DUAL GRAPH CROSS-DOMAIN FEW-SHOT LEARNING FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Yuxiang Zhang[1,2], *Wei Li*[1,2,*], *Mengmeng Zhang*[1,2], *Ran Tao*[1,2]

[1]School of Information and Electronics, Beijing Institute of Technology
[2] Beijing Key Lab of Fractional Signals and Systems

ABSTRACT

Most domain adaptation (DA) methods focus on the case where the source data (SD) and target data (TD) with the same classes are obtained by the same sensor in cross-scene hyperspectral image (HSI) classification tasks. However, the classification performance is significantly reduced when there are new classes in TD. In addition, domain alignment is carried out based on local spatial information in most methods, rarely taking into account the non-local spatial information (non-local relationships) with strong correspondence. A Dual Graph Cross-domain Few-shot Learning (DG-CFSL) framework is proposed, trying to make up for the above shortcomings by combining Few-shot Learning (FSL) with domain alignment. Both SD with all label samples and TD with a few label samples are implemented for FSL episodic training. Meanwhile, Intra-domain Distribution Extraction block (IDE-block) is designed to characterize and aggregate the intra-domain non-local relationships. Furthermore, feature- and distribution-level cross-domain graph alignments are used to mitigate the impact of domain shift on FSL. Experimental results on two public HSI data sets demonstrate the effectiveness of the proposed method.

Index Terms— Hyperspectral Image Classification, Cross-Scene, Few-shot Learning, Domain Adaption, Graph Neural Network (GNN).

1. INTRODUCTION

In practical remote sensing applications, cross-scene classification task is often encountered, in which training and testing samples are the source domain data (SD) with sufficient labels and target domain data (TD) with a small amount or even no labels, respectively. Most domain adaptation (DA) frameworks aim to adapt the same classes of knowledge learned from SD to TD [1, 2, 3, 4], Zhu et al. proposed the Deep Subdomain Adaption Network (DSAN) that defined the concept of sub-domains and used local MDD (LMMD) to align the relevant sub-domains respectively [5], and the adaptive performance decreases significantly when new classes are present in TD. In recent years, Few-shot Learning (FSL) has been widely concerned due to its effectiveness in identifying new unseen classes using a few training samples. As a supervised meta-learning approach, FSL adopts the episodic learning approach, which completes classification without changing the trained model in the testing stage [6, 7]. Chen et al. gave the classifier-baseline and meta-baseline, and explored a simple process, meta-learning over a whole classification pre-trained model on its evaluation metric [8]. Contrast learning was combined with FSL, and infoPacth loss was designed to eliminate inductive bias in SD [9]. FSL has been used in HSI classification to alleviate the requirement of producing larger training sets. Liu et al. proposed deep few-shot learning (DFSL) to classify hyperspectral image (HSI) with few labeled data [10]. Relation Network for HSI FSC (RN-FSC) selected a few TD samples as fine-tuning dataset to fine-tune the SD training model [11].

At present, most cross-domain HSI classification methods based on DA focus on the same sensor and classes. In addition, the current cross-scene classification work is based on local spatial information extracted by traditional convolution operator to characterize and align the data distribution. However, in general, there is no strong correspondence between local spatial relationships between two scenes. Therefore, it is necessary to consider non-local relationships between classes to represent the distribution of domain, and further design an effective non-local relationships (i.e., graph data) alignment strategy to alleviate the influence of domain shift.

In order to overcome the above two limitations, a Dual Graph Cross-domain Few-shot Learning (DG-CFSL) framework is proposed, which is based on the mechanism of combining DA and FSL. Firstly, the episodic learning pattern of FSL is implemented on SD and TD, which is to build a meta-task (i.e., support set and query set). The features of them are mapped to the same dimension through the mapping layer. And then the similarity between the support set and query set is learned using the metric function. Furthermore, the Intra-domain Distribution Extraction block (IDE-block) of SD and TD is designed, and Graph Optimal transport (GOT) is employed for the cross-domain alignment from two aspects of graph feature- and graph distribution-level.

This paper is supported by the Beijing Natural Science Foundation (Grant no. JQ20021), National Natural Science Foundation of China (Grant nos. 61922013, 61421001 and U1833203). (Email: liwei089@ieee.org)

2. PROPOSED APPROACH

The flowchart of the proposed DG-CFSL is shown in Fig. 1, which contains three parts, i.e., feature extraction, few-shot learning and domain alignment. Domain alignment is composed of IDE-block and GOT, in which the domain distribution characteristics are obtained and domain shift is reduced.

Assuming that $\mathbf{X}_s \in \mathbb{R}^{d_s}$ with C_s classes and $\mathbf{X}_t \in \mathbb{R}^{d_t}$ with C_t classes are data from SD and TD, respectively. Here, d_s, d_t denote the dimension. There are two cross-domain FSL tasks: SD FSL task $\mathcal{T}_s$ and TD FSL task $\mathcal{T}_t$. All data in SD FSL task is labeled, and TD FSL task is separated into TD training data $\mathcal{T}_t^{tr}$ with a few labeled samples and TD testing data $\mathcal{T}_t^{te}$ with unlabeled samples.

2.1. Feature Extraction

Due to different sensors of SD and TD, the spectral resolutions of the samples are inconsistent. Therefore, the mapping layers, $\mathbf{M}_s(\bullet)$ for SD and $\mathbf{M}_t(\bullet)$ for TD, are used to map two domains to the same dimension d_{map}. The deep residual 3D CNN network $f_{emb}(\bullet)$ is used to extract the d_{emb}-dimensional features (d_{map} and d_{emb} are set to 100 and 128 in the experiment respectively).

2.2. Source and Target Few-shot Learning

FSL of SD and TD are performed simultaneously in each iteration. Taking SD as an example, the embedded features $f_{emb}(\mathbf{M}_s(\mathbf{X}_s))$ are divided into support set and query set $\mathcal{Q}_s$, where the support set serves as the labeled training set. In each episode, FSL is performed by calculating Euclidean distance between the embedded features of query set and the prototype of each class , and the prediction loss of the query set is minimized. The predicted probability output of a query sample $\mathbf{x}_q$ is obtained by the calculated Euclidean distance through softmax function,

$$P(\hat{y}_q|\mathbf{x}_q) = Softmax\left(-ED\left(\mathbf{x}_q, \mathbf{x}_s^{mean(c)}\right)\right) \quad (1)$$

where $ED(\bullet)$ denotes an Euclidean distance function, $\mathbf{x}_s^{mean(c)}$ is the mean value of c-th class support set, also known as the c-th class prototype ($c \in C_s$). The FSL loss of a query sample in SD is calculated by cross entropy loss,

$$\mathcal{L}_{CE}^s(P(\hat{y}_q|\mathbf{x}_q), y_q) = -y_q \log P(\hat{y}_q|\mathbf{x}_q). \quad (2)$$

The calculation process for $\mathcal{L}_{CE}^t$ is similar to that above, but note that L_{CE}^t is obtained based on the meta-task constructed using $\mathcal{T}_t^{tr}$ after data augment.

2.3. Domain Alignment with IDE-block

Given the effect of domain shift on classification performance in the procedure of FSL episodic training, domain alignment is one of the effective measures. However, most methods are based on local spatial information for statistical feature alignment, ignoring the importance of non-local spatial relationships between samples. Therefore, for each meta-task (C-way K-shot), the similarity between support embedded features and query embedded features is used to construct the feature graph (FG) of SD and TD, respectively. $\mathbf{G}_f^{s/t} = (\mathbf{V}_f^{s/t}, \mathbf{E}_f^{s/t}) = \{(\mathbf{v}_{f(i)}^{s/t}, \mathbf{e}_{f(i,j)}^{s/t})\}_{i/j=1}^{C\times K+T}$ (s/t represents SD or TD, $C \times K$ and T represent the number of support samples and query samples, respectively). In order to capture the distribution-level characteristics from the FG, an IDE-block is designed to generate the distribution graph (DG), $\mathbf{G}_d^{s/t} = (\mathbf{V}_d^{s/t}, \mathbf{E}_d^{s/t}) = \{(\mathbf{v}_{d(i)}^{s/t}, \mathbf{e}_{d(i,j)}^{s/t})\}_{i/j=1}^{C\times K+T}$.

Feature Graph: The embedded features are separated into support set $\mathcal{S}_s$ and query set $\mathcal{Q}_s$ as the initialization of node features in FG,

$$\mathbf{v}_{f(i)}^{s0} = f_{emb}(\mathbf{M}_s(\mathbf{x}_i^s)) \quad (3)$$

where the node $\mathbf{v}_{f(i)}^{s0} \in \mathbb{R}^{d_{emb}}$ is the feature-level information. The edge in FG is constructed by calculating the similarity between two nodes,

$$\mathbf{e}_{f(i,j)}^{s0} = f_{FG}\left(\left(\mathbf{v}_{f(i)}^{s0} - \mathbf{v}_{f(j)}^{s0}\right)^2\right) \quad (4)$$

where $f_{FG}(\bullet)$ consists of two Conv2d-BN-ReLU blocks to enhance the nonlinear representation of non-local relationships.

Feature-to-Distribution Aggregation: In order to capture the intra-domain distribution-level information, the similarity between nodes on the edges of FG is aggregated as the node features of DG,

$$\mathbf{v}_{d(i)}^{s1} \leftarrow MLP\left(\left[\mathbf{e}_{f(i,j)}^{s0}, \mathbf{v}_{d(i)}^{s0}\right]_{j=1}^{C\times K}\right) \quad (5)$$

where MLP consists of a fully connected layer and an ReLU, which is used to aggregate the FG edge features $\mathbf{e}_{f(i,j)}^{s0}$ and the initialization of DG node features $\mathbf{v}_{d(i)}^{s0}$ to obtain the new DG node features $\mathbf{v}_{d(i)}^{s1}$, the value of j-th in $\mathbf{v}_{d(i)}^{s1} \in \mathbb{R}^{C\times K}$ represents the non-local relationship between node i and j, $[\bullet,\bullet]$ is the concatenation operator, and $\mathbf{v}_{d(i)}^{s0}$ is the initialization of DG node features,

$$\mathbf{v}_{d(i)}^{s0} = \begin{cases} \left[\delta\left(y_i^s, y_j^s\right)\right]_{j=1}^{C\times K} & \mathbf{x}_i^s \in \mathcal{S}_s \\ \left[\frac{1}{CK}, \cdots, \frac{1}{CK}\right] & \mathbf{x}_i^s \in \mathcal{Q}_s \end{cases} \quad (6)$$

where $\delta(\bullet,\bullet)$ denotes the Kronecker delta function which outputs one when the labels of two support samples are consistent (i.e., y_i^s=y_j^s) and zero when $y_i^s \neq y_j^s$.

Distribution Graph: The similarity between non-local relationships is acquired by calculating the DG node features $\mathbf{v}_{d(i)}^{s1}$,

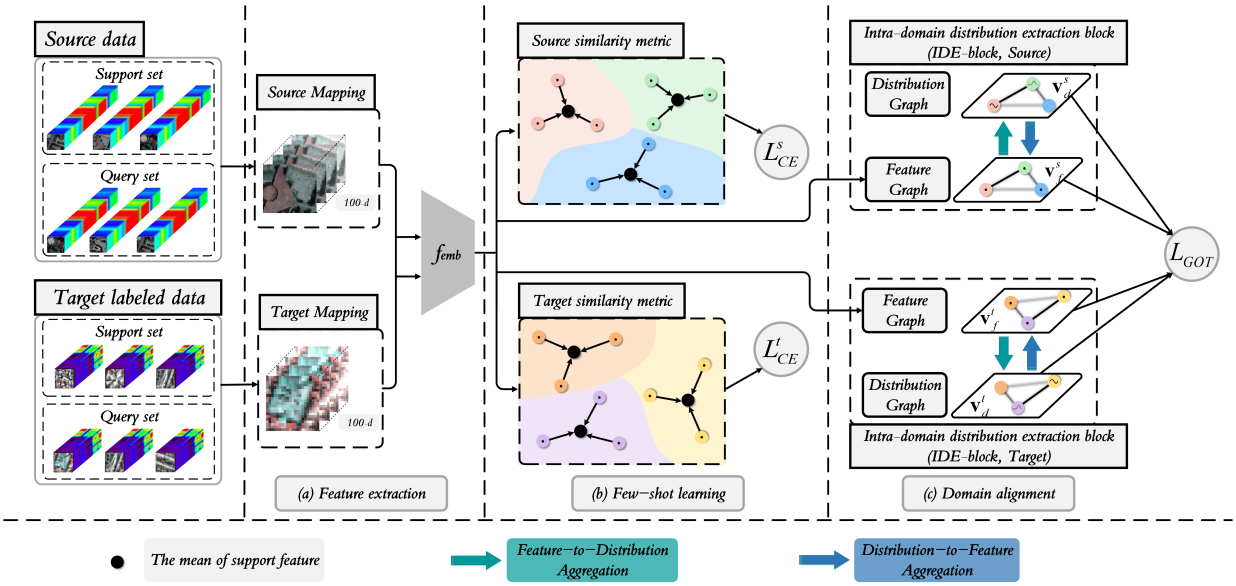

Fig. 1. Flowchart of the proposed DG-CFSL, including feature extraction, few-shot learning and domain alignment.

$$\mathbf{e}^{s1}_{d(i,j)} = f_{DG}\left(\left(\mathbf{v}^{s1}_{d(i)} - \mathbf{v}^{s1}_{d(j)}\right)^2\right) \tag{7}$$

where $f_{DG}(\bullet)$ is two Conv2d-BN-ReLU blocks. In the produced DG, the high-order distribution information is accessed based on the non-local relationships among samples in a meta-task.

Distribution-to-Feature Aggregation: The FG node features are updated by aggregating the DG edge features $\mathbf{e}^{s1}_{d(i,j)}$ and the initialization of FG node features $\mathbf{v}^{s0}_{f(i)}$,

$$\mathbf{v}^{s1}_{f(i)} \leftarrow MLP\left(\left[\left(\sum_{j=1}^{C\times K+T} \mathbf{e}^{s1}_{d(i,j)}\mathbf{v}^{s0}_{p(j)}\right), \mathbf{v}^{s0}_{f(i)}\right]_{i=1}^{C\times K+T}\right) \tag{8}$$

where MLP is two Conv2d-BN-ReLU blocks. IDE-block corresponding to TD is the same as the above process.

Cross-domain graph alignment: To reduce domain shift by domain alignment, Wasserstein distance (WD) for node matching, and Gromov-Wasserstein distance (GWD) for edge matching are calculated by $\mathbf{v}^{s1}_{f(i)}$ and $\mathbf{v}^{s1}_{d(i)}$. In order to achieve the best integration of WD and GWD, a transformation matrix shared by WD and GWD is considered [12]. GOT distance is defined as,

$$\begin{aligned} GOT\left(\mathbf{G}^s_{f/d}, \mathbf{G}^t_{f/d}\right) = \sum_{i,\hat{i},j,\hat{j}} \Big(&\mathbf{T}_{ij}\cdot c\left(\mathbf{v}^s_i, \mathbf{v}^t_j\right) \\ &+ \mathbf{T}_{\hat{i}\hat{j}}\cdot L\left(\mathbf{v}^s_i, \mathbf{v}^t_j, \hat{\mathbf{v}}^s_i, \hat{\mathbf{v}}^t_j\right)\Big) \end{aligned} \tag{9}$$

where $\mathbf{T}_{ij}$ and $\mathbf{T}_{\hat{i}\hat{j}}$ are the transformation matrix of two pairs of nodes, $c(\mathbf{x}^s, \mathbf{x}^t)$ is the cross-domain cost matrix obtained by the cosine distance, $L\left(\mathbf{v}^s_i, \mathbf{v}^t_j, \hat{\mathbf{v}}^s_i, \hat{\mathbf{v}}^t_j\right)$ is the cost function evaluating the intra-graph structural similarity between two pairs of nodes $(\mathbf{v}^s_i, \hat{\mathbf{v}}^s_i)$ and $\left(\mathbf{v}^t_j, \hat{\mathbf{v}}^t_j\right)$. Thus, GOT loss for cross-domain graph alignment is formulated as,

$$\mathcal{L}_{GOT} = GOT\left(\mathbf{G}^s_f, \mathbf{G}^s_f\right) + GOT\left(\mathbf{G}^s_d, \mathbf{G}^s_d\right). \tag{10}$$

Then the total loss of DG-CFSL is defined as follows,

$$\mathcal{L}_{total} = \mathcal{L}^s_{CE} + \mathcal{L}^t_{CE} + \lambda_1\mathcal{L}_{GOT} \tag{11}$$

where λ_1 is the regularization parameter. In the proposed DG-CFSL framework, $\mathcal{T}^{tr}_t$ is regarded as the support set and $\mathcal{T}^{te}_t$ as the query set in the testing stage, and nearest neighbor classifier is applied to predict labels for the query features of TD output by the embedded feature extractor.

3. EXPERIMENTAL RESULTS

In this section, the Chikusei dataset is used as source scene, and the Pavia University and Salinas datasets are used as target scene. Several state-of-the-art deep DA and FSL-based algorithms are employed for comparison algorithms, 3DCNN [13], Deep Subdomain Adaption Network (DSAN) [5], Dynamic Adversarial Adaptation Network (DAAN) [14], Distribution Propagation Graph Network (DPGN) [15], DFSL [10], and RN-FSC [11].

3.1. Experiment Data

Source scene: The Chikusei data has 128 bands in the spectral range from 363 nm to 1018 nm, with a size of 2517×2335 and a spatial resolution of 2.5m. There are 19 classes, including urban and rural areas.

Target scene: (1) The Pavia University data has 103 spectral bands in the spectral range from 430 nm to 860nm. The size of image is 610×340 pixels with 9 classes, the spatial resolution is 1.3m. (2) The Salinas data has 200 bands, with 512×217 pixels and a spatial resolution of 3.7 m. The ground truth contains 16 classes.

Table 1. Class-specific and overall classification accuracy (%) for the Pavia University data (5 labeled samples from TD).

Class	Samples	Classification algorithms						
		3DCNN [13]	DAAN [14]	DSAN [5]	DPGN [15]	DFSL [10]	RN-FSC [11]	DG-CFSL
Asphalt	6631	51.31	93.38	91.15	84.09	69.05	69.42	78.93 ±8.97
Meadows	18549	70.69	95.10	94.10	62.90	85.45	92.70	88.23 ±6.58
Gravel	2099	36.10	43.26	29.16	0.05	56.83	49.95	64.90±8.59
Trees	3064	68.03	92.30	96.90	89.10	89.52	92.09	91.98±3.66
Sheets	1345	92.91	99.63	99.63	99.7	99.29	98.45	98.18±2.58
Bare soil	5029	43.67	54.32	27.36	56.33	70.50	57.66	78.35±9.48
Bitumen	1330	69.74	11.28	58.35	65.94	71.36	69.36	80.48±8.40
Bricks	3682	57.52	0.46	54.02	66.54	58.29	64.07	72.08±9.92
Shadow	947	97.77	88.38	94.72	0.00	97.02	98.81	96.20±5.00
OA(%)		62.57	76.54	78.43	64.42	78.23	79.84	**83.61±2.82**
KC		0.53	0.68	0.70	0.55	0.70	0.72	**0.79±0.03**

Table 2. Class-specific and overall classification accuracy (%) for the Salinas data (5 labeled samples from TD).

Class	Samples	Classification algorithms						
		3DCNN [13]	DAAN [14]	DSAN [5]	DPGN [15]	DFSL [10]	RN-FSC [11]	DG-CFSL
Brocoli green weeds 1	2009	98.95	99.75	92.83	87.72	94.86	96.45	98.87±1.77
Brocoli green weeds 2	3726	97.02	96.16	99.49	99.49	98.68	99.15	98.89±1.55
Fallow	1976	83.55	50.86	61.13	79.76	93.65	85.85	95.72±4.17
Fallow rough plow	1394	92.18	99.71	98.71	98.34	99.39	98.49	98.57±1.92
Fallow smooth	2678	83.38	86.56	99.22	80.13	90.6	82.67	89.60±8.56
Stubble	3959	89.67	99.92	99.97	99.92	97.64	97.29	99.08±0.98
Celery	3579	99.16	96.51	97.29	99.86	99.33	99.39	98.67±1.49
Grape untrained	11271	68.15	91.79	89.22	50.84	78.34	71.59	77.48±9.18
Soil vinyard develop	6203	98.23	83.15	85.44	89.03	90.19	88.16	99.04±1.22
Corn senesced green weeds	3278	33.13	70.99	78.92	81.24	61.84	69.72	75.94±8.46
Lettuce romaine 4wk	1063	75.47	46.25	99.34	89.46	96.45	89.29	95.91±3.49
Lettuce romaine 5wk	1927	94.29	95.49	71.35	99.17	93.11	94.03	98.89±1.83
Lettuce romaine 6wk	916	98.47	99.34	99.67	99.56	99.50	99.45	98.63±1.05
Lettuce romaine 7wk	1070	92.90	98.60	72.52	98.87	97.53	96.58	97.27±8.77
Vinyard untrained	7268	62.29	57.18	62.36	59.75	77.46	69.30	72.77±8.20
Vinyard vertical trellis	1807	97.45	83.40	74.60	77.69	85.98	81.86	91.88±6.41
OA(%)		80.43	84.06	85.35	78.67	87.67	83.84	**88.63±2.30**
KC		0.78	0.82	0.84	0.76	0.86	0.82	**0.87±0.03**

3.2. Experimental Setting and Performance

The input of DG-CFSL is set as patch size of 9×9. Adaptive Moment Estimation(Adam) is used as the optimization scheme. The learning rate is set to 1e-2, and the regularization parameter λ_1 is 1. The number of training iterations is 10000. For each meta-task of C-way K-shot in episodic training, C is set as the number of classes of TD, 9 for the Pavia University and 16 for the Salinas. K is regarded as that K labeled samples are randomly selected to form a support set $\mathcal{S}$ in each episode, and K for SD FSL and TD FSL is set to 1 in all experiments. In addition, the number of the query samples in $\mathcal{Q}$ is set to 19 to evaluate the learned classifier. Note that C classes are randomly extracted from SD when constructing a meta-task. Furthermore, 5 labeled samples are selected from each class of TD for TD FSL, and the data is augmented by adding gaussian random noise.

Tables 1-2 report class-specific accuracy (CA), the overall accuracy (OA) and the Kappa coefficient (KC) of the above methods in three TDs. For all the classification methods, 5 labeled samples are randomly selected for each class as supervised samples in the TD. Compared with 3DCNN [13], which only uses a few TD samples training, DSAN [5] and RN-FSC [11] and DG-CFSL, which ustilize SD knowledge and design domain alignment blocks, provide a 5% to 10% improvement in OA. It suggests that when there are a few samples in TD, transfering SD knowledge to TD helps improve the classification accuracy of the model on TD. The performance of FSL-based methods, DFSL [10] and RN-FSC [11], is improved by 4% to 8% over unsupervised DA methods, because the learning method of constructing meta-tasks in FSL for episodic training has stronger generalization ability and learns more fully SD knowledge for TD. This demonstrates that FSL method trained by meta-learning idea can better solve the problem of few labeled samples. Due to the design of domain alignment block, DG-CFSL perform better than DPGN, DFSL and RN-FSC that only focus on FSL. In particular, the spectral shift of some classes is effectively reduced to improve the classification performance, such as the 6-th class (Bare soil) 7-th class (Bitumen) and 8-th class (Bricks) in Pavia University data.

4. CONCLUSIONS

In this paper, Dual Graph Cross-domain Few-shot Learning (DG-CFSL) has been proposed to solve the issues of cross-scene classification, which performs both FSL and domain alignment under the condition that all labeled samples in SD is available and few labeled samples are in TD. Specifically, the episodic learning pattern of FSL is used to learn the class separability metric space of SD and TD respectively, and DA is considered to reduce the negative impact of domain shift on FSL. An Intra-domain Distribution Extraction Block (IDE-block) is designed to characterize and aggregate intra-domain non-local relationships, and graph alignment is performed from feature- and distribution-level. The experimental results show that the proposed method has presented significant improvements over the state-of-the-art models.

5. REFERENCES

[1] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan, "Learning transferable features with deep adaptation networks," in *International conference on machine learning*. PMLR, 2015, pp. 97–105.

[2] Xin Jin, Cuiling Lan, Wenjun Zeng, and Zhibo Chen, "Style normalization and restitution for domaingeneralization and adaptation," *arXiv preprint arXiv:2101.00588*, 2021.

[3] Muhammad Awais, Fengwei Zhou, Hang Xu, Lanqing Hong, Ping Luo, Sung-Ho Bae, and Zhenguo Li, "Adversarial robustness for unsupervised domain adaptation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 8568–8577.

[4] Yuxiang Zhang, Wei Li, Mengmeng Zhang, Ying Qu, Ran Tao, and Hairong Qi, "Topological structure and semantic information transfer network for cross-scene hyperspectral image classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[5] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Guolin Ke, Jingwu Chen, Jiang Bian, Hui Xiong, and Qing He, "Deep subdomain adaptation network for image classification," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 4, pp. 1713–1722, 2020.

[6] Jake Snell, Kevin Swersky, and Richard Zemel, "Prototypical networks for few-shot learning," in *NIPS*, 03 2017, p. 4077–4087.

[7] Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Y. Bengio, "Manifold mixup: Better representations by interpolating hidden states," in *ICML*, 05 2019, pp. 6438–6447.

[8] Yinbo Chen, Zhuang Liu, Huijuan Xu, Trevor Darrell, and Xiaolong Wang, "Meta-baseline: Exploring simple meta-learning for few-shot learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9062–9071.

[9] Chen Liu, Yanwei Fu, Chengming Xu, Siqian Yang, Jilin Li, Chengjie Wang, and Li Zhang, "Learning a few-shot embedding model with contrastive learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, pp. 8635–8643.

[10] L. Bing, X. Yu, A. Yu, P. Zhang, and R. Wang, "Deep few-shot learning for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2290–2304, 2018.

[11] K. Gao, B. Liu, X. Yu, J. Qin, and X. Tan, "Deep relation network for hyperspectral image few-shot classification," *Remote Sensing*, vol. 12, no. 6, pp. 923, 2020.

[12] Liqun Chen, Zhe Gan, Yu Cheng, Linjie Li, Lawrence Carin, and JJ (Jingjing) Liu, "Graph optimal transport for cross-domain alignment," in *International Conference on Machine Learning (ICML 2020)*, July 2020.

[13] Ying Li, Haokui Zhang, and Qiang Shen, "Spectral–spatial classification of hyperspectral imagery with 3d convolutional neural network," *Remote Sensing*, vol. 9, pp. 67, 01 2017.

[14] Chaohui Yu, Jindong Wang, Yiqiang Chen, and Meiyu Huang, "Transfer learning with dynamic adversarial adaptation network," in *2019 IEEE International Conference on Data Mining (ICDM)*, 2019, pp. 778–786.

[15] Ling Yang, Liangliang Li, Zilun Zhang, Xinyu Zhou, Erjin Zhou, and Yu Liu, "Dpgn: Distribution propagation graph network for few-shot learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13390–13399.