Interpolation-based Contrastive Learning for Few-Label Semi-Supervised Learning

Xihong Yang*, Xiaochang Hu*, Sihang Zhou, Xinwang Liu†, *Senior Member, IEEE*, En Zhu†

Abstract—Semi-supervised learning (SSL) has long been proved to be an effective technique to construct powerful models with limited labels. In the existing literature, consistency regularization-based methods, which force the perturbed samples to have similar predictions with the original ones have attracted much attention for their promising accuracy. However, we observe that, the performance of such methods decreases drastically when the labels get extremely limited, e.g., 2 or 3 labels for each category. Our empirical study finds that the main problem lies with the drift of semantic information in the procedure of data augmentation. The problem can be alleviated when enough supervision is provided. However, when little guidance is available, the incorrect regularization would mislead the network and undermine the performance of the algorithm. To tackle the problem, we (1) propose an interpolation-based method to construct more reliable positive sample pairs; (2) design a novel contrastive loss to guide the embedding of the learned network to change linearly between samples so as to improve the discriminative capability of the network by enlarging the margin decision boundaries. Since no destructive regularization is introduced, the performance of our proposed algorithm is largely improved. Specifically, the proposed algorithm outperforms the second best algorithm (Comatch) with 5.3% by achieving 88.73% classification accuracy when only two labels are available for each class on the CIFAR-10 dataset. Moreover, we further prove the generality of the proposed method by improving the performance of the existing state-of-the-art algorithms considerably with our proposed strategy. The corresponding code is available at https://github.com/xihongyang1999/ICL_SSL.

Index Terms—Semi-supervised learning, contrastive learning, interpolation-based method, few-label.
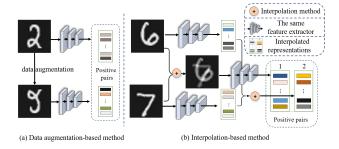

Figure 1: Illustration of the positive sample pair construction process. Different from the existing works which construct positive sample pairs with data augmentation, we construct positive sample pairs with interpolation operations. Specifically, given two unlabeled images, the integration of the sample embeddings '1' and the embedding of the sample integration '2' are acquired as a positive sample pair.

I. INTRODUCTION

IN recent years, machine learning has developed rapidly and achieved remarkable performance in many fields like, image classification [1], [2], object detection [3], [4], semantic segmentation [5], [6], and clustering [7]–[14]. Convolutional neural networks (CNNs) have attracted the attention of many researchers. The success of most of these deep neural networks depends heavily on a large number of high-quality labeled datasets [2], [15], [16].

However, collecting labeled data can consume a lot of resources which is un-affordable to countless everyday learning demands in modern society. Therefore, deep learning algorithms which can achieve appropriate performance with

tractable supervision have been a hot research spot in recent years. Specifically, deep semi-supervised learning (SSL) algorithms, which seek to improve the performance of deep learning models on datasets with only limited labeled data by leveraging large amounts of unlabeled data, are an important branch in this family. This has led to a plethora of SSL methods designed for various fields [17]–[23].

Among all the deep semi-supervised learning algorithms, consistency regularization based methods treat the original input and its augmented version as positive pairs, which is a form of contrastive learning [24]–[31]. These consistency regularization-based methods follow a common assumption that ever after data augmentation, the classifier could output the same class probability for an unlabeled sample, which means data augmentation will not change the semantic. The input image should be more similar to its augmented version than other images. Under this assumption, researchers perturb the input samples by conducting data augmentation to generate similar samples of the original data.

The mentioned algorithms have contributed remarkable performance improvement to improve the learning accuracy when only a few labeled data are available. However, we observe that, when the number of labeled data gets extremely small, e.g., 2 to 3 labels for each category, the performance of the existing algorithms would drop drastically. For example, to the CIFAR-10 dataset whose scale for training samples is $50,000$ and 10 categories, the performance of the state-of-the-art algorithm MixMatch [28] can achieve the top-1 accuracy of 86.47% when 250 labeled data is available. Nevertheless,

X. Yang, X. Liu and E. Zhu are with School of Computer, National University of Defense Technology, Changsha, 410073, China. (E-mail: {yangxihong, xinwangliu, enzhu} @nudt.edu.cn)

X. Hu and S. Zhou are with College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China.(E-mail: huxiaochang17@nudt.edu.cn, sihangjoe@gmail.com)

*: Equal contribution. †: Corresponding author.

Table I: Classification accuracy of two state-of-the-art semi-supervised algorithms, i.e., MixMatch [28] and Mean-Teacher [27], on CIFAR-10 dataset with $30, 40, 250, 500$ and 1000 labels.

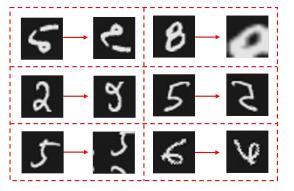| Method		30	40	250	500	1000
MixMatch	[28]	50.10	59.08	86.47	89.33	90.79
Mean-Teacher	[27]	24.51	24.93	52.49	70.15	80.12

Figure 2: Representative examples of semantic information drift caused by inappropriate data augmentation of MINIST samples.

Figure 3: Illustration of classification results with different data augmentations on the MINIST dataset.

the performance of the same algorithm drops to 50.10% when only 30 labeled samples are available. The similar phenomenon happens to the Mean-Teacher [27] algorithm whose performance drop by more than a half when the label number decreases from 250 to 30. More experimental results can be found in Table I.

According to our analysis, one of the main reasons that cause large performance decrease lies with the semantic information drift during data augmentation. Taking the samples in the MINIST dataset for example, when the vertical flip is applied to the samples, the labels of "6"s and "9"s, "2"s and "5"s can easily get changed. This would challenge the rationality of the information consistency assumption of existing methods. This problem could be alleviated when relatively abundant label information is available. However, when the label information is extremely lacked, the performance of the corresponding algorithms could decrease a lot.

In this paper, to solve the problem of semantic information drift caused by data augmentation-based positive sample pair construction, we propose a novel interpolation-based positive sample pair construction fashion. Generally, our design roots from the observation that the margin of decision boundaries would get larger if the prediction of the network could change linearly [32], [33]. Under the circumstance of semi-supervised learning, when the label is extremely limited, we seek to improve the discriminative capability of the network by forcing the embedding of the network to change linearly. Specifically, given two unlabeled images, on the one hand, we embed the samples separately into the latent space. On the other hand, we conduct image-level interpolation for an integrated image and do the embedding with the same network. Then, by combining the embedding of the interpolated images with
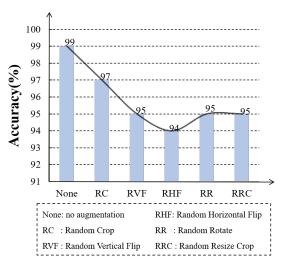
the interpolation of the embeddings, we construct a positive sample pair. In our setting, the negative sample pairs are the embedding pair of different samples. By forcing the positive sample pairs to be close to each other in the latent space and the negative sample pairs to get far away from each other, we enlarge the margin of decision boundaries, thus improving the performance of the algorithm. To achieve the goal, we further propose a novel contrastive learning-based loss function to guide the network for better learning. We name the resultant algorithm Interpolation Contrastive Learning Semi-Supervised Learning (ICL-SSL).

