# Lips-SpecFormer: Non-Linear Interpolable Transformer for Spectral Reconstruction using Adjacent Channel Coupling (Paper# 268)

Abhishek Kumar Sinha and S. Manthira Moorthi

Space Applications Center, Indian Space Research Organization, Ahmedabad

## Motivation

The motivation to propose Lips-SpecFormer is to overcome the limitations of using spectral wise self-attention to learn spectral dependencies. Firstly, To apply self-attention along the spectral dimension on the $X \in \mathbf{R}^{H \times W \times C}$ shaped feature map, the corresponding spectral attention coefficient using estimated key $K \in \mathbf{R}^{C \times HW}$ and query $Q \in \mathbf{R}^{C \times HW}$ is computed as $A_{ij} = \sum_{k=0}^{HW-1} Q_{i,k} K_{k,j}^T$. It squeezes the spatio-spectral context between two channels to a single scalar value causing the information loss. Second, the $L_2$ Lipschitz constant of self-attention is bounded by the variance of the input resulting in larger sensitivity.

## Architecture

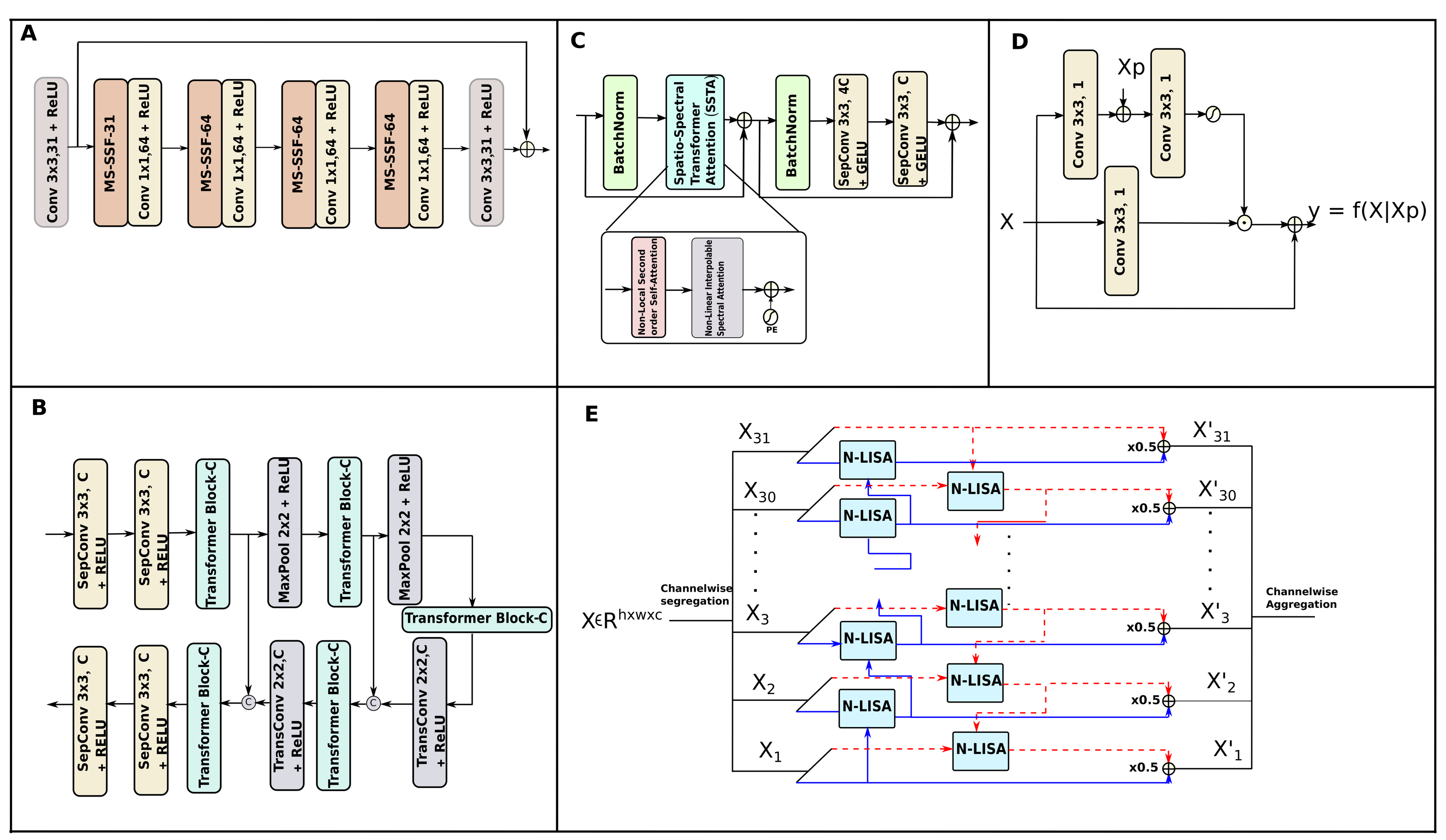


Figure 1. A: End-to-end transformer network. B: MSSSFB-C: Multi-Scale Spatio Spectral Feature Block with C number of input channels. C: Transformer block with C number of input channels. D: N-LISA: Non-Linear Interpolable Spectral Attention architecture. E: Architecture of spectral attention using N-LISA.

