AUTONOMOUS SOURCE KNOWLEDGE SELECTION IN MULTI-DOMAIN ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Unsupervised multi-domain adaptation plays a key role in transfer learning by leveraging acquired rich source information from multiple source domains to solve target task from an unlabeled target domain. However, multiple source domains often contain much redundant or unrelated information which can harm transfer performance, especially when in massive-source domain settings. It is urgent to develop effective strategies for identifying and selecting the most transferable knowledge from massive source domains to address the target task. In this paper, we propose a multi-domain adaptation method named *Auto*nomous Source Knowledge *S*election (AutoS) to autonomosly select source training samples and models, enabling the prediction of target task using more relevant and transferable source information. The proposed method employs a density-driven selection strategy to choose source samples during training and to determine which source models should contribute to target prediction. Simulteneously, a pseudo-label enhancement module built on a pre-trained multimodal modal is employed to mitigate target label noise and improve self-supervision. Experiments on real-world datasets indicate the superiority of the proposed method.

1 INTRODUCTION

Domain adaptation has achieved significant progress in transfer learning by addressing data scarcity in the target domain through the utilization of knowledge from source domain(s). Typical domain adaptation methods involve various techniques to overcome distribution shifts and modal gaps, including approaches based on feature alignment (Bai et al., 2024) and those relying on model fine-tuning (Zhang et al., 2023b). Feature alignment–based methods focus on matching source and target data in a latent feature space to bridge their gaps, commonly employing techniques such as maximum mean discrepancy (MMD) (Ning et al., 2025), Kullback–Leibler (KL) divergence (Schlachter et al., 2025), Wasserstein distance (He et al., 2024), and adversarial learning (Fang et al., 2024). In contrast, model fine-tuning methods (Zhang et al., 2023c; Li et al., 2024) primarily rely on pseudo-label estimation and self-training strategies to adapt the model, making label denoising techniques a critical component for ensuring performance (Litrico et al., 2023).

To enhance the generality of domain adaptation in handling complex scenarios, category shifts have been considered, leading to the development of partial (Kong et al., 2022), open-set (Wan et al., 2024), and universal (Qu et al., 2024) domain adaptation, where out-of-distribution detection techniques serve as an effective solution. Simultaneously, extracting transferable knowledge from multiple source domains (Ma et al., 2024) and modalities (Zhang et al., 2023a; 2025) has emerged as a prominent approach for leveraging richer information to enhance transfer performance. To integrate knowledge, linear combination is a widely used approach for fusing features or predictions from multiple source domains or modalities, including both simple averaging (Zhao et al., 2020), weighted averaging (Dong et al., 2021) and federated learning (Huang et al., 2023) strategies.

However, knowledge from different sources also introduce challenges, particularly when dealing with massive numbers of samples or domains, as shown in Figure 1. Previous multi-domain adaptation methods typically combine information from all source domains, but seldom focus on selecting one or more relevant domains. Limited research has explored reducing the number of transferred source domains, and such distillation processes often involve manual operations. One represent work is sample and source distillation, (Li et al., 2023). It selects source samples and domains

that are more similar to the target domain to enhance transfer performance. However, this method requires training source models twice, once for selection and once for adaptation, which is impractical when dealing with massive source domains due to the high computational cost. Automatically and effectively selecting source knowledge from massive domains remains an urgent challenge in multi-domain adaptation.

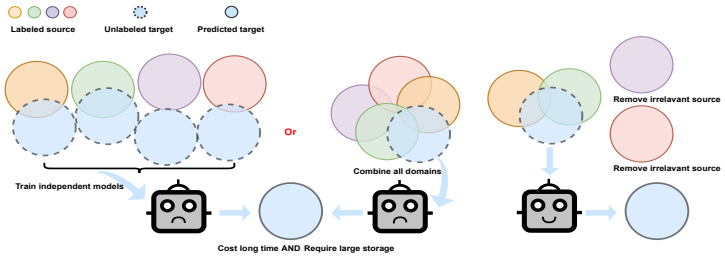


Figure 1: Illustration of multi-domain adaptation. The left figure exemplifies that training independent models by matching each source-target pair or by combining all domains can cause high computational and storage costs, while overlooking the fact that irrelevant sources can degrade transfer performance. The right figure depicts an ideal solution in which irrelevant source domains are removed while relevant ones are retained.

Considering the mentioned gaps, in this paper, we propose an autonomous source knowledge selection method to achieve target-oriented domain adaptation, which reduces the number of source domains involved in transfer without degrading performance. The proposed method introduces a density-controlled sample collection strategy to construct an intermediate domain composed of high-confidence source and target samples. Simultaneously, guided by the proportion of confident target samples and the target data density relative to each source domain, source domains with low similarity are removed during training and subsequent adaptation. In this way, multi-domain adaptation can be achieved by automatically identifying useful source domains while progressively discarding irrelevant domains during training, thereby avoiding the introduction of noisy knowledge and facilitating learning, particularly when dealing with massive source domains.

Our contributions can be summarized as: (1) An autonomous source knowledge selection method for multi-domain adaptation. This benefits the target domain by pruning redundant or dissimilar source domains and suppressing noisy information. (2) A target-oriented multi-domain adaptation approach for enhancing downstream task performance through prompt tuning across different modalities. This enables domain adaptation without data matching, making it flexible for transfer learning with or without access to source data. (3) A density-controlled sample collection strategy for gathering high-confident source and target samples. This benefits domain adaptation by constructing an intermediate domain and introducing self-supervision for the target task.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 details the proposed autonomous source knowledge selection method; Section 4 presents experiments and analysis on real-world datasets; and Section 5 concludes with directions for future research.

2 Related Work

2.1 Multi-Domain Adaptation

Multi-domain adaptation achieve great progress in recent. Contrastive adversarial learning (Wilson et al., 2023) deals with multi-source time series domain adaptation by aligning cross-source label information. In this framework, adversarial learning guides a domain classifier to distinguish whether a sample is from the source or target domain, while contrastive learning enforces intra-class closer together and inter-class separation. Prototype-based mean teacher (Belal et al., 2024) employs class prototypes to encode specific information from multiple domains, where the contrastive loss is used to align intra-class knowledge while separating inter-class knowledge across domains. Multiple adaptation network (Lu et al., 2025) addresses multi-source and multi-target domain adaptation,

which uses multiple alignment strategies to align both features and classifiers that are relevant. Style information transfer is also considered to fully leverage knowledge from multiple target domains.

2.2 Federated Learning

Federated learning is a widely used strategy that avoids training on independent source domains while leveraging the benefits of multiple sources. Universal federated domain adaptation (Liu et al., 2024) defines a hot-learning strategy with contrastive label disambiguation, which solves category shifts through one-hot outputs of source models and detects unknown categories using a cluster-level strategy built based on the consensus knowledge across source and target domains. Heterogeneous fuzzy domain adaptation (Li et al., 2025) integrates federated training and fuzzy logic to enable domain adaptation without requiring access to source data or models. A transformation module is introduced to address heterogeneity, while self-knowledge distillation is employed to federally construct the target model by simulating the predictions of source models.

2.3 Multimodal Domain Adaptation

Distilling multimodal foundation model (Tang et al., 2024b) achieves source-free domain adaptation through a two-step process, including customizing the vision–language model via prompt learning to minimize mutual information and distilling knowledge based on the target domain. Text-image alignment network (Kondapaneni et al., 2024) extends stable diffusion to leverage perceptual knowledge for predicting visual tasks from text-based generative prompts. It employs model personalization and caption modification to adapt the pre-trained model to the target domain, achieving improvements over unaligned baselines. Text-free graph foundation model (Yu et al., 2025) introduces a novel structure alignment framework to learn multi-domain knowledge from graphs originating in multiple source domains. It adapts to unseen target domains by incorporating a set of structure tokens and dual prompts, thereby unifying domain-specific information with structural knowledge.

3 The Proposed Autonomous Source Knowledge Selection Method

3.1 Overview

The whole proposed method is displayed in Figure 2. The framework consists of two stages. The first stage, illustrated in the upper figure, focuses on source model training with target-driven autonomous source knowledge selection. In this phase, both source and target data are transformed into a shared latent feature space, where the similarity between each source and the target domain is evaluated based on data density. High-confidence samples are retained while irrelevant or low-similarity samples are discarded, and source domains identified as irrelevant are progressively removed during further training. The proposed density-controlled selection is designed not only to reduce redundancy but also to suppress noisy knowledge, thereby ensuring that the training process exploits the most transferable information. Notably, all source models are initialized from a single shared model and updated in a federated manner, rather than being trained independently. This design avoids maintaining independent models for each source domain, thereby reducing memory requirements and computational costs. The federated model is then employed as the target model for adaptation. The second stage, illustrated in the lower figure, involves target model training with cross-modality transfer. In this phase, knowledge selected from the previous stage is combined with a pseudo-label enhancement module and cross-modal prompts to mitigate label noise, enabling the target model to adapt without direct reliance on the full set of source data or models.

3.2 Source Model Training

Denote multiple source domains as $\{\mathcal{D}_k^s = \{\boldsymbol{x}_{ki}^s, \boldsymbol{y}_{ki}^s\}_{i=1}^{n_k}\}_{k=1}^K$, and the target domain as $\mathcal{D}^t = \{\boldsymbol{x}_i^t\}_{i=1}^n$ To handle multiple domains without training independent source models, we first initialize a global model composed of a feature extraction module Φ and a decision layer P. For each source domain, source samples are fed into Φ to extract features in space $\mathbb{R}^d$, which are then passed to P to obtain predictions in $\mathbb{R}^{\mathcal{C}}$. By minimizing the errors between the ground truth labels and the

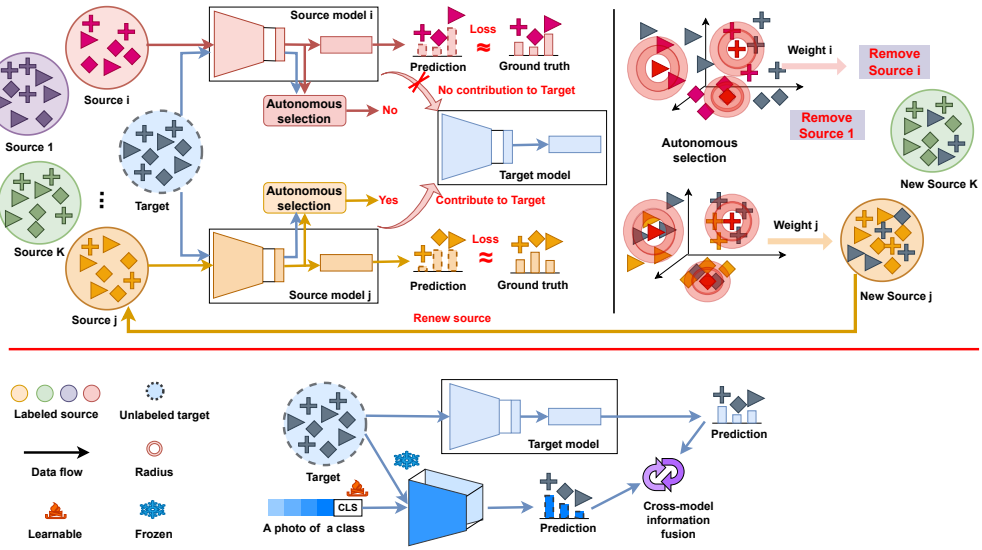


Figure 2: Whole framework of the proposed autonomous source knowledge selection. The upper figure indicates source model training with target-driven autonomous source knowledge selection, the lower figure displays target model training with cross-modality transfer.

predictions, the model is updated by:

$$\Phi_k, P_k = \underset{\substack{\Phi,P \\ (\boldsymbol{x}_k^s, \tilde{\boldsymbol{y}}_k^s)\in\mathcal{D}_k^s}}{\arg\min}\ \mathcal{L}(P(\Phi(\boldsymbol{x}_k^s)), \mathcal{LS}(\boldsymbol{y}_k^s)),$$
$$\mathcal{L} = -\frac{1}{n_k}\sum_{i=1}^{n_k}\mathcal{LS}(\boldsymbol{y}_{ki}^s)\log(P(\Phi(\boldsymbol{x}_{ki}^s))), \tag{1}$$
$$\mathcal{LS}(\boldsymbol{y}_k^s) = (1-\mu)\boldsymbol{y}_k^s + \frac{\mu}{|\mathcal{C}|}, \tilde{\boldsymbol{y}}_k^s, \boldsymbol{y}_k^s \in \mathbb{R}^{\mathcal{C}}.$$

$\mathcal{LS}$ is the label smoothing operation to accelerate training speed.

3.3 Autonomous Source Knowledge Selection

To select more relevant information, we define a density-controlled selection strategy to autonomously collects highly similar source and target samples. Following previous findings (Wang et al., 2022) that cluster centers are close to the mean values of normalized classifier weight vectors, we first collect cluster centers as:

$$\boldsymbol{f}_{kc}^s = \text{Norm}(P_k)c, c = 1, \cdots, \mathcal{C}; k = 1, \cdots, K. \tag{2}$$

The target clustering label is:

$$\boldsymbol{y}_k^t = \underset{c}{\arg\min}(\text{Dis}(\Phi\boldsymbol{x}^t, \boldsymbol{f}_{kc}^s)), c = 1, \cdots, \mathcal{C}; k = 1, \cdots, K. \tag{3}$$

The radius of each source cluster and the radius and density of target data corresponding to source cluster centers are then calculated as:

$$\boldsymbol{r}_{kc}^s = \text{Rd}(\text{Dis}(\Phi(\boldsymbol{x}_k^s), \boldsymbol{f}_{kc}^s)\mathbb{I}_{\boldsymbol{y}_k^s=c}), \boldsymbol{r}_{kc}^t = \text{Rd}(\text{Dis}(\Phi(\boldsymbol{x}^t), \boldsymbol{f}_{kc}^s)\mathbb{I}_{\boldsymbol{y}_k^t=c}), \tag{4}$$

Dis denotes the operation of computing the distance between source samples in the same cluster and their cluster center. Rd means the operation for calculating cluster radius using different distance metrics, such as average distance, root mean square distance and maximum radius. Unless otherwise specified, this work employs the average radius computed using cosine distance. Furthermore, thresholds for selecting highly confident source and target samples are defined as:

$$d_{kc}^s = \alpha\boldsymbol{r}_{kc}^s + s_{adj}, d_{kc}^t = \alpha\boldsymbol{r}_{kc}^s - t_{adj}, c = 1, \cdots, \mathcal{C}; k = 1, \cdots, K. \tag{5}$$

α, s_{adj} and t_{adj} are the parameters that control the radius used to assign values separating similar and dissimilar samples under different distance metrics. For any source and target sample, if $\text{Dis}(\Phi(\boldsymbol{x}_k^s), \boldsymbol{f}_{kc}^s)_{\mathbb{I}_{\boldsymbol{y}_k^s=c}} < d_{kc}^s$, we define the source sample as a confident sample, if $\text{Dis}(\Phi(\boldsymbol{x}^t), \boldsymbol{f}_{kc}^s)_{\arg\min_{\boldsymbol{x}^t}(\text{Dis}(\boldsymbol{x}^t, \boldsymbol{f}_{kc}^s))} < d_{kc}^t$, we define the target sample as a confident sample. Denote the selected source samples as $\mathcal{U}^s = \{\boldsymbol{x}_{ki}^s, \boldsymbol{y}_{ki}^s\}_{i=1}^{n_k^s}$ and target samples as $\mathcal{U}^t = \{\boldsymbol{x}_i^t, \boldsymbol{y}_{ki}^t\}_{i=1}^{n_k^t}$, the selected high-confidence samples are grouped together to update the original source domain, retaining confident source samples, incorporating confident target samples, and discarding low-confidence source samples.

To further define if a source domain can be removed during further training, target data density can be expressed as:

$$\rho_{kc}^t = \frac{n_{kc}^t}{\frac{\pi^{\mathbb{R}^{d/2}}}{\Gamma(\mathbb{R}^{d/2}+1)} \boldsymbol{r}_{kc}^t}, c = 1, \cdots, \mathcal{C}; k = 1, \cdots, K. \tag{6}$$

Then, two weights are defined as:

$$\begin{aligned} \omega_{k1} = \frac{n_k^t}{n}, \omega_{k2} = \frac{1}{\mathcal{C}} \sum_{c=1}^{\mathcal{C}} (1 - \frac{1}{1 + e^{\rho_{kc}^t}}), \\ \omega_k = \lambda \omega_{k1} + (1 - \lambda) \omega_{k2}, c = 1, \cdots, \mathcal{C}; k = 1, \cdots, K. \end{aligned} \tag{7}$$

The rules for removing irrelevant source domains are defined as:

$$\begin{aligned} \text{Keep}(\mathcal{D}_k^s) = \begin{cases} TRUE, \text{ if } \omega_k >= \frac{1}{K} - \sigma; \\ FALSE, \text{ if } \omega_k < \frac{1}{K} - \sigma; \end{cases} \\ c = 1, \cdots, \mathcal{C}; k = 1, \cdots, K. \end{aligned} \tag{8}$$

Then we can get renewed source domains $\{\mathcal{D}'^s_k = \{\boldsymbol{x}_{ki}^s, \boldsymbol{x}_j^t, \boldsymbol{y}_{ki}^s, \boldsymbol{y}_{kj}^t\}_{i,j=1}^{n_k^s, n_k^t}\}_{k=1}^{K'}, \{\boldsymbol{x}_k^s, \boldsymbol{y}_k^s\} \in \mathcal{U}^s, \{\boldsymbol{x}^t, \boldsymbol{y}_k^t\} \in \mathcal{U}^t, K' \in [1, K]$ for further training.

3.4 Target Model Adaptation

To adapt target model, based on the autonomously selected source domains, target model is first defined as the linear combination of models learned in equation equation 1:

$$\Phi = \sum_{k=1}^{K'} \omega_k \cdot \Phi_k, P = \sum_{k=1}^{K'} \omega_k \cdot P_k, K' \in [1, K]. \tag{9}$$

Then the possibility of target label is predicted as:

$$\mathcal{P}^t = P(\Phi(\boldsymbol{x}^t)). \tag{10}$$

At this stage, gaps remain between the pre-trained models P and Φ. To bridge these gaps, we employ pseudo-labels generated by a frozen foundation model (e.g., CLIP) to self-supervise the adaptation of the target model, where only prompts are fine-tuned to take the benefits of cross-modality knowledge. Denote foundation model as Ψ, and original text prompts as $\{\mathcal{T}_c\}_{c=1}^{\mathcal{C}}$, the prediction on target domain can be expressed as:

$$\mathcal{P}_{FM} = \Psi(\boldsymbol{x}^t)_{vis} \cdot (\Psi(\mathcal{T}_c)_{txt})^T. \tag{11}$$

Following previous findings in (Tang et al., 2024a; Li et al., 2025) which maintain prediction consistency under a structural causal model, the external structural causal factor is defined by maximizing the correlation between the representations of random variables and the predictions of the target model, while the refinement of the desired latent factor is conditionally equivalent to a self-supervised information bottleneck. Denote latent random variables of $\{\mathcal{T}_c\}_{c=1}^{\mathcal{C}}$ as $\mathcal{P}_V$ following distribution of $\mathcal{P}_{FM}$, and the bridging values connecting text prompts and $\mathcal{P}_{FM}$ as $\mathcal{P}'_V$, the loss function of fine-tuning prompts with a Gaussian distribution is defined as following by minimizing the errors between target model prediction $\mathcal{P}^t$ and foundation model prediction $\mathcal{P}_{FM}$:

$$\begin{aligned} \mathcal{L}_{ex} = \frac{1}{n} \sum_{i=1}^{n} \beta[(\frac{(\mathcal{P}_{FM}^i - \mathcal{P}^{ti})^2}{\text{diag}(\mathcal{P}_{FM}^i)} + \log|\text{diag}(\mathcal{P}_{FM}^i)|) + \gamma \mathcal{KL}(g(\mathcal{P}'_V) \| \mathcal{P}_V)], \\ g(\mathcal{P}'_V) = \frac{1}{\text{diag}(\mathcal{P}_V)} \odot \mathcal{P}_V. \end{aligned} \tag{12}$$

$g(\cdot)$ is a learnable function providing a probability distribution with no information loss.

The cross-entropy loss for self-supervising target model is then defined as:

$$\mathcal{L}_{in} = -\frac{1}{n}\sum_{i=1}^{n}\theta\mathcal{P}_{FM}\log\mathcal{P}^t + \delta\sum\bar{\mathcal{P}}^t\log\bar{\mathcal{P}}^t + \sum_{c=1}^{C}\mathcal{KL}(\frac{1}{n}\sum_{i=1}^{n}(\mathcal{P}^{ti})_c\|\frac{1}{C}), \bar{\mathcal{P}}^t = \frac{1}{n}\sum_{i=1}^{n}\mathcal{P}^{ti}. \tag{13}$$

Target label is finally predicted as:

$$\boldsymbol{y}^t = \arg\max\ \text{SoftMax}(P(\Phi(\boldsymbol{x}^t))). \tag{14}$$

4 Experiments and Analysis

Datasets: The proposed method is evaluated on four real-world visual classification datasets: Office31, OfficeHome, DomainNet126 and DomainNet. Office31 contains three domains with 31 shared categories, denoted as A (Amazon), D (DSLR), and W (Webcam). OfficeHome consists of four domains with 65 shared categories, denoted as A (Art), C (Clipart), P (Product) and R (Real-World). DomainNet126 includes four domains with 126 shared categories, including C (Clipart), P (Painting), R (Real), and S (Sketch). DomainNet has six domains with 345 shared categories, denoted as C (Clipart), I (Infograph), P (Painting), Q (Quickdraw), R (Real), and S (Sketch).

Parameters: For a fair comparison, the target model adopts ResNet50 as its backbone, while CLIP with a vision transformer backbone is used as a frozen foundation model to generate target predictions that guide model adaptation under self-supervision. SGD optimizer with momentum 0.9 is used to update the parameters. The initial learning rate $\eta = 0.001$, it changes linearly with training epochs as $\eta = \frac{\eta_0}{(1+10p)^{0.75}}$ for all datasets. The batchsize is 64. The trade-off parameters are set as $\beta = 0.003$, $\lambda = 0.5$, $\gamma = 0.5$, $\theta = 0.4$, $\delta = 1.0$ and $\sigma = \frac{1}{2p}$. For average radius, $\alpha = 1$. s_{adj} is defined as the median distance between source samples and their corresponding cluster centers within the same class, t_{adj} is set to one-third of the source radius for each cluster.

All experiments are carried out using Pytorch 1.12.1+cu113 on a NVIDIA RTX-A5500 GPU.

Baselines: Baselines used in this work include methods based on ResNet: CAiDA (Dong et al., 2021)(NeurIPS), FixBi (Na et al., 2021)(CVPR), SSD (Li et al., 2023)(TCYB), Co-MDA (Liu et al., 2023)(TCSVT), DCL (Tian et al., 2023)(TCSVT), GSDE (Westfechtel et al., 2024)(WACV), SEAL (Xia et al., 2024)(AAAI), MPA (Chen et al., 2024)(NeurIPS), KGCDE (Wong et al., 2024)(PR), DSACDIC (Zhao et al., 2024)(WACV) LCFD (Tang et al., 2024a)(ArXiv), DAMP (Du et al., 2024)(CVPR), Ucon-SFDA (Xu et al., 2025)(ICLR), FuzHDA (Li et al., 2025)(TFS) ProDe (Tang et al., 2025)(ICLR) TIGM (Zhu et al., 2025)(CVPR); and methods based on vision transformer: DeiT (Touvron et al., 2021)(ICML), CDTrans (Xu et al., 2021)(ICLR), SSRT (Sun et al., 2022)(CVPR), DSiT (Sanyal et al., 2023)(ICCV).

In the proposed method AutoS, when ResNet is used as the backbone, it is fully fine-tuned, whereas when a Vision transformer is used, the backbone is frozen. "AutoS/sf" denotes AutoS under the source-free setting, while "AutoS*" denotes AutoS based on a frozen Vision Transformer backbone.

Results: Tables 1 and 2 show the transfer performance on four datasets. The proposed AutoS consistently outperforms most baselines. On Office31 and DomainNet126, AutoS trained with source data surpasses its source-free model AutoS/sf, whereas on OfficeHome the source-free setting performs better. Compared with multi-domain feature-alignment methods such as SSD and MPA, AutoS achieves higher accuracy while using fewer source knowledge. Against self-supervision approaches, such as CAiDA, Ucon-SFDA and methods that leverage large language models, like SEAL and LCFD, AutoS benefits from a target-confident sample selection strategy that effectively separates high- and low-confidence target samples for adaptation. Relative to federated-learning baselines Co-MDA and FuzHDA, AutoS still delivers superior performance on most tasks while relying only on selected source knowledge.

Ablation Study: Table 3 shows the performance of AutoS on OfficeHome using an ablation strategy. "FedAvg" refers to a model trained without source-knowledge selection, combining domains through simple federated averaging. "w/o TarCof" indicates training without adding target samples

Table 1: Results of AutoS on datasets Office31 and OfficeHome.

