CoI^{2}A: Collaborative Inter-domain and Intra-domain Alignments for Multisource Domain Adaptation

Chen Lin, Zhenfeng Zhu, Shenghui Wang, Zhenwei Shi, *Senior Member, IEEE*, and Yao Zhao, *Fellow, IEEE*

***Abstract*— In the remote sensing information interpretation tasks, compared with collecting lots of high-quality image labels for the target domain, a large amount of labeled remote sensing data from multiple source domains are generally available without any extra cost. In this article, our work focuses on how to exploit the rich knowledge obtained from multiple source domains to guide the interpretation of the target scene, and we propose a novel framework called collaborative interdomain and intradomain alignments for multisource domain adaptation (MDA), namely CoI^{2}A, in which interdomain and intradomain alignments are well collaborated to reduce the distribution divergence across sources and target. To reduce the discrepancy across sources, the intersource alignment is proposed to map multiple sources into a unified representation space. In addition, the cross-domain attention is introduced to enforce the intra-class compactness of the target. Interdomain alignment aligns each source with target domain separately with the help of cross-domain attention. As for the intradomain alignment, the multihead attentive representations of the target obtained by cross-domain attention are correlated into a unified one. The experimental results obtained from different scene classification tasks demonstrate the superiority of our model.**

***Index Terms*— Class-aware alignment, interdomain alignment, intradomain alignment, multisource domain adaptation (MDA), scene classification.**

NOMENCLATURE

Notation	*Description*
$\mathcal{D}^{s_i} = (X^{s_i}, Y^{s_i})$	ith labeled source domains.
$\mathcal{D}^t = (X^t)$	Unlabeled target domain.
$X^{s_i} = \{x_n^{s_i}\}_{n=1}^{N^{s_i}}$	Set of N^{s_i} samples for the ith source domain.
$Y^{s_i} = \{y_n^{s_i}\}_{n=1}^{N^{s_i}}$	Set of labels for the ith source domain.
$X^t = \{x_n^t\}_{n=1}^{N^t}$	Set of N^t target samples.
$P^{s_i}(x, y)$	Distribution of the ith source domain.
$P^t(x, y)$	Distribution of the target domain.
K	Total number of categories.
$X_k^{s_i}$	Set of ith source domain samples belonging to kth category.
$N_k^{s_i}$	Total number of samples in $X_k^{s_i}$.
$\tilde{\mathcal{D}}^{s_i} = (\tilde{X}^{s_i}, \tilde{Y}^{s_i})$	M new generated labeled source domains.
$\tilde{X}^{s_i} = \{\tilde{x}_n^{s_i}\}_{n=1}^{N^{s_i}}$	Set of N^{s_i} samples for the ith new synthesized source domain.
$\tilde{Y}^{s_i} = \{\tilde{y}_n^{s_i}\}_{n=1}^{N^{s_i}}$	Set of labels for the ith new synthesized source domain.
$F^{\text{pre}}(x)$ and $C_i^{\text{pre}}(F(x))$	Embedding for x in feature layer and classification layer during pretraining for intersource alignment.
$F(x)$ and $C_i(F(x))$	Embedding for x in feature layer and classification layer.

I. INTRODUCTION

REMOTE sensing scene classification is one of the hotspots in the field of remote sensing, dedicating to decoding very high-resolution (VHR) images accurately into required geographic information under limited human and material resources [1]. In the real-world scenario, a large amount of labeled remote sensing data, also called by source data in our work, is easily available without additional labeling cost. However, due to different acquisition methods and surface feature styles between the existing remote sensing data (source data) and the remote sensing data that need to be classified (target data),[1] the classification results of target data on the classifier directly trained from source data are not unsatisfactory. How to narrow the difference between source data and target data becomes one of the core issues.

Manuscript received 9 June 2023; revised 20 September 2023; accepted 16 October 2023. Date of publication 20 October 2023; date of current version 3 November 2023. This work was supported in part by the Science and Technology Innovation 2030—New Generation Artificial Intelligence Major Project under Grant 2018AAA0102100 and in part by the National Natural Science Foundation of China under Grant 61976018. *(Corresponding author: Zhenfeng Zhu.)*

Chen Lin, Zhenfeng Zhu, Shenghui Wang, and Yao Zhao are with the Beijing Key Laboratory of Advanced Information Science and Network Technology and the Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China (e-mail: cchenlin@bjtu.edu.cn; zhfzhu@bjtu.edu.cn; shwang@bjtu.edu.cn; yzhao@bjtu.edu.cn).

Zhenwei Shi is with the School of Astronautics, Beihang University, Beijing 100083, China (e-mail: shizhenwei@buaa.edu.cn).

Digital Object Identifier 10.1109/TGRS.2023.3326156

[1]Note that since the source data and target data may have different modalities, we will uniformly refer to these data as domains throughout this article.

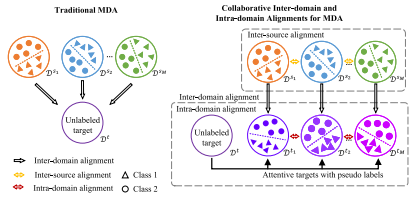

Fig. 1. Schematic of MDA. (Left) Traditional MDA methods generally perform independently domain alignment between each source and target. (Right) Proposed CoI^{2}A collaborates the three alignment strategies, including intersource, interdomain, and intradomain alignment, in a unified framework.

Unsupervised domain adaptation (UDA) aims to narrow the distribution discrepancy between source domain and target domain [2], reducing the labeling burden on the target domain. Unlike UDA with only one single-source domain available, multisource domain adaptation (MDA) extends DA by collecting labeled data from multiple source domains with different distributions [3]. MDA is more feasible in practice, and has received considerable attention in many practical applications [4], [5]. Since different types of remote sensing data generally contains diverse data structures and rich ground object information, they can reflect the types of landforms from various angles. For this reason, we mainly focus on leveraging MDA to make full use of the characteristics of different source domains for the classification of target domain.

