Neurocomputing 619 (2025) 129114

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Domain knowledge boosted adaptation: Leveraging vision-language models for multi-source domain adaptation

Yuwei He [a], Juexiao Feng [b], Guiguang Ding [b,*], Yuchen Guo [c], Tao He [d]

[a] The 15th Research Institute of China Electronics Technology Group Corporation, Beijing, China
[b] School of Software, Tsinghua University, Beijing, China
[c] Institute for Brain and Cognitive Sciences, BNRist, Tsinghua University, Beijing, China
[d] GRG Banking Equipment Co., Ltd., Guangzhou, China

ARTICLE INFO

Communicated by W. Zhou

Keywords:
Deep learning
Multi-source domain adaptation
Prompt learning

ABSTRACT

Multi-source domain adaptation (MSDA) aims to adapt a model trained on multiple labeled source domains to an unlabeled target domain. Existing MSDA methods primarily focus on reducing domain gaps by aligning the source domains with the target domain, either jointly or separately. However, these methods often distort semantic-related features and overlook the valuable domain-related information present in diverse domains. In this paper, we propose a novel MSDA method called Domain Knowledge Boosted Adaptation (DKBA) that leverages domain-related information to enhance model performance. Firstly, we employ prompt learning to embed domain-related information learned from a pretrained vision-language model into prompt embeddings. These embeddings serve as conditional priors, allowing the classification model to adaptively embed semantic-related features and obtain domain-invariant semantic features without excessively aligning domains. Our proposed DKBA approach achieves state-of-the-art results on four MSDA datasets, highlighting its effectiveness in leveraging domain knowledge for improved adaptation performance.

1. Introduction

The success of deep learning heavily relies on the assumption that the training and test data share the same distribution [1]. However, in practical applications, the testing data often have a different distribution from the training data, leading to significant performance degradation. Additionally, the training data are collected from various domains, and ignoring the distribution discrepancies within the training data can also result in performance degradation.

The issue of domain bias has elevated the research significance of Multi-Source Domain Adaptation (MSDA). MSDA addresses the crucial requirement of leveraging knowledge obtained from multiple source domains and applying it to a related yet distinct target domain, where the data in the target domain are unlabeled. The primary objective of MSDA is to mitigate the decrease in accuracy caused by domain shifts. Existing methods have naturally emerged with the idea of enhancing accuracy by reducing domain shifts across different domains. These methods either unify all domains jointly or align each source domain separately with the target domain. Joint domain unifying methods aim to minimize the disparities between the target domain and the combined source domains [2,3], while separate domain alignment methods align the target domain with each source domain pair-wise, generating multiple source-specific classification outcomes for a target sample [4,5]. Both of these domain alignment-based approaches have demonstrated certain levels of success.

However, the process of reducing discrepancies between domains can inadvertently result in the loss of both semantic-related and domain-related information. These pieces of information are crucial for the performance of the target domain. To address this challenge, existing methods have incorporated targeted designs. Some methods focus on preserving class discriminability while reducing domain shifts to mitigate the loss of semantic information [6–8]. Other methods introduce artificial priors to maintain domain relationship information and mitigate the loss of domain-related information [2,9,10]. However, these supplementary tasks and prior information are often devised based on researchers' experience, which inherently imposes certain limitations.

In this paper, we propose Domain Knowledge Boosted Adaptation (DKBA) for multi-source domain adaptation, as depicted in Fig. 1. DKBA focuses on learning and leveraging domain-related information to enhance category classification in the target domain.

* Corresponding author.
E-mail address: dinggg@tsinghua.edu.cn (G. Ding).

https://doi.org/10.1016/j.neucom.2024.129114
Received 28 April 2024; Received in revised form 16 September 2024; Accepted 1 December 2024
Available online 8 December 2024

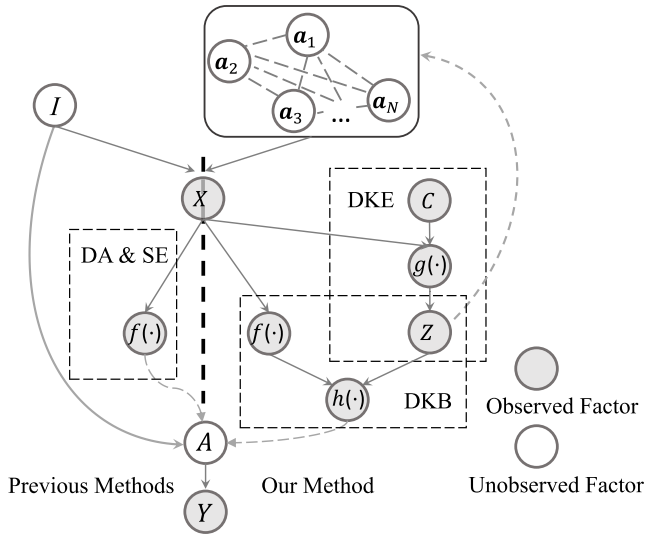

Fig. 1. We present our method using a causal graph to illustrate the underlying framework. An image X is generated as a result of two independent factors: (1) Semantic content I; and (2) Various visual elements $\{\mathbf{a}_i\}_{i=1}^{N}$, such as geometric shapes and colors. The domain we aim to comprehend consists of images with specific visual element patterns. Previous methods typically tackle domain alignment (DA) and semantic embedding (SE) simultaneously to establish a model f, capable of generating an ideal domain-invariant representation for accurate image category prediction. In our proposed method, DKBA, we adopt a two-step approach. Firstly, we conduct a domain knowledge embedding (DKE) procedure, where we independently learn the domain knowledge contained in the embedding Z. This procedure involves a vision-language model g and specially designed text prompts C. Subsequently, we proceed with a domain knowledge boosting (DKB) procedure, where we combine the learned embedding with the original feature extractor f to predict the image category.

To capture domain-related information, we utilize a vision-language (VL) model. VL models establish a shared continuous space for images and text, enabling us to extract desired image information features, specifically domain-related information, by providing appropriate query text. To obtain these information features, we introduce the prompt learning method, where query text is represented as learnable prompt embeddings. The prompt comprises three components: a learnable semantic-related context, a learnable domain-related context, and an unlearnable manually designed context ("An image of [CLASS], a [DOMAIN] image"). By employing a contrastive objective, we align the text features of prompt embeddings with their corresponding image features, effectively incorporating semantic and domain-related information into the learnable prompts. By leveraging the domain-related information, we assign pseudo-labels to images from the unlabeled domain. Subsequently, we utilize the domain-related prompt embeddings as conditional priors to parameterize the classification model. This enables the classifier to adapt to the target samples and enhance the classification performance by leveraging the domain knowledge.

Our contributions are summarized as follows:

- We propose Domain Knowledge Boosted Adaptation (DKBA) for multi-source domain adaptation, which effectively disentangles the learning of domain-related information and utilizes this information to boost semantic feature learning. By doing so, DKBA mitigates the negative impact of domain alignment on semantic learning and significantly improves the upper limit of classification accuracy.
- To the best of our knowledge, we are the first to leverage prompt learning-based knowledge to enhance the classification model in the context of multi-source domain adaptation.
- We extensively evaluate our method on four benchmark datasets. The state-of-the-art performance further validates the effectiveness of DKBA in multi-source domain adaptation.

2. Related works

2.1. Multi-source domain adaptation

In the field of multi-source domain adaptation (MSDA), the techniques for unifying domain distributions have primarily been derived from single-source domain adaptation (SSDA) methods. SSDA methods aim to enhance task performance by reducing domain distribution discrepancies. This is achieved through various techniques, such as minimizing metrics that evaluate distribution discrepancies [11,12] or employing adversarial learning to make the features indistinguishable across domains [13,14].

MSDA has given rise to two types of methods: joint domain unifying methods [2,9,15–17] and separate domain alignment methods [4,5,18–20]. Joint domain unifying methods explore domain unification at various levels, including the extractor, feature, and classifier levels. On the other hand, separate domain alignment methods address MSDA challenges by considering the interactions between the target domain and each source domain pairwise. For example, the Dynamic Generator With Attention (DGWA) method [20] utilizes dynamic parameters to adapt across different source and target domains. Recent studies have investigated the complementarity between different alignment strategies to enhance task performance [21]. However, the domain alignment adaptation procedure can lead to a degradation of semantic-related information and domain-related information. To mitigate this issue, several information-preserving methods have been proposed [6–8,22,23]. For instance, the Guided Discrimination and Correlation Subspace Learning (GDCSL) method enhances class discriminability by optimizing class scatter metrics, aligning with our goals of robust class boundaries [22]. Similarly, the Domain and Class Mutual Learning (DMAL) framework categorizes features into domain-specific and class-specific types, providing a detailed framework for feature management in adaptation scenarios [23]. Additionally, for preserving domain-related information, approaches like [2,9,10] incorporate predefined domain relationships into the learning process using graph convolution. In contrast, our method leverages prompt learning to obtain domain-related information and does not require predefined prior knowledge, thereby improving the flexibility of the approach.

2.2. Prompt learning with vision language model

Prompt learning, introduced by Petroni [24], has been extensively studied in NLP [25] It involves prepending instructions to the input and pre-training the language model to improve downstream task performance. While manually defined prompts have been used by Petroni et al. [24], they may not be optimal or appropriate, leading to inaccurate instructions. To obtain more accurate knowledge estimation from language models, methods have been proposed to automatically explore optimal prompts [25,26]. Recently, prompts have been integrated into vision-language models for learning generic visual representations [27–29]. For instance, CLIP [28] achieves state-of-the-art visual representations by pre-training a vision-language model on 400 million image–text pairs. Additionally, Zhou et al. [29] introduce CoOp, which uses continuous representations to automatically learn task-relevant prompts. In the context of domain adaptation, Ge et al. [30] designed domain-agnostic and domain-specific prompts to address distribution shift in unsupervised domain adaptation (UDA). Chen et al. [31] adopt a similar prompt designing strategy for multi-source domain adaptation (MSDA). However, these works directly adopt the vision-language model as the final classifier, which limits the ability to design the model specifically for the target domain and may restrict performance improvement in the target domain. In contrast, our approach utilizes the domain knowledge provided by vision-language models to enhance our own classifiers, allowing us to achieve better results.

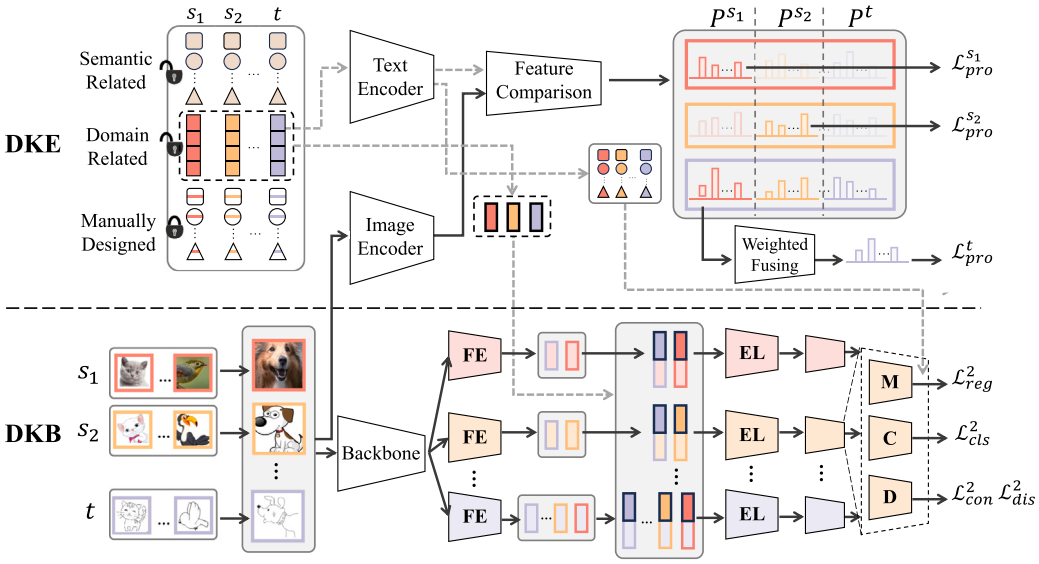

Fig. 2. An overview of our proposed DKBA. Our MSDA approach consists of two main steps: Domain Knowledge Embedding (DKE) and Domain Knowledge Boosting (DKB). In the DKE step, we leverage prompt learning on a pretrained vision language model to embed domain-related information into prompts. This learning procedure enables the vision language models to classify images from different domains by comparing features. In the DKB step, we concatenate the domain knowledge-containing prompts with the features extracted by the feature extractor (FE). This allows the subsequent embedding layer (EL) to adaptively embed the concatenated features, resulting in domain-invariant semantic features. Furthermore, we utilize the text features of the prompts to regularize the embedded features from EL. This regularization ensures semantic feature consistency and maximizes the utility of domain knowledge in the classification process.

3. Methods

In unsupervised multi-source domain adaptation, we consider a scenario where there are N labeled source domains and one unlabeled target domain, each exhibiting distinct data distributions.

Given a source domain set $\mathcal{S}$, the labeled images from the source domains are represented as $\{(X^{s_i}, Y^{s_j})\}_{j=1}^{|S|}$, where $X^{s_j} = \{x_i^{s_j}\}_{i=1}^{|X^{s_j}|}$ denotes the images and $Y^{s_j} = \left\{y_i^{s_j}\right\}_{i=1}^{|Y^{s_j}|}$ denotes the category labels. Similarly, the unlabeled target images are denoted as $X^t = \{x_i^t\}_{i=1}^{|X^t|}$. Noted that all domains share the same category set, and K represents the total number of classes. The objective of unsupervised multi-source domain adaptation (MSDA) is to develop a classifier that effectively operates on the target domain by leveraging the labeled source data and unlabeled target data.

