

Received September 21, 2021, accepted October 19, 2021, date of publication October 26, 2021, date of current version November 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3123281

# Metric Information Matrix for Maximum Mean Discrepancy for Domain Adaptation

WENJUAN REN<sup>1</sup>, SHIE ZHOU<sup>2</sup>, ZHANPENG YANG<sup>1</sup>, QUAN SHI<sup>1</sup>, XIAN SUN<sup>1</sup>, AND LUYI YANG<sup>2</sup>

<sup>1</sup>Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China

<sup>2</sup>Foreign Language Department, Information and Engineering University, Zhengzhou 450000, China

Corresponding authors: Wenjuan Ren (renwj@aircas.ac.cn) and Zhanpeng Yang (zhanpengyang@mail.ie.ac.cn)

**ABSTRACT** In this paper, we focus the problem of unsupervised domain adaptation which transfers knowledge from a well-labeled source domain to an unlabeled target domain with distinctive distributions. Based on Gromov-Hausdorff's theory, we proposed two kinds of feature mappings in the model of joint distribution adaptation by embedding the original feature subspace to a common subspace. It can be seen as a part of feature embedding used for the models based feature alignment. Our experiments show that constructed mappings have the abilities to alleviate the feature discrepancy and mitigate the distribution shift between source domain and target domains.

**INDEX TERMS** Domain adaptation, metric information matrix, maximum mean discrepancy, Toeplitz matrix, convolutional filter mask.

## I. INTRODUCTION

In some application scenarios, one can only access to information on source domains, but there doesn't exist direct information about a target domain. Domain adaptation (DA) aims at training a model in source domain and applying it to a new target domain. The key of most domain adaptation methods is to learn a transformation on the features to reduce the discrepancy of the distributions between the source and the target datasets. There are different situations in real-world problems: 1) The marginal distributions are different, while the conditional distributions are similar. 2) The marginal distributions are similar, while the conditional distributions are different. 3) Both the marginal and the conditional distributions are different. The approaches of domain adaptation fall into two major categories: based on optimization and based on deep learning.

Maximum mean discrepancy (MMD) is a widely adopted tool based on optimization to measure the discrepancy between the source and target domain distributions. A popular domain adaptation is feature-based, which projects different domain data into a shared subspace to minimize their discrepancy, usually measured by MMD [1]. Li *et al.* proposed a novel heterogeneous domain adaptation method that can optimal both feature discrepancy and distribution divergence in a unified objective function, and the notion of progressive alignment was given to learn a new transferable feature space by dictionary-sharing coding and align the distribution gaps

The associate editor coordinating the review of this manuscript and approving it for publication was Seifedine Kadry.

on the new space [2]. Joint distribution adaptation (JDA) and its generalized discriminative joint probability MMD (DJP-MMD) were proposed to measuring the distribution shift between domains [4], [5].

Deep domain method based on neural network currently provide some new solutions to domain adaptation. Hoffman *et al.* derived some new normalized solutions with strong theoretical guarantees for the cross-entropy loss and other similar losses, and present a number of novel contributions to the multiple-source adaptation problem [6]. In order to character the common feature subspace in more details, an Adversarial Tight Match method with a novel distance loss named Maximum Density Divergence, was proposed in adversarial domain adaptation which enjoys the benefits of both adversarial training and metric learning [7]. Moreover, a quintessential domain adaptation framework on deep learning is the Deep Transfer Network (DTN), where the highly flexible deep neural networks are used to implement such a distribution matching process [8].

In this paper we focus on the approaches based on MMD for domain adaptation. We introduced a graph structure, named metric information matrix, which can be used to character the geometric information of a set or a feature space. Then we used the invariant property to character the local metric information of a feature space for domain adaptation. We also provided some additional notions such as Gromov-Wasserstein barycenters and  $k$ -nearest neighbour graph to improve the MMD-like algorithms based metric matrices. The main contributions of the paper are:

(1) We introduce the notion of metric information matrix constructed by some metric matrices, Toeplitz Matrix and Markov Matrix. It can be used as a based block to represent the geometric structural information of the source domain and target domain.

(2) We provide a strategy that can help us to search a common subspace where one can optimize easily the between-domain transferability and the between-class discriminability by using MMD-like methods. Moreover, associated with MMD, it can be seen as a part of feature embedding in the adversarial domain adaptation models. An experiment is given to verify the effectiveness of the hybrid adversarial model.

(3) One can combine the proposed approach with the MMD-like approaches in a straightforward way by transform the original domains to a common subspace. And we show by extensive experiments that the proposed method outperforms the baseline methods in accuracy on different datasets.

