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

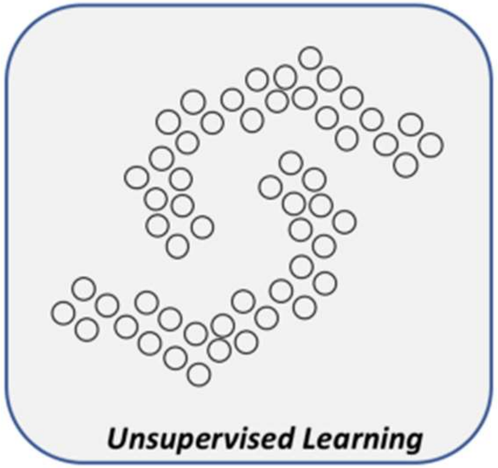
ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

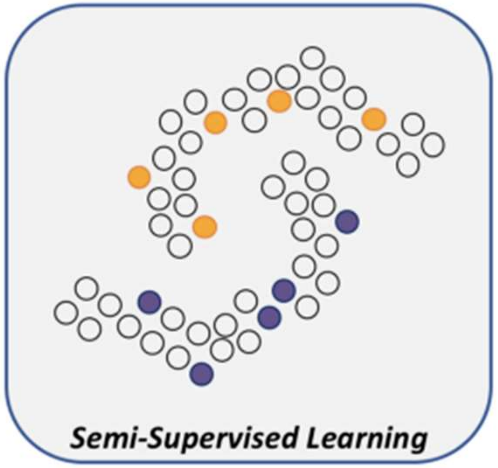
Semi-Supervised Learning

Class Imbalance & Inconsistent Distribution

Semi-supervised Learning (SSL)

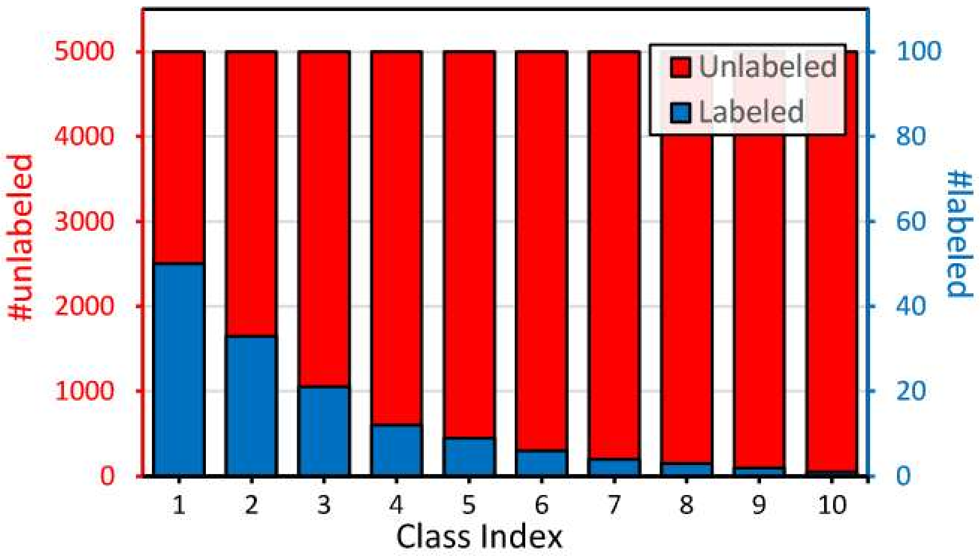
$$\arg\min_{\theta} \underbrace{\sum_{(x,y)\in D_L} \mathcal{L}_s(x,y;\theta) + \sum_{x\in D_U}}_{} \mathcal{L}_u(x;\theta)$$





Background

Class Imbalance & Inconsistent Distribution



Related Work

Class Imbalance

- re-sampling
- re-weighting

Inconsistent Distribution

- ?

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

Smoothed Adaptive Weighting for Imbalanced Semi-Supervised Learning: Improve Reliability Against Unknown Distribution Data

Zhengfeng Lai [*1] Chao Wang [*2] Henrry Gunawan [1] Sen-Ching Cheung [3] Chen-Nee Chuah [1]

ICML 2022

Motivation

~~re-sampling~~

As **re-sampling** based techniques may introduce ***significant computational cost !***

re-weighting

(*easy*) ① **The unlabeled data share similar distribution as the labeled set.**

(*difficult*) ② **The distribution of unlabeled data is unknown.**

(*easy*) ① The unlabeled data share similar distribution as the labeled set.

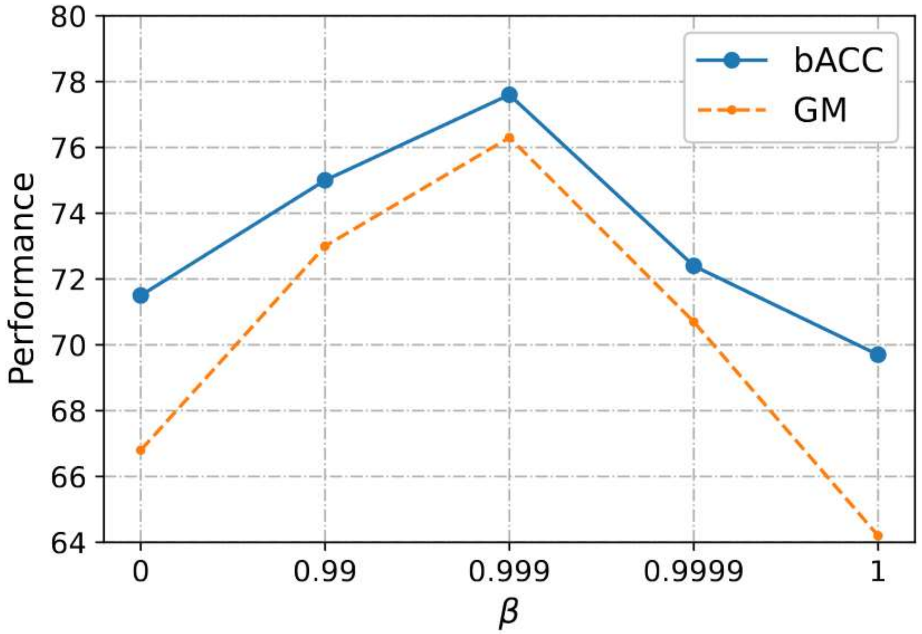
The consistency loss:

