Targeted Representation Alignment for Open-World Semi-Supervised Learning

Ruixuan Xiao[1], Lei Feng[2], Kai Tang[1], Junbo Zhao[1], Yixuan Li[3], Gang Chen[1], Haobo Wang[1]*
[1]Zhejiang University, China
[2]Singapore University of Technology and Design, Singapore
[3]University of Wisconsin-Madison, USA
{xiaoruixuan,tk0819,j.zhao,cg,wanghaobo}@zju.edu.cn, lfengqaq@gmail.com, sharonli@cs.wisc.edu

Abstract

Open-world Semi-Supervised Learning aims to classify unlabeled samples utilizing information from labeled data, while unlabeled samples are not only from the labeled known categories but also from novel categories previously unseen. Despite the promise, current approaches solely rely on hazardous similarity-based clustering algorithms and give unlabeled samples free rein to spontaneously group into distinct novel class clusters. Nevertheless, due to the absence of novel class supervision, these methods typically suffer from the representation collapse dilemma—features of different novel categories can get closely intertwined and indistinguishable, even collapsing into the same cluster and leading to degraded performance. To alleviate this, we propose a novel framework TRAILER which targets to attain an optimal feature arrangement revealed by the recently uncovered neural collapse phenomenon. To fulfill this, we adopt targeted prototypes that are pre-assigned uniformly with maximum separation and then progressively align the representations to them. To further tackle the potential downsides of such stringent alignment, we encapsulate a sample-target allocation mechanism with coarse-to-fine refinery that is able to infer label assignments with high quality. Extensive experiments demonstrate that TRAILER outperforms current state-of-the-art methods on generic and fine-grained benchmarks. The code is available at https://github.com/Justherozen/TRAILER.

1. Introduction

The resounding achievements of deep learning heavily rely on large-scale and accurately-annotated training data, which is labor-intensive and time-consuming. To alleviate this, *semi-supervised learning (SSL)* has emerged as a prominent alternative [4, 55, 60], which exploits unlabeled data for boosted performance. Despite the success, current SSL methods are primarily based on the closed-world assumption that labeled and unlabeled data share the same predefined label space. Nevertheless, this assumption rarely holds in real-world applications, where the model may encounter novel classes previously unseen in unlabeled data. To relax this assumption, *novel category discovery (NCD)* was first proposed [15, 21, 22] which assumes all unlabeled data strictly belong to disjoint novel classes. Recently, a more practical setting, *open-world semi-supervised learning (open-world SSL)*, also referred to as *generalized category discovery (GCD)*, has received huge attention [5, 46, 50] by generalizing standard SSL and NCD. It assumes that unlabeled data contain both known classes from the labeled data and similar yet distinct novel classes.

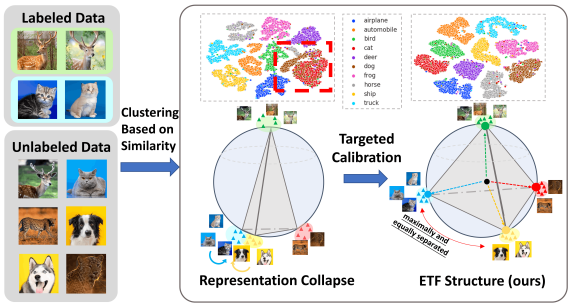


Figure 1. **Left:** Scenario of open-world SSL where unlabeled samples are from known and novel classes. **Right:** Visualization without and with our targeted calibration mechanism. Without calibration, representations tend to get jumbled and inseparable, as in the left where dog and cat categories converge into the same cluster.

The goal of open-world SSL is to classify the already-known classes from the labeled data, as well as to discover new classes in unlabeled data without annotations and assign instances to them. Nascent efforts have been made to address this practical yet under-investigated task [20, 33, 41, 50]. The essence of these recent works is to rashly rely on the similarity-based [5, 41] or contrastive-based [46] clustering algorithm in feature space to spontaneously

*Corresponding author.

group different samples into separate clusters. Some methods further integrate self-training techniques with generated pseudo-labels for boosted performance [20, 33]. However, due to the missing supervision and semantic shift for novel categories, the self-organized clustering step can get overfitted and biased towards the known classes, leading to a representation collapse dilemma— the feature distance between one novel category and another category may get too close and finally become inseparable, even converging within a single cluster and thus jeopardizing generalization performance, as illustrated in the left of Figure 1.

On the other hand, a recent study has unveiled a feature-classifier alignment phenomenon termed *neural collapse (NC)* [38]. It indicates that under ideal conditions, the last-layer features of the same class tend to collapse into their within-class mean. Meanwhile, these within-class means of all classes will be aligned with classifiers and formed as a simplex equiangular tight frame (ETF) at the end of training, as elaborated in Section 3.1. Intuitively, such an elegant NC phenomenon affords an optimal feature structure with minimized within-class variance and maximized between-class variance [36], coinciding with what we fervently desire. Nevertheless, in open-world SSL, the absence of novel class annotation can hinder the representation learning and thus disrupt the natural induction of NC.

In an effort to salvage this and accomplish the appealing feature arrangement of NC, we propose a novel framework, **T**argeted **R**epresentation **A**l**I**gnment for open-world semi-supervised **LE**a**R**ning (dubbed **TRAILER**). Key to our method, TRAILER leverages a targeted classifier to progressively align embeddings to the pre-assigned ETF structure, eventually fulfilling the optimal state of uniformly and maximally separated embeddings. Despite the promise, this calibration procedure is still susceptible to false alignment stemming from unreliable pseudo-labels. To mitigate this, we integrate a hierarchical sample-target allocation strategy that comprises two sequential steps: (i)-rough sample assignment based on the optimal transport mechanism; (ii)-label refinery by casting the known-novel separation problem to a specific weakly-supervised paradigm termed PU learning [2, 13, 62]. Such an allocation mechanism can provide more precise pseudo-labels and thus facilitate representation calibration. Through our visualized results in Figure 1 and Section 4.3, TRAILER does indeed induce discriminative and non-skewed embeddings. Empirically, TRAILER surpasses state-of-the-art approaches on different evaluation benchmarks, *e.g.*, improves the best baseline by **3.1%**, **2.9%**, and **1.5%** overall accuracy on CIFAR-10, CIFAR-100 and ImageNet-100 datasets, respectively.

2. Related Work

Closed-World Semi-Supervised Learning has been widely explored [4, 44, 55, 66] to learn from limited labeled and massive unlabeled data. Pseudo labeling [31, 42] and consistency regularization [30, 44, 55] are two prominent techniques. The former is based on self-training that utilizes model predictions as optimization targets [40, 56], while the latter enforces model to produce consistent outputs for different views of the same image [3, 4, 43, 55]. However, these methods are primarily based on the closed-world assumption that labeled and unlabeled data share the same pre-defined label space, which limits their real-world application where unlabeled samples from previously unseen classes may emerge. Open-set SSL [19, 24, 25, 45] relaxes this by simply rejecting all unseen class instances, but still fails to classify them into distinct novel categories.

Open-World Semi-Supervised Learning further accommodates real-world scenarios [5, 20, 46]. It is also known as generalized category discovery (GCD) [50] and is a natural extension of novel category discovery (NCD). NCD is first introduced in [21] and improved in later works [15, 22] to discover new categories by leveraging the knowledge from labeled known categories. However, NCD assumes that all unlabeled samples belong to novel classes and thus fails to recognize known class samples. Most recently, open-world SSL generalizes NCD to identify unlabeled samples from both known and novel classes [5]. To tackle this, several incipient solutions have been proposed [20, 59], which typically involve similarity-based representation learning and then involve sample clustering to generate pseudo-labels for self-training. ORCA [5] is the pioneering work that integrates an uncertainty-based margin loss to control the intra-class variance of different classes. OpenCon [46] further derives a semi-supervised constrastive mechanism to produce compact feature space. Despite the promise, these methods solely rely on spontaneous clustering based on sample similarity, overlooking potential risks that representations of different classes possibly converge into the same cluster, thus leading to the representation collapse dilemma.

Neural Collapse describes an elegant alignment phenomenon [38], which states that the last-layer features will collapse into their within-class centers and these centers together with classifiers will eventually form a simplex equiangular tight frame (ETF). Stemming from the appealing symmetry, a plethora of studies have been triggered to explain this phenomenon theoretically [26, 48, 58, 65]. NC is proved to be the global optimality under ideal conditions with the CE [18, 26] and the MSE [23, 64] loss functions. The NC phenomenon has also been investigated in some specific scenarios like imbalanced learning [54, 57, 63] noisy label learning [37, 52], transfer learning [17] and incremental learning [58]. In our work, we investigate the opportunities of inducing such optimal feature arrangement of NC through targeted calibration for open-world SSL.

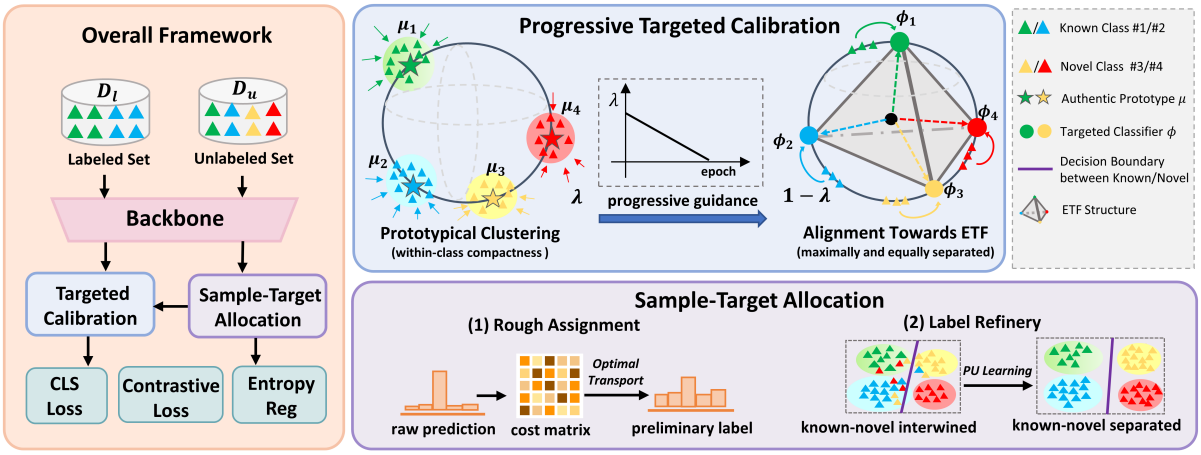


Figure 2. Overview of TRAILER. For representation calibration, we leverage the targeted classifier ϕ^{etf} to align the feature towards the optimal arrangement of ETF progressively while maintaining the within-class compactness with prototypical clustering. For sample-target allocation, the unlabeled samples are provided with reliable pseudo-labels with rough assignment and subsequent label refinery.

3. Method

Problem Definition. In open-world SSL, we assume that the training dataset $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$ comprises two subsets: a labeled set $\mathcal{D}_l = \{x_i, y_i\}_{i=1}^m$ where the input image $x_i \in \mathbb{R}^d$ and the ground-truth label y_i come from a set of known classes $\mathcal{C}_{known}$, and an unlabeled set $\mathcal{D}_u = \{x_i\}_{i=1}^n$ where the potential ground-truth label is from a set of classes $\mathcal{C}_{all}$. In closed-world SSL, the unlabeled samples strictly come from known classes, i.e., $\mathcal{C}_{all} = \mathcal{C}_{known}$. As for open-world SSL which is more practical, the unlabeled samples can come from either known classes $\mathcal{C}_{known}$ or novel classes, i.e., $\mathcal{C}_{known} \subset \mathcal{C}_{all}$. We consider $\mathcal{C}_{novel} = \mathcal{C}_{all} \backslash \mathcal{C}_{known}$ as the set of novel classes. Following previous setting [5, 15, 46], we assume the number of novel classes $|\mathcal{C}_{novel}|$ is known apriori, which can also be estimated using the off-the-shelf methods [21, 50] as discussed in Section 4.3.

Learning Goal. For classification problem with $K = |\mathcal{C}_{all}|$ classes, open-world SSL aims to obtain a model $f_{\theta,\phi} : \mathbb{R}^d \mapsto \mathbb{R}^K$ that recognizes samples from both known and novel classes based on $\mathcal{D}_l$ and $\mathcal{D}_u$. Specifically, $f_{\theta,\phi} = h_\phi \circ g_\theta$ can be decomposed into a representation backbone g_θ and a classification head h_ϕ with weights $\phi = [\phi_1, ..., \phi_k] \in \mathbb{R}^{P \times K}$, for which we adopt a prototypical one following [15] with logit $h(z) = \text{Norm}(z)^\top \text{Norm}(\phi)$. Formally, given the weakly and strongly augmented view x_i and x_i' of the same image in batch $\mathcal{B}$, we define $z_i, z_i' \in \mathbb{R}^P$ as the last-layer P-dimensional features and $p_i = \text{softmax}(h(z_i))$ as output probability. Normalization for features is denoted as $\text{Norm}(z) = z/||z||_2$ and omitted in the subsequent.

