Class-Balanced Loss Based on Effective Number of Samples

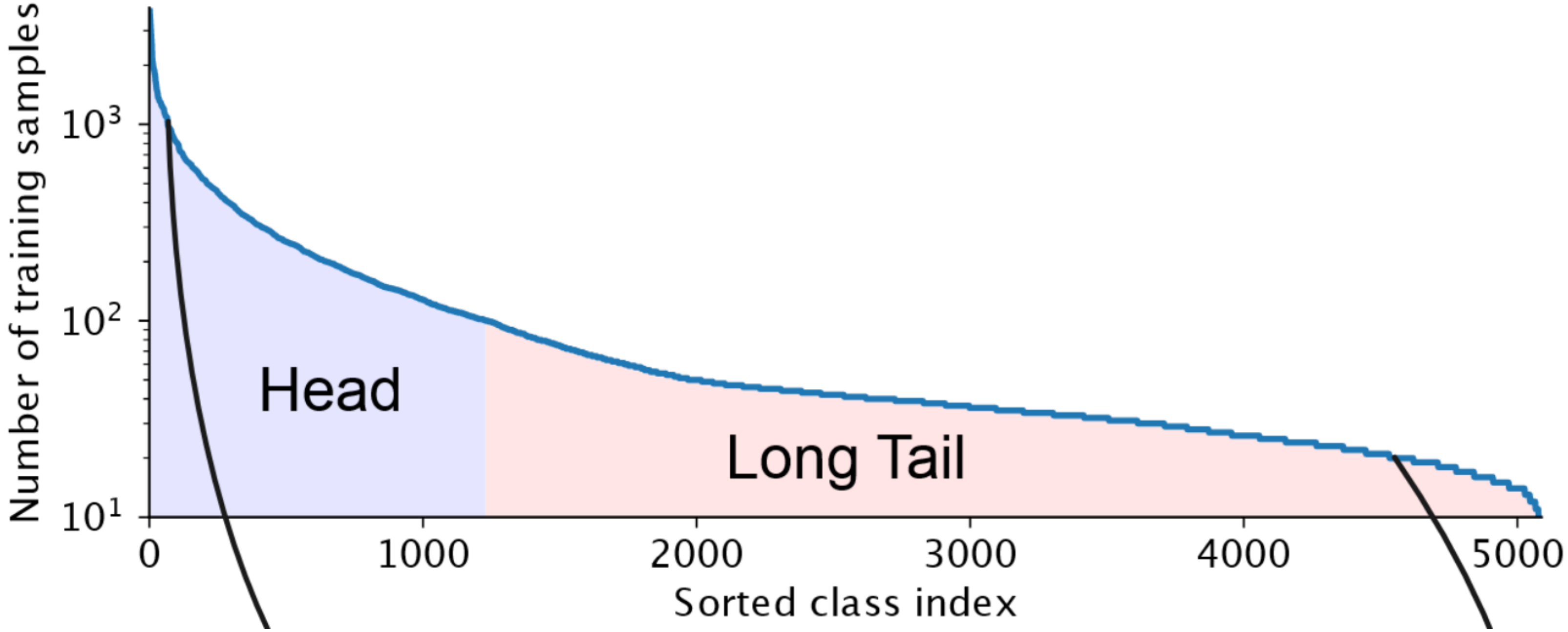
Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, Serge Belongie

