Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Dual collaboration for decentralized multi-source domain adaptation*

Yikang WEI[1,2], Yahong HAN[‡1,2]

[1]*College of Intelligence and Computing, Tianjin University, Tianjin 300350, China*

[2]*Tianjin Key Lab of Machine Learning, Tianjin University, Tianjin 300350, China*

E-mail: yikang@tju.edu.cn; yahong@tju.edu.cn

Received June 29, 2022; Revision accepted Sept. 22, 2022; Crosschecked Nov. 2, 2022

Abstract: The goal of decentralized multi-source domain adaptation is to conduct unsupervised multi-source domain adaptation in a data decentralization scenario. The challenge of data decentralization is that the source domains and target domain lack cross-domain collaboration during training. On the unlabeled target domain, the target model needs to transfer supervision knowledge with the collaboration of source models, while the domain gap will lead to limited adaptation performance from source models. On the labeled source domain, the source model tends to overfit its domain data in the data decentralization scenario, which leads to the negative transfer problem. For these challenges, we propose dual collaboration for decentralized multi-source domain adaptation by training and aggregating the local source models and local target model in collaboration with each other. On the target domain, we train the local target model by distilling supervision knowledge and fully using the unlabeled target domain data to alleviate the domain shift problem with the collaboration of local source models. On the source domain, we regularize the local source models in collaboration with the local target model to overcome the negative transfer problem. This forms a dual collaboration between the decentralized source domains and target domain, which improves the domain adaptation performance under the data decentralization scenario. Extensive experiments indicate that our method outperforms the state-of-the-art methods by a large margin on standard multi-source domain adaptation datasets.

Key words: Multi-source domain adaptation; Data decentralization; Domain shift; Negative transfer
https://doi.org/10.1631/FITEE.2200284 **CLC number:** TP181

1 Introduction

Deep learning methods need labeled data on a large scale to learn knowledge (Han et al., 2021; Yang Y et al., 2021; Gao et al., 2022; Zhao YB et al., 2022) from data. To reduce the cost of labeling data, unsupervised multi-source domain adaptation (UMDA) uses multiple labeled source domains and an unlabeled target domain to train a model to perform well on the unlabeled target domain. The conventional UMDA methods (Xu et al., 2018; Peng et al., 2019; Zhao SC et al., 2019, 2021; Venkat et al., 2020; Li RH et al., 2021; Pu et al., 2021; Zhou et al., 2021) need to access the data from the labeled source domain and unlabeled target domain simultaneously to reduce the domain gap. However, due to data privacy and storage cost concerns, the data from different domains are isolated and cannot be accessed by other domains. This decentralized multi-source domain adaptation (DMDA) (Liang et al., 2020; Feng et al., 2021) scenario presents a great challenge for the conventional UMDA methods.

The existing DMDA methods (Liang et al., 2020; Feng et al., 2021) solve the data

‡ Corresponding author

* Project supported by the National Nature Science Foundation of China (Nos. 61876130 and 61932009) and the Starry Night Science Fund of Zhejiang University Shanghai Institute for Advanced Study

ORCID: Yikang WEI, https://orcid.org/0000-0003-3113-9817; Yahong HAN, https://orcid.org/0000-0003-2768-1398

decentralization challenge and domain shift problem by fine-tuning the source models or distilling supervision knowledge from source models in the unlabeled target domain. The source-free methods (Li R et al., 2020; Liang et al., 2020; Ahmed et al., 2021) pre-train the source model using only the labeled source domain data and then fine-tune the source model on the unlabeled target domain by using pseudo-labels. The source-free training paradigm can transfer knowledge from the source domain to the target domain without accessing the source domain data, but the source-free paradigm leads to limited adaptation performance due to the lack of collaboration between the decentralized source domains and target domain. For example, the pre-trained source models tend to overfit the distribution of source domain data without the collaboration of the target domain, which reduces the transfer performance and leads to the negative transfer problem on the target domain. The other DMDA methods (Peng et al., 2020; Feng et al., 2021; Wu and Gong, 2021) collaboratively train the local source models and local target model on the decentralized domains to obtain a global model that performs well on the target domain. These collaborative methods focus on training the local target model on the decentralized unlabeled target domain by distilling supervision knowledge, for example, the high-confidence pseudo-labels generated by the local source models (Feng et al., 2021; Wu and Gong, 2021), while the domain shift problem reduces the domain adaptation performance by distilling knowledge from the local source models.

There are two obstacles to DMDA. First, the data distribution discrepancy between the source domains and target domain leads to limited adaptation performance on the decentralized target domain. The source models inevitably generate inaccurate pseudo-labels for fine-tuning or distilling knowledge in the unlabeled target domain, which reduces the domain adaptation performance. Alleviating the domain shift problem on the decentralized unlabeled target domain should be considered. Second, the source models tend to overfit their domain data under the data decentralization scenario, which leads to the source models shifting away from the distribution of target domain data. This domain shift problem reduces the transferable performance of the source models and leads to the negative transfer problem when applying the source models to the target domain. A method for overcoming the negative transfer problem in the decentralized source domains should also be considered.

For the above DMDA challenges, we propose a dual collaboration framework for decentralized multi-source domain adaptation (DC-MDA). Under the dual collaboration framework, models from different domains are used as the bridge for reducing the domain gap, while the data from different domains are still kept decentralized and cannot be accessed by other domains. On the decentralized target domain, we propose a domain weighted soft label (DWSL) strategy to generate soft labels by integrating the predictions from multiple local source models according to the confidence of predictions, which can alleviate the inaccurate pseudo-label problem on the target domain. Meanwhile, to alleviate the domain shift problem, we propose consistency regularization (CR) to constrain the consistency of cross-domain prediction for different views of the same sample. Entropy minimization (EM) is used to increase the confidence on the low-confidence target domain data. On the decentralized source domains, we propose model consistency regularization (MCR) to regularize the local source model toward the target domain model during training, which alleviates the negative transfer problem. Then we aggregate the local source models and local target model to obtain a global model, and repeat the training process on the decentralized source domains and target domain. The training procedures on the decentralized source domain with the collaboration of the target domain model and on the decentralized target domain with the collaboration of source domain models form a dual collaboration for reducing the domain gap in a decentralized data scenario.

Overall, the contributions can be summarized as follows:

1. A dual collaboration framework is proposed to reduce the domain gap between the decentralized source domains and target domain.

2. On the decentralized target domain, we propose a DWSL strategy to generate pseudo-labels and improve the discrimination of the target domain model by fully using the unlabeled target domain data.

3. On the decentralized source domains, we regularize local source models toward the target domain

model during training to reduce the domain gap.

4. Experiments have been conducted on standard multi-source domain adaptation datasets, indicating the superior performance of our DC-MDA.

2 Related works

2.1 Unsupervised domain adaptation

The conventional unsupervised domain adaptation (UDA) methods aim to transfer knowledge from the labeled source domain to the unlabeled target domain. The existing UDA methods focus mainly on the single source domain scenario (Ganin and Lempitsky, 2015; Long et al., 2015; Li S et al., 2018). However, in the real world, there may exist multiple source domains related to the target domain, which can be used to conduct domain adaptation from multiple source domains to a single target domain.

The existing UMDA methods conduct domain alignment between the target domain and multiple source domains by adversarial learning based methods (Zhao H et al., 2018; Zhao SC et al., 2020) or discrepancy-based methods (Xu et al., 2018; Peng et al., 2019). Other UMDA methods conduct domain alignment by graph modeling (Wang H et al., 2020), a sample selection strategy (Yang LY et al., 2020), or CR methods (Zhou et al., 2021). Different from the single source domain adaptation scenario, the discrepancies of different source domains should be considered to conduct domain alignment in the multi-source scenario (Li RH et al., 2021; Nguyen et al., 2021; Zuo et al., 2021).

The conventional UMDA methods usually need to access source domain data and target domain data simultaneously. For the adversarial learning based methods, the adversarial training between source domain data and target domain data is conducted for domain alignment. For the discrepancy-based methods, MMD alignment is conducted on the source domain data and target domain data to reduce the distribution discrepancy. While considering the data privacy and storage cost, the data from different domains are decentralized and cannot be exchanged across domains. We focus on the more challenging DMDA scenario.

2.2 Source-free methods

Source-free methods (Kundu et al., 2020, 2021; Li R et al., 2020; Liang et al., 2020; Ahmed et al., 2021; Liu et al., 2021; Qiu et al., 2021; Xia et al., 2021) assume that source domain data cannot be accessed on the target domain, while the well-trained source domain model can be accessed on the target domain to conduct domain adaptation. The source-free methods solve the data decentralization challenge using pre-training on the source domain and then fine-tuning on the target domain. The representative work of source-free methods is Source HypOthes Transfer (SHOT) (Liang et al., 2020), which implicitly aligns representations of target domain data to the source hypothesis by freezing the classifier of the source model and fine-tuning the feature extractor on the target domain.

