



UNIVERSITÀ DEGLI STUDI
DI GENOVA

Object Detection with ZED Mini on Unity

Aurora Bertino, Sara Romano, Chiara Saporetti

Index

01 Introduction and Motivation

02 Hardware and Software

03 Methods

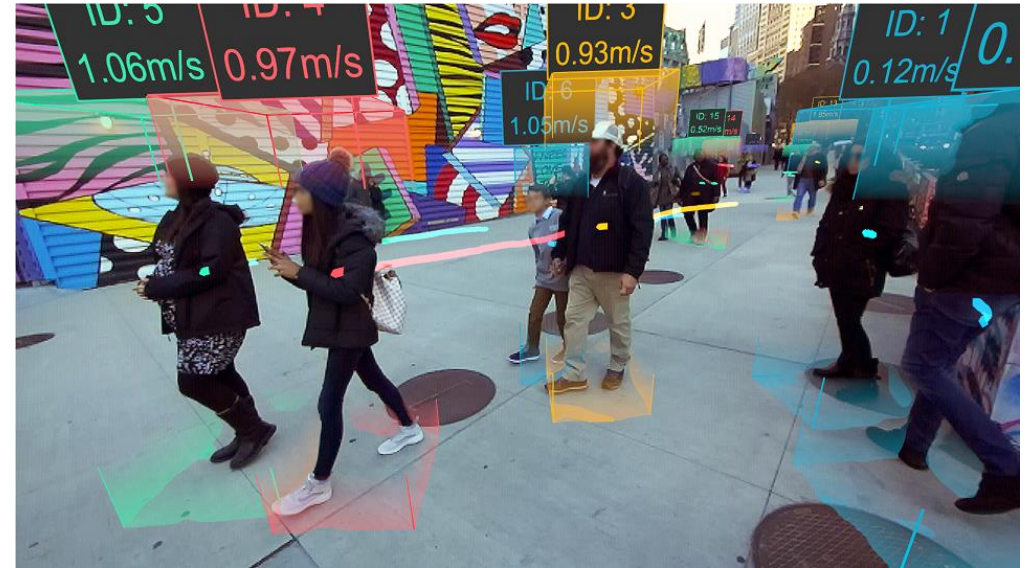
04 Results and Future Work



Introduction and motivation

Introduction

- In recent times, object detection and pose estimation have gained significant attention in the context of robotic vision applications.
- Computer-Vision is a fast growing application field in robotics:
 - camera sensors
 - computational power
 - machine learning
- We propose a system that performs object detection in Unity with the ZED Mini camera.



Motivation

- Cameras and robots are technologically "ready" to do object detection, but algorithms to do so are not always present in the SDKs.
- Stereolabs provides object detection only for the ZED 2 Camera




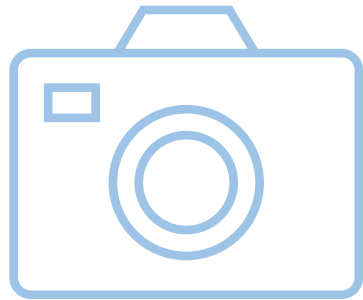
Problem

- Make possible to perform object detection in the Unity ecosystem using ZED Mini camera



Considered approaches

- Unity + ZED 2 API
- Tensorflow + ZED
- Unity + Vuforia
- **Unity + Barracuda** 



Hardware and Software

Hardware

- ZED Mini Camera



Video Output

Output Resolution	Side by Side 2x (2208x1242) @15fps 2x (1920x1080) @30fps 2x (1280x720) @60fps 2x (672x376) @100fps
-------------------	---

Output Format	YUV 4:2:2
---------------	-----------

Field of View	Max. 90° (H) x 60° (V) x 100° (D)
---------------	-----------------------------------

RGB Sensor Type	1/3" 4MP CMOS
-----------------	---------------

Active Array Size	2688x1520 pixels per sensor (4MP)
-------------------	-----------------------------------

Focal Length	2.8mm (0.11") - f/2.0
--------------	-----------------------

Shutter	Electronic synchronized rolling shutter
---------	---

Interface	USB 3.0 Type-C port
-----------	---------------------

Depth Sensing

Baseline	63 mm (2.4")
----------	--------------

Depth Range	0.10 m to 15 m (0.3 to 49 ft)
-------------	-------------------------------

Depth Map Resolution	Native video resolution (in Ultra mode)
----------------------	---

Depth Accuracy	< 1.5% up to 3m < 7% up to 15m
----------------	-----------------------------------

Software

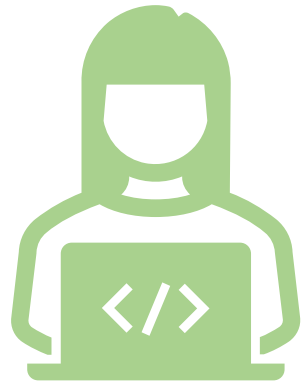
- ZED APIs
- Unity (on Windows with C#)
- Barracuda package
- Zed-Unity plugin
- Tiny Yolo v2

System requirements

- Win 10, Win 8, Win 7
- Ubuntu 18.0/16.04

SDK Requirements

- Dual-core 2.3GHz or faster
- Minimum 4GB RAM
- Nvidia GPU (1) Compute capability ≥ 3.0



Methods



ZED M

ZED

SDK, Unity plugin



Video
fetching

Image pre-
processing



With
Barracuda Package

Image post-
processing

Depth
Sensing

Visualization

Fetching ZED video stream in Unity

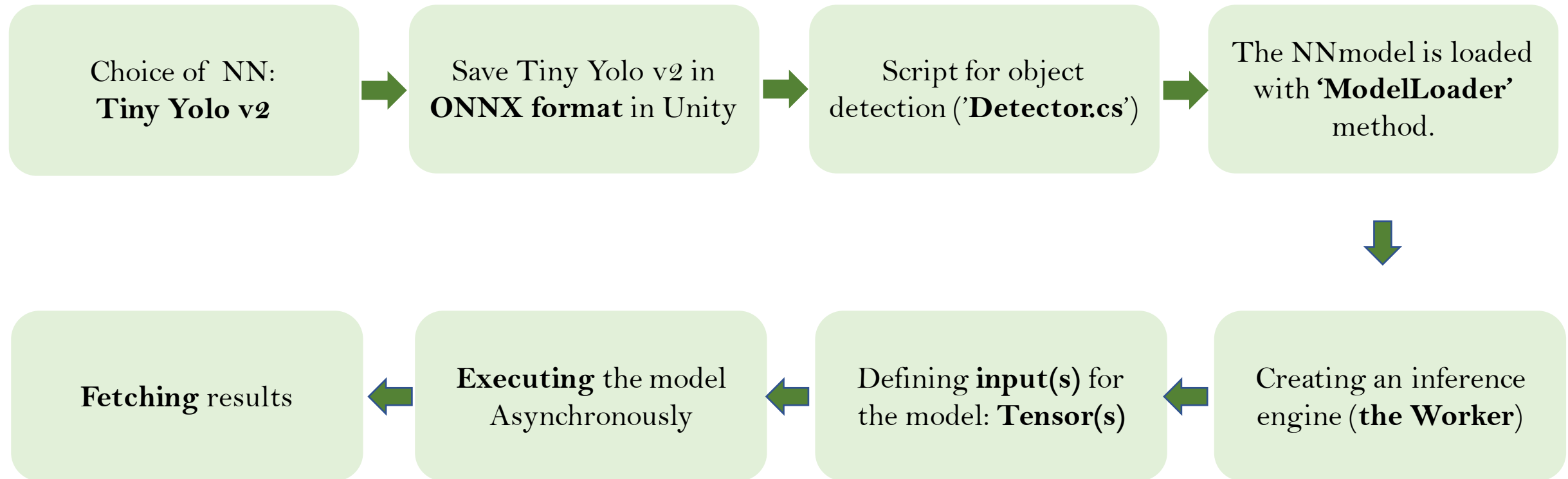
In Unity GUI:

- Import the ZED Camera from the **ZED SDK**

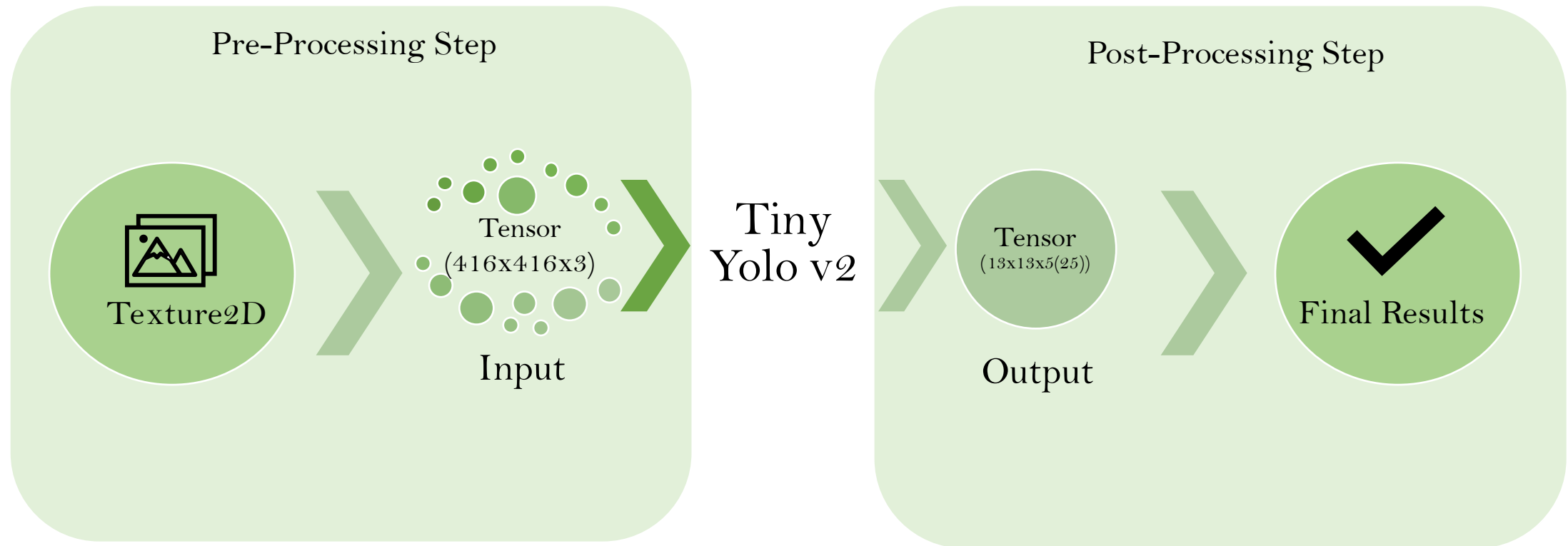
In the Script (CameraData.cs):

- Retrieve the camera frames on the CPU in **ZedMat** format.
- Marshal copy the result in a matpointer.
- Define a **Texture2D** variable to be used in “LoadRawTextureData”.

Importing a NN in Unity using Barracuda



Detector



Pre-processing

CameraData.cs



Texture2D
(2560x720)



CROP & SQUARE



Texture2D
(720x720)



SCALE &
ROTATE(0°)



Texture2D
(416x416)

Detector.cs

Creating a Tensor (416 x 416 x 3):
Height x Width x [r,g,b].



Tiny yolo v2

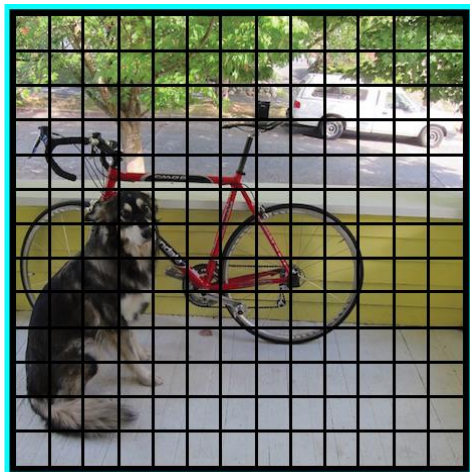
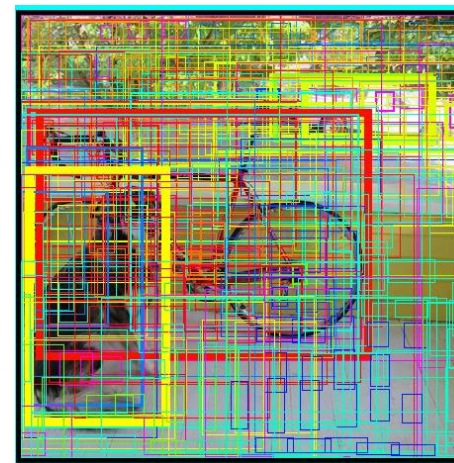
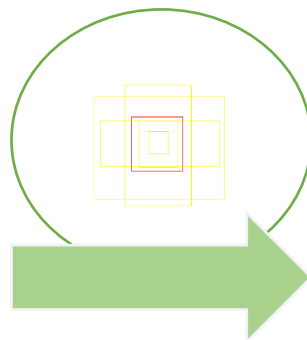


Image divided into
13 x 13 grid cells

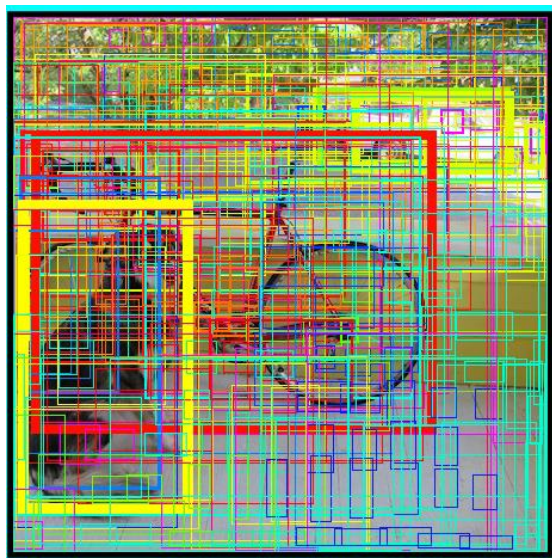


Each cell predicts
5 bounding boxes.
Each of them is described
by 25 data elements.

Each Bounding box has:

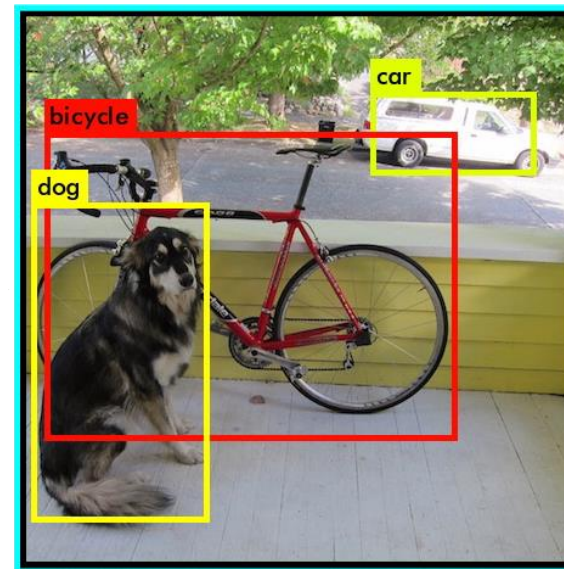
- (x, y, h, w) ;
- Confidence score;
- The probabilities of the classes (which are 20).

Post-Processing



BOXES
FILTERED
ACCORDING TO
THEIR
LOCALIZATION

BOX FILTERED
ACCORDING TO
THEIR
CLASSIFICATION



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

Depth sensing with ZED APIs

Retrieve the depth information from the ZED, exploiting the Stereo Camera. With APIs:

- **Depth map:** store a distance value (Z) for each pixel (X, Y).
- **Point cloud:** stores its data on 4 channels (X, Y, Z, color).
- **Retrieval:** similar to camera image (ZedMat MAT_32F_C4 for point cloud, MAT_32F_C1 for depth)





ZED M

ZED

SDK, Unity plugin



Video
fetching

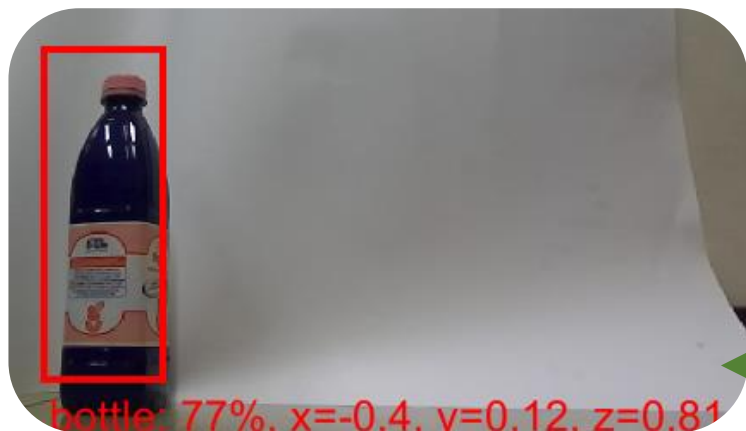
Image pre-
processing



With
Barracuda Package

Depth
Sensing

Image post-
processing



Final remark

- Stereolabs provides many applications but we had to develop new code from scratch.
- Scripts for our purpose do not exist, so the whole initial work was to study the single ZED Manager functions and their usage. Examples:
 - `CreateTextureImageType()` vs `RetrieveImage()`
 - `ZEDMat()`

