OpenLDN: Learning to Discover Novel Classes for Open-World Semi-Supervised Learning

Mamshad Nayeem Rizve[1], Navid Kardan[1], Salman Khan[2], Fahad Shahbaz Khan[2], and Mubarak Shah[1]

[1] Center for Research in Computer Vision, UCF, USA
[2] Mohamed bin Zayed University of AI, UAE
{nayeemrizve, kardan}@knights.ucf.edu, {salman.khan, fahad.khan}@mbzuai.ac.ae, shah@crcv.ucf.edu

Abstract. Semi-supervised learning (SSL) is one of the dominant approaches to address the annotation bottleneck of supervised learning. Recent SSL methods can effectively leverage a large repository of unlabeled data to improve performance while relying on a small set of labeled data. One common assumption in most SSL methods is that the labeled and unlabeled data are from the same data distribution. However, this is hardly the case in many real-world scenarios, which limits their applicability. In this work, instead, we attempt to solve the challenging open-world SSL problem that does not make such an assumption. In the open-world SSL problem, the objective is to recognize samples of known classes, and simultaneously detect and cluster samples belonging to novel classes present in unlabeled data. This work introduces OpenLDN that utilizes a pairwise similarity loss to discover novel classes. Using a bi-level optimization rule this pairwise similarity loss exploits the information available in the labeled set to implicitly cluster novel class samples, while simultaneously recognizing samples from known classes. After discovering novel classes, OpenLDN transforms the open-world SSL problem into a standard SSL problem to achieve additional performance gains using existing SSL methods. Our extensive experiments demonstrate that OpenLDN outperforms the current state-of-the-art methods on multiple popular classification benchmarks while providing a better accuracy/training time trade-off. Code: https://github.com/nayeemrizve/OpenLDN

Keywords: Open-world, Semi-supervised learning, Novel classes

1 Introduction

Deep learning methods have made significant progress on challenging supervised learning tasks [28,61,27,10,13]. However, the supervised learning paradigm assumes access to large amounts of manually labeled data which is time-consuming and expensive to acquire. Several approaches have been proposed to address this problem, including semi-supervised learning [62,46,6], active learning [36,21,55],

self-supervised learning [18,14,26], transfer learning [56,73,37], and few-shot learning [20,66,58,53]. Among them, semi-supervised learning (SSL) is one of the dominant approaches which reduces the amount of annotation required by taking advantage of a large collection of unlabeled data.

Even though recent SSL methods [6,5,59,69] have achieved promising results, one of their primary assumptions is that both labeled and unlabeled data come from the same distribution. However, this assumption is difficult to satisfy in many real-world scenarios (open-world problems e.g. [4,33]). For instance, unlabeled data is commonly mined from the web sources which can include examples from unknown classes. It has been established that training with such examples generally deteriorates the performance of the standard SSL methods [47,16] . To mitigate the negative impact of unlabeled samples from unknown (novel) classes, different solutions have been proposed [23,16,75]. However, their main motivation is to merely ignore novel class samples to prevent performance degradation on known classes. In contrast, recently ORCA [7] generalizes the SSL problem with novel classes, where the objective is not only to retain the performance on known classes but also to recognize samples of novel classes. This realistic SSL setup is called open-world SSL problem and is the focus of this work.

This work proposes OpenLDN which employs a pairwise similarity loss to discover novel classes. This loss solves a pairwise similarity prediction task that determines whether an image pair belongs to the same class or not. Essentially, this task is akin to unsupervised clustering problem [11,67], thereby promoting novel class discovery by identifying coherent clusters. The fundamental challenge for solving pairwise similarity is to determine similarity relationship between a pair of images without accessing their class labels. One common way to overcome this challenge is to estimate the pairwise similarity relationship based on pretrained unsupervised/self-supervised features [24,7]. However, this process is computationally expensive. To avoid dependency on unsupervised/self-supervised pretraining, instead, we exploit the information available in the labeled examples from known classes for solving the pairwise similarity prediction task, and introduce a pairwise similarity prediction network to generate the similarity scores between a pair of images. To update the parameters of this network, we resort to a bi-level optimization rule [3,23], which transfers the information available in the labeled examples of known classes to utilize them in learning unknown classes. In particular, we *implicitly* optimize the parameters of the similarity prediction network based on the cross-entropy loss on labeled examples. This way, we solve the pairwise similarity prediction task without relying on unsupervised/self-supervised pretraining, which makes the overall training more efficient while providing substantial performance gains.

Learning pairwise similarity relationship based on output probabilities leads to implicitly discovering clusters according to the most probable class, hence, the discovery of novel classes. Once we learn to recognize novel classes, we can generate pseudo-labels for novel class samples. This subsequently allows us to transform the open-world SSL problem into a closed-world SSL problem by utilizing the generated pseudo-labels of unlabeled samples to incorporate novel

class samples into the labeled set. This unique perspective of transforming the open-world problem into a closed-world one is particularly powerful since it allows us to leverage any off-the-shelf closed-world SSL method to achieve further improvements. However, one shortcoming of this strategy is that the generated pseudo-labels for novel classes tend to be noisy, which can in turn impede the subsequent training. To address this issue, we introduce *iterative pseudo-labeling*, a simple and efficient way to handle the noisy estimation of pseudo-labels.

In summary, our key contributions are: (1) we propose a novel algorithm, OpenLDN, to solve open-world SSL. OpenLDN applies a bi-level optimization rule to determine pairwise similarity relationship without relying on pretrained features, (2) we propose to transform the open-world SSL into a closed-world SSL problem by discovering novel classes; this allows us to leverage any off-the-shelf closed-world SSL method to further improve performance, and (3) we introduce *iterative pseudo-labeling*, a simple and efficient method to handle noisy pseudo-labels of novel classes, (4) our experiments show that OpenLDN outperforms the existing state-of-the-art methods by a significant margin.

2 Related Works

Semi-Supervised Learning: SSL is a popular approach to handle label annotation bottleneck in supervised learning [22,32,44,35,50,15,9]. Generally, these methods are developed for a closed-world setup, where unlabeled set only contains samples from the known classes. The two most dominant approaches for closed-world SSL are consistency regularization [54,41,45,63] and pseudo-labeling [43,57,2,51]. The consistency regularization based methods minimize a consistency loss between differently augmented versions of an image to extract salient features from the unlabeled samples. Pseudo-labeling based methods generate pseudo-labels for the unlabeled samples by a network trained on labeled data, and subsequently training on them in a supervised manner. Finally, the hybrid approaches [6,5,59] combine both consistency regularization and pseudo-labeling.

