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, and David Zhang®, Fellow, IEEE

Abstract—In many practical transfer learning scenarios, the feature distribution is different across the source and target domains (i.e., nonindependent identical distribution). Maximum mean discrepancy (MMD), as a domain discrepancy metric, has achieved promising performance in unsupervised domain adaptation (DA). We argue that the MMD-based DA methods ignore the data locality structure, which, up 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-based intermediate domain generation learning called Manifold Criterion guided Transfer Learning (MCTL) is proposed in this paper. The merits of the proposed MCTL are fourfold: 1) the concept of manifold criterion (MC) is first proposed as a measure validating the distribution matching across domains, and DA 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 is presented, such that both the global and local discrepancies can be effectively and positively reduced; and 4) a simplified version of MCTL called MCTL-S is presented under a perfect domain generation assumption for more generic transfer learning scenario. Experiments on a number of benchmark visual transfer tasks demonstrate the superiority of the proposed MC guided generative transfer method, by comparing with the other state-of-the-art methods. The source code is available in https://github.com/wangshanshanCQU/MCTL.

Index Terms—Discrepancy metric, domain adaptation (DA), domain generation, manifold criterion (MC), transfer learning.

I. INTRODUCTION

STATISTICAL machine learning models rely heavily on the assumption that the data used for training and testing 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]–[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. In addition, 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 (DA) [8], [9], which aim to learn domain-invariant models across source and target domains, have been investigated. DA [10]–[12], as one kind of transfer learning (TL) perspective, addresses the problem that data are from two related but different domains [13], [14]. DA establishes knowledge transfer from the labeled source domain to the unlabeled target domain by exploring the domain-invariant structures that bridge different domains with substantial distribution discrepancy. In terms of the accessibility of target data labels in transfer learning, DA methods can be divided into three categories: supervised [15], [16], semisupervised [5], [17], [18], and unsupervised [19]–[21].

In this paper, we focus on unsupervised transfer learning, where the target data labels are unavailable in the 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, which 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). In addition, in the framework of deep transfer learning [24], the 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 the imanifold assumption that the data with local similarity are with similar labels. Motivated by the manifold assumption, a novel manifold criterion (MC) is proposed in this paper, which is similar but very different from the conventional manifold algorithms that the MC actually acts as a generative transfer criterion for the unsupervised DA.

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 the manifold assumption [26], an objective generative transfer metric, 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 $x'_j$ by using the source data is also close to $x_j$, we recognize that the generated intermediate domain data share 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 are 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 the graph Laplacian regularization [27], which reflects the affinity of the raw data.

Specifically, with the proposed MC, a Manifold Criterion guided Transfer Learning (MCTL) is proposed, which aims to pursue a latent common subspace via a projection matrix $P$ for source and target domains. In the common subspace, a generative transfer matrix $Z$ 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 an unsupervised manner. The findings and analysis show that the proposed MC can be used to reduce the local domain discrepancy.

In addition, 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 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 paper are fourfold.

1) We propose an unsupervised manifold criterion generative transfer learning method, which aims to generate a new intermediate target domain that holds similar distribution with true target data by leveraging the source data as basis. The proposed 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.

2) In order to keep the global distribution consistency, a global generative discrepancy metric (GGDM), which 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.

3) For improving the correlation between the source data and the generative target data, the LRC regularization on the transfer matrix $Z$ is integrated in MCTL, such that the block-diagonal property can be utilized for preventing the domain transfer from distortion and negative transfer.

4) Under the MCTL framework, for a more generic case, a simplified version of MCTL (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 MC. In Section IV, the proposed MCTL method and optimization are formulated. In Section V, the MCTL-S 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, this paper is concluded in Section IX.

II. RELATED WORK

A. Shallow Transfer Learning

A lot of transfer learning methods are proposed to tackle the heterogeneous DA problems. Generally, these methods can be divided into the following three categories.

1) 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 support vector machine (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 and Marconcini [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 and a domain transfer MKL [5] which learn an SVM classifier and a kernel function simultaneously for classifier adaptation. Zhang and Zhang [31] proposed a robust classifier transfer method [extreme domain adaptation (EDA)] that was modeled based on extreme learning machine (ELM) and manifold regularization for visual recognition.

2) Feature Augmentation-/Transformation-Based Approaches: Li et al. [32] proposed a heterogeneous feature augmentation that tends to learn a transformed feature space for DA. Kulis et al. [9] proposed an asymmetric regularized cross-domain transform method for learning a transformation metric. Hoffman et al. [33] proposed a Max-Margin Domain Transforms that optimized a category specific transformation for domain transfer. Gong et al. [34] proposed a Geodesic Flow Kernel method that 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 across the geodesic between source and target data is sampled, and the source data are projected into the subspaces for discriminative classifier learning. An unsupervised feature transformation approach, transfer component analysis [11], was proposed to discover the common features having the same marginal distribution by using MMD as nonparametric discrepancy metric. MMD [23], [36], [37] is often used in transfer learning. Long et al. [38] proposed a transfer sparse coding approach to construct the robust sparse representations by using empirical MMD as the distance measure. The transfer joint matching proposed by Long et al. [19] tends to learn a nonlinear transformation by minimizing the MMD-based distribution discrepancy.

3) Feature Representation-Based Approaches: Different from the above-mentioned methods, DA is achieved by representing across domain features. Jhuo et al. [39] proposed a robust domain adaptation with low-rank (RDALR) method, in which the source data are reconstructed with target domain by using low-rank modeling. Similarly, Shao et al. [40] proposed a latent transfer subspace learning (LTSL) method by prelearning a subspace using principal component analysis (PCA) or linear discriminant analysis (LDA), and then, the low-rank representation across domain is modeled. Zhang et al. [41], [42] proposed the Latent Sparse Domain Transfer (LSDT) and Discriminative Kernel Transfer Learning methods for visual adaptation, by jointly learning a subspace projection and sparse reconstruction across domain. Furthermore, Xu et al. [43] proposed a discriminative transfer subspace learning (DTSL) method that combines the low-rank and sparse constraints 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 DA 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]–[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 the ImageNet for high-level feature extractor. Tzeng et al. [44] proposed a DDC method that simultaneously achieves knowledge transfer between domains and tasks by using CNN. Long et al. [25] proposed a deep adaptation network (DAN) method by imposing the MMD loss on the high-level features across domains. In addition, Long et al. [21] also proposed a residual transfer network (RTN) that 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 a large annotated image set. In addition, Hu et al. [24] proposed a non-CNN-based deep transfer metric learning method to learn a set of hierarchical nonlinear transformations for achieving the cross-domain visual recognition.

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

In this paper, 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 preextracted deep features.

C. Differences Between MCTL and Other Reconstruction Transfer Methodologies

The proposed MCTL is partly related by reconstruction transfer methods, such as LTSL [40], LSDT [41], and DTSL [43], 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.

