

Novel Class Discovery for Hyperspectral Image via Class-Relation Perceptive Distillation With Prototype-Level Clustering Prediction

Chunyan Yu[✉], Senior Member, IEEE, Xiaowen Zhao, Yulei Wang[✉], Member, IEEE,
and Xiaoqiang Lu[✉], Senior Member, IEEE

Abstract—Confronted with the increasing emergency of hyperspectral remote sensing categories in the dynamic environment, traditional classification models that depend on fixed-category labeled data encounter difficulties on new classes recognition. Novel class discovery (NCD) aims to discover unknown class-disjoint novel classes in an unlabeled dataset with the preexisting knowledge of known classes. Notably, the critical goal of NCD is to ensure recognition accuracy of known classes while identifying new ones. In this article, we propose a class-relation perceptive distillation with prototype-level clustering prediction (CRPD-PCP) network for NCD of hyperspectral image (HSI). The proposed framework comprises an initial training stage (ITS) and a NCD stage (NCDS) with two essential modules. Specifically, we present the CRPD module, which imposes a similarity constraint on the prediction of the distribution of new class data over the models of two stages. With the CRPD operated on the NCDS, our model effectively captures class relation information in spectral-spatial domain between known and novel classes of HSI to avoid forgetting old knowledge. Besides, we establish the PCP module to generate high-confidence pseudolabels for unlabeled novel classes. To be specific, we progressively cluster samples with the same spectral angular distance from the perspective of prototypes, and the self-supervised prototype-level knowledge distillation (KD) strategy in PCP facilitates effective identification of new categories. Experiments conducted on four datasets demonstrate that the CRPD-PCP model generates superior performance compared to other NCD methods for HSI. Our code will be released at <https://github.com/Chirsyey/CRPD-PCP.git>

Index Terms—Hyperspectral image (HSI), knowledge distillation (KD), novel class discovery (NCD), relationship perception.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) captures spectral information across a range of wavelengths to provide detailed information for target recognition, which supports extensive application in the fields of geological surveying, agricultural

Received 7 May 2025; revised 27 August 2025; accepted 13 September 2025. Date of publication 17 September 2025; date of current version 2 October 2025. The work of Chunyan Yu was supported by the National Natural Science Foundation of China under Grant 62471079. (Corresponding author: Yulei Wang.)

Chunyan Yu, Xiaowen Zhao, and Yulei Wang are with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: yucy@dlmu.edu.cn; bothwo@163.com; wangyulei@dlmu.edu.cn).

Xiaoqiang Lu is with the College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China.

Digital Object Identifier 10.1109/TGRS.2025.3611081

monitoring, and environment assessment [1]. HSI classification (HSIC) [2], [3], [4], [5], [6], [7], [8], [9], [10] serves as a pivotal task in HSI interpretation and aims to predict the class label of each pixel effectively. Present developments in deep learning had significantly improved the accuracy and efficiency of HSIC methods, particularly in terms of feature extraction and model training. Typically, convolutional neural networks (CNNs) [11], [12], [13] exploited the hierarchical structure of data through convolutional layers and had been widely employed for HSIC. Paul et al. [14] combined 3D-CNN, 2D-CNN with spatial pyramid pooling, which effectively extracted joint spectral-spatial features and enhanced multiscale representation. The parallel multiinput network [15] leveraged implicit spectral spatial information in HSI through four parallel convolutional branches, deriving multilevel spatial features. A three branches-model Alkhatib et al. [16] adopted a multiscale 3D-CNN with three-branch fusion and Reddy et al. [17] introduced a novel approach based on a 3D-CNN augmented with a self-attention mechanism. Recently, transformer-based approaches had achieved substantial advancements in HSIC. Among the remarkable methods, the convolution transformer mixer (CTMixer) [18] introduced a convolutional multihead self-attention framework that effectively bridged local feature extraction and global contextual modeling. By integrating wavelet transforms with downsampling, the WaveFormer method [19] achieved lossless feature map decompression, which strengthened the interaction between structural and shape features for HSIC. Zhang et al. [20] presented a multidimensional spectral transformer incorporating channelwise correlation, and a dual-branch multiscale Transformer Gong et al. [21] was employed to address local information loss in large-scale. Recently, Mamba-based solutions for HSIC have gained popularity due to the effectiveness in processing long sequence data. MorpMamba [22] combined morphological operations with a state space framework to convert HSI patches into spatial-spectral tokens with separable convolutions. In [23], a dual-branch model with local-global spatial and short-long spectral perception was proposed for HSIC, which incorporated Mamba and a gate attention unit for efficient spatial-spectral feature merging. Currently, knowledge distillation (KD) grew up as an effective method for HSIC, and demonstrated by Feng et al. [24] in a cross-domain few-shot learning framework which combined meta-knowledge

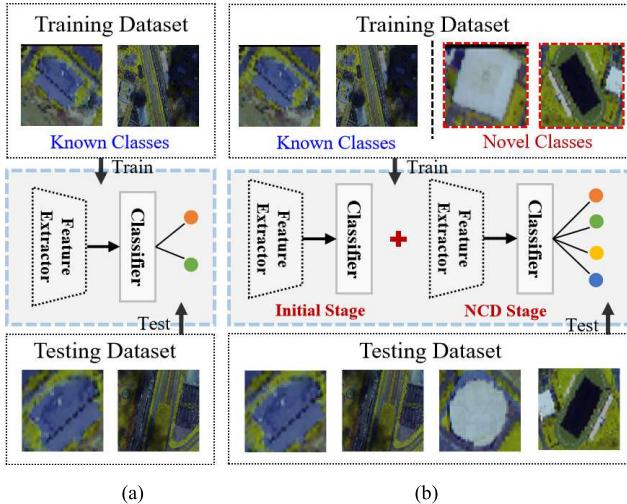


Fig. 1. Illustration of two typical classification paradigms of HSIC. (a) Traditional method for HSIC. (b) NCD for HSIC.

extraction. The dual-branch spatial and spectral transformer network [25] captured long-range dependencies in HSI with decoupled KD to overcome the inherent limitation of CNNs, which merely extracted local features through fixed receptive fields. Wu et al. [26] proposed central spectral self-distillation (CSSD), which separated the spectral information at the center of the patch and applied the pure spectral information for self-distillation.

With the development of HSI data acquisition capability of remote sensing technology, new categories of ground objects have emerged continuously, and the requirement for new class identification has become increasingly urgent for HSIC. Novel class discovery (NCD) [27] methods leverage knowledge from known categories to enhance the identification of novel classes, providing an effective approach for accurately recognizing previously unseen classes in complex HSIC scenarios. The primary goal of NCD is to accurately recognize and classify previously unseen classes, assigning them to specific categories rather than labeling them as “unknown.” To distinguish the difference between traditional method [28], [29] and NCD in the HSIC domain, we present simplified flowcharts for the two typical classification paradigms in Fig. 1. As can be observed, traditional HSIC method typically consists of a feature extractor for discriminative extraction and a classifier for label prediction of each pixel. Generally speaking, the traditional pattern is confined to classifying known and labeled categories, lacking the capacity to recognize novel categories. In contrast, NCD breaks the limitations of fixed category classification and classifies new categories that are not presented in the training stage. Notably, the known and new categories in the NCD task are disjoint. As illustrated in Fig. 1(b), the approach of NCD involves two stages, the initial stage is responsible for extracting knowledge from known categories, the second stage aims to identify and determine samples which belong to the novel classes. For NCD, the key tasks include retaining knowledge of the known classes while simultaneously learning about the unlabeled new classes. The dual focus is essential for effectively distinguishing between old and new categories.

The NCD task was put forward to identify unlabeled novel classes in NCD stage with the knowledge of known classes in initial stage. The early applications of NCD were explored in transfer learning [30], [31], [32], [33], [34]. Later methods focused on keeping the new category with the same representation space with the known samples, realizing the exploitation of the existing knowledge. Specifically, neighborhood contrastive learning (NCL) [25] introduced distinct contrastive learning objectives to capture the consistency between instances and relationship in a shared representation space. GCD [35] adopted vision transformers with contrastive representation learning for the open-world setting to address the constrained setting of NCD. Unified objective function (UNO) [36] proposed a multiview self-labeling strategy to generate pseudolabels and discovery novel classes. rKD [37] introduced relationship representation, which adaptively promoted knowledge transfer based on the semantic similarity between the novel and known classes. Some other methods tried to constrain labeled samples from unlabeled samples to achieve class separation, OpenMix [38] enhanced training by generating mixed images from labeled and unlabeled examples, producing more reliable pseudolabels than direct unlabeled predictions to prevent model overfitting. By leveraging the disjoint relationship of labeled and unlabeled data through an interclass constraint, IIC [39] achieved class separation in the embedding space. Additionally, deep transfer clustering (DTC) [40] learned temporal ensembling and consistency at different stages, enabling knowledge transfer between known and novel classes. Self-Cooperation KD (SCKD) method [41] constructed disjoint spaces for known and novel classes with self-cooperation learning, which alleviates the effect of class imbalance on NCD. Two-step prompt tuning (TSPT) approach [42] addressed class-incremental NCD by rehearsal-free and plug-and-play prompt tuning, and the clustering mechanism was employed during the NCD. In the remote sensing field, Zhou et al. [43] incorporated a novel positive pair filtering mechanism, which effectively identified reliable remote sensing sample pairs from unlabeled novel classes. Li et al. [44] integrated an ensemble of base learners with a modified projection, enhancing both base-class recognition and NCD under limited labeled data conditions. SPECIAL Pang et al. [45] proposed a novel zero-shot HSIC framework based on CLIP, which aimed to eliminate the need for manual annotations and achieve zero-shot learning. Although current approaches have advanced in general remote sensing, NCD in hyperspectral domains remains unexplored.

In this article, we present the class-relation perceptive distillation with prototype-level clustering prediction (CRPD-PCP) network for NCD of HSI. The proposed model not only captures class relationships information between novel and known classes but also enhances pseudolabel confidence through prototype clustering. Specifically, the CRPD-PCP model includes the initial training stage (ITS) and the NCD stage (NCDS). In ITS, the classification model identifies labeled known classes, and the model in NCDS processes partially labeled known classes together with unlabeled new classes. To capture categorical differences between novel and known classes in HSI, the model in NCDS integrates

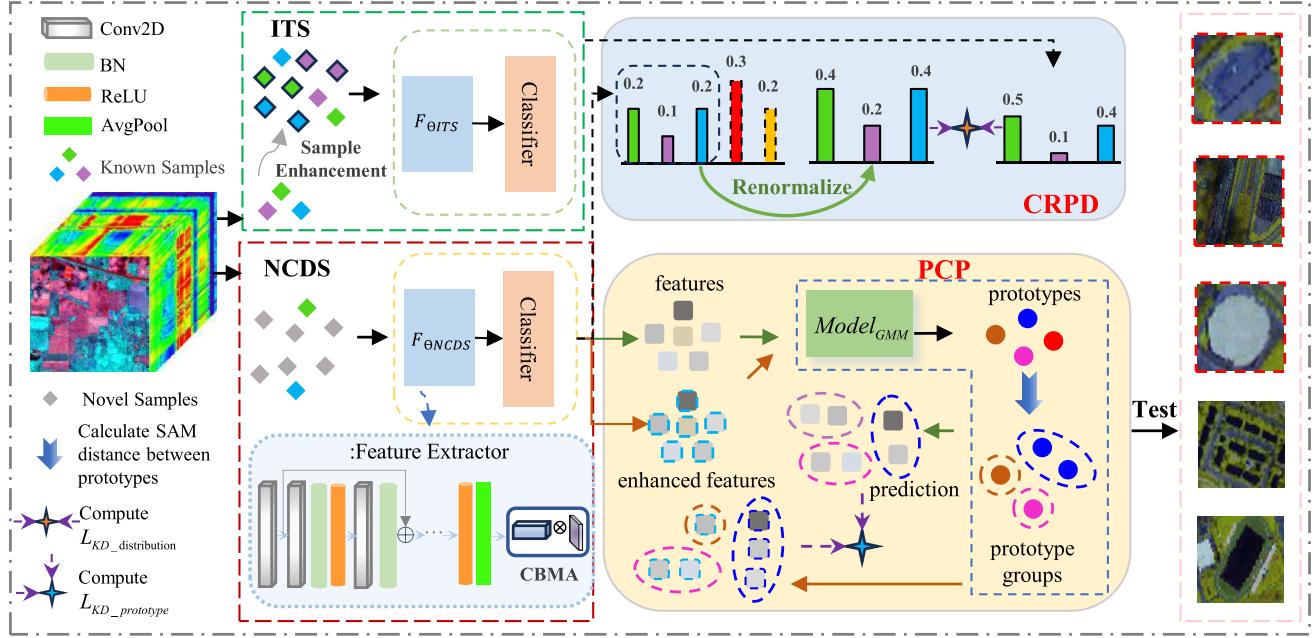


Fig. 2. Flowchart of the proposed CRPD-PCP network. Our model consists of sequential stages including the ITS and the NCDS. First, the model of ITS is trained with the selected samples of the known classes for recognition. Next, we build the model in NCDS by completing knowledge integration with the samples from the novel data. Crucially, the CRPD module is established to guarantee relationship knowledge transferring and integration between novel and known classes, and the PCP module is constructed to achieve the prediction of novel class pseudolabels for the subsequent learning.

a category relationship perceptive distillation module CRPD, which generates relationship distillation constraints for new classes through the combined prediction distributions of the models in ITS and NCDS. Additionally, NCDS includes a PCP module, which depends on clustering with prototypes and generates pseudolabels on prototype distances computed through spectral angle mapper (SAM) distance. Besides, the designed self-supervised prototype-level KD in PCP further enhances pseudolabel reliability of NCD for HSI.

