Boosting Dense Long-Tailed Object Detection from Data-Centric View -Supplementary Material-

Weichen Xu, Jian Cao(✉), Tianhao Fu, Hongyi Yao, and Yuan Wang

Peking University, Beijing, China
{xuweichen1999,tianhaofu1,yhy}@stu.pku.edu.cn, {caojian}@ss.pku.edu.cn

Appendix 1. Alternative Calculations of the Copy Times

In this appendix, we consider alternative calculations of the copy times t_g and demonstrate that these can be equally effective. In Eq. (2), We set the copy times for all rare class instances to ρ. We define it as hard mode. In other cases, we use the long-tailed scaling factor s_{c_g} defined in Eq. (3). In particular, we set $\varepsilon = 3$, $\gamma = 5$ and $\mu = 2$. In addition, We subtract 1 from s_{c_g}, so that the range of s_{c_g} lies between 0 and 3. In soft mode, copy times t_g is calculated as:

$$t_g = \text{round}(s_{c_g}), \tag{5}$$

where round stands for rounding function. In this way, the copy times gradually increase with the rarity r_{c_g}, and the maximum is 3. For random mode, After calculating the copy times for all instances using Eq. (5), Only one instance is randomly selected for Copy-Move from the instances with copy times greater than 0. Similarly, only the maximum copy times greater than 0 are selected for max mode. For all experiments, we use ResNet-50 as backbone for 24 epochs. Table 6 shows the experimental results. It can be seen that the performance of all modes is higher than baseline, and the hard mode used in the paper has achieved the best performance in overall AP. This proves the effectiveness of our proposed Copy-Move and that precise rule is not crucial.

Table 6: More calculations of the copy times t_g.

mode	AP	AP_r	AP_c	AP_f
Baseline	27.5	20.2	26.1	32.4
hard	**28.1**	20.8	**26.7**	32.8
soft	27.8	**21.0**	26.1	32.8
random	28.0	20.9	26.3	**33.0**
max	28.0	20.9	26.6	32.7

Appendix 2. Training Details

We add mixup [1], InstaBoost [2], and Copy-Paste [3] to baseline. For mixup, we use the same multi-scale as REDet during training. For InstaBoost, we trained in the same way as [2], i.e., the shorter edge of the input image is smaller than 800 pixels, and the longer edge is smaller than 1333 pixels with a probability of 0.5. As for Copy-Paste, we train with large scale jittering on the image size of 1024×1024 mentioned in [3]. All models are evaluated at the same size, i.e., the shorter edge of the input image is smaller than 800 pixels, and the longer edge is smaller than 1333 pixels.

Appendix 3. Performance on COCO Dataset

On COCO, we calculate r_{c_g} using Eq. (1) with $t = 0.01$. We trained RetinaNet with ResNet-50 for 12 epochs. The Right part of Table 7 shows results. Both Copy-Move and LTTSS can improve the performance, but not obvious. We think this is because COCO is relatively balanced.

Table 7: The supplementary experiments on the COCO dataset.

	COCO, 12 epochs		
method	AP	AP_{50}	AP_{75}
no special augmentation + ATSS (Baseline)	39.50	58.61	42.38
Copy-Move + ATSS	39.62 **(+0.12)**	58.67	42.58
no special augmentation + LTTSS	39.73 **(+0.23)**	58.86	42.62
Copy-Move + LTTSS	39.77 **(+0.27)**	58.84	42.84
Copy-Paste + ATSS	37.00 **(-2.50)**	55.40	39.70

References

1. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: International Conference on Learning Representations. (2018)
2. Fang, H.S., Sun, J., Wang, R., Gou, M., Li, Y.L., Lu, C.: Instaboost: Boosting instance segmentation via probability map guided copy-pasting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). (2019)
3. Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.Y., Cubuk, E.D., Le, Q.V., Zoph, B.: Simple copy-paste is a strong data augmentation method for instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2021) 2918–2928