
Class-Balanced Loss Based on Effective Number of Samples

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

Cornell University, Cornell Tech, Google Brain, Alphabet Inc.

CVPR 2019

2020.05.14

최영제

1. Introduction

2. Related Works

3. Proposed methodology

4. Another method

1 Introduction

Background

- 분류 문제를 해결할 때 real data set은 long-tail distribution을 띄는 경향을 보임
→ class 별로 sample size가 다르며 소수 sample을 보유한 class가 많은 경우 long-tail의 모양을 띠
- Real data set에서 deep learning의 경우 class imbalance가 심한 경우 성능이 안좋은 경향을 보임

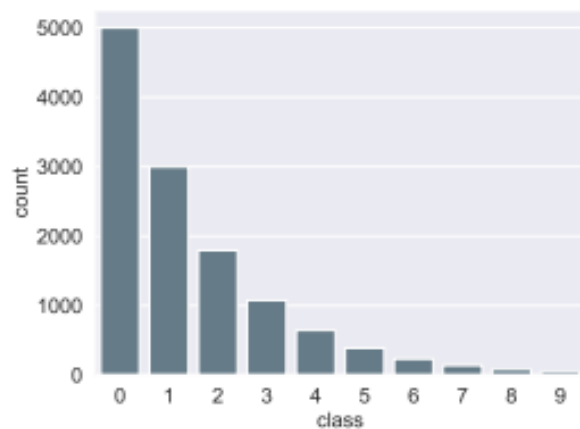


Figure 01 : Long-tail distribution example

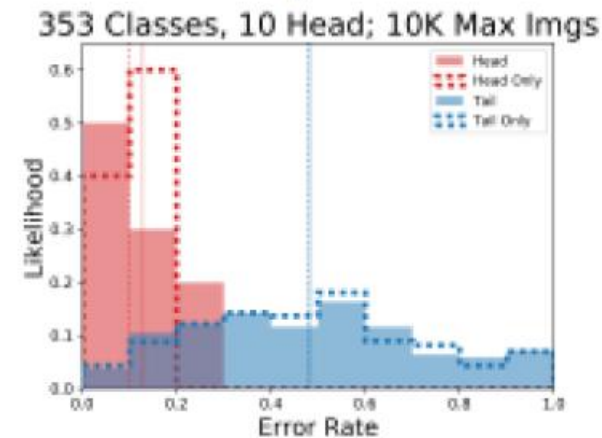
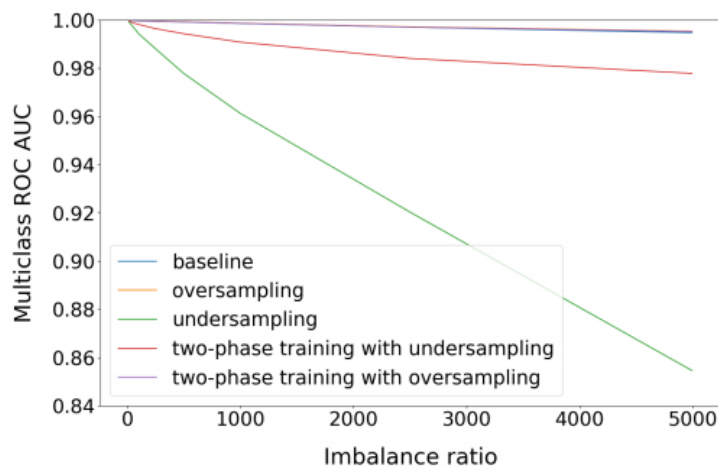


