TSP-Mamba: The Travelling Salesman Problem Meets Mamba for Image Super-resolution and Beyond – Supplementary Materials

Kun Zhou[1] Xinyu Lin[2] Jiangbo Lu[1]*
[1]SmartMore Corporation [2]SSE, CUHK-Shenzhen
kunzhou@link.cuhk.edu.cn, linxinyu0327@gmail.com, jiangbo.lu@gmail.com

1. More Ablation Studies

In this section, we will discuss more fine-gain designs and comprehensive ablated analysis of our proposed TSP-Mamba, including the choices of starting position, overlapped or non-overlapped window partition, TSP-LSTM without global 2D-SSM and other local scanning schemes (i.e., Hilbert, Zigzag and Z-shaped pipelines). Later on, we will further evaluate the impacts of either ablating TSP-LSTM or replacing it with window attentions.

[**a. Starting pos.:**] We compare our default top-left setup (TSP) with two counterparts starting from the center pixel (**TSP-C$_1$** (1,1) and **TSP-C$_2$** (2,2) for a 4×4 patch). TSP-C$_{1,2}$ yield similar results to our default version (termed as TSP). [**b. Partition:**] For efficiency, we use a *non-overlapping partition* scheme in TSP-LSTM. The overlapping version (**O-TSP**) shows only a slight improvement (0.02dB↑) but larger FLOPs. [**c. Multi TSP-LSTM** (Multi-TSP-LSTM) **without 2D-SSM:**] Stacking multiple TSP-LSTM layers may lack global modeling ability. As can be seen in Tab. below, Multi-TSP-LSTM gets degraded performance. This experiment underscores the necessity of integrating local TSP-LSTM with global 2D-SSM. [**d. Scanning:**] We trained three additional models replacing TSP with Hilbert, Zigzag, and Z-Shaped scannings. Results below show TSP (*content-adaptive*) outperforms the other three *content-invariant* scanning methods with only extra 0.8G FLOPs for building TSP paths. *Models in **bold** use the TSP-Mamaba-Tiny structure (431K) but different scan strategies.

Urban100 ×4	**TSP-C$_1$**	**TSP-C$_2$**	**O-TSP**	Multi-TSP-LSTM	**TSP**
PSNR (Param.)	26.60	26.59	26.63	26.24(433K)	26.61
FLOPs (G)	16.2	16.2	18.5	20.7	16.2

Urban100 ×4	**Hilbert Scan**	**Zigzag**	**Z-Shaped**	**TSP**
PSNR/FLOPs(G)	26.49/15.4	26.41/15.4	26.49/15.4	26.61/16.2

We trained two additional models: (i) **M9**, which drops TSP-LSTM, (ii) **M10**, which replaces TSP-LSTM with 4×4 window attention. The results below show that M8 (default setting in Tab.3) is 0.48dB/0.12dB higher than M9 and M10.

*Corresponding author

Urban100 ×4	**M9** (no TSP-LSTM)	**M10** (Win. Attn.)	**M8** (TSP-LSTM)
PSNR/SSIM	26.13/0.7881	26.49/0.7976	26.61/0.7992
Param./FLOPs(G)	398K / 12.1	442K / 21.7	431K / 16.2

2. Detailed Results of Classic SISR.

In our main paper, we report the averaged performance of different classic SISR methods, including SwinIR [4], SRFormer-L [5], HAT [1], RGT [2], MambaIR-L [3]. Here, we provide the detailed result on each benchmark.

Class SISR (×2)	Set5	Set14	BSD100	Urban100	Manga109
SwinIR	38.42	34.46	32.53	33.81	39.92
SRFormer-L	38.51	34.44	32.57	34.09	40.07
HAT	**38.63**	**34.86**	**32.62**	34.45	40.26
RGT	38.59	34.83	**32.62**	**34.47**	**40.34**
MambaIR-L	38.57	34.67	32.58	34.15	40.28
TSP-Mamba-L	**38.78**	**35.09**	**32.69**	**34.70**	**40.64**
Class SISR (×3)	Set5	Set14	BSD100	Urban100	Manga109
SwinIR	34.97	30.93	29.46	29.75	35.12
SRFormer-L	35.02	30.94	29.48	30.04	35.26
HAT	35.07	31.08	29.54	30.23	35.53
RGT	**35.15**	**31.13**	**29.55**	**30.28**	**35.55**
MambaIR-L	35.08	30.99	29.51	29.93	35.43
TSP-Mamba-L	**35.26**	**31.49**	**29.58**	**30.54**	**35.94**
Class SISR (×4)	Set5	Set14	BSD100	Urban100	Manga109
SwinIR	32.92	29.09	27.92	27.46	32.03
SRFormer-L	32.93	29.08	27.94	27.68	32.21
HAT	33.04	**29.23**	**28.00**	27.97	32.48
RGT	**33.12**	**29.23**	**28.00**	**27.98**	**32.50**
MambaIR-L	33.03	29.20	27.98	27.68	32.32
TSP-Mamba-L	**33.31**	**29.57**	**28.05**	**28.24**	**33.00**

References

[1] Xiangyu Chen, Xintao Wang, Wenlong Zhang, Xiangtao Kong, Yu Qiao, Jiantao Zhou, and Chao Dong. Hat: Hybrid attention transformer for image restoration. *arXiv preprint arXiv:2309.05239*, 2023. 1

[2] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, and Xiaokang Yang. Recursive generalization transformer for image super-resolution. In *ICLR*, 2024. 1

[3] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong

Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. *arXiv preprint arXiv:2402.15648*, 2024. 1

[4] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 1

[5] Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. Srformer: Permuted self-attention for single image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12780–12791, 2023. 1