LEARNING RGB-D FEATURE EMBEDDINGS FOR UNSEEN OBJECT INSTANCE SEGMENTATION

Yu Xiang, 10/12/2020

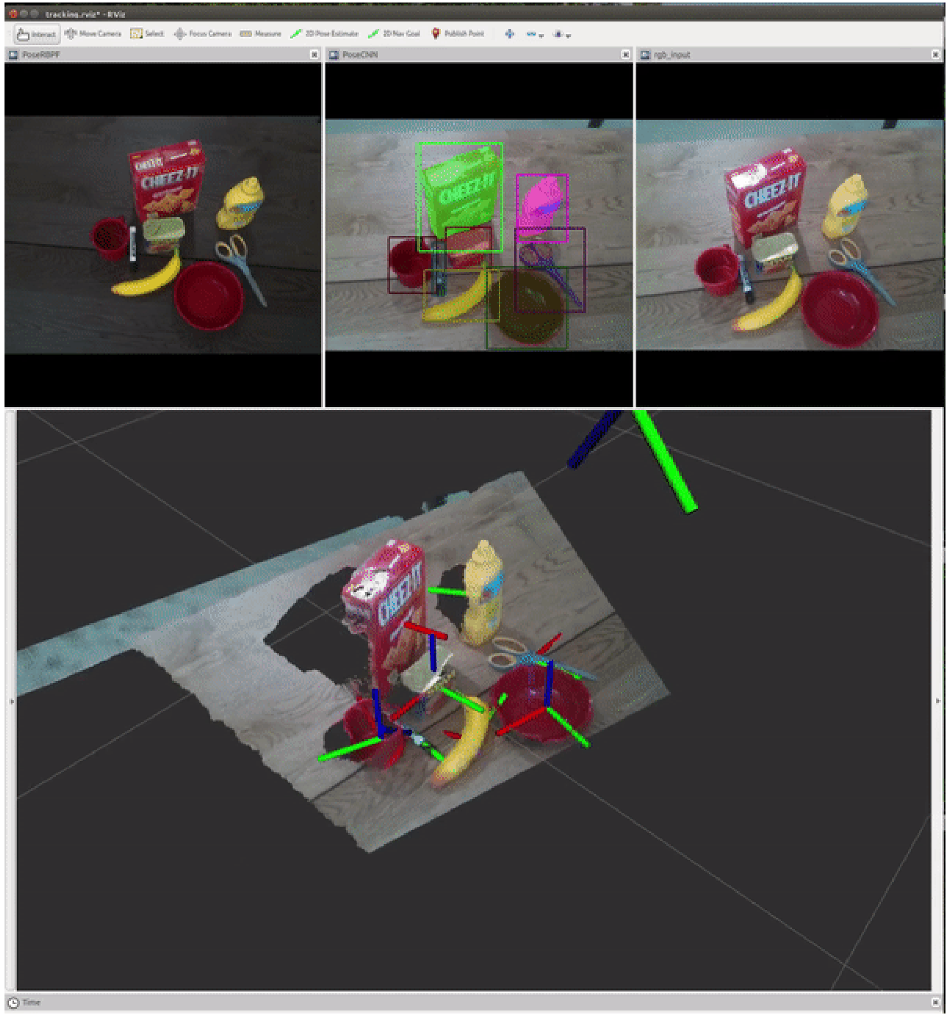
ROBOTS IN UNSTRUCTURED ENVIRONMENTS

How can a robot manipulate objects in this cluttered kitchen?

MODEL-BASED OBJECT RECOGNITION

3D models

Not scalable



PoseCNN + PoseRBPF

Xiang et al. RSS'18
Deng et al. RSS'19

NVIDIA

UNSEEN OBJECT INSTANCE SEGMENTATION

- Can we train a model to segment unseen objects in images?

It is difficult to obtain 3D model for every object.

