Assignment #2 (Basic Detection)

Matjaz Zupancic Muc
IBB 22/23, FRI, UL
mm1706@student.uni-lj.si

I. Introduction

In this work we compare performance of two popular real-time detection algorithms Viola-Jones (VJ) and YOLO. Algorithms are evaluated on the task of ear detection. We vary parameters of VJ and report performance (mean intersection over union (mean IoU) and mean accuracy precision (mAP@[0;0.01;1])) for each parameter set. We experiment with a parameter called "numDetections" and show that is somewhat correlated with the confidence of the model. Finally we take a look at a subset of predictions and try to elaborate on the results.

II. Related Work

[1] proposes a method for visual object detection which is both fast and achieves high detection rates. Speed of the model is achieved by a special image representation (Integral Image) and use of a classifier cascade, which is capable of performing quick and reliable rejections of the background regions of the image.

[2] proposes a real time ear detector, they train classifiers to detect left ears, right ears and ears in general. AdaBoost is used to train the classifier cascade. Positive training samples contain profile or almost profile images, the negative samples consist of wallpaper images. They show that a high detection and low False detection rate is achieved.

[3] proposes a so called YOLO model, which follows a regression based approach to object detection. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. A model is both fast and achieves high mAP. They show that the model outperforms methods such as DPM and R-CNN when challenged with object detection on artwork.

III. Methodology

We setup the OpenCV VJ implementation and PyTorch YOLO implementation. We use ear cascade provided by [2] and pre-trained ear detection weights for YOLO. We fine-tune parameters of VJ, namely the scaleFactor (specifies by how much the image size is reduced at each image scale) and minNeighbors (specifies how many neighbors each candidate rectangle should have to retain it). We evaluate the performance (mean IoU and mAP@[0;0.01;1]) for a detector which utilizes the combination of both left and right ear cascade (left and right ear cascade is applied on a single image, both detections are stored) and YOLO. In order to compute mAP@[0;0.01;1] for VJ we take advantage of the numDetections parameter of OpenCV VJ implementation. Parameter specifies the number of neighboring positively classified rectangles that were joined together to form a specific predicted bounding box. We compute the correlation between IoU and numDetections and compare it to the correlation of IoU and confidence levels returned by YOLO.

IV. Experiments

We report the mean IoU and mAP@[0;0.01;1] for VJ and YOLO. For VJ we vary the following two parameters scaleFactor and minNeighbors. We used the provided test dataset. Images are converted to Gray scale before they are feed to VJ and YOLO.

V. Results and Discussion

A. Results

TABLE I
Joined Ear Cascade Performance

scaleFfactor	minNeighbors	mean IoU	mAP
1.05	1	0.214	0.245
1.05	2	0.316	0.231
1.1	1	0.32	0.229
1.05	3	0.376	0.218
1.05	4	0.439	0.214
1.1	2	0.446	0.211
1.05	5	0.473	0.206
1.1	3	0.514	0.198
1.05	6	0.52	0.198
1.2	1	0.419	0.191
1.05	7	0.543	0.189
1.1	4	0.545	0.174
1.2	2	0.559	0.172
1.3	1	0.471	0.163
1.1	5	0.564	0.16
1.1	6	0.579	0.15
1.2	3	0.603	0.144
1.1	7	0.594	0.14
1.3	2	0.556	0.133
1.2	4	0.606	0.129
1.2	5	0.603	0.111
1.3	3	0.595	0.107
1.2	6	0.614	0.1
1.3	4	0.601	0.082
1.2	7	0.609	0.08
1.3	5	0.599	0.058
1.3	6	0.589	0.043
1.3	7	0.602	0.038

B. Discussion

Fig. 3 shows the relationship between the IoU and confidence for each prediction. Correlation between between IoU-numPredictions (VJ) is equal to 0.28. We see that correla-

