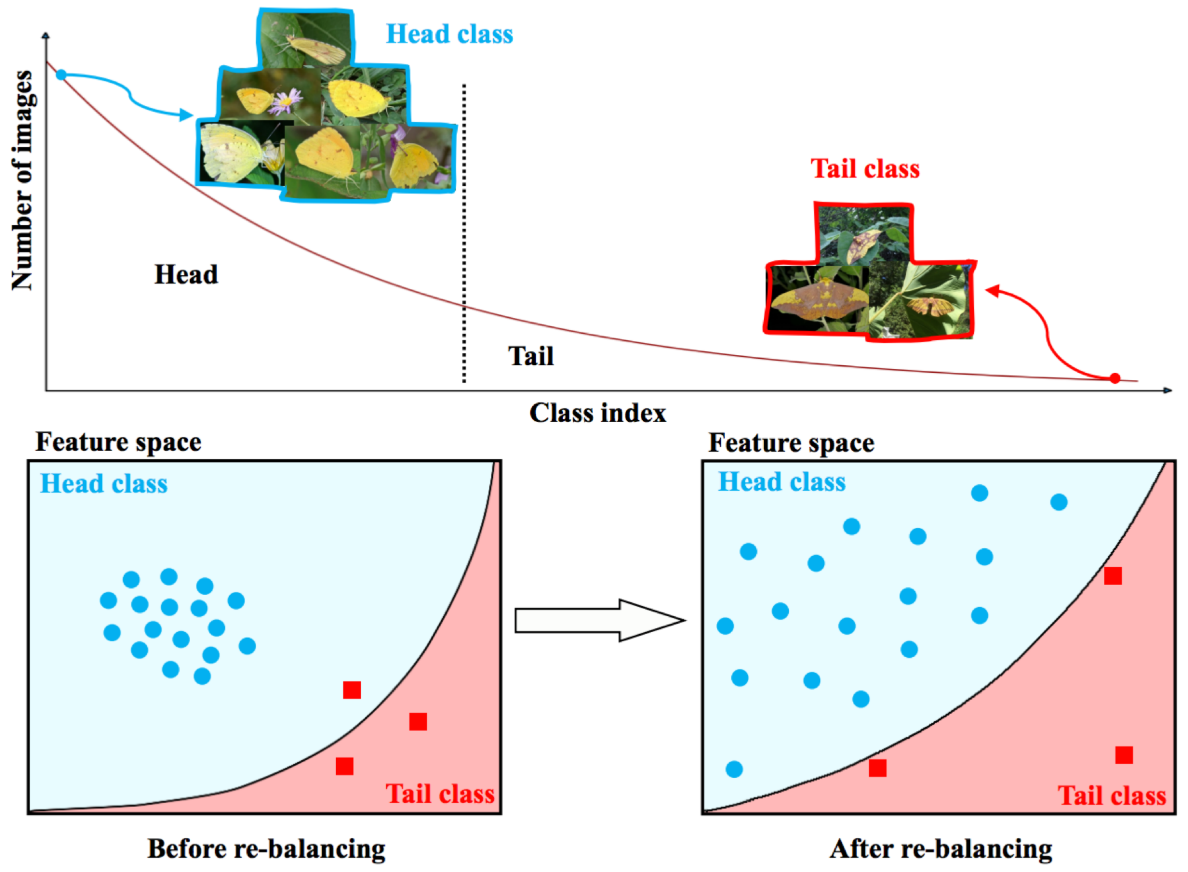
BBN: Bilateral-Branch Network with Cumulative Learning for Long-Tailed Visual Recognition

Boyan Zhou[1] Quan Cui[1,2] Xiu-Shen Wei[1]* Zhao-Min Chen[1,3]

