



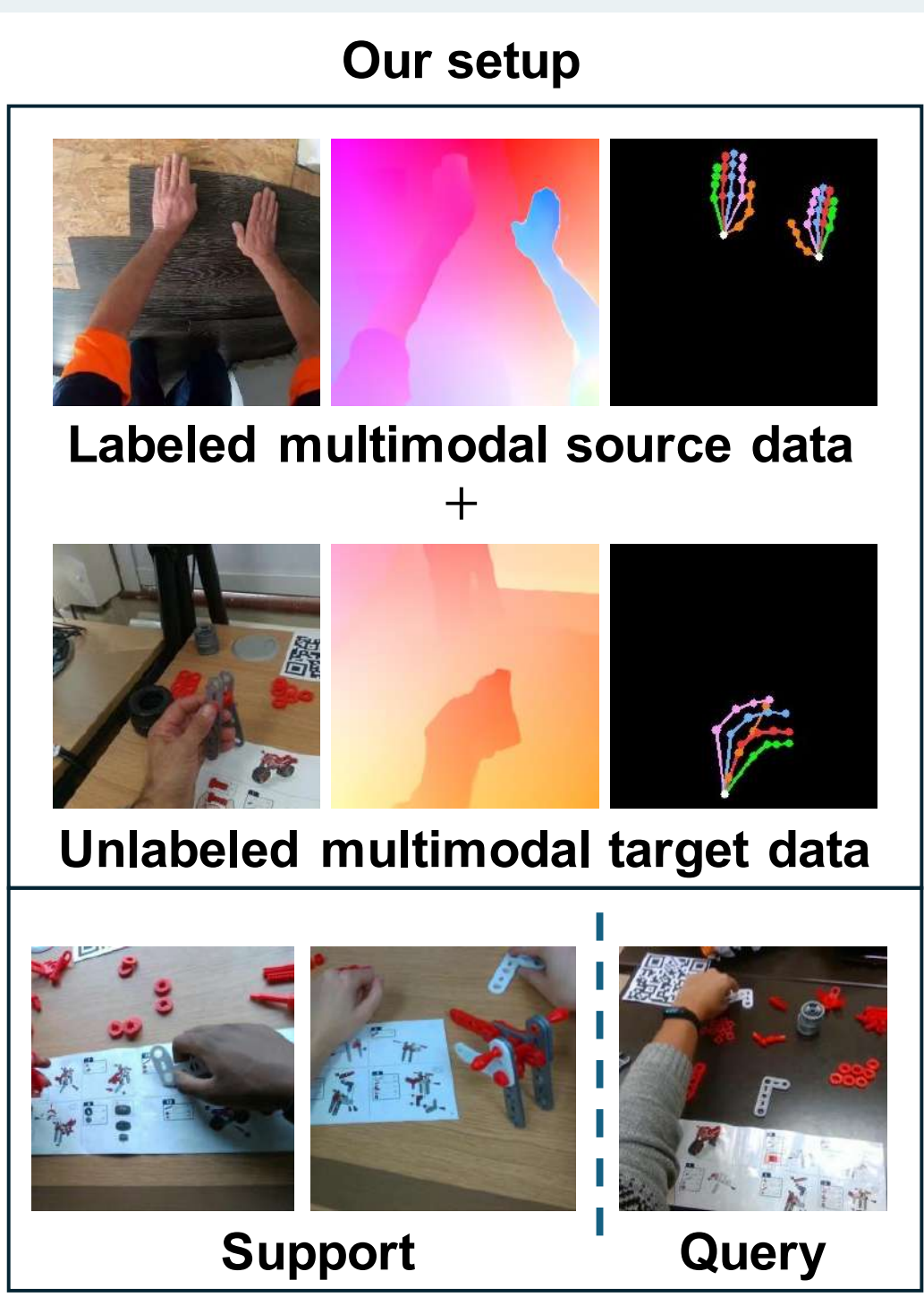
Overview

- Address a novel challenging, but practical problem: CD-FSL with unlabeled target and multimodal input
- Propose MM-CDFSL, a novel approach for CDFSL for egocentric action recognition
- Achieve SOTA in both accuracy & inference cost

Problem Setup

Previous Related Problem Setup

- ❑ Few-Shot
 - MAML [ICML'17], ProtoNet [NeurIPS'17]
- ❑ Cross-Domain Few-Shot
 - BS-CDFSL [ECCV'20]
- ❑ Cross-Domain Few-Shot w/ unlabeled target
 - STARTUP [ICLR'21], Dynamic Distill [NeurIPS'21], CDFSL-V [ICCV'23]

