Looking for PhD Position

WeChat rednote

Here I am!

wdtang0705@gmail.com

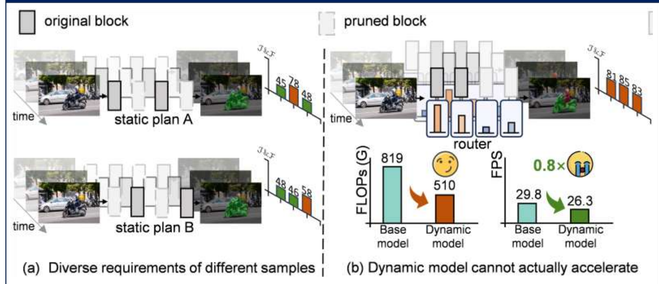
Efficient Video Object Segmentation and Tracking with Recurrent Dynamic Submodel

Weidong Tang, Zhiyuan Liang, Xinyan Wan, Chen Zhu, Zhaopan Xu, Pengfei Zhou, Yan Song, Yang You, Wangbo Zhao†
XDU, NUS, USTC, HIT

CVPR JUNE 3-7, 2026 DENVER COLORADO

WICKED IDEA GENERATING...

Motivation



- We observe that **different video frames favor distinct structures**, necessitating a dynamic approach.

Sample-adaptive Model

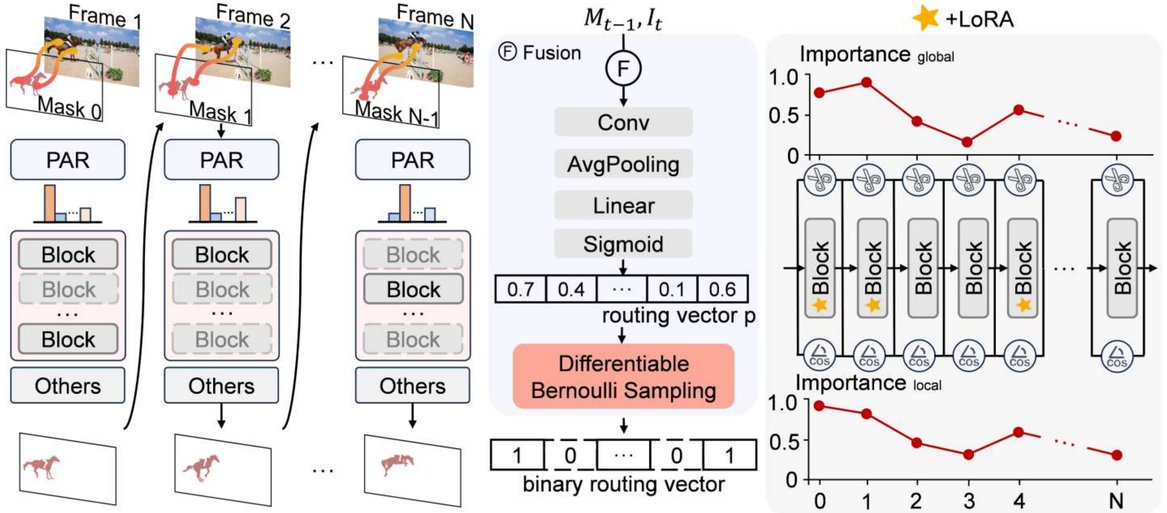
- However, current dynamic models **fail to deliver real acceleration on ViTs** due to severe per-block routing overhead.

Coarse-grained ViT-agnostic Routing

- Furthermore, they simply apply image-level logic, **completely ignoring the crucial temporal correlations in videos.**

Temporal Information

Overall Pipeline of the **Recurrent Dynamic Submodel**



Only execute necessary blocks; only fine-tune critical blocks.

Prediction-Aware Router (PAR):
A single, lightweight router that fuses the previous frame's mask (temporal prior) with current visual features. It determines the optimal submodel for the entire network in one step, eliminating the latency of dense per-block routing.

Importance-Aware LoRA (I-LoRA):
Evaluates the significance of each block via both global output impact and local feature transformation. It allocates trainable LoRA adapters only to the most critical blocks, drastically reducing fine-tuning costs.

Sophisticated Analysis.

Gang, hop on.

Experiments

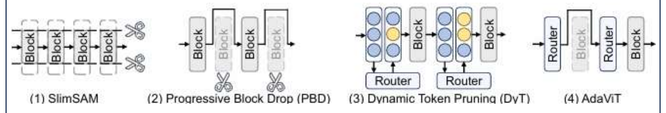
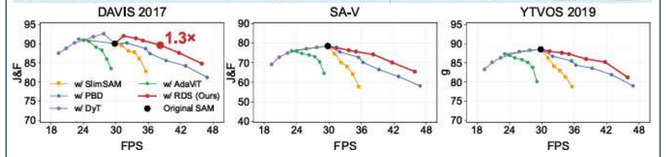


Table 1. **Performance comparison with other methods.** This table compares the performance of different methods in terms of trainable parameters, training set size, GPU usage, number of iterations, FLOPs, FPS, and $\mathcal{J}\&\mathcal{F}$ score [28] on DAVIS 2017.

Method	Trainable param. (M)	Train set size (%)	GPUs	Iter. (K)	FLOPs (G)↓	FPS ↑	$\mathcal{J}\&\mathcal{F}$ ↑
SAM2	224	100	256	200	819	29.8	90.0
SlimSAM	147	>5	8	38	547	33.4	87.8
PBD	7.6	<0.03	1	5	**500**	**39.8**	85.6
DyT	20.5	<0.03	1	5	629	22.3	**90.2**
AdaViT	9.6	<0.03	1	5	528	26.3	88.3
RDS (Ours)	**7.6**	**<0.03**	1	5	**500**	38.2	89.6



Achieves **Pareto Frontier** across all benchmarks.