1) **Reconstruction Transfer:** As the name implies, a reconstruction matrix is expected for domain correspondence. In LTSL, subspace projection $W$ is prelearned by off-the-shelf methods, such as PCA and LDA. Then, projected source data $WX_s$ are used to reconstruct the projected target data $WX_T$ via the low-rank constraint. The subspace may be suboptimal leading to a possible local optimum of $Z$. Furthermore, the LSDT method was proposed for realizing DA by exploiting the cross-domain sparse reconstruction in some latent subspace, simultaneously. The DTSL was proposed by posing the hybrid regularization of sparsity and low-rank constraints for learning a more robust reconstruction transfer matrix. Reconstruction transfer always expresses the target domain by leveraging the 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.

2) **Generative Transfer:** The proposed MCTL method introduces a generative transfer learning concept, which aims to realize an intermediate domain generation by constructing an MC loss. The motivation is that the DA 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 the following.

1) DA is recognized to be a domain generation problem, instead of a domain alignment problem.
2) The MC 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.

3) **Similarity and Relationship:** The reconstruction transfer and generative transfer are similar and related in the following 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 DA.
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 the manifold hypothesis works but under the data of independent identically distribution (i.i.d.). Therefore, we could have a try to build an MC 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 the manifold learning and regularization, we propose a novel MC that is utilized as the generative discrepancy metric.

In semisupervised learning (SSL), manifold regularization is often used but under i.i.d. condition. However, transfer learning is different from SSL that domain the data do not satisfy the i.i.d. condition. In this paper, it should be figured out that if the intermediate domain can be generated via the MC guided objective function, then the distribution of the generated intermediate domain and the true target domain is recognized to be matched.

The idea of MC is described in Fig. 2. We observe that a projection matrix $P$ is first learned for some common subspace projection, and then, a generative transfer matrix $Z$ is learned for the 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 have similar affinity structure with the true target domain, i.e., MC is satisfied, we can have a conclusion that the generative data shares similar distribution with the target domain.

Notably, different from the reconstruction-based DA methods, in this paper, we tend to generate an intermediate domain by leveraging the 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 the transfer learning models. To be frank, 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 comply with the manifold assumption originated from the true target domain data. If the MC is satisfied (i.e., i.i.d. is achieved), then DA or distribution alignment is completed, which is the principle of MCTL.

### IV. MANIFOLD CRITERION GUIDED TRANSFER LEARNING

#### A. Notations

In this paper, source and target domains are defined by the subscripts $S$ and $T$. Training set of source and target
domains are defined as $\phi(X_S) \in R^{m \times n_S}$ and $\phi(X_T) \in R^{m \times n_T}$. $\phi(X_GT) \in R^{m \times n_{GT}}$ denotes the generative target domain, where $\phi$ denotes an implicit but generic transformation, $m$ denotes dimensionality, and $n_S$ and $n_T$ denote the number of samples in source and target domains, respectively. Let $X = [X_S; X_T]$, then $\phi(X) \in R^{m \times n}$, where $n = n_S + n_T$. Let $P \in R^{m \times d}$ be the basis transformation that maps raw data space from $R^m$ to a latent subspace $R^d$. $Z \in R^{n \times n_{GT}}$ represents the generative transfer matrix, $I$ denotes the identity matrix, and $\| \cdot \|_F$ and $\| \cdot \|_2$ denote the $l_F$- and $l_2$-norms, respectively. The superscript $T$ denotes the transpose operator, and $Tr(\cdot)$ denotes the matrix trace operator.

In RKHS, the kernel Gram matrix $K$ is defined as $[K_{i,j}] = < \phi(x_i), \phi(x_j)>_{\Phi1} = \phi(x_i)^T \phi(x_j) = k(x_i, x_j)$, where $k$ is a kernel function. Let $K = \phi(X)^T \phi(X)$, $K_S = \phi(X)^T \phi(X_S)$, and $K_T = \phi(X)^T \phi(X_T)$, and it is easy to get that $K \in R^{n \times n}$, $K_S \in R^{n \times n_S}$, and $K_T \in R^{n \times n_T}$.

B. Problem Formulation

In this section, the proposed MCTL method is presented in Fig. 2, in which the distribution between the Generated intermediate Target domain ($D_{GT}$) and the true Target 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 MC as domain discrepancy metric. Specifically, two generative discrepancy metrics (LGDM versus GGDM) for measuring the domain discrepancy locally and globally are proposed. Overall, the model is composed of three items. The first item is the MC-based LGDM loss that is used to measure the local domain discrepancy with the MC by exploiting the locality of target data. The second item is the GGDM loss that is applied to minimize the global domain discrepancy of marginal distributions between the generated intermediate target domain and the true target domain. The third item is the LRC regularization (low-rank constraint) that is carried out to keep the generalization of $Z$. A detailed MCTL method is described in the follows.

1) MC-Based Local Generative Discrepancy Metric: The MC-based LGDM loss is used to enhance the distribution consistency between source and target domains indirectly, by constraining the generative target data with MC. For convenience, $\phi(x_{GT}^p)$ is defined as a sample in $\phi(X_{GT})$, and $\phi(x_T^i)$ is defined as a sample in $\phi(X_T)$. We claim that the distribution consistency between $\phi(X_{GT})$ and $\phi(X_T)$ is achieved, i.e., domain transfer is done, only if the two sets satisfy the following MC, which can be formulated as:

$$LGDM(D_{GT}, D_T) = \frac{1}{n_T} \sum_{i=1}^{n_T} \frac{1}{n_T} \sum_{i=1}^{n_T} \| \phi(x_{GT}^p) - \phi(x_T^i) \|_2^2$$

$$= Tr(\phi(X_{GT})D(\phi(X_{GT})^T)$$

$$+ Tr(\phi(X_T)D(\phi(X_T)^T)$$

$$- 2 Tr(\phi(X_{GT})W(\phi(X_T)^T)$$

where $W \in R^{n_{GT} \times n_T}$ is the affinity matrix, described as

$$W_{pq} = \begin{cases} 1, & \text{if } x_{GT}^p \in NN_k(x_T^i) \text{ or } x_T^i \in NN_k(x_{GT}^p) \\ 0, & \text{otherwise} \end{cases}$$

and $NN_k(x)$ represents the $k$th nearest neighbors of sample $x$. The matrix $D \in R^{n_{GT} \times n_T}$ is a diagonal matrix with entries $D_{pp} = \sum_{q} W_{pq}$, $p = 1, \ldots, n_T$. As claimed earlier, $\Phi1^T = \Phi1^T \phi(X)^T$, the projected source data and target data can be expressed as $\Phi1^T \phi(x_{GT})$ and $\Phi1^T \phi(x_T)$. By substituting $\phi(X_{GT}) = \phi(X_S)Z$ and the Gram matrix after projection (i.e., $\Phi1^T K_S$ and $\Phi1^T K_T$ ) into (1), the MC-based LGDM loss can be further formulated as

$$\min_{\phi, Z} \frac{1}{(n_T)^2} Tr(\phi1^T K_S ZD(\phi1^T K_S Z)^T) + \frac{1}{(n_T)^2} Tr(\phi1^T K_T D(\phi1^T K_T)^T) - \frac{2}{(n_T)^2} Tr(\phi1^T K_S Z W(\phi1^T K_T)^T).$$