CVPR 2019

2020.08.05 윤주열

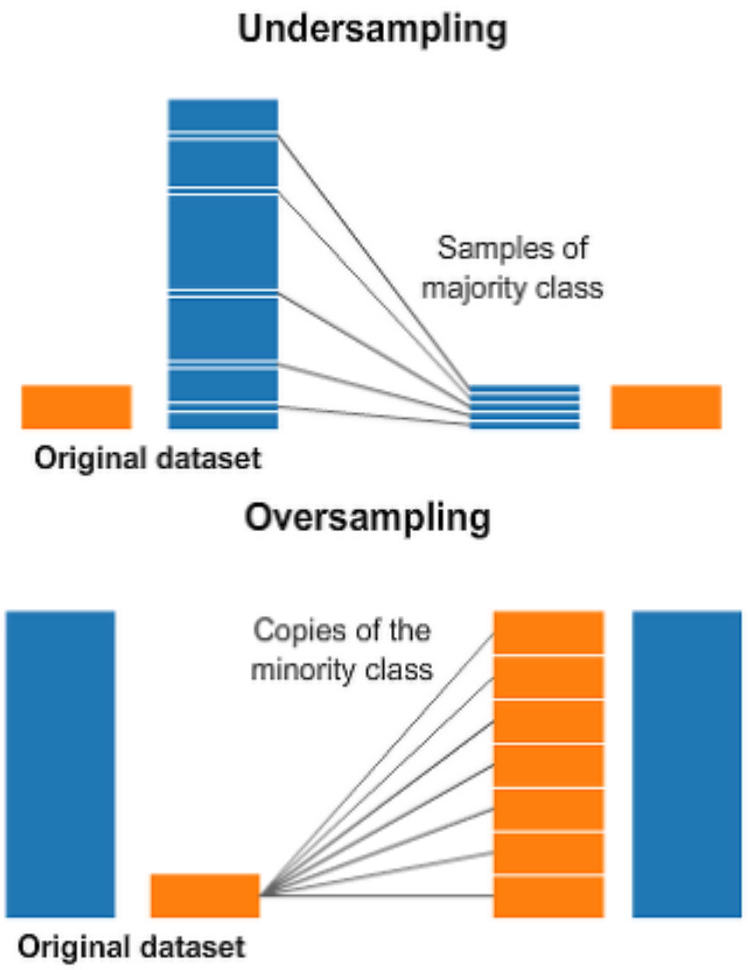
Introduction

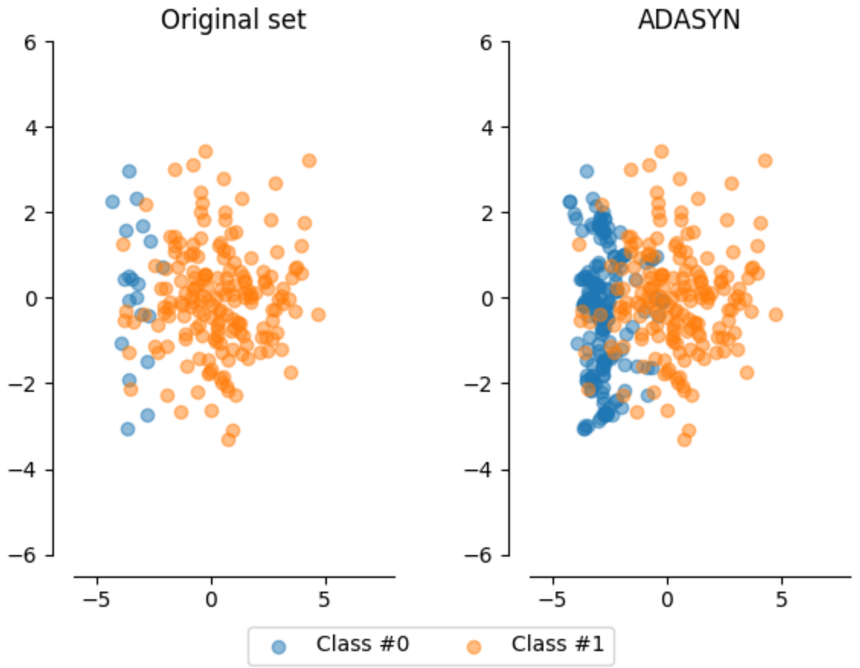
- Long tailed distributions in real world datasets harms performance for weakly represented classes.



Related Work

- Resampling





Related Work

- Cost-sensitive re-weighting (Class-balanced loss)

CROSSENTROPYLOSS

```
CLASS torch.nn.CrossEntropyLoss(weight: Optional[torch.Tensor] = None, size_average=None,
      ignore_index: int = -100, reduce=None, reduction: str = 'mean')
```

[SOURCE]

$$\text{loss}(x, class) = -weight[class] \log\left(\frac{\exp(x[class])}{\sum_j \exp(x[j])}\right)$$

Commonly used weights : Inverse frequency, Square rooted inverse frequency(“smoothed”)

- Focal Loss

$$\text{FL}(p_t) = -(1 - p_t)^\gamma \log(p_t).$$

Method

- Intuition: Given a training set, we can count the "effective number" of samples the set has.