Recent works [47,16] demonstrate that the presence of novel class samples in the unlabeled set negatively impacts the performance on known classes. Different solutions have been proposed to address this issue [23,16,75]. A weight function is trained in [23] to down-weight the novel class samples. Novel class samples are filtered out in [16] based on confidence scores. Weighted batch normalization is introduced in [75] to achieve robustness against novel class samples. However, none of these methods attempt to solve the challenging open-world SSL problem, where the objective is to detect samples of novel classes and classify them. To the best of our knowledge, ORCA [7] is the only method that addresses this issue by introducing an uncertainty-aware adaptive margin based cross-entropy loss to mitigate excessive influence of known classes at early stages of training. However, to discover novel classes, ORCA relies on self-supervised pretraining, which is computationally costly. To overcome the reliance on self-supervised pretraining, the pairwise similarity loss in OpenLDN exploits the information available in labeled examples from known classes using a bi-level optimization rule.

Novel Class Discovery: Novel class discovery problem [25,24,30,29,19,77,74,76,31] is closely related to unsupervised clustering [71,70,68,64]. The key difference between the novel class discovery and unsupervised clustering is that the former relies on an extra labeled set to learn the novel classes. To discover novel classes, [24] performs self-supervised pretraining followed by solving a pairwise similarity prediction task based on the rank statistics of the self-supervised features. [25] extends the deep clustering framework to discover novel classes. Pairwise similarity prediction task is also applied in [30,29] to categorize novel classes by transferring knowledge from the known classes. While novel class discovery methods generally use multiple objective functions, [19] simplifies this using multi-view pseudo-labeling and training with cross-entropy loss. The key difference between open-world SSL and novel class discovery is that the former does not assume that unlabeled data only contains novel class samples. Hence, novel class discovery methods are not readily applicable to open-world SSL problem. Additionally, our experimentation shows that OpenLDN outperforms the appropriately modified novel class discovery methods for open-world SSL by a considerable margin.

3   Method

To identify unlabeled samples from both known and novel classes, we introduce a pairwise similarity loss to implicitly cluster the unlabeled data into known and novel classes. This implicit clustering induces discovery of novel classes which is complemented by a cross-entropy loss and an entropy regularization term. Next, we generate pseudo-labels for novel class samples to transform the original open-world SSL problem into a closed-world SSL problem. This transformation allows us to take benefit of existing off-the-shelf closed-world SSL methods to learn on both known and novel classes, delivering further gains. An overview of our approach is provided in Fig. 1. In the following, we present the problem formulation and provide the details of our approach.

3.1   Problem Formulation

We denote scalars as a, vectors as $\mathbf{a}$, matrices as $\mathbf{A}$, and sets as $\mathbb{A}$. In a matrix, the first index always represents rows and the second index represents columns. Further, $\mathbf{A}_{i,:}$ and $\mathbf{A}_{:,k}$ refer to the ith row and k^{th} column in $\mathbf{A}$, respectively.

In open-world SSL problem we assume that there is a labeled set, $\mathbb{S}_L$, and an unlabeled set, $\mathbb{S}_U$. Let, $\mathbb{S}_L = \{\mathbf{x}_i^l, \mathbf{y}_i^l\}_{i=1}^{n_l}$ represent the labeled dataset with n_l samples, where $\mathbf{x}_i^l$ is a labeled sample and $\mathbf{y}_i^l$ is its corresponding label, which belongs to one of the c_l known classes. Similarly, $\mathbb{S}_U = \{\mathbf{x}_i^u\}_{i=1}^{n_u}$, consists of n_u unlabeled samples, where $\mathbf{x}_i^u$ belongs to one of the c_u classes, where c_u is the total number of classes in $\mathbb{S}_U$. In conventional closed-world SSL setting, it is assumed that the class categories for both labeled and unlabeled data are the same. However, in open-world SSL framework, $\mathbb{S}_U$ contains some samples that do not belong to any of the known classes. The samples belonging to unknown classes are called novel class samples where each sample belongs to one of the c_n novel classes, i.e., in open-world setting $c_u = c_l + c_n$.
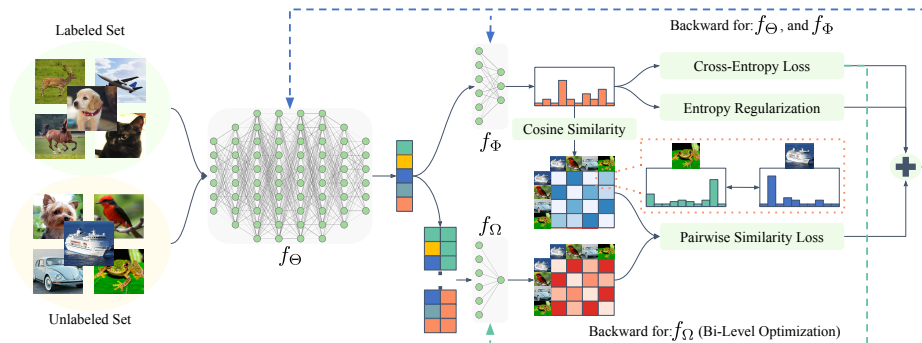
Fig. 1: *OpenLDN Overview - Learning to Discover Novel classes:* A set of labeled and unlabeled images are provided to the feature extractor, f_Θ, to obtain feature embeddings. The embeddings are passed to the classifier, f_Φ, to obtain output probabilities. We compute pairwise cosine similarity scores from the output probabilities for every possible pair in a batch. In parallel, the pairwise similarity prediction network, f_Ω, also outputs similarity scores based on pairs of feature embeddings. Afterwards, we calculate pairwise similarity loss (Eq. 2) to promote the discovery of novel classes. We also compute cross-entropy (CE) loss (Eq. 6) and entropy regularization loss (Eq. 7) to complement the pairwise similarity loss by learning from labeled and pseudo-labeled samples and avoiding trivial solutions, respectively. Next, we update the parameters of f_Θ and f_Φ to minimize the overall loss. Then we compute CE loss using only the labeled samples with updated f_Θ and f_Φ. Finally, we utilize a bi-level optimization rule to update f_Ω based on this CE loss (Eq. 4). The bi-level optimization rule helps to optimize f_Ω by transferring feature similarities from known to unknown classes.

3.2 Learning to Discover Novel Classes

To discover novel classes OpenLDN leverages a neural network, f_Θ, parameterized with Θ, as feature extractor. The feature extractor generates a feature embedding by projecting an input image $\mathbf{x}$ into the embedding space, $\mathbf{z} \in \mathbb{R}^d$, i.e., $f_\Theta : \mathbb{X} \mapsto \mathbb{Z}$. Here, $\mathbb{X}$, and $\mathbb{Z}$ are sets of input images and feature embeddings, respectively. Next, to recognise samples from novel classes, as well as to classify the samples from known classes, we apply a classifier, f_Φ, parameterized with Φ. This classifier projects the embedding vector $\mathbf{z}$ into an output classification space, $f_\Phi : \mathbb{Z} \mapsto \mathbb{R}^{c_l+c_n}$. In this output space, the first c_l logits correspond to the known classes, and the remaining c_n logits belong to novel classes. The softmax probability scores, $\hat{\mathbf{y}} \in \mathbb{R}^{c_l+c_n}$, are obtained from these output scores using softmax activation function, i.e., $\hat{\mathbf{y}} = \text{Softmax}(f_\Phi \circ f_\Theta(\mathbf{x}))$.

