ScratchDet: Training Single-Shot Object Detectors from Scratch -Supplementary Material-

Rui Zhu[1,4]*, Shifeng Zhang[2*], Xiaobo Wang[1], Longyin Wen[3], Hailin Shi[1]†, Liefeng Bo[3], Tao Mei[1]
[1]JD AI Research, China; [2]CASIA, UCAS, China; [3]JD Digits, USA; [4]Sun Yat-sen University, China
{zhurui10,wangxiaobo8,longyin.wen,shihailin,liefeng.bo,tmei}@jd.com, shifeng.zhang@nlpr.ia.ac.cn

1. Complete Object Detection Results

We show the complete object detection results of the proposed ScratchDet method on the PASCAL VOC 2007 `test` set, PASCAL VOC 2012 `test` set and MS COCO `test-dev` set in Table 1, Table 2 and Table 3, respectively. Among the results of all published methods, our ScratchDet achieves the best performance on these three detection datasets, *i.e.*, 86.3% mAP on the PASCAL VOC 2007 `test` set, 86.3% mAP on the PASCAL VOC 2012 `test` set and 39.1% AP on the MS COCO `test-dev` set. And we select some detection examples on the PASCAL VOC 2007 `test` set, the PASCAL VOC 2012 `test` set and the MS COCO `test-dev` in Figure 1, Figure 2, and Figure 3, respectively. Different colors of the bounding boxes indicate different object categories. Our method works well with the occlusions, truncations, inter-class interference and clustered background.

Table 1. Object detection results on the PASCAL VOC 2007 `test` set. All models use Root-ResNet-34 as the backbone network.

Method	Data	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
ScratchDet300	07+12	80.4	86.0	87.7	77.8	73.9	58.8	87.4	88.4	88.2	66.4	84.3	78.4	84.0	87.5	88.3	83.6	57.3	80.3	79.9	87.9	81.2
ScratchDet300+	07+12	84.1	90.0	89.2	83.6	80.0	70.1	89.3	89.5	89.0	73.0	86.9	79.8	87.4	90.1	89.3	87.1	63.3	86.9	83.5	88.9	83.4
ScratchDet300	COCO+07+12	84.0	87.9	89.3	85.6	79.8	69.4	89.1	89.2	88.5	73.2	87.5	81.7	88.4	89.5	88.7	86.3	63.1	84.5	84.3	88.1	85.6
ScratchDet300+	COCO+07+12	86.3	90.4	89.6	88.4	85.4	78.9	90.1	89.3	89.5	77.4	89.7	83.9	89.1	90.3	89.5	88.3	68.1	87.6	85.9	87.4	87.7

Table 2. Object detection results on the PASCAL VOC `2012 test` set. All models use Root-ResNet-34 as the backbone network.

Method	Data	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
ScratchDet300	07++12	78.5	90.1	86.8	74.5	66.3	54.0	83.7	82.6	91.6	64.1	83.1	67.7	90.1	87.6	87.8	85.7	56.9	81.7	74.6	87.2	75.3
ScratchDet300+	07++12	83.6	92.2	90.3	82.6	73.9	68.1	86.8	90.5	93.9	70.3	88.0	72.3	92.3	91.5	91.0	90.3	63.6	87.6	77.4	89.9	80.2
ScratchDet300	COCO+07++12	82.1	91.7	89.3	79.1	71.9	62.7	85.7	85.3	93.9	68.8	87.2	68.7	91.9	90.6	90.9	88.2	61.2	84.7	79.2	89.7	81.0
ScratchDet300+	COCO+07++12	86.3	94.0	91.8	86.0	78.9	75.6	88.6	91.3	95.1	74.0	90.0	73.0	93.6	93.0	92.6	91.9	69.7	90.2	80.9	91.8	83.7

Table 3. Object detection results on the MS COCO `test-dev` set. All models use Root-ResNet-34 as the backbone network.

Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
ScratchDet300	32.7	52.2	34.9	13.0	35.6	49.0	29.3	43.9	45.7	20.6	50.8	65.3
ScratchDet300+	39.1	59.2	42.6	23.1	43.5	51.0	33.1	53.3	58.3	36.6	63.4	74.5

*Equally-contributed and this work is done at JD AI Research.
†Corresponding author.

