Towards Realistic Semi-Supervised Learning

Mamshad Nayeem Rizve, Navid Kardan, and Mubarak Shah

Center for Research in Computer Vision, UCF, USA
{nayeemrizve, kardan}@knights.ucf.edu, shah@crcv.ucf.edu

Abstract. Deep learning is pushing the state-of-the-art in many computer vision applications. However, it relies on large annotated data repositories, and capturing the unconstrained nature of the real-world data is yet to be solved. Semi-supervised learning (SSL) complements the annotated training data with a large corpus of unlabeled data to reduce annotation cost. The standard SSL approach assumes unlabeled data are from the same distribution as annotated data. Recently, a more realistic SSL problem, called open-world SSL, is introduced, where the unannotated data might contain samples from unknown classes. In this paper, we propose a novel pseudo-label based approach to tackle SSL in open-world setting. At the core of our method, we utilize sample uncertainty and incorporate prior knowledge about class distribution to generate reliable class-distribution-aware pseudo-labels for unlabeled data belonging to both known and unknown classes. Our extensive experimentation showcases the effectiveness of our approach on several benchmark datasets, where it substantially outperforms the existing state-of-the-art on seven diverse datasets including CIFAR-100 ($\sim$17%), ImageNet-100 ($\sim$5%), and Tiny ImageNet ($\sim$9%). We also highlight the flexibility of our approach in solving novel class discovery task, demonstrate its stability in dealing with imbalanced data, and complement our approach with a technique to estimate the number of novel classes. Code: https://github.com/nayeemrizve/TRSSL

Keywords: Semi-supervised learning, Open-world, Uncertainty

1 Introduction

Deep learning systems have made tremendous progress in solving many challenging vision problems [25,24,11,20,54,1]. However, most of this progress has been made in controlled environments, which limits their application in real-world scenarios. For instance, in classification, we should know all the classes in advance. However, many real-world problems cannot be expressed with this constraint, where we constantly encounter new concepts while exploring an unconstrained environment. A practical learning model should be able to properly detect and handle new situations. Open-world problems [59,4,33,23,9,29,34] try to model this unconstrained nature of real-world data.

Despite abundance of real-world data, it is often required to annotate raw data before passing it to supervised models, which is quiet costly. One of the

dominant approaches to reduce the cost of annotation is semi-supervised learning (SSL) [63,6,43,49,62], where the objective is to leverage a set of unlabeled data in conjunction with a limited labeled set to improve performance. Following [9], in this work, we consider the unlabeled set to possibly contain samples from unknown (novel) classes that are not present in the labeled set. This problem is called open-world SSL [9]. Here, the goal is to identify novel-class samples and classify them, as well as to improve known-class performance by utilizing unlabeled known-class samples.

At first sight, the major difficulty with open-world SSL might be related to breaking the closed-world assumption. In fact, it is common knowledge that presence of samples from novel classes deteriorates the performance of standard SSL methods drastically [51,14]. This leads to introduction of new approaches that mitigate this issue based on identifying, and subsequently reducing the effect of novel class samples to generalize SSL to more practical settings [21,14,69]. However, open-world SSL requires *identifying and assigning samples to novel classes*, which contrasts with this simpler objective of ignoring them. To the best of our knowledge ORCA [9] is the only prior work that proposes a solution for this challenging problem, where the authors also demonstrate that open-world SSL problem cannot be solved by simple extensions of existing SSL approaches.

Improving upon ORCA, this paper introduces a streamlined approach for open-world SSL problem, which does not require careful design choices for multiple objectives, and does not rely on feature initialization. Our approach substantially improves state-of-the-art performance on multiple datasets (Fig. 1). Furthermore, distinctly from previous work, our algorithm can naturally handle arbitrary class distributions such as imbalanced data. Finally, we propose a method to estimate the number of unknown classes for more practical applications.
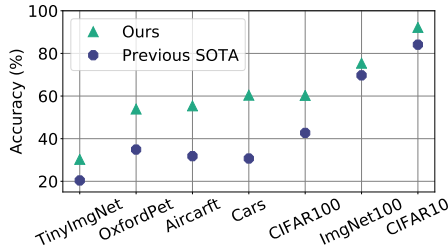

Fig. 1: Performance of our proposed method with respect to previous SOTA method on Tiny ImageNet, Oxford-IIIT Pet, FGVC-Aircraft, Stanford-Cars, CIFAR-100, ImageNet-100, and CIFAR-10 datasets respectively.

For solving the open-world SSL problem, we employ an intuitive pseudo-labeling approach. Our pseudo-label generation process takes different challenges associated with the open-world SSL problem—simultaneously classifying samples from both known and unknown classes, and handling arbitrary class distribution—into account. Furthermore, we incorporate sample uncertainty into pseudo-label learning to address the unreliable nature of generated pseudo-labels. We make two major technical contributions in this work: (1) we propose a novel pseudo-label generation method, which takes advantage of the prior knowledge about class distribution and generate pseudo-labels accordingly using Sinkhorn-Knopp algorithm [61,10,67,2]. Our proposed solution can take advantage of any arbitrary data distribution which includes imbalanced distributions. (2) we in-

troduce a novel uncertainty-guided temperature scaling technique to address the unreliable nature of the generated pseudo-labels. Additionally, we propose a simple yet effective method for estimating the number of novel classes, allowing for a more realistic application of our method. Our extensive experimentation on four standard benchmark datasets and also three additional fine-grained datasets demonstrate that the proposed method significantly outperforms the existing works (Fig. 1). Finally, our experimentation with data imbalance (Sec. 4.3) signifies that the proposed method can work satisfactorily even when no prior knowledge is available about the underlying class distribution.

2   Related Works

Open-World Learning To address the unconstrained nature of real-world data, multiple research directions have been explored. In this work, we refer to all these different approaches as open-world learning method. Open-set recognition (OSR) [59,27,45], open-world recognition (OWR) [4,7,66,29], out-of-distribution detection [45,44,68,32,31], and novel class discovery (NCD) [26,23,22,70,17] are some of the notable open-world learning approaches.

