Computing and Informatics, Vol. 39, 2020, 318–339, doi: 10.31577/cai_2020_1-2_318

DEEP CONVOLUTION AND CORRELATED MANIFOLD EMBEDDED DISTRIBUTION ALIGNMENT FOR FOREST FIRE SMOKE PREDICTION

Yaoli Wang, Xiaohui Liu

College of Information and Computer
Taiyuan University of Technology
Jinzhong, 030600, China
e-mail: wangyaoli@tyut.edu.cn, 1662010618@qq.com

Maozhen Li*

The Key Laboratory of Service Computing and Embedded Systems
Tongji University
Shanghai, China
e-mail: Maozhen.Li@gmail.com

Wenxia Di

Foreign Languages Department
Taiyuan Normal University
Jinzhong, 030600, China
e-mail: wendy_cn@263.net

Lipo Wang

School of Electrical and Electronic Engineering
Nanyang Technological University
639798, Singapore
e-mail: elpwang@ntu.edu.sg

* Corresponding author

Abstract. This paper proposes the deep convolution and correlated manifold embedded distribution alignment (DC-CMEDA) model, which is able to realize the transfer learning classification between and among various small datasets, and greatly shorten the training time. First, pre-trained Resnet50 network is used for feature transfer to extract smoke features because of the difficulty in training small dataset of forest fire smoke; second, a correlated manifold embedded distribution alignment (CMEDA) is proposed to register the smoke features in order to align the input feature distributions of the source and target domains; and finally, a trainable network model is constructed. This model is evaluated in the paper based on satellite remote sensing image and video image datasets. Compared with the deep convolutional integrated long short-term memory (DC-ILSTM) network, DC-CMEDA has increased the accuracy of video images by 1.50 %, and the accuracy of satellite remote sensing images by 4.00 %. Compared the CMEDA algorithm with the ILSTM algorithm, the number of iterations of the former has decreased to 10 times or less, and the algorithm complexity of CMEDA is lower than that of ILSTM. DC-CMEDA has a great advantage in terms of convergence speed. The experimental results show that DC-CMEDA can solve the problem of small sample smoke dataset detection and recognition.

Keywords: Transfer learning, domain adaptation, deep convolution, small dataset, forest fire smoke

1 INTRODUCTION

The research on smoke image detection technology mainly focuses on the classification and recognition of smoke images through deep learning training. This method has high classification accuracy, but a strong dependence on the dataset. It requires that the training and test data meet independent and identical distributions, and it needs enough training samples available. However, smoke images of different scenes and different resolutions cannot meet the same and independent distribution. It is also very expensive, and difficult to obtain many labeled sample images in different scenes. Therefore, this type of method requires training of various models, which is time-consuming and labor-intensive. In addition, there is an overfitting phenomenon in small sample scenario training. With the development of artificial intelligence, smoke detection technology based on transfer learning has a good application prospect. This technology is used to learn and train with large sample smoke image data, and to popularize what is collected in small sample smoke images, which not only reduces the time for model training, but also prevents overfitting.

The existing methods mainly proceed from the perspective of deep transfer learning. Literature [1] proposed a feature migration method based on isomorphic data by taking ImageNet dataset as source data and using VGG16 model, which provided a feasible method for smoke detection and identification; literature [2] used the pre-trained VGG16 network on the ImageNet dataset for effective smoke feature

extraction, and proposed an integrated long-term and short-term memory network, which uses this network to fuse smoke features in segments. Finally, a trainable deep neural network model is constructed, which can be used for forest fire smoke detection. This method uses pre-trained VGG16 model feature extraction, which requires high requirements for sample dataset, and the model has a shallow network depth and insufficient extraction of certain features, which will result in a low accuracy of recognition and classification.

Literature [3] proposed a recursive convolutional neural network based on RNN and successfully applied it in the field of video smoke detection; literature [4] introduced the Inception-v3 network trained on the ImageNet dataset, first remove and reset the last full connection layer in the Inception-v3 network, then freeze all parameters of the convolutional layer and pooling layer in the previous hidden layer, and then use the collected small smoke dataset for training to fine-tune the reset full connection layer. Finally, the smoke detection model of deep transfer learning convolutional neural network is obtained. This method reduces the number of datasets in the process of fine-tuning the fully connected layer, but it is still unable to accurately identify the classification for small sample datasets; and the deep network is complex, with many parameters, and the fine-tuning still takes a long time.

Another popular research area of transfer learning [5] at present is domain adaptation [6], which has a good effect on solving the problem of source domain data calibration target domain data, achieves data classification, and has advantages in the number of parameters and time consumption. The probability distribution adaptation method is mainly carried out from three aspects, namely: edge distribution adaptation, conditional distribution adaptation, and joint distribution adaptation. The earliest application of conditional distribution adaptation [7] to transfer learning was achieved by domain adaptive of conditional probability models through feature subsets, and then the Conditional Transition Component (CTC) [8] was modeled to make the method get developed. Transfer Component Analysis (TCA) [9] applies edge distribution adaptation to transfer learning, and some scholars have extended the Transfer Component Analysis (TCA), such as ACA [10], DTMKL [12], DME [13], CMD [14, 11], etc. Joint Distribution Alignment (JDA) [15] considers the edge distribution and the conditional distribution at the same time, and the effect is better, but the importance of the edge distribution and the conditional distribution is not considered, and the default weights are equal. The proposal of Balanced Distribution Adaptation (BDA) [16] improves the Joint Distribution Alignment (JDA). It considers the distribution adaptability between domains and enables the weight of each class to be changed adaptively. This type of method registers the probability distribution, but lacks consideration of subspace alignment. Therefore, its transfer effect is limited for smoke detection and recognition.

In addition, the subspace learning method is a method of transforming the source domain and the target domain to the same subspace, and then building a unified model. It solves the problem of domain adaptation mainly from two aspects, which are statistical property transformation and manifold learning [17]. In terms of statistical characteristics transformation, the Subspace Alignment (SA) [18] directly

reduces the distance between the two domains by optimizing the mapping function that converts the source domain subspace to the target subspace to bring the source domain subspace and the target domain subspace closer together. The Subspace Distribution Alignment (SDA) [19] adaptively expands SA by increasing the subspace variance without considering the local attributes of the subspace, and ignores the conditional distribution alignment; CORrelation Alignment (CORAL) [20] aligns subspaces in second-order statistics without considering distribution alignment. Scatter Component Analysis (SCA) [21] minimizes the divergence between them by converting the samples into a set of subspaces. In terms of manifold learning, the Sample Geodesic Flow (SGF) [21] treats the problem of domain adaptation as an incremental "walking" problem, and samples a limited number of points in the manifold space to construct a geodesic flow; Geodesic Stream Kernel (GFK) [18] extends the idea of sampling points in the manifold, and proposes a learning method for inter-domain geodesic stream kernels; Domain Invariant Projection (DIP) [23, 14] passed Grasmanian manifolds are used for domain adaptation, but conditional distribution alignment is ignored; Heilinger distance is used to approximate the geodesic distance in Riemann space, and a Statistical Manifold (SM) is proposed [24]. These methods solve the problem of domain adaptation from the perspective of subspace learning, but they lack the registration of probability distributions and still have limitations for smoke detection and recognition.