To mitigate the adverse effects of domain alignment and leverage domain-related information to enhance task performance, we propose a method called Domain Knowledge Boosted Adaptation (DKBA). The overall framework, depicted in Fig. 2, consists of two steps: domain knowledge embeddings and domain knowledge boosting. In this section, we will explain the principles and technical details of these steps.

3.1. Domain knowledge embedding

We employ CLIP [28] as our vision-language model to capture domain-related information. CLIP [28] is trained using image–text pairs in a contrastive manner, where each input text describes the main content of its corresponding image. As a result, the text can effectively serve as a prompt to represent the image information. The prompt can be in the form of a sequence of discrete or continuous tokens, with the continuous tokens being optimized to capture the desired information type more effectively. This enables us to leverage the power of CLIP's pre-trained model to extract domain-related information and enhance our multi-source domain adaptation framework. To capture domain-related information, the prompt of class k from domain d consists of three components: a learnable semantic-related context, a learnable domain-related context, and an unlearnable manually designed context. The prompt can be represented as follows:

$$\mathbf{c}_k^d = [\mathbf{v}]_1^k[\mathbf{v}]2^k \dots [\mathbf{v}]_{M_1}^k[\mathbf{d}]_1^d[\mathbf{d}]_2^d \dots [\mathbf{d}]_{M_2}^d[\mathrm{MD}]_k^d. \tag{1}$$

Here, M_1 and M_2 represent the numbers of semantic-related and domain-related context tokens, respectively. The semantic-related context captures invariant category semantic information that is shared across all domains. On the other hand, the domain-related context is designed to capture variant image styles and is specific to each domain. Additionally, the unlearnable manually designed context MD provides a basic description of the corresponding image, such as "An image of [class name of k], a [domain name of d] image." This manually designed context helps achieve a good matching result between image and text features during the early stages of prompt optimization.

For a specific domain d, the probability that a sample belongs to the kth category can be obtained based on its prompt $\{\mathbf{c}_k^d\}_{k=1}^K$. This probability is calculated using the following equation:

$$P^d\left(\hat{y}_i = k \mid \mathbf{x}_i\right) = \frac{\exp\left(\left\langle g^{te}\left(\mathbf{c}_k^d\right), g^{ie}\left(\mathbf{x}_i\right)\right\rangle / T\right)}{\sum_{j=1}^K \exp\left(\left\langle g^{te}\left(\mathbf{c}_j^d\right), g^{ie}\left(\mathbf{x}_i\right)\right\rangle / T\right)}. \tag{2}$$

Here, g^{te} and g^{ie} correspond to the text encoder and image encoder, respectively. The $\langle\cdot,\cdot\rangle$ represents the cosine similarity. The temperature parameter T is used to control the sharpness of the probability distribution. Each domain's prompts correspond to a classification branch. Therefore, based on the prediction probabilities, the prompts can be optimized using a classification loss, such as cross-entropy loss. For the prompts of a source domain s_j, they can be directly optimized with ground truth labels using the following loss function:

$$\mathcal{L}_{pro}^{s_j} = -\mathbb{E}_{\mathbf{x}\in X^{s_j}} \log P^{s_j}\left(\hat{y}_i^{s_j} = y_i^{s_j}\right). \tag{3}$$

For the prompts of the target domain t, pseudo labels are generated by weighting and summing the prediction results from the source branches and the target branch. This process is represented by the following equation:

$$y_i^t = \arg\max_k \sum_d^{d\in S\cup\{t\}} w_i^d P^d\left(\hat{y}_i^t = k \mid \mathbf{x}_i^t\right), \tag{4}$$

Here, $\sum_{d\in S\cup\{t\}} w_i^d = 1$. The weight of each branch is considered at both the sample level and domain level. At the sample level, the weight of each sample $w_{i,sam}^d$ is determined by the prediction certainty, calculated as $w_{i,sam}^d = e^{-H(P^d(\mathbf{x}_i^t))}$, where $H(\cdot)$ represents the entropy of a probability distribution. At the domain level, the weight of each domain branch w_{dom}^d is determined by the domain similarity with the target domain. The similarity is evaluated by calculating the cosine similarity between the averaged domain-related prompts: $\langle \frac{1}{M_2}\sum_{i=1}^{M_2}\mathbf{d}_i^d, \frac{1}{M_2}\sum_{i=1}^{M_2}\mathbf{d}_i^t\rangle$. Based on these two levels, the weight w_i^d is calculated as follows:

$$w_i^d = \frac{w_{i,sam}^d w_{dom}^d}{\sum_d^{d\in S\cup\{t\}} \exp(w_{i,sam}^d w_{dom}^d)}. \tag{5}$$

Only samples whose maximum prediction probability is larger than a fixed threshold τ are used to optimize the prompts of the target domain. This optimization is done using the following loss function:

$$\mathcal{L}_{pro}^t = -\mathbb{E}_{\mathbf{x}\in X^t}\mathbb{I}\Big\{\sum_d^{d\in S\cup\{t\}} w_i^d P^d\left(\hat{y}_i^t = y_i^t \mid \mathbf{x}_i^t\right) \geq \tau\Big\} \cdot \log P^t\left(\hat{y}_i^u = y_i^t \mid \mathbf{x}_i^t\right). \tag{6}$$

Here, $\mathbb{I}\cdot$ is an indicator function. With the prompt optimization procedure described above, domain knowledge is encoded in the domain-related prompt embeddings.

3.2. Domain knowledge boosting

After optimizing the prompts, we obtain three important components that we utilize in tuning our classification model. These components are: (1) Pseudo-labels of the target domain: $Y^t = \{y_i^t\}_{i=1}^{|Y^t|}$; (2) Domain-related prompt embeddings: $\{\mathbf{d}^d\}_{d\in S\cup\{t\}}$, where $\mathbf{d}^d = \frac{1}{M_2}\sum_{i=1}^{M_2}\mathbf{d}_i^d$. Text features: $\{\mathbf{z}_k^d\}_{k=1}^K$, extracted from the prompt embeddings of each domain, where $\mathbf{z}_k^d = g^{te}(\mathbf{c}_d^k)$. To leverage the domain knowledge contained in the prompt embeddings, we adopt a data processing procedure inspired by the text encoder in CLIP. This procedure involves disentangling semantic information and domain information at the encoder input and then jointly modeling the two types of information.

