



# Balanced Discriminative Transfer Feature Learning for Visual Domain Adaptation

**Abstract:** Transfer learning aims to transfer source models to a target domain. Leveraging the feature matching can alleviate the domain shift effectively, but this process ignores the relationship of the marginal distribution matching and the conditional distribution matching. Simultaneously, the discriminative information of both domains is also neglected, which is important for improving the performance on the target domain. In this paper, we propose a novel method called Balanced Discriminative Transfer Feature Learning for Visual Domain Adaptation (BDTFL). The proposed method can adaptively balance the relationship of both distribution matchings and capture the category discriminative information of both domains. Therefore, balanced feature matching can achieve more accurate feature matching and adaptively adjust itself to different scenes. At the same time, discriminative information is exploited to alleviate category confusion during feature matching. And with assistance of the category discriminative information captured from both domains, the source classifier can be transferred to the target domain more accurately and boost the performance of target classification. Extensive experiments show the superiority of BDTFL on popular visual cross-domain benchmarks.

**Keywords:** transfer learning; domain adaptation; distribution adaptation; discriminative information

SU Limin<sup>1</sup>, ZHANG Qiang<sup>2</sup>, LI Shuang<sup>1</sup>,  
Chi Harold LIU<sup>1</sup>

(1. Beijing Institute of Technology, Beijing 100081, China;

2. ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTECOM.202004011

[http://kns.cnki.net/kcms/detail/34.1294.](http://kns.cnki.net/kcms/detail/34.1294.TN.20201205.1851.002.html)

TN.20201205.1851.002.html, published online  
December 5, 2020

Manuscript received: 2019-12-28

**Citation** (IEEE Format): L. M. Su, Q. Zhang, S. Li, et al., "Balanced discriminative transfer feature learning for visual domain adaptation," *ZTE Communications*, vol. 18, no. 4, pp. 78 - 83, Dec. 2020. doi: 10.12142/ZTECOM.202004011.

## 1 Introduction

Labeled data are very important for supervised domain adaptation. However, it is time-consuming and expensive to annotate data manually in real application. To address this limitation, transfer learning becomes an effective alternative choice by applying the knowledge extracted from an available well-labeled source domain to a target domain. There is a common assumption in transfer learning that source data and target data belong to related but different distributions<sup>[1]</sup>. This can provide a promising approach to the scenarios that suffer from a shortage of labeled data<sup>[1-5]</sup>.

Most existing distribution adaptation methods achieve either domain-wise alignment, class-wise alignment or both of them. Both theoretical verification and experiment results indicate the adapting marginal distribution and conditional distribution of both domains could obtain better performance. However, domain-wise alignment and class-wise alignment are often considered as equally important. In this case, the discrepancy of these two distributions is ignored. In fact, marginal dis-

tribution matching and conditional distribution matching should be given different weights under different scenarios. Under the circumstance that the correlation between the source domain and the target domain is lower, marginal distribution needs to be more valued. When the target domain is closely similar to the source domain, conditional distribution should be given more attention. If we notice this situation and give different weights to these two domain matching, the alignment of source and target domains will be more accurate.

Achieving domain matching simultaneously brings a new problem that different categories may be wrongly matched. Therefore, we should not only achieve feature matching, but also need the learned feature to be class-discriminative. Once the feature of samples is class-discriminative, points belonging to the same class would be clustered into compact clusters and different clusters would move away from each other, which is helpful to classification of the model.

In order to solve above problems, we propose a novel method called Balanced Distribution Transfer Feature Learning

(BDTFL). BDTFL can not only balance the relationship of both distribution matchings, but also encourage the learned features to be class-discriminative. On one hand, for different transfer learning tasks, BDTFL can effectively adjust relative weight of both distribution matchings. On the other hand, we also alleviate category confusion during domain-wise alignment by leveraging category discriminative information of both domains. When the learning feature appears class-discriminative, discriminative information of samples is exploited to boost the classification performance. Several existing methods can be regarded as special cases of BDTFL. To evaluate the performance of BDTFL, we conduct comprehensive experiments on several visual cross-domain classification tasks. The results verify that our method can outperform the baselines significantly.

We summarize our contributions as the following three points:

(1) We propose a novel transfer learning method BDTFL to balance the marginal and conditional distribution matching. BDTFL can adaptively adjust relationship of both distribution matching and achieve more accurate alignment. Thus, several transfer learning methods can be treated as special cases of BDTFL.

(2) Our method adopts the triplet loss to exploit the discriminative information among categories. When the learned features with discriminative information are extracted to match the marginal and conditional distributions across domains, we encourage the smaller distance between samples with the same label and the larger distance between samples with different labels, which improves classification effect of the model on the target domain.

(3) Comprehensive experiments on several visual cross-domain datasets demonstrate the superiority compared with other state-of-the-art methods.

## 2 Related Work

The prior work has achieved a promising process in the field of domain adaptation. Existing domain adaptation methods can be roughly divided into two categories: instance reweighting and feature extraction. Our proposed method BDTFL belongs to the feature-based learning methods. Thus, we will focus on feature extraction approaches to discuss.

A transfer component analysis (TCA)<sup>[6]</sup> approach is proposed to minimize the distance between source and target domains using the maximum mean discrepancy metric<sup>[7]</sup> by searching transfer component. Furthermore, LONG et al.<sup>[8]</sup> put forward joint distribution adaptation (JDA) to jointly match marginal and conditional distributions of source domain and target domain in a principal dimension reduction procedure. Based on JDA, transfer joint matching (TJM), proposed in Ref. [9], reweights source instances to reduce differ-

ence between two domains.

To capture the discriminative information of two domains, Ref. [10] adds the supervised source class cluster to the domain matching process. Furthermore, a domain-invariant and class-discriminative (DICD) approach<sup>[11]</sup> tries to simultaneously maximize the inter-class dispersion and minimize the intra-class variance of source and target data to mitigate the domain shift. Inspired by DICD, our method explores class-discriminative information during the feature matching in the latent space.

However, the above methods ignore the relationship of two distribution matching for different scenes. They just simply and roughly add metrics together. Different from the above methods, our method can adjust the relationship of two distribution matching according to different scenes and capture class-discriminative information of two domains.

## 3 Proposed Approach

In this section, we first demonstrate the problem definition and then show how to learn adaptive domain-invariant feature representation with category discriminative information in our method.

### 3.1 Preliminary and Motivation

In unsupervised domain adaptation problems, given source labeled data  $\mathcal{D}_s = \{(x_{si}, y_{si})\}_{i=1}^{n_s}$  and target unlabeled data  $\mathcal{D}_t = \{(x_{tj})\}_{j=1}^{n_t}$ , where  $n_s$  and  $n_t$  are the numbers of source and target samples respectively, it is assumed that  $x_{si}, x_{tj} \in X, (X \subset \mathbb{R}^m)$  and  $y_{si} \in Y (Y \subset \mathbb{R})$ .  $y_{si}$  is the label of the source sample  $x_{si}$ ;  $X$  is the feature space and  $Y$  is the label space.

Since there exists a domain shift across domains, it is obvious that marginal distributions satisfy  $P_s(x_s) \neq P_t(x_t)$  with condition distributions  $P_s(y_s|x_s) \neq P_t(y_t|x_t)$ . We aim to find a discriminative feature transformation  $\mathbf{A}$ , utilizing adaptive joint feature matching to mitigate the domain shift. In this way, the classifier can be well transferred to the target domain, which results in better target classification.

During the feature extraction process, we focus on the following two points:

(1) Effective feature extraction can alleviate domain shift. For different scenes, the relationship of the marginal distribution and conditional distribution should be taken into consideration. When source and target domains match in the embedding space, our method balances the relationship of the marginal distribution and conditional distribution, which is different from JDA.

(2) To further improve the performance of the model, the samples sharing the same label are expected to gather closer and clusters with different labels to be far away from each other. Therefore, we extract the category-discriminative informa-

tion by minimizing the intra-class dispersion and maximizing the inter-class compactness.

### 3.2 Learning Adaptive Domain-Invariant Features

We are devoted to extract adaptive domain-invariant feature representation such that the source model trained by source data can always be well applied to target data in different scenarios. In this way, we search a linear domain-invariant project to discover shared subspace where the domain shift is effectively alleviated.

#### 3.2.1 Joint Feature Matching

Since the source and target domains are related but different, it is essential to match the marginal distributions and conditional distributions across domains to achieve overlap between the source domain and target domain. The metric Maximum Mean Discrepancy is adopted to measure the distribution distance. The mean discrepancy of the marginal distributions and conditional distributions is represented as:

$$\begin{aligned} \mathcal{Z}_d &= \mathcal{Z}_0 + \sum_{k=1}^C \mathcal{Z}_k = \\ & \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \mathbf{A}^T x_{si} - \frac{1}{n_t} \sum_{j=1}^{n_t} \mathbf{A}^T x_{tj} \right\|^2 + \\ & \sum_{k=1}^C \left\| \frac{1}{n_s^k} \sum_{x_{si} \in \mathcal{D}_s^k} \mathbf{A}^T x_{si}^k - \frac{1}{n_t^k} \sum_{x_{tj} \in \mathcal{D}_t^k} \mathbf{A}^T x_{tj}^k \right\|^2 = \\ & \text{Tr}(\mathbf{A}^T \mathbf{X} (\mathbf{P}_0 + \sum_{k=1}^C \mathbf{P}_k) \mathbf{X}^T \mathbf{A}), \end{aligned} \quad (1)$$

where  $\mathbf{X}_S$  and  $\mathbf{X}_T$  are concatenated by column as  $\mathbf{X}$ . The marginal distribution distance matrix  $\mathbf{P}_0$  and the conditional distribution distance matrix  $\mathbf{P}_k$  can be formulated respectively as:

$$\mathbf{P}_0 = \begin{bmatrix} \frac{1}{n_s^2} \mathbf{1}_{n_s \times n_s} & -\frac{1}{n_s n_t} \mathbf{1}_{n_s \times n_t} \\ -\frac{1}{n_s n_t} \mathbf{1}_{n_t \times n_s} & \frac{1}{n_t^2} \mathbf{1}_{n_t \times n_t} \end{bmatrix}, \quad (2)$$

$$(\mathbf{P}_k)_{ij} = \begin{cases} \frac{1}{(n_s^k)^2}, & \text{if } x_i, x_j \in \mathcal{D}_s^k, \\ -\frac{1}{n_s^k n_t^k}, & \text{if } x_i \in \mathcal{D}_s^k, x_j \in \hat{\mathcal{D}}_t^k, \\ -\frac{1}{n_s^k n_t^k}, & \text{if } x_i \in \hat{\mathcal{D}}_t^k, x_j \in \mathcal{D}_s^k, \\ \frac{1}{(n_t^k)^2}, & \text{if } x_i, x_j \in \hat{\mathcal{D}}_t^k, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where  $\mathbf{1}_{m \times n}$  is defined as  $m \times n$  matrix filled with one.

We need to minimize Eq. (1) to achieve marginal and conditional distribution matching in extracted feature subspace. However, in this process, the marginal distribution and conditional distribution are equally important. In fact, their importance varies from different scenarios. The relationship between marginal distribution matching and conditional distribution matching should be balanced for different scenarios.

#### 3.2.2 Balanced Feature Matching

BDTFL innovatively adjusts the relative magnitude of marginal distribution matching and conditional distribution matching to adapt different scenarios. Specifically, through a factor  $\rho$ , we adaptively adjust the relative magnitude of marginal distribution matching and conditional distribution matching. In this way, BDTFL can satisfy different requirements of various scenarios.

By defining  $\mathbf{P} = \mathbf{P}_0 + \rho \sum_{k=1}^C \mathbf{P}_k$ , the specific form of Eq. (1) is formulated as:

$$\begin{aligned} \mathcal{Z}_d &= \mathcal{Z}_0 + \rho \sum_{k=1}^C \mathcal{Z}_k = \\ \text{Tr}(\mathbf{A}^T \mathbf{X} (\mathbf{P}_0 + \rho \sum_{k=1}^C \mathbf{P}_k) \mathbf{X}^T \mathbf{A}) &= \\ \text{Tr}(\mathbf{A}^T \mathbf{X} \mathbf{P} \mathbf{X}^T \mathbf{A}). \end{aligned} \quad (4)$$

When the domain discrepancy is too large, there exists large diversity of class-wise distributions across domains. Class-wise alignment would then bring about serious class confusion. As a result, the class-wise alignment does not make any sense and even causes negative transfer problem. Thus, the class-wise alignment should be weakened in this case. However, if the target domain is closely similar to the source domain, the class-wise alignment should be enhanced. In that case, we should pay more attention to conditional distribution matching. The application scenarios can be greatly expanded by balancing marginal distribution matching and conditional distribution matching. We should estimate the similarity of the source and target domains according to different scenarios and then select  $\rho$ . In this paper, the values of  $\rho$  are selected according to the performance in experiments.

#### 3.2.3 Learning Category-Discriminative Information

Completing the above steps, we simultaneously align marginal and conditional distributions across domains to alleviate the domain shift. We also expect the learned feature representations in the latent shared subspace to be discriminative. On one hand, if samples sharing the same label are more concentrated and clusters sharing different labels are more dispersed, the source classifier can be well generalized to the target domain and achieves better performance on the target domain. On the other hand, with the assistance of the discriminative information captured from both domains, conditional distribu-

tion matching can be more accurate.

Inspired by triplet loss<sup>[11-13]</sup>, we adopt the strategy that makes the samples sharing the same label close and clusters with different labels far away from each other in the extracted feature subspace. The corresponding matrix of the distance loss of source data can be formulated as:

$$\mathcal{Z}_s = \sum_{k=1}^C \frac{n_s}{n_s} \sum_{\mathbf{y}_{si}, \mathbf{y}_{sj} = k} \|\nu_{si} - \nu_{sj}\|^2 + \sum_{\mathbf{y}_{si} \neq \mathbf{y}_{sj}} \|\nu_{si} - \nu_{sj}\|^2 = \text{Tr}(\mathbf{A}^T \mathbf{X}_s (\mathbf{Q}_1^S - \mathbf{Q}_2^S) \mathbf{X}_s^T \mathbf{A}), \quad (5)$$

where  $\mathbf{Q}_1^S$  and  $\mathbf{Q}_2^S$  can be calculated as:

$$(\mathbf{Q}_1^S)_{ij} = \begin{cases} n_s, & \text{if } i = j, \\ \frac{n_s}{n_s^k}, & \text{if } i \neq j, \mathbf{y}_{si} = \mathbf{y}_{sj} = k, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

$$(\mathbf{Q}_2^S)_{ij} = \begin{cases} n_s - n_s^k, & \text{if } i = j, \mathbf{y}_{si} = k, \\ -1, & \text{if } i \neq j, \mathbf{y}_{si} \neq \mathbf{y}_{sj}, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

In the target domain, we utilize the pseudo label to make similar derivation and definition. We define  $\mathbf{Q}_1 = \text{diag}(\mathbf{Q}_1^S, \mathbf{Q}_1^T)$  and  $\mathbf{Q}_2 = \text{diag}(\mathbf{Q}_2^S, \mathbf{Q}_2^T)$ , and the whole category-discriminative loss for both domains is obtained as:

$$\mathcal{Z}_c = \mathcal{Z}_s + \mathcal{Z}_t = \text{Tr}(\mathbf{A}^T \mathbf{X} (\mathbf{Q}_1 - \mathbf{Q}_2) \mathbf{X}^T \mathbf{A}) = \text{Tr}(\mathbf{A}^T \mathbf{X} \mathbf{Q} \mathbf{X}^T \mathbf{A}), \quad (8)$$

where  $\mathbf{Q} = \mathbf{Q}_1 - \mathbf{Q}_2$ .

The minimization of Eq. (8) encourages intra-class compactness and inter-class dispersion to enhance the discriminative characteristic of the learned representation.

### 3.2.4 Overall Framework

We incorporate Eqs. (4) and (8) into one formulation, which represents the overall loss of BDTFL. Then we minimize this overall loss as:

$$\min_{\mathbf{A}} \mathcal{Z}_d + \mathcal{Z}_c + \beta \|\mathbf{A}\|^2 \quad \text{s.t. } \mathbf{A}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{A} = \mathbf{I}_d, \quad (9)$$

where  $\mathbf{I}_d$  represents an identity matrix of dimension  $d$  and  $\mathbf{H} = \mathbf{I}_{(n_s + n_t)} - [1/(n_s + n_t)] \mathbf{1}_{(n_s + n_t) \times (n_s + n_t)}$  represents the centering matrix. Besides,  $\beta$  is the regularization coefficient for avoiding overfitting.

If we incorporate the matrix forms of Eqs. (4) and (8), Eq. (9) can be rewritten as:

$$\min_{\mathbf{A}} \text{Tr}(\mathbf{A}^T \mathbf{X} (\mathbf{P} + \mathbf{Q}) \mathbf{X}^T \mathbf{A}) + \beta \|\mathbf{A}\|^2 \quad \text{s.t. } \mathbf{A}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{A} = \mathbf{I}_d. \quad (10)$$

The constrained optimization problem (10) can be regarded as a generalized eigen-decomposition problem. The optimal solution *w.r.t.*  $\mathbf{A}$  ( $\mathbf{A} \in \mathbb{R}^{m \times d}$ ) can be solved as:

$$(\mathbf{X} (\mathbf{P} + \mathbf{Q}) \mathbf{X}^T + \beta \mathbf{I}_m) \mathbf{A} = \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{A} \Theta, \quad (11)$$

where  $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_d) \in \mathbb{R}^{d \times d}$  is a diagonal matrix with Lagrange Multipliers. The optimization problem (11) is also a generalized eigen-decomposition problem. And the optimal solution is the generalized eigenvectors of Eq. (11) corresponding to the  $d$ -smallest eigenvalues. BDTFL can be kernelized to be applied to nonlinear scenarios. Through the above steps, we achieve the joint feature matching simultaneously retaining the discriminative information in BDTFL.

Recently, deep neural networks have been applied into transfer learning. Fine-tuning can be effectively used to transfer knowledge extracted from a source dataset to a target task. Several factors that influence the performance of fine-tuning are systematically investigated in Ref. [14], including the distribution of source and target data. In recent works on deep domain adaptation, discrepancy measures are embedded into deep architectures to reduce the gap between the source domain and target domain. Theoretically, BDTFL can be combined with the deep neural network to learn better feature representation, which will be further explored in the future.

## 4 Experiments

In this section, we evaluate the performance of BDTFL with massive state-of-the-art shallow domain adaptation methods on several popular visual cross-domain benchmarks.

### 4.1 Description of Datasets

Two widely-used benchmarks are adopted in our experiments, including Office-31(DeCAF7)<sup>[15]</sup> and CMU-PIE<sup>[16]</sup>. Office-31 includes 4 652 images with 31 categories composed of three domains: Amazon (A), Webcam (W) and DSLR (D). CMU-PIE contains more than 40 000 face images from 68 individuals.

Considering different pose factors, we select 5 out of 13 poses as Ref. [8]. The detail of the datasets is shown in **Table 1**.

▼ **Table 1. Cross-domain datasets used in the experiments**

Dataset	Number of Samples	Number of Features	Number of Classes	Domain
Office-31 (DeCAF7)	4 652	4 096	31	A, W, D
CMU-PIE	11 554	1 024	68	C05, C07, C09, C27, C29

### 4.2 Comparison Methods

We compare our method with the following seven methods:

- 1-Nearst Neighbor Classifier (INN)<sup>[17]</sup>;
- Pricipal Component Analysis (PCA)<sup>[18]</sup>;
- Geodesic Flow Kernel (GFK)<sup>[19]</sup>;
- Transfer Component Analysis (TCA)<sup>[6]</sup>;
- Joint Distribution Adaptation (JDA)<sup>[8]</sup>;
- Subspace Alignment (SA)<sup>[20]</sup>;
- Discriminative Transfer Subspace Learning (DTSL)<sup>[21]</sup>.

Among these seven methods, the first two are traditional transfer learning methods and the others are state-of-the-art shallow transfer learning methods.

### 4.3 Implementation Details

INN is acted as the final classifier, while INN, PCA, GFK, TCA, SA, DTSL and BDTFL are acted as feature extractors. We search  $\rho$  in  $\{0.001, 0.01, \dots, 100\}$  for BDTFL. After the reduction of dimensionality, the feature number  $d$  is fixed as 100. When extracting the feature, we use the linear kernel in all the experiments. The iteration number  $T$  in the experiments is fixed as 10. The target classification accuracy is considered as the evaluation metric of the methods.

### 4.4 Experiment Results

The following experiment results present the superiority of the proposed method in this paper.

(1) Results on the Office-31 dataset (**Table 2**): On almost all the tasks, BDTFL achieves great advantage compared with the others. On  $A \rightarrow W$  and  $W \rightarrow A$  tasks, our method obtains a significant performance improvement of 7.66% and 5.19% compared to the best baseline SA. The target average classification accuracy of our method reaches 70.47%, 2.57% more than the best baseline. BDTFL performs better than JDA on each task, which verifies the discriminative information is vital to the transfer feature learning.

(2) Results on the CMU-PIE dataset (**Table 3**): On the most tasks, BDTFL performs better than the others. The target aver-

▼Table 2. Accuracy (%) on Office-31(Decaf,) Datasets

Task/Method	INN	PCA	GFK	TCA	JDA	SA	DTSL	BDTFL
A→D	59.6	60.6	52.0	56.2	56.8	61.0	60.0	62.65
A→W	54.0	55.6	48.2	54.8	58.1	59.5	54.5	58.62
D→A	42.4	44.6	41.8	44.3	44.8	46.9	46.6	54.56
D→W	90.9	91.8	86.5	92.1	95.7	95.1	94.3	96.60
W→A	40.8	42.0	38.6	42.1	46.2	46.6	45.6	51.79
W→D	97.8	98.6	87.5	95.6	97.6	98.2	94.0	98.59
Average	64.3	65.5	59.1	64.2	66.5	67.9	65.8	70.47

INN: 1-Nearst Neighbor Classifier  
 BDTFL: Balanced Discriminative Transfer Feature Learning for Visual Domain Adaptation  
 DTSL: Discriminative Transfer Subspace Learning  
 GFK: Geodesic Flow Kernel  
 JDA: Joint Distribution Adaptation  
 PCA: Pricipal Component Analysis  
 SA: Subspace Alignment  
 TCA: Transfer Component Analysis

▼Table 3. Accuracy (%) on CMU-PIE Datasets

Task/Method	INN	PCA	GFK	TCA	JDA	SA	DTSL	BDTFL
C05→C07	26.1	24.8	26.2	40.8	58.6	26.8	65.9	67.3
C05→C09	26.6	25.2	27.3	41.8	52.0	28.2	64.1	68.1
C05→C27	30.7	29.3	31.2	59.6	83.7	30.9	82.0	89.9
C05→C29	16.7	16.3	17.6	29.4	47.7	19.6	54.9	58.6
C07→C05	24.5	24.2	25.2	41.8	60.6	26.4	45.0	59.8
C07→C09	46.6	45.5	47.4	51.5	60.2	48.0	53.5	62.7
C07→C27	54.1	53.4	54.3	64.7	75.4	54.3	71.4	80.4
C07→C29	26.5	25.4	27.1	33.7	40.9	28.2	48.0	49.6
C09→C05	21.4	21.0	21.8	34.7	50.9	23.2	52.5	57.7
C09→C07	41.0	40.5	43.2	47.7	56.1	44.3	55.6	61.1
C09→C27	46.5	46.1	46.4	56.2	68.0	46.2	77.5	76.1
C09→C29	26.2	25.3	26.7	33.2	40.3	28.9	54.1	53.2
C27→C05	33.0	32.0	34.2	55.6	81.0	36.3	81.5	86.8
C27→C07	62.7	61.0	62.9	67.8	82.8	63.8	85.4	86.6
C27→C09	73.2	72.2	73.4	75.9	87.2	73.2	82.2	86.5
C27→C29	37.2	35.1	37.4	40.3	49.9	38.1	72.6	72.8
C29→C05	18.5	18.9	20.4	27.0	47.5	23.4	52.2	59.1
C29→C07	24.2	23.4	24.6	30.0	44.8	25.5	49.4	53.6
C29→C09	28.3	27.2	28.5	30.0	48.1	28.6	58.5	57.6
C29→C27	31.2	30.3	31.3	33.6	56.5	31.2	64.3	66.0
Average	34.8	33.9	35.4	44.8	59.6	36.3	63.5	67.7

INN: 1-Nearst Neighbor Classifier  
 BDTFL: Balanced Discriminative Transfer Feature Learning for Visual Domain Adaptation  
 DTSL: Discriminative Transfer Subspace Learning  
 GFK: Geodesic Flow Kernel  
 JDA: Joint Distribution Adaptation  
 PCA: Pricipal Component Analysis  
 SA: Subspace Alignment  
 TCA: Transfer Component Analysis

age classification accuracy is 67.7%, 4.2% more than the best baseline DTSL. On  $C05 \rightarrow C27$  and  $C29 \rightarrow C05$  tasks, our method obtains significant performance improvements of 6.2% and 6.9% compared to the best baseline DTSL.

## 5 Conclusions

In this paper, we propose a novel method called BDTFL, which adaptively balances the relationship of the marginal distribution matching and the conditional distribution matching and captures the category discriminative information. The adaptive feature matching encourages the overlap between the source domain and target domain to be more accurate. And the category discriminative information can make samples sharing the same label close and clusters with different labels far away from each other, which can further boost the conditional distribution matching across domains. The discriminative information is exploited to alleviate category confusion during the domain-wise alignment. With the assistance of discriminative information, the source classifier can be well transferred to the target domain and boost target classification performance. Our experiments verify the superiority of BDTFL.



**References**

- [1] PAN S J, YANG Q. A survey on transfer learning [J]. *IEEE transactions on knowledge and data engineering*, 2010, 22(10): 1345 – 1359. DOI: 10.1109/TKDE.2009.191
- [2] GONG B Q, GRAUMAN K, SHA F. Connecting the dots with landmarks: discriminatively learning domain-invariant features for unsupervised domain adaptation [C]//30th International Conference on International Conference on Machine Learning. Atlanta, USA: Curran Associates, 2013: 222 – 230
- [3] DUAN L X, TSANG I W, XU D. Domain transfer multiple kernel learning [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2012, 34(3): 465 – 479. DOI: 10.1109/TPAMI.2011.114
- [4] TAN B, ZHANG Y, PAN S J, et al. Distant domain transfer learning [C]//31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI Press, 2017: 2604 – 2610
- [5] CHANG W-C, WU Y X, LIU H X, et al. Cross-domain kernel induction for transfer learning [C]//31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI Press, 2017: 1763 – 1769
- [6] PAN S J, TSANG I W, KWOK J T, et al. Domain adaptation via transfer component analysis [J]. *IEEE transactions on neural networks*, 2011, 22(2): 199 – 210. DOI: 10.1109/TNN.2010.2091281
- [7] GRETTON A, BORGWARDT K M, RASCH M, et al. A kernel method for the two-sample-problem [C]//19th International Conference on Neural Information Processing Systems. Kitakyushu, Japan: Springer, 2007: 513 – 520
- [8] LONG M S, WANG J M, DING G G, et al. Transfer feature learning with joint distribution adaptation [C]//IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013: 2200 – 2207. DOI: 10.1109/ICCV.2013.274
- [9] LONG M S, WANG J M, DING G G, et al. Transfer joint matching for unsupervised domain adaptation [C]//IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014: 1410 – 1417. DOI: 10.1109/CVPR.2014.183
- [10] BAKTASHMOTLAGH M, HARANDI M T, LOVELL B C, et al. Unsupervised domain adaptation by domain invariant projection [C]//IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013: 769 – 776. DOI: 10.1109/ICCV.2013.100
- [11] LI S, SONG S J, HUANG G, et al. Domain invariant and class-discriminative feature learning for visual domain adaptation [J]. *IEEE transactions on image processing*, 2018, 27(9): 4260 – 4273. DOI: 10.1109/TIP.2018.2839528
- [12] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering [EB/OL]. (2015-03-12)[2018-01-01]. <https://arxiv.org/abs/1503.03832>
- [13] XIAO Q Q, LUO H, ZHANG C. Margin sample mining loss: a deep learning based method for person Re-identification [EB/OL]. (2017-10-02)[2018-01-01]. <https://arxiv.org/abs/1710.00478>
- [14] LI X Y, HERRANZ L, JIANG S Q. Multifaceted analysis of fine-tuning in a deep model for visual recognition [J]. *IMS transactions on data science*, 2020, 1(1): 1 – 22. DOI: 10.1145/3319500
- [15] SAENKO K, KULIS B, FRITZ M, et al. Adapting visual category models to new domains [C]//11th European Conference on Computer Vision. Heraklion, Greece: IEEE, 2010: 213 – 226
- [16] SIM T, BAKER S, BSAT M. The CMU pose, illumination, and expression (PIE) database [C]//Fifth IEEE International Conference on Automatic Face and Gesture Recognition. Washington, USA: IEEE, 2002. DOI: 10.1109/AFGR.2002.1004130
- [17] FUKUNAGA K, NARENDRA P M. A branch and bound algorithm for computing k-nearest neighbors [J]. *IEEE transactions on computers*, 1975, C-24(7): 750 – 753. DOI: 10.1109/T-C.1975.224297
- [18] JOLLIFFE I T. *Principal component analysis* [M]. New York, USA: Springer-Verlag, 2011: 1094 – 1096. DOI: 10.1007/b98835
- [19] GONG B, SHI Y, SHA F, et al. Geodesic flow kernel for unsupervised domain adaptation [C]//IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012: 2066 – 2073
- [20] FERNANDO B, HABRARD A, SEBBAN M, et al. Unsupervised visual domain adaptation using subspace alignment [C]//IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013: 2960 – 2967. DOI: 10.1109/ICCV.2013.368
- [21] XU Y, FANG X Z, WU J, et al. Discriminative transfer subspace learning via low-rank and sparse representation [J]. *IEEE transactions on image processing*, 2016, 25(2): 850 – 863. DOI: 10.1109/TIP.2015.2510498

**Biographies**

**SU Limin** is pursuing the M.Sc. degree in computer science at Beijing Institute of Technology, China. Her research interests include machine learning and transfer learning.

**ZHANG Qiang** received his B.S. degree in computer science from Hefei University of Technology, China in 1995. Working with ZTE Corporation, he was the director of the research team of TMN Standards from 2003 to 2006 and has been an NMS/OSS solution manager since 2007.

**LI Shuang** (shuangli@bit.edu.cn) received his Ph.D. degree in Department of Automation, Tsinghua University, China in 2018. He is an assistant professor in the School of Computer Science and Technology, Beijing Institute of Technology, China. He was a visiting research scholar at the Department of Computer Science, Cornell University, USA from November 2015 to June 2016. His main research interests include machine learning and deep learning, especially in transfer learning and domain adaptation.

**Chi Harold LIU** receives the Ph.D. degree from Imperial College, UK in 2010, and the B.Eng. degree from Tsinghua University, China in 2006. He is currently a full professor and the vice dean at the School of Computer Science and Technology, Beijing Institute of Technology, China. He is also the director of IBM Mainframe Excellence Center (Beijing), IBM Big Data Technology Center, and National Laboratory of Data Intelligence for China Light Industry. Before moving to academia, he joined IBM Research, China as a staff researcher and project manager, after working as a postdoctoral researcher at Deutsche Telekom Laboratories, Germany and a visiting scholar at IBM T. J. Watson Research Center, USA. His current research interests include the Internet of Things (IoT), big data analytics, mobile computing and deep learning. He received the Distinguished Young Scholar Award in 2013, IBM First Plateau Invention Achievement Award in 2012, and IBM First Patent Application Award in 2011, and was interviewed by EWeb.com as the Featured Engineer in 2011. He has published more than 80 prestigious conference and journal papers and owned more than 14 EU/USA/UK/China patents. He serves as the area editor for *KSI Transactions on Internet and Information Systems* and the book editor for six books published by Taylor & Francis Group, USA and China Machinery Press. He also served as the general chair of IEEE SECON'13 workshop on IoT Networking and Control, IEEE WCNC'12 workshop on IoT Enabling Technologies, and ACM UbiComp'11 Workshop on Networking and Object Memories for IoT. Moreover, he served as the consultant to Asian Development Bank, Bain & Company, and KPMG, USA, and the peer reviewer for Qatar National Research Foundation and National Science Foundation, China. He is a senior member of IEEE.