In 2018, Wang et al. proposed Manifold Embedded Distribution Alignment (MEDA) from the perspective of probability distribution adaptation method and subspace learning method [25], not only using the principle of structural risk minimization to learn the domain invariant classifier on manifold domain, but also aligning the edge distribution and conditional distribution dynamically, which provides a feasible method for quantitative calculation of adaptive factors. This method achieves good results in transfer learning classification, but still has the limitation of lacking feature extraction [26, 27] and original feature space alignment in solving smoke detection and recognition.

In view of the above problems, this paper proposes a smoke detection model by combining deep CNN and improved MEDA. This model can realize the transfer learning classification between various small datasets, and greatly reduce the time complexity. Firstly, pre-trained Resnet50 network is used for feature transfer to extract smoke features because of the difficulty in training small dataset of forest fire smoke; secondly, a correlated manifold embedded distribution alignment (CMEDA) is proposed to register the smoke features in order to align the input feature distributions of the source and target domains; finally, a trainable network model is constructed.

2 DATA FEATURES EXTRACTION IN ALL DOMAINS

This paper explores and compares several CNN models with different parameter settings for forest fire smoke detection, namely, AlexNet, Resnet, VGG, and GoogleNet.

In these models, AlexNet and GoogleNet use large convolution kernels with sizes of 11×11 and 7×7, and with step sizes of 3 and 5, which may ignore important features of the smoke area. Although VGG uses a small convolution kernel with a size of 3×3 and a step size of 1, VGG shallow network depth is not conducive to processing and extracting the features of each pixel of the smoke image, and VGG takes up a large space with a size of 528 MB. In addition, compared with Resnet34, Resnet50 replaces two convolution kernels of 3×3 with convolution kernels of sizes 1×1, 3×3, and 1×1. With similar time complexity, it is beneficial to process and extract the features of each pixel of smoke images, with higher accuracy and less computation. The parameter comparison between Resnet50 and other CNN models is shown in Table 1. As seen from Table 1, the top-1 accuracy, top-5 accuracy, and top-5 test error rate of Resnet50 on the ImageNet dataset are better than those of other state-of-the-art architectures of models. Therefore, this paper improves the architecture of the Resnet50 model based on the problem of forest fire smoke detection. The paper first uses the ImageNet dataset and small sample satellite remote sensing image dataset to fine-tune the parameters of the Resnet50 network model, then transfers the trained parameters corresponding to the convolutional layer and removes the fully connected layer to obtain the transfer learning network model, and finally, uses the above model for feature extraction of smoke and non-smoke images in the source and target domains.

Model	Parameter Amount/Million	Top-1 Accuracy (%)	Top-5 Accuracy (%)	Top-5 Error Rate (%)
GoogleNet	60	69.8	89.3	6.7
AlexNet	7	57.5	80.3	16.4
VGG16	138	70.5	91	7.3
Resnet50	256	75.9	92.9	5.25

Table 1. Comparison of Resnet50 and other CNN model parameters

The transfer learning model based on Resnet50 network in this paper is shown in Figure 1. This model transfers the trained parameters of the model on the ImageNet dataset and the small sample satellite Remote Sensing (RS) image dataset, removes the fully connected layers, and finally performs feature extraction on the smoke and images in the two domains. As shown in Figure 1 on the left, the smoke feature extraction model used in this paper is mainly composed of convolutional layers and downsampling layers. The model contains a total of 49 convolutional layers and 4 downsampling layers, of which the first segment is a convolutional layer composed of a $7 \times 7 \times 64$ convolution kernel; the second segment consists of three bottleneck structures, each of which contains three convolutional layers consisting of $1 \times 1 \times 64$, $3 \times 3 \times 64$, and $1 \times 1 \times 256$ convolution kernels; the third segment (unshown in Figure 1) is composed of 4 bottleneck structures, and each bottleneck structure contains three convolution layers consisting of $1 \times 1 \times 128$, $3 \times 3 \times 128$, and $1 \times 1 \times 512$ convolution kernels; the fourth segment (unshown in Figure 1) is composed of 6 bottleneck structures, and each bottleneck structure contains three

convolution layers consisting of $1 \times 1 \times 256$, $3 \times 3 \times 256$, and $1 \times 1 \times 1\,024$ convolution kernels; the fifth segment is composed of three bottleneck structures, and each bottleneck structure contains three convolution layers consisting of $1 \times 1 \times 512$, $3 \times 3 \times 512$, and $1 \times 1 \times 2\,048$ convolution kernels.

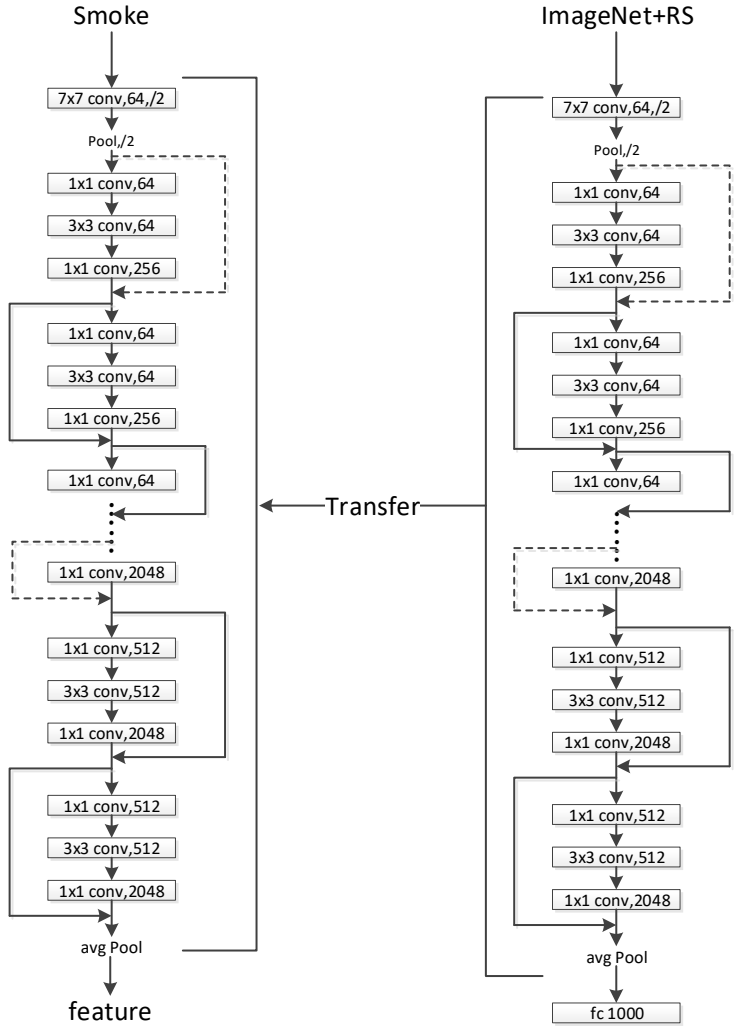

Figure 1. Transfer learning model based on Resnet50 network

This model is mainly obtained from the Resnet50 network, and the corresponding Resnet50 network training parameters are loaded at the same time. First, the convolutional layer is constructed based on the Resnet50 network; second, the smoke dataset is used as input to obtain the parameters of the convolutional layer in the

trained Resnet50 network on ImageNet; finally, image feature extraction is performed.

