

# SEMI-SUPERVISED LONG-TAILED RECOGNITION USING ALTERNATE SAMPLING

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Long tailed recognition is confronted by two intervening challenges, *i.e.*, the sample scarcity in the tail classes and the imbalanced class distribution. The class geometry in feature space mainly suffers from the data scarcity, while imbalance distribution biases the decision boundary of classes. Previous work makes assumptions on the underneath geometric structure of the tail classes to address the data scarcity challenge, and resorts to class balanced sampling or reweighting to address the data imbalance challenge. We advocate to leverage the readily available unlabeled data in a semi-supervised setting to approach to long tailed recognition. An alternate sampling strategy is then introduced to overcome the two challenges in a single framework. The feature embedding (geometric structure) and classifier are updated in an iterative fashion. The extra unlabeled data, regularized by a consistency loss, leads to a better geometric structure. The class-balanced sampling is implemented to train the classifier such that it is not affected by the imbalance distribution or the quality of pseudo labels. We demonstrate significant accuracy improvements over other competitive methods on two datasets, where we improve on tail classes without much, if at all, degradations on head classes.

## 1 INTRODUCTION

Many real world applications need to deal with data of a long-tailed distribution, where the frequency of samples from different classes are imbalanced. As shown in Figure 1a(Top), several highly populated classes take up most of the labeled samples, and some of the classes only have very few samples during training. Two fundamental challenges of long-tailed distribution are the scarcity of data samples in the tail classes as well as the imbalanced training data distribution.

On the one hand, the data scarcity of tail classes limits the intra-class variations presented in the training data, which makes it extremely difficult to learn the correct geometric structure of those classes. On the other hand, with an imbalanced training data distribution, tail classes contribute little in the training loss, and hence the class decision boundaries is prone to be biased. In an extreme case, where two classes have 10,000 and 1 samples respectively, the model could totally fail on the second class with the training error less than 0.01%.

These two challenges intervening together make the long-tailed problem specifically hard to address. Methods focusing on few-shot learning have been introduced to address the data scarcity problem through data augmentation Wang et al. (2018); Hariharan & Girshick (2017); Liu et al. (2022) or class geometry transfer Liu et al. (2021; 2020). This, in fact, relies on an assumption that the geometry of tail classes should follow the existing training data or more specifically the head classes.

Many works in the long-tailed learning literature Lin et al. (2017); Cao et al. (2019); Kang et al. (2020b); Zhou et al. (2020) tried to mitigate the imbalance issue by re-sampling the training data to be a balanced distribution or calibrating the sample weights in calculating the loss. However, the designing of this calibration can be difficult to avoid negative impact on head classes. Furthermore, without considering the scarcity challenge, the tail class performance can not be improved significantly.

We propose to eliminate the geometry limitation of training data by leveraging massive unlabeled real data in training to help improve the long-tailed recognition accuracy. This yields a more re-

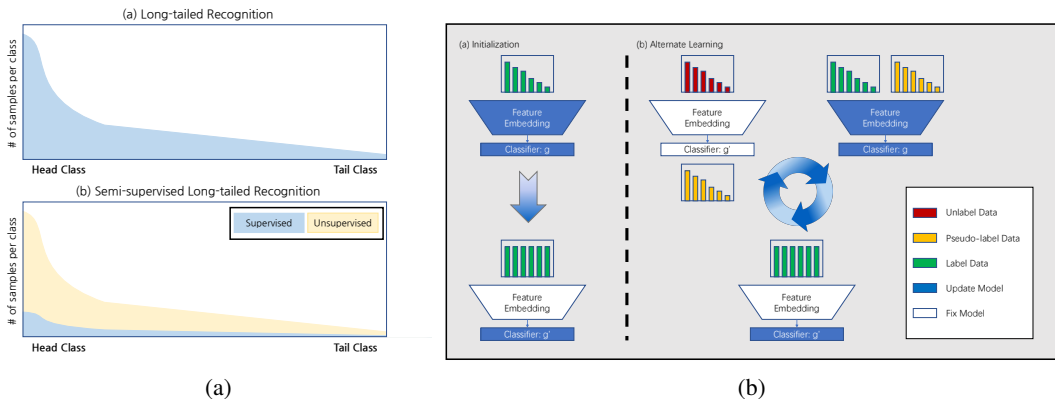


Figure 1: (a): Comparison of different recognition paradigms: Top: typical data distribution over classes in long-tailed recognition; Bottom: the proposed Semi-supervised long-tail recognition setting. (b) Left: Initialization procedure. A recognition model is first trained with random sampling. After that the feature embedding is used to train a new classifier with class-balanced sampling. Right: Diagram of alternate learning in three stages.

alistic and challenging setting, namely semi-supervised long-tailed recognition. As shown in Figure 1a(Bottom), we assume a long-tailed data distribution of the overall dataset and both the labeled and unlabeled subsets of training data follow the same underlying long-tailed data distribution.

This setting generally resembles a realistic data collection and annotation workflow. After collecting the raw data, one has no knowledge of its class distribution before annotation. As it is expensive to annotate the full corpus, a common practice is to randomly sample a subset for annotation under a given labeling budget. When the raw data follows a long-tailed class distribution, we should expect the same in the labeled subset.

To address the semi-supervised long-tailed recognition problem, An alternate learning framework is presented to update the feature embedding and the classifier in an iterative fashion. By decoupling the training procedure, it can leverage the advantages of extra data for class geometry and balance sampling for class boundaries, while avoid the unlabeled data balancing and pseudo label misleading difficulties. Furthermore, the iterative strategy is especially suitable for the semi-supervised setting, as pseudo label accuracy can be improved gradually.

Inspired by Kang et al. (2020b), we decouple the recognition model into a feature embedding and a classifier, and train them with random sampling and class-balanced sampling respectively. In the stage of the feature embedding updating, unsupervised data is used to mitigate the data scarcity problem. Following similar designs in semi-supervised learning literature Laine & Aila (2016); Tarvainen & Valpola (2017), unlabelled samples are assigned with pseudo labels and then incorporated in fine-tuning the feature embedding with a regularization term limiting its potential negative impacts.

