Open-Domain Semi-Supervised Learning via Glocal Cluster Structure Exploitation

Zekun Li, Lei Qi, Yawen Li, Yinghuan Shi*, Yang Gao

Abstract—Semi-supervised learning (SSL) aims to reduce the heavy reliance of current deep models on costly manual annotation by leveraging a large amount of unlabeled data in combination with a much smaller set of labeled data. However, most existing SSL methods assume that all labeled and unlabeled data are drawn from the same feature distribution, which can be impractical in real-world applications. In this study, we take the initial step to systematically investigate the open-domain semi-supervised learning setting, where a feature distribution mismatch exists between labeled and unlabeled data. In pursuit of an effective solution for open-domain SSL, we propose a novel framework called **GlocalMatch**, which aims to exploit both **glo**bal and lo**cal** (*i.e.*, glocal) cluster structure of open-domain unlabeled data. The glocal cluster structure is utilized in two complementary ways. Firstly, GlocalMatch optimizes a Glocal Cluster Compacting (GCC) objective, that encourages feature representations of the same class, whether with in the same domain or across different domains, to become closer to each other. Secondly, GlocalMatch incorporates a Glocal Semantic Aggregation (GSA) strategy to produce more reliable pseudo-labels by aggregating predictions from neighboring clusters. Extensive experiments demonstrate that GlocalMatch outperforms the state-of-the-art SSL methods significantly, achieving superior performance for both in-domain and out-of-domain generalization. The code is released in https://github.com/nukezil/GlocalMatch.

Index Terms—Semi-supervised learning, distribution mismatch, cluster structure, pseudo-labeling.

1 Introduction

Semi-Supervised Learning (SSL) [1] is one of the fundamental paradigms in machine learning, aimed at enhancing model performance by incorporating unlabeled data, which are often much easier to obtain with little human labor. Given only a small fraction of labeled data, advanced deep SSL methods have exhibited outstanding results in various vision tasks, including image classification [2], object detection [3], and semantic segmentation [4]. SSL has also achieved success in other tasks involving diverse data types beyond images [5]–[9]. Despite the successful applications of SSL, it is important to note that most of these methods rely on the essential prerequisite that all labeled and unlabeled samples are drawn from the same distribution. However, in real-world tasks, it is often challenging, even impossible, to obtain a perfectly matched unlabeled dataset due to the sheer volume of data. Researchers have observed that SSL methods may exhibit poor performance when faced with unlabeled data containing classes unknown in the labeled data [10]. This situation, where there is a mismatch in class distribution between labeled and unlabeled data, is also referred to as Open-Set Semi-Supervised Learning [11]. To address this challenge, various methods have been proposed to alleviate the negative effects caused by unlabeled samples from unknown classes [11]–[18].

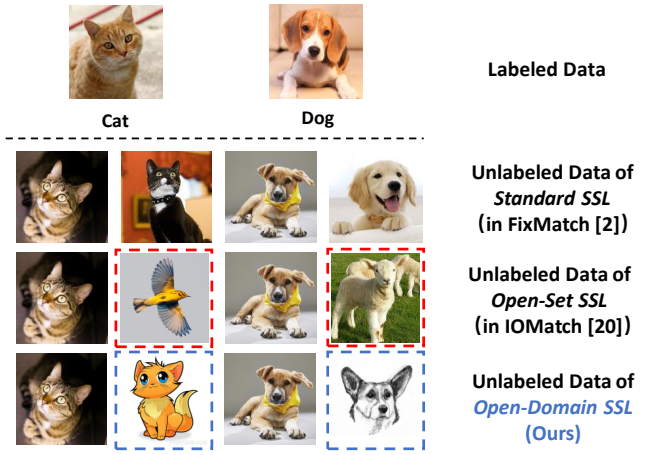


Fig. 1. Illustration of various SSL settings. (1) For standard SSL, labeled and unlabeled data are sampled from the *same* distribution. (2) For open-set SSL under *class distribution mismatch*, unlabeled data may contain classes unknown in labeled data (denoted by the red boxes). (3) For open-domain SSL under *feature distribution mismatch*, unlabeled data may contain samples from different domains than labeled data, *i.e.*, out-of-domain samples (denoted by the blue boxes).

Distinguished from the prior studies, we focus on addressing another realistic issue of SSL with a mismatch in feature distribution between labeled and unlabeled data. This issue is also critical to ensure the effectiveness of SSL in real-world applications, but has received limited investigation in existing research. The feature distribution mismatch problem is prevalent in practical scenarios due to the diverse domains of massive unlabeled data, which can be collected at different times, from various locations, and through different means. For example, the unlabeled data

* *Yinghuan Shi is the corresponding author.*
Zekun Li, Yinghuan Shi, and Yang Gao are with the State Key Laboratory for Novel Software Technology and the National Institute of Healthcare Data Science, Nanjing University, Nanjing 210093, China (e-mail: lizekun@smail.nju.edu.cn; syh@nju.edu.cn; gaoy@nju.edu.cn).
Lei Qi is with the School of Computer Science and Engineering and the Key Lab of Computer Network and Information Integration, Southeast University, Nanjing 211189, China (e-mail: qilei@seu.edu.cn).
Yawen Li is with the School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: warmly0716@126.com).

might be a mixture of low-resolution and high-resolution images, real images and synthetic images, or images with different artistic styles. Moreover, the task becomes even more demanding when all the labeled samples come exclusively from one of the domains and the domain labels of the unlabeled samples are unavailable. Inspired by the name convention of open-set SSL, we term such a challenging and realistic setting as *Open-Domain Semi-Supervised Learning*. As depicted in Fig. 1, the distinction in various SSL settings lies in the nature of the distribution mismatch between labeled and unlabeled data.

To demonstrate how out-of-domain samples within unlabeled data will affect the performance of SSL, we provide a case study using two popular datasets, CIFAR [19] and STL [20], which consist of low-resolution and high-resolution natural images, respectively. We compare the results obtained in the standard SSL setting, where both labeled and unlabeled samples are from the CIFAR dataset, with the results in the open-domain SSL setting, where the unlabeled data includes additional samples from the STL dataset, as shown in Fig. 2. For the classic standard SSL method, FixMatch [2], learning with additional out-of-domain unlabeled samples leads to much lower in-domain performance. More surprisingly, the model has almost lost all its out-of-domain generalization capability. Similar phenomena can be observed even in the latest state-of-the-art methods like SoftMatch [21]. The significant performance degradation arises from the pseudo-labeling mechanism, which is widely adopted in the mainstream SSL methods. During the early stage of training, the model will be inevitably biased towards the labeled domain, resulting in unreliable pseudo-labels for the out-of-domain unlabeled samples. Due to the lack of labeled samples from the corresponding domains to provide reliable supervision, these erroneous pseudo-labels become difficult to correct. As a consequence, the model will suffer from severe confirmation bias and can hardly generalize to out-of-domain testing data.

In light of the aforementioned challenge, we endeavor to enhance the traditional pseudo-labeling mechanism by leveraging the cluster structure of unlabeled data, instead of just learning from the instance-level semantic information. Given that the unlabeled samples are collected from different domains, the cluster structure should be examined from both local and global perspectives: From the local perspective, within each domain, the samples will form small clusters with high semantic consistency; From the global perspective, across different domains, clusters with similar semantics should be relatively close to each other. The glocal cluster structure will be harnessed to facilitate the learning of representation and classification simultaneously.

In this work, we propose a novel open-domain SSL framework named GlocalMatch. It exploits the glocal cluster structure through two complementary components: the Glocal Cluster Compacting (GCC) objective and the Glocal Semantic Aggregation (GSA) strategy. To optimize the GCC objective, we periodically perform K-Means clustering on all unlabeled data. At a local level, samples within each cluster are optimized to be closer to their respective centroids, while at a global scale, clusters exhibiting similar semantics are adjusted to be closer to one another. Simultaneously, the glocal cluster structure is employed for enhancing pseudo-labeling

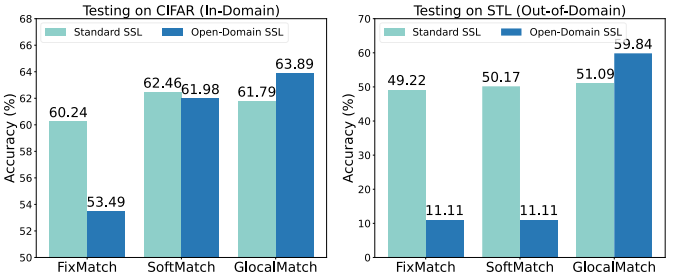


Fig. 2. We present the performance of models trained with different methods under different settings. For standard SSL, the labeled and unlabeled data are all from CIFAR. Only 1% of the samples are labeled. For open-domain SSL, the unlabeled data contain additional samples from STL. We report the classification accuracy on the testing sets of CIFAR and STL, respectively.

through the GSA strategy. Concretely, we assign a one-hot class label to each cluster by establishing complete bipartite connections between cluster centroids and class prototypes. This assignment can be formulated as a minimum-cost flow problem. Consequently, the pseudo-label of each sample is refined by aggregating semantic information not only from its own cluster but also from neighboring clusters. Using the glocal cluster structure as a bridge, these two components can mutually reinforce each other during the training process.

Extensive experiments have been conducted across various open-domain SSL scenarios with multiple datasets, where the proposed method, GlocalMatch, is compared with the latest state-of-the-art SSL methods. Fig. 2 offers a quick glance at the results, which demonstrate that GlocalMatch effectively mitigates the adverse impacts of out-of-domain samples and can even leverage them to achieve performance improvements. The achievement stems from two aspects: Firstly, the refinement process results in high-confidence pseudo-labels, which greatly enhances their reliability. Secondly, even samples with low confidence can also contribute to the exploitation of the glocal cluster structure. In summary, GlocalMatch can more accurately and effectively leverage open-domain unlabeled data.

Our contributions are summarized as follows:

- To the best of our knowledge, we are the first to systematically investigate the realistic yet challenging setting of open-domain SSL, where a feature distribution mismatch exists between labeled and unlabeled data.
- We propose a novel open-domain SSL framework, GlocalMatch, which aims to exploit the cluster structure of unlabeled data from both local and global perspectives. The glocal cluster information is utilized for boost the representation and classification simultaneously.
- We introduce two complementary components: the glocal cluster compacting objective for representation and the glocal semantic aggregation strategy for classification. Taking the glocal cluster structure as a bridge, they can enhance each other during training.

The rest of this paper is organized as follows. We discuss previous research relevant to the open-domain SSL problem in Section 2. In Section 3, we begin by introducing the preliminaries and the overall framework of GlocalMatch, and then explore the specific technical aspects. Our experimental results, along with the related analysis and discussions, are covered in Section 4, and we conclude in Section 5.

2 Related Work

2.1 Realistic Semi-Supervised Learning

In the era of deep learning, semi-supervised learning has garnered significant attention due to the imperative for extensive training data. In the context of deep SSL, consistency regularization [22] and pseudo-labeling [23] are two mainstream techniques that have been widely employed in prior studies [24]–[27]. Among more recent works, FixMatch [2] stands out as one of the most influential SSL methods, known for its simplicity and effectiveness. It enhances consistency regularization through a stronger form of data augmentation and incorporates confidence-based pseudo-labeling. FlexMatch [28] and FreeMatch [29] adjust the class-specific confidence thresholds based on varying learning difficulties. SoftMatch [21] proposes weighting unlabeled samples based on their confidence to address the quantity-quality trade-off problem of pseudo-labeling. Drawing inspiration from the advancements in contrastive learning [30]–[32], some methods leverage instance-level feature similarity in auxiliary learning objectives [33]–[36]. There are also graph-based methods achieving high performance on structured data [37]–[39]. For more comprehensive reviews on SSL theories and methods, we refer readers to [40]–[42].

While numerous positive results have been achieved, most existing SSL methods rely on the condition that labeled and unlabeled data share the exact same distribution. In realistic scenarios, however, a distribution mismatch between labeled and unlabeled data is common, which can lead to serious performance degradation in SSL methods [10]. The class distribution mismatch arises from discrepancies in label spaces, where the unlabeled data may contain new classes unknown in the labeled data. This setting, known as open-set SSL, has attracted growing attention. Researchers have put forth various strategies to alleviate the negative effects of such outliers from unknown classes. Certain open-set SSL methods employ an intuitive detect-and-exclude strategy, aiming to identify outliers and subsequently remove them from consideration [11], [12], [15]. On the other hand, alternative approaches recognize the potential value of outliers and utilize them in diverse ways [14], [16]–[18].

The feature distribution mismatch occurs when the unlabeled data may contain out-of-domain samples, which we refer to as open-domain SSL. This problem is also crucial for ensuring the performance of SSL methods in real-world tasks, but has not yet been thoroughly studied. Existing works, Huang et al. [43] and Jia et al. [44], explore a simplified scenario in which all unlabeled data are drawn from a single different domain. Proposed for such a setting, CAFA [43] and BDA [44] aim to align the distribution of unlabeled data to that of labeled data. Specifically, CAFA [43] achieves feature alignment through adversarial training and BDA [44] designs a weighted pseudo-labeling mechanism for distribution adaptation. While CAFA [43] and BDA [44] perform well when dealing with a single different domain in unlabeled data, their applicability diminishes in the more realistic open-domain SSL setting, as the absence of ground-truth domain labels and the amalgamation of multiple unknown domains will significantly compromise the effectiveness of distribution adaptation.

TABLE 1
Comparison of Open-Domain SSL and Existing Related Settings

Setting	Limited Labels	Domain Shift	Unknown Domain Information
Standard SSL [2]	✓	✗	✗
UDA [57]	✗	✓	✗
Huang's [43] and Jia's [44]	✓	✓	✗
Open-Domain SSL (Ours)	✓	✓	✓

2.2 Learning with Data from Different Domains

In open-domain SSL, models are expected to exploit unlabeled data from different domains. A learning problem related to it is Unsupervised Domain Adaptation (UDA) [45], where models are trained with a set of labeled "source" samples and a set of unlabeled "target" samples to enable generalization in the "target" domain. Motivated by seminal UDA theories [46]–[48], existing studies pursue diverse methods to reduce the domain discrepancy. A mainstream branch of works [49]–[51] proposes explicitly minimizing various discrepancy metrics, such as maximum mean discrepancy (MMD) [52] and its variants. Another branch of works [53]–[55] leverages the adversarial training paradigm to learn domain-invariant feature representations. Additionally, some researchers have observed that SSL and UDA share a common learning paradigm with different configurations of labeled and unlabeled data [56]. Taking this aspect into consideration, a unified approach called AdaMatch [56] has been proposed, aiming to encompass both standard SSL and UDA tasks.

Although UDA and open-domain SSL both involve learning from unlabeled samples and dealing with domain shifts, there are fundamental differences between the two settings. Firstly, in UDA, we have access to the domain information, which means that we know all the labeled samples are drawn from the source domain and all the unlabeled samples are from the target domain. However, in open-domain SSL, the domain information is absent as we cannot identify the domain label of each unlabeled sample during training. Secondly, UDA typically assumes the availability of plentiful labeled source data, whereas in open-domain SSL, the number of labeled samples is quite limited. Therefore, despite achieving high performance on standard UDA tasks, the strong UDA methods cannot be directly applied to the open-domain SSL setting due to the challenges posed by the scarcity of labeled data and the absence of domain information.

In Table 1, we summarize the distinctions between our proposed open-domain SSL setting and existing ones across three dimensions: (1) "Limited Labels": Whether only a limited number of labeled samples are available during training; (2) "Domain Shift": Whether there is a feature distribution mismatch (or domain shift) between labeled and unlabeled data; (3) "Unknown Domain Information": Whether it is unknown from which domain each unlabeled sample comes. Each of these dimensions loosens the constraints on training data, making our setting more realistic but also more challenging.

3 METHODOLOGY

The main idea of GlocalMatch is to exploit the glocal cluster structure of open-domain unlabeled data from both local and global perspectives. In this section, we will delve into the depth of the proposed framework.

In Section 3.1, we begin by providing a formal definition of open-domain semi-supervised learning and presenting an overview of GlocalMatch. Then, in Section 3.2, we elaborate on the assumption of glocal cluster structure, which is the core motivation of GlocalMatch. Moving on to Section 3.3, we introduce the glocal cluster compacting objective for optimizing the feature representations. Next, in Section 3.4, we provide details on how the glocal cluster structure aids in refining pseudo-labels through glocal semantic aggregation. Finally, in Section 3.5, we discuss the overall training procedure of GlocalMatch.

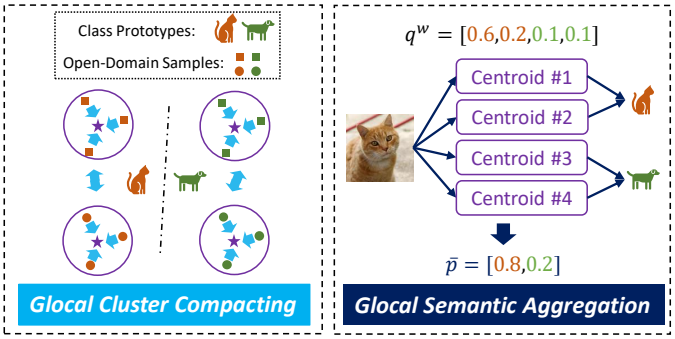


Fig. 3. We offer an intuitive illustration of the core ideas within in our proposed GlocalMatch framework. Through the Glocal Cluster Compacting (GCC) objective, we enhance the compactness of feature representations associated with each class. For the Glocal Semantic Aggregation (GSA) strategy, we take into account the semantics of centroids and the similarity between samples and centroids (represented by $\boldsymbol{q}^w$) to produce glocal structural pseudo-labels (denoted as $\hat{\boldsymbol{p}}$).

3.1 Preliminaries and Overview

We define an open-domain semi-supervised learning task, where the training set consists of N^l labeled samples and N^u unlabeled samples. As in standard SSL, we assume that $N^l \ll N^u$. For training, we use mini-batches comprising labeled and unlabeled data. Let $\mathcal{X} = \{(\boldsymbol{x}_i, y_i) : i \in (1, \ldots, B_l)\}$ represent a batch of B_l labeled samples, where $\boldsymbol{x}_i$ is a training sample and y_i is the corresponding label. Additionally, let $\mathcal{U} = \{\boldsymbol{u}_i : i \in (1, \ldots, B_u)\}$ represent a batch of B_u unlabeled samples. The labeled samples $\mathcal{X}$ and a portion of the unlabeled samples $\mathcal{U}^{in}$ are drawn from the same domain, and we refer to $\mathcal{U}^{in}$ as in-domain samples. Conversely, the remaining portion of the unlabeled samples $\mathcal{U}^{out}$ are drawn from different domain(s) and are referred to as out-of-domain samples. Technically, we have $\mathcal{U}^{in} \cup \mathcal{U}^{out} = \mathcal{U}$ and $\mathcal{U}^{in} \cap \mathcal{U}^{out} = \varnothing$. Moreover, the domain information is unavailable during training, which means that we do not know whether an unlabeled sample $\boldsymbol{u}_i$ belongs to $\mathcal{U}^{in}$ or $\mathcal{U}^{out}$. It is assumed that all involved domains share the same label space, and the total number of classes is L.

Given a labeled batch $\mathcal{X}$, we apply a random weak transformation function $\mathcal{T}_w(\cdot)$ to obtain the weakly augmented samples. A base encoder network $\mathcal{F}(\cdot)$ is employed to extract the feature representations from these samples, *i.e.*, $\boldsymbol{h}_i^l = \mathcal{F}(\mathcal{T}_w(\boldsymbol{x}_i)) \in \mathbb{R}^D$. A fully-connected classifier $\phi(\cdot)$ maps the feature $\boldsymbol{h}_i^l$ into the semantic label prediction, *i.e.*, $\boldsymbol{p}_i^l = \phi(\boldsymbol{h}_i^l)$. The labeled batch are used to optimize the networks with the standard cross-entropy loss $\mathrm{H}(\cdot)$:

$$\mathcal{L}_s(\mathcal{X}) = \frac{1}{B_l} \sum_{i=1}^{B_l} \mathrm{H}(y_i, \boldsymbol{p}_i^l). \tag{1}$$

Additionally, we adopt a non-linear projection head $\mathcal{G}(\cdot)$ to obtain the normalized low-dimensional embedding $\boldsymbol{z}_i^l = \mathcal{G}(\boldsymbol{h}_i^l)/\|\mathcal{G}(\boldsymbol{h}_i^l)\| \in \mathbb{R}^d$. For an unlabeled batch $\mathcal{U}$, we apply both the weak and strong augmentation with $\mathcal{T}_w(\cdot)$ and $\mathcal{T}_s(\cdot)$. The same operations as above are performed to obtain $\boldsymbol{h}_i^w$ and $\boldsymbol{z}_i^w$ for the weakly augmented samples $\mathcal{T}_w(\boldsymbol{u}_i)$; $\boldsymbol{h}_i^s$ and $\boldsymbol{z}_i^s$ for the strongly augmented samples $\mathcal{T}_s(\boldsymbol{u}_i)$. For the weakly augmented images, the semantic label predictions are obtained by $\boldsymbol{p}_i^w = \mathrm{DA}(\phi(\boldsymbol{h}_i^w))$, where $\mathrm{DA}(\cdot)$ stands for the distribution alignment strategy as in [58] to balance the distribution of the model's predictions and thus prevent them from collapsing to certain classes. For the strongly augmented images, $\boldsymbol{p}_i^s = \phi(\boldsymbol{h}_i^s)$.

GlocalMatch exploits the glocal cluster structure of open-domain unlabeled data with two components. We illustrate the core ideas in Fig. 3. Firstly, it optimizes the glocal cluster compacting (GCC) objective, encouraging samples of the same class to become closer in the feature space, even if they are from different domains. Secondly, the novel glocal semantic aggregation (GSA) strategy is introduced to produce more reliable pseudo-labels, alleviating confirmation bias. The two components, for representation and classification, are simultaneously optimized in GlocalMatch. They can progressively enhance each other, facilitated by the glocal cluster structure acting as a bridge.

3.2 Assumption on Glocal Cluster Structure

Before delving into the technical details, we first elaborate on the glocal cluster assumption, which serves as the core motivation behind GlocalMatch.

The cluster assumption in SSL states that data points belonging to the same cluster should be of the same class [1]. As highlighted in [40], this assumption can be considered as a necessary condition for SSL: if the data points (both labeled and unlabeled) cannot be meaningfully clustered, it is impossible for an SSL method to improve on a supervised learning method. For the standard SSL setting, where all labeled and unlabeled samples are drawn from the same domain, the cluster assumption has been implicitly or explicitly relied upon in deep SSL methods [59]–[61]. For a model trained on a single domain while dealing with data from multiple unknown domains, it has been observed that samples of different domains are tend to form domain-specific intrinsic structures [62]–[64]. Within each intrinsic structure, the cluster assumption still holds [64]. This observation inspires us to examine the overall cluster structure of open-domain samples from the local perspective.