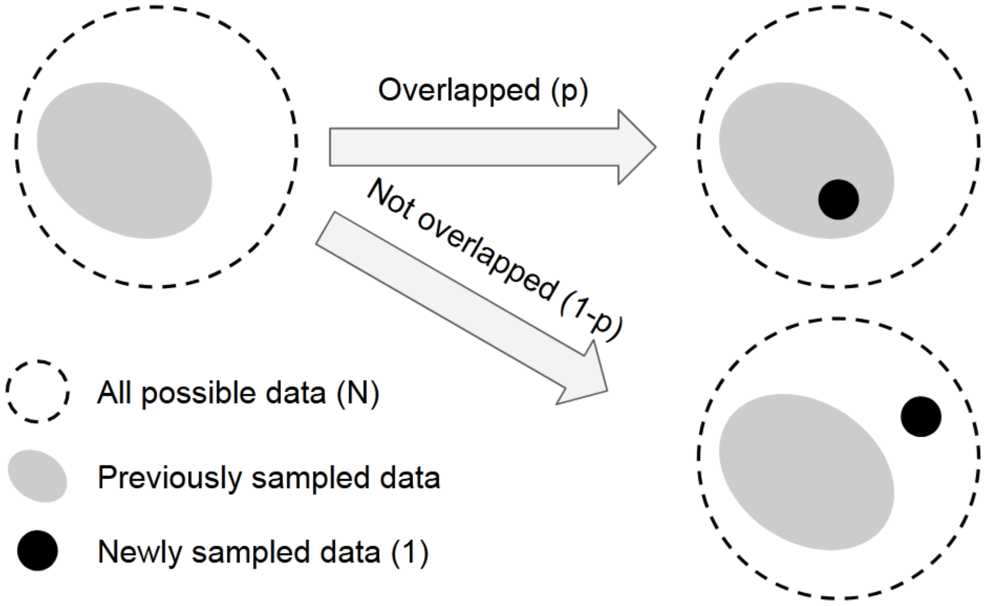


Figure 2. Giving the set of all possible data with volume N and the set of previously sampled data, a new sample with volume 1 has the probability of p being overlapped with previous data and the probability of $1-p$ not being overlapped.

Definition 1 (Effective Number)**.** The *effective number* of samples is the expected volume of samples.

Proposition 1 (Effective Number)**.** $E_n = (1-\beta^n)/(1-\beta)$, where $\beta = (N-1)/N$.

$$p = E_{n-1}/N$$

$$E_n = pE_{n-1}+(1-p)(E_{n-1}+1) = 1+\frac{N-1}{N}E_{n-1}. \quad (1)$$

Assume $E_{n-1} = (1-\beta^{n-1})/(1-\beta)$ holds, then

$$E_n = 1+\beta\frac{1-\beta^{n-1}}{1-\beta} = \frac{1-\beta+\beta-\beta^n}{1-\beta} = \frac{1-\beta^n}{1-\beta}. \quad (2)$$

$$N = \lim_{n\to\infty} E_n = 1/(1-\beta).$$

Method

- Class-Balanced Loss

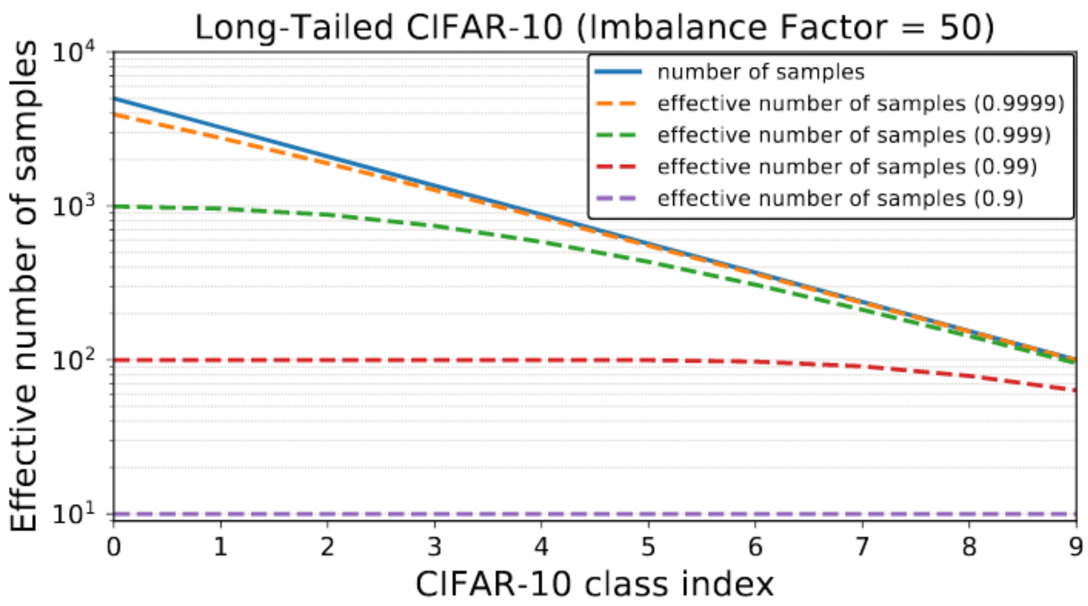
  use the effective number to give weights to different classes.

  applicable to different types of loss.

