APLA: CLASS-IMBALANCED SEMI-SUPERVISED LEARNING WITH ADAPTIVE PSEUDO-LABELING AND LOSS ADJUSTMENT

Anonymous authors
Paper under double-blind review

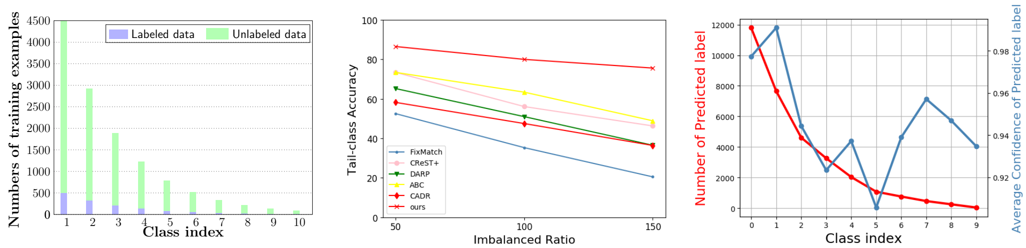
ABSTRACT

Semi-supervised learning (SSL) can substantially improve the performance of deep neural networks by utilizing unlabeled data when labeled data is scarce. Existing SSL algorithms implicitly assume that the class distribution of labeled datasets and unlabeled datasets are balanced, which means the different classes have the same numbers of training samples. However, they can hardly perform well on minority classes(the classes with few training examples) when the class distribution of training data is imbalanced, since the pseudo-labels learned from unlabeled data tend to be biased toward majority classes(the classes with a large number of training examples). To alleviate this issue, we propose a method called Adaptive Pseudo-labeling and Loss Adjustment (APLA) for class-imbalanced semi-supervised learning (CISSL), which includes Class-Aware Pseudo-label Thresholding (CAPT) that can utilize the imbalanced unlabeled data by dynamically adjusting the threshold for selecting pseudo-labels, and Class-Aware Loss Adjustment (CALA) that can mitigate the bias in both supervised loss and unsupervised loss. According to the experiments, APLA can deliver much higher accuracy than benchmark methods under various CISSL scenarios.

1 INTRODUCTION

Semi-supervised learning (SSL)(Chapelle et al., 2006) is a paradigm that can improve learning performance with a few labeled data by using additional unlabeled examples as auxiliaries compared to supervised learning. It provides a way to explore the latent patterns from extra unlabeled examples, alleviating the need for a large number of labels. The state-of-the-art (SOTA) SSL algorithms (Berthelot et al., 2019; 2020; Sohn et al., 2020) often construct a model with a common assumption that the class distribution of the training data is balanced, which means the different classes have the same numbers of training samples. Imbalanced data, however, is widely existing in many realistic scenarios, which leads to the poor performance of SSL algorithms. According to the recent research (Yang & Xu, 2020), the model trained on imbalanced data are easily biased towards majority classes which have a large number of training examples, and far away from minority classes which have few training examples. Although there are several class-imbalanced learning (CIL) algorithms proposed, they are designed for supervised learning and do not exploit unlabeled data, which means that they can not be simply combined with SSL algorithms under class-imbalanced semi-supervised learning (CISSL). There have been a few studies (Yang & Xu, 2020; Kim et al., 2020; Wei et al., 2021; Lee et al., 2021; Hu et al., 2022) on CISSL, but the improvements are mainly at the cost of additional computational overhead, and overfitting on minority-class data or losing information on majority-class data due to the re-sampling of data. Thus, how to efficiently utilize the labeled data and unlabeled data become the core challenge of CISSL.

The main idea of existing CISSL methods can be divided into two categories. One is to improve the quality of pseudo-labels generated by the initial SSL models from SSL perspective. The other is to mitigate the class-imbalanced loss by introducing an auxiliary balanced classifier from CIL perspective. The former relies on that the unlabeled data are more balanced than labeled data, while the latter is not capable of rebalancing the imbalanced bias caused by a large number of pseudo-labels. Due to the imbalanced training data(see Fig. 1(a)), SSL algorithms have to face great challenge to generalize the minority classes which has few training examples. Pseudo-labels,



(a) Long-tailed distribution

(b) Tail-class accuracy

(c) Number and average confidence of predicted pseudo-labels

Figure 1: (a) Class distribution of labeled and unlabeled data on CIFAR-10-LT under the imbalance ratio $\gamma_l = \gamma_u = 50$, where the most majority class has 50× more labeled and unlabeled training samples than the most minority class. (b) Performance of tail-class (the three classes with least training samples) accuracy(%) on CIFAR-10-LT dataset under class imbalance ratio 50, 100, and 150 with 10% of labels available. Our proposed method improves the tail-class accuracy. (c) The number and average confidence of prediction on unlabeled data in SSL algorithm FixMatch (Sohn et al., 2020). Only samples with larger confidence than fixed threshold (0.95) can be imputed.

generated by a model trained on labeled data, are commonly leveraged in SSL algorithms. Although the large number of unlabeled data can help alleviate the degeneration caused by imbalanced data (Yang & Xu, 2020), the pseudo-labels generated by an initial model trained with imbalanced data tend to be biased toward majority classes and deteriorates the model quality. Most SSL methods, i.e., MixMatch, ReMixMatch and FixMatch (Berthelot et al., 2019; 2020; Sohn et al., 2020) have not been evaluated on imbalanced class distribution.

As shown in Fig .1(c), the pseudo-labels of FixMatch tend to have high confidence on majority classes while having low confidence on minority classes. Due to the reason that FixMatch only selects the samples with confidence larger than 0.95, it is obvious that the pseudo-labels are biased toward majority classes and will degenerate the model performance.

According to the discovery that the confidence of pseudo labels varies according to the scenario of different imbalance ratio of labeled data γ_l and that of unlabeled data γ_u, we propose a novel CISSL algorithms called Adaptive Pseudo-labeling and Loss Adjustment (APLA) that can effectively use unlabeled data. Fig.1(b) showcases the effectiveness of APLA in the tail-class. First, we use Class-Aware Pseudo-label Thresholding (CAPT) to re-balance the biased pseudo-labels (Section 4.1). CAPT can efficiently mitigates the imbalanced pseudo-labels by dynamically adjusting the pseudo-label threshold, higher for majority classes and lower for minority classes. Furthermore, even if we can get unlabeled data with balanced pseudo-labels, the imbalance still exists in labeled data. Thus, we compensate for the imbalanced loss of minority classes by introducing Class-Aware Loss Adjustment (CALA), adjusting both supervised loss and unsupervised loss for per class (Section 4.2). To further reduce the training cost, we incorporate CAPT and CALA with an extra auxiliary balanced classifier followed by the success of Lee et al. (2021) (Section 4.3). Experiments demonstrate that our work can deliver much higher accuracy than benchmark methods under various CISSL scenarios. Our main contribution are summarized as follows:

- We proposed a new CISSL method called Adaptive Pseudo-labeling and Loss Adjustment (APLA). It significantly improves the precision of pseudo-labels by Class-Aware Pseudo-label Thresholding (CAPT), and remove the bias from both the supervised loss and unsupervised loss by Class-Aware Loss Adjustment (CALA) .
- CAPT adapts the threshold for selecting pseudo-labels according to the class distribution in CISSL.
- CALA adjusts the loss according to the distribution of labeled data and unlabeled data, avoiding the learning performance degeneration when unlabeled data is much more balanced than labeled data.

