Appendix:
Instant-Teaching: An End-to-End Semi-Supervised Object Detection Framework

Qiang Zhou, Chaohui Yu, Zhibin Wang, Qi Qian, Hao Li
Alibaba Group
{jianchong.zq, huakun.ych, zhibin.waz, qi.qian, lihao.lh}@alibaba-inc.com

A. Learning Schedules

We provide more details on different learning schedules used in our experiments.

A.1. MS-COCO: Quick

Training

- **Batch size:** 16.
- **LR decay:** [0.01 ($\leq$120k), 0.001 ($\leq$160k), 0.0001 ($\leq$180k)].
- **Data processing:** Short edge size is sampled between 500 and 800 if the long edge is less than 1024 after resizing.
- **Batch per image for training Faster-RCNN head:** 64.

Testing

- **Data processing:** Short edge size is fixed to 800 if the long edge is less than 1024 after resizing.
- **Score threshold for testing Faster-RCNN head:** 0.001.

A.2. MS-COCO: Standard, [n] $\times$

Training

- **Batch size:** 16.
- **LR decay (1×):** [0.01 ($\leq$120k), 0.001 ($\leq$160k), 0.0001 ($\leq$180k)].
- **LR decay (2×):** [0.01 ($\leq$240k), 0.001 ($\leq$320k), 0.0001 ($\leq$360k)].
- **LR decay (3×):** [0.01 ($\leq$420k), 0.001 ($\leq$500k), 0.0001 ($\leq$540k)].
- **Data processing:** Short edge size is fixed to 800 if the long edge is less than 1333 after resizing.
- **Batch per image for training Faster-RCNN head:** 512.

Testing

- **Data processing:** Short edge size is fixed to 800 if the long edge is less than 1333 after resizing.
- **Score threshold for testing Faster-RCNN head:** 0.001.

A.3. PASCAL VOC

Training

- **Batch size:** 16.
- **LR decay:** [0.01 ($\leq$120k), 0.001 ($\leq$160k), 0.0001 ($\leq$180k)].
- **Data processing:** Short edge size is fixed to 600 if the long edge is less than 1000 after resizing.
- **Batch per image for training Faster-RCNN head:** 256.

Testing

- **Data processing:** Short edge size is fixed to 600 if the long edge is less than 1000 after resizing.
- **Score threshold for testing Faster-RCNN head:** 0.001.

B. Hyperparameters

In this section, we provide descriptions of the hyperparameters used in our experiments, as shown in Tabel 1. Unless otherwise specified, we use the same hyperparameters in both the MS-COCO and PASCAL VOC experiments.

Hyperparameters	Description	Value
λ	The bounding box regression loss weight	1.0
λ_u	The unsupervised loss weight	1.0
τ	The confidence threshold	0.9
α_m	The coefficient of $Beta$ distribution	1.0
λ_m	The mixing coefficient of Mixup	$Beta(\alpha_m, \alpha_m)$
LR	The initial learning rate	0.01
Momentum	The momentum used in SGD	0.9
Weight decay	The weight decay	$1e{-}4$
Training steps	The total training steps	180k
Batch size	The batch size	16
Batch ratio	The ratio between labeled and unlabeled images in a batch	1:1

Table 1. Descriptions of the hyperparameters used in our experiments.

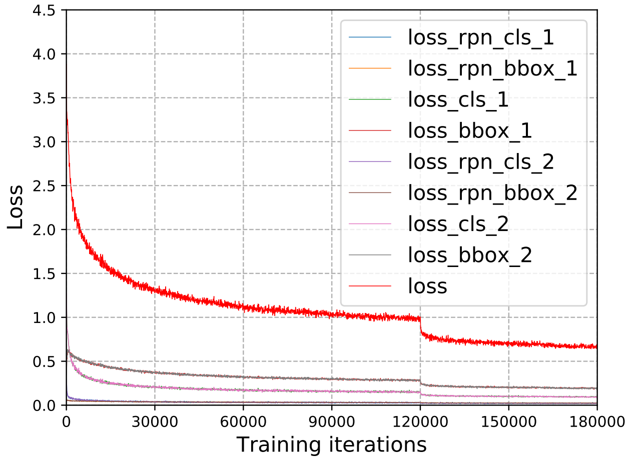

Figure 1. Value change of loss w.r.t. training iterations.

C. Convergence Analysis

In this section, we empirically evaluate the convergence of Instant-Teaching*. As shown in Fig. 1, we can observe that both the two models trained with our Instant-Teaching* method can reach a steady convergence. These results demonstrate that our Instant-Teaching* can not only achieve state-of-the-art results, it can also be trained easily.

D. Ablation Study on Backbone

In this section, we verify the effect of different backbones on our Instant-Teaching* framework. From Table 2, we replace the ResNet-50 backbone and test the efficacy of the supervised baseline, Instant-Teaching, and Instant-Teaching* method on the 2% protocol respectively. We can observe that our Instant-Teaching* can reach better performance with a more powerful backbone. In other words, it is easy to elevate the performance of our Instant-Teaching* framework by using a more powerful backbone.

E. Experimental Config

For the detailed experimental configuration, please refer to the attached "`config.py`". We will release the source code soon.

Methods	Backbone	2% COCO
Supervised	R50-FPN	12.70±0.15
Instant-Teaching		20.70±0.30 (+8.00)
Instant-Teaching*		22.45±0.15 (+9.75)
Supervised	R101-FPN	15.80±0.50
Instant-Teaching		22.10±0.15 (+6.30)
Instant-Teaching*		23.50±0.20 (+7.70)
Supervised	X101-32x4d-FPN	16.60±0.20
Instant-Teaching		22.40±0.15 (+5.80)
Instant-Teaching*		24.20±0.15 (+7.60)
Supervised	R2N-101-FPN	17.4±0.30
Instant-Teaching		23.5±0.15 (+6.10)
Instant-Teaching*		25.9±0.20 (+8.50)

Table 2. Comparison of mAP for different semi-supervised methods with different backbones on the 2% MS-COCO protocol. The value in brackets represents the mAP improvement compared to the corresponding supervised model.