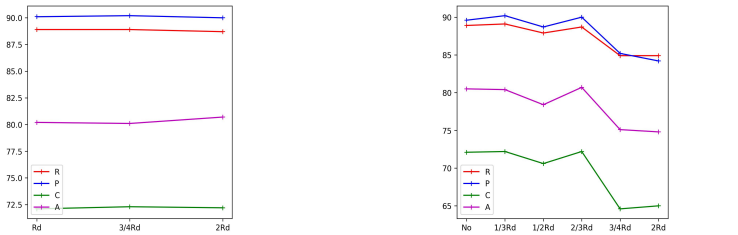
Method	D	W	A	Avg	Method	R	P	C	A	Avg
CAiDA	**99.8**	98.9	75.8	91.5	CAiDA	84.2	84.7	60.5	75.2	76.2
SSD	**99.8**	**99.1**	76.0	91.6	SSD	83.2	81.2	64.5	72.5	75.4
Co-MDA	96.3	95.3	75.3	89.0	Co-MDA	83.9	85.3	64.0	74.4	76.9
GSDE	97.8	97.8	78.8	91.7	GSDE	82.1	81.6	59.2	71.6	73.6
FixBi	97.5	97.7	79.1	91.4	MPA	85.7	86.2	54.9	74.8	75.4
SEAL	97.2	92.2	77.1	88.8	SEAL	84.3	82.7	57.8	73.4	74.6
KGCDE	99.8	99.2	79.4	92.8	KGCDE	85.4	83.7	68.3	75.6	78.3
DCL	97.3	95.3	77.2	89.9	DAMP	86.9	89.1	60.1	76.6	78.2
DSACDIC	95.9	96.8	75.9	89.5	TIGM	82.5	82.6	60.7	70.7	74.1
LCFD	93.4	92.5	82.8	89.7	LCFD	89.7	**90.2**	72.2	80.7	83.2
Ucon-SFDA	97.4	97.1	77.1	90.5	Ucon-SFDA	80.9	82.3	61.9	69.5	73.6
FuzHDA	99.2	97.0	83.5	**93.2**	FuzHDA	87.9	89.9	71.4	81.2	82.6
ProDe	96.5	93.9	79.4	89.9	ProDe	89.0	89.7	64.9	80.6	81.1
AutoS	97.0	97.0	**83.7**	92.6	AutoS	88.7	90.0	72.2	80.7	82.9
AutoS/sf	96.6	94.6	82.8	91.3	AutoS/sf	**89.7**	89.7	**72.4**	**80.8**	**83.2**

Table 2: Results of AutoS on datasets DomainNet126 and DomainNet.

Method	C	P	R	S	Avg	Method	C	I	P	Q	R	S	Avg
GSDE	83.4	77.8	**91.0**	**79.9**	83.1	CAiDA	63.6	20.7	54.3	19.3	71.2	51.6	46.8
DAMP	74.5	76.2	88.4	71.0	77.5	SSD	**67.2**	21.7	52.4	20.8	67.8	55.3	47.5
LCFD	79.3	77.5	88.1	75.3	80.1	DSiT*	55.3	23.4	47.1	17.6	63.5	45.5	42.1
Ucon-SFDA	72.2	69.6	81.0	65.4	72.1	DeiT*	41.4	39.9	38.9	19.6	39.0	41.5	36.7
FuzHDA	82.4	80.0	88.7	76.9	82.0	SSRT*	49.8	46.3	45.0	29.3	48.8	52.1	45.2
TIGM	74.7	72.1	86.2	67.7	75.2	CDTrans*	48.7	**48.4**	46.4	**30.7**	45.6	51.5	45.2
AutoS	**83.7**	**81.0**	90.0	78.1	**83.2**	AutoS	64.1	22.5	50.2	8.2	61.1	47.5	42.3
AutoS/sf	82.1	79.7	89.2	75.0	81.5	AutoS*	64.6	32.7	**58.1**	6.2	**69.4**	**56.5**	**47.9**

Table 3: Ablation study of AutoS on OfficeHome.

Method	R	P	C	A	Avg
Fedavg	88.6	89.4	71.8	80.4	82.6
w/o TarCof	**88.9**	89.6	72.1	80.5	82.8
w/o $\mathcal{L}$	88.1	89.4	70.1	79.1	81.7
w/o $\mathcal{L}_{in}$	76.5	73.5	51.9	64.6	66.6
w/o $\mathcal{L}_{ex}$	88.2	89.3	70.1	80.2	82.0
AutoS	88.7	**90.0**	**72.2**	**80.7**	**82.9**



(a) Size of selected source data (b) Size of selected target data

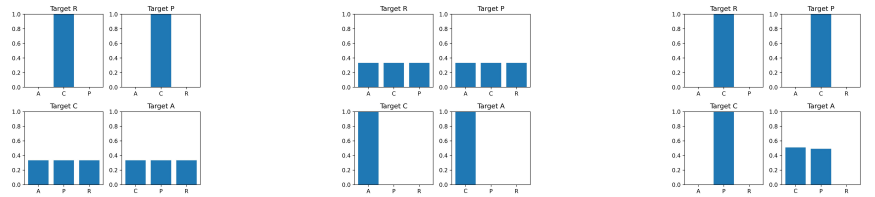
Figure 3: Transfer performance with different size of selected source and target data.

to the selected source domains. AutoS achieves the best performance. The model with no source-target interaction ($\mathcal{L}_{in}$) gives the lowest accuracy, indicating the need to bridge domain gaps. The next most critical factor is source supervision ($\mathcal{L}$), showing the value of supervision from source labels. Target pseudo-labels, enhanced by the frozen foundation model and fine-tuned prompts ($\mathcal{L}_{ex}$), provide the third-highest gain.

Figure 3 shows the influence of the size of selected source and target confident samples on OfficeHome. "Rd" denotes the average radius of the source clusters. For the source data, removing samples whose distance to their cluster centers exceeds the mean radius does not affect transfer performance, indicating that the source domains contain redundant knowledge. For the target data, the best transfer performance is achieved when using confident target samples whose distance to the mean source radius is less than one third of that radius.

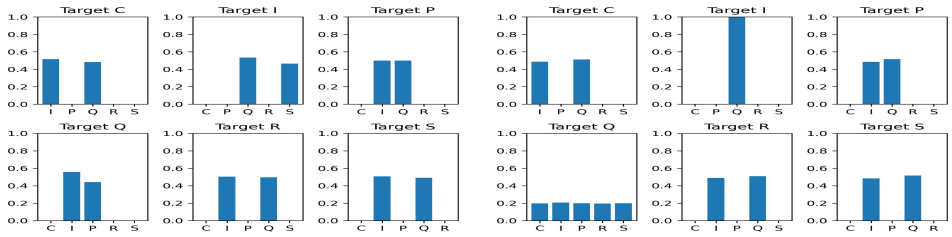
Table 4: Performance of AutoS with different radius metric.

	ACCURACY					TIME				
Method	R	P	C	A	Avg	R	P	C	A	Avg
RMS	88.7	**90.1**	69.9	78.0	81.7	**7718.3**	9205.8	7033.1	4867.5	7206.2
MAX	**88.8**	89.7	72.0	79.8	82.6	8314.4	8289.1	8299.8	5141.1	7511.1
MEAN	88.7	90.0	**72.2**	**80.7**	**82.9**	9130.3	**6998.4**	**6492.5**	**4351.5**	**6743.2**



(a) Average radius (b) Root mean square radius (c) Maxium radius

Figure 4: Transfer performance with different distance metrics for defining radius of clusters, taking OfficeHome as example.



(a) Sected by ResNet (b) Selected by ViT

Figure 5: Autonomously select source domains in DomainNet.

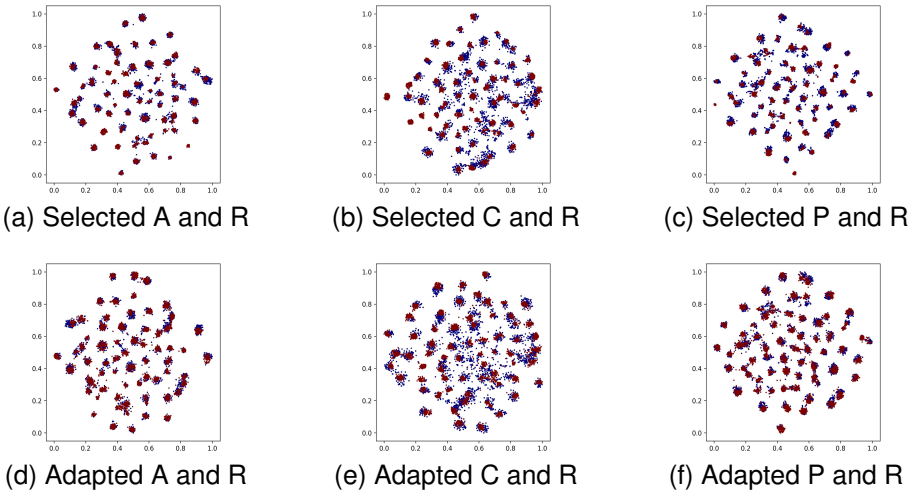
Table 4 reports the transfer performance and computational time of AutoS with different radius metrics. "RMS" is the root mean square distance, "MAX" is the maximum radius, and "MEAN" is the average radius. The results show that the average radius obtains the best performance while requiring the least running time. Figures 4 and 5 show the autonomously selected source domains using density-controlled strategy. For most tasks, the proposed AutoS achieves better performance than multi-domain adaptation baselines that use all sources, while requiring only one or two source domains.

