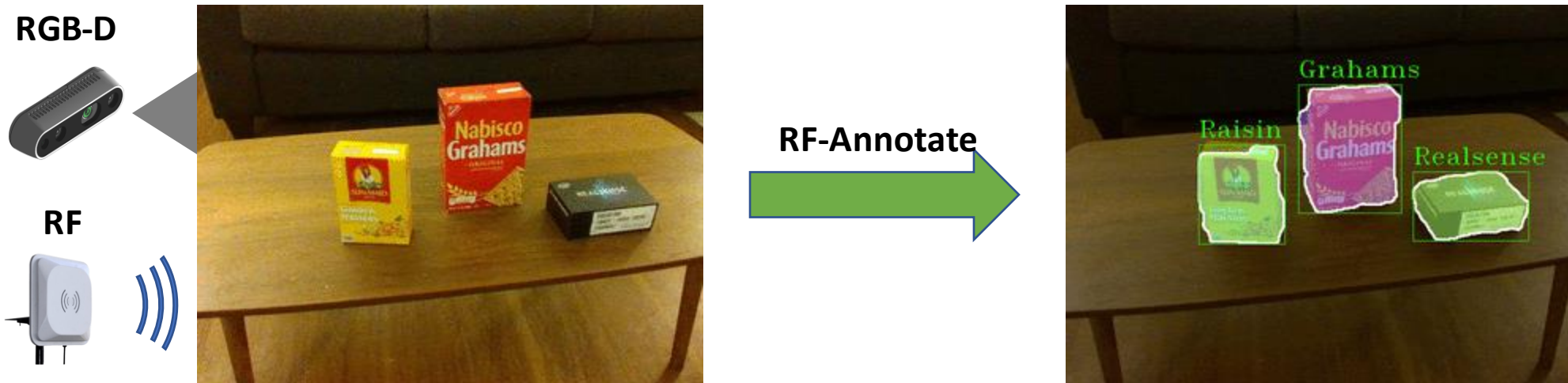




RF-Annotate: Automatic RF-Supervised Image Annotation of Common Objects in Context



Emerson Sie and Deepak Vasisht



UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN



We need large diverse datasets for generalizable robots



We need large diverse datasets for generalizable robots

ARID [ICRA'18]



OCID [ICRA'19]



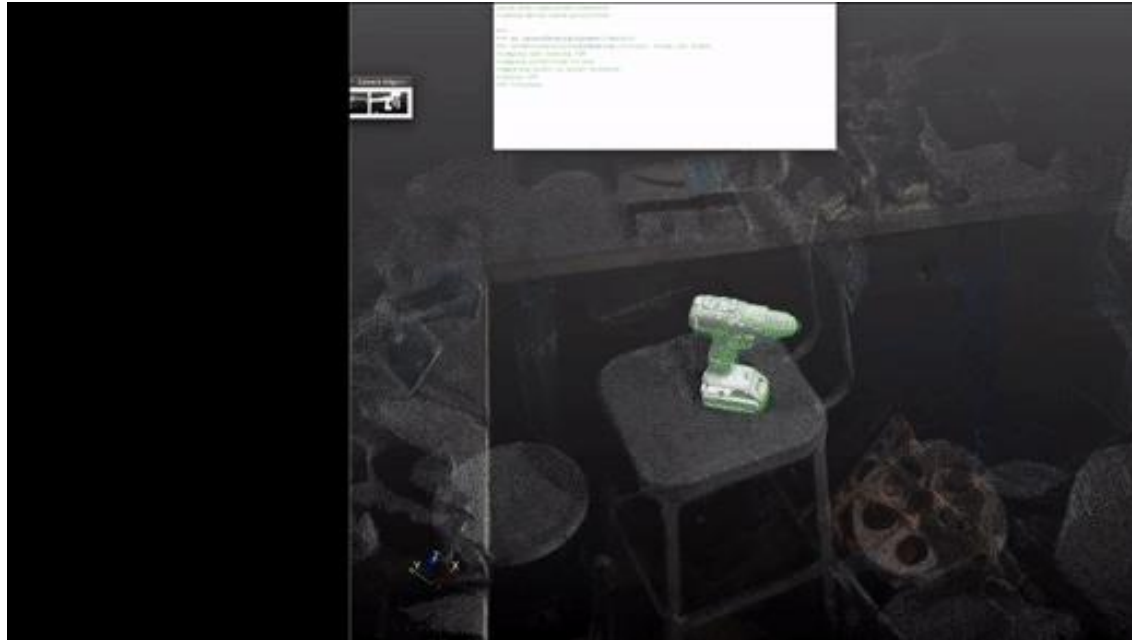
OpenLORIS [ICRA'20]



