Separated and Independent Contrastive Learning on Labeled and Unlabeled Samples: Boosting Performance on Long-tail Semi-supervised Learning

Dongyoung Kim[1]
kimdongyoung0218@hallym.ac.kr
Won-Sook Lee[2]
wslee@uottawa.ca
Jeong-Gun Lee[†1]
jeonggun.lee@hallym.ac.kr

[1] Department of Computer Engineering, Hallym University, Chuncheon, South Korea.
[2] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada.

Abstract

Conventional semi-supervised learning (SSL) encounters challenges in effectively addressing issues associated with long-tail problems, primarily stemming from imbalances within a dataset. Previous semi-supervised approaches incorporating contrastive learning relied on unlabeled samples to apply the contrastive learning method. Consequently, to identify positive samples from unlabeled ones, they needed to make pseudo-labels, but inaccurate pseudo-labels lead to confirmation bias toward majority classes in long-tail datasets. Therefore, we try to obtain meaningful information from labeled samples which include accurate labels. In this paper, we introduce **S**eperated **I**ndependent **C**ontrastive **S**emi-**S**upervised **L**earning (**SICSSL**) for long-tail, which leverages a supervised contrastive learning approach for labeled samples and unlabeled samples separately and independently to enhance performance. In our experiments, employing labeled samples for contrastive learning yields superior performance compared to the contrastive learning using only unlabeled samples.

1 Introduction

Recently, semi-supervised learning in deep learning has become a solution to decrease the need of time-consuming and labor-intensive labeling. The semi-supervised learning method incorporates a substantial number of unlabeled samples alongside a restricted set of labeled samples for training deep learning models. Typical semi-supervised learning has been developed mainly under the assumption of uniform data distribution among class samples. In reality, this assumption is often invalid, as real-world datasets commonly exhibit a ‘long-tail’ problem. The long-tail problem arises when there is a large gap between the number

† Corresponding author.

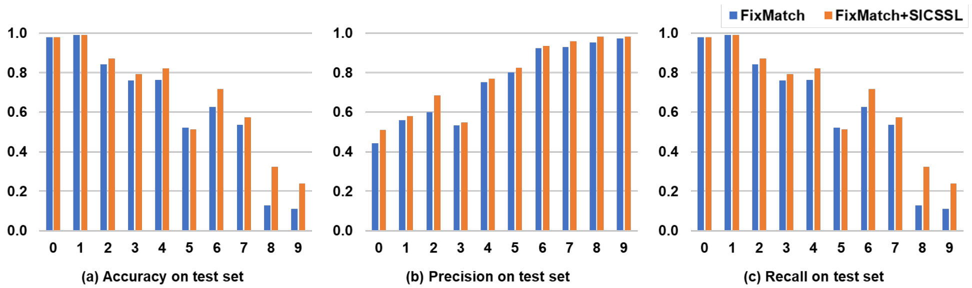

Figure 1: Experimental results on a CIFAR-10 public dataset. The X-axis represents the index of the class and the Y-axis represents accuracy, precision and recall with each corresponding to figures (a), (b) and (c), respectively. The blue bar shows the performance results of the original FixMatch and the orange bar shows the results of the proposed method. When applying our proposed method to FixMatch, the model shows the performance enhancement on all the class indices (x-axis shows the class indexes: lower is major and higher is minor). Note that the proposed method obtains noticeable improvements in recall for the minor classes (indices 8 and 9).

of data samples in the majority class and the minority class. Consequently, employing standard semi-supervised learning methods to train the model for long-tail problems can result in a 'confirmation bias' toward the majority classes. As a response to this challenge, numerous researchers have proposed several methods recently to address long-tail semi-supervised learning tasks[14, 20, 23, 25, 32].

Contrastive learning has been introduced as a means to train models through representation learning without relying on a classifier layer [6, 7, 13, 15, 16, 30]. While supervised contrastive learning (SupCon) [19] is a well-known approach that incorporates label information for training to utilize more positive samples than conventional unlabeled contrastive learning approaches. SupCon tends to exhibit confirmation bias toward majority classes in long-tail datasets, primarily due to the lack of positive samples in the minority class [24]. Class-aware Contrastive Semi-Supervised Learning (CCSSL) [33] has predominantly focused on only adapting a contrastive methodology to unlabeled samples. Therefore, in long-tail semi-supervised learning, contrastive learning for labeled samples has gathered comparatively less attention in the existing research landscape.