Table 5 reports the running time and GPU memory usage of the proposed AutoS compared with two multi-domain baselines, SSD and FuzHDA. The reported running time includes both source and target model training. It shows that AutoS requires less time and lower memory than the baselines.

Table 5: Running time and GPU memory usage.

Method	R	P	C	A	Avg	GPU
SSD	62476.2	41824.8	42081.2	42190.0	47143.0	12415.0
FuzHDA	**7314.0**	7243.1	7323.8	7467.3	7337.0	8195.0
AutoS	9130.3	**6998.4**	**6492.5**	**4351.5**	**6743.2**	**8157.0**

Visualization: Figure 6 illustrates the distributions of the selected source and target samples, along with the adaptation of the complete source and target domains using task R from OfficeHome as an example. It shows that the three source domains align well with the target domain. The reason is that AutoS autonomously selects source domains during training, the target model initially learns from all sources but later removes those defined as dissimilar. This autonomous selection strategy allows the method to emphasize the most relevant source domains at the appropriate time while still preserving transferable knowledge from all sources.



(a) Selected A and R (b) Selected C and R (c) Selected P and R

(d) Adapted A and R (e) Adapted C and R (f) Adapted P and R

Figure 6: T-SNE of adapted source and target data on task R from OfficeHome.

5 Conclusion and Future Work

This work introduces an autonomous source knowledge selection method that dynamically identifies relevant and irrelevant source domains with respect to the target domain during training. First, the method applies a federated learning strategy to leverage transferable knowledge from multiple sources and designs a density-controlled selection strategy to determine relevant domains while assigning combination weights for federated aggregation of source models in adapting to the target. Furthermore, rather than relying on data matching, the method adapts the target domain through self-supervision using pseudo-labels enhanced by a frozen foundation model, while fine-tuning only the prompts. This design makes the approach flexible for domain adaptation both with and without access to source data.

In futuer study, we will extend the proposed method to deal with more complex settings such as label shifts and modality shifts

LLMs Usage

The authors used ChatGPT for grammar and spelling checks only, with prompt ”Proofread the sentences”.

REFERENCES

Shuanghao Bai, Min Zhang, Wanqi Zhou, Siteng Huang, Zhirong Luan, Donglin Wang, and Badong Chen. Prompt-based distribution alignment for unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pp. 729–737, 2024.

Atif Belal, Akhil Meethal, Francisco Perdigon Romero, Marco Pedersoli, and Eric Granger. Multi-source domain adaptation for object detection with prototype-based mean teacher. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1277–1286, 2024.

Haoran Chen, Xintong Han, Zuxuan Wu, and Yu-Gang Jiang. Multi-prompt alignment for multi-source unsupervised domain adaptation. *Advances in Neural Information Processing Systems*, 36, 2024.

Jiahua Dong, Zhen Fang, Anjin Liu, Gan Sun, and Tongliang Liu. Confident anchor-induced multi-source free domain adaptation. *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, 34, December 6 - 14 2021.

Zhekai Du, Xinyao Li, Fengling Li, Ke Lu, Lei Zhu, and Jingjing Li. Domain-agnostic mutual prompting for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23375–23384, 2024.

Yuchun Fang, Chen Chen, Wei Zhang, Jiahua Wu, Zhaoxiang Zhang, and Shaorong Xie. Prototype learning for adversarial domain adaptation. *Pattern Recognition*, 155:110653, 2024.

Yifei He, Haoxiang Wang, Bo Li, and Han Zhao. Gradual domain adaptation: Theory and algorithms. *Journal of Machine Learning Research*, 25(361):1–40, 2024.

Wenke Huang, Mang Ye, Zekun Shi, He Li, and Bo Du. Rethinking federated learning with domain shift: A prototype view. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16312–16322. IEEE, 2023.

Neehar Kondapaneni, Markus Marks, Manuel Knott, Rogério Guimaraes, and Pietro Perona. Text-image alignment for diffusion-based perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13883–13893, 2024.

Lingjing Kong, Shaoan Xie, Weiran Yao, Yujia Zheng, Guangyi Chen, Petar Stojanov, Victor Akinwande, and Kun Zhang. Partial disentanglement for domain adaptation. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 11455–11472. PMLR, 2022.

Keqiuyin Li, Jie Lu, Hua Zuo, and Guangquan Zhang. Multidomain adaptation with sample and source distillation. *IEEE Transactions on Cybernetics*, 2023.

Keqiuyin Li, Jie Lu, Hua Zuo, and Guangquan Zhang. Fuzzy domain adaptation from heterogeneous source teacher models. *IEEE Transactions on Fuzzy Systems*, 2025.

Xinyao Li, Jingjing Li, Fengling Li, Lei Zhu, and Ke Lu. Agile multi-source-free domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 13673–13681, 2024.

Mattia Litrico, Alessio Del Bue, and Pietro Morerio. Guiding pseudo-labels with uncertainty estimation for test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Xinhui Liu, Wei Xi, Wen Li, Dong Xu, Gairui Bai, and Jizhong Zhao. Co-mda: Federated multi-source domain adaptation on black-box models. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

Xinhui Liu, Zhenghao Chen, Luping Zhou, Dong Xu, Wei Xi, Gairui Bai, Yihan Zhao, and Jizhong Zhao. Ufda: Universal federated domain adaptation with practical assumptions. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pp. 14026–14034, 2024.

Yuwu Lu, Haoyu Huang, Xue Hu, Zhihui Lai, and Xuelong Li. Multiple adaptation network for multi-source and multi-target domain adaptation. *IEEE Transactions on Multimedia*, 2025.

Guangzhi Ma, Jie Lu, and Guangquan Zhang. Multi-source domain adaptation with interval-valued target data via fuzzy neural networks. *IEEE Transactions on Fuzzy Systems*, 2024.

Jaemin Na, Heechul Jung, Hyung Jin Chang, and Wonjun Hwang. Fixbi: Bridging domain spaces for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1094–1103, 2021.

Liangbo Ning, Zuowei Zhang, Weiping Ding, Dian Shao, and Yining Zhu. Multilevel distribution alignment for multisource universal domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

Sanqing Qu, Tianpei Zou, Lianghua He, Florian Röhrbein, Alois Knoll, Guang Chen, and Changjun Jiang. Lead: Learning decomposition for source-free universal domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23334–23343, 2024.

Sunandini Sanyal, Ashish Ramayee Asokan, Suvaansh Bhambri, Akshay Kulkarni, Jogendra Nath Kundu, and R Venkatesh Babu. Domain-specificity inducing transformers for source-free domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18928–18937, 2023.

Pascal Schlachter, Simon Wagner, and Bin Yang. Memory-efficient pseudo-labeling for online source-free universal domain adaptation using a gaussian mixture model. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 6425–6434. IEEE, 2025.

Tao Sun, Cheng Lu, Tianshuo Zhang, and Haibin Ling. Safe self-refinement for transformer-based domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7191–7200, 2022.

Song Tang, Wenxin Su, Mao Ye, Jianwei Zhang, and Xiatian Zhu. Unified source-free domain adaptation. *arXiv preprint arXiv:2403.07601*, 2024a.

Song Tang, Wenxin Su, Mao Ye, and Xiatian Zhu. Source-free domain adaptation with frozen multimodal foundation model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23711–23720, 2024b.

Song Tang, Wenxin Su, Yan Gan, Mao Ye, Jianwei Zhang, and Xiatian Zhu. Proxy denoising for source-free domain adaptation. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

Qing Tian, Heyang Sun, Shun Peng, Yuhui Zheng, Jun Wan, and Zhen Lei. Dcl: Dipolar confidence learning for source-free unsupervised domain adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(6):4342–4353, 2023.

Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 10347–10357. PMLR, 2021.

Fuli Wan, Han Zhao, Xu Yang, and Cheng Deng. Unveiling the unknown: Unleashing the power of unknown to known in open-set source-free domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24015–24024, 2024.

Rui Wang, Zuxuan Wu, Zejia Weng, Jingjing Chen, Guo-Jun Qi, and Yu-Gang Jiang. Cross-domain contrastive learning for unsupervised domain adaptation. *IEEE Transactions on Multimedia*, 25: 1665–1673, 2022.

Thomas Westfechtel, Hao-Wei Yeh, Dexuan Zhang, and Tatsuya Harada. Gradual source domain expansion for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1946–1955, 2024.

