A Transfer Learning Method for Aircrafts Recognition

Hongbo Li(✉), Bin Guo, Tong Gao, and Hao Chen

Harbin Institute of Technology, Harbin, People's Republic of China
drbobo@hit.edu.cn

Abstract. An effective method for recognizing aircrafts with different resolutions is proposed. Since training aircraft samples and test aircraft samples are imaging in different resolutions, different satellites and different imaging conditions, they obey different distributions. The Feature Subspace Alignment and Balanced Distribution Adaptation (FSA-BDA) method is proposed to solve this problem. Different from other transfer learning methods, it considers both spatial alignment and probability adaptation, so that, the probability distribution of the source domain data and the target domain data is as consistent as possible in the same feature space. The method first performs FSA, which maps the source domain and the target domain data to a low-dimensional common mapping space through different mapping matrices for preserving the structural information. Secondly, the BDA method is used to properly adapt the marginal probability and the conditional probability through the weight adjustment, which can leverage the importance of the marginal and conditional distribution discrepancies. This paper aims at recognizing three types of aircrafts, which are B52, F15 and F16 aircrafts. The experimental results show that the proposed method is better than several state-of-the-art methods.

Keywords: Aircrafts recognition · Probability adaptation · Transfer learning

1 Introduction

In recent years, computer vision and related technologies have developed rapidly. Image recognition technology has become a hot topic in research. As the number of remote sensing satellites increases, the amount of data on aircraft targets based on optical satellite remote sensing images is also increasing. However, the acquisition time, weather conditions, imaging angles and image spatial resolutions of the aircraft in optical images are usually different. Moreover, due to the different resolutions, the data difference is very large. Therefore, it is necessary to study how to automatically identify aircraft targets based on a small amount of labeled data.

The representative of the earlier aircraft image recognition research is Dudani S A who proved that the aircraft was automatically identified [1]. In [2], a coarse-to-fine strategy is proposed. And the fine matching phase uses the shape feature to complete the matching of the candidate target. A reconstruction-based similarity measure is proposed [3], which transforms the type recognition problem into a reconstruction problem. In recent years, a novel airport detection and aircraft recognition method that

S. Han et al. (Eds.): AICON 2019, LNICST 286, pp. 175–185, 2019.
https://doi.org/10.1007/978-3-030-22968-9_16

is based on the two-layer visual saliency analysis model and support vector machines is proposed for high-resolution broad-area remote-sensing images [4]. The aircraft's landmark detection is used to address the aircraft type recognition problem [5], which needs fewer labeled data and alleviates the work of human annotation. A novel aircraft type recognition framework based on deep convolutional neural networks is proposed [6]. The idea of fine-grained visual classification is proposed, which attempts to make full use of the features from discriminative object parts [7].

However, none of these methods considers the difference between the training sample and the test sample, and some methods require a large number of labeled data. However, in actual situations, the data obtained are often very different due to the different types of satellites, and the data with labels is very rare. In order to solve the problem of large differences in training and test samples, a transfer learning method was proposed.

Aiming at the research goal of this paper, a Feature Subspace Alignment and Balanced Distribution Adaptation (FSA-BDA) algorithm is proposed to solve this problem. The method first performs feature subspace alignment, and obtains a mapping matrix for the source domain data and the target domain data respectively, and maps the source domain and the target domain data to a low-dimensional common mapping space through different mapping matrices. After the spatial alignment is achieved, the BDA [8] method is used to properly adapt the edge probability and the conditional probability through the weight setting, so that their distribution in the same mapping space is as uniform as possible, and a better effect is obtained.

There are two points different from the other transfer learning methods. First, consider both spatial alignment and probability adaptation, so that the distribution of source and target domains in the same mapping space is as uniform as possible. Second, when the proposed method is spatially aligned, two different mapping matrices are used to map the source and target domains respectively, and the structural information in the source data and the target data can be retained.

Due to the good nature of Gabor features, we extract Gabor features to identify aircrafts. We first proposed FSA algorithm which achieves space alignment and then aligned both marginal distribution and conditional distribution by BDA method, so that the new method FSA-BDA which we proposed can increase the accuracy rate of aircrafts recognition.

