NEURAL INFORMATION PROCESSING SYSTEMS

Fourier-enhanced Implicit Neural Fusion Network for Multispectral and Hyperspectral Image Fusion

Yu-Jie Liang †[1]
yujieliang0219@gmail.com

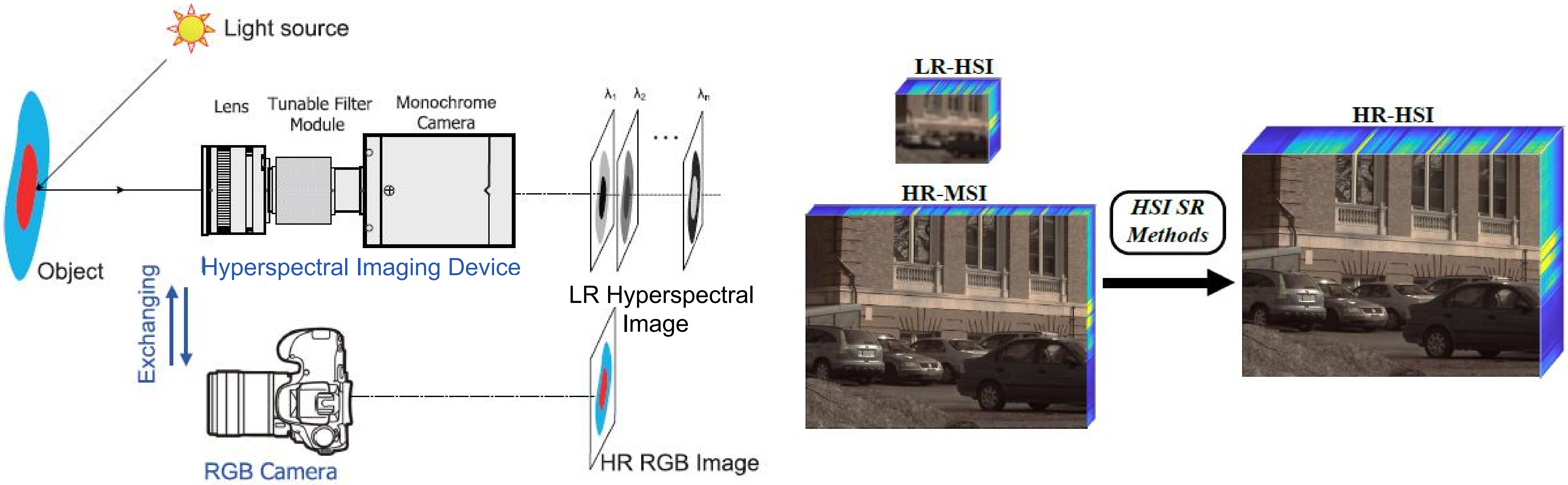
Zihan Cao †[1]
iamzihan666@gmail.com

Shangqi Deng [2]
shangqideng0124@gmail.com

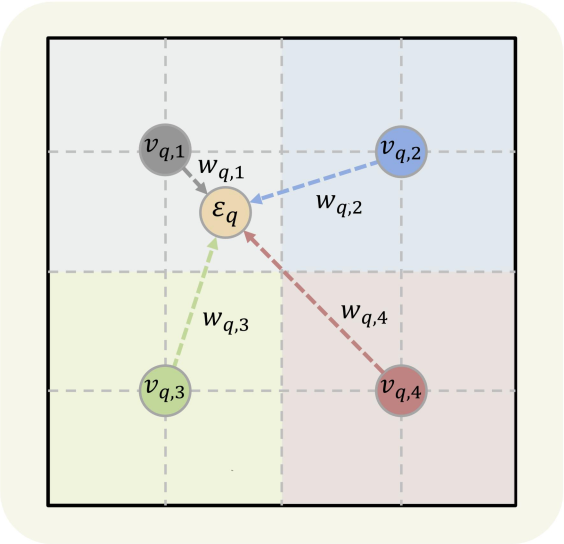
Hong-Xia Dou [3]