In Fig. 1, FixMatch obtains high recall and low precision on the major class and low recall and high precision on minor class because of confirmation bias toward major class. The main goal of imbalanced semi-supervised learning is to increase recall and precision on the minor class and the major class, respectively. In this paper, to obtain better performance on precision and recall, simultaneously, we propose **S**eparated **I**ndependent **C**ontrastive **S**emi-**S**upervised **L**earning for Long-Tail called **SICSSL**. In the proposed SICSSL, two separate and independent contrastive loss terms are used for the disjoint labeled and unlabeled samples. The added separated contrastive loss terms on labeled samples and unlabeled samples can lead to the prevention of overfitting on the classifier layer while working as a regularizer. As shown in Fig. 1, when adopting the proposed SICSSL in FixMatch, the model achieves improved performance. Particularly, in the minor classes such as the class indexes of 8 and 9, the model obtains higher accuracy than that without the SICSSL. Note that the adoption of the SICSSL leads to enhanced recall in minority classes and enhanced precision in majority classes.

The proposed SICSSL is simple and effective, requiring only the addition of a multi-layer perceptron (MLP) layer for efficient model training by incorporating supervised contrastive learning. The simplicity allows our method to seamlessly integrate with any semi-supervised learning architecture such as ABC [23] and ACR [32].

2 Related Works

• **Semi-supervised Learning :** Generally, supervised learning requires a large number of labeled data samples to achieve meaningful accuracy performance. However, collecting a sufficiently large number of labeled samples to train a model is time-consuming and labor-intensive. To obtain satisfactory performance without enough labeled samples, semi-supervised learning has been proposed to use not only a small amount of labeled samples but also a large number of unlabeled data samples for training a model. The main idea of semi-supervised learning is to use pseudo-labels and consistency loss function for employing unlabeled samples.

MixMatch [3] and ReMixMatch [2] use soft pseudo-labels for unlabeled samples. They use distribution alignment and sharpening to generate pseudo-labels. FixMatch [29] is a popular method in semi-supervised learning and they use one-hot pseudo-labels for unlabeled samples. On the other hand, FlexMatch [34] and FreeMatch [31] use an adaptive threshold for each class to employ the unlabeled samples more effectively so that a model can utilize diverse unlabeled samples compared to previous semi-supervised approaches.

• **Long-tail Problem :** Recently, long-tail problems have gained attention in many applications since the long-tail problems are unavoidable in real-world scenarios. The primary challenge associated with long-tail problems is the confirmation bias toward majority classes [4, 12, 18]. Re-weighting [10, 27] and re-sampling [1, 5, 17, 21, 26] are typical methods to improve performance in long-tail problems. DARP [20] and DASO [25] employ a pseudo-label refinement method for long-tail semi-supervised learning tasks to reduce incorrect pseudo-label assignments and ensure improved model performance. ABC [23] employs a masking strategy to mitigate confirmation bias for both labeled and unlabeled samples. Adaptive consistency regularizer (ACR) [32] utilizes two-branch network and adaptive adjustment method to refine pseudo-labels.

• **Contrastive Learning :** Contrastive learning is a powerful technique for learning informative representations from unlabeled data. MoCo [16] and SimCLR [6] propose to utilize representation learning on unlabeled samples for training a model. MoCo uses a momentum encoder and a queue memory bank to make different latent features for comparison with anchor features. Different from MoCo, SimCLR utilizes negative samples only from the mini-batch stage without a momentum encoder and queue. SupCon [19] uses the supervised manner in contrastive learning to utilize more positive samples compared to unsupervised contrastive learning methods. TSC [24] uses a target cluster to solve the confirmation bias when using supervised contrastive learning in long-tail recognition problems. CCSSL [33] employs a contrastive learning approach in semi-supervised learning. The method utilizes supervised contrastive learning for unlabeled samples through pseudo-labeling.

3 Motivation

In the CCSSL[33], authors utilize contrastive learning only on 'unlabeled' samples to improve the performance. In addition, to control the influence of contrastive learning loss on the unlabeled samples, they reflect re-weighting approach in a contrastive loss term. The main problem of CCSSL in imbalanced semi-supervised learning is that CCSSL uses inaccurate pseudo-labels to determine the 'positive samples'. In consequence, in the imbalanced semi-supervised learning, the probability of generating inaccurate pseudo-labels becomes higher compared to the uniform semi-supervised learning and these inaccurate pseudo-labels can lead to performance degradation. To better exploit the benefit of contrastive learning, in this paper, we focus on how to utilize 'labeled samples' for the contrastive learning.

Intuitively, the supervised contrastive learning can lead to confirmation bias toward majority classes due to the gap in sample sizes between majority and minority classes. Through the experiments, however, we show that SICSSL which utilizes labeled samples for contrastive learning can lead to improving performance.

The main difference between conventional contrastive learning approaches for an imbalanced problem and our approach is the role of the contrastive learning: Previous SupCon [19] and TSC [24] use contrastive learning for pre-training a backbone network. Then, they train the model using linear probing for a classifier layer or using the fine-tune of all the layers. In our case, however, we use the contrastive learning loss term as a 'regularizer' of improving the generalization ability of a model by reducing overfitting since the conventional semi-supervised learning loss such as a FixMatch loss term and the potentially competing contrastive loss term are used simultaneously for training.