Considering the different imaging locations and collection methods of remote sensing data, domain shift exists commonly among different datasets. For the scenario of multisource domain, a simple and straightforward approach is to combine all source domains and then use a single-source DA method. However, this kind of method ignores the differences between source domains and tents to result in a suboptimal model [6]. As illustrated in Fig. 1, although some traditional MDA methods [7], [8] (left) take different source domains into account in a manner of many to one, they lack of considering the discrepancy of different sources as well as how to strike a good balance between interdomain alignment and inter source alignment. What's more, the rich detailed information and complex composition for VHR images lead to the fact that the ground objects in different scenes may be the same, and the differences between the ground objects in the same scene may be magnified. However, the current MDA algorithms for scene classification only consider the alignment between domains, and ignore the alignment between the category information of multiple domains [9], [10].

To address the above issues, we attempt to establish a novel framework for MDA whose motivation is well illustrated in the right of Fig. 1. Compared with traditional MDA methods, the proposed framework not only considers the discrepancy across sources, but also makes full use of the discriminant information from sources to effectively promote the class-aware alignment between each source and target to reduce the domain shift between the source and the target. In brief, our main contributions can be summarized as follows.

1) Toward the MDA task, we propose a collaborative inter-domain and intra-domain alignments network for scene classification, also named by CoI^{2}A, in which the intersource, interdomain, and intradomain alignments are well collaborated to perform cross-domain adaption.
2) To achieve interdomain alignment technically, cross-domain attention associated with source semantic information is proposed to guide to generate multihead attentive representation, thus facilitating the category-aware alignment between multisource and target.
3) To further integrate multihead attentive representations of target, a favorable intradomain alignment via shared representation learning is used to maintain the consistence of the distributions among them.
4) Extensive experiments are conducted on several remote sensing datasets, and the experimental results demonstrate that the proposed model outperforms other existing approaches on the benchmark datasets.

II. Related Work

A. Remote Sensing Scene Classification

Remote sensing scene classification aims to infer the ground object category of the observation target according to the distribution and texture structure of the ground object contained in the remote sensing images. Currently, the existing remote sensing scene classification methods can be roughly divided into hand-crafted feature-based [11], [12] and deep feature-based [13], [14].

Although the substantial progress in remote sensing scene classification has been made, the current works still face the challenge of poor transferability in classification. To achieve automatic interpretation, UDA-based methods are widely used in remote sensing scene classification [15], [16]. In practical applications, due to the need for using multiple remote sensing of different modalities, many circumstances faced with more than one source domain [17]. MB-NET [18] is the first method to apply MDA in remote sensing scene classification, in which a multibranch network is proposed using the mean of multiple features to reduce the distance between different domains. Lu et al. [19] introduced a multisource compensation network (MSCN) to solve MDA problem of nonshared classes. Although previous works have made some success, they mostly achieve domain-level alignment and ignore the misalignment across sources.

B. Multisource Domain Adaptation

The study on MDA originated from adaptive support vector machines (A-SVM) proposed by Yang et al. [20], which aggregated the classifiers of each source domain to adjust to the classification model in target domain. Blitzer et al. [21] proposed the generalization boundary of transfer learning for the first time, which laid the theoretical foundation for MDA. Depending on the generalization boundary, more relevant source domains can be selected. Mansour et al. [22] pointed out that the prediction results of target data could be represented by the weighted combination results of multiple source domain distributions. So far, the approaches based on the weighted combination of source classifiers has been widely used in MDA.

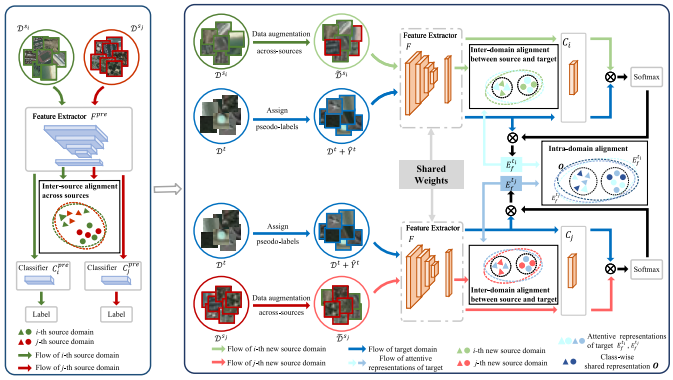

Fig. 2. Framework of the proposed collaborative interdomain and intradomain alignments for MDA, i.e., CoI^{2}A. Both the feature extractor F^{pre} and F use Resnet-50 [25] as backbone, while F^{pre} is pretrained on ImageNet and F is fine-tuned on the well pretrained F^{pre}. C_i^{pre} and C_i are independent MLP classifiers. $\mathbf{O} \in \mathbb{R}^{K \times d_f}$ associates the multihead attentive representations of target $\{E_f^{t_i}\}_{i-1}^{M}$ by learning a shared representation.

Most recently, deep-learning-based MDA methods have been well developed. Peng et al. [23] established the DomainNet, the largest DA dataset to date, and proposed a moment-matching model (M^{3}SDA) to dynamically align moment distance of feature distributions to transfer the knowledge learned from multiple labeled source domains to unlabeled target domains. Wang et al. [24] explored the interaction between domains and proposed a MDA network. In general, the previous methods mainly focus on aligning sources with target in a manner of many to one, while our model is capable of collaborating three alignment strategies including intersource, interdomain, and intradomain in a unified framework.

III. Overview of the Proposed MDA Framework

A. Notations

For scene classification task with multiple sources, there are M labeled source domains $\{\mathcal{D}^{s_i}\}_{i=1}^{M}$ and an unlabeled target domain $\mathcal{D}^t$. For the ith source domain $\mathcal{D}^{s_i}$, let $X^{s_i} = \{x_n^{s_i}\}_{n=1}^{N^{s_i}}$ and $Y^{s_i} = \{y_n^{s_i}\}_{n=1}^{N^{s_i}}$ indicate the observed N^{s_i} images and corresponding labels drawn from the source distribution $P^{s_i}(x, y)$, Similarly, the unlabeled target data $X^t = \{x_n^t\}_{n=1}^{N^t}$ are drawn from the target distribution $P^t(x, y)$ without label observation, where N^t denotes the number of target images. Note that there are two general assumptions in multisource cross-domain classification task: 1) all domains share the same categories in our scene classification task, i.e., $y_n^{s_i} \in \mathcal{Y}$, $y_n^t \in \mathcal{Y}$, where $\mathcal{Y} = \{1, 2, \ldots, K\}$ is the class label space and K is the number of classes and 2) domains have different distributions, i.e., $P^{s_i}(x, y) \neq P^{s_j}(x, y) \neq P^t(x, y)$ $\forall i, j \in 1, 2, \ldots, M$. Our goal is to learn a flexible transfer model to predict target samples with the help from multisources. To make the description of other notations more concise, we list them in Nomenclature section.

B. Framework

The overall framework of the proposed CoI^{2}A, is shown in Fig. 2. As we can see, it mainly consists of three alignment strategies that are briefly descried as follows.

1) *Intersource Alignment Across Sources:* The purpose of the above intersource alignment is to coordinate multiple sources into a compatible representation space, while alleviating the source discrepancy across sources due to different imaging methods.
2) *Interdomain Alignment Between Source and Target:* For each source $\mathcal{D}^{s_i}$, $i = 1, \ldots, M$, and its cocunterpart target $\mathcal{D}^t$, a category-aware interdomain alignment between them is implemented with the guidance of cross-domain attention, where $\tilde{\mathcal{D}}^{s_i}$ denotes the augmentation of $\mathcal{D}^{s_i}$ by the means of data exchange across sources.
3) *Intradomain Alignment:* To promote consistent distribution among the multihead attentive representations $\{E_f^{t_i}\}_{i=1}^{M}$ of the target, a favorable intradomain alignment is adopted to align semantically them via a unified representation that is constrained by the pseudo-label on the target domain.

IV. Methodology

A. Intersource Alignment

Due to the differences in imaging modes, environment, and other factors, there exists inevitably discrepancy not only between each source domain and target domain, but also between any two source domains in MDA. To address this issue, the shared backbone network (feature extractor) F^{pre} using ResNet-50 [25] as backbone is first pretrained as shown in Fig. 2 to establish a macro multisource alignment. In pretraining, the maximum mean discrepancy (MMD) is used to narrow down the gap between any pair of sources $\mathcal{D}^{s_i}$ and $\mathcal{D}^{s_j}$, and we have

$$\mathcal{L}_{\text{mmd}} = \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \mathbf{MMD}\left(F^{\text{pre}}(X^{s_i}), F^{\text{pre}}(X^{s_j})\right) \tag{1}$$

with $\mathbf{MMD}(F^{\text{pre}}(X^{s_i}), F^{\text{pre}}(X^{s_j}))$ being defined by

$$\mathbf{MMD}(F^{\text{pre}}(X^{s_i}), F^{\text{pre}}(X^{s_j})) := \sup_{\phi \sim \mathcal{H}} \|\mathbb{E}_{X^{s_i} \sim \mathcal{P}_i}[\phi(F^{\text{pre}}(X^{s_i}))] - \mathbb{E}_{X^{s_j} \sim \mathcal{P}_j}[\phi(F^{\text{pre}}(X^{s_j}))]\|_{\mathcal{H}}^2 \tag{2}$$

where P^{s_i} and P^{s_j} are two distributions that $X^{s_i} \in \mathcal{D}^{s_i}$ and $X^{s_j} \in \mathcal{D}^{s_j}$ obey, respectively, $\phi(\cdot)$ denotes the mapping function, and $\|\cdot\|_{\mathcal{H}}$ is the RKHS.

To enhance the discriminative ability in the embedding space $F^{\text{pre}}(\cdot)$, we also pretrain an independent multilayer perceptron (MLP) classifier C_i^{pre} for each source domain $\mathcal{D}^{s_i}$, $i = 1, \ldots, M$, with the cross-entropy loss as follows:

$$\mathcal{L}_{\text{clc}} = -\frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_{(x^{s_i}, y^{s_i}) \sim (X^{s_i}, Y^{s_i})} \left[\mathbf{y}^{s_i} \log\left(C_i^{\text{pre}}\left(F^{\text{pre}}\left(x^{s_i}\right)\right)\right)\right] \tag{3}$$

where $\mathbf{y}^{s_i}$ is an one-shot encoded vector of y^{s_i}.

By further combining (1) and (3), we have the following macro multisource alignment loss $\mathcal{L}_s$:

$$\mathcal{L}_s = \mathcal{L}_{\text{clc}} + \mathcal{L}_{\text{mmd}}. \tag{4}$$

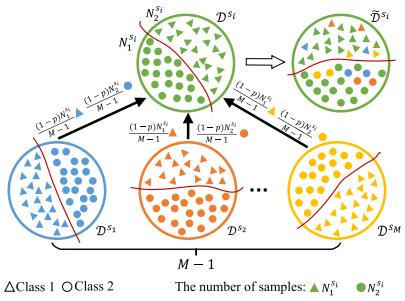

Fig. 3. Sketch map of data augmentation across sources. For source domain $\{\mathcal{D}^{S_i}\}_{i=1}^{M}$, appropriate samples are selected from the remaining source domains for replacement to form a new augmented source domain $\{\tilde{\mathcal{D}}^{Si}\}_{i=1}^{M}$. Best viewed in color.

Explicitly, $\mathcal{L}_s$ not only facilitates the eliminating of the gaps among different source domains caused by the above discrepancy, but also enables the embedding space $F^{\text{pre}}(\cdot)$ more discriminative.

B. Data Augmentation and Pseudo-Labels Generation

1) Data Augmentation Across-Sources: To gap the cross-source bias and promote the compatibility of the same category from multiple source domains, we come up with a simple but effective way for data augmentation. Specifically, as shown in Fig. 3, let $X_k^{S_i} \in X^{S_i}$ represent the set of $N_k^{S_i}$ source samples belonging to the kth category of source $\mathcal{D}^{S_i}$, where $N_k^{S_i}$ is the total number of $X_k^{S_i}$ and $N^{S_i} = \sum_{k=1}^{K} N_k^{S_i}$. We randomly leave the sample with percentage p (in our work, p is set to 80%) of the original $X_k^{S_i}$ unchanged, and then replace the remaining samples with $((1-p)N_k^{S_i})/(M-1)$ samples randomly drawn from samples belonging to the kth category in other $M-1$ source domains.