## II. MATHEMATICAL FOUNDATIONS

Let  $\mathcal{D}_s$  be the source domain with  $n_s$  labeled samples  $\{X_s, Y_s\} = \{(X_{s,i}, Y_{s,i})\}_{i=1}^{n_s}$ . Let  $\mathcal{D}_t$  be the target domain with  $n_t$  unlabeled samples  $X_t = \{x_{t,j}\}_{j=1}^{n_t}$ , where  $x \in \mathbb{R}^{d \times 1}$  is the feature tensor, and let  $y$  be the label, with  $y \in \{1, \dots, C\}$  for  $C$ -class classification. In this paper we only consider homogeneous transfer learning, i.e.,  $\mathcal{X}_s = \mathcal{X}_t$  and  $\mathcal{Y}_s = \mathcal{Y}_t$ . Domain adaptation aims to search a mapping  $H$  that brings  $H(X_s)$  and  $H(X_t)$  together in some means, so that the well-trained classifier works well on both  $H(X_s)$  and  $H(X_t)$ .

Consider a feature mapping  $g$  that maps  $x$  to a feature space. The general objective function of domain adaptation can be defined as following

$$\min_h d_{S,T} + \lambda \mathcal{R}(h), \quad (1)$$

where  $d_{S,T} = d(P(X_s, Y_s), P(X_t, Y_t))$  is a discrepancy metric between the source and target domain distributions,  $\mathcal{R}(h) = \|h\|_F^2$  controls the mapping complexity, and  $\lambda$  is a regularization parameter.

### A. MATHEMATICAL FOUNDATIONS

The metric matrix theory is an important research object in metric measure geometry, and it is a tool to study the Gromov's theory of convergence and concentration of metrics and measures. By the definition of distance matrix, it is easy to see that any translation, rotation, or reflection transform will lead to the same metric matrix such as Euclidean distance matrix. We can use this property to character the local metric information of a set (or a domain).

Let  $(\mathcal{X}, d)$  be a metric space. We introduce the following definition to character the metric information of a domain data.

#### 1) METRIC INFORMATION MATRIX

For a metric space  $\mathcal{X}$  and a natural number  $N$ , the distance matrix  $\mathcal{M}_N(\mathcal{X})$  of  $\mathcal{X}$  in order  $N$  is defined to be the set of

symmetric matrices  $\{d_{\mathcal{X}}(x_i, x_j)\}_{ij}^N$  of order  $N$ , where  $x_i, i = 1, 2, \dots, N$  run over all points in  $\mathcal{X}$ . By the definition of metric (distance) matrix, we can obtain a set of  $N \times N$  matrices for order  $N$ . Then for a feature space  $(X, \mathcal{D})$  of the domain we can define a metric information matrix (MIM) by

$$\mathcal{M}_N(X) := \left\{ \mathcal{D}(x_i, x_j) \right\}_{i,j=1}^N, \quad x_{i,j} \in X \quad (2)$$

where  $\mathcal{D}$  is a metric. If  $\mathcal{X}$  is compact, then the distance matrix  $\mathcal{K}_N(\mathcal{X})$  is compact. It can be used to character the geometry feature of a set, and by Gromov-Housdorff (or Gromov-Wasserstein) distance one can measure the distance between the two metric information matrices.

#### 2) TOEPLITZ MATRIX

A Toeplitz matrix or diagonal-constant matrix is a matrix in which each descending diagonal from left to right is constant.

#### 3) MARKOV MATRIX

If the probability of moving from  $i$  to  $j$  in one time step is  $Pr(j|i)$ , the stochastic matrix  $P$  is given by using  $P_{i,j}$  as the  $i$ -th row and  $j$ -th column element.

#### 4) GROMOV-WASSERSTEIN BARYCENTERS

Gromov-Wasserstein barycenters of measured similarity matrices  $(C_s)_{s=1}^S, C_s \in \mathbb{R}^{N_s \times N_s}$  can be defined as a F chet mean

$$\min_{C \in \mathbb{R}^{N \times N}} \sum_s \lambda_s GW_\varepsilon(C, C_s, p, p_s). \quad (3)$$

For a large class of losses (KL loss or  $L_2$  loss), one can compute the global minimizer by the Proposition 3 in paper [3].

#### 5) MAXIMUM MEAN DISCREPANCY

Maximum mean discrepancy is a distance associated with the embedding probabilities in a reproducing kernel Hilbert space. For the embedding of  $P$  and  $Q$ , Maximum Mean Discrepancy  $MMD_k(P, Q)$  can be defined as the distance between the corresponding mean elements, i.e.,

$$MMD_k(P, Q) = \|\mu_P - \mu_Q\|_{\mathcal{H}}. \quad (4)$$

## B. DOMAIN ADAPTATION

Domain adaptation aims at learning a model in the source domain and applying it to the target domain. Many domain adaptation methods try to reduce the distribution shift between the source and the target datasets. A discriminative joint probability maximum mean discrepancy for domain adaption was proposed in paper [4]. The object function for domain adaptation of DJP-MMD can be defined as

$$\begin{aligned} \arg \min_A & \left\| A^T X_s N_s - A^T X_t N_t \right\|_F^2 \\ & - \mu \left\| A^T X_s M_s - A^T X_t M_t \right\|_F^2 + \lambda \|A\|_F^2 \\ \text{s.t. } & A^T X H X^T A = I, \end{aligned} \quad (5)$$

where  $H = I - 1_n$  is the centering matrix, in which  $n = n_s + n_t$  and  $1_n \in \mathbb{R}^{n \times n}$  is a matrix with all elements being  $\frac{1}{n}$ .

