

Semi-supervised Domain Adaptation via Prototype-based Multi-level Learning

Xinyang Huang^{1,2}, Chuang Zhu^{1*}, Wenkai Chen¹

¹School of Artificial Intelligence, Beijing University of Posts and Telecommunications

²School of Artificial Intelligence, Xidian University

hsinyanghuang7@gmail.com, {czhu, wkchen}@bupt.edu.com

Abstract

In semi-supervised domain adaptation (SSDA), a few labeled target samples of each class help the model to transfer knowledge representation from the fully labeled source domain to the target domain. Many existing methods ignore the benefits of making full use of the labeled target samples from multi-level. To make better use of this additional data, we propose a novel Prototype-based Multi-level Learning (ProML) framework to better tap the potential of labeled target samples. To achieve intra-domain adaptation, we first introduce a pseudo-label aggregation based on the intra-domain optimal transport to help the model align the feature distribution of unlabeled target samples and the prototype. At the inter-domain level, we propose a cross-domain alignment loss to help the model use the target prototype for cross-domain knowledge transfer. We further propose a dual consistency based on prototype similarity and linear classifier to promote discriminative learning of compact target feature representation at the batch level. Extensive experiments on three datasets, including *DomainNet*, *VisDA2017*, and *Office-Home*, demonstrate that our proposed method achieves state-of-the-art performance in SSDA. Our code is available at <https://github.com/bupt-ai-cz/ProML>.

1 Introduction

The deep neural network has achieved great success in various visual tasks. However, due to the high cost of labeled data it requires and the degradation of model performance when deploying models in a new domain (target), unsupervised domain adaptation (UDA) has been proposed. To further improve the model performance, many semi-supervised domain adaptation (SSDA) works are proposed by adding a few (e.g., one sample per class) labeled target samples based on UDA. Compared to UDA, the key to SSDA is whether the model can make better use of the additional limited target samples to help the model better fit the learned features from the source domain to the target domain.

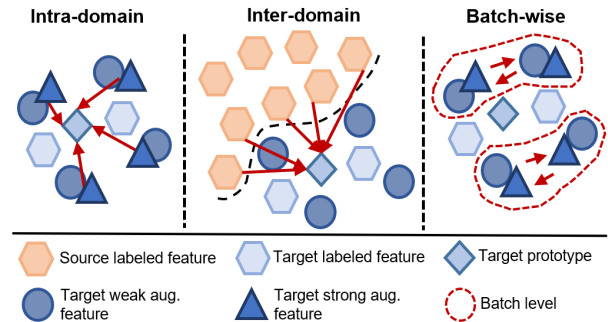


Figure 1: Overview of our Prototype-based Multi-level Learning (ProML) framework. Our ProML framework helps models transfer knowledge using the target prototype from (1) intra-domain level, (2) inter-domain level, and (3) batch level. The arrows represent the feature alignment, and *aug.* represents augmentation.

In recent years, to help model knowledge transfer, some existing SSDA studies make better use of unlabeled samples in target domains by combining self-supervised learning [Pérez-Carrasco *et al.*, 2022] and semi-supervised learning [Li *et al.*, 2021a; Yan *et al.*, 2022]. However, they ignore the potential of labeled samples in the target domain. Although [Li *et al.*, 2021b; Singh, 2021] use few target samples by building a prototype, they only use the prototype from one level unilaterally and ignore the additional knowledge learned from other levels (e.g. intra-domain level). At the same time, it is not robust to use only a very limited number of label or pseudo-label samples to build prototypes. To fill this gap, we propose a robust prototype-based multi-level learning (ProML) framework. Our prototype is updated by both labeled and unlabeled target samples which helps the model to transfer knowledge across domains more robustly. Our ProML makes full use of the target prototype from three levels: intra-domain, inter-domain, and batch-wise to help the model better transfer knowledge through target samples, as shown in Figure 1.

Specifically, at the intra-domain level, we propose a pseudo-labeling strategy based on the optimal transport to help the model align the data distribution of labeled and unlabeled samples in the target domain, obtaining a more compact intra-domain feature distribution and robust pseudo-labels. This method improves the pseudo-labels of target samples by resolving the optimal transport plan between the weakly aug-

*Corresponding Author

mented target samples and the target prototype. At the same time, the transport plan is also applied to the strongly augment view to form a consistency constraint. At the inter-domain level, we cross-align the source samples with the same class of the target prototype. Inter-domain prototype alignment helps the model learn better about cross-domain knowledge transfer and category alignment. At the batch level, different from work [Yan *et al.*, 2022], we consider a mini-batch of samples to calculate class correlation matrices between predictions with different augments from two perspectives of prediction probability and prototype similarity, increasing the correlation in the same class and reducing the correlation from different classes. Classifiers from different perspectives represent features comprehensive to learn the relationship between the batch-wise target samples. Combining the above three, through our ProML framework, the model can capture the knowledge of target domain samples from different levels, to better enable the model to learn more comprehensive and complementary domain-adaptive knowledge.

Our main contributions can be summarized as follows:

(1) We propose a novel Prototype-based Multi-level Learning (ProML) framework, which exploits the potential of labeled target samples by making full use of the target prototype from multiple levels.

(2) We propose a pseudo-label aggregation based on intra-domain optimal transport to SSDA, which helps the model form a more compact target domain robustly.

(3) We propose a batch-wise dual consistency, which helps the model learn more distinctive target representations from different perspectives.

(4) Experiments have shown that our ProML implements a new state-of-the-art result in most SSDA problems.

