Manifold Criterion Guided Transfer Learning via Intermediate Domain Generation

Lei Zhang, *Senior Member, IEEE,* Shanshan Wang, Guang-Bin Huang, *Senior Member, IEEE,* Wangmeng Zuo, *Senior Member, IEEE,* Jian Yang, *Member, IEEE,* David Zhang, *Fellow, IEEE*

***Abstract*—In many practical transfer learning scenarios, the feature distribution is different across the source and target domains (i.e. non-$i.i.d.$). Maximum mean discrepancy (MMD), as a domain discrepancy metric, has achieved promising performance in unsupervised domain adaptation (DA). We argue that MMD-based DA methods ignore the data locality structure, which, to some extent, would cause the negative transfer effect. The locality plays an important role in minimizing the nonlinear local domain discrepancy underlying the marginal distributions. For better exploiting the domain locality, a novel local generative discrepancy metric (LGDM) based intermediate domain generation learning called Manifold Criterion guided Transfer Learning (MCTL) is proposed in this paper. The merits of the proposed MCTL are four-fold: 1) the concept of manifold criterion (MC) is first proposed as a measure validating the distribution matching across domains, and domain adaptation is achieved if the MC is satisfied; 2) the proposed MC can well guide the generation of the intermediate domain sharing similar distribution with the target domain, by minimizing the local domain discrepancy; 3) a global generative discrepancy metric (GGDM) is presented, such that both the global and local discrepancy can be effectively and positively reduced; 4) a simplified version of MCTL called MCTL-S is presented under a perfect domain generation assumption for more generic learning scenario. Experiments on a number of benchmark visual transfer tasks demonstrate the superiority of the proposed manifold criterion guided generative transfer method, by comparing with other state-of-the-art methods. The source code is available in https://github.com/wangshanshanCQU/MCTL.**

***Index Terms*—Transfer Learning, domain adaptation, manifold criterion, discrepancy metric, domain generation.**

I. Introduction

STATISTICAL machine learning models rely heavily on the assumption that the data used for training and test are drawn from the same or similar distribution, i.e. independent identical distribution ($i.i.d.$). However, in real world, it is impossible to guarantee that assumption. Hence, in visual recognition tasks, classifier or model usually does not work well because of data bias between the distributions of the training and test data [1], [2] [3], [4], [5], [6], [7]. The domain discrepancy constitutes a major obstacle in training the predictive models across domains. For example, an object recognition model trained on labeled images may not generalize well on the testing images under various variations in the pose, occlusion, or illumination. In Machine Learning this problem is labeled as domain mismatch. Failing to model such a distribution shift may cause significant performance degradation. Also, the models trained with only a limited number of labeled patterns are usually not robust for pattern recognition tasks. Furthermore, manual labeling of sufficient training data for diverse application domains may be prohibitive. However, by leveraging the labeled data drawn from another sufficiently labeled source domain that describes related contents with target domain, establishing an effective model is possible. Therefore, the challenging objective is how to achieve knowledge transfer across domains such that the distribution mismatch is reduced. Underlying techniques for addressing this challenge, such as domain adaptation [8] [9], which aims to learn domain-invariant models across source and target domain, has been investigated.

Domain adaptation (DA) [10], [11], [12] as one kind of transfer learning (TL) perspective, addresses the problem that data is from two related but different domains [13], [14]. Domain adaptation establishes knowledge transfer from the labeled source domain to the unlabeled target domain by exploring domain-invariant structures that bridge different domains with substantial distribution discrepancy. In terms of the accessibility of target data labels in transfer learning, domain adaptation methods can be divided into three categories: supervised [15], [16], semi-supervised [17], [18], [5] and unsupervised [19], [20], [21].

In this paper, we focus on unsupervised transfer learning where the target data labels are unavailable in transfer model learning phase. Unsupervised setting is more challenging due to the common data scarcity problem. In unsupervised transfer learning [22], Maximum Mean Discrepancy (MMD) [23] is widely used and has achieved promising performance. MMD, that aims at minimizing the domain distribution discrepancy, is generally exploited to reduce the difference of conditional distributions and marginal distributions across domains by utilizing the unlabeled domain data in a Reproducing Kernel Hilbert Space (RKHS). Also, in the framework of deep trans-

This work was supported by the National Science Fund of China under Grants (61771079, 91420201 and 61472187), Chongqing Natural Science Fund (No. cstc2018jcyjAX0250), the 973 Program No.2014CB349303, and Program for Changjiang Scholars. *(Corresponding author: Lei Zhang)*

L. Zhang and S. Wang are with the School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044,China. (E-mail: leizhang@cqu.edu.cn, wangshanshan@cqu.edu.cn).

G.B. Huang is with School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. (E-mail: egbhuang@ntu.edu.sg).

W.M. Zuo is with School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. (E-mail: wmzuo@hit.edu.cn).

J. Yang is with School of Computer Science and Technology, Nanjing University of Science and Technology, China. (E-mail: csjyang@njust.edu.cn).

D. Zhang is with School of Science and Engineering, Chinese University of Hong Kong (Shenzhen), Shenzhen, China. (E-mail: davidzhang@cuhk.edu.cn).

fer learning [24], MMD-based adaptation layers are further integrated in deep neural networks to improve the transferable capability between the source and domains [25].

MMD actually acts as a discrepancy metric or criterion to evaluate the distribution mismatch across domains and works well in aligning the global distribution. However, it only considers the domain discrepancy and generally ignores the intrinsic data structure of target domain, e.g. local structure just as Fig.1(b). It is known that geometric structure is indispensable for domain distance minimization, which, thus, can well exploit the internal local structure of target data. Particularly, in unsupervised learning, the local structure of target data often plays a more important role than the global structure. This is originated from manifold assumption that the data with local similarity is with similar labels. Motivated by manifold assumption, a novel manifold criterion (MC) is proposed in our work, which is similar but very different from conventional manifold algorithms that the MC actually acts as a generative transfer criterion for unsupervised domain adaptation.

Intuitively, we hold the assumption that if a new target domain can be automatically generated by using the source domain data, the domain transfer issue can be naturally addressed. To this end, a criterion that measures the generative effect can be explored. In this paper, considering the locality property of target data, we wish that the generative target data should hold similar local structure with the true target domain data. Naturally, motivated by manifold assumption [26], an objective generative transfer metric, manifold criterion (MC), is proposed. Suppose that two samples x_i and x_j in target domain are close to each other, and if the generative target sample x_i^g by using the source data is also close to x_j, we recognize that the generated intermediate domain data shares similar distribution with the target domain. This is the basic idea of our generative transfer learning in this paper.

But how to construct the generative target domain? From the perspective of manifold learning, we expect that the new target data is generated by using a locality structure preservation metric. This idea can be interpreted under the commonly investigated case of independent identically distribution ($i.i.d.$) that the affinity structure in high-dimensional space can still be preserved in some projected low-dimensional subspace (i.e. manifold structure embedding). In general, the internal intrinsic structure can remain unchanged by using graph Laplacian regularization [27], which reflects the affinity of the raw data.

Specifically, with the proposed manifold criterion, a **M**anifold **C**riterion guided **T**ransfer **L**earning (MCTL) is proposed, which aims to pursue a latent common subspace via a projection matrix $\boldsymbol{\mathcal{P}}$ for source and target domain. In the common subspace, a generative transfer matrix $\boldsymbol{\mathcal{Z}}$ is solved by leveraging the source domain data and the MC generative metric, for a new generative data that holds similar marginal distribution with source data in a unsupervised manner. The findings and analysis show that the proposed manifold criterion can be used to reduce the local domain discrepancy.

Additionally, in MCTL model, the embedding of low-rank constraint (LRC) on the transfer matrix ensures that the data from source domains can be well interpreted during generation, which can show an approximated block-diagonal

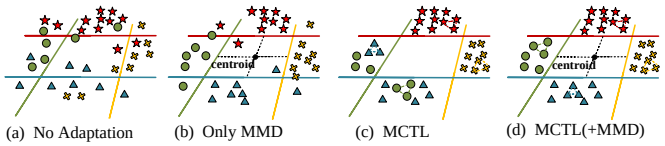

Fig. 1: Motivation of MCTL. The lines represent the classification boundary of source domain. The centroid represents the geometric center of all data points.

property. With the LRC exploited, the local structure based MC can be guaranteed as we wish without distortion [28].

The idea of our MCTL is described in Fig.2. In summary, the main contribution and novelty of this work are fourfold:

- We propose a unsupervised manifold criterion generative transfer learning (MCTL) method, which aims to generate a new intermediate target domain that holds similar distribution with true target data by leveraging source data as basis. The proposed manifold criterion (MC) is modeled by a novel local generative discrepancy metric (LGDM) for local cross-domain discrepancy measure, such that the local transfer can be effectively aligned.
- In order to keep the global distribution consistency, a global generative discrepancy metric (GGDM), that offers a linear method to compare the high-order statistics of two distributions, is proposed to minimize the discrepancy between the generative target data and the true target data. Therefore, the local and global affinity structures across domains are simultaneously guaranteed.
- For improving the correlation between the source data and the generative target data, LRC regularization on the transfer matrix $\boldsymbol{\mathcal{Z}}$ is integrated in MCTL, such that the block-diagonal property can be utilized for preventing the domain transfer from distortion and negative transfer.
- Under the MCTL framework, for a more generic case, a simplified version of MCTL (i.e. MCTL-S) method is proposed, which constrains that the generative data should be seriously consistent with the target domain in a simple yet generic manner. Interestingly, with this constraint, the LGDM loss in MCTL-S is naturally degenerated into a generic manifold regularization.

The remainder of this paper is organized as follows. In Section II, we review the related work in transfer learning. In Section III, we present the preliminary idea of the proposed manifold criterion. In Section IV, the proposed MCTL method and optimization are formulated. In Section V, the simplified version of MCTL is introduced and preliminarily analyzed. In Section VI, the classification method is described. In Section VII, the experiments in cross-domain visual recognition are presented. The discussion is presented in Section VIII. Finally, the paper is concluded in Section IX.

II. Related Work

A. Shallow Transfer Learning

A lot of transfer learning methods are proposed to tackle heterogeneous domain adaptation problems. Generally, these methods can be divided into three categories in the follows.

Classifier based approaches. A generic way is to directly learn a common classifier on auxiliary domain data by leveraging a few labeled target data. Yang et al. [29] proposed an adaptive SVM (A-SVM) to learn a new target classifier $f^T(x)$ by supposing that $f^T(x) = f^S(x) + \Delta f(x)$, where the classifier $f^S(x)$ is trained with the labeled source samples and $\Delta f(x)$ is the perturbation function. Bruzzone et al. [30] developed an approach to iteratively learn the SVM classifier by labeling the unlabeled target samples and simultaneously removing some labeled samples in the source domain. Duan et al. [8] proposed an adaptive multiple kernel learning (AMKL) for consumer video event recognition from annotated web videos. Also, a domain transfer MKL (DTMKL) [5], which learn a SVM classifier and a kernel function simultaneously for classifier adaptation. Zhang et al. [31] proposed a robust classifier transfer method (EDA) which was modelled based on ELM and manifold regularization for visual recognition.

Feature augmentation/transformation based approaches. Li et al. [32] proposed a heterogeneous feature augmentation (HFA)which tends to learn a transformed feature space for domain adaptation. Kulis et al. [9] proposed an asymmetric regularized cross-domain transform (ARC-t) method for learning a transformation metric. In [33], Hoffman et al. proposed a Max-Margin Domain Transforms (MMDT) which a category specific transformation was optimized for domain transfer. Gong et al. proposed a Geodesic Flow Kernel (GFK) [34] method which integrates an infinite number of linear subspaces on the geodesic path to learn the domain-invariant feature representation. Gopalan et al. [35] proposed an unsupervised method (SGF) for low dimensional subspace transfer in which a group of subspaces along the geodesic between source and target data is sampled, and the source data is projected into the subspaces for discriminative classifier learning. An unsupervised feature transformation approach, Transfer Component Analysis (TCA) [11], was proposed to discover common features having the same marginal distribution by using Maximum Mean Discrepancy (MMD) as non-parametric discrepancy metric. MMD [23], [36], [37] is often used in transfer learning. Long et al. [38] proposed a Transfer Sparse Coding (TSC) approach to construct robust sparse representations by using empirical MMD as the distance measure. The Transfer Joint Matching (TJM) proposed by Long et al. [19] tends to learn a non-linear transformation by minimizing the MMD based distribution discrepancy.