4 Preliminary

• **Problem setting :** In the proposed method, we use two datasets that are labeled ($\mathcal{D}^l \in \{(x_i^l, y_i^l)\}_{i=1}^N$) and unlabeled ($\mathcal{D}^u \in \{(x_j^u)\}_{j=1}^M$) dataset. N and M are the number of labeled samples and unlabeled samples, respectively. In addition, N_1 and M_1 are the number of samples in major class on labeled samples and unlabeled samples, respectively. For the i-th labeled sample, it has ground-truth information y_i^l which is an one-hot vector.

• **Semi-supervised learning :** Many existing semi-supervised learning algorithms utilize supervised loss for labeled samples and consistency loss for unlabeled samples, respectively. The goal of those algorithms is to minimize the supervised loss and consistency loss. In this paper, we adopt a FixMatch approach to solve the semi-supervised learning problem. The loss functions can be defined as follows:

$$\mathcal{L}_{CE} = -\frac{1}{B}\sum_{i=1}^{B} \ell_{CE}(C(E(x_i^l)), y_i^l) \tag{1}$$

$$\mathcal{L}_{con} = -\frac{1}{\mu B}\sum_{j=1}^{\mu B} \mathbb{1}(\max(C(E(\alpha(x_j^u)))) \geq \tau) \cdot \ell_{CE}(C(E(\mathcal{A}(x_j^u))), \hat{y}_j^u) \tag{2}$$

$$\mathcal{L}^{back} = \mathcal{L}_{CE} + \lambda_{con} \cdot \mathcal{L}_{con} \tag{3}$$

In Eq. (1) - (3), $\mathcal{L}_{CE}$ denotes a cross-entropy loss and $\mathcal{L}_{con}$ is a consistency loss. $\mathcal{A}(\cdot)$, $\alpha(\cdot)$, $C(\cdot)$ and $E(\cdot)$ are strong augmentation, weak augmentation, the classifier layer and encoder backbone layer, respectively. Pseudo-label ($\hat{y}_j^u$) is a one-hot vector generated by applying an argmax function to $C(E(\alpha(x_j^u)))$ for calculating consistency loss.

• **InfoNCE for contrastive learning :** In contrastive learning approaches, selecting positive samples is crucial to improve performance. MoCo and SimCLR do not use labeled information so that they use an augmentation strategy to derive positive samples from an anchor sample. Consequently, the number of positive samples is proportional to the number of augmentations.

On the other hand, in supervised contrastive learning, to determine the positive samples from an anchor sample, label information from other samples can be used. In this situation, the number of positive samples depends on the number of augmentations and the number of the samples in a same class. The loss terms for self-supervised contrastive learning ($\mathcal{L}^{self}$) and supervised contrastive learning ($\mathcal{L}^{sup}$) can be described as follows:

$$\mathcal{L}^{self} = -\sum_{i \in I} \log\left(\frac{\exp(z_i \cdot z_i^p / T)}{\sum_{a \in A_{(i)}} \exp(z_i \cdot z_a / T)}\right) \tag{4}$$

$$\mathcal{L}^{sup} = -\sum_{i \in I} \frac{1}{|P_{(i)}|} \sum_{p \in P_{(i)}} \log\left(\frac{\exp(z_i \cdot z_i^p / T)}{\sum_{a \in A_{(i)}} \exp(z_i \cdot z_a / T)}\right) \tag{5}$$

In Eq. (4) and Eq. (5), z_i, z_i^p and z_a are a vector from an anchor sample, positive sample and all samples in mini-batch excluding anchor sample, respectively. T is a scaling factor to make smoothing or sharpening of similarity values. I, $P_{(i)}$ and $A_{(i)}$ are the set of the indices for mini-batches, the set of latent vectors from positive samples in the i-th mini-batch and the set of latent vectors from all samples in the i-th mini-batch excluding the anchor sample (z_i), respectively. In self-supervised contrastive learning, the number of positive samples is equal to "the number of augmentations - 1" while it is determined by both labeled class information and augmentations in supervised contrastive learning.

• **Augmentation strategy for labeled samples :** Traditionally, prior works solely utilize unlabeled samples for contrastive learning, thereby omitting the need for strong augmentation in labeled samples. However, our experiments reveal that employing supervised contrastive learning on labeled samples with strong augmentation significantly enhances performance compared with the case where this augmentation is not used.

To provide the alternative perspective in comparison with the case of using only weak augmentation, we adopt a policy of strong augmentation, mirroring the approach used with unlabeled samples. Leveraging label information in labeled samples allows us to identify more positive samples. Consequently, we can employ the original supervised contrastive technique to train the model without relying on pseudo-labels.