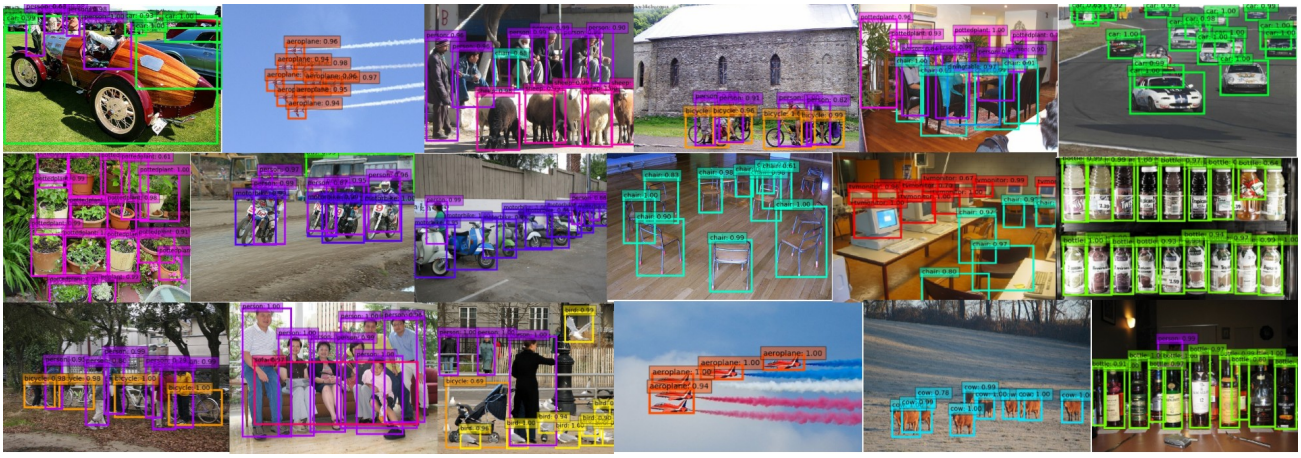


Figure 1. Qualitative results of ScratchDet300 on the PASCAL VOC 2007 `test` set (corresponding to 84.0% mAP). The training data is 07+12+COCO.

Figure 2. Qualitative results of ScratchDet300 on the PASCAL VOC 2012 `test` set (corresponding to 82.1% mAP). The training data is 07++12+COCO.

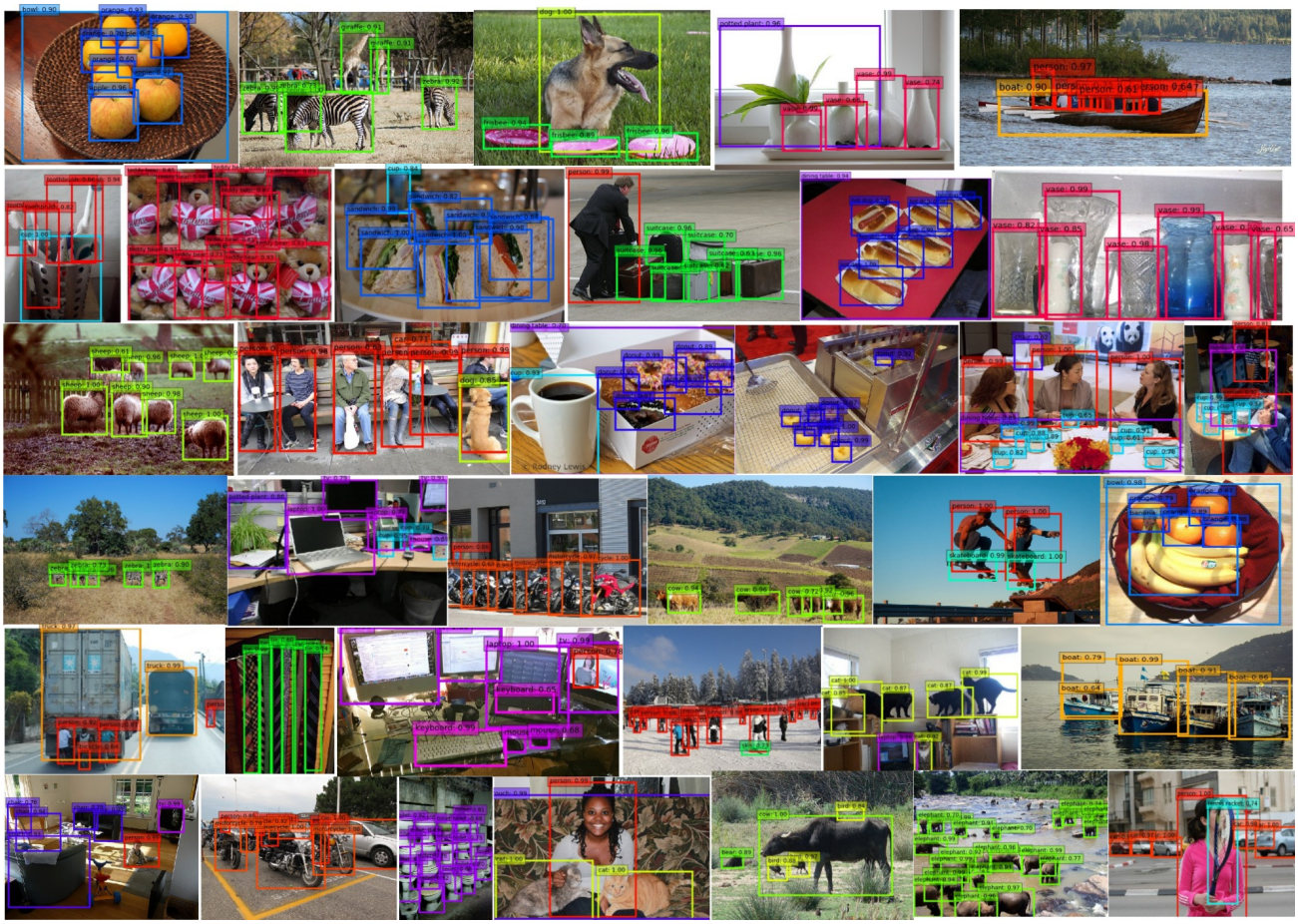
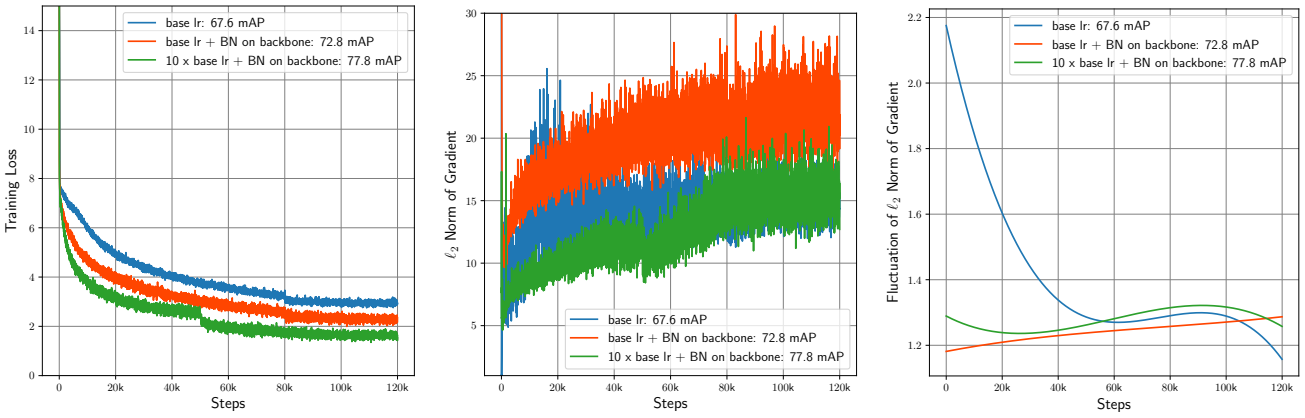


