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

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

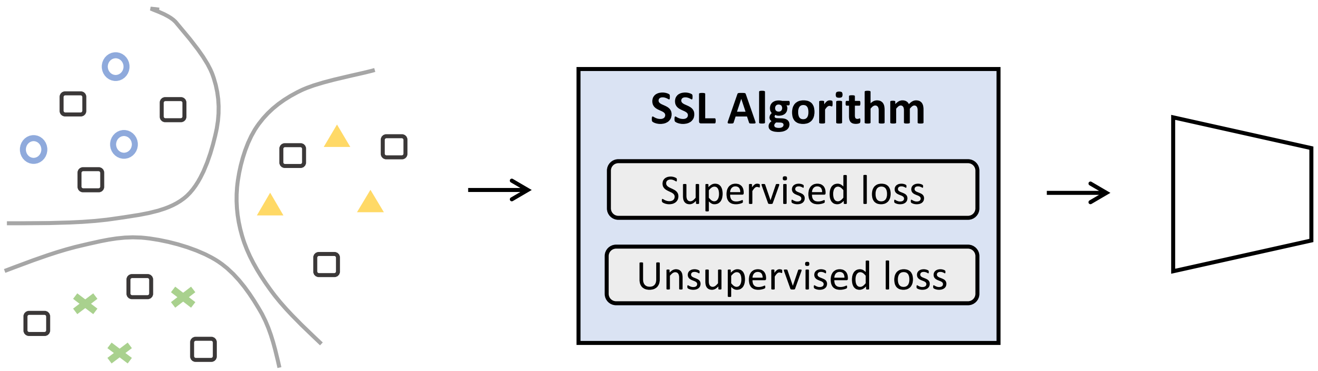
Class-Imbalanced Semi-Supervised Learning with Adaptive Thresholding

Lan-Zhe Guo [1] **Yu-Feng Li** [1]

ICML 2022

Semi-Supervised Learning

- Learning a classifier from labeled and unlabeled examples



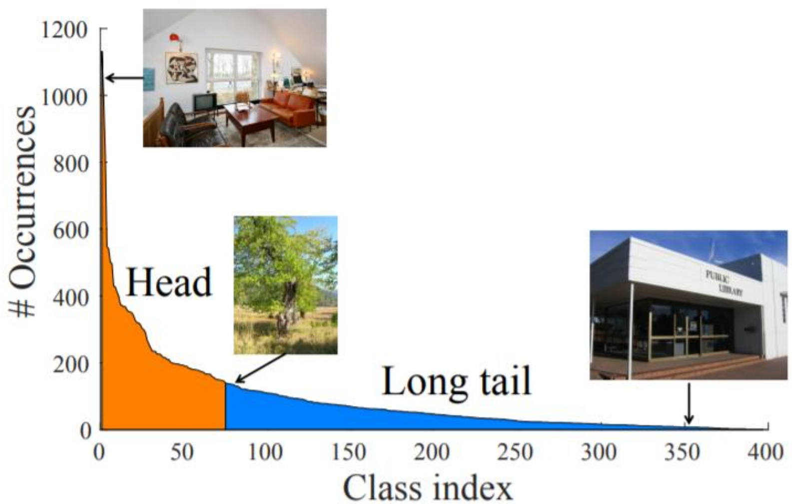
- The most representative method: FixMatch

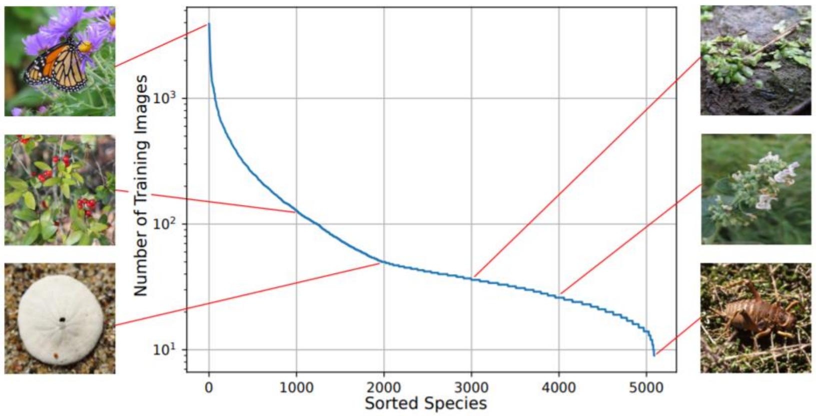
$$\frac{1}{B}\sum_{b=1}^{B}\mathrm{H}(p_b, p_{\mathrm{m}}(y \mid \alpha(x_b))) + \frac{1}{\mu B}\sum_{b=1}^{\mu B}\mathbb{1}(\max(q_b) \geq \tau)\,\mathrm{H}(\hat{q}_b, p_{\mathrm{m}}(y \mid \mathcal{A}(u_b)))$$

- An implicit assumption: the data is **class-balanced**.