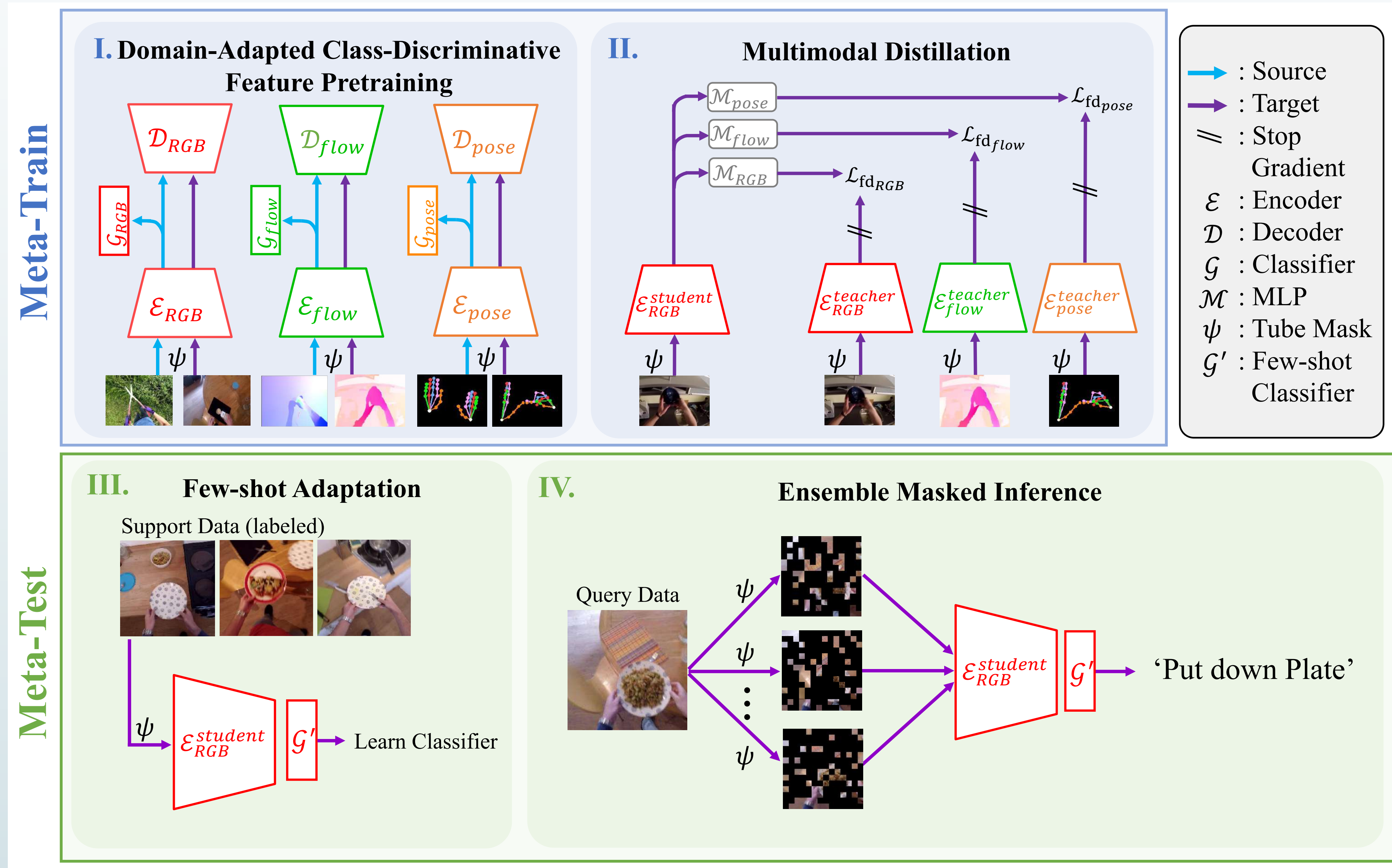


Meta-Training
(all m modalities)
Source Dataset: D_S
Unlabeled Target Dataset: D_{T_u}

Meta-Test
(only RGB)
Target Dataset: D_T
Support Set: S (N -way K -shot)
Query Set: Q (N classes)