By the above random data exchange among multiple sources, we obtain the new augmented source domains as $\{\tilde{\mathcal{D}}^{S_i}\}_{i=1}^{M}$. And for the new ith source domain $\tilde{\mathcal{D}}^{S_i}$, we have $\tilde{X}^{S_i} = \{\tilde{x}_n^{S_i}\}_{n=1}^{N^{S_i}}$ and $\tilde{Y}^{S_i} = \{\tilde{y}_n^{S_i}\}_{n=1}^{N^{S_i}}$. Intuitively, compared to the original source domains $\{\mathcal{D}^{S_i}\}_{i=1}^{M}$, the new augmented source domains $\{\tilde{\mathcal{D}}^{S_i}\}_{i=1}^{M}$ will be more favorable to boost the generalization of the learned model for the subsequent target classification.

2) Pseudo-Labels Generation: Inspired by active learning, reliable pseudo-labels for the target samples are expected for self-supervised learning in target domain to achieve category-aware alignment. In particular, to carry out the interdomain alignment between $\mathcal{D}^{S_i}$ and target $\mathcal{D}^t$, $i = 1, \ldots, M$, let us assume that K target class prototypes $\{c_k^{t_i}\}_{k=1}^{K}$ are available, where t_i means that these target prototypes receive the knowledge from the ith source domain $\mathcal{D}^{S_i}$. For the input target sample x_n^t, $n = 1, \ldots, N^t$, the corresponding pseudo-label $\hat{y}_n^{t_i}$ is obtained by assigning the category of the prototype to which it is closest. In real implementation, to import the category knowledge of the source domain $\tilde{\mathcal{D}}^{S_i}$ into the target class prototypes, its class centroids $\{\tilde{c}_k^{S_i}\}_{k=1}^{K}$ are used to initialize the target class prototypes, where $\tilde{c}_k^{S_i} = \sum_{n=1}^{N^{S_i}} \mathbb{1}(\tilde{y}_n^{S_i} = k) F(\tilde{x}_n^{S_i})/N_k^{S_i}$ and $\mathbb{1}$ denotes a "0–1" indicator.

It should be noted that the target class prototypes $\{c_k^{t_i}\}_{k=1}^{K}$ given above are updated at every training epoch using the class centroids of the samples with same pseudo-label in target domain, and we have $c_k^{t_i} = \sum_{n=1}^{N^t} \mathbb{1}(\hat{y}_n^{t_i} = k) F(x_n^t)/N_k^t$, where $\hat{y}_n^{t_i}$ is the pseudo-label of x_n^t and N_k^t is the number of target samples that belong to the kth pseudo-label. In addition, those samples which are far from the affiliated prototype are discarded without being assigned a pseudo label.

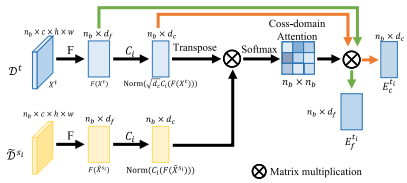

Fig. 4. Details of cross-domain attention module. The cross-domain attention is calculated on a set of Q, K, and V, where $Q^i = \text{Norm}(C_i(F(\tilde{X}^{S_i})))$ and $K^i = \sqrt{d_c}\text{Norm}(C_i(F(X^t)))$, $V_f = F(X^t)$ and $V_c = C_i(F(X^t))$. Norm$(\cdot)$ means an operator of row normalization.

C. Interdomain Alignment With Cross-Domain Attention Guidance

1) Cross-Domain Attention: To improve the discriminability of the representation learned for the target domain, a cross-domain attention mechanism is proposed. By associating the semantic information of the source domain to the target domain in the form of attention, it helps to strengthen the within-class compactness of the target domain. As we know, the general attention function is computed on a set of queries, keys, and values (Q, K, V), which is formulated as

$$\textbf{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{5}$$

where d_k is the dimension of query and key and $\boldsymbol{\alpha} = \text{softmax}(QK^\top/\sqrt{d_k})$ denotes the attention map.

Here, unlike traditional attention mechanism utilizing the same inputs into the query and key, we take $Q^i = \text{Norm}(C_i(F(\tilde{X}^{S_i}))) \in \mathbb{R}^{n_b \times d_c}$ and $K^i = \sqrt{d_c}\text{Norm}(C_i(F(X^t))) \in \mathbb{R}^{n_b \times d_c}$ as query and key, respectively, to establish the cross-domain attention map $\boldsymbol{\alpha}^i$ of the ith attention head as illustrated in Fig. 4, where Norm$(\cdot)$ means an operator of row normalization, n_b denotes the number of the input, and d_c is the dimension of query and key. Thus, for each element $\alpha_{uv}^i \in \boldsymbol{\alpha}^i \ \forall u, v \in \{1, 2, \ldots, n_b\}$, we have

$$\alpha_{uv}^i = \frac{C_i\left(F\left(\tilde{x}_u^{S_i}\right)\right) \cdot C_i\left(F\left(x_v^t\right)\right)}{\|C_i\left(F\left(\tilde{x}_u^{S_i}\right)\right)\|_2 \|C_i\left(F\left(x_v^t\right)\right)\|_2}. \tag{6}$$

Clearly, α_{uv}^i tends to measure the impact of the uth source semantic information on the vth target sample. The score of α_{uv} indicates the similarity of $\tilde{x}_u^{S_i}$ and x_v^t. The higher the score, the more likely it is that $\tilde{x}_u^{S_i}$ and x_v^t are of the same category.

Based on the attention maps $\{\boldsymbol{\alpha}^i\}_{i=1}^{M}$, we can obtain the multihead attentive representations $\{E_f^{t_i}\}_{i=1}^{M}$ and $\{E_c^{t_i}\}_{i=1}^{M}$ for target domain in feature layer and classification layer, respectively, where $E_f^{t_i} = \textbf{Attention}(Q^i, K^i, V_f)$ with $V_f = F(X^t) \in \mathbb{R}^{n_b \times d_f}$ and $E_c^{t_i} = \textbf{Attention}(Q^i, K^i, V_c)$ with $V_c = C_i(F(X^t)) \in \mathbb{R}^{n_b \times d_c}$.

2) Interdomain Alignment Between Source and Target: To better align each source with target, a two-space alignment is designed with the guidance of cross-domain attention, i.e., distribution matching in feature space and semantic space, respectively.

