Paper ID: 4937

Bridging Local–Global Dissonance: Learning from Compressive Measurements for Hyperspectral Reconstruction

Xian-Hua Han
Rikkyo University, Japan

RIKKYO AI

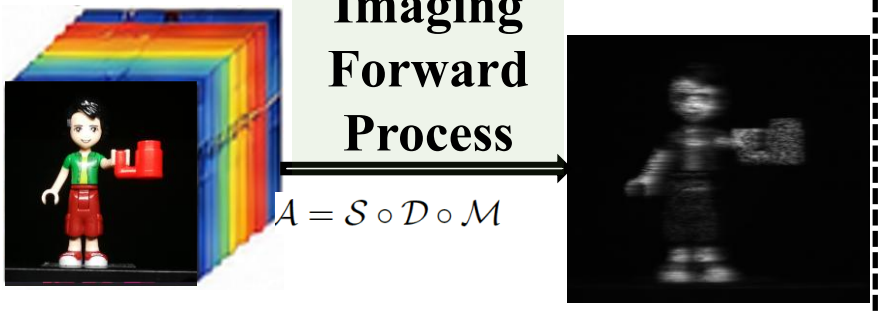
ICML
International Conference
On Machine Learning

RIKKYO UNIVERSITY
Graduate School of Artificial Intelligence and Science

Introduction & Motivation

Snapshot Compressive Imaging (SCI):

→Capture a full **hyperspectral (HS) cube** into a **single 2D Snapshot**



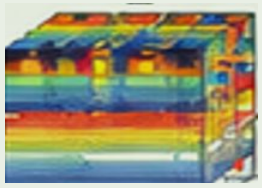
A fundamental Local-Global Dissonance

$\mathcal{M}$ **(Modulation)**

Coded Aperture

$\mathcal{D}$ **(Dispersion)**

Wavelength-dependent shear



$\mathcal{S}$ **(Integration)**

Spectral sum

Locally modulated measurements vs. **Globally entangled** spectral structures

HSI Reconstruction: A **severely ill-posed** inverse problem due to **information loss and noise**

Limitations of Existing Methods

CNN-based methods:

Strong local modeling but **weak long-range** spectral spatial reasoning

Transformer-based Methods:

Global attention amplifies **artifacts** and **destroys** mask-guided **local structures** under compressive sensing

Hybrid models:

Local and global representations are **loosely coupled** and **not aligned** with CASSI physics.

Existing unfolding priors:

Lack physics consistency, causing unstable optimization under severe under-sampling

Proposed Method

HSRA (Hierarchical Scale-Reconciling Architecture) :
A **physics-consistent multi-scale prior**

->Progressively reconciles **local sufficiency** and **global coherence** according to the CASSI forward model

MK-PTM: Preserves **fine local** spectral–spatial details via multi-kernel local mixing $\Longleftrightarrow \mathcal{M}$

LWIM: Enables efficient **mid-scale latent** interaction for dispersion-aware reasoning $\Longleftrightarrow \mathcal{D}$

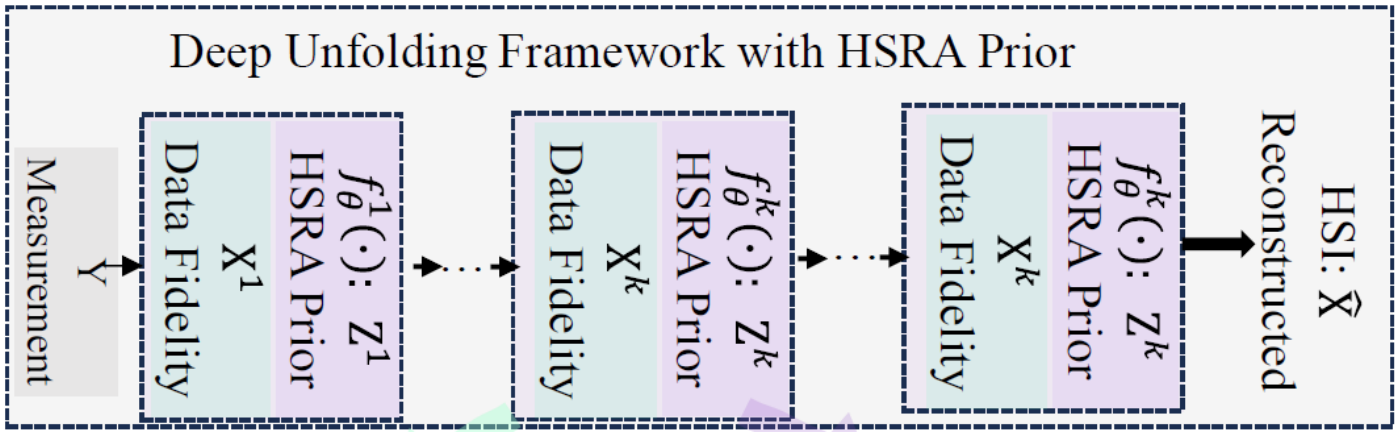
HM-SSA: Progressively **expands attention granularity** for Stable global coherence $\Longleftrightarrow \mathcal{D} \circ \mathcal{S}$

Proposed Method

HSRA is embedded into a **physics-consistent deep unfolding framework**

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathcal{A}(\mathbf{X}) - \mathbf{Y}\|_2^2 + \lambda\mathcal{R}(\mathbf{X})$$

Learned Prior: $f_\theta^k(\cdot)$ Using HSRA