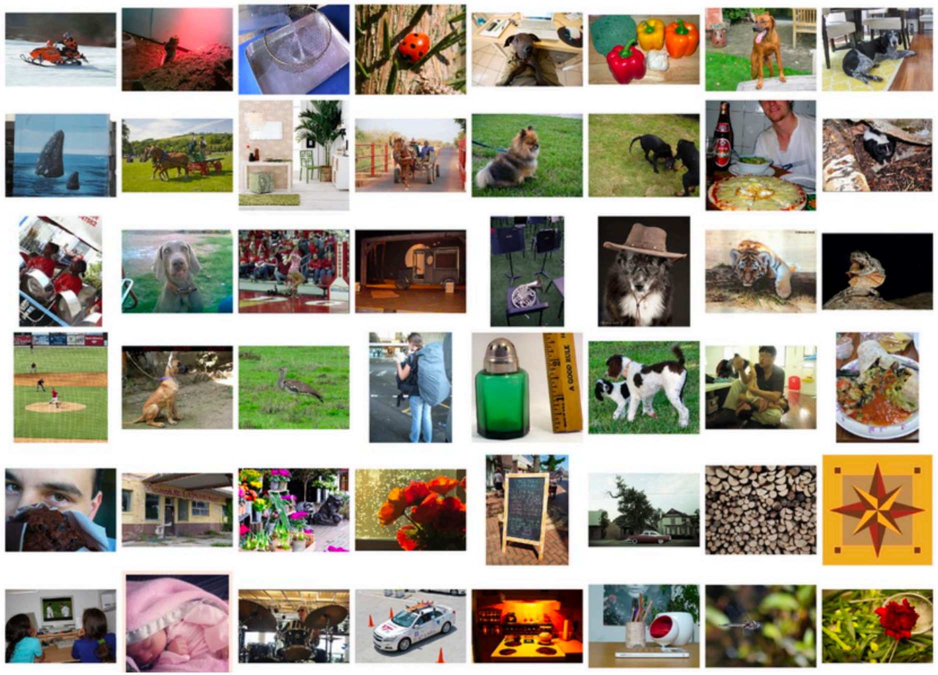
SEGMENTATION ENABLES GRASPING

Unseen Object Segmentation + GraspNet

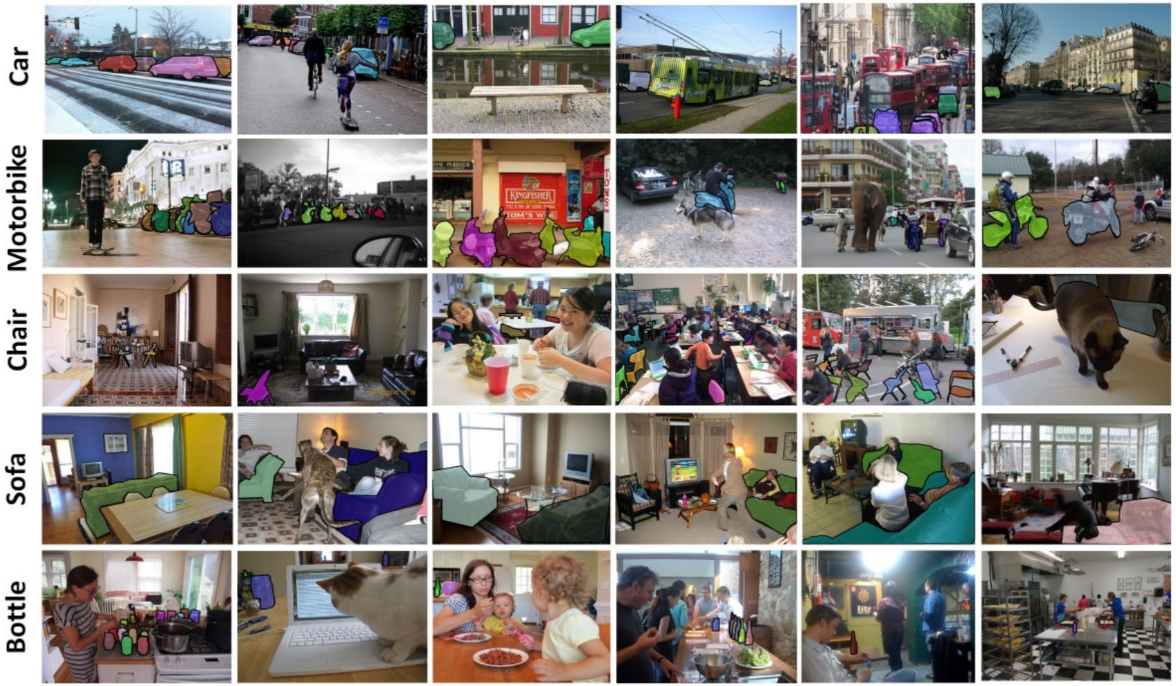
Xie et al. CoRL’19
Mousavian et al. ICCV’19

LEARNING THE CONCEPT OF "OBJECT"

- Learning from data



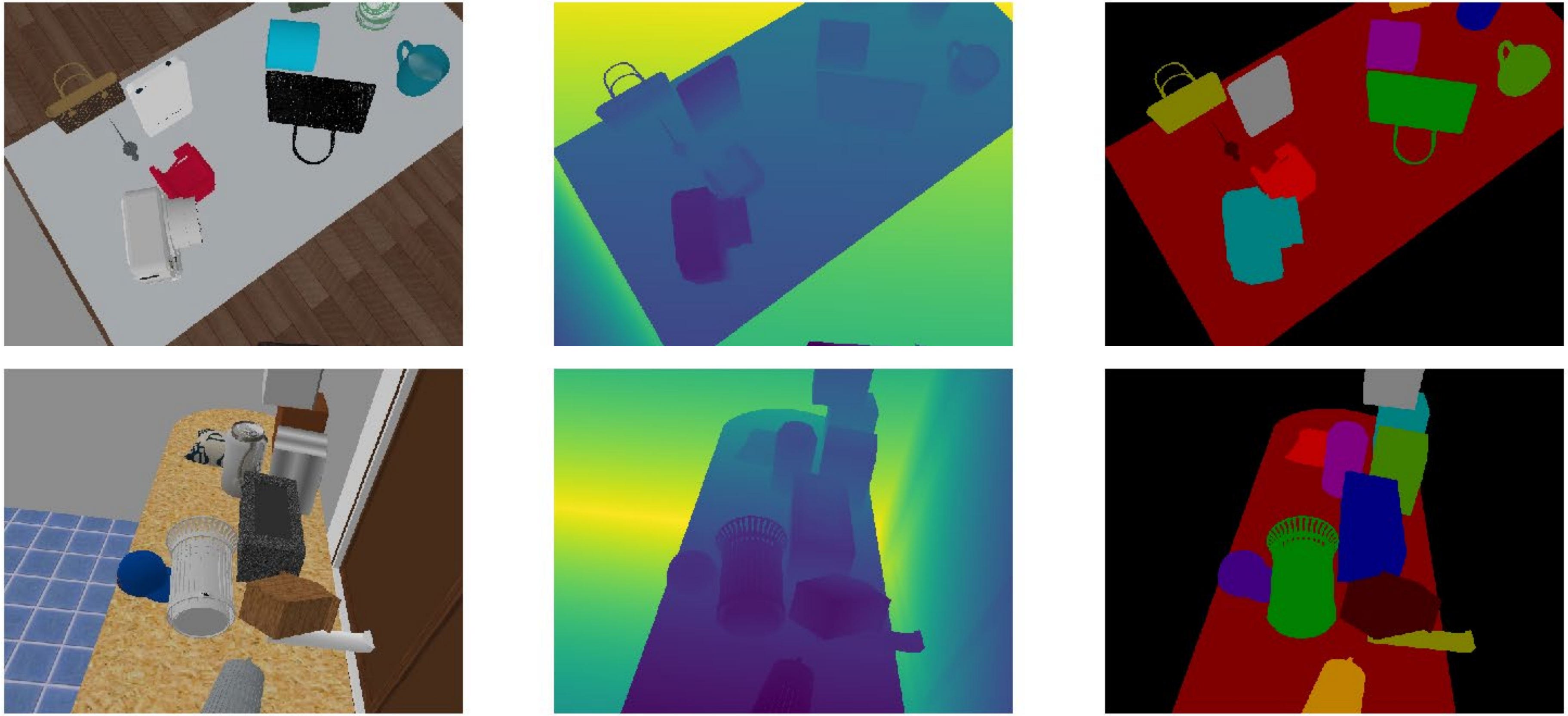
ImageNet Deng et al. CVPR'09



COCO Dataset Lin et al. ECCV'14

Internet Images, not suitable for indoor robotic settings

LEARNING FROM SYNTHETIC DATA



RGB | Depth | Instance Label

40,000 scenes
7 RGB-D images per scene

ShapeNet objects in the PyBullet simulator

Xie et al. CoRL'19

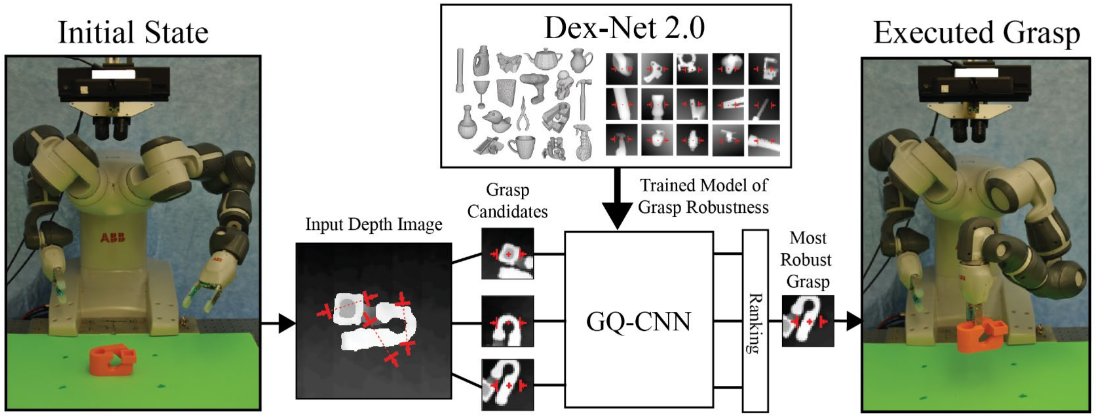
Need to deal with the sim-to-real gap

PREVIOUS WORKS: LEARNING FROM DEPTH

- Synthetic depth generalizes better to the real depth images

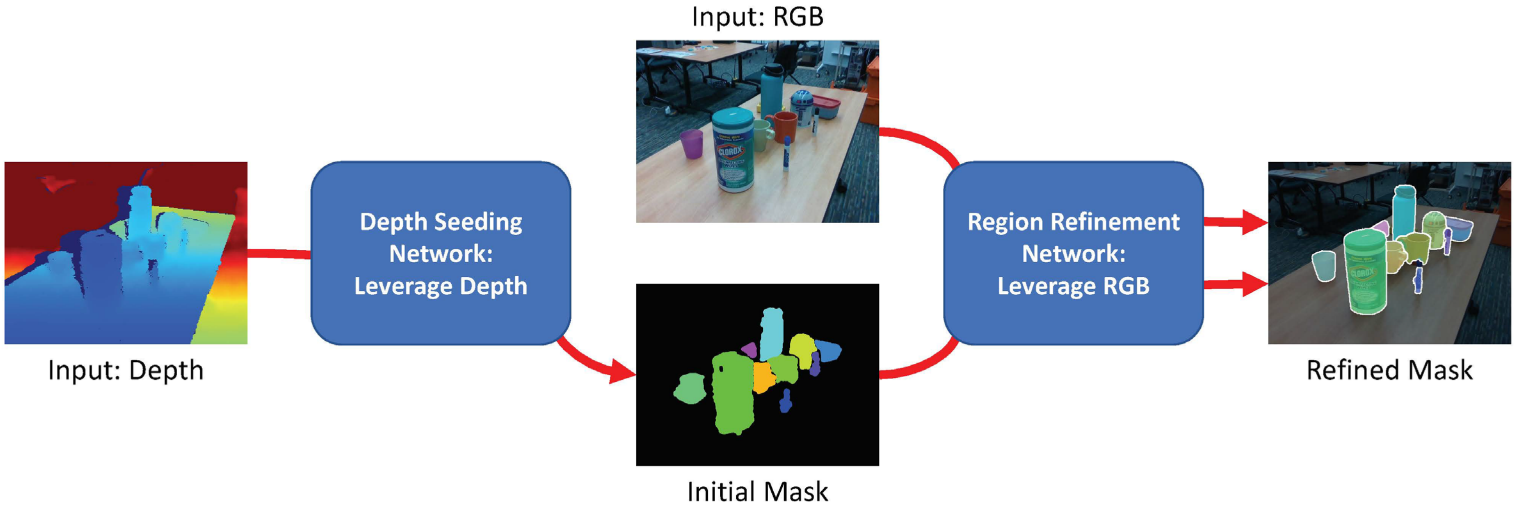
Dex-Net 2.0

Mahler et al. RSS’17



UOIS-Net

Xie et al. CoRL’19

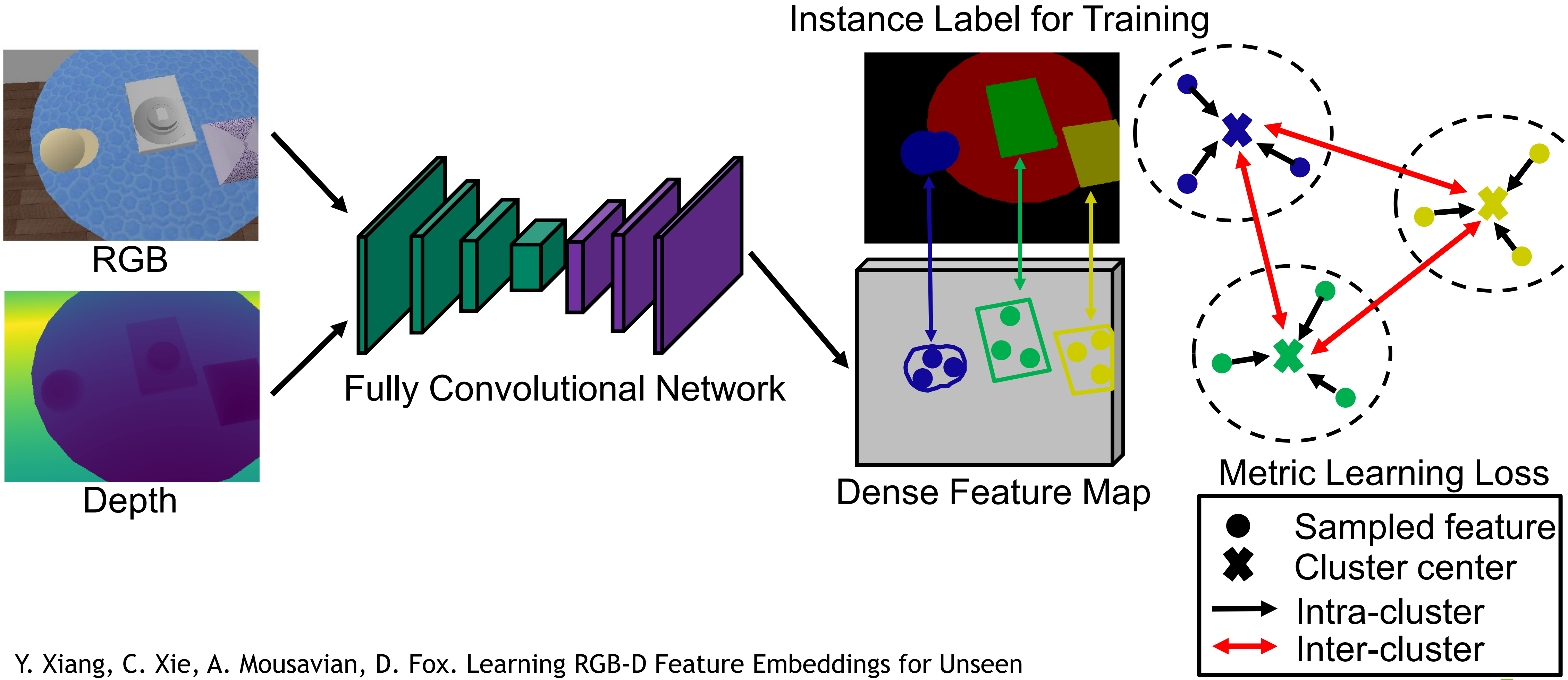


CAN WE UTILIZE NON-PHOTOREALISTIC SYTHETIC RGB IMAGES?

- Depth is not good for transparent objects or thin objects

ClearGrasp
Sajjan et al. ICRA'20

OUR WORK: LEARNING RGB-D FEATURE EMBEDDINGS FOR SEGMENTATION



Y. Xiang, C. Xie, A. Mousavian, D. Fox. Learning RGB-D Feature Embeddings for Unseen Object Instance Segmentation. arXiv:2007.15157, 2020.

METRIC LEARNING LOSS FUNCTION

- Intra-cluster loss function

$$\mu^k = \frac{\sum_{i=1}^{N} \mathbf{x}_i^k}{\|\sum_{i=1}^{N} \mathbf{x}_i^k\|}$$

Spherical mean

$$d(\mu^k, \mathbf{x}_i^k) = \frac{1}{2}(1 - \mu^k \cdot \mathbf{x}_i^k)$$

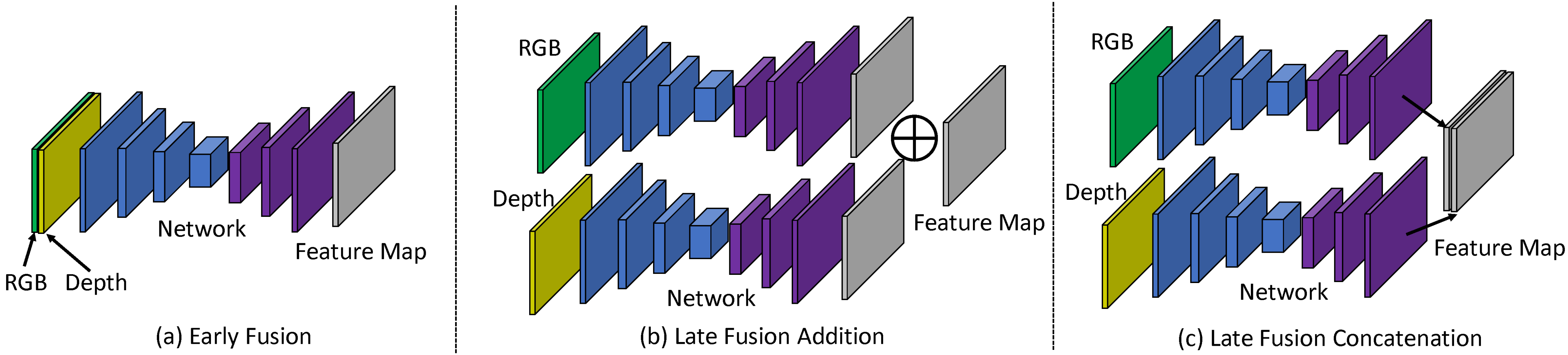
Cosine distance

$$\ell_{\text{intra}} = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{N} \frac{1\left\{d(\mu^k, \mathbf{x}_i^k) - \alpha \geq 0\right\} \; d^2(\mu^k, \mathbf{x}_i^k)}{\sum_{i=1}^{N} 1\left\{d(\mu^k, \mathbf{x}_i^k) - \alpha \geq 0\right\}}$$