3 CORRELATED MANIFOLD EMBEDDED DISTRIBUTION ALIGNMENT ALGORITHM

From the perspective of the probability distribution adaptation method and the subspace learning method, MEDA uses the principle of structural risk minimization to learn the domain-invariant classifier on the manifold domain, and dynamically aligns the edge distribution and conditional distribution. Therefore, the drift between domains is greatly reduced. The specific flow of the algorithm is shown in Figure 2.

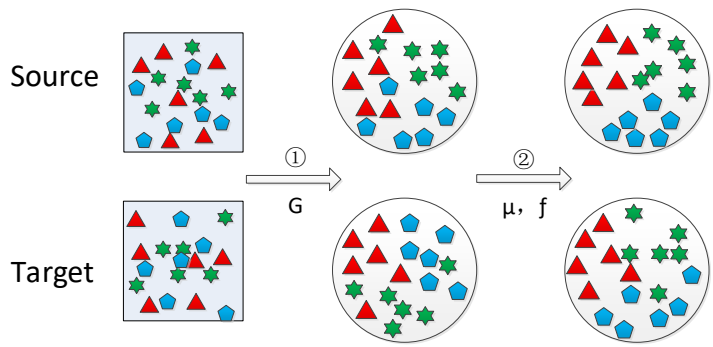

Figure 2. The idea of the manifold distribution registration method. ① Transform the features in the original space into the manifold space by learning the manifold kernel **G**. ② Align dynamically the distributions through learning the adaptive factor μ, and learn the final domain-invariant classifier f through structural risk minimization in the manifold space.

Compared with the features of the original space, the features of the manifold space have some good geometric structures, which can avoid the distortion of the features. Therefore, to eliminate the degenerate feature transformation, the manifold feature learning is an important processing step. When learning manifold feature transformations, MEDA uses d-dimensional subspaces to model the data domain, and then embeds these subspaces into the manifold **G**.

After obtaining the manifold characteristics, in order to dynamically measure the relative importance of the edge distribution and the conditional distribution, MEDA introduced an adaptive factor to adaptively balance the two distributions. In formal language, the adaptive distributed adaptation $\overline{D_f}$ can be expressed as:

$$\overline{D_f}\left(\mathrm{D}_s, \mathrm{D}_t\right) = (1-\mu)\, D_f\left(P_s, P_t\right) + \mu \sum_{c=1}^{C} D_f^{(c)}\left(Q_s, Q_t\right) \tag{1}$$

where $\mu \in [0,1]$ represents the adaptive factor, $c \in \{1, \cdots, C\}$ is a category indication. $D_f(P_s, P_t)$ represents edge distribution adaptation, $D_f^{(c)}(Q_s, Q_t)$ indicates a conditional distribution adaptation to category c.

MEDA uses Maximum Mean Discrepancy (MMD) to calculate the difference between two probability distributions. The MMD distance between two probability distributions p and q is defined as $d^2(p,q) = \left(\mathrm{E}_p\left[\phi\left(\mathbf{Z}_s\right)\right] - \mathrm{E}_q\left[\phi\left(\mathbf{Z}_t\right)\right]\right)_{\mathcal{H}_K}^2$, where $\mathcal{H}_K$ is the Reproducing Kernel Hilbert space (RKHS) expanded by the feature map $\phi(\bullet)$, and $\mathrm{E}(\bullet)$ is the mean of the embedded samples. Finally, MEDA summarizes the manifold learning and dynamic distribution alignment, and learns the final domain-invariant classifier through structural risk minimization:

$$f = \underset{f \in \sum_{i=1}^{n} \mathcal{H}_K}{\arg\min}\ l\left(f\left(g\left(\mathbf{x}_i\right)\right), y_i\right) + \eta \|f\|_K^2 + \lambda \overline{D_f}\left(\mathrm{D}_s, \mathrm{D}_t\right) + \rho \overline{R_f}\left(\mathrm{D}_s, \mathrm{D}_t\right). \tag{2}$$

In the formula, $g\left(\mathbf{x}_i\right)$ represents learning manifold features, $\overline{D_f}\left(\mathrm{D}_s, \mathrm{D}_t\right)$ represents dynamically aligned edge distribution and conditional distribution, and $\overline{R_f}\left(\mathrm{D}_s, \mathrm{D}_t\right)$ is a regularization term. This part can better learn the geometric properties of the closest point in the manifold space. MEDA is the first attempt to deal with degraded feature conversion and unassessed distribution alignment challenges. It has achieved a good result in classification accuracy, but it still has limitations in the detection and recognition of smoke images. If the input features of the two fields are first distributed and aligned in the original space before manifold feature learning, the two fields will be better registered and the classification accuracy will be higher. Therefore, this paper proposes the CMEDA module. The CMEDA module adds an input feature distribution alignment section to the MEDA module. The input feature distribution alignment first removes the feature correlation of the source domain, then re-associates the target domain, and finally adds the association of the target domain to the source characteristics. The process is shown in Figure 3.

In the original space, the input feature distributions of the source and target domains are aligned by comparing the second-order statistics of the two domains. This method can minimize the domain offset. In order to minimize the distance between the second-order statistics (covariance) of the two domains, this paper performs a linear transformation on the original source features and uses the Frobenius norm as the matrix distance metric, as follows:

$$\min_{\mathbf{A}} \left\|\mathbf{C}_{\hat{S}} - \mathbf{C}_T\right\|_F^2 = \min_{\mathbf{A}} \left\|\mathbf{A}^T \mathbf{C}_S \mathbf{A} - \mathbf{C}_T\right\|_F^2. \tag{3}$$

Calculated $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}_S \mathbf{E} = \left(\mathbf{U}_S \mathbf{\Sigma}_S{}^{\frac{1}{2}} \mathbf{U}_S{}^T\right)\left(\mathbf{U}_{T[1:r]} \mathbf{\Sigma}_{T[1:r]}{}^{\frac{1}{2}} \mathbf{U}_{T[1:r]}{}^T\right). \tag{4}$$

$\mathbf{U}_S \mathbf{\Sigma}_S{}^{\frac{1}{2}} \mathbf{U}_S{}^T$ in $\mathbf{A}$ can be regarded as removing the feature correlation of the source domain, and $\mathbf{U}_{T[1:r]} \mathbf{\Sigma}_{T[1:r]}{}^{\frac{1}{2}} \mathbf{U}_{T[1:r]}{}^T$ can be regarded as re-associating the target domain and adding the association of the target domain to the source characteristic.