From (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 (2) and the manifold embedding or regularization is that we aim to produce the i.i.d. assumption with an MC, 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 GGDM is proposed by minimizing the discrepancy as follows:

$$GGDM(D_{GT}, D_T) = \frac{1}{n_T} \sum_{i=1}^{n_T} \| (\phi(x_{GT}^i) - \phi(x_T^i)) \|_2^2$$

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 domains by using a projection matrix $P$. Therefore, by projecting $\phi(X_{GT})$ and $\phi(X_T)$ to the subspace, the GGDM loss after projection can be formulated as follows. Considering that $\phi(X_{GT}) = \phi(X_S)Z$, by substituting it in (3), there is

$$GGDM(D_{GT}, D_T) = \frac{1}{n_T} \sum_{i=1}^{n_T} \| P^T (\phi(X_{GT}^i) - \phi(X_T^i)) \|_2^2$$

$$= \frac{1}{n_T} \| P^T (\phi(X_S)Z - \phi(X_T)) \|_2^2$$

where $I$ represents a full one column vector.

The projection matrix $P$ is a linear transformation, which can be represented as some linear combination of the training data, i.e., $P^T = \Phi1^T \phi(X)^T$, where $\Phi$ denotes the linear combination coefficient matrix. Then, the projected source data can be expressed as $\Phi1^T \phi(x_{GT})$, and the projected target data can be expressed as $\Phi1^T \phi(x_T)$. With the kernel trick, the inner product of implicit transformation is represented as
a Gram matrix, from raw space to RKHS. As described in Section IV-A, let $K_S = \phi(X)^T \phi(X)$ and $K_T = \phi(X_T)^T \phi(X_T)$, the source domain and target domains can be expressed simply as $\Phi^T K_S$ and $\Phi^T K_T$, respectively. Therefore, the GGDM loss is formulated as

$$\min_{\Phi, Z} \frac{1}{n_T} \left\| \Phi^T (K_S Z - K_T) \right\|_2^2.$$  

(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 a critical role in the proposed model. In this paper, a general transfer matrix $Z$ 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 $Z$ as a choice for enhancing the domain correlation. In our MCTL, LRC, which 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 lie 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 $Z$ is approximately block-wise. Therefore, LRC regularization is necessary. Considering the nonconvexity property of rank function which is NP-hard, the nuclear norm $\|Z\|_*$ is used as a rank approximation in this paper.

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

$$\min_{\Phi, Z} \frac{1}{(n_T)^2} Tr(\Phi^T K_S Z D(\Phi^T K_S Z)^T)$$

$$+ \frac{1}{(n_T)^2} Tr(\Phi^T K_T D(\Phi^T K_T)^T)$$

$$- \frac{2}{(n_T)^2} Tr(\Phi^T K_S Z W(\Phi^T K_T)^T)$$

$$+ \frac{1}{n_T} \left\| \Phi^T (K_S Z - K_T) \right\|_2^2$$

$$+ \lambda_1 \|Z\|_*$$

s.t. $\Phi^T K \Phi = I$.  

(6)

where $\tau$ and $\lambda_1$ are the tradeoff parameters. The rows of $P$ are required to be orthogonal and normalized to unit norm for preventing trivial solutions by enforcing $P^T P = I$, which can be further rewritten as $\Phi^T K \Phi = I$, an equality constraint. Obviously, the model is nonconvex 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 $\Phi$ and $Z$ 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 $Z$ is fixed, a general eigenvalue decomposition is used for solving $\Phi$. Second, when $\Phi$ is fixed, the inexact augmented Lagrangian multiplier (IALM) and gradient descent are used to solve $Z$. In the following, the optimization details of the proposed method are presented.

By introducing an auxiliary variable $J$, the problem (6) can be written as follows. Furthermore, with the augmented Lagrangian function [52], the model can be written as

$$\min_{\Phi, Z, J} \frac{1}{(n_T)^2} (Tr(\Phi^T K_S Z D(\Phi^T K_S Z)^T)$$

$$+ Tr(\Phi^T K_T D(\Phi^T K_T)^T) - 2Tr(\Phi^T K_S Z W(\Phi^T K_T)^T))$$

$$+ \frac{\tau}{(n_T)^2} \|Z\|_*$$

$$+ \lambda_1 \|Z\|_*$$

$$- K_T Z^T (K_S)^T + K_T (K_T)^T \Phi + \lambda_1 \|J\|_*$$

$$+ Tr(\mathcal{R}_1^T (Z - J)) + \frac{\mu}{2} \left\| Z - J \right\|_F^2$$

(7)

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

In the following, we present how to optimize the three variables $\Phi$, $J$, and $Z$ in the problem (7) based on the eigenvalue decomposition, IALM, and gradient descent in stepwise.

1) Update $\Phi$: By frozen $Z$ and $J$, $\Phi$ can be solved as

$$\Phi^* = \arg \min_{\Phi} \frac{1}{(n_T)^2} (Tr(\Phi^T K_S Z D(\Phi^T K_S Z)^T)$$

$$+ Tr(\Phi^T K_T D(\Phi^T K_T)^T)$$

$$- 2Tr(\Phi^T K_S Z W(\Phi^T K_T)^T))$$

$$+ \frac{\tau}{(n_T)^2} \|Z\|_*$$

$$- K_T Z^T (K_S)^T + K_T (K_T)^T \Phi$$

$$s.t. \quad \Phi^T K \Phi = I.$$  

(8)

We can derive the solution $\Phi_K$ of the $K$th iteration in columnwise. To obtain the $i$th column vector in $\Phi_K$ by setting the partial derivative of problem (8) with respect to $\Phi_{K(i,i)}$ to be zero, there is

$$\frac{1}{(n_T)^2} (K_S Z D Z^T (K_S)^T + K_T D (K_T)^T - K_S Z W (K_T)^T)$$

$$- K_T W Z^T (K_S)^T \Phi_{K(i,i)} + \frac{\tau}{(n_T)^2} (K_S Z^T (K_S)^T)$$

$$- K_S Z^T (K_T)^T - K_T Z^T (K_S)^T + K_T (K_T)^T \Phi_{K(i,i)}$$

$$= - \lambda K_{K(i,i)}.$$  

(9)

It is clear that $\Phi_K$ can be obtained by solving an eigen-decomposition problem, and $\Phi_{K(i,i)}$ is the $i$th eigenvector corresponding to the $i$th smallest eigenvalue.

