# Vision Mamba Mender

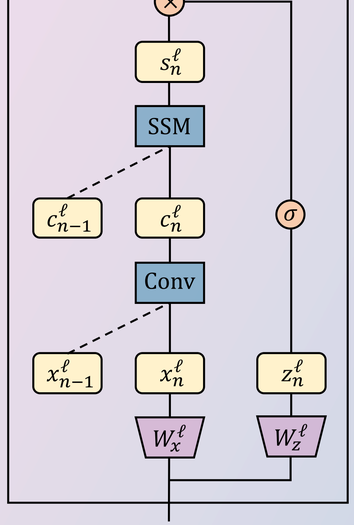
Jiacong Hu, Anda Cao, Zunlei Feng*, Shengxuming Zhang, Yi Wang, Lingxiang Jia, Mingli Song

ZHEJIANG UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

## Introduction

Vision Mamba Mender, a systematic approach for understanding the workings of Mamba, identifying flaws within, and subsequently optimizing model performance. Specifically, we present methods for predictive correlation analysis of Mamba's hidden states from both internal and external perspectives, along with corresponding definitions of correlation scores, aimed at understanding the workings of Mamba in visual recognition tasks and identifying flaws therein. Additionally, tailored repair methods are proposed for identified external and internal state flaws to eliminate them and optimize model performance.




$$x_n^{(\ell)} = \mathrm{SiLU}\left(h_n^{(\ell)-1} \cdot W_x^{(\ell)}\right)$$
$$a_1(\ell), c_2^{(\ell)}, \cdots, c_n^{(\ell)} = \mathrm{cacusal-ConvlD}\left(x_1^{(\ell)}, x_2^{(\ell)}, \cdots, x_n\right)$$
$$s_n^{(\ell)} = \mathrm{selective-SSM}\left(c_1^{(\ell)}, c_2^{(\ell)}, \cdots, c_n^{(\ell)}\right)$$
$$z_n^{(\ell)} = \mathrm{SiLU}\left(h_n^{(\ell)-1} \cdot W_z^{(\ell)}\right)$$
$$y_n^{(\ell)} = \left(s_n^{(\ell)} \odot z_n^{(\ell)}\right) \cdot W_y^{(\ell)}$$
$$h_n^{(\ell)} = h_n^{(\ell)-1} + y_n^{(\ell)}$$

## Where Do Flaws Occur?

### External State Correlation Analysis

**Grad-ESC:** Given the predictive distribution $p$ output by the Mamba model and the true class $k$ of the input sample, the calculation process for External State Correlation $\mathbf{e}^{(\ell,s)}$ of the states $\{s_n^{(\ell)}\}_{n=1}^N$ output by the selective SSM module is as follows:

$$\mathbf{e}^{(\ell,s)} = \mathcal{R}\left(\bar{s}_1^{(\ell)}, \bar{s}_2^{(\ell)}, \cdots, \bar{s}_N^{(\ell)}\right), \bar{s}_n^{(\ell)} = \mathbb{E}_D\left(g^{(\ell,s)} \odot s_n^{(\ell)}\right), g^{(\ell,s)} = \frac{1}{N}\sum_{n=1}^N \frac{\partial p^k}{\partial s_n^{(\ell)}}$$

***Definition 1 (External Correlation Score).*** *Given a pre-trained Mamba model $\mathcal{F}(\cdot)$, an input image $i$, foreground annotations $m$ of the input image, and external state correlation $\mathbf{e}^{(\ell,s)}$ computed through the proposed method, the external correlation score is defined as follows:*

$$ECS\left(\mathbf{e}^{(\ell,s)}\right) = \underbrace{\frac{\mathrm{softmax}\left(\mathcal{F}\left(\mathbf{e}^{(\ell,s)+} \odot i\right)\right)}{\mathrm{softmax}\left(\mathcal{F}\left(\mathbf{e}^{(\ell,s)-} \odot i\right)\right)}}_{\text{perturbation test}} \times \underbrace{\mathrm{IoU}\left(\mathbf{e}^{(\ell,s)+}, m\right)}_{\text{segmentation test}}$$