hongxiadou1991@126.com

Liang-Jian Deng* [1]
liangjian.deng@uestc.edu.cn

Multispectral and Hyperspectral Image Fusion



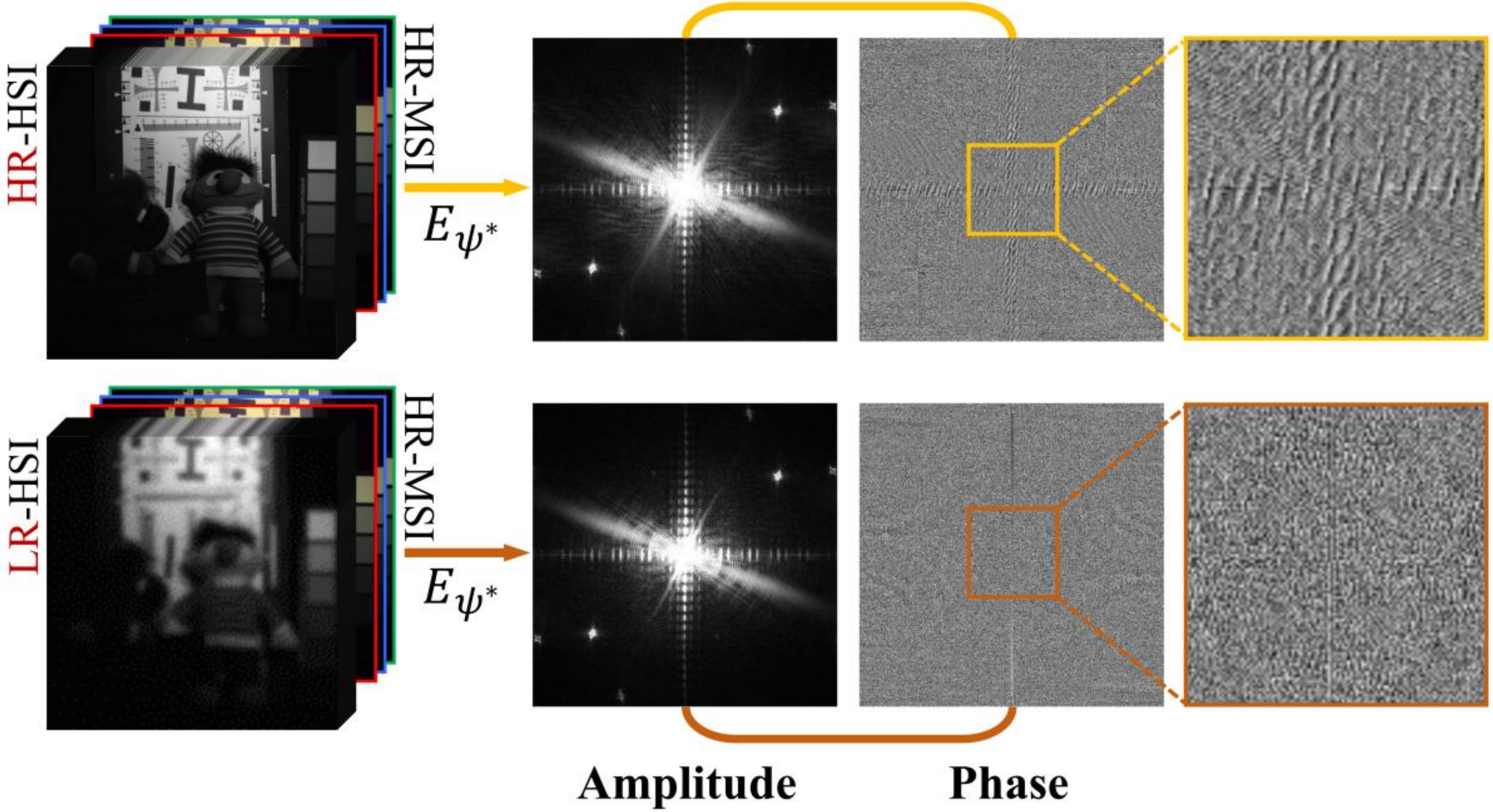
Preliminary and Motivation



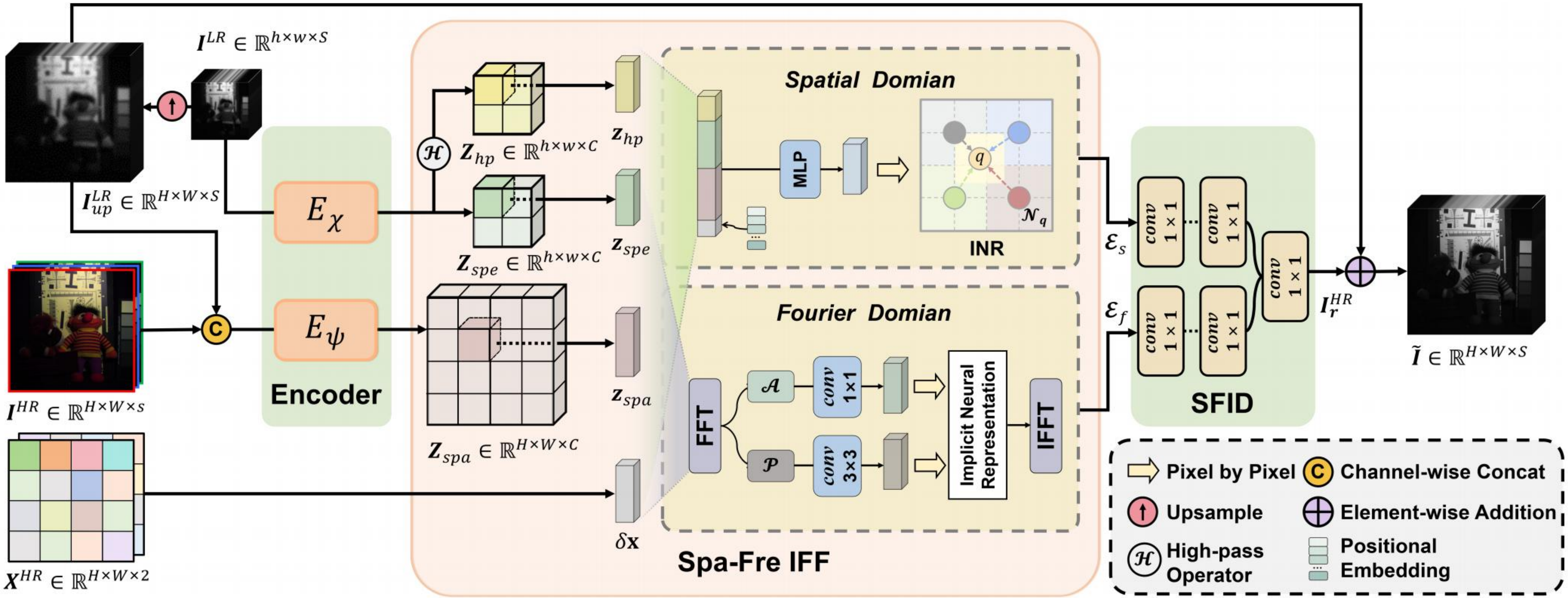
$$\hat{\mathbf{I}}(\mathbf{x}_q) = \sum_{i \in \mathcal{N}_q} w_{q,i} \mathbf{v}_{q,i}$$

$$w_{q,i} = S_i / S$$

$$\mathbf{v}_{q,i} = \phi_\theta(\mathbf{z}_i, \mathbf{x}_q - \mathbf{x}_i)$$



FeINFN Framework



Spatial-Frequency Implicit Fusion Function

- **INR Encoder Networks**

  The INR encoders try to extract spatial and spectral latent codes:

$$\mathbf{Z}_{spe} = E_{\chi}(\mathbf{I}^{LR}), \quad \mathbf{Z}_{spa} = E_{\psi}\left(\mathrm{Cat}(\mathbf{I}_{up}^{LR}, \mathbf{I}^{HR})\right)$$

