Mask-guided Spectral-wise Transformer for Efficient Hyperspectral Image Reconstruction

Yuanhao Cai 1,2,*, Jing Lin 1,2,*, Xiaowan Hu 1,2,
Haoqian Wang 1,2,†, Xin Yuan 3, Yulun Zhang 4,
Radu Timofte 4, and Luc Van Gool 4
Shenzhen International Graduate School, Tsinghua University,
Shenzhen Institute of Future Media Technology,
Westlake University, ETH Z¨urich

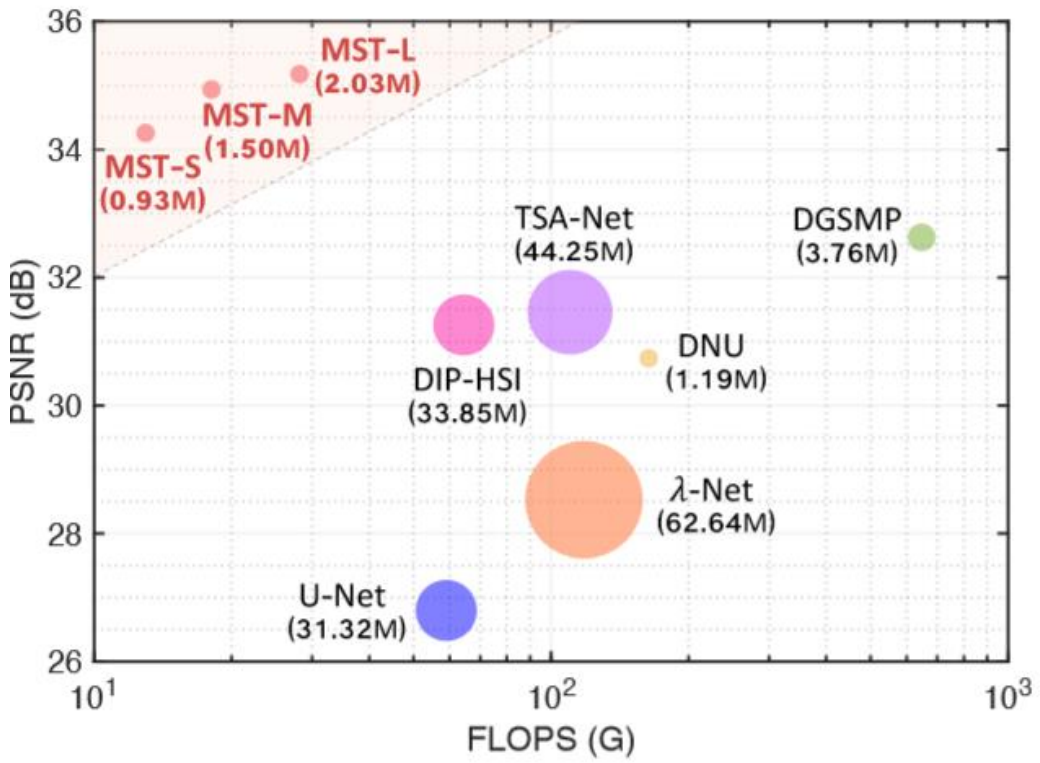
MST++: Multi-stage Spectral-wise Transformer For Efficient Spectral Reconstruction

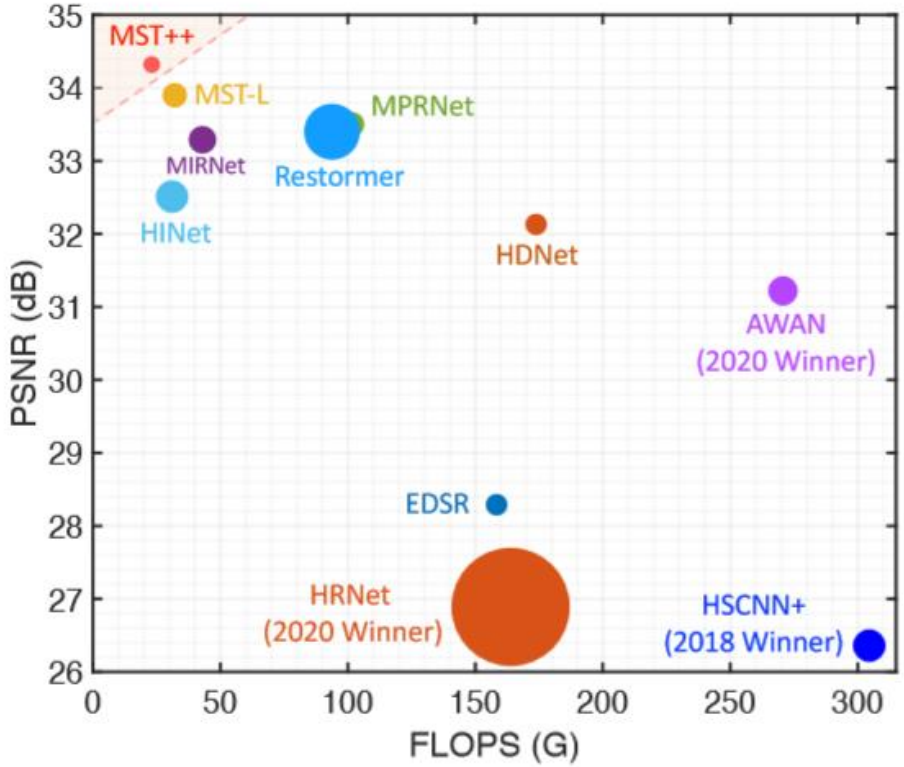
Yuanhao Cai 1,*, Jing Lin 1,*, Zudi Lin 2,
Haoqian Wang 1,†, Yulun Zhang 3, Hanspeter Pfister 2,
Radu Timofte 3,4, Luc Van Gool 3
Shenzhen International Graduate School,
Tsinghua University,
Harvard University, CVL, ETH Z¨urich, 4 CAIDAS, JMU W¨urzburg