However, the local cluster structure alone is insufficient to train an SSL model that generalizes well across different domains. From the global perspective, it is expected that clusters of the same class from different domains should be

aligned close to each other. This global cluster structure is not evident and is easily disrupted when the model is not well-trained. The case study in Fig. 2 demonstrates that the unreliable pseudo-labeling mechanism can severely disrupt the cluster structure of unknown domains, resulting in a significant loss of out-of-domain generalizability.

In GlocalMatch, we design GCC and GSA to effectively utilize the glocal cluster structure while preventing its disruption. Specially, the optimization of GCC objective aligns both samples with each local clusters and clusters exhibiting similar semantics to be closer to each other. The GSA strategy prevents unreliable pseudo-labels from disrupting the glocal cluster structure. At the same time, as the model's generalization capability improves on open-domain data, the glocal cluster structure becomes increasingly prominent.

3.3 Glocal Cluster Compacting for Representation

Broadly, the glocal cluster compacting (GCC) objective is aimed to enhance the intra-class compactness of feature representations, thereby rendering the feature space more discriminative. Considering that the unlabeled data can encompass samples from multiple distinct domains, the optimization of the GCC objective takes into account both local and global perspectives.

3.3.1 Compacting from Local Perspective

During training, we periodically perform K-Means clustering on all the unlabeled samples with the projected embeddings $\{\boldsymbol{z}_i^w\}_{i=1}^{N_u}$ of their weakly augmented views. The samples are clustered into K clusters represented by their centroids $\mathcal{C} = [\boldsymbol{c}_1; \cdots; \boldsymbol{c}_K] \in \mathbb{R}^{K\times d}$. The clustering assignment matrix is formulated as $\mathcal{A} \in \{0,1\}^{N_u\times K}$:

$$\mathcal{A}_{i,j} = \begin{cases} 1 & \text{if } \boldsymbol{u}_i \text{ is assigned to } \boldsymbol{c}_j; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

It is obvious that $\sum_j \mathcal{A}_{i,j} = 1$ for each i. Besides, we denote the submatrix with respect to the unlabeled samples within a mini-batch as $\mathcal{A}' \in \{0,1\}^{B_u\times K}$. Using the cosine similarity function, which is defined as $\text{sim}(\boldsymbol{a},\boldsymbol{b}) = \boldsymbol{a}^\top\boldsymbol{b}/\|\boldsymbol{a}\|\|\boldsymbol{b}\|$, the probability distribution that the i-th weakly and strongly augmented sample is assigned to each cluster can be estimated as $\boldsymbol{q}_i^w$ and $\boldsymbol{q}_i^s$, in which

$$\boldsymbol{q}_{i,j}^{w/s} = \frac{\exp\left(\text{sim}(\boldsymbol{z}_i^{w/s}, \boldsymbol{c}_j)/T\right)}{\sum_{j'=1}^{K}\exp\left(\text{sim}(\boldsymbol{z}_i^{w/s}, \boldsymbol{c}_{j'})/T\right)}, \tag{3}$$

where T is the temperature parameter controlling the concentration degree.

By combining the prototypical contrastive loss [65] with the weak-to-strong consistency technique, the GCC loss for a mini-batch of unlabeled data can be expressed as

$$\mathcal{L}_c(\mathcal{U}) = \frac{1}{B_u}\sum_{i=1}^{B_u}\text{H}(\widehat{\mathcal{W}}_i, \boldsymbol{q}_i^s), \tag{4}$$

where $\widehat{\mathcal{W}} \in \mathbb{R}^{B_u\times K}$ is the target assignment matrix. Specifically, $\widehat{\mathcal{W}}$ should encode the glocal cluster structure information, which represents the optimal similarity of each sample to all the centroids. Therefore, $\widehat{\mathcal{W}}$ should adhere to the constraints that $\widehat{\mathcal{W}}_{i,j} \geq 0$ and $\sum_j \widehat{\mathcal{W}}_{i,j} = 1$ for any i.

As with the previous works [63], [65], [66] that also exploit the cluster structure for representation learning, we can directly employ the K-Means clustering assignment $\mathcal{A}'$ as the target $\widehat{\mathcal{W}}$. In this way, the loss primarily optimizes the compactness from a local perspective, encouraging the samples to be closer to their respective centroids. Considering such local cluster structure is essential because samples within the same cluster often belong to the same class. However, relying solely on the local structure is insufficient, as clusters of the same class might be dispersed widely when they are from different domains. Below, we present how we also encode the global cluster structure into the target.

3.3.2 Integrating with Global Perspective

Regarding the global cluster structure, clusters with similar semantics should be drawn closer to other. Given that we can only process a mini-batch of unlabeled samples per iteration, we accomplish this indirectly by bringing samples with similar semantics together. Technically, we maintain a set of class prototypes $\mathcal{P} = [\boldsymbol{\mu}_1; \cdots; \boldsymbol{\mu}_L] \in \mathbb{R}^{L\times d}$, which is calculated with the projected embeddings of all the labeled samples:

$$\boldsymbol{\mu}_c = \frac{1}{|\mathcal{I}_c^l|}\sum_{i\in\mathcal{I}_c^l}\boldsymbol{z}_i^l, \tag{5}$$

where $\mathcal{I}_c^l$ denotes the indices of labeled samples belonging to the c-th class. Then, the similarity distribution between the j-th centroid to the class prototypes, denoted as $\widetilde{\boldsymbol{p}}_j$, is computed by

$$\widetilde{\boldsymbol{p}}_{j,c} = \frac{\exp\left(\text{sim}(\boldsymbol{c}_j, \boldsymbol{\mu}_c)/T\right)}{\sum_{c'=1}^{L}\exp\left(\text{sim}(\boldsymbol{c}_j, \boldsymbol{\mu}_{c'})/T\right)}. \tag{6}$$

For the i-th unlabeled sample, the semantic prediction $\boldsymbol{p}_i^w$ is produced by the classifier on its weakly augmented view. We adhere to a simple principle: if a sample is predicted to belong to a certain class, it should be drawn closer to the centroids that are near the prototype of that class. Therefore, we obtain the global target matrix, denoted as $\mathcal{W}^{global} \in \mathbb{R}^{B_u\times K}$, in which

$$\mathcal{W}_{i,j}^{global} = \text{Normalize}(\boldsymbol{p}_i^w \cdot \widetilde{\boldsymbol{p}}_j), \tag{7}$$

where

$$\text{Normalize}(\mathcal{W})_i = \frac{\mathcal{W}_i}{\sum_j \mathcal{W}_{i,j}}. \tag{8}$$

By adopting $\mathcal{W}^{global}$ as the target, samples exhibiting similar semantics will be attracted toward similar centroids, bringing them closer to each other. As clusters are formed by samples, it follows that clusters sharing similar semantics will naturally be drawn closer to one another.

Integrating the local target $\mathcal{W}^{local} = \mathcal{A}' \in \{0,1\}^{B_u\times K}$ with the global target $\mathcal{W}^{global}$, we finally reach the glocal target assignment matrix:

$$\widehat{\mathcal{W}} = \text{Normalize}(\mathcal{W}^{local} + \mathcal{W}^{global}). \tag{9}$$

The GCC loss (Eq. (4)) employs the glocal target $\widehat{\mathcal{W}}$ to exploits the glocal cluster structure for representation learning.

It is noteworthy that we perform K-Means clustering and define the loss on the projected embeddings $\boldsymbol{z}_i$'s instead

of directly on the feature representations $\boldsymbol{h}_i$'s. Such design is informed by prior research on self-supervised representation learning [67]. This study empirically demonstrates that a non-linear projection head substantially improves the quality of feature representations generated by the layer preceding it, which implicitly aids classification with the more discriminative feature space. In the following paragraphs, we explore how to explicitly enhance pseudo-labeling by utilizing the glocal cluster structure.

3.4 Glocal Semantic Aggregation for Classification

The primary challenge in open-domain SSL emerges from incorrect pseudo-labels assigned to out-of-domain unlabeled samples, particularly in the initial training stages. Due to the absence of labeled samples from the corresponding domains to offer dependable supervision, rectifying these errors for the classifier itself can be quite challenging. In response to this challenge, we design a glocal semantic aggregation (GSA) strategy, which generates an alternative set of pseudo-labels based on the glocal cluster structure, for complementing those predicted by the classifier.

3.4.1 Formulation of Centroid Matching

We start by associating semantics with the unsupervised clusters. Specifically, given that a majority of samples within each cluster belong to the same class, the first step is to determine the corresponding class labels for the clusters. In the projection space, the clusters are represented by their centroids $\mathcal{C}$, and the semantic information is encoded with in the class prototypes $\mathcal{P}$. The problem equals matching each centroid to a certain prototype. Formally, our objective is to obtain a matching matrix $\mathcal{Q} \in \{0,1\}^{K\times L}$, where

$$\mathcal{Q}_{j,c} = \begin{cases} 1 & \text{if } \boldsymbol{c}_j \text{ matches } \boldsymbol{\mu}_c; \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

with the constraint

$$\sum_{c'=1}^{k} \mathcal{Q}_{j,c'} = 1, \forall j \in \{1, \cdots, K\}. \tag{11}$$

Each matrix $\mathcal{Q}$ satisfying the above constraint corresponds to a label prediction for the clusters. Additionally, we regulate the minimum number of clusters allocated to each class to prevent the issue of collapsing, wherein most of the clusters are assigned to few dominant classes. The optimal matching matrix $\mathcal{Q}^*$ minimizes the total sum of pairwise distances between the matched clusters and classes:

$$\begin{aligned} \mathcal{Q}^* = \arg\min_{\mathcal{Q}} & \sum_{j=1}^{K}\sum_{c=1}^{L} \mathcal{Q}_{j,c} \|\boldsymbol{c}_j - \boldsymbol{\mu}_c\| \\ \text{s.t.} \quad & \sum_{c'=1}^{L} \mathcal{Q}_{j,c'} = 1, \sum_{j'=1}^{K} \mathcal{Q}_{j',c} \geq \lfloor \eta \frac{K}{L} \rfloor, \end{aligned} \tag{12}$$

where $\eta \in [0,1]$ is the hyperparameter controlling the lower-bound size of cluster sets belonging to each class.

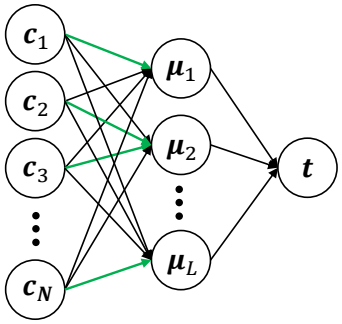


Fig. 4. The flow network graph G constructed by the centroids $\{\boldsymbol{c}_j\}$ and the class prototypes $\{\boldsymbol{\mu}_c\}$. $f(\boldsymbol{c}_j, \boldsymbol{\mu}_c) > 0$ (denoted by a green arc in the figure) indicates that the j-th cluster is assigned to the c-th class.