Garrett Wilson, Janardhan Rao Doppa, and Diane J Cook. Calda: Improving multi-source time series domain adaptation with contrastive adversarial learning. *IEEE transactions on pattern analysis and machine intelligence*, 45(12):14208–14221, 2023.

Wai Keung Wong, Yuwu Lu, Zhihui Lai, and Xuelong Li. Graph correlated discriminant embedding for multi-source domain adaptation. *Pattern Recognition*, pp. 110538, 2024.

Mingxuan Xia, Junbo Zhao, Gengyu Lyu, Zenan Huang, Tianlei Hu, Gang Chen, and Haobo Wang. A separation and alignment framework for black-box domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pp. 16005–16013, 2024.

Gezheng Xu, Hui Guo, Li Yi, Charles Ling, Boyu Wang, and Grace Yi. Revisiting source-free domain adaptation: a new perspective via uncertainty control. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

Tongkun Xu, Weihua Chen, Pichao Wang, Fan Wang, Hao Li, and Rong Jin. Cdtrans: Cross-domain transformer for unsupervised domain adaptation. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

Xingtong Yu, Zechuan Gong, Chang Zhou, Yuan Fang, and Hui Zhang. Samgpt: Text-free graph foundation model for multi-domain pre-training and cross-domain adaptation. In *Proceedings of the ACM on Web Conference 2025*, pp. 1142–1153, 2025.

Ningyuan Zhang, Jie Lu, Keqiuyin Li, Zhen Fang, and Guangquan Zhang. Release the powers of prompt tuning: Cross-modality prompt transfer. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

Pingping Zhang, Shiqi Wang, Meng Wang, Jiguo Li, Xu Wang, and Sam Kwong. Rethinking semantic image compression: Scalable representation with cross-modality transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023a.

Wenyu Zhang, Li Shen, and Chuan-Sheng Foo. Rethinking the role of pre-trained networks in source-free domain adaptation. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 18841–18851, 2023b.

Yixin Zhang, Zilei Wang, and Weinan He. Class relationship embedded learning for source-free unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7619–7629, 2023c.

Sicheng Zhao, Guangzhi Wang, Shanghang Zhang, Yang Gu, Yaxian Li, Zhichao Song, Pengfei Xu, Runbo Hu, Hua Chai, and Kurt Keutzer. Multi-source distilling domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pp. 12975–12983, New York, USA, February 7 - 12 2020.

Yewei Zhao, Hu Han, Shiguang Shan, and Xilin Chen. Deep subdomain alignment for cross-domain image classification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2820–2829, 2024.

Ronghang Zhu, Mengxuan Hu, Weiming Zhuang, Lingjuan Lyu, Xiang Yu, and Sheng Li. Revisiting source-free domain adaptation: Insights into representativeness, generalization, and variety. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25688–25697, 2025.

A APPENDIX

A.1 PROOF OF THE OBJECTIVE FUNCTION FOR STRUCTURAL CAUSAL MODEL

Given target predictions provided by source federal model and frozen foundation model as $\mathcal{P}^t$ and $\mathcal{P}_{FM}$ respectively, latent values of text prompts $\{\mathcal{T}_c\}_{c=1}^{\mathcal{C}}$ as V, latent random values following distribution of $\mathcal{P}^t$ and $\mathcal{P}_{FM}$ as $\mathcal{P}_Y$ and $\mathcal{P}_V$, the intermediate bottleneck values connecting latent text prompts V and $\mathcal{P}_{FM}$ as $\mathcal{P}'_V$. To ensure the relationship between target model and text prompts, we following previous works (Tang et al., 2024a; Li et al., 2025) to maximize the correlation between the latent random values of text prompts V and target model predictions $\mathcal{P}_Y$, which is:

$\max \mathcal{I}(\mathcal{P}_Y, V) = \max \mathcal{I}(\mathcal{P}_Y, \mathcal{P}_V), \mathcal{P}_V = \Psi(V)_{txt}$.

Based on the statement about information loss under dimensional compression, for random values Z, X and Y, there is a function $h(\cdot)$ satisfies that $X \in \mathcal{R}^m \overset{h(\cdot)}{\to} Y \in \mathcal{R}^{\mathcal{C}}$, if $\mathcal{C} < m$, then:

(1). There is $X' \neq X$ satisfying $h(X') = h(X)$;

(2) If $Z \to X \to Y$ a Markov chain, then $\mathcal{I}(Z, Y) \leq \mathcal{I}(Z, X)$.

Let $Z = \mathcal{P}_Y$, $Y = \mathcal{P}_V$ and $X = V$, then we have

$$\mathcal{I}(\mathcal{P}_Y, \mathcal{P}_V) \leq \mathcal{I}(\mathcal{P}_Y, V). \tag{15}$$

Simultaneously, for random values Z, X and Y, there are intermediate value Z' and intermediate value Y' satisfying $Z' \overset{g(\cdot)}{\to} Y'$, and $Z \to Z' \to (Y, Y')$ while $g(\cdot)$ is reversible and uncompressed, then:

(1). $\mathcal{I}(Z, Y') = \mathcal{I}(Z, g(Z')) = \mathcal{I}(Z, Z')$

(2). If $Z' \to Y' \to Y$ a Markov chain, then $\mathcal{I}(Y', Y) \leq \mathcal{I}(Z', Y)$.

Then we have:

$$\mathcal{I}(Z, Y) = \mathcal{I}(Z, Z') - \mathcal{I}(Z', Y) \leq \mathcal{I}(Z, Z') - \mathcal{I}(Y', Y), Y' = g(Z'). \tag{16}$$

Let $Z' = \mathcal{P}'_v$, $Y' = \mathcal{P}'_Y = g(\mathcal{P}'_V)$, we have:

$$\begin{aligned} \mathcal{I}(\mathcal{P}_Y, \mathcal{P}_V) &= \mathcal{I}(\mathcal{P}_Y, \mathcal{P}'_V) - \mathcal{I}(\mathcal{P}'_V, \mathcal{P}_V) \\ &\leq \mathcal{I}(\mathcal{P}_Y, \mathcal{P}'_V) - \mathcal{I}(\mathcal{P}'_Y, \mathcal{P}_V) \\ &= \mathcal{I}(\mathcal{P}_Y, \mathcal{P}'_V) - \mathcal{I}(g(\mathcal{P}'_V), \mathcal{P}_V). \end{aligned} \tag{17}$$

The upper bound of $\mathcal{I}(\mathcal{P}_Y, \mathcal{P}'_V)$ is calculated as:

$$\begin{aligned} \mathcal{I}(\mathcal{P}_Y, \mathcal{P}'_V) &\equiv H(\mathcal{P}'_V) - H(\mathcal{P}'_V|\mathcal{P}_Y) \equiv H(\mathcal{P}_Y) - H(\mathcal{P}_Y|\mathcal{P}'_V) \\ &= -\sum \mathbb{P}(\mathcal{P}_Y) \log \mathbb{P}(\mathcal{P}_Y) - \sum \mathbb{P}(\mathcal{P}'_V, \mathcal{P}_Y) \log \frac{\mathbb{P}(\mathcal{P}'_V, \mathcal{P}_Y)}{\mathbb{P}(\mathcal{P}'_V)} \\ &= \mathbb{E}[\log \mathbb{P}(\mathcal{P}_Y)] + \mathbb{E}[\log \frac{\mathbb{P}(\mathcal{P}'_V, \mathcal{P}_Y)}{\mathbb{P}(\mathcal{P}'_V)}] \\ &= \mathbb{E}[\log \mathbb{P}(\mathcal{P}'_V)] + \mathbb{E}[\log \frac{\mathbb{P}(\mathcal{P}_Y, \mathcal{P}'_V)}{\mathbb{P}(\mathcal{P}_Y)}] \\ &\geq \text{const} + \mathbb{E}[\log \mathbb{P}(\mathcal{P}'_V|\mathcal{P}_Y))], \end{aligned} \tag{18}$$

Calculation of $\mathcal{I}(g(\mathcal{P}'_V), \mathcal{P}_V)$ is expressed as:

$$\begin{aligned} \mathcal{I}(g(\mathcal{P}'_V), \mathcal{P}_V) &= \mathcal{KL}(g(\mathcal{P}'_V) \| \mathcal{P}_V), \\ g(\mathcal{P}'_V) &= \frac{1}{\text{diag}(\mathcal{P}_Y)} \odot \mathcal{P}_V. \end{aligned} \tag{19}$$

Then we can get Equation 12 by modeling $\mathbb{P}(\mathcal{P}'_V|\mathcal{P}_Y)$ as Gaussian distribution $\mathbb{N}(\mathcal{P}_Y, \text{diag}(\mathcal{P}_Y))$ and updating the random values with $\mathcal{P}^t$ and $\mathcal{P}_{FM}$ during training.