## Lipschitz constant of N-LISA

Let the loss function for transformer network be $\mathcal{L}_t = \rho_t + \frac{\gamma}{2} \|w_t\|_2^2$ at time $t$, where $\rho$ is the data fidelity term and $\gamma$ is $L_2$ regularisation parameter. Assume that there are $N$ numbers of $k \times k$ convolution filters in the neural net. The upper bound on the magnitude of Jacobian $\|J_{i,j}\|_2$ after $T$ iterations is given by,

$$\|J_{i,j}\|_2 \leq \left( \prod_{k=j+1}^{i} \sqrt{\frac{2\rho_T}{\gamma N}} \|V_k^F \odot \sigma(F_k^F(Q_k, K_{k-1}))\|_2 \right) \left\| \frac{\partial Y_j}{\partial X_j} \right\|_2, \ for\ i > j\ and,$$

$$\|J_{i,j}\|_2 \leq \left( \prod_{k=i}^{j-1} \sqrt{\frac{2\rho_T}{\gamma N}} \|V_k^B \odot \sigma(F_k^B(Q_k, K_{k+1}))\|_2 \right) \left\| \frac{\partial Y_j}{\partial X_j} \right\|_2, \ for\ i < j.$$

## Experiments and Results

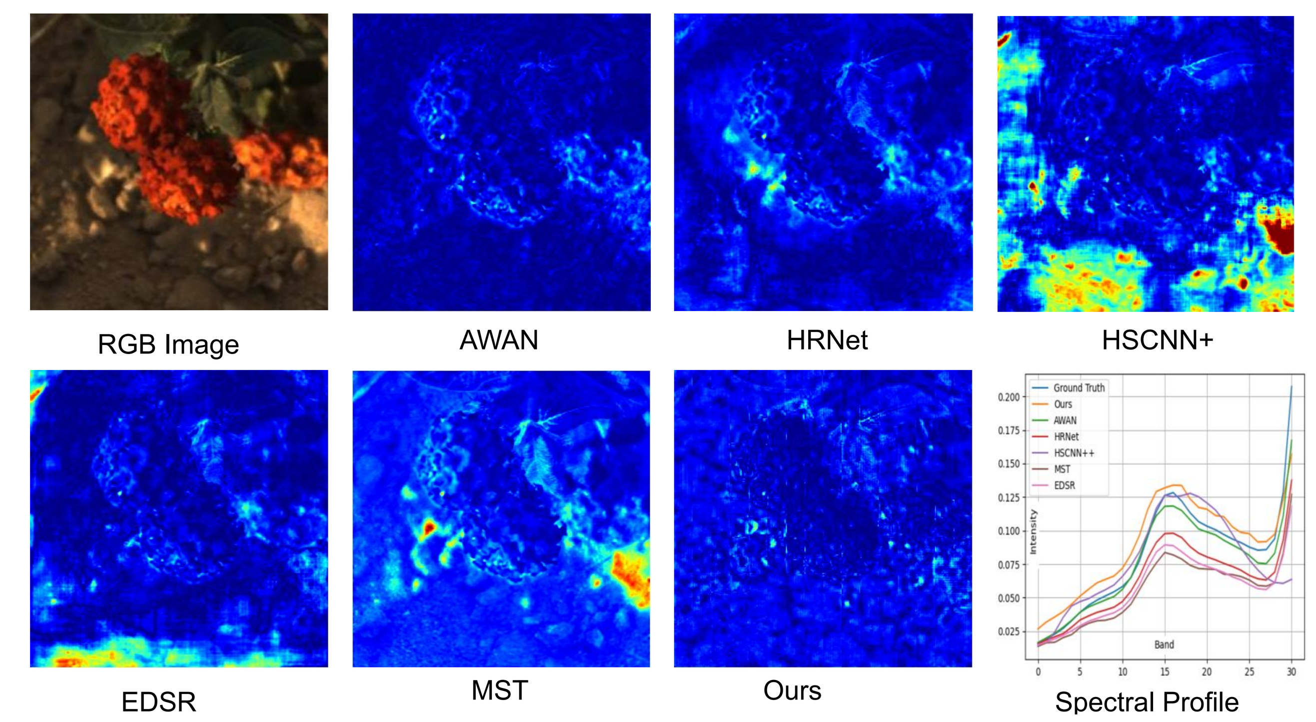
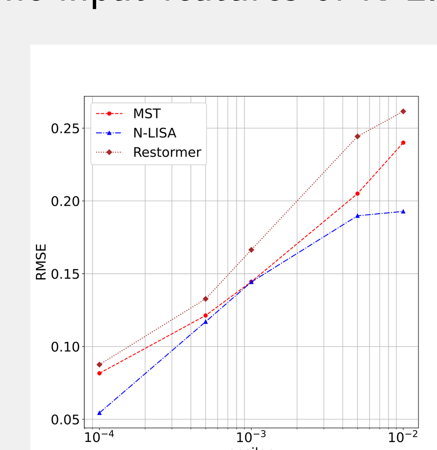



Figure 2. Illustration of residual map in the spectral band predicted by different methods. Spectral profile compares the spectral profiles generated by different methods.

