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

# FCC: Feature Clusters Compression for Long-Tailed Visual Recognition

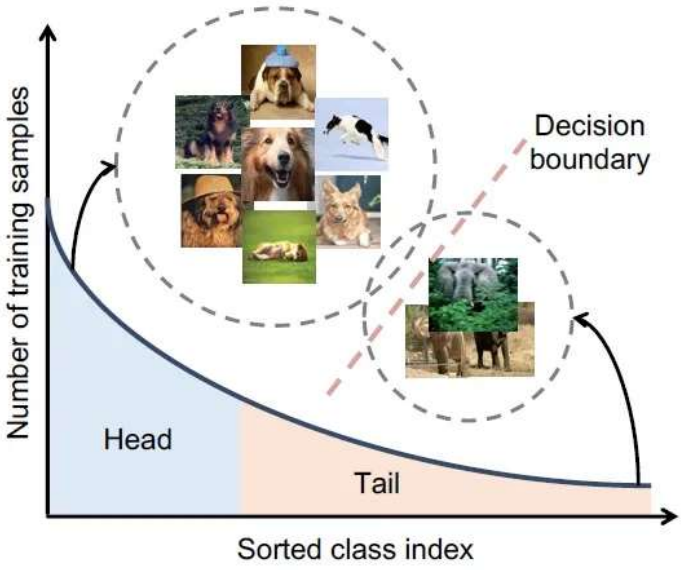
Jian Li[1], Ziyao Meng[2], Daqian Shi[3], Rui Song[1], Xiaolei Diao[3], Jingwen Wang[1], Hao Xu[1,*]
[1]Jilin University, [2]University of Minho, [3]University of Trento
lijianjlu@126.com, id9272@alunos.uminho.pt, daqian.shi@unitn.it, songrui20@mails.jlu.edu.cn,
xiaolei.diao@unitn.it, wangjingwen_jlu@163.com, xuhao@jlu.edu.cn

CVPR 2023

# Related Work

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

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

- 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

# FCC: Feature Clusters Compression for Long-Tailed Visual Recognition

Jian Li[1], Ziyao Meng[2], Daqian Shi[3], Rui Song[1], Xiaolei Diao[3], Jingwen Wang[1], Hao Xu[1,*]
[1]Jilin University, [2]University of Minho, [3]University of Trento
lijianjlu@126.com, id9272@alunos.uminho.pt, daqian.shi@unitn.it, songrui20@mails.jlu.edu.cn,
xiaolei.diao@unitn.it, wangjingwen_jlu@163.com, xuhao@jlu.edu.cn

CVPR 2023

# Motivation

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

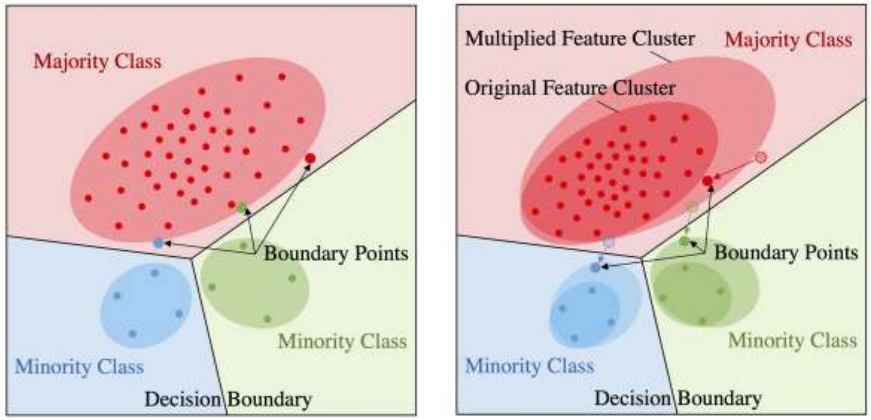



Figure 1. (a) DNNs can map backbone features into different clusters, while minority classes are mapped into sparse clusters compared to majority classes. The sparsity causes boundary points mapped far from their clusters or even cross the boundary. (b) FCC can compress original features compared with multiplied features, which makes these features are mapped closer together. Because the decision boundary remains unchanged in test phase, boundary points will be brought back within the boundary.

# Method

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

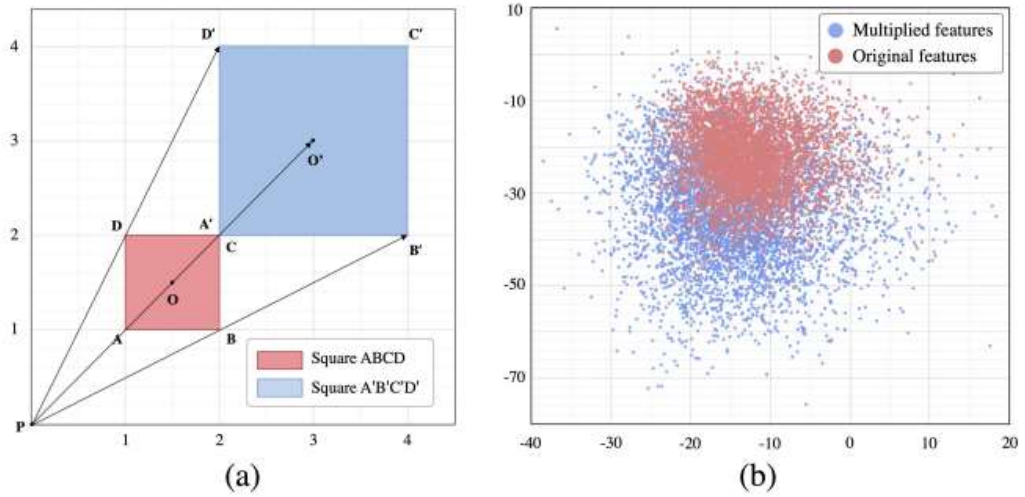



Figure 2. (a) The square $ABCD$ will be transformed to the square $A'B'C'D'$ when its vertex coordinates are multiplied by 2. (b) Visualization of the original and multiplied features of class 0. Experiment is conducted on CIFAR-10-LT-100, where FCC with $\gamma$ of 0.5 is used to ResNet-32.

expanding to $N$-dimensional space, the distance is also shorten by $\tau$ times and the density is enlarged to $\tau^N$ times.

$$f_M^i = f_O^i * \tau_i \tag{1}$$

where $f_M^i$ and $f_O^i$ denote the multiplied and original features of class $i$, respectively. For the scaling factor $\tau_i$, we define three strategies for setting it to control compression degrees of each class, as followings:

# Method


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

$$f_M^i = f_O^i * \tau_i \tag{1}$$

where $f_M^i$ and $f_O^i$ denote the multiplied and original features of class $i$, respectively. For the scaling factor $\tau_i$, we define three strategies for setting it to control compression degrees of each class, as followings:

**Uniform compression.** Set the same $\tau_i$ for all classes as:

$$\tau_i = 1 + \gamma \tag{2}$$

**Equal difference compression.** $\tau_i$ is reduced in sequence from majority to minority classes, as following:

$$\tau_i = 1 + \gamma * (1 - i/C) \tag{3}$$

**Half compression.** Equal difference compression is only used for top 50% or bottom 50% classes, otherwise $\tau_i$ is set to 1 for other classes, it can be formulated as following:

$$\tau_i = 1 + \gamma * (1 - i/C) * \varphi((-1)^\beta * (i - C/2)) \tag{4}$$

where $\gamma > 0$ is a scaling hyper-parameter, $C$ is the number of classes, and $i \in [0, C)$ is the index of class. $\varphi(\cdot)$ is 1 when its parameter is negative, otherwise it is 0. And $\beta$ is 0 when only compress top 50% classes, otherwise it is 1.

# Prove



**Setting:** The FC network contains an input layer with 3 neurons, a hidden layer with 3 neurons $\{a_1, a_2, a_3\}$ and a output layer with 2 neurons $\{o_1, o_2\}$. $\{\tau x_1, \tau x_2, \tau x_3\}$ and $\{x_1, x_2, x_3\}$ are the multiplied and original features, respectively, and they both belong to class 1. The scaling factor of class 1 is $\tau$ ($\tau > 1$). $\{y_1, y_2\}$ and $\{y_1{}', y_2{}'\}$ are outputs of the multiplied and original features produced by the FC network, respectively. $\{w_{i1}, w_{i2}, w_{i3}\}$ and $b_i$ are weights and bias of the neuron $a_i$ ($i \in \{1, 2, 3\}$), respectively. $\{n_{j1}, n_{j2}, n_{j3}\}$ and $z_j$ are weights and bias of the neuron $o_j$ ($j \in \{1, 2\}$), respectively.

**Target:** If the FC network can normally work, the classification result of the original feature will be equal to that of the multiplied feature, i.e., $y_1{}' > y_2{}'$ when $y_1 > y_2$.

The outputs of the multiplied feature can be expressed as follows:

$$\begin{aligned} y_i = {} & n_{i1}(\tau w_{11}x_1 + \tau w_{12}x_2 + \tau w_{13}x_3) + n_{i1}b_1 + \\ & n_{i2}(\tau w_{21}x_1 + \tau w_{22}x_2 + \tau w_{23}x_3) + n_{i2}b_2 + \\ & n_{i3}(\tau w_{31}x_1 + \tau w_{32}x_2 + \tau w_{33}x_3) + n_{i3}b_3 + \\ & z_i \end{aligned} \tag{5}$$

where $i \in \{1, 2\}$, then we denote $(y_1 - y_2)$ as $\eta$, $(w_{11}x_1 + w_{12}x_2 + w_{13}x_3)$ as $X_1$, $(w_{21}x_1 + w_{22}x_2 + w_{23}x_3)$ as $X_2$, $(w_{31}x_1 + w_{32}x_2 + w_{33}x_3)$ as $X_3$ and $(n_{11}b_1 + n_{12}b_2 + n_{13}b_3 + z_1) - (n_{21}b_1 + n_{22}b_2 + n_{23}b_3 + z_2)$ as $B$, further $\eta$ is converted as follows:

$$\eta = \tau k_1 X_1 + \tau k_2 X_2 + \tau k_3 X_3 + B \tag{6}$$

# Prove



$$\eta = \tau k_1 X_1 + \tau k_2 X_2 + \tau k_3 X_3 + B \qquad (6)$$

a (decision) plane in geometric space when $\eta = 0$.

$y_1 > y_2, \eta > 0$ and Eq. (6) can be formulated as follows:

$$\begin{cases} \tau d_1 X_1 + \tau d_2 X_2 + \tau d_3 X_3 > 1, & B < 0 \\ \tau d_1 X_1 + \tau d_2 X_2 + \tau d_3 X_3 < 1, & B > 0 \end{cases} \qquad (7)$$

$$\eta' = k_1 X_1 + k_2 X_2 + k_3 X_3 + B \qquad (8)$$

When $\eta' = 0$, Eq. (8) can be formulated as follows:

$$d_1 X_1 + d_2 X_2 + d_3 X_3 = 1 \qquad (9)$$

# Prove



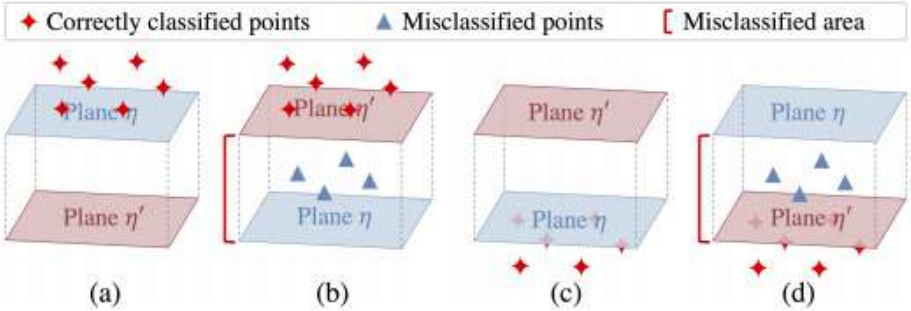



Figure 3. Relationship between planes $\eta$ and $\eta'$ and feature points in geometric space. When feature points are above plane $\eta$, (a) plane $\eta'$ is below plane $\eta$, or (b) plane $\eta'$ is above plane $\eta$. When feature points are below plane $\eta$, (c) plane $\eta'$ is above plane $\eta$, or (d) plane $\eta'$ is below plane $\eta$.

$$\eta = \tau k_1 X_1 + \tau k_2 X_2 + \tau k_3 X_3 + B \tag{6}$$

$$\begin{cases} \tau d_1 X_1 + \tau d_2 X_2 + \tau d_3 X_3 > 1, & B < 0 \\ \tau d_1 X_1 + \tau d_2 X_2 + \tau d_3 X_3 < 1, & B > 0 \end{cases} \tag{7}$$

$$\eta' = k_1 X_1 + k_2 X_2 + k_3 X_3 + B \tag{8}$$

$$d_1 X_1 + d_2 X_2 + d_3 X_3 = 1 \tag{9}$$

$B < 0$, the point $(X_1, X_2, X_3)$ is above plane $\eta$ based on Eq. (7). If plane $\eta'$ is below plane $\eta$, the point is also above plane $\eta'$, as shown in Fig. 3a, so $d_1 X_1 + d_2 X_2 + d_3 X_3 > 1$ in Eq. (9), and then we can get $\eta' > 0$ (i.e., $y_1' > y_2'$) based on Eqs. (8) and (9). That implies the FC can normally work on this point. If plane $\eta'$ is above plane $\eta$, the point might be above or below plane $\eta'$, as shown in Fig. 3b. The point can also be correctly classified when it is above plane $\eta'$ since $d_1 X_1 + d_2 X_2 + d_3 X_3 > 1$, but when it is below plane $\eta'$, $d_1 X_1 + d_2 X_2 + d_3 X_3 < 1$ and $y_1' < y_2'$, which means the FC will misclassify the point.

$B > 0$, …

$B = 0$, **True**

**Target:** If the FC network can normally work, the classification result of the original feature will be equal to that of the multiplied feature, i.e., $y_1' > y_2'$ when $y_1 > y_2$.

# Experiment


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

| Method | CIFAR-10-LT-50 | | | CIFAR-10-LT-100 | | | CIFAR-100-LT-50 | | | CIFAR-100-LT-100 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Raw | FCC | Incr | Raw | FCC | Incr | Raw | FCC | Incr | Raw | FCC | Incr |
| Baseline (Vanilla ResNet32) [10] | 22.99% | 19.78% | +3.21% | 27.59% | 24.08% | +3.51% | 57.38% | 54.83% | +2.55% | 60.92% | 58.93% | +1.99% |
| Focal loss (ICCV 2017) [21] | 23.29% | 20.49% | +2.80% | 27.94% | 26.23% | +1.71% | 57.25% | 55.24% | +2.01% | 62.29% | 58.63% | **+3.66%** |
| CB Focal loss (CVPR 2019) [7] | 22.63% | 21.37% | +1.26% | 25.63% | 25.37% | +0.26% | 56.79% | 54.84% | +1.95% | 61.28% | 59.42% | +1.86% |
| CBCE (CVPR 2019) [7] | 21.48% | 19.51% | +1.97% | 27.50% | 24.15% | +3.35% | 56.58% | 54.60% | +1.98% | 61.56% | 59.59% | +1.97% |
| BSCE (NeurIPS 2020) [27] | 17.84% | 16.85% | +0.99% | 21.78% | 20.87% | +0.91% | 52.47% | 52.61%* | -0.14% | 58.55% | 57.30% | +1.25% |
| CELS (CVPR 2016) [28] | 22.70% | 18.97% | +3.73% | 27.49% | 26.40% | +1.09% | 56.96% | 54.80% | +2.16% | 61.93% | 60.13% | +1.80% |
| CELAS (CVPR 2021) [39] | 21.42% | 19.17% | +2.25% | 27.45% | 24.53% | +2.92% | 57.23% | 55.34% | +1.89% | 61.95% | 60.78% | +1.17% |
| LDAM (NeurIPS 2019) [3] | 21.47% | 21.06% | +0.41% | 26.58% | 26.35% | +0.23% | 56.94% | 56.54% | +0.40% | 61.26% | 60.83% | +0.43% |
| CDT [35] | 18.04% | 17.12% | +0.92% | 21.36% | 20.32% | +1.04% | 56.41% | 56.37% | +0.04% | 60.76% | 60.84%* | -0.08% |
| CB sampling (ICLR 2020) [13] | 22.31% | 21.06% | +1.25% | 27.02% | 26.51% | +0.51% | 60.67% | 59.24% | +1.43% | 66.47% | 64.75% | +1.72% |
| SR sampling (ICLR 2020) [13] | 20.89% | 20.41% | +0.48% | 28.03% | 25.82% | +2.21% | 57.94% | 55.83% | +2.11% | 63.26% | 61.60% | +1.66% |
| PB sampling (ICLR 2020) [13] | 21.11% | 19.76% | +1.35% | 25.16% | 23.70%* | +1.46% | 55.15% | 53.33% | +1.82% | 60.61% | 58.98% | +1.63% |
| Input Mixup (ICLR 2018) [36] | 21.39% | 17.48% | **+3.91%** | 25.84% | 22.44% | +3.40% | 54.48% | 51.35% | +3.13% | 59.14% | 55.81% | +3.33% |
| Manifold Mixup (ICML 2019) [31] | 21.24% | 19.97% | +1.27% | 23.58% | 22.89%* | +0.69% | 56.24% | 51.35% | **+4.89%** | 61.48% | 60.35% | +1.13% |
| Remix (ECCV 2020) [6] | 20.53% | 17.00% | +3.53% | 25.95% | 22.03% | **+3.92%** | 54.25% | 51.36% | +2.89% | 59.16% | 56.23% | +2.93% |
| CB sampling+DRS | 19.86% | 18.4% | +1.46% | 23.36% | 21.91% | +1.45% | 54.28% | 52.93% | +1.35% | 58.32% | 57.00% | +1.32% |
| SR sampling+DRS | 20.49% | 19.16% | +1.33% | 25.59% | 24.09% | +1.50% | 55.92% | 54.11% | +1.81% | 59.73% | 57.29% | +2.44% |
| PB sampling+DRS | 19.73% | 18.44% | +1.29% | 24.58% | 22.70% | +1.88% | 54.56% | 53.19% | +1.37% | 58.82% | 57.29% | +1.53% |
| BSCE+DRW | 18.79% | 17.74% | +1.05% | 21.88% | 20.73% | +1.15% | 53.68% | 53.46% | +0.22% | 57.63% | 57.37% | +0.26% |
| CELAS+DRW | 22.48% | 19.19% | +3.29% | 27.20% | 23.97% | +3.23% | 56.70% | 55.01% | +1.69% | 61.31% | 59.93% | +1.38% |
| CDT+DRW | 18.45% | 17.81% | +0.64% | 21.82% | 20.83% | +0.99% | 53.70% | 53.32%* | +0.38% | 57.76% | 57.54%* | +0.22% |
| cRT (ICLR 2020) [13] | 20.01% | 19.62% | +0.39% | 22.81% | 22.36% | +0.45% | 54.92% | 55.02% | -0.10% | 58.37% | 58.17% | +0.20% |
| DiVE (ICCV 2021) [11] | 17.34% | 15.93% | +1.41% | 21.32% | 19.99% | +1.33% | 50.19% | 50.63% | -0.44% | 55.84% | 54.73% | +1.11% |
| LTR-WB +WD&Max (CVPR 2022) [1] | – | – | – | – | – | – | – | – | – | 47.40% | 46.50%* | +0.90% |
| SADE (NeurIPS 2022) [37] | – | – | – | – | – | – | – | – | – | 51.02% | 50.58%* | +0.44% |
| NCL (CVPR 2022) [18] | **12.92%** | **12.72%*** | +0.20% | **14.50%** | **14.20%*** | +0.30% | **41.67%** | **41.56%*** | +0.11% | **46.14%** | **45.49%*** | +0.65% |

Table 1. Top-1 error rates comparisons between raw methods and those with FCC on **long-tailed CIFAR**. The results are presented in the order of baseline, re-weighting, re-sampling, mixup, two-stage training and multi-expert methods. $*$ denotes $\gamma$ of 0.1 is used in FCC.

# Experiment


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

| Method | ImageNet-LT | | | iNaturalist 2018 | | |
|---|---|---|---|---|---|---|
| | Raw | FCC | Incr | Raw | FCC | Incr |
| ResNet10/32 [10] | 61.07% | 60.60% | +0.47% | 72.49% | 71.99% | +0.50% |
| Focal loss [21] | 63.10% | 62.71% | +0.39% | – | – | – |
| CBCE [7] | 60.92% | 60.86% | +0.06% | 69.85% | 69.12% | +0.73% |
| LDAM-DRW [3] | 63.53% | 63.25% | +0.28% | **59.62%** | **59.54%** | +0.08% |
| BBN [40]† | 51.80% | 50.72% | +1.08% | – | – | – |
| cRT [13] | 58.20% | 56.59% | +1.61% | 64.38% | 63.86% | +0.52% |
| $\tau$-norm [13] | 66.10% | 64.48% | +1.62% | 76.39% | 75.49% | +0.90% |
| DiVE [11] | 56.93% | 56.32% | +0.61% | – | – | – |
| RIDE [32] | 55.72% | 55.49% | +0.23% | – | – | – |
| SADE [37]⋆ | **41.08%** | **39.47%** | +1.61% | – | – | – |
| NCL [18]† | 47.32% | 45.34% | **+1.98%** | 63.46% | 61.17% | **+2.29%** |

Table 2. Top-1 error rates comparisons. † and ⋆ indicate the backbone is ResNet-50 and ResNeXt-50, respectively.

# Experiment—— Compression strategies

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

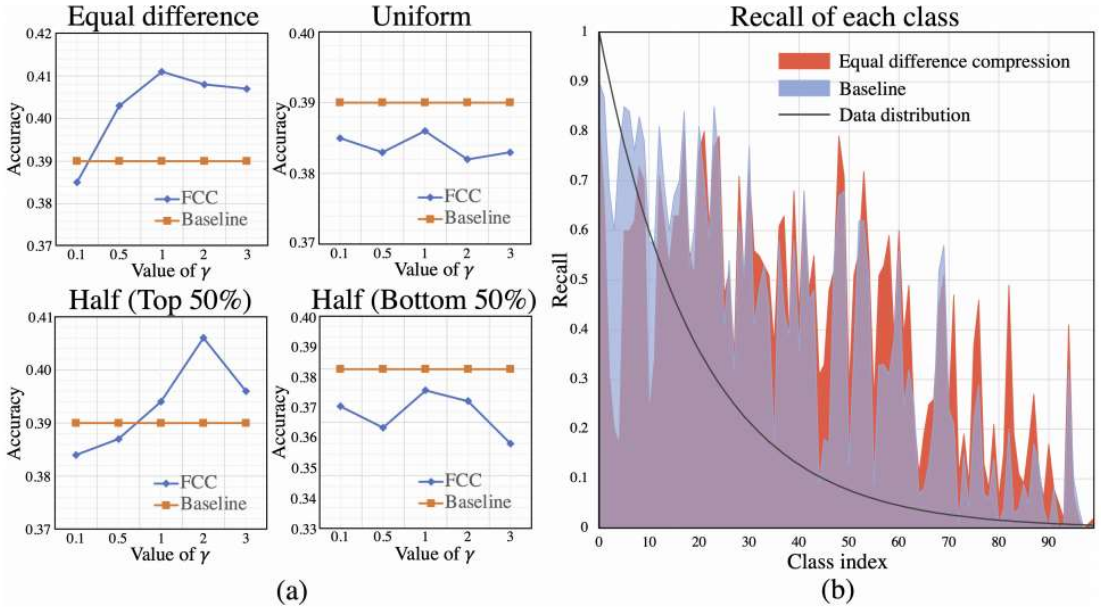



Figure 4. (a) Accuracy comparisons of each compression strategy. (b) Recall comparisons of each class between baseline (vanilla ResNet-32) and FCC with equal difference compression ($\gamma = 1$). Analysis is conducted on CIFAR-100-LT-100.

$$\tau_i = 1 + \gamma * (1 - i/C)$$

$$\tau_i = 1 + \gamma$$

$$\tau_i = 1 + \gamma * (1 - i/C) * \varphi((-1)^{\beta} * (i - C/2))$$

# Experiment——Impact of FCC on boundary points

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

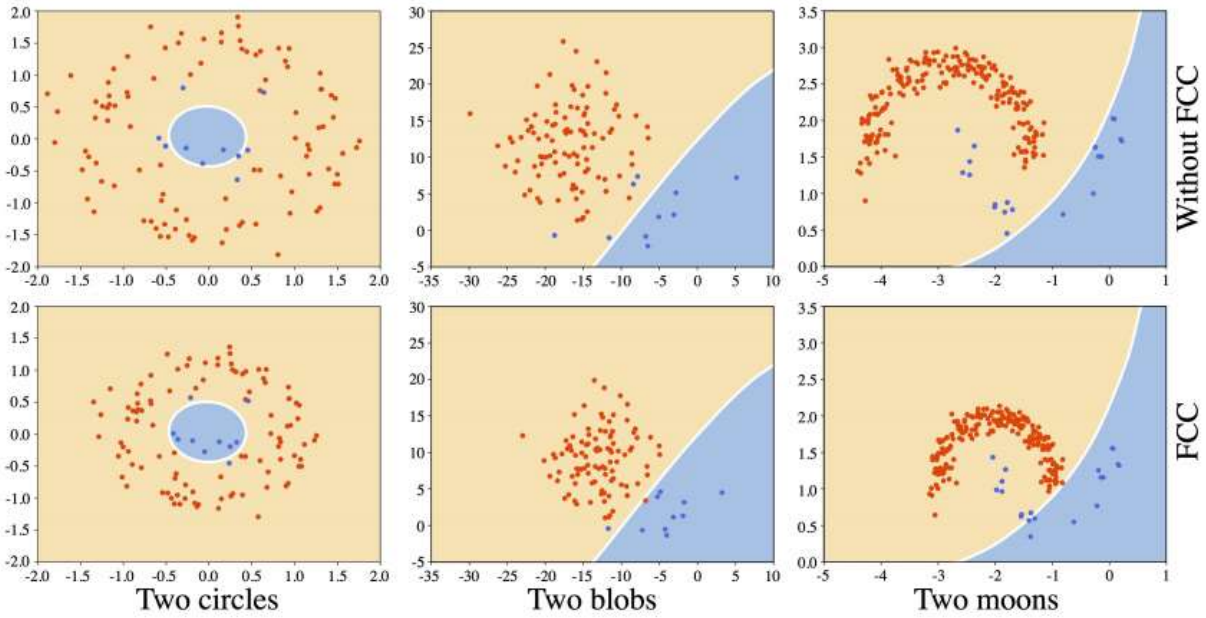



Figure 6. Impact of FCC on boundary points. Majority and minority classes are plotted in orange and blue, respectively. FCC can bring the points of minority classes back within the boundary.


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

# Thanks