2019 Intelligent Sensing Summer School (London) September, 2-6

The CORSMAL challenge

Team 3

Deep Robot with Mask R-CNN (Convolutional Neural Network) Object Detection

Asaduddin Bilal Mohammed

Zico Pratama Putra

Centre for Intelligent Sensing

Queen Mary University of London

CORSMAL

Collaborative object recognition, shared manipulation and learning

Outline

- Motivation
- Related Work
- Framework Overview
- The need for real training data
- Machine Learning
- Evaluation
- Conclusion

MOTIVATION

Objectives

- **Explore machine learning libraries**
- **Experiment iteration**
- **Communicate with real-world robot**

Challenge

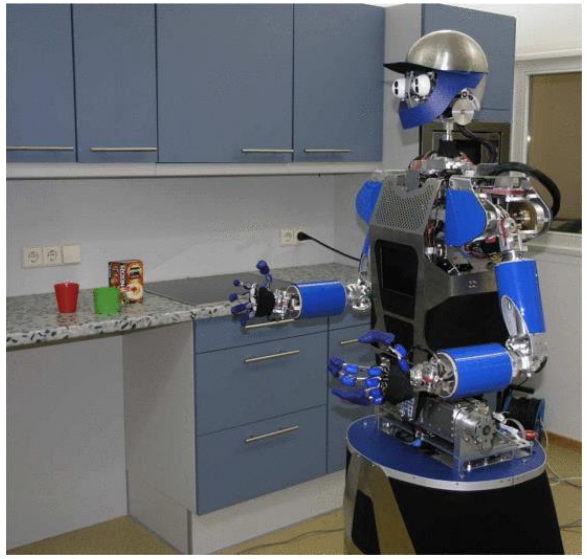
- The Robot had to be able to sense and grip uncertainties and object occlusions
- Constraining each object's design and sensor choices, included physical properties and grip position

Related Work

Humanoid Grasping

Humanoid Robot HRP3
(Kaneko et al., 2008)

Boston Dynamic's Atlas
(Boston Dynamics, 2018)



ARMAR-III (Asfour et al., 2006),

CORSMAL
Collaborative object recognition,
shared manipulation and learning

RBG Based Object Tracking

Open Pose (Cao, 2017)

- Multiple people
- Great Accuracy
- General Purpose

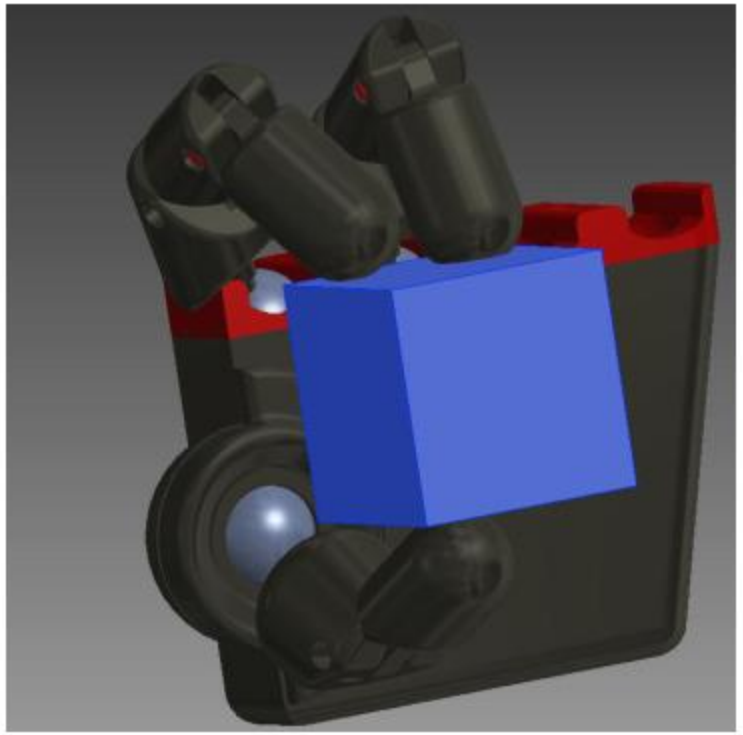
RBG Based Object Tracking

Dense Pose (Guller et al. 2018)

- Surface representation
- 5 millions manually annotated poitns

Deep Learning for Object Detection

Multi-fingered robotic hand for grasping tasks (Bezak, 2014)