Warm-up Phase. Similar to previous works [5, 33], we also include a commonly adopted warm-up strategy at the start of training. The overall objectives to cluster known classes and discover potential novel samples consist of three losses: 1) the cross-entropy loss (CE) $\mathcal{L}_{ce}$ on labeled samples $\mathcal{D}_l$ which is elaborated later in Section 3.1; 2) the self-supervised contrastive loss $\mathcal{L}_{con} = -\log \frac{\exp(z_i^\top z_i'/\tau)}{\sum_{j=1}^n \exp(z_i^\top z_j'/\tau)}$ on all samples; and 3) the entropy regularization $\mathcal{L}_{ent} = \sum_{k=1}^K \overline{p}^k \log(\overline{p}^k)$ for preventing the trivial solution of assigning all instances to the same class, where $\overline{p} = \frac{1}{|\mathcal{B}|} \sum_{x_i \in \mathcal{B}} p_i$ is the average probability of the batch. The overall losses for warm-up are formulated,

$$\mathcal{L}_{warm} = \mathcal{L}_{ce} + \mathcal{L}_{con} + \alpha \mathcal{L}_{ent} \tag{1}$$

where α is a weighting parameter. After warm-up, we also would like to group unlabeled samples into distinct clusters and assign them pseudo-labels for self-training. As mentioned earlier, the crucial challenge is how to abstain from the representation collapse dilemma and obtain discriminative features without novel class annotations. To fulfill this, we first draw inspiration from neural collapse phenomenon.

3.1. Neural Collapse for Representation Calibration

Recent works have disclosed the Neural Collapse phenomenon [17, 38] which corresponds to an optimal feature-classifier alignment towards a simplex equiangular tight frame (ETF). This offers a uniformly and equally separated feature arrangement that is desirable for open-world SSL.

Definition 1 (Simplex ETF) *A Simplex equiangular tight frame (ETF) refers to a collection of K equal-length and maximally-equiangular P-dimensional embedding vectors $\mathbf{E} = [e_1, ..., e_K] \in \mathbb{R}^{P \times K}$ which satisfies:*

$$\mathbf{E} = \sqrt{\frac{K}{K-1}} \mathbf{U} \left(\mathbf{I}_K - \frac{1}{K} \mathbf{1}_K \mathbf{1}_K^\top \right) \tag{2}$$

where $\mathbf{I}_K$ is the identity matrix, $\mathbf{1}_K$ is an all-ones vector, and $\mathbf{U} \in \mathbb{R}^{P\times K}(P \geq K)$ allows a rotation.

All vectors in a simplex ETF $\mathbf{E}$ have an equal ℓ_2 norm and the same pair-wise maximal equiangular angle $-\frac{1}{K-1}$ [57],

$$e_{k_1}^\top e_{k_2} = \frac{K}{K-1}\delta_{k_1,k_2} - \frac{1}{K-1}, \forall k_1, k_2 \in [1, K] \quad (3)$$

where $\delta_{k_1,k_2} = 1$ when $k_1 = k_2$ and 0 otherwise. The NC phenomenon can be described by the following properties under ideal conditions: (1) The last-layer features converge to their within-class means, and the within-class means of all classes converge to different vertices of a simplex ETF. (2) The normalized classifier weights converge to the same simplex ETF. Notably, this process is implicit and natural since it is induced with common optimizers such as SGD, rather than enforced by an explicit objective.

As discussed above, NC has provided an optimal ETF structure for feature arrangement with minimal within-class covariance and maximum between-class covariance (i.e., maximized Fisher discriminant ratio [16]), which coincides with our goal of deriving discriminative and non-skewed embeddings. However, NC is delicate and can fail to be naturally induced under imperfect conditions [37, 57]. For open-world SSL, the missing novel class annotations can also lead to biased features and break the NC optimality.

To remedy this, we resort to effectuating an explicit NC and fulfilling the appealing feature arrangement with representation calibration, which is achieved by leveraging a targeted prototypical classifier h_{etf}. In specific, h_{etf} is pre-assigned with weights $\phi^{etf} \in \mathbb{R}^{P\times K}$, which is a random simplex ETF with optimal structure as Definition 1 and is fixed rather than learnable during training. By assuming a known label $\hat{y}_i$ for each sample x_i, the standard CE loss $\mathcal{L}_{ce}$ is then replaced as $\mathcal{L}_{etf}$, which is formulated,

$$\mathcal{L}_{etf}(x_i, \hat{y}_i) = -\log \frac{\exp\left(z_i{}^\top \cdot \phi_{\hat{y}_i}^{etf}/\tau\right)}{\sum_{k=1}^{K}\exp\left(z_i{}^\top \cdot \phi_k^{etf}/\tau\right)} \quad (4)$$

where τ is the temperature parameter. Through $\mathcal{L}_{etf}$, the feature z_i is drawn closer to its corresponding target $\phi_{\hat{y}_i}^{etf}$ and pulled away from other class vectors of ETF structure.

For the labeled sample (x_i, y_i), we directly set $\hat{y}_i = y_i$. As for unlabeled data without gound-truth labels, we design a sample-target allocation strategy to generate pseudo-labels $\hat{y}$, whose details are shown in Section 3.2.

Progressive Alignment. Notably, if we directly optimize the loss $\mathcal{L}_{etf}$ after the warm-up phase, the provisionally clustered samples would be forcibly dragged to the corresponding targeted vector $\phi_{\hat{y}}^{etf}$. As a result, the already-learned clustering compactness can get compromised and the within-cluster feature distribution can get jumbled.

To address this, we intend to align the representations in a progressive manner to preserve the compactness within each cluster during calibration. Formally, given sample x_i with pseudo-label $\hat{y}_i$, we also maintain the authentic prototypes $\mu_k = \text{Norm}(\text{Avg}_{\hat{y}_i=k}(z_i))$ for each class $1 \leq k \leq K$. To preserve the compactness of cluster structure, a prototypical clustering loss is also integrated,

$$\mathcal{L}_{proto}(z_i, \hat{y}_i) = -\log \frac{\exp\left(z_i{}^\top \cdot \mu_{\hat{y}_i}/\tau\right)}{\sum_{k=1}^{K}\exp\left(z_i{}^\top \cdot \mu_k/\tau\right)} \quad (5)$$

The overall classification loss is then formulated,

$$\mathcal{L}_{cls} = \lambda\mathcal{L}_{proto} + (1-\lambda)\mathcal{L}_{etf} \quad (6)$$

where λ is a calibration strength parameter that is gradually ramped down from 1 to 0. With such a progressive manner, the embeddings are cautiously and smoothly drawn closer to the targeted optimal arrangement of ETF while the within cluster compactness is preserved.

3.2. Hierarchical Sample-Target Allocation

In previous steps, we involve targeted classifiers to regulate representation learning. Nevertheless, our stringent calibration still holds unexpected risks. Lacking supervision, the calibration process for unlabeled samples largely relies on the quality of generated pseudo-labels. When the pseudo-labels are unreliable, the feature alignment procedure can be misguided towards the wrong cluster. To overcome this, we introduce our hierarchical sample-target allocation mechanism with course-to-fine assignment in what follows.

Rough Assignment through Optimal Transport. For pseudo-label generation, one may adopt model prediction $\arg\max_{1\leq k\leq K} h^k(z_i)$ directly. However, this may lead to degenerate solutions where data points of different categories map into a single category [1], thus resulting in representation collapse. To mitigate this, we adopt a widely-used equipartition constraint [6, 15] for rough label assignment, which is attained by solving an optimal transport problem.

Formally, given a batch $\mathcal{B}$ of b unlabeled samples with the logit matrix $\mathbf{P} = [h(z_1), ..., h(z_b)]$ whose columns are output logits by h_{etf} and the label-assignment matrix $\mathbf{Q} = [q_1, ..., q_b] \in [0, 1]^{K\times b}$ whose columns are the to-be-assigned soft labels q_i. $\mathbf{Q}$ is estimated by,

$$\begin{aligned} \mathbf{Q} &= \max_{\mathbf{Q}\in\Gamma} \text{Tr}(\mathbf{Q}^\top\mathbf{P}) + \epsilon\,\text{H}(\mathbf{Q}) \\ \text{s.t. } \Gamma &= \{\mathbf{Q} \in \mathbb{R}_+^{K\times b} \mid \mathbf{Q}\mathbf{1}_b = \frac{1}{K}\mathbf{1}_k, \mathbf{Q}^\top\mathbf{1}_K = \frac{1}{b}\mathbf{1}_b\} \end{aligned} \quad (7)$$

where $\text{H}(\cdot)$ is the entropy function, ϵ is a hyperparameter and $\text{Tr}(\cdot)$ is the trace function. The solution of Eq. (7) is obtained using the Sinkhorn-Knopp algorithm [11]. Please refer to Appendix A for more optimization details.

Label Refinery with Known-Novel Separation. After the rough assignment, we obtain unpolished soft labels $q \in [0,1]^K$. However, the aforementioned skewed predictions towards known classes stemming from overfitting remain unresolved. To rectify this, we prioritize the subtask of precisely demarcating known and novel classes for label refinery, which is achieved by including an auxiliary binary head h_{aux} and reformulating this binary task as the classical positive unlabeled (PU) learning problem [2, 13, 62].

PU learning tackles a special case of binary classification where the model has access to a labeled set containing merely positive samples and an unlabeled set with both positive and negative samples. Thus, the task of distinguishing known and novel classes can be intuitively converted to PU learning by regarding known and novel as positive and negative respectively. A flurry of methods has been designed for PU learning [10, 27, 34], among which we adopt a prior-free variational algorithm [8] with loss $\mathcal{L}_{aux}$ to optimize the newly devised h_{aux}. See Appendix A for more details.

After that, we elicit the binary prediction $\hat{y}^{aux}$ for h_{aux}, where 1 and 0 indicate prediction of known and novel respectively. Then we leverage $\hat{y}^{aux}$ to refine the label q,

$$q^j = \begin{cases} q^j & \text{if } \mathbb{I}(j \in \mathcal{C}_{known}) = \hat{y}^{aux} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where q^j is the j-th entry of q, $1 \leq j \leq K$ and $\mathbb{I}(\cdot)$ is the indicator function. That is, we retain the known/novel class components of soft label q based on prediction $\hat{y}^{aux}$ and discard the remaining inconsistent components. The refined q are then utilized for the final hard pseudo-label generation $\hat{y}_i = \arg\max_{1 \leq k \leq K} q^k$ in $\mathcal{L}_{cls}$ of Eq. (6).

Pseudo-Label Filtering. While the proposed sample-target allocation mechanism can enhance the quality of pseudo-labels, the remaining obstinate noisy pseudo-labels can still impede model training, especially in the early stages when the model output is not accurate enough. To neutralize this, we further utilize a high-confidence selection to select clean pseudo-labels from noisy ones for $\mathcal{D}_u$. Formally, for each class $1 \leq j \leq K$, we select a clean subset $\mathcal{D}^j_{sel}$ from $\mathcal{D}^j_u = \{(x, \hat{y}) \in \mathcal{D}_u | \hat{y} = j\}$ class by class with the highest confidence scores in the first R percent,

$$\mathcal{D}^j_{sel} = \{(x, \hat{y}) \in \mathcal{D}^j_u | \ rank(q^j) < R\%\} \tag{9}$$

Then the integral selected set is merged as $\mathcal{D}_{sel} = \cup_{j=1}^{K} \mathcal{D}^j_{sel}$. The classification loss in Eq. (6) is then merely valid on $\mathcal{D}_{sel}$. In practice, we set a small R to ensure the precision of selected samples at the beginning and gradually increase it to integrate more clean samples. Finally, the overall training loss is given by,

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \mathcal{L}_{aux} + \mathcal{L}_{con} + \alpha \mathcal{L}_{ent} \tag{10}$$

where $\mathcal{L}_{con}$ and $\mathcal{L}_{ent}$ are described in Eq. (1). Through this procedure, Our sample-target allocation mechanism is able to provide pseudo-labels with high quality and guide the representation calibration toward the correct targets. The overview of TRAILER is illustrated in Figure 2.

4. Experiment

In this section, we present the main results and part of the ablation results to verify the effectiveness of TRAILER. More experimental details and results, such as the results with less labeled data, can be found in Appendix.

4.1. Setup

Datasets. We first evaluate TRAILER on three datasets CIFAR-10, CIFAR-100 [29] and ImageNet-100, which is randomly sub-sampled with 100 classes from ImageNet [12] following [5]. For each dataset, We first divide classes into the first 50% known and the rest 50% novel classes following [5]. Then we randomly select a portion (50%) of samples from the known classes as labeled $\mathcal{D}_l$, and the remaining along with all novel class samples are regarded as unlabeled $\mathcal{D}_u$. We also conduct experiments with different novel class ratios and on the fine-grained semantic shift benchmarks (SSB) [51], including CUB-200 [53], Stanford Cars [28], FGVC-Aircraft [35], and a naturally long-tailed Herbarium 19 [47] datasets, as discussed in Section 4.3.

Baselines. We compare TRAILER with various open-world SSL baselines, including ORCA [5], GCD [50], OpenCon [46], NACH [20], OpenLDN [41], GPC [61] and OpenNCD [33]. We also provide comparisons with SSL, open-set SSL, and NCD baselines following [20, 33]. For traditional and open-set SSL baselines including FixMatch [44], DS^{3}L [19], and CGDL [45], we extend them to be applicable to novel classes by selecting samples with lower confidence scores as novel and applying k-means clustering to obtain results. For NCD methods including DTC [21], RankStats [22], and UNO+ [15], they are extended to classify known classes by using Hungarian algorithm to match some discovered classes with known classes. For performances of baselines, we directly adopt reported results from receptive papers or from [33], except OpenLDN was re-tested under transductive setting. All compared methods are implemented based on pre-trained ResNet using SimCLR [9] following [5], except DTC and OpenCon which have specialized pre-training procedure. We run all experiments three times and report the averaged results.