### III. DOMAIN ADAPTATION WITH INFORMATION MATRIX

To facilitate domain adaptation, we improve MMD-like models by minimize the object function on a common subspace. Generally speaking, for any  $N \leq |X|$ ,  $X$  and  $\mathcal{M}_N(X)$  cannot be placed in one-to-one correspondence. If  $N \neq |X|$ , then MMD algorithms based on labels and pseudo-labels doesn't work on this subspace. If  $N = |X|$ ,  $\mathcal{M}_N(X)$  is also called similarity matrix which is indeed a symmetric nonnegative adjacency matrix with zero diagonal. In this paper we construct a neighbour graph of domains and perform MMD-like models for domain adaptation on the graph.

#### A. I-MIM METHOD

##### 1) NEAREST NEIGHBOUR GRAPH

For a given set  $V = (v_1, v_2, \dots, v_n)$ , a sparse affinity matrix  $A \in \mathbb{R}^{n \times n}$  with elements

$$a_{ij} := \begin{cases} [\mathcal{D}(v_i, v_j)]_+^\gamma, & \text{if } i \neq j \wedge v_i \in NN_k(v_j), \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

is constructed, where  $NN_k$  denotes the set of  $k$  nearest neighbors in  $X$ , and  $\gamma$  is parameter following recent work on manifold-based search [9]. Because this matrix is constructed by a metric function, we call it metric information matrix (i-MIM).

A linear mapping is given as

$$\mathcal{L}(X) = \gamma \cdot X + (1 - \gamma) \cdot \mathcal{M}(X), \quad (7)$$

where  $\gamma$  is a trade-off parameter.

For clarity, we summarize the learning algorithms in Algorithms 1.

TABLE 1. i-MIM Algorithm 1.

<b>Algorithm 1:</b> i-MIM method.
<b>Input:</b> Data $X_s, Y_s, X_t, Y_t, \gamma$ .
1. Construct nearest neighbour graph by manifold-based search method and calculate metric information matrix $\mathcal{M}_s, \mathcal{M}_t$ ;
2. Optimize DJP-MMD for feature matrices.
3. Take a linear transformation by Eq. (7);
<b>Output:</b> $Y_t$ , estimated target domain labels.

##### 2) ii-MIM

We introduce a modified algorithm for domain adaptation that can be summarized as follows. We construct a nearest neighbour graph on the source domain and target domain. Then we perform the discriminative joint probability maximum mean discrepancy for domain adaptation on two graphs. The idea is developed below, while a graphical overview of the proposed approach is shown in Figure 1. Then the modified object function of the discriminative joint probability maximum mean discrepancy for domain adaptation can be defined as

$$\begin{aligned} \arg \min_A & \left\| A^T \mathcal{M}_s N_s - A^T \mathcal{M}_t N_t \right\|_F^2 \\ & - \mu \left\| A^T \mathcal{M}_s M_s - A^T \mathcal{M}_t M_t \right\|_F^2 + \lambda \|A\|_F^2 \\ \text{s.t. } & A^T \mathcal{L}_{s,t} H \mathcal{M}_{s,t}^T A = I. \end{aligned} \quad (8)$$

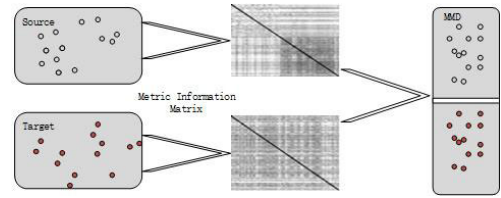


FIGURE 1. An illustrative example of the i-MIM method.

#### B. II-MIM METHOD

The i-MIM only characters the metric relationship of points on each domain and that will miss the information of the original data. We provide a strategy to mapping the source and target data into a common subspace. We construct an information matrix which can be constructed as a nearest neighbor graph or as a matrix with some special properties such as Toeplitz matrix. We introduce the definition of convolutional filter masks proposed in [10].

##### 1) CONVOLUTIONAL FILTER MASKS

A convolutional filter mask  $W$  is a sequence support in  $\{0, 1, \dots, s\}$  for some  $s \in \mathbb{N}$  called filter length, involving only  $s + 1$  free parameters. Such a filter mask  $w = (w_k)_{k=-\infty}^{\infty}$  support in  $\{0, 1, \dots, s\}$  for some  $s \in \mathbb{N}$  satisfies  $w_k = 0$  for  $k \notin [0, s]$  and convoluting with it leads to a Toeplitz type  $(D + s) \times D$  convolutional matrix  $T$ .

