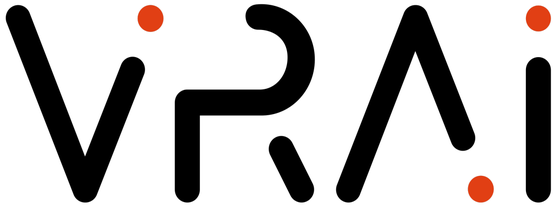
Staff detection with marker from a top-view camera

Iezzi Christian - Incicco Emanuele

UNIVERSITÀ
POLITECNICA
DELLE MARCHE



OUTLINE

- Why this project?
- ArUco
- Dataset construction
- State-of-the-art Neural Networks
  - YOLOv3
  - YOLOv5
- Performance
- Future developments

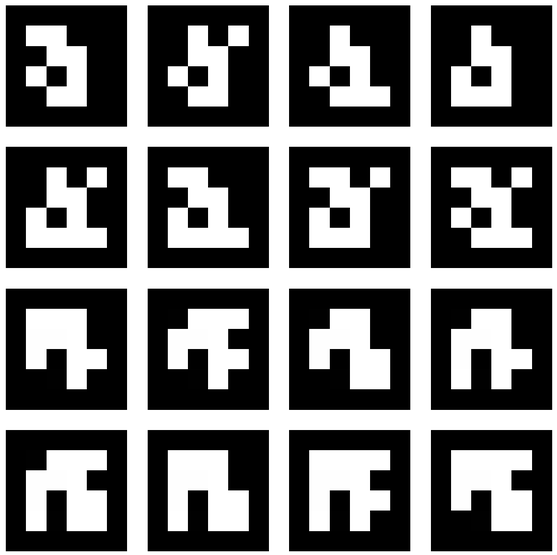
WHY THIS PROJECT?

The project was born from the idea of wanting to recognize the staff of a certain sale point from a top-view camera.

We therefore want, first of all, to obtain a dataset that simulates this situation, on which we can then train state-of-the-art neural networks for the object detection task.