2 Related Works

Unsupervised domain adaptation. Unsupervised domain adaptation (UDA) [Pan *et al.*, 2010] has made some exciting achievements. It aims to transfer the knowledge learned from the labeled source domain to the unlabeled target domain. The methods based on feature space alignment [Long *et al.*, 2017; Sun and Saenko, 2016] usually use joint distribution to make the two domains as close as possible in the feature space, to reduce the differences between the two domains. Some methods based on GAN [Chen *et al.*, 2019; Long *et al.*, 2018; Pei *et al.*, 2018] are also popular. They calibrate the distribution of the source domain and target domain by generating codes that cannot be distinguished from the perspective of the discriminator, which is trained to classify the target domain. Despite significant advances in UDA, UDA methods do not perform well in SSDA [Saito *et al.*, 2019], which is our main reason for focusing on SSDA.

Semi-supervised domain adaptation. In SSDA, the problem assumes that there are a few labeled samples in the target domain. It can be thought of as a combination of semi-supervised learning (SSL) and DA. Some methods [Saito *et al.*, 2019; Li *et al.*, 2021a] propose using Adversarial training to adjust source and target distributions, clustering similar target samples together through different clustering strategies, and separating different samples to reduce intra-domain gaps.

Some methods [Li *et al.*, 2021b; Singh, 2021] use the idea of self-supervised learning to build target prototypes and help model transformation across domains by comparing learning methods. They only build the prototype with very limited labels or high-noise pseudo-labels, ignoring the importance of different sample relationships. The methods based on SSL [Pérez-Carrasco *et al.*, 2022] use different losses to enhance the consistency between the feature representations of unlabeled samples and [Yan *et al.*, 2022] standardizes the consistency of different views of target domain samples at three levels, which facilitates learning more representative target features from each other. However, they all ignore the importance of making further use of the labeled target samples. In this work, we use the prototype to tap the potential of labeled target samples from multiple levels to help the model learn comprehensive and complementary feature representations.

3 Methodology

In this section, we first specify the definition and notation of SSDA and then introduce the proposed Prototype-based Multi-level Learning (ProML) framework from three levels.

3.1 Framework

In SSDA, we sample datasets from two different distributions. In this setting, we can access labeled source samples $\mathcal{D}_s = \{(x_{si}, y_{si})\}_{i=1}^{N_s}$ sampled from the source distribution $P_s(X, Y)$. We also have a limited number of labeled target samples $\mathcal{D}_t^l = \{(x_{ti}^l, y_{ti}^l)\}_{i=1}^{N_t^l}$ and a large number of unlabeled samples $\mathcal{D}_t^u = \{(x_{ti}^u)\}_{i=1}^{N_t^u}$ both from the target distribution $P_t(X, Y)$. This two distribution satisfy: $P_s(Y) = P_t(Y)$ and $P_s(X|Y) \neq P_t(X|Y)$. The framework is composed of a feature extractor G and a linear classifier F . An outline of our ProML framework is illustrated in Figure 2.

Following [Sohn *et al.*, 2020], we feed the labeled source samples and the labeled target samples into G to obtain their feature representations $f_s, f_t^l \in \mathbb{R}^d$, and then get the probability prediction $p_s, p_t^l \in \mathbb{R}^c$ through F . Similarly, we generate two different views, for each unlabeled target sample x_{ti}^u by weak and strong augment, represented as x_{ti}^{uw} and x_{ti}^{us} . Then the target samples of these two views are fed to the same feature extractor G to generate representations f_{ti}^{uw} and f_{ti}^{us} . Finally, the probability prediction p_{ti}^{uw}, p_{ti}^{us} are obtained through the same classifier F . We can calculate the standard cross-entropy loss by using labeled sample pairs from two domains and consider a simple pseudo-label regularization with different views of unlabeled samples. Specifically, we optimize the following baseline:

$$L_{\text{base}} = - \left(\sum_{x \in x_s \cup x_t^l} y \log p + \sum_{x \in x_t^u} \mathbb{1}_{\{p_t^{uw} \geq \tau_1\}} \log p_t^{us} \right), \quad (1)$$

where $\mathbb{1}$ is an indicator function, y is the label of labeled samples x , and τ_1 is the pseudo-label threshold of the regularization. In the following, we will further introduce our framework from three different levels.