Outline

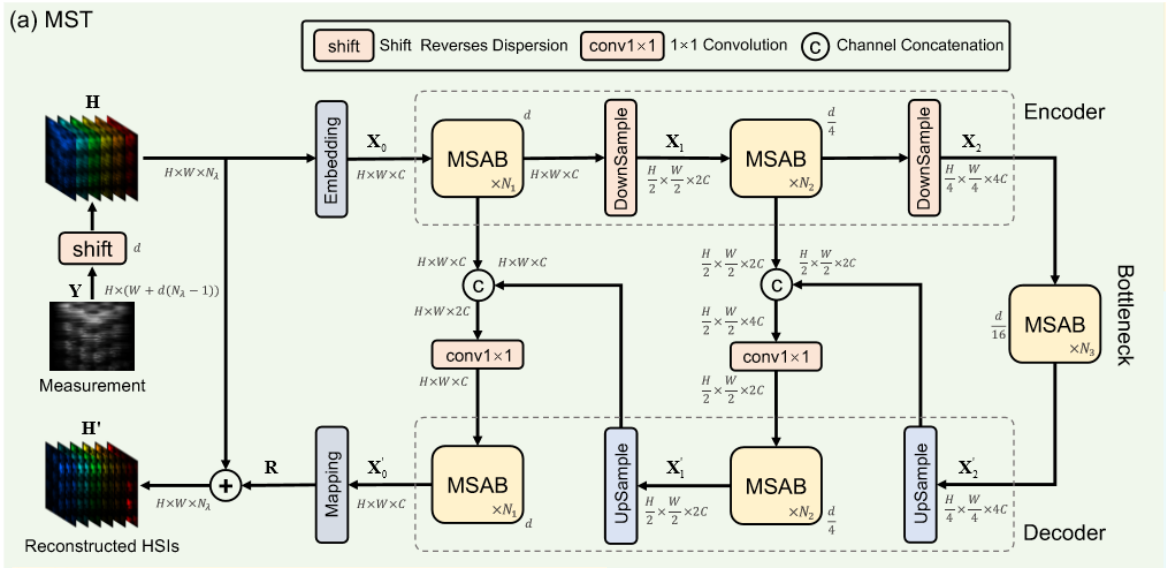
- Analysis
  - Propose a new method, MST, for HSI reconstruction.
  - Present a novel self-attention, S-MSA, to capture the inter-spectra similarity and dependencies of HSIs.
  - Our MST dramatically outperforms SOTA methods on all scenes in simulation while requiring much cheaper Params and FLOPS.
- Analysis
  - Proposed a novel framework, MST++, for SR.
  - Validate a series of natural image restoration models on this SR task.
  - Quantitative and qualitative experiments demonstrate that our MST++ dramatically outperforms SOTA methods while requiring much cheaper Params and FLOPS..
- Experiments

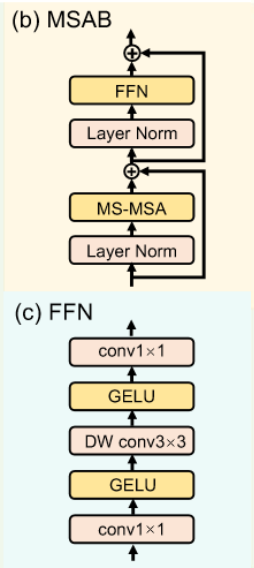
Outline

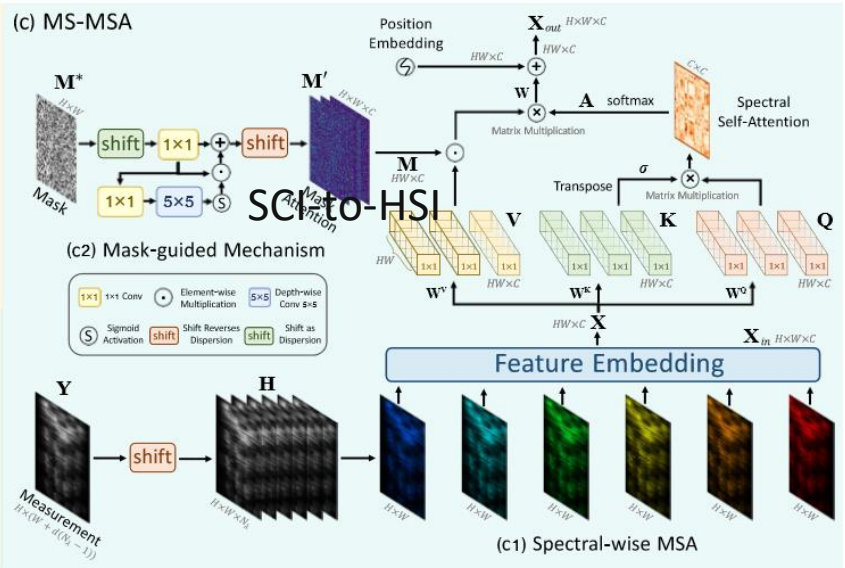

MST-L
(2.03M)
MST-M
(1.50M)
MST-S
(0.93M)
TSA-Net
(44.25M)
DGSMP
(3.76M)
DIP-HSI
(33.85M)
DNU
(1.19M)
λ-Net
(62.64M)
U-Net
(31.32M)
PSNR (dB)
FLOPS (G)


MST++
MST-L
MPRNet
MIRNet
Restormer
HINet
HDNet
AWAN
(2020 Winner)
EDSR
HRNet
(2020 Winner)
HSCNN+
(2018 Winner)
PSNR (dB)
FLOPS (G)

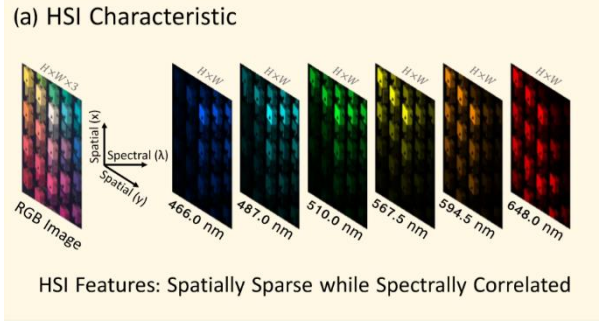
Mask-guided Spectral-wise Transformer (MST)

