BigDetection: A Large-scale Benchmark for Improved Object Detector Pre-training

Likun Cai[1*] Zhi Zhang[2] Yi Zhu[2] Li Zhang[1] Mu Li[2] Xiangyang Xue[1]
[1]Fudan University [2]Amazon Inc.

A. Pseudo Annotation Details

We provide the implementation details of pseudo annotation generation mechanism. As mentioned in paper, we adopt a multi-teacher strategy to improve the credibility of pseudo annotations. There are two teacher models utilized in our work as shown in Tab. 1. For both teachers, we set the score threshold as 0.5 and the NMS threshold as 0.6. After training, two teacher models achieve 23.1 and 24.1 AP on `bigdet_val` respectively, which are quite close. In Fig. 1, we show the comparison of AP results between two teacher models on single class. It can be seen that two teacher models have significant difference on single class, and the results gap even reaches 20 AP for class "*squid_(food)*". Mixing the outputs of multiple teacher models will fix the AP gap on single class, thereby improving the quality of pseudo annotations.

	Model	Schedule	Score	NMS
T1	CenterNet2	8×	0.5	0.6
T2	Cascade R-CNN	8×	0.5	0.6

Table 1. Basic setting for two teacher models. For both models, an 8× training schedule is adopted, and score and NMS threshold are set as 0.5 and 0.6.

B. Generalization Ability

One of the great benefits of pre-training on a large-scale dataset is a well-trained model only needs a few target labels to perform considerably well. Here, we show BigDetection pre-training is helpful across a variety of dataset sizes and semantic domains, and helps data efficiency.

Following the partially labeled data setting mentioned in paper, CenterNet2 [5] with FPN is adopted for fair comparison. The finetuning is done on PASCAL VOC [3] and Cityscapes [1] using 1% and 5% samples of train split. Tab. 2 compares the results of ImageNet [2] pre-trained model (Supervised), OpenImages [4] pre-trained model and BigDetection pre-trained model. Comparing with existing largest detection pre-training dataset, model pre-trained on BigDetection has better performance when dealing with insufficient training data.

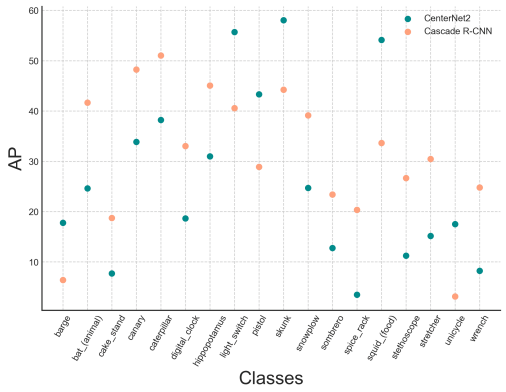

Figure 1. Comparison of AP results between different teacher models on several classes.

Methods	VOC 1%	VOC 5%	Cityscapes 1%	Cityscapes 5%
ImageNet	23.4	50.2	12.4	24.3
OpenImages	58.3	67.1	23.3	30.1
BigDetection	**64.6**	**72.6**	**31.8**	**38.9**

Table 2. Comparing with different pre-trained models under multiple partially labeled datasets.

C. Qualitative Results

In this part, we visualize detection results on images from COCO validation set that contain multiple small-scale objects. CenterNet2 [5] pre-trained on OpenImages [4] and BigDetection with 8× schedule are adopted for comparison. As shown in Fig. 2, left column shows detection results of OpenImages pre-trained model, and right column shows re-

*Work done during an internship at Amazon.

(a) OpenImages results

(b) BigDetection results.

(c) OpenImages results

(d) BigDetection results

(e) OpenImages results

(f) BigDetection results

Figure 2. Visual comparison on results of OpenImages and BigDetection pre-trained models. Left column shows detection results of OpenImages. Right column shows detection results of BigDetection. First row: "*knife*", "*carrot*" and "*cup*" classes have several small-scale objects that are not captured by OpenImages pre-trained model, even "*spoon*" is misclassified as "*fork*". Most of these objects present in the detection results of BigDetection pre-trained model with correct label predictions. Second row: results of small-scale object "*suitcase*". These objects are either not detected, or are misclassified by OpenImages pre-trained model, while almost all of them are captured by BigDetection pre-trained model. Third row: results of classes "*cup*" and "*wine glass*". BigDetection pre-trained model can better capture these small-scale objects while OpenImages pretrained model cannot.

sults of BigDetection pre-trained model.

In first row (Fig. 2a and Fig. 2b), for "*knife*", "*carrot*" and "*cup*" classes, several small-scale objects are not captured by OpenImages pre-trained model, even "*spoon*" is misclassified as "*fork*". However, most of these objects present in the detection results of BigDetection pre-trained model with correct class predictions. In second row (Fig. 2c and Fig. 2d), detection results on small-scale object "*suitcase*" is illustrated. These objects are either not detected, or are misclassified by OpenImages pre-trained model, while almost all of them are captured by our system. Finally, in third row (Fig. 2e and Fig. 2f), results of classes "*cup*" and "*wine glass*" are shown. BigDetection pre-trained model can better capture these small-scale objects while OpenImages pre-trained model cannot. In summary, BigDetection pre-trained model has better performance when facing small-scale objects, including capturing more small-scale objects and more accurate class prediction.

D. Synsets of BigDetection

In addition, we provide the final synsets of BigDetection dataset, which are obtained by our manual data-cleaning and careful designed category mapping principles. The synsets file *bigdetection_synsets.txt* is available at https://github.com/amazon-research/bigdetection.

References

[1] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 1

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1

[3] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111(1):98–136, 2015. 1

[4] Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, et al. The open images dataset v4. *International Journal of Computer Vision*, 128(7):1956–1981, 2020. 1

[5] Xingyi Zhou, Vladlen Koltun, and Philipp Krähenbühl. Probabilistic two-stage detection. *arXiv preprint arXiv:2103.07461*, 2021. 1