Data Fidelity Update
(**Physical-Driven**)

$X^{k+1} = Z^k - \mu\mathcal{A}^T(\mathcal{A}Z^k - Y)$

Consistency with CASSI measurement

Prior Update
(**HSRA Network**)

$Z^{k+1} = f_\theta^{k+1}(X^{k+1})$

Scale-aware reconciliation
(Local -> Mid -> Global

Key Insight

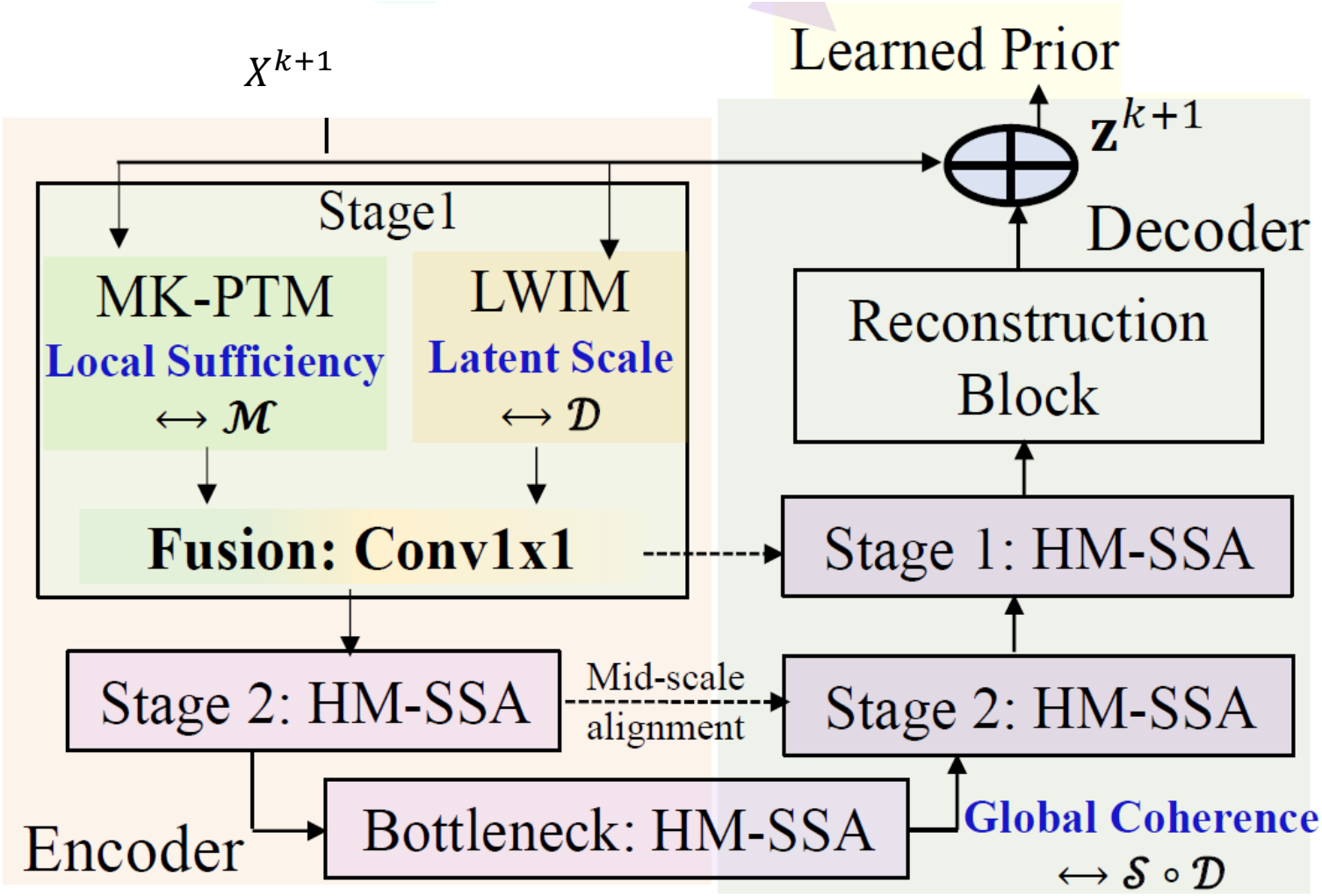
Instead of uniform global modelling, HSRA performs progressive scale-aware reconciliation:

Local Sufficiency -> Mid-scale Interaction -> Global Coherence

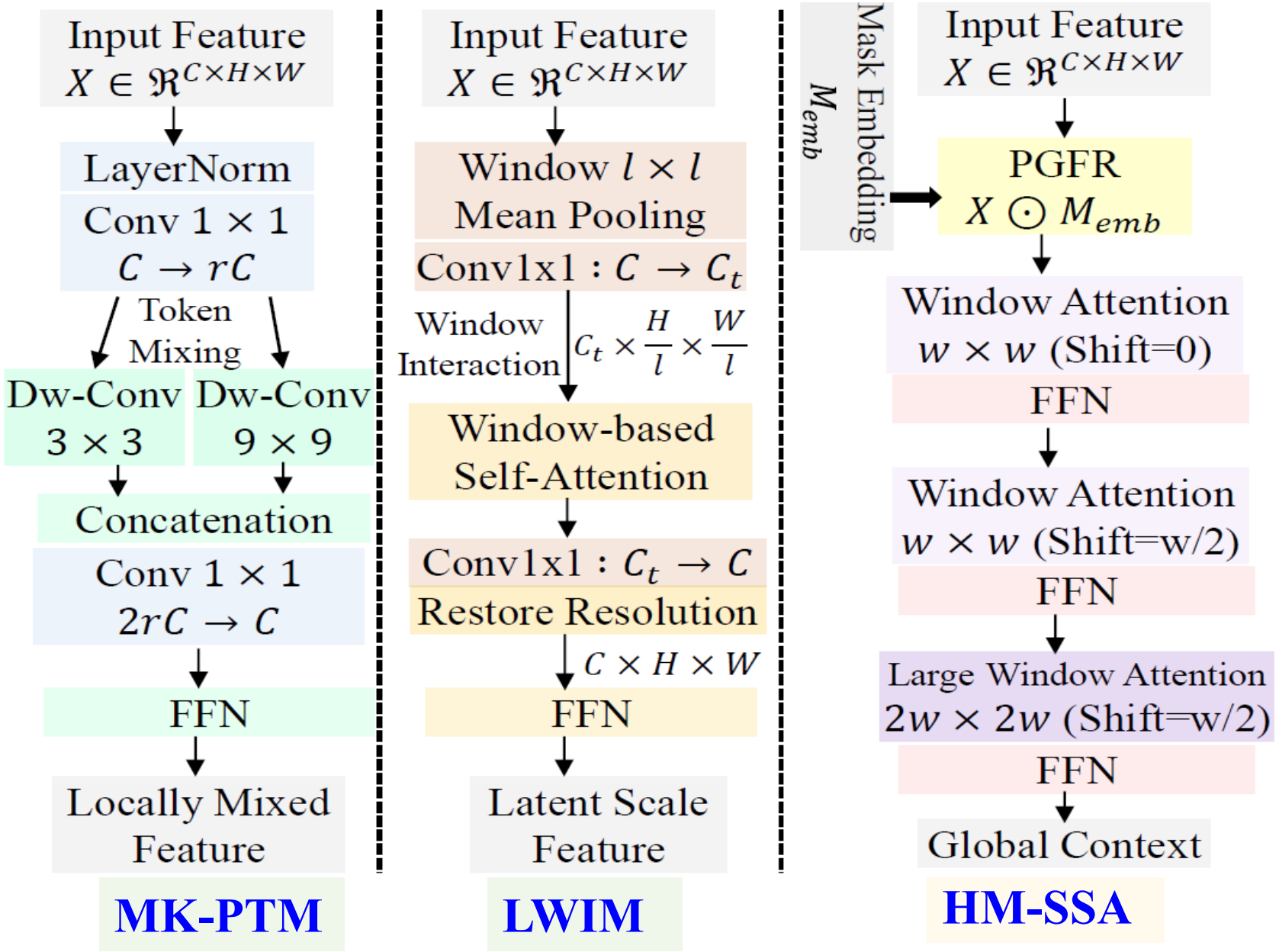
Hierarchical Scale-Reconciling Architecture