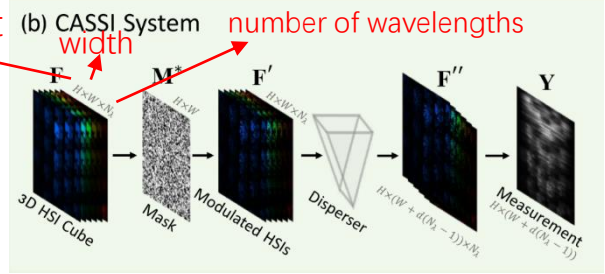






SCI to HSI





height
width
number of wavelengths

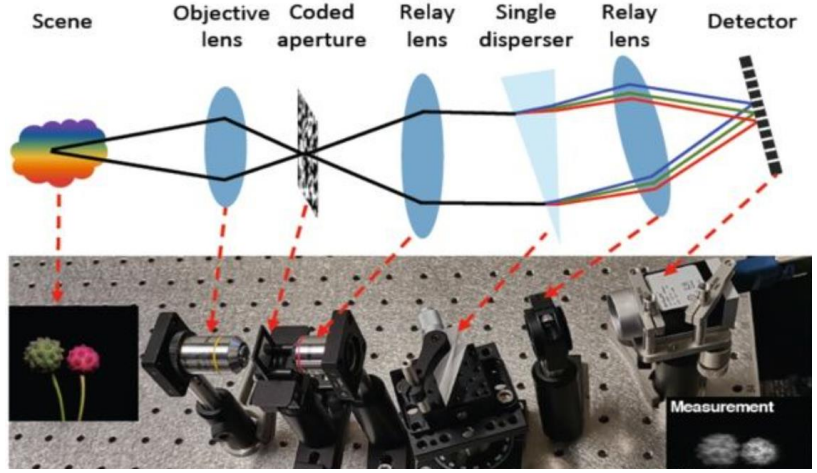
$$\mathbf{F}'(:,:,n_\lambda) = \mathbf{F}(:,:,n_\lambda) \odot \mathbf{M}^*$$

$$\mathbf{M}^* \in \mathbb{R}^{H \times W}$$

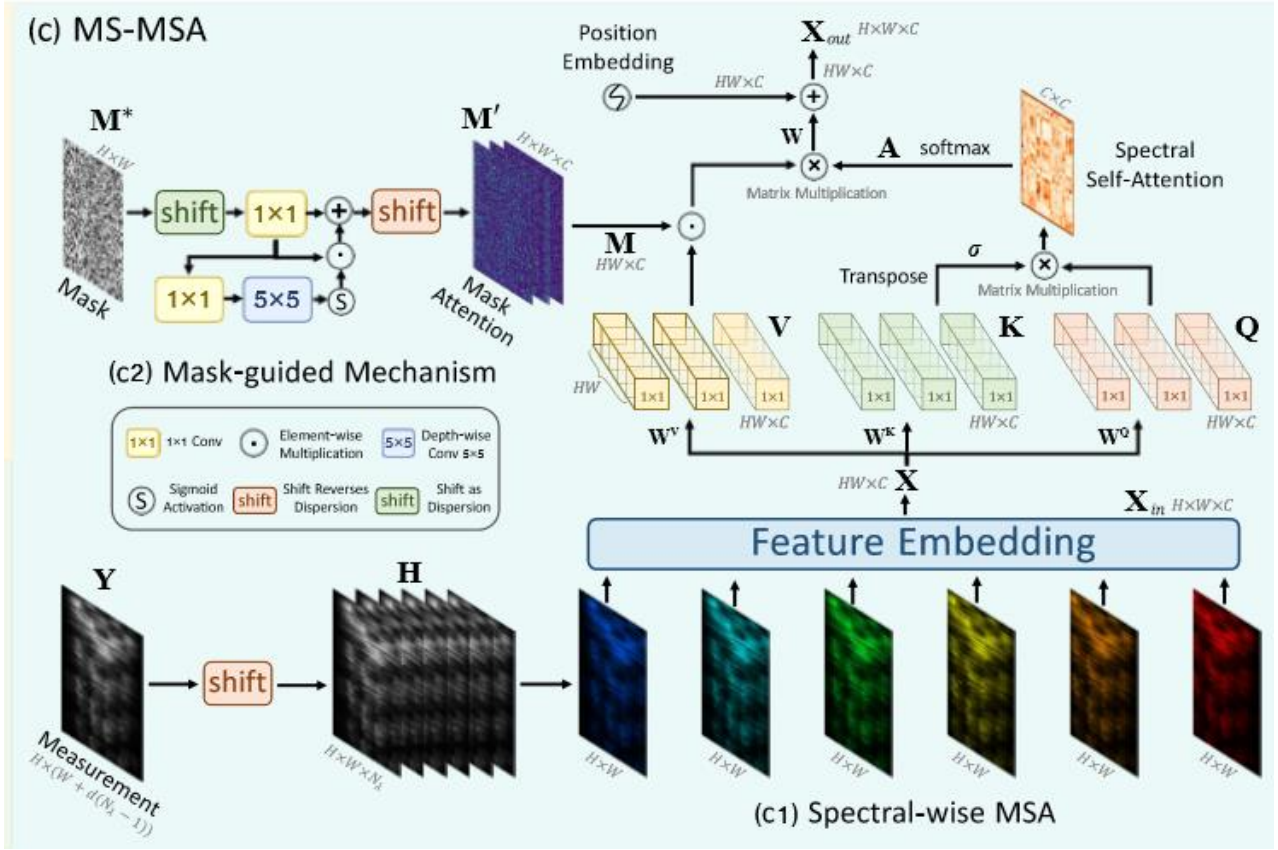
$$\mathbf{F}''(u,v,n_\lambda) = \mathbf{F}'(x, y + d(\lambda_n - \lambda_c), n_\lambda)$$

$$\mathbf{Y} = \sum_{n_\lambda=1}^{N_\lambda} \mathbf{F}''(:,:,n_\lambda) + \mathbf{G}$$

imaging noise



.Mask-guided Mechanism



$$\mathbf{M}_s(x, y, n_\lambda) = \mathbf{M}^*(x, y + d(\lambda_n - \lambda_c))$$

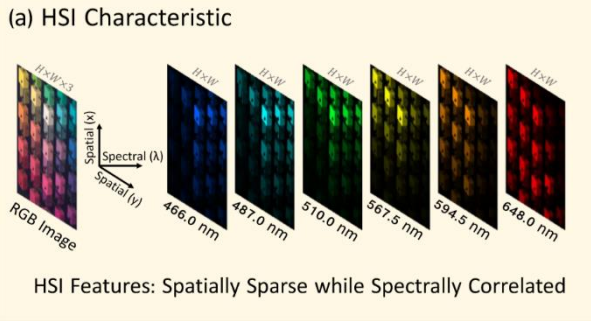
$$\mathbf{M}_s \in \mathbb{R}^{H\times(W+d(N_\lambda-1))\times N_\lambda}$$