2 Related Work

According to the definition of transfer learning, it can be divided into multi-task learning, cross-domain learning and learning under different data distribution [9]. According to the label of data in the target domain, the transfer learning can also be divided into supervised transfer learning and transfer learning. In this paper, we pay attention to the difficult unsupervised transfer learning that the target domain is completely unlabeled. About the domain adaptation problem, some methods align the

space, projecting source data and target data into a new space, such as SA [10] and SDA [11]. But they do not consider the distribution alignment.

Some methods align distribution. For example, the STL [12] reduces the distance of the conditional probability distribution of the source domain and the target domain, thereby completing the migration learning. TCA [13] adapts marginal distribution, and JDA [14] adapts marginal distribution and conditional distribution at the same time.

However, sometimes it cannot achieve satisfaction effect by only considering space alignment or probability distribution adaptation. Therefore, we propose FSA-BDA algorithm. This method considers both spatial alignment and probability adaptation. First, the data in the source domain and the target domain are mapped to the same feature space, and then the balanced distribution adaptation method is used to simultaneously adapt the two domains which are marginal distribution and conditional distribution according to a certain weight. So this method can achieve a better effect.

3 Proposed Method

The method first performs Feature Subspace Alignment (FSA), which maps the source domain and the target domain data to a low-dimensional common mapping space through different mapping matrices. And then, the BDA [8] is used to properly adapt the edge probability and the conditional probability through the weight setting, so that their distribution in the same mapping space is as uniform as possible.

3.1 Feature Spatial Alignment

For data with different distributions, although they are of the same type, the difference in the characteristic representation of the data causes their spatial distribution inconsistency. The FSA finds the appropriate mapping change by solving the feature representation of the data, and maps the source domain data and the target domain data into a common space through feature transformation. The FSA algorithm can construct a low-dimensional common mapping space. The specific algorithm is as follows.

Suppose the label space $C = \{c_1, c_2, \ldots, c_c\}$. The source domain and the target domain data are $\mathbf{D}_s$ and $\mathbf{D}_t$, respectively, $\mathbf{D}_s \neq \mathbf{D}_t$. And the problem is to get the label $\mathbf{Y}_t$ of the target domain. The source domain $\mathbf{X}_s \in \mathbf{R}^{m \times p}$ contains m samples, and $\mathbf{Y}_s$ is the label of $\mathbf{X}_s$. The target domain $\mathbf{X}_s \in \mathbf{R}^{m \times p}$ contains n samples, and p is the dimension of the sample feature. Learning the transformation matrix $\mathbf{T} = \{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_d\} \in \mathbf{R}^{p \times d}$ from the above given conditions, the low-dimensional projection of the original data via the transformation matrix is $\mathbf{Z}_s \in \mathbf{R}^{m \times d}$ and $\mathbf{Z}_t \in \mathbf{R}^{n \times d}$, where d is the dimension of the new feature. Mapping function of the two mapping source domains, the mapping of the target domain function. Find the mapping function $\mathbf{T}_s : \mathbf{X}_s \rightarrow \mathbf{Z}_S \in \mathbf{R}^{m \times d}$ of the source domain, the mapping function $\mathbf{T}_t : \mathbf{X}_t \rightarrow \mathbf{Z}_t \in \mathbf{R}^{n \times d}$ of the target domain.

To get the appropriate mapping changes, minimization (1) is able to align data from different fields in the mapping space.

$$\mathbf{F}_1 = \sum_{\mathbf{x}_i \in D_s} \sum_{\mathbf{x}_j \in D_t} \left\| \mathbf{T}_s^T \mathbf{x}_i - \mathbf{T}_t^T \mathbf{x}_j \right\|_2^2 W(i,j) \tag{1}$$

Where $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are derived from different fields and measure the similarity between $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$. The similarity matrix $\mathbf{W} = \begin{Bmatrix} 0 & \mathbf{W}_a^{s,t} \\ \mathbf{W}_a^{t,s} & 0 \end{Bmatrix}$ is $(m+n) \times (m+n)$. $W_a(i,j)$ represents the similarity of $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$.

