Deep Learning-Based Cross-Sensor Domain Adaptation Under Active Learning for Land Cover Classification

Indrajit Kalita, *Graduate Student Member, IEEE*, Runku Nikhil Sai Kumar, and Moumita Roy, *Member, IEEE*

***Abstract*— Cross-sensor remote-sensing data have a significant impact on degrading the performance of traditional land cover classification (LCC) models. This occurs due to different probability distributions of the collected data from different satellites (having diverse image resolutions and different geographical locations). To resolve this, a cross-sensor domain adaptation (DA) strategy is investigated by considering two source → target scenarios using hyperspectral and aerial image datasets. At the onset, a feature extraction (FE) along with a "stacking of sample" (whenever required) strategy is proposed to balance the cross-sensor data in terms of feature dimensions and the available number of samples. Thereafter, a standard deviation (SD)-based active learning (AL) technique is investigated by exploiting the labeled source images to get the "most-informative" target samples. Finally, the labeled source and "most-informative" target samples are merged to train a classifier which is then used to predict the land cover classes under a multi-sensor framework. Experimental results are found to be promising for the proposed scheme to handle the DA problem under a cross-sensor environment.**

***Index Terms*— Active learning (AL), convolutional neural network, cross-sensor domain adaptation (DA), land cover classification (LCC).**

I. Introduction

With the rapid exposure to advance remote-sensing technology in the past few decades, a massive volume of remotely sensed (RS) images are captured using different types of sensing devices, such as satellites (multispectral or hyperspectral), manned, and unmanned aircrafts. These huge amounts of images collected from multiple sensors (having different spectral, spatial, and temporal features) can be beneficial for the creation of accurate land cover classification (LCC) maps in a near real-time scenario. These maps can be utilized for precise monitoring of the Earth and ecosystem, which have a vast spectrum of applications like crop monitoring, deforestation control, soil contamination, water pollution, biogeochemical cycling, and thermal mapping [1].

In literature, the traditional LCC has been categorized in two different approaches: pixel-level LCC [2] and scene-level LCC [3]. Pixel-level strategy focuses on the pattern acquired from pixels, while the scene-level approaches considered the whole object or scene to extract features. However, both the categories are considered in the cross-sensor (multi-sensor) environment. Therefore, the classification approach refereed in the manuscript is named as LCC to generalize the problem.

Deep learning (DL) models (mainly supervised) are applied proficiently for the automatic generation of LCC maps considering homogeneous sensor data [3]– [6]. Scott *et al.* and Kalita *et al.* investigated a deep convolutional neural network (DCNN) approach; whereas, Hong *et al.* [5] explored the graph convolutional neural network. Furthermore, Li *et al.* [6] investigated a deep alignment network using the remote sensing knowledge graph for the aerial images. However, due to the rapid development of the RS-equipment, the images collected by the satellites significantly vary in terms of different spectral responses, ground sampling distances, and inherent noise of the instruments [7], [8]. In this regard, Hong *et al.* [8] investigated a classification technique considering the heterogeneous data (multi-sensor images in pixel-level and scene-level) for LCC. Nonetheless, DL models trained over images collected by one sensing device (i.e., source domain) underperforms in generation of LCC maps using images captured from the different sensors (i.e., target domain). This problem is well known as domain adaptation (DA) [4], [9] in remote-sensing environment. The state-of-the-art DA approaches are based on an ideal situation in which the resolution and dimension of data, as well as data-collecting satellites, are similar across the domains. However, under multi-sensor environment, a more complex situation with significantly different source and target domains is a common phenomenon [7]. For example, a classification model trained over the low-resolution pixel-level source images (e.g., Landsat-8 and MODIS) datasets performs poorly for the very high-resolution scene-level satellites image datasets (e.g., Ikonos-2 and Quickbird). In literature, it is observed that the development of such a technique for LCC (in multi-sensor environment from pixel-level to scene-level and vice versa) is seldom explored. This motivates us to develop effective techniques for LCC by resolving cross-sensor DA problem in presence of limited number (or in absence) of labeled target information.

The objective of this work is to develop effective techniques for cross-sensor DA problem in LCCs using RS images. In this regard, the present work has initially addressed the issue of the difference in feature dimension between hyperspectral and aerial images. Following this, an active learning (AL) has been proposed to select "most-informative" target samples for labeling from the target domain to subdue the effect of distribution differences between the source-target domains. Finally, a classifier, trained using the labeled source and "most-informative" target samples, is utilized to predict the land

Manuscript received September 17, 2021; revised November 7, 2021; accepted November 10, 2021. Date of publication November 23, 2021; date of current version January 10, 2022. *(Corresponding author: Moumita Roy.)*

The authors are with the Department of Computer Science and Engineering, Indian Institute of Information Technology Guwahati, Guwahati 781015, India (e-mail: indrakalita09@gmail.com; moumita2009.roy@gmail.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/LGRS.2021.3130285, provided by the authors.

Digital Object Identifier 10.1109/LGRS.2021.3130285

cover classes under an adaptive environment. The contribution of the proposed work can be highlighted under three aspects.

1) A solution for the multi-sensor DA problem (in pixel-level and scene-level) in LCC has been investigated.
2) A sample selection technique (stacking of sample) has been explored to balance the number of sample images under a cross-sensor environment.
3) A standard deviation-based AL (SDAL) technique has been investigated to identify "most-informative" target samples.

II. Related Work

In literature, the hyperspectral DA (HDA) can be sub-divided as data transformation [10], [11] and classification-based approaches [12], [13]. The former approaches can handle the problem of DA by projecting the samples from both the domains into a common subspace to neutralize their distribution differences, whereas the latter have utilized the source-target domain data to improve the performance of the models [12], [13]. It is worth mentioning, hyperspectral imaging is greatly affected by the spectral variability of the spectral signatures due to the various illumination and topological conditions, atmospheric factors, etc., [14]. Hong *et al.* [14] has proposed an augmented linear mixing model to consider such spectral variabilities. However, in the current investigation, the impact of spectral variability is not considered over the performance of the proposed DA model, and the variabilities between the aerial and hyperspectral data (HSD) in terms of dimensions, the number of bands, and training samples are considered during the investigation.

In scene level, the problem of DA in LCC using aerial images can be categorized viz. unsupervised, semi-supervised, and AL-based strategy. The unsupervised scheme follows the data transformation-based approaches to obtain the common subspace between the source–target regions [9], [15]. The semi-supervised approaches explore DA techniques by considering the limited amount of labeled samples from the target domain [16]. Similarly, AL techniques detect "most-informative" target samples for manual labeling [4].

The present state-of-the-art DA methods are primarily concerned with the homogeneous sensors' data (with similar feature space and same dimensions). The difference in data distribution across domains is mainly caused by atmospheric conditions and seasonal changes. However, in the remote-sensing domain, the source-target regions can contain completely different types of information in terms of feature space, resolution (Low/VHR), band varieties, or dimensionalities [7], [17]. Mateo *et al.* [7] explored a cross-sensor DA technique for cloud detection utilizing Landsat-8 and Proba-V images. Furthermore, Huang *et al.* [17] have investigated a multi-sensor DA problem using the active and passive satellite data.

III. Proposed Methodology

The proposed methodology has been carried out into three phases: feature extraction (FE), AL, and classification using two source→target scenarios: Hyperspectral→Aerial ($H \rightarrow A$) and Aerial→Hyperspectral ($A \rightarrow H$).

The explanation of proposed methodology along with graphical representation (in Fig. 1) is presented in Sections II-A–II-C.

A. Feature Extraction (FE)

In the present work, the source and target domains have a significant difference in terms of dimensionality of the patterns, number of bands, and number of available training samples. Here, FE algorithms ensure that the source–target samples are processed in the same dimensional feature space so that data can be handled by a common classifier. However, it has been observed that in some cases, the number of source samples is significantly higher than the few labeled target samples (as evident in $H \rightarrow A$ scenario). Therefore, data prepossessing of the source samples is required, prior to FE. Furthermore, for each adaptive source→target scenario, a separate FE technique has been involved due to significant differences between the samples across the domains. Let assume, $I^s = \{(\overrightarrow{x_i^s}, y_i^s)\}_{i=1}^m$ denotes the set of m number of source samples with C number of different class labels, where $y_i^s \epsilon \{1, 2, 3, \ldots, C\}$; similarly, $I^t = \{\overrightarrow{x_i^t}\}_{i=1}^n$ denotes the set of n number of target samples.

1) FE for $H \rightarrow A$ Scenario: In the current scenario, it is observed that the number of source samples (i.e., 119 258) is significantly higher than the limited number of labeled target samples (i.e., 837). In such cases, a model trained with HSD has become incompetent to learn from few target samples (to be obtained after AL) of aerial images. In this regard, a sample selection strategy has been developed to reduce the number of samples in HSD set, without losing any information. The algorithm (i.e., stacking of sample approach) and its visualization are available in Algorithm S-1 and Figure S-1 (in the supplementary material). Here, all the samples from the set I_s are distributed in distinct buckets (number of buckets, b is same as C) corresponding to their class labels. Thereafter, the f number of samples (f is same as number of feature bands in HSD) from each bucket is chosen at random to obtain one composite sample of dimension $f \times f$, and is added to the set I_{sp} as a new sample. The selected samples are eliminated from the bucket, and the process continues until the number of samples remaining in the bucket has attained a value less than f. The method will generate a new source domain dataset, (I_{sp}) with mp number of processed source samples, $I_{sp} = \{\overrightarrow{x_i^{sp}}\}_{i=1}^{mp}$; providing $\overrightarrow{x_i^{sp}}$ is the ith available sample in I_{sp}, and ($mp < m$).

After the processing of the HSD data, the FE strategy has been applied separately over the source and target samples (denoted by I_{sp} and I_t, respectively) to obtain sample sets of same dimensions. The FE for the source domain (with set I_{sp}) has been carried out by training a scratch DCNN model (as shown in in Fig. S-1). It contains three convolution layers (CNN1, CNN2, and CNN3), two dense layers (dense1 and dense2), and one classification layer (denoted by classification block). Each convolution layer is followed by a max-pooling (MP1, MP2, and MP3) operation. The dimension of the last dense layer (dense2) is d and the input dimension of the network is $f \times f$. The entire model is trained for a C number of classes. After training, the model is utilized as a feature

extractor to extract d-dimensional features for the samples in the set I_{sp}. Here, $\overrightarrow{\tilde{x}_i^{sp}}$ denotes the ith source sample with extracted d-dimensional features.

In the target domain, to extract the knowledgeable features, the pretrained VGG-16 [18] model is used as a feature extractor. In this regard, the extracted $d1$ dimensional target sample $\overrightarrow{\hat{x}_i^t}$ has been derived from target sample $\overrightarrow{x_i^t}$. Thereafter, an autoencoder (AE) neural network [19] is applied on the sample $\overrightarrow{\hat{x}_i^t}$ for further reduction of the feature dimension. The AE with one hidden layer of dimension d is trained using the $d1$ dimensional target samples $\overrightarrow{\hat{x}_i^t}$ as input. After training, the d-dimensional extracted target sample $\overrightarrow{\tilde{x}_i^t}$ has been obtained from the encoding layer of the trained AE. The entire architecture is presented in Fig. S-1 (denoted using the red color boxes).

2) FE for $A \rightarrow H$ Scenario: In the current scenario, the data prepossessing of source samples is not required due to its limited quantity (i.e., 837). However, the feature space dimension of each source sample (in scene-level) with three bands (i.e., $256 \times 256 \times 3$ for aerial images) is substantially larger than the feature space dimension of each target samples (in pixel-level) with 102 bands (i.e., 1×102 for HSD). Here, to reduce the feature-space dimensions of the source samples, initially a pretrained VGG-16 [18] model has been fine-tuned with the available source samples. Here, a new dense layer with the $d1$ dimension is introduced before the classification layer, and the final classification layer in the pretrained architecture is substituted with C number of neurons. Following this, the generated fine-tuned VGG-16 model is being used to extract the relevant features with reduced shape. In this regard, the fine-tuned model is used to derive the d-dimensional extracted sample $\overrightarrow{\tilde{x}_i^s}$ from the source sample $\overrightarrow{x_i^s}$. The FE phase is depicted in Fig. S-2.

The FE for all the target samples ($\overrightarrow{x_i^t}$ of dimension $f = 102$) can be accomplished by training an AE (with encoding dimension d) to maintain its dimension to be the similar as the feature dimension of the source samples. After training, the d-dimensional extracted target sample $\overrightarrow{\tilde{x}_i^t}$ from $\overrightarrow{x_i^t}$ has acquired from the encoding layer of the AE. The architecture of AE is depicted in Fig. S-2 (denoted using the red color boxes).

B. AL to Identify the "Most-Informative" Samples

In the present work, a SDAL strategy has been investigated to select few labeled target samples (cost number of "most-informative" target samples). The key objective of this strategy is to select the samples for active labeling from all the classes present in the target domain. The motivation and algorithmic representation of SDAL technique is described in Section S-1 and Algorithm S-2, respectively. In the algorithm, initially, all the source samples with extracted features ($\overrightarrow{\tilde{x}_i^s}$) are assigned to b number of different source buckets based on their actual class labels. Thereafter, the standard deviation (SD) between the features of each source sample is computed. In addition, the mean of the SD for all the source samples present in each bucket is also calculated. This bucket-wise mean value

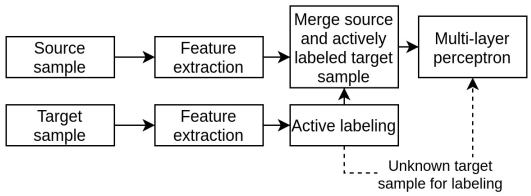

Fig. 1. Graphical representation of the entire model investigated in the letter.

is regarded as the representative score for the corresponding source bucket. In this regard, the representative score for jth source bucket is r_j^s. On the other hand, a soft-score is also obtained for all the target samples using their extracted features ($\overrightarrow{\tilde{x}_i^t}$). For this purpose, the SD of ith target sample ($\overrightarrow{\tilde{x}_i^t}$) is used as the soft-score (sf_i^t) of that respective sample. Thereafter, the absolute difference between the soft-score of each target sample and all representative scores of the source buckets is obtained. Here, the class label corresponding to the source bucket having the lowest difference is selected for soft labeling of each target sample. Moreover, the soft-score (sf_i^t) of the ith target sample is updated by this lowest difference value. Finally, the target samples are distributed to tb number of target buckets using their soft-labels. It may be noted that the number of distinctive soft labels for the target sample may also be fewer than the number of distinctive source labels. It means that the value of tb is less than or equal to b. For identifying the "most-informative" samples from the target domain, all samples in each target bucket are initially sorted in ascending order of their soft-scores (sf_i^t). Thereafter, in each iteration, the proposed strategy selects one target sample from each bucket for active labeling, by considering the top-most and bottom-most samples in alternative iterations. As a result, in each iteration, tb number of samples have been selected for active labeling. It implies that the cost of AL after the first iteration is tb. This process is continued until the cost of the algorithm is below the predefined cost, i.e., a prefixed number target samples for possible labeling. However, in each iteration, the selected target samples have been removed from the buckets.

C. Training of the Final Classification Model

In the proposed methodology, the source samples with extracted features and the set of actively labeled "most-informative" target samples with extracted features are merged to form the final training set. This labeled training set is utilized to train a multilayer perceptron model. Here, the number of input and output neurons are the same as the dimension of the extracted features (d) and the number of classes available in the training set (C), respectively. Finally, the trained classifier is used to predict the final labels of the land cover classes under the cross-sensor adaptive environment.

IV. Experimental Results

The effectiveness of the proposed method is assessed through experimentation using aerial and hyperspectral image datasets (HIDs). The experimental details (such as model training environment, parameters, and hyperparameters) of

the proposed architecture are explained in the supplementary material (Section S-2). A description of the two datasets is given below.

1) Aerial Image Merge Dataset (AIMD): The aerial image merge dataset (AIMD) is a combination of two aerial image datasets, namely aerial image dataset (AID) [20] and Asian subcontinent dataset (ASCD) [4]. The descriptions of the AID and ASCD datasets can be found in the supplementary material (Section S-1). In the current investigation, the images of the two classes (name: "Bareland" and "Meadow") are selected from the AID dataset and the images of the "Trees" and "Water" classes are selected from ASCD dataset to form the AIMD dataset. Finally, the AIMD dataset contains four classes (i.e., "Bareland," "Medow," "Trees," and "Water").

2) Hyperspectral Image Dataset (HID): The University of Pavia is a HID recorded over the Pavia campus in Northern Italy. The dataset consists of images with dimensions of $1096 \times 715 \times 102$ (102 denotes the spectral band) and nine different land cover classes. In the current study, samples from four classes are chosen i.e., "Water," "Trees," "Meadow," and "Bareland" for experimentation.

In the current investigation, two cross-domain scenarios (i.e., $H \rightarrow A$ and $A \rightarrow H$) have shared four similar classes (i.e., "Water," "Trees," "Meadow," and "Bareland"). The total number of samples from AIMD and HID datasets are 837 and 119 258, respectively. The class-wise sample distribution of both the datasets is depicted in Table S-1. Moreover, the two datasets (AIMD and HID) are made publicly available and can be accessed in the GitHub repository (link: *https://github.com/indrakalita/MultiSensor-DA*).

A. Result and Analysis

To validate the proposed and existing methods, the precision (P), recall (R), F1-score (F), and overall accuracy (OA)-values are considered as the performance measuring indices.

1) Assessment Analysis of the Performance of SDAL With Varying Cost: In the current investigation, the performance of the proposed SDAL scheme is evaluated for five different AL costs, with values 4, 8, 12, 16, and 20, respectively. The methods are evaluated for both $H \rightarrow A$ and $A \rightarrow H$ scenarios, and the evaluation statistics are presented in Tables S-2 and S-3, and graphical representations for the same are presented in Fig. S-3. From the statistics, it is observed that the highest values for $H \rightarrow A/A \rightarrow H$ scenarios are obtained as 53%/89% (at cost 4/20), 53%/75% (at cost 20/16), 51%/74% (at cost 20), and 61%/97% (at cost 20) for P, R, F, and OA-indices.

2) Comparative Analysis of Results Obtained Using the Proposed Approach and the Other State-of-the-Art Approaches: To evaluate the performance, the proposed SDAL scheme (with cost 4, 8, 12, 16, and 20) is compared with the five state-of-the-art schemes: 1) heterogeneous feature augmentation (HFA) [21]; 2) sub-space co-projection (SCP) [22]; 3) extreme learning machine-based heterogeneous domain adaptation (EHDA) [13]; 4) asymmetric adaptation neural network (AANN) [9]; and 5) ensemble decision-based heterogeneous domain adaptation (EDHDA) [4]. Out of these five compared schemes, three adaptation techniques (HFA, SCP, and EHDA) have used the HSD, and two techniques (AANN, EDHDA) have used aerial images for model performance evaluation. As already mentioned, no investigation has been observed (in literature) by considering the multi-sensor DA problem in LCC. Therefore, the authors have compared the proposed scheme with the DA works investigated separately on hyperspectral and aerial images. To maintain homogeneity during experimentation with the proposed approach (SDAL), the same training ecosystem in terms of source and target samples under a cross-sensor environment has been generated for all the five state-of-the-art techniques. Moreover, cross-sensor adaptation using the random selection method (RSL in Section S-1) is also compared with the SDAL algorithm. The results in tabular/bar graph representations of the comparative analysis have been presented in Tables/Figs. S-4/S-4 and S-5/S-5 for $H \rightarrow A$ and $A \rightarrow H$ scenarios, respectively.

a) Comparative analysis of results considering $H \rightarrow A$ scenarios: Analyzing Table S-4 and Fig. S-4, it is observed that for varying cost-values (cost = 4, 8, 12, 16, and 20), SDAL has outperformed the HFA by a margin of 18%–34% in terms of P, 14%–20% in terms of R and F, and by 15%–30% in the case of OA-index, respectively. Similarly, in comparison to SCP/EHDA/AANN/EDHDA, the performance of SDAL is found to be superior by a margin of 10%–30%/4%–27%/1%–24%/2%–28% in terms of P, 10%–15%/1%–12%/1%–8%/6%–8% in the case of R, 5%–12%/5%–17%/1%–11%/1%–8% for F, and of 15%–24%/13%–21%/5%–18%/6%–21% in case of OA-index, respectively. Moreover, the SDAL is more effective than RSL to deal with cross-sensor DA problems, and it outperforms RSL in terms of P, R, F, and OA- measures by a margin of 14%–45%, 3%–18%, 11%–19%, and 4%–19%, respectively, for the four different values of cost (i.e., 4, 12, 16, and 20). However, the SDAL technique is observed to be the comparable with RSL scheme for cost = 8. Here, the RSL scheme performs better than SDAL scheme by a margin of 2% in terms of P, 9% in terms of R, and 6% in terms of F-index. In terms of OA-index, SDAL is appeared to be significant by a margin of 4%. It is worth mentioning that SDAL has an edge over RSL in terms of overall accuracy (OA) for all five different cost values.

b) Comparative analysis of results considering $A \rightarrow H$ scenarios: By analyzing the results depicted in Table S-5 and Fig. S-5, it is evident that the proposed SDAL has significantly outperformed the considered state-of-the-art techniques for all varying cost-values (i.e., 4, 8, 12, 16, and 20). In comparison with HFA and SCP techniques, SDAL has achieved superior performance with significant margins of 6%–51% and 27%–47% in the case of P, by 22%–46% and 23%–37% in case of R, by 23%–56% and 18%–43% in case of F, and by margins of 15%–51% and 18%–52% in case of OA-index, respectively. For the rest state-of-the-art methods (EHDA, AANN, EDHDA) also, the proposed SDAL has significantly outperformed all, in terms of P, R, F, and OA. During the investigation, SDAL is found to be significant than RSL, by margins of 5%–49% in P, 1%–29% in R, 1%–43% in F, and 1%–25% in terms of OA-index for cost values of 4, 8, 2, and 16. However, it is observed that for higher cost-value (i.e., cost = 20), RSL has an edge over the scheme SDAL in terms

of R (by a margin of 6%), F-measures (by a margin of 7%), P (margin of 5%), and OA (margin of 1%)-measures. It has been observed that the performance of the proposed SDAL scheme is quite significant as compared to the RSL scheme for low cost values.

3) Ablation Analysis of the Proposed Model: In the experiment, the proposed model is decomposed to its distinct components (stacking of sample technique, different FE approaches) to evaluate the contribution of each component on the final architecture. In this regard, a synthetic model (i.e., Model A) is generated by considering two distinct FE strategies for the hyperspectral (HID) and aerial image (AIMD) datasets. Here, for the HID set, an AE is employed to extract 64-dimension of features from a feature space of 102-dimension. The same AE-based strategy is used for both the cases ($H \rightarrow A$ and $A \rightarrow H$), where HID is used as source and target domain dataset, respectively. In the case of the AIMD set, a pretrained VGG16 model is used to extract 4096-D features from the $256 \times 256 \times 3$ image dataset. This is followed by an AE model to further reduce the feature space into 64-dimensions. The same FE strategy is used for both the cases ($A \rightarrow H$ and $H \rightarrow A$), where AIMD is used as source and target domain dataset, respectively. In addition, Model A is updated by incorporating data augmentation (by putting noise and rotating images) on the AIMD to balance the sample ratio between the two datasets. In Table S-6, this method is referred to as Model B. Both the models are tested considering the cost as 20. These two models (Model A and Model B) are compared to the proposed model (refereed as proposed in Table S-6), and the results show that the proposed approach outperforms Model A/Model B by a margin of 34%/43%, 22%/28%, 30%/38%, and 28%/32% in terms of P, R, F, and OA-indices, respectively for $H \rightarrow A$ scenario. A similar observation has also been obtained for $A \rightarrow H$ scenario From the analysis, the authors are motivated to include different FE strategies (including stacking of sample approach) while taking label information (if available) into the proposed framework.

V. Conclusion and Future Scope

In the present work, a cross-sensor domain adaptation strategy for automated LCC has been explored using a deep neural network in integration with an AL strategy. The findings of the investigation over the collection of two different satellite image datasets confirm its potential to produce promising outcomes, relative to the other state-of-the-art techniques. In addition, it has been observed that the selection technique of the "most-informative" target sample through AL is highly stable as compared to the RSL approach. From this investigation, it can be comprehended that the proposed SDAL is significantly effective for the low-cost scenario under a cross-sensors DA environment. In the future, the similar investigation may be carried out using multispectral images. Moreover, the scheme can be made cost effective by investigating semi-supervised/unsupervised scheme instead of AL-based scheme. Further, the spectral unmixing process for spectral variability in the DA problem of LCC has also been kept as an interesting scope of future works.

References

[1] S. K. Meher and D. A. Kumar, "Ensemble of adaptive rule-based granular neural network classifiers for multispectral remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2222–2231, May 2015.

[2] M. Imani and H. Ghassemian, "Feature extraction using weighted training samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1386–1387, Jul. 2015.

[3] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land–cover classification of high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, Apr. 2017.

[4] I. Kalita and M. Roy, "Deep neural network-based heterogeneous domain adaptation using ensemble decision making in land cover classification," *IEEE Trans. Artif. Intell.*, vol. 1, no. 2, pp. 167–180, Oct. 2020.

[5] 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.

[6] Y. Li, D. Kong, Y. Zhang, Y. Tan, and L. Chen, "Robust deep alignment network with remote sensing knowledge graph for zero-shot and generalized zero-shot remote sensing image scene classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 145–158, Sep. 2021.

[7] G. Mateo-Garcia, V. Laparra, D. Lopez-Puigdollers, and L. Gomez-Chova, "Cross-sensor adversarial domain adaptation of Landsat-8 and Proba-V images for cloud detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 747–761, 2021.

[8] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[9] N. Ammour, L. Bashmal, Y. Bazi, M. M. A. Rahhal, and M. Zuair, "Asymmetric adaptation of deep features for cross-domain classification in remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 597–601, Apr. 2018.

[10] Y.-R. Yeh, C.-H. Huang, and Y.-C. F. Wang, "Heterogeneous domain adaptation and classification by exploiting the correlation subspace," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2009–2018, May 2014.

[11] C. Wang and S. Mahadevan, "Heterogeneous domain adaptation using manifold alignment," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2011, vol. 22, no. 1, p. 1541.

[12] X. Li, L. Zhang, B. Du, L. Zhang, and Q. Shi, "Iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2022–2035, May 2017.

[13] L. Zhou and L. Ma, "Extreme learning machine-based heterogeneous domain adaptation for classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1781–1785, Nov. 2019.

[14] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[15] L. B. E. Riz and B. Demir, "Domain adaptation based on deep denoising auto-encoders for classification of remote sensing images," *Proc. SPIE*, vol. 10004, Oct. 2016, Art. no. 100040K.

[16] T. Postadjian, A. L. Bris, H. Sahbi, and C. Malle, "Domain adaptation for large scale classification of very high resolution satellite images with deep convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 3623–3626.

[17] X. Huang, S. Ali, S. Purushotham, J. Wang, C. Wang, and Z. Zhang, "Deep multi-sensor domain adaptation on active and passive satellite remote sensing data," in *Proc. 1st KDD Workshop Deep Learn. Spatiotemporal Data, Appl., Syst. (DeepSpatial)*, Jan. 2020.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[19] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. New Delhi, India: Pearson, 2010.

[20] G.-S. Xia *et al.*, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[21] L. Duan, D. Xu, and I. Tsang, "Learning with augmented features for heterogeneous domain adaptation," 2012, *arXiv:1206.4660*.

[22] M. Xiao and Y. Guo, "Semi-supervised subspace co-projection for multi-class heterogeneous domain adaptation," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2015, pp. 525–540.