Firstly, we use a feature extractor with a multi-branch structure [21] to extract semantic-related features. The feature extractor f consists of a shared backbone and $|S|+1$ feature extractor layers. Each layer is responsible for processing a specific set of domains, denoted as O^j. The domain set O^j (where $j \neq |S|+1$) includes the jth source domain s_j and the target domain t, represented as $O^j = s_j, t$. The domain set $O^{|S|+1}$ of the $(|S|+1)$th layer contains all the domains, given by $O^{|S|+1} = S\cup\{t\}$.

Given an image $\mathbf{x}_i^d$ from the domain d, the feature extracted from the jth branch of the feature extractor is denoted as $\mathbf{h}_i^{d,j} = f^j(\mathbf{x}_i^d)$. We then concatenate $\mathbf{h}_i^{d,j}$ with its corresponding domain prompt embedding $\mathbf{d}^d$ and pass the concatenated feature through an embedding layer h^j. The embedding layer adaptively embeds the concatenated features based on the domain knowledge contained in $\mathbf{d}^d$, effectively disentangling domain-related information from the semantic-related information in $\mathbf{h}_i^{d,j}$.

Based on the resulting embedding feature $\mathbf{e}_i^{d,j} = h^d(\mathbf{h}_i^{d,j}, \mathbf{d}^d)$, we apply a classification layer and utilize cross-entropy loss $\mathcal{L}_{cls}^j$ as our classification loss to optimize the model. Additionally, since the domain-related prompt $\mathbf{d}^d$ is encoded into text features $\{\mathbf{z}_k^d\}_{k=1}^K$, we introduce regularizations to ensure semantic consistency between the text features and the embedding features.

To achieve this, we employ a regularization approach inspired by CLIP, where we classify $\mathbf{e}_i^{d,j}$ through feature comparison. This regularization encourages the embedding features to align with the semantic information captured by the text features, enhancing the overall semantic classification performance. Specifically, with another feature mapping layer m^j, the regularization function is defined as follows:

$$\mathcal{L}_{reg}^j = -\frac{1}{|O^j|}\sum_d^{d\in O^j}\mathbb{E}_{\mathbf{x}\in X^d}\log P^j\left(\hat{y}_i^d = y_i^d\right),$$
$$P^j\left(\hat{y}_i = k \mid \mathbf{x}_i^d\right) = \frac{\exp\left(\left\langle \mathbf{z}_k^d, m^j\left(\mathbf{e}_i^{d,j}\right)\right\rangle/T\right)}{\sum_{r=1}^K \exp\left(\left\langle \mathbf{z}_r^d, m^j\left(\mathbf{e}_i^{d,j}\right)\right\rangle/T\right)}. \tag{7}$$

To further ensure semantic consistency between domains, We minimize the domain distribution discrepancies by confusing a domain discriminator D^j. The loss function to optimize the discriminator is defined as:

$$\mathcal{L}_{dis}^j = \sum_{d\in O^j\setminus t}\mathbb{E}_{\mathbf{x}\in X^d}\left[D^j(\mathbf{h}) - 0\right]^2 + \mathbb{E}_{\mathbf{x}\in X^t}\left[D^j(\mathbf{h}) - 1\right]^2. \tag{8}$$

The confusing loss for the feature extractor is defined as:

$$\mathcal{L}_{con}^j = \sum_{d\in O^j\setminus t}\mathbb{E}_{\mathbf{x}\in X^d}\left[D^j(\mathbf{h}) - \frac{1}{2}\right]^2 + \mathbb{E}_{\mathbf{x}\in X^t}\left[D^j(\mathbf{h}) - \frac{1}{2}\right]^2. \tag{9}$$

Ultimately, we use the following loss function $\mathcal{L}$ to optimize our classification model:

$$\mathcal{L} = \sum_{j=1}^{|S|+1}(\mathcal{L}_{cls}^j + \mathcal{L}_{con}^j + \mathcal{L}_{dis}^j + \alpha\mathcal{L}_{reg}^j), \tag{10}$$

where α is used to control the strength of $\mathcal{L}_{reg}^j$.

During testing, the final prediction for a target domain sample is calculated as the weighted average of the prediction results from each classification branch. The calculation of the weights is similar to that in Eq. (5), taking into account both the prediction certainty and the domain similarity.

4. Experiment and analysis

4.1. Datasets and experimental settings

We evaluate our approach using four experimental benchmarks: Office-Caltech10 [32], Office-31 [33], Office-Home [34], and DomainNet [4]. Office-Caltech10 consists of 10 categories and 4 domains (DSLR (D), Webcam (W), Amazon (A), and Caltech (C)), totaling 2533 images. Office-31 contains 31 categories and 4652 images across 3 domains (DSLR (D), Webcam (W), and Amazon (A)). Office-Home comprises 65 categories and 4 domains (Art (A), Clipart (C), Product (P), and Real world (R)), with 15,500 images featuring common categories such as fork and table. DomainNet includes 569,010 images of 345 categories in 6 domains (Clipart (C), Infograph (I), Painting (P), Quickdraw (Q), Real (R), and Sketch (S)).

For these datasets, we follow the experimental settings of [2,35]. For the prompt learning, we utilize CLIP whose encoder architecture is ResNet-101 [1]. Stochastic gradient descent (SGD) is utilized for optimization of prompts, with a learning rate of 3e-3. Regarding the classification model, we utilize a pre-trained ResNet-101 on ImageNet [36] as the backbone. The feature extractor layer, embedding layer, and mapping layer are all one-layer fully connected layers. For model optimization, we employ SGD with a learning rate of 1e-4. Both the number of prompt tokens, $M1$ and $M2$, are set to 16. The temperature parameter T in Eq. (2) and Eq. (7) is set to 1. Additionally, the pseudo-label threshold τ in Eq. (6) is set to 0.5.

4.2. Comparisons with the state-of-the-art

In general, the compared methods can be categorized into three groups:

Table 1
Comparison with state-of-the-art models on Office Caltech10 dataset.

Standards	Methods	→D	→W	→A	→C	Avg
SC	Source-only	98.3	99.0	86.1	87.8	92.8
	DAN [11]	98.2	99.3	94.8	89.7	95.5
MS	DAN [11]	99.1	99.5	91.6	89.2	94.8
	JAN [12]	99.4	99.4	91.8	91.2	95.5
	DCTN [18]	99.0	99.4	92.7	90.2	95.3
	MCD [6]	99.1	99.5	92.1	91.5	95.6
	M^3SDA [4]	99.2	99.5	94.5	92.2	96.4
	CMSS [3]	99.3	99.6	96.6	93.7	97.2
	STEM [35]	**100**	**100**	98.4	94.2	98.2
	SSG [9]	**100**	**100**	99.0	94.2	98.3
	MLAN [21]	**100**	**100**	99.1	94.7	98.5
	DKBA(ours)	**100**	**100**	**99.2**	**95.1**	**98.6**

Table 2
Comparison with state-of-the-art models on Office-31 dataset.