• **Seperated and independent supervised contrastive learning for semi-supervised learning:** In this paper, we use two independent contrastive loss terms on labeled and unlabeled samples. For unlabeled samples, we adopt a CCSSL contrastive loss term to re-weight the samples with inaccurate pseudo-labels. Our total loss term can be described as follows:

$$\mathcal{L}_l^{sup} = -\sum_{i \in I^l} \frac{1}{|P_{(i)}^l|} \sum_{p \in P_{(i)}^l} \log\left(\frac{\exp(z_i \cdot z_i^p / T)}{\sum_{a \in A_{(i)}} \exp(z_i \cdot z_a / T)}\right) \tag{6}$$

$$\mathcal{L}_u^{sup} = -\sum_{j \in I^u} \frac{1}{|P_{(j)}^u|} \sum_{\hat{p} \in P_{(j)}^u} w_{j,p} \cdot \log\left(\frac{\exp(z_j \cdot z_j^{\hat{p}} / T)}{\sum_{a \in A_{(j)}} \exp(z_j \cdot z_a / T)}\right) \tag{7}$$

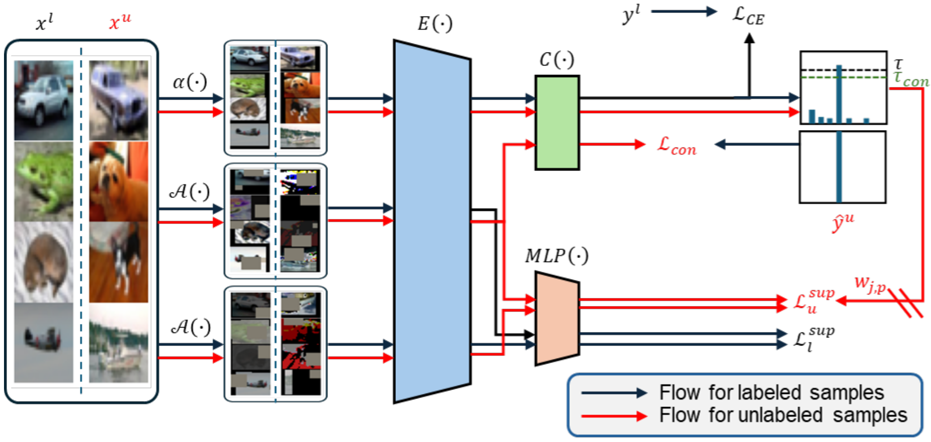

Figure 2: Framework for our proposed SICSSL. To utilize the contrastive learning approach, we adopt a multi-layer perceptron to generate a latent vector. The main component of SICSSL is that the contrastive learning loss terms are calculated separately and independently for labeled and unlabeled samples.

$$w_{j,p}^{mask} = \begin{cases} 1 & \text{if } j = p, \\ 1 & \text{if } z_j \text{ and } z_p \text{ are from the samples with same pseudo-label} \\ & \text{and } q_j > \tau_{con} \text{ and } q_p > \tau_{con}, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

$$w_{j,p} = \begin{cases} q_j \cdot q_p \cdot w_{j,p}^{mask} & \text{if } j \neq p, \\ w_{j,p}^{mask} & \text{otherwise.} \end{cases} \tag{9}$$

$$\mathcal{L}^{total} = \mathcal{L}^{back} + \lambda_l^{sup} \cdot \mathcal{L}_l^{sup} + \lambda_u^{sup} \cdot \mathcal{L}_u^{sup} \tag{10}$$

In Eq. (6), (7) and (8), j and p are the indices of an anchor and a positive sample, respectively, in a mini-batch for unlabeled samples. I^l, and I^u are the sets of the indices for mini-batches on labeled samples and unlabeled samples. $P_{(i)}^l$ and $P_{(j)}^u$ are the sets of indices for the latent vectors from positive samples in i-th mini-batch for labeled samples and j-th mini-batch for unlabeled samples, respectively. q_j and q_p are softmax outputs of a weakly augmented anchor and a weakly augmented positive sample. The softmax output is used as a confidence measure to determine the inclusion of the given input sample in loss evaluation, according to a confidence threshold (τ_{con}).

In Eq. (10), there is no interaction between the two contrastive learning loss terms for labeled samples and unlabeled samples. Therefore, there is no pull and push between labeled samples and unlabeled samples. With the loss term given in Eq. (6), we try to utilize accurate labeled samples for contrastive learning, which can obtain the accurate positive samples.

In conventional contrastive learning in supervised learning tasks, two-step training is used: (1) they pre-train the model using contrastive learning, and (2) they fine-tune the model using supervised learning with cross-entropy loss term. However, in our method, the contrastive learning loss term is used simultaneously with loss terms for semi-supervised

Table 1: Comparison between previous works and proposed methods with various imbalanced ratios (γ). The symbol, '*', indicates an unstable training result.