3.4.2 Solving as Minimum Cost Flow Problem

As mentioned in [68], the form of the constraints in the centroid matching problem (*i.e.*, Eq. (12)) renders it equivalent to a Minimum Cost Flow (MCF) problem [69]. A general MCF problem is defined upon a flow network graph $G = (V, E)$. Each node $u \in V$ is associated with a value $b(v)$ indicating whether it is a supply node ($b(v) > 0$) or a demand node ($b(v) < 0$). A feasible MCF problem requires that $\sum_{v\in V} b(v) = 0$. For each directed edge $e = (v, w) \in E$, $f(v, w)$ indicates the amount of flow on the edge; $c(v, w)$ indicates the cost for shipping one unit flow on the edge; and $u(v, w)$ indicates the capacity of the edge. The MCF problem is to find the optimal flow f^* which satisfies all the supplies and demands, and has the minimum total cost:

$$\begin{aligned} f^* = \arg\min_{f} & \sum_{(v,w)\in E} c(v,w) \cdot f(v,w) \\ \text{s.t.} \quad & 0 \leq f(v,w) \leq u(v,w), \forall (v,w) \in E \\ & \sum_{w\in V} f(v,w) - \sum_{w\in V} f(w,v) = b(v), \forall v \in V. \end{aligned} \tag{13}$$

In our centroid matching problem, each centroid $\boldsymbol{c}_j$ corresponds to a supply node with $b(\boldsymbol{c}_j) = 1$, and each class prototype $\boldsymbol{\mu}_c$ corresponds to a demand node with $b(\boldsymbol{\mu}_c) = -\lfloor \eta K/L \rfloor$. Then G is a complete bipartite graph, in which there is an edge for each $(\boldsymbol{c}_j, \boldsymbol{\mu}_c)$ pair. We define the cost on edge $(\boldsymbol{c}_j, \boldsymbol{\mu}_c)$ as $c(\boldsymbol{c}_j, \boldsymbol{\mu}_c) = \|\boldsymbol{c}_j - \boldsymbol{\mu}_c\|$. To satisfy the constrains involving the supplies and demands, we add an virtual demand node $\boldsymbol{t}$ with the demand $b(\boldsymbol{t}) = \lfloor \eta K \rfloor - K$. There are additional edges from each class prototype $\boldsymbol{\mu}_c$ to $\boldsymbol{t}$ with zero cost and no connections between the centroids $\boldsymbol{c}_j$ with $\boldsymbol{t}$. We illustrate the flow network graph in Fig. 4. Since $b(v), \forall v \in V$ is always integer-valued, the optimal flow f^* is also integer-valued, *i.e.*, $f^*(\boldsymbol{c}_j, \boldsymbol{\mu}_c) \in \{0, 1\}$ for each j and c, which is proved in [68]. Therefore, the optimal matching $\mathcal{Q}^*$ corresponds to the optimal flow f^*, that is $\mathcal{Q}^*_{j,c} = f^*(\boldsymbol{c}_j, \boldsymbol{\mu}_c)$. When it comes to implementation, the MCF problem can be efficiently solved with the network simplex algorithm [70].

3.4.3 Aggregation for Structural Pseudo-labels

With the optimal matching matrix $\mathcal{Q}^*$, we have assigned each cluster to a certain class. In a straightforward manner, we can employ the one-hot vector $\mathcal{Q}^*_j$ as the structural

pseudo-label for each unlabeled sample in the j-th cluster. However, such approach might be suboptimal for samples located at the boundaries of clusters. Considering this, we aggregate the semantic information from all the clusters to generate the structural pseudo-label $\bar{\boldsymbol{p}}_i$ for each unlabeled sample $\boldsymbol{u}_i$:

$$\bar{\boldsymbol{p}}_i = \frac{1}{L}\sum_{j=1}^{L} \boldsymbol{q}_{i,j}^{w} \cdot \mathcal{Q}_i^*, \tag{14}$$

where $\boldsymbol{q}_{i,j}^{w}$ is the normalized similarity between the sample $\boldsymbol{u}_i$ and the centroid $\boldsymbol{c}_j$, which has been defined in Eq. (3).

We use $\bar{\boldsymbol{p}}_i$ to refine the original pseudo-label predicted by the classifier:

$$\widehat{\boldsymbol{p}}_i = \alpha \boldsymbol{p}_i^{w} + (1-\alpha)\bar{\boldsymbol{p}}_i, \tag{15}$$

where $\alpha \in [0,1]$ is the hyperparameter controlling the degree of refinement. Integrating the glocal cluster structure, $\widehat{\boldsymbol{p}}_i$'s are more reliable pseudo-labels for open-domain unlabeled samples:

$$\mathcal{L}_u(\mathcal{U}) = \frac{1}{B_u}\sum_{i=1}^{B_u} \mathbb{1}(\max_c(\widehat{\boldsymbol{p}}_{i,c}) > \tau) \cdot \mathrm{H}(\widehat{\boldsymbol{p}}_i, \boldsymbol{p}_i^{s}), \tag{16}$$

where τ is the threshold of confidence to filter out unreliable pseudo-labels.

In the GSA strategy, accurately assigning the unsupervised clusters to semantic classes is of paramount importance. To accomplish this, we formulate the task as a constrained centroid matching problem that is equivalent to the MCF problem. Another straightforward approach involves directly utilizing the centroid-prototype similarity distribution $\widetilde{\boldsymbol{p}}_{j,c}$ (from Eq. (6)) to determine the class labels of clusters. However, this simplistic approach without constraints fails to address the issue of collapsing, leading to incorrect structural pseudo-labels. Furthermore, this approach would exacerbate confirmation bias since we employ $\widetilde{\boldsymbol{p}}_{j,c}$ to construct the global target for representation learning.

At last, it is worth noting that the GCC objective and the GSA strategy serve as two complementary components within the GlocalMatch framework, working in tandem to progressively boost one another. On one hand, the GSA strategy is utilized to produce more reliable pseudo-labels for training the classifier, thus providing essential semantic information to optimize the GCC objective. On the other hand, as the glocal cluster compacting is enhanced, the accuracy of structural pseudo-labels also experiences a corresponding increase.

3.5 Training Procedure of GlocalMatch

At last, we present how we exploit the glocal cluster structure via the GCC objective and the GSA strategy in the GlocalMatch framework.

We maintain two memory buffers to store the projected embeddings of all labeled and unlabeled samples. In each iteration, the memory buffers are updated with the new embeddings from the current mini-batches. When all the labeled samples have been processed, we update the class prototypes. And once all the unlabeled samples have been processed, we perform K-Means clustering and attach semantic class labels to the updated centroids through centroid matching. We employ the following unified loss for the concurrent learning of both representation and classification:

$$\mathcal{L} = \mathcal{L}_{\mathrm{s}} + \lambda_c \mathcal{L}_c + \lambda_u \mathcal{L}_u, \tag{17}$$

where λ_c and λ_u are the balancing factors controlling the trade-off for each loss part. Across different tasks, we empirically find that setting $\lambda_c = \lambda_u = 1$ is a simple yet effective choice. The detailed training procedure in each iteration is presented in Algorithm 1, where we define some functions to stand for multiple equations (*e.g.* , aggregate_semantics$(\cdots)$), please refer to the corresponding comments in the code lines.

4 Experiments

4.1 Experimental Setup

4.1.1 Datasets

We have developed three evaluation benchmarks for open-domain SSL with public multi-domain datasets of different scales, namely *CIFAR-STL* [71], *PACS* [72], and *DomainNet* [73]. *CIFAR-STL* is created by combining low-resolution (original 32×32 pixels) images from CIFAR-10 [19] (*i.e.*, the *CIFAR* domain) with high-resolution images (original 96×96 pixels) from STL-10 [20] (*i.e.*, the *STL* domain). It comprises 9 classes of animals and vehicles that are shared between the two domains. *PACS* consists of 7 classes of images from 4 domains: *Art Painting*, *Cartoon*, *Photo*, and *Sketch*. *DomainNet* is a more complex and challenging datasets, which contains images of 345 classes from 6 domains: *Clipart*, *Infograph*, *Painting*, *Quickdraw*, *Real*, and *Sketch*. In the experiments, we resize all the images to a fixed size of 96×96 pixels. Some exampling images are presented in Fig. 5.

4.1.2 Evaluation Protocol

In order to comprehensively evaluate the performance of various methods under the open-domain SSL setting, we consider the classification accuracy across multiple testing sets.

Firstly, we consider the *in-domain* testing set, where the testing samples are collected from the same domain as the labeled data. The *in-domain* accuracy assesses the capacity of methods to leverage open-domain unlabeled data for enhancing classification within the original domain. Furthermore, we construct the *out-of-domain* testing set, comprising testing samples from the domain(s) different from the labeled data. The *out-of-domain* accuracy measures the models' ability generalize to the domain(s) lacking any annotations. Besides, we also report the *overall* accuracy on a testing set containing all the aforementioned in-domain and out-of-domain samples.

Using the uniform evaluation protocol, we compare our proposed GlocalMatch with the latest state-of-the-art baseline methods of standard SSL, including FixMatch [2], FlexMatch [28], AdaMatch [56], FreeMatch [29], and SoftMatch [21]. We additionally compare GlocalMatch with existing solutions for SSL involving feature distribution mismatch, specifically CAFA [43] and BDA [44].

Algorithm 1 Training Procedure of GlocalMatch in Each Iteration

Input: $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{B_l}$ and $\{\boldsymbol{u}_i\}_{i=1}^{B_u}$: Labeled and unlabeled samples. $\mathcal{T}_w(\cdot)$ and $\mathcal{T}_s(\cdot)$: Weak and strong augmentation. $\mathcal{F}(\cdot)$: Base encoder. $\mathcal{G}(\cdot)$: Projection head. $\phi(\cdot)$: Classifier. τ: Confidence threshold. λ_c, λ_u: Weights of losses. $\mathcal{Z}^l$, $\mathcal{Z}^w$, $\mathcal{Z}^s$: Embeddings of all labeled and unlabeled samples. n_i: Index number of current iteration.

if $\lceil N_l/B_l \rceil \mid n_i$ **then**
  $\mathcal{P} = \text{update\_class\_prototypes}(\mathcal{Z}^l)$ ▷ *Once all the labeled samples processed, update the class prototypes (Eq. 5).*
end if
if $\lceil N_u/B_u \rceil \mid n_i$ **then**
  $\mathcal{C}, \mathcal{A} = \text{k\_means}(\mathcal{Z}^w)$ ▷ *Once all the unlabeled samples processed, perform K-Means clustering (Eq. 2).*
  $\mathcal{Q}^* = \text{centroid\_matching}(\mathcal{C}, \mathcal{P})$ ▷ *And solve the centroid matching (Eq. 12-13).*
end if
$\boldsymbol{h} = \mathcal{F}(\mathcal{T}_w(\boldsymbol{x}))$, $\boldsymbol{h}^w = \mathcal{F}(\mathcal{T}_w(\boldsymbol{u}))$, $\boldsymbol{h}^s = \mathcal{F}(\mathcal{T}_s(\boldsymbol{x}))$ ▷ *Obtain the features of the labeled and unlabeled samples.*
$\boldsymbol{z} = \mathcal{G}(\boldsymbol{h})$, $\boldsymbol{z}^w = \mathcal{G}(\boldsymbol{h}^w)$, $\boldsymbol{z}^s = \mathcal{G}(\boldsymbol{h}^s)$ ▷ *Map the features into the projection space.*
$\boldsymbol{p} = \phi(\boldsymbol{h})$, $\boldsymbol{p}^w = \phi(\boldsymbol{h}^w)$, $\boldsymbol{p}^s = \phi(\boldsymbol{h}^s)$ ▷ *The classifier produces semantic predictions.*
$\widehat{W} = \text{get\_glocal\_target}(\mathcal{C}, \mathcal{A}, \mathcal{P}, \boldsymbol{p}^w)$ ▷ *Obtain the glocal target matrix for the GCC optimization (Eq. 6-9)*
$\widehat{\boldsymbol{p}} = \text{aggregate\_semantics}(\mathcal{C}, \mathcal{P}, \mathcal{Q}^*, \boldsymbol{p}^w)$ ▷ *Refine the pseudo-labels via the GSA strategy (Eq. 14-15)*
$\mathcal{L}_s(\mathcal{X}) = \frac{1}{B}\sum_{i=1}^{B} \mathrm{H}(y_i, \boldsymbol{p}_i)$ ▷ *Calculate the supervised loss.*
$\mathcal{L}_c(\mathcal{U}) = \frac{1}{B_u}\sum_{i=1}^{B_u} \mathrm{H}(\widehat{\mathcal{W}}_i, \boldsymbol{q}_i^s)$ ▷ *Calculate the compactness loss.*
$\mathcal{L}_u(\mathcal{U}) = \frac{1}{B_u}\sum_{i=1}^{B_u} \mathbb{1}(\max_c(\widehat{\boldsymbol{p}}_{i,c}) > \tau) \cdot \mathrm{H}(\widehat{\boldsymbol{p}}_i, \boldsymbol{p}_i^s)$ ▷ *Calculate the unlabeled pseudo-labeling loss.*

Output: The overall loss $\mathcal{L} = \mathcal{L}_s + \lambda_c\mathcal{L}_c + \lambda_u\mathcal{L}_u$ to update the network parameters.

