MST++: Multi-stage Spectral-wise Transformer for Efficient Spectral Reconstruction

Winner of NTIRE 2022 Challenge on Spectral Reconstruction from RGB

Yuanhao Cai[1,*], Jing Lin[1,*], Zudi Lin[2], Haoqian Wang[1]
Hanspeter Pfister[2], Yulun Zhang[3], Radu Timofte[3,4], Luc Van Gool[3]

Shenzhen International Graduate School, Tsinghua University[1]
Harvard University[2] ETH Zürich[3] JMU Würzburg[4]

Overview

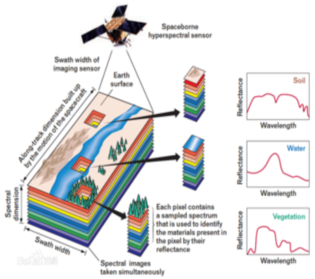
- Introduction
- Method
- Experiments

Introduction

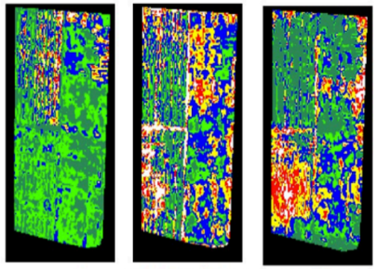
Introduction

Application

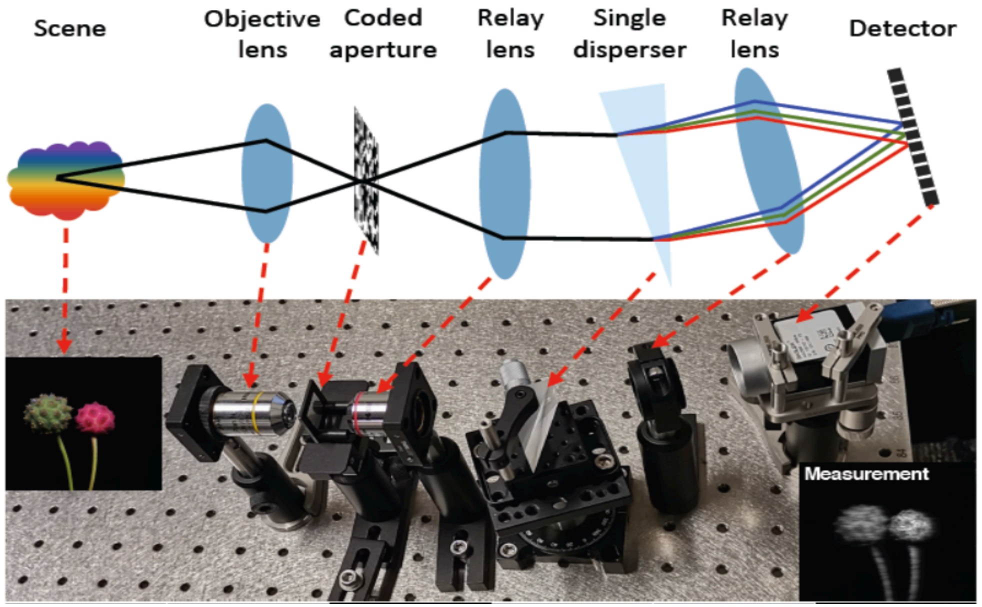
Medical Imaging



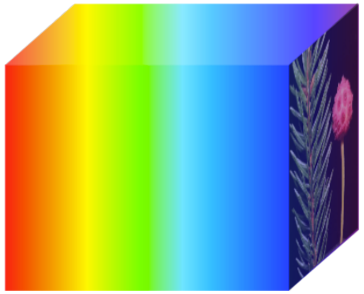
Object Tracking



Remote Sensing



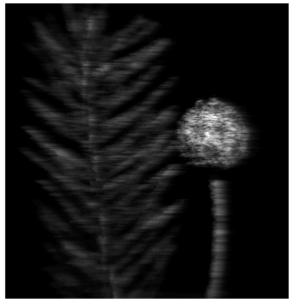
Coded Aperture Snapshot Spectral Imaging (CASSI) System



3D HSI Cube

Compress

Reconstruct



Compressed Measurement

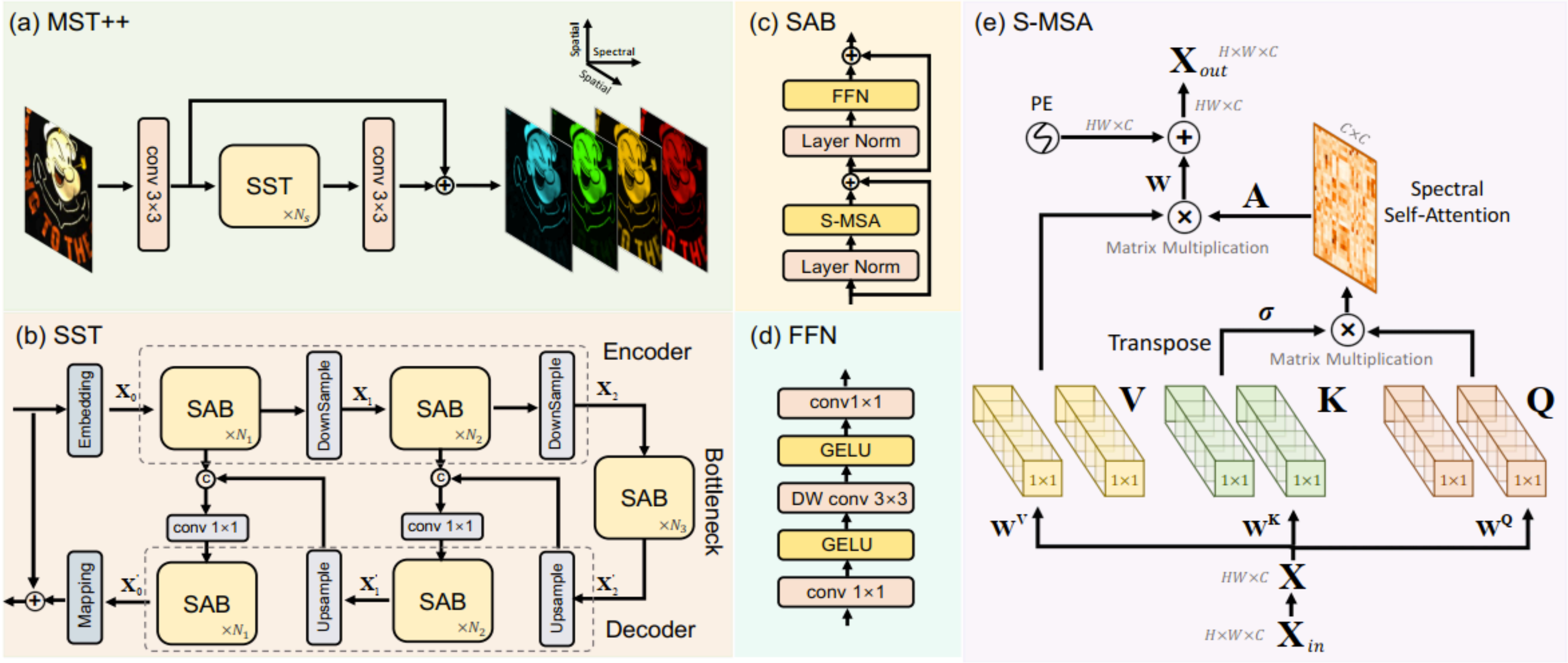
Model-based Methods

- Time consuming
- Poor representing capacity
- Unsatisfactory performance

CNN-based Methods

- Show limitations in capturing long-range dependencies

Method



Network Architecture

- Cascaded by N_s SSTs
- Each SST adopts a three-level U-shaped structure

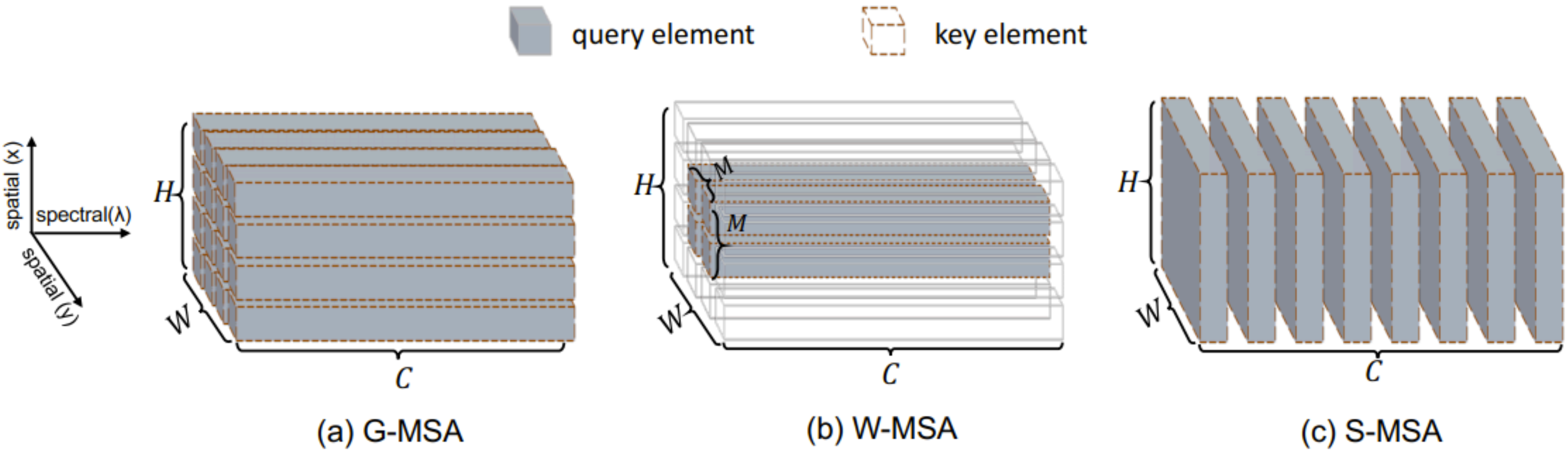
Spectral-wise MSA

- Based on HSI characteristic
- Treats each spectral feature map as a token to calculate self-attention

$$\mathbf{Q} = \mathbf{X}\mathbf{W}^{\mathbf{Q}}, \mathbf{K} = \mathbf{X}\mathbf{W}^{\mathbf{K}}, \mathbf{V} = \mathbf{X}\mathbf{W}^{\mathbf{V}}$$

$$\mathbf{A}_j = \mathrm{softmax}(\sigma_j \mathbf{K}_j^{\mathrm{T}} \mathbf{Q}_j), \quad head_j = \mathbf{V}_j \mathbf{A}_j$$

$$\text{S-MSA}(\mathbf{X}) = \big(\underset{j=1}{\overset{N}{\mathrm{Concat}}}(head_j)\big)\mathbf{W} + f_p(\mathbf{V})$$



We compare computational complexity of S-MSA, global MSA, and window-based MSA as

$$O(\mathrm{G-MSA}) = 2(HW)^2C, \qquad O(\mathrm{W-MSA}) = \frac{HW}{M^2}(2(M^2)^2C) = 2M^2HWC,$$

$$O(\mathrm{S-MSA}) = N((C/N)^2HW + (C/N)^2HW)^2 = \frac{2HWC^2}{N}$$

Only our S-MSA can enjoy linear computational complexity and global receptive fields jointly.

