

# A Hybrid Gray Wolf Optimizer for Hyperspectral Image Band Selection

Yulei Wang<sup>✉</sup>, Member, IEEE, Qingyu Zhu, Haipeng Ma, and Haoyang Yu<sup>✉</sup>, Member, IEEE

**Abstract**—High spectral dimensionality of hyperspectral image (HSI) has brought great redundancy for data processing. Band selection (BS), as one of the most commonly used dimension reduction (DR) techniques, attempts to remove the redundant spectral bands, while maintaining good classification or detection rate for later applications. Gray wolf optimizer (GWO) algorithm is a meta-heuristic algorithm, and it is used for HSI BS. However, the convergence factor of the basic GWO is linearly decreased, leading to a slower convergence speed and increasing the probability of falling into local optimality. This article proposes a new hybrid gray wolf optimizer (HGWO) algorithm for HSI BS, which uses adaptive decreasing convergence factor instead of linear convergence factor to improve GWO convergence rate and combines category separability for initialization to avoid local optimality. Five nonlinear functions are used to test the convergence of the proposed HGWO algorithm, compared with the state-of-the-art optimization algorithms. Finally, the experimentations are performed on three widely used real hyperspectral datasets for HSI classification, and the experimental results show that band subsets selected by the proposed HGWO algorithm can obtain better classification accuracy compared with other global optimization algorithms.

**Index Terms**—Band selection (BS), classification, global optimization, hybrid gray wolf optimizer (HGWO), hyperspectral image (HSI).

## I. INTRODUCTION

HYPERSPECTRAL images have rich spectral information and are widely used in many fields, such as classification and target recognition. The high dimensionality of hyperspectral images (HSIs) could enhance the ability to distinguish features; however, the high correlation between bands would also bring great difficulty for later processing, such as increasing complexity of subsequent processing algorithms and leading to “Hughes” phenomenon. At present, dimension

Manuscript received September 2, 2021; revised February 27, 2022; accepted April 6, 2022. Date of publication April 18, 2022; date of current version May 4, 2022. The work of Yulei Wang was supported in part by the National Natural Science Foundation of China under Grant 61801075; in part by the China Postdoctoral Science Foundation under Grant 2020M670723; in part by the Open Research Funds of State Key Laboratory of Integrated Services Networks, Xidian University, under Grant ISN20-15; and in part by the Fundamental Research Funds for the Central Universities under Grant 3132022232. The work of Haipeng Yu was supported by the National Natural Science Foundation of China under Grant 42101350. (*Corresponding author: Haoyang Yu*.)

The authors are with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: wangyulei@dlmu.edu.cn; 1067336725@qq.com; mhp0520@163.com; yuhy@dlmu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2022.3167888

reduction (DR) of hyperspectral is mainly partitioned into two branches: extraction and selection of data. The former is feature extraction based on transformation, and the original high-dimensional HSI data is projected into a low-dimensional space through various transformations, discarding some unimportant properties. However, this type of method transforms the image, so that the transformed data no longer have original physical attributes, which is not conducive to understand the original data. The latter is based on non-transformed feature selection, also called band selection (BS), which selects representative band subset from original hyperspectral bands. Compared with feature extraction, BS can retain the physical information of the original data and is a simple and effective DR method. However, there are some scientific problems for hyperspectral BS. For example, how to determine the number of bands in selected band subsets; how to select high-quality band subsets without label information; how to reduce computational complexity when searching the optimal band subsets; how to select the most favorable band subset for classification or detection. This article did research on the last question.

Since BS algorithms could inherit the physical interpretability of the original data, it is the preferred method for DR. Some BS methods are applied to hyperspectral classification [1]–[7], and some are applied to target detection [8]–[11]. According to whether there are labeled samples or not, the existing BS methods can be further divided into three categories: supervised [12], [13], semi-supervised [14], [15], and unsupervised [16]–[20]. Supervised refers to training and learning based on labeled samples, and then performing BS. Semi-supervised seeks the representative band subset using both the limited labeled and some unlabeled samples. As a matter of fact, in practical applications, it is difficult to obtain labeled samples. Hence, unsupervised BS without prior information has drawn more attention.

In the past few years, scholars have researched many methods of BS, which can be generally divided into two aspects. One is based on the traditional algorithms, and the other is based on deep learning. The Traditional traditional algorithms can be divided into information-based, cluster-based, and search-based algorithms. Information-based band selection is a basic method. This method relies on the priority arrangement of bands by specified criteria, such as information entropy, variance, signal-to-noise ratio, and information dispersion, etc., [4]. The bands are sorted according to these criteria, and then selected according to the sequencing.

As a result, this method is usually unsupervised. Cluster-based BS methods are also often applied. Bands are clustered according to a certain rule, and then the most representative band is selected according to a certain criterion in each cluster [6], [21]. The last is the global search algorithm, which aims to find the optimal solution of the objective function; this article introduces this method in detail. In addition to the traditional methods, deep learning has also been applied to the selection of hyperspectral bands in recent years, and have achieved good results [22]–[26].

Recently, global optimization algorithms are widely used for hyperspectral BS, such as genetic algorithm (GA) [27], firefly algorithm (FA) [28], gravity search algorithm (GSA) [29], and particle swarm optimization (PSO) [30], [31]. GA has many parameters, the execution is complicated, and it is easy to fall into local optimum, resulting in bad global search performance. As for the FA algorithm, the discovery rate is low, the solution accuracy is not high, and the convergence speed is slow. GSA has slow convergence speed and unsatisfactory global search effect. As a new heuristic algorithm, the GWO algorithm was proposed in 2014. Compared with the existing optimization algorithms, it has the characteristics of fewer adjustment parameters, fast convergence, and high execution, but it still has some problems; the convergence speed is slow when solving multimodal functions, and it is easy to fall into the shortcomings of local extremes. The variations in the GWO algorithm are proposed recently to overcome the above deficiencies. Gai *et al.* [32] proposed to combine GWO and GA. Xu and Su [33] combined GWO with the cuckoo search algorithm. Medjahed *et al.* [34] applied GWO to HSI BS in 2016, opening a new meta-heuristic algorithm for hyperspectral BS. In [35], based on the information gain and the spectral curve of the hyperspectral dataset, the band subset decomposition technique is improved combined with GWO. In [17], the optimal feature selection is based on the optimization process of GWO algorithm and maximum entropy principle. In [36], GWO is applied to target detection in HSIs. Due to the redundancy of HSI, many dimensionality reduction methods have been developed [37]–[42].

The traditional GWO BS algorithm has a random initialization for band subset. To get better band subset for classification, this article proposed a new HGWO algorithm for hyperspectral BS, which uses relative category separability to obtain the initial population, and a new nonlinear convergence factor is then used instead of linear convergence factor to achieve an adaptively search speed. Optimal band subsets obtained by HGWO and other global optimization algorithms are, respectively, used for classification by support vector machine (SVM) on three real HSI datasets. The experimental results show that band subset obtained by the proposed HGWO could get better classification effect.

The rest of this article is organized as follows: Section II presents and discusses the proposed HGWO approach. In Section III, the convergence performance of HGWO and other four global optimization algorithms are tested through five nonlinear functions to prove the effectiveness of the proposed algorithm. Then, Section IV describes the experimental

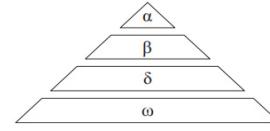


Fig. 1. Social hierarchy of gray wolves.

results and the performance of the proposed method. Finally, the conclusion is given in Section V.

## II. PROPOSED ALGORITHM

GWO is a typical global optimization algorithm with few parameters and is easy to implement. This article proposes a new hybrid gray wolf algorithm to obtain an optimized band subset for hyperspectral classification.

### A. Basics of GWO

Gray wolves strictly abide by social hierarchical relationship. As shown in Fig. 1, the leader wolf is called  $\alpha$ , who is responsible for making decisions on activities such as predation and habitation. The second wolf is  $\beta$  who is responsible for assisting  $\alpha$  in making decisions, and  $\beta$  becomes a candidate for  $\alpha$  when  $\alpha$  ages or passes away.  $\delta$  is the bottom end of the wolf pack and usually obeys other wolves. The social hierarchy of gray wolves is shown in Fig. 1.

In mathematical model, each gray wolf individual represents a candidate in population. The optimal solution in population is recorded as  $\alpha$ , the second optimal solution is recorded as  $\beta$ , the third optimal solution is recorded as  $\delta$ , and the other candidate solutions are collectively called  $\omega$ . Gray in each iteration of the wolf algorithm,  $\alpha$ ,  $\beta$ , and  $\delta$  are used as the first three optimal solutions to determine the position of the prey, command  $\omega$  to update, and constantly approach the prey.

Assuming that the gray wolf population is  $nPop$ ,  $NB$  is the dimension of search space, and the position of the  $i$ -th gray wolf is represented as  $X_i = (x_i^1, x_i^2, \dots, x_i^{NB})$ , the gray wolf will gradually approach and surround it when searching for prey, and the mathematical model of this behavior is as follows:

$$\mathbf{D} = |\mathbf{C}X_p(t) - X(t)| \quad (1)$$

$$X(t+1) = X_p(t) - \mathbf{AD} \quad (2)$$

where  $\mathbf{D}$  is the distance between gray wolf individual and prey,  $t$  is the current iteration, and  $X_p(t)$  and  $X(t)$  represent the position vectors of the prey and gray wolf, respectively.  $\mathbf{A}$  and  $\mathbf{C}$  are the coefficient vectors, and the calculation formulas are as follows:

$$\mathbf{A} = 2a(\mathbf{r}_1 - 1) \quad (3)$$

$$\mathbf{C} = 2\mathbf{r}_2 \quad (4)$$

where  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are the random vectors in  $[0,1]$ , and  $a$  is the convergence factor linearly decreasing from 2 to 0. The calculation of  $a$  is as follows:

$$a = 2 - 2t/t_{max} \quad (5)$$

where  $t_{max}$  is the maximum number of iterations.

In surrounding behavior, changing the values of  $\mathbf{A}$  and  $\mathbf{C}$  can realize gray wolf's search for prey. The setting of  $\mathbf{r}_1$  and  $\mathbf{r}_2$  is random, so the gray wolf can search for many directions around prey.  $|\mathbf{A}|$  represents the value of  $\mathbf{A}$ . We force the gray wolf to launch the behavior of attacking the prey by making  $|\mathbf{A}| < 1$ , and the gray wolves to search for a new prey by making  $|\mathbf{A}| > 1$ . Here, gray wolves diverge from the prey to find a potential optimal solution.

Assuming that  $\alpha$ ,  $\beta$ , and  $\delta$  have a strong ability to identify potential prey positions (optimal solutions). In each iteration, keep the best three gray wolves in current population, and then update the positions of other search gray wolves ( $\omega$ ) based on their location information. The mathematical model of this behavior can be expressed as follows:

$$\mathbf{D}_\alpha = |\mathbf{C}_1 \mathbf{X}_\alpha(t) - \mathbf{X}(t)| \quad (6)$$

$$\mathbf{D}_\beta = |\mathbf{C}_2 \mathbf{X}_\beta(t) - \mathbf{X}(t)| \quad (7)$$

$$\mathbf{D}_\delta = |\mathbf{C}_3 \mathbf{X}_\delta(t) - \mathbf{X}(t)| \quad (8)$$

$$\mathbf{X}_1(t+1) = \mathbf{X}_\alpha(t) - \mathbf{A}_1 \mathbf{D}_\alpha \quad (9)$$

$$\mathbf{X}_2(t+1) = \mathbf{X}_\beta(t) - \mathbf{A}_2 \mathbf{D}_\beta \quad (10)$$

$$\mathbf{X}_3(t+1) = \mathbf{X}_\delta(t) - \mathbf{A}_3 \mathbf{D}_\delta \quad (11)$$

$$\mathbf{X}(t+1) = [\mathbf{X}_1(t+1) + \mathbf{X}_2(t+1) + \mathbf{X}_3(t+1)]/3 \quad (12)$$

where  $\mathbf{X}_\alpha(t)$ ,  $\mathbf{X}_\beta(t)$ , and  $\mathbf{X}_\delta(t)$  represent the position vectors of  $\alpha$ ,  $\beta$ , and  $\delta$ , respectively.  $\mathbf{D}_\alpha$ ,  $\mathbf{D}_\beta$ , and  $\mathbf{D}_\delta$  represent the distance between current candidate gray wolf and the best three wolves.

Through the iteration of the above formulas, the gray wolf position is continuously updated, and prey is finally captured. Accordingly, the position and objective of the  $\alpha$  wolf are the optimal solutions.

### B. Objective Function

In this article, an objective function is designed, which can effectively evaluate candidate subsets and return their optimal benign metric, which combines class separability distance and classification accuracy to better realize the application of band subsets to HSI classification.

Given an HSI dataset  $\text{HIM} = \{x_1, x_2, \dots, x_n\} \in R^{l \times n}$ , where  $n$  represents the total number of pixels, and  $l$  represents the number of bands, with  $k$  classes  $\{\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_k\}$ , the class separability criterion of band can be mathematically defined as follows:

$$\mathbf{S}_b = \sum_{i=1}^k P_i (m_i - m)(m_i - m)^T \quad (13)$$

$$\mathbf{S}_w = \sum_{i=1}^k P_i \left( \frac{1}{N_i} \sum_{x_i \in C_i} (x_i - m_i)(x_i - m_i)^T \right) \quad (14)$$

where  $\mathbf{S}_b$  represents the inter-class dispersion matrix,  $\mathbf{S}_w$  represents the intra-class dispersion matrix,  $P_i$  denotes the prior probabilities of category  $i$ ,  $m_i$  indicates the mean vector of the  $i$ th sample set,  $m$  indicates the total mean vector of all kinds of sample sets, and  $N_i$  indicates the number of pixels in category  $C_i$ .

Classification always expects  $\mathbf{S}_w$  to be as small as possible and  $\mathbf{S}_b$  to be as large as possible. The relative distance between

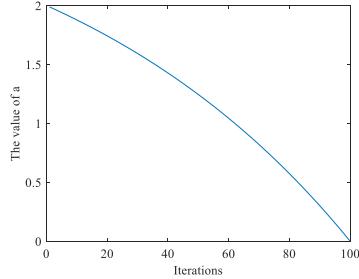


Fig. 2. Curve of the convergence factor.

categories determined according to this criterion can be used as a measure of category separability. The relative distance defined in this article is shown as follows:

$$J = \text{tr}(\mathbf{S}_w^{-1} \mathbf{S}_b) \quad (15)$$

where  $\text{tr}$  represents the trace of matrix.

### C. Proposed HGWO Algorithm

This section will introduce the proposed HGWO in detail. The HGWO algorithm has two innovations: 1) it improves the convergence factor of GWO and 2) it uses category separability to measure the performance of initial population.

1) *Improve the Convergence Factor*: The convergence factor determines the search ability of GWO. The convergence factor of the basic GWO algorithm linearly converges from 2 to 0. Linear convergence factor cannot perform a global search well and is easy to fall into a local search. Generally, in the global optimization algorithm, the search space is large in the early stage and it needs more time to find the optimal value; after several iterations, it converges faster to get the optimal value. Since the exponential function has the similar characteristic as the above description, an adaptive nonlinear change convergence factor (similar as the exponential function) shown in (16) is proposed, which could get an adaptive nonlinear convergence factor with the same range of [0,2] as the original linear convergence factor

$$a = 2 - 2 \cdot \left( e^{\frac{t}{t_{\max}}} - 1 \right) / (e - 1). \quad (16)$$

For a visual understanding of value  $a$  in different iterations, (16) is plotted as Fig. 2. The value of  $a$  decreases nonlinearly with the increase in the number of iterations and decreases slowly at the initial stage of the iteration with a little shrinkage range, which realizes fine search and changes rapidly at the later stage of the iteration to realize fast search.

2) *Algorithm Implementation*: Hyperspectral data usually have strong band correlation and high data redundancy. This article proposes an improved gray wolf algorithm for BS. Determine the initial population according to category separability, and select initial three wolves according to the objective function, and then update iteratively. Fig. 3 shows the flowchart of the proposed HGWO method.

### III. CONVERGENCE PERFORMANCE TEST

Five nonlinear test functions in [43] are used to test the convergence performance of the proposed HGWO algorithm,

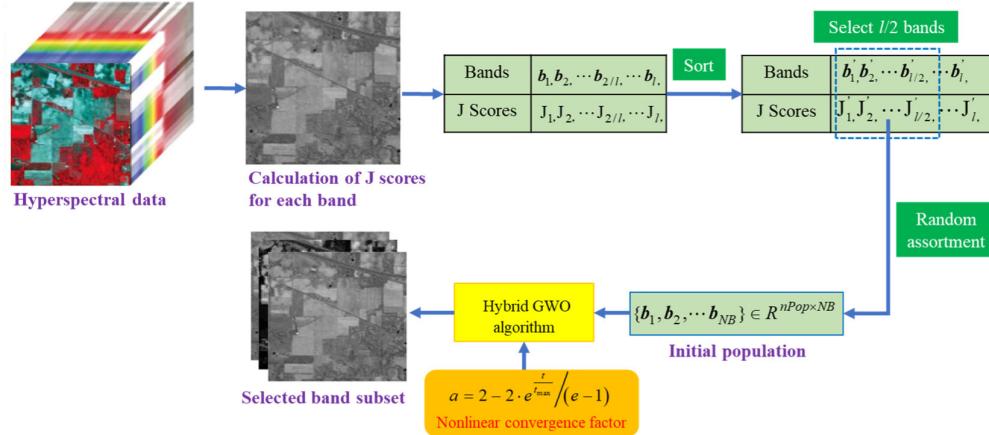


Fig. 3. Flowchart of the proposed method. Where nPop is the population size and NB is the number of selected bands.

TABLE I  
TEST FUNCTION TABLES

Function name	Function	Range	$F_{\min}$
Sphere	$f_1(x) = \sum_{i=1}^n x_i^2$	[-100,100]	0
Griewank	$f_2(x) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	[-600,600]	0
Rosenbrock	$f_3(x) = \sum_{i=1}^{n-1} \left[ 100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2 \right]$	[-20,20]	0
Rastrigin	$f_4(x) = \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i) + 10]$	[-5.12,5.12]	0
Ackley	$f_5(x) = -20 \exp(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) - \exp\left[\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right] + 20 + e$	[-32,32]	0

TABLE II  
COMPARISON RESULTS OF FIVE TEST FUNCTIONS

Function	PSO		GA		GWO		HGWO	
	Ave	Std	Ave	Std	Ave	Std	Ave	Std
f1	1.2666e+03	3.3703e+02	0.00017	0.00011	4.8278e-36	8.4730e-37	<b>2.8319e-40</b>	<b>2.3253e-40</b>
f2	7.4000e-03	4.3000e-03	0.00727	1.4879e-5	1.2426e-6	8.2797e-7	<b>0</b>	<b>0</b>
f3	1.0421e-6	1.8699e-5	0.00006	2.0302	1.3341e-5	1.1054e-4	<b>2.8610e-6</b>	<b>1.4211e-5</b>
f4	4.8863e-10	9.2442e-10	1.9899	0.9479	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
f5	21.4906	1.5560	20.9315	0.5955	19.0602	0.9628	<b>15.7152</b>	<b>0.8736</b>

compared with PSO, GA, and basic GWO. The detailed information of the test functions is shown in Table I.

The parameters are set as follows: the population size is set to 50, the maximum number of iterations is set to 500, the independent operation is set to 30 times, and the average values and standard deviations of the optimal solution are recorded. All the algorithms have the same initial population to ensure the fairness of the experimental results. The comparison results are listed in Table II. It can be shown that the proposed HGWO has a better performance in convergence and search

success rate on all five functions, compared with other three algorithms.

To better analyze the convergence performance of different algorithms, Fig. 4 gives a visual comparison, including the graphics of five functions and the convergence curves by different algorithms. From the right sub-figures, it could be easily seen that both GA and PSO are prone to fall into local optimum, and the proposed HGWO has a faster convergence rate and a higher success rate in contrast to the other three algorithms. Besides, the proposed HGWO could also reach

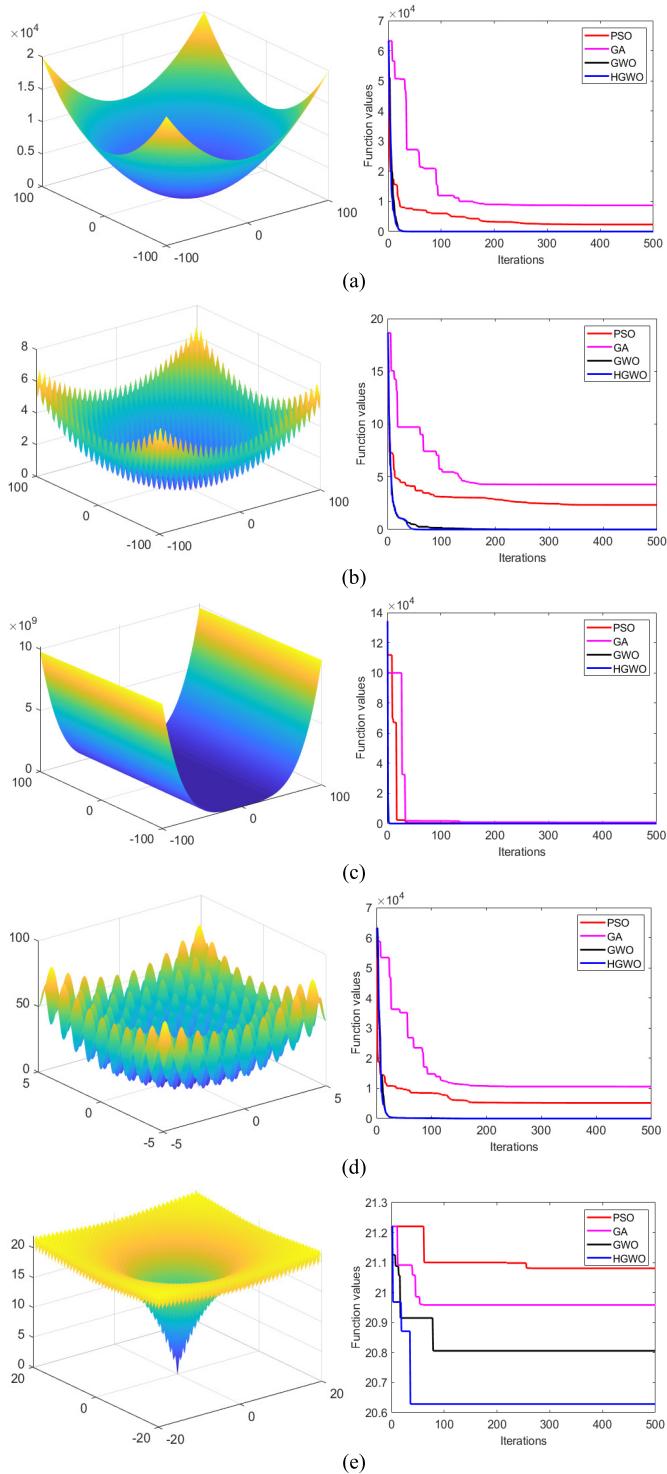


Fig. 4. Comparison of function diagrams and convergence curves. (a) f1. (b) f2. (c) f3. (d) f4. (e) f5.

the optimal value with fewer iterations, compared with other algorithms.

As shown in each sub-figure of Fig. 4, the left side is the three-dimensional distribution diagram of test function, and the right side is the convergence curve of different optimization algorithms for the global search of test function. HGWO has a better convergence effect than PSO for every test function. GA has the worst convergence effect in f1, f2, and f4.

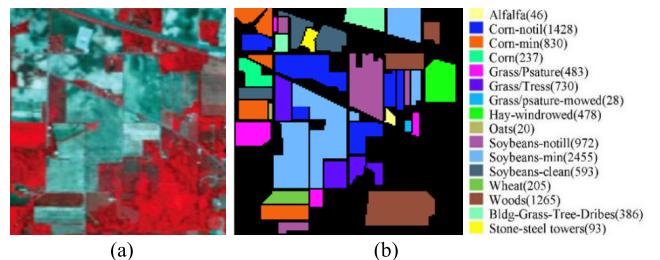


Fig. 5. Indian Pines with 16 classes, with (a) pseudo-color of Indian Pines and (b) color ground-truth image with class labels.

The convergence effect of HGWO in f3 and f4 is similar to GWO. f3 has the fastest convergence speed. The results show that compared with other optimization algorithms, HGWO has good convergence and a higher success rate when mining the global optimal solution.

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

For hyperspectral remote sensing images, BS is an important technology of DR, and the evaluation criteria of DR are mainly for the subsequent application-oriented perspective. Ground feature classification is one of the most important applications. Three HSI datasets obtained by different sensors are selected for verification.

##### A. Datasets' Description

1) *Indian Pines*: Indian Pines is a set of public data provided by the Remote Sensing Application Laboratory, Purdue University, West Lafayette, IN, USA. The data were acquired by the AVIRIS sensor in the agricultural area of Northwest Indiana in 1992. It has 224 spectral bands with a size of  $145 \times 145$  pixels from 0.4 to  $2.5 \mu\text{m}$ , and the spectral resolution is 10 nm with 20 m spatial resolution per pixel, including 20 water absorption bands, i.e., 104–108, 150–163, and 220. After removing the noise and water absorption band, there are 200 bands remaining. Fig. 5 shows the pseudo-color image of the Indian Pines dataset and 16 categories' distribution labels.

2) *Pavia University*: Pavia University was acquired by ROSIS sensor at Pavia University in Northern Italy in 2001. Its spectral range is 0.43–0.86  $\mu\text{m}$ , the image size is  $610 \times 340$  with 1.3 m spatial resolution per pixel, and there are 103 bands remaining after removing 12 noise bands. Fig. 6 shows the pseudo-color image and nine categories' distribution labels of the Pavia University dataset.

3) *Salinas*: Salinas is a public hyperspectral dataset acquired by the AVIRIS sensor over Salinas Valley, CA, USA, in 1998. The spectral resolution is 10 nm with a spatial resolution of 3.7 m per pixel, and the image size is  $512 \times 227$ , including 20 water absorption bands, i.e., 108–112, 154–167, and 224. After removing water absorption bands, there are 204 bands remaining. Fig. 7 shows the pseudo-color image of Salinas and 16 categories' distribution labels. The detailed description of three HSI datasets is shown in Table III.

##### B. Experimental Results' Analysis

In this section, the experimental results are verified by hyperspectral classification using SVM algorithm. Gaussian

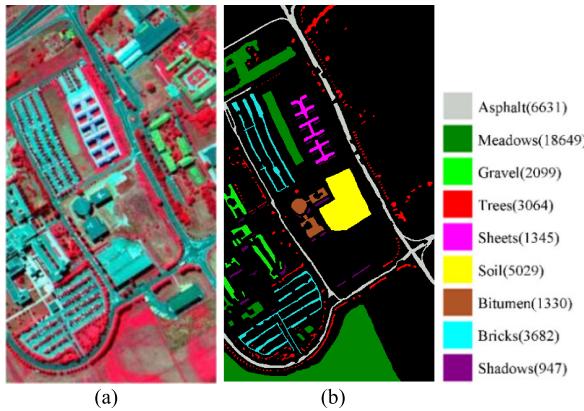


Fig. 6. Pavia University with nine classes, with (a) pseudo-color of Pavia University and (b) Color ground-truth image with class labels.

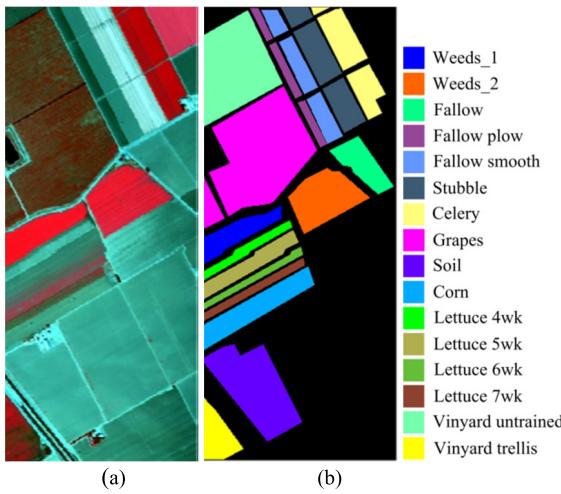


Fig. 7. Salinas with 16 classes, with (a) pseudo-color of Salinas and (b) color ground-truth image with class labels.

radial basis kernel function is selected, 10% of the data points in each category are randomly selected as the training set, and the rest are used as the testing set [6], [21]. The same training set was used in one experiment to classify the band subsets selected by different algorithms to ensure fairness, and the training set was randomly selected in each experiment. The classification results and deviations of 20 experiments were recorded in the table.

The  $J$  values of each band of the three datasets are plotted in Fig. 8(a)–(c), and it can be seen that the  $J$  values of different bands for the same dataset are different, and the distribution of  $J$  values in different datasets is obviously different, and then finding band subsets with large  $J$  values is more conducive to classification.

To see the effectiveness of the proposed algorithm, different BS algorithms are used for comparison in this section using three hyperspectral datasets. Two state-of-the-art algorithms PSO [44] and GA [27] are used for comparison. Besides, since the proposed HGWO is a kind of GWO-based algorithm, four similar algorithms are also compared in this article, including traditional GWO, IGWO-1, IGWO-2 [34], and NGWO (nonlinear GWO). The IGWO-1 algorithm is an improved

GWO algorithm where the  $J$  value is initialized, but the convergence factor is not improved. The IGWO-2 algorithm is another improved GWO algorithm where OA is used as the objective function. The NGWO algorithm, just as its name implies, is the nonlinear version of the GWO algorithm.

### 1) Band Selection and Classification for Indian Pines:

In this section, the performance comparison is given using the Indian Pines dataset. Fig. 9 shows the selection and classification results for the Indian Pines dataset. Fig. 9(a) and (b) shows OA and AA performances of different BS algorithms when the number of band changes, and it could be seen that the proposed HGWO classification performance is better than other methods in most cases by improving convergence factor and combining the initial population with class separability criterion, the IGWO-2 algorithm sometimes goes better than the proposed HGWO especially for AA performance; this is also proved in the latter comparison in Table IV that the overall AA of IGWO-2 is a little better than HGWO. Because of the uncertainty of cross-variation, GA has the worst classification performance, which corresponds to the result of the test functions. When the number of selected bands is small, as the number of selected bands increases, the classification performance increases faster. However, when the number of bands increased to 20 or even larger, OA and AA performances tend to be stable. Fig. 9(c) and (d) shows the box plots of OA and AA when the number of selected bands is 26.

Fig. 10 shows the distributions of band index selected by seven different BS algorithms ( $NB = 26$ ).

---

### Algorithm HGWO

---

**Input:** Hyperspectral image data:  $HIM$

number of selected bands:  $NB$

population size:  $nPop$

initial population:  $Pop$

maximum iteration number:  $t_{max}$

index (sequence number) of selected bands  $index$

**Initialization:**

Calculate  $J$  of each band according to (13)–(15) and arranged in descending order to  $J$ , then select  $l/2$  bands in order from  $J$  to  $index$ , finally, randomly select band subsets from  $index$  with the number of  $nPop$  groups of  $NB$  as the initial population:

$$Pop = \{b_1, b_2, \dots, b_{NB}\} \in R^{nPop \times NB}$$

**Output:** selected band subsets.

**Steps:**

1. Calculate the objective function of everyone  $\{F_1, F_2, \dots, F_{nPop}\}$  according to (13)–(15).

2. Three individuals with optimal  $F$  were selected as  $X_\alpha, X_\beta, X_\delta$ .

3. **while**  $t < t_{max}$  **do**

$$t = t + 1$$

Calculate  $a$  according to (16)

**for**  $i = 1: nPop$  **do**

**for**  $j = 1: NB$  **do**

Update  $A, C$  according to (3) and (4)

Calculate the objective function  $F$  of all

search agents

Update  $X_\alpha, X_\beta, X_\delta$  according to (6)–(8)

Update the position  $X(i, j)$  according to (9)–(12)

**end for**

**end for**

**end while**

4. Return band subsets.

---

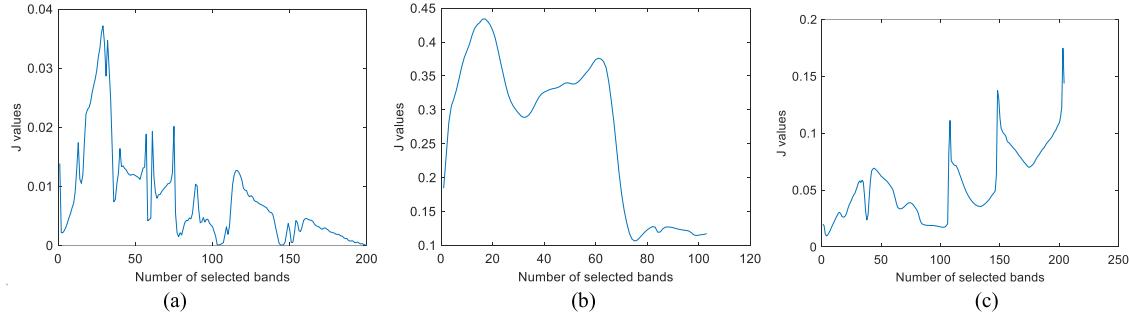
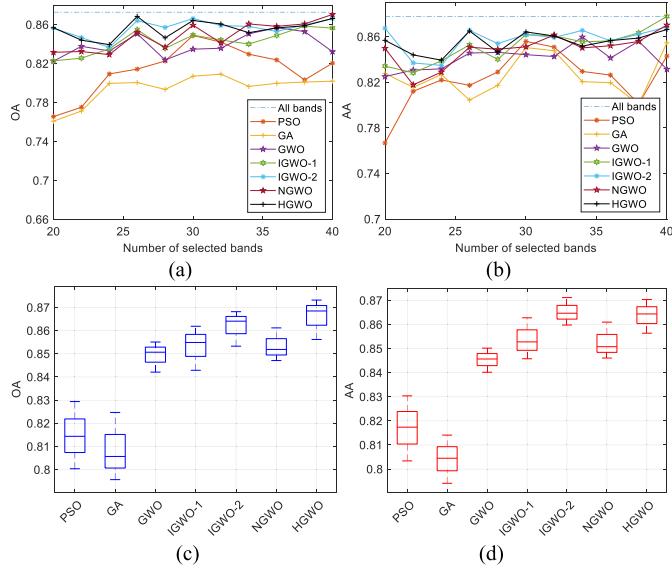
Fig. 8.  $J$  values of each band of (a) Indian Pines, (b) Pavia University, and (c) Salinas.

Fig. 9. Classification results of the Indian Pines dataset. (a) and (b) OA and AA curves by SVM for different numbers of selected bands, respectively. (c) and (d) Box plots of OA and AA when 26 bands are selected.

The classification maps of all the bands and four algorithms are shown in Fig. 11. For a fair comparison, the classification maps are generated by fixed training and testing sets for all the algorithms.

Table IV shows each classification accuracy of 16 categories using all the bands and 26 selected bands by seven BS algorithms. To see the robustness of different BS algorithms, 20 independent runs are recorded to compute the average OA, AA, Kappa, and the corresponding standard deviations. It can be seen that compared with other methods, the proposed HGWO has higher classification accuracy and better stability, and the classification accuracy of every category of HGWO is almost higher than other methods. When 26 bands are selected, the classification accuracy of HGWO is not much different from all the bands; meanwhile, calculation complexity can be reduced, which proves the necessity of BS.

From all the results in Figs. 9–11 and Table IV, it is proven that except for the IGWO-2 algorithm, the proposed HGWO has improved the classification of almost all the categories and obtained the best values of OA, AA, and Kappa, compared with other algorithms. The IGWO-2 algorithm has similar performance as the proposed HGWO, and the best

TABLE III  
DESCRIPTION OF THREE HSI DATASETS

	Indian Pines	Pavia University	Salinas
Size	$145 \times 145$	$610 \times 340$	$512 \times 227$
Bands	200	103	204
Categories	16	9	16
Resolution	20 m	1.3 m	3.7 m
Sensor	AVIRIS	ROSIS	AVIRIS

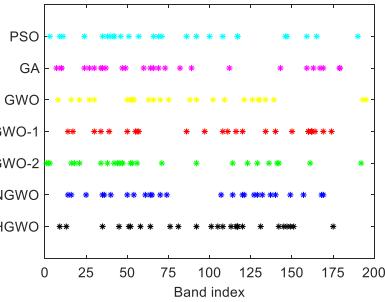


Fig. 10. Distributions of 26 bands selected by different algorithms.

performance for 16 categories appears alternately between IGWO-2 and HGWO; however, the proposed HGWO has much less computational complexity compared with IGWO-2, which will be discussed in Section IV-D.

2) *Band Selection and Classification for Pavia University*: In this section, the performance comparison is given using the Pavia University dataset. Fig. 12 shows the selection and classification results for the Pavia University dataset. Fig. 12(a) and (b) shows OA and AA performances of different BS algorithms for different NB values, and the same conclusion could be given that HGWO has the best classification performance. When  $NB = 22$ , the best classification accuracy would be achieved. Fig. 12(c) and (d) also shows the box plots of OA and AA when the number of selected bands is 22, proving the superiority of the proposed HGWO algorithm. Fig. 13 shows the distributions of band index selected by four different BS algorithms ( $NB = 22$ ).

The classification maps of all the bands and seven BS algorithms are shown in Fig. 14 (when  $NB = 22$ ). The classification maps are generated by fixed training and testing sets for all the algorithms. Table V shows each classification

TABLE IV  
CLASSIFICATION ACCURACY OF ALL BANDS AND DIFFERENT ALGORITHMS ON INDIAN PINES

Category	All bands	PSO	GA	GWO	IGWO-1	IGWO-2	NGWO	HGWO
1	98.00 $\pm$ 5.7	83.26 $\pm$ 6.1	87.50 $\pm$ 7.1	88.89 $\pm$ 5.2	92.58 $\pm$ 5.1	93.97 $\pm$ 4.8	<b>94.12 <math>\pm</math> 5.2</b>	88.24 $\pm$ 5.1
2	84.39 $\pm$ 2.4	75.52 $\pm$ 4.7	72.96 $\pm$ 4.8	81.29 $\pm$ 3.1	78.47 $\pm$ 3.8	80.39 $\pm$ 3.1	72.87 $\pm$ 3.8	<b>83.78 <math>\pm</math> 3.7</b>
3	81.28 $\pm$ 4.9	67.90 $\pm$ 7.2	72.46 $\pm$ 7.4	72.89 $\pm$ 4.9	77.31 $\pm$ 5.3	76.13 $\pm$ 4.2	75.45 $\pm$ 4.9	<b>82.35 <math>\pm</math> 4.6</b>
4	65.87 $\pm$ 4.3	62.07 $\pm$ 5.8	56.80 $\pm$ 7.4	68.38 $\pm$ 5.8	58.74 $\pm$ 5.5	60.32 $\pm$ 4.9	63.39 $\pm$ 6.9	<b>63.43 <math>\pm</math> 6.1</b>
5	90.00 $\pm$ 2.7	80.95 $\pm$ 4.1	80.51 $\pm$ 4.4	<b>91.24 <math>\pm</math> 2.9</b>	89.16 $\pm$ 4.1	88.65 $\pm$ 3.6	87.61 $\pm$ 3.3	85.78 $\pm$ 2.7
6	94.44 $\pm$ 4.1	90.53 $\pm$ 4.3	91.91 $\pm$ 5.2	94.30 $\pm$ 5.1	94.00 $\pm$ 5.7	<b>97.27 <math>\pm</math> 4.8</b>	93.65 $\pm$ 5.8	95.60 $\pm$ 5.3
7	99.98 $\pm$ 0.4	91.20 $\pm$ 0.8	84.62 $\pm$ 0.8	99.81 $\pm$ 0.4	97.23 $\pm$ 0.5	<b>99.89 <math>\pm</math> 0.6</b>	97.56 $\pm$ 0.6	92.31 $\pm$ 0.7
8	99.01 $\pm$ 2.7	97.02 $\pm$ 2.7	95.36 $\pm$ 3.4	98.19 $\pm$ 3.1	97.50 $\pm$ 2.6	<b>98.50 <math>\pm</math> 3.1</b>	98.98 $\pm$ 2.1	98.42 $\pm$ 1.7
9	76.67 $\pm$ 1.1	66.67 $\pm$ 1.1	60.00 $\pm$ 1.7	76.92 $\pm$ 1.4	72.50 $\pm$ 1.0	81.82 $\pm$ 0.9	76.92 $\pm$ 1.3	<b>84.50 <math>\pm</math> 0.8</b>
10	76.97 $\pm$ 4.2	70.25 $\pm$ 5.3	70.30 $\pm$ 4.9	70.88 $\pm$ 5.2	76.50 $\pm$ 4.2	76.36 $\pm$ 3.9	70.94 $\pm$ 4.2	<b>76.72 <math>\pm</math> 5.1</b>
11	89.05 $\pm$ 6.7	83.81 $\pm$ 8.1	82.42 $\pm$ 8.3	87.42 $\pm$ 7.2	86.90 $\pm$ 7.2	88.68 $\pm$ 7.3	85.26 $\pm$ 7.3	<b>88.69 <math>\pm</math> 6.5</b>
12	83.09 $\pm$ 4.3	79.13 $\pm$ 5.0	77.48 $\pm$ 5.7	80.57 $\pm$ 3.5	77.15 $\pm$ 4.0	<b>84.39 <math>\pm</math> 3.9</b>	80.31 $\pm$ 4.1	83.06 $\pm$ 4.1
13	96.70 $\pm$ 3.2	94.00 $\pm$ 2.4	95.60 $\pm$ 3.1	98.94 $\pm$ 2.8	97.67 $\pm$ 2.8	<b>98.95 <math>\pm</math> 3.6</b>	96.92 $\pm$ 3.5	98.50 $\pm$ 3.1
14	96.99 $\pm$ 2.4	92.61 $\pm$ 3.1	91.36 $\pm$ 4.1	96.16 $\pm$ 4.7	<b>97.28 <math>\pm</math> 3.2</b>	96.98 $\pm$ 4.6	95.48 $\pm$ 3.2	96.76 $\pm$ 2.7
15	73.25 $\pm$ 4.2	65.00 $\pm$ 5.7	60.57 $\pm$ 6.7	71.19 $\pm$ 3.9	72.04 $\pm$ 4.7	71.88 $\pm$ 5.1	<b>73.75 <math>\pm</math> 4.3</b>	68.45 $\pm$ 4.1
16	94.44 $\pm$ 2.0	96.44 $\pm$ 2.4	90.24 $\pm$ 3.2	<b>98.23 <math>\pm</math> 3.1</b>	93.75 $\pm$ 2.7	97.56 $\pm$ 3.2	97.50 $\pm$ 3.1	97.50 $\pm$ 2.8
OA (%)	87.16 $\pm$ 1.5	81.16 $\pm$ 2.1	80.49 $\pm$ 1.9	85.07 $\pm$ 2.2	85.17 $\pm$ 1.9	86.51 $\pm$ 2.0	85.20 $\pm$ 2.1	<b>86.85 <math>\pm</math> 1.4</b>
AA (%)	87.51 $\pm$ 1.3	81.02 $\pm$ 1.9	79.38 $\pm$ 2.2	85.96 $\pm$ 1.7	84.92 $\pm$ 1.7	<b>86.98 <math>\pm</math> 2.3</b>	85.04 $\pm$ 1.5	86.51 $\pm$ 1.6
Kappa (%)	84.87 $\pm$ 1.4	77.38 $\pm$ 1.7	77.18 $\pm$ 2.1	81.48 $\pm$ 1.9	82.36 $\pm$ 1.7	83.92 $\pm$ 2.1	82.67 $\pm$ 1.4	<b>84.28 <math>\pm</math> 1.9</b>

TABLE V  
CLASSIFICATION ACCURACY OF ALL BANDS AND DIFFERENT ALGORITHMS ON PAVIA UNIVERSITY

Category	All bands	PSO	GA	GWO	IGWO-1	IGWO-2	NGWO	HGWO
1	96.99 $\pm$ 1.1	96.93 $\pm$ 1.9	97.12 $\pm$ 2.2	91.83 $\pm$ 1.4	96.99 $\pm$ 1.1	92.66 $\pm$ 2.3	93.21 $\pm$ 2.1	<b>97.42 <math>\pm</math> 1.9</b>
2	97.18 $\pm$ 0.5	96.90 $\pm$ 0.4	96.80 $\pm$ 1.7	94.35 $\pm$ 1.3	<b>97.19 <math>\pm</math> 0.9</b>	94.77 $\pm$ 0.7	95.81 $\pm$ 0.4	96.77 $\pm$ 0.4
3	87.90 $\pm$ 3.2	84.84 $\pm$ 4.8	75.07 $\pm$ 4.9	89.91 $\pm$ 3.8	89.34 $\pm$ 3.0	86.73 $\pm$ 2.1	85.96 $\pm$ 1.9	<b>89.50 <math>\pm</math> 2.1</b>
4	93.78 $\pm$ 1.6	91.68 $\pm$ 1.1	83.60 $\pm$ 2.1	93.40 $\pm$ 1.9	81.01 $\pm$ 1.1	93.79 $\pm$ 1.9	<b>95.14 <math>\pm</math> 1.4</b>	92.29 $\pm$ 1.6
5	98.93 $\pm$ 0.1	99.84 $\pm$ 0.8	99.53 $\pm$ 0.7	98.89 $\pm$ 1.1	99.23 $\pm$ 0.8	99.26 $\pm$ 1.1	98.64 $\pm$ 0.6	<b>99.38 <math>\pm</math> 0.4</b>
6	89.18 $\pm$ 1.7	82.45 $\pm$ 2.1	79.29 $\pm$ 3.7	80.91 $\pm$ 3.1	82.14 $\pm$ 2.8	81.78 $\pm$ 2.4	82.11 $\pm$ 1.9	<b>83.22 <math>\pm</math> 1.7</b>
7	82.87 $\pm$ 2.4	71.25 $\pm$ 3.1	71.65 $\pm$ 3.9	78.95 $\pm$ 2.7	78.61 $\pm$ 2.1	<b>82.12 <math>\pm</math> 2.1</b>	78.93 $\pm$ 2.4	79.06 $\pm$ 2.6
8	87.84 $\pm$ 0.3	88.36 $\pm$ 0.7	87.40 $\pm$ 1.7	83.02 $\pm$ 1.2	87.19 $\pm$ 0.8	83.80 $\pm$ 1.4	83.57 $\pm$ 0.5	<b>89.85 <math>\pm</math> 0.7</b>
9	99.78 $\pm$ 0.1	99.07 $\pm$ 1.2	98.71 $\pm$ 1.8	98.92 $\pm$ 1.9	99.78 $\pm$ 0.4	98.80 $\pm$ 1.4	99.64 $\pm$ 0.4	<b>99.98 <math>\pm</math> 0.9</b>
OA (%)	94.21 $\pm$ 0.3	92.07 $\pm$ 0.6	90.84 $\pm$ 1.8	91.41 $\pm$ 1.3	92.37 $\pm$ 0.5	92.59 $\pm$ 1.2	91.49 $\pm$ 0.5	<b>93.57 <math>\pm</math> 0.4</b>
AA (%)	92.71 $\pm$ 0.5	90.15 $\pm$ 0.9	87.73 $\pm$ 1.0	89.57 $\pm$ 1.3	90.16 $\pm$ 0.6	90.41 $\pm$ 1.1	90.34 $\pm$ 0.6	<b>91.94 <math>\pm</math> 0.7</b>
Kappa (%)	91.30 $\pm$ 0.6	88.11 $\pm$ 0.6	86.45 $\pm$ 1.9	89.31 $\pm$ 1.4	88.47 $\pm$ 0.7	89.07 $\pm$ 1.2	89.54 $\pm$ 0.9	<b>89.71 <math>\pm</math> 0.6</b>

accuracy of nine categories using all the bands and 26 selected bands by PSO, GA, GWO, IGWO-1, IGWO-2, NGWO, and HGWO.

Due to the guidance of class separability criterion on the initial population, the classification accuracy of HGWO in each category is higher than 75% when 22 bands are selected. Compared with other six BS algorithms, HGWO has higher classification accuracy and better stability.

3) *Band Selection and Classification for Salinas:* In this section, the performance comparison is given using the Salinas dataset. Fig. 15 shows the selection and classification results,

and the same conclusion could be given that HGWO has the best classification performance mostly; however, IGWO-2 sometimes goes higher than the proposed HGWO especially in Fig. 15(a) for OA comparison; this will also be compared in detail in Table VI for each category. When NB = 28, the best classification accuracy would be achieved. Fig. 15(c) and (d) also shows the box plots of OA and AA when the number of selected bands is 28, proving the superiority of the proposed HGWO algorithm.

Fig. 16 shows the distributions of band index selected by four different BS algorithms (NB = 28).

TABLE VI  
CLASSIFICATION ACCURACY OF ALL BANDS AND DIFFERENT ALGORITHMS ON SALINAS

Category	All bands	PSO	GA	GWO	IGWO-1	IGWO-2	NGWO	HGWO
1	99.93 $\pm$ 0.4	99.34 $\pm$ 0.4	98.89 $\pm$ 1.2	98.99 $\pm$ 1.1	98.75 $\pm$ 0.6	98.98 $\pm$ 0.8	99.31 $\pm$ 0.7	<b>99.91 <math>\pm</math> 0.6</b>
2	99.94 $\pm$ 0.5	99.29 $\pm$ 0.3	99.29 $\pm$ 0.9	99.12 $\pm$ 0.7	99.87 $\pm$ 0.7	99.12 $\pm$ 0.4	<b>99.95 <math>\pm</math> 0.3</b>	99.84 $\pm$ 0.4
3	97.30 $\pm$ 0.2	99.70 $\pm$ 1.8	99.11 $\pm$ 1.7	98.87 $\pm$ 1.5	99.11 $\pm$ 1.4	<b>99.77 <math>\pm</math> 0.5</b>	99.32 $\pm$ 0.8	98.09 $\pm$ 1.1
4	99.09 $\pm$ 0.8	98.07 $\pm$ 0.7	96.82 $\pm$ 0.9	99.18 $\pm$ 1.1	99.34 $\pm$ 0.8	99.34 $\pm$ 0.8	<b>99.34 <math>\pm</math> 0.4</b>	98.95 $\pm$ 0.2
5	99.36 $\pm$ 0.4	99.12 $\pm$ 0.8	99.26 $\pm$ 1.5	99.30 $\pm$ 0.6	99.02 $\pm$ 0.9	99.30 $\pm$ 1.1	<b>99.30 <math>\pm</math> 0.5</b>	99.26 $\pm$ 0.3
6	99.88 $\pm$ 0.3	98.98 $\pm$ 0.7	99.52 $\pm$ 0.7	99.96 $\pm$ 0.9	99.97 $\pm$ 0.8	99.92 $\pm$ 0.7	99.22 $\pm$ 0.5	<b>99.98 <math>\pm</math> 0.7</b>
7	99.99 $\pm$ 0.6	99.46 $\pm$ 0.4	99.32 $\pm$ 0.6	99.90 $\pm$ 0.8	99.97 $\pm$ 0.4	98.86 $\pm$ 0.4	99.13 $\pm$ 0.9	<b>99.97 <math>\pm</math> 0.1</b>
8	89.04 $\pm$ 2.4	81.84 $\pm$ 2.4	82.28 $\pm$ 3.8	83.53 $\pm$ 2.8	84.06 $\pm$ 2.8	<b>84.83 <math>\pm</math> 1.8</b>	83.68 $\pm$ 1.6	84.17 $\pm$ 1.4
9	99.44 $\pm$ 0.9	99.37 $\pm$ 0.8	99.57 $\pm$ 0.7	99.57 $\pm$ 1.1	99.60 $\pm$ 0.6	99.64 $\pm$ 0.9	99.62 $\pm$ 0.8	<b>99.89 <math>\pm</math> 0.5</b>
10	97.00 $\pm$ 0.8	97.84 $\pm$ 0.8	97.03 $\pm$ 1.9	94.56 $\pm$ 1.8	97.46 $\pm$ 1.7	<b>98.55 <math>\pm</math> 1.5</b>	98.32 $\pm$ 1.6	97.85 $\pm$ 1.4
11	96.69 $\pm$ 0.9	99.06 $\pm$ 1.7	99.06 $\pm$ 1.9	97.50 $\pm$ 2.1	98.69 $\pm$ 0.9	<b>99.58 <math>\pm</math> 1.7</b>	96.50 $\pm$ 1.9	97.76 $\pm$ 0.8
12	99.46 $\pm$ 0.7	97.58 $\pm$ 0.9	97.00 $\pm$ 1.8	99.77 $\pm$ 0.9	99.61 $\pm$ 1.7	99.54 $\pm$ 1.2	99.77 $\pm$ 0.6	<b>99.84 <math>\pm</math> 0.9</b>
13	99.99 $\pm$ 0.5	96.98 $\pm$ 0.6	95.95 $\pm$ 0.9	98.89 $\pm$ 1.2	98.85 $\pm$ 0.3	99.25 $\pm$ 0.5	99.31 $\pm$ 0.7	<b>99.78 <math>\pm</math> 0.2</b>
14	96.73 $\pm$ 1.1	97.74 $\pm$ 1.7	90.81 $\pm$ 1.8	97.54 $\pm$ 0.8	97.41 $\pm$ 1.4	<b>97.95 <math>\pm</math> 0.5</b>	97.74 $\pm$ 1.2	97.42 $\pm$ 0.7
15	84.73 $\pm$ 2.7	73.45 $\pm$ 3.7	78.80 $\pm$ 3.9	78.53 $\pm$ 3.1	76.18 $\pm$ 2.3	80.00 $\pm$ 2.1	82.22 $\pm$ 2.1	<b>85.71 <math>\pm</math> 2.1</b>
16	97.23 $\pm$ 1.4	99.41 $\pm$ 1.3	98.17 $\pm$ 1.7	97.82 $\pm$ 1.3	98.82 $\pm$ 1.9	<b>99.63 <math>\pm</math> 1.2</b>	98.18 $\pm$ 1.2	98.32 $\pm$ 1.6
OA (%)	92.56 $\pm$ 0.6	91.02 $\pm$ 0.6	90.81 $\pm$ 1.1	91.11 $\pm$ 1.2	92.28 $\pm$ 0.7	92.14 $\pm$ 1.1	91.96 $\pm$ 0.3	<b>92.52 <math>\pm</math> 0.4</b>
AA (%)	97.24 $\pm$ 0.5	96.08 $\pm$ 0.7	95.65 $\pm$ 1.8	96.44 $\pm$ 1.3	96.67 $\pm$ 0.5	97.14 $\pm$ 1.3	96.93 $\pm$ 0.4	<b>97.21 <math>\pm</math> 0.5</b>
Kappa (%)	91.59 $\pm$ 0.7	89.24 $\pm$ 0.9	88.99 $\pm$ 1.2	89.4 $\pm$ 1.3	91.39 $\pm$ 0.5	91.25 $\pm$ 1.3	90.51 $\pm$ 0.4	<b>91.63 <math>\pm</math> 0.6</b>

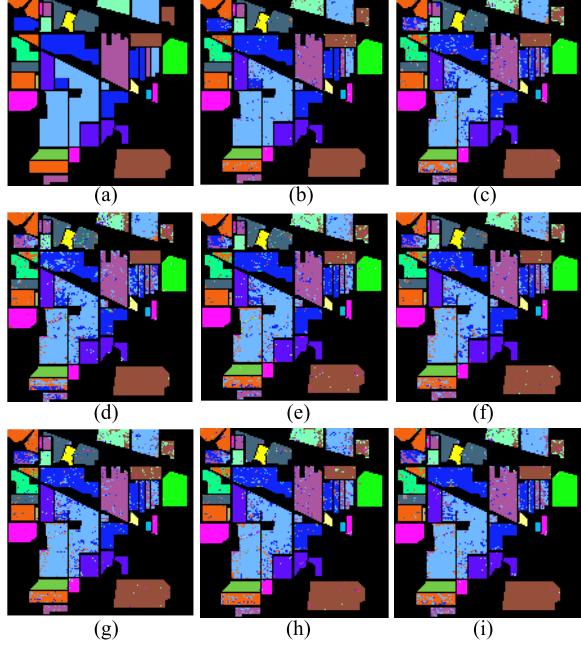


Fig. 11. (a) Color ground-truth image with class labels of Indian Pines. (b)–(i) Classification results of 26 bands were selected on Indian Pines by all the bands, PSO, GA, GWO, IGWO-1, IGWO-2, NGWO, and HGWO, respectively.

The classification maps of all the bands and seven BS algorithms are shown in Fig. 17 (when NB = 28). The classification maps are generated by fixed training and testing sets for all the algorithms.

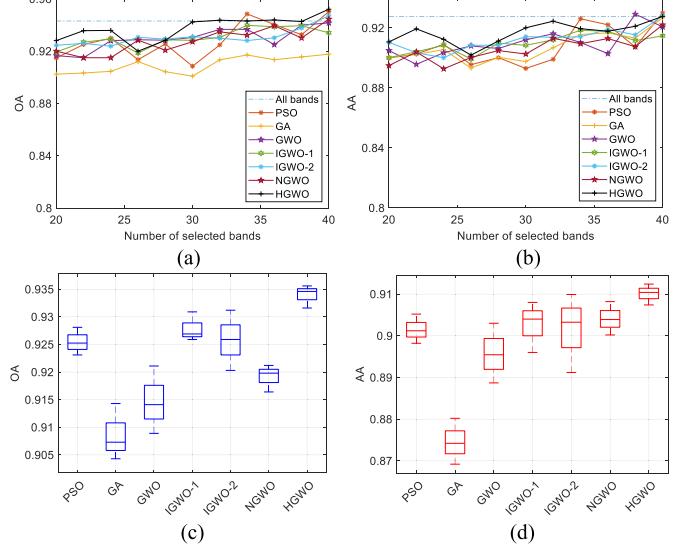


Fig. 12. Classification results of the Pavia University dataset. (a) and (b) OA and AA curves by SVM for different numbers of selected bands, respectively. (c) and (d) Box plots of OA and AA when 22 bands are selected.

The experiments for the Salinas dataset have similar conclusion as the Indian Pines dataset in Section I. It is proven that except for IGWO-2, the proposed HGWO has improved the classification of almost all the categories and obtained the best values of OA, AA, and Kappa, compared with other algorithms. The IGWO-2 algorithm has similar performance as the proposed HGWO, and the best performance for 16 categories appears alternately between IGWO-2 and HGWO;

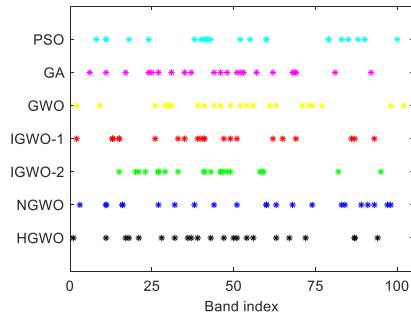


Fig. 13. Index of 22 bands selected by different algorithms.

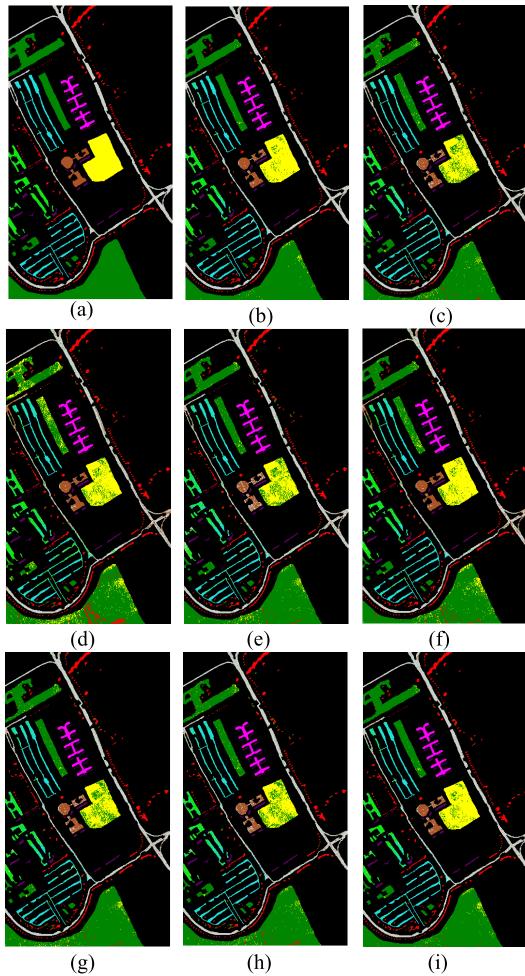


Fig. 14. (a) Color ground-truth image with class labels of Pavia University. (b)–(i) Classification results of 22 bands were selected on Pavia University by all the bands, PSO, GA, GWO, IGWO-1, IGWO-2, NGWO, and HGWO, respectively.

however, the proposed HGWO has much less computational complexity compared with IGWO-2, which will be discussed in Section IV-D.

#### C. Effect of Hyperparameter $nPop$ on Accuracy

In the mini code of the proposed algorithm,  $nPop$  is the population size in the initial part, and it is set to  $nPop = 30$  in our experiment. However, this parameter might impact the

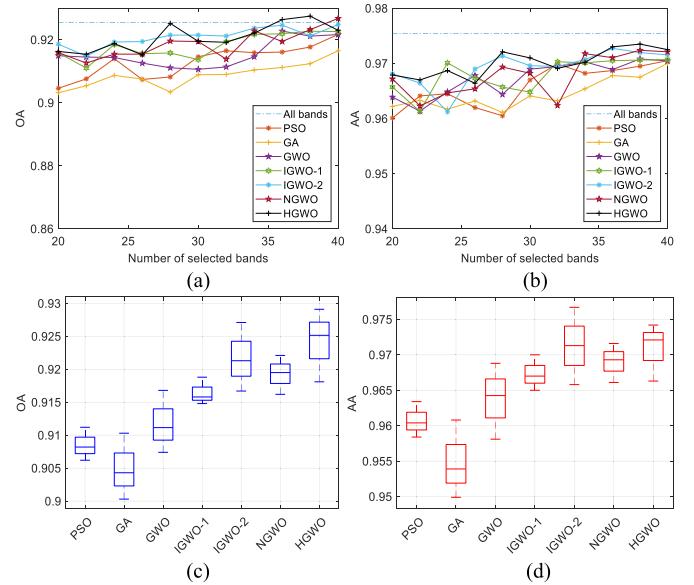


Fig. 15. Classification results of the Salinas dataset. (a) and (b) OA and AA curves by SVM for different numbers of selected bands, respectively. (c) and (d) Box plots of OA and AA when 28 bands are selected.

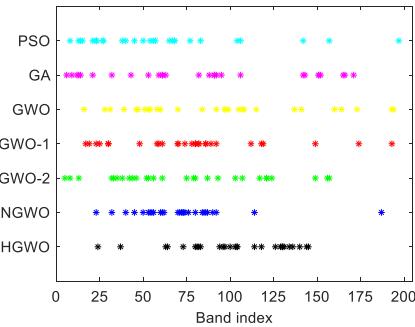


Fig. 16. Index of 28 bands selected by different algorithms.

performance of the algorithm not only on the computing time but also on the accuracy performance. In this section, the analysis of the effect of hyperparameter  $nPop$  on accuracy is given using the Indian Pines dataset, and the analysis of the effect on computational time would be given in Section IV-D.

The OA and AA curves by SVM for different numbers of  $nPop$  on Indian Pines are plotted in Fig. 18. We can see from the OA curve in Fig. 18(a) that the classification accuracy reached the maximum value when  $nPop = 30$  or 50. However, the computational complexity will be increased when  $nPop = 50$ , and classification accuracy was not much different from  $nPop = 30$ , so the number of populations is chosen to be  $nPop = 30$  for this experiment.

#### D. Comparison of Computational Complexity

To verify the efficiency of the proposed algorithm, this section compares the computational complexity of different BS algorithms on three different hyperspectral datasets. The experiments were run 20 times, and the average computational time was recorded as shown in Table VII(a)–(c) for different  $nPop$  values. In order for a fair comparison, the number

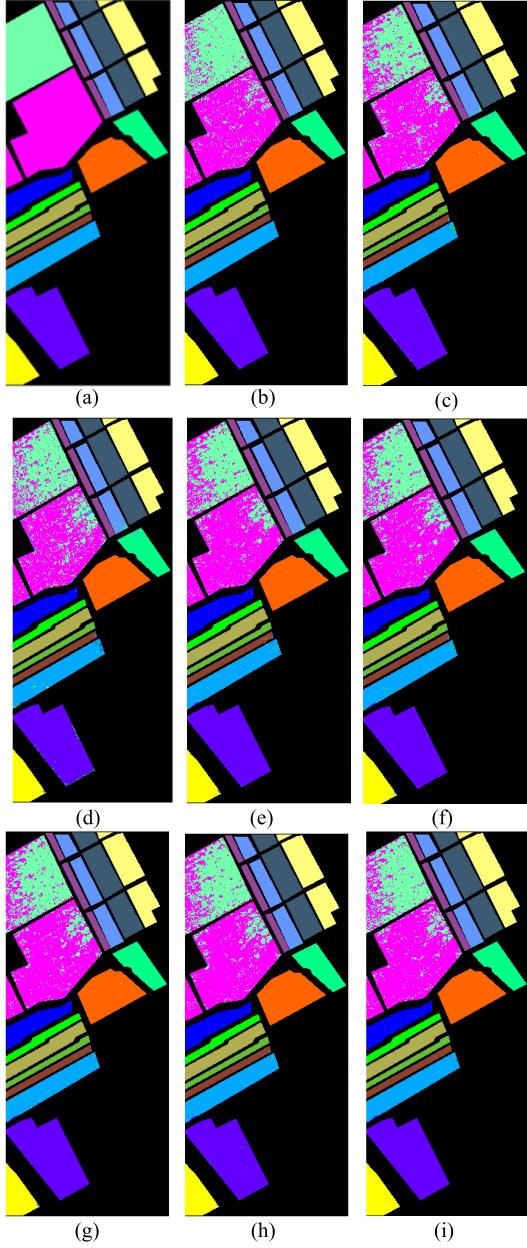


Fig. 17. (a) Color ground-truth image with class labels of Salinas. (b)–(i) Classification results of 28 bands were selected on Salinas by all the bands, PSO, GA, GWO, IGWO-1, IGWO-2, NGWO, and HGWO, respectively.

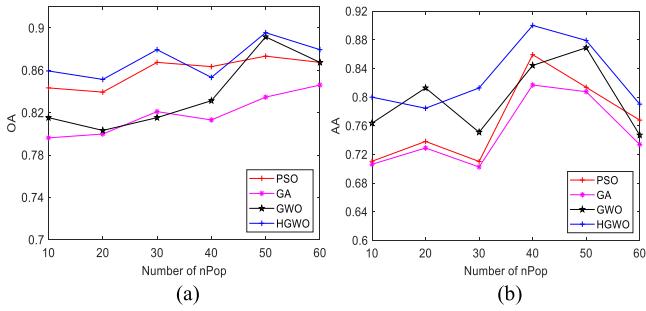


Fig. 18. (a) and (b) OA and AA curves by SVM for different numbers of nPop on Indian Pines.

of bands is the same for each dataset. It can be seen that NGWO has the shortest computation time among all the

TABLE VII

- (a) AVERAGE COMPUTATIONAL TIME OF DIFFERENT BS METHODS ON THE INDIAN PINES DATASET. (b) AVERAGE COMPUTATIONAL TIME OF DIFFERENT BS METHODS ON THE PAVIA UNIVERSITY DATASET.  
(c) AVERAGE COMPUTATIONAL TIME OF DIFFERENT BS METHODS ON THE SALINAS DATASET

Algorithms	Computing time for Indian Pines (NB=26)			
	nPop=20	nPop=30	nPop=40	nPop=50
PSO/second	41.1	67.1	85.7	106.8
GA/second	139.9	207.5	286.0	362.7
GWO/second	39.9	59.6	77.6	97.8
GWO-1/second	44.3	63.7	88.2	113.9
GWO-2/second	921.3	1731.4	2614.9	3264.5
NGWO/second	<b>38.4</b>	<b>57.6</b>	<b>74.8</b>	<b>95.5</b>
HGWO/second	40.3	61.3	82.0	102.9

Algorithms	Computing time for Pavia University (NB=22)			
	nPop=20	nPop=30	nPop=40	nPop=50
PSO/second	196.9	307.0	412.8	519.9
GA/second	1195.5	1416.6	1873.8	2391.3
GWO/second	198.3	309.1	414.6	511.8
IGWO-1/second	206.4	328.2	431.3	5430.7
IGWO-2/second	23307.9	32451.8	39786.5	49871.3
NGWO/second	<b>195.2</b>	<b>303.1</b>	<b>406.6</b>	<b>502.3</b>
HGWO/second	201.6	319.0	425.4	520.3

Algorithms	Computing time for Salinas (NB=28)			
	nPop=20	nPop=30	nPop=40	nPop=50
PSO/second	155.6	239.2	309.9	393.8
GA/second	1143.7	1648.3	1993.4	2364.3
GWO/second	154.8	231.1	307.2	386.3
IGWO-1/second	171.3	250.4	325.9	418.0
IGWO-2/second	16470.8	25735.7	33267.8	42677.9
NGWO/second	<b>149.7</b>	<b>228.5</b>	<b>304.3</b>	<b>378.4</b>
HGWO/second	158.5	239.7	312.9	401.2

algorithms, because it does not calculate the  $J$  value and changes the convergence factor to accelerate convergence. Although HGWO has a slightly lower computation time than NGWO, it can achieve the best classification performance on all three datasets. Therefore, the proposed algorithm ensures the classification superiority while the calculation time is acceptable.

## V. CONCLUSION

As a non-transformed feature selection, BS has been widely used to select representative band subset and reduce computational complexity for later process. There are many scientific problems for hyperspectral BS: 1) how to determine the number of bands for the selected band subset; 2) how to selected

high-quality band subsets without label information; 3) how to reduce the computing time when searching the optimal band subset because most of the BS algorithms are searching algorithms and sometimes would fall into local optimality; and 4) how to evaluate the quality of the selected band subset. As for point 1, the most widely used method is using visual dimensionality (VD) to determine the number of bands to be selected, and point 2 is usually called unsupervised BS, which are not included in this article. This article focuses on the 3rd and 4th problems and develops a new meta-heuristic algorithm, a hybrid GWO (HGWO), for hyperspectral BS. There are two main innovations of the proposed HGWO algorithm:

- 1) Nonlinear convergence factor to realize quick and fine search. In view of the fact that the linear convergence factor of GWO is easy to fall into local optimal, the proposed HGWO algorithm uses a new nonlinear convergence factor instead of the linear factor, which could search slowly at the initial stage of the iteration with a little shrinkage range to realize fine search, and then with a quick search speed at the later stage of the iteration to realize fast search. Using five functions to test the performance of the convergence factor in Section III, the proposed HGWO has the best convergence performance through all five functions, as shown in Fig. 4 and Table II. Further comparison for hyperspectral BS has been conducted on three real HSIs, and the computing time using the proposed HGWO is better than the conventional GWO and GA, similar to PSO, but the classification performance using band subset obtained by HGWO is better than the PSO algorithm;
- 2) Adding category separability to help HGWO become a BS method beneficial to classification. From the results, it could be seen that the classification performances are compared with other BS algorithms, which could be proved by Tables IV–VI and Figs. 9, 11, 12, 14, 15, and 17, for all three HSI datasets. The results of comparative experiments on three typical hyperspectral datasets successfully confirmed that the proposed HGWO algorithm is more effective than several existing methods;
- 3) HGWO and IGWO-2 comparison. The proposed HGWO is undisputed better performed than PSO, GA, GWO, GWO-1, and NGWO. However, it has similar performance as IGWO-2 especially for the Indian Pines data and Salinas datasets. Through computational time comparison, the proposed HGWO has been proven to be much less computational complexity than the IGWO-2 algorithm, showing its superior time performance with more than 23 times speed-up ratio. This is because the main difference between HGWO and IGWO-2 is the use of nonlinear convergence factor, and it has been discussed in 1) that the nonlinear convergence factor could gain quick and fine search.

## REFERENCES

- [1] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 51–64, Jan. 2017.
- [2] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral band selection by multitask sparsity pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 631–644, Feb. 2015.
- [3] W. Xia, B. Wang, and L. Zhang, "Band selection for hyperspectral imagery: A new approach based on complex networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1229–1233, Sep. 2013.
- [4] Y. Wang, L. Wang, H. Xie, and C.-I. Chang, "Fusion of various band selection methods for hyperspectral imagery," *Remote Sens.*, vol. 11, no. 18, p. 2125, 2019.
- [5] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, Jan. 2011.
- [6] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 1–13, Oct. 2018.
- [7] J. Yue, L. Fang, H. Rahmani, and P. Ghamisi, "Self-supervised learning with adaptive distillation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.
- [8] Y. Wang *et al.*, "Constrained-target band selection for multiple-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6079–6103, Aug. 2019.
- [9] W. Xie, Y. Li, J. Lei, J. Yang, C.-I. Chang, and Z. Li, "Hyperspectral band selection for spectral–spatial anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3426–3436, May 2020.
- [10] S. Yang, Z. Shi, and W. Tang, "Robust hyperspectral image target detection using an inequality constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3389–3404, Jun. 2015.
- [11] X. Shang *et al.*, "Target-constrained interference-minimized band selection for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 6044–6064, Jul. 2020.
- [12] X. Cao, T. Xiong, and L. Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 1–5, 2016.
- [13] L. C. B. Dos Santos, S. J. F. Guimaraes, A. D. A. Araujo, and J. A. Dos Santos, "Unsupervised hyperspectral band selection based on spectral rhythm analysis," presented at the 27th SIBGRAPI Conf. Graph., Patterns Images, 2014.
- [14] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2956–2969, May 2015.
- [15] X. Bai, Z. Guo, Y. Wang, Z. Zhang, and J. Zhou, "Semisupervised hyperspectral band selection via spectral–spatial hypergraph model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2774–2783, Jun. 2015.
- [16] F. Li, P. Zhang, and L. Huchuan, "Unsupervised band selection of hyperspectral images via multi-dictionary sparse representation," *IEEE Access*, vol. 6, pp. 71632–71643, 2018.
- [17] F. Xie, C. Lei, F. Li, D. Huang, and J. Yang, "Unsupervised hyperspectral feature selection based on fuzzy c-means and grey wolf optimizer," *Int. J. Remote Sens.*, vol. 40, no. 9, pp. 3344–3367, 2018.
- [18] X. Cao, C. Wei, Y. Ge, J. Feng, J. Zhao, and L. Jiao, "Semi-supervised hyperspectral band selection based on dynamic classifier selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1289–1298, Apr. 2019.
- [19] X. Luo, Z. Shen, R. Xue, and H. Wan, "Unsupervised band selection method based on importance-assisted column subset selection," *IEEE Access*, vol. 7, pp. 517–527, 2019.
- [20] J. Tschanterl *et al.*, "MIMR-DGSA: Unsupervised hyperspectral band selection based on information theory and a modified discrete gravitational search algorithm," *Inf. Fusion*, vol. 51, pp. 189–200, Nov. 2019.
- [21] H. Zhao, L. Bruzzone, R. Guan, F. Zhou, and C. Yang, "Spectral–spatial genetic algorithm-based unsupervised band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9616–9632, Nov. 2021.
- [22] J. Feng *et al.*, "Deep reinforcement learning for semisupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2021.
- [23] Y. Cai, X. Liu, and Z. Cai, "BS-Nets: An end-to-end framework for band selection of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1969–1984, Nov. 2020.
- [24] P. R. Lorenzo, L. Tulczyjew, M. Marcinkiewicz, and J. Nalepa, "Hyperspectral band selection using attention-based convolutional neural networks," *IEEE Access*, vol. 8, pp. 42384–42403, 2020.

- [25] J. Feng *et al.*, "Dual-graph convolutional network based on band attention and sparse constraint for hyperspectral band selection," *Knowl.-Based Syst.*, vol. 231, Nov. 2021, Art. no. 107428.
- [26] N. He, L. Fang, and A. Plaza, "Hybrid first and second order attention Unet for building segmentation in remote sensing images," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 1–12, Apr. 2020.
- [27] D. Saqui, J. H. Saito, and L. Jorge, "Methodology for band selection of hyperspectral images using genetic algorithms and Gaussian maximum likelihood classifier," presented at the CSCI, Las Vegas, NV, USA, Dec. 15–17, 2016.
- [28] H. Su, B. Yong, and Q. Du, "Hyperspectral band selection using improved firefly algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, Jan. 2016.
- [29] A. Zhang *et al.*, "Hyperspectral band selection using crossover-based gravitational search algorithm," *IET Image Process.*, vol. 13, no. 2, pp. 280–286, Feb. 2019.
- [30] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 773–777, May 2017.
- [31] H. Su, Q. Du, G. Chen, and P. Du, "Optimized hyperspectral band selection using particle swarm optimization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2659–2670, Jun. 2014.
- [32] C. Q. Wendong Gai, J. Liu, and J. Zhang, "An improved grey wolf algorithm for global optimization," presented at the 30th CCDC, Shen Yang, China, Jun. 7 2018.
- [33] H. Xu, X. Liu, and J. Su, "An improved grey wolf optimizer algorithm integrated with cuckoo search," in *Proc. 9th IEEE Int. Conf. Intell. Data Acquisition Adv. Comput. Syst., Technol. Appl. (IDAACS)*, Bucharest, Romania, Sep. 2017, pp. 21–23.
- [34] S. A. Medjahed, T. A. Saadi, A. Benyettou, and M. Ouali, "Gray wolf optimizer for hyperspectral band selection," *Appl. Soft Comput.*, vol. 40, pp. 178–186, Mar. 2016.
- [35] F. Xie, F. Li, C. Lei, and L. Ke, "Representative band selection for hyