Hybrid Spectral Denoising Transformer with Guided Attention

Zeqiang Lai[1] Chenggang Yan[2] Ying Fu[1]*

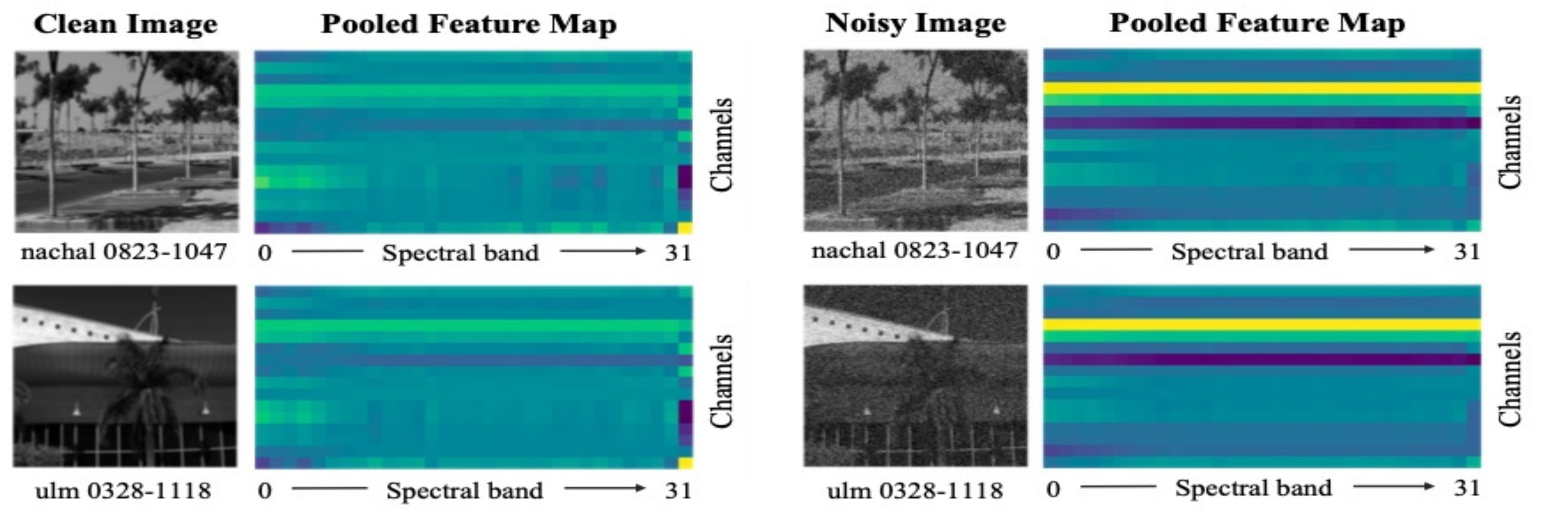
[1] Beijing Institute of Technology, [2] Hangzhou Dianzi University

ICCV23 PARIS

Introduction

Motivation

- Most existing HSI denoising method requires sperate training for different types of HSIs with different bands or data sources.
- 3D CNN deals with arbitrary HSIs but introduce large parameters.
- HSI Transformers brings performance improvements, but existing works only focus spatial or channel instead of spectral interactions.
- HSIs often exhibit beneficial fixed structures, e.g., relative intensity correlations of different bands for objects.



Contributions

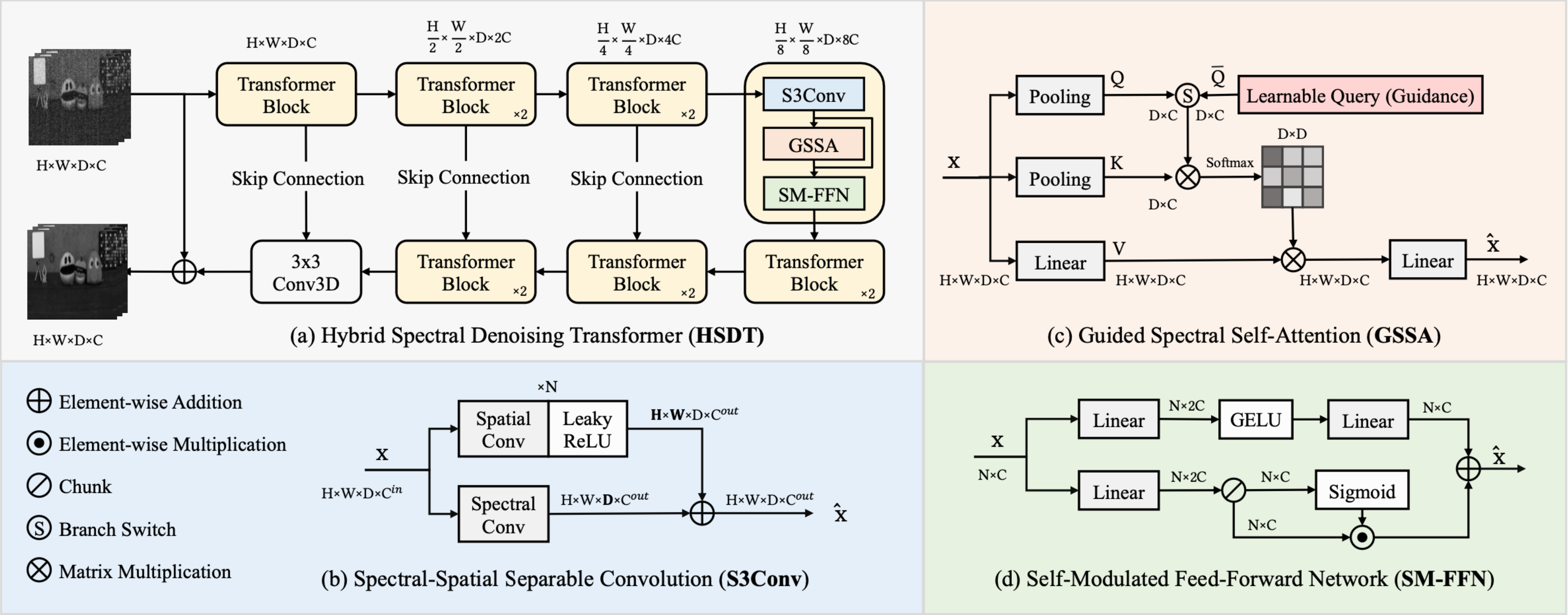
- We propose a **Hybrid Spectral Denoising Transformer** (HSDT) with a novel **Guided Spectral Self Attention** (GSSA) that incorporates learnable queries encoding the global statistics of HSIs.
- We propose **Spectral-Spatial Separable Convolution** (S3Conv) and **Self-Modulated FFN** (SM-FNN) for efficient and effective feature extraction and transformation.

Highlights

- **Flexibility :** Train **one model, solve all** HSIs with different bands and captured by different cameras.
- **Performance :** Up to 1 dB PSNR improvement across noise settings.
- **Fast Convergence :** 1 epoch to reach 39.5 PSNR on ICVL Gaussian 50. 3 epochs surpasses QRNN3D trained with 30 epochs.
- **Lightweight :** HSDT-S achieves comparable performance against the SOTA with 0.13M parameters. HSDT-M outperforms the SOTA by a large margin with only 0.52M parameters.

Method

Overall Architecture

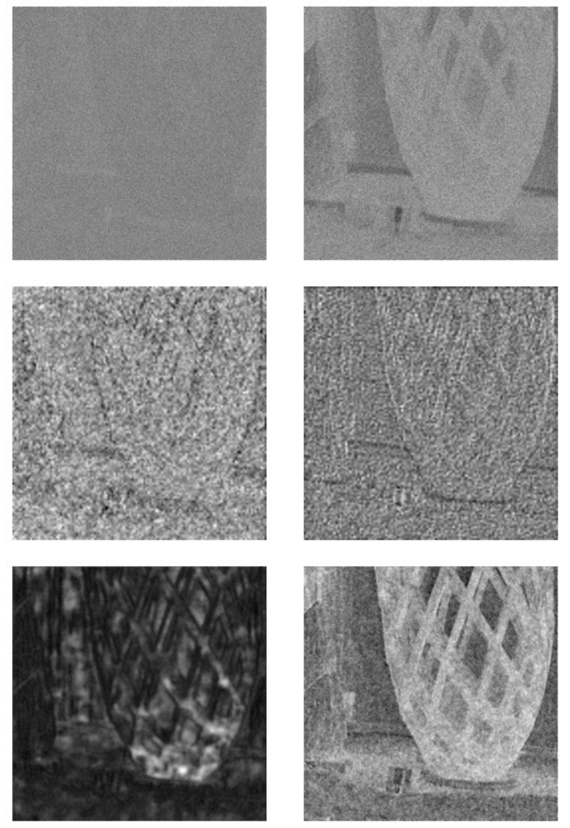


(a) Hybrid Spectral Denoising Transformer (HSDT)
(b) Spectral-Spatial Separable Convolution (S3Conv)
(c) Guided Spectral Self-Attention (GSSA)
(d) Self-Modulated Feed-Forward Network (SM-FFN)

Network Details

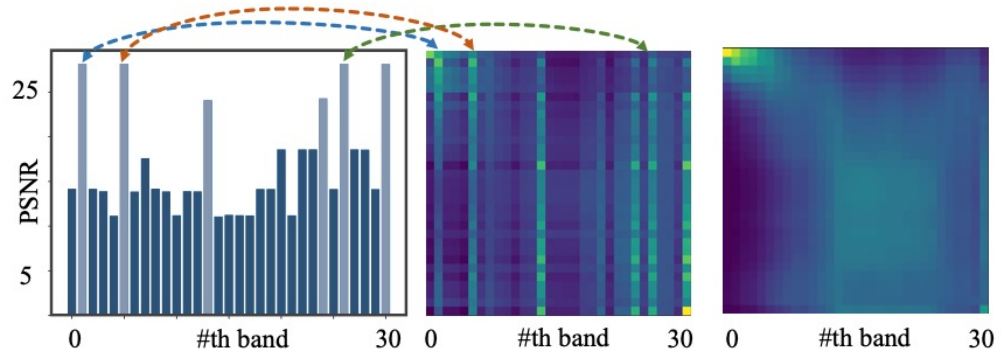
	Input	3D	Dim	QK Proj	V Proj	Attn
MDTA	HWC	×	C	Linear + DWConv	Linear+ DWConv	Reshape
MGSA	HWDC	✓	D	Linear + DWConv	Linear+ DWConv	Reshape
MS-MSA	HWC	×	C	Linear	Linear + Mask-Attn*	Reshape
GSA	HWC	×	C	Linear	Linear	Reshape
GSSA (Ours)	**HWDC**	**✓**	**D**	**None**	**Linear**	**Pooling**

(1) **Comparison** of different attention design



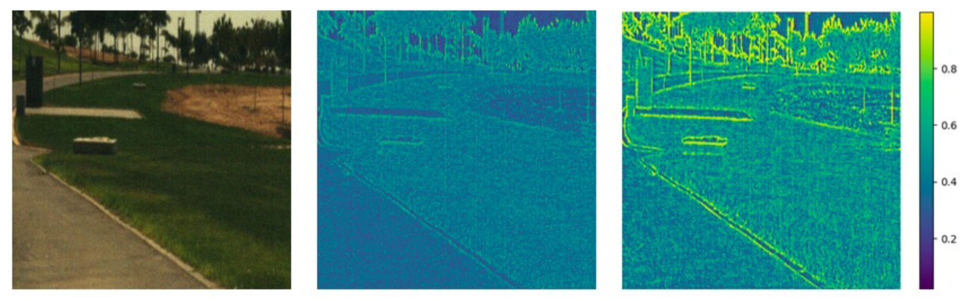
Conv3D S3Conv

(2) **S3Conv** produce discriminative features



(a) SNR of each band (b) Attn w/ LQ (c) Attn w/o LQ

(3) **GSSA** guides the model to pay attention to more informative bands with higher SNR.