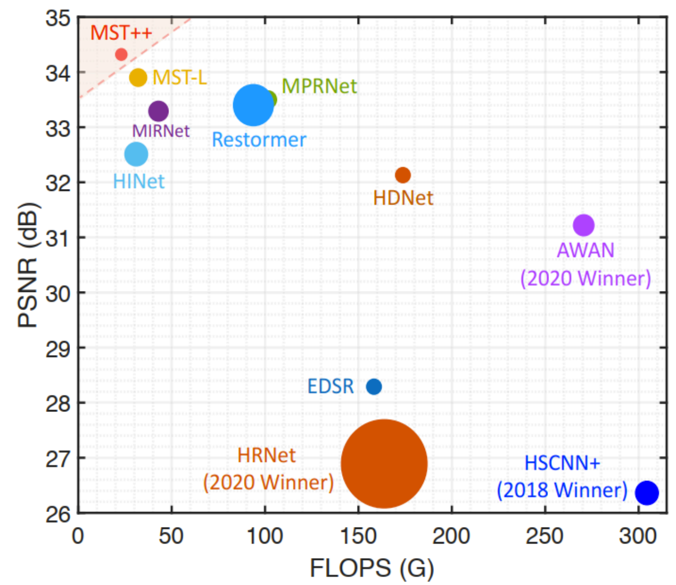
Experiments

Experiments

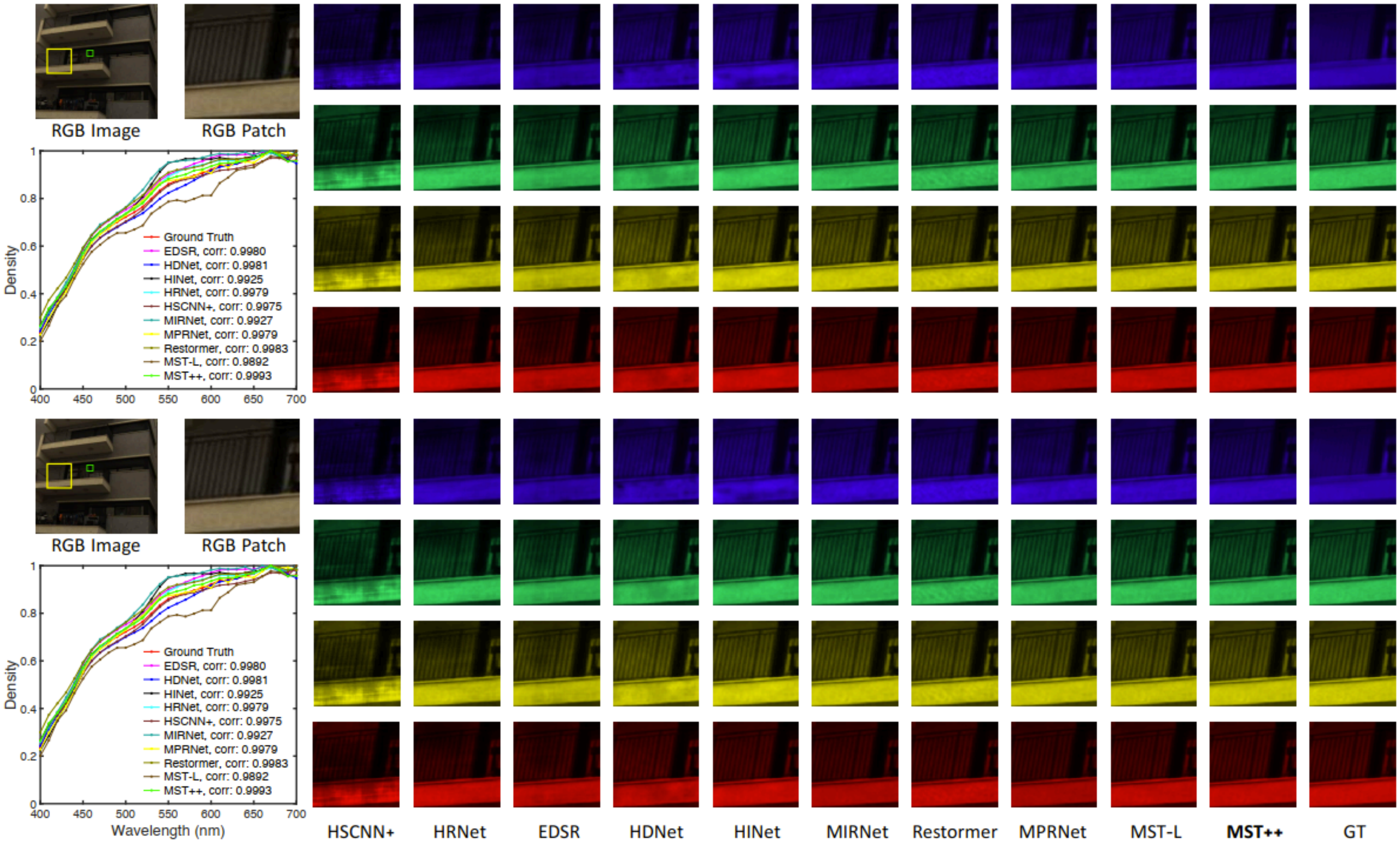
	NTIRE 2022 HSI Dataset - `Valid`					NTIRE 2022 HSI Dataset - `Test`		
Method	Params (M)	FLOPS (G)	MRAE	RMSE	PSNR	Username	MRAE	RMSE
HSCNN+ [67]	4.65	304.45	0.3814	0.0588	26.36	pipixia	0.2434	0.0411
HRNet [88]	31.70	163.81	0.3476	0.0550	26.89	uslab	0.2377	0.0391
EDSR [45]	2.42	158.32	0.3277	0.0437	28.29	orange_dog	0.2377	0.0376
AWAN [36]	4.04	270.61	0.2500	0.0367	31.22	askldklasfj	0.2345	0.0361
HDNet [29]	2.66	173.81	0.2048	0.0317	32.13	HSHAJii	0.2308	0.0364
HINet [21]	5.21	31.04	0.2032	0.0303	32.51	ptdoge_hot	0.2107	0.0365
MIRNet [84]	3.75	42.95	0.1890	0.0274	33.29	test_pseudo	0.2036	0.0324
Restormer [83]	15.11	93.77	0.1833	0.0274	33.40	gkdgkd	0.1935	0.0322
MPRNet [85]	3.62	101.59	0.1817	0.0270	33.50	deeppf	0.1767	0.0322
MST-L [13]	2.45	32.07	0.1772	0.0256	33.90	mialgo_ls	0.1247	0.0257
MST++	**1.62**	**23.05**	**0.1645**	**0.0248**	**34.32**	**MST++***	**0.1131**	**0.0231**