Open-set recognition methods aim to identify novel class samples during inference to avoid assigning them to one of the known/seen classes. One of the early works on OSR was proposed in [59], where a one-vs-all strategy was applied to prevent assigning novel class samples to known classes. [27] extends OSR to multi-class setup by using probabilistic modeling to adjust the classification boundary. Instead of designing robust models for OSR, ODIN [45] detects novel class samples (out-of-distribution) based on difference in output probabilities caused by changing the softmax temperature and adding small controlled perturbations to the inputs. Even though OSR is a related problem, the focus of this work is more general where our goal is to not only detect novel class samples but also to cluster them.

OWR methods such as [4] work in an incremental manner, where once the model determines instances from novel classes an oracle can provide class labels for unknown samples to incorporate them into the seen set. To incorporate new classes, [66] maintains a dynamic list of exemplar samples for each class, and unknown examples are detected by finding the similarity with these exemplars. Finally, authors in [29] propose contrastive clustering and energy based unknown sample detection for open-world object detection. The key difference between these methods and ours is that we do not rely on an oracle to learn novel classes.

NCD methods are most closely related to our task. The main objective of NCD methods is to cluster novel class samples in the unlabeled set. To this end, authors in [26] leverage the information available in the seen classes by training a pairwise similarity prediction network that they later apply to cluster novel class samples. Similar to their approach, a pairwise similarity task is solved to discover novel classes based on a novel rank statistics in [22]. Most NCD methods rely on multiple objective functions and require some sort of feature pretraining approach. This is addressed in [17] by utilizing multi-view pseudo-

labeling and overclustering while only relying on cross-entropy loss. The main difference between NCD problem and our task is that we do not assume unlabeled data only includes novel class samples. Besides, in contrast to most of these methods, our proposed solution requires only one loss function and does not make architectural changes to treat seen and novel classes differently. Additionally, our extensive experimentation demonstrates that extension of these methods is not very effective for open-world SSL problem.

Semi-Supervised Learning Extensive research has been conducted on closed-world SSL [19,28,46,36,53,13,10,58,41,48,63,43,60,6,5,62]. The closed-world SSL methods achieve impressive performance on standard benchmark datasets. However, these methods assume that the unlabeled data only contains samples from seen classes, which is very restrictive. Moreover, recent works [51,14] suggest that presence of novel class samples deteriorates performance of SSL methods. Robust SSL methods [21,14,69] address this issue by filtering out or reweighting novel class samples. The realistic open-world SSL problem as proposed in [9] requires clustering the novel class samples which is not addressed by robust SSL methods. To the best of our knowledge, ORCA [9] is the only prior work that solves this challenging problem. ORCA achieves very promising performance in comparison to other novel class discovery or robust SSL based baselines. However, to solve this problem ORCA leverages self-supervised pretraining and multiple objective functions. In a concurrent work, [56] proposes a solution to open-world SSL that does not rely on feature pretraining. However, similar to ORCA, their approach relies on multiple objectives. In contrast, our proposed solution outperforms ORCA by a large margin without relying on either of them.

3   Method

Similar to standard closed-world SSL, the training data for open-world SSL problem consists of a labeled set, $\mathbb{D}_L$, and an unlabeled set, $\mathbb{D}_U$. The labeled set, $\mathbb{D}_L$ encompasses N_L labeled samples s.t. $\mathbb{D}_L = \left\{ \mathbf{x}_l^{(i)}, \mathbf{y}_l^{(i)} \right\}_{i=1}^{N_L}$, where $\mathbf{x}_l^{(i)}$ is an input and $\mathbf{y}_l^{(i)}$ is its corresponding label (in one-hot encoding) belonging to one of the $\mathbb{C}_L$ classes. On the other hand, the unlabeled set, $\mathbb{D}_U$, consists of N_U (in practice, $N_U \gg N_L$) unlabeled samples s.t. $\mathbb{D}_U = \left\{ \mathbf{x}_u^{(i)} \right\}_{i=1}^{N_U}$, where $\mathbf{x}_u^{(i)}$ is a sample without any label that belongs to one of the $\mathbb{C}_U$ classes. The primary distinction between the closed-world and open-world SSL formulation is that the closed-world SSL assumes $\mathbb{C}_L = \mathbb{C}_U$, whereas in open-world SSL $\mathbb{C}_L \subset \mathbb{C}_U$. We refer to $\mathbb{C}_U \setminus \mathbb{C}_L$, as novel classes, $\mathbb{C}_N$. Note that unlike previous works on novel class discovery problem [22,17,70], we do not need to know the number of novel classes, $|\mathbb{C}_N|$, in advance. During test time, the objective is to assign samples from novel classes to their corresponding novel class in $\mathbb{C}_N$, and to classify the samples from seen classes into one of the $|\mathbb{C}_L|$ classes.

In the following subsections, we first introduce our class-distribution-aware pseudo-label based training objective to classify the samples from seen classes, while attributing the samples from novel classes to their respective categories (Sec. 3.1). After that, we introduce uncertainty-guided temperature scaling to incorporate reliability of pseudo-labels into the learning process (Sec. 3.2).
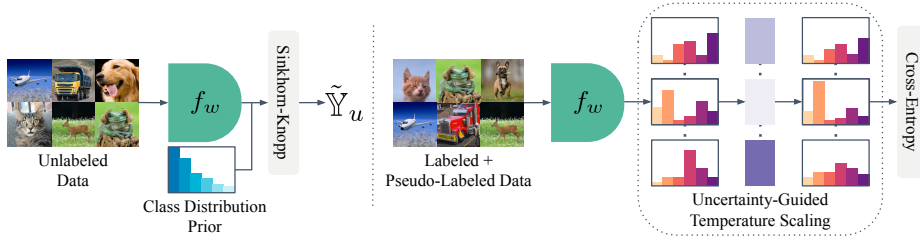
Fig. 2: *Training Overview:* **Left:** generating pseudo-labels. Our model generates pseudo-labels for the unlabeled samples using Sinkhorn-Knopp while taking class distribution prior into account. **Right:** reliable training with both labeled and unlabeled samples. We use the ground-truth labels and generated pseudo-labels to train in a supervised manner. To address the unreliable nature of pseudo-labels in open-world SSL, we apply uncertainty-guided temperature scaling (darker color refers to higher uncertainty).

