AREA: Adaptive Reweighting via Effective Area for Long-Tailed Classification

Xiaohua Chen[1,2] Yucan Zhou[1*] Dayan Wu[1] Chule Yang[3]
Bo Li[1,2] Qinghua Hu[4] Weiping Wang[1,2]
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3] Defense Innovation Institute, Military Academy of Sciences [4]Tianjin University
{chenxiaohua, zhouyucan, wudayan, libo, wangweiping}@iie.ac.cn
yangchule@126.com huqinghua@tju.edu.cn

A. Appendix

A.1. N-Based Weight vs N^{eff}-Based Weight on CIFAR-100

We also draw the N^{eff}-based weights of the CIFAR-100-LT under different imbalance ratios $\lambda \in \{200, 50, 20, 10\}$. The results are shown in Figure 1 - 4, respectively. The "Original_weight" means N-based weights.

A.2. Effective Area on CIFAR-100-LT

We draw the effective area N^{eff} of the CIFAR-100-LT under different imbalance ratios $\lambda \in \{200, 100, 50, 20\}$. The results of CIFAR-100-LT are shown in Figure 5 - 8. The "Original" means the statistical number N. To explicitly exhibit the difference between N and N^{eff}, we take the results on CIFAR-100-LT with $\lambda = 100$ (shown in Figure 5) as an example, and obtain the following observations:

Firstly, the effective area can be larger, smaller, and equal to the original sample number. For example, the N of "Apple" are 500 while the N^{eff} is larger than the 616.28. For "Lion", the N^{eff} is 64.87, which is smaller than N . For the tail class, the N is near to the N^{eff}.

Secondly, the effective area distribution varies in different training epochs, i.e., N^{eff} can be adaptively adjusted. This is because as the training procedure proceeded, the features are optimized. Correspondingly, the correlations of samples are updated, resulting in the changeable effective area. Besides, N^{eff} is more flexible as it is a real number.

Thirdly, the category with more samples usually has a larger N^{eff} than N. For example, the "Apple" and "Cloud" are both head categories ($N > 100$). The N^{eff} of "Apple" and "Cloud" are larger than the N 116.28 (616.28-500) and 16.42 (187.42-171), respectively. We argue that it is because the category "Apple" has sufficient samples with rich diversity, such as more backgrounds, angles, colors, etc.

A.3. Effective Area on Balanced CIFAR-10/100

We use ResNet-32 as our backbone, which is trained by SGD with a momentum of 0.9 and a weight decay 1×10^{-4}. We train the model 160 epochs with batch size 64. The initial learning rate is 0.1, and the linear warm-up learning rate schedule is adopted. Besides, we decay the learning rate by 0.1 at the 80^{th} and 120^{th} epoch.

As shown in Figure 9, we draw N^{eff} with different epochs on balanced CIFAR-10 and CIFAR-100. Specifically, for balanced CIFAR-100, we draw the N^{eff} in epoch=180,190 and reorder the categories according to the size of N^{eff}. We can see that statistically balanced data sets may be imbalanced in the feature space.