- **Spatial-Frequency Implicit Fusion Function**

Spatical Domain:

$$\mathbf{z}_{hp} = \mathcal{H}(\mathbf{z}_{spe})$$

$$\mathbf{w}_{q,i}, \mathbf{v}_{q,i} = \phi_{\theta}(\mathbf{z}_{spe}, \mathbf{z}_{spa}, \mathbf{z}_{hp}, \gamma(\delta\mathbf{x}))$$

$$\boldsymbol{\varepsilon}_s = \sum_{i \in \mathcal{N}_q} \overline{\mathbf{w}}_{q,i} * \mathbf{v}_{q,i}.$$

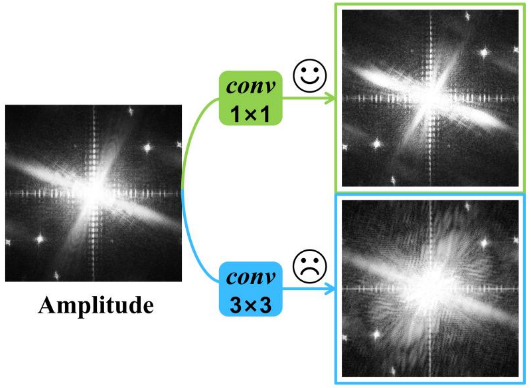
$$\delta\mathbf{x} = \{\mathbf{x}_q - \mathbf{x}_{q,i}\}_{i \in \mathcal{N}_q}$$

$$\gamma(\delta\mathbf{x}) = [\sin(2^0\delta\mathbf{x}), \cos(2^0\delta\mathbf{x}), \cdots, \sin(2^{L-1}\delta\mathbf{x}), \cos(2^{L-1}\delta\mathbf{x})]$$

Frequency Domain: We employ FFT to transform latent codes from the spatial domain to the frequency domain, handling amplitude and phase separately:

$$\mathbf{w}_{q,i}^{\mathcal{A}}, \mathbf{v}_{q,i}^{\mathcal{A}} = \phi_{\alpha}^{\mathcal{A}}(\mathcal{A}(\mathbf{f}_{spe}), \mathcal{A}(\mathbf{f}_{spa}), \delta\mathbf{x}) \qquad \mathcal{A}_f' = \sum_{i \in \mathcal{N}_q} \overline{\mathbf{w}}_{q,i}^{\mathcal{A}} * \mathbf{v}_{q,i}^{\mathcal{A}}$$