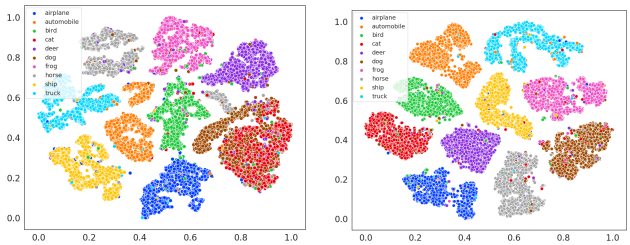
Evaluation Metrics. Following [5, 33], we evaluate the transductive performance on both known and novel classes. For novel classes, we utilize the Hungarian algorithm to derive the optimal assignments and then calculate accuracy

Methods	CIFAR-10			CIFAR-100			ImageNet-100		
	Known	Novel	All	Known	Novel	All	Known	Novel	All
FixMatch [44]	71.5	50.4	49.5	39.6	23.5	20.3	65.8	36.7	34.9
DS^{3}L [19]	77.6	45.3	40.2	55.1	23.7	24.0	71.2	32.5	30.8
CGDL [45]	72.3	44.6	39.7	49.3	22.5	23.5	67.3	33.8	31.9
DTC [21]	53.9	39.5	38.3	31.3	22.9	18.3	25.6	20.8	21.3
RankStats [22]	86.6	81.0	82.9	36.4	28.4	23.1	47.3	28.7	40.3
SimCLR [9]	58.3	63.4	51.7	28.6	21.1	22.3	39.5	35.7	36.9
ORCA [5]	88.2	90.4	89.7	66.9	43.0	48.1	89.1	72.1	77.8
GCD [50]	78.4	79.7	79.1	68.5	33.5	45.2	82.3	58.3	68.2
OpenLDN [41]	92.3	86.8	88.6	67.3	40.4	49.4	74.5	45.7	55.3
OpenCon [46]	90.4	91.1	89.3	69.1	47.8	52.7	90.6	80.8	83.8
NACH [20]	89.5	92.2	91.3	68.7	47.0	52.1	91.0	75.5	79.6
OpenNCD [33]	88.4	90.6	90.1	**69.7**	43.4	49.3	90.0	77.5	81.6
TRAILER (Ours)	**93.4**	**95.0**	**94.4**	**69.7**	**48.7**	**55.6**	**91.4**	**82.4**	**85.3**

Table 1. Accuracy comparison of known, novel, and all classes on CIFAR-10, CIFAR-100 and ImaneNet-100 dataset. The dataset is composed of 50% known classes and 50% novel classes, with 50% of the known classes labeled. **Bold** entries indicate superior results.

Ablation	CIFAR-10			CIFAR-100		
	Known	Novel	All	Known	Novel	All
TRAILER	93.4	**95.0**	**94.4**	69.7	**48.7**	**55.6**
w/o Rep Calibration	94.1	58.9	70.6	69.2	14.9	33.4
with $\lambda = 0$	93.9	91.8	92.5	70.2	46.6	54.6
w/o Rough Assign	92.9	92.3	92.5	69.3	37.9	48.3
w/o Label Refinery	**95.2**	89.6	91.5	**70.3**	45.4	53.7
w/o PL Filtering	92.2	90.5	91.0	69.9	43.6	52.2

Table 2. Ablation results on CIFAR-10 and CIFAR-100. 'Rep' and 'PL' indicate representation and pseudo-label respectively.



(a) Without calibration. (b) With calibration.

Figure 3. T-SNE feature visualization on CIFAR-10. Different colors represent the corresponding ground-truth classes. Without calibration, the dog and cat classes collapse into the same cluster.

$\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(y_i = perm(\hat{y}_i))$ where $perm(\cdot)$ is the matching permutation. The overall joint accuracy is also measured by Hungarian match using both known and novel classes.

Implementation Details. For CIFAR experiments, we use ResNet-18 as the backbone. The model is trained for 200 epochs with warm-up of 15 epochs. We employ the Adam optimizer with a batch size of 200 and an initial learning rate of $5e^{-4}$, which decays by a cosine scheduler. The temperature τ is set as 0.3 and 0.5 for CIFAR-10 and CIFAR-100. We linearly ramp down λ from 1 to 0 and ramp up selection ratio $R\%$ from 0.3 to 0.9 for pseudo-label filtering. The label refinery is enabled after training h_{aux} for 15 epochs. For the optimization of Sinkhorn-Knopp, we adopt hyperparameters from [15] and set $\epsilon = 0.05$ with 3 iterations. We put more implementation details for ImageNet-100 and fine-grained dataset in Appendix A.

4.2. Main Reults

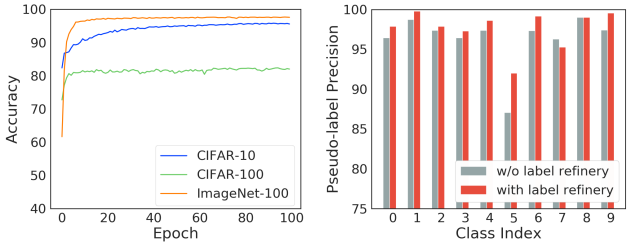
Table 1 shows the comparison results on the CIFAR-10, CIFAR-100, and ImageNet-100 datasets, including the accuracy of the 'Known', 'Novel', and 'All' classes. TRAILER significantly outperforms all the rivals by a considerable margin in terms of overall accuracy on different datasets. Specifically, on CIFAR-10, TRAILER improves upon the best baseline by 1.1%, 2.8%, and 3.1% for the 'Known', 'Novel', and 'All' classes respectively. Moreover, on the more challenging CIFAR-100 and ImageNet-100 datasets with larger label space, while most baselines demonstrate a notable performance drop, TRAILER shows robustness and reaches an extraordinary trade-off between the known and novel classes. It consistently maintains superior performance with 2.9% and 1.5% overall enhancement. These results empirically demonstrate the efficacy of our proposed TRAILER for the task of open-world SSL.

4.3. Analysis

Results on Semantic Shift Benchmarks. We further conduct experiments on fine-grained semantic shift benchmarks (SSB) [51] following previous work [50, 61]. For fair comparisons, we adopt the same training and compar-

Methods	CUB			Stanford Cars			FGVC-Aircraft			Herbarium 19		
	Known	Novel	All	Known	Novel	All	Known	Novel	All	Known	Novel	All
k-means	38.9	32.1	34.3	10.6	13.8	12.8	14.4	16.8	16.0	12.9	12.8	12.9
RankStats+ [22]	51.6	24.2	33.3	61.8	12.1	28.3	36.4	22.2	26.9	55.8	12.8	27.9
UNO+ [15]	49.0	28.1	35.1	70.5	18.6	35.5	56.4	32.2	40.3	53.7	14.7	28.3
ORCA [5]	45.6	30.2	35.3	50.1	10.7	23.5	31.8	17.1	22.0	30.9	15.5	20.9
GCD [50]	56.6	48.7	51.3	57.6	29.9	39.0	41.1	46.9	45.0	51.0	27.0	35.4
OpenCon [46]	63.8	52.1	54.7	**78.6**	32.7	49.1	-	-	-	**58.9**	28.6	39.3
GPC [61]	58.2	53.1	55.4	59.2	32.8	42.8	42.5	47.9	46.3	51.7	27.9	36.5
TRAILER (Ours)	**71.3**	**61.9**	**65.1**	71.7	**47.6**	**55.4**	**62.6**	**50.5**	**54.5**	57.0	**37.8**	**44.5**

Table 3. Results on the Semantic Shift Benchmarks. **Bold** entries indicate superior results and blank ones indicate results are not provided.