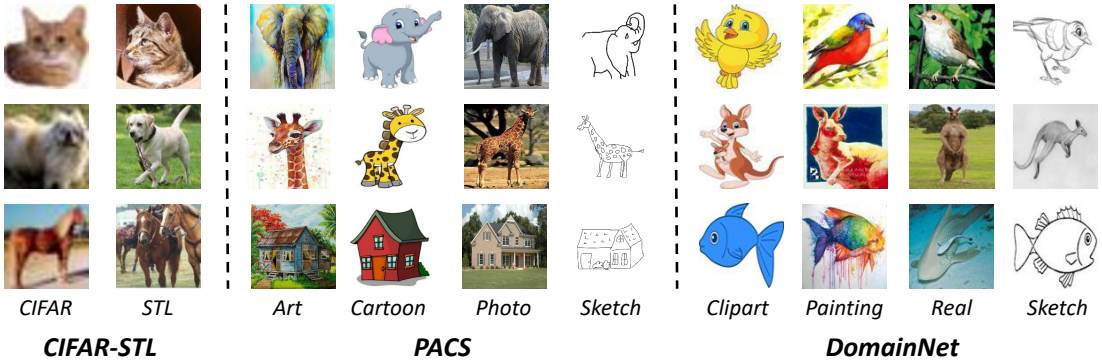


Fig. 5. Examples of open-domain samples from our benchmarks.

4.1.3 Implementation Details

For GlocalMatch, we adopt the WRN-37-2 [74] as the base encoder network, and a MLP with one hidden layer as the projection head. Following the standard setup of mainstream SSL methods [2], the network is optimized using a standard SGD optimizer with the momentum of 0.9 and weight decay of 5×10^{-4}. We adopt an initial learning rate of 0.03 with a cosine learning rate delay scheduler as $lr = 0.03\cos(7\pi n_i/16N_i)$, where n_i is the index number of current iteration and N_{it} is the number of total training iterations. We set $\lambda_c = \lambda_u = 1$, $B_l = B_u = 64$, $\alpha = 0.9$, $\eta = 0.9$, $T = 0.2$ and $N_i = 204800$ across different tasks. A critical hyperparameters is the number of K-Means clusters K. We adopt a unified criterion by setting $K = 100L$, where L represents the total number of classes in the current task. The value of confidence threshold varies slightly across different datasets: For the CIFAR-STL and PACS benchmarks, we adopt a higher threshold, $\tau = 0.95$, to filter out unreliable pseudo-labels. However, for the DomainNet benchmark, we should use $\tau = 0.75$ to fully utilize open-domain unlabeled samples, as the dataset is so complex and challenging that the average confidence is much lower. We employ the *Faiss* [75] and *NetworkX* libraries [76], for efficient implementation of K-Means clustering and MCF problem solving.

For the baseline methods, we employ the implementations from the USB [42] codebase, which provide optimal method-specific hyperparameters. The proposed GlocalMatch is also implemented using the same codebase. To ensure fair comparisons, all the methods adopt the same backbone network. Additionally, the optimizer, learning rate scheduler, and common hyperparameters such as B_l, B_u, and N_{it} are all consistently configured.

4.2 Main Results

4.2.1 CIFAR-STL

First, we provide an overview of the benchmark created using the *CIFAR-STL* dataset. We construct the training set using the original training images of CIFAR-10 and STL-10. Specifically, we randomly select 500 images per shared class from both CIFAR-10 and STL-10 datasets. As a result, the training set comprises a total of $9,000$ images. Subsequently, we randomly choose 5 and 25 samples per class from each domain to serve as labeled data, while the remaining samples from both domains are used as unlabeled data. Considering the labeled domain and the number of labels, there are a total of 4 distinct training data splits. The testing set is likewise selected at random from the original testing

TABLE 2
Performance (including *in*-domain, *out*-of-domain, and over*all* accuracy, %) on the CIFAR-STL benchmark.

Number of Labels		45 (5 labels per class)					
Labeled Domain		CIFAR			STL		
Test Set		In	Out	All	In	Out	All
FixMatch [2]	NeurIPS'20	52.14 ± 8.96	14.55 ± 4.83	33.34 ± 6.87	53.83 ± 6.40	11.11 ± 0.00	35.01 ± 5.83
FlexMatch [28]	NeurIPS'21	46.48 ± 1.88	11.11 ± 0.01	28.79 ± 0.93	55.49 ± 3.33	17.23 ± 3.74	36.36 ± 2.57
AdaMatch [56]	ICLR'22	58.83 ± 2.58	11.12 ± 0.01	34.98 ± 1.29	61.68 ± 4.79	11.14 ± 0.02	36.41 ± 2.39
FreeMatch [29]	ICLR'23	50.15 ± 1.11	11.80 ± 1.00	30.98 ± 0.73	55.21 ± 4.51	17.25 ± 5.53	36.23 ± 3.49
SoftMatch [21]	ICLR'23	60.64 ± 0.95	11.10 ± 0.04	35.87 ± 0.49	66.48 ± 1.90	11.10 ± 0.00	38.79 ± 0.95
CAFA [43]	NeurIPS'21	31.09 ± 1.72	18.01 ± 1.45	24.55 ± 1.54	38.86 ± 1.07	21.58 ± 1.89	30.22 ± 0.56
BDA [44]	ICML'23	34.88 ± 2.32	14.58 ± 1.31	24.73 ± 1.74	41.37 ± 0.55	22.73 ± 1.24	32.05 ± 0.44
GlocalMatch	**Ours**	**61.45 ± 1.73**	**56.80 ± 2.24**	**59.13 ± 1.95**	**69.00 ± 1.24**	**53.05 ± 2.07**	**61.03 ± 0.82**

Number of Labels		225 (25 labels per class)					
Labeled Domain		CIFAR			STL		
Test Set		In	Out	All	In	Out	All
FixMatch [2]	NeurIPS'20	74.55 ± 1.03	14.54 ± 4.84	44.55 ± 2.45	81.10 ± 0.55	11.11 ± 0.00	46.11 ± 0.28
FlexMatch [28]	NeurIPS'21	70.58 ± 0.79	13.58 ± 2.59	42.08 ± 1.49	78.18 ± 1.79	20.12 ± 1.31	49.15 ± 0.70
AdaMatch [56]	ICLR'22	72.51 ± 0.33	11.15 ± 0.02	41.79 ± 0.22	81.39 ± 1.28	28.52 ± 1.94	54.97 ± 1.45
FreeMatch [29]	ICLR'23	70.27 ± 0.53	14.29 ± 4.48	42.28 ± 2.34	78.99 ± 1.09	13.78 ± 1.37	46.38 ± 0.15
SoftMatch [21]	ICLR'23	71.61 ± 0.29	11.30 ± 0.21	41.46 ± 0.19	81.09 ± 0.86	15.33 ± 2.94	48.21 ± 1.72
CAFA [43]	NeurIPS'21	62.41 ± 0.68	21.41 ± 0.92	41.79 ± 0.79	70.39 ± 1.03	25.04 ± 0.72	47.71 ± 0.87
BDA [44]	ICML'23	66.33 ± 0.53	19.44 ± 0.73	42.89 ± 0.79	72.81 ± 0.34	26.99 ± 0.62	49.90 ± 0.25
GlocalMatch	**Ours**	**77.24 ± 0.45**	**72.64 ± 1.19**	**74.94 ± 0.51**	**81.42 ± 0.16**	**66.33 ± 0.12**	**73.86 ± 0.14**

TABLE 3
Performance (including *in*-domain, *out*-of-domain, and over*all* accuracy, %) on the PACS benchmark.

Number of Labels	35 (5 labels per class)											
Labeled Domain	Art			Cartoon			Photo			Sketch		
Test Data	In	Out	All	In	Out	All	In	Out	All	In	Out	All
FixMatch [2]	41.44	21.68	25.73	73.02	13.41	27.39	71.50	16.74	25.88	66.12	18.29	37.09
AdaMatch [56]	36.44	26.89	28.85	72.81	26.38	37.27	72.77	21.12	29.74	72.36	16.98	38.77
SoftMatch [21]	43.91	22.66	27.01	75.68	22.96	35.33	76.00	18.14	27.80	69.79	20.88	40.13
GlocalMatch	**70.46**	**42.86**	**48.52**	**75.82**	**28.10**	**39.30**	**82.53**	**37.00**	**44.60**	**74.99**	**29.72**	**47.53**

Number of Labels	105 (15 labels per class)											
Labeled Domain	Art			Cartoon			Photo			Sketch		
Test Data	In	Out	All	In	Out	All	In	Out	All	In	Out	All
FixMatch [2]	67.26	22.13	31.38	**88.30**	39.99	51.32	84.63	18.24	29.33	86.79	14.83	43.14
AdaMatch [56]	74.68	29.04	38.41	85.57	37.73	48.95	83.67	16.98	28.12	85.94	19.22	45.47
SoftMatch [21]	72.80	28.54	37.61	87.52	42.49	53.05	85.52	15.76	27.44	87.34	28.02	51.36
GlocalMatch	**75.90**	**48.01**	**53.72**	87.97	**51.16**	**59.80**	**88.64**	**41.67**	**49.51**	**89.01**	**48.24**	**64.28**

images of both datasets, containing 500 testing samples per class for each domain. Each experiment is repeated three times with different random seeds. The reported results in Tab. 2 include both the mean accuracy (on *in*-domain, *out*-of-domain, and *all* testing data) and the standard deviation.

It is evident that our proposed GlocalMatch substantially outperforms the baseline methods across all evaluation metrics. The results demonstrate that GlocalMatch effectively mitigates the adverse effects of out-of-domain samples on classification within the original domain, leading to a notable increase in in-domain accuracy. Furthermore, GlocalMatch showcases remarkable performance in the unknown domain without any annotations, as evidenced by its impressive out-of-domain accuracy. In stark contrast, the baseline methods are barely able to generalize to the unknown domain. The reason for this could be attributed to the fact that both *CIFAR* and *STL* datasets are relatively simple with a small number of classes and limited intra-domain variance. This makes the baseline methods prone to overfitting the labeled domain, thereby leading to the occurrence of severe confirmation bias. Therefore, we further consider the more challenging benchmarks, *PACS* and *DomainNet*, that encompass more complex domains and a large number of classes.

4.2.2 PACS

The *PACS* benchmark involves 4 domains, comprising a total of $9,991$ images belonging to 7 classes. It is notable that the sample counts within each class of every domain are imbalanced. Because the original dataset does not provide training and testing splits, we randomly split the entire

TABLE 4
Performance (including *in*-domain, *out*-of-domain, and over*all* accuracy, %) on the DomainNet-65 benchmark.