Feature representation based approaches. Different from those methods above, domain adaptation is achieved by representing across domain features. Jhuo et al. [39] proposed a RDALR method, in which the source data is reconstructed with target domain by using low-rank modeling. Similarly, Shao et al. [40] proposed a LTSL method by pre-learning a subspace using PCA or LDA, then low-rank representation across domain is modeled. Zhang et al. [41], [42] proposed Latent Sparse Domain Transfer (LSDT) and Discriminative Kernel Transfer Learning (DKTL) methods for visual adaptation, by jointly learning a subspace projection and sparse reconstruction across domain. Further, Xu et al. [43] proposed a DTSL method, which combines the low-rank and sparse constraint on the reconstruction matrix.

In this paper, the proposed method is different from the existing shallow transfer learning methods that a generative transfer idea is motivated, which tends to achieve domain adaptation by generating an intermediate domain that has similar distribution with the true target domain.

B. Deep Transfer Learning

Deep learning, as a data-driven transfer learning method, has witnessed a great achievements in many fields [44], [45], [46], [47]. However, when solving domain data problems by deep learning technology, massive labeled training data are required. For the small-size tasks, deep learning may not work well. Therefore, deep transfer learning methods have been studied.

Donahue et al. [48] proposed a deep transfer method for small-scale object recognition, and the convolutional network (AlexNet) was trained on ImageNet. Similarly, Razavian et al. [49] also proposed to train a network based on ImageNet for high-level feature extractor. Tzeng et al. [44] proposed a DDC method which simultaneously achieves knowledge transfer between domains and tasks by using CNN. Long et al. [25] proposed a deep adaptation network (DAN) method by imposing MMD loss on the high-level features across domains. Additionally, Long et al. [21] also proposed a residual transfer network (RTN) which tends to learn a residual classifier based on softmax loss. Oquab et al. [46] proposed a CNN architecture for middle level feature transfer, which is trained on large annotated image set. Additionally, Hu et al. [24] proposed a non-CNN based deep transfer metric learning (DTML) method to learn a set of hierarchical nonlinear transformations for achieving cross-domain visual recognition.

Recently, GAN inspired adversarial domain adaptation has been preliminarily studied. Tzeng et al. proposed a novel ADDA method [50] for adversarial domain adaptation, in which CNN is used for adversarial discriminative feature learning, and achieves the state-of-the-art performance.

In this work, although the proposed MCTL method is a shallow transfer learning paradigm, the competitive capability comparing to these deep transfer learning methods has been validated on the pre-extracted deep features.

C. Differences Between MCTL and Other Reconstruction Transfer Methodologies

The proposed MCTL is partly related by reconstruction transfer methods, such as DTSL [43], LSDT [41] and LTSL [40], but essentially different from them. These methods aim to learn a common subspace where a feature reconstruction matrix between domains is learned for adaptation. Sparse reconstruction and low-rank based constraints were considered, respectively. Different from reconstruction transfer, the proposed MCTL is a generative transfer learning paradigm, which is partly inspired by the idea of GAN [51] and manifold learning. The differences and relations are as follows.

Reconstruction Transfer. As the name implies, a reconstruction matrix is expected for domain correspondence. In LTSL, subspace projection $\mathbf{W}$ is pre-learned by off-the-shelf methods such as PCA, LDA, etc. Then projected source data $\mathbf{WX_S}$ is used to reconstruct the projected target data $\mathbf{WX_T}$

via low-rank constraint. The subspace may be suboptimal leading to a possible local optimum of $\mathcal{Z}$. Further, the LSDT method was proposed for realizing domain adaptation by exploiting cross-domain sparse reconstruction in some latent subspace, simultaneously. The DTSL was proposed by posing hybrid regularization of sparsity and low-rank constraints for learning a more robust reconstruction transfer matrix. Reconstruction transfer always expresses target domain by leveraging source domain, however, this expression is not accurate due to the limited number of target domain data in calculating the reconstruction error loss, and the robustness is decreased.

Generative Transfer. The proposed MCTL method introduces a generative transfer learning concept, which aims to realize an intermediate domain generation by constructing a Manifold Criterion loss. The motivation is that the domain adaptation problem can be solved by generating a similar domain that shares the same distribution with the true target domain. The essential differences of our work from reconstruction lie in that: (1) Domain adaptation is recognized to be a domain generation problem, instead of a domain alignment problem. (2) The manifold criterion loss is well constructed for generation, instead of the least-square based reconstruction error loss. In addition, the GGDM based global domain discrepancy loss and LRC regularization are also integrated in MCTL for global distribution discrepancy reduction and domain correlation enhancement, simultaneously.

Similarity and Relationship. The reconstruction transfer and generative transfer are similar and related in three aspects. (1) Both aim at pursuing a more similar domain with the target data by leveraging the source domain data. (2) Both are unsupervised transfer learning, which do not need the data label information in domain adaptation. (3) Both have similar model formulation and solvers for obtaining the domain correspondence matrix and transformation.

III. Manifold Criterion Preliminary

Manifold learning [20], [27] as a typical unsupervised learning method has been widely used. Manifold hypothesis means that an intrinsic geometric low-dimensional structure is embedded in high-dimensional feature space and the data with affinity structure own similar labels. This demonstrates that manifold hypothesis works but under the data of independent identically distribution ($i.i.d.$). Therefore, we could have a try to build a manifold criterion to measure the $i.i.d.$ condition (i.e. domain discrepancy minimization) and guide the transfer learning across domains through an intermediate domain.

In this paper, manifold hypothesis is used in the process of generating domain as shown in Fig.2. Essentially different from manifold learning and regularization, we propose a novel manifold criterion (MC) that is utilized as generative discrepancy metric. In semi-supervised learning (SSL), manifold regularization is often used but under $i.i.d.$ condition. However, transfer learning is different from SSL that domain data does not satisfy $i.i.d.$ condition. In this paper, it should be figure out that if the intermediate domain can be generated via the manifold criterion guided objective function, then the distribution of the generated intermediate domain and the true target domain is recognized to be matched.

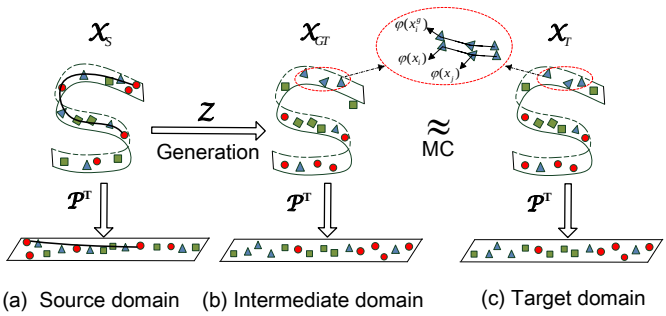

Fig. 2: Illustration of the proposed Manifold Criterion Guided Transfer Learning (MCTL). (a) represents the source domain $\boldsymbol{\mathcal{X}}_S$ which is used to generate an intermediate target domain $\boldsymbol{\mathcal{X}}_{GT}$ shown as (b), that is similar to the true target domain $\boldsymbol{\mathcal{X}}_T$ shown in (c). The intermediate domain generation is carried out by the learned generative matrix $\boldsymbol{\mathcal{Z}}$ based on the manifold criterion (MC) in an unsupervised manner. MC interprets the distribution discrepancy, which implies that if the local discrepancy is minimized, the distribution consistency is then achieved. Further, a projection matrix $\boldsymbol{\mathcal{P}}$ is learned for domain feature embedding. Notably, the $\varphi(.)$ is used as the implicit mapping function of data, which can be kernelized in implementation with inner product.

The idea of manifold criterion is described in Fig.2. We observe that a projection matrix $\boldsymbol{\mathcal{P}}$ is first learned for some common subspace projection, and then a generative transfer matrix $\boldsymbol{\mathcal{Z}}$ is learned for intrinsic structure preservation and distribution discrepancy minimization between the true target data and generative target data by source domain data. That is, if the generative data has similar affinity structure with the true target domain, i.e. manifold criterion is satisfied, we can have a conclusion that the generative data shares similar distribution with target domain. Notably, different from reconstruction based domain adaptation methods, in this work, we tend to generate an intermediate domain by leveraging source domain, i.e. generative transfer instead of reconstruction transfer.

Moreover, we show Fig.1 to imply that MC (local) and MMD (global) can be jointly considered in transfer learning models. Frankly, the idea of this paper is intuitive, simple and easy to follow. The key point lies in that how to generate the intermediate domain data such that the generated data complies with manifold assumption originated from the true target domain data. If the manifold criterion is satisfied (i.e. $i.i.d.$ is achieved), then domain adaptation or distribution alignment is completed, which is the principle of MCTL.

IV. MCTL: Manifold Criterion Guided Transfer Learning

A. Notations

In this paper, source and target domain are defined by subscript S and T. Training set of source and target domain are defined as $\varphi(\mathbf{X}_S) \in \mathbf{R}^{m\times n_S}$ and $\varphi(\mathbf{X}_T) \in \mathbf{R}^{m\times n_T}$. $\varphi(\mathbf{X}_{GT}) \in \mathbf{R}^{m\times n_T}$ denotes generative target domain, where φ denotes an implicit but generic transformation, m denotes dimensionality, n_S and n_T denote the number

of samples in source and target domain, respectively. Let $\mathbf{X} = [\mathbf{X}_S, \mathbf{X}_T]$, then $\varphi(\mathbf{X}) \in \mathbf{R}^{m\times n}$, where $n = n_S + n_T$. Let $\boldsymbol{\mathcal{P}} \in \mathbf{R}^{m\times d}(m \geq d)$ be the basis transformation that maps raw data space from $\mathbf{R}^m$ to a latent subspace $\mathbf{R}^d$. $\boldsymbol{\mathcal{Z}} \in \mathbf{R}^{n_S\times n_T}$ represents generative transfer matrix, $\mathbf{I}$ denotes identity matrix, $\| \bullet \|_F$ and $\| \bullet \|_2$ denote l_F-norm and l_2-norm, respectively. The superscript $\mathbf{T}$ denotes transpose operator and $\mathbf{Tr}(\bullet)$ denotes matrix trace operator.

In RKHS, the kernel Gram matrix $\mathcal{K}$ is defined as $\left[\mathbf{K}\right]_{i,j} =< \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) >_{\mathcal{H}} = \varphi(\mathbf{x}_i)^H\varphi(\mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j)$, where k is a kernel function. In the following sections, let $\mathbf{K}{=}\varphi(\mathbf{X})^{\mathbf{T}}\varphi(\mathbf{X})$, $\mathbf{K}_S = \varphi(\mathbf{X})^{\mathbf{T}}\varphi(\mathbf{X}_S)$ and $\mathbf{K}_T = \varphi(\mathbf{X})^{\mathbf{T}}\varphi(\mathbf{X}_T)$, and it is easy to get that $\mathbf{K} \in \mathbf{R}^{n\times n}$, $\mathbf{K}_S \in \mathbf{R}^{n\times n_S}$ and $\mathbf{K}_T \in \mathbf{R}^{n\times n_T}$.

B. Problem Formulation

In this section, the proposed MCTL method is presented in Fig.2, in which the same distribution between the **G**enerated intermediate **T**arget domain (D_{GT}) and the true **T**arget domain (D_T) under common subspace is what we expected. That is, the intermediate target domain is generated to share the approximated distribution as the true target domain by exploiting the proposed Manifold Criterion as domain discrepancy metric. Specifically, two generative discrepancy metrics (LGDM vs. GGDM) for measuring the domain discrepancy locally and globally are proposed. Overall, the model is composed of three items. The 1^{st} item is MC-based LGDM loss which is used to measure the local domain discrepancy with the manifold criterion by exploiting the locality of target data. The 2^{nd} item is the GGDM loss which is applied to minimize the global domain discrepancy of marginal distributions between the generated intermediate target domain and the true target domain. The 3^{rd} item is the LRC regularization (low-rank constraint) which is carried out to keep the generalization of $\boldsymbol{\mathcal{Z}}$. A detailed MCTL method is described in the follows.

