2021 Intelligent Sensing Winter School

An Affordance Detection Pipeline for Resource-Constrained Devices

Tommaso Apicella[1,2], Andrea Cavallaro[2], Riccardo Berta[1], Paolo Gastaldo[1], Francesco Bellotti[1] and Edoardo Ragusa[1]

[1] DITEN, University of Genoa

[2] CIS, Queen Mary University of London

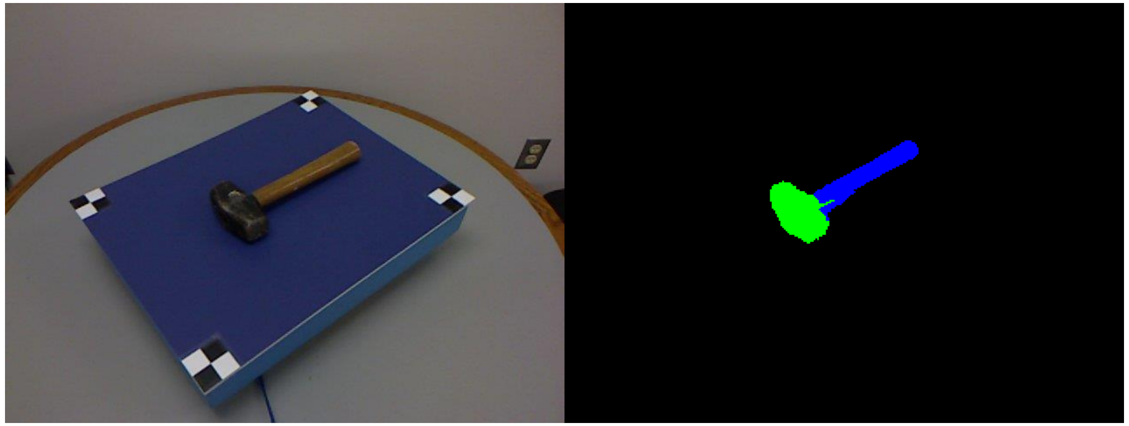
CIS centre for intelligent sensing

Università di Genova

Queen Mary University of London

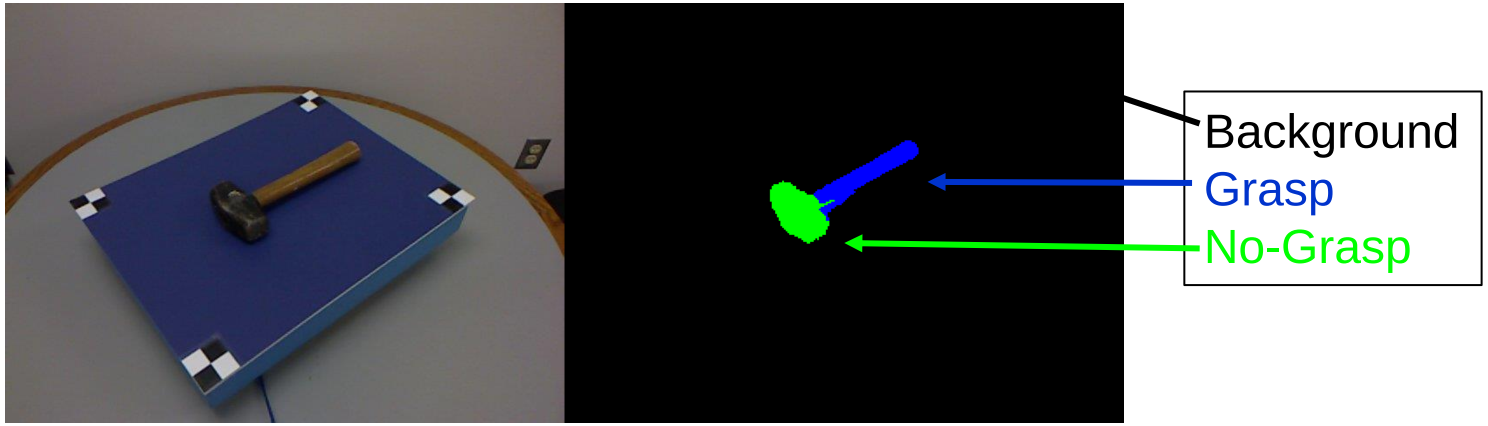
Affordance detection

Segmenting parts of objects based on potential interaction with a human.



