# Trajectory-aware Shifted State Space Models for Online Video Super-Resolution

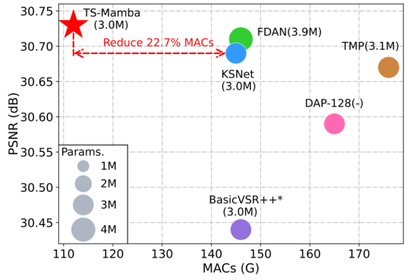
Qiang Zhu[1], Xiandong Meng[1*], Yuxuan Jiang[2], Fan Zhang[2], David Bull[2], Shuyuan Zhu[3*], Bing Zeng[3], Ronggang Wang[1,4*]
[1]Pengcheng Laboratory, [2]University of Bristol, [3]University of Electronic Science and Technology of China, [4]Peking University

ICLR

## Motivation

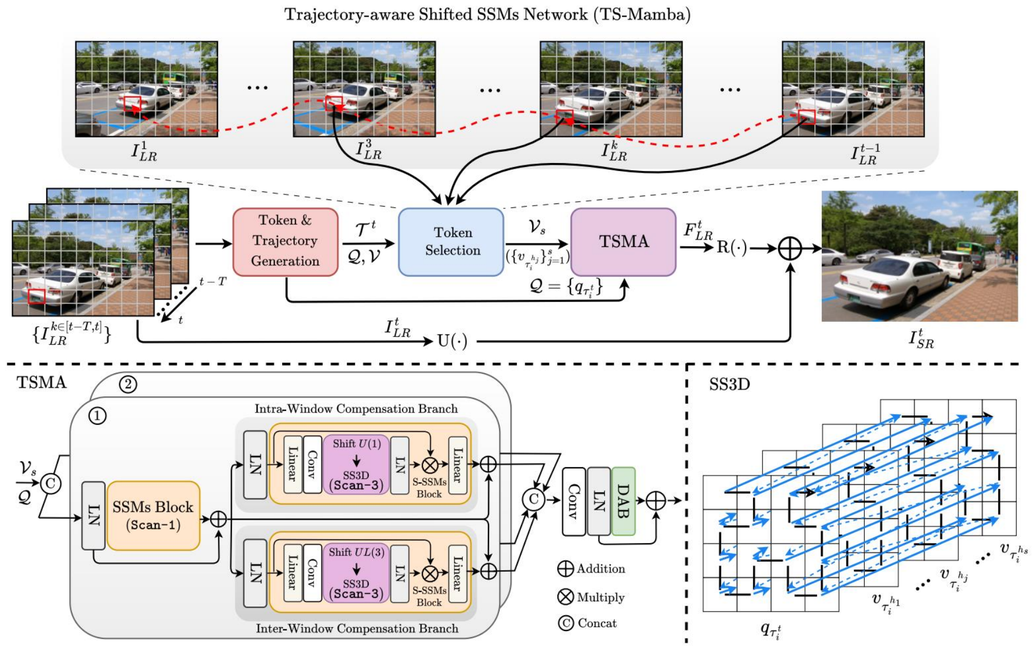
- **Online VSR**: Restoring the current high-resolution video frame based on **temporally previous frames** with **low latency** and **low computational complexity**.
- **Shortcomings of Existing Methods**:
  (1) They predominantly use **short-term temporal information** based on CNN which restricts their ability to further enhance reconstruction quality.
  (2) They often bring **heavy computational overhead** when incorporates long-term temporal alignment, resulting in challenges for real-time applications.

## Contributions




- The **first SSMs-based online VSR model** that aggregates long-term spatio-temporal information from previous frames at the token level to restore HR frames.
- The **first time to introduce video trajectories into Mamba** to select the most similar tokens to construct the trajectory-aware shifted Mamba.
- The **novel shifted SSMs blocks** are designed to strengthen the local spatial continuity of Mamba.

## Methodology




- The core designs in our **TS-Mamba**:

**a)** **Trajectory Generation & Selection** extracts tokens from frames and constructs their temporal trajectories. The most similar tokens are selected along trajectories for aggregation.

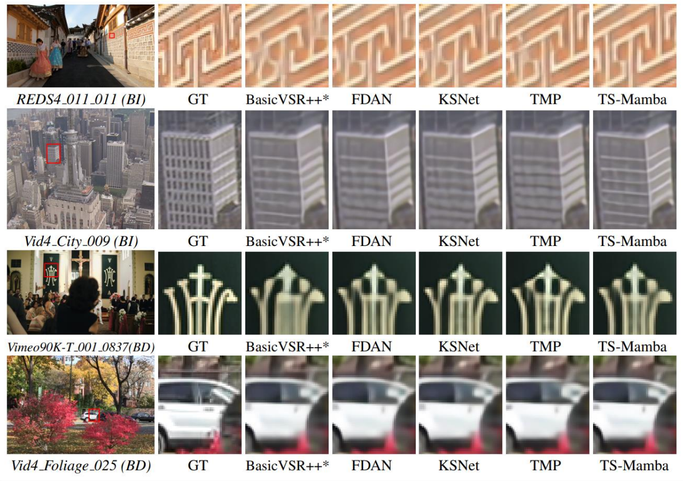
**b)** **Trajectory-aware Shifted Mamba Aggregation** (**TSMA**) designs **Shifted SSMs Blocks** with **Shifting operations** and Hilbert scannings to compensate for the **Intra-Window** and **Inter-Window losses** of Hilbert scannings, strengthening local spatial continuity of Mamba.

