Soft Anchor-Point Object Detection
– Supplementary Material –

Chenchen Zhu, Fangyi Chen, Zhiqiang Shen, and Marios Savvides

Carnegie Mellon University, Pittsburgh PA 15213, USA
{chenchez,fangyic,zhiqians,marioss}@andrew.cmu.edu

S1 Discussion

Besides the ablation studies of the main paper, we ask more questions and conduct additional experiments to further understand our proposed training strategy and SAPD. We follow the same experimental setting as in the ablation studies in Section 4.1. All models are using the ResNet-50 [1] backbone.

S1.1 Soft-Weighting during Training or Testing?

Previous work like FCOS [5] applied soft-weighting in testing. Specifically, FCOS predicts the "center-ness" masks from extra network branches and the final score is computed by multiplying the predicted center-ness with the corresponding classification score. Differently, our soft-weighting scheme is applied during the training phase to down-weight anchor points' contribution to the network loss. In other words, FCOS is trained to predict the "center-ness" function but we are using the function to directly reweight the loss of anchor points.

	SW(ours)	CN(on cls.)	CN(on reg.)	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
FSAF [6]				35.9	55.0	37.9	19.8	39.6	48.2
	✓			**37.0**	55.8	39.5	20.5	40.1	48.5
		✓		36.1	55.2	38.1	20.3	40.0	47.4
			✓	36.5	55.5	38.9	21.0	40.1	48.3
	✓		✓	36.8	55.2	39.4	20.6	40.2	48.0

Table S1. Performance comparison between soft-weighting the loss during training by soft-weighted anchor points and down-weighting the confidence score during testing by predicted "center-ness". **SW**: soft-weighted anchor points, **CN**: center-ness, "on cls.": center-ness branch on the classification branch, "on reg.": center-ness branch on the regression branch.

For comparison between soft-weighting in training vs. in testing, we implement the "center-ness" mask branches attached to our baseline FSAF module [6] using the official code and optimize them the same way as [5]. Performances are reported in Table S1. Our soft-weighting scheme in training is more effective than the various versions of center-ness weighting in testing. The best version of

center-ness improves the AP by 0.6% while our soft-weighting scheme achieves a 1.1% AP gain. We think the reason is that soft-weighting during training is directly addressing the false attention issue, in which anchor points with poorly aligned features for precise localization are down-weighted. But in center-ness weighting, the anchor points are still contributing equally to the network loss, which is forcing all anchor points to perform equally well no matter how good are their feature representations. So the soft-weighting during testing is not fully resolving the false attention issue. This is further verified by the fact that if our soft-weighting scheme is applied on top of the center-ness weighting we can observe another improvement (see 4th and 5th entries in Table S1). However, if we compare between 2nd and 5th entries, applying center-ness weighting on top of our soft-weighting scheme is not improving the performance, which indicates that our soft-weighting scheme alone can work well to suppress the poorly localized detections. Therefore, we believe reweighting the anchor loss is more close to the essence of suppressing poorly localized detections than reshaping the confidence score during inference.

S1.2 Which Loss to Reweight?

In Eq. (2) of the main paper, the anchor point loss is the summation of the classification focal loss and the localization IoU loss for positive samples. By default, our soft-weighting scheme is applied to both classification and localization losses in the soft-weighted anchor points. In this section, we study the effect of applying our soft-weighting scheme to only the classification (cls) loss or the localization (loc) loss. Results are reported in Table S2. If we only reweight the single cls loss or loc loss, the performance becomes even worse than the baseline. The possible reason is that down-weighting a loss causes the network to focus on optimizing the other unweighted loss and the network is biased to be good at a single task. But the detection problem requires the network to be balanced for both proper classification and localization abilities.

	cls	loc	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
FSAF [6]			35.9	55.0	37.9	19.8	39.6	48.2
FSAF+SW	✓		33.3	50.6	35.4	18.3	37.7	43.1
FSAF+SW		✓	35.6	55.2	37.6	19.7	39.7	45.8
FSAF+SW	✓	✓	**37.0**	55.8	39.5	20.5	40.1	48.5

Table S2. The effect of applying our soft-weighting scheme to only the classification (cls) loss, or only the localization (loc) loss, or the summation of classification and regression loss (cls+loc) for the soft-weighted anchor points. **SW**: soft-weighed anchor points.

Therefore, compared to previous detection methods that reweighting only the classification loss [3, 2, 4], our soft-weighting scheme is more comprehensive and balanced.

S2 Visualization of Feature Selection Network

We visualize more examples of the soft-selected pyramid levels from the feature selection network in Figure S1. The feature selection network predicts the per-level "participation" degree for each instance to fully explore the power of feature pyramid and it is agnostic to the instance class, being general for a variety of objects including animals, human, food, vehicle, furniture, etc. Using features from multiple pyramid levels for detection is better than the online feature selection strategy in the FSAF module [6], which only chooses a single level to assign the instance when training the network.

Fig. S1. More visualization of the soft-selection weights from the feature selection network. Weights (the top-left red bars) ranging from 0 to 1 of five pyramid levels (P_3 to P_7) are predicted for each instance (blue box). The more filled a red bar is, the higher the weight is. *Best viewed in digital version and zoomed in.*

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
2. Law, H., Deng, J.: Cornernet: Detecting objects as paired keypoints. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 734–750 (2018)
3. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
4. Sun, F., Kong, T., Huang, W., Tan, C., Fang, B., Liu, H.: Feature pyramid reconfiguration with consistent loss for object detection. IEEE Transactions on Image Processing **28**(10), 5041–5051 (2019)
5. 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 (2019)
6. Zhu, C., He, Y., Savvides, M.: Feature selective anchor-free module for single-shot object detection. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)