Motivation

- However, the real-world is **class-imbalanced**...



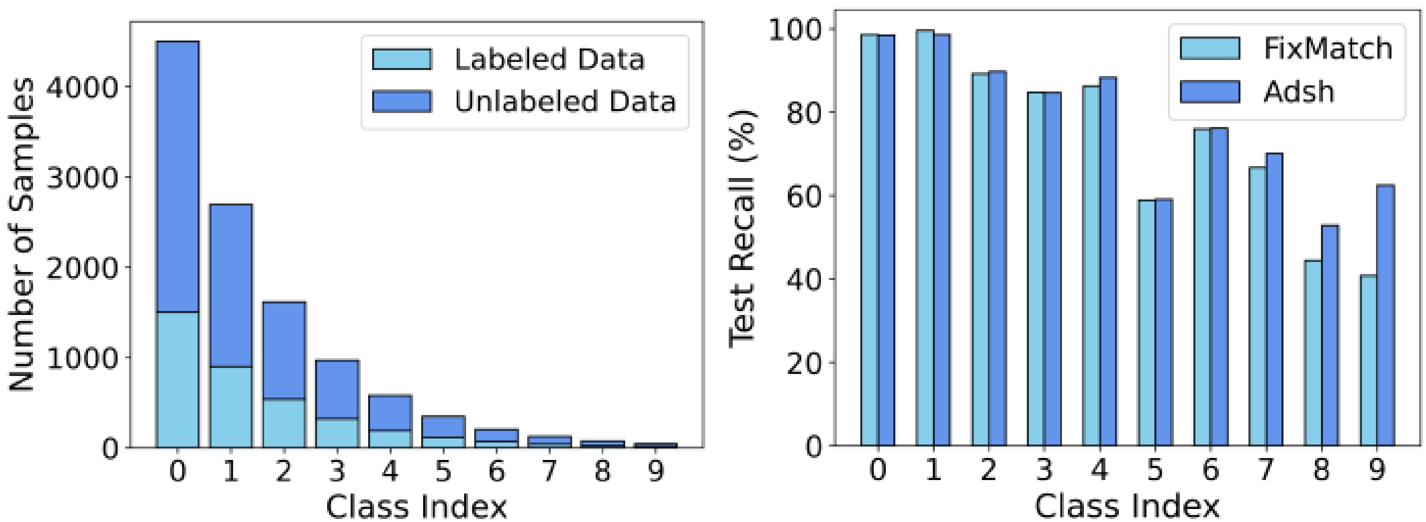


Places [Wang et al. 2017]
There are much more
"dining room" than "library"

Species [Van Horn et al. 2019]
There are much more "dog"
than "panda"

Motivation

- By adopting a **fixed threshold**, the sota SSL method suffers from **overfitting to majority classes**, leading to a **low recall rates on minority classes**.



(a) Imbalanced Dataset (b) Test Recall (%)

- Solution: **class-dependent threshold**
    - Using an adaptive class-dependent threshold to select pseudo labels.

The Proposed Method

- Reformulation of SSL objective function

s_k is used to control how many pseudo labels should be selected for class k.

$$
\begin{aligned}
\min_{\hat{\mathbf{y}},\mathbf{s},\theta} \quad & \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} -y_{i,k}\log f(\mathbf{y}=k|\alpha(\mathbf{x}_i^l);\theta) \\
& + \frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{K}[-\hat{y}_{i,k}\log f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta) \\
& \boxed{-s_k\hat{y}_{i,k}]} \\
\text{s.t.} \quad & \hat{\mathbf{y}}_i = [\hat{y}_{i,1},\cdots,\hat{y}_{i,K}] \in \{0,1\}^K \\
& 0 \le \mathbf{1}^\top \hat{\mathbf{y}}_i \le 1 \\
& s_k > 0, \ \ \forall 1 \le k \le K
\end{aligned}
$$

- It requires to answer the following two questions：
  - **Why** s_k can be regarded a class-dependent threshold ?
  - **How** to find the optimal s_k ?