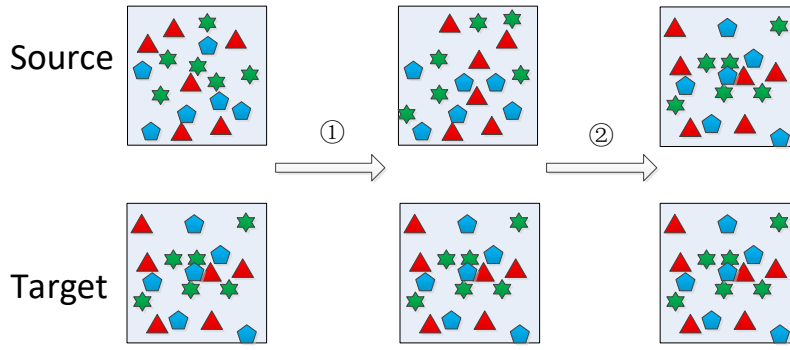

Figure 3. Associated alignment input feature distribution. ① Remove the feature correlation of the source domain and keep the target domain unchanged. ② Re-associate the target domain, add the correlation of the target domain to the source characteristics, and obtain the feature distribution which is aligned between the source domain and the target domain.

Therefore, the CMEDA module consists of three parts: the first part is the association alignment input feature distribution, the second part is the manifold feature learning, and the third part is the dynamic alignment of the edge distribution and the conditional distribution. The overall flowchart of the module is shown in Figure 4.

In manifold feature learning, $\mathbf{S}_s$ and $\mathbf{S}_t$ are used to represent the subspace of the source and target domains after Principal Component Analysis (PCA), respectively, then $\mathbf{G}$ can be regarded as a set of all d-dimensional subspaces. Each d-dimensional primitive subspace can be viewed as a point on $\mathbf{G}$ and the geodetic line $\{\mathbf{\Phi}(t) : 0 \leq t \leq 1\}$ between two points can form a path between two subspaces.

Let $\mathbf{S}_s = \mathbf{\Phi}(0)$, $\mathbf{S}_t = \mathbf{\Phi}(1)$, finding a geodesic from $\mathbf{\Phi}(0)$ to $\mathbf{\Phi}(1)$ is equivalent to transforming the original features into a space of infinite dimensions, and finally reducing the drift between domains. In particular, features in a manifold space can be represented as $\mathbf{Z} = \mathbf{\Phi}(t)^T\mathbf{X}$. The inner product of the transformed features $\mathbf{z}_i$ and $\mathbf{z}_j$ defines a semi-positive definite GFK.

$$\langle \mathbf{z}_i, \mathbf{z}_j \rangle = \int_0^1 \left(\mathbf{\Phi}(t)^T\mathbf{x}_i\right)^T \left(\mathbf{\Phi}(t)^T\mathbf{x}_j\right) dt = \mathbf{x}_i{}^T\mathbf{G}\mathbf{x}_j. \tag{5}$$

Therefore, through $\mathbf{Z} = \sqrt{\mathbf{G}}\mathbf{X}$, the features in the original space can be transformed into Grassmann manifold space, and the kernel $\mathbf{G}$ can be efficiently calculated by matrix singular value decomposition.

In addition, because the target domain data D_t has no labels, it is not feasible to directly evaluate the conditional probability distribution $Q_t = Q_t(y_t|\mathbf{Z}_t)$ of

the target domain, and when the number of samples is large enough, $Q_t(\mathbf{Z}_t|y_t)$ and Q_t have a good similarity. The class conditional probability $Q_t(\mathbf{Z}_t|y_t)$ is used to approximate Q_t. To approximate $Q_t(\mathbf{Z}_t|y_t)$, a weak classifier is trained on the source domain D_s, and then this weak classifier is used to predict on D_t to obtain pseudo-labels in the target domain. Since the confidence of these pseudo-labels may not be high, this weak classifier can iteratively modify the prediction results.

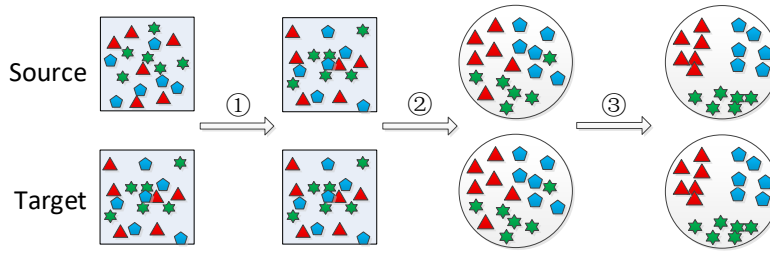

Figure 4. CMEDA flowchart. ① Remove the feature correlation of the source domain and add the correlation of the target domain to the source features to obtain the feature distribution of the source domain aligned with the target domain in the original space. ② Transform the distribution-aligned features in the original space into the manifold space by learning the manifold kernel. ③ Align adaptively the distribution in the manifold space by learning adaptive factors, and learn the final domain-invariant classifier through structural risk minimization.

4 FOREST FIRE SMOKE DETECTION METHOD BASED ON DC-CMEDA

The main purpose of the forest fire smoke detection method is to construct a classification algorithm which is able to realize the migration of trace datasets to achieve forest fire smoke detection. The structure of the DC-CMEDA model is shown in Figure 5. The model first uses Resnet50 to extract N-dimensional features from the source and target domain data. Then the source and target domain features are processed by the CMEDA module, that is, the input distribution features are aligned, and the structural risk minimization method is used to learn the domain-invariant classifier in the Grassmann manifold. Meanwhile, dynamic distribution alignment is performed by considering the different importance of edge distribution and conditional distribution, and finally, transfer classification of smoke images from source domain to target domain is realized.

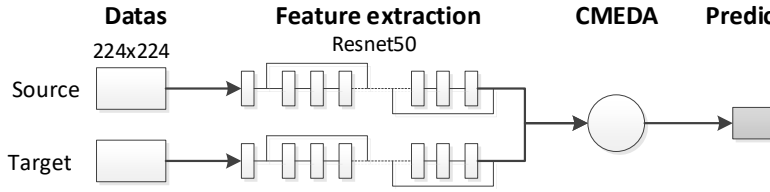

Figure 5. DC-CMEDA model structure

In transfer learning, we use the Resnet50 model based on the ImageNet dataset as a CNN model to extract smoke features for each image. Experiments show that the network trained on the ImageNet dataset has better generalization ability.