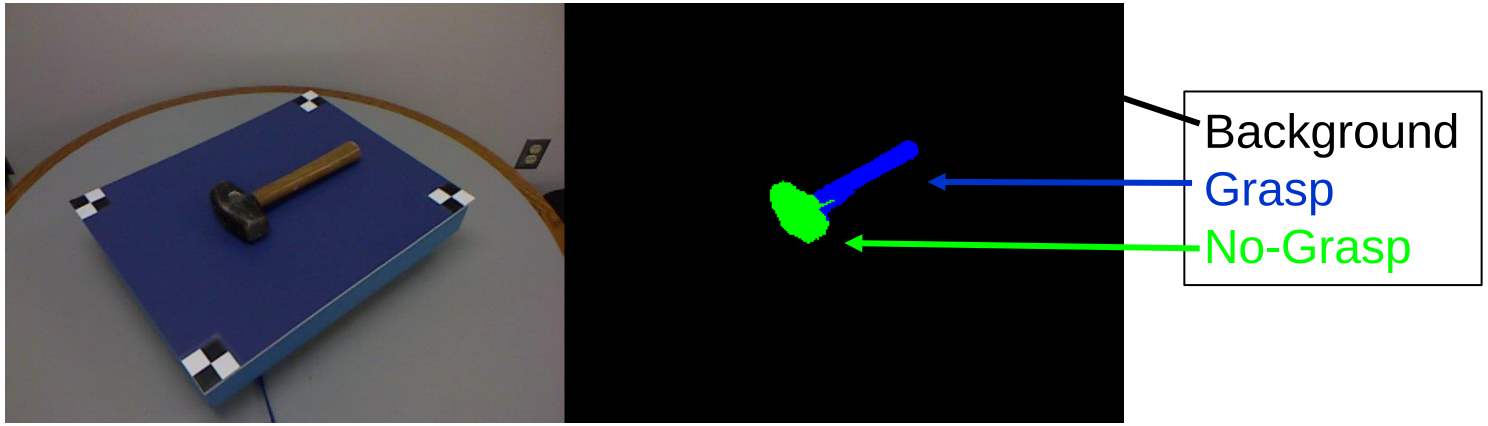
Affordance detection

Segmenting parts of objects based on potential interaction with a human.



Affordance detection

Segmenting parts of objects based on potential interaction with a human.



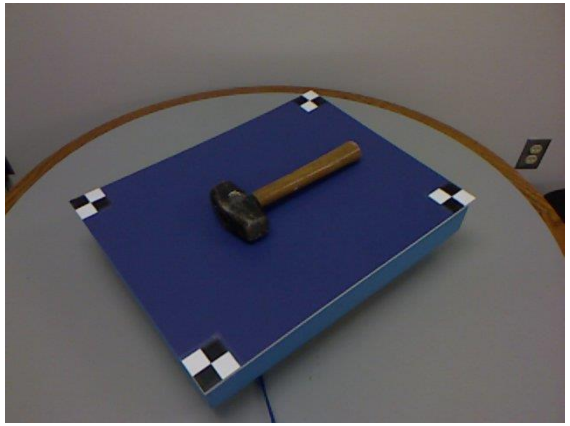
Wearable context (prosthetic):

- Resource-constrained devices
- Robotic solutions require remarkable computational power
- Human-in-the-loop application

Baseline

Models for portable embedded systems [1] assume objects to be:

- In foreground
- Completely visible

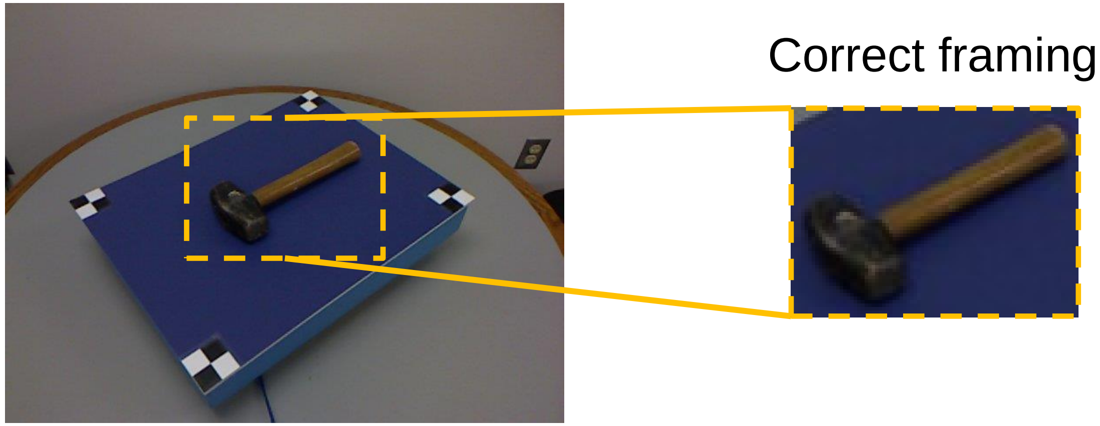


[1] Ragusa, E. et al., *"Hardware-Aware Affordance Detection for Application in Portable Embedded Systems",* IEEE Access, 2021.

