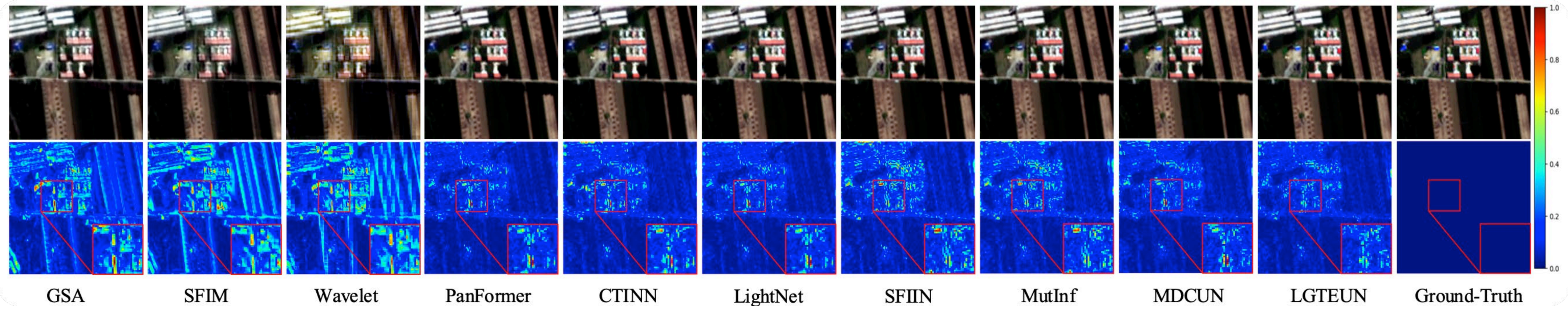
$$\mathcal{F}(\mathbf{X}_g)(u,v) = \frac{1}{\sqrt{HW}}\sum_{h=0}^{H-1}\sum_{w=0}^{W-1}\mathbf{X}_g(h,w)e^{-j2\pi(\frac{h}{H}u + \frac{w}{W}v)},$$

$$\mathcal{A}(\mathbf{X}_g)(u,v) = \sqrt{R^2(\mathbf{X}_g)(u,v) + I^2(\mathbf{X}_g)(u,v)},$$

$$\mathcal{P}(\mathbf{X}_g)(u,v) = arctan[\frac{I(\mathbf{X}_g)(u,v)}{R(\mathbf{X}_g)(u,v)}],$$

$$\mathbf{F}_g = \mathcal{F}^{-1}(DConv(\mathcal{A}(\mathbf{X}_g)), DConv(\mathcal{P}(\mathbf{X}_g))),$$

Fig: Qualitative comparison



Experiments

- **Quantitative comparison**

Method	WorldView-3					WorldView-2					GaoFen-2				
	PSNR↑	SSIM↑	Q8↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	Q4↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	Q4↑	SAM↓	ERGAS↓
GSA	22.5164	0.6343	0.5742	0.1106	7.8267	33.5975	0.8899	0.5681	0.0573	2.5402	36.0557	0.8838	0.5517	0.0641	3.5758
SFIM	21.4154	0.5415	0.4525	0.1147	8.8553	32.6334	0.8728	0.5159	0.0597	3.1919	34.7715	0.8572	0.4584	0.0657	4.2073
Wavelet	21.4464	0.5656	0.5271	0.1503	9.1545	32.1992	0.8500	0.4577	0.0638	3.3799	33.9208	0.8197	0.4033	0.0695	4.6445
PanFormer	30.4772	0.9368	0.9316	0.0672	3.1830	41.3581	0.9731	0.8236	0.0241	1.0617	44.8540	0.9805	0.8865	0.0271	1.3334
CTINN	31.8564	0.9518	0.9460	0.0660	2.7421	41.2015	0.9735	0.8149	0.0246	1.0880	44.2942	0.9784	0.8716	0.0293	1.4148
LightNet	32.0018	0.9525	0.9472	0.0639	2.6853	41.5589	0.9739	0.8220	0.0237	1.0382	44.6876	0.9787	0.8741	0.0279	1.3510
SFIIN	31.6587	0.9492	0.9435	0.0652	2.8016	41.9489	0.9752	0.8108	0.0229	1.0084	44.7248	0.9802	0.8721	0.0280	1.3361
MutInf	31.8298	0.9523	0.9469	0.0636	2.7526	41.9522	0.9760	0.8258	0.0227	1.0153	44.8305	0.9800	0.8836	0.0277	1.3394
MDCUN	31.2978	0.9429	0.9363	0.0661	2.9295	42.3351	0.9772	0.8370	0.0216	0.9638	45.5677	0.9825	0.8915	0.0252	1.2249
LGTEUN	32.2188	0.9545	0.9494	0.0605	2.6286	42.6837	0.9786	0.8415	0.0208	0.9280	45.8364	0.9840	0.8973	0.0247	1.1824