During the transfer from the source domain to the target domain, the CMEDA module aligns the input distribution features of the two domains to improvement of the accuracy of image detection. CMEDA not only learned the domain-invariant classifier on the manifold domain by using the principle of structural risk minimization, but also dynamically aligned the edge distribution and conditional distribution. Experiments show that the CMEDA network significantly increases the accuracy of smoke detection compared with the MEDA network.

This method performs feature extraction and transfer classification on images in two domains (satellite remote sensing and video images) and it outputs two types of results (smoked and smokeless). The main steps of the experimental algorithm are as follows.

Algorithm DC-CMEDA

Input: Source domain dataset $\{\mathbf{Simage}_i : 1 \leq i \leq M\}$,
Target domain dataset $\{\mathbf{Timage}_j : 1 \leq j \leq N\}$

Output: Classifier f

1: Preprocess Source and Target domain datasets: Adjust $\mathbf{Simage}_i$ and $\mathbf{Timage}_j$ to resolutions of $3 \times 224 \times 224$ and transform randomly and normalize them
2: Construct transfer learning model based on Resnet50 network
3: Perform feature extraction on $\mathbf{Simage}_i$ and $\mathbf{Timage}_j$, get feature matrices $\mathbf{X}_s$ and $\mathbf{X}_t$, and get source domain label $\mathbf{y}_s$
4: Obtain the feature distribution of the source domain aligned with the target domain in the original space by $\mathbf{X}_s{}' = \mathbf{X}_s * \mathbf{A}$, and get data matrix $\mathbf{X} = (\mathbf{X}_s{}', \mathbf{X}_t)$
5: Train a weak classifier using $\mathbf{D}_s$, then apply the classifier to predict pseudo-label $\hat{\mathbf{y}}_t$ in target domain $\mathbf{D}_s$
6: **repeat**
7: Calculate the adaptive factor μ using Formula (1) and obtain f via Formula (2)
8: Update the label of $\mathrm{D}_t : \hat{\mathbf{y}}_t = f(\mathbf{X}_t)$
9: **until** Convergence
10: **return** Classifier f

5 EXPERIMENTAL RESULTS AND ANALYSES

5.1 Experimental Dataset

This paper studies the transfer learning classification technology between smoke images of different resolution videos in a forest fire video monitoring system. Satellite remote sensing images and video images are selected as experimental data. For satellite remote sensing (RS) images, a large amount of data are available, the cost is low, and it is easy to apply in the real world. When a relatively obvious smoke image is detected, it means that the fire is already very large, and timely feedback and disaster relief cannot be achieved. The video images can just overcome the shortcomings of satellite remote sensing (RS) images which are unable to offer real-time feed back. Video images can quickly capture the fire situation and give real-time feedback, but they still have the disadvantage that they are unable to be automatically interpreted. The transfer learning of satellite remote sensing (RS) images and video images can give full play to their respective advantages and achieve the effect of real-time and automatic interpretation. The dataset in this paper is derived from the Academy of Forestry of Shanxi Province, China.

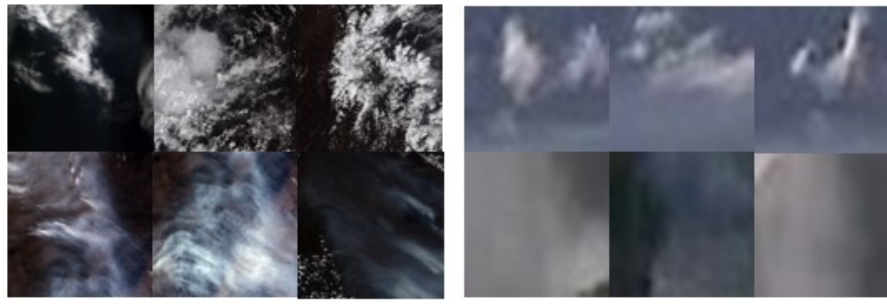

Figure 6. Image sample map

5.2 Experimental Evaluation Criteria

This paper uses Precision, Recall, and the Harmony Mean F1 of both Precision and Recall to measure the network performance, so that images with smoke are positive and images with no smoke are negative. The calculation formula is as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{8}$$

$$F1 = \frac{2TP}{2TP + FP + FN}. \tag{9}$$

TP means to predict positive class as positive class; TN means to predict negative classes as negative classes; FP means to predict negative class as positive class; FN means to predict positive class as negative class.

5.3 Results and Analyses

The followings are four comparative experiments in this paper. Experiment 1 shows the comparison of the parameters on the video image sample dataset and the satellite remote sensing sample dataset by applying the Resnet50 method and other methods. Experiment 2 gives a comparison of detection results of several domain adaptive methods based on Resnet50 network. Experiment 3 compares DC-CMEDA-based forest fire smoke detection methods with other state-of-the-art methods. Experiment 4 gives a comparison of the effects of Resnet modules on the DC-CMEDA algorithm.

The experimental dataset includes a small sample set of 200 satellite remote sensing (RS) images and 200 video images. Different image distributions represent two different domains. Each domain includes 100 images with smoke and 100 images with no smoke.

5.3.1 Experiment 1

In transfer learning, the Resnet50 network is compared with other CNN networks. Tables 2 and 3 show the comparison of the parameters on the video image sample dataset and the satellite remote sensing sample dataset by applying the Resnet50 method and other methods. In this paper, the cross-validation method is used to divide the sample set in each domain into training set, validation set, and test set in proportion. Among them, the training set accounts for 50 % of the total sample (50 images with smoke and 50 images with no smoke), and the validation set and test set each account for 25 % (each contains 25 images with smoke and 25 images with no smoke).

As can be seen from Table 2, the accuracy of the video image sample dataset using AlexNet and GoogleNet is the lowest, and the false positive and false negative score the worst. The detection results using VGG16 are better than the detection results using AlexNet and GoogleNet. But compared with the Resnet50 model, the accuracy of VGG16 is still lower and the false positive is higher. Among AlexNet, GoogleNet, VGG16, and Resnet50, Resnet50 has achieved the best results, which are 16.67 % in false positive, 11.54 % in false negative, and 86.00 % in accuracy.

Table 3 shows the satellite remote sensing verification set. The accuracy of the AlexNet, GoogleNet, and VGG16 models is also lower than that of the Resnet50

Model	False Positive (%)	False Negative (%)	Accuracy (%)
AlexNet	30.43	29.62	70.00
GoogleNet	26.09	25.93	74.00
VGG16	24.00	16.00	80.00
Resnet50	16.67	11.54	86.00

Table 2. Accuracy of video image sample set during training

model, and the false positive and false negative scores are also relatively poor compared to the Resnet50. That is, the experimental verification of the satellite remote sensing set using Resnet50 has relatively good results, of which false positive is 16.67 %, false negative is 19.23 %, and the accuracy is 82.00 %. Therefore, the performance of the Resnet50 model is better than other models.