$$E_{n_i} = (1-\beta_i^{n_i})/(1-\beta_i), \text{ where } \beta_i = (N_i - 1)/N_i$$

$$\alpha_i \propto 1/E_{n_i} \qquad \sum_{i=1}^{C} \alpha_i = C$$

$$\mathrm{CB}(\mathbf{p}, y) = \frac{1}{E_{n_y}} \mathcal{L}(\mathbf{p}, y) = \frac{1-\beta}{1-\beta^{n_y}} \mathcal{L}(\mathbf{p}, y), \qquad (6)$$



Experiments

- Tested with three kinds of loss (Softmax Loss, Sigmoid CrossEntropy, Focal Loss)
- Tested with CIFAR, iNaturalist, ImageNet datasets.

Dataset Name	# Classes	Imbalance
Long-Tailed CIFAR-10	10	10.00 - 200.00
Long-Tailed CIFAR-100	100	10.00 - 200.00
iNaturalist 2017	5,089	435.44
iNaturalist 2018	8,142	500.00
ILSVRC 2012	1,000	1.78

Table 1. Datasets that are used to evaluate the effectiveness of class-balanced loss. We created 5 long-tailed versions of both CIFAR-10 and CIFAR-100 with imbalance factors of 10, 20, 50, 100 and 200 respectively.

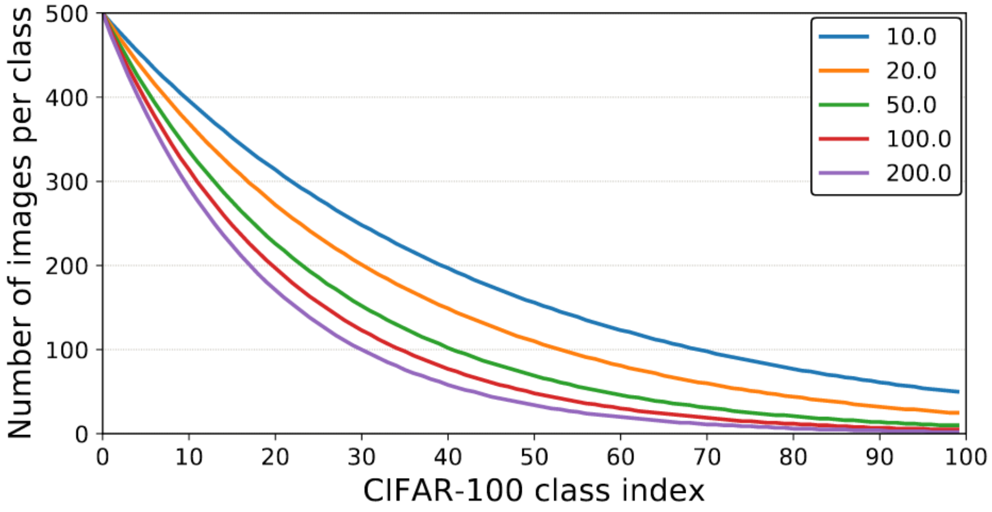


Figure 4. Number of training samples per class in artificially created long-tailed CIFAR-100 datasets with different imbalance factors.

Experiments

- Test results for CIFAR dataset with different imbalance severity.

Dataset Name	Long-Tailed CIFAR-10						Long-Tailed CIFAR-100					
Imbalance	200	100	50	20	10	1	200	100	50	20	10	1
Softmax	**34.32**	29.64	25.19	17.77	13.61	6.61	65.16	61.68	56.15	48.86	44.29	29.07
Sigmoid	34.51	**29.55**	23.84	**16.40**	**12.97**	**6.36**	64.39	**61.22**	55.85	48.57	44.73	**28.39**
Focal ($\gamma = 0.5$)	36.00	29.77	**23.28**	17.11	13.19	6.75	65.00	61.31	55.88	48.90	44.30	28.55
Focal ($\gamma = 1.0$)	34.71	29.62	23.29	17.24	13.34	6.60	**64.38**	61.59	**55.68**	**48.05**	**44.22**	28.85
Focal ($\gamma = 2.0$)	35.12	30.41	23.48	16.77	13.68	6.61	65.25	61.61	56.30	48.98	45.00	28.52
Class-Balanced	**31.11**	**25.43**	**20.73**	**15.64**	**12.51**	**6.36***	**63.77**	**60.40**	**54.68**	**47.41**	**42.01**	**28.39***
Loss Type	SM	Focal	Focal	SM	SGM	SGM	Focal	Focal	SGM	Focal	Focal	SGM
β	0.9999	0.9999	0.9999	0.9999	0.9999	-	0.9	0.9	0.99	0.99	0.999	-
γ	-	1.0	2.0	-	-	-	1.0	1.0	-	0.5	0.5	-

Table 2. Classification error rate of ResNet-32 trained with different loss functions on long-tailed CIFAR-10 and CIFAR-100. We show best results of class-balanced loss with best hyperparameters (SM represents Softmax and SGM represents Sigmoid) chosen via cross-validation. Class-balanced loss is able to achieve significant performance gains. $*$ denotes the case when each class has same number of samples, class-balanced term is always 1 therefore it reduces to the original loss function.