As for the distribution matching in feature space, we finely align the multiple sources with the multihead attentive representations that inherit the pseudo-label of the corresponding target samples in the feature space by

$$\mathcal{L}_{\text{inter}}^{f} = \frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{K}\Big\|\mathbb{E}_{\tilde{X}_k^{s_i}\sim\tilde{P}_k^{s_i}}\left[\phi\left(F\left(\tilde{X}_k^{s_i}\right)\right)\right] - \mathbb{E}_{X_k^{t}\sim P_k^{t}}\left[\phi\left(E_{f,k}^{t_i}\right)\right]\Big\|_{\mathcal{H}}^{2} \quad (7)$$

where $E_{f,k}^{t_i}$ is drawn from target samples belonging to the kth pseudo-label. Explicitly, the cross-domain attention used as prior knowledge from source domain is nicely injected into MMD to guide the category-aware alignment.

In addition to the distribution matching of source and target in feature space, the matching of them is also expected to be semantically consistent, and we have

$$\mathcal{L}_{\text{inter}}^{s} = \frac{1}{M}\sum_{i=1}^{M}\text{KL}\left(C_i\left(F\left(\tilde{X}^{s_i}\right)\right)\middle\|E_c^{t_i}\right) \quad (8)$$

where $\text{KL}(\cdot\|\cdot)$ denotes the Kullback-Leibler divergence.

Combining (7) and (8), the objective of the interalignment between each source and target is formulated as

$$\mathcal{L}_{\text{inter}} = \mathcal{L}_{\text{inter}}^{f} + \mathcal{L}_{\text{inter}}^{s}. \quad (9)$$

D. Intradomain Alignment

To mitigate the domain discrepancy between source and target, each of sources is separately aligned with the same target domain via the interdomain loss $\mathcal{L}_{\text{inter}}$ by (9). But for the multihead attentive representations of target $\{E_f^{t_i}\}_{i=1}^{M}$, we still need to make a further integration which we name by intradomain alignment.

Particularly, instead of adopting the concatenation of the above multihead attention for integration as usual, a shared representation $\mathbf{O} \in \mathbb{R}^{K\times d_f}$ is explored to help correlate the multihead attentive representations of target into a unified one

$$\mathcal{L}_{\text{intra}} = \underbrace{\sum_{i=1}^{M}\left\|\mathbf{M}E_f^{t_i} - \mathbf{O}\right\|^2}_{\text{Association constraint}} + \underbrace{\lambda\left\|\mathbf{O} - \mathbf{C}^t\right\|^2}_{\text{Fitting constraint}} \quad (10)$$

where $\mathbf{M} = [m_{k,n}] \in \mathbb{R}^{K\times n_b}$ is a class-wise indicator with $m_{k,n} = 1/n_k$ when $\hat{y}_n^{t_i} = k$, and 0, otherwise. $\mathbf{C}^t = [c_1, c_2, \ldots, c_K] \in \mathbb{R}^{K\times d_f}$ is the target class prototype with $c_k = (1/M)\sum_{i=1}^{M} c_k^{t_i}$. As we can see from (10), the first term means to seek a class-wise shared representation to associate the multihead attentive representations of target, and the second term attempts to comply the class-wise shared representation with the target class prototype.

For target prediction with multiple sources, different sources have different influences on target domain. Thus, a weight learning method is developed to evaluate the contributions of different sources

$$\omega_i = \frac{\sum_{k=1}^{K}\text{sim}\left(c_k^{t_i}, \tilde{c}_k^{s_i}\right)}{\sum_{i=1}^{M}\sum_{k=1}^{K}\text{sim}\left(c_k^{t_i}, \tilde{c}_k^{s_i}\right)} \quad (11)$$

where $\text{sim}(c_k^{t_i}, \tilde{c}_k^{s_i})$ denotes the cosine similarity between target and source class prototype.

In addition to interdomain alignment and intradomain alignment, it is important to obtain the discriminative information from each source. Through supervised classification of labeled sources, discriminative features can be generated, and thus affect the adaptation. In this way, we use the cross-entropy loss to train the classifiers

$$\mathcal{L}_{\text{ce}} = -\sum_{i=1}^{M}\omega_i\mathbb{E}_{(x^{\tilde{s}_i}, y^{\tilde{s}_i})\sim(X^{\tilde{s}_i}, Y^{\tilde{s}_i})}[\mathbf{y}^{\tilde{s}_i}\log\left(C_i(F(x^{\tilde{s}_i}))\right)] \quad (12)$$

where $\mathbf{y}^{\tilde{s}_i}$ is one-shot encoded vector of $y^{\tilde{s}_i}$.

By further combining (9), (10), and (12), the total objective for the unified framework can be formulated as

$$\mathcal{L} = \beta\mathcal{L}_{\text{inter}} + \gamma\mathcal{L}_{\text{intra}} + \mathcal{L}_{\text{ce}} \quad (13)$$

where β and γ are two trade-off coefficients.

In the inference, the goal is to accurately classify a given target sample x^t, and its inferred label y^t can be given by

$$y^t = \sum_{i=1}^{M}\omega_i C_i(F(x^t)). \quad (14)$$

V. Experimental Results and Analysis

A. Experiment Setting

1) Datasets: We validate the effectiveness of our method for multisource scene classification on four remote sensing datasets. AID [26], Merced [11], NWPU [27], and PatternNet [28] were acquired over different regions with different sensors. Note that there are some shared categories in the four datasets. Following the setting in [18], we select twelve common classes to form the four datasets. As the same scene images may be denoted by different names due to different annotation ways, we have uniform name for each of the same scene image. Especially, the playground and stadium in AID are recombined to form the game space, the airplane and airport in NWPU are reassembled into airfield, the new farm in NWPU consists of circular farmland and rectangular farmland, the Stadium and ground track field in NWPU are combined to form game space, and the tennis court and football field in PatternNet compose the game space.

2) Implementation Details: The ResNet-50 [25] pretrained on ImageNet and follows by a single fully connected (FC) layer, are used as the backbone to extract features. The classifiers exploit another FC operation (MLP) to predict the results. Source domains and target domains share all the feature extractor parameters. We use stochastic gradient descent (SGD) with momentum of 0.9 and the learning rate of 10^{-3} as optimizer to train the network.