Robot handling using Active Vision

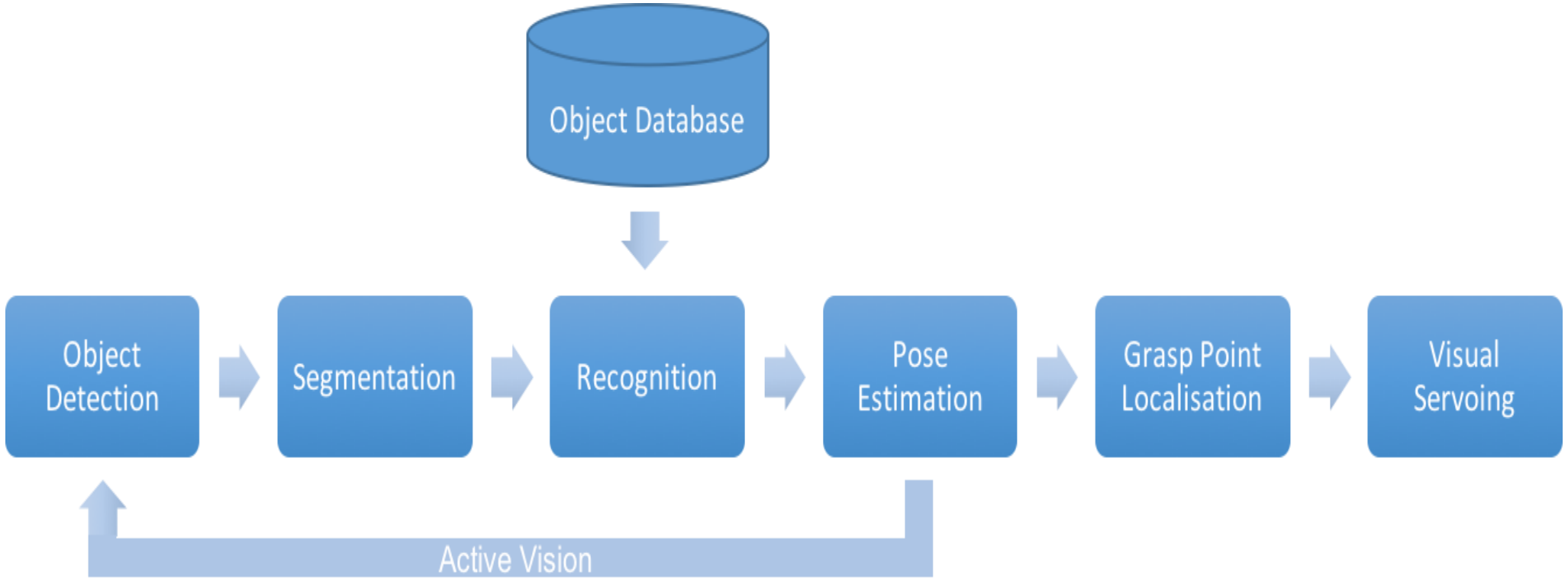


FIGURE 1: Traditional active vision pipeline for object manipulation.

Framework Overview

Sensor Camera Solution

Intel RealSense D435i

- **Depth Sensor**
- **RBG sensor**
- **Infrared projector**

What Depth Camera Data Looks Like

Pixel has 4 values (Red, Green, Blue, Depth)