1) MC based Local Generative Discrepancy Metric: The MC based local generative discrepancy metric (LGDM) loss is used to enhance the distribution consistency between source and target domain indirectly, by constraining the generative target data with manifold criterion. For convenience, $\varphi(x^p_{GT})$ is defined as a sample in $\varphi(\mathbf{X}_{GT})$ and $\varphi(x^q_T)$ is defined as a sample in $\varphi(\mathbf{X}_T)$. We claim that the distribution consistency between $\varphi(\mathbf{X}_{GT})$ and $\varphi(\mathbf{X}_T)$ is achieved, i.e. domain transfer is done, only if two sets satisfy the following manifold criterion, which can be formulated as

$$\begin{aligned} LGDM(D_{GT}, D_T) &= \sum_{p,q}^{n_T} W_{pq} \left\|\varphi(x^p_{GT}) - \varphi(x^q_T)\right\|_2^2 \\ &= Tr(\varphi(\mathbf{X}_{GT})\mathbf{D}(\varphi(\mathbf{X}_{GT})^{\mathbf{T}}) \\ &\quad + Tr(\varphi(\mathbf{X}_T)\mathbf{D}(\varphi(\mathbf{X}_T)^{\mathbf{T}}) \\ &\quad - 2Tr(\varphi(\mathbf{X}_{GT})\mathbf{W}(\varphi(\mathbf{X}_T)^{\mathbf{T}}) \end{aligned} \tag{1}$$

where $\mathbf{W} \in R^{n_T\times n_T}$ is the affinity matrix described as $W_{pq} = \begin{cases} 1, if\ x^p_{GT} \in NN_k(x^q_T) or\ x^q_T \in NN_k(x^p_{GT}) \\ 0,\ otherwise \end{cases}$ and $NN_k(\mathbf{x})$ represents the k^{th} nearest neighbors of sample $\mathbf{x}$. The matrix $\mathbf{D} \in R^{n_T\times n_T}$ is a diagonal matrix with entries $D_{pp} = \sum_q W_{pq}$, $p = 1, ..., n_T$. As claimed before, $\boldsymbol{\mathcal{P}}^{\mathbf{T}}{=}\boldsymbol{\Phi}^{\mathbf{T}}\varphi(\mathbf{X})^{\mathbf{T}}$, the projected source data and target data can be expressed as $\boldsymbol{\Phi}^{\mathbf{T}}\varphi(\mathbf{X})^{\mathbf{T}}\varphi(\mathbf{X}_S)$ and $\boldsymbol{\Phi}^{\mathbf{T}}\varphi(\mathbf{X})^{\mathbf{T}}\varphi(\mathbf{X}_T)$. By substituting $\varphi(\mathbf{X}_{GT}) = \varphi(\mathbf{X}_S)\boldsymbol{\mathcal{Z}}$ and the Gram matrix after projection (i.e. $\boldsymbol{\Phi}^{\mathbf{T}}\mathbf{K}_S$ and $\boldsymbol{\Phi}^{\mathbf{T}}\mathbf{K}_T$) into Eq. (1), the MC based LGDM loss can be further formulated as

$$\begin{aligned} &\min_{\boldsymbol{\Phi},\boldsymbol{\mathcal{Z}}} \frac{1}{(n_T)^2} Tr(\boldsymbol{\Phi}^{\mathbf{T}}\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{D}(\boldsymbol{\Phi}^{\mathbf{T}}\mathbf{K}_S\boldsymbol{\mathcal{Z}})^{\mathbf{T}}) \\ &\quad + \frac{1}{(n_T)^2} Tr(\boldsymbol{\Phi}^{\mathbf{T}}\mathbf{K}_T\mathbf{D}(\boldsymbol{\Phi}^{\mathbf{T}}\mathbf{K}_T)^{\mathbf{T}}) \\ &\quad - \frac{2}{(n_T)^2} Tr(\boldsymbol{\Phi}^{\mathbf{T}}\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{W}(\boldsymbol{\Phi}^{\mathbf{T}}\mathbf{K}_T)^{\mathbf{T}}) \end{aligned} \tag{2}$$

From Eq.(2), the motivation is clearly demonstrated which tends to achieve local structure consistency (i.e. manifold consistency) between the generative target data and the true target data. The intrinsic difference between Eq.(2) and the manifold embedding or regularization is that we aim to produce the $i.i.d.$ assumption with a manifold criterion, while the conventional manifold learning relies on this assumption.

2) Global Generative Discrepancy Metric Loss: In order to reduce the distribution mismatch between the generative target data and the true target data, a generic MMD for global generative discrepancy metric (GGDM) is proposed by minimizing the discrepancy as follows.

$$GGDM(D_{GT}, D_T) = \frac{1}{n_T}\sum_{i=1}^{n_T} \left\|(\varphi(\mathbf{X}^i_{GT}){-}\varphi(\mathbf{X}^i_T))\right\|_2^2 \tag{3}$$

where D_{GT} and D_T denote the distribution of generated target domain and true target domain, respectively. However, model may not transfer knowledge directly and it is unclear where a test sample is from (source or target domain) if there is not a common subspace. We consider to find a latent common subspace for source and target domain by using a projection matrix $\boldsymbol{\mathcal{P}}$. Therefore, by projecting $\varphi(\mathbf{X}_{GT})$ and $\varphi(\mathbf{X}_T)$ to the subspace, the GGDM loss after projection can be formulated as follows. Considering that $\varphi(\mathbf{X}_{GT}) = \varphi(\mathbf{X}_S)\boldsymbol{\mathcal{Z}}$, by substituting it in the equation, there is

$$\begin{aligned} GGDM(D_{GT}, D_T) &= \frac{1}{n_T}\sum_{i=1}^{n_T} \left\|\boldsymbol{\mathcal{P}}^{\mathbf{T}}(\varphi(\mathbf{X}^i_{GT}){-}\varphi(\mathbf{X}^i_T))\right\|_2^2 \\ &= \frac{1}{n_T}\left\|\mathbf{P}^{\mathbf{T}}(\varphi(\mathbf{X}_S)\boldsymbol{\mathcal{Z}} - \varphi(\mathbf{X}_T))\mathbf{1}\right\|_2^2 \end{aligned} \tag{4}$$

where $\mathbf{1}$ represents a full one column vector.

The projection matrix $\boldsymbol{\mathcal{P}}$ is a linear transformation, which can be represented as some linear combination of the training data, i.e. $\boldsymbol{\mathcal{P}}^{\mathbf{T}}{=}\boldsymbol{\Phi}^{\mathbf{T}}\varphi(\mathbf{X})^{\mathbf{T}}$, where $\boldsymbol{\Phi}$ denotes the linear combination coefficient matrix. Then the projected source data can be expressed as $\boldsymbol{\Phi}^{\mathbf{T}}\varphi(\mathbf{X})^{\mathbf{T}}\varphi(\mathbf{X}_S)$ and the projected target data can be expressed as $\boldsymbol{\Phi}^{\mathbf{T}}\varphi(\mathbf{X})^{\mathbf{T}}\varphi(\mathbf{X}_T)$. With the kernel trick, the inner product of implicit transformation is represented as Gram matrix, from raw space to RKHS. As described in section 4.1, let $\mathbf{K}_S = \varphi(\mathbf{X})^{\mathbf{T}}\varphi(\mathbf{X}_S)$ and $\mathbf{K}_T = \varphi(\mathbf{X})^{\mathbf{T}}\varphi(\mathbf{X}_T)$, the source domain and target domain

can be expressed simply as $\mathbf{\Phi^T K}_S$ and $\mathbf{\Phi^T K}_T$, respectively. Therefore, the GGDM loss is formulated as

$$\min_{\mathbf{\Phi},\boldsymbol{\mathcal{Z}}} \frac{1}{n_T} \left\| \mathbf{\Phi^T}(\mathbf{K}_S\boldsymbol{\mathcal{Z}} - \mathbf{K}_T)\mathbf{1} \right\|_2^2 \tag{5}$$

3) LRC for Domain Correlation Enhancement: In domain transfer, the loss functions are designed for interpreting the generative target data and the true target data. Significantly, the generative target data plays an critical role in the proposed model. In this work, a general transfer matrix $\boldsymbol{\mathcal{Z}}$ is used to bridge the source domain data and the generative data (intermediate result). It is known that for structural consistency between different domains is our goal, therefore, it is natural to consider the low-rank structure of $\boldsymbol{\mathcal{Z}}$ as a choice for enhancing the domain correlation. In our MCTL, low-rank constraint (LRC), that is effective in showing the global structure of different domain data, is finally used. The LRC regularization ensures that the data from different domains can be well interlaced during domain generation, which is significant to reduce the disparity of domain distributions. Furthermore, if the projected data lies in the same manifold, each sample in target domain can be represented by its neighbors in source domain. This requires that the generative transfer matrix $\boldsymbol{\mathcal{Z}}$ is approximately block-wise. Therefore, LRC regularization is necessary. Considering the non-convexity property of rank function which is NP-hard, the nuclear norm $||\boldsymbol{\mathcal{Z}}||_*$ is used as a rank approximation in this work.

4) Completed Model of MCTL: By reviewing the MC based LGDM loss in Eq.(2), the GGDM loss in Eq.(5), and the LRC regularization, the objective function of our MCTL method is finally formulated as follows.

$$\begin{aligned}
&\min_{\mathbf{\Phi},\boldsymbol{\mathcal{Z}}} \frac{1}{(n_T)^2} Tr(\mathbf{\Phi^T K}_S\boldsymbol{\mathcal{Z}}\mathbf{D}(\mathbf{\Phi^T K}_S\boldsymbol{\mathcal{Z}})^\mathbf{T}) \\
&\quad + \frac{1}{(n_T)^2} Tr(\mathbf{\Phi^T K}_T\mathbf{D}(\mathbf{\Phi^T K}_T)^\mathbf{T}) \\
&\quad - \frac{2}{(n_T)^2} Tr(\mathbf{\Phi^T K}_S\boldsymbol{\mathcal{Z}}\mathbf{W}(\mathbf{\Phi^T K}_T)^\mathbf{T}) \\
&\quad + \tau \frac{1}{n_T} \left\| \mathbf{\Phi^T}(\mathbf{K}_S\boldsymbol{\mathcal{Z}} - \mathbf{K}_T)\mathbf{1} \right\|_2^2 \\
&\quad + \lambda_1 ||\boldsymbol{\mathcal{Z}}||_* \\
&s.t. \mathbf{\Phi^T K \Phi} = \mathbf{I}
\end{aligned} \tag{6}$$

where τ and λ_1 are the trade-off parameters. The rows of $\boldsymbol{\mathcal{P}}$ are required to be orthogonal and normalized to unit norm for preventing trivial solutions by enforcing $\boldsymbol{\mathcal{P}}^\mathbf{T}\boldsymbol{\mathcal{P}} = \mathbf{I}$, which can be further rewritten as $\mathbf{\Phi^T K \Phi} = \mathbf{I}$, an equality constraint. Obviously, the model is non-convex with respect to two variables, but can be solved with the variable alternating strategy, and the optimization algorithm is formulated.

C. Optimization

There are two variables $\mathbf{\Phi}$ and $\boldsymbol{\mathcal{Z}}$ in the MCTL model (6), therefore an efficient variable alternating optimization strategy is naturally considered, i.e. one variable is solved while frozen the other one. First, when $\boldsymbol{\mathcal{Z}}$ is fixed, a general Eigen-value decomposition is used for solving $\mathbf{\Phi}$. Second, when $\mathbf{\Phi}$ is fixed, the inexact augmented Lagrangian multiplier (IALM) and gradient descent are used to solve $\boldsymbol{\mathcal{Z}}$. In the following, the optimization details of the proposed method are presented.

By introducing an auxiliary variable $\boldsymbol{\mathcal{J}}$, the problem (6) can be written as follows. Furthermore, with the augmented Lagrange function [52], the model can be written as

$$\begin{aligned}
&\min_{\mathbf{\Phi},\boldsymbol{\mathcal{Z}},\boldsymbol{\mathcal{J}}} \frac{1}{(n_T)^2} (Tr(\mathbf{\Phi^T K}_S\boldsymbol{\mathcal{Z}}\mathbf{D}(\mathbf{\Phi^T K}_S\boldsymbol{\mathcal{Z}})^\mathbf{T}) \\
&\quad + Tr(\mathbf{\Phi^T K}_T\mathbf{D}(\mathbf{\Phi^T K}_T)^\mathbf{T}) - 2Tr(\mathbf{\Phi^T K}_S\boldsymbol{\mathcal{Z}}\mathbf{W}(\mathbf{\Phi^T K}_T)^\mathbf{T})) \\
&\quad + \frac{\tau}{(n_T)^2} \mathbf{\Phi^T}(\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1}(\mathbf{K}_S\boldsymbol{\mathcal{Z}})^\mathbf{T} - \mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1}(\mathbf{K}_T)^\mathbf{T} \\
&\quad - \mathbf{K}_T\mathbf{1}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T} + \mathbf{K}_T\mathbf{1}(\mathbf{K}_T)^\mathbf{T})\mathbf{\Phi} + \lambda_1 ||\boldsymbol{\mathcal{J}}||_* \\
&\quad + Tr(\boldsymbol{\mathcal{R}}_1^\mathbf{T}(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{J}})) + \frac{\mu}{2}(||\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{J}}||_F^2)
\end{aligned} \tag{7}$$

where $\mathbf{1}$ represents a full one matrix instead of a full one vector as the problem (6) is unfolded. $\boldsymbol{\mathcal{R}}_1$ denotes the Lag-multiplier and μ is a penalty parameter.