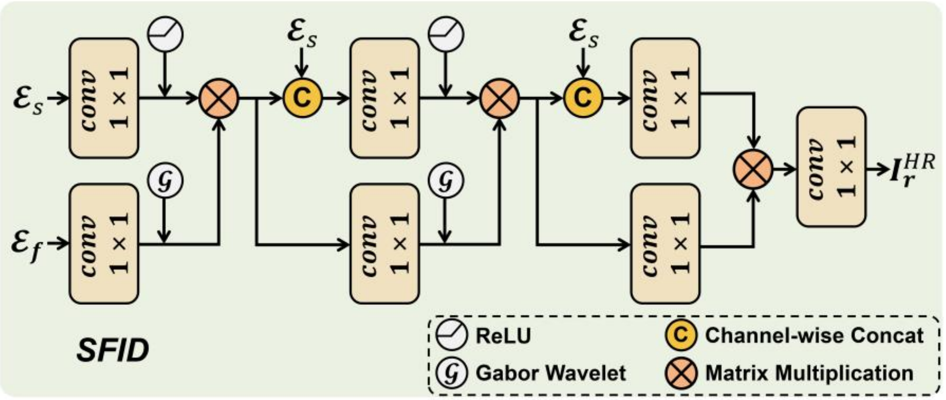
$$\mathbf{w}_{q,i}^{\mathcal{P}}, \mathbf{v}_{q,i}^{\mathcal{P}} = \phi_{\beta}^{\mathcal{P}}(\mathcal{P}(\mathbf{f}_{spe}), \mathcal{P}(\mathbf{f}_{spa}), \delta\mathbf{x}) \qquad \mathcal{P}_f' = \sum_{i \in \mathcal{N}_q} \overline{\mathbf{w}}_{q,i}^{\mathcal{P}} * \mathbf{v}_{q,i}^{\mathcal{P}}$$



Spatial-Frequency Interactive Decoder

time domain feature

frequency domain feature



Theorem 1. ***The complex Gabor wavelet activation*** $\mathcal{G}(\mathbf{x}) = e^{j\omega_0 \mathbf{x}} e^{-|\upsilon_0 \mathbf{x}|^2}$ ***has the time-frequency tightness property . Moreover, from the perspective of signal spectrum analysis, this activation helps the decoder learn the optimal bandwidths.***

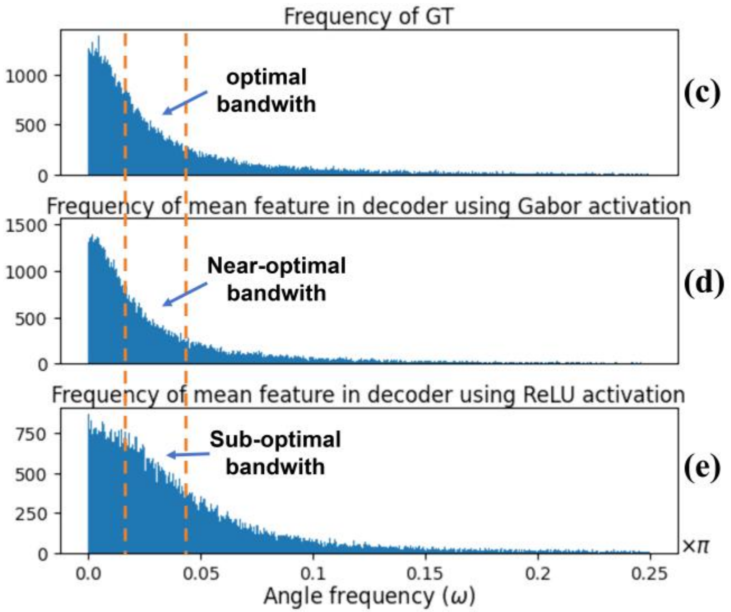
Time domain Tightness:

$|\mathcal{G}(\mathbf{x})|$ is primarily concentrated around **x**=0 due to the exponential decay term.