(a) Binary accuracy of h_{aux}. (b) Pseudo-label precision.

Figure 4. (a) Accuracy of binary prediction between known and novel classes in first 100 epochs. (b) Per-class pseudo-label precision for unlabeled samples on CIFAR-10 with and without label refinery. Class indexes 0-4 are known while 5-9 are novel.

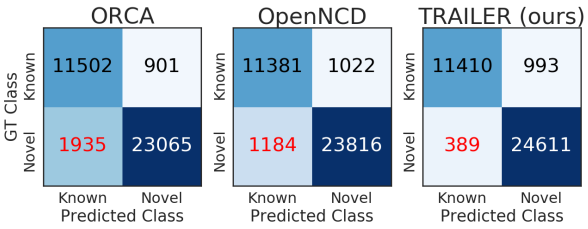


Figure 5. Confusion matrix between known and novel classes on CIFAR-10. TRAILER can alleviate biased predictions.

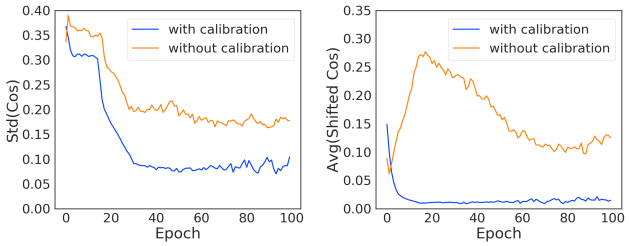
isons protocol from [50] and utilize ViT-B/16 backbone [14] with DINO pre-trained weights [7]. As shown in Table 3, it can be observed that TRAILER exceeds the baselines by a substantial lead in all metrics, with improvements of 8.8%, 14.8%, 2.6%, and 9.2% for 'Novel' accuracy on these datasets. These results consistently demonstrate its superior performance, even under challenging fine-grained tasks.

Benefits of Representation Calibration. To explore the benefits of representation calibration, we conduct ablations and equip TRAILER with different representation learning strategies: 1) *TRAILER w/o Rep Calibration* which dismiss the feature alignment of the targeted classifier and instead utilize the vanilla learnable classifier; 2) *TRAILER with* $\lambda = 0$ which discards the progressive manner for direct calibration. From Table 2, although TRAILER w/o Rep Calibration remains competitive on known classes, it suffers from a calamitous performance plunge for novel classes, which indicates the efficacy of representation calibration. TRAILER with $\lambda = 0$ achieves decent performance, but still underperforms TRAILER perceptibly, indicating the advantage of the progressive manner.

We further visualize the representation with and without representation calibration on CIFAR-10 using t-SNE [49] in Figure 3 and in Appendix B. It can be shown that representation calibration facilitates inducing compact and non-skewed representations, discarding which can lead to representation collapse—the representations are biased towards known classes, where the dog class (novel, brown) and part of horse class samples (novel, grey) collapse into the cluster of the cat class (known, red), as shown in Figure 3a.

Effect of Sample-Target Allocation. Next, we ablate the effectiveness of the sample-target allocation mechanism in Table 2. For the coarse-level assignment, we test the variant *TRAILER w/o Rough Assign* which sticks to the vanilla softmax function for rough label assignment of q and suffers from rigorous performance degradation, especially on CIFAR-100. For the fine-level refinery, we test the variant *TRAILER w/o Label Refinery* which discards the refinery procedure of auxiliary head h_{aux}. Although such a variant exhibits slightly better performance for the known classes, it significantly lags behind TRAILER for the novel classes.

To demonstrate it more intuitively, we visualize the accuracy of binary prediction for h_{aux} between known and novel classes in Figure 4a. We can find that h_{aux} quickly achieves high prediction accuracies and such accurate separation between known and novel indeed refines the rough assignment supplementarily, finally enhancing the pseudo-label precision on 9 of 10 classes, as shown in Figure 4b. Such a mechanism also alleviates the biased prediction, as shown in Figure 5 where a non-trivial amount of novel samples are



(a) Equiangularity (b) Maximal separation

Figure 6. Visualizations on the (a) equiangularity and (b) maximal separation properties of neural collapse on CIFAR-10 in the first 100 epochs. Representation calibration enables TRAILER to achieve feature arrangement with equal and maximal separation.

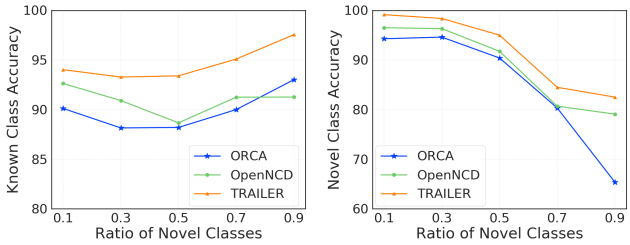
Ablation	Estimation of K	Known	Novel	All
TRAILER	Known (100)	69.7	**48.7**	**55.6**
ORCA	Known (100)	66.9	43.0	48.1
TRAILER	Estimated (124)	**71.1**	44.3	53.3
ORCA	Estimated (124)	66.3	40.0	46.4

Table 4. Ablations with unknown class number K on CIFAR-100.

misidentified as known (the number in red) for baselines.

Empirical Visualizations for Neural Collapse. In the previous section, we have provided the embedding visualization. Here we supply some quantitative results in addition. Recall that two core properties of NC are equiangularity and maximal separation. Given the centered class-means $\hat{z}^k$ for features from all classes $1 \leq k \leq K$, for equiangularity, we calculate the standard deviation of the cosines between each pair $\mathrm{Std}_{k \neq k'}(\cos(\hat{z}^k, \hat{z}^{k'}))$. For maximal separation, we calculate the average value $\mathrm{Avg}_{k \neq k'}(\cos(\hat{z}^k, \hat{z}^{k'}) + \frac{1}{K-1})$ since the optimal pair-wise angle is $-\frac{1}{K-1}$ as discussed in Eq. (3). The trends of these two metrics on the CIFAR-10 dataset with and without representation calibration are shown in Figure 6. Notably, equiped with representation calibration, the standard deviations of the cosines and shifted average cosine values approach zero, which indicates that TRAILER achieves the engaging feature structure that is close to neural collapse.

Unknown Number of Novel Classes. The previous experiments follow previous protocols and assume that the class number is known apriori [5, 15, 46], which is sometimes impractical in real-world applications and brings challenges for both TRAILER and the other baselines. A simple resolution is to first use estimation methods [32, 39] to derive the number of classes before deployment, for example, k-estimation with Brent's algorithm from [50] or cross-validating using labeled probe set from [5, 21, 46]. On



(a) Accuracy for known classes. (b) Accuracy for novel classes.