Stage 1 of Encoder
MK-PTM: Local sufficiency
LWIM: Mid-range interaction

Other Stages of Encoder/Decoder
HM-SSA: Global structural coherence



Hierarchical Scale-Reconciling Architecture



Experimental Results

Dataset: CAVE (Training), 10 Scenes from KAIST (Test)

5 Real Snapshot Measurements **Three Model Variants**

Comparisons of End-2-End Methods on Simulated Dataset

Methods	GFLOPs	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	Avg
	E2E methods											
TSA-Net (Meng et al., 2020b)	110.06	32.03	31.00	32.25	39.19	29.39	31.44	30.32	29.35	30.01	29.59	31.46
		0.892	0.858	0.915	0.953	0.884	0.908	0.878	0.888	0.890	0.874	0.894
HDNet (Hu et al., 2022)	154.76	35.14	35.67	36.03	42.30	32.69	34.46	33.67	32.48	34.89	32.38	34.97
		0.935	0.940	0.943	0.969	0.946	0.952	0.926	0.941	0.942	0.937	0.943
MST-S (Cai et al., 2022b)	12.96	34.71	34.45	35.32	41.50	31.90	33.85	32.69	31.69	34.67	31.82	34.26
		0.930	0.925	0.943	0.967	0.933	0.943	0.911	0.933	0.939	0.926	0.935
MST-M (Cai et al., 2022b)	18.07	35.15	35.19	36.26	42.48	32.49	34.28	33.29	32.40	35.35	32.53	34.94
		0.937	0.935	0.950	0.973	0.943	0.948	0.921	0.943	0.942	0.935	0.943
MST-L (Cai et al., 2022b)	28.15	35.40	35.87	36.51	42.27	32.77	34.80	33.66	32.67	35.39	32.50	35.18
		0.941	0.944	0.953	0.973	0.947	0.955	0.925	0.948	0.949	0.941	0.948
CST-S (Cai et al., 2022a)	11.67	34.78	34.81	35.42	41.84	32.29	34.49	33.47	32.89	34.96	32.14	34.71
		0.930	0.931	0.944	0.967	0.939	0.949	0.922	0.945	0.944	0.932	0.940
CST-M (Cai et al., 2022a)	16.91	35.16	35.60	36.57	42.29	32.82	35.15	33.85	33.52	35.28	32.84	35.31
		0.938	0.942	0.953	0.972	0.948	0.956	0.927	0.952	0.946	0.940	0.947
CST-L (Cai et al., 2022a)	27.81	35.82	36.54	37.39	42.28	33.40	35.52	34.44	33.83	35.92	33.36	35.85
		0.947	0.952	0.959	0.972	0.953	0.962	0.937	0.961	0.951	0.948	0.954
CST-L+ (Cai et al., 2022a)	40.01	35.96	36.84	38.16	42.44	33.25	35.72	34.86	34.34	36.51	33.09	36.12
		0.949	0.955	0.962	0.975	0.955	0.963	0.944	0.961	0.957	0.945	0.957
S^2-Tran (Wang et al., 2025)	27.21	36.17	37.57	37.29	42.96	34.40	**36.44**	35.41	34.50	36.54	33.57	36.48
		0.949	0.958	0.957	0.975	0.960	0.965	0.946	0.963	0.959	0.952	0.958
DWMT (Luo et al., 2024)	46.71	36.46	37.75	38.47	44.23	33.99	36.17	35.22	34.56	37.41	**34.00**	36.82
		0.957	**0.963**	0.965	0.984	0.963	**0.970**	0.949	**0.968**	**0.965**	**0.959**	0.964
HSRA-S (Our)	9.01	35.85	36.92	37.45	43.93	34.15	35.06	34.97	33.57	36.30	32.60	36.08
		0.947	0.949	0.955	0.978	0.956	0.955	0.941	0.951	0.947	0.938	0.952
HSRA-M (Our)	13.05	36.46	37.49	38.64	44.63	34.46	35.71	35.92	34.07	37.24	33.11	36.77
		0.954	0.957	0.961	0.983	0.961	0.963	0.953	0.959	0.958	0.947	0.960
HSRA-L (Our)	27.34	**36.90**	**38.12**	**39.69**	**45.38**	**35.34**	36.23	**36.32**	**34.82**	**38.09**	**34.00**	**37.49**
		0.958	0.962	**0.968**	**0.986**	**0.967**	0.969	**0.956**	0.966	0.964	0.957	**0.965**

Experimental Results

Dataset: CAVE (Training), 10 Scenes from KAIST (Test)
5 Real Snapshot Measurements

