ARTICLE

Transductive Transfer Dictionary Learning Algorithm for Remote Sensing Image Classification

Jiaqun Zhu[1], Hongda Chen[2], Yiqing Fan[1] and Tongguang Ni[1,2,*]

[1]Aliyun School of Big Data, Changzhou University, Changzhou, 213164, China

[2]Hua Lookeng Honors College, Changzhou University, Changzhou, 213164, China

*Corresponding Author: Tongguang Ni. Email: ntg@cczu.edu.cn

ABSTRACT

To create a green and healthy living environment, people have put forward higher requirements for the refined management of ecological resources. A variety of technologies, including satellite remote sensing, Internet of Things, artificial intelligence, and big data, can build a smart environmental monitoring system. Remote sensing image classification is an important research content in ecological environmental monitoring. Remote sensing images contain rich spatial information and multi-temporal information, but also bring challenges such as difficulty in obtaining classification labels and low classification accuracy. To solve this problem, this study develops a transductive transfer dictionary learning (TTDL) algorithm. In the TTDL, the source and target domains are transformed from the original sample space to a common subspace. TTDL trains a shared discriminative dictionary in this subspace, establishes associations between domains, and also obtains sparse representations of source and target domain data. To obtain an effective shared discriminative dictionary, triple-induced ordinal locality preserving term, Fisher discriminant term, and graph Laplacian regularization term are introduced into the TTDL. The triplet-induced ordinal locality preserving term on sub-space projection preserves the local structure of data in low-dimensional subspaces. The Fisher discriminant term on dictionary improves differences among different sub-dictionaries through intra-class and inter-class scatters. The graph Laplacian regularization term on sparse representation maintains the manifold structure using a semi-supervised weight graph matrix, which can indirectly improve the discriminative performance of the dictionary. The TTDL is tested on several remote sensing image datasets and has strong discrimination classification performance.

KEYWORDS

Classification; dictionary learning; remote sensing image; transductive transfer learning

1 Introduction

The ecological environment is closely related to human life. A good ecological environment is the foundation of human survival and health. At present, maintaining a healthy ecological environment has been the consensus of people all over the world. Ecological environment monitoring refers to the monitoring activities carried out to accurately, timely, and comprehensively reflect the ecological environment status and its changing trend with the objects of mountains, water, forests, fields, lakes,

and grasses. The rapid development of satellite and aviation technology makes the application of remote sensing technology an important part of ecological environmental monitoring [1]. A large amount of real-time and reliable ecological environmental image information can be obtained by using satellite remote sensing images. Remote sensing image classification is the process of real-time scene classification and recognition based on the content extracted from images. Improving the accuracy of scene image classification can bring many conveniences to ecological environmental monitoring. Remote sensing image classification labels an image as a high-level semantic class [2]. Different from pixel-based semantic segmentation, it only pays attention to which pixels correspond to a certain feature. In scene classification, a small-scale image has its semantic information and can be classified into a certain semantic class through semantic features. Therefore, the overall cognition of a remote sensing image constitutes the scene class. The texture and spatial features of a scene, as well as the relationship between the objects, enable not only to describe the underlying physical features of the image, but also to characterize the semantic information of images. Dictionary learning has achieved success in this field due to its sparse representation and reconstruction capabilities. For example, Soltani-Farani et al. [3] developed a spectral and contextual characteristics-based dictionary learning algorithm for multispectral resolution sample classification. The linear combinations of two features of spectral are used as the common elements in the dictionary. Vu et al. [4] developed a dictionary learning model characterized by shared and class-specific dictionaries. Especially, the Fisher discrimination and low-rank constraints are enforced for the learned dictionaries. Geng et al. [5] used an online dictionary learning algorithm with the special atom's selection strategy. The particle swarm optimization algorithm was adopted in the model update phase.

In scene classification, some different scene classes may contain similar feature representations and spatial texture structures [6]. For example, commercial and residential areas often contain houses, vehicles, highways, and interchanges. Bridges often contain roads, vehicles, and other objects. There will also be large differences in characteristics in the same class of scene, such as parking lots with cars and parking lots without cars. Thus, how to effectively and efficiently classify scenes is a challenging problem. In addition, nature scenes are rich and diverse, and even the same ground object may exhibit different characteristics under different time and space conditions, which results in "the same object with different spectrum" and "different objects with the same spectrum". Such differences make most remote sensing application fields lack suitable labeled sample sets. At present, most remote sensing image processing problems based on artificial intelligence adopt supervised learning algorithms. The supervised learning algorithms require sufficiently fine-grained data labeling of the target image, and use these manually labeled remote sensing samples to train effective automatic classifiers on data with the same feature distribution. Therefore, it is a challenge to utilize traditional supervised learning algorithms to build a universal and reusable remote sensing image processing model [7].

Transfer learning is a type of algorithm of using the relevant dataset (called source domain) to calibrate the target domain when the source domain with sufficient class labels and the target domain with insufficient labels (or without class labels) have different feature distributions [8]. In practical applications, a large number of pre-labeled images or open-source public datasets are readily available, which can be used as the source domain. Due to the difference in objective factors such as sensors, algorithms, atmospheric conditions, solar radiation, ground features, building styles, and vegetation characteristics of data acquisition, there are often differences between different domains. To solve this problem, Zhou et al. [9] proposed the correlation alignment (CORAL) algorithm to achieve domain adaptation and capture the texture of the image structure. Tuia et al. [10] developed the spectral feature alignment (SFA) algorithm to use singular value decomposition for calculating the mapping function of feature representation so that the two domains can learn common features independent of their

respective domains in a common latent space. Wang et al. [11] developed the structural correspondence learning algorithm to model the correlation of features.

In this paper, we propose a transductive transfer dictionary learning (TTDL) algorithm for remote sensing image classification. By TTDL, the source and target domains are transformed into a common subspace, and the samples in the two domains are re-encoded to have similar feature distributions. In this subspace, a shared dictionary is trained to establish the association between two domains and obtain the sparse representations of two domain samples. Also, the triplet-induced ordinal locality preserving term on subspace projection, Fisher discriminant term on dictionary and graph Laplacian regularization term on sparse representation are introduced into the TTDL algorithm. The triplet-induced ordinal locality preserving term considers the ranking information of each sample's neighborhood, so it can more accurately describe the local geometry of the data. Fisher's criterion directly constrains the intra-class distance and inter-class distance of the learned dictionary, instead of constraining the sparse representation. Its direct benefit is that the dictionary atoms of the same class can be more compact, so the similarity between different classes can be greatly reduced and the reconstruction ability and discriminative ability of the dictionary can be enhanced. Following the principle that the sparse representation of the same class should be as similar as possible and the sparse representation of different classes should be as different as possible, the graph Laplacian regularization term on sparse representation is constructed. The semi-supervised weight graph matrix in the source domain is built using the known class labels. Since the weight graph matrix in the target domain is unknown, it appears as a variable in the model optimization and reaches its optimal solution when the algorithm converges. Finally, TTDL obtains the discriminative dictionary and sparse representation, to better complete the remote sensing image classification across datasets.

The contributions of this paper are as follows:

(1) A remote sensing image classification algorithm for cross dataset transfer learning is proposed. Different domains are projected into the subspace to eliminate distribution differences. And the shared dictionary is established the relationship between two domains.

(2) Fisher's criterion on dictionary improves the intra-class compactness and inter-class differences of sub-dictionary. In this way, one sub-dictionary reconstructs a certain class of training samples, which can enhance the discrimination of the dictionary.

