Assignment 2: Basic Detection

Stefanela Stevanović
IBB 22/23, FRI, UL
ss51676@student.uni-lj.si

I. INTRODUCTION

The goal of this assignment is to set up two popular detection methods: Viola-Jones/Haar cascade detector and YOLO (v5) and evaluate them on ear detection task using existing weights.

II. METHODOLOGY

Viola-Jones and YOLO detection methods were set-up and evaluated by calculating the Intersection over Union between the detected object box and the supplies ground-truths. IoU (Intersection over Union) is a term used to describe the extent of overlap of two boxes and is calculated as:

$$IoU(\%) = \frac{area\ of\ overlap\ of\ two\ boxes}{area\ of\ union\ of\ two\ boxes} \cdot 100 \quad (1)$$

IoU has value between 0% and 100%, where 0% means no overlap and 100% means complete overlap of two detection boxes. The accuracy of predictions for one method was calculated as:

$$ACC\ (\%) = \frac{TP}{TP+FP+FN} \cdot 100 \quad (2)$$

Where TP represents number of true positives (number of correct predictions), FP number of false positives (detected object is not a wanted object) and FN number of false negatives (wanted object is not detected on an image). Number of true positives is determined based on selected IoU threshold, such that predictions with IoU ≥ threshold are considered true positives and predictions with IoU ≤ threshold are false

Precision (P) is defined as the the number of true positives over the number of true positives plus the number of false positives, while recall (R) is defined as the number of true positives over the number of true positives plus the number of false positives.

$$P = \frac{TP}{TP+FP} \quad (3)$$

$$R = \frac{TP}{TP+FN} \quad (4)$$

III. EXPERIMENTS

The dataset used in the experiment is 500 images of human ears belonging to one class. The code was written in the Python programming language (version 3.9.7) in the Spyder environment.

YOLO v5 detector was set up by running *detect.py* script from git repository https://github.com/ultralytics/yolov5 and the detection was evalued by calculating IoU (1) between returned labels and supplied weights for each detection. Afterwards average IoU was calculated, so that for images were there is no detection IoU is 0, and accuracy was calculated with threshold 0.5

The Viola-Jones detection method was set up by loading XMLs of trained cascades and by using the built-in OpenCV detectMultiScale function on input images with three input parameters: scaleFactor, minNeighbors and minSize. scaleFactor is used to create an image pyramid, a multi-scale representation of an image, such that the detection can be scale-invariant. The image pyramid is performed by downsampling the image using neighboring pixels, after which a detection window shifts around the whole image at each scale to detect the ear. minNeighbors is a parameter specifying how many neighbors each candidate rectangle should have to retain it, while minSize defines minimum possible object size in pixels, such that objects smaller than that are ignored.

The VJ detection was run several times for different parameter values in order to find the set of parameters that give the best results. Combinations of the following parameter values were used: scaleFactor = { 1.1, 1.05, 1.025 }, minNeighbors = { 3, 4, 5 ,6 } and minSize = { (0,0), (15,15), (30,30) }. The most optimal set of parameters was selected based on the calculated average IoU (1), and calculated accuracy (2) with threshold 0.5.

After that, precision-recall graphs were made for different values of VJ parameters and YOLO baseline, where precision and recall were calculated using

formulas (3) and (4). For chosen optimal VJ parameter combination, five best representative VJ predictions in respect to IoU values and ten failed VJ predictions were shown.

IV. RESULTS AND DISCUSSION

In this part of assignment the results obtained in the experiments described in chapter III are presented and discussed.

A. Results

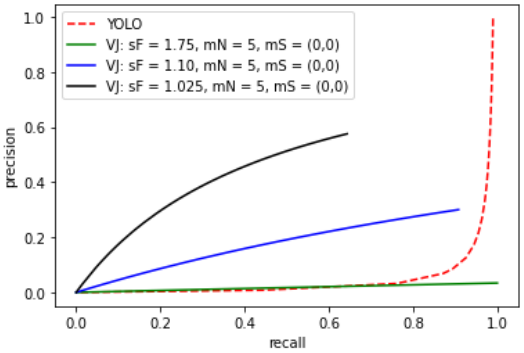
YOLO detection method has 98% accuracy and 82,5% average IoU. Table 1 shows accuracy (%) and average IoU (%) for different for VJ with different parameters.