2 RELATED WORK

This work is related to class-imbalanced learning and semi-supervised learning.

2.1 CLASS-IMBALANCED LEARNING

Datasets in real-world often exhibits class-imbalanced label distribution (Van Horn et al., 2018) and make the standard learning models hard to generalize. Several Class-imbalanced learning(CIL) methods (Cui et al., 2019; Huang et al., 2020; Liu et al., 2019; Cao et al., 2019; Ren et al., 2020; Zhou et al., 2020; Menon et al., 2021) have been proposed to address this problem. Most of them (Cao et al., 2019; Cui et al., 2019; Huang et al., 2020; Liu et al., 2019) handle with quantity imbalance, where the distribution of training examples from different classes is imbalanced, such as the long-tailed distribution (Van Horn et al., 2018; Gupta et al., 2019) and step imbalanced distribution (Buda et al., 2018). The solution can be categorized as re-sampling (He & Garcia, 2009; Pouyanfar et al., 2018; Xu et al., 2021), re-weighting (Buda et al., 2018; Byrd & Lipton, 2019; Cui et al., 2019; Park et al., 2021; Huang et al., 2020; Cao et al., 2019; Menon et al., 2021), synthetic samples (Chou et al., 2020; Chawla et al., 2002), meta learning (Ren et al., 2020; Shu et al., 2019), transfer learning (Liu et al., 2019; Yin et al., 2019; Jamal et al., 2020) and decoupling representation and classifier (Zhou et al., 2020; Kang et al., 2020; Zhong et al., 2021). They are designed for supervised learning and do not exploit unlabeled data. Compared with these supervised methods, our method mitigate the imbalance in labeled data by utilizing extra unlabeled data.

2.2 SEMI-SUPERVISED LEARNING

Several SSL methods have been proposed to utilizes unlabeled data in recent years. Entropy minimization (Grandvalet & Bengio, 2005) is designed to make the prediction made by training models have high confidence, and prevent the class distribution of predictive results from being too flat and having no tendency. Consistency regularization (Sajjadi et al., 2017; Laine & Aila, 2017; Tarvainen & Valpola, 2017; Miyato et al., 2019; Xie et al., 2020) are designed to make the predictive results to have consistency under various disturbances, which improves the generalization of SSL algorithms by using extra unlabeled data. Holistic methods (Berthelot et al., 2019; 2020; Sohn et al., 2020; Zhang et al., 2021) gain large accuracy improvement in SSL algorithms by combining the entropy minimization and the consistency regularization. But these methods use a fixed threshold to select pseudo-labels and highly rely on the quality of pseudo-labels. When training data is class-imbalanced, these methods lead to performance degradation as the model produces low quality pseudo-labels. Our method uses CAPT to generate pseudo-labels with high accuracy by selecting pseudo-labels with a dynamical threshold.

2.3 CLASS-IMBALANCED SEMI-SUPERVISED LEARNING

There have been a few studies on class-imbalanced semi-supervised learning (CISSL) (Yang & Xu, 2020; Kim et al., 2020; Wei et al., 2021; Lee et al., 2021; Hu et al., 2022). The representatives are DARP (Kim et al., 2020), CReST (Wei et al., 2021),ABC (Lee et al., 2021) and CADR (Hu et al., 2022). DARP refines biased pseudo-labels via a convex optimization. CReST, a self-training (Rosenberg et al., 2005) technique, alleviates class imbalance by using pseudo-labeled unlabeled data points classified as minority classes with a higher precision than those classified as majority classes. These two pseudo-label based algorithms rely on getting a lot of pseudo-labels with high confidence to mitigates the imbalance, which makes them inappropriate when unlabeled data is more imbalanced than labeled data. ABC, a simple auxiliary balanced classifier, mitigates the bias of the learning model caused by class imbalance using extra regularization term. But, ABC only assumes that the class distributions between labeled data and unlabeled data are the same, which is difficult to verify since we have no idea about the true class distribution of unlabeled data in real-world scenarios. CADR dynamically decreases the pseudo-label thresholding for minority classes and removes the bias from the supervised model training by using Class-Aware Propensity (CAP). But the assumption of CADR is that the distribution of unlabeled data is balanced, which is too restrictive to be adapted to a variety of class-imbalanced scenarios. Our method mitigates the imbalance in both CIL perspective and SSL perspective, generating pseudo-labels with high quality and training a more balanced classifier.

3 Preliminary and Background

This section provides notations used in this paper and gives a brief review of SSL algorithms with a fixed threshold.

3.1 Problem Setting

The training data of the CISSL task consists of n labeled examples $\mathcal{D}_l = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ and m unlabeled examples $\mathcal{D}_u = \{x_{n+1}, \cdots, x_{n+m}\}$. Generally, $m >> n, x \in \mathcal{X} \in \mathbb{R}^D, y \in \mathcal{Y} = \{1, \cdots, C\}$ where D is the number of input dimension and C is the number of output classes in training examples. We denote the number of data points in class C under $\mathcal{D}_l$ and $\mathcal{D}_u$ as n_c and m_c, respectively, i.e., $\sum_{c=1}^{C} n_c = n$ and $\sum_{c=1}^{C} m_c = m$. The ratio of the class imbalance of $\mathcal{D}_l$ and $\mathcal{D}_u$ are denoted as $\gamma_l = \frac{n_1}{n_c}$ and $\gamma_u = \frac{m_1}{m_c}$. In class-imbalanced scenarios, $\gamma_l \gg 1, \gamma_u \geq 1$. In general, we assume that $\mathcal{D}_l$ and $\mathcal{D}_u$ have the same distribution, i.e., $\gamma_l = \gamma_u$. But there are some cases where $\mathcal{D}_l$ and $\mathcal{D}_u$ have different distributions, i.e., $\gamma_l \neq \gamma_u$. The aim of the CISSL algorithms is to find an appropriate learning model $f(x; \theta) : \{\mathcal{X}; \Theta\} \rightarrow \mathcal{Y}$ parameterized by $\theta \in \Theta$ from imbalanced training data to mitigate the generalization risk.