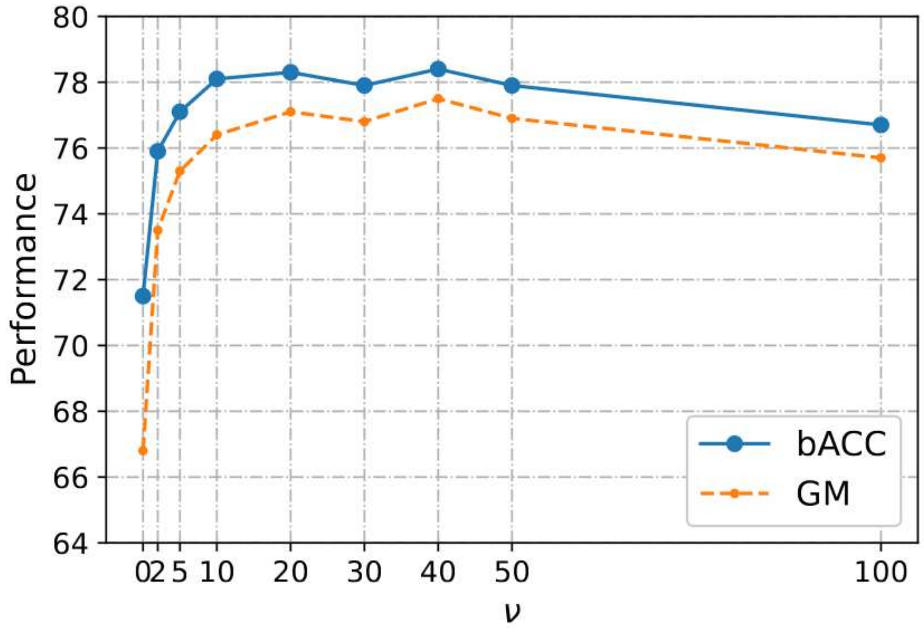
$$\mathcal{L}_c(x;\theta) := \sum_{k=1}^{C} p(x;\theta)_k \cdot \log(h(\mathtt{pt}(x);\theta)_k).$$

The weighted consistency loss:

$$\mathcal{L}_{cw}(x;w,\theta) := \sum_{k=1}^{C} \boxed{w_k} \cdot p(x;\theta)_k \cdot \log(h(\mathtt{pt}(x);\theta)_k)$$

The smoothing functions:





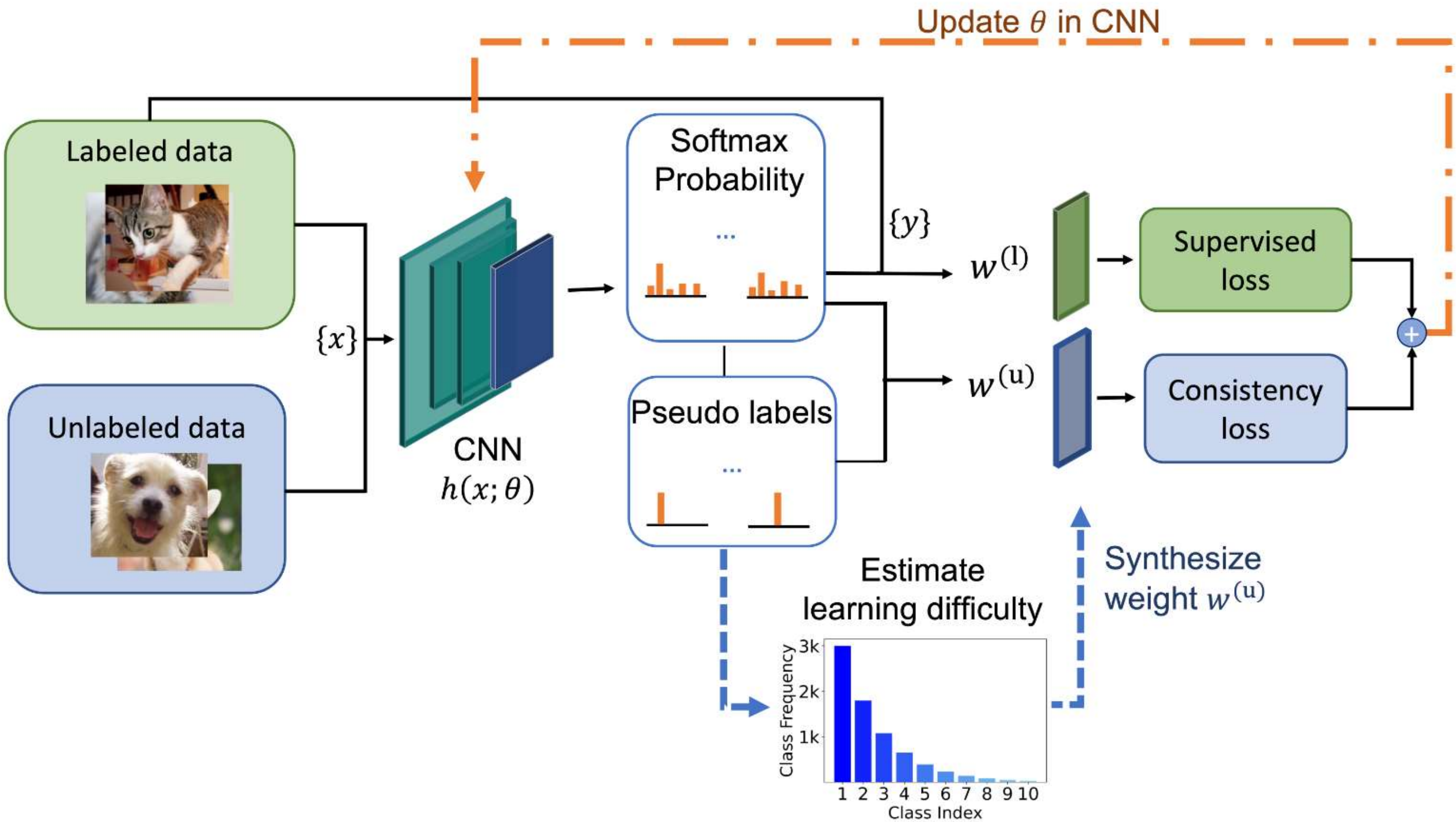
$$w_k \propto 1/E_k, \text{ where } E_k = (1-\beta^{n_k})/(1-\beta)$$

