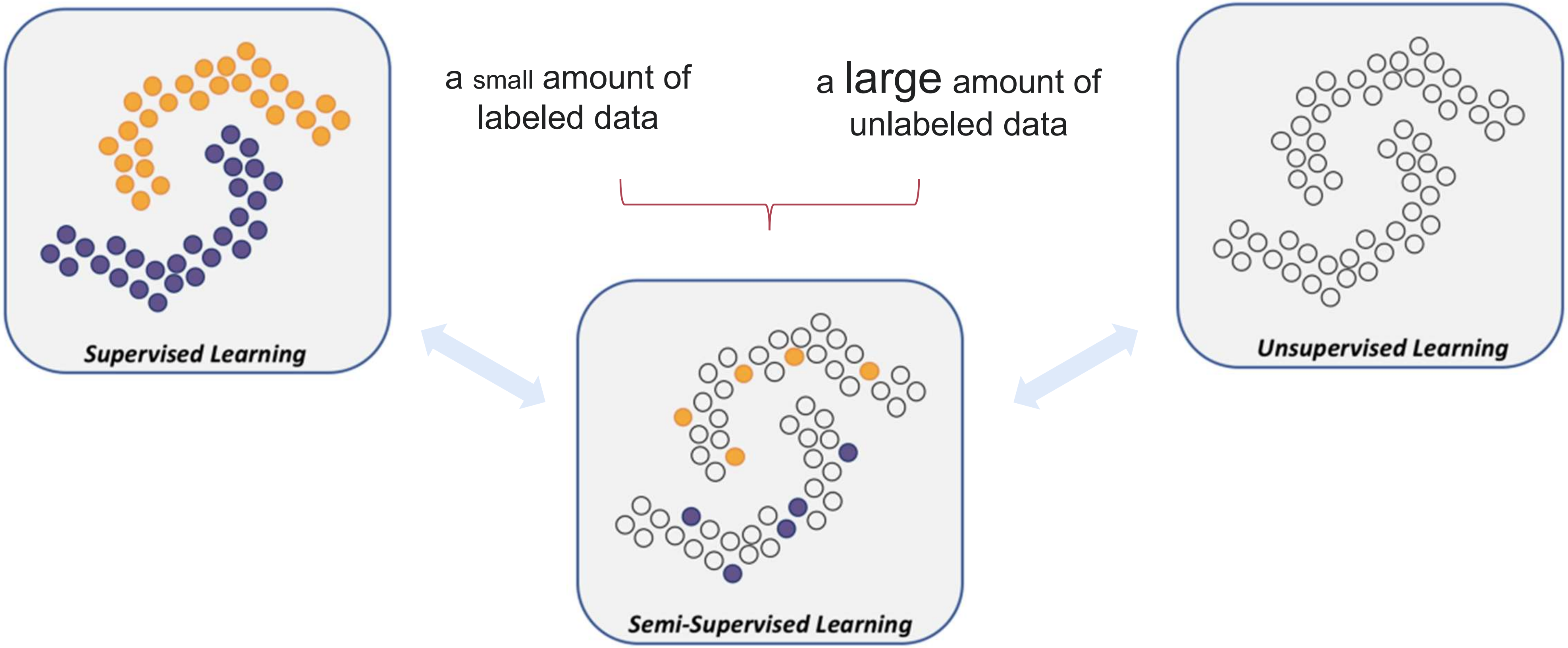
模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

IMBALANCED SEMI-SUPERVISED LEARNING

Background

Semi-supervised Learning （SSL）



Background

Imbalanced Semi-supervised Learning （SSL)

assumption： The distribution of labeled and/or unlabeled data are class-balanced.

↓

assumption： The labeled and unlabeled data are in the same class-imbalance distribution.

The test set is class-balanced

模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

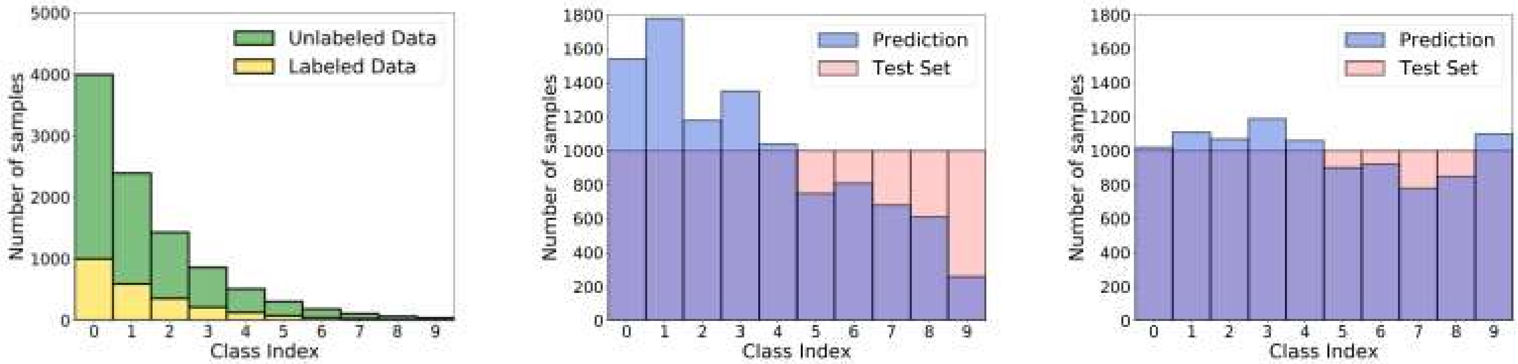
ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

ABC: Auxiliary Balanced Classifier for Class-imbalanced Semi-supervised Learning

Hyuck Lee Seungjae Shin Heeyoung Kim
Department of Industrial and Systems Engineering, KAIST
{dlgur0921, tmdwo0910, heeyoungkim}@kaist.ac.kr

NIPS 2021

Related work



(a) Class-imbalanced training set (b) ReMixMatch (c) Proposed algorithm

Figure 1: Predictions on a class-balanced test set using ReMixMatch (b) and the proposed algorithm (c) trained on a class-imbalanced training set (a).

Experiments on CIFAR-10-LT. Training set is class-imbalanced with imbalance ratio γ = 100.

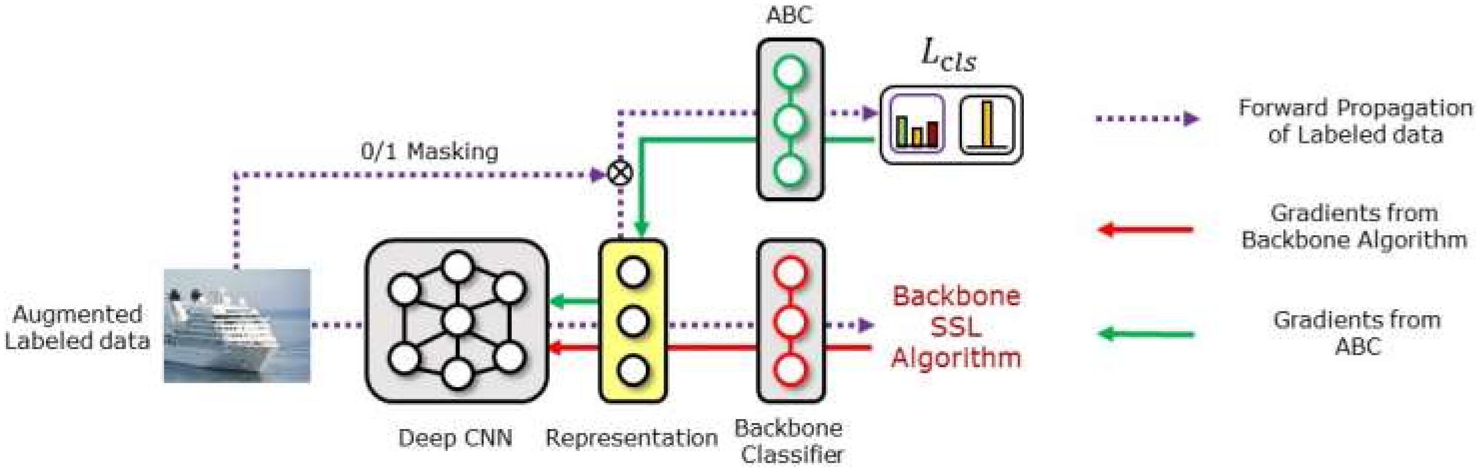


Figure 2: Overall procedure for balanced training of the ABC with a $0/1$ mask

$$L_{cls} = \frac{1}{B}\sum_{b=1}^{B} M\left(x_b\right) \mathbf{H}\left(p_s\left(y|\alpha\left(x_b\right)\right), p_b\right),$$

$$M\left(x_b\right) = \mathcal{B}\left(\frac{N_L}{N_{y_b}}\right),$$

Method

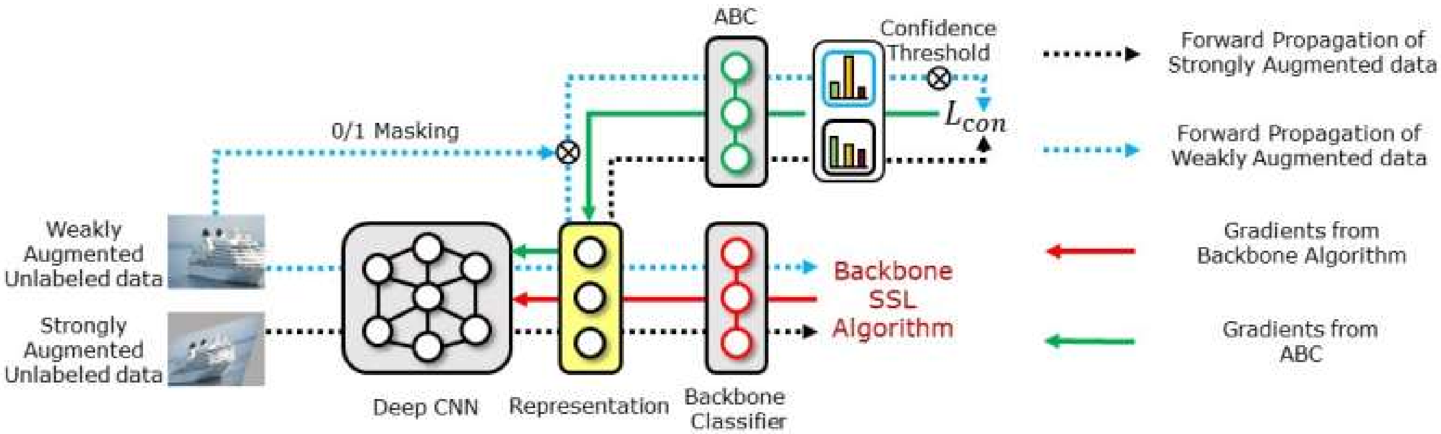


Figure 5: Overall procedure of consistency regularization for the ABC

$$L_{con} = \frac{1}{B}\sum_{b=1}^{B}\sum_{k=1}^{2} M\left(u_b\right)\mathbf{I}\left(\max\left(q_b\right) \geq \tau\right)\mathbf{H}\left(p_s\left(y|\mathcal{A}_k\left(u_b\right)\right), q_b\right),$$

$$M\left(u_b\right) = \mathcal{B}\left(\frac{N_L}{N_{\hat{q}_b}}\right),$$

Method

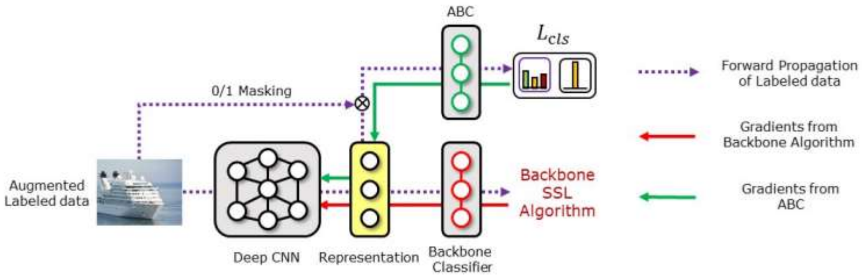


Figure 2: Overall procedure for balanced training of the ABC with a 0/1 mask

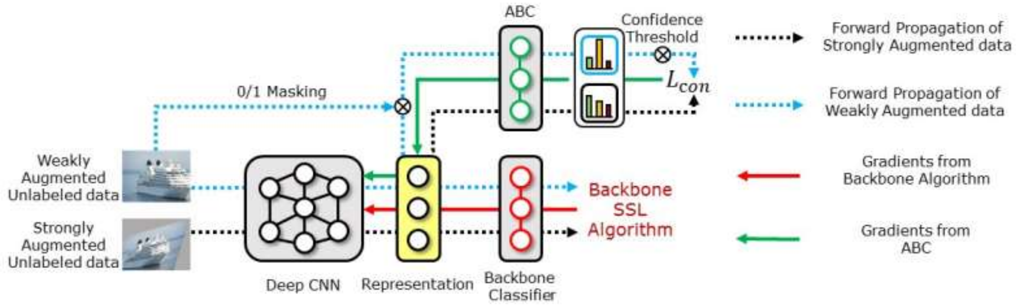


Figure 5: Overall procedure of consistency regularization for the ABC

$$L_{cls} = \frac{1}{B}\sum_{b=1}^{B} M(x_b)\,\mathbf{H}(p_s(y|\alpha(x_b)), p_b),$$

$$M(x_b) = \mathcal{B}\left(\frac{N_L}{N_{y_b}}\right),$$

$$L_{con} = \frac{1}{B}\sum_{b=1}^{B}\sum_{k=1}^{2} M(u_b)\,\mathbf{I}(\max(q_b) \geq \tau)\,\mathbf{H}(p_s(y|\mathcal{A}_k(u_b)), q_b),$$

$$M(u_b) = \mathcal{B}\left(\frac{N_L}{N_{\hat{q}_b}}\right),$$

$$L_{total} = L_{cls} + L_{con} + L_{back}.$$

Method

ABC

Algorithm 1 Pseudo code of the proposed algorithm

Input: $\mathcal{MB}_{\mathcal{X}} = \{(x_b, y_b) : b \in (1, ..., B)\} \subset \mathcal{X}$, $\mathcal{MB}_{\mathcal{U}} = \{(u_b) : b \in (1, \ldots, B)\} \subset \mathcal{U}$
Output: Classification model $f : \mathbb{R}^d \rightarrow \{1, ...L\}$
Parameters : $\boldsymbol{\theta}$ (Parameters of Wide ResNet-28-2 and ABC)

while Training **do**
  for $b = 1$ to B **do**
    $\alpha(x_b) =$ Augment(x_b)
    $\alpha(u_b) =$ WeakAugment(u_b)
    $\mathcal{A}_k(u_b) =$StrongAugment$_k(x_b)$, $k = 1, 2$
    Predicted class distribution for $\alpha(x_b) = p_s(y|\alpha(x_b))$
    Generate $0/1$ mask $M(x_b)$.
    Soft pseudo label $q_b = p_s(y|\alpha(u_b))$
    if $\max(q_b) \geq 0.95$ **then**
      Predicted class distribution for $\mathcal{A}_k(u_b) = p_s(y|\mathcal{A}_k(u_b)), k = 1, 2$
      Generate $0/1$ mask $M(u_b)$.
    end if
    Loss from the backbone L_{back} += backbone$(\alpha(x_b), \alpha(u_b), \mathcal{A}_k(u_b))$
  end for
  Calculate the classification loss L_{cls} and consistency regularization loss L_{con}.
  Total Loss $L_{total} = L_{cls} + L_{con} + L_{back}$
  $\Delta\boldsymbol{\theta} \propto \nabla_{\boldsymbol{\theta}} L_{total}, \quad \boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \Delta\boldsymbol{\theta}$
end while

Experiment

Table 1: Overall accuracy/minority-class-accuracy under the main setting

	CIFAR-10-LT	SVHN-LT	CIFAR-100-LT
Algorithm	$\gamma = 100$, $\beta = 20\%$	$\gamma = 100$, $\beta = 20\%$	$\gamma = 20$, $\beta = 40\%$
Vanilla	$55.3_{\pm 1.30}$ / $33.9_{\pm 1.88}$	$77.0_{\pm 0.67}$ / $63.3_{\pm 1.25}$	$40.1_{\pm 1.15}$ / $25.2_{\pm 0.95}$
VAT [24]	$55.3_{\pm 0.88}$ / $28.2_{\pm 1.55}$	$81.3_{\pm 0.47}$ / $68.2_{\pm 0.88}$	$40.4_{\pm 0.34}$ / $24.8_{\pm 0.38}$
BALMS [27]	$70.7_{\pm 0.59}$ / $69.8_{\pm 1.03}$	$87.6_{\pm 0.53}$ / $85.0_{\pm 0.67}$	$50.2_{\pm 0.54}$ / $42.9_{\pm 1.03}$
FixMatch [29]	$72.3_{\pm 0.33}$ / $53.8_{\pm 0.63}$	$88.0_{\pm 0.30}$ / $79.4_{\pm 0.54}$	$51.0_{\pm 0.20}$ / $32.8_{\pm 0.41}$
w/ CReST+PDA [34]	$76.6_{\pm 0.46}$ / $61.4_{\pm 0.85}$	$89.1_{\pm 0.69}$ / $81.7_{\pm 1.18}$	$51.6_{\pm 0.29}$ / $36.4_{\pm 0.46}$
w/ DARP [18]	$73.7_{\pm 0.98}$ / $57.0_{\pm 2.12}$	$88.6_{\pm 0.19}$ / $80.5_{\pm 0.54}$	$51.4_{\pm 0.37}$ / $33.9_{\pm 0.77}$
w/ DARP+cRT [18]	$78.1_{\pm 0.89}$ / $66.6_{\pm 1.55}$	$89.9_{\pm 0.44}$ / $83.5_{\pm 0.61}$	$54.7_{\pm 0.46}$ / $41.2_{\pm 0.42}$
w/ ABC	$\mathbf{81.1}_{\pm 0.82}$ / $\mathbf{72.0}_{\pm 1.77}$	$\mathbf{92.0}_{\pm 0.38}$ / $\mathbf{87.9}_{\pm 0.73}$	$\mathbf{56.3}_{\pm 0.19}$ / $\mathbf{43.4}_{\pm 0.42}$
ReMixMatch [3]	$73.7_{\pm 0.39}$ / $55.9_{\pm 0.87}$	$89.8_{\pm 0.42}$ / $82.8_{\pm 0.68}$	$54.0_{\pm 0.29}$ / $37.1_{\pm 0.37}$
w/ CReST+PDA [34]	$75.7_{\pm 0.34}$ / $59.6_{\pm 0.76}$	$90.9_{\pm 0.20}$ / $85.2_{\pm 0.39}$	$54.6_{\pm 0.48}$ / $38.1_{\pm 0.69}$
w/ DARP [18]	$74.4_{\pm 0.41}$ / $56.9_{\pm 0.67}$	$90.2_{\pm 0.22}$ / $83.5_{\pm 0.40}$	$54.5_{\pm 0.33}$ / $37.7_{\pm 0.58}$
w/ DARP+cRT [18]	$78.5_{\pm 0.61}$ / $66.4_{\pm 1.68}$	$92.1_{\pm 0.48}$ / $87.6_{\pm 0.75}$	$55.1_{\pm 0.45}$ / $43.6_{\pm 0.58}$
w/ ABC	$\mathbf{82.4}_{\pm 0.45}$ / $\mathbf{75.7}_{\pm 1.18}$	$\mathbf{93.9}_{\pm 0.16}$ / $\mathbf{92.5}_{\pm 0.4}$	$\mathbf{57.6}_{\pm 0.26}$ / $\mathbf{46.7}_{\pm 0.50}$

Experiment

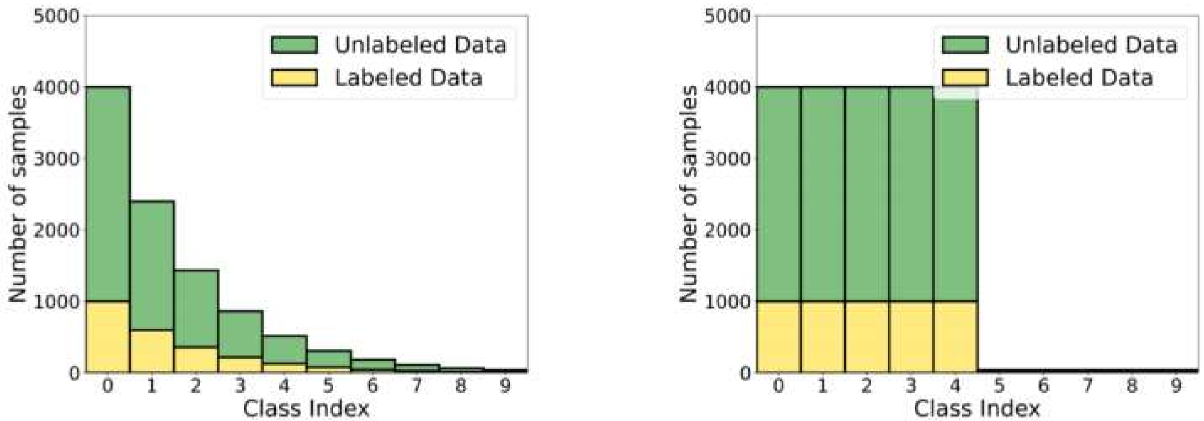
Table 2: Overall accuracy/minority-class accuracy for CIFAR-10 under various settings

CIFAR-10-LT				
Algorithm	$\gamma = 100, \beta = 10\%$	$\gamma = 100, \beta = 30\%$	$\gamma = 50, \beta = 20\%$	$\gamma = 150, \beta = 20\%$
FixMatch [29]	$70.0_{\pm 0.59}$ / $48.9_{\pm 1.04}$	$74.9_{\pm 0.63}$ / $58.2_{\pm 1.28}$	$81.2_{\pm 0.07}$ / $70.7_{\pm 0.36}$	$68.5_{\pm 0.60}$ / $45.8_{\pm 1.15}$
w/ CReST+PDA [34]	$73.9_{\pm 0.40}$ / $58.9_{\pm 1.14}$	$77.6_{\pm 0.73}$ / $64.0_{\pm 1.39}$	$83.3_{\pm 0.10}$ / $75.7_{\pm 0.39}$	$70.0_{\pm 0.82}$ / $49.4_{\pm 1.52}$
w/ DARP+cRT [18]	$74.6_{\pm 0.98}$ / $59.2_{\pm 2.12}$	$79.0_{\pm 0.25}$ / $67.7_{\pm 0.95}$	$83.6_{\pm 0.42}$ / $77.1_{\pm 1.19}$	$73.2_{\pm 0.85}$ / $57.1_{\pm 1.13}$
w/ ABC	$\mathbf{77.2}_{\pm 1.60}$ / $\mathbf{65.7}_{\pm 2.85}$	$\mathbf{81.5}_{\pm 0.29}$ / $\mathbf{72.9}_{\pm 0.96}$	$\mathbf{85.2}_{\pm 0.51}$ / $\mathbf{80.2}_{\pm 0.64}$	$\mathbf{77.1}_{\pm 0.46}$ / $\mathbf{64.4}_{\pm 0.92}$
ReMixMatch [3]	$71.5_{\pm 0.51}$ / $52.2_{\pm 1.08}$	$75.8_{\pm 0.10}$ / $59.4_{\pm 0.17}$	$81.5_{\pm 0.17}$ / $70.7_{\pm 0.32}$	$69.9_{\pm 0.23}$ / $48.4_{\pm 0.60}$
w/ CReST+PDA [34]	$73.8_{\pm 0.32}$ / $56.6_{\pm 0.43}$	$78.6_{\pm 0.73}$ / $64.8_{\pm 1.49}$	$83.9_{\pm 0.26}$ / $75.4_{\pm 0.52}$	$71.3_{\pm 0.77}$ / $50.8_{\pm 1.59}$
w/ DARP+cRT [18]	$75.9_{\pm 1.20}$ / $62.1_{\pm 3.10}$	$81.0_{\pm 0.16}$ / $70.7_{\pm 0.72}$	$84.5_{\pm 0.80}$ / $77.8_{\pm 1.67}$	$73.9_{\pm 0.59}$ / $57.4_{\pm 1.45}$
w/ ABC	$\mathbf{79.8}_{\pm 0.36}$ / $\mathbf{70.8}_{\pm 0.92}$	$\mathbf{84.3}_{\pm 1.03}$ / $\mathbf{80.6}_{\pm 0.97}$	$\mathbf{87.5}_{\pm 0.31}$ / $\mathbf{84.6}_{\pm 1.19}$	$\mathbf{80.6}_{\pm 0.66}$ / $\mathbf{72.1}_{\pm 1.51}$

Table 4: Overall accuracy/minority-class accuracy for the large-scale LSUN dataset

LSUN, $\gamma = 100$, $\beta = 20\%$					
Algorithm	w/ -	w/ cRT [17]	w/ DARP [18]	w/ DARP+cRT [18]	w/ ABC
FixMatch [29]	73.1 / 55.3	77.0 / 71.5	71.0 / 51.8	75.8 / 69.5	**78.9 / 75.5**
ReMixMatch [3]	69.4 / 49.1	75.4 / 69.5	65.6 / 44.1	72.1 / 67.5	**76.9 / 69.5**

Experiment



(a) Long-tailed imbalance (b) Step imbalance

Figure 6: Long-tailed imbalance and step imbalance

Table 3: Overall accuracy/minority-class accuracy on CIFAR-10 under a step imbalance setting

CIFAR-10-Step, $\gamma = 100$, $\beta = 20\%$				
Algorithm	w/ -	w/ CReST+PDA [34]	w/ DARP+cRT [18]	w/ ABC
FixMatch [29]	$54.0_{\pm 0.84}$ / $11.8_{\pm 1.71}$	$71.1_{\pm 0.78}$ / $48.2_{\pm 2.26}$	$69.8_{\pm 1.51}$ / $45.1_{\pm 2.70}$	**75.9**$_{\pm 0.49}$ / **57.0**$_{\pm 1.07}$
ReMixMatch [3]	$60.8_{\pm 0.10}$ / $25.1_{\pm 1.28}$	$64.6_{\pm 0.97}$ / $33.5_{\pm 2.05}$	$72.3_{\pm 1.77}$ / $50.6_{\pm 3.53}$	**76.4**$_{\pm 1.70}$ / **65.7**$_{\pm 1.30}$

Experiment

(a)ABC (without SSL backbone) (b)FixMatch+ABC (c) ReMixMatch+ABC

Figure 3: t-SNE of the proposed algorithm and the ABC (without SSL backbone)

Table 5: Ablation study for ReMixMatch+ABC on CIFAR-10-LT, $\gamma = 100$, $\beta = 20\%$

Ablation study	Overall	Minority
ReMixMatch+ABC (proposed algorithm)	**82.4**	**75.7**
Without gradually decreasing the parameter of $\mathcal{B}(\cdot)$ for consistency regularization	81.8	74.6
Without consistency regularization for the ABC	79.4	66.9
Without using the 0/1 mask for the consistency regularization loss L_{con}	79.0	69.2
Without using the 0/1 mask for the classification loss L_{cls}	74.4	57.8
Without using the confidence threshold τ for consistency regularization	74.3	75.4
Using hard pseudo labels for consistency regularization	70.2	75.1
Without training backbone (ABC without SSL backbone)	68.7	56.2
Training the ABC with a re-weighting technique	81.2	74.1
Decoupled training of the backbone and ABC	79.5	72.3

Background

Imbalanced Semi-supervised Learning （SSL)

assumption： The distribution of labeled and/or unlabeled data are class-balanced.

↓

assumption： The labeled and unlabeled data are in the different class-imbalance distribution.

The test set is class-balanced

模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

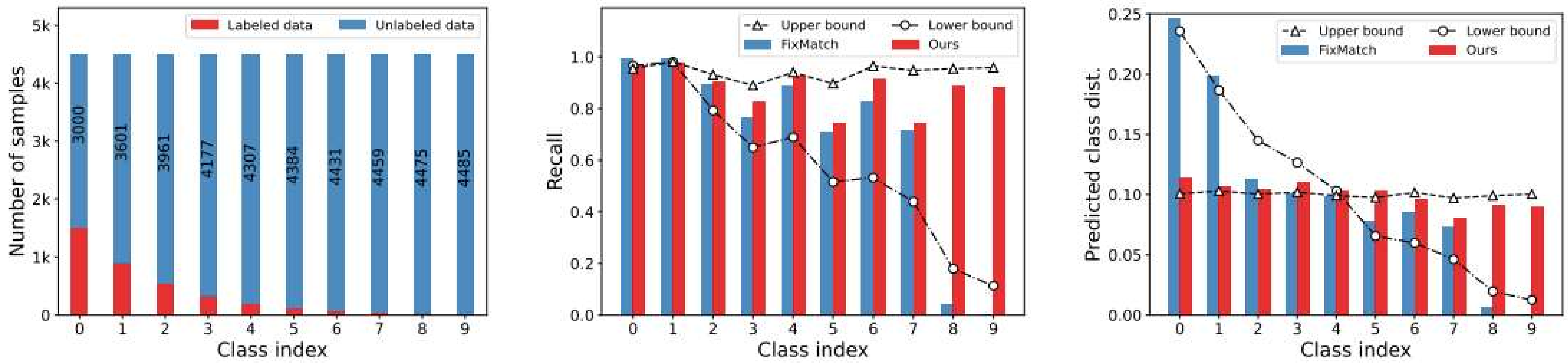
IMBALANCED SEMI-SUPERVISED LEARNING WITH BIAS ADAPTIVE CLASSIFIER

Renzhen Wang[1], Xixi Jia[2], Quanziang Wang[1], Yichen Wu[3], Deyu Meng[1,4,5*]
[1]Xi'an Jiaotong University, [2]Xidian University, [3]City University of Hong Kong
[4]Macau University of Science and Technology, [5]Peng Cheng Laboratory
{rzwang,dymeng}@mail.xjtu.edu.cn

ICLR 2023

Related work

Imbalanced Semi-supervised Learning



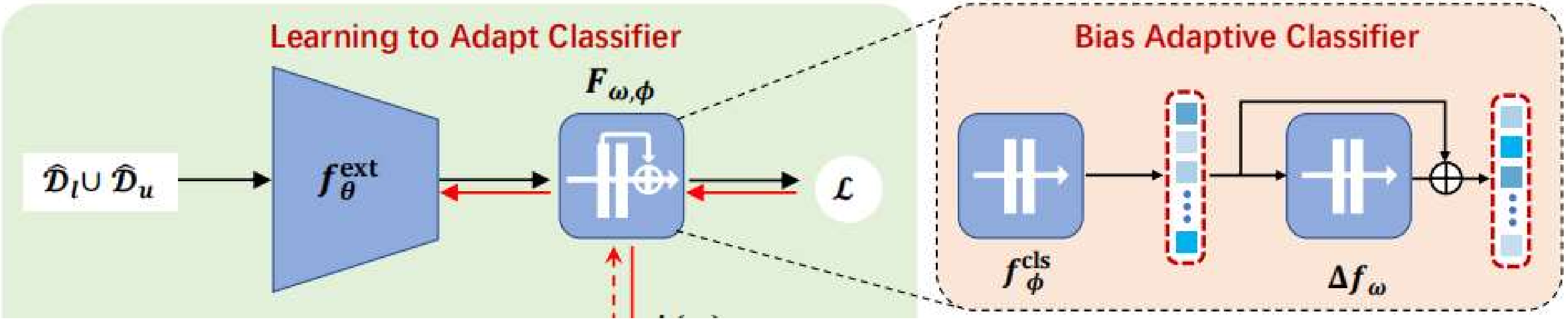
(a) Class distribution (b) Per-class recall (c) Predicted class distribution

Experiments on CIFAR-10-LT. Labeled set is class-imbalanced with imbalance ratio γ = 100, while the whole training data remains balanced.

Method

Bias Adaptive Classifier

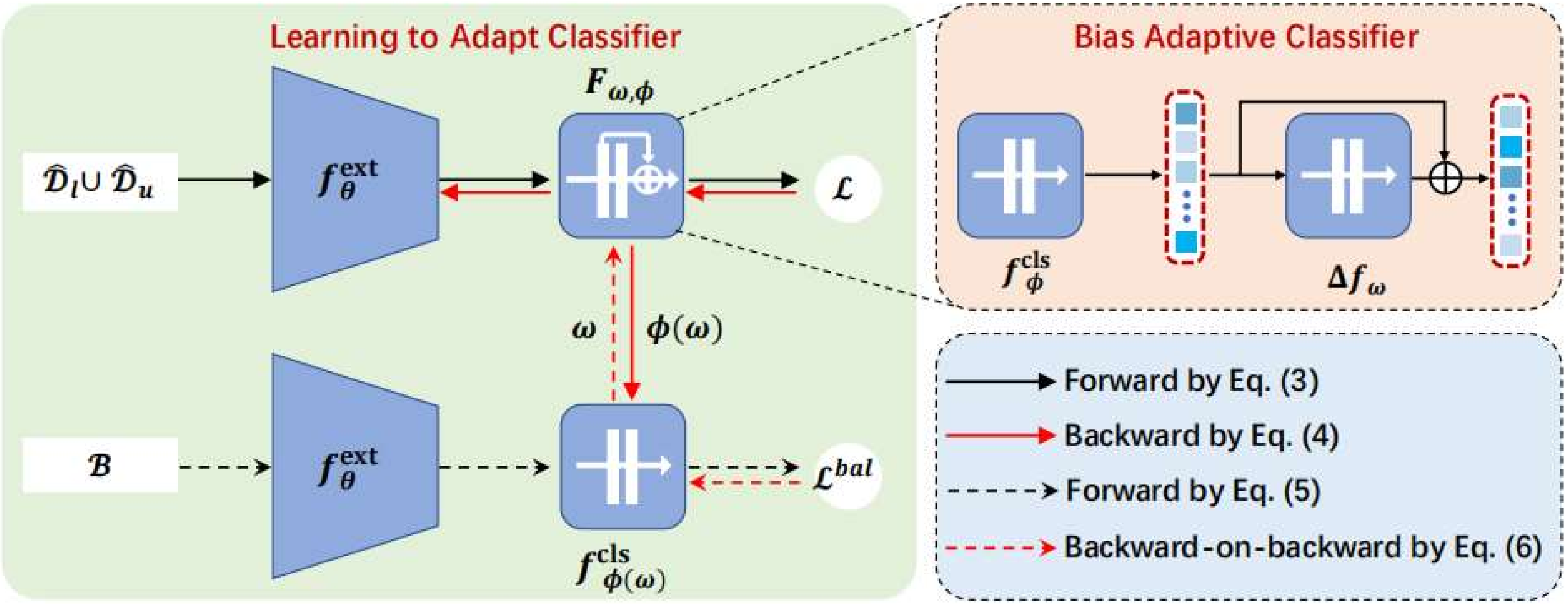
$$F_{\omega,\phi}(\mathbf{z}) = (\mathbf{I} + \Delta f_\omega) \circ f_\phi^{\text{cls}}(\mathbf{z}),$$



$$\mathcal{L} = \frac{1}{|\hat{\mathcal{D}}_l|} \sum_{\mathbf{x}_i \in \hat{\mathcal{D}}_l} \mathrm{H}(\hat{\Psi}(\mathbf{x}_i), \mathbf{y}_i) + \frac{1}{|\hat{\mathcal{D}}_u|} \sum_{\mathbf{x}_i \in \hat{\mathcal{D}}_u} \lambda_i \mathrm{H}(\hat{\Psi}(\mathbf{x}_i), \hat{\mathbf{y}}_i).$$

Method

Bias Adaptive Classifier



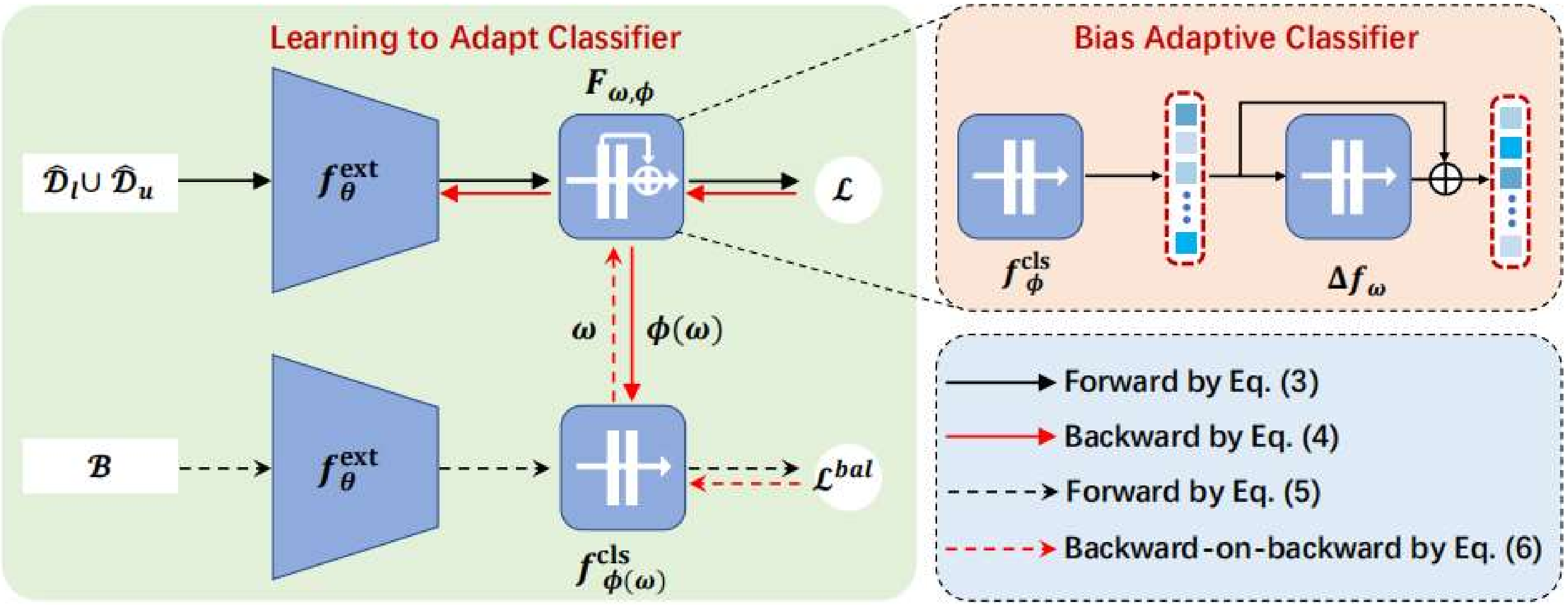
$$\mathcal{L} = \frac{1}{|\hat{\mathcal{D}}_l|}\sum_{\mathbf{x}_i \in \hat{\mathcal{D}}_l} \mathrm{H}(\hat{\Psi}(\mathbf{x}_i), \mathbf{y}_i) + \frac{1}{|\hat{\mathcal{D}}_u|}\sum_{\mathbf{x}_i \in \hat{\mathcal{D}}_u} \lambda_i \mathrm{H}(\hat{\Psi}(\mathbf{x}_i), \hat{\mathbf{y}}_i). \quad (3)$$

$$(\theta^{t+1}, \phi^{t+1}(\omega)) = (\theta^t, \phi^t) - \alpha \nabla_{\theta,\phi}\mathcal{L}, \quad (4)$$

$$\mathcal{L}^{bal} = \frac{1}{|\mathcal{B}|}\sum_{\mathbf{x}_i \in \mathcal{B}} \mathrm{H}(f_{\phi^{t+1}(\omega)}^{\mathrm{cls}} \circ f_{\theta^{t+1}}^{\mathrm{ext}}(\mathbf{x}_i), \mathbf{y}_i), \quad (5)$$

$$\omega^{t+1} = \omega^t - \eta \nabla_\omega \mathcal{L}^{bal}, \quad (6)$$

Complexity Analysis



However, in the backward pass, the second-order gradient of ω only requires to unroll the gradient graph of the linear classifier f_ϕ^{cls}.

$$\frac{\#\text{Params}(f_\phi^{cls})}{\#\text{Params}(\Psi)} \times \text{training time of one full backward pass.}$$

Method

ParNeC 模式识别与神经计算研究组 PAttern Recognition and NEural Computing

Theoretical Analysis

$$\mathcal{L}^{bal} = \frac{1}{|\mathcal{B}|} \sum_{\mathbf{x}_i \in \mathcal{B}} \mathrm{H}(f^{\text{cls}}_{\phi^{t+1}(\omega)} \circ f^{\text{ext}}_{\theta^{t+1}}(\mathbf{x}_i), \mathbf{y}_i),$$

In consequence, we have

$$\begin{aligned}
\omega^{t+1} &= \omega^t - \eta \nabla \mathcal{L}^{bal}(\theta, \phi^{t+1})|_{\omega^t} \\
&= \omega^t + \eta\alpha \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \mathcal{L}_i(\theta,\phi)}{\partial\phi\partial\omega}^T |_{\phi^t,\omega^t} \frac{\partial \mathcal{L}^{bal}(\theta,\phi)}{\partial\phi}|_{\phi^{t+1}} \\
&= \omega^t + \eta\alpha \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \Delta f_i}{\partial\omega}|_{\omega^t} \left(\frac{\partial^2 \mathcal{L}_i(\theta,\phi)}{\partial\phi\partial\Delta f_i}^T |_{\phi^t} \frac{\partial \mathcal{L}^{bal}(\theta,\phi)}{\partial\phi}|_{\phi^{t+1}} \right),
\end{aligned}$$

Denote by $\Xi_i = \frac{\partial \mathcal{L}_i(\theta,\phi)}{\partial \Delta f_i}$, *then Eq.* ($\boxed{9}$) *becomes*

$$\omega^{t+1} = \omega^t + \eta\alpha \frac{1}{n} \sum_{i=1}^{n} G_i \frac{\partial \Delta f_i}{\partial \omega}|_{\omega^t},$$

where

$$G_i = \left\langle \frac{\partial \Xi_i}{\partial\phi}|_{\phi^t}, \frac{\partial \mathcal{L}^{bal}(\theta,\phi)}{\partial\phi}|_{\phi^{t+1}} \right\rangle.$$

As the training loss is

$$\mathcal{L}_i(\theta,\phi) = \log \sum_{k=1}^{d} e^{z_{i,k}+\Delta f_{i,k}} - z_{i,c_i} - \Delta f_{i,c_i},$$

where c_i is the class label of the i-th sample. Therefore

$$\Xi_{i,k} = \begin{cases} \frac{e^{z_{i,k}+\Delta_{i,k}}}{\sum_{s=1}^{d} e^{z_{i,s}+\Delta_{i,s}}}, & k \neq c_i \\ \frac{e^{z_{i,k}+\Delta_{i,k}}}{\sum_{s=1}^{d} e^{z_{i,s}+\Delta_{i,s}}} - 1, & k = c_i \end{cases} = \mathbf{p}_i - \mathbf{y}_i$$

Meanwhile, the upper level loss is defined as

$$\mathcal{L}^{bal}(\theta,\phi) = -\frac{1}{m} \sum_{j=1}^{m} \mathcal{L}_i^{bal}(\theta,\phi),$$

$$G_i = \frac{\partial(\mathbf{p}_i - \mathbf{y}_i)}{\partial\phi}|_{\phi^t}^T \left(\frac{1}{m} \sum_{j=1}^{m} \frac{\partial \mathcal{L}_j^{\text{bal}}(\theta,\phi)}{\partial\phi}|_{\phi^{t+1}}\right)$$

Method

Algorithm 1 learning to adapt classifier during training

Input: labeled / unlabeled training data $\mathcal{D}_l$ / $\mathcal{D}_u$, labeled / unlabeled batch size n / m, max iterations T

Output: classification network parameters $\{\theta, \phi\}$

Initialize $\{\theta^0, \phi^0\} \leftarrow \{\theta, \phi\}$ and $\omega^0 \leftarrow \omega$.
for $t = 0$ **to** T **do**
    $\hat{\mathcal{D}}_l = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{n} \leftarrow \text{SampleMiniBatch}(\mathcal{D}_l, n)$.
    $\hat{\mathcal{D}}_u = \{\mathbf{x}_i\}_{i=1}^{m} \leftarrow \text{SampleMiniBatch}(\mathcal{D}_u, m)$.
    $\mathcal{B} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{n} \leftarrow \text{SampleMiniBatch}(\mathcal{D}_l, n)$.
    Estimate pseudo-label $\hat{\mathbf{y}}_i$ for $\mathbf{x}_i \in \hat{\mathcal{D}}_u$.
    Compute lower-level loss $\mathcal{L}$ by Eq. (3).
    Update network parameters $\{\theta^{t+1}, \phi^{t+1}\}$ by Eq. (4).
    Compute upper-level loss $\mathcal{L}^{bal}$ by Eq. (5).
    Update bias attractor parameters ω^{t+1} by Eq. (6).
end for

Experiment

We adopt balanced accuracy (bACC) and geometric mean scores (GM) as the evaluation metrics.

Methods	CIFAR-10 ($\gamma_l = \gamma_u$)		CIFAR-10 ($\gamma_l \neq \gamma_u$)	
	$\gamma = 100$	$\gamma = 150$	$\gamma_u = 1$ (uniform)	$\gamma_u = 100$ (reversed)
Vanilla	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$	$55.6_{\pm0.43}$ / $44.0_{\pm0.98}$	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$
w/ Re-sampling	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$	$52.2_{\pm0.05}$ / $38.2_{\pm1.49}$	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$
w/ LDAM-DRW	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$	$57.9_{\pm0.20}$ / $50.4_{\pm0.30}$	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$
w/ cRT	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$	$59.3_{\pm0.10}$ / $54.6_{\pm0.72}$	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$
MixMatch	$64.8_{\pm0.28}$ / $49.0_{\pm2.05}$	$62.5_{\pm0.31}$ / $42.5_{\pm1.68}$	$41.5_{\pm0.76}$ / $12.0_{\pm1.34}$	$47.9_{\pm0.09}$ / $20.5_{\pm0.85}$
w/ DARP	$67.9_{\pm0.14}$ / $61.2_{\pm0.15}$	$65.8_{\pm0.52}$ / $56.5_{\pm2.08}$	$86.7_{\pm0.80}$ / $86.2_{\pm0.82}$	$72.9_{\pm0.24}$ / $71.0_{\pm0.32}$
w/ SaR	$66.8_{\pm0.92}$ / $59.9_{\pm1.32}$	$64.4_{\pm2.21}$ / $57.3_{\pm1.95}$	$68.4_{\pm3.20}$ / $62.0_{\pm2.17}$	$65.5_{\pm1.01}$ / $64.2_{\pm0.95}$
w/ DASO	$69.8_{\pm1.10}$ / $69.3_{\pm1.07}$	$66.5_{\pm1.99}$ / $65.4_{\pm2.25}$	$75.5_{\pm0.48}$ / $74.6_{\pm0.67}$	$65.7_{\pm1.01}$ / $62.0_{\pm1.23}$
w/ ABC	$75.7_{\pm0.76}$ / $74.7_{\pm0.47}$	$68.5_{\pm0.40}$ / $56.4_{\pm1.50}$	$72.1_{\pm0.53}$ / $41.2_{\pm4.40}$	$62.9_{\pm0.36}$ / $59.9_{\pm0.60}$
w/ L2AC (ours)	$\mathbf{76.6}_{\pm0.73}$ / $\mathbf{75.7}_{\pm1.08}$	$\mathbf{72.1}_{\pm0.62}$ / $\mathbf{70.3}_{\pm0.93}$	$\mathbf{87.2}_{\pm0.09}$ / $\mathbf{86.7}_{\pm0.08}$	$\mathbf{74.0}_{\pm0.82}$ / $\mathbf{72.9}_{\pm1.01}$
FixMatch	$71.5_{\pm0.72}$ / $66.8_{\pm1.51}$	$68.4_{\pm0.15}$ / $59.9_{\pm0.43}$	$68.9_{\pm1.95}$ / $42.8_{\pm8.11}$	$65.5_{\pm0.05}$ / $26.0_{\pm0.44}$
w/ DARP	$75.5_{\pm0.05}$ / $73.0_{\pm0.09}$	$70.4_{\pm0.25}$ / $64.9_{\pm0.17}$	$85.4_{\pm0.55}$ / $85.0_{\pm0.65}$	$74.9_{\pm0.51}$ / $72.3_{\pm1.13}$
w/ CReST+	$77.5_{\pm0.15}$ / $76.1_{\pm0.15}$	$72.1_{\pm0.74}$ / $68.9_{\pm1.29}$	N/A	N/A
w/ SaR	$77.6_{\pm0.42}$ / $75.9_{\pm0.76}$	$71.5_{\pm0.23}$ / $66.9_{\pm0.25}$	$85.9_{\pm0.68}$ / $85.3_{\pm0.53}$	$78.3_{\pm0.34}$ / $76.1_{\pm0.21}$
w/ DASO	$78.3_{\pm0.55}$ / $76.5_{\pm0.57}$	$74.6_{\pm0.74}$ / $71.7_{\pm0.52}$	$87.9_{\pm0.41}$ / $87.7_{\pm0.43}$	$79.5_{\pm0.91}$ / $78.9_{\pm0.96}$
w/ ABC	$80.2_{\pm0.42}$ / $78.9_{\pm1.29}$	$74.7_{\pm1.04}$ / $72.2_{\pm1.45}$	$81.3_{\pm0.34}$ / $80.2_{\pm0.36}$	$70.3_{\pm0.50}$ / $67.9_{\pm0.70}$
w/ L2AC (ours)	$\mathbf{82.1}_{\pm0.57}$ / $\mathbf{81.5}_{\pm0.64}$	$\mathbf{77.6}_{\pm0.53}$ / $\mathbf{75.8}_{\pm0.71}$	$\mathbf{89.5}_{\pm0.18}$ / $\mathbf{89.2}_{\pm0.19}$	$\mathbf{82.2}_{\pm1.23}$ / $\mathbf{81.7}_{\pm1.36}$

It can be observed that L2AC significantly improves MixMatch and FixMatch at least **9**% absolute gain on bACC and at least **14**% on GM for all settings.

Experiment

Methods	CIFAR-100 ($\gamma_l = \gamma_u$)		STL-10 (γ_u =N/A)	
	$\gamma_l = 10$	$\gamma_l = 20$	$\gamma_l = 10$	$\gamma_l = 20$
FixMatch	$55.1_{\pm0.09}$ / $46.7_{\pm0.53}$	$49.5_{\pm0.38}$ / $34.2_{\pm1.01}$	$69.6_{\pm0.60}$ / $62.6_{\pm1.11}$	$65.5_{\pm0.05}$ / $26.0_{\pm0.44}$
w/ DARP	$56.3_{\pm0.25}$ / $48.2_{\pm0.73}$	$50.2_{\pm0.18}$ / $36.0_{\pm0.60}$	$72.9_{\pm0.24}$ / $69.5_{\pm0.18}$	$74.9_{\pm0.51}$ / $72.3_{\pm1.13}$
w/ ABC	$58.2_{\pm0.08}$ / $51.8_{\pm0.25}$	$\mathbf{53.1}_{\pm0.19}$ / $42.2_{\pm0.82}$	$78.2_{\pm0.35}$ / $77.3_{\pm0.30}$	$72.7_{\pm0.08}$ / $70.6_{\pm0.22}$
w/ DASO	$\mathbf{58.3}_{\pm0.39}$ / $51.4_{\pm0.80}$	$53.0_{\pm0.27}$ / $39.5_{\pm1.45}$	$78.2_{\pm0.63}$ / $77.4_{\pm0.53}$	$75.4_{\pm0.81}$ / $74.4_{\pm1.00}$
w/ L2AC (ours)	$57.8_{\pm0.19}$ / $\mathbf{52.1}_{\pm0.31}$	$52.6_{\pm0.13}$ / $\mathbf{43.0}_{\pm0.45}$	$\mathbf{79.9}_{\pm0.52}$ / $\mathbf{79.1}_{\pm0.49}$	$\mathbf{77.0}_{\pm0.65}$ / $\mathbf{75.8}_{\pm0.68}$

It is worth noting that the unlabeled set of STL-10 is noisy as it contains samples that do not belong to any of the classes in the labeled set.

L2AC significantly outperforms ABC and DASO on both bACC and GM, which demonstrates that it has greater potential to be applied in the practical SSL scenarios.

Experiment

Methods	bACC/GM	Methods	bACC/GM
Vanilla	$38.3_{\pm 0.05}$ / $29.9_{\pm 0.08}$	DARP (Kim et al., 2020a)	$45.5_{\pm 0.32}$ / $37.5_{\pm 0.04}$
cRT (Kang et al., 2020)]	$39.3_{\pm 0.21}$ / $33.7_{\pm 0.37}$	ABC (Lee et al., 2021)	$47.0_{\pm 0.26}$ / $39.2_{\pm 0.34}$
FixMatch Sohn et al. (2020)	$44.9_{\pm 0.11}$ / $35.7_{\pm 0.66}$	L2AC (ours)	$\mathbf{48.8}_{\pm 0.19}$ / $\mathbf{40.6}_{\pm 0.17}$

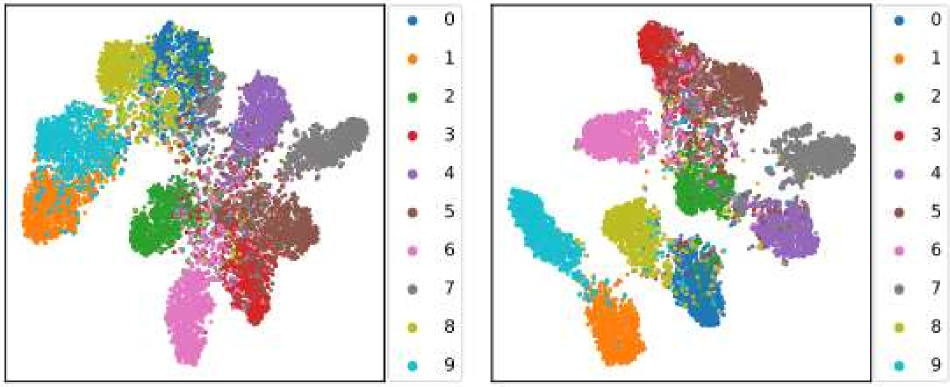
This further verifies the efficacy of our proposed method toward the real-world imbalanced SSL applications.

Experiment

CIFAR-10 ($\gamma_l = \gamma_u = 100$)	$\beta = 1$	$\beta = 5$	$\beta = 10$	β=20	$\beta = 30$
FixMatch	54.9 / 16.5	65.1 / 35.5	69.0 / 53.9	72.0 / 62.2	76.5 / 74.3
w/ L2AC (ours)	62.8 / 55.8	75.9 / 74.1	79.3 / 78.4	80.8 / 79.9	83.6 / 83.2
STL-10 ($\gamma_l = 10, \gamma_u =$ N/A)	**$\beta = 5$**	**$\beta = 10$**	**$\beta = 20$**	**β=40**	**$\beta = 60$**
FixMatch	46.5 / 19.9	48.8 / 27.0	58.2 / 39.8	67.2 / 60.7	69.2 / 67.6
w/ L2AC (ours)	62.8 / 57.1	66.5 / 62.9	72.6 / 70.6	77.0 / 75.7	78.8 / 77.9

In such an extremely biased scenario, our L2AC significantly improves FixMatch by around **16**% over bACC and **37**% over GM.

Experiment



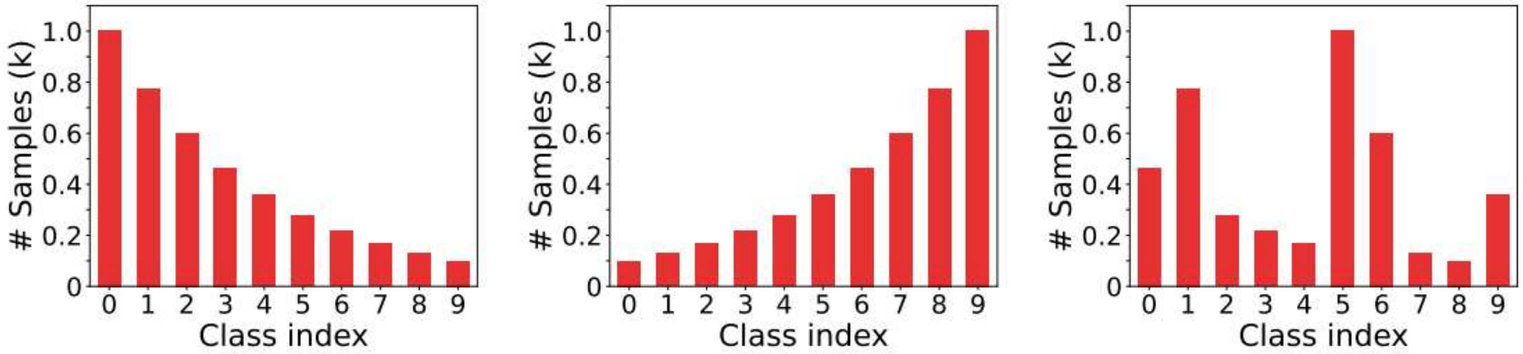
(a) FixMatch (b) L2AC (ours)

Figure 5: t-SNE visualization of training data for (a) FixMatch and (b) L2AC. L2AC helps to discriminate tail classes from majority ones.

Table 5: Ablation study.

Methods	CIFAR-10 ($\gamma_l = 100$)	
	$\gamma_u = 100$	1/100 (reversed)
FixMatch	71.5 / 68.8	65.5 / 26.0
FixMatch w/ bias attractor	73.9 / 70.7	66.6 / 44.8
L2AC w/o bi-level training	78.4 / 76.6	79.3 / 78.0
L2AC (ours)	82.1 / 81.5	82.2 / 81.7

Experiment



(a) Test-1 (b) Test-2 (c) Test-3

Table 6: Imbalanced test set results. ACC: accuracy for all samples.

Methods	Test-1			Test-2			Test-3		
	bACC	GM	ACC	bACC	GM	ACC	bACC	GM	ACC
FixMatch (Sohn et al., 2020)	72.4	66.3	86.1	72.7	66.9	56.6	72.3	65.7	75.6
w/ DARP (Kim et al., 2020a)	74.8	72.5	86.3	75.5	73.2	63.9	75.6	73.3	77.4
w/ CReST (Wei et al., 2021)	77.8	76.5	86.3	77.2	74.8	68.9	77.5	76.3	80.3
w/ ABC (Lee et al., 2021)	80.2	79.2	**88.1**	80.2	79.0	71.7	80.1	79.0	82.8
w/ L2AC (ours)	**82.6**	**82.0**	87.2	**82.4**	**81.8**	**78.6**	**82.7**	**82.1**	**83.9**

All these models are trained on CIFAR-10 with imbalanced ratio γ = γl = γu = 100.

Thanks