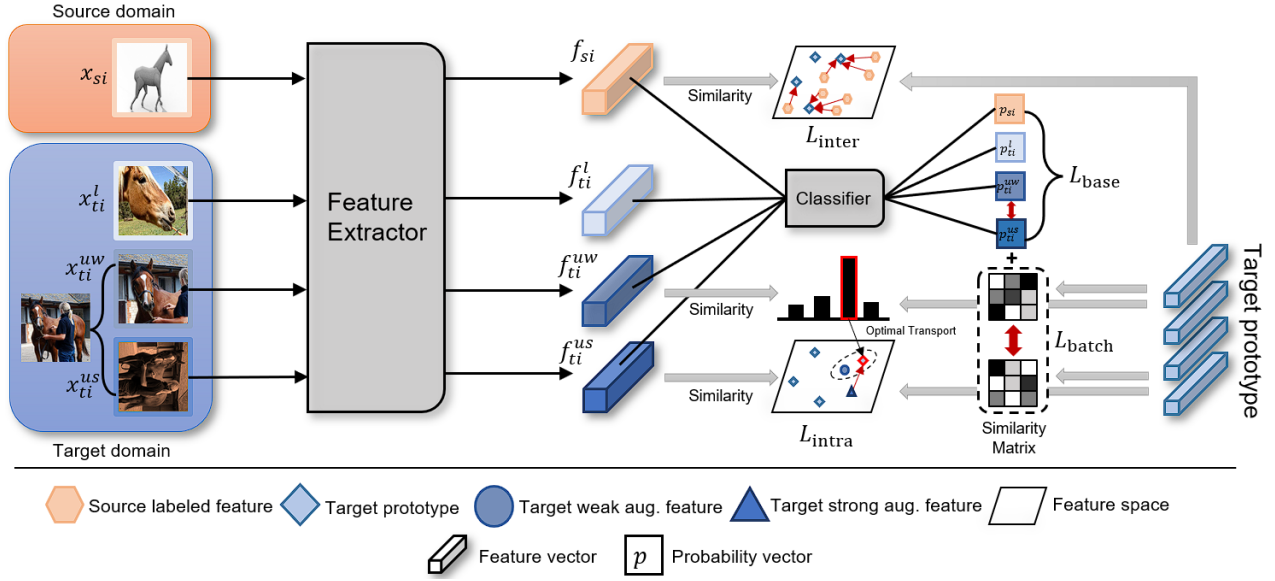


Figure 2: The structure of our ProML framework. First, the target samples are weakly and strongly augmented and then pass through the classifier together with the source samples to calculate the base loss. For the intra-domain level, the weakly augmented target samples generate pseudo-labels with the optimal transfer plan computed with the target prototype and compute the consistency loss with the strongly augmented samples. For the inter-domain level, the similarity loss between source samples and the target prototype of corresponding categories is computed to achieve cross-domain knowledge transfer. Finally, the dual consistency loss of the two augmented views in each mini-batch is considered from the perspective of linear and prototype-based classifiers.

3.2 Intra-domain Pseudo-label Aggregation

A compact target domain can help the model make better use of pseudo-labels and cross-domain knowledge transfer. We propose a novel pseudo-label aggregation strategy to align the unlabeled and labeled target data robustly and accurately.

To make full use of the labeled target samples, we initialize the target prototype:

$$\mathbf{c}_t^k = \frac{\sum_{i=1}^{N_t^l} \mathbb{1}_{\{y_{ti}^l=k\}} f_{ti}^l}{\sum_{i=1}^{N_t^l} \mathbb{1}_{\{y_{ti}^l=k\}}}, \quad (2)$$

where y_{ti}^l is the label of the i -th labeled target sample. The optimal transport (OT) is used to align the inter-domain level representation for domain adaptation [Courty *et al.*, 2017; Yan *et al.*, 2022]. It finds the optimal coupling plan γ_0 according to the given transport function to minimize the total transport cost. To make better use of the additional data provided by SSDA, we propose to apply OT to the intra-domain feature aggregation instead of the inter-domain. For labeled and weakly augmented unlabeled samples, we assume that:

$$\gamma_0 = \arg \min_{\gamma \in \mathcal{B}} \langle \gamma, \mathbf{C}^w \rangle_F, \quad (3)$$

$$\mathcal{B} = \left\{ \gamma \in (\mathbb{R}^+)^{N_s \times N_t^u} \mid \gamma \mathbf{1}_{N_t^u} = \mu_s, \gamma^\top \mathbf{1}_{N_s} = \mu_t \right\}, \quad (4)$$

where $\gamma_{0i,j}$ means the transport plan between the i -th labeled sample and the j -th unlabeled sample in the target domain, $\langle \cdot, \cdot \rangle_F$ represents Frobenius inner product, $\mathbf{1}_d$ is 1 's d -dimensional vector, $\mu_s \in \mathbb{R}^{N_s}$, $\mu_t \in \mathbb{R}^{N_t^u}$ are the empirical

distributions of labeled and unlabeled samples respectively, and the default is a uniform distribution. $\mathbf{C}^w \in \mathbb{R}^{N_s \times N_t^u}$ is a cost matrix for indicating each transportation. In consideration of the robustness and accuracy of the transportation, we propose to use the target prototype to replace the labeled samples in the transport plan. This not only makes the feature representation more representative but also alleviates the negative alignment caused by the lack of some classes in the minibatch of samples. The cost matrix is as follows:

$$\mathbf{C}_{i,j}^w = 1 - \mathbf{c}_t^{i\top} f_{tj}^{uw}, \quad (5)$$

where each element in the cost matrix $\mathbf{C}_{i,j}^w$ represents the degree of difference between the i -th target prototype \mathbf{c}_t^i and the j -th weak augmented target feature f_{tj}^{uw} . The resulting OT plan γ_0 helps the model better align the distribution of different samples in the target domain.

To further improve the robustness of the prototype, we add unlabeled target samples to the update of the prototype and regard the OT plan as a supplement to the pseudo-label strategy. Specifically, we consider the following pseudo-label strategy:

$$\tilde{y}_i^{uw} = \begin{cases} \arg \max(\tilde{p}_{ti}^{uw}), & \text{if } \tilde{p}_{ti}^{uw} \geq \tau_1 \\ \gamma_{0i}, & \text{if } \tilde{p}_{ti}^{uw} < \tau_1 \text{ and } \tilde{p}_{ti}^{uw} \geq \tau_2 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where τ_2 is the confidence threshold for using a pseudo-label supplementation strategy, \tilde{p}_{ti}^{uw} is the maximum prediction probability of the i -th unlabeled weak sample in all categories, and $\arg \max(\cdot)$ represents the category corresponding to the prediction probability. When the distribution of unlabeled data is close to that of labeled data, this module can

better use the prototype initialized by labeled data to assign pseudo-labels to unlabeled data through the proposed strategy. We make better use of labeled data to improve the robustness of prototypes through the pseudo-label strategy.