Algorithm	CIFAR-10-LT ($N_1 = 1000, M_1 = 4000$)			CIFAR-100-LT ($N_1 = 200, M_1 = 300$)		
	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$	$\gamma = 20$	$\gamma = 40$	$\gamma = 60$
FixMatch [29]	72.41±1.19	62.72±1.56	59.08±1.15	48.48±1.02	44.03±0.69	41.50±0.41
w. DASO [25]	75.64±0.42	65.90±2.90	62.57±1.44	49.26±0.45	44.52±0.38	40.70±1.15
w. CCSSL [33]	75.18±0.41	67.16±1.34	62.04±1.44	50.19±0.84	45.73±0.46	42.96±0.67
w. SICSSL(Ours)	**76.86±0.44**	**68.37±0.46**	**63.45±0.56**	**53.34±0.13**	**48.37±0.27**	**45.60±0.27**
ABC [23]	85.20±0.26	80.63±1.04	77.70±0.48	54.58±0.49	49.33±0.45	46.51±0.45
w. CCSSL [33]	85.48±0.68	80.78±0.89	**78.01±0.41**	54.64±0.22	49.47±0.37	46.81±0.40
w. SICSSL(Ours)	**85.86**±0.36	**81.57±0.72**	77.46±0.97	**57.44±0.28**	**51.38±0.23**	**48.27±0.49**
ACR [32]	82.91±0.44	78.93±1.13	75.93±0.90	54.20±0.14	49.73±0.43	47.30±0.62
w. CCSSL [33]	83.04±0.32	79.49±1.04	76.73±1.53	55.53±0.05	50.69±0.69	*17.21±22.87
w. SICSSL(Ours)	**84.96±0.38**	**81.63±0.93**	**78.29±0.94**	**56.66±0.26**	**52.17±0.41**	**49.38±0.33**

Table 2: Comparison between previous works and proposed methods with various settings on the number of labeled samples and unlabeled samples on the major class (N_1 and M_1).

Algorithm	CIFAR-10-LT ($\gamma = 100$)			CIFAR-100-LT ($\gamma = 20$)		
	$N_1 = 250$ $M_1 = 4750$	$N_1 = 500$ $M_1 = 4500$	$N_1 = 1000$ $M_1 = 4000$	$N_1 = 100$ $M_1 = 400$	$N_1 = 150$ $M_1 = 350$	$N_1 = 200$ $M_1 = 300$
FixMatch [29]	52.32±4.09	56.72±1.92	62.72±1.56	42.30±0.66	46.91±0.90	48.48±1.02
w. DASO [25]	**57.74±0.92**	61.85±2.44	65.90±2.90	42.96±0.25	47.27±0.26	49.26±0.45
w. CCSSL [33]	56.98±2.39	**62.72±1.36**	67.16±1.34	44.49±0.24	47.85±0.26	50.19±0.84
w. SICSSL (Ours)	53.28±1.67	60.68±0.21	**68.37**±0.46	**47.08±0.88**	**50.79±0.29**	**53.34±0.13**
ABC [23]	72.69±2.08	77.10±0.41	80.63±1.04	46.08±0.62	51.69±0.42	54.58±0.49
w. CCSSL [33]	**74.18±2.23**	76.86±1.64	80.78±0.89	47.60±0.51	52.37±0.17	54.64±0.22
w. SICSSL(Ours)	74.03±3.19	**77.89±0.46**	**81.57±0.72**	**50.53±0.48**	**55.76±0.40**	**57.44±0.28**
ACR [32]	67.43±0.96	72.95±0.86	78.93±0.90	47.54±0.36	52.02±0.35	54.20±0.14
w. CCSSL [33]	**68.08±1.32**	73.97±1.41	79.49±1.04	49.03±0.54	52.02±0.35	55.53±0.05
w. SICSSL(Ours)	66.79±1.43	**77.02±1.18**	**81.63±0.93**	**49.89±1.06**	**54.25±0.63**	**56.66±0.26**

learning. In this case, the contrastive loss term works as a 'regularization' term and it suppresses confirmation bias in a 'classifier layer'.

Figure 2 illustrates the framework of SICSSL. Unlike the conventional SupCon [19], the model utilizes the supervised contrastive loss term on labeled samples simultaneously and independently with the semi-supervised loss term.

In the ablation study, we try to compare the two types of contrastive learning approaches on labeled and unlabeled samples. With obtained experimental results, we show that the separated and independent contrastive learning approach is more suited for imbalanced semi-supervised learning. For the trainings, λ_l^{sup} and λ_u^{sup} are set to 1.0 for CIFAR-10, CIFAR-100 and ImageNet-127 public datasets. For STL-10 dataset, λ_l^{sup} and λ_u^{sup} are set to 0.1.