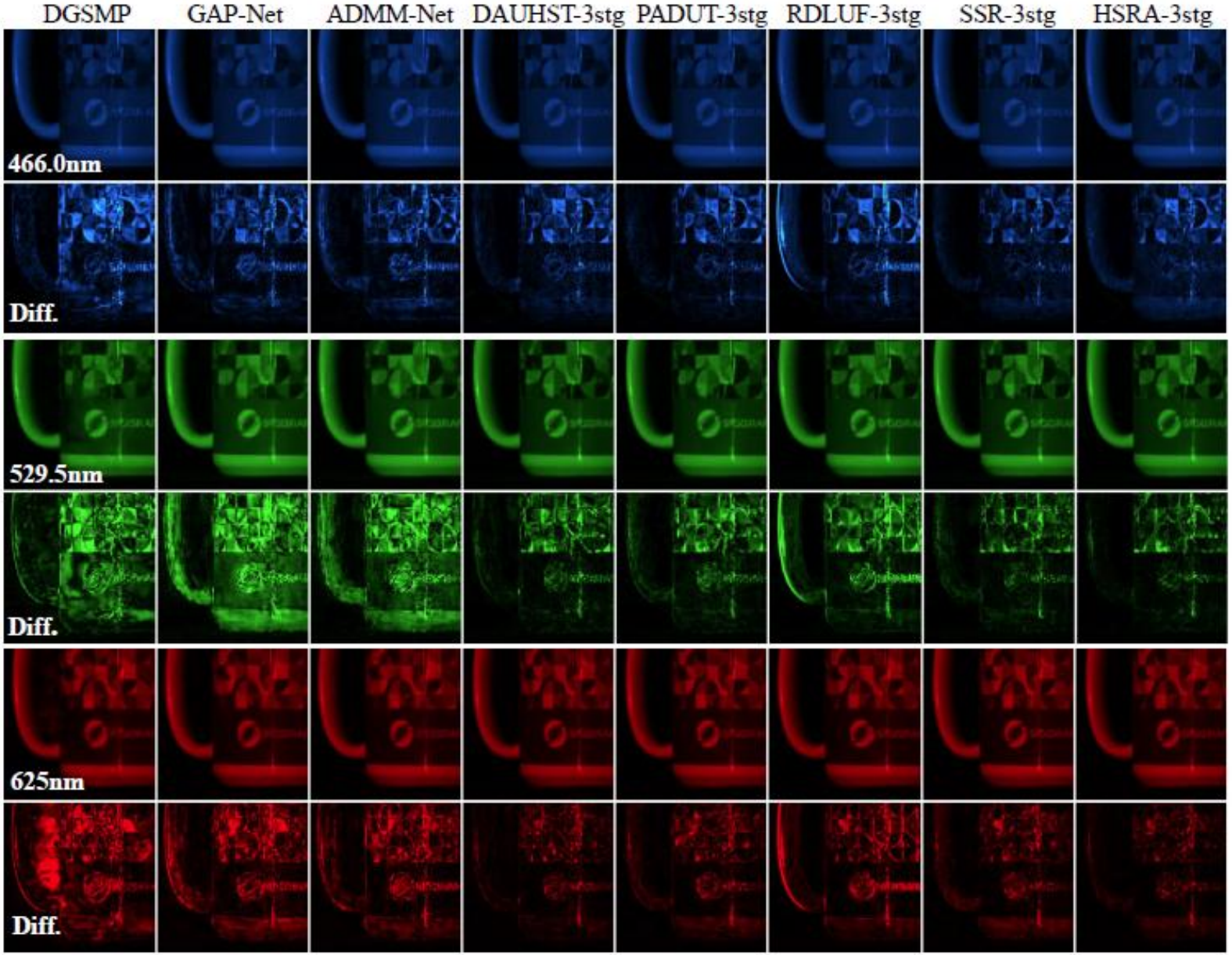
Comparisons of Deep Unfolding Models→ New SOTA

DAUHST-3stg (Cai et al., 2022c)	27.17	36.56 0.953	37.92 0.960	39.36 0.968	44.97 0.985	34.82 0.964	36.22 0.968	35.99 0.953	34.24 0.963	38.50 0.967	33.63 0.954	37.22 0.963
PADUT-3stg (Li et al., 2023)	22.91	36.16 0.953	37.83 0.963	39.55 0.971	44.43 0.985	34.56 0.965	35.56 0.967	35.61 0.951	33.70 0.963	38.14 0.965	33.18 0.950	36.87 0.963
RDLUF-3stg (Dong et al., 2023)	38.45	36.67 0.953	38.48 0.965	40.63 0.971	46.04 0.986	34.63 0.963	36.18 0.966	35.85 0.951	34.37 0.963	38.98 0.966	33.73 0.950	37.56 0.963
SSR-3stg (Zhang et al., 2024b)	28.38	**38.11** 0.967	39.91 0.975	42.33 0.979	46.86 0.991	36.42 0.974	37.23 0.975	37.49 0.967	35.59 0.973	41.20 0.979	34.99 0.965	39.01 0.974
HSRA-DUM-3stg (Our)	37.39	37.85 **0.970**	40.85 0.981	42.82 0.981	47.21 0.992	37.01 0.978	37.69 0.980	38.30 0.971	36.07 0.977	40.96 0.979	35.25 0.969	39.40 0.978
HSRA-DUM-4stg (Our)	49.82	37.81 **0.970**	**41.25** **0.983**	**43.40** **0.982**	**47.83** **0.993**	**37.51** **0.980**	**38.13** **0.981**	**38.72** **0.974**	**36.50** **0.979**	**41.48** **0.981**	**35.44** **0.970**	**39.81** **0.979**

Best Overall: 39.81 dB/ 0.982 (HSEA-4stg)

Experimental Results

Visual Comparisons: Three bands and Difference Images



Experimental Results: Ablation Study

Ablation Study: Introduce **scale-specific evaluation metrics** to quantitatively analyze representation behavior across **different regimes**