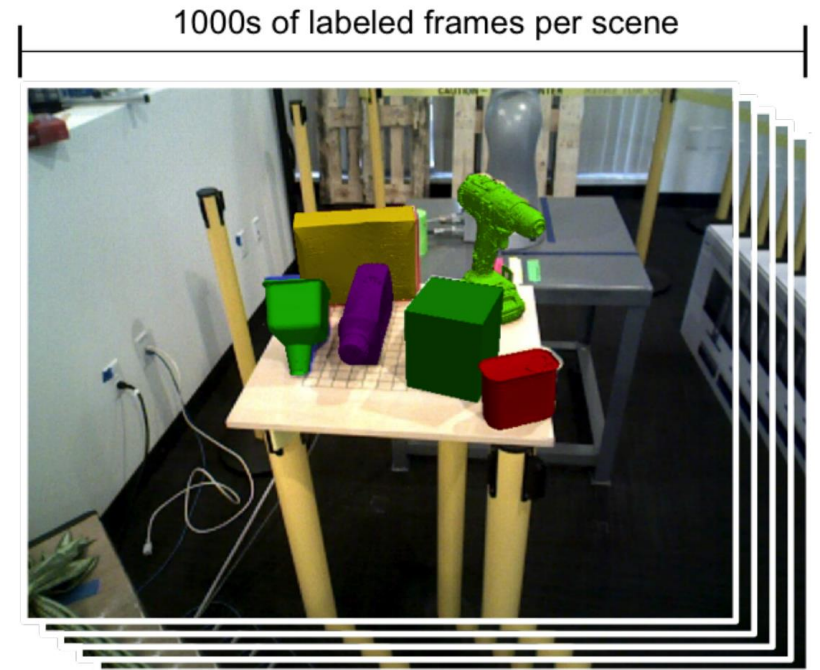
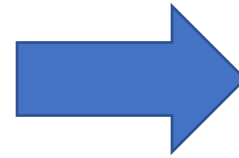
Time



Can existing annotation methods keep up with demand?



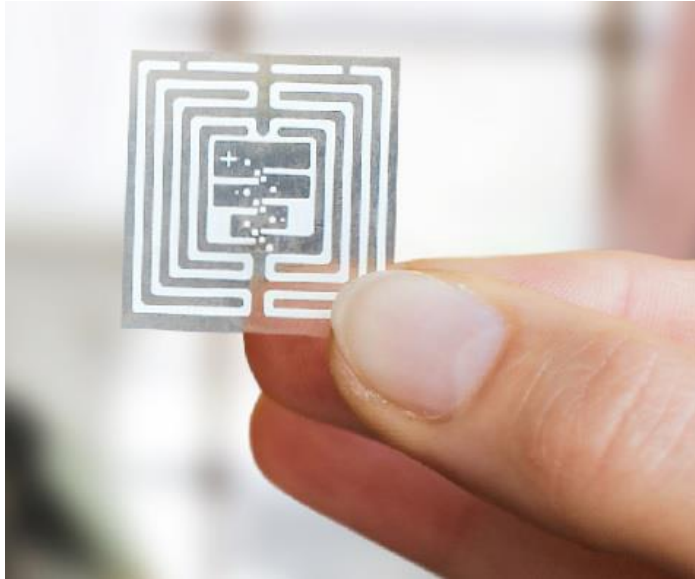
LabelFusion [2]



Human input remains a bottleneck



What can we do about this?



