A Dataset for Semantic and Instance Segmentation of Modern Fruit Orchards

Tieqiao Wang, Abhinav Jain, Liqiang He, Cindy Grimm, Sinisa Todorovic
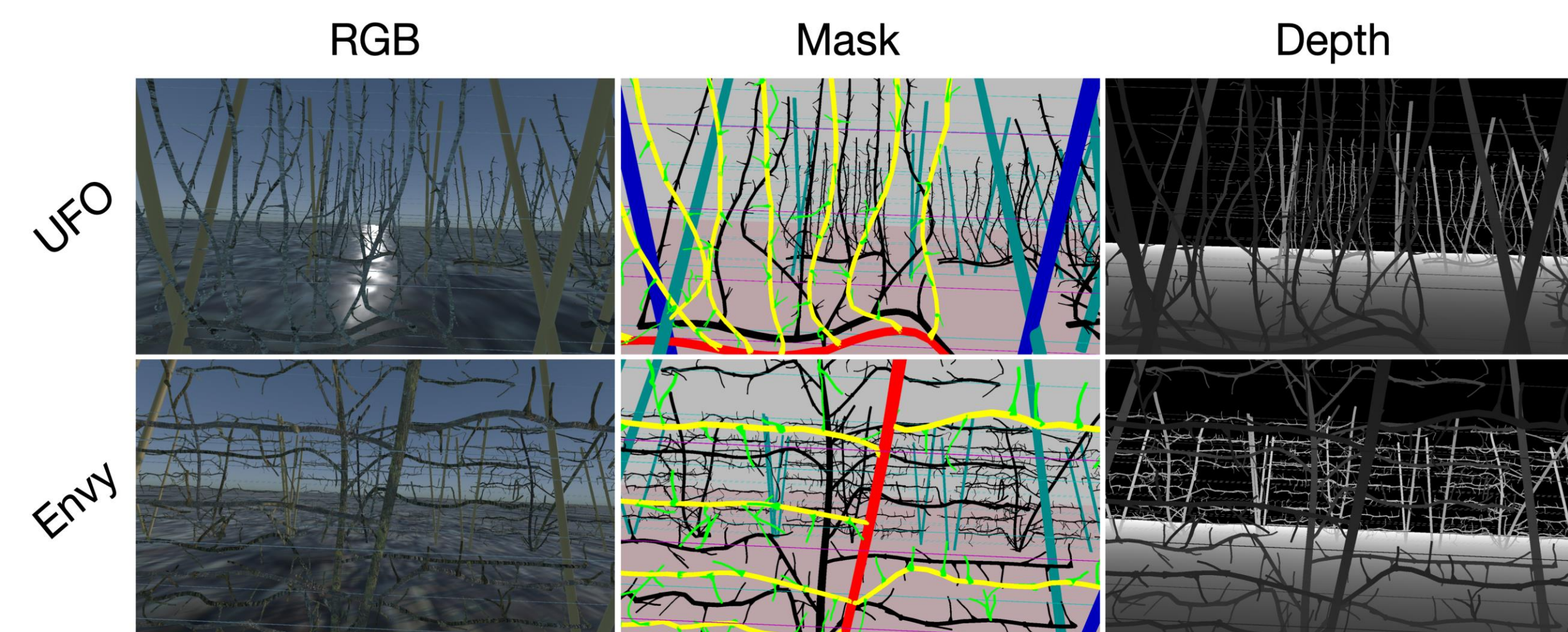
Data & Code

Goal:

➤ Benchmark a large dataset for semantic and instance segmentation of elongated thin objects in cluttered scenes, e.g., branches of fruit trees in modern orchards.

Dataset:

➤ **Synthetic data**: 5,000 images per tree type – Envy apple and UFO cherry tree type – with ground-truth depths and segmentation masks of 9 semantic classes.

RGB Mask Depth

➤ **Real-world data**: 132 RGB videos of Cherry trees (68,760 frames) and 91 videos of Apple trees (91,470 frames)


➤ **Additional data**: 500 synthetic meshes of trees for 3D reconstruction and simulation.