After the feature embedding fine-tuning, class-balanced sampling is used to update the classifier, avoiding the bias on decision boundaries. When updating the classifier, only labeled data is used to get around the difficulty of applying correctly class-balanced sampling on unlabeled data, aligning with the intuition that the classifier needs more robust supervision than the feature embedding. These two steps are iteratively performed until convergence.

To summarize, in this paper, 1) we resort to semi-supervised learning to help improve long-tailed recognition accuracy and identify practical gap of current semi-supervised recognition datasets due to their well-balanced unlabeled subset; 2) we propose a new recognition paradigm named semi-supervised long-tailed recognition better resembling real-world data collection and annotation workflow; 3) we propose a new alternative sampling method to address the semi-supervised long-tailed recognition and demonstrate significant improvements on several benchmarks.

## 2 RELATED WORK

**Long-tailed recognition** has been recently studied a lot Wang et al. (2017); Oh Song et al. (2016); Lin et al. (2017); Zhang et al. (2017); Liu et al. (2019); Wang & Hebert (2016). Several approaches

have been proposed, including metric learning Oh Song et al. (2016); Zhang et al. (2017), loss weighting Lin et al. (2017), and meta-learning Wang & Hebert (2016), or ensemble models Xiang et al.; Wang et al. (2020); Zhang et al. (2022). Early attempts Oh Song et al. (2016); Zhang et al. (2017); Lin et al. (2017) design dedicated loss functions to mitigate the data imbalanced problem. For example, lift loss Oh Song et al. (2016) introduces margins between many training samples. As reported by Liu et al. (2019), when applied to long-tailed recognition, many of these methods improved accuracy of the few-shot group, but at the cost of lower accuracy over the many-shot classes.

Other methods, e.g. LDAM-DRW Cao et al. (2019) replace cross-entropy loss with LDAM loss. This adds a calibration factor to the original cross-entropy loss. When combined with loss re-weighting, it improves the accuracy in all splits in long-tailed recognition. However, it can not be easily generalized to semi-supervised learning. Because both the calibration factor and the loss weight are calculated based on the number of samples of each class.

Ensemble methods Xiang et al.; Wang et al. (2020); Zhang et al. (2022) train different experts for different splits of classes based on the distribution. The optimal splits can be hard to design and vary from one dataset to another, though good performance is achieved.

Following the success of contrastive learning, several works Kang et al. (2020a); Cui et al. (2021); Zhu et al. (2022) introduce the supervised contrastive loss Khosla et al. (2020) to long-tailed learning with balancing techniques, e.g. re-weighting or re-sampling. These methods mitigate the imbalance class decision boundary difficulty with the help of contrastive learning, but leave the scarcity problem unaddressed.

Recent works Ma et al. (2021); Tian et al. (2022) propose to language models to improve the long-tailed recognition performance. However, introducing language models and data is beyond the scope of this paper, where the long-tailed problem is discussed as a pure computer vision task.

Yang & Xu (2020) rethinks the value of labels in imbalance learning. As part of the discussion, semi-supervised learning is included. However, only the basic pseudo label solution and simple datasets, such as CIFAR and SVHN, are discussed.

Several works Kang et al. (2020b); Zhou et al. (2020) with improved long-tailed recognition share the observation that feature embedding and the classifier should be trained with different sampling strategies. In this work, we adopt our method on this observation to learn the feature embedding model with random sampling and train the classifier with class-balanced sampling. This design is further closely compatible with semi-supervised learning under alternate learning.

**Semi-supervised learning** has been extensively discussed in recognition discipline Laine & Aila (2016); Rasmus et al. (2015); Tarvainen & Valpola (2017). One common observation is to optimize the traditional cross-entropy loss together with a regularization term that regulates the perturbation consistency of unlabelled data.

Ladder net Rasmus et al. (2015) is introduced to minimise the reconstruction loss between the network outputs from a given sample and its perturbation. It is then simplified in Laine & Aila (2016) as two temporal modules:  $\Pi$ -Model and Temporal Ensembling. The Temporal Ensembling encourages the output of the network from unlabeled data to be similar to its counterpart from previous training epoch. More recently, Mean Teacher Tarvainen & Valpola (2017) extends it by assembling along training. Instead of storing previous predictions, they assemble a Teacher model by calculating the moving average of the training network, i.e. the Student. The Teacher is then used to provide the consistency of predictions to the Student.

In addition to that, MA-DNN Chen et al. (2018) introduces a memory module to maintain the category prototypes and provide regularization for learning with unlabeled data. Label propagation Li et al. (2018) is also considered with the help of label graph. More recently, Mixmatch Berthelot et al. (2019) and Fixmatch Sohn et al. (2020) improve the performance by introducing powerful data augmentations and perturbation consistencies.

All the semi-supervised methods above do not separate labeled data during semi-supervised training. In fact, it is beneficial to combine labeled data and unlabeled data in a certain proportion Laine & Aila (2016); Tarvainen & Valpola (2017). However, without further knowledge, we have no insight how to deal with this combination when long-tailed distribution is included. Furthermore, long-

tailed learning methods require calibration or re-sampling based on the class distribution. This combination of labeled and unlabeled data makes the distribution unstable. In result, this is not suitable for long-tailed recognition.

Recently, Salsa Rebuffi et al. (2020) proposes to decouple the supervised learning from semi-supervised training. Our method follows the alternate training scheme from it, because it is surprisingly compatible with long-tailed learning. In practice, our method differs from Salsa in the following aspects.

First, we adopt class-balanced sampling in supervised learning to deal with the long-tailed distribution. Second, we use supervised learning instead of self-supervised learning as initialization. We find that self-supervised learning results in inferior performance in long-tailed scenario. Third, the re-initialization is not needed. Because our initialization is already from supervised learning, there is not a specific starting point to re-initialize the model. In fact, this enhances the soft constraint between the two stages in Rebuffi et al. (2020).