5 Experiments

• **Dataset :** We use well-known public datasets to assess performance: CIFAR-10, CIFAR-100 [22], STL-10 [8] and imagenet-127 [28] datasets with long-tail distribution. We partition the training dataset into labeled and unlabeled samples. Here, γ, N_1 and M_1 denote the imbalanced ratio, number of labeled samples on the major class and number of unlabeled samples on the major class, respectively. The imbalance ratio can be calculated by $\gamma = \frac{\#\ of\ the\ major}{\#\ of\ the\ minor}$.

• **Model architecture :** A Wide ResNet-based (WRN) CNN architecture is employed along with the FixMatch to generate pseudo-labels for unlabeled samples. An MLP layer is added to generate latent vectors ($z_i, z_j, z_i^p, z_j^{\hat{p}}$ and z_a) for calculating contrastive learning loss terms in Eq. (6) and (7).

Table 3: Comparison between previous works and proposed methods on STL-10 public dataset. The symbol, '†', indicates that SICSSL obtains lower performance compared to DASO. However, SICSSL produces better results than FixMatch with DASO when SICSSL is combined with DASO.

Algorithm	$\gamma_l = 10, \gamma_u = N/A$		$\gamma = 20, \gamma_u = N/A$	
	$N_1 = 150$ $M_1 = 100k$	$N_1 = 450$ $M_1 = 100k$	$N_1 = 150$ $M_1 = 100k$	$N_1 = 450$ $M_1 = 100k$
FixMatch [29]	56.08±1.83	64.54±0.94	48.19±1.16	56.38±2.80
w. DASO [25]	**63.75±0.08**	68.44±1.19	**56.00±0.46**	**64.88±1.59**
w. CCSSL [33]	56.82±1.94	71.42±0.79	49.56 ± 0.71	58.07±0.89
w. SICSSL(Ours)	†57.81±2.31	**72.29±0.93**	†47.94±1.63	†60.13±1.81
DASO [25]	63.75±0.08	68.44±1.19	**56.00±0.46**	64.88±1.59
w. SICSSL(Ours)	**64.27±0.29**	**71.13±0.56**	55.90±1.81	**66.30±0.60**
ABC [23]	64.74±0.19	71.93±0.44	60.36±0.80	68.48±1.85
w. CCSSL [33]	65.83±0.02	72.24±1.39	60.70±0.04	69.32±1.49
w. SICSSL(Ours)	**67.04±0.29**	**73.02±0.39**	**60.91±0.01**	**72.07±0.71**
ACR [32]	64.62±1.00	72.92±0.83	61.41±0.72	69.38±2.88
w. CCSSL [33]	64.33±1.18	72.91±0.14	**63.42±1.26**	69.92±2.13
w. SICSSL(Ours)	**66.08±0.41**	**74.68±0.28**	62.90±0.96	**72.45±0.60**

Table 4: Influence of SICSSL in ImageNet-127 public dataset on FixMatch and ACR. The symbol, '*', indicates unstable training result.

Algorithm	Image Size	
	32×32	64×64
FixMatch [29]	44.88	46.85
w. DASO [25]	45.18	47.04
w. CCSSL [33]	44.58	*23.61
w. SICSSL(Ours)	**46.41**	**49.12**
ACR [32]	41.20	45.44
w. CCSSL [33]	*21.90	44.91
w. SICSSL(Ours)	**42.95**	**47.90**

• **Training strategy :** An Adam optimizer is used with a 0.002 learning rate without a scheduling algorithm. Training is conducted on the dataset for 250,000 iterations (500 epochs × 500 iterations for each epoch). To enhance the performance, we also use the exponential moving average (EMA) with the momentum of 0.999. The numbers of labeled samples and unlabeled samples in each batch are 64 and 64, respectively. The confidence thresholds, τ in Eq. (2) and τ_{con} in Eq. (8), for deciding when to use pseudo-labels for the consistency loss and the contrastive loss are set at 0.95 and 0.9, respectively. For the augmentation strategy, we apply both weak and strong augmentation techniques to both labeled and unlabeled samples.

Typically, strong augmentation is used exclusively for unlabeled samples to enhance features through consistency loss. However, in this work, supervised contrastive learning is applied to labeled samples and strong augmentation is used also in labeled samples in order to generate diverse views from the same anchors. All evaluations are performed independently three times using different random seeds to ensure the reliability of the results.

6 Results

Table 1 and 2 present accuracy comparisons between the proposed method and previous works while varying "imbalanced ratio" (γ) and the "number of samples on major class in labeled and unlabeled dataset" (N_1 and M_1) on CIFAR-10-LT and CIFAR-100-LT. We adjust the imbalanced ratio and labeled ratio ($\frac{N_1}{N_1+M_1}$) to control the difficulty of a classification task. A higher imbalanced ratio and lower labeled ratio make the problem more challeng-

Table 5: The influence of separated and independent contrastive learning approaches in long-tail semi-supervised tasks in CIFAR-100-LT datasets ($\gamma = 20$, $N_1 = 200$ and $M_1 = 300$).