Figure 7. Accuracy comparisons for known and novel classes on CIFAR-10 with different novel class ratios ranging from 0.1 to 0.9.

CIFAR-100 the former gets exactly 100 while the latter derives 124 and was adopted by us for further exploration. We then utilize this estimation result to rerun the experiment by initializing the corresponding number of classifier vectors. As shown in Table 4, TRAILER is able to achieve competitive performance even when the number of classes is not known in advance. We further provide additional results with different estimated class numbers K from 80 to 130 in Appendix B to demonstrate the robustness of TRAILER.

Effect of Novel Class Ratio. Here we further evaluate the performance of TRAILER on the CIFAR-10 dataset with different ratios of novel classes from 0.1 to 0.9. As shown in Figure 7, we can see that TRAILER displays robust performance when the ratio of novel classes grows larger from 0.1 to 0.9 and consistently manifests superior performance than other baselines on both known and novel classes.

5. Conclusion

In this work, we propose a novel framework TRAILER for open-world SSL. We first take inspiration from the recently discovered neural collapse phenomenon and intend to attain its appealing feature arrangement with minimal within-class and maximum between-class covariance. To achieve this, we adopt a targeted classifier and align representations towards its pre-assigned optimal structure in a progressive manner. To mitigate the potential risk of misalignment, we tackle the sample-target assignment for unlabeled samples with a coarse-to-fine allocation mechanism. Comprehensive experiments show that TRAILER enhances the baselines by a notable margin on both generic and fine-grained benchmarks. We hope our work can inspire future research to leverage the appealing properties of the neural collapse phenomenon for tasks in open-world scenarios.

6. Acknowledgments

This work is supported by the Pioneer R&D Program of Zhejiang (No. 2024C01035).

References

[1] Yuki Markus Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. In *ICLR*. OpenReview.net, 2020. 4

[2] Jessa Bekker and Jesse Davis. Learning from positive and unlabeled data: a survey. *Mach. Learn.*, 109(4):719–760, 2020. 2, 5

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

[4] David Berthelot, Nicholas Carlini, Ian J. Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. In *NeurIPS*, pages 5050–5060, 2019. 1, 2

[5] Kaidi Cao, Maria Brbic, and Jure Leskovec. Open-world semi-supervised learning. In *ICLR*. OpenReview.net, 2022. 1, 2, 3, 5, 6, 7, 8

[6] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *NeurIPS*, 2020. 4

[7] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, pages 9630–9640. IEEE, 2021. 7

[8] Hui Chen, Fangqing Liu, Yin Wang, Liyue Zhao, and Hao Wu. A variational approach for learning from positive and unlabeled data. In *NeurIPS 2020*, 2020. 5

[9] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607. PMLR, 2020. 5, 6

[10] Xuxi Chen, Wuyang Chen, Tianlong Chen, Ye Yuan, Chen Gong, Kewei Chen, and Zhangyang Wang. Self-pu: Self boosted and calibrated positive-unlabeled training. In *ICML*, pages 1510–1519. PMLR, 2020. 5

[11] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *NeurIPS*, pages 2292–2300, 2013. 4

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255. IEEE Computer Society, 2009. 5

[13] François Denis. PAC learning from positive statistical queries. In *Algorithmic Learning Theory, 9th International Conference, ALT '98, Otzenhausen, Germany, October 8-10, 1998, Proceedings*, pages 112–126. Springer, 1998. 2, 5

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*. OpenReview.net, 2021. 7

[15] Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. A unified objective for novel class discovery. In *ICCV*, pages 9264–9272. IEEE, 2021. 1, 2, 3, 4, 5, 6, 7, 8

[16] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936. 4

[17] Tomer Galanti, András György, and Marcus Hutter. On the role of neural collapse in transfer learning. In *ICLR*. OpenReview.net, 2022. 2, 3

[18] Florian Graf, Christoph D. Hofer, Marc Niethammer, and Roland Kwitt. Dissecting supervised constrastive learning. In *ICML*, pages 3821–3830. PMLR, 2021. 2

[19] Lan-Zhe Guo, Zhenyu Zhang, Yuan Jiang, Yu-Feng Li, and Zhi-Hua Zhou. Safe deep semi-supervised learning for unseen-class unlabeled data. In *ICML*, pages 3897–3906. PMLR, 2020. 2, 5, 6

[20] Lan-Zhe Guo, Yi-Ge Zhang, Zhi-Fan Wu, Jie-Jing Shao, and Yu-Feng Li. Robust semi-supervised learning when not all classes have labels. In *NeurIPS*, 2022. 1, 2, 5, 6

[21] Kai Han, Andrea Vedaldi, and Andrew Zisserman. Learning to discover novel visual categories via deep transfer clustering. In *ICCV*, pages 8400–8408. IEEE, 2019. 1, 2, 3, 5, 6, 8

[22] Kai Han, Sylvestre-Alvise Rebuffi, Sébastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Automatically discovering and learning new visual categories with ranking statistics. In *ICLR*. OpenReview.net, 2020. 1, 2, 5, 6, 7

[23] X. Y. Han, Vardan Papyan, and David L. Donoho. Neural collapse under MSE loss: Proximity to and dynamics on the central path. In *ICLR*. OpenReview.net, 2022. 2

[24] Zongyan Han, Zhenyong Fu, Shuo Chen, and Jian Yang. Contrastive embedding for generalized zero-shot learning. In *CVPR*, pages 2371–2381. Computer Vision Foundation / IEEE, 2021. 2

[25] Zongyan Han, Zhenyong Fu, Shuo Chen, and Jian Yang. Semantic contrastive embedding for generalized zero-shot learning. *Int. J. Comput. Vis.*, 130(11):2606–2622, 2022. 2

[26] Wenlong Ji, Yiping Lu, Yiliang Zhang, Zhun Deng, and Weijie J. Su. An unconstrained layer-peeled perspective on neural collapse. In *ICLR*. OpenReview.net, 2022. 2

[27] Ryuichi Kiryo, Gang Niu, Marthinus Christoffel du Plessis, and Masashi Sugiyama. Positive-unlabeled learning with non-negative risk estimator. In *NeurIPS*, pages 1675–1685, 2017. 5

[28] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *ICCV*, pages 554–561. IEEE Computer Society, 2013. 5

[29] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. *Master's thesis, Department of Computer Science, University of Toronto*, 2009. 5

[30] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. In *ICLR*. OpenReview.net, 2017. 2

[31] Dong-Hyun Lee. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. *ICML 2013 Workshop : Challenges in Representation Learning (WREPL)*, 2013. 2

[32] Collin Leiber, Lena G. M. Bauer, Benjamin Schelling, Christian Böhm, and Claudia Plant. Dip-based deep embedded

clustering with k-estimation. In *KDD*, pages 903–913. ACM, 2021. 8

[33] Jiaming Liu, Yangqiming Wang, Tongze Zhang, Yulu Fan, Qinli Yang, and Junming Shao. Open-world semi-supervised novel class discovery. In *IJCAI*, pages 4002–4010. ijcai.org, 2023. 1, 2, 3, 5, 6