RFID (2000s)



Tile Tracker (2012)



Apple AirTag (2021)

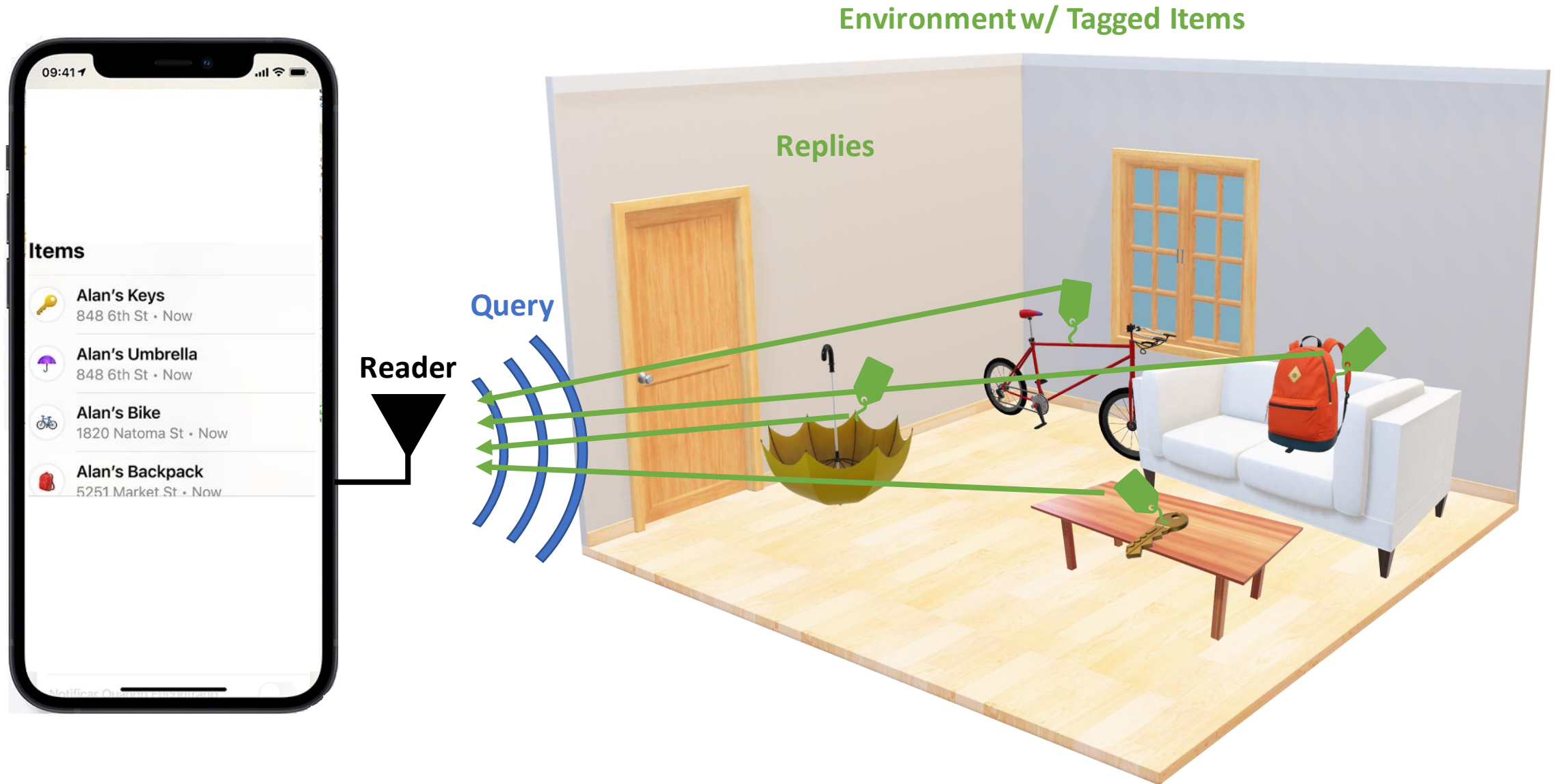


Time

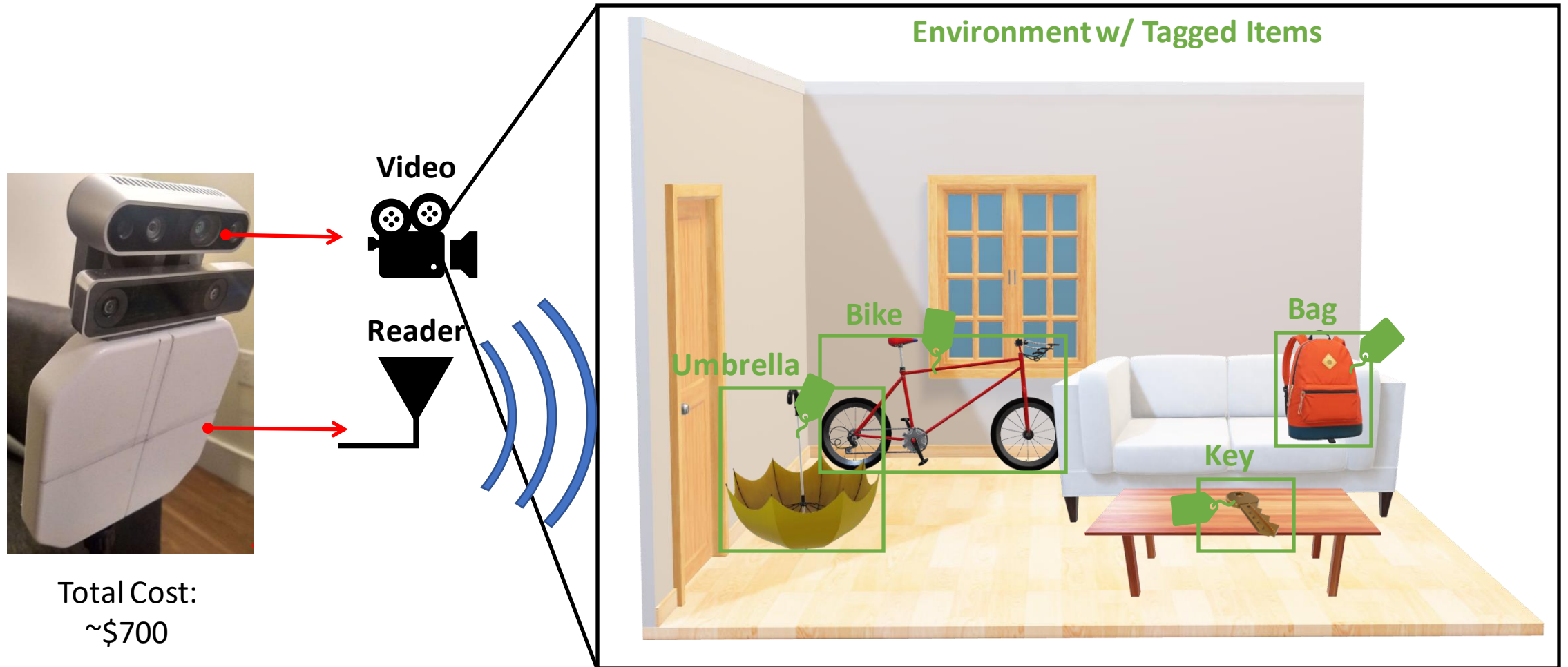
Wireless tracker tags are becoming increasingly common!



How do tracker tags work?



This Talk: Use them to annotate ~100s of video frames at once



a) at the *pixel* level and b) no clicks necessary



This Talk: Use them to annotate ~100s of video frames at once

Tagged Items



3-5¢ per RFID

Experimental Setup



Why RFID? Because they are battery-free and very cheap.



How to generate object instance masks?

Use category-agnostic instance segmentation on depth.