Experiments

- Test results for CIFAR dataset with different imbalance severity.

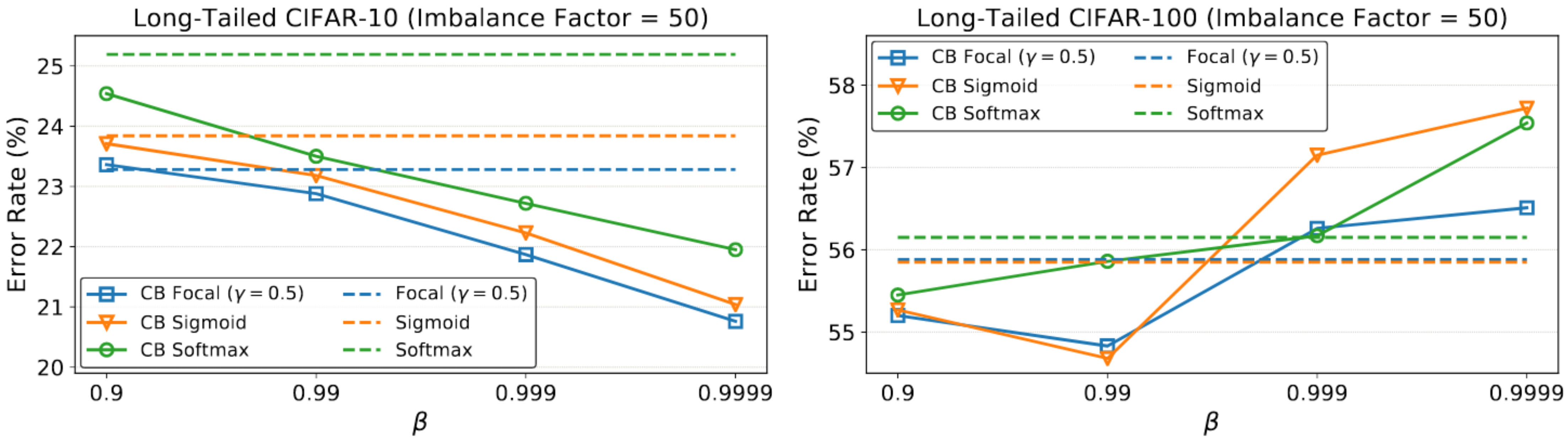


Figure 5. Classification error rate when trained with and without the class-balanced term. On CIFAR-10, class-balanced loss yields consistent improvement across different β and the larger the β is, the larger the improvement is. On CIFAR-100, $\beta = 0.99$ or $\beta = 0.999$ improves the original loss, whereas a larger β hurts the performance.

Experiments

- Test results for large-scale datasets.

					iNaturalist 2017		iNaturalist 2018		ILSVRC 2012	
Network	Loss	β	γ	Input Size	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
ResNet-50	Softmax	-	-	224×224	45.38	22.67	42.86	21.31	23.92	7.03
ResNet-101	Softmax	-	-	224×224	42.57	20.42	39.47	18.86	22.65	6.47
ResNet-152	Softmax	-	-	224×224	41.42	19.47	38.61	18.07	21.68	5.92
ResNet-50	CB Focal	0.999	0.5	224×224	41.92	20.92	38.88	18.97	22.71	6.72
ResNet-101	CB Focal	0.999	0.5	224×224	39.06	18.96	36.12	17.18	21.57	5.91
ResNet-152	CB Focal	0.999	0.5	224×224	38.06	18.42	35.21	16.34	20.87	5.61
ResNet-50	CB Focal	0.999	0.5	320×320	38.16	18.28	35.84	16.85	21.99	6.27
ResNet-101	CB Focal	0.999	0.5	320×320	34.96	15.90	32.02	14.27	20.25	5.34
ResNet-152	CB Focal	0.999	0.5	320×320	33.73	14.96	30.95	13.54	19.72	4.97

Table 3. Classification error rate on large-scale datasets trained with different loss functions. The proposed class-balanced term combined with focal loss (CB Focal) is able to outperform softmax cross-entropy by a large margin.