HyperTransformer: A Textural and Spectral Feature Fusion Transformer for Pansharpening

Wele Gedara Chaminda Bandara and Vishal M. Patel

Johns Hopkins University | Whiting School of Engineering | Baltimore, MD

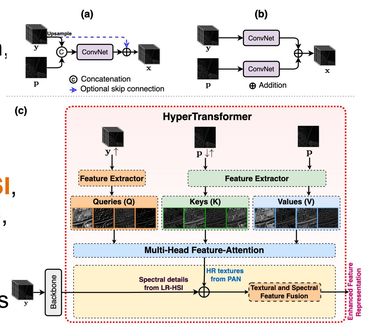
Problem Statement

Input: High-resolution panchromatic image (PAN) & Low-resolution hyperspectral image (LR-HSI).
Output: High-resolution hyperspectral image (HR-HSI).

Existing Methods vs. Proposed Method

Existing methods simply concatenates PAN and LR-HIS in,

a) Image-domain, OR
b) Feature-domain.

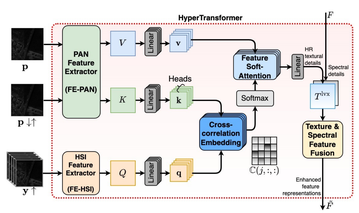


Our ***HyperTransformer*** utilize feature representations of **LR-HSI**, **PAN↓↑**, and **PAN** as **Queries (Q)**, **Keys (K)** and **Values (V)** in an attention mechanism to transfer most relevant HR texture features to spectral features of LR-HSI.

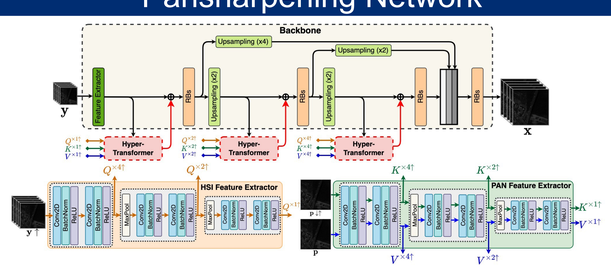
What is inside HyperTransformer?

HyperTransformer consists of four main modules,

1. **Feature extractors** for PAN and LR-HIS,
2. Multi-Head **feature soft-attention** module,
3. **Texture-Spectral feature fusion** module.



Pansharpening Network



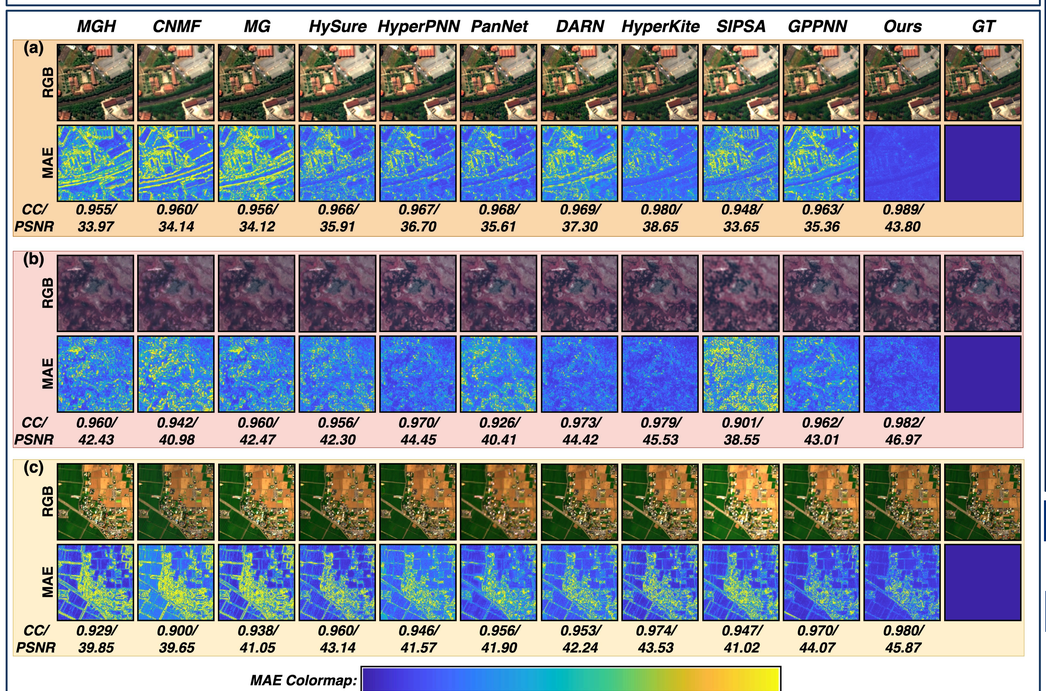
- HyperTransformer is applied at **three scales**: ×1, × 2, × 4.
- HyperTransformer is optimized using three loss terms: (1) **Reconstraction loss** (L1 loss), (2) **VGG perceptual loss**, and (3) **Transfer perceptual loss**.

