

# CA-SSL: Class-Agnostic Semi-Supervised Learning for Detection & Segmentation

Lu Qi<sup>1,3</sup>, Jason Kuen<sup>2</sup>, Zhe Lin<sup>2</sup>, Jiuxiang Gu<sup>2</sup>, Fengyun Rao<sup>1</sup>, Dian Li<sup>1</sup>,  
Weidong Guo<sup>1\*</sup>, Zhen Wen<sup>1</sup>, Ming-Hsuan Yang<sup>3</sup>, Jiaya Jia<sup>4</sup>

<sup>1</sup>QQ Browser Lab, Tencent

<sup>2</sup>Adobe Research

<sup>3</sup>The University of California, Merced

<sup>4</sup>The Chinese University of Hong Kong

## 1 Introduction

In this supplementary document, we provide the results from an additional experiment to demonstrate the effectiveness of our class-agnostic semi-supervised learning framework, CA-SSL:

- The comparison between using Objects365 [2] ground-truth annotations and the pseudo labels generated by our COCO-trained labeling detector, for pre-training.

as well as some experimental details for the three training stages of CA-SSL:

- The hyperparameter details for the Pseudo Labeler Training, Warmup Training, and Finetuning stages.

## 2 Additional experiments

**Comparison with Objects365 annotations.** Table 1 shows the comparison between using the class-agnostic pseudo labels (obtained using our COCO-trained pseudo labeler) and the large-scale class-specific ground-truth annotations of Objects365 [2], for warmup training. Compared to human-crafted large-scale class-specific annotations annotated by humans, the class-agnostic pseudo labels generated by our labeling detector are more effective. The reason is that the annotations in Objects365 dataset are usually partially/incompletely-annotated with many missing objects. It leads to more conflicting and ambiguous training signals in the Late Pretraining stage, thus causing the model to perform worse on the final downstream performance. Conversely, our generated pseudo labels are more reliable because they closely approximate the annotations of COCO dataset, which suffer significantly less from the problem of missing annotations.

---

\* Weidong Guo is the corresponding author.

Method	AP <sup>det</sup>	AP <sub>50</sub> <sup>det</sup>	AP <sub>75</sub> <sup>det</sup>
Baseline	46.8	66.2	50.8
Objects365 annotations	50.1	67.2	53.4
Pseudo Labels (ours)	<b>50.7</b>	<b>68.6</b>	<b>54.7</b>

**Table 1.** Comparison between using Objects365 [2] and the class-agnostic pseudo labels obtained with CA-SSL for warmup training. Here We report the downstream object detection task results obtained after finetuning the models.

**Ablation for long-tail problem.** For brevity, here we report the detection performance for the most rare COCO class “hair dryer”: class-specific detection achieves 4.7 AP, while class-agnostic detection (Ours) achieves 14.6 AP.

### 3 Hyperparameter details

#### 3.1 Pseudo Labeler Training

We strictly follow the training configs of Entity Segmentation [1]. The training epochs and batch sizes are 36 and 16. The optimizer is AdamW with a learning rate (LR) of 0.01. We decay the LR with 0.1 after 33 and 35 epochs. The longer edge size of the images is 1333. The shorter edge sizes of the images are randomly sampled from the range between 640 and 800 with a stride of 32.

#### 3.2 Warmup Training

We first use the pseudo labeler trained in the pseudo labeler training stage to generate pseudo labels on unlabeled images. In the pseudo-label generation step, we set the score threshold  $\delta$  to 0.2 to determine whether a particular prediction can be regarded as a pseudo label.

After that, we perform warmup training of the target detector with a setting similar to that of pseudo labeler training. For warmup training, we additionally consider the following: (1) the use of unlabeled image split; (2) strong data augmentation; (3) a longer training schedule with 60 epochs.

#### 3.3 Finetuning

Other than the differences resulted from using class-specific annotations on COCO `train2017` and 36 training epochs, the other training details in the finetuning stage are identical to those of warmup training.

## References

1. Qi, L., Kuen, J., Wang, Y., Gu, J., Zhao, H., Lin, Z., Torr, P., Jia, J.: Open-world entity segmentation. In: arXiv (2021) [2](#)
2. Shao, S., Li, Z., Zhang, T., Peng, C., Yu, G., Zhang, X., Li, J., Sun, J.: Objects365: A large-scale, high-quality dataset for object detection. In: ICCV (2019) [1](#), [2](#)