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

$$J_{K+1} = \min_{J_K} \lambda_1 \|J_K\|_*$$

$$+ Tr(\mathcal{R}_1^T (Z_K - J_K))$$

$$+ \frac{\mu K}{2} \left\| Z_K - J_K \right\|_F^2.$$  

(10)
where

\[ \frac{\nabla}{\partial} J_{\text{of } Z} \text{ with respect to } \]

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

It can be further rewritten as

\[ J_{K+1} = \min_{J_K} \lambda_1 \| J_K \|_s + \frac{\mu K}{2} \| J_K - \left( Z_K + \frac{R_{1K}}{\mu K} \right) \|_F^2. \]  

(11)

Problem (11) can be efficiently solved using the singular value thresholding (SVT) operator [53], which contains two major steps. First, singular value decomposition is conducted on matrix \( S = Z_K + (R_{1K}/\mu K) \), and get \( S = U_S \sum_S V_S \), where \( \sum_S = \text{diag}((\sigma_i)_{1 \leq i \leq r}) \), \( \sigma_i \) is the singular value with rank \( r \). Second, the optimal solution \( J_{K+1} \) is then obtained by thresholding the singular values as \( J_{K+1} = U_S \Omega_{1/\mu K}(\sum_S)V_S \), where \( \Omega_{1/\mu K}(\sum_S) = \text{diag}((\sigma_i - (1/\mu K))_+) \), and \([*]_+\) denotes the positive value operator.

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

\[ \min_{Z} \frac{1}{2} \left( \frac{1}{n_T} (\text{Tr} (\Phi^T K_S Z \Phi^T K_S Z)^T) - 2 \text{Tr} (\Phi^T K_S Z W (\Phi^T K_T)^T) + \text{Tr} (R_1^T (Z - J)) \right) \]

\[ + \frac{\mu}{2} \| Z - J \|^2_F + \frac{\tau}{n_T^2} \Phi^T (K_S Z 1^T Z (K_S)^T) + \frac{2}{n_T^2} (K_S^T \Phi^T K_T W)^T + \frac{2\tau}{n_T^2} (K_S^T \Phi^T K_T W)^T \]

\[ - K_S Z 1 (K_T)^T - K_T 1 Z (K_S)^T \Phi \].

(12)

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

\[ Z_{K+1} = Z_K - a \cdot \nabla(Z) \]

where \( \nabla(Z) \) denotes the gradient, which is calculated as

\[ \nabla(Z) = \frac{2}{n_T^2} ((K_S)^T \Phi^T K_S Z D - (K_S)^T \Phi^T K_T W) \]

\[ + R_1 + \mu (Z - J) + \frac{2\tau}{n_T^2} (K_S^T \Phi^T K_T W) \]

\[ - \frac{2\tau}{n_T^2} (K_S^T \Phi^T K_T W)^T + \frac{2\tau}{n_T^2} (K_S^T \Phi^T K_T W)^T \].

(14)

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

V. 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, in this section, by using the MC and considering the generic manifold embedding, for model simplicity, we rewrite an MCTL-S, as illustrated in Fig. 3.

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 are strictly the same as the true target data, i.e., \( X_{\text{GT}} = X_T \) \((D_{\text{GT}} \text{ coincides with } D_T)\). then MCTL-S is formulated as

\[ \min_{\Phi, Z} \frac{1}{2} \text{Tr} (\Phi^T K_S Z L (\Phi^T K_S Z)^T) \]

\[ + \tau \left( \frac{1}{n_T} \Phi^T (K_S Z - K_T) \right)^2 + \lambda_1 || Z ||_s \]

(15)

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

The following experimental results in Tables VIII and IX also prove that both the harsh MCTL-S model and the MCTL can achieve good performance. This demonstrates that the MC-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 $\Phi$ and $Z$, the optimization procedure of the MCTL-S method is illustrated as follows.

1) Update $\Phi$: In the MCTL-S model, by frozen $Z$ and $\mathcal{J}$, the derivative of the objective function (15) with respect to $\Phi_{K(i,:)}$ is set as zero, and there is

$$
\nabla_{\Phi_{K(i,:)}} J = 0
$$

$$
\frac{2}{(n_T)^2}(K_S^T \mathcal{L}_T (K_S)^T) \Phi_{K(i,:)} + \frac{\tau}{(n_T)^2} (K_S^T \mathcal{L}_T (K_S)^T) - K_S^T (K_T)^T - K_T^T (K_S)^T + K_T^T (K_T)^T) \Phi_{K(i,:)} = -\lambda K \Phi_{K(i,:)}.
$$

Therefore, $\Phi_K$ in iteration $K$ can be obtained by solving an eigenvalue decomposition problem, and $\Phi_{K(i,:)}$ is the $i$th eigenvector corresponding to the $i$th smallest eigenvalue.

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

3) Update $Z$: The variable $Z$ can be updated according to Section IV-C.3 by using the gradient descent algorithm. The gradient with respect to $Z$ can be expressed as

$$
\nabla J(Z) = \frac{4}{(n_T)^2} (K_S)^T \Phi_T^T K_S \mathcal{L}_T + R_1 + \mu(Z - J)
$$

$$
+ \frac{2\tau}{(n_T)^2} (K_S)^T \Phi_T^T K_S \mathcal{L}_T (K_S)^T (K_S)^T (K_T)^T + K_T^T (K_T)^T) \Phi_{K(i,:)}.
$$

VI. CLASSIFICATION

For classification, the projected source data and target data can be represented as $X'_S = \Phi^T \phi(X)$ and $X'_T = \Phi^T \phi(X)$. Then, the existing classifiers (e.g., SVM, least square method [55], and SRC [56]) can be trained on the domain aligned and augmented training data $[X'_S, X'_T]$ with label $Y = [Y_S, Y_T]$ 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 DTS [43], the classifier is trained only on $X'_S$ with label $Y_S$. Finally, classification on those unlabeled target test data, i.e., $X'_{T_u} = \Phi^T \phi(X)' \phi(X'_{T_u})$, is achieved, and the recognition accuracy is reported and compared.

VII. EXPERIMENTS

In this section, the experiments on several benchmark data sets [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 data sets [38]; 2) cross-pose face recognition: Multi-PiE face data set; and 3) cross-domain handwritten digit recognition: USPS, SEMEION, and MNIST data sets. Several related transfer learning methods based on feature transformation and reconstruction, such as GFK [34], SGF [35], LTSL [40], LSDT [41], DTS [43], and SA [60], have been compared and discussed.

A. CROSS-DOMAIN OBJECT RECOGNITION

For cross-domain object/image recognition, five benchmark data sets are used, where several sample images in 4DA office data set are shown in Fig. 4.