Depth Camera



RGB



Depth

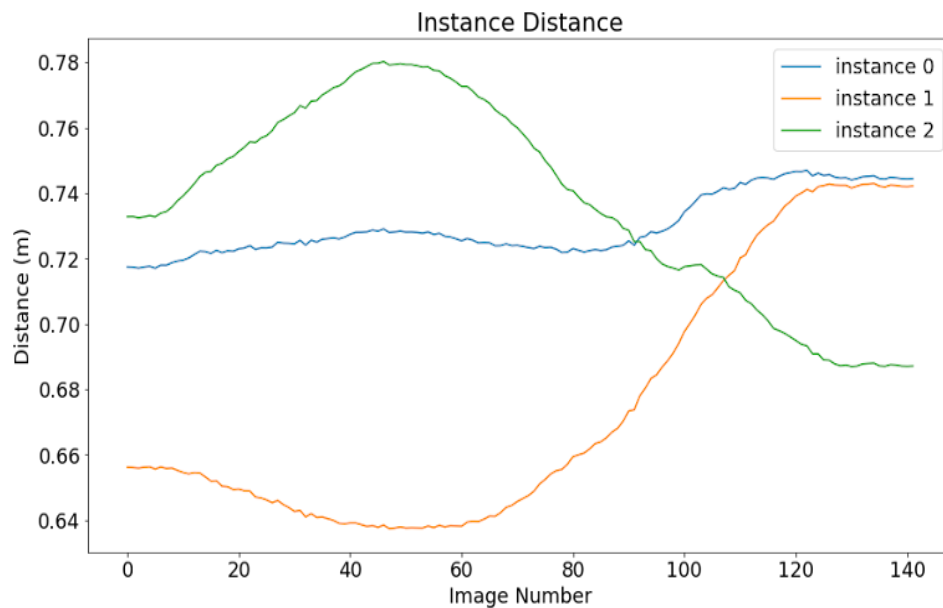
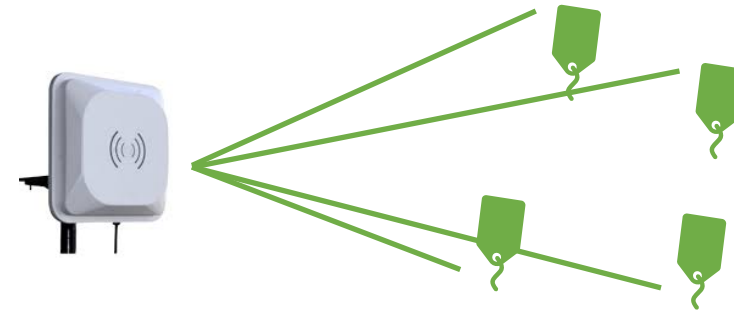
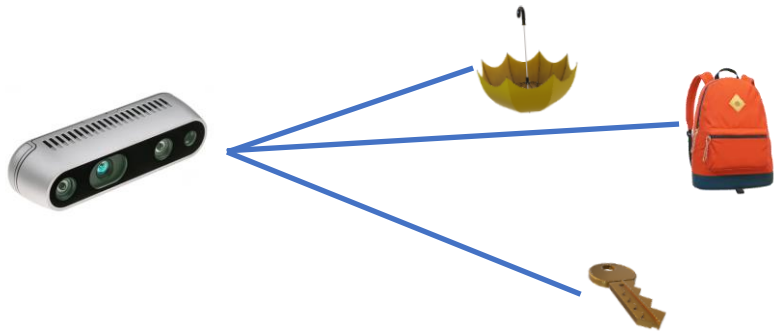
[3]



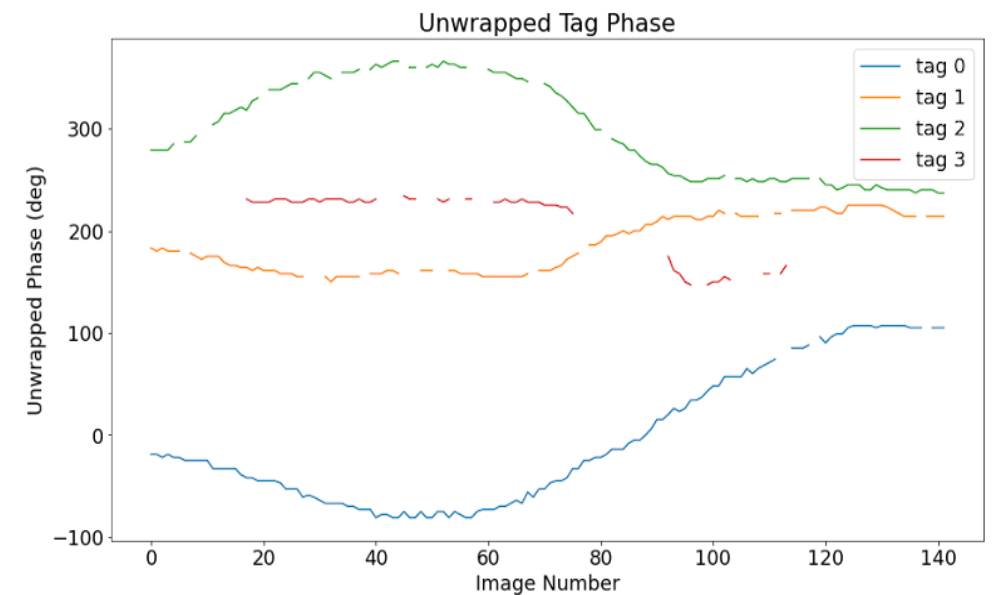
Instance Proposals

How to assign tag label to each instance mask?