Figure 02 : error rate per class

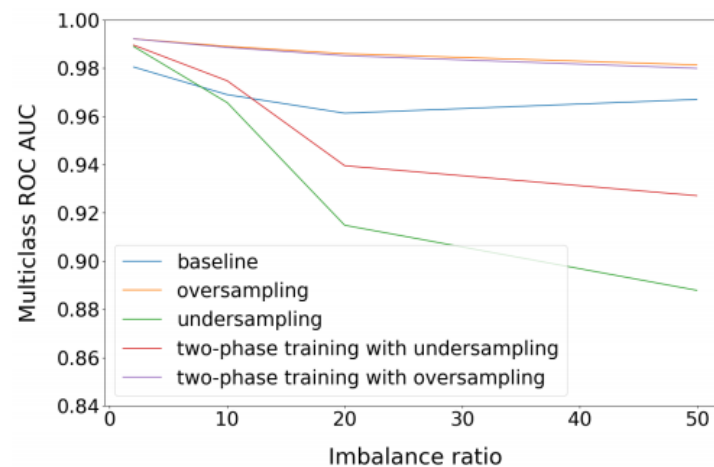
2 Related Works

Two strategies in class imbalanced problem

- Class imbalanced problem을 다루는 두 가지 방법은 re-sampling과 re-weighting(= cost-sensitive learning)이 존재.
- 먼저 re-sampling의 경우 over sampling method와 under sampling method가 있으며 통상적으로 over sampling method가 성능이 높음
→ Figure 03의 two-phase training이란 balanced data로 학습 후 output layer만 original data로 fine-tuning하는 것을 말함
- 정형 데이터의 경우 oversampling 기법으로 [SMOTE, ADASYN, SMOTE+Tomek](#) 등의 방법이 있음



(a) MNIST



(b) CIFAR-10

Figure 03 : model performances of different re sampling methods

2 Related Works

Two strategies in class imbalanced problem

- 또 다른 방법인 re-weighting(= cost-sensitive learning)은 data를 변형하는 것이 아닌 loss에 weighting을 주어 학습하는 것을 말함
- Cost-sensitive learning은 sample 별로 다른 weight를 주는 방법과 class 별로 다른 weight를 주는 경우로 나뉨
- 먼저 sample 별로 다른 error를 반영하는 loss function으로는 대표적으로 focal-loss가 존재함[1]
→ focal loss의 핵심은 쉬운(=confidence가 높은) sample일 수록 더 적은 loss를 반영하여 학습하는 것

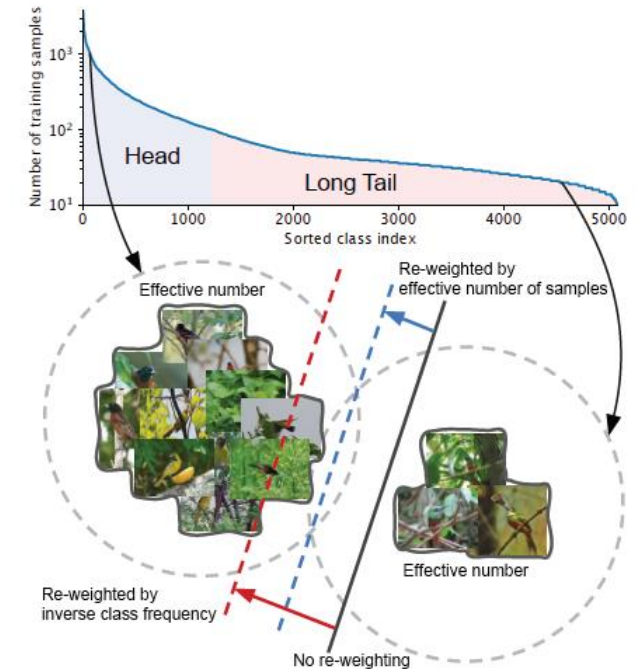
$$\text{CE}(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1 - p) & \text{otherwise.} \end{cases} \quad p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise,} \end{cases} \quad \text{CE}(p, y) = \text{CE}(p_t) = -\log(p_t). \quad \text{FL}(p_t) = -(1 - p_t)^\gamma \log(p_t).$$

- 다음으로 class 별 re-weighting method는 주로 sample size 비율의 역수를 weight로 삼아서 학습하는 방법을 사용하였음
→ if proportion per class = [0.2, 0.8], then weight per class = [5, 1.25]
→ majority class와 minority class의 error를 다르게 반영하여 불균형을 완화시키는 것이 목표