The training loss of an SSL algorithm usually consists of supervised loss $\mathcal{L}_s$ and unsupervised loss $\mathcal{L}_u$ with a weight parameter λ,i.e., $\mathcal{L}_s + \lambda\mathcal{L}_u$. $\mathcal{L}_s$ is calculated by $\mathcal{D}_l$ and $\mathcal{L}_u$ is calculated by $\mathcal{D}_l$. Typically, $\mathcal{L}_s$ can be written as follows:

$$\mathcal{L}_s = \frac{1}{|\mathcal{D}_l|} \sum_{x \in \mathcal{D}_l} H(y, f(x; \theta)) \tag{1}$$

here H is the function of cross entropy loss and $f(x; \theta)$ is the predicted probabilities of input x produced by the model f.

Different choices for the unsupervised loss $\mathcal{L}_u$ lead to different semi-supervised learning algorithms. Typically, there are two ways to construct unsupervised loss. One is designed to optimize a regularization that does not depend on labels, such as consistency regularization (Miyato et al., 2019; Xie et al., 2020). The other one is designed to generate pseudo-labels with high confidence to formulate a loss with unlabeled data (Lee, 2013; Berthelot et al., 2019; 2020; Sohn et al., 2020). Next, we will introduce a recent SSL work (Sohn et al., 2020) to illustrate how to generate pseudo-labels and construct unsupervised loss $\mathcal{L}_u$.

3.2 FixMatch: An SSL algorithm with Fixed Pseudo-label Threshold

Due to its great success in SSL field, we choose FixMatch as the backbone to make sure that our proposed method can utilize the high-quality representations learned by the backbone. In order to better illustrate our proposed approach, we briefly introduce FixMatch in this section.

FixMatch (Sohn et al., 2020) combines consistency regularization and pseudo-labeling with a simple framework as well as using weak and strong augmentation for consistency regularization separately. For supervised loss $\mathcal{L}_s$, FixMatch computes standard cross-entropy loss on a weakly augmented version of labeled examples $A_w(x)$ from the labeled set $\mathcal{D}_l$. For unsupervised loss $\mathcal{L}_u$, FixMatch first generates pseudo-labels using the model's predictions on weakly augmented unlabeled examples. The predicted label is only retained as pseudo-label if the highest class probability of model prediction is greater than the threshold $\tau = 0.95$. The model is then trained to predict the pseudo-label when fed with the strongly augmented version $A_s(x)$. The total loss can be written as follows:

$$\mathcal{L}_s = \frac{1}{|\mathcal{D}_l|} \sum_{x \in \mathcal{D}_l} H(y, f(A_w(x); \theta)) \tag{2}$$

$$\mathcal{L}_u = \frac{1}{|\mathcal{D}_u|} \sum_{x \in \mathcal{D}_u} 1(\max(f(A_w(x); \theta) > \tau)) H(f(A_w(x); \theta), f(A_s(x); \theta)) \tag{3}$$

$$\mathcal{L} = \mathcal{L}_s + \lambda\mathcal{L}_u \tag{4}$$

As discussed in Section 1, the traditional SSL model trained with class-imbalanced data is biased toward majority classes and far from minority classes. Unlabeled data predicted as majority classes tend to have high confidence, and data predicted as minority classes tend to have low confidence. Traditional SSL algorithms like FixMatch using a fixed threshold τ to select imputation samples, which aims to select correct pseudo labels and discard noise ones. Samples predicted as minority class tend to be eliminated while samples predicted as majority classes tend to be selected in Fig. 3 (Shown in Appendix B). According to Wei et al. (2021), the discarded samples that are predicted as minority classes still hold high precision, while the retained samples that are predicted as majority classes have low precision and many of them may have wrong pseudo-labels, which leads to the bias in the calculation of $\mathcal{L}_u$.

The standard cross entropy loss formulates as:

$$H(y, f(x;\theta)) = -\log \frac{f_y(x;\theta)}{\sum_{c=1}^{C} f_c(x;\theta)} \tag{5}$$

As mentioned in Section 1, the learning classifier using standard cross entropy loss would be biased towards the majority classes due to the reason that the distribution of training data is heavily class-imbalanced. FixMatch suffers it in both the supervised and unsupervised loss terms, and causes the degradation of learning performance with class-imbalanced training data. We propose Adaptive Pseudo-labeling and Loss Adjustment (APLA) framework to mitigate the bias in both supervised loss and unsupervised loss in the next section.

4 Adaptive Pseudo-labeling and Loss Adjustment Framework

The APLA framework can be decomposed into two parts: Class-Aware Pseudo-label Thresholding (CAPT) for unsupervised loss and Class-aware Loss Adjustment (CALA) for both supervised loss and unsupervised loss.

4.1 Class-Aware Pseudo-label Thresholding

To alleviate the imbalance of prediction and generate more correct pseudo-label, Class-Aware Pseudo-label Thresholding (CAPT) utilizes the imbalanced unlabeled data by dynamical adjusting the pseudo-label threshold. Let $q_{m,t}$ be the model's predicted class distribution given a weakly-augmented version of a unlabeled image m at time step t, $\sigma_t(c)$ denotes the selected pseudo-label of class c at time step t. We get a class-aware pseudo-label threshold $\mathcal{T}_t(c)$ for class c as:

$$\sigma_t(c) = \sum_{m=1}^{M} 1(\max(q_{m,t}) > \tau_c) \cdot 1(\arg\max(q_{m,t}) = c) \tag{6}$$

$$temp_c = \frac{\sigma_t(c)}{\max\{\max_c \sigma_t, M - \sum_c \sigma_t\}} \cdot \gamma_c \tag{7}$$

$$\mathcal{T}_t(c) = \mathcal{M}(temp_c) \cdot \tau. \tag{8}$$

where γ_c is a hyperparameter for class c and $\mathcal{M}$ is a non-linear mapping function. According to the Zhang et al. (2021), we set $\mathcal{M}(x) = \frac{x}{2-x}$, making the samples predicted as minority classes have lower pseudo-label threshold. We replace the fixed threshold τ in FixMatch with $\mathcal{T}_t(c)$ for class c. Generally, our CAPT sets a higher threshold for majority classes and a lower threshold for minority classes, encouraging the model using more unlabeled data with correct pseudo-labels that are predicted as minority classes. By using CAPT, we rebalance the biased pseudo-labels and mitigate the imbalance labels in training data.