##### 2) ii-MIM

Let  $(X_t, N_t)$  be the finite state space of the target domain, where  $N_t$  is the cardinality of  $X_t$ . Let  $(X_s, N_s)$  be the finite state space of the source domain, where  $N_s$  is the cardinality of  $X_s$ . We then construct the similarity matrices or the convolutional filter mask associated with Gromov-Wasserstein barycenter of the source domain and the target domain and denote as  $M_s := \mathcal{M}(X_s)$  and  $M_t := \mathcal{M}(X_t)$ . Using the information matrices  $M_s, M_t$  to describe the Markov chains, then we can construct the stochastic matrix  $\mathcal{S}_s := \mathcal{S}(M_s)$  and  $\mathcal{S}_t := \mathcal{S}(M_t)$ . Then the linear mapping is given as

$$\mathcal{L}(X) = \gamma \cdot X + (1 - \gamma) \cdot \mathcal{S}[\mathcal{M}](X), \quad (9)$$

where  $\gamma$  is a trade-off parameter. The matrix  $\mathcal{S} \cdot \mathcal{M}$  is named the metric information matrix (ii-MIM).

Then we perform the discriminative joint probability maximum mean discrepancy for domain adaptation on two graphs. These ideas are developed below, while a graphical overview of the proposed approach is shown in Figure 2. Then the modified object function of the discriminative joint probability maximum mean discrepancy for domain adaptation can be defined as

$$\begin{aligned} \arg \min_A & \left\| A^T \mathcal{L}(X_s) N_s - A^T \mathcal{L}(X_t) N_t \right\|_F^2 \\ & - \mu \left\| A^T \mathcal{L}(X_s) M_s - A^T \mathcal{L}(X_t) M_t \right\|_F^2 + \lambda \|A\|_F^2 \\ \text{s.t. } & A^T \mathcal{L}(X_{s,t}) H \mathcal{L}(X_{s,t})^T A = I. \end{aligned} \quad (10)$$

The optimal problems can be optimized by the algorithm provided in the paper [4].

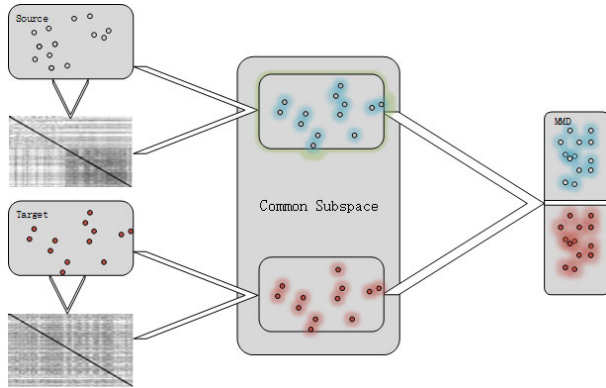


FIGURE 2. An illustrative example of the *i*-MIM method.

TABLE 2. *ii*-MIM Algorithm 2.

Algorithm 2: <i>ii</i> -MIM method.
<b>Input:</b> Data $X_s, Y_s, X_t, Y_t, \gamma$ .
1. Calculate Euclidean distance matrix $\mathcal{M}_s, \mathcal{M}_t$ ;
2. Calculate Gromov-Wasserstein barycenters $\mathcal{GW}(\mathcal{M}_s, \mathcal{M}_t)$ ;
3. Construct Toeplitz matrix induced by $\mathcal{GW}(\mathcal{M}_s, \mathcal{M}_t)$ ;
4. Generalize the convolutional filter masks by Eq. 9;
5. Optimize DJP-MMD for feature matrices.
<b>Output:</b> $\hat{Y}_t$ , estimated target domain labels.

## IV. EXPERIMENTS

In this section, some experiments are given to show the effectiveness of the modified algorithms.

### A. DATASETS

#### 1) OBJECT RECOGNITION

Office+Caltech includes four domains: Caltech (C), Amazon (A), Webcam (W), and DSLR (D). In this paper we adopt the SURF features, and obtain  $4 \times 3 = 12$  different domain adaptation tasks by randomly selecting a source domain and a target domain.

#### 2) COIL

COIL includes 20 objects with 1,440 images. The images of each object are taken 5 degrees apart as the object was rotated on a turntable, and each object has 72 images of  $32 \times 32$  pixels. The dataset is partitioned into two equal subsets (COIL1 and COIL2) with different distributions.

#### 3) DIGIT RECOGNITION

USPS and MNIST are two public digit recognition datasets with different resolutions. Our experiments used the public USPS and MNIST datasets released by Long *et al.* [5], which randomly sampled 1,800 images in USPS and 2,000 images in MNIST. They both have 10 classes of digits, with different distributions.