TABLE I

CLASSIFICATION ACCURACY (%) ON CROSS-DOMAIN TASKS WITH TWO SOURCES. THE BEST RESULT AMONG ALL COMPETITIVE METHODS IS HIGHLIGHTED WITH RED TYPE, WHILE THE SECOND PERFORMANCE IS MARKED IN GREEN

Standards	Method	N, P→A	A, P→N	A, M→P	N, A→M	Avg.
Source-only	-	93.56	88.16	89.67	90.58	90.49
Single best	$DDC_{(2014)}$ [29]	87.26	90.54	89.34	92.53	89.92
	$DAN_{(2015)}$ [30]	92.46	90.84	96.55	93.08	93.23
	Deep $Coral_{(2016)}$ [36]	92.85	88.82	96.28	90.58	92.13
	$CAN_{(2019)}$ [31]	97.05	92.71	97.32	94.59	95.42
Source combine	$CAN_{(2019)}$ [31]	97.54	93.46	98.12	95.43	96.13
	$CoVi_{(2022)}$ [32]	96.99	94.02	97.85	96.22	96.27
Multi source	$M^3SDA_{(2019)}$ [23]	94.01	91.33	95.46	92.34	93.29
	$MFSAN_{(2019)}$ [7]	96.96	92.77	99.05	95.51	96.07
	$LCt\text{-}MSDA_{(2020)}$ [24]	85.81	87.53	90.68	85.62	87.41
	$T\text{-}SVDNet_{(2021)}$ [33]	94.85	87.43	94.52	91.75	93.47
	$SSG_{(2022)}$ [34]	97.22	94.86	98.82	95.50	96.60
	$RRL_{(2023)}$ [35]	97.64	94.00	98.83	96.25	96.68
	Ours	97.90	95.26	99.45	96.67	97.32

3) Baselines: The following standards are adopted for comparison.

1) *Source Only:* All the source domains are leveraged to train the network and directly tests the network with target domain.
2) *Single Best:* The best performance of a single-source domain is obtained by the single-source domain adaptation methods.
3) *Source Combine:* All the source domains are combined into one domain to conduct single-source domain adaptation.
4) *Multisource:* All the source domains are employed to conduct MDA.

The compared methods belonging to single-best and source-combined include DDC [29], DAN [30], CAN [31], and CoVi [32], those belonging to multisource include M^3SDA [23], MFSAN [7], LCt-MSDA [24], T-SVDNet [33], SSG [34], and RRL [35].

B. Comparison With the State-of-the-Art

1) Cross-Domain Task With Two Sources: In the case of cross-domain task with two sources, four transfer tasks: N, P → A, A, P → N, A, M → P; N, A → M are performed to evaluate our method. Table I shows the classification accuracy on the four tasks of remote sensing datasets. The proposed model outperforms other state-of-the-art methods, which demonstrates the effectiveness of our method. In general, our method improves the mean accuracy to 97.32%, 0.64% higher than the next highest method RRL [35].

Meanwhile, t-SNE [37] is used to visualize the distribution of the learned representation. For the transfer task N, A → M, Fig. 5 shows the changes of feature distribution before and after domain adaptation. The discrepancy across source domains in Fig. 5(a) is relatively small. But there is a large shift between the multisource and the target. After adopting MDA, the multisource and the target are favorably combined, which means that our model can learn domain-invariant features well.

In addition to visualizing the representation distribution, we also calculate the discrepancy between two domains via $\mathcal{A}$-distance [38]. Fig. 6 illustrates the $\mathcal{A}$-distance between sources and target obtained by the source-only setting and our model for the task N, A → M. It can be seen that our model achieves the lowest $\mathcal{A}$-distance, which demonstrates CoI^{2}A has significant ability to reduce the discrepancy between different domains.

(a) (b)

Fig. 5. Visualizations on task N, A → M from (a) source-only setting and (b) our CoI^{2}A. The black dots denote the target representation (M). The red dots and yellow dots are source representations of N and A, respectively.

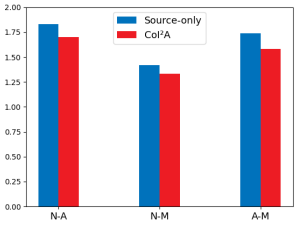

Fig. 6. $\mathcal{A}$-distance between N and A, N and M, and A and M.

TABLE II

CLASSIFICATION ACCURACY (%) ON CROSS-DOMAIN TASK WITH THREE SOURCES. THE BEST RESULT AMONG ALL COMPETITIVE METHODS IS HIGHLIGHTED WITH RED TYPE, WHILE THE SECOND PERFORMANCE IS MARKED IN GREEN

Standards	Method	M,P,N→A	A,M,P→N	N,A,M→P	P,A,N→M	Avg
Source-only	-	92.21	85.91	94.63	91.17	90.98
Single best	$DDC_{(2014)}$ [29]	90.54	87.26	92.53	89.34	91.42
	$DAN_{(2015)}$ [30]	92.46	90.84	96.55	94.25	93.53
	Deep $Coral_{(2016)}$ [36]	92.85	88.82	96.28	90.58	92.13
	$CAN_{(2019)}$ [31]	97.05	92.71	97.32	95.43	95.63
Source combine	$CAN_{(2019)}$ [31]	98.01	93.25	97.55	96.53	96.34
	$CoVi_{(2022)}$ [32]	96.43	92.33	98.03	96.36	95.29
Multi source	$M^3SDA_{(2019)}$ [23]	92.12	89.74	95.05	93.56	92.62
	$MFSAN_{(2019)}$ [7]	96.11	91.19	98.37	96.33	95.50
	$LCt\text{-}MSDA_{(2020)}$ [24]	93.61	86.36	91.70	86.17	89.46
	$T\text{-}SVDNet_{(2021)}$ [33]	91.46	93.31	97.65	91.41	93.46
	$SSG_{(2022)}$ [34]	97.25	95.57	98.24	96.68	96.79
	$RRL_{(2023)}$ [35]	96.83	93.81	98.91	97.79	96.84
	Ours	98.64	94.33	99.62	97.92	97.63

2) Cross-Domain Tasks With Three Sources: Four experiments on cross-domain task with three sources: A, M, P → N, M, P, N → A, P, N, A → M, N, A, M → P are also conducted. Table II lists the experiment results of our CoI^{2}A and other baselines on the four transfer tasks. The average accuracy of our model is 97.63%, achieving the best performance and surpassing all the compared methods. Compared with the latest pseudo-label-based approach SSG [34], the proposed CoI^{2}A achieves the competitive performance under most cross-domain tasks. By comparison with other baselines, it is clearly shown that the proposed CoI^{2}A can be well capable of adapting to the MDA task on remote sensing datasets. We also observed that for the task M, P, N → A, the result of using three sources is worse than that of using two sources, probably because multiple sources bring some noise information resulting in performance degradation.