TABLE 1
Accuracy and average IoU for different sets of VJ parameters

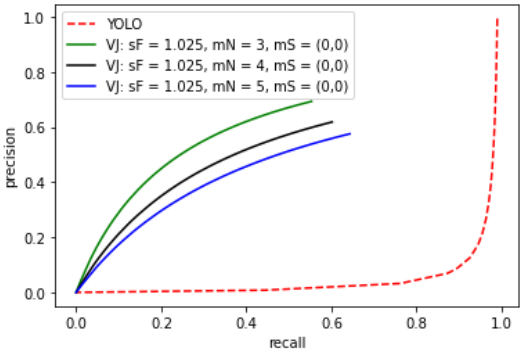
Scale Factor	Min Neighbors	Min Size	Acc (%)	Average IoU (%)
1.1	3	(0,0)	32.0	23.4
1.1	3	(15,15)	32.0	23.4
1.1	3	(30,30)	30.2	22.2
1.1	4	(0,0)	27.8	20.5
1.1	4	(15,15)	27.8	20.5
1.1	4	(30,30)	25.8	19.2
1.1	5	(0,0)	25.6	18.7
1.1	5	(15,15)	25.6	18.7
1.1	5	(30,30)	23.8	17.5
1.1	6	(0,0)	23.9	17.5
1.1	6	(15,15)	23.9	17.5
1.1	6	(30,30)	22.0	16.2
1.05	3	(0,0)	33.0	24.6
1.05	3	(15,15)	32.9	24.5
1.05	3	(30,30)	32.2	24.1
1.05	4	(0,0)	33.3	24.4
1.05	4	(15,15)	33.3	24.3
1.05	4	(30,30)	31.5	26.2
1,05	5	(0,0)	32.8	23.7
1,05	5	(15,15)	32.1	23.1
1.05	5	(30,30)	31.2	22.9
1.05	6	(0,0)	32.0	23.0
1.05	6	(15,15)	32.9	23.7
1.05	6	(30,30)	30.0	22.0
1.025	3	(0,0)	30.8	24.1
1.025	3	(15,15)	32.3	25.1
1.025	3	(30,30)	33.4	26.1
1.025	4	(0,0)	33.1	25.4
1.025	4	(15,15)	34.0	26.0
1.025	4	(30,30)	34.0	26.2
1.025	5	(0,0)	34.2	25.9
1.025	5	(15,15)	34.9	26.3
1.025	**5**	**(30,30)**	**34.9**	**26.7**
1.025	6	(0,0)	34.7	26.1
1.025	6	(15,15)	35.1	26.3
1.025	6	(30,30)	34.4	26.0

Figure 1 shows precision-recall plots for YOLO baseline and different VJ parameters. 1a) shows how plot changes with the change of scalec factor (sF), 1b) shows how plot changes with the change of minumum neighbors (mN) and 1c) shows how plot changes with the change of minimum object size (mS).

a)



b)



c)

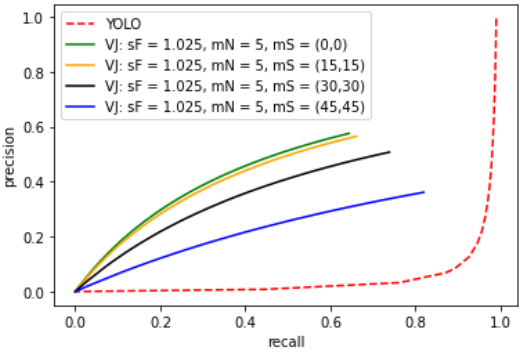


Fig.1. Precision-recall plot for YOLO baseline and VJ with different: a) scaleFactor (sF); b) number of minNeighbors (mN) and c) minSize (mS).