Similarly, minimizing Eq. (2) can constrain similar data in the source domain to be mapped to similar locations, while unsimilar data is separated in the mapped space.

$$\mathbf{F}_2 = \sum_{x_i \in D_s} \sum_{x_j \in D_s} \left\| \mathbf{T}_s^T \mathbf{x}_i - \mathbf{T}_s^T \mathbf{x}_j \right\|_2^2 W_s(i,j) \tag{2}$$

Where $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ come from different domains, and $W_s(i,j)$ measures the similarity between $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$.

Similarly, minimizing Eq. (3) can constrain similar data in the target domain to be mapped to similar locations, while unsimilar data is separated.

$$\mathbf{F}_3 = \sum_{\mathbf{x}_i \in D_t} \sum_{\mathbf{x}_j \in D_t} \left\| \mathbf{T}_t^T \mathbf{x}_i - \mathbf{T}_t^T \mathbf{x}_j \right\|_2^2 W_t(i,j) \tag{3}$$

Where $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are from the target domain. If $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are neighbors in the original field of the target domain, they should remain in the neighbors in the mapped space, otherwise they should be away from each other in the mapped space.

Define the objective function as:

$$C(\mathbf{T}_s, \mathbf{T}_t) = F_1 + \alpha F_2 + (1-\alpha) F_3 \tag{4}$$

Where F_1 can make similar data in different fields be aligned to similar positions in the mapping space, F_2 and F_3 maintain the original manifold structure of each domain data in the low-dimensional mapping space. In order to optimizing $C(\mathbf{T}_s, \mathbf{T}_t)$, Definition $\mathbf{T}^T = [\mathbf{T}_S^T, \mathbf{T}_T^T]$, $\mathbf{X}^T = [\mathbf{X}_S^T, \mathbf{X}_T^T]$ and expand the equation as follows.

$$\begin{aligned} F_1 &= \sum_{\mathbf{x}_i \in D_s} \sum_{\mathbf{x}_j \in D_t} \left\| \mathbf{T}_s^T \mathbf{x}_i - \mathbf{T}_t^T \mathbf{x}_j \right\|_2^2 W(i,j) \\ &= 2\mathbf{T}^T \mathbf{X}(\mathbf{P} - \mathbf{W}) \mathbf{X}^T \mathbf{T} \\ &= 2\mathbf{T}^T \mathbf{X} \mathbf{L} X^T \mathbf{T} \end{aligned} \tag{5}$$

Where $\mathbf{P}$ is a diagonal matrix, $P(i,j) = \sum_j W(i,j)$, $\mathbf{L} = \mathbf{P} - \mathbf{W}$.

$$\begin{aligned} \mathrm{F}_2 &= \sum_{\mathbf{x}_\mathrm{i}\in D_s}\sum_{\mathbf{x}_\mathrm{j}\in D_s}\left\|\mathbf{T}_s^{\mathrm{T}}\mathbf{x}_i-\mathbf{T}_s^{\mathrm{T}}\mathbf{x}_j\right\|_2^2 W_s(\mathrm{i},\mathrm{j}) \\ &= 2\mathbf{T}^{\mathrm{T}}\mathbf{X}(\mathbf{P}_s-\mathbf{W}_s)\mathbf{X}^{\mathrm{T}}\mathbf{T} \\ &= 2\mathbf{T}^{\mathrm{T}}\mathbf{X}\mathbf{L}_s\mathbf{X}^{\mathrm{T}}\mathbf{T} \end{aligned} \tag{6}$$

Where $\mathbf{P}_s$ is a diagonal matrix, $\mathrm{P}_s(\mathrm{i},\mathrm{j})=\sum_j \mathrm{W}_\mathrm{s}(\mathrm{i},\mathrm{j})$, $\mathbf{L}_\mathrm{s}=\mathbf{P}_\mathrm{s}-\mathbf{W}_\mathrm{s}$.