Baseline

Models for portable embedded systems [1] assume objects to be:

- In foreground
- Completely visible

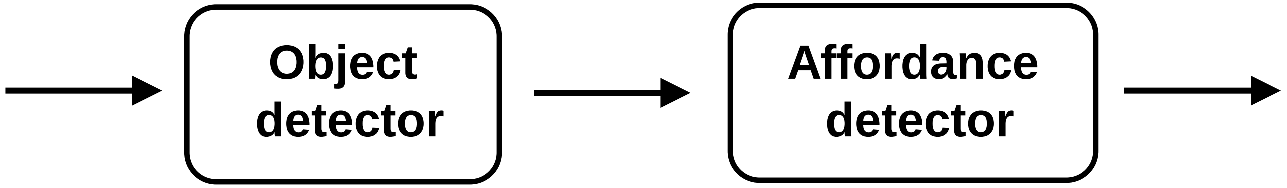


Framing issue: the position of the camera is indirectly controlled by a human

[1] Ragusa, E. et al., *"Hardware-Aware Affordance Detection for Application in Portable Embedded Systems",* IEEE Access, 2021.

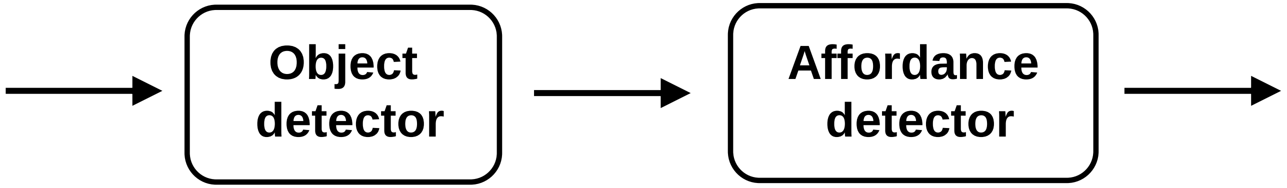
Contributions

- Overcome framing issues

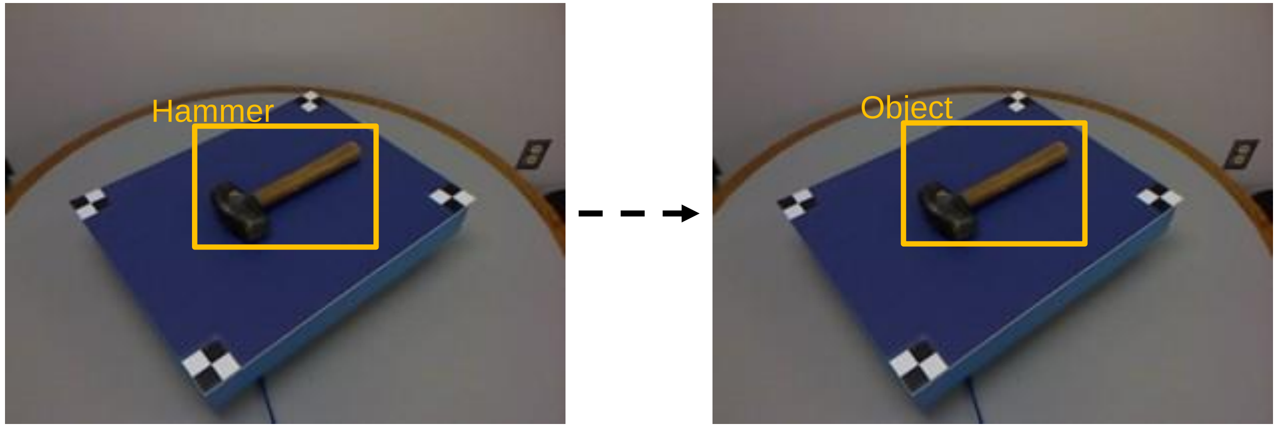


Contributions

- Overcome framing issues

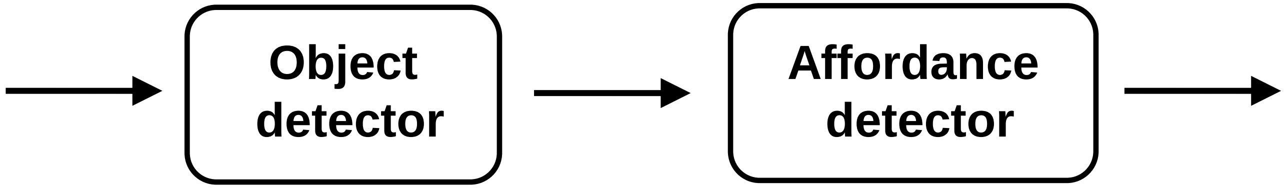


