Topological Structure and Semantic Information Transfer Network for Cross-Scene Hyperspectral Image Classification

Yuxiang Zhang, *Student Member, IEEE*, Wei Li, *Senior Member, IEEE*, Mengmeng Zhang, Ying Qu, *Member, IEEE*, Ran Tao, *Senior Member, IEEE*, and Hairong Qi, *Fellow, IEEE*

***Abstract*— Domain adaptation techniques have been widely applied to the problem of cross-scene hyperspectral image (HSI) classification. Most existing methods use convolutional neural networks (CNNs) to extract statistical features from data and often neglect the potential topological structure information between different land cover classes. CNN-based approaches generally only model the local spatial relationships of the samples, which largely limits their ability to capture the nonlocal topological relationship that would better represent the underlying data structure of HSI. In order to make up for the above shortcomings, a Topological structure and Semantic information Transfer network (TSTnet) is developed. The method employs the graph structure to characterize topological relationships and the graph convolutional network (GCN) that is good at processing for cross-scene HSI classification. In the proposed TSTnet, graph optimal transmission (GOT) is used to align topological relationships to assist distribution alignment between the source domain and the target domain based on the maximum mean difference (MMD). Furthermore, subgraphs from the source domain and the target domain are dynamically constructed based on CNN features to take advantage of the discriminative capacity of CNN models that, in turn, improve the robustness of classification. In addition, to better characterize the correlation between distribution alignment and topological relationship alignment, a consistency constraint is enforced to integrate the output of CNN and GCN. Experimental results on three cross-scene HSI datasets demonstrate that the proposed TSTnet performs significantly better than some state-of-the-art domain-adaptive approaches. The codes will be available from the website: https://github.com/YuxiangZhang-BIT/IEEE_TNNLS_TSTnet.**

***Index Terms*— Cross-scene, distribution alignment, domain adaption, graph alignment, graph convolutional network (GCN), hyperspectral image (HSI) classification.**

Manuscript received March 4, 2021; revised June 20, 2021 and August 16, 2021; accepted August 31, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61922013, Grant U1833203, and Grant 62001023; in part by the Beijing Natural Science Foundation under Grant JQ20021 and Grant L191004; and in part by the China Postdoctoral Science Foundation under Grant BX20200058. *(Corresponding author: Wei Li.)*

Yuxiang Zhang, Wei Li, Mengmeng Zhang, and Ran Tao are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China, and also with the Beijing Key Laboratory of Fractional Signals and Systems, Beijing Institute of Technology, Beijing 100081, China (e-mail: zyx829625@163.com; liwei089@ieee.org; mengmengzhang@bit.edu.cn; rantao@bit.edu.cn).

Ying Qu and Hairong Qi are with the Advanced Imaging and Collaborative Information Processing Group, Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996 USA (e-mail: yqu3@vols.utk.edu; hqi@utk.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3109872.

Digital Object Identifier 10.1109/TNNLS.2021.3109872

NOMENCLATURE

Abbreviations:

HSI	Hyperspectral image.
SD	Source data.
TD	Target data.
DA	Domain adaptation.
GCN	Graph convolutional network.
GOT	Graph optimal transmission.
MMD	Maximum mean difference.
WD	Wasserstein distance.
GWD	Gromov–Wasserstein distance.

Variables:

$\mathbf{X}_s$ and $\mathbf{X}_t$	Source and target data.
$\mathbf{G}$	Undirected graph.
$\mathbf{V}$	Node set of the graph.
$\mathbf{E}$	Set of edges connected by nodes.
$\mathbf{Z}_s$ and $\mathbf{Z}_t$	SD and TD features extracted by CNN.
$\mathbf{G}_s$ and $\mathbf{G}_t$	SD and TD graph features extracted by GCN.
$P(\mathbf{X}_s)$ and $P(\mathbf{X}_t)$	Marginal distribution of SD and TD.

I. INTRODUCTION

IN RECENT years, the use of various remote sensing data [e.g., HSI, synthetic aperture radar (SAR), and light detection and ranging (LiDAR)] for the task of land cover classification has become a popular subject. Compared with multispectral images, HSI has rich spectral information, which can identify the materials and classes of land cover more accurately. On the other hand, most existing classification approaches, especially deep learning-based methods, need a large training set, which is both time-consuming and labor-intensive to collect. In addition, both traditional and deep learning algorithms, such as representation-based methods [1]–[3], least-squares regression-based methods [4]–[6], and convolutional neural network (CNN)-based methods [7]–[10], are only suitable for fixed scenes, that is, training samples and testing samples are independently and identically distributed. As a result, the methods cannot be transferred to other scenes, and it is impossible to test the data obtained in real time without labels. This is the main problem faced by the cross-scene classification task. The training samples and testing samples are, respectively, labeled SD and unlabeled TD. The

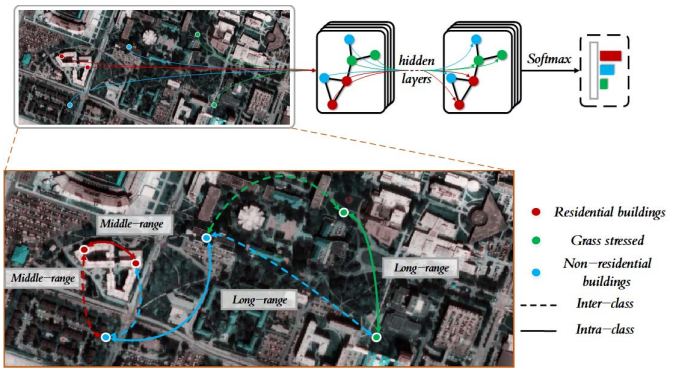

Fig. 1. Schematic of GCN and the middle- and long-range topological relationships.

purpose of this task is to use SD to train the model, transfer the shared knowledge in SD to TD, and classify TD to realize the migration of models to different scenes.

The acquisition process of HSI is inevitably affected by various factors, such as the different sensor nonlinearities, different seasons, and different weather conditions [11], which lead to differences in spectral reflectance between SD and TD of the same land cover class. Therefore, when directly using SD to classify TD, the problem of spectral shift is often encountered. Fortunately, transfer learning can be helpful to solve this problem and has been widely used for cross-scene classification. DA as a case of transductive transfer learning sets the source task and target task to be the same, while SD and TD are different but related [12], which can reduce the spectral shift in feature-level and learn domain-invariant models. Distribution alignment-based methods [13]–[15], feature selection-based methods [16]–[18], and subspace learning-based methods are three types of traditional DA methods [19]–[21].

Many deep learning methods use an adaptive layer to realize the adaptation of SD and TD, thereby making the classification performance of models on TD better. A deep adaptation network (DAN) was proposed in [22], which added three adaptive layers at the same time, and each layer adopted the multiple-kernel maximum mean discrepancy (MK-MMD) metric. Zhu *et al.* [23] proposed the deep subdomain adaption network (DSAN), which defined the concept of subdomains and used local MDD (LMMD) to align the relevant subdomains, respectively. Physically constrained transfer learning through shared abundance space (PCTL-SAS) for HSI classification was proposed [24], which bridged the gap between SD and TD by projecting the HSI data from SD and TD into a shared abundance space based on their own physical characteristics and had extraordinary performance in real-world sensing scenarios.

Topological properties reflect the nonlocal structure of the data manifold. For example, objects such as buildings, water, and vegetation serve as nodes, and the topological relations of these nodes in the spatial distribution form the unique characteristics in the urban scene. Scenes of the same type, e.g., Houston 2013 and Houston 2018, as shown in Section IV-A, have similar topological relationships that describe the characteristics. The topological relationship reflecting the overall characteristics is completely not limited by the coordinates and still maintains the original properties when deformation occurs. This invariant feature is beneficial to the classification performance, especially for cross-scene with domain shift.

In cross-scene classification, the domain shift of SD and TD in different scenes is usually aligned with the local semantic information extracted by CNN. However, the local spatial relationship between the two scenes does not have a strong correspondence, and the inherent topological relationship between land cover classes needs to be utilized and better aligns two domains from the nonlocal spatial relationship. Recently, GCN has been proposed and successfully applied to the representation and analysis of data in non-Euclidean space (i.e., graph structure data) by modeling the relationships between samples (or vertices). Fig. 1 shows the structure diagram of GCN. In Fig. 1, compared to CNN only extracting local spatial information in HSI patches, GCN constructs a graph for middle- and long-range land cover classes, which fully considers the inherent topological relationships. From the perspective of constructing convolution operators, GCN is divided into the convolution theorem-based GCN (i.e., spectral method) [25], [26] and aggregation function-based GCN (i.e., spatial method) [27]. At present, GCN-based methods are not as popular as CNN in HSI classification, and there are only a few related works. In all these works, the methods of constructing graphs can be divided into two types: constructing graphs with pixels as nodes and constructing graphs with an average spectrum of superpixels as nodes. The first method builds an accurate graph based on the original spectral characteristics and spatial topological relationships. However, there are too many nodes in the graph, and the calculation cost is high, where the related work is miniGCN [28]. The second method uses segmentation algorithms, such as the simple linear iterative cluster (SLIC), to integrate spatial texture information into the construction of graphs. There are fewer nodes in the graphs, which requires a small calculation consumption. However, it is possible to build inaccurate graphs when the superpixel contains multiple classes. CNN-enhanced GCN [29] and multiscale dynamic GCN [30] adopted the second method of constructing graphs.

The cross-scene HSI classification methods mentioned above have the following three limitations.

1) All existing models are designed from the level of data statistical feature distribution to reduce the spectral shift. However, they have never been considered using the topological structure information of the land cover classes to reduce the spectral shift, that is, GCN is used in HSI cross-scene classification.
2) Most methods rely too much on the quality of the superpixels obtained by the segmentation algorithm, which seriously affects the accuracy of graphs. In addition, the construction of the graph is separated from the training process of the model. A fixed graph is used in the training process, and the node features of the graph cannot be updated adaptively.
3) CNN is used for local spatial information extraction in all models, but the limited ability to model topological

relationships between land cover classes of CNN limits its performance in cross-scene HSI classification.

In order to solve the above issues, a Topological structure and Semantic information Transfer network (**TSTnet**) is proposed, which reduces the spectral shift from two aspects: distribution alignment and graph alignment. Specifically, MMD is used to align the marginal distribution of the two domains, and graph optimal transport (GOT) [31] is used to align the nodes and edges of the graph of the two domains. In addition, the deep features extracted by CNN are used to construct the subgraph of the two domains. With the training of the network, the subgraph is dynamically adjusted, which is better than the method of directly using superpixels to construct the graph.

To the best of our knowledge, this is the first work to introduce GCN to cross-scene HSI classification. The main contributions of this work are summarized as follows.

1) In the proposed TSTnet, the domain alignment is carried out from the perspectives of both statistical distribution and topological structure, which are experimentally for the cross-scene HSI classification task.
2) The process of dynamically updating the graph achieves more accurate feature embedding and graph construction, improves the graph features extracted by GCN, and helps align the topological structure of the two domains.
3) The method of using deep features of CNN to construct the graph combines abundance semantic information and topological structure information, where CCN extracts local spatial information, and GCN extracts nonlocal spatial information. This combination is beneficial to improve the spatial perception capacity of the model.

The rest of this article is organized as follows. Section II reviews recent development in areas of DA and GCN. Section III elaborates on the theoretical framework of the proposed TSTnet. The extensive experiments and analyses are presented in Section IV. Finally, conclusions are drawn in Section V.

II. Related Work

Some techniques used in the proposed method are described in this section. Assume that $\mathbf{X}_s = \{\mathbf{x}_i^s\}_{i=1}^{n_s} \in \mathbb{R}^d$ and $\mathbf{X}_t = \{\mathbf{x}_i^t\}_{i=1}^{n_t} \in \mathbb{R}^d$ are data from SD and TD, respectively, and $\mathbf{Y}_s$ and $\mathbf{Y}_t$ are the corresponding class labels. Note that, there is no $\mathbf{Y}_t$ in the unsupervised case. Here, d, n_s, and n_t denote the dimension of data, the number of source samples, and the number of target samples, respectively. The marginal distribution of SD and TD is expressed as $P(\mathbf{X}_s)$ and $P(\mathbf{X}_t)$, respectively. A graph is represented as $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{G}$ denotes an undirected graph, $\mathbf{V}$ is the node set of the graph, and $\mathbf{E}$ indicates the set of edges connected by nodes $\mathbf{v}_i$ and $\mathbf{v}_j$. Some important abbreviation and variables with their descriptions are presented in Nomenclature.

A. Graph Sampling Aggregation Network (GraphSAGE)

The definition of the convolution operator of the convolution theorem-based GCN depends on the graph Laplacian matrix. Let $\mathbf{L}$ denote the graph Laplacian matrix, $\mathbf{A}$ denote the adjacency matrix, which reflects the relationship between a node and its neighbors, $\mathbf{D}$ denote the degree matrix, which is a diagonal matrix, and $\mathbf{D}_{ii}$ denote the degree of the ith node and is defined as $\mathbf{D}_{ii} = \sum_{j=1}^{n} \mathbf{A}_{i,j}$. The spectral convolution of GCN is defined as follows:

$$\mathbf{H}^{(\ell+1)} = \sigma\left(\widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(\ell)}\mathbf{W}^{(\ell)}\right) \tag{1}$$

where $\mathbf{H}^{(\ell)}$ and $\mathbf{W}^{(\ell)}$ indicate the ℓth layer output and trainable parameters, $\widetilde{\mathbf{D}}^{-(1/2)}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-(1/2)}$ is the symmetrical normalized form of $\mathbf{L}$, which prevents the gradient from exploding or disappearing in the deep GCN, $\sigma(\bullet)$ is an activation function, $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, and $\widetilde{\mathbf{D}}_{i,i} = \sum_j \widetilde{\mathbf{A}}_{i,j}$.

The aggregation function-based GCN defines the graph convolution operator from a spatial perspective. Adjacent nodes are randomly sampled in the GraphSAGE, and the number of adjacent nodes in each node is smaller than a given number of samples. Fig. 2 illustrates GraphSAGE forward propagation schematic, with $\mathbf{v}_1$ and $\mathbf{v}_2$ as the central nodes, the first-order neighbors nodes ($k = 1$) are randomly sampled, and only the sampled nodes are regarded as related nodes. Then, GraphSAGE aggregates adjacent nodes to update the feature expressions of central nodes. It can be seen from Fig. 2 that, during the information dissemination, the node information is extended to k-order neighbors after aggregation for k times. For example, the node $\mathbf{v}_1$ in the third layer has aggregated its first-order ($k = 1$) and second-order ($k = 2$) information.

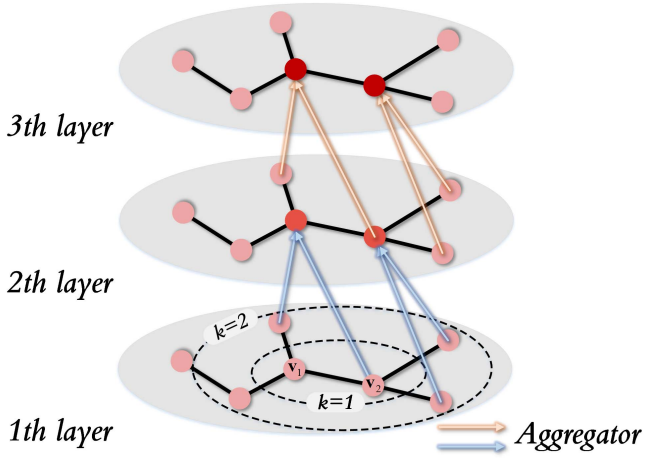

Fig. 2. GraphSAGE forward propagation schematic.

B. Graph Optimal Transport

Cross-domain alignment is expressed as a graph matching problem and solved by calculating the distance similarity based on graph. GOT uses the cosine distance to measure the similarity between nodes and uses two optimal transport distances: WD for node matching and GWD for edge matching. The formula of WD can be described as

$$\mathcal{D}_{\mathrm{WD}}[P(\mathbf{X}_s), P(\mathbf{X}_t)] = \inf_{\gamma \in \Pi[P(\mathbf{X}_s), P(\mathbf{X}_t)]} E_{(\mathbf{x}^s, \mathbf{x}^t)\sim\gamma}[c(\mathbf{x}^s, \mathbf{x}^t)]$$
$$c\left(\mathbf{x}_i^s, \mathbf{x}_j^t\right) = 1 - \frac{\left(\mathbf{x}_i^s\right)^T \mathbf{x}_j^t}{\left\|\mathbf{x}_i^s\right\|_2 \left\|\mathbf{x}_j^t\right\|_2} \tag{2}$$

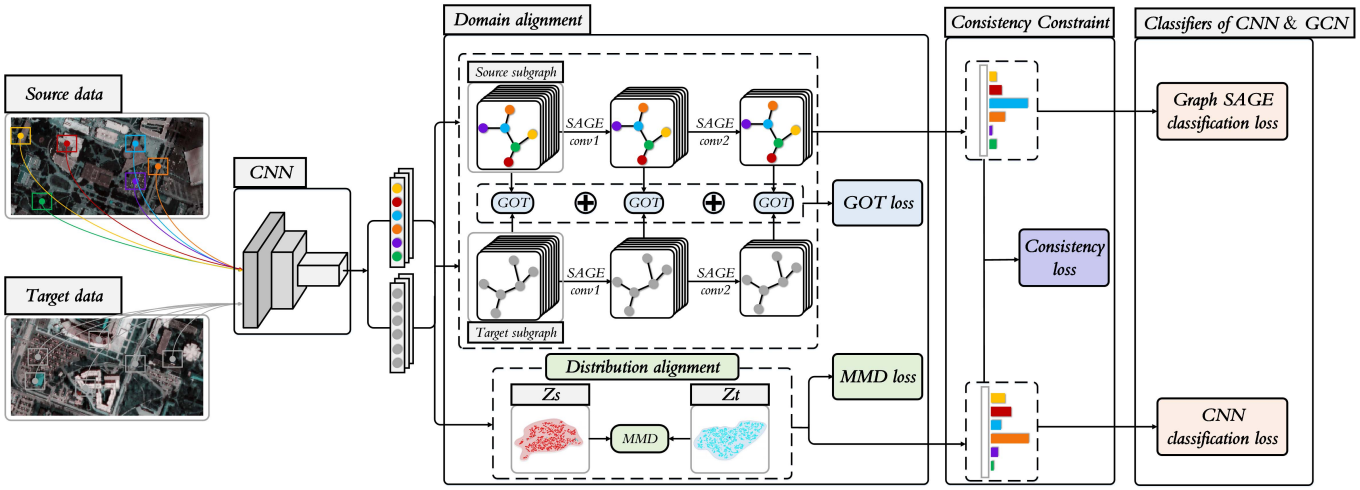

Fig. 3. Flowchart of the proposed TSTnet, including graph alignment, distribution alignment, and consistency constraint. Different colors in the SD represent labels from different classes, and gray in the TD represents no any label information.

where $c(\mathbf{x}^s, \mathbf{x}^t)$ is the cross-domain cost matrix obtained by the cosine distance and $\Pi[P(\mathbf{X}_s), P(\mathbf{X}_t)]$ denotes all the joint distributions $\gamma(\mathbf{x}^s, \mathbf{x}^t)$. GWD can be stated as

$$\mathcal{D}_{\mathrm{GWD}}[P(\mathbf{X}_s), P(\mathbf{X}_t)] = \inf_{\gamma \in \Pi[P(\mathbf{X}_s), P(\mathbf{X}_t)]} E_{\substack{(\mathbf{x}^s, \mathbf{x}^t) \sim \gamma \\ (\hat{\mathbf{x}}^s, \hat{\mathbf{x}}^t) \sim \gamma}} \left[L\left(\mathbf{x}_i^s, \mathbf{x}_j^t, \hat{\mathbf{x}}_i^s, \hat{\mathbf{x}}_j^t\right) \right]$$
$$L\left(\mathbf{x}_i^s, \mathbf{x}_j^t, \hat{\mathbf{x}}_i^s, \hat{\mathbf{x}}_j^t\right) = \left\| c_1\left(\mathbf{x}_i^s, \hat{\mathbf{x}}_i^s\right) - c_2\left(\mathbf{x}_j^t, \hat{\mathbf{x}}_j^t\right) \right\| \quad (3)$$

where $L(\mathbf{x}_i^s, \mathbf{x}_j^t, \hat{\mathbf{x}}_i^s, \hat{\mathbf{x}}_j^t)$ is the cost function evaluating the intragraph structural similarity between two pairs of nodes $(\mathbf{x}_i^s, \hat{\mathbf{x}}_i^s)$ and $(\mathbf{x}_j^t, \hat{\mathbf{x}}_j^t)$. A transport plan $\mathbf{T}$ shared by WD and GWD was proposed to integrate the WD and GWD and obtain the GOT distance. The details of GOT can be seen in [31].

C. Maximum Mean Discrepancy

Among the existing methods, MMD [32], as an effective nonparametric distance metric, is one of the most widely used strategies to measure the distribution difference between SD and TD. The MMD distance of two different domains is described as

$$\mathrm{MMD}(\mathbf{X}_s, \mathbf{X}_t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi\left(\mathbf{x}_i^s\right) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi\left(\mathbf{x}_j^t\right) \right\|_H^2. \quad (4)$$

The meaning of the formula is to calculate the mean discrepancy between SD and TD in reproducing kernel Hilbert space (RKHS). The smaller the value of MMD, the more similar the two domains, and vice versa. By square expansion of the formula, the inner product in RKHS space can be converted into a kernel function, so the MMD can be calculated directly through the kernel function.

III. Proposed Domain Adaptation Method

The flowchart of the proposed TSTnet is shown in Fig. 3, which contains three parts, i.e., domain alignment (graph alignment and distribution alignment), the consistency constraint, and classifiers of CNN and GCN. Specifically, patches from SD (with labels) and patched from TD (without any label) are randomly selected and put into CNN to extract features with abundant semantic information. The features of SD and TD extracted by CNN are denoted as $\mathbf{Z}_s$ and $\mathbf{Z}_t$, which are used for the distribution alignment and subgraph construction of the two domains.

In the GCN branch, GraphSAGE is adopted as the graph convolution layer, and the features of each layer are aligned by GOT. The graph features of SD and TD extracted by GraphSAGE are denoted as $\mathbf{G}_s$ and $\mathbf{G}_t$. Finally, $\mathbf{Z}_s$ and $\mathbf{G}_s$ are used for probability prediction, which is imposed a constraint of distribution consistency. In the final test stage, CNN that can extract domain invariant features is used to classify TD.

A. Domain Alignment

The MMD distance is used in the marginal distribution alignment. In practice, the unbiased estimation of MMD is to compare the squared distance between the empirical kernel mean embeddings as

$$\mathcal{L}_{\mathrm{MMD}}(\mathbf{Z}_s, \mathbf{Z}_t) = \frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} r\left(\mathbf{z}_i^s, \mathbf{z}_j^s\right) + \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} r\left(\mathbf{z}_i^t, \mathbf{z}_j^t\right) - \frac{2}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} r\left(\mathbf{z}_i^s, \mathbf{z}_j^t\right) \quad (5)$$

where the kernel $r(\cdot, \cdot)$ represents inner product of the features in RKHS.

As for the alignment of topological structure information, first, the mean aggregation function is adopted in GraphSAGE, and its formula is

$$\mathbf{G}_{\mathbf{v}}^{\ell} \leftarrow \sigma\left(\mathbf{W}^{\ell-1} \cdot \mathrm{Mean}\left(\left\{\mathbf{G}_{\mathbf{v}}^{\ell-1}\right\} \cup \left\{\mathbf{G}_{\mathbf{u}}^{\ell-1}, \forall \mathbf{u} \in \mathcal{N}(\mathbf{v})\right\}\right)\right) \quad (6)$$

where $\mathbf{v}$ and $\mathbf{u}$ are the central node and neighboring nodes, $\mathcal{N}(\mathbf{v})$ is the neighborhood function, which controls the order of selecting neighboring nodes, $\sigma(\bullet)$ is the nonlinear activation function, $\mathbf{W}^{\ell-1}$ is the weight matrix of the ℓth layer, and Mean$(\bullet)$ is to add the central node and the selected

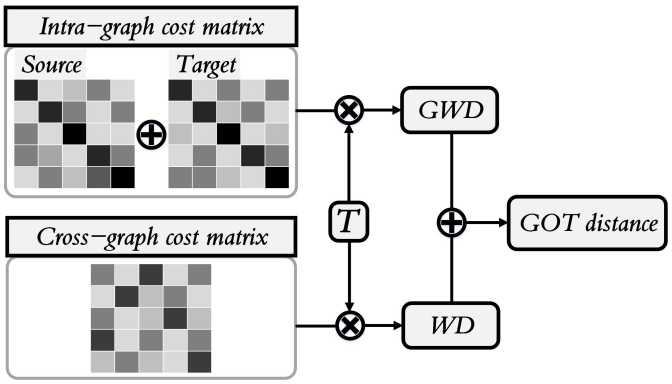

Fig. 4. Calculation process of GOT.

neighboring nodes. Two layers of GraphSAGE based on the mean aggregation function are used in TSTnet, and then, WD and GWD are calculated for the graph features of each layer. It is worth noting that WD measures the distance of node features between two domains without considering the topological information. On the other hand, GWD compares graph structures by measuring the distance between a pair of nodes in each intragraph. The fusion of these two distances can effectively consider the node and edge information, so as to achieve better graph matching. GWD only considers the similarity between $c_1(\mathbf{x}_i^s, \hat{\mathbf{x}}_i^s)$ and $c_2(\mathbf{x}_j^t, \hat{\mathbf{x}}_j^t)$ but does not consider node representation and cannot be directly applied to graph alignment. In order to achieve the best integration of WD and GWD, a transformation matrix $\mathbf{T}$ shared by WD and GWD was proposed [31]. Fig. 4 illustrates the calculation process of GOT. The GOT distance is defined as

$$\begin{aligned}\mathrm{GOT}(\mathbf{G}_s, \mathbf{G}_t) = \sum_{i,\hat{i},j,\hat{j}} \big(\mathbf{T}_{ij}\gamma \cdot c\big(\mathbf{g}_i^s, \mathbf{g}_j^t\big) \\ + (1-\gamma)\cdot \mathbf{T}_{\hat{i}\hat{j}} L\big(\mathbf{g}_i^s, \mathbf{g}_j^t, \hat{\mathbf{g}}_i^s, \hat{\mathbf{g}}_j^t\big)\big) \quad (7)\end{aligned}$$

where $\mathbf{T}_{ij}$ and $\mathbf{T}_{\hat{i}\hat{j}}$ are the transformation matrices of two pairs of nodes [i.e., $(\mathbf{g}_i^s, \mathbf{g}_j^t)$ and $(\hat{\mathbf{g}}_i^s, \hat{\mathbf{g}}_j^t)$] and γ is set to 0.5 in the experiment. Thus, the GOT loss can be formulated as

$$\mathcal{L}_{\mathrm{GOT}}(\mathbf{Z}_s, \mathbf{Z}_t) = \sum_{\ell=0}^{2} \mathrm{GOT}\big(\mathbf{G}_s^\ell, \mathbf{G}_t^\ell\big) \quad (8)$$

where $\mathbf{G}^\ell$ is the ℓth layer output of GCN, $\ell = 0$ represents the original subgraph features constructed by $\mathbf{Z}_s$ and $\mathbf{Z}_t$, and $\ell = 1$ and $\ell = 2$ are the graph convolution features of the first layer and the second layer, respectively.

B. Consistency Constraint

Different from simple feature fusion, in order to better make the distributed alignment and topological relationship alignment work together in the training process and to learn better domain invariant features, the consistency constraint of the probability prediction distribution is introduced into TSTnet. The cross-entropy function can be used to calculate the difference between the learning model distribution and the training distribution, so it is used to evaluate the distribution consistency

$$\mathcal{L}_{\mathrm{consistency}}(\mathbf{Z}_s, \mathbf{G}_s) = -\frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{c=1}^{C} p_i^c \log q_i^c \quad (9)$$

where C is the number of class, p_i^c and q_i^c are the probability prediction of CNN and GCN, and $\mathcal{L}_{\mathrm{consistency}}(\bullet)$ denotes the constraint of distribution consistency on the probability predictions of CNN and GCN.

C. Classifiers of CNN and GCN

The cross-entropy loss is used in both CNN and GCN when predicting the probability of SD and using the label to calculate the error. The cross-entropy loss of the labeled SD $\mathbf{x}_i^s$ is defined as

$$\mathcal{L}_y(\mathbf{p}_i, \mathbf{y}_i) = -\sum_{c=1}^{C} y_i^c \log p_i^c \quad (10)$$

where $\mathbf{y}_i$ is the one-hot encoding of the label information of $\mathbf{x}_i^s$ and $\mathbf{p}_i$ is the predicted probability output obtained by the softmax function. Therefore, the classification loss of CNN and GCN on SD is defined as

$$\begin{aligned}\mathcal{L}_{\mathrm{CNN}}(\mathbf{X}_s, \mathbf{Y}_s) &= \frac{1}{n_s}\sum_{i=1}^{n_s}\mathcal{L}_y\big(S\big(\mathbf{z}_i^s\big), \mathbf{y}_i\big) \\ \mathcal{L}_{\mathrm{GCN}}(\mathbf{Z}_s, \mathbf{Y}_s) &= \frac{1}{n_s}\sum_{i=1}^{n_s}\mathcal{L}_y\big(S\big(\mathbf{g}_i^s\big), \mathbf{y}_i\big) \quad (11)\end{aligned}$$

where $S(\bullet)$ denotes the softmax function.

Integrating above loss functions, the total loss of TSTnet is defined as follows:

$$\begin{aligned}\mathcal{L}(\mathbf{X}_s, \mathbf{X}_t, \mathbf{Y}_s) = \mathcal{L}_{\mathrm{CNN}}(\mathbf{X}_s, \mathbf{Y}_s) + \mathcal{L}_{\mathrm{GCN}}(\mathbf{Z}_s, \mathbf{Y}_s) \\ + \mathcal{L}_{\mathrm{consistency}}(\mathbf{Z}_s, \mathbf{G}_s) + \lambda_1 \mathcal{L}_{\mathrm{MMD}}(\mathbf{Z}_s, \mathbf{Z}_t) \\ + \lambda_2 \mathcal{L}_{\mathrm{GOT}}(\mathbf{Z}_s, \mathbf{Z}_t) \quad (12)\end{aligned}$$

where λ_1 and λ_2 are the regularization parameters controlling the different contributions of distribution alignment and graph alignment.

The main idea of CNN-based DA is to reduce the spectral shift from the perspective of aligning the data statistical feature distribution, and the topological structure information extracted by GCN has never been considered for DA. In order to make up for this shortcoming, the proposed TSTnet cooperatively reduces the spectral shift and learns domain invariant features via distribution alignment and graph alignment. In addition, a new method of constructing graphs is proposed, which is to use the semantic features of the spatial patch information as node features and dynamically adjust the subgraphs of the two domains during the training process with the change of node features extracted by CNN.

D. Alignment Performance of TSTnet

The most concerned problem in cross-scene HSI classification tasks is the spectral shift. In other words, there are differences in spectral reflectance between SD and TD of the same land cover class. Fig. 5(a), (c), (e), (g), (i), and (k)

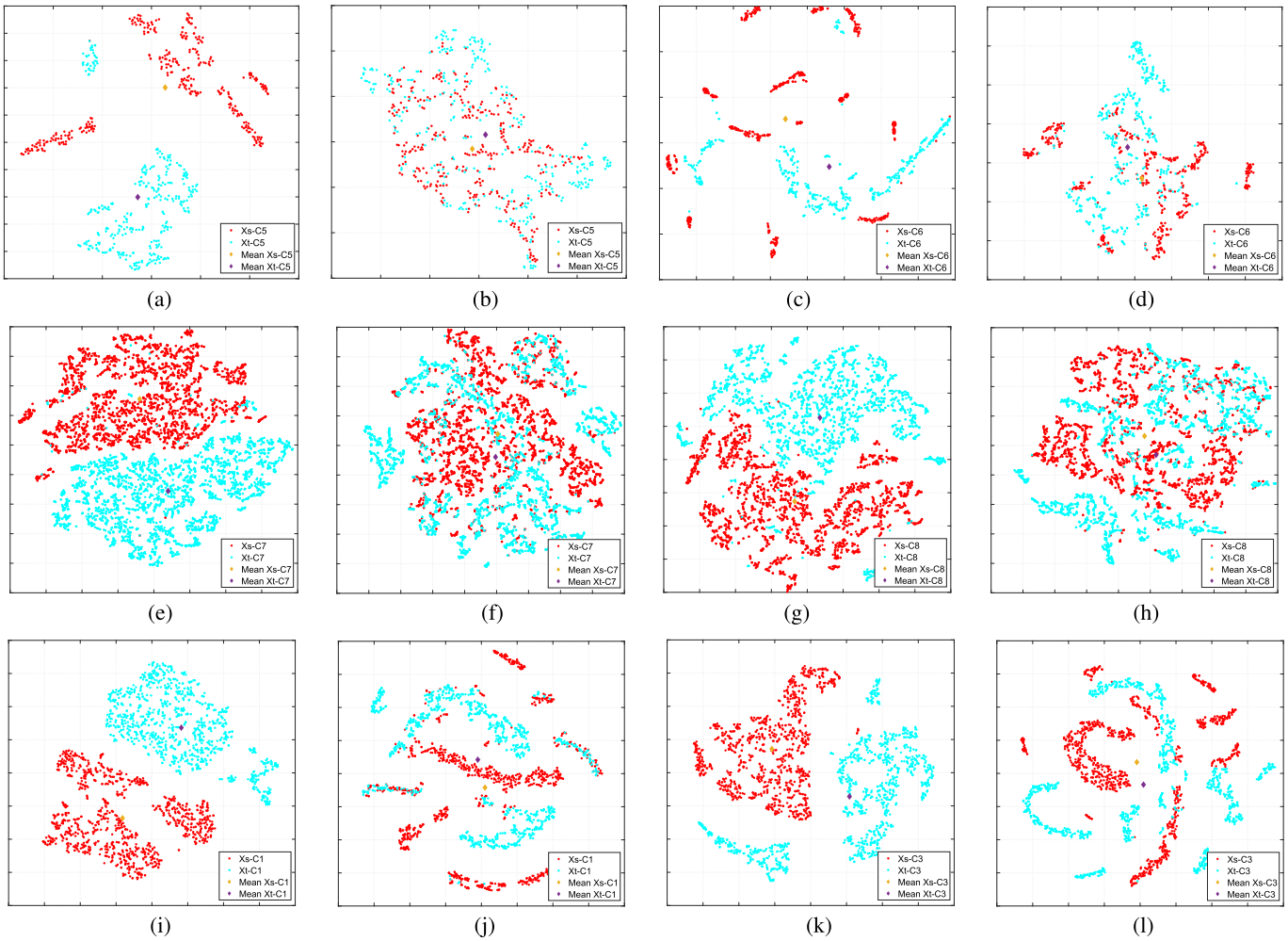

Fig. 5. Alignment performance of the proposed TSTnet using the Houston data, HyRANK data, and S-H data (OS-original samples and AS-aligned samples). (a) OS from Houston (Class 5). (b) AS by TSTnet (Class 5). (c) OS from Houston (Class 6). (d) AS by TSTnet (Class 6). (e) OS from HyRANK (Class 7). (f) AS by TSTnet (Class 7). (g) OS from HyRANK (Class 8). (h) AS by TSTnet (Class 8). (i) OS from S-H (Class 1). (j) AS by TSTnet (Class 1). (k) OS from S-H (Class 3). (l) AS by TSTnet (Class 3).

illustrates this phenomenon, which is the 2-D visualization of the original samples obtained from the Houston data, HyRANK data, and Shanghai–Hangzhou (S-H) data (the data are detailed in Section IV-A), the red and blue points represent SD and TD, respectively, and the purple and yellow points are the centroids of the distribution. It is obvious that the distribution of original samples from the same class in the two domains is inconsistent, and their centroids are far apart.

Through the proposed transfer learning scheme, SD and TD data are projected into the domain invariant subspace, so that the model trained on SD can be directly used for TD without data labeling or model retraining. This is mainly because the spectral shift is greatly reduced under proper physical constraints. To further demonstrate alignment performance, 256-dimensional domain invariant features of the trained TSTnet are outputted, and the distribution of these features is shown in Fig. 5. It is observed that the spectral shift on all datasets is greatly alleviated, and the distribution of the same class of SD and TD overlaps with each other, particularly the fifth class of Houston data and the seventh and eighth classes of HyRANK. Taking the fifth class of Houston data as an example [see Fig. 5(a) and (b)], compared to Fig. 5(a), the separate distribution of the fifth class of SD and TD in the original feature space is overlapped in the alignment space, as shown in Fig. 5(b). This phenomenon indicates that the domain shift is alleviated in the proposed TSTnet by representing the topological relationship of land cover classes and coordinating the distribution and graph alignment. Compared with the other two datasets, the S-H dataset is extremely noisy. For example, the third class should contain more than several types of vegetation. Therefore, even if the domain differences are reduced in this dataset, the distribution of the same class cannot overlap with each other.

IV. Experimental Results and Discussion

In this section, three cross-scene HSI datasets, i.e., the Houston data, HyRANK data, and S-H data, are conducted to verify the effectiveness of the proposed TSTnet. Several classic and state-of-the-art unsupervised deep domain adaptive algorithms are employed for comparison algorithms, including support vector machine (SVM), which is a general nonlinear

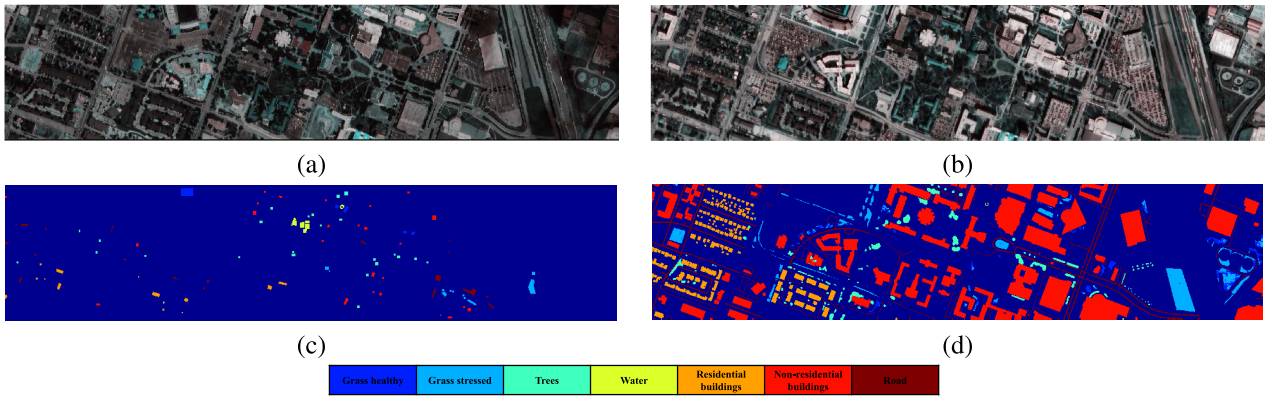

Fig. 6. Pseudocolor image and ground-truth map of Houston. (a) Pseudocolor image of Houston 2013. (b) Pseudocolor image of Houston 2018. (c) Ground-truth map of Houston 2013. (d) Ground-truth map of Houston 2018.

TABLE I

NUMBER OF SOURCE AND TARGET SAMPLES FOR THE HOUSTON DATASET AND 5% OF TRAINING LABELS FROM SD

Class		Number of Samples	
No.	Name	Houston 2013 (Source)	Houston 2018 (Target)
1	Grass healthy	345	1353
2	Grass stressed	365	4888
3	Trees	365	2766
4	Water	285	22
5	Residential buildings	319	5347
6	Non-residential buildings	408	32459
7	Road	443	6365
Total		2530	53200

TABLE II

NUMBER OF SOURCE AND TARGET SAMPLES FOR THE HYRANK DATASET AND 5% OF TRAINING LABELS FROM SD

Class		Number of Samples	
No.	Name	Dioni (Source)	Loukia (Target)
1	Dense Urban Fabric	1262	206
2	Mineral Extraction Sites	204	54
3	Non Irrigated Arable Land	614	426
4	Fruit Trees	150	79
5	Olive Groves	1768	1107
6	Coniferous Forest	361	422
7	Dense Sderophyllous Vegetation	5035	2996
8	Sparce Sderophyllous Vegetation	6374	2361
9	Sparcely Vegetated Areas	1754	399
10	Rocks and Sand	492	453
11	Water	1612	1393
12	Coastal Water	398	421
Total		20024	10317

version, and its kernel adopts radial basis function (RBF), DAN [22], DeepCoral [33], DSAN [23], DAAN [34], multi-representation adaptation network (MRAN) [35], and HTCNN [36]. A recent work on GCN-based HSI classification, CEGCN [29], is also employed. The comparison algorithms and TSTnet use SD and TD data and SD labels during training and do not use TD labels information. In the experiment, 5% of samples are selected from the SD of the three datasets, and the data enhancement operations of rotation and noise are performed. The class-specific accuracy (CA), the overall accuracy (OA), and the Kappa coefficient (KC) are employed to evaluate the classification performance.

A. Experiment Data

Houston: The dataset includes Houston 2013 [37] and Houston 2018 [38] scenes that were obtained by different sensors on the University of Houston campus and its vicinity in different years. The Houston 2013 dataset is composed of 349 $\times$ 1905 pixels, including 144 spectral bands; the wavelength range is 380–1050 nm; and the image spatial resolution is 2.5 m. The Houston 2018 dataset has the same wavelength range but contains 48 spectral bands, and the image has a spatial resolution of 1 m. There are seven consistent classes in their scene. We extract 48 spectral bands (wavelength range 0.38$\sim$1.05 μm) from the Houston 2013 scene corresponding to the Houston 2018 scene and select the overlapping area of 209 $\times$ 955. The name of land cover classes and the number of samples are listed in Table I. In addition, their pseudocolor representations and ground-truth maps are shown in Fig. 6.

HyRANK: The HyRANK dataset has been developed in the framework of the International Society for Photogrammetry and Remote Sensing (ISPRS) Scientific Initiatives [39]. The satellite HSI collected by the Hyperion sensor (EO-1, USGS) has 176 spectral bands. The two labeled scenes are Dioni and Loukia, which are composed of 250 $\times$ 1376 pixels and 249 $\times$ 945 pixels, respectively. There are 12 consistent classes, which are listed in Table II. The vegetation classes are prone to be misclassified. The pseudocolor representations and ground-truth maps of two scenes are shown in Fig. 7.

S-H: The Shanghai and Hangzhou datasets were captured by the EO-1 Hyperion hyperspectral sensor, which retains 198 bands after removing the bad bands. The size of the Shanghai data image is 1600 $\times$ 230, which includes roads, buildings, plants, and the waters of the Yangtze River and Huangpu River. The Hangzhou data image size is 590 $\times$ 230, including roads, buildings, plants, West Lake, and the Qiantang River basin. Three common land cover classes were selected for cross-scene classification, which is listed in Table III, and their pseudocolor representations and ground-truth maps are shown in Fig. 8.

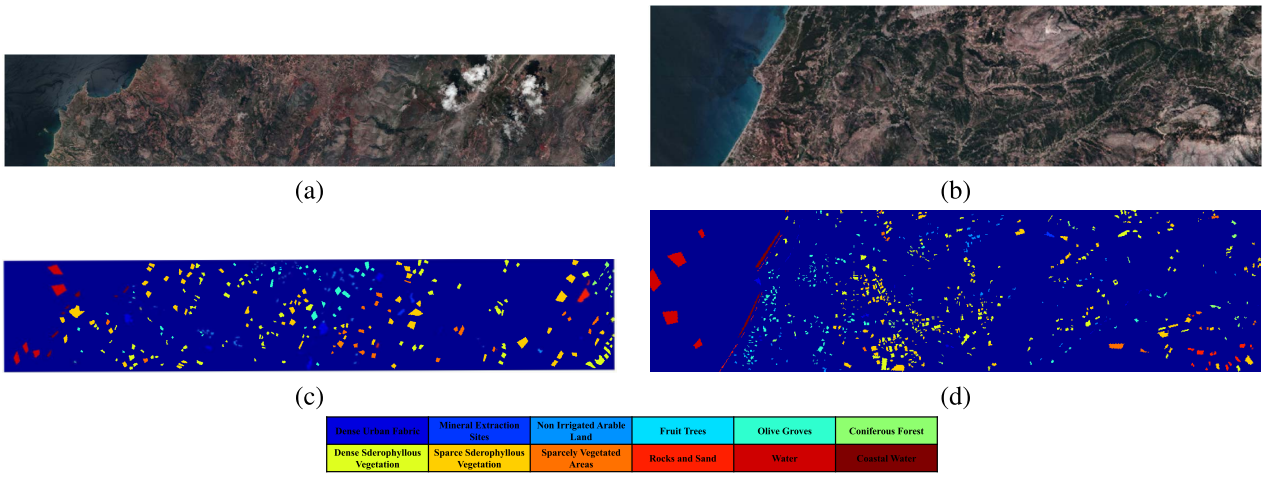

Fig. 7. Pseudocolor image and ground-truth map of HyRANK. (a) Pseudocolor image of Dioni. (b) Pseudocolor image of Loukia. (c) Ground-truth map of Dioni. (d) Ground-truth map of Loukia.

TABLE III
NUMBER OF SOURCE AND TARGET SAMPLES FOR THE S-H DATASET AND 5% OF TRAINING LABELS FROM SD

Class		Number of Samples	
No.	Name	Hangzhou (Source)	Shanghai (Target)
1	Water	18043	123123
2	Land/Building	77450	161689
3	Plant	40207	83188
Total		135700	368000

TABLE IV
PARAMETER TUNING OF γ FOR THE PROPOSED TSTNET USING THE HOUSTON DATA

Data set	Houston 2018				
	0.1	0.3	0.5	0.7	0.9
OA (%)	72.28±4.52	75.13±1.92	**75.34±1.96**	73.13±3.12	71.09±1.06
KC (κ)	55.37±5.24	59.43±3.44	**59.69±3.56**	57.42±4.12	53.41±2.91

TABLE V
PARAMETER TUNING OF THE BASE LEARNING RATE η_0 FOR THE PROPOSED TSTNET USING THE THREE EXPERIMENTAL DATA

Data set	Base learning rate η_0				
	1e-5	1e-4	1e-3	1e-2	1e-1
Houston	34.4	38.88	70.4	**76.65**	14.96
HyRANK	23.1	53.94	**63.94**	59.52	59.26
Shanghai-Hangzhou	**90.19**	59.96	78.31	83.52	74.81

TABLE VI
NETWORK CONFIGURATION OF TSTNET

TSTnet	CNN	GraphSAGE
Input Dimension	bz×13×13×d	-
Semantic Feature Extraction	3×3 Conv2d BN2d LeakyReLU	-
	bz×11×11×32	-
	3×3 Conv2d BN2d LeakyReLU	-
	bz×9×9×32	-
	3×3 Conv2d BN2d LeakyReLU	-
	bz×7×7×64	-
	3×3 Conv2d BN2d LeakyReLU	-
	bz×5×5×64	-
Topological Structure Extraction	-	bz×1600
	-	SAGEConv(aggr=mean) BN1d
	-	bz×64
	-	SAGEConv(aggr=mean) BN1d
	-	bz×32
Multilayer Perceptron	FC ReLU Dropout(0.5)	FC ReLU Dropout(0.5)
	bz×4096	bz×1024
	FC ReLU Dropout(0.5)	FC ReLU Dropout(0.5)
	bz×256	bz×1024
	FC	FC
Output Dimension	C	C

B. Parameter Tuning

In the proposed TSTnet, adjustable hyperparameters are the regularization parameters and base learning rate, i.e., λ_1, λ_2, and η_0. The tenfold cross-validation is adopted to select the optimal parameters. The value ranges of the regularization parameters and base learning rate are $\{1e-4, 1e-3, 1e-2, 1e-1, 1e+0, 1e+1, 1e+2\}$ and $\{1e-5, 1e-4, 1e-3, 1e-2, 1e-1\}$.

λ_1 and λ_2 are important hyperparameters of TSTnet, which controls the contribution of semantic information and topological structure information to cross-domain classification. Fig. 9 shows the changing trend of the classification accuracy of TSTnet in three experimental datasets with different combinations of λ_1 and λ_2 (indicated by OA). In order to reduce the pressure of hyperparameter optimization, we choose suboptimal parameters. It can be seen from Fig. 9 that the proposed method with parameter λ_1 in the range of $1e+0$ to $1e+1$ can achieve suboptimal classification performance,

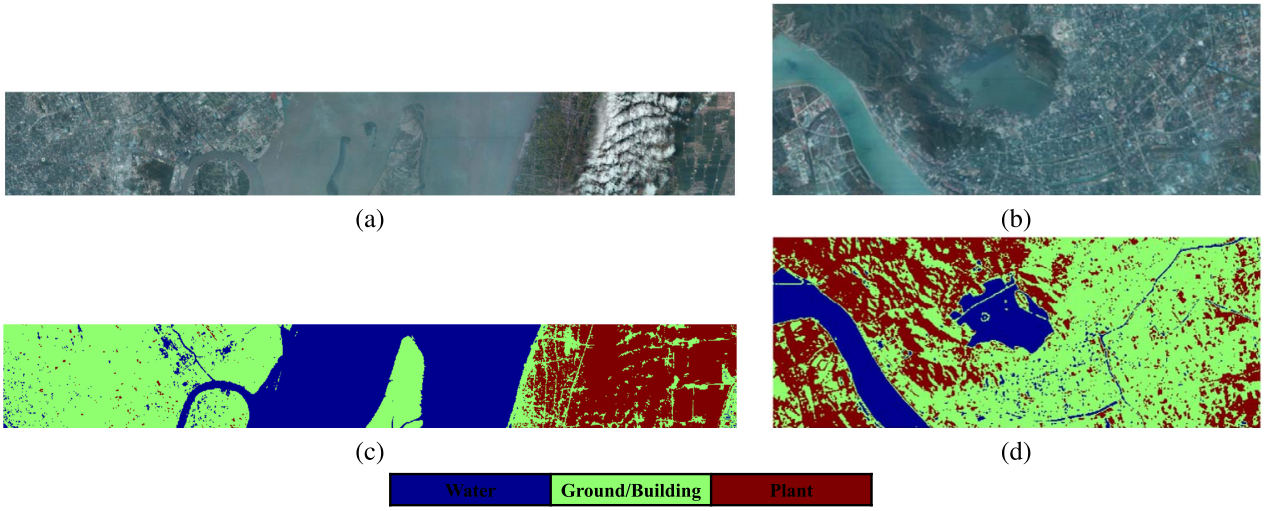

Fig. 8. Pseudocolor image and ground-truth map of S-H. (a) Pseudocolor image of Shanghai. (b) Pseudocolor image of Hangzhou. (c) Ground-truth map of Shanghai. (d) Ground-truth map of Hangzhou.

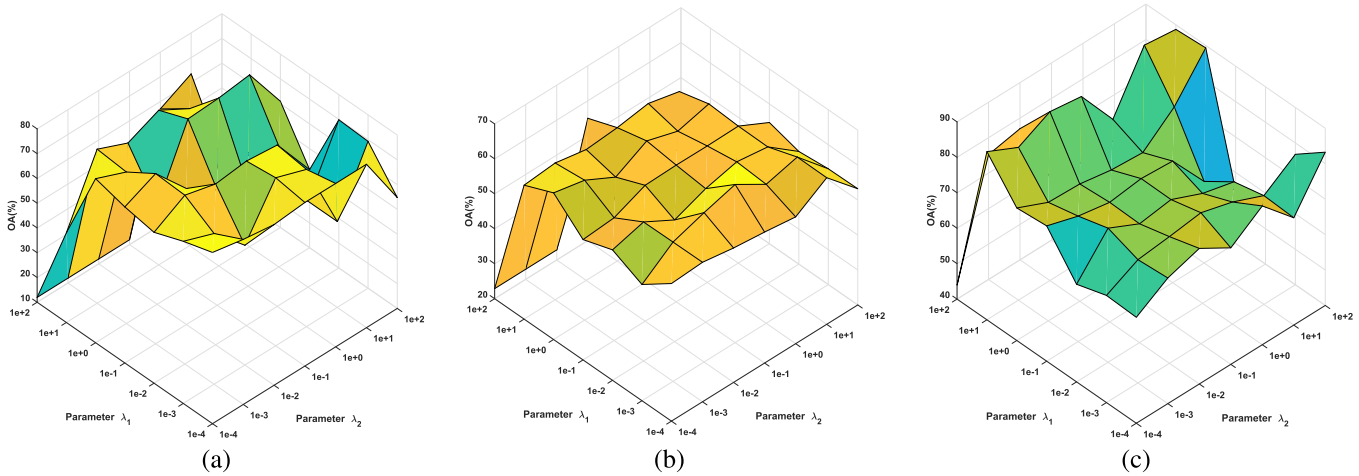

Fig. 9. Parameter tuning of λ_1 and λ_2 for the proposed TSTnet using all the three experimental data. (a) Houston. (b) HyRANK. (c) S-H.

which indicates that TSTnet is less sensitive to λ_1, and λ_2 has its own suboptimal range in the three datasets. The above analysis confirms that the topological structure information in the proposed TSTnet plays a more important role in learning the domain invariant subspace than the semantic information. Therefore, the suboptimal parameters λ_1 and λ_2 corresponding to the Houston data are $1e+0$ and $1e+1$; for the HyRANK data, and they are $1e+0$ and $1e-1$; for S-H data, they are $1e+0$ and $1e+2$, respectively.

The hyperparameter γ is used to adjust the weight of WD and GWD in the GOT distance. The classification accuracy of TSTnet on Houston data with different weights is listed in Table IV, which indicates that the node distance and edge distance of the two domains can cooperate better under the adjustment of equal weight.

The learning rate controls the step of gradient descent in the training process and also affects the learning effect of the model. The effect of different base learning rates on TSTnet in all datasets is tested, and Table V provides the classification results. The optimal base learning rate corresponding to the Houston data is $1e-2$; for the HyRANK data, it is $1e-3$; and for the S-H data, it is $1e-5$.

The proposed method is implemented using the Pytorch platform, and the torch geometric library is used to implement GraphSAGE. Table VI lists the network configuration of the proposed TSTnet, where Conv, BN, FC, and bz stand for convolution, batch normalization, full connection layer, and batch size. In the semantic feature extraction part, Conv2d and BN2d use the Kaiming distribution and constant 1 initialization parameters, respectively, the FC in the multilayer perceptron uses the normal distribution $N(0, 0.001)$, and the rest part utilizes the default parameter initialization in the library. The minibatch stochastic gradient descent (SGD) with momentum of 0.9 is taken as optimization scheme, and the strategy of changing the learning rate is adjusted according to the following formula: $\eta_w = (\eta_0/(1+\alpha w)^\beta)$, where w is the training progress linearly changing from 0 to 1, η_0 is the base learning rate, $\alpha = 10$, and $\beta = 0.75$. Moreover, the batch size in the training progress is set to 100, the ℓ_2-norm regularization is set as $1e-4$, which is used for weight decay

TABLE VII
ABLATION COMPARISON OF EACH MODULE OF TSTNET AND USING 5% LABELS FROM SD

Data set	Houston 2018			Loukia			Shanghai		
	CNN(MMD)	TSTnet (no consistency)	TSTnet	CNN(MMD)	TSTnet (no consistency)	TSTnet	CNN(MMD)	TSTnet (no consistency)	TSTnet
OA (%)	62.25±4.71	72.93±2.75	**75.34±1.96**	54.00±1.68	62.33±2.36	**63.31±3.05**	83.46±3.03	89.19±1.99	**90.36±1.35**
KC (κ)	40.64±5.47	54.65±7.48	**59.69±3.56**	40.40±2.23	54.26±2.89	**55.37±3.82**	75.04±4.03	83.27±3.34	**85.27±2.02**

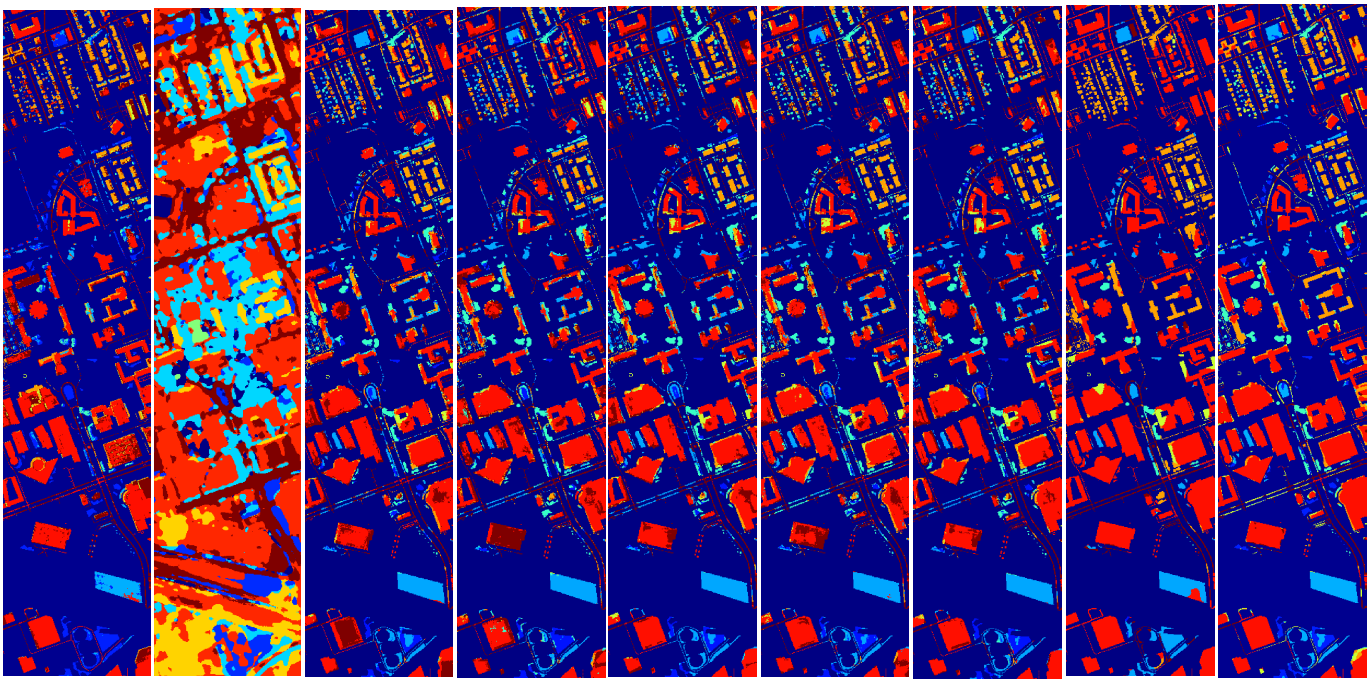

Fig. 10. Visualization and classification maps using 5% labels from SD for the target scene Houston 2018 obtained with different methods, including (a) SVM (65.40%), (b) CEGCN [29] (65.14%), (c) DAN [22] (66.05%), (d) DeepCoral [33] (61.26%), (e) DAAN [34] (66.59%), (f) MRAN [35] (66.95%), (g) DSAN [23] (67.08%), (h) HTCNN [36] (74.72%), and (i) TSTnet (76.12%).

and can stabilize network training to reduce overfitting, and the number of training iterations is 500.

C. Ablation Study

The contributions of the important components of TSTnet are evaluated. Looking back at the loss function [see (12)], TSTnet mainly includes two DA strategies, graph alignment and distribution alignment, and the consistency constraint term is designed to achieve better collaboration between them. The ablation analysis is performed on three datasets to demonstrate the validity of each item.

First, in order to verify the effectiveness of GCN and the graph alignment, only CNN and the distribution alignment are kept for comparison, denoted as CNN (MMD), and second, only the consistency constraint is deleted from TSTnet for comparison to prove the validity of this constraint, denoted as TSTnet (no consistency). The mean and standard deviation of ten experiments performed by the two ablation models are listed in Table VII. The performance of CNN (MMD) that only performs distribution alignment is significantly lower than TSTnet, which is due to the limitation of the ability of CNN to model sample relationships and the limitation of a single DA strategy. The OA of TSTnet without consistency constraint is 1%∼2% lower than TSTnet on all datasets, which reflects that the consistency constraint item contributes to the collaboration of graph alignment and distribution alignment, and improves the learned domain invariant features.

D. Performance on Cross-Scene HSI Classification

Furthermore, three cross-scene HSI datasets are employed to evaluate the performance of TSTnet. Comparison algorithms include CEGCN, DAN, DeepCoral, DSAN, DAAN, MRAN, and HTCNN. The optimal base learning rate of all comparison algorithms is selected from $\{1e-5, 1e-4, 1e-3, 1e-2, 1e-1\}$, and the number of training iterations is set to 500. In particular, CEGCN has one hyperparameter, the segmentation scale controls the number of superpixels, and its optimal value is selected from {100, 200, 300, 400, 500}. DAN, DeepCoral, DSAN, DAAN, and MRAN have one regularization parameter. In the following experiment, the candidate regularization parameter set is $\{1e-4, 1e-3, 1e-2, 1e-1, 1e+0, 1e+1, 1e+2\}$ and using cross-validation to find the corresponding optimal parameters. For a fair comparison, all comparison methods employ the same

TABLE VIII

CLASS-SPECIFIC AND OVERALL CLASSIFICATION ACCURACY (%) USING 5% LABELS FROM SD OF DIFFERENT METHODS FOR THE TARGET SCENE HOUSTON 2018

Class	Classification algorithms								
	SVM	CEGCN	DAN	DeepCoral	DAAN	MRAN	DSAN	HTCNN	TSTnet
1	99.78	8.28	55.95	48.71	61.57	56.39	57.95	4.85	85.03±4.34
2	65.30	42.49	72.18	64.81	76.94	75.57	67.90	71.57	68.73±4.90
3	25.70	75.27	62.87	61.28	66.67	68.00	71.69	35.75	52.82±4.38
4	100	40.91	100	50.00	72.73	63.64	81.82	53.64	100±0
5	73.26	58.95	56.33	52.42	52.76	66.50	61.79	54.40	61.71±5.88
6	69.12	71.74	74.11	65.42	69.94	68.54	70.26	90.80	84.44±3.37
7	43.63	61.84	31.80	47.43	54.23	54.36	54.53	44.05	53.07±5.08
OA (%)	64.67	65.14	66.05	61.26	66.59	66.95	67.08	74.72	**75.34±1.96**
KC (κ)	45.59	46.04	47.82	42.87	50.83	51.51	52.05	55.24	**59.69±2.56**

TABLE IX

CLASS-SPECIFIC AND OVERALL CLASSIFICATION ACCURACY (%) USING 5% LABELS FROM SD OF DIFFERENT METHODS FOR THE TARGET SCENE LOUKIA

Class	Classification algorithms								
	SVM	CEGCN	DAN	DeepCoral	DAAN	MRAN	DSAN	HTCNN	TSTnet
1	14.08	0	10.68	11.65	18.45	11.65	26.21	5.34	19.58±5.14
2	40.47	0	40.74	0	14.81	0	18.52	0	44.44±8.30
3	8.22	0	23.47	14.32	7.51	3.52	23.71	45.07	20.77±9.44
4	1.27	6.33	3.80	5.06	10.13	0	0	0	4.01±3.82
5	1.45	3.34	17.71	32.34	10.12	29.09	48.69	18.61	51.54±6.77
6	36.48	0	4.98	0	5.21	40.52	45.97	2.61	14.73±6.12
7	62.62	96.09	62.05	72.03	81.91	64.12	60.58	77.77	66.67±2.78
8	54.55	0	65.14	76.15	72.94	57.65	67.26	62.22	66.02±3.81
9	95.49	0	63.16	15.04	7.02	71.18	37.09	6.27	78.57±9.04
10	6.62	0	0	0	0.22	23.18	4.86	0	13.06±5.99
11	100	100	100	100	100	100	100	100	100±0
12	100	0	95.96	100	100	100	100	97.86	100±0
OA (%)	54.73	41.81	56.31	60.84	60.47	58.32	60.92	58.63	**63.31±3.05**
KC (κ)	45.20	21.99	45.94	50.93	49.71	49.14	52.44	47.47	**55.37±3.82**

architecture (for HTCNN, we use VGG16Net, and the others use ResNet50). All convolutional and pooling layers are fine-tuned from the ImageNet pretrained model, and the classifier layer is trained through backpropagation.

Spectral data of SD and TD are normalized by row with ℓ_2 norm for training and testing before network training. Tables VIII–X provide class-specific classification performance of aforementioned approaches on the three target scenes and report the average precision and standard deviation. In addition, the performance in the HyRANK dataset is poor since some classes are similar, such as the fourth class (Fruit Trees), the fifth class (Olive Groves), and the sixth class (Coniferous Forest), and the topological relationships of the land cover classes in the forest scene are obviously weaker than that in the urban scene. Furthermore, in order to demonstrate the interchangeability of SD and TD, experimental results with Housto2018 as SD and Houston 2013 as TD are listed in Table XI. It is obvious that TSTnet can still get the best cross-scene classification performance.

Classification maps are further shown in Figs. 10–12. In these maps, labeled pixels are displayed as ground truth and unlabeled pixels as backgrounds. It is clear that the proposed TSTnet obtains less noise and more accurate results in some areas of the classification maps, such as the first class (grass healthy) and sixth class (nonresidential buildings) in the Houston 2018 data. Compared with other methods, sixth class (nonresidential buildings) in the classification map of TSTnet is not classified as seventh class (road), which indicates that TSTnet has captured the adjacent topological relationship between them in the training process, so it is obviously better than the domain adaptive method considering only statistical features. In addition, the background pixels in the CEGCN classification map are also classified because it segments the entire HSI and then classifies HSI based on superpixels.

Fig. 13 illustrates the OA of all methods with different percentages of training samples. In the experiment, according to the total SD of the three datasets, the number of training samples in the Houston dataset varies from 10% to 100%, and the HyRANK and S-H datasets are set from 1% to 10%. For Houston and HyRank datasets, TSTNet is significantly better than other domain adaptive methods under different training sample conditions. However, TSTnet performed poorly when the training samples are 1% and 3%, mainly because the data noise is extremely serious and TSTnet cannot learn the domain invariant features describing the two domains in the case of a very small number of samples.

In addition, the computational complexity of each algorithm is evaluated. All the experiments are carried out using Pytorch on an Intel Core i7-9700 (16-GB RAM) powered with Nvidia GTX 1060 GPU with a 6-GB memory. The time of one epoch training is listed in Table XII, which contains one forward propagation and one backward propagation with a 100-sized batch, and the number of training samples in the SD is the same as in the previous experiments. As listed in Table XII,

TABLE X

CLASS-SPECIFIC AND OVERALL CLASSIFICATION ACCURACY (%) USING 5% LABELS FROM SD OF DIFFERENT METHODS FOR THE TARGET SCENE SHANGHAI

Class	Classification algorithms								
	SVM	CEGCN	DAN	DeepCoral	DAAN	MRAN	DSAN	HTCNN	TSTnet
1	91.54	81.58	89.48	93.56	92.11	85.36	91.90	89.24	93.47±0.71
2	33.01	33.89	78.41	77.59	77.73	82.06	84.83	37.14	84.58±3.74
3	100	99.73	99.22	99.35	99.32	99.62	95.45	100	96.96±4.65
OA (%)	67.73	64.73	86.82	87.86	87.42	87.13	89.60	68.78	**90.36±1.35**
KC (κ)	54.15	49.56	80.03	81.67	80.99	80.39	84.08	55.34	**85.27±2.02**

TABLE XI

CLASS-SPECIFIC AND OVERALL CLASSIFICATION ACCURACY (%) USING 5% LABELS FROM SOURCE SCENE HOUSTON 2018 FOR THE TARGET SCENE HOUSTON 2013

Class	Classification algorithms								
	SVM	CEGCN	DAN	DeepCoral	DAAN	MRAN	DSAN	HTCNN	TSTnet
1	91.01	92.17	86.96	59.13	87.83	84.35	94.20	91.01	89.39±8.30
2	37.26	64.38	95.62	86.85	97.53	96.71	94.79	95.89	92.38±3.18
3	67.12	37.26	81.10	96.16	98.08	96.44	86.58	85.21	85.70±3.30
4	95.79	2.11	12.63	0	12.63	23.16	49.47	12.63	83.02±7.25
5	68.34	73.67	63.95	67.71	68.03	59.87	73.04	67.71	67.52±2.84
6	71.81	98.77	100	100	99.75	100	99.75	98.77	96.27±2.96
7	100	96.84	96.61	63.21	97.97	95.49	98.42	100	97.70±4.38
OA (%)	75.97	69.64	79.88	70.20	83.44	82.37	87.11	81.94	**88.38±2.35**
KC (κ)	71.86	64.28	76.28	64.86	80.47	79.22	84.85	78.72	**86.36±2.77**

TABLE XII

EXECUTION TIME (IN SECONDS) OF ONE EPOCH TRAINING OF DIFFERENT METHODS

Data set	CEGCN	DAN	DeepCoral	DAAN	MRAN	DSAN	HTCNN	TSTN
Houston	2.11	22.79	29.59	22.84	29.94	21.98	21.81	25.15
HyRANK	1.64	18.82	24.06	18.64	23.71	17.76	17.98	19.73
Shanghai-Hangzhou	5.58	65.19	82.30	64.32	81.42	62.02	63.83	68.53

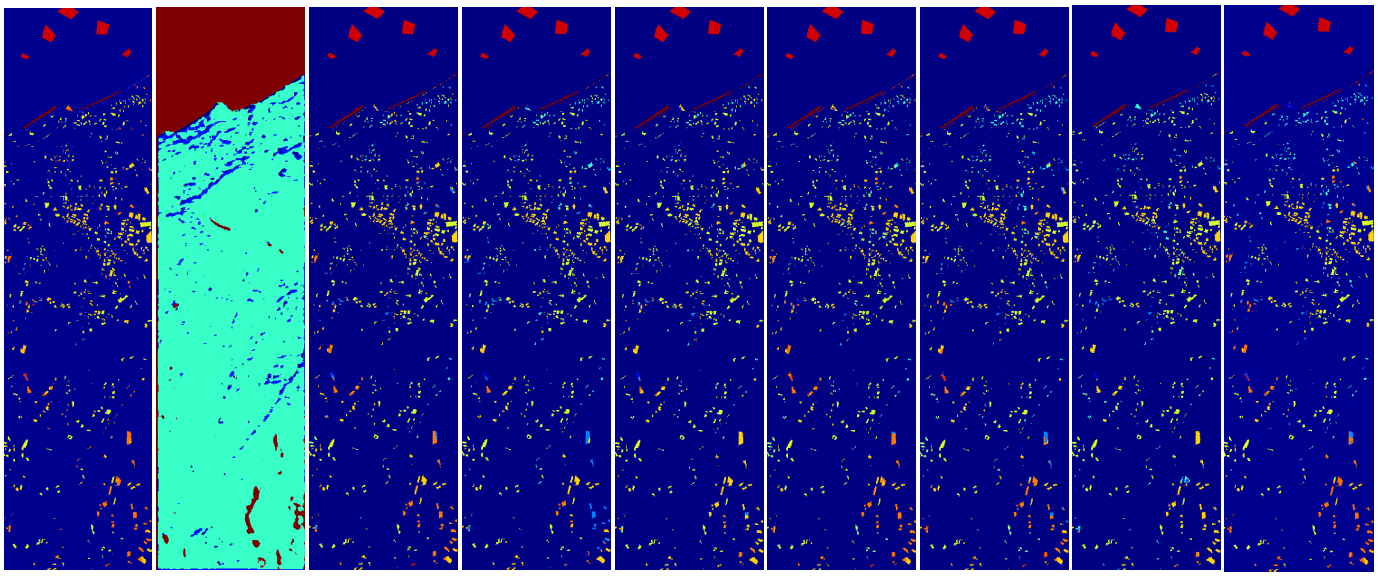

Fig. 11. Visualization and classification maps using 5% labels from SD for the target scene Loukia obtained with different methods, including (a) SVM(54.73%), (b) CEGCN [29] (41.81%), (c) DAN [22] (56.31%), (d) DeepCoral [33] (60.84%), (e) DAAN [34] (60.47%), (f) MRAN [35] (58.32%), (g) DSAN [23] (60.92%), (h) HTCNN [36] (58.63%), and (i) TSTnet (63.47%).

CEGCN takes the shortest time (the segmentation time is not included). This is because CEGCN takes the whole HSI as input, which allows the calculation to be performed in parallel. Except for the longest time-consuming DeepCoral and MRAN, the training time of TSTnet is similar to other methods. In TSTnet, the more time-consuming is the SACEConv layer,

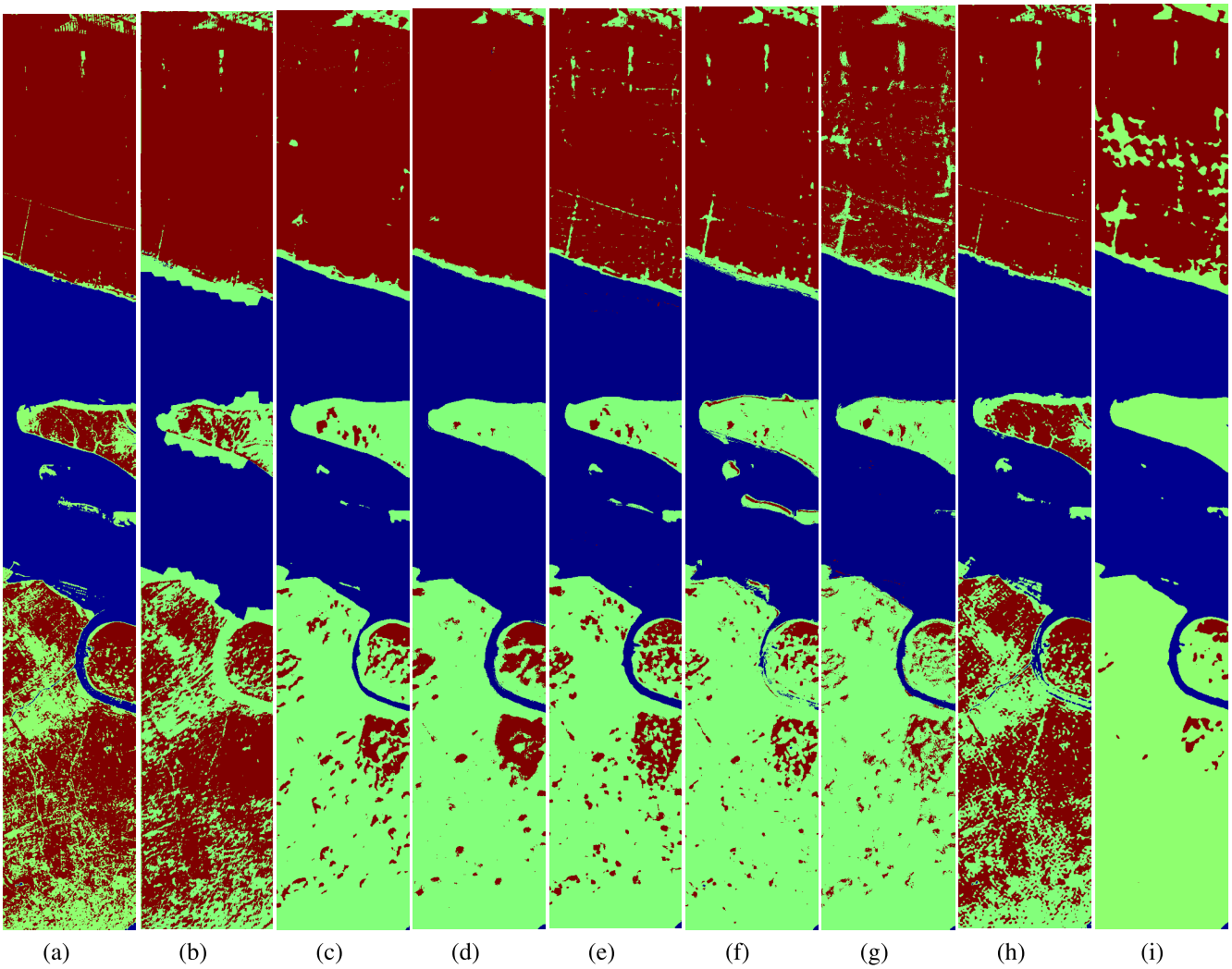

Fig. 12. Visualization and classification maps using 5% labels from SD for the target scene Shanghai obtained with different methods, including (a) SVM (67.73%), (b) CEGCN [29] (64.73%), (c) DAN [22] (86.82%), (d) DeepCoral [33] (87.86%), (e) DAAN [34] (87.42%), (f) MRAN [35] (87.13%), (g) DSAN [23] (89.60%), (h) HTCNN [36] (68.78%), and (i) TSTnet (91.03%).

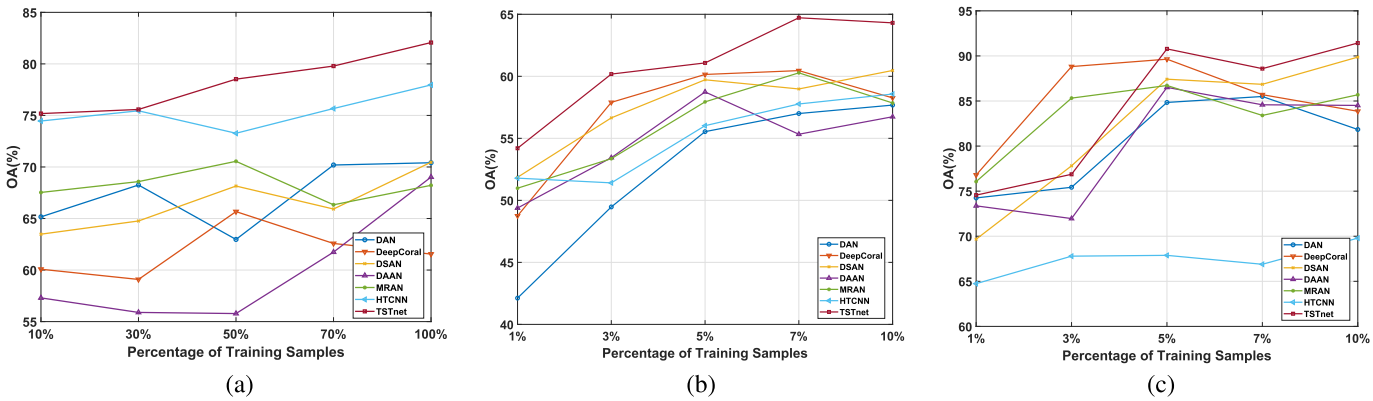

Fig. 13. Overall classification accuracy of all methods versus different percentages of training samples for the three experimental data. (a) Houston. (b) HyRANK. (c) S-H.

where the number of thousands of nodes makes it time-consuming to propagate and update the node information.

V. Conclusion

In this article, an efficient **TSTnet** was proposed for improving the domain adaptive methods and reducing the spectral shift in cross-scene HSI classification. Graph alignment and distribution alignment work cooperatively, in which GOT and MMD are employed. In addition, the GCN is applied to cross-scene HSI classification, where the subgraphs are dynamically constructed via deep features with abundant semantic information. Furthermore, a consistency constraint is designed

to achieve better collaboration between distribution alignment and topological relationship alignment. Comprehensive experiments on three cross-scene HSI datasets verified the effectiveness of the proposed TSTnet in reducing domain shift and significantly improving cross-scene HSI classification performance.

REFERENCES

[1] K. Li, G. Zhang, X. Li, and J. Xie, "Face recognition based on improved Retinex and sparse representation," *Procedia Eng.*, vol. 15, pp. 2010–2014, Dec. 2011.

[2] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 471–478.

[3] W. Li, Y. Zhang, N. Liu, Q. Du, and R. Tao, "Structure-aware collaborative representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7246–7261, Sep. 2019.

[4] Y. Zhang, W. Li, H.-C. Li, R. Tao, and Q. Du, "Discriminative marginalized least-squares regression for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3148–3161, May 2020.

[5] W.-Y. Ren, M. Tang, Y. Peng, and G.-H. Li, "Semi-supervised image classification via nonnegative least-squares regression," *Multimedia Syst.*, vol. 23, no. 6, pp. 725–738, Nov. 2017.

[6] H. F. Grahn and P. Geladi, "Detection, classification, and quantification in hyperspectral images using classical least squares models," *Techn. Appl. Hyperspectral Image Anal.*, vol. 27, pp. 181–202, Sep. 2007.

[7] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.

[8] M. Zhang, W. Li, Q. Du, L. Gao, and B. Zhang, "Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 100–111, Jan. 2020.

[9] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.

[10] X. Zhao *et al.*, "Joint classification of hyperspectral and LiDAR data using hierarchical random walk and deep CNN architecture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7355–7370, Oct. 2020.

[11] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 456–460, Feb. 2001.

[12] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[13] Z. Sun, C. Wang, H. Wang, and J. Li, "Learn multiple-kernel SVMs for domain adaptation in hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1224–1228, Sep. 2013.

[14] J. Xia, N. Yokoya, and A. Iwasaki, "Ensemble of transfer component analysis for domain adaptation in hyperspectral remote sensing image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 4762–4765.

[15] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semi-supervised transfer component analysis for domain adaptation in remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3550–3564, Jul. 2015.

[16] J. Blitzer, R. T. Mcdonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Sydney, NSW, Australia, Jul. 2007, pp. 22–23.

[17] Q. Gu, Z. Li, and J. Han, "Joint feature selection and subspace learning," in *Proc. 32nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1294–1299.

[18] J. Li, J. Zhao, and K. Lu, "Joint feature selection and structure preservation for domain adaptation," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 1697–1703.

[19] H. Sun, S. Liu, S. Zhou, and H. Zou, "Unsupervised cross-view semantic transfer for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 13–17, Jan. 2016.

[20] C. Luo and L. Ma, "Manifold regularized distribution adaptation for classification of remote sensing images," *IEEE Access*, vol. 6, pp. 4697–4708, 2018.

[21] T. Liu, X. Zhang, and Y. Gu, "Unsupervised cross-temporal classification of hyperspectral images with multiple geodesic flow kernel learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9688–9701, Dec. 2019.

[22] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 97–105.

[23] Y. Zhu *et al.*, "Deep subdomain adaptation network for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2021.

[24] Y. Qu, R. K. Baghbaderani, W. Li, L. Gao, Y. Zhang, and H. Qi, "Physically constrained transfer learning through shared abundance space for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Jan. 8, 2021, doi: 10.1109/TGRS.2020.3045790.

[25] J. Bruna, W. Zaremba, A. Szlam, and Y. Lecun, "Spectral networks and locally connected networks on graphs," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Apr. 2014, pp. 1–14.

[26] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2016, pp. 3844–3852.

[27] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2017, pp. 1025–1035.

[28] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[29] Q. Liu, L. Xiao, J. Yang, and Z. Wei, "CNN-enhanced graph convolutional network with pixel- and superpixel-level feature fusion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Nov. 24, 2020, doi: 10.1109/TGRS.2020.3037361.

[30] S. Wan, C. Gong, P. Zhong, B. Du, L. Zhang, and J. Yang, "Multiscale dynamic graph convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3162–3177, May 2019.

[31] L. Chen, Z. Gan, Y. Cheng, L. Li, L. Carin, and J. J. Liu, "Graph optimal transport for cross-domain alignment," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Jul. 2020, pp. 1542–1553.

[32] B. Schölkopf, J. Platt, and T. Hofmann, "A kernel method for the two-sample-problem," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 513–520.

[33] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, Jul. 2016, pp. 443–450.

[34] C. Yu, J. Wang, Y. Chen, and M. Huang, "Transfer learning with dynamic adversarial adaptation network," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Sep. 2019, pp. 778–786.

[35] Y. Zhu *et al.*, "Multi-representation adaptation network for cross-domain image classification," *Neural Netw.*, vol. 119, pp. 214–221, Nov. 2019.

[36] X. He, Y. Chen, and P. Ghamisi, "Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3246–3263, May 2019.

[37] C. Debes *et al.*, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.

[38] B. Le Saux, N. Yokoya, R. Hansch, and S. Prasad, "2018 IEEE GRSS data fusion contest: Multimodal land use classification [technical committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 1, pp. 52–54, Mar. 2018.

[39] K. Karantzalos, C. Karakizi, Z. Kandylakis, and G. Antoniou, "HyRANK hyperspectral satellite dataset I (version v001)," Int. Soc. Photogramm. Remote Sens., Tech. Rep., 2018, doi: 10.5281/zenodo.1222202.