For each minibatch, we calculate the feature average of labeled and pseudo-labeled target samples for clustering, and then use the exponential moving average to update the target prototype during training:

$$\mathbf{c}_t^k = \alpha \mathbf{c}_t^k + (1 - \alpha) \tilde{\mathbf{c}}_t^k, \quad (7)$$

where $\tilde{\mathbf{c}}_t^k$ is the target prototype clustered by target samples with labels and pseudo-labels in this minibatch. Under the guidance of the optimal coupling γ_0 , the feature representation of each strongly augment view target sample can have a consistent mapping plan with the weak augment view, forming a consistency constraint:

$$L_{\text{intra}} = \langle \gamma_0, \mathbf{C}^s \rangle_F, \quad (8)$$

where \mathbf{C}^s is the similarity matrix between the strong augment target sample and the target prototype similar to Equation 5. As mentioned above, the model can effectively align the distribution of the target prototype and unlabeled samples and pay more attention to the details of the intra-domain level.

3.3 Inter-domain Alignment

Knowledge transfer is an essential capability of models in SSDA. For the cross-domain level, we can use the relationship between the target prototype and the source sample to naturally transfer the knowledge of the model in the source domain to the target domain.

Specifically, for each category of target prototype, we can calculate the softmax of the similarity of the corresponding target prototype in the feature space according to the category of the source sample:

$$\mathbf{s}_i^k = \frac{\exp(\text{sim}(f_{si}, \mathbf{c}_t^k) / T_1)}{\sum_{i=1}^{N_s} \exp(\text{sim}(f_{si}, \mathbf{c}_t^k) / T_1)}, \quad (9)$$

where $\text{sim}(\cdot, \cdot)$ means cosine similarity, f_{si} is the feature of i -th source samples, T_1 is a scale temperature, and N_s is the number of source samples. Then, we can calculate the cross-domain prototype alignment loss for each source sample:

$$L_{\text{inter}} = - \sum_{k=1}^C \sum_{i=1}^{N_s} \mathbb{1}(y_s = k) \log \mathbf{s}_i^k, \quad (10)$$

where C is the number of classes. Our prototype is not only composed of labeled targets but also adopts a more robust pseudo-label update strategy to alleviate negative transfer caused by the dispersion of target samples.

3.4 Batch-wise Dual Consistency

To make the model more comprehensively learn the representation in the target domain, different from [Yan *et al.*, 2022], we consider the dual relationship between target features at the batch level. We increase the confidence difference between different views by sharpening the confidence and then model the clustering of each class as the classification confidence and prototype similarity for all samples of

that class in the batch. Maintain the consistency of the strong and weak views of the class allocation as the positive pair in the same batch, and reduce the similarity between different classes as the negative pair, instead of the sample-wise contrastive learning.

Given the prediction $\mathbf{P}^{uw} \in \mathbb{R}^{N_t \times C}$ of the target sample of a batch, we use the sharpening function to reduce the entropy of label distribution, widening the gap between different confidence, and enhancing the contrast between different views:

$$\hat{p}_i^{uw} = \frac{p_i^{uw \frac{1}{T_2}}}{\sum_{j=1}^C p_j^{uw \frac{1}{T_2}}}, \quad (11)$$

where T_2 is a temperature hyperparameter, C is the number of classes. However, it is one-sided to optimize only from the perspective of linear classifiers. Due to the existence of the target prototype, we propose that batch level relationships be learned dually not only from the perspective of linear classifier but also from the perspective of prototype similarity. Since linear classifier can assign learnable parameters to each class, while prototype-based classifier only relies on excellent feature representation, we calculate the cross-correlation matrix of strong and weak views from the perspective of linear classifier and prototype-based classifier respectively:

$$\begin{aligned} \mathbf{R}_l^{ws} &= \hat{\mathbf{P}}^{uw\top} \hat{\mathbf{P}}^{us} \\ \mathbf{R}_p^{ws} &= \mathbf{S}^{uw\top} \mathbf{S}^{us}, \end{aligned} \quad (12)$$

where $\hat{\mathbf{P}}^{uw}$, $\hat{\mathbf{P}}^{us}$ are sharpening batch probability matrices and \mathbf{S}^{uw} , \mathbf{S}^{us} are batch similarity matrices with the prototype from weak and strong views similar to Equation 9. \mathbf{R}_l^{ws} , \mathbf{R}_p^{ws} are asymmetric matrices, and each element represents two similarities of different views at the batch level. From this, the dual contrast loss of batch class can be calculated:

$$\begin{aligned} L_{\text{batch}} &= \frac{1}{2C} \left(\underbrace{\|\phi(\mathbf{R}_l^{ws}) - \mathbf{I}\|_1 + \|\phi(\mathbf{R}_l^{ws\top}) - \mathbf{I}\|_1}_{\text{linear classifier}} + \right. \\ &\quad \left. \underbrace{\|\phi(\mathbf{R}_p^{ws}) - \mathbf{I}\|_1 + \|\phi(\mathbf{R}_p^{ws\top}) - \mathbf{I}\|_1}_{\text{prototype-based classifier}} \right), \end{aligned} \quad (13)$$

where $\phi(\cdot)$ is a normalized function that keeps the row total as 1. $\mathbf{I} \in \mathbb{R}^{C \times C}$ is the identity matrix, and $\|\cdot\|_1$ represents the sum of the absolute values of the matrix.

Dual consistency can benefit the framework in learning cross-domain knowledge from two perspectives. Linear classifiers can use learnable parameters to focus more on the distinguishing dimensions represented by features while suppressing unrelated feature dimensions by assigning higher or lower weights to different dimensions. In contrast, prototype-based classifiers cannot take advantage of this which requires more discriminative feature representation [Xu *et al.*, 2022].