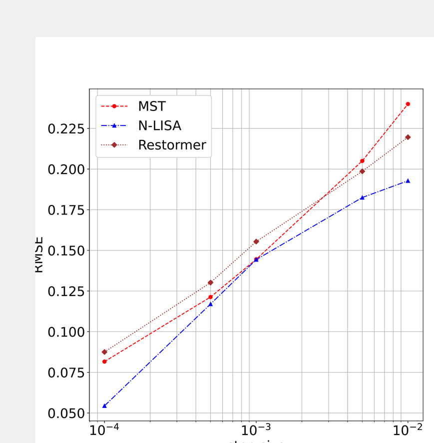
| Method | Params | Flops | CAVE | | NTIRE 2020 | | NTIRE 2022 | |
|---|---|---|---|---|---|---|---|---|
| | | | RMSE | SAM | MRAE | RMSE | MRAE | RMSE |
| Bicubic | - | - | 0.1689 | 34.382 | 0.1745 | 0.0506 | 0.2005 | 0.0712 |
| HSCNN+ | 4.65 | 266.84 | 0.0353 | 12.208 | 0.0684 | 0.0182 | 0.3814 | 0.0588 |
| HRNet | 31.70 | 143.51 | 0.0298 | 8.150 | 0.0682 | 0.0178 | 0.3476 | 0.0550 |
| EDSR | 2.42 | 142.53 | 0.0384 | 8.755 | 0.0707 | 0.0162 | 0.3277 | 0.0437 |
| AWAN | 4.04 | 231.29 | 0.0375 | 8.654 | 0.0678 | 0.0175 | 0.2500 | 0.0367 |
| HD-Net | 2.66 | 173.81 | 0.0326 | 8.314 | 0.0722 | 0.0176 | 0.2047 | 0.0317 |
| MPRNet | 3.62 | 101.59 | 0.0294 | 7.864 | 0.0722 | 0.0168 | 0.1817 | 0.0270 |
| MST | 2.45 | **26.29** | 0.0289 | 7.812 | 0.0747 | 0.0173 | 0.1772 | **0.0256** |
| **Ours** | **1.18** | 36.84 | **0.0246** | **7.661** | **0.0669** | **0.0158** | **0.1767** | 0.0301 |

## Adversarial Robustness

From previous studies such as Zhang et. al. [5], it is known that Lipschitz stability also imparts adversarial robustness. Therefore, we evaluate the performance of three transformer architectures for Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD) attacks. For both FGSM and PGD-20 attacks, we adjust the step-sizes and perturb the input features of N-LISA from NTIRE-2022 dataset.




(a) Robustness analysis under FGSM attack




(b) Robustness analysis under PGD-20 attack

Figure 3. Comparison of adversarial robustness under FGSM and PGD-20 attacks.

## N-LISA vs Spectral MSA

| Attention | NTIRE-2020 | | NTIRE-2022 | |
|---|---|---|---|---|
| | MRAE | RMSE | MRAE | RMSE |
| MSA | 0.1120 | 0.0420 | 0.2420 | 0.0512 |
| N-LISA | **0.0669** | **0.0158** | **0.1767** | **0.0301** |

Table 1. Quantitative results for different attention layers.

| $j$ | $\|\|J_{0,j}\|\|_2$ | $\|\|J_{5,j}\|\|_2$ | $\|\|J_{20,j}\|\|_2$ |
|---|---|---|---|
| 0 | 3.894 | 0.0011 | 0.0013 |
| 5 | 0.050 | 0.428 | 0.0011 |
| 20 | 0.00043 | 0.0008 | 0.426 |

Table 2. Some of 2-Lipschitz constant for spectral self-attention of MST for different perturbed channel.

| $j$ | $\|\|J_{0,j}\|\|_2$ | $\|\|J_{5,j}\|\|_2$ | $\|\|J_{20,j}\|\|_2$ |
|---|---|---|---|
| 0 | 0.431 | 0 | 0 |
| 5 | 0 | 0.770 | 0 |
| 20 | 0 | 0 | 0.776 |

Table 3. Some of 2-Lipschitz constant for spectral self-attention of N-LISA for different perturbed channel.

Tables 2 and 3 show the estimated Lipschitz constants for multihead-self-attention and N-LISA respectively. While 2-Lipschitz constant of diagonal Jacobian elements are found to be comparable, unlike multihead spectral self-attention, any perturbation in a given channel is not propagated to other channels in the feature maps of N-LISA.

## References and Contact

**Contact**: aks@sac.isro.gov.in

[1] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. 2022.

[2] Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, and Fei Liu. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from rgb images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1894–1903, 2020.

[3] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017.

[4] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 518–525, 2017.

[5] Bohang Zhang, Du Jiang, Di He, and Liwei Wang. Rethinking lipschitz neural networks and certified robustness: A boolean function perspective. In *Conference on Neural Information Processing Systems*, 2022.