Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training

Hongkai Zhang[1,2], Hong Chang[1,2], Bingpeng Ma[2],
Naiyan Wang[3], and Xilin Chen[1,2]

[1] Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, Chinese Academy of Sciences
[2] University of Chinese Academy of Sciences, [3] TuSimple
hongkai.zhang@vipl.ict.ac.cn, changhong@ict.ac.cn, bpma@ucas.ac.cn, winsty@gmail.com, xlchen@ict.ac.cn

1 Effectiveness on One-stage Detectors

Considering that dynamic training is a general viewpoint, we also try to apply the dynamic design on one-stage object detectors. Specifically, we choose a representative one-stage detector RetinaNet [21] to validate the effectiveness of our method.

Table 1. Effectiveness of the dynamic design on RetinaNet with ResNet-50-FPN as backbone on COCO `val` set.

Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
RetinaNet	35.6	55.4	38.4	20.3	39.5	46.5
Dynamic RetinaNet	36.3	55.5	38.8	20.7	39.9	47.5

As shown in Table 1, our dynamic design brings 0.7 points higher box AP than the RetinaNet baseline. It should be noted that RetinaNet has already changed the β of SmoothL1 Loss to a small value (0.11), so in our experiment, adjusting β to a slightly smaller value (0.05 with DSL) has little effect. Moreover, since the input of RetinaNet is the pre-defined anchors, the distribution of input is relatively fixed. So the impact of DLA can be regarded as using a more reasonable IoU threshold (e.g. 0.55) for training. We believe our method can be applied to other one-stage detectors if their inputs are more dynamic, like Guided Anchoring [34].

2 Experimental Results on PASCAL VOC dataset

To further demonstrate the effectiveness of our method, we conduct experiments on PASCAL VOC [1] dataset with the same hyperparameters as on MS COCO

Table 2. Experimental results using ResNet-50-FPN backbone on PASCAL VOC2007 `test` set.

Method	AP	AP_{50}	AP_{60}	AP_{70}	AP_{80}	AP_{90}
Faster R-CNN	47.2	76.9	71.0	59.9	39.2	8.1
Dynamic R-CNN	48.7	76.9	71.5	60.9	42.1	11.6

dataset. We use the union of VOC2007 and VOC2012 `trainval` as the training set and report the results on the VOC2007 `test` set. As shown in Table 2, Dynamic R-CNN improves the Faster R-CNN baseline by 1.5 points AP and 3.5 points AP_{90}. Thus we reach similar conclusions as from the results on COCO dataset, which validates the effectiveness and universality of Dynamic R-CNN.

References

1. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. IJCV **88**(2), 303–338 (2010)