3.1 Class-Distribution-Aware Pseudo-Labeling

To achieve the dual objective of open-world SSL problem, i.e., identifying samples from the seen classes and clustering the samples from novel classes, we design a single classification objective. To this end, we utilize a neural network, f_w, to map the input data $\mathbf{x}$ into the output space of class scores (logits), $\mathbf{z} \in \mathbb{R}^{|\mathbb{C}_L|+|\mathbb{C}_N|}$, s.t. $f_w : \mathbb{X} \to \mathbb{Z}$; here, $\mathbb{X}$ is the set of input data and $\mathbb{Z}$ is the set of output logits. In our setup, the first $|\mathbb{C}_L|$ entries of the class score vector (logits), $\mathbf{z}$, correspond to seen classes and the remaining $|\mathbb{C}_N|$ elements correspond to novel classes. Finally, we transform these logits to probability distribution, $\hat{\mathbf{y}}$, using softmax activation function: $\hat{\mathbf{y}}_j = \exp(\mathbf{z_j}) / \sum_k \exp(\mathbf{z_k})$.

The neural network, f_w, can be trained using cross-entropy loss if the labels for all the input samples are available. However, in open-world SSL problem the samples in $\mathbb{D}_U$ lack label. To address this issue, pseudo-labels, $\tilde{\mathbf{y}}_u \in \hat{\mathbb{Y}}_u$, are generated for all unlabeled samples. After that, cross entropy loss is applied to train the model using the available ground-truth labels, $\mathbf{y}_l \in \mathbb{Y}_l$, and generated pseudo-labels. Here, we assume one-hot encoding for $\mathbf{y}_l$ and $\mathbb{Y}$ denotes the set of all labels, where $\mathbb{Y} = \mathbb{Y}_l \cup \hat{\mathbb{Y}}_u$. Now, the cross-entropy loss is defined using,

$$\mathcal{L}_{ce} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} \mathbf{y}_j^{(i)} \log \hat{\mathbf{y}}_j^{(i)}, \tag{1}$$

where, $C = |\mathbb{C}_L| + |\mathbb{C}_N|$ is total number of classes, $N = N_L + N_U$ is the total number of samples, $\mathbf{y} \in \mathbb{Y}$, and $\mathbf{y}_j^{(i)}$ is the jth element of the class label vector, $\mathbf{y}^{(i)}$, for training instance i.

Next, we discuss the class-distribution-aware pseudo-label generation process. Since pseudo-label generation process is inherently ill-posed, we can guide this process by injecting an inductive bias. To this end, we propose to generate pseudo-labels in such a way that the class distribution of generated pseudo-labels should follow the underlying class distribution of samples. More formally, we enforce the following constraint:

$$\forall j \sum_i^{N_U} \tilde{\mathbf{y}}_j^{(i)} = N_U^{C_j}, \tag{2}$$

where, $N_U^{C_j}$ is the number of samples in jth class.

One common strategy to satisfy this objective is to apply an entropy max-imization term coupled with optimizing a pairwise similarity score objective [9,64]. However, this approach implicitly assumes that the classes are balanced and optimizing the pairwise objective requires a good set of initial features; be-sides, coordinating these two objectives requires careful design. This paper pur-sues a more streamlined approach by generating pseudo-labels such that they directly satisfy the constraints in Eq. 2. Fortunately, this constrained pseudo-label generation problem is inherently a transportation problem [30,8], where we want to assign unlabeled samples to one of the seen/novel classes based on output probabilities. Such an assignment can be captured with an assignment matrix, $\mathbf{A}$, which can be interpreted as (normalized) pseudo-labels. Following Cuturi's notation [15], every such assignment $\mathbf{A}$, called a transport matrix, that satisfies the constraint in Eq. 2 is a member of a transportation polytope, $\mathcal{A}$.

$$\mathcal{A} := \left\{ \mathbf{A} \in \mathbb{R}^{N_U \times C} | \forall j \sum \mathbf{A}_{:,j} = \frac{N_U^{C_j}}{N_U}, \forall i \sum \mathbf{A}_{i,:} = \frac{1}{N_U} \right\}. \tag{3}$$

Note that every transport matrix $\mathbf{A}$ is a joint probability, therefore, it is a normalized matrix. By considering the cross-entropy cost of assigning unlabeled samples based on model predictions to different classes, an optimal solution can be found within the transportation polytope $\mathcal{A}$. More formally, we solve $\min_{\mathbf{A} \in \mathcal{A}} -Tr(\mathbf{A}^T \log(\hat{\mathbf{Y}}_U/N_U))$ optimization problem, where $\hat{\mathbf{Y}}_U$ is the matrix of output probabilities generated by the model for the unlabeled samples. Unfor-tunately, enforcing the constraint described in Eq. 2 is non-trivial for novel classes since we do not know the specific order of novel classes. To address this issue, we need to solve a permutation problem while obtaining the optimal assignment matrix, $\mathbf{A}$. To this end, we introduce a permutation matrix $\mathbf{P}_\pi$ and reformulate the optimization problem as $\min_{\mathbf{A} \in \mathcal{A}} -Tr((\mathbf{A}\mathbf{P}_\pi)^T \log(\hat{\mathbf{Y}}_U/N_U))$. Here, the per-mutation matrix $\mathbf{P}_\pi$ reorders the columns of the assignment matrix. We estimate the permutation matrix $\mathbf{P}_\pi$ from the order of the marginal of output probabili-ties $\hat{\mathbf{Y}}_U$. This simple reordering ensures that per class constraint is aligned with the output probabilities. After determining the permutation, finding the optimal solution for $\mathbf{A}$ becomes an instance of the optimal transport problem. Hence, can be solved using Sinkhorn-Knopp algorithm. Cuturi [15] proposes a fast version of Sinkhorn-Knopp algorithm. In particular, [15] shows that a fast estimation of the optimal assignment can be obtained by:

$$\mathbf{A} = \text{diag}(\mathbf{m})(\hat{\mathbf{Y}}_U/N_U)^\lambda \text{diag}(\mathbf{n}), \tag{4}$$

where λ is a regularization term that controls the speed of convergence versus precision of the solution, vectors $\mathbf{m}$ and $\mathbf{n}$ are used for scaling $\hat{\mathbf{Y}}_U/N_U$ so that the transportation matrix $\mathbf{A}$ is also a probability matrix. This is an itereative procedure where $\mathbf{m}$ and $\mathbf{n}$ are updated according to the following rules:

$$\mathbf{m} \leftarrow [(\hat{\mathbf{Y}}_U/N_U)^\lambda \mathbf{n}]^{-1}, \mathbf{n} \leftarrow [\mathbf{m}^T (\hat{\mathbf{Y}}_U/N_U)^\lambda]^{-1}. \tag{5}$$

