Missing Modality Robustness in Semi-Supervised Multi-Modal Semantic Segmentation

Harsh Maheshwari, Yen-Cheng Liu, Zsolt Kira

Georgia Tech

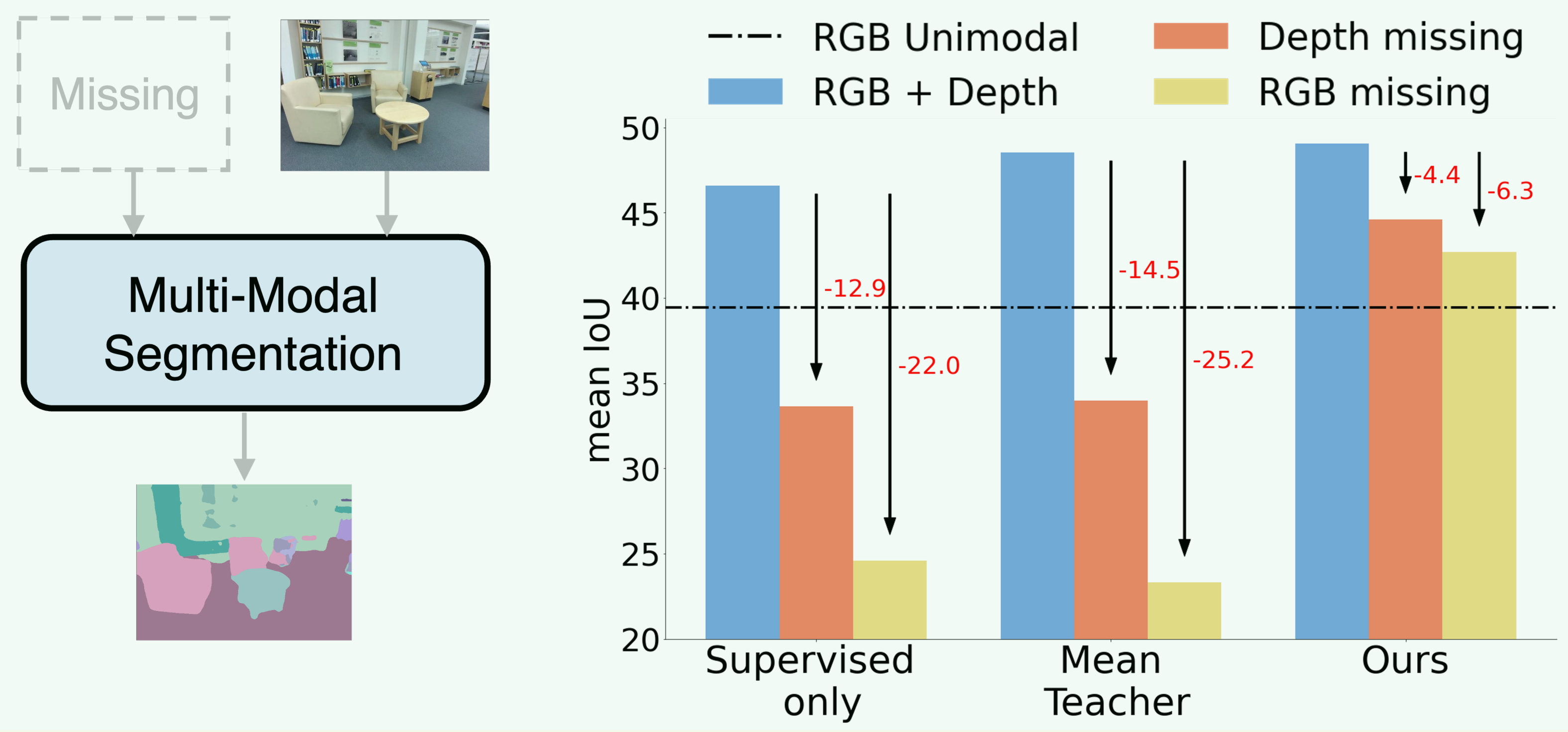
Making multi-modal segmentation more useful

1. Fusion algorithms should work well in low-label regime as labels are scarce - Simpler fusion algorithm
2. Fusion algorithms should work well even if a modality is missing at test time - Missing modality robustness

Solution: Enhancing missing modality robustness in semi-supervised (low-label) multi-modal segmentation by proposing:

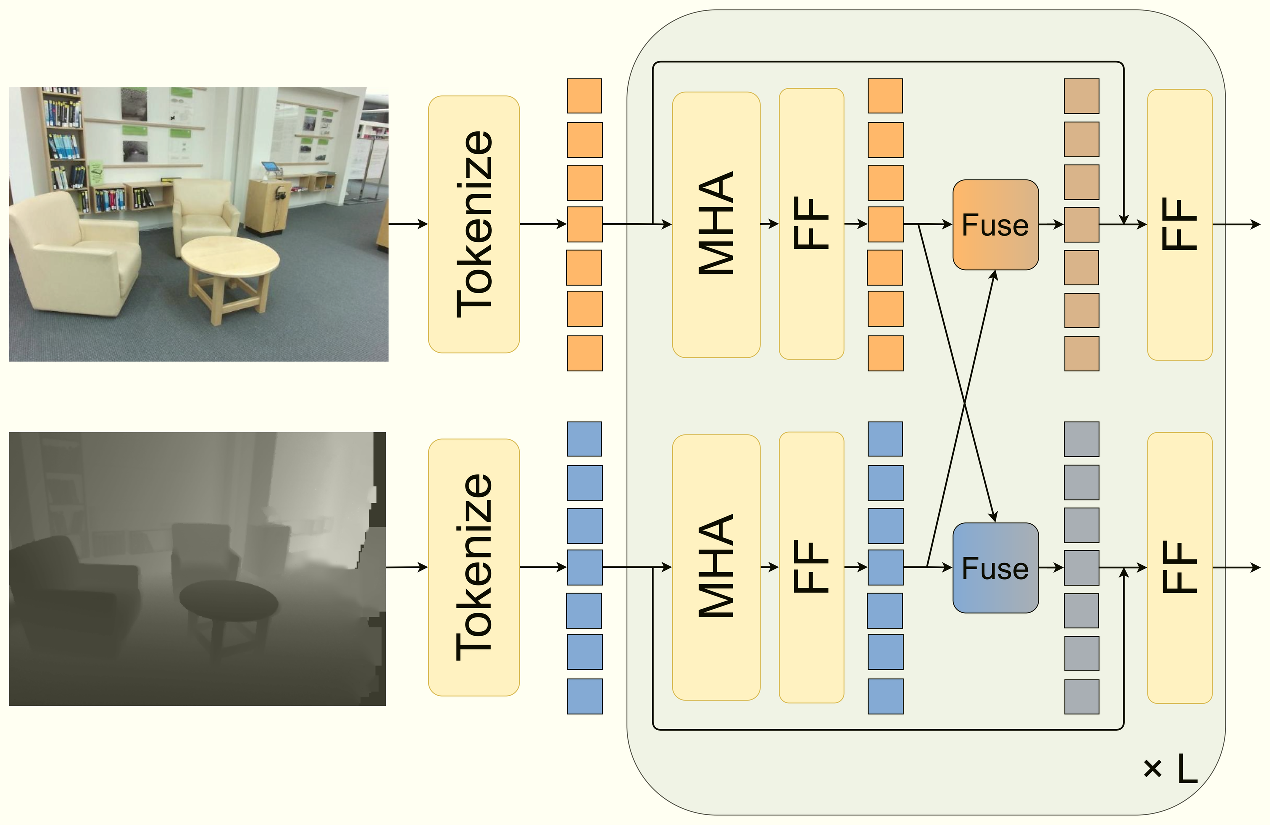
(a) a simpler fusion algorithm, Linear Fusion, that surpasses others with no extra trainable parameters

(b) a semi-supervised framework, M3L, that not only improves multi-modal segmentation performance but also makes the model robust to missing modalities.



