# PU-KBS: A Robust Positive and Unlabeled Learning Framework With Key Band Selection for One-Class Hyperspectral Image Classification

Ziying Liu, Hengwei Zhao, *Student Member, IEEE*, Xinyu Wang, *Member, IEEE*, Shaoyu Wang, Jingtao Li, and Yanfei Zhong, *Senior Member, IEEE*

*Abstract*—Positive and unlabeled (PU) learning is aimed at building a binary classifier to distinguish the target from the background using only the known positive samples, which is an advanced solution for the hyperspectral target detection (HTD) task. However, when PU learning (PUL) meets complex hyperspectral scenarios, there are two main challenges: 1) How to estimate the class prior accurately? The class prior, i.e., the target proportion, is an important prior for PUL to learn the discriminant boundary, but it is difficult to estimate in hyperspectral imagery, due to the interclass spectral similarity and 2) How to remove redundancy and improve the discriminative features of the target? The diagnostic spectral feature extraction is important for the weakly supervised PUL models as it can help with separating the target from the background. In this article, to tackle these challenges, a robust PUL framework with key band selection (PU-KBS) is proposed, which is modeled as an end-to-end and class prior free PUL framework, where the accurate class prior and the most discriminative key band subset are jointly initialized and iteratively updated until reaching the optimal result by evolutionary search. Meanwhile, a deep PUL detector is introduced for guiding the subsequent search direction and discriminative deep feature extraction. The proposed PU-KBS framework was verified using different hyperspectral datasets, where accurate class prior estimation, diagnostic spectral characteristics, and robust detection results could be obtained simultaneously by the PU-KBS framework. Furthermore, the improvement in band selection interpretability and detection performance was proven experimentally.

*Index Terms*—Hyperspectral image, one-class classification (OCC), positive and unlabeled learning (PUL).

## NOMENCLATURE

| | |
|---|---|
| $NP$ | Population size. |
| $N$ | Number of bands. |
| $M$ | Number of selected bands. |
| $E$ | Number of employed/onlooker bees. |
| $S$ | Number of scout bees. |
| $x_{im}$ | $m$th band selected from the $i$th solution. |
| $\mathbf{U} \in \mathbb{R}^{1 \times M}$ | Upper bounds of the band subspace and class prior. |
| $\mathbf{L} \in \mathbb{R}^{1 \times M}$ | Lower bounds of the band subspace and class prior. |
| $F_i$ | $i$th cost function of the possible solutions. |
| $\pi_P$ | Class prior of PU learning. |

## I. INTRODUCTION

**H**YPERSPECTRAL imagery collects abundant spectral information of different ground objects, so it has unique advantages in target detection, with broad application potential in fields such as camouflage target recognition [1], specific crop extraction [2], plant disease detection [3], invasive species identification [4], and disaster area extraction [5].

In the past decades, various methods have been proposed for identifying a single target/class of interest from hyperspectral remote sensing images, which is essentially a binary classification (BC) problem [6], i.e., separating the target class from the background. However, in complex scenarios, it is difficult to annotate complete negative samples containing all the background classes, resulting in only positive samples being obtained [7]. Assuming that only the spectra of the positive class are known, essentially, it is a one-class classification (OCC) problem [6] while is defined as a hyperspectral target detection (HTD) problem [8] in the hyperspectral search field, both of which are aimed at building a binary classifier with only target/positive samples. Their differences and details are discussed in Section II-B.

In this study, we mainly focus on OCC methods for identifying a specific target/class from the background, which can be roughly divided into main categories, according to the data used in the training: 1) the first category is positive ($P$) learning, where only positive samples are involved in the

training, e.g., one-class support vector machine (OCSVM) [9], support vector data description (SVDD) [10] and 2) the second category is positive and unlabeled (PU) learning, where unlabeled samples are introduced along with positive samples to train a binary classifier [11], [12], e.g., biased SVM (BSVM) [7], classifier-independent PU learning (PUL) [13], positive and background learning (PBL) [14], deep PUL (deepPUL), deep PBL (deepPBL) [15], unbiased PU (uPU) [16], and nonnegative PU (nnPU) [17]. Compared with $P$ learning, PUL can further improve the identification accuracy since it mines additional information for the background/negative class from a large amount of unlabeled data, which helps to train a more robust binary classifier.

Multiple ways have been developed for mining negative priors/knowledge in PUL, e.g., selecting negative samples with high confidence from unlabeled data [18], a biased classifier [7], post-threshold calibration [13], [14], [15], and risk estimation [16], [17]. However, most PUL methods require prior knowledge of the target, termed the class prior ($\pi_P$), which is the proportion of the target in the scene [19]. This is assumed to be known to help to learn an optimal PU classifier. Taking the widely used and advanced unbiased risk estimation PUL method [17] as an example, the class prior $\pi_P$ is an important parameter for PUL since it is directly used in calculating the risk, which varies from 0 to 1.

When applying PUL in a complex hyperspectral remote sensing scenario that features abundant land-cover classes in the background and high interclass spectral similarity, e.g., agriculture and forestry, there are still two main challenges encountered: *1) How to estimate the class prior accurately in a complex scene?* As proved in the risk defined in (2), the *class prior* ($\pi_P$) is a crucial parameter. The experiments depicted in Fig. 1(a) also illustrate that the accuracy of a PU classifier depends on the estimation of $\pi_P$. $\pi_P$ is usually defined manually in advance, but can be inaccurate in real applications [20] and *2) How to remove redundant information and improve the discriminative features of the target?* Hyperspectral data contain a significant amount of redundant information, and selecting key features is crucial for applying PUL to hyperspectral imagery, since it is a weakly supervised task [21]. As an example, as shown in Fig. 1(b), the accuracy of the PU classifier is also sensitive to the number of bands, and the best accuracy is not obtained in the case of using all of the bands.

Although the two main challenges could be roughly solved by class prior assumption and band selection method respectively, the two-stage solution would cause error accumulation and may further affect the detection performance. Considering class prior estimation and key band selection (KBS) are essentially both parameter optimization problems, it could be modeled in the same PUL framework as a multiparameter joint optimization problem, which can avoid cumulative errors through an end-to-end approach, helping with enhancing the separation of target and background.

In this article, to address the above problems, we propose a robust PUL framework with KBS (PU-KBS) for one-class hyperspectral image classification, which is aimed at modeling and solving three tasks in one framework jointly, i.e.,
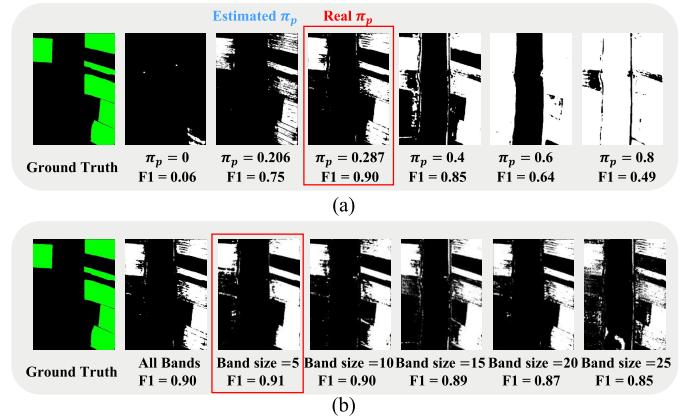


Fig. 1. Main challenges when applying PUL to hyperspectral imagery. (a) Class prior estimation. (b) Redundant information removal.

PUL, class prior estimation, and KBS. Specifically, for a hyperspectral image and several positive samples, PU-KBS can autonomously obtain a robust PU classifier, an accurate class prior, and a key band subset for the target of interest. In PU-KBS, to extract discriminative features of the target and reduce the redundancy, a KBS operation is first designed to find the most discriminative and diagnostic spectral bands of the target, while a deep PU classifier follows to further extract the discriminative deep features. As for the class prior, it is modeled as a parameter search problem, which is encoded with the key bands' numbers in the same vector, randomly initialized and iteratively updated until reaching the optimal result. By wrapping the KBS and the class prior estimation together with a deep PU classifier, the proposed PU-KBS framework can be treated as a class prior free PU classifier, as well as a target detection driven band selector. The main contributions of this work are summarized as follows:

1) A Robust PUL Framework: The proposed PU-KBS framework is designed for complex scenarios, and can autonomously identify targets in a data-driven manner without any manual prior. It is more robust than the advanced HTD and OCC methods, which was proven in the experiments conducted in this study.

2) Joint Estimation of the Class Prior with PUL: The class prior, i.e., the proportion of the target, is modeled and jointly estimated in PU-KBS. The experiments conducted in this study showed that the class prior estimation and the PU classifier can promote each other, and can reach an optimal result simultaneously.

3) Joint Determination of the Key Band Subset with PUL: The diagnostic spectral characteristics of the target of interest can be jointly obtained in PU-KBS. Ablation experiments confirmed that joint modeling is better than separate modeling. In addition, the selected key bands are shown to increase the separability of the target and background.

The rest of this article is organized as follows. Section II gives the related background. Section III describes the general idea and the details of the proposed PU-KBS framework. The experiments and analysis with hyperspectral datasets and practical hyperspectral data are presented in Section IV, while

the ablation experiments and a discussion are presented in Section V. Finally, Section VI draws our conclusion.

## II. RELATED WORKS

### A. OCC for Hyperspectral Imagery

OCC is an efficient specific target class extraction method that learns a binary classifier with only the target of interest samples, where the class of the target of interest can be regarded as the positive class, and the other kinds of classes are all classified as the negative class. OCC classifiers can be divided into two categories, according to whether they need unlabeled samples for training: 1) positive ($P$) learning classifiers and 2) PUL classifiers. The detailed formula for OCC is expressed as follows. $x$ denotes the input image, $y = 1$ denotes the positive target data, and $y = 0$ denotes the other negative data. Therefore, the marginal distribution of the positive class, negative class, and unlabeled data (both positive and negative) can be denoted, respectively, as $P_P(x) = P(x|y = 1)$, $P_N(x) = P(x|y = -1)$, and $P(x)$. $\pi_P = P(y = +1)$ represents the class prior. The objective of a $P$ classifier is to learn a binary classifier $f$ from the positive samples $P_P(x)$, while the objective of a PU classifier is to learn a binary classifier $f$ from the positive samples $P_P(x)$ and unlabeled samples $P(x)$.

*1) Positive (P) Learning:* $P$ learning only requires positive examples for training, in methods such as OCSVM and SVDD. SVDD distinguishes the target by modeling the minimum volume hypersphere in the feature space, while OCSVM maximizes the distance from the separating hyperplane to the origin in the feature space corresponding to the kernel. Deep SVDD (DSVDD) [22] was further derived with the development of a deep learning model. However, the decision boundary of $P$ classifiers is difficult to find, and empirical parameters are also required to balance the degree of fit.

*2) PUL:* In order to further improve the OCC accuracy, unlabeled data can be added to help the training, which is denoted as PUL [23]. The unlabeled data can include both positive and negative samples.

PUL [11], [12] has already been applied in many fields of remote sensing. The representative heuristic method uses a two-step strategy to identify reliable negative samples (or positive samples) from unlabeled samples and then trains a binary classifier, using the positive samples as well as the selected negative samples. However, the detection result can be seriously affected by the reliability of the selected negative samples.

Another one-step approach is to train a binary classifier with both PU samples, treating the unlabeled samples with smaller cost weights as noisy (positive samples included) negative samples. This approach is used in methods such as the BSVM method, but extra training is needed for the cost parameter estimation. Based on the assumption of unlabeled samples as weighted positive samples and negative samples, a PUL framework was introduced. However, this method may be contrary to the actual data distribution, owing to the assumption of both types of samples being completely random. The PBL [14] framework was further developed,

which is more suitable for application in remote sensing, but it does rely on a constant to calibrate the positive class posterior probability. These methods can be transformed into the deepPUL and deepPBL methods [15] by introducing a deep learning classifier.

Risk estimation is a necessary part of the PUL task. In the ideal case where the negative samples are known, the risk estimator of the traditional Bayesian binary classifier is defined as

$$\widehat{R}_{PN}(f) = \pi_P \widehat{R}_P^+(f) + (1 - \pi_P)\widehat{R}_N^-(f) \tag{1}$$

where $f$ represents the classifier, $\widehat{R}_P^+(f)$ and $\widehat{R}_N^-(f)$ denote the risk of classifying positive samples as positive and classifying negative samples as negative, respectively. However, in PUL, due to the lack of negative samples, referring to the equation $P(x) = \pi_p P_P(x) + (1 - \pi_p)P_N(x)$, $\widehat{R}_N^-(f)$ in (1) is replaced using $\widehat{R}_U^-(f)$ and $\widehat{R}_P^-(f)$ in the following equation, which represent classifying unlabeled samples as negative and positive samples as negative, respectively, and are weighted by the class prior $\pi_P$:

$$\widehat{R}_{PU}(f) = \pi_P \widehat{R}_P^+(f) + \widehat{R}_U^-(f) - \pi_P \widehat{R}_P^-(f) \tag{2}$$

which represents the unbiased risk estimator (uPU) of PUL.

*3) The Problem When OCC is Applied to Hyperspectral Imagery:* All of the one-step PUL-based OCC methods mentioned above require a class prior to provide help for the negative sample acquisition, which assigns weights to the unlabeled data as a preprocessing step [11]. Meanwhile, the true value of the class prior is unknown in practical applications, and is generally estimated with an estimation method. However, the class prior assumption is usually uncertain, due to the spectral variability of hyperspectral imagery, and this estimation error can affect the OCC results.

### B. OCC Versus Other Related Concepts

For identifying a specific target, there are some similar concepts to OCC in the field of hyperspectral image processing, such as BC and HTD. In this part, we mainly focus on the comparison of these concepts and highlight the differences between them.

*1) OCC Versus BC:* OCC is a special case of BC to solve the problem of single-class identification. As shown in Fig. 2(a), supervised BC learns a standard binary classifier using both positive and negative annotations to distinguish between the positive class and negative class [24]. In contrast, as shown in Fig. 2(b), OCC learns a binary classifier using only positive annotations, to distinguish the target (positive class) from the background (negative class). Therefore, OCC is a kind of weakly supervised BC task, which is more challenging, due to the lack of negative samples [11].

In complex scenes, OCC can greatly save on the annotation workload since only a small number of positive samples are required [2]. Meanwhile, in contrast, supervised BC usually requires annotation of all the negative classes, which is time-consuming and difficult to achieve in reality [25]. For example, when mapping invasive tree species in a forest, in addition to the samples of the target of interest, a complete set of negative samples containing all the background tree
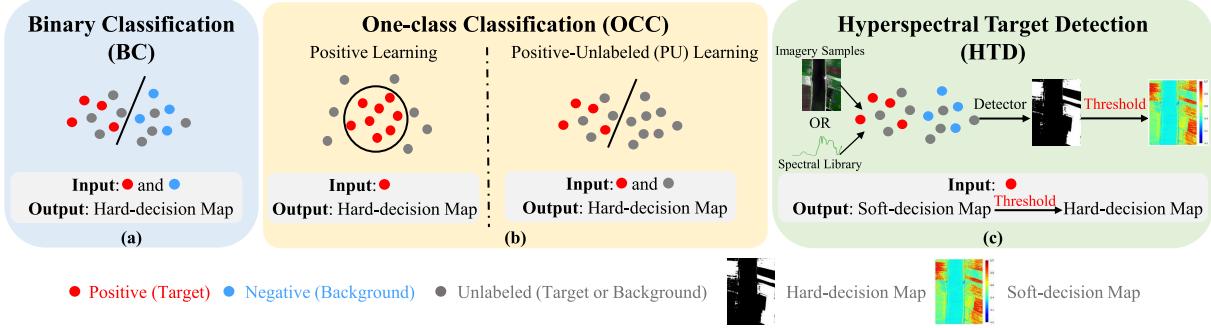
Fig. 2. Comparison between OCC and related concepts (a) BC, (b) OCC, and (c) HTD.

species is required for supervised BC, which is costly for labeling. In contrast, OCC has a natural advantage in complex scenes [4].

*2) OCC Versus HTD:* As for HTD, only the HTD methods using partial target knowledge [26] are compared in this section. As shown in Fig. 2(c), first, the HTD methods usually generate an intensity or soft classification map by the detector [27], where the target responses are enhanced while the background responses are suppressed [8], which is followed by additional thresholds to sperate target from the background [28]. This is because HTD was originally used to identify targets from low spatial resolution data with mixed pixels. In contrast, OCC methods are aimed at obtaining a certain decision boundary, so that their output is usually a hard decision map of targets [7]. Generally speaking, OCC and HTD are similar tasks that are aimed at separating the target (positive class) from the background with only positive samples or spectral signatures. OCC is a broader concept in machine learning, with extensive applications in the fields of text classification [29], medical diagnosis [30], recommendation systems [31], etc.

With the development of hyperspectral imaging and platforms, hyperspectral images with a large volume and high spatial resolution are leading to new challenges, such as the severe intraclass spectral variability and interclass spectral similarity [32], [33], [34], and thus a more robust target detector is necessary. Taking invasive tree species mapping as an example, the spectra of the target class have high variability due to the lighting, shadows, and the texture of the canopy, while the spectra between the target and background (other tree species) have high spectral similarity, due to their similar basic material compositions [35], [36]. Recently, deep learning has shown potential due to its automatic feature extraction ability [27], [37]. The advanced deep PUL-based OCC methods are possible solutions for complex scenarios, where the substantial unlabeled samples contain both positive and negative information, which helps with training a more robust classifier. Deep features can also suppress the spectral variability to some extent [3], [38], and could well be a major trend of robust HTD in the future.

## III. PROPOSED METHOD

### A. PU-KBS Framework

The robust PU-KBS (see Fig. 3) is discussed in this section, which is aimed at automatically determining the class prior and

the key bands during the OCC process. As shown in Fig. 3, for a given hyperspectral cube $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{P \times N}$ ($P$ pixels and $N$ bands) and a few positive samples (unknown class prior $\pi_P$), the proposed PU-KBS framework can obtain the optimal results and the key bands that reveal the diagnostic features of the target of interest. Meanwhile, the essential parameter of PUL—$\pi_P$—is encoded and jointly estimated during the learning process. Specifically, PU-KBS includes two main phases: 1) a class prior $\pi_{Pi}$ and a band subset $\mathbf{X}_i = [\mathbf{x}_1, \ldots, \mathbf{x}_M] \in \mathbb{R}^{P \times M}$ are first randomly initialized within the search space to allow the deep PUL network to learn a classifier and 2) search and update of the new class prior $\pi_{Pi+1}$ and the new band subset $\mathbf{X}_i$ with the current optimal fitness by the use of evolutionary search. These phases are iterated until an optimal PU classifier is achieved.

It should be noted that the proposed method mainly focuses on the following two aspects. The first aspect is how to estimate the class prior jointly during PUL. This is an important prior and sensitive parameter for PUL, which is usually assumed to be known but is very hard to estimate manually. As for PUL, an important feature of our work is the exploration of PUL with a known class prior to PUL without a class prior, instead of designing a new PU network. Second, PU-KBS is aimed at obtaining the key bands of the target of interest, which are discriminative compared to the background. This, on the one hand, can help to reduce the redundant information for PUL and, on the other hand, the central wavelength of the optimal bands has guiding value for the design of an economic multispectral sensor aimed at specific targets.

### B. Encoding and Initialization of the Class Prior and Key Bands for Joint Learning

In order to optimize the class prior $\pi_{Pi}$ and key band subset $\mathbf{x}_i \in \mathbb{R}^{1 \times M}$ simultaneously, the indices for the $\pi_{Pi}$ value (the $(M + 1)$st dimension) and the positions of each key band (the first $M$ dimensions) are encoded in the modeling process, which are combined into the vector $x_{im}$ and randomly initialized and iteratively optimized together in their respective search spaces. In addition, the search space for the key bands is determined by a coarse-to-fine neighborhood grouping strategy, which can help to speed up the search and reduce the possibility of correlation within the selected bands.

Specifically, since a large search space for the spectral range would lead to low search efficiency, in order to accelerate the speed of the search, as shown in Fig. 4, a coarse-to-fine
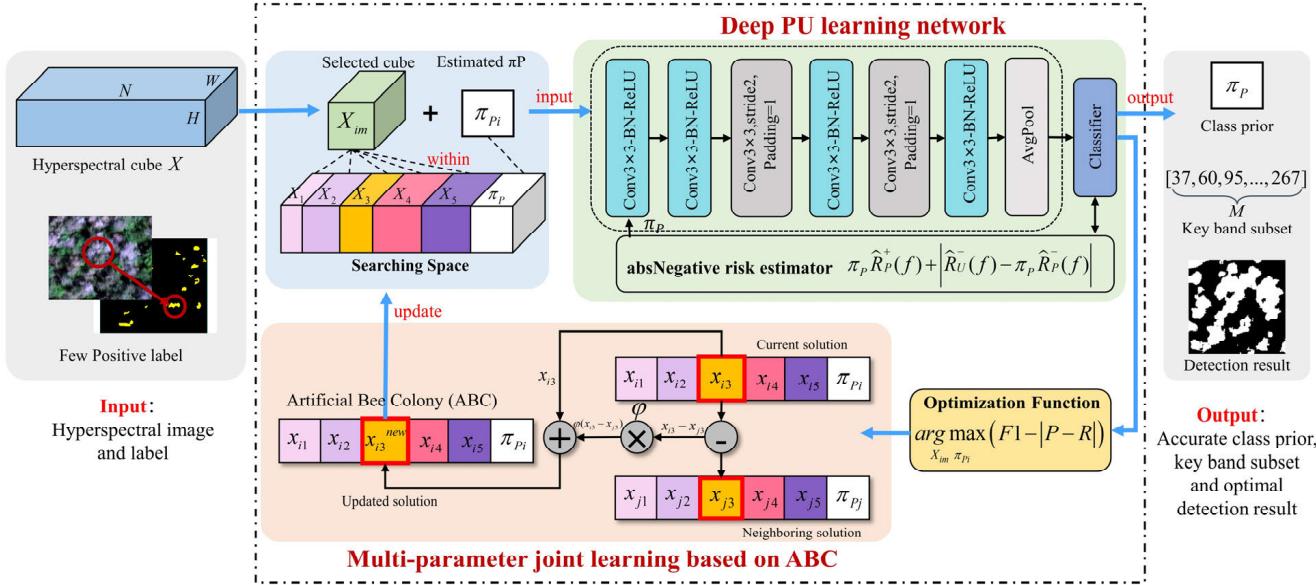
Fig. 3. Framework of the proposed robust PU-KBS. Two main phases are included: 1) encoding and initialization phase and 2) search and update phase.
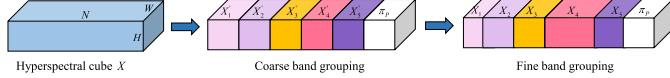


Fig. 4. Schematic of neighborhood band grouping.

neighborhood band grouping [39] strategy is used to determine the limit of the selected key band subset by fully exploring the bands' spectral contextual information.

The neighborhood band grouping can be roughly divided into coarse and fine grouping parts. For the coarse band grouping, the band space of $\mathbf{X}$ is divided into $M$ coarse groups as evenly as possible, which is a uniform grouping process. For the fine band grouping, the aim is to obtain a more accurate grouping result by refining the coarse result based on the similarity of adjacent bands. By comparing the similarity between the current band and the cluster center by the Euclidean distance, the band group label $\mathbf{T} \in \mathbb{R}^{1 \times N}$ is iteratively updated as the nearest label, so as to redivide the band space with lower similarity and redundancy. The bands of the $m$th group can be obtained from the following equation:

$$\mathbf{X}_m = \{x_i\}, \quad \mathbf{T}_i = m. \tag{3}$$

Therefore, the initial position of each band is determined within the search subspace, and for the class prior parameter, this is (0,1). The 0th generation of solutions is initialized randomly between the upper bound $\mathbf{U} \in \mathbb{R}^{1 \times M}$ and lower bound $\mathbf{L} \in \mathbb{R}^{1 \times M}$, which can be defined as shown in the following equation:

$$x_{im} = \mathbf{L}_m + \text{rand}(0, 1)(\mathbf{U}_m - \mathbf{L}_m) \tag{4}$$

where $i = 1, 2, \ldots, NP; m = 1, 2, \ldots, M+1$. There are $NP$ solutions, each containing $M$ selected bands, plus an increased dimension for the class prior. We take the $i$th solution as an example, where $x_{im}$ denotes the $m$th band selected from its

[1]http://rsidea.whu.edu.cn/resource_sharing.htm

subspace for the band subset $\mathbf{x_i} \in \mathbb{R}^{1 \times M}$ in consequence, while $x_{iM+1}$ indicates the estimated class prior $\pi_{Pi}$.

### C. Deep PUL Classifier

Any detection model can be embedded in the detection phase of the proposed PU-KBS framework, which can help to obtain the optimal search parameters (class prior and key band subset) as well as the optimal training parameters for the detector by guiding the direction of the joint optimization process. In addition, synthesizing the previous analysis and discussion (in Section II-B), due to the excellent performance of the risk estimation-based PUL methods, a deep PUL network module [4] (the ITreeDet code is available at) is introduced as the target detector for the optimization function calculation, which can automatically extract robust features from the imagery, so as to reduce the impact of the spectral variability problem.

Specifically, as presented in (5), the risk estimator is set as absNegative (absPU), which can avoid network overfitting for PU samples by keeping the risk estimation of the negative class greater than 0 through the absolute term. Meanwhile, for the connection of the two main phases, the network input $\mathbf{X_i}$ is filtered from the original input $\mathbf{X}$ by the searched candidate $M$ bands, and the searched class prior $\pi_{Pi}$ is used as the critical parameter of PUL

$$\tilde{R}_{\text{PU}}(f) = \pi_P \widehat{R}_P^+(f) + \left| \widehat{R}_U^-(f) - \pi_P \widehat{R}_P^-(f) \right|. \tag{5}$$

In general, the deep PU network regards the input candidate class prior and key band subset as the model parameters and image data reference, respectively. Meanwhile, their detection performance is used as a guide for the further search direction.

### D. Searching Class Prior and Key Band Search by Evolutionary computation

Although the current optimal detection result can be obtained by inputting the initial candidate solution into the

deep PU network, it will not be the global optimal solution. The deep PU classifier is still sensitive to these parameters (class prior and key band subset). Therefore, it is necessary to continue updating them as the input classifier parameters, which can further optimize the training parameters of the classifier and obtain a better detection result. For the updating of these parameters, evolutionary computation (EC) [40] is introduced for the optimization of the encoded vector $x_{im}$, which is achieved by estimating the global optimal solution through greedy search within the iterations. Two main problems are to be solved, including: 1) how to build the optimization function and 2) how to realize the optimization search. The optimization function is based on a new discovery, while the optimization search is achieved by an advanced evolutionary search method.

*1) Optimization Function:* In order to obtain the optimal parameters (class prior and key band subset) and balance the accuracy metrics, a new optimization function is designed using three evaluation metrics, based on the discovery of the class prior [41], which is also the bridge between the detection model and the optimization search. The involved metrics are the precision ($P$), recall ($R$), and $F1$-score ($F1$), calculated by the true positive (TP), false positive (FP), and false negative (FN) pixels. The corresponding formulas and the influence of the class prior on the three metrics at an interval of 0.01 are shown in Fig. 5. As shown, when the estimated $\pi_P$ is less than the truth, $P$ is larger than $R$. In contrast, when the estimated $\pi_P$ is larger than the truth, $P$ is smaller than $R$. The larger the estimated bias, the worse the $F1$, i.e., the $F1$ is basically optimal when the estimated $\pi_P$ approximates the truth. As per these change rules, the optimization function (OPT) can be calculated as shown in the following equation:

$$\text{OPT} = \max\left[\frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} - \left|\frac{\text{TP}(\text{FN} - \text{FP})}{(\text{TP} + \text{FP})(\text{TP} + \text{FN})}\right|\right] \quad (6)$$

when the class prior approaches the truth, the absolute value of (precision–recall) is smaller, which can be used as an interference term to help the EC search quickly and accurately locate the actual range of the class prior, while also helping with balancing the $P$ and $R$ values, thus further improving the detection results.

*2) Optimal Band Subset and Class Prior Search:* For the joint optimization of the class prior and key bands, EC search is introduced to automatically optimize these parameters within the set search space iteratively by the designed optimization function. Here, we take the artificial bee colony (ABC) algorithm [42] as an example, due to the advantage of only a few insensitive parameters being involved.

There are three kinds of bees involved and exchanging information in the ABC algorithm. Employed bees search and update each position of the solution in the neighborhood based on the memorized quality and position information. Onlooker bees receive the memorized information and then choose the solution to exploit by its quality using roulette wheel selection. Scout bees emerge when the trail counter $l$ reaches or exceeds the limit $L = \text{round}(0.6ME)$, where round(.) means the rounding operation. The $i$th optimization function of the
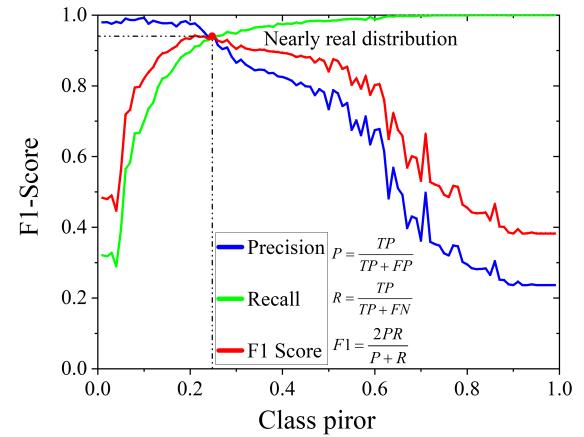


Fig. 5. Influence of the class prior to the detection metrics.

possible solutions is recorded as $\text{OPT}_i$, calculated by the accuracy combination of the detection model. The updated solutions for the iterations in the neighborhood are produced as shown in the following equation:

$$x_{im}^{\text{new}} = x_{im} + \varphi(x_{im} - x_{jm}), \quad j \neq i \quad (7)$$

where $\varphi$ is a random number uniformly distributed in the range $[-1, 1]$ and $j$ is a random solution number that is different from $i$. If the position of the new solution breaks the bound constraint, a bound handling process is performed, i.e., if the position is over/under the upper/lower bounds, it is set as the corresponding bound value. The greedy selection between the original and new solutions is applied using the fitness (8) evaluated by the optimization function (6), so as to update the memorized solution position to a better one, where mean(.) is the averaging operation in (8). Meanwhile, the selection probability (9) is determined by the fitness, and the trial counter $l$ for each solution is reset to 0 if the position is changed; otherwise, the value is increased by one

$$F_i = e^{\frac{\text{OPT}_i}{\text{mean(OPT)}}}, \quad i = 1, 2, \dots, NP \quad (8)$$

$$PB_i = \frac{F_i}{\sum_{i=1}^{NP} F_i}. \quad (9)$$

The employed, onlooker, and scout bees exchange and cycle until the maximum iterations *Iter* is reached, where the best band subset $\mathbf{x_b} \in \mathbb{R}^{1 \times M}$ and the most accurate class prior $\pi_{Pb}$ are output as the optimal solution.

### E. Implementation Issues

Based on the above two phases, a robust PU-KBS is proposed. According to the band subset size $M$, the first $M$ searched upper $\mathbf{U}$ and lower $\mathbf{L}$ bounds of the bands are obtained by coarse-to-fine neighborhood grouping, while the last dimension for $\pi_P$ is in the range of [0, 1]. The fitness of the possible EC search solution is calculated by the optimization function of the detection model, which guides the further search direction. After reaching the set generation *Iter* of the EC search (ABC algorithm), the optimal band subset and the estimated class prior to within the iterations are obtained using the validation set. By verification with the test set under the

## TABLE I
### DETAILS OF THE HYPERSPECTRAL IMAGE DATASETS

| Image attribute | WHU-Hi-LongKou | PWD | Indian Pines |
|---|---|---|---|
| Size | $550 \times 400$ | $1567 \times 717$ | $145 \times 145$ |
| Spectral bands | 270 | 274 | 200 |
| Wavelength range | 400–1000 nm | 400–1000 nm | 400-2500 nm |
| Training samples | 100 | 500 | 40 |
| Unlabeled samples | 4000 | 5000 | 400 |
| Validation samples | 100/100 | 500/500 | 40/40 |
| Test samples | 63112/140530 | 64486/208371 | 478/9771 |
| $E$ | 30 | 30 | 30 |
| $S$ | 10 | 10 | 10 |
| *Iter* | 20 | 10 | 20 |
| *Epochs* | 200 | 300 | 200 |



Fig. 6. WHU-HI-LongKou dataset. (a) Image. (b) Ground truth. (b) Spectral curves.



Fig. 7. PWD dataset. (a) Image. (b) Training set. (c) Test set. (d) Spectral curves.

*Epochs* iterations setting, the corresponding evaluation indicators and detection result are the final output. The pseudo-code for the PU-KBS framework is listed in Algorithm 1.

---

**Algorithm 1** Pseudo-Code of the PU-KBS Framework

1. **Input** $\mathbf{X} \in \mathrm{R}^{P \times N}$, $M$;
2. Partition bands into subspaces as (3), determine boundary $\mathbf{U}$ and $\mathbf{L}$;
3. **While $t < $ Iter**
   Initialize solutions (4), calculate loss function (6), set t=0;
4. For employed bees, update solutions as (7), update trials l;
5. For onlooker bees, update solutions as (7) and (9), update l;
6. **If $l \geq L$**
   For scout bees, initialize solutions as (4), reset l=0;
7. t=t+1;
   **End While**
8. **Output:** $\mathbf{x_b}$, $\pi_{Pb}$, detection accuracy and detection result.

---

In the designed framework, subspace grouping helps reduce the possibility of similarity within the selected bands and improves the search efficiency. The ABC algorithm involves only a few parameters and also has a low sensitivity to these parameters, and the positive and negative feedback mechanism effectively avoids falling into local optima. What is more, the ABC algorithm is suitable for solving high-dimensional problems. In the deep PUL network, only the labeled samples of the target class and unlabeled information are required for the training, and the robust feature extractor can decrease the impact of spectral variability.

## IV. EXPERIMENTS AND ANALYSIS

### A. Hyperspectral Datasets

In order to prove the universality of the proposed framework, two types of hyperspectral datasets were used in the experimental validation: the WHU-HI-Longkou dataset and the Indian Pines dataset at the dataset level and the pine wilt disease (PWD) detection dataset at the application level. The details of the hyperspectral image datasets are given in Table I.

1) *WHU-Hi-LongKou Dataset:* The WHU-Hi-LongKou dataset [32] is a standard dataset for precise crop classification. The broad-leaf soybean class was chosen as
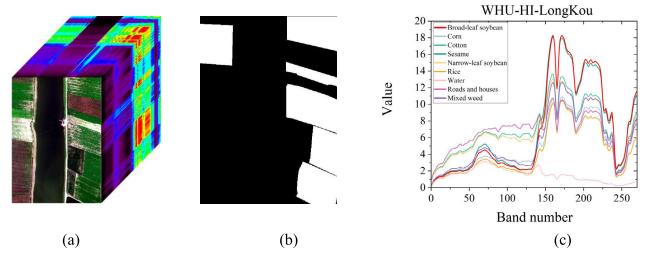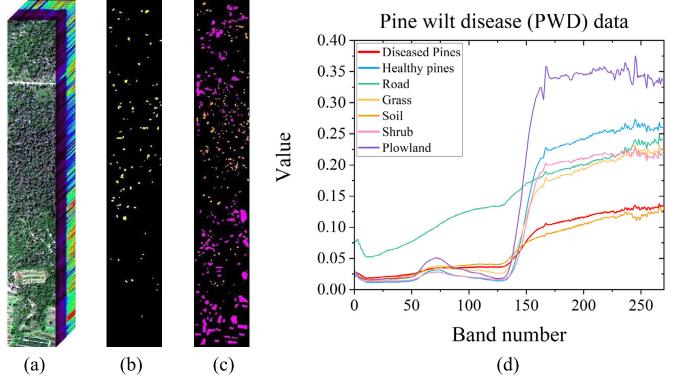
the target class, for which an overview and the spectral curves are provided in Fig. 6. The dataset details are provided in Table I.

2) *PWD Detection Dataset:* PWD is a widespread and hard-to-control irreversible disease leading to rapid wilting of the infected pine trees, the spread of which has caused serious damage to both the economy and ecological security [43]. Therefore, early prevention and treatment are particularly important. Due to the large-scale and multitemporal properties, remote sensing methods have gradually taken the place of traditional manual methods of detection. The detection of pine trees with PWD is a kind of target detection problem at the application level. The PWD dataset (see Fig. 7) was acquired at Yantai Hill, Hexi, Shandong, China [3]. The dataset details are provided in Table I. The main pine species in the study area are *Pinus massoniana* Lamb and *Pinus armandii* Franch, some of which have mild or moderate PWD (yellow/brown needles), with the healthy pines having green needles. In addition to the pine species, there is also soil, road, vegetation (e.g., grasses and shrubs), etc.

3) *Indian Pines Dataset:* Indian Pines dataset is a classical hyperspectral dataset containing 16 land-cover classes with similar spectrum. The hay-windrowed class is selected as the target class, for which an overview and the spectral curves are provided in Fig. 8. The dataset details are provided in Table I as well.

### B. Experimental Settings

1) *Comparison Methods:* We compared the proposed PU-KBS framework with the classical and advanced
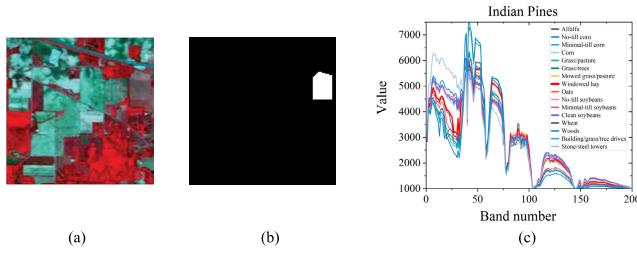
Fig. 8. Indian Pines dataset. (a) Image. (b) Ground truth. (b) Spectral curves.

P, PU, and HTD methods that were mentioned in Section II, including: 1) P learning classifiers, i.e., OCSVM and DSVDD; 2) PUL classifiers, i.e., BSVM, PUL, PBL, deepPUL, deepPBL, and the risk estimation based methods—uPU and nnPU; and 3) HTD methods, i.e., the matched filter (MF) [44], constrained energy minimization (CEM) [45], ensemble-based CEM (ECEM) [46], and a dual sparsity constrained (DSC) approach [47]. The thresholds of the HTD methods were selected using Otsu's thresholding method [48].

2) *Metrics:* Four evaluation metrics are used in this article to evaluate the detection performance: precision (P) (10), recall (R) (11), F1-score (F1) (12) (as described in Section III-D), and the area under curve (AUC). Precision represents the user's accuracy, which indicates the correct detection rate of the pixels that are certified as positive. Recall is the producer's accuracy, which can reflect the positive proportion of the pixels that are correctly detected. The F1-score is the harmonic average of the precision and recall, which can measure the comprehensive result of the model. The AUC metric computed by the receiver operating characteristic (ROC) curve is mainly used to assess the model's generalizability and confidence. Due to the comprehensiveness of the F1-score metric, it is considered the main detection accuracy reference. All the experimental accuracy results are the average of five repeated measurements. Meanwhile, all the experiments were implemented in Python 3.8 and run on Linux CentOS 7.2.1511, using an Nvidia Tesla P100 GPU

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$F1\text{-score} = \frac{2PR}{P + R}. \tag{12}$$

### C. Experiment Results and Analysis for the WHU-Hi-LongKou Dataset

In this experiment, the Longkou dataset, which is a standard dataset, was used to analyze the robustness of the different P, PU, and HTD methods under different numbers of bands, where the band subset size was set from 2 to 10 with an interval of 2. Note that, for a fair comparison, the band subsets used in the comparison methods were selected using the improved sparse subspace clustering (ISSC) method [49] (which has a suboptimal performance, as explained in Section V-B).

The quantitative results obtained under different numbers of selected bands are shown in Fig. 9(a)–(d). The quantitative results obtained using six selected bands are listed in Table II (left), where the top-three results are highlighted with different colors, and the class prior parameter of the compared PUL methods is set as both the estimated value and the true value (as explained in Section V-A). PU-KBS obtains the best F1-score values and moderate AUC metric results, while also obtaining high P and R values at the same time, which indicates that PU-KBS shows good model robustness under different band subset sizes. Moreover, inaccurate class prior estimation leads to an overall decrease in the detection accuracy, i.e., the values of P and R are more unbalanced, and P is higher and R is lower, due to the low and inaccurate class prior estimation (as shown in Fig. 5). Furthermore, the moderate AUC values indicate the stability and relatively high confidence of the PU-KBS detection model. The performance of the P classifiers is usually lower than that of the PU classifiers, due to the lack of negative samples as a reference. In addition, the qualitative results obtained using six selected bands are shown in Fig. 9(e)–(t). The detection region of PU-KBS is more complete, while the distribution maps of the traditional methods generally contain some noisy points, due to the limited feature representation ability.

### D. Experiment Results and Analysis for the PWD Detection Dataset

In this experiment, the PWD dataset, which is an application dataset, was applied to analyze the robustness of the different P, PU, and HTD methods under different numbers of bands in complex scenes, where the band subset size was set from 5 to 30 with an interval of 5. Similarly, for a fair comparison, the band subsets used in the comparison methods were selected using the ISSC method (which has a suboptimal performance, as explained in Section V-B).

The quantitative results under different numbers of selected bands are shown in Fig. 10(a)–(d), and the quantitative results obtained using ten selected bands are listed in Table II (medium). The top-three results are again highlighted with different colors. The class prior parameter of the compared PUL methods was set as both the estimated value and the true value (as explained in Section V-A). PU-KBS obtains the best F1-score and relatively high R and AUC values under all band subset sizes. Significantly, the R-value is more important than the P-value in disease area detection applications. This suggests that the PU-KBS framework could find as many potentially diseased areas as possible for prewarning monitoring, which would be influenced by inaccurate class prior estimation. Moreover, an abnormal phenomenon of low F1-score with a high AUC value can be found in some of the HTD methods (MF, CEM, and ECEM), which could be related to the error accumulation caused by the detection threshold segmentation step. In addition, the qualitative results obtained using ten selected bands are shown in Fig. 10(e)–(t). Within the manually labeled ground truth, the red box area represents the PWD area, while the yellow box represents the non PWD area. Within the red box area, the misdetection rate of PU-KBS is the lowest, as shown by the most complete PWD area.
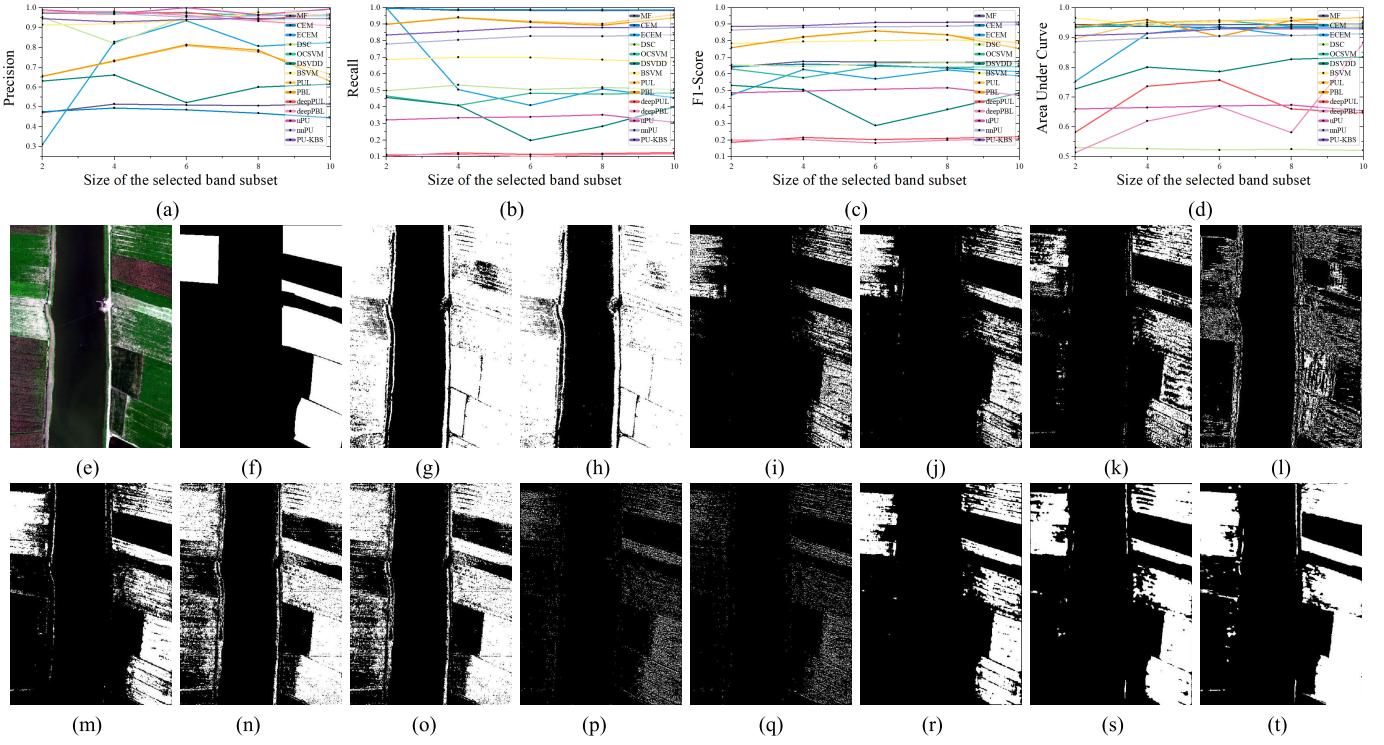
Fig. 9. Detection model comparison results for the WHU-HI-LongKou dataset. (a)–(d) Quantitative accuracy metrics. (g)–(t) Visual detection results obtained using six selected bands, where (g)–(s) are obtained by the selection result of the ISSC method, and (m)–(s) are obtained using the true class prior as the input parameter. (a) Precision. (b) Recall. (c) F1-Score. (d) AUC. (e) Original image. (f) Ground truth. (g) MF. (h) CEM. (i) ECEM. (j) DSC. (k) OCSVM. (l) DSVDD. (m) BSVM. (n) PUL. (o) PBL. (p) deepPUL. (q) deepPBL. (r) uPU. (s) nnPU. (t) PU-KBS.

TABLE II

COMPARISON WITH DIFFERENT TARGET DETECTION METHODS, WHERE THE BAND SUBSET WAS SELECTED BY ISSC, AND THE CLASS PRIOR OF THE PUL METHODS WAS SET AS BOTH THE ESTIMATED AND TRUE VALUES, WITH THE TOP-THREE RESULTS MARKED IN RED, GREEN, AND BLUE, RESPECTIVELY

| Type | Method (ISSC bands) | WHU-HI-LongKou dataset (six selected bands) | | | | The PWD dataset (ten selected bands) | | | | Indian Pines dataset (ten selected bands) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | P | R | F1 | AUC | P | R | F1 | AUC | P | R | F1 | AUC |
| HTD methods | MF | 0.5095 | 0.9839 | 0.6713 | 0.9427 | 0.4195 | 0.9987 | 0.5908 | 0.8762 | 0.8571 | 0.5397 | 0.6624 | 0.9075 |
| | CEM | 0.4857 | 0.9871 | 0.6511 | 0.9327 | 0.4103 | 0.9987 | 0.5816 | 0.8760 | 0.9430 | 0.5188 | 0.6694 | 0.9945 |
| | ECEM | 0.9348 | 0.4100 | 0.5700 | 0.9364 | 0.8999 | 0.2824 | 0.4299 | 0.8219 | 0.9370 | 0.4665 | 0.6229 | 0.9956 |
| | DSC | 0.9549 | 0.5054 | 0.6609 | 0.5222 | 0.2354 | 0.9942 | 0.3806 | 0.5176 | 0.4236 | 0.9916 | 0.5936 | 0.5042 |
| P learning methods | OCSVM | 0.9722 | 0.4834 | 0.6457 | 0.9575 | 0.7267 | 0.4413 | 0.5491 | 0.8477 | 0.9538 | 0.6046 | 0.7401 | 0.9901 |
| | DSVDD | 0.5224 | 0.1983 | 0.2875 | 0.7856 | 0.7809 | 0.8751 | 0.8254 | 0.8111 | 0.1398 | 0.3410 | 0.1983 | 0.4235 |
| PU learning methods (with estimated class prior) | BSVM | 0.9766 | 0.6672 | 0.7928 | 0.9567 | 0.9830 | 0.6211 | 0.7612 | 0.9567 | 0.8357 | 0.9998 | 0.9105 | 0.9993 |
| | PUL | 0.6831 | 0.9582 | 0.7976 | 0.8977 | 0.8262 | 0.2890 | 0.4282 | 0.9504 | 0.9358 | 0.9142 | 0.9249 | 0.9982 |
| | PBL | 0.6854 | 0.9571 | 0.7988 | 0.8986 | 0.8134 | 0.2383 | 0.3687 | 0.9267 | 0.9245 | 0.8452 | 0.8831 | 0.9986 |
| | deepPUL | 0.9726 | 0.1006 | 0.1824 | 0.9084 | 0.9370 | 0.1379 | 0.2405 | 0.6018 | 0.9737 | 0.7631 | 0.8553 | 0.8240 |
| | deepPBL | 0.9762 | 0.0922 | 0.1685 | 0.7427 | 0.9321 | 0.1195 | 0.2118 | 0.4176 | 0.9414 | 0.8164 | 0.8701 | 0.9568 |
| | uPU | 0.9995 | 0.2919 | 0.4518 | 0.6459 | 0.9771 | 0.4434 | 0.6100 | 0.7201 | 0.9571 | 0.9435 | 0.9502 | 0.9706 |
| | nnPU | 0.9821 | 0.6389 | 0.7742 | 0.8168 | 0.8508 | 0.8398 | 0.8452 | 0.8971 | 0.9633 | 0.9331 | 0.9479 | 0.9657 |
| PU learning methods (with true class prior) | BSVM | 0.9360 | 0.6989 | 0.8002 | 0.9518 | 0.8744 | 0.7145 | 0.7864 | 0.9667 | 0.8383 | 0.9979 | 0.9112 | 0.9993 |
| | PUL | 0.8088 | 0.9171 | 0.8596 | 0.9575 | 0.8329 | 0.2926 | 0.4331 | 0.9316 | 0.9084 | 0.9540 | 0.9306 | 0.9987 |
| | PBL | 0.8129 | 0.9124 | 0.8598 | 0.9034 | 0.8260 | 0.2496 | 0.3833 | 0.9093 | 0.9289 | 0.8472 | 0.8862 | 0.9845 |
| | deepPUL | 0.9760 | 0.1128 | 0.2022 | 0.7571 | 0.9478 | 0.1550 | 0.2642 | 0.6491 | 0.9609 | 0.7749 | 0.8569 | 0.9212 |
| | deepPBL | 0.9611 | 0.1006 | 0.1822 | 0.6680 | 0.8712 | 0.1397 | 0.2407 | 0.6126 | 0.9358 | 0.8297 | 0.8705 | 0.9414 |
| | uPU | 0.9994 | 0.3396 | 0.5069 | 0.6697 | 0.9269 | 0.4850 | 0.6368 | 0.7366 | 0.9628 | 0.9289 | 0.9455 | 0.9636 |
| | nnPU | 0.9524 | 0.8284 | 0.8836 | 0.9045 | 0.8642 | 0.8437 | 0.8538 | 0.9013 | 0.9799 | 0.9184 | 0.9482 | 0.9587 |
| Proposed | PU-KBS | 0.9401 | 0.8818 | 0.9100 | 0.9283 | 0.8627 | 0.8657 | 0.8642 | 0.9115 | 0.9817 | 0.9446 | 0.9626 | 0.9718 |

Meanwhile, for the misclassification within the yellow box area, where the test negative samples are classified as positive, the false alarm rate of PU-KBS is also the lowest.

### E. Experiment Results and Analysis for the Indian Pines Dataset

In this experiment, the Indian Pines dataset, which is a standard dataset with limited samples, was applied to analyze the robustness of the different $P$, PU, and HTD methods under different numbers of bands in complex scenes, where the band subset size was set from 5 to 30 with an interval of 5. Meanwhile, for a fair comparison, the band subsets used in the comparison methods were selected using the ISSC method (which has a suboptimal performance, as explained in Section V-B).

The quantitative results under different numbers of selected bands are shown in Fig. 11(a)–(d), and the quantitative results obtained using ten selected bands are listed in Table II (right).
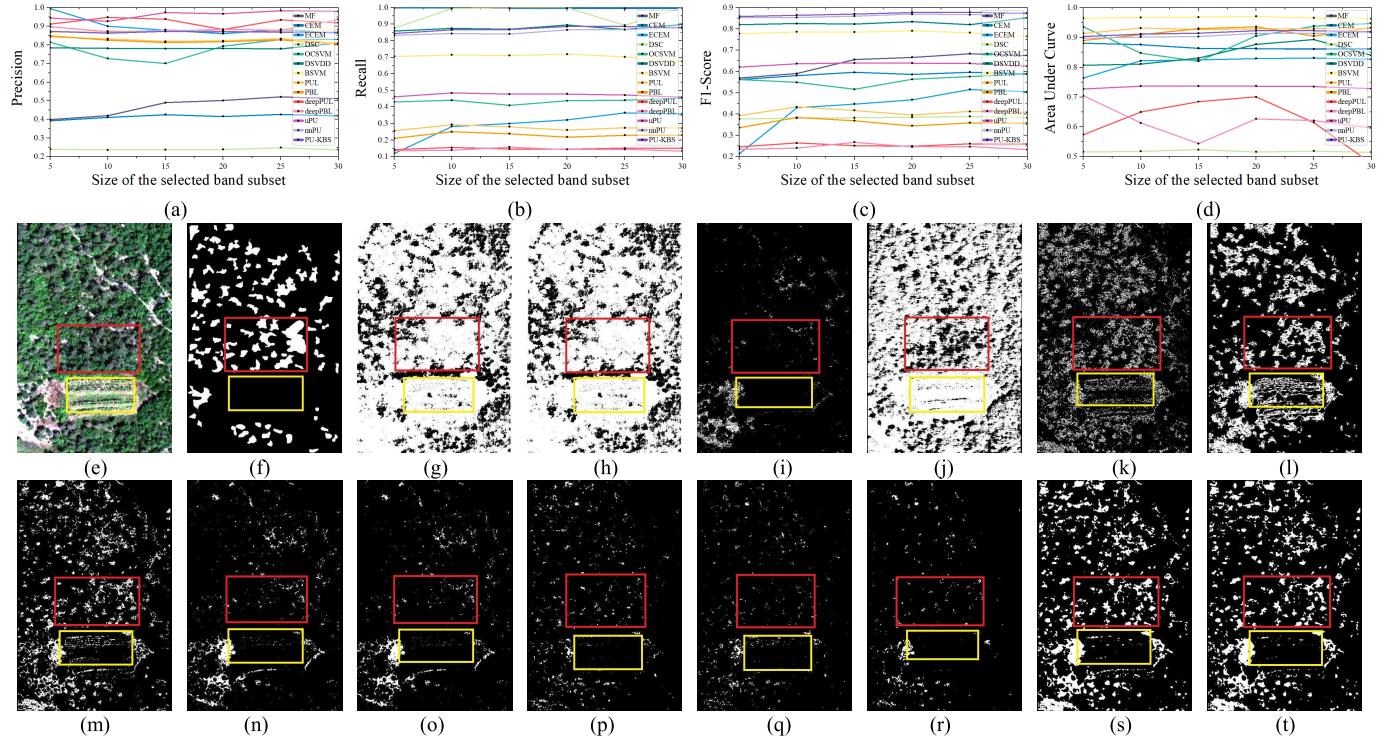
Fig. 10. Detection model comparison results for the PWD dataset. (a)–(d) Quantitative accuracy metrics. (g)–(t) Visual detection results obtained using ten selected bands, where (g)–(s) are obtained by the selection result of the ISSC method, and (m)–(s) are obtained using the true class prior as the input parameter. (a) Precision. (b) Recall. (c) $F$1-Score. (d) AUC. (e) Original image. (f) Ground truth. (g) MF. (h) CEM. (i) ECEM. (j) DSC. (k) OCSVM. (l) DSVDD. (m) BSVM. (n) PUL. (o) PBL. (p) deepPUL. (q) deepPBL. (r) uPU. (s) nnPU. (t) PU-KBS.
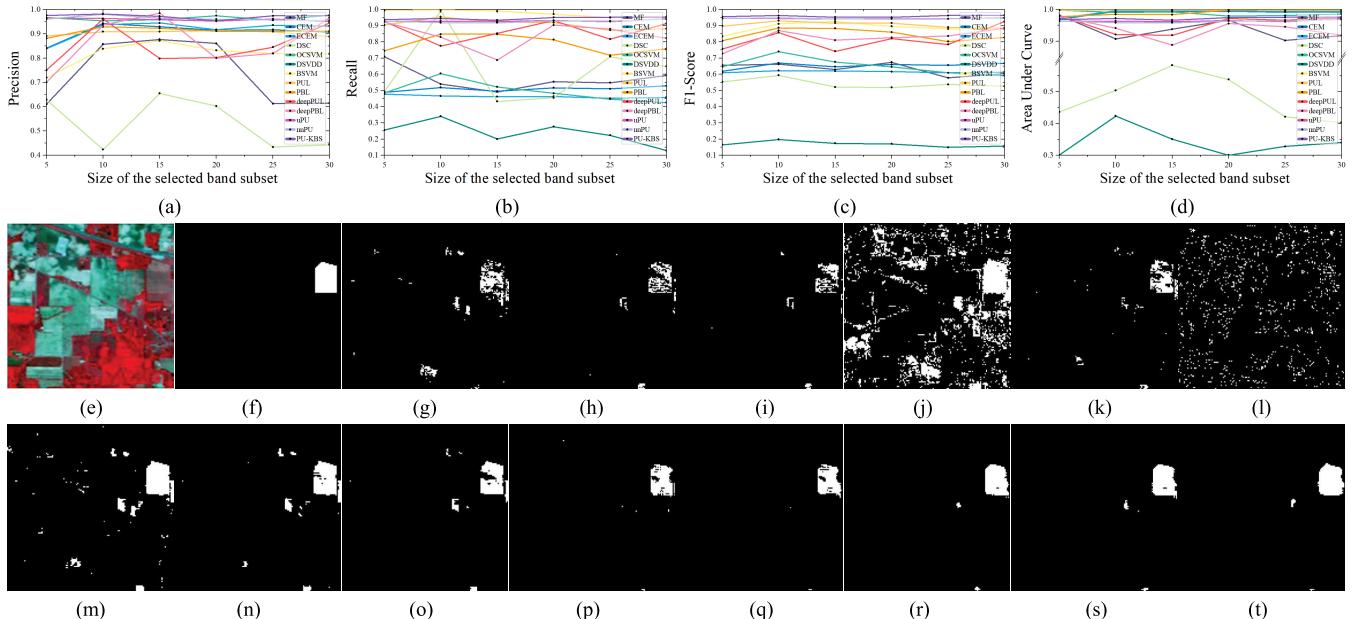


Fig. 11. Detection model comparison results for the Indian Pines dataset. (a)–(d) Quantitative accuracy metrics. (g)–(t) Visual detection results obtained using six selected bands, where (g)–(s) are obtained by the selection result of the ISSC method, and (m)–(s) are obtained using the true class prior as the input parameter. (a) Precision. (b) Recall. (c) $F$1-Score. (d) AUC. (e) Original image. (f) Ground truth. (g) MF. (h) CEM. (i) ECEM. (j) DSC. (k) OCSVM. (l) DSVDD. (m) BSVM. (n) PUL. (o) PBL. (p) deepPUL. (q) deepPBL. (r) uPU. (s) nnPU. (t) PU-KBS.

The top-three results are again highlighted with different colors. The class prior parameter of the compared PUL methods was set as both the estimated value and the true value (as explained in Section V-A). PU-KBS obtains the best $F$1-score and relatively high $P$ and $R$ values under different band subset sizes. It can be found that the detection performance of uPU has an apparent improvement compared with the other two datasets, which is because the risk of the negative class has been estimated to be negative in the other datasets, further resulting in the network overfitting and has influenced the

TABLE III
ESTIMATED CLASS PRIORS FOR THE THREE DATASETS

| WHU-HI-LongKou | | | PWD | | | Indian Pines | | |
|---|---|---|---|---|---|---|---|---|
| Method | Bands | Value | Method | Bands | Value | Method | Bands | Value |
| Truth | -- | 0.287 | Truth | -- | 0.245 | Truth | -- | 0.023 |
| KMPE | -- | 0.206 | KMPE | -- | 0.173 | KMPE | -- | 0.039 |
| PU-KBS | -- | -- | PU-KBS | 5 | 0.229 | PU-KBS | 5 | 0.024 |
|  | 2 | 0.293 |  | 10 | 0.255 |  | 10 | 0.022 |
|  | 4 | 0.278 |  | 15 | 0.260 |  | 15 | 0.032 |
|  | 6 | 0.289 |  | 20 | 0.269 |  | 20 | 0.030 |
|  | 8 | 0.282 |  | 25 | 0.270 |  | 25 | 0.021 |
|  | 10 | 0.278 |  | 30 | 0.262 |  | 30 | 0.027 |

TABLE IV
PARAMETER SETTINGS OF THE COMPARISON METHODS

| Method | Major parameter |
|---|---|
| SSIM | win_size=11 |
| $k$-means | max_iter=300 |
| FNGBS | Z=0.1 |
| ISSC | λ=0.01 |
| SpaBS | λ=0.5 |
| GA | NP=60, w=0.8, Pc=0.9 |
| DE | NP=60, F=0.001, CR=0.3 |
| PSO | NP=60, $c_1$=0.5, $c_2$=0.5 |

detection result seriously. While the estimated negative risk basically remains positive due to the very small samples for training in the Indian Pines dataset, the detection results of uPU are relatively high and similar to that of the nnPU method. In addition, the qualitative results obtained using ten selected bands are shown in Fig. 11(e)–(t). PU-KBS has the most complete detection result within the ground truth area, and also the least misclassified result within the nontarget area.

## V. ABLATION EXPERIMENTS AND DISCUSSION

### A. Ablation Experiment for the Estimated Class Prior

This ablation experiment was conducted to compare the estimated class prior value of PU-KBS under different band subset sizes with the true value and the traditional estimated class prior value. The WHU-HI-Longkou dataset and Indian Pines dataset can provide the true target distribution for comparison, due to the complete annotation, while the proportion of diseased pines in the PWD dataset is unknown, so the test value was estimated by the intersection point of the accuracy metric curves (as shown in Fig. 5).

The class prior estimation results are listed in Table III, where the proposed method is compared with the traditional estimation method of the kernel mixture proportion estimation (KMPE) algorithm [50], whose KM2 value is commonly used to determine the target class prior. As shown, the estimation errors of PU-KBS under different band sizes are all within the reasonable 0.05 range [41], while the KM2 value is not accurate and shows a big difference.

### B. Ablation Experiment for the Key Band Subset

To compare the KBS results of PU-KBS with those of typical band selection methods, different kinds of band selection methods were used following the same deep PU network (absPU), including: 1) ranking-based methods, i.e., maximum variance principal component analysis (MVPCA) [51] and spatial/spectral structural similarity (SSIM) [52]; 2) clustering-based methods, i.e., $k$-means clustering ($k$-means) [53] and fast neighborhood grouping for hyperspectral band selection (FNGBS) [39]; and 3) sparsity-based methods, i.e., ISSC [49] and sparse representation based band selection (SpaBS) [54]. Moreover, the $\pi_P$ parameters of the benchmark methods were set as the true values. The main parameter settings are listed in Table IV.

The detection results for the WHU-HI-Longkou dataset under different numbers of selected bands are shown in Fig. 12. The quantitative detection results of the PWD dataset
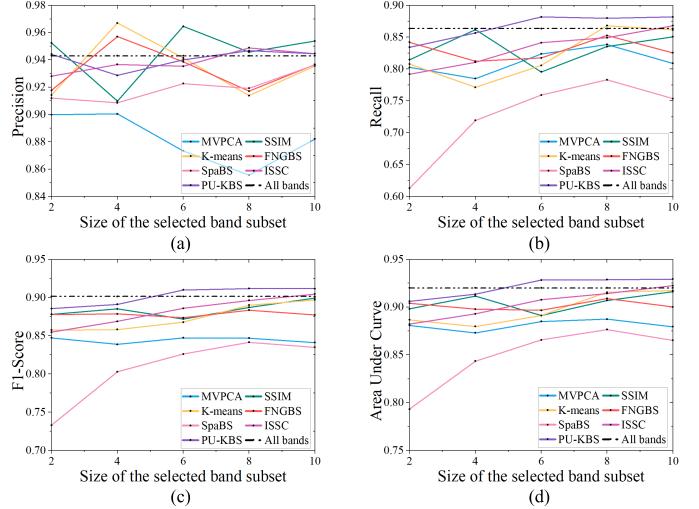


Fig. 12. Results of the comparison with different band selection methods for the WHU-HI-LongKou dataset. (a) Precision. (b) Recall. (c) $F$1-Score. (d) AUC.

and the Indian Pines dataset are shown in the Appendix as a Supplementary Material. The result values for the WHU-HI-Longkou dataset (using six selected bands), the PWD dataset (using ten selected bands), and the Indian Pines dataset (using ten selected bands) are listed in Table V (left), (medium), and (right), where the top-three results are highlighted with different colors. The proposed PU-KBS framework obtains high and stable values of $R$, $F$1-score, and AUC, and it exceeds the performance obtained using all bands when using six selected bands. For the $F$1-score values, PU-KBS outperforms the other methods in all sizes of band subsets, which demonstrates the stability and robustness of PU-KBS in KBS. Meanwhile, ISSC shows the second-best performance, but is slightly inferior to SSIM and FNGBS with 2–4 bands, while stably maintaining the second-best performance with 6–10 bands. For the AUC values, PU-KBS again achieves the best score, while ISSC shows a similar trend, maintaining a top-three performance in most cases. Therefore, the band selection results of ISSC were used for the detection model comparison.

### C. Ablation Experiment for Different EC Search Methods

To compare the search performance of different evolutionary algorithms, the search strategy of the PU-KBS framework was changed to other EC methods, besides the ABC algorithm, as an ablation experiment. The other EC search methods were genetic algorithm (GA) [55], differential evolution (DE)

TABLE V

COMPARISON WITH DIFFERENT BAND SELECTION METHODS, WHERE THE DETECTOR WAS ABSPU, WITH THE TOP-THREE RESULTS MARKED IN RED, GREEN, AND BLUE, RESPECTIVELY

| Method | WHU-HI-LongKou dataset (six selected bands) | | | | The PWD dataset (ten selected bands) | | | | Indian Pines dataset (ten selected bands) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P | R | F1 | AUC | P | R | F1 | AUC | P | R | F1 | AUC |
| All-bands | 0.9429 | 0.8635 | 0.9015 | 0.9200 | 0.8682 | 0.8542 | 0.8611 | 0.9070 | 0.9568 | 0.9268 | 0.9416 | 0.9624 |
| MVPCA | 0.8734 | 0.8236 | 0.8473 | 0.8849 | 0.8592 | 0.8470 | 0.8531 | 0.9020 | 0.9442 | 0.9276 | 0.9358 | 0.9625 |
| SSIM | 0.9646 | 0.7954 | 0.8719 | 0.8912 | 0.8564 | 0.8652 | 0.8608 | 0.9102 | 0.9606 | 0.9280 | 0.9440 | 0.9631 |
| *k*-means | 0.9414 | 0.8058 | 0.8678 | 0.8912 | 0.8729 | 0.8505 | 0.8615 | 0.9061 | 0.9499 | 0.9276 | 0.9386 | 0.9626 |
| FNGBS | 0.9386 | 0.8174 | 0.8735 | 0.8967 | 0.8675 | 0.8497 | 0.8585 | 0.9048 | 0.9374 | 0.9289 | 0.9330 | 0.9629 |
| SpaBS | 0.9227 | 0.7592 | 0.8261 | 0.8657 | 0.8612 | 0.8480 | 0.8545 | 0.9028 | 0.9354 | 0.9477 | 0.9414 | 0.9722 |
| ISSC | 0.9354 | 0.8414 | 0.8859 | 0.9076 | 0.8688 | 0.8544 | 0.8615 | 0.9072 | 0.9656 | 0.9280 | 0.9464 | 0.9632 |
| PU-KBS | 0.9401 | 0.8818 | 0.9100 | 0.9283 | 0.8627 | 0.8657 | 0.8642 | 0.9115 | 0.9817 | 0.9446 | 0.9626 | 0.9718 |
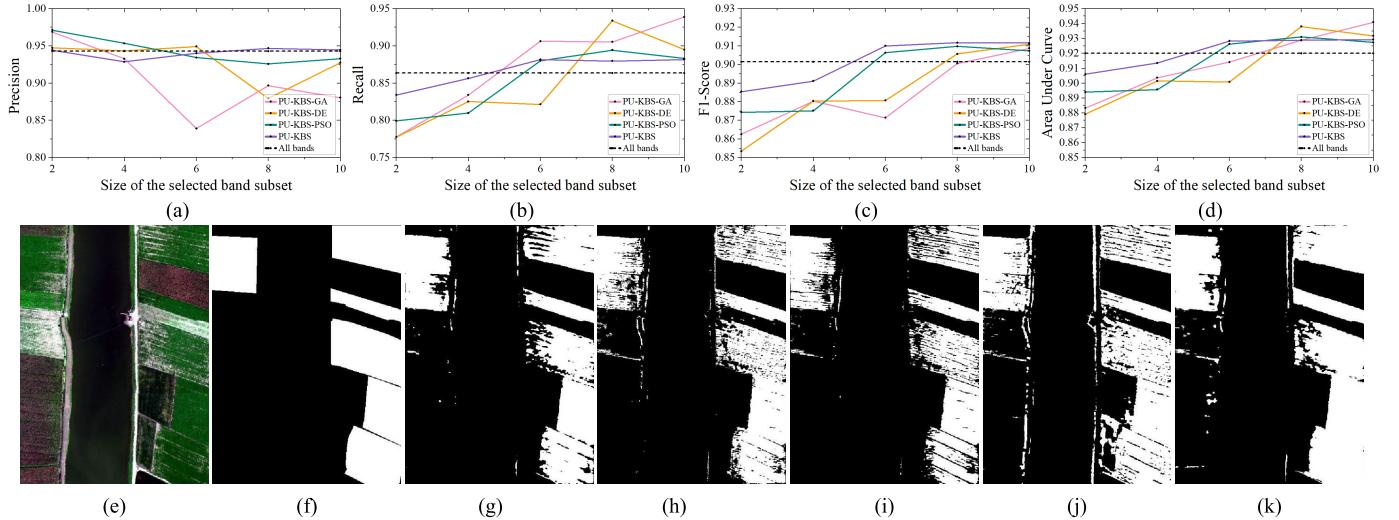


Fig. 13. Results of the ablation experiment for different EC Search methods with the WHU-HI-LongKou dataset. (a)–(d) Quantitative accuracy metrics. (f) Ground truth. (g)–(k) Visible detection results, where (h)–(k) were obtained using six selected bands. (a) Precision. (b) Recall. (c) *F*1-Score. (d) AUC. (e) Original image. (f) Ground truth. (g) All bands. (h) PU-KBS-GA. (i) PU-KBS-DE. (j) PU-KBS-PSO. (k) PU-KBS-ABC.

TABLE VI

ABLATION RESULTS FOR DIFFERENT EC SEARCH METHODS, WITH THE TOP-THREE RESULTS MARKED IN RED, GREEN, AND BLUE, RESPECTIVELY

| Benchmark | WHU-HI-LongKou dataset (six selected bands) | | | | |
|---|---|---|---|---|---|
| | P | R | F1 | AUC | Search times |
| PU-KBS-GA | 0.8390 | 0.9063 | 0.8714 | 0.9141 | 1200 |
| PU-KBS-DE | 0.9493 | 0.8214 | 0.8807 | 0.9008 | 2400 |
| PU-KBS-PSO | 0.9344 | 0.8802 | 0.9063 | 0.9262 | 1259 |
| PU-KBS–ABC | 0.9401 | 0.8818 | 0.9100 | 0.9283 | 1169 |
| All-bands | 0.9429 | 0.8635 | 0.9015 | 0.9200 | - |

[56], and particle swarm optimization (PSO) [57]. The main parameter settings are listed in Table IV.

The detection results for the WHU-HI-Longkou dataset under different numbers of selected bands are shown in Fig. 13(a)–(d). The distribution maps and the result values obtained using six selected bands are shown in Fig. 13(e)–(k) and Table VI, where the top-three results are highlighted with different colors. The quantitative detection results of the PWD dataset and the Indian Pines dataset are shown in the Appendix as a Supplementary Material. The performance of the ABC algorithm is superior in both search results and search efficiency, and it can obtain the optimal detection results with

the least number of search times under the same population and iteration settings. Furthermore, the performance of the ABC algorithm has been reported to be comparable with these EC search methods or even better in the current literature [42], [58], [59]. In addition, the ABC algorithm also has fewer control parameters and is efficient for the local and global optimization of multimodal and multivariable problems, such as high-dimensional hyperspectral data.

### D. Discussion on the Selected Key Bands for the PWD Detection Dataset

In order to investigate the diagnostic spectral characteristics of pines with PWD, we compared the ten selected band results of PU-KBS and the other band selection methods for the PWD dataset (as described in Section V-B). This comparison can also prove the application value of the KBS.

The ten selected band results of the benchmarking methods and PU-KBS are visualized in Fig. 14(a)–(g) by line charts, where they are compared with the original PWD spectral curve [see Fig. 14(h)]. The corresponding quantitative spectral angle (SA) values are shown in Fig. 14(i). As shown, the ten selected band results of PU-KBS have the most similar trend and accurate turning points, compared to the original PWD spectral curve, which can also be verified by the lowest
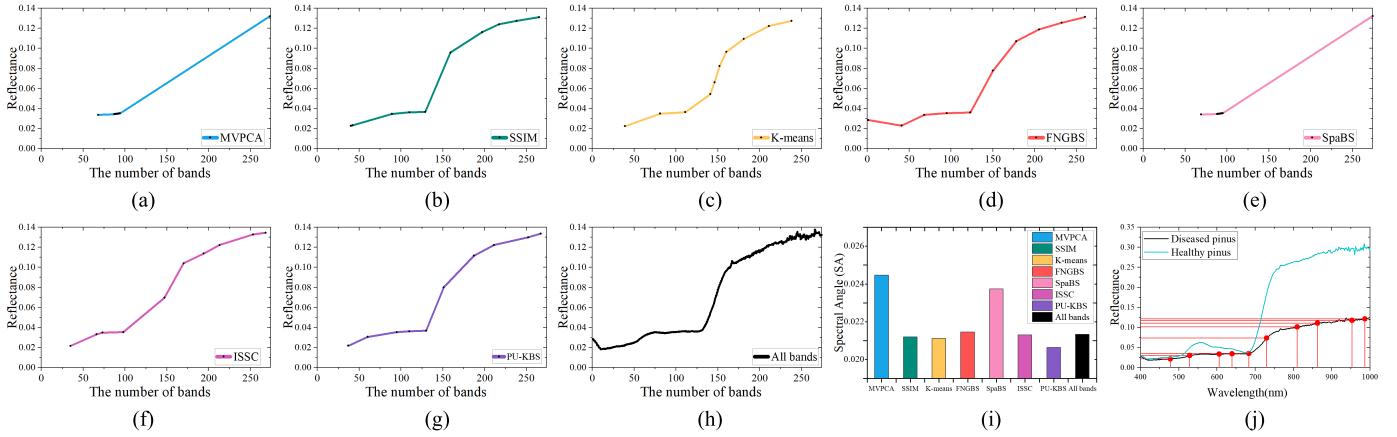
Fig. 14. Visualization and analysis of the ten selected band results for the PWD dataset. (a)–(g) Ten selected band results of the different band selection methods. (a) MVPCA. (b) SSIM. (c) k-means. (d) FNGBS. (e) SpaBS. (f) ISSC. (g) PU-KBS. (i) Average SA. (j) Spectral curves of healthy and diseased pines.

SA value. Although the band selection results of some of the other methods also look similar to the original spectral curve, their results have the characteristics of local accumulation (SSIM, ISSC) or an overall uniform distribution (k-means, FNGBS), and the turning point offset can affect the spectral curve description.

The spectral curves of the healthy and moderately diseased pines are shown in Fig. 14(j), where the located red points are the ten selected band subsets of PU-KBS. The diagnostic spectral characteristics for PWD detection can be obtained by the 10-band result of PU-KBS. The wavelength range for the study can be divided into the green (500–580 nm), red (620–760 nm), and near-infrared (760–900 nm) regions. In the green spectrum, the peak reflectance of diseased pines is lower and moving to the longer wavelength, so as to distinguish PWD by reflectance difference at 530 and 606 nm. As for the red spectrum, the peak reflectance trend is the opposite of the green spectrum, where the bands at 680 and 730 nm can respectively describe the turning point and turning degree of the PWD spectrum. For the infrared spectrum, the great difference between reflectance, i.e., 810 and 863 nm, allows easy separation of the diseased pines from the healthy pines. It can be found that the PU-KBS-selected bands are similar to the manual analysis results for PWD detection, which demonstrates the correct guidance and application value of the diagnostic spectral characteristics acquisition.

## VI. CONCLUSION

In this paper, in order to solve the two main problems when the PUL methods are applied to hyperspectral imageries within complex scenarios—class prior estimation and redundant information removal, a robust PU-KBS framework has been proposed. The candidate class prior and key band subset are encoded in the same vector for jointly initializing and updating by the evolutionary algorithm through the optimization function constructed by the detection result of the PU network detector iteratively until the optimal detection results are obtained. Five experiments based on a standard dataset and ap-plication data that compared with the classical OCC methods and HTD methods were designed and conducted,

which demonstrated that the end-to-end framework is superior to the two-stage solution, the proposed PU-KBS framework can acquire an accurately estimated class prior and key band subset for the specific target of interest, as well as the robust detector.

In the future, how to improve the proposed framework from wrapper-based to embedded-based by closely linking the detectors and evolutionary search should be considered, meanwhile, the current model complexity of the proposed PU-KBS framework still has room to fall. Furthermore, the band selection consistency across image strips could be considered.

## REFERENCES

[1] H. Bi, C. Zhang, K. Wang, J. Tong, and F. Zheng, "Rethinking camouflaged object detection: Models and datasets," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5708–5724, Sep. 2022.

[2] L. Lei, X. Wang, Y. Zhong, H. Zhao, X. Hu, and C. Luo, "DOCC: Deep one-class crop classification via positive and unlabeled learning for multi-modal satellite imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 105, Dec. 2021, Art. no. 102598.

[3] J. Li, X. Wang, H. Zhao, X. Hu, and Y. Zhong, "Detecting pine wilt disease at the pixel level from high spatial and spectral resolution UAV-borne imagery in complex forest landscapes using deep one-class classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102947.

[4] H. Zhao et al., "Mapping the distribution of invasive tree species using deep one-class classification in the tropical Montane landscape of Kenya," *ISPRS J. Photogramm. Remote Sens.*, vol. 187, pp. 328–344, May 2022.

[5] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, p. 111716, May 2020.

[6] B. Mack and B. Waske, "In-depth comparisons of MaxEnt, biased SVM and one-class SVM for one-class classification of remote sensing data," *Remote Sens. Lett.*, vol. 8, no. 3, pp. 290–299, Mar. 2017.

[7] R. Piiroinen, F. E. Fassnacht, J. Heiskanen, E. Maeda, B. Mack, and P. Pellikka, "Invasive tree species detection in the eastern arc mountains biodiversity hotspot using one class classification," *Remote Sens. Environ.*, vol. 218, pp. 119–131, Dec. 2018.

[8] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.

[9] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

[10] C. Sanchez-Hernandez, D. S. Boyd, and G. M. Foody, "One-class classification for mapping a specific land-cover class: SVDD classification of fenland," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 1061–1073, Apr. 2007.

[11] C. Elkan and K. Noto, "Learning classifiers from only positive and unlabeled data," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2008, pp. 213–220.

[12] B. Liu, Y. Dai, X. Li, W. S. Lee, and P. S. Yu, "Building text classifiers using positive and unlabeled examples," in *Proc. 3rd IEEE Int. Conf. Data Mining*, Jan. 2003, pp. 179–186.

[13] Y. Lu and L. Wang, "How to automate timely large-scale mangrove mapping with remote sensing," *Remote Sens. Environ.*, vol. 264, Oct. 2021, Art. no. 112584.

[14] W. Li, Q. Guo, and C. Elkan, "Can we model the probability of presence of species without absence data?" *Ecography*, vol. 34, no. 6, pp. 1096–1105, Dec. 2011.

[15] W. Li, Q. Guo, and C. Elkan, "One-class remote sensing classification from positive and unlabeled background data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 730–746, 2021.

[16] M. C. du Plessis, G. Niu, and M. Sugiyama, "Convex formulation for learning from positive and unlabeled data," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1386–1394.

[17] R. Kiryo, G. Niu, M. C. Du Plessis, and M. Sugiyama, "Positive-unlabeled learning with non-negative risk estimator," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1674–1684.

[18] G. M. Foody, A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd, "Training set size requirements for the classification of a specific class," *Remote Sens. Environ.*, vol. 104, no. 1, pp. 1–14, Sep. 2006.

[19] H. Zhao, X. Wang, J. Li, and Y. Zhong, "Class prior-free positive-unlabeled learning with Taylor variational loss for hyperspectral remote sensing imagery," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 16827–16836.

[20] J. Bekker and J. Davis, "Learning from positive and unlabeled data: A survey," *Mach. Learn.*, vol. 109, pp. 719–760, Jan. 2020.

[21] A. Ritter, E. Wright, W. Casey, and T. Mitchell, "Weakly supervised extraction of computer security events from Twitter," in *Proc. 24th Int. Conf. World Wide Web*, May 2015, pp. 896–905.

[22] L. Ruff, R. A. Vandermeulen, N. Goernitz, L. Deecke, and M. Kloft, "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.

[23] M. C. Du Plessis, G. Niu, and M. Sugiyama, "Analysis of learning from positive and unlabeled data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 703–711.

[24] X. Deng, W. Li, X. Liu, Q. Guo, and S. Newsam, "One-class remote sensing classification: One-class vs. binary classifiers," *Int. J. Remote Sens.*, vol. 39, no. 6, pp. 1890–1910, Mar. 2018.

[25] G. Ward, T. Hastie, S. Barry, J. Elith, and J. R. Leathwick, "Presence-only data and the EM algorithm," *Biometrics*, vol. 65, no. 2, pp. 554–563, Jun. 2009.

[26] C.-I. Chang, *Real-Time Progressive Hyperspectral Image Processing*. Cham, Switzerland: Springer, 2016.

[27] Z. Chen et al., "Global to local: A hierarchical detection algorithm for hyperspectral image target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5544915.

[28] C.-I. Chang, "Hyperspectral target detection: Hypothesis testing, signal-to-noise ratio, and spectral angle theories," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5505223.

[29] H. J. Shin, D.-H. Eom, and S.-S. Kim, "One-class support vector machines—An application in machine fault detection and classification," *Comput. Ind. Eng.*, vol. 48, no. 2, pp. 395–408, Mar. 2005.

[30] P. Yang, X.-L. Li, J.-P. Mei, C.-K. Kwoh, and S.-K. Ng, "Positive-unlabeled learning for disease gene identification," *Bioinformatics*, vol. 28, no. 20, pp. 2640–2647, Oct. 2012.

[31] C.-J. Hsieh, N. Natarajan, and I. Dhillon, "PU learning for matrix completion," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2445–2453.

[32] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H$^2$) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sens. Environ.*, vol. 250, Dec. 2020, Art. no. 112012.

[33] Y. Tang et al., "An object fine-grained change detection method based on frequency decoupling interaction for high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5600213.

[34] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503314.

[35] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.

[36] 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.

[37] Z. Chen, D. Hong, and H. Gao, "Grid network: Feature extraction in anisotropic perspective for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[38] C. Zhao et al., "Hyperspectral image classification with multi-attention transformer and adaptive superpixel segmentation-based active learning," *IEEE Trans. Image Process.*, vol. 32, pp. 3606–3621, 2023.

[39] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5028–5039, Jun. 2021.

[40] X. Ma et al., "A multiobjective evolutionary algorithm based on decision variable analyses for multiobjective optimization problems with large-scale variables," *IEEE Trans. Evol. Comput.*, vol. 20, no. 2, pp. 275–298, Apr. 2016.

[41] H. Zhao, Y. Zhong, X. Wang, and H. Shu, "One-class risk estimation for one-class hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* vol. 61, 2023, Art. no. 5518017.

[42] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Appl. Math. Comput.*, vol. 214, no. 1, pp. 108–132, 2009.

[43] K. Fukuda, "Physiological process of the symptom development and resistance mechanism in pine wilt disease," *J. Forest Res.*, vol. 2, no. 3, pp. 171–181, Aug. 1997.

[44] X. Jin, S. Paswaters, and H. Cline, "A comparative study of target detection algorithms for hyperspectral imagery," *Proc. SPIE*, vol. 7334, pp. 682–693, Feb. 2009.

[45] X. Geng, L. Ji, K. Sun, and Y. Zhao, "CEM: More bands, better performance," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1876–1880, Nov. 2014.

[46] R. Zhao, Z. Shi, Z. Zou, and Z. Zhang, "Ensemble-based cascaded constrained energy minimization for hyperspectral target detection," *Remote Sens.*, vol. 11, no. 11, p. 1310, Jun. 2019.

[47] D. Shen, X. Ma, H. Wang, and J. Liu, "A dual sparsity constrained approach for hyperspectral target detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2022, pp. 1963–1966.

[48] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[49] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, Jun. 2015.

[50] H. Ramaswamy, C. Scott, and A. Tewari, "Mixture proportion estimation via kernel embeddings of distributions," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2052–2060.

[51] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.

[52] B. Xu, X. Li, W. Hou, Y. Wang, and Y. Wei, "A similarity-based ranking method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9585–9599, Nov. 2021.

[53] M. Ahmad, I. U. Haq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using k-means clustering," *Int. J. Eng. Technol.*, vol. 3, no. 6, pp. 606–614, 2011.

[54] S. Li and H. Qi, "Sparse representation based band selection for hyperspectral images," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2693–2696.

[55] S. Mirjalili, "Genetic algorithm," in *Evolutionary Algorithms and Neural Networks*. Cham, Switzerland: Springer, 2019, pp. 43–55.

[56] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[57] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, 1995, pp. 1942–1948.

[58] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Appl. Soft Comput.*, vol. 8, no. 1, pp. 687–697, Jan. 2008.

[59] F. S. Abu-Mouti and M. E. El-Hawary, "Overview of artificial bee colony (ABC) algorithm and its applications," in *Proc. IEEE Int. Syst. Conf. SysCon*, Mar. 2012, pp. 1–6.

**Ziying Liu** received the B.S. degree from Wuhan University, Wuhan, China, in 2022, where she is currently pursuing the M.S. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS).

Her major research interests include hyperspectral image preprocessing.

**Shaoyu Wang** received the Ph.D. degree from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, in 2022.

He is now conducting postdoc research works at the College of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea. His major research interests include hyperspectral information processing and remote sensing of GHG fluxes.

**Jingtao Li** received the B.S. degree from the School of Geography and Information Engineering, China University of Geosciences, Wuhan, China, in 2021. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan.

His major research interests include anomaly detection in remote sensing.

**Hengwei Zhao** (Student Member, IEEE) received the B.S. degree in surveying and mapping engineering from the School of Resources and Civil Engineering, Northeastern University, Shenyang, China, in 2019. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

His major research interests include weakly supervised remote sensing image processing.

**Yanfei Zhong** (Senior Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

Since 2010, he has been a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. He organized the Intelligent Data Extraction, Analysis and Applications of Remote Sensing (RSIDEA) research group. He has published more than 100 research papers in international journals, such as *Remote Sensing of Environment*, *ISPRS Journal of Photogrammetry and Remote Sensing*, and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. His research interests include hyperspectral remote sensing information processing, high-resolution remote sensing image understanding, and geoscience interpretation for multisource remote sensing data and applications.

Dr. Zhong is a fellow of the Institution of Engineering and Technology (IET). He was a recipient of the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). He won the Second-Place Prize in 2013 IEEE GRSS Data Fusion Contest and the Single-view Semantic 3-D Challenge of the 2019 IEEE GRSS Data Fusion Contest, respectively. He is currently serving as an Associate Editor for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and the *International Journal of Remote Sensing*.

**Xinyu Wang** (Member, IEEE) received the B.S. degree in photogrammetry and remote sensing and the Ph.D. degree in communication and information systems from Wuhan University, Wuhan, China, in 2014 and 2019, respectively.

Since 2019, he has been an Associate Research Fellow at the School of Remote Sensing and Information Engineering, Wuhan University. His major research interests include hyperspectral data processing and applications.