RGB-D Video Object Segmentation via Enhanced Multi-store Feature Memory

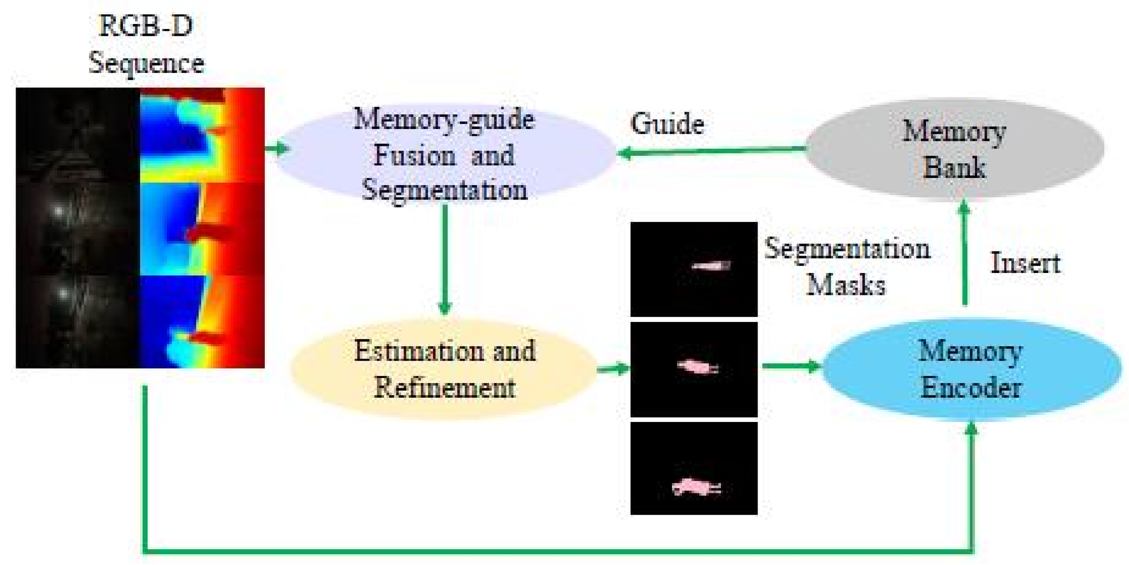
ICMR 2024

Boyue Xu, Yi Xu, Ruichao Hou✉, Tongwei Ren, Gangshan Wu

State Key Laboratory for Novel Software Technology, Nanjing University

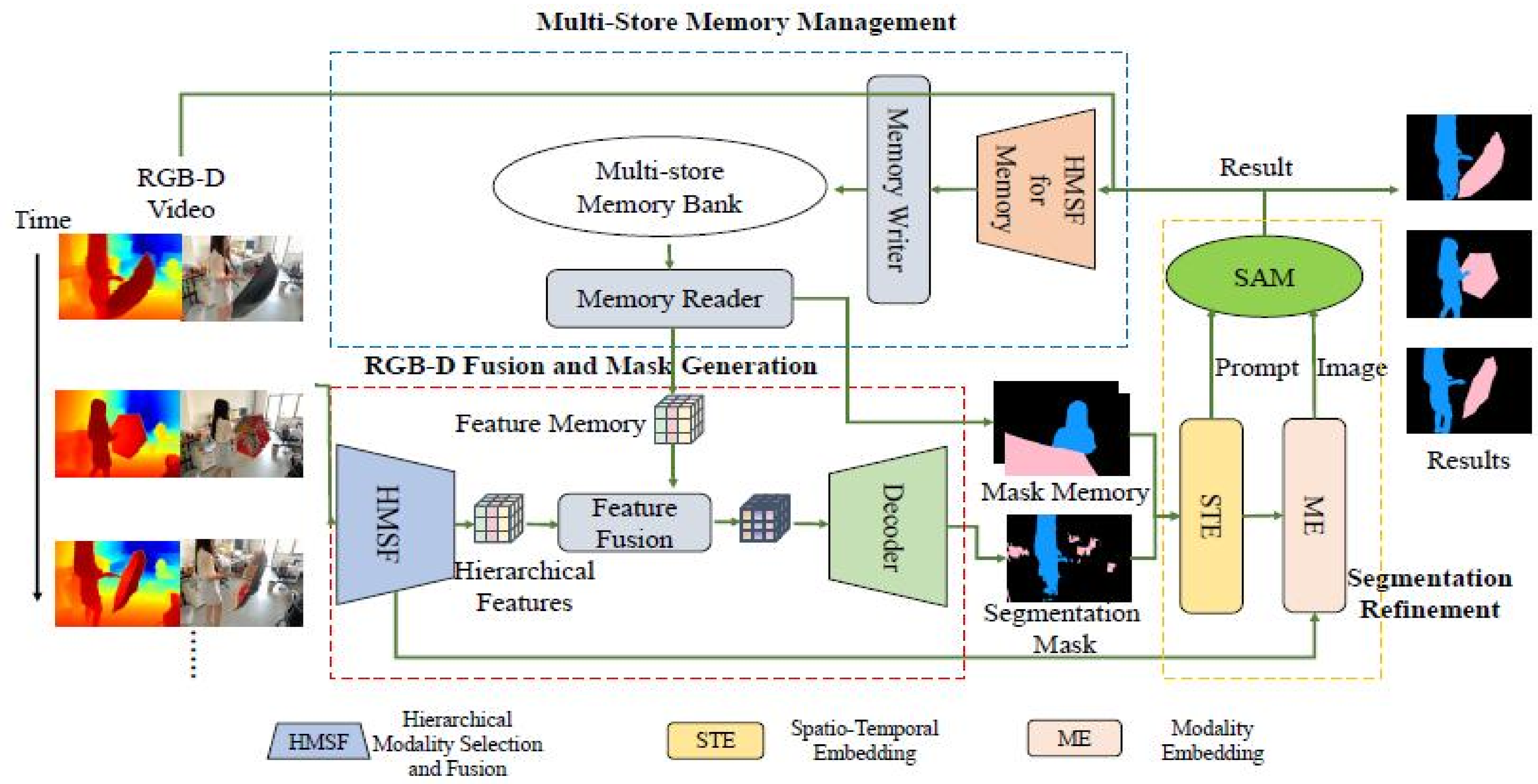
Introduction

RGB-Depth(RGB-D) Video Object Segmentation (VOS) combines RGB and depth data. RGB offers visual details like color and texture, while depth providing the distance from the camera to the object.
We propose a RGB-D VOS method via multi-store feature memory for robust segmentation. Specifically, we design the hierarchical modality selection and fusion and a segmentation refinement module, significantly improving the robustness of segmentation.



Method

The proposed method comprises three modules: RGB-D fusion and mask generation module, segmentation refinement module, and multi-store memory management module. Specifically, the RGB-D fusion and mask generation module aims to fuse RGB-D dual-modality features and integrate them with multi-store feature memory to produce segmentation results. The segmentation refinement module flexibly utilizes the SAM to refine segmentation results and ensure more accurate results as memory to guide subsequent segmentation. The multi-store memory management module encodes and stores both RGB-D images and segmentation results as feature memory.