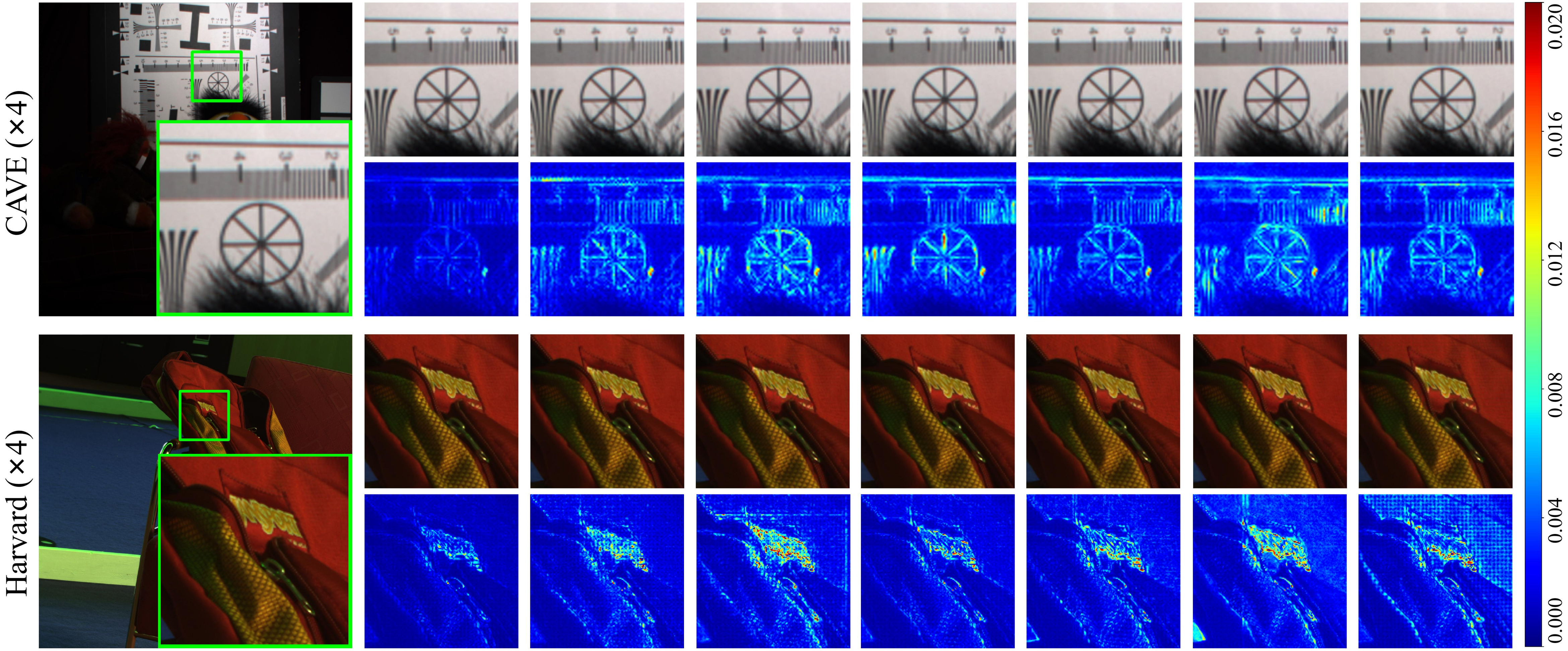
Frequency domain Tightness:

$$\mathcal{F}[\mathcal{G}(\mathbf{x})] = \int e^{j\omega_0 \mathbf{x}} e^{-|\upsilon_0 \mathbf{x}|^2} e^{-j\omega \mathbf{x}} d\mathbf{x}.$$