Our overall objective to discover novel classes while recognizing known classes consists of three losses: a) a pairwise similarity loss $\mathcal{L}_{pair}$, b) a cross-entropy (CE) loss $\mathcal{L}_{ce}$, and c) an entropy regularization term $\mathcal{L}_{reg}$. The pairwise similarity loss helps the network to discover novel classes, whereas the CE loss helps classify known classes and novel classes by utilizing the groundtruth labels and generated pseudo-labels, while the entropy regularization helps in avoiding tirvial solutions.

The overall objective function to discover novel classes is as follows:

$$\mathcal{L}_{nov} = \mathcal{L}_{pair} + \mathcal{L}_{ce} + \mathcal{L}_{reg}. \tag{1}$$

After training with $\mathcal{L}_{nov}$, samples that are assigned to any of the last c_n logits are considered novel class samples.

Pairwise Similarity Loss: Discovering novel classes is a core component of our proposed method, which is an unsupervised clustering problem, that can be expressed as a pairwise similarity prediction task [11,67]. In particular, there can only be two possible relationships between a pair of images with respect to clusters, either they belong to the same cluster or not. However, to solve a pairwise similarity prediction task supervision is needed. Previous methods [7,24] try to overcome this problem by generating pairwise pseudo-labels for all pairs of images by finding the nearest neighbors (labeled as members of the same cluster) based on pretrained features. However, such an approach is computationally expensive and suffers from noisy estimation of nearest neighbors.

In sharp contrast to this approach, instead of relying on unsupervised/self-supervised pretraining to obtain labels for the pairwise similarity prediction task, we learn to estimate the pairwise similarity scores based on the available groundtruth annotations which are more reliable. To this end, we introduce a pairwise similarity prediction network, f_Ω, parameterized by Ω. Given a pair of embedding vectors, f_Ω outputs a pairwise similarity score, i.e., $f_\Omega : \mathbb{Z} \times \mathbb{Z} \mapsto [0, 1]$. The pairwise similarity score from f_Ω can be used as the supervisory signal for minimizing the pairwise similarity loss. To this end, given a batch of images, we compute the cosine similarity of output probabilities between all pairs of images. After that, for our pairwise similarity loss, we minimize l_2 loss between the computed cosine similarity scores of output probabilities and the estimated pairwise similarity scores from f_Ω. Note that minimizing pairwise similarity loss for cosine similarities of output probabilities is crucial since this will implicitly lead to formation of clusters based on maximum probability scores, consequently, the recognition of novel classes. The pairwise similarity loss is as follows:

$$\mathcal{L}_{pair} = \sum_{i \neq j} \left(\mathrm{Sim}(\hat{\mathbf{Y}}_{i,:}, \hat{\mathbf{Y}}_{j,:}) - f_\Omega(\mathbf{Z}_{i,:}, \mathbf{Z}_{j,:}) \right)^2, \tag{2}$$

where, $\hat{\mathbf{Y}}$ is the output probability matrix, $\mathbf{Z}$ is the matrix of feature embeddings, $\mathrm{Sim}(.,.)$ denotes the cosine similarity function.

To optimize the parameters of f_Ω, we devise a bi-level optimization procedure [3]. Since we do not have access to the labels of any unlabeled sample, especially the samples from the novel classes, we utilize the groundtruth labels of the labeled examples belonging to known classes. The main motivation behind this bi-level optimization is to acquire a set of parameters Ω that do not deteriorate the performance of $f_\Phi \circ f_\Theta$ on known classes. Thereby, we optimize f_Ω based on the cross-entropy loss computed on labeled examples. The optimization procedure is as follows:

First, we update the parameters of feature extractor and classifier with the combined loss introduced in Eq. 1 to discover novel classes.

$$(\Theta^*, \Phi^*) = (\Theta, \Phi) - \alpha_{(\Theta, \Phi)} \nabla_{(\Theta, \Phi)} \mathcal{L}_{nov}(\Theta, \Phi, \Omega). \tag{3}$$

where, $\alpha_{(\Theta,\Phi)}$ refers to the learning rate for optimizing the parameters Θ and Φ.

Next, we use the supervised cross-entropy loss, $\mathcal{L}_{ce}^l = -\sum_i \sum_k \mathbf{Y}_{i,k} \log \hat{\mathbf{Y}}_{i,k}$, computed over the labeled examples to update the parameters of f_Ω. Here, $\mathbf{Y}$ is the matrix of groundtruth labels. The update rule is according to,

$$\Omega^* = \Omega - \alpha_\Omega \nabla_\Omega \mathcal{L}_{ce}^l(\Theta^*, \Phi^*), \tag{4}$$

where, α_Ω is the learning rate for optimizing the parameters Ω. Since Ω is not explicit in the objective $\mathcal{L}_{ce}^l(\Theta^*, \Phi^*)$ in Eq. 4, we perform a bi-level optimization to calculate $\nabla_\Omega \mathcal{L}_{ce}^l(\Theta^*, \Phi^*)$. This nested optimization is available in most modern deep learning packages that support automatic differentiation [49,1].

This bi-level optimization procedure ensures that the parameters of f_Ω are updated in such a way that the classification performance on known classes does not deteriorate since this is one of the primary objectives in open-world SSL.

Learning with Labeled and Pseudo-Labeled Data: In the above, we introduced the pairwise similarity loss to recognise novel classes by solving a pairwise similarity prediction task. Recall that our aim is to recognise novel classes in the unlabeled set and to classify the known classes, at the same time. This problem only allows access to a limited amount of annotations for known classes. The straightforward way to utilize these available annotations would be to minimize a cross-entropy loss on the labeled samples. However, this approach can create a strong bias towards known classes because of their strong training signal [7]. To mitigate this bias and to utilize the unlabeled samples more efficiently, we generate pseudo-labels for all the unlabeled data. The generated pseudo-labels can be used with groundtruth labels for minimizing cross-entropy loss.