Results

HyperTransformer improves SOTA results in CC/SAM/RMSE/ERGAS/PSNR by ~0.9/26.9/32.6/29.4/13.3%, ~0.3/19.2/11.9/14.0/3.2%, ~0.6/12.7/13.6/13.8/4.1% on Pavia Center, Botswana, and Chikusei datasets.

Method	Pavia Center Dataset [34]					Botswana Dataset [38]					Chikusei Dataset [53]				
	CC	SAM	RMSE $\times 10^{-2}$	ERGAS	PSNR	CC	SAM	RMSE $\times 10^{-2}$	ERGAS	PSNR	CC	SAM	RMSE $\times 10^{-2}$	ERGAS	PSNR
PCA [18] PERS-2014	0.845	8.92	3.45	6.64	31.26	0.943	2.38	1.98	2.22	40.03	0.887	6.99	2.47	7.71	30.98
GFPCA [25] DataFusion-2014	0.902	8.31	3.98	7.44	29.09	0.901	2.66	2.45	2.75	37.83	0.883	4.76	1.98	7.00	30.96
BF [43] JSTSP-2015	0.918	9.60	3.44	6.63	30.22	0.931	2.47	1.88	2.34	40.01	0.903	5.15	1.94	6.62	37.89
BFS [42] TGRS-2015	0.925	8.10	3.05	6.00	31.09	0.932	2.39	1.85	2.32	40.15	0.917	4.69	1.72	6.39	37.99
SFIM [27] IJRS-2000	0.946	6.76	2.55	5.43	32.61	0.932	3.44	2.81	2.25	39.58	0.928	3.79	1.43	6.43	39.55
GS [3] TGRS-2007	0.961	6.62	2.55	4.95	32.93	0.946	2.34	1.93	2.17	40.14	0.733	5.64	2.96	8.17	35.13
GSA [20] US Pat.-2000	0.950	7.15	2.34	4.70	33.52	0.955	2.04	1.59	1.85	41.89	0.943	3.52	1.42	4.30	41.38
MGH [2] PERS-2006	0.955	6.81	2.25	4.77	33.97	0.960	2.07	1.54	1.69	42.43	0.929	3.82	1.45	6.40	39.85
CNMF [55] TGRS-2011	0.960	6.64	2.20	4.39	34.14	0.942	2.61	1.73	2.10	40.98	0.900	4.72	1.91	5.75	39.65
MG [1] TGRS-2002	0.956	6.55	2.20	4.45	34.12	0.960	2.02	1.51	1.68	42.47	0.938	3.81	1.52	4.41	41.05
HySure [35] TGRS-2014	0.966	6.13	1.80	3.77	35.91	0.956	2.15	1.46	1.77	42.30	0.960	2.98	1.13	**3.69**	43.14
HyperPNN [14] JST-RS-2019	0.967	6.09	1.67	3.82	36.70	0.970	1.67	1.15	1.44	**44.45**	0.946	3.97	1.11	4.77	41.57
PanNet [49] ICCV-2017	0.968	6.36	1.83	3.89	35.61	0.926	2.17	1.53	2.82	40.41	0.956	3.79	**0.88**	5.32	41.90
DARN [58] TGRS-2020	**0.969**	6.43	**1.56**	**3.95**	**37.30**	**0.973**	**1.58**	**1.09**	**1.35**	44.42	0.953	3.60	1.05	4.44	42.24
HyperKite [7] TGRS-2021	**0.980**	**5.61**	**1.29**	**2.85**	**38.65**	**0.979**	**1.46**	**1.01**	**1.21**	**45.53**	**0.974**	**2.85**	1.03	**3.62**	**43.53**
SIPSA [24] CVPR-2021	0.948	**5.27**	2.38	4.52	33.65	0.901	2.34	2.20	2.54	38.55	0.947	2.87	1.06	5.09	41.02
GPPNN [46] CVPR-2021	0.963	6.52	1.91	4.05	35.36	0.962	1.90	1.36	1.65	43.01	**0.970**	**2.75**	**0.66**	4.24	**44.07**
Ours	**0.989**	**3.85**	**0.87**	**2.01**	**43.80**	**0.982**	**1.18**	**0.89**	**1.04**	**46.97**	**0.980**	**2.40**	**0.57**	**3.12**	**45.87**

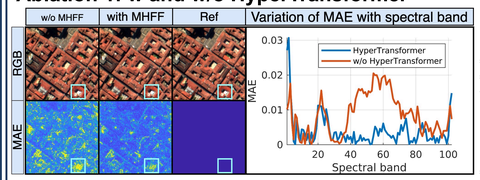
*Higher values of CC and PSNR, and lower values of SAM, RMSE, and ERGAS indicate good pansharpening performance. The ideal values of CC, SAM, RMSE, ERGAS, and PSNR are 1, 0, 0, 0, and ∞, respectively. Color convention: **best**, **2nd-best**, and **3rd-best**.



