ICLR

Towards Calibrated Deep Clustering Network

Yuheng Jia[1,2] Jianhong Cheng[1], Hui Liu[2], Junhui Hou[3]

1 School of Computer Science and Engineering, Southeast University

2 School of Computing & Information Sciences, Saint Francis University

3 Department of Computer Science, City University of Hong Kong

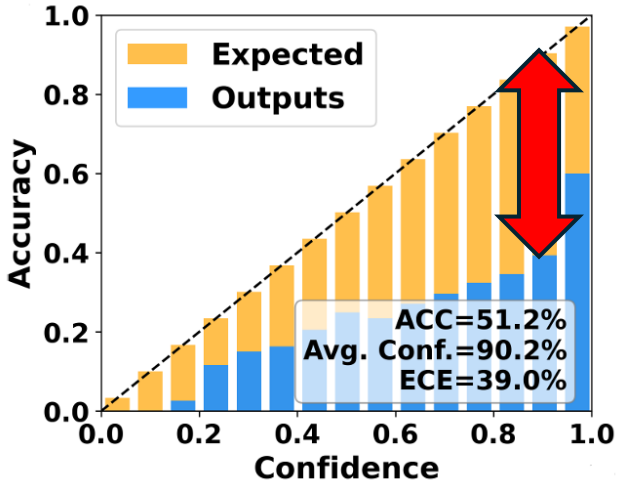
Deep Clustering

Crucial Task

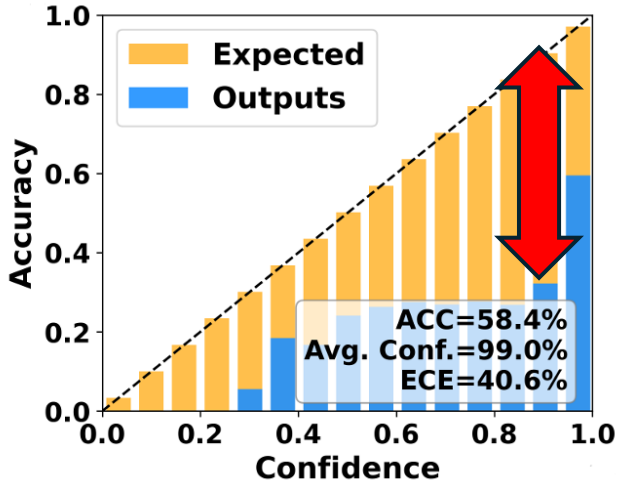
- computer vision
- medical diagnostics
- autonomous driving

Challenge

- Overconfidence. Model confidence scores are significantly higher than the actual prediction accuracy.
- Fixed confidence thresholds introduce more noisy pseudo-labels as confidence increases during training.



Reliability diagram of SCAN[1]



Reliability diagram of SPICE[2]

[1] Scan: Learning to classify images without labels. ECCV20.
[2] Spice: Semantic pseudo-labeling for image clustering. TIP22.

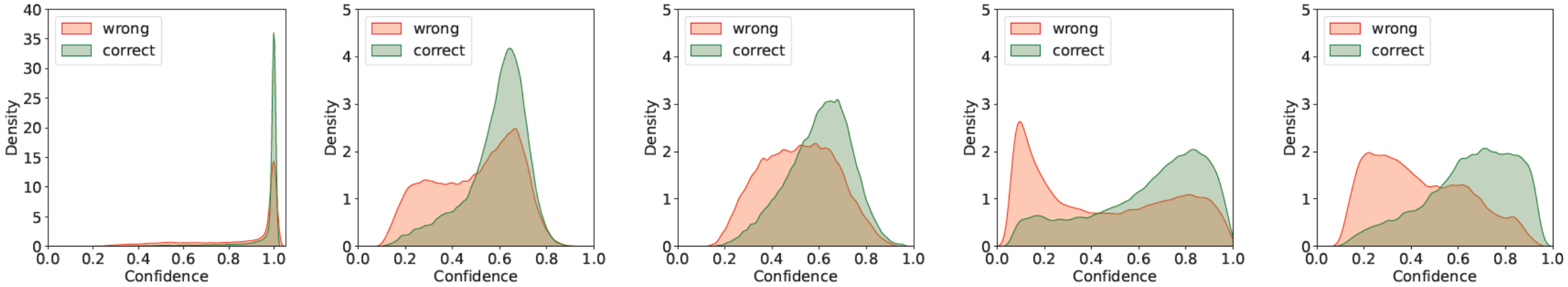
Existing Methods

Post-calibration methods

- The requirement for a labeled validation set does not hold in deep clustering.

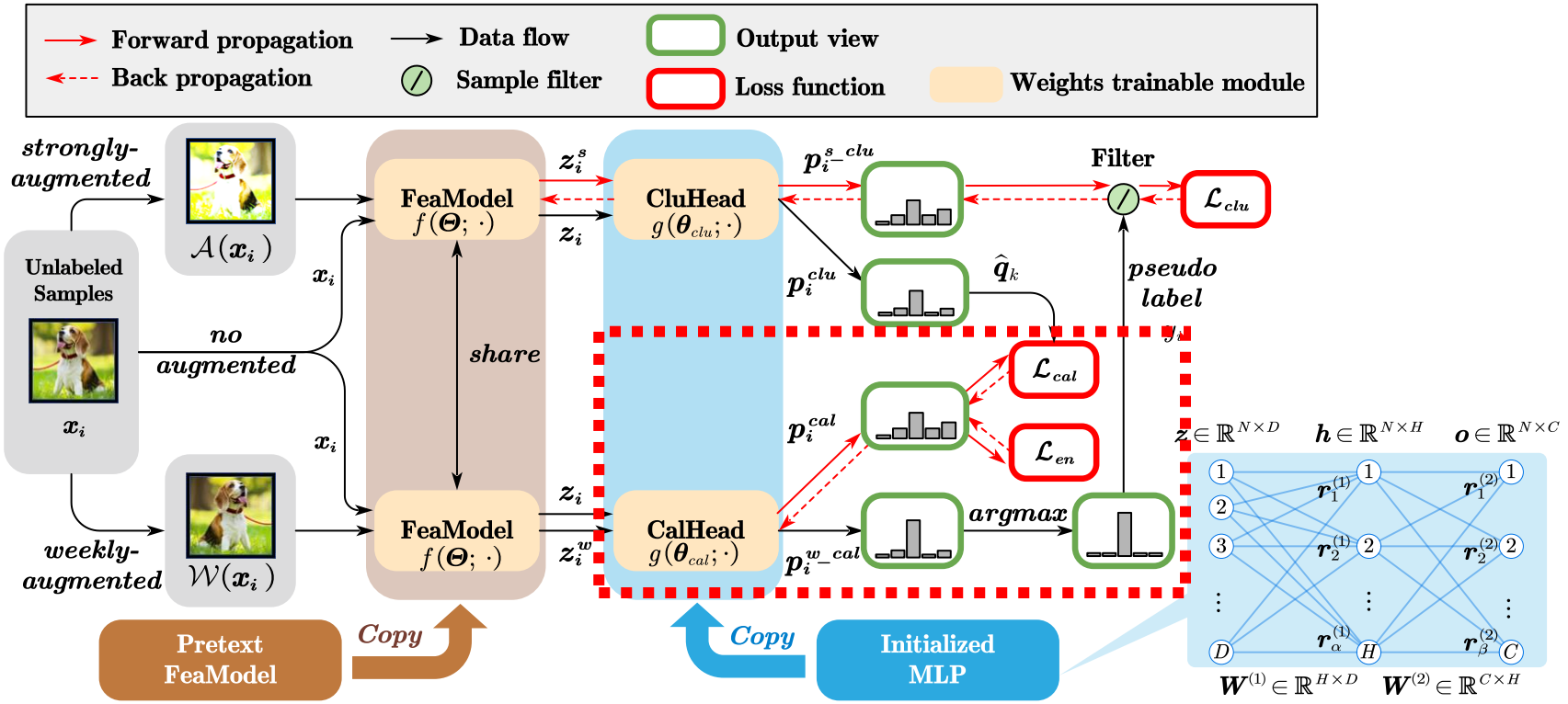
Regularization-based methods

- Label Smoothing (LS), Focal Loss (FL) and L1 Norm (L1) penalizes all predictions, including correct ones.



(a) SCAN (b) LS (c) FL (d) L1 (e) CDC-Cal(Ours)

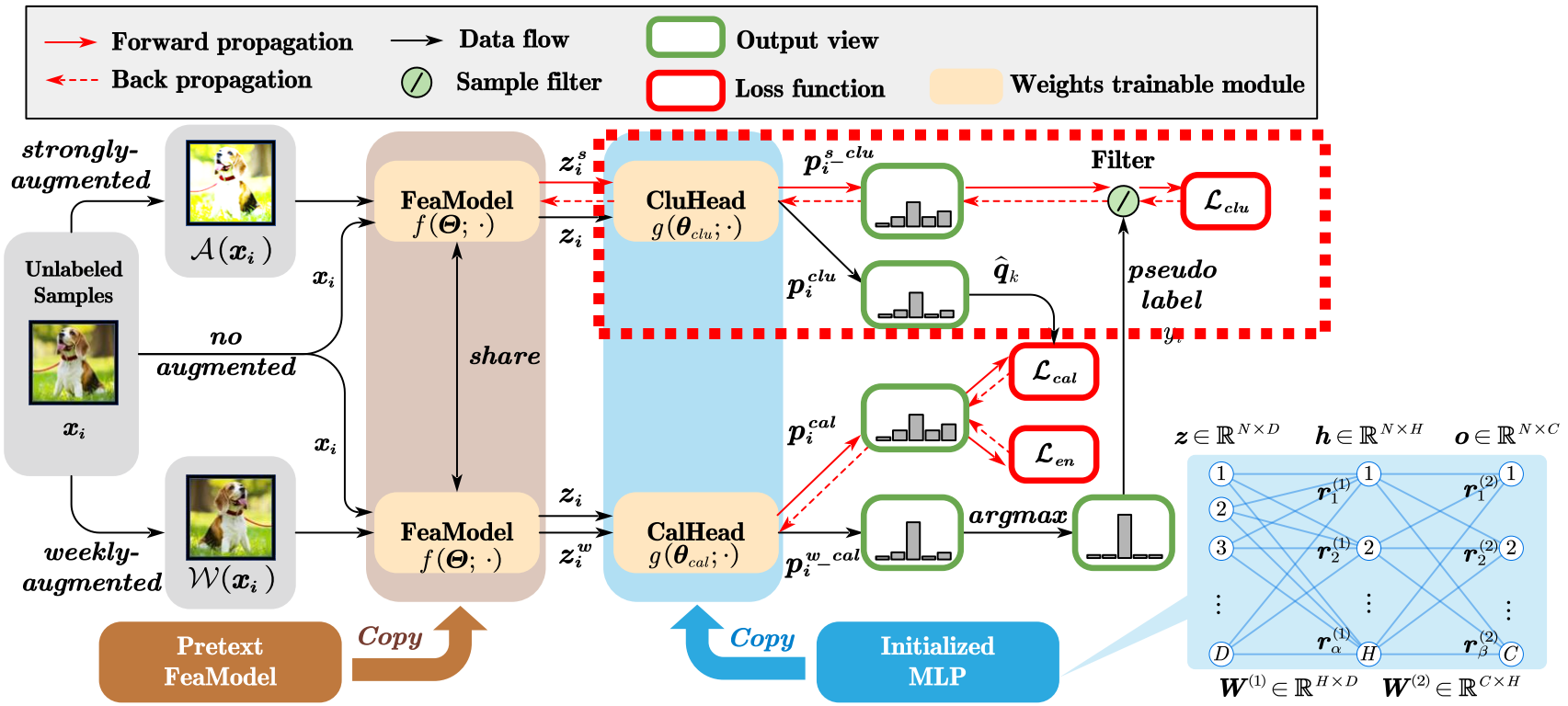
The Proposed Method



1. Calibration Head

- Adjust confidence scores from the Clustering Head to align predicted confidence with actual accuracy.

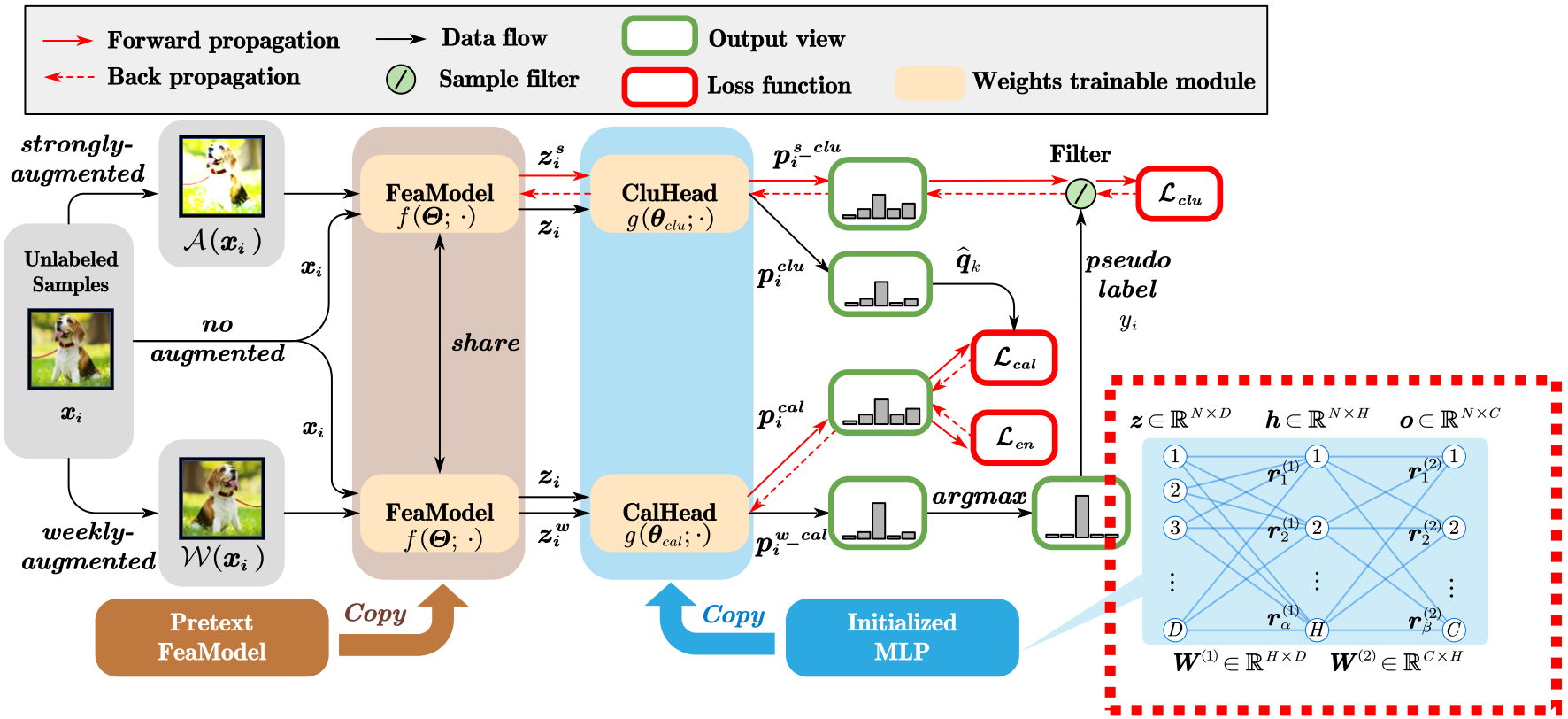
The Proposed Method



2. Clustering Head

- Dynamically select high-confidence samples based on calibrated confidence for pseudo-label training.

The Proposed Method



3. Feature Prototype-based Initialization Strategy

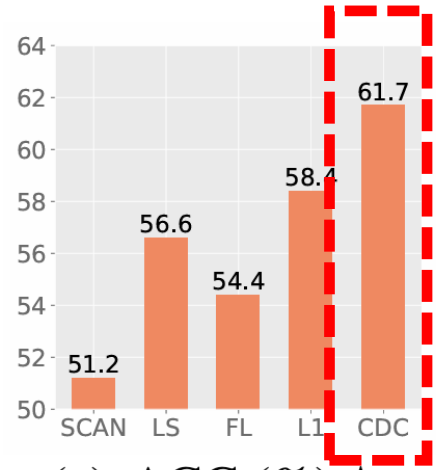
- Help accelerate training convergence and enhance robustness.

Results on Six Benchmark Datasets

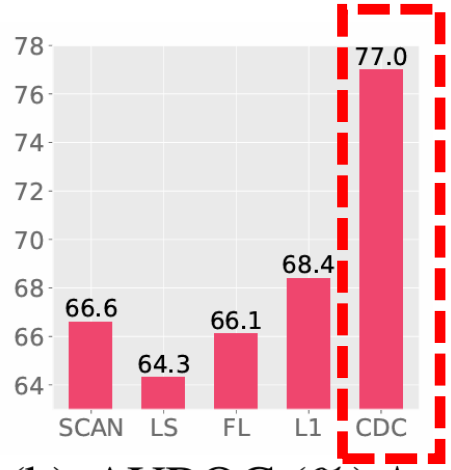
Table 1: The clustering performance ACC, ARI (%) and calibration error ECE (%) of various deep clustering methods trained on six image benchmarks. The best and second-best results are highlighted in **bold** and underlined, respectively. ↑ (↓) means the higher (resp. lower), the better.

Method	CIFAR-10			CIFAR-20			STL-10			ImageNet-10			ImageNet-Dogs			Tiny-ImageNet		
	ACC↑	ARI↑	ECE↓	ACC↑	ARI↑	ECE↓	ACC↑	ARI↑	ECE↓	ACC↑	ARI↑	ECE↓	ACC↑	ARI↑	ECE↓	ACC↑	ARI↑	ECE↓
K-means	22.9	4.9	N/A	13.0	2.8	N/A	19.2	6.1	N/A	24.1	5.7	N/A	10.5	2.0	N/A	2.5	0.5	N/A
MoCo-v2	82.9	64.9	N/A	50.7	26.2	N/A	68.8	45.5	N/A	56.7	30.9	N/A	62.8	48.1	N/A	25.2	11.0	N/A
Simsiam	70.7	53.1	N/A	33.0	16.2	N/A	49.4	34.9	N/A	78.4	68.8	N/A	44.2	27.3	N/A	19.0	8.4	N/A
BYOL	57.0	47.6	N/A	34.7	21.2	N/A	56.3	38.6	N/A	71.5	54.1	N/A	58.2	44.2	N/A	11.2	4.6	N/A
DMICC	82.8	69.0	N/A	46.8	29.1	N/A	80.0	62.5	N/A	96.2	91.6	N/A	58.7	43.8	N/A	-	-	-
ProPos	94.3	88.4	N/A	61.4	45.1	N/A	86.7	73.7	N/A	96.2	91.8	N/A	77.5	67.5	N/A	29.4	17.9	N/A
CoNR	93.2	86.1	N/A	60.4	44.3	N/A	92.6	84.6	N/A	96.4	92.2	N/A	**79.4**	66.7	N/A	30.8	18.4	N/A
DivClust	81.9	68.1	-	43.7	28.3	-	-	-	-	93.6	87.8	-	52.9	37.6	-	-	-	-
CC	85.2	72.8	6.2	42.4	28.4	29.7	80.0	67.7	11.9	90.6	85.3	8.1	69.6	56.0	19.3	12.1	5.7	**3.2**
TCC	90.6	73.3	-	49.1	31.2	-	81.4	68.9	-	89.7	82.5	-	59.5	41.7	-	-	-	-
TCL	88.7	78.0	-	53.1	35.7	-	86.8	75.7	-	89.5	83.7	-	64.4	51.6	-	-	-	-
SeCu-Size	90.0	81.5	8.1	52.9	38.4	13.1	80.2	63.1	9.9	-	-	-	-	-	-	-	-	-
SeCu	92.6	85.4	4.9	52.7	39.7	41.8	83.6	69.3	6.5	-	-	-	-	-	-	-	-	-
SCAN-2	84.1	74.1	10.9	50.0	34.7	37.1	87.0	75.6	7.4	95.1	89.4	2.7	63.3	49.6	26.4	27.6	15.3	27.4
SCAN-3	90.3	80.8	6.7	51.2	35.6	39.0	91.4	82.5	6.6	97.0	93.6	1.5	72.2	58.7	19.5	25.8	13.4	48.8
SPICE-2	84.4	70.9	15.4	47.6	30.3	52.3	89.6	79.2	10.1	92.1	83.6	7.8	64.6	47.7	35.3	30.5	16.3	48.5
SPICE-3	91.5	83.4	7.8	58.4	42.2	40.6	93.0	85.5	6.3	95.9	91.2	4.1	67.5	52.6	32.5	29.1	14.7	N/A
CDC-Clu (Ours)	94.9	89.4	1.4	**61.9**	**46.7**	28.0	**93.1**	**85.8**	4.8	97.2	94.0	1.8	79.3	**70.3**	17.1	**34.0**	**20.0**	37.8
CDC-Cal (Ours)	**94.9**	**89.5**	**1.1**	61.7	46.6	**4.9**	93.0	85.6	**0.9**	**97.3**	**94.1**	**0.8**	79.2	70.0	**7.7**	33.9	19.9	11.0
Supervised	89.7	78.9	4.0	71.7	50.2	11.0	80.4	62.2	10.0	99.2	98.3	0.9	93.1	85.7	0.9	47.7	24.3	5.1
+MoCo-v2	94.1	87.5	2.4	83.2	68.4	6.7	90.5	80.7	3.5	99.9	99.8	0.4	99.5	99.0	0.9	53.8	30.9	8.4

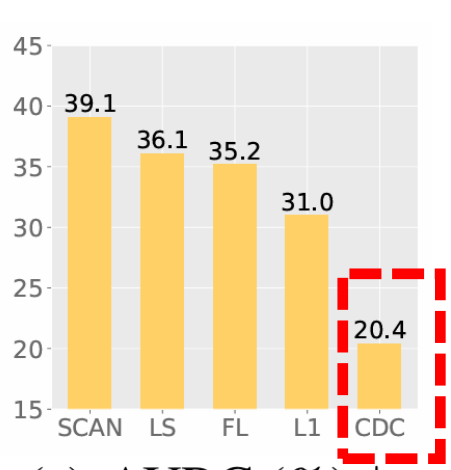
Failure Rejection Ability on CIFAR-20



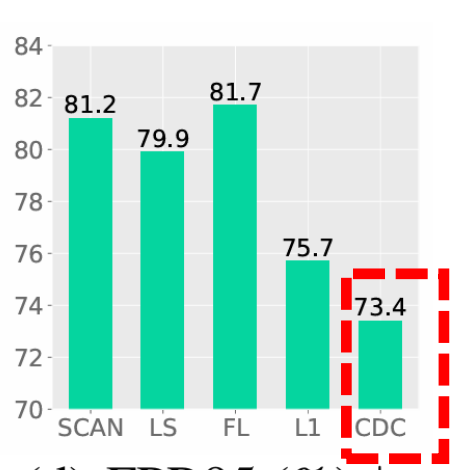
(a) ACC (%) ↑



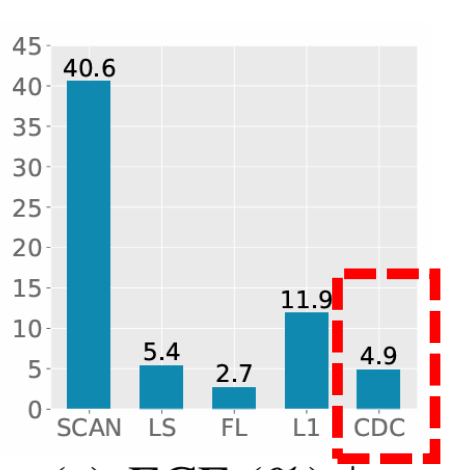
(b) AUROC (%) ↑



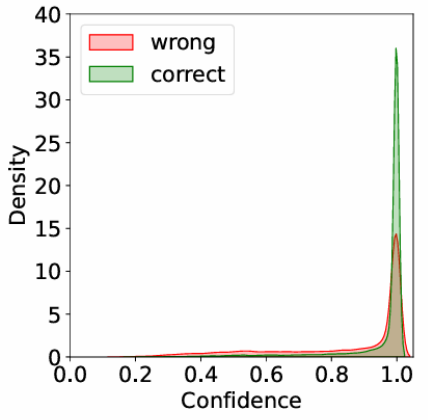
(c) AURC (%) ↓



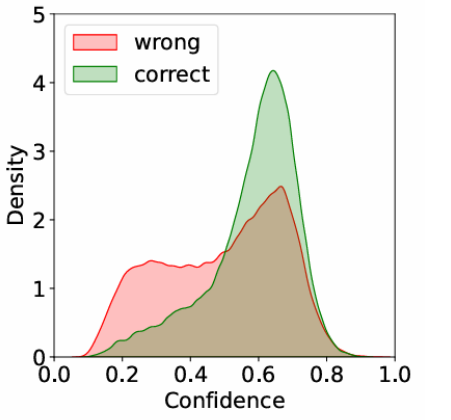
(d) FPR95 (%) ↓



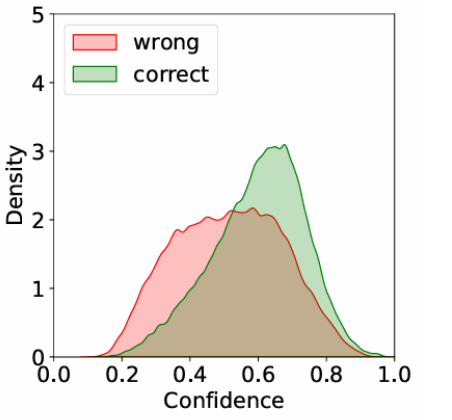
(e) ECE (%) ↓



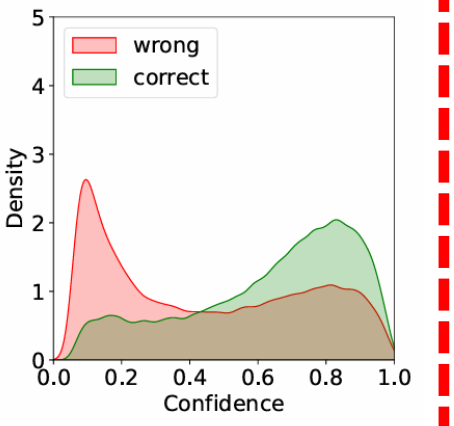
(f) SCAN



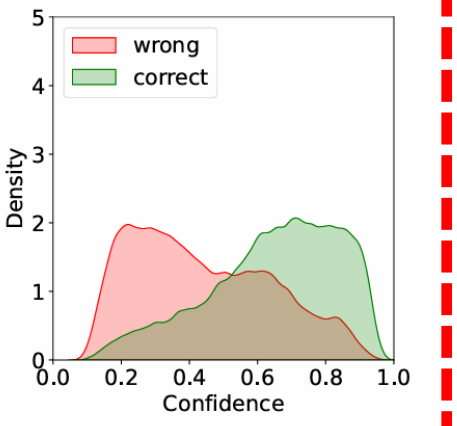
(g) LS



(h) FL

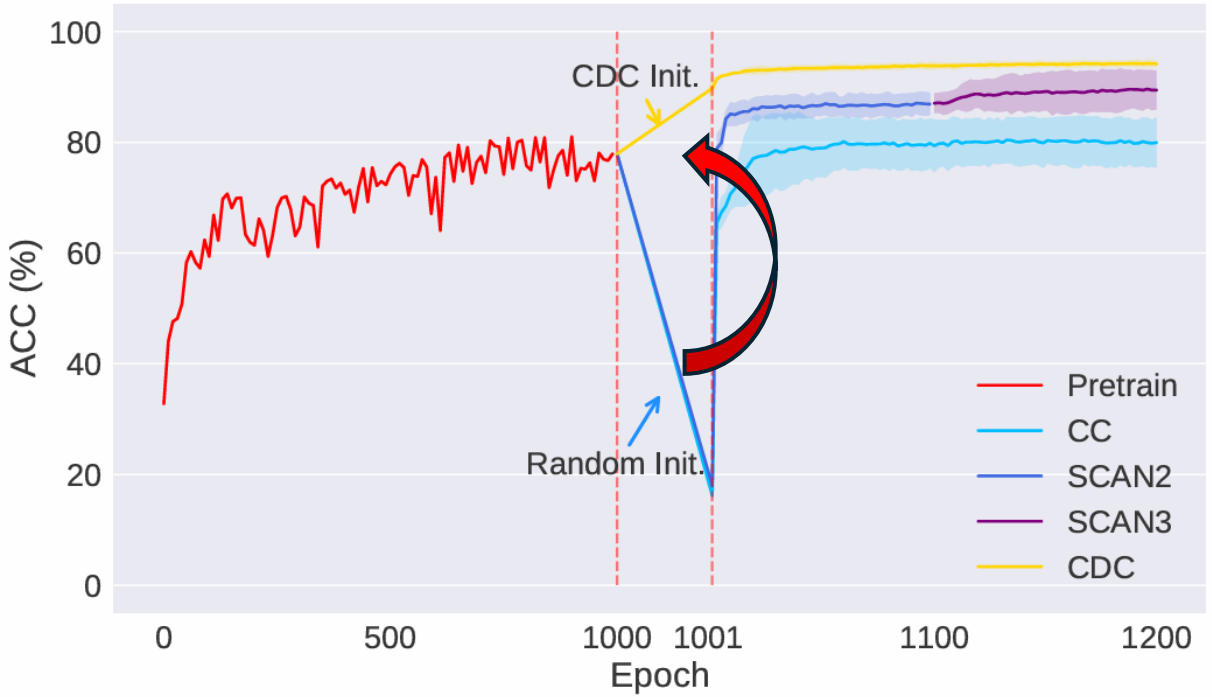


(i) L1

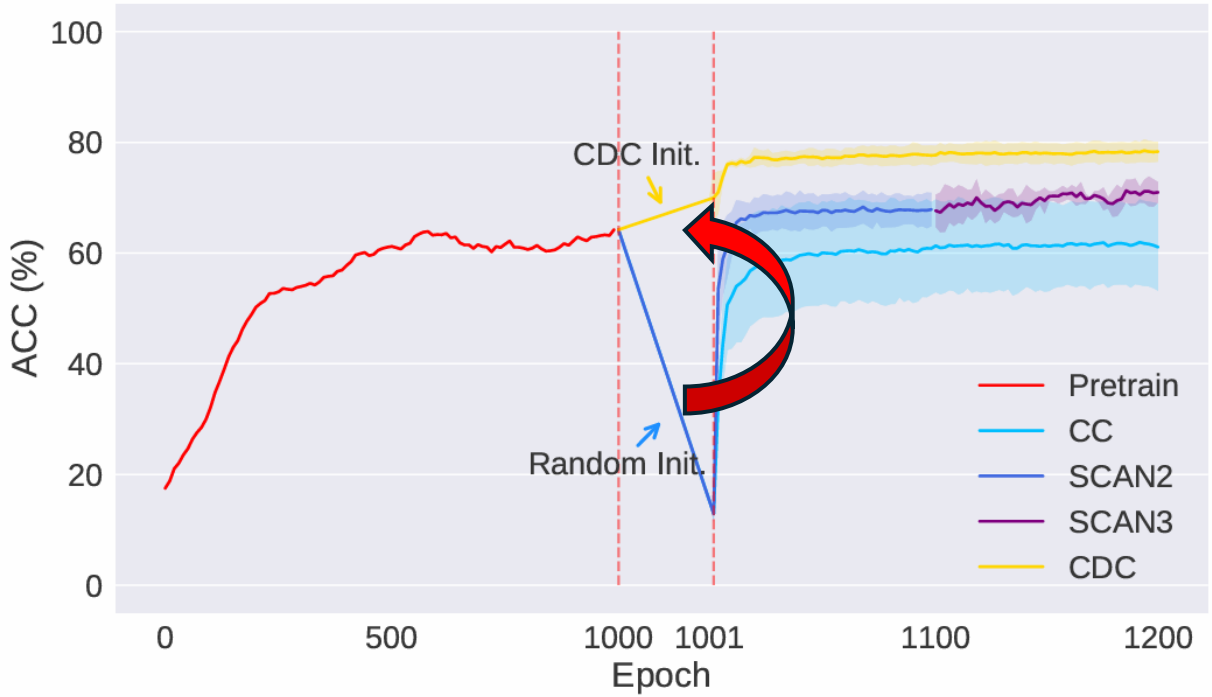


(j) CDC-Cal(Ours)

Training Process



(a) CIFAR-10



(b) ImageNet-Dogs

- Fewer training stages
- Better initialization strategy
- More stable performance improvement

THANKS

GitHub:
https://github.com/ChengJianH/CDC
Email:
chengjh@seu.edu.cn