Decoupling Representation and Classifier for Long-Tailed Recognition

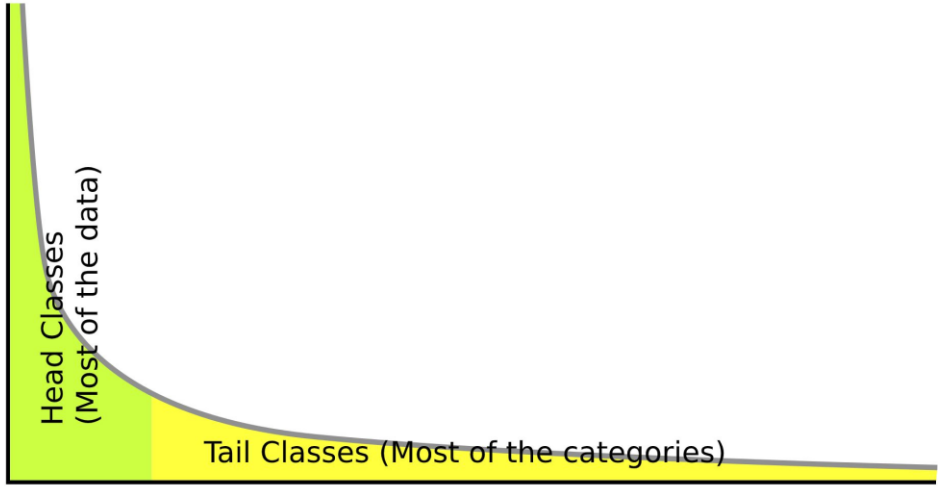
Bingyi Kang1, Saining Xie , Marcus Rohrbach , Zhicheng Yan , Albert Gordo , Jiashi Feng , Yannis Kalantidis

ICLR 2020

Background

What is long-tailed data ?

The danger of long-tailed data



The basic solutions to long-tailed data

- **re-sampling**
- **re-weighting**
- **transfer learning**

Motivation

The ambiguity of aforementioned methods （joint learning）

It **unclear** how the long-tailed recognition ability is achieved—is it from learning **a better representation or better classifier** decision boundaries?

Joint learning means learning feature representation and classifier at the same time

Decouple Representation and Classifier

To answer this question, the author decouple long-tail recognition into representation learning and classification.

Benefits

- Decoupling representation learning and classification find that : instance-balanced sampling learns the best and most generalizable representations.

- With good representations learned, it is also possible to achieve strong long-tailed recognition ability by adjusting only the classifier.

- The work achieve significantly higher accuracy than well established state-of-the-art methods.

Representation learning

1. Instance-balanced sampling， q=1, IB

Each sample has an **equal** probability of being selected

$$p_j = \frac{n_j^q}{\sum_{i=1}^{C} n_i^q},$$

2. Class-balanced sampling， q=0, CB

Each class is selected **equally**, and then samples are selected from the classes

3. Square-root sampling， q = ½

It's essentially a **variation** on the **two previous** sampling methods

Representation learning

4. Progressively-balanced sampling, PB

$$p_j^{\mathrm{PB}}(t) = (1 - \frac{t}{T})p_j^{\mathrm{IB}} + \frac{t}{T}p_j^{\mathrm{CB}},$$

According to the number of iterations (epoch) in training, PB is **based on sample equalization** (IB) and **category equalization** (CB) sampling

Classifier learning

1. Classifier Re-training, (CRT)

keeping the **representations fixed**, randomly re-initialize and **optimize the classifier** weights for a small number of epochs using **class-balanced sampling**.

2. Nearest Class Mean classifier,(NCM)

It first calculates the mean of the learned features for each category, and then performs a **nearest neighbor search** to determine the category.

Classifier learning

3. τ -normalized classifier (τ -normalized)

