Ear segmentation using Mask R-CNN

Assignment #2

Image Based Biometrics 2020/21, Faculty of Computer and Information Science, University of Ljubljana

Novak Marko

***Abstract*—Mask R-CNN is currently one of the top-performing frameworks for object detection and segmentation. This report covers its use for ear segmentation on AWE database [1]. Source code used in this report is available on GitHub [2].**

I. Introduction

Mask R-CNN has been developed by researchers at Facebook AI Research as an upgrade to existing Faster R-CNN solution [3]. The key improvement in Mask R-CNN is that loss for segmentation masks is calculated in a class-independent way.

This report uses Mask R-CNN implementation by Matterport [4].

II. Methodology

AWE dataset comes with bounding boxes and ear masks. For best results with Mask R-CNN, each mask is split according to bounding boxes, so that we get an array of multiple masks per image. For faster training, these masks are stored as numpy files on start and reused on each run. For training and detection, images are scaled to 512px and padded with zeros until square.

The model has been trained over 30 epochs, each with 100 steps and 8 images per step, and learning rate 0,001. Training images from AWE dataset have been randomly rotated between -25 and 25 degrees, and horizontally flipped in 50% cases to increase the sample size for training. To reduce training time, initial weights for the model have been loaded from a model pre-trained on COCO dataset (`--weights coco`).

III. Results

Resulting model has been tested on test images from AWE dataset. While ROC curve 1 appears to be comparable with current R-CNN models used for object detection, I was certainly hoping for a higher true positive rate.

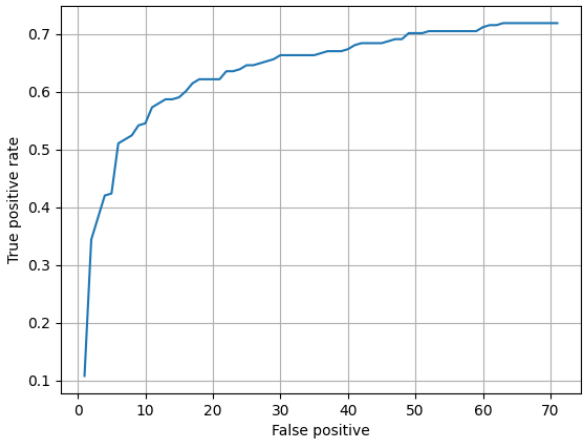

Figure 1: ROC curve for the resulting model.

IV. Conclusion

R-CNN models take significant time to train, which is why I opted to start from a pre-trained model and use a higher learning rate. Based on preliminary tests, this approach would likely provide better results if trained from scratch, on a larger number of steps and with lower learning rate.

V. Addendum

After spending additional time and training model starting from COCO dataset weights with learning rate 0,000 1, 30 epochs, 375 steps, and 8 images per step, the resulting model achieved results as shown in Figure 2. Compared to the original model, the second model manages to better score true and false matches.

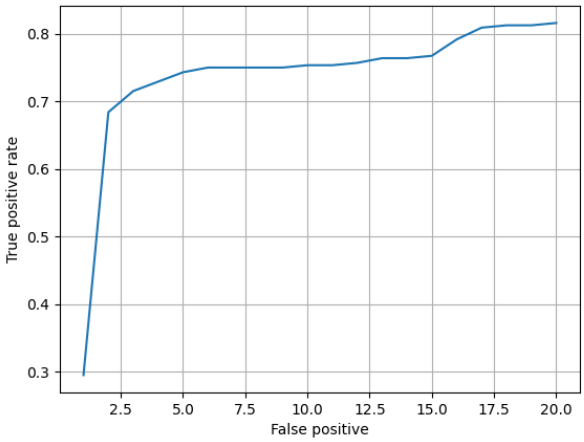

Figure 2: ROC curve for the second model.

References

[1] Z. Emersic, V. Struc, and P. Peer, "Ear recognition: More than a survey," *CoRR*, vol. abs/1611.06203, 2016. [Online]. Available: http://arxiv.org/abs/1611.06203

[2] M. Novak, "Mask rcnn." [Online]. Available: https://github.com/HitkoDev/Mask_RCNN

[3] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," *CoRR*, vol. abs/1703.06870, 2017. [Online]. Available: http://arxiv.org/abs/1703.06870

[4] W. Abdulla, "Mask r-cnn for object detection and instance segmentation on keras and tensorflow," https://github.com/matterport/Mask_RCNN, 2017.