Uncertainty principle: $|\omega_0| \cdot \upsilon_0 \geq \frac{1}{4\pi}$



Experiments



(a) GT (b) Ours (c)MoGDCN (e) 3DT-Net (e) DSPNet (f) DHIF (g) MIMO-SST (h) PSRT

Experiments

Methods	CAVE ×4					Harvard ×4				
	PSNR(↑)	SAM(↓)	ERGAS(↓)	SSIM(↑)	#params	PSNR(↑)	SAM(↓)	ERGAS(↓)	SSIM(↑)	#params
Bicubic	34.33±3.88	4.45±1.62	7.21±4.90	0.944±0.029	–	38.71±4.33	2.53±0.67	4.45±1.81	0.948±0.027	–
CSTF-FUS [22]	34.46±4.28	14.37±5.30	8.29±5.29	0.866±0.075	–	39.15±3.45	6.93±2.69	4.66±1.81	0.914±0.049	–
LTTR [9]	35.85±3.49	6.99±2.55	5.99±2.92	0.956±0.029	–	40.88±3.94	4.01±1.27	4.03±2.18	0.957±0.035	–
LTMR [8]	36.54±3.30	6.71±2.19	5.39±2.53	0.963±0.021	–	42.06±3.56	3.51±0.99	3.59±2.03	0.970±0.020	–
IR-TenSR [45]	35.61±3.45	12.30±4.68	5.90±3.05	0.945±0.027	–	40.47±3.04	4.36±1.52	5.57±1.57	0.963±0.014	–
ResTFNet [24]	45.58±5.47	2.82±0.70	2.36±2.59	0.993±0.006	2.387M	45.94±4.35	2.61±0.69	2.56±1.32	0.985±0.008	2.387M
SSRNet [52]	48.62±3.92	2.54±0.84	1.63±1.21	0.995±0.002	**0.027M**	48.00±3.36	2.31±0.60	2.30±1.42	0.987±0.007	**0.027M**
HSRNet [16]	50.38±3.38	2.23±0.66	1.20±0.75	0.996±0.001	0.633M	48.29±3.03	2.26±0.56	1.87±0.81	0.988±0.006	0.633M
MogDCN [10]	51.63±4.10	2.03±0.62	1.11±0.82	0.997±0.002	6.840M	47.89±4.09	2.11±0.52	1.89±0.82	0.988±0.007	6.840M
Fusformer [15]	49.98±8.10	2.20±0.85	2.50±5.21	0.994±0.011	0.504M	47.87±5.13	2.84±2.07	2.04±0.99	0.986±0.010	0.467M
DHIF [17]	51.07±4.17	2.01±0.63	1.22±0.97	0.997±0.002	22.462M	47.68±3.85	2.32±0.53	1.95±0.92	0.988±0.007	22.462M
PSRT [7]	50.47±6.19	2.19±0.64	2.06±3.71	0.996±0.003	0.247M	47.96±3.21	2.18±0.55	1.89±0.86	0.988±0.006	0.247M
3DT-Net [25]	51.38±4.18	2.16±0.70	1.14±1.00	0.996±0.003	3.464M	47.78±4.42	**2.04±0.51**	1.98±0.86	**0.989±0.006**	3.464M
DSPNet [35]	51.18±3.92	2.15±0.64	1.13±0.82	0.997±0.002	6.064M	48.29±3.16	2.30±0.55	1.93±0.93	0.988±0.006	6.064M
MIMO-SST [11]	50.98±3.39	2.23±0.70	1.18±0.73	0.997±0.002	4.983M	47.08±5.56	2.09±0.53	2.07±0.82	0.988±0.007	4.983M
FeINFN(Ours)	**52.47±4.10**	**1.91±0.59**	**0.98±0.74**	**0.998±0.002**	3.165M	**49.06±3.15**	2.10±0.53	**1.78±0.75**	0.989±0.007	3.165M

Experiments

Table 2: Quantitative comparisons with other up-sampling methods on the CAVE (×4) dataset.

Methods	PSNR(↑)	SAM(↓)	ERGAS(↓)	SSIM(↑)	#params
Bilinear	52.23±4.40	1.92±0.60	1.03±0.86	0.997±0.0021	3.119M
Bicubic	52.22±4.31	1.95±0.61	1.02±0.82	0.997±0.0021	3.119M
Pixel Shuffle	52.26±4.37	**1.90±0.59**	1.02±0.85	0.997±0.0022	3.057M
Our	**52.47±4.10**	1.91±0.59	**0.98±0.74**	**0.998±0.0015**	3.165M

Table 3: Quantitative comparisons with reduced models on the CAVE (×4) dataset. $\mathcal{S}$ & $\mathcal{F}$ mean the domain difference.