### Internal State Correlation Analysis

**Grad-ISC:** Given the predicted distribution $p$ outputted by the Mamba model and the true class $k$ of the input sample, let's take the example of the $n$-th state $x_n^{(\ell)}$, which is the output of matrix $W_x^{(\ell)}$. The computation process for the corresponding Internal State Correlation $\mathbf{i}_n^{(\ell,x)}$ is as follows:

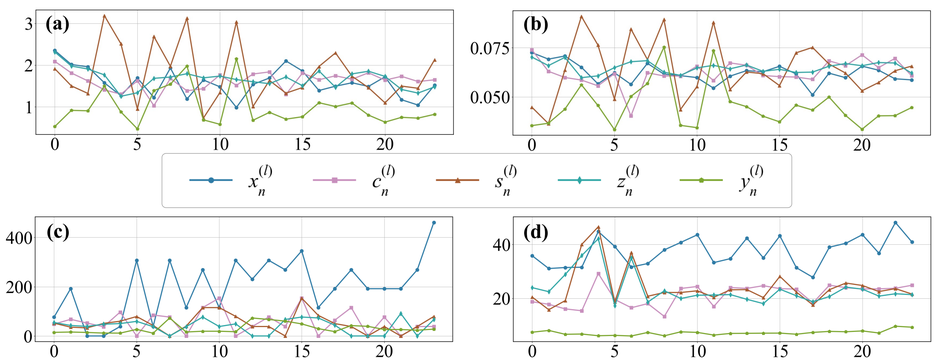
$$\mathbf{i}_n^{(\ell,x)} = g_n^{(\ell,x)} \odot s_n^{(\ell)}, g_n^{(\ell,x)} = \frac{\partial p^k}{\partial x_n^{(\ell)}}$$

***Definition 2 (Internal Correlation Score).*** *Given $J$ samples belonging to the same class and the internal state correlation $\mathbf{i}_n^{(\ell,x)}$ computed using the proposed method for a particular sample, the Internal Correlation Score is defined as follows:*

$$ICS\left(\mathbf{i}_n^{(\ell,x)}\right) = \underbrace{\mathbb{E}_D\left(\frac{1}{j}\sum_{j=1}^J \mathbf{i}_{n,j}^{(\ell,x)+}\right)}_{\text{simplicity}} \times \underbrace{\mathbb{E}_D\left(\frac{\frac{1}{j}\sum_{j=1}^J \mathbf{i}_{n,j}^{(\ell,x)+}}{\mathbf{i}_{n,1}^{(\ell,x)+} \oplus \mathbf{i}_{n,2}^{(\ell,x)+} \oplus \cdots \oplus \mathbf{i}_{n,J}^{(\ell,x)+}}\right)}_{\text{homogeneity}}$$

### Results of Correlation Analysis

Comparison of state correlation scores across different blocks between simple and difficult samples. (a) and (b) show the external state correlation scores for simple and difficult samples, respectively. (c) and (d) present the internal state correlation scores for simple and difficult samples, respectively.




## How to Repair Flaws?

### External State Correlation Repair

We focus on repairing the external correlation flaws of the states $c_n^{(\ell)}$ and $s_n^{(\ell)}$ output by the Conv and SSM modules in the deeper blocks. Specifically, we identify difficult samples from the training set and then constrain the external correlations of the hidden states using object annotations $m$:

$$Loss_e = \mathbb{E}_{HW}\left(\mathbf{e}^{(\ell,c)+} \odot m\right) + \mathbb{E}_{HW}\left(\mathbf{e}^{(\ell,s)+} \odot m\right)$$