At the same time, due to the setting of DA, the prediction of the linear classifier may be biased toward the source domain. Comparing with target prototype similarity can alleviate the inaccurate contrast caused by over-fitting the source domain. We combine the two to encourage the model to learn a more discriminative and accurate relationship between batch level objectives from different perspectives.

Method	R→C		R→P		P→C		C→S		S→P		R→S		P→R		Mean	
	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
S+T	55.6	60.0	60.6	62.2	56.8	59.4	50.8	55.0	56.0	59.5	46.3	50.1	71.8	73.9	56.9	60.0
DANN	58.2	59.8	61.4	62.8	56.3	59.6	52.8	55.4	57.4	59.9	52.2	54.9	70.3	72.2	58.4	60.7
ENT	65.2	71.0	65.9	69.2	65.4	71.1	54.6	60.0	59.7	62.1	52.1	61.1	75.0	78.6	62.6	67.6
MME	70.0	72.2	67.7	69.7	69.0	71.7	56.3	61.8	64.8	66.8	61.0	61.9	76.1	78.5	66.4	68.9
BiAT	73.0	74.9	68.0	68.8	71.6	74.6	57.9	61.5	63.9	67.5	58.5	62.1	77.0	78.6	67.1	69.7
APE	70.4	76.6	70.8	72.1	72.9	76.7	56.7	63.1	64.5	66.1	63.0	67.8	76.6	79.4	67.6	71.7
Con ² DA	71.3	74.2	71.8	72.1	71.1	75.0	60.0	65.7	63.5	67.1	65.2	67.1	75.7	78.6	68.4	71.4
CDAC	77.4	79.6	74.2	75.1	75.5	79.3	67.6	69.9	71.0	73.4	69.2	72.5	80.4	81.9	73.6	76.0
DECOTA	79.1	80.4	74.9	75.2	76.9	78.7	65.1	68.6	72.0	72.7	69.7	71.9	79.6	81.5	73.9	75.6
CLDA	76.1	77.7	75.1	75.7	71.0	76.4	63.7	69.7	70.2	73.7	67.1	71.1	80.1	82.9	71.9	75.3
ECACL	75.3	79.0	74.1	77.3	75.3	79.4	65.0	70.6	72.1	74.6	68.1	71.6	79.7	82.4	72.8	76.4
MCL	77.4	79.4	74.6	76.3	75.5	78.8	66.4	70.9	74.0	74.7	70.7	72.3	82.0	83.3	74.4	76.5
ProML	78.5	80.2	75.4	76.5	77.8	78.9	70.2	72.0	74.1	75.4	72.4	73.5	84.0	84.8	76.1	77.4

Table 1: Accuracy (%) on *DomainNet* under the settings of 1-shot and 3-shot using ResNet34 as backbone networks.

Method	1-shot	3-shot
S+T	60.2	64.6
ENT	63.6	72.7
MME	68.7	70.9
APE	78.9	81.0
CDAC	69.9	80.6
DECOTA	64.9	80.7
ECACL	81.1	83.3
MCL	86.3	87.3
ProML	87.6	88.4

Table 2: Mean Class-wise Accuracy (MCA)(%) on *VisDA2017* using ResNet34 as the backbone network.

3.5 Overall Framework and Training Objective

To sum up, the overall training objectives of the framework are as follows:

$$L_{\text{all}} = L_{\text{base}} + \lambda_{\text{intra}}L_{\text{intra}} + \lambda_{\text{inter}}L_{\text{inter}} + \lambda_{\text{batch}}L_{\text{batch}}, \quad (14)$$

where λ_{intra} , λ_{inter} and λ_{batch} are the hyper-parameters that balance different levels. We train the model in our framework by employing the overall training loss described in Equation 14.

4 Experiments

4.1 Experimental Setup

Datasets. We verified our ProML framework on three popular SSDA datasets. **DomainNet** is the latest large-scale multi-source domain adaptive dataset, with 6 domains and 345 categories [Peng *et al.*, 2019]. According to [Saito *et al.*, 2019], four fields (Real, Clipart, Painting, Sketch) and 126 categories were selected for SSDA evaluation. **VisDA2017** consists of 150k synthetic images and 55k real images, including two domains and 12 categories [Peng *et al.*, 2017]. For each category of each dataset, we randomly select one or three labeled samples (1-shot or 3-shot) as labeled target samples. **Office-Home** is also a mainstream domain adaptive dataset, including 4 domains (Real, Clipart, Art, Product) and 65 classes [Venkateswara *et al.*, 2017]. We follow the standard of most SSDA work [Saito *et al.*, 2019], and report the

overall accuracy as an indicator of *DomainNet* and *Office-Home*, and report the average class accuracy (MCA) as an evaluation indicator of *VisDA2017*.

Implementation details.¹ To ensure the fairness of the experiment, similar to [Saito *et al.*, 2019; Li *et al.*, 2021a], we use ResNet34 [He *et al.*, 2016] pre-trained on Imagenet as the backbone of the model. The settings of batch size, optimizer, feature size, and learning rate are also consistent with [Saito *et al.*, 2019]. Similar to [Yan *et al.*, 2022], RandomFlip and RandomCrop are used as weak image augment methods, and RandAugment [Cubuk *et al.*, 2020] is used as strong augment methods. The momentum α used to update prototypes is set to 0.9, and the threshold τ_1 is set to 0.95 and τ_2 is set to 0.4 for *VisDA2017*, 0.3 for *DomainNet* and 0.1 for *Office-Home*. For the OT solver, we solve it through Sinkhorn-Knopp iteration [Fratras *et al.*, 2021]. The temperature T_1 and T_2 in the similarity function are set to 0.05 and 0.1. Balance hyperparameter λ_{intra} and λ_{inter} is set to 1. λ_{batch} is set to 1 for *DomainNet*, 1 for *Office-Home*, and 0.1 for *VisDA2017*.

