An Exemplar-Based Multi-View Domain Generalization Framework for Visual Recognition

Li Niu, Wen Li, Dong Xu, *Senior Member, IEEE*, and Jianfei Cai, *Senior Member, IEEE*

***Abstract*— In this paper, we propose a new exemplar-based multi-view domain generalization (EMVDG) framework for visual recognition by learning robust classifier that are able to generalize well to arbitrary target domain based on the training samples with multiple types of features (i.e., multi-view features). In this framework, we aim to address two issues simultaneously. First, the distribution of training samples (i.e., the source domain) is often considerably different from that of testing samples (i.e., the target domain), so the performance of the classifiers learnt on the source domain may drop significantly on the target domain. Moreover, the testing data are often unseen during the training procedure. Second, when the training data are associated with multi-view features, the recognition performance can be further improved by exploiting the relation among multiple types of features. To address the first issue, considering that it has been shown that fusing multiple SVM classifiers can enhance the domain generalization ability, we build our EMVDG framework upon exemplar SVMs (ESVMs), in which a set of ESVM classifiers are learnt with each one trained based on one positive training sample and all the negative training samples. When the source domain contains multiple latent domains, the learnt ESVM classifiers are expected to be grouped into multiple clusters. To address the second issue, we propose two approaches under the EMVDG framework based on the consensus principle and the complementary principle, respectively. Specifically, we propose an EMVDG_CO method by adding a co-regularizer to enforce the cluster structures of ESVM classifiers on different views to be consistent based on the consensus principle. Inspired by multiple kernel learning, we also propose another EMVDG_MK method by fusing the ESVM classifiers from different views based on the complementary principle. In addition, we further extend our EMVDG framework to exemplar-based multi-view domain adaptation (EMVDA) framework when the unlabeled target domain data are available during the training procedure. The effectiveness of our EMVDG and EMVDA frameworks for visual recognition is clearly demonstrated by comprehensive experiments on three benchmark data sets.**

***Index Terms*— Domain generalization, domain adaptation, latent domain discovery, multi-view learning.**

I. Introduction

In the field of visual recognition, the data distributions of the training data and the testing data are usually quite different, in which the training set (*resp.*, the testing set) is referred to as the source domain (*resp.*, the target domain). Recently, abundant domain adaptation approaches [1]–[14] were proposed to reduce the data distribution mismatch between the source domain and the target domain explicitly. Nevertheless, the target domain samples are often unavailable during the training procedure and this problem is named domain generalization [15]. In comparison with domain adaptation, domain generalization aims to learn robust classifiers that can generalize well to arbitrary target domain. More recently, several domain generalization approaches [15]–[18] were also developed to enhance the generalization capability of the classifiers learnt on the source domain. For more details about domain generalization and adaptation, please refer to Section II.

Most of the existing approaches for domain generalization or domain adaptation only utilize one type of feature in the training and the testing stage. In fact, when the training and testing data are associated with multiple types of features, the recognition performance can be enhanced by exploiting the relation among multiple types of features (see Section II for the details). Some recently proposed domain adaptation approaches [7], [19]–[21] are based on multiple types of features, which aim to tackle with the data distribution mismatch and simultaneously exploit the relation among multiple types of features. Blitzer *et al.* [19] use canonical correlation analysis (CCA) to learn the projection matrices, based on which the classifiers learnt on the source domain are adapted to the target domain. In [20], different weights are assigned to the training samples based on the maximum mean discrepancy (MMD), while the prediction scores obtained on multiple views are expected to be consistent. Yang and Gao [21] incorporate an MMD-based regularizer into the CCA framework. The approach in [7] can be used to learn the kernel weights to cope

Manuscript received June 14, 2016; revised September 13, 2016; accepted September 26, 2016. This work was supported in part by the Rapid-Rich Object Search Laboratory within the Nanyang Technological University, Singapore, through the National Research Foundation, Singapore, under its Interactive Digital Media Strategic Research Programme, in part by the Singapore MoE Tier 2 under Grant ARC42/13, and in part by the Faculty of Engineering and Information Technologies, The University of Sydney, under the Faculty Research Cluster Program.

L. Niu is with the Interdisciplinary Graduate School, Nanyang Technological University, Singapore 639798 (e-mail: lniu002@ntu.edu.sg).

W. Li is with the Computer Vision Laboratory, ETH Zürich, 8092 Zürich, Switzerland (e-mail: liwen@vision.ee.ethz.ch).

D. Xu is with the School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: dong.xu@sydney.edu.au).

J. Cai is with School of Computer Science & Engineering, Nanyang Technological University, Singapore 639798 (e-mail: asjfcai@ntu.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2016.2615469

with the domain distribution mismatch by treating each view as a kernel. However, the above multi-view approaches [7], [19]–[21] require the target domain samples in the training stage, which are not available in the domain generalization scenario.

To this end, we propose an exemplar-based multi-view domain generalization (EMVDG) framework by utilizing multi-view source domain data to learn robust classifiers, which are able to generalize well to the arbitrary target domain. On the one hand, our approach is inspired by the recent work [16], which demonstrates that fusing multiple SVM classifiers can enhance the domain generalization capability. In particular, our EMVDG framework builds upon ESVMs [22] with each SVM classifier learnt based on one positive training sample together with all the negative training samples. According to the assumptions in [16], [23], and [24], the source domain may contain multiple hidden latent domains. Thus, the ESVM classifiers, which correspond to the positive samples belonging to the same hidden latent domain, are expected to be similar. Therefore, the ESVM classifiers can be grouped into multiple clusters, which can be achieved by using low-rank techniques [e.g., nuclear norm-based regularizer or low-rank representation (LRR)].

On the other hand, in order to take full advantage of multi-view features, we propose two methods under the EMVDG framework based on the consensus principle and the complementary principle [25], respectively. Without loss of generality, the consensus principle expects the information of multiple views to be consistent while the complementary principle assumes that each view may contain some information, which are missing in the other views, so that multiple views can be jointly used to make the data representation more comprehensive. In this paper, for the consensus principle, we enforce the consistency of inherent cluster structures on different views by adding a co-regularizer, which uses LRR [26] based on the weight vectors of ESVM classifiers. This method is named EMVDG_CO. For the complementary principle, we linearly combine multiple kernels on different views as in multiple kernel learning (MKL) [27], and simultaneously enforce the dual matrix, which consists of the dual vectors of ESVM classifiers, to be low rank by adding a nuclear norm-based regularizer. We refer to this approach as EMVDG_MK. For both methods, alternating optimization algorithms are developed to solve the nontrivial optimization problems.

Our major contributions of this paper can be summarized as follows. First, we propose an effective EMVDG framework including two methods EMVDG_CO and EMVDG_MK. To the best of our knowledge, this is the first work to explore the domain generalization problem in the multi-view scenario. Second, we further extend our EMVDG framework to exemplar-based multi-view domain adaptation (EMVDA) for domain adaptation, which can utilize the unlabeled target domain data.

This paper differs from our preliminary conference version [28] in the following aspects. First, in [28], we only discussed the EMVDG_CO method and its corresponding domain adaptation method EMVDA_CO based on the consensus principle, which are referred to as MVDG and MVDA in [28], respectively. In this paper, we additionally explore the complementary principle and propose a new EMVDG_MK method, which leads to a more general EMVDG framework including both EMVDG_CO and EMVDG_MK methods. Second, we also extend our newly proposed EMVDG_MK method to EMVDA_MK for domain adaptation in an EMVDA framework. Third, moreover, we employ new encoding method (i.e., Fisher vectors) instead of using bag-of-words (BOWs) based on improved dense trajectory (IDT) descriptors to improve the action recognition performance on the ACT4^2 and online RGBD action data set (ORGBD) data sets.

II. Related Work

This paper is related to the domain generalization methods [15], [16]. In [15], the marginal distribution mismatch between different latent domains is reduced while the conditional distribution on each view is maintained. But the approach proposed in [15] requires domain labels, which may not be available in real world applications. Among the existing domain generalization methods, our work is more related to [16], which builds upon ESVMs [22] to explore the low-rank structure in positive source domain samples. However, the above approaches [15], [16] only focus on one type of feature, while our work focuses on domain generalization in the multi-view scenario.

This paper is also related to the latent domain discovering methods [23], [24]. In [24], the training samples on the source domain are clustered into different hidden latent domains. In [23], the distribution mismatch between each pair of different latent domains is maximized. After the latent domains are discovered, the classifiers learnt based on each hidden latent domain are fused, and then the integrated classifier is applied to the testing data. However, the above methods require the number of hidden latent domains and these methods do not discuss how to employ multiple types of features effectively.

In this paper, our EMVDG framework is also extended to EMVDA for domain adaptation. Therefore, we discuss some existing domain adaptation approaches here. In general, the domain adaptation approaches can be classified into classifier-based approaches [5], [7], feature-based approaches [1]–[4], and instance-reweighting methods [6]. Interested readers can refer to [29] for more details. However, the above works do not discuss how to utilize the source domain samples with multi-view features. As mentioned in Section I, some domain adaptation approaches [7], [19]–[21] can be used in the multi-view scenario. However, their methods require the target domain samples in the training stage, which are not available for domain generalization.

Finally, this paper is related to the multi-view learning approaches [27], [30]–[33]. In general, the existing multi-view learning methods mainly rely on either the consensus principle or the complementary principle [25]. For the consensus principle, the approach in [30] first uses the kernel canonical correlation analysis (KCCA) to transform the training and testing features, and then learns SVM classifiers by using the transformed features, while the work in [31] formulates this two-stage approach as one unified optimization problem. Ding and Fu [32] proposed to learn a common low-rank subspace among multiple types of features. For the complementary principle, the linear combination of multiple kernels on different types of features is used to improve the performance in the MKL methods [27], [34]. In addition, some multi-view semi-supervised learning approaches [35], [36] have also been proposed. For manifold-based approaches, a semi-supervised Laplacian regularizer is incorporated into KCCA in [37], while the average matrix of multiple Laplacian matrices based on multi-view features is used in a semi-supervised learning method in [36]. In co-training, Blum and Mitchell [35] select the confident unlabeled training samples by utilizing the classifier learnt on one view and add these confident samples to the labeled training set for learning the classifier on the other view in an iterative way. For more details about multi-view learning, please refer to the recent survey [25]. However, all the above multi-view learning methods assume that the data distribution of the training data and the testing data are the same, while our frameworks do not have this assumption.

III. Exemplar-Based Multi-View Domain Generalization

In this section, an EMVDG framework is proposed. In the following, we first introduce domain generalization with ESVMs briefly in Section III-A, and then introduce our two methods under the EMVDG framework: EMVDG_CO in Section III-B and EMVDG_MK in Section III-C.

For better presentation, we denote a matrix/vector by using an uppercase/lowercase letter in boldface. $\mathbf{1}_n, \mathbf{0}_n \in \mathbb{R}^n$ are used to denote the n-dim column vectors with the entries of all ones and all zeros, respectively. Moreover, we use $\mathbf{1}$ and $\mathbf{0}$ to replace $\mathbf{1}_n$ and $\mathbf{0}_n$ when the dimensionality is obvious. Similarly, we use $\mathbf{I}$ and $\mathbf{O}$ to denote the identity matrix and the matrix of all zeros, respectively. We use the superscript $'$ to denote the transpose of a vector/matrix and $\mathbf{A}^{-1}$ to denote the inverse matrix of $\mathbf{A}$. Moreover, we represent the elementwise product between two matrices by using $\mathbf{A} \circ \mathbf{B}$. The inequality $\mathbf{a} \leq \mathbf{b}$ is used to denote that $a_i \leq b_i$ for $i = 1, \ldots, n$.

In this paper, we explore the multi-view domain generalization problem in the binary classification scenario. Suppose the source domain contains n positive training samples and m negative training samples, in which each sample is associated with V types of features, each positive training sample can be denoted as $\mathbf{x}_i^+ = (\mathbf{x}_i^{1+}, \ldots, \mathbf{x}_i^{V+})$, $i = 1, \ldots, n$, and each negative training sample can be denoted as $\mathbf{x}_j^- = (\mathbf{x}_j^{1-}, \ldots, \mathbf{x}_j^{V-})$, $j = 1, \ldots, m$.

A. Domain Generalization With Exemplar SVMs

Domain generalization targets at learning robust classifiers that are able to generalize well to the arbitrary target domain by utilizing the source domain samples, which can be achieved by fusing multiple SVM classifiers as discussed in Section II. Specifically, when the source domain data are assumed to be sampled from multiple latent domains, the latent domain labels are given (i.e., in [15]) or obtained by using latent domain discovering methods [23], [24]. Then, the classifiers learnt based on each latent domain are integrated to predict the target domain data. Since the training samples within each latent domain are with more coherent data distribution, each classifier corresponding to each latent domain should be more discriminative and the integrated classifier should be more robust to the various data distribution of the unseen target domain.