Color image from a color sensor

Each image has RGB Value associated with it

Depth image from a depth sensor

Each pixel Depth (distance from camera) value

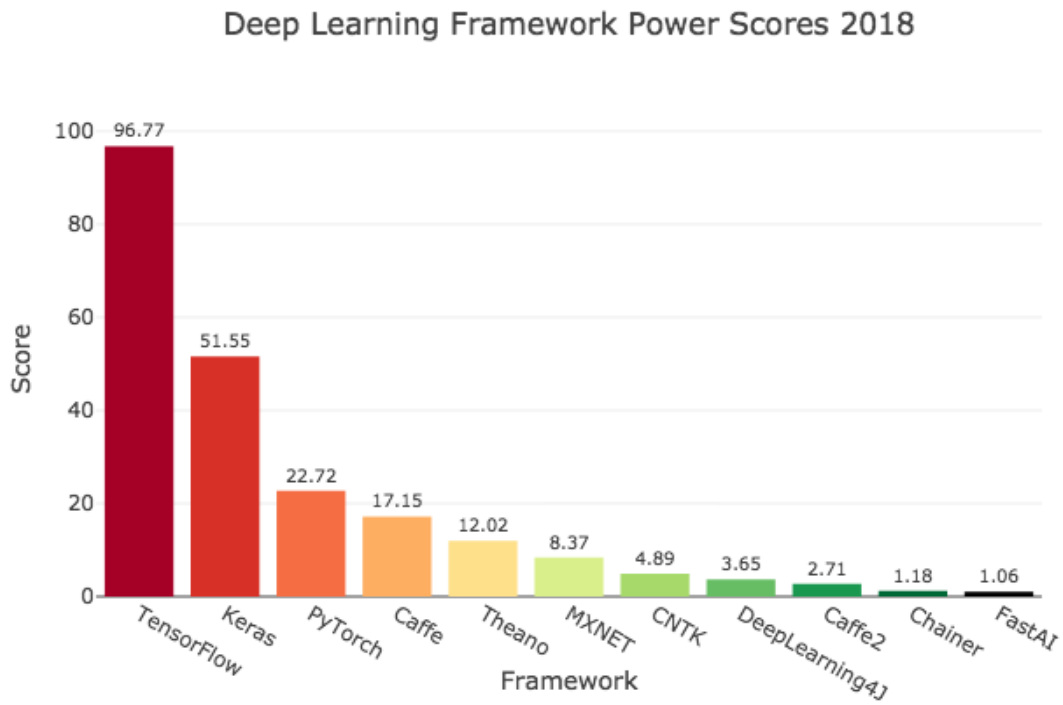
Implementation

Deep Learning Libraries

- **TensorFlow**

TensorFlow is an interface for expressing machine learning algorithms

- Google Search
- Google Photos
- Street View and Google Maps
- Google Translate



Ref: https://towardsdatascience.com/deep-learning-framework-power-scores-2018-23607ddf297a

CORSMAL
Collaborative object recognition, shared manipulation and learning

Deep Learning Libraries & Programming Language

- **OpenCV**
- **COCO Trained Model Dataset (mask_rcnn_inception_v2_coco) (Lin, 2015)**

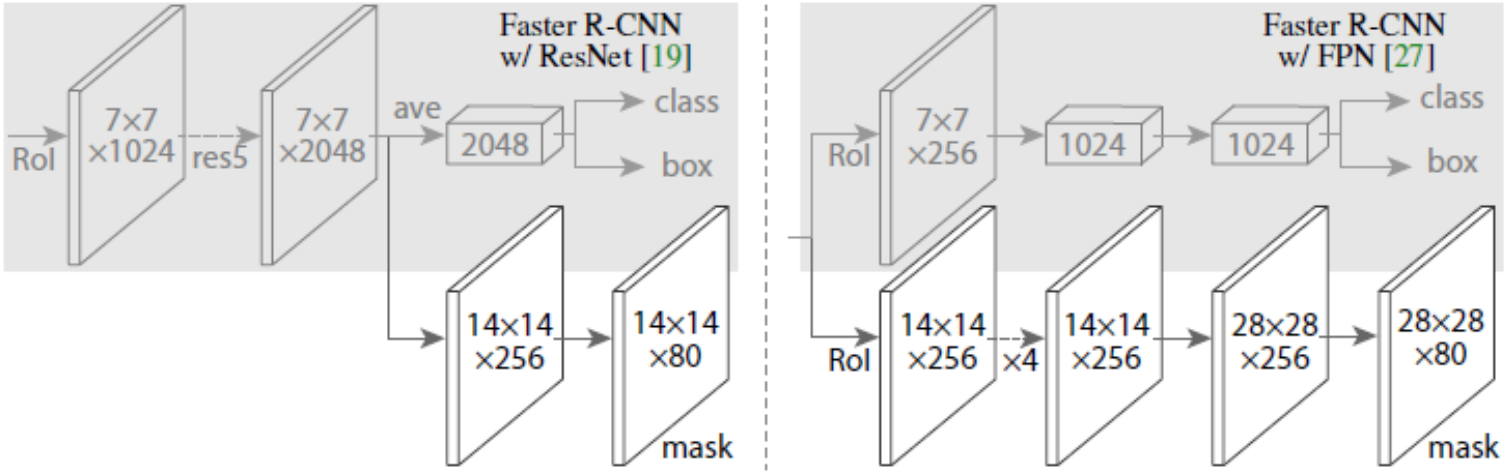
Mask API provides segmentation masks for every object instance

- **Python**

Ref: https://towardsdatascience.com/deep-learning-framework-power-scores-2018-23607ddf297a

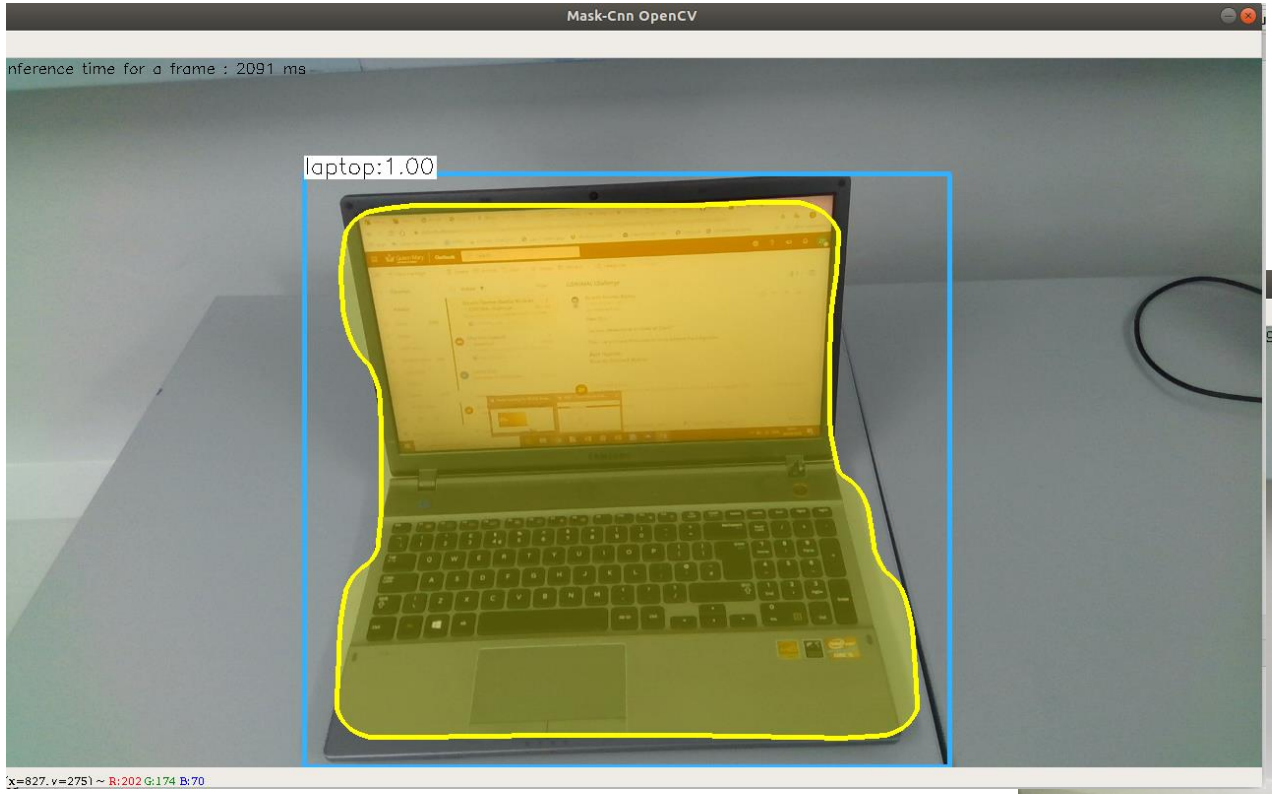
Object Instance Segmentation

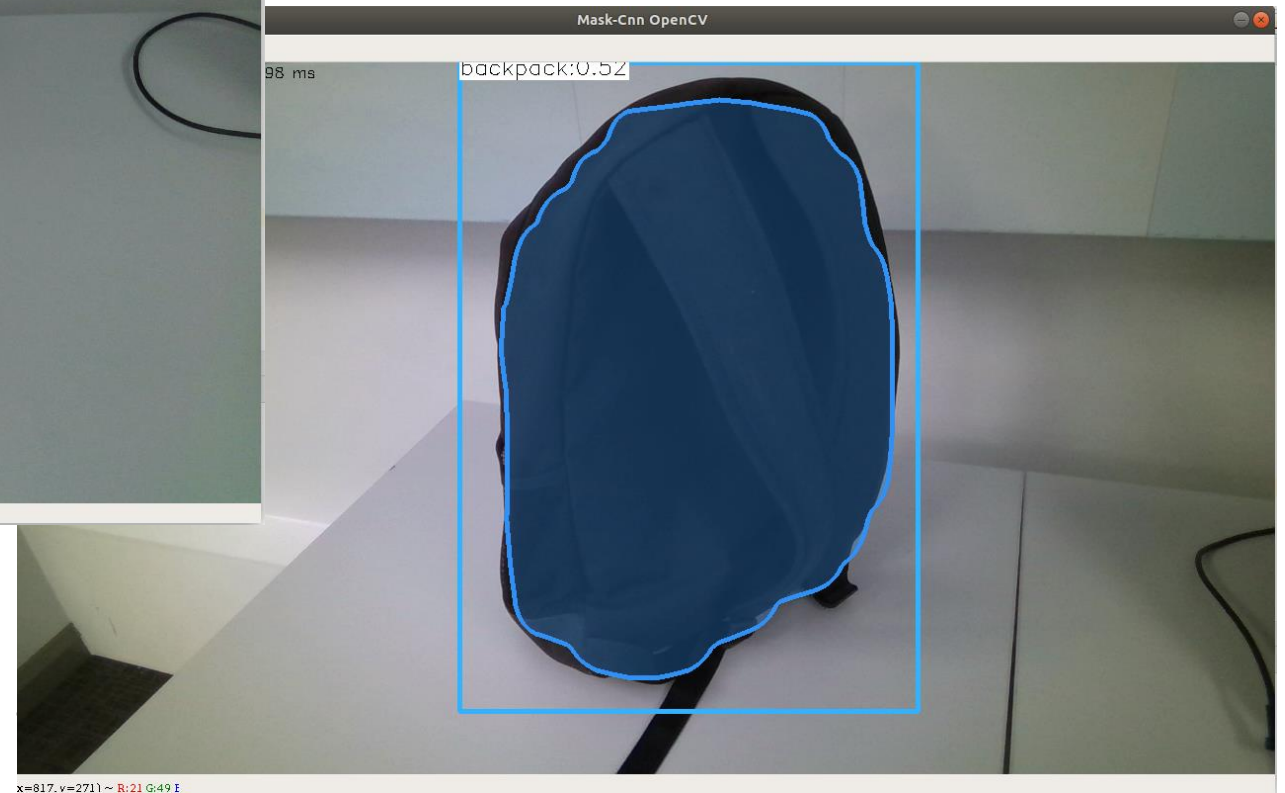
- **Mask R-CNN (Kaiming He et al, 2018)**

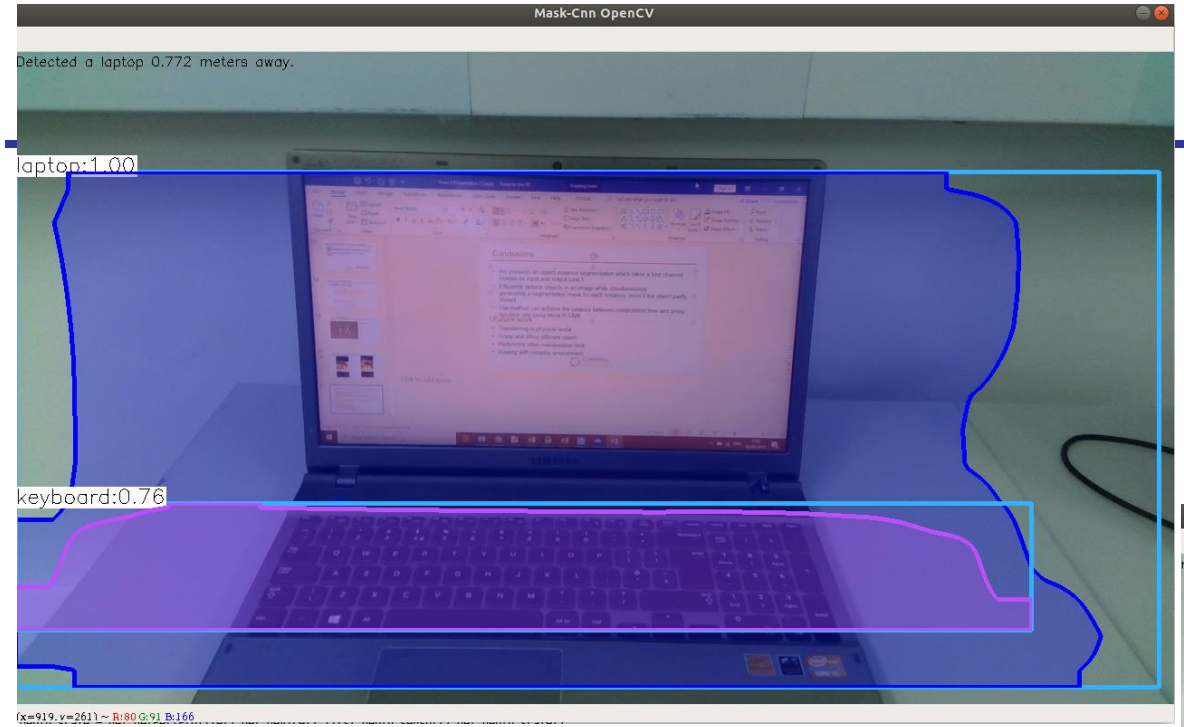


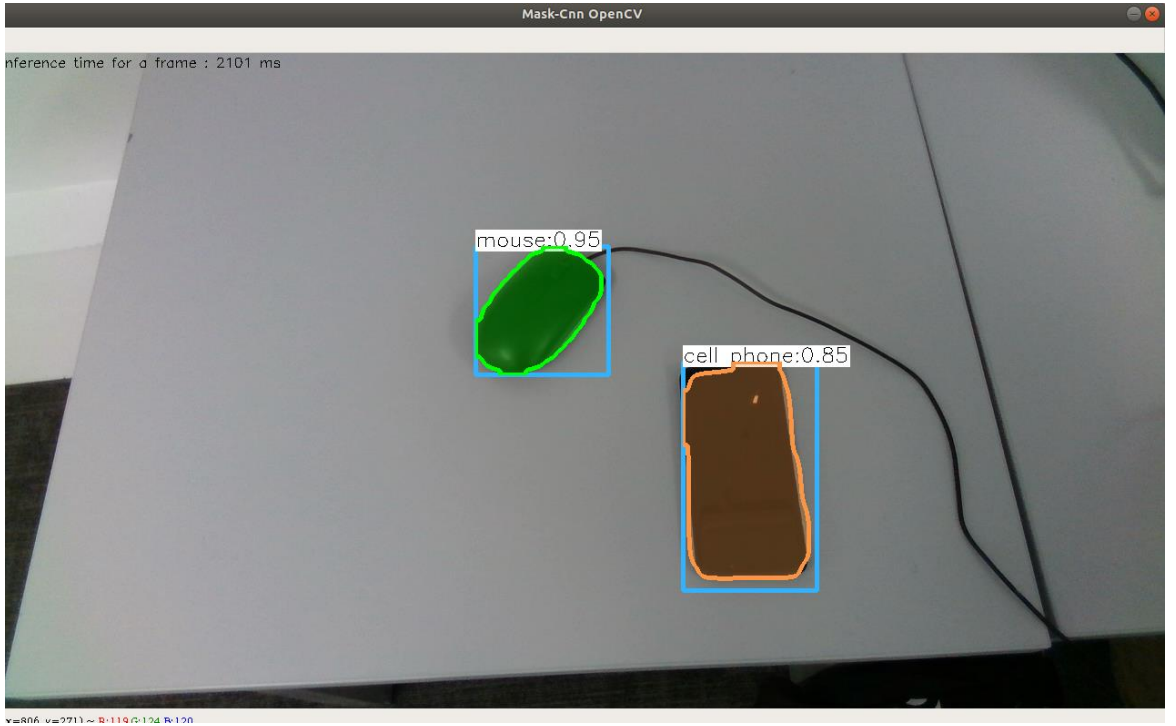
Mask R-CNN Architecture (Kaiming He et al, 2018)