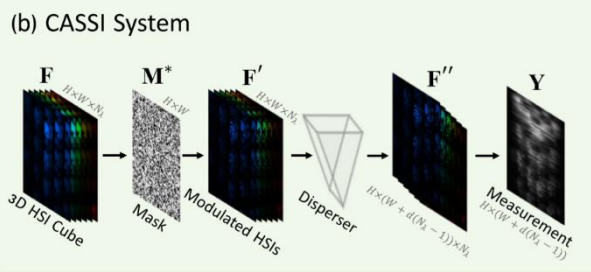
sigmoid activation　　mapping function

$$\mathbf{M}'_s = (\mathbf{W}_1\mathbf{M}_s) \odot (1 + \delta(f_{dw}(\mathbf{W}_2\mathbf{W}_1\mathbf{M}_s)),$$

learnable parameters of the two layers

Mask-guided Mechanism





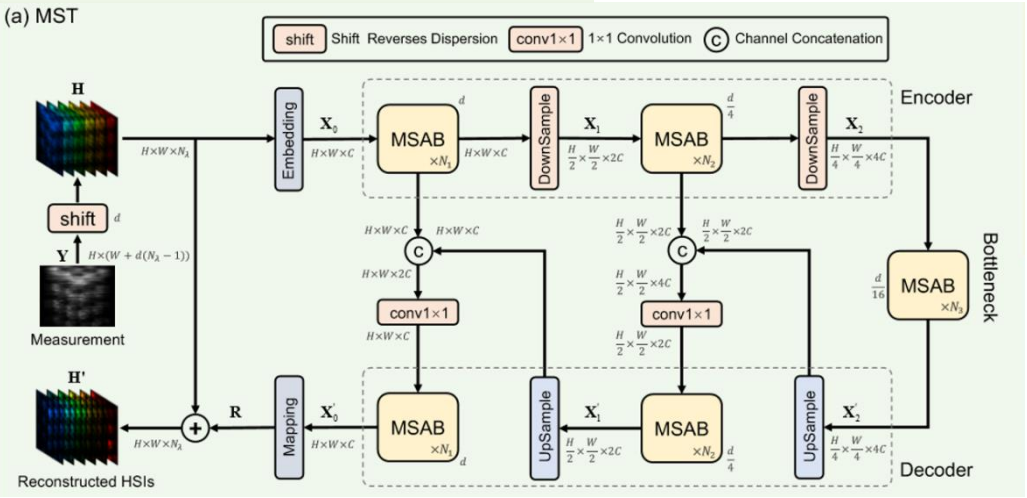
To spatially align the mask attention map $\mathbf{F}'$ $\mathbf{H}$

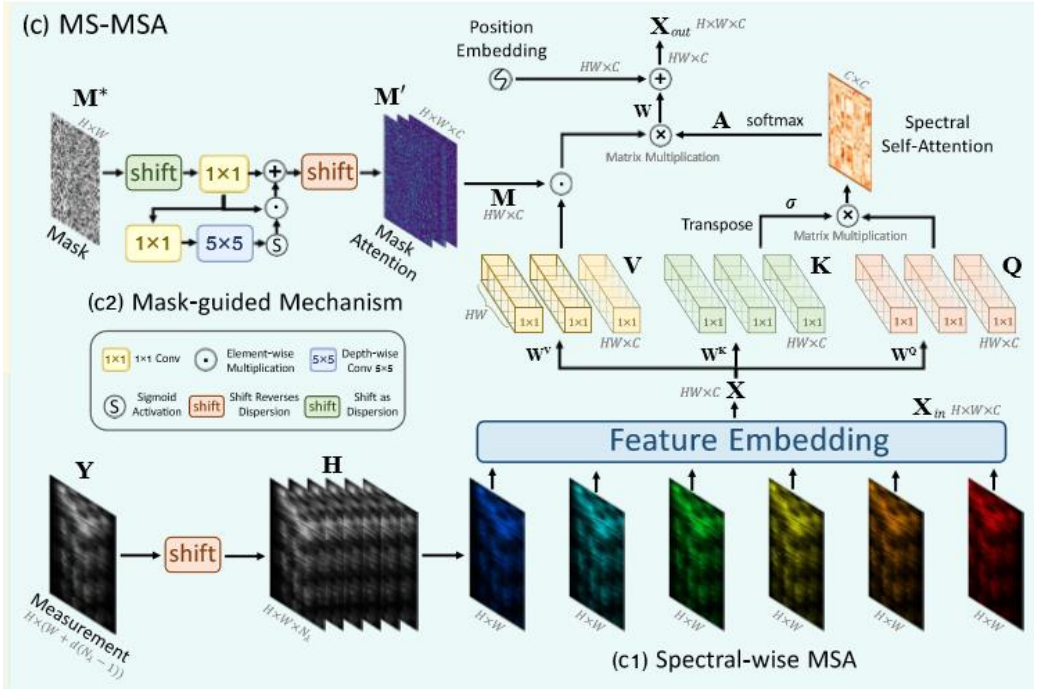
$$\mathbf{M}'(x, y, n_\lambda) = \mathbf{M}'_s(x, y - d(\lambda_n - \lambda_c), n_\lambda)$$

$\mathbf{M}' \longrightarrow \mathbf{M} \in \mathbb{R}^{HW \times C}$ to match the dimensions of V

