# SoDa$^2$: Single-Stage Open-Set Domain Adaptation via Decoupled Alignment for Cross-Scene Hyperspectral Image Classification

Yiwen Liu, Minghua Wang, *Member, IEEE*, Jing Yao, *Member, IEEE*, Xin Zhao, *Member, IEEE*, and Gemine Vivone, *Senior Member, IEEE*

***Abstract*—Cross-scene hyperspectral image (HSI) classification stands as a fundamental research topic in remote sensing, with extensive applications spanning various fields. Owing to the inclusion of unknown categories in the target domain and the existence of domain shift across different scenes, open-set domain adaptation (OSDA) techniques are commonly employed to address cross-scene HSI classification. However, existing OS cross-scene HSI classification methods still face two critical challenges: 1) domain shift issues arising from the direct alignment of mixed spectral–spatial features and 2) high computational costs caused by two-stage training strategies. To address these issues, this article proposes a single-stage open-set domain adaptation method based on decoupled alignment (SoDa$^2$) for cross-scene HSI classification. A contribution-aware dual-modality feature extraction is customized to disentangle the characteristics from spectral sequence signals and spatial details, selectively and adaptively enhancing discriminative features. The decoupled alignment module minimizes the maximum mean discrepancy (MMD) to independently reduce the spectral discrepancy and the spatial discrepancy between the source and target domains, extracting more fine-grained domain-invariant features. A cost-effective single-stage dual-branch framework is designed to learn MMD-constrained aligned features and constraint-free intrinsic features for adaptive distinction between known and unknown classes. This framework employs a Gaussian mixture model (GMM) to model the squared cosine similarity distribution between the two feature types, enabling OS recognition without prior knowledge of unknown classes. Extensive experiments on three groups of HSI datasets demonstrate that SoDa$^2$ outperforms state-of-the-art methods, achieving superior classification accuracy and model transferability for OS cross-scene tasks. The SoDa$^2$ code will be available at https://github.com/liuyiwen523/SoDa2**

***Index Terms*—Alignment, cross-scene, hyperspectral image (HSI) classification, open-set domain adaptation (OSDA), single-stage.**

Received 9 March 2026; revised 24 May 2026; accepted 24 June 2026. Date of publication 26 June 2026; date of current version 13 July 2026. This work was supported in part by the National Natural Science Foundation of China under Grant 62571271 and Grant 62201552, in part by the Natural Science Foundation of Tianjin under Grant 24JCQNJC01890, and in part by the Fundamental Research Funds for the Central Universities. *(Corresponding author: Minghua Wang.)*

Yiwen Liu, Minghua Wang, and Xin Zhao are with the Institute of Robotics and Automatic Information System (IRAIS), College of Artificial Intelligence, and Tianjin Key Laboratory of Intelligent Robotics (tjKLIR), Nankai University, Tianjin 300071, China (e-mail: liuyiwen@mail.nankai.edu.cn; wangminghua@nankai.edu.cn; zhaoxin@nankai.edu.cn).

Jing Yao is with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: jasonyao92@gmail.com).

Gemine Vivone is with the Institute of Methodologies for Environmental Analysis (CNR-IMAA), National Research Council, 85050 Tito, Italy (e-mail: gemine.vivone@imaa.cnr.it).

Digital Object Identifier 10.1109/TGRS.2026.3707690

## I. Introduction

HYPERSPECTRAL image (HSI), as one of the key technologies in remote sensing, integrates both spectral and spatial information and is widely applied in various fields such as precision agriculture, environmental monitoring, and mineral identification [1], [2], [3], [4], [5], [6], [7], [8], [9]. Among these applications, HSI classification serves as a core task of HSI, with the primary objective of assigning a class label to each pixel based on its spectral and spatial characteristics [10], [11], [12]. With the rapid advancement of deep learning, an increasing number of researchers have adopted methods such as convolutional neural network (CNN), transformer, and autoencoder (AE) to perform HSI classification more effectively [13], [14], [15], [16], [17], [18].

Although the above methods achieve remarkable performance, they rely on the assumption that training and testing datasets are independently and identically distributed [19]. In real-world applications, however, HSIs collected at different times or locations often exhibit distribution discrepancies due to variations in sensor parameters, imaging conditions, and atmospheric effects [20], [21]. This phenomenon is commonly referred to as the domain shift problem [22]. When a model trained on a well-labeled source domain is directly applied to a target domain with domain shift, its classification accuracy typically degrades significantly [23]. Domain adaptation (DA), as a kind of transfer learning, is an effective method to solve the above problems [24]. DA methods aim to reduce the distribution discrepancy between the source and target domains by learning domain-invariant features that are insensitive to domain variations [25]. Among them, classical DA approaches based on discrepancy minimization and adversarial learning have been widely studied and successfully applied in HSI classification [26], [27].

Most of the aforementioned DA methods focus on closed-set, where the source and target domains are assumed to share the same set of categories [28], [29]. However, in real applications, the target domain often contains unknown classes that do not exist in the source domain, leading to an open-set (OS) [30], [31]. In this context, OSDA faces two main challenges: 1) addressing the domain shift between the source and target domains and 2) accurately distinguishing known classes from unknown ones in the target domain. Currently, an increasing number of researchers have begun to explore OSDA. Saito et al. [32] proposed the open set DA by backpropagation (OSBP) model, which constructs a generator and a discriminator to address domain shift through adversarial training while distinguishing known and unknown classes. Liu et al. [33] introduced the separate to adapt (STA) model, which employs a binary classifier to separate known and unknown classes and a multiclass classifier to classify known samples, using adversarial learning to mitigate domain shift effects. Li et al. [34] proposed a causality-based framework, adjustment and alignment (ANNA), which uses a front-door adjustment module to correct biased learning in the source domain and applies decoupled causal alignment to align cross-domain distributions. Bi et al. [35] developed an OSDA method based on a weighted generative adversarial network and dynamic thresholding (WGDT), employing a domain discriminator to prevent negative transfer and a class-anchor weighting strategy to identify unknown classes effectively.

It is worth noting that most existing OSDA methods tackle the domain shift through domain adversarial strategies [36], [37], [38]. By constructing an adversarial game between the domain discriminator and the feature extractor, the overall distribution discrepancy between the source and target domains is minimized to achieve feature alignment. Traditional methods commonly concatenate or directly mix the spectral features and spatial features to form a unified fusion feature and then perform the overall adversarial alignment, that is, hybrid alignment. Qie et al. [39] constructed a spectral–spatial joint feature extraction structure to extract spatial features and spectral features, respectively, and fuse them at the feature level to extract domain-invariant features through adversarial learning. Xin et al. [25] designed a transformer-based network for extracting global spectral–spatial features and designed a domain discriminator on the mixed features to achieve the extraction of domain-invariant features. Nevertheless, as illustrated in Fig. 1, in HSI data, the spectral feature and the spatial feature are not only distributed independently across source and target domains but also contribute differently to the final classification results [40]. When the interdomain discrepancy is substantial, spectral information tends to easily distinguish classes. In contrast, when spectral features exhibit high similarity, such as asphalt and asphalt pavement, spatial detail features become more critical. Therefore, directly aligning mixed spectral–spatial features in a unified space via conventional adversarial training or similar approaches may undermine the cross-domain transferability and classification accuracy of the model.

In addition, most of the existing OSDA methods adopt a dual-classifier architecture combined with a two-stage training strategy, as shown in Fig. 2. These approaches typically involve two classifiers: one for known classes and another for unknown classes. In the first stage (the pretraining phase), the known-class classifier is trained using source domain data. In the second stage, both source and target domain data are used to train the unknown-class classifier. Although this dual-classifier, two-stage training framework has attracted considerable attention due to its strong performance, it still suffers from complex training procedures and high computational costs.

To address the aforementioned challenges, this article proposes a single-stage open-set domain adaptation method based on decoupled alignment (SoDa$^2$) for HSI classification. The main contributions are summarized as follows.

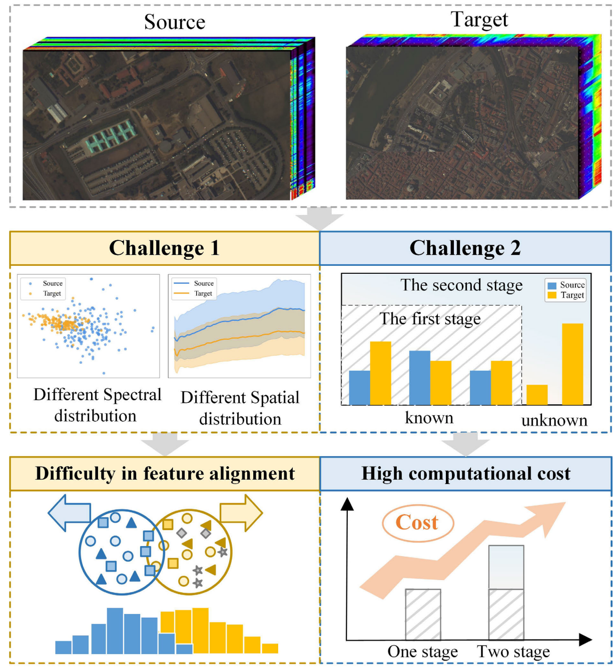



Fig. 1. Key challenges of OSDA in HSI. Direct alignment of mixed spectral–spatial features causes the domain shift problem. Moreover, the two-stage training strategy further increases computational costs.

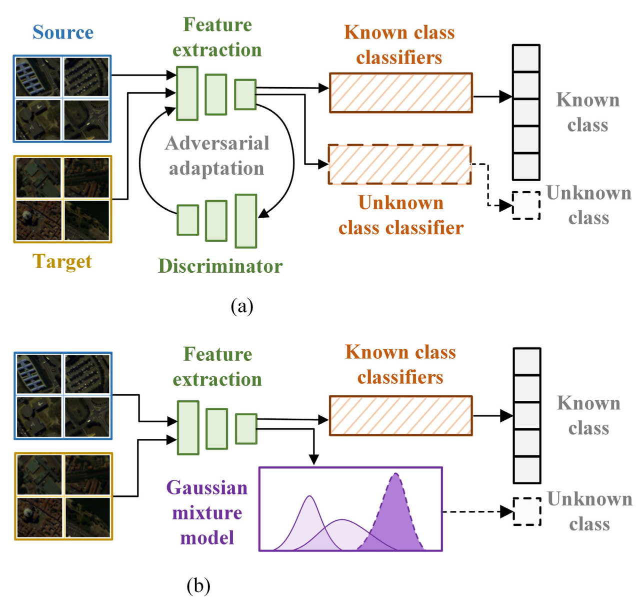



Fig. 2. Comparison of open set domain adaptive models. (a) Two-stage OSDA: First, train a known class classifier, and then train a separate unknown class classifier to recognize unknown classes. (b) Single-stage OSDA (SoDa$^2$): Train only a known class classifier, while using a GMM to identify unknown classes, achieving single-stage processing.

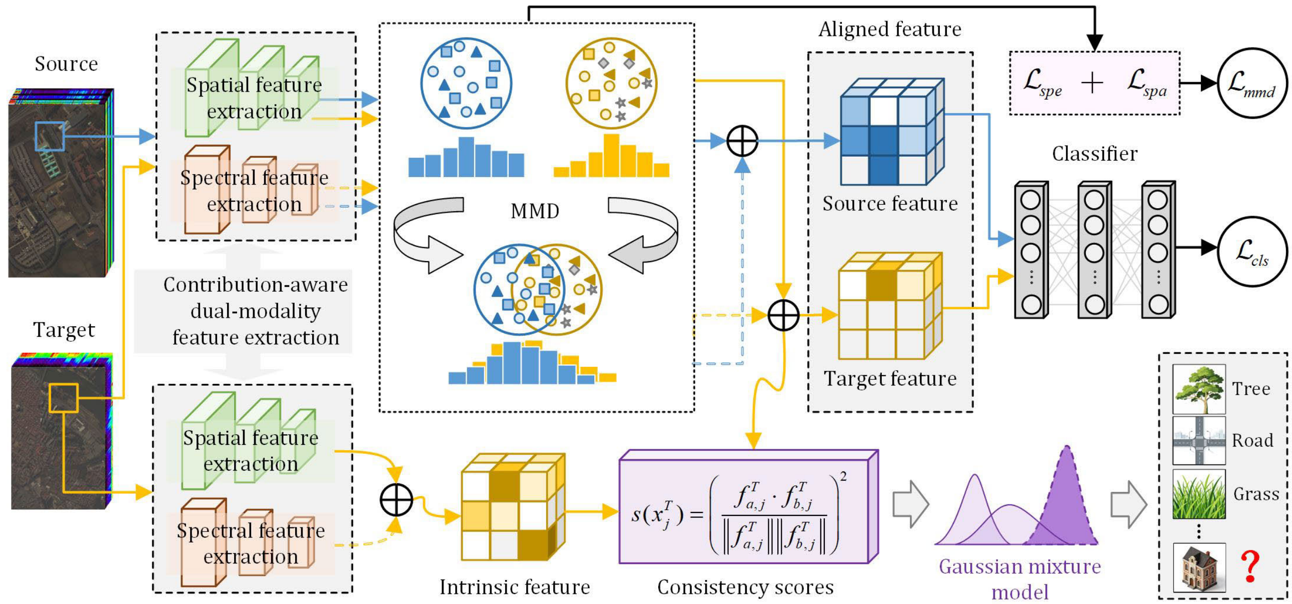



Fig. 3. Framework diagram of a SoDa$^2$. Blue lines indicate the source domain, and yellow lines represent the target domain. Solid and dashed lines after the contribution-aware dual-modality feature extraction module represent the extracted spectral and spatial features, respectively.

1) We propose a SoDa$^2$ method for HSI classification to address the challenges of domain shift and unknown classes under open-set cross-scene conditions. A contribution-aware dual-modality feature extraction and decoupled alignment are constructed to collaboratively extract independent domain-invariant features from spectral sequences and spatial details. A single-stage training framework is designed to adaptively distinguish between known and unknown classes.

2) To learn more independent domain-invariant features, we design a dedicated decoupled alignment module that separately aligns spectral and spatial features across different scenes. Specifically, the proposed method minimizes the maximum mean discrepancy (MMD) between the source and target domains along the spectral and spatial dimensions, thereby effectively alleviating domain shift in HSI classification and reducing cross-domain feature distribution discrepancies in a targeted manner.

3) We design a single-stage training framework based on dual-branch feature extraction, which adaptively distinguishes between known and unknown classes in a cost-effective manner. This framework facilitates the learning of aligned features constrained via MMD and intrinsic features free from alignment constraints. To achieve effective separation of known and unknown classes without prior knowledge of the latter, a Gaussian mixture model (GMM) is employed to model the distribution of squared cosine similarity values between the aligned and intrinsic features of each target sample.

The remainder of this article is organized as follows. Section II elaborates our proposed method SoDa$^2$. Section III reports the experiments in detail and discusses the experimental results. Finally, a brief conclusion is drawn in Section IV.

## II. Proposed Method

The source domain data are defined as $D^S = \{(X^S, Y^S)\} = \{(x_i^S, y_i^S)\}_{i=1}^{n^S}$, where $n^S$ denotes the number of samples in the source domain. The source domain data follow the distribution $P$, and the number of categories in the source domain is $C^S$. The target domain data are defined as $D^T = \{X^T\} = \{x_j^T\}_{j=1}^{n^T}$, where $n^T$ represents the number of samples in the target domain. The target domain data follow the distribution $Q$, where $Q \neq P$. The number of categories in the target domain is $C^T$, and it can be expressed as $C^T = C^S + C^{Unk}$, where $C^{Unk}$ denotes the number of unknown classes.

Our goal is to train a model that can accurately classify target domain samples belonging to the shared categories $C^S$, while simultaneously identifying target domain samples that belong to the unknown categories $C^{Unk}$. To achieve this, we design the SoDa$^2$. First, the contribution-aware dual-modality feature extraction is developed to separately learn the spectral and spatial contribution-aware features of each sample. Next, a decoupled alignment module minimizes the feature distribution discrepancy between the source and target domains for the shared categories. Finally, an open-set recognition module, based on the GMM, identifies known and unknown classes, with the known classes being further classified by a classifier to complete the final classification task. The overall framework of the model is illustrated in Fig. 3.

### A. Contribution-Aware Dual-Modality Feature Extraction

Considering that a single modality feature is insufficient to comprehensively represent the characteristics of ground objects, we design a contribution-aware dual-modality feature extraction. This module consists of two parts: spectral feature extraction and spatial feature extraction, which operate in parallel to extract both spectral and spatial information, as illustrated in Fig. 4.

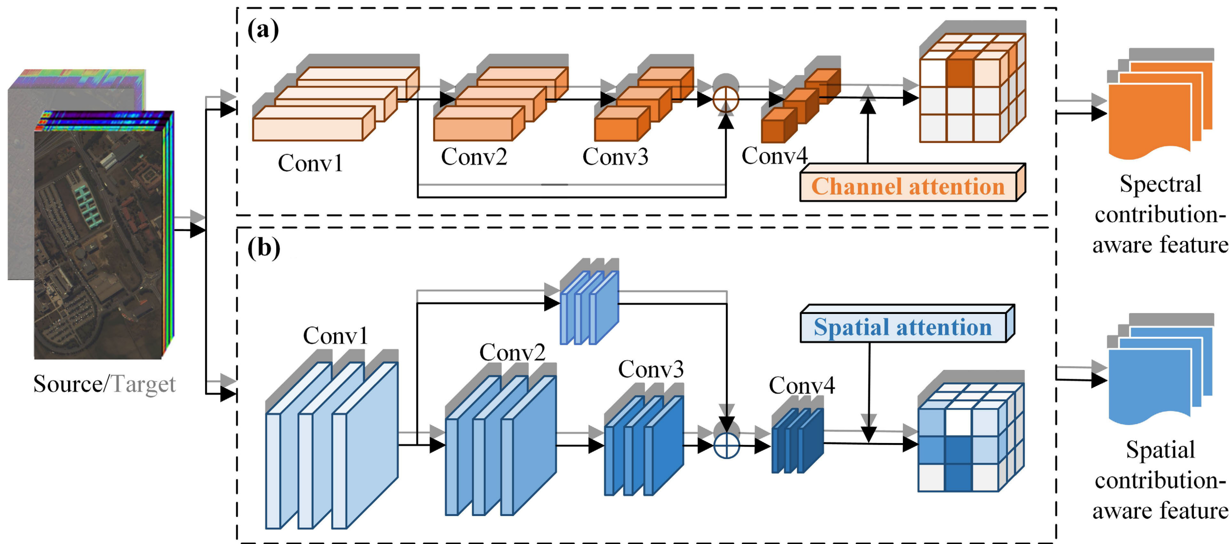



Fig. 4. Framework diagram of contribution-aware dual-modality feature extraction. The black lines represent the source domain paths, while the gray lines represent the target domain paths. (a) Spectral feature extraction: Captures the spectral features of HSI through CNN and channel attention mechanism. (b) Spatial feature extraction: Captures the spatial features of HSI through CNN and spatial attention mechanism.

In the spectral feature extraction module, the input data obtain the basic representation through the initial convolution layer, and then gradually captures higher order information through two convolution layers. To balance the fine-grained band information and the higher order spectral information, the model builds cross-layer residual connections between the output of the first and third convolution layers. After fusing the features from these two layers, the result is fed into the final convolutional layer to obtain the spectral features, which can be expressed as follows:

$$F_{spe} = H_{fuse}\left(Conv(X), Conv^3(X)\right) \tag{1}$$

where $X$ represents the input data, $F_{spe}$ represents the extracted spectral features, $H_{fuse}$ denotes the residual fusion operation, $Conv(\cdot)$ represents the convolution operation, and $Conv^3$ indicates three successive convolutional layers.

In the spatial feature extraction module, the model extracts the spatial information through three layers of 3-D convolutional layers to fully capture the spatial structure and local texture features. To enhance the compensatory effect of shallow spatial information on deep features, the features obtained from the first layer of convolution are used as the residual branch. After additional convolutional transformation, they are fused with the output of the third layer of convolution to ensure that the shallow local details are retained in the deep spatial features. Subsequently, this residual branch will be combined with the output of the third convolutional layer to ensure that the shallow local details are retained in the deep spatial features. The fused spatial features are further processed by the fourth convolutional layer to obtain the final spatial representation, which can be expressed as follows:

$$F_{spa} = H_{fuse}\left(Conv^2(X), Conv^3(X)\right) \tag{2}$$

where $F_{spa}$ denotes the extracted spatial features, and $Conv^2(X)$ represents the residual features obtained after two convolutional operations.

HSI contains massive amounts of information, including a certain portion of redundant data. To select the spectral bands and spatial details that contribute significantly to the classification task, this model introduces an attention mechanism to extract contribution-aware features. The spectral features are processed through a channel attention module, which adaptively reweights each channel according to its importance, thereby emphasizing the most informative spectral information. The spatial features are processed through a spatial attention module, which assigns higher weights to important spatial locations to strengthen the representation of key regions. The spectral and spatial contribution-aware feature after attention enhancement can be expressed as follows:

$$\tilde{F}_{spe} = A_c\left(F_{spe}\right), \tilde{F}_{spa} = A_s\left(F_{spa}\right) \tag{3}$$

where $\tilde{F}_{spe}$ represents the spectral contribution-aware feature of the channel attention mechanism $A_c(\cdot)$ and $\tilde{F}_{spa}$ represents the spatial contribution-aware features after the spatial attention mechanism $A_s(\cdot)$.

### B. Decoupled Alignment

To overcome performance degradation caused by distribution discrepancies between the source and target domains, we design a decoupled alignment module. The core idea of this module is to independently align spectral and spatial contribution-aware features at the feature level, thereby enhancing the generalization ability and classification accuracy of the model on target-domain samples. Traditional adversarial DA methods typically align features in a coarse-grained manner. Such global alignment strategies struggle to simultaneously accommodate the heterogeneity and independently distributed nature of spectral and spatial modalities, which may lead to cross-modal feature confusion. Therefore, our module adopts a discrepancy-based DA approach to reduce feature distribution differences.

Common discrepancy-based methods include Kullback–Leibler (KL) divergence [41], MMD [42], and correlation alignment (CORAL) [43]. KL divergence relies on accurate estimation of probability density functions for both domains, which is difficult to achieve in high-dimensional feature spaces. CORAL aligns distributions by matching second-order statistics, yet linear second-order metrics are often insufficient for describing the complex nonlinear properties of hyperspectral features. In contrast, MMD employs kernel mappings within a reproducing kernel Hilbert space (RKHS) to flexibly capture nonlinear high-dimensional distributions. This makes MMD particularly well-suited for cross-domain alignment in hyperspectral tasks. Accordingly, this study imposes separate MMD constraints on the spectral and spatial contribution-aware features to reduce the distribution discrepancy between the source and target domains within their respective feature subspaces. The decoupled alignment loss for spectral contribution-aware features is formulated as follows:

$$\mathcal{L}_{spe} = \left\| \frac{1}{|\tilde{F}^S_{spe}|} \sum_{f^S \in \tilde{F}^S_{spe}} \phi\left(f^S\right) - \frac{1}{|\tilde{F}^T_{spe}|} \sum_{f^T \in \tilde{F}^T_{spe}} \phi\left(f^T\right) \right\|^2_H \tag{4}$$

where $\phi(\cdot)$ denotes the nonlinear feature mapping into the RKHS and $H$ represents the RKHS. Similarly, the decoupled alignment loss for spatial contribution-aware features is computed as follows:

$$\mathcal{L}_{spa} = \left\| \frac{1}{|\tilde{F}^S_{spa}|} \sum_{f^S \in \tilde{F}^S_{spa}} \phi\left(f^S\right) - \frac{1}{|\tilde{F}^T_{spa}|} \sum_{f^T \in \tilde{F}^T_{spa}} \phi\left(f^T\right) \right\|^2_H. \tag{5}$$

Therefore, the overall decoupled alignment loss is defined as follows:

$$\mathcal{L}_{mmd} = \mathcal{L}_{spe} + \mathcal{L}_{spa}. \tag{6}$$

After achieving decoupled alignment, the spatial and spectral contribution-aware features are concatenated along the channel dimension to obtain fused domain-invariant features.

### C. Open Set Recognition Module

To effectively distinguish known classes from unknown classes in the target domain, we design an open-set recognition module. The core idea of this module is to construct two parallel feature-extraction branches and quantify the consistency between the representations generated for the same target sample. By measuring the semantic agreement between these two representations, the module enables reliable identification of unknown-class samples.

As illustrated in Fig. 5, the open-set recognition module consists of two branches: an aligned feature encoder and an intrinsic feature encoder. The aligned feature encoder aims to capture semantic information shared between the source and target domains for known classes. Its key mechanism is the introduction of MMD as a cross-domain alignment constraint. By minimizing the feature distribution discrepancy between domains in the RKHS, this branch is encouraged to learn domain-invariant representations. For a target-domain sample $X^T$, its aligned feature is formulated as follows:

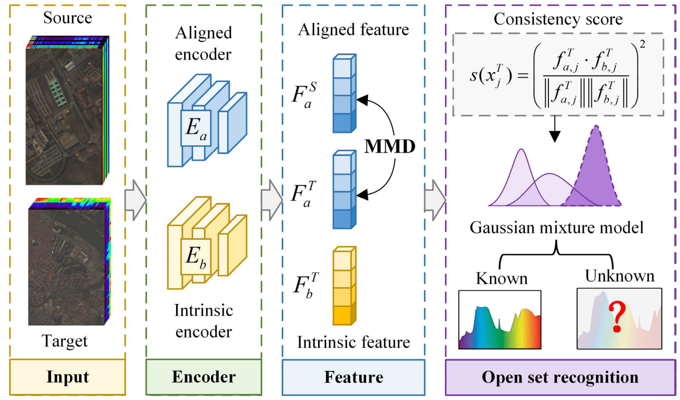



Fig. 5. Open set recognition module. The aligned encoder $E_a$ extracts domain-invariant features under MMD constraints, while the intrinsic encoder $E_b$ captures target-specific information. The consistency score is computed as the squared cosine similarity between the aligned and intrinsic features. GMM is employed to distinguish between known and unknown classes.

$$F^T_a = E_a\left(X^T; \theta_a\right) \tag{7}$$

where $F^T_a$ denotes the aligned feature of the target sample, $E_a(\cdot)$ represents the aligned feature encoder—comprising a spectral feature extractor and a spatial feature extractor, and $\theta_a$ is its parameter set.

In contrast, the intrinsic feature branch operates without any cross-domain constraints and is designed to capture target-domain intrinsic information, with special emphasis on discriminative characteristics of unknown classes. This branch relies solely on target-domain data for optimization, thus avoiding the potential suppression of target-specific information caused by domain-alignment constraints. The intrinsic feature is given by the following:

$$F^T_b = E_b\left(X^T; \theta_b\right) \tag{8}$$

where $F^T_b$ denotes the intrinsic feature, $E_b(\cdot)$ is the intrinsic feature encoder (sharing the same architecture as $E_a$ but with independent parameters), and $\theta_b$ is its parameter set.

To measure the semantic correlation of feature representations from the two branches, the module employs cosine similarity as the consistency metric. For a target sample $x^T_j$, the cosine similarity between its aligned feature $f^T_{a,j} \in F^T_a$ and intrinsic feature $f^T_{b,j} \in F^T_b$ is computed as follows:

$$sim\left(x^T_j\right) = \frac{f^T_{a,j} \cdot f^T_{b,j}}{\|f^T_{a,j}\|\|f^T_{b,j}\|}. \tag{9}$$

This value reflects the directional alignment of the two feature vectors. In an open-set setting, known-class samples are dominated by shared cross-domain semantics due to the decoupled alignment constraint, resulting in low consistency between aligned and intrinsic features. In contrast, unknown-class samples lack source-domain supervision, making intrinsic features the primary descriptors of their semantic content. Consequently, they tend to exhibit higher directional consistency with aligned features. To enhance discriminability and ensure non-negativity, we square the cosine similarity to obtain the final consistency score by the following equation:

$$s\left(x_j^T\right) = \left(\frac{f_{a,j}^T \cdot f_{b,j}^T}{\|f_{a,j}^T\|\|f_{b,j}^T\|}\right)^2. \tag{10}$$

Based on the consistency scores $s(X^T)$, this method assumes that all target-domain samples follow a mixture distribution composed of $K$ Gaussian components. These components represent two types of samples: those belonging to known classes, which generally exhibit lower consistency scores, and those belonging to unknown classes, which usually present higher scores. The distribution of the scores is modeled by a GMM, expressed as follows:

$$p(s) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(s \mid \mu_k, \sigma_k^2\right) \tag{11}$$

where $\pi_k$, $\mu_k$, and $\sigma_k$ denote the mixing coefficient, mean, and variance of the $k$th Gaussian component, respectively.

After fitting the model using the Expectation–Maximization algorithm, samples are classified by comparing the means of the Gaussian components. Samples assigned to the component with the largest mean are identified as unknown, while the remaining samples are considered known. The decision rule is written as follows:

$$\mathcal{D}(x_t) = \begin{cases} \text{Unknown}, & \text{if } \arg\max_k \gamma_k\left(s\left(x_j^T\right)\right) = \arg\max_k \mu_k \\ \text{Known}, & \text{otherwise} \end{cases} \tag{12}$$

where $\gamma_k(s(x_j^T))$ is the posterior probability that the score $s(x_j^T)$ belongs to the $k$th Gaussian component.

This probabilistic mechanism adaptively captures the distribution patterns of consistency scores and avoids the need for explicitly defining unknown-class characteristics, thereby enabling automatic open-set sample recognition without prior knowledge.

### D. Loss Function

After distinguishing known and unknown samples in the target domain through the open-set recognition module, the samples identified as belonging to known classes are further classified by a known-class classifier. This classifier is trained on the source domain data and is capable of discriminating between different known classes. It is optimized using the standard cross-entropy loss function, formulated as follows:

$$\mathcal{L}_{cls} = -\frac{1}{n^S} \sum_{i=1}^{n^S} \sum_{c=1}^{C^s} y_{i,c}^S \log\left(p\left(c|f_i^S\right)\right) \tag{13}$$

where $y_{i,c}^S$ denotes the ground-truth label indicating whether the $i$th source sample belongs to class $c$, and $p(c|f_i^S)$ represents the probability that the multimodal feature $f_i^S$ of the $i$th source sample is assigned to class $c$.

It is worth noting that the proposed method adopts a single-stage training paradigm, integrating open-set recognition and known-class classification into a unified end-to-end optimization framework. Unlike conventional two-stage approaches, all modules in our framework are jointly optimized during training. During training, each mini-batch contains both source-domain and target-domain samples. The overall loss function is defined as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \alpha \mathcal{L}_{mmd} \tag{14}$$

where $\mathcal{L}_{total}$ denotes the total loss of the proposed SoDa$^2$ algorithm and $\alpha$ is the weighting coefficient for the decoupled alignment loss.

The specific procedure for the proposed single-stage training model SoDa$^2$ is shown in Algorithm 1.

## III. Experiments

In this section, we conducted experiments on three groups of commonly used cross-scene HSI datasets to evaluate the effectiveness of the proposed SoDa$^2$ method. First, we present a detailed introduction to the datasets, model parameters, and evaluation metrics. Subsequently, a comprehensive qualitative and quantitative comparison is made between the proposed method SoDa$^2$ and the existing state-of-the-art techniques. Finally, through a series of detailed ablation experiments and parameter analyses, we verify the role of each core module and determine the optimal parameter configuration.

### A. Dataset Description

Experiments were conducted on six HSI datasets, namely Pavia University (PU), Pavia Center (PC), Houston 2013 (HU13), Houston 2018 (HU18), Ziyuan1-02D Yancheng (ZY), and GaoFen-5 Yancheng (GF). These six HSI datasets were further divided into three cross-scenario tasks: PU–PC, HU13–HU18, and ZY–GF.

1) *PU–PC:* The PU dataset was acquired by the reflective optics system imaging spectrometer (ROSIS) sensor developed by the German Aerospace Center. The imaging area covers a university in Pavia, Northern Italy, and the contiguous regions surrounding this academic site. The spatial resolution of PU is 1.3 m, with a spatial dimension of $610 \times 340$ pixels and 103 bands included. The PC dataset is also derived from the ROSIS sensor, with its imaging area located in the central region of Pavia in northern Italy. It shares the same spatial resolution of 1.3 m as the PU dataset, and has a spatial dimension of $1096 \times 715$ pixels with 102 spectral bands. During the experiment, PU was employed as the source domain dataset, while PC served as the target domain dataset. A total of 102 spectral bands shared by both PU and PC were selected as the input spectral bands. The classes of tree, asphalt, brick, bitumen, shadow, meadow, and bare soil were designated as the known categories. In addition, the tiles class is defined as the unknown class to simulate the open-set scenario.

2) *HU13–HU18:* The HU13 and HU18 datasets were derived, respectively, from the IEEE Geoscience and Remote Sensing Society data fusion contests held in 2013 and 2018. Both datasets cover the imaging area of the University of Houston campus and its surrounding regions, with a consistent spatial resolution of 1 m. Specifically, the HU13 dataset has an image size of $349 \times 1905$ pixels and encompasses 144 spectral bands, whereas the HU18 dataset features an image

**Algorithm 1** Specific Procedure of SoDa$^2$

**Input:** Source domain data $D^S = \{(X^S, Y^S)\}$, Target domain data $D^T = \{X^T\}$, Decoupled alignment loss weight $\alpha$, Number of Gaussian components $K$.

**Output:** Class probabilities of target-domain samples.

1: Feed the source-domain data $X^S$ into the aligned encoder to obtain the spectral contribution-aware features $\tilde{F}^S_{spe,a}$ and spatial contribution-aware features $\tilde{F}^S_{spa,a}$ through equation (1), equation (2), and equation (3);
2: Feed the target-domain data $X^T$ into the aligned encoder to obtain the spectral contribution-aware features $\tilde{F}^T_{spe,a}$ and spatial contribution-aware features $\tilde{F}^T_{spa,a}$ through equation (1), equation (2), and equation (3);
3: Compute the decoupled losses for spectral and spatial contribution-aware features using equation (4) and (5);
4: Obtain the total decoupled loss $\mathcal{L}_{mmd}$ as in equation (6);
5: The spectral contribution-aware feature $\tilde{F}^S_{spe,a}$ and spatial contribution-aware feature $\tilde{F}^S_{spa,a}$ are concatenated by channels to obtain the aligned feature $F^S_a$;
6: Input $F^S_a$ into the classifier to obtain class probabilities and calculate the classification loss $\mathcal{L}_{cls}$ using equation (13);
7: Calculate the overall loss $\mathcal{L}_{total}$ via equation (14) and update model parameters by minimizing this loss;
8: Feed the target-domain data $X^T$ into the intrinsic encoder to extract spectral contribution-aware features $\tilde{F}^T_{spe,b}$ and spatial contribution-aware features $\tilde{F}^T_{spa,b}$;
9: The $\tilde{F}^T_{spe,a}$ and $\tilde{F}^T_{spa,a}$ are concatenated by channels to obtain the aligned feature $F^T_a$;
10: The $\tilde{F}^T_{spe,b}$ and $\tilde{F}^T_{spa,b}$ are concatenated by channels to obtain the intrinsic feature $F^T_b$;
11: Calculate the consistency score $s$ between the $F^T_a$ and $F^T_b$ via equation (10);
12: Fit a GMM to the consistency scores according to equation (11), obtaining the means of $K$ Gaussian components;
13: According to equation (12), identify samples belonging to unknown classes as those corresponding to high-mean components, while samples associated with lower-mean components are considered known classes;
14: Feed the identified known samples into the trained classifier to obtain the final class probabilities for target-domain samples.

size of 209 × 955 pixels with 48 spectral bands. In this study, HU13 was designated as the source domain data, while HU18 was treated as the target domain data. The 48 spectral bands shared by both HU13 and HU18 were selected as the model input. In the experiment, the evergreen trees and deciduous trees in HU18 were merged into a tree class. Seven categories, namely grass healthy, grass stressed, trees, water, residential buildings (RBs), non-RBs (NRBs), and road, were defined as the known classes. The remaining 12 classes in the HU18 dataset were designated as the unknown classes.

3) *ZY–GF:* Both the ZY and GF datasets capture wetland scenes in Yancheng, Jiangsu, China, with image sizes of 1398 × 942 pixels and 1175 × 585 pixels, respectively. The ZY dataset is acquired by the Ziyuan1-02D satellite, while the GF dataset is collected by the GaoFen-5 satellite. The spatial resolution of both datasets is 30 m, and they both contain 147 spectral bands. In the experiment, the ZY dataset was designated as the source domain, and the GF dataset as the target domain. Architecture, sea, and offshore water are identified as the known classes shared by the ZY and GF datasets, with the other four categories exclusive to the GF dataset being assigned to the unknown class set.

## B. Setup

1) *Implementation Details:* All experiments are conducted using Python 3.8.20 with PyTorch 1.12.0. The computing environment is configured with CUDA 11.6, and all models are trained on an NVIDIA RTX A6000 GPU. During training, the stochastic gradient descent (SGD) optimizer was used, with a learning rate of 0.001, a weight decay of $1e^{-3}$, and a momentum of 0.9. The number of Gaussian mixture components in the GMM is set to K = 2. For a fair comparison, all baseline methods are implemented under the same experimental environment and adopt the identical hyperparameter settings, including the same optimizer, learning rate, weight decay, momentum, batch size, and training epochs, ensuring consistent experimental conditions across all methods.

2) *Metrics:* To evaluate the performance of different methods, we adopt three metrics: the mean accuracy of known classes (OS*), the accuracy of unknown classes (UNK), and their harmonic mean (HOS). The corresponding formulations are given as follows:

$$\begin{aligned} \mathrm{OS}^* &= \frac{1}{|C^s|}\sum_{c\in C^s}\mathrm{Acc}_c \\ \mathrm{UNK} &= \mathrm{Acc}_u \\ \mathrm{HOS} &= \frac{2 \times \mathrm{OS}^* \times \mathrm{UNK}}{\mathrm{OS}^* + \mathrm{UNK}} \end{aligned} \tag{15}$$

where $\mathrm{Acc}_c$ represents the classification accuracy on known classes and $\mathrm{Acc}_u$ represents the classification accuracy on unknown classes.

## C. Experimental Results

To verify the effectiveness of the proposed method, we compare it with six OSDA approaches, including OSBP [32], STA [33], DAMC [44], UADAL [45], ANNA [34], MTS [28], and WGDT [35]. Tables I–III report the experimental results of these methods on the PU–PC, HU13–HU18, and ZY–GF tasks, respectively. The evaluation metrics include OS*, UNK, and HOS, where higher values indicate better performance. In the tables, the best results are highlighted in bold, while the second-best results are underlined.

Table I shows the results on the PU–PC task, with visualizations in Fig. 6. In terms of overall performance, the

TABLE I
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON THE PU–PC TASK

| Method | Venue | Tree | Asphalt | Brick | Bitumen | Shadow | Meadow | Bare Soil | OS* | UNK | HOS |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OSBP [32] | ECCV'18 | 47.5 | 57.8 | 50.8 | 30.9 | 92.5 | 60.6 | 33.7 | 53.4 | 9.3 | 15.8 |
| STA [33] | CVPR'19 | 80.4 | 6.7 | 49.6 | 11.5 | 98.6 | 43.0 | 24.8 | 45.0 | 53.4 | 48.8 |
| DAMC [44] | TMM'21 | 71.2 | 51.4 | 49.2 | 16.6 | 90.0 | 62.6 | 24.2 | 52.2 | 25.2 | 34.0 |
| UADAL [45] | NeurIPS'22 | 91.3 | 77.5 | 44.9 | 12.1 | 99.8 | **77.3** | 3.5 | 58.1 | 83.9 | 68.6 |
| ANNA [34] | CVPR'23 | 57.2 | 75.9 | 56.9 | 19.4 | 98.3 | 21.2 | 41.2 | 52.9 | 86.1 | 65.5 |
| MTS [28] | TCSVT'24 | **98.6** | 63.2 | **67.1** | **67.8** | 98.5 | 4.1 | 32.0 | 61.6 | 24.8 | 35.4 |
| WGDT [35] | TGRS'25 | 96.1 | 81.0 | 47.1 | 47.0 | **100.0** | 64.9 | 46.4 | 68.9 | 86.3 | 76.6 |
| SoDa$^2$ | **Ours** | 97.1 | **94.0** | 66.0 | 23.5 | 90.2 | 75.4 | **47.0** | **70.5** | **94.3** | **80.7** |

TABLE II
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON THE HU13–HU18 TASK

| Method | Venue | Grass healthy | Grass stressed | Trees | Water | RB | NRB | Road | **OS*** | **UNK** | **HOS** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OSBP [32] | ECCV'18 | 76.4 | 38.1 | 48.3 | 64.7 | 14.8 | 37.2 | 55.4 | 47.8 | 7.0 | 12.2 |
| STA [33] | CVPR'19 | 92.7 | 51.4 | **63.6** | 30.5 | **85.0** | 11.2 | 12.7 | 49.6 | 57.4 | 53.2 |
| DAMC [44] | TMM'21 | 66.5 | 49.8 | 18.3 | 70.3 | 44.2 | 11.2 | 30.1 | 41.5 | 27.5 | 33.1 |
| UADAL [45] | NeurIPS'22 | 91.9 | 57.9 | 46.0 | 1.9 | 25.8 | 9.0 | 5.7 | 34.0 | **75.0** | 46.8 |
| ANNA [34] | CVPR'23 | 91.3 | 29.6 | 26.1 | 24.4 | 8.6 | 7.8 | **62.9** | 35.8 | 35.0 | 35.4 |
| MTS [28] | TCSVT'24 | 94.1 | 38.2 | 46.8 | 10.2 | 62.8 | 22.5 | 49.7 | 46.3 | 25.2 | 32.6 |
| WGDT [35] | TGRS'25 | 74.5 | 34.3 | 48.1 | **90.6** | 65.7 | **64.7** | 3.7 | 54.4 | 70.2 | 61.4 |
| SoDa$^2$ | **Ours** | **98.5** | **65.5** | 60.2 | 62.0 | **85.0** | 41.3 | 0.5 | **59.0** | 72.3 | **65.0** |

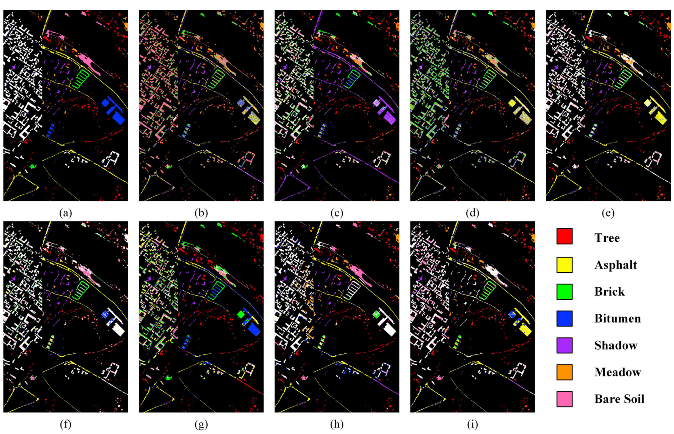



Fig. 6. Visualization of classification results for the PU–PC task. (a) Ground-truth. (b) OSBP. (c) STA. (d) DAMC. (e) UADAL. (f) ANNA. (g) MTS. (h) WGDT. (i) SoDa$^2$.

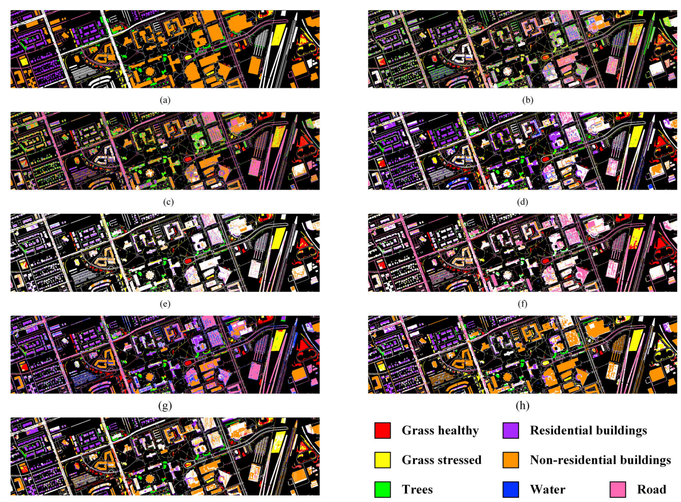



Fig. 7. Visualization of classification results for the HU13–HU18 task. (a) Ground-truth. (b) OSBP. (c) STA. (d) DAMC. (e) UADAL. (f) ANNA. (g) MTS. (h) WGDT. (i) SoDa$^2$.

proposed SoDa$^2$ achieves the best results, with an HOS of 80.7%, which is significantly higher than those of the competing methods. For unknown-class recognition, SoDa$^2$ attains an UNK metric of 94.3%, outperforming all comparison methods. This demonstrates its strong ability to identify unknown categories under open-set scenarios. For known-class classification, SoDa$^2$ achieves an OS* of 1.6% higher than WGDT and shows more competitive performance across multiple categories. Among the compared methods, WGDT achieves the second-best overall performance. UADAL and ANNA perform relatively well in unknown-class recognition but show lower accuracy in known-class classification. In contrast, MTS exhibits the opposite trend, with poor performance in unknown-class recognition. OSBP and DAMC show weaker ability to identify unknown classes, resulting in lower overall metrics. Overall, SoDa$^2$ achieves leading performance in both known- and unknown-class recognition. It maintains a favorable balance between the two aspects, which validates the effectiveness of the proposed method.

Table II presents the experimental results of different methods on the HU13–HU18 task. The visualizations of all methods are shown in Fig. 7. Considering all

TABLE III
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON THE ZY–GF TASK

| Method | Venue | Architecture | Sea | Offshore water | OS* | UNK | HOS |
|---|---|---|---|---|---|---|---|
| OSBP [32] | ECCV'18 | 65.8 | 56.3 | 73.4 | 65.2 | 15.5 | 25.0 |
| STA [33] | CVPR'19 | 6.4 | **100.0** | 52.4 | 52.9 | 68.9 | 59.9 |
| DAMC [44] | TMM'21 | 67.5 | 75.2 | 65.3 | 69.3 | 15.8 | 25.8 |
| UADAL [45] | NeurIPS'22 | 74.2 | 5.5 | 82.6 | 54.1 | 57.0 | 55.5 |
| ANNA [34] | CVPR'23 | 66.7 | 99.8 | 88.7 | 85.1 | 20.0 | 32.4 |
| MTS [28] | TCSVT'24 | 34.2 | 91.2 | 91.4 | 72.3 | 64.5 | 68.1 |
| WGDT [35] | TGRS'25 | **83.7** | 99.6 | 81.2 | 88.2 | 90.2 | 89.2 |
| SoDa$^2$ | **Ours** | 79.7 | 99.5 | **97.0** | **92.1** | **97.4** | **94.7** |

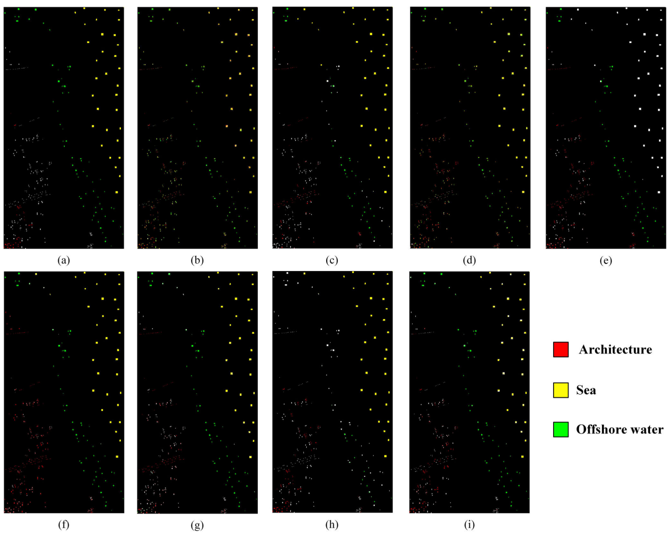



Fig. 8. Visualization of classification results for the ZY–GF task. (a) Groundtruth. (b) OSBP. (c) STA. (d) DAMC. (e) UADAL. (f) ANNA. (g) MTS. (h) WGDT. (i) SoDa$^2$.

evaluation metrics, the proposed SoDa$^2$ method exhibits the best overall classification performance. SoDa$^2$ obtains an HOS of 65.0%, outperforming all compared methods. This demonstrates its strong capability in both known-class classification and unknown-class recognition tasks. For the known-class classification task, SoDa$^2$ achieves the highest OS* value of 59.0% among all methods. From a class-wise perspective, SoDa$^2$ attains the best classification accuracy on the Grass healthy, Grass stressed, and Residential buildings categories. Moreover, SoDa$^2$ also maintains strong performance on most remaining categories. In many cases, it ranks as the second-best method. This indicates its stability in known-class recognition. Although the UNK metric of SoDa$^2$ is 2.7% lower than that of UADAL, the difference is relatively small. Meanwhile, SoDa$^2$ achieves higher accuracy on known classes. When both aspects are considered, SoDa$^2$ still shows superior overall performance. In contrast, while UADAL attains the highest accuracy in unknown-class recognition, its performance on known-class classification is relatively limited. WGDT shows stable performance in both OS* and UNK metrics, but still underperforms SoDa$^2$. Other methods, including MTS, STA, ANNA, DAMC, and OSBP, achieve competitive results in some categories. However, their overall performance remains inferior to that of SoDa$^2$. Overall, the results demonstrate that the proposed method is highly competitive across multiple evaluation metrics. In particular, it significantly improves unknown-class discrimination while preserving strong known-class recognition ability.

Table III reports the classification performance of different methods on the ZY–GF task. The corresponding visualization results are shown in Fig. 8. Overall, SoDa$^2$ exhibits the optimal performance in terms of the comprehensive evaluation metric, with an HOS of 94.7%, which is the highest among all competing methods. For known-class recognition, SoDa$^2$ attains the highest OS* value of 92.1%, indicating its discriminative ability for known categories. In the unknown-class recognition task, SoDa$^2$ also achieves the best performance, demonstrating its capability to identify unknown classes. Among the compared methods, WGDT performs well in both known-class classification and unknown-class recognition, achieving the second-best HOS value. ANNA shows reasonable accuracy on known classes but suffers from low accuracy in unknown-class recognition, which limits its overall performance. The STA method achieves a recognition accuracy of 100% on the Sea category. However, its performance on the architecture category is relatively poor. MTS also performs poorly on the architecture category. Other methods, such as UADAL, DAMC, and OSBP, perform poorly on at least one task, which results in inferior overall performance. In summary, the proposed SoDa$^2$ method not only achieves the highest accuracy in both known and unknown class recognition on the ZY–GF task but also maintains an excellent balance between the two, further validating its effectiveness in cross-source open-set HSI classification scenarios.

To intuitively compare the performance of different methods on the HOS metric across various tasks, we visualize the experimental results using radar charts, as shown in Fig. 9. The polar axis represents the value of the HOS metric, ranging from 0 to 100, where values closer to the outer circle indicate better performance. Each radial direction corresponds to one comparison method. Specifically, Fig. 9(a) shows the results on the PU–PC task, Fig. 9(b) on the HU13–HU18 task, and Fig. 9(c) on the ZY–GF task. As can be observed from the radar charts, the proposed SoDa$^2$ achieves the highest HOS scores across all three tasks, which verifies its

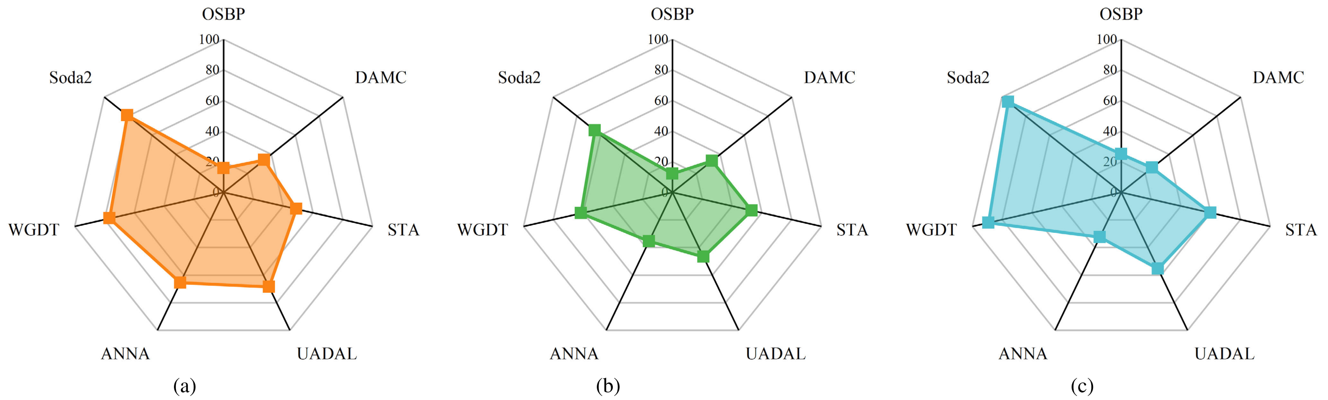



Fig. 9. Visualization of HOS using different comparison methods. The polar axis ranges from 0 to 100, where a larger value indicates better overall performance. (a) PU–PC. (b) HU13–HU18. (c) ZY–GF.

TABLE IV
RESULTS OF LOSS FUNCTION ABLATION EXPERIMENTS

| Task | $\mathcal{L}_{cls}$ | $\mathcal{L}_{spe}$ | $\mathcal{L}_{spa}$ | OS* | UNK | HOS |
|---|---|---|---|---|---|---|
| PU-PC | ✓ | | | 59.4 | 66.7 | 62.8 |
| | ✓ | ✓ | | 58.1 | 84.2 | 68.8 |
| | ✓ | | ✓ | 63.8 | 78.7 | 70.5 |
| | ✓ | ✓ | ✓ | 70.5 | 94.3 | 80.6 |
| HU13-HU18 | ✓ | | | 51.2 | 31.4 | 39.0 |
| | ✓ | ✓ | | 65.8 | 30.4 | 41.6 |
| | ✓ | | ✓ | 56.9 | 44.8 | 50.1 |
| | ✓ | ✓ | ✓ | 59.0 | 72.3 | 65.0 |
| ZY-GF | ✓ | | | 58.3 | 47.1 | 52.1 |
| | ✓ | ✓ | | 63.1 | 52.0 | 57.0 |
| | ✓ | | ✓ | 69.0 | 88.8 | 77.7 |
| | ✓ | ✓ | ✓ | 92.1 | 97.4 | 94.7 |

excellent capability in open-set recognition and its balanced overall performance.

### *D. Ablation Study*

To further investigate the contribution of each component, ablation experiments are conducted on three cross-scene tasks. We first analyze the impact of different loss terms. The overall loss function of the proposed method consists of three components: the classification loss $\mathcal{L}_{cls}$, the spectral decoupled alignment loss $\mathcal{L}_{spe}$, and the spatial decoupled alignment loss $\mathcal{L}_{spa}$. To examine the effect of each loss component on model performance, we design the following comparison settings: using only the classification loss, using the classification loss combined with the spectral decoupled alignment loss, using the classification loss combined with the spatial decoupled alignment loss, and using all three loss terms, i.e., the complete model. The results of the loss-function ablation experiments are reported in Table IV.

The experimental results indicate that, across three cross-scene HSI tasks, the proposed SoDa$^2$ model consistently achieves the highest HOS values. This demonstrates that the designed loss terms can effectively work in a complementary manner to enhance the overall category recognition capability in open-set cross-scene classification. When only the classification loss is employed, the model yields the lowest HOS, with particularly poor performance in unknown-class recognition. This is mainly attributed to the distribution discrepancy between the source and target domains, which leads to domain shift and severely limits the model's generalization ability to unknown samples. After incorporating the spectral decoupled alignment loss, both UNK and HOS exhibit stable improvements on the PU–PC and ZY–GF tasks. This observation suggests that aligning cross-domain spectral feature distributions facilitates the learning of more domain-invariant representations, thereby enhancing the discriminability of unknown classes. Similarly, adding the spatial decoupled alignment loss separately also improves model performance. For example, in the HU13–HU18 tasks, adding the spatial decoupled alignment loss significantly improves OS*, indicating that spatial structure alignment helps improve classification accuracy for known classes. When the classification loss, spectral decoupled alignment loss, and spatial decoupled alignment loss are jointly optimized, all evaluation metrics reach their best performance across the three tasks. The synergistic optimization of these loss components substantially enhances the robustness and overall performance of the model in complex cross-scene open-set classification scenarios, thereby validating the rationality and necessity of the proposed loss-function design.

Subsequently, we modified the multimodal feature extraction module and compared two different strategies for combining feature fusion and attention mechanisms. The first strategy involves fusing spatial and spectral features first and then applying a spatial-channel attention mechanism. The second strategy, which is the proposed method in this article, first enhances the two types of features separately using spatial-channel attention mechanisms, followed by weighted fusion. This experiment aims to validate the impact of the order of the attention mechanism and feature fusion operations on the efficiency of multimodal information integration and the final recognition performance. The experimental results are shown in Table V.

As can be seen from the experimental results in Table V, in the three cross-scene HSI classification tasks, the proposed method SoDa$^2$ outperforms its variant SoDa$^2$_F in terms

TABLE V
ABLATION EXPERIMENT RESULTS OF FEATURE EXTRACTION MODULE

| Metric | PU-PC | | HU13-HU18 | | ZY-GF | |
|---|---|---|---|---|---|---|
| | SoDa$^2$_F | SoDa$^2$ | SoDa$^2$_F | SoDa$^2$ | SoDa$^2$_F | SoDa$^2$ |
| OS* | 64.6 | 70.5 | 60.0 | 59.0 | 91.4 | 92.1 |
| UNK | 78.3 | 94.3 | 37.4 | 72.3 | 61.6 | 97.4 |
| HOS | 70.5 | 80.7 | 46.1 | 65.0 | 73.5 | 94.7 |

TABLE VI
EXPERIMENTAL RESULTS WITH DIFFERENT NUMBERS OF $K$

| Task | Metric | $K$ | | | |
|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 |
| PU-PC | OS* | 70.5 | 64.7 | 68.1 | 72.3 |
| | UNK | 94.3 | 71.4 | 50.5 | 28.3 |
| | HOS | 80.6 | 67.9 | 58.0 | 40.7 |
| HU13-HU18 | OS* | 59.0 | 56.0 | 57.5 | 70.5 |
| | UNK | 72.3 | 38.0 | 26.4 | 15.6 |
| | HOS | 65.0 | 45.3 | 36.2 | 25.6 |
| ZY-GF | OS* | 92.1 | 87.4 | 97.8 | 96.3 |
| | UNK | 97.4 | 80.8 | 66.1 | 51.7 |
| | HOS | 94.7 | 84.0 | 78.9 | 67.2 |

TABLE VII
EXPERIMENTAL RESULTS WITH DIFFERENT NUMBERS OF $\alpha$

| $\alpha$ | OS* | UNK | HOS | $\alpha$ | OS* | UNK | HOS |
|---|---|---|---|---|---|---|---|
| 1 | 69.8 | 75.8 | 72.6 | 7 | 65.7 | 93.3 | 77.1 |
| 2 | 65.0 | 84.7 | 73.6 | 8 | 65.2 | 95.5 | 77.5 |
| 3 | 60.9 | 91.0 | 73.0 | 9 | 66.5 | 95.9 | 78.6 |
| 4 | 68.1 | 81.0 | 74.0 | 10 | 70.5 | 94.3 | 80.6 |
| 5 | 68.1 | 82.8 | 74.7 | 11 | 67.9 | 89.9 | 77.4 |
| 6 | 63.3 | 92.4 | 75.1 | 12 | 67.2 | 84.3 | 74.8 |

TABLE VIII
TRAINING AND TESTING TIME COMPARISON

| Method | PU-PC | | HU13-HU18 | | ZY-GF | |
|---|---|---|---|---|---|---|
| | train | test | train | test | train | test |
| OSBP | 262.31 | 26.33 | 167.92 | 123.77 | 83.39 | 1.59 |
| DAMC | 661.22 | 51.86 | 417.84 | 228.21 | 178.13 | 2.65 |
| STA | 499.17 | 41.42 | 439.37 | 212.54 | 204.47 | 2.52 |
| UADAL | 2531.75 | 23.41 | 12550.67 | 154.94 | 603.07 | 1.46 |
| ANNA | 257.25 | 22.16 | 239.42 | 113.41 | 112.06 | 1.42 |
| WGDT | 258.20 | 12.38 | 243.54 | 69.74 | 97.88 | 0.91 |
| SoDa$^2$ | **140.97** | **10.44** | **132.80** | **58.93** | **49.91** | **0.87** |

of HOS index. For the PU–PC task, SoDa$^2$ improves OS* by 5.9% and UNK by 16.0% compared with SoDa$^2$_F. For the HU13–HU18 task, although SoDa$^2$_F achieves a slightly higher OS*, SoDa$^2$ obtains a substantial improvement in UNK. For the ZY–GF task, SoDa$^2$ gains 0.7% in OS* and 35.8% in UNK, leading to a remarkable 21.2% increase in HOS. That is, the method of performing attention mechanisms separately before fusion is better than the method of fusing first and then performing attention mechanisms, which verifies the effectiveness of the multimodal feature extraction module structure adopted.

## E. Parameter Analysis

To further investigate the influence of key parameters on model performance, we conduct a series of parameter analysis experiments.

1) *Analysis of Parameter K:* We first analyze the effect of the number of Gaussian mixture components $K$ in the GMM. The experiments are carried out on three cross-scene tasks. We vary $K$ within the set {2, 3, 4, 5} and evaluate the corresponding changes in OS*, UNK, and HOS on the PU–PC, HU13–HU18, and ZY–GF tasks. The experimental results are summarized in Table VI. The experimental results demonstrate that different numbers of Gaussian mixture components have distinct effects on model performance. As $K$ increases, both UNK and HOS exhibit a decreasing trend across all three tasks. When $K = 5$, the HOS value reaches its lowest level, indicating that an excessively large number of mixture components tends to bias the model toward classifying most samples as known classes. This behavior degrades the model's generalization ability to unknown samples and consequently weakens its discriminative capability for unknown categories. Therefore, to achieve a favorable balance between known- and unknown-class recognition, we set $K = 2$ in all experiments.

2) *Analysis of Parameter $\alpha$:* To investigate the influence of the decoupled alignment loss coefficient on model performance, we conduct a parameter sensitivity analysis on the PU–PC dataset with different values of $\alpha$. Specifically, $\alpha$ is varied within the range {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. The experimental results are reported in Table VII, including the three evaluation metrics OS*, UNK, and HOS. From the results, it can be observed that as $\alpha$ increases, the UNK metric shows an upward trend, indicating that strengthening the decoupled alignment loss helps the model better recognize unknown classes. In contrast, the OS* metric remains relatively stable across different values of $\alpha$, exhibiting only minor fluctuations. With respect to the HOS metric, a gradual improvement is observed as $\alpha$ increases, reflecting an overall enhancement in balanced performance. The HOS value reaches its maximum when $\alpha = 10$, corresponding to the best overall model performance. When $\alpha$ is further increased to 11 and 12, HOS slightly decreases, suggesting that an excessively large DA weight may suppress the overall performance. Therefore, we set $\alpha = 10$ in all experiments.

3) *Analysis of Complexity:* To verify the efficiency of the proposed single-stage framework, we compare the training time and testing time. All experiments are conducted on an NVIDIA RTX A6000 GPU with a consistent batch size and data settings. The results are summarized in Table VIII. As shown in the table, the proposed SoDa$^2$ achieves the shortest training and testing times across all tasks compared with competing methods. In all three cross-scene adaptation tasks, SoDa$^2$ maintains outstanding computational efficiency. Specifically, it only requires 49.91 s for model training and

10.44 s for inference on the ZY–GF task, which is remarkably faster than other competitors. $SoDa^2$ also obtains the optimal time consumption performance on the remaining two tasks.

## IV. Conclusion

This article proposes a single-stage open-set domain adaptation method with decoupled alignment ($SoDa^2$) for cross-scene HSI classification. A contribution-aware dual-modality feature extraction module is designed to jointly capture spectral and spatial contribution-aware features. To learn domain-invariant representations, we construct a decoupled alignment module. MMD is employed to independently minimize the distribution discrepancies between the source domain and target domain in spectral and spatial feature spaces, respectively. An open-set recognition module is further designed, in which unconstrained intrinsic features extracted from the target domain are compared with MMD-aligned domain-invariant features through a feature consistency measurement. Based on the resulting consistency scores, a GMM is constructed to effectively distinguish known and unknown classes. Extensive experiments conducted on three groups of HSI datasets demonstrate that the proposed $SoDa^2$ method achieves superior performance in terms of classification accuracy for both known and unknown classes. In addition, comprehensive ablation studies and parameter analyses are conducted. Nevertheless, it is observed that $SoDa^2$ achieves relatively lower accuracy on a few known classes with subtle interclass differences. This limitation will be further addressed in our future work to better capture fine-grained class distinctions.

## References


[1] Z. Zha, B. Wen, X. Yuan, J. Zhou, and C. Zhu, "Hyperspectral image denoising via nonlocal rank residual modeling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[2] W. Dong, T. Yang, J. Qu, T. Zhang, S. Xiao, and Y. Li, "Joint contextual representation model-informed interpretable network with dictionary aligning for hyperspectral and LiDAR classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 11, pp. 6804–6818, Nov. 2023.

[3] M. Wang et al., "Cross-satellite hierarchical multimodal denoising for hyperspectral imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 232, pp. 94–108, Feb. 2026.

[4] Z. Zha et al., "Nonlocal structured sparsity regularization modeling for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5510316.

[5] M. Wang, B. Shang, J. Yao, R. Zhao, and X. Zhao, "Dual-interaction spatiotemporal fusion network for agricultural remote sensing imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 20, no. 2, pp. 1–15, Mar. 2026.

[6] R. Hou et al., "Early detection of pine wilt disease by combining pigment and moisture content indices using UAV-based hyperspectral imagery," *Remote Sens.*, vol. 17, no. 11, p. 1833, May 2025.

[7] J. Li, K. Zheng, W. Liu, Z. Li, H. Yu, and L. Ni, "Model-guided coarse-to-fine fusion network for unsupervised hyperspectral image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[8] D. Peng et al., "Newly established forests dominated global carbon sequestration change induced by land cover conversions," *Nature Commun.*, vol. 16, no. 1, p. 6570, Jul. 2025.

[9] J. Hu et al., "Mapping 10-m harvested area in the major winter wheat-producing regions of China from 2018 to 2022," *Sci. Data*, vol. 11, no. 1, p. 1038, Sep. 2024.

[10] V. Kumar, R. S. Singh, M. Rambabu, and Y. Dua, "Deep learning for hyperspectral image classification: A survey," *Comput. Sci. Rev.*, vol. 53, Aug. 2024, Art. no. 100658.

[11] Y. Zhang, W. Li, W. Jia, M. Zhang, R. Tao, and S. Liang, "Cross-domain hyperspectral image classification based on bi-directional domain adaptation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 12, pp. 12038–12051, Dec. 2025.

[12] Y. Zhang, P. Duan, L. Liang, X. Kang, J. Li, and A. Plaza, "PFS3F: Probabilistic fusion of superpixel-wise and semantic-aware structural features for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 9, pp. 8723–8737, Sep. 2025.

[13] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, Jul. 2020.

[14] X. Yang, W. Cao, Y. Lu, and Y. Zhou, "Hyperspectral image transformer classification networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5528715.

[15] G. Jaiswal, R. Rani, H. Mangotra, and A. Sharma, "Integration of hyperspectral imaging and autoencoders: Benefits, applications, hyperparameter tunning and challenges," *Comput. Sci. Rev.*, vol. 50, Nov. 2023, Art. no. 100584.

[16] L. Pang, J. Yao, K. Li, J. Zhou, D. Meng, and X. Cao, "SPECIAL: Zero-shot hyperspectral image classification with CLIP," 2025, *arXiv:2501.16222*.

[17] J. Yao et al., "Semi-active convolutional neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5537915.

[18] J. Hu et al., "High-precision inversion of vegetation parameters in the AI era: Integrating hyperspectral remote sensing and deep learning," *Innov.*, vol. 6, no. 6, Jun. 2025, Art. no. 100868.

[19] C. Zhao, B. Qin, S. Feng, W. Zhu, L. Zhang, and J. Ren, "An unsupervised domain adaptation method towards multi-level features and decision boundaries for cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5546216.

[20] J. Zhang, W. Li, W. Sun, Y. Zhang, and R. Tao, "Locality robust domain adaptation for cross-scene hyperspectral image classification," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 121822.

[21] Z. Jiang et al., "Cross-domain hyperspectral image classification," *Pattern Recognit.*, vol. 168, Dec. 2025, Art. no. 111836.

[22] Y. Qi, D. Liu, J. Zhang, and Y. Zhang, "A shift reduction domain generalization network for hyperspectral image cross-domain classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025.

[23] Z. Feng, S. Tong, S. Yang, X. Zhang, and L. Jiao, "Pseudo-label-assisted subdomain adaptation for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 6, pp. 4729–4744, Jun. 2024.

[24] J. Peng, Y. Huang, W. Sun, N. Chen, Y. Ning, and Q. Du, "Domain adaptation in remote sensing image classification: A survey," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9842–9859, 2022.

[25] Z. Xin et al., "Feature disentanglement based domain adaptation network for cross-scene coastal wetland hyperspectral image classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 129, May 2024, Art. no. 103850.

[26] Z. Li et al., "Supervised contrastive learning-based unsupervised domain adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5524017.

[27] Y. Huang et al., "Cross-scene wetland mapping on hyperspectral remote sensing images using adversarial domain adaptation network," *ISPRS J. Photogramm. Remote Sens.*, vol. 203, pp. 37–54, Sep. 2023.

[28] D. Chang, A. Sain, Z. Ma, Y.-Z. Song, R. Wang, and J. Guo, "Mind the gap: Open set domain adaptation via mutual-to-separate framework," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 6, pp. 4159–4174, Jun. 2024.

[29] J. Li, L. Yang, Q. Wang, and Q. Hu, "WDAN: A weighted discriminative adversarial network with dual classifiers for fine-grained open-set domain adaptation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 9, pp. 5133–5147, Sep. 2023.

[30] J. Li, L. Yang, and Q. Hu, "Enhancing multi-source open-set domain adaptation through nearest neighbor classification with self-supervised vision transformer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2648–2662, 2024.

[31] M. Sun, R. Zhao, Z. Huang, S. Ding, and J. Liu, "IT-OSE: Exploring optimal sample size for industrial data augmentation," *IEEE Trans. Ind. Informat.*, vol. 22, no. 6, pp. 5542–5552, Jun. 2026.

[32] K. Saito, S. Yamamoto, Y. Ushiku, and T. Harada, "Open set domain adaptation by backpropagation," in *Computer Vision–(ECCV)*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Cham, Switzerland: Springer, 2018, pp. 156–171.

[33] H. Liu, Z. Cao, M. Long, J. Wang, and Q. Yang, "Separate to adapt: Open set domain adaptation via progressive separation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2922–2931.

[34] W. Li, J. Liu, B. Han, and Y. Yuan, "Adjustment and alignment for unbiased open set domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 24110–24119.

[35] K. Bi et al., "Open-set domain adaptation for hyperspectral image classification based on weighted generative adversarial networks and dynamic thresholding," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5507717.

[36] J. Zheng et al., "Open-set domain adaptation for scene classification using multi-adversarial learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 208, pp. 245–260, Feb. 2024.

[37] Y. Zhang, H. Zhang, B. Chen, J. Zheng, and H. Pan, "Integrating intrinsic information: A novel open set domain adaptation network for cross-domain fault diagnosis with multiple unknown faults," *Knowl.-Based Syst.*, vol. 299, Sep. 2024, Art. no. 112100.

[38] J. Zhang, J. Liu, L. Shi, B. Pan, and X. Xu, "An open set domain adaptation network based on adversarial learning for remote sensing image scene classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 1365–1368.

[39] Y. Qie et al., "Parallel adversarial domain adaptation for cross-dataset hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5533115.

[40] Y. Huang et al., "Two-branch attention adversarial domain adaptation network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5540813.

[41] S. Ji, Z. Zhang, S. Ying, L. Wang, X. Zhao, and Y. Gao, "Kullback–Leibler divergence metric learning," *IEEE Trans. Cybern.*, vol. 52, no. 4, pp. 2047–2058, Apr. 2022.

[42] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006.

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

[44] T. Shermin, G. Lu, S. W. Teng, M. Murshed, and F. Sohel, "Adversarial network with multiple classifiers for open set domain adaptation," *IEEE Trans. Multimedia*, vol. 23, pp. 2732–2744, 2021.

[45] J. Jang, B. Na, D. H. Shin, M. Ji, K. Song, and I.-C. Moon, "Unknown-aware domain adversarial learning for open-set domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 16755–16767.

**Yiwen Liu** received the B.E. degree in automation and the M.E. degree in control science and engineering from Taiyuan University of Technology, Taiyuan, China, in 2021 and 2025, respectively. She is currently pursuing the Ph.D. degree in intelligent science and technology with Nankai University, Tianjin, China.

Her research interests include hyperspectral image classification, deep learning, and transfer learning.

**Minghua Wang** (Member, IEEE) received the B.S. degree from the School of Automation, Harbin Institute of Technology (HIT), Harbin, China, in 2016, and the Ph.D. degree in control science and engineering from HIT in 2021.

She was a Visiting Ph.D. Student with CNRS, Grenoble INP, GIPSA-Laboratory, Université Grenoble Alpes, Grenoble, France, from 2019 to 2020. She was with the Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), from 2021 to 2023. She is currently an Associate Professor at the College of Artificial Intelligence, Nankai University, Tianjin, China. Her research interests include remote sensing image processing, noise removal, anomaly detection, machine learning, and deep learning.

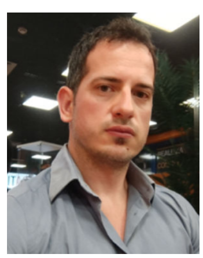
**Jing Yao** (Member, IEEE) received the Ph.D. degree in mathematics from Xi'an Jiaotong University, Xi'an, China, in 2021.

From 2019 to 2020, he was a Visiting Student at the Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany, and the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. Since 2021, he has been an Assistant Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include artificial intelligence for hyperspectral imaging and vision and language models for multimodal remote sensing.

Dr. Yao was a recipient of the IEEE Geoscience and Remote Sensing Society Highest Impact Paper Award in 2025. He has been included in Stanford's List of World's Top 2% Scientists since 2024.

**Xin Zhao** (Member, IEEE) received the B.S. degree in control theory and control engineering from Nankai University, Tianjin, China, in 1991, the M.S. degree in control theory and control engineering from Shenyang Institute of Automation, CAS, Shenyang, China, in 1994, and the Ph.D. degree in control theory and control engineering from Nankai University in 1997.

He joined Nankai University as a Faculty Member in 1997, where he is currently a Professor and the Vice Dean of the College of Artificial Intelligence. He is also with the Institute of Intelligence Technology and Robotic Systems, Shenzhen Research Institute of Nankai University, Shenzhen, China. His research interests include micromanipulators, microsystems, and mathematical biology.



**Gemine Vivone** (Senior Member, IEEE) received the B.Sc. (summa cum laude), M.Sc. (summa cum laude), and Ph.D. (Hons.) degrees in information engineering from the University of Salerno, Salerno, Italy, in 2008, 2011, and 2014, respectively.

He is currently a Senior Researcher with the National Research Council, Tito Scalo, Italy. His main research interests include image fusion, statistical signal processing, deep learning, and classification and tracking of remotely sensed images.

Dr. Vivone received the IEEE GRSS Early Career Award in 2021, the Symposium Best Paper Award from the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) in 2015, and the Best Reviewer Award from IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2017. He is also the Editor-in-Chief of IEEE Geoscience and Remote Sensing eNewsletter, an Area Editor of *Information Fusion* (Elsevier), and an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He is also an Advisory Board Member of *ISPRS Journal of Photogrammetry and Remote Sensing* and an Editorial Board Member of *Remote Sensing* (MDPI). He is listed in the World's Top 2% Scientists by Stanford University.