Re-normalize the category boundary in the classifier to achieve equilibrium

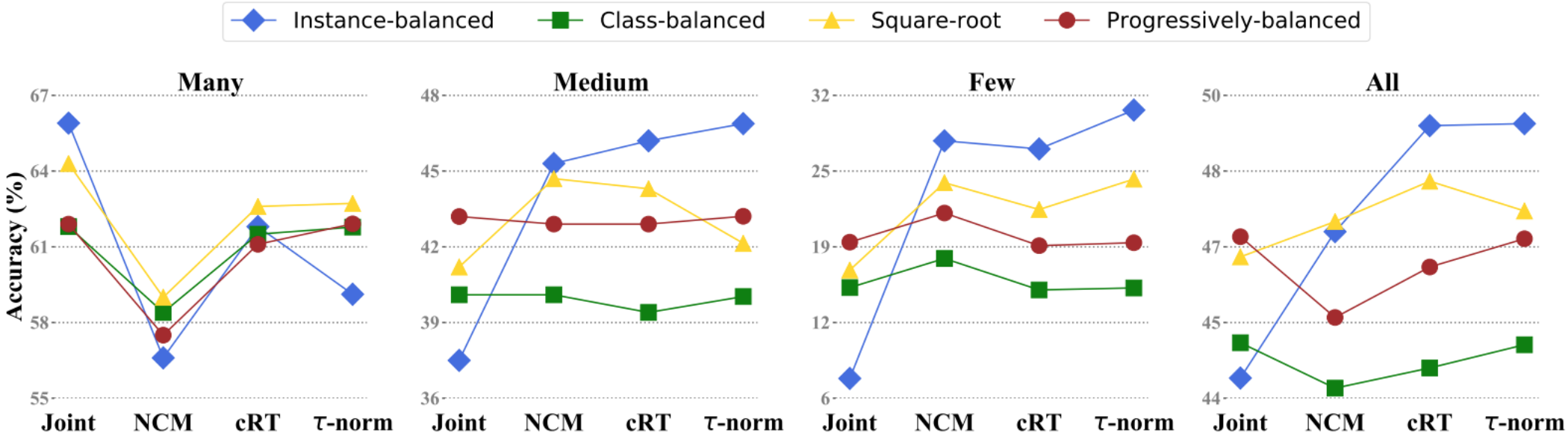
$$\tilde{w}_i = \frac{w_i}{||w_i||^\tau}$$

4. Learnable weight scaling (LWS)

Fixed classification weight Wi, a weighted parameter Fi is learned for each class

$$\widetilde{w_i} = f_i * w_i, \text{where } f_i = \frac{1}{||w_i||^\tau}.$$

Experiments



Experiments

Table 2: Long-tail recognition accuracy on ImageNet-LT for different backbone architectures. † denotes results directly copied from Liu et al. (2019). * denotes results reproduced with the authors' code. ** denotes OLTR with our representation learning stage.

Method	ResNet-10	ResNeXt-50	ResNeXt-152
FSLwF† (Gidaris & Komodakis, 2018)	28.4	-	-
Focal Loss† (Lin et al., 2017)	30.5	-	-
Range Loss† (Zhang et al., 2017)	30.7	-	-
Lifted Loss† (Oh Song et al., 2016)	30.8	-	-
OLTR† (Liu et al., 2019)	35.6	-	-
OLTR*	34.1	37.7	24.8
OLTR**	37.3	46.3	50.3
Joint	34.8	44.4	47.8
NCM	35.5	47.3	51.3
cRT	**41.8**	**49.5**	52.4
τ-normalized	40.6	**49.4**	**52.8**
LWS	**41.4**	**49.9**	**53.3**

Experiments

Table 3: Overall accuracy on iNaturalist 2018. Rows with † denote results directly copied from Cao et al. (2019). We present results when training for 90/200 epochs.

Method	ResNet-50	ResNet-152
CB-Focal†	61.1	-
LDAM†	64.6	-
LDAM+DRW†	68.0	-
Joint	61.7/65.8	65.0/69.0
NCM	58.2/63.1	61.9/67.3
cRT	65.2/67.6	68.5/71.2
τ-normalized	65.6/**69.3**	68.8/**72.5**
LWS	65.9/**69.5**	69.1/72.1

Table 4: Results on Places-LT, starting from an ImageNet pre-trained ResNet152. † denotes results directly copied from Liu et al. (2019).

Method	Many	Medium	Few	All
Lifted Loss†	41.1	35.4	24.0	35.2
Focal Loss†	41.1	34.8	22.4	34.6
Range Loss†	41.1	35.4	23.2	35.1
FSLwF†	43.9	29.9	29.5	34.9
OLTR†	44.7	37.0	25.3	35.9
Joint	**45.7**	27.3	8.2	30.2
NCM	40.4	37.1	27.3	36.4
cRT	42.0	37.6	24.9	36.7
τ-normalized	37.8	**40.7**	**31.8**	**37.9**
LWS	40.6	39.1	28.6	37.6

Thanks