#### 4) MODULATION RECOGNITION

Experiments are conducted on the public dataset named RadioML [11], which is commonly used for domain

adaptation modulation recognition. It contains 11 modulation classes, 8PSK, BPSK, QPSK, QAM16, QAM64, AM-DSB, CPFSK, GFSK, LFM, WBFM, PAM4. Each example waveform is of size  $2 \times 128$ . GNU Radio models for time varying multi-path fading of the channel impulse response, random walk drifting of carrier frequency oscillator and additive Gaussian white noise are added, and the signals are passed through harsh channels models which introduce unknown scale, translation and dilation. RadioML consists of two parts, named RadioML2016.10a (10a) and RadioML2016.04c (04c). Each class contains 10,000 samples for 20 signal-to-noise ratios levels.

### B. SETTING

In our experiments we compared the models for unsupervised domain adaptation tasks, JPDA [4], TCA [12], JDA [5], DICE [13], PACET [14] and BDA (which used the A-distance [15] to compute the weight, instead of grid search in [16]), adversarial discriminative domain adaptation [17] and adversarial unsupervised domain adaptation [18]. We chose *k*-nearest neighbor model as the classifier. The parameter settings in [12] were used for TCA, JDA and BDA. We fixed  $p = 100$  and  $T = 20$  in all experiments, and the regularization parameter  $\lambda = 1$  with linear kernel for Office+Caltech dataset,  $\lambda = 0.1$  with primal kernel for other datasets. In the experiment of deep domain adaptation we used an improved neural network model in our previous work [19] to feature extraction.

### C. EXPERIMENTAL RESULTS

#### 1) DOMAIN CLASSIFICATION ACCURACY

The classification accuracies of the following algorithms are given in Table 1 and Table 2. The accuracies of proposed algorithms of the *i*-MIM and the *ii*-MIM are respectively 56.87% and 56.50%, which both outperform the baseline methods. It clearly demonstrates the effectiveness of the proposed methods. Furthermore, we can obtain that:

(1) The *i*-MIM method always performs better in the feature subspace constructed by nearest neighbour graph, and it shows that the geometry feature subspace can replace the original domain as a latent subspace to perform the MMD-like methods.

(2) As can be seen from the results of the *ii*-MIM method, convolutional filter masks can be used to construct a linear transform to map the original feature space to a common subspace which is more easier to optimize the domain adaptation approaches based on MMD.

(3) In the frame of deep domain adaptation the proposed method based on MIM-like feature embedding still outperforms the baseline methods (ADDA, GR-AUDA or FA-AUDA) in a 1-D feature vector space extracted by a deep neural network.

(4) The feature embeddings associated with metric information matrix can obtain a more transferrable and also



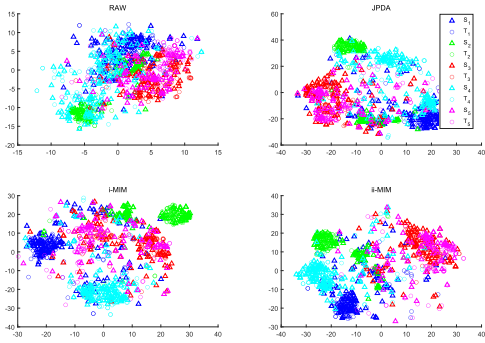


FIGURE 3. t-SNE visualization of the first five classes' distributions before and after different DA approaches, when transferring Caltech (source) to Amazon (target).

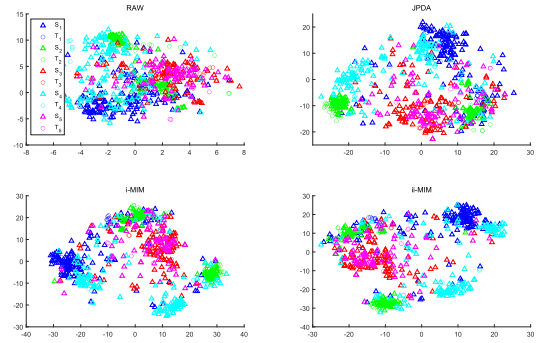


FIGURE 6. t-SNE visualization of the first five classes' distributions before and after different DA approaches, when transferring Caltech (source) to Dslr(target).

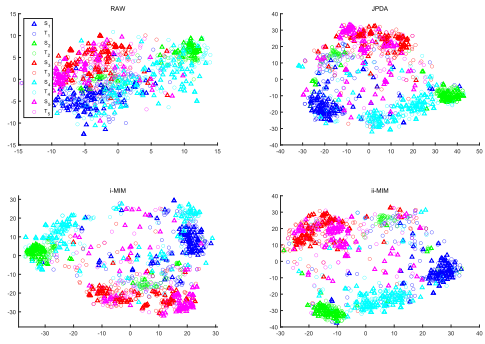


FIGURE 4. t-SNE visualization of the first five classes' distributions before and after different DA approaches, when transferring Amazon (source) to Caltech (target).