Standards	Methods	→D	→W	→A	Avg
SB	Source-only	99.0	95.3	50.2	81.5
	DAN [11]	99.0	96.0	54.0	83.0
	RTN [37]	99.6	96.8	51.0	82.5
	ADDA [38]	99.4	95.3	54.6	83.1
SC	DAN [11]	98.8	96.2	54.9	83.3
	RTN [37]	99.2	95.8	53.4	82.8
	JAN [12]	99.4	95.9	54.6	83.3
	ADDA [38]	99.2	96.0	55.9	83.7
	MCD [6]	99.5	96.2	54.4	83.4
MS	MDAN [15]	99.2	95.4	55.2	83.3
	DCTN [18]	99.6	96.9	54.9	83.8
	MDDA [5]	99.2	97.1	56.2	84.2
	LtC-MSDA [2]	99.6	97.2	56.9	84.6
	MOST [39]	**100**	98.7	60.6	86.4
	SSG [9]	**100**	**99.5**	71.3	90.3
	MLAN [21]	99.7	99.0	76.1	91.6
	DKBA(ours)	99.8	99.1	**78.5**	**92.5**

Table 3
Comparison with state-of-the-art models on Office-Home dataset.

Standards	Methods	→A	→C	→P	→R	Avg
SB	Source-only	65.3	49.6	79.7	75.4	67.5
	DAN [11]	68.2	56.5	80.3	75.9	70.2
	CORAL [40]	67.0	53.6	80.3	76.3	69.3
SC	DAN [11]	68.5	59.4	79.0	82.5	72.4
	CORAL [40]	68.1	58.6	79.5	82.7	72.2
MS	MFSAN [41]	72.1	62.0	80.3	81.8	74.1
	SImpAl [8]	70.8	56.3	80.2	81.5	72.2
	DARN [42]	70.0	68.4	82.8	83.9	76.3
	SSG [9]	75.6	68.0	84.2	84.3	78.0
	MLAN [21]	75.3	64.0	84.6	84.8	77.2
	MPA [31]	74.8	54.9	86.2	85.7	75.4
	DKBA(ours)	**76.1**	**68.9**	**90.2**	**87.1**	**80.6**

- Single Best (SB): The methods determine the optimal source domain for achieving the highest performance in adapting to the target domain.
- Source Combine (SC): The methods consider multiple source domains as a single domain for adaptation.
- Multi-Source (MS): The methods utilize multiple source domains for domain adaptation, taking into account the discrepancies among the different source domains (see Tables 1–4).

On the Office-Caltech10 and Office-31 datasets, DKBA demonstrates improvement and achieves state-of-the-art performance, with average accuracies of 98.6% and 92.5% respectively. For Office-31 datasets, Although our method slightly trails the SSG model in the →D and →W tasks, with accuracies of 99.8% and 99.1% respectively, this can be attributed to the smaller domain discrepancies in these tasks. In such settings, the architectural nuances, such as the graph correlation layers used by SSG, may play a more significant role in enhancing model performance. In contrast, for the →A task, which involves a larger domain discrepancy, our method significantly outperforms SSG by achieving an accuracy of 78.5%. This improvement highlights the effectiveness of our approach in managing larger domain shifts, which is critical for practical applications.

Table 4
Comparison with state-of-the-art models on DomainNet dataset.

Standards	Methods	→C	→I	→P	→Q	→R	→S	Avg
SB	Source Only	39.6	8.2	33.9	11.8	41.6	23.1	26.4
	DAN [11]	39.1	11.4	33.3	16.2	42.1	29.7	28.6
	RTN [37]	35.3	10.7	31.7	13.1	40.6	26.5	26.3
	JAN [12]	35.3	9.1	32.5	14.3	43.1	25.7	26.7
	ADDA [38]	39.5	14.5	29.1	14.9	41.9	30.7	28,4
SC	Source Only	47.6	13.0	38.1	13.3	51.9	33.7	32.9
	DAN [11]	45.4	12.8	36.2	15.3	48.6	34.0	32.1
	RTN [37]	44.2	12.6	35.3	14.6	48.4	31.7	31.1
	JAN [12]	40.9	11.1	35.4	12.1	45.8	32.3	29.6
	ADDA [38]	47.5	11.4	36.7	14.7	49.1	33.5	32.2
	MCD [6]	54.3	22.1	45.7	7.6	58.4	43.5	38.5
MS	MDAN [15]	52.4	21.3	46.9	8.6	54.9	46.5	38.4
	M^3SDA [4]	58.6	26.0	52.3	6.3	62.7	49.5	42.6
	MDDA [5]	59.4	23.8	53.2	12.5	61.8	48.6	43.2
	CMSS [3]	64.2	28.0	53.6	16.0	63.4	53.8	46.5
	LtC-MSDA [2]	63.1	28.7	56.1	16.3	66.1	53.8	47.4
	DAEL [43]	70.8	26.5	57.4	12.2	65.0	60.6	48.7
	SSG [9]	68.7	24.8	55.7	18.4	68.8	56.3	48.8
	DCTN [18]	69.6	27.5	57.3	17.8	72.5	55.3	49.8
	DRT [16]	71.0	31.6	61.0	12.3	71.4	60.7	51.3
	MLAN [21]	71.4	29.3	59.5	**28.4**	73.9	58.7	53.5
	MPA [31]	65.2	**47.3**	62.0	10.2	**82.0**	57.9	54.1
	DKBA(ours)	**72.3**	42.1	**63.1**	12.9	80.8	**63.3**	**55.8**

The Office-Home dataset poses a significant challenge due to its substantial domain discrepancy, making it particularly difficult for methods to achieve high accuracy. However, even in this challenging scenario, KBDA maintains a clear advantage, achieving an impressive accuracy of 80.7%. DKBA achieves an average accuracy of 53.8% on six transfer tasks of DomainNet, showcasing the improved performance of STD over existing MSA methods. When compared to MPA, there is a slight decrease in performance for the '→I' and '→R' tasks. These tasks involve images that closely resemble natural scenes, which are precisely the types of images that CLIP excels at recognizing. Consequently, when the target domain primarily consists of such images, fine-tuning the CLIP model instead of using another feature extractor may yield superior results. In addition, for the '→Q' tasks, we observed a noticeable decrease in performance compared to MLAN. This decline can primarily be attributed to the substantial domain gap between the Quickdraw target domain and the other source domains. The unique characteristics of the Quickdraw data, which consists of sketch-based images, differ significantly from the photographic images in the source domains. Consequently, the domain information from the source domains provides limited assistance in enhancing performance on the task. In such scenarios, alternative strategies that improve the quality of pseudo-labels can be more effective in boosting performance. For instance, MLAN employs a source-guided K-means clustering approach to enhance the quality of pseudo-labels in the target domain.