Model	False Positive (%)	False Negative (%)	Accuracy (%)
AlexNet	29.17	26.92	72.00
GoogleNet	26.09	25.93	74.00
VGG16	17.39	25.93	78.00
Resnet50	16.67	19.23	82.00

Table 3. Accuracy of satellite remote sensing sample set during training

Based on the analyses and summaries of the results in Tables 2 and 3, the following conclusions can be drawn.

1. Learning model parameters through transfer learning can reduce the number of samples required for a dataset. But for too small sample datasets, even the most advanced CNN networks have low accuracy in recognition and classification detection due to the inability to fully adjust the parameters.
2. The model parameters in the experiment are obtained from the pre-trained model migration based on the ImageNet dataset, so there are fewer satellite remote sensing smoke images in the ImageNet dataset. Therefore, comparing Table 2 with Table 3, it is found that the accuracy under the satellite remote sensing verification set is significantly lower than the accuracy under the video image verification set.
3. The accuracy of both the satellite remote sensing verification set and the video image verification set is better than other CNN networks due to the advantages of the size of the convolution kernel and the depth of the network in the Resnet50 network model.

5.3.2 Experiment 2

This part uses the Resnet50 network along with JDA, BDA, GFK, MEDA and CMEDA methods for testing and comparison. Table 4 shows the compared results of false positives, false negatives, and accuracy of various detection and recognition

methods of the satellite remote sensing sample set as the source domain and the video image sample set as the target domain. Table 5 shows the comparison between the source domain, the video image sample set and the target domain, the satellite remote sensing sample set.

As seen from Table 4, when the satellite remote sensing sample set is used as the source domain and the video image sample set as the target domain, the transfer effect of MEDA is obviously better than that of GFK, TDA and BDA either from the perspective of false positives and false negatives or from the perspective of accuracy. However, when MEDA is compared with the CMEDA module, the combined CMEDA has a better effect in experimental verification. In CMEDA, false positive is 4.81 %, false negative is 3.13 %, and the accuracy is 96.00 %.

Model	False Positive (%)	False Negative (%)	Accuracy (%)
GFK	18.18	11.88	85.00
JDA	11.32	9.57	89.50
BDA	8.82	8.16	91.50
MEDA	7.29	3.85	94.50
CMEDA	4.81	3.13	96.00

Table 4. Transfer accuracy of satellite remote sensing images to video images

As seen from Table 5, when the video image sample set is used as the source domain and the satellite remote sensing sample set as the target domain, the transfer effect of CMEDA is much better than that of GFK, TDA, BDA, and even better than MEDA. In CMEDA, false positive is 11.76 %, false negative is 9.18 %, the accuracy is 89.50 %, and the accuracy is significantly improved by 2.00 % comparing with that of MEDA.

Model	False Positive (%)	False Negative (%)	Accuracy (%)
GFK	21.43	18.63	80.00
JDA	16.19	15.79	84.00
BDA	15.89	13.98	85.00
MEDA	13.08	11.83	87.50
CMEDA	11.76	9.18	89.50

Table 5. Transfer accuracy of video images to satellite remote sensing images

Based on the analyses and summaries of the results in Tables 4 and 5, the conclusions are drawn as follows.

1. GFK, JDA, and BDA solve the domain adaptation problem by taking either subspace learning method only or probability distribution adaptation method only into consideration while MEDA applies the two methods simultaneously, and thus has a better transfer effect.

2. The input feature distributions of the source and target domains in CMEDA are aligned in the original space before the manifold feature learning is conducted, so the improved CMEDA module performs better than MEDA.

3. The parameters of the CNN network model are obtained from transfering a pre-trained model based on the ImageNet dataset, so the accuracy of transfering from video image to satellite remote sensing image is lower than that of transfering from satellite remote sensing image to video image.

5.3.3 Experiment 3

This part compares the deep convolutional long-term short-term memory network (DC-ILSTM) proposed by Wei et al. in [2] and the smoke detection method using convolution and recursive network proposed by Filonenko et al. in [3] to verify the detection effect of the DC-CMEDA method shown in Tables 6 and 7.

Table 6 shows the accuracy test results transfered from satellite remote sensing images to video images. After having tested the two methods in [2] and [3], this paper uses satellite remote sensing images to train the model and to fine-tune the parameters, and uses video images as the test set to verify the detection effect. Table 6 shows that the DC-CMEDA method proposed in this paper has an efficient detection effect, and the accuracy is as high as 96.0 %.

Performance	Filonenko's Method [3]	DC-ILSTM [2]	DC-CMEDA
Accuracy (%)	93.5	94.5	96.0
Precision (%)	93.3	93.4	94.9
Recall (%)	94.2	96.2	96.8
F1 (%)	93.7	94.8	95.8

Table 6. Accuracy test results transfered from satellite remote sensing images to video images

Table 7 shows the test results of the accuracy transfered from video images to satellite remote sensing images. Models are trained with video images, the parameters are fine-tuned, and satellite remote sensing images are used as the test set to verify the detection effect. As shown in Table 7, after our having tested the two methods in [2] and [3], and our own method DC-CMEDA, the DC-CMEDA method proposed in this paper has an efficient detection effect, and the accuracy is as high as 89.5 %.

The analyses and summaries of the results in Tables 6 and 7 show that in terms of small sample datasets, the DC-CMEDA method proposed in this paper has a more efficient detection effect than other most advanced methods. Take the precision, recall, and the harmonic mean F1 of the former two into consideration, the proposed DC-CMEDA method performs the best.

Performance	Filonenko's Method [3]	DC-ILSTM [2]	DC-CMEDA
Accuracy (%)	83.0	85.5	89.5
Precision (%)	82.5	86.2	88.1
Recall (%)	82.5	83.5	90.8
F1 (%)	82.5	84.8	89.4

Table 7. Accuracy test results transfered from video images to satellite remote sensing images

5.3.4 Experiment 4

Under the condition of applying video images and satellite remote sensing smoke images small datasets to DC-CMEDA, the convergence accuracy and convergence speed of the DC-CMEDA-based algorithms are contrasted after combining the CMEDA module with module Resnet34 and module Resnet50, respectively. The experimental results are shown in Figure 7. The red line chart in Figure 7 shows the accuracy of each iteration in the process of learning the video image sample dataset and predicting the satellite remote sensing sample dataset. The blue line chart shows the accuracy of each iteration in the process of learning the satellite remote sensing sample dataset and predicting the video image sample dataset.