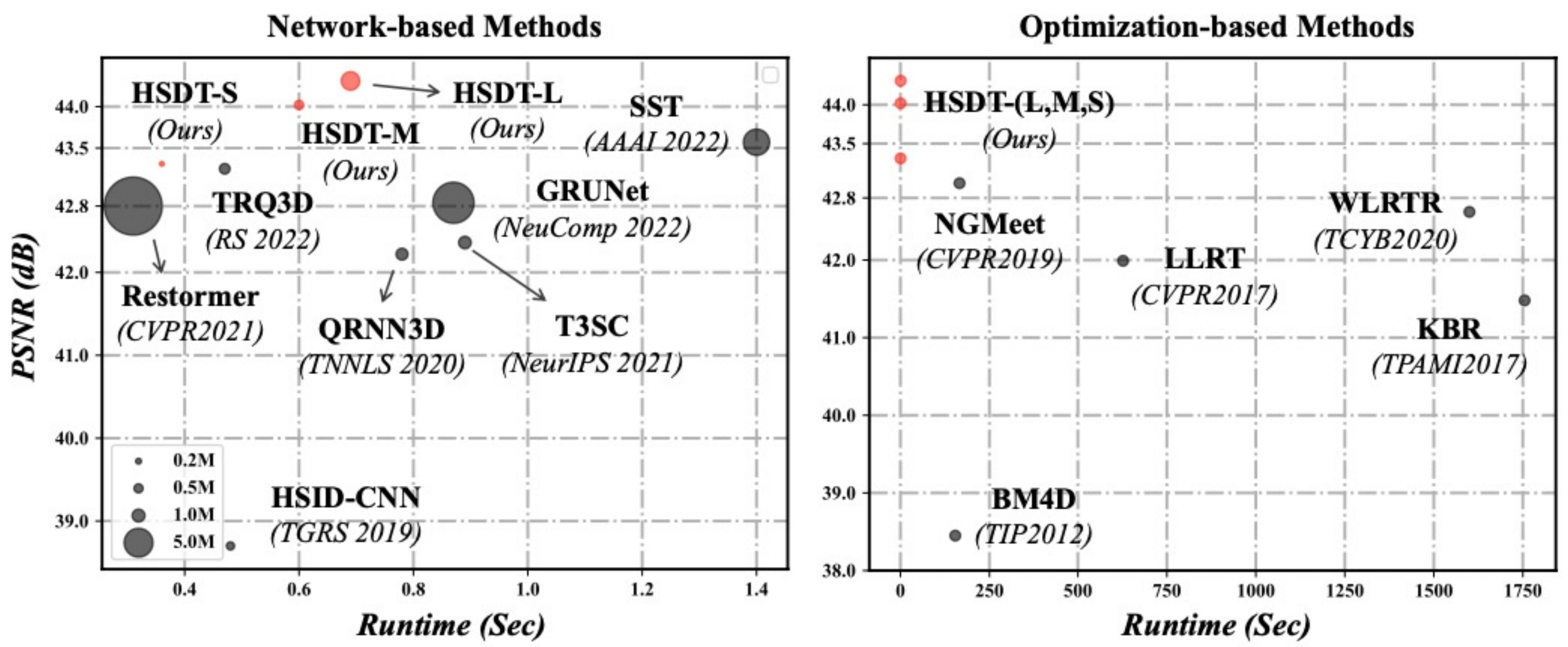


(a) Noisy HSI (b) Feature map (c) Modulation Weight

(4) **SM-FFN** amplifies the features in high-information-density regions.

Experiments

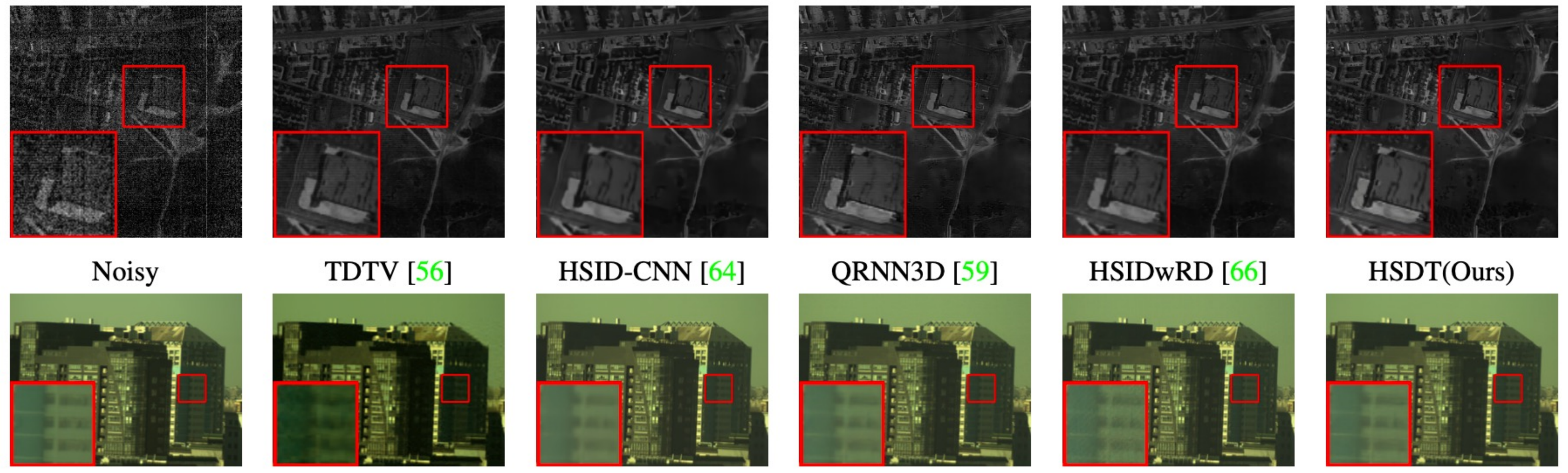
Runtime Comparison



Quantitative Comparison

	Params (M)	Runtime (s)	Gaussian 50	Gaussian Blind	Complex Stripe	Complex Mixture	RealHSI
Optimization SOTA	-	166	40.26 dB	42.23 dB	37.67 dB	34.77 dB	31.14 dB
Network SOTA	4.14	1.4	41.41 dB	42.81 dB	41.27 dB	39.19 dB	31.23 dB
HSDT-S (Ours)	**0.13**	0.36	41.16 dB	42.57 dB	41.11 dB	40.22 dB	-
HSDT-M (Ours)	0.52	0.60	41.82 dB	43.32 dB	41.28 dB	40.46 dB	-
HSDT-L (Ours)	2.09	**0.69**	**42.09** dB	**43.59** dB	**42.02** dB	**41.07** dB	**31.42** dB

Visual Comparison



Noisy, TDTV [56], HSID-CNN [64], QRNN3D [59], HSIDwRD [66], HSDT(Ours)

Ground Truth, TDTV [56], HSID-CNN [64], QRNN3D [59], HSIDwRD [66], HSDT(ours)