1) Results on 4DA Office Data Set (Amazon, DSLR, Webcam1 and Caltech 2562) [34]: Four domains, such as Amazon (A), DSLR (D), Webcam (W), and Caltech (C), are included in the 4DA data set, which contains ten object classes. In our experiment, the configuration is followed in [34], where 20 samples per class are selected from Amazon—eight samples per class from DSLR, Webcam, and Caltech when they are used as source domains—three samples per class are chosen when they are used as target training data, while the rest of the 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 proposed MCTL ranks the second (54%) in average but slightly inferior to LTLSD-LDA (54.9%). The reason may be that the discrimination of LDA helps improve the performance because LTLSD-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.

1http://www.eecs.berkeley.edu/~mfritz/domainadaptation/
2http://www.vision.caltech.edu/Image_Datasets/Caltech256/
TABLE I
RECOGNITION ACCURACY (%) OF DIFFERENT DA IN 4DA SETTING

| 4DA Tasks | Naive Comb | HPA [18] | ARC-1 [9] | MMDT [33] | SGP [38] | GFK [34] | SA [60] | LTSL -PCA [40] | LTSL -LDA [40] | LSDT [41] | MCTL |
|-----------|-----------|----------|-----------|-----------|----------|----------|--------|----------------|----------------|------------|------|
| A → D     | 55.9      | 52.7     | 50.2      | 56.7      | 46.9     | 50.9     | 58.1   | 50.4           | 59.1           | 52.9       | 56.1|
| C → D     | 55.8      | 51.9     | 50.6      | 56.5      | 50.2     | 55.0     | 56.6   | 49.5           | 59.6           | 56.0       | 57.3|
| W → D     | 55.1      | 51.7     | 71.3      | 67.0      | 78.6     | 75.0     | 82.6   | 83.2           | 32.6           | 75.7       | 73.4|
| A → C     | 32.0      | 31.1     | 37.0      | 36.4      | 37.5     | 39.6     | 38.4   | 41.5           | 39.8           | 42.2       | 43.0|
| W → C     | 30.4      | 29.4     | 31.9      | 32.2      | 32.9     | 32.8     | 34.1   | 36.7           | 38.5           | 36.9       | 37.5|
| B → C     | 31.7      | 31.0     | 33.5      | 34.1      | 32.9     | 33.9     | 35.8   | 36.2           | 36.7           | 37.6       | 37.8|
| D → A     | 45.7      | 45.8     | 42.5      | 44.9      | 44.9     | 46.2     | 45.8   | 45.7           | 47.4           | 46.6       | 47.9|
| W → A     | 45.6      | 45.9     | 43.4      | 47.7      | 43.0     | 46.2     | 44.8   | 41.9           | 47.8           | 46.6       | 48.8|
| C → W     | 60.3      | 60.5     | 55.9      | 63.8      | 54.2     | 57.0     | 60.7   | 50.4           | 59.5           | 57.6       | 59.6|
| D → W     | 62.1      | 62.1     | 78.3      | 74.1      | 76.6     | 80.2     | 84.8   | 81.0           | 78.3           | 83.1       | 82.1|
| A → 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 DA OF THE SEVENTH LAYER IN 4DACNN SETTING

| 4DA-CNN Task(s) | SourceOnly | Naive Comb | SGP [35] | TCA | GFK [34] | LTSL [40] | LSCT [41] | MCTL |
|-----------------|------------|------------|----------|-----|----------|-----------|-----------|------|
| A → D           | 81.3       | 94.1       | 92.0     | 82.8| 94.3     | 93.5      | 94.6      | 94.8 |
| C → D           | 77.6       | 92.8       | 97.6     | 94.9| 91.9     | 93.9      | 94.6      | 94.8 |
| W → D           | 96.2       | 98.9       | 97.6     | 99.4| 98.5     | 98.8      | 99.3      | 99.3 |
| A → C           | 79.3       | 83.4       | 77.4     | 81.2| 79.1     | 85.4      | 87.0      | 87.1 |
| W → C           | 68.1       | 81.2       | 76.8     | 75.5| 76.1     | 82.6      | 84.2      | 84.7 |
| D → C           | 74.3       | 82.7       | 78.2     | 79.6| 77.5     | 84.8      | 86.2      | 86.4 |
| D → A           | 81.8       | 90.9       | 88.0     | 90.4| 90.1     | 91.9      | 92.5      | 92.7 |
| W → A           | 73.4       | 90.6       | 86.8     | 85.6| 85.6     | 91.0      | 91.7      | 92.1 |
| C → A           | 86.5       | 90.3       | 89.3     | 92.1| 88.4     | 90.9      | 92.5      | 92.7 |
| C → W           | 67.8       | 90.6       | 87.8     | 88.1| 86.4     | 90.8      | 93.5      | 93.1 |
| D → W           | 95.1       | 98.0       | 95.7     | 96.9| 96.5     | 97.8      | 98.3      | 98.5 |
| A → 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 |

Fig. 6. Some examples from COIL-20 data set.

Fig. 7. Some examples from MSRC and VOC 2007 data sets.

2) Results on 4DA-CNN Data Set (Amazon, DSLR, WebCam, and Caltech 256) [61], [63]: In 4DA-CNN data set, the CNN features are extracted by feeding the raw 4DA data (ten object classes) into the well-trained convolutional neural network (AlexNet with five convolutional layers and three fully connected layers) on ImageNet [63]. The features from the sixth and seventh 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 seventh layer are experimented.

The recognition accuracies by using the seventh 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 RTN [21], DAN [25], DDC [44], and AlexNet [63]. 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 RTN but still better than other three deep transfer learning models. The comparison shows that the proposed
MCTL, as a shallower transfer learning method, has a good competitiveness.

3) Results on COIL-20 Data Set (Columbia Object Image Library [64]): The COIL-20 data set, shown in Fig. 6, contains 20 objects with 1440 gray-scale images (72 multipose 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 data set is divided into two subsets C1 and C2, with each two 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 domains 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%).

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 data set6 with 337 subjects is used. Each subject contains four different sessions with 15 poses, 20 illuminations, and six expressions. The facial images in Sessions 1 and 2 of one person are shown in Fig. 8. In our experiment, we select the first 60 subjects from Sessions 1 and 2. As a result, a smaller session 1 (S1) with seven 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 the domain data. Notably, the raw image pixels are used as features. Specifically, the experimental configurations are set as follows.

SI: 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 five 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 the source data and target training data in the two sessions, respectively. The remaining ten facial images are used as target test data.