Dataset	Algorithm	Performance Measure		
		Acc.	Major Acc.	Minor Acc.
CIFAR-100-LT	Baseline [23]	54.58±0.49	68.08±0.13	41.08±0.87
	w. CCSSL [53]	54.64±0.22	67.10±0.49	41.37±0.05
	w. UniCSSL	57.10±0.23	**69.81**±**0.30**	44.39±0.75
	w. SICSSL (Ours)	**57.44**±**0.28**	69.61±0.50	**45.26**±**0.06**

ing to achieve higher accuracy. Our SICSSL can be utilized easily with previous methods (FixMatch, ABC and ACR) and obtains higher performance than those without SICSSL.

Also, we train the models on STL-10 (Table 3) and ImageNet-127 (Table 4) datasets to validate the further extensibility of the SICSSL. In STL-10 dataset, the SICSSL leads to performance enhancement on various imbalanced settings and size of labeled samples. Note that SICSSL can not outperform the DASO when the SICSSL is adopted directly in FixMatch. However, the benefit of SICSSL is that SICSSL can be adapted easily in various semi-supervised learning. So, SICSSL can be fused with DASO and in this case SICSSL leads to performance enhancement in the most cases as shown in "DASO w. SICSSL" in Table 3. In the ImageNet-127 dataset with 32×32 and 64×64 image sizes, the proposed SICSSL obtains better performance than those without the SICSSL in FixMatch and ACR. When CCSSL is adopted in FixMatch and ACR, the model obtains unstable results as shown in Table 1 and 4 (marked by '*'). From the experiments, we expect that the SICSSL provides more stable training convergence than CCSSL when used for the contrastive learning with imbalanced semi-supervised learning tasks.

• **Discussion :** In Table 1, 2 and 3, the model obtains lower performance compared to those without SICSSL approach when the number of labeled samples is extremely small (e.g., $N_1 = 250$ in CIFAR-10 and $N_1 = 150$ in STL-10 dataset). In CIFAR-10-LT with $N_1 = 250$, the number of samples in the minor class is just '2' ($N_{10} = 2$). In this extreme case, the probability of causing confirmation bias toward major class is high even SICSSL is used. In STL-10 dataset, the probability of having confirmation bias is higher than in CIFAR-10 dataset since the gap in sample size between labeled and unlabeled samples is huge. From the experiments, we observed that the superiority of our approach cannot be guaranteed in such extreme cases.

When we train the model using FixMatch with previous methods and SICSSL, SICSSL obtains lower performance compared to DASO († in Table 3). Since the SICSSL does not use a pseudo-label refinement approach, SICSSL can be utilized with DASO. When adopt SICSSL in DASO, the model outperforms in the most cases in STL-10 dataset without the case of $N_1 = 150$ and $\gamma = 20$. To prevent performance degradation, we have to search for suitable hyper-parameters in the case of having the extremely small number of labeled samples.

• **Influence of Separated Independent Contrastive Learning on Labeled and Unlabeled Samples :** In this work, we proposed a 'separated and independent' contrastive learning approach rather than using a unified contrastive learning where the contrastive learning is applied to an unified set of mixed labeled and unlabeled samples. In supervised contrastive learning, selecting accurate positive samples is important for developing a high accuracy model. Using the separated and independent approach, we can guarantee robust accurate contrastive learning on labeled samples which are not mixed with inaccurate positive samples from unlabeled samples. As shown in Table 5, SICSSL outperforms previous works and 'UniCSSL' (Unified Constrastive Semi-Supervised Learning) which employs an unified contrastive learning loss term on a mixed set of labeled and unlabeled samples. It is noteworthy

Table 6: The influence of different augmentation strategies in CIFAR-100-LT dataset ($\gamma = 20$, $N_1 = 200$ and $M_1 = 300$) with SICSSL. Strategy 1 (S-1) is to use strong augmentations in a semi-supervised learning for unlabeled samples and Strategy 2 (S-2) is to use SimCLR-based augmentations.

Augmentation Strategy	Performance Measure		
	Acc.	Major Acc.	Minor Acc.
S-1	**56.90±0.39**	**68.89±0.29**	**44.91±0.81**
S-2	55.16±0.30	67.20±0.28	43.12±0.83

that our SICSSL achieves higher performance in minority classes.

• **Augmentation Strategy on Labeled Samples for Contrastive Learning :** Typically, in contrastive learning approaches for vision tasks, three augmentation strategies, namely 'random crop and resize', 'color distortions', and 'Gaussian blur', are commonly employed to generate diverse views (used as augmentations in SimCLR). As shown in Table 6, the use of the strong augmentation strategy (S-1) yields better performance compared to the SimCLR-based augmentation strategy (S-2). Strong augmentation used in S-1 consists of Cutout [11], CTAugment [2] and RandAugment [9]. This is the same with the strong augmentation strategy used in training for unlabeled samples.