Most source-free methods focus on the single-source scenario, which ignores the difference in multiple source domains in the multi-source scenario. Recently, Data frEe multi-sourCe unsupervISed domaIn adaptatiON (DECISION) (Ahmed et al., 2021) improves SHOT in the multi-source scenario. DECISION combines the predictions and features from different source models with the learnable domain weights to obtain better performance.

The source-free methods solve the data decentralization challenge by pre-training on the source domain and then fine-tuning on the target domain. This approach lacks collaboration between the decentralized source and target domains during training. Different from the source-free methods, our dual collaboration framework can reduce the domain gap between the decentralized source domains and target domain in a collaborative manner.

2.3 Unsupervised federated domain adaptation methods

Unsupervised federated domain adaptation (UFDA) (Peng et al., 2020) aims to solve the DMDA problem using the federated learning (FL) paradigm (McMahan et al., 2017; Yang Q et al., 2020). The FL paradigm has been well explored with respect to the data decentralization challenge, which assumes that decentralized data cannot leave their local clients while the models from different clients can be exchanged. There are some works (Peng et al., 2020; Feng et al., 2021; Wu and Gong, 2021;

Wang B et al., 2022) that explore the DMDA problem under the FL paradigm. Federated adversarial domain adaptation (FADA) (Peng et al., 2020) solves the DMDA problem by federated adversarial alignment and representation disentanglement. Collaborative optimization and aggregation (COPA) (Wu and Gong, 2021) solves the DMDA problem by learning the domain-invariant representation and using hybrid batch-instance normalization to improve the generalization. Self-supervised federated domain adaptation (SFDA) (Wang B et al., 2022) uses the similarity between the source and the target domain class centroids to generate pseudo-labels for the unlabeled target domain.

Knowledge distillation based decentralized domain adaptation (KD3A) (Feng et al., 2021) is the representative work that solves the DMDA problem by generating a pseudo-label on the unlabeled target domain and aggregating the local models with domain weights. The knowledge vote (KV) generates a high-quality pseudo-label by voting to avoid inaccurate labels. The consensus focus filters the irrelevant and malicious source domains by allocating domain weights according to the contributions of different domains. KD3A provides a framework for solving the DMDA problem under the FL paradigm by iteratively training and aggregating local feature extractors from different domains and keeping the classifiers from different domains specifically.

Different from the source-free methods, the FL-based methods can collaboratively train the isolated source domain data and target domain data. The data from different domains are kept isolated. However, the models from different domains can be exchanged across domains to reduce the domain gap in the data decentralization scenario. Our method follows the UFDA setup. Different from the existing UFDA methods, our dual collaboration method trains the local target model with the collaboration of local source models to distill supervision knowledge, and trains the local source models with the collaboration of the local target model to overcome the negative transfer problem. In the model aggregation stage, the feature extractors and classifiers from different domains are simultaneously aggregated to reduce the model discrepancy between the source domains and target domain.

3 Methodology

3.1 Problem setting and definition

We focus on the DMDA setup, which assumes that the decentralized data from different domains cannot be exchanged, while the models trained on the decentralized domains can be exchanged for domain adaptation. In this work, we consider the general K-way image classification task. There are m labeled source domains $\{\mathcal{L}_{S_j}\}_{j=1}^{m}$ and an unlabeled target domain $\mathcal{U}_{\mathrm{T}} = \{\boldsymbol{x}_{\mathrm{T}}^{i}\}_{i=1}^{N_{\mathrm{T}}}$, where $\mathcal{L}_{S_j} = \{\boldsymbol{x}_{S_j}^{i}, y_{S_j}^{i}\}_{i=1}^{N_{S_j}}$ with N_{S_j} samples. The $\boldsymbol{x}_{S_j}^{i} \in \mathcal{X}_{S_j}$ and $y_{S_j}^{i} \in \mathbb{R}^{K}$ denote the i^{th} image and the corresponding label from the S_j^{th} source domain, respectively. The $\boldsymbol{x}_{\mathrm{T}}^{i} \in \mathcal{X}_{\mathrm{T}}$ denotes the i^{th} unlabeled target domain image. The purpose of DMDA is to obtain a model that performs well on the unlabeled target domain under the data decentralization scenario. Under our DC-MDA framework, each source domain and target domain leverages its domain data to train a local model. The local model can be denoted as $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ for the m source domains and $\mathcal{M}_{\mathrm{T}}$ for the target domain. Each local model is composed of a feature extractor g and a classifier f. After aggregating local source models $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ and local target model $\mathcal{M}_{\mathrm{T}}$, we obtain a global model $\mathcal{M}_{\mathrm{G}}$.

3.2 Framework

The overall framework of our method is shown in Fig. 1. For a training round, we train the local source model $\mathcal{M}_{S_j}$ on the source domain by cross-entropy loss ℓ_{ce} and the MCR loss ℓ_{mcr} regularized by local target model $\mathcal{M}_{\mathrm{T}}$ in step 1. Then we upload the trained local source models $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ to the target domain. On the target domain, we distill knowledge from $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ using the loss ℓ_{dwsl} and increase the discrimination of the local target model $\mathcal{M}_{\mathrm{T}}$ using CR loss ℓ_{cr} on high-confidence data and EM loss ℓ_{em} on low-confidence data in step 2. Then, we aggregate $\mathcal{M}_{\mathrm{T}}$ and $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ to obtain a global model $\mathcal{M}_{\mathrm{G}}$ in step 3. The global model $\mathcal{M}_{\mathrm{G}}$ is uploaded to the source domains and target domain to initialize the local source models $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ and local target model $\mathcal{M}_{\mathrm{T}}$ to conduct the next round collaborative training. The local target model $\mathcal{M}_{\mathrm{T}}$ is also uploaded to the source domains to regularize the local source models $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ on the next round. In this way, the decentralized data from different

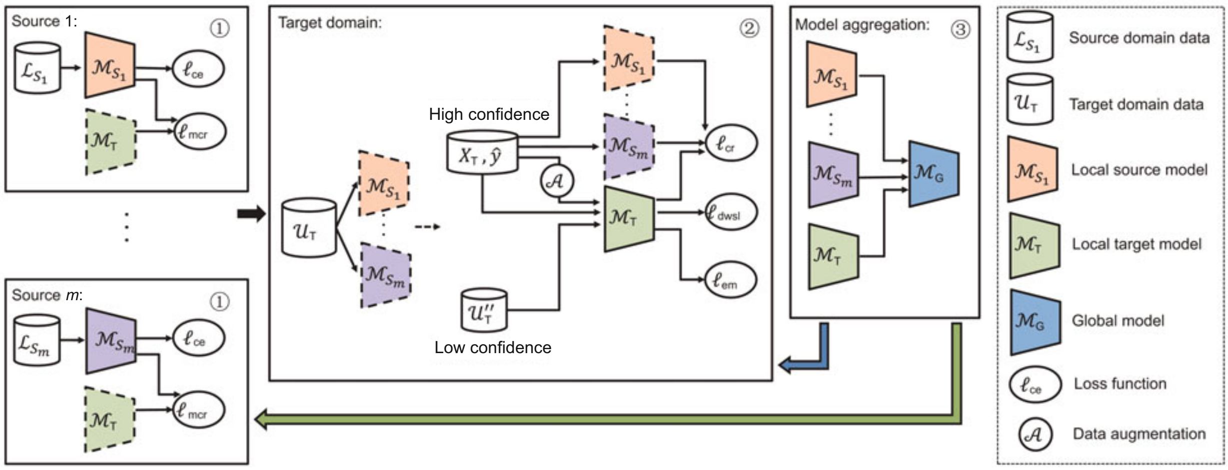

Fig. 1 The framework of dual collaboration decentralized multi-source domain adaptation (DC-MDA) for solving decentralized multi-source domain adaptation (DMDA)

There are three steps in a round. In step 1, we train the local source models $\mathcal{M}_{S_j}$ on source domain $\mathcal{L}_{S_j}$ with the collaboration of local target model $\mathcal{M}_\mathrm{T}$. Then the local source models $\{\mathcal{M}_{S_j}\}_{j=1}^m$ are uploaded to the target domain. In step 2, we train the local target model $\mathcal{M}_\mathrm{T}$ on target domain $\mathcal{U}_\mathrm{T}$ with the collaboration of local source models $\{\mathcal{M}_{S_j}\}_{j=1}^m$. In step 3, we aggregate the local source models $\{\mathcal{M}_{S_j}\}_{j=1}^m$ and local target model $\mathcal{M}_\mathrm{T}$ to obtain global model $\mathcal{M}_\mathrm{G}$. Then, global model $\mathcal{M}_\mathrm{G}$ is uploaded to the source domains and target domain, and is used to initialize the local source models and local target model. The local target domain model $\mathcal{M}_\mathrm{T}$ is uploaded to source domains to regularize the training of local source models on the next round. On the next round, the above steps 1–3 are repeated

domains remain in their domain and the trained local models are used as bridges to conduct domain adaptation. During the first round, the source domain lacks a local target domain model, and we train only local source domain models using cross-entropy loss ℓ_{ce}. The details of local target model training and local source model training are shown in the following subsections.