$$\begin{aligned} \mathrm{F}_3 &= \sum_{\mathrm{x}_\mathrm{i}\in D_t}\sum_{\mathrm{x}_\mathrm{j}\in D_t}\left\|\mathbf{T}_t^{\mathrm{T}}\mathbf{x}_i-\mathbf{T}_t^{\mathrm{T}}\mathbf{x}_j\right\|_2^2 W_t(\mathrm{i},\mathrm{j}) \\ &= 2\mathbf{T}^{\mathrm{T}}\mathbf{X}(\mathbf{D}_t-\mathbf{W}_t)\mathbf{X}^{\mathrm{T}}\mathbf{T} \\ &= 2\mathbf{T}^{\mathrm{T}}\mathbf{X}\mathbf{L}_t\mathbf{X}^{\mathrm{T}}\mathbf{T} \end{aligned} \tag{7}$$

Where $\mathbf{P}_t$ is a diagonal matrix, $\mathrm{P}_t(\mathrm{i},\mathrm{j})=\sum_j \mathrm{W}_t(\mathrm{i},\mathrm{j})$, $\mathbf{L}_t=\mathbf{P}_t-\mathbf{W}_t$.

Substituting Eqs. (5), (6) and (7) into (4), by mathematical principles of matrix and regularization, Eq. (8) can be obtained.

$$\begin{aligned} \mathrm{C}(\mathbf{T}_S,\mathbf{T}_\mathrm{T}) &= \mathrm{F}_1+\alpha\mathrm{F}_2+(1-\alpha)\mathrm{F}_3 \\ &= \mathbf{T}_t\mathbf{X}\mathbf{L}\mathbf{X}^{\mathrm{T}}\mathbf{T}+\alpha\mathbf{T}^T\mathbf{X}\mathbf{L}_s\mathbf{X}^{\mathrm{T}}\mathbf{T}+(1-\alpha)\mathbf{T}^T\mathbf{X}\mathbf{L}_t\mathbf{X}^{\mathrm{T}}\mathbf{T} \end{aligned} \tag{8}$$

Where α is used to control the contribution of the manifold regularization terms in the two fields. The target vector that needs to be solved in the problem is

$$\begin{aligned} \mathbf{F}^* &= \arg_{\mathrm{T}_t,\mathrm{T}_s}\min\ \mathrm{F}_1+\alpha\mathrm{F}_2+(1-\alpha)\mathrm{F}_3 \\ &= \arg_{\mathrm{T}_t,\mathrm{T}_s}\min\ \mathbf{T}_\mathrm{t}\mathbf{X}\mathbf{L}\mathbf{X}^{\mathrm{T}}\mathbf{T}+\alpha\mathbf{T}^T\mathbf{X}\mathbf{L}_\mathrm{s}\mathbf{X}^{\mathrm{T}}\mathbf{T}+(1-\alpha)\mathbf{T}^T\mathbf{X}\mathbf{L}_\mathrm{t}\mathbf{X}^{\mathrm{T}}\mathbf{T} \end{aligned} \tag{9}$$

Solving the k minimum eigenvalues of (9) gives the obtained mapping function. It can be seen that Eq. (9) is a generalized Rayleigh quotient problem. According to the solution spectrum framework, its solution is the eigenvector corresponding to the largest generalized eigenvalue of Eq. (10).

$$\mathbf{X}(\mathbf{L}+\alpha\mathbf{L}_\mathrm{s}+(1-\alpha)\mathbf{L}_\mathrm{t})\mathbf{X}^{\mathrm{T}} \tag{10}$$

When d > 1, the feature vector corresponding to the first d largest non-zero eigenvalues can be used to form a transformed projection matrix. We can get $\boldsymbol{Z_S}\in\boldsymbol{F_S D_S}$ and $\boldsymbol{Z_T}\in\boldsymbol{F_T D_T}$, which map the two domains to the same space.

3.2 Balanced Distribution Adaptation

After spatial alignment, the BDA [8] method is used to match the edge probability and the conditional probability with different weight ratios, thereby aligning the probability distributions of the features in the same space.