With the models continuously optimized along alternate learning, our method achieves superior performance while maintains the same amount of training epochs as fine-tuning simply on pseudo labels.

### 3 METHOD

In this section, we will introduce the proposed method to semi-supervised long-tailed recognition. The semi-supervised long-tailed recognition problem is first defined, and some notations are clarified. The decoupling strategy of long-tailed recognition is then discussed. This is also the initialization phase of our method. After that, the alternate learning scheme with 3 stages is fully discussed.

#### 3.1 SEMI-SUPERVISED LONG-TAILED RECOGNITION

We start by defining the semi-supervised long-tailed recognition problem. Consider an image recognition problem with a labeled training set  $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i); i = 1, \dots, N\}$ , where  $x_i$  is an example and  $y_i \in \{1, \dots, C\}$  its label, where  $C$  is the number of classes. For semi-supervised learning, there is also an unsupervised training subset  $\mathcal{U} = \{\mathbf{x}_i; i = 1, \dots, M\}$ .

Although the labels of data in  $\mathcal{U}$  are not available, every sample has its label from  $\{1, \dots, C\}$ . For class  $j$ , we have  $n_j$  samples from  $\mathcal{D}$  and  $m_j$  samples from  $\mathcal{U}$ . With the assumption that supervised and unsupervised data follow the same distribution, we have the fact  $\frac{n_j}{N} = \frac{m_j}{M}, \forall j$ . The testing set, on the other hand, in order to evaluate the performance on every class without bias, is balanced sampled on all classes in  $\{1, \dots, C\}$ .

#### 3.2 MODEL DECOUPLING AND DATA SAMPLING

A CNN model combines a feature embedding  $\mathbf{z} = f(\mathbf{x}; \theta) \in \mathbb{R}^d$ , and a classifier  $g(\mathbf{z}) \in [0, 1]^C$ . Embedding  $f(\mathbf{x}; \theta)$  is implemented by several convolutional layers of parameters  $\theta$ . The classifier operates on the embedding to produce a class prediction  $\hat{y} = \arg \max_i g_i(\mathbf{z})$ . In this work, we adopt the popular linear classifier  $g(\mathbf{z}) = \nu(\mathbf{W}\mathbf{z} + \mathbf{b})$ , where  $\nu$  is the softmax function.

Standard (random sampling) training of the CNN lies on mini-batch SGD, where each batch is randomly sampled from training data. A class  $j$  of  $n_j$  training examples has probability  $\frac{n_j}{N}$  of being represented in the batch. Without loss of generality, we assume classes sorted by decreasing cardinality, i.e.  $n_i \leq n_j, \forall i > j$ . In the long-tailed setting, where  $n_1 \gg n_C$ , the model is not fully trained on classes of large index  $j$  (tail classes) and under-fits. This can be avoided with recourse to non-uniform sampling strategies, the most popular of which is class-balanced sampling. This samples each class with probability  $\frac{1}{C}$ , over-sampling tail classes.

Kang et al. (2020b); Zhou et al. (2020) shows that while classifier benefits from class-balanced sampling, feature embedding is more robust in random sampling. Practically, Kang et al. (2020b) achieves this by decoupling the training into two stages, and train the feature embedding with random sampling in the first stage, and classifier the second with class-balanced sampling.

### 3.3 INITIALIZATION

The initialization of the proposed method follows the decoupling from Kang et al. (2020b). The two-stage initialization is illustrated in Figure 1b(Left). A CNN model is first trained with random sampling. A feature embedding  $\mathbf{z} = f(\mathbf{x}; \theta) \in \mathbb{R}^d$ , and a classifier  $g'(\mathbf{z}) \in [0, 1]^C$  are acquired. After convergence, the classifier is re-initialized and trained with class-balanced sampling, with the feature embedding fixed. This results in a class-balanced classifier  $g(\mathbf{z}) \in [0, 1]^C$ . Both the feature embedding and the classifier are trained on the supervised training subset  $\mathcal{D}$ .

### 3.4 ALTERNATE LEARNING

After obtaining an initialized model, most semi-supervised learning methods fine-tune the model on a combination of supervised and unsupervised samples. This is, however, incompatible with our long-tailed recognition model. When applied on unsupervised data, we have no ground truth for class-balanced sampling. One can make a sacrifice by relying on pseudo labels assigned by the initialized model. But the effectiveness will depend on the accuracy of pseudo labels.

It is even worse when considering the fact that long-tailed models usually have better performance on highly populated classes and worse on few-shot classes. Class-balanced sampling over-samples few-shot classes, while down-samples many-shot. This means, in general, the worse part of pseudo labels contributes more to the training loss than it should be, while the better part contributes less.

Another difficulty is the model compatibility when combining the long-tailed model to semi-supervised learning methods. Many semi-supervised learning methods evolve the model and pseudo labels at the same time. For example, Mean Teacher Tarvainen & Valpola (2017) assembles the teacher model by moving average and trains the student with consistency loss. When it comes to long-tailed model, it is not clear when we should update the feature embedding or classifier. And it is also difficult to incorporate both random and class-balanced sampling.

Inspired by Rebuffi et al. (2020), which separates supervised learning apart from semi-supervised learning, we propose an alternate learning scheme. The supervised training on data  $\mathcal{D}$ , and semi-supervised training on data  $\mathcal{D} \cup \mathcal{U}$  are carried out in an alternate fashion together with model decoupling and different data sampling strategies.

In practice, after initialization, we have a feature embedding  $\mathbf{z} = f(\mathbf{x}; \theta)$ , a classifier  $g'(\mathbf{z})$  trained with random sampling, and a classifier  $g(\mathbf{z})$  trained with class-balanced sampling. In Kang et al. (2020b), only  $g(\mathbf{z})$  is used in testing. However, we keep the randomly trained classifier  $g'(\mathbf{z})$  for further usage. The training scheme iterates among 3 stages for  $N$  loops, which are shown in Figure 1b(Right).

