

Fully Supervised and Guided Distillation for One-Stage Detectors

— Supplementary Material

Deyu Wang¹, Dongchao Wen^{1,*}{0000-0001-7311-1842}, Junjie Liu¹, Wei Tao¹, Tse-Wei Chen²,
Kinya Osa², Masami Kato²

¹Canon Information Technology (Beijing) Co., LTD, China

{wangdeyu, wendongchao, liujunjie, taowei}@canon-ib.com.cn

²Device Technology Development Headquarters, Canon Inc., Japan

twchen@ieee.org

More detection analysis

Fig 1 and Fig 2 demonstrate the visualization of false detection composition and performance for non-distilled and distilled student model of Tiny-YOLO distilled by Darknet53 teacher network with the analysis tool [1].

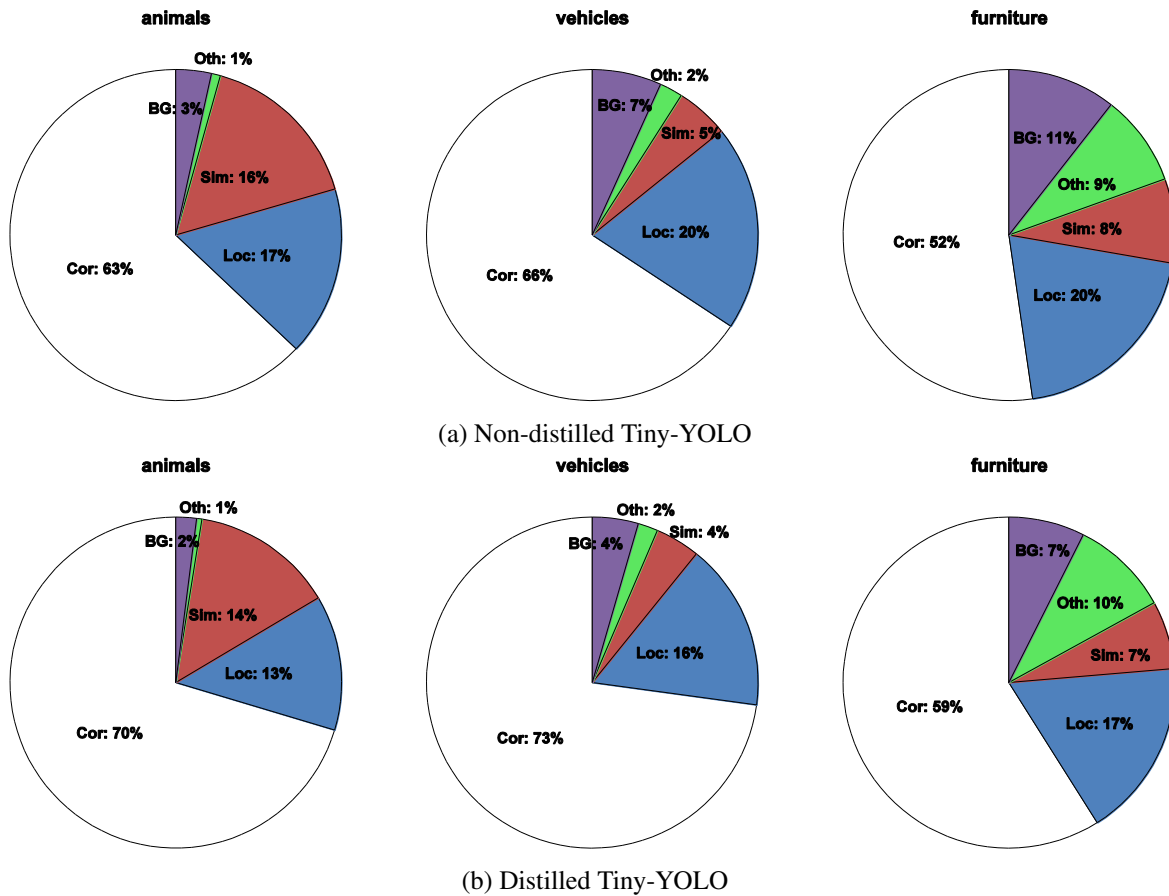


Figure 1: Analysis of detection for Tiny-YOLO distilled with Darknet53 and non-distilled Tiny-YOLO on the Pascal VOC 2007 dataset. Correct fraction of detections (Cor), false positives due to poor localization (Loc), confusion with similar objects (Sim), confusion with other objects (Oth) and error in background (BG).