After MA algorithm, the feature space $\chi_S = \chi_T$ and the label space $\mathbf{y}_S = \mathbf{y}_T$, the marginal distribution $P_s(\mathbf{Z}_S) \neq P_t(\mathbf{Z}_T)$ the conditional distribution $P_s(\mathbf{y}_s|\mathbf{Z}_S) \neq P_t(\mathbf{y}_t|\mathbf{Z}_T)$.

The BDA method reduces the difference in marginal distribution and conditional distribution. According to a certain proportion. It can be expressed as follow.

$$\mathrm{D}(\mathbf{Z}_S|\mathbf{Z}_T) = (1-\beta)\mathrm{D}(P_s(\mathbf{Z}_S), P_t(\mathbf{Z}_T)) + \beta\mathrm{D}(P_s(\mathbf{y}_s|\mathbf{Z}_S), P_t(\mathbf{y}_t|\mathbf{Z}_T)) \tag{11}$$

Where $\beta \in [0,1]$ is a balance factor to leverage the different importance of distributions. When $\beta \to 0$, it means the datasets are more dissimilar, so the marginal distribution is more dominant; when $\beta \to 1$ it reveals the datasets are similar, so the conditional distribution is more important to adapt. Therefore, the balance factor β can adaptively leverage the importance of each distribution and lead to good results

Using the class conditional distribution $P_t(Z_T|y_t)$ approximate $P_t(y_t|Z_T)$ according to the sufficient statistics when sample sizes are large. In order to compute $P_t(Z_T|y_t)$, we apply prediction on D_t using some base classifier trained on D_s to get the soft labels for D_t. The soft labels may be less reliable, so we iteratively refine them. Using maximum mean discrepancy (MMD) [14] to compute the marginal and conditional distribution divergences Eq. (11) can be represented as:

$$\mathrm{D}(\mathbf{Z}_S, \mathrm{Z}_T) = (1-\beta)\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_{s_i} - \frac{1}{m}\sum_{j=1}^{m}\mathbf{z}_{t_j}\right\|_H^2 + \beta\sum_{c=1}^{C}\left\|\frac{1}{n_c}\sum_{z_{s_i}\in Z_S^{(c)}}\mathbf{z}_{s_i} - \frac{1}{m_c}\sum_{z_{s_i}\in Z_T^{(c)}}\mathbf{z}_{t_j}\right\|_H^2 \tag{12}$$

Where $Z_S^{(c)}$ and $Z_T^{(c)}$ denote the samples belonging to class c in source and target domain, respectively. n_c and m_c, denote the number of samples. And, n and m stand for the number of samples in the source or target domain. H denotes reproducing kernel Hilbert space. By mathematical principles, Eq. (12) can be formalized as follows.

$$\begin{gathered}\min\ tr\left(\boldsymbol{A}^T\boldsymbol{Z}\left((1-\beta)\boldsymbol{K_0} + \beta\sum_{c=1}^{C}\boldsymbol{K_c}\right)\boldsymbol{Z}^T\boldsymbol{A}\right) + \lambda\|\boldsymbol{A}\|_F^2 \\ s.t. \boldsymbol{A}^T\boldsymbol{Z}\boldsymbol{H}\boldsymbol{Z}^T\boldsymbol{A} = \boldsymbol{I}\end{gathered} \tag{13}$$

Where λ is the regularization parameter with $\|.\|_F^2$ the Frobenius norm. What's more, $\boldsymbol{K_0}$ and $\boldsymbol{K_c}$ are MMD matrices and can be constructed as follows.