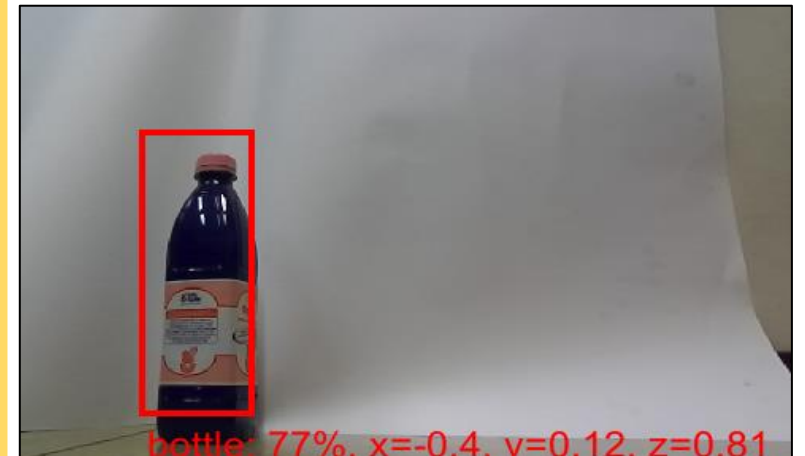
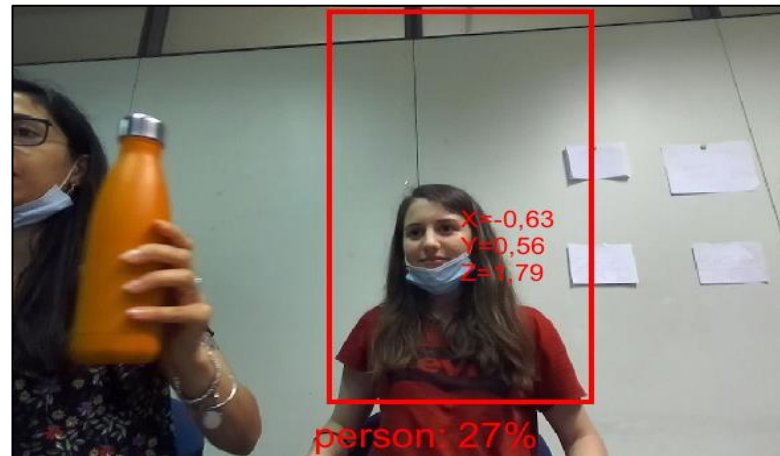
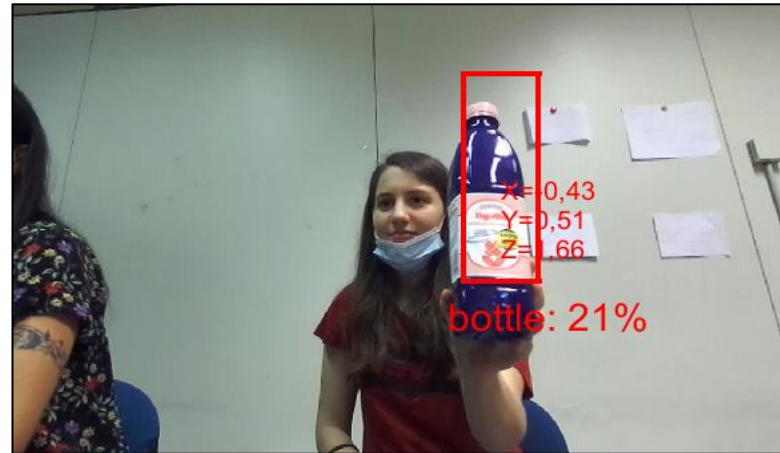


Results and future work

Results

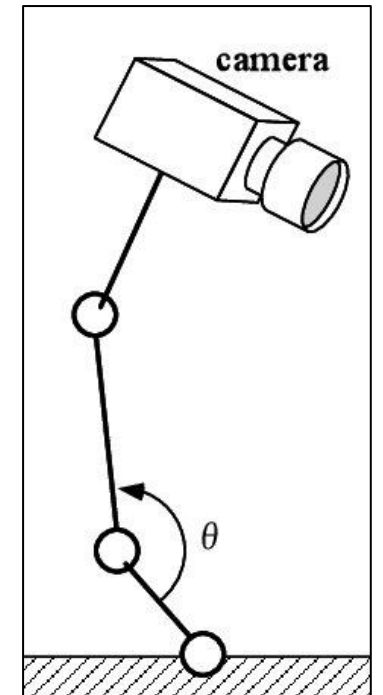
Tested on:

- recorded videos in SVO format (30fps) and on the real camera (60fps).
- objects present in the laboratory and ourselves.



Future work

- **First step** of a bigger project!
- Add **other Neural Networks**, (in ONNX or custom) with possibility to choose between them.
- Build an assembling environment:
 - For **humans**: attach the AR/VR headset (maybe on the GPU for higher fps).
 - For a **robotic** manipulator



Conclusions

- **Object detection on Unity with the ZED Mini:**
 - Video stream fetching (ZED → Texture2D).
 - Pre-processing of the image (Texture2D → Tensor).
 - Post processing of the Neural network output (Tensor → BB for the GUI).
- **Good results** at 30 fps.
- **Adaptability** of the project to other neural networks
- Adds **new functionality** to ZED Mini (APIs only for ZED2).
- Possibility to add **AR** / **VR** headset.



ZED



Thank you for your attention!