With the above-mentioned 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 the other DA methods in handling such pose

| Tasks | SVM | TSL | RDALR [62] | DTSL [43] | LTS [40] | LSDT [41] | MCTL |
|-------|-----|-----|------------|-----------|---------|----------|------|
| C1 → C2 | 82.7 | 80.0 | 80.7 | 84.6 | 75.4 | 81.7 | 84.8 |
| C2 → 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 |

| Tasks | SVM | TSL | RDALR [62] | DTSL [43] | LTS [40] | LSDT [41] | MCTL |
|-------|-----|-----|------------|-----------|---------|----------|------|
| M → V | 37.1 | 32.4 | 37.5 | 38.0 | 38.0 | 47.4 | 47.4 |
| V → M | 56.5 | 43.2 | 62.3 | 56.4 | 61.1 | 63.9 | 64.8 |
| Average | 46.5 | 37.8 | 45.9 | 47.2 | 52.6 | 55.6 | 56.1 |

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

3) Results on COIL-20 Data Set3 (Columbia Object Image Library [64]): The COIL-20 data set, shown in Fig. 6, contains 20 objects with 1440 gray-scale images (72 multipose 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 data set is divided into two subsets C1 and C2, with each two 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 domains 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%).

4) Results on MSRC4 and VOC 20075 Data Sets [43]: The MSRC data set, shown in Fig. 7, contains 4323 images with 18 classes, and the VOC 2007 data set contains 5011 images with 20 concepts. The two data sets share six semantic classes: airplane, bicycle, bird, car, cow, and sheep. We follow [19] to construct a cross-domain image data set MSRC versus VOC (M → 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 data sets: VOC versus MSRC (V → M). All images are uniformly rescaled to 256 pixels, and 128-D dense scale-invariant feature transformation (SIFT) features using the VLFeat open-source package are extracted. Then, K-means clustering is used to obtain a 240-D codebook. In this way, the source and target domains’ data are constructed to share the same label set. The experimental results of different DA methods are shown in Table IV, from which we observe that the performance of our method is 0.5% higher than the state-of-the-art LSDT method and 3.5% higher than the LTS method in average cross-domain recognition performance.

---

3http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
4http://research.microsoft.com/en-us/projects/objectclassrecognition
5http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007
6http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html
TABLE V
RECOGNITION ACCURACY (%) OF DIFFERENT DA METHODS ON FACE RECOGNITION ACROSS POSES

| Tasks      | Naive Comb | A-SVM | SGF [35] | GFK [34] | SA [60] | LTSL [40] | LSCT [41] | MCTL |
|------------|------------|-------|----------|----------|---------|-----------|-----------|------|
| S1 (0° → 60°) | 61.0       | 57.0  | 53.7     | 61.0     | 51.3    | 56.0      | 59.7      | 65.3 |
| S2 (0° → 60°) | 62.7       | 62.7  | 55.0     | 58.7     | 62.7    | 62.7      | 63.3      | 70.0 |
| S1 + S2 (0° → 60°) | 60.2    | 60.1  | 53.8     | 56.3     | 61.7    | 60.2      | 61.7      | 68.3 |
| S1 → 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 DA ON HANDWRITTEN DIGITS RECOGNITION

| Tasks      | Naive Comb | A-SVM | SGF [35] | GFK [34] | SA [60] | LTSL [40] | LSCT [41] | MCTL |
|------------|------------|-------|----------|----------|---------|-----------|-----------|------|
| M → U      | 78.8       | 78.3  | 79.2     | 82.6     | 78.8    | 83.2      | 79.3      | 87.8 |
| S → U      | 83.6       | 76.8  | 77.5     | 82.7     | 82.5    | 83.6      | 84.7      | 84.8 |
| M → S      | 51.9       | 70.3  | 51.6     | 70.5     | 74.4    | 72.8      | 69.1      | 74.0 |
| U → S      | 65.3       | 74.5  | 70.9     | 76.7     | 74.6    | 65.3      | 67.4      | 83.0 |
| U → M      | 71.7       | 73.2  | 71.1     | 74.9     | 72.9    | 71.7      | 70.5      | 81.2 |
| S → 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 |

C. Cross-Domain Handwritten Digits Recognition

Three handwritten digits data sets, including MNIST (M),7 USPS (U),8 and SEMEION (S)9, with ten classes from digit 0 ∼ 9 are used for evaluating the proposed MCTL. The MNIST data set consists of 70000 instances of 28 × 28, the USPS data set consists of 9298 examples of 16 × 16, and the SEMEION data set consists of 2593 images of 16 × 16. The MNIST data set is cropped into 16 × 16. Several images from three data sets are shown in Fig. 9. Each data set is used as source and target domains alternatively, and six cross-domain tasks are explored. In addition, 100 samples per class from source domain and ten samples per class from target domain are randomly selected for training; five random splits are used, and the average classification accuracies are reported in Table VI. From the results, we observe that our MCTL outperforms the 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 the state-of-the-art method, the comprehensive superiority of MCTL in all data sets is clearly demonstrated in Table VII, which shows the mean value of all the cross-domain tasks in the data sets. From the results, we can observe that our MCTL outperforms the state-of-the-art LTSL and LSCT about 2.8% in average performance on all the transfer tasks explored in this paper.

VIII. DISCUSSION

A. Analysis of MCTL-S

When the condition $X_{GT} = X_T$ 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 data set between MCTL and MCTL-S are presented in Table VIII, and the comparisons on COIL-20 data set are shown in Table IX. From the results, we observe that the proposed MCTL and the harsh MCTL-S perform 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 tradeoff coefficients $\tau$ and $\lambda_1$ are fixed as 1 in experiments. Dimensions of common subspace is set as $d = n$. The Gaussian kernel function $k(x_i, x_j) = \exp(- \frac{||x_i - x_j||^2}{2\sigma^2})$ is used, where $\sigma$ can be tuned for different
are two tradeoff parameters. However, 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, MMD loss-based GGDM term, MC loss-based LGDM term, and LRC regularization term, are included. For better interpreting the effect of 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 the validity of the proposed MCTL because local discrepancy metric is deserved for transfer learning.

| COIL-20 | MCTL | MCTL-S |
|---------|------|--------|
| C_1 \(\rightarrow\) C_2 | 84.83 | 85.00 |
| C_2 \(\rightarrow\) C_1 | 83.67 | 83.67 |
| Average | 84.25 | 84.34 |

### TABLE IX

| 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 | 87.06 |
| W \(\rightarrow\) C | 84.73 | 84.74 |
| D \(\rightarrow\) C | 86.37 | 86.34 |
| D \(\rightarrow\) A | 92.66 | 92.66 |
| 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 VIII

| 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 | 87.06 |
| W \(\rightarrow\) C | 84.73 | 84.74 |
| D \(\rightarrow\) C | 86.37 | 86.34 |
| D \(\rightarrow\) A | 92.66 | 92.66 |
| 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 |

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 \(\lambda_1\) by fixing \(\tau = 1\). Fig. 10(d) shows the parameter analysis of \(\tau\) by fixing \(\lambda_1 = 1\). For tuning the two parameters simultaneously, we have also provided the 3-D surface on the COIL-20 data set in Fig. 12(a) (C1 \(\rightarrow\) C2) and (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 Algorithm 1 is presented. Algorithm 1 includes three basic steps: update \(\mathcal{Z}\), update \(\mathcal{J}\), and update \(\Phi\). The computation of \(\mathcal{J}\) involves eigendecomposition and matrix multiplication, and the complexity is \(O(n^3)\). The computation of updating \(\mathcal{J}\) and \(\mathcal{Z}\) 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.