Another aspect of our pseudo-label generation is inducing perturbation invariant features. Generally learning invariant features is achieved by minimizing a consistency loss that minimizes the distance between the output representation of two transformed versions of the same image [58,6,65]. To achieve this, for the unlabeled data, given image $\mathbf{x}$, we generate two augmented versions of this image, $\mathbf{x}_{\tau_1} = \tau_1(\mathbf{x})$, and $\mathbf{x}_{\tau_2} = \tau_2(\mathbf{x})$, where $\tau_1(.)$, and $\tau_2(.)$ are two stochastic transformations. The generated pseudo-labels for these two augmented images are $\tilde{\mathbf{y}}_{\tau_1}$, and $\tilde{\mathbf{y}}_{\tau_2}$, respectively. To learn transformation invariant representation using cross-entropy loss, we treat $\tilde{\mathbf{y}}_{\tau_2}$ as the corresponding pseudo-label of $\mathbf{x}_{\tau_1}$ and vice versa. This cross pseudo-labeling encourages learning of perturbation invariant features without introducing a new loss function.

Finally, in its original formulation Sinkhorn-Knopp algorithm generates hard pseudo-labels [15]. However, recent literature [10] reports better performance by applying soft pseudo-labels. In our work we utilize a mixture of soft and hard pseudo-labels (mixed pseudo-labels), which we found to be beneficial (Sec. 4.3). To be specific, to encourage confident learning for novel classes, we generate hard pseudo-labels for unlabeled samples which are strongly assigned to novel classes. For the rest of the unlabeled samples, we use soft pseudo-labels.

3.2 Uncertainty-Guided Temperature Scaling

Since we generate pseudo-labels by relying on the confidence scores of the network, final performance is affected by their reliability. We can capture the reliability of prediction confidences by measuring their uncertainty. One simple way to do that in the standard neural networks is to perform Monte Carlo sampling in the network parameter space [18] or in the input space [3,55]. Since we do not want to modify the network parameters, we decide to perform stochastic sampling in input space. To this end, we apply stochastic transformations on input data and estimate the sample uncertainty, $u(.)$, by calculating the variance over the applied stochastic transformations [16,50,55]:

$$u(\mathbf{x}) = \text{Var}(\hat{\mathbf{y}}) = \frac{1}{\mathcal{T}} \sum_{i=1}^{\mathcal{T}} (\hat{\mathbf{y}}_{\tau_i} - \text{E}(\hat{\mathbf{y}}))^2, \tag{6}$$

where, $\hat{\mathbf{y}}_{\tau_i} = \text{Softmax}(f_w(\tau_i(\mathbf{x})))$, $\tau_i(.)$ represents a stochastic transformation applied to the input $\mathbf{x}$, and $\text{E}(\hat{\mathbf{y}}) = \frac{1}{\mathcal{T}} \sum_{i=1}^{\mathcal{T}} \hat{\mathbf{y}}_{\tau_i}$.

Next, we want to incorporate this uncertainty information into our training process. One strategy to achieve this is to select more reliable pseudo-labels by filtering out unreliable samples based on their uncertainty score [55]. However, two potential drawbacks of this approach are introducing a new hyperparameter and discarding a portion of available data. Therefore, to tackle both of these drawbacks, we introduce uncertainty-guided temperature scaling.

Recall that in our training we use softmax probabilities for cross-entropy loss. Temperature scaling is a strategy to modify the softness of the output probability distribution. In standard softmax probability computation, the temperature

value is set to 1. A higher value of temperature increases the entropy or uncertainty of the softmax probability, whereas a lower value makes it more certain. Existing works [9,17,12,35] apply a fixed temperature value (whether high or low) as a hyperparameter. In contrast, we propose to use a different temperature for each sample during the course of training which is influenced by the certainty of its pseudo-label. The main idea is that if the network is certain about its prediction on a particular sample we make this prediction more confident and vice versa. Based on this idea we modify the softmax probability computation in the following way:

$$\hat{\mathbf{y}}_j^{(i)} = \frac{\exp(\mathbf{z}_j^{(i)}/\mathrm{u}(\mathbf{x}^{(i)}))}{\sum_k \exp(\mathbf{z}_k^{(i)}/\mathrm{u}(\mathbf{x}^{(i)}))}, \tag{7}$$

where $\mathrm{u}(\mathbf{x}^{(i)})$ is the uncertainty of sample $\mathbf{x}^{(i)}$ that is obtained from Eq. 6.

In practice, the sample uncertainties calculated by Eq. 6 have low magnitudes. Therefore, we normalize these uncertainty values across the entire dataset before plugging them into Eq. 7.

Our training algorithm is provided in supplementary materials.

4   Experiments and Results

4.1   Experimental Setup

In the following, we describe our experimental setup including dataset descriptions, implementation details, evaluation details, and specifics of our baselines.