4.2 Analysis of Experimental Results

In the experiments, we compare our ProML with two baselines: S+T uses only source and labeled target data, and ENT [Grandvalet and Bengio, 2004] uses entropy minimization for unlabeled samples; and several popular SSDA methods, i.e., DANN [Ganin *et al.*, 2016], MME [Saito *et al.*, 2019], BiAT [Jiang *et al.*, 2020], APE [Kim and Kim, 2020], Con²DA [Pérez-Carrasco *et al.*, 2022], CDAC [Li *et al.*, 2021a], DECOTA [Yang *et al.*, 2021], CLDA [Singh, 2021], ECACL [Li *et al.*, 2021b] and MCL [Yan *et al.*, 2022].

DomainNet. As shown in Tab. 1, our proposed ProML achieves 76.1% and 77.4% average accuracy and SOTA performance in 7 scenarios of 1-shot and 3-shot, respectively.

VisDA2017. *VisDA2017* demonstrates the validity of ProML in Tab. 2, our ProML achieves 87.6% MCA in 1-shot and 88.4% in 3-shot, which are superior to the SOTA methods.

Office-Home. As shown in Tab. 3, ProML outperforms the existing SOTA methods in both 1-shot and 3-shot scenarios, with accuracy reaching 74.6% and 77.8%, respectively.

¹<https://bupt-ai-cz.github.io/ProML/>

Method	R→C	R→P	R→A	P→R	P→C	P→A	A→P	A→C	A→R	C→R	C→A	C→P	Mean
1-shot													
S+T	52.1	78.6	66.2	74.4	48.3	57.2	69.8	50.9	73.8	70.0	56.3	68.1	63.8
DANN	53.1	74.8	64.5	68.4	51.9	55.7	67.9	52.3	73.9	69.2	54.1	66.8	62.7
ENT	53.6	81.9	70.4	79.9	51.9	63.0	75.0	52.9	76.7	73.2	63.2	73.6	67.9
MME	61.9	82.8	71.2	79.2	57.4	64.7	75.5	59.6	77.8	74.8	65.7	74.5	70.4
APE	60.7	81.6	72.5	78.6	58.3	63.6	76.1	53.9	75.2	72.3	63.6	69.8	68.9
CDAC	61.9	83.1	72.7	80.0	59.3	64.6	75.9	61.2	78.5	75.3	64.5	75.1	71.0
DECOTA	56.0	79.4	71.3	76.9	48.8	60.0	68.5	42.1	72.6	70.7	60.3	70.4	64.8
MCL	67.0	85.5	73.8	81.3	61.1	68.0	79.5	64.2	81.2	78.4	68.5	79.3	74.0
ProML	67.5	86.1	73.7	81.9	61.4	69.3	79.7	64.5	81.7	79.0	69.1	80.5	74.6
3-shot													
S+T	55.7	80.8	67.8	73.1	53.8	63.5	73.1	54.0	74.2	68.3	57.6	72.3	66.2
DANN	57.3	75.5	65.2	69.2	51.8	56.6	68.3	54.7	73.8	67.1	55.1	67.5	63.5
ENT	62.6	85.7	70.2	79.9	60.5	63.9	79.5	61.3	79.1	76.4	64.7	79.1	71.9
MME	64.6	85.5	71.3	80.1	64.6	65.5	79.0	63.6	79.7	76.6	67.2	79.3	73.1
APE	66.4	86.2	73.4	82.0	65.2	66.1	81.1	63.9	80.2	76.8	66.6	79.9	74.0
CDAC	67.8	85.6	72.2	81.9	67.0	67.5	80.3	65.9	80.6	80.2	67.4	81.4	74.2
CLDA	66.0	87.6	76.7	82.2	63.9	72.4	81.4	63.4	81.3	80.3	70.5	80.9	75.5
DECOTA	70.4	87.7	74.0	82.1	68.0	69.9	81.8	64.0	80.5	79.0	68.0	83.2	75.7
MCL	70.1	88.1	75.3	83.0	68.0	69.9	83.9	67.5	82.4	81.6	71.4	84.3	77.1
ProML	71.0	88.6	75.8	83.8	68.9	72.5	83.9	67.8	82.2	82.3	72.1	84.1	77.8

Table 3: Accuracy (%) on *Office-Home* under the settings of 1-shot and 3-shot using ResNet34 as the backbone network.

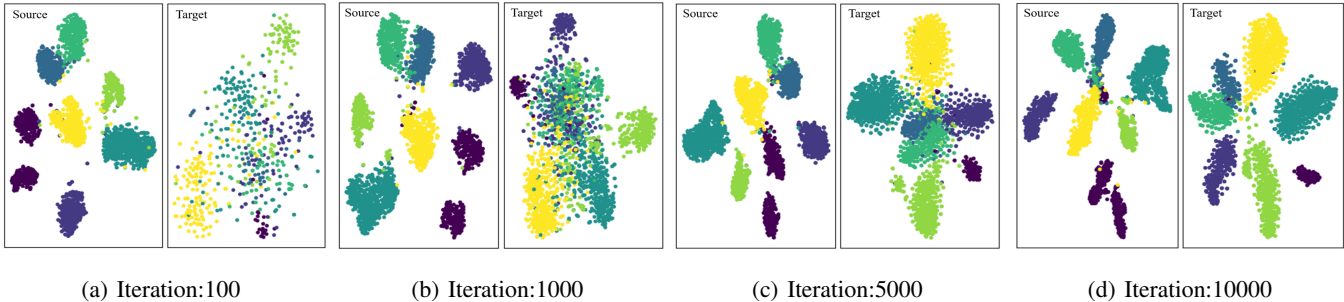


Figure 3: The t-SNE visualization of the feature alignment progress through our ProML during the training time on *DomainNet* P→R.