Motivation and Key Challenges:

➤ Automating orchard tasks, like pruning fruit trees, demands detailed tree-structure understanding to achieve robotic manipulation.
➤ There are no large-scale, labeled datasets of orchard scenes.
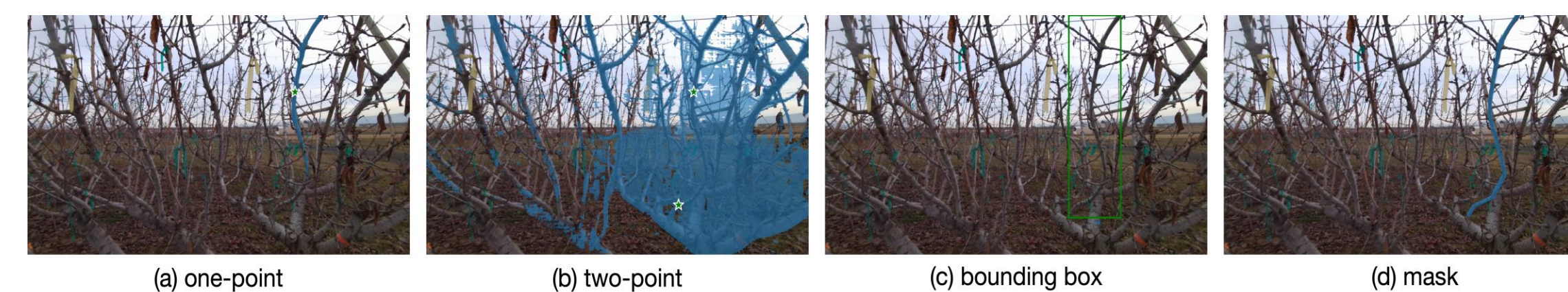➤ Manual labeling of tree branches is challenging and prone to errors.

Two Tasks:

➤ **Unsupervised Domain Adaptation for Segmentation.** Our approach – Semantically-Guided Depth Refinement – improves synthetic-to-real data adaptation for semantic segmentation.
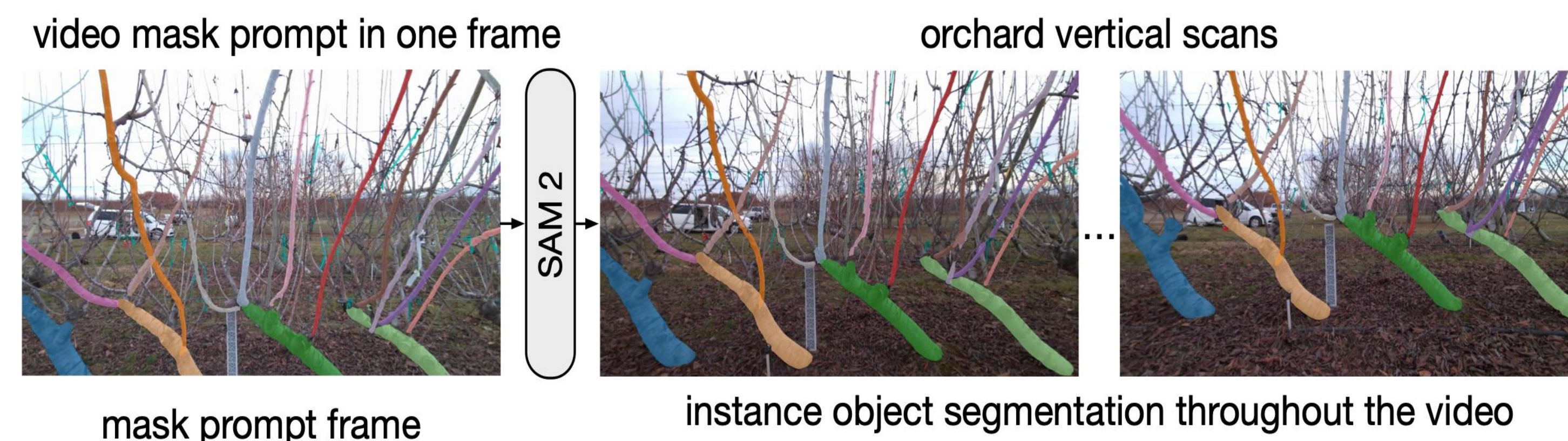