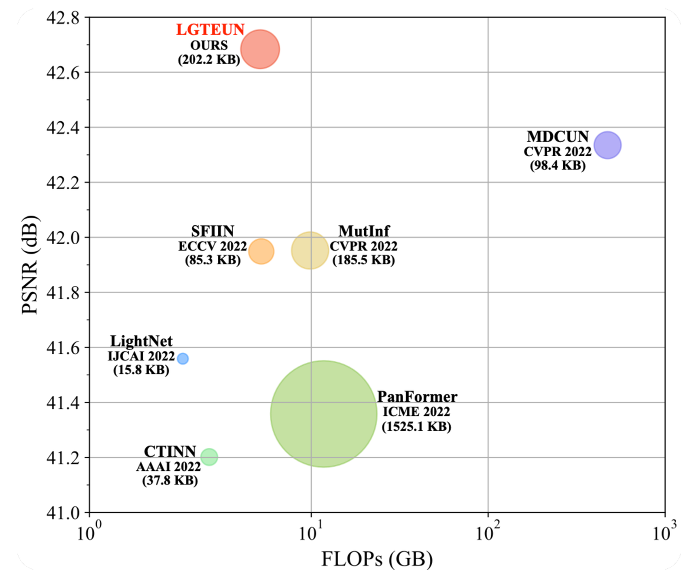
- **A key hyperparameter setting**

Data Set	Metric	Stage 1	Stage 2	Stage 3	Stage 4
WorldView-3	PSNR↑	32.0339	32.2188	32.068	32.0042
	SSIM↑	0.9532	0.9545	0.9535	0.9527
	Q8↑	0.9481	0.9494	0.9487	0.9480
	SAM↓	0.0605	0.0605	0.0603	0.0612
	ERGAS↓	2.6765	2.6286	2.6678	2.6898
	Time (s/img)	0.0070	0.0133	0.0205	0.0262
	Params (KB)	270.2	540.0	809.9	1079.7
	FLOPs (GB)	9.52	19.04	28.56	38.08
WorldView-2	PSNR↑	42.600	42.6837	42.4771	42.1634
	SSIM↑	0.9784	0.9786	0.9781	0.9767
	Q4↑	0.8398	0.8415	0.8383	0.8329
	SAM↓	0.0209	0.0208	0.0213	0.0222
	ERGAS↓	0.9358	0.928	0.9573	0.9787
	Time (s/img)	0.0065	0.0137	0.0204	0.0254
	Params (KB)	101.2	202.2	303.2	404.2
	FLOPs (GB)	2.57	5.14	7.71	10.28

- **Full-resolution test**

Method	Full-resolution Test		
	D_λ↓	D_S↓	QNR↑
GSA	0.0094	0.1076	0.8839
SFIM	0.0094	0.1061	0.8854
Wavelet	0.0552	0.1330	0.8193
PanFormer	0.0191	0.0416	0.9400
CTINN	0.0123	0.0442	0.9440
LightNet	0.0185	0.0282	0.9539
SFIIN	0.0198	0.0352	0.9457
MutInf	0.0163	0.0420	0.9423
MDCUN	0.0747	0.1673	0.7708
LGTEUN	0.0162	0.0310	0.9532

- **Performance-efficiency comparison**



- **Efficiency comparison**

Data Set	Metric	GSA	SFIM	Wavelet	PanFormer	CTINN	LightNet	SFIIN	MutInf	MDCUN	LGTEUN
WorldView-3	Time (s/img)	0.0482	0.0591	0.0562	0.0160	0.0426	0.0019	0.0529	0.1083	0.1747	0.0133
	Params (KB)	–	–	–	1532.8	38.3	16.3	85.8	185.8	140.9	540.0
	FLOPs (GB)	–	–	–	11.92	2.68	2.02	5.25	9.87	479.54	19.04
GaoFen-2	Time (s/img)	0.0216	0.0301	0.0271	0.0257	0.0431	0.0017	0.0528	0.1141	0.1017	0.0129
	Params (KB)	–	–	–	1530.3	37.8	15.8	85.3	185.5	98.3	202.2
	FLOPs (GB)	–	–	–	11.77	2.65	1.95	5.22	9.85	473.19	5.14

- **Ablation study**

Setting		Reduced-resolution Test					Full-resolution Test		
Local Branch	*Global Branch*	PSNR↑	SSIM↑	Q8↑	SAM↓	ERGAS↓	D_λ↓	D_S↓	QNR↑
✗	✓	31.9309	0.9519	0.9468	0.0636	2.7102	0.0177	0.0364	0.9465
✓	✗	31.9742	0.9525	0.9468	0.0618	2.7029	0.0170	0.0349	0.9486
✓	✓	32.2188	0.9545	0.9494	0.0605	2.6286	0.0162	0.0310	0.9532

Contact

Mail: msli@mail.sdu.edu.cn

Towards Intelligence Mechanism Lab

School of Software, Shandong University, China

Codebase **Homepage**