Two strategies in class imbalanced problem

- 그러나 앞서 말한 방법들은 모두 단점을 갖고 있음
- re-sampling에서 under sampling의 경우 data의 손실 발생, oversampling의 경우 sample이 커져 time cost 및 overfitting 가능성 증가
- class별 re-weighting method에서 단순 proportion의 역수로 weight를 줄 경우 model의 majority class의 분류 능력을 감소시킬 가능성 존재
→ 그 이유는 weight가 sample size에 따라 선형적으로 변하기 때문
- 따라서 class별로 다른 weight를 주되 단순 sample size가 아닌 effective number에 기반하여 weight를 반영하는 것이 논문의 핵심

What is effective number?



3_{SIL} Proposed methodology

Effective number of samples

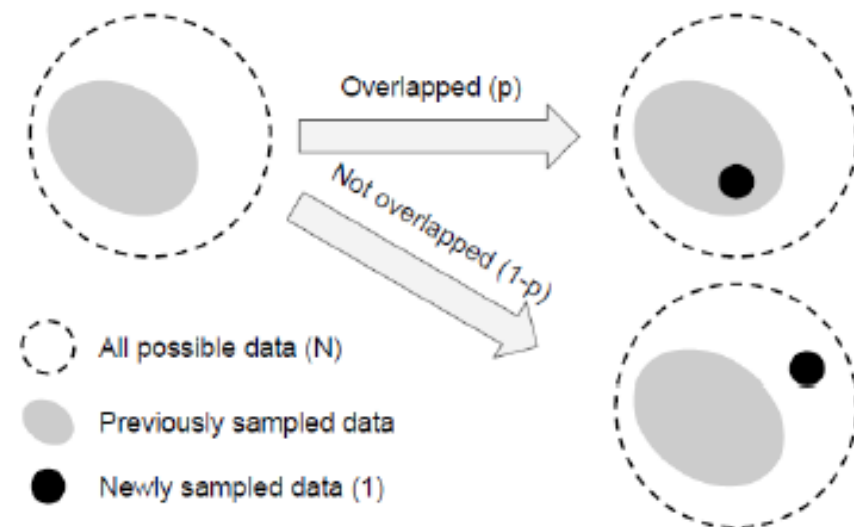
- Data sample의 effective number 란 N개의 data가 있을 때 중복, 유사한 sample을 제외한 영향력 있는 sample들의 개수를 뜻함
→ 간단한 예로, 1부터 10까지 5번 복원 추출을 진행했을 때 [1, 8, 2, 9, 3]이 추출될 경우 sample size와 effective number는 모두 5
→ 반면 [1, 1, 2, 2, 3]이 추출될 경우 effective number는 3의 값을 가짐
- Random covering 개념을 적용하면 전체 데이터(N)에서 n번째 sample을 추출할 경우 p의 확률로 sampled data와 중복(=coverage 안에 포함)되며 또한 (1-p)의 확률로 새로운 sample이 추출됨
- Effective number는 이러한 새로운 sample이 나올 경우 counting함