4.2 Class-Aware Loss Adjustment

Although we use CAPT to rebalance the biased pseudo-labels, it still exists imbalance in both labeled data and unlabeled data. Similar to (Ren et al., 2020; Menon et al., 2021), we choose Class-Aware Loss Adjustment (CALA) to replace the standard cross entropy loss:

$$H_{CALA}(y, f(x;\theta)) = -\log \frac{f_y(x;\theta) + \beta \log \pi_y}{\sum_{c=1}^{C} f_c(x;\theta) + \beta \log \pi_c} \tag{9}$$

where π_y is the estimate of class prior $\mathbb{P}(y)$ and $\beta > 0$ is a tuning parameter to be chosen based on holdout calibration. If the class distribution is balanced, $\pi_c = 1, c \in \{1, \cdots, C\}$ and $H_{CALA}(y, f(x;\theta)) = H(y, f(x;\theta))$. Using this loss encourages large margins between the true label and other negative labels.

Due to the reason that $\mathcal{D}_l$ and $\mathcal{D}_u$ may not have the same class distribution, π for supervised loss π^l and unsupervised loss π^u are computed separately. Due to the reason that only the distribution of labeled data is already determined, we set $\hat{P}(Y_u)$ as the estimated class distribution of unlabeled data at time step t and $P(Y_l)$ as the class distribution of labeled data. $P(Y_l)$ can be computed by the class imbalance γ_l, and $\hat{P}_t(Y_u)$ is computed by the prediction of the training model, which means the predicted distribution of unlabeled data at time step t. Then π^l and π_t^u can be computed as follows:

$$\pi^l = P(Y_l), \quad \pi_t^u = \hat{P}_t(Y_u) \tag{10}$$

$$p_t(c) = \sum_{m=1}^{M} 1(\arg\max(q_{m,t}) = c) \tag{11}$$

$$\hat{P}_t(y_u = c) = \frac{p_t(c)}{p_t(1)}, \quad P(y_l = c) = \frac{n_c}{n_1} \tag{12}$$

where $p_t(c)$ reflects the number of unlabeled data to be predicted as class c at time step t. We encourage the model to generate more balanced prediction on unlabeled data and compensate the bias caused by the imbalance of labeled and unlabeled data by using CALA.

4.3 Extra Auxiliary Balanced Classifier

To further reduce the training cost, we incorporate CAPT and CALA with an extra auxiliary balanced classifier inspired by the success of Zhou et al. (2020); Lee et al. (2021). As the work in Lee et al. (2021), the extra auxiliary balanced classifier is trained simultaneously with the backbone algorithms, so it can share high-quality representations learned from all data points with the backbone algorithms. Note that the classification loss for the extra auxiliary balanced classifier is also adjusted by CAPT and CALA. We train the proposed algorithm with loss for the extra auxiliary balanced classifier $\mathcal{L}_{eabc}$ and the loss for the backbone $\mathcal{L}_{back}$. The total loss function $\mathcal{L}_{sum}$ is expressed as:

$$\mathcal{L}_s = \frac{1}{|\mathcal{D}_l|} \sum_{x \in \mathcal{D}_l} H_{CALA}(y, f(A_w(x);\theta)) \tag{13}$$

$$\mathcal{L}_u = \frac{1}{|\mathcal{D}_u|} \sum_{x \in \mathcal{D}_u} 1(\max(f(A_w(x);\theta) > \mathcal{T}_t(c)))H_{CALA}(f(A_w(x);\theta), f(A_s(x);\theta)) \tag{14}$$

$$\mathcal{L}_{back} = \mathcal{L}_s + \lambda \mathcal{L}_u \tag{15}$$

$$\mathcal{L}_{sum} = \mathcal{L}_{eabc} + \mathcal{L}_{back} \tag{16}$$

5 Experiments

In this section, we evaluate various algorithms including SSL, CIL and CISSL under various scenarios for class-imbalanced classification problems. We first provide description of our experimental setups in Section 5.1. We then give empirical evaluations on our proposal and other compared methods under various scenarios in Section 5.2. Finally, we present ablation study to help understand the superiority of our proposal in Section 5.3.

5.1 Experimental setup

We choose CIFAR-10 (Krizhevsky, 2009) and SHVN (Netzer et al., 2011) as the basic datasets to create various class-imbalanced datasets with the class-imbalance ratio of labeled data γ_l and the class-imbalanced ratio of unlabeled data γ_u. There are two types of class imbalance, the long-tailed imbalance where the number of data points exponential decline from the largest class to the smallest class, i.e. $n_k = n_1 * \gamma^{\frac{1-k}{L-1}}$, and the step imbalance (Buda et al., 2018) where the majority classes have same number of data points and the minority classes also have the same number of data points. Two types of class imbalance for these datasets are illustrated in Appendix B. We choose $n_1 = 500, m_1 = 4500$ for CIFAR-10-LT and $n_1 = 1000, m_1 = 4000$ for CIFAR-10-Step and SHVN-Step.

Our method is compared with the performance of these algorithms, including:

- WRN-28-2 (Zagoruyko & Komodakis, 2016) (Vanilla algorithm): The basic Deep CNN is trained on only labeled data with the simple cross-entropy loss.
- MiSLAS (Zhong et al., 2021) (CIL algorithm): The SOTA CIL algorithm uses MixUp (Zhang et al., 2019) and label-aware smoothing to handle different degrees of overconfidence for classes and reduce dataset bias by shift learning on the batch normalization layer in the decoupling framework, without using extra unlabeled data.
- MixMatch (Berthelot et al., 2019), FixMatch (Sohn et al., 2020) (SSL algorithms): The SOTA SSL algorithms combined consistency regularization and pseudo-labels have achieved great success in SSL, without taking class imbalance into account.
- DARP (Kim et al., 2020) (CISSL algorithm): The algorithm uses DARP to refine the pseudo-labels obtained by SSL algorithms, e.g. FixMatch .
- CReST (Wei et al., 2021) (CISSL algorithm): The algorithm alleviates the class imbalance by selecting pseudo-labeled unlabeled instance classified as minority classes with a higher confidence than those classified as majority classes.
- ABC (Lee et al., 2021) (CISSL algorithm): The algorithm provides auxiliary balanced classifier to rebalance the biased model by introducing extra regularization terms.
- CADR (Hu et al., 2022)(CISSL algorithm): The algorithm removes the bias from both the supervised model training end—by using Class-Aware Prospesity (CAP), and the unlabeled data imputation end—by using Class-Aware Imputation (CAI).

All experiments are trained with batch size 64 for $250,000$ iterations. We use the Adam optimizer (Kingma & Ba, 2015) with a learning rate of 0.002, and use Cutout (Devries & Taylor, 2017) and RandomAugment (Cubuk et al., 2019) for strong data augmentation, following the approach provided in Lee et al. (2021). As suggested by Berthelot et al. (2019), we evaluate the performance of these algorithms using an exponential moving average of the parameters over iterations with a decay rate of 0.999, instead of scheduling the learning rate. In Tables 1-2, we use the overall accuracy and the accuracy only for tail class as performance measures in long-tailed setting. In Table 3 we use the the overall accuracy and the accuracy only for minority classes as performance measures in step-imbalanced setting. Each experiment is repeated five times with the long-tailed imbalance setting and three times with the step-imbalance setting. We report the average and standard deviation of the performance measures.