The experimental results show that (1) the convergence accuracy and convergence speed of the DC-CMEDA algorithm combined with Resnet50 module are better than those of the DC-CMEDA algorithm combined with Resnet34 module. Taking convergence speed and small sample datasets into consideration, this paper does not discuss cases combined with other Resnet size except Resnet34 and Resnet50. (2) The domain adaptive algorithm in the DC-CMEDA has converged once the number of iterations is less than 10 times. It is clear that under the same accuracy condition, the number of CMEDA iterations is much less than 150 times of the ILSTM algorithm, an example as in deep convolution integration [2], and the algorithm complexity of CMEDA is lower than that of ILSTM. The convergence speed of DC-CMEDA is much higher than that of other deep convolution integrated smoke detection methods [2, 3].

6 CONCLUSION

Aiming at solving the problem of small sample datasets in certain scenes in the forest fire smoke detection, this paper proposes a DC-CMEDA model. This model not only performs feature extraction on a small sample dataset of forest fires on a deep transfer learning architecture, it also proposes a smoke feature registration in combination with the CMEDA module. In the experiments, the model was evaluated based on satellite remote sensing image and video image datasets, and compared with various state-of-the-art forest fire smoke detection methods. The results show that, in terms of detection performance, the model discussed in this paper detects smoke faster and meanwhile, the detection accuracy is higher than other methods.

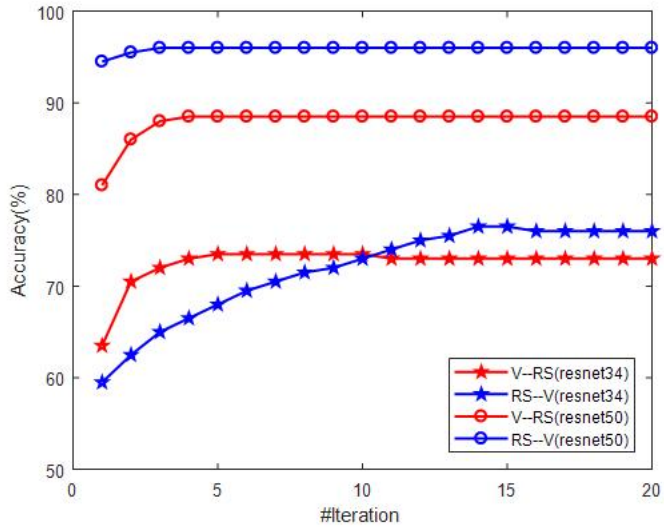

Figure 7. Convergence rate comparison

Among the various methods of smoke detection, the use of domain adaptive methods has not been reported. This paper first attempts to combine domain adaptation with deep transfer learning to solve the smoke detection problem. In the next stage of our work, we need to further optimize the model to improve the accuracy of forest fire smoke detection.

Acknowledgements

This study is funded by Joint Research Fund for Overseas Chinese Scholars and Scholars in Hong Kong and Macao (Grant No. 61828601), Natural Science Foundation of Shanxi Province (Grant No. 201801D121141), and Provincial Program on Key Research Projects of Shanxi (Social Development Area, Grant No. 201903D321003).

REFERENCES

[1] Wang, W.—Mao, W.—He, J.—Dou, Z.: Smoke Recognition Based on Deep Transfer Learning. Journal of Computer Applications, Vol. 37, 2017, No. 11, pp. 3176–3181.

[2] Wei, X.—Wu, S.—Wang, Y.: Forest Fire Smoke Detection Model Based on Deep Convolution Long Short-Term Memory Network. Journal of Computer Applications, Vol. 39, 2019, No. 10, pp. 2883–2887.

[3] Filonenko, A.—Kurnianggoro, L.—Jo, K. H.: Smoke Detection on Video Sequences Using Convolutional and Recurrent Neural Networks. In: Nguyen, N., Papadopoulos, G., Jędrzejowicz, P., Trawiński, B., Vossen, G. (Eds.): Computational Collective Intelligence (ICCCI 2017). Springer, Cham, Lecture Notes in Computer Science, Vol. 10449, 2017, pp. 558–566, doi: 10.1007/978-3-319-67077-5_54.

[4] Han, C.—Ma, J.—Wu, W.—Chen, J.: Smoke Image Detection Based on Deep Transfer Learning. Journal of Wuhan Textile University, Vol. 32, 2019, No. 2, pp. 65–71.

[5] Pan, S. J.—Yang, Q.: A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, Vol. 22, 2010, No. 10, pp. 1345–1359, doi: 10.1109/TKDE.2009.191.

[6] Wang, J. et al.: Everything about Transfer Learning and Domain Adapation. Available at: `http://transferlearning.xyz`, 2018.

[7] Satpal, S.—Sarawagi, S.: Domain Adaptation of Conditional Probability Models via Feature Subsetting. In: Kok, J. N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (Eds.): Knowledge Discovery in Databases: PKDD 2007. Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 4702, 2017, pp. 224–235, doi: 10.1007/978-3-540-74976-9_23.

[8] Gong, M.—Zhang, K.—Liu, T.—Tao, D.—Glymour, C.—Schölkopf, B.: Domain Adaptation with Conditional Transferable Components. Proceedings of the 33rd International Conference on Machine Learning (ICML'16), 2016, Vol. 48, pp. 2839–2848.

[9] Pan, S. J.—Tsang, I. W.—Kwok, J. T.—Yang, Q.: Domain Adaptation via Transfer Component Analysis. IEEE Transactions on Neural Networks, Vol. 22, 2011, No. 2, pp. 199–210, doi: 10.1109/TNN.2010.2091281.

[10] Dorri, F.—Ghodsi, A.: Adapting Component Analysis. IEEE 12th International Conference on Data Mining (ICDM 2012), 2012, pp. 846–851, doi: 10.1109/ICDM.2012.85.

[11] Lan, Z.—Sourina, O.—Wang, L. P.—Scherer, R.—Müller-Putz, G. R.: Domain Adaptation Techniques for EEG-Based Emotion Recognition: A Comparative Study on Two Public Datasets. IEEE Transactions on Cognitive and Developmental Systems, Vol. 11, 2019, No. 1, pp. 85–94, doi: 10.1109/TCDS.2018.2826840.

[12] Duan, L.—Tsang, I. W.—Xu, D.: Domain Transfer Multiple Kernel Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, 2012, No. 3, pp. 465–479, doi: 10.1109/TPAMI.2011.114.

[13] Baktashmotlagh, M.—Harandi, M.—Salzmann, M.: Distribution-Matching Embedding for Visual Domain Adaptation. The Journal of Machine Learning Research, Vol. 17, 2016, No. 1, pp. 3760–3789.

[14] Zellinger, W.—Lughofer, E.—Saminger-Platz, S.—Grubinger, T.—Natschläger, T.: Central Moment Discrepancy (CMD) for Domain-Invariant Representation Learning. International Conference on Learning Representations (ICLR 2017), 2017, arXiv:1702.08811.