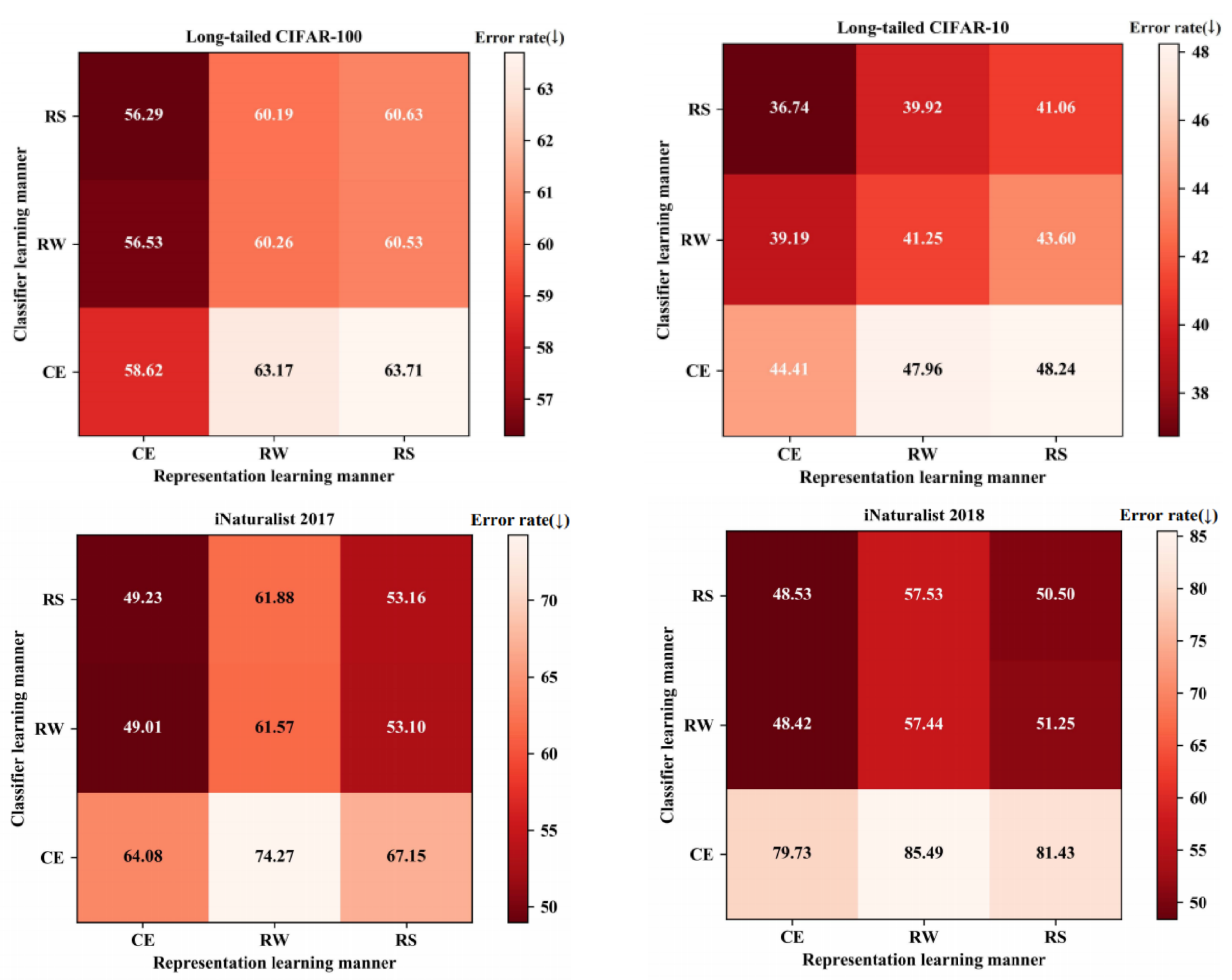
[1]Megvii Technology [2]Waseda University [3]Nanjing University

CVPR SEATTLE WASHINGTON JUNE 16-18 2020

Introduction & Observations



Our work focuses on tackling the visual recognition task of long-tailed data distribution. In this paper, we reveal the mechanism of re-balancing strategies is to significantly promote classifier learning but unexpectedly damage the representative ability of the deep features.



In this figure, "CE" (Cross-Entropy), "RW" (Re-Weighting) and "RS" (Re-Sampling) are the conducted learning manners.

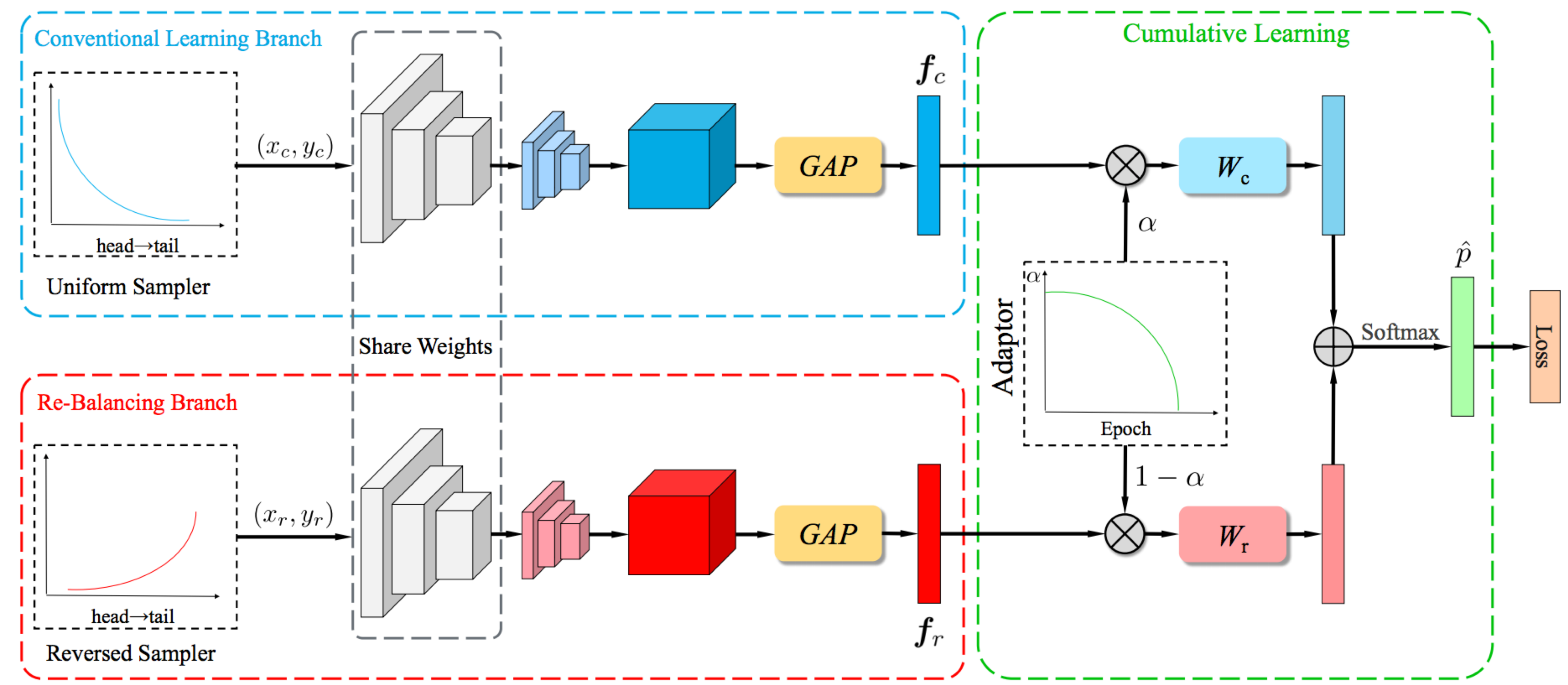
- ✔ When fixing the representation (comparing error rates of three blocks in the vertical direction), error rates of classifiers trained with RW/RS are reasonably lower than CE.
- ✔ When fixing the classifier (comparing error rates in the horizontal direction), the representations trained with CE surprisingly get lower error rates than those with RW/RS.

Proposed Bilateral-Branch Network & Cumulative Learning Strategy

Our BBN consists of three main components, i.e., **conventional learning branch**, **re-balancing branch** and **cumulative learning strategy.** Both branches use the same residual network structure and share all the weights except for the last residual block.

- ✔ Conventional Learning Branch:
  A uniform sampler is applied to obtain sample (x_c, y_c) as the input data from the original data distribution.
- ✔ Re-Balancing Branch:
  A reversed sampler is employed to acquire sample (x_r, y_r) according to the following manner:
  $$P_i = \frac{w_i}{\sum_{j=1}^{C} w_j},$$
  where P_i denotes the sampling possibility for the i-th class, $\omega_i = \frac{N_{max}}{N_i}$ and C is the number of classes
- ✔ Cumulative Learning Strategy:
  It is designed to first learn the universal patterns and then pay attention to the tail data gradually by controlling the weights for features produced by two branches and the classification loss:

$$\alpha = 1 - \left(\frac{T}{T_{max}}\right)^2 \qquad \begin{aligned} \mathbf{z} &= \alpha \boldsymbol{W}_c^\top \boldsymbol{f}_c + (1-\alpha)\boldsymbol{W}_r^\top \boldsymbol{f}_r \\ \mathcal{L} &= \alpha E(\hat{\boldsymbol{p}}, y_c) + (1-\alpha)E(\hat{\boldsymbol{p}}, y_r) \end{aligned}$$



Experiments & Visualization Results

Figure 1. Histogram of the measurement for the compactness of intra-class representations on the CIFAR-10-IR50 dataset.

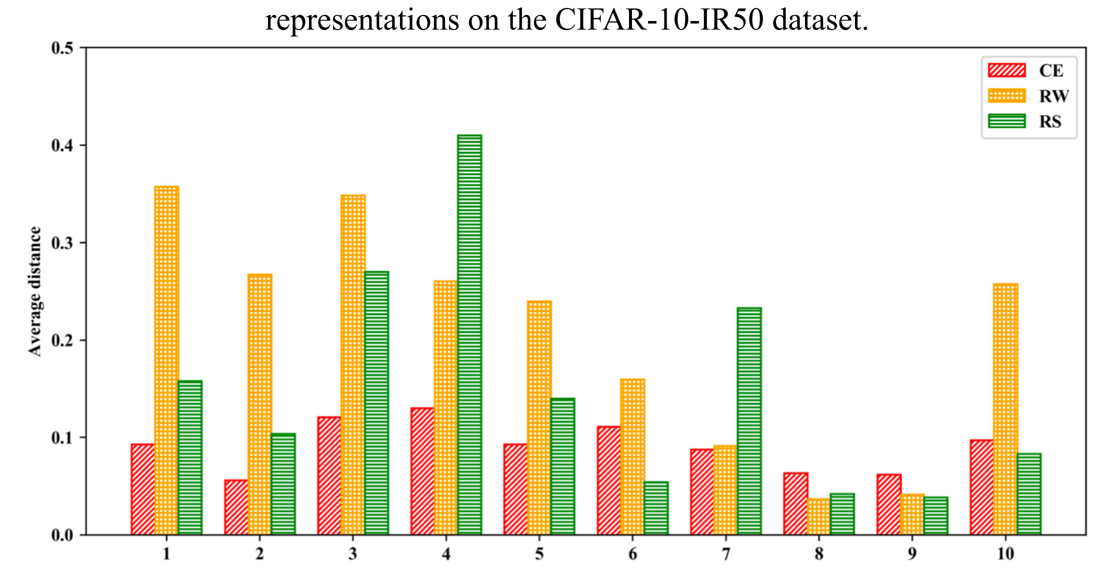


Figure 2. L2-norm of classifier weights for different learning manners.

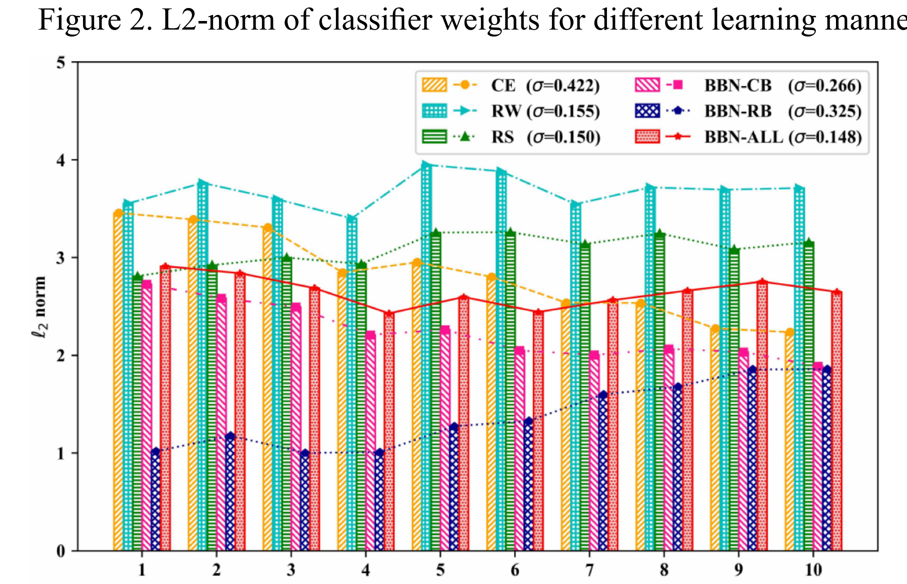


Table 1. Top-1 error rates on long-tailed CIFAR-10 and CIFAR-100.

Dataset	Long-tailed CIFAR-10			Long-tailed CIFAR-100		
Imbalance ratio	100	50	10	100	50	10
CE	29.64	25.19	13.61	61.68	56.15	44.29
Focal [17]	29.62	23.28	13.34	61.59	55.68	44.22
Mixup [35]	26.94	22.18	12.90	60.46	55.01	41.98
Manifold Mixup [29]	27.04	22.05	12.97	61.75	56.91	43.45
Manifold Mixup (two samplers)	26.90	20.79	13.17	63.19	57.95	43.54
CE-DRW [3]	23.66	20.03	12.44	58.49	54.71	41.88
CE-DRS [3]	24.39	20.19	12.62	58.39	54.52	41.89
CB-Focal [5]	25.43	20.73	12.90	60.40	54.83	42.01
LDAM-DRW [3]	22.97	18.97	11.84	57.96	53.38	41.29
Our BBN	**20.18**	**17.82**	**11.68**	**57.44**	**52.98**	**40.88**

Table 2. Top-1 error rates on large-scale long-tailed datasets.

Dataset	iNaturalist 2018	iNaturalist 2017
CE	42.84	45.38
CE-DRW [3]	36.27	40.48
CE-DRS [3]	36.44	40.12
CB-Focal [5]	38.88	41.92
LDAM-DRW* [3]	32.00	–
LDAM-DRW [3]	35.42	39.49
LDAM-DRW [3] (2×)	33.88	38.19
Our BBN	33.71	36.61
Our BBN (2×)	**30.38**	**34.25**

Contributions & Conclusions

- ✔ For studying long-tailed problems, we explored how class re-balancing strategies influenced representation learning and classifier learning of deep networks, and revealed that they can promote classifier learning significantly but also damage representation learning to some extent.
- ✔ Motivated by this, we proposed a Bilateral-Branch Network (BBN) with a specific cumulative learning strategy to take care of both representation learning and classifier learning for exhaustively improving the recognition performance of long-tailed tasks.
- ✔ By conducting extensive experiments, we proved that our BBN could outperform state-of-the-art results on long-tailed benchmarks, including the large-scale datasets iNaturalist17 and iNaturalist18.
- ✔ In the future, we attempt to tackle the long-tailed detection problems with our BBN model.