5.2 Results

5.2.1 CIFAR-10-LT under $\gamma_l = \gamma_u$

We first evaluate the algorithms with $\gamma_l = \gamma_u$, which is the most common scenarios that labeled and unlabeled data are sampled from the same distribution. In order to produce convincing results, we compare our work with the existing SSL, CIL and CISSL algorithms on CIFAR-10-LT with various imbalance ratio. From the results in Table 1, we can observe that in most cases SSL methods perform better than class-imbalanced learning method by using extra unlabeled data. The other CISSL methods achieve good performance among compared methods since they consider both unlabeled data and imbalanced distribution. It is noticeable that our proposal consistently achieves the best performance in all settings with various imbalance ratios of training examples. For example, our APLA preforms 11.5% better in overall accuracy and 43.1% better in tail-class accuracy than FixMatch upon CIFAR-10-LT($\gamma_l = \gamma_u = 100$). The improvement of APLA is more significant as the imbalanced ratio increases.

Table 1: Overall accuracy/ tail-class (the three classes with least training samples) accuracy with the long-tailed imbalanced setting. SSL denotes semi-supervised learning and CIL denotes class-imbalanced learning.

			CIFAR-10-LT($\gamma_l = \gamma_u$)		
Algorithm	SSL	CIL	$\gamma_l = 50$	$\gamma_l = 100$	$\gamma_l = 150$
Vanilla	-	-	$49.3_{\pm 1.68}$ / $23.0_{\pm 3.44}$	$44.2_{\pm 0.37}$ / $10.3_{\pm 2.22}$	$40.4_{\pm 1.10}$ / $5.1_{\pm 1.83}$
MiSLAS	-	✓	$60.0_{\pm 0.38}$ / $45.1_{\pm 2.79}$	$53.0_{\pm 0.11}$ / $28.4_{\pm 1.39}$	$48.6_{\pm 0.86}$ / $20.5_{\pm 2.15}$
MixMatch	✓	-	$62.5_{\pm 1.46}$ / $22.1_{\pm 3.80}$	$56.7_{\pm 1.05}$ / $7.6_{\pm 2.56}$	$52.2_{\pm 2.07}$ / $7.9_{\pm 3.21}$
FixMatch	✓	-	$76.0_{\pm 0.96}$ / $52.5_{\pm 3.67}$	$68.7_{\pm 0.70}$ / $35.3_{\pm 2.58}$	$63.2_{\pm 0.32}$ / $20.5_{\pm 0.20}$
w/ CReST+	✓	-	$81.0_{\pm 0.51}$ / $73.4_{\pm 1.47}$	$74.5_{\pm 0.61}$ / $56.1_{\pm 1.56}$	$72.3_{\pm 0.70}$ / $46.3_{\pm 2.52}$
w/ DARP	✓	-	$79.9_{\pm 0.12}$ / $65.2_{\pm 0.59}$	$73.9_{\pm 0.96}$ / $51.0_{\pm 1.98}$	$68.4_{\pm 0.23}$ / $36.5_{\pm 1.13}$
w/ ABC	✓	-	$82.4_{\pm 0.51}$ / $73.4_{\pm 2.05}$	$77.2_{\pm 0.54}$ / $63.4_{\pm 1.87}$	$72.4_{\pm 0.81}$ / $48.9_{\pm 3.65}$
W/ CADR	✓	-	$76.8_{\pm 0.51}$ / $58.2_{\pm 1.33}$	$70.6_{\pm 0.71}$ / $47.5_{\pm 1.62}$	$66.4_{\pm 0.73}$ / $36.4_{\pm 2.17}$
w/ Ours	✓	-	**82.8**$_{\pm 0.46}$ / **84.4**$_{\pm 2.18}$	**80.2**$_{\pm 0.61}$ / **78.4**$_{\pm 1.12}$	**75.5**$_{\pm 1.25}$ / **72.2**$_{\pm 3.75}$

5.2.2 CIFAR-10-LT UNDER $\gamma_l \neq \gamma_u$

We then evaluate the algorithms with $\gamma_l \neq \gamma_u$, which is not unusual in realistic scenarios where labeled and unlabeled data are sampled from the different distribution. In this case, it is also hard to know the real distribution of unlabeled data. So, for the training model, the imbalance ratio γ_u of unlabeled data is an unknown parameter. Generally, the performance is related to the class-imbalanced ratio of all training data. The accuracy of the SSL model should increase as the γ_u decreases, which means the overall distribution of training data becomes more balanced. But in Table 2, an interesting observation is that for a fixed γ_l, MixMatch, FixMatch, FixMatch+CReST+ and FixMatch+ABC suffer a great performance degradation when $\gamma_u = 1$, which means the most balanced unlabeled dataset. This is because that these algorithms does not take the distribution of unlabeled data into consideration. As shown in Fig 3 (Shown in Appendix A), when unlabeled data is balanced, these algorithms easily mistake a large number of unlabeled data from tail class as data from head class, leading the performance degradation of tail class. The results in Table 2 also show that our method performs better than FixMatch and other CISSL algorithms on all settings. Especially in CIFAR-10-LT ($\gamma_l = 100, \gamma_u = 1$), our APLA is 19.7% better in overall accuracy and 60.1% better in tail class accuracy than FixMatch, and 10.7% better in overall accuracy and 28.7% better in tail class accuracy than FixMatch+ABC.

Table 2: Overall accuracy/tail-class (the three classes with smallest training samples) accuracy under the long-tailed setting($\gamma_l \neq \gamma_u$). SSL denotes semi-supervised learning and CIL denotes class-imbalanced learning.