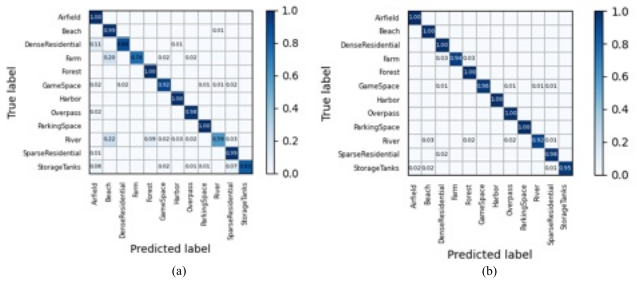

Fig. 7. Visualization of the confusion matrix of merced for the task P, N, A → M. (a) Source-only. (b) Our proposed CoI^{2}A.

TABLE III

ACCURACY (%) OF ABLATION STUDY ON THE EFFECTIVENESS OF $\mathcal{L}_s$, $\mathcal{L}_{\text{inter}}$, AND $\mathcal{L}_{\text{intra}}$

$\mathcal{L}_s$	$\mathcal{L}_{inter}$	$\mathcal{L}_{intra}$	→A	→N	→P	→M
×	√	√	96.36	92.48	96.51	96.33
√	×	×	93.65	88.81	93.74	93.08
√	×	√	94.41	89.40	94.78	94.21
√	√	×	96.20	91.49	97.52	96.25
√	√	√	**98.64**	**94.33**	**99.62**	**97.92**

For the task P, N, A → M, Fig. 7 shows the confusion matrices of classification in target domain under two kinds of settings: source-only and our proposed CoI^{2}A. As we can see from Fig. 7(a), the errors of classification on River and Farm are relatively large, especially it is difficult for the classifiers to distinguish River from Beach. After interdomain and intradomain alignments, the errors are obviously reduced and the accuracy of CoI^{2}A on River increases from 0.59 to 0.92, as well as the accuracy rates of other categories have a great improvement. The above results further confirm the effectiveness of the proposed CoI^{2}A.

C. Performance Analysis

1) Ablation Studies: Several experiments are conducted on three-source cross-domain tasks A, M, P → N, M, P, N → A, P, N, A → M, N, A, M → P to demonstrate the role of each part of our method. We mainly analyze the effectiveness of intersource alignment $\mathcal{L}_s$ in (4), interdomain alignment between source and target $\mathcal{L}_{\text{inter}}$ and intradomain alignment $\mathcal{L}_{\text{intra}}$ in (13). One of these losses in turn is removed to reflect its contribution to the learning performance, and the specific results are shown in Table III.

Obviously, without using any one of alignments, the adaptation performance degrades in most cases. Especially, the accuracies on M, P, N → A, P, N, A → M, and N, A, M → P are improved by about 2.0% with $\mathcal{L}_{\text{intra}}$. Moreover, by adding $\mathcal{L}_{\text{inter}}$, the interdomain alignment across source and target improves the accuracy on each of tasks by about 4.0%. Meanwhile, we can find that the interdomain alignment makes more contributions than the intradomain alignment, since the distribution discrepancy between multisource and target has already been well reduced by interdomain alignment. As a supplement, the intradomain alignment is leveraged to further improve the performance of the proposed model.

Experiments are conducted on the tasks with three sources to demonstrate the effectiveness of the cross-domain attention.

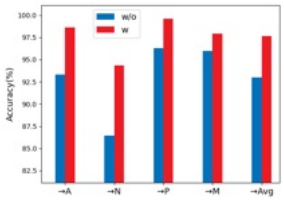

Fig. 8. Effectiveness of cross-domain attention. "w/o" denotes the model without using cross-domain attention. "w" means the proposed model.

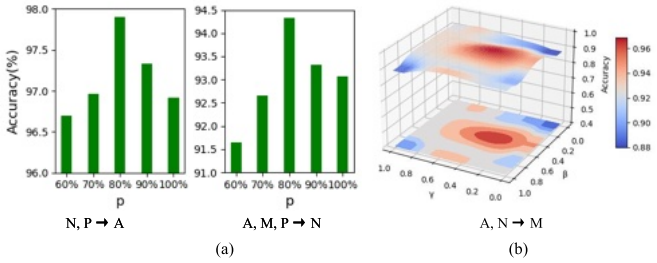

Fig. 9. Sensitivity analysis. (a) Sensitivity analysis of p on different tasks: A, M, P → N, and N, P → A. (b) Sensitivity analysis of β and γ on task A, N → M.

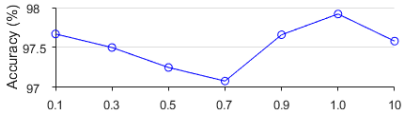

Fig. 10. Performance on different trade-off coefficients of the fitting constraint term on the task P, N, A → M.

As illustrated in Fig. 8, for comparison with directly aligning the multisource with target (i.e., w/o cross-domain attention), we can obtain about 4.60% improvement in average accuracy using cross-domain attention, which confirms that the proposed cross-attention is indeed effective for improving the ability of adaptation.

2) Hyperparameters Analysis: To investigate the effect of cross-source data augmentation, sample complexity experiments are conducted. For each source domain, p is set to 0.6:0.1:1.0. Note that if the number of the sample in the other domain is less than the one to be selected, the entire samples will be selected. As shown in Fig. 9(a), we can observe that the accuracy first keeps increasing and reaches the best at $p = 80\%$ when p varies from 60% to 80%, and then it decreases by a large margin when p is set to 100%, i.e., without applying cross source data augmentation. Explicitly, it indicates that the proposed cross-source data augmentation is essentially beneficial to boosting the model training.

Furthermore, we also evaluate the sensitivity of β and γ that are used to balance the interdomain alignment loss and intradomain alignment loss. Experiments on task A, N → M are conducted while varying β and γ, and the final sensitivity map is shown in Fig. 9(b). It can be noticed that the performance of the model degrades as β and γ approach 0, which means that the interdomain and intradomain constraints are crucial for the model. When β is set to 0.5 and γ equals to 0.3, the accuracy achieves the best. Thus, we set β as 0.5 and γ as 0.3.