The main contributions of this paper are listed as follows:

- We find that semantic information drift is one of the main problems that cause the performance of existing consistency regularization-based semi-supervised algorithms to decrease drastically when extremely limited labeled data is provided.
- We propose an interpolation-based positive sample construction method and a novel contrastive loss function to solve the problem and improve the learning accuracy.
- Our experimental results on the benchmark datasets verify the superior performance of the proposed algorithms against the state-of-the-art algorithms. We also show the generality of our proposed algorithm by enhancing the performance of the existing advanced algorithms steadily with our method.

II. RELATED WORK

In this section, we first define the main notations and then review several semi-supervised learning (SSL) methods related to our method ICL-SSL.

A. Notations Definition

Given a dataset $\mathcal{D} = \mathcal{X} \cup \mathcal{U}$, where $\mathcal{X} = \{(x_1, y_1), \cdots, (x_m, y_m)\}$ is an labeled sub-dataset, $\mathcal{U} = u_{m+1}, \cdots, u_{m+n}\}$ is a unlabeled sub-dataset, $n \gg m$ and y_m is encoded by one-hot, we define a classification model as

Table II: Notation summary

Notations	Meaning
b	Bathsize
C	The number of classes
$\mathcal{X}^B = (x_i, y_i)$	Labeled sub-dataset with b
$\mathcal{U}^B = \{u_1, \cdots, u_b\}$	Unlabeled sub-dataset matrix with b
$x, u \in R^C \times R^H \times R^W$	Input samples
$y \in \{0, 1\}^C$	Label with C classes encoded by one-hot
$q \in R^C$	Predicted category probability distribution
$f(\cdot)$	The encoder network
$h(\cdot)$	The classifier
$g(\cdot)$	The projector head
$z \in R^D = g(f(x)) = F(x)$	Normalized low-dimensional representation

$p(y|x; \theta)$, which outputs a distribution over class labels y for an input x with parameters θ. For the model $p(y|x; \theta)$, it is concatenated by a encoder network $f(\cdot)$ and a classification head $h(\cdot)$ before softmax function. Meanwhile, after the encoder network $f(\cdot)$, we set a projection head $g(\cdot)$, outputting the normalized low-dimensional representation $z = g(f(\cdot))$. To simplify, $F(\cdot)$ is defined as $g(f(\cdot))$. For more detailed definitions, please refer to Table II.

B. Contrastive Learning

Thanks to leveraging unlabeled data for model training, contrastive learning attracts much attention of some researchers and becomes a hot spot recently [34]–[37]. It is a widely adopted form of self-supervised learning [34], [38]–[42], which can be used to optimize the task of instance discrimination. Instead of training a classification, contrastive learning is to maximize the similarities of positive pairs and minimize the similarities of negative pairs. It is important to learn the invariance with different views generated by data augmentations. The contrastive learning loss on unlabeled data can be described as follows:

$$-\log \frac{exp(F(DA(x_i)) \cdot F(DA(x_i))/T)}{\sum_{j=1}^{N} exp((F(DA(x_i)) \cdot F(DA(x_j))/T)}, \quad (1)$$

where T is a temperature parameter [31]. $DA(\cdot)$ denotes the stochastic data augmentation function. $F(\cdot)$ is the simplified presentation of the encoder network $f(\cdot)$ and the project head $g(\cdot)$. In recent methods, through designing a memory bank, MoCo [35] maintains the consistency of the negative sample pairs. SimCLR [34] calculates the pairwise similarity between two similar samples from the images in the same batch, which pushes the negative samples away while pulling the positive samples. Consistency regularization can be interpreted as a special form of contrastive learning, in which only positive samples are included.

C. Consistency Regularization

Consistency regularization utilizes the assumption that the classifier should output the same prediction for the unlabeled data even after it is augmented. Data augmentation is a frequent regularization technique in semi-supervised learning. Through various data augmentation methods, consistency regularization generates a copy of the sample regarded as a similar

sample to the original data. In the simplest form, prior work [43] adds the following consistency regularization loss on unlabeled samples:

$$||p(y|DA(x); \theta) - p(y|DA(x); \theta)||_2^2, \quad (2)$$

where $DA(\cdot)$ is a stochastic data augmentation. With the use of an exponential moving average (EMA) model, Mean-Teacher [27] replaces one of the terms in Eq.2, which provides a more stable target. To maximally alter the output class distribution, Virtual Adversarial Training (VAT) [44] uses an adversarial transformation in place of $DA(\cdot)$. More recently, a form of consistency regularization is utilized in Mixmatch [28] by using random horizontal flips and crops for the input samples. Unsupervised data augmentation (UDA) [45], ReMixMatch [30] and FixMatch [29] have been proposed with the use of weak and strong data augmentations. Generally speaking, through a weakly-augmented unlabeled sample, they generate a pseudo label and enforce consistency against the strongly-augmented version of the same input. The above consistency regularization models are based on data augmentation to generate positive samples. Although promising performance has been achieved, we observe that the discriminative capability of previews methods is limited since they would suffer from the semantic information drift issue. Therefore, the constructed samples are no longer similar. Instead of carefully designing data augmentations to utilize consistency regularization, we use an interpolation-based method to obtain positive pairs, which will avoid the semantic information drift caused by data augmentations.