**Stage 1: Label assignment.** In this stage, pseudo labels are assigned for the unsupervised subset  $\mathcal{U}$ . The feature embedding  $f(\mathbf{x}; \theta)$  and class-balanced classifier  $g(\mathbf{z})$  are used. The choice of classifier is equivalent to the long-tailed model when tested for better overall accuracy. The unsupervised subset with pseudo labels is  $\hat{\mathcal{U}} = \{(\mathbf{x}_i, \hat{y}_i); i = 1, \dots, M\}$ , where  $\hat{y}_i$  are pseudo labels.

**Stage 2: Semi-supervised training.** After label assignment, we have pseudo labels for all unsupervised data. The model is fine-tuned on the combination of true and pseudo labels, i.e. on  $\mathcal{D} \cup \hat{\mathcal{U}}$ . In this stage, random sampling is used to update the feature embedding  $f(\mathbf{x}; \theta)$  and the randomly-trained classifier  $g'(\mathbf{z})$ . The classification is optimized by cross-entropy loss:

$$\mathcal{L}_{CE} = \sum_{(\mathbf{x}_i, y_i) \in \mathcal{D} \cup \hat{\mathcal{U}}} -\log g'_{y_i}(f(\mathbf{x}_i; \theta)), \quad (1)$$

where  $g'_{y_i}$  is the  $y_i$ -th element of  $g'$ .

In semi-supervised learning literature, a regularization loss is usually applied to maintain the consistency for unlabeled data. This consistency loss captures the fact that data points in the neighborhood usually share the same label. We adopt this idea and implement the temporal consistency from Laine & Aila (2016). In practice, the class probabilities are acquired from  $g'$ . Given the class probability  $p^{e-1}$  from epoch  $e-1$ , and the class probability  $p^e$  from epoch  $e$ , the loss is KL-divergence between

Table 1: Results(Accuracy in %) on CIFAR-10-SSLT. ResNet-18 is used for all methods. The best of each column is highlighted with **red background**, the second with **green**, and the third with **blue**.

Method	Imbalance factor=100				Imbalance factor=1000			
	Overall	Many-Shot	Medium-Shot	Few-Shot	Overall	Many-Shot	Medium-Shot	Few-Shot
LDAM-DRW (L) Cao et al. (2019)	67.4	79.7	54.2	68.1	46.2	70.3	36.3	35.6
Pseudo-Label + L	69.6	69.7	55.1	80.2	48.4	74.0	39.3	36.0
Mean Teacher Tarvainen & Valpola (2017) + L	69.9	69.7	57.3	79.4	48.3	75.7	41.4	32.9
Decoupling (D) Kang et al. (2020b)	64.0	91.1	63.0	44.4	45.8	86.5	47.2	14.4
Pseudo-Label + D	68.9	92.7	70.8	49.8	46.5	89.0	47.0	14.2
Ours	71.3	89.5	67.7	60.2	66.7	84.4	69.4	51.4

the two.

$$\mathcal{L}_{consist} = \sum_{(\mathbf{x}_i, y_i) \in \mathcal{D} \cup \tilde{\mathcal{U}}} \sum_j p_j^{e-1} \log \frac{p_j^{e-1}}{p_j^e}, \quad (2)$$

where  $p_j^{e-1}$  and  $p_j^e$  are the  $j$ -th element of  $p^{e-1}$  and  $p^e$  respectively.

Overall, the semi-supervised learning loss is the combination of the two.

$$\mathcal{L}_{semi} = \mathcal{L}_{CE} + \lambda \mathcal{L}_{consist}. \quad (3)$$

**Stage 3: Supervised training.** We update the class-balanced classifier  $g(\mathbf{z})$  based on the refined feature embedding, which is fine-tuned with semi-supervised learning in Stage 2. Specifically, the fine-tuning is applied with class-balanced sampling and only on the supervised subset  $\mathcal{D}$ . In this stage, only classifier is updated. The feature embedding is fixed and only used in forwarding. Given the class-balanced version of supervised subset  $\mathcal{D}'$ , the cross-entropy loss for classification is

$$\mathcal{L}_{sup} = \sum_{(\mathbf{x}_i, y_i) \in \mathcal{D}'} -\log g_{y_i}(f(\mathbf{x}_i; \theta)), \quad (4)$$

where  $g_{y_i}$  is the  $y_i$ -th element of  $g$ .

### 3.5 INSIGHT OF THE DESIGN

**Feature embedding** is trained with random sampling and semi-supervised learning. This is consistent with long-tailed model in the sampling scheme. It also follows the fact that feature embedding is less prone to noisy labels. Actually, in self-supervised learning literature Gidaris et al. (2018); He et al. (2020); Chen et al. (2020), the feature embedding can even be learned without labels.

**Classifier** is learned with class-balance sampling and only supervised data. This is again the same as the supervised version. And by avoiding fitting the classifier on pseudo labels, we prevent the wrong labels from propagating through the whole training process. Given the fact that the pseudo labels are provided by the classifier, if classifier is still optimize on those, wrong labels can be easily maintained in the fine-tuned version of the classifier.

Training the classifier only on labeled data also avoids the dilemma of class-balancing on unlabeled data. Without ground truth labels, class-balanced sampling can only rely on pseudo labels, which are not perfect. And the fact that pseudo labels have more errors on few-shot classes is specially not suitable for class-balanced sampling. Because when few-shot classes are over-sampled, those errors are also scaled up during training.

## 4 EXPERIMENTS

### 4.1 DATASETS

We manually curate two semi-supervised long-tailed recognition benchmarks.