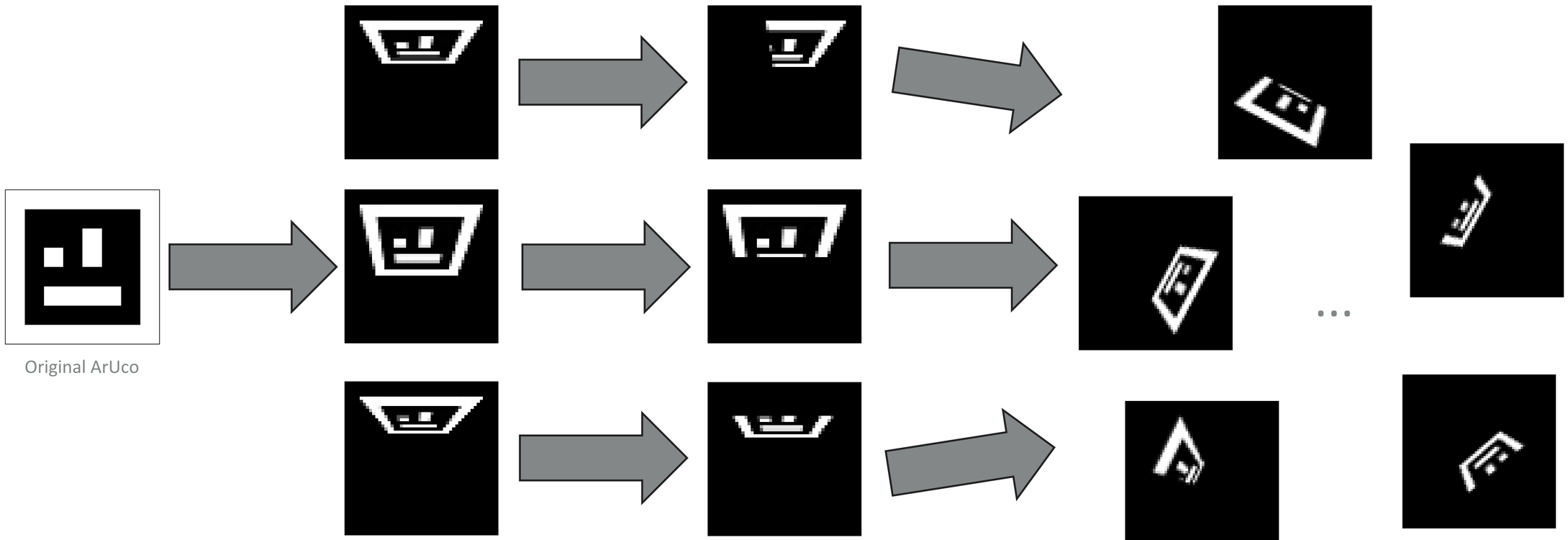
ArUco

An ArUco marker is a square marker composed by a black border and an inner binary matrix which determines its identifier.



Some ArUco markers

DATASET



Original ArUco

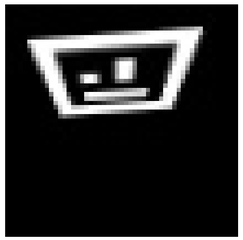
Some perspective transformations

Some random edge occlusions

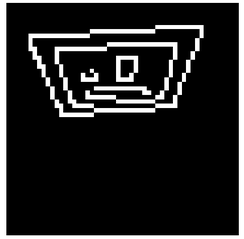
Some rotation and flip tranformations

DATASET

For each ArUco transformation



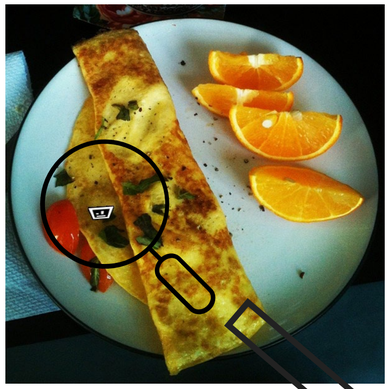
Canny Edge Detection



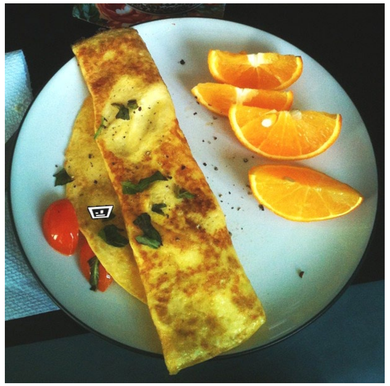
ArUco mask

COCO dataset

Choose random images from COCO dataset



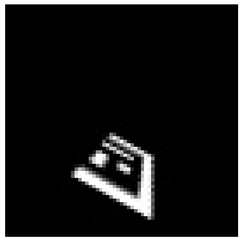
Make the overlay



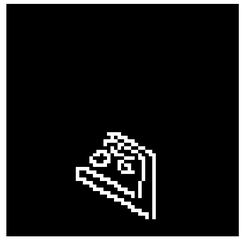
Randomly make a brightness transformation

DATASET (Occluded ArUco)