- Inter-cluster loss function

$$\ell_{\text{inter}} = \frac{2}{K(K-1)} \sum_{k<k'} \left[\delta - d(\mu^k, \mu^{k'})\right]_+^2$$

FUSING RGB AND DEPTH



MEAN SHIFT CLUSTERING

- von Mises-Fisher (vMF) mean shift for unit length vectors

Kobayashi and Otsu. ICPR'10

- Find local maxima of the von Mises-Fisher distribution

$$p(\mathbf{x}; \mu, \kappa) = C(\kappa) \exp(\kappa \mathbf{x}^T \mu)$$

Algorithm 1: von Mises-Fisher mean shift clustering

Input: Feature embedding matrix $\mathbf{X} \in \mathbb{R}^{n \times C}$, κ, ϵ, number of seed m, number of iteration T

Sample m initial clustering centers from $\mathbf{X}$ as the m furthest points, denote it as $\mu^{(0)} \in \mathbb{R}^{m \times C}$;

for $t \leftarrow 1$ **to** T **do**

  Compute weight matrix $\mathbf{W} \leftarrow \exp(\kappa \mu^{(t-1)} \mathbf{X}^T)$;

  Update $\mu^{(t)'} \leftarrow \mathbf{WX}$;

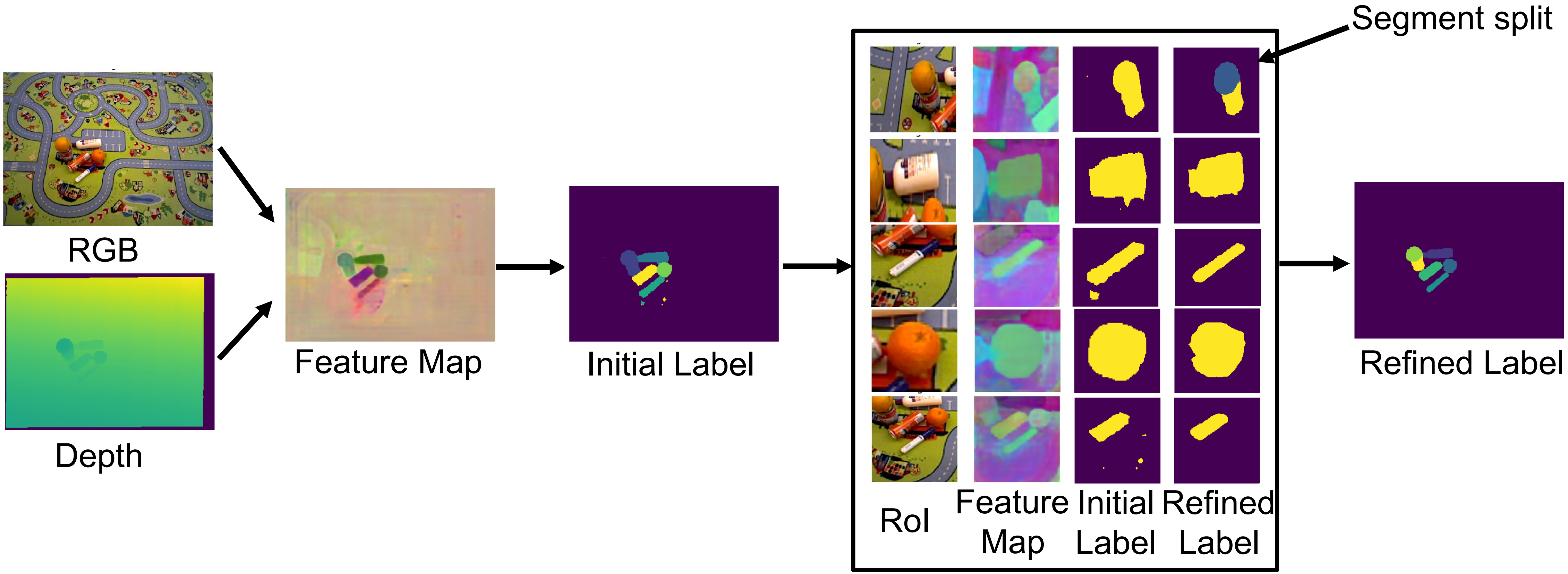
  Normalize each row vector in $\mu^{(t)'}$ to obtain $\mu^{(t)}$;

end

Merge cluster centers in $\mu^{(T)}$ with cosine distance smaller than ϵ ;

Assign each pixel to the closest cluster center ;