- Leverage human-in-the-loop feature: the object class is known by the human

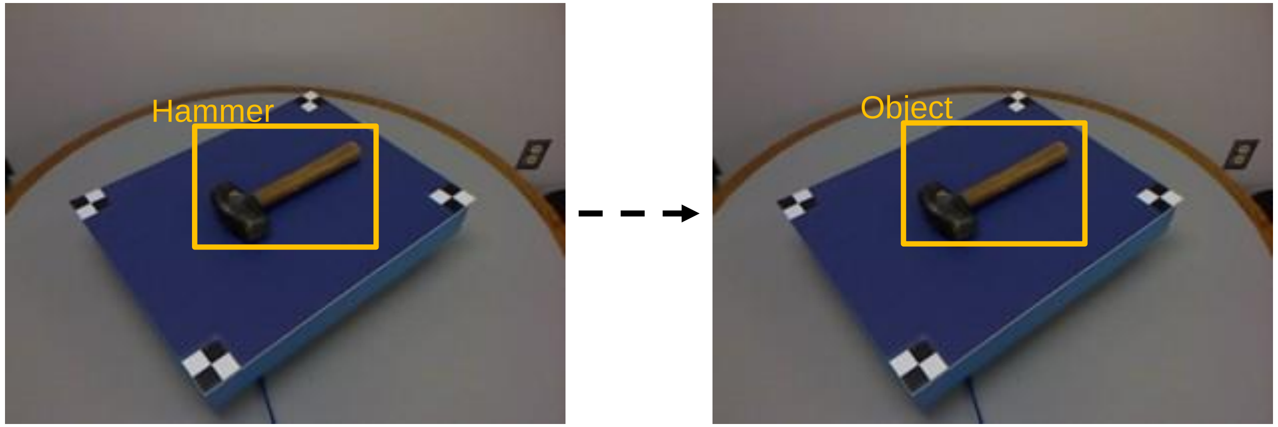


Contributions

- Overcome framing issues

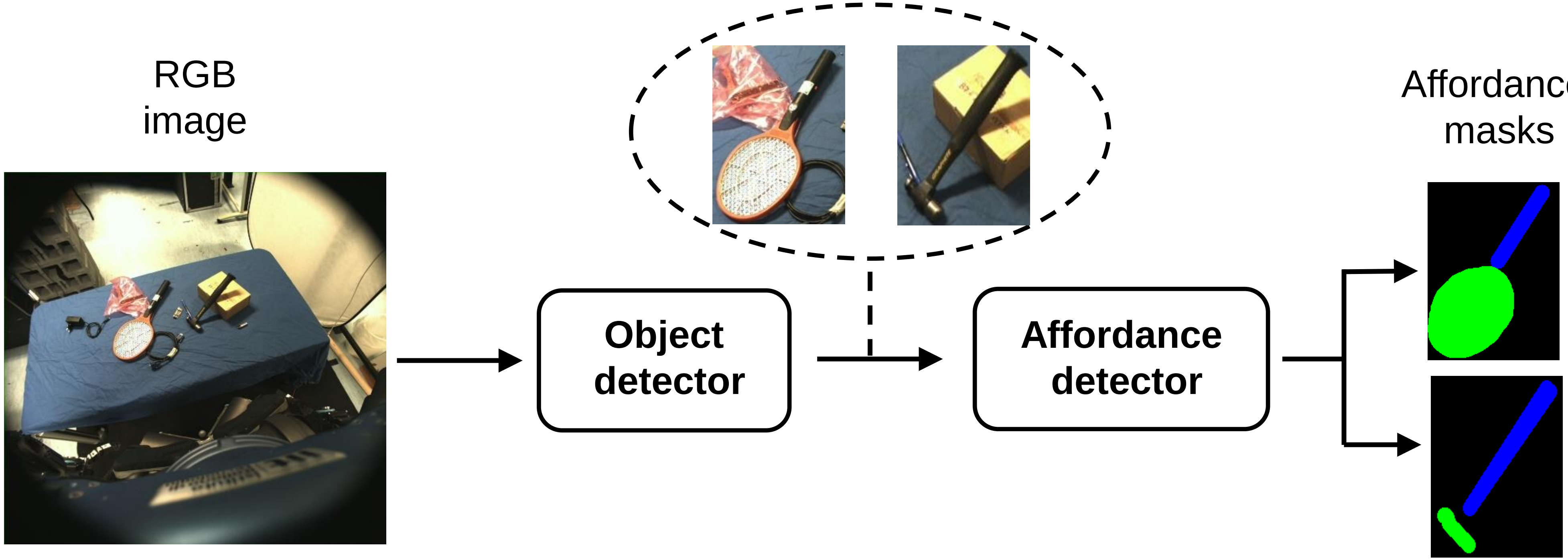


- Leverage human-in-the-loop feature: the object class is known by the human



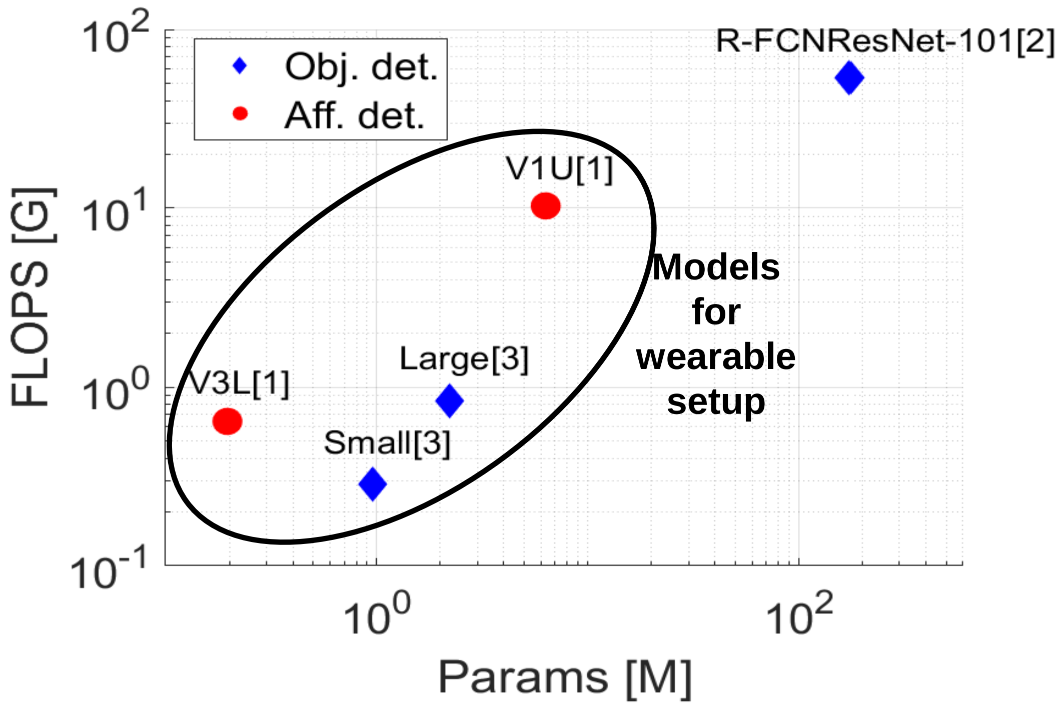
- Target resource-constrained devices employing lightweight models

Proposed Method



Considered models

Comparison: wearable vs robotic



[1] Ragusa, E. et al., *“Hardware-Aware Affordance Detection for Application in Portable Embedded Systems”,* IEEE Access, 2021.