[15] Long, M.—Wang, J.—Ding, G.—Sun, J.—Yu, P. S.: Transfer Feature Learning with Joint Distribution Adaptation. 2013 IEEE International Conference on Computer Vision (ICCV 2013), 2013, pp. 2200–2207, doi: 10.1109/ICCV.2013.274.

[16] Wang, J.—Chen, Y.—Hao, S.—Feng, W.—Shen, Z.: Balanced Distribution Adaptation for Transfer Learning. 2017 IEEE International Conference on Data Mining (ICDM 2017), 2017, pp. 1129–1134, doi: 10.1109/ICDM.2017.150.

[17] Gopalan, R.—Li, R.—Chellappa, R.: Domain Adaptation for Object Recognition: An Unsupervised Approach. IEEE International Conference on Computer Vision (ICCV 2011), 2011, pp. 999–1006, doi: 10.1109/ICCV.2011.6126344.

[18] Fernando, B.—Habrard, A.—Sebban, M.—Tuytelaars, T.: Unsupervised Visual Domain Adaptation Using Subspace Alignment. Proceedings of the IEEE International Conference on Computer Vision (ICCV 2013), 2013, pp. 2960–2967, doi: 10.1109/ICCV.2013.368.

[19] Sun, B.—Saenko, K.: Subspace Distribution Alignment for Unsupervised Domain Adaptation. 26th British Machine Vision Conference (BMVC 2015), 2015, Art. No. 24, 10 pp., doi: 10.5244/C.29.24.

[20] Sun, B.—Feng, J.—Saenko, K.: Return of Frustratingly Easy Domain Adaptation. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI 2016), 2016, pp. 2059–2065.

[21] Ghifary, M.—Balduzzi, D.—Kleijn, W. B.—Zhang, M.: Scatter Component Analysis: A Unifed Framework for Domain Adaptation and Domain Generalization. IEEE Transactions on Patern Analysis and Machine Intelligence, Vol. 39, 2017, No. 7, pp. 1414–1430, doi: 10.1109/TPAMI.2016.2599532.

[22] Gong, B.—Shi, Y.—Sha, F.—Grauman, K.: Geodesic Flow Kernel for Unsupervised Domain Adaptation. IEEE Conference on Computer Vision and Patern Recognition (CVPR 2012), 2012, pp. 2066–2073, doi: 10.1109/CVPR.2012.6247911.

[23] Baktashmotlagh, M.—Harandi, M. T.—Lovell, B. C.—Salzmann, M.: Unsupervised Domain Adaptation by Domain Invariant Projection. Proceedings of the IEEE International Conference on Computer Vision (ICCV 2013), 2013, pp. 769–776, doi: 10.1109/ICCV.2013.100.

[24] Baktashmotlagh, M.—Harandi, M. T.—Lovell, B. C.—Salzmann, M.: Domain Adaptation on the Statistical Manifold. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (ICCV 2014), 2014, pp. 2481–2488, doi: 10.1109/CVPR.2014.318.

[25] Wang, J.—Feng, W.—Chen, Y.—Yu, H.—Huang, M.—Yu, P. S.: Visual Domain Adaptation with Manifold Embedded Distribution Alignment. Proceedings of the 26th ACM International Conference on Multimedia (MM 2018), Seoul, Republic of Korea, 2018, pp. 402–410, doi: 10.1145/3240508.3240512.

[26] Huang, Z.—Li, M.—Chousidis, C.—Mousavi, A.—Jiang, C.: Schema Theory-Based Data Engineering in Gene Expression Programming for Big Data Analytics. IEEE Transactions on Evolutionary Computation, Vol. 22, 2018, No. 5, pp. 792–804, doi: 10.1109/TEVC.2017.2771445.

[27] Wang, L. P.—Wang, Y.—Chang, Q.: Feature Selection Methods for Big Data Bioinformatics: A Survey from the Search Perspective. Methods, Vol. 111, 2016, pp. 21–31, doi: 10.1016/j.ymeth.2016.08.014.

Yaoli Wang received his B.S. and M.Sc. degrees from the Northwestern Polytechnical University, China and his Ph.D. degree from the Taiyuan University of Technology, China. His research interests include intelligent techniques with applications to communications, image/video processing, robots. He is (co-)author of over 40 papers. He holds two Chinese patents. He is Member of the Expert Committee on Embedded Systems of the China Electronic Society, Member of the Seventh Forest and Grassland Fire Prevention Committee of the China Forestry Society, and Member of the Seventh Forest Fire Control Branch of the China Fire Control Association.

Xiaohui Liu received her B.Sc. degree from the Taiyuan University of Science and Technology, China. Her research interest is intelligent techniques with applications to robots. She is now studying for her M.Sc. degree at the Taiyuan University of Technology.

Maozhen Li is Professor in the Department of Electronic and Computer Engineering, Brunel University London, UK. He is also a Visiting Professor of Tongji University, Shanghai, China. He received the Ph.D. from the Institute of Software, Chinese Academy of Sciences, in 1997. His main research interests include high performance computing, big data analytics and intelligent systems with applications to smart grid, smart manufacturing and smart cities. He has over 180 research publications in these areas including 4 books. He has served over 30 IEEE conferences and is on the editorial board of a number of journals. He is a Fellow of the British Computer Society (BCS) and the Institute of Engineering and Technology (IET).

Wenxia Di received her B.Sc. degree from the Inner Mongolian Normal University, China. Her research interests include intelligent techniques with applications to speech signal processing. She is (co-)author of over 10 papers. She has co-authored 2 books. She is Member of English Voice Association.

Lipo Wang received his B.Sc. degree from the National University of Defense Technology, China, and Ph.D. from the Louisiana State University, USA. His research interests include intelligent techniques with applications to communications, image/video processing, biomedical engineering, and data mining. He is (co-)author of over 270 papers, of which more than 90 are in journals. He holds a U.S. patent in neural networks and a Chinese patent in VLSI. He has co-authored 2 monographs and (co-)edited 15 books. He was/will be a keynote/panel speaker for 21 international conferences. He is/was Associate Editor/Editorial Board Member of 30 international journals, including 3 IEEE Transactions, and the guest editor for 10 journal special issues. He is an AdCom Member (2010–2015) of the IEEE Computational Intelligence Society (CIS) and served as CIS Vice President for Technical Activities and Chair of Emergent Technologies Technical Committee. He is Member of the Board of Governors of the International Neural Network Society (2011–2016) and was an AdCom member of the IEEE Biometrics Council. He served as Chair of Education Committee, IEEE Engineering in Medicine and Biology Society (EMBS). He was President of the Asia-Pacific Neural Network Assembly (APNNA) and received the APNNA Excellent Service Award. He was founding Chair of both the EMBS Singapore Chapter and CIS Singapore Chapter. He serves/served as chair/committee member of over 200 international conferences.