3.3 Local target domain model training

3.3.1 Domain weighted soft label

For training the local target model on the decentralized target domain, we propose a novel DWSL strategy to generate soft labels for unlabeled data with the collaboration of local source models.

Different strategies can be used to generate pseudo-labels for the unlabeled target domain data using the local source models. The first strategy uses the max prediction p_{i^*} on the unlabeled target domain data $\boldsymbol{x}_\mathrm{T}^i$ from local source models as the pseudo-label $\hat{y}_i = \arg\max(p_{i^*})$. Due to the domain shift, this max strategy will lead to inaccurate pseudo-labels on the target domain. The other strategy is to average the predictions from different local source models. This mean strategy can capture the distribution of different classes to overcome the inaccurate pseudo-label problem, but leads to class confusion due to smoothing class distribution. The vote strategy, such as the KV used in KD3A, uses the most frequent prediction as the pseudo-labels. However, the above pseudo-label strategies ignore the discrepancy between source domains and the target domain.

To use the complementary predictions from different local source models, we average the predictions from different local source models by different domain weights $\{\alpha_{S_j}\}_{j=1}^m$ to obtain a weighted prediction:

$$\hat{p}_i = \frac{1}{m}\sum_{j=1}^{m}\alpha_{S_j}\mathcal{M}_{S_j}(\boldsymbol{x}_\mathrm{T}^i). \tag{1}$$

However, the weights of different models may be hard to determine under the data decentralization scenario. The prediction entropy indicates the prediction confidence for the unlabeled target domain data $\boldsymbol{x}_\mathrm{T}^i$. So, we calculate the source domain weights $\{\alpha_{S_j}\}_{j=1}^m$ according to the entropy $\{h_{S_j}\}_{j=1}^m$ from multiple local source models:

$$\alpha_{S_j} = \frac{\exp(1/h_{S_j})}{\sum_{k=1}^m \exp(1/h_{S_k})}, \tag{2}$$

where $h_{S_j} = -\delta(\mathcal{M}_{S_j}(\boldsymbol{x}_{\mathrm{T}}^i)) \ln(\delta(\mathcal{M}_{S_j}(\boldsymbol{x}_{\mathrm{T}}^i)))$.

The weighted prediction leads to class distribution confusion due to averaging. To reduce class confusion, we sharpen the weighted prediction to obtain the final soft label:

$$\hat{y}_{\mathrm{T}}^i = \frac{\hat{p}_i/\tau}{\sum_{k=1}^{K} \hat{p}_i^k/\tau}, \tag{3}$$

where the temperature parameter τ sharpens the domain weighted mean prediction when $\tau < 1.0$.

Then we use the sharpened DWSL, whose largest class probability $\hat{y}_{\mathrm{T}}^{i*}$ is larger than a threshold T as the high-confidence soft label. The threshold can filter the low-confidence pseudo-labels.

Once the high-confidence soft pseudo-labels are obtained by the DWSL strategy, we can train the local target model using cross-entropy:

$$\ell_{\mathrm{dwsl}} = -E_{\boldsymbol{x}_{\mathrm{T}}^i \in \mathcal{U}_{\mathrm{T}}}[\mathbb{I}(\hat{y}_{\mathrm{T}}^{i*} \geq T)\hat{y}_T^{i*} \ln(\delta(\mathcal{M}_{\mathrm{T}}(\boldsymbol{x}_{\mathrm{T}}^i)))], \tag{4}$$

where $\delta(\cdot)$ denotes the softmax function and $\mathbb{I}$ denotes the indicator function.

3.3.2 Consistency regularization

To improve the discrimination of the local target model, we use CR on the target domain. CR comes from the cluster assumption, which requires that the output of the classifier should be consistent between input images and their perturbed versions (Miyato et al., 2018).

The cluster assumption can cause the decision boundaries to be far away from the high-density data regions, which leads to better discrimination in the target domain and alleviates the domain shift problem. We use the strong image augmentations RandAugment (Cubuk et al., 2020) to generate the perturbed version of the unlabeled target domain image $\boldsymbol{x}_{\mathrm{T}}^i$ as $\mathcal{A}(\boldsymbol{x}_{\mathrm{T}}^i)$. Then we use the square of the ℓ_2 norm to ensure the consistency between the prediction from the online local target model on $\mathcal{A}(\boldsymbol{x}_{\mathrm{T}}^i)$ and the offline local source models on $\boldsymbol{x}_{\mathrm{T}}^i$. The CR loss is as follows:

$$\begin{aligned}&\ell_{\mathrm{cr}} = \\ &- E_{\boldsymbol{x}_{\mathrm{T}}^i \in \mathcal{U}_{\mathrm{T}}}[\mathbb{I}(\hat{y}_{\mathrm{T}}^{i*} \geq T)\left\|\hat{y}_{\mathrm{T}}^i - \delta(\mathcal{M}_{\mathrm{T}}(\mathcal{A}(\boldsymbol{x}_{\mathrm{T}}^i)))\right\|_2^2].\end{aligned} \tag{5}$$

3.3.3 Entropy minimization

For the low-confidence unlabeled target domain data, whose largest class probability $\hat{y}_{\mathrm{T}}^{i*}$ is lower than a threshold T, we use EM (Saito et al., 2019) to make sure that the local target model is confident on the prediction:

$$\begin{aligned}&\ell_{\mathrm{em}} = \\ &- E_{\boldsymbol{x}_{\mathrm{T}}^i \in \mathcal{U}_{\mathrm{T}}}[\mathbb{I}(\hat{y}_{\mathrm{T}}^{i*} < T)\delta(\mathcal{M}_{\mathrm{T}}(\boldsymbol{x}_{\mathrm{T}}^i)) \ln(\delta(\mathcal{M}_{\mathrm{T}}(\boldsymbol{x}_{\mathrm{T}}^i)))].\end{aligned} \tag{6}$$

For the target domain, the overall loss during local target domain training can be shown as follows:

$$\ell_{\mathrm{tar}} = \ell_{\mathrm{dwsl}} + \ell_{\mathrm{cr}} + \ell_{\mathrm{em}}. \tag{7}$$

On the isolated target domain, the supervised knowledge from the source domains is transferred to the target domain by pseudo-labels generated by the source models and the consistency between the source models and the target model, which can reduce the domain gap between the labeled source domains and unlabeled target domain.

3.4 Local source domain model training

For the j^{th} source domain, there are ground truth labels for training, so the cross-entropy loss is adopted for training the local source model:

$$\ell_{\mathrm{ce}} = -E_{\boldsymbol{x}_{S_j}^i \in \mathcal{L}_{S_j}}[y_{S_j}^i \ln(\delta(\mathcal{M}_{S_j}(\boldsymbol{x}_{S_j}^i)))]. \tag{8}$$

To solve the overfitting problem during local source model training, we use the local target model to regularize the local source model. MCR is proposed as follows:

$$\ell_{\mathrm{mcr}} = \left\|g_{S_j} - g_{\mathrm{T}}\right\|_2^2, \tag{9}$$

where we use the square of the ℓ_2 norm to regularize the local source domain feature extractor g_{S_j}, and do not shift away from the local target feature extractor g_{T} too much to relieve the negative transfer problem.

Detailed analysis of the negative transfer problem on the decentralized source domains and the effect of model regularization is shown in the experiments section.

For the source domain, the overall training loss can be shown as follows:

$$\ell_{\mathrm{src}} = \ell_{\mathrm{ce}} + \lambda\ell_{\mathrm{mcr}}, \tag{10}$$

where λ is a hyper-parameter to control the ratio of ℓ_{mcr}.

On the isolated source domain, the source model is regularized to the target domain model, which can reduce the domain gap between the source domain and target domain on the model level.