Number of Labels	260 (4 labels per class)											
Labeled Domain	Clipart			Painting			Real			Sketch		
Test Data	In	Out	All	In	Out	All	In	Out	All	In	Out	All
FixMatch [2]	50.46	23.21	30.03	14.67	8.79	10.26	17.94	5.34	8.49	23.28	6.39	10.62
AdaMatch [56]	53.44	29.26	35.31	17.03	9.85	11.64	25.54	13.40	16.44	26.56	15.62	18.36
SoftMatch [21]	48.72	28.48	33.54	16.21	8.31	10.28	26.46	12.38	15.90	32.82	23.76	26.03
GlocalMatch	**57.64**	**31.32**	**37.90**	**26.46**	**15.87**	**19.80**	**40.62**	**24.17**	**28.28**	**33.64**	**24.75**	**26.97**

dataset into training and testing sets in an approximate $9:1$ ratio. From the training set of $8,981$ images, we randomly choose 5 or 15 samples per class from each domain as labeled data and use the rest as unlabeled data. Similarly, we report the *in*-domain, *out*-of-domain, and over*all* performance in Tab. 3, presented by the mean accuracy of three random runs.

The results substantiate the robustness of GlocalMatch in scenarios where the unlabeled data include out-of-domain samples from multiple domains. In the *PACS* benchmark, the *Art* domain stands out as particularly challenging due to the high diversity of images within it. However, GlocalMatch continues to exhibit high performance, surpassing its strongest rival by an impressive margin of 19.67% in overall accuracy for the 5-label-per-class task. In average of all the 4 labeled domains, GlocalMatch outperforms the previous SOTA method by 11.34% and 14.47% in overall accuracy, with 5 and 15 labels per class, respectively.

4.2.3 DomainNet

The full *DomainNet* dataset is at a large scale, demanding significant computational resources for training. Additionally, as highlighted in prior studies [71], [77], some domains and classes in the original dataset suffer from noisy labels. To address this, we construct a subset of DomainNet by excluding domains and classes with noisy labels and a limited number of samples. As a result, the *DomainNet-65* benchmark used in this work comprise 65 classes from 4 domains, namely *Clipart*, *Painting*, *Real*, and *Sketch*. Each class within every domain contains 80 images for training and 15 images for testing. From the training set, we randomly choose 4 samples as labeled data. The results are presented in Tab. 4.

In such a complex and challenging benchmark, GlocalMatch still excel across all the tasks. On average, it achieves an improvement in overall accuracy of 6.80% over the previous SOTA method, which further verifies the effectiveness and robustness of GlocalMatch.

4.2.4 Summary of Results

Finally, we provide a summary of the results obtained from the three benchmarks. To be specific, we present the average overall accuracy across all labeled domains within each benchmark. The results in Tab. 5 demonstrate the significant advantage of GlocalMatch over baseline methods in terms of generalization performance: By labeling an extremely limited proportion of samples from a single domain, GlocalMatch can effectively learn general semantic knowledge that benefits the classification on multiple different domains.

TABLE 5
Comprehensive performance (represented by average overall accuracy, %) across the three benchmarks.

Dataset	CIFAR-STL		PACS		DomainNet
Labels per Class	5	25	5	15	4
FixMatch [2]	34.18	45.33	29.02	38.79	14.85
FlexMatch [28]	32.58	45.62	-	-	-
AdaMatch [56]	35.69	48.38	33.65	40.23	20.44
FreeMatch [29]	33.61	44.33	-	-	-
SoftMatch [21]	37.33	44.84	32.57	42.36	21.44
GlocalMatch	**60.08**	**74.40**	**44.99**	**56.83**	**28.24**

TABLE 6
Results (including in-domain and out-of-domain accuracy, %) of ablation study on the CIFAR-STL benchmark.

Number of Labels	45 (5 labels per class)			
Labeled Domain	CIFAR		STL	
Test Data	In	Out	In	Out
FixMatch	53.49	11.11	58.66	11.11
GM w/o GCC	54.78	26.15	59.86	24.47
GM w/o GSA	59.13	32.73	62.42	30.95
GM w/o $\mathcal{W}^{global}$	63.20	51.22	65.09	49.40
GM w/o CM	62.13	46.67	63.82	44.53
GM w/o *proj.*	61.24	53.64	65.18	52.60
GlocalMatch	**63.89**	**59.64**	**68.86**	**55.49**

4.3 Ablation Study

We verify the effectiveness of the components and corresponding designs in GlocalMatch, with the experiments conducted on the *CIFAR-STL* benchmark. The results are presented in Tab. 6.

We use FixMatch as the baseline, which includes only $\mathcal{L}_s$ and $\mathcal{L}_u$ (without refining pseudo-labels via GSA). Firstly, we remove the GCC objective $\mathcal{L}_c$ from GlocalMatch (denoted as "GM w/o GCC") and notice a considerable drop in both the in-domain and out-of-domain accuracy. This indicates that the glocal compactness of clusters is critical. Next, we disable the GSA strategy by setting $\alpha = 1$ so that the original $\boldsymbol{p}^w$'s are utilized as pseudo-labels without refinement. This absence of GSA also leads to significant performance degradation, particularly in terms of the out-of-domain accuracy. Therefore, the efficacy of the two core components of GlocalMatch, GCC and GSA, has been validated.

Furthermore, the results demonstrate the significance of certain technical designs in GlocalMatch. In GCC, it is important to take into account the global cluster information encoded in $\mathcal{W}^{global}$ for generalizing to out-of-domain samples. In GSA, we solve the Centroid Matching (CM) problem to establish the association of semantics with clusters.

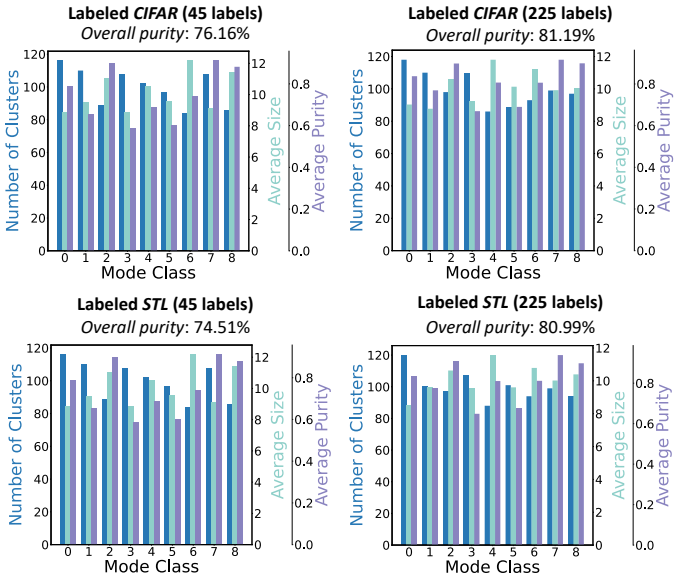


Fig. 6. A quantitative analysis on clustering. We present the quantitative metrics of K-Means clusters, including the number of clusters, average size, and average purity for each class in CIFAR-STL.

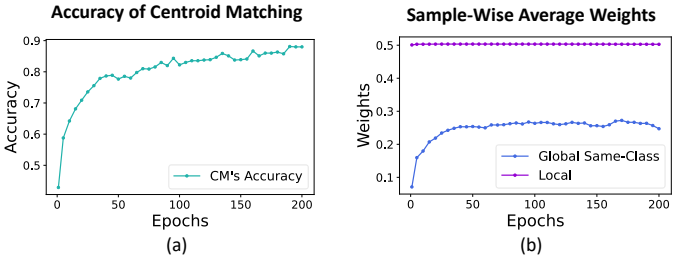


Fig. 7. Visualization of GlocalMatch training procedure on CIFAR-STL (Labeled CIFAR, 225 labels). We present (a) the accuracy of centroid matching algorithm, and (b) the weights of local and global same-class centroids, to demonstrate the gradual enhancement of the glocal cluster structure.

Adopting the trivial approach using the centroid-prototype similarity (referred as "GM w/o CM") will greatly undermine the out-of-domain performance. At last, we observe the importance of the projection space. If we omit the projection head and directly optimize the GCC objective in the feature space for classification, the performance will also be negatively impacted.

4.4 Further Analysis and Discussions

4.4.1 Analyses on Glocal Clustering

We use K-Means to explore the local cluster structure of unlabeled open-domain samples. In spite of its simplicity, we demonstrate that it can effectively offer valuable structural information for semi-supervised learning. We first introduce the concept of a "mode class" for each cluster, which refers to the class that the majority of samples within that cluster belong to. It can be seen as the ground-truth semantic label for each cluster. Next, we define the "purity" of each cluster as the ratio of samples belonging to the mode class. In Fig. 6, we show the average purity of clusters of each class. We also present the number of clusters and average size for each class. Even when there are extremely limited labels available during training, the overall purity of clusters is still high,

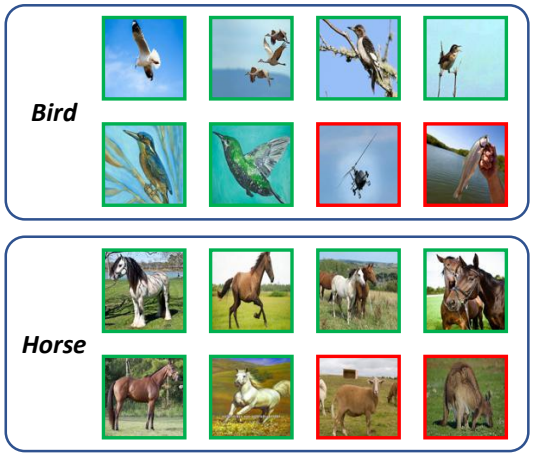


Fig. 8. A qualitative analysis on clustering. Wrongly clustered samples are denoted by red boxes. The visually similar backgrounds lead the model to misidentify "Helicopter" as "Bird" and "Sheep" as "Horse".

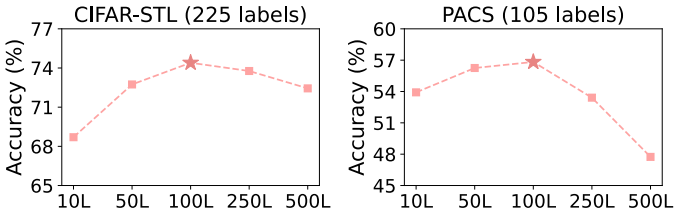


Fig. 9. Average overall accuracy (%) on the CIFAR-STL benchmark with different numbers of clustering centroids. The results show that $K = 100L$ is an appropriate choice across different benchmarks.

which validates the existence of the cluster structure from the local perspective.

To demonstrate the existence and gradual enhancement of the cluster structure from the global perspective, we visualize the accuracy of centroid matching as Fig. 7(a) and the sample-wise average weights of local and global same-class centroids in the normalized glocal clustering targets $\widehat{\mathcal{W}}$ as Fig. 7(b). For each sample, the local centroid is the centroid of its local cluster, and the global same-class centroids are the other centroids sharing the same class with its local centroid. The sum of weights of global same-class centroids indicates the global structural information correctly utilized.

In addition to the above quantitative analyses, we also provide a qualitative analysis to reveal where glocal clustering could fail. In Fig. 8, we visualize some examples which are assigned to the wrong clusters, denoted by red boxes. As the clustering procedure is built upon semantic similarity, mistakes may arise from the visual similarity of backgrounds, such as the "blue sky" and "grass ground" in the examples. To further improve the clustering performance, incorporating representation disentanglement techniques [78]–[80] to extract more informative features could be beneficial.

4.4.2 Number of Clustering Centroids

Experiments on the *CIFAR-STL* and *PACS* benchmarks demonstrate that the number of clustering centroids, K, is critical for the performance of GlocalMatch. Specifically, we traverse the values $\{10L, 50L, 100L, 250L, 500L\}$, where L is the number of classes. The average overall accuracy is presented in Fig. 9.