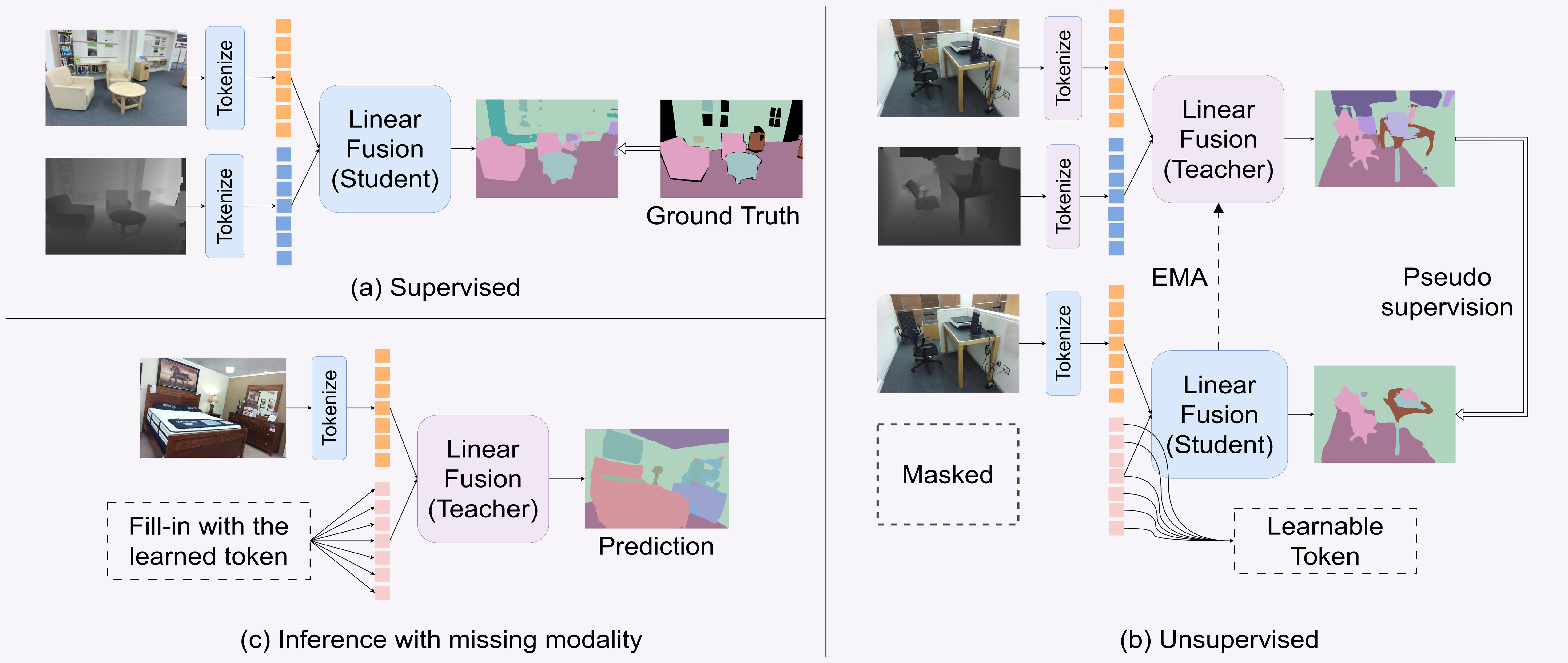
(a) Linear Fusion

Linear Fusion aggregates the tokens of the two modalities by taking a weighted average. This simple algorithm doesn't require any additional trainable parameter and is more effective than prior learnable fusion algorithms.