D. The Interpolation-based Method

Mixup [32] is an effective data augmentation strategy for image classification in computer vision [46]–[49]. It linearly interpolates the input samples and their labels on the input data and label spaces.

$$\begin{aligned}
\lambda &\sim Beta(\alpha, \beta), \\
\lambda' &= max(\lambda, 1 - \lambda), \\
x' &= \lambda' x_1 + (1 - \lambda') x_2, \\
y' &= \lambda' y_1 + (1 - \lambda') y_2,
\end{aligned} \quad (3)$$

where the α and β are the parameter of Beta distribution, $\lambda \in [0, 1]$. The interpolations of input samples should lead to interpolations of the associated labels. In this manner, Mixup could extend the training distribution. It is recently achieved state-of-the-art performance through different tasks and network architectures. In [50], the interpolations are performed in the input space. In order to improve model performance, [51] is proposed to measure the realism of latent space interpolations in unsupervised learning. [33] performs the interpolation between input and pseudo-labels. Although the above methods are verified to be effective, they will still change the construction method of consistency regularized positive sample pairs. Therefore, how to solve the semantic information drift in consistency regularization is an open question. Different from the above approaches, we propose an interpolation-based method term ICL-SSL to construct positive
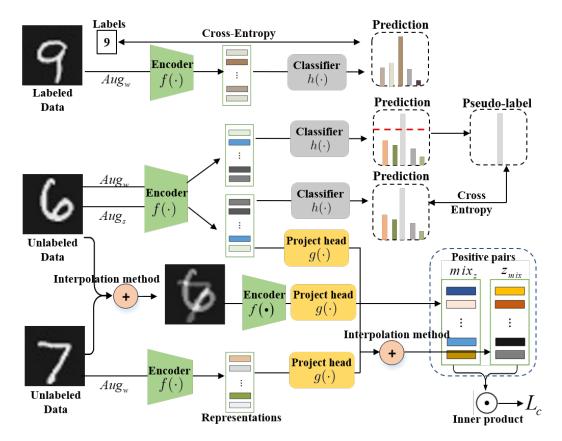
Figure 4: Illustration of Interpolation-based Contrastive Learning Semi-Supervised Learning (ICL-SSL) mechanism. Following the definition in [29], Aug_w denotes the weak augmentation of the original input image. Aug_s denotes strong augmentation. Specifically, given two unlabeled images u_i and u_j, we firstly embed the samples separately into the latent space. Then, we conduct image-level interpolation for an integrated image and do the embedding with the same network. z_{mix} and mix_z are the positive embeddings pair constructed by ICL-SSL. By combining these two positive embeddings, we design a novel contrastive loss L_c and force these two embeddings to change linearly between samples, improving the discriminative capability of the network. Therefore, our network would be guided to learn more discriminative embeddings with the interpolation-based method.

sample pairs. Without using data augmentation to construct positive sample pairs, ICL-SSL is performed between the input samples and the representations, thus avoiding semantic information drift.

III. METHOD

In this section, we introduce our proposed semi-supervised learning method. Firstly, we will explore the reason for the performance degradation under few labels via some experiments on MINIST dataset. Through the exploratory experiment, we analyze that the semantic information of the input samples will be drifted after some inappropriate data augmentations, thus limiting the performance. After that, to address this issue, we introduce an interpolation-based method ICL-SSL under few labels to construct more reliable positive sample pairs. Finally, we will detail the designed contrastive loss of ICL-SSL.

A. Semantic information drift

Although promising performances have been achieved by the existing algorithms, we observe that when the number of labeled data gets extremely small, e.g. 2 to 3 labels for each category, the performance of the existing algorithms would decrease drastically. The detailed observation is shown in Table. I. Therefore, we conduct experiments to explore the reason to cause the performance dropping.

Consistency regularization is an essential piece for many state-of-the-art semi-supervised learning methods [28]–[31]. A common assumption of consistency regularization is that the classifier should output the same class probability of an unlabeled sample even if it is augmented.

In several SSL methods [28]–[30], when training data is not enough for generalization, data augmentation is a technique to apply consistency regularization. MixMatch [28] processes the input samples through random horizontal and random crops. The weak data augmentation method uses horizontal flips and vertical flips to process unlabeled samples in FixMatch [29].

Through experiments shown in Fig. 2, we find that some data augmentations will change the semantic information about the input samples, leading to a decrease in the semantic similarity of the constructed samples damaging the SSL training. We visualize the result of data augmentation. It can be

Algorithm 1 Interpolation-based Contrastive Learning Semi-Supervised Learning(ICL-SSL)

Input: Labeled data $X = (x_1, y_1), (x_2, y_2), ..., (x_b, y_b)$, unlabeled data $U = (u_1, u_2, ..., u_b)$, Beta distribution parameter α for feature interpolation, Batch size b, Epoch number e

1: **while** $e <$ Epoch **do**
2:    **for** $i = 1$ to b **do**
3:       $y_b = p(y|x)$;
4:       $q_i = p(y|u_i)$;
5:       $z_i = g(f(u_i))$;
6:    **end for**
7:    **for** $i \in 1, . . . ,$ b and $j \in 1, . . . ,$ b **do**
8:       $\lambda = Beta(\alpha, \alpha)$;
9:       $u_{mix} = \lambda * u_i + (1 - \lambda) * u_j$;
10:      $z_{mix} = g(f(u_{mix}))$;
11:      $mix_z = \lambda * z_i + (1 - \lambda) * z_j$;
12:    **end for**
13: **end while**
14: Calculate classification loss via Eq.8, 10 and 11

---

found that the semantic information of the input samples has been changed. Fig. 2 shows that under one data augmentation (random vertical flip), the semantic information of "7"s and "2"s, "6"s and "9"s, "2"s and "5"s can easily get changed. As a result, the quality of the constructed positive samples decreases or the construction fails, which in turn affects the performance of the model. To further verify the effect of data augmentation, we implement experiments on the MINIST dataset.

As shown in Fig. 5(b), MINIST is a dataset composed of handwritten numbers, which is commonly used in deep learning research. MINIST consists of 60000 training data and 10000 test data. Aiming to reduce the influence of irrelevant factors (e.g. complex structure of training model) to the performance, we explore the semantic information drift problem caused by data augmentation with two-layer MLPs.

From the empirical analysis, we observe that the accuracy is decreased by 5.0% after the random horizontal flip argumentation on MINIST. As a consequence, after random vertical flips, the accuracy decreases by 4.0% shown in Fig. 3. Additionally, we also explore rotation, random re-cropping and random cropping, the result shows that those data augmentations will also limit the performance of the model.

The experiment on MINIST can illustrate that during SSL training, some inappropriate data augmentations will change the semantic information of the input samples. Therefore, the semantic correlation of positive sample pairs will be destroyed by inappropriate data augmentations. When the label information is lacking, the incorrect regularization caused by data augmentation would mislead the network and limit the algorithm performance.

B. ICL-SSL

To solve the semantic information drift problem, we proposed a novel interpolation contrastive learning Semi-supervised learning method termed ICL-SSL. Specifically, ICL-SSL does not change the semantic information during the positive pair construction of consistency regularization. In the following, we first obtain the low-dimensional representation z

of the unlabeled sample. Then, we describe the interpolation-based positive sample pairs construction method and loss function in detail.