**CIFAR-10-SSLT.** For easy comparison and ablation, we compose a lightweight semi-supervised long-tailed dataset based on CIFAR-10 Krizhevsky et al. (2009). Following Cao et al. (2019), we randomly sample the training set of CIFAR-10 under an exponential function with imbalance ratios

Table 2: Results(Accuracy in %) on ImageNet-SSLT (imbalance factor=125). ResNet-18/50 are used for all methods. For many-shot  $t > 100$ , for medium-shot  $t \in (10, 100]$ , and for few-shot  $t \leq 10$ , where  $t$  is the number of labeled samples. The best of each column is highlighted with red background, the second with green, and the third with blue.

Method	ResNet-18				ResNet-50			
	Overall	Many-Shot	Medium-Shot	Few-Shot	Overall	Many-Shot	Medium-Shot	Few-Shot
LDAM-DRW (L) Cao et al. (2019)	21.3	42.6	27.0	8.6	24.9	51.2	31.1	9.9
Pseudo-Label + L	17.6	22.4	20.9	12.6	23.9	44.0	30.0	11.1
Mean Teacher Tarvainen & Valpola (2017) + L	21.3	41.8	28.1	7.6	25.6	49.1	31.8	11.7
Decoupling (D) Kang et al. (2020b)	24.8	53.9	31.1	8.7	27.2	58.5	34.2	9.8
Pseudo-Label + D	25.3	47.6	32.1	11.1	27.7	52.2	34.7	12.4
Ours	26.5	52.0	33.9	10.7	29.0	57.1	36.5	12.3

in  $\{100, 1000\}$  (the ratio of most populated class to least populated). The unsupervised subset is collected from Tiny Images Torralba et al. (2008) following the strategy introduced in Yang & Xu (2020). The class distribution of unlabeled data is always the same as the labeled one, with 5 times larger. For better description and comparison, we assign the 10 classes into 3 splits: many-shot, medium-shot, few-shot, with many-shot the most populated 3 classes, medium-shot the medium 3, and few-shot the least 4 classes. The supervised subset is more imbalance than those used by traditional long-tailed tasks (usually 10 to 100 imbalance ratio). This is designed to emphasize the importance of semi-supervised learning in the case of extreme long-tailed data distribution.

**ImageNet-SSLT.** To evaluate the effectiveness of semi-supervised long-tailed recognition methods on large-scale datasets, we assemble a challenging dataset from ImageNet (ILSVRC-2012) Deng et al. (2009). The supervised subset is sampled with Lomax distribution with shape parameter  $\alpha = 6$ , scale parameter  $\lambda = 1000$ . It contains 41, 134 images from 1000 classes, with the maximum of 250 images per class and the minimum of 2 samples. The unsupervised subset is sampled under the same distribution with an unsupervised factor 4, i.e.  $|\mathcal{U}| = 4|\mathcal{D}|$ . The 1000 classes are divided into 3 splits based on the amount of labeled data  $n$ : many-shot ( $n > 100$ ), medium-shot ( $10 < n \leq 100$ ), few-shot ( $n \leq 10$ ). In result, the dataset has 140 many-shot, 433 medium-shot, and 427 few-shot classes. Methods are evaluated under all classes and each class split.

## 4.2 NETWORK ARCHITECTURE

ResNet-18 He et al. (2016) is used on both CIFAR-10-SSLT and ImageNet-SSLT for fast experiments and comparison. ResNet-50 He et al. (2016) is used on ImageNet-SSLT to show how methods scale up to larger networks.

## 4.3 COMPARISON METHODS

To our best knowledge, there is no available method designated for semi-supervised long-tailed recognition. We explore typical long-tailed recognition methods and semi-supervised recognition methods, and combine them as baselines.

**Long-tailed recognition.** We consider two long-tailed methods, one for loss calibration and the other for re-sampling. LDAM-DRW Cao et al. (2019) converts cross-entropy loss to LDAM loss with calibration factors based on class counts. It further regulates the loss with a loss weight also from class counts. Decoupling Kang et al. (2020b) decouples the training of embedding and classifier with different sampling strategies. This is also the initialization in our method.

**Semi-supervised recognition.** Pseudo-Label is a basic semi-supervised learning algorithm and can be easily combined with other models. It contains two phases. The first phase is initialization, the recognition model is trained on labeled data. Predictions of the initialized model are assigned on unlabeled data, i.e. pseudo labels. The initialized model is then trained or fine-tuned on the combination of labeled and unlabeled data. In practice, we combine Pseudo-Label method with the two long-tailed recognition models to create two semi-supervised long-tailed recognition baselines. Pseudo-Label combined with LDAM-DRW is the method used in Yang & Xu (2020). Mean Teacher Tarvainen & Valpola (2017) is a well-known semi-supervised learning method. It contains a Student model that is trained with SGD and a Teacher model that is updated with moving average of the Student. It is, however, unclear how to train it with Decoupling. We only implement LDAM loss with Student training.

Table 3: (a) Ablation results(Accuracy in %) on CIFAR-10-SSLT, Imbalance factor 100 is used. Sampling methods are denoted as R for random, and C for class-balanced. The last two method names shows where the embedding is trained. (b) Pseudo label accuracy on unlabeled training subset. CIFAR-10-SSLT with imbalance ratio 100 is used. Compared to testing set, the unsupervised subset is not balanced. In result, the overall accuracy is higher than that on testing set, because of the domination of many-shot classes. The results in many/medium/few-shot splits are more useful.

(a)					(b)				
Method	Overall	Many-Shot	Medium-Shot	Few-Shot	Loop	Overall	Many-Shot	Medium-Shot	Few-Shot
R + R	50.9	93.0	57.8	14.1	0	87.7	92.3	63.0	41.8
C + R	61.2	91.3	62.6	37.6	1	87.9	92.3	64.0	48.1
C + C	63.3	91.2	64.4	41.6	2	87.8	92.1	64.7	52.2
$\mathcal{D} \cup \mathcal{U}'$	70.1	89.6	68.7	56.5	3	87.8	91.8	65.3	55.8
$\mathcal{D}$	63.3	91.6	61.9	43.2	4	87.7	91.6	65.8	57.8

#### 4.4 TRAINING DETAIL

In initialization, the feature embedding is trained with 200 epochs, and classifier is learned in 10 epochs after that. Stage 2 contains 40 epochs of fine-tuning of the embedding on the whole dataset. In 5 loops of stages, it is in total 200 epochs of embedding fine-tuning. There are also 10 epochs of classifier fine-tuning in Stage 3 per loop. In semi-supervised learning loss (3),  $\lambda = 1$  is used. SGD optimizer with learning rate of 0.1 is used with cosine annealing during training in all stages. The momentum is 0.9, and weight decay is 0.0005. All comparison methods are implemented with the hyper-parameters in their papers. The codes from authors are used when available.