[2] Nguyen, A. et al., *“Object-Based Affordances Detection With Convolutional Neural Networks and Dense Conditional Random Fields”,* IEEE /RSJ IROS, 2017.

[3] Howard, A. et al., *“Searching for MobileNetV3”,* IEEE/CVF ICCV, 2019.

Datasets

UMD [4]

- Simple setting
- Same resolution
- Blue rotating support
- Single object per image

IIT-AFF [2]

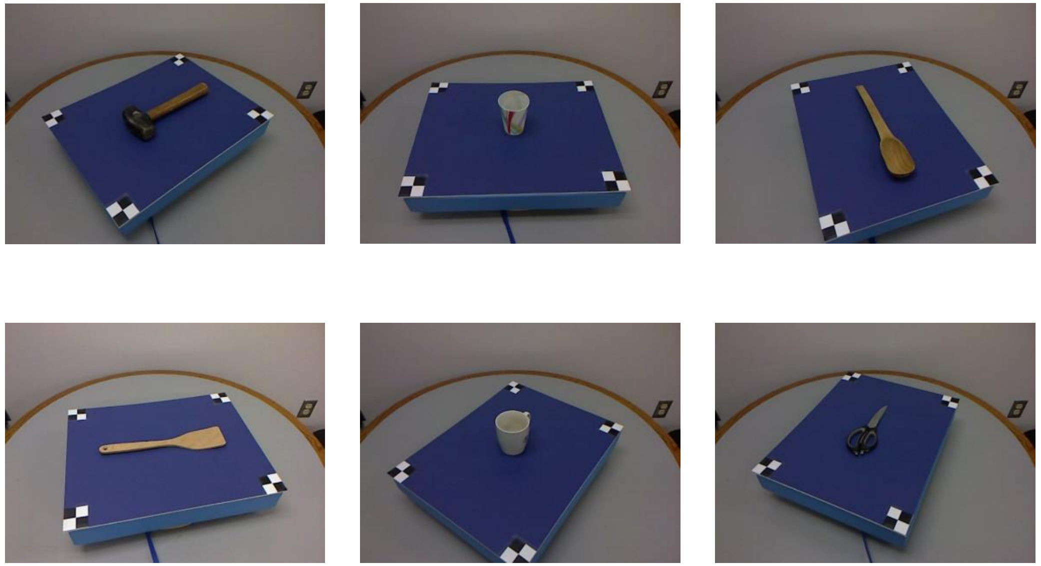
- Challenging setting
- Different resolutions
- Different supports and scenes
- Multiple objects per image