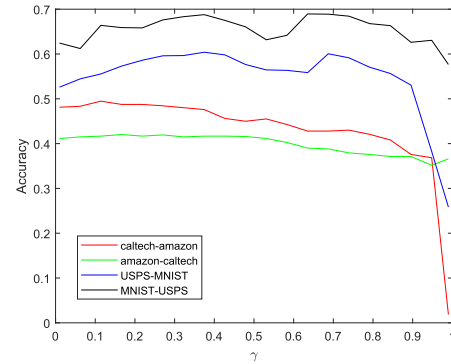


FIGURE 7. Average classification accuracies with the trade-off parameter  $\gamma$ .

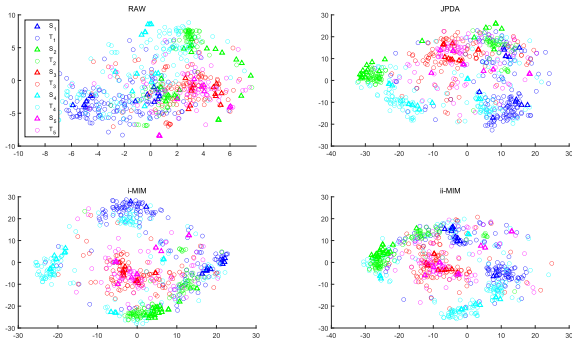


FIGURE 5. t-SNE visualization of the first five classes' distributions before and after different DA approaches, when transferring Dslr (source) to Caltech (target).

more discriminative feature mapping for cross-domain visual adaptation.

## 2) FEATURE VISUALIZATION

We adopt the t-SNE method [20] to visualize the learned feature on the C-W task, W-C task, D-C task and C-D task on Office+Caltech dataset. The visualization results are plotted in Figure 3-6. From the results, we can obtain the following observations:

(1) Figure 3 shows the results of the first five classes' data distributions when transferring Caltech (source) to Amazon

(target), before and after different distribution adaptation approaches, where RAW denotes the raw data distribution. For the raw distribution, the samples of the source and the target domains are mixed together. After domain adaptation, the proposed approaches bring data distributions of the source and the target domains together, and also keeps samples from different classes well-separated.

(2) Compared with JPDA, the proposed methods based MIM not only align the distributions between the source and target data but also enhance the discriminative power of the target data (See  $S_4 \rightarrow T_4$ ).

(3) i-MIM and ii-MIM can both align the distributions between the source and target data. And i-MIM is a better feature representation than ii-MIM and JPDA (See  $(S_3 \rightarrow T_3)$  or  $(S_5 \rightarrow T_5)$  in Figure 3).

## 3) PARAMETERS SENSITIVITY AND CONVERGENCE

We consider the parameter sensitivity and convergence of the proposed models on different datasets. The main adjustable parameter  $\gamma$  was studied for two classes approaches base MIM. The result is shown in Figure 7. The method of ii-MIM is robust when  $\gamma \in [0.1, 0.6]$ . Another parameter is the metric used in metric information matrix. Here we compare 12 different kinds of distances, Euclidean, suare-dEuclidean, seuclidean, mahalanobis, cityblock, minkowski,

TABLE 3. Classification accuracy of the algorithms.

Dateset	Source	Target	TCA	JDA	BDA	JPDA	i-MIM	ii-MIM
COIL	COIL1	COIL2	88.47	89.31	89.44	92.08	<b>95.14</b>	94.03
	COIL2	COIL1	85.83	88.47	88.33	89.86	<b>94.31</b>	93.61
Office+Caltech	C	A	38.20	44.78	44.57	47.60	49.06	<b>49.16</b>
		W	38.64	41.69	40.34	45.76	46.78	<b>47.12</b>
		D	41.40	45.22	45.22	46.50	46.50	<b>49.04</b>
	A	C	37.76	39.36	39.27	40.78	41.50	<b>41.59</b>
		W	37.63	37.97	37.97	40.68	<b>44.36</b>	41.69
		D	33.12	39.49	<b>40.76</b>	36.94	40.13	38.22
	W	C	29.30	31.17	31.43	34.55	34.82	<b>34.91</b>
		A	30.06	32.78	32.46	33.82	<b>35.92</b>	33.82
		D	87.26	89.17	89.17	88.54	<b>91.72</b>	89.91
	D	C	31.70	31.52	31.17	34.73	35.17	<b>35.26</b>
		A	32.15	33.09	33.19	34.66	34.34	<b>34.55</b>
		W	86.10	89.49	89.49	91.19	<b>91.53</b>	<b>91.53</b>
USPS+MNIST	USPS	MNIST	51.05	59.65	59.90	58.80	<b>60.80</b>	60.60
	MNIST	USPS	56.28	67.28	67.39	66.78	<b>67.78</b>	<b>68.89</b>
Average			50.31	53.78	53.76	55.20	<b>56.87</b>	56.50

