SCIENCE CHINA
Information Sciences

• LETTER •

October 2025, Vol. 68, Iss. 10, 209102:1–209102:2
https://doi.org/10.1007/s11432-024-4586-y

Progressive joint distribution alignment network for cross-scene hyperspectral image classification

Zhuojun XIE[1,2], Puhong DUAN[3], Xudong KANG[3*], Wang LIU[2] & Shutao LI[2*]

[1]*Hunan Vanguard Group Corporation Limited, Changsha 410100, China*
[2]*College of Electrical and Information Engineering, Hunan University, Changsha 410082, China*
[3]*School of Robotics, Hunan University, Changsha 410082, China*

Received 25 April 2024/Revised 27 July 2024/Accepted 13 June 2025/Published online 26 September 2025

Citation Xie Z J, Duan P H, Kang X D, et al. Progressive joint distribution alignment network for cross-scene hyperspectral image classification. Sci China Inf Sci, 2025, 68(10): 209102, https://doi.org/10.1007/s11432-024-4586-y

Unsupervised cross-scene hyperspectral image classification (HSIC) aims to apply knowledge acquired from a source scene to recognize the similar target scenes, even when there are no labeled samples available for the new scene.

Nowadays, numerous studies have been made for cross-scene hyperspectral image classification. For example, Yang et al. [1] extended the manifold alignment methods based on the assumption of local geometric similarity across scenes, achieving simultaneous alignment of spatial and spectral features. Ye et al. [2] employed dictionary learning to project cross-scene spectral features into a shared latent space, aligning spectral distributions and mitigating the influence of domain shift on classification performance. Although these traditional methods mentioned above can ensure the consistency of feature distributions across scenes by learning linear or nonlinear transformations, the use of handcraft features limits their ability to capture impressive features from diverse data. Because of the superior performance of the deep learning model in feature extraction, it has been broadly applied in domain adaptation. Zhang et al. [3] designed a novel discriminative cooperative alignment method to decrease the geometric and statistical shift, which utilized maximum mean discrepancy to align the global marginal distribution and the subspace alignment. Tang et al. [4] introduced the concept of adversarial learning and designed an HSI cross-scene classification framework that achieves both domain-level and category-level feature alignment.

Although these methods improve the classification performance of hyperspectral images (HSIs), there are still several issues for cross-scene HSIC. (1) Existing feature alignment methods heavily depend on high-quality pseudo-labels. However, during the initial training phase, the generated pseudo-labels tend to be noisy due to domain shift. These low-quality pseudo-labels can lead to negative transfer, thereby reducing classification accuracy. (2) The hyperspectral image is susceptible to various factors like lighting, atmospheric conditions, and imaging sensors, leading to spectral differences of the same land cover. Therefore, single domain-level or category-level alignment while ignoring sample-level differences leads to the unsatisfied classification results in the target scene.

To tackle these issues, a novel unsupervised domain adaptation (UDA) framework named PJANet is proposed, which can efficiently narrow the distribution gap among various scenes. Specifically, a convolutional neural network (CNN) is applied to extract important features from different scenes at first, in which a classic spectral-spatial attention (SS-att) mechanism is applied to obtain discriminative features. Then, since the spectral shift in the early stages of training may lead to a decrease in model fitting accuracy, a novel progressive joint distribution alignment (PJDA) strategy is designed to achieve finer distribution alignment of cross scenes. The DeepCoral method is employed to match the feature distribution across different scenes in the early stages of training, which can reduce domain distribution differences and enhance the consistency of representations for similar land covers across different scenes. In addition, we utilize the feature consistency concept to obtain pseudo-labels for the target scene by calculating feature similarity across scenes. After an initial phase of feature alignment, we utilized the pseudo-labels from the target scene and introduced contrastive learning to develop a cross-sample alignment (CSA) module. This module efficiently minimizes the distances between similar samples from different scenes while increasing the distances between dissimilar samples. As a result, it further mitigates domain shift, resulting in superior cross-scene HSIC performance. Experiments on three widely used datasets confirm the efficacy of the proposed PJANet.

Proposed PJANet for cross-scene hyperspectral image classification. Figure 1 presents the flowchart of our method, including feature extraction and progressive joint distribution alignment. Specifically, we select two HSI patches $\mathbf{I}_s \in R^{p\times p\times B}$ and $\mathbf{I}_t \in R^{p\times p\times B}$ from labeled source scene D_s and unlabeled target scene D_t, respectively, where p is the input image size and B denotes the amount of spectral bands. They are fed into the feature extractor at the same time. In order to further enhance characterization capabilities of the feature extractor, an SS-att mechanism

---

* Corresponding author (email: xudong_kang@163.com, shutao_li@hnu.edu.cn)

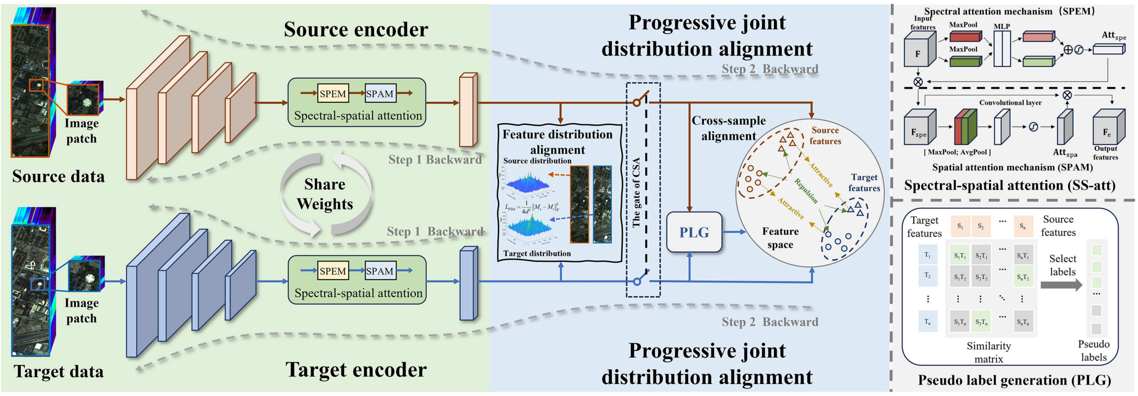

Figure 1 (Color online) Flowchart of the proposed progressive joint alignment network. CSA is short for cross-sample alignment. During the initial training stage, the feature distribution alignment (FDA) is used to minimize the disparity across scenes. After performing several epochs, the FDA and the CSA are simultaneously utilized to help the model achieve finer alignment.

is adopted in the feature extraction process. Furthermore, a novel progressive joint distribution alignment strategy is proposed to narrow the distribution gap across scenes, which mainly contains three steps, i.e., FDA, CSA, and progressive joint distribution alignment. First, the FDA is applied to match the feature distribution across scenes. Then, according to the feature similarity across scenes, we can estimate pseudo-labels in the target scene. After optimizing the model with FDA for a period of time, the CSA is employed to align the sample distribution. Finally, the FDA and CSA are jointly used to simultaneously achieve domain-level and sample-level alignment.

Step 1. In order to learn the latent representation of cross-scene hyperspectral images, a deep CNN, i.e., Resnet-50, is utilized to obtain discriminative features, where an efficient attention mechanism is employed to enhance representation ability and suppress the unimportant information. Then, a linear layer with a rectified linear activation unit (ReLU) is used to enhance the non-linear representation of the extracted features. Specifically, the Resnet-50 is applied as the feature extractor f, i.e., $\mathbf{F}_s = f(\mathbf{I}_s)$ and $\mathbf{F}_t = f(\mathbf{I}_t)$. The detailed steps of the spectral-spatial feature extraction are presented in Appendix B.

Step 2. Due to the presence of spectral shift phenomena in cross-scenes, the training model with source data cannot obtain satisfactory classification results on the target scene. In this work, the progressive joint distribution alignment strategy is proposed to capture the domain-invariant information. Specifically, in the preliminary stage of the training process, the FDA method is utilized to match the feature distribution of different scenes. After several epochs of training, both the FDA and CSA are employed concurrently to align the sample distribution and feature distribution across scenes. In this scenario, the PJDA strategy achieves domain-level and sample-level alignment simultaneously, thereby further enhancing classification performance across scenes. The detailed steps can be found in Appendix C.

Feature extraction loss: During the feature extraction process, a widely used classification loss, i.e., cross entropy loss, is employed to update weights of the proposed network. The feature extraction loss can be expressed in (B5).

FDA loss: The DeepCoral loss is introduced to align the feature distribution across scenes. Specifically, the covariance matrices of cross-domain features $\mathbf{M}_s$ and $\mathbf{M}_t$ are first calculated. Then, the DeepCoral loss L_{FDA} is expressed as $L_{\mathrm{FDA}} = \frac{1}{4d^2}\left\|\mathbf{M}_s - \mathbf{M}_t\right\|_F^2$.

CSA loss: To achieve finer sample-level alignment, a contrastive loss is introduced to learn significant feature representation and enhance feature discrimination. Therefore, for each $\mathbf{I}_s^i \in Q_s$, we can construct a set of positive sample pairs $Q_t^{i+} = \left\{\mathbf{I}_t^j \in Q_t \middle| M_{ij} = 1\right\}$ and negative sample pairs $Q_t^{i-} = \left\{\mathbf{I}_t^j \in Q_t \middle| M_{ij} = -1\right\}$, where Q_s and Q_t are the source and target batches, respectively.

Experimental results on several hyperspectral datasets (shown in Appendix E) reveal that our method can exceed the cutting-edge classification techniques with regard to visual and objective results.

Acknowledgements This work was supported by National Key R&D Program of China (Grant No. 2021YFA0715203), National Natural Science Foundation of China (Grant Nos. 62201207, 62371185), Natural Science Foundation of Hunan Province (Grant Nos. 2021JJ022, 2023JJ40163), and Science and Technology Innovation Program of Hunan Province (Grant No. 2023RC3124).

Supporting information Appendixes A–F. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Yang H L, Crawford M M. Spectral and spatial proximity-based manifold alignment for multitemporal hyperspectral image classification. IEEE Trans Geosci Remote Sens, 2016, 54: 51–64
2 Ye M, Qian Y, Zhou J, et al. Dictionary learning-based feature-level domain adaptation for cross-scene hyperspectral image classification. IEEE Trans Geosci Remote Sens, 2017, 55: 1544–1562
3 Zhang Y, Li W, Tao R, et al. Cross-scene hyperspectral image classification with discriminative cooperative alignment. IEEE Trans Geosci Remote Sens, 2021, 59: 9646–9660
4 Tang X, Li C, Peng Y. Unsupervised joint adversarial domain adaptation for cross-scene hyperspectral image classification. IEEE Trans Geosci Remote Sens, 2022, 60: 5536415