Following the common practice [43,2,51], we generate pseudo-labels based on the network output probabilities. To reduce the possibility of erroneous training with unreliable pseudo-labels, we generate pseudo-labels only for sufficiently confident predictions. In addition, our pseudo-label based cross-entropy learning satisfies another commonly used objective in SSL works i.e., consistency regularization. This objective encourages perturbation invariant output distribution so that the decision boundaries lie in low density regions [12,65]. One way to satisfy this objective is to minimize the divergence between output probabilities of two randomly transformed versions of an image. However, it adds another term to the loss, and consequently a new hyperparameter. A more elegant way is to use the generated pseudo-label from one transformed version of an image as the target for the other version. We utilize the pseudo-labels generated from a weakly transformed version of an image $\mathbf{x}^w$ as the target for its strongly augmented version, $\mathbf{x}^s$. We state our pseudo-label generation process below:

$$\mathbb{S}_{PL} = \{(\mathbf{x}_i^s, \mathbb{1}_{\max(\hat{\mathbf{Y}}_{i,:}^w)}(\hat{\mathbf{Y}}_{i,:}^w)|\max(\hat{\mathbf{Y}}_{i,:}^w) > \tau\}, \tag{5}$$

where, $\tau = 0.5$ (midpoint in binary classification) to avoid per-dataset finetuning.

Once the pseudo-labels are generated we combine them with the groundtruth labels, $\mathbb{S} = \mathbb{S}_{PL} \cup \mathbb{S}_L$, and train the model using cross-entropy loss. In practice, we combine these two sets within a batch. Let $\mathbb{S}_B$ denote a batch, the cross-entropy loss on this set is defined as:

$$\mathcal{L}_{ce} = -\sum_{i \in \mathbb{S}_B} \sum_{k=1}^{c_u} \tilde{\mathbf{Y}}_{i,k} \log \hat{\mathbf{Y}}_{i,k}, \tag{6}$$

where, $\tilde{\mathbf{Y}}$ is the matrix of one-hot encoded ground-truth labels and the generated pseudo-labels.

Entropy Regularization: One of the well-known drawbacks of assigning unlabeled data to distinct categories based on a discriminative (such as cross-entropy) loss is that it can lead to a trivial solution where all the unlabeled samples are assigned to the same class [72,8,7,19]. Our pairwise similarity loss suffers from the same problem since such a solution will also minimize our pairwise similarity loss in Eq. 2. To address this, we incorporate an entropy regularization term in our training objective. One way to achieve this would be to apply entropy regularization to the output of each sample independently. However, this way of entropy minimization leads to substantial changes in the individual output probabilities which results in arbitrary class assignments for novel class samples. To avoid this problem, we apply entropy regularization over an aggregated statistic, in our case, the mean of the sample probabilities of an entire batch. This entropy regularization term prevents a single class from dominating the entire batch, where most of unlabeled samples are only assigned to one class. This term does not interfere with the balanced class assignments. The entropy regularization is defined as,

$$\mathcal{L}_{reg} = \sum_{k=1}^{c_u} \bar{\mathbf{y}}_k \log \bar{\mathbf{y}}_k, \tag{7}$$

where, $\bar{\mathbf{y}} = \frac{1}{b} \sum_{i=1}^{b} \hat{\mathbf{Y}}_{i,:}$ is the average probability of the batch, and b denotes the number of examples in a batch.

3.3   Closed-World Training with Iterative Pseudo-Labeling

Once we discover novel classes in the unlabeled data we can reformulate the open-world SSL problem as a closed-world one to improve performance. To this end, we generate the pseudo-labels for all novel class samples using Eq. 8:

$$\check{\mathbf{Y}} = \mathbb{1}_{\max(\hat{\mathbf{Y}}_{i,:})}(\hat{\mathbf{Y}}_{i,:}). \tag{8}$$

Next, using generated pseudo-labels, we add novel class samples to the labeled set. At this point, we are able to apply any standard closed-world SSL method [6,69,59,63]. Unfortunately, pseudo-labels tend to contain noise that can hamper the performance. To mitigate the negative impact of noise, we propose to perform pseudo-labeling during the closed-world SSL training in an iterative manner. This new iterative pseudo-labeling approach can be related to EM algorithm. From this perspective, we iteratively attempt to update the pseudo-labels (expectation step), and train the network by minimizing the loss on those updated pseudo-labels (maximization step). It is important to note that OpenLDN, including the final closed-world SSL retraining, is computationally lighter or comparable to the unsupervised/self-supervised pretraining based approaches (Sec. 4.2). Besides, the transformation from the open-world SSL problem to a closed-world problem is a general solution which can be applied to other methods as well. We provide our overall training algorithm in supplementary materials.

4   Experimental Evaluation

Datasets: To demonstrate the effectiveness of OpenLDN, we conduct experiments on five common benchmark datasets: CIFAR-10 [38], CIFAR-100 [39], ImageNet-100[17], Tiny ImageNet [42], and Oxford-IIIT Pet dataset [48]. Both CIFAR-10 and CIFAR-100 datasets contain 60K images (split into 50K/10k train/test set), and they have 10 and 100 categories, respectively. ImagNet1-100 dataset contains 100 image categories from ImageNet. Tiny ImageNet contains 100K/10K training/validation images from 200 classes. Finally, Oxford-IIIT Pet contains images from 37 categories split into 3680/3669 train/test set. In our experiments, we divide each of these datasets based on the percentage of known and novel classes. We consider the first c_l classes as known and the rest as novel. For known classes, we randomly select a portion of data to construct the labeled set and add the rest to the unlabeled set along with all novel class samples.

Implementation Details: We use ResNet-18 [28] as the feature extractor in all of our experiments except the experiments on ImageNet-100 dataset where we use ResNet-50. We instantiate our pairwise similarity prediction network, f_Ω, with an MLP consisting of a single hidden layer of dimension 100. The classifier, f_Φ, is a single linear layer. To discover novel classes, we train for 50 epochs with a batch size of 200 (480 for ImageNet-100) in all the experiments. We always use Adam optimizer [34]. For training the feature extractor and the classifier, we set the learning rate to $5e^{-4}$ ($1e^{-2}$ for ImageNet-100). For the pairwise similarity prediction network, we use a learning rate of $1e^{-4}$. We use two popular closed-world SSL methods, Mixmatch[6] and UDA[69], for second stage closed-world SSL training. For this closed-world training, to preserve data balance, we select an equal number of pseudo-labels for each novel class. For iterative pseudo-labeling, we generate pseudo-labels every 10 epochs. Additional implementation details are available in the supplementary materials.

Evaluation Metrics: We report standard accuracy for known classes. In addition, following [25,24,7,19], we report clustering accuracy on novel classes. We leverage the Hungarian algorithm [40] to align the predictions and groundtruth labels before measuring the classification accuracy. Finally, we also report the joint accuracy on the novel and known classes by using the Hungarian algorithm.