#### 4.5 RESULTS

**CIFAR-10-SSLT** results are shown in Table 1 with imbalance ratio 100 and 1000. Our methods outperforms all other methods in overall accuracy. Among all 6 splits, we get the best performance 2 times, the second 1 time, and the third 2 times.

Our initialized model is equivalent to Decoupling, which shows the worst performance among all methods. Alternate learning improves the overall performance more than 7% when imbalance factor is 100, and 20% with imbalance factor 1000. Most of the improvement is from medium and few-shot classes. The larger improvement on the more imbalanced distribution shows that our method is more effective with more skewed dataset. When Pseudo-Label is added upon Decoupling, around 5% improvement is achieved with imbalance factor 100. But this improvement diminishes when the data is more imbalanced. This implies the fact that Pseudo-Label is more sensitive to bad tail class labelling.

With the improvement upon Pseudo-Label, our method has the same amount of training epochs on unsupervised data. The extra calculation in our methods compared to Pseudo-Label is from Stage 1 and 2. However, the classifier training is only on supervised data, and only the linear classifier is updated. And label assignment does not involve any back-propagation. The extra time on these two stages are trivial compared to the training of the whole model on the whole dataset.

LDAM-DRW provides very competitive results without any semi-supervised learning methods when imbalance factor is 100. However, it scales up bad when combined with semi-supervised techniques. By adding Pseudo-Label, it only improves 2% of overall accuracy. After looking at the splits results, we find that it improves the few-shot performance at the cost of many-shot. We believe this is because the wrong balancing factor introduced in LDAM loss. It does not match the true distribution, and skews the training process. Mean Teacher makes little difference from Pseudo-Label on LDAM-DRW.

**ImageNet-SSLT** results are shown in Table 2. Our methods outperforms all baseline methods with both ResNet-18 and -50 architectures. Among all 6 splits, we get the best performance 2 times, the second 3 times, and the third 1 time. The ImageNet-SSLT setting is really challenging that all of the methods give below 30% overall accuracy. In fact, our method is the only one that improves the few-shot performance while maintains the many-shot accuracy.

On ImageNet-SSLT, Pseudo-Label based methods (Pseudo-label +L/D) lose efficacy, because they improve few-shot performance marginally with big sacrifice on many-shot, which is not observed



when Pseudo-Label is used on CIFAR-10-SSLT. This is due to the bad many-shot pseudo-label quality on ImageNet-SSLT. Unlike CIFAR-10-SSLT, where the initialized model has 90% of accuracy on many-shot, many-shot performance on ImageNet-SSLT is only around 50%. These wrong labels can mislead the training and lower the performance of Pseudo-Label methods. Our method, on the other hand, updates the pseudo labels iteratively, and is less prone to this problem.

Specifically, adding Pseudo-Label on LDAM-DRW decreases the overall performance. This can be explained by the fact that the balancing factor in it does not match the true distribution. Mean Teacher improves upon LDAM-DRW when ResNet-50 is used. But it is still not as good as ours.

#### 4.6 ABLATIONS

We further study the training choices of alternate learning. This consists of two parts, i.e. the sampling choices and semi-supervised learning choices. Results on CIFAR-10-SSLT with imbalance factor 100 are listed in Table 3a.

**Sampling choice.** Currently, during alternate learning we use random sampling in Stage 2 and class-balance sampling in Stage 3. This is consistent with long-tailed recognition Kang et al. (2020b). However, other combinations are possible. Results are listed in the first 3 lines of Table 3a, with naming format: {sampling in Stage 2}+{sampling in Stage 3}. None of the 3 alternatives can beat the initialized model (Decoupling). This is expected. When the classifier is randomly trained (“R+R” and “C+R”), the model performs bad on few-shot classes. This will in turn harm the training of embedding by pseudo labels on unsupervised subset. “C+C” trains the feature embedding with class-balanced sampling. However, it is balancing on pseudo labels, which can be wrong. The results show that this balancing yields inferior feature embedding.

**Semi-supervised learning choice.** We train feature embedding with the whole dataset, i.e.  $\mathcal{D} \cup \mathcal{U}'$ , and the classifier with labeled subset  $\mathcal{D}$ . Other combinations can also be investigated. The classifier can also be semi-supervised trained, i.e. on  $\mathcal{D} \cup \mathcal{U}'$ . At the same time, feature embedding is trained with or without  $\mathcal{U}'$ . We show the results in the last 2 lines of Table 3a. In these two experiments, the classifier is always trained on  $\mathcal{D} \cup \mathcal{U}'$ . The difference is whether  $\mathcal{U}'$  is used for embedding learning. Compared to the regular setting, where the classifier is trained on  $\mathcal{D}$ , when we train it on  $\mathcal{D} \cup \mathcal{U}'$ , the performance is slightly lower. This can be explained by the fact that wrong pseudo labels in  $\mathcal{U}'$  can be propagated through loops if the classifier is optimized on them. This is especially true for few-shot classes, where the accuracy is low. Because of class-balanced sampling, the impact of few-shot classes is amplified. When compared to Table 1, the main performance drop is from few-shot classes. This confirms our assumption. However, when we further remove the unsupervised training of embedding, the performance drops a lot. It is even worse than the initialized model (Decoupling). In this case, the feature embedding should be equivalent to that of the initialization. The only difference is the classifier. This further proves the fact that fine-tuning classifier on pseudo-labels harms the performance.