For each ArUco transformation



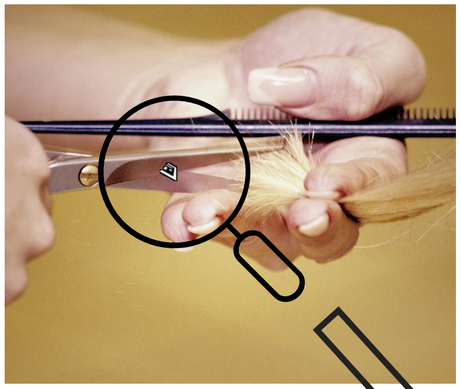
Canny Edge Detection



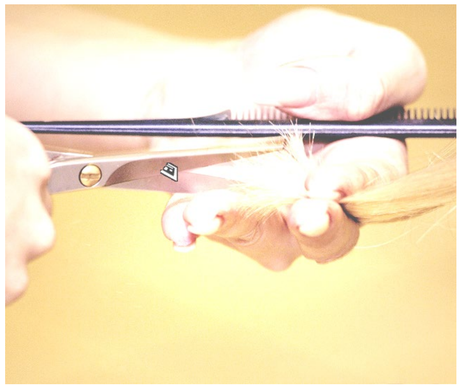
ArUco mask

COCO dataset

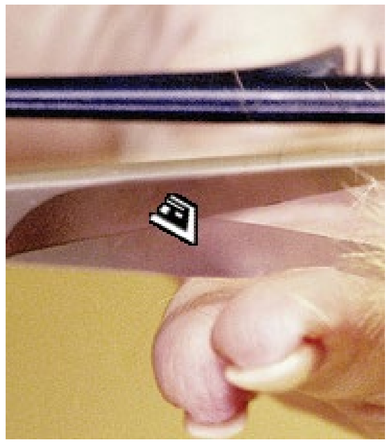
Choose random images from COCO dataset



Make the overlay



Randomly make a brightness transformation



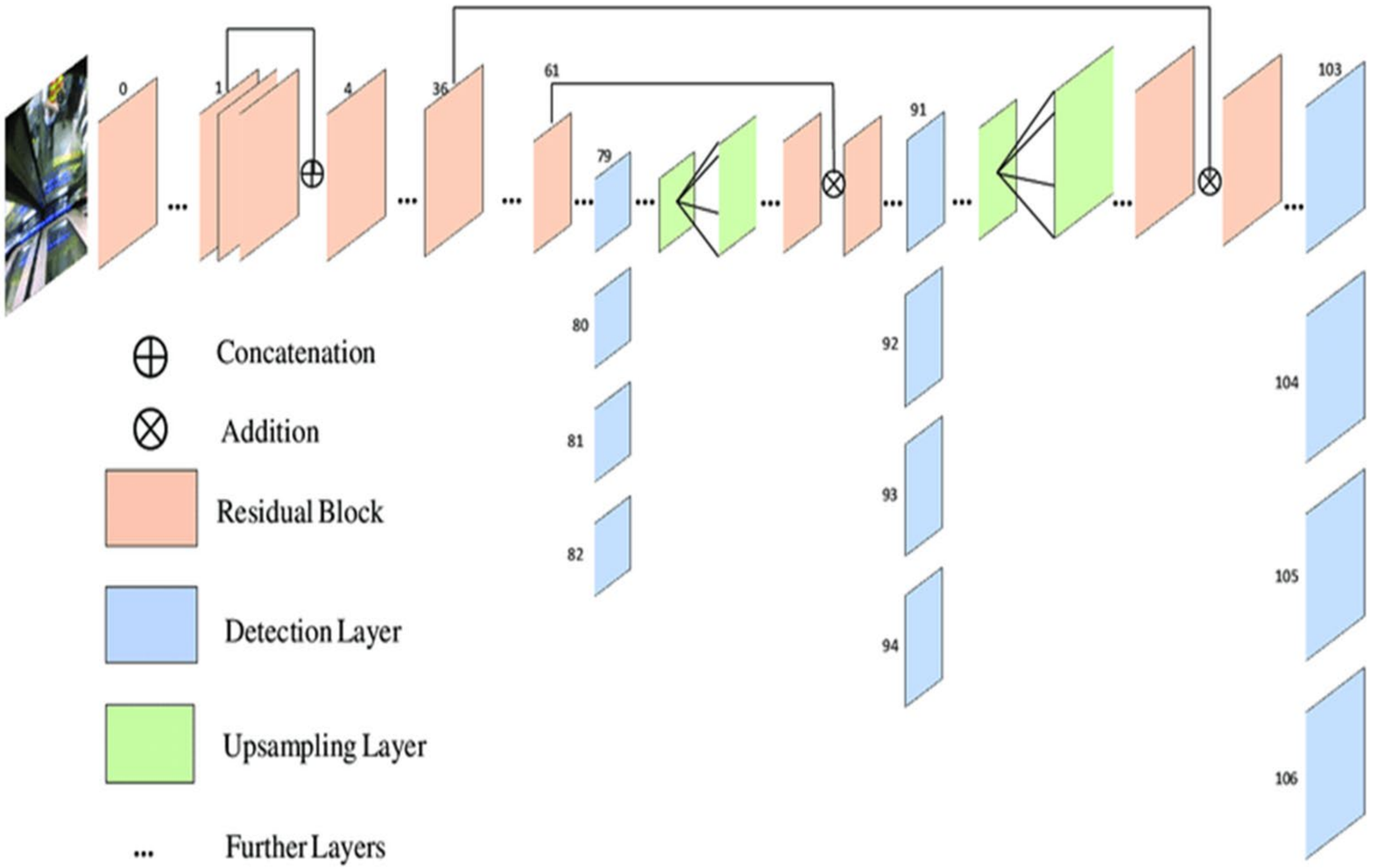
STATE-OF-THE-ART NEURAL NETWORKS