4.1   Results

CIFAR-10, CIFAR-100, and ImageNet-100 Experiments: We present our experimental results on CIFAR-10, CIFAR-100, and ImageNet-100 datasets in Tab. 1. We conduct experiments with 50% novel classes on all three datasets, where we include 50% labeled data from known classes. We report additional results with less labeled data in supplementary material. For comparison, we primarily use the scores reported in [7]. Furthermore, as another competitive baseline, we modify a recent novel class discovery method, UNO [19], and include its performance for comparison. Tab. 1 shows that both OpenLDN-MixMatch and OpenLDN-UDA significantly outperform novel class discovery

Method	CIFAR10			CIFAR100			ImageNet100		
	Known	Novel	All	Known	Novel	All	Known	Novel	All
FixMatch[59]	71.5	50.4	49.5	39.6	23.5	20.3	65.8	36.7	34.9
DS³L[23]	77.6	45.3	40.2	55.1	23.7	24.0	71.2	32.5	30.8
CGDL[60]	72.3	44.6	39.7	49.3	22.5	23.5	67.3	33.8	31.9
DTC [25]	53.9	39.5	38.3	31.3	22.9	18.3	25.6	20.8	21.3
RankStats[24]	86.6	81.0	82.9	36.4	28.4	23.1	47.3	28.7	40.3
UNO[19]	91.6	69.3	80.5	68.3	36.5	51.5	–	–	–
ORCA[7]	88.2	90.4	89.7	66.9	43.0	48.1	89.1	**72.1**	77.8
OpenLDN-MixMatch	95.2	92.7	94.0	73.5	46.8$_{\uparrow 3.8}$	60.1$_{\uparrow 8.6}$	–	–	–
OpenLDN-UDA	**95.7**$_{\uparrow 4.1}$	**95.1**$_{\uparrow 4.7}$	**95.4**$_{\uparrow 5.7}$	**74.1**$_{\uparrow 5.8}$	44.5	59.3	**89.6**$_{\uparrow 0.5}$	68.6$_{\downarrow 3.5}$	**79.1**$_{\uparrow 1.3}$

Table 1: Accuracy on **CIFAR-10**, **CIFAR-100**, and **ImageNet-100** datasets with 50% classes as known and 50% classes as novel.

Method	Tiny ImageNet			Oxford-IIIT Pet		
	Known	Novel	All	Known	Novel	All
DTC [25]	28.8	16.3	19.9	20.7	16.0	13.5
RankStats [24]	5.7	5.4	3.4	12.6	11.9	11.1
UNO [19]	46.5	15.7	30.3	49.8	22.7	34.9
OpenLDN-MixMatch	52.3	19.5	36.0	**67.1**$_{\uparrow 17.3}$	27.3	47.7
OpenLDN-UDA	**58.3**$_{\uparrow 11.8}$	**25.5**$_{\uparrow 9.8}$	**41.9**$_{\uparrow 11.6}$	66.8	**33.1**$_{\uparrow 10.4}$	**50.4**$_{\uparrow 15.5}$

Table 2: Accuracy on **Tiny ImageNet** and **Oxford-IIIT Pet** datasets with 50% classes as known and 50% classes as novel.

methods (DTC [25], RankStats [24], and UNO [19]) that have been modified for open-world SSL task. OpenLDN also outperforms other baseline methods: FixMatch [59], DS³L [23], and CGDL [60]. These results showcase the efficacy of OpenLDN, where it outperforms previous state-of-the-art (ORCA [7]) by about 4.7-7.5% absolute improvement on different evaluation metrics on CIFAR-10 dataset. We observe a similar pattern on CIFAR-100 dataset, where OpenLDN outperforms ORCA and UNO by 12% and 8.6% respectively on the joint task of classifying known and novel classes. We also notice a similar trend on ImageNet-100 dataset. In parallel to outperforming all the baselines methods, OpenLDN achieves a modest 1.3% improvement over ORCA. These results validate the effectiveness of OpenLDN in solving open-world SSL problem.

Tiny ImageNet and Oxford-IIIT Pet Experiments: We also conduct additional experiments on the challenging Tiny ImageNet dataset, where the total number of classes is significantly larger than CIFAR-10, CIFAR-100, and ImageNet-100 datasets. Moreover, to further demonstrate the effectiveness of OpenLDN, we also conduct experiments on a fine-grained dataset, i.e., Oxford-IIIT Pet. The results of these experiments are presented in Tab. 2. On Tiny ImageNet dataset we observe that OpenLDN significantly outperforms DTC and RankStat. Furthermore, OpenLDN-UDA achieves ~60% relative improvement on novel classes over UNO. OpenLDN-MixMatch also achieves a significant improvement over UNO on novel and all classes. Furthermore, on fine-

EntReg	SimLoss	CWT	ItrPL	Known	Novel	All
✗	✓	✗	✗	66.7	—	33.4
✓	✗	✗	✗	66.2	26.6	46.2
✓	✓	✗	✗	66.2	40.3	53.3
✓	✓	✓	✗	73.9	44.9	59.1
✓	✓	✓	✓	73.5	46.8	60.1

Table 3: Ablation study on **CIFAR-100** with 50% classes as known and 50% classes as novel. Here, **EntReg** refers to entropy regularization, **SimLoss** means pairwise similarity loss, **CWT** refers to closed-world SSL training, and **ItrPL** denotes iterative pseudo-labeling. Each component of OpenLDN contributes towards final performance.

grained Oxford-IIIT Pet dataset, we make a similar comparison and observe that OpenLDN significantly outperforms all three novel class discovery methods by a large margin. To be precise, OpenLDN-Mixmatch achieves 12.8% absolute improvement over UNO on the joint classification task and similarly, OpenLDN-UDA achieves 15.5% absolute improvement. Experiments on both of these datasets demonstrate that OpenLDN can scale up to a large number of classes and is also effective for challenging fine-grained classification task.

4.2   Ablation and Analysis

We conduct extensive ablation studies on the CIFAR-100 dataset, with 50% labels, to study the contribution of different components of OpenLDN. The results are presented in Tab. 3. In this table, the first row demonstrates that without entropy regularization OpenLDN is unable to detect novel classes. We contribute this to overpowering of a single class (sec. 3.2). Next, we evaluate the impact of our pairwise similarity loss with bi-level optimization rule. We observe that without pairwise similarity loss the performance of OpenLDN degrades by 13.7% on novel classes which makes it the most critical component of our proposed solution. We also observe that our pairwise similarity loss does not sacrifice known class performance to improve the performance on novel classes. This outcome is expected since one of the objectives of our bi-level optimization rule is to retain performance on known classes (sec. 3.2). The fourth row demonstrates the effectiveness of transforming the open-world SSL problem into a closed-world one. Here, we observe that with this component we obtain a significant improvement in known class performance. In addition, we also notice a significant improvement in novel class performance. Interestingly, on this dataset, OpenLDN outperforms ORCA [7] (on joint classification task) even without the subsequent closed-world SSL training. Finally, Tab. 3 shows that including iterative pseudo-labeling proves to be effective, where we observe ∼2% performance boost on novel classes. In conclusion, this extensive ablation study empirically validates the effectiveness of different components of our solution.

