Cross-Domain Classification of Multisource Remote Sensing Data Using Fractional Fusion and Spatial-Spectral Domain Adaptation

Xudong Zhao, *Student Member, IEEE*, Mengmeng Zhang, Ran Tao, *Senior Member, IEEE*,
Wei Li, *Senior Member, IEEE*, Wenzhi Liao, *Senior Member, IEEE*, and Wilfried Philips, *Senior Member, IEEE*

***Abstract*—Limitation of labeled samples has always been a challenge for hyperspectral image (HSI) classification. In real remote sensing applications, we encounter a situation where an HSI scene is not labeled at all. To solve this problem, cross-domain learning methods are developed by utilizing another HSI scene with similar land covers and sufficient labeled samples. However, the disparity between HSI scenes is still a challenge in reducing the classification performance, which may be affected by variations in illumination and weather. As a robust supplement to these variations, light detection and ranging (LiDAR) data provide stable elevation and spatial information. In this article, we propose a multisource cross-domain classification method using fractional fusion and spatial-spectral domain adaptation to reduce the disparity between scenes. First, the spatial information of HSI is preserved by fractional differential masks. Then, the LiDAR data are utilized for spectral alignment of HSI. The utilization of LiDAR data reduces the pixel-level disparity between scenes. At last, a spatial-spectral domain adaptation network is proposed to reduce domain shift at the feature level and extract discriminative spatial-spectral features. Experimental results on HSI and LiDAR scenes show 5% –10% improvements in overall accuracy compared with the state-of-the-art methods.**

***Index Terms*—Cross-domain classification, fractional fusion (FrF), hyperspectral image (HSI), light detection and ranging (LiDAR), spatial-spectral domain adaptation (SSDA).**

I. Introduction

EXPONENTIAL growth of multisource remote sensing data has created a compelling demand for automatical analysis and interpretation of datasets [1], [2]. Artificial intelligence provides great opportunities for multisource remote sensing data analysis [3], [4]. Hyperspectral image (HSI)-based multisource remote sensing data have been systematically applied for monitoring land-use and land-cover classification, target detection, and environmental changes [5]–[8]. Specifically, HSI can provide detailed spectral information for uniquely discriminating materials of interest [9]. However, the acquisition of labeled samples is time-consuming and laborious or even infeasible [10].

In real remote sensing applications, we encounter the situation where an HSI scene to be classified (target domain) is not labeled at all because of the labor and natural limitation. Meanwhile, other similar scenes (source domain) sharing the same land covers may have sufficient labeled samples. A nature idea is to exploit the information of labeled samples in the source scene to help target scene classification. This task is called cross-domain classification [11], [12]. For instance, if a city suffered from a natural disaster, both the field condition and disaster losses need assessment through HSI-based multisource remote sensing data and classification techniques. On-site labeling samples in the investigated scene are not realistic, but some scenes captured from similar cities are easy to find and with sufficient labeled samples. These cities share similar land covers and make it possible to transfer knowledge between similar scenes.

A straightforward method for cross-domain HSI classification is training classifiers on the source scene to classify the target scene [13]. However this simple way suffers from spectral shift, i.e., pixels of the same land cover may vary in spectral reflectance profiles from two scenes [14]. As shown in Fig. 1, given the same material, the spectral reflectance profiles from two scenes may be significantly shifted due to the variation of sensor altitude, atmospheric condition, illumination, etc. [15], [16]. Fig. 1 illustrates the spectral reflectance of two HSI scenes, e.g., Trento in Italy (source domain) and Houston in the USA (target domain). Red and blue curves represent spectral reflectance of buildings in the source and target domains, respectively. Due to the spectral shift phenomenon, the classifiers trained using labeled training samples from the source domain perform poorly on the testing samples from the target domain.

The key issue of cross-domain classification is to reduce the spectral shift of source and target scenes. One significant reason for spectral shift is illumination, classical classification tasks

Manuscript received 27 March 2022; revised 30 April 2022 and 3 June 2022; accepted 9 July 2022. Date of publication 13 July 2022; date of current version 27 July 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62001023 and Grant 61922013, and in part by the Beijing Natural Science Foundation under Grant JQ20021. *(Xudong Zhao and Mengmeng Zhang contributed equally to this work.) (Corresponding author: Ran Tao.)*

Xudong Zhao is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China, and also with the Image Processing and Interpretation, IMEC Research Group, Ghent University, 9000 Ghent, Belgium (e-mail: zhaoxudong@bit.edu.cn).

Mengmeng Zhang, Ran Tao, and Wei Li are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: 7520200002@bit.edu.cn; rantao@bit.edu.cn; liwei089@ieee.org).

Wenzhi Liao is with the Flanders Make, 3920 Lommel, Belgium, and also with the Department of TELIN, Ghent University, 9000 Ghent, Belgium (e-mail: wenzhi.liao@gmail.com).

Wilfried Philips is with the Image Processing and Interpretation, IMEC Research Group, Ghent University, 9000 Ghent, Belgium (e-mail: wilfried.philips@ugent.be).

Digital Object Identifier 10.1109/JSTARS.2022.3190316

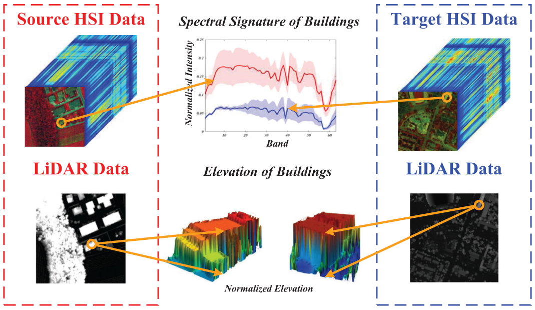


Fig. 1. Spectral shift of HSI and robust elevation feature of LiDAR in cross-domain classification. Solid lines depict mean spectral signature of specific land cover in HSI, and the shaded part indicates standard deviation of corresponding samples.

include atmospheric compensation routines to provide estimates of the average ground leaving reflectance retrievals [17]. However, per-pixel solar and sky ancillary illumination information varies in scenes and leads to poor classification performance [18]. To compensate for the pixel-level shift, a guidance auxiliary image must be robust to the variations in illumination.

As a robust supplement to these variations, light detection and ranging (LiDAR) data provide stable elevation and spatial information acquired at any time of the day and under adverse weather conditions [19]. LiDAR provides meaningful elevation and spatial information regardless of illumination variations between scenes. As shown in Fig. 1, the normalized elevation information and spatial shape of buildings in the source and target scenes are robust to environmental variations and maintain discrimination against other land covers. For instance, buildings shadowed by clouds in the target scene show shifted spectral reflectance, but the elevations still fall into the same range. In this regard, the diagnostic spectral information of HSI enables fine-grained classification of observed objects, while the vertical structural information of LiDAR can be applied for obscuration (cloud/shadow) area exploration. Notably, the sensor development enables HSI and LiDAR to be captured simultaneously. Utilizing HSI and LiDAR data of the same ground sample distance (GSD), we can obtain per-pixel solar and sky ancillary illumination information [20]. Recently, there has been an emergence of joint feature extraction methods being applied to HSI-based multisource classification [21]–[25]. In [25], a novel CNN denoiser was designed to regularize HSI and multispectral image fusion. These multimodality data come with their unique advantages over single modality and improve the classification accuracy [26]. However, researchers have not investigated whether LiDAR data can aid HSI for cross-domain pixel-level classification. By integrating the discriminative spectral feature of HSI and the robust elevation feature of LiDAR, we aim at leveraging the LiDAR data to improve cross-domain HSI classification by reducing the per-pixel spectral shift.

Incorporating LiDAR reduces data-level per-pixel spectral shift caused by illumination variation, but there exist feature-level disparities between source and target scenes. With the aid of information from the source domain and the related features, the target domain can be adapted with the source domain and classified by domain adaptation (DA) techniques [27]–[29]. Inspired by the traditional DA method in machine learning, DA technology has been introduced to HSI classification to reduce the data shift and distribution bias [30]–[32]. The maximum mean discrepancy criterion is widely used for first-order statistic alignment [33], [34]. To reduce the distribution bias, transfer component analysis (TCA) was introduced to cross-domain HSI classification [33]. In [35], correlation alignment (Coral) was proposed to align the correlation of scenes. In [36], stratified transfer learning (STL) was proposed to learn common subspace by exploiting the intra-affinity of classes.

Recently, deep DA technique achieved promising performance in cross-domain HSI classification tasks [37]–[39]. In [38], an unsupervised cross-domain HSI classification method was proposed based on adversarial DA. In [39], an augmented associative learning-based DA method was proposed for HSI classification. In [37], an efficient and effective model was created for HSI classification by implementing open-set DA and generative adversarial network (GAN). Although these DA methods reduce the disparity between scenes and improve the generalization ability, the discrimination of land covers is reduced in embedding space as well. Specifically, cross-domain HSI using GAN [37] only utilizes spatial information by convolutional layers while the discriminative spectral information is missing.

Focusing on the above challenges in cross-domain multimodal classification, this article aims at reducing spectral shift by data-level fractional fusion and modifying feature-level adaptation by spatial-spectral DA. In the proposed fractional fusion and spatial-spectral domain adaptation (FrF-SSDA), first a fractional differential mask (FrDM) is utilized to enhance spatial information of HSI. Then, we propose a fractional fusion method to fuse LiDAR data and HSI. Because of the robustness of LiDAR to illumination variation, the spectral shift of source and target scenes is reduced. Second, to improve cross-domain HSI classification by feature-level DA, we modified the GAN for spatial-spectral feature extraction and domain adaptation. The joint use of HSI and LiDAR can not only reduce spectral shift between multimodal datasets but also utilize the discriminative spectral feature of HSI and the robust elevation feature of LiDAR data. Finally, three groups of HSI and LiDAR data are used for cross-domain classification experiments, and the results compared with competitive methods indicate the effectiveness of the proposed FrF-SSDA.

The main contributions can be highlighted as follows.

1) A fractional fusion stage leveraging LiDAR data is proposed to reduce per-pixel spectral shift of HSI and improve cross-domain multimodal classification. The classical FrDM is modified for comprehensive spatial-spectral information extraction. Combined with LiDAR data, the shift of spectral signature for the same land cover in source and target domains is reduced.
2) A spatial-spectral domain-adaptive network is proposed to reduce the feature-level disparity. Based on the spatial adaptation ability of GAN, the modified SSDA network extracts sequential spectral features from the fused data, which aligns domains in feature space while improving feature discrimination.

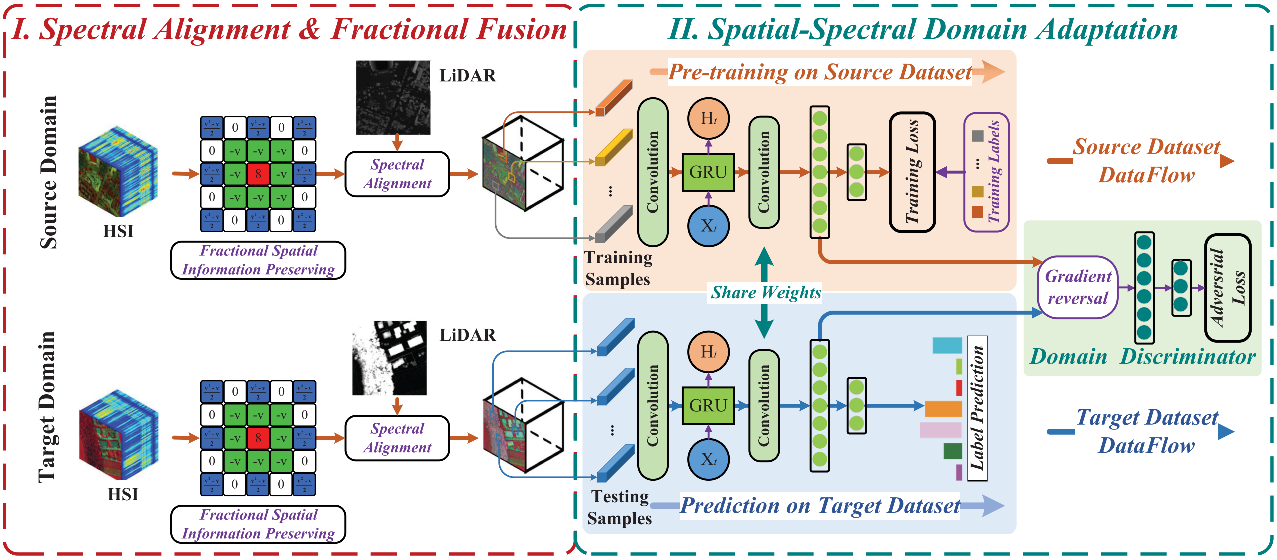


Fig. 2. Flowchart of proposed FrF-SSDA framework.

The rest of this article is organized as follows. The proposed framework is introduced in Section II. In Section III, experimental results and analysis are presented. Finally, Section IV concludes this article.

II. Proposed Framework

In this article, an FrF-SSDA method is proposed for multimodal cross-domain classification. The overall flowchart of the method is shown in Fig. 2.

The proposed FrF-SSDA consists of 1) a LiDAR-aided spectral alignment and fractional fusion part to reduce spectral shift, and 2) an SSDA network for feature-level adaptation. In Section II-A, first the problem statement and motivations of the proposed FrF-SSDA are depicted. Then, the two major parts of the proposed FrF-SSDA are depicted in Sections II-B and II-C, respectively. Finally, the training and inference procedures are introduced in Section II-D.

A. *Problem Statement and Motivations*

For a clear presentation, we first introduce the cross-domain classification problem statement and notations used in the proposed FrF-SSDA. In the cross-domain HSI classification task, the source scenes consist of registered HSI image $\mathbf{H}^S$ and LiDAR image $\mathbf{L}^S$, which are captured together and of the same GSD. The target domain consists of the registered HSI image $\mathbf{H}^T$ and LiDAR image $\mathbf{L}^T$. In cross-domain classification, the source and target scenes share the same land covers but only the source scenes have sufficient labeled samples. The cross-domain classification aims at exploiting information of the labeled samples in source scenes to classify the target scenes.

Affected by the capturing environment variations especially illumination changes, the spectral signatures of same land cover in $\mathbf{H}^S$ and $\mathbf{H}^T$ may be shifted. But the LiDAR data $\mathbf{L}^S$ and $\mathbf{L}^T$ provide meaningful elevation and spatial information regardless of illumination variations. Thus, one of the motivations is to reduce the per-pixel spectral shift caused by illumination changes by leveraging LiDAR data. To align the pixel-level spectral reflection of $\mathbf{H}^S$ and $\mathbf{H}^T$, a fractional fusion stage is designed using $\mathbf{H}^S, \mathbf{L}^S$ and $\mathbf{H}^T, \mathbf{L}^T$. The fused scenes are represented as $\mathbf{F}^S$ and $\mathbf{F}^T$. The motivation of this LiDAR data-aided spectral alignment and fractional fusion stage is to reduce data-level spectral shift, which is detailed in Section II-B.

To reduce feature-level disparity between $\mathbf{F}^S$ and $\mathbf{F}^T$, an SSDA network is designed. Given the labeled sample set $\mathbf{X}^S = \{x_1, x_2, \ldots, x_{ns}\}$ in the fused source scene $\mathbf{F}^S$ with their labels $\mathbf{Y}^S = \{y_1, y_2, \ldots, y_{ns}\}$, the target of cross-domain classification is predicting the labels of $\mathbf{X}^T = \{x_1, x_2, \ldots, x_{nt}\}$ in target scene $\mathbf{F}^T$. Considering both feature alignment and discrimination, the proposed SSDA aims to align scenes by projecting source and target features to shared feature space and improve the effectiveness of the classifier on the target scenes. Only the source labels $\mathbf{Y}^S$ are utilized and the DA from $\mathbf{X}^S$ to $\mathbf{X}^T$ is applied. Without using any target labels $\mathbf{Y}^T$, the test set consists of target samples $\mathbf{X}^T$. This feature-level SSDA process is detailed in Section II-C.

B. *Fractional Fusion*

As shown in part I of Fig. 2, the first stage of proposed method is leveraging LiDAR data to reduce per-pixel spectral shift and fractional fusion to preserve both spatial and spectral information.

HSI can provide detailed spectral information for uniquely discriminating materials of interest, but the spatial and textural information is usually limited. To preserve the valuable spatial information, an FrDM [40] step is applied for spatial information enhancement and preservation. To realize the cross-domain classification, discriminative spatial-spectral information is significant. We proposed FrF to improve the discrimination of spectral information while preserving spatial information.

Given an HSI image $\mathbf{H} \in \mathbb{R}^{r\times c\times b}$, the spatial fractional information of kth channel of $\mathbf{H}$ is preserved as $[\nabla^v \mathbf{H}_k]_j = ([D_1^v \mathbf{H}_k]_j, [D_2^v \mathbf{H}_k]_j)^T$, where $[\cdot]_j$ denotes the jth column, and D_1^v and D_2^v are the horizontal and vertical finite fractional difference operators, respectively, which are defined as

$$\begin{aligned}[D_1^v \mathbf{H}_k]_j &= \sum_{l=0}^{J-1} \omega_l^{(v)} \mathbf{H}_k(r+l, s)\\ [D_2^v \mathbf{H}_k]_j &= \sum_{l=0}^{J-1} \omega_l^{(v)} \mathbf{H}_k(r, s+l)\end{aligned} \tag{1}$$

where $\omega_l^{(v)} = (-1)^{l+1} C_l^v$, $C_l^v = \frac{\Gamma(v+1)}{\Gamma(l+1)\Gamma(v-l+1)}$ denotes the generalized binomial coefficient, $\Gamma[\cdot]$ is the Gamma function, and $J \geq 3$ is a positive integer.

With the preserved spatial information $\nabla^v \mathbf{H}$, the spectral shift problem maintains to be reduced. To compensate the illumination variations, LiDAR data are utilized to modulate the preserved HSI on a per-pixel basis. The per-pixel illumination intensity can be measured using the robust LiDAR source and utilized to align spectral-spatial information of HSI. The objective is to adjust the target spectral energy and make it statistically more similar to the source HSI with similar land covers. The alignment coefficient $\triangle$ is obtained from per-pixel solar and sky ancillary illumination information of LiDAR. With LiDAR data $\mathbf{L} \in \mathbb{R}^{r\times c}$ and overall intensity of HSI $\mathbf{I} = \sum_{k=1}^{b} \frac{1}{b} \widehat{\mathbf{H}}^{(k)}$, the coefficient $\triangle$ is defined as

$$\triangle = (\mathbf{L} - \mu(\mathbf{L}))\frac{\sigma(\mathbf{I})}{\sigma(\mathbf{L})} + \mu(\mathbf{I}) \tag{2}$$

where μ is the mean of the 2-D image while σ means the variance.

With the spectral alignment coefficient, we adjust the HSI scenes $\widehat{\mathbf{H}}^S$ and $\widehat{\mathbf{H}}^T$ to reduce the spectral shift caused by per-pixel illumination variations. The aligned data by fractional fusion are

$$\mathbf{F}^{(k)} = \nabla^v \mathbf{H}_k + (\Delta - \mathbf{I})/2 \tag{3}$$

where $\widehat{\mathbf{H}}^{(k)}$ and $\mathbf{F}^{(k)}$ are the kth band of the preserved HSI and aligned data $\mathbf{F}$, respectively. The source and target scenes after alignment and fusion are represented as $\mathbf{F}^S$ and $\mathbf{F}^T$, respectively.

More than considering the discriminative spectral signature of HSI images for cross-domain classification, preserving the spatial information of HSI and leveraging LiDAR data to reduce per-pixel spectral shift of HSI are also significant. The FrDM is applied to transfer the geometry of the HSI image (especially for spatial details and texture) into the fused multisource image, which models that the fractional-order gradient feature of the fused image should be consistent with that of the HSI image. Then, LiDAR-based alignment coefficients are utilized for spectral alignment of the spatial-spectral fused data. As the solid lines in Fig. 3 shows, even with the same land-cover buildings, the spectral reflectance of source (e.g., Houston13 in Fig. 3) and target scenes $\mathbf{H}^T$ (e.g., Trento in Fig. 3) are different. Due to the capturing environment changes (e.g., illumination variations), the spectral shift is a major obstacle that leads to poor classification performance on the target scenes. As shown in Fig. 3, the proposed fractional fusion and the LiDAR-based per-pixel alignment stage reduce the spectral shift and maintain the aligned spectral-spatial information. Compared with the original spectral feature, the fused data not only show a reduced spectral shift but also show a smaller standard deviation, which improves the discrimination of spectral information.

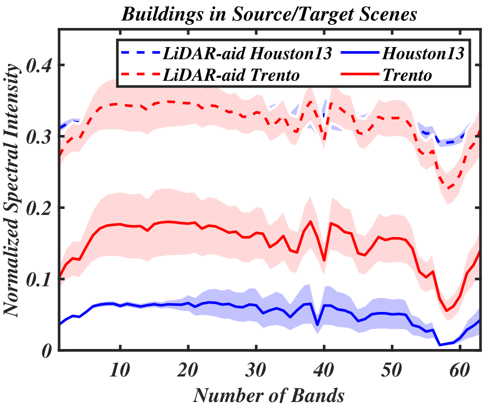


Fig. 3. Effects of the proposed fractional fusion stage. Solid lines depict the mean spectral signature of HSI, and the dashed lines represent the aligned mean spectral signature of fused data. The shaded part indicates the standard deviation.

C. Spatial-Spectral Domain Adaptation

Section II-A demonstrates that fusing HSI and LiDAR can reduce the pixel-level spectral shift caused by illumination variation, but there exists feature-level disparity between source and target domains. As shown in part II of Fig. 2, the SSDA network is designed to generate domain-invariant and class-discriminative features. In this stage, two aspects are ensured including 1) projecting source and target features to a shared feature space to reduce the feature-level domain shift, and 2) improving the discrimination of spectral features by recurrent layers.

Given the fused source and target scenes, the labeled training set $\mathbf{X}^S = \{x_1, x_2, \ldots, x_{ns}\} \in \mathbb{R}^{1\times 1\times b}$ in the fused source scene $\mathbf{F}^S \in \mathbb{R}^{r\times c\times b}$ and their labels $\mathbf{Y}^S = \{y_1, y_2, \ldots, y_{ns}\} \in \{1, 2, \ldots, Nc\}$ are used for pretraining, where $r \times c$ is image sizes while b being the number of band. Then, by utilizing unlabeled sample set $\mathbf{X}^T = \{x_1, x_2, \ldots, x_{nt}\} \in \mathbb{R}^{1\times 1\times b}$ in the target scene $\mathbf{F}^T \in \mathbb{R}^{r\times c\times b}$, the motivation of cross-domain classification is predicting the labels $\mathbf{Y}^S = \{y_1, y_2, \ldots, y_{ns}\} \in \{1, 2, \ldots, Nc\}$ for $\mathbf{X}^T$.

Considering both feature alignment and feature discrimination, the objective of the proposed SSDA is to align source and target scenes by projecting features to shared feature space and improve the effectiveness of classifier on the target scene. The designed SSDA consists of feature extractor $\mathcal{G}_f(\cdot;\theta_f)$, domain discriminator $\mathcal{G}_d(\cdot;\theta_d)$, and classifier $\mathcal{G}_c(\cdot;\theta_c)$, with their corresponding parameters $\theta_f, \theta_d,$ and θ_c.

In a classical GAN, convolutional layers are applied to extract the local 2-D spatial information. However, the spatial receptive field produced by stacking multiple convolutional kernels is limited, resulting in long-range dependencies that are still not adequately captured. In addition, the lack of the ability to extract

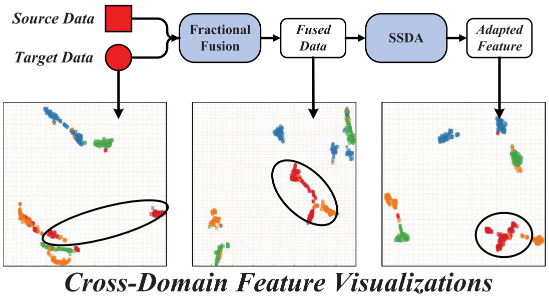


Fig. 4. Visualization of aligned features by the proposed FrF-SSDA framework.

sequential spectral information results in the loss of spectral information, which limited the DA and classification performance. To modify the baseline GAN method, a spatial-spectral feature extractor is designed to project the source and target features to a shared feature space in the proposed SSDA. The feature extractor $\mathcal{G}_f$ consists of convolutional blocks to maintain complete spatial information, and gate recurrent unit (GRU) [41] blocks to acquire long-distance nonlocal dependencies between spectra while with clearer spatial location information and more detailed sequential spectral information. $\mathcal{G}_f$ projects the input samples into a shared feature space, which allows the classifier $\mathcal{G}_c$ to predict the label of source and target samples in the same manner. For instance, given a labeled sample in source domain $x_i \in \mathbf{X}^S$ with its label $y_i \in \mathbf{Y}^S$, the pretraining classification loss is

$$\begin{aligned} L_c^i(\theta_f, \theta_c) &= L_c\left(\mathcal{G}_c\left(\mathcal{G}_f\left(x_i; \theta_f\right); \theta_c\right), y_i\right) \\ &= \log \frac{1}{\mathcal{G}_c(\mathcal{G}_f(x_i;\theta_f);\theta_c)_{y_i}}. \end{aligned} \quad (4)$$

Given $\mathbf{X}^S$ and $\mathbf{X}^T$ from fused data, the objective of the feature extraction part is to learn the spectral features and drive the target samples under the support of the source samples. As shown in Fig. 4, with the spatial-spectral feature extractor applied, the preserved spatial information is extracted by convolutional blocks while the GRUs emphasize the discriminative spectral information. By training the network in an adversarial manner, the domain shift between source and target samples is reduced.

The domain discriminator $\mathcal{G}_d(\cdot;\theta_d)$ is designed to cripple the ability to detect whether the samples belong to the source or target domains. And then, the domain discriminator can adjust the classifier to fit the source samples closely and drive the target samples under the support of the source domain. Then, the domain loss is

$$\begin{aligned} L_d^i(\theta_f, \theta_d) &= L_d\left(\mathcal{G}_d\left(\mathcal{G}_f\left(x_j; \theta_f\right); \theta_d\right), d_i\right) \\ &= d_i \log \frac{1}{\mathcal{G}_c\left(\mathcal{G}_f\left(x_j; \theta_f\right); \theta_d\right)} + \cdots \\ &\quad + (1 - d_i) \log \frac{1}{1 - \mathcal{G}_c\left(\mathcal{G}_f\left(x_j; \theta_f\right); \theta_d\right)} \end{aligned} \quad (5)$$

Algorithm 1: FrF-SSDA Framework.

```
Require: Source data $\mathbf{H}^S, \mathbf{L}^S$, source labels $\mathbf{Y}^S$, target data $\mathbf{H}^T, \mathbf{L}^T$, training epochs $Ep$.
Ensure: Target labels $\mathbf{Y}^T$.
Preserve fractional spatial information as (1).
Spectral alignment leveraging LiDAR $\widehat{\mathbf{L}}^S, \widehat{\mathbf{L}}^T$ as (2).
Fractional fuse and extract training samples $\mathbf{X}^S$ and $\mathbf{X}^T$.
Initialize all weights and bias terms
while epoch $< Ep$ do
  for $i$ from 1 to $n$ do
    Train $\mathbf{X}^S$ and compute $L_c^i(\theta_f, \theta_c)$ as (4).
    Train $\mathbf{X}^T$ and compute $L_d^i(\theta_f, \theta_d)$ as (5).
    Compute overall loss as (6).
    Backpropagation and update weights.
  end for
end while
Predict target labels $\mathbf{Y}^T$.
```

where d_j is the domain label for the jth sample and $d_j = 0$ means that x_j is from the source domain while $d_j = 1$ for the target domain. To realize DA by backpropagation, a gradient reversal layer is $\mathcal{R}(x)$ used [27].

Considering both domain adaptation and discriminative feature extraction, the complete optimization objective is

$$E = \frac{1}{ns}\sum_{i=1}^{ns} L_c(\mathcal{G}_c) - \lambda \frac{1}{ns}\sum_{i=1}^{ns} L_d - \lambda \frac{1}{nt}\sum_{j=1}^{nt} L_d \quad (6)$$

where λ is the weight of discriminator and is adaptively controlled by the ratio parameter $\lambda(l) = \frac{2}{1+exp(-10\frac{l}{\gamma n})} - 1$ with l being the number of training epoch. By updating the parameters $\theta_f, \theta_c,$ and θ_d in the training process, the proposed SSDA aligns source and target scenes, adapts classifier $\mathcal{G}_c$ to the target scene, and predicts the target labels $\mathbf{Y}^T$.

The proposed SSDA modifies the baseline GAN method and employs the convolutional GRU in the feature extractor to acquire long-distance nonlocal dependencies between spectra. While the convolutional layers maintain complete spatial information, more detailed sequential spectral information is captured by GRU. As shown in Fig. 4, the spatial-spectral feature extraction step improves the discrimination of different kinds of land covers. The domain-adaptive network aligns domains in the same feature space and reduces the feature-level domain shift.

D. Training and Inference Procedures

The training and inference procedures of the proposed FrF-SSDA are divided into four steps, as Fig. 5 shows. Parameters of convolutional layers and GRU are listed in the figure for example.

1) *Step 1: LiDAR-aided spectral alignment and fractional fusion.* We first designed a v-order FrDM to preserve spatial information of the HSI scenes. Then, LiDAR images are incorporated to compensate for the per-pixel illumination variations. In the process of spectral alignment, both the

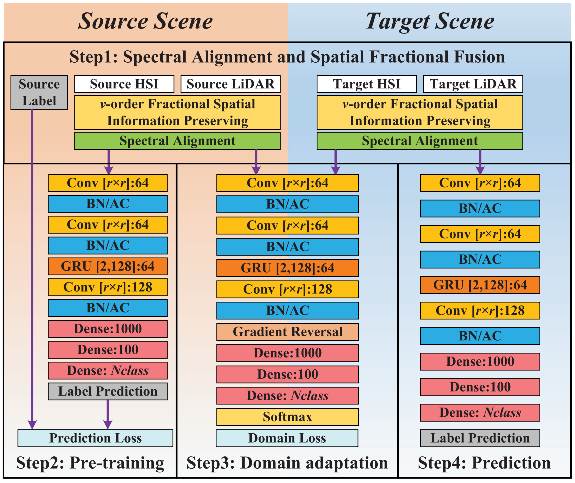


Fig. 5. Network structure, and training and inference processes of the proposed FrF-SSDA framework.

HSI and LiDAR data are normalized. Finally, each preserved HSI band is adjusted using its LiDAR-aided alignment coefficient to reduce the per-pixel spectral shift. This step fuses the discriminative spectral feature of HSIs and robust elevation information of LiDAR images, preserving the valuable spatial information of HSIs and reducing the spectral shift.

2) *Step 2: Pretraining on the source scene.* After the fractional fusion stage in step 1, normalization is applied again to scale the input of the SSDA network. The pretraining on the source scene trains the feature extractor and predictor on the source scene to classify the source samples correctly. $\mathcal{G}_f$ takes each sample $x_i \in \mathbf{X}^S$ as input and learn the spectral-spatial feature. A standard cross-entropy loss is used as (4). This step makes the network fit the source features, and works in an adversarial manner with the following step 3 to align the source and target scenes at feature level.

3) *Step 3: Domain discriminating.* With the initialized θ_f by step 2, θ_d is adjusted in an adversarial manner to maximize the discrimination loss. This step is executed in both source and target domains and aligns the extracted spatial-spectral features. By minimizing $E(\theta_f, \theta_c, \theta_d)$, parameters are updated to train the proposed network.

4) *Step 4: Inference of the target samples.* After training the feature extractor and domain discriminator in steps 2 and 3, the inference process fixes the parameters and predicts the target labels $\mathbf{Y}^T$. The detailed training and inference procedures of the proposed FrF-SSDA are summarized in Algorithm 1.

III. Experimental Results and Comparison

A. Data Description and Preparation

To verify the effectiveness of the FrF-SSDA, three groups of multimodal scenes are utilized for experimental analysis.

1) Houston Scenes: They consist of two urban scenes acquired over the University of Houston campus, Houston, TX, USA. The IEEE GRSS DFC released these HSI and LiDAR data in 2013[1] and 2018,[2] respectively. The spatial resolution and spectral bands of Houston13–Houston18 scenes are different. We first reserve the overlap area of two scenes containing similar land covers, and then reduce the spatial resolution of Houston18 to that of Houston13 by down-sampling. The 48 co-occurrence spectral bands of HSIs in Houston13 and Houston18 are extracted. The HSI and LiDAR data are shown in Fig. 8 while the labeled samples are listed in Table II. To evaluate the effectiveness of cross-temporal classifiers, seven general land-cover categories of interest are selected for example.

2) Nashua–Hanover Scenes: They consist of two urban scenes acquired over the city of Nashua and Hanover, USA, which comes from G-LiHT data.[3] The Nashua dataset supplies coregistered LiDAR and HSI with 450×1050 pixels at a GSD of 1 m. The Hanover dataset supplies coregistered LiDAR and HSI with 700×620 pixels at a GSD of 1 m. The HSI data contain 114 spectral channels ranging from 0.42 to 0.95 μm. The HSIs and LiDAR data are shown in Fig. 9 with seven land-cover categories of interest.

3) Houston–Trento Scenes: They consist of two HSI and LiDAR scenes acquired over different areas. One set of HSI and LiDAR-based digital surface model (DSM) is Houston13 and another named Trento was acquired over a rural area in Trento, Italy [42]. We extract the 64 co-occurrence spectral bands of HSIs in Houston13 and Trento. The HSIs and LiDAR data are shown in Fig. 10 with four land-cover categories of interest.

B. Experimental Setup

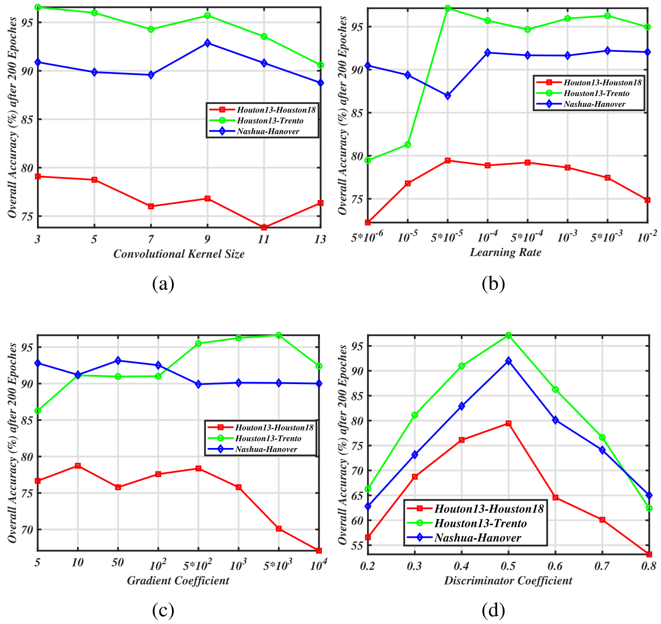
1) Implementation Details: The programs are implemented using MATLAB and Python 3.6. The networks are constructed using Pytorch. Experiments are conducted on a personal computer equipped with Ubuntu18.04 and NVIDIA GeForce RTX 2080 Ti. To quantify the experimental results, three commonly used evaluation metrics are adopted including overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa). In the training phase, we use Adam optimizer. No regularization term has been applied in cost functions (4)–(6). The learning rate can be updated by multiplying the initial learning rate by $1 - \frac{\text{epoch}}{\text{Epochs}}$, where the basic learning rate is 5×10^{-3} and the weight decay is 10^{-3}. Pixel-by-pixel input strategy is used in training the network with the input batch size being 128.

2) Algorithm Configuration: The proposed FrF-SSDA contains four parts as follows. In the fractional fusion and spectral alignment part, $v = 0.5$ order FrDM is applied for spatial information preservation. $n = 3$ approximate expression is used. In the SSDA network, each convolutional block contains a convolutional layer with $r \times r$ filter, a batch normalization layer, and a ReLU activation layer. Each GRU contains two recurrent

[1] [Online]. Available: http://www.grss-ieee.org/community/technical-committees/data-fusion/2013-ieee-grss-data-fusion-contest/

[2] [Online]. Available: http://www.grss-ieee.org/community/technical-committees/data-fusion/2018-ieee-grss-data-fusion-contest/

[3] [Online]. Available: https://glihtdata.gsfc.nasa.gov



(a) (b)

(c) (d)

Fig. 6. Classification performance of the proposed FrF-SSDA with different parameters. (a) Convolution kernel size $r \times r$, (b) Learning rate. (c) Gradient Coefficient. (d) Discriminator Coefficient.

TABLE I
ABLATION ANALYSIS OF THE PROPOSED FRF-SSDA USING DIFFERENT FUSION STAGES (%)

Methods	RAW	HPF	GS	BRO	WAVE	GRW	SFPSD	FrF
Source: Houston13, Target: Houston18								
OA	70.00	65.96	76.63	75.53	66.63	75.17	77.31	**79.45**
AA	64.59	67.45	73.55	62.84	65.70	61.84	74.02	**79.07**
Kappa	0.4861	0.4443	0.6023	0.5696	0.4033	0.5708	0.6127	**0.6755**
Source: Nashua, Target: Hanover								
OA	81.49	83.25	85.59	77.36	82.77	87.44	88.61	**93.16**
AA	65.12	64.04	75.04	62.10	70.09	74.31	**75.64**	74.93
Kappa	0.7440	0.7019	0.7953	0.6847	0.7219	0.8126	0.8354	**0.9035**
Source: Houston13, Target: Trento								
OA	67.79	67.38	67.45	88.21	82.30	88.67	89.73	**97.14**
AA	64.79	64.55	65.79	62.62	70.13	67.81	69.54	**94.21**
Kappa	0.4570	0.4616	0.4778	0.7687	0.6522	0.8061	0.8213	**0.9432**

layers with 128 features in the hidden state. There are three dense blocks with the size of output feature $1000, 100, P$. The first two include a linear layer, ReLU, and batch normalization layer. The last layer only contains a linear layer and is then fed into the softmax function for classification. The domain discriminator shares the same network setting with the predictor except for a gradient reverse layer. The Gradient parameter is initialed as $\gamma = 500$.

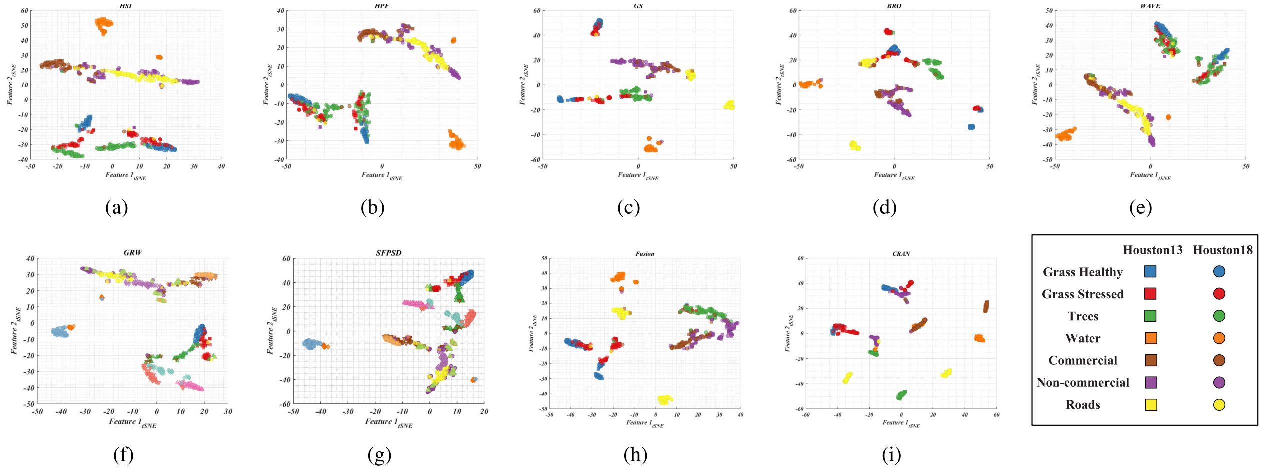
3) Parameter Analysis: To validate the effectiveness and sensitivity of parameters involved in the proposed FrF-SSDA, experimental analysis using varying parameters are compared in Fig. 6.

The effect of convolution kernels sizes $r \times r$ is depicted in Fig. 6(a). The researched range is constrained from 3×3 to 13×13. As shown in Fig. 6(a), the input blocks with different sizes yield different classification performances. Small sizes yield better performance, and all three tasks obtain the best classification results with 3×3 convolutional kernels. In Fig. 6(b), the OA after 200 training epochs shows that learning rate lr affects the convergence of the learning process. The search range is constrained from 5×10^{-6} to 10^{-2}. The algorithm using a small learning rate, e.g., $5 \times 10^{-6}, 10^{-5}$, cannot converge and results in poor performance. Large learning rates 10^{-2} means large fluctuations in the objective function and results in an unstable training process. For the gradient coefficient γ in $\mathcal{G}_d$, the search range is constrained from 5 to 10^4, which indicates that the classification performance varies with the ratio of adapted information. A larger γ means a larger increase in the speed of domain loss, which transfers more source information to target scenes. Fig. 6(c) shows $\gamma = 500$ is optimal, which means a proper weight of information adaptation. Furthermore, a hard discriminator between source and target samples is also important hyperparameter, Fig. 6(d) shows the discriminator coefficients perform best when $t = 0.5$ on all the three datasets. These parameters can be set by cross-validation on the available training set in practical applications. The optimal parameters in the researched three cross-domain tasks are $r \times r = 3 \times 3, \text{lr} = 5e - 5, \gamma = 500$.

C. Ablation Analysis

To demonstrate the effectiveness of FrF-SSDA, the extracted features are shown with visualized feature maps. The discrimination of different land covers is improved step by step while the extracted features are aligned between source and target scenes.

As listed in Table I, the proposed fractional fusion stage is compared with other fusion methods. Simple HSI and LiDAR feature splicing or stacking operations are highly susceptible to redundant information stacking, and cannot reduce the spectral shift. As shown in Fig. 3, the source and target scenes are aligned by the proposed fusion stage. The proposed fractional fusion stage reduces data-level spectral shift by incorporating LiDAR scenes. The t-distributed stochastic neighbor embedding feature visualization algorithm is used to visualize the similarity of data [49]. As shown in Fig. 7, the proposed fractional fusion stage improves the feature discrimination of different land covers. As listed in Table I, addition (ADD), high-pass filtering (HPF) [43], Gram–Schmidt pan-sharpening (GS) [44], Brovey transform (BRO) [45], additive wavelet transform (WAVE) [46], generalized random walks (GRW) [47], and subpixel registration (SFPSD) [48] are used to fuse HSI and LiDAR data. However, the classical fusion methods cannot reduce the spectral shift, which results in unsatisfactory feature discrimination and classification performance. As shown in Fig. 7, the red, green, and blue circles/squares, respectively, show that the stressed grass, healthy grass, and tree land cover are well separated by the proposed fractional fusion stage. These land covers are with similar spectral signatures but different elevations, which results in misclassification using original HSI data. With other competitive fusion methods, these land covers are still difficult to distinguish, while the proposed fractional fusion stage improves feature discrimination well. Furthermore, compared with the baseline GAN model, the proposed SSDA network included GRU blocks to acquire long-distance nonlocal dependencies between spectra while with detailed sequential spectral information. As shown



(a) (b) (c) (d) (e)

(f) (g) (h) (i)

Fig. 7. Feature visualization of different fusion methods on Houston13–Houston18 datasets. (a) Original HSI. (b) HPF [43]. (c) GS [44]. (d) Brovey [45]. (e) Wavelet [46]. (f) GRW [47]. (g) SFPSD [48]. (h) Fractional fusion. (i) SSDA feature.

TABLE II
COMPARISON OF THE CLASSIFICATION ACCURACY (%) USING THE HOUSTON13–HOUSTON18 SCENES

Class	Samples (Train/ Test)		SVM [13]		STL [36]		Coral [35]		MEDA [31]		JGSA [32]		JDA [30]		DSAN [50]		TsTNet [51]		GAN [37]		FrF-SSDA	
	Source: Houston13	Target: Houston18	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	Proposed
1. Grass Healthy	(345/0)	(0/1353)	99.78	96.38	87.21	97.78	99.78	96.45	78.71	100.00	87.66	99.78	79.45	90.54	62.31	67.55	85.03	70.29	76.99	81.52	70.04	95.42
2. Grass Stressed	(365/0)	(0/4888)	62.54	39.73	93.00	61.87	64.22	39.50	75.98	26.47	31.94	29.09	94.66	67.92	77.50	45.62	68.73	77.62	32.51	43.80	35.80	41.49
3. Trees	(365/0)	(0/2007)	17.90	77.91	53.90	88.36	17.82	77.48	63.12	22.05	75.85	57.09	66.96	79.79	74.55	88.03	52.82	78.45	10.91	58.64	66.37	59.83
4. Water	(285/0)	(0/22)	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
5. Commercial	(319/0)	(0/5347)	48.34	10.08	66.02	76.73	47.95	9.73	85.52	1.95	65.49	55.17	63.96	25.81	3.53	65.92	79.65	84.89	3.53	65.92	79.65	84.89
6. Noncommercial	(408/0)	(0/32459)	91.43	88.37	72.56	65.12	91.38	88.65	13.53	4.16	49.92	54.76	67.76	71.80	91.92	86.64	84.37	83.52	91.92	86.64	84.37	83.52
7. Roads	(443/0)	(0/6365)	19.43	78.15	64.62	12.16	18.49	85.60	75.24	89.58	76.95	86.11	60.75	82.31	72.58	39.28	15.87	88.31	72.58	39.28	15.87	88.31
OA			72.23	74.48	72.24	61.71	72.19	74.26	38.16	19.61	55.39	57.48	69.28	68.98	71.54	73.37	70.00	79.45	71.54	73.37	70.00	**79.45**
AA			62.78	70.09	76.76	71.72	62.81	69.63	70.30	49.17	69.69	68.86	76.22	74.02	55.49	67.97	64.59	79.07	55.49	67.97	64.59	**79.07**
Kappa			49.60	56.32	56.99	41.24	49.46	55.87	0.3057	0.1356	0.3989	0.4189	0.5362	0.5262	47.23	56.67	48.61	67.55	47.23	56.67	48.61	**67.55**

in Fig. 7(f) and (g), the features from source and target domains are gathered and show better discrimination. Compared with the GAN-based unsupervised domain adaptation (UDA) models, the proposed SSDA gains 5% improvements shown in quantitative evaluations.

D. Experimental Results and Comparison

To validate the effectiveness of the proposed FrF-SSDA, experimental results on three cross-domain multisource data are compared with other competitive classifiers. The compared methods include support vector machine (SVM) [13], STL [36], Coral [35], joint distribution adaptation (JDA) [30], manifold embedded distribution alignment (MEDA) [31], joint geometrical and statistical alignment (JGSA) [32], deep subdomain adaptation network (DSAN) [50], Topological structure and Semantic information Transfer network [51], and GAN [37]. The experimental setups of compared methods are optimized as suggested. All compared methods use original image patches of HSI and LiDAR data as inputs without data augmentation.

The proposed FrF-SSDA is adapted to align the source and target scenes, and then uses the trained classifier to predict the labels of target samples. The first experiment is a cross-temporal task that trains the classifiers on Houston13 and tests on Houston18. The second and third tasks are cross-site classification, which is trained using the Nashua dataset and tested on Hanover dataset, trained using the Houston13 dataset, and tested on Trento dataset, respectively. The number of training samples is shown in Tables II–IV. For comprehensive information extraction from the source scene, all the samples with corresponding labels are used. Parameters of the competitive algorithms are optimized and the same training and testing samples are used for a fair comparison.

The qualitative results of competitive methods and proposed FrF-SSDA are shown in Figs. 8–10(f)–(h) with corresponding accuracies in Tables II–IV. From the comparison between classification using the original HSI scenes and the fused scenes, there is an improvement of about 10%. The fractional fusion stage preserves the spatial information and reduces the spectral shift between scenes, which improves the generalization of the proposed classifier. The traditional method: SVM is susceptible to spectral shift and variation, which further leads to unacceptable results. Specifically, in Fig. 8(d), it can be seen that the lack of information fusion and DA makes it difficult for SVM to predict the label of target samples, which leads to low accuracies in Table II. Classical statistical DA methods, e.g., STL and Coral, rely on model and lead to unstable performance. Without accurately learned models using HSI and LiDAR, the information loss results in a negative effect on the feature discrimination. Based on the GANs, GAN and proposed FrF-SSDA can enhance the alignment between source and target domains. By incorporating source label information into the target domain,

TABLE III
COMPARISON OF THE CLASSIFICATION ACCURACY (%) USING THE NASHUA–HANOVER SCENES

Class	Samples (Train/ Test)		SVM [13]		STL [36]		Coral [35]		MEDA [31]		JGSA [32]		JDA [30]		DSAN [50]		TsTNet [51]		GAN [37]		FrF-SSDA	
	Source: Nashua	Target: Hanover	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	Proposed
1. Trees	(12713/0)	(18605/0)	98.20	88.74	94.85	95.37	81.17	92.72	56.21	67.12	40.46	42.32	94.09	94.54	91.11	98.22	88.64	91.25	96.00	93.81	89.43	98.09
2. Grass	(7528/0)	(4410/0)	60.27	100.00	50.48	100.00	43.06	100.00	75.71	95.62	91.34	89.73	57.44	99.98	100.00	61.07	100.00	92.22	40.29	96.64	97.51	100.00
3. Water	(16323/0)	(5425/0)	100.00	100.00	100.00	99.98	100.00	100.00	11.01	100.00	99.98	99.54	99.93	99.96	100.00	100.00	100.00	100.00	100.00	89.13	100.00	100.00
4. White roof	(698/0)	(860/0)	28.02	63.95	30.12	67.67	27.67	61.86	84.65	96.63	90.93	95.23	44.53	83.14	49.49	51.40	63.26	73.95	97.44	71.40	29.53	29.53
5. Grey roof	(3417/0)	(4338/0)	84.26	88.31	77.89	84.00	86.40	86.70	4.36	1.38	73.72	75.40	67.61	79.16	81.17	75.93	89.03	90.69	7.05	62.47	92.55	87.51
6. Roads	(17476/0)	(4955/0)	93.42	67.49	51.93	79.07	41.39	73.18	11.58	15.30	46.76	85.11	51.10	72.51	89.69	94.03	80.03	89.45	7.47	80.67	29.81	79.31
7. Bare soil	(8821/0)	(764/0)	23.56	16.75	27.49	18.19	19.50	19.63	98.95	79.32	25.13	29.06	27.75	26.18	45.93	27.23	45.63	56.81	56.41	28.80	17.02	27.62
OA			88.82	86.59	80.14	90.91	72.06	89.07	43.01	61.07	58.68	64.65	79.62	89.61	81.92	88.66	87.00	91.02	67.89	86.56	81.49	**91.99**
AA			69.68	75.04	61.82	77.76	57.03	76.30	48.93	65.05	66.90	73.77	63.21	79.35	66.20	72.55	78.94	**84.91**	57.81	74.70	65.12	74.58
Kappa			0.8383	0.8139	0.7138	0.8719	0.6096	0.8475	0.2761	0.4908	0.4953	0.5631	0.7086	0.8542	0.8083	0.8363	0.8027	0.8754	0.5309	0.8084	0.7440	**0.8861**

TABLE IV
COMPARISON OF THE CLASSIFICATION ACCURACY (%) USING THE HOUSTON13–TRENTO SCENES

Class	Samples (Train/ Test)		SVM [13]		STL [36]		Coral [35]		MEDA [31]		JGSA [32]		JDA [30]		DSAN [50]		TsTNet [51]		GAN [37]		FrF-SSDA	
	Source: Houston13	Target: Trento	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	GS	HSI	Proposed
1. Trees	(365/0)	(0/9123)	100.00	100.00	100.00	100.00	100.00	100.00	12.57	99.87	98.89	99.79	100.00	100.00	34.44	71.46	92.20	89.07	69.63	78.92	78.24	99.74
2. Buildings	(319/0)	(0/2903)	0.17	92.56	69.51	98.07	0.17	92.39	92.63	75.71	55.43	77.09	50.67	95.83	93.90	97.21	59.31	80.22	91.94	90.87	0.17	95.97
3. Ground	(650/0)	(0/479)	77.24	0.00	0.00	0.00	70.98	0.00	97.08	96.66	92.69	96.24	0.00	0.00	100.00	97.29	93.94	2.08	99.38	92.48	99.16	93.11
4. Roads	(443/0)	(0/3174)	61.06	67.77	38.31	51.89	74.04	68.46	76.97	11.12	40.45	68.62	65.91	56.62	20.60	42.03	4.47	69.65	0.41	67.23	81.60	88.03
OA			72.94	89.04	78.81	86.85	75.37	89.15	43.01	77.33	78.83	89.17	80.91	87.39	42.56	71.14	73.27	82.53	62.49	79.13	67.79	**97.14**
AA			59.62	65.08	51.96	62.49	61.30	65.21	69.81	70.84	71.87	85.44	54.15	63.11	62.24	76.99	62.48	60.26	65.34	82.37	64.79	**94.21**
Kappa			0.5448	0.8042	0.6266	0.7661	0.5827	0.8061	0.3395	0.6145	0.6377	0.8157	0.6649	0.7760	0.2844	0.5341	0.3989	0.6761	0.4275	0.6391	0.4570	**0.9432**

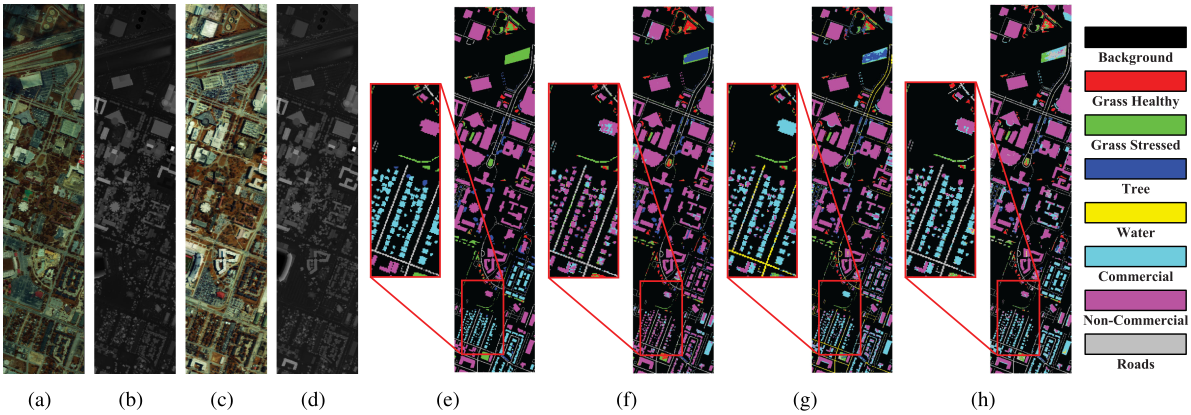


Fig. 8. Houston13–Houston18 cross-domain classification maps. (a) Source HSI (Rband:25, Gband:14, Bband:6). (b) Source LiDAR. (c) Target HSI (Rband:25, Gband:14, Bband:6). (d) Target LiDAR. (e) Ground truth. (f) SVM [13] (74.48%). (g) GAN [37] (73.37%). (h) FrF-SSDA(79.45%).

the GAN and FrF-SSDA improve the classification performance significantly. However, the long-range dependencies are still not adequately captured. In addition, the lack of ability to extract sequential spectral information results in the loss of spectral information, which limited the DA performance.

The proposed FrF-SSDA generalizes GAN by combining GRU, which maintains complete spatial information and acquires long-distance nonlocal dependencies between spectra. With clearer spatial location information and more detailed sequential spectral information, the proposed FrF-SSDA reduces spectral shift and aligns source and target domains at feature level. In the cross-temporal task, the environmental conditions vary, and part of the spatial distribution changes. For land covers with similar spectral features in Table II, e.g., commercial and noncommercial areas, the FrF-SSDA obtains promising accuracies 84.89% and 83.52%. Although the classification of these spectrally similar classes is much more difficult due to the presence of disparities between source and target domains, the proposed method performs well for most land covers. Furthermore, even in the cross-scene classification task with completely different land covers, the proposed FrF-SSDA performs best among the competitive methods.

1) Robustness to Number of Training Samples: A robust method is capable of adapting to various conditions. In real cross-domain HSI classification tasks, the training set in source scenes is difficult to obtain, which results in the small sample size problem. To verify the robustness of proposed FrF-SSDA to training sample size, we test the three groups of cross-domain classification experiments using different numbers of training samples, as shown in Fig. 11. In the experiment, the per-class numbers of training samples are set to 50–300. With a small sample size in the source scene, the proposed FrF-SSDA reflects the best performance among all the compared methods. In Fig. 11(c), using 50 training pixels, the proposed method still

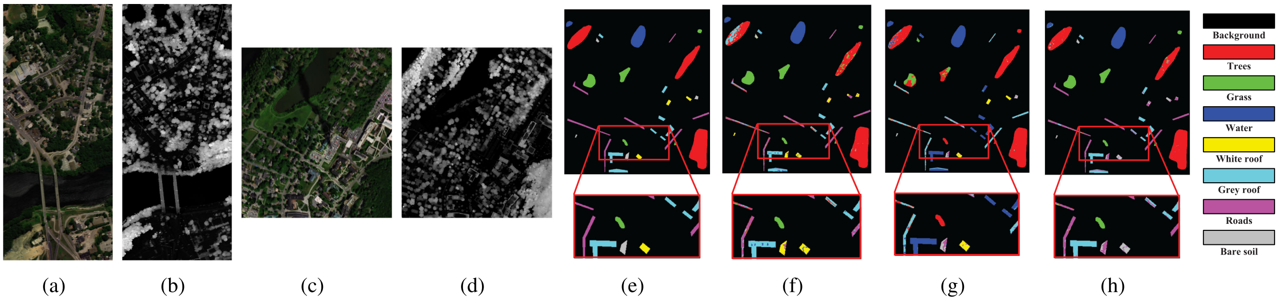


Fig. 9. Nashua–Hanover cross-domain classification maps. (a) Source HSI (Rband:54, Gband:33, Bband:15). (b) Source LiDAR. (c) Target HSI (Rband:54, Gband:33, Bband:15). (d) Target LiDAR. (e) Ground truth. (f) SVM [13] (86.59%). (g) GAN [37] (86.56%). (h) FrF-SSDA(93.16%).

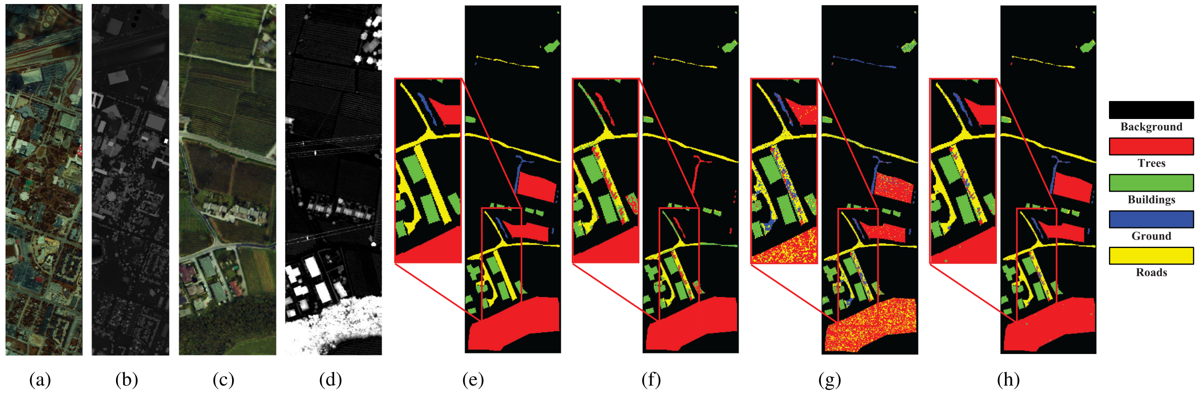


Fig. 10. Houston13–Trento cross-domain classification maps. (a) Source HSI (Rband:25, Gband:14, Bband:6). (b) Source LiDAR. (c) Target HSI (Rband:25, Gband:14, Bband:6). (d) Target LiDAR. (e) Ground truth. (f) SVM [13] (89.04%). (g) GAN [37] (79.13%). (h) FrF-SSDA(97.14%).

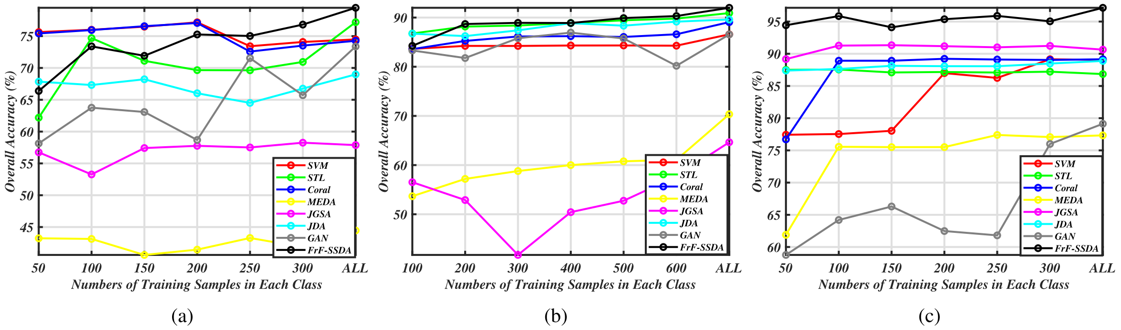


Fig. 11. Classification performance of the proposed FrF-SSDA using different number of training samples. (a) Train on Houston13 and test on Houston18. (b) Train on Nashua and test on Hanover. (c) Train on Houston13 and test on Trento.

provides 95% OA while those of other methods are all below 90%.

2) Robustness to Noise: Actual multisensor remote sensing datasets often suffer from various degradation, noise effects, or variabilities in the process of imaging, which reduce the classification accuracy [52]. To verify the robustness of FrF-SSDA to the noise that may exist in practice and compare it with other comparison methods, we studied the impact of different degrees of data noise on the classification results, as shown in Fig. 12. In the experiment, signal-dependent Gaussian white noise with different degrees of zero means is added satisfying the following model:

$$E' = E(1 + \sigma G) \tag{7}$$

where E and E' are the original and noisy datasets, respectively, and G denotes the Gaussian white noise, with the power of noise is 0 dBW. σ denoting noisy level that varies from 0.1 to 0.7. In Fig. 12, the FrF-SSDA shows robust classification performance under the influence of different noise intensities. Even if affected by a strong noise level of 0.7, the FrF-SSDA still obtains an OA of more than 80% on the Nashua–Hanover dataset.

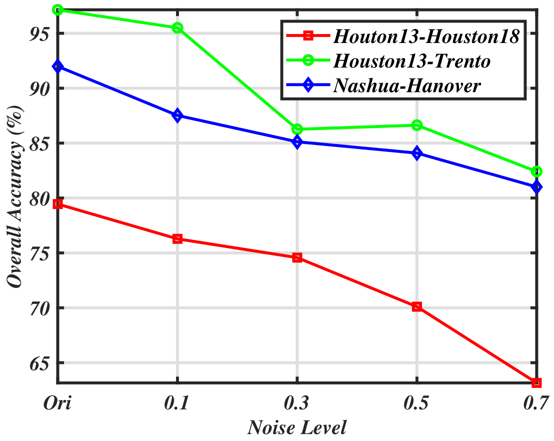


Fig. 12. Classification performance under different noise level.

TABLE V
MODEL SIZES(M)

	Total params (M)
GAN [37]	1.53
DSAN [50]	24.52
TsTNet [51]	7.78
FrF-SSDA	2.46

TABLE VI
ELAPSED TIME (SECONDS) OF METHODS

Methods	Source	Houston13	Nashua	Houston13
	Target	Houston18	Hanover	Trento
SVM	Train	3.12	20.64	0.37
	Test	3.22	15.32	0.44
STL	Train	9.55	60.54	4.97
	Test	9.04	63.41	5.02
Coral	Train	3.35	20.87	0.44
	Test	3.43	19.64	0.50
MEDA	Train	450.45	1128.55	237.95
	Test	3.59	20.49	3.58
JGSA	Train	80.96	420.21	78.31
	Test	12.76	40.89	10.95
JDA	Train	18.63	99.74	15.82
	Test	10.20	53.27	9.78
GAN	Train	92.56	451.35	61.03
	Test	1.67	4.01	0.88
FrF-SSDA	Train	112.49	657.82	74.74
	Test	1.77	4.74	0.94

E. Computational Cost Analysis

To evaluate the computational cost of the proposed FrF-SSDA and state-of-the-art (SOTA) deep learning methods, both the model sizes and running time are analyzed. As listed in Table V, the number of parameters is shown to evaluate the model sizes. SOTA methods with deep layers are relatively complicated, costing more computational resources compared with the proposed FrF-SSDA network. Further, replacing the convolutional layer in GAN with GRU layers can achieve higher accuracy.

The training and testing costs of the competitive methods are listed in Table VI to analyze the computational cost of the proposed method. The same hardware and software configurations are used to compare methods. As listed in Table VI, the training process of the proposed FrF-SSDA is more time-consuming than testing the target scene for the training procedures of pretraining on source scenes and DA. Because of the computation of large correlation matrices of MEDA and JGSA, they cost more time when the number of training samples in the source scenes increases. The baseline network GAN and proposed FrF-SSDA cost less time because only the researched samples are used for training.

IV. CONCLUSION

Aiming at reducing data-level spectral shift using multimodal data and improving cross-domain multimodal classification by feature-level DA, an FrF-SSDA network is proposed. In the proposed FrF-SSDA, first an FrDM operator is utilized to preserve spatial information of HSI firs. Then, LiDAR is leveraged and fused with HSI to reduce the spectral shift. The joint use of HSI and LiDAR not only reduces the data-level spectral shift but also utilizes the discriminative spectral feature of HSI and the robust elevation feature of LiDAR data. Then, to improve feature-level SSDA and generate domain-adaptive features, a spatial-spectral adversarial network is designed. The designed SSDA extracts discriminative features for classification while reducing the domain shift. Finally, three multimodal datasets are used for cross-domain classification experiments, and the results are compared with competitive methods indicate the effectiveness. However, the proposed method still need further research because of specific limitations. For example, the performance declines with extremely few training samples, which needs more effort. Furthermore, extending the proposed FrF-SSDA for few-shot learning of scenes with unknown land covers requires more research.

REFERENCES

[1] P. Ghamisi *et al.*, “Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art,” *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 6–39, Mar. 2019.

[2] X. Zhu, F. Cai, J. Tian, and T. K. Williams, “Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions,” *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 527.

[3] L. Zhang and L. Zhang, “Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities,” *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 2, pp. 270–294, Jun. 2022.

[4] 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.

[5] M. Schmitt and X. Zhu, “Data fusion and remote sensing: An ever-growing relationship,” *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 4, pp. 6–23, Dec. 2016.

[6] X. Zhao, R. Tao, W. Li, W. Philips, and W. Liao, “Fractional Gabor convolutional network for multisource remote sensing data classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5503818.

[7] 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.

[8] D. Hong *et al.*, “SpectralFormer: Rethinking hyperspectral image classification with transformers,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615.

[9] R. Tao, X. Zhao, W. Li, H. Li, and Q. Du, “Hyperspectral anomaly detection by fractional Fourier entropy,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.

[10] H. Wu and S. Prasad, “Semi-supervised deep learning using pseudo labels for hyperspectral image classification,” *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.

[11] J. Peng, W. Sun, L. Ma, and Q. Du, "Discriminative transfer joint matching for domain adaptation in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 972–976, Jun. 2019.

[12] H. Liu, W. Li, X. Xia, M. Zhang, C. Gao, and R. Tao, "Spectral shift mitigation for cross-scene hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6624–6638, Jun. 2021.

[13] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[14] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[15] M. Ye, Y. Qian, J. Zhou, and Y. Tang, "Dictionary learning-based feature-level domain adaptation for cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1544–1562, Mar. 2017.

[16] A. Zare and K. Ho, "Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2013.

[17] M. K. Griffin and H. K. Burke, "Compensation of hyperspectral data for atmospheric effects," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 29–54, 2003.

[18] E. J. Ientilucci, "Leveraging LiDAR data to aid in hyperspectral image target detection in the radiance domain," *Proc. SPIE*, vol. 8390, 2012, Art. no. 839007.

[19] C. Mallet and F. Bretar, "Full-waveform topographic LiDAR: State-of-the-art," *ISPRS J. Photogrammetry Remote Sens.*, vol. 64, no. 1, pp. 1–16, 2009.

[20] M. Brell, K. Segl, L. Guanter, and B. Bookhagen, "Hyperspectral and LiDAR intensity data fusion: A framework for the rigorous correction of illumination, anisotropic effects, and cross calibration," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2799–2810, May 2017.

[21] 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.

[22] 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.

[23] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep Siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, Apr. 2020.

[24] S. Li, R. Dian, L. Fang, and J. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.

[25] R. Dian, S. Li, and X. Kang, "Regularizing hyperspectral and multispectral image fusion by CNN denoiser," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1124–1135, Mar. 2021.

[26] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[27] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.

[28] X. Tong *et al.*, "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111322.

[29] L. Zhang, M. Lan, J. Zhang, and D. Tao, "Stagewise unsupervised domain adaptation with adversarial self-training for road segmentation of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5609413.

[30] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2200–2207.

[31] J. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, and P. S. Yu, "Visual domain adaptation with manifold embedded distribution alignment," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 402–410.

[32] J. Zhang, W. Li, and P. Ogunbona, "Joint geometrical and statistical alignment for visual domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1859–1867.

[33] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semisupervised 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.

[34] Y. Zhang, W. Li, R. Tao, J. Peng, Q. Du, and Z. Cai, "Cross-scene hyperspectral image classification with discriminative cooperative alignment," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9646–9660, Nov. 2021.

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

[36] J. Wang, Y. Chen, L. Hu, X. Peng, and S. Yu Philip, "Stratified transfer learning for cross-domain activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2018, pp. 1–10.

[37] S. Nirmal, V. Sowmya, and K. Soman, "Open set domain adaptation for hyperspectral image classification using generative adversarial network," in *Inventive Communication and Computational Technologies*. Berlin, Germany: Springer, 2020, pp. 819–827.

[38] X. Ma, X. Mou, J. Wang, X. Liu, J. Geng, and H. Wang, "Cross-dataset hyperspectral image classification based on adversarial domain adaptation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4179–4190, May 2021.

[39] M. Chen, L. Ma, W. Wang, and Q. Du, "Augmented associative learning-based domain adaptation for classification of hyperspectral remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6236–6248, Oct. 2020.

[40] Y. Pu, J. Zhou, and X. Yuan, "Fractional differential mask: A fractional differential-based approach for multiscale texture enhancement," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 491–511, Feb. 2010.

[41] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[42] B. Rasti, P. Ghamisi, and R. Gloaguen, "Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3997–4007, Jul. 2017.

[43] P. Chavez *et al.*, "Comparison of three different methods to merge multiresolution and multispectral data—Landsat TM and spot panchromatic," *Photogrammetric Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.

[44] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent, 6,011,875, 2000.

[45] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images. II. channel ratio and 'chromaticity' transformation techniques," *Remote Sens. Environ.*, vol. 22, no. 3, pp. 343–365, 1987.

[46] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3834–3843, Nov. 2009.

[47] J. Ma, Z. Zhou, B. Wang, L. Miao, and H. Zong, "Multi-focus image fusion using boosted random walks-based algorithm with two-scale focus maps," *Neurocomputing*, vol. 335, pp. 9–20, 2019.

[48] G. Xie, M. Wang, Z. Zhang, S. Xiang, and L. He, "Near real-time automatic sub-pixel registration of panchromatic and multispectral images for pan-sharpening," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3674.

[49] L. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

[50] 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.

[51] Y. Zhang, W. Li, M. Zhang, Y. Qu, R. Tao, and H. Qi, "Topological structure and semantic information transfer network for cross-scene hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2021.3109872.

[52] D. Hong, N. Yokoya, J. Chanussot, and 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.

Xudong Zhao (Student Member, IEEE) received the B.S. degree in science and technology in 2016 from Electronic Information Department, Beijing Institute of Technology, Beijing, China, where he is currently working toward the Ph.D. degree in information and communication engineering. He is also working toward second Ph.D. degree in computer science engineering with Ghent University, Ghent Belgium.

His research interests include fractional signal processing and multisource remote sensing.

Mengmeng Zhang received the B.S. degree in computer science and technology from the Qingdao University of Science and Technology, Qingdao, China, in 2014, and the Ph.D. degree in control science and engineering from the Beijing University of Chemical Technology, Beijing, China, in 2019.

She is a Postdoctoral Researcher with the School of Information and Electronics, Beijing Institute of Technology, Beijing. Her research interests include remote sensing image processing and pattern recognition.

Ran Tao (Senior Member, IEEE) received the B.S. degree from the Electronics Engineering Institute of PLA, Hefei, China, in 1985, and the M.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1990 and 1993, respectively, all in engineering.

He was a Senior Visiting Scholar with the University of Michigan, Ann Arbor, MI, USA, and the University of Delaware, Newark, DE, USA, in 2001 and 2016, respectively. He has been a Chief-Professor of the Creative Research Groups with the National Natural Science Foundation of China since 2014, and was a Chief-Professor of the Program for Changjiang Scholars and Innovative Research Team in University during 2010–2012. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. His current research interests include fractional Fourier transform and its applications, theory, and technology for radar and communication systems.

Dr. Tao is a Fellow of the Institute of Engineering and Technology and the Chinese Institute of Electronics. He was the recipient of the National Science Foundation of China for Distinguished Young Scholars in 2006, the Distinguished Professor of Changjiang Scholars Program in 2009, and the First Prize of Science and Technology Progress in 2006 and 2007, and the First Prize of Natural Science in 2013, both awarded by the Ministry of Education. He is currently the Vice-Chair of the IEEE China Council. He is also the Vice-Chair of the International Union of Radio Science (URSI) China Council and a member of Wireless Communication and Signal Processing Commission, URSI.

Wei Li (Senior Member, IEEE) received the B.E. degree in telecommunications engineering from Xidian University, Xi’an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-Sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent one year as a Postdoctoral Researcher with the University of California, Davis, CA, USA. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China.

Dr. Li was the recipient of the 2015 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS) for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE SIGNAL PROCESSING LETTERS, and JSTARS. He was the Guest Editor for the special issue of the *Journal of Real-Time Image Processing*, *Remote Sensing*, and JSTARS.

Wenzhi Liao (Senior Member, IEEE) received the B.Sc. degree in mathematics from Hainan Normal University, Haikou, China, in 2006, the Ph.D. degree in engineering from the South China University of Technology, Guangzhou, China, in 2012, and the second Ph.D. degree in computer science engineering from Ghent University, Ghent, Belgium, in 2012.

From 2012 to 2019, he was a Postdoctoral Research Fellow first with Ghent University and then with the Research Foundation Flanders, Vlaanderen, Belgium. From February 2020 to January 2022, he was with Sustainable Materials Management, Flemish Institute for Technological Research, Mol, Belgium. Since February 2022, he has been with Flanders Make, focusing on smart vision for Industry 4.0. He is also a Guest Professor with Ghent University. His research interests include image processing, pattern recognition, remote sensing, and material recycling, with focus on mathematical morphology, multisensor data fusion, hyperspectral image restoration, and AI for recycling.

Dr. Liao was the recipient of the Best Paper Challenge Awards in both the 2013 IEEE GRSS Data Fusion Contest and the 2014 IEEE GRSS Data Fusion Contest. He is an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Wilfried Philips (Senior Member, IEEE) was born in Aalst, Belgium, in 1966. He received the Diploma degree in electrical engineering and the Ph.D. degree in applied sciences from Ghent University, Belgium, in 1989 and 1993, respectively.

He is currently a Senior Full Professor with the Department of Telecommunications and Information Processing, Ghent University, where he heads the Image Processing and Interpretation Research Group. He also leads the activities in image processing and sensor fusion within the research institute IMEC. He is also a Co-Founder of the Senso2Me company, which provides Internet of Things solutions for elderly care. His main research interests include image and video quality improvement and estimation, real-time computer vision, and sensor data processing.