JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

JILIN UNIVERSITY

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

Email: lijianjlu@126.com

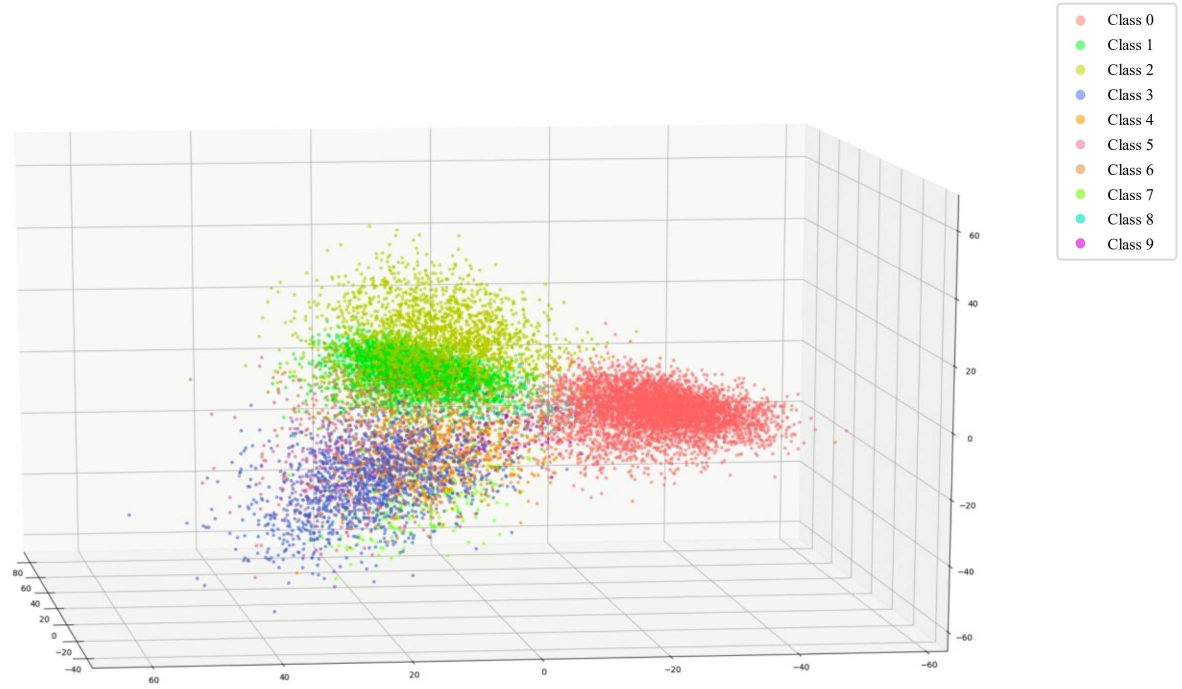
Paper Tag: THU-PM-332

FCC: Feature Clusters Compression for Long-Tailed Visual Recognition

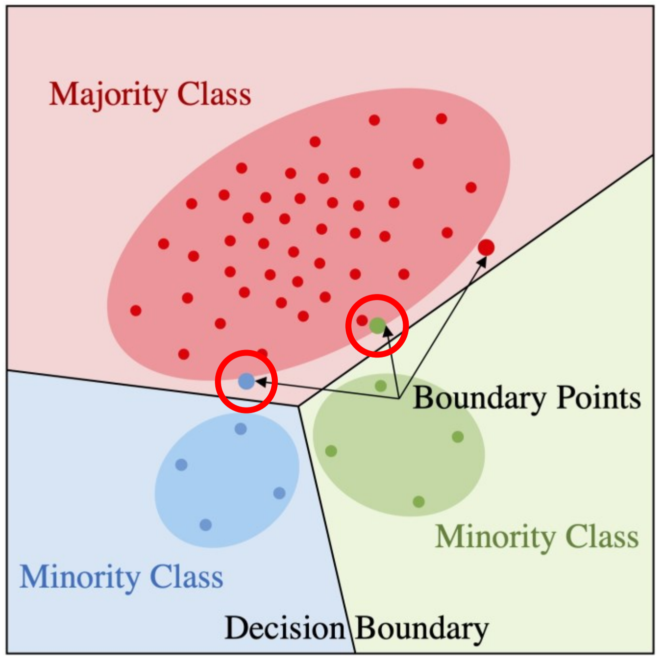
A simple and generic method for long-tailed visual recognition.

Motivation

DNNs can map samples into different feature clusters, while minority classes are mapped into sparser clusters than majority classes.



Visualization of backbone features on CIFAR-10-LT-50 dataset



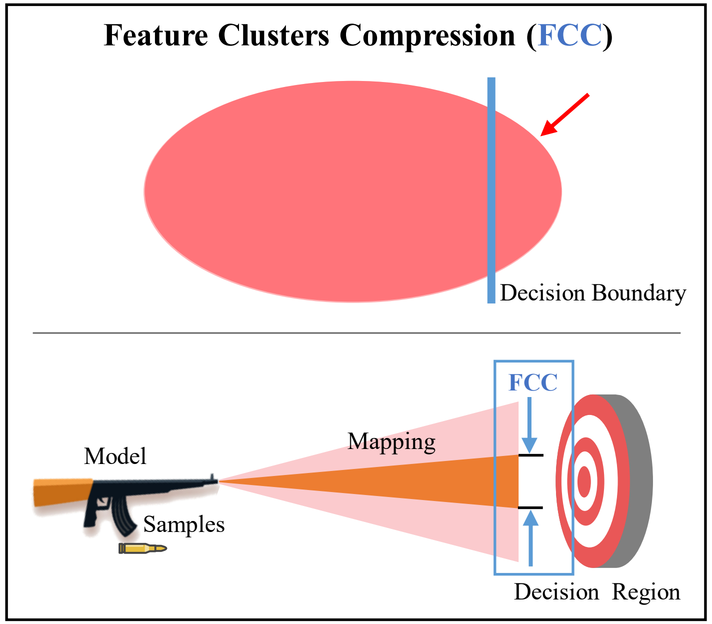
Boundary points of minority classes easily cross the boundary

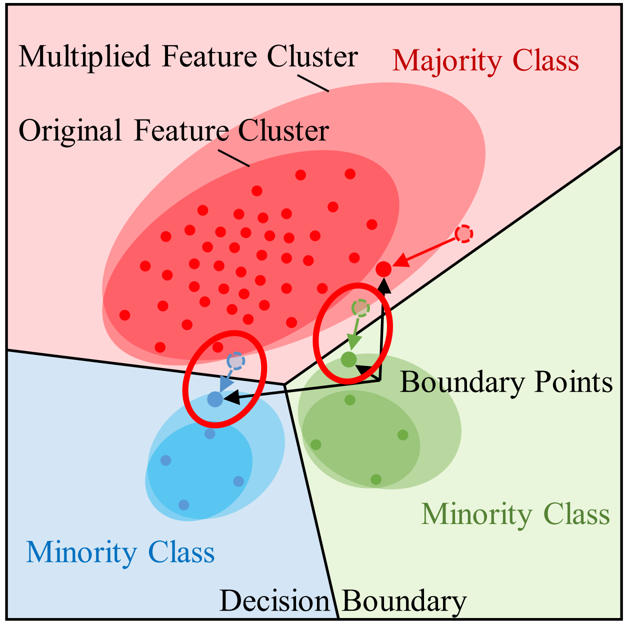
FCC: Feature Clusters Compression for Long-Tailed Visual Recognition

A simple and generic method for long-tailed visual recognition.

Motivation

The proposed FCC can compress backbone feature clusters, such that test samples are mapped closer together and do not easily cross the decision boundary.





FCC: Feature Clusters Compression for Long-Tailed Visual Recognition

A simple and generic method for long-tailed visual recognition.

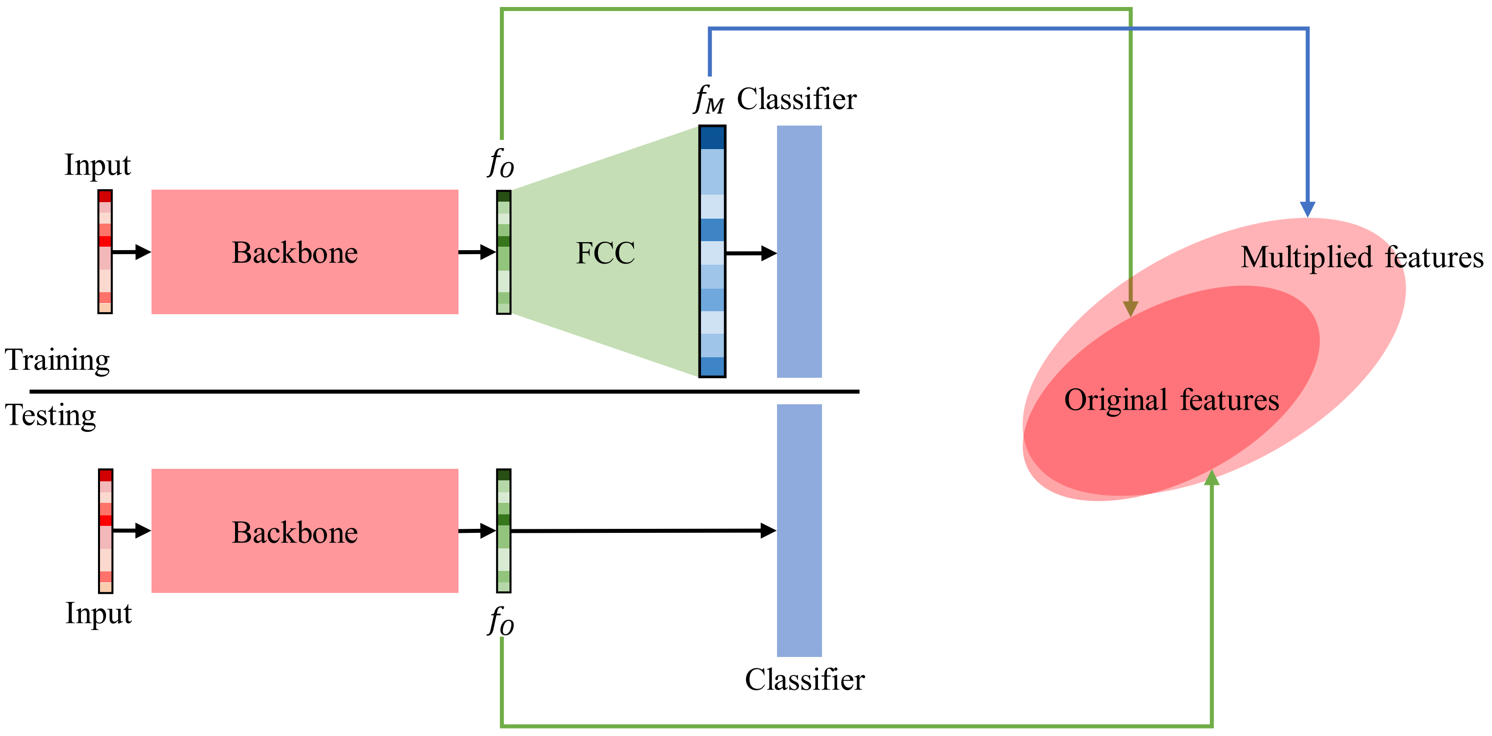
Methodology

FCC can be easily achieved by only multiplying original backbone features $\boldsymbol{f_O^i}$ of class i by a scaling factor $\boldsymbol{\tau_i}$ (> 1) and further feed the multiplied features $\boldsymbol{f_M^i}$ to the classifier in each training batch, as followings:

The Implementation of FCC

$$f_M^i = f_O^i * \tau_i \quad (1)$$

$$\tau_i = 1 + \gamma \quad (2)$$



Methodology

FCC can be easily achieved by multiplying backbone features by a specific scaling factor.

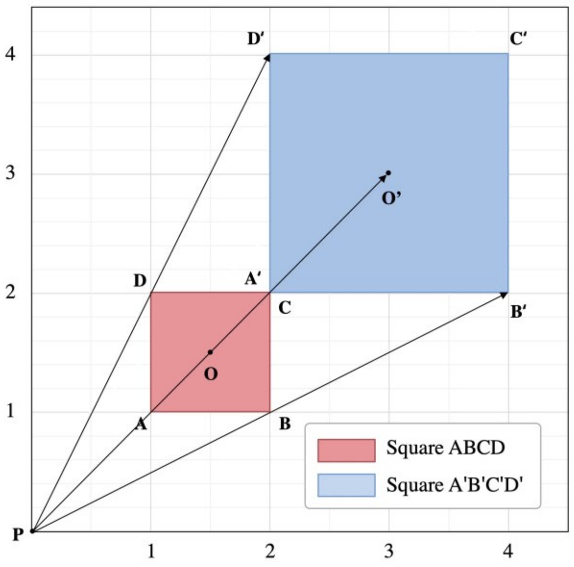
Principle of FCC

The Implementation of FCC

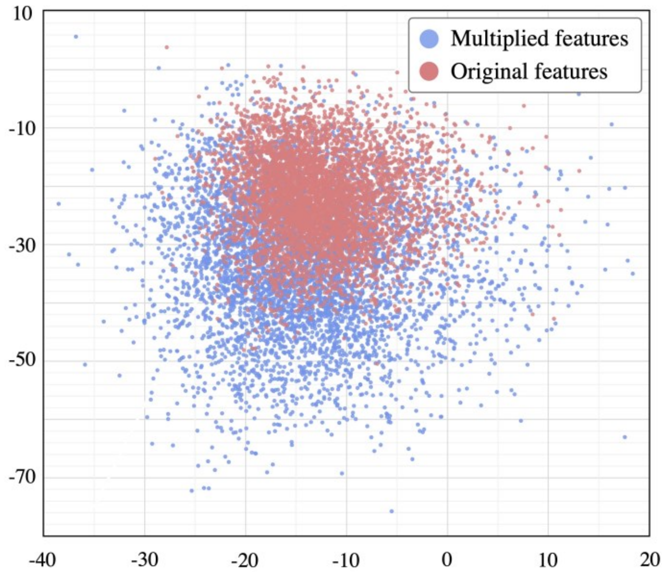
$$f_M^i = f_O^i * \tau_i \quad (1)$$

Example of Compression

$$\begin{aligned} &\text{Point A } (x_A,\ y_A)*\tau \rightarrow \text{Point A' } (\tau x_A,\ \tau y_A) \\ &\text{Point B } (x_B,\ y_B)*\tau \rightarrow \text{Point B' } (\tau x_B,\ \tau y_B) \\ &\text{Point C } (x_C,\ y_C)*\tau \rightarrow \text{Point C' } (\tau x_C,\ \tau y_C) \\ &\text{Point D } (x_D,\ y_D)*\tau \rightarrow \text{Point D' } (\tau x_D,\ \tau y_D) \end{aligned} \quad (2)$$



(a) Principle of FCC



(b) Visualization of features by PCA

Methodology

FCC can be easily achieved by multiplying backbone features by a specific scaling factor.

Different Compression Strategies

For the scaling factor τ_i, we define three strategies for setting it to control compression degrees of each class, as followings:

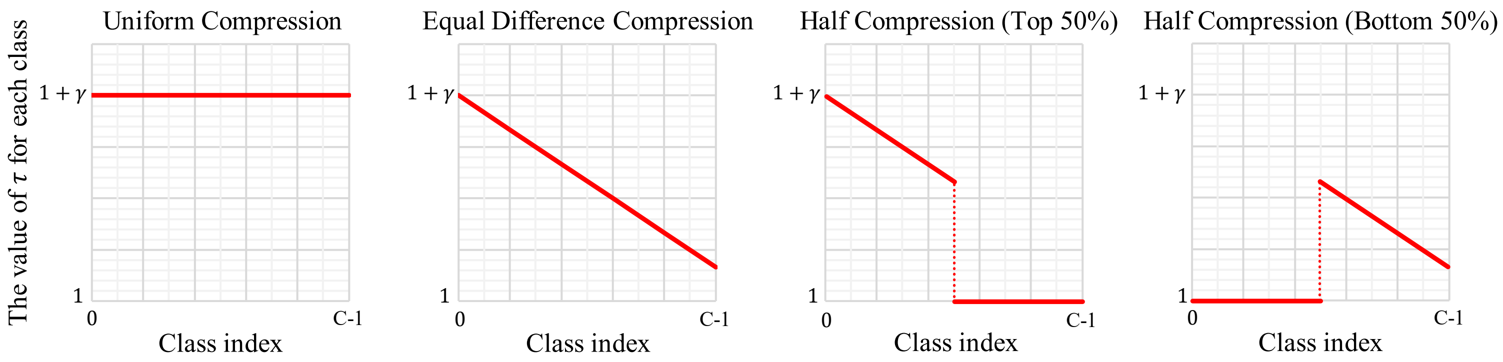
- **Uniform compression.** Set the same τ_i for all classes as:

$$\tau_i = 1 + \gamma$$

- **Equal difference compression.** τ_i is reduced in sequence from majority to minority classes, as following:

$$\tau_i = 1 + \gamma * (1 - \frac{i}{C})$$

- **Half compression.** Equal difference compression is only used for top or bottom 50% classes, otherwise τ_i is set to 1 for other classes.

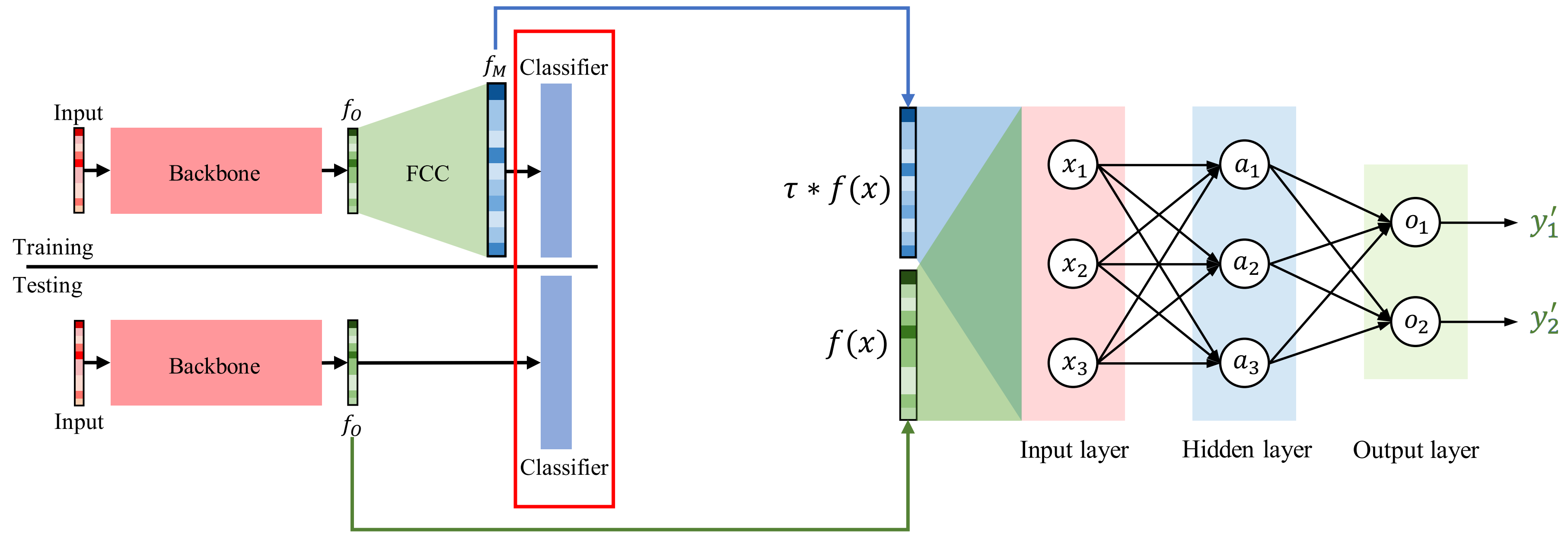


Methodology

FCC can be easily achieved by multiplying backbone features by a specific scaling factor.

Feasibility Demonstration of FCC

The input of the classifier is different in training and testing phases. It is indispensable to demonstrate the classifier can also normally work on the original features in testing stage.



FC can normally work

$if\ \eta = y_1 - y_2 > 0,$

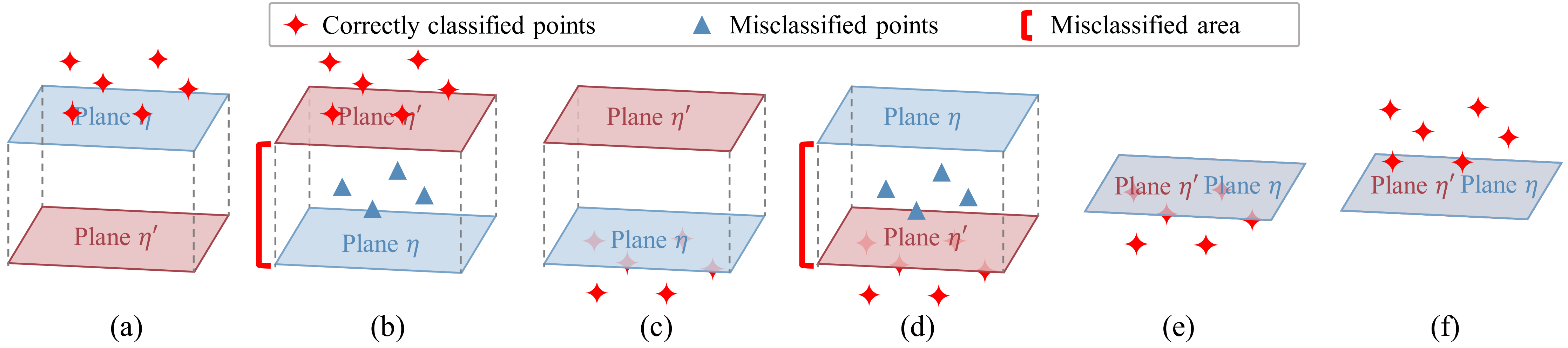
$then\ \eta' = y_1' - y_2' > 0$

Methodology

FCC can be easily achieved by multiplying backbone features by a specific scaling factor.

Feasibility Demonstration of FCC

- η and η' are decision planes in feature space when they are equal to 0.
- Planes η and η' are parallel in feature space.
- FC cannot normally work when original features fall into 'misclassified area'.
- It is a positive relationship between the size of 'misclassified area' and scaling factor τ.
- 'Misclassified area' does not affect the overall performance when τ is set to an appropriate value, since few features will fall into this area.



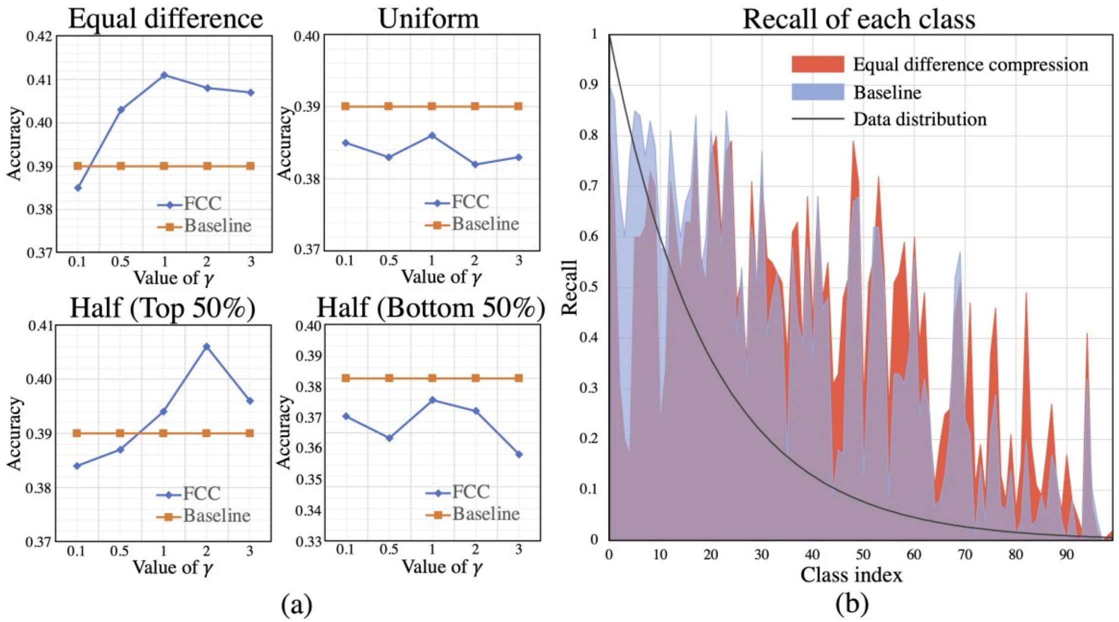
Analysis of FCC

Exploring the characteristics of FCC.

Compression Strategies

We evaluate different compression strategies on CIFAR-10-LT-100 dataset.

- Equal difference compression outperforms other strategies.
- Equal difference compression achieves a remarkable improvement on minority classes.



Analysis of FCC

Exploring the characteristics of FCC.

Hyper-parameter γ

- For CIFAR-10-LT, γ of 0.5 achieves the best performance.
- For CIFAR-100-LT, γ of 1 obtains the best result.
- For ImageNet-LT, γ of 0.1 produces the best result.
- The optimal γ for different datasets is inconsistent, but γ of 0.1, 0.5 and 1 are generally the best choices.

γ	CIFAR-10-LT-50			CIFAR-10-LT-100			CIFAR-100-LT-50			CIFAR-100-LT-100			ImageNet-LT		
	Raw	FCC	Incr.	Raw	FCC	Incr.	Raw	FCC	Incr.	Raw	FCC	Incr.	Raw	FCC	Incr.
0.1	22.99%	23.21%	-0.22%	27.59%	28.74%	-1.15%	57.38%	57.15%	+0.23%	60.92%	61.51%	-0.59%	61.07%	**60.60%**	**+0.47%**
0.5	22.99%	**19.78%**	**+3.21%**	**27.59%**	**24.08%**	**+3.51%**	57.38%	56.97%	+0.41%	60.92%	59.64%	+1.28%	61.07%	60.79%	+0.28%
1	22.99%	20.46%	+2.53%	27.59%	24.45%	+2.14%	57.38%	**54.84%**	**+2.54%**	**60.92%**	**58.93%**	**+1.99%**	61.07%	61.25%	-0.18%
2	22.99%	23.93%	-0.94%	27.59%	27.31%	+0.28%	57.38%	55.27%	+2.11%	60.92%	59.22%	+1.70%	61.07%	61.73%	-0.66%
3	22.99%	20.78%	+2.21%	27.59%	30.11%	-2.52%	57.38%	56.96%	+0.42%	60.92%	59.31%	+1.61%	61.07%	63.88%	-2.81%

Top-1 error rates comparisons between raw methods and those using FCC with different γ.

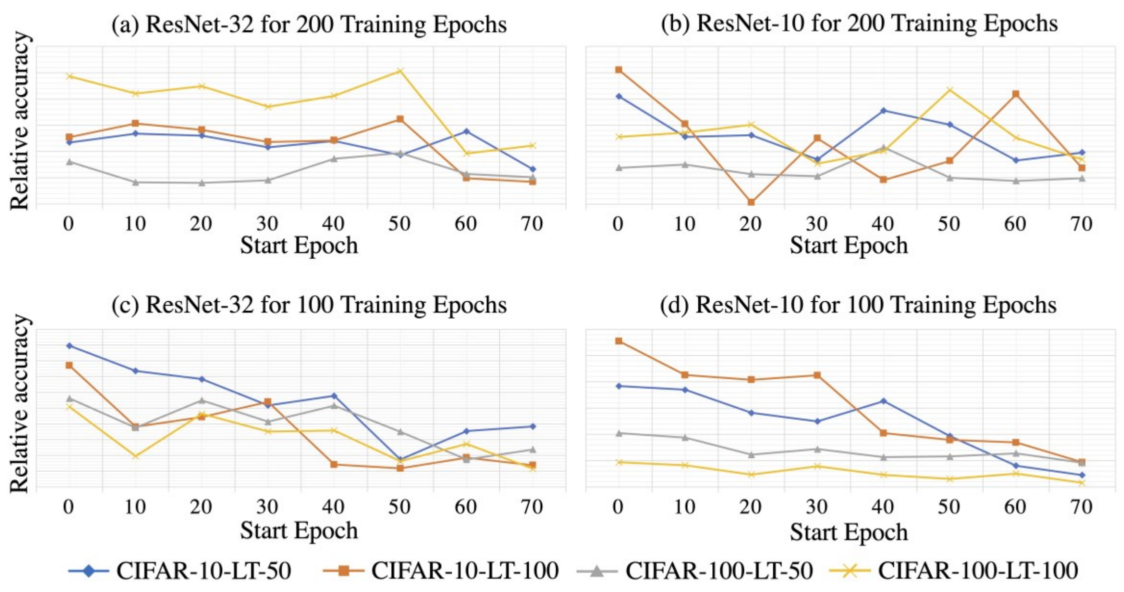
Analysis of FCC

FCC can be easily achieved by multiplying backbone features by a specific scaling factor.

When to Start FCC in Training Phase?

We train ResNet-10/32 for 100/200 epochs on long-tailed CIFAR datasets, in which FCC is used from $[10,20,30,40,50,60,70]^{th}$ epoch.

- For 200 training epochs, using FCC from the 50^{th} epoch generally yields the best results.
- For 100 training epochs, 0^{th} epoch achieves the best results.
- Sufficient epochs need to be reserved for FCC to compress features under fewer training epochs.
- The weaker networks require more epochs for FCC.



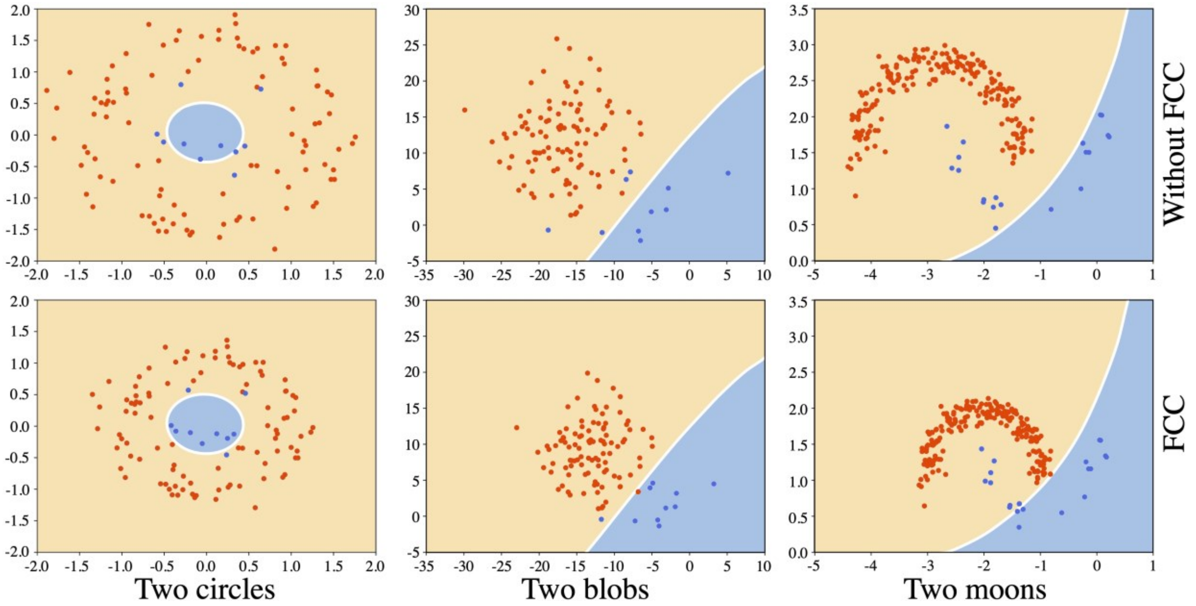
Analysis of FCC

FCC can be easily achieved by multiplying backbone features by a specific scaling factor.

Impact of FCC on Boundary Points

We present the results of FCC on three other imbalanced datasets, which are created based on commonly used datasets from scikit-learn, including two circles, two blobs and two moons.

- The results show that FCC can compress feature clusters.
- FCC can make boundary points of minority classes back within the boundary.



Experimental Results

Overall performance comparisons between raw methods and those with FCC.

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%

Top-1 error rates comparisons between raw methods and those with FCC on long-tailed CIFAR datasets.

Experimental Results

Overall performance comparisons between raw methods and those with FCC.

Results

- Extensive experiments have been conducted on four popular datasets, including CIFAR-10-LT, CIFAR-100-LT, ImageNet-LT and iNaturalist 2018.
- FCC is applied to **28** advanced long-tailed methods.
- For CIFAR-10/100-LT, in **98** experimental groups, FCC significantly improves **94** of them by an average of 1.55% (**4.89% max** and 0.04% min).
- For ImageNet-LT, FCC improves the performance by an average 0.9% (**1.98% max** and 0.06% min).
- For iNaturalist 2018, FCC reinforces the performance by an average of 0.84% (**2.29% max** and 0.08% min).

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%
τ-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%**

Top-1 error rates comparisons between raw methods and those with FCC.W

Conclusions

- We tackle long-tailed visual recognition from a novel perspective of increasing the density of backbone features.

- We propose a Feature Clusters Compression (FCC) to improve the density of backbone features, and it can be easily achieved and friendly combined with existing long-tailed methods to boost them.

- Extensive experiments demonstrate FCC applied to existing methods achieves significant performance improvement and state-of-the-art results on four popular datasets.

Paper and Code

FCC: A simple and generic method for long-tailed visual recognition.

Paper and Code

- Paper link:
  https://openaccess.thecvf.com/content/CVPR2023/papers/Li_FCC_Feature_Clusters_Compression_for_Long-Tailed_Visual_Recognition_CVPR_2023_paper.pdf

- Code link:
  https://github.com/lijian16/FCC

Paper

Code

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

吉林大学
JILIN UNIVERSITY

THANK YOU!

Jian Li