3.5 Model aggregation

After the local training, e.g., 1 epoch, the local source models $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ and the local target model $\mathcal{M}_{\mathrm{T}}$ are aggregated by parameter averaging:

$$\mathcal{M}_{\mathrm{G}}=\frac{1}{m+1}\left(\sum_{j=1}^{m}\frac{N_{S_j}}{N_{\mathrm{tot}}}\mathcal{M}_{S_j}+\frac{N_{\mathrm{T}}}{N_{\mathrm{tot}}}\mathcal{M}_{\mathrm{T}}\right), \quad (11)$$

where $N_{\mathrm{tot}}=\sum_{j=1}^{m}N_{S_j}+N_{\mathrm{T}}$ denotes the sum of the source and target domain samples.

In the next round, the global model $\mathcal{M}_{\mathrm{G}}$ is used to initialize the local models $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ on the source domains and $\mathcal{M}_{\mathrm{T}}$ on the target domain. In the next round, the local training on the source domains and target domain will be repeated.

The overall algorithm that trains E rounds is shown in Algorithm 1.

4 Experiments

4.1 Datasets

We conduct experiments on the standard multi-source domain adaptation datasets including Digit-5, Office31, Office-Caltech10, and DomainNet.

1. Digit-5 is a dataset with digit images of 10 classes from five domains: MNIST (mt), USPS (up), MNIST-M (mm), SVHN (sv), and Synthetic (sy). Each domain contains 25 000 images for training and 9000 for testing. USPS uses 29 752 images for training and 1860 for testing.

2. Office31 contains 4110 images of 31 classes from three domains: Amazon (A), Webcam (W), and DSLR (D). We split the training set and testing set by the ratio of 8:2 on each class.

3. Office-Caltech10 contains 2533 images of 10 classes from four domains: Amazon (A), Webcam (W), DSLR (D), and Caltech-256 (C). We split the training set and testing set by the ratio 8:2 following KD3A (Feng et al., 2021).

4. DomainNet is the largest real-world domain adaptation dataset with 0.6 million images of 345 classes from six domains: Clipart (clp), Infograph (inf), Painting (pnt), Quickdraw (qdr), Real (rel), and Sketch (skt). The partition of the training set and testing set follows the common approach (Peng et al., 2019).

Algorithm 1 DC-MDA method

Input: Source domain data $\{\mathcal{L}_{S_j}\}_{j=1}^{m}$ and target domain data $\mathcal{U}_{\mathrm{T}}$
Output: Global model $\mathcal{M}_{\mathrm{G}}$
1: Initialize the local models with the same initial parameters
2: **for** t=1 to E **do**
3: **for** j=1 to m **do**
4: Train local source model $\mathcal{M}_{S_j}$ with loss ℓ_{src} (Eq. (10)) on source domain $\mathcal{L}_{S_j}$
5: **end for**
6: Upload local source models $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ to the target domain
7: Train $\mathcal{M}_{\mathrm{T}}$ with loss ℓ_{tar} (Eq. (7)) on target domain $\mathcal{U}_{\mathrm{T}}$
8: Aggregate $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ and $\mathcal{M}_{\mathrm{T}}$ to obtain $\mathcal{M}_{\mathrm{G}}$ with Eq. (11) // Model aggregation
9: Upload global model $\mathcal{M}_{\mathrm{G}}$ to the source domains and target domain to initialize the local models $\{\mathcal{M}_{S_j}\}_{j=1}^{m}$ and $\mathcal{M}_{\mathrm{T}}$, and upload the local target model $\mathcal{M}_{\mathrm{T}}$ to source domains
10: **end for**
11: **return** $\mathcal{M}_{\mathrm{G}}$

4.2 Baseline methods

In our experiments, we compare our method with the state-of-the-art source-free methods and UFDA methods in the DMDA setup. The source-free methods include SHOT (Liang et al., 2020) and DECISION (Ahmed et al., 2021). The UFDA methods include FADA (Peng et al., 2020), SFDA (Wang B et al., 2022), COPA (Wu and Gong, 2021), and KD3A (Feng et al., 2021). The results of conventional UMDA methods in the UMDA setup are also reported, including the adversarial learning based methods DANN (Ganin and Lempitsky, 2015), MDAN (Zhao H et al., 2018), MDDA (Zhao SC et al., 2020), and DCTN (Xu et al., 2018). The discrepancy-based methods are DAN (Long et al., 2015), M3SDA (Peng et al., 2019), and ABMSDA (Zuo et al., 2021). The sample selection method is CMSS (Yang LY et al., 2020), and the other state-of-the-art methods are DAEL (Zhou et al., 2021), LtC-MSDA (Wang H et al., 2020), T-SVDNet (Li RH et al., 2021), and STEM (Nguyen et al., 2021).

4.3 Implementation details

Following the previous setup (Feng et al., 2021), we use a three-layer convolutional neural network (CNN) as the backbone for the Digit-5 dataset. The local model is optimized by the stochastic gradient descent (SGD) optimizer with 0.9 momentum, and the cosine schedule is used to decay the learning rate from 0.05 to 0 during training. The batch size is 128, the confidence threshold T is 0.9, and the temperature parameter τ is 0.8.

For the Office31 dataset, we use ResNet50 pre-trained on ImageNet as the backbone. For the Office-Caltech10 and DomainNet datasets, we use ResNet101 pre-trained on ImageNet as the backbone. The batch size is 32 for Office31, 16 for Office-Caltech10, and 40 for DomainNet. SGD is used as an optimizer with 0.9 momentum and the cosine schedule is used to decay the learning rate from 0.01 to 0. The confidence threshold T used in DWSL is 0.9 and the temperature parameter τ is 0.4.

On the Digit-5, Office31, Office-Caltech10, and DomainNet datasets, the λ of ℓ_{mcr} in ℓ_{src} is 0.001. We train the local model 1 epoch on their domain and then aggregate these local models to conduct the next round of training. The total training round E is 100.

In our experiments, we choose one domain as an unlabeled target domain and the remaining domains as the labeled source domains. Each experiment is repeated five times on different random seeds. The mean accuracy and standard deviations are reported for the experiments. We use PyTorch to implement our method and conduct the experiments with 6× NVIDIA RTX 3090 GPU.

4.4 Digit classification

The results on the Digit-5 dataset are shown in Table 1. For the easy domain, e.g., MNIST, the source-only method, which directly uses the model trained on the source domains, can achieve 92.3% accuracy, while for the hard domains, e.g., MNIST-M and SVHN, the source-only method achieves only 63.7% accuracy on MNIST-M and 71.5% on SVHN due to the domain gap. The conventional UMDA methods, such as the adversarial learning based methods DANN and MDDA and the discrepancy-based methods DAN and M3SDA, conduct domain alignment to reduce the domain gap and achieve better performance. The state-of-the-art methods, e.g., ABMSDA, T-SVDNet, and STEM, consider the contributions of different source domains and achieve better performance, which indicates that combining different source domains with domain weights is important. However, these methods need to access the source domain data and target domain data simultaneously and cannot work under the data decentralization scenario. Our method even outperforms these methods by overcoming domain shift and data decentralized challenges simultaneously.

For source-free methods, SHOT and DECISION achieve 90.2% and 94.0% average accuracy by fine-tuning the source domain model on the unlabeled target domain. SHOT treats the multi-source domain scenario as single source domain adaptation by simply averaging the predictions from multiple models transferring from a single source domain to the target domain. DECISION achieves 94% accuracy by weighting the predictions from different source models, which indicates the importance of distinguishing the contributions of different source models, while source-free methods cannot collaboratively train source and target models. Our method outperforms DECISION by collaboratively training the source domains and target domain.

For the UFDA methods, KD3A achieves 92.0% accuracy by using knowledge distillation on the target domain and weighted aggregation. Different from SFDA, which uses class centroids to generate pseudo-labels, our method uses high-confidence pseudo-labels to train the target domain model to overcome the inaccurate pseudo-label problem. Different from KD3A, we use DWSL to obtain the weighted soft label and fully use the unlabeled target domain data to improve the specific performance of the local target domain model. Simultaneously, local source domain models are regularized by the local target domain model to overcome the negative transfer problem, which will be analyzed in detail in the ablation study section. So, as shown in Table 1, our method achieves 97.3% average accuracy, outperforming the existing DMDA methods, e.g., FADA, SFDA, COPA, and KD3A, by a large margin. Our method even outperforms the conventional UMDA methods, which can access the source domain data and target domain data simultaneously.