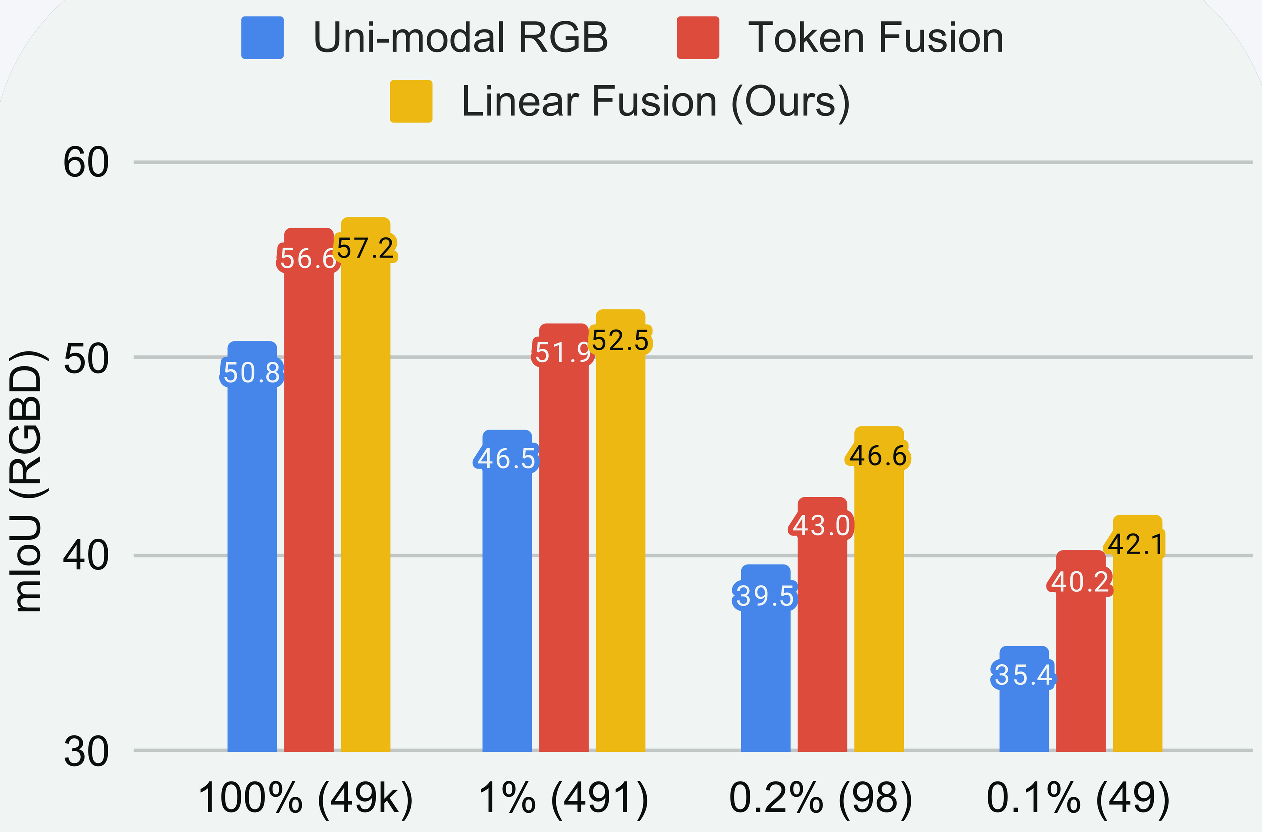


(b) M3L

The proposed M3L (Multi-modal teacher for Masked Modality Learning) semi-supervised framework

