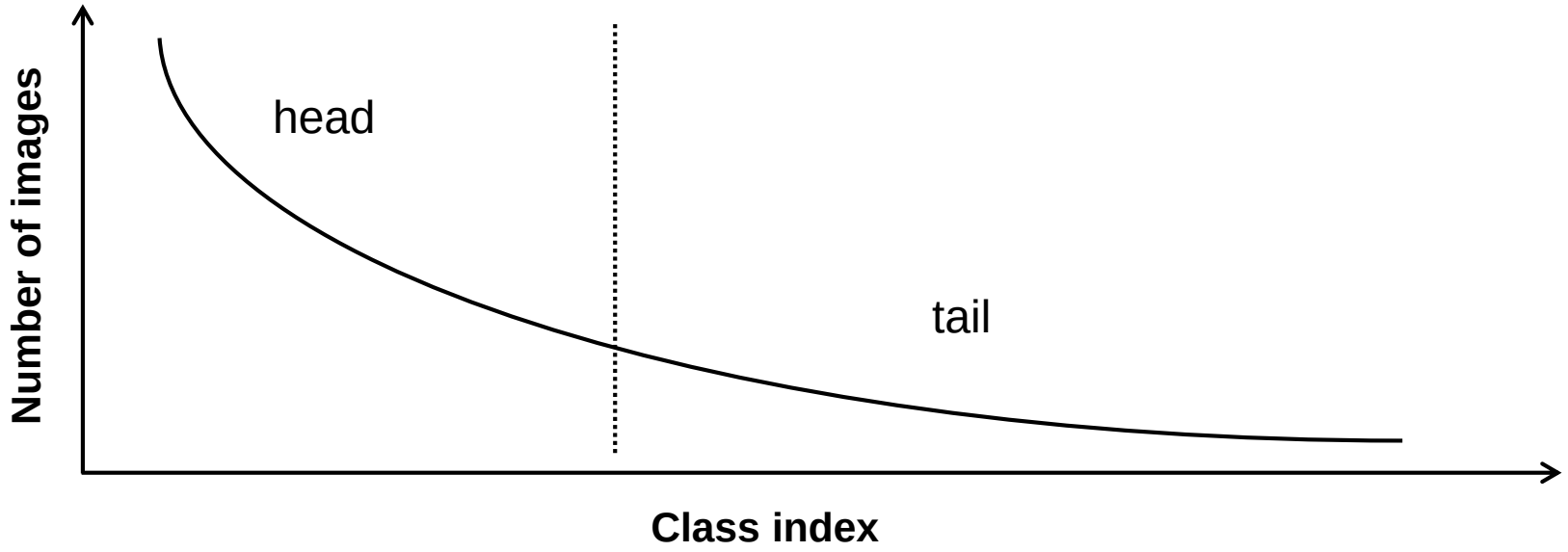
南京航空航天大學

Nanjing University of Aeronautics and Astronautics

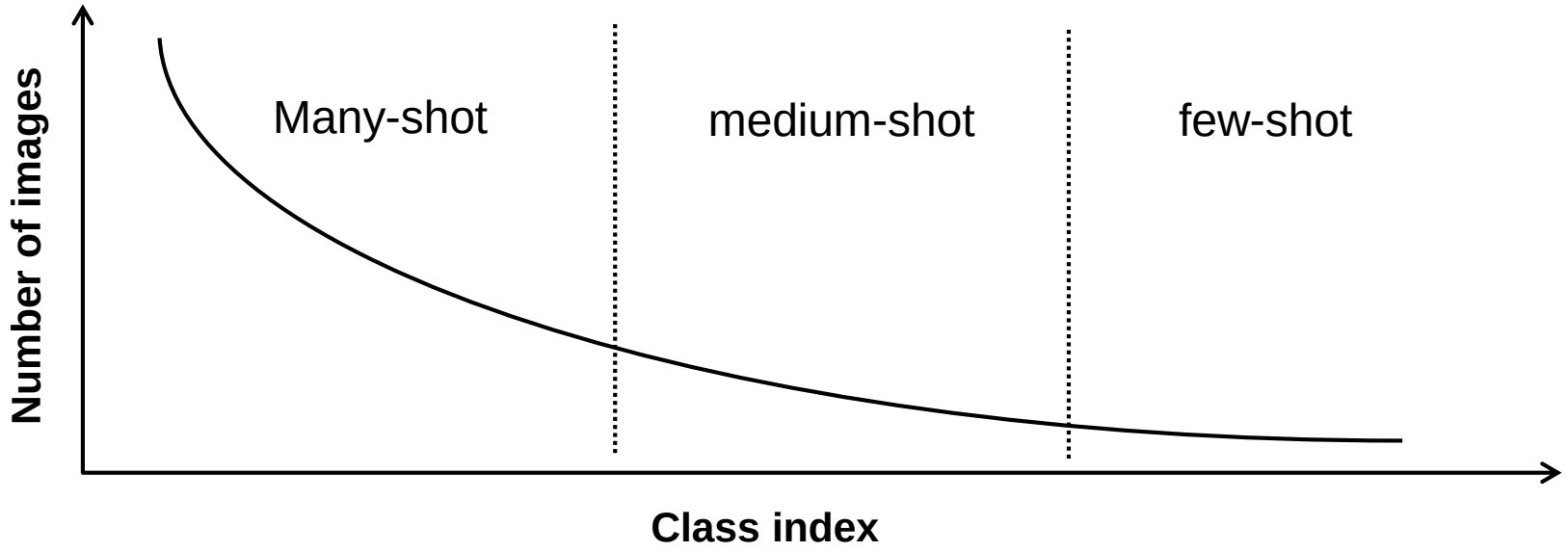
Long-Tailed Visual Recognition

QingHeng Zhang

Long-Tailed Distributions



Long-Tailed Distributions



RELATED WORKS

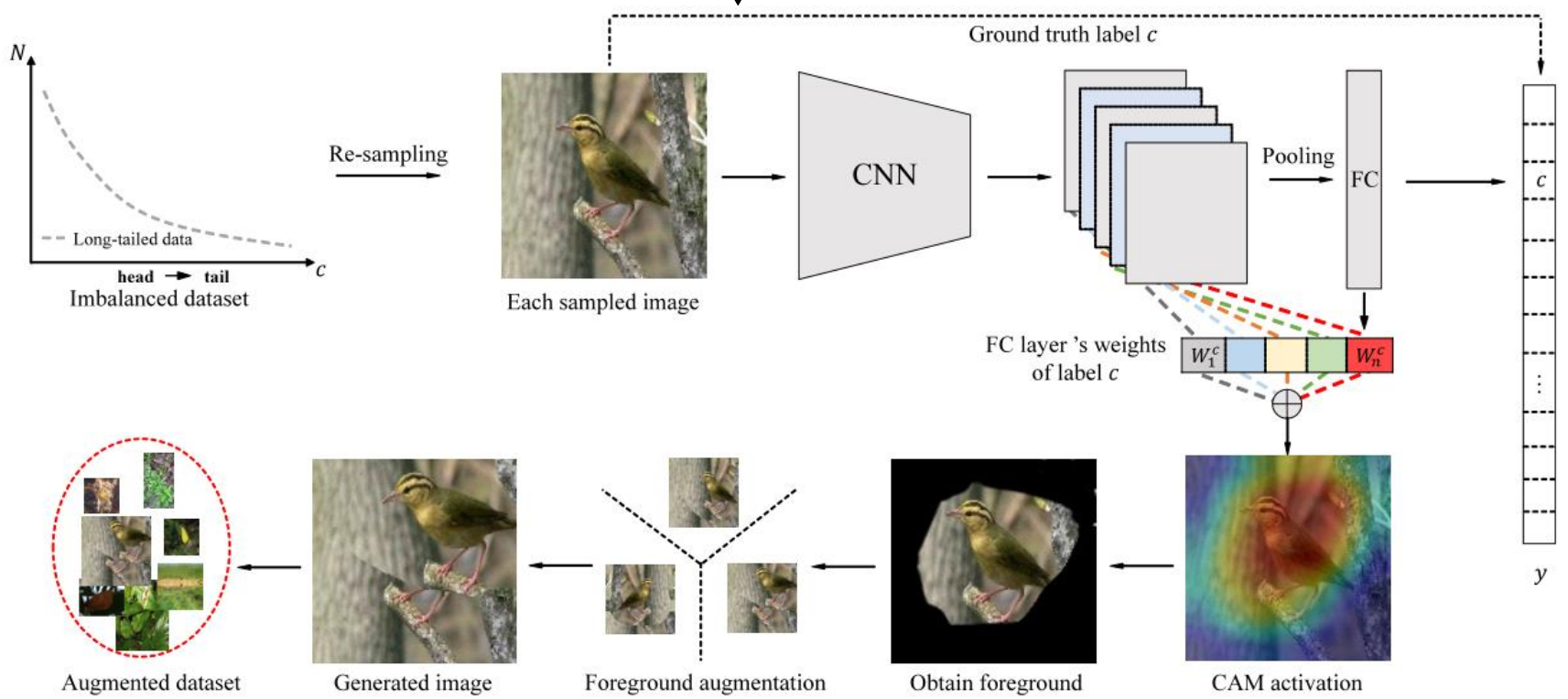
南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Re-sampling

- Random over/under sampling
- Distribution-aware sampling
  - Decoupling Representation and Classifier (2019) $\longrightarrow$ $p_j = \frac{n_j^q}{\sum_{i=1}^{C} n_i^q}$
- Class Activation Map based sampling
  - Bag of Tricks for Long-Tailed Visual Recognition(2021)

Re-weighting

Multi-Expert



RELATED WORKS

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

Re-sampling

- Random over/under sampling
- Distribution-aware sampling
  - Decoupling Representation and Classifier (2019)
- Class Activation Map based sampling
  - Bag of Tricks for Long-Tailed Visual Recognition(2021)

Re-weighting

- Focal loss (2017)
- Class-balanced loss (2019)
- LDAM loss (2019)
- Equalization loss (2020)

$$\mathrm{FL}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t)$$

$$\mathrm{CB}(\mathbf{p}, y) = \frac{1}{E_{n_y}}\mathcal{L}(\mathbf{p}, y) = \frac{1-\beta}{1-\beta^{n_y}}\mathcal{L}(\mathbf{p}, y)$$

$$\mathcal{L}_{\mathrm{LDAM}}((x, y); f) = -\log\frac{e^{z_y-\Delta_y}}{e^{z_y-\Delta_y}+\sum_{j\neq y}e^{z_j}}$$

$$\text{where } \Delta_j = \frac{C}{n_j^{1/4}} \text{ for } j \in \{1, \ldots, k\}$$

Multi-Expert

$$L_{SEQL} = -\sum_{j=1}^{C} y_j \log(\tilde{p}_j) \quad \text{where} \quad \tilde{p}_j = \frac{e^{z_j}}{\sum_{k=1}^{C}\tilde{w}_k e^{z_k}}$$

$$\tilde{w}_k = 1 - \beta T_\lambda(f_k)(1-y_k)$$

RELATED WORKS

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

Re-sampling

- Random over/under sampling
- Distribution-aware sampling
  - Decoupling Representation and Classifier (2019)
- Class Activation Map based sampling
  - Bag of Tricks for Long-Tailed Visual Recognition(2021)

Re-weighting

- Focal loss (2017)
- Class-balanced loss (2019)
- LDAM loss (2019)
- Equalization loss (2020)

Multi-Expert

- **LFME**: **L**earning **F**rom **M**ultiple **E**xperts (2020)
- **BBN**: **B**ilateral-**B**ranch **N**etwork (2020)
- **RIDE**: **R**out**I**ng **D**iverse **E**xperts (2021)

ours

RELATED WORKS

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

Re-sampling
- Random over/under sampling
- Distribution-aware sampling
  - Decoupling Representation and Classifier (2019)
- Class Activation Map based sampling
  - Bag of Tricks for Long-Tailed Visual Recognition(2021)

Re-weighting
- Focal loss (2017)
- Class-balanced loss (2019)
- LDAM loss (2019)
- Equalization loss (2020)

Multi-Expert
- **LFME**: **L**earning **F**rom **M**ultiple **E**xperts (2020)
- **BBN**: **B**ilateral-**B**ranch **N**etwork (2020)
- **RIDE**: **R**out**I**ng **D**iverse **E**xperts (2021)

ours

LFME:Learning From Multiple Experts

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

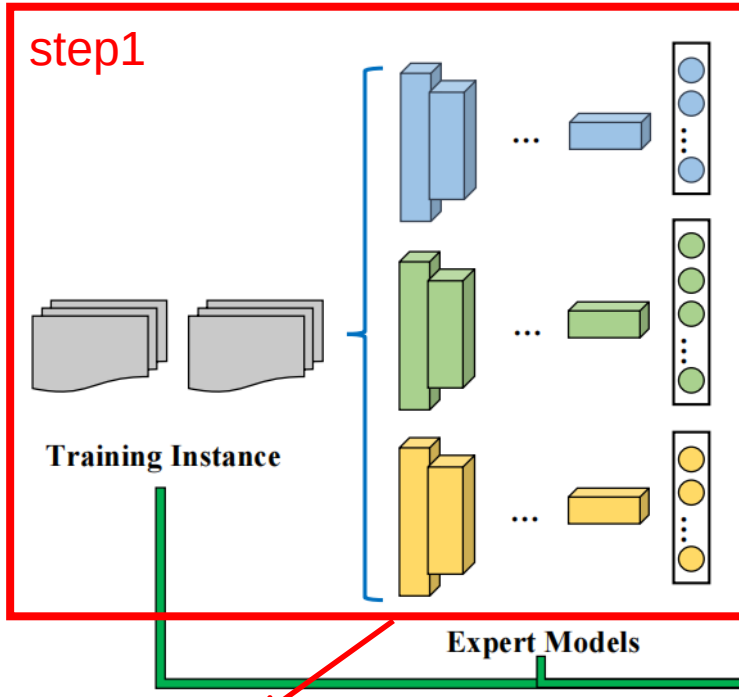
observation1:

Training a CNN on **a balanced dataset** with **fewer samples** sometimes leads to superior performances than on a long-tailed dataset with more samples.

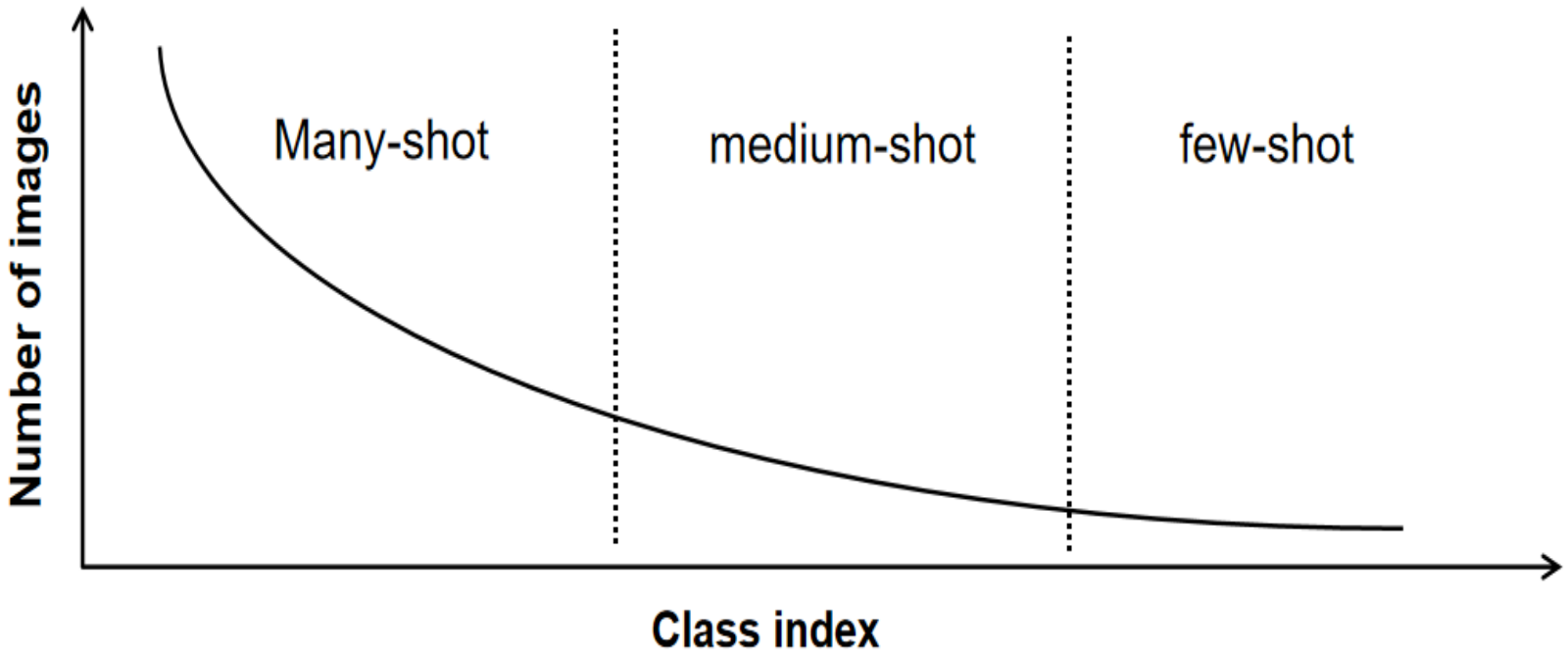
step1. Train “Experts”

1) provide output logits’ distribution for knowledge transfer
2) provide output confidence as instance difficulty cues.

setp2. Train “Student” based on prior knowledge(Experts)



sorted by numbers



LFME:Learning From Multiple Experts

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

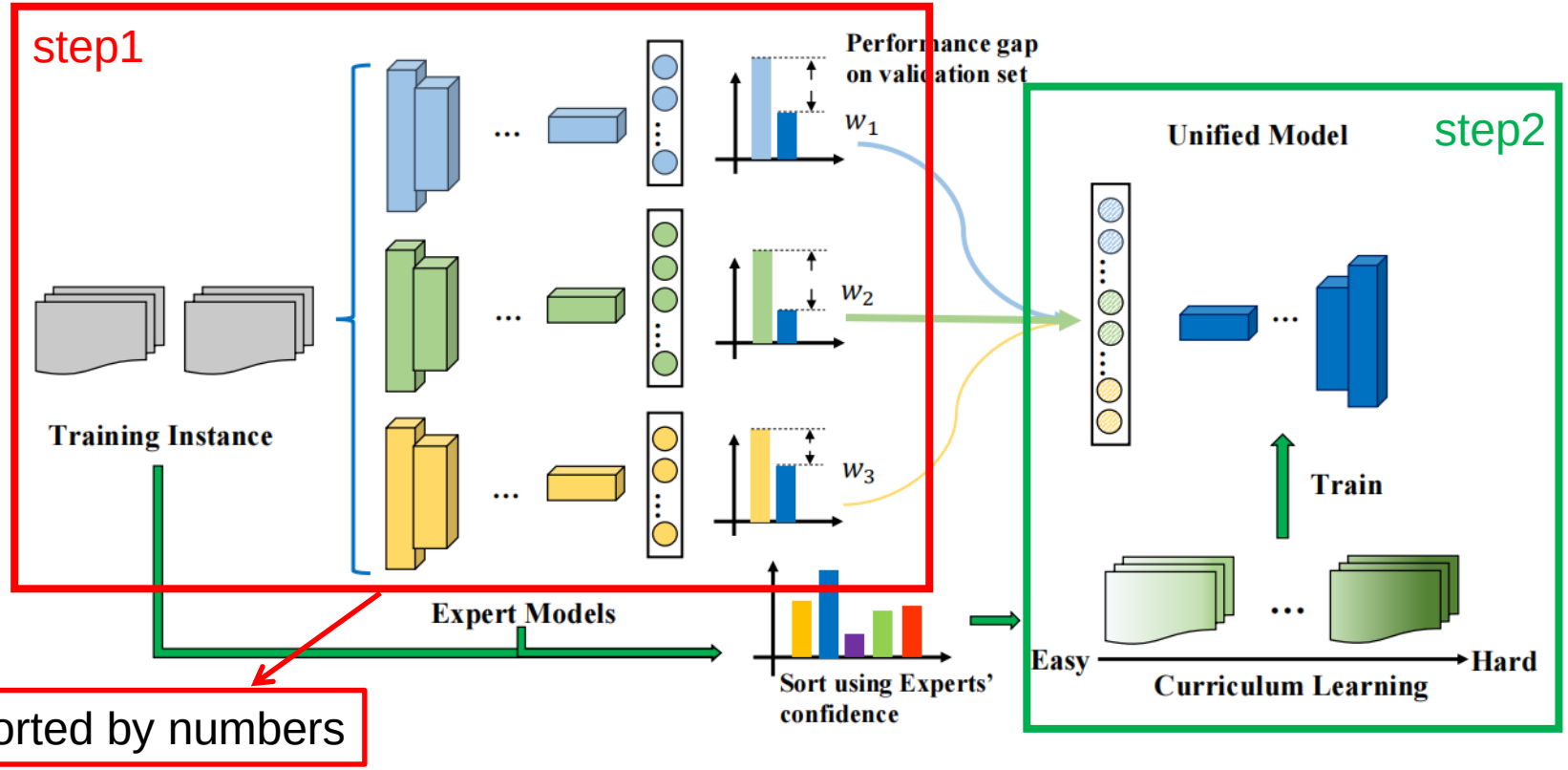
observation:

Training a CNN on **a balanced dataset** with **fewer samples** sometimes leads to superior performances than on a long-tailed dataset with more samples.

step1. Train “Experts”

1) provide output logits’ distribution for knowledge transfer
2) provide output confidence as instance difficulty cues.

setp2. Train “Student” based on prior knowledge(Experts)



RELATED WORKS

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

Re-sampling
- Random over/under sampling
- Distribution-aware sampling
  - Decoupling Representation and Classifier (2019)
- Class Activation Map based sampling
  - Bag of Tricks for Long-Tailed Visual Recognition(2021)

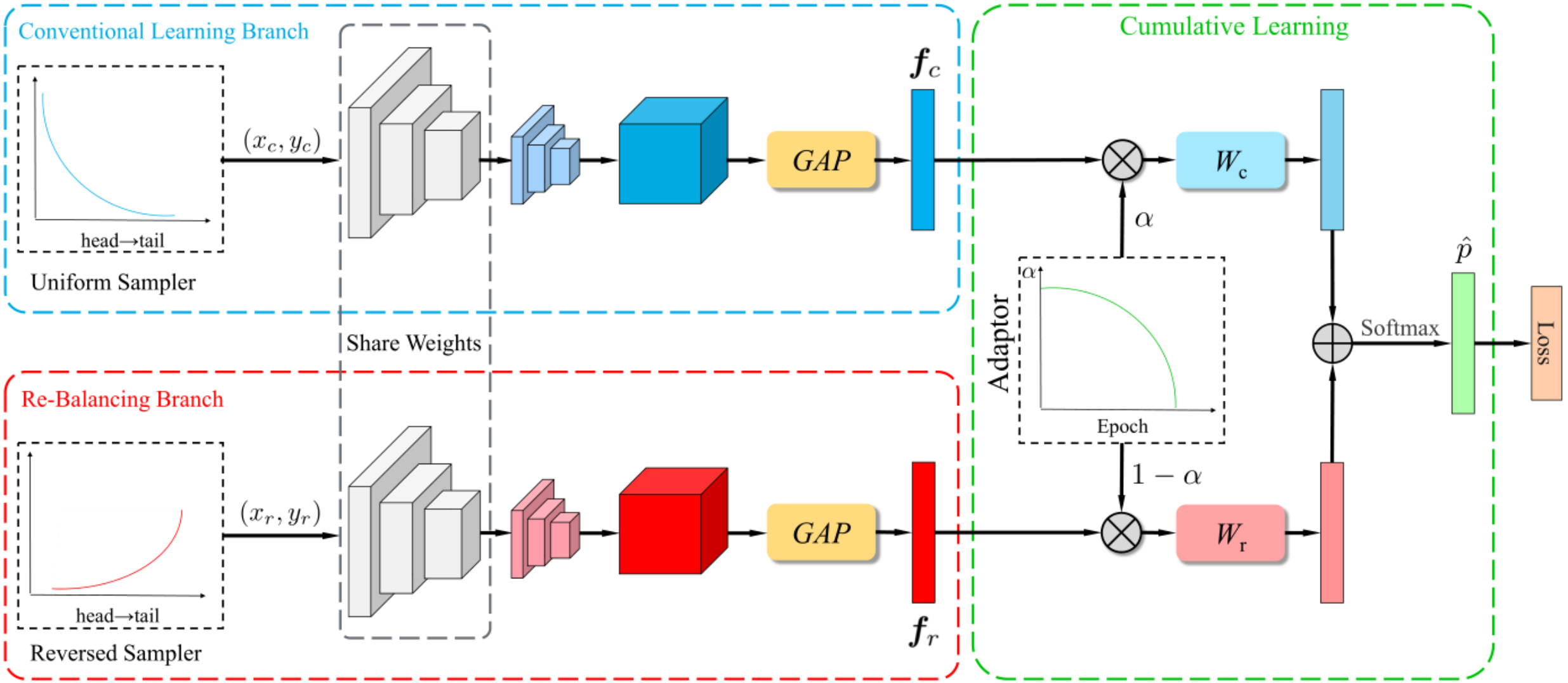
Re-weighting
- Focal loss (2017)
- Class-balanced loss (2019)
- LDAM loss (2019)
- Equalization loss (2020)

Multi-Expert
- **LFME**: **L**earning **F**rom **M**ultiple **E**xperts (2020)
- **BBN**: **B**ilateral-**B**ranch **N**etwork (2020)
- **RIDE**: **R**out**I**ng **D**iverse **E**xperts (2021)

ours

BBN: Bilateral-Branch Network

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

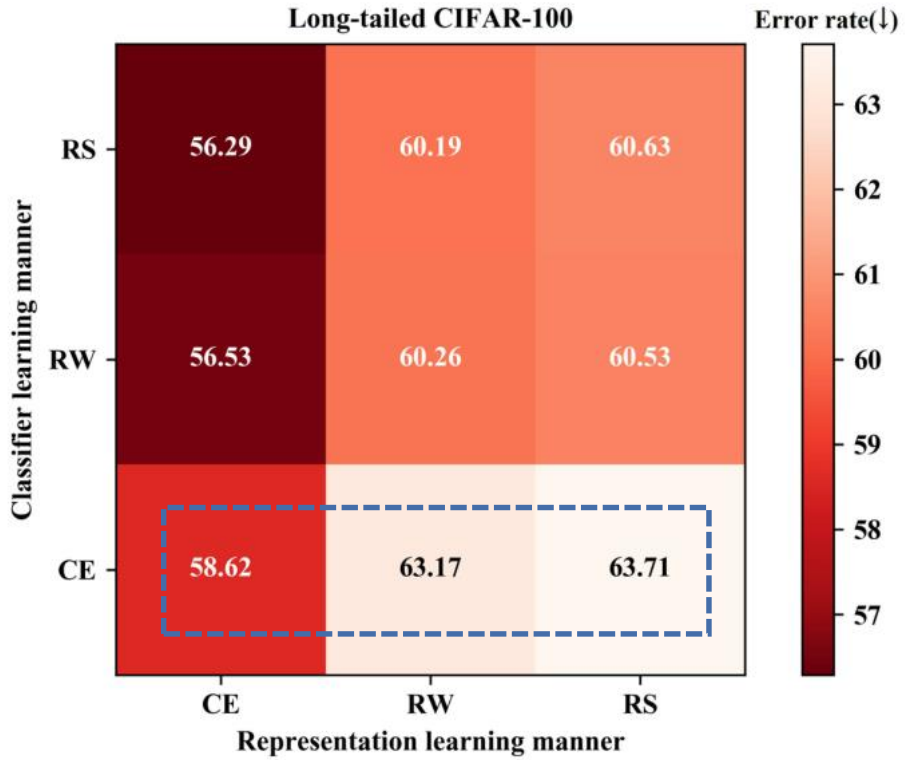


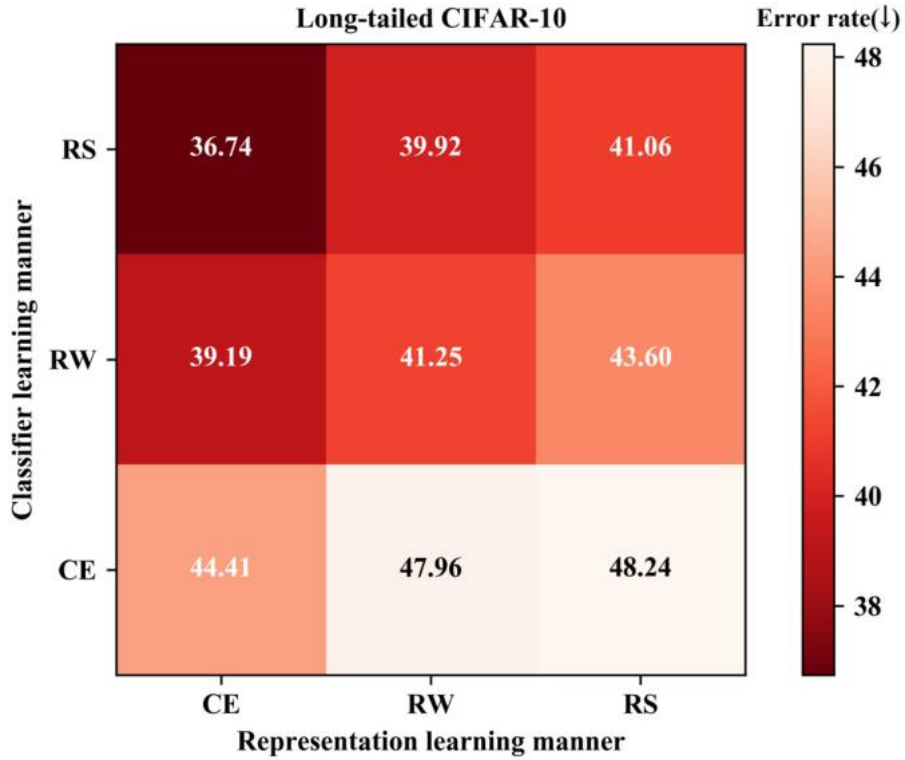
BBN: Bilateral-Branch Network

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

observation:

Long-tailed dataset can learn a very good feature representation.





Summary

LFME:

Each expert focuses on a relatively balanced group such as **many-shot** classes, **medium-shot** classes, and **few-shot** classes.

BBN:

Each expert focuses on the **head** and the **tail** classes respectively.

RIDE: (evaluation)

Each expert in **BBN** and **LFME** has **no** balanced access to the **entire** dataset.

What did RIDE do ?

RELATED WORKS

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

Re-sampling
- Random over/under sampling
- Distribution-aware sampling
  - Decoupling Representation and Classifier (2019)
- Class Activation Map based sampling
  - Bag of Tricks for Long-Tailed Visual Recognition(2021)

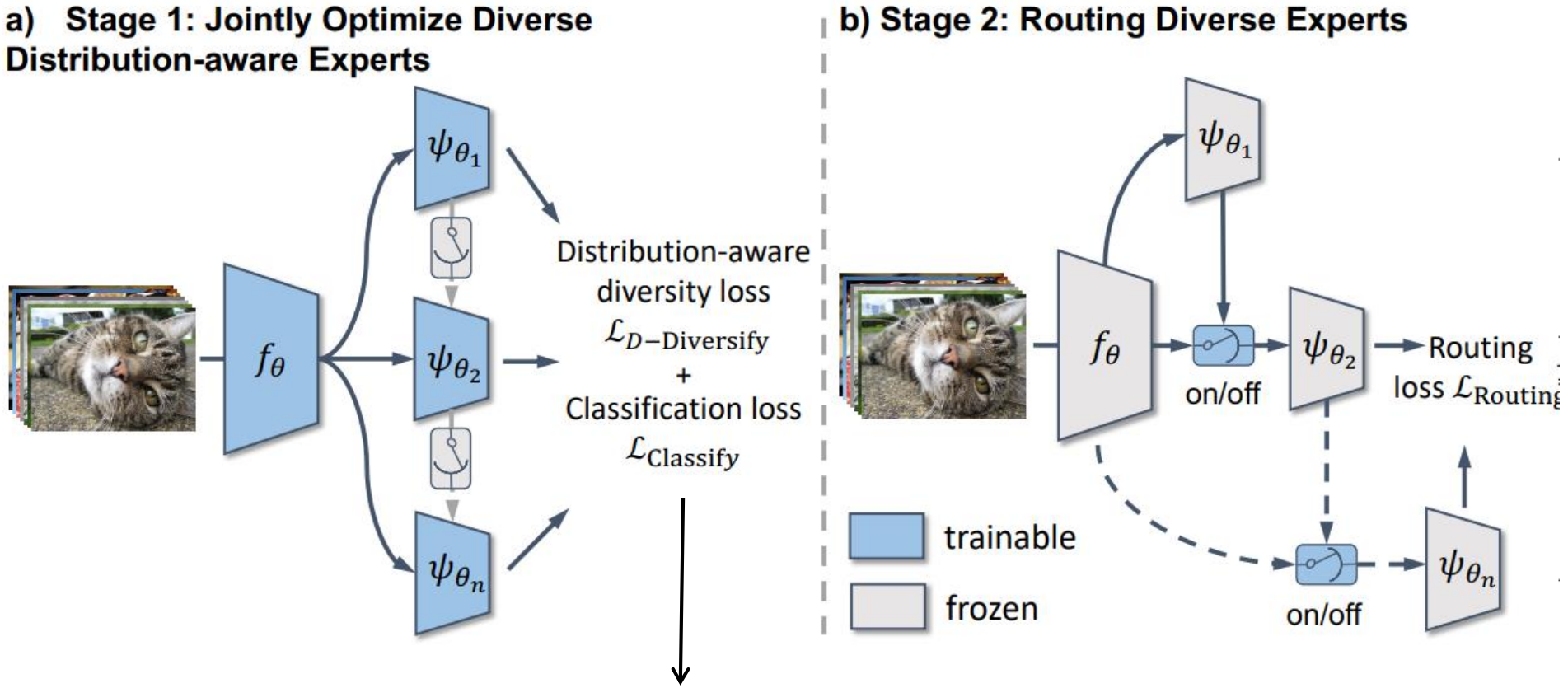
Re-weighting
- Focal loss (2017)
- Class-balanced loss (2019)
- LDAM loss (2019)
- Equalization loss (2020)

Multi-Expert
- **LFME**: **L**earning **F**rom **M**ultiple **E**xperts (2020)
- **BBN**: **B**ilateral-**B**ranch **N**etwork (2020)
- **RIDE**: **R**out**I**ng **D**iverse **E**xperts (2021)

ours

RIDE: RoutIng Diverse Experts

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



$$\mathcal{L}^{\text{individual}}(x, y) = \sum_{i=1}^{n} \mathcal{L}\left(\psi_{\theta_i}\left(f_\theta(x)\right), y\right)$$

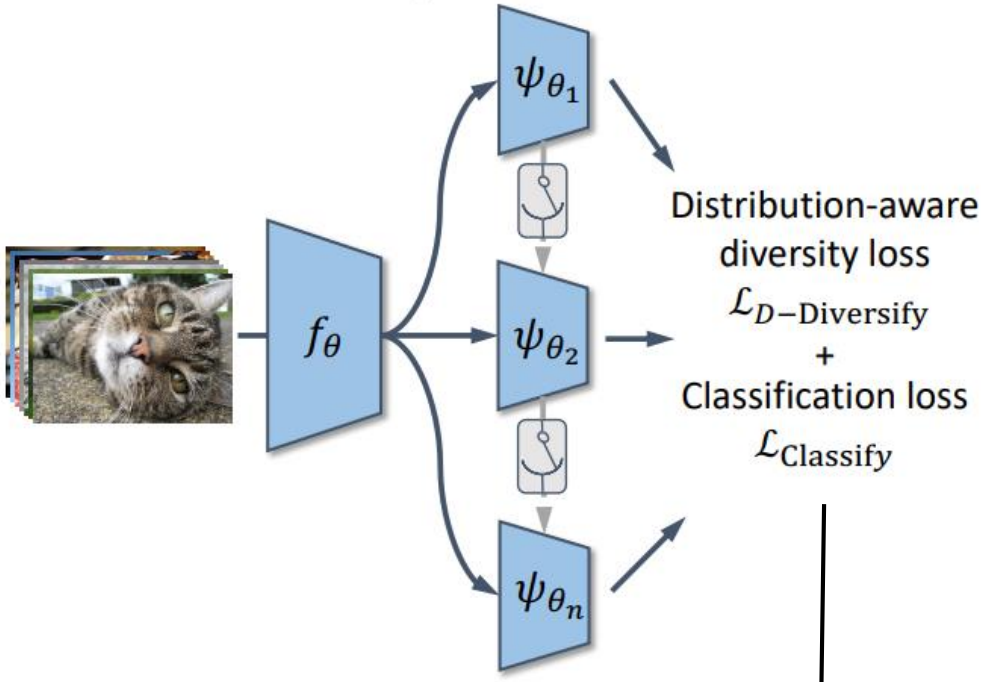
$$\mathcal{L}^{\text{collaborative}}(x, y) = \mathcal{L}\left(\frac{1}{n}\sum_{i=1}^{n} \psi_{\theta_i}\left(f_\theta(x)\right), y\right)$$

RIDE: RoutIng Diverse Experts

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

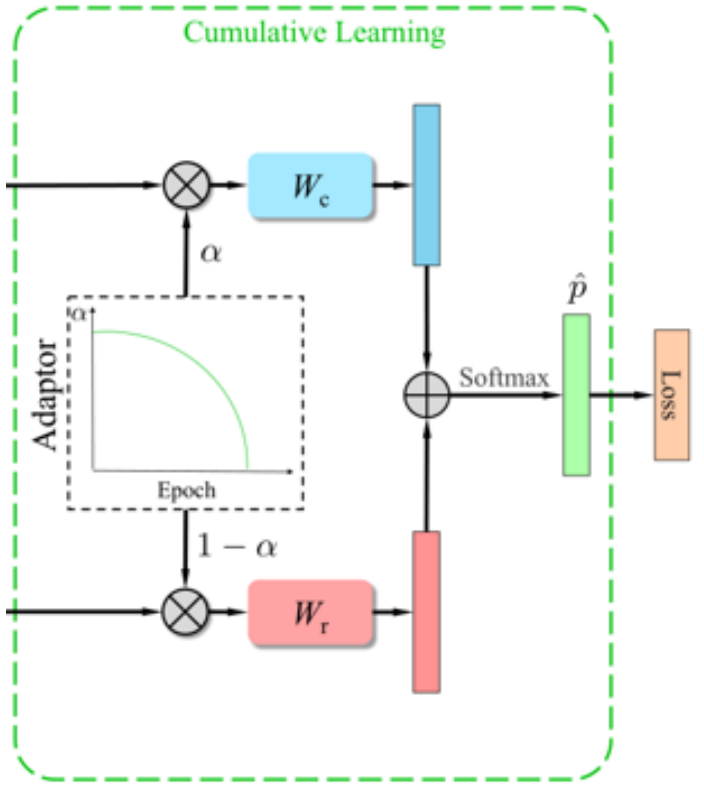
RIDE:

a) Stage 1: Jointly Optimize Diverse Distribution-aware Experts



$$\mathcal{L}^{\text{individual}}(x, y) = \sum_{i=1}^{n} \mathcal{L}\left(\psi_{\theta_i}\left(f_\theta(x)\right), y\right)$$

BBN:



$$\mathcal{L}^{\text{collaborative}}(x, y) = \mathcal{L}\left(\frac{1}{n}\sum_{i=1}^{n} \psi_{\theta_i}\left(f_\theta(x)\right), y\right)$$

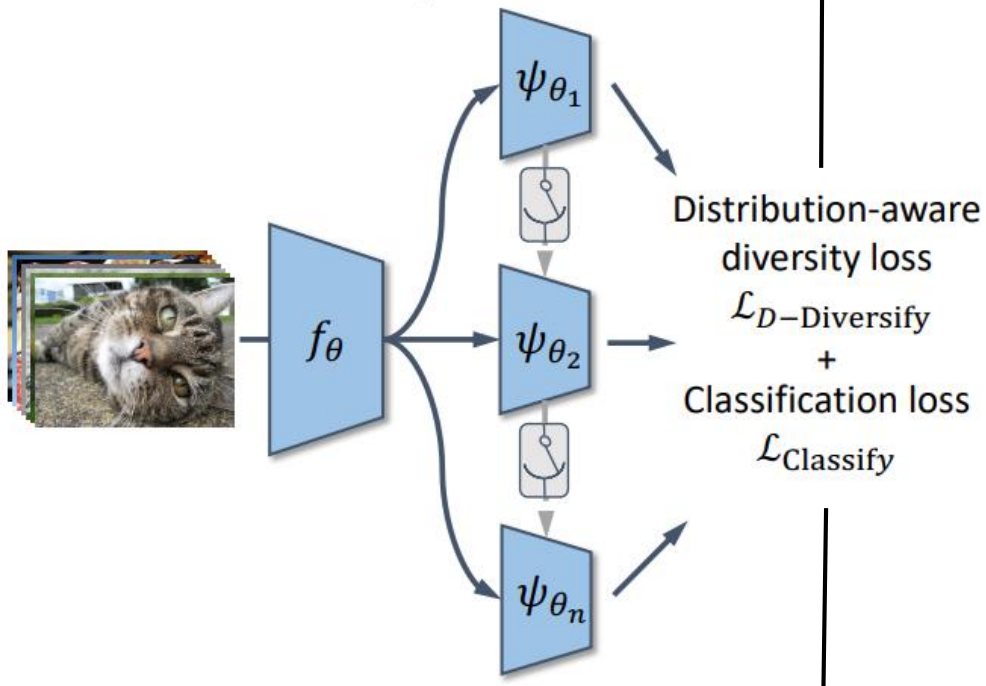
RIDE: RoutIng Diverse Experts

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

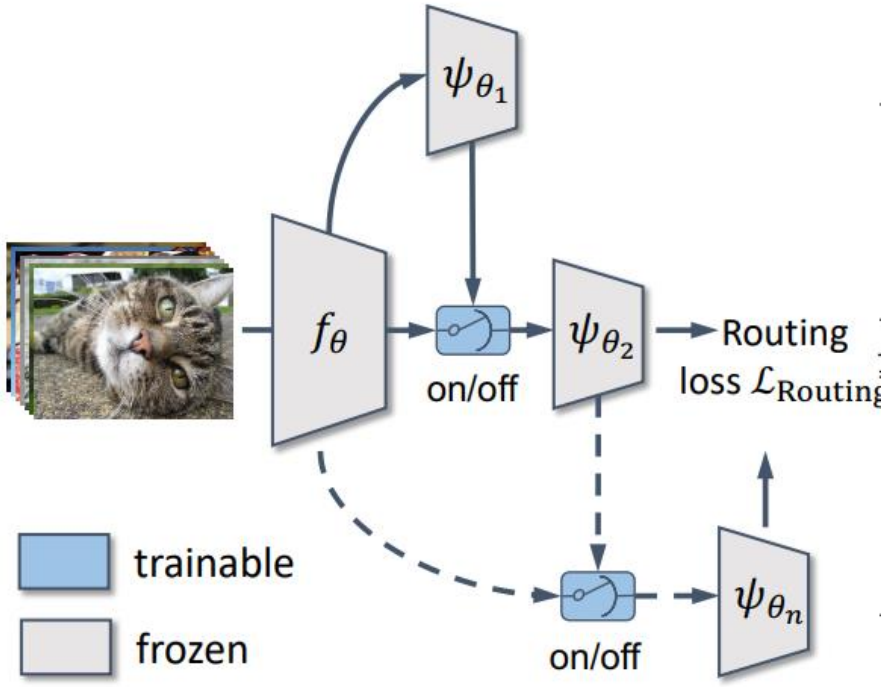
$$\mathcal{L}_{\text{D-Diversify}}(x, y; \theta_i) = \frac{-1}{n-1} \sum_{j=1, j\neq i}^{n} D_{\text{KL}}\left(\mathbf{p}^{(i)}(x, y) \| \mathbf{p}^{(j)}(x, y)\right)$$

Each expert focuses on different classes.

a) Stage 1: Jointly Optimize Diverse Distribution-aware Experts



b) Stage 2: Routing Diverse Experts



$$\mathcal{L}^{\text{individual}}(x, y) = \sum_{i=1}^{n} \mathcal{L}\left(\psi_{\theta_i}\left(f_\theta(x)\right), y\right)$$

$$\mathcal{L}^{\text{collaborative}}(x, y) = \mathcal{L}\left(\frac{1}{n}\sum_{i=1}^{n} \psi_{\theta_i}\left(f_\theta(x)\right), y\right)$$

Summary

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

LFME:

Each expert focuses on a relatively **balanced** group such as many-shot classes, medium-shot classes, and few-shot classes.

BBN:

Each expert focuses on the **head** and the **tail** classes respectively.

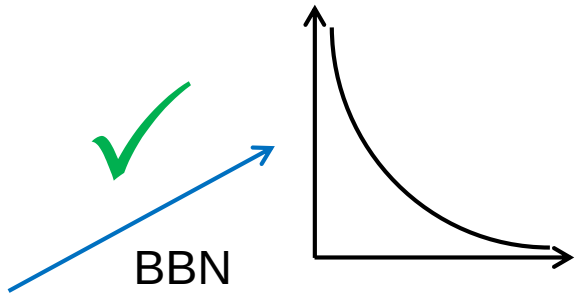
RIDE: (evaluation)

Each expert in **BBN** and **LFME** has **no** balanced access to the **entire** dataset.

What did RIDE do ?

RIDE:

Each expert in **RIDE** has a balanced access to the **entire** dataset, and each expert focuses on **different** classes.



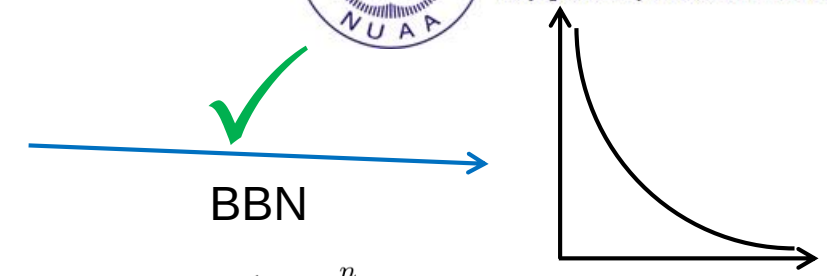
$$\mathcal{L}_{\text{D-Diversify}}(x, y; \theta_i) = \frac{-1}{n-1} \sum_{j=1, j\neq i}^{n} D_{\text{KL}}\left(\mathbf{p}^{(i)}(x, y) \| \mathbf{p}^{(j)}(x, y)\right)$$

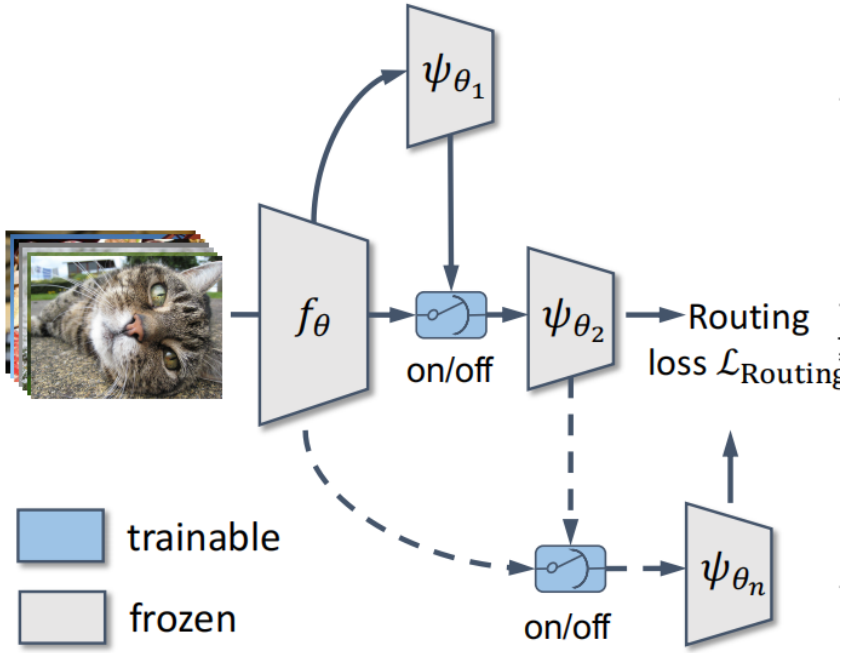
Summary

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

RIDE:

Each expert in **RIDE** has a balanced access to the **entire** dataset, and each expert focuses on different classes.





$$\mathcal{L}_{\text{D-Diversify}}(x, y; \theta_i) = \frac{-1}{n-1} \sum_{j=1, j\neq i}^{n} D_{\text{KL}}\left(\mathbf{p}^{(i)}(x, y) \| \mathbf{p}^{(j)}(x, y)\right)$$

0.4 < threshold(0.5)

(0.8 + 0.4) / 2 = 0.6 > threshold(0.5)

Diversity → KL

Train each expert independently

RELATED WORKS

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

Re-sampling
- Random over/under sampling
- Distribution-aware sampling
  - Decoupling Representation and Classifier (2019)
- Class Activation Map based sampling
  - Bag of Tricks for Long-Tailed Visual Recognition(2021)

Re-weighting
- Focal loss (2017)
- Class-balanced loss (2019)
- LDAM loss (2019)
- Equalization loss (2020)

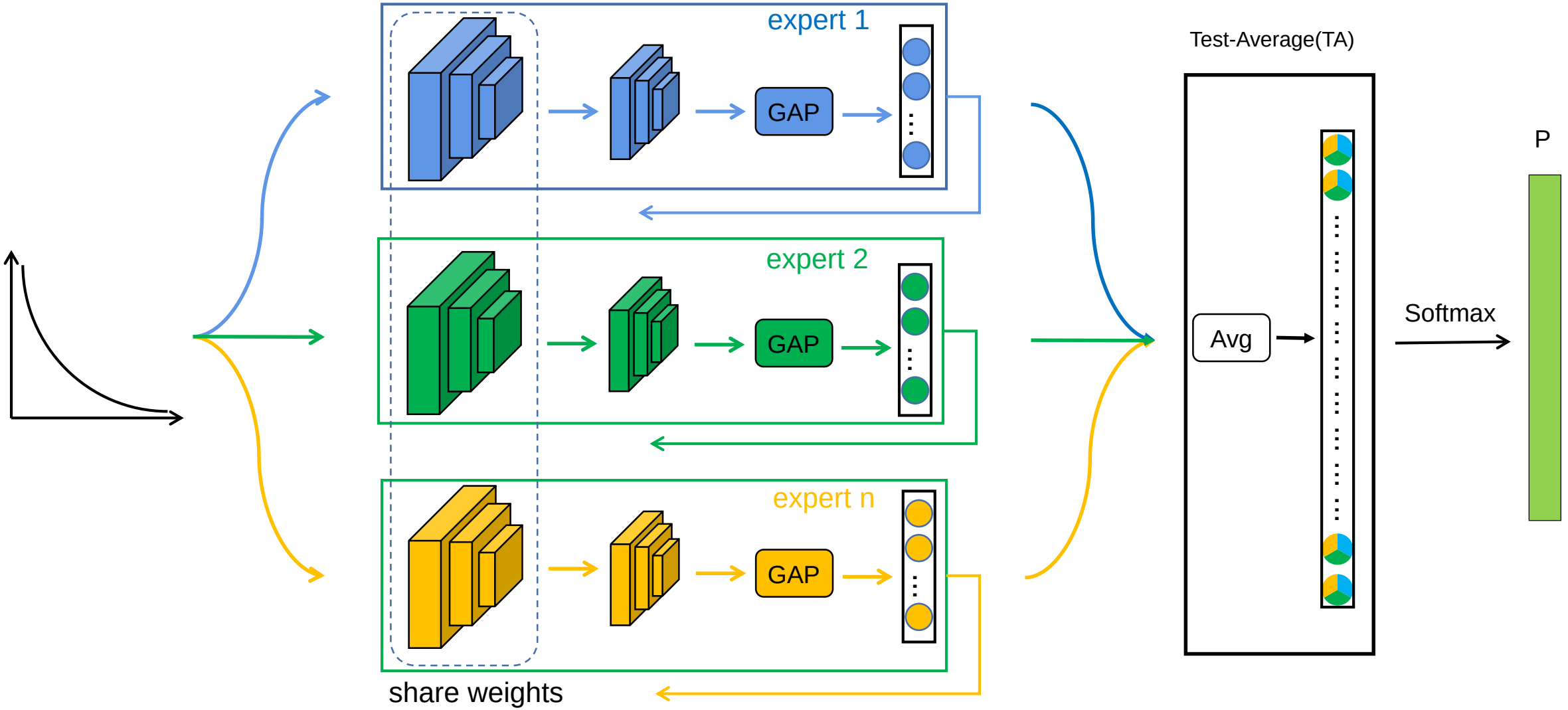
Multi-Expert
- **LFME**: **L**earning **F**rom **M**ultiple **E**xperts (2020)
- **BBN**: **B**ilateral-**B**ranch **N**etwork (2020)
- **RIDE**: **R**out**I**ng **D**iverse **E**xperts (2021)

→ ours

Our Framwork

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

Insight : Training each expert on the **entire** dataset and **independently.**



Comparisons

Method	Data for experts	learning manner of experts
LFME	independent	independent
BBN	reverse	dependent
RIDE	same	dependent
Ours	same	independent

Table 1.Comparisons between Ours and three SOTA multi-expert networks.

Experiments

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

Method	Long-tailed CIFAR10				Long-tailed CIFAR100			
Imbalance radio	200	100	50	10	200	100	50	10
Class Balance(19)	68.89	74.57	79.27	87.49	36.23	39.60	45.32	57.99
LDAM(19)	-	77.03	-	88.16	-	42.04	-	58.71
Equalization(20)	-	-	-	-	-	43.38	-	-
LFME(20)	-	-	-	-	-	43.8	-	-
BBN(20)	-	79.82	82.18	88.32	-	42.56	47.02	59.12
RIDE(21)	-	-	-	-	-	49.1	-	-
CAM(21)	-	80.03	83.59	-	-	47.83	51.69	-
SOTA	68.89	80.03	83.59	88.32	36.23	49.1	51.69	59.12
Ours	**77.87**	**82.19**	**85.14**	**90.15**	**46.82**	**51.53**	**56.23**	**65.27**

Table 2.The top-1 accuracy on Cifar10,100-LT with ResNet-32.

All results are copied from papers.

Experiments

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

Method	Long-tailed ImageNet
OLTR(19)	35.60
LDAM(19)	36.00
Decoupling(19)	41.80
Feature space augmentation(20)	35.20
Equalization Loss(20)	36.44
LFME(20)	38.80
Class Activation Map(21)	43.13
Ours	**43.60**

Table 3.The top-1 accuracy on ImageNet-LT with ResNet-10.

All results are copied from papers.

Ablation

Method	Long-tailed CIFAR10				Long-tailed CIFAR100			
SOTA	68.89	80.03	83.59	88.32	36.23	49.1	51.69	59.12
Ours(w/o diversity)	69.28	76.21	81.46	87.48	37.53	42.04	48.10	59.00
Ours(w/o TA)	75.00	79.91	83.11	88.64	43.26	47.36	51.76	61.86
Ours(diversity+TA**)**	**77.87**	**82.19**	**85.14**	**90.15**	**46.82**	**51.53**	**56.23**	**65.27**

Table 4.Ablation study on diversity and TA on Cifar10,100-LT with ResNet-32.

Experiments

cifar10
Imbalance radio = 50

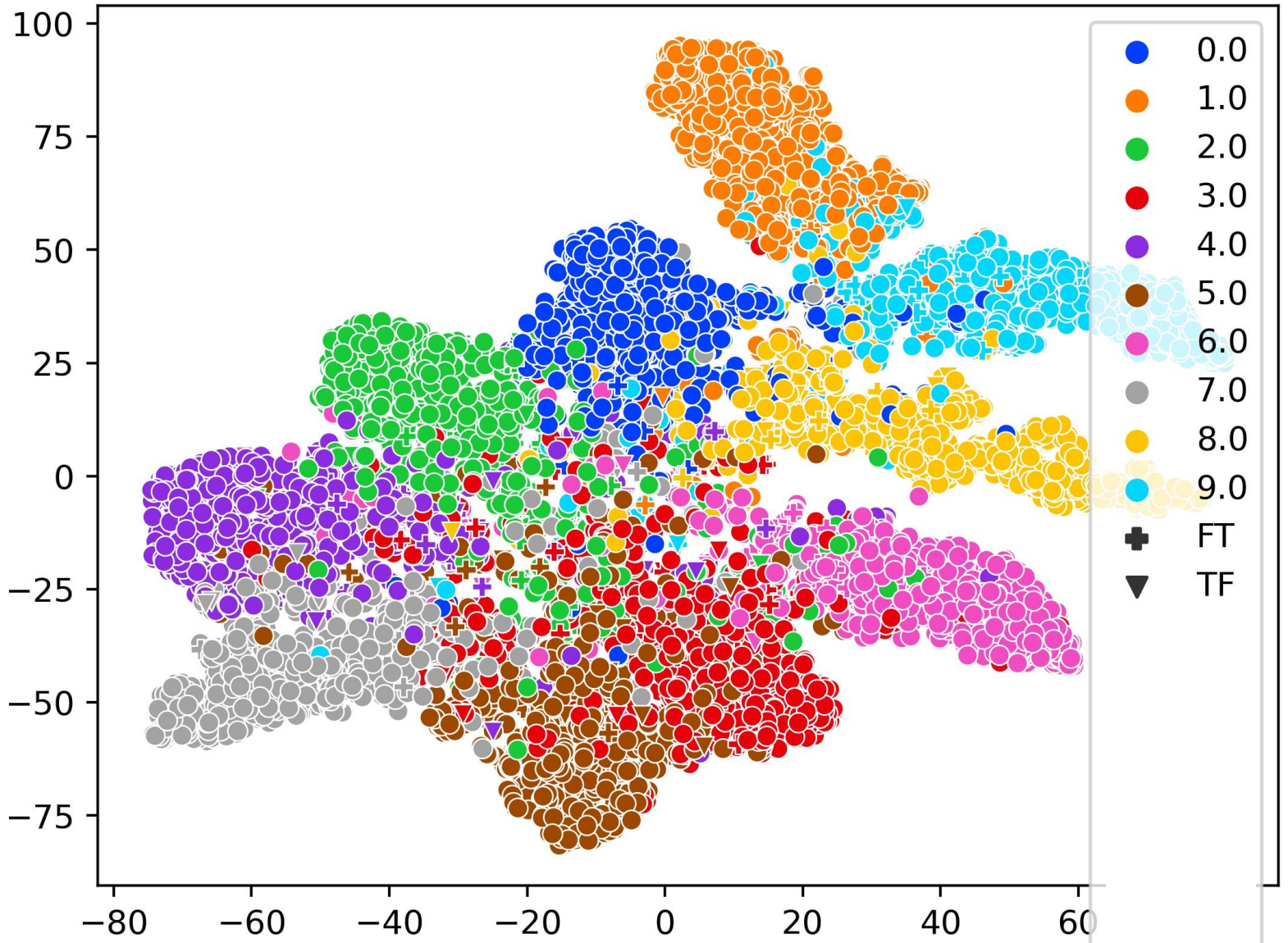
```
avg Result: Prec@1 85.140 Prec@5 99.180
avg Class Accuracy: [0.962,0.983,0.846,0.798,0.867,0.795,0.846,0.800,0.819,0.798]
**********************************************************************************
each experts:
expert1 Result: Prec@1 82.730 Prec@5 98.850
expert1 Class Accuracy: [0.941,0.965,0.803,0.763,0.830,0.783,0.820,0.777,0.815,0.776]
expert2 Result: Prec@1 81.820 Prec@5 99.020
expert2 Class Accuracy: [0.950,0.975,0.820,0.749,0.842,0.745,0.821,0.757,0.777,0.746]
expert3 Result: Prec@1 82.660 Prec@5 98.960
expert3 Class Accuracy: [0.948,0.977,0.812,0.775,0.838,0.766,0.823,0.766,0.779,0.782]
**********************************************************************************
```

expert1 VS Avg :
TF :expert True and Avg Flase 237
FT :expert False and Avg True 478

(478-237)/100=2.41
82.73 + 2.41 = 85.14

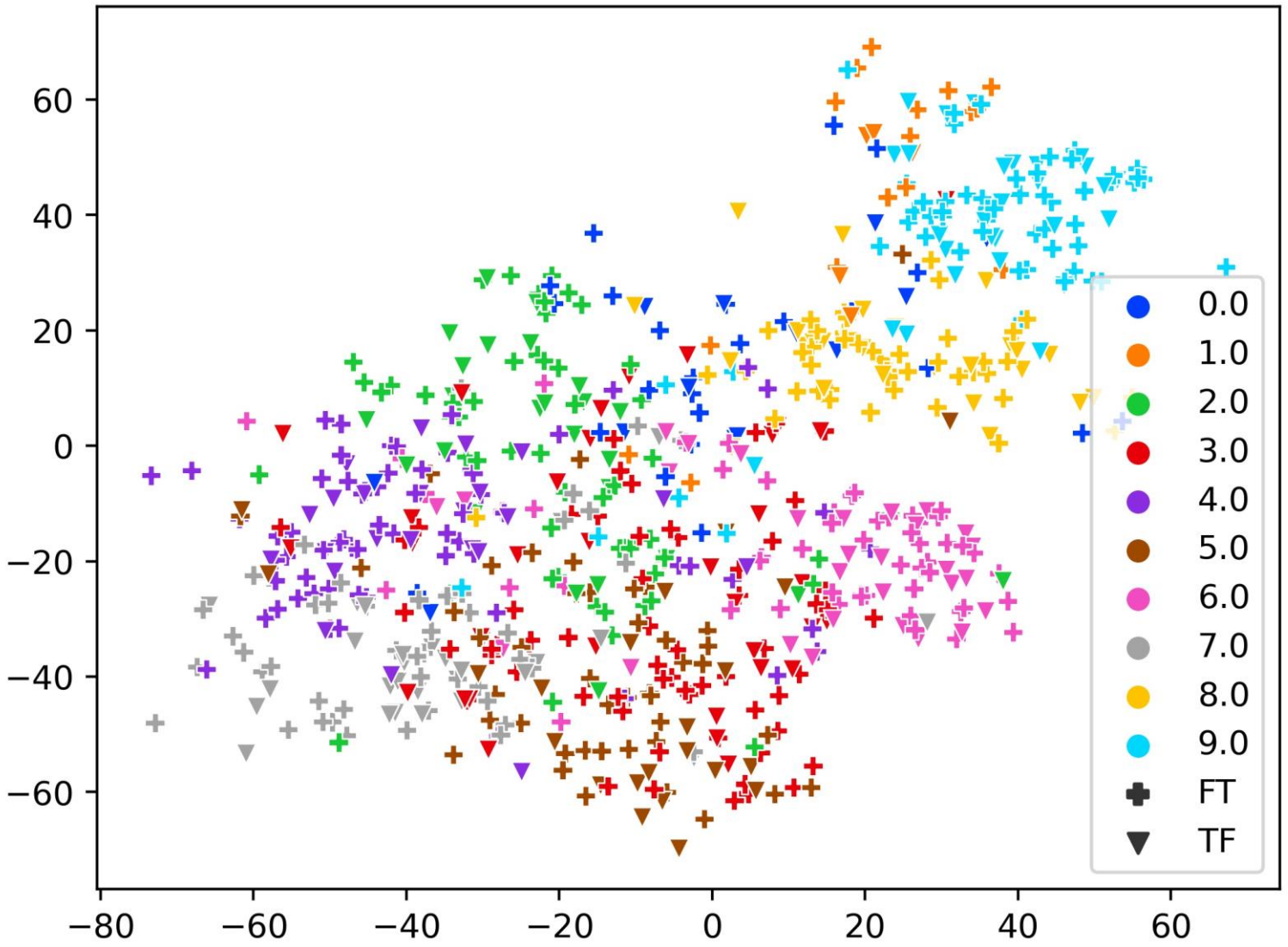
Experiments

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



Experiments

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



: 478

: 237

Summary

LFME:

Each expert focuses on a relatively balanced group such as **many-shot** classes, **medium-shot** classes, and **few-shot** classes.

BBN:

Each expert focuses on the **head** and the **tail** classes respectively.

RIDE:

Each expert has a balanced access to the **entire** dataset, and each expert focuses on **different** classes.

Ours:

Each expert has a balanced access to the **entire** dataset, and each expert focuses on **all** classes.

Advantages

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

- The independent **training manner** brings **diversity** to experts, and the Test-Average(TA) makes the network more **robust**.
- The **diversity** and the **Test-Average** are working at the same time.
- Our network is **One-stage** model, while LFME and RIDE aren’t.
- Our network outperforms **SOTA** methods.

THANKS