$$(\boldsymbol{K_0})_{ij} = \begin{cases} \frac{1}{n^2} & \mathbf{z}_i, \mathbf{z}_j \in \mathbf{D}_s \\ \frac{1}{m^2} & \mathbf{z}_i, \mathbf{z}_j \in \mathbf{D}_t \\ -\frac{1}{mn} & \text{otherwise} \end{cases} \tag{14}$$

$$(\boldsymbol{K_c})_{ij} = \begin{cases} \frac{1}{n_c^2} & \mathbf{z}_i, \mathbf{z}_j \in \mathbf{D}_s^{(c)} \\ \frac{1}{m_c^2} & \mathbf{z}_i, \mathbf{z}_j \in \mathbf{D}_t^{(c)} \\ -\frac{1}{m_c n_c} & \begin{cases} \mathbf{z}_i \in \mathbf{D}_t^{(c)}, \mathbf{z}_j \in \mathbf{D}_s^{(c)} \\ \mathbf{z}_i \in \mathbf{D}_s^{(c)}, \mathbf{z}_j \in \mathbf{D}_t^{(c)} \end{cases} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Where $\boldsymbol{Z}$ stands for the input data matrix composed of source data and target data, $\boldsymbol{A}$ denotes the transformation matrix, I is the identity matrix, and $\boldsymbol{H} = \boldsymbol{I} - (1/n)\mathbf{1}$.

Denote $\Phi = (\varphi_1, \varphi_2, \cdots, \varphi_d)$ as Lagrange multipliers, then Lagrange function for (13) is

$$L = tr\left(\boldsymbol{A}^T\boldsymbol{Z}\left((1-\beta)\boldsymbol{K_0} + \beta\sum_{c=1}^{C}\boldsymbol{K_c}\right)\boldsymbol{Z}^T\boldsymbol{A}\right) + \lambda\|\boldsymbol{A}\|_F^2 + tr((\boldsymbol{I} - \boldsymbol{A}^T\boldsymbol{Z}\boldsymbol{H}\boldsymbol{Z}^T\boldsymbol{A})\boldsymbol{\Phi}) \tag{16}$$

Set derivative $\partial L/\partial \boldsymbol{A} = 0$, the optimization can be derived as a generalized eigen decomposition problem.

$$\left(\boldsymbol{Z}\left((1-\beta)\boldsymbol{K_0} + \beta\sum_{c=1}^{C}\boldsymbol{K_c}\right)\boldsymbol{Z}^T + \lambda\boldsymbol{I}\right)\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{H}\boldsymbol{Z}^T\boldsymbol{A}\boldsymbol{\Phi} \tag{17}$$

Finally, the optimal transformation matrix $\boldsymbol{A}$ can be obtained by solving this equation and finding its d smallest eigenvectors.

By the FSA-BDA method, the source domain feature data and the target domain feature data can be mapped to the same space, and their distribution is also aligned by the BDA method.

4 Experiment and Analysis

4.1 Data Sets

The data sets are aircrafts with different resolutions under different imaging conditions obtained from open satellite maps. Parts of the dataset images are shown in Fig. 1. And the left of F15 aircrafts, F16 aircrafts and B52 aircrafts is 0.5 m resolution. The right of them is 1 m resolution.

The differences of the aircraft in different resolutions are very large. We use the labeled image in the source domain to identify the unlabeled image in the target domain. There are 200 B52 aircrafts, 180 F15 aircrafts and 210 F16 aircrafts in the source domain which are 0.5 m resolution. There are 125 B52 aircrafts, 125 F15 aircrafts and 125 F16 aircrafts in the source domain which are 1 m resolution.

(a)F15

(b)F16

(c)B52

Fig. 1. Parts of the dataset. (a), (b) and (c) shows the three types of aircrafts in images in different resolutions.

4.2 Contrast Methods

The FSA-BDA is compared with five related methods, which are 1-Nearest Neighbor Classifier (NN), Principal Component Analysis(PCA) + NN, Transfer Component Analysis (TCA) + NN [13], Balanced Distribution Adaptation (BDA) + NN [8], Subspace Alignment (SA) + NN [10]. The 1NN classifier is trained on the source domain for identifying target domain ships.

These methods involve the subspace base k or d and the regularization parameters. We set d = k = 110 and $\lambda = 0.6, \alpha = 0.7, \beta = 0.3$ by ten cross-validation. The number of iterations of BDA and the proposed method is set to T = 15. What is more, the classification accuracy is used as an evaluation index, which is calculated as the ratio of the number of correctly identified samples to the total number of input samples in the target domain.