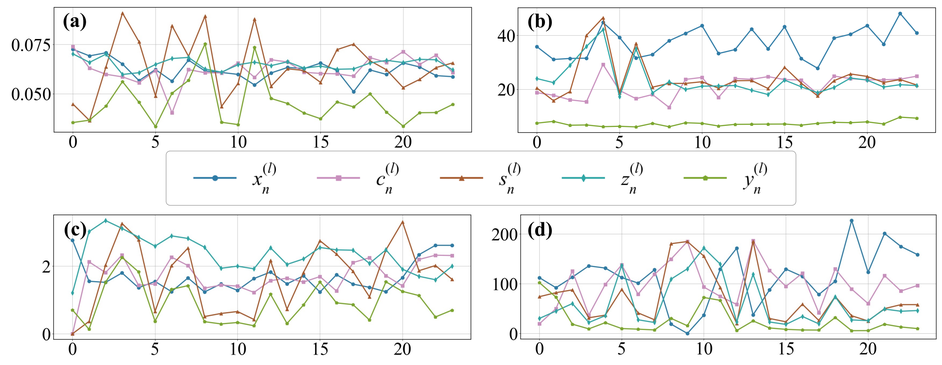
### Internal State Correlation Repair

We focus on repairing the internal correlation flaws of the states $x_n^{(\ell)}$ output by the linear mapping $W_x^{(\ell)}$ within deeper blocks. Specifically, we leverage the templates $\mathbf{i}_n^{(\ell,x)+}$ to constrain the internal correlations of $x_n^{(\ell)}$:

$$Loss_c = \mathbb{E}_D\left(\mathbf{i}_n^{(\ell,x)+} \odot \mathbf{i}_n^{(\ell,x)+}\right), \mathbf{i}_n^{(\ell,x)+} = \frac{1}{j}\sum_{j=1}^J \mathbf{i}_{n,j}^{(\ell,x)+}$$

### Results of Flaw Repair

Repairing SOTA Vision Mamba Models. (a) and (b) show the state correlation scores before flaw repair. (c) and (d) present the state correlation scores after flaw repair.




| | ViM-T | VMamba-T | SiMBA-S | EMamba-T | LocalVim-T |
|---|---|---|---|---|---|
| **ImageNet-50** | | | | | |
| **Base** | 76.44 | 79.96 | 81.32 | 81.44 | 75.08 |
| **+Ext** | 78.48+2.04 | 81.08+1.12 | 84.48+3.16 | 82.68+1.24 | 78.68+3.60 |
| **+Int** | 78.20+1.76 | 82.36+2.40 | 84.64+3.32 | 83.24+1.80 | 78.20+3.12 |
| **+All** | 79.68+3.24 | 82.28+2.32 | 86.52+5.20 | 83.32+1.88 | 80.56+5.48 |
| **ImageNet-300** | | | | | |
| **Base** | 75.11 | 75.04 | 68.08 | 74.67 | 70.39 |
| **+Ext** | 77.87+2.76 | 76.01+0.97 | 69.73+1.65 | 75.03+0.36 | 73.09+2.70 |
| **+Int** | 77.76+2.65 | 76.31+1.27 | 69.93+1.85 | 75.55+0.88 | 72.71+2.32 |
| **+All** | 79.53+4.42 | 76.75+1.71 | 70.58+2.50 | 75.66+0.99 | 74.84+4.45 |
| **ImageNet-1K** | | | | | |
| **Base** | 71.64 | 67.54 | 51.06 | 67.35 | 57.70 |
| **+Ext** | 73.02+1.38 | 68.34+0.80 | 52.12+1.06 | 67.62+0.27 | 59.30+1.60 |
| **+Int** | 72.79+1.15 | 68.67+1.13 | 52.23+1.17 | 67.89+0.54 | 59.90+2.20 |
| **+All** | 73.30+1.66 | 68.68+1.14 | 52.24+1.18 | 67.84+0.49 | 60.83+3.13 |