The neural network that was used is YOLO (You Only Look Once), in particular YOLOv3 and YOLOv5 were used.

YOLO is a convolutional neural network (CNN) for the real-time object detection.

The algorithm applies a single neural network to the full image dividing it into regions.

YOLO V3



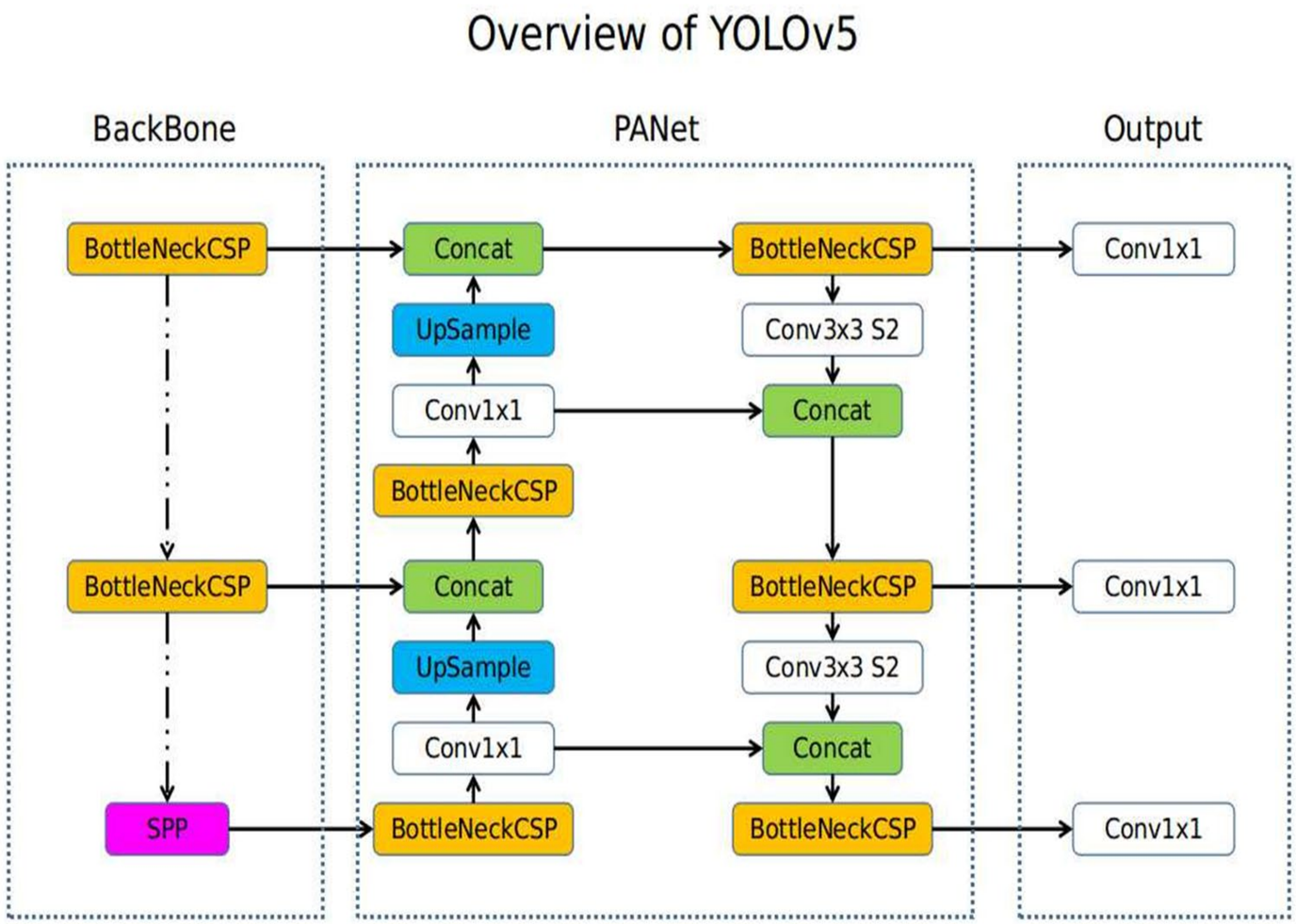
YOLOv3 uses Darknet-53 as its backbone feature extractor.