Impact of Closed-World Training: In an attempt to further investigate the impact of closed-world SSL training, we perform a t-SNE visualization of probability outputs of the novel class samples on CIFAR-10 dataset in Fig. 2. Following
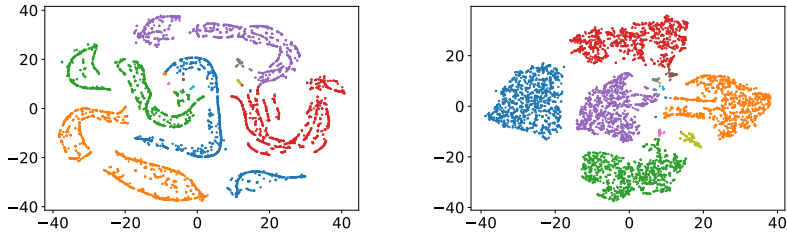
Fig. 2: Impact of closed-world training. t-SNE visualization of novel class probabilities on **CIFAR-10**: (left to right) before and after closed-world training.

our general setup, in this experiment, we consider 50% classes as novel. The results in Fig. 2 show that after novel class discovery training the novel classes form very distinct clusters. However, since novel classes are learned through auxiliary losses, without any direct supervision, there is some overlap between different classes. After training with a closed-world SSL method (MixMatch) we observe that these overlaps fade away and novel classes become well-separated and form compact clusters. This analysis further validates the complimentary effect of incorporating a closed-world SSL method after discovering novel classes.

Effect of Changing Pairwise Similarity Estimation: In another set of experiments, to analyze the effectiveness of our pairwise similarity estimation with bi-level optimization rule we conduct experiments with alternate pairwise similarity estimation methods on CIFAR-100 dataset. The results are reported in Tab. 4. In this table, to provide a baseline to compare other pairwise similarity estimation techniques, we include the performance of OpenLDN without any pairwise similarity loss in the first row. The next row in the table demonstrates the

Pairwise Sim. Est.	Known	Novel	All
No similarity	66.2	26.6	46.2
Soft Cosine	64.5	10.3	37.4
Hard Cosine (0.50)	53.7	2.1	27.9
Hard Cosine (0.95)	54.2	17.3	35.8
Nearest Neighbor	**66.4**	31.7	49.1
OpenLDN	66.2	**40.3**	**53.3**

Table 4: Results with alternate pairwise similarity estimation methods on **CIFAR-100** dataset with 50% classes as known and 50% classes as novel.

performance of OpenLDN when the pairwise similarity is directly estimated from the cosine similarity of features. Surprisingly this method of estimating pairwise similarity performs worse than our baseline without any pairwise similarity loss (first row). We hypothesize this phenomenon to instability of cosine similarity of features without any feature pretraining. In the next set of experiments (third and fourth rows), we utilize a hard version of this pairwise similarity estimation method, where we threshold the pairwise feature similarities using two different thresholds (0.50 and 0.95). After that, we minimize a binary cross-entropy loss as the pairwise similarity loss. The results of these two experiments are reported in third and fourth row, where a further drop in performance is observed. One possible explanation is that without any feature pretraining this kind of pairwise similarity estimation leads to a lot of false positives (from novel

classes) which in turn reduces the score of known classes. Finally, in the next row of the table, we use a nearest neighbor based pairwise similarity estimation technique similar to ORCA [7]. As indicated in the table, this nearest neighbor based similarity estimation improves over the baseline without any pairwise similarity loss. However, our pairwise similarity estimation based on the bi-level optimization rule outperforms this nearest neighbor based estimation technique by a significant margin; we obtain ∼9% absolute improvement on novel classes. These set of experiments further validate our claim that common pairwise similarity estimation techniques are not potent without feature pretraining, whereas, the proposed pairwise similarity estimation in OpenLDN is able to learn these pairwise similarities effectively.

Computational Cost Analysis: One of the primary advantages of OpenLDN is that unlike ORCA it does not require any feature initialization technique. This enables OpenLDN to be computationally efficient in comparison to ORCA. To demonstrate the efficiency of OpenLDN we report the performance on CIFAR-100 dataset across different training budgets in Fig. 3. We notice that, ORCA with SimCLR pre-

Fig. 3: Accuracy with respect to wall-clock time on **CIFAR-100** dataset. OpenLDN outperforms ORCA in less than 3 hours.

training takes ∼ 28 hours with our computing resources. We also observe that OpenLDN outperforms ORCA under 3 hours on this dataset even without the closed-world SSL training. On the other hand, the closed-world SSL training reaches a reasonable performance very fast and improves relatively slowly over time. Therefore, if computational budget is of concern we can stop training at an earlier stage without making a noticeable trade-off in performance.

Unknown Number of Novel Classes: In our experiments, we assume that the number of novel classes is known in advance. This follows novel class discovery methods [25,24,19], and exiting work on open-world SSL [7]. However, this is a limiting assumption since in real-world applications the number of novel classes is rarely known a priori. Therefore, estimating the number of novel classes is crucial for wider adoption. To the best of our knowledge, DTC [25] is the only method that proposes a solution for estimating the number of novel classes. However, in our experiments we find that DTC only works for small number of unknown classes and fails to estimate the number of novel classes for CIFAR-100 dataset, where 50% of classes are novel. Therefore, instead of estimating the number of novel classes by DTC, we analyze the performance of OpenLDN with the assumption that a reasonable method for estimating the number of novel classes is provided. We further assume that the hypothetical method will either overestimate or underestimate the number of novel classes. We conduct two sets of experiments to investigate both these cases. The results are demonstrated in Fig. 4. We observe that performance of OpenLDN is stable over a
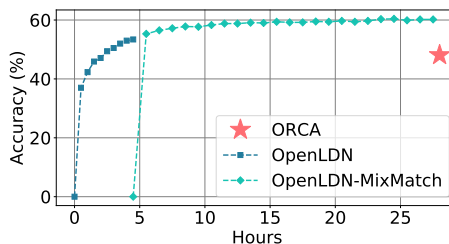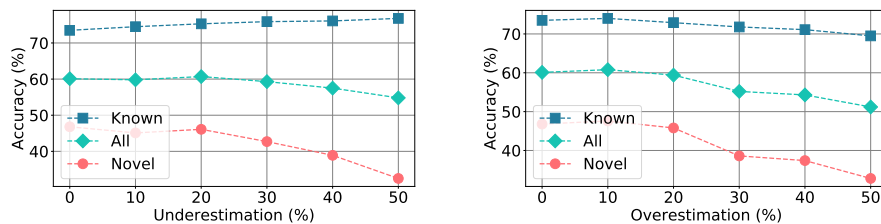
Fig. 4: Performance on **CIFAR-100** dataset with unknown number of novel classes. We set first 50% classes as known and remaining 50% classes as novel.

wide range of estimation errors. We also notice that even with 50% estimation error OpenLDN outperforms ORCA on CIFAR-100 dataset. In summary, these experiments demonstrate that OpenLDN can be applied in a more realistic setup if a reasonable method for estimating the number of novel classes is available.