TABLE 7
Performance on CIFAR-STL under the standard SSL setting.

Number of Labels	45 (5 labels per class)						225 (25 labels per class)					
Labeled Domain	CIFAR			STL			CIFAR			STL		
Test Data	In	Out	All	In	Out	All	In	Out	All	In	Out	All
FixMatch [2]	56.24	46.22	51.23	57.75	17.46	37.61	71.78	52.60	61.19	79.80	25.04	52.42
SoftMatch [21]	**63.09**	50.63	**56.86**	67.13	20.63	43.88	74.13	58.49	66.31	**82.13**	28.01	55.07
GlocalMatch	61.49	**51.09**	56.29	**69.00**	**25.72**	**47.36**	**76.00**	**60.48**	**68.24**	81.09	**33.47**	**57.28**

TABLE 8
Performance on CIFAR-STL with all out-of-domain unlabeled samples.

Number of Labels	45 (5 labels per class)						225 (25 labels per class)					
Labeled Domain	CIFAR			STL			CIFAR			STL		
Test Data	In	Out	All	In	Out	All	In	Out	All	In	Out	All
AdaMatch [56]	36.24	23.90	30.07	36.38	21.44	28.91	53.47	33.71	43.59	55.38	14.66	35.02
SoftMatch [21]	31.98	21.50	26.74	38.22	20.02	29.12	52.36	39.40	45.88	53.73	18.79	35.26
GlocalMatch	**48.93**	**58.03**	**53.48**	**46.67**	**46.27**	**46.47**	**61.84**	**67.16**	**64.50**	**66.24**	**60.74**	**63.49**

TABLE 9
Performance on VisDA2017 under the setting of [44].

Methods	150 labels		300 labels	
	S/R	R/S	S/R	R/S
Mean Teacher [25]	84.15 ± 1.08	73.68 ± 1.00	86.90 ± 0.61	76.90 ± 0.46
FixMatch [2]	78.46 ± 4.15	67.10 ± 9.46	82.88 ± 0.85	71.74 ± 0.45
FlexMatch [28]	83.43 ± 1.74	67.90 ± 1.77	88.09 ± 0.53	75.17 ± 1.34
UASD [12]	85.58 ± 1.55	78.49 ± 0.41	89.58 ± 0.79	81.82 ± 0.68
CAFA [43]	83.95 ± 1.79	72.89 ± 1.03	87.81 ± 0.47	76.48 ± 0.72
BDA [44]	85.92 ± 1.16	79.15 ± 0.39	89.85 ± 0.71	82.27 ± 0.60
GlocalMatch	**88.17 ± 0.85**	**83.44 ± 0.51**	**93.24 ± 0.69**	**85.08 ± 0.43**

TABLE 10
Performance on the more complex DomainNet-126 benchmark.

Number of Labels	504 (4 labels per class)							
Labeled Domain	Clipart		Painting		Real		Sketch	
Test Data	In	All	In	All	In	All	In	All
FixMatch [2]	42.48	20.36	12.22	6.71	21.84	10.36	22.38	10.16
AdaMatch [56]	47.59	24.68	16.17	8.29	24.90	15.20	29.65	16.75
SoftMatch [21]	45.95	23.17	15.41	12.70	25.06	14.44	28.49	17.38
GlocalMatch	**51.43**	**26.31**	**23.17**	**15.28**	**39.37**	**22.38**	**33.65**	**19.68**

In GlocalMatch, we rely on the clusters produced by K-Means to optimize feature representations and assist pseudo-labeling. Therefore, it is crucial to ensure that as many samples as possible within each cluster belong to the same class. Supported by experimental results, we found that choosing an appropriately larger number of clusters helps improve the purity of clusters, reducing the negative impact of intra-domain variance. On the other hand, setting K to be too large can also negatively impact the classification performance, particularly when there are only a limited number of samples for each class within a single domain. This is because it may results in meaningless clusters containing only one sample. The experimental results support that $100L$ is appropriate to achieve high performance.

4.4.3 Scalability on Various Settings

Although this work mainly focuses on the open-domain SSL problem where unlabeled data comprise samples drawn from different domains, we would like to state that the proposed GlocalMatch is actually a general framework to utilize unlabeled samples collected in various scenarios. As presented in Tab. 7, GlocalMatch can achieve comparable or better in-domain performance compared with the SOTA standard SSL method under the standard SSL setting. Besides, GlocalMatch can generalize better to the out-of-domain samples even if it has never seen such samples during training, no matter labeled or unlabeled.

We also evaluate the methods in the scenario where all the unlabeled data come from a different domain than the labeled data, mirroring the specific case studied in [43] and [44]. In addition to the evaluation on the CIFAR-STL benchmark (Tab. 8), we extend our experiments to match the setup used in [44], showing the scalability of GlocalMatch on another large-scale benchmark, VisDA2017 [81] (Tab. 9).

For validating the scalability of GlocalMatch on more challenging open-domain SSL tasks, we construct a considerably more intricate benchmark, *DomainNet-126*, which contains $137,486$ unlabeled samples of 126 classes. In addition to the massive scale of the unlabeled dataset, there is a significant issue of severe class imbalance caused by the long-tailed distribution, further adding to the complexity. In spite of the high complexity, GlocalMatch continues to showcase its effectiveness on this challenging benchmark, as shown in Tab. 10.

4.4.4 Computation Efficiency

When compared to the simplest baseline method FixMatch, GlocalMatch introduces extra computation primarily from the K-Means clustering. Theoretical analysis suggests that the average time complexity of K-Means is approximately $O(N_u K)$. In practice, when running on a single NVIDIA RTX 2080 Ti GPU, GlocalMatch incurs approximately 12%, 29%, and 62% additional training time compared to FixMatch on *CIFAR-STL*, *PACS*, and *DomainNet*, respectively.

The additional memory overhead comes from the memory buffers storing the projected embeddings. For a task involving approximately $N_u = 10^5$ images and an embedding dimension of $d = 128$, the GPU memory cost is merely around 50 MB. This amount is significantly lower than the initial GPU memory consumption during network training and can be considered negligible.

4.4.5 Generalizability across Diverse Domains

It is meaningful to think about how different the domains could be to remain the generalizability of GlocalMatch. In this regard, we believe a necessary condition is that the involved domains should share some common semantically related visual elements, such as similar edge shapes of objects in the same class, even in the presence of significant visual diversity among the domains. Additionally, based on extensive experimental results, we observe that the smaller the differences between the domains in terms of visual appearances, the better the performance of cross-domain generalization. This observation aligns with existing theoretical analyses [82] in the literature.

4.4.6 Limitations and Future Work

Finally, we discuss the limitations of our current work and suggest potential directions for improvement.

More Efficient Training: In the GlocalMatch framework, we utilize the offline K-Means algorithm, which involves clustering all samples' embeddings. However, it might be less adaptable for larger datasets. To address this, we could explore incorporating online clustering methods [83], [84] into GlocalMatch.

More Realistic Settings: In the real-world SSL applications, the class and feature distribution mismatch may exist at the same time. Therefore, it is meaningful to consider the open-set and open-domain problems in a unified SSL setting. Besides, due to the inherent nature of data, real-world applications may also encounter challenges related to fine-grained class categories and long-tail distribution. We will make efforts to explore more practical and challenging scenarios in the future.

Theoretical Analyses: We will focus on conducting theoretical analyses of factors influencing the generalization and applicability of semi-supervised learning algorithms in real-world scenarios. Based on existing experimental results, such factors may include the intra-domain feature diversity of labeled and unlabeled data, as well as the differences between different domains. The theoretical analyses will assist us in constructing *safe* SSL algorithms [13], [85], [86], guaranteeing they perform no worse when training on additional out-of-domain unlabeled samples.

5 Conclusion

In this paper, we take the first step to systematically investigate the open-domain semi-supervised learning problem, where the feature distribution mismatch problem exists between labeled and unlabeled data. In order to tackle this practical yet challenging problem, we analyze why existing methods based on pseudo-labeling fail generalizing to out-of-domain samples. Then we propose a novel framework, GlocalMatch, which aims to exploit both local and glocal cluster structure of open-domain unlabeled data. Two complementary components, namely the glocal cluster compacting (GCC) objective and the glocal semantic aggregation (GSA) strategy, are introduced for the simultaneous learning of discriminative feature representation and reliable pseudo-label production. Extensive experiments have been conducted, and the results demonstrate the significant superiority of GlocalMatch compared with all the baseline methods across the tasks at different scales.

Acknowledgments

This work is supported by NSFC Program (62222604, 62206052, 62192783) and Jiangsu Natural Science Foundation Project (BK20210224).

References

[1] O. Chapelle, B. Scholkopf, and A. Zien, *Semi-Supervised Learning*. MIT Press, 2006.
[2] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," in *NeurIPS*, 2020.
[3] Y.-C. Liu, C.-Y. Ma, Z. He, C.-W. Kuo, K. Chen, P. Zhang, B. Wu, Z. Kira, and P. Vajda, "Unbiased teacher for semi-supervised object detection," in *ICLR*, 2021.
[4] L. Yang, L. Qi, L. Feng, W. Zhang, and Y. Shi, "Revisiting weak-to-strong consistency in semi-supervised semantic segmentation," in *CVPR*, 2023.
[5] F. He, F. Nie, R. Wang, H. Hu, W. Jia, and X. Li, "Fast semi-supervised learning with optimal bipartite graph," *IEEE TKDE*, vol. 33, no. 9, pp. 3245–3257, 2021.
[6] B. Zhang, Q. Qiang, F. Wang, and F. Nie, "Fast multi-view semi-supervised learning with learned graph," *IEEE TKDE*, vol. 34, no. 1, pp. 286–299, 2022.
[7] Y. Mi, W. Liu, Y. Shi, and J. Li, "Semi-supervised concept learning by concept-cognitive learning and concept space," *IEEE TKDE*, vol. 34, no. 5, pp. 2429–2442, 2022.
[8] Z. Wang, L. Zhang, R. Wang, F. Nie, and X. Li, "Semi-supervised learning via bipartite graph construction with adaptive neighbors," *IEEE TKDE*, vol. 35, no. 5, pp. 5257–5268, 2023.
[9] Y. Li, H. Xiong, Q. Wang, L. Kong, H. Liu, H. Li, J. Bian, S. Wang, G. Chen, D. Dou, and D. Yin, "Coltr: Semi-supervised learning to rank with co-training and over-parameterization for web search," *IEEE TKDE*, 2023.
[10] A. Oliver, A. Odena, C. A. Raffel, E. D. Cubuk, and I. Goodfellow, "Realistic evaluation of deep semi-supervised learning algorithms," in *NeurIPS*, 2018.
[11] Q. Yu, D. Ikami, G. Irie, and K. Aizawa, "Multi-task curriculum framework for open-set semi-supervised learning," in *ECCV*, 2020, pp. 438–454.
[12] Y. Chen, X. Zhu, W. Li, and S. Gong, "Semi-supervised learning under class distribution mismatch," in *AAAI*, 2020.
[13] L.-Z. Guo, Z.-Y. Zhang, Y. Jiang, Y.-F. Li, and Z.-H. Zhou, "Safe deep semi-supervised learning for unseen-class unlabeled data," in *ICML*, 2020.
[14] J. Huang, C. Fang, W. Chen, Z. Chai, X. Wei, P. Wei, L. Lin, and G. Li, "Trash to treasure: harvesting ood data with cross-modal matching for open-set semi-supervised learning," in *ICCV*, 2021, pp. 8310–8319.
[15] K. Saito, D. Kim, and K. Saenko, "Openmatch: Open-set consistency regularization for semi-supervised learning with outliers," in *NeurIPS*, 2021.
[16] R. He, Z. Han, X. Lu, and Y. Yin, "Safe-student for safe deep semi-supervised learning with unseen-class unlabeled data," in *CVPR*, 2022.
[17] ——, "Safer-student for safe deep semi-supervised learning with unseen-class unlabeled data," *IEEE TKDE*, 2023.
[18] Z. Li, L. Qi, Y. Shi, and Y. Gao, "Iomatch: Simplifying open-set semi-supervised learning with joint inliers and outliers utilization," in *ICCV*, 2023.
[19] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.
[20] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *AISTATS*, 2011.
[21] H. Chen, R. Tao, Y. Fan, Y. Wang, J. Wang, B. Schiele, X. Xie, B. Raj, and M. Savvides, "Softmatch: Addressing the quantity-quality trade-off in semi-supervised learning," in *ICLR*, 2023.
[22] P. Bachman, O. Alsharif, and D. Precup, "Learning with pseudo-ensembles," in *NeurIPS*, 2014.
[23] D. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *ICML Workshop*, 2013.
[24] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in *ICLR*, 2017.