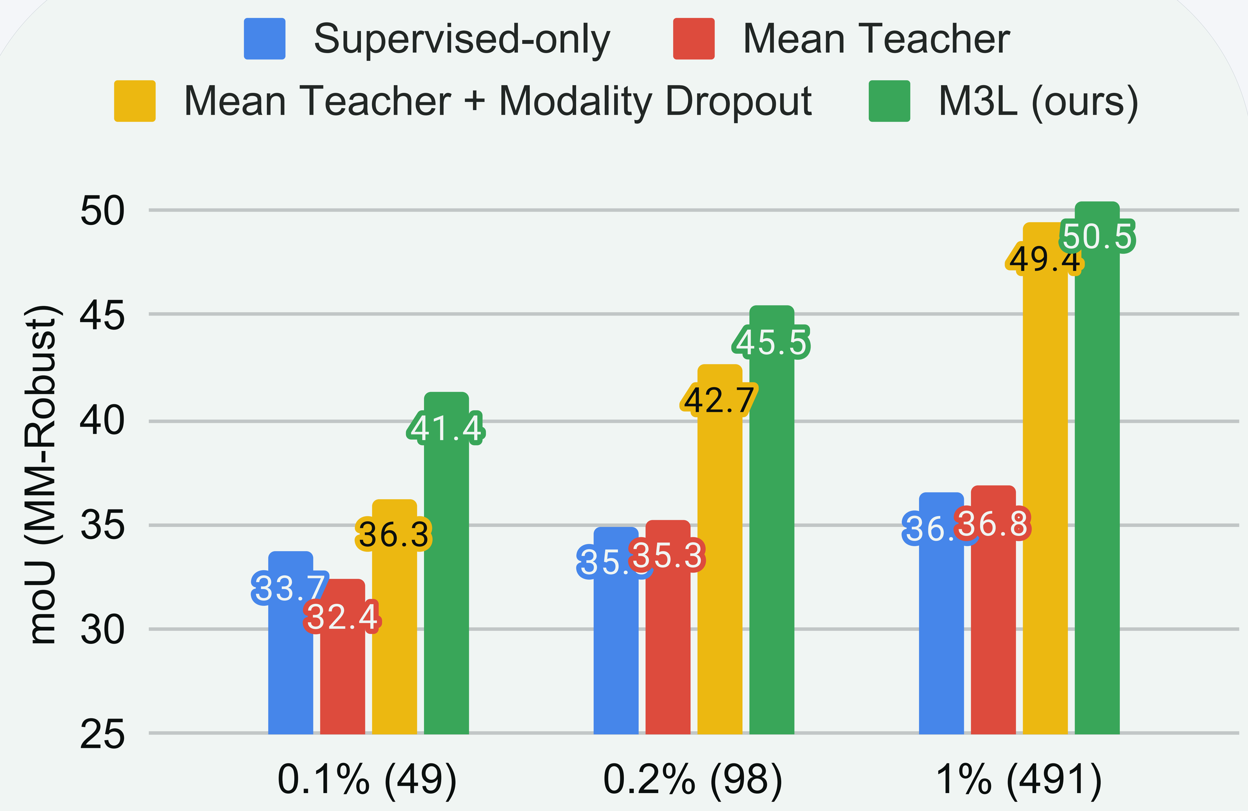


Results



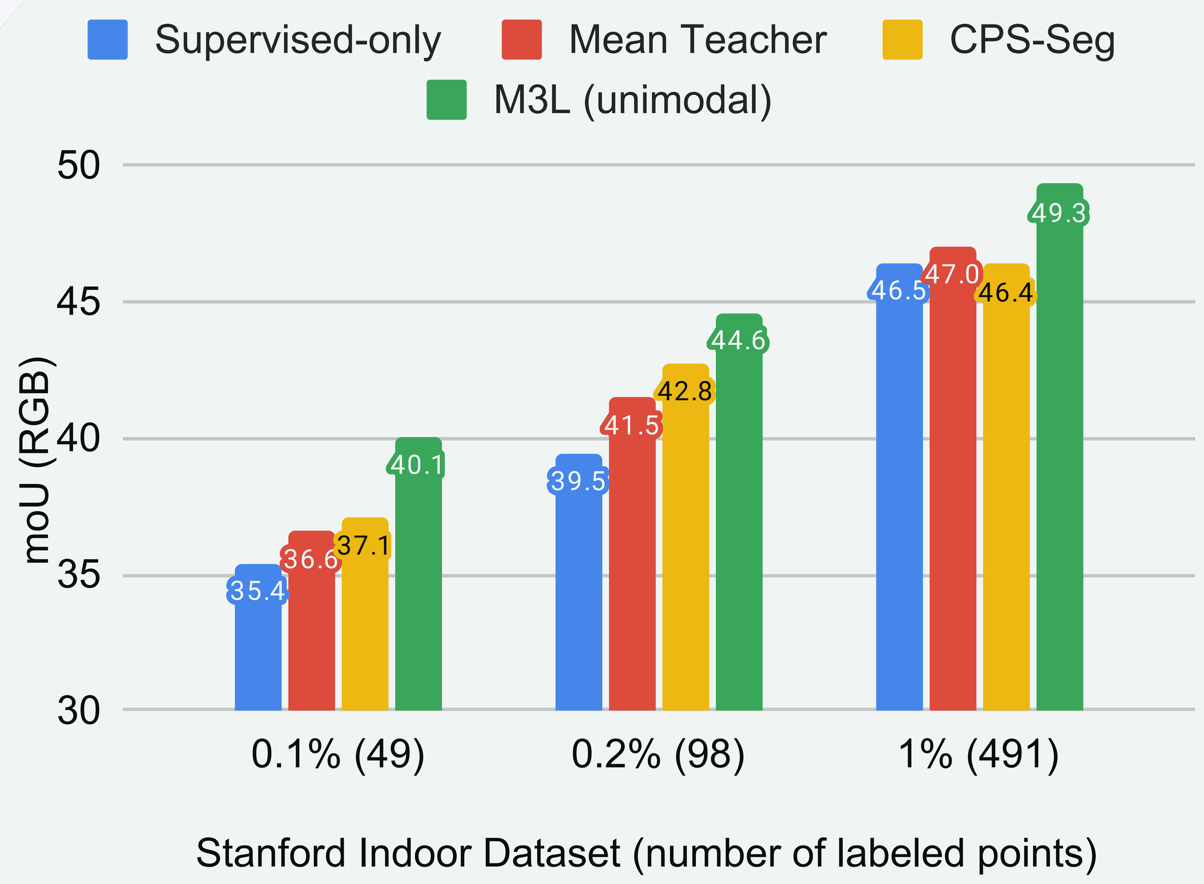
Linear Fusion outperforms Token Fusion [1] in RGBD performance, especially in low-label regime

[1] Multimodal Token Fusion for Vision Transformers, Wang et al., CVPR 2022



M3L successfully utilizes unlabeled data to make the model robust to missing modalities and improve segmentation performance

*MM-Robust is the average of RGBD, RGB-missing and Depth-missing performance



M3L effectively utilizes additional modality during training to improve single modality inference, beating unimodal semi-supervised algorithms (Mean Teacher and CPS [2])

*CPS-Seg is Segformer architecture trained with CPS framework

[2] Semi-Supervised Semantic Segmentation With Cross Pseudo Supervision, Chen et al., CVPR 2021