Category	Configuration	GFLOPs	PSNR↑	SSIM↑	SAM↓	LPSNR(8)↑	LPSNR(16)↑	ERGAS↓
Homogeneous Token Mixing	All-ConvFormer	8.78	35.40	0.944	8.110	40.995	40.132	23.591
	All-MK-PTM	8.41	35.47	0.944	8.361	40.862	40.028	23.258
	All-SwinFormer (8×8)	8.80	34.98	0.941	7.666	41.079	40.137	24.209
	All-HM-SSA	9.47	35.11	0.943	7.712	4.103	40.213	24.123
Local Sufficiency	MK-PTM, r=0.5	8.71	35.90	0.950	7.780	41.297	40.447	22.106
	MK-PTM, 3×3 only	8.64	35.92	0.950	8.077	41.246	40.404	21.930
	r=1, kernels {3, 9}	–	≥ +0.16	+0.002	↓	↑	↑	↓
Mid-Scale Interaction	W/o LWIM	8.23	35.89	0.948	8.112	41.142	40.303	22.068
	With LWIM (ours)	–	+0.19	+0.004	↓	↑	↑	↓
Global Coherence	W/o large-window SA ($2w$)	6.64	35.65	0.945	8.304	40.890	40.040	22.631
	Full HM-SSA	–	+0.43	+0.007	↓	↑	↑	↓
Full Model	**HSRA-S (Ours)**	**9.01**	**36.08**	**0.952**	**7.684**	**41.530**	**40.661**	**21.542**

SAM: Local spectral fidelity at the pixel level

LPSNR (8/16): Mid-scale structural consistency over different local regions

ERGAS: Global spatial-spectral coherence->measuring the overall consistency of reconstructed HS structures

Experimental Results: Ablation Study

Ablation Study: Introduce **scale-specific evaluation metrics** to quantitatively analyze representation behavior across **different regimes**

Category	Configuration	GFLOPs	PSNR↑	SSIM↑	SAM↓	LPSNR(8)↑	LPSNR(16)↑	ERGAS↓
Homogeneous Token Mixing	All-ConvFormer	8.78	35.40	0.944	8.110	40.995	40.132	23.591
	All-MK-PTM	8.41	35.47	0.944	8.361	40.862	40.028	23.258
	All-SwinFormer (8×8)	8.80	34.98	0.941	7.666	41.079	40.137	24.209
	All-HM-SSA	9.47	35.11	0.943	7.712	4.103	40.213	24.123
Local Sufficiency	MK-PTM, r=0.5	8.71	35.90	0.950	7.780	41.297	40.447	22.106
	MK-PTM, 3×3 only	8.64	35.92	0.950	8.077	41.246	40.404	21.930
	r=1, kernels {3, 9}	–	≥ +0.16	+0.002	↓	↑	↑	↓
Mid-Scale Interaction	W/o LWIM	8.23	35.89	0.948	8.112	41.142	40.303	22.068
	With LWIM (ours)	–	+0.19	+0.004	↓	↑	↑	↓
Global Coherence	W/o large-window SA ($2w$)	6.64	35.65	0.945	8.304	40.890	40.040	22.631
	Full HM-SSA	–	+0.43	+0.007	↓	↑	↑	↓
Full Model	**HSRA-S (Ours)**	**9.01**	**36.08**	**0.952**	**7.684**	**41.530**	**40.661**	**21.542**

SAM: Local spectral fidelity at the pixel level

LPSNR (8/16): Mid-scale structural consistency over different local regions

ERGAS: Global spatial-spectral coherence->measuring the overall consistency of reconstructed HS structures

Takeaways

- ✓ **Addresses the fundamental Local-Global dissonance in CASSI**
- ✓ **Physics-consistent, hierarchical, and scale-reconciling**
- ✓ **Superior accuracy, robustness, and efficiency**
- ✓ **General and extensible to various inverse imaging tasks**

Code will be publicly available at: https://github.com/hanxhua1216/HSRA

Thank You