$$E_n = pE_{n-1} + (1 - p)(E_{n-1} + 1)$$

- 여기서 p의 값은 n-1번째 effective number를 전체 sample(N)로 나눈 값

$$p = \frac{E_{n-1}}{N}$$

→ 즉 위의 예제에서 6번째 sample을 뽑을 때 확률 p는 $3/10 = 0.3$



3_{SIL} Proposed methodology

Class balanced loss

- Effective number 공식에 p 를 대입하여 정리하며 아래와 같음

$$E_n = 1 + \frac{N-1}{N} E_{n-1} \quad {}^*p = \frac{E_{n-1}}{N}$$

- 여기서 $(N-1)/N$ 을 β 로 치환하고 논문의 가정을 활용하면 Effective number는 다음과 같이 정리됨

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

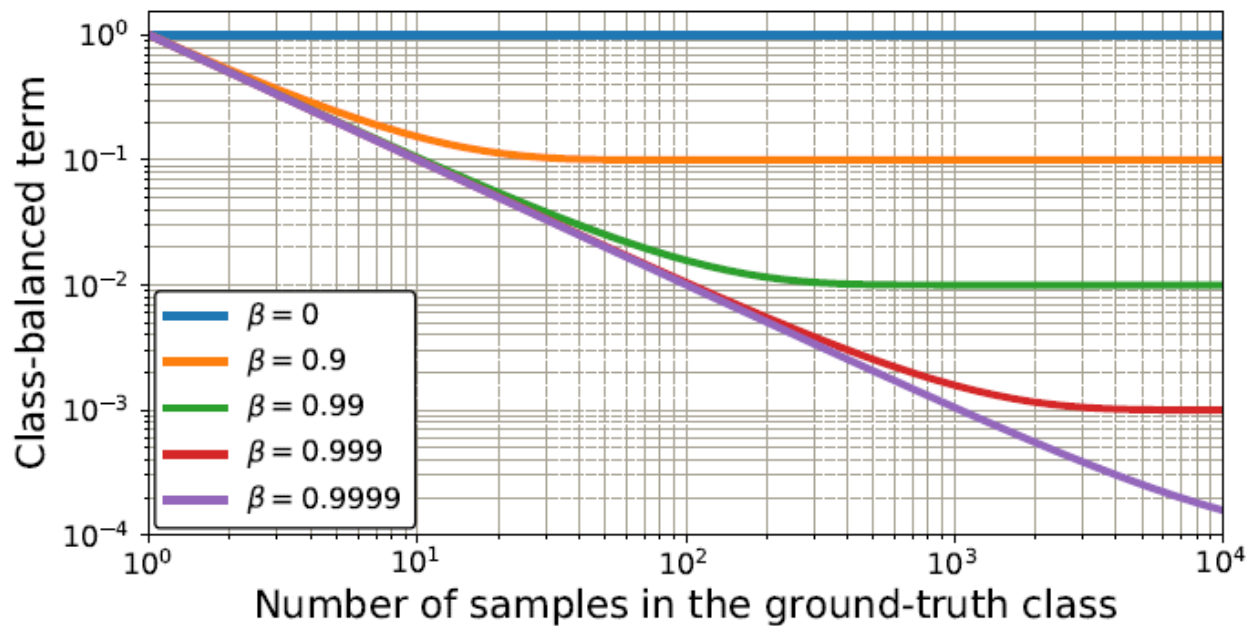
$$E_n = 1 + \beta \frac{1 - \beta^{n-1}}{1 - \beta} = \frac{1 - \beta + \beta - \beta^n}{1 - \beta} = \frac{1 - \beta^n}{1 - \beta}$$

- 구해진 effective number는 sample의 개수와 비례하기 때문에 effective number의 역수를 class balance term으로 사용
- 최종적으로 논문에서 제안하는 class balanced loss는 다음과 같음

$$CB(p, y) = \frac{1}{E_{n_y}} \mathcal{L}(p, y) = \frac{1 - \beta}{1 - \beta^{n_y}} \mathcal{L}(p, y) \quad * \mathcal{L}(p, y) = \text{softmax, bce, focal loss etc}$$

Class balanced loss

- CB loss를 사용하면 class proportion의 역수를 사용할 때 선형적으로 변하는 weight 문제를 해결할 수 있는가?
→ β 값에 따라 sample 숫자가 일정 수준 이상이면 더 이상 weight은 감소하지 않음



Experiments

- 5가지 data set을 사용하였으며 CIFAR-10, 100의 경우 인위적으로 imbalance data를 만들어 학습 (test는 기존과 동일)
→ imbalance factor는 largest class를 smallest class로 나눈 값