Figure 3. Qualitative results of ScratchDet300 on the MS COCO `test-dev` set (corresponding to 32.7% mAP). The training data is COCO `trainval35k`.



(a) Loss Value (b) L2 Norm of Gradient (c) Fluctuation of L2 Norm of Gradient

Figure 4. Analysis of the optimization landscape of SSD after adding BatchNorm on the backbone subnetwork. We plot (a) the training loss value, (b) L2 Norm of gradient and (c) the fluctuation of L2 Norm of gradient of three detectors. The blue curve represents the original SSD, the red and green curves represent the SSD trained with BatchNorm on the backbone network using base learning rate and $10\times$ base learning rate, respectively. It is the similar trend with the curves of adding BatchNorm on the detection head subnetwork.

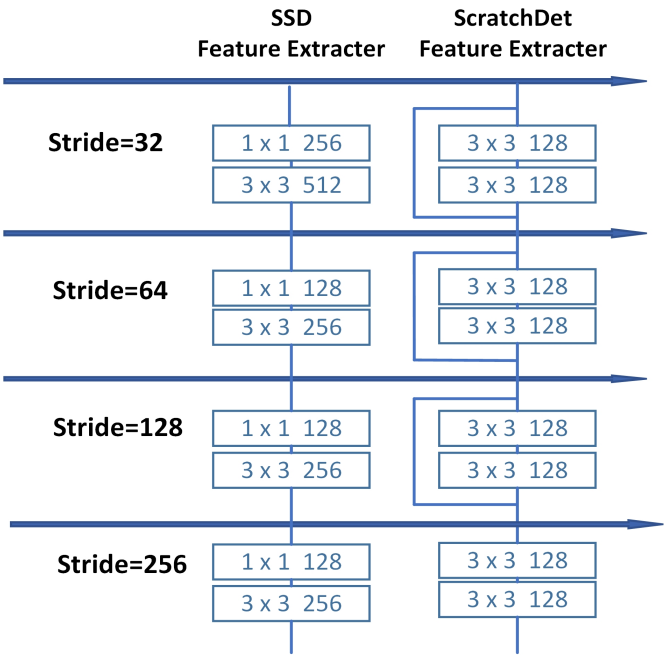


Figure 5. Comparison of the extra added layers between SSD and ScratchDet. This change brings less parameters and computions.