(3) The unsupervised triplet graph in the triplet-induced ordinal locality preserving term is used to exploit the data local structure information. The semi-supervised weight graph matrix is used to maintain the manifold structure. Thereby, the discriminative ability of subspace and sparse representation can be enhanced.

2 Related Work

2.1 Dictionary Learning

Let $\mathbf{Y} \in \mathbf{R}^{n \times N}$ represent the training data set containing N samples of C classes, where the dimension of samples is n. The dictionary learning [12] is constructed as,

$$\min_{\mathbf{D},\mathbf{A}} \|\mathbf{Y} - \mathbf{DA}\|_2^2 + \lambda \|\mathbf{A}\|_p, \tag{1}$$

where $\mathbf{D} \in \mathbf{R}^{n \times K}$ and $\mathbf{A} \in \mathbf{R}^{K \times N}$ are the dictionary and sparse representation on $\mathbf{Y}$, respectively. $\|\mathbf{Y} - \mathbf{DA}\|_2^2$ represents the reconstruction error term. $\|\mathbf{A}\|_p$ represents the regularization constraint of ℓ_p-norm of sparse representation $\mathbf{A}$. $\lambda > 0$ is the trade-off parameter. Eq. (1) is not a convex optimal solution. $\mathbf{D}$ and $\mathbf{A}$ can be solved by an alternate iterative strategy.

In classification tasks, the new test sample $\mathbf{y}_{new}$ can be classified by,

$$identity\left(\mathbf{y}_{new}\right) = \arg\min_l \left\| \mathbf{y}_{new} - \mathbf{D}\sigma_l\left(\mathbf{a}_{new}\right) \right\|_2, \tag{2}$$

The function $\sigma_l\left(\mathbf{a}_{new}\right)$ returns a vector, whose internal non-zero elements are related to the class of data.

2.2 Fisher Discrimination Dictionary Learning

Fisher discrimination dictionary learning (FDDL) [13] algorithm aims to obtain a structured dictionary of training data. Using Fisher's criterion, the learned sparse representations belonging to different classes have a large spatial distance, while the sparse representations belonging to the same class have a small spatial distance. The objective function of FDDL is,

$$\min_{\mathbf{D},\mathbf{A}} \frac{1}{2} \sum_{l=1}^{C} r\left(\mathbf{Y}_l, \mathbf{D}, \mathbf{A}_l\right) + \lambda_1 \|\mathbf{S}\|_1 + \frac{\lambda_2}{2} g(\mathbf{A}), \tag{3}$$

where $r\left(\mathbf{Y}_l, \mathbf{D}, \mathbf{A}_l\right)$ is the structure dictionary term. $\mathbf{Y}_l$ and $\mathbf{A}_l$ are the class label and sparse representation of the l-th class, respectively. $g\left(\mathbf{A}\right)$ is the Fisher's criterion term on sparse representation. $r\left(\mathbf{Y}_l, \mathbf{D}, \mathbf{A}_l\right)$ and $g\left(\mathbf{A}\right)$ are defined as,

$$r\left(\mathbf{Y}_l, \mathbf{D}, \mathbf{A}_l\right) = \left\| \mathbf{Y}_l - \mathbf{D}\mathbf{A}_l \right\|_F^2 + \left\| \mathbf{Y}_l - \mathbf{D}\mathbf{A}_l \right\|_F^2 + \sum_{j\neq l} \left\| \mathbf{D}_j \mathbf{A}_l^j \right\|_F^2, \tag{4}$$

$$g\left(\mathbf{A}\right) = \sum_{l=1}^{C} \left(\sum_{j=1}^{N_j} \left\| \mathbf{a}_l^j - \mathbf{m}_l \right\|_2^2 - N_l \left\| \mathbf{m}_l - \mathbf{m} \right\|_F^2 \right) + \|\mathbf{A}\|_F^2, \tag{5}$$

where $\mathbf{D}_j$ is the j-th class sub-dictionary. $\mathbf{A}_l^j$ is the sparse representation of $\mathbf{A}_l$ over $\mathbf{D}_j$. m_l and m are the mean vector of $\mathbf{A}_l$ and $\mathbf{A}$, respectively.

3 Transductive Transfer Dictionary Learning Algorithm

3.1 Objective Function

Let $\mathbf{Y}^s = \left[\mathbf{y}_1^s, \mathbf{y}_2^s, \ldots, \mathbf{y}_{N_s}^s\right] \in \mathbf{R}^{n \times N_s}$ and $\mathbf{Y}^t = \left[\mathbf{y}_1^t, \mathbf{y}_2^t, \ldots, \mathbf{y}_{N_t}^t\right] \in \mathbf{R}^{n \times N_t}$ are the source and target domains, respectively, where N_s and N_t are the size of samples in the source and target domains, respectively, $N = N_s + N_t$. TTDL uses the orthographic projection matrices $\mathbf{P}_s \in \mathbf{R}^{m \times n}$ and $\mathbf{P}_t \in \mathbf{R}^{m \times n}$ ($m < n$) to project the source and target domains into the low-dimensional common subspace, respectively. The samples of two domains are re-encoded with the same or similar feature distributions in the subspace. Meanwhile, TTDL trains a discriminative dictionary $\mathbf{D}$ of domain-invariant features under the framework of dictionary learning, and also obtains the sparse representation $\mathbf{A}_s$ and $\mathbf{A}_t$ for the source and target domains, respectively. Let $\tilde{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y}^s & 0 \\ 0 & \mathbf{Y}^t \end{bmatrix}$, $\tilde{\mathbf{P}} = [\mathbf{P}_s, \mathbf{P}_t]$, $\tilde{\mathbf{A}} = [\mathbf{A}_s, \mathbf{A}_t]$, the objective function of TTDL is represented by,

$$\min_{\tilde{\mathbf{P}}, \tilde{\mathbf{A}}, \mathbf{D}, \mathbf{Q}_t} \left\| \tilde{\mathbf{P}}\tilde{\mathbf{Y}} - \mathbf{D}\tilde{\mathbf{A}} \right\|_F^2 + \theta_1 \left\| \tilde{\mathbf{A}} \right\|_F^2 + \theta_2 \Phi(\tilde{\mathbf{P}}) + \theta_3 Z\left(\mathbf{D}\right) + \theta_4 \Gamma(\tilde{\mathbf{A}}, \mathbf{Q}_t), \tag{6}$$

where $\mathbf{Q}_t$ is the weight graph matrix in the target domain. Through $\tilde{\mathbf{P}}$, different domain data is projected into the subspace to eliminate the distribution difference between two domains. Also the shared dictionary $\mathbf{D}$ is established the association between two domains. $\Phi(\tilde{\mathbf{P}})$, $Z\left(\mathbf{D}\right)$ and $\Gamma(\tilde{\mathbf{A}}, \mathbf{Q}_t)$ are functions on variables $\tilde{\mathbf{P}}$, $\mathbf{D}$, $\tilde{\mathbf{A}}$ and $\mathbf{Q}_t$, respectively, which helps to establish the relationship between

two domains and improve the discriminative ability of dictionary learning. $\theta_1, \theta_2, \theta_3$ and θ_4 are trade-off parameters, to coordinate the functions of each item in the objective function.

$\Phi(\tilde{\mathbf{P}})$ is the triplet-induced ordinal locality preserving term on projection matrix. We think that the samples in the low dimensional subspace should keep their local structure information in the original space. $\Phi(\tilde{\mathbf{P}})$ considers the ranking information of samples' neighborhood, so it can more accurately describe the local geometric structure of data. First, we construct the k-nearest neighbor set ϖ_i for each sample $\tilde{\mathbf{y}}_i$. The sample $\tilde{\mathbf{y}}_i$ with two neighbors $\tilde{\mathbf{y}}_u$ and $\tilde{\mathbf{y}}_l$ builds a triplet $(\tilde{\mathbf{y}}_i, \tilde{\mathbf{y}}_u, \tilde{\mathbf{y}}_l)$. We build the asymmetric distance matrix $\mathbf{C}^i$ for $\tilde{\mathbf{y}}_i$, and its element C_{ul}^i is computed as $C_{ul}^i = \left\| \tilde{\mathbf{y}}_i - \tilde{\mathbf{y}}_u \right\|_2^2 - \left\| \tilde{\mathbf{y}}_i - \tilde{\mathbf{y}}_l \right\|_2^2$. Then we build the similarity matrix $\mathbf{G} \in \mathbf{R}^{N \times N}$ as,

$$g_{il} = \begin{cases} \sum_{u \in \Omega_i} C_{ul}^i, & l \in \varpi_i \\ 0, & l \notin \varpi_i \end{cases} \tag{7}$$

The triplet-induced ordinal locality preserving term $\Phi(\tilde{\mathbf{P}})$ is represented as,

$$\min_{\tilde{\mathbf{P}}} \sum_{i=1}^{N} \sum_{l=1}^{N} g_{il} \left\| \tilde{\mathbf{P}}^T \tilde{\mathbf{y}}_i - \tilde{\mathbf{P}}^T \tilde{\mathbf{y}}_l \right\|^2, \tag{8}$$

$s.t.\ \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T = \mathbf{I},$

Denote $\Delta \in \mathbf{R}^{N \times N}$ is the Laplacian matrix $\Delta = \mathbf{M} - \dfrac{\mathbf{G} + \mathbf{G}^T}{2}$. $\mathbf{M}$ is the diagonal matrix, where

$$M_{ii} = \sum_{l} \frac{g_{il} + g_{li}}{2}. \tag{9}$$

Eq. (8) can be represented as,

$$\min_{\tilde{\mathbf{P}}} Tr\left(\tilde{\mathbf{P}} \tilde{\mathbf{Y}} \Delta \tilde{\mathbf{Y}}^T \tilde{\mathbf{P}}^T \right), \tag{10}$$

$s.t.\ \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T = \mathbf{I}.$

$Z(\mathbf{D})$ is the Fisher discriminant term on dictionary. To ensure the dictionary atoms have intra-class compactness and inter-class differences, we implement Fisher discriminant criteria on the dictionary $\mathbf{D}$. $Z(\mathbf{D})$ is represented as,

$$Z(\mathbf{D}) = Tr\left(\Omega_w(\mathbf{D}) - \Omega_B(\mathbf{D})\right). \tag{11}$$

where $\Omega_w(\mathbf{D})$ and $\Omega_B(\mathbf{D})$ are the intra-class scatter and the inter-class scatter on $\mathbf{D}$, respectively,

$$\Omega_w(\mathbf{D}) = \sum_{l=1}^{C} \left(\mathbf{D}_l - \tilde{\mathbf{D}}_l \right) \left(\mathbf{D}_l - \tilde{\mathbf{D}}_l \right)^T, \tag{12}$$

$$\Omega_B(\mathbf{D}) = \sum_{l=1}^{C} K_l \left(\mathbf{D}_l - \tilde{\mathbf{D}} \right) \left(\mathbf{D}_l - \tilde{\mathbf{D}} \right)^T, \tag{13}$$

where K_l is the number of the lth class atoms. $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{D}}_l$ are the mean values of $\mathbf{D}$ and $\mathbf{D}_l$, respectively.

Let $\mathbf{E}_l = \mathbf{1}_{K_l} \mathbf{1}_{K_l}^T \in \mathbf{R}^{K_l \times K_l}$, $\mathbf{E} = \mathbf{1}_K \mathbf{1}_K^T \in \mathbf{R}^{K \times K}$, we have $\tilde{\mathbf{D}}_l = \mathbf{D}_l \mathbf{E}_l / K_l, \tilde{\mathbf{D}} = \mathbf{D}\mathbf{E}/K$, $\Omega_w(\mathbf{D})$ and $\Omega_B(\mathbf{D})$ can be written as,

$$\Omega_w(\mathbf{D}) = \sum_{l=1}^{C} \left(\mathbf{D}_l \mathbf{D}_l^T - \frac{1}{K_l} \mathbf{D}_l \mathbf{E}_l \mathbf{D}_l^T \right), \tag{14}$$

$$\Omega_B(\mathbf{D}) = \sum_{l=1}^{C} K_l(\tilde{\mathbf{D}}_l\tilde{\mathbf{D}}_l^T - 2\tilde{\mathbf{D}}_l\tilde{\mathbf{D}}^T + \tilde{\mathbf{D}}\tilde{\mathbf{D}}^T). \tag{15}$$

Substituting Eqs. (14) and (15) into Eq. (11), $Z(\mathbf{D})$ is represented as,

$$Z(\mathbf{D}) = Tr\left(\mathbf{DBD}^T\right), \tag{16}$$

where $\mathbf{B} = \mathbf{I}_K - Diag\left(\dfrac{2}{K_1}\mathbf{E}_1, \dfrac{2}{K_2}\mathbf{E}_2, \ldots, \dfrac{2}{K_C}\mathbf{E}_C\right) + \dfrac{1}{K}\mathbf{E}$.

$\Gamma\left(\tilde{\mathbf{S}}, \mathbf{Q}_t\right)$ is the graph Laplacian regularization term on sparse representation. We think that the sparse representation of the same class should be as similar as possible, and the sparse representation of different classes should be as different as possible. The weight graph matrix $\mathbf{Q}_s$ is built on the source domain $\mathbf{Y}^s$. The element q_{ij}^s in $\mathbf{Q}_s$ is defined as,

$$q_{ij}^s = \begin{cases} 1, & \mathbf{y}_i^s \text{ and } \mathbf{y}_i^s \text{ are of the same class} \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

Because the class label of $\mathbf{Y}^t$ is unknown, TTDL estimates the weight graph matrix $\mathbf{Q}^t$ in the target domain. $\Gamma(\tilde{\mathbf{A}}, \mathbf{Q}_t)$ is represented as,

$$\min_{\tilde{\mathbf{A}}, \mathbf{Q}^t} \sum_{i=1}^{N_S} \sum_{j=1}^{N_S} q_{ij}^s \left\|\tilde{\mathbf{a}}_i^s - \tilde{\mathbf{a}}_j^s\right\|^2 + \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} q_{ij}^t \left\|\tilde{\mathbf{a}}_i^t - \tilde{\mathbf{a}}_j^t\right\|^2 + \delta \left\|\mathbf{Q}_t\right\|_F^2, \tag{18}$$

$$s.t. \ \mathbf{q}_i^t \mathbf{1} = 1, \ q_{ij}^t \geq 0$$

where δ is the trade-off parameter. Let $\mathbf{L} \in \mathbf{R}^{N \times N}$ be the Laplacian matrix with $\mathbf{L} = \mathbf{W} - \tilde{\mathbf{Q}}$, $\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_t \end{bmatrix}$ and $\mathbf{W}$ be the diagonal matrix with $W_{ii} = \sum_j q_{ij}$.

$$\min_{\tilde{\mathbf{A}}, \mathbf{Q}^t} Tr\left(\tilde{\mathbf{A}}\mathbf{L}\tilde{\mathbf{A}}^T\right) + \delta \left\|\mathbf{Q}_t\right\|_F^2, \tag{19}$$

$$s.t. \mathbf{q}_{ij}^t \mathbf{1} = 1, q_{ij}^t \geq 0$$

Combining Eqs. (8), (11), and (19) into Eq. (6), the objective function of TTDL is re-written as,

$$\min_{\tilde{\mathbf{P}}, \tilde{\mathbf{A}}, \mathbf{D}, \mathbf{Q}_t} \left\|\tilde{\mathbf{P}}\tilde{\mathbf{Y}} - \mathbf{D}\tilde{\mathbf{A}}\right\|_F^2 + \theta_1 \left\|\tilde{\mathbf{A}}\right\|_F^2 + \theta_2 Tr\left(\tilde{\mathbf{P}}\tilde{\mathbf{Y}}\Delta\tilde{\mathbf{Y}}^T\tilde{\mathbf{P}}^T\right) + \theta_3 Tr\left(\mathbf{DBD}^T\right) + \theta_4 Tr\left(\tilde{\mathbf{A}}\mathbf{L}\tilde{\mathbf{A}}^T\right) + \delta \left\|\mathbf{Q}_t\right\|_F^2,$$

$$s.t. \ \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T = \mathbf{I},$$

$$\mathbf{q}_{ij}^t \mathbf{1} = 1, \ q_{ij}^t \geq 0,$$

$$\|d_k\|_2^2 \leq 1, \ \forall k \tag{20}$$

3.2 Optimization

For the objective function of the TTDL algorithm, four variables $\tilde{\mathbf{P}}$, $\tilde{\mathbf{A}}$, $\mathbf{D}$ and $\mathbf{Q}_t$ should be optimized. We adopt the alternately optimization strategy and have the following four steps:

(1) $\tilde{\mathbf{P}}$ is optimized by fixing $\tilde{\mathbf{A}}$, $\mathbf{D}$ and $\mathbf{Q}_t$,

$$\min_{\tilde{\mathbf{P}}} \left\|\tilde{\mathbf{P}}\tilde{\mathbf{Y}} - \mathbf{D}\tilde{\mathbf{A}}\right\|_F^2 + \theta_2 Tr\left(\tilde{\mathbf{P}}\tilde{\mathbf{Y}}\Delta\tilde{\mathbf{Y}}^T\tilde{\mathbf{P}}^T\right),$$

$$s.t. \ \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T = \mathbf{I}, \tag{21}$$

According to [14], there exists the matrices $\tilde{\mathbf{R}} \in \mathbf{R}^{N \times m}$, $\mathbf{Z} \in \mathbf{R}^{N \times K}$, which has the following form: $\tilde{\mathbf{P}} = (\tilde{\mathbf{Y}}\tilde{\mathbf{R}})^{T}$, $\mathbf{D} = \tilde{\mathbf{P}}\tilde{\mathbf{Y}}\mathbf{Z}$, and $\tilde{\mathbf{K}} = \tilde{\mathbf{Y}}^{T}\tilde{\mathbf{Y}}$. Then we have,

$$\min_{\tilde{\mathbf{R}}} \left\| \tilde{\mathbf{R}}^{T}\tilde{\mathbf{K}}\left(\mathbf{I} - \mathbf{Z}\tilde{\mathbf{A}}\right) \right\|_{F}^{2} + \theta_{2} Tr\left(\tilde{\mathbf{R}}^{T}\tilde{\mathbf{K}}\Delta\tilde{\mathbf{K}}^{T}\tilde{\mathbf{R}}\right),$$

$s.t.\ \tilde{\mathbf{R}}^{T}\tilde{\mathbf{K}}\tilde{\mathbf{R}} = \mathbf{I}.$ (22)

We have the closed-form solution of $\tilde{\mathbf{R}}$ as,

$$\tilde{\mathbf{R}} = \tau \Sigma^{-1/2}\sigma,$$ (23)

where $\tilde{\mathbf{K}} = \tau \Sigma \tau^{T}$. σ is obtained by,

$$\min_{\sigma} Tr\left(\sigma^{T}\mathbf{H}\sigma\right),$$

$s.t.\ \sigma^{T}\sigma = \mathbf{I},$ (24)

where $\mathbf{H} = \Sigma^{1/2}\tau^{T}\left(\left(\mathbf{I} - \mathbf{Z}\tilde{\mathbf{A}}\right)\left(\mathbf{I} - \mathbf{Z}\tilde{\mathbf{A}}\right)^{T} + \theta_{2}\Delta\tau\Sigma^{1/2}\right)$. Obviously, σ has the closed-form solution.

(2) $\mathbf{D}$ can be optimized by fixing $\tilde{\mathbf{P}}$, $\tilde{\mathbf{A}}$ and $\mathbf{Q}_{t}$,

$$\min_{\mathbf{D}} \left\| \tilde{\mathbf{P}}\tilde{\mathbf{Y}} - \mathbf{D}\tilde{\mathbf{A}} \right\|_{F}^{2} + \theta_{3} Tr\left(\mathbf{DBD}^{T}\right),$$

$s.t.\ \|d_{k}\|_{2}^{2} \le 1,\ \forall k.$ (25)

Using the Lagrange dual approach, $\mathbf{D}$ is obtained by,

$$\mathbf{D} = \tilde{\mathbf{P}}\tilde{\mathbf{Y}}\tilde{\mathbf{A}}^{T}\left(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^{T} + \theta_{3}\mathbf{B}\right)^{-1}.$$ (26)

(3) $\tilde{\mathbf{S}}$ is optimized by fixing $\tilde{\mathbf{P}}$, $\mathbf{D}$ and $\mathbf{Q}_{t}$,

$$\min_{\tilde{\mathbf{A}}} \left\| \tilde{\mathbf{P}}\tilde{\mathbf{Y}} - \mathbf{D}\tilde{\mathbf{A}} \right\|_{F}^{2} + \theta_{1} \left\| \tilde{\mathbf{A}} \right\|_{F}^{2} + \theta_{4} Tr\left(\tilde{\mathbf{A}}\mathbf{L}\tilde{\mathbf{A}}^{T}\right).$$ (27)

We re-write Eq. (27) in terms of the row vector $\tilde{\mathbf{a}}_{i}$ in $\tilde{\mathbf{A}}$ as,

$$\min_{\tilde{\mathbf{a}}_{i}} \left\| \left(\tilde{\mathbf{P}}\tilde{\mathbf{Y}}\right)_{i} - \mathbf{D}\tilde{\mathbf{a}}_{i} \right\|_{2}^{2} + \theta_{1} \left\| \tilde{\mathbf{a}}_{i} \right\|_{2}^{2} + \theta_{4} \sum_{j} q_{i,j} \left\| \tilde{\mathbf{a}}_{i} - \tilde{\mathbf{a}}_{j} \right\|_{2}^{2}.$$ (28)

The closed-form solution of $\tilde{\mathbf{a}}_{i}$ is,

$$\tilde{\mathbf{a}}_{i} = \left(\mathbf{D}^{T}\mathbf{D} + \theta_{4}\sum_{j} q_{i,j}\mathbf{I} + \theta_{1}\mathbf{I}\right)^{-1}\left(\mathbf{D}^{T}\left(\tilde{\mathbf{P}}\tilde{\mathbf{Y}}\right)_{i} + \theta_{4}\sum_{j} q_{i,j}\tilde{\mathbf{a}}_{j}\right).$$ (29)

(4) $\mathbf{Q}_{t}$ can be optimized by fixing $\tilde{\mathbf{P}}$, $\mathbf{D}$ and $\tilde{\mathbf{A}}$, we have,

$$\min_{\mathbf{Q}_{t}} \theta_{4} \sum_{i=1}^{N_{t}} \sum_{j=1}^{N_{t}} q_{ij}^{t} \left\| \tilde{\mathbf{a}}_{i}^{t} - \tilde{\mathbf{a}}_{j}^{t} \right\|^{2} + \delta \|\mathbf{Q}_{t}\|_{F}^{2},$$

$s.t.\ \mathbf{q}_{ij}^{t}\mathbf{1} = 1,\ q_{ij}^{t} \ge 0,$ (30)

We define $h_{i,j} = \dfrac{\theta_{4} \left\| \tilde{\mathbf{a}}_{t,i} - \tilde{\mathbf{a}}_{t,j} \right\|_{2}^{2}}{2\delta}$, Eq. (30) is represented as,

$$\min_{q_{ij}^{t}} \left\| \mathbf{q}_{i}^{t} + \mathbf{h}_{i} \right\|_{2}^{2},$$

$s.t.\ \mathbf{q}_{i}^{t}\mathbf{1} = 1,\ q_{ij}^{t} \ge 0,\ \forall i,j$ (31)

The closed-form solution of $\mathbf{q}_i^t$ is,

$$\mathbf{q}_i^t = \left(\frac{1 + \sum_{j=1}^{k} \hat{h}_{ij}}{k} \mathbf{1} - \mathbf{h}_i \right)_+ , \tag{32}$$

where $(\cdot)_+$ means the elements in $\mathbf{q}_i^t$ are nonnegative. The element $\hat{h}_{ij}$ of $\hat{\mathbf{h}}_i$ is the same as that in $\mathbf{h}_i$ with the ascending order.

It is noted that according to [14], δ can be computed by,

$$\delta = \frac{1}{N^t} \sum_{i=1}^{N^t} \left(\frac{k}{2} \hat{h}_{ik+1} - \frac{1}{2} \sum_{j=1}^{k} \hat{h}_{ij} \right). \tag{33}$$

The solving procedure of Eq. (20) is summarized in Algorithm 1.

Algorithm 1: TTDL algorithm

Input: The labeled source domain $\mathbf{Y}^s$ and unlabeled target domain $\mathbf{Y}^t$, parameters θ_1, θ_2, θ_3, and θ_4;

Output: $\tilde{\mathbf{P}}$, $\tilde{\mathbf{A}}$, $\mathbf{D}$, and $\mathbf{Q}_t$.

Calculate the similarity matrix $\mathbf{G}$ via Eq. (7) and weight graph matrix $\mathbf{Q}^s$ via Eq. (17);
While not converge do
  Update $\tilde{\mathbf{P}}$ via Eqs. (22)–(24);
  Update $\mathbf{D}$ via Eq. (26);
  Update $\tilde{\mathbf{A}}$ via Eq. (30);
  Update $\mathbf{Q}_t$ via Eq. (32);
end while
Compute the parameter δ via Eq. (33)
Obtain the optimal $\tilde{\mathbf{P}}$, $\tilde{\mathbf{A}}$, $\mathbf{D}$, and $\mathbf{Q}_t$.

---

4 Experiments

4.1 Datasets

The TTDL algorithm is evaluated on real-world remote sensing datasets: RSSCN7 [1], Ucmerced land [2], Aerial image dataset (AID) [15], and SIRI-WHU dataset [16]. The RSSCN7 dataset consists of 2800 images of seven scene classes. There are 400 images of each class, in which the size of each image is 400×400 pixels. The Ucmerced land dataset consists of 2100 aerial scene images from 21 classes, with the size of 256×256 pixels. The AID dataset is a large scale aerial scene dataset, which consists of 10,000 remote sensing scene images of 30 classes, with the size of 600×600 pixels. The SIRI-WHU dataset is also an aerial scene dataset, composed of 12 classes of aerial scene images, in which each class contains 200 images, with the size of 200×200 pixels. To efficiently represent the remote sensing images, we adopt two different types of deep feature representations: ResNet50 and VGG-VD-16 [7]. They are extracted by the two convolution neural networks (CNN). The dimensions of the two deep features are 2048.

In the experiment, we design three cross-domain remote sensing image scene classification tasks: Ucmerced→RSSCN7 (named U→R), AID→RSSCN7 (named A→R), and SIRI-WHU→RSSCN7 (named S→R), referring to $\mathbf{Y}^s→\mathbf{Y}^t$. To match the classes in RSSCN7, we select corresponding similar classes to form the source domain, the detailed information of three cross-dataset scene classification tasks is shown in Table 1. Following the ID sequence in Table 1, some sample images of

corresponding classes in four remote sensing image datasets are shown in Fig. 1. In the experiments, we randomly select 80% of the images in $\mathbf{Y}^s$ and 5% of the images in $\mathbf{Y}^t$ for model training, and the rest of images in $\mathbf{Y}^t$ are used for testing. Comparison algorithms include: sparse representation-based nearest neighbor (SRNN) [17], maximum mean discrepancy (MMD) [18], generalized joint distribution adaptation (G-JDA) [19], and maximum independence domain adaptation (MIDA) [20], and transfer independently together (TIT) [21]. SRNN is a basic sparse representation algorithm, while MMD, G-JDA, MIDA, and TIT are transfer learning algorithms. The regularization and kernel parameters are set in $\left\{2^{-6}, 2^{-5}, \ldots, 2^6\right\}$. The TTDL algorithm involves four parameters, which are set in $\left\{10^{-3}, 10^{-2}, \ldots, 10^3\right\}$. The size of the sub-dictionary is the same as the number of training images in each class. The experiment is repeated ten times, with each run's classification accuracy being recorded.

Table 1: The information of selected subclass in four sensing image datasets

| ID	RSSCN7		Ucmerced land		SIRI-WHU		AID	
	Subclass	No.	Subclass	No.	Subclass	No.	Subclass	No.
1	Grass	400	Agriculture	100	Agriculture	200	Farmland	370
2	Field	400	Baseball diamond	100	Meadow	200	Baseball diamond	220
3	Industry	400	Building	100	Industrial	200	Industrial	390
4	River/Lake	400	River	100	River	400	River	410
5	Forest	400	Forest	100	Park	200	Forest	250
6	Resident	400	Medium-residential	300	Residential	200	Medium-residential	290

(a)
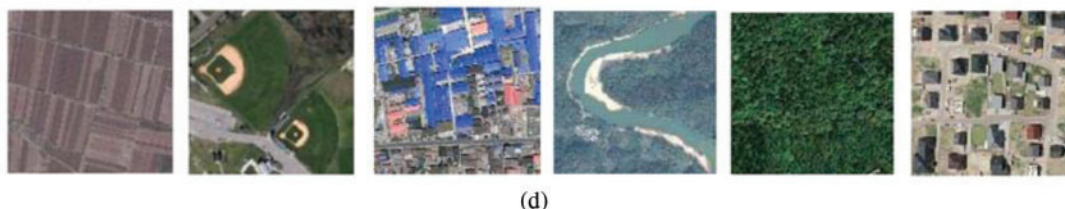
(b)

(c)
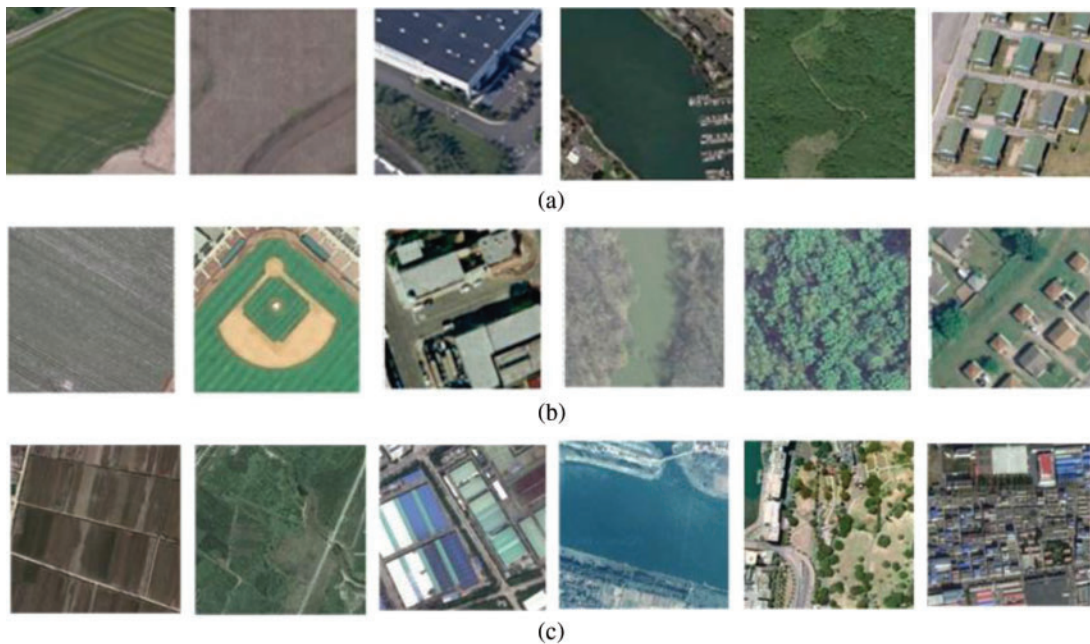
Figure 1: (Continued)

Figure 1: Sample images in four dataset, (a) RSSCN7, (b) Ucmerced land, (c) SIRI-WHU, (d) AID

4.2 Performance Comparison

We compare the TTDL algorithm on three cross-domain remote sensing image classification tasks. The accuracy results in each class are shown in Tables 2–4. Analyzing the experimental results, we can see that the TTDL algorithm achieves the best classification accuracy. The baseline algorithm SRNN cannot obtain satisfactory performance in all cross-domain remote sensing image classification tasks. Since SRNN is only trained on the source domain, it cannot be used directly on the RSSCN7 dataset due to differences in data distribution across different domains. G-JDA aligns cross-domain edge feature probability distributions and conditional feature probability distributions in the new feature space by applying their respective projection matrices. TIT also uses landmarks to select representative samples for cross-domain feature matching while maintaining the manifold structure using graphs. Both MMD and MIDA project different domains into the common subspace to achieve cross-domain feature alignment. The above algorithms except TTDL do not consider triplet-induced ordinal locality preserving term, Fisher discriminant term, and graph Laplacian regularization term in transfer learning. The TTDL algorithm not only maintains the local structure in the subspace, but also makes similar sparse representations as similar as possible, to enhance the discriminative performance of sparse representations. Additionally, TTDL minimizes the intra-class scatter of atoms and maximizes the inter-class scatter of atoms, which can greatly promote the discriminative performance of the dictionary.

Table 2: Accuracy (%) with its standard deviation of each class in the U→R task (The bolds indicate the best results)

Features	Classes	SRNN	MMD	G-JDA	MIDA	TIT	TTDL
ResNet50	Grass	64.76 (1.76)	83.72 (1.78)	82.90 (1.03)	88.10 (1.44)	88.03 (1.35)	**89.14** (1.00)
	Field	58.07 (1.13)	74.76 (1.69)	74.29 (0.81)	78.17 (0.96)	77.52 (1.78)	**81.32** (0.74)
	Industry	50.76 (1.30)	62.68 (2.10)	68.77 (1.39)	70.01 (1.55)	68.07 (0.96)	**71.76** (1.03)
	River/Lake	52.34 (1.15)	78.76 (1.05)	83.03 (1.65)	85.37 (2.16)	84.50 (1.03)	**88.36** (0.82)
	Forest	88.54 (1.40)	89.99 (2.09)	90.99 (2.04)	91.18 (1.69)	90.39 (1.63)	**91.49** (1.42)
	Resident	46.58 (1.30)	57.03 (1.56)	69.18 (1.93)	71.14 (2.26)	68.76 (1.69)	**74.68** (1.09)

(Continued)

Table 2 (continued)

Features	Classes	SRNN	MMD	G-JDA	MIDA	TIT	TTDL
VGG-VD-16	Grass	62.69 (1.11)	82.44 (1.87)	83.14 (1.42)	84.41 (1.24)	83.84 (0.70)	**87.46** (0.73)
	Field	56.17 (0.65)	75.5 (1.04)	73.95 (0.90)	77.48 (0.68)	76.42 (0.80)	**78.43** (0.84)
	Industry	48.66 (1.55)	61.33 (0.96)	67.85 (1.80)	67.87 (0.77)	66.08 (1.19)	**69.75** (1.08)
	River/Lake	50.29 (0.78)	76.40 (1.71)	81.71 (1.86)	83.67 (1.86)	82.24 (1.08)	**86.50** (0.93)
	Forest	84.10 (1.32)	87.97 (2.00)	87.45 (1.16)	88.83 (0.85)	88.01 (1.14)	**90.01** (0.77)
	Resident	44.80 (1.35)	54.41 (1.13)	67.38 (1.00)	68.67 (1.22)	66.86 (1.19)	**72.98** (0.69)

Table 3: Accuracy (%) with its standard deviation of each class in the A→R task (The bolds indicate the best results)

Features	Classes	SRNN	MMD	G-JDA	MIDA	TIT	TTDL
ResNet50	Grass	67.33 (1.44)	84.22 (1.85)	85.09 (1.60)	89.22 (1.89)	89.31 (1.92)	**90.13** (1.66)
	Field	60.59 (1.21)	77.99 (1.08)	79.65 (0.93)	80.41 (2.35)	82.60 (1.81)	**83.13** (1.05)
	Industry	54.80 (2.09)	66.02 (1.84)	70.77 (1.60)	74.94 (1.90)	74.75 (1.52)	**75.70** (1.04)
	River/Lake	55.07 (1.27)	80.47 (1.37)	86.24 (1.92)	87.06 (0.85)	85.88 (2.08)	**90.70** (1.07)
	Forest	88.40 (1.55)	90.14 (1.45)	91.31 (1.45)	91.85 (1.43)	92.99 (0.94)	**93.65** (0.93)
	Resident	50.75 (1.28)	70.36 (0.93)	73.81 (1.97)	72.65 (1.56)	73.02 (1.49)	**76.52** (0.83)
VGG-VD-16	Grass	65.49 (1.40)	82.55 (1.91)	83.62 (1.22)	87.19 (1.31)	86.23 (1.76)	**89.25** (0.97)
	Field	58.52 (0.82)	75.41 (1.58)	76.10 (2.18)	78.87 (1.52)	80.50 (2.49)	**82.74** (1.09)
	Industry	50.73 (0.86)	63.31 (2.08)	67.68 (2.10)	70.17 (1.15)	71.64 (0.98)	**72.83** (1.54)
	River/Lake	50.49 (1.85)	75.55 (2.97)	83.26 (1.79)	84.85 (1.77)	83.67 (0.56)	**86.84** (1.13)
	Forest	87.27 (1.15)	89.72 (1.70)	90.07 (1.45)	90.63 (1.36)	90.39 (2.03)	**91.27** (1.10)

(Continued)

Table 3 (continued)

Features	Classes	SRNN	MMD	G-JDA	MIDA	TIT	TTDL
	Resident	48.35 (1.78)	70.08 (1.70)	71.80 (0.84)	73.90 (2.58)	70.76 (1.86)	**74.65** (0.83)

Table 4: Accuracy (%) with its standard deviation of each class in the S→R task (The bolds indicate the best results)

Features	Classes	SRNN	MMD	G-JDA	MIDA	TIT	TTDL
ResNet50	Grass	60.09 (1.59)	80.23 (1.26)	81.87 (1.50)	84.85 (1.45)	84.02 (1.49)	**86.22** (1.61)
	Field	49.81 (1.50)	71.76 (1.08)	73.11 (0.90)	76.18 (1.23)	76.69 (1.34)	**78.34** (1.24)
	Industry	45.89 (1.06)	66.00 (0.85)	66.81 (1.42)	67.32 (0.86)	65.83 (1.83)	**70.68** (1.08)
	River/Lake	50.30 (1.58)	70.98 (0.70)	80.21 (0.95)	81.31 (1.14)	82.05 (1.32)	**85.78** (1.64)
	Forest	80.11 (0.69)	81.65 (1.74)	84.55 (1.89)	85.53 (1.11)	84.97 (1.27)	**87.89** (1.44)
	Resident	42.58 (1.29)	55.26 (1.91)	70.71 (2.15)	68.45 (1.82)	68.08 (1.19)	**72.43** (1.11)
VGG-VD-16	Grass	58.56 (1.42)	78.25 (1.36)	77.23 (1.63)	80.63 (2.16)	81.07 (0.86)	**83.98** (1.02)
	Field	47.34 (0.98)	68.58 (2.16)	70.92 (1.14)	73.61 (1.23)	73.66 (1.56)	**75.29** (0.90)
	Industry	43.17 (1.11)	62.24 (0.78)	64.12 (2.01)	65.05 (2.16)	64.59 (1.08)	**68.82** (1.14)
	River/Lake	50.98 (2.42)	69.13 (2.36)	80.86 (0.63)	80.53 (2.19)	80.34 (0.86)	**83.22** (1.02)
	Forest	77.68 (1.22)	79.35 (1.04)	82.44 (2.22)	83.29 (1.51)	83.94 (1.91)	**85.38** (1.15)
	Resident	40.56 (1.99)	52.07 (1.97)	69.16 (2.41)	67.15 (1.04)	67.40 (1.99)	**71.61** (1.53)

The classification results using ResNet50 and VGG-VD-16 features are comparable. As we know, the contained information in remote sensing scene images is often closely related to its class. The traditional feature information such as color, texture, space, and spectral information is insufficient for remote sensing images. Especially when the features corresponding to some classes are not significant

enough, the accuracy of the classifier will be reduced. Thus, the deep features are effective in our experiments.

Fig. 2 records the classification performance of each algorithm in three transfer learning tasks. We can see that the average accuracy of TTDL is the highest among all sub-classes. For example, in the U→R task, using the ResNet50 features, the average accuracy of TTDL is 22.62% higher than the SRNN algorithm, and 3.25% higher than the second best. In the A→R task, using the VGG-VD-16 features, the average accuracy of TTDL is 22.17% higher than the SRNN algorithm, and 1.88% higher than the MIDA algorithm. In the S→R task, using the ResNet50 features, the average accuracy of TTDL algorithm is 25.53% higher than the SRNN algorithm, and 3.28% higher than the second best. These results show that TTDL is effective in three classification tasks of U→R, A→R, and S→R. Thus, the dictionary learning framework combined with subspace learning, Fisher discriminant, and local information preserving is a good choice for cross-dataset remote sensing image classification.
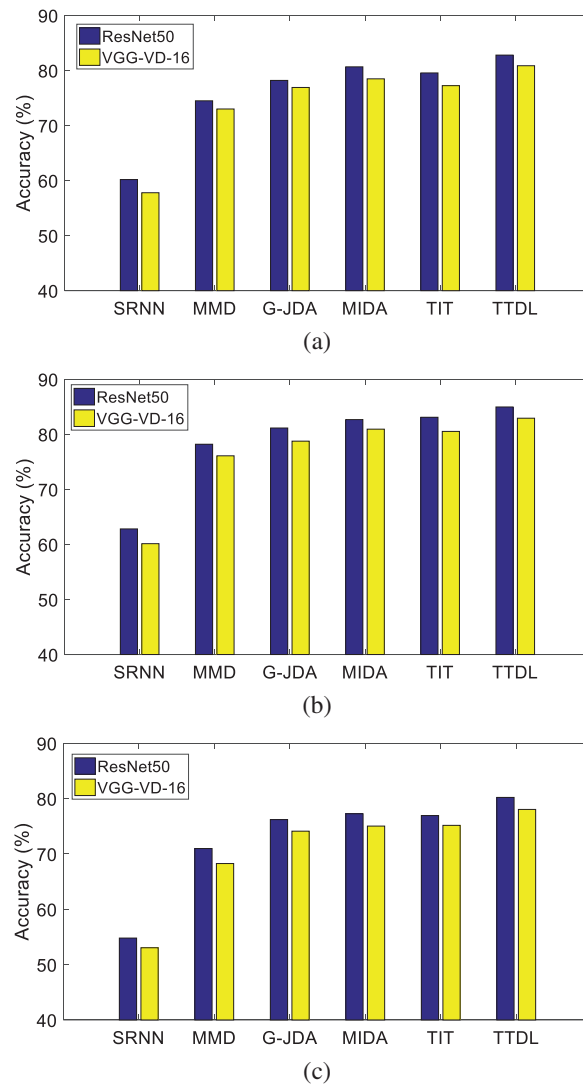

Figure 2: Accuracy comparison of all algorithms in the, (a) U→R task, (b) A→R task, (c) S→R task

4.3 Model Analysis

We show the confusion matrix using the ResNet50 feature in Tables 5–7. The value in confusion matrices means the accuracy (%) of TTDL in each class. The testing data of each scene class in the RSSCN7 dataset consists of 380 images. We can see that in the U→R task, TTDL classifies scene classes of Grass, River/Lake, and Forest over 80%, and the accuracy of Industry is low. The reason is mainly that Industry and Resident have the high similarity in the Ucmerced land and RSSCN7 datasets. The performances of TTDL in the A→R and S→R tasks show similar results. The classification performance in Grass, River/Lake, and Forest, are a higher than that of other classes.

Table 5: Confusion matrix of TTDL on the U→R task

	Grass	Field	Industry	River/Lake	Forest	Resident
Grass	89.14	5.04	0	2.50	3.32	0
Field	13.32	81.32	0	3.71	1.65	0
Industry	0	3.21	71.76	9.47	0	15.56
River/Lake	2.25	2.06	3.33	88.36	1.03	2.97
Forest	0	4.83	0	3.68	91.49	0
Resident	0	0	25.32	0	0	74.68

Table 6: Confusion matrix of TTDL on the A→R task

	Grass	Field	Industry	River/Lake	Forest	Resident
Grass	90.13	3.01	0	2.13	3.82	0.91
Field	10.24	83.13	0	4.79	1.84	0
Industry	0	3.06	75.70	6.29	0	14.95
River/Lake	3.11	2.11	2.29	90.70	0.26	1.53
Forest	0	4.11	0	2.24	93.65	0
Resident	0	0.95	20.79	0.89	0.85	76.52

Table 7: Confusion matrix of TTDL on the S→R task

	Grass	Field	Industry	River/Lake	Forest	Resident
Grass	86.22	6.41	0	3.92	3.45	0
Field	14.21	78.34	0.24	4.51	2.07	0.63
Industry	0	2.49	70.68	8.99	0	17.84
River/Lake	2.58	2.84	4.58	85.78	2.06	2.16
Forest	0	4.84	0.46	6.81	87.89	0
Resident	0	0.47	26.05	0.26	0.79	72.43

4.4 Ablation Experiment

To further analysis three components in TTDL, we show the ablation experiment results using ResNet50 feature in Table 8. TTDL with $\theta_2 = 0$ means $\Phi(\tilde{P})$ is removed. TTDL with $\theta_3 = 0$ means $Z(\mathbf{D})$ is removed. TTDL with $\theta_4 = \delta = 0$ means $\Gamma(\tilde{\mathbf{A}}, \mathbf{Q}_t)$ is removed. The experimental results validate the effectiveness of triplet-induced ordinal locality preserving term, Fisher discriminant term, and graph Laplacian regularization term in TTDL. The triplet-induced ordinal locality preserving term finds a suitable projection subspace to preserve the data structure. The Fisher discriminant term builds a discriminant dictionary to bridge two different domains. The graph Laplacian regularization term learns the discriminant sparse representation. With the joint learning of three terms, the TTDL algorithm has achieved satisfactory results in cross-dataset remote scene image classifications.

Table 8: Component evaluation in TTDL

	$\theta_2 = \theta_3 = \theta_4 = 0$	$\theta_2 = 0$	$\theta_3 = 0$	$\theta_4 = \delta = 0$	TTDL
U→R	53.23	77.28	76.17	78.92	81.55
A→R	67.20	77.36	75.29	76.13	80.70
S→R	60.04	75.26	77.20	76.45	81.87

5 Conclusion

The ecological environment has become one of the root factors affecting human health. Environmental and health management is a comprehensive and complex work across departments, fields, and disciplines. The rapid development of aerospace, satellite remote sensing, and data communication technology makes the application of remote sensing technology in ecological environmental monitoring more extensive. Reducing the workload of manual annotation and achieving high-precision classification of remote sensing images are difficult problems in ecological environmental monitoring. Benefiting from the previous manual labeling work, there are a large number of labeled datasets in the source domain. Because of the large differences between different datasets, it is difficult to achieve the ideal classification by directly training the classifier with these labeled datasets. To solve this problem, this paper proposes a transductive transfer dictionary learning algorithm TTDL. To obtain the representation of different domain samples, TTDL uses a subspace projection strategy to eliminate the distribution difference. In TTDL, the triplet induced ordinal locality preserving term, Fisher discriminant term, and graph Laplacian regularization term are introduced, so that the dictionary has intra-class compactness and inter-class differences. The TTDL algorithm has achieved satisfactory results in remote sensing image classification across datasets. Our work in the next stage includes integrating statistical information from data into subspace projection to eliminate feature distribution differences. For remote sensing images, the class imbalance often exists; it will lead to a high misclassification rate for classes with fewer samples. How to solve this problem is also our future research work. In addition, the transfer learning algorithm proposed in this paper is only applied to a single feature perspective. How to use multiple feature perspective data to accurately describe the data structure has important research significance and application value.

Availability of Data and Materials: Four public datasets RSSCN7, Ucmerced land, SIRI-WHU, and AID are used in this study. The RSSCN7 dataset can be downloaded in https://github.com/palewithout/RSSCN7. The Ucmerced land dataset can be downloaded in http://weegee.vision.ucmerced.edu/datasets/landuse.html. The SIRI-WHU dataset can be downloaded in https://figshare.com/articles/dataset/SIRI_WHU_Dataset/8796980. The AID dataset can be downloaded in https://paperswithcode.com/dataset/aids.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Zou, Q., Ni, L., Zhang, T., Wang, Q. (2015). Deep learning based feature selection for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters, 12(11),* 2321–2325.
2. Yang, Y., Newsam, S. (2010). Bag-of-visual-words and spatial extensions for land-use classification. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 270–279. San Jose.
3. Soltani-Farani, A., Rabiee, H. R., Hosseini, S. A. (2014). Spatial-aware dictionary learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 53(1),* 527–541.
4. Vu, T. H., Monga, V. (2017). Fast low-rank shared dictionary learning for image classification. *IEEE Transactions on Image Processing, 26(11),* 5160–5175.
5. Geng, H., Wang, L., Liu, P. (2014). Dictionary learning for large-scale remote sensing image based on particle swarm optimization. *2014 12th International Conference on Signal Processing (ICSP)*, pp. 784–789. Hangzhou, China.
6. Zheng, Z., Zhong, Y., Su, Y., Ma, A. (2022). Domain adaptation via a task-specific classifier framework for remote sensing cross-scene classification. *IEEE Transactions on Geoscience and Remote Sensing, 60(2),* 1–13.
7. Zhang, J., Liu, J., Pan, B., Shi, Z. (2020). Domain adaptation based on correlation subspace dynamic distribution alignment for remote sensing image scene classification. *IEEE Transactions on Geoscience and Remote Sensing, 58(11),* 7920–7930.
8. Zhu, L., Ma, L. (2016). Class centroid alignment based domain adaptation for classification of remote sensing images. *Pattern Recognition Letters, 83(11),* 124–132.
9. Zhou, Z., Wu, Y., Yang, X., Zhou, Y. (2022). Neural style transfer with adaptive auto-correlation alignment loss. *IEEE Signal Processing Letters, 29(4),* 1027–1031. https://doi.org/10.1109/LSP.2022.3165758
10. Tuia, D., Marcos, D., Camps-Valls, G. (2016). Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization. *ISPRS Journal of Photogrammetry and Remote Sensing, 120(10),* 1–12.
11. Wang, D., Wu, J., Yang, J., Jing, B., Zhang, W. et al. (2021). Cross-lingual knowledge transferring by structural correspondence and space transfer. *IEEE Transactions on Cybernetics, 52(7),* 6555–6566.
12. Jiang, Z., Lin, Z., Davis, L. S. (2013). Label consistent K-SVD: Learning a discriminative dictionary for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(11),* 2651–2664.
13. Yang, M., Zhang, L., Feng, X., Zhang, D. (2014). Sparse representation based fisher discrimination dictionary learning for image classification. *International Journal of Computer Vision, 109(5),* 209–232.
14. Qi, L., Huo, J., Fan, X., Shi, Y., Gao, Y. (2018). Unsupervised joint subspace and dictionary learning for enhanced cross-domain person re-identification. *IEEE Journal of Selected Topics in Signal Processing, 12(6),* 1263–1275.
15. Xia, G. S., Hu, J., Hu, F., Shi, B., Bai, X. et al. (2017). AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing, 55(7),* 3965–3981.

16. Zhao, B., Zhong, Y., Xia, G. S., Zhang, L. (2015). Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing, 54(4),* 2108–2123.

17. Zou, J., Li, W., Du, Q. (2015). Sparse representation-based nearest neighbor classifiers for hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters, 12(12),* 2418–2422.

18. Sejdinovic, D., Sriperumbudur, B., Gretton, A., Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics, 41(5),* 2263–2291.

19. Hsieh, Y. T., Tao, S. Y., Tsai, Y. H. H., Yeh, Y. R., Wang, Y. C. F. (2016). Recognizing heterogeneous cross-domain data via generalized joint distribution adaptation. *2016 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. Seattle.

20. Yan, K., Kou, L., Zhang, D. (2017). Learning domain-invariant subspace using domain features and independence maximization. *IEEE Transactions on Cybernetics, 48(1),* 288–299.

21. Li, J., Lu, K., Huang, Z., Zhu, L., Shen, H. T. (2018). Transfer independently together: A generalized framework for domain adaptation. *IEEE Transactions on Cybernetics, 49(6),* 2144–2155.