7 Conclusion

This paper proposed a model which exploits the effectiveness of supervised contrastive learning on the separated two sample sets, labeled samples and unlabeled samples, for semi-supervised learning on a long-tail imbalanced dataset. While previous researches in the semi-supervised learning with contrastive learning have predominantly concentrated on leveraging unlabeled samples, our experiments revealed that applying contrastive learning to labeled samples enhances performance in long-tail semi-supervised learning. We also investigated the impact of the contrastive learning strategies and the augmentation strategies on the accuracy performance in the ablation studies. Our results demonstrated that the proposed SICSSL produces better performance than previous work and it can be easily and orthogonally adapted with recently proposed imbalanced semi-supervised learning approaches, such as ABC and ACR.

8 Acknowledgement

This work is the results of a study on the "Leaders in INdustry-university Cooperation 3.0" Project, supported by the Ministry of Education and National Research Foundation of Korea (202403460001).

References

[1] Shin Ando and Chun Yuan Huang. Deep over-sampling framework for classifying imbalanced data. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part I 10*, pages 770–785. Springer, 2017.

[2] David Berthelot, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv preprint arXiv:1911.09785*, 2019.

[3] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32, 2019.

[4] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in neural information processing systems*, 32, 2019.

[5] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[7] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15750–15758, 2021.

[8] Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 215–223. JMLR Workshop and Conference Proceedings, 2011.

[9] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020.

[10] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277, 2019.

[11] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.

[12] Qi Dong, Shaogang Gong, and Xiatian Zhu. Imbalanced deep learning by minority class incremental rectification. *IEEE transactions on pattern analysis and machine intelligence*, 41(6):1367–1381, 2018.

[13] Lijie Fan, Sijia Liu, Pin-Yu Chen, Gaoyuan Zhang, and Chuang Gan. When does contrastive learning preserve adversarial robustness from pretraining to finetuning? *Advances in neural information processing systems*, 34:21480–21492, 2021.

[14] Yue Fan, Dengxin Dai, Anna Kukleva, and Bernt Schiele. Cossl: Co-learning of representation and classifier for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14574–14584, 2022.

[15] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

[16] 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*, pages 9729–9738, 2020.

[17] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. Learning deep representation for imbalanced classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5375–5384, 2016.

[18] Salman Khan, Munawar Hayat, Syed Waqas Zamir, Jianbing Shen, and Ling Shao. Striking the right balance with uncertainty. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 103–112, 2019.

[19] 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.

[20] 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. *Advances in neural information processing systems*, 33:14567–14579, 2020.

[21] Jaehyung Kim, Jongheon Jeong, and Jinwoo Shin. M2m: Imbalanced classification via major-to-minor translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13896–13905, 2020.

[22] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. *Technical report*, 2009.

[23] Hyuck Lee, Seungjae Shin, and Heeyoung Kim. Abc: Auxiliary balanced classifier for class-imbalanced semi-supervised learning. *Advances in Neural Information Processing Systems*, 34:7082–7094, 2021.

[24] Tianhong Li, Peng Cao, Yuan Yuan, Lijie Fan, Yuzhe Yang, Rogerio S Feris, Piotr Indyk, and Dina Katabi. Targeted supervised contrastive learning for long-tailed recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6918–6928, 2022.

[25] Youngtaek Oh, Dong-Jin Kim, and In So Kweon. Daso: Distribution-aware semantics-oriented pseudo-label for imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9786–9796, 2022.

[26] Seulki Park, Youngkyu Hong, Byeongho Heo, Sangdoo Yun, and Jin Young Choi. The majority can help the minority: Context-rich minority oversampling for long-tailed classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6887–6896, 2022.

[27] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *International conference on machine learning*, pages 4334–4343. PMLR, 2018.

[28] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115: 211–252, 2015.

[29] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020.

[30] Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*, pages 9929–9939. PMLR, 2020.

[31] Yidong Wang, Hao Chen, Qiang Heng, Wenxin Hou, Yue Fan, Zhen Wu, Jindong Wang, Marios Savvides, Takahiro Shinozaki, Bhiksha Raj, et al. Freematch: Self-adaptive thresholding for semi-supervised learning. *arXiv preprint arXiv:2205.07246*, 2022.

[32] Tong Wei and Kai Gan. Towards realistic long-tailed semi-supervised learning: Consistency is all you need. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3469–3478, 2023.

[33] Fan Yang, Kai Wu, Shuyi Zhang, Guannan Jiang, Yong Liu, Feng Zheng, Wei Zhang, Chengjie Wang, and Long Zeng. Class-aware contrastive semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14421–14430, 2022.

[34] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34: 18408–18419, 2021.