Foveal-Mamba: Inside-Out Ring Scanning with Recurrent Offset Prediction for Visual State Space Models

Yi-Kuan Hsieh[1], Kuan-Chuan Peng[2], Xin Li[3], Ming-Ching Chang[3], Chi-Chia Sun[4], Yu-Chee Tseng[1], Jun-Wei Hsieh[1]

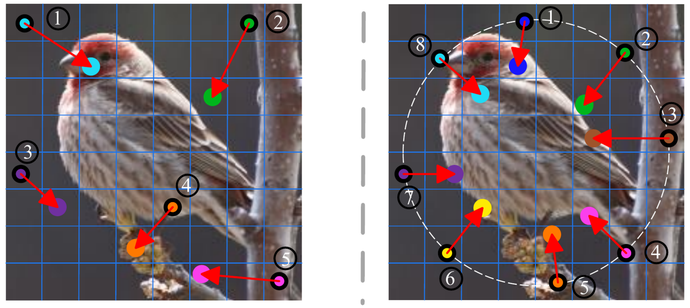
國立陽明交通大學 NATIONAL YANG MING CHIAO TUNG UNIVERSITY · UNIVERSITY AT ALBANY STATE UNIVERSITY OF NEW YORK · MITSUBISHI ELECTRIC RESEARCH LABORATORIES

NEURAL INFORMATION PROCESSING SYSTEMS 2026

Why Foveal-Mamba?

- Existing deformable visual Mamba methods predict offsets mainly from local spatial responses.
- Local offset prediction is weakly coupled with recurrent state evolution.
- Off-center objects and competing instances require coherent, object-aware sampling trajectories.

(a) DAMamba: independent local offset prediction | (b) Foveal-Mamba: Inside-Out Ring Scan + RS-OPN



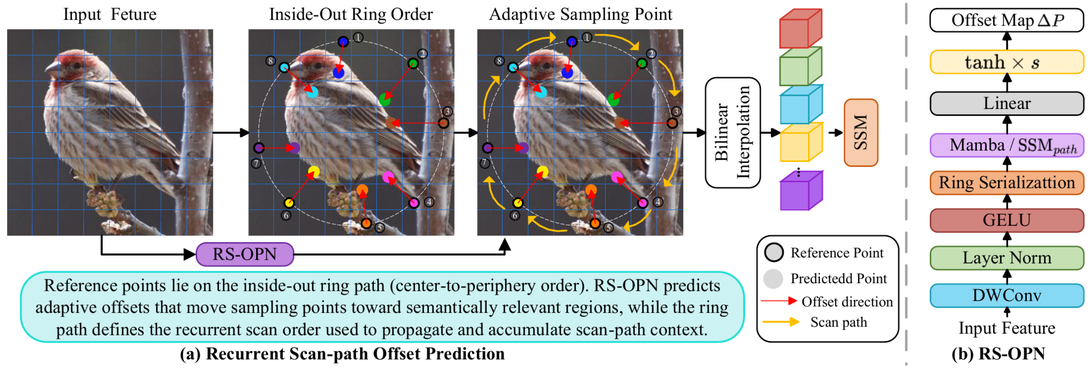
Original token position (before offset) · Offset direction (predicted displacement) · Final displaced position (after offset)

Key Contributions

1. **Recurrent scan-path offset prediction**
   Predict offsets from accumulated recurrent context.
2. **Content-adaptive inside-out prior**
   Adapt sampling to off-center and multiple objects.
3. **Coupled sampling and SSM aggregation**
   Share the ring path between RS-OPN and main SSM.

Method Overview

One shared ring path couples recurrent offset prediction with deformable sampling and main SSM feature mixing. Learned offsets adapt to image content.