	L_{intra}	L_{inter}	L_{batch}	1-shot	3-shot
①	✗	✗	✗	74.1	78.1
②	✓	✗	✗	77.6	82.9
③	✗	✓	✗	76.2	81.4
④	✗	✗	✓	75.1	80.0
⑤	✗	✓	✓	83.1	84.6
⑥	✓	✗	✓	82.6	85.2
⑦	✓	✓	✗	83.9	86.0
⑧	✓	✓	✓	87.6	88.4

Table 4: Ablation studies of ProML’s different components. We report the MCA (%) on *VisDA2017* under the settings of 1-shot and 3-shot using a ResNet34 backbone.

4.3 Additional Analysis

Ablation studies. We performed ablation experiments on *VisDA2017* at the settings of 1-shot and 3-shot, as shown in Tab. 4. Row ② ③ ④ show that each component can be signif-

	linear pred.	proto. pred.	update proto.	1-shot	3-shot
①	✓	✗	✗	84.6	85.8
②	✗	✓	✓	80.7	84.2
③	✓	✗	✓	85.8	86.7
④	✓	✓	✗	82.0	85.1
⑤	✓	✓	✓	87.6	88.4

Table 5: Ablation study on the effectiveness of prototype-based contrast and update of our ProML on *VisDA2017* under the settings of 1-shot and 3-shot using a ResNet34 backbone.

icantly improved. Row ⑤ ⑥ ⑦ show that each combination still improves performance, indicating the universality of the proposed module. Due to the integrity of our framework, one aspect of the lack of consideration may not be optimal in performance, but the best performance can be achieved when all components are activated.

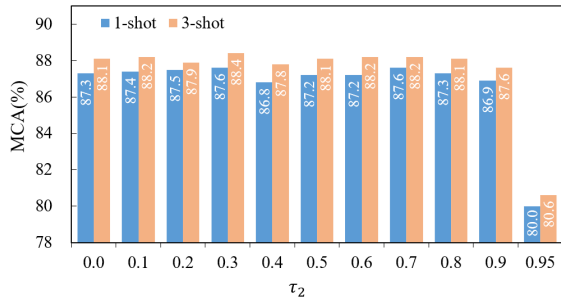


Figure 4: Mean Class Accuracy (MCA)(%) under different pseudo-label confidence thresholds on *VisDA2017* for 1-shot and 3-shot, where $\tau_2 = 0.95$ means no intra-domain OT.

Method	Epoch 100		Epoch 1000		Epoch 10000		Mean	
	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
Linear	23.4	36.0	53.1	67.2	65.6	76.5	65.6	75.6
Proto.	18.8	37.5	56.3	64.0	65.6	74.9	65.5	74.1
Our	23.4	37.5	57.3	68.1	68.1	78.1	66.9	76.8

Table 6: The label accuracy (%) of different pseudo label methods under the 1-shot and 3-shot settings of *DomainNet C→S*. *Linear* represents the linear classifier probability, *Proto.* represents the prototype similarity, and *Our* represents our intra-domain OT strategy.

Prototype-based contrast and update. Tab. 5 shows ablation studies of prototype-based classifier and updated procedure. Row ① shows only linear prediction, and its performance degrades a lot, which proves that the knowledge learned only by using a linear classifier is limited. The comparison with row ③ proves that the model will benefit from the dynamic update of the prototype. The global static prototype that only uses labeled data is not robust, so its generalization ability is limited. Row ② shows only prototype-based predictions and dynamic updates during training. Its performance is much lower, which means that without the help of a linear classifier, the prototype-based classifier cannot learn consistency under limited label data settings, so it cannot make full use of the additional knowledge brought by prototypes. Row ④ only adds the prototype-based classifier on row ①, but its performance does not increase but decreases, which indicates that biased prototypes will bring negative knowledge transfer. Row ⑤ achieves the optimal performance after integrating all methods, which shows that our method can complement knowledge from multi-level.

Intra-domain OT pseudo-labels. Tab. 6 shows the pseudo-label accuracy of different methods. By making the target domain more compact, our strategy gives more accurate pseudo-labels at different training stages than simple methods only using maximum linear classifier prediction (*Linear*) or maximum prototype similarity (*Proto.*) under the same confidence threshold τ_2 in Equation 6.

Convergence analysis. To further analyze the convergence of our method, we describe the feature t-SNE [Van der Maaten and Hinton, 2008] visualization of source and target domains during different training times in Figure 3. We randomly selected seven categories of the feature on *Domain-*

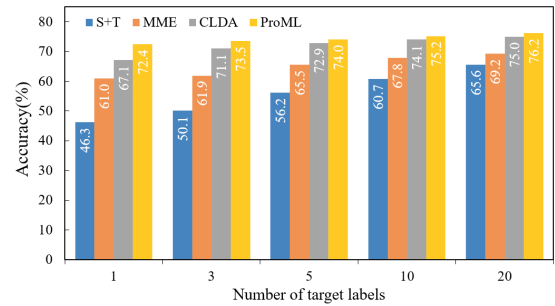


Figure 5: Accuracy with different numbers of labeled samples per class in target domain on *DomainNet R→S*.