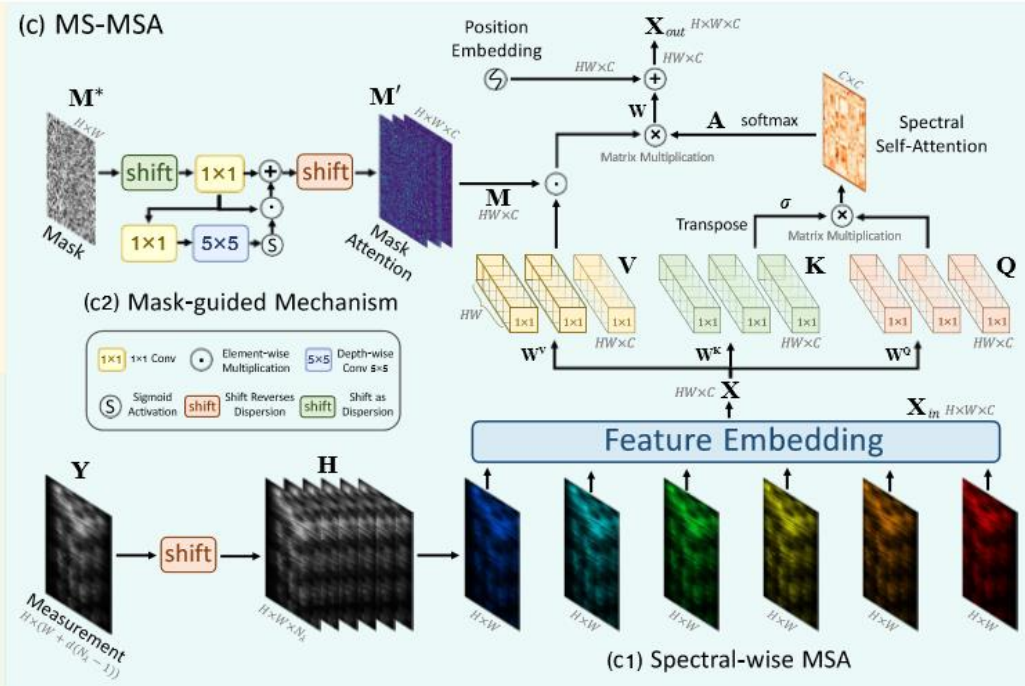
$$\mathbf{M} = [\mathbf{M}_1, \ldots, \mathbf{M}_N].$$

$$head_j = (\mathbf{M}_j \odot \mathbf{V}_j)\mathbf{A}_j.$$





Spectral-wise Multi-head Sefl-Attention (S-MSA)



$$\mathbf{X}_{in} \in \mathbb{R}^{H\times W\times C}$$
$$\mathbf{X} \in \mathbb{R}^{HW\times C}$$

$$\mathbf{V} \in \mathbb{R}^{HW\times C} \quad \mathbf{K} \in \mathbb{R}^{HW\times C} \quad \mathbf{Q} \in \mathbb{R}^{HW\times C}$$

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

$$\mathbf{Q}=[\mathbf{Q}_1,\ldots,\mathbf{Q}_N], \quad \mathbf{K}=[\mathbf{K}_1,\ldots,\mathbf{K}_N] \quad \mathbf{V}=[\mathbf{V}_1,\ldots,\mathbf{V}_N]$$

$$d_h = \frac{C}{N}$$

treats each spectral representation as a token and calculates the self-attention for each headj

analyze the computational complexity

$$O(\text{S-MSA}) = \frac{2HWC^2}{N}, \; O(\text{G-MSA}) = 2(HW)^2C,$$
$$O(\text{W-MSA}) = 2(M^2)^2(\frac{HW}{M^2})C = 2M^2HWC,$$

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

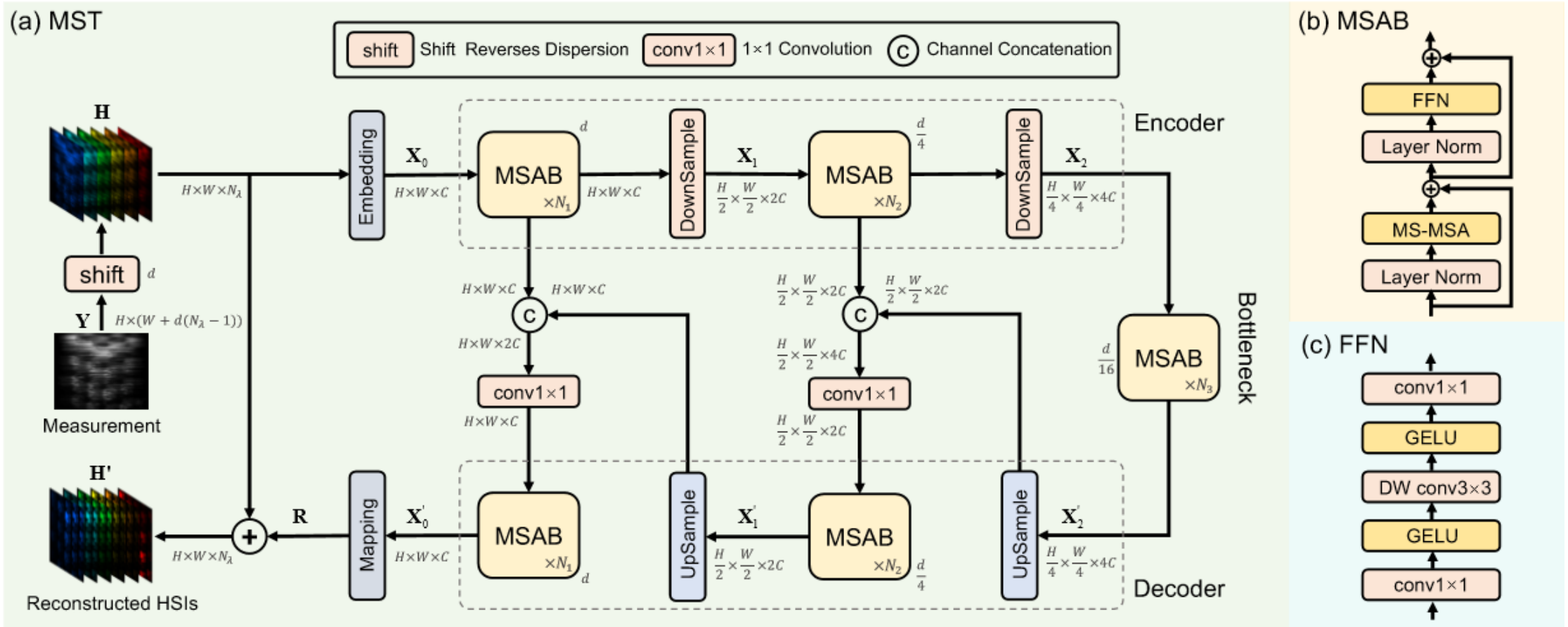
$$\text{S-MSA}(\mathbf{X}) = \left(\underset{j=1}{\overset{N}{\text{Concat}}}(head_j)\right)\mathbf{W} + f_p(\mathbf{V})$$