4.5 Object recognition

The results on the Office31 and Office-Caltech10 datasets are shown in Tables 2 and 3, respectively. As we can see, the source-only method achieves 80.2% average accuracy on Office31 and 92.8% on Office-Caltech10. The conventional UMDA methods, e.g., M3SDA, DCTN, and LtC-MSDA, achieve better performance on the target domain compared with the source-only method by aligning

Table 1 Classification accuracy (mean±std) on the Digit-5 dataset

Method	Classification accuracy (%)					Avg (%)
	mt	mm	sv	sy	up	
Source-only	92.3±0.9	63.7±0.8	71.5±0.8	83.4±0.8	90.7±0.5	80.3
DAN (Long et al., 2015)	96.3±0.5	63.8±0.7	62.5±0.7	85.4±0.8	94.2±0.9	80.4
DANN (Ganin and Lempitsky, 2015)	97.6±0.8	71.3±0.6	63.5±0.8	85.3±0.8	92.3±0.9	82.0
DCTN (Xu et al., 2018)	96.2±0.8	70.5±1.2	77.6±0.4	86.8±0.8	92.8±0.3	84.8
M3SDA (Peng et al., 2019)	98.4±0.7	72.8±1.1	81.3±0.9	89.6±0.6	96.1±0.8	87.7
MDAN (Zhao H et al., 2018)	97.2±1.0	75.7±0.8	82.2±0.8	85.2±0.6	93.3±0.5	86.7
MDDA (Zhao SC et al., 2020)	98.8±0.4	78.6±0.6	79.3±0.8	89.7±0.7	93.9±0.5	88.1
CMSS (Yang LY et al., 2020)	99.0±0.1	75.3±0.6	88.4±0.6	93.7±0.2	97.7±0.1	90.8
ABMSDA (Zuo et al., 2021)	99.3±0.1	73.4±0.7	88.2±0.5	97.7±0.2	97.1±0.9	91.1
LtC-MSDA (Wang H et al., 2020)	99.0±0.4	85.6±0.8	83.2±0.6	93.0±0.5	98.3±0.4	91.8
T-SVDNet (Li RH et al., 2021)	99.3±0.1	91.2±0.7	84.9±1.5	95.7±0.3	98.6±0.2	93.9
STEM (Nguyen et al., 2021)	99.4	89.7	89.9	97.5	98.4	95.0
DAEL (Zhou et al., 2021)	99.5±0.0	93.8±0.1	92.5±0.2	**97.9±0.0**	98.7±0.8	96.5
SHOT (Liang et al., 2020)	98.2±0.4	80.2±0.4	84.5±0.3	91.1±0.2	97.1±0.3	90.2
DECISION (Ahmed et al., 2021)	99.2	93.0	82.6	97.5	97.8	94.0
FADA (Peng et al., 2020)	91.4±0.7	62.5±0.7	50.5±0.3	71.8±0.5	91.7±1.0	73.6
SFDA (Wang B et al., 2022)	99.1	72.3	86.0	90.4	98.1	89.2
COPA (Wu and Gong, 2021)	99.4	89.8	91.0	97.5	**99.2**	95.4
KD3A (Feng et al., 2021)	99.2±0.1	87.3±0.2	85.6±0.2	89.4±0.3	98.5±0.3	92.0
DC-MDA	**99.6±0.0**	**97.1±0.2**	**93.2±1.1**	97.6±0.4	98.8±0.1	**97.3**

The best results are in bold

Table 2 Classification accuracy (mean) on the Office31 dataset

Method	Accuracy (%)			Avg (%)
	A	D	W	
Source-only	51.6	92.0	97.1	80.2
DCTN (Xu et al., 2018)	54.9	99.6	96.9	83.8
M3SDA (Peng et al., 2019)	55.4	99.4	96.2	83.7
MDAN (Zhao H et al., 2018)	55.2	99.2	95.4	83.3
MDDA (Zhao SC et al., 2020)	56.2	99.2	97.2	84.2
LtC-MSDA (Wang H et al., 2020)	56.9	99.6	97.2	84.6
SHOT (Liang et al., 2020)	75.0	97.8	94.9	89.3
DECISION (Ahmed et al., 2021)	75.4	99.6	98.4	91.1
KD3A (Feng et al., 2021)	71.9	99.8	98.7	90.1
DC-MDA	**75.8**	**100**	**100**	**91.9**

The best results are in bold

Table 3 Classification accuracy (mean) on the Office-Caltech10 dataset

Method	Accuracy (%)				Avg (%)
	A	C	D	W	
Source-only	86.1	87.8	98.3	99.0	92.8
DCTN (Xu et al., 2018)	92.7	90.2	99.0	99.4	95.3
M3SDA (Peng et al., 2019)	94.5	92.2	99.2	99.5	96.4
CMSS (Yang LY et al., 2020)	96.6	93.7	99.3	99.6	97.2
STEM (Nguyen et al., 2021)	**98.4**	94.2	**100**	**100**	98.2
SHOT (Liang et al., 2020)	96.4	96.2	98.5	99.7	97.7
DECISION (Ahmed et al., 2021)	95.9	95.9	**100**	99.6	98.0
FADA (Peng et al., 2020)	84.2	88.7	87.1	88.1	87.1
COPA (Wu and Gong, 2021)	95.8	94.6	99.6	99.8	97.5
KD3A (Feng et al., 2021)	97.4	96.4	98.4	99.7	97.9
DC-MDA	96.9	**97.2**	**100**	**100**	**98.5**

The best results are in bold

features from the source domain and target domain, which indicates that reducing the domain gap will greatly improve the performance on the target domain. Meanwhile, STEM achieves 98.2% average accuracy on Office-Caltech10 by considering the contribution of each source domain, which indicates the importance of using domain weights.

In the data decentralization scenario, our method consistently outperforms the source-free methods and UFDA methods. On the hard domains such as Amazon (A) on the Office31 dataset and Caltech-256 (C) on the Office-Caltech10 dataset, our method outperforms DECISION due to the advantage of dual collaborative training. Meanwhile, our method outperforms the UFDA methods KD3A and COPA because we fully explore the unlabeled target domain under domain shift to improve the discrimination of the target domain model and overcome the negative transfer problem on the source domain. On the easy domains, such as DSLR (D) and Webcam (W) on the Office31 and Office-Caltech10 datasets, our method even achieves 100% accuracy. Compared with other DMDA methods, our method achieves the best accuracy of 91.9% on Office31 and 98.5% on Office-Caltech10, which indicates the effect of our dual collaboration framework in the DMDA setup.

4.6 Results on DomainNet

The results on the DomainNet dataset are shown in Table 4. The source-only method leads to poor performance on the target domain, especially in the hard domain Quickdraw, due to the huge domain gap between the source domains and target domain. For the conventional UMDA methods, e.g., M3SDA and DCTN, domain alignment is used to reduce domain shift, while the results are slightly better or even worse than those of the source-only method. This negative transfer problem may be caused by ignoring the discrepancies of different source domains. ABMSDA achieves 47.6% average accuracy by weighted different source domains. DAEL achieves 48.7% average accuracy by using pseudo-labels and ensuring the consistency between the input images and their augmented views to overcome the overfitting problem. Different from these conventional UMDA methods, our method is designed for a decentralized data scenario and uses dual collaboration to obtain a global model that performs well on the target domain. As shown in Table 4, our DC-MDA method even outperforms the conventional UMDA methods on the more challenging DMDA setup.

For the source-free methods, SHOT and DECISION achieve 44.2% and 45.9% average accuracy respectively on six domains by fine-tuning the source-only models. This fine-tuning paradigm leads to poor performance compared with collaborative training sources and the target domain.

For UFDA methods, KD3A alleviates the negative transfer problem from the irrelevant domains by assigning low weights during the model aggregating stage and achieves 51.1% average accuracy, while the negative transfer problem of the irrelevant domain at the pseudo-label generation stage is ignored. The prediction from an irrelevant source domain model will lead to inaccurate pseudo-labels. Different from KD3A, we use DWSL to relieve the negative influence of the irrelevant domains during the pseudo-label generation stage. Meanwhile, we regularize the local source domain models to avoid the negative transfer problem during source model training. The target domain and source domain collaboratively improve the performance of local models on the target domain. Our method achieves 23.0% accuracy on the Quickdraw domain, outperforming KD3A by 6.6%. Our method achieves 52.5% average accuracy and outperforms the source-free and UFDA methods.