5   Limitations

In this work, we focus on the more general open-world SSL problem, where the unlabeled set can include unknown class samples. However, our proposed solution is based on a few assumptions. Most notably, following prior works, we assume that the known classes are similar to novel classes, thus they will share some relevant information that can be exploited for discovering novel classes. However, in some extreme cases, this assumption might be violated. Besides, in our entropy regularization term, we encourage the outputs to be uniform. For imbalanced data, this can lead to suboptimal results. One solution is to apply the prior target distribution [52] instead of uniform distribution in the entropy regularization loss. This prompts a new avenue of research that investigate novel methods to estimate prior target distribution in open-world environments.

6   Conclusion

In this work, we propose OpenLDN to solve open-world SSL problem. OpenLDN utilizes a pairwise similarity loss to discover and cluster novel classes by solving a pairwise similarity prediction task. One advantage of our solution is that the proposed pairwise similarity objective does not rely on any additional self-supervised pretraining, but instead, it exploits the information readily available in the labeled set using a bi-level optimization rule. Moreover, our solution brings in a unique perspective towards solving open-world SSL problems by transforming it into a closed-world SSL problem. This insight provides the opportunity to leverage all recent advancements in closed-world SSL methods readily in the context of open-world SSL. Finally, we introduce iterative pseudo-labeling as a simple and effective tool to address the noise present in generated pseudo-labels for novel classes without adding any significant computational overhead. OpenLDN is able to outperform the state-of-the-art open-world SSL methods while incurring a lower computational cost. We demonstrate the superior performance of OpenLDN over a wide range of vision datasets.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), https://www.tensorflow.org/, software available from tensorflow.org 7
2. Arazo, E., Ortego, D., Albert, P., O'Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2020) 3, 7
3. Bard, J.F.: Practical bilevel optimization: algorithms and applications, vol. 30. Springer Science & Business Media (2013) 2, 6
4. Bendale, A., Boult, T.: Towards open world recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1893–1902 (2015) 2
5. Berthelot, D., Carlini, N., Cubuk, E.D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C.: Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. In: International Conference on Learning Representations (2020) 2, 3
6. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. In: Advances in Neural Information Processing Systems 32, pp. 5049–5059. Curran Associates, Inc. (2019) 1, 2, 3, 8, 9
7. Cao, K., Brbic, M., Leskovec, J.: Open-world semi-supervised learning. In: International Conference on Learning Representations (2022), https://openreview.net/forum?id=O-r8LOR-CCA 2, 3, 6, 7, 8, 9, 10, 11, 13
8. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 132–149 (2018) 8
9. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. Advances in Neural Information Processing Systems **33** (2020) 3
10. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017) 1
11. Chang, J., Wang, L., Meng, G., Xiang, S., Pan, C.: Deep adaptive image clustering. In: Proceedings of the IEEE international conference on computer vision. pp. 5879–5887 (2017) 2, 6
12. Chapelle, O., Zien, A.: Semi-supervised classification by low density separation. In: AISTATS. vol. 2005, pp. 57–64. Citeseer (2005) 7
13. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018) 1
14. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. arXiv preprint arXiv:2002.05709 (2020) 2

15. Chen, T., Kornblith, S., Swersky, K., Norouzi, M., Hinton, G.E.: Big self-supervised models are strong semi-supervised learners. Advances in Neural Information Processing Systems **33** (2020) 3

16. Chen, Y., Zhu, X., Li, W., Gong, S.: Semi-supervised learning under class distribution mismatch. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 3569–3576 (2020) 2, 3

17. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009) 9

18. Doersch, C., Gupta, A., Efros, A.A.: Unsupervised visual representation learning by context prediction. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1422–1430 (2015) 2

19. Fini, E., Sangineto, E., Lathuilière, S., Zhong, Z., Nabi, M., Ricci, E.: A unified objective for novel class discovery. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9284–9292 (2021) 4, 8, 9, 10, 13

20. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. Proceedings of Machine Learning Research, vol. 70, pp. 1126–1135. PMLR, International Convention Centre, Sydney, Australia (06–11 Aug 2017), http://proceedings.mlr.press/v70/finn17a.html 2

21. Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: International Conference on Machine Learning. pp. 1183–1192. PMLR (2017) 1

22. Gammerman, A., Vovk, V., Vapnik, V.: Learning by transduction. In: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. p. 148–155. UAI'98, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1998) 3

23. Guo, L.Z., Zhang, Z.Y., Jiang, Y., Li, Y.F., Zhou, Z.H.: Safe deep semi-supervised learning for unseen-class unlabeled data. In: International Conference on Machine Learning. pp. 3897–3906. PMLR (2020) 2, 3, 10

24. Han, K., Rebuffi, S.A., Ehrhardt, S., Vedaldi, A., Zisserman, A.: Automatically discovering and learning new visual categories with ranking statistics. In: International Conference on Learning Representations (2020) 2, 4, 6, 9, 10, 13

25. Han, K., Vedaldi, A., Zisserman, A.: Learning to discover novel visual categories via deep transfer clustering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8401–8409 (2019) 4, 9, 10, 13

26. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9729–9738 (2020) 2

27. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017) 1

28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 1, 9

29. Hsu, Y.C., Lv, Z., Kira, Z.: Learning to cluster in order to transfer across domains and tasks. In: International Conference on Learning Representations (2018), https://openreview.net/forum?id=ByRWCqvT- 4

30. Hsu, Y.C., Lv, Z., Schlosser, J., Odom, P., Kira, Z.: Multi-class classification without multi-class labels. In: International Conference on Learning Representations (2019), https://openreview.net/forum?id=SJzR2iRcK7 4

31. Jia, X., Han, K., Zhu, Y., Green, B.: Joint representation learning and novel category discovery on single-and multi-modal data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 610–619 (2021) 4

32. Joachims, T.: Transductive inference for text classification using support vector machines. In: Icml. vol. 99, pp. 200–209 (1999) 3

33. Kardan, N., Stanley, K.O.: Mitigating fooling with competitive overcomplete output layer neural networks. In: 2017 International Joint Conference on Neural Networks (IJCNN). pp. 518–525. IEEE (2017) 2

34. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) 9

35. Kingma, D.P., Mohamed, S., Rezende, D.J., Welling, M.: Semi-supervised learning with deep generative models. In: Advances in neural information processing systems. pp. 3581–3589 (2014) 3

36. Konyushkova, K., Sznitman, R., Fua, P.: Learning active learning from data. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017), https://proceedings.neurips.cc/paper/2017/file/8ca8da41fe1ebc8d3ca31dc14f5fc56c-Paper.pdf 1

37. Kornblith, S., Shlens, J., Le, Q.V.: Do better imagenet models transfer better? In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2661–2671 (2019) 2

38. Krizhevsky, A., Nair, V., Hinton, G.: Cifar-10 (canadian institute for advanced research) 9

39. Krizhevsky, A., Nair, V., Hinton, G.: Cifar-100 (canadian institute for advanced research) 9

40. Kuhn, H.W.: The hungarian method for the assignment problem. Naval research logistics quarterly 2(1-2), 83–97 (1955) 9

41. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. In: ICLR (Poster). OpenReview.net (2017) 3

42. Le, Y., Yang, X.: Tiny imagenet visual recognition challenge. CS 231N 7(7), 3 (2015) 9

43. Lee, D.H.: Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks (2013) 3, 7

44. Liu, B., Wu, Z., Hu, H., Lin, S.: Deep metric transfer for label propagation with limited annotated data. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 0–0 (2019) 3

45. Miyato, T., ichi Maeda, S., Koyama, M., Ishii, S.: Virtual adversarial training: A regularization method for supervised and semi-supervised learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 41, 1979–1993 (2018) 3

46. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE transactions on pattern analysis and machine intelligence 41(8), 1979–1993 (2018) 1

47. Oliver, A., Odena, A., Raffel, C., Cubuk, E.D., Goodfellow, I.J.: Realistic evaluation of deep semi-supervised learning algorithms. arXiv preprint arXiv:1804.09170 (2018) 2, 3

48. Parkhi, O.M., Vedaldi, A., Zisserman, A., Jawahar, C.V.: Cats and dogs. In: IEEE Conference on Computer Vision and Pattern Recognition (2012) 9

49. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017) 7

50. Pu, Y., Gan, Z., Henao, R., Yuan, X., Li, C., Stevens, A., Carin, L.: Variational autoencoder for deep learning of images, labels and captions. In: Advances in neural information processing systems. pp. 2352–2360 (2016) 3

51. Rizve, M.N., Duarte, K., Rawat, Y.S., Shah, M.: In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=-ODN6SbiUU 3, 7

52. Rizve, M.N., Kardan, N., Shah, M.: Towards realistic semi-supervised learning. In: European Conference on Computer Vision (2022) 14

53. Rizve, M.N., Khan, S., Khan, F.S., Shah, M.: Exploring complementary strengths of invariant and equivariant representations for few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10836–10846 (2021) 2

54. Sajjadi, M., Javanmardi, M., Tasdizen, T.: Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems 29, pp. 1163–1171. Curran Associates, Inc. (2016) 3

55. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A coreset approach. In: International Conference on Learning Representations (2018), https://openreview.net/forum?id=H1aIuk-RW 1

56. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 806–813 (2014) 2

57. Shi, W., Gong, Y., Ding, C., MaXiaoyu Tao, Z., Zheng, N.: Transductive semi-supervised deep learning using min-max features. In: The European Conference on Computer Vision (ECCV) (September 2018) 3

58. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. In: Advances in neural information processing systems. pp. 4077–4087 (2017) 2

59. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 596–608. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper/2020/file/06964dce9addb1c5cb5d6e3d9838f733-Paper.pdf 2, 3, 8, 10

60. Sun, X., Yang, Z., Zhang, C., Ling, K.V., Peng, G.: Conditional gaussian distribution learning for open set recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13480–13489 (2020) 10

61. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826 (2016) 1

62. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. arXiv preprint arXiv:1703.01780 (2017) 1

63. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30, pp. 1195–1204. Curran Associates, Inc. (2017) 3, 8

64. Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M., Van Gool, L.: Scan: Learning to classify images without labels. In: European Conference on Computer Vision. pp. 268–285. Springer (2020) 4

65. Verma, V., Kawaguchi, K., Lamb, A., Kannala, J., Bengio, Y., Lopez-Paz, D.: Interpolation consistency training for semi-supervised learning. arXiv preprint arXiv:1903.03825 (2019) 7

66. Vinyals, O., Blundell, C., Lillicrap, T., kavukcuoglu, k., Wierstra, D.: Matching networks for one shot learning. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems 29, pp. 3630–3638. Curran Associates, Inc. (2016), http://papers.nips.cc/paper/6385-matching-networks-for-one-shot-learning.pdf 2

67. Wu, J., Long, K., Wang, F., Qian, C., Li, C., Lin, Z., Zha, H.: Deep comprehensive correlation mining for image clustering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8150–8159 (2019) 2, 6

68. Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: International conference on machine learning. pp. 478–487. PMLR (2016) 4

69. Xie, Q., Dai, Z., Hovy, E., Luong, M.T., Le, Q.V.: Unsupervised data augmentation for consistency training. arXiv preprint arXiv:1904.12848 (2019) 2, 8, 9

70. Yang, B., Fu, X., Sidiropoulos, N.D., Hong, M.: Towards k-means-friendly spaces: Simultaneous deep learning and clustering. In: international conference on machine learning. pp. 3861–3870. PMLR (2017) 4

71. Yang, J., Parikh, D., Batra, D.: Joint unsupervised learning of deep representations and image clusters. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5147–5156 (2016) 4

72. YM., A., C., R., A., V.: Self-labelling via simultaneous clustering and representation learning. In: International Conference on Learning Representations (2020), https://openreview.net/forum?id=Hyx-jyBFPr 8

73. Zamir, A.R., Sax, A., Shen, W., Guibas, L.J., Malik, J., Savarese, S.: Taskonomy: Disentangling task transfer learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3712–3722 (2018) 2

74. Zhao, B., Han, K.: Novel visual category discovery with dual ranking statistics and mutual knowledge distillation. Advances in Neural Information Processing Systems **34**, 22982–22994 (2021) 4

75. Zhao, X., Krishnateja, K., Iyer, R., Chen, F.: Robust semi-supervised learning with out of distribution data. arXiv preprint arXiv:2010.03658 (2020) 2, 3

76. Zhong, Z., Fini, E., Roy, S., Luo, Z., Ricci, E., Sebe, N.: Neighborhood contrastive learning for novel class discovery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10867–10875 (2021) 4

77. Zhong, Z., Zhu, L., Luo, Z., Li, S., Yang, Y., Sebe, N.: Openmix: Reviving known knowledge for discovering novel visual categories in an open world. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9462–9470 (2021) 4