$$\mathbf{X}_{out} \in \mathbb{R}^{H\times W\times C}$$

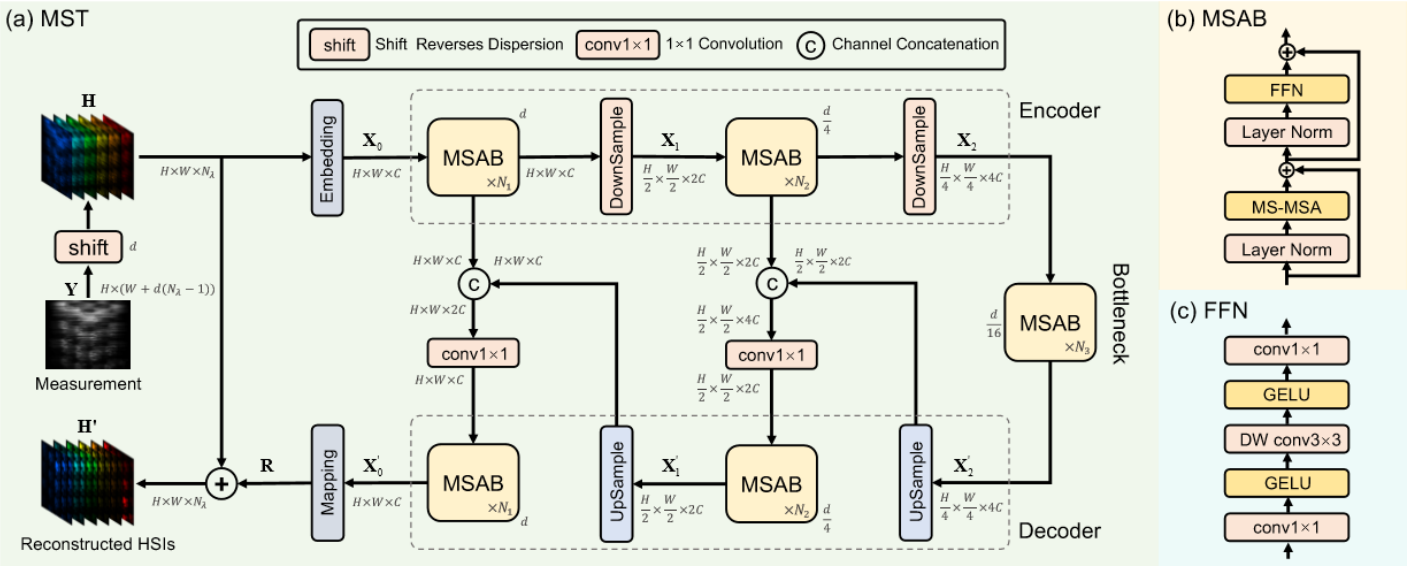
The receptive field of our S-MSA is global and not limited to position specific windows.

Mask-guided Spectral-wise Transformer (MST)

Mask-guided Self-Attention Block (MSAB)



Mask-guided Spectral-wise Transformer (MST)



$$\mathbf{F}''(u,v,n_\lambda) = \mathbf{F}'(x, y + d(\lambda_n - \lambda_c), n_\lambda)$$

$$\mathbf{H}(x,y,n_\lambda) = \mathbf{Y}(x, y - d(\lambda_n - \lambda_c)).$$

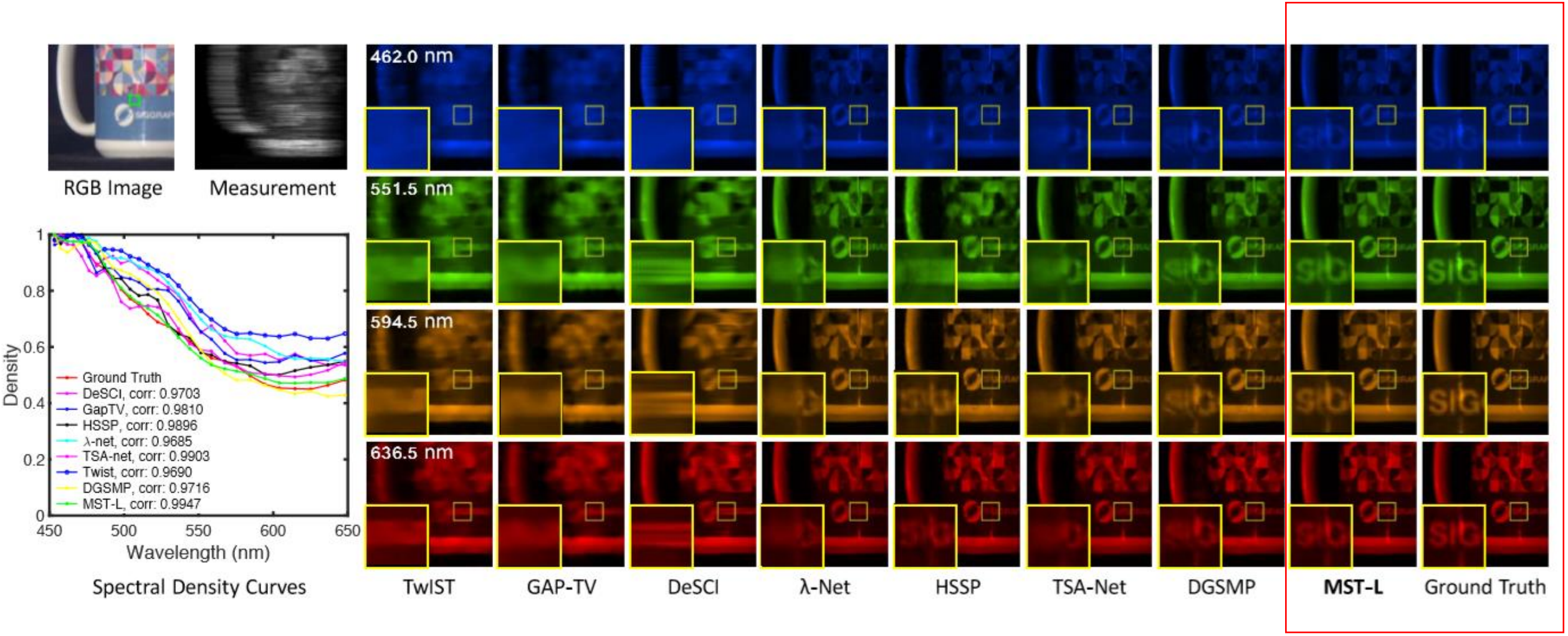
- MST exploits a conv3×3 (convolution with kernel size=3) layer to map H into feature X_0
- $\mathbf{X}_i \in \mathbb{R}^{\frac{H}{2^i} \times \frac{W}{2^i} \times 2^i C}$
- X2 passes through the bottleneck that consists of N3 MSABs.
- We follow the spirit of U-Net and design a symmetrical structure as the decoder.
- $\mathbf{H}' = \mathbf{H} + \mathbf{R}.$