TWO-STAGE CLUSTERING



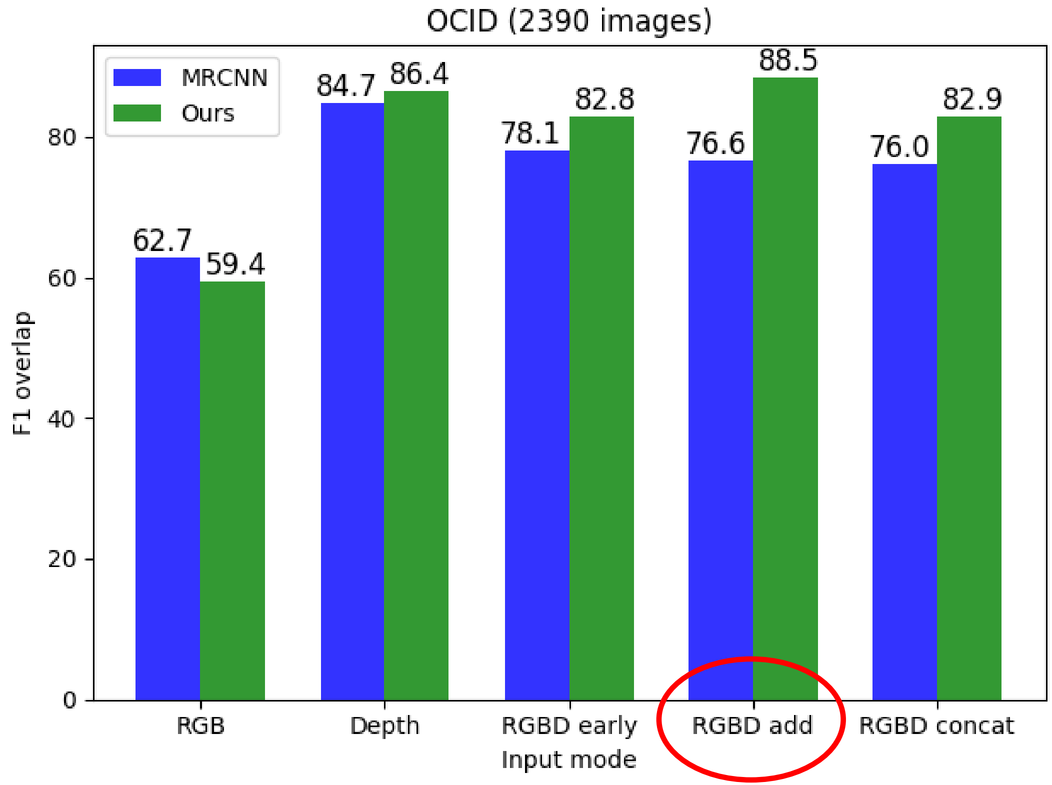
EXPERIMENTS: DATASETS

- Object Cluster Indoor Dataste (OCID), 2,390 RGB-D images — Sushi et al. ICRA'19

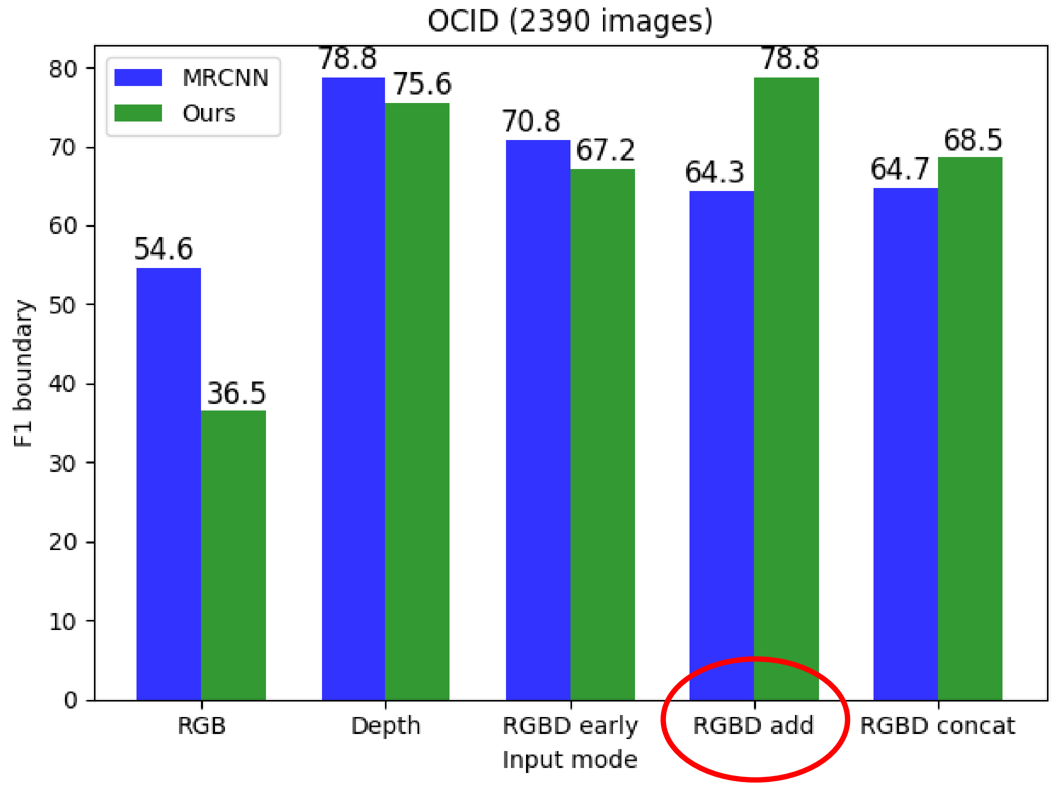
- Object Segmentation Database (OSD), 111 RGB-D images — Richtsfeld et al. IROS'12