$\mathcal{S}$	$\mathcal{F}$	PSNR(↑)	SAM(↓)	ERGAS(↓)	SSIM(↑)	#params
✓	✗	52.11±4.22	1.95±0.59	1.04±0.82	0.998±0.0017	2.869M
✗	✓	47.86±3.42	3.49±1.30	1.67±1.13	0.995±0.0020	2.940M
✓	✓	**52.47±4.10**	**1.91±0.59**	**0.98±0.74**	**0.998±0.0015**	3.165M

Table 4: Quantitative comparisons with different activation functions in SFID on the CAVE (×4) dataset.

Nonlinear	PSNR(↑)	SAM(↓)	ERGAS(↓)	SSIM(↑)
ReLU	52.03±3.84	2.00±0.59	1.02±0.74	**0.998±0.0013**
GELU	51.96±3.88	2.01±0.60	1.03±0.75	0.998±0.0014
Leaky ReLU	51.98±3.92	2.01±0.60	1.03±0.76	0.998±0.0014
Our	**52.47±4.10**	**1.91±0.59**	**0.98±0.74**	0.998±0.0015

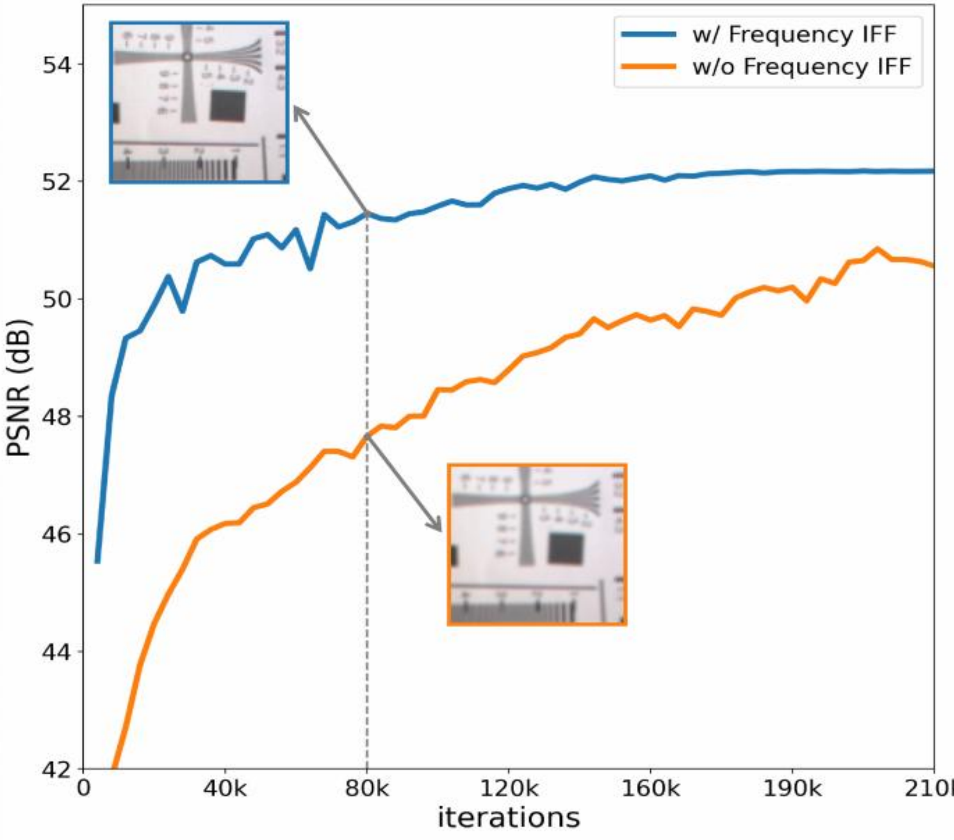


Figure 7: Changes in PSNR on the CAVE dataset of our FeINFN over iterations with and without the "Fourier Domain". The Frequency IFF can help the network learn the high-frequency details and converge faster.

Thanks for your attention !

Paper

Code

Datasets