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

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

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

[28] B. Zhang, Y. Wang, W. Hou, H. Wu, J. Wang, M. Okumura, and T. Shinozaki, "Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling," in *NeurIPS*, 2021.

[29] Y. Wang, H. Chen, Q. Heng, W. Hou, Y. Fan, Z. Wu, J. Wang, M. Savvides, T. Shinozaki, B. Raj, B. Schiele, and X. Xie, "Freematch: Self-adaptive thresholding for semi-supervised learning," in *ICLR*, 2023.

[30] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *CVPR*, 2018.

[31] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. E. Hinton, "Big self-supervised models are strong semi-supervised learners," in *NeurIPS*, 2020.

[32] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *CVPR*, 2020.

[33] J. Li, C. Xiong, and S. C. Hoi, "Comatch: Semi-supervised learning with contrastive graph regularization," in *ICCV*, 2021.

[34] M. Zheng, S. You, L. Huang, F. Wang, C. Qian, and C. Xu, "Simmatch: Semi-supervised learning with similarity matching," in *CVPR*, 2022.

[35] F. Yang, K. Wu, S. Zhang, G. Jiang, Y. Liu, F. Zheng, W. Zhang, C. Wang, and L. Zeng, "Class-aware contrastive semi-supervised learning," in *CVPR*, 2022.

[36] I. Nassar, M. Hayat, E. Abbasnejad, H. Rezatofighi, and G. Haffari, "Protocon: Pseudo-label refinement via online clustering and prototypical consistency for efficient semi-supervised learning," in *CVPR*, 2023.

[37] J. Yu, D. Tao, and M. Wang, "Adaptive hypergraph learning and its application in image classification," *IEEE TIP*, vol. 21, no. 7, pp. 3262–3272, 2012.

[38] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2017.

[39] J. Li, Y. Huang, H. Chang, and Y. Rong, "Semi-supervised hierarchical graph classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 6265–6276, 2022.

[40] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[41] X. Yang, Z. Song, I. King, and Z. Xu, "A survey on deep semi-supervised learning," *IEEE TKDE*, vol. 35, no. 9, pp. 8934–8954, 2023.

[42] Y. Wang, H. Chen, Y. Fan, W. Sun, R. Tao, W. Hou, R. Wang, L. Yang, Z. Zhou, L.-Z. Guo, H. Qi, Z. Wu, Y.-F. Li, S. Nakamura, W. Ye, M. Savvides, B. Raj, T. Shinozaki, B. Schiele, J. Wang, X. Xie, and Y. Zhang, "Usb: A unified semi-supervised learning benchmark for classification," in *NeurIPS*, 2022.

[43] Z. Huang, C. Xue, B. Han, J. Yang, and C. Gong, "Universal semi-supervised learning," in *NeurIPS*, 2021.

[44] L.-H. Jia, L.-Z. Guo, Z. Zhou, J.-J. Shao, Y. Xiang, and Y.-F. Li, "Bidirectional adaptation for robust semi-supervised learning with inconsistent data distributions," in *ICML*, 2023.

[45] G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," *ACM TIST*, vol. 11, no. 5, pp. 1–46, 2020.

[46] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in *NeurIPS*, 2007.

[47] Y. Mansour, M. Mohri, and A. Rostamizadeh, "Domain adaptation with multiple sources," in *NeurIPS*, 2008.

[48] C. Cortes, Y. Mansour, and M. Mohri, "Learning bounds for importance weighting." in *NeurIPS*, 2010.

[49] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *ICML*, 2015.

[50] M. Long, H. Zhu, J. Wang, and M. Jordan, "Deep transfer learning with joint adaptation networks," in *ICML*, 2017.

[51] G. Kang, L. Jiang, Y. Yang, and A. Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," in *CVPR*, 2019.

[52] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *JMLR*, vol. 13, no. 1, pp. 723–773, 2012.

[53] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *CVPR*, 2017.

[54] M. Long, Z. Cao, J. Wang, and M. Jordan, "Conditional adversarial domain adaptation," in *NeurIPS*, 2018.

[55] Z. Du, J. Li, H. Su, L. Zhu, and K. Lu, "Cross-domain gradient discrepancy minimization for unsupervised domain adaptation," in *CVPR*, 2021.

[56] D. Berthelot, R. Roelofs, K. Sohn, N. Carlini, and A. Kurakin, "Adamatch: A unified approach to semi-supervised learning and domain adaptation," in *ICLR*, 2022.

[57] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *ICML*, 2015.

[58] D. Berthelot, N. Carlini, E. D. Cubuk, A. Kurakin, K. Sohn, H. Zhang, and C. Raffel, "Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring," in *ICLR*, 2020.

[59] K. Kamnitsas, D. Castro, L. L. Folgoc, I. Walker, R. Tanno, D. Rueckert, B. Glocker, A. Criminisi, and A. Nori, "Semi-supervised learning via compact latent space clustering," in *ICML*, 2018.

[60] M. N. Rizve, N. Kardan, and M. Shah, "Towards realistic semi-supervised learning," in *ECCV*, 2022.

[61] E. Fini, P. Astolfi, K. Alahari, X. Alameda-Pineda, J. Mairal, M. Nabi, and E. Ricci, "Semi-supervised learning made simple with self-supervised clustering," in *CVPR*, 2023.

[62] K. Saito, D. Kim, S. Sclaroff, and K. Saenko, "Universal domain adaptation through self supervision," in *NeurIPS*, 2020.

[63] X. Yue, Z. Zheng, S. Zhang, Y. Gao, T. Darrell, K. Keutzer, and A. S. Vincentelli, "Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation," in *CVPR*, 2021.

[64] H. Tang, X. Zhu, K. Chen, K. Jia, and C. L. P. Chen, "Towards uncovering the intrinsic data structures for unsupervised domain adaptation using structurally regularized deep clustering," *IEEE TPAMI*, 2021.

[65] J. Li, P. Zhou, C. Xiong, and S. C. Hoi, "Prototypical contrastive learning of unsupervised representations," in *ICLR*, 2021.

[66] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised learning of visual features by contrasting cluster assignments," in *NeurIPS*, 2020.

[67] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *ICML*, 2020.

[68] P. S. Bradley, K. P. Bennett, and A. Demiriz, "Constrained k-means clustering," Microsoft Research, Redmond, Tech. Rep. MSR-TR-2000-65, 2000.

[69] D. P. Bertsekas, *Linear network optimization*. MIT Press Cambridge, MA, 1991.

[70] Z. Király and P. Kovács, "Efficient implementations of minimum-cost flow algorithms," *arXiv preprint arXiv:1207.6381*, 2012.

[71] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang, "Domain adaptive ensemble learning," *IEEE TIP*, vol. 30, pp. 8008–8018, 2021.

[72] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Deeper, broader and artier domain generalization," in *ICCV*, 2017.

[73] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *ICCV*, 2019.

[74] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *BMVC*, 2016.

[75] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *IEEE TBD*, vol. 7, no. 3, pp. 535–547, 2019.

[76] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using networkx," Los Alamos National Lab.(LANL), Los Alamos, NM (United States), Tech. Rep., 2008.

[77] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko, "Semi-supervised domain adaptation via minimax entropy," in *ICCV*, 2019.

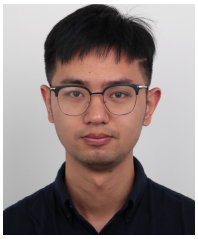
[78] Y.-C. Liu, Y.-Y. Yeh, T.-C. Fu, S.-D. Wang, W.-C. Chiu, and Y.-C. F. Wang, "Detach and adapt: Learning cross-domain disentangled deep representation," in *CVPR*, 2018.

[79] X. Peng, Z. Huang, X. Sun, and K. Saenko, "Domain agnostic learning with disentangled representations," in *ICML*, 2019.

[80] W. Deng, L. Zhao, Q. Liao, D. Guo, G. Kuang, D. Hu, M. Pietikäinen, and L. Liu, "Informative feature disentanglement

for unsupervised domain adaptation," *IEEE TMM*, vol. 24, pp. 2407–2421, 2021.
[81] X. Peng, B. Usman, N. Kaushik, J. Hoffman, D. Wang, and K. Saenko, "Visda: The visual domain adaptation challenge," *arXiv preprint arXiv:1710.06924*, 2017.
[82] I. Redko, E. Morvant, A. Habrard, M. Sebban, and Y. Bennani, "A survey on domain adaptation theory: learning bounds and theoretical guarantees," *arXiv preprint arXiv:2004.11829*, 2020.
[83] X. Zhan, J. Xie, Z. Liu, Y.-S. Ong, and C. C. Loy, "Online deep clustering for unsupervised representation learning," in *CVPR*, 2020.
[84] Q. Qian, Y. Xu, J. Hu, H. Li, and R. Jin, "Unsupervised visual representation learning by online constrained k-means," in *CVPR*, 2022.
[85] Y.-F. Li and Z.-H. Zhou, "Towards making unlabeled data never hurt," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 1, pp. 175–188, 2014.
[86] H. Schmutz, O. Humbert, and P.-A. Mattei, "Don't fear the unlabelled: safe semi-supervised learning via debiasing," in *ICLR*, 2023.

Zekun Li is currently pursuing his Ph.D. degree with the Department of Computer Science and Technology, Nanjing University, China. His research interests include various label-efficient learning methods, especially semi-supervised learning.



Lei Qi is currently an Assistant Researcher with the School of Computer Science and Engineering, Southeast University, China. His current research interests include some ML methods, such as domain adaptation, semi-supervised learning, unsupervised learning, and meta-learning. For applications, he mainly focuses on person re-identification and image segmentation.

Yawen Li Yawen Li is an associate professor at the School of Economics and Management, Beijing University of Posts and Telecommunications. She received her Ph.D. from Tsinghua University in 2018. Her research interest focuses on green innovation, knowledge graph analysis, and data mining.

Yinghuan Shi is currently an Associate Professor at the Department of Computer Science and Technology, Nanjing University, and he is also affiliated with National Institute of Healthcare Data Science, Nanjing University. He received the B.Sc. and Ph.D. degrees both from Computer Science, Nanjing University, in 2007 and 2013, respectively. His research interests include machine learning, pattern recognition, and medical image analysis.

Yang Gao (Senior Member, IEEE) is a Professor in the Department of Computer Science and Technology, Nanjing University. He is currently directing the Reasoning and Learning Research Group in Nanjing University. He has published more than 100 papers in top-tired conferences and journals. He also serves as Program Chair and Area Chair for many international conferences. His current research interests include artificial intelligence and machine learning.