4.7 Ablation study

4.7.1 Contribution of each loss

We study the contributions of different losses in our method. Tables 5–8 show the contributions of

Table 4 Classification accuracy (mean±std) on the DomainNet dataset

Method	Classification accuracy (%)						Avg (%)
	clp	inf	pnt	qdr	rel	skt	
Source-only	52.1±0.5	23.1±0.3	47.7±1.0	13.3±0.7	60.7±0.3	46.5±0.6	40.6
DCTN (Xu et al., 2018)	48.6±0.7	23.4±0.6	48.8±0.6	7.2±0.5	53.5±0.6	47.3±0.5	38.1
M3SDA (Peng et al., 2019)	58.6±0.5	26.0±0.9	52.3±0.6	6.3±0.6	62.7±0.5	49.5±0.8	42.6
MDAN (Zhao H et al., 2018)	60.3±0.4	25.0±0.4	50.3±0.4	8.2±2.0	61.5±0.6	51.3±0.6	42.8
MDDA (Zhao SC et al., 2020)	59.4±0.6	23.8±0.8	53.2±0.6	12.5±0.6	61.8±0.5	48.6±0.8	43.2
CMSS (Yang LY et al., 2020)	64.2±0.2	28.0±0.2	53.6±0.4	16.0±0.1	63.4±0.2	53.8±0.4	46.5
ABMSDA (Zuo et al., 2021)	66.9±0.2	25.3±0.1	55.8±0.2	18.2±0.1	64.1±0.1	55.2±0.1	47.6
LtC-MSDA (Wang H et al., 2020)	63.1±0.5	**28.7±0.7**	56.1±0.5	16.3±0.5	66.1±0.6	53.8±0.6	47.4
T-SVDNet (Li RH et al., 2021)	66.1±0.4	25.0±0.8	54.3±0.7	16.5±0.9	65.4±0.5	54.6±0.6	47.0
DAEL (Zhou et al., 2021)	70.8±0.1	26.5±0.1	57.4±0.3	12.2±0.7	65.0±0.2	**60.6±0.3**	48.7
SHOT (Liang et al., 2020)	61.7	22.2	52.6	12.2	67.7	48.6	44.2
DECISION (Ahmed et al., 2021)	61.5	21.6	54.6	18.9	67.5	51.0	45.9
FADA (Peng et al., 2020)	45.3±0.7	16.3±0.8	38.9±0.7	7.9±0.4	46.7±0.4	26.8±0.4	30.3
KD3A (Feng et al., 2021)	72.5±0.6	23.4±0.4	**60.9±0.7**	16.4±0.3	72.7±0.6	**60.6±0.3**	51.1
DC-MDA	**73.6±0.1**	24.4±0.3	60.0±0.3	**23.0±0.3**	**73.1±0.2**	60.6±0.1	**52.5**

The best results are in bold

$\ell_{\rm dwsl}$, $\ell_{\rm tar}$, and $\ell_{\rm tar}+\ell_{\rm mcr}$ on the Digit-5, Office31, Office-Caltech10, and DomainNet datasets. As we can see, the $\ell_{\rm dwsl}$, $\ell_{\rm tar}$, and $\ell_{\rm tar}+\ell_{\rm mcr}$ consistently improve the average accuracy on the Digit-5, Office-Caltech10, and DomainNet datasets. For Digit-5 and Office-Caltech10, $\ell_{\rm tar}$ fully uses the unlabeled target domain data to enhance the discrimination of the local target model, which significantly improves the average accuracy. For Office31 and DomainNet, the number of classes is large and the class distribution is valuable. The average accuracy of $\ell_{\rm dwsl}$ is better than that of KD3A on Office31 and DomainNet by generating soft labels, which can capture the class distribution. $\ell_{\rm tar}$ improves the discrimination of the local target model and $\ell_{\rm mcr}$ improves the transferable performance of the local source models. The average accuracy is further improved by combining $\ell_{\rm tar}$ and $\ell_{\rm mcr}$ to reduce the domain shift on the source domain and target domain.

4.7.2 Pseudo-label strategy

We study the different pseudo-label strategies in our framework. As shown in Fig. 2a, four pseudo-label strategies are considered. The max strategy uses the max confident prediction from different source models as the pseudo-label. The mean strategy uses the mean prediction from different source models as the soft-label. The KV strategy proposed by KD3A uses the vote strategy to generate the soft-label. The DWSL strategy we propose outperforms other strategies by a large margin on hard domains such as MNIST-M and SVHN. The max strategy cannot capture the correlation information of classes and leads to inaccurate pseudo-labels due to domain shift, which is more serious on the hard domains. The mean strategy, KV strategy, and DWSL strategy can capture the class distribution information and achieve better performance than the max strategy on the hard domains. The mean strategy and KV strategy ignore the discrepancies of different source

Table 5 The accuracy of different losses on the Digit-5 dataset

Loss	Accuracy (%)					Avg (%)
	mt	mm	sv	sy	up	
$\ell_{\rm dwsl}$	99.3	80.6	85.0	92.0	98.2	91.0
$\ell_{\rm tar}$	**99.6**	95.8	91.2	97.2	**98.8**	96.5
$\ell_{\rm tar}+\ell_{\rm mcr}$	**99.6**	**97.1**	**93.2**	**97.6**	**98.8**	**97.3**

The best results are in bold

Table 6 The accuracy of different losses on the Office31 dataset

Loss	Accuracy (%)			Avg (%)
	A	D	W	
$\ell_{\rm dwsl}$	73.1	98.9	99.4	90.5
$\ell_{\rm tar}$	73.7	98.9	98.7	90.4
$\ell_{\rm tar}+\ell_{\rm mcr}$	**75.8**	**100**	**100**	**91.9**

The best results are in bold

Table 7 The accuracy of different losses on the Office-Caltech10 dataset

Loss	Accuracy (%)				Avg (%)
	A	C	D	W	
$\ell_{\rm dwsl}$	**97.9**	96.0	96.8	98.3	97.3
$\ell_{\rm tar}$	97.4	94.2	**100**	**100**	97.9
$\ell_{\rm tar}+\ell_{\rm mcr}$	96.9	**97.2**	**100**	**100**	**98.5**

The best results are in bold

Table 8 The accuracy of different losses on the DomainNet dataset

Loss	Accuracy (%)						Avg (%)
	clp	inf	pnt	qdr	rel	skt	
$\ell_{\rm dwsl}$	72.5	23.9	58.4	21.4	70.5	59.0	51.0
$\ell_{\rm tar}$	73.4	**24.8**	58.9	22.3	71.3	60.5	51.9
$\ell_{\rm tar}+\ell_{\rm mcr}$	**73.6**	24.4	**60.0**	**23.0**	**73.1**	**60.6**	**52.5**

The best results are in bold

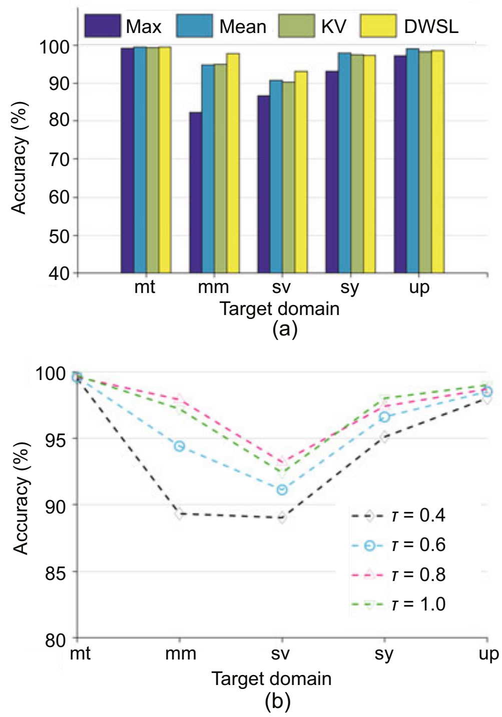

Fig. 2 The accuracy of different pseudo-label strategies in our framework (a) and the sensitivity of hyper-parameter τ (b)

models. Our DWSL strategy uses entropy information from different source models to generate domain weights, and we use the temperature parameter τ to sharpen the soft-label for better discrimination performance.

We further analyze the sensitivity of the temperature parameter τ ranging from 0.4 to 1.0 in Fig. 2b. We find that the accuracy does not change too much, which indicates that our method is not sensitive to the value of τ in this range.

4.7.3 Negative transfer problem

We further analyze the negative transfer problem on the source domains and the effect of regularization. As shown in Fig. 3, we evaluate the accuracy of different local source models and the initial global models on the target domain data to analyze the transferable performance on the target domain. In our experiments, MNIST-M is the target domain and the remaining domains are the source domains on the Digit-5 dataset. In Fig. 3a, the local source models, e.g., trained on the SVHN and Synthetic domains, perform worse on the target domain compared with the initial global model during training. The local source models shifting away from the target domain will lead to inaccurate pseudo-labels on the target domain. Meanwhile, aggregating these local source models, which perform worse on the target domain, will reduce the performance of the global model on the target domain.