Datasets We conduct experiments on four commonly used computer vision benchmark datasets: CIFAR-10 [38], CIFAR-100 [39], ImageNet-100 [57] and Tiny ImageNet [42]. The datasets are selected in increasing order of difficulty based on the number of classes. We also evaluate our method on three drastically different fine-grained classification datasets: Oxford-IIIT Pet [52], FGVC-Aircraft [47], and Stanford-Cars [37]. A detailed description of these datasets is provided in supplementary materials. For all the datasets, we use the first 50% classes as seen and the remaining 50% classes as novel. We use 10% data from the seen classes as the labeled set and use the remaining 90% data, along with the samples from novel classes, as unlabeled set for our experiments on standard benchmark datasets. For fine-grained datasets, we use 50% data from seen classes as labeled. Additional results with other data percentage are provided in the supplementary materials.

Implementation Details Following ORCA [9], for a fair comparison, we use ResNet-50 [25] for ImageNet-100 experiments and ResNet-18 [25] for all the other experiments. We apply l_2 normalization to the weights of the last linear layer. For CIFAR-10, CIFAR-100, and Tiny ImageNet experiments, we train our model for 200 epochs. For the other datasets, we train these model for 100 epochs. We use a batchsize of 256 for all of our experiments except ImageNet-100 where similar to [9] we use a batchsize of 512. For optimizing the network parameters, we use SGD optimizer with momentum. We use a cosine annealing based learning rate

scheduler accompanied by a linear warmup, where we set the base learning rate to 0.1 and set the warmup length to 10 epochs. For network parameters, we set the weight decay to 1e-4. Following [10], we set the value of λ to 0.05 and perform 3 iterations for pseudo-label generation using the Sinkhorn-Knopp algorithm. Additional implementation details are provided in supplementary materials.

Evaluation Details For evaluation, we report standard classification accuracy on seen classes. On novel classes, we report clustering accuracy following [9,22,17,23]. To this end, we consider the class prediction as cluster ID. Next, we use the Hungarian algorithm [40] to match cluster IDs with ground-truth classes. Once the matches are obtained, we calculate classification accuracy with the corresponding cluster IDs. Besides, if a novel class sample gets assigned to one of the seen classes, we consider that as a misclassified prediction and remove that sample before matching the cluster IDs with ground-truth class labels. We also report clustering accuracy for all the classes.

Comparison Details We compare the performance of our method on CIFAR-10, CIFAR-100, and ImageNet-100 datasets with the results reported in [9]. The remaining four datasets do not have any publicly available evaluation for open-world SSL problem. Therefore, we extend three recent novel class discovery methods [22,23,17] to open-world SSL setting using publicly available codebase. For [22,23], we extend the unlabeled head to include logits for seen classes by following [9]. However, neither of these methods has any explicit classification loss for seen classes in the unlabeled head. Therefore, there is no straightforward way to map the seen class samples into their corresponding class logits. For reporting scores on seen classes, we use the Hungarian algorithm for these two methods. In [17], pseudo-labels are generated for the novel class samples on the unlabeled head. To make it compatible with open-world SSL setting, we generate pseudo-labels from the concatenated prediction of the labeled and unlabeled heads during training. Since this method has explicit classification loss, we report standard classification accuracy on seen classes.

4.2   Main Results

Standard Benchmark Datasets We compare our method with existing literature on open-world SSL problem [9] and other related approaches that have been modified for this problem in Tab. 1 and 2. On CIFAR-10 we observe that our proposed method outperforms ORCA [9] on both seen and novel classes by 12.1% and 4.1%, respectively. Our proposed method also outperforms other novel class discovery methods [23,22,17] by a large margin. The same trend is observed for FixMatch [62] (a state-of-the-art closed-world SSL method). Finally, our proposed method outperforms DS^{3}L[21], a popular robust SSL method. Interestingly, improvement of our proposed method is more prominent on CIFAR-100 dataset, which is more challenging because of the higher number of classes. On CIFAR-100 dataset, our proposed method outperforms ORCA by around 20% on novel classes and 16% on seen classes. Noticeably, we observe that UNO[17] marginally outperforms ORCA on this dataset. However, our proposed method

Table 1: Average accuracy on the **CIFAR-10**, **CIFAR-100**, and **ImageNet-100** datasets with 50% classes as seen and 50% classes as novel. The results are averaged over three independent runs.

Method	CIFAR-10			CIFAR-100			ImageNet-100		
	Seen	Novel	All	Seen	Novel	All	Seen	Novel	All
FixMatch[62]	64.3	49.4	47.3	30.9	18.5	15.3	60.9	33.7	30.2
DS³L[21]	70.5	46.6	43.5	33.7	15.8	15.1	64.3	28.1	25.9
DTC[23]	42.7	31.8	32.4	22.1	10.5	13.7	24.5	17.8	19.3
RankStats[22]	71.4	63.9	66.7	20.4	16.7	17.8	41.2	26.8	37.4
UNO[17]	86.5	71.2	78.9	53.7	33.6	42.7	66.0	42.2	53.3
ORCA[9]	82.8	85.5	84.1	52.5	31.8	38.6	83.9	60.5	69.7
Ours	94.9	89.6	92.2	68.5	52.1	60.3	82.6	67.8	75.4

Table 2: Average accuracy on the **Tiny ImageNet**, **Oxford-IIIT Pet**, **FGVC-Aircraft**, and **Stanford-Cars** datasets with 50% classes as seen and 50% classes as novel. The results are averaged over three independent runs.

Method	Tiny ImageNet			Oxford-IIIT Pet			FGVC-Aircraft			Stanford-Cars		
	Seen	Novel	All	Seen	Novel	All	Seen	Novel	All	Seen	Novel	All
DTC[23]	13.5	12.7	11.5	20.7	16.0	13.5	16.3	16.5	11.8	12.3	10.0	7.7
RankStats[22]	9.6	8.9	6.4	12.6	11.9	11.1	13.4	13.6	11.1	10.4	9.1	6.6
UNO[17]	28.4	14.4	20.4	49.8	22.7	34.9	44.4	24.7	31.8	49.0	15.7	30.7
Ours	39.5	20.5	30.3	70.9	36.1	53.9	69.5	41.2	55.4	83.5	37.1	60.4

outperforms UNO by a significant margin. Next, we evaluate on two variants of ImageNet: ImageNet-100, and Tiny ImageNet. We observe a similar trend on ImageNet-100 dataset, where we observe an overall improvement of 5.7% over ORCA. After that, we conduct experiments on challenging Tiny ImageNet dataset. This dataset is more challenging than CIFAR-100 and ImageNet-100 dataset since it has 200 classes. Besides, without transfer learning, even the performance of supervised methods is relatively low on this dataset. Overall, our proposed method outperforms the second best method, UNO, by 9.9%, which is almost 50% relative improvement on this challenging dataset. The results on these four datasets demonstrate that *the proposed method not only outperforms previous methods but also excels in scenarios where the number of classes is significantly higher which is always a challenge for clustering methods.*