$$w_k \propto \frac{n}{n+\nu \cdot n_k}, \text{ for } k = 1, \dots, C$$

——***Classbalanced loss based on effective number of samples. (CVPR2019)*** (*difficult*) ② **The distribution of unlabeled data is unknown.**

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

SAW: Smoothed adaptive weighting against unknown distribution data



SAW: Smoothed adaptive weighting against unknown distribution data

Algorithm 1 The SAW Framework

Data: Labeled data $\{(x_i, y_i)\}_{i=1}^m$, unlabeled data $\{x_j\}_{j=1}^n$, weight in supervised loss $w^{(\mathrm{l})}$, number of classes C, learning rate η, max epoch T

Initialize uniform weights in consistency loss $w^{(\mathrm{u})}$ and model parameter θ

while $t < T$ **do**

  for $1, \ldots K$ **do**

    Sample batches label data $\{x, y\}$

    Sample batches unlabeled data $\{x\}$

    Update $\theta \leftarrow \theta - \eta \cdot \nabla \mathcal{L}(\theta, w^{(\mathrm{u})})$

  end for

  Compute pseudo labels $p(x; \theta)$ and estimate distribution $(n_1, n_2, \ldots, n_C)$

  Adjust the distribution $(\hat{n}_1, \hat{n}_2, \ldots, \hat{n}_C)$

  Update weight $w^{(\mathrm{u})}$ based on smoothed weighting function

end while

Return: θ

$$\sum_{i=1}^m \mathcal{L}_{lw}(x_i^{(\mathrm{l})}, y_i^{(\mathrm{l})}; w^{(\mathrm{l})}, \theta) + \sum_{j=1}^n \mathcal{L}_{cw}(x_j^{(\mathrm{u})}; w^{(\mathrm{u})}, \theta)$$

$$\hat{n}_k = \max(n_k, 1)$$

$$w_k \propto 1/E_k, \text{ where } E_k = (1 - \beta^{n_k})/(1 - \beta)$$

Experiment (*easy*) ① The unlabeled data share similar distribution as the labeled set.

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

Table 4. Comparison of classification performance on CIFAR10-LT under $\gamma = \gamma_l = \gamma_u$ (hold-out test set is of reversed distribution). The evaluation criterion is bACC/GM. The best results are in bold.

Algorithm	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$
ReMixMatch (Berthelot et al., 2020)	71.0±0.55 / 83.5±0.29	54.7±0.51 / 74.4±0.47	41.5±1.69 / 66.4±1.22
ReMixMatch + DARP (Kim et al., 2020)	66.9±0.75 / 80.5±0.46	49.7±1.55 / 70.5±0.90	35.8±1.81 / 60.9±2.42
ReMixMatch + CReST (Wei et al., 2021)	64.3±0.25 / 75.7±0.34	51.2±0.92 / 72.1±0.85	39.2±1.46 / 65.8±1.88
ReMixMatch + SAW	**86.3**±0.61 / **86.1**±0.64	**77.0**±0.59 / **76.0**±0.42	**71.5**±0.30 / **68.9**±0.26
FixMatch (Sohn et al., 2020)	70.5±0.26 / 82.2±0.31	51.0±1.65 / 71.5±1.24	38.5±1.15 / 63.4±0.31
FixMatch + DARP (Kim et al., 2020)	72.2±0.62 / 82.8±0.17	57.6±0.36 / 74.8±0.48	46.5±1.26 / 68.1±0.10
FixMatch + CReST (Wei et al., 2021)	69.4±0.35 / 80.1±0.41	52.4±0.32 / 70.3±0.28	42.9±1.45 / 67.4±1.07
FixMatch + SAW	**78.7**±0.77 / **84.2**±0.36	**64.3**±1.96 / **76.4**±0.88	**57.5**±2.83 / **70.5**±1.50

Table 9. Comparison of bACC on CIFAR10-LT under $\gamma = \gamma_l = \gamma_u$. The hold-out test set is of the **same** distribution of the training set (imbalanced).

	CIFAR10-LT		
Algorithm	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$
FixMatch (Sohn et al., 2020)	89.2±0.59	85.4±0.59	83.1±0.33
FixMatch + DARP (Kim et al., 2020)	93.1±1.95	90.1±0.98	88.5±0.89
FixMatch + CReST (Wei et al., 2021)	86.7±1.98	84.3±1.20	82.4±1.46
FixMatch + SAW	**94.7**±1.23	**92.4**±0.35	**90.5**±0.47

Experiment (*difficult*) ② The distribution of unlabeled data is unknown.

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

Table 2. Comparison of classification performance on CIFAR10-LT under $\gamma_l = 100, \gamma_u \neq \gamma_l$ (hold-out test set is balanced). The evaluation criterion is bACC/GM. The best results are in bold.

Algorithm	$\gamma_u = 1$	$\gamma_u = 50$	$\gamma_u = 150$
ReMixMatch (Berthelot et al., 2020)	48.3±0.14 / 19.5±0.85	75.1±0.43 / 71.9±0.77	72.5±0.10 / 68.2±0.32
ReMixMatch* (Berthelot et al., 2020)	85.0±1.35 / 84.3±1.55	77.0±0.12 / 74.7±0.04	72.8±0.10 / 68.8±0.21
ReMixMatch* + DARP (Kim et al., 2020)	**89.7**±0.15 / **89.4**±0.17	77.4±0.22 / 75.0±0.25	73.2±0.11 / 69.2±0.31
ReMixMatch* + CReST (Wei et al., 2021)	45.9±1.27 / 20.1±1.99	70.2±0.45 / 65.8±0.71	65.4±0.34 / 62.9±0.15
ReMixMatch* + SAW	88.3±0.15 / 88.9±0.10	**80.3**±0.36 / **79.6**±0.40	**74.0**±0.94 / **72.4**±0.94
FixMatch (Sohn et al., 2020)	68.9±1.95 / 42.8±8.11	73.9±0.25 / 70.5±0.52	69.6±0.60 / 62.6±1.11
FixMatch + DARP (Kim et al., 2020)	**85.4**±0.55 / **85.0**±0.65	77.3±0.17 / 75.5±0.21	72.9±0.24 / 69.5±0.18
FixMatch + CReST (Wei et al., 2021)	60.2±1.34 / 35.9±2.50	65.8±0.78 / 67.1±0.84	60.1±1.44 / 51.4±1.68
FixMatch + SAW	83.9±0.44 / 83.3±0.47	**81.5**±2.25 / **80.9**±2.30	**76.8**±0.31 / **75.4**±0.37

Table 7. Comparison of classification performance on CIFAR10-LT under $\gamma_l = 100, \gamma_u \neq \gamma_l$ (hold-out test set is of reversed distributions). The evaluation criterion is bACC/GM. The best results are in bold.

	CIFAR-10 ($\gamma_l = 100$)		
Algorithm	$\gamma_u = 1$	$\gamma_u = 50$	$\gamma_u = 150$
ReMixMatch (Berthelot et al., 2020)	70.0±1.39 / 48.3±0.53	61.4±0.21 / 77.3±0.41	48.6±1.62 / 71.3±0.67
ReMixMatch* + DARP (Kim et al., 2020)	67.3±1.63 / 49.5±2.01	54.5±1.30 / 72.6±0.58	44.6±1.69 / 67.7±1.16
ReMixMatch* + SAW	**84.0**±0.05 / **83.7**±0.42	**72.7**±0.56 / **79.6**±1.11	**59.8**±1.03 / **73.6**±1.19
FixMatch (Sohn et al., 2020)	70.0±1.39 / 48.3±0.53	61.0±1.61 / 76.4±0.52	48.0±2.26 / 70.4±0.88
FixMatch + DARP (Kim et al., 2020)	68.1±0.37 / 48.5±1.46	59.5±1.41 / 75.2±1.08	55.1±1.31 / 73.8±0.43
FixMatch + SAW	**84.0**±0.39 / **83.4**±0.42	**76.4**±6.71 / **81.5**±1.97	**63.3**±0.37 / **75.4**±1.23

Experiment

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

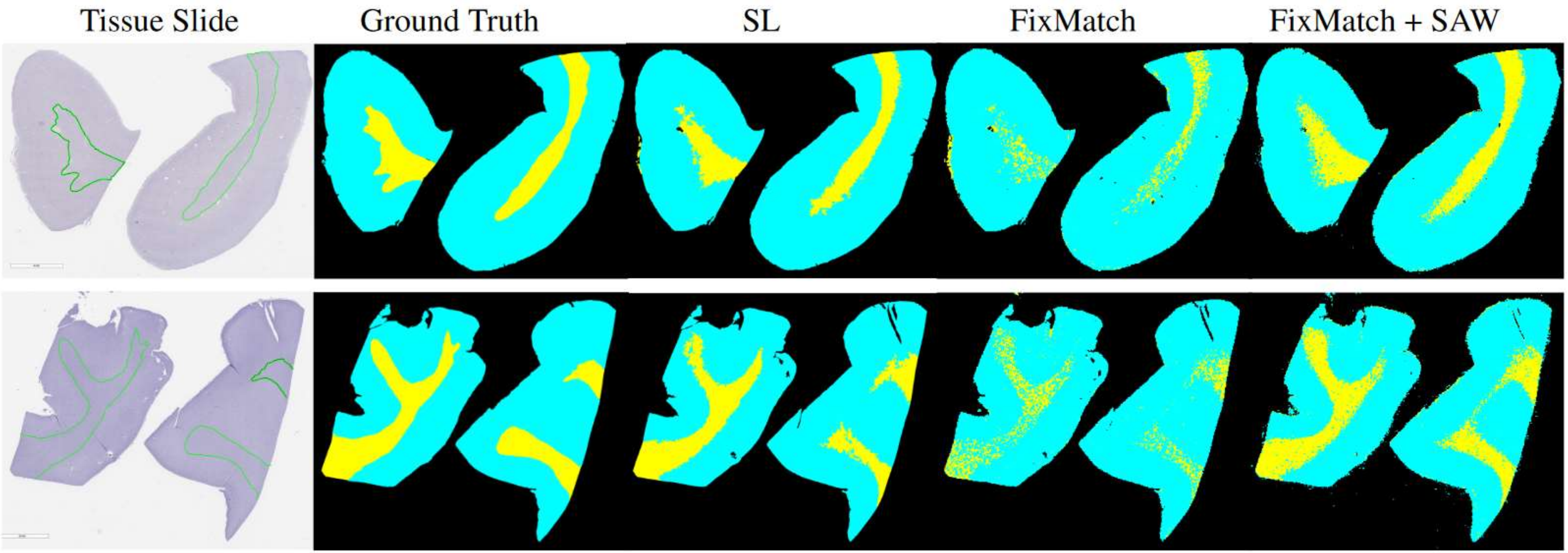


Figure 5. Segmentation masks visualization: GM, WM, and background are indicated by cyan, yellow, and black, respectively. SL refers to train supervised learning (U-Net (Oskal et al., 2019)) on all 20 pathology images with their annotations. The SSL results are only using 0.1% area of 2 pathology images selected from the training set.

On Non-Random Missing Labels in Semi-Supervised Learning

Xinting Hu[1] Yulei Niu[1]* Chunyan Miao[1] Xian-Sheng Hua[2] Hanwang Zhang[1]
[1]Nanyang Technological University, [2]Damo Academy, Alibaba Group
xinting001@e.ntu.edu.sg, yn.yuleiniu@gmail.com
{ascymiao, hanwangzhang}@ntu.edu.sg, xiansheng.hxs@alibaba-inc.com

ICLR 2022

Motivation

~~Missing Completely At Random (MCAR)~~

Due to the imbalanced human preferences for the "class"!

Missing Not At Random (MNAR)

$$\mathbb{E}[\hat{y}] = \mathbb{E}[y|\hat{\theta}] = \sum_{(x,y)\in D_L} y \cdot P(y|x) = \sum_{(x,y)\in D} y \cdot P(y|x, \underline{M=0})$$

$$\neq \sum_{(x,y)\in D} y \cdot P(y|x) = \mathbb{E}[y],$$

Labeled data $(M=0)$

Unlabeled data $(M=1)$

Class-Aware Doubly Robust (CADR) Estimator

① for labeled data

$$\hat{\theta}_{\text{CADR}} = \arg\min_{\theta} \mathcal{L}_{\text{CADR}} = \arg\min_{\theta} \boxed{\mathcal{L}_{\text{CAP}}} + \boxed{\mathcal{L}_{\text{CAI}}} + \boxed{\mathcal{L}_{\text{supp}}}$$

② for unlabeled data

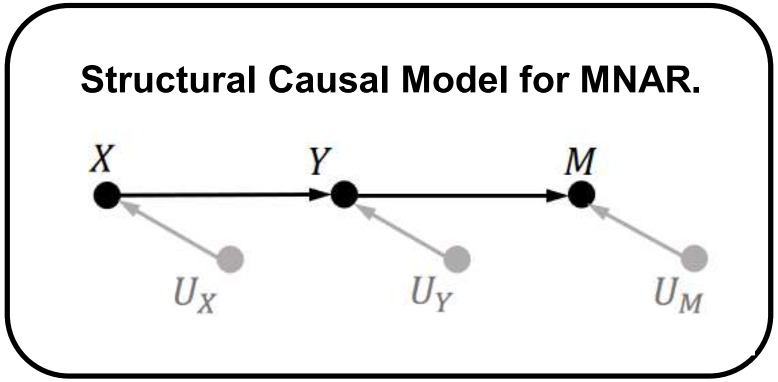
③ to guarantee the unbiasedness

① Class-Aware Propensity (CAP)

Traditional SSL methods:

$$\hat{\theta} = \arg\max_{\theta} \log P(Y|X, M = 0; \theta) = \arg\max_{\theta} \sum_{(x,y)\in D_L} \log P(y|x; \theta)$$

Structural Causal Model for MNAR:



$$P(X|Y, M = 0) = P(X|Y)$$

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

① Class-Aware Propensity (CAP)

$$\hat{\theta} = \arg\max_{\theta} \log P(X|Y;\theta) = \arg\max_{\theta} \sum\nolimits_{(x,y)\in D_L} \log P(x|y;\theta) \tag{5}$$

$$= \arg\max_{\theta} \sum\nolimits_{(x,y)\in D_L} \log \frac{P(y|x;\theta)P(x;\theta)}{P(y;\theta)} \tag{6}$$

$$= \arg\max_{\theta} \sum\nolimits_{(x,y)\in D_L} \log \frac{P(y|x;\theta)}{P(y;\theta)} \tag{7}$$

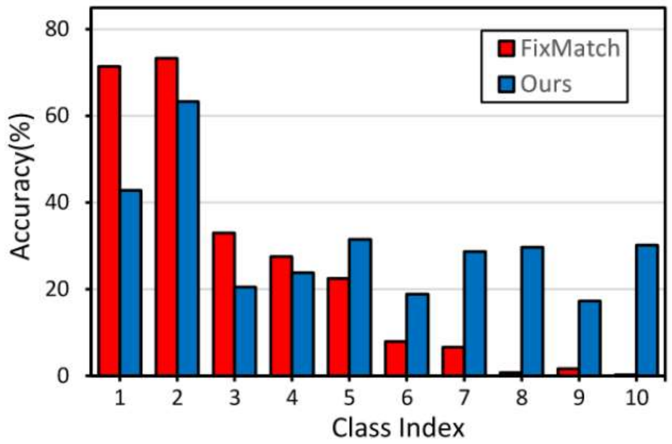
$$= \arg\max_{\theta} \sum\nolimits_{(x,y)\in D_L} \log P(y|x;\theta) \cdot \frac{\log P(y|x;\theta) - \boxed{\log P(y;\theta)}}{\log P(y|x;\theta)} \tag{8}$$

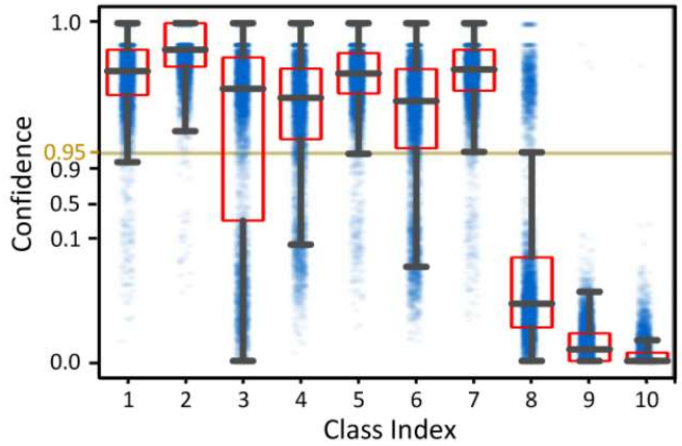
$$\triangleq \arg\max_{\theta} \sum\nolimits_{(x,y)\in D_L} \log P(y|x;\theta) \cdot \frac{1}{s(x,y)}. \tag{9}$$

EMA: $\boxed{\hat{P}(Y) \leftarrow \mu\hat{P}(Y) + (1-\mu)P(Y;B_t,\theta_t)}$

② Class-Aware Imputation (CAI)

CAP *vs* FixMatch:





Class-aware threshold:

$$\tau(x) = \tau_o \cdot \left(\frac{\hat{P}(C_x)}{\max_{y\in\{1,\cdots,C\}} \hat{P}(y)}\right)^{\beta}, \quad C_x = \arg\max_y P(y|x;\theta)$$

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

③ A supplementary loss

Rewrite the training objective of CAP and CAI:

$m^{(i)}$ — the missing state

$p^{(i)}$ — $\frac{\log P(y|x;\theta)}{\log P(y|x;\theta) - \log P(y;\theta)}$

$q^{(i)}$ — is the imputed label

$$\mathcal{L}_{\text{CAP}} = \frac{1}{N} \sum_{i=1,\cdots,N} \frac{(1 - m^{(i)})\mathcal{L}_s(x^{(i)}, y^{(i)})}{p^{(i)}}$$

$$\mathcal{L}_{\text{CAI}} = \frac{1}{N} \sum_{i=1,\cdots,N} (m^{(i)}\mathcal{L}_u(x^{(i)}, q^{(i)})\, \mathbb{I}(\text{con}(q^{(i)}) > \tau(x^{(i)})) + (1 - m^{(i)})\mathcal{L}_s(x^{(i)}, y^{(i)})),$$

$$\mathcal{L}_{\text{supp}} = \frac{1}{N} \sum_{i=1,\cdots,N} (1 - m^{(i)} - \frac{1 - m^{(i)}}{p^{(i)}})\mathcal{L}_u(x^{(i)}, q^{(i)})\, \mathbb{I}(\text{con}(q^{(i)}) > \tau) - \frac{1}{N} \sum_{i=1,\cdots,N} (1 - m^{(i)})\mathcal{L}_s(x^{(i)}, y^{(i)}),$$

Comparison Results

Method	CIFAR-10 γ = 20	CIFAR-10 50	CIFAR-10 100	CIFAR-100 50	CIFAR-100 100	CIFAR-100 200	STL-10 50	STL-10 100	mini-ImageNet 50	mini-ImageNet 100
Π Model	21.59	27.54	30.39	24.95	29.93	33.91	31.89	34.69	11.77	15.30
MixMatch	26.63	31.28	28.02	37.82	41.32	42.92	28.98	28.31	13.12	18.30
ReMixMatch	41.84	38.44	38.20	42.45	39.71	39.22	41.33	39.55	22.64	23.50
FixMatch	56.26	65.61	72.28	50.51	48.82	50.62	47.22	57.01	23.56	26.57
+ CREST	51.10	55.40	63.60	40.30	46.30	49.60	–	–	–	–
+ DARP	63.14	70.44	74.74	38.87	40.49	44.15	39.66	39.72	–	–
+ CADR (Ours)	**79.63**	**93.79**	**93.97**	**59.53**	**60.88**	**63.30**	**70.29**	**76.70**	**29.07**	**32.78**

Table 1: A Comparison of mean accuracies (%). We alter the imbalance ratio γ of labeled data and leave the unlabeled data balanced ($\gamma_u = 1$). We keep $N_{max} = \gamma$ so that the least number of labeled data among all the classes is always 1.

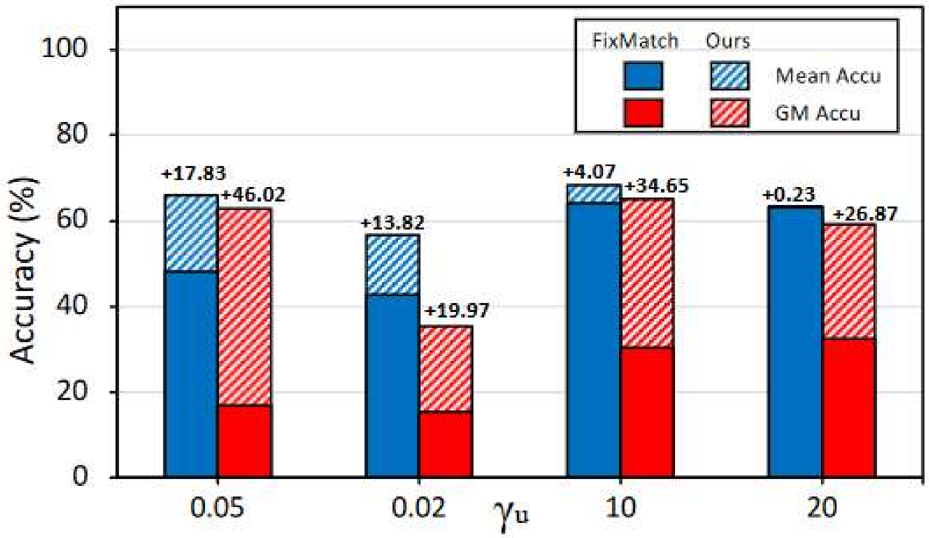
Ablation study

	CIFAR-10			CIFAR-100			STL-10		mini-ImageNet	
Method	γ= *20*	*50*	*100*	*50*	*100*	*200*	*50*	*100*	*50*	*100*
FixMatch	56.26	65.61	72.28	50.51	48.82	50.62	47.22	57.01	23.56	26.57
w/ CAP	79.38	89.50	93.95	55.72	58.53	<u>63.07</u>	69.97	**77.69**	<u>28.54</u>	<u>32.23</u>
w/ CAI	79.02	88.15	93.86	<u>58.55</u>	<u>59.80</u>	58.26	**70.64**	71.21	27.15	30.94
w/o CADR	<u>79.43</u>	<u>89.53</u>	**94.10**	57.93	59.50	62.78	70.17	75.44	25.54	31.66
w/ CADR	**79.63**	**93.79**	<u>93.97</u>	**59.53**	**60.88**	**63.30**	<u>70.29</u>	<u>76.70</u>	**29.07**	**32.78**

Table 2: The individual performance of our proposed Class-Aware Propensity (CAP) and Class-Aware Imputation (CAI) alone and together in trivial combination (w/o CADR) and CADR combination (w/ CADR). We marked the **best** and <u>second-best</u> accuracies.

Experiment

More Settings and Baselines.



Method	Accuracy (%)
FixMatch	65.61
+ CREST	55.40 (-10.21)
+ DARP	70.44 (+4.83)
+ re-weighting	66.03 (+0.42)
+ re-sampling	65.49 (-0.12)
+ LA loss (τ=1)	60.22 (-5.39)
+ DASH	65.62 (+0.01)
+ CAP (Ours)	89.50 (+23.89)
+ CAI (Ours)	88.15 (+22.54)
+ CADR (Ours)	**93.79** (+28.18)

Table 3: Comparison with multiple baseline methods. Experiments are conducted on CIFAR-10 ($\gamma = N_{max} = 50$).

Thanks

Experiment (*easy*) ① The unlabeled data share similar distribution as the labeled set.

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

Table 1. Comparison of classification performance on CIFAR10-LT under $\gamma = \gamma_l = \gamma_u$ (hold-out test set is balanced). The evaluation criterion is bACC/GM. The best results are in bold.

Algorithm	SSL	RB	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$
Wide ResNet-28-2 (Oliver et al., 2018)	-	-	65.2±0.05 / 61.1±0.09	58.8±0.13 / 51.0±0.11	55.6±0.43 / 44.0±0.98
Re-sampling (Japkowicz, 2000)	-	✓	64.3±0.48 / 60.6±0.67	55.8±0.47 / 45.1±0.30	52.2±0.05 / 38.2±1.49
LDAM-DRW (Cao et al., 2019)	-	✓	68.9±0.07 / 67.0±0.08	62.8±0.17 / 58.9±0.60	57.9±0.20 / 50.4±0.30
cRT (Kang et al., 2020)	-	✓	67.8±0.13 / 66.3±0.15	63.2±0.45 / 59.9±0.40	59.3±0.10 / 54.6±0.72
ReMixMatch (Berthelot et al., 2020)	✓	-	81.5±0.26 / 80.2±0.32	73.8±0.38 / 69.5±0.84	69.9±0.47 / 62.5±0.35
ReMixMatch + cRT (Kang et al., 2019)	✓	✓	86.8±0.50 / 86.5±0.49	81.4±0.41 / 80.7±0.45	78.9±0.84 / 77.8±0.94
ReMixMatch + DARP (Kim et al., 2020)	✓	-	82.1±0.14 / 80.8±0.09	75.8±0.09 / 72.6±0.24	71.0±0.27 / 64.5±0.68
ReMixMatch + DARP + cRT (Kim et al., 2020)	✓	✓	87.3±0.61 / 87.0±0.11	83.5±0.07 / 83.1±0.09	79.7±0.54 / 78.9±0.49
ReMixMatch + CReST (Wei et al., 2021)	✓	✓	85.2±0.19 / 84.9±0.25	76.2±0.31 / 75.1±0.28	71.4±0.23 / 67.5±0.40
ReMixMatch + SAW	✓	-	86.3±0.61 / 86.1±0.64	77.0±0.59 / 76.0±0.42	71.5±0.30 / 68.9±0.26
ReMixMatch + SAW + cRT (Kang et al., 2019)	✓	✓	**87.6**±0.21 / **87.4**±0.26	**85.4**±0.32 / **83.9**±0.21	**79.9**±0.15 / **79.9**±0.12
FixMatch (Sohn et al., 2020)	✓	-	79.2±0.33 / 77.8±0.36	71.5±0.72 / 66.8±1.51	68.4±0.15 / 59.9±0.43
FixMatch + DARP (Kim et al., 2020)	✓	-	81.8±0.24 / 80.9±0.28	75.5±0.05 / 73.0±0.09	70.4±0.25 / 64.9±0.17
FixMatch + CReST (Wei et al., 2021)	✓	✓	83.0±0.39 / 81.5±0.17	75.7±0.38 / 72.7±0.85	70.8±0.25 / 64.5±0.31
FixMatch + CReST + & LA (Wei et al., 2021)	✓	✓	85.6±0.36 / 81.9±0.45	**81.2**±0.70 / 74.5±0.99	71.9±2.24 / 64.4±1.75
FixMatch + SAW	✓	-	84.0±0.10 / 83.6±0.12	77.5±0.63 / 76.3±0.80	71.6±0.35 / 69.7±0.46
FixMatch + SAW + & LA (Menon et al., 2020)	✓	✓	**86.2**±0.15 / **83.9**±0.35	80.7±0.15 / **77.5**±0.21	**73.7**±0.06 / **71.2**±0.17

Experiment

(*easy*) ① **The unlabeled data share similar distribution as the labeled set.**

(*difficult*) ② **The distribution of unlabeled data is unknown.**

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

Table 3. Comparison of classification performance on CIFAR100-LT under $\gamma = \gamma_l = \gamma_u$, and STL-10 where the distribution of unlabeled data is unknown. Hold-out test set is balanced. The evaluation criterion is bACC/GM. The best results are in bold.

Algorithm	CIFAR100-LT ($\gamma = \gamma_l = \gamma_u$) $\gamma = 10$	CIFAR100-LT ($\gamma = \gamma_l = \gamma_u$) $\gamma = 20$	STL-10 $\gamma_l = 10$	STL-10 $\gamma_l = 20$
ReMixMatch (Berthelot et al., 2020)	59.2±0.03 / 52.1±0.13	53.5±0.03 / 42.3±0.13	67.8±0.45 / 61.1±0.92	60.1±1.18 / 44.9±1.52
ReMixMatch* + DARP (Kim et al., 2020)	59.8±0.20 / 52.9±0.41	54.4±0.07 / 44.2±0.07	79.4±0.07 / 78.2±0.10	70.9±0.44 / 67.0±1.62
ReMixMatch* + CReST (Wei et al., 2021)	59.6±0.32 / 52.5±0.24	53.9±0.15 / 43.9±0.19	65.3±0.23 / 59.3±0.41	55.8±2.05 / 40.2±2.39
ReMixMatch* + SAW	**61.8**±0.06 / **56.9**±0.40	**55.3**±0.26 / **46.3**±0.65	**82.0**±0.55 / **81.0**±0.64	**79.2**±0.44 / **77.9**±0.52
FixMatch (Sohn et al., 2020)	60.1±0.05 / 54.4±0.11	54.0±0.04 / 44.4±0.17	72.9±0.09 / 69.6±0.01	63.4±0.21 / 52.6±0.09
FixMatch + DARP (Kim et al., 2020)	61.1±0.23 / 56.4±0.28	54.9±0.05 / 46.4±0.41	77.8±0.33 / 76.5±0.40	69.9±1.77 / 65.4±3.07
FixMatch + CReST (Wei et al., 2021)	60.8±0.32 / 54.9±0.45	54.5±0.21 / 44.8±0.19	70.5±0.75 / 67.8±0.32	60.5±2.13 / 50.3±2.56
FixMatch + SAW	**62.1**±0.25 / **58.0**±0.20	**55.7**±0.10 / **49.4**±0.40	**78.3**±0.25 / **77.0**±0.19	**71.9**±0.81 / **69.0**±0.81

Experiment

(*easy*) ① **The unlabeled data share similar distribution as the labeled set.**

Table 9. Comparison of bACC on CIFAR10-LT under $\gamma = \gamma_l = \gamma_u$. The hold-out test set is of the **same** distribution of the training set (imbalanced).

	CIFAR10-LT		
Algorithm	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$
FixMatch (Sohn et al., 2020)	$89.2_{\pm 0.59}$	$85.4_{\pm 0.59}$	$83.1_{\pm 0.33}$
FixMatch + DARP (Kim et al., 2020)	$93.1_{\pm 1.95}$	$90.1_{\pm 0.98}$	$88.5_{\pm 0.89}$
FixMatch + CReST (Wei et al., 2021)	$86.7_{\pm 1.98}$	$84.3_{\pm 1.20}$	$82.4_{\pm 1.46}$
FixMatch + SAW	$\mathbf{94.7}_{\pm 1.23}$	$\mathbf{92.4}_{\pm 0.35}$	$\mathbf{90.5}_{\pm 0.47}$

(*difficult*) ② **The distribution of unlabeled data is unknown.**

Table 8. Comparison of bACC on CIFAR10-LT under $\gamma_l = 1$ (labaled set is **balanced**) but $\gamma_u \neq \gamma_l$. The hold-out test set is of the **flipped** distribution of the training unlabeled set.

	CIFAR10-LT		
γ_u	50	100	150
FixMatch (Sohn et al., 2020)	$77.8_{\pm 0.36}$	$66.8_{\pm 1.51}$	$59.9_{\pm 0.43}$
FixMatch + DARP (Kim et al., 2020)	$68.4_{\pm 1.36}$	$55.6_{\pm 3.22}$	$52.1_{\pm 2.07}$
FixMatch + SAW	$\mathbf{82.4}_{\pm 0.49}$	$\mathbf{75.2}_{\pm 1.46}$	$\mathbf{70.1}_{\pm 0.94}$