

Supplementary Materials of "Sampling Techniques for Large-Scale Object Detection from Sparsely Annotated Objects"

Yusuke Niitani Takuya Akiba Tommi Kerola Toru Ogawa Shotaro Sano Shuji Suzuki
Preferred Networks, Inc.

{niitani, akiba, tommi, ogawa, sano, ssuzuki}@preferred.jp

1. More Results for the Competition

We give detailed experimental results of our submission to Google AI Open Images Competition 2018. For the competition, we added various techniques on top of *part-aware sampling*. The improvements made by each technique is found in Table 1.

In Table 2, we show the results of our single best model and ensemble of models. The best ensemble of models includes models fine-tuned exclusively on rare categories. These expert models improves the scores for rare categories because the single best model performs poorly on rare categories due to huge class imbalance in the dataset. During ensembling, we further boost performance by prioritizing certain models based on their validation scores so that outputs of weaker networks do not degrade the ensemble of prediction. We do not show the result on the validation set with this technique because the validation ground truth is used to tune parameters. For the results of the other top competitors, we only have the scores on the test set. We visualize a sample of our model’s detections in Figure 1.

References

- [1] Y. Gao, X. Bu, Y. Hu, H. Shen, T. Bai, X. Li, and S. Wen. Solution for large-scale hierarchical object detection datasets with incomplete annotation and data imbalance. *arXiv:1810.06208*, 2018.
- [2] Y. Zhu, C. Zhao, J. Wang, X. Zhao, Y. Wu, H. Lu, et al. Couplet: Coupling global structure with local parts for object detection. In *ICCV*, 2017.

Table 1: Performance of a single model with single scale testing on the validation split with bells and whistles.

	validation mAP
Baseline	64.5
+ Part-aware sampling	65.2 (+0.7)
+ 16 epochs	65.8 (+0.6)
+ Context head [2]	66.0 (+0.2)
+ SENet-154 and additional anchors	67.5 (+1.5)

Table 2: Ensemble of models with test-time augmentation. The validation score for the other competitors’ methods are not available.

	val	public test	private test
Single best (Ours)	69.95	55.81	53.43
Ensemble best w/o val tuning (Ours)	74.07	62.34	58.48
Ensemble best (Ours)		62.88	58.63
Private LB 1st place		<u>61.71</u>	58.66
Private LB 3rd place [1]		62.16	58.62

Figure 1: Visualization of our trained model for the images in the test set of Open Images Dataset v4. The best single model included in our submission to Google AI Open Images Competition 2018 is used for this visualization. We set the score threshold to 0.5.