Experiments

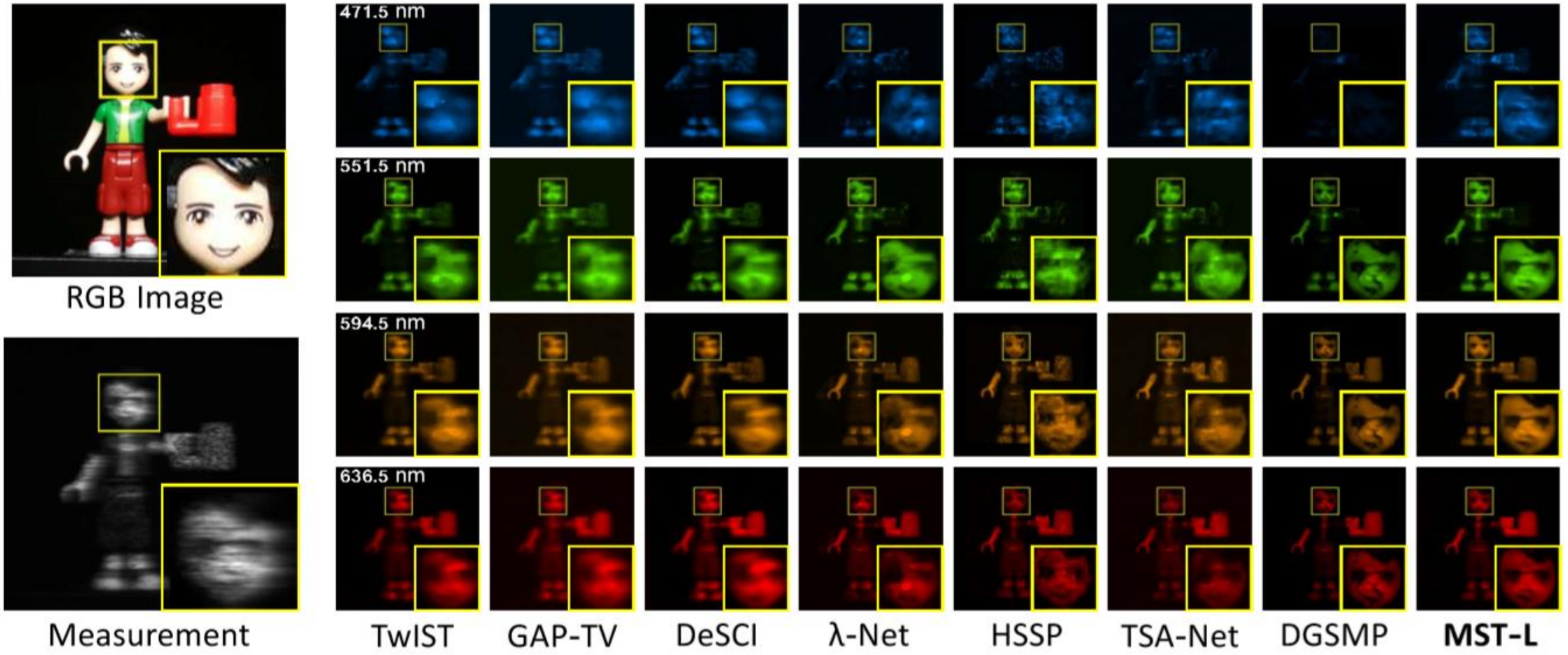
Scene	TwIST [4]		GAP-TV [57]		DeSCI [30]		λ-net [39]		HSSP [49]		DNU [50]		DIP-HSI [38]		TSA-Net [36]		DGSMP [20]		MST-S		MST-M		MST-L	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	25.16	0.700	26.82	0.754	27.13	0.748	30.10	0.849	31.48	0.858	31.72	0.863	32.68	0.890	32.03	0.892	33.26	0.915	34.71	0.930	35.15	0.937	**35.40**	**0.941**
2	23.02	0.604	22.89	0.610	23.04	0.620	28.49	0.805	31.09	0.842	31.13	0.846	27.26	0.833	31.00	0.858	32.09	0.898	34.45	0.925	35.19	0.935	**35.87**	**0.944**
3	21.40	0.711	26.31	0.802	26.62	0.818	27.73	0.870	28.96	0.823	29.99	0.845	31.30	0.914	32.25	0.915	33.06	0.925	35.32	0.943	36.26	0.950	**36.51**	**0.953**
4	30.19	0.851	30.65	0.852	34.96	0.897	37.01	0.934	34.56	0.902	35.34	0.908	40.54	0.962	39.19	0.953	40.54	0.964	41.50	0.967	**42.48**	**0.973**	42.27	**0.973**
5	21.41	0.635	23.64	0.703	23.94	0.706	26.19	0.817	28.53	0.808	29.03	0.833	29.79	0.900	29.39	0.884	28.86	0.882	31.90	0.933	32.49	0.943	**32.77**	**0.947**
6	20.95	0.644	21.85	0.663	22.38	0.683	28.64	0.853	30.83	0.877	30.87	0.887	30.39	0.877	31.44	0.908	33.08	0.937	33.85	0.943	34.28	0.948	**34.80**	**0.955**
7	22.20	0.643	23.76	0.688	24.45	0.743	26.47	0.806	28.71	0.824	28.99	0.839	28.18	0.913	30.32	0.878	30.74	0.886	32.69	0.911	33.29	0.921	**33.66**	**0.925**
8	21.82	0.650	21.98	0.655	22.03	0.673	26.09	0.831	30.09	0.881	30.13	0.885	29.44	0.874	29.35	0.888	31.55	0.923	31.69	0.933	32.40	0.943	**32.67**	**0.948**
9	22.42	0.690	22.63	0.682	24.56	0.732	27.50	0.826	30.43	0.868	31.03	0.876	34.51	0.927	30.01	0.890	31.66	0.911	34.67	0.939	35.35	0.942	**35.39**	**0.949**
10	22.67	0.569	23.10	0.584	23.59	0.587	27.13	0.816	28.78	0.842	29.14	0.849	28.51	0.851	29.59	0.874	31.44	0.925	31.82	0.926	32.53	0.935	**32.50**	**0.941**
Avg	23.12	0.669	24.36	0.669	25.27	0.721	28.53	0.841	30.35	0.852	30.74	0.863	31.26	0.894	31.46	0.894	32.63	0.917	34.26	0.935	34.94	0.943	**35.18**	**0.948**