Evaluation & Result





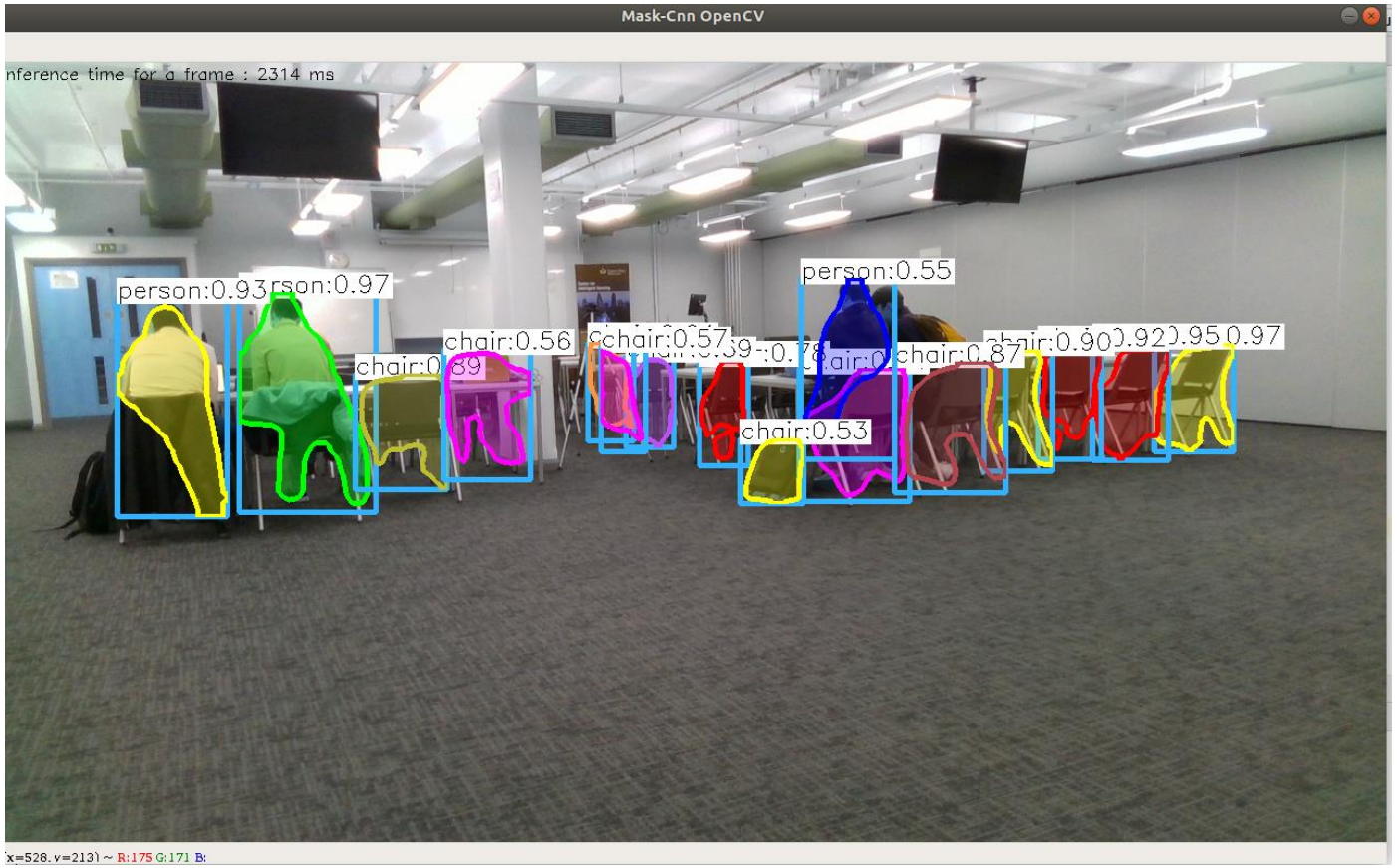




CORSMAL

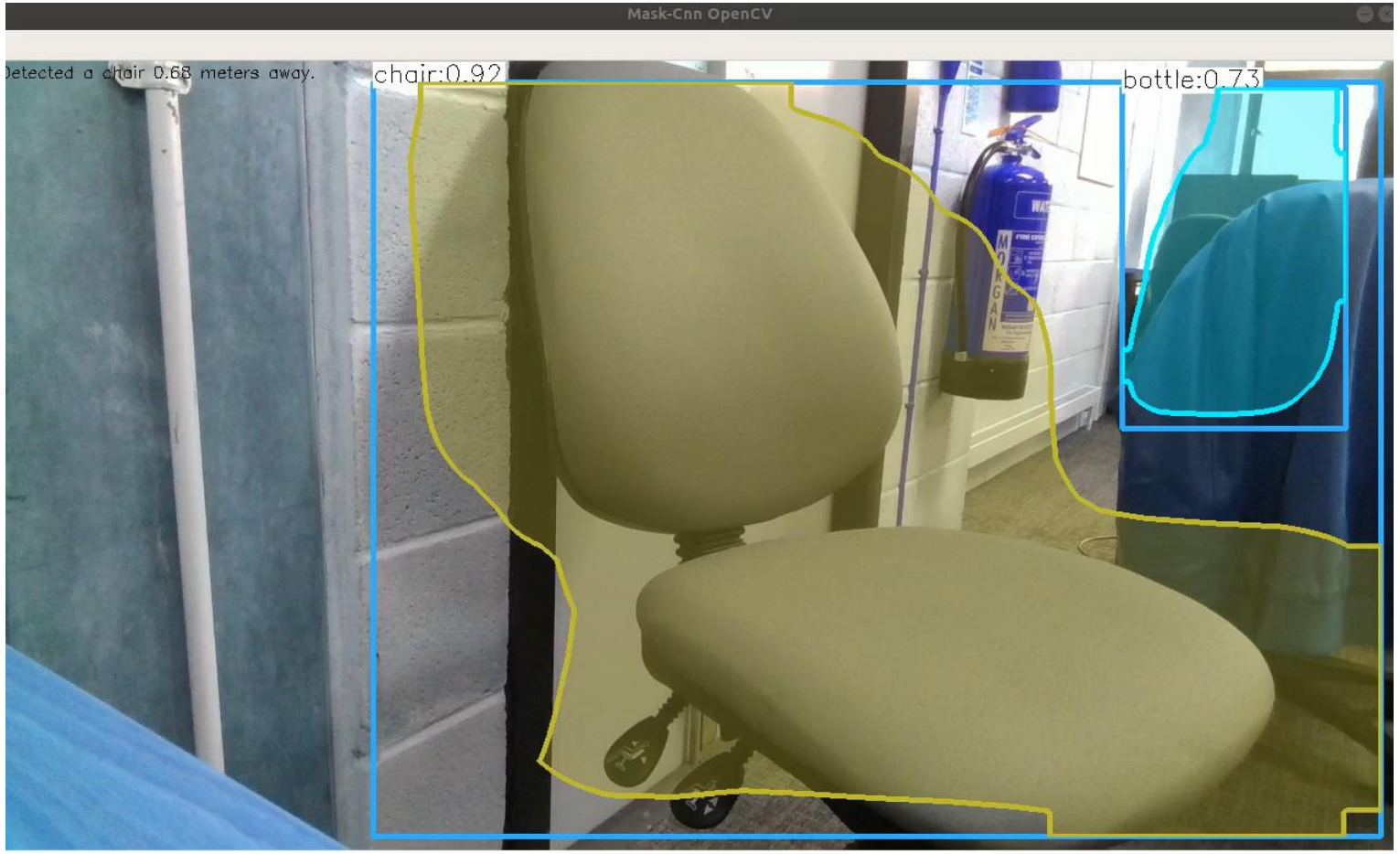
Collaborative object recognition,
shared manipulation and learning

Detecting multiple images



CORSMAL
Collaborative object recognition,
shared manipulation and learning

Distance Accuracy



Conclusions

- We presents an object instance segmentation which takes a 4 channel images as input and output Line 1
- Efficiently detects objects and measure the distance while simultaneously generating a segmentation mask for each instance, even if the object partly shown
- Our method can achieve the balance between computation time and grasp success rate using Mask R-CNN

Future work

- Transferring to physical world
- Grasp and lifting different object
- Performing other manipulation task
- Dealing with complex environment

CORSMAL
Collaborative object recognition, shared manipulation and learning