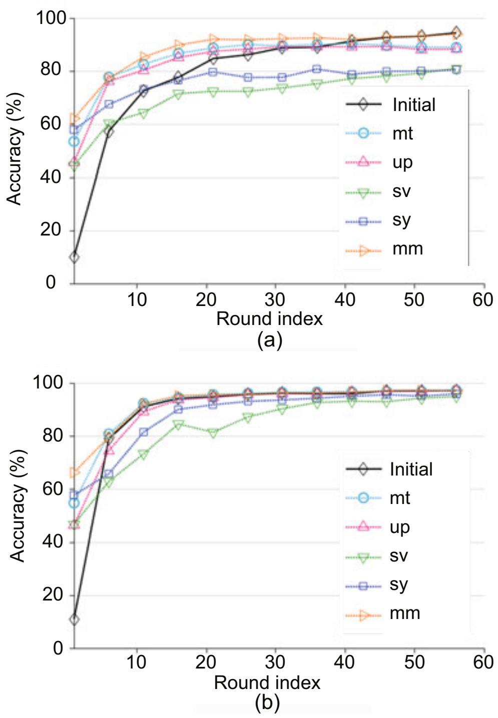

Fig. 3 The accuracy of the local source domain model on the target domain after local training without the collaboration of the local target model (a) and with the collaboration of the local target model by using $\ell_{\rm mcr}$ (b)

To solve the negative transfer problem, we use the local target model to regularize the local source model. As shown in Fig. 3b, the transferable performance of the local source models, e.g., those trained on the source domains SVHN and Synthetic, is not much worse than that of the initial model during training. The negative transfer problem can be relieved by using $\ell_{\rm mcr}$ regularization. As shown in Table 5, the adaptation performance on hard domains, such as MNIST-M and SVHN, is further improved by using $\ell_{\rm mcr}$ on the source domains. By using $\ell_{\rm mcr}$, the adaptation performance on other datasets is also improved, as shown in Tables 6–8, which indicates the effect of $\ell_{\rm mcr}$ on relieving the negative transfer problem.

4.7.4 Model aggregation strategy

We further study the aggregating strategies including aggregating only the feature extractors and aggregating the feature extractors and classifiers. As shown in Tables 9–12, by aggregating classifiers (AC), the average accuracy of the global model is slightly better than that of aggregating only the feature extractor. In several hard domains, such as SVHN in Digit-5, Amazon in Office31, and Quickdraw in DomainNet, aggregating the classifiers leads

Table 9 Contributions of aggregating classifiers (AC) for Digit-5

	Accuracy (%)					Avg (%)
	mt	mm	sv	sy	up	
w/o AC	**99.7**	**97.3**	92.2	**97.8**	**98.9**	97.2
with AC	99.6	97.1	**93.2**	97.6	98.8	**97.3**

The best results are in bold

Table 10 Contributions of aggregating classifiers (AC) for Office31

	Accuracy (%)			Avg (%)
	A	D	W	
w/o AC	73.7	**100**	98.7	90.8
with AC	**75.8**	**100**	**100**	**91.9**

The best results are in bold

to better performance by reducing the classifier discrepancy. In contrast, with KD3A, which uses domain-specific classifiers and aggregates only feature extractors, we aggregate the feature extractors and classifiers of different local models to obtain better performance.

4.7.5 Feature visualization

As shown in Fig. 4, we visualize the features of the source domains and target domain by t-SNE (van der Maaten and Hinton, 2008) to verify the effect of our method on reducing domain shift. Figs. 4a and 4d illustrate the features extracted by the source-only model. Without using target domain data, the source-only model cannot extract discriminated target features (red points). Figs. 4b and 4e illustrate the features extracted by the global model, which uses our DWSL strategy to train the unlabeled target domain. The target features (red points) are aligned to the clusters of source features. Figs. 4c and 4f illustrate the features extracted by the global model, which uses our DC-MDA method. The target

Table 11 Contributions of aggregating classifiers (AC) for Office-Caltech10

	Accuracy (%)				Avg (%)
	A	C	D	W	
w/o AC	95.8	96.9	**100**	**100**	98.2
with AC	**96.9**	**97.2**	**100**	**100**	**98.5**

The best results are in bold

Table 12 Contributions of aggregating classifiers (AC) for DomainNet

	Accuracy (%)						Avg (%)
	clp	inf	pnt	qdr	rel	skt	
w/o AC	72.3	**24.4**	57.8	21.6	72.4	59.5	51.3
with AC	**73.6**	**24.4**	**60.0**	**23.0**	**73.1**	**60.6**	**52.5**

The best results are in bold

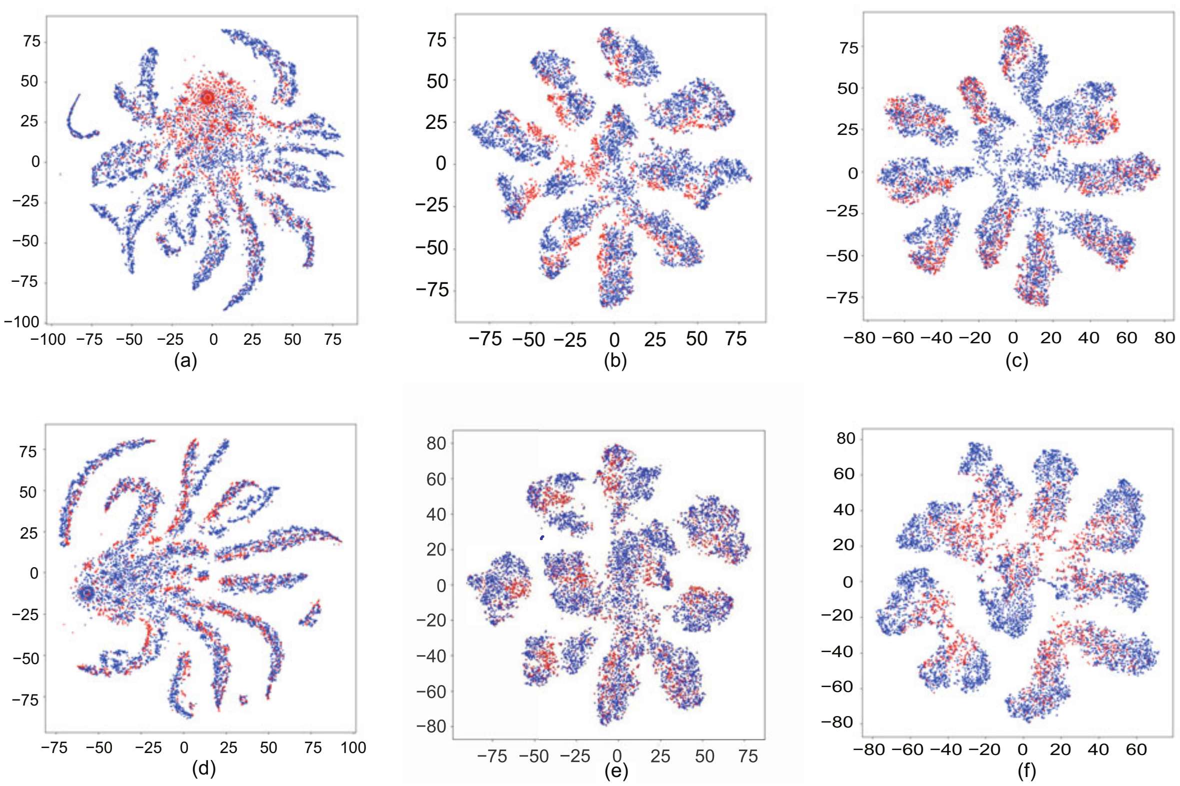

Fig. 4 Visualization of the features (blue points) from the source domains and the features (red points) from the target domain on the Digit-5 dataset

The first row indicates the target domain MNIST-M; the second row indicates the target domain SVHN. The features are extracted by the source-only model in (a) and (d), using the DWSL to train the target model in (b) and (e), and using our DC-MDA method in (c) and (f). Reference to color refer to the online version of this figure features are aligned to source features and become more discriminating on the feature space, which leads to better classification performance.

5 Conclusions

In this paper, we propose a novel DC-MDA method for DMDA. We overcome the data decentralization and domain shift challenges by training the target domain model with the collaboration of source domain models and training the source domain models with the collaboration of the target model. Our DC-MDA method does not collect data from different domains and uses only the models from different domains to overcome the data decentralization challenges and solve the domain shift problem. For the domain shift challenge, our method improves the discrimination of the local target model and the transferable performance of the local source models with collaboration from other domain models to obtain a model that works well on the target domain. Under the DMDA setup, our method significantly improves the domain adaptation performance on the target domain.

Contributors

Yikang WEI designed the research and drafted the paper. Yahong HAN helped organize the paper. Yikang WEI revised and finalized the paper.

