

# Probabilistic Anchor Assignment with IoU Prediction for Object Detection: Supplementary Material

## 1 Appendix

### 1.1 Training Details

We train our models with 8 GPUs each of which holds two images during training. The parameters of Batch Normalization layers [2] are frozen as is a common practice. All backbones are pre-trained with ImageNet dataset [3]. We set the initial learning rate to 0.01 and decay it by a factor of 10 at 90k and 120k iterations for the 135k setting and at 120k and 160k for the 180k setting. For the 180k setting the multi-scale training strategy (resizing the shorter side of input images to a scale randomly chosen from 640 to 800) is adopted as is also a common practice. The momentum and weight decay are set to 0.9 and  $1e-4$  respectively. Following [1] we use the learning rate warmup for the first 500 iterations. It is noted that multiplying individual localization losses by the scores of an auxiliary task (in our case, this is predicted IoUs with corresponding GT boxes, and centerness scores when using the centerness prediction as in [4, 5]), which is also applied in previous works [4, 5], helps train faster and leads to a better performance.

### 1.2 Network architecture

Here we provide Figure 1 for a visualization of our network architecture. It is a modified RetinaNet architecture with a single anchor per spatial location which is exactly the same as models used in FCOS [4] and ATSS [5]. The only difference is that the additional head in our model predicts IoUs of predicted boxes whereas FCOS and ATTS models predict centerness scores.

### 1.3 More Ablation Studies

We conduct additional ablation studies regarding the effects of  $\text{topk } \mathcal{K}$  and the default anchor scale. All the experiments in the main paper are conducted with  $\mathcal{K} = 9$  and the default anchor scale of 8. The anchor size for each pyramid level is determined by the product of its stride and the default anchor scale<sup>1</sup>. Table 1 shows the results on different default anchor scales. It shows that the proposed probabilistic anchor assignment is robust to both  $\mathcal{K}$  and anchor sizes.

---

<sup>1</sup> So with the default anchor scale 8 and a feature pyramid of strides from 8 to 128, the anchor sizes are from 64 to 1024.

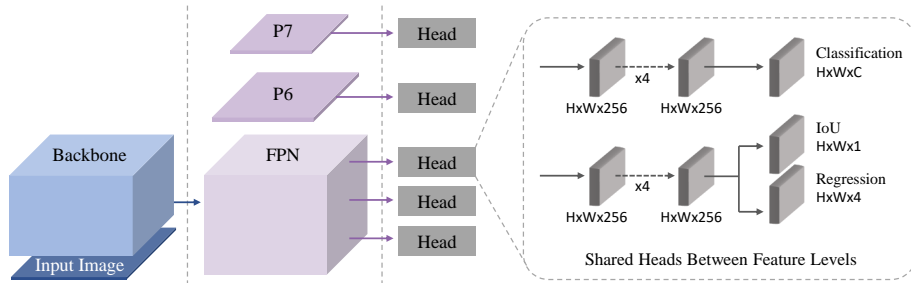


Fig. 1: The proposed model architecture. It has the same structure as FCOS [4] and ATSS [5] but predicts the IoU of detected boxes instead of the centerness.

Table 1: Ablation studies on COCO minival set with Res50 backbone. **Left:** Comparison of different topk  $\mathcal{K}$  values. **Right:** Comparison of different default anchor scales.

topk $\mathcal{K}$	AP	AP50	AP75
5	40.5	58.9	43.5
9	40.8	59.1	44.0
18	40.4	58.7	43.5
25	40.7	58.9	43.9

default anchor scale	AP	AP50	AP75
4	40.8	59.9	44.0
6	40.7	59.5	43.8
8	40.8	59.1	44.0
10	40.8	59.3	43.9

#### 1.4 More Visualization of Anchor Assignment

We visualize the proposed anchor assignment during training. Figure 2 shows anchor assignment results on COCO training set. Figure 3 shows anchor assignment results on a non-COCO image.

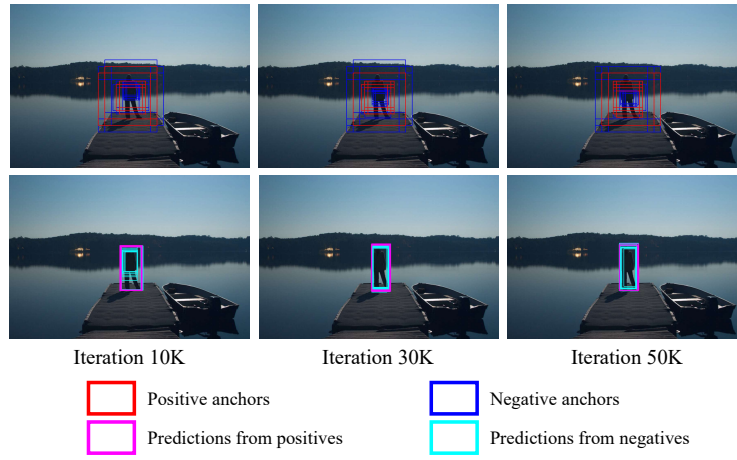
#### 1.5 Visualization of Detection Results

We visualize detection results on COCO minival set in Figure 4.

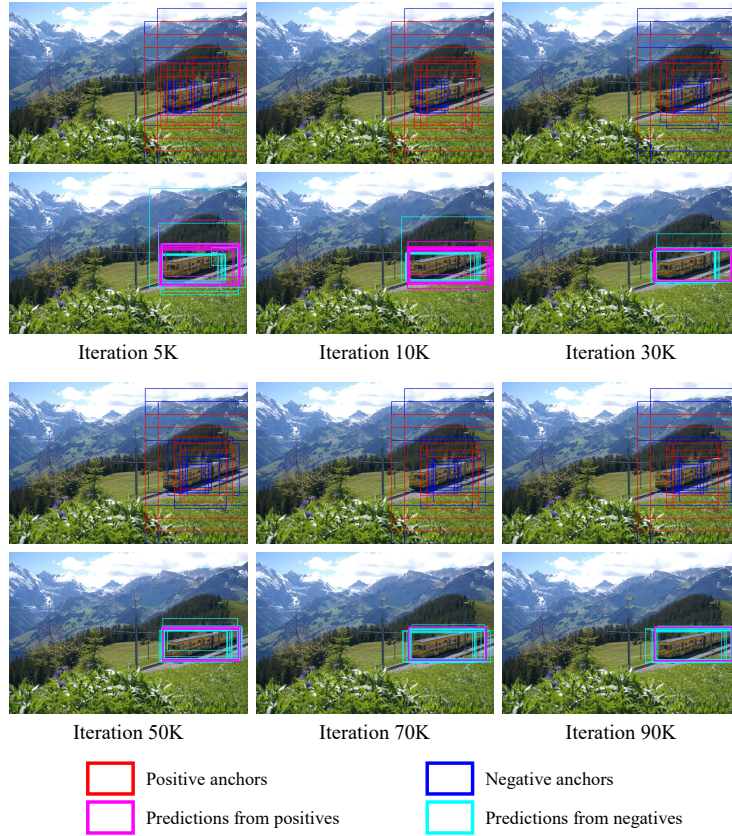
## References

1. Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., He, K.: Accurate, large minibatch sgd: Training imagenet in 1 hour. arXiv preprint arXiv:1706.02677 (2017)
2. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
3. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision **115**(3), 211–252 (2015)

4. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 9627–9636 (2019)
5. Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z.: Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. arXiv preprint arXiv:1912.02424 (2019)



(a) Image ID: 10388



(b) Image ID: 87156

Fig. 2: Evolution of anchor assignment and predicted boxes during training.

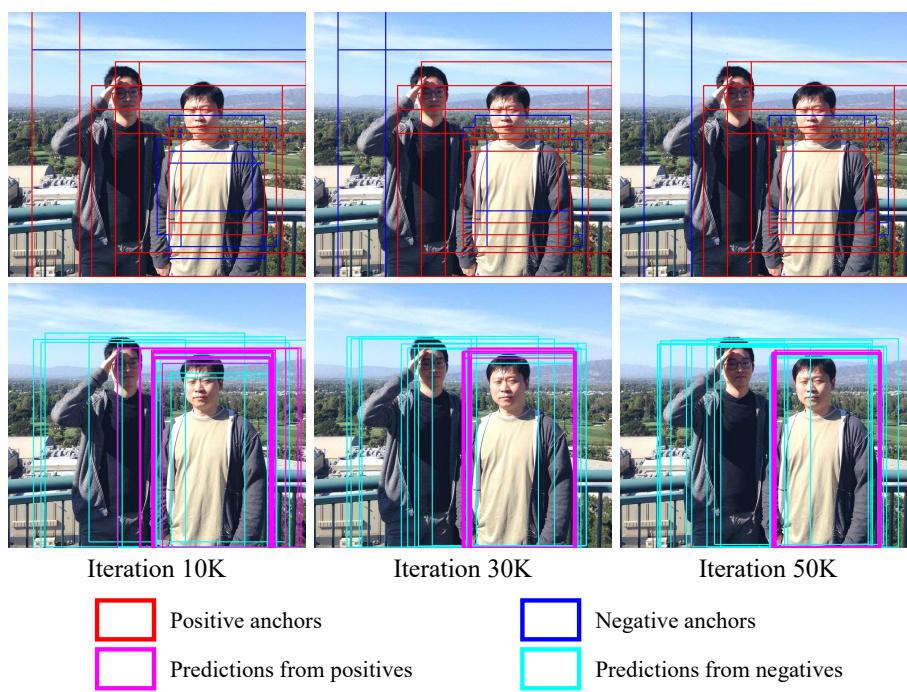


Fig. 3: Evolution of anchor assignment and predicted boxes on a non-COCO image.

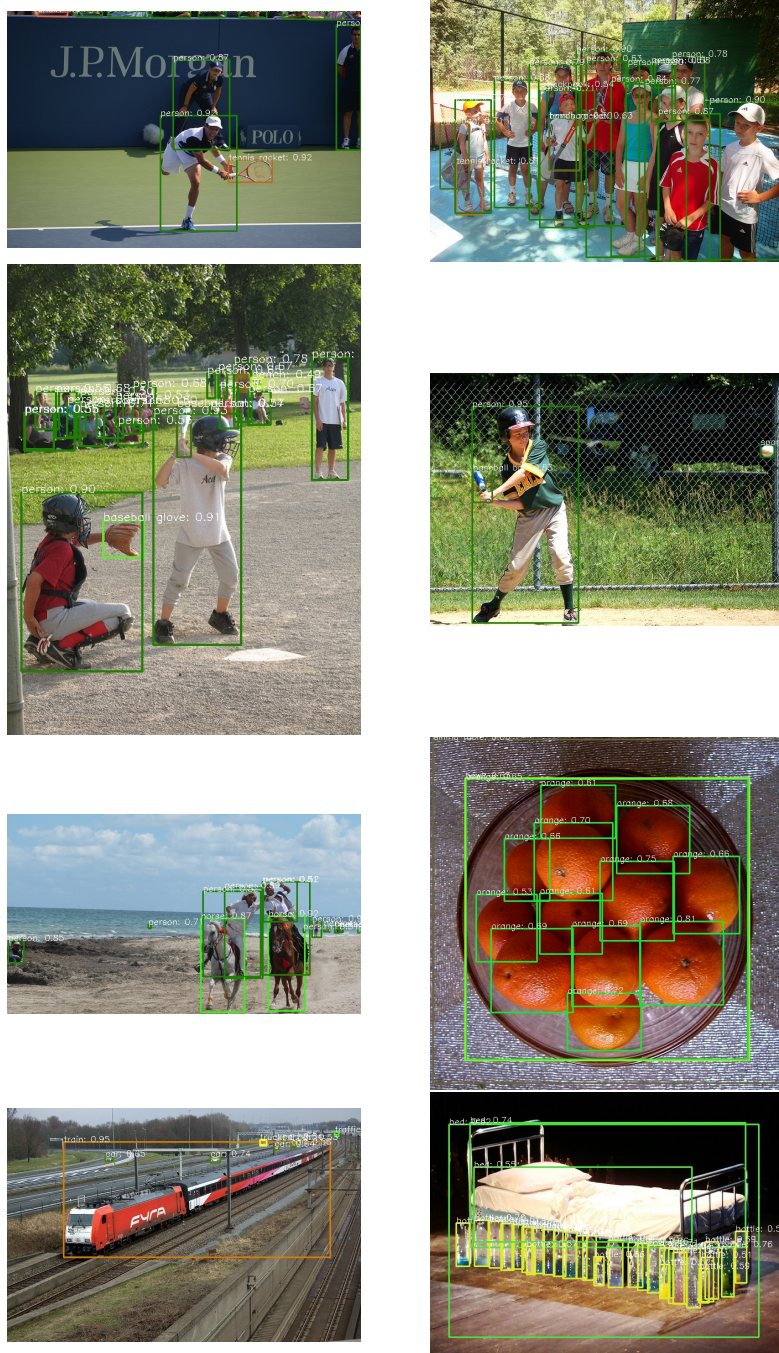


Fig. 4: Visualization of detection results on images of COCO minival set.