In summary, our observations and conclusions are as follows: Firstly, Source-Combine methods consistently outperform Single-Best methods, indicating that incorporating data from different domains can significantly enhance task performance. Secondly, Multi-Source methods exhibit the highest performance, underscoring the importance of addressing the discrepancies among source domains. Additionally, our experiments included methods like Maximum Classifier Discrepancy (MCD) and SImpAl, which focus on reducing domain shifts while preserving class discriminability. These methods underscore the importance of leveraging discriminative information from multiple domains. Thirdly, methods that align domains separately, such as DCTN, M^3SDA, and MDDA, generally outperform methods that align domains

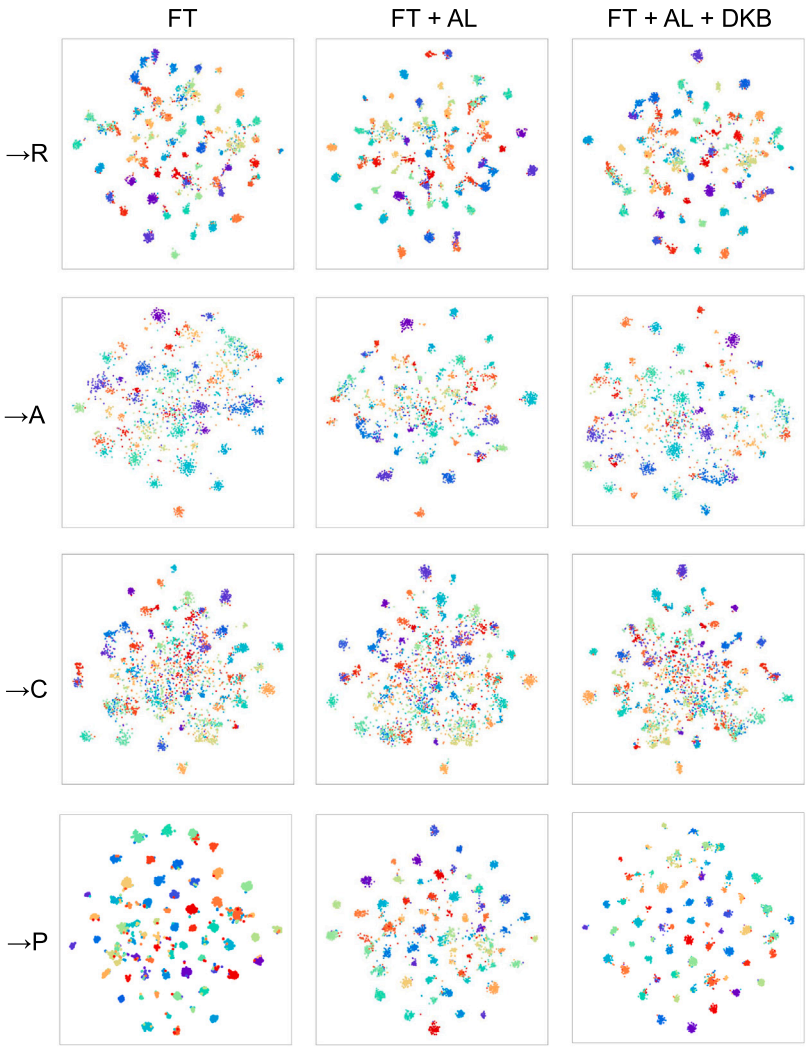

Fig. 3. The t-SNE visualization for the embedding features from FT, FT + AL and FL + AL + DKB on the Office-Home dataset.

jointly, such as MDAN, LtC-MSDA, and DRT. Interestingly, MLAN, which combines both alignment strategies, outperforms both individual approaches. This could be attributed to the fact that having more tunable parameters enhances the model's representation ability for specific tasks. For instance, the state-of-the-art jointly aligning based method DRT achieves improvement primarily through additional source-specific parameters in the backbone. Consequently, since MPA only optimizes a limited number of parameters, its classification performance is inherently limited. Finally, the state-of-the-art performance of KBDA highlights the advantage of its domain arrangement strategy, further emphasizing the importance of effective domain knowledge in achieving superior results.

4.3. Domain knowledge importance evaluation

To showcase the significance of domain knowledge, we assess the impact of different domain-related prompts on the classification performance of CLIP. Our experiments are conducted on the Office-Home dataset, and the results are presented in Table 5. For instance, considering the entry (→A, Clipart), it indicates that in the MSDA task →A, the text features used for classifying the target domain Art are encoded from domain related prompts from the domain Clipart. Analyzing the results, we observe that even without ground truth labels to optimize prompts, employing prompts corresponding to the target domain yields significantly superior classification performance compared to using prompts from other domains. This finding underscores the importance of domain knowledge for accurate target domain classification and validates the rationale behind our research motivation.

4.4. Ablation study

Table 6 provides an analysis of each component of our method on the Office-Home dataset. Firstly, ULSP refers to directly applying CLIP with UnLearnable Semantic related Prompts ("an image of [CLASS]") for classification. Adding Unlearnable Domain related Prompts (ULDP, "an [DOMAIN] class") into ULSP does not lead to performance improvement. However, the performance does improve when utilizing "Learnable Semantic related and Domain related Prompts," which enables CLIP to generate pseudo labels of higher quality. Further performance improvement is achieved by leveraging the pseudo labels from the target domain and the ground truth labels from the source domains to finetune the classification model.

Comparing the performance of finetuning with just Adversarial Learning (AL) to our Domain Knowledge Boosting (DKB) strategy, we observe an average improvement of 0.7%. This demonstrates the effectiveness of DKB. To visually demonstrate the impact of DKB, we employ t-SNE [44] to visualize the features from the $(|S|+1)$th feature extractor layer of FT, (FT + AL) and (FT + AL + DKB). Fig. 3 presents the visualization results for the Office-Home dataset. With our knowledge boosting strategy, the features of the same class exhibit a more pronounced clustering effect, indicating improved discriminability.

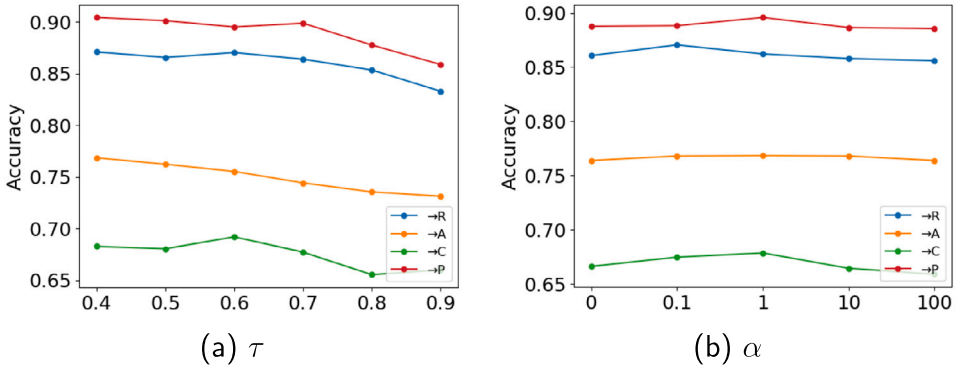

Fig. 4. Sensitivity analysis experiments of τ and α.