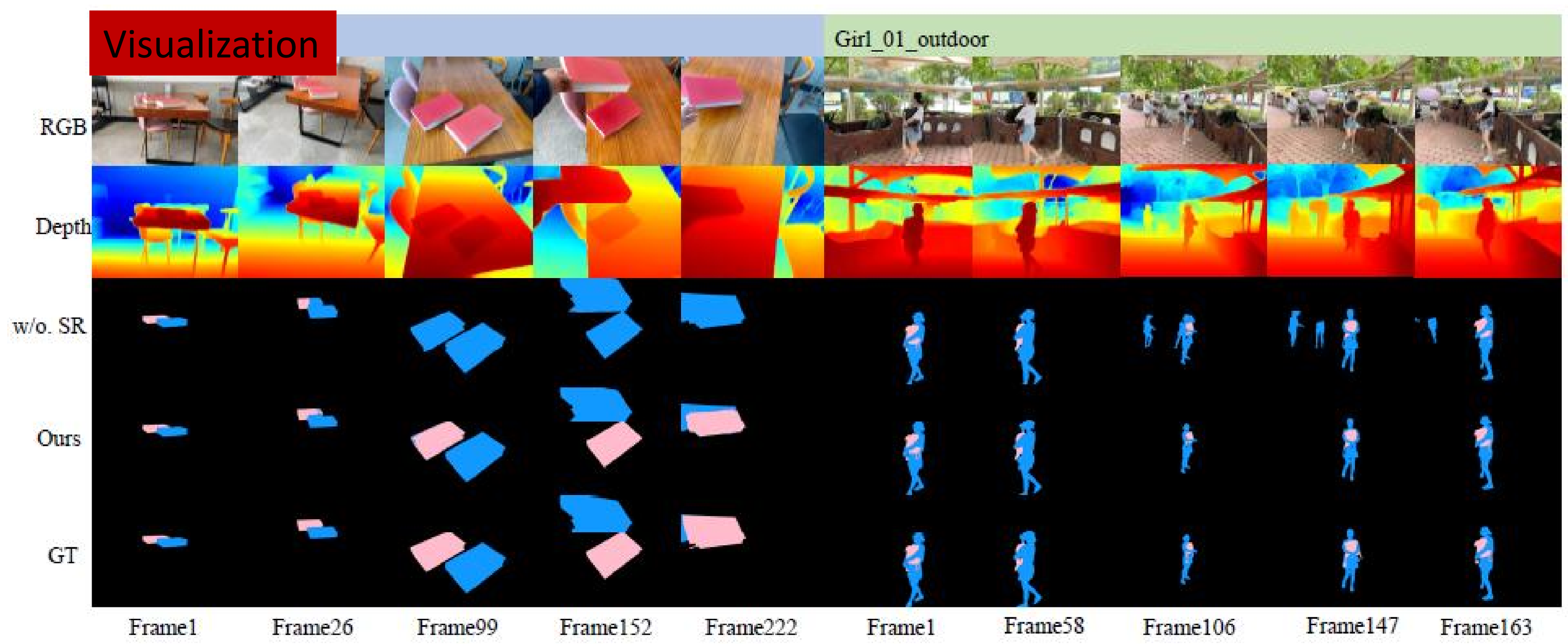


Experiments

Dataset: ARKitTrack

Metrics: Region similarity (J), Contour accuracy(F), J&F

Ablation	STE	ME	$\mathcal{J}_M\uparrow$	$\mathcal{F}_M\uparrow$	$\mathcal{J}\&\mathcal{F}\uparrow$
			0.617	0.680	0.649
✓			0.637	0.691	0.664
✓	✓		0.651	0.702	0.677
✓	✓	✓	**0.673**	**0.723**	**0.698**



Comparison with the SOTA: The method is superior to all the RGB-D VOS methods

Ablation Study: The experiment demonstrates the effectiveness of each component of the method

Comparison Methods	Year	$\mathcal{J}_M\uparrow$	$\mathcal{F}_M\uparrow$	$\mathcal{J}\&\mathcal{F}\uparrow$
STCN [5]	2021	0.489	0.560	0.525
AOT [41]	2021	0.555	0.627	0.582
RPCM [38]	2022	0.492	0.527	0.509
QDMN [21]	2022	0.276	0.337	0.306
XMem [4]	2022	0.541	0.565	0.553
STCN_RGBD [5]	2021	0.498	0.574	0.537
XMem_RGBD [4]	2022	0.617	0.680	0.649
SAMTrack_RGBD [7]	2023	0.445	0.463	0.454
ARKitTrack [46]	2023	0.625	0.698	0.662
Ours	-	**0.673**	**0.723**	**0.698**

南京大學 NANJING UNIVERSITY

xuby@smail.nju.edu.cn
rc_hou@smail.nju.edu.cn

MAGUS MediA recoGnition and UnderStanding