Proposed Method



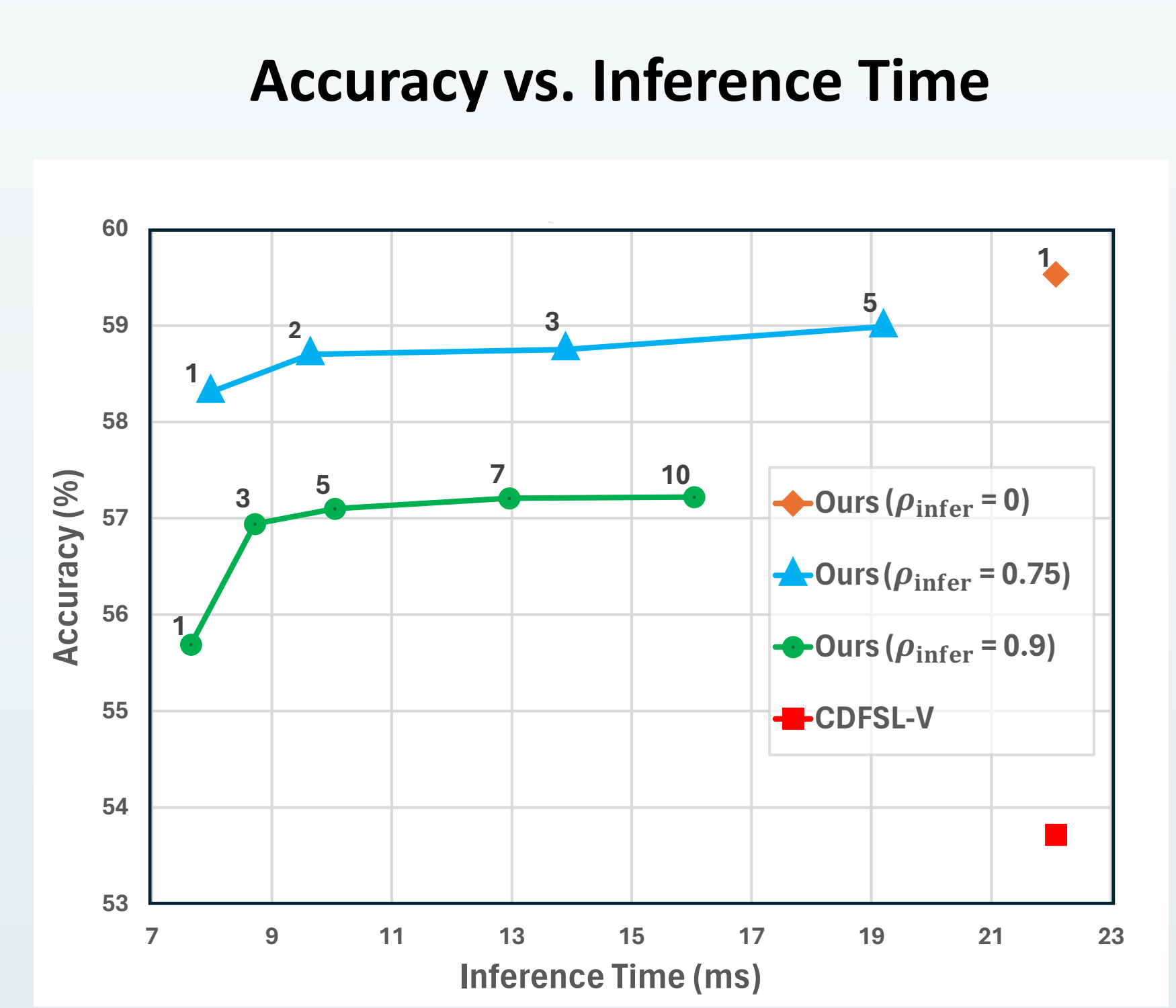
Issues

- 1. Domain Adaptability**
 - Solely rely on RGB
 - Using multimodal is unexplored
- 2. Inference Cost**
 - Process desely-sampled frames
 - Computational cost for resource limited devices

Experimental Results

Few-shot Accuracy & Inference Cost on EPIC, MECCANO, WEAR

Method	Runtime (ms)	GFLOPs	Memory (MiB)	1-shot			5-shot		
				EPIC	MEC	WEAR	EPIC	MEC	WEAR
Random Initialization				29.20±.37	23.10±.24	25.96±.27	40.28±.42	27.04±.28	38.71±.36
VideoMAE [NeurIPS'22]				35.07±.41	27.75±.31	44.65±.38	47.13±.43	35.92±.33	63.92±.35
STARTUP++ [ICLR'21]	22.1	68.5	2782	35.18±.43	26.84±.30	39.15±.35	50.24±.45	34.05±.31	59.88±.36
Dynamic Distill++ [NeurIPS'21]				36.96±.43	27.87±.30	35.84±.32	53.78±.47	37.87±.33	56.23±.35
CDFSL-V [ICCV'23]				38.17±.44	26.03±.29	39.11±.35	53.72±.91	35.64±.32	58.27±.36
Ours	9.64	37.0	968	41.97±.46	28.34±.30	51.25±.40	58.70±.90	37.80±.46	69.57±.37



Domain Adaptability & Class-Discriminiveness

$\mathcal{L}_{recon}^{source}$	$\mathcal{L}_{recon}^{target}$	$\mathcal{L}_{ce}^{source}$	1-shot	5-shot
✓	✓		35.42	49.82
✓		✓	40.50	56.43
✓	✓	✓	41.97	58.70

Multimodal Distillation

Method	1-shot	5-shot
Only RGB Training	46.17	67.19
RGB+Pose	49.39	67.90
Ours	51.25	69.57

Limitations & Future Work

- ❖ Multimodal data for both source and target
 - Missing modality cases during training
- ❖ Equally distilling multiple modalities
 - Dynamical adjustment of distillation weights according to the modality's relevance in the target domain