In the following, we present how to optimize the three variables $\mathbf{\Phi}$, $\boldsymbol{\mathcal{J}}$, and $\boldsymbol{\mathcal{Z}}$ in the problem (7) based on Eigen-value decomposition, IALM and gradient descent in step-wise.

1) Update $\mathbf{\Phi}$: By frozen $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\mathcal{J}}$, $\mathbf{\Phi}$ can be solved as

$$\begin{aligned}
&\mathbf{\Phi}^* = \arg\min_{\mathbf{\Phi}} \frac{1}{(n_T)^2} (Tr(\mathbf{\Phi^T K}_S\boldsymbol{\mathcal{Z}}\mathbf{D}(\mathbf{\Phi^T K}_S\boldsymbol{\mathcal{Z}})^\mathbf{T}) \\
&\quad + Tr(\mathbf{\Phi^T K}_T\mathbf{D}(\mathbf{\Phi^T K}_T)^\mathbf{T}) - 2Tr(\mathbf{\Phi^T K}_S\boldsymbol{\mathcal{Z}}\mathbf{W}(\mathbf{\Phi^T K}_T)^\mathbf{T})) \\
&\quad + \frac{\tau}{(n_T)^2} \mathbf{\Phi^T}(\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T} - \mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1}(\mathbf{K}_T)^\mathbf{T} \\
&\quad - \mathbf{K}_T\mathbf{1}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T} + \mathbf{K}_T\mathbf{1}(\mathbf{K}_T)^\mathbf{T})\mathbf{\Phi} \\
&s.t. \mathbf{\Phi^T K \Phi} = \mathbf{I}
\end{aligned} \tag{8}$$

We can derive the solution $\mathbf{\Phi}_K$ of the K^{th} iteration in column-wise. To obtain the i^{th} column vector in $\mathbf{\Phi}_K$, by setting the partial derivative of problem (8) with respect to $\mathbf{\Phi}_{K(:,i)}$ to be zero, there is

$$\begin{aligned}
&\frac{1}{(n_T)^2} (\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{D}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T} + \mathbf{K}_T\mathbf{D}(\mathbf{K}_T)^\mathbf{T} - \mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{W}(\mathbf{K}_T)^\mathbf{T} \\
&- \mathbf{K}_T\mathbf{W}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T})\mathbf{\Phi}_{K(:,i)} + \frac{\tau}{(n_T)^2} (\mathbf{K}_S\boldsymbol{\mathcal{Z}}1\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T} \\
&- \mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1}(\mathbf{K}_T)^\mathbf{T} - \mathbf{K}_T\mathbf{1}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T} + \mathbf{K}_T\mathbf{1}(\mathbf{K}_T)^\mathbf{T})\mathbf{\Phi}_{K(:,i)} \\
&= -\lambda\mathbf{K}\mathbf{\Phi}_{K(:,i)}
\end{aligned} \tag{9}$$

It is clear that $\mathbf{\Phi}_K$ can be obtained by solving an Eigen-decomposition problem, and $\mathbf{\Phi}_{K(:,i)}$ is the i^{th} eigenvector corresponding to the i^{th} smallest eigenvalue.

2) Update $\mathcal{J}$: By frozen $\mathbf{\Phi}$ and $\boldsymbol{\mathcal{Z}}$, the problem is solved with respect to $\boldsymbol{\mathcal{J}}$. After dropping out the irrelevant terms with respect to $\boldsymbol{\mathcal{J}}$, $\boldsymbol{\mathcal{J}}_{\boldsymbol{K+1}}$ in iteration $K+1$ can be solved as

$$\begin{aligned}
\boldsymbol{\mathcal{J}}_{K+1} = &\min_{\boldsymbol{\mathcal{J}}_K} \lambda_1 \parallel \boldsymbol{\mathcal{J}}_K \parallel_* + \mathbf{Tr}(\boldsymbol{\mathcal{R}}_{1K}^T(\boldsymbol{\mathcal{Z}}_K - \boldsymbol{\mathcal{J}}_K)) \\
&+ \frac{\mu_K}{2} \parallel \boldsymbol{\mathcal{Z}}_K - \boldsymbol{\mathcal{J}}_K \parallel_F^2
\end{aligned} \tag{10}$$

It can be further rewritten as

$$\boldsymbol{\mathcal{J}}_{K+1} = \min_{\boldsymbol{\mathcal{J}}_K} \lambda_1 \parallel \boldsymbol{\mathcal{J}}_K \parallel_* + \frac{\mu_K}{2} \parallel \boldsymbol{\mathcal{J}}_K - (\boldsymbol{\mathcal{Z}}_K + \frac{\boldsymbol{\mathcal{R}}_{1K}}{\mu_K}) \parallel_F^2 \tag{11}$$

Problem (11) can be efficiently solved using the singular value thresholding (SVT) operator [53], which contains two major steps. First, singular value decomposition (SVD) is conducted on matrix $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{Z}}_K + \frac{\boldsymbol{\mathcal{R}}_{1K}}{\mu_K}$, and get $\boldsymbol{\mathcal{S}} = \mathbf{U_S}\sum_\mathbf{S}\mathbf{V_S}$,

Algorithm 1 The Proposed MCTL

Input: $\boldsymbol{\mathcal{X}}_S \in \mathcal{R}^{m\times n_S}$, $\boldsymbol{\mathcal{X}}_T \in \mathcal{R}^{m\times n_T}$, τ, λ_1
Procedure:
1. Compute $\boldsymbol{\mathcal{K}}_T = \varphi(\boldsymbol{\mathcal{X}})^T\varphi(\boldsymbol{\mathcal{X}}_T)$, $\boldsymbol{\mathcal{K}}_S = \varphi(\boldsymbol{\mathcal{X}})^T\varphi(\boldsymbol{\mathcal{X}}_S)$, $\boldsymbol{\mathcal{K}} = \varphi(\boldsymbol{\mathcal{X}})^T\varphi(\boldsymbol{\mathcal{X}})$, $\boldsymbol{\mathcal{X}} = [\boldsymbol{\mathcal{X}}_S, \boldsymbol{\mathcal{X}}_T]$
2.Initialize: $\boldsymbol{\mathcal{J}}$=$\boldsymbol{\mathcal{Z}}$=**0**
3. **While** not converge **do**
  3.1 **Step1**: Fix $\boldsymbol{\mathcal{J}}$ and $\boldsymbol{\mathcal{Z}}$, and update $\boldsymbol{\Phi}$ by solving eigenvalue decomposition problem (9).
  3.2 **Step2**: Fix $\boldsymbol{\Phi}$, and update $\boldsymbol{\mathcal{Z}}$ using IALM:
    3.2.1. Fix $\boldsymbol{\mathcal{Z}}$ and update $\boldsymbol{\mathcal{J}}$ by using the singular value thresholding (SVT) [53] operator on problem (11).
    3.2.2. Fix $\boldsymbol{\mathcal{J}}$ and update $\boldsymbol{\mathcal{Z}}$ according to gradient descent operator, i.e. Equation (13).
  3.3 Update the multiplier $\boldsymbol{\mathcal{R}}_1$:
    $\boldsymbol{\mathcal{R}}_1 = \boldsymbol{\mathcal{R}}_1 + \mu(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{J}})$
  3.4 Update the parameter μ:
    $\mu = min(\mu \times 1.01, max_\mu)$
  3.5 Check convergence
end while
Output: $\boldsymbol{\Phi}$ and $\boldsymbol{\mathcal{Z}}$.

where $\sum_{\mathbf{S}} = diag(\{\sigma_i\}_{1\le i\le r})$, σ_i is the singular value with rank r. Second, the optimal solution $\boldsymbol{\mathcal{J}}_{K+1}$ is then obtained by thresholding the singular values as $\boldsymbol{\mathcal{J}}_{K+1} = \mathbf{U_S}\boldsymbol{\Omega}_{(\mathbf{1}/\mu_\mathbf{k})}(\sum_{\mathbf{S}})\mathbf{V_S}$, where $\boldsymbol{\Omega}_{(\mathbf{1}/\mu_\mathbf{k})}(\sum_{\mathbf{S}}) = diag(\{\sigma_i - (1/\mu_k)\}_+)$, and $\{\bullet\}_+$ denotes the positive value operator.

3) Update $\mathcal{Z}$: By frozen $\boldsymbol{\Phi}$ and $\boldsymbol{\mathcal{J}}$, the problem is solved with respect to $\boldsymbol{\mathcal{Z}}$. By dropping out those terms independent of $\boldsymbol{\mathcal{Z}}$ in (7), there is

$$\begin{aligned}\min_{\boldsymbol{\mathcal{Z}}} &\frac{1}{(n_T)^2}(Tr(\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{D}(\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_S\boldsymbol{\mathcal{Z}})^\mathbf{T}) \\ &- 2Tr(\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{W}(\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_T)^\mathbf{T})) + Tr(\mathbf{R}_1^\mathbf{T}(\boldsymbol{\mathcal{Z}} - \mathbf{J})) \\ &+ \frac{\mu}{2}(||\boldsymbol{\mathcal{Z}} - \mathbf{J}||_F^2) + \frac{\tau}{(n_T)^2}\boldsymbol{\Phi}^\mathbf{T}(\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T} \\ &- \mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1}(\mathbf{K}_T)^\mathbf{T} - \mathbf{K}_T\mathbf{1}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T})\boldsymbol{\Phi}\end{aligned} \tag{12}$$

We can see from problem (12) that it is hard to obtain a closed-form solution of $\boldsymbol{\mathcal{Z}}$. Therefore, the general gradient descent operator [54] is used, and the solution of $\boldsymbol{\mathcal{Z}}_{K+1}$ in the $(K+1)^{th}$ iteration is presented as

$$\boldsymbol{\mathcal{Z}}_{K+1} = \boldsymbol{\mathcal{Z}}_K - \alpha \bullet \bigtriangledown(\boldsymbol{\mathcal{Z}}) \tag{13}$$

where $\bigtriangledown(\boldsymbol{\mathcal{Z}})$ denotes the gradient, which is calculated as

$$\begin{aligned}\nabla(\boldsymbol{\mathcal{Z}}) = &\frac{2}{(n_T)^2}((\mathbf{K}_S)^\mathbf{T}\boldsymbol{\Phi}\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{D} - (\mathbf{K}_S)^\mathbf{T}\boldsymbol{\Phi}\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_T\mathbf{W}) \\ &+ \mathbf{R}_1 + \mu(\boldsymbol{\mathcal{Z}} - \mathbf{J}) + \frac{2\tau}{(n_T)^2}(\mathbf{K}_S)^\mathbf{T}\boldsymbol{\Phi}\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1} \\ &- \frac{2\tau}{(n_T)^2}(\mathbf{K}_S)^\mathbf{T}\boldsymbol{\Phi}\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_T\mathbf{1}\end{aligned} \tag{14}$$

In detail, the iterative optimization procedure of the proposed MCTL is summarized in **Algorithm 1**.

V. MCTL-S: Simplified Version of MCTL

As illustrated in MCTL, which aims to minimize the distribution discrepancy between the generative target data and the true target data as close as possible, by using the manifold criterion. In this section, considering the generic manifold embedding, for model simplicity, we rewrite a simplified version of MCTL (MCTL-S in short) as illustrated in Fig.3.

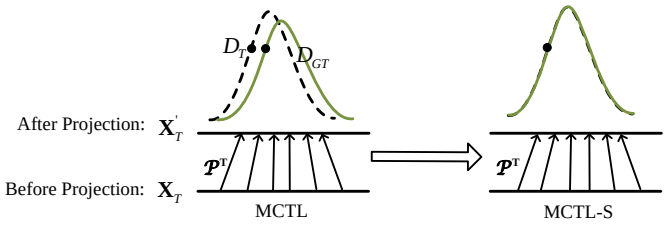

Fig. 3: Difference between MCTL (left) and MCTL-S (right). In MCTL, there is error between the true target domain D_T and the generative target domain D_{GT}. In MCTL-S, the D_{GT} is supposed to be coincided with the true target domain D_T.