In our ICL-SSL method, the representations are extracted by encoder network $f(\cdot)$. Concretely, for any two unlabeled samples u_i, u_j in a batch of unlabeled sub-dataset $\mathcal{U}^B$, we could obtain their normalized representations z_i, z_j with ℓ^2-norm:

$$z_i = F(u_i), z_i = \frac{z_i}{||z_i||_2},$$
$$z_j = F(u_j), z_j = \frac{z_j}{||z_j||_2}, \tag{4}$$

where $F(\cdot)$ is defined as $g(f(\cdot))$, a simple form of encoder network $f(\cdot)$ and project head $g(\cdot)$.

After that, we perform interpolation operations on the normalized low-dimensional feature representations z_i and z_j.

$$mix_z = \lambda z_i + (1 - \lambda)z_j$$
$$= \lambda F(u_i) + (1 - \lambda)F(u_j), \tag{5}$$

where mix_z denotes the interpolated representation of z_i and z_j, λ is generated by Beta distribution. Simultaneously, unlike the above steps, we first perform an interpolation operation in the sample space (u_i, u_j) and then get the normalized low-dimensional feature:

$$z_{mix} = F(\lambda u_i + (1 - \lambda)u_j), \tag{6}$$

where z_{mix} is the representation of interpolated input data u_i and u_j. The constructed positive sample pair can be presented as follows:

$$[mix_z, z_{mix}]. \tag{7}$$

The framework of our proposed ICL-SSL is shown in Fig. 4. ICL-SSL is a semantic-agnostic positive sample construction method. Specifically, we generate one positive sample from the features z_{mix} obtained by interpolating two inputs, and the other mix_z from interpolating the two features of the input. By this setting, both of these positive samples contain the original semantic information of each input (u_i, u_j). It has demonstrated that the interpolation operation has the effect to push the decision boundaries away from the class boundaries in [32], [33]. In this manner, with the utilization of our ICL-SSL, the margin decision boundaries would get larger, thus improving the discriminative capability of the network under few labels.

C. Loss function

The loss of ICL-SSL mainly consists of three parts: the supervised classification loss L_x, the unsupervised classification loss L_u and the contrastive loss L_c.

In detail, L_x is the supervised classification loss on the labeled data, which is defined as the cross-entropy between the ground-truth labels and the model's predictions:

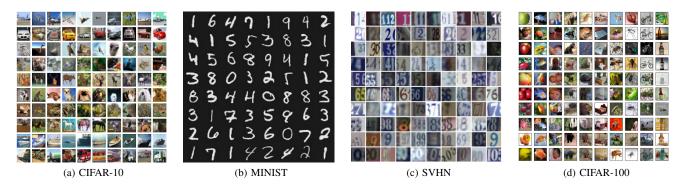$$L_x = \frac{1}{B} \sum_{b=1}^{B} H(y_b, p(y|x_b)), \tag{8}$$

(a) CIFAR-10   (b) MINIST   (c) SVHN   (d) CIFAR-100

Figure 5: Illustration of the CIFAR-10, MINIST, SVHN, CIFAR-100 dataset.

where x_b denotes the labeled data in $\mathcal{X}^B$. $p(y|x_b)$ is the output of the classifier. H is the cross-entropy between the two distributions y_b and $p(y|x_b)$.

For the unlabeled data, its pseudo label $\hat{q}_b$ is generated by the classification head $h(\cdot)$ and the $Softmax$ function. The formula can be described as:

$$\hat{q}_b = Softmax(h(f(u_b))). \tag{9}$$

L_u is defined as the cross-entropy between the pseudo-labels and the model's predictions. It can be calculated by:

$$L_u = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \ell \left(max\left(\hat{q}_b \right) \geq \tau \right) H \left(\hat{q}_b, p\left(y | u_b \right) \right), \tag{10}$$

where $\hat{q}_b$ is the predicted probability of pseudo labels. ℓ is the function to calculate the loss. When the largest class probability is above the threshold τ, the loss will be calculated. Meanwhile, μ is used to count the number of valid unlabeled samples. H is the cross-entropy between q_b and $p(y|u_b)$.

Through the positive sample pairs $[mix_z, z_{mix}]$ constructed by interpolation strategy, the contrastive loss can be computed as:

$$L_c = -log \frac{exp((mix_z \cdot z_{mix})/T)}{\sum_{k=1}^{B} I_{\lfloor k \neq i \rfloor} exp((z_i \cdot z_j)/T)}, \tag{11}$$

where T is a temperature parameter. Similar to SimCLR [34], we do not sample negative samples explicitly. Instead, we treat the other examples within a minibatch as negative samples. $I_{\lfloor k \neq i \rfloor \in 0,1}$ is an indicator function. When $k = i$, the value of I is set to 1. The similarity between positive is measured by the inner product. This loss is calculated across all positive samples in a batch. The contrastive loss encourages the model to produce similar representations for positive samples and pushes the negative samples away. The relation of the embedding changes linearly due to the proposed positive sample pairs constructed method. By minimizing Eq.11, the margin decision boundaries will be enlarged, thus improving the discriminative of the network.

In summary, the loss function of ICL-SSL can be computed by:

$$L = L_x + L_u + \alpha L_c, \tag{12}$$

Table III: Dataset summary

Dataset	Size	Train Set	Test Set	Class	Type
SVHN	32×32	73257	26032	10	image
MINIST	28×28	60000	10000	10	image
CIFAR-10	32×32	50000	10000	10	image
CIFAR-100	32×32	50000	10000	100	image

where L_x represents the supervised loss and L_u is the unsupervised loss. α is a trade-off hyper-parameter to control the weight of the total loss. The detailed learning procedure of ICL-SSL is shown in Algorithm 1.

IV. EXPERIMENT

We evaluate the effectiveness of ICL-SSL on several semi-supervised learning benchmarks. We focus on the most challenging label-scare scenario where few labels are available, e.g., 2 or 3 labels for each category. At the same time, our ablation study teases apart the contribution of ICL-SSL components. In addition, we further verify the generality of the proposed method by improving the performance of the existing state-of-the-art algorithms considerably with our proposed strategy.

A. Implementation details

1) Datasets & Metric: The proposed algorithms are experimentally evaluated on SVHN [52], CIFAR-10 [53] and CIFAR-100 [53] datasets.

- The CIFAR-10 dataset consists of 60000 images of size 32×32. The training set of CIFAR-10 consists of 50000 images and the test set consists of 10000 images. The dataset includes ten classes, including images of natural objects such as horse, deer, fork, car and aircraft.
- The CIFAR-100 dataset is similar to the CIFAR-10 dataset and contains 60000 images of the size 32×32. The 100 classes in the CIFAR-100 are grouped into 20

superclasses. Each class consists of 500 training images and 100 testing images.