[34] Chuan Luo, Pu Zhao, Chen Chen, Bo Qiao, Chao Du, Hongyu Zhang, Wei Wu, Shaowei Cai, Bing He, Saravanakumar Rajmohan, and Qingwei Lin. PULNS: positive-unlabeled learning with effective negative sample selector. In *AAAI*, pages 8784–8792. AAAI Press, 2021. 5

[35] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew B. Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *CoRR*, abs/1306.5151, 2013. 5

[36] Aleix M. Martínez and Avinash C. Kak. PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(2):228–233, 2001. 2

[37] Duc Anh Nguyen, Ron Levie, Julian Lienen, Eyke Hüllermeier, and Gitta Kutyniok. Memorization-dilation: Modeling neural collapse under noise. In *ICLR*. OpenReview.net, 2023. 2, 4

[38] Vardan Papyan, X. Y. Han, and David L. Donoho. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 117(40):24652–24663, 2020. 2, 3

[39] Dan Pelleg and Andrew W. Moore. X-means: Extending k-means with efficient estimation of the number of clusters. In *ICML*, pages 727–734. Morgan Kaufmann, 2000. 8

[40] Zhongzheng Ren, Raymond A. Yeh, and Alexander G. Schwing. Not all unlabeled data are equal: Learning to weight data in semi-supervised learning. In *NeurIPS*, 2020. 2

[41] Mamshad Nayeem Rizve, Navid Kardan, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. Openldn: Learning to discover novel classes for open-world semi-supervised learning. In *ECCV*, pages 382–401. Springer, 2022. 1, 5, 6

[42] Chuck Rosenberg, Martial Hebert, and Henry Schneiderman. Semi-supervised self-training of object detection models. In *7th IEEE Workshop on Applications of Computer Vision / IEEE Workshop on Motion and Video Computing (WACV/MOTION 2005), 5-7 January 2005, Breckenridge, CO, USA*, pages 29–36. IEEE Computer Society, 2005. 2

[43] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. In *NeurIPS*, pages 1163–1171, 2016. 2

[44] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *NeurIPS*, 2020. 2, 5, 6

[45] Xin Sun, Zhenning Yang, Chi Zhang, Keck Voon Ling, and Guohao Peng. Conditional gaussian distribution learning for open set recognition. In *CVPR*, pages 13477–13486. Computer Vision Foundation / IEEE, 2020. 2, 5, 6

[46] Yiyou Sun and Yixuan Li. Opencon: Open-world contrastive learning. *Trans. Mach. Learn. Res.*, 2023, 2023. 1, 2, 3, 5, 6, 7, 8

[47] Kiat Chuan Tan, Yulong Liu, Barbara Ambrose, Melissa Tulig, and Serge J. Belongie. The herbarium challenge 2019 dataset. *CoRR*, abs/1906.05372, 2019. 5

[48] Christos Thrampoulidis, Ganesh Ramachandra Kini, Vala Vakilian, and Tina Behnia. Imbalance trouble: Revisiting neural-collapse geometry. In *NeurIPS*, 2022. 2

[49] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9 (11), 2008. 7

[50] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *CVPR*, pages 7482–7491. IEEE, 2022. 1, 2, 3, 5, 6, 7, 8

[51] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. In *ICLR*. OpenReview.net, 2022. 5, 6

[52] Jiaheng Wei, Zhaowei Zhu, Hao Cheng, Tongliang Liu, Gang Niu, and Yang Liu. Learning with noisy labels revisited: A study using real-world human annotations. In *ICLR*. OpenReview.net, 2022. 2

[53] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona. Caltech-UCSD Birds 200. Technical Report CNS-TR-2010-001, California Institute of Technology, 2010. 5

[54] Liang Xie, Yibo Yang, Deng Cai, and Xiaofei He. Neural collapse inspired attraction-repulsion-balanced loss for imbalanced learning. *Neurocomputing*, 527:60–70, 2023. 2

[55] Qizhe Xie, Zihang Dai, Eduard H. Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. In *NeurIPS*, 2020. 1, 2

[56] Yi Xu, Lei Shang, Jinxing Ye, Qi Qian, Yu-Feng Li, Baigui Sun, Hao Li, and Rong Jin. Dash: Semi-supervised learning with dynamic thresholding. In *ICML*, pages 11525–11536. PMLR, 2021. 2

[57] Yibo Yang, Shixiang Chen, Xiangtai Li, Liang Xie, Zhouchen Lin, and Dacheng Tao. Inducing neural collapse in imbalanced learning: Do we really need a learnable classifier at the end of deep neural network? In *NeurIPS*, 2022. 2, 4

[58] Yibo Yang, Haobo Yuan, Xiangtai Li, Zhouchen Lin, Philip H. S. Torr, and Dacheng Tao. Neural collapse inspired feature-classifier alignment for few-shot class-incremental learning. In *ICLR*. OpenReview.net, 2023. 2

[59] Bo Ye, Kai Gan, Tong Wei, and Min-Ling Zhang. Bridging the gap: Learning pace synchronization for open-world semi-supervised learning. *CoRR*, abs/2309.11930, 2023. 2

[60] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. In *NeurIPS 2021*, pages 18408–18419, 2021. 1

[61] Bingchen Zhao, Xin Wen, and Kai Han. Learning semi-supervised gaussian mixture models for generalized category discovery. In *ICCV*, pages 16623–16633, 2023. 5, 6, 7

[62] Yunrui Zhao, Qianqian Xu, Yangbangyan Jiang, Peisong Wen, and Qingming Huang. Dist-pu: Positive-unlabeled learning from a label distribution perspective. In *CVPR*, pages 14441–14450. IEEE, 2022. 2, 5

[63] Zhisheng Zhong, Jiequan Cui, Yibo Yang, Xiaoyang Wu, Xiaojuan Qi, Xiangyu Zhang, and Jiaya Jia. Understanding imbalanced semantic segmentation through neural collapse. In *CVPR*, pages 19550–19559. IEEE, 2023. 2

[64] Jinxin Zhou, Xiao Li, Tianyu Ding, Chong You, Qing Qu, and Zhihui Zhu. On the optimization landscape of neural collapse under MSE loss: Global optimality with unconstrained features. In *ICML*, pages 27179–27202. PMLR, 2022. 2

[65] Zhihui Zhu, Tianyu Ding, Jinxin Zhou, Xiao Li, Chong You, Jeremias Sulam, and Qing Qu. A geometric analysis of neural collapse with unconstrained features. In *NeurIPS*, pages 29820–29834, 2021. 2

[66] Zhaowei Zhu, Tianyi Luo, and Yang Liu. The rich get richer: Disparate impact of semi-supervised learning. In *ICLR*. OpenReview.net, 2022. 2