[2] Nguyen, A. et al., *“Object-Based Affordances Detection With Convolutional Neural Networks and Dense Conditional Random Fields”,* IEEE/RSJ IROS, 2017.

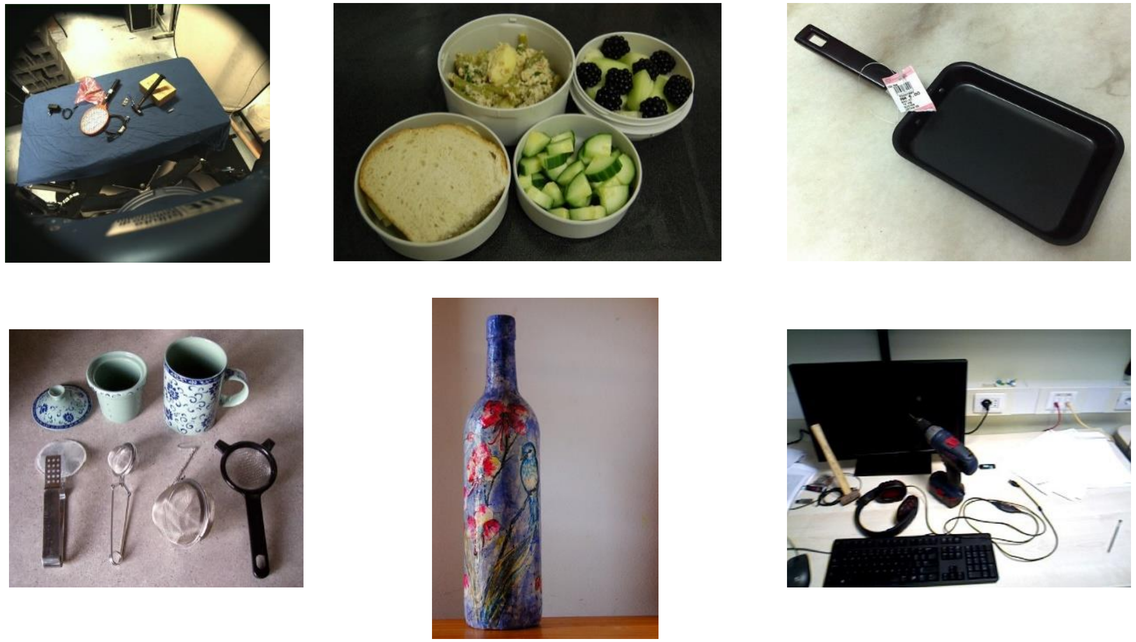
[4] Myers, A. et al., *“Affordance Detection of Tool Parts from Geometric Features”,* IEEE ICRA, 2015.

Datasets

UMD [4]



IIT-AFF [2]

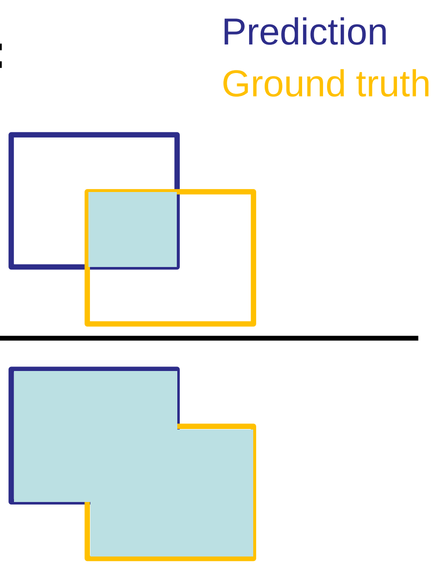


[2] Nguyen, A. et al., *“Object-Based Affordances Detection With Convolutional Neural Networks and Dense Conditional Random Fields”,* IEEE/RSJ IROS, 2017.

[4] Myers, A. et al., *“Affordance Detection of Tool Parts from Geometric Features”,* IEEE ICRA, 2015.

