– Supplementary material –
Exploring Classification Equilibrium in Long-Tailed Object Detection

1. Implementation details

Here we elaborate on the more detailed implementations for the experiments in ablation study.

Feature memory module. The memory size M is set to 80, which can store five latest instance features for each category. Thus, we initialize the memory features by randomly selecting 5 instances per class. In addition, an image may contain multiple instances of the same category (*e.g.*, > 20 instances) on LVIS. For simplicity and efficiency, we randomly select 5 instances to update the feature memory for each category that contains more than 5 instances in an image during the training.

Dataset-based EBL. For dataset-based EBL, we use the number of the training instances of each category to compute the loss margin as follow:

$$\delta_{yy'} = log(\frac{n_{y'}}{n_y}), \tag{1}$$

where n_y is the number of the training instances of category y. As the score-based EBL, we use a small value $\hat{n}_{C+1}$ to replace n_{C+1} (*i.e.*, the number of the training instances of background) to reduce the false positive cases. In our experiments, $\hat{n}_{C+1}$ is set to 1.

Dataset-based MFS. The bounding box generator and the feature memory module of dataset-based MFS are the same as those of score-based MFS, while the probabilistic sampler of dataset-based MFS uses the sampling probability based on the number of the training instances of each category. Namely, the sampling probability is computed as follow:

$$p_y = \frac{f(n_y)}{\sum_{y'} f(n_{y'})}, \tag{2}$$

where:

$$f(n_y) = \frac{1}{n_y}. \tag{3}$$

Repeat factor sampling (RFS). RFS over-samples the images containing the tail classes by increasing the sampling rate for these categories. As demonstrated in [1], we use the best setting of RFS that over-samples the categories that appear in less than 0.1% of the total images.

References

[1] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5356–5364, 2019.