Qualitative results of different pansharpening algorithms on **(a) Pavia Center**, **(b) Botswana**, and **(c) Chikusei datasets.**

- **RGB =** RGB synthesized image from HIS, **MAE =** Mean Absolute Error.
- Our HyperTransformer achieves significantly **lower MAE** than all the other methods.
- These visual results further demonstrate the excellent ability of our HyperTransformer to extract fine details more effectively.

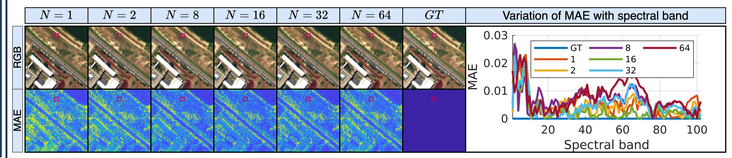
Ablation 1: w and w/o HyperTransformer



B/L: Base network
N: Number of heads (N)

N	CC	SAM	RMSE	ERGAS	PSNR
B/L	0.981	4.88	0.0133	2.84	38.71
1	0.975	4.32	0.0103	2.31	40.59
2	0.976	4.19	0.0095	2.18	42.52
8	**0.987**	4.06	0.0092	2.13	**43.20**
16	**0.989**	**3.85**	**0.0087**	**2.01**	**43.80**
32	**0.988**	**4.02**	**0.0090**	**2.10**	**43.47**
64	**0.987**	**4.04**	**0.0091**	**2.12**	43.19

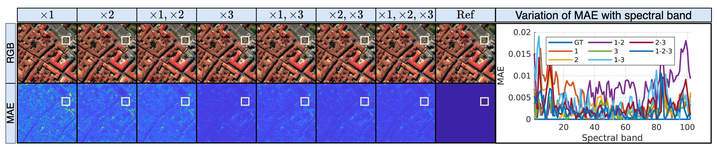
Ablation 2: Number of global descriptors



N	CC	SAM	RMSE	ERGAS	PSNR
B/L	0.981	4.88	0.0133	2.84	38.71
1	0.975	4.32	0.0103	2.31	40.59
2	0.976	4.19	0.0095	2.18	42.52
8	**0.987**	4.06	0.0092	2.13	**43.20**
16	**0.989**	**3.85**	**0.0087**	**2.01**	**43.80**
32	**0.988**	**4.02**	**0.0090**	**2.10**	**43.47**
64	**0.987**	**4.04**	**0.0091**	**2.12**	43.19

Ablation 3: Multi-scale features

×1	×2	×4	CC	SAM	RMSE	ERGAS	PSNR
B/L			0.956	4.86	0.0204	3.90	35.81
✓	✗	✗	0.975	4.76	0.0149	3.00	38.42
✗	✓	✗	0.985	4.29	0.0108	2.40	40.80
✓	✓	✗	0.985	4.41	0.0109	2.38	41.02
✗	✗	✓	**0.986**	**4.01**	0.0098	2.21	42.88
✓	✗	✓	**0.988**	**3.96**	**0.0092**	**2.20**	**43.58**
✗	✓	✓	**0.988**	**3.85**	**0.0089**	**2.09**	**43.60**
✓	✓	✓	**0.989**	**3.85**	**0.0087**	**2.01**	**43.80**



Ablation 4: Different loss terms

L1: Reconstruction loss
$L_{vgg\text{-}per}$: VGG perceptual loss
$L_{t\text{-}per}$: Transfer perceptual loss

L_1	$\mathcal{L}_{vgg\text{-}per}$	$\mathcal{L}_{t\text{-}per}$	CC	SAM	RMSE	ERGAS	PSNR
✓	✗	✗	**0.987**	**4.07**	**0.0091**	**2.12**	**43.00**
✓	✓	✗	**0.988**	**4.01**	**0.0090**	**2.09**	**43.60**
✓	✓	✓	**0.989**	**3.85**	**0.0087**	**2.01**	**43.80**

Acknowledgement

This work was supported by NSF CAREER award 2045489.

Conclusion

- We proposed a novel textural-spectral feature fusion network for pansharpening.
- Extensirve experiemnts conducted on three datasets demonstrate the superiority of our HyperTransfroemr over SOTA approaches on quantitative and qualitative evaluations.

SCAN ME