Metrics

Intersection over Union (IoU):



$$\text{IoU} = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

Mean Average Precision (mAP):

- Area under Precision-Recall curve

F_β^w score [5]:

$$F_\beta^w = (1+\beta^2)\frac{Precision^w \cdot Recall^w}{\beta^2 \cdot Precision^w + Recall^w}$$

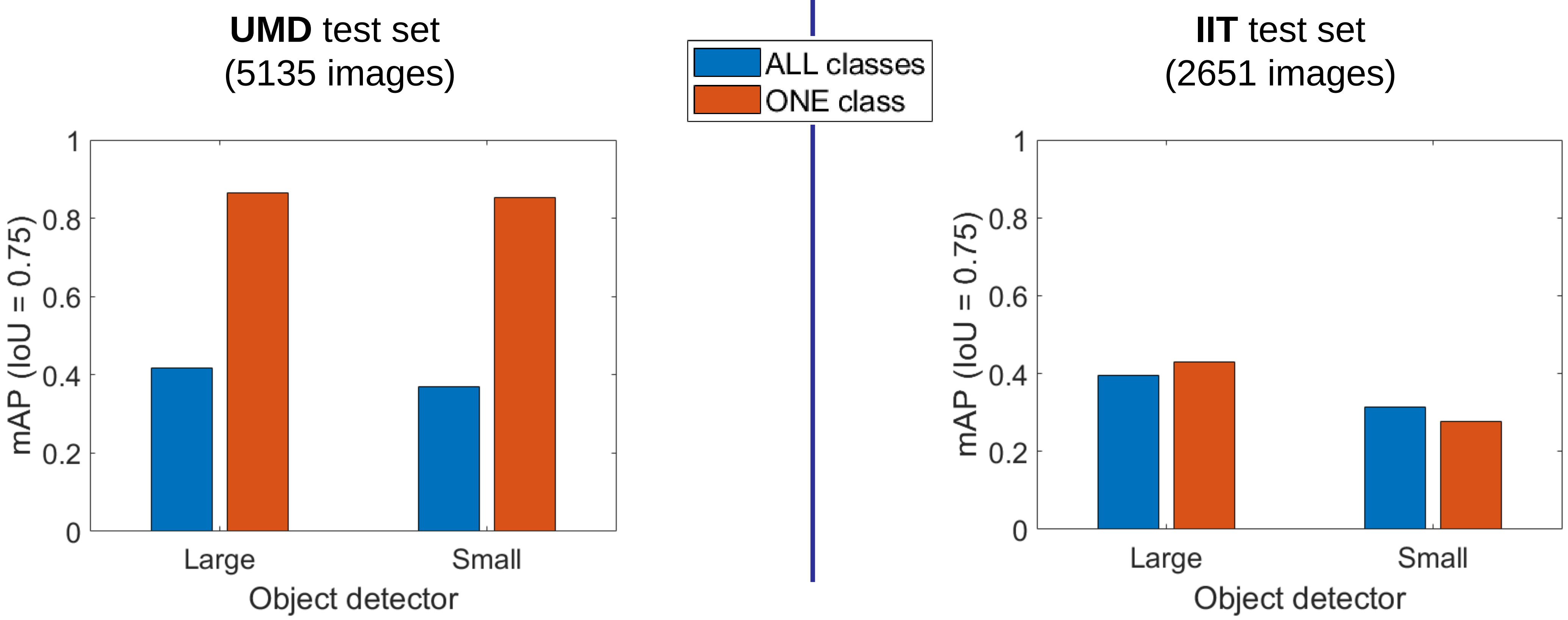
w:

- weights the dependency between ground truth pixels
- weights the errors based on distance with respect to ground truth

[5] Margolin, R. et al., *"How to evaluate foreground maps,"* IEEE Conference on Computer Vision and Pattern Recognition, 2014.

Object detection results

Detection improvement leveraging human-in-the-loop (ONE class)