The Proposed Method

- Adaptive threshold

Given a model f, it satisfies

$$\hat{y}_{i,k} = \begin{cases} 1, & \textbf{\textit{if}}\ k = \operatorname{argmax} \dfrac{f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta)}{\exp(-s_k)} \\ & \dfrac{f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta)}{\exp(-s_k)} \geq 1. \\ 0, & \textbf{\textit{otherwise}}. \end{cases}$$

Lemma 4.2. *If* $\exp(s_k - s_{k'}) > \frac{f(\mathbf{y}=k'|\alpha(\mathbf{x}_i^u);\theta)}{f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta)}$ *holds for all* k *and* k' *that satisfy* $f(\mathbf{y} = k|\alpha(\mathbf{x}_i^u);\theta) > f(\mathbf{y} = k'|\alpha(\mathbf{x}_i^u);\theta)$, *then we have:* $\operatorname{argmax} \frac{f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta)}{\exp(-s_k)} = \operatorname{argmax} f(\mathbf{y} = k|\alpha(\mathbf{x}_i^u);\theta)$.

If $\exp(s_k - s_{k'}) > \frac{f(\mathbf{y}=k'|\alpha(\mathbf{x}_i^u);\theta)}{f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta)}$, the adaptive threshold can be obtained

$$\mathbb{I}(\max(\mathbf{q}_i) \geq \exp(-s_{\hat{\mathbf{y}}_i^u}))$$

The Proposed Method

- Finding the optimal s_k
  - If the ground-truth label distribution is known

    The threshold s_k can be solved has the same class distribution as the ground-truth y^*

$$\sum_{i=1}^{M} \mathbb{I}(f(\mathbf{y}=k|\alpha(\mathbf{x}_i^u);\theta) \geq \exp(-s_k))$$

$$= \frac{\sum_{i=1}^{M} \mathbb{I}(f(\mathbf{y}=1|\alpha(\mathbf{x}_i^u);\theta) \geq \exp(-s_1))}{\gamma_k}$$

$$\gamma_k = \frac{\sum_{i=1}^{M} y_{i,1}^*}{\sum_{i=1}^{M} y_{i,k}^*}$$

  - If the ground-truth class distribution is unknown

    The threshold s_k can be solved to make sure the same percentage of pseudo labels are selected for each class

$$\rho = \frac{\sum_{i=1}^{M} \mathbb{I}(f(\mathbf{y}=1|\alpha(\mathbf{x}_i^u);\theta) \geq \tau_1)}{length(C_1)}$$

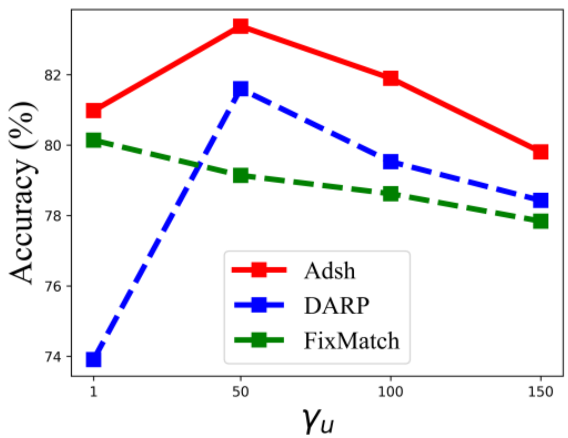
Experiments

Table 1. Comparison of classification performance (Accuracy (%)) on imbalanced CIFAR-10 dataset under three different imbalance ratio: $\gamma = 50, 100, 150$ and two different numbers of labeled data: $N_1 = 1500$, $M_1 = 3000$ and $N_1 = 500$, $M_1 = 4000$. The best results are indicated in bold.