A. Formulation of MCTL-S

With the description of Fig.3 (right), suppose an extreme case of $perfect$ domain generation, that is, the generated target data is strictly the same as the true target data, i.e. $\mathbf{X}_{GT} = \mathbf{X}_T$ (D_{GT} coincides with D_T), then MCTL-S is formulated as,

$$\begin{aligned}\min_{\boldsymbol{\Phi},\boldsymbol{\mathcal{Z}}} &\frac{2}{(n_T)^2}Tr(\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{L}(\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_S\boldsymbol{\mathcal{Z}})^\mathbf{T}) \\ &+ \tau\left\|\frac{1}{n_T}\boldsymbol{\Phi}^\mathbf{T}(\mathbf{K}_S\boldsymbol{\mathcal{Z}} - \mathbf{K}_T)\mathbf{1}\right\|_2^2 \\ &+ \lambda_1||\boldsymbol{\mathcal{Z}}||_*\end{aligned} \tag{15}$$

where $\mathbf{L}$=$\mathbf{D} - \mathbf{W}$ is the conventional Laplacian matrix. Also, the objective function (15) contains three items such as the MC based LGDM loss, the GGDM loss and LRC regularization. From the MC-S loss term in Equation (15), we observe a generic manifold regularization term with Laplacian matrix. Therefore, the MC loss can be degenerated into a conventional manifold constraint by implying $\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_T$=$\boldsymbol{\Phi}^\mathbf{T}\mathbf{K}_S\boldsymbol{\mathcal{Z}}$, which shows that MCTL-S model is harsher than MCTL model.

The following experimental results in Table VIII and IX also prove that both the harsh MCTL-S model and the MCTL can achieve good performance. This demonstrates that manifold criterion based intermediate domain generation is a very effective scheme for transfer learning.

B. Optimization of MCTL-S

MCTL-S has a similar mechanism with MCTL, therefore, the MCTL-S optimization is almost the same as MCTL. With two updating steps for $\boldsymbol{\Phi}$ and $\boldsymbol{\mathcal{Z}}$, the optimization procedure of the MCTL-S method is illustrated as follows.

• Update $\boldsymbol{\Phi}$. In the MCTL-S model, by frozen $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\mathcal{J}}$, the derivative of the objective function (15) w.r.t. $\boldsymbol{\Phi}_{K(:,i)}$ is set as zero, there is

$$\begin{aligned}&\frac{2}{(n_T)^2}(\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{L}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T})\boldsymbol{\Phi}_{K(:,i)} + \frac{\tau}{(n_T)^2}(\mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T} \\ &- \mathbf{K}_S\boldsymbol{\mathcal{Z}}\mathbf{1}(\mathbf{K}_T)^\mathbf{T} - \mathbf{K}_T\mathbf{1}\boldsymbol{\mathcal{Z}}^\mathbf{T}(\mathbf{K}_S)^\mathbf{T} + \mathbf{K}_T\mathbf{1}(\mathbf{K}_T)^\mathbf{T})\boldsymbol{\Phi}_{K(:,i)} \\ &= -\lambda\mathbf{K}\boldsymbol{\Phi}_{K(:,i)}\end{aligned} \tag{16}$$

Therefore, $\boldsymbol{\Phi}_K$ in iteration K can be obtained by solving an Eigenvalue decomposition problem, and $\boldsymbol{\Phi}_{K(:,i)}$ is the i^{th} eigenvector corresponding to the i^{th} smallest eigenvalue.

Fig. 4: Some images from 4DA datasets

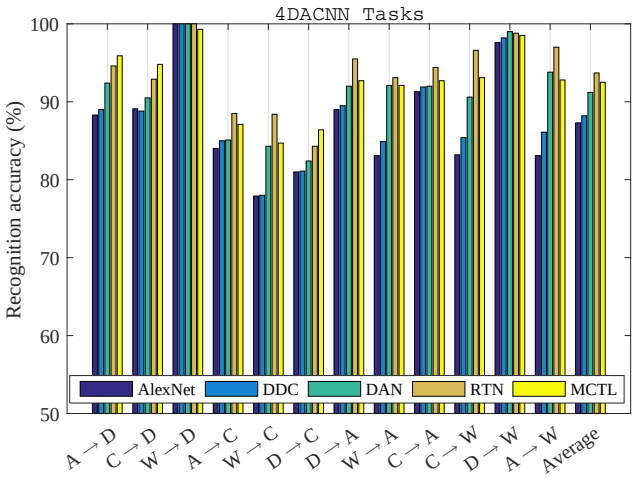

Fig. 5: Comparison with deep transfer learning methods

• Update $\mathcal{J}$. The variable $\mathcal{J}$ can be effectively solved by the singular value thresholding (SVT) operator [53], which is similar to the problem (11).

• Update $\mathcal{Z}$. The variable $\mathcal{Z}$ can be updated according to section 4.3.3 by using gradient descent algorithm. The gradient with respect to $\mathcal{Z}$ can be expressed as

$$\nabla(\mathcal{Z}) = \frac{4}{(n_T)^2}(\mathbf{K}_S)^{\mathbf{T}}\mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}}\mathbf{K}_S\mathcal{Z}\mathbf{L} + \mathbf{R}_1 + \mu(\mathcal{Z} - \mathbf{J}) + \frac{2\tau}{(n_T)^2}((\mathbf{K}_S)^{\mathbf{T}}\mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}}\mathbf{K}_S\mathcal{Z}1 - (\mathbf{K}_S)^{\mathbf{T}}\mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}}\mathbf{K}_T1) \tag{17}$$

VI. Classification

For classification, the projected source data and target data can be represented as $\mathcal{X}_S{}' = \mathbf{\Phi}^{\mathbf{T}}\varphi(\mathcal{X})^{\mathbf{T}}\varphi(\mathcal{X}_S)$, $\mathcal{X}_T{}' = \mathbf{\Phi}^{\mathbf{T}}\varphi(\mathcal{X})^{\mathbf{T}}\varphi(\mathcal{X}_S)\mathcal{Z}$. Then, existing classifiers (e.g., SVM, least square method [55], SRC [56]) can be trained on the domain aligned and augmented training data $[\mathcal{X}_S{}', \mathcal{X}_T{}']$ with label $\mathcal{Y} = [\mathcal{Y}_S, \mathcal{Y}_T]$ by following the experimental setting as LSDT [41]. Notably, for the COIL-20, MSRC and VOC2007 experiments, in order to follow the same experimental setting with DTSL [43], the classifier is trained only on $\mathcal{X}_S{}'$ with label $\mathcal{Y}_S$. Finally, classification on those unlabeled target test data, i.e. $\mathcal{X}_{Tu}{}' = \mathbf{\Phi}^{\mathbf{T}}\varphi(\mathcal{X})^{\mathbf{T}}\varphi(\mathcal{X}_{Tu})$, is achieved, and the recognition accuracy is reported and compared.

VII. Experiments

In this section, the experiments on several benchmark datasets [57] have been exploited for evaluating the proposed MCTL method, including (1) cross-domain object recognition [58], [59]: 4DA office data, 4DA-CNN office data, COIL-20 data, and MSRC-VOC 2007 datasets [38]; (2) cross-pose face recognition: Multi-PIE face dataset; (3) cross-domain handwritten digit recognition: USPS, SEMEION and MNIST datasets. Several related transfer learning methods based on feature transformation and reconstruction, such as SGF [35], GFK [34], SA [60], LTSL [40], DTSL [43], and LSDT [41] have been compared and discussed.

Fig. 6: Some examples from COIL-20 dataset

A. Cross-domain Object Recognition

For cross-domain object/image recognition, 5 benchmark datasets are used, where several sample images in 4DA office dataset are shown in Fig. 4, several sample images in COIL-20 object dataset are shown in Fig. 6, several sample images in MSRC and VOC 2007 datasets are described in Fig. 7.

Results on 4DA Office dataset (Amazon, DSLR, Webcam[1] and Caltech 256[2]) [34]:

Four domains such as Amazon (A), DSLR (D), Webcam (W), and Caltech (C) are included in 4DA dataset, which contains 10 object classes. In our experiment, the configuration is followed in [34] where 20 samples per class are selected from Amazon, 8 samples per class from DSLR, Webcam and Caltech when they are used as source domains; 3 samples per class are chosen when they are used as target training data, while the rest data in target domains are used for testing. Note that the 800-bin SURF features [34], [61] are extracted.

The recognition accuracies are reported in Table I, from which we observe that the propose MCTL ranks the second (54%) in average but slightly inferior to LTSL-LDA (54.9%). The reason may be that the discrimination of LDA helps

[1] http://www.eecs.berkeley.edu/~mfritz/domainadaptation/
[2] http://www.vision.caltech.edu/Image_Datasets/Caltech256/

TABLE I: Recognition accuracy (%) of different domain adaptation in 4DA Setting

4DA Tasks	Naive Comb	HFA [15]	ARC-t [9]	MMDT [33]	SGF [35]	GFK [34]	SA [60]	LTSL -PCA [40]	LTSL -LDA [40]	LSDT [41]	**MCTL**
$A \rightarrow D$	55.9	52.7	50.2	56.7	46.9	50.9	55.1	50.4	**59.1**	52.9	56.1
$C \rightarrow D$	55.8	51.9	50.6	56.5	50.2	55.0	56.6	49.5	**59.6**	56.0	57.3
$W \rightarrow D$	55.1	51.7	71.3	67.0	78.6	75.0	82.3	**82.6**	**82.6**	75.7	73.4
$A \rightarrow C$	32.0	31.1	37.0	36.4	37.5	39.6	38.4	41.5	39.8	42.2	**43.0**
$W \rightarrow C$	30.4	29.4	31.9	32.2	32.9	32.8	34.1	36.7	**38.5**	36.9	37.5
$D \rightarrow C$	31.7	31.0	33.5	34.1	32.9	33.9	35.8	36.2	36.7	37.6	**37.8**
$D \rightarrow A$	45.7	45.8	42.5	46.9	44.9	46.2	45.8	45.7	**47.4**	46.6	47.0
$W \rightarrow A$	45.6	45.9	43.4	47.7	43.0	46.2	44.8	41.9	47.8	46.6	**48.8**
$C \rightarrow A$	45.3	45.5	44.1	49.4	42.0	46.1	45.3	49.3	**50.4**	47.7	42.8
$C \rightarrow W$	60.3	60.5	55.9	**63.8**	54.2	57.0	60.7	50.4	59.5	57.6	59.6
$D \rightarrow W$	62.1	62.1	78.3	74.1	78.6	80.2	**84.8**	81.0	78.3	83.1	82.1
$A \rightarrow W$	62.4	61.8	55.7	**64.6**	54.2	56.9	60.3	52.3	59.5	57.2	55.7
Average	48.5	47.4	49.5	52.5	49.7	51.6	53.7	51.5	**54.9**	53.3	54.0

TABLE II: Recognition accuracy (%) of different domain adaptation of the 7^{th} layer in 4DACNN Setting

4DA-CNN Tasks(f7)	SourceOnly	Naive Comb	SGF [35]	TCA	GFK [34]	LTSL [40]	LSDT [41]	**MCTL**
$A \rightarrow D$	81.3	94.1	92.0	82.8	94.3	94.5	**96.0**	95.9
$C \rightarrow D$	77.6	92.8	92.4	87.9	91.9	93.5	94.6	**94.8**
$W \rightarrow D$	96.2	98.9	97.6	**99.4**	98.5	98.8	99.3	99.3
$A \rightarrow C$	79.3	83.4	77.4	81.2	79.1	85.4	87.0	**87.1**
$W \rightarrow C$	68.1	81.2	76.8	75.5	76.1	82.6	84.2	**84.7**
$D \rightarrow C$	74.3	82.7	78.2	79.6	77.5	84.8	86.2	**86.4**
$D \rightarrow A$	81.8	90.9	88.0	90.4	90.1	91.9	92.5	**92.7**
$W \rightarrow A$	73.4	90.6	86.8	85.6	85.6	91.0	91.7	**92.1**
$C \rightarrow A$	86.5	90.3	89.3	92.1	88.4	90.9	92.5	**92.7**
$C \rightarrow W$	67.8	90.6	87.8	88.1	86.4	90.8	**93.5**	93.1
$D \rightarrow W$	95.1	98.0	95.7	96.9	96.5	97.8	98.3	**98.5**
$A \rightarrow W$	71.6	91.1	88.1	84.4	88.6	91.5	**92.9**	92.8
Average	79.4	90.4	87.5	87.0	87.8	91.1	92.4	**92.5**

TABLE III: Recognition accuracy (%) of different domain adaptation methods on COIL-20

Tasks	SVM	TSL	RDALR [62]	DTSL [43]	LTSL [40]	LSDT [41]	**MCTL**
$C1 \rightarrow C2$	82.7	80.0	80.7	84.6	75.4	81.7	**84.8**
$C2 \rightarrow C1$	84.0	75.6	78.8	**84.2**	72.2	81.5	83.7
Average	83.3	77.8	79.7	**84.4**	73.8	81.6	84.3

MSRC VOC 2007

Fig. 7: Some examples from MSRC and VOC 2007 datasets

improve the performance, because LTSL-PCA only achieves 51.5%, and our MCTL also outperforms other methods. Notably, the 4DA task is a challenging benchmark, which attracts many competitive approaches for evaluation and comparison. Therefore, excellent baselines have been achieved.

Results on 4DA-CNN dataset (Amazon, DSLR, Webcam and Caltech 256) [63], [61]:

In 4DA-CNN dataset, the CNN features are extracted by feeding the raw 4DA data (10 object classes) into the well trained convolutional neural network (AlexNet with 5 convolutional layers and 3 fully connected layers) on ImageNet [63]. The features from the 6^{th} and 7^{th} layers (i.e. DeCAF [48]) are explored. The feature dimensionality is 4096. In experiments, a standard configuration and protocol is used by following [34]. In this paper, the features of the 7^{th} layer are experimented. The recognition accuracies by using the 7^{th} layer outputs for 12 cross-domain tasks are shown in Table II, from which we can observe that the average recognition accuracy of the proposed method shows the best performance. The superiority of generative transfer learning is demonstrated. We can see that our MCTL outperforms LTSL-LDA, this may be because there has been a better discrimination of CNN features, and discriminative learning may not significantly work.

The compared methods in Table II are shallow transfer learning. It is interesting to compare with deep transfer learning methods, such as AlexNet [63], DDC [44], DAN [25] and RTN [21]. The comparison is described in Fig.5, from which we can observe that our proposed method ranks the second in average performance (92.5%), which is inferior to the residual transfer network (RTN), but still better than other three deep transfer learning models. The comparison shows that the proposed MCTL, as a shallow transfer learning

TABLE IV: Recognition accuracy (%) of different domain adaptation methods on MSRC and VOC 2007 Datasets

Tasks	SVM	TSL	RDALR [62]	DTSL [43]	LTSL [40]	LSDT [41]	**MCTL**
$M \rightarrow V$	37.1	32.4	37.5	38.0	38.0	**47.4**	**47.4**
$V \rightarrow M$	55.5	43.2	62.3	56.4	**67.1**	63.9	64.8
Average	46.3	37.8	49.9	47.2	52.6	55.6	**56.1**

Fig. 8: Facial images of one person from CMU Multi-PIE

Fig. 9: Some images from handwritten digits datasets

method, has a good competitiveness.

Results on COIL-20 dataset[3]: Columbia Object Image Library [64]:

The COIL-20 dataset contains 20 objects with 1440 gray scale images (72 multi-pose images per object). The image size is 128×128 of 256 gray levels. In experiments, by following the experimental protocol in [43], the size of each image is cropped into 32×32 and the dataset is divided into two subsets C1 and C2, with each 2 quadrants are included. Specifically, the C1 set contains the directions of $[0°, 85°]$ and $[180°, 265°]$, from quadrants 1 and 3. The C2 set contains the directions of $[90°, 175°]$ and $[270°, 355°]$, from quadrants 2 and 4. The two subsets are distribution different but relevant in semantic, and therefore come to a DA problem. By taking C1 and C2 as source and target domain alternatively, the cross-domain recognition rates of different methods are shown in Table III, from which we see that the proposed MCTL (84.3%) is a little inferior to DTSL (84.4%), but shows a superior performance over other related methods, especially the recent LSDT method (81.6%).

Results on MSRC[4] and VOC 2007[5] datasets: [43]:

The MSRC dataset contains 4323 images with 18 classes and the VOC 2007 dataset contains 5011 images with 20 concepts. The two datasets share 6 semantic classes: airplane, bicycle, bird, car, cow and sheep. We follow [19] to construct a cross-domain image dataset MSRC vs. VOC ($M \rightarrow V$) by selecting 1269 images from MSRC as the source domain, and 1530 images from VOC 2007 as the target domain. Then we switch the two datasets: VOC vs. MSRC ($V \rightarrow M$). All images are uniformly rescaled to 256 pixels, and 128-dimensional dense SIFT (DSIFT) features using the VLFeat open source package are extracted. Then K-means clustering is used to obtain a 240-dimensional codebook. In this way, the source and target domain data are constructed to share the same label set. The experimental results of different domain adaptation methods are shown in Table IV, from which we observe that the performance of our method is 0.5% higher than state-of-the-art LSDT method and 3.5% higher than LTSL method in average cross-domain recognition performance.

[3]http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php

[4]http://research.microsoft.com/en-us/projects/objectclassrecognition

[5]http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007

B. Cross-poses Face Recognition

It is known that 3D pose change in faces is a nonlinear transfer problem, general recognition models are very sensitive to pose change. Therefore, it is challenging to handle the pose based face recognition issue. In this section, the popular CMU Multi-PIE face dataset[6] with 337 subjects is used. Each subject contains 4 different sessions with 15 poses, 20 illuminations, and 6 expressions. The facial images in Session 1 and Session 2 of one person are shown in Fig. 8. In our experiment, we select the first 60 subjects from Session 1 and Session 2. As a result, a smaller session 1 (S1) with 7 images of different poses per class under neutral expression and a smaller session 2 (S2) that is similar to S1 but under smile expression are constructed as domain data. Notably, the raw image pixels are used as features. Specifically, the experimental configurations are set as follows.

S1: One frontal face (0°) per subject is used as source data, one 60° posed face is used as the target training data, and the remaining 5 facial images are used as the target test data.

S2: The experimental configuration is the same as S1.

S1+S2: The two frontal faces (0°) and the two 60° posed faces under neutral and smile expression are used as source data and target training data in the two sessions, respectively. The remaining 10 facial images are used as target test data.

S1 $\rightarrow$ **S2**: S1 is used as source data, the frontal and 60° posed faces in S2 are used as the target training data, and the remaining data in S2 are used as test data.

With above settings, the recognition accuracies of different methods have been shown in Table V. It is clear that the proposed method performs significantly better, which is 5% over other DA methods in handling such pose variation based nonlinear transfer problem. This also demonstrates that the proposed intermediate domain generation based transfer learning can better interpret local generative discrepancy metric (LGDM) and improve the nonlinear local transfer problem. The manifold criterion is then validated.

[6]http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html

TABLE V: Recognition accuracy (%) of different domain adaptation methods on face recognition across poses

Tasks	Naive Comb	A-SVM	SGF [35]	GFK [34]	SA [60]	LTSL [40]	LSDT [41]	**MCTL**
$S1$ ($0° \rightarrow 60°$)	61.0	57.0	53.7	61.0	51.3	56.0	59.7	**65.3**
$S2$ ($0° \rightarrow 60°$)	62.7	62.7	55.0	58.7	62.7	62.7	63.3	**70.0**
$S1 + S2$ ($0° \rightarrow 60°$)	60.2	60.1	53.8	56.3	61.7	60.2	61.7	**68.3**
$S1 \rightarrow S2$	93.6	94.3	92.5	96.7	98.3	97.2	95.8	**98.7**
Average	69.4	68.5	63.8	67.0	68.5	70.3	70.1	**75.6**

TABLE VI: Recognition accuracy (%) of different domain adaptation on handwritten digits recognition

Tasks	Naive Comb	A-SVM	SGF [35]	GFK [34]	SA [60]	LTSL [40]	LSDT [41]	**MCTL**
$M \rightarrow U$	78.8	78.3	79.2	82.6	78.8	83.2	79.3	**87.8**
$S \rightarrow U$	83.6	76.8	77.5	82.7	82.5	83.6	84.7	**84.8**
$M \rightarrow S$	51.9	70.5	51.6	70.5	**74.4**	72.8	69.1	74.0
$U \rightarrow S$	65.3	74.5	70.9	76.7	74.6	65.3	67.4	**83.0**
$U \rightarrow M$	71.7	73.2	71.1	74.9	72.9	71.7	70.5	**81.2**
$S \rightarrow M$	67.6	69.3	66.9	74.5	72.9	67.6	70.0	**74.0**
Average	69.8	73.8	69.5	77.0	76.0	74.0	73.5	**80.8**

TABLE VII: Average performance of all transfer tasks

All Transfer Tasks	LTSL [40]	LSDT [41]	MCTL
Average (%)	69.45	71.08	73.88

C. Cross-domain Handwritten Digits Recognition

Three handwritten digits datasets including MNIST (M)[7], USPS (U)[8] and SEMEION (S)[9] with 10 classes from digit $0 \sim 9$ are used for evaluating the proposed MCTL. The MNIST dataset consists of 70,000 instances of 28×28, the USPS dataset consists of 9298 examples of 16×16, and the SEMEION dataset consists of 2593 images of 16×16. The MNIST dataset is cropped into 16×16. Several images from three datasets are shown in Fig. 9. Each dataset is used as source and target domain alternatively, and 6 cross-domain tasks are explored. Also, 100 samples per class from source domain and 10 samples per class from target domain are randomly selected for training. 5 random splits are used, and the average classification accuracies are reported in Table VI. From the results, we observe that our MCTL outperforms other state-of-the-art methods with 3%, and the significant superiority is therefore proved.

From the whole experiments on 4DA, 4DA-CNN, COIL-20, MSRC and VOC2007, Multi-PIE, and Handwritten digits, we can see that the proposed MCTL shows competitive performance. Although our MCTL shows very slight improvement on several tasks by comparing to state-of-the-art method, the comprehensive superiority of MCTL in all datasets is clearly demonstrated in Table VII, which shows the mean value of all the cross-domain tasks in the datasets. From the results, we can observe that our MCTL outperforms state-of-the-art LTSL and LSDT about 2.8% in average performance on all the transfer tasks explored in this paper.

[7] http://yann.lecun.com/exdb/mnist/
[8] http://www-i6.informatik.rwth-aachen.de/~keysers/usps.html
[9] http://archive.ics.uci.edu/ml/datasets/Semeion+Handwritten+Digit

VIII. Discussion

A. Analysis of MCTL-S

When the condition $\mathbf{X}_{GT} = \mathbf{X}_T$ is strictly satisfied, i.e. perfect domain generation, our model is degenerated into the MCTL-S model, which can be simply formulated as problem (15). The MC-S loss is more similar to a generic manifold regularization, which is built in an ideal condition focusing on the locality structure. Under this case, domain generation relies more on local manifold, regardless of the global property. Therefore, the performance of the MCTL-S with ideal and perfect condition will degrade when global shift of domain data is encountered. The GGDM loss that measures the global structure can be an effective relaxation.

The experimental comparisons on 4DACNN dataset between MCTL and MCTL-S are presented in Table VIII and the comparisons on COIL-20 dataset are shown in Table IX. From the results, we observe that the proposed MCTL and the harsh MCTL-S performs similar performance. This demonstrates that domain generation is a feasible way for unsupervised domain transfer learning. It is also encouraging for us to use deep generative method (e.g. GAN) for transfer learning in the future. The potential problem of GAN is that the similar high-level semantic information across domain may be generated, but the distribution may still be inconsistent.

B. Parameter Setting and Ablation Analysis

In our method, the trade-off coefficients τ and λ_1 are fixed as 1 in experiments. Dimensions of common subspace is set as $d = n$. The Gaussian kernel function $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(- \| \mathbf{x}_i - \mathbf{x}_j \|^2 / 2\sigma^2)$ is used, where σ can be tuned for different tasks, e.g. $\sigma = 1.2$ for 4DA-CNN and $\sigma = 0.8$ for COIL-20. But the linear kernel function is adopted for discussion as it can effectively avoid the influence of kernel parameter. The least square classifier [55] is used in DA experiments except that in COIL-20 experiment, the SVM classifier is used because of its good performance.