Our MST++ significantly outperforms SOTA CNN-based methods with cheaper memory and computational costs

We plot PSNR-Params-FLOPS comparisons, where MST++ takes up the upper-left corner



Experiments



Experiments

Method	Baseline	SW-MSA	W-MSA	G-MSA	**S-MSA**
MRAE	0.3177	0.2839	0.2624	0.1821	**0.1645**
RMSE	0.0453	0.0399	0.0375	0.0271	**0.0248**
Params (M)	1.30	1.60	1.60	1.60	1.62
FLOPS (G)	17.68	24.10	24.10	25.11	23.05

(a) Ablation study of different self-attention mechanisms.

N_s	1	2	3	4
MRAE	0.1761	0.1716	**0.1645**	0.1711
RMSE	0.0266	0.0269	**0.0248**	0.0265
Params (M)	0.55	1.08	1.62	2.16
FLOPS (G)	8.10	15.57	23.05	30.52

(b) Ablation study of stage number N_s.

Ablation study of different MSA modules

Our S-MSA achieves the most significant improvement while requiring cheapest memory and computational costs

Ablation study of stage number N_s

The performance is improved as the stage number increases

Please note that MST++ is based on our previous work “Mask-guided Spectral-wise Transformer for Effcient Hyperspectral Image Reconstruction” that has been accepted by CVPR 2022 main conference

Thanks for Watching

MST ： https://github.com/caiyuanhao1998/MST

MST++ ： https://github.com/caiyuanhao1998/MST-plus-plus