Data-free Knowledge Distillation for Object Detection (Supplementary Material)

Akshay Chawla, Hongxu Yin, Pavlo Molchanov and Jose Alvarez
NVIDIA
Santa Clara, CA
akshaych@alumni.cmu.edu, {dannyy, pmolchanov, josea}@nvidia.com

1. Hyperparameters for DIODE

Table 1 provides the hyper parameters used for DIODE inversion of Yolo-V3 detector pre-trained on MS-COCO dataset.

Hyperparameter	value
Φ	Yolo-V3
Φ_{stu}	Yolo-V3
Φ_{verif}	Yolo-V3-Tiny
Dataset of Φ,Φ_{verif}	MS-COCO
N (batchsize)	$256 \rightarrow 96 \rightarrow 48$
Resolution (H=W)	$160 \rightarrow 320 \rightarrow 416$
optimizer family	Adam(lr, $\beta_1 = 0, \beta_2 = 0$)
optimizer lr	$0.2 \rightarrow 0.002 \rightarrow 0.0002$
scheduler	cosine decay w/ warmup
iterations	$2500 \rightarrow 1500 \rightarrow 1000$
α_{TV}	75.0
α_{BN}	0.1
α_{l2}	1e-7

Table 1: Hyperparameters for using DIODE to invert an off-the-shelf Yolo-V3 detector pre-trained on MS-COCO. Multiple hyperparameters in a row represent values used for multi-resolution synthesis on dimensions $160 \rightarrow 320 \rightarrow 416$.

2. Hyperparameters for $\mathbf{Y}_{fp}$ sampling

False positive sampling ($\mathbf{Y}_{fp}$ sampling) is an approach to aggregate semantically relevant false positive objects that appear with high confidence during the DIODE optimization process as described in section 3.1 in main paper. Its hyperparameters are listed in Table 2. In order to stabilize the process, we have a warmup period where no predictions are aggregated. Early exit when the image has stabilised and no more objects are appearing with high confidence. Predictions with a confidence greater than "min confidence" that have less than "overlap IoU" with an existing target are considered false positives and added to targets. We also filter out very small (area $\leq$ Min area) and very large (area $\geq$ Max area) predictions.

Hyperparameter	value
Warmup	800
Total iterations	4000
Early exit	2800
Min confidence	0.2
Overlap IoU	0.35
Min area	1%
Max area	85%

Table 2: Hyperparameters for $\mathbf{Y}_{fp}$ sampling for DIODE on Yolo-V3 object detector pre-trained on MS-COCO.

3. Hyperparameters for Knowledge Distillation

Table 3 provides hyper parameters used for distilling student detector from teacher detector.

Hyperparameter	value
Φ	Yolo-V3
Φ_{stu}	Yolo-V3
batchsize	64
Image size	training: 416, eval: 640
optimizer	Adam(0.001, $\beta_1 = 0.9, \beta_2 = 0.999$)
scheduler	Cosine decay w/ warm up
iterations	500k

Table 3: Hyperparameters for mimic learning knowledge distillation on Yolo-V3.