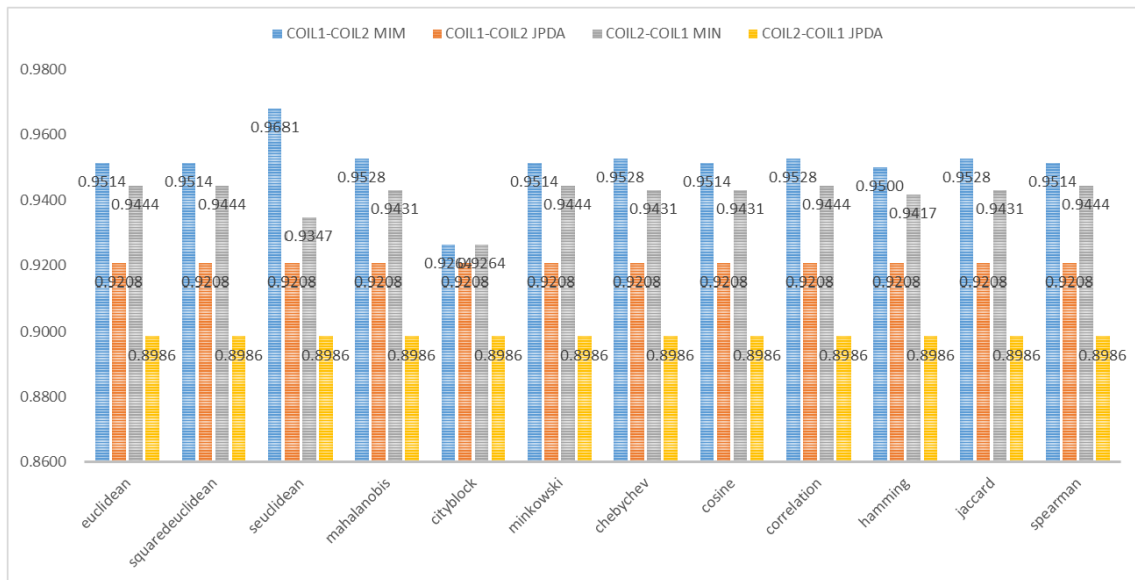


FIGURE 8. Classification accuracies with the different metric matrices.

TABLE 4. Classification accuracy of the algorithms in deep domain adaptation.

Dateset	Source	Target	GR-AUDA	FA-AUDA	AUDA	CYCAD	AUDA-MIM
RADIOML	RadioML04c	RadioML10a	50.10	46.80	10.15	46.39	<b>55.81</b>
	RadioML10a	RadioML04c	49.20	<b>70.20</b>	51.46	21.47	55.93

chebychev, consine, correlation, hamming, jaccard and spearman. It shows that the distance in MIM is important for domain adaptation, but there is no clear evidence to show which distance is the best one. In the view of average accuracy cosine distance and Euclidean distance are more robust.

4) ABLATION STUDY

We also conducted the ablation study to check if the MMD based on i-MIM can indeed improve the ability of the domain adaptation. The JDP-MMD was used as a baseline. When embedded in DA, the classification accuracies of the

approaches based on different metric matrices are shown in Figure 8. Compared with the baseline method, the proposed algorithms can improve classification accuracies on the tasks COIL2-COIL1 and COIL1-COIL2.

V. CONCLUSION

In this paper, we constructed two metric information matrices which could be as basic components of feature mappings for the models of domain adaptation. Combined with MMD-like model, the feature space induced by metric information matrix is more simple and effective to measure the

discrepancy between different domains. The feature mapping can also be as a basic block of a deep domain adaptation neural network such as Adversarial Discriminative Domain Adaptation. Experiments on different classification datasets verified the superiority of proposed methods.

Furthermore, the proposed models are not the state-of-the-art models. For example, DICE [13] and PACET [14] are both more effective than ours for domain adaptation in the experiments. In the future work, we will try to borrow ideas from PACET and DICE to modified our models.