Use sensor motion and correlate spatial profiles.



Bipartite Matching



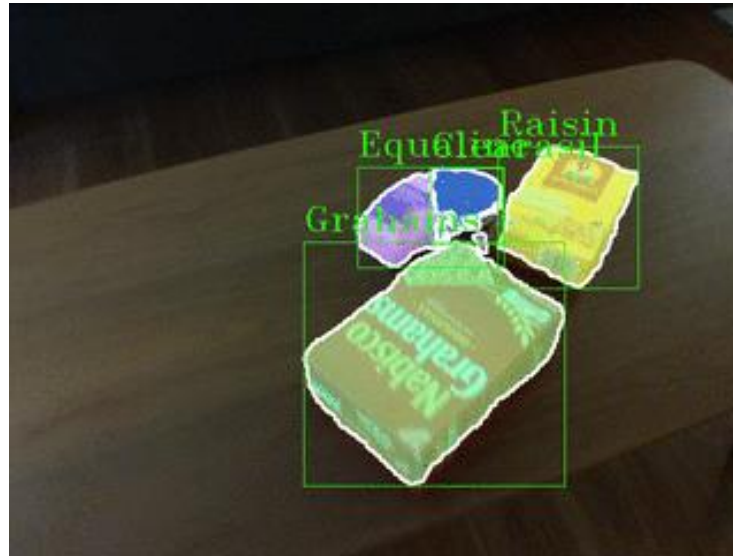
Qualitative Results

Evaluated on >3000 frames of objects arranged in varying levels of clutter [4].

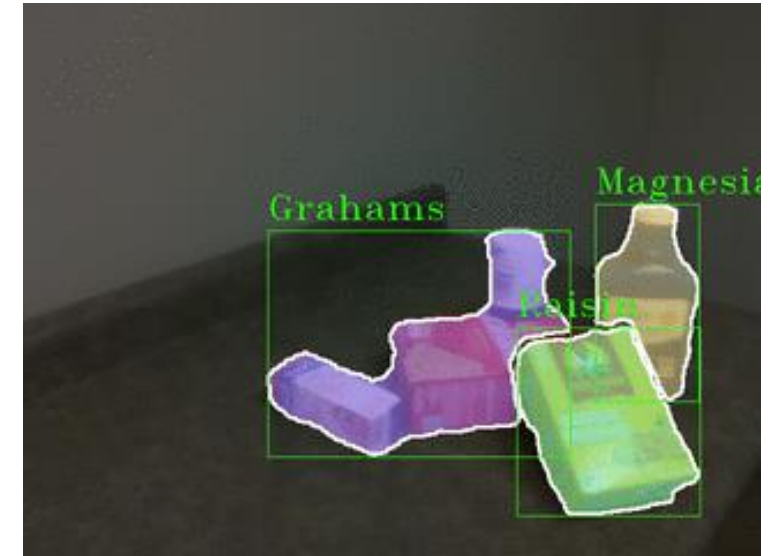
Separated



Touching



Stacked



Increasing Object Closeness/Difficulty

Quantitative Results

Compared output of pipeline with human-annotated videos as ground truth.

Metric	Separated	Touching	Stacked
Instance Recall	0.96	0.86	0.91
Matching Accuracy	1	0.8	0.88

Key Takeaways

- Perfect matching accuracy when objects are physically separated.
- Limiting factor is mask generation [3], not RF component.

Condition		Mask Overlap (F)		Recall@0.75
Separated	↑	0.88	↑	0.86
Touching	→	0.78	↓	0.62
Stacked	→	0.75	→	0.77

Under-segmentation

