CoLA: Co-Calibrated Logit Adjustment for Long-Tailed Semi-Supervised Learning

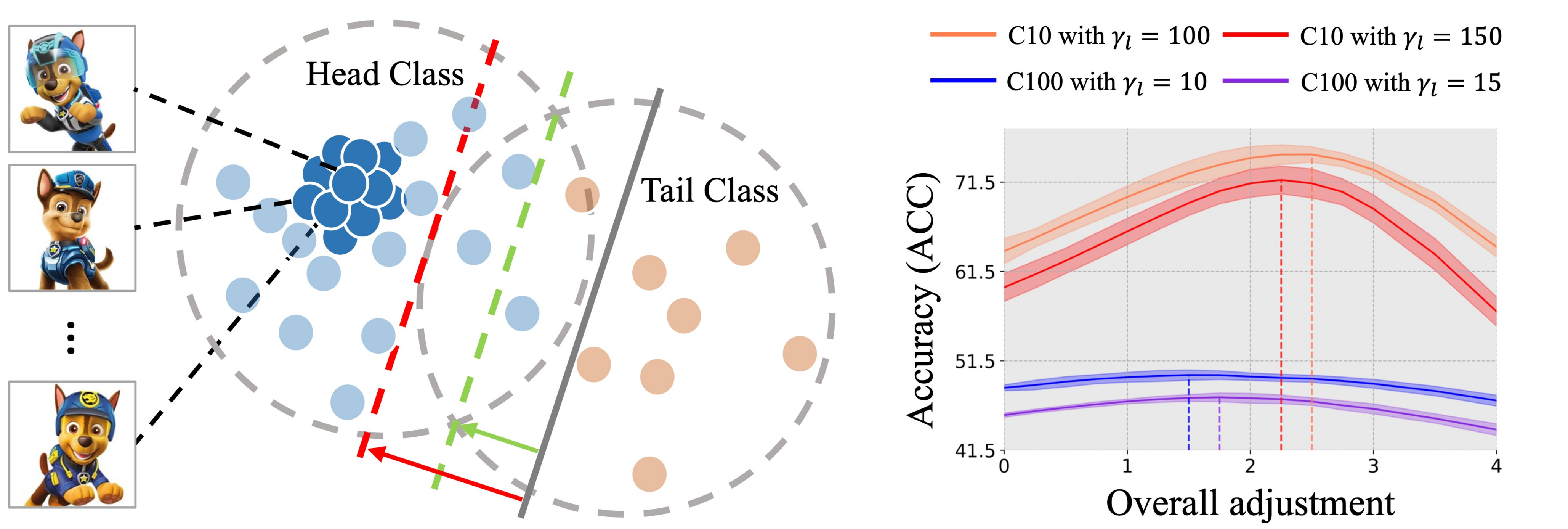
Qian Shao, Qiyuan Chen, Jiahe Chen, Zepeng Li, Qianqian Tang, Hongxia Xu†, Jian Wu†

Motivations

Two critical limitations exist in Logit Adjustment (LA) design:

(1) Reliance on simple frequency counting may overestimate the prevalence of head classes due to sample redundancy, leading to harmful over-suppression (red line in Figure (a));

(2) Ignore the interplay between class-wise adjustment and overall adjustment (Figure (b)).



(a) Long-tailed binary classification (b) Relations between adjustment and ACC

De-Duplicated Distribution Estimation (DDDE)

DDDE resolves over-suppression by quantifying the Effective Number (EN) of samples using the effective rank (erank) of feature representations:

(1) Gather the representations of high-confidence samples for each class to form a feature matrix $\boldsymbol{Z}_y$.

(2) Compute the singular values of $\boldsymbol{Z}_y$ to determine the energy distribution.

(3) Use the exponential of the Shannon entropy of this energy distribution as the effective rank:

$$\text{erank}(\boldsymbol{Z}_y) = \exp\left(-\sum_{i=1}^{m_y} p(i)\log p(i)\right).$$

(4) Normalize the effective counts across all classes to yield a robust, de-duplicated estimate of the unlabeled class distribution $\hat{P}_{Y_u}(y)$:

$$\hat{P}_{Y_u}(y) = \text{erank}(\boldsymbol{Z}_y) / \sum_{k\in\mathcal{Y}} \text{erank}(\boldsymbol{Z}_y).$$

Logit Meta-Calibration (LMC)

LMC formulates the overall adjustment strength τ as a learnable parameter rather than a static guess:

(1) Construct a proxy validation set via resampling the labeled data to match $\hat{P}_{Y_u}(y)$:

$$\mathbb{P}\left(\left(\boldsymbol{x}_i^l, y_i\right) \text{ is selected}\right) = \frac{\hat{P}_{Y_u}(y_i)}{N_{y_i}} \Big/ \max_{y\in\mathcal{Y}}\left(\frac{\hat{P}_{Y_u}(y)}{N_y}\right).$$

(2) The optimal strength τ^* is then found by minimizing the Cross-Entropy (CE) loss on this proxy set:

$$\tau^* = \underset{\tau}{\text{argmin}} \frac{1}{V}\sum_{i=1}^{V} \mathcal{L}_{\text{CE}}\left(y_i^v, \sigma\left(z\left(\alpha(\boldsymbol{x}_i^v)\right) - \tau\cdot\boldsymbol{p}\right)\right).$$

Main Experimental Results

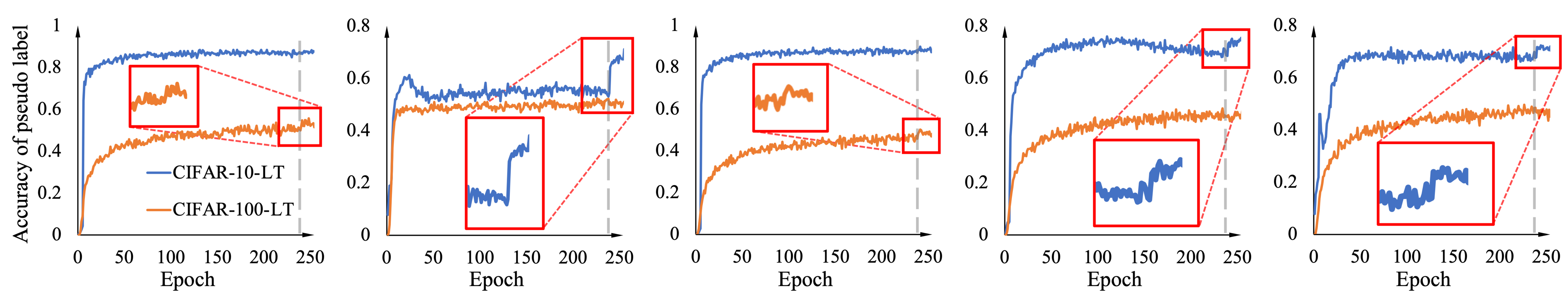
Comparison with other LTSSL methods

Method	CIFAR-10-LT					CIFAR-100-LT				
	CON	UNI	REV	MID	HT	CON	UNI	REV	MID	HT
Supervised	53.27 ± 8.94	61.18 ± 2.74	61.73 ± 2.55	53.48 ± 8.66	53.64 ± 8.84	47.12 ± 1.65	47.09 ± 1.66	46.95 ± 1.63	46.83 ± 1.58	47.01 ± 1.59
FixMatch	71.93 ± 4.05	71.24 ± 2.58	68.30 ± 3.14	70.10 ± 4.35	66.50 ± 6.84	55.89 ± 1.94	47.40 ± 1.66	55.83 ± 2.03	55.65 ± 2.01	56.59 ± 1.94
Logit Adjustment-Based										
ACR	80.89 ± 2.92	82.16 ± 1.02	83.49 ± 1.15	79.74 ± 4.75	78.42 ± 4.45	$\underline{58.31 \pm 1.77}$	48.98 ± 1.45	$\underline{59.21 \pm 1.17}$	$\underline{57.66 \pm 1.64}$	$\underline{58.76 \pm 1.31}$
CPE	80.28 ± 3.18	82.59 ± 1.13	84.01 ± 0.91	80.80 ± 3.05	79.63 ± 3.86	53.12 ± 2.19	46.89 ± 1.47	58.61 ± 1.12	54.42 ± 2.35	57.15 ± 1.56
Meta-Expert	81.33 ± 2.53	$\underline{83.12 \pm 1.09}$	$\underline{85.03 \pm 0.54}$	$\underline{80.85 \pm 3.20}$	$\underline{79.87 \pm 3.44}$	54.60 ± 2.21	47.85 ± 1.55	58.59 ± 1.74	55.09 ± 2.08	57.64 ± 1.75
Sim-Pro	$\underline{81.57 \pm 2.29}$	80.86 ± 0.83	74.71 ± 1.57	77.69 ± 4.22	76.78 ± 2.95	58.30 ± 1.60	48.06 ± 1.78	57.36 ± 1.89	57.55 ± 1.69	58.36 ± 1.57
Loss Reweighting-Based										
DeBiasPL	72.51 ± 4.40	71.64 ± 3.13	68.68 ± 3.50	70.67 ± 4.63	67.30 ± 7.04	56.37 ± 1.74	47.92 ± 1.67	56.60 ± 2.09	56.16 ± 1.73	56.98 ± 1.69
SAW	76.57 ± 3.73	81.34 ± 0.69	74.83 ± 1.56	74.39 ± 4.97	76.70 ± 2.40	57.09 ± 1.72	48.20 ± 1.62	57.50 ± 1.72	56.93 ± 1.93	57.87 ± 1.83
Resampling-Based										
ABC	80.15 ± 2.57	81.08 ± 0.88	81.83 ± 1.01	78.70 ± 4.23	79.40 ± 3.21	57.06 ± 1.56	49.07 ± 1.52	58.48 ± 1.14	57.31 ± 1.70	58.75 ± 1.57
CReST	75.16 ± 3.84	82.00 ± 1.56	83.60 ± 2.69	79.78 ± 4.02	77.71 ± 5.93	56.18 ± 1.71	47.50 ± 1.63	57.95 ± 1.51	57.47 ± 1.37	56.43 ± 1.49
Data Mixing-Based										
BEM	75.27 ± 3.76	80.07 ± 1.53	77.77 ± 1.96	80.29 ± 3.45	73.37 ± 4.36	57.58 ± 1.67	49.22 ± 1.27	58.30 ± 1.67	57.45 ± 1.79	58.05 ± 1.51
CoSSL	80.52 ± 2.59	79.42 ± 1.64	74.38 ± 2.34	79.32 ± 3.44	72.55 ± 4.34	56.90 ± 1.33	$\underline{49.57 \pm 1.36}$	57.23 ± 1.80	57.30 ± 1.55	58.06 ± 1.68
Distribution Alignment-Based										
DARP	74.11 ± 3.48	71.59 ± 2.72	68.84 ± 3.06	70.80 ± 4.37	67.08 ± 6.80	56.73 ± 1.66	47.82 ± 1.60	56.39 ± 1.91	55.82 ± 1.99	56.93 ± 1.84
RDA	73.14 ± 3.42	77.51 ± 1.03	71.80 ± 1.38	64.58 ± 5.04	70.84 ± 4.03	56.32 ± 1.99	47.64 ± 1.81	56.16 ± 1.86	55.53 ± 2.31	57.06 ± 1.97
Others										
DASO	70.16 ± 4.74	74.75 ± 1.94	76.47 ± 1.72	68.32 ± 5.58	68.61 ± 5.39	57.32 ± 1.82	47.89 ± 1.59	57.93 ± 1.97	57.27 ± 1.84	58.16 ± 1.92
CoLA (Ours)	$\mathbf{81.87 \pm 2.70}$	$\mathbf{83.66 \pm 1.29}$	$\mathbf{85.61 \pm 1.56}$	$\mathbf{81.86 \pm 3.41}$	$\mathbf{80.65 \pm 3.32}$	$\mathbf{59.04 \pm 1.59}$	$\mathbf{50.26 \pm 1.23}$	$\mathbf{60.39 \pm 1.22}$	$\mathbf{58.71 \pm 1.58}$	$\mathbf{59.89 \pm 1.45}$

Ablation of DDDE and LMC

Dataset	CIFAR-10-LT					CIFAR-100-LT				
$(k_{\max}, k_{\min})$	(1,10)	(~,~)	(10,1)	(5,10)	(10,5)	(1,100)	(100,1)	(~,~)	(50,100)	(100,50)
(γ_l, γ_u)	(100,100)	(100,1)	(100,100)	(100,100)	(100,100)	(10,10)	(10,1)	(10,10)	(10,10)	(10,10)
w/o D-1	83.12	82.03	80.57	82.34	81.30	56.23	48.07	59.33	55.69	57.83
w/o D-2	83.56	82.31	83.42	83.06	82.11	55.41	47.68	59.02	55.84	57.59
w/o D-4	82.64	81.39	75.44	81.27	79.25	53.32	46.05	57.36	54.22	56.14
w/o D-L	84.66	83.38	84.77	84.50	83.86	60.16	50.63	60.79	59.64	60.18
w/ D-L (Ours)	**85.04**	**84.83**	**86.84**	**85.16**	**84.42**	**60.42**	**51.28**	**61.40**	**60.10**	**61.17**

Visualization of pseudo-label accuracy variation



(a) Consistent distribution (b) Uniform distribution (c) Reversed distribution (d) Middle distribution (e) Head-tail distribution