- The SVHN dataset includes 73257 training data and 26032 test data of size 32 × 32. Besides, each example is a close-up image of house numbers from 0 to 9.

Detailed dataset statistics are summarized in Table III. We use the accuracy metric to evaluate the classification performance.

2) Experiment Settings: All experiments are implemented with an NVIDIA 1080Ti GPU on PyTorch platform. Following SSL evaluation methods, we evaluate our method on standard SSL benchmarks with the "Wide-ResNet-28" model from [54]. Compared with other methods, our model focuses on the challenging label-scare scenario e.g., 2 or 3 labels for each category. For CIFAR-10 and SVHN datasets, we train them for 300 epochs until convergence, the batch size chosen by us is 64. Due to the limited computing resources, the batch size of the all comparison experiments on CIFAR-100 dataset is set to 16. The weight parameter α to control loss is set to 0.5, and the parameter μ of the batch size for the control of unlabeled data is set to 1. The learning rate is set to 0.03 for CIFAR-10, CIFAR-100 and SVHN. The threshold τ is set to 0.95. Besides, our network is trained using SGD optimizer. For our proposed method, we adopt the source data of CoMatch [31]. To alleviate the impact of randomness, we evaluate the models on 5 runs for each number of labeled points with different random seeds.

In Sub-Section "Transfer to other models", the algorithms are implemented with an NVIDIA 1080Ti GPU on PyTorch platform with 40, 250, 500, and 1000 labels on CIFAR-10 dataset. Three state-of-the-art algorithms are compared in our transferring experiments, including MixMatch [28], Mean-Teacher [27] and VAT [44]. For those algorithms, we reproduce results by adopting their source code with the original settings. The code for the compared algorithms can be downloaded from the authors' website: MixMatch [1], Mean-Teacher [2], VAT[3]. Specifically, the training epoch is set as 300. The learning rate of the optimizer is set as 0.002 for MixMatch, 0.003 for Mean-Teacher, and 0.01 for VAT.

B. Comparison with the State-of-the-Art Algorithms

In this section, six state-of-the-art semi-supervised algorithms are compared to verify the effectiveness of ICL-SSL. The information for the compared algorithms is listed as follows:

(1) **CoMatch** [31]: The class probabilities and low-dimensional embeddings are jointly learned in CoMatch. Through imposing a smoothness constraint to the class probabilities, the quality of pseudo labels could be improved. Overall, CoMatch combines the pseudo-based model, the contrast-loss-based model and the graph-based model to improve the model performance in the case of few labels.

(2) **FixMatch** [29]: For the labeled image FixMatch utilize weak-augmentation to generate the pseudo label. Additionally, for the unlabeled image, the pseudo label is obtained by the high-confidence prediction. And then, the network is trained to predict the pseudo label with the strongly augmented version of the same image.

(3) **MixMatch** [28]: MixMatch jointly optimizes two losses: the supervised loss and unsupervised loss. In detail, cross-entropy is chosen for the supervised loss. The unsupervised loss is the mean square error (MSE) between predictions and generated pseudo labels. MixMatch constructs pseudo labels by data augmentation. With the use of the sharpen function $Sharpen(\cdot)$, MixMatch could improve the quality of pseudo labels. In addition, Mixup is added in the training process, which can construct virtual samples through interpolation.

(4) **Virtual Adversarial Training(VAT)** [44]: VAT is based on data perturbation. It replaces data augmentation with adversarial transformations. The adversarial transformation can lead to a lower classification error.

(5) π**-model** [26]: For the same image, data augmentation is used to apply consistency regularization. The loss of π-model contains the supervised loss and the unsupervised loss. Specifically, the supervised loss is defined as the cross-entropy loss, and the unsupervised loss is the unsupervised consistency loss.

(6) **Mean-Teacher** [27]: Mean-Teacher is a student-teacher-approach for SSL. The teacher model is based on the average weights of a student model in each update step. In Mean-Teacher, the mean square error loss (MSE) is used as its consistency loss between two predictions. Besides, it uses the exponential moving average (EMA) to update, because the EMA is only updated once per epoch, which can control the model update speed.

C. Performance Comparison

1) CIFAR-10: To demonstrate the superiority of ICL-SSL, we conduct performance comparison experiments for our proposed ICL-SSL and 4 baselines, including Mean-Teacher [27], MixMatch [28], FixMath [29] and CoMatch [31]. For CIFAR-10 dataset, we evaluate the accuracy of above methods with a varying number of labeled data from 20 to 40. The results are reported in Table. IV. For fairness, we create 5 runs for each number of labeled points with different random seeds to alleviate the influence of randomness. We can observe that our method ICL-SSL outperforms all other methods by a significant margin, taking the result on only 2 labeled data in each class for example, ICL-SSL could reach an accuracy of 88.73%. For comparison, at 20 labels the second best algorithm (CoMatch [31]) achieves an accuracy 83.43%, which is 5.30% lower than ICL-SSL. ICL-SSL can achieve higher accuracy by using fewer labels.

2) SVHN: Moreover, we implement comparison experiments on SVHN dataset. The comparison algorithms contains π model [26], Mean-Teacher [27], MixMatch [28], FixMatch [29]. The quantity of labels is 250 to 1000. The results can be seen in TableV. With different random seeds, we evaluate the models on 5 runs for each number of labeled data. We could observe that ICL-SSL outperforms all compared methods
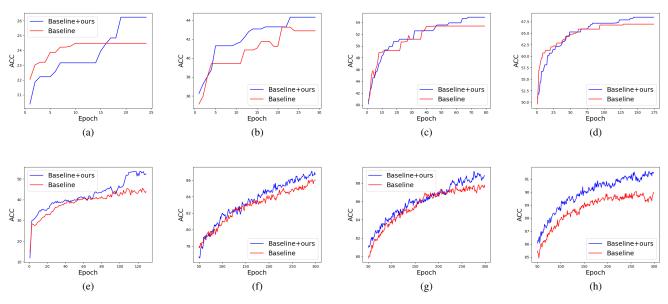

Figure 6: Performance variation when the number of labeled data changes from 40 to 1000 on the CIFAR-10 dataset. (a) - (d) are the classification accuracy of Mean-Teacher [27] and (e) - (h) are the results of MixMatch [28]. The blue curve denotes the accuracy enhanced by our positive sample pair construction mechanism, and the red curve represents the accuracy of the original model.
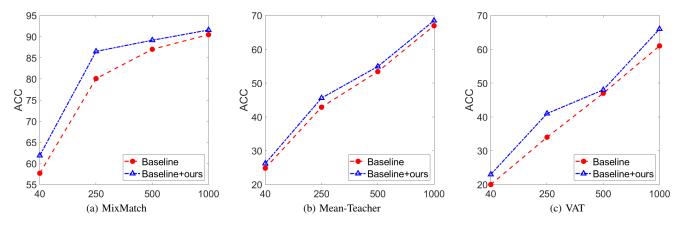

Figure 7: Performance comparison of different state-of-the-art methods on CIFAR-10 dataset with a varying number of labeled data. The blue curve denotes the accuracy enhanced by our positive sample pair construction mechanism, and red curve represents the accuracy of the original model, respectively.