In MCTL model, three items such as MMD loss based GGDM term, MC loss based LGDM term and LRC regularization term are included. For better interpreting the effect of

TABLE VIII: Recognition accuracy (%) in 4DACNN dataset

4DA-CNN Tasks	MCTL	MCTL-S
$A \rightarrow D$	95.67	95.71
$C \rightarrow D$	94.69	94.72
$W \rightarrow D$	99.25	99.29
$A \rightarrow C$	87.11	87.05
$W \rightarrow C$	84.73	84.74
$D \rightarrow C$	86.37	86.34
$D \rightarrow A$	92.66	92.65
$W \rightarrow A$	92.06	92.07
$C \rightarrow A$	92.68	92.06
$C \rightarrow W$	93.08	93.04
$D \rightarrow W$	98.49	98.51
$A \rightarrow W$	92.79	92.83
$Average$	92.47	92.47

TABLE IX: Recognition accuracy (%) in COIL-20

COIL-20	MCTL	MCTL-S
$C1 \rightarrow C2$	84.83	85.00
$C2 \rightarrow C1$	83.67	83.67
Average	84.25	84.34

TABLE X: Results of ablation analysis

Tasks	MCTL	no LGDM	no LRC	no GGDM
$C1 \rightarrow C2$	**77.0**	73.0	76.7	76.8
$M \rightarrow U$	71.0	70.0	67.0	**73.0**
$V \rightarrow M$	70.2	70.1	70.1	**70.3**
$Average$	72.7	71.0	71.2	**73.4**

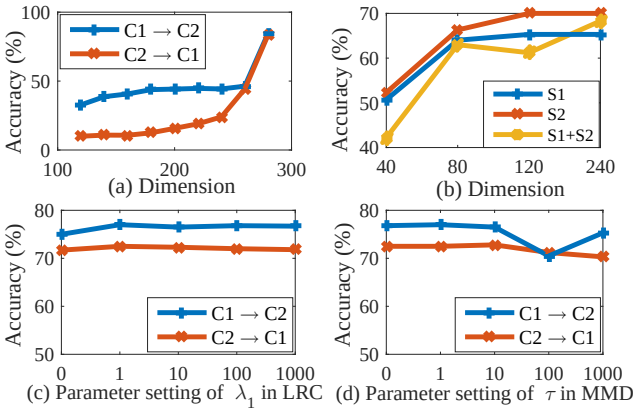

Fig. 10: Dimensionality and parameter sensitivity analysis

each term, the ablation analysis by removing one of them is discussed. Therefore, some extra experiments on the COIL-20 object recognition task (i.e. $C1 \rightarrow C2$), Handwritten Digits recognition task (i.e. $M \rightarrow U$) and MSRC-VOC 2007 image recognition task (i.e. $V \rightarrow M$) are studied for ablation analysis. The experimental results are shown in Table X. We can observe that the LGDM loss plays more important role than GGDM loss, with 2.4% improvement in average. This is reasonable because in many real cross-domain tasks, global transfer may result in negative transfer, due to the local bias problem of domain discrepancy. This further demonstrates the superiority and validity of the proposed MCTL because local discrepancy metric is deserved for transfer learning.

C. Model Dimensionality and Parameter Analysis

Dimensionality Analysis. In MCTL model, a latent common subspace $\boldsymbol{\mathcal{P}}$ is learned. Therefore, the performance variation with different subspace dimensions is studied on the COIL-20 ($C1 \rightarrow C2$ and $C1 \rightarrow C2$) and CMU Multi-PIE face datasets including $S1$, $S2$, and $S1+S2$ tasks. The performance curve with increasing number of the dimensionality d is shown in Fig. 10 (a) and (b). Generally, the recognition performance can be improved with increasing number of dimension.

Parameter Sensitivity Analysis. In MCTL model, there are two trade-off parameters τ and λ_1 involved in parameter tuning. To have an insight of their sensitivity to model, the parameter sensitivity analysis is studied on COIL-20 ($C1 \rightarrow C2$ and $C2 \rightarrow C1$) task by tuning the parameters from $\{0, 1, 10, 100, 1000\}$, respectively. Fig. 10 (c) shows the parameter analysis of λ_1 by fixing $\tau = 1$. Fig. 10 (d) shows the parameter analysis of τ by fixing $\lambda_1 = 1$. For tuning the two parameters simultaneously, we have also provided the 3D surface on COIL-20 dataset in Fig.12 (a) ($C1 \rightarrow C2$) and Fig.12 (b) ($C2 \rightarrow C1$). We can see that the model is robust to the model parameters, without serious fluctuation.

D. Computational Complexity and Time Analysis

In this section, the computational complexity of the Algorithm 1 is presented. The algorithm includes three basic steps: update $\boldsymbol{\mathcal{Z}}$, update $\boldsymbol{\mathcal{J}}$, and update $\boldsymbol{\Phi}$. The computation of $\boldsymbol{\Phi}$ involves eigen-decomposition and matrix multiplication, and the complexity is $O(n^3)$. The computation of updating $\boldsymbol{\mathcal{J}}$ and $\boldsymbol{\mathcal{Z}}$ is $O(n^2)$. Suppose that the number of iterations is T, then the total computational complexity of MCTL can be expressed as $O(Tn^3) + O(Tn^2)$. It is noteworthy that the complexity of Gram matrix computation is not included, because it can be computed in advance without computing in Algorithm 1.

Further, Table XI shows the computational time comparisons on CMU Multi-PIE data ($S1 \rightarrow S2$) and handwritten digits data ($M \rightarrow U$). From Table XI, we observe that the proposed MCTL has also a low computational time. We should claim that the proposed method is better used together with deep models for large-scale data, due to the stronger feature representation capability of deep methods with large-scale data. Notably, all algorithms in experiments are implemented in computer of Intel i5-4460 CPU, 3.20GHz, and 16GB RAM.

E. Model Visualization and Convergence

In this section, the visualization and convergence will be discussed. Pose alignment is a difficult task. Therefore, for better insight of the MCTL model, the feature visualization is explored. We have shown the visualization of CMU PIE. The first row in Fig. 11 illustrates the pose transfer process under Session 1 via MCTL, from which we observe that the generated intermediate domain data by source data inherits similar distribution property of target data.

Further, COIL-20 and handwritten digits datasets are also exploited. The second row of Fig. 11 shows the pose transfer

TABLE XI: Computational time analysis and recognition accuracy (%)

Tasks	SGF [35]	GFK [34]	SA [60]	LTSL [40]	**MCTL**
$S1 \rightarrow S2$	$10.9s$ (92.5%)	$1.5s$ (96.7%)	$4.18s$ (98.3%)	$7.21s$ (97.2%)	$7.62s$ (97.3%)
$M \rightarrow U$	$75s$ (79.2%)	$12.2s$ (82.6%)	$30.5s$ (78.8%)	$62.1s$ (83.2%)	$98.8s$ (87.8%)

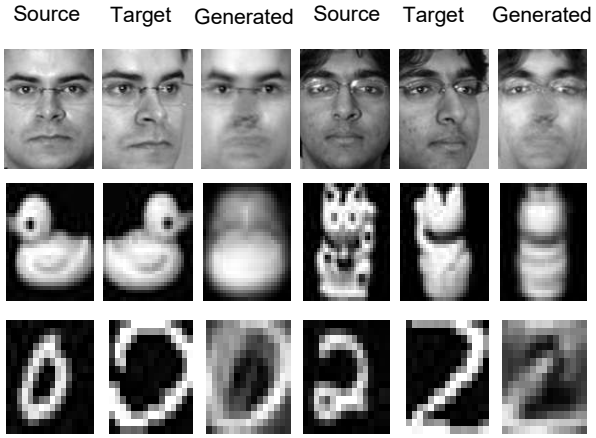

Fig. 11: Visualization of MCTL alignment

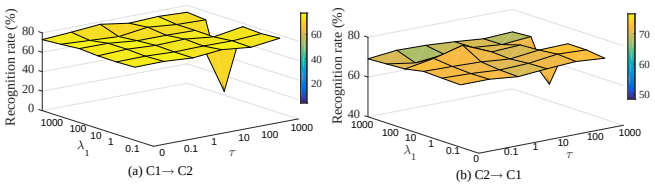

Fig. 12: Parameter sensitivity analysis

process, and the generative data shows a compromise of source and target data in visual disparity. Similarly, the visualization of the generated handwritten digits (intermediate domain) by taking MNIST as the source domain and SEMEION as target domain is shown in the third row of Fig. 11. The effect of domain generation is clearly shown.

Additionally, the convergence of our MCTL method is explored by observing the variation of the objective function. In the experiments, the number of iterations is set to be 15, and the variation of the objective function (i.e. F_{min}) is described in Fig. 13. It is clear that the objective function decreases to a constant value after several iterations, by running the algorithm, on COIL-20 ($C1 \rightarrow C2$) and 4DACNN ($A \rightarrow D$), respectively. Also, the convergence of each term in the MCTL, such as F_{MC} (i.e. MC based LGDM loss), F_{MMD} (i.e. GGDM loss), and $F_{\boldsymbol{Z}}$ (i.e. LRC regularization) are also presented in Fig. 13. We can observe the fast convergence of MCTL after several iterations. Notably, the optimization solver in this paper may not be optimal selection, and the performance may be further fine-tuned with better solvers.

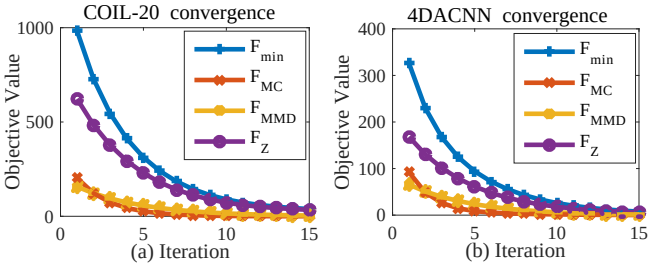

Fig. 13: Convergence of MCTL algorithm

IX. Conclusion

In this paper, we propose a new transfer learning perspective with intermediate domain generation. Specifically, a Manifold Criterion Guided Transfer Learning (MCTL) method is introduced. In previous work, MMD is commonly used for global domain discrepancy minimization and achieves good performance in domain adaptation. However, an open problem, that MMD neglects the locality geometric structure of domain data, is preserved. In order to overcome the bottleneck, motivated by manifold criterion, MCTL is proposed, which aims at generating a new intermediate domain sharing similar distribution with the true target domain. The manifold criterion (MC) implies that the domain adaptation is achieved if MC is satisfied (i.e. minimal domain discrepancy). The rationale behind MC is that if the locality structure is preserved between the generated intermediate domain and the true target domain, then the $i.i.d.$ condition is achieved. Finally, with a MC based LGDM loss, GGDM loss and LRC regularization jointly constructed, MCTL is established. Extensive experiments on benchmark DA datasets demonstrate the superiority of the proposed method over several state-of-the-art DA methods.

Acknowledgment

The authors would like to thank the editor and the anonymous reviewers for their valuable comments and suggestions.

References

[1] H. V. Nguyen, H. T. Ho, V. M. Patel, and R. Chellappa, "Dash-n: Joint hierarchical domain adaptation and feature learning," *IEEE Trans. Image Process*, vol. 24, no. 12, pp. 5479–5491, 2015.
[2] G. Csurka, "Domain adaptation for visual applications: A comprehensive survey," *arXiv*, 2017.
[3] J. Hoffman, B. Kulis, T. Darrell, and K. Saenko, "Discovering latent domains for multisource domain adaptation," in *ECCV*. Springer, 2012, pp. 702–715.
[4] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowle. Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
[5] L. Duan, I. W. Tsang, and D. Xu, "Domain transfer multiple kernel learning," *IEEE Trans. PAMI*, vol. 34, no. 3, pp. 465–479, 2012.
[6] W. Li, Z. Xu, D. Xu, D. Dai, and L. Van Gool, "Domain generalization and adaptation using low rank exemplar svms," *IEEE Trans. PAMI*, 2017.

[7] R. Gopalan, R. Li, and R. Chellappa, "Unsupervised adaptation across domain shifts by generating intermediate data representations," *IEEE Trans. PAMI*, vol. 36, no. 11, pp. 2288–2302, 2014.

[8] L. Duan, D. Xu, I. W. Tsang, and J. Luo, "Visual event recognition in videos by learning from web data," in *CVPR*, 2010, pp. 1959–1966.

[9] B. Kulis, K. Saenko, and T. Darrell, "What you saw is not what you get: Domain adaptation using asymmetric kernel transforms," in *CVPR*, 2011, pp. 1785–1792.

[10] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," 2010.

[11] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[12] L. Duan, I. W. Tsang, D. Xu, and S. J. Maybank, "Domain transfer svm for video concept detection," in *CVPR*, 2009, pp. 1375–1381.

[13] M. Baktashmotlagh, M. T. Harandi, B. C. Lovell, and M. Salzmann, "Unsupervised domain adaptation by domain invariant projection," in *ICCV*, 2013, pp. 769–776.

[14] S. Ben-David, J. Blitzer, K. Crammer, F. Pereira *et al.*, "Analysis of representations for domain adaptation," *NIPS*, vol. 19, p. 137, 2007.

[15] L. Duan, D. Xu, and I. Tsang, "Learning with augmented features for heterogeneous domain adaptation," *arXiv*, 2012.

[16] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction." in *AAAI*, vol. 8, 2008, pp. 677–682.

[17] Y. Li, J. Liu, H. Lu, and S. Ma, "Learning robust face representation with classwise block-diagonal structure," *IEEE Trans. Information Forensics and Security*, vol. 9, no. 12, pp. 2051–2062, 2014.

[18] T. Yao, Y. Pan, C.-W. Ngo, H. Li, and T. Mei, "Semi-supervised domain adaptation with subspace learning for visual recognition," in *ICCV*, 2015, pp. 2142–2150.

[19] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *CVPR*, 2014, pp. 1410–1417.

[20] M. Baktashmotlagh, M. T. Harandi, B. C. Lovell, and M. Salzmann, "Domain adaptation on the statistical manifold," in *CVPR*, 2014, pp. 2481–2488.

[21] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *NIPS*, 2016, pp. 136–144.

[22] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," *arXiv*, 2014.

[23] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, A. J. Smola *et al.*, "A kernel method for the two-sample-problem," *NIPS*, p. 513, 2007.

[24] J. Hu, J. Lu, and Y.-P. Tan, "Deep transfer metric learning," in *CVPR*, 2015, pp. 325–333.

[25] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *ICML*, 2015, pp. 97–105.

[26] J. Hoffman, T. Darrell, and K. Saenko, "Continuous manifold based adaptation for evolving visual domains," in *CVPR*, 2014, pp. 867–874.

[27] X. Yang, M. Wang, R. Hong, Q. Tian, and Y. Rui, "Enhancing person re-identification in a self-trained subspace," *TOMM*, vol. 13, no. 3, pp. 27:1–27:23, 2017.

[28] T. Kanamori, S. Hido, and M. Sugiyama, "Efficient direct density ratio estimation for non-stationarity adaptation and outlier detection," in *NIPS*, 2009, pp. 809–816.

[29] J. Yang, R. Yan, and A. G. Hauptmann, "Cross-domain video concept detection using adaptive svms," in *ACM MM*, 2007, pp. 188–197.

[30] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A dasvm classification technique and a circular validation strategy," *IEEE Trans. PAMI*, vol. 32, no. 5, pp. 770–787, 2010.

[31] L. Zhang and D. Zhang, "Robust visual knowledge transfer via extreme learning machine-based domain adpatation," *IEEE Trans. Image Processing*, vol. 25, no. 3, pp. 4959–4973, 2016.

[32] W. Li, L. Duan, D. Xu, and I. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *IEEE Trans. PAMI*, vol. 36, no. 6, pp. 1134–1148, 2014.

[33] J. Hoffman, E. Rodner, J. Donahue, B. Kulis, and K. Saenko, "Asymmetric and category invariant feature transformations for domain adaptation," *IJCV*, vol. 109, no. 1, pp. 28–41, 2014.

[34] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *CVPR*, 2012, pp. 2066–2073.

[35] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *Proc. ICCV*, 2011, pp. 999–1006.

[36] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Scholköpf, and A. Smola, "A kernel two-sample test," *JMLR*, pp. 723–773, 2012.

[37] A. Iyer, J. S. Nath, and S. Sarawagi, "Maximum mean discrepancy for class ratio estimation: Convergence bounds and kernel selection." in *ICML*, 2014, pp. 530–538.

[38] M. Long, G. Ding, J. Wang, J. Sun, Y. Guo, and P. S. Yu, "Transfer sparse coding for robust image representation," in *ICCV*, 2013, pp. 407–414.

[39] I.-H. Jhuo, D. Liu, D. Lee, and S.-F. Chang, "Robust visual domain adaptation with low-rank reconstruction," in *CVPR*, 2012, pp. 2168–2175.

[40] M. Shao, D. Kit, and Y. Fu, "Generalized transfer subspace learning through low-rank constraint," *IJCV*, vol. 109, no. 1-2, pp. 74–93, 2014.

[41] L. Zhang, W. Zuo, and D. Zhang, "Lsdt: Latent sparse domain transfer learning for visual adaptation," *IEEE Trans Image Processing*, vol. 25, no. 3, pp. 1177–1191, 2016.

[42] L. Zhang, S. K. Jha, T. Liu, and G. Pei, "Discriminative kernel transfer learning via l2,1-norm minimization," in *IJCNN*, 2016.

[43] Y. Xu, X. Fang, J. Wu, X. Li, and D. Zhang, "Discriminative transfer subspace learning via low-rank and sparse representation." *IEEE Trans Image Processing*, vol. 25, no. 2, pp. 850–863, 2015.

[44] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *ICCV*, 2015, pp. 4068–4076.

[45] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *ICML*, 2011, pp. 513–520.

[46] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *CVPR*, 2014, pp. 1717–1724.

[47] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer learning from deep features for remote sensing and poverty mapping," *arXiv*, 2015.

[48] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," in *ICML*, 2014, pp. 647–655.

[49] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "Cnn features off-the-shelf: an astounding baseline for recognition," in *CVPR*, 2014, pp. 806–813.

[50] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," 2017, cVPR.

[51] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014, pp. 2672–2680.

[52] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," *NIPS*, pp. 612–620, 2011.

[53] J. F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *Siam Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2008.

[54] L. Rosasco, A. Verri, M. Santoro, S. Mosci, and S. Villa, "Iterative projection methods for structured sparsity regularization," *Computation*, 2009.

[55] T. Kanamori, S. Hido, and M. Sugiyama, "A least-squares approach to direct importance estimation," *JMLR*, vol. 10, no. Jul, pp. 1391–1445, 2009.

[56] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. PAMI*, vol. 31, no. 2, pp. 210–227, 2009.

[57] A. Gaidon, G. Zen, and J. A. Rodriguez-Serrano, "Self-learning camera: Autonomous adaptation of object detectors to unlabeled video streams," *arXiv*, 2014.

[58] B. Gong, K. Grauman, and F. Sha, "Learning kernels for unsupervised domain adaptation with applications to visual object recognition," *IJCV*, vol. 109, no. 1-2, pp. 3–27, 2014.

[59] D. Liu, G. Hua, and T. Chen, "A hierarchical visual model for video object summarization," *IEEE Trans. PAMI*, vol. 32, no. 12, pp. 2178–2190, 2010.

[60] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *ICCV*, 2014, pp. 2960–2967.

[61] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," 2011, pp. 999–1006.

[62] S. F. Chang, D. T. Lee, D. Liu, and I. Jhuo, "Robust visual domain adaptation with low-rank reconstruction," in *CVPR*, 2013, pp. 2168–2175.

[63] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *NIPS*, pp. 1097–1105, 2012.

[64] C. Rate and C. Retrieval, "Columbia object image library (coil-20)," *Computer*, 2011.

Lei Zhang (M'14-SM'18) received his Ph.D degree in Circuits and Systems from the College of Communication Engineering, Chongqing University, Chongqing, China, in 2013. He was selected as a Hong Kong Scholar in China in 2013, and worked as a Post-Doctoral Fellow with The Hong Kong Polytechnic University, Hong Kong, from 2013 to 2015. He is currently a Professor/Distinguished Research Fellow with Chongqing University. He has authored more than 70 scientific papers in top journals and conferences, including the IEEE T-NNLS, IEEE T-IP, IEEE T-MM, IEEE T-IM, IEEE T-SMCA, Neurocomputing, Information Fusion, etc. His current research interests include machine learning, pattern recognition, computer vision and intelligent systems. Dr. Zhang was a recipient of Outstanding Reviewer Award of Sensor Review Journal in 2016, Outstanding Doctoral Dissertation Award of Chongqing, China, in 2015, Hong Kong Scholar Award in 2014, Academy Award for Youth Innovation of Chongqing University in 2013 and the New Academic Researcher Award for Doctoral Candidates from the Ministry of Education, China, in 2012.

Shanshan Wang received BE and ME from the ChongQing University in 2010 and 2013, respectively. She is currently pursuing the Ph.D. degree at ChongQing University. Her current research interests include machine learning, pattern recognition, computer vision.

Guangbin Huang (M'98-SM'04) received the B.Sc. degree in applied mathematics and the M.Eng. degree in computer engineering from Northeastern University, China, in 1991 and 1994, respectively, and Ph.D. degree in electrical engineering from Nanyang Technological University, Singapore, in 1999. During undergraduate period, he also concurrently studied with the Applied Mathematics Department and Wireless Communication Department, Northeastern University, China. He serves as an Associate Editor of Neurocomputing, Neural Networks, Cognitive Computation, and the IEEE Transactions on Cybernetics. His current research interests include machine learning, computational intelligence, and extreme learning machines. He was a Research Fellow with the Singapore Institute of Manufacturing Technology, from 1998 to 2001, where he has led/implemented several key industrial projects (e.g., a Chief Designer and a Technical Leader of Singapore Changi Airport Cargo Terminal Upgrading Project). From 2001, he has been an Assistant Professor and an Associate Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He received the best paper award of the IEEE Transactions on Neural Networks and Learning Systems in 2013.

Wangmeng Zuo received the Ph.D. degree in computer application technology from the Harbin Institute of Technology, Harbin, China, in 2007. From July 2004 to December 2004, from November 2005 to August 2006, and from July 2007 to February 2008, he was a Research Assistant at the Department of Computing, Hong Kong Polytechnic University, Hong Kong. From August 2009 to February 2010, he was a Visiting Professor in Microsoft Research Asia. He is currently an Associate Professor in the School of Computer Science and Technology, Harbin Institute of Technology. Dr. Zuo has published more than 60 papers in top tier academic journals and conferences. His current research interests include image modeling and blind restoration, discriminative learning, biometrics, and 3D vision. Dr. Zuo is an Associate Editor of the IET Biometrics. He is a senior member of the IEEE.

Jian Yang received the PhD degree from Nanjing University of Science and Technology (NUST), on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a postdoctoral researcher at the University of Zaragoza. From 2004 to 2006, he was a Postdoctoral Fellow at Biometrics Centre of Hong Kong Polytechnic University. From 2006 to 2007, he was a Postdoctoral Fellow at Department of Computer Science of New Jersey Institute of Technology. Now, he is a Chang-Jiang professor in the School of Computer Science and Technology of NUST. He is the author of more than 100 scientific papers in pattern recognition and computer vision. His journal papers have been cited more than 4000 times in the ISI Web of Science, and 9000 times in the Web of Scholar Google. His research interests include pattern recognition, computer vision and machine learning. Currently, he is/was an associate editor of Pattern Recognition Letters, IEEE Trans. Neural Networks and Learning Systems, and Neurocomputing. He is a Fellow of IAPR.

David Zhang (F'09) graduated in Computer Science from Peking University in 1974. He received his MSc in 1982 and his PhD in 1985 in Computer Science from the Harbin Institute of Technology (HIT), respectively. From 1986 to 1988 he was a Postdoctoral Fellow at Tsinghua University and then an Associate Professor at the Academia Sinica, Beijing. In 1994 he received his second PhD in Electrical and Computer Engineering from the University of Waterloo, Ontario, Canada. He is a Chair Professor since 2005 at the Hong Kong Polytechnic University where he is the Founding Director of the Biometrics Research Centre (UGC/CRC) supported by the Hong Kong SAR Government in 1998. He also serves as Visiting Chair Professor in Tsinghua University, and Adjunct Professor in Peking University, Shanghai Jiao Tong University, HIT, and the University of Waterloo. He is the Founder and Editor-in-Chief, International Journal of Image and Graphics (IJIG); Book Editor, Springer International Series on Biometrics (KISB); Organizer, the International Conference on Biometrics Authentication (ICBA); Associate Editor of more than ten international journals including IEEE TRANSACTIONS and so on; and the author of more than 10 books, over 300 international journal papers and 30 patents from USA/Japan/HK/China. Professor Zhang is a Croucher Senior Research Fellow, Distinguished Speaker of the IEEE Computer Society, and a Fellow of both IEEE and IAPR.