Supplementary Material for Paper titled by Transformation Invariant Few-Shot Object Detection

Aoxue Li Zhenguo Li
Huawei Noah's Ark Lab, China
lax@pku.edu.cn, Li.Zhenguo@huawei.com

1. Training Details

As in meta-learning approaches [20,23], we train our approach in two phases: In the first base training phase, we train our full model with the whole base class training data. In the second finetuning phase, we add the few-shot samples from novel classes into model training. To balance training samples for base and novel classes, we randomly select K labeled instances for each base class from the whole training dataset and combine them with the few-shot novel class samples to form a new finetuning dataset for the second phase. The full model is finetuned with the small finetuning dataset. For fair comparison with FA, we use the same optimization strategy as in [23].

2. Implementation Details

Our model is based on the state-of-the-art few-shot object detection approach, i.e., FA [23]. The approach consists of two branches: a guidance extraction branch and a query prediction branch. For the guidance extraction branch, we add the proposed TGC loss to minimize the difference between guidance vectors of original images and their transformed variants. The guidance vectors are extracted from the same guidance extractor as in [23]. It is implemented as a convolutional neural network (i.e. ResNet-101) followed by a sigmoid function. For the query prediction branch, we produce RoI features of original images conditioned on the RoI proposals extracted from their transformed variants. These proposals are extracted by using the backbone network and the region proposal network as in [23].

3. Additional Results under Semi-Supervised FSOD Scenario

Table 1 provides the comparative results for semi-supervised few-shot object detection on the third novel class set of PASCAL VOC dataset. Our approach is shown to be more effective than two competing baselines. In particular, it achieves comparable or even better results than fully-supervised approaches in some cases. This also validates its effectiveness for semi-supervised few-shot object detection.

Supervision	Model	Novel Class Set 3				
		1-shot	2-shot	3-shot	5-shot	10-shot
Fully	MRCNN [24]	14.3	18.2	27.5	41.2	48.1
	TFA/w.fc [20]	15.7	27.2	34.7	40.8	44.6
	TFA/w.cos [20]	17.9	27.2	34.3	40.8	45.6
	FA [23]	21.2	30.0	37.2	43.8	49.6
Semi-25%labeled	TIGE(ours)	18.5	25.2	26.9	35.5	46.6
	TIQP(ours)	18.4	25.6	26.6	34.9	46.3
	TIP(ours)	**19.2**	**26.9**	**27.7**	**36.2**	**47.7**
Semi-50%labeled	TIGE(ours)	20.4	28.4	31.8	40.4	49.3
	TIQP(ours)	20.3	28.2	32.1	40.2	48.7
	TIP(ours)	**21.5**	**29.7**	**33.2**	**41.6**	**50.1**

Table 1. Comparative results for semi-supervised few-shot object detection on the PASCAL VOC dataset. We evaluate the performance on three different sets of novel categories. This table provides the results on the third novel class set, while results of the other two novel class sets are provided in Table 5 in the main text. The mean average precision (%) on the novel classes is used as the evaluation metric of this dataset. The reported results are averaged over multiple runs.