SVHN with 250, 500, and 1000 labeled data. For example, ICL-SSL exceeds MixMatch by 3.46% with 250 labels.

3) CIFAR-100: To further investigate the effectiveness of our proposed model, we conduct experiments on CIFAR-100 dataset. Table. V reports the performance of the four methods with 200, 400, and 1000 labels. From those results, we can observe that, our proposed ICL-SSL could achieve better performance compared with other state-of-the-art algorithms. Taking the result with 200 lables for example, ICL-SSL exceeds FixMatch [29] by 4.75%.

Through the above experiments, our method outperforms all the existing methods in the case of few labels. The reason is that other methods use data augmentation to generate positive sample pairs, easily leading to incorrect regularization. Different from them, our ICL-SSL aims to improve the discriminative capability from two aspects. Firstly, we proposed an interpolation-based method to construct more reliable positive sample pairs, thus alleviating the incorrect regularization. Additionally, we design a contrastive loss to guide the embedding to change linearly in samples, which could enlarge the margin

Table IV: Accuracy comparison with other state-of-the-art methods on five different folds, including Mean-Teacher [27], MixMatch [28], FixMatch [29] and CoMatch [31] on CIFAR-10 dataset. The red and blue values indicate the best and the runner-up results.

Method		20 labels	30 labels	40 labels
Mean-Teacher	[27]	21.79±0.57	24.51±0.35	24.93±0.62
MixMatch	[28]	38.51±8.48	50.10±5.81	59.08±3.04
FixMatch	[29]	72.63±5.37	86.65±3.56	89.69±4.58
CoMatch	[31]	83.43±9.20	88.68±3.79	90.14±2.86
ICL-SSL	**Ours**	88.73±5.69	90.30±3.10	91.78±2.23

Figure 8: Sensitivity analysis of the hyper-parameter α on CIFAR-10 dataset with 20 labels.

decision boundaries. In summary, we proposed ICL-SSL that could improve the discriminative capability of the network and achieves the top-level performance on CIFRA-10, SVHN, and CIFAR-100 dataset.

D. Time Cost

As shown in Table VIII, we compare the training and the inference time of ICL-SSL and other state-of-the-art algorithms, including MixMatch [28], FixMatch [29], and CoMatch [31]. The results are the average training time for 300 epochs with 20 labels on CIFAR-10 dataset. We observe that the training and the inference time of ICL-SLL are 193.81 seconds and 1.08 seconds, respectively. From the Table.VIII we find that the computational efficiency of the proposed algorithm is comparable to the MixMatch and FixMatch and is much faster than that of CoMatch.

E. Ablation Study

In this section, we implement extensive ablation studies to examine the effect of different components in ICL-SSL. Due to the number of experiments in our ablation study, we perform the study with 20 and 40 labels split from CIFAR-10 dataset. The parameter settings are kept the same with comparison experiments, and the results are shown in Table. VII.

Effective of contrastive loss

To further investigate the superiority of the proposed contrastive loss, we experimentally compare our method. Here, we denote the FixMatch [29] as the baseline. With the experimental results, in the case of few labels, the model performance achieves better performance than that of baselines. Taking the result on CIFAR-10 with 20 labels for example, the accuracy exceeds the baseline by 16.1% performance increment. From the empirical analysis, it benefits from the contrastive loss to guide the embedding of the network to change linearly between samples to improve the discriminative capability of the network.

Effective of interpolation-based positive samples construction method.

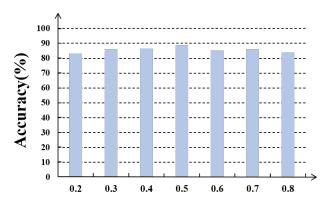Additionally, we verify the effectiveness of the interpolation-based positive samples construction method. As shown in TableVII, we can observe that the accuracy would decrease from 88.73% to 56.91%. The above experiments demonstrate the effectiveness of the interpolation-based positive samples construction method.

F. Sensitivity Analysis

Further, we investigate the effect of hyper-parameters α. As shown in Fig. 8, we observe that the classification accuracy will not fluctuate greatly when the α is varying. This demonstrates that our model ICL-SSL is insensitive to the variation of the hyper-parameter α.

G. Transferring to other models

To verify the generality of our proposed ICL-SSL, we transfer our method to the existing state-of-the-art algorithms. We implement our method into other semi-supervised learning models (MixMatch [28], VAT [44], Mean-Teacher [27]). All the experiments are implemented with CIFAR-10 dataset. Experiments are carried out on the number of labeled data from 40, 250, 500 and 1000. Here, we denote the baseline and the baseline with our method ICL-SSL as "B" and "B+O", respectively.

From Fig. 7, we have observed as follows: 1) The models could achieve better performance with our method. 2) As shown in Table VI, taking the results in MixMatch [28] for example, our method could improve the classification accuracy by 4.02% on 40 labeled data and 2.14% on 500 labels on CIFAR-10 dataset, respectively. In conclusion, the experiment results show that ICL-SSL can improve the model performance in other semi-supervised models. Moreover, in Fig. 6, we further show that other state-of-the-art methods could obtain higher accuracy with our proposed strategy during the training process.

V. CONCLUSION

In this work, we propose an interpolation-based method termed ICL-SSL to construct reliable positive sample pairs, thus alleviating the semantic information drift with extreme labels (e.g., 2 or 3 labels for each class). Specifically, ICL-SSL is a semantic-agnostic method. We interpolate the input

Table V: Accuracy comparison with π model [26], Mean-Teacher [27], MixMatch [28], and FixMatch [29] on CIFAR-100 and SVHN dataset. The red and blue values indicate the best and the runner-up results. The average and std values of the five fold cross validation are reported.