Furthermore, Table XI shows the computational time comparisons on the CMU Multi-PIE data (S1 \(\rightarrow\) S2) and the 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, 3.20 GHz, and 16-GB RAM.

### E. Model Visualization and Convergence

In this section, the visualization and convergence will be discussed. Pose alignment is a difficult task. Therefore, for
TABLE XI

| Tasks | SGF [35] | GPK [34] | SA [60] | LTSL [40] | MCTL |
|-------|----------|----------|---------|------------|------|
| S1 → S2 | 10.9s (92.5%) | 1.5s (96.7%) | 4.18s (98.3%) | 7.21s (97.2%) | 7.62s (97.3%) |
| M → U | 7.5s (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.

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 inherit similar distribution property of the target data.

Furthermore, COIL-20 and handwritten digits data sets are also exploited. The second row of Fig. 11 shows the pose transfer process, and the generative data show 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 the target domain is shown in the third row of Fig. 11. The effect of domain generation is clearly shown.

In addition, 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., $F_{\text{min}}$) 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 ($C_1 \rightarrow C_2$) and 4DACNN ($A \rightarrow D$), respectively. In addition, the convergence of each term in the MCTL, such as $F_{\text{MC}}$ (i.e., MC-based LGDM loss), $F_{\text{MMD}}$ (i.e., GGDM loss), and $F_{\text{Z}}$ (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.

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 DA. However, an open problem, which MMD neglects the locality geometric structure of domain data, is preserved. In order to overcome the bottleneck, motivated by MC, MCTL is proposed, which aims at generating a new intermediate domain sharing similar distribution with the true target domain. The MC implies that the DA 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 an MC-based LGDM loss, GGDM loss, and LRC regularization jointly constructed, MCTL is established. Extensive experiments on benchmark DA data sets 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, Dec. 2015.

[2] G. Csurka. (2017). “Domain adaptation for visual applications: A comprehensive survey.” [Online]. Available: https://arxiv.org/abs/1702.05374

[3] J. Hoffman, B. Kulis, T. Darrell, and K. Saenko, “Discovering latent domains for multisource domain adaptation,” in Proc. ECCV. Singapore: Springer, 2012, pp. 702–715.

[4] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[5] L. Duan, I. W. Tsang, and D. Xu, “Domain transfer multiple kernel learning,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 3, pp. 465–479, Mar. 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. Pattern Anal. Mach. Intell., vol. 40, no. 5, pp. 1114–1127, May 2018.
[7] R. Gopalan, R. Li, and R. Chellappa, “Unsupervised adaptation across domain shifts by generating intermediate data representations,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 11, pp. 2288–2302, Nov. 2014.
[8] L. Duan, D. Xu, I. W. Tsang, and J. Luo, “Visual event recognition in videos by learning from Web data,” in Proc. CVPR, Jun. 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 Proc. CVPR, Jun. 2011, pp. 1785–1792.
[10] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, “Adapting visual category models to new domains,” in Proc. Eur. Conf. Comput. Vis., 2010, pp. 213–226.
[11] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain adaptation via transfer component analysis,” IEEE Trans. Neural Netw., vol. 22, no. 2, pp. 199–210, Feb. 2011.
[12] L. Duan, I. W. Tsang, D. Xu, and S. J. Maybank, “Domain transfer SVM for video concept detection,” in Proc. CVPR, Jun. 2009, pp. 1375–1381.
[13] M. Baktashmotlagh, M. T. Harandi, B. C. Lovell, and M. Salzmann, “Unsupervised domain adaptation by domain invariant projection,” in Proc. IJCV, Dec. 2013, pp. 767–796.
[14] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, “Analysis of representations for domain adaptation,” in Proc. NIPS, 2007, pp. 137–144.
[15] L. Duan, D. Xu, and I. W. Tsang. (2012). “Learning transferable models to new domains,” in Proc. CVPR, Jun. 2010, pp. 213–226.
[16] J. Hoffman, E. Rodner, J. Donahue, B. Kulis, and K. Saenko, “Asymmetry and category invariant feature transformations for domain adaptation,” Int. J. Comput. Vis., vol. 109, nos. 1–2, pp. 28–41, Aug. 2014.
[17] B. Gong, Y. Shi, F. Sha, and K. Grauman, “Geodesic flow kernel for unsupervised domain adaptation,” in Proc. CVPR, Jun. 2012, pp. 2066–2073.
[18] R. Gopalan, R. Li, and R. Chellappa, “Domain adaptation for object recognition: An unsupervised approach,” in Proc. ICCV, Nov. 2011, pp. 999–1006.
[19] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, “A kernel two-sample test,” J. Mach. Learn. Res., vol. 13, pp. 723–773, Mar. 2012.
[20] A. Iyer, S. Nath, and S. Sarawagi, “Maximum mean discrepancy for class ratio estimation: Convergence bounds and kernel selection,” in Proc. ICML, Jun. 2014, pp. 530–538.
[21] M. Long, G. Ding, J. Wang, J. Sun, and P. S. Yu, “Transfer sparse coding for robust image representation,” in Proc. ICCV, Jun. 2013, pp. 407–414.
[22] I.-H. Juo, D. Liu, D. Lee, and S.-F. Chang, “Robust visual domain adaptation with low-rank reconstruction,” in Proc. CVPR, Jun. 2012, pp. 2168–2175.
[23] M. Shao, D. Kit, and Y. Fu, “Generalized transfer subspace learning through low-rank constraint,” Int. J. Comput. Vis., vol. 109, nos. 1–2, pp. 74–93, 2014.
[24] L. Zhang, W. Zuo, and D. Zhang, “LSDT: Latent sparse domain transfer learning for visual adaptation,” IEEE Trans. Image Process., vol. 25, no. 3, pp. 1177–1191, Mar. 2016.
[25] L. Zhang, S. K. Jha, T. Liu, and G. Pei, “Discriminative Kernel transfer learning via l2,1-norm minimization,” in Proc. ICNN, Jul. 2016, pp. 2220–2227.
[26] Y. Xu, X. Fang, J. Wu, X. Li, and D. Zhang, “Discriminative transfer subspace learning via low-rank and sparse representation,” IEEE Trans. Image Process., vol. 25, no. 2, pp. 850–863, Feb. 2016.
[27] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, “Simultaneous deep transfer across domains and tasks,” in Proc. ICCV, Dec. 2015, pp. 4068–4076.
[28] X. Glorot, A. Bordes, and Y. Bengio, “Domain adaptation for large-scale sentiment classification: A deep learning approach,” in Proc. ICML, Jun. 2011, pp. 513–520.
[29] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Learning and transferring mid-level image representations using convolutional neural networks,” in Proc. CVPR, Jun. 2014, pp. 1717–1724.
[30] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon. (2015). “Transfer learning from deep features for remote sensing and poverty mapping.” [Online]. Available: https://arxiv.org/abs/1510.00098
[31] J. Donahue et al., “DeCAF: A deep convolutional activation feature for generic visual recognition,” in Proc. ICML, Jun. 2014, pp. 647–655.
[32] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN features off-the-shelf: An astounding baseline for recognition,” in Proc. CVPR, Jun. 2014, pp. 512–519.
[33] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial discriminative domain adaptation,” in Proc. CVPR, Jul. 2017, pp. 7172–7176.
[34] I. Goodfellow et al., “Generative adversarial nets,” in Proc. NIPS, 2014, pp. 2672–2680.
[35] Z. Lin, R. Liu, and Z. Su, “Linearized alternating direction method with adaptive penalty for low-rank representation,” in Proc. NIPS, 2011, pp. 612–620.
[36] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” SIAM J. Optim., vol. 20, no. 4, pp. 1956–1982, Mar. 2010.
[37] L. Rosasco, A. Verri, M. Santoro, S. Mosci, and S. Villa, “Iterative projection methods for structured sparsity regularization,” Massachusetts Inst. Technol., Cambridge, U.K., Tech. Rep. TR-2009-050, 2009.
[38] T. Kanamori, S. Hido, and M. Sugiyama, “A least-squares approach to direct importance estimation,” J. Mach. Learn. Res., vol. 10, pp. 1391–1445, Jul. 2009.
[39] I. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 10, pp. 210–227, Feb. 2009.
[40] A. Gaidon, G. Zen, and J. A. Rodriguez-Serrano. (2014). “Self-learning camera: Autonomous adaptation of object detectors to unlabeled video streams.” [Online]. Available: https://arxiv.org/abs/1406.4296
[41] B. Gong, K. Grauman, and F. Sha, “Learning kernels for unsupervised domain adaptation with applications to visual object recognition,” Int. J. Comput. Vis., vol. 109, nos. 1–2, pp. 3–27, Aug. 2014.
Lei Zhang (M’14–SM’18) received the 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. He was 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 TNNSLS, IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON MULTIMEDIA (TMM), IEEE TRANSACTIONS ON CYBERNETICS (TCYB), IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT (TIM), and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS (TSMCA). His current research interests include machine learning, pattern recognition, computer vision, and intelligent systems.