Besides, to analyze the sensitivity of the trade-off parameter λ in 10, we vary it from 0.1 to 10, the obtained performance is shown in Fig. 10. It can be seen that the performance changes within 1% gain. As the result is optimal when λ is 1, we set λ to 1 in all the experiments.

VI. Conclusion

In this article, toward remote sensing scene classification, we propose a novel framework named CoI^{2}A for MDA, in which three alignments strategies, i.e., intersource, interdomain, and intradomain alignment, collaborate to work off the source shift and domain shift. The cross-domain attention is proposed to improve the discriminability of the representation learned for target. With the guidance of cross domain, each source aligns with the target in a category-aware way. And to further integrate the multihead representations of target based on cross attention, intradomain alignment is adopted to promote consistency across multihead representations of target.

References

[1] Y. Zhong, X. Wang, S. Wang, and L. Zhang, "Advances in spaceborne hyperspectral remote sensing in China," *Geo-Spatial Inf. Sci.*, vol. 24, no. 1, pp. 95–120, Jan. 2021.

[2] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, vol. 19, 2006, pp. 1–8.

[3] S. Sun, H. Shi, and Y. Wu, "A survey of multi-source domain adaptation," *Inf. Fusion*, vol. 24, pp. 84–92, Jul. 2015.

[4] H. Liao, "Speaker adaptation of context dependent deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 7947–7951.

[5] X. Yao, S. Zhao, P. Xu, and J. Yang, "Multi-source domain adaptation for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3253–3262.

[6] H. Zhao, S. Zhang, G. Wu, J. M. F. Moura, J. P. Costeira, and G. J. Gordon, "Adversarial multiple source domain adaptation," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.

[7] Y. Zhu, F. Zhuang, and D. Wang, "Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 5989–5996.

[8] R. Xu, Z. Chen, W. Zuo, J. Yan, and L. Lin, "Deep cocktail network: Multi-source unsupervised domain adaptation with category shift," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3964–3973.

[9] J. Tao, D. Song, S. Wen, and W. Hu, "Robust multi-source adaptation visual classification using supervised low-rank representation," *Pattern Recognit.*, vol. 61, pp. 47–65, Jan. 2017.

[10] C. Chen, W. Xie, Y. Wen, Y. Huang, and X. Ding, "Multiple-source domain adaptation with generative adversarial nets," *Knowl.-Based Syst.*, vol. 199, Jul. 2020, Art. no. 105962.

[11] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2010, pp. 270–279.

[12] B. Zhao, Y. Zhong, L. Zhang, and B. Huang, "The Fisher kernel coding framework for high spatial resolution scene classification," *Remote Sens.*, vol. 8, no. 2, p. 157, Feb. 2016.

[13] S. Chaib, H. Liu, Y. Gu, and H. Yao, "Deep feature fusion for VHR remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4775–4784, Aug. 2017.

[14] J. Xie, N. He, L. Fang, and A. Plaza, "Scale-free convolutional neural network for remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6916–6928, Sep. 2019.

[15] S. Song, H. Yu, Z. Miao, Q. Zhang, Y. Lin, and S. Wang, "Domain adaptation for convolutional neural networks-based remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1324–1328, Aug. 2019.

[16] Y. Li, W. Hu, H. Li, H. Dong, B. Zhang, and Q. Tian, "Aligning discriminative and representative features: An unsupervised domain adaptation method for building damage assessment," *IEEE Trans. Image Process.*, vol. 29, pp. 6110–6122, 2020.

[17] M. Xu, M. Wu, K. Chen, C. Zhang, and J. Guo, "The eyes of the gods: A survey of unsupervised domain adaptation methods based on remote sensing data," *Remote Sens.*, vol. 14, no. 17, p. 4380, Sep. 2022.

[18] M. Al Rahhal, Y. Bazi, T. Abdullah, M. Mekhalfi, H. AlHichri, and M. Zuair, "Learning a multi-branch neural network from multiple sources for knowledge adaptation in remote sensing imagery," *Remote Sens.*, vol. 10, no. 12, p. 1890, Nov. 2018.

[19] X. Lu, T. Gong, and X. Zheng, "Multisource compensation network for remote sensing cross-domain scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2504–2515, Apr. 2020.

[20] J. Yang, R. Yan, and A. G. Hauptmann, "Cross-domain video concept detection using adaptive svms," in *Proc. 15th ACM Int. Conf. Multimedia*, Sep. 2007, pp. 188–197.

[21] J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. Wortman, "Learning bounds for domain adaptation," in *Proc. 20th Int. Conf. Neural Inf. Process. Syst.*, vol. 20, 2007, pp. 1–8.

[22] Y. Mansour, M. Mohri, and A. Rostamizadeh, "Domain adaptation with multiple sources," in *Proc. 21st Int. Conf. Neural Inf. Process. Syst.*, vol. 21, 2008, pp. 1–8.

[23] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1406–1415.

[24] H. Wang, M. Xu, B. Ni, and W. Zhang, "Learning to combine: Knowledge aggregation for multi-source domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 727–744.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[26] G.-S. Xia et al., "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[27] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

[28] W. Zhou, S. Newsam, C. Li, and Z. Shao, "PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 197–209, Nov. 2018.

[29] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, *arXiv:1412.3474*.

[30] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 97–105.

[31] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4888–4897.

[32] J. Na, D. Han, H. J. Chang, and W. Hwang, "Contrastive vicinal space for unsupervised domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 92–110.

[33] R. Li, X. Jia, J. He, S. Chen, and Q. Hu, "T-SVDNet: Exploring high-order prototypical correlations for multi-source domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9971–9980.sss

[34] J. Yuan et al., "Self-supervised graph neural network for multi-source domain adaptation," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 3907–3916.

[35] S. Chen, L. Zheng, and H. Wu, "Riemannian representation learning for multi-source domain adaptation," *Pattern Recognit.*, vol. 137, May 2023, Art. no. 109271.

[36] B. Sun and K. Saenko, "Deep CORAL: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 443–450.

[37] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.

[38] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, nos. 1–2, pp. 151–175, May 2010.