a) Multi-modal Fusion Architecture   b) Semantically-Guided Depth Refinement   c) UDA Training Pipeline
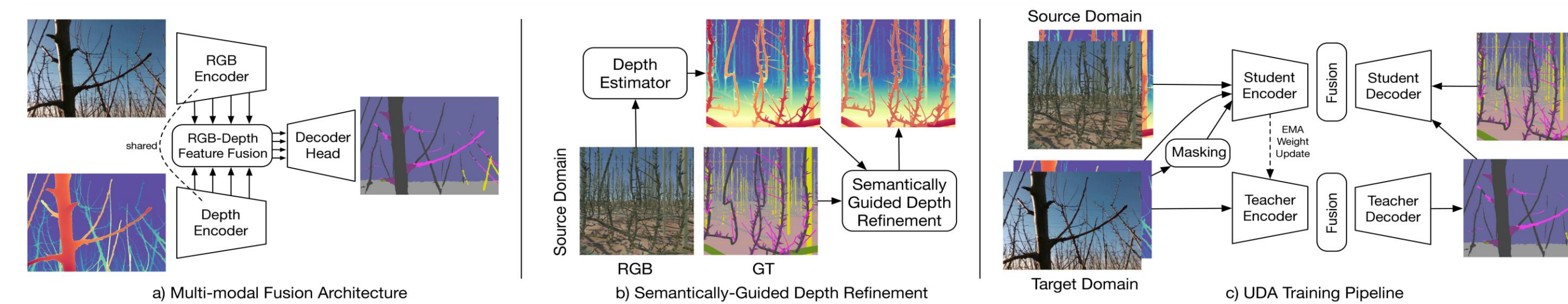
➤ **Instance Segmentation under Supervised Setting.** Our approach – SAM-M2F – enables annotation propagation throughout the video.
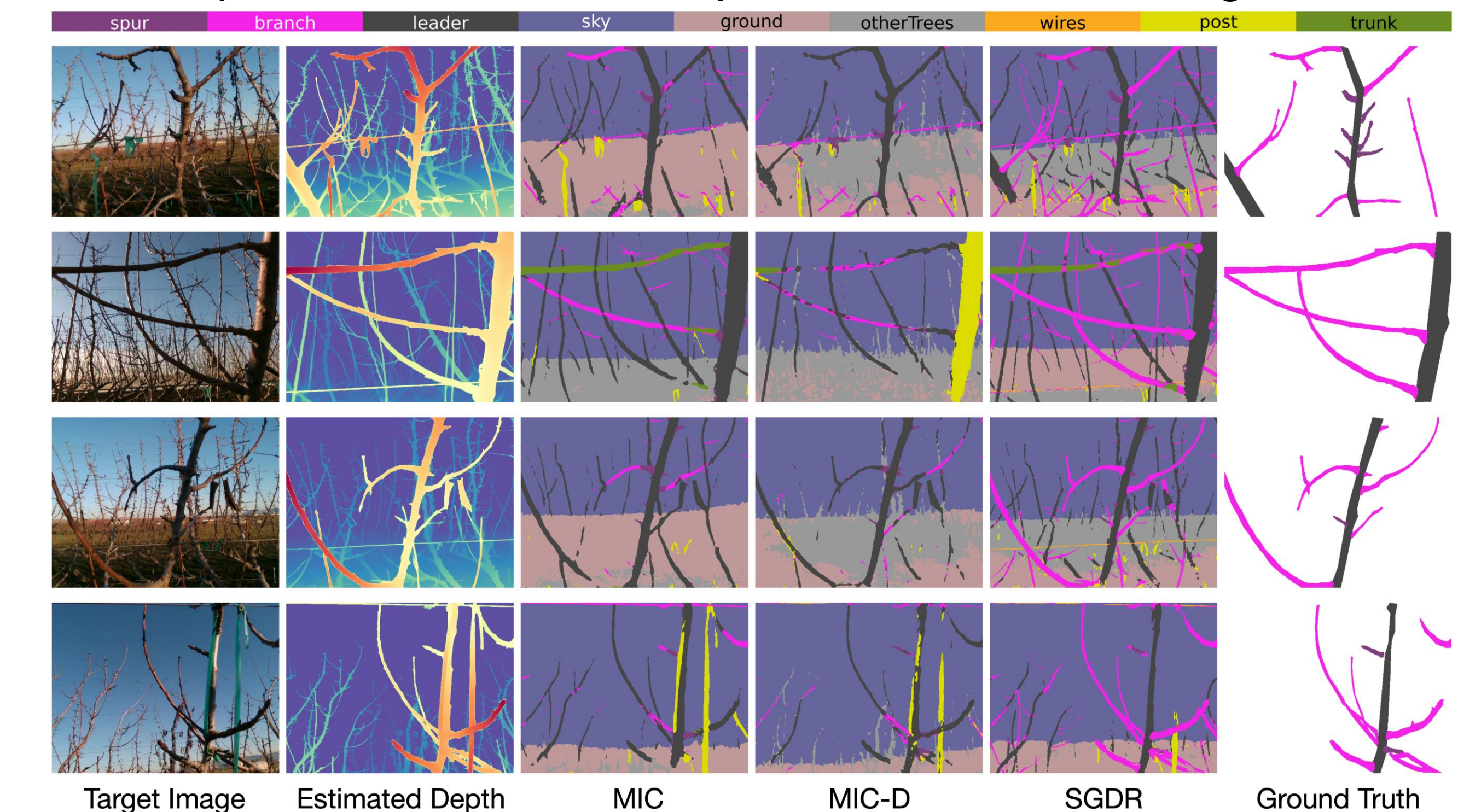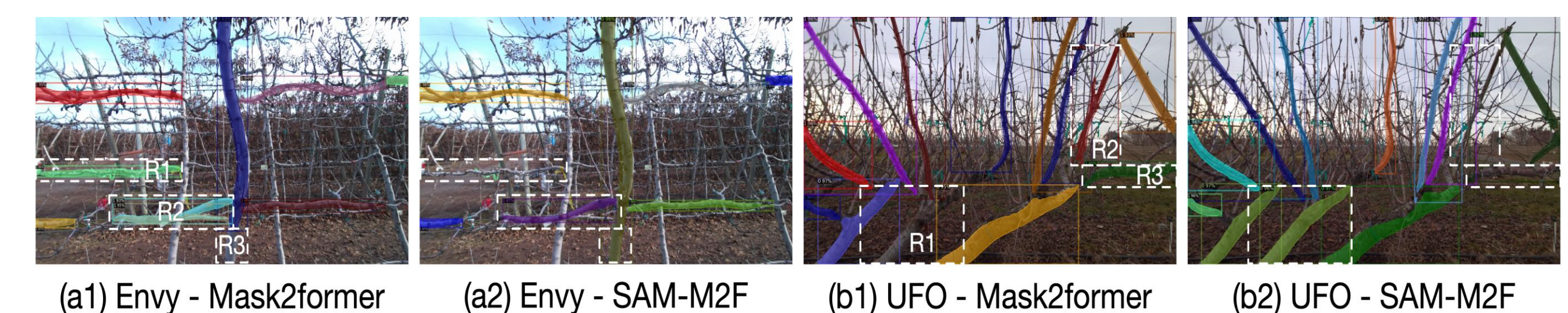
Baseline: SAM2 with visual prompt