			CIFAR-10-LT($\gamma_l = 100$)		
Algorithm	SSL	CIL	$\gamma_u = 1$	$\gamma_u = 50$	$\gamma_u = 150$
Vanilla	-	-	$44.2_{\pm 0.37}$ / $10.3_{\pm 2.22}$	$44.2_{\pm 0.37}$ / $10.3_{\pm 2.22}$	$44.2_{\pm 0.37}$ / $10.3_{\pm 2.22}$
MiSLAS	-	✓	$53.0_{\pm 0.11}$ / $28.4_{\pm 1.39}$	$53.0_{\pm 0.11}$ / $28.4_{\pm 1.39}$	$53.0_{\pm 0.11}$ / $28.4_{\pm 1.39}$
MixMatch	✓	-	$36.7_{\pm 0.56}$ / $1.0_{\pm 0.55}$	$56.6_{\pm 0.52}$ / $13.1_{\pm 2.92}$	$56.2_{\pm 1.35}$ / $11.8_{\pm 3.70}$
FixMatch	✓	-	$65.7_{\pm 0.52}$ / $23.1_{\pm 0.24}$	$71.8_{\pm 1.12}$ / $41.2_{\pm 3.42}$	$67.7_{\pm 0.77}$ / $33.3_{\pm 2.62}$
w/ CReST+	✓	-	$76.1_{\pm 1.62}$ / $62.1_{\pm 3.01}$	$79.4_{\pm 1.48}$ / $68.6_{\pm 0.95}$	$72.1_{\pm 2.36}$ / $46.2_{\pm 4.37}$
w/ DARP	✓	-	$76.7_{\pm 0.13}$ / $65.5_{\pm 0.41}$	$74.3_{\pm 0.29}$ / $63.4_{\pm 0.39}$	$71.1_{\pm 0.13}$ / $48.1_{\pm 0.39}$
w/ ABC	✓	-	$74.7_{\pm 0.75}$ / $54.5_{\pm 2.52}$	$79.2_{\pm 0.46}$ / $65.3_{\pm 1.92}$	$74.7_{\pm 0.27}$ / $65.1_{\pm 1.77}$
w/ CADR	✓	-	$82.9_{\pm 1.57}$ / $71.3_{\pm 1.36}$	$76.0_{\pm 1.08}$ / $58.4_{\pm 0.95}$	$68.9_{\pm 1.61}$ / $42.7_{\pm 1.69}$
w/ Ours	✓	-	**85.4**$_{\pm 0.94}$ / **83.2**$_{\pm 2.94}$	**81.8**$_{\pm 0.47}$ / **81.8**$_{\pm 1.56}$	**76.9**$_{\pm 0.60}$ / **73.3**$_{\pm 2.47}$

5.2.3 CIFAR-10-STEP AND SVHN-STEP UNDER $\gamma_l = \gamma_u$

We also evaluate the algorithms with the step imbalance setting, where the half of the classes have few training data. This setting assumes a more severely imbalanced class distribution than the long-tailed imbalance settings. We omit CADR since it can not be applied to this situation. The experiment results of CIFAR-10-Step and SHVN-Step in Table 3 show that APLA outperforms other algorithms in minority class accuracy, and the overall accuracy of APLA is a slightly lower than ABC in SVHN-Step. This may be the reason that SVHN is an simple dataset and the adaptive threshold may easily degenerate the majority classes accuracy in SVHN-Step. The more quantitative comparison can be seen in Appendix C.

Table 3: Overall accuracy/Minority-class accuracy on CIFAR-10 and SVHN under step imbalanced setting.

	CIFAR-10-Step	SVHN-Step
Algorithm	$\gamma_l = \gamma_u = 100$	$\gamma_l = \gamma_u = 100$
FixMatch	$54.0_{\pm 0.84}$ / $11.8_{\pm 1.71}$	$79.8_{\pm 1.34}$ / $61.5_{\pm 2.76}$
w/ CReST+	$71.1_{\pm 0.78}$ / $48.2_{\pm 2.26}$	$86.6_{\pm 0.19}$ / $76.3_{\pm 0.23}$
w/ DARP	$67.9_{\pm 1.98}$ / $43.0_{\pm 2.12}$	$85.3_{\pm 0.19}$ / $67.9_{\pm 0.40}$
w/ ABC	$75.9_{\pm 0.49}$ / $57.0_{\pm 1.07}$	$\mathbf{90.6}_{\pm 0.17}$ / $85.6_{\pm 0.35}$
w/Ours	$\mathbf{76.8}_{\pm 0.65}$ / $\mathbf{74.9}_{\pm 1.42}$	$88.3_{\pm 0.74}$ / $\mathbf{87.4}_{\pm 0.70}$

5.3 ABLATION STUDY

We also conduct an ablation study on CIFAR-10-LT in the main setting to investigate the effect of each component of the proposed algorithm. The results for APLA are presented in Table 4, where each row indicates the proposed algorithm with the described conditions in that row.

Table 4: Ablation study for APLA on CIFAR-10-LT, $\gamma_l = \gamma_u = 100$

Ablation study	Overall	Tail-class
APLA (proposed algorithm)	**80.2**	**80.0**
Without CAPT	78.5	76.3
Without CALA	76.2	60.2
Without extra auxiliary classifier	74.2	54.9

6 CONCLUSION

We introduced the Adaptive Pseudo-labeling and Loss Adjustment (APLA), a simple but effective framework, which is attached to a state-of-the-art SSL algorithm, for CISSL. First, we proposed Class-Aware Pseudo-label Thresholding (CAPT) to generate correct pseudo-labels by dynamically adjust the pseudo-label threshold for different classes. We effectively mitigated the imbalance of the pseudo-labels by CAPT. Then, we proposed Class-Aware Loss Adjustment (CALA) to compensate the huge bias caused by imbalanced training data in both supervised loss and unsupervised loss. Finally, we combined CAPT and CALA by using extra auxiliary balanced classifier. The experimental results obtained under various setting demonstrate that our proposed method outperforms other CISSL algorithms. We also conducted a qualitative analysis and an ablation study to verify the contribution of each component of the proposed algorithm. How to full utilize unlabeled data with SSL models in realistic scenarios has attracted great attention in recent years. CISSL is a representative problem that using unlabeled data with huge imbalance. Our work proposed a novel scheme for the problem. One limitation of our scheme is that APLA may degrade the majority classes accuracy in rare occasion. We will deal with these problem in future work. The code of this paper will be released after the review process.

REFERENCES

David Berthelot, Nicholas Carlini, Ian J. Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, 2019.

David Berthelot, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, 2020.

Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259, 2018.

Jonathon Byrd and Zachary Chase Lipton. What is the effect of importance weighting in deep learning? In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pp. 872–881. PMLR, 2019.

Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Aréchiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 1565–1576, 2019.

Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. Introduction to semi-supervised learning. In Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien (eds.), *Semi-Supervised Learning*, pp. 1–12. The MIT Press, 2006.

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.*, 16:321–357, 2002.

Hsin-Ping Chou, Shih-Chieh Chang, Jia-Yu Pan, Wei Wei, and Da-Cheng Juan. Remix: Rebalanced mixup. In *Computer Vision - ECCV 2020 Workshops - Glasgow, UK, August 23-28, 2020, Proceedings, Part VI*, 2020.

Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. Randaugment: Practical data augmentation with no separate search. *CoRR*, abs/1909.13719, 2019.