Table 5
Domain knowledge importance evaluation on Office-Home dataset.

	→A	→C	→P	→R
Art	**75.4**	68.2	65.2	71.8
Clipart	54.0	**60.6**	51.9	57.5
Product	76.6	77.8	**88.5**	85.6
Real	78.7	78.9	80.1	**86.5**

Table 6
Ablation study on Office-Home dataset.

Method	→A	→C	→P	→R	Avg
ULSP	71.2	55.5	85.8	85.5	74.5
ULSP + ULDP	69.6	53.2	85.5	78.0	71.6
LSP + LDP	75.4	60.6	88.5	86.5	77.8
FT	75.3	67.2	89.0	86.7	79.6
FT + AL	75.6	67.9	89.6	86.9	80.0
FT + AL + DKB	**76.1**	**68.9**	**90.2**	**87.1**	**80.7**

4.5. Sensitivity analysis

In this section, we explore the sensitivity of two hyperparameters: the threshold τ for selecting pseudo labels and the parameter α that controls the strength of text feature regularization. Fig. 4(a) and 4(b) illustrate the sensitivity analysis conducted on the Office-Home dataset. Fig. 4(a) shows that an optimal balance between label quality and data volume yields the best results for the tasks, emphasizing the need for careful selection of the threshold τ to ensure high-quality pseudo labels and maximize data utilization. Fig. 4(b) demonstrates that using text features for appropriate constraints improves the model's performance. However, excessive constraint strength can have a negative impact on task performance, possibly due to differences in feature encoding between CLIP's image encoder and our feature extractor.

5. Conclusion

In this study, we introduced Domain Knowledge Boosted Adaptation (DKBA), a novel approach to multi-source domain adaptation that effectively leverages embedded domain knowledge to enhance classification performance. By innovatively integrating a vision-language model with prompt learning, DKBA capitalizes on the rich semantic and domain-specific information available in diverse data sources. This method marks a significant departure from traditional MSDA techniques, which often fail to fully utilize domain-specific information and may lead to semantic distortion.

The strengths of our approach are evidenced by its superior performance across multiple benchmark datasets, where it consistently outperforms existing state-of-the-art methods. By leveraging domain-specific prompts, DKBA not only preserves but enhances semantic integrity, leading to more robust and domain-invariant feature representations. This is particularly advantageous in complex adaptation scenarios involving significant domain discrepancies.

The implications of our work for the field are substantial. DKBA provides a framework that can be adapted to other domain adaptation challenges, particularly those involving unstructured or loosely labeled data. Our findings underscore the importance of incorporating domain knowledge into the adaptation process, a strategy that could be beneficial across various applications of machine learning. For future work, we plan to refine our approach by incorporating category-level and instance-level domain information. This could potentially lead to even more tailored and precise adaptations, further improving performance. Additionally, exploring alternative architectures and learning strategies that could complement or enhance the prompt-learning mechanism may yield interesting avenues for research. Finally, expanding the applicability of our approach to other forms of data beyond images, such as text or video, could broaden its utility and impact in the field of domain adaptation.

CRediT authorship contribution statement

Yuwei He: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Juexiao Feng:** Writing – original draft, Methodology, Formal analysis. **Guiguang Ding:** Supervision, Funding acquisition, Conceptualization. **Yuchen Guo:** Writing – review & editing, Supervision, Formal analysis. **Tao He:** Writing – review & editing, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the invaluable resources provided by the 15th Research Institute of China Electronics Technology Group Corporation. We also express our appreciation for the support from the National Natural Science Foundation of China (Grant Nos. 61925107 and 62021002), the Zhejiang Provincial Natural Science Foundation of China (Grant No. LDT23F01013F01), the Key R&D Program of Xinjiang, China (Grant No. 2022B01006), and the Postdoctoral Science Foundation, China (Grant No. 2024M750565).

Data availability

Data will be made available on request.

References

[1] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[2] H. Wang, M. Xu, B. Ni, W. Zhang, Learning to combine: Knowledge aggregation for multi-source domain adaptation, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16, Springer, 2020, pp. 727–744.

[3] L. Yang, Y. Balaji, S.-N. Lim, A. Shrivastava, Curriculum manager for source selection in multi-source domain adaptation, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16, Springer, 2020, pp. 608–624.

[4] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, B. Wang, Moment matching for multi-source domain adaptation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1406–1415.

[5] S. Zhao, G. Wang, S. Zhang, Y. Gu, Y. Li, Z. Song, P. Xu, R. Hu, H. Chai, K. Keutzer, Multi-source distilling domain adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, (07) 2020, pp. 12975–12983.

[6] K. Saito, K. Watanabe, Y. Ushiku, T. Harada, Maximum classifier discrepancy for unsupervised domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3723–3732.

[7] X. Chen, S. Wang, M. Long, J. Wang, Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation, in: International Conference on Machine Learning, PMLR, 2019, pp. 1081–1090.

[8] N. Venkat, J.N. Kundu, D. Singh, A. Revanur, et al., Your classifier can secretly suffice multi-source domain adaptation, Adv. Neural Inf. Process. Syst. 33 (2020) 4647–4659.

[9] J. Yuan, F. Hou, Y. Du, Z. Shi, X. Geng, J. Fan, Y. Rui, Self-supervised graph neural network for multi-source domain adaptation, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 3907–3916.

[10] W.K. Wong, Y. Lu, Z. Lai, X. Li, Graph correlated discriminant embedding for multi-source domain adaptation, Pattern Recognit. 153 (2024) 110538.

[11] M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks, in: International Conference on Machine Learning, PMLR, 2015, pp. 97–105.

[12] M. Long, H. Zhu, J. Wang, M.I. Jordan, Deep transfer learning with joint adaptation networks, in: International Conference on Machine Learning, PMLR, 2017, pp. 2208–2217.

[13] L. Hu, M. Kan, S. Shan, X. Chen, Duplex generative adversarial network for unsupervised domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1498–1507.

[14] Z. Pei, Z. Cao, M. Long, J. Wang, Multi-adversarial domain adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, (1) 2018.

[15] H. Zhao, S. Zhang, G. Wu, J.M. Moura, J.P. Costeira, G.J. Gordon, Adversarial multiple source domain adaptation, Adv. Neural Inf. Process. Syst. 31 (2018).

[16] Y. Li, L. Yuan, Y. Chen, P. Wang, N. Vasconcelos, Dynamic transfer for multi-source domain adaptation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10998–11007.