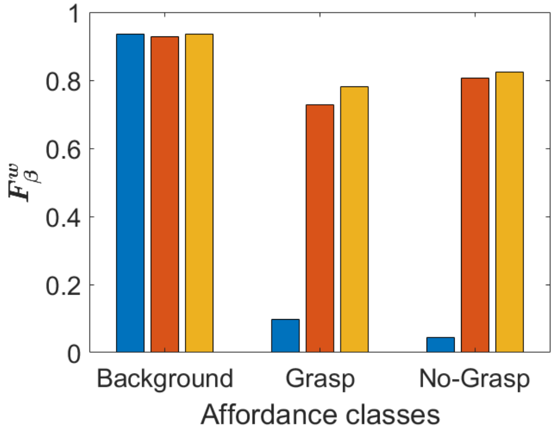


Affordance detection results

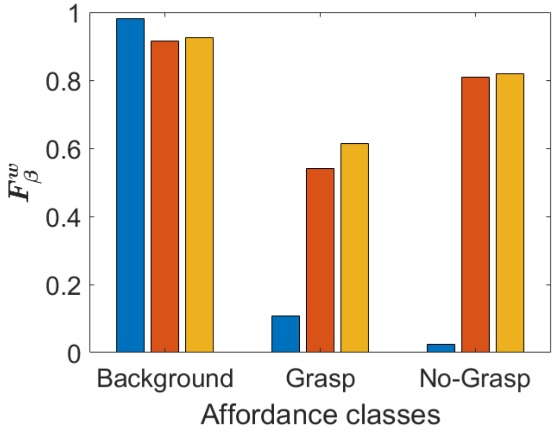
Improvement employing proposed method (Obj. det. + Aff. det.) vs baseline (only Aff. det.)

UMD test set

V1U

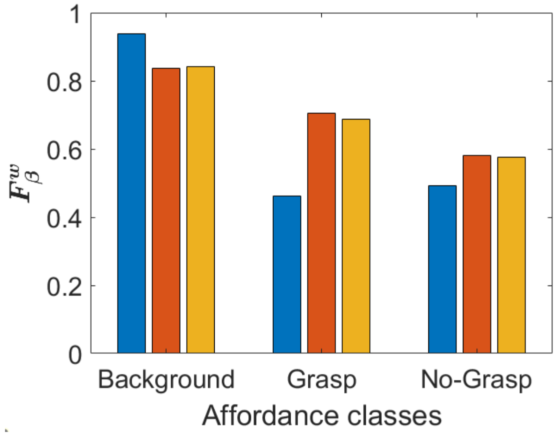


V3L

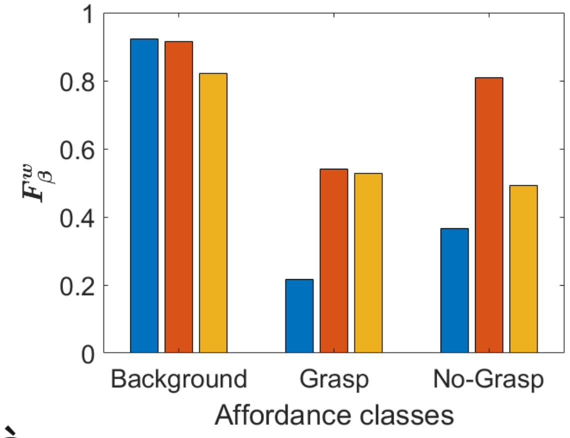


IIT test set

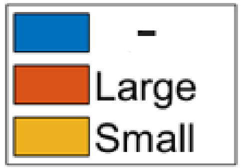
V1U



V3L



Obj. Det.

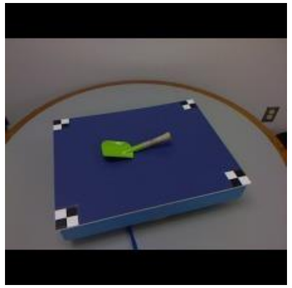
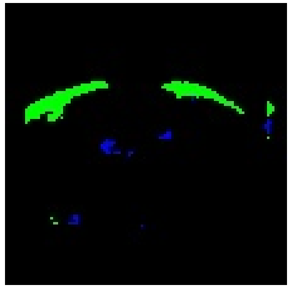
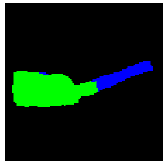
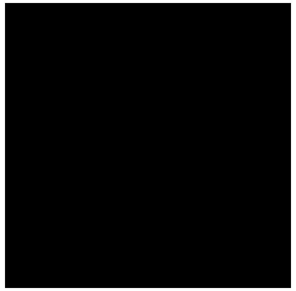
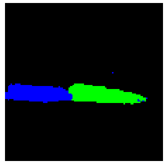
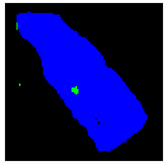
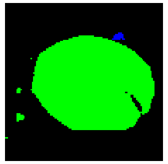


centre for intelligent sensing

Università di Genova

Queen Mary University of London

Qualitative results

Image	Baseline	Proposed method	
	Aff. Det.	Obj. det.	Aff. det.
			
		  	  

Conclusions

- Pipeline to overcome framing issue
- Object detection improvement leveraging human-in-the-loop
- Affordance detection improvement
- Target resource-constrained devices employing lightweight models

tommaso.apicella@edu.unige.it
t.apicella@qmul.ac.uk

https://github.com/SEAlab-unige/ICECS-2021

CIS centre for intelligent sensing
Università di Genova
Queen Mary University of London