However, in the real-world scenario, the variance of training samples is likely to be affected by complicated hidden factors that often overlap and interact with each other. Considering that it is a very challenging task to explicitly discover the hidden latent domains, low-rank ESVM (LRESVM) was proposed in [16], which utilizes the low-rank structure of positive source domain data. It is worth noting that this approach builds upon the ESVMs [22] with each SVM classifier learnt based on one positive source domain sample and all negative source domain samples. ESVM targets at capturing the specific feature of individual positive training sample, which has been widely used in many computer vision tasks such as object detection [22], image retrieval [38], and feature encoding [39]. By using $f_i(\mathbf{x}) = \mathbf{w}_i'\mathbf{x}$ to denote the ESVM classifier learnt based on the ith positive sample $\mathbf{x}_i^+$ and all the negative samples[1] $\{\mathbf{x}_j^-|_{j=1}^m\}$ (we only focus on single-view learning in this section, so the superscript v is omitted for ease of presentation), the formulation for learning n ESVMs can be written as

$$\begin{aligned}\min_{\mathbf{w}_i,\xi_i,\epsilon_{ij}} \quad & \frac{1}{2}\sum_{i=1}^{n} \|\mathbf{w}_i\|^2 + C\sum_{i=1}^{n}\xi_i + C\sum_{i=1}^{n}\sum_{j=1}^{m}\epsilon_{ij} \\ \text{s.t.} \quad & \mathbf{w}_i{}'\mathbf{x}_i^+ \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \\ & \mathbf{w}_i{}'\mathbf{x}_j^- \leq -1 + \epsilon_{ij}, \quad \epsilon_{ij} \geq 0, \quad \forall i, \forall j \end{aligned} \tag{1}$$

where C is a tradeoff parameter, ξ_i's and ϵ_{ij}'s are the slack variables, and $\|\mathbf{w}_i\|^2$ is the regularizer to control the complexity of $\mathbf{w}_i$.

As the positive samples belonging to the same hidden latent domain should be similar, the work in [16] enforces the prediction score matrix $\bar{\mathbf{G}} \in \mathcal{R}^{n\times n}$, in which $\bar{G}_{ij}$ is the prediction score by using the jth ESVM classifier on the ith positive training sample, to be low rank by employing a nuclear norm-based regularizer. However, this approach only considers the training data with one type of feature. When the training data are associated with multiple types of features, we demonstrate that it is useful to exploit the relation among multiple types of features based on the consensus principle in Section III-B or the complementary principle in Section III-C.

B. Exemplar-Based Multi-View Domain Generalization With Co-Regularizer

In this section, we propose our EMVDG_CO method by taking advantage of multi-view features based on the consensus principle, in which an ESVM classifier is learnt for each positive sample on each view. Specifically, we use $f_i^v(\mathbf{x}^v) = \mathbf{w}_i^{v\prime}\mathbf{x}^v$ to denote the ESVM classifier learnt based on $\mathbf{x}_i^{v+}$ and $\{\mathbf{x}_j^{v-}|_{j=1}^m\}$ on the vth view. We also use $\mathbf{W}^v = [\mathbf{w}_1^v, \ldots, \mathbf{w}_n^v]$ to

[1] We do not employ the bias term explicitly. Instead, we augment each feature vector with an extra element of 1.

denote the weight matrix consisting of all the ESVM classifiers learnt on the vth view.

1) Formulation: Since the positive samples belonging to the same hidden latent domain should be similar, their corresponding ESVM classifiers should also be similar to each other, and thus the weight vectors $\mathbf{w}_i^v$'s on each view can be grouped into multiple clusters. In this paper, such a cluster structure is exploited by utilizing the LRR [26] technique. According to LRR [26], the weight matrix on each view can be reconstructed by using itself as a dictionary, i.e., $\mathbf{W}^v = \mathbf{W}^v\mathbf{Z}^v + \mathbf{E}^v$, in which $\mathbf{Z}^v \in \mathbb{R}^{n\times n}$ is the representation matrix and $\mathbf{E}^v$ is the reconstruction error. Note that the representation matrix $\mathbf{Z}^v$ encodes the cluster structure of ESVM classifiers [26], in which the between-cluster (*resp.*, within-cluster) entries of $\mathbf{Z}^v$ are generally sparse (*resp.*, dense).

On the one hand, in LRR, the representation matrices $\mathbf{Z}^v$'s are expected to be low rank. Moreover, by jointly learning $\mathbf{W}^v$ and low-rank matrix $\mathbf{Z}^v$ using $\mathbf{W}^v = \mathbf{W}^v\mathbf{Z}^v + \mathbf{E}^v$, $\mathbf{W}^v$ is also expected to be low rank when the error term $\mathbf{E}^v$ is close to zero. In such a case, the weight vectors $\mathbf{w}_i^v$'s corresponding to the positive training samples belonging to the same hidden latent domain are expected to be similar, which is consistent with our motivation.

On the other hand, when the training data are associated with multiple types of features, the cluster structures of $\mathbf{W}^v$'s on different views are expected to be consistent according to the consensus principle. Based on our LRR, in which the cluster structure of $\mathbf{W}^v$ is encoded in $\mathbf{Z}^v$, such consistency can be easily introduced by enforcing $\mathbf{Z}^v$'s on multiple views to be close to each other based on our new co-regularizer $\sum_{v,\tilde{v}:v\neq\tilde{v}}\|\mathbf{Z}^v - \mathbf{Z}^{\tilde{v}}\|_F^2$.

To this end, we formulate our EMVDG_CO method as

$$\min_{\substack{\mathbf{Z}^v,\mathbf{W}^v,\mathbf{E}^v\\ \xi_i^v,\epsilon_{ij}^v}} \sum_{v=1}^{V}\left(\frac{1}{2}\|\mathbf{W}^v\|_F^2 + C\sum_{i=1}^{n}\xi_i^v + C\sum_{i=1}^{n}\sum_{j=1}^{m}\epsilon_{ij}^v\right) + \sum_{v=1}^{V}\left(\lambda_2\|\mathbf{E}^v\|_F^2 + \lambda_3\|\mathbf{Z}^v\|_*\right) + \frac{\gamma}{2}\sum_{v,\tilde{v}:v\neq\tilde{v}}\|\mathbf{Z}^v - \mathbf{Z}^{\tilde{v}}\|_F^2 \qquad (2)$$

$$\text{s.t. } \mathbf{w}_i^{v\prime}\mathbf{x}_i^{v+} \geq 1 - \xi_i^v, \quad \xi_i^v \geq 0, \quad \forall v, \forall i \qquad (3)$$

$$\mathbf{w}_i^{v\prime}\mathbf{x}_j^{v-} \leq -1 + \epsilon_{ij}^v, \quad \epsilon_{ij}^v \geq 0, \quad \forall v, \forall i, \forall j \qquad (4)$$

$$\mathbf{W}^v = \mathbf{W}^v\mathbf{Z}^v + \mathbf{E}^v, \quad \forall v \qquad (5)$$

where ξ_i^v, ϵ_{ij}^v are the slack variables, $\|\mathbf{W}^v\|_F^2$ is the regularizer for controlling the complexity of ESVM classifiers, and C, λ_2, λ_3, and γ are the tradeoff parameters. The nuclear norm-based regularizer $\|\mathbf{Z}^v\|_*$ is used to enforce $\mathbf{Z}^v$ to be low rank, and the regularizer $\|\mathbf{E}^v\|_F^2$ is employed to enforce the reconstruction error $\mathbf{E}^v$ to approach zeros.

2) Optimization: For better optimizing the problem in (2), an intermediate variable $\mathbf{G}^v$ is introduced for each $\mathbf{W}^v$. Instead of employing LRR on $\mathbf{W}^v$ as in (2), we employ LRR on $\mathbf{G}^v$ while enforcing $\mathbf{G}^v$ to be close to $\mathbf{W}^v$ by adding the regularizer $\|\mathbf{W}^v - \mathbf{G}^v\|_F^2$. In particular, we reach the following formulation:

$$\min_{\substack{\mathbf{Z}^v,\mathbf{W}^v,\mathbf{G}^v\\ \mathbf{E}^v,\xi_i^v,\epsilon_{ij}^v}} \sum_{v=1}^{V}\left(\frac{1}{2}\|\mathbf{W}^v\|_F^2 + C\sum_{i=1}^{n}\xi_i^v + C\sum_{i=1}^{n}\sum_{j=1}^{m}\epsilon_{ij}^v\right) + \sum_{v=1}^{V}\left(\lambda_1\|\mathbf{W}^v - \mathbf{G}^v\|_F^2 + \lambda_2\|\mathbf{E}^v\|_F^2 + \lambda_3\|\mathbf{Z}^v\|_*\right) + \frac{\gamma}{2}\sum_{v,\tilde{v}:v\neq\tilde{v}}\|\mathbf{Z}^v - \mathbf{Z}^{\tilde{v}}\|_F^2 \qquad (6)$$

$$\text{s.t. } \mathbf{w}_i^{v\prime}\mathbf{x}_i^{v+} \geq 1 - \xi_i^v, \quad \xi_i^v \geq 0, \quad \forall v, \forall i \qquad (7)$$

$$\mathbf{w}_i^{v\prime}\mathbf{x}_j^{v-} \leq -1 + \epsilon_{ij}^v, \quad \epsilon_{ij}^v \geq 0, \quad \forall v, \forall i, \forall j \qquad (8)$$

$$\mathbf{G}^v = \mathbf{G}^v\mathbf{Z}^v + \mathbf{E}^v, \quad \forall v \qquad (9)$$

in which λ_1 is a tradeoff parameter. It is obvious that the problem in (6) can reduce to the problem in (2) when λ_1 approaches $+\infty$. The problem in (6) can be solved by an alternative approach, in which two sets of variables $\{\mathbf{Z}^v, \mathbf{E}^v\}$ and $\{\mathbf{W}^v, \mathbf{G}^v, \xi_i^v, \epsilon_{ij}^v\}$ are updated alternatively until the objective value of (6) converges.

a) Update $\mathbf{Z}^v$ and $\mathbf{E}^v$: When $\mathbf{W}^v$, $\mathbf{G}^v$, ξ_i^v, and ϵ_{ij}^v are fixed, the problem in (6) becomes the following problem:

$$\min_{\mathbf{Z}^v,\mathbf{E}^v} \sum_{v=1}^{V}\left(\lambda_2\|\mathbf{E}^v\|_F^2 + \lambda_3\|\mathbf{Z}^v\|_*\right) + \frac{\gamma}{2}\sum_{v,\tilde{v}:v\neq\tilde{v}}\|\mathbf{Z}^v - \mathbf{Z}^{\tilde{v}}\|_F^2 \qquad (10)$$

$$\text{s.t. } \mathbf{G}^v = \mathbf{G}^v\mathbf{Z}^v + \mathbf{E}^v, \quad \forall v \qquad (11)$$

which can be solved by utilizing the inexact augmented Lagrange multiplier (ALM) method [40]. In particular, we introduce the auxiliary variable $\mathbf{P}^v$ (*resp.*, $\mathbf{Q}^v$) to replace $\mathbf{Z}^v$ in $\|\mathbf{Z}^v\|_*$ [*resp.*, $\mathbf{Z}^v$ in the constraint (11)], and arrive at the augmented Lagrangian function as

$$\begin{aligned}\mathcal{L} = {} & \sum_{v=1}^{V}\left(\lambda_2\|\mathbf{E}^v\|_F^2 + \lambda_3\|\mathbf{P}^v\|_*\right) + \frac{\gamma}{2}\sum_{v,\tilde{v}:v\neq\tilde{v}}\|\mathbf{Z}^v - \mathbf{Z}^{\tilde{v}}\|_F^2 \\ & + \sum_{v=1}^{V}\langle\mathbf{S}^v, \mathbf{Z}^v - \mathbf{P}^v\rangle + \sum_{v=1}^{V}\langle\mathbf{T}^v, \mathbf{Z}^v - \mathbf{Q}^v\rangle \\ & + \sum_{v=1}^{V}\langle\mathbf{R}^v, \mathbf{G}^v - \mathbf{G}^v\mathbf{Q}^v - \mathbf{E}^v\rangle + \frac{\mu}{2}\sum_{v=1}^{V}\|\mathbf{Z}^v - \mathbf{P}^v\|_F^2 \\ & + \frac{\mu}{2}\sum_{v=1}^{V}\|\mathbf{Z}^v - \mathbf{Q}^v\|_F^2 + \frac{\mu}{2}\sum_{v=1}^{V}\|\mathbf{G}^v - \mathbf{G}^v\mathbf{Q}^v - \mathbf{E}^v\|_F^2\end{aligned} \qquad (12)$$

where $\mu > 0$ is a penalty parameter, and $\mathbf{S}^v$, $\mathbf{T}^v$, and $\mathbf{R}^v$ are the Lagrangian multipliers. The objective function in (12) can be minimized by using the inexact ALM approach, i.e., updating the variables $\{\mathbf{P}^v, \mathbf{Q}^v, \mathbf{Z}^v, \mathbf{E}^v\}$'s, the Lagrangian multipliers $\{\mathbf{S}^v, \mathbf{T}^v, \mathbf{R}^v\}$'s, and the penalty parameter μ in the augmented Lagrangian function (12) iteratively until the termination criterion is met. In the following, we will describe how to update $\mathbf{P}^v$, $\mathbf{Q}^v$, $\mathbf{Z}^v$, and $\mathbf{E}^v$ when fixing other variables one by one while the methods for updating $\mathbf{S}^v$, $\mathbf{T}^v$, $\mathbf{R}^v$, and μ are trivial and can be directly found in Algorithm 1.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

When fixing the other variables, the subproblem for updating $\{\mathbf{P}^v|_{v=1}^V\}$ is independent with respect to each $\mathbf{P}^v$, so we solve each $\mathbf{P}^v$ separately. After omitting and adding some constants, we reach the objective function with respect to $\mathbf{P}^v$ in (22), which can be solved by employing the singular value thresholding (SVT) algorithm [41].

When fixing the other variables, the subproblem for updating $\{\mathbf{Q}^v|_{v=1}^V\}$ is independent with respect to each $\mathbf{Q}^v$, so we solve each $\mathbf{Q}^v$ separately. By setting the derivative of the subproblem with respect to $\mathbf{Q}^v$ to zeros, we can easily obtain the solution to $\mathbf{Q}^v$ as in (23).

When fixing the other variables, the subproblem for updating $\{\mathbf{Z}^v|_{v=1}^V\}$ can be rewritten as (24) after omitting the terms without $\mathbf{Z}^v$s and using $\mathbf{H}^v$ to replace $(1/2)(\mathbf{P}^v + \mathbf{Q}^v - (1/\mu)(\mathbf{S}^v + \mathbf{T}^v))$, which has a close-form solution based on the vectorization of $\mathbf{Z}^v$.

When fixing the other variables, the subproblem for updating $\{\mathbf{E}^v|_{v=1}^V\}$ is independent with respect to each $\mathbf{E}^v$, so we solve each $\mathbf{E}^v$ separately. By setting the derivative of the subproblem with respect to $\mathbf{E}^v$ to zeros, the solution to $\mathbf{E}^v$ can be easily obtained as in (25). The steps to solve (12) are summarized in Algorithm 1.

Algorithm 1 Solving (12) With Inexact ALM

```
Input: $\mathbf{G}^v, \lambda_2, \lambda_3, \gamma$
Initialize $\mathbf{Z}^v = \mathbf{E}^v = \mathbf{S}^v = \mathbf{T}^v = \mathbf{R}^v = \mathbf{O}$, $\rho = 0.1$, $\mu = 0.1$, $\mu_{max} = 10^6$, $\nu = 10^{-5}$, $N_{iter} = 10^6$.
for $t = 1 : N_{iter}$ do
    For $v = 1, \ldots, V$, update $\mathbf{P}^v$ by solving
        $$\mathbf{P}^v = \arg\min_{\mathbf{P}^v} \lambda_3 \|\mathbf{P}^v\|_* + \frac{\mu}{2}\|\mathbf{P}^v - (\mathbf{Z}^v + \frac{\mathbf{S}^v}{\mu})\|_F^2. \tag{22}$$
    For $v = 1, \ldots, V$, update $\mathbf{Q}^v$ by
        $$\mathbf{Q}^v = (\mathbf{I} + \mathbf{G}^{v\prime}\mathbf{G}^v)^{-1}(\mathbf{G}^{v\prime}(\mathbf{G}^v - \mathbf{E}^v + \frac{\mathbf{R}^v}{\mu}) + \mathbf{Z}^v + \frac{\mathbf{T}^v}{\mu}). \tag{23}$$
    For $v = 1, \ldots, V$, update $\mathbf{Z}^v$ by solving
        $$\min_{\mathbf{Z}^v} \frac{\gamma}{2} \sum_{v,\tilde{v}:v\neq\tilde{v}} \|\mathbf{Z}^v - \mathbf{Z}^{\tilde{v}}\|_F^2 + \sum_{v=1}^{V} \mu\|\mathbf{Z}^v - \mathbf{H}^v\|_F^2, \tag{24}$$
        where $\mathbf{H}^v = \frac{1}{2}(\mathbf{P}^v + \mathbf{Q}^v - \frac{1}{\mu}(\mathbf{S}^v + \mathbf{T}^v))$.
    For $v = 1, \ldots, V$, update $\mathbf{E}^v$ by
        $$\mathbf{E}^v = \frac{\mu(\mathbf{G}^v - \mathbf{G}^v\mathbf{Q}^v) + \mathbf{R}^v}{2\lambda_2 + \mu}. \tag{25}$$
    For $v = 1, \ldots, V$, update $\mathbf{S}^v$, $\mathbf{T}^v$, and $\mathbf{R}^v$ by
        $$\mathbf{S}^v = \mathbf{S}^v + \mu(\mathbf{Z}^v - \mathbf{P}^v), \tag{26}$$
        $$\mathbf{T}^v = \mathbf{T}^v + \mu(\mathbf{Z}^v - \mathbf{Q}^v), \tag{27}$$
        $$\mathbf{R}^v = \mathbf{R}^v + \mu(\mathbf{G}^v - \mathbf{G}^v\mathbf{Q}^v - \mathbf{E}^v). \tag{28}$$
    Update the parameter $\mu$ by $\mu = \min(\mu_{max}, (1+\rho)\mu)$.
    Break if $\|\mathbf{G}^v - \mathbf{G}^v\mathbf{Q}^v - \mathbf{E}^v\|_\infty < \nu$, $\|\mathbf{Z}^v - \mathbf{P}^v\|_\infty < \nu$, $\|\mathbf{Z}^v - \mathbf{Q}^v\|_\infty < \nu$, $\forall v$.
end for
Output: $\mathbf{Z}^v$.
```

b) Update $\mathbf{W}^v$, $\mathbf{G}^v$, ξ_i^v, ϵ_{ij}^v: When fixing $\mathbf{Z}^v$, we equivalently replace $\mathbf{E}^v$ by $\mathbf{G}^v - \mathbf{G}^v\mathbf{Z}^v$ and rewrite the problem in (6) as

$$\min_{\substack{\mathbf{W}^v, \mathbf{G}^v \\ \xi_i^v, \epsilon_{ij}^v}} \sum_{v=1}^{V} \left(\frac{1}{2}\|\mathbf{W}^v\|_F^2 + C\sum_{i=1}^{n}\xi_i^v + C\sum_{i=1}^{n}\sum_{j=1}^{m}\epsilon_{ij}^v \right.$$
$$\left. + \lambda_1\|\mathbf{W}^v - \mathbf{G}^v\|_F^2 + \lambda_2\|\mathbf{G}^v - \mathbf{G}^v\mathbf{Z}^v\|_F^2 \right) \tag{13}$$
$$\text{s.t. } \mathbf{w}_i^{v\prime}\mathbf{x}_i^{v+} \geq 1 - \xi_i^v, \quad \xi_i^v \geq 0, \quad \forall v, \forall i, \tag{14}$$
$$\mathbf{w}_i^{v\prime}\mathbf{x}_j^{v-} \leq -1 + \epsilon_{ij}^v, \quad \epsilon_{ij}^v \geq 0, \quad \forall v, \forall i, \forall j. \tag{15}$$

It can be observed that the above problem contains V independent subproblems corresponding to V views. So we solve each subproblem by alternatively updating two sets of variables $\{\mathbf{W}^v, \xi_i^v, \epsilon_{ij}^v\}$ and $\mathbf{G}^v$ until the objective value of (13) converges. In particular, when fixing $\mathbf{G}^v$, the problem with respect to $\mathbf{W}^v$, ξ_i^v, and ϵ_{ij}^v can be separated into n independent subproblems with each related to one ESVM classifier. Thus, we have the following subproblem with respect to the ith ESVM classifier:

$$\min_{\mathbf{w}_i^v, \xi_i^v, \epsilon_{ij}^v} \frac{1}{2}\|\mathbf{w}_i^v\|^2 + C\left(\xi_i^v + \sum_{j=1}^{m}\epsilon_{ij}^v\right) + \lambda_1\|\mathbf{w}_i^v - \mathbf{g}_i^v\|^2 \tag{16}$$
$$\text{s.t. } \mathbf{w}_i^{v\prime}\mathbf{x}_i^{v+} \geq 1 - \xi_i^v, \quad \xi_i^v \geq 0 \tag{17}$$
$$\mathbf{w}_i^{v\prime}\mathbf{x}_j^{v-} \leq -1 + \epsilon_{ij}^v, \quad \epsilon_{ij}^v \geq 0, \quad \forall j \tag{18}$$

where $\mathbf{g}_i^v$ is the ith column vector of $\mathbf{G}^v$. We introduce the dual variables $\{\hat{\alpha}^+, \hat{\beta}^+\}$ and $\{\hat{\alpha}_j^-, \hat{\beta}_j^-\}$'s for the constraints in (17) and (18), respectively, and obtain the dual form of (16) as

$$\min_{\hat{\boldsymbol{\alpha}}} \hat{\boldsymbol{\alpha}}' \frac{\mathbf{K}_i^v \circ (\mathbf{y}\mathbf{y}')}{2(1+2\lambda_1)}\hat{\boldsymbol{\alpha}} + \left[\frac{2\lambda_1(\mathbf{X}_i^{v\prime}\mathbf{g}_i^v) \circ \mathbf{y}}{1+2\lambda_1} - \mathbf{1}\right]' \hat{\boldsymbol{\alpha}}$$
$$\text{s.t. } \mathbf{0} \leq \hat{\boldsymbol{\alpha}} \leq C\mathbf{1} \tag{19}$$

where $\mathbf{X}_i^v = [\mathbf{x}_i^{v+}, \mathbf{x}_1^{v-}, \ldots, \mathbf{x}_m^{v-}]$, $\mathbf{K}_i^v = \mathbf{X}_i^{v\prime}\mathbf{X}_i^v$, $\hat{\boldsymbol{\alpha}} = [\hat{\alpha}^+, \hat{\alpha}_1^-, \ldots, \hat{\alpha}_m^-]'$, and $\mathbf{y} = [1, -\mathbf{1}_m']'$. The problem in (19) is a quadratic programming (QP) problem, which can be solved efficiently by using the SMO algorithm [42], i.e., updating one selected dual variable in each iteration. With obtained $\hat{\boldsymbol{\alpha}}$, $\mathbf{w}_i^v$ can be recovered by using the following equation:

$$\mathbf{w}_i^v = \frac{1}{1+2\lambda_1}(2\lambda_1\mathbf{g}_i^v + \mathbf{X}_i^v(\mathbf{y} \circ \hat{\boldsymbol{\alpha}})). \tag{20}$$

When $\mathbf{W}^v$, ξ_i^v, and ϵ_{ij}^v are fixed, we have a closed-form solution for updating $\mathbf{G}^v$. In particular, by setting the derivative of (13) with respect to $\mathbf{G}^v$ to zeros, we can easily obtain the updating equation of $\mathbf{G}^v$ as

$$\mathbf{G}^v = \lambda_1\mathbf{W}^v\left(\lambda_2(\mathbf{I} - \mathbf{Z}^v)(\mathbf{I} - \mathbf{Z}^v)' + \lambda_1\mathbf{I}\right)^{-1}. \tag{21}$$

The whole algorithm is listed in Algorithm 2.

During the testing procedure, inspired by the prediction method in [16], given a test sample, we average the higher prediction scores of this sample obtained by using the exemplar classifiers on each view. By representing each test sample as $\mathbf{u} = (\mathbf{u}^1, \ldots, \mathbf{u}^V)$ with $\mathbf{u}^v$ being the vth view feature,

Algorithm 2 Exemplar-Based Multi-View Domain Generalization With Co-Regularizer

```
Input: Training data $\{\mathbf{x}_i^{v+}|_{i=1}^{n}\}$ and $\{\mathbf{x}_j^{v-}|_{j=1}^{m}\}$ with $V$ views.
Initialize[2] $\mathbf{G}^v$'s.
repeat
   Use Algorithm 1 to update $\mathbf{Z}^v$'s.
   repeat
      Solve $n$ independent subproblems in the dual form (19)
      and then recover $\mathbf{W}^v$ using (20) on each view.
      Update $\mathbf{G}^v$ by using (21) on each view.
   until The objective function of (13) converges.
until The objective function of (6) converges.
Output: The learnt classifier $\mathbf{W}^v$'s.
```

we formulate the final prediction score of $\mathbf{u}$ as

$$f(\mathbf{u}) = \frac{1}{V}\sum_{v=1}^{V}\frac{1}{|\Gamma(\mathbf{u}^v)|}\sum_{i:i\in\Gamma(\mathbf{u}^v)} f_i^v(\mathbf{u}^v) \tag{29}$$

where $f_i^v(\mathbf{u}^v)$ is the prediction score of $\mathbf{u}^v$ by using the ith ESVM classifier $\mathbf{w}_i^v$, and $\Gamma(\mathbf{u}^v)$ is the index set of ESVM classifiers, which obtain the top prediction scores on $\mathbf{u}^v$. Following [16], the cardinality of $\Gamma(\mathbf{u}^v)$ (i.e., $|\Gamma(\mathbf{u}^v)|$) is set as 5 in our experiments. By using this prediction method, we conjecture that this test sample is predicted by the ESVM classifiers learnt based on the positive training samples, which may come from the most relevant hidden latent domain. Consequently, the integrated classifier $f(\mathbf{u})$ in (29) is expected to generalize well to the arbitrary target domain.

C. Exemplar-Based Multi-View Domain Generalization Based on MKL

Inspired by MKL [27], in this section, we propose our EMVDG_MK approach by exploiting multi-view features based on the complementary principle. Specifically, in our multi-view scenario, multiple types of features may have complementary information, and thus it is beneficial to fuse the classifiers learnt on different views. By treating each view as a kernel, our problem can be considered as an MKL problem.

1) Formulation: Inspired by [27], we first write the primal form of MKL based on hard-margin[3] SVM with V-view features as

$$\min_{\mathbf{d},\mathbf{w}} \sum_{v=1}^{V}\frac{\|\mathbf{w}^v\|^2}{d_v} \tag{30}$$

$$\text{s.t. } \tilde{y}_i\sum_{v=1}^{V}\mathbf{w}^{v'}\mathbf{x}_i^v \ge 1, \quad \forall i$$

$$\mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \ge \mathbf{0} \tag{31}$$

where $\mathbf{d} = [d_1,\dots,d_V]'$, $\tilde{y}_i$ is the label of the ith training sample, $\mathbf{x}_i^v$ is the vth type of feature of the ith training sample, and $\mathbf{w}^v$ is the SVM classifier on the vth view. From (30), we can observe that the SVM classifiers $\mathbf{w}^v$s on different views are integrated based on the complementary principle.

By introducing dual variables α_is for the constraints in (31) and setting the derivative of the Lagrangian form with respect to each $\mathbf{w}^v$ to zeros, we can easily obtain the following equation:

$$\mathbf{w}^v = d_v\mathbf{X}^v(\boldsymbol{\alpha}\circ\tilde{\mathbf{y}}), \quad \forall v \tag{32}$$

where $\mathbf{X}^v = [\mathbf{x}_1^v,\dots,\mathbf{x}_{\tilde{n}}^v]$ with $\tilde{n}$ being the number of training samples, $\boldsymbol{\alpha} = [\alpha_1,\dots,\alpha_{\tilde{n}}]'$, and $\tilde{\mathbf{y}} = [\tilde{y}_1,\dots,\tilde{y}_{\tilde{n}}]'$. By substituting (32) back into the Lagrangian form of (30), we can obtain the dual form of (30) as the following min-max optimization problem:

$$\min_{\mathbf{d}}\max_{\boldsymbol{\alpha}} \; -\frac{1}{2}\sum_{v=1}^{V} d_v\boldsymbol{\alpha}'(\mathbf{K}^v\circ(\tilde{\mathbf{y}}\tilde{\mathbf{y}}'))\boldsymbol{\alpha} + \mathbf{1}'\boldsymbol{\alpha}$$

$$\text{s.t. } \boldsymbol{\alpha} \ge \mathbf{0}$$

$$\mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \ge \mathbf{0} \tag{33}$$

where $\mathbf{K}^v = \mathbf{X}^{v'}\mathbf{X}^v$ is the kernel matrix on the vth view. From the dual form in (33), we can observe that multiple kernels on different views are linearly combined with the coefficient $\mathbf{d}$ based on the complementary principle.

In (32), the ESVM classifiers on different views of the same positive sample share the same dual vector $\boldsymbol{\alpha}$, so we have in total n dual vectors, each of which corresponds to one positive sample. By using $\boldsymbol{\alpha}_i$ to denote the dual vector of the ESVM classifiers corresponding to the ith positive training sample, we can formulate our MKL problem with V views as

$$\min_{\mathbf{d}}\max_{\boldsymbol{\alpha}_i} \; -\frac{1}{2}\sum_{i=1}^{n}\sum_{v=1}^{V} d_v\boldsymbol{\alpha}_i'\mathbf{M}_i^v\boldsymbol{\alpha}_i + \sum_{i=1}^{n}\mathbf{1}'\boldsymbol{\alpha}_i$$

$$\text{s.t. } \boldsymbol{\alpha}_i \ge \mathbf{0}, \quad \forall i$$

$$\mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \ge \mathbf{0} \tag{34}$$

in which $\mathbf{d}$ is the same as defined in the paragraph below (30) and $\mathbf{M}_i^v = \mathbf{K}_i^v\circ(\mathbf{y}\mathbf{y}')$ with $\mathbf{y}$ and $\mathbf{K}_i^v$ being the same as defined in the paragraph below (19).

Recall that the positive training samples are likely to come from multiple hidden latent domains. When the jth positive training sample and the kth training sample come from the same latent domain, $\mathbf{X}_j^v$ and $\mathbf{X}_k^v$ should be similar, and the weight vectors of their corresponding ESVM classifiers (i.e., $\mathbf{w}_j^v$ and $\mathbf{w}_k^v$) should also be similar, as discussed in Section III-B. Moreover, similar to (32), we can easily derive that $\mathbf{w}_i^v = d_v\mathbf{X}_i^v(\boldsymbol{\alpha}_i\circ\mathbf{y})$, based on which we can infer that the dual vectors $\boldsymbol{\alpha}_j$ and $\boldsymbol{\alpha}_k$ should be similar when $\mathbf{w}_j^v$ is similar to $\mathbf{w}_k^v$ and $\mathbf{X}_j^v$ is similar to $\mathbf{X}_k^v$. Based on the above discussions, the dual vectors $\boldsymbol{\alpha}_i$'s can be organized into multiple hidden clusters. By denoting the dual matrix as $\mathbf{A} = [\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n] \in \mathcal{R}^{(m+1)\times n}$, we add a nuclear norm-based regularizer $\|\mathbf{A}\|_*$ to (34) to enforce $\mathbf{A}$ to be low rank, and

[2] We initialize $\mathbf{G}^v$ by using the weight vector of the exemplar classifier learnt based on the ith positive sample and all the negative samples on the vth view as its ith column vector.

[3] Our formulation can be similarly derived when using soft-margin SVM. Here, we use parameter-free hard-margin SVM for simplicity. Moreover, we do not employ the bias term explicitly. Instead, we augment each feature vector with an extra element of 1.

arrive at our final formulation

$$\begin{aligned}\min_{\mathbf{d}}\max_{\mathbf{A}}\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{\upsilon=1}^{V} d_\upsilon \boldsymbol{\alpha}_i' \mathbf{M}_i^\upsilon \boldsymbol{\alpha}_i + \sum_{i=1}^{n} \mathbf{1}'\boldsymbol{\alpha}_i - \zeta\|\mathbf{A}\|_* \\ \text{s.t. } & \boldsymbol{\alpha}_i \geq \mathbf{0}, \quad \forall i \\ & \mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \geq \mathbf{0} \end{aligned} \tag{35}$$

in which ζ is a tradeoff parameter.

2) Optimization: The problem in (35) is not easy to be optimized due to the regularizer $\|\mathbf{A}\|_*$, so we introduce an intermediate variable $\mathbf{B}$ and apply the low-rank regularizer on $\mathbf{B}$ instead of $\mathbf{A}$ and enforce $\mathbf{B}$ to be close to $\mathbf{A}$. Then, we reach the following formulation:

$$\begin{aligned}\min_{\mathbf{d}}\max_{\mathbf{A},\mathbf{B}}\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{\upsilon=1}^{V} d_\upsilon \boldsymbol{\alpha}_i' \mathbf{M}_i^\upsilon \boldsymbol{\alpha}_i + \sum_{i=1}^{n} \mathbf{1}'\boldsymbol{\alpha}_i \\ & - \zeta_1\|\mathbf{B}\|_* - \frac{\zeta_2}{2}\|\mathbf{A}-\mathbf{B}\|_F^2 \\ \text{s.t. } & \boldsymbol{\alpha}_i \geq \mathbf{0}, \quad \forall i \\ & \mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \geq \mathbf{0} \end{aligned} \tag{36}$$

in which ζ_1 and ζ_2 are two tradeoff parameters. It is obvious that the problem in (36) can reduce to the problem in (35) when ζ_2 approaches $+\infty$. Since the objective function in (36) is concave with respect to $\mathbf{B}$ and convex with respect to $\mathbf{d}$, $\min_{\mathbf{d}}$ and $\max_{\mathbf{B}}$ can be exchanged [43]. Then, we can rewrite (36) as

$$\begin{aligned}\max_{\mathbf{B}}\min_{\mathbf{d}}\max_{\mathbf{A}}\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{\upsilon=1}^{V} d_\upsilon \boldsymbol{\alpha}_i' \mathbf{M}_i^\upsilon \boldsymbol{\alpha}_i + \sum_{i=1}^{n} \mathbf{1}'\boldsymbol{\alpha}_i \\ & - \zeta_1\|\mathbf{B}\|_* - \frac{\zeta_2}{2}\|\mathbf{A}-\mathbf{B}\|_F^2 \\ \text{s.t. } & \boldsymbol{\alpha}_i \geq \mathbf{0}, \quad \forall i \\ & \mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \geq \mathbf{0}. \end{aligned} \tag{37}$$

Note that the inner problem of (37) with respect to $\mathbf{d}$ and $\mathbf{A}$ can be reformulated as a convex problem, and will be discussed in the proof of Proposition 1. Therefore, we solve (37) by using an alternating optimization approach. In particular, we alternatively update two sets of variables $\mathbf{B}$ and $\{\mathbf{A}, \mathbf{d}\}$ until the objective of (37) converges.

a) Update $\mathbf{B}$: When fixing $\mathbf{A}$ and $\mathbf{d}$, the problem in (37) reduces to the following problem:

$$\min_{\mathbf{B}} \zeta_1\|\mathbf{B}\|_* + \frac{\zeta_2}{2}\|\mathbf{A}-\mathbf{B}\|_F^2 \tag{38}$$

which can be solved by employing the SVT algorithm [41].

b) Update $\mathbf{A}, \mathbf{d}$: When fixing $\mathbf{B}$, the problem in (37) can reduce to the following problem:

$$\begin{aligned}\min_{\mathbf{d}}\max_{\boldsymbol{\alpha}_i}\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{\upsilon=1}^{V} d_\upsilon \boldsymbol{\alpha}_i' \mathbf{M}_i^\upsilon \boldsymbol{\alpha}_i + \sum_{i=1}^{n} 1'\boldsymbol{\alpha}_i \\ & - \frac{\zeta_2}{2}\sum_{i=1}^{n}\|\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i\|^2 \\ \text{s.t. } & \boldsymbol{\alpha}_i \geq \mathbf{0}, \quad \forall i \\ & \mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \geq \mathbf{0} \end{aligned} \tag{39}$$

in which β_i is the ith column of $\mathbf{B}$.

Interestingly, the primal form of (39) is closely related to the primal form of MKL in (30), which is described as follows.

Proposition 1: The primal form of (39) can be written as

$$\min_{\substack{\mathbf{d},\mathbf{w}_i^\upsilon \\ \tilde{\xi}_i,\tilde{\epsilon}_{ij}}} \frac{1}{2}\sum_{i=1}^{n}\sum_{\upsilon=1}^{V}\frac{\|\mathbf{w}_i^\upsilon\|^2}{d_\upsilon} + \frac{1}{2\zeta_2}\left(\sum_{i=1}^{n}\tilde{\xi}_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{\epsilon}_{ij}^2\right) \tag{40}$$

$$\text{s.t. } \sum_{\upsilon=1}^{V}\mathbf{w}_i^{\upsilon\prime}\mathbf{x}_i^{\upsilon+} \geq (1+\zeta_2\beta_i^+) - \tilde{\xi}_i, \quad \forall i \tag{41}$$

$$\sum_{\upsilon=1}^{V}\mathbf{w}_i^{\upsilon\prime}\mathbf{x}_j^{\upsilon-} \leq -(1+\zeta_2\beta_{ij}^-) + \tilde{\epsilon}_{ij}, \quad \forall i, \forall j \tag{42}$$

$$\mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \geq \mathbf{0} \tag{43}$$

where β_i^+'s and β_{ij}^-'s are newly introduced variables, and $\tilde{\xi}_i$'s and $\tilde{\epsilon}_{ij}$'s are the slack variables.

Proof: We prove that the dual form of (40) can be equivalently written as (39). After introducing the dual variables α_i^+'s for the constraints in (41) and α_{ij}^-'s for the constraints in (42), we arrive at the Lagrangian form of (40) as

$$\begin{aligned}\mathcal{L}_{\mathbf{w}} = {} & \frac{1}{2}\sum_{i=1}^{n}\sum_{\upsilon=1}^{V}\frac{\|\mathbf{w}_i^\upsilon\|^2}{d_\upsilon} + \frac{1}{2\zeta_2}\left(\sum_{i=1}^{n}\tilde{\xi}_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{\epsilon}_{ij}^2\right) \\ & - \sum_{i=1}^{n}\alpha_i^+\left(\sum_{\upsilon=1}^{V}\mathbf{w}_i^{\upsilon\prime}\mathbf{x}_i^{\upsilon+} - 1 - \zeta_2\beta_i^+ + \tilde{\xi}_i\right) \\ & + \sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_{ij}^-\left(\sum_{\upsilon=1}^{V}\mathbf{w}_i^{\upsilon\prime}\mathbf{x}_j^{\upsilon-} + 1 + \zeta_2\beta_{ij}^- - \tilde{\epsilon}_{ij}\right). \end{aligned} \tag{44}$$

By setting the derivatives of $\mathcal{L}_{\mathbf{w}}$ with respect to $\tilde{\xi}_i$, $\tilde{\epsilon}_{ij}$, and $\mathbf{w}_i^\upsilon$ to zeros separately, we obtain $\tilde{\xi}_i = \zeta_2\alpha_i^+$, $\tilde{\epsilon}_{ij} = \zeta_2\alpha_{ij}^-$, and the following equation:

$$\mathbf{w}_i^\upsilon = d_\upsilon \mathbf{X}_i^\upsilon(\boldsymbol{\alpha}_i \circ \mathbf{y}) \tag{45}$$

in which $\mathbf{X}_i^\upsilon$ and $\mathbf{y}$ are the same as defined in the paragraph below (19), and $\boldsymbol{\alpha}_i = [\alpha_i^+, \alpha_{i1}^-, \ldots, \alpha_{im}^-]'$ corresponds to the dual vector in (39). By substituting (45) back into (44), we can obtain the dual form of (40) as

$$\begin{aligned}\min_{\mathbf{d}}\max_{\boldsymbol{\alpha}_i}\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{\upsilon=1}^{V} d_\upsilon \boldsymbol{\alpha}_i' \mathbf{M}_i^\upsilon \boldsymbol{\alpha}_i + \sum_{i=1}^{n}(\mathbf{1}+\zeta_2\boldsymbol{\beta}_i)'\boldsymbol{\alpha}_i \\ & - \frac{\zeta_2}{2}\sum_{i=1}^{n}\|\boldsymbol{\alpha}_i\|^2 \\ \text{s.t. } & \boldsymbol{\alpha}_i \geq \mathbf{0}, \quad \forall i \\ & \mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \geq \mathbf{0} \end{aligned} \tag{46}$$

where $\boldsymbol{\beta}_i = [\beta_i^+, \beta_{i1}^-, \ldots, \beta_{im}^-]'$ corresponds to $\boldsymbol{\beta}_i$ in (39). After adding a constant term $-(\zeta_2/2)\sum_{i=1}^{n}\|\boldsymbol{\beta}_i\|^2$ in (46) followed by some simplifications, we can arrive at the exact form of (39). Therefore, we complete the proof here. ■

The problem in (40) is jointly convex with respect to $\mathbf{d}$, $\mathbf{w}_i^\upsilon$'s, $\tilde{\xi}_i$'s, and $\tilde{\epsilon}_{ij}$'s, so the global optimum can be achieved by using an alternative optimization approach. Specifically, when $\mathbf{d}$ is fixed, we solve $\boldsymbol{\alpha}_i$ in the dual form in (39) and then recover $\mathbf{w}_i^\upsilon$ by using (45). The subproblems with respect to each $\boldsymbol{\alpha}_i$ are

Algorithm 3 Exemplar-Based Multi-View Domain Generalization Based on MKL

```
Input: Training data $\{\mathbf{x}_i^{\upsilon+}|_{i=1}^n\}$ and $\{\mathbf{x}_j^{\upsilon-}|_{j=1}^m\}$ with $V$ views.
Initialize[4] $\mathbf{A}$, $\mathbf{d} = \frac{1}{V}\mathbf{1}$.
repeat
    Update $\mathbf{B}$'s by solving the problem in (38).
    repeat
        Update $\boldsymbol{\alpha}_i$'s by solving $n$ independent subproblems in the inner problem of (39) and then recover $\mathbf{w}_i^\upsilon$'s by using (45) on each view.
        Update $\mathbf{d}$ by using (47).
    until The objective function of (39) converges.
until The objective function of (37) converges.
Output: The learnt classifier $\mathbf{W}^\upsilon$'s.
```

independent to each subproblem being a QP problem, which can be solved efficiently by using the SMO algorithm [42]. When $\mathbf{w}_i^\upsilon$'s are fixed, we introduce a dual variable τ for the constraint $\mathbf{1}'\mathbf{d} = 1$ in (43) and set the derivative of the Lagrangian form with respect to d_υ to zero, which leads to $d_\upsilon = ((\sum_{i=1}^n \|\mathbf{w}_i^\upsilon\|^2)/(2\tau))^{1/2}$. Considering $\mathbf{1}'\mathbf{d} = 1$ and the equation in (45), we can easily obtain the closed-form solution for d_υ as

$$d_\upsilon = \frac{\sqrt{\sum_{i=1}^n \|\mathbf{w}_i^\upsilon\|^2}}{\sum_{\upsilon=1}^V \sqrt{\sum_{i=1}^n \|\mathbf{w}_i^\upsilon\|^2}} = \frac{\sqrt{\sum_{i=1}^n d_\upsilon^2 \boldsymbol{\alpha}_i' \mathbf{M}_i^\upsilon \boldsymbol{\alpha}_i}}{\sum_{\upsilon=1}^V \sqrt{\sum_{i=1}^n d_\upsilon^2 \boldsymbol{\alpha}_i' \mathbf{M}_i^\upsilon \boldsymbol{\alpha}_i}}. \tag{47}$$

The whole algorithm of EMVDG_MK is listed in Algorithm 3. In the testing stage, we use the same prediction method as for EMVDG_CO [see (29)] in Section III-B.

IV. Extending Our EMVDG Framework for Domain Adaptation

When we have unlabeled target domain samples in the training stage, our EMVDG framework can be extended to EMVDA by utilizing the unlabeled data for domain adaptation. Specifically, we further add a Laplacian regularizer, such that the prediction scores of target domain samples obtained by using the learnt ESVM classifiers should satisfy the smoothness constraint. This regularizer has proved to be effective for domain adaptation [44]. To be exact, when two target domain samples are similar, their prediction scores obtained by using the same set of SVM classifiers should be close to each other. We extend our EMVDG_CO and EMVDG_MK methods to EMVDA_CO and EMVDA_MK, respectively.

A. Exemplar-Based Multi-View Domain Adaptation With Co-Regularizer

We add a Laplacian regularizer to the objective function of our EMVDG_CO method [i.e., (6)] and formulate the objective function of our EMVDA_CO approach as

$$\begin{aligned}\min_{\substack{\mathbf{Z}^\upsilon,\mathbf{W}^\upsilon,\mathbf{G}^\upsilon\\ \mathbf{E}^\upsilon,\xi_i^\upsilon,\epsilon_{ij}^\upsilon}} \sum_{\upsilon=1}^V \Bigg(&\frac{1}{2}\|\mathbf{W}^\upsilon\|_F^2 + C\sum_{i=1}^n \xi_i^\upsilon + C\sum_{i=1}^n\sum_{j=1}^m \epsilon_{ij}^\upsilon \\ &+ \lambda_1\|\mathbf{W}^\upsilon - \mathbf{G}^\upsilon\|_F^2 + \lambda_2\|\mathbf{E}^\upsilon\|_F^2 + \lambda_3\|\mathbf{Z}^\upsilon\|_* \Bigg) \\ &+ \frac{\gamma}{2}\sum_{\upsilon,\tilde{\upsilon}:\upsilon\neq\tilde{\upsilon}} \|\mathbf{Z}^\upsilon - \mathbf{Z}^{\tilde{\upsilon}}\|_F^2 + \theta\sum_{\upsilon=1}^V \Omega(\mathbf{W}^\upsilon, \mathbf{L}^\upsilon, \mathbf{U}^\upsilon)\end{aligned} \tag{48}$$

$$\text{s.t. } \mathbf{w}_i^{\upsilon\prime}\mathbf{x}_i^{\upsilon+} \geq 1 - \xi_i^\upsilon, \quad \xi_i^\upsilon \geq 0, \quad \forall\upsilon, \forall i \tag{49}$$

$$\mathbf{w}_i^{\upsilon\prime}\mathbf{x}_j^{\upsilon-} \leq -1 + \epsilon_{ij}^\upsilon, \quad \epsilon_{ij}^\upsilon \geq 0, \quad \forall\upsilon, \forall i, \forall j \tag{50}$$

$$\mathbf{G}^\upsilon = \mathbf{G}^\upsilon\mathbf{Z}^\upsilon + \mathbf{E}^\upsilon, \quad \forall\upsilon \tag{51}$$

where θ is a tradeoff parameter, $\Omega(\mathbf{W}^\upsilon, \mathbf{L}^\upsilon, \mathbf{U}^\upsilon) = \mathrm{tr}(\mathbf{W}^{\upsilon\prime}\mathbf{U}^\upsilon\mathbf{L}^\upsilon\mathbf{U}^{\upsilon\prime}\mathbf{W}^\upsilon)$ is the Laplacian regularizer, in which $\mathbf{U}^\upsilon = [\mathbf{u}_1^\upsilon, \ldots, \mathbf{u}_N^\upsilon]$ is the target domain samples with N being the total number of unlabeled target domain samples and $\mathbf{u}_i^\upsilon$ being the υth type of feature of the ith target domain sample, and $\mathbf{L}^\upsilon$ is the Laplacian matrix constructed based on the target domain samples on the υth view. Note that we use the nearest neighbor graph to construct the Laplacian matrices $\mathbf{L}^\upsilon$s based on cosine similarity as suggested in [45].

We can solve the problem in (48) similar to that for solving (6). The only difference lies in that when updating $\mathbf{W}^\upsilon$ on the υth view, compared with (16), the subproblem with respect to the ith exemplar classifier has an additional Laplacian regularizer, which is written as

$$\begin{aligned}\min_{\mathbf{w}_i^\upsilon,\xi_i^\upsilon,\epsilon_{ij}^\upsilon} &\frac{1}{2}\|\mathbf{w}_i^\upsilon\|^2 + C\left(\xi_i^\upsilon + \sum_{j=1}^m \epsilon_{ij}^\upsilon\right) + \lambda_1\|\mathbf{w}_i^\upsilon - \mathbf{g}_i^\upsilon\|^2 \\ &+ \theta\mathbf{w}_i^{\upsilon\prime}\mathbf{U}^\upsilon\mathbf{L}^\upsilon\mathbf{U}^{\upsilon\prime}\mathbf{w}_i^\upsilon\end{aligned} \tag{52}$$

$$\text{s.t. } \mathbf{w}_i^{\upsilon\prime}\mathbf{x}_i^{\upsilon+} \geq 1 - \xi_i^\upsilon, \quad \xi_i^\upsilon \geq 0 \tag{53}$$

$$\mathbf{w}_i^{\upsilon\prime}\mathbf{x}_j^{\upsilon-} \leq -1 + \epsilon_{ij}^\upsilon, \quad \epsilon_{ij}^\upsilon \geq 0, \quad \forall j \tag{54}$$

which can also be solved in the dual form by using the SMO algorithm [42].

B. Exemplar-Based Multi-View Domain Adaptation Based on MKL

Similar to Section IV-A, we also add a Laplacian regularizer to the objective function of our EMVDG_MK method [i.e., (35)]. Recall that $\mathbf{w}_i^\upsilon = d_\upsilon\mathbf{X}_i^\upsilon(\boldsymbol{\alpha}_i \circ \mathbf{y})$ [see (45)], so we can derive the Laplacian regularizer $\Omega(\mathbf{W}^\upsilon, \mathbf{L}^\upsilon, \mathbf{U}^\upsilon) = \mathrm{tr}(\mathbf{W}^{\upsilon\prime}\mathbf{U}^\upsilon\mathbf{L}^\upsilon\mathbf{U}^{\upsilon\prime}\mathbf{W}^\upsilon) = d_\upsilon^2\sum_{i=1}^n \boldsymbol{\alpha}_i'(\mathbf{X}_i^{\upsilon\prime}\mathbf{U}^\upsilon\mathbf{L}^\upsilon\mathbf{U}^{\upsilon\prime}\mathbf{X}_i^\upsilon \circ (\mathbf{y}\mathbf{y}'))\boldsymbol{\alpha}_i$. Similar to the regularizer $\|\mathbf{w}_i^\upsilon\|^2$ in (40), we assign the weight $(1)/(d_\upsilon)$ to the Laplacian regularizer on the υth view. After denoting $\hat{\mathbf{K}}_i^\upsilon = \mathbf{X}_i^{\upsilon\prime}\mathbf{U}^\upsilon$ and adding the weighted Laplacian regularizer to (35), we formulate our EMVDA_MK method as

$$\begin{aligned}\min_{\mathbf{d}}\max_{\mathbf{A}} \quad &-\frac{1}{2}\sum_{i=1}^n\sum_{\upsilon=1}^V d_\upsilon\boldsymbol{\alpha}_i'\mathbf{M}_i^\upsilon\boldsymbol{\alpha}_i + \sum_{i=1}^n \mathbf{1}'\boldsymbol{\alpha}_i - \zeta\|\mathbf{A}\|_* \\ &-\frac{\vartheta}{2}\sum_{i=1}^n\sum_{\upsilon=1}^V d_\upsilon\boldsymbol{\alpha}_i'\big(\hat{\mathbf{K}}_i^\upsilon\mathbf{L}^\upsilon\hat{\mathbf{K}}_i^{\upsilon\prime} \circ (\mathbf{y}\mathbf{y}')\big)\boldsymbol{\alpha}_i \\ \text{s.t.} \quad &\boldsymbol{\alpha}_i \geq \mathbf{0}, \quad \forall i \\ &\mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \geq \mathbf{0}\end{aligned} \tag{55}$$

[4] We initialize $\mathbf{A}$ with its ith column vector being the dual vector of exemplar classifiers learnt based on the averaged kernel from V views, which are obtained based on the ith positive sample and all the negative samples.

where ϑ is a tradeoff parameter. After denoting $\hat{\mathbf{M}}_i^v = (\mathbf{K}_i^v + \vartheta \hat{\mathbf{K}}_i^v \mathbf{L}^v \hat{\mathbf{K}}_i^{v\prime}) \circ (\mathbf{y}\mathbf{y}')$, we can simplify (55) as

$$\begin{aligned}\min_{\mathbf{d}} \max_{\mathbf{A}} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{v=1}^{V} d_v \boldsymbol{\alpha}_i' \hat{\mathbf{M}}_i^v \boldsymbol{\alpha}_i + \sum_{i=1}^{n} \mathbf{1}'\boldsymbol{\alpha}_i - \zeta \|\mathbf{A}\|_* \\ \text{s.t. } & \boldsymbol{\alpha}_i \geq \mathbf{0}, \quad \forall i \\ & \mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \geq \mathbf{0} \end{aligned} \tag{56}$$

which shares a similar form with (35) except that we replace $\mathbf{M}_i^v$ by $\hat{\mathbf{M}}_i^v$. So the algorithm for solving (56) is similar to that for solving (35). The only difference lies in that when fixing $\mathbf{B}$ and updating $\{\mathbf{A}, \mathbf{d}\}$, the primal form of the subproblem can be written as

$$\begin{aligned}\min_{\substack{\mathbf{d}, \mathbf{w}_i^v \\ \tilde{\xi}_i, \tilde{\epsilon}_{ij}}} \quad & \frac{1}{2}\sum_{i=1}^{n}\sum_{v=1}^{V} \frac{\|\mathbf{w}_i^v\|^2}{d_v} + \frac{1}{2\zeta_2}\left(\sum_{i=1}^{n} \tilde{\xi}_i^2 + \sum_{i=1}^{n}\sum_{j=1}^{m} \tilde{\epsilon}_{ij}^2\right) \\ \text{s.t. } & \sum_{v=1}^{V} \mathbf{w}_i^{v\prime} \psi\left(\mathbf{x}_i^{v+}\right) \geq \left(1 + \zeta_2 \beta_i^+\right) - \tilde{\xi}_i, \quad \forall i \\ & \sum_{v=1}^{V} \mathbf{w}_i^{v\prime} \psi\left(\mathbf{x}_j^{v-}\right) \leq -\left(1 + \zeta_2 \beta_{ij}^-\right) + \tilde{\epsilon}_{ij}, \quad \forall i, \forall j \\ & \mathbf{1}'\mathbf{d} = 1, \quad \mathbf{d} \geq \mathbf{0} \end{aligned} \tag{57}$$

where $\psi(\cdot)$ is the feature mapping function induced by the kernel $(\mathbf{K}_i^v + \vartheta \hat{\mathbf{K}}_i^v \mathbf{L}^v \hat{\mathbf{K}}_i^{v\prime})$. The problem in (57) shares a similar form with (40) except that we apply the feature mapping function $\psi(\cdot)$ on $\mathbf{x}_i^{v+}$'s and $\mathbf{x}_j^{v-}$'s. Therefore, when updating $\mathbf{A}$ (*resp.*, $\mathbf{d}$), we replace $\mathbf{M}_i^v$ in (39) [*resp.*, (47)] by $\hat{\mathbf{M}}_i^v$.

V. Experiments

In this section, the effectiveness of our EMVDG and EMVDA frameworks for human action recognition and object recognition is demonstrated by extensive experiments on three benchmark data sets. In particular, we show that our EMVDG (*resp.*, EMVDA) framework outperforms all the state-of-the-art baselines in Section V-A (*resp.*, Section V-B). We also provide the insightful analysis on why our two methods under the EMVDG framework are effective. Moreover, we take the Office-Caltech data set as an example to show that the performance can be further improved by using more types of features in Section V-C.

A. Domain Generalization

1) Experimental Settings: All methods are evaluated for the human action recognition task on two benchmark data sets: ACT4^2 [46] and ORGBD [47].

The ACT4^2 data set consists of 2648 RGB-D videos from 14 action categories, which are captured from four camera viewpoints. As suggested in [46], the samples captured from each camera viewpoint are treated as one domain. Then, the videos from two domains and the remaining two domains are merged as the source domain and the target domain, respectively, which leads to in total six settings.

The ORGBD [47] contains the RGB-D videos from seven action categories. This data set has three sets with each set containing 112 videos, in which Set 3 is captured in one environment, while Set 1 and Set 2 are captured in another environment. In order to evaluate all methods for cross-environment human action recognition, two sets captured in different environments are merged as the source domain and the remaining one is treated as the target domain. Thus, we have a total of two settings, that is, Set 1 and 3 (*resp.*, Set 2 and 3) for training and Set 2 (*resp.*, Set 1) for testing.

For human action recognition on the ACT4^2 and ORGBD data sets, two types of features (i.e., RGB and depth) are used in the experiments. In particular, for each pair of RGB and depth videos in both ACT4^2 and ORGBD data sets, we extract the IDT descriptors [48]. Compared with the preliminary conference version of this paper [28], we use the Fisher vector encoding method instead of BOW to encode the IDT descriptors. Specifically, following [2], we train 256 Gaussian mixture models based on the IDT descriptors from the training videos, and then extract a 109 056-dim Fisher vector for each training and testing video. Finally, we perform PCA to reduce the dimension of Fisher vectors to 10 000.

Moreover, all methods are also evaluated for the object recognition task on the benchmark data set Office-Caltech [2]. The images in the Office-Caltech data set are from four domains, that is, Caltech-256 (C), Amazon (A), Webcam (W), and Digital SLR (D). Following the experimental setting in [2], the ten common categories among the four domains are used, which consists of a total of 2533 images. As suggested in [16] and [23], we mix D and W (*resp.*, C, D, and W; A and C) as the source domain and the remaining domains are used as the target domain, which leads to three experimental settings in total. For each image, we extract the 4096-dim DeCAF$_6$ feature [49] and the 4096-dim Caffe$_6$ [50] feature as two-view features.

2) Baselines: We compare EMVDG_CO and EMVDG_MK methods with two basic baselines, i.e., SVM [51] and ESVM (ESVM) [22], as well as three sets of baseline methods: the multi-view learning approaches, the domain generalization approaches, and the latent domain discovering approaches. For SVM, the classifiers are trained on each view, and then, we fuse the prediction scores from two views for the final prediction. For ESVM, one ESVM classifier is trained for each positive training sample on each view, and then, we use the same prediction method as in (29).

The multi-view learning baseline methods contain SVM-2K [31], KCCA [30], low-rank common subspace (LRCS) [32], and MKL [27] by utilizing two types of features, i.e., RGB/DeCAF$_6$ features and depth/Caffe$_6$ features.

The domain generalization baseline methods include LRESVM [16] and domain-invariant component analysis (DICA) [15]. Under the multi-view setting, LRESVM and DICA are employed on each view, and then, we fuse the prediction scores from multiple views.

The latent domain discovering methods contain [23] and [24]. We learn the SVM classifiers for each discovered latent domain, followed by employing two prediction strategies named "ensemble" and "match" as

TABLE I

AVERAGE ACCURACIES (%) OVER MULTIPLE SETTINGS OF DIFFERENT APPROACHES ON EACH DATA SET WITHOUT USING THE TARGET DOMAIN SAMPLES DURING THE TRAINING PROCEDURE. WE DENOTE THE BEST RESULTS IN BOLDFACE

Dataset	ACT4^2	ORGBD	Office-Caltech
SVM [51]	68.10	62.05	84.52
ESVM [22]	69.11	62.95	86.14
LRCS [32]	70.81	66.07	85.28
SVM-2K [31]	70.34	65.63	86.10
KCCA [30]	69.56	63.84	86.33
MKL [27]	69.98	65.18	86.50
DICA [15]	69.53	66.52	86.12
LRESVM [16]	71.18	67.42	87.04
[23](match)	70.05	65.63	86.47
[23](ensemble)	69.28	66.52	86.06
[24](match)	68.60	61.16	85.75
[24](ensemble)	68.66	65.63	84.81
Sub-Cate [52]	69.90	64.74	86.64
EMVDG_CO_sim	72.10	67.86	87.72
EMVDG_CO	**74.22**	69.20	88.13
EMVDG_MK	73.08	**70.54**	**88.33**

in [16]. Similar to latent domains discovering algorithms, subcategorization methods aim at discovering subcategories within each category, which can also be applied to our task. Therefore, we include the discriminative subcategorization (Sub-Cate) method [52] as a baseline. For all the above methods, we employ them on each view, and then average the prediction scores from two views.

Moreover, in order to validate the co-regularizer in (2), we additionally report the results of a simplified version of our EMVDG_CO method, which is named EMVDG_CO_sim, in which the co-regularizer $\sum_{v,\tilde{v}:v\neq\tilde{v}} \|\mathbf{Z}^v - \mathbf{Z}^{\tilde{v}}\|_F^2$ is removed by setting γ to 0.

For performance evaluation, the recognition accuracy is used for all approaches. For our EMVDG_CO method, the parameters are empirically fixed as $C = 0.1$, $\lambda_1 = 100$, $\lambda_2 = 10$, $\lambda_3 = 0.1$, and $\gamma = 100$ for all settings on all data sets. For our EMVDG_MK method, the parameters are empirically fixed as $\zeta_1 = 10$, $\zeta_2 = 10000$ for all settings on all data sets. For the baselines, the optimal parameters are chosen based on their best performance on the testing set. Due to the space limitation, only the average accuracy over the 3 (*resp.*, 6, 2) settings for the Office-Caltech (*resp.*, ACT4^2, ORGBD) data set is reported.

3) Results: We summarize the experimental results in Table I, from which we observe that ESVM outperforms SVM, which indicates the effectiveness of fusing multiple ESVM classifiers to enhance the domain generalization ability.

Multi-view learning approaches LRCS, SVM-2K, KCCA, and MKL achieve better results than SVM, because they exploit the relation among multiple types of features. LRESVM, DICA, and Sub-Cate are all better than SVM, which indicates that it is useful to exploit the intrinsic structure when the training data are sampled from multiple latent domains. The latent domain discovering approaches [23], [24] using the "match" or "ensemble" strategy generally outperform SVM, which shows the effectiveness of discovering the latent domains.

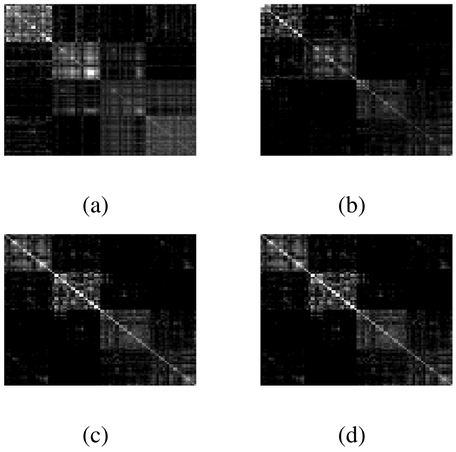


Fig. 1. Illustration of the learnt representation matrices $\mathbf{Z}^v$s on two views for the action "Put On" on the ACT4^2 data set when treating the camera viewpoint 1 and 4 (*resp.*, 2 and 3) as the source (*resp.*, target) domain. (a) Z^{RGB} without co-reg. (b) Z^{depth} without co-reg. (c) Z^{RGB} with co-reg. (d) Z^{depth} with co-reg.

Another observation is that EMVDG_CO_sim is better than ESVM on all data sets. Since ESVM can be treated as a special case of our EMVDG_CO_sim method without employing LRR, the results indicate the benefits of exploiting the low-rank structure of positive training samples for domain generalization. Our EMVDG_CO method achieves better results than its simplified version EMVDG_CO_sim, which shows that our new co-regularizer $\sum_{v,\tilde{v}:v\neq\tilde{v}} \|\mathbf{Z}^v - \mathbf{Z}^{\tilde{v}}\|_F^2$ is effective. Therefore, it is useful to jointly exploit the cluster structures from multiple views.

Our EMVDG_CO and EMVDG_MK methods outperform all the baseline methods on all three data sets, which indicates that our EMVDG framework can improve the domain generalization ability and utilize multiple types of features effectively. Note that there is no consistent winner in our EMVDG framework. In particular, our EMVDG_CO method achieves the best result on the ACT4^2 data set, while our EMVDG_MK method outperforms EMVDG_CO on the ORGBD data set and achieves the comparable result on the Office-Caltech data set.

4) Analysis on the Learnt Representation Matrices Using EMVDG_CO_sim and EMVDG_CO: In order to demonstrate how our EMVDG_CO method exploits the latent domains of positive training samples in an intuitive way, we take the ACT4^2 data set as an example to compare the representation matrices $\mathbf{Z}^v$s (i.e., $\mathbf{Z}^{\text{RGB}}$ and $\mathbf{Z}^{\text{depth}}$) learnt by using our EMVDG_CO method and its simplified version EMVDG_CO_sim in Fig. 1, which correspond to MVDG and MVDG (without co-reg) in the preliminary conference version respectively. Recall that the representation matrix $\mathbf{Z}^v$ encodes the cluster structure of exemplar classifiers, in which the between-cluster (*resp.*, within-cluster) entries are generally sparse (*resp.*, dense). Therefore, in ideal cases, $\mathbf{Z}^v$ should be block-diagonal with each block representing a latent domain. From Fig. 1, we observe that all four representation matrices exhibit block-diagonal structure, which indicates that it is effective to discover hidden latent domains by employing LRR on each view. It is worth noting that although only two domains (i.e., camera viewpoint 1 and camera viewpoint 4)

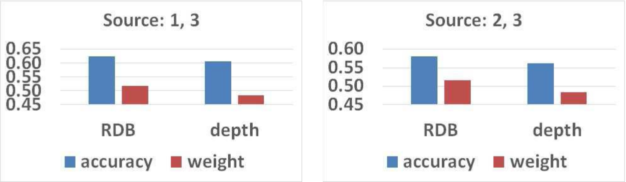


Fig. 2. Illustration of the kernel combination weights and the accuracies of SVM based on each kernel corresponding to RGB and depth features on the two settings on the ORGBD data set.

are merged as the source domain, there are in fact four blocks in Fig. 1, which means that totally four latent domains are discovered. We conjecture that actors are likely to put on clothes from two opposite directions with each direction leading to a latent domain. Thus, the videos captured from each camera viewpoint actually contain two latent domains. Another observation is that with our newly proposed co-regularizer, the two representation matrices learnt by using our EMVDG_CO method are more consistent and also exhibit relatively clearer block-diagonal structure than those learnt by using the simplified version EMVDG_CO_sim. This result demonstrates the benefits of using our co-regularizer. We have similar observations for the other scenarios.

5) Analysis on the Learnt Kernel Weights Using EMVDG_MK: From Table I, we observe that our EMVDG_MK method outperforms our EMVDG_CO method on the ORGBD data set, which could be explained as follows. Since two types of features (i.e., RGB and depth) are used on the ORGBD data set, we conjecture that one of them (i.e., RGB or depth) is more discriminative, so that assigning higher weight for more discriminative features will help better exploit the latent domain structure and learn more robust classifiers, which leads to better performance. To this end, we analyze the learnt kernel weights **d** in (35) by taking the two settings on the ORGBD data set as examples.

To capture the relation between the kernels constructed from different types of features and the learnt kernel weights, we additionally report the accuracies of SVM by using only RGB or depth features. When the performance of SVM obtained based on one feature is higher than the other one, the corresponding kernel is expected to be more discriminative and the weight assigned to this kernel is expected to be higher. With regards to the kernel weights, we have a set of learnt kernel combination weights for each category, as our EMVDG_MK method is under the binary classification setting. For better representation, we report the average of the learnt kernel weights over all categories. To this end, we illustrate the accuracies of SVM and the learnt kernel weights in Fig. 2, from which we observe that the SVM classifiers trained based on the RGB features achieve better performance on both settings. Moreover, higher weights are correctly assigned to the RGB kernel by EMVDG_MK on both settings, which demonstrates that our EMVDG_MK method can select more discriminative kernels.

B. Domain Adaptation

1) Experimental Settings: We use the same experimental settings as in Section V-A except that we additionally use the unlabeled target domain data in the training process.

2) Baselines: We compare our EMVDA framework, including EMVDA_CO and EMVDA_MK methods, with three sets of baseline methods: the domain adaptation approaches and the multi-view semi-supervised learning approaches as well as the existing multi-view domain adaptation approaches.

The domain adaptation baselines are kernel mean matching (KMM) [6], domain adaptive SVM (DASVM) [5], domain-invariant projection (DIP) [3], subspace alignment (SA) [4], transfer component analysis (TCA) [53], sampling geodesic flow (SGF) [1], and geodesic flow kernel (GFK) [2]. The above domain adaptation approaches are employed on each view, followed by fusing the prediction scores from two views using the late fusion strategy.

Our EMVDA framework is also compared with multi-view semi-supervised learning approaches Co-LapSVM [36] and Co-training [35], together with the multi-view domain adaptation approaches including multi-view transfer learning (MVTL_LM) [20], Coupled [19], multi-view discriminant transfer (MDT) [21], and domain transfer multiple kernel learning (DTMKL) [7], which exploit the relation among multiple types of features and simultaneously cope with the domain distribution mismatch. In addition, we compare our EMVDA framework with LRCS [32] by using the target domain samples as the dictionary as suggested in [32].

Compared with EMVDG_CO, our EMVDA_CO method has an extra parameter θ, which is empirically set as 10^{-5} for all settings on all data sets. Similarly, compared with EMVDG_MK, our EMVDA_MK method has an extra parameter ϑ, which is empirically set as 10^{-7} for all settings on all data sets. For the baselines, the optimal parameters are chosen based on their best results on the testing set. Due to the space limitation, only the average accuracy over the three (*resp.*, 6, 2) settings for the Office-Caltech (*resp.*, ACT4^2, ORGBD) data set is reported.

3) Results: We summarize the experimental results in Table II. The results of SVM from Table I are also included for comparison. It can be observed that the domain adaptation approaches DASVM, KMM, SA, DIP, GFK, TCA, and SGF outperform SVM, which indicates the advantage of reducing the domain distribution mismatch between the source domain and the target domain.

We also observe that the multi-view semi-supervised learning approaches Co-LapSVM and Co-training as well as the multi-view domain adaptation approaches Coupled, MVTL_LM, MDT, and DTMKL are generally better than the multi-view learning approaches reported in Table I, which demonstrates the effectiveness of utilizing the unlabeled target domain samples. Another observation is that the multi-view domain adaptation approaches are generally better than or comparable with other domain adaptation approaches, which shows the advantage of additionally exploiting the relation among multiple views. LRCS also achieves better results by using the target domain data as the dictionary, when compared with its corresponding results without using the target domain data.

TABLE II

AVERAGE ACCURACIES (%) OVER MULTIPLE SETTINGS OF DIFFERENT APPROACHES ON EACH DATA SET AFTER UTILIZING THE TARGET DOMAIN SAMPLES DURING THE TRAINING PROCEDURE. THE BEST RESULTS ARE DENOTED IN BOLDFACE

Dataset	ACT4^2	ORGBD	Office-Caltech
SVM [51]	68.10	62.05	84.52
DASVM [5]	70.61	65.63	85.60
KMM [6]	71.74	65.18	86.34
TCA [53]	70.29	64.74	85.79
SA [4]	71.85	67.42	86.79
DIP [3]	71.49	66.07	86.58
GFK [2]	70.31	65.63	86.22
SGF [1]	71.31	62.50	85.78
Co-training [35]	73.36	67.86	87.96
Co-LapSVM [36]	73.76	68.31	88.20
Coupled [19]	73.16	68.31	86.48
MVTL_LM [20]	73.80	62.95	87.76
MDT [21]	72.66	67.42	86.87
DTMKL [7]	73.57	68.31	88.07
LRCS [32]	73.04	67.42	86.12
EMVDA_CO	**76.18**	70.09	**91.04**
EMVDA_MK	75.08	**71.88**	89.84

Our EMVDA_CO (*resp.*, EMVDA_MK) method outperforms our EMVDG_CO (*resp.*, EMVDG_MK) method reported in Table I, which indicates the benefits of utilizing the unlabeled target domain data during the training procedure. Moreover, our EMVDA_CO and EMVDA_MK methods outperform all the baselines on all data sets. Our EMVDA_CO method achieves the best results on the ACT4^2 and Office-Caltech data set, while our EMVDA_MK achieves the best result on the ORGBD data set.

C. Utilizing Multiple Types of Features

Although we only use two types of features (i.e., RGB/depth features for human action recognition and Decaf$_6$/Caffe$_6$ for object recognition) in Sections V-A and V-B, our EMVDG and EMVDA frameworks can be readily used for multiple types of features. When employing more types of features, our EMVDG_MK and EMVDA_MK methods are much more efficient than our EMVDG_CO and EMVDA_CO methods, which can be explained as follows. For our EMVDG_CO method, we need to update $\mathbf{W}^v$s and $\mathbf{G}^v$s on each view as indicated in Algorithm 2, and update $\mathbf{Z}^v$s by solving the subproblems on each view as indicated in Algorithm 1. So the training time of our EMVDG_CO method increases linearly with the number of views. In contrast, for our EMVDG_MK method, the most time-consuming steps are to solve the problem in (38) and the inner problem of (39), and their time complexity is irrelevant to the number of views, as indicated in Algorithm 3. So the extra training time of our EMVDG_MK method with multiple types of features is much less than that of EMVDG_CO. The analysis of the time complexity for EMVDA_CO and EMVDA_MK is similar to that for EMVDG_CO and EMVDG_MK, respectively.

To compare EMVDG_MK (*resp.*, EMVDA_MK) with EMVDG_CO (*resp.*, EMVDA_CO) in terms of the training time and accuracy when using different numbers of views, we take the Office-Caltech data set as an example to conduct experiments on a server machine with Intel Xeon 3.2-GHz CPUs and 16-GB RAM using a single thread. Besides the Decaf$_6$ and Caffe$_6$ features, we additionally use Decaf$_7$ and Caffe$_7$ features, which leads to four types of features in total. The average training time over three settings of our four methods is reported in Table III, from which we can observe that the training time of EMVDG_CO and EMVDA_CO approximately increases linearly as the number of views increases while the training time of EMVDG_MK and EMVDA_MK increases much less. We also report the average accuracies over three settings of our four methods in Table IV, from which we observe that the performances of all four methods are improved after employing two more types of features. When using four types of features, EMVDG_CO (*resp.*, EMVDA_CO) achieves better result than EMVDG_MK (*resp.*, EMVDA_MK). However, our EMVDG_MK and EMVDA_MK methods are much more efficient.

TABLE III

AVERAGE TRAINING TIME (s) OF OUR EMVDG AND EMVDA FRAMEWORKS ON THE OFFICE-CALTECH DATA SET BY EMPLOYING TWO-VIEW OR FOUR-VIEW FEATURES

Method	EMVDG_CO	EMVDG_MK	EMVDA_CO	EMVDA_MK
2 views	2795.2013	399.4078	3245.8470	570.6233
4 views	4815.8333	439.0771	7630.5167	643.8519

TABLE IV

AVERAGE ACCURACIES (%) OF OUR EMVDG AND EMVDA FRAMEWORKS ON THE OFFICE-CALTECH DATA SET BY EMPLOYING TWO-VIEW OR FOUR-VIEW FEATURES

Method	EMVDG_CO	EMVDG_MK	EMVDA_CO	EMVDA_MK
2 views	88.13	88.33	91.04	89.84
4 views	91.54	90.63	93.26	92.20

VI. CONCLUSION

In this paper, an EMVDG framework has been proposed for visual recognition. Our framework can enhance the domain generalization capability to the arbitrary target domain and simultaneously exploit the relation among multiple types of features. Moreover, our EMVDG framework has been further extended to a new domain adaptation framework named EMVDA by additionally using the unlabeled target domain samples in the training process. The effectiveness of our EMVDG and EMVDA frameworks has been demonstrated by extensive experiments for visual recognition on three benchmark data sets.

REFERENCES

[1] R. Gopalan, R. Li, and R. Chellappa, “Domain adaptation for object recognition: An unsupervised approach,” in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 999–1006.

[2] B. Gong, Y. Shi, F. Sha, and K. Grauman, “Geodesic flow kernel for unsupervised domain adaptation,” in *Proc. IEEE Conf. CVPR*, Jun. 2012, pp. 2066–2073.

[3] M. Baktashmotlagh, M. T. Harandi, B. C. Lovell, and M. Salzmann, “Unsupervised domain adaptation by domain invariant projection,” in *Proc. IEEE ICCV*, Dec. 2013, pp. 769–776.

[4] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, “Unsupervised visual domain adaptation using subspace alignment,” in *Proc. IEEE ICCV*, Dec. 2013, pp. 2960–2967.

[5] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A DASVM classification technique and a circular validation strategy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 770–787, May 2010.

[6] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schölkopf, "Correcting sample selection bias by unlabeled data," in *Proc. 20th Annu. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2006, pp. 601–608.

[7] 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.

[8] L. Cheng and S. J. Pan, "Semi-supervised domain adaptation on manifolds," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2240–2249, Dec. 2014.

[9] L. Niu, W. Li, and D. Xu, "Exploiting privileged information from web data for action and event recognition," *Int. J. Comput. Vis.*, vol. 118, no. 2, pp. 130–150, Jun. 2016.

[10] L. Niu, J. Cai, and D. Xu, "Domain adaptive fisher vector for visual recognition," in *Proc. 14th Eur.Conf. Comput. Vis.*, The Netherlands, Oct. 2016, pp. 550–566.

[11] L. Duan, D. Xu, I. W.-H. Tsang, and J. Luo, "Visual event recognition in videos by learning from Web data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1667–1680, Sep. 2012.

[12] L. Duan, D. Xu, and I. W. Tsang, "Domain adaptation from multiple sources: A domain-dependent regularization approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 504–518, Mar. 2012.

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

[14] W. Li, L. Niu, and D. Xu, "Exploiting privileged information from web data for image categorization," in *Proc. 13th Eur. Conf. Comput. Vis.*, Zürich, Switzerland, Sep. 2014, pp. 437–452.

[15] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain generalization via invariant feature representation," in *Proc. 30th IEEE Int. Conf. Mach. Learn.*, Atlanta, GA, USA, Jun. 2013, pp. 10–18.

[16] Z. Xu, W. Li, L. Niu, and D. Xu, "Exploiting low-rank structure from latent domains for domain generalization," in *Proc. 13th Eur. Conf. Comput. Vis.*, Zürich, Switzerland, Sep. 2014, pp. 628–643.

[17] L. Niu, W. Li, and D. Xu, "Visual recognition by learning from Web data: A weakly supervised domain generalization approach," in *Proc. 28th IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 2774–2783.

[18] L. Niu, W. Li, D. Xu, and J. Cai, "Visual recognition by learning from web data via weakly supervised domain generalization," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.

[19] J. Blitzer, S. Kakade, and D. P. Foster, "Domain adaptation with coupled subspaces," in *Proc. Int. Conf. Artif. Intell. Statist.*, Lauderdale, FL, USA, Apr. 2011, pp. 173–181.

[20] D. Zhang, J. He, Y. Liu, L. Si, and R. Lawrence, "Multi-view transfer learning with a large margin approach," in *Proc. 17th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, San Diego, CA, USA, Aug. 2011, pp. 1208–1216.

[21] P. Yang and W. Gao, "Multi-view discriminant transfer learning," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, Beijing, China, Aug. 2013, pp. 1848–1854.

[22] T. Malisiewicz, A. Gupta, and A. A. Efros, "Ensemble of exemplar-SVMs for object detection and beyond," in *Proc. ICCV*, Nov. 2011, pp. 89–96.

[23] B. Gong, K. Grauman, and F. Sha, "Reshaping visual datasets for domain adaptation," in *Proc. 27th Annu. Conf. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 2013, pp. 1286–1294.

[24] J. Hoffman, B. Kulis, T. Darrell, and K. Saenko, "Discovering latent domains for multisource domain adaptation," in *Proc. 12th Eur. Conf. Comput. Vis.*, Florence, Italy, Oct. 2012, pp. 702–715.

[25] C. Xu, D. Tao, and C. Xu. (Apr. 2013). "A survey on multi-view learning." [Online]. Avaliable: https://arxiv.org/abs/1304.5634

[26] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 663–670.

[27] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the smo algorithm," in *Proc. 21th Int. Conf. Mach. Learn.*, Banff, Canada, Jul. 2004, p. 6.

[28] L. Niu, W. Li, and D. Xu, "Multi-view domain generalization for visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 4193–4201.

[29] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual categorization: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 5, pp. 1019–1034, May 2015.

[30] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, 2004.

[31] J. D. R. Farquhar, D. R. Hardoon, H. Meng, J. Shawe-Taylor, and S. Szedmak, "Two view learning: SVM-2K, theory and practice," in *Proc. 19th Annu. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2005, pp. 355–362.

[32] Z. Ding and Y. Fu, "Low-rank common subspace for multi-view learning," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Dec. 2014, pp. 110–119.

[33] A. Iosifidis, A. Tefas, and I. Pitas, "View-invariant action recognition based on artificial neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 412–424, Mar. 2012.

[34] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, Jan. 2004.

[35] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. Theory*, New York, NY, USA, Jul. 1998, pp. 92–100.

[36] V. Sindhwani, P. Niyogi, and M. Belkin, "A co-regularization approach to semi-supervised learning with multiple views," in *Proc. 22nd Int. Conf. Mach. Learn. Workshop Learn. Multiple Views*, Bonn, Germany, Aug. 2005, pp. 74–79.

[37] M. B. Blaschko, C. H. Lampert, and A. Gretton, "Semi-supervised laplacian regularization of kernel canonical correlation analysis," in *Proc. 9th Eur. Conf. Mach. Learn. Knowl. Discovery*, Antwerp, Belgium, Sep. 2008, pp. 133–145.

[38] A. Shrivastava, T. Malisiewicz, A. Gupta, and A. A. Efros, "Data-driven visual similarity for cross-domain image matching," *ACM Trans. Graph.*, vol. 30, no. 6, p. 154, Dec. 2011.

[39] J. Zepeda and P. Perez, "Exemplar svms as visual feature encoders," in *Proc. 28th IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 3052–3060.

[40] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[41] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[42] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Microsoft Res. Tech. Rep. MSR-TR-98-14, Apr. 1998.

[43] Y.-F. Li, I. W. Tsang, J. T. Kwok, and Z.-H. Zhou, "Tighter and convex maximum margin clustering," in *Proc. 12th Int. Conf. Artif. Intell. Statist.*, Clearwater Beach, FL, USA, Apr. 2009, pp. 344–351.

[44] J. Donahue, J. Hoffman, E. Rodner, K. Saenko, and T. Darrell, "Semi-supervised domain adaptation with instance constraints," in *Proc. CVPR*, 2013, pp. 668–675.

[45] G. Ye, D. Liu, I.-H. Jhuo, and S.-F. Chang, "Robust late fusion with rank minimization," in *Proc. CVPR*, Jun. 2012, pp. 3021–3028.

[46] Z. Cheng, L. Qin, Y. Ye, Q. Huang, and Q. Tian, "Human daily action analysis with multi-view and color-depth data," in *Proc. 12th Eur. Conf. Comput. Vis., Workshop Consum. Depth Cameras for Comput. Vis.*, Firenze, Italy, Oct. 2012, pp. 52–61.

[47] G. Yu, Z. Liu, and J. Yuan, "Discriminative orderlet mining for real-time recognition of human-object interaction," in *Proc. 12th Asian Conf. Comput. Vis.*, Singapore, Nov. 2014, pp. 50–65.

[48] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3551–3558.

[49] J. Donahue *et al.*, "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. 31st IEEE Int. Conf. Mach. Learn.*, Beijing, China, Jun. 2014, pp. 647–655.

[50] Y. Jia *et al.*. (Jun. 2014). "Caffe: Convolutional architecture for fast feature embedding." [Online]. Avaliable: https://arxiv.org/abs/1408.5093

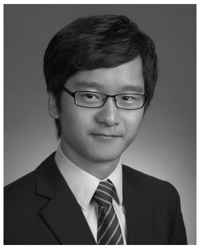
[51] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[52] M. Hoai and A. Zisserman, "Discriminative sub-categorization," in *Proc. 26th IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, Jun. 2013, pp. 1666–1673.

[53] 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.

Li Niu received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2011. He is currently pursuing the Ph.D. degree with the Interdisciplinary Graduate School, Nanyang Technological University, Singapore.

His current research interests include machine learning and computer vision.



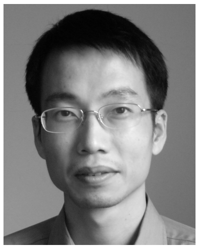
Wen Li received the B.S. and M.Eng degrees from the Beijing Normal University, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree from the Nanyang Technological University, Singapore, in 2015.

He is currently a Post-Doctoral Researcher with the Computer Vision Laboratory, ETH Zürich, Switzerland. His current interests include transfer learning, multi-view learning, multiple kernel learning, and their applications in computer vision.

Dong Xu (M'07–SM'13) received the B.Eng. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2001 and 2005, respectively.

During his Ph.D. study, he also worked at the Microsoft Research Asia and The Chinese University of Hong Kong for more than two years. He was a Post-Doctoral Research Scientist with Columbia University from 2006 to 2007, and a Faculty Member with Nanyang Technological University, from 2007 to 2015. He is currently a professor (Chair in computer engineering) with the School of Electrical and Information Engineering, The University of Sydney, Australia. He has authored over 100 papers in IEEE Transactions and top tier conferences. His current research interests include computer vision, multimedia, machine learning, and biomedical image analysis.

Dr. Xu co-authored work on transfer learning for video event recognition received the Best Student Paper Award in the IEEE International Conference on Computer Vision and Pattern Recognition in 2010, Another his co-authored work also won the IEEE TRANSACTIONS ON MULTIMEDIA Prize Paper Award in 2014.



Jianfei Cai (S'98–M'02–SM'07) received the Ph.D. degree from the University of Missouri, Columbia, MO, USA.

He is currently an Associate Professor and has served as the Head of Visual and Interactive Computing Division and the Head of Computer Communication Division With the school of Computer Science & Engineering, Nanyang Technological University, Singapore. He has authored over 170 technical papers in international journals and conferences. His current research interests include computer vision, visual computing, and multimedia networking.

Dr. Cai has been actively participating in program committees of various conferences. He has served as the leading Technical Program Chair for the IEEE International Conference on Multimedia and Expo 2012 and the leading General Chair for the Pacific-rim Conference on Multimedia 2012. Since 2013, he has been serving as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING. He has also served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2006 to 2013.