Method	λ-net [39]	DNU [50]	DIP-HSI [38]	TSA-Net [36]	DGSMP [20]	MST-S	MST-M	MST-L
PSNR	28.53	30.74	31.26	31.46	32.63	**34.26**	**34.94**	**35.18**
SSIM	0.841	0.863	0.894	0.894	0.917	**0.935**	**0.943**	**0.948**
Params (M)	62.64	1.19	33.85	44.25	3.76	**0.93**	**1.50**	**2.03**
FLOPS (G)	117.98	163.48	64.42	110.06	646.65	**12.96**	**18.07**	**28.15**

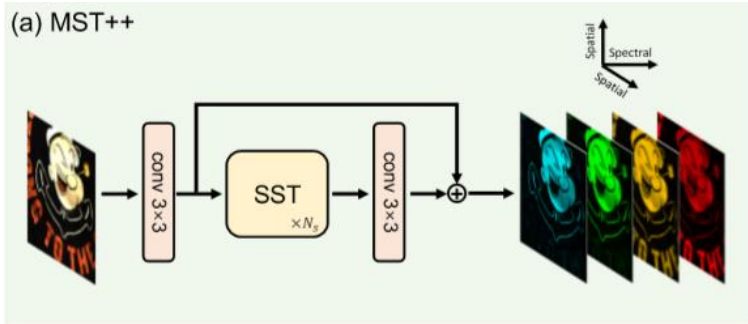
Experiments

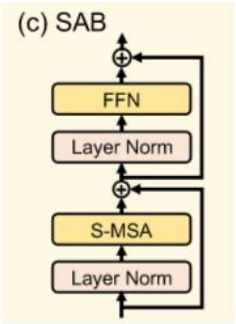


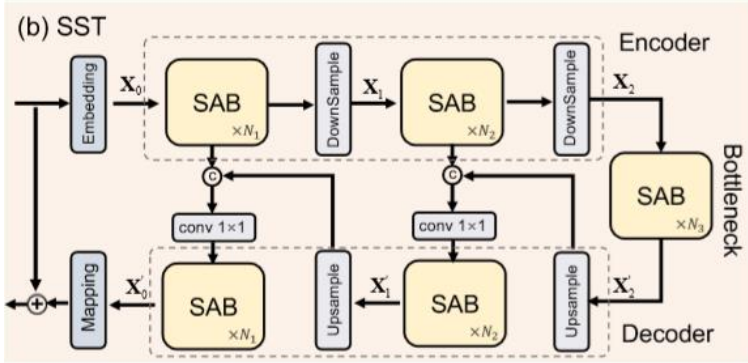
Experiments

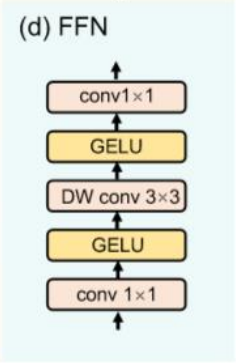


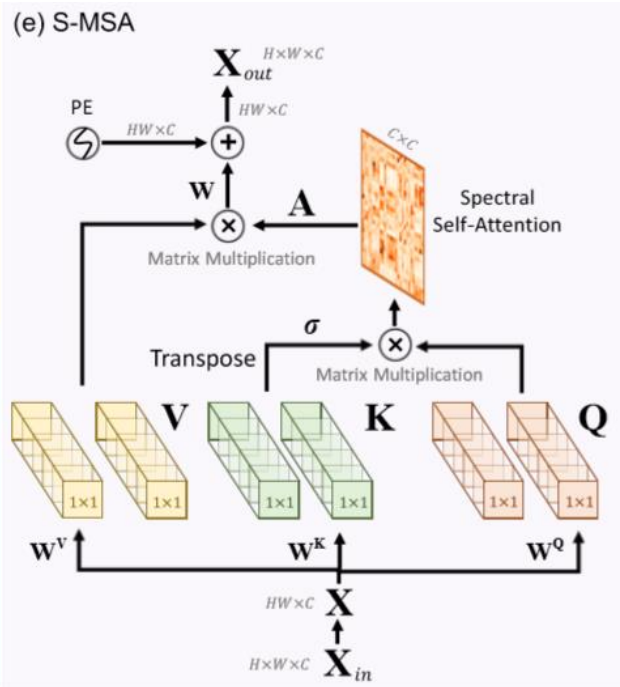
Mask-guided Spectral-wise Transformer (MST++)

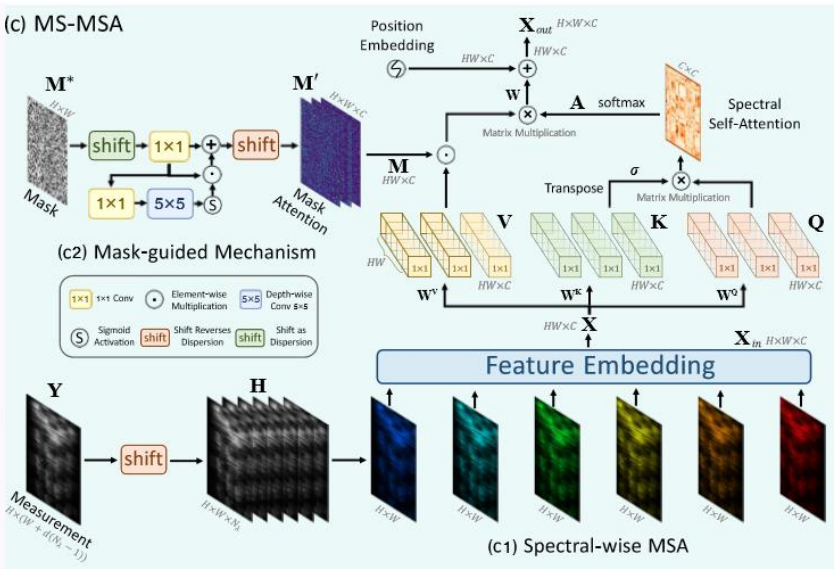












Experiments

	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**