A.2 ALGORITHM OF THE PROPOSED AUTOS

The whole algorithm of the proposed AutoS is summarized as in Algorithm 1:

Algorithm 1 AutoS: Autonomous source knowledge selection.

Input: Source domains $\{\mathcal{D}_k^s = \{\boldsymbol{x}_{ki}^s, \boldsymbol{y}_{ki}^s\}_{i=1}^{n_k}\}_{k=1}^K$, target domain $\mathcal{D}^t = \{\boldsymbol{x}_i^t\}_{i=1}^n$;
Initialization: Feature extraction module Φ, classifier P, text prompts $\{\mathcal{T}_c\}_{c=1}^{\mathcal{C}}$;
for $\epsilon = 1, \epsilon < \mathcal{I}, \epsilon++$, **do**
  Φ_k, $P_k \leftarrow \Phi$, P: Train source independent models Φ_k, P_k as in equation 1;
  $\{\boldsymbol{f}_{kc}^s\}_{c=1}^{\mathcal{C}} \leftarrow P_k$: Collect the kth source cluster centers equation 2;
  $\boldsymbol{y}_k^t \leftarrow \boldsymbol{x}^t$, $\boldsymbol{f}_{kc}^s$: Collect target clustering label as in equation 3;
  $\boldsymbol{r}_{kc}^s \leftarrow \boldsymbol{x}_k^s$,$\boldsymbol{f}_{kc}^s$, $\boldsymbol{r}_{kc}^t \leftarrow \boldsymbol{x}^t$, $\boldsymbol{f}_{kc}^s$: Calculate cluster radius of source and target data as in equation 4;
  $\boldsymbol{d}_{kc}^s$, $\boldsymbol{d}_{kc}^t \leftarrow \boldsymbol{r}_{kc}^s$: Define thresholds to identify confident source and target samples as in equation 5;
  $\rho_{kc}^t \leftarrow \boldsymbol{r}_{kc}^t$: Define thresholds to identify confident source and target samples as in equation 6;
  Renew source domains by adding selected target samples while removing irrelevant source samples;
  ω_k: Calculate weights to select relevant source domains as in equation 7;
  Keep($\mathcal{D}_k^s$): Select relevant source domains as in equation 8;
  Φ, $P \leftarrow \Phi_k$, P_k, $k \in K'$: Get target model as in equation 9;
  $\mathcal{P}^t \leftarrow \boldsymbol{x}^t$: Collect target prediction as in equation 10;
  $\mathcal{P}^{FM} \leftarrow \boldsymbol{x}^t$, $\mathcal{T}_c$: Collect frozen foundation model prediction as in equation 11;
  $\mathcal{T} \leftarrow \mathcal{P}^{FM}$, $\mathcal{P}^t$, $\mathcal{P}^V$, $\mathcal{P}'^V$: Update text prompts as in equation 12;
  Φ, $P \leftarrow \mathcal{P}^{FM}$, $\mathcal{P}^t$: Update target model as in equation 13;
end for
Output: Target label $\boldsymbol{y}_t$ via equation 14

A.3 SELECTION STATUS OF SOURCE DOMAINS WITH TRAINING PROGRESS

This work employs cosine distance to computing the distance between source/target samples and their cluster centers, which is expressed as:

$$\text{Dis} = \frac{\boldsymbol{x}_{(\cdot)}^* \cdot \boldsymbol{f}_{kc}^s}{\|\boldsymbol{x}_{(\cdot)}^*\| \cdot \|\boldsymbol{f}_{kc}^s\|}, * \in s, t, (\cdot) = k \text{ if } * = s$$

Average radius is expressed as:

$$\text{Rd} = \frac{1}{n_{kc}} \sum_{i=1}^{n_{kc}} \frac{\boldsymbol{x}_{(\cdot)}^* \cdot \boldsymbol{f}_{kc}^s}{\|\boldsymbol{x}_{(\cdot)}^*\| \cdot \|\boldsymbol{f}_{kc}^s\|}, * \in s, t, (\cdot) = k \text{ if } * = s$$

n_{kc} is the number of kth source samples belonging to cth class.

Root mean square radius is expressed as:

$$\text{Rd} = \sqrt{\frac{1}{n_{kc}} \sum_{i=1}^{n_{kc}} (\frac{\boldsymbol{x}_{(\cdot)}^* \cdot \boldsymbol{f}_{kc}^s}{\|\boldsymbol{x}_{(\cdot)}^*\| \cdot \|\boldsymbol{f}_{kc}^s\|})^2}, * \in s, t, (\cdot) = k \text{ if } * = s$$

If employing root mean square radius, $\alpha = 1.5$.

Maximum radius is expressed as:

$$\text{Rd} = \max_{i \in [1, n_{kc}]} \frac{\boldsymbol{x}_{(\cdot)}^{*i} \cdot \boldsymbol{f}_{kc}^s}{\|\boldsymbol{x}_{(\cdot)}^{*i}\| \cdot \|\boldsymbol{f}_{kc}^s\|}, * \in s, t, (\cdot) = k \text{ if } * = s$$

Autonomous source selection status is displayed as Figures 7, 8 and 9.

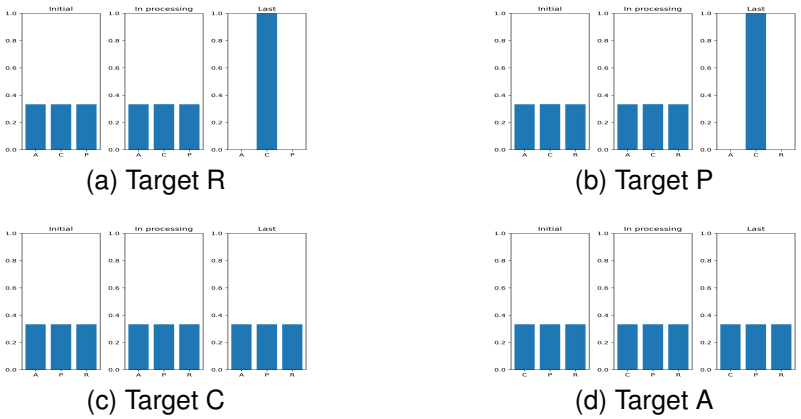


Figure 7: Source domain selection status using average radius of OfficeHome.

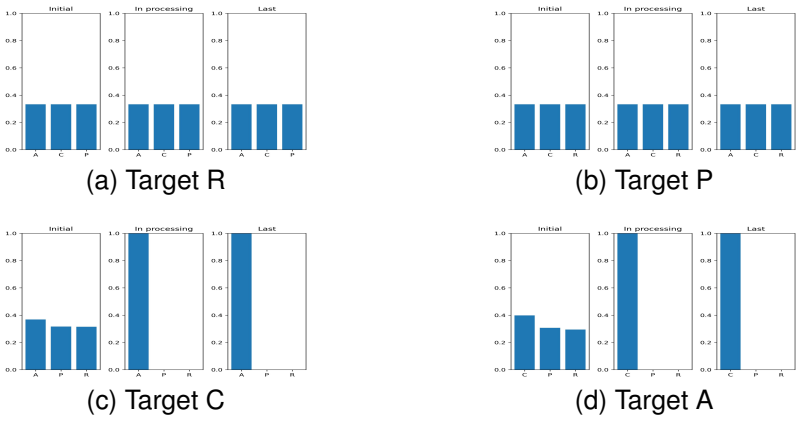


Figure 8: Source domain selection status using root mean square radius of OfficeHome.

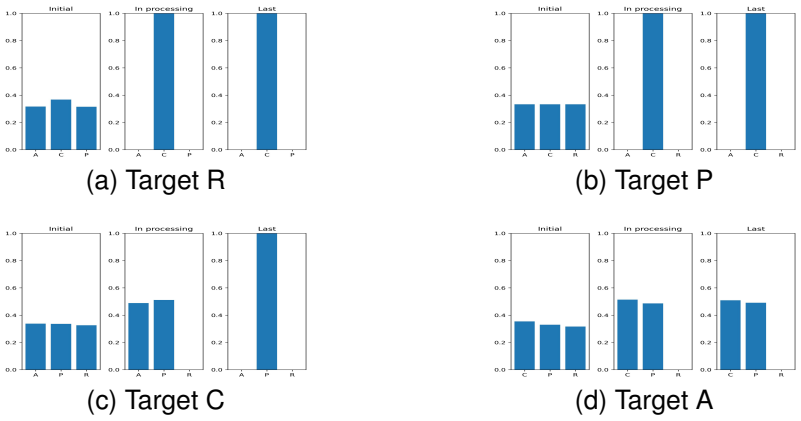


Figure 9: Source domain selection status using maximum radius of OfficeHome.