## REFERENCES

- [1] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, pp. 723–773, Mar. 2012.
- [2] J. Li, K. Lu, Z. Huang, L. Zhu, and H. T. Shen, "Heterogeneous domain adaptation through progressive alignment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1381–1391, May 2019, doi: [10.1109/TNNLS.2018.2868854](https://doi.org/10.1109/TNNLS.2018.2868854).
- [3] G. Peyré, M. Cuturi, and J. Solomon, "Gromov-Wasserstein averaging of kernel and distance matrices," in *Proc. 33rd Int. Conf. Mach. Learn.*, Jun. 2016, pp. 2664–2672.
- [4] W. Zhang and D. Wu, "Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8, doi: [10.1109/IJCNN48605.2020.9207365](https://doi.org/10.1109/IJCNN48605.2020.9207365).
- [5] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207, doi: [10.1109/ICCV.2013.274](https://doi.org/10.1109/ICCV.2013.274).
- [6] J. Hoffman, M. Mohri, and N. Zhang, "Algorithms and theory for multiple-source adaptation," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, Montréal, QC, Canada, Dec. 2018, pp. 8256–8266.
- [7] J. Li, E. Chen, Z. Ding, L. Zhu, K. Lu, and H. T. Shen, "Maximum density divergence for domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 3918–3930, Nov. 2021, doi: [10.1109/TPAMI.2020.2991050](https://doi.org/10.1109/TPAMI.2020.2991050).
- [8] X. Zhang, F. Xinnan Yu, S.-F. Chang, and S. Wang, "Deep transfer network: Unsupervised domain adaptation," 2015, *arXiv:1503.00591*.
- [9] A. Iscen, G. Toliás, Y. Avrithis, T. Furon, and O. Chum, "Efficient diffusion on region manifolds: Recovering small objects with compact CNN representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 926–935, doi: [10.1109/CVPR.2017.105](https://doi.org/10.1109/CVPR.2017.105).
- [10] D. X. Zhou, "Universality of deep convolutional neural network," *Appl. Comput. Harmon. Anal.*, vol. 48, no. 2, pp. 787–794, 2020.
- [11] T. O. J. Shea. *Deepsig*. Accessed: 2016. [Online]. Available: <http://www.deepsig.io/>
- [12] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011, doi: [10.1109/TNN.2010.2091281](https://doi.org/10.1109/TNN.2010.2091281).
- [13] J. Liang, R. He, Z. Sun, and T. Tan, "Aggregating randomized clustering-promoting invariant projections for domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 5, pp. 1027–1042, May 2019, doi: [10.1109/TPAMI.2018.2832198](https://doi.org/10.1109/TPAMI.2018.2832198).
- [14] J. Liang, R. He, Z. Sun, and T. Tan, "Exploring uncertainty in pseudo-label guided unsupervised domain adaptation," *Pattern Recognit.*, vol. 96, Dec. 2019, Art. no. 106996.
- [15] J. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, and P. S. Yu, "Visual domain adaptation with manifold embedded distribution alignment," in *Proc. 26th ACM Int. Conf. Multimedia*, Seoul, South Korea, Oct. 2018, pp. 402–410, doi: [10.1145/3240508.3240512](https://doi.org/10.1145/3240508.3240512).
- [16] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 1129–1134, doi: [10.1109/ICDM.2017.150](https://doi.org/10.1109/ICDM.2017.150).
- [17] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971, doi: [10.1109/CVPR.2017.316](https://doi.org/10.1109/CVPR.2017.316).
- [18] Q. Wang, P. Du, X. Li, J. Yang, and G. Wang, "Adversarial unsupervised domain adaptation for cross scenario waveform recognition," *Signal Process.*, vol. 171, Jun. 2020, Art. no. 107526.

[19] R. Lin, W. Ren, X. Sun, Z. Yang, and K. Fu, "A hybrid neural network for fast automatic modulation classification," *IEEE Access*, vol. 8, pp. 130314–130322, 2020, doi: [10.1109/ACCESS.2020.3009471](https://doi.org/10.1109/ACCESS.2020.3009471).

[20] L. van D. Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, Oct. 2014.



**WENJUAN REN** received the B.Sc. degree from Xidian University, Xi'an, China, in 2005, and the Ph.D. degree from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 2011. She is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. Her research interests include signal processing, radar technology, communication engineering, and data mining.



**SHIE ZHOU** received the M.A. degree in english language and literature from Zhengzhou University, Zhengzhou, China, in 2008. She is currently working as an Associate Professor with the Foreign Language Department, PLA Strategic Support Force Information and Engineering University, with a research focus on college English education and translation.



**ZHANPENG YANG** received the Ph.D. degree from the Beijing Institute of Technology, in 2016. From 2016 to 2020, he worked as a Postdoctoral Scholar with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include functional analysis, signal processing, and differential equations.



**QUAN SHI** received the B.Sc. and M.Sc. degrees from Beihang University, Beijing, China, in 2017. He is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include functional analysis, signal processing, and differential equations.



**XIAN SUN** received the B.Sc. degree from Beihang University, Beijing, China, in 2004, and the M.Sc. and Ph.D. degrees from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 2006 and 2009, respectively. He is currently a Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include deep learning, computer vision, and remote sensing image understanding.



**LUYI YANG** received the B.A. and M.A. degrees in english language and literature from Beijing Foreign Studies University, Beijing, China, in 2012 and 2016, respectively. She is currently working as a Lecturer with the PLA Strategic Support Force Information, PLA Strategic Support Force Information and Engineering University, with a research focus on college English education and foreign military studies.

...