南京航空航天大學
Nanjing University of Aeronautics and Astronautics

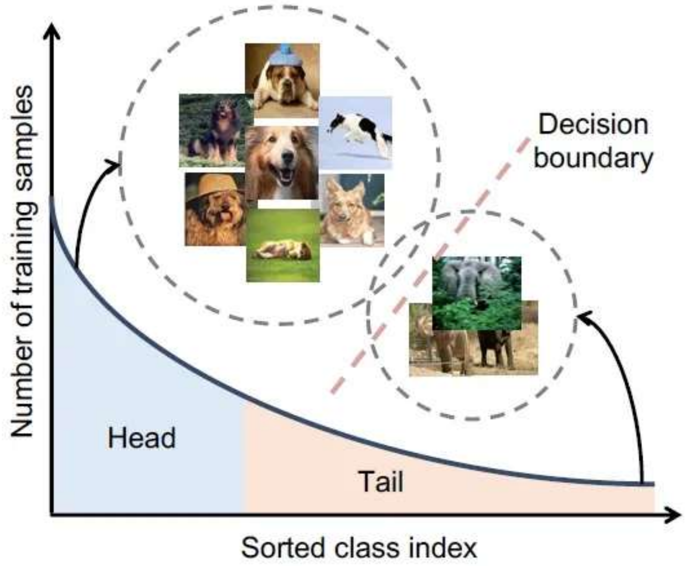
Subclass-balancing Contrastive Learning for Long-tailed Recognition

Chengkai Hou[1], Jieyu Zhang[2], Haonan Wang[3], Tianyi Zhou[4]*
[1]Jilin university, [2]University of Washington, [3]National University of Singapore
[4] University of Maryland
houck20@mails.jlu.edu.cn, jieyuz2@cs.washington.edu, haonan.wang@u.nus.edu
tianyi@umd.edu

ICCV 2023

Related Work

Long-tailed Distribution



- **Re-sampling**: over /undersampling/ balanced-sampling
- **Re-weighting**: assign higher weights to tail classes
- **Logit-adjustment**: add a prior related to n_i to adjust margin
- **Multi-expert**: multiple networks, sub-datasets, loss, distillation

---

- **Contrastive learning**: class centers (KCL, PCL); augmentation to bring classes closer (BCL); augmentation to bring each other closer (GMCL); CIIP
- **Others**: combining distillation learning, transfer learning, new metrics, feature generation, LLM-generated

- Train on a long-tailed dataset; test on a balanced dataset【Top-1 acc/error】
- Lead to the distorted embedding space and the biased classifier

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Subclass-balancing Contrastive Learning for Long-tailed Recognition

Chengkai Hou[1], Jieyu Zhang[2], Haonan Wang[3], Tianyi Zhou[4]*
[1]Jilin university, [2]University of Washington, [3]National University of Singapore
[4] University of Maryland
houck20@mails.jlu.edu.cn, jieyuz2@cs.washington.edu, haonan.wang@u.nus.edu
tianyi@umd.edu

ICCV 2023

Motivation

Methods

Drawbacks

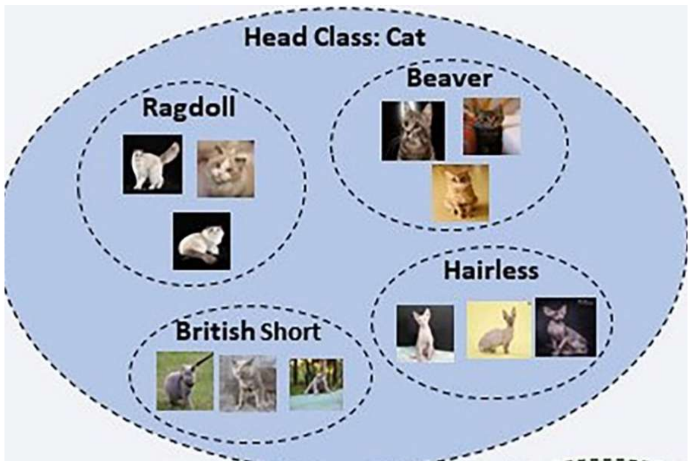
Re-sampling: over /undersampling/ balanced-sampling

Re-weighting: assign higher weights to tail classes

Contrastive learning: class centers (SCL、KCL)

Sacrifice instance balance

Ignore the potential rich semantic structure of head classes



Motivation

Contrastive learning: class centers (SCL、KCL) ------ instance unbalance

$$\mathcal{L}_{SCL} = \sum_{i=1}^{N} -\frac{1}{|\tilde{P}_i|} \sum_{z_p \in \tilde{P}_i} \log \frac{\exp(z_i \cdot z_p^\top / \tau)}{\sum_{z_a \in \tilde{V}_i} \exp(z_i \cdot z_a^\top / \tau)}$$

$$\mathcal{L}_{KCL} = \sum_{i=1}^{N} -\frac{1}{k+1} \sum_{z_p \in \tilde{P}_i^k} \log \frac{\exp(z_i \cdot z_p^\top / \tau)}{\sum_{z_a \in \tilde{V}_i} \exp(z_i \cdot z_a^\top / \tau)}$$

$p(c) = \frac{k}{n_c}$ the imbalance ratio is $\frac{n_1}{n_C} = 100$

$p(C) = 1$ $p(1) = \frac{k}{100k} = 0.01$

When the instances of head class are selected once, that of tail class may already be trained 100 times.

Method

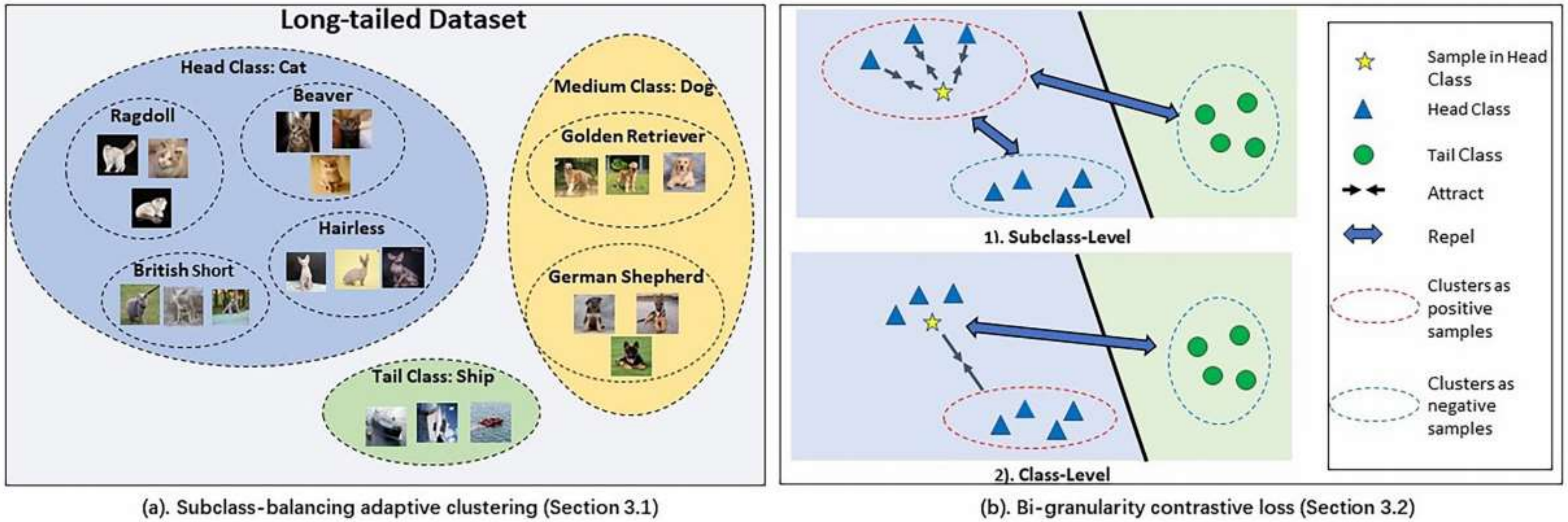


Figure 1: Illustration of subclass-balancing contrastive learning (SBCL). It initially divides the head classes into multiple subclasses of comparable size. Then, during training, SBCL builds each sample to be closer to samples from the same subclass than samples from different subclasses but the same class, which are also made to be closer than samples from different classes.

Method

Algorithm 1 Subclass-balancing Adaptive Clustering

Require: Sample set $\mathcal{S} = \{x_i\}_{i=1}^{n}$; A threshold M; The number of iterations K;

$M = \max(n_C, \delta)$

Ensure: Cluster assignments for samples in $\mathcal{S}$

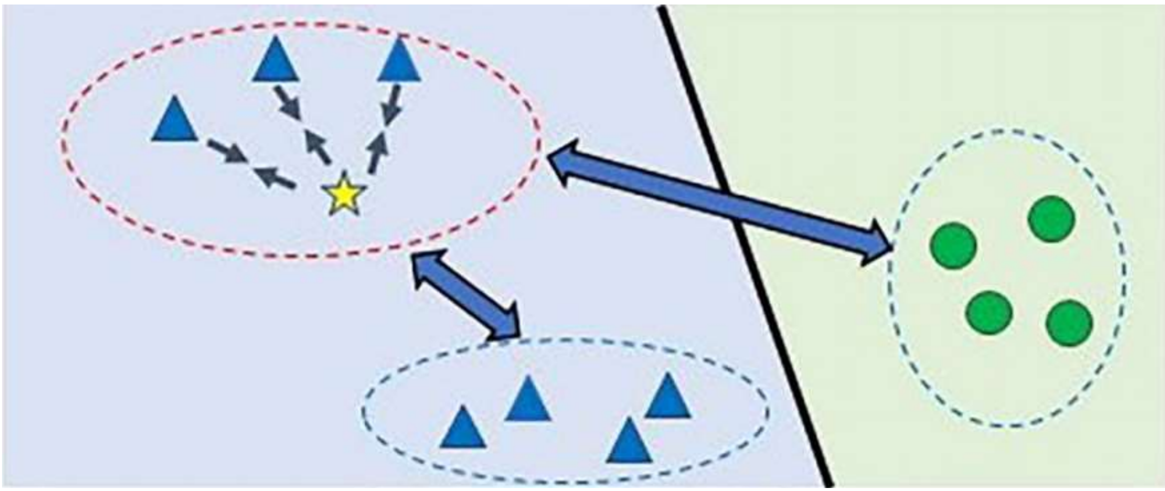
for $k = 0$ to K **do**
  if $k = 0$ **then**
    Choose the cluster centers y_j which are farther away from previously selected centers.
  else
    Update the cluster centers $y_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i$. ▷ n_j is the number of samples in a cluster
  end if
  Construct the cluster center set $\mathcal{C} = \{y_j\}_{j=1}^{m}$. ▷ m is the number of cluster centers
  while $\mathcal{S} \neq \phi$ **do** ▷ Assign samples to centers y_i
    Select the most similar pair $(x_i, y_j) = \arg\max_{x \in \mathcal{S}, y \in \mathcal{C}} \text{cosine-similarity}(x, y)$.
    Assign the sample x_i to the center y_j.
    Delete the assigned sample x_i from the sample set $\mathcal{S} = \mathcal{S} / \{x_i\}$.
    if $n_j \geq M$ **then** ▷ Sample number in a cluster exceeds the threshold M
      Delete the cluster center y_j from the cluster center set $\mathcal{C} = \mathcal{C} / \{y_j\}$.
    end if
  end while
end for

Apply clustering algorithm to classes which contain multiple instances while the tail classes remain unchanged.

Method

A direct consequence of **replacing class label in SCL/KCL with cluster label** is that we no longer distinguish instances from different head classes, and therefore the **boundaries between classes might be blurry**, leading to sub-optimal feature space

$$\mathcal{L}_{SCL} = \sum_{i=1}^{N} -\frac{1}{|\tilde{P}_i|} \sum_{z_p \in \tilde{P}_i} \log \frac{\exp(z_i \cdot z_p^\top / \tau)}{\sum_{z_a \in \tilde{V}_i} \exp(z_i \cdot z_a^\top / \tau)}$$



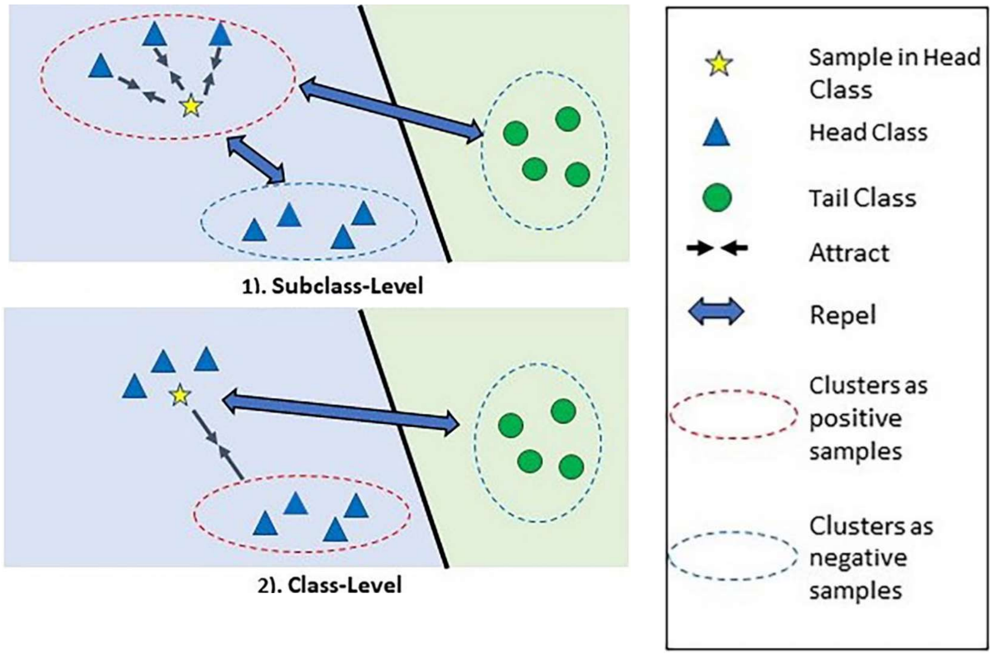
Method

$$\mathcal{L}_{SBCL} = -\sum_{i=1}^{N}\Big(\frac{1}{|\tilde{M}_i|}\sum_{z_p \in \tilde{M}_i} \log \frac{\exp(z_i \cdot z_p^\top/\tau_1)}{\sum_{z_a \in \tilde{V}_i} \exp(z_i \cdot z_a^\top/\tau_1)}$$

$$+\beta \frac{1}{|\tilde{P}_i| - |M_i|}\sum_{z_p \in \tilde{P}_i/M_i} \log \frac{\exp(z_i \cdot z_p^\top/\tau_2)}{\sum_{z_a \in \tilde{V}_i/M_i} \exp(z_i \cdot z_a^\top/\tau_2)}\Big)$$

$$M_i = \{z_j \in P_i : \Gamma_{y_i}(x_i) = \Gamma_{y_i}(x_j)\}$$

$$P_i = \{z_j \in V_i : y_j = y_i\}$$



Method

$$\mathcal{L}_{SBCL} = -\sum_{i=1}^{N}\Big(\frac{1}{|\tilde{M}_i|}\sum_{z_p\in\tilde{M}_i}\log\frac{\exp(z_i\cdot z_p^\top/\tau_1)}{\sum_{z_a\in\tilde{V}_i}\exp(z_i\cdot z_a^\top/\tau_1)}$$

$$+\beta\frac{1}{|\tilde{P}_i|-|M_i|}\sum_{z_p\in\tilde{P}_i/M_i}\log\frac{\exp(z_i\cdot z_p^\top/\tau_2)}{\sum_{z_a\in\tilde{V}_i/M_i}\exp(z_i\cdot z_a^\top/\tau_2)}\Big)$$

Low temperature τ 1: Encourage concentrated distribution of features.

High temperature τ 2: Make the feature distribution uniform.

Expect instances of **the same subclass to form more concentrated clusters** in the feature space than instances of the same class

$$\tau_2(c) = \tau_1\cdot\exp\left(\frac{\phi(c)}{\frac{1}{C}\sum_{i=1}^{C}\phi(i)}\right) \qquad \phi(c) = \frac{\sum_{i=1}^{n_c}\|z_i - t_c\|_2}{n_c\log(n_c+\alpha)}$$

Method

Algorithm 2 Training Algorithm

Require: Dataset $\mathcal{D} = \{x_i, y_i\}_{i \in [n]}$; ; The update interval of cluster assignment K; The number of warm-up epoch T_0; The total number of epoch T; The hyperparameters β and δ.

Ensure: A trained feature extractor $f_\theta(\cdot)$

1: Initialize the model parameters θ
2: Train $f_\theta(\cdot)$ with SCL/KCL for T_0 epochs ▷ Warm-up stage
3: **for** $t = T_0$ to T **do**
4: **if** $t\%K == 0$ **then** ▷ Update cluster and termperature
5: Update the cluster assignment based on the current feature extractor $f_\theta(x)$
6: Update the temperture τ_2 for each head class using Eq. 5 and Eq. 6
7: **end if**
8: Train $f_\theta(\cdot)$ using Eq. 4 ▷ Subclass-balancing contrastive learning
9: **end for**

Experiment

Table 1: Performance comparison on ImageNet-LT and iNaturalist 2018 datasets. Top-1 accuracy of ResNet-50 [26] is reported. The "Many", "Medium" ,"Few" and "All" denotes different groups. † denotes our reproduced results of PCL, SwAV, and BYOL based on their official code. Other baselines' results on ImageNet-LT and iNaturalist 2018 are copied from Li *et al*. [41].

Backbone	ImageNet-LT				iNaturalist 2018			
Methods	Many	Medium	Few	All	Many	Medium	Few	All
CE	64.0	33.8	5.8	41.6	72.2	63.0	57.2	61.7
Focal loss [44]	51.0	40.8	20.8	43.7	-	-	-	61.3
CB-Focal [15]	-	-	-	-	-	-	-	61.1
LDAM-DRW [6]	60.4	46.9	30.7	49.8	-	-	-	64.6
OLTR [46]	35.8	32.3	21.5	32.2	59.0	64.1	64.9	63.9
τ-norm [34]	56.6	44.2	27.4	46.7	71.1	68.9	69.3	69.3
cRT [34]	58.8	44.0	26.1	47.3	73.2	68.8	66.1	68.2
LWS [34]	57.1	45.2	29.3	47.7	71.0	69.8	68.8	69.5
PCL† [40]	34.7	26.1	12.3	27.5	48.5	45.9	41.7	44.5
SwAV† [8]	37.5	28.3	15.6	30.1	51.9	48.4	43.7	47.0
BYOL† [21]	37.7	28.9	16.3	30.6	52.3	48.6	44.1	47.2
SCL [36]	61.4	47.0	28.2	49.8	-	-	-	66.4
KCL [33]	62.4	49.0	29.5	51.5	-	-	-	68.6
TSC [41]	63.5	49.7	30.4	52.4	72.6	70.6	67.8	69.7
SBCL	**63.8**	**51.3**	**31.2**	**53.4**	**73.3**	**71.9**	**68.6**	**70.8**

Experiment

Table 2: Performance comparison on CIFAR-100-LT. Top-1 accuracy of the ResNet-32 [26] under different imbalance ratios is reported. We also report the accuracy of our re-implemented important baselines (†) in same setting on CIFAR-100-LT. The columns of "Statistic (IR 100)" are the results of different disjoint subsets on CIFAR-100-LT with imbalance ratio being 100.

Method	Imbalance Ratio			Statistic (IR 100)		
	100	50	10	Many	Medium	Few
CE	38.3	43.9	55.7	65.2	37.1	9.1
CB-CE [15]	38.6	44.6	57.1	-	-	-
Focal Loss [44]	38.4	44.3	55.8	65.3	38.4	8.1
CB-Focal [15]	38.7	45.2	58.0	65.0	37.6	10.3
CE-DRW [6]	41.4	45.3	58.1	-	-	-
CE-DRS [6]	41.6	45.5	58.1	-	-	-
LDAM [6]	39.6	45.0	56.9	-	-	-
LDAM-DRW [6]	42.0	46.6	58.7	61.5	41.7	20.2
M2m-ERM [37]	42.9	-	58.2	-	-	-
M2m-LDAM [37]	43.5	-	57.6	-	-	-
cRT [34]	43.3	46.8	58.1	64.0	44.8	18.1
LWS [34]	43.1	46.4	58.1	-	-	-
SCL [36] †	42.1	45.2	54.8	62.8	42.0	18.4
KCL [33]	42.8	46.3	57.6	-	-	-
KCL†	42.8	46.4	57.5	63.4	42.5	19.2
TSC [41]	43.8	47.4	**59.0**	-	-	-
TSC†	43.5	47.6	58.7	63.7	43.2	20.4
SBCL	**44.9**	**48.7**	57.9	**64.4**	**45.3**	**22.2**

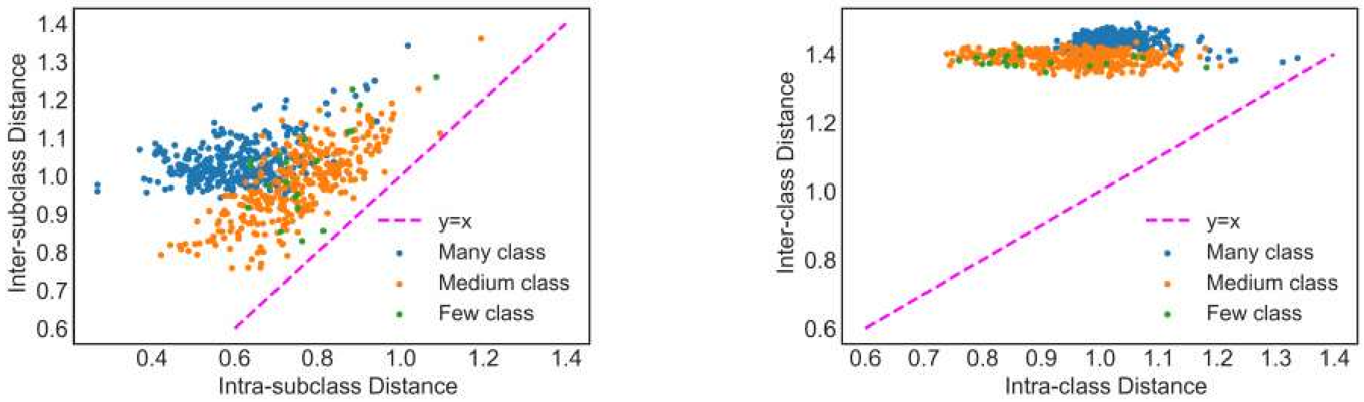
Table 3: Object detection results on PASCAL VOC.

Method	ImageNet			ImageNet-LT		
	AP_{50}	AP	AP_{75}	AP_{50}	AP	AP_{75}
random init.	60.2	33.8	33.1	60.2	33.8	33.1
CE	81.3	53.7	59.2	76.5	48.5	51.0
CL [24]	81.3	56.1	62.7	78.2	51.5	56.5
KCL [33]	**82.3**	55.5	62.1	79.7	52.6	57.9
SBCL	81.9	**56.2**	**62.8**	**80.6**	**53.4**	**58.8**

Table 5: Ablation study on different components of SBCL.

Warm-up	Adaptive cluster	Dynamic temperature	CIFAR-100-LT		
Imbalance Ratio			100	50	10
	✓	✓	44.0	47.9	57.2
✓		✓	43.8	47.2	56.5
✓	✓		43.8	47.8	57.0
✓	✓	✓	**44.9**	**48.7**	**57.9**

Experiment



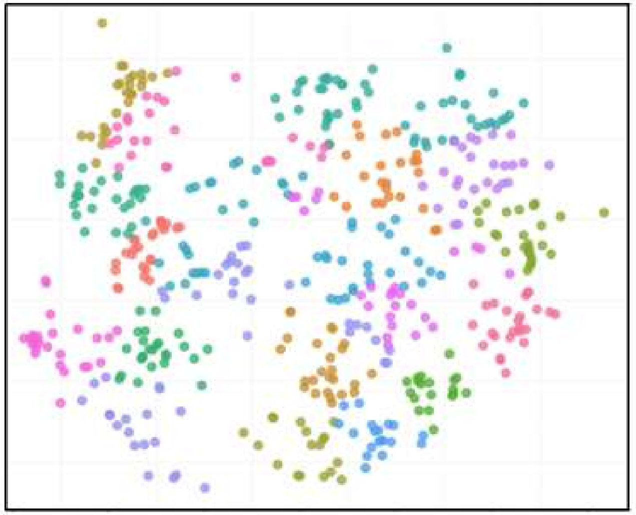
(a) Feature distance in subclasses.

(b) Feature distance in classes.

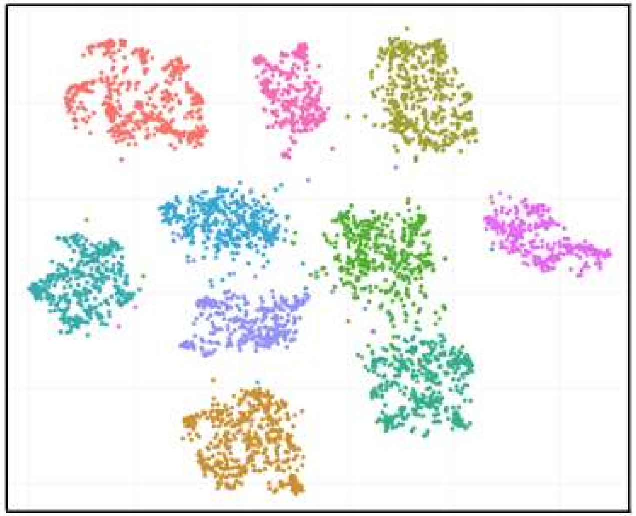
Figure 2: Feature distance of subclasses and classes on CIFAR-100-LT with imbalance ratio 100. We randomly sample instances from many-shot and medium-shot classes so that the size of each equals to that of few-shot classes.

$$\mathbf{D}(z_i, S) = \frac{1}{|S|} \sum_{z_j \in S} \|z_i - z_j\|_2$$

Experiment



(a) Distribution of subclasses on a head class.



(b) Distribution of head classes on the dataset.

Figure 3: Feature distribution of subclasses and classes on CIFAR-100-LT with imbalance ratio 100. Color represents subclasses/classes.

Thanks