EFFECT OF THE INPUT MODE

Mask R-CNN. He et al. CVPR'17

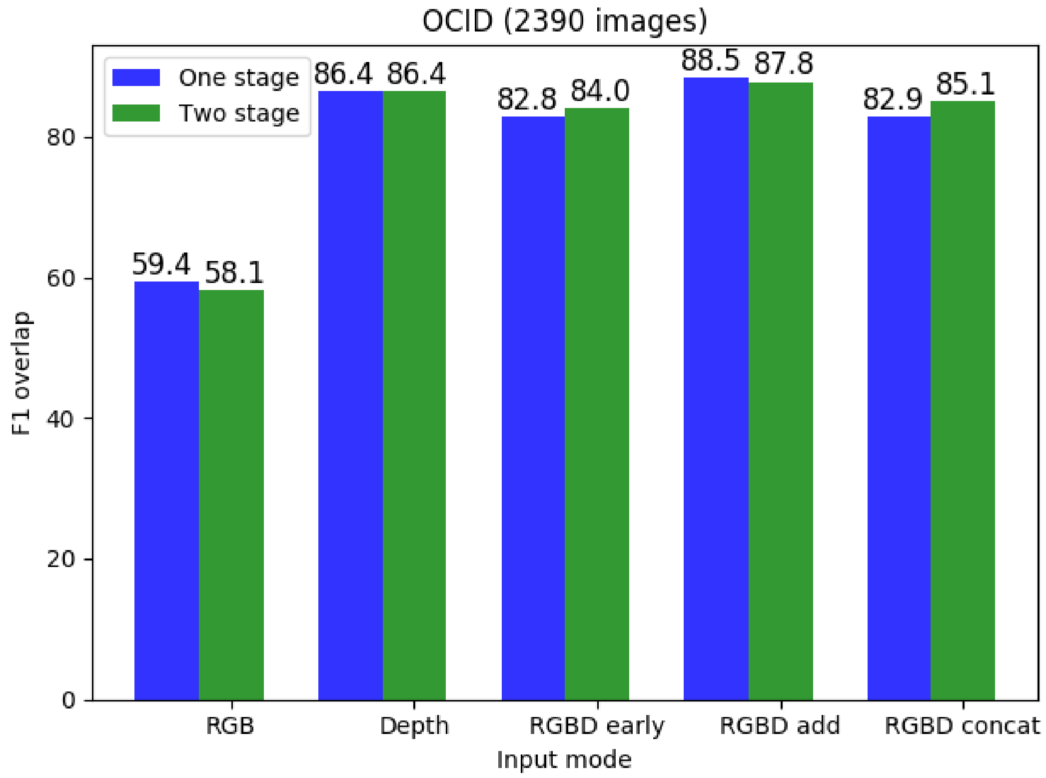


F-score overlap

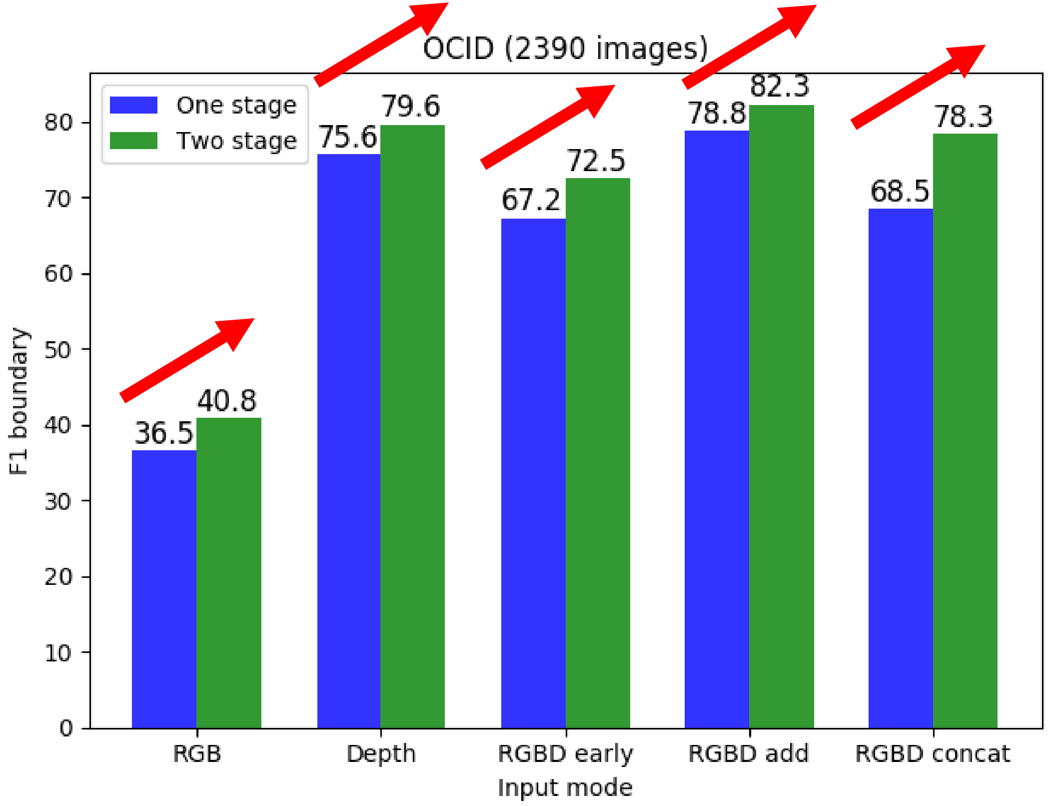


F-score boundary

EFFECT OF THE TWO-STAGE CLUSTERING

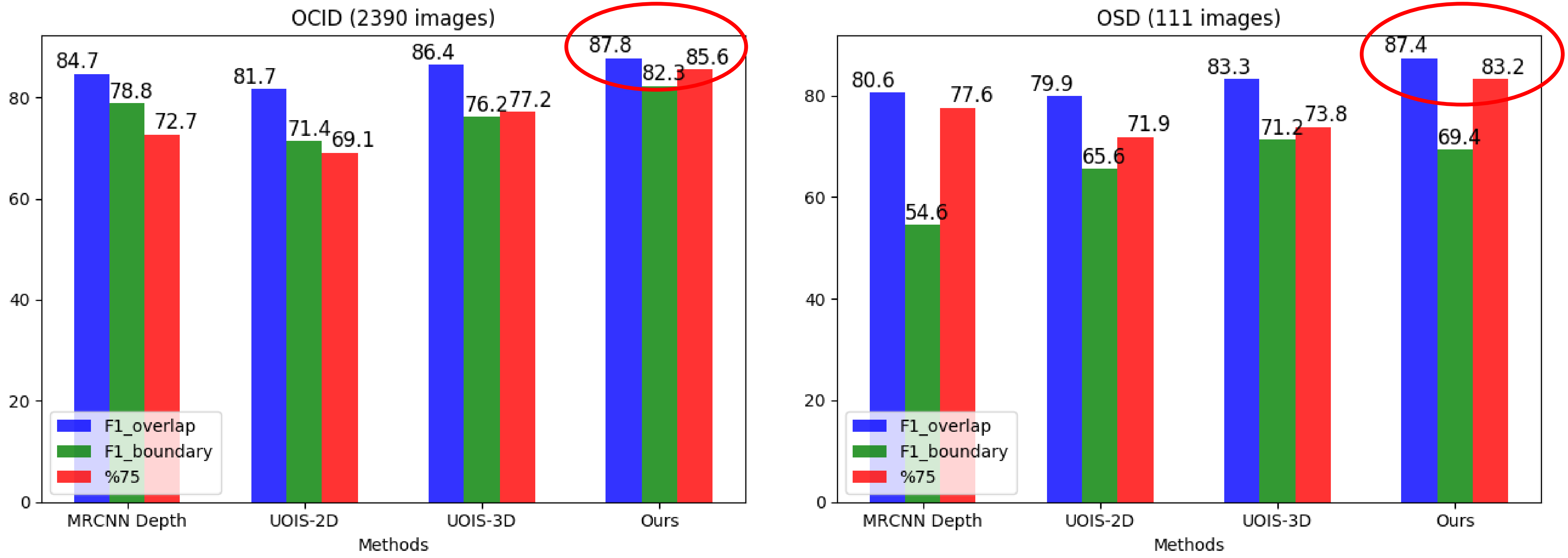


F-score overlap

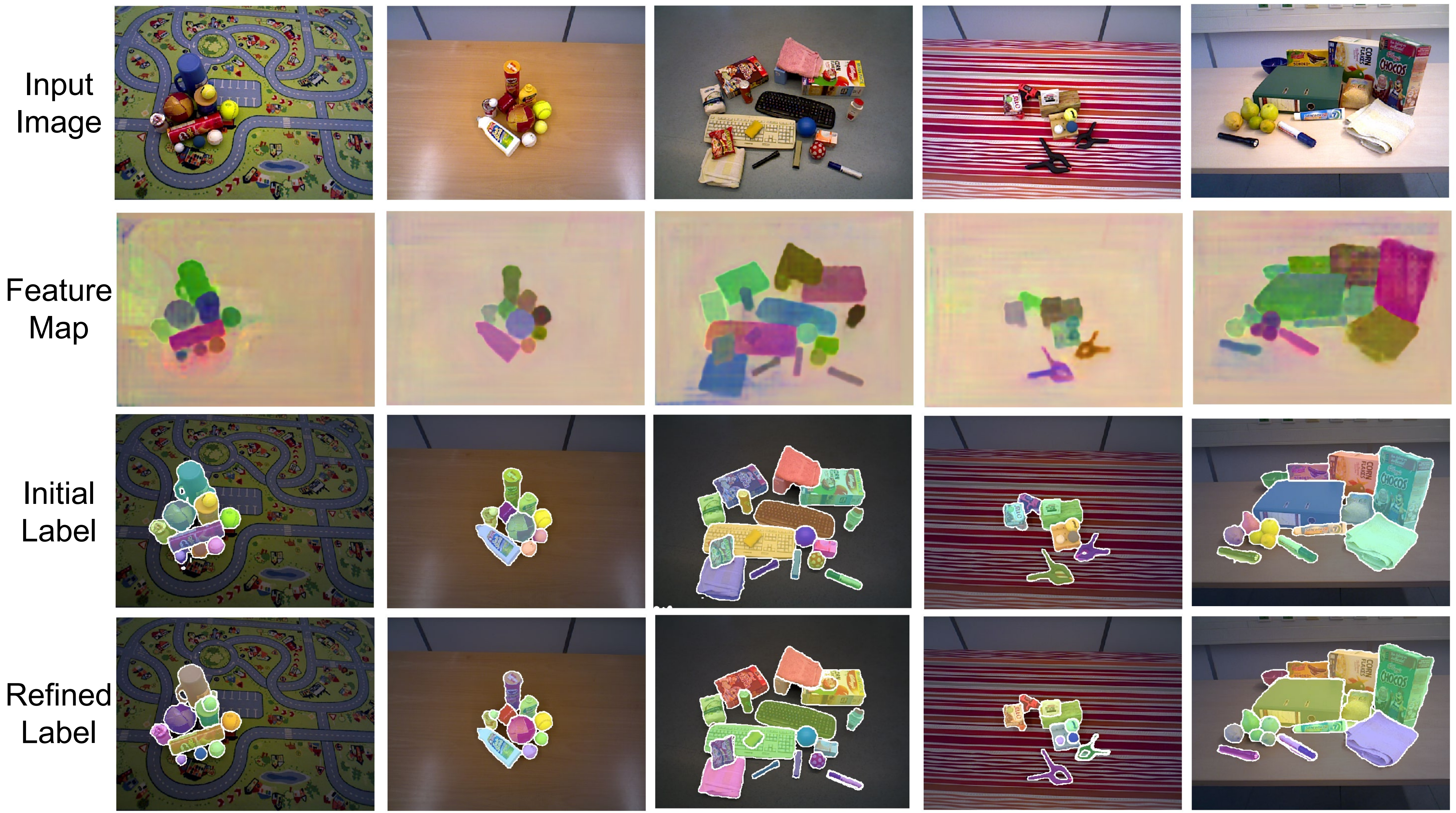


F-score boundary

COMPARISON TO STATE-OF-THE-ARTS



Mask R-CNN. He et al. CVPR’17
UOIS-2D. Xie et al. CoRL’19
UOIS-3D. Xie et al. arXiv:2007.08073


Input Image
Feature Map
Initial Label
Refined Label

FAILURE CASES

Input Image

Final Label

ANECDOTAL EXAMPLE ON TRANSPARENT OBJECTS

ClearGrasp
Sajjan et al. ICRA'20

CONCLUSION

- Learning RGB-D feature embeddings from synthetic data with a metric learning loss that transfers well to the real world
- Adding non-photorealistic RGB images to Depth can still improve in our method
- Using RGB images can handle objects with bad or missing depth information such as transparent, flat or thin objects

Questions?