(a) one-point   (b) two-point   (c) bounding box   (d) mask

Our approach: SAM-M2F

video mask prompt in one frame   orchard vertical scans

mask prompt frame   instance object segmentation throughout the video

Results:

➤ Unsupervised Domain Adaptation - Semantic Segmentation

Target Image   Estimated Depth   MIC   MIC-D   SGDR   Ground Truth

Method	IoU			Acc			Overall		
	spur	branch	leader	spur	branch	leader	aAcc	mIoU	mAcc
MIC [32]	28.7	37.7	**76.5**	45.3	39.4	**87.1**	69.7	47.6	57.3
MIC-D	27.0	29.2	64.3	39.4	30.0	76.4	59.7	40.2	48.6
SGDR	**33.6**	**58.6**	76.4	**55.6**	**69.5**	79.5	**74.9**	**56.2**	**68.2**

➤ Supervised Learning - Instance Segmentation

(a1) Envy - Mask2former   (a2) Envy - SAM-M2F   (b1) UFO - Mask2former   (b2) UFO - SAM-M2F

Method	AP	AP50	AP75	APm	APl	leader	branch	post
Mask2Former [14]	54.289	81.997	54.246	24.341	56.321	57.290	31.146	**74.431**
SAM-M2F	**64.095**	**91.145**	**79.060**	53.384	66.157	**74.922**	**43.086**	74.277