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge J. Belongie. Class-balanced loss based on effective number of samples. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 9268–9277. Computer Vision Foundation / IEEE, 2019.

Terrance Devries and Graham W. Taylor. Improved regularization of convolutional neural networks with cutout. *CoRR*, abs/1708.04552, 2017.

Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. In *Actes de CAP 05, Conférence francophone sur l'apprentissage automatique - 2005, Nice, France, du 31 mai au 3 juin 2005*, 2005.

Agrim Gupta, Piotr Dollar, and Ross Girshick. LVIS: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.*, 21(9):1263–1284, 2009.

Xinting Hu, Yulei Niu, Chunyan Miao, Xian-Sheng Hua, and Hanwang Zhang. On non-random missing labels in semi-supervised learning. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*, 2022.

Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. Deep imbalanced learning for face recognition and attribute prediction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(11):2781–2794, 2020.

Muhammad Abdullah Jamal, Matthew Brown, Ming-Hsuan Yang, Liqiang Wang, and Boqing Gong. Rethinking class-balanced methods for long-tailed visual recognition from a domain adaptation perspective. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, June 13-19, 2020*, 2020.

Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, 2020.

Jaehyung Kim, Youngbum Hur, Sejun Park, Eunho Yang, Sung Ju Hwang, and Jinwoo Shin. Distribution aligning refinery of pseudo-label for imbalanced semi-supervised learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, Department of Computer Science, University of Toronto, 2009.

Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

Dong-Hyun Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *In ICML Workshop on Challenges in Representation Learning, 2013*, 2013.

Hyuck Lee, Seungjae Shin, and Heeyoung Kim. ABC: auxiliary balanced classifier for class-imbalanced semi-supervised learning. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021*, 2021.

Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X. Yu. Large-scale long-tailed recognition in an open world. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 2537–2546. Computer Vision Foundation / IEEE, 2019.

Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.

Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 41(8):1979–1993, 2019.

Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. In *Deep Learning and Unsupervised Feature Learning Workshop, Advances in Neural Information Processing Systems 2011, NeurIPS 2011*, 2011.

Seulki Park, Jongin Lim, Younghan Jeon, and Jin Young Choi. Influence-balanced loss for imbalanced visual classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 735–744, October 2021.

Samira Pouyanfar, Yudong Tao, Anup Mohan, Haiman Tian, Ahmed S. Kaseb, Kent Gauen, Ryan Dailey, Sarah Aghajanzadeh, Yung-Hsiang Lu, Shu-Ching Chen, and Mei-Ling Shyu. Dynamic sampling in convolutional neural networks for imbalanced data classification. In *IEEE 1st Conference on Multimedia Information Processing and Retrieval, MIPR 2018, Miami, FL, USA, April 10-12, 2018*, pp. 112–117. IEEE, 2018.

Jiawei Ren, Cunjun Yu, Shunan Sheng, Xiao Ma, Haiyu Zhao, Shuai Yi, and Hongsheng Li. Balanced meta-softmax for long-tailed visual recognition. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Chuck Rosenberg, Martial Hebert, and Henry Schneiderman. Semi-supervised self-training of object detection models. In *7th IEEE Workshop on Applications of Computer Vision / IEEE Workshop on Motion and Video Computing (WACV/MOTION 2005), 5-7 January 2005, Breckenridge, CO, USA*, pp. 29–36. IEEE Computer Society, 2005.

Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, Barcelona, Spain,December 5-10, 2016,*, 2017.

Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. Meta-weight-net: Learning an explicit mapping for sample weighting. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp. 1917–1928, 2019.

Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, December 4-9, 2017,*, 2017.

Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8769–8778, 2018.

Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan L. Yuille, and Fan Yang. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021*, 2021.

Qizhe Xie, Zihang Dai, Eduard H. Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Zhengzhuo Xu, Zenghao Chai, and Chun Yuan. Towards calibrated model for long-tailed visual recognition from prior perspective. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021*, 2021.

Yuzhe Yang and Zhi Xu. Rethinking the value of labels for improving class-imbalanced learning. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Xi Yin, Xiang Yu, Kihyuk Sohn, Xiaoming Liu, and Manmohan Chandraker. Feature transfer learning for face recognition with under-represented data. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 5704–5713. Computer Vision Foundation / IEEE, 2019.

Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*, 2016.

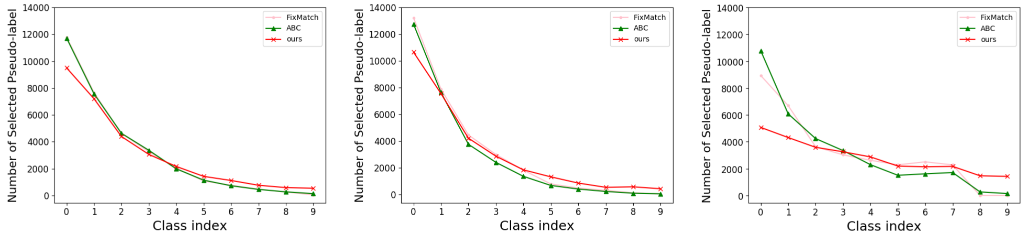
Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 18408–18419, 2021.

Hongyi Zhang, Moustapha Cissé, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2019.

Zhisheng Zhong, Jiequan Cui, Shu Liu, and Jiaya Jia. Improving calibration for long-tailed recognition. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021*, 2021.

Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. BBN: bilateral-branch network with cumulative learning for long-tailed visual recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, June 13-19, 2020*, 2020.

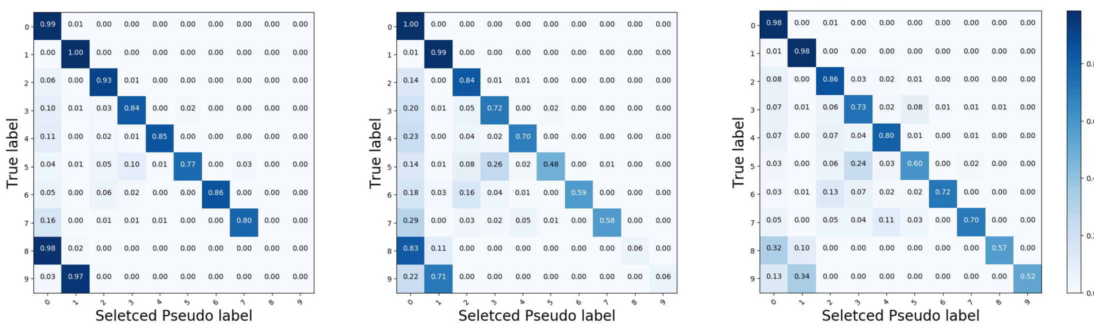
A THE IMBALANCE IN PSEUDO-LABELS



(a) $\gamma_l = \gamma_u = 50$ (b) $\gamma_l = \gamma_u = 100$ (c) $\gamma_l = 100, \gamma_u = 1$

Figure 2: The number of selected pseudo-labels for different classes in CIFAR-10-LT under different imbalanced ratio.