*Corresponding author

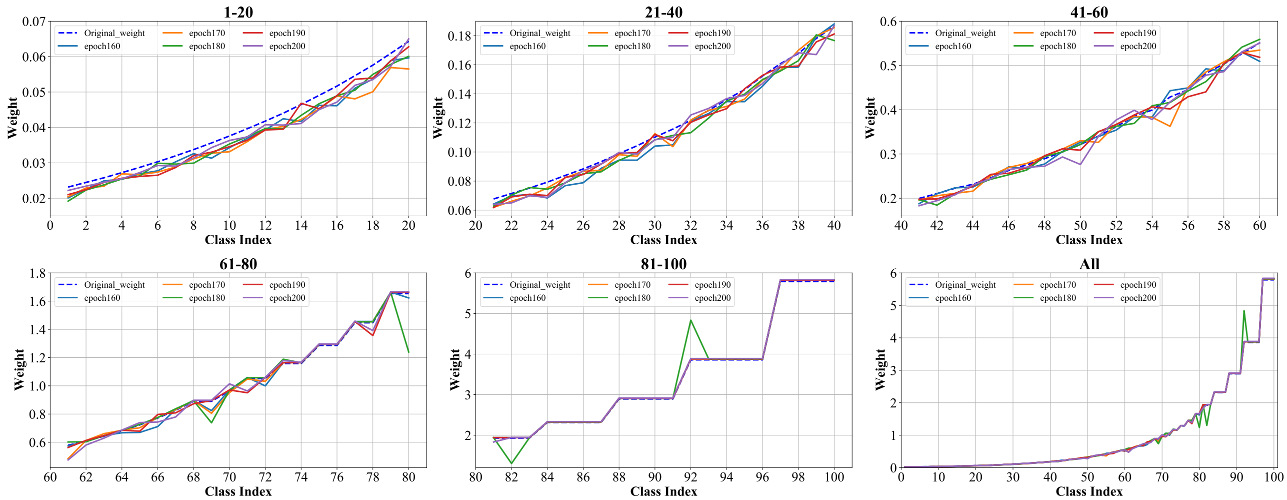

Figure 1. N-based weights vs N^{eff}-based weights on CIFAR-100-LT with λ=200.

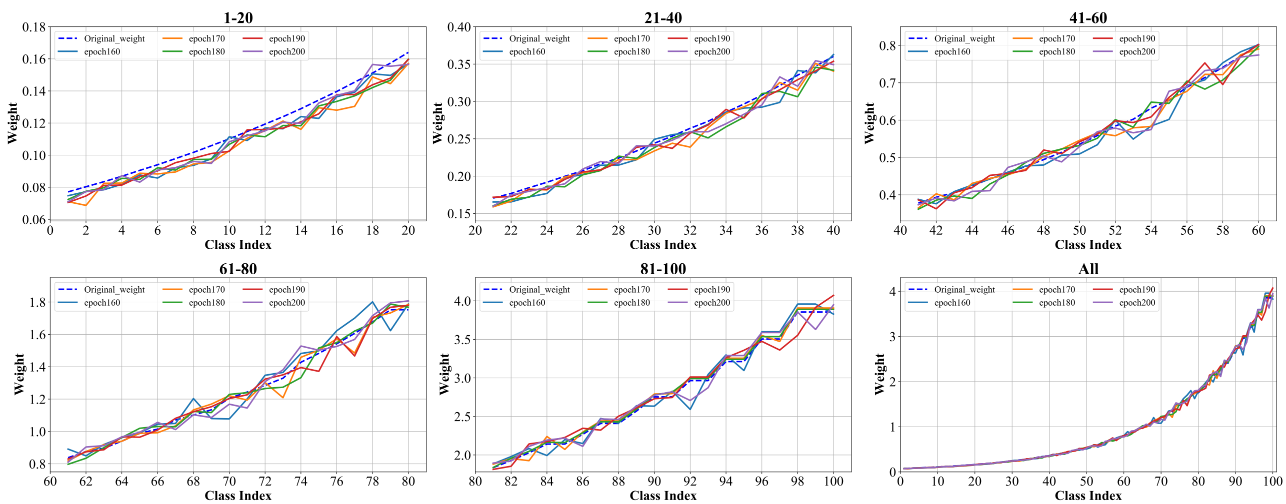

Figure 2. N-based weights vs N^{eff}-based weights on CIFAR-100-LT with λ=50.