Method		CIFAR 100			SVHN		
		200 labels	400 labels	800 labels	250 labels	500 labels	1000 labels
π Model	[26]	8.53±0.25	11.67±0.37	17.64±1.06	42.66±0.91	53.33±1.39	65.90±0.03
Mean-Teacher	[27]	7.11±0.06	11.54±0.28	17.82±0.09	42.70±1.79	55.71±0.53	67.71±1.22
MixMatch	[28]	4.55±0.45	17.68±0.07	26.75±1.13	92.12±0.06	94.53±0.43	95.13±0.04
FixMatch	[29]	9.31±0.08	24.44±0.35	28.12±0.30	95.45±0.07	95.73±0.15	95.94±0.10
ICL-SSL	Ours	14.06±0.52	26.52±1.20	33.81±0.63	95.58±0.14	95.80±0.12	96.05±0.14

Table VI: Applying the interpolation positive sample pair construction mechanism to other state-of-the-art semi-supervised learning algorithms, including MixMatch [28], VAT [44] and Mean-teacher [27], on the CIFAR-10 dataset. The blue values represent the results enhanced by our positive sample pair construction mechanism, and the black values are the results of the original model. 'B' and 'B+O' represent the baseline and the baseline with our method, respectively.

Method		40 labels		250 labels		500 labels		1000 labels	
		B	B+O	B	B+O	B	B+O	B	B+O
VAT	1	20.00	23.00	34.00	41.00	47.00	48.00	61.00	66.00
MixMatch	2	57.86	61.88	86.06	86.50	87.00	89.14	90.46	91.56
Mean-Teacher	3	24.86	26.24	42.88	45.58	53.40	54.90	66.98	68.48

Table VII: Ablation comparisons of ICL-SSL mechanism. The results are reported with 20 labels on CIFAR-10 dataset.

Ablation	20 labels	40 labels
ICL-SSL	88.73	91.78
ICL-SSL without contrastive loss	72.63	89.69
ICL-SSL without the interpolation-based method	56.91	70.89

Table VIII: Training and inference time comparison on CIFAR-10 dataset with 20 labels.

Method		Training Time (s)	Inference Time (s)
MixMatch	[28]	121.50 ± 0.21	0.60 ± 0.13
FixMatch	[29]	155.37 ± 0.10	3.46 ± 0.07
CoMatch	[31]	571.65 ± 0.15	1.30 ± 0.03
ICL-SLL	Ours	193.81 ± 0.14	1.08 ± 0.62

images and their representations in image-level and latent space, respectively. Besides, the designed contrastive loss will guide the embeddings changing linearly between samples and thus get a larger margin decision boundary. Benefiting from this mechanism, the discriminative capability of the network can be improved with extreme labels. Extensive experiments demonstrate the effectiveness and generality of our ICL-SSL.

In the future, we will try to extend ICL-SSL to other fields (e.g. graph semi-supervised node classification). Besides, as we analyzed in section IV-D, although our proposed algorithm is as efficient as other state-of-the-art contrastive algorithms, its efficiency still needs to be improved to suit even larger scale datasets. Therefore, how to reduce the training time is also a future work direction.

REFERENCES

[1] R. He, Z. Han, X. Lu, and Y. Yin, "Safe-student for safe deep semi-supervised learning with unseen-class unlabeled data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 14 585–14 594.

[2] X. Hu, Y. Zeng, X. Xu, S. Zhou, and L. Liu, "Robust semi-supervised classification based on data augmented online elms with deep features," *Knowledge-Based Systems*, vol. 229, p. 107307, 2021.

[3] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[4] I. Bekkerman and J. Tabrikian, "Target detection and localization using mimo radars and sonars," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3873–3883, 2006.

[5] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, no. 1, pp. 98–136, 2015.

[6] S. Zhou, D. Nie, E. Adeli, J. Yin, J. Lian, and D. Shen, "High-resolution encoder–decoder networks for low-contrast medical image segmentation," *IEEE Transactions on Image Processing*, vol. 29, pp. 461–475, 2019.

[7] L. Li, S. Wang, X. Liu, E. Zhu, L. Shen, K. Li, and K. Li, "Local sample-weighted multiple kernel clustering with consensus discriminative graph," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[8] S. Wang, X. Liu, L. Liu, S. Zhou, and E. Zhu, "Late fusion multiple kernel clustering with proxy graph refinement," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[9] S. Zhou, X. Liu, M. Li, E. Zhu, L. Liu, C. Zhang, and J. Yin, "Multiple kernel clustering with neighbor-kernel subspace segmentation," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 4, pp. 1351–1362, 2019.

[10] S. Wang, X. Liu, L. Liu, W. Tu, X. Zhu, J. Liu, S. Zhou, and E. Zhu, "Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9776–9785.

[11] S. Zhou, E. Zhu, X. Liu, T. Zheng, Q. Liu, J. Xia, and J. Yin, "Subspace segmentation-based robust multiple kernel clustering," *Information Fusion*, vol. 53, pp. 145–154, 2020.

[12] S. Wang, X. Liu, E. Zhu, C. Tang, J. Liu, J. Hu, J. Xia, and J. Yin, "Multi-view clustering via late fusion alignment maximization." in *IJCAI*, 2019, pp. 3778–3784.

[13] S. Zhou, X. Liu, J. Liu, X. Guo, Y. Zhao, E. Zhu, Y. Zhai, J. Yin, and W. Gao, "Multi-view spectral clustering with optimal neighborhood laplacian matrix," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 6965–6972.

[14] S. Wang, X. Liu, X. Zhu, P. Zhang, Y. Zhang, F. Gao, and E. Zhu, "Fast parameter-free multi-view subspace clustering with consensus anchor guidance," *IEEE Transactions on Image Processing*, vol. 31, pp. 556–568, 2021.

[15] D. Mahajan, R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, and L. Van Der Maaten, "Exploring the limits of weakly supervised pretraining," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 181–196.

[16] R. He, Z. Han, and Y. Yin, "Towards safe and robust weakly-supervised anomaly detection under subpopulation shift," *Knowledge-Based Systems*, p. 109088, 2022.

[17] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE transactions on cybernetics*, vol. 48, no. 2, pp. 648–660, 2017.

[18] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A semisupervised recurrent convolutional attention model for human activity recognition," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 5, pp. 1747–1756, 2019.

[19] E. Yu, J. Sun, J. Li, X. Chang, X.-H. Han, and A. G. Hauptmann, "Adaptive semi-supervised feature selection for cross-modal retrieval," *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1276–1288, 2018.

[20] Z. Zhang, T. W. Chow, and M. Zhao, "Trace ratio optimization-based semi-supervised nonlinear dimensionality reduction for marginal manifold visualization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 5, pp. 1148–1161, 2012.