Fine-Grained Datasets Finally, we evaluate our method on three fine-grained classification datasets with different number of classes. This evaluation is particularly important since fine-grained classification captures challenges associated with many real-world applications. We hypothesize that, fine-grained classification is a harder problem for open-world semi-supervised learning since the novel classes are visually similar to seen classes. In these experiments we compare the performance of the proposed method with three novel class discovery methods, DTC[23], RankStat [22], and UNO[17]. We report our results in Tab. 2. Once again our method outperforms all three methods on these fine-grained classifi-

cation datasets by a significant margin. To be specific, in overall, the proposed method achieves 50-100% relative improvement compared to the second best method UNO. Together, our previous results combined with these fine-grained results, showcase the efficacy of our proposed method and indicate a wider application for more practical settings.

Table 3: Ablataion study on **CIFAR-10**, **CIFAR-100**, and **Tiny ImageNet** datasets with 50% classes as seen and 50% classes as novel. Here, **UTS** refers to uncertainty-guided temperature scaling, **MPL** refers to mixed pseudo-labeling, and **Oracle** refers to having prior knowledge about the number of novel classes.

UTS	MPL	Oracle	CIFAR-10			CIFAR-100			Tiny ImageNet		
			Seen	Novel	All	Seen	Novel	All	Seen	Novel	All
✗	✗	✓	96.0	84.4	90.2	69.2	46.5	57.9	38.1	17.5	28.1
✓	✗	✓	95.0	86.6	90.8	69.4	46.6	57.9	41.3	16.0	29.2
✗	✓	✓	95.8	87.9	91.9	66.9	48.1	57.5	34.9	21.0	28.2
✓	✓	✗	94.9	89.6	92.2	65.5	44.2	54.8	40.3	19.3	30.2
✓	✓	✓	94.9	89.6	92.2	68.5	52.1	60.3	39.5	20.5	30.3

4.3 Ablation and Analysis

To investigate the impact of different components, we conduct extensive ablation study on CIFAR-10, CIFAR-100, and Tiny ImageNet datasets. We report the results in Tab. 3. The first row depicts the performance of our proposed method without uncertainty-guided temperature scaling, and mixed pseudo-labeling. Here, we can see that our proposed method is able to achieve reasonable performance solely based on distribution-aware pseudo-labels. Next, we investigate the impact of removing mixed pseudo-labeling. We observe that the performance on novel classes drops considerably; 3% on CIFAR-10, 5.5% on CIFAR-100, and 4.5% on the Tiny ImageNet dataset. This shows that mixed pseudo-labeling encourages confident learning for novel classes and is a crucial component of our method. After that, we investigate the effect of uncertainty-guided temperature scaling. We observe that the overall performance on all three datasets drops from 0.3%-2.8%. We also observe that the performance degradation is more severe on harder datasets (6.9% relative degradation on Tiny ImageNet compared to 4.6% on CIFAR-100). Next, we report scores with the estimated number of novel classes (Sec. 4.3) for completeness (without Oracle in Tab. 3). We observe that even with the estimated number of novel classes, our method greatly outperforms ORCA and UNO. Our ablation study as a whole demonstrates that every component of our proposed method is crucial and makes a noticeable contribution to the final performance while achieving their designated goal.

Estimating Number of Novel Classes A realistic semi-supervised learning system should make minimal assumption about the nature of the problem. For open-world SSL problem, determining the number of novel classes is a crucial step since without explicit determination of the number of classes either a

Table 4: Estimation of the number of novel classes. The table shows the estimated number of classes vs the actual number of classes in different datasets.

Dataset	GT	Estimated	Error
CIFAR-10	10	10	0%
CIFAR-100	100	117	17%
ImageNet-100	100	139	39%
Tiny ImageNet	200	192	−4%

method will have to assume that the number of novel classes is known in advance or set an upper limit for the number of novel classes. A more practical approach is to estimate the number of unkown classes. Therefore, this work proposes a solution to explicitly estimate the number of novel classes. To this end, we leverage self-supervised features from SimCLR [12].

To estimate the number of novel classes, we perform k-means clustering on SimCLR features with different values of k. We determine the optimal k by evaluating the performance of generated clusters on the labeled samples. We empirically find that this approach generally underestimates the number of novel classes. This is to be expected since clustering accuracy usually decreases with

Fig. 3: Accuracy as a function of class estimation error on **CIFAR-100** dataset.

increasing number of clusters due to assignment of labeled samples to unknown clusters. To mitigate this issue, we perform a sample reassignment technique, where we reallocate the labeled samples assigned to unknown clusters (misclassified samples) to their nearest labeled clusters based on their distance from the cluster centers. Additional details are provided in the supplementary materials.

We report the performance of our estimation method in Tab. 4. We observe that on all four datasets our proposed estimation method leads to reasonable performance. In addition to this, we conduct a series of experiments on CIFAR-100 dataset to determine the sensitivity of the proposed method to the novel class estimation error. The results are reported in Fig. 3 where we observe that our proposed method performs reasonably well over a wide range of estimation error. Please note that even with 25% overestimation and underestimation errors, our proposed method outperforms ORCA and UNO (Tab. 1). These results reaffirms the practicality of the proposed solution.