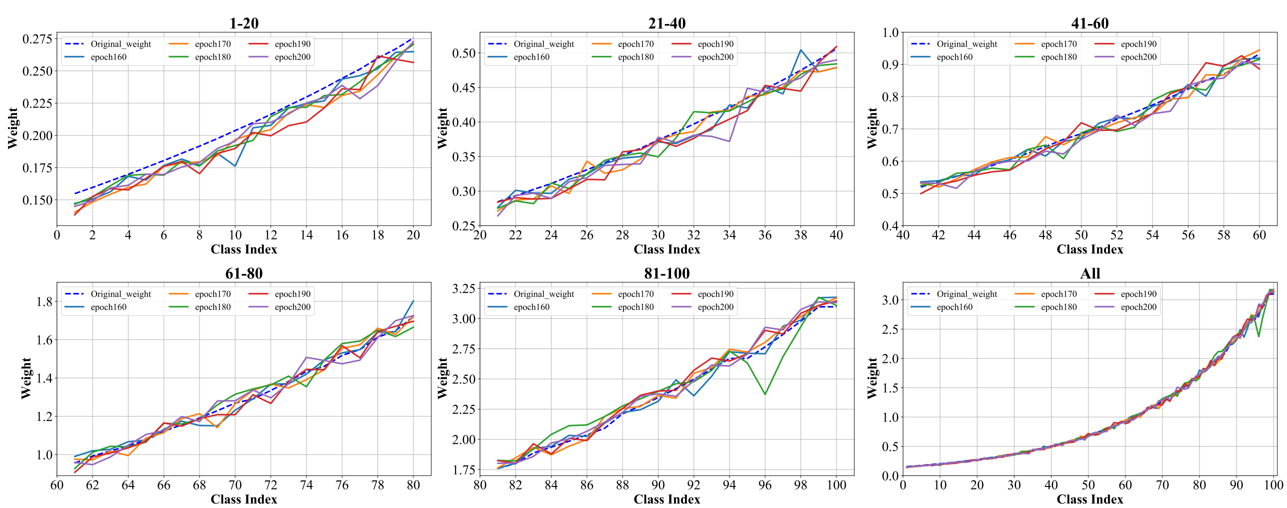

Figure 3. N-based weights vs N^{eff}-based weights on CIFAR-100-LT with λ=20.

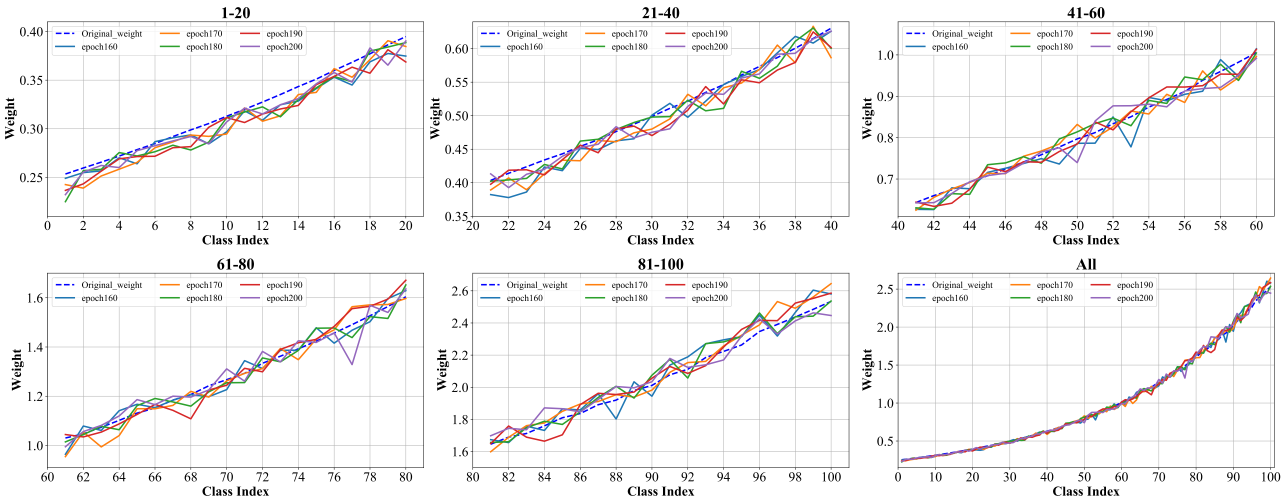

Figure 4. N-based weights vs N^{eff}-based weights on CIFAR-100-LT with λ=10.