The main contributions of this article are summarized as follows.

- 1) We propose the class-relation perceptive distillation with PCP network to implement new class recognition. To our knowledge, the CRPD-PCP model is the first attempt to address NCD of HSI. Facilitated with the presented classes relationship distillation constraints and PCP module to discover novel class, our model effectively preserves classification performance for known classes when identifying new classes.
- 2) We present the CRPD module to capture the spectral-spatial relationship transfer between novel and known categories. By imposing a relation constrained KD based on the prediction distribution of new class data of the ITS and NCDS, our model maintains the integrity of relative relationships while assimilating information from new classes, which are beneficial for avoiding forgetting old knowledge effectively.
- 3) We present the PCP module that explores SAM distances between new class samples through class prototype-level clustering, generating pseudo labels from distance similarities in NCDS. Besides, the self-supervised KD constraint boosts the generalization of NCD model and

enhances the pseudolabel reliability of the unlabeled data for the subsequent learning.

The remainder of this article is organized as follows. In Section II, the details of the proposed CRPD-PCP architecture are described. Section III provides the experimental results and analysis, and conclusions are drawn in Section IV.

II. PROPOSED APPROACH

As shown in Fig. 2, the overall architecture of the CRPD-PCP comprises the two stages along with two principal components: the CRPD and the PCP modules. The ITS constructs the initial model from labeled known categories and serves as the foundation for NCDS. To discover novel classes effectively, the model of NCDS processes both unlabeled new categories and selected known classes. The presented CRPD module enforces prediction consistency between ITS and NCDS models with novel classes through distillation constraint, which captures the relationship information between new and known categories. The presented PCP module generates high confidence pseudolabels for novel classes through class prototype-level KD, enhancing the discriminative representation of novel class during NCDS. Further details regarding each component are described in Sections II-A–II-D.

A. Problem Definition

The training stage for the NCD of HSI includes the ITS and NCDS, the training set of HSI data $D_{\text{train}} = \{D_{\text{ITS}}, D_{\text{NCDS}}\}$ is composed of the samples in ITS denoted as $D_{\text{ITS}} = \bigcup_{t=1}^K \{(x_i^t, y_i^t)\}_{i=1}^{N_t}$ and the samples in NCDS called as $D_{\text{NCDS}} = \bigcup_{m \in C_{\text{selected}}} \{(x_m^m, y_m^m)\}_{i=1}^{N_m} \cup D^u$, $C_{\text{selected}} \subseteq \{1, \dots, K\}$, where $D^u = \{x_j\}_{j=1}^{N_u}$. In D_{ITS} , N_t denotes the number of samples of the t th

TABLE I
BACKBONE STRUCTURE OF THE CRPD-PCP MODEL

	Layer Name	Input size	Kernel sets	Output size
Feature Extractor	Conv1	$20 \times 9 \times 9$	$3 \times 3, 64$	$64 \times 9 \times 9$
	Block_1	$64 \times 9 \times 9$	$3 \times 3, 64$ $3 \times 3, 64$	$64 \times 9 \times 9$
	Block_2	$64 \times 9 \times 9$	$3 \times 3,$ $128 \times 3 \times 3,$ 128	$128 \times 5 \times 5$
	Block_3	$128 \times 5 \times 5$	$3 \times 3,$ $256 \times 3 \times 3,$ 256	$256 \times 3 \times 3$
	Average pool	$256 \times 3 \times 3$		$256 \times 1 \times 1$
	CBAM	$256 \times 1 \times 1$		$256 \times 1 \times 1$
Classifier	Fully Connection	1×256		$1 \times K$ in ITS $1 \times (m+Q)$ in NCDS

known class, and (x_i^t, y_i^t) means the i th sample of t th class in ITS and the corresponding class label. In addition, the number of known classes in D_{ITS} are K . While in D_{NCDS} , C_{selected} represents the classes in D_{ITS} selected for NCDS, and Q denotes the number of novel classes discovered. Particularly, x_j means the j th unlabeled novel class samples named D^u and N_n denotes the number of novel samples. For NCD of HSI, the initial model employs D_{ITS} to train during ITS, and the NCDS model is established with the samples from D_{NCDS} . After two stages of training, we select some known categories from D_{ITS} and D^u randomly to form the test set D_{TEST} randomly, and test the performance of the update model in NCDS on the recognition of all the categories.

B. Cross-Class Relationship Perception Distillation Module

In our model, the CRPD module enforces distillation constraints to extract relationship information between new and known classes, providing essential support for NCD. Specifically, the shared backbone network in both ITS and NCDS extracts spatial-spectral features of HSI. By enforcing decision level distributional constraints, CRPD enhances the generalization capability for novel class recognition.

1) *Backbone Network*: The backbone network consists of a feature extractor F_θ and a unified classifier G_ϕ . In our paper, we employ ResNet combined with the convolutional block attention module (CBAM) [46] for the implementation of F_θ . Specifically, F_θ is divided into six main components, and the specific structure is shown in Table I. G_ϕ is implemented with a fully connected layer with node number varies dynamically according to the number of classification categories in the ITS and NCDS.

To train the model of ITS for K -class supervised classification, the multiclass cross-entropy loss function as shown in (1) is applied accordingly

$$L_{\text{cls}} = -\frac{1}{N_t} \sum_{t=1}^K \sum_{i=1}^{N_t} \log(F_{\theta_{\text{ITS}}}(x_i^t)) \quad (1)$$

where $F_{\theta_{\text{ITS}}}(x_i^t)$ represents the predicted probability of the t th class.

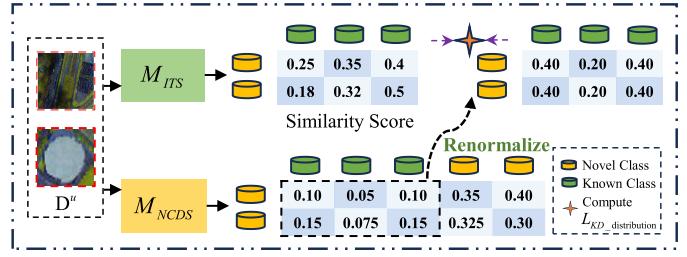


Fig. 3. Schematic representation of the CRPD module.

2) *CRPD*: The specific pattern of CRPD is shown as Fig. 3, which guides the model in effectively identifying novel classes. To enhance the ability of the NCDS model to discover unlabeled novel classes, the pretrained classification model in ITS is employed to predict novel class samples, which aims to acquire the similarity relational information between novel classes and known categories. The operation is defined as follows:

$$p_1 = \text{softmax}(M_{\text{ITS}}(x_j)/T) \quad (2)$$

where p_1 denotes the predicted similarity score of the model in ITS for novel classes, T represents the temperature parameter and M_{ITS} represents the model in training process.

The model in NCDS named as M_{NCDS} performs category similarity score prediction for novel classes following the same procedure as for known categories. Since the input data contain both known and novel classes, a re-normalization operation is applied to obtain the prediction distribution of novel classes relative to known categories in the NCDS model. The prediction is generated as follows:

$$p_2 = \text{Renormalize}(\text{softmax}(M_{\text{NCDS}}(x_j)/T)) \quad (3)$$

where p_2 denotes the output distributions produced when novel classes are processed in NCDS.

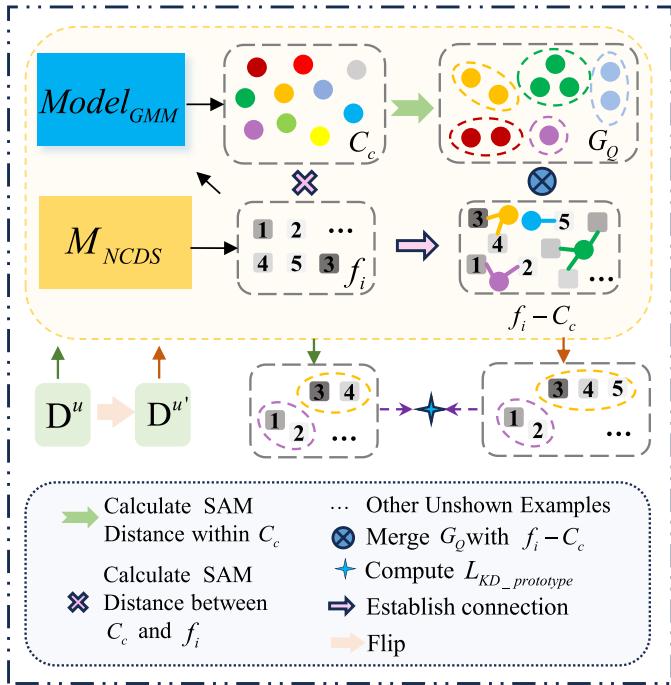
To effectively employ cross-class relational information between novel and known classes, KD constraints are activated between the distribution predictions from the ITS and the re-normalized predictions in NCDS. Through consistent alignment of relational information, the model in NCDS achieves enhanced generalization performance for novel class recognition. The loss function as shown in (4) is applied accordingly

$$L_{\text{KD_distribution}} = D_{\text{KL}}(p_1 \| p_2) = \sum_{k=1}^K p_1^k \log\left(\frac{p_1^k}{p_2^k}\right). \quad (4)$$

C. PCP Module

Inspired by prototype learning, our model presents the PCP module to overcome the challenge of predicting unlabeled novel classes labels in HSI. As illustrated in Fig. 4, the class prototype clustering mechanism generates high-confidence pseudolabels by measuring SAM distances between novel samples of HSI. Meanwhile, the self-supervised KD improves the reliability of the generated pseudolabels, which contribute to robust discovering novel classes for NCD of HSI.

First, the feature extractor in NCDS generates instance features for unlabeled novel class samples, which directly serve as



input features for Gaussian mixture models (GMMs) clustering. Capitalizing on the adaptability of GMM to hyperspectral data characteristics, a set of class prototypes representing each novel class distribution in the shared feature space are derived. Features and class prototypes are generated according to the following equations:

$$f_j = F_{\theta_{NCDS}}(x_j), \quad (j = 1, \dots, N_n) \quad (5)$$

$$C_c = \text{GMM}(f_j), \quad (c = 1, \dots, P) \quad (6)$$

where f_j represents the feature of x_j with the feature extractor and C_c represents the class prototypes and P denotes the number of class prototypes.

Next, f_j is gathered by evaluating the SAM cosine similarities with C_c , notably, all f_j that are comparable similar to a given prototype C_c are assigned to the same category, with the assignment probability computed as follows:

$$\text{SAM}(f_j, C_c) = \frac{f_j \cdot C_c}{\|f_j\| \cdot \|C_c\|} \quad (7)$$

$$P(f_j \in C_c) = \frac{\exp(\text{SAM}(f_j, C_c))}{\sum_{t=1}^j \exp(\text{SAM}(f_t, C_c))}. \quad (8)$$

In this way, SAM quantified spectral similarity between feature f_j and C_c ranging [0, 1], with higher values denoting closer resemblance.

Further, to generate high-confidence pseudolabels for unlabeled data, class prototypes divide the feature space by computing SAM distances to form distinct prototype groups. All instances linked to prototypes within the same group are assigned consistent pseudolabels, facilitating categorization of novel classes. Specific operations are defined as follows:

$$\{G_1, G_2, \dots, G_Q\} = \text{SAM}\{C_1, C_2, \dots, C_P\}. \quad (9)$$

Algorithm 1 of the Proposed CRPD-PCP Method

Stage	Input: D_{ITS}
ITS	For $n = 1$ to Epoch do For $b = 1$ to Batch do $L_{cls} \leftarrow$ Calculate the loss with Equ. (1) $\theta, \phi = \leftarrow -\nabla_{\theta, \phi}(L_{cls})$ End End $\theta_{ITS}, \phi_{ITS} \leftarrow$ Parameters of F_θ and G_ϕ for model of ITS Output: M_{NCDS} with θ_{ITS} and ϕ_{ITS}
NCDS	Input: D_{NCDS} $\bigcup_{k=1}^{N_n} \{p_1^k\} \leftarrow$ Calculate with Equ. (2) For $n = 1$ to Epoch do For $b = 1$ to Batch do $\bigcup_{k=1}^{N_n} \{p_2^k\} \leftarrow$ Calculate with Equ. (3) $L_{KD_distribution} \leftarrow$ Calculate the loss with $L_{KD_prototype} \leftarrow$ Calculate the loss with Equ. (10) $L \leftarrow$ Calculate the loss with Equ. (11) $\theta, \phi = \leftarrow -\nabla_{\theta, \phi}(L)$ End End Output: M_{NCDS} with θ_{NCDS} and ϕ_{NCDS}
	Input: D_{TEST} Prediction with M_{NCDS} Output: Classification result data and images for each category

Lastly, the prototype-level clustering distillation loss improves novel class recognition by maintaining consistency in a self-supervised way. In the implementation, we employ the original prototype groups G_Q and the flipped counterparts G'_Q . The KD loss enforces similar feature distributions for both original and transformed image prototypes with the following equation:

$$L_{KD_prototype} = \frac{1}{Q} \sum_{q=1}^Q \left[\text{KL}(M^{NCDS}(G_q) \| M^{NCDS}(G'_q)) + \text{KL}(M^{NCDS}(G'_q) \| M^{NCDS}(G_q)) \right]. \quad (10)$$

In summary, the loss function L is defined in the following equation:

$$L = \alpha L_{cls} + \beta L_{KD_distribution} + \gamma L_{KD_prototype} \quad (11)$$

where α, β , and γ represent different loss coefficients.

D. Algorithm of CRPD-PCP Model

In this section, we describe the specific input, output, and formulas in the following algorithm of the CRPD-PCP method, including the ITS and the NCDS.

TABLE II
SUMMARY OF FOUR HSI DATASETS

Data set	Houston	Botswana	Salinas	Pavia
Areas	Houston, USA	Botswana	California, USA	Pavia, Italy
Sensor	ITRES CASI-1500	HYPERION	AVIRIS	ROSIS
Spatial Size	349×1905	256×1476	512×614	610×340
Bands	144	145	204	103
Classes	15	14	16	9

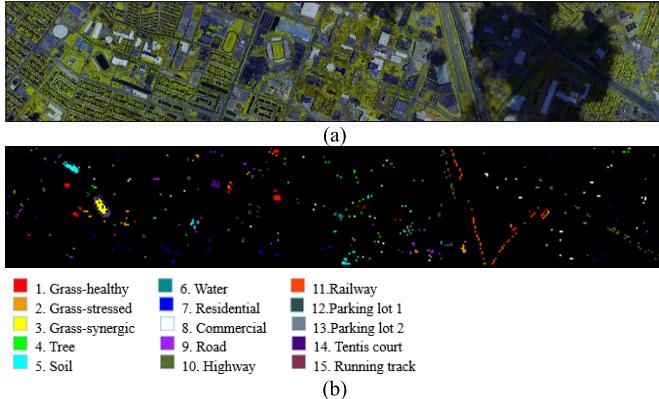


Fig. 5. Data of Houston. (a) False-color image (bands 123, 68, and 9). (b) Ground truth image.

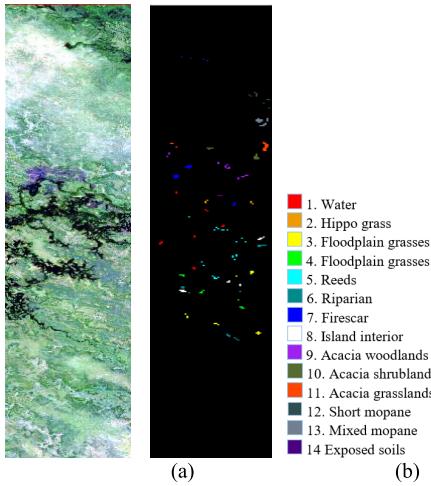


Fig. 6. Data of Botswana. (a) False color image (bands 10, 67, and 114). (b) Ground truth image.

III. EXPERIMENT AND ANALYSIS

A. Data Description

We evaluate the performance of the CRPD-PCP method on four wide HSI datasets, the key information of the datasets are summarized in Table II.

1) *Houston*: The Houston dataset was acquired by the ITRES CASI-1500 sensor and covered the University of Houston campus. The image resolution is 349 × 1905 with 144 spectral bands. Fig. 5(a) and (b) display the false-color and ground truth images with 15 land-cover categories, respectively.

2) *Botswana*: The Botswana dataset was collected by the HYPERION sensor on the EO-1 satellite between 2001 and

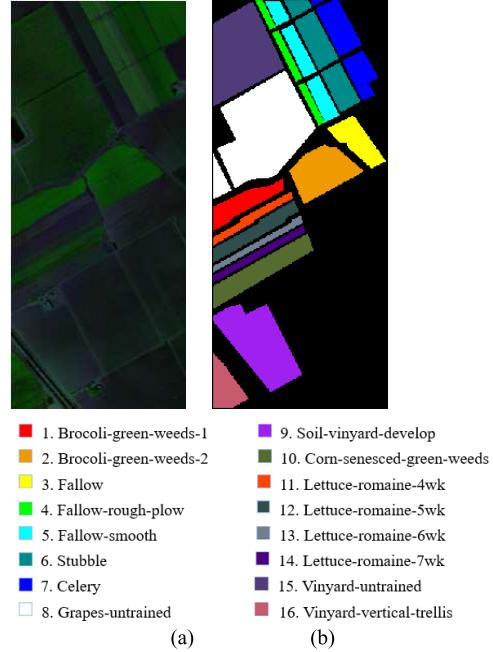


Fig. 7. Data of Salinas. (a) False color image (bands 150, 80, and 172). (b) Ground truth image.

2004, capturing the Okavango Delta region. With 145 spectral bands spanning 400–2500 nm, the Botswana dataset includes 14 land cover categories and 256 × 1476 pixels. Fig. 6(a) and (b) presents the false-color and ground truth maps, respectively.

3) *Salinas*: The Salinas dataset was acquired over Salinas Valley in California with the 224-band AVIRIS sensor. The dataset has a size of 512 × 217 and a spatial resolution of approximately 3.7 m. After eliminating water absorption and low SNR bands, 204 bands remain in this dataset. The false-color and ground truth images with 16 labeled land-cover classes are shown in Fig. 7(a) and (b), respectively.

4) *Pavia*: The Pavia dataset was captured by the Reflective Optics System Imaging Spectrometer (ROSIS) over the city of Pavia which covers the University of Pavia campus. The image has a resolution of 610 × 340 pixels. After removing noisy bands, the spectral bands decreased from 115 to 103. The bands span the wavelength range from 430 to 860 nm, with a spatial resolution of 1.3 m. The dataset encompassed nine categories as presented in Fig. 8(a) and (b) which display the false-color and ground truth images, respectively.

B. Experimental Configuration

All experiments are conducted in an environment built on Pytorch 1.7 and Python 3.8 with an Nvidia GeForce GTX 1050 Ti GPU, an Intel I5-8400 CPU, and 8 GB of RAM. In the preprocessing phase, the HSI datasets are extracted 20 principal components with Principal Component Analysis. The size of samples is 9 × 9, the learning rates are set to 0.001 for the Houston, Botswana, and Pavia datasets, while the rate is set to 0.0001 for the Salinas dataset. The training epochs are configured as 70 for Houston and Pavia, 130 for Botswana, and 80 for Salinas. A consistent batch size of 128 is applied across all datasets, with the Adam optimizer employed for parameter

TABLE III
COMPARISON OF CLASS PERFORMANCE OF DIFFERENT APPROACHES FOR THE HOUSTON DATA

	Class	UNO	rKD	DTC	NCL	TSPT	SCKD	CRPD-PCP
Known Classes	Tree	5.45 ± 6.98	98.52 ± 2.13	96.54 ± 3.91	100 ± 0.00	82.35 ± 1.87	83.25 ± 1.76	93.25 ± 1.93
	Soil	8.67 ± 5.11	5.35 ± 6.84	99.07 ± 1.31	48.90 ± 3.91	86.72 ± 2.03	87.62 ± 2.01	72.56 ± 0.69
	Water	7.51 ± 4.43	33.48 ± 38.93	91.02 ± 4.43	97.97 ± 0.15	84.19 ± 1.95	85.18 ± 1.92	61.60 ± 14.27
	Residential	11.45 ± 7.26	55.35 ± 28.62	86.03 ± 9.54	29.43 ± 5.49	88.56 ± 2.11	89.43 ± 2.15	70.91 ± 6.15
	Highway	12.52 ± 12.37	37.78 ± 26.60	77.67 ± 5.71	36.78 ± 1.62	71.42 ± 2.68	76.32 ± 2.65	80.83 ± 11.49
	Railway	31.11 ± 26.63	4.99 ± 9.98	63.18 ± 4.28	78.18 ± 0.22	92.83 ± 1.75	92.15 ± 1.68	67.29 ± 5.13
	Parking lot1	18.33 ± 9.96	69.29 ± 26.12	90.15 ± 12.49	56.88 ± 6.24	65.37 ± 3.21	67.89 ± 3.12	93.39 ± 3.15
Novel Classes	Tentis court	20.38 ± 29.36	2.85 ± 5.58	99.44 ± 0.32	84.65 ± 0.27	94.69 ± 1.52	94.76 ± 1.53	99.07 ± 0.96
	Grass-healthy	23.44 ± 26.87	4.70 ± 9.40	11.94 ± 16.70	27.89 ± 1.55	75.28 ± 2.74	75.31 ± 2.87	93.41 ± 2.33
	Grass-stressed	0.34 ± 0.38	30.78 ± 37.86	34.45 ± 42.19	84.11 ± 0.97	68.15 ± 3.09	68.75 ± 2.23	82.15 ± 10.82
	Grass-synergic	0.72 ± 0.90	5.79 ± 11.57	36.52 ± 44.74	75.21 ± 1.90	89.32 ± 2.25	72.46 ± 2.79	98.36 ± 1.57
	Commercial	2.15 ± 2.40	15.82 ± 31.61	9.65 ± 11.82	58.17 ± 0.73	70.64 ± 2.86	64.06 ± 3.42	77.28 ± 6.21
	Road	0.80 ± 0.92	2.19 ± 4.38	30.62 ± 37.51	42.88 ± 2.06	59.73 ± 3.26	83.59 ± 2.19	82.75 ± 7.46
	Parking lost2	2.94 ± 2.62	4.80 ± 9.59	25.84 ± 31.69	36.54 ± 2.20	85.81 ± 2.31	70.15 ± 2.98	59.96 ± 21.38
Performance	Running track	5.06 ± 4.69	7.02 ± 13.96	16.58 ± 20.31	98.02 ± 1.34	76.45 ± 2.93	69.52 ± 3.05	97.42 ± 3.90
	OA	10.53 ± 2.78	23.02 ± 3.58	58.65 ± 11.02	56.31 ± 2.01	75.31 ± 3.14	80.05 ± 3.21	84.57 ± 2.59
	Kappa	3.69 ± 3.21	16.59 ± 4.03	55.26 ± 12.03	52.65 ± 2.17	73.56 ± 1.37	76.42 ± 1.43	83.31 ± 2.80

TABLE IV
COMPARISON OF CLASS PERFORMANCE OF DIFFERENT APPROACHES FOR THE BOTSWANA DATA

	Class	UNO	rKD	DTC	NCL	TSPT	SCKD	CRPD-PCP
Known Classes	Reeds	29.65 ± 15.84	88.18 ± 7.21	84.83 ± 16.38	71.48 ± 3.24	83.65 ± 1.92	85.29 ± 1.86	34.57 ± 11.07
	Riparian	24.54 ± 41.43	53.16 ± 31.34	61.56 ± 34.65	98.99 ± 0.67	87.31 ± 2.05	88.74 ± 1.98	48.70 ± 20.19
	Firescar	1.54 ± 2.67	62.86 ± 40.36	80.39 ± 38.46	57.67 ± 0.39	75.42 ± 2.68	93.56 ± 1.57	99.61 ± 0.01
	Island interior	14.04 ± 24.32	9.47 ± 18.95	68.37 ± 36.42	43.10 ± 1.34	92.17 ± 1.63	75.31 ± 2.72	78.72 ± 5.84
Novel Classes	Acacia woodlands	24.84 ± 37.21	0.32 ± 0.64	76.11 ± 37.51	81.93 ± 9.80	68.93 ± 3.15	87.62 ± 1.93	94.08 ± 6.04
	Acacia grasslands	47.21 ± 37.41	61.38 ± 28.37	90.03 ± 11.44	98.72 ± 0.37	79.25 ± 2.37	86.43 ± 2.21	88.70 ± 9.01
	Exposed soils	0.26 ± 0.46	39.37 ± 48.22	58.11 ± 47.46	96.92 ± 0.31	85.68 ± 2.11	82.75 ± 2.28	0.29 ± 0.59
	Water	9.26 ± 16.04	4.95 ± 9.91	44.81 ± 45.67	51.39 ± 0.27	70.36 ± 2.89	94.18 ± 1.49	99.65 ± 0.01
	Hippo grass	0.09 ± 0.15	4.70 ± 9.40	44.30 ± 45.48	32.98 ± 0.01	90.12 ± 1.85	73.52 ± 2.85	98.61 ± 0.79
	Floodplain grasses1	0.01 ± 0.01	2.23 ± 4.45	38.01 ± 46.55	100 ± 0.00	65.74 ± 3.32	83.37 ± 2.33	92.75 ± 8.07
	Floodplain grasses2	7.24 ± 12.54	11.57 ± 23.13	0.65 ± 1.30	54.05 ± 2.29	81.59 ± 2.24	91.25 ± 1.68	98.70 ± 0.90
	Acacia shrublands	2.47 ± 4.27	17.47 ± 34.94	39.76 ± 48.69	41.45 ± 0.16	73.82 ± 2.76	79.64 ± 2.41	93.39 ± 6.34
	Short mopane	0.28 ± 0.48	0.33 ± 0.66	39.78 ± 48.72	92.44 ± 0.54	88.47 ± 1.98	84.59 ± 2.25	97.46 ± 2.06
Performance	Mixed mopane	10.88 ± 18.85	0.01 ± 0.01	0.15 ± 0.30	81.04 ± 0.30	62.95 ± 3.48	83.16 ± 2.22	98.43 ± 1.19
	OA	21.15 ± 13.89	25.77 ± 3.40	51.33 ± 19.45	58.99 ± 1.76	80.12 ± 3.25	84.26 ± 2.97	87.36 ± 2.80
	Kappa	14.85 ± 1.10	19.11 ± 3.71	49.31 ± 21.10	55.53 ± 1.88	78.05 ± 1.42	81.52 ± 1.35	81.97 ± 3.04

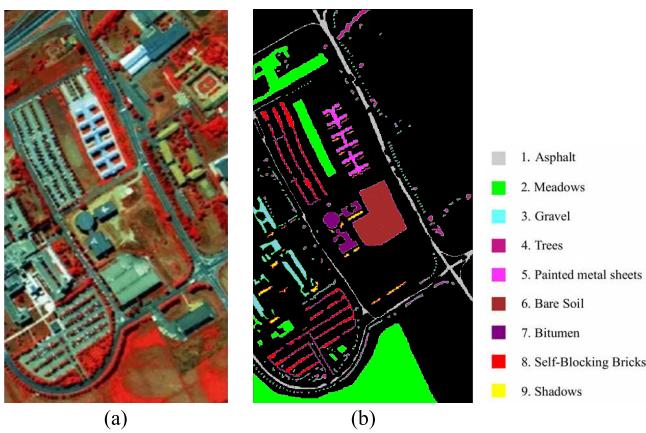


Fig. 8. Data of Pavia. (a) False color image (bands 60, 27, and 17). (b) Ground truth image.

updates. Additionally, the loss UNO coefficients α , β , and γ are set to 0.4, 0.4, and 0.2 for Houston, Botswana, and Pavia, and 0.3, 0.5, and 0.2 for Salinas, respectively. In our experiments, the conventional evaluation metrics include overall accuracy OA and the Kappa coefficient.

C. Comparison Results and Analysis

In this section, we investigate the performance of our proposed model compared with other six approaches on all the datasets. The compared methods include NCL, UNO, rKD, DTC, TSPT, and SCKD. To ensure the comparability of experimental results and demonstrate the superiority of our method, the category selection for all methods is the same on the three different stages. In order to ensure reliable and comparable results, each NCD method on different dataset is repeated five times. The training set comprises 10% randomly sampled data, and all the sample are employed for the testing. K in ITS is set to 8, 5, 7 and 4 for the Houston, Botswana, Salinas, and Pavia datasets, respectively. For NCDS, Q is set to 7, 9, 9, and 5 for the four datasets. All methods adopt fixed categories during ITS to ensure a fair comparison. Classification results are summarized in Tables III–VI and visualized with color-coded maps in Figs. 5–8.

For the Houston dataset, Grass-healthy, Grass-stressed, Grass-synergic, Commercial, Road, Parking lot2, and Running Track are selected as the seven unlabeled categories in the NCDS, while the remaining eight categories serve as known

TABLE V
COMPARISON OF CLASS PERFORMANCE OF DIFFERENT APPROACHES FOR THE SALINAS DATA

	Class	UNO	rKD	DTC	NCL	TSPT	SCKD	CRPD-PCP
Known Classes	Brocoli-green-weeds-1	2.93 ± 3.36	10.73±13.95	60.73± 8.02	99.28 ± 0.94	77.36 ± 0.89	79.52 ± 3.27	99.32 ± 1.20
	Brocoli-green-weeds-2	11.99 ± 9.76	3.15 ± 4.50	68.20±39.45	76.82 ± 0.25	80.72 ± 3.56	82.36 ± 0.65	89.36±13.03
	Fallow-rough-plow	4.10 ± 1.57	82.68±28.03	60.06±44.79	59.63 ± 3.72	83.15 ± 1.24	76.78 ± 5.42	89.08±13.44
	Soil-vinyard-develop	12.55 ± 4.72	99.39 ± 0.60	58.29±39.61	66.20±10.98	78.69 ± 4.11	84.21 ± 2.18	94.23 ± 3.58
	Corn-senesced-green-weeds	7.60 ± 4.18	2.29±2.98	63.58±44.64	56.02 ± 0.14	71.28 ± 2.07	74.43 ± 6.93	98.70 ± 2.45
	Lettuce-romaine-5wk	22.40 ± 5.62	81.14 ± 8.50	68.42 ± 3.46	84.87 ± 1.88	86.34 ± 5.23	88.15 ± 1.47	96.69 ± 0.97
Novel Classes	Vinyard-vertical-trellis	13.45±10.82	68.30±37.00	62.31±46.04	96.30 ± 0.47	64.57 ± 0.98	67.68 ± 4.05	85.19 ± 5.34
	Fallow	10.18 ± 4.64	57.41±26.06	63.63±43.14	89.17 ± 1.19	67.43 ± 6.78	71.35 ± 0.92	91.84 ± 5.55
	Fallow-smooth	5.11 ± 8.23	2.73 ± 5.01	4.16 ± 6.39	77.87 ± 7.51	81.59 ± 2.89	83.47 ± 7.26	82.51±22.16
	Stubble	2.88 ± 1.10	9.21 ± 14.95	20.78±39.34	54.09 ± 0.15	60.38 ± 1.56	64.21 ± 3.68	59.81±23.20
	Celery	12.24±22.20	10.02±12.98	20.51±39.72	87.72 ± 0.01	75.72 ± 7.34	78.56 ± 1.89	98.62 ± 1.48
	Grapes-untrained	3.83 ± 2.57	0.17 ± 0.25	21.13±39.11	56.28 ± 0.32	69.16 ± 3.01	72.89 ± 5.11	85.23±24.90
	Lettuce-romaine-4wk	5.88 ± 5.51	0.09 ± 0.08	9.44 ± 15.46	97.63 ± 3.05	84.25 ± 0.76	85.63 ± 2.74	75.19±38.07
Performance	Lettuce-romaine-6wk	12.29±21.95	4.54 ± 5.63	16.90±22.45	85.08 ± 7.26	63.93 ± 4.89	66.34 ± 0.58	88.43 ± 1.65
	Lettuce-romaine-7wk	5.57 ± 4.30	0.32 ± 0.54	23.51±37.87	35.03 ± 1.15	73.54 ± 2.31	75.27 ± 6.35	90.88±12.80
Performance	OA	11.19 ± 3.82	32.31 ± 2.11	40.06±25.47	74.64 ± 15.51	73.28 ± 8.11	76.55 ± 2.09	82.44± 2.63
	Kappa	9.05 ± 5.33	23.59 ± 2.51	39.13±32.38	71.68 ± 16.58	68.35 ± 1.03	72.42 ± 7.56	81.45 ± 2.94

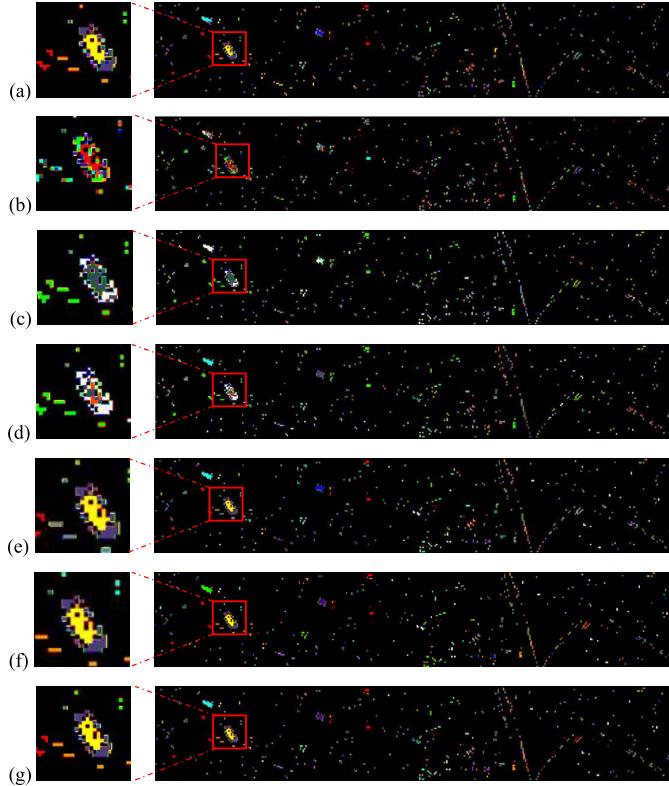


Fig. 9. Classification of maps obtained via different HSIIC methods for the Houston data. (a) NCL. (b) UNO. (c) rKD. (d) DTC. (e) TSPT. (f) SCKD. (g) CRPD-PCP.

classes with prior knowledge during the ITS. Table III reports the classification results for each land cover category, and the corresponding classification maps are illustrated in Fig. 9. As can be observed, the proposed model achieves higher classification accuracy than other methods for the known classes with 80.83% for Highway, 93.39% for Parking lot1, and 99.07% for Tennis court. In the NCDS, the CRPD-PCP method achieves an OA of 84.57% and a Kappa of 83.31% across all categories.

Our approach exhibits slightly lower accuracy in only three categories and outperforms other approaches in the remaining four categories. To be specific, the model acquires limited performance for the categories including Water and Parking lot2 with the accuracy below 70%. Analysis of the predicted and ground truth maps reveals that the model misclassifies some Water pixels as Tree and confuses Parking lot2 with Grass-healthy and Parking lot1.

For the experiment of Botswana data, Water, Hippo grass, Floodplain grasses1, Floodplain grasses2, Acacia shrublands, short mopane, and mixed mopane are defined as the seven unknown categories in the NCDS, while the remaining seven categories serve as prior-known classes. The classification results are shown in Fig. 10 and Table IV. Except for Acacia grasslands and the exposed soils, all unknown categories in the NCDS achieve the OA more than 90%, outperforming other models and demonstrating the strong capability of CRPD-PCP in NCD. For unknown class predictions, some exposed soils pixels are misclassified as Acacia grasslands and Mixed mopane. Meanwhile, parts of Reeds, with a classification accuracy of 34.57%, and Riparian, with an accuracy of 48.70%, are incorrectly assigned to Floodplain grasses2 and Hippo grass, respectively. However, Firescar and Acacia woodlands achieve higher accuracy than other methods. Although the accuracy of exposed soils performs poorly, CRPD-PCP achieves an overall OA of 87.36% and a Kappa of 81.97%, further validating the effectiveness in NCD of HSI.

The third experiment is performed on the Salinas dataset, the novel classes in the NCDS include Fallow, Stubble, Lettuce-romaine-4wk, Lettuce-romaine-6wk, Fallow-smooth, Lettuce-romaine-7wk, Celery, Grapes-untrained, and Vinyard-untrained, and the other seven categories serve as known classes in the ITS. The classification results are presented in Table V and Fig. 11. In ITS, the CRPD-PCP model achieves higher accuracy than other methods on five initial classes. The other two classes Soil-vinyard-develop and Vinyard-vertical-trellis achieve lower accuracy than other

TABLE VI
COMPARISON OF CLASS PERFORMANCE OF DIFFERENT APPROACHES FOR THE PAVIA DATA

	Class	UNO	rKD	DTC	NCL	TSPT	SCKD	CRPD-PCP
Known Classes	Asphalt	3.37 ± 2.43	92.2 ± 2.5	87.98 ± 1.25	100 ± 0.00	89.45 ± 3.23	97.45 ± 0.23	98.15 ± 0.85
	Meadows	60.62 ± 43.5	83.76 ± 6.12	99.59 ± 0.22	98.86 ± 0.01	99.27 ± 1.06	99.98 ± 0.65	99.89 ± 0.02
	Gravel	0.46 ± 0.65	98.69 ± 0.36	45.28 ± 2.93	99.42 ± 0.50	82.37 ± 4.24	57.79 ± 3.14	81.91 ± 18.20
	Trees	31.58 ± 41.66	0.11 ± 0.03	85.84 ± 2.15	98.81 ± 0.02	89.62 ± 7.15	92.56 ± 1.17	95.48 ± 0.72
Novel Classes	Painted metal sheets	21.69 ± 41.98	0.86 ± 0.06	74.65 ± 14.05	6.83 ± 3.14	98.92 ± 1.83	99.20 ± 1.24	99.45 ± 0.11
	Bare Soil	0.12 ± 0.16	3.93 ± 2.51	91.88 ± 3.77	6.10 ± 0.09	73.20 ± 7.93	46.95 ± 2.36	80.27 ± 22.18
	Bitumen	0.81 ± 0.98	0.09 ± 0.03	23.9 ± 10.75	2.78 ± 9.32	91.35 ± 0.63	95.49 ± 3.83	86.20 ± 1.81
	Self-Blocking Bricks	0.15 ± 0.15	1.86 ± 2.17	46.09 ± 1.71	6.26 ± 3.02	24.12 ± 3.54	89.14 ± 2.37	69.90 ± 7.90
Performance	Shadows	6.16 ± 8.71	9.63 ± 3.14	97.99 ± 0.63	9.58 ± 2.13	89.02 ± 3.13	92.47 ± 1.31	97.49 ± 0.37
	OA	31.78 ± 18.43	55.66 ± 3.03	85.46 ± 1.45	71.12 ± 3.02	86.58 ± 2.21	89.68 ± 2.34	93.95 ± 2.94
	Kappa	10.56 ± 9.92	39.13 ± 3.35	80.63 ± 1.95	59.36 ± 0.42	81.91 ± 3.27	85.92 ± 1.56	90.67 ± 4.08

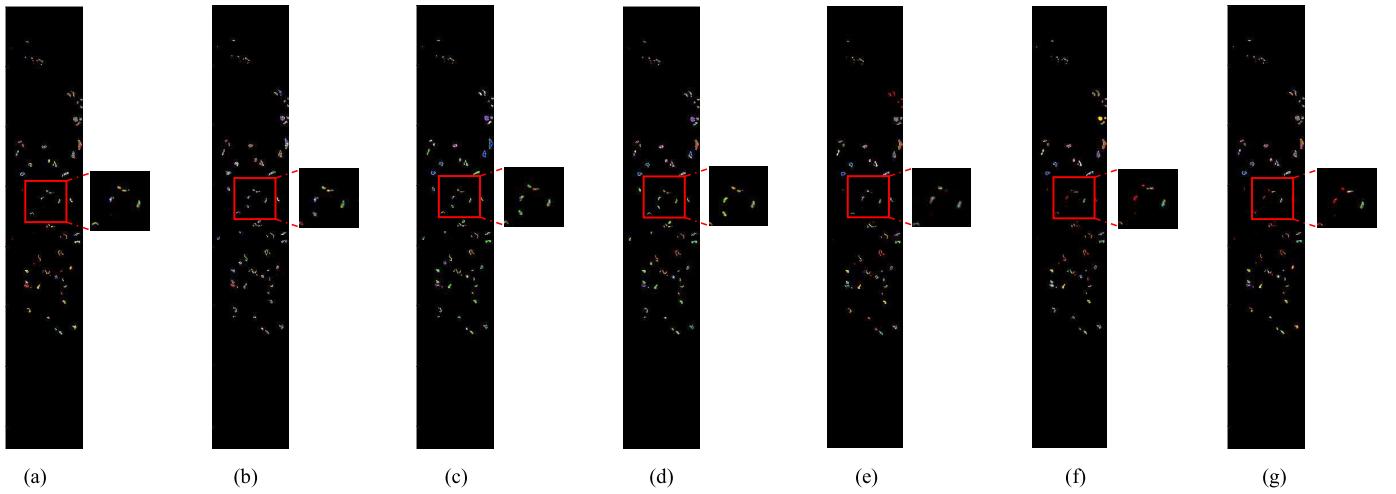


Fig. 10. Classification maps obtained via different HSIIC methods for the Botswana data. (a) NCL. (b) UNO. (c) rKD. (d) DTC. (e) TSPT. (f) SCKD. (g) CRPD-PCP.

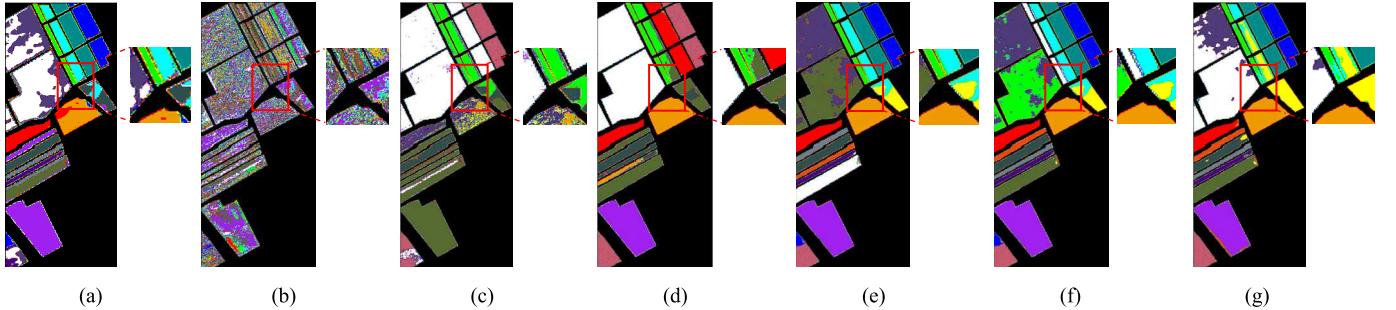


Fig. 11. Classification maps obtained via different HSIIC methods for the Salinas data. (a) NCL. (b) UNO. (c) rKD. (d) DTC. (e) TSPT. (f) SCKD. (g) CRPD-PCP.

methods. In NCDS, the performance difference between Grapes-untrained and Lettuce-romaine-5wk remains below 10%, demonstrating the capability of our model to transfer knowledge from known classes. The Vinyard-untrained class achieves 71.39% classification accuracy, with edge pixels misclassified as Grapes-untrained. Additionally, our model demonstrates inferior performance on the Lettuce-romaine-4wk class with OA of 75.19%, while NCL shows emphasis on novel class clustering. The CRPD-PCP outperforms other methods for Celery, Stubble, Fallow-smooth, Lettuce-romaine-6wk, Fallow, and Lettuce-romaine-7wk. The CRPD-PCP

method acquires an OA of 84.57% and a Kappa of 83.31%, clearly manifesting the strong performance in NCD and overall classification.

Table VI and Fig. 12 present the results of the Pavia dataset. As observed, the four known classes including Asphalt, Meadows, Gravel, and Trees achieve an accuracy exceeding 80%. Among them, the accuracy of Meadows and Asphalt are 99.89% and 98.15%. For NCDS, five novel categories such as Shadows, Bare Soil, and Painted metal sheets outperform other methods. Notably, Bare soil and Self-Blocking Bricks suffer misclassified as meadows, and painted metal sheets

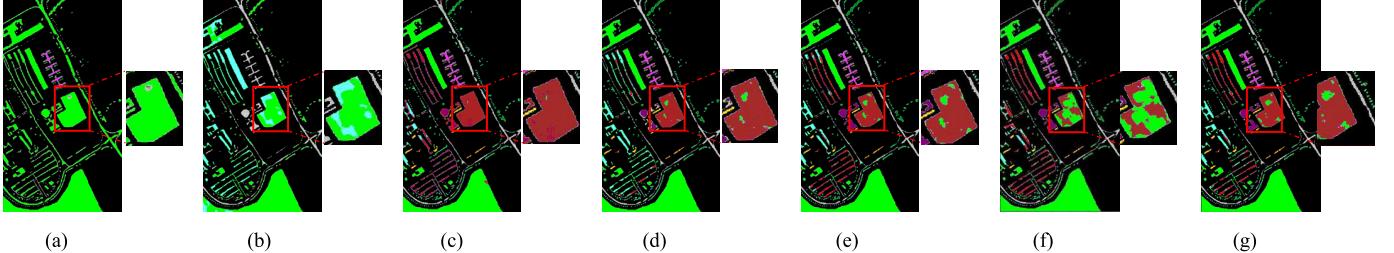


Fig. 12. Classification maps obtained via different HSIIC methods for the Pavia data. (a) NCL. (b) UNO. (c) rKD. (d) DTC. (e) TSPT. (f) SCKD. (g) CRPD-PCP.

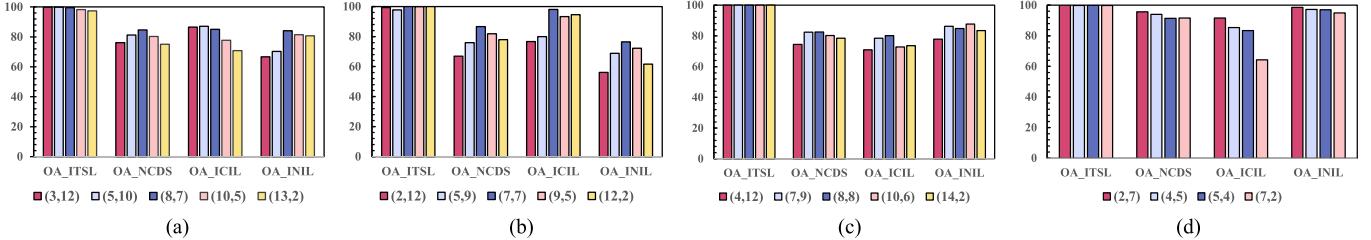


Fig. 13. Accuracy results generated with the CRPD-PCP with K and Q in the CRPD-PCP for the four datasets. (a) Houston. (b) Botswana. (c) Salinas. (d) Pavia.

achieves an accuracy of 99.45%. In contrast, the proposed CRPD-PCP method achieves an OA of 93.95% and a Kappa of 90.67, confirming the superiority and effectiveness on NCD for HSI.

D. Impact of the Class Number in the ITS

To explore the impact of different category numbers in NCDS, our experiments are designed with two stage category number combinations denoted as (K, Q) , where K and Q represent the number of known categories in ITS and new categories in NCDS, respectively. For Houston dataset, the settings are $(3, 12)$, $(5, 10)$, $(8, 7)$, $(10, 5)$, $(13, 2)$, for Botswana and Salinas datasets, they are $(2, 12)$, $(5, 9)$, $(7, 7)$, $(9, 5)$, $(12, 2)$ and $(4, 12)$, $(7, 9)$, $(8, 8)$, $(10, 6)$, $(14, 2)$. For the Pavia data, the settings are $(2, 7)$, $(4, 5)$, $(5, 4)$, $(7, 2)$. Besides, four extended metrics OA_ITSL, OA_INIL, OA_ICIL, and OA_NCDS are employed for a more comprehensive assessment. Notably, OA_NCDS aligns with the OA metric commonly used in similar studies. In the subsequent analysis, OA in tables and figures refers to OA_NCDS unless otherwise backbone network under supervised learning conditions. Experimental results indicate that varying the number of known categories has negligible impact on OA_ITSL performance, with classification accuracies consistently exceeding 98% across all four datasets. OA_NCDS evaluates capability in recognizing unlabeled novel classes while preserving knowledge of previously known classes. The Houston dataset shows the variations of OA_NCDS between 75% and 85%, while the Botswana dataset reaches peak accuracy at 86.7%. The Salinas dataset maintains stable accuracy near 80%. OA_NCDS values for Houston and Botswana initially increase then decrease with growing known category numbers, whereas Salinas remains relatively unaffected by such variations. Additionally, denoting the classification accuracy of new classes under the new class

discovery model, OA_INIL reflects the generalization capability when handling novel categories. Experimental results show that OA_INIL values for all four datasets exhibit an initial increase followed by decrease as the number of known categories grows specified. Experimental results on four extended metrics are shown as Fig. 13. Specifically, OA_ITSL evaluates the classification accuracy of known classes within the ITS framework, demonstrating the discriminative capability of the OA_ICIL serves as a key metric for evaluating new class discovery model performance on known classes. The Botswana and Salinas datasets exhibit similar OA_ICIL trends, initially rising then declining as known category numbers increase. The Botswana dataset exhibits known knowledge forgetting with larger disparities between known and novel category numbers, achieving optimal integration only when both category numbers balance at 7. The Salinas dataset shows lowest OA_ICIL with highly dissimilar category numbers, peaking at 7 known and 9 new categories. The Houston dataset demonstrates an inverse relationship, where higher known and new category ratios correspond to lower OA_ICIL values, indicating proportional known knowledge forgetting in the CRPD-PCP framework. For Pavia, OA_ITSL maintains high stability for 95%. As the combination of known and novel category numbers shifts, OA_NCDS exhibits a trend of initial ascent followed by decline. A relatively high value emerges with $(4, 5)$, which indicates a balanced quantity between known and novel categories. OA_INIL fluctuates with the increase in known category counts. Regarding OA_ICIL, similar to trends observed in the Botswana and Salinas datasets, optimal performance is achieved when known and novel category counts are balanced, as seen with 5 known and 4 novel categories. In cases of substantial disparities in category counts, the known knowledge forgetting arises, demonstrating that the count of category balance exists as an influence on the integration of known knowledge in the model of NCDS.

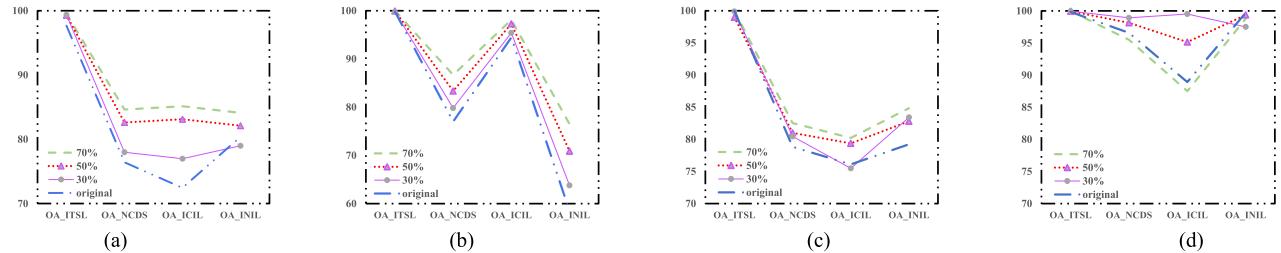


Fig. 14. Compared OAs generated with the different prototype numbers for the four datasets. (a) Houston. (b) Botswana. (c) Salinas. (d) Pavia.

TABLE VII
ABLATION STUDY ON THE FOUR HSI DATASETS

	Module				Accuracy	
	Baseline	CBMA	CRPD	PCP	OA_ICIL	OA_NCDS
Houston	✓				74.74	75.96
	✓	✓			68.60	63.74
	✓		✓		72.50	68.10
	✓			✓	78.07	80.55
	✓	✓	✓		72.31	69.50
	✓	✓		✓	85.12	83.48
	✓		✓	✓	81.93	80.70
	✓	✓	✓	✓	85.13	84.57
Botswana	✓				96.88	84.96
	✓	✓			94.47	81.76
	✓		✓		96.21	84.37
	✓			✓	96.23	84.16
	✓	✓	✓		93.42	80.70
	✓	✓		✓	97.00	81.28
	✓		✓	✓	95.05	79.50
	✓	✓	✓	✓	98.11	87.36
Salinas	✓				73.63	75.66
	✓	✓			72.82	74.29
	✓		✓		73.56	74.13
	✓			✓	79.68	77.40
	✓	✓	✓		81.85	81.81
	✓	✓		✓	78.76	79.44
	✓		✓	✓	83.01	80.38
	✓	✓	✓	✓	80.21	82.44
Pavia	✓				75.28	69.14
	✓	✓			80.17	77.34
	✓		✓		77.49	75.27
	✓			✓	75.27	80.32
	✓	✓	✓		77.36	73.24
	✓	✓		✓	82.23	90.79
	✓		✓	✓	74.07	86.62
	✓	✓	✓	✓	85.37	93.95

E. Ablation Study

Ablation experiments on four datasets are conducted in this section, with the conventional NCD framework serving as baseline. The experimental design incorporates CBMA, CRPD, and PCP modules for comparative evaluation. Results in Table VII reveal that different modules produce significant variations in both OA_NCDSM and OA_ICIL metrics compared to the baseline, the ablation study demonstrates complementary functionality among the proposed modules. As can be observed, CBAM consistently enhances feature representation, improving OA metrics for Houston and Botswana datasets while showing positive effects in most Salinas cases. PCP exhibits particularly strong standalone performance, achieving notable results across datasets with special effectiveness in Houston and Salinas. The analysis of module combinations shows that CBAM working with

PCP not only produces significant accuracy gains for Salinas but also achieves a balanced performance between OA_ICIL improvement and minor reductions in other Botswana metrics. CRPD displays more variable performance, with its combination with PCP surpassing standalone CRPD results in Houston and Salinas but not in Botswana. As for Pavia, the OA_ICIL values exhibit variations with different module combinations, ranging from 74.07% to 85.37%, indicating the differential impacts of Baseline, CBMA, CRPD, and PCP on known class. The OA_NCDS values vary from 69.14% to 93.95%, which reflects the capability to recognize novel classes and to retain known class knowledge. Notably, the combination of multiple modules enhances the accuracy for both metrics and highlight the importance of rational module selection and integration.

F. Analysis of Prototype-Level Clustering Reliability

To investigate the effectiveness of the prototype-level clustering method by analyzing four OA metrics under different prototype number, we design the following experiments. The number of prototypes is increased from the actual number of classes to 30%, 50%, and 70% of the original class of samples. Results shown as Fig. 14 demonstrate that employing multiple prototypes per class enhances accuracy across OA_INIL, OA_ICIL, and OA_NCDS for four datasets, which indicated that the prototype-level clustering method have a positive guiding effect on NCD. The model of ITS show classification capability, resulting in minimal effect to OA_ITSL across different prototype numbers. Performance improves initially with more prototypes, but stops increasing when prototype numbers exceed 50% of the sample size. For OA_NCDS, all four datasets show only marginal improvements between 50% and 70% prototype ratios compared to the 30% and 50% range. The value of OA_ICIL for the Botswana dataset tends to reach stability for 97.32%, while that for the Houston and Salinas datasets increase to a certain extent and gradually converge for around 84%. All four datasets show converging OA_INIL values as prototype numbers increase. For the Pavia dataset, OA_ITSL stays stable with the all the numbers of prototypes, while OA_INIL first rises and then falls as prototypes increase. OA_NCDS increases when prototypes reach 50% of the sample size, indicating that moderate prototypes improve the recognition of novel classes. In short, the prototype-level clustering method helps NCD for Pavia, and the prototype numbers should be controlled to balance the performance metrics.

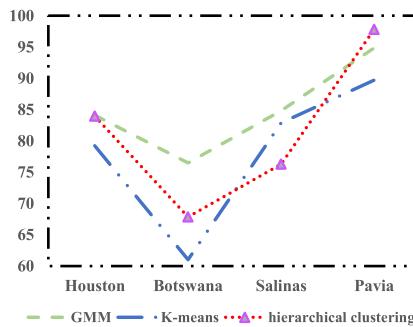


Fig. 15. Performance comparison for the four datasets with different cluster methods.

G. Analysis of Different Clustering Algorithms

We evaluate the impact of different clustering methods on prototype level clustering as depicted in Fig. 15. The experimental setup compares three fundamental clustering approaches, including K-means which partitions data through center refinement to minimize within-cluster variance. While simple and computationally efficient, this method shows strong dependence on initial center positions. GMM represents data through a combination of Gaussian distributions and is suitable for complex data distributions. Hierarchical clustering creates layered groups by merging smaller groups, removing the requirement to set the group count in advance. The numerical results of experiments across four datasets demonstrate the impact of prototype clustering operations performed during NCDS. We specifically evaluate the OA_INIL metric to validate how effectively our clustering approach preserves knowledge in prototype-level clustering. For Houston, GMM and hierarchical clustering achieve nearly identical accuracy at 84.57%. On the Botswana, GMM reaches 77.43% accuracy, significantly outperforming K-means at 62.18% and hierarchical clustering at 68.32%. The Salinas shows comparable performance between GMM and K-means, both achieving approximately 83% accuracy. The performance of Pavia varies across different methods, yet all results demonstrate an upward trend, especially with the accuracy of GMM reaching nearly 96%. The GMM method consistently achieves higher accuracy than K-means and hierarchical clustering in unlabeled clustering of new categories. The experiments demonstrate the GMM approach more effectively captures the complex spectral spatial characteristics in HSI due to the ability to model overlapping clusters. The results confirm GMM better for clustering unlabeled prototype for the NCD of HSI tasks, outperforming the K-means and the hierarchical clustering algorithms.

H. Visualization of the Category Distribution

We employ the t-distributed stochastic neighbor embedding (t-SNE) algorithm to visualize the category feature distribution of our method. Specifically, we conducted experiments on the Houston, Botswana, Salinas, and Pavia datasets, where the final cluster numbers were 15, 14, 16, and 9, respectively. Fig. 16 visually compares the original hyperspectral data with ground truth labels against the obtained

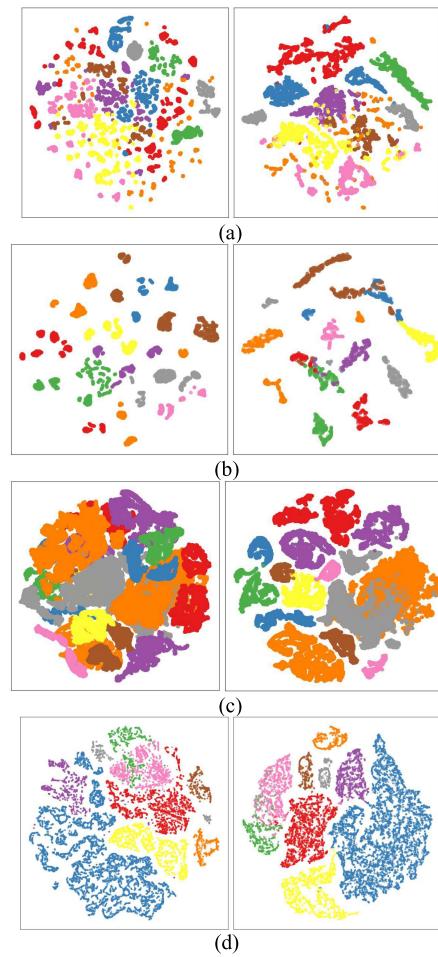


Fig. 16. Diagram of T_NSE for the four datasets. (a) Houston. (b) Botswana. (c) Salinas. (d) Pavia.

features with the CRPD-PCP model. As can be observed, the results demonstrate a remarkable transformation from initially disordered patterns to well-separated clusters that precisely match the semantic categories. The maps of all four datasets demonstrate the consistent improvement where the processed features form compact and distinct groupings with clear boundaries between different classes. Briefly speaking, the visualization provides believable evidence of the effectiveness of the CRPD-PCP model in learning discriminative representation.

I. Analysis of Different Ratios in Training Stage

In this section, we evaluate the impact of training data proportions on model performance by examining two key ratios. Specifically, *Rate1* represents the proportion of the training set to the test set and sets as 5%, 10%, and 15%. *Rate2* is denoted by the proportion of known classes in NCDS which evaluated at 10%, 15%, and 20%. The results on different datasets are reported in Fig. 17 and Table VIII. For the Houston dataset, experimental results demonstrate remarkable stability in OA_SNIT performance, which consistently maintains 100% accuracy regardless of *Rate2* increases when *Rate1* remains fixed. The OA_NCDS metric demonstrates complex dynamic behavior with distinct pattern variations across different *Rate1*

TABLE VIII
RESULTS OF DIFFERENT RATE1 AND RATE2 FOR FOUR DATASETS

DataSet	Rate1	Rate2	OA_ITSL	OA_NCDS	OA_INIL	OA_ICIL
Houston	5%	10%	98.87	71.26	66.87	76.53
	5%	15%	98.79	77.81	83.29	71.22
	5%	20%	98.74	78.10	83.83	71.21
	10%	10%	99.51	84.57	84.11	85.13
	10%	15%	99.45	82.65	86.48	78.04
	10%	20%	99.55	82.94	84.56	75.30
	15%	10%	99.73	83.63	83.53	83.74
	15%	15%	99.72	81.72	86.24	76.29
	15%	20%	99.74	84.96	94.81	73.12
	5%	10%	66.39	54.09	48.89	59.91
Botswana	5%	15%	64.47	47.14	21.70	75.55
	5%	20%	72.99	64.81	61.38	68.64
	10%	10%	100	87.36	76.49	98.11
	10%	15%	99.89	89.11	82.32	96.71
	10%	20%	99.91	91.51	90.04	93.16
	15%	10%	99.86	90.38	85.82	95.46
	15%	15%	99.88	90.28	86.09	94.97
	15%	20%	99.88	92.83	90.79	95.10
	10%	5%	99.99	82.44	84.83	80.21
	10%	10%	99.98	91.93	94.53	88.28
Salinas	10%	15%	99.98	93.46	98.01	87.07
	15%	5%	99.98	89.73	94.97	82.36
	15%	10%	99.97	91.67	97.93	82.92
	15%	15%	99.99	91.93	98.09	83.30
	5%	10%	99.40	90.34	97.88	71.71
	5%	15%	99.78	91.22	98.41	73.49
Pavia	5%	20%	99.88	95.18	98.86	86.11
	10%	10%	98.43	93.97	99.08	81.34
	10%	15%	99.97	90.31	98.77	69.44
	10%	20%	99.75	93.95	97.26	85.37
	15%	10%	99.93	92.21	96.11	82.58
	15%	15%	99.95	87.57	97.04	64.21
	15%	20%	99.89	92.34	99.27	75.23
	5%	10%	99.40	90.34	97.88	71.71
	5%	15%	99.78	91.22	98.41	73.49
	5%	20%	99.88	95.18	98.86	86.11

settings. At the 5% training ratio, performance follows a characteristic rise-and-fall trajectory, whereas both 10% and 15% configurations show an initial performance decline before subsequent improvement. The optimal balance emerges at 10% for both ratios, achieving peak OA_NCDS accuracy of 84.57% while preserving robust performance across all evaluation metrics. The Botswana dataset presents greater sensitivity to parameter adjustments. While the maximum OA_NCDS value of 89.11% occurs at the setting of Rate1 to 10% and Rate2 to 20%, this configuration demonstrates compromised performance in OA_ICIL measurements. Evaluation of the Salinas dataset reveals fundamental limitations at the setting of Rate1 to 5% due to insufficient training sample availability.

Subsequent analysis of Rate1 configurations identifies the combination with Rate1 of 10% and Rate2 of 5% as delivering optimal model performance. Likewise, for the Pavia dataset, OA_SNIT remains unaffected by variations in the two ratios. When the Rate1 keeps constant as Rate2 increases, OA_ICIL and OA_NCDS exhibit a trend of initially decreasing followed by an increase. As observed, the optimal configuration is Rate1 at 5% and Rate2 at 20%. Briefly, the Houston and Botswana datasets both converge on a 10% setting for both ratios, while Salinas attains optimal performance with Rate1 at 10% and Rate2 at 5%, while Pavia achieves satisfactory results with the setting of Rate1 to 5% and Rate2 to 20%. The research outcomes identify dataset-specific

TABLE IX
EFFICIENCY COMPARISON WITH DIFFERENT MODELS

Method	Params(M)	FLOPs(G)
UNO	27.65	2.3
rKD	11.68	0.23
DTC	17.98	7.32
NCL	1.13	1.3
TSPT	87.92	18
SCKD	92.23	21.35
CRPD-PCP	1.25	0.42

ratio configurations is beneficial for the CRPD-PCP model learning while ensuring the model stability in diverse NCD for HSI.

J. Efficiency Comparison With Different Models

In this section, we evaluate the computational efficiency of the compared models by analyzing the Params and FLOPs. As shown in Table IX, both TSPT and SCKD models have higher computational complexity to achieve NCD of intricate patterns than other models. The rKD model that leverages KD to enhance inference efficiency has 11.68 MB of Params and 0.23 MB of FLOPs. Conversely, the DTC model incurs high FLOPs due to the clustering operations, which restrict the applicability in the subsequent NCD task. While the NCL model serves as a viable lightweight approach, it compromises

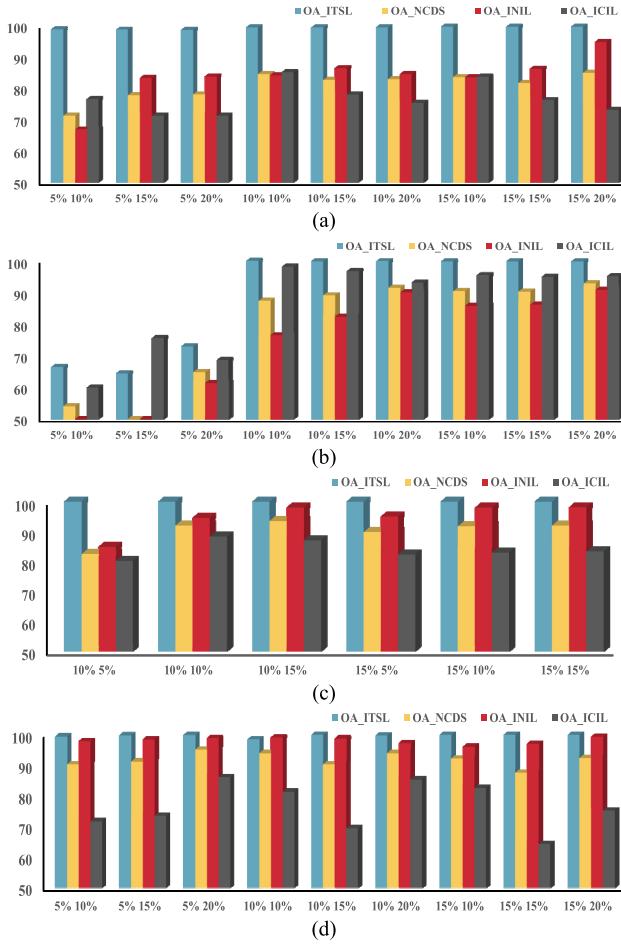


Fig. 17. Accuracy results with different *Rate1* and *Rate2* in the CRPD-PCP for the four datasets. (a) Houston. (b) Botswana. (c) Salinas. (d) Pavia.

TABLE X
TRAINING AND CLUSTER TIME CONSUMPTION
OF DIFFERENT APPROACHES (S)

Dataset	Time	UNO	rKD	DTC	NCL	TSPT	SCKD	DRDC-PCP
Houston	Train	132.43	284.72	210.9	77.45	198.27	327.19	320.32
	Cluster	27	33.08	39.89	23.47	8.87	10.23	22.48
Botswana	Train	99.89	190.11	220.59	66.37	145.90	70.11	73.28
	Cluster	12.79	33.90	40.64	25.78	10.32	12.94	16.75
Salinas	Train	648.89	1545.9	590.89	132.99	97.45	681.45	558.27
	Cluster	52.43	45.89	30.67	21.63	31.45	79.06	43.67
Pavia	Train	110.35	200.56	185.72	85.23	160.41	300.63	280.45
	Cluster	20.15	30.22	32.58	18.64	9.35	11.47	19.36

accuracy in NCD scenarios. In contrast, the proposed CRPD-PCP model has 1.25 MB of Params and 0.42 MB of FLOPs, which demonstrate lightweight performance while effectively facilitating NCD.

We also reported the training time and the clustering time for both the proposed method and the comparative methods in Table X. As observed, while the UNO approach exhibits a relatively high training time for the Houston dataset, our method achieves a well-balanced training time of 320.32 s, along with a clustering time of 22.48 s. For the Botswana dataset, the NCL method achieves low training time of 66.37 s, while has the cluster time of 25.78 s. The TSPT generates the lowest cluster time, although it has a training time

TABLE XI
COMPARISON ON LOSS COEFFICIENTS FOR THE FOUR HSI DATASETS

	Coefficient			Accuracy			
	α	β	γ	OA_NCDS	OA_ITSL	OA_INIL	OA_ICIL
Houston	0.4	0.4	0.2	84.57	84.11	85.12	85.13
	0.4	0.4	0.1	80.32	85.31	80.32	74.32
	0.4	0.2	0.2	83.29	63.67	72.45	80.24
	0.2	0.4	0.2	82.11	78.08	80.49	79.14
Botswana	0.4	0.4	0.2	87.36	70.89	97.28	98.11
	0.4	0.4	0.1	82.54	65.32	92.17	87.32
	0.4	0.2	0.2	85.19	58.46	90.33	77.49
	0.2	0.4	0.2	83.77	62.18	91.55	92.88
Salinas	0.3	0.5	0.2	82.44	99.99	94.83	82.44
	0.3	0.5	0.1	78.32	98.55	92.16	63.24
	0.3	0.3	0.2	80.17	95.33	90.22	78.33
	0.1	0.5	0.2	75.94	93.12	80.21	74.37
Pavia	0.4	0.4	0.2	93.95	99.75	97.25	85.37
	0.4	0.4	0.1	89.22	98.33	94.18	80.15
	0.4	0.2	0.2	91.55	95.67	90.44	70.34
	0.2	0.4	0.2	87.88	94.21	88.73	78.42

of 145.90 s. Our approach remains competitively efficient in both the training and clustering stages. For the Salinas dataset, our method outperforms the UNO and rKD methods with lower training times. Additionally, while our approach requires more time compared to the NCL and TSPT methods, it delivers superior performance overall and indicates that our method strikes a satisfactory balance between accuracy and efficiency. For the Pavia dataset, the NCL and UNO demonstrate improved training performance due to the simple structures. While our approach requires 280.45 s for training and 19.36 s for clustering, striking a balance between performance and efficiency. Notably, the proposed DRDC-PCP method maintains consistency in time consumption for all the four datasets. In summary, our method not only excels in accuracy but also maintains a commendable computational efficiency, demonstrating a well-balanced trade-off between the two critical factors.

K. Analysis of Different Coefficients of the Loss UNO

In this section, we conduct experiments to demonstrate the influence of different coefficients of the loss function across the four datasets. As shown in Table XI, for the Houston dataset, the OA_ITSL decreases when the loss coefficient β is reduced, indicating that class relationship distillation is essential for effective cross-class learning. For the Botswana dataset, a reduction in γ leads to a significant decline in OA_ICIL, which emphasizes the impact on spatial-spectral feature extraction. In the Salinas dataset, a decrease in the α coefficient results in the most substantial drop in OA_NCDS and confirms the importance of the basic cross-entropy loss. Additionally, for the Pavia dataset, OA_ICIL declines with a reduction in β , which verifies that the class relationship constraints are particularly critical for feature aggregation. Overall, the experimental results demonstrate that appropriate coefficients for the loss function contribute to achieving satisfactory performance.

L. Analysis of Confusion Matrix

To comprehensively evaluate the generalization and classification performance of CRPD-PCP, we report the confusion

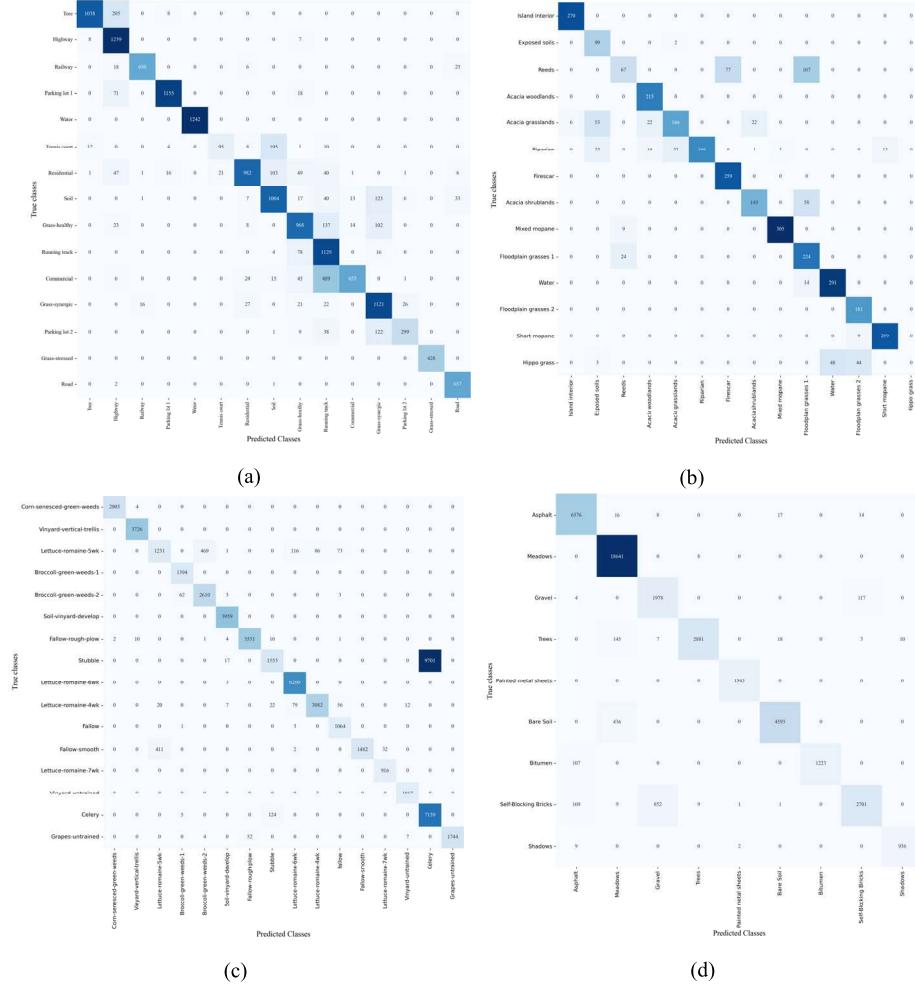


Fig. 18. Confusion matrix for the four datasets. (a) Houston. (b) Botswana. (c) Salinas. (d) Pavia.

matrix comparison on the four datasets shown as Fig. 18. For the Houston dataset, our model demonstrates robust discriminative capability for major categories such as road and water. But the confusion between similar land-covers like residential and commercial highlights the challenges posed by complex landscapes. Most categories in the Botswana dataset are well classified including Island interior and Acacia woodlands, while the misclassifications occur between Acacia grasslands and Riparian zones, reflecting the effect of the inherent spectral overlaps for NCD. Due to the dynamic spectral changes, the Salinas dataset encounters difficulties in recognizing crop at various growth stages. Nevertheless, our approach achieves high accuracy in identifying stable land-covers like Grapes-untrained. For the Pavia dataset, CRPD-PCP shows strong performance in classifying prominent categories such as Meadows and Bare Soil, yet error arises between gravel and asphalt due to the reflectance.

IV. CONCLUSION

To solve the NCD task of HSI, we present the CRPD-PCP network, which consists of two key components including the CRPD module and the PCP module. The CRPD module enforces distillation constraints by leveraging relationships

between new and known categories during NCD stage. With the relation perception mechanism, our model effectively reduces recognition gaps between categories and meanwhile preserves knowledge of known classes. Besides, the PCP module generates pseudolabels by analyzing SAM distance similarities between new class samples during prototype-level clustering. The self-supervised KD constraint in PCP enhances pseudolabel reliability while strengthening the model generalization in complex scenarios. Experimental results on four popular HSI datasets demonstrate the effectiveness of the CRPD-PCP framework in NCD and highlight the potential advantage in the field of hyperspectral interpretation analysis.

In the future, we aim to extend the CRPD-PCP model with well-designed graph distillation memory block for the challenging large-scale datasets that contain multiple class types and diverse features. For the multimodal remote sensing data, we intend to obtain semantic mapping with knowledge graph involved network, enabling the NCD model to adaptively distinguish between known, discovered, and unknown classes that emerge in real scenario. Besides, we plan to investigate more lightweight network architectures and efficient distillation strategies to further reduce the computational costs of the NCD model for real-world applications.

REFERENCES

- [1] 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.
- [2] C. Yu, Y. Zhu, M. Song, Y. Wang, and Q. Zhang, "Unseen feature extraction: Spatial mapping expansion with spectral compression network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Jun. 2024, Art. no. 5521915.
- [3] H. Yang, H. Yu, K. Zheng, J. Hu, T. Tao, and Q. Zhang, "Hyperspectral image classification based on interactive transformer and CNN with multilevel feature fusion network," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [4] W. Sun et al., "A simple and effective spectral-spatial method for mapping large-scale coastal wetlands using China ZY1-02D satellite hyperspectral images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 104, Dec. 2021, Art. no. 102572.
- [5] K. Liu et al., "Mapping coastal wetlands using transformer in transformer deep network on China ZY1-02D hyperspectral satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3891–3903, 2022.
- [6] C. Zhou, Z. He, J. Dong, Y. Li, J. Ren, and A. Plaza, "Low-rank and sparse representation meet deep unfolding: A new interpretable network for hyperspectral change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, May 2025, Art. no. 5513516.
- [7] C. Zhou, Z. He, A. Lou, and A. Plaza, "RGB-to-HSV: A frequency-spectrum unfolding network for spectral super-resolution of RGB videos," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Feb. 2024, Art. no. 5609318.
- [8] J. Li, K. Zheng, L. Gao, Z. Han, Z. Li, and J. Chanussot, "Enhanced deep image prior for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, Jan. 2025, Art. no. 5504218.
- [9] J. Li, K. Zheng, L. Gao, L. Ni, M. Huang, and J. Chanussot, "Model-informed multistage unsupervised network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Apr. 2024, Art. no. 5516117.
- [10] C. Yu, Y. Zhu, Y. Wang, E. Zhao, Q. Zhang, and X. Lu, "Concern with center-pixel labeling: Center-specific perception transformer network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, May 2025, Art. no. 5514614.
- [11] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.
- [12] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, Jan. 2017.
- [13] B. Liu, X. Yu, P. Zhang, A. Yu, Q. Fu, and X. Wei, "Supervised deep feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1909–1921, Apr. 2018.
- [14] A. Paul, S. Bhoumik, and N. Chaki, "SSNET: An improved deep hybrid network for hyperspectral image classification," *Neural Comput. Appl.*, vol. 33, no. 5, pp. 1575–1585, Mar. 2021.
- [15] H. Zhong, L. Li, J. Ren, W. Wu, and R. Wang, "Hyperspectral image classification via parallel multi-input mechanism-based convolutional neural network," *Multimedia Tools Appl.*, vol. 81, no. 17, pp. 24601–24626, Jul. 2022.
- [16] M. Q. Alkhatab et al., "Tri-CNN: A three branch model for hyperspectral image classification," *Remote Sens.*, vol. 15, no. 2, p. 316, Jan. 2023.
- [17] C. K. K. Reddy et al., "Enhancing precision agriculture and land cover classification: A self-attention 3D convolutional neural network approach for hyperspectral image analysis," *IEEE Access*, vol. 12, pp. 125592–125608, 2024.
- [18] J. Zhang, Z. Meng, F. Zhao, H. Liu, and Z. Chang, "Convolution transformer mixer for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [19] M. Ahmad, U. Ghous, M. Usama, and M. Mazzara, "WaveFormer: Spectral-spatial wavelet transformer for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.
- [20] K. Zhang, Z. Tan, J. Sun, B. Zhu, Y. Yang, and Q. Lv, "A multidimensional spectral transformer with channel-wise correlation for hyperspectral image classification," *Appl. Sci.*, vol. 13, no. 9, p. 5482, Apr. 2023.
- [21] Z. Gong, X. Zhou, and W. Yao, "MultiScale spectral-spatial convolutional transformer for hyperspectral image classification," *IET Image Process.*, vol. 18, no. 13, pp. 4328–4340, Nov. 2024.
- [22] M. Ahmad et al., "Spatial-spectral morphological mamba for hyperspectral image classification," *Neurocomputing*, vol. 636, Jul. 2025, Art. no. 129995.
- [23] Z. Pan, C. Li, A. Plaza, J. Chanussot, and D. Hong, "Hyperspectral image classification with mamba," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2024, Art. no. 5602814.
- [24] S. Feng, H. Zhang, B. Xi, C. Zhao, Y. Li, and J. Chanussot, "Cross-domain few-shot learning based on decoupled knowledge distillation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Oct. 2024, Art. no. 5534414.
- [25] Z. Zhong, E. Fini, S. Roy, Z. Luo, E. Ricci, and N. Sebe, "Neighborhood contrastive learning for novel class discovery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, p. 867–875.
- [26] H. Wu, Z. Xue, S. Zhou, and H. Su, "Overcoming granularity mismatch in knowledge distillation for few-shot hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, Jan. 2025, Art. no. 5503517.
- [27] R. Zhang, Y. Zhao, Z. Yin, D. Li, and Z. Wu, "A deep learning-based novel class discovery approach for automatic modulation classification," *IEEE Commun. Lett.*, vol. 27, no. 11, pp. 3018–3022, Nov. 2023.
- [28] V. Joeveek, T. Hemalatha, and K. P. Soman, "Determining an efficient supervised classification method for hyperspectral image," in *Proc. Int. Conf. Adv. Recent Technol. Commun. Comput.*, Oct. 2009, pp. 384–386.
- [29] J. Wen, W. Yan, and W. Lin, "Supervised linear manifold learning feature extraction for hyperspectral image classification," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 3710–3713.
- [30] Y.-C. Hsu, Z. Lv, and Z. Kira, "Learning to cluster in order to transfer across domains and tasks," 2017, *arXiv:1711.10125*.
- [31] Y.-C. Hsu, Z. Lv, J. Schlosser, P. Odom, and Z. Kira, "Multi-class classification without multi-class labels," 2019, *arXiv:1901.00544*.
- [32] M. Yang, Y. Zhu, J. Yu, A. Wu, and C. Deng, "Divide and conquer: Compositional experts for generalized novel class discovery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14248–14257.
- [33] N. Pu, Z. Zhong, and N. Sebe, "Dynamic conceptional contrastive learning for generalized category discovery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7579–7588.
- [34] X. Wen, B. Zhao, and X. Qi, "Parametric classification for generalized category discovery: A baseline study," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 16590–16600.
- [35] S. Vaze, K. Hant, A. Vedaldi, and A. Zisserman, "Generalized category discovery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 7482–7491.
- [36] E. Fini, E. Sangineto, S. Lathuiliere, Z. Zhong, M. Nabi, and E. Ricci, "A unified objective for novel class discovery," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9284–9292.
- [37] P. Gu, C. Zhang, R. Xu, and X. He, "Class-relation knowledge distillation for novel class discovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023.
- [38] P. Gu, C. Zhang, R. Xu, and X. He, "Class-relation knowledge distillation for novel class discovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Aug. 2023, Art. no. 15017.
- [39] W. Li, Z. Fan, J. Huo, and Y. Gao, "Modeling inter-class and intra-class constraints in novel class discovery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 3449–3458.
- [40] K. Han, A. Vedaldi, and A. Zisserman, "Learning to discover novel visual categories via deep transfer clustering," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8401–8409.
- [41] Y. Wang, Z. Chen, and D. Yang, "Self-cooperation knowledge distillation for novel class discovery," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland*: Springer, 2024, pp. 459–476.
- [42] J. An, Z. Du, H. Zhang, and D. Wu, "TSPT: Two-step prompt tuning for class-incremental novel class discovery," *Knowl.-Based Syst.*, vol. 319, Jun. 2025, Art. no. 113603.
- [43] Y. Zhou, H. Zhu, C. Xu, R. Zhang, G. Hua, and W. Yang, "Class-incremental novel category discovery in remote sensing image scene classification via contrastive learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 9214–9225, 2024.
- [44] Z. Li, F. Lu, J. Zou, L. Hu, and H. Zhang, "Generalized few-shot meets remote sensing: Discovering novel classes in land cover mapping via hybrid semantic segmentation framework," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2024, pp. 2744–2754.
- [45] L. Pang, J. Yao, K. Li, and X. Cao, "SPECIAL: Zero-shot hyperspectral image classification with CLIP," 2025, *arXiv:2501.16222*.
- [46] Y. Liu, Y. Li, Z. He, and A. Plaza, "Dual-branch collaborative learning with adaptive feature calibration for hyperspectral image classification," 2024, *arXiv:2403.05451*.