YOLOv3 increased the AP for small objects.

In YOLOv3, given an input image, what you get is a multiscale prediction.

Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

YOLO V5



The YOLO v5 has three important parts:
-Backbone
-Neck
-Head

In YOLO v5 the CSP (Cross Stage Partial Networks) are used as a backbone feature extractor.

For the model Neck YOLO v5 uses PANet (Path Aggregation Network).

https://github.com/ultralytics/yolov5

UNIVERSITÀ POLITECNICA DELLE MARCHE

VRAI

YOLOv3

Performance

Class	Images	Labels	Precision	Recall	mAP@.5	mAP@.5:.95
all	600	600	0.975	0.969	0.989	0.87
aruco0	600	40	0.975	0.96	0.991	0.876
aruco1	600	40	0.902	0.95	0.981	0.864
aruco2	600	40	0.949	0.923	0.968	0.86
aruco3	600	40	1	1	0.995	0.868
aruco4	600	40	0.928	0.967	0.979	0.863
aruco5	600	40	1	0.961	0.993	0.89
aruco6	600	40	1	0.995	0.995	0.868
aruco7	600	40	0.942	1	0.99	0.866
aruco8	600	40	0.998	0.975	0.995	0.868
aruco9	600	40	1	0.998	0.995	0.869
aruco10	600	40	0.977	0.975	0.995	0.908
aruco11	600	40	1	0.983	0.955	0.886
aruco12	600	40	0.974	0.925	0.986	0.845
aruco13	600	40	1	0.977	0.995	0.872
aruco14	600	40	0.974	0.943	0.981	0.845