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

- CIFAR-10, 100 results table

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	-

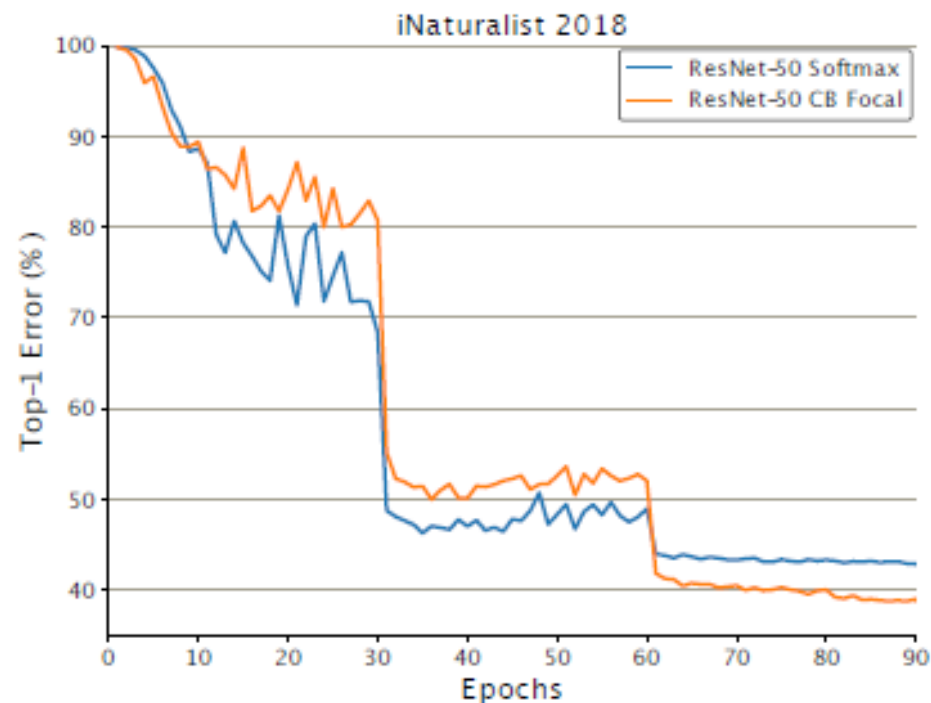
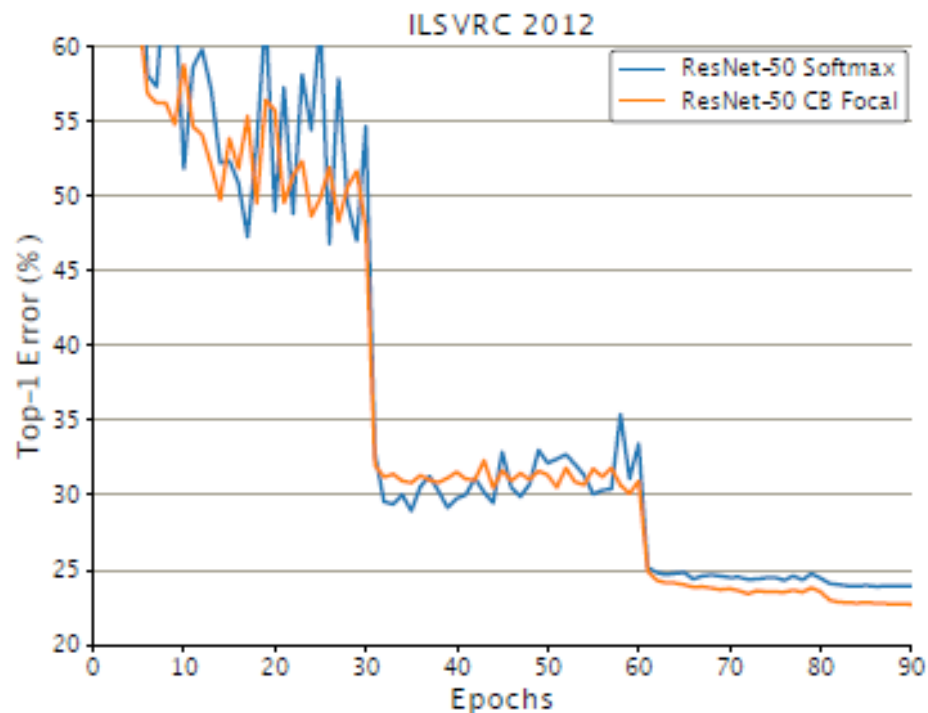
Experiments