Dr. Zhang was a recipient of the Outstanding Reviewer Award for more than ten journals, the New Academic Researcher Award for Doctoral Candidates from the Ministry of Education, China, in 2012, the Academy Award for Youth Innovation of Chongqing University in 2013, the Hong Kong Scholar Award in 2014, and the Outstanding Doctoral Dissertation Award of Chongqing, China, in 2015.

Shanshan Wang received the B.E. and M.E. degrees from Chongqing University, Chongqing, China, in 2010 and 2013, respectively, where she is currently pursuing the Ph.D. degree. Her current research interests include machine learning, pattern recognition, and computer vision.

Guang-Bin 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 the 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 the Wireless Communication Department, Northeastern University, Shenyang, China. From 1998 to 2001, he was a Research Fellow with the Singapore Institute of Manufacturing Technology, Singapore, where he has led/implemented several key industrial projects, including the Chief Designer and the Technical Leader of the Singapore Changi Airport Cargo Terminal Upgrading Project. Since 2001, he has been an Assistant Professor and an Associate Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University. His current research interests include machine learning, computational intelligence, and extreme learning machines.

Dr. Huang received the Best Paper Award of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS in 2013. He serves as an Associate Editor for Neurocomputing, Neural Networks, Cognitive Computation, and the IEEE TRANSACTIONS ON CYBERNETICS.

Wangmeng Zuo (SM’14) received the Ph.D. degree in computer application technology from the Harbin Institute of Technology, Harbin, China, in 2007. From 2004 to 2008, he was a Research Assistant with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. From 2009 to 2010, he was a Visiting Professor with Microsoft Research Asia, Beijing, China. He is currently an Associate Professor with the School of Computer Science and Technology, Harbin Institute of Technology. He 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 3-D vision.

Dr. Zuo is an Associate Editor of the IET Biometrics.

Jian Yang received the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology, (NUST), Nanjing, China, in 2002. In 2003, he was a Post-Doctoral Researcher with the University of Zaragoza, Zaragoza, Spain. From 2004 to 2006, he was a Post-Doctoral Fellow with the Biometrics Centre, The Hong Kong Polytechnic University, Hong Kong. From 2006 to 2007, he was a Post-Doctoral Fellow with the Department of Computer Science, New Jersey Institute of Technology, Newark, NY, USA. He is currently a Chang-Jiang Professor with the School of Computer Science and Technology, 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 current research interests include pattern recognition, computer vision, and machine learning.

Dr. Yang is a fellow of International Association for Pattern Recognition (IAPR). He has been an Associate Editor of the Pattern Recognition Letters, IEEE TRANSACTIONS NEURAL NETWORKS AND LEARNING SYSTEMS, and Neurocomputing.

David Zhang (F’09) received the bachelor’s degree in computer science from Peking University, Beijing, China, in 1974, the M.Sc. and Ph.D. degrees in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 1994.

From 1986 to 1988, he was a Post-Doctoral Fellow with Tsinghua University, Beijing, and then an Associate Professor with the Academia Sinica, Beijing. Since 2005, he has been a Chair Professor with The Hong Kong Polytechnic University, Hong Kong, where he is currently the Founding Director of the Biometrics Research Centre, supported by the Hong Kong SAR Government in 1998. He is also a Visiting Chair Professor with Tsinghua University and an Adjunct Professor with Peking University, Shanghai Jiao Tong University, Shanghai, China, HIT, and the University of Waterloo. He is an Editor of the book International Series on Biometrics (Springer, KISB). He has authored more than ten books and over 300 international journal papers. He holds 30 patents from USA/Japan/Hong Kong/China.

Dr. Zhang is a Croucher Senior Research Fellow. He is a fellow International Association for Pattern Recognition (IAPR). He is the Founder and the Editor-in-Chief of the International Journal of Image and Graphics. He is also a Distinguished Speaker of the IEEE Computer Society. He is also an Associate Editor of more than ten international journals, including IEEE TRANSACTIONS and so on. He is an Organizer of the International Conference on Biometrics Authentication.