The observation for three sub-set of object categories (animals, vehicles and furniture) is show in Fig 1, which demonstrates that our algorithm can help the student network to improve recall signicantly, and effectively reduce false detections, especially for the error in BG. Besides, the false detection composition reveals that our algorithm not only promotes the localization ability (Loc) of the student network, but also reduces the errors in background (BG) as well as the confusion in the similar category (Sim) and other category (Oth). In Fig 2, the x-axis from left to right reflects detection outputs of high to low confidence.

*Dongchao Wen is corresponding author.

Obviously, for these three object categories, the detection performance of distilled student network has been improved and their recalls are also increased, especially for furniture category.

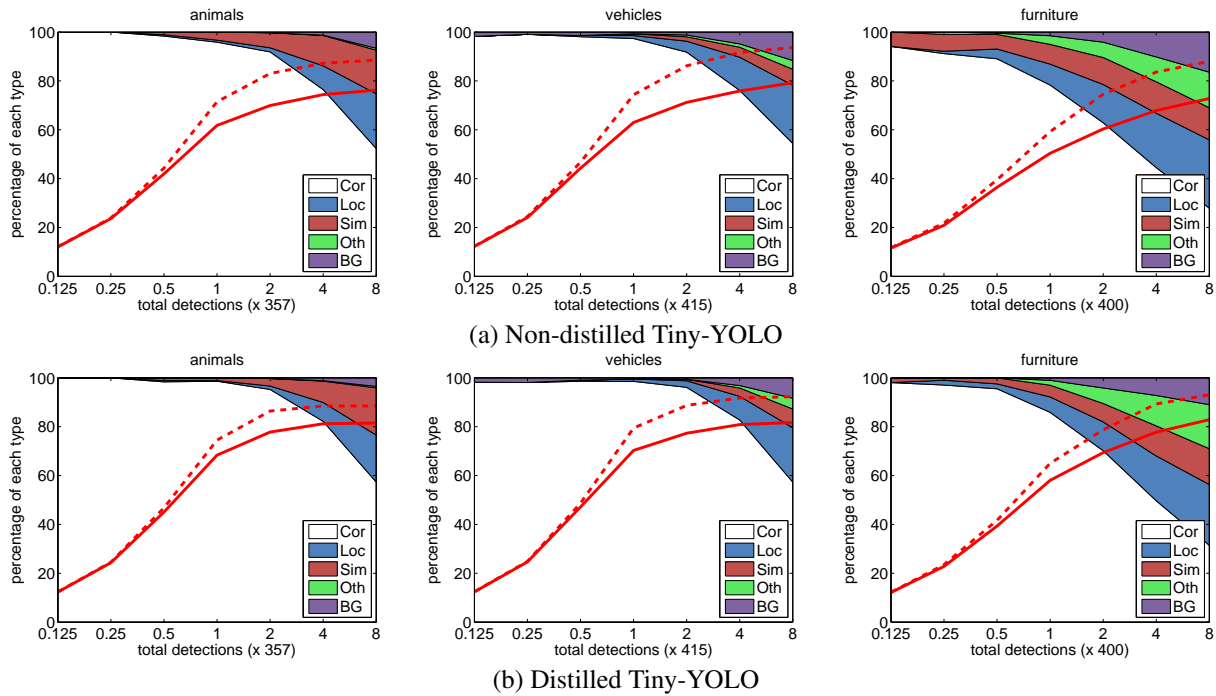


Figure 2: Performance for distilled and non-distilled detector on vehicles, animals, and furniture for Tiny-YOLO based student on Pascal VOC 2007 dataset. Line plots show recall as the number of detections increases, where dashed and solid red line are based on weak criteria (0.1 jaccard overlap) and strong criteria (0.5 jaccard overlap) respectively.

Table 1: Class-wise performance comparison on Pascal VOC 2007 test dataset by using Darknet53, Resnet50 to distill Tiny-YOLO, Resnet18 and MobileNetV2 respectively (mAP, %). -D means that the student network distilled with the teacher network based on our FSGD algorithm.