[17] S. Wang, B. Wang, Z. Zhang, A.A. Heidari, H. Chen, Class-aware sample reweighting optimal transport for multi-source domain adaptation, Neurocomputing 523 (2023) 213–223, http://dx.doi.org/10.1016/j.neucom.2022.12.048.

[18] R. Xu, Z. Chen, W. Zuo, J. Yan, L. Lin, Deep cocktail network: Multi-source unsupervised domain adaptation with category shift, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3964–3973.

[19] J. Liu, J. Li, K. Lu, Coupled local–global adaptation for multi-source transfer learning, Neurocomputing 275 (2018) 247–254, http://dx.doi.org/10.1016/j.neucom.2017.06.051.

[20] Y. Lu, H. Huang, B. Zeng, Z. Lai, X. Li, Multi-source and multi-target domain adaptation based on dynamic generator with attention, IEEE Trans. Multimed. (2024).

[21] Y. Xu, M. Kan, S. Shan, X. Chen, Mutual learning of joint and separate domain alignments for multi-source domain adaptation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 1890–1899.

[22] Y. Lu, W.K. Wong, B. Zeng, Z. Lai, X. Li, Guided discrimination and correlation subspace learning for domain adaptation, IEEE Trans. Image Process. 32 (2023) 2017–2032.

[23] J. Huang, N. Xiao, L. Zhang, Balancing transferability and discriminability for unsupervised domain adaptation, IEEE Trans. Neural Netw. Learn. Syst. 35 (4) (2022) 5807–5814.

[24] F. Petroni, T. Rocktäschel, S. Riedel, P. Lewis, A. Bakhtin, Y. Wu, A. Miller, Language models as knowledge bases? in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 2463–2473.

[25] Z. Jiang, F.F. Xu, J. Araki, G. Neubig, How can we know what language models know? Trans. Assoc. Comput. Linguist. 8 (2020) 423–438.

[26] Z. Zhong, D. Friedman, D. Chen, Factual probing is [mask]: Learning vs. Learning to recall, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 5017–5033.

[27] C. Jia, Y. Yang, Y. Xia, Y.-T. Chen, Z. Parekh, H. Pham, Q. Le, Y.-H. Sung, Z. Li, T. Duerig, Scaling up visual and vision-language representation learning with noisy text supervision, in: International Conference on Machine Learning, PMLR, 2021, pp. 4904–4916.

[28] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning transferable visual models from natural language supervision, in: International Conference on Machine Learning, PMLR, 2021, pp. 8748–8763.

[29] K. Zhou, J. Yang, C.C. Loy, Z. Liu, Learning to prompt for vision-language models, Int. J. Comput. Vis. 130 (9) (2022) 2337–2348.

[30] C. Ge, R. Huang, M. Xie, Z. Lai, S. Song, S. Li, G. Huang, Domain adaptation via prompt learning, IEEE Trans. Neural Netw. Learn. Syst. (2023).

[31] H. Chen, X. Han, Z. Wu, Y.-G. Jiang, Multi-prompt alignment for multi-source unsupervised domain adaptation, Adv. Neural Inf. Process. Syst. 36 (2024).

[32] B. Gong, Y. Shi, F. Sha, K. Grauman, Geodesic flow kernel for unsupervised domain adaptation, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 2066–2073.

[33] K. Saenko, B. Kulis, M. Fritz, T. Darrell, Adapting visual category models to new domains, in: Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11, Springer, 2010, pp. 213–226.

[34] H. Venkateswara, J. Eusebio, S. Chakraborty, S. Panchanathan, Deep hashing network for unsupervised domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5018–5027.

[35] V.-A. Nguyen, T. Nguyen, T. Le, Q.H. Tran, D. Phung, Stem: An approach to multi-source domain adaptation with guarantees, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9352–9363.

[36] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee, 2009, pp. 248–255.

[37] M. Long, H. Zhu, J. Wang, M.I. Jordan, Unsupervised domain adaptation with residual transfer networks, Adv. Neural Inf. Process. Syst. 29 (2016).

[38] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7167–7176.

[39] T. Nguyen, T. Le, H. Zhao, Q.H. Tran, T. Nguyen, D. Phung, Most: Multi-source domain adaptation via optimal transport for student-teacher learning, in: Uncertainty in Artificial Intelligence, PMLR, 2021, pp. 225–235.

[40] B. Sun, J. Feng, K. Saenko, Correlation alignment for unsupervised domain adaptation, in: Domain Adaptation in Computer Vision Applications, Springer, 2017, pp. 153–171.

[41] Y. Zhu, F. Zhuang, D. Wang, Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, (01) 2019, pp. 5989–5996.

[42] J. Wen, R. Greiner, D. Schuurmans, Domain aggregation networks for multi-source domain adaptation, in: International Conference on Machine Learning, PMLR, 2020, pp. 10214–10224.

[43] K. Zhou, Y. Yang, Y. Qiao, T. Xiang, Domain adaptive ensemble learning, IEEE Trans. Image Process. 30 (2021) 8008–8018.

[44] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).

Yuwei He received the Ph.D. degree s from the School of Software, Tsinghua University, Beijing, China in 2021, respectively. He was a Post-Doctoral Researcher with the School of Software, Tsinghua University, from 2021 to 2024. He is currently an algorithm researcher with the 15th Research Institute of China Electronics Technology Group Corporation. His current research centers on the area of domain adaptation, medical image analysis, and computer vision.

Juexiao Feng received the B.S. degree from the School of Software, Wuhan University of Technology, Wuhan, China, in 2020. He is currently pursuing the Ph.D. degree in software engineering with Tsinghua University, Beijing. His research interest includes mainly domain adaptation and open world detection.

Guiguang Ding received his Ph.D. degree in electronic engineering from Xidian University, China, in 2014. He is currently a professor of School of Software, Tsinghua University. Before joining school of software in 2006, he has been a postdoctoral research fellow in the Department of Automation, Tsinghua University. He has published over 100 papers in major journals and conferences, including the PR, IEEE TIP, TMM, TKDE, SIG IR, AAAI, ICML, IJCAI, CVPR, and ICCV. His current research centers on the area of multimedia information retrieval, computer vision and machine learning.

Yuchen Guo received the B.Sc. and Ph.D. degrees from the School of Software, Tsinghua University, Beijing, China, in 2018 and 2013, respectively. He was a Post-Doctoral Researcher with the Department of Automation, Tsinghua University, from 2018 to 2020. He is currently an Associate Researcher with the Beijing National Research Center for Information Science and Technology, Tsinghua University. His research interest focuses on brain-inspired artificial intelligence.

Tao He received the Ph.D. degree from the School of Software, Tsinghua University, Beijing, China, in 2023. He is currently a Post-Doctoral Researcher with GRG Banking Equipment Co., Ltd. His current research centers on the area of computer vision, pedestrian re identification, and large language model.