Multisource Domain Attribute Adaptation Based on Adaptive Multikernel Alignment Learning

Xuesong Wang, *Member, IEEE,* Wanwan Huang, Yuhu Cheng, *Member, IEEE*, Qiang Yu, and Zhongliang Wei

***Abstract*—For attribute-based zero-shot learning (ZSL), the attribute classifiers learned previously on the training images may not be usable for the testing images due to that the training and testing images may follow different data distributions. Since domain adaptation learning can effectively perform knowledge transfer under the circumstance of different data distributions, we proposed a novel ZSL method, referred to as multisource domain attribute adaptation based on adaptive multikernel alignment learning (A-MKAL), from the point of view of classifier adaptation. Considering there may be a large difference between object classes, we adopt the clustering method to group the training images according to the class–class correlation measured by the whitened cosine similarity, thus multiple source domains are created. The created multiple source domains are then combined into one weighted source domain to participate in the distribution discrepancy match across domains. In order to adapt the attribute classifier learned on the well-defined source domains to the target domain (the training image set), we designed the A-MKAL by applying the centered kernel alignment to align the attribute kernel matrix and the kernel function of adaptive multiple kernel learning. Experiments on Shoes, OSR, and AWA datasets show that, compared with state-of-the-art methods, our proposed method yields more accurate classification.**

***Index Terms*—Adaptive multikernel learning (A-MKL), attribute adaptation, centered kernel alignment (CKA), multisource domain adaptation, zero-shot learning (ZSL).**

I. Introduction

ATTRIBUTES are deemed as a kind of high-level semantic properties that can be artificially labeled and observed in an image [1]. An object usually has many attributes and based on which, the object class can be recognized. Different objects could have attributes in common. For example, the attribute "spots" could be owned by different objects such as "giraffe," "cow," "dalmatian," "leopard," and "seal." Consequently, by describing commonly shared attributes for all the object classes, attribute knowledge learned previously to new object classes is transferrable, which can not only effectively improve the performance of image retrieval [2]–[4] and activity recognition [5]–[7], but also help resolve the problem of zero-shot learning (ZSL) [8],[9].

ZSL, i.e., zero-shot image classification, is a widely attended research direction in pattern recognition and computer vision. Unlike traditional image classification problems, the object classes emerging at the testing stage do not participate in training progress of the zero-shot image classifier, i.e., training samples are insufficient to cover all the object classes. To transfer attribute knowledge from the known (also known as seen or training) classes to unknown (also known as unseen or testing) classes in the field of zero-shot image classification, it is required to build a bridge connection from low level features to class labels in terms of attributes. At the moment there are a couple of mainstream attribute-based classification approaches for ZSL, namely indirect attribute prediction (IAP) and direct attribute prediction (DAP) [10]. The main working flow of IAP and DAP can be described as the following two stages [10], [11]. At the training stage, a multiclass object classifier or multiple binary attribute classifiers are learned according to the samples of seen classes. At the testing stage, the learned object classifier or attribute classifiers are directly applied to predict the classes or attributes of unseen images. Since the seen and unseen images may follow different data distributions, this kind of "hard knowledge transfer" without any adjustment will inevitably lead to domain shift problems. Researchers have found that across classes, one single attribute varies greatly [11]. For example, although the attribute "spots" is shared by cow and leopard, the appearance of attribute *spots* including the color, size, shape, and location on the body may show large discrepancy. Thus, the attribute classifier for *spots* learned on cow may not be usable on leopard.

Transfer learning is a machine learning method for cross-domain learning, with the aim to seek the useful knowledge from one or more source domains related to a target domain and use it in helping the target task learning. As a typical transfer learning technique, domain adaptation (DA) learning [12], [13] can effectively perform knowledge transfer under the circumstance of different data distributions. In DA learning, the relationship between tasks and the past learning experience is used to accelerate the learning of new

Manuscript received September 19, 2017; revised November 28, 2017; accepted December 27, 2017. Date of publication January 25, 2018; date of current version April 15, 2020. This work was supported by the National Natural Science Foundation of China under Grant 61772532 and Grant 61472424. This paper was recommended by Associate Editor Q. Wang. *(Corresponding author: Yuhu Cheng.)*

X. Wang, W. Huang, and Y. Cheng are with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: wangxuesongcumt@163.com; 15252011972@163.com; chengyuhu@163.com).

Q. Yu is with the School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: yuqiangcumt@163.com).

Z. Wei is with the College of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 230001, China (e-mail: zhlwei@aust.edu.cn).

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/TSMC.2018.2791603

tasks, which can minimize the effects of distribution discrepancy to a great extent. Therefore, DA is an effective solution to resolve the domain shift problem in attribute-based ZSL. Han *et al.* [14] proposed a method named image attribute adaptation (IAA), which aims to adapt attribute knowledge from a source image set to a target image set. In such a way, appropriate attributes for novel images in a target domain are predicted. In the framework of IAA [14], a nonlinear mapping function of multiple base kernels to map every training image from both source and target sets to a reproducing kernel Hilbert space (RKHS), where the cross-domain distribution discrepancy is reduced via maximum mean discrepancy (MMD) [15]–[17]. Most ZSL reports make use of attributes as semantic embedding space. Or alternatively, the word vector serves as another promising embedding space for ZSL [18]–[20]. By exploiting simultaneously both types of semantic embedding spaces, Kodirov *et al.* [21] proposed an unsupervised DA (UDA) method to solve the domain shift problem in ZSL. Specifically, a regularized sparse coding framework is formulated that uses the target domain class labels' projections in the semantic space (attribute space or word vector space) to regularize the learned target domain projection and thus the projection domain shift problem is solved effectively [20]. Similarly, Ji *et al.* [22] proposed a manifold regularized cross-modal embedding DA approach MCME-DA for ZSL. It can be observed that except the attribute annotation for images, both UDA and MCME-DA additionally require word vector description stemmed from a linguistic knowledge database, e.g., Wikipedia.

In summary, the aforementioned IAA, UDA, and MCMC-DA belong to feature representation adaptation, which aims to search a new feature representation where the shared features between source and target domains are enhanced and independent features are weakened. Except feature representation adaptation, classifier adaptation is another important technique in the field of DA learning. Yang *et al.* [23] have shown that adaptive support vector machine (A-SVM) allows for a faster learning of concepts of a new domain, compared with starting from zero or applying a classifier from old domain directly. Through A-SVM, Liu and Kovashka [24] proposed a method for adapting an attribute classifier learned on one domain to another, which includes five different versions such as "Adapt," "Adapt-Simi," "Adapt-Disc," "Adapt-Both," and "Adapt-Rand." Apparently, the defects of A-SVM are twofold: 1) the perturbation function is based on one single kernel and 2) the unlabeled samples in the target domain cannot be utilized. Fortunately, the adaptive multiple kernel learning (A-MKL) framework proposed by Duan *et al.* [25] can address the above problems. In order to better select suitable kernel functions for A-MKL, we proposed a novel adaptive multiple kernel alignment learning (A-MKAL) by introducing the centered kernel alignment (CKA) and further used it to adapt the attribute classifier.

Compared with the feature representation adaptation, the classifier adaptation is much simpler. Therefore, we proposed a novel method to attain the goal of attribute adaptation from the point of view of classifier adaptation in this paper. Moreover, the existing attribute adaptation methods including the aforementioned IAA, UDA, MCMC-DA, and A-SVM-based methods belong to single-source transfer, i.e., all seen images from different object classes are viewed as a single one source domain. This kind of single-source domain processing manner is simple, but the difference between classes is ignored. If we simplistically force all seen images to be treated as a source domain, the negative transfer will inevitably occur when there is a large difference between object classes and thus the zero-shot image classification accuracy is affected. Therefore, a novel ZSL method is proposed, which is here referred to as multisource domain attribute adaptation based on A-MKAL (MDAA+A-MKAL), in this paper.

More specifically, our main contributions are outlined as the following four aspects.

1) We presented a ZSL method (MDAA+A-MKAL) by viewing the seen and unseen image sets as the source and target domains, respectively, and further by using the classifier adaptation technique to adapt the attribute classifiers learned previously on the seen image sets.
2) According to the criterion of class–class correlations, we proposed a multisource domain creation strategy for the seen images through clustering. The class–class correlation is measured by the whitened cosine similarity (WCS) based on known class-attribute matrix.
3) We proposed a domain weighting method based on the relevance probability between attribute and domain. Through domain weighting, the created multiple source domains are combined into one weighted source domain so as to participate in the distribution discrepancy match across domains.
4) We designed an adaptive attribute classifier (A-MKAL) by applying the CKA criterion to align the attribute kernel matrix and the kernel function of A-MKL. The objective function of A-MKAL consists of three parts: a) kernel function metric; b) domain distribution discrepancy metric; and c) classifier loss function.

This paper is arranged as follows. The proposed MDAA+A-MKAL is described in Section II. Experiments on three datasets are reported in Section III, followed by the conclusion in Section IV.

II. Proposed MDAA+A-MKAL

A. Framework of MDAA+A-MKAL

The proposed MDAA+A-MKAL approach mainly consists of six stages, as shown in Fig. 1. Stage I is the creation of source domain. It is a much convenience and simple way to view each seen object class as a source domain, and thus multiple source domains can be created. However, this kind of creation manner has the following twofold defects.

1) If there are more seen object classes, there are more attribute classifiers based on the trainings samples from each seen class, causing heavy workload.
2) The correlation between object classes is ignored.

Therefore, aiming at the above defects, we adopt the clustering method to group the images from seen classes according to the class–class correlation. Here, the class–class correlation is measured through the WCS based on the known class-attribute

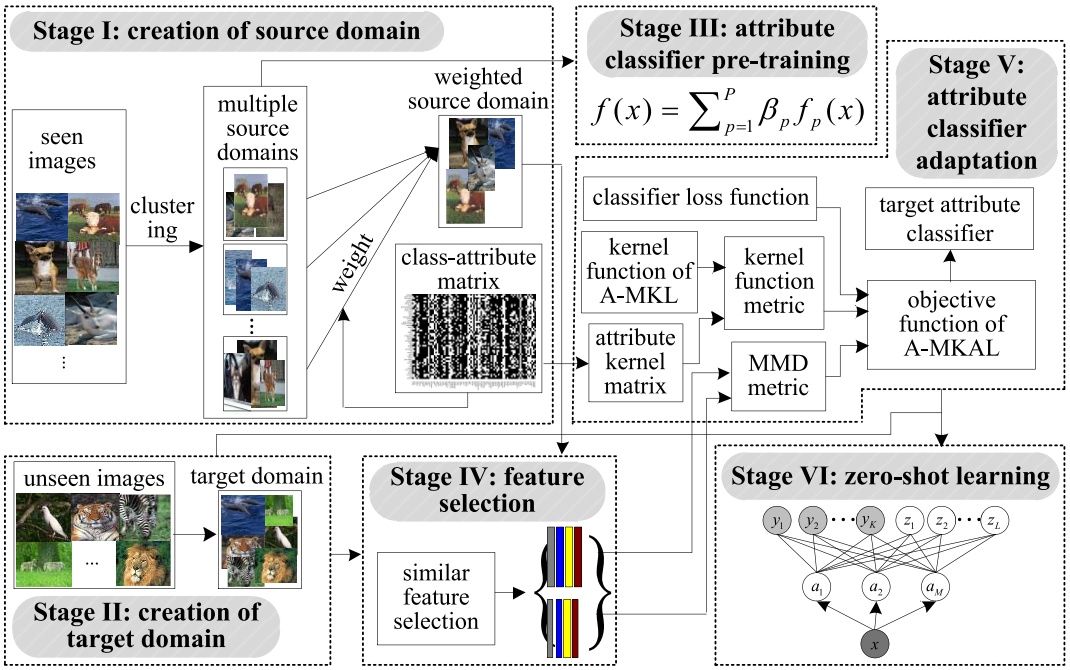


Fig. 1. Framework of the proposed MMDA+A-MKAL.

matrix. Thus, multiple source domains can be obtained through merging similar object classes. In order to perform the subsequent feature selection, a part of samples are further selected from each source domain through domain weighting to generate a weighted source domain. Stage II is the creation of target domain. According to the setting defined by Han *et al.* [14], we manually label some target images with attributes to form the set of labeled target training images. Thus, the created target domain contains a small number of attribute-labeled images and a large number of unlabeled images. Stage III is the attribute classifier pretraining. On each source domain created in stage II, an attribute classifier $f_p(x)$ for each source domain is trained. Then, multiple attribute classifiers are linearly combined to obtain $f(x) = \sum_{p=1}^{P} \beta_p f_p(x)$, where P is the number of source domains and β_p is the domain combination coefficient that needs to be learned in stage V. Stage IV is the feature selection. Although there is distribution discrepancy between the source and target domains, there are some similarities. In order to reduce the distribution discrepancy across domains to some extent and further better perform the attribute transfer from the source to target domain, it is necessary to select similar features [26] across domains before matching the distribution discrepancy. Stage V is the attribute classifier adaptation. In order to achieve better classification performance, we adopt A-MKAL to linearly combine multiple base kernels and thus to model the perturbation function. The final objective function of A-MKAL consists three parts: 1) kernel function metric; 2) domain distribution discrepancy metric; and 3) classifier loss function. The kernel function metric such as CKA aims to obtain the optimal combination coefficients of base kernel functions by calculating the alignment between the A-MKL kernel function and the target kernel matrix. The ideal target kernel matrix used in this paper is the attribute kernel matrix that is stemmed from the prior knowledge: class-attribute matrix. Based on the similar features across the source and target domains after feature selection, we employ the MMD metric to reduce the mismatch between the data distributions of two domains. Through solving the objective function of A-MKAL, the perturbation function can be obtained and thus the adaptive multikernel attribute classifier based on kernel alignment as well. Stage VI is ZSL. The conventional DAP or IAP attribute prediction model (DAP model is considered in this paper) can be used to predict the class labels of unseen images. We now elaborate some main steps of the proposed framework in the following sections.

B. Creation of Source Domain

Since similar object classes have attribute sharing characteristics, we employ the attribute knowledge to create multiple source domains. WCS, which reflects pattern correlation, is often applied for pattern recognition stemming from biometric matching [27]. By referring to our method proposed in [28], we try to measure the correlation between object classes.

Let f^{y_i} and f^{y_j}, respectively, be vectors containing certain class information of training classes y_i and y_j. Here f^{y_i} and f^{y_j} can be obtained according to the class-attribute matrix. The row and column of class-attribute matrix denote the object classes and attributes, respectively. If an object class has an attribute, the element of class-attribute matrix is 1, otherwise 0. In fact, f^{y_i} is the row vector of class-attribute matrix. With WCS, the class–class correlation is defined as [27], [28],

$$\delta_{\mathrm{WC}}\left(f^{y_i}, f^{y_j}\right) = \frac{\left(G^T f^{y_i}\right)^T \left(G^T f^{y_j}\right)}{||G^T f^{y_i}|| \cdot ||G^T f^{y_j}||} \tag{1}$$

where G is whitened transformation matrix, i.e., whitened operator, and $||\cdot||$ is norm operator. G is calculated by covariance matrix Σ [28]

$$\Sigma = E\big[(\chi - M_0)(\chi - M_0)^T\big] \tag{2}$$

where $E(\cdot)$ is expectation and $M_0 = E(\chi)$ is the global average. The covariance matrix can be simplified by using the principal component analysis method [28]

$$\Sigma = \Psi \Lambda \Psi^T \tag{3}$$

where Ψ represents the orthogonal matrix with eigenvectors G in column form and Λ is a diagonal matrix with eigen-values of G as elements. Under the action of whitened transformation $G = \Psi \Lambda^{-1/2}$, (1) is now written as [28]

$$\delta_{\mathrm{WC}}\big(f^{y_i}, f^{y_j}\big) = \frac{(f^{y_i})^T \Sigma^{-1} (f^{y_j})^T}{||G^T f^{y_i}|| \cdot ||G^T f^{y_j}||}. \tag{4}$$

According to the obtained class–class correlations, we adopt the hierarchical clustering [29]–[31] to group the seen images. It is known that there are three linkage criteria for the hierarchical clustering method: 1) single-linkage; 2) complete-linkage; and 3) average-linkage. Compared with the former two linkage criteria, the average-linkage can avoid the sensitivity to outlier and reduce the calculation error. Therefore, the average-linkage is used in this paper. At first, we merge pair of the most similar object classes to form an observation set. Then, we employ the average-linkage criterion [31] to calculate the distance (class–class correlation) between sets and merge the nearest two sets of observations into a new set. In this manner, the iteration stops until the distance value between the nearest two sets is higher than a predefined threshold ε, i.e., $\delta_{\mathrm{WC}} \geq \varepsilon$.

An attribute of a certain specification can describe one of the aspects in an object class and classes may have commonly shared attributes. Therefore, a specific attribute may be commonly used by multiple source domains. The degree of importance for one attribute is different for each domain [31]. As an example, for the attribute "four-legged" and for two different domains "hoofed animal" and "marine animal," with the assumption that the weights of attribute four-legged are ω_1 for hoofed animal and ω_2 for marine animal, respectively, we have $\omega_1 > \omega_2$ obviously [32]. By weighting various source domains, the effective training images for the pretraining of attribute classifiers can be selected more reliably. Obviously, there is some relevance between an attribute and a domain. By referring to our method proposed in [32], such relevance is quantified and used to define weight of each attribute.

The proportion of images that have one attribute initially enables the indication, the importance degree for that attribute on classification. The weight of an attribute can be decided by statistical counts on attributes of training images [32]. Therefore, the relevance probability $p(a_m, \mathrm{rel})$ and the irrelevance probability $p(a_m, \mathrm{irrel})$ are used to calculate attribute weights, with [31]

$$p(a_m, \mathrm{rel}) = \frac{\mathrm{Count}\big(a_m = a_m^p \in s_p\big)}{\mathrm{Count}\big(a_m = a_m^p\big)} \tag{5}$$

$$p(a_m, \mathrm{irrel}) = \frac{\mathrm{Count}\big(a_m = a_m^p \notin s_p\big)}{\mathrm{Count}\big(a_m = a_m^p\big)} \tag{6}$$

where a_m^p represents the values of attribute a_m in the pth source domain s_p, $\mathrm{Count}(a_m = a_m^p \in s_p)$ and $\mathrm{Count}(a_m = a_m^p \notin s_p)$ are the numbers of images where the value of a_m is equal to and is not equal to a_m^p in source domain s_p, respectively, and $\mathrm{Count}(a_m = a_m^p)$ represents the total number of images where attribute a_m exists in source domain s_p. For attribute a_m, its weight for domain s_p is defined as [32]

$$\omega_p^{a_m} = \frac{p(a_m, \mathrm{rel})}{p(a_m, \mathrm{irrel})}. \tag{7}$$

The weight denotes the degree of relevance between attribute a_m and domain s_p. A higher $\omega_p^{a_m}$ means the closer relevance between a_m and s_p. In practical usage, we perform normalization on the weight as

$$\hat{\omega}_p^{a_m} = \frac{\omega_p^{a_m}}{\sum_p \omega_p^{a_m}}. \tag{8}$$

For attribute a_m, if we want to select a total of n_A images from P source domains to generate the weighted source domain, the numbers of images selected from the P source domain are $n_A \hat{\omega}_1^{a_m}$, $n_A \hat{\omega}_2^{a_m} \ldots$, and $n_A \hat{\omega}_P^{a_m}$, respectively.

C. Feature Selection

In order to make the weighted source and target domains more close, the feature selection method proposed by Liu and Kovashka [24] is adopted to select similar features across domains for reducing the mismatch of distribution discrepancy across domains. In addition, lower dimensions can reduce computational burden and thus improve learning efficiency.

Let images in the weighted source and target domains contain attribute a_m be D_S^+ and D_T^+, and the images that do not contain attribute a_m be D_S^- and D_T^-. The mean and variance of feature dimension j on D_S^+ and D_T^+ are calculated as follows [24]:

$$\mu\big(j, D_S^+\big) = \frac{1}{\big|D_S^+\big|} \sum_{x_i \in D_S^+} x_{i,j} \tag{9}$$

$$\sigma^2\big(j, D_S^+\big) = \frac{1}{\big|D_S^+\big|} \sum_{x_i \in D_S^+} \big(x_{i,j} - \mu\big(j, D_S^+\big)\big)^2 \tag{10}$$

where $x_{i,j}$ denotes the jth element of image x_i. The scores s_j^+ and s_j^- used to measure the similarity of feature dimensions between the weighted source and target domains are defined as [24]

$$s_j^+ = \frac{\big|\mu\big(j, D_S^+\big) - \mu\big(j, D_T^+\big)\big|^2}{\sigma^2\big(j, D_S^+\big) + \sigma^2\big(j, D_T^+\big)} \tag{11}$$

$$s_j^- = \frac{\big|\mu\big(j, D_S^-\big) - \mu\big(j, D_T^-\big)\big|^2}{\sigma^2\big(j, D_S^-\big) + \sigma^2\big(j, D_T^-\big)}. \tag{12}$$

As in [24], all feature dimensions j with respect to scores s_j^+ and s_j^- separately are ranked in an ascending order, and pick N_s feature dimensions with the smallest ranks. Thus, the most similar feature dimensions between the weighted source and target domains are selected.

D. A-MKL Learning Based on Centered Kernel Alignment

The A-MKL is a kind of classifier that can be used to address the DA issue [25]. In A-MKL, the target attribute classifier is adapted from an existing classifier $f(x)$ trained based on samples from source domains. The target attribute classifier $\hat{f}(x)$ consists of the pretrained classifier $f(x)$ and the perturbation function $\Delta f(x)$, i.e.,

$$\hat{f}(x) = f(x) + \Delta f(x). \tag{13}$$

The kernel function in the perturbation function is the linear combination of multiple base kernel functions, namely

$$K = \sum_{q=1}^{Q} \eta_q K_q(x_i, x_j) \tag{14}$$

where η_q is the linear combination coefficient of base kernel function which satisfies $\eta_q \geq 0$ and $\sum_{q=1}^{Q} \eta_q = 1$, each base kernel function $K_q(x_i, x_j)$ is induced from the nonlinear feature mapping function $\varphi_q(x)$, i.e., $K_q(x_i, x_j) = \varphi_q(x_i)\varphi_q(x_j)$, where Q denotes the total number of base kernel functions.

Let β_p be the linear combination coefficient of pretrained attribute classifier $f_p(x)$, w_q be the model parameter, and b be the bias item, the decision function of A-MKL is defined as follows [25]:

$$\hat{f}(x) = \sum_{p=1}^{P} \beta_p f_p(x) + \underbrace{\sum_{q=1}^{Q} \eta_q w_q^T \varphi_q(x) + b}_{\Delta f(x)}. \tag{15}$$

In A-MKL, the MMD criterion is adopted to measure the mismatch between the data distributions of two domains. The minimized distribution discrepancy between the source domain D^A and the target domain D^T is written as

$$\mathrm{MMD}\left(D^A, D^T\right) = \left\| \frac{1}{n_A} \sum_{i=1}^{n_A} \varphi\left(x_i^A\right) - \frac{1}{n_T} \sum_{i=1}^{n_T} \varphi\left(x_i^T\right) \right\|_H \tag{16}$$

where x_i^A and x_i^T are images in source and target domains, respectively, n_A and n_T are the total numbers of images, respectively, and $\|\cdot\|_H$ denotes the L_2-norm in an RKHS. Since $K_q(x_i, x_j) = \varphi_q(x_i)\varphi_q(x_j)$, the square of MMD can be written as

$$\mathrm{MMD}^2\left(D^A, D^T\right) = \Omega^2(\eta) = h^T \eta \tag{17}$$

where $\eta = [\eta_1, \ldots, \eta_Q]$ and $h = [\mathrm{tr}(K_1\eta), \ldots, \mathrm{tr}(K_Q\eta)]^T$. The objective function of A-MKL can be expressed as

$$\min_{\eta} G(\eta) = \min_{\eta} \left(\frac{1}{2}\Omega^2(\eta) + \theta J(\eta) \right) \tag{18}$$

where θ is the balancing parameter used to balance the MMD metric and the classifier loss function and

$$\begin{aligned} \min_{\eta} \ J(\eta) = \min_{w_q, \beta, b, \xi_i} & \left(\frac{1}{2} \left(\sum_{q=1}^{Q} \eta_q \|w_q\|^2 + \lambda \|\beta\|^2 \right) \right. \\ & \left. + \ C \sum_{i=1}^{n_A+n_T} \xi_i \right) \\ \text{s.t.} \ \ y_i \hat{f}(x_i) \geq 1 - \xi_i, \xi_i \geq 0 & \end{aligned} \tag{19}$$

Algorithm 1 MMDA+A-MKAL

Input: the training images x_i^A, the testing images x_i^T, the training object classes y_i, the testing object classes z_k, the image attributes a_m, the class-attribute matrix, the base kernel functions K_q, the control parameters including ε, N_s, n_A, n_T, Q, θ, λ, and C.
1. Calculate the class-class correlation $\delta_{\mathrm{WC}}(f^{y_i}, f^{y_j})$ through Eq. (4) according to the class-attribute matrix and obtain P source domains: $s_1, s_2, \ldots, s_P$.
2. Train the attribute classifier $f_p(x)$ on each source domain.
3. Calculate the weight of attribute a_m for source domain s_p through Eq. (8) according to the class-attribute matrix, thus the weighted source domain is created.
4. Manually label some testing images with attributes to construct the target domain.
5. Select N_s similar features across the weighted source and target domains by using Eqs. (11)-(12).
6. Construct the kernel function K through Eq. (14) according to multiple predefined base kernel functions K_q.
7. Solve Eq. (22) to obtain parameters η, β_p, w_q and b.
8. Obtain the target attribute classifier $\hat{f}(x)$ by using Eq. (20).
9. Assign the best output class from all testing classes to a testing image x_i^T by using DAP or IAP model based on the learned target attribute classifier.
Output: The predicted class labels for testing images.

where $\sum_{i=1}^{n_A+n_T} \xi_i$ denotes the total classifying error of classifier $\hat{f}(x)$, $\beta = [\beta_1, \ldots, \beta_p]$, λ is the regularization parameter, and C is the penalty factor.

By introducing Lagrange multiple $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_{n_A+n_T}]$, the target attribute classifier is expressed as

$$\hat{f}(x) = \sum_{i=1}^{n_A+n_T} \alpha_i y_i \left(\sum_{q=1}^{Q} \eta_q K_q(x_i, x) + \frac{1}{\lambda} f^T(x_i) f(x) \right) + b. \tag{20}$$

Each target attribute classifier is obtained through individual training of each A-MKL. In order to take full advantage of the known attribute knowledge and thus to improve A-MKL performance, the CKA [33], [34] is used to learn the optimal kernel function of A-MKL. Each image has multiple attributes that may be highly correlated. Suppose $L_m = \langle a_m, a_{m'} \rangle$ be the kernel matrix over the attributes, the alignment between attribute kernel matrix and kernel function of A-MKL is defined by [32]

$$\rho(K, L) = \frac{\mathrm{tr}(KL)}{\sqrt{\mathrm{tr}(LL)}\sqrt{\mathrm{tr}(KK)}}. \tag{21}$$

Since $\sqrt{\mathrm{tr}(LL)}$ is not associated with parameter η, $\sqrt{\mathrm{tr}(LL)}$ can be ignored during the optimization solving process. Moreover, since the roles of $\sqrt{KK}$ and $\Omega(\eta)$ are duplicate to some extent, $\sqrt{KK}$ also can be excluded from the process. Therefore, by combining (18) and (21), the final objective function of A-MKAL can be written as

$$\min_{\eta} G(\eta) = \min_{\eta} \left(\frac{1}{2}(\Omega(\eta) - \mathrm{tr}(KL))^2 + \theta J(\eta) \right). \tag{22}$$

E. Algorithm Steps

In summary, the detailed steps of the proposed MMDA+A-MKAL are summarized as Algorithm 1.

III. Experiments

A. Description of Datasets

Datasets including Shoes [35], OSR [36], and AWA [10] are adopted to evaluate the performance of our proposed MMDA+A-MKL. These datasets cover three distinct fields of objects, landscape, and animals, which can fully verify the suitability of MMDA+A-MKL to different application fields. Each image in the Shoes dataset is described by two different feature types: 1) a 960-dimensional gist descriptor and 30-dimensional laboratory color histogram. For the OSR dataset, 512-dimensional gist descriptor is applied as image features. According to [21], [22], and [24], we adopt the 4096-D deep learning (DL) features named DeCAF [37], [38] as the image feature for the AWA dataset since these features are already available online for comparison. In our experiments, there are total four base kernel functions are used, which are linear kernel, Gaussian kernel, polynomial kernel, and sigmoid kernel. All the experiments have been performed in MATLAB R2014a using a computer with a dual Intel Xeon E5-2660 V3, 48-GB memory, and a 12-GB NVIDIA Tesla K40c graphic card.

B. Parameter Settings

As noted in Section II, control parameters of the proposed MMDA+A-MKL include the threshold ε and the dimension of similar features between the weighted source and target domains. In this section, we will analyze the effects of the two control parameters on the performance of MMDA+A-MKL and provide a general guidance for setting such parameters.

First, we study the relationship between threshold and classification performance. For MMDA+A-MKL, the threshold directly determines the number of clustering and further that of source domains. The experiment is designed as follows: $N_s = 0$; for the Shoes dataset, six object classes including "athletic-shoes," "boots," "clogs," "high_heels," "rain_boots," and "sneakers" are selected as training classes, with the remaining four classes as testing classes. The size of training and testing images are 6000 and 4000, respectively; for the OSR dataset, we choose "tallbuilding," "insidecity," "street," and "mountain" to constitute the training classes. The number of training and testing images are 1330 and 1358, respectively; ten object classes including "chimpanzee," "giant panda," "hippopotamus," "humpback+whale," "leopard," "pig," "raccoon', "rat," "seal," and "persian+cat" are selected from the AWA dataset as testing classes, with the remaining 40 classes as training classes. There are total 4000 and 1000 images in the training and testing image sets, respectively. Fig. 2 shows the clustering binary trees on different datasets where the horizontal axis represents all object classes in the source domain, i.e., training classes. Because of the limited space, we use numbers to represent the object classes in AWA. The numbers on the horizontal axis in Fig. 2(c) are consistent with the serial numbers of object classes in AWA. For example, "1" and "40" denote object classes "antelope" and "deer," respectively. The vertical axis represents the similarity between object classes, i.e., δ_{WC}. Take the Shoes dataset as an example, Fig. 2(a) shows the result of clustering binary tree. When the threshold ε is 0.5, you can draw a line parallel to the horizontal axis at the vertical axis of 0.5, thus the number of intersections by this line and the binary tree means the obtained number of source domains.

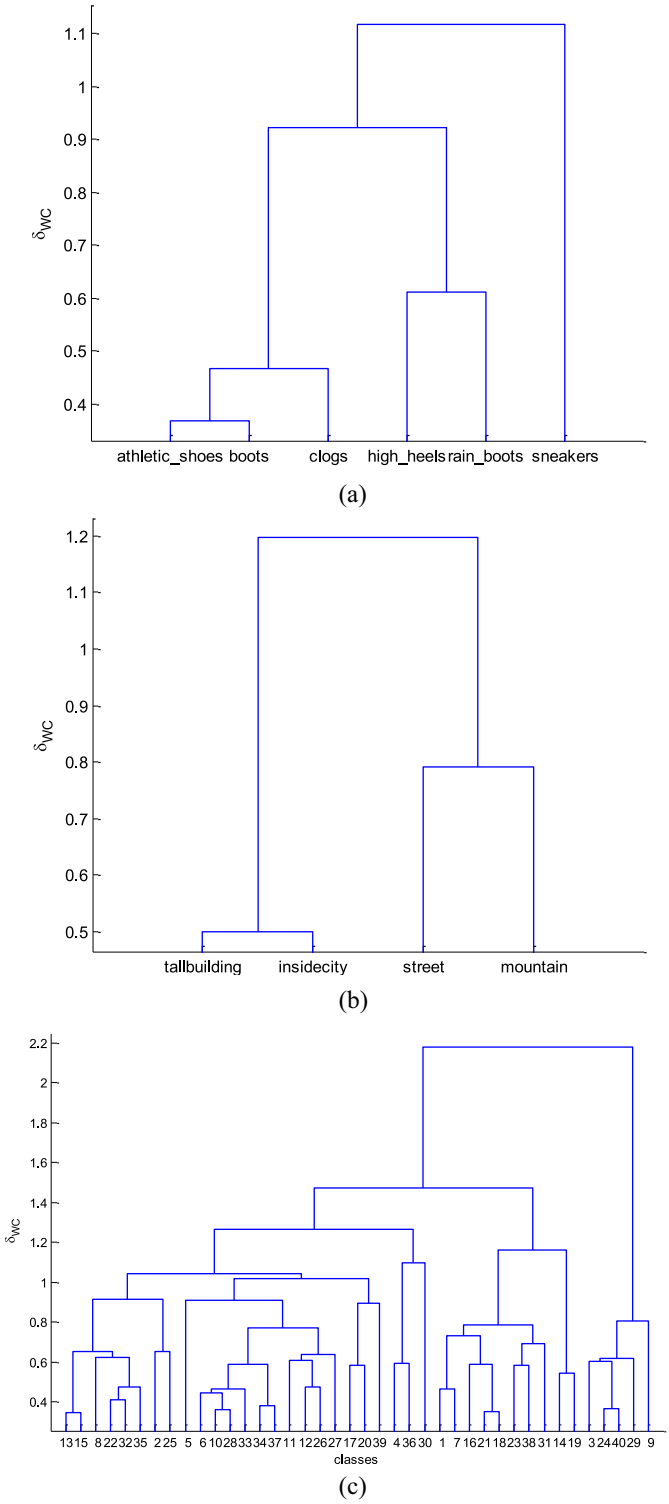


Fig. 2. Clustering binary tree. (a) Shoes dataset. (b) OSR dataset. (c) AWA dataset.

Classification accuracy [in terms of acc (%)] on the testing classes and computation duration [in terms of time (s)] by ZSL

TABLE I
INFLUENCE OF ε ON CLASSIFICATION PERFORMANCE

Dataset	ε=0			ε=0.5			ε=0.8		
	P	acc (%)	time (s)	P	acc (%)	time (s)	P	acc (%)	time(s)
Shoes	6	37.54	447.64	4	39.86	435.67	3	**41.82**	421.72
OSR	4	38.25	282.43	3	42.41	364.71	2	**44.36**	249.34
AWA	40	23.51	954.32	27	24.72	808.55	10	28.30	624.17
Dataset	ε=1.2			ε=1.6			ε=2.2		
	P	acc (%)	time (s)	P	acc (%)	time (s)	P	acc (%)	time (s)
Shoes	1	35.59	404.48	1	35.59	404.48	1	35.59	404.48
OSR	1	37.80	199.51	1	37.80	199.51	1	37.80	199.51
AWA	4	**30.12**	411.25	2	25.37	386.70	1	22.67	379.52

TABLE II
INFLUENCE OF N_S ON CLASSIFICATION PERFORMANCE

Dataset	N_s=200		N_s=400		N_s=600		N_s=800		N_s=960	
	acc (%)	time (s)	acc (%)	time (s)	acc (%)	time (s)	acc (%)	time (s)	acc (%)	time (s)
Shoes	31.97	362.03	35.15	379.74	38.04	398.55	**46.21**	405.02	41.82	421.72
Dataset	N_s=100		N_s=200		N_s=300		N_s=400		N_s=512	
	acc (%)	time (s)	acc (%)	time (s)	acc (%)	time (s)	acc (%)	time (s)	acc (%)	time (s)
OSR	26.26	185.04	33.64	190.71	40.07	199.38	**47.14**	241.28	44.36	249.34
Dataset	N_s=500		N_s=1000		N_s=1200		N_s=1500		N_s=2000	
	acc (%)	time (s)	acc (%)	time (s)	acc (%)	time (s)	acc (%)	time (s)	acc (%)	time (s)
AWA	32.43	242.73	35.21	268.49	38.27	280.42	**43.31**	295.36	38.03	319.26
Dataset	N_s=3000		N_s=4096							
	acc (%)	time (s)	acc (%)	time (s)						
AWA	34.54	369.17	30.12	411.25						

TABLE III
COMPARISON OF CLASSIFICATION ACCURACY (%) ON TESTING CLASSES (SHOES DATASET)

N/*O*	8/2	7/3	6/4	5/5	4/6	3/7	2/8	avg.
DAP [10]	41.75	32.55	27.16	19.07	15.87	12.85	11.91	23.02
IAA [14]	45.31	42.54	40.62	35.81	24.26	17.65	10.27	30.92
UDA [21]	48.17	37.92	31.05	22.47	17.58	13.82	12.23	26.18
MCME-DA [22]	50.10	35.43	28.28	19.22	16.77	14.38	12.60	25.25
Adapt-Simi [24]	47.72	44.14	41.36	36.55	31.75	21.05	16.01	34.08
MDAA+A-MKL	48.10	44.47	42.05	37.46	32.07	21.26	16.24	34.52
MDAA+A-MKAL	**51.46**	**47.32**	**44.57**	**42.24**	**35.16**	**30.52**	**21.32**	**38.94**

TABLE IV
COMPARISON OF CLASSIFICATION ACCURACY (%) ON TESTING CLASSES (OSR DATASET)

N/*O*	6/2	5/3	4/4	3/5	2/6	avg.
DAP [10]	38.04	20.51	19.25	17.78	15.79	22.27
IAA [14]	66.21	52.14	32.50	25.40	19.94	39.24
UDA [21]	60.55	41.76	25.17	18.69	15.21	32.28
MCME-DA [22]	62.81	41.31	26.73	20.29	16.97	33.62
Adapt-Simi [24]	67.87	60.48	33.42	26.30	17.42	41.10
MDAA+A-MKL	67.95	60.72	33.64	26.47	17.73	41.30
MDAA+A-MKAL	**71.27**	**61.26**	**35.48**	**27.70**	**19.38**	**43.02**

under different thresholds are shown in Table I, where P is the number of source domains. It can be observed from Fig. 3 and Table I as follows.

1) As ε increases, all classification accuracies on three datasets increase first and then decrease.
2) When the threshold is lower enough, for example $\varepsilon = 0$, the number of source domains equals that of training classes. Since the class–class correlation is not taken into account, the classification accuracy in this case is lower.
3) When the threshold is higher enough, for example $\varepsilon = 2.2$, all training classes are clustered into one source domain, i.e., $P = 1$. Thus, the weighted source domain is created by randomly selecting n_A images from clustered source domain and these n_A images may not cover all training classes. In addition, the domain combination has not been implemented in the process of attribute classifier pretraining due to $P = 1$. Therefore, the classification accuracy in this case is the lowest, which also illustrates inferiority of single source attribute adaptation.
4) The more number of source domains, the more attribute classifiers that need pretraining. Therefore, the computation amount increases inversely with ε.

TABLE V
COMPARISONS OF CLASSIFICATION ACCURACY (%) ON TESTING CLASSES (AWA DATASET)

N/O	45/5	44/6	43/7	42/8	41/9	40/10	avg.
DAP [10]	50.05	46.28	45.01	42.44	41.2	32.42	42.90
IAA [14]	54.62	51.91	46.34	41.50	39.53	32.64	44.42
UDA [21]	58.02	53.50	50.16	44.32	41.54	37.85	47.57
MCME-DA [22]	58.71	56.2	53.42	44.37	37.96	31.12	46.96
Adapt-Simi [24]	57.90	54.43	52.74	47.35	44.25	38.04	49.12
MDAA+A-MKL	58.47	54.77	53.05	47.56	44.54	38.72	49.52
MDAA+A-MKAL	**61.05**	**57.81**	**54.79**	**50.22**	**46.36**	**41.01**	**51.87**

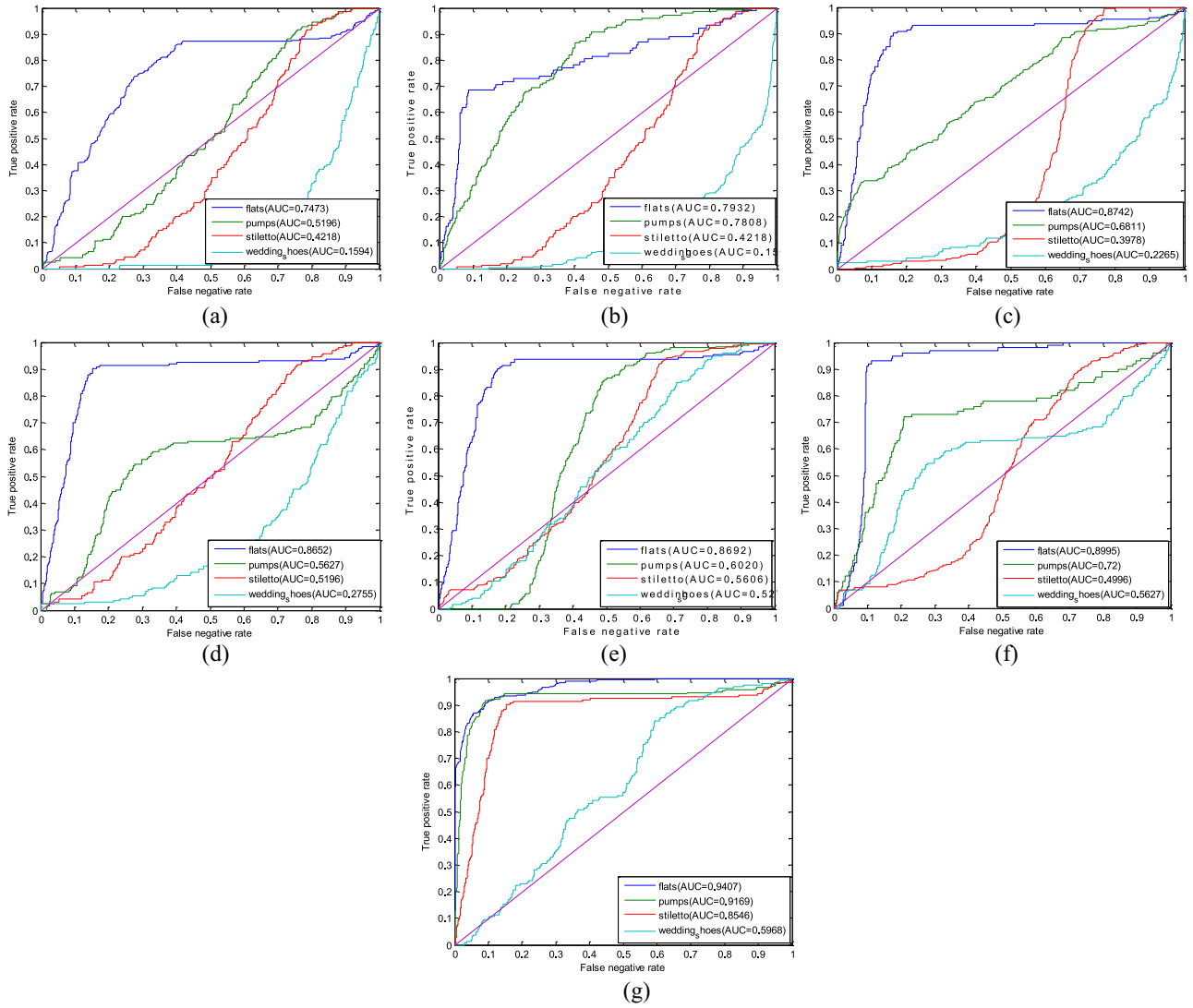


Fig. 3. ROC curves and AUC values (Shoes dataset). (a) DAP. (b) UDA. (c) MCME-DA. (d) IAA. (e) Adapt-Simi. (f) MDAA+A-MKL. (g) MDAA+A-MKAL.

Based on the above analysis, in the following experiments, we set $\varepsilon = 0.8$ for Shoes and OSR datasets and $\varepsilon = 1.2$ for AWA dataset.

Then we study the influence of the dimension of similar features on the classification performance. Table II shows the classification accuracy and computation duration with variation of N_s. It is observed as follows.

1) As N_s increases, all classification accuracies on three datasets rises up first and then decrease.
2) Whether N_s is lower or higher enough, the classification accuracies on all datasets are relatively low. The reasons are as follows: many similar features between the weighted source and target domains are not selected when N_s is low, which will lead to information loss. On the other hand, the selected similar features are redundant when N_s is high.
3) The dimension rise of similar domains extends the duration of attribute classifier adaptation progress. Therefore, computation increases accordingly with N_s.

Based on the above analysis, in the following experiments, we set $N_s = 800$, $N_s = 400$, and $N_s = 1500$ for Shoes, OSR, and AWA datasets, respectively.

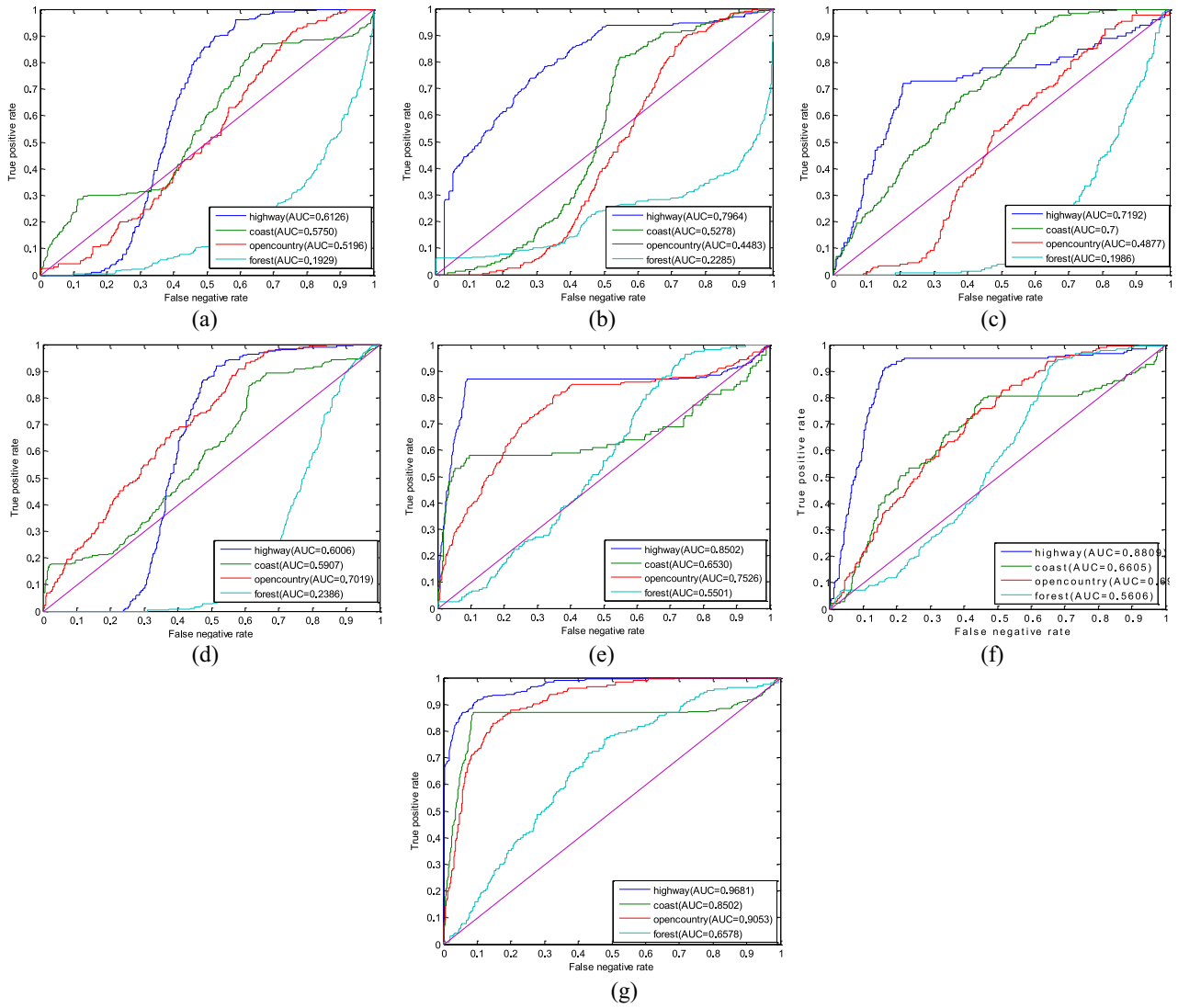


Fig. 4. ROC curves and AUC values (OSR dataset). (a) DAP. (b) UDA. (c) MCME-DA. (d) IAA. (e) Adapt-Simi. (f) MDAA+A-MKL. (g) MDAA+A-MKAL.

C. Zero-Shot Image Classification

In order to evaluate MDAA+A-MKAL performance in the zero-shot image classification, the following six methods for comparison are studied: DAP [10], IAA [14], UDA [21], MCME-DA [22], Adapt-Simi [24], and MDAA+A-MKL. MDAA+A-MKL is a special case of MDAA+A-MKAL, i.e., the CKA is not adopted in MDAA+A-MKAL.

According to the experimental procedure shown in [9] and [39], we study the relation between number of testing classes and classification performance of the zero-shot image. Specifically in each group of experiment, different numbers of training classes and testing classes are selected out. Take the Shoes dataset as an example, Table III illustrates that the total seven groups of experiments were taken. To reduce the effect of random factors on classification performance, C^O_{N+O}-fold zero-shot classification experiments are taken for each group of experiment, where N and O represent the numbers of training and testing classes respectively. According to the mentioned experimental setup, seven times of C^O_{N+O}-fold (C^2_{10} = 45-fold, C^3_{10} = 120-fold, C^4_{10} = 210-fold, C^5_{10} = 252-fold, C^6_{10} = 210-fold, C^7_{10} = 120-fold, and C^8_{10} = 45-fold) experiments were carried out on the Shoes dataset. Due to high number of animal classes, O classes from the ten predefined testing classes in Section III-B per experiment are picked up only. In other words, six times of C^O_{10}-fold (C^5_{10} = 252-fold, C^6_{10} = 210-fold, C^7_{10} = 120-fold, C^8_{10} = 45-fold, C^9_{10} = 10-fold, and C^{10}_{10} = 1-fold) experiments were carried out on the AWA dataset.

Tables III–V comparatively list classification accuracy on these testing classes on different datasets, where "avg." denotes the average accuracy of all groups of experiments for each method. It can be observed as follows.

1) No matter what the number of testing classes is, the classification accuracies of MDAA+A-MKAL achieve the highest among the seven approaches.
2) Compared with DAP, all the attribute adaptation methods such as IAA, UDA, MCME-DA, Adapt-Simi, MDAA+A-MKL, and MDAA+A-MKAL, yield higher average classification accuracies on the three datasets,

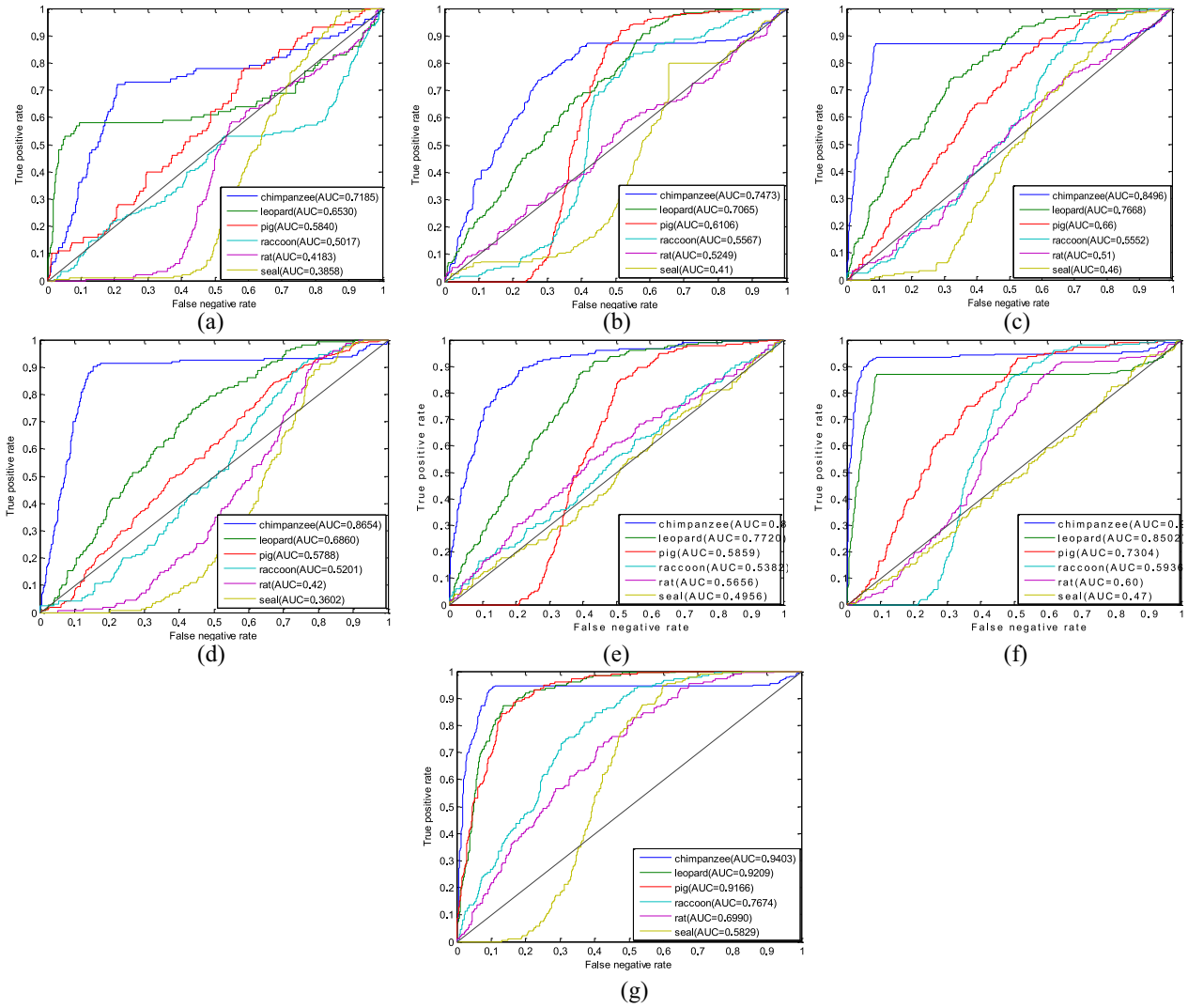


Fig. 5. ROC curves and AUC values (AWA dataset). (a) DAP. (b) UDA. (c) MCME-DA. (d) IAA. (e) Adapt-Simi. (f) MDAA+A-MKL. (g) MDAA+A-MKAL.

which illustrates the adapted attribute classifiers can indeed improve the performance of attribute-based ZSL.

3) Compared with single source attribute adaptation methods including IAA, UDA, MCME-DA, and Adapt-Simi, our proposed MDAA+A-MKL and MDAA+A-MKAL yield higher average classification accuracies on the three datasets. This is because multisource transfer learning technique adopted in our method can avoid the negative transfer to some extent.
4) Since the kernel function of A-MKL and the attribute kernel matrix are not aligned, the average classification accuracies of MDAA+A-MKL of all the three datasets are lower than that of MDAA+A-MKAL.

The classification accuracy will reveal the relation between the correctly classified number of testing images and the total number of testing images; however, the relation between specificity (false negative rate) and sensitivity (true positive rate) cannot be reflected. Therefore the receiver operator characteristic (ROC) curve and the area under curve (AUC) value are further introduced in experiment to evaluate the zero-shot image classification performance. Figs. 3–5 present the ROC curves and AUC values, illustrating the following observations.

1) All ROC curves of MDAA+A-MKAL are closer to the left corner of the coordinator and AUC values are the highest among the seven methods, which indicates better performance of MDAA+A-MKAL on zero-shot image classification.
2) Except MDAA+A-MKAL, the AUC values of the other six approaches in comparison on some testing classes are below 0.5, meaning that the six methods on ZSL are in some cases interior even to a random predictive model.

IV. Conclusion

ZSL technique aims at object classification problems that no training images are offered for testing classes. For ZSL, attribute transfer serves as a central role in recognizing testing images. The commonly used ZSL methods directly apply the attribute classifiers learned on training images to predict the attributes of training images without any adjustment. Since the training and testing images may follow different data

distributions, the domain shift problem will inevitably occur. By using the DA learning technique, we proposed a novel ZSL method called MMDA+A-MKAL. The main advantages are as follows.

1) The manner of viewing the training images as multiple source domains simultaneously considers the similarity and difference between classes, which can effectively avoid the negative transfer problem compared with single-source transfer.
2) Selecting similar features from the source and target domains is helpful not only for reducing the mismatch of distribution discrepancy across domains, but also for improving learning efficiency.
3) The usage of CKA is helpful for learning the optimal kernel function of A-MKL through aligning the ideal attribute kernel matrix and the kernel function. Thus, the designed A-MKAL can better adapt the attribute classifier learned on the source domain to the target domain and help improving the performance of attribute prediction and zero-shot recognition of testing images.

For MMDA+A-MKAL, it requires that manually label some testing images with attributes to construct the target domain. If the labeled attribute is wrong, the zero-shot classification accuracy will be affected. Feature representation adaptation does not require a partial attribute-labeled sample in the target domain. Recently, DL [40] as an attractive research in machine learning field, is found to be able to automatically learn representative features from data via stacking multilayer nonlinear units. As our future work, we plan to study how to apply DL-based feature representation adaptation to address the attribute adaptation in ZSL. In addition, the source and target domain images share the same view in this paper. Currently, image features include color, texture, shape, etc. If we treat different kinds of feature as multiple views of image, then multiview learning [41] can well integrate information provided by each view. Therefore, attribute adaptation based on multiview transfer learning is another future work.

REFERENCES

[1] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1778–1785.

[2] N. Passalis and A. Tefas, "Learning neural bag-of-features for large-scale image retrieval," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 10, pp. 2641–2652, Oct. 2017.

[3] Y. H. Xu *et al.*, "Attribute hashing for zero-shot image retrieval," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Hong Kong, 2017, pp. 133–138.

[4] W. Lei and G. Q. Xiao, "Image retrieval using two-dimensional inverted index and semantic attributes," in *Proc. IEEE Int. Conf. Softw. Eng. Service Sci.*, Beijing, China, 2016, pp. 679–682.

[5] J. Wannenburg and R. Malekian, "Physical activity recognition from smartphone accelerometer data for user context awareness sensing," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 12, pp. 3142–3149, Dec. 2017.

[6] J. L. Yu and J. F. Sun, "Multiactivity 3-D human pose tracking in incorporated motion model with transition bridges," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2017.2664891.

[7] J. Liu, B. Kuipers, and S. Savarese, "Recognizing human actions by attributes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Colorado Springs, CO, USA, 2011, pp. 3337–3344.

[8] C. Z. Luo, Z. T. Li, K. Z. Huang, J. S. Feng, and M. Wang, "Zero-shot learning via attribute regression and class prototype rectification," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 637–648, Feb. 2018.

[9] X. S. Wang, C. Chen, Y. H. Cheng, and Z. Jane Wang, "Zero-shot image classification based on deep feature extraction," *IEEE Trans. Cogn. Develop. Syst.*, to be published, doi: 10.1109/TCDS.2016.2632178.

[10] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, 2009, pp. 951–958.

[11] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 453–465, Mar. 2014.

[12] X. Y. Wang, H. X. Li, Y. Li, F. M. Shen, and F. Porikli, "Robust and real-time deep tracking via multi-scale domain adaptation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Hong Kong, 2017, pp. 1338–1343.

[13] K. Murasaki, H. Yonemoto, and K. Sudo, "Domain adaptation of articulated pose estimation via synthetic pose prior," in *Proc. 15th IAPR Int. Conf. Mach. Vis. Appl.*, Nagoya, Japan, 2017, pp. 137–140.

[14] Y. Han *et al.*, "Image attribute adaptation," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 1115–1126, Jun. 2014.

[15] S. F. Zou, Y. B. Liang, H. V. Poor, and X. H. Shi, "Nonparametric detection of anomalous data streams," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp. 5785–5797, Nov. 2017.

[16] H. Sun, S. Liu, S. L. Zhou, and H. X. Zou, "Transfer sparse subspace analysis for unsupervised cross-view scene model adaptation," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 7, pp. 2901–2909, Jul. 2016.

[17] M. J. Gangeh *et al.*, "Computer aided theragnosis using quantitative ultrasound spectroscopy and maximum mean discrepancy in locally advanced breast cancer," *IEEE Trans. Medical Imag.*, vol. 35, no. 3, pp. 778–790, Mar. 2016.

[18] I. Alexiou, T. Xiang, and S. G. Gong, "Exploring synonyms as context in zero-shot action recognition," in *Proc. IEEE Int. Conf. Image Process.*, Phoenix, AZ, USA, 2016, pp. 4190–4194.

[19] S. Naha and Y. Wang, "Beyond verbs: Understanding actions in videos with text," in *Proc. Int. Conf. Pattern Recognit.*, Cancún, Mexico, 2017, pp. 1833–1838.

[20] S. Reed, Z. Akata, H. L. Lee, and B. Schiele, "Learning deep representations of fine-grained visual descriptions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 49–58.

[21] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, "Unsupervised domain adaptation for zero-shot learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, 2015, pp. 2452–2460.

[22] Z. Ji, Y. Yu, Y. Pang, J. Guo, and Z. Zhang, "Manifold regularized cross-modal embedding for zero-shot learning," *Inf. Sci.*, vol. 378, pp. 48–58, Feb. 2017.

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

[24] S. Q. Liu and A. Kovashka, "Adapting attributes by selecting features similar across domains," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2016, pp. 1–8.

[25] L. X. Duan, D. Xu, I. W.-H. Tsang, and J. B. Luo, "Visual event recognition in videos by learning from Web data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, 2010, pp. 1959–1966.

[26] Q. Wang, J. Z. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.

[27] C. Liu, "The Bayes decision rule induced similarity measures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1086–1090, Jun. 2007.

[28] X. S. Wang, C. Chen, and Y. H. Cheng, "Zero-shot learning by exploiting class-related and attribute-related prior knowledge," *IET Comput. Vis.*, vol. 10, no. 6, pp. 483–492, Sep. 2016.

[29] R. R. Yager, "Intelligent control of the hierarchical agglomerative clustering process," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 30, no. 6, pp. 835–845, Dec. 2000.

[30] S. Sambasivam and N. Theodosopoulos, "Advanced data clustering methods of mining Web documents," *Issues Inf. Sci. Inf. Technol.*, vol. 3, no. 1, pp. 563–579, Jan. 2006.

[31] R. Alvarez, A. Moser, and C. A. Rahmann, "Novel methodology for selecting representative operating points for the TNEP," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2234–2242, May 2017.

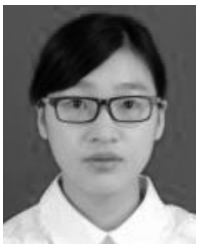
[32] Y. H. Cheng, X. Qiao, and X. S. Wang, "An improved indirect attribute weighted prediction model for zero-shot image classification," *IEICE Trans. Inf. Syst.*, vol. E99, no. 2, pp. 435–442, Feb. 2016.

[33] C. Cortes, M. Mohri, and A. Rostamizadeh, "Algorithms for learning kernels based on centered alignment," *J. Mach. Learn. Res.*, vol. 13, pp. 795–828, Mar. 2012.

[34] A. M. Álvarez-Meza, D. Cárdenas-Peña, and G. Castellanos-Domínguez, "MRI discrimination by inter-slice similarities and kernel-based centered alignment," in *Proc. Int. Work Conf. Bio-Inspired Intell.*, 2014, pp. 149–153.
[35] T. L. Berg, A. C. Berg, and J. Shih, "Automatic attribute discovery and characterization from noisy Web data," in *Proc. Europe Conf. Comput. Vis.*, 2010, pp. 663–676.
[36] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, May/Jun. 2001.
[37] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.
[38] J. Donahue *et al.*, "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.
[39] Y. H. Cheng, X. Qiao, X. S. Wang, and Q. Yu, "Random forest classifier for zero-shot learning based on relative attribute," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2017.2677441.
[40] Y. Liu *et al.*, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018.
[41] X. L. Li, M. L. Chen, F. P. Nie, and Q. Wang, "A multiview-based parameter free framework for group detection," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 2017, pp. 4147–4153.

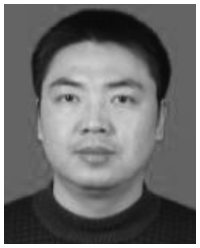
Xuesong Wang (M'15) received the Ph.D. degree in control science and engineering from the China University of Mining and Technology, Xuzhou, China, in 2002.

She is currently a Professor with the School of Information and Control Engineering, China University of Mining and Technology. Her current research interests include machine learning, bioinformatics, and artificial intelligence.



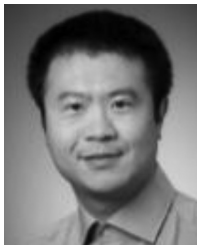
Wanwan Huang received the B.S. degree in automation from the Anhui University of Science and Technology, Huainan, China, in 2015. She is currently pursuing the master's degree with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China

Her current research interest includes zero-shot learning.



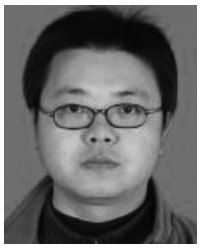
Yuhu Cheng (M'15) received the Ph.D. degree in control science and engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2005.

He is currently a Professor with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China. His current research interests include machine learning and intelligent systems.



Qiang Yu received the Ph.D. degree in electrical engineering from the University of Bundeswehr Muenchen, Munich, Germany, in 2012.

He is currently an Associate Professor with the School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou, China. His current research interests include intelligent learning and systems.



Zhongliang Wei received the M.S. degree in computer application from the Anhui University of Science and Technology, Huainan, China, in 2011.

He is currently a Lecturer with the College of Computer Science and Engineering, Anhui University of Science and Technology. His current research interests include data mining and machine learning.