Data Imbalance Even though most standard benchmark vision datasets are class balanced, in real-world this is hardly the case. Instead, real-world data often demonstrates long-tailed distribution. Since our proposed method can take any arbitrary distribution into account for generating pseudo-labels, it can naturally take imbalance into account. To demonstrate the effectiveness of our proposed method on imbalanced data, we conduct experiments on CIFAR-100 dataset and
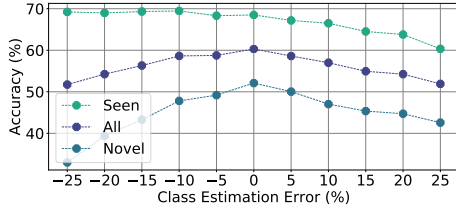
Table 5: Performance on **CIFAR-100** dataset with different imabalance factors (**IF**) with 50% classes as seen and 50% classes as novel.

Method	IF=10			IF=20		
	Seen	Novel	All	Seen	Novel	All
Balanced Class Distribution Prior	48.4	28.6	38.9	44.4	22.9	33.8
Imbalanced Class Distribution Prior	50.5	30.8	41.0	48.8	24.6	36.9
Estimated Class Distribution Prior	50.2	31.3	41.3	44.2	24.0	35.3

report the results in Tab. 5. We observe that for both imbalance factors (exponential) of 10 and 20, our proposed method with imbalance class distribution prior improves over the balanced prior baseline by 1.1% and 3.1%, respectively. We also conduct another set of experiments where we assume no access to class distribution prior. To this end, we propose a simple extension of our method to address imbalance problem. In cases where we do not have access to the prior information about the distribution of classes, to train our model, we start with a class-balanced prior. Next, we iteratively update our prior based on the latest posterior class distribution after every few epochs. The results are reported in the last row of Tab. 5. We observe that our simple estimation technique performs reasonably well and outperforms the class-balanced baseline with a noticeable margin. In summary, these experiments validate that our proposed method can effectively take advantage of underlying data distribution and work reasonably well even when we do not have access to the class distribution prior.
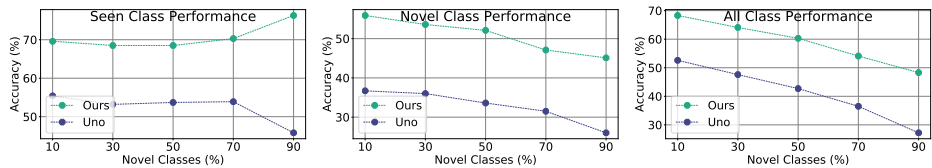
Fig. 4: Accuracy on seen (left), novel (middle), and all classes (right), as a function of different percentage of novel classes on the **CIFAR-100** dataset.

Different Percentage of Novel Classes In all of our experiments, we consider 50% classes as seen and the remaining 50% as novel. To further investigate how our method performs under different conditions, we vary percentages of novel classes. We conduct this experiment on CIFAR-100 dataset. The results are presented in Fig. 4, where we vary the number of novel classes from 10% to 90%. For this analysis, we compare the performance with UNO. The left figure in Fig. 4 shows that our performance on seen classes remains relatively the same as we increase the percentage of novel classes. Furthermore, we observe that our seen class accuracy increases considerably when the percentage of novel classes is very high (90%) which is to be expected since this is an easier classification task for seen classes. However, for UNO, we notice a significant performance drop as the number of novel classes increases which shows that UNO is not suf-

ficiently stable for this challenging setup. On novel classes (Fig. 4-middle), as we expect, we observe a steady drops in performance as the number of novel classes increase. However, as depicted in this graph, even at a very high novel class ratio, our proposed method can successfully provide a very good performance. Note that, we do not include ORCA in this experiment since their code is not publicly available. However, a similar analysis for ORCA is available in their supplementary materials with 50% labeled data. We observe that our novel class performance is noticeably higher than ORCA even though we only employ 10% labeled data. Finally, in Fig. 4-right we observe that the overall performance degrades predictably as we increase the percentage of of novel classes.

Novel Class Discovery In this work, we propose a general solution for open-world SSL problem which can be easily modified for the novel class discovery problem, where the principal assumption is that the unlabeled data contains only novel class samples. In this set of experiments we apply our proposed method on the novel class discovery task by generating pseudo-labels only for novel classes. We do not make any other

Table 6: Performance on novel class discovery task on **CIFAR-100** dataset with 50% classes as seen and 50% classes as novel.

Method	Novel
k-means	28.3
DTC[23]	35.9
RankStats[22]	39.2
RankStats+[22]	44.1
UNO[17]	52.9
Ours	57.5

modification to the original method for this task. The findings from these experiments are reported in Tab. 6. We conduct experiments on CIFAR-100-50, i.e., 50 classes are set as novel. For comparison, we use the results reported in UNO [17]. To the best of our knowledge, UNO reports the best scores for this particular experimental setup. Tab. 6 demonstrates that the porposed method outperforms k-means, DTC [23], RankStats [22], and RankStats+ by a significant margin. Importantly, our method also outperforms the current state-of-the-art method for novel class discovery, UNO, by 4.6%. Interestingly, this experiment demonstrates that *our proposed method is a versatile solution which can be readily applied to novel class discovery problem.*

5   Conclusion

In this work, we propose a practical method for open-world SSL problem. Our proposed method generates pseudo-labels according to class distribution prior to solve open-world SSL problem in realistic settings with arbitrary class distributions. We extend our method to handle practical scenarios where neither the number of unkown classes nor the class distribution prior is available. Furthermore, we introduce uncertainty-guided temperature scaling to improve the reliability of pseudo-label learning. Our extensive experiments on seven diverse datasets demonstrate the effectiveness of our approach, where it significantly improves the state-of-the-art. Finally, we show that our method can be readily applied to novel class discovery problem to outperform the existing solutions.