**Accuracy on unsupervised training subset.** In Stage 1, we assign pseudo labels for all samples in  $\mathcal{U}$ . Table 3b reveals how the accuracy changes along loops in all splits. Few-shot split performance improves much faster than others. This proves the effectiveness of our alternate learning scheme, and explains why our method outperforms the baselines by a large margin in few-shot classes. The unsupervised subset has a long-tailed distribution, so the overall performance is dominated by many-shot. However, alternate learning still gets benefits from the improvement on few-shot split. Accuracy on different splits is more useful when we analyze how the model evolves during training.

## 5 CONCLUSION

This work introduces the semi-supervised long-tailed recognition problem. It extends the long-tailed problem with unsupervised data. With the property of labeled and unlabeled data obeying the same distribution, this problem setting follows the realistic data collection and annotation. A method based on alternate learning is proposed. By separating supervised training from semi-supervised and decoupling the sampling methods, it incorporates the decoupling training scheme in long-tailed recognition with semi-supervised learning. Experiments show that the proposed method outperforms all baselines. When results are split based on class cardinality, the method exhibits its robustness to defective pseudo labels. This is especially true for few-shot classes.

## REFERENCES

- David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. *arXiv preprint arXiv:1905.02249*, 2019.
- Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *arXiv preprint arXiv:1906.07413*, 2019.
- Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.
- Yanbei Chen, Xiatian Zhu, and Shaogang Gong. Semi-supervised deep learning with memory. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 268–283, 2018.
- Jiequan Cui, Zhisheng Zhong, Shu Liu, Bei Yu, and Jiaya Jia. Parametric contrastive learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 715–724, 2021.
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.
- Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*, 2018.
- Bharath Hariharan and Ross Girshick. Low-shot visual recognition by shrinking and hallucinating features. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3018–3027, 2017.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9729–9738, 2020.
- Bingyi Kang, Yu Li, Sa Xie, Zehuan Yuan, and Jiashi Feng. Exploring balanced feature spaces for representation learning. In *International Conference on Learning Representations*, 2020a.
- Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. In *Eighth International Conference on Learning Representations (ICLR)*, 2020b.
- Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33:18661–18673, 2020.
- Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.
- Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016.
- Qimai Li, Xiao-Ming Wu, and Zhichao Guan. Generalized label propagation methods for semi-supervised learning. 2018.
- Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.
- Bo Liu, Haoxiang Li, Hao Kang, Gang Hua, and Nuno Vasconcelos. Gistnet: a geometric structure transfer network for long-tailed recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8209–8218, 2021.

- Bo Liu, Haoxiang Li, Hao Kang, Gang Hua, and Nuno Vasconcelos. Breadcrumbs: Adversarial class-balanced sampling for long-tailed recognition. In *European Conference on Computer Vision*, pp. 637–653. Springer, 2022.
- Jialun Liu, Yifan Sun, Chuchu Han, Zhaopeng Dou, and Wenhui Li. Deep representation learning on long-tailed data: A learnable embedding augmentation perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2970–2979, 2020.
- Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X Yu. Large-scale long-tailed recognition in an open world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2537–2546, 2019.
- Teli Ma, Shijie Geng, Mengmeng Wang, Jing Shao, Jiasen Lu, Hongsheng Li, Peng Gao, and Yu Qiao. A simple long-tailed recognition baseline via vision-language model. *arXiv preprint arXiv:2111.14745*, 2021.
- Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4004–4012, 2016.
- Antti Rasmus, Harri Valpola, Mikko Honkala, Mathias Berglund, and Tapani Raiko. Semi-supervised learning with ladder networks. *arXiv preprint arXiv:1507.02672*, 2015.
- Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Semi-supervised learning with scarce annotations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 762–763, 2020.
- 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. *arXiv preprint arXiv:2001.07685*, 2020.
- Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv preprint arXiv:1703.01780*, 2017.
- Changyao Tian, Wenhai Wang, Xizhou Zhu, Jifeng Dai, and Yu Qiao. VI-ltr: Learning class-wise visual-linguistic representation for long-tailed visual recognition. In *European Conference on Computer Vision*, pp. 73–91. Springer, 2022.
- Antonio Torralba, Rob Fergus, and William T Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 30(11):1958–1970, 2008.
- Grant Van Horn, Oisín 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*, pp. 8769–8778, 2018.
- Xudong Wang, Long Lian, Zhongqi Miao, Ziwei Liu, and Stella Yu. Long-tailed recognition by routing diverse distribution-aware experts. In *International Conference on Learning Representations*, 2020.
- Yu-Xiong Wang and Martial Hebert. Learning to learn: Model regression networks for easy small sample learning. In *European Conference on Computer Vision*, pp. 616–634. Springer, 2016.
- Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Learning to model the tail. In *Advances in Neural Information Processing Systems*, pp. 7029–7039, 2017.
- Yu-Xiong Wang, Ross Girshick, Martial Hebert, and Bharath Hariharan. Low-shot learning from imaginary data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7278–7286, 2018.
- Liuyu Xiang, Guiguang Ding, and Jungong Han. Learning from multiple experts: Self-paced knowledge distillation for long-tailed classification.

Yuzhe Yang and Zhi Xu. Rethinking the value of labels for improving class-imbalanced learning. *arXiv preprint arXiv:2006.07529*, 2020.

Xiao Zhang, Zhiyuan Fang, Yandong Wen, Zhifeng Li, and Yu Qiao. Range loss for deep face recognition with long-tailed training data. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5409–5418, 2017.

Yifan Zhang, Bryan Hooi, Lanqing Hong, and Jiashi Feng. Self-supervised aggregation of diverse experts for test-agnostic long-tailed recognition. *Advances in Neural Information Processing Systems*, 3, 2022.

Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9719–9728, 2020.

Jianggang Zhu, Zheng Wang, Jingjing Chen, Yi-Ping Phoebe Chen, and Yu-Gang Jiang. Balanced contrastive learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6908–6917, 2022.