Fig.2 presents the prediction of unlabeled data of FixMatch, FixMatch+ABC and FixMatch+APLA trained on CIFAR-10-LT with different imbalanced ratio. The imbalance of pseudo-labels is more imbalanced than the true distribution of unlabeled data in FixMatch and FixMatch+ABC. The imbalanced ratio of pseudo-labels increases as the imbalance ratio increases in FixMatch and FixMatch+ABC. The huge mismatch between the distribution of unlabeled data and pseudo-labels leading to the huge performance degeneration of FixMatch and FixMatch +ABC in Fig. 2(c).



(a)FixMatch (b)FixMatch+ABC (c) FixMatch + APLA

Figure 3: Confusion matrices of the selected pseudo-labels on the unlabeled data of CIFAR-10-LT under imbalance ratio $\gamma_l = 100, \gamma_u = 1$.

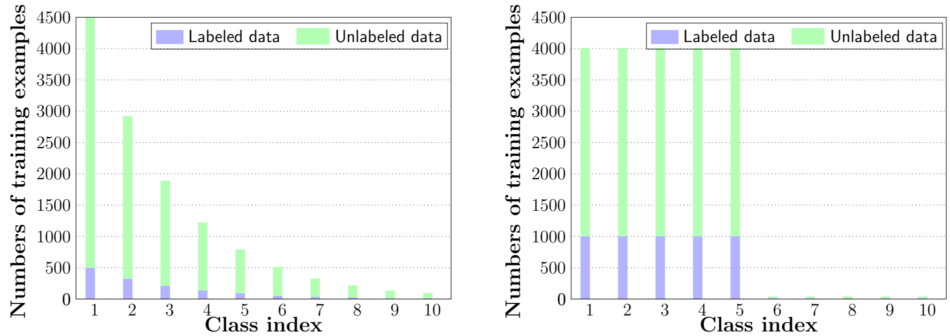
The results in Fig. 3 show that the pseudo-labels generated by FixMatch and FixMatch+ABC are biased towards majority classes. For example, almost all the unlabeled data that belong to class 9 are predicted wrongly as class 1 in FixMatch. Our APLA achieves a more unbiased confusion matrix of selected pseudo-labels on the unlabeled data.

B TWO TYPES OF CLASS IMBALANCE FOR THE CONSIDERED DATASETS

Two types of class imbalance for the considered datasets are illustrated in Fig. 4. In Fig. 4(a), we set $\gamma_l = \gamma_u = 50, n_1 = 500, m_1 = 4500$. In Fig. 4(b), we set $\gamma_l = \gamma_u = 100, n_1 = 1000, m_1 = 4000$. We can aslo see that each minority class of step-imbalance setting has a very small amount of data in Fig. 4(b). Existing SSL algorithms can be hardly perform well on minority class under step imbalanced settings due to the scarce data in minority class.

C MORE QUANTITATIVE COMPARISON

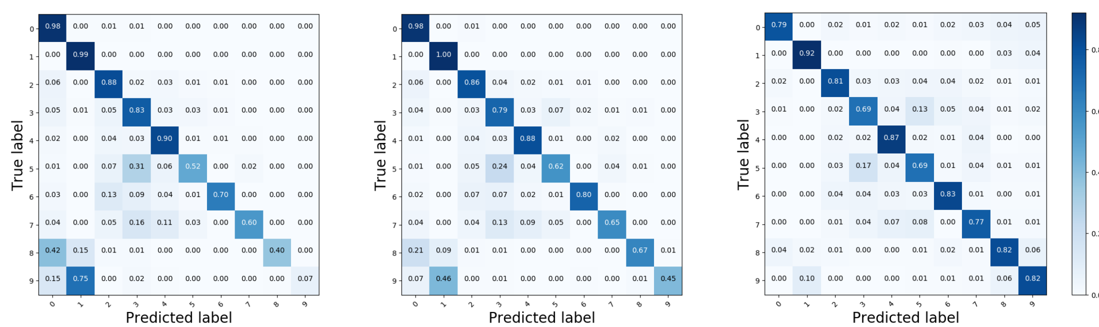
Fig. 5 presents the confusion matrices of FixMatch, FixMatch+ABC, and FixMatch+APLA trained on CIFAR-10-LT, $\gamma_l = \gamma_u = 100, n_1 = 500, m_1 = 4500$. FixMatch and FixMatch+ABC often



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

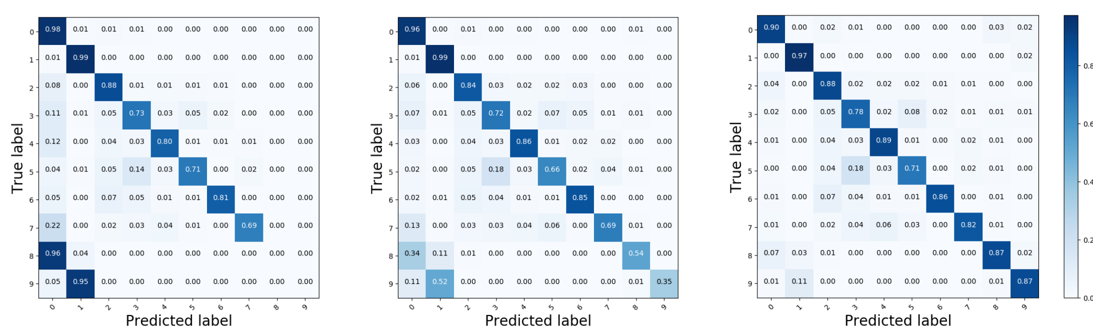
Figure 4: Long-tailed imbalance and step imbalance

misclassified test data points in the tail-classes as the data point in the head-classes (e.g., classes 8 and 9 into classes 0 and 1). In contrast, FixMatch+APLA classified the test data points in the tail-class with higher accuracy, and produced a significantly more balanced class-distribution than FixMatch and FixMatch+ABC. This phenomenon is even more significant in Fig. 6, when unlabeled data is balanced and labeled data is imbalanced.



(a)FixMatch (b)FixMatch+ABC (c) FixMatch + APLA

Figure 5: Confusion matrices of the prediction on the test set of CIFAR-10-LT under imbalance ratio $\gamma_l = \gamma_u = 100$.



(a)FixMatch (b)FixMatch+ABC (c) FixMatch + APLA

Figure 6: Confusion matrices of the prediction on the test set of CIFAR-10-LT under imbalance ratio $\gamma_l = 100, \gamma_u = 1$.