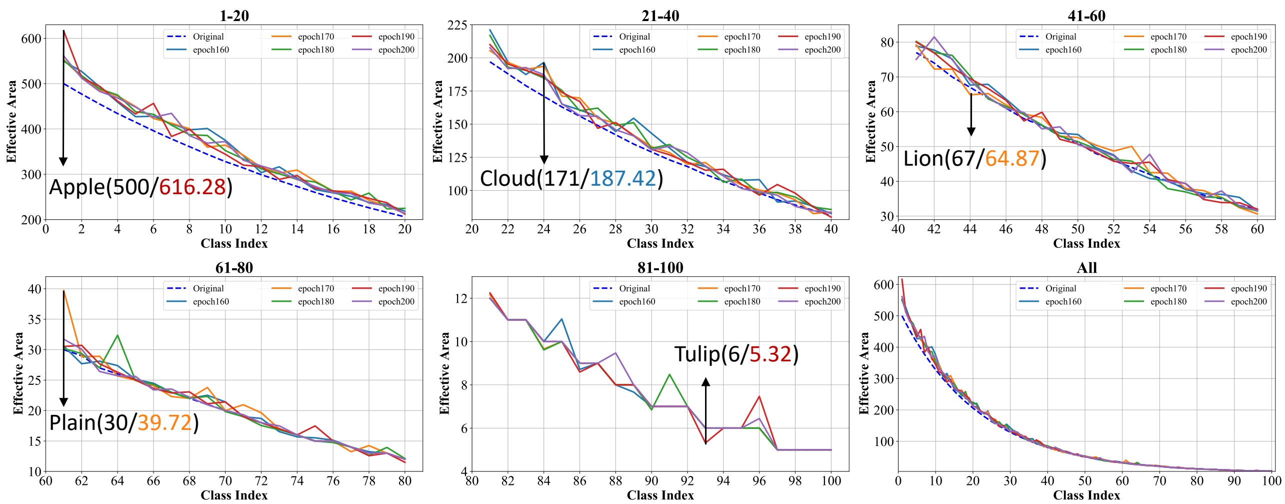

Figure 5. Effective area N^{eff} in different epochs on CIFAR-100-LT with λ=100.

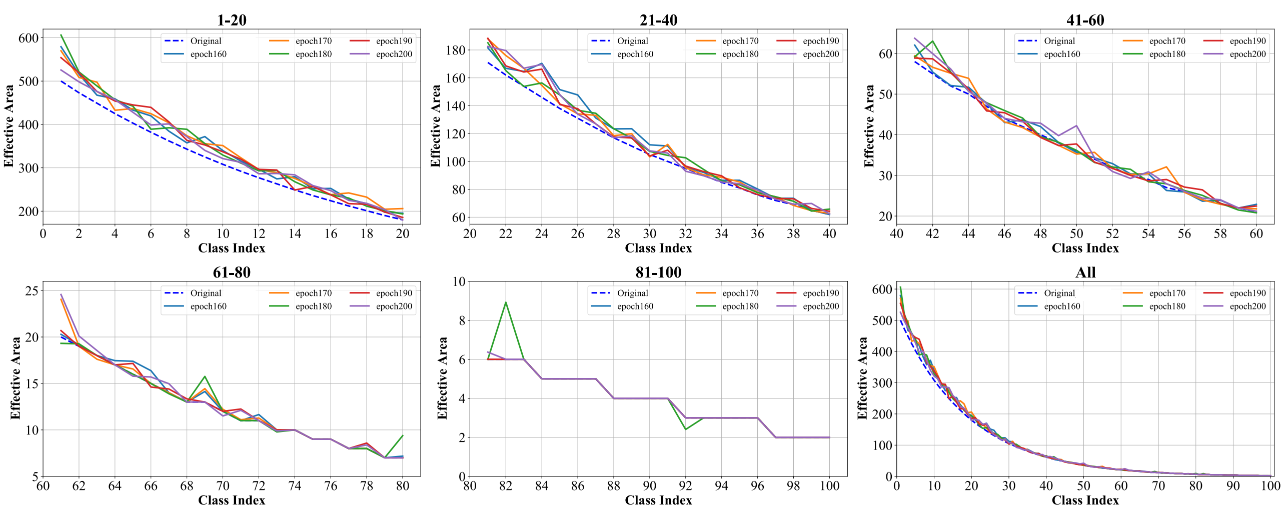

Figure 6. Effective area N^{eff} in different epochs on CIFAR-100-LT with λ=200.

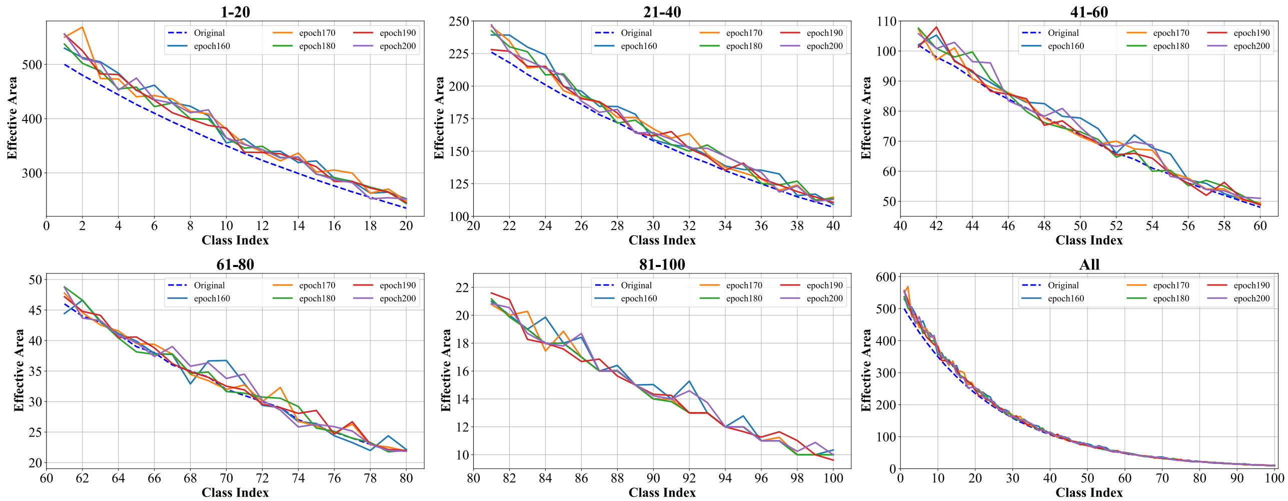

Figure 7. Effective area in different epochs on CIFAR-100-LT with λ=50.

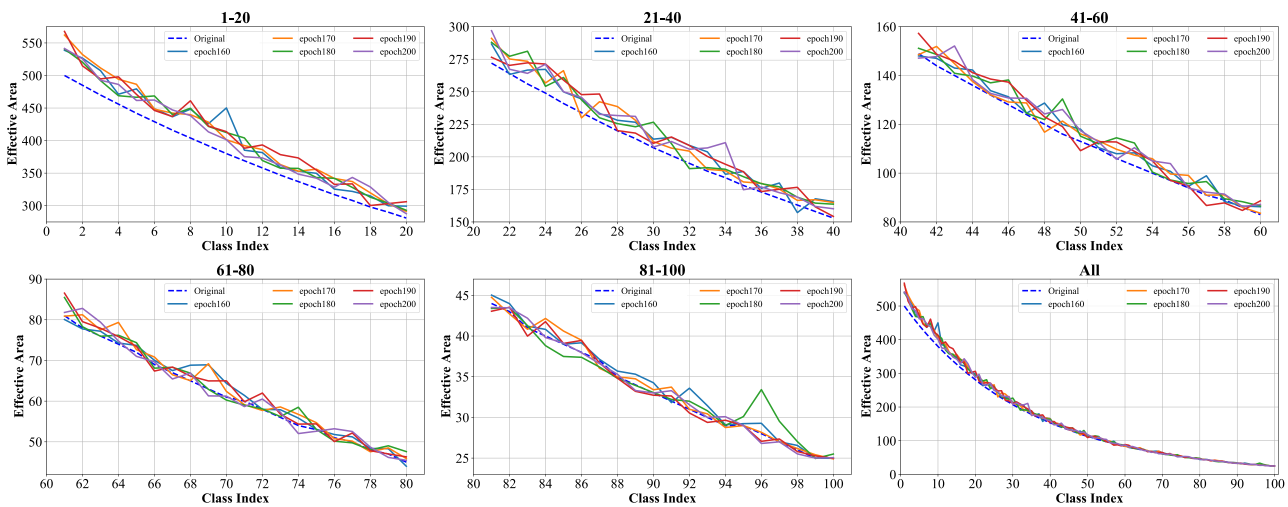

Figure 8. Effective area N^{eff} in different epochs on CIFAR-100-LT with λ=20.

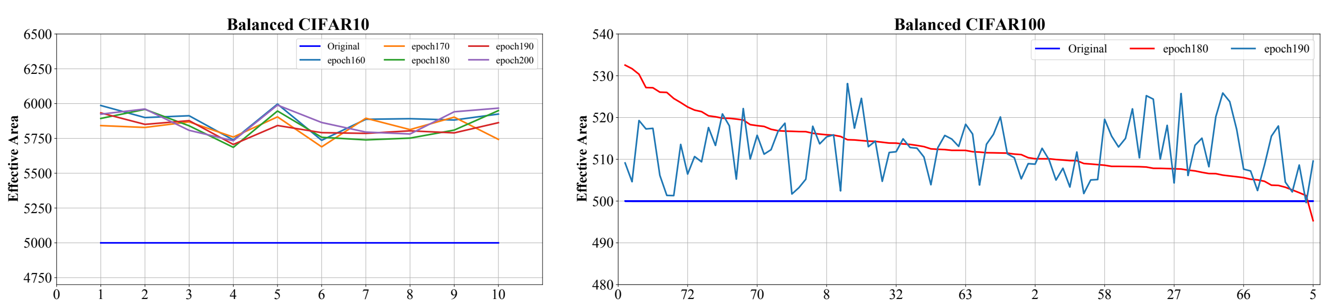

Figure 9. Effective area N^{eff} on balanced CIFAR-10 and CIFAR-100.