[21] Z. Zhang, F. Li, L. Jia, J. Qin, L. Zhang, and S. Yan, "Robust adaptive embedded label propagation with weight learning for inductive classification," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3388–3403, 2017.

[22] H. Zhang, Z. Zhang, M. Zhao, Q. Ye, M. Zhang, and M. Wang, "Robust triple-matrix-recovery-based auto-weighted label propagation for classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4538–4552, 2020.

[23] Q. Ye, J. Yang, T. Yin, and Z. Zhang, "Can the virtual labels obtained by traditional lp approaches be well encoded in wlr?" *IEEE transactions on neural networks and learning systems*, vol. 27, no. 7, pp. 1591–1598, 2015.

[24] R. Jozefowicz, O. Vinyals, M. Schuster, N. Shazeer, and Y. Wu, "Exploring the limits of language modeling," *arXiv preprint arXiv:1602.02410*, 2016.

[25] M. Sajjadi, M. Javanmardi, and T. Tasdizen, "Regularization with stochastic transformations and perturbations for deep semi-supervised learning," *Advances in neural information processing systems*, vol. 29, pp. 1163–1171, 2016.

[26] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," *arXiv preprint arXiv:1610.02242*, 2016.

[27] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *arXiv preprint arXiv:1703.01780*, 2017.

[28] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," *arXiv preprint arXiv:1905.02249*, 2019.

[29] K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffel, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *arXiv preprint arXiv:2001.07685*, 2020.

[30] D. Berthelot, N. Carlini, E. D. Cubuk, A. Kurakin, K. Sohn, H. Zhang, and C. Raffel, "Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring," *arXiv preprint arXiv:1911.09785*, 2019.

[31] J. Li, C. Xiong, and S. C. Hoi, "Comatch: Semi-supervised learning with contrastive graph regularization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9475–9484.

[32] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[33] V. Verma, K. Kawaguchi, A. Lamb, J. Kannala, Y. Bengio, and D. Lopez-Paz, "Interpolation consistency training for semi-supervised learning," *arXiv preprint arXiv:1903.03825*, 2019.

[34] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[35] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9729–9738.

[36] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," *arXiv preprint arXiv:2103.03230*, 2021.

[37] X. Chen and K. He, "Exploring simple siamese representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15 750–15 758.

[38] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3733–3742.

[39] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[40] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton, "Big self-supervised models are strong semi-supervised learners," *arXiv preprint arXiv:2006.10029*, 2020.

[41] P. Chen, T. Ma, X. Qin, W. Xu, and S. Zhou, "Data-efficient semi-supervised learning by reliable edge mining," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9192–9201.

[42] Y. Liu, W. Tu, S. Zhou, X. Liu, L. Song, X. Yang, and E. Zhu, "Deep graph clustering via dual correlation reduction," in *Proc. of AAAI*, 2022.

[43] M. Sajjadi, M. Javanmardi, and T. Tasdizen, "Regularization with stochastic transformations and perturbations for deep semi-supervised learning," *Advances in neural information processing systems*, vol. 29, pp. 1163–1171, 2016.

[44] T. Miyato, S.-i. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: a regularization method for supervised and semi-supervised learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1979–1993, 2018.

[45] Q. Xie, Z. Dai, E. Hovy, M.-T. Luong, and Q. V. Le, "Unsupervised data augmentation for consistency training," *arXiv preprint arXiv:1904.12848*, 2019.

[46] T. Lucas, C. Tallec, Y. Ollivier, and J. Verbeek, "Mixed batches and symmetric discriminators for gan training," in *International Conference on Machine Learning*. PMLR, 2018, pp. 2844–2853.

[47] D. Hendrycks, N. Mu, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan, "Augmix: A simple data processing method to improve robustness and uncertainty," *arXiv preprint arXiv:1912.02781*, 2019.

[48] H. Guo, "Nonlinear mixup: Out-of-manifold data augmentation for text classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 4044–4051.

[49] H. Guo, Y. Mao, and R. Zhang, "Augmenting data with mixup for sentence classification: An empirical study," *arXiv preprint arXiv:1905.08941*, 2019.

[50] Y. Tokozume, Y. Ushiku, and T. Harada, "Between-class learning for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5486–5494.

[51] D. Berthelot, C. Raffel, A. Roy, and I. Goodfellow, "Understanding and improving interpolation in autoencoders via an adversarial regularizer," *arXiv preprint arXiv:1807.07543*, 2018.

[52] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," 2011.

[53] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.

[54] A. Oliver, A. Odena, C. Raffel, E. D. Cubuk, and I. J. Goodfellow, "Realistic evaluation of deep semi-supervised learning algorithms," *arXiv preprint arXiv:1804.09170*, 2018.

Xihong Yang is recommended for admission to the National University of Defense Technology (NUDT) as a master's student with excellent grades and competition awards. He is working hard to pursue his master degree. His current research interests include semi-supervised learning, self-supervised learning and graph neural networks.

Xiaochang Hu is pursuing his Ph.D. degree in College of Intelligence Science and Technology, National University of Defense Technology (NUDT), China. His research interests include semi-supervised learning, robotics, and outdoor scene understanding. He has published several papers in international journals and conferences such as Knowledge-Based Systems, Journal of Field Robotics etc.

Sihang Zhou received his PhD degree from School of Computer, National University of Defense Technology (NUDT), China. He is now lecturer at College of Intelligence Science and Technology, NUDT. His current research interests include machine learning and medical image analysis. Dr. Zhou has published 40+ peer-reviewed papers, including IEEE T-IP, IEEE T-NNLS, IEEE T-MI, Information Fusion, Medical Image Analysis, AAAI, MICCAI, etc.

Xinwang Liu received his PhD degree from National University of Defense Technology (NUDT), China. He is now Professor of School of Computer, NUDT. His current research interests include kernel learning and unsupervised feature learning. Dr. Liu has published 60+ peer-reviewed papers, including those in highly regarded journals and conferences such as IEEE T-PAMI, IEEE T-KDE, IEEE T-IP, IEEE T-NNLS, IEEE T-MM, IEEE T-IFS, ICML, NeurIPS, ICCV, CVPR, AAAI, IJCAI, etc. He serves as the associated editor of Information Fusion Journal. More information can be found at https://xinwangliu.github.io/.

En Zhu received his Ph.D. degree from National University of Defense Technology (NUDT), China. He is now a professor at the School of Computer Science, NUDT, China. His main research interests are pattern recognition, image processing, machine vision and machine learning. Dr. Zhu has published 60+ peer-reviewed papers, including those in IEEE T-CSVT, IEEE T-NNLS, PR, AAAI, IJCAI, etc. He was awarded China National Excellence Doctoral Dissertation.