Net R→S for clearer visualization. Figure 3(a) clearly shows the initial domain difference between the source and the target domain. Early feature descriptions often show many misaligned source and target clusters, so the model initially performs poorly in the target domain. As the training progresses, it can be seen from Figure 3(b) and Figure 3(c) that our method converges and aggregates the feature of the target domain. Accumulated good target feature is obtained and it proves that our method can obtain a compact target domain distribution, as shown in Figure 3(d).

Parameter sensitivity. We will use the pseudo-labels threshold τ_2 of intra-domain OT strategy in Equation 6 as a variable, analyze changes in MCA with *VisDA2017* under the settings of 1-shot and 3-shot, and experimental results support our argument. As shown in Figure 4, with the change of the threshold, MCA floats within 1%, which demonstrates the robustness of the strategy. When this strategy is not used, i.e. $\tau_2 = 0.95$, MCA decreased significantly. This result is reasonable and shows that this strategy can help the model make use of unlabeled target samples.

Impact of the number of target labels. We studied the performance impact of different target label numbers on *Domain R→S*. From Figure 5, all methods can improve performance by labeling more target samples. By contrast, our ProML always achieves optimal performance in all cases. This confirms that our method can make use of a flexible number of labeled target samples to help model knowledge transfer.

5 Conclusion

In this paper, we propose a novel prototype-based multi-level learning (ProML) framework to make better use of target samples. ProML leverages prototypes constructed from target samples at three levels, (i) intra-domain level, aligns labeled and unlabeled sample distributions within the target domain using pseudo-label aggregation help models based on optimal transport, (ii) inter-domain level, it aligns source and target domain, and (iii) batch level, it learns compact classes and partition clusters from two dual perspectives. Extensive experimental studies have demonstrated its advantages.

Acknowledgements

This work was supported by the National Key R&D Program of China (2021ZD0109802), by the National Natural Science Foundation of China (81972248).

References

- [Chen *et al.*, 2019] Chao Chen, Zhihong Chen, Boyuan Jiang, and Xinyu Jin. Joint domain alignment and discriminative feature learning for unsupervised deep domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3296–3303, 2019.
- [Courty *et al.*, 2017] Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. *Advances in Neural Information Processing Systems*, 30, 2017.
- [Cubuk *et al.*, 2020] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020.
- [FAtlas *et al.*, 2021] Kilian Atlas, Thibault Séjourné, Rémi Flamary, and Nicolas Courty. Unbalanced minibatch optimal transport; applications to domain adaptation. In *International Conference on Machine Learning*, pages 3186–3197. PMLR, 2021.
- [Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.
- [Grandvalet and Bengio, 2004] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems*, 17, 2004.
- [He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [Jiang *et al.*, 2020] Pin Jiang, Aming Wu, Yahong Han, Yunfeng Shao, Meiyu Qi, and Bingshuai Li. Bidirectional adversarial training for semi-supervised domain adaptation. In *IJCAI*, pages 934–940, 2020.
- [Kim and Kim, 2020] Taekyung Kim and Changick Kim. Attract, perturb, and explore: Learning a feature alignment network for semi-supervised domain adaptation. In *European conference on computer vision*, pages 591–607. Springer, 2020.
- [Li *et al.*, 2021a] Jichang Li, Guanbin Li, Yemin Shi, and Yizhou Yu. Cross-domain adaptive clustering for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2505–2514, 2021.
- [Li *et al.*, 2021b] Kai Li, Chang Liu, Handong Zhao, Yulun Zhang, and Yun Fu. Ecacl: A holistic framework for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8578–8587, 2021.
- [Long *et al.*, 2017] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017.
- [Long *et al.*, 2018] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31, 2018.
- [Pan *et al.*, 2010] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE transactions on neural networks*, 22(2):199–210, 2010.
- [Pei *et al.*, 2018] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. In *Thirty-second AAAI conference on artificial intelligence*, 2018.
- [Peng *et al.*, 2017] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge, 2017.
- [Peng *et al.*, 2019] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1406–1415, 2019.
- [Pérez-Carrasco *et al.*, 2022] Manuel Pérez-Carrasco, Pavlos Protopapas, and Guillermo Cabrera-Vives. Con²da: Simplifying semi-supervised domain adaptation by learning consistent and contrastive feature representations. *arXiv preprint arXiv:2204.01558*, 2022.
- [Saito *et al.*, 2019] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. Semi-supervised domain adaptation via minimax entropy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8050–8058, 2019.
- [Singh, 2021] Ankit Singh. Clda: Contrastive learning for semi-supervised domain adaptation. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 5089–5101. Curran Associates, Inc., 2021.
- [Sohn *et al.*, 2020] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020.
- [Sun and Saenko, 2016] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pages 443–450. Springer, 2016.
- [Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

- [Venkateswara *et al.*, 2017] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5018–5027, 2017.
- [Xu *et al.*, 2022] Hai-Ming Xu, Lingqiao Liu, Qiuchen Bian, and Zhen Yang. Semi-supervised semantic segmentation with prototype-based consistency regularization. *Advances in neural information processing systems*, 2022.
- [Yan *et al.*, 2022] Zizheng Yan, Yushuang Wu, Guanbin Li, Yipeng Qin, Xiaoguang Han, and Shuguang Cui. Multi-level consistency learning for semi-supervised domain adaptation. *arXiv preprint arXiv:2205.04066*, 2022.
- [Yang *et al.*, 2021] Luyu Yang, Yan Wang, Mingfei Gao, Abhinav Shrivastava, Kilian Q Weinberger, Wei-Lun Chao, and Ser-Nam Lim. Deep co-training with task decomposition for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8906–8916, 2021.