$$\text{Precision} = \frac{T_P}{T_P + F_P}$$

$$\text{Recall} = \frac{T_P}{T_P + F_N}$$

Performance

YOLOv5

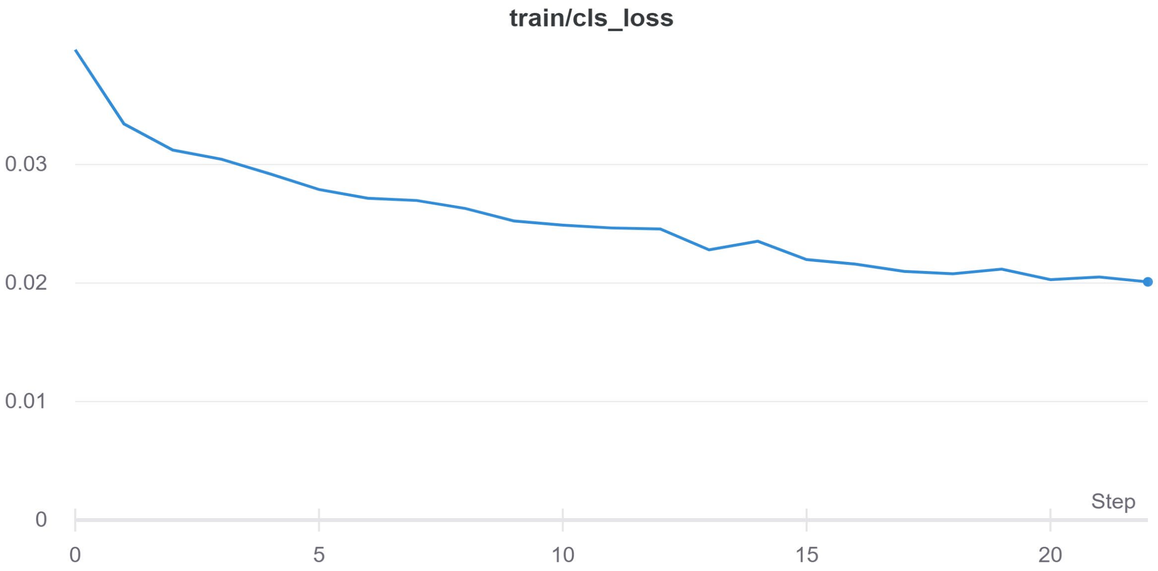
Class	Images	Labels	Precision	Recall	mAP@.5	mAP@.5:.95
all	600	600	0.784	0.724	0.803	0.654
aruco0	600	40	0.604	0.75	0.611	0.501
aruco1	600	40	0.614	0.75	0.761	0.62
aruco2	600	40	1	0.437	0.74	0.609
aruco3	600	40	0.806	0.825	0.897	0.689
aruco4	600	40	0.542	0.725	0.651	0.54
aruco5	600	40	0.916	0.813	0.913	0.763
aruco6	600	40	0.879	0.729	0.878	0.66
aruco7	600	40	0.779	0.375	0.637	0.534
aruco8	600	40	0.926	0.941	0.978	0.822
aruco9	600	40	0.604	0.775	0.703	0.591
aruco10	600	40	0.891	0.975	0.97	0.806
aruco11	600	40	0.833	0.775	0.908	0.744
aruco12	600	40	0.826	0.948	0.95	0.749
aruco13	600	40	0.869	0.499	0.79	0.657
aruco14	600	40	0.668	0.55	0.661	0.526