References

1. Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., Zhang, L.: Bottom-up and top-down attention for image captioning and visual question answering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018) 1

2. Asano, Y., Patrick, M., Rupprecht, C., Vedaldi, A.: Labelling unlabelled videos from scratch with multi-modal self-supervision. Advances in Neural Information Processing Systems **33**, 4660–4671 (2020) 2

3. Ayhan, M.S., Berens, P.: Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks. In: International conference on Medical Imaging with Deep Learning (2018) 7

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) 1, 3

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) 4

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) 2, 4, 7

7. Boult, T.E., Cruz, S., Dhamija, A.R., Gunther, M., Henrydoss, J., Scheirer, W.J.: Learning and the unknown: Surveying steps toward open world recognition. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 9801–9807 (2019) 3

8. Brenier, Y.: D'ecomposition polaire et r'earrangement monotone des champs de vecteurs. CR Acad. Sci. Paris Sér. I Math. **305**, 805–808 (1987) 6

9. 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 1, 2, 4, 6, 8, 9, 10

10. Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A.: Unsupervised learning of visual features by contrasting cluster assignments. arXiv preprint arXiv:2006.09882 (2020) 2, 4, 7, 9

11. 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

12. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. arXiv preprint arXiv:2002.05709 (2020) 8, 12

13. 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) 4

14. 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, 4

15. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems **26**, 2292–2300 (2013) 6, 7

16. Feinman, R., Curtin, R.R., Shintre, S., Gardner, A.B.: Detecting adversarial samples from artifacts. arXiv preprint arXiv:1703.00410 (2017) 7

17. 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) 3, 4, 8, 9, 10, 14

18. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059. PMLR (2016) 7

19. 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) 4

20. Girshick, R.: Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 1440–1448 (2015) 1

21. 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, 4, 9, 10

22. 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) 3, 4, 9, 10, 14

23. 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) 1, 3, 9, 10, 14

24. 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

25. 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, 8

26. 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- 3

27. Jain, L.P., Scheirer, W.J., Boult, T.E.: Multi-class open set recognition using probability of inclusion. In: European Conference on Computer Vision. pp. 393–409. Springer (2014) 3

28. Joachims, T.: Transductive inference for text classification using support vector machines. In: Icml. vol. 99, pp. 200–209 (1999) 4

29. Joseph, K., Khan, S., Khan, F.S., Balasubramanian, V.N.: Towards open world object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5830–5840 (2021) 1, 3

30. Kantorovich, L.: On translation of mass. In: Dokl. AN SSSR. vol. 37, p. 20 (1942) 6

31. Kardan, N., Shah, M., Hill, M.: Self-joint supervised learning. In: International Conference on Learning Representations (2021) 3

32. Kardan, N., Sharma, A., Stanley, K.O.: Towards consistent predictive confidence through fitted ensembles. In: 2021 International Joint Conference on Neural Networks (IJCNN). pp. 1–9. IEEE (2021) 3

33. Kardan, N., Stanley, K.O.: Fitted learning: Models with awareness of their limits. arXiv preprint arXiv:1609.02226 (2016) 1

34. 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) 1

35. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. arXiv preprint arXiv:2004.11362 (2020) 8
36. 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) 4
37. Krause, J., Stark, M., Deng, J., Fei-Fei, L.: 3d object representations for fine-grained categorization. In: 4th International IEEE Workshop on 3D Representation and Recognition (3dRR-13). Sydney, Australia (2013) 8
38. Krizhevsky, A., Nair, V., Hinton, G.: Cifar-10 (canadian institute for advanced research) 8
39. Krizhevsky, A., Nair, V., Hinton, G.: Cifar-100 (canadian institute for advanced research) 8
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) 4
42. Le, Y., Yang, X.: Tiny imagenet visual recognition challenge. CS 231N **7**(7), 3 (2015) 8
43. Lee, D.H.: Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks (2013) 2, 4
44. Lee, K., Lee, K., Lee, H., Shin, J.: A simple unified framework for detecting out-of-distribution samples and adversarial attacks. Advances in neural information processing systems **31** (2018) 3
45. Liang, S., Li, Y., Srikant, R.: Enhancing the reliability of out-of-distribution image detection in neural networks. arXiv preprint arXiv:1706.02690 (2017) 3
46. 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) 4
47. Maji, S., Kannala, J., Rahtu, E., Blaschko, M., Vedaldi, A.: Fine-grained visual classification of aircraft. Tech. rep. (2013) 8
48. 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) 4
49. 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) 2
50. Mukherjee, S., Awadallah, A.: Uncertainty-aware self-training for few-shot text classification. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 21199–21212. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper/2020/file/f23d125da1e29e34c552f448610ff25f-Paper.pdf 7
51. 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, 4
52. Parkhi, O.M., Vedaldi, A., Zisserman, A., Jawahar, C.V.: Cats and dogs. In: IEEE Conference on Computer Vision and Pattern Recognition (2012) 8
53. 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) 4

54. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017) 1

55. 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 7

56. Rizve, M.N., Kardan, N., Khan, S., Khan, F.S., Shah, M.: Openldn: Learning to discover novel classes for open-world semi-supervised learning. In: European Conference on Computer Vision (2022) 4

57. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision **115**(3), 211–252 (2015) 8

58. 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) 4, 7

59. Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boult, T.E.: Toward open set recognition. IEEE transactions on pattern analysis and machine intelligence **35**(7), 1757–1772 (2012) 1, 3

60. 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) 4

61. Sinkhorn, R., Knopp, P.: Concerning nonnegative matrices and doubly stochastic matrices. Pacific Journal of Mathematics **21**(2), 343–348 (1967) 2

62. Sohn, K., Berthelot, D., Li, C.L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., Raffel, C.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. arXiv preprint arXiv:2001.07685 (2020) 2, 4, 9, 10

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) 2, 4

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) 6

65. Verma, V., Lamb, A., Kannala, J., Bengio, Y., Lopez-Paz, D.: Interpolation consistency training for semi-supervised learning. In: IJCAI (2019) 7

66. Xu, H., Liu, B., Shu, L., Yu, P.: Open-world learning and application to product classification. In: The World Wide Web Conference. pp. 3413–3419 (2019) 3

67. 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 2

68. Zaeemzadeh, A., Bisagno, N., Sambugaro, Z., Conci, N., Rahnavard, N., Shah, M.: Out-of-distribution detection using union of 1-dimensional subspaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9452–9461 (2021) 3

69. Zhao, X., Krishnateja, K., Iyer, R., Chen, F.: Robust semi-supervised learning with out of distribution data. arXiv preprint arXiv:2010.03658 (2020) 2, 4

70. 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 (CVPR). pp. 9462–9470 (June 2021) 3, 4