TABLE II
YOLO PERFORMANCE

mean IoU	mAP
0.76	0.71

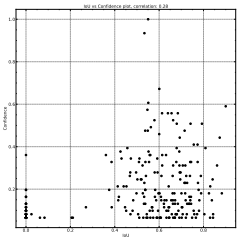

Fig. 1. VJ

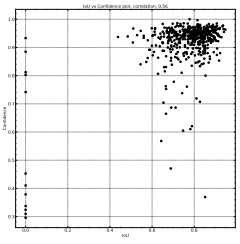

Fig. 2. VJ

Fig. 3. IoU vs Confidence

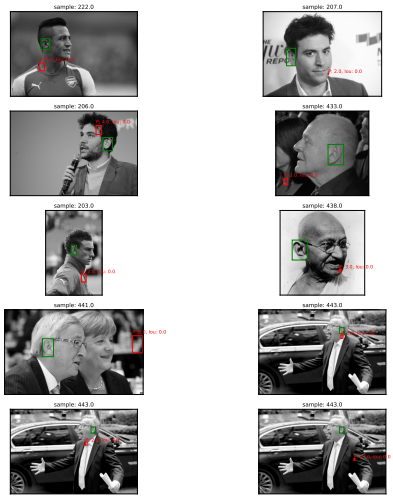

Fig. 4. VJ Worst 10 predictions

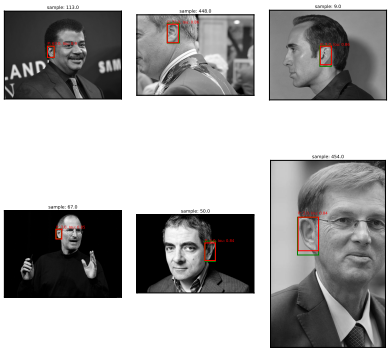

Fig. 5. VJ

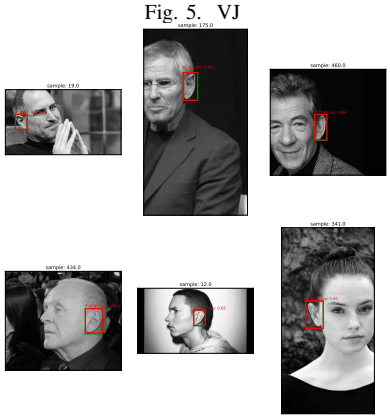

Fig. 6. YOLO

Fig. 7. Top 6 predictions

tion is positive, but is only 1/2 of the correlation of IoU-Confidence for YOLO. We decide to use the numPredictions to compute precision-recall curves and use them to compute mAP@[0;0.01;1], despite the fact that it is likely not the best estimate. From table I we can conclude that as minNeighbors increases the mean IoU increases, this is because the predicted detections have high support, because of that the recall is small and consequently the estimated mAP@[0;0.01;1] is small. Small scaleFactor in combination with a few minNeighbors results in high mAP, this is likely because few minNeighbors allow the model to make some predictions (i.e recall goes up), small scaleFactor allows the model to be certain about the predictions it makes (i.e predictions are likely good). Table II shows us that the performance of YOLO is far above the best VJ. YOLO is able to make great predictions (high mIoU) while being able to maintain a high mAP@[0;0.01;1]. Figure 4 shows the 10 worst predictions made by VJ with highest mAP@[0;0.01;1]. We see that since minNeighbors is equal to 1 the predicted bounding boxes are bad, if we instead plotted the VJ with highest mean IoU the worst prediction would have an IoU of 0.43. It seems like the model picks up small shaded regions of the image as ears. For example on sample 206 the model picks up hairline, if we look closely we see that the shape of it somewhat resembles the ear shape. On sample 203 the model picks up a part of the jacket which also resembles an ear (if looked from the side). On sample 441 the model picks up background, which again is in a shape of the ear. We also see that the model picks up super small regions, this could be avoided by also specifying the minSize parameter of algorithm. Figure 7 shows 6 best YOLO predictions and 6 best predictions made by VJ with highest mAP@[0;0.01;1]. We see that the results of the two models are comparable. 5 out of 6 best predictions by VJ do not have a pronounced background texture and ears are of relatively high resolution, this may be the reason why VJ does not struggle on this images.

VI. CONCLUSION

We evaluated VJ for different values of parameterss and found that as the mean IoU increases, mAP@[0;0.01;1] decreases. We show that YOLO outperforms the best VJ by a large factor. Finally we determined the quality of the numDetections parameter of VJ when used as a confidence measure.

REFERENCES

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. Ieee, 2001, pp. I–I.

[2] M. Castrillón-Santana, J. Lorenzo-Navarro, and D. Hernández-Sosa, "An study on ear detection and its applications to face detection," in *Conference of the Spanish Association for Artificial Intelligence*. Springer, 2011, pp. 313–322.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.