**c)** **Selective Scanning along Temporal Dimension** (**SS3D**) scans the current token and selected tokens to convert spatio-temporal neighboring pixels into **1D token sequence**, preserving local spatial information and progressively capturing global temporal patters.

## Quantitative Results

| Category | Methods | Supp. Frame | R.T. | Run.↓ (ms) | FPS↑ (1/s) | MACs↓ (G) | Params.↓ (M) | BI degradation REDS4(RGB)↑ (PSNR/SSIM) | BI degradation Vid4(Y)↑ (PSNR/SSIM) | BD degradation Vimeo-90K-T(Y)↑ (PSNR/SSIM) | BD degradation Vid4(Y)↑ (PSNR/SSIM) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Offline | BasicVSR++ | P+F | ✗ | 77 | 13.0 | 418 | 7.3 | 32.39/0.9069 | 27.79/0.8400 | 38.21/0.9550 | 29.04/0.8753 |
| | MIA-VSR | P+F | ✗ | 318 | 3.1 | 3220 | 16.5 | 32.78/0.9220 | 28.20/0.8507 | - | - |
| | IART | P+F | ✗ | 180 | 5.6 | 5020 | 13.4 | 32.90/0.9138 | 28.26/0.8517 | 38.62/0.9579 | 29.68/0.8884 |
| | VSRM | P+F | ✗ | 223 | 4.5 | 2174 | 17.1 | 33.11/0.9162 | 28.44/0.8552 | - | - |
| Online | RRN | P | ✓ | 34 | 29.4 | 193 | 3.4 | 28.82/0.8234 | 25.85/0.7660 | 36.69/0.9432 | 27.69/**0.8488** |
| | BasicVSR++* | P | ✓ | 40 | 25.0 | 146 | **3.0** | 30.44/0.8686 | 27.06/0.8173 | 37.11/0.9464 | 27.49/0.8426 |
| | DAP-128 | P | ✓ | 38 | 26.3 | 165 | - | 30.59/0.8703 | - | 37.29/0.9476 | - |
| | FDAN | P | ✓ | 34 | 29.4 | 146 | 3.9 | 30.71/0.8723 | 27.14/0.8206† | **37.36**/0.9483† | **27.76**/0.8471 |
| | KSNet | P | ✓ | 31 | 32.3 | 145 | **3.0** | 30.69/0.8724 | 27.14/0.8208 | 37.34/**0.9490** | 27.63/0.8444† |
| | TMP | P | ✓ | **25** | **40.1** | 176 | 3.1 | 30.67/0.8710 | 27.10/0.8167 | 37.33/0.9481 | 27.61/0.8428 |
| | VSRM* | P | ✓ | 31 | 32.7 | 136 | 3.1 | 30.64/0.8701 | 27.10/0.8163 | 37.28/0.9477 | 27.57/0.8423 |
| | **TS-Mamba** | P | ✓ | 29 | 33.5 | **112** | **3.0** | **30.73/0.8727** | **27.17/0.8209** | **37.36**/0.9482 | 27.70/0.8473 |

## Visual Results




## Ablation Studies

| Models | PSNR(dB)↑ / SSIM↑ | Params.(M)↓ | Run.(ms)↓ | MACs(G)↓ |
|---|---|---|---|---|
| (v1.1) *w/o* Trajectory | 30.45 / 0.8678 | 1.7 | 20 | 84 |
| (v1.2) *w/o* $\mathcal{L}_{trj}$ | 30.70 / 0.8721 | 3.0 | 29 | 112 |
| (v1.3) *w/o* IntraWCB | 30.58 / 0.8702 | 2.8 | 25 | 97 |
| (v1.4) *w/o* InterWCB | 30.61 / 0.8706 | 2.8 | 25 | 97 |
| (v1.5) *w/o* IntraWCB+InterWCB | 30.52 / 0.8689 | 2.4 | 21 | 85 |
| (v1.6) *w/o* $U(1)/D(1)$ | 30.65 / 0.8710 | 3.0 | 27 | 112 |
| (v1.7) *w/o* $UL(3)/DL(3)$ | 30.67 / 0.8714 | 3.0 | 27 | 112 |
| (v1.8) *w/o* (v1.6) + (v1.7) | 30.61 / 0.8702 | 3.0 | 25 | 111 |
| **TS-Mamba (ours)** | 30.73 / 0.8727 | 3.0 | 29 | 112 |