Figure 2 shows 5 best representative Viola-Jones images in respect of IoU value with VJ parameters scaleFactor = 1.025, minNeighbors = 5, minSize = (30,30), while table 2 shows these IoU values obtained with VJ and IoU values with YOLO for same 5 images. Images of 10 failed Viola-Jones predictions with same parameters are shown on figure 3.

Fig. 2. Best Viola-Jones predictions

TABLE 2
Comparison of average IoU for 5 best VJ predictions with YOLO.

Image	**VJ IoU (%)**	**YOLO IoU (%)**
a)	93.0	96.0
b)	92.4	83.4
c)	90.5	90.4
d)	90.4	83.2
e)	87.3	81.7

Fig. 3. Failed VJ predictions

B. Discussion

The YOLO detection method gives much better accuracy (98%) than Viola-Jones with any set of parameters shown in Table 1. For scale factor parameter, lower value means that smaller steps are used for resizing which increases the chance of detection, but this also makes the algorithm work slower. Higher number of minNeighbors results in less detections but with higher quality (more of them being true positive). Increasing the minSize parameter shows the same pattern of behavior as minNeighbors, because ignoring objects smaller than defined size results in fewer detections with greater number of TPs. An increase in the number of neighbors with smaller values of the scale factor (1.1 and 1.05) led to a decrease in accuracy, while with a scale factor of 1.025 this pattern changes and the accuracy begins to increase significantly with an increase in the number of minimum neighbors. The combination (1.025, 5, (30,30)) was taken as the most optimal set of parameters because it gives the highest value of average IoU (26.7%) and the second highest value of accuracy (34.9%). Only the combination (1.025, 6, (15,15)) gives greater accuracy (35.1%), but due to the 6 min neighbors and the low scale factor VJ algorithm with these parameters requires a lot of time, and therefore preference was given to the previous combination of parameters, which was further used in the work .

Figure 1 shows precision-recall plots for YOLO baseline and different VJ parameters. These plots have an unusual shape, showing that as the threshold decreases, both values increase. This is explained by the way TPs and FPs are calculated. For smaller threshold values, a greater number of detections are accepted as true positives, thus resulting in fewer FPs. The number of FNs does not change with the change of threshold because this number represents the number of images on which there is no detection and it is constant for one set of VJ parameters. Due to the constant number of FNs, the numerator in the formula for recall, which is the number of TPs, will increase faster with a decrease in the threshold than the denominator, which is the sum of TPs and FNs, which leads to an increase in recall. Also, in precision formula, the denominator is constant, so by reducing the threshold, only the numerator increases, which leads to an increase in the precision value.

Figure 2 shows 5 best VJ predictions and their IoU values are shown in Table 2. Table 2 shows that the IoU value for most of these images is over 90%, with the highest value of 93%, which is considered high accuracy. For the same images, YOLO managed to detect the desired object with great accuracy, however, it should be borne in mind that these are not the most accurate YOLO predictions, ie that VJ and YOLO will most accurately detect the desired object on different images.

In the figure 3 examples of failed VJ ear predictions are shown. Some of the reasons for failure are: blurred ear due to “background blur” effect (a), poor image resolution (b), improper lighting in the form of too much highlight (c) or too much shadow (d), the upper part of ear covered with hair (e) and (f), face facing the camera due to which the contours of the ear are less visible (g), rotated ear (h), the earphone in the ear (i) and the lower part of the ear covered by the hand (j).

V. CONCLUSION

This assignment showed that Viola Jones/Haar cascade algorithm faced many challenges. It proved to be sensitive to lighting, poor image resolution, rotated objects of interest, objects partially covered by another object,etc. The YOLO detection method eliminates many of these shortcomings of the Haar cascade, providing greater detection accuracy.