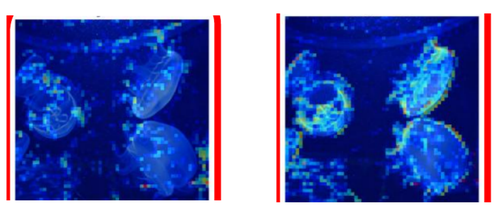


Reference points lie on the inside-out ring path (center-to-periphery order). RS-OPN predicts adaptive offsets that move sampling points toward semantically relevant regions, while the ring path defines the recurrent scan order used to propagate and accumulate scan-path context.

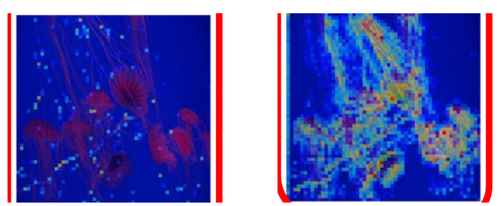
(a) Recurrent Scan-path Offset Prediction

Earlier Object Focus at Stage 1

Earlier object contours



More consistent multi-instance responses



Why recurrent context matters
Accumulated scan-path states guide offsets to coordinate sampling across object regions.

Center-first, content-adaptive sampling
The center sets the order; learned offsets adapt sampling to off-center and multiple objects.

Main Results

(A) ImageNet-1K classification

Model	Top-1	Params	FLOPs
VMamba-T	82.6	30M	5.6G
DefMamba-T	83.5	26M	4.8G
DAMamba-T*	83.8	26M	4.8G
Foveal-Mamba-T	**84.3**	**25M**	**4.6G**
DAMamba-S*	84.8	45M	10.3G
Foveal-Mamba-S	**85.2**	**44M**	**10.2G**

* Reported DAMamba values were not reproduced under our training setting. A100 throughput (T/S): ours 872/482 vs. DAMamba 1018/595 images/s. Input size: 224 × 224.

(B) ADE20K segmentation

Model	SS mIoU	MS mIoU
DAMamba-T	50.3	51.2
Foveal-Mamba-T	**50.8**	**51.6**
DAMamba-S	51.2	52.0
Foveal-Mamba-S	**51.8**	**52.5**

UPerNet, single-scale (SS) / multi-scale (MS) testing.

(C) COCO detection & segmentation

Model	Box AP	Mask AP
DAMamba-T	48.5	43.4
Foveal-Mamba-T	**48.9**	**43.7**
DAMamba-S	49.8	44.5
Foveal-Mamba-S	**50.2**	**44.9**

Mask R-CNN, 1× schedule. Results: paper Tables 1–3.

Path Regularization (Table 8)

$$L = L_{task} + \lambda_p L_{path} + \lambda_c L_{curv} + \lambda_m L_{mag}$$

Regularization on mini-ImageNet	Top-1 (%)
None	88.21
Smoothness	88.32
Smoothness + curvature	88.39
Smoothness + curvature + magnitude	**88.55**

Ablation: Why It Works

Component-wise ablation on mini-ImageNet. Top-1 accuracy (%).

Setting	Ring OPN	RS-OPN	Ring SSM	Path reg.	Top-1
DAMamba-style OPN	No	No	No	No	87.48
+ Ring order only	Yes	No	No	No	87.69
+ RS-OPN only	No	Yes	No	No	87.83
+ Ring + RS-OPN	Yes	Yes	No	No	88.06
+ Ring-ordered main SSM	Yes	Yes	Yes	No	88.21
Full Foveal-Mamba	**Yes**	**Yes**	**Yes**	**Yes**	**88.55**

Full model: +1.07 percentage points. Path regularization adds no inference-time modules.

Scan Path Ablation

Same Mamba-in-OPN, different scan order

Scan path	Top-1	SS	Box AP
Raster	88.09	50.3	48.5
Reverse raster	88.04	50.2	48.4
Serpentine	88.15	50.4	48.5
Outside-in ring	88.23	50.4	48.5
Inside-out ring	**88.55**	**50.8**	**48.9**

Top-1: mini-ImageNet. SS: ADE20K. Paper Table 7.

Contact: yikuan0725@gmail.com Corresponding author: jwhsieh@nycu.edu.tw

T: Tiny S: Small