The results show that different cluster radius metrics play a key role in autonomously selecting source domains. The mean radius is more likely to select relevant domains because the proportion of target samples falling within the cluster radius contributes to the computation of selection weights. The root mean square radius and the maximum radius generally provide larger radius values, resulting in more target samples being added to the updated source domains and thereby increasing label noise when training the federated model.

A.4 Description of compared baselines

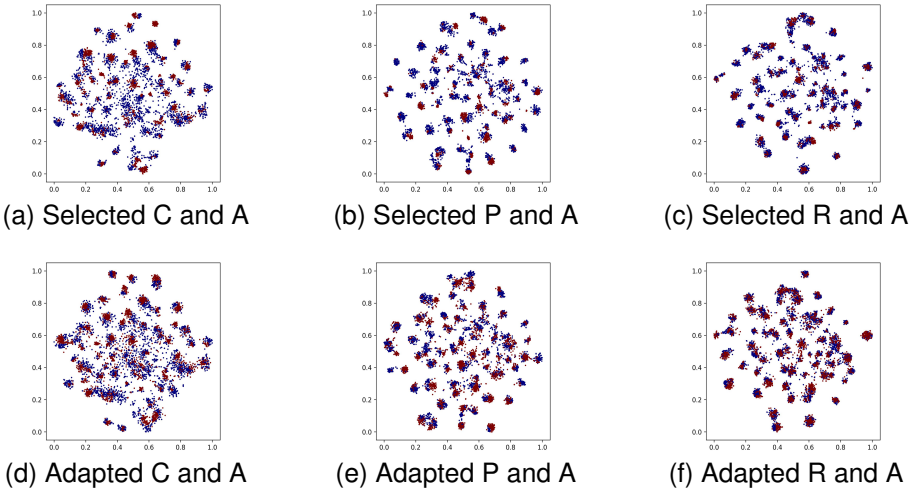
Settings and core techniques of the baselines are described in Table 6:

Table 6: The baselines. "SF" means source-free, "MS" means multi-source, "BB" means backbone structure, while "FFM" indicates frozen foundation model.

Baseline	Venue	SF	MS	BB	FFM
CAiDA Dong et al. (2021)	NeurIPS'21	✓	✓	RN	×
FixBi Na et al. (2021)	CVPR'21	✓	×	RN	×
DeiT Touvron et al. (2021)	ICML'21	✓	×	ViT	×
CDTrans Xu et al. (2021)	ICLR'21	✓	×	ViT	×
SSRT Sun et al. (2022)	CVPR'22	✓	×	ViT	×
DSiT Sanyal et al. (2023)	ICCV'23	✓	×	ViT	×
SSD Li et al. (2023)	TCYB'23	×	✓	RN	×
Co-MDA Liu et al. (2023)	TCSVT'23	✓	✓	RN	×
DCL Tian et al. (2023)	TCSVT'23	✓	×	RN	×
GSDE Westfechtel et al. (2024)	WACV'24	✓	×	RN	×
MPA Chen et al. (2024)	NeurIPS'23	×	✓	RN	✓
SEAL Xia et al. (2024)	AAAI'24	✓	✓	RN	×
KGCDE Wong et al. (2024)	PR'24	×	✓	RN	×
DSACDIC Zhao et al. (2024)	WACV'24	×	×	RN	×
LCFD Tang et al. (2024a)	ArXiv'24	✓	×	RN	✓
DAMP Du et al. (2024)	CVPR'24	✓	×	RN	✓
Ucon-SFDA Xu et al. (2025)	ICLR'24	✓	×	RN	×
FuzHDA Li et al. (2025)	TFS'25	✓	✓	RN	✓
ProDe Tang et al. (2025)	ICLR'25	✓	×	RN	×
TIGM Zhu et al. (2025)	CVPR'25	✓	×	RN	×

A.5 Visualization of tasks from OfficeHome with and without source data

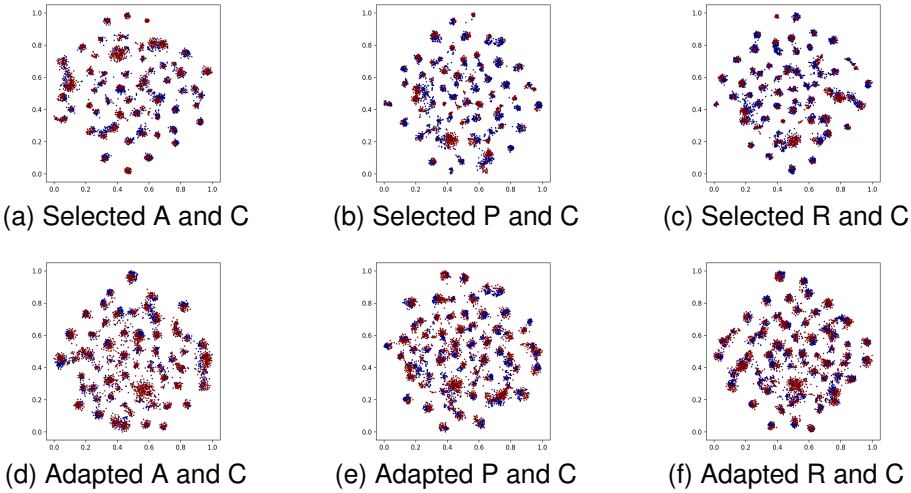
T-SNE visualization of all tasks from OfficeHome is displayed as following Figures.



(a) Selected C and A (b) Selected P and A (c) Selected R and A

(d) Adapted C and A (e) Adapted P and A (f) Adapted R and A

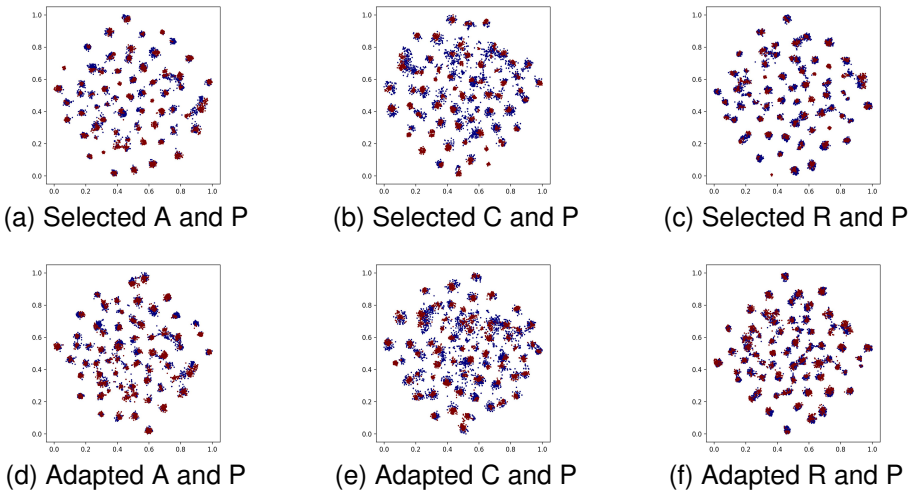
Figure 10: T-SNE of adapted source and target data on task A from OfficeHome.



(a) Selected A and C (b) Selected P and C (c) Selected R and C

(d) Adapted A and C (e) Adapted P and C (f) Adapted R and C

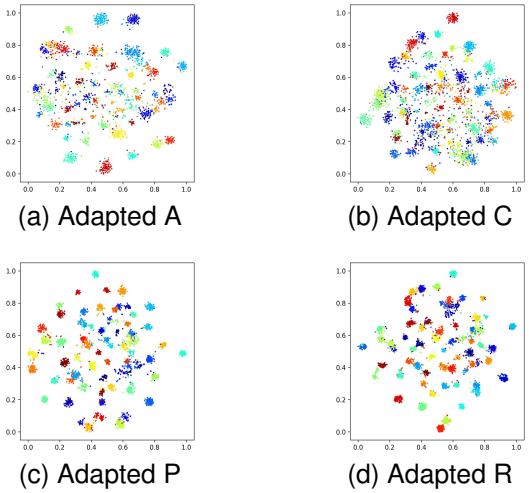
Figure 11: T-SNE of adapted source and target data on task C from OfficeHome.



(a) Selected A and P (b) Selected C and P (c) Selected R and P

(d) Adapted A and P (e) Adapted C and P (f) Adapted R and P

Figure 12: T-SNE of adapted source and target data on task P from OfficeHome.



(a) Adapted A (b) Adapted C

(c) Adapted P (d) Adapted R

Figure 13: T-SNE of target data under source-free setting from OfficeHome.