4.3 Experimental Results and Analysis

Results of experiments are discussed in this section. The performance of FSA-BDA algorithm is measured by classification accuracy and MMD distance. The following is the specific description.

Classification Accuracy. The classification accuracy of the three types of aircrafts shown in Table 1 and Fig. 2. The abscissas 1, 2 and 3 represent B52, F15 and F16 aircrafts respectively.

Table 1. Recognition accuracy.

Classification	B52	F15	F16	Average
FSA-BDA	97.6%	96.8%	98.4%	97.6%
JDA	93.6%	92.8%	95.2%	93.9%
TCA	89.6%	88.8%	91.2%	89.9%
SA	88.0%	87.2%	90.4%	88.5%
PCA	78.4%	76.8%	81.6%	78.9%
1NN	75.2%	73.6%	78.4%	75.7%

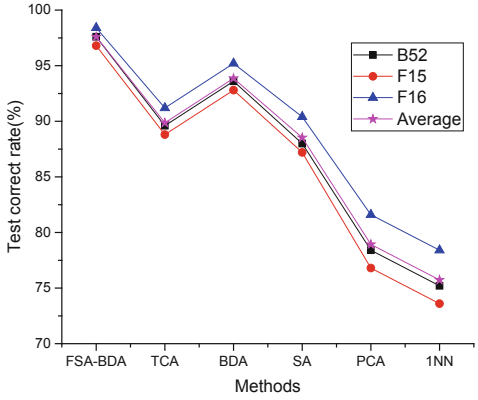
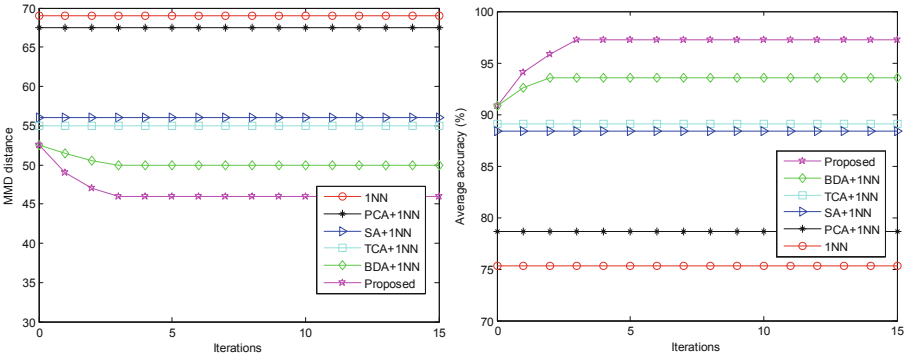


Fig. 2. Recognition accuracy.

The results are reported in Fig. 2 and Table 1 respectively. As illustrated, our method significantly outperforms the existing methods. The average recognition accuracy of the six methods is 97.6%, 93.86%, 89.86%, 88.53%, 78.93% and 75.73%. The accuracy of FSA-BDA on three aircrafts is 97.6%, which performs better than BDA, TCA, SA, PCA and 1NN. PCA and 1NN are traditional methods. SA only considers spatial alignment in adaptive domain problems. BDA minimizes the marginal and conditional allocation differences according to a certain weight between domains, while TCA can only minimize marginal differences. However, neither BDA nor TCA considers space adjustments. FSA-BDA considers spatial alignment and distribution differences. Therefore, theoretical analysis can explain the rationality of experimental results.

MMD Distance. MMD distance [14] is used to verify the distribution suitability of FSA-BDA and the other five methods. As shown in Fig. 3, the MMD distance and the average accuracy of NN, PCA, TCA, BDA, SA, and FSA-BDA increased with the number of iterations are shown. The abscissas stand for iterations.

The proposed method is compared with state-of-the-art methods in Fig. 3. Experiments demonstrate the efficacy of our method. All transfer learning methods can reduce the MMD distance. When only NN is used, the distance between the source domain and the target domain is large. Since TCA minimizes marginal differences, SA



(a) MMD distance (b) Recognition accuracy

Fig. 3. Performance of methods. Fig. (a) and (b) shows MMD distance and recognition accuracy.

aligns space, and BDA minimizes marginal and conditional distribution differences. FSA-BDA not only adjusts the space between the source and target domains, but also minimizes differences in marginal and conditional distribution according to a certain weight. And it can be found that the smaller the MMD distance, the higher the correct recognition rate. The MMD distance of FSA-BDA is significantly smaller than other methods. It shows FSA-BDA achieves the best performance in the aircrafts recognition problems.

5 Conclusion

In this paper, we propose a method called the Feature Subspace Alignment and Balanced Distribution Adaptation (FSA-BDA) for recognizing aircrafts with different resolutions. This method not only considers spatial alignment, but also considers probabilistic adaptation, so it achieves better results. The feature subspace alignment (FSA) maps the source domain and the target domain data to a low-dimensional common mapping space through different mapping matrices, in order to preserve the structural information in the source data and the target data. The BDA method is used to properly adapt the edge probability and the conditional probability through the weight setting, so that their distribution in the same mapping space is as uniform as possible. Extensive experiments demonstrate that the FSA-BDA method is better than the state-of-the-art transfer learning methods. In the future, we will continue to study the aircraft target recognition problem combined with transfer learning, such as the targets from different sources.

Acknowledgements. This work was supported in part by a grant from the Defense Industrial Technology Development Program (No. JCKY2016603C004).

References

1. Dudani, S.A., Breeding, K.J., McGhee, R.B.: Aircraft identification by moment invariants. IEEE Trans. Comput. **100**(1), 39–46 (1977)
2. Liu, G., Sun, X., et al.: Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior. IEEE Geosci. Remote Sens. Lett. **10**(3), 573–577 (2013)
3. Wu, Q., Sun, H., Sun, X., et al.: Aircraft recognition in high-resolution optical satellite remote sensing images. IEEE Geosc. Remote Sens. Lett. **12**(1), 112–116 (2015)
4. Zhang, L., Zhang, Y.: Airport detection and aircraft recognition based on two-layer saliency model in high spatial resolution remote-sensing images. IEEE J. Sele. Top. Appl. Earth Obs. Remote Sens. **10**(4), 1511–1524 (2017)
5. Zhao, A., Fu, K., Wang, S., et al.: Aircraft recognition based on landmark detection in remote sensing images. IEEE Geosci. Remote Sens. Lett. **14**(8), 1413–1417 (2017)
6. Zuo, J., Xu, G., Fu, K., et al.: Aircraft type recognition based on segmentation with deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett. **15**(2), 282–286 (2018)
7. Fu, K., Dai, W., Zhang, Y., et al.: MultiCAM: multiple class activation mapping for aircraft recognition in remote sensing images. Remote Sens. **11**(5), 544 (2019)
8. Wang, J., Chen, Y.: Balanced distribution adaptation for transfer learning. In: 2017 IEEE International Conference on Data Mining, pp. 1129–1134 (2017)
9. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
10. Fernando, B., Habrard, A., Sebban, M., Tuytelaars, T.: Unsupervised visual domain adaptation using subspace alignment. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2960–2967 (2013)
11. Sun, B., Saenko, K.: Subspace: distribution alignment for unsupervised domain adaptation. In: BMVC, p. 24:1 (2015)
12. Wang, J., Chen, Y., Hu, L., Peng, X., Philip, S.Y.: Stratified transfer learning for cross-domain activity recognition. In: 2018 IEEE International Conference on Pervasive Computing and Communications, pp. 1–10 (2018)
13. Pan, S.J., Tsang, I.W., Kwok, J.T., et al.: Domain adaptation via transfer component analysis. IEEE Trans. Neural Netw. **22**(2), 199–210 (2011)
14. Long, M., Wang, J., Ding, G., Sun, J., Yu, P.S.: Transfer feature learning with joint distribution adaptation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2200–2207 (2013)