$$\text{Precision} = \frac{T_P}{T_P + F_P}$$

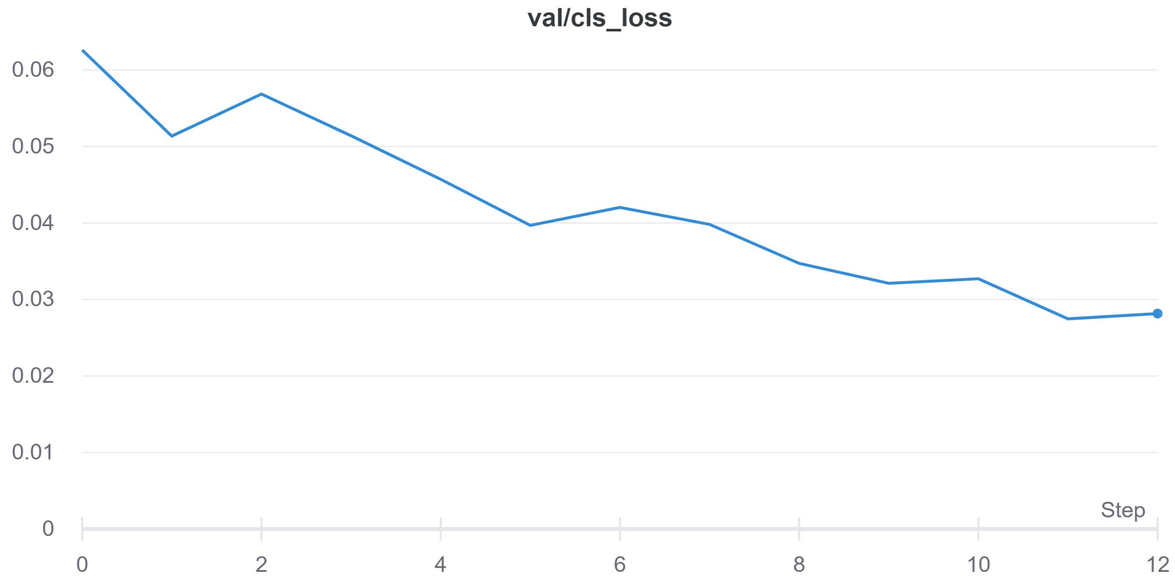
$$\text{Recall} = \frac{T_P}{T_P + F_N}$$

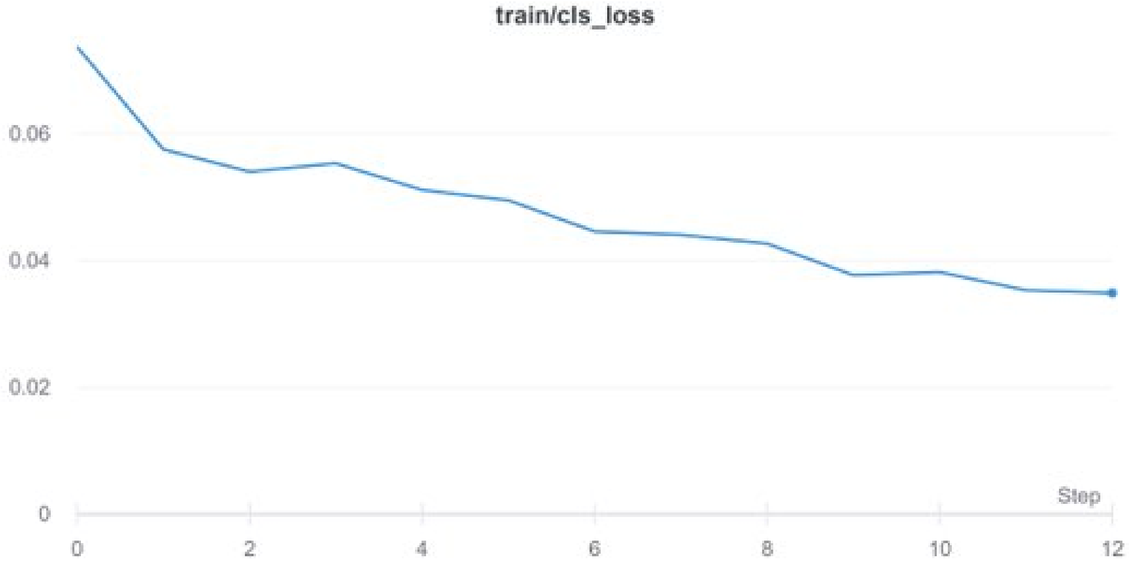
Performance

YOLOv3



YOLOv5





FUTURE DEVELOPMENTS

Dataset construction:

- ✔ Implemention of rotation, flip and brightness transformations
- ✔ Implemention of ArUco edge occlusions
- Implemention of other occlusions (for instance, occlusions of the corners or of the centre of the ArUco).
- Real dataset of the staff from a top-view camera

State-of-the-art Neural Networks:

- ✔ YOLOv3
- ✔ YOLOv5
- Using other state-of-the-art networks, for instance SSD.

THANK YOU FOR YOUR ATTENTION!