## A TRAINING DETAIL

In initialization, the feature embedding is trained with 200 epochs, and classifier is learned in 10 epochs after that. Stage 2 contains 40 epochs of fine-tuning of the embedding on the whole dataset. In 5 loops of stages, it is in total 200 epochs of embedding fine-tuning. There are also 10 epochs of classifier fine-tuning in Stage 3 per loop. In semi-supervised learning loss,  $\lambda = 1$  is used.

SGD optimizer with learning rate of 0.1 is used with cosine annealing during training in all stages. The momentum is 0.9, and weight decay is 0.0005.

All comparison methods are implemented with the hyper-parameters in their papers. The codes from authors are used when available.

## B MORE ABLATIONS

**Accuracy on unsupervised training subset.** In Stage 1, we assign pseudo labels for all samples in  $\mathcal{U}$ . Table 4 reveals how the accuracy changes along loops in all splits. Few-shot split performance improves much faster than others. This proves the effectiveness of our alternate learning scheme, and explains why our method outperforms the baselines by a large margin in few-shot classes.

The unsupervised subset has a long-tailed distribution, so the overall performance is dominated by many-shot. However, alternate learning still gets benefits from the improvement on few-shot split. Accuracy on different splits is more useful when we analyze how the model evolves during training.

**More results.** Table 6 compares our method to Salsa and FixMatch. The gains are significant for all data splits. And ResNet-32 results are shown in Table 5.

Table 4: Pseudo label accuracy on unlabeled training subset. CIFAR-10-SSLT with imbalance ratio 100 is used. Compared to testing set, the unsupervised subset is not balanced. In result, the overall accuracy is higher than that on testing set, because of the domination of many-shot classes. The results in many/medium/few-shot splits are more useful.

Loop	Overall	Many-Shot	Medium-Shot	Few-Shot
0	87.7	92.3	63.0	41.8
1	87.9	92.3	64.0	48.1
2	87.8	92.1	64.7	52.2
3	87.8	91.8	65.3	55.8
4	87.7	91.6	65.8	57.8

Table 5: Results(Accuracy in %) on CIFAR-10-SSLT, imbalance factor 100. ResNet-32 is used for all methods.

Method	Overall	Many-Shot	Medium-Shot	Few-Shot
LDAM-DRW	77.34	91.8	74.3	73.7
Pseudo-Label + L	81.1	87.6	75.6	80.7
Decoupling	68.2	91.6	66.9	49.2
Ours	83.1	92.0	77.7	80.4

Table 6: Results(Accuracy in %) on CIFAR-10-SSLT, imbalance factor 100. ResNet-18 is used for all methods.

Method	Overall	Many-Shot	Medium-Shot	Few-Shot
Salsa	59.6	82.5	60.7	41.5
FixMatch	64.1	83.6	62.4	50.6
Ours	71.3	89.5	67.7	60.2

Table 7: Results(Accuracy in %) on iNaturalist2018-SSLT. ResNet-50 are used for all methods. For many-shot  $t > 100$ , for medium-shot  $t \in (10, 100]$ , and for few-shot  $t \leq 10$ , where  $t$  is the number of labeled samples.

Method	Overall	Many-Shot	Medium-Shot	Few-Shot
Decoupling	27.9	54.1	41.7	24.8
Pseudo-Label + Decoupling	26.3	39.9	35.8	24.3
Ours	28.4	49.5	38.7	26.1

## C INATURALIST2018-SSLT

**Dataset.** We further curate a benchmark for semi-supervised long-tailed recognition based on iNaturalist 2018 (Van Horn et al. (2018)). iNaturalist 2018 is a long-tailed dataset sampled from natural distribution. We follow the distribution in both of the labeled and unlabeled subset. More specifically, Samples in each class is randomly down-sampled one-fifth of the total number as labeled data, and the remains are assigned as unsupervised subset. Classes with less than 2 labeled samples are eliminated. In result, iNaturalist2018-SSLT contains 8080 classes, with labeled samples from 200 to 2, and the unsupervised subset is 4 times larger.

Classes are divided into three splits based on the number of labeled samples: many-shot ( $[100, +\infty)$ ), medium-shot ( $[10, 100)$ ), and few-shot ( $[2, 10)$ ). It is a extremely long-tailed dataset, with 134 many-shot classes, 1220 medium-shot classes, and 7010 few-shot classes.

**Results.** Results are shown in Table 7. Our method is the only one that improves the overall performance upon baseline. iNaturalist2018-SSLT is different from our other benchmarks in the amount of few-shot classes. It has a very long tail taking up 87% of the label space. This makes the dataset especially hard when combined with unsupervised data.

With the inferior quality of predictions, we see significant drop of Pseudo-Label method in many-shot split. In fact, Pseudo-Label decreases the accuracy of baselines in all splits. Our method mitigates this problem, and improve the few-shot performance. Given the fact that most classes are in few-shot split, our method is the only one that increase the overall performance.

**Comparison among benchmarks.** From CIFAR-10-SSLT to ImageNet-SSLT and iNaturalist2018-SSLT, the datasets have more and more classes and few-shot classes. In result, they are more and more challenging. This challenge makes Pseudo-Label method ineffective. From CIFAR-10-SSLT to ImageNet-SSLT, the shortcoming first appears in many-shot splits. On ImageNet-SSLT, Pseudo-Label improves the few-shot performance with a sacrifice of many-shot performance. Our method is more robust to this difficulty. It keeps the many-shot performance while improves the few-shot performance. On iNaturalist2018-SSLT, the Pseudo-Label improvement on few-shot split also dis-

appears, and the drop on many-shot is big. Our method, however, can still improve the few-shot performance and control the drop of many-shot compared to the baseline.

All of these results show that semi-supervised long-tailed recognition is a challenging problem. Given the fact that this problem follows the natural workflow of data collecting, we believe it deserves more attention in the literature.