Compliance with ethics guidelines

Yikang WEI and Yahong HAN declare that they have no conflict of interest.

References

Ahmed SM, Raychaudhuri DS, Paul S, et al., 2021. Unsupervised multi-source domain adaptation without access to source data. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.10103-10112. https://doi.org/10.1109/CVPR46437.2021.00997

Cubuk ED, Zoph B, Shlens J, et al., 2020. Randaugment: practical automated data augmentation with a reduced search space. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops, p.702-703. https://doi.org/10.1109/CVPRW50498.2020.00359

Feng HZ, You ZY, Chen MH, et al., 2021. KD3A: unsupervised multi-source decentralized domain adaptation via knowledge distillation. Pro 38th Int Conf on Machine Learning, p.3274-3283. https://doi.org/10.48550/arXiv.2011.09757

Ganin Y, Lempitsky V, 2015. Unsupervised domain adaptation by backpropagation. Pro 32nd Int Conf on Machine Learning, p.1180-1189.

Gao Z, Guo LM, Ren TW, et al., 2022. Pairwise two-stream ConvNets for cross-domain action recognition with small data. *IEEE Trans Neur Netw Learn Syst*, 33(3):1147-1161. https://doi.org/10.1109/TNNLS.2020.3041018

Han YH, Wu AM, Zhu LC, et al., 2021. Visual commonsense reasoning with directional visual connections. *Front Inform Technol Electron Eng*, 22(5):625-637. https://doi.org/10.1631/FITEE.2000722

Kundu JN, Venkat N, Rahul MV, et al., 2020. Universal source-free domain adaptation. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.4544-4553. https://doi.org/10.1109/CVPR42600.2020.00460

Kundu JN, Kulkarni A, Singh A, et al., 2021. Generalize then adapt: source-free domain adaptive semantic segmentation. Proc IEEE/CVF Int Conf on Computer Vision, p.7046-7056. https://doi.org/10.1109/ICCV48922.2021.00696

Li R, Jiao QF, Cao WM, et al., 2020. Model adaptation: unsupervised domain adaptation without source data. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.9641-9650. https://doi.org/10.1109/CVPR42600.2020.00966

Li RH, Jia X, He JZ, et al., 2021. T-SVDNet: exploring high-order prototypical correlations for multi-source domain adaptation. Proc IEEE/CVF Int Conf on Computer Vision, p.9991-10000. https://doi.org/10.1109/ICCV48922.2021.00984

Li S, Song SJ, Wu C, 2018. Layer-wise domain correction for unsupervised domain adaptation. *Front Inform Technol Electron Eng*, 19(1):91-103. https://doi.org/10.1631/FITEE.1700774

Liang J, Hu DP, Feng JS, 2020. Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation. Proc 37th Int Conf on Machine Learning, p.6028-6039.

Liu YA, Zhang W, Wang J, 2021. Source-free domain adaptation for semantic segmentation. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.1215-1224. https://doi.org/10.1109/CVPR46437.2021.00127

Long MS, Cao Y, Wang JM, et al., 2015. Learning transferable features with deep adaptation networks. Proc 32nd Int Conf on Machine Learning, p.97-105.

McMahan B, Moore E, Ramage D, et al., 2017. Communication-efficient learning of deep networks from decentralized data. Proc 20thInt Conf on Artificial Intelligence and Statistics, p.1273-1282.

Miyato T, Maeda Si, Koyama M, et al., 2018. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE Trans Patt Anal Mach Intell*, 41(8):1979-1993. https://doi.org/10.1109/tpami.2018.2858821

Nguyen VA, Nguyen T, Le T, et al., 2021. STEM: an approach to multi-source domain adaptation with guarantees. Proc IEEE/CVF Int Conf on Computer Vision, p.9352-9363. https://doi.org/10.1109/ICCV48922.2021.00922

Peng XC, Bai QX, Xia XD, et al., 2019. Moment matching for multi-source domain adaptation. Proc IEEE/CVF Int Conf on Computer Vision, p.1406-1415. https://doi.org/10.1109/ICCV.2019.00149

Peng XC, Huang ZJ, Zhu YZ, et al., 2020. Federated adversarial domain adaptation. Proc 8th Int Conf on Learning Representations, p.1-19. https://doi.org/10.48550/arXiv.1911.02054

Pu SL, Zhao W, Chen WJ, et al., 2021. Unsupervised object detection with scene-adaptive concept learning. *Front Inform Technol Electron Eng*, 22(5):638-651. https://doi.org/10.1631/FITEE.2000567

Qiu Z, Zhang YF, Lin HB, et al., 2021. Source-free domain adaptation via avatar prototype generation and adaptation. Proc 30th Int Joint Conf on Artificial Intelligence, p.2921-2927. https://doi.org/10.24963/ijcai.2021/402

Saito K, Kim D, Sclaroff S, et al., 2019. Semi-supervised domain adaptation via minimax entropy. Proc IEEE/CVF Int Conf on Computer Vision, p.8050-8058. https://doi.org/10.1109/ICCV.2019.00814

van der Maaten L, Hinton G, 2008. Visualizing data using t-SNE. *J Mach Learn Res*, 9(86):2579-2605.

Venkat N, Kundu JN, Singh DK, et al., 2020. Your classifier can secretly suffice multi-source domain adaptation. Proc 34th Int Conf on Neural Information Processing Systems, 33:4647-4659. https://doi.org/10.48550/arXiv.2103.11169

Wang B, Li G, Wu C, et al., 2022. A framework for self-supervised federated domain adaptation. *EURASIP J Wirel Commun Network*, 2022(1):37. https://doi.org/10.1186/s13638-022-02104-8

Wang H, Xu MH, Ni BB, et al., 2020. Learning to combine: knowledge aggregation for multi-source domain adaptation. Proc 16th European Conf on Computer Vision, p.727-744.

Wu GL, Gong SG, 2021. Collaborative optimization and aggregation for decentralized domain generalization and adaptation. Proc IEEE/CVF Int Conf on Computer Vision, p.6484-6493. https://doi.org/10.1109/iccv48922.2021.00642

Xia HF, Zhao HD, Ding ZM, 2021. Adaptive adversarial network for source-free domain adaptation. Proc IEEE/CVF Int Conf on Computer Vision, p.9010-9019. https://doi.org/10.1109/ICCV48922.2021.00888

Xu RJ, Chen ZL, Zuo WM, et al., 2018. Deep cocktail network: multi-source unsupervised domain adaptation with category shift. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.3964-3973. https://doi.org/10.1109/CVPR.2018.00417

Yang LY, Balaji Y, Lim SN, et al., 2020. Curriculum manager for source selection in multi-source domain adaptation. Proc 16th European Conf on Computer Vision, p.608-624. https://doi.org/10.1007/978-3-030-58568-6_36

Yang Q, Liu Y, Cheng Y, et al., 2020. Federated Learning. Springer, Cham, Germany.

Yang Y, Zhuang YT, Pan YH, 2021. Multiple knowledge representation for big data artificial intelligence: framework, applications, and case studies. *Front Inform Technol Electron Eng*, 22(12):1551-1558. https://doi.org/10.1631/FITEE.2100463

Zhao H, Zhang SH, Wu GH, et al., 2018. Adversarial multiple source domain adaptation. Proc 32nd Int Conf on Neural Information Processing Systems, p.8568-8579.

Zhao SC, Li B, Yue XY, et al., 2019. Multi-source domain adaptation for semantic segmentation. Proc 33rd Int Conf on Neural Information Processing Systems, p.655.

Zhao SC, Wang GZ, Zhang SH, et al., 2020. Multi-source distilling domain adaptation. Proc AAAI on Conf Artificial Intelligence, 34(7):12975-12983. https://doi.org/10.1609/aaai.v34i07.6997

Zhao SC, Li B, Xu PF, et al., 2021. Madan: multi-source adversarial domain aggregation network for domain adaptation. *Int J Comput Vis*, 129(8):2399-2424. https://doi.org/10.1007/s11263-021-01479-3

Zhao YB, Zhang H, Gao Z, et al., 2022. A temporal-aware relation and attention network for temporal action localization. *IEEE Trans Image Process*, 31:4746-4760. https://doi.org/10.1109/TIP.2022.3182866

Zhou KY, Yang YX, Qiao Y, et al., 2021. Domain adaptive ensemble learning. *IEEE Trans Image Process*, 30:8008-8018. https://doi.org/10.1109/TIP.2021.3112012

Zuo YK, Yao HT, Xu CS, 2021. Attention-based multi-source domain adaptation. *IEEE Trans Image Process*, 30:3793-3803. https://doi.org/10.1109/TIP.2021.3065254