- iNaturalist2017, 2018, ImageNet results table

					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

Experiments

- Loss plot comparison per epochs



Learning imbalanced datasets with label distribution aware margin loss

- LDAM-loss[2], NIPS 2019

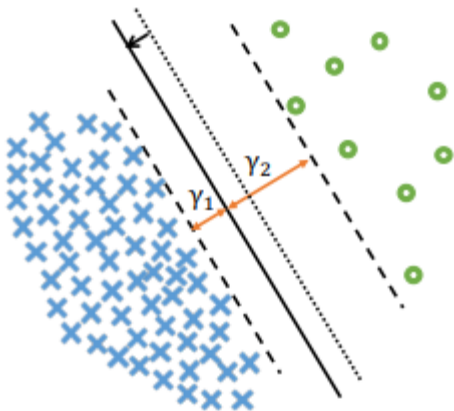


Figure 1: For binary classification with a linearly separable classifier, the margin γ_i of the i -th class is defined to be the minimum distance of the data in the i -th class to the decision boundary. We show that the test error with the uniform label distribution is bounded by a quantity that scales in $\frac{1}{\gamma_1 \sqrt{n_1}} + \frac{1}{\gamma_2 \sqrt{n_2}}$. As illustrated here, fixing the direction of the decision boundary leads to a fixed $\gamma_1 + \gamma_2$, but the trade-off between γ_1, γ_2 can be optimized by shifting the decision boundary. As derived in Section 3.1, the optimal trade-off is $\gamma_i \propto n_i^{-1/4}$ where n_i is the sample size of the i -th class.

Dataset	Imbalanced CIFAR-10				Imbalanced CIFAR-100			
Imbalance Type	long-tailed		step		long-tailed		step	
Imbalance Ratio	100	10	100	10	100	10	100	10
ERM	29.64	13.61	36.70	17.50	61.68	44.30	61.45	45.37
Focal [Lin et al., 2017]	29.62	13.34	36.09	16.36	61.59	44.22	61.43	46.54
LDAM	26.65	13.04	33.42	15.00	60.40	43.09	60.42	43.73
CB RS	29.45	13.21	38.14	15.41	66.56	44.94	66.23	46.92
CB RW [Cui et al., 2019]	27.63	13.46	38.06	16.20	66.01	42.88	78.69	47.52
CB Focal [Cui et al., 2019]	25.43	12.90	39.73	16.54	63.98	42.01	80.24	49.98
HG-DRS	27.16	14.03	29.93	14.85	-	-	-	-
LDAM-HG-DRS	24.42	12.72	24.53	12.82	-	-	-	-
M-DRW	24.94	13.57	27.67	13.17	59.49	43.78	58.91	44.72
LDAM-DRW	22.97	11.84	23.08	12.19	57.96	41.29	54.64	40.54

- [1] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).
- [2] Cao, K., Wei, C., Gaidon, A., Arechiga, N., & Ma, T. (2019). Learning imbalanced datasets with label-distribution-aware margin loss. In Advances in Neural Information Processing Systems (pp. 1565-1576).
- Figure 01 : Van Horn, G., & Perona, P. (2017). The devil is in the tails: Fine-grained classification in the wild. arXiv preprint arXiv:1709.01450.
- Figure 02 : Cao, K., Wei, C., Gaidon, A., Arechiga, N., & Ma, T. (2019). Learning imbalanced datasets with label-distribution-aware margin loss. In Advances in Neural Information Processing Systems (pp. 1565-1576).
- Figure 03 : Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks, 106, 249-259.

Q&A