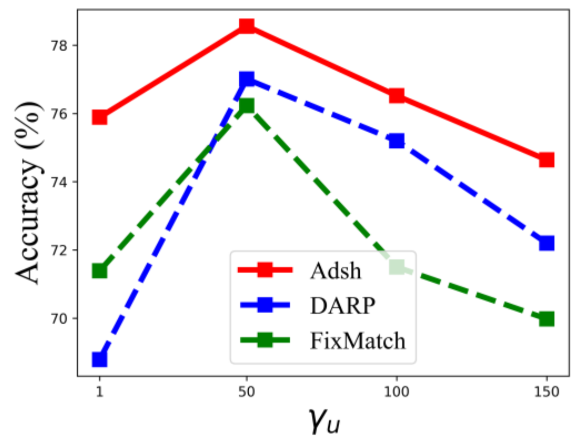
Imbalanced CIFAR-10 Dataset						
	$N_1 = 1500, M_1 = 3000$			$N_1 = 500, M_1 = 4000$		
Algorithm	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$
Supervised	65.23 ± 0.05	58.94 ± 0.13	55.63 ± 0.38	51.31 ± 0.34	45.82 ± 0.41	40.90 ± 0.39
CBL	65.52 ± 0.31	58.52 ± 0.45	52.36 ± 0.58	51.94 ± 0.71	46.22 ± 0.92	41.58 ± 1.24
Re-Sampling	64.53 ± 0.39	56.34 ± 0.42	53.21 ± 0.51	51.96 ± 0.65	48.13 ± 1.25	40.26 ± 1.88
cRT	67.82 ± 0.14	63.43 ± 0.45	59.56 ± 0.44	56.28 ± 1.45	48.11 ± 0.79	45.02 ± 1.08
LDAM	68.91 ± 0.10	63.15 ± 0.24	58.68 ± 0.30	56.41 ± 0.92	49.27 ± 0.88	45.10 ± 0.75
Mean-Teacher	68.84 ± 0.82	61.33 ± 0.28	54.79 ± 0.31	56.34 ± 1.68	48.55 ± 0.77	45.32 ± 1.20
MixMatch	73.59 ± 0.46	65.03 ± 0.26	62.71 ± 0.29	65.32 ± 1.20	56.41 ± 1.96	52.38 ± 1.88
ReMixMatch	78.96 ± 0.29	72.88 ± 0.12	68.61 ± 0.40	76.83 ± 0.98	70.12 ± 1.23	59.58 ± 1.30
FixMatch	79.10 ± 0.14	71.50 ± 0.31	68.47 ± 0.15	77.34 ± 0.96	68.45 ± 0.94	60.10 ± 0.82
DARP	81.60 ± 0.31	75.23 ± 0.14	69.31 ± 0.26	76.72 ± 0.46	69.41 ± 0.50	61.23 ± 0.31
CReST	82.03 ± 0.26	75.08 ± 0.41	69.84 ± 0.39	76.18 ± 0.36	69.50 ± 0.70	60.81 ± 0.55
Adsh	$\mathbf{83.38} \pm \mathbf{0.06}$	$\mathbf{76.52} \pm \mathbf{0.35}$	$\mathbf{71.49} \pm \mathbf{0.30}$	$\mathbf{79.27} \pm \mathbf{0.38}$	$\mathbf{70.97} \pm \mathbf{0.46}$	$\mathbf{62.04} \pm \mathbf{0.51}$

$$\gamma_l = \frac{N_1}{N_K} \qquad \gamma_u = \frac{M_1}{M_K}$$

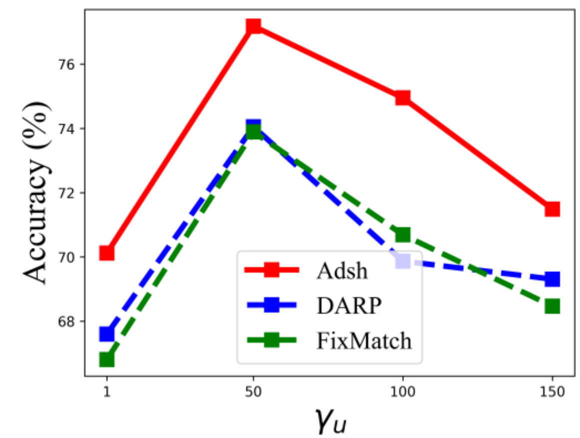
Experiments



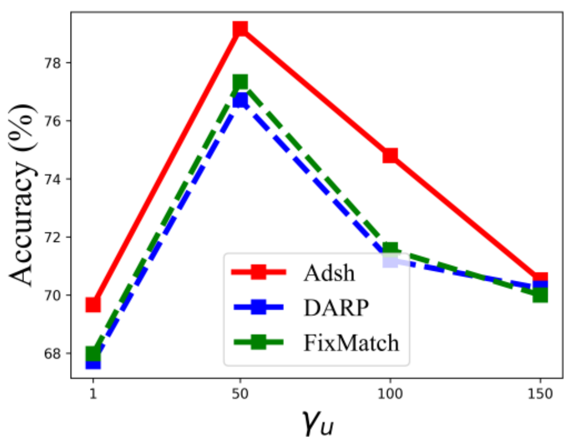
(a) $N_1 = 1500, \gamma_l = 50$



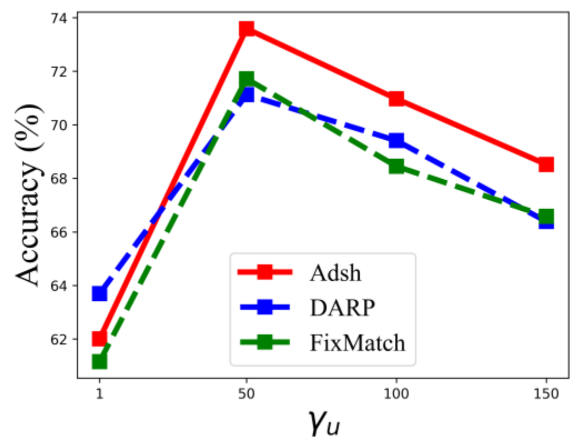
(b) $N_1 = 1500, \gamma_l = 100$



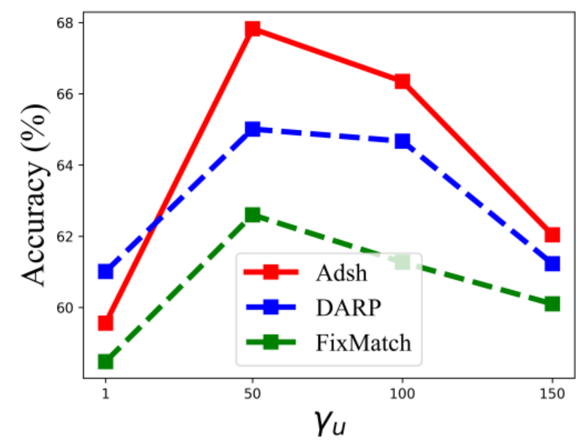
(c) $N_1 = 1500, \gamma_l = 150$



(d) $N_1 = 500, \gamma_l = 50$

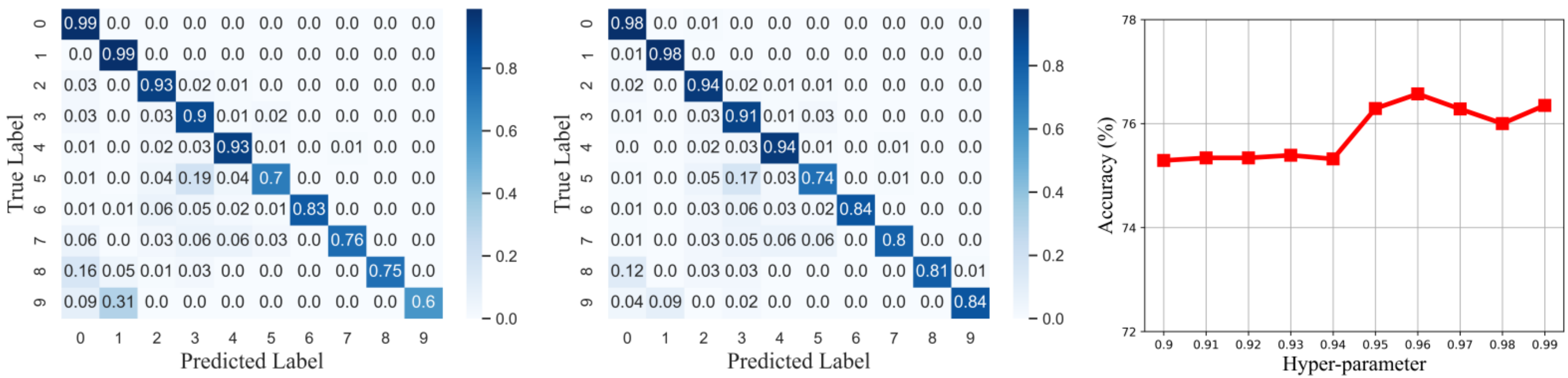


(e) $N_1 = 500, \gamma_l = 100$



(f) $N_1 = 500, \gamma_l = 150$

Experiments



(a) Confusion Matrix of FixMatch (b) Confusion Matrix of Adsh (c) Hyper-Parameter Sensitivity

	SVHN	STL-10	
Algorithm	$\gamma = 100$	$\gamma_l = 10$	$\gamma_l = 20$
ReMixMatch	88.91 ± 0.32	67.43 ± 0.43	60.82 ± 0.93
FixMatch	89.34 ± 0.20	73.25 ± 0.21	63.54 ± 0.21
DARP	90.15 ± 0.46	76.97 ± 0.45	68.87 ± 0.66
CReST	89.90 ± 0.64	76.30 ± 0.38	69.43 ± 0.89
Adsh	$\mathbf{92.13 \pm 0.39}$	$\mathbf{79.25 \pm 0.41}$	$\mathbf{71.03 \pm 0.20}$

THANKS