Model	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
TinyYOLO	57.1	65.3	70.4	43.3	38.1	21.8	65.1	69.0	69.8	34.7	56.0	56.7	66.0	77.5	74.3	65.8	29.4	53.7	55.7	68.8	60.7
TinyYOLO-D (Resnet50)	61.3	67.9	74.1	51.0	48.3	27.6	72.4	73.6	74.1	40.4	56.4	60.6	67.7	77.5	74.7	68.6	33.2	54.8	64.4	72.7	65.6
	+4.2	+2.6	+3.7	+7.7	+10.2	+5.8	+7.3	+4.6	+4.3	+5.7	+0.4	+3.9	+1.7	0.0	+0.4	+2.8	+3.8	+1.1	+8.7	+3.9	+4.9
TinyYOLO-D (Darknet50)	65.7	69.3	77.8	57.3	51.9	30.7	75.6	75.8	80.5	43.2	61.3	68.5	75.1	83.0	77.6	71.0	36.1	63.8	69.6	76.7	68.4
	+8.6	+4.0	+7.4	+14.0	+13.8	+8.9	+10.5	+6.8	+10.7	+8.5	+5.3	+11.8	+9.1	+5.5	+3.3	+5.2	+6.7	+10.1	+13.9	+7.9	+7.7
MobileNetV2	68.6	73.1	77.9	63.1	55.4	39.9	79.0	80.6	81.0	49.5	67.2	69.8	72.8	83.4	80.2	74.8	39.5	63.1	69.9	80.3	71.3
MobileNetV2-D (Resnet50)	70.1	70.9	80.5	65.4	58.5	42.7	78.7	81.8	85.2	52.3	67.6	68.8	75.1	84.1	81.1	76.4	42.3	65.9	71.9	82.0	71.2
	+1.5	-2.2	+2.6	+2.3	+3.1	+2.8	-0.3	+1.2	+4.2	+2.8	+0.4	-1.0	+2.3	+0.7	+0.9	+1.6	+2.8	+2.8	+2.0	+1.7	-0.1
MobileNetV2-D (Darknet53)	71.4	73.4	80.1	68.2	59.6	42.4	78.5	81.8	85.8	52.1	70.4	67.7	79.9	85.2	82.3	75.7	44.8	68.5	76.5	82.8	72.3
	+2.8	+0.3	+2.2	+5.1	+4.2	+2.5	-0.5	+1.2	+4.8	+2.6	+3.2	-2.1	+7.1	+1.8	+2.1	+0.9	+5.3	+5.4	+6.6	+2.5	+1.0
Resnet18	69.2	69.3	79.7	66.2	54.8	41.9	77.9	79.8	84.6	48.5	66.3	67.6	79.2	82.3	79.9	75.7	42.9	68.5	71.6	77.6	69.5
Resnet18-D (Resnet50)	71.0	73.9	80.9	67.0	58.0	44.3	79.9	82.3	85.7	52.0	66.5	70.2	78.0	83.9	81.7	76.9	43.5	66.4	72.9	83.6	72.7
	+1.8	+4.6	+1.2	+0.8	+3.2	+2.4	+2.0	+2.5	+1.1	+3.5	+0.2	+2.6	-1.2	+1.6	+1.8	+1.2	+0.6	-2.1	+1.3	+6.0	+3.2
Resnet18-D (Darknet53)	73.5	75.7	80.9	71.9	63.1	43.7	82.6	83.7	84.7	55.6	72.1	73.4	82.6	85.4	82.6	77.4	47.0	73.3	78.1	84.4	72.6
	+4.3	+6.4	+1.2	+5.7	+8.3	+1.8	+4.7	+3.9	+0.1	+7.1	+5.8	+5.8	+3.4	+3.1	+2.7	+1.7	+4.1	+4.8	+6.5	+6.8	+3.1

Class-wise performance analysis

Table 1 shows the class-wise performance details on Pascal VOC 2007 dataset by using Darknet53 and Resnet50 to distill Tiny-YOLO, Resnet18 and MobileNetV2 respectively. These lightweight student networks get significant improvement in most of categories, especially for the Tiny-YOLO-D (Darknet53) and Resnet18-D (Darknet53), these two distilled models get fully improvement in all categories and have respectively 8.56% and 4.35% absolute gains in mAP.

References

1. D. Hoiem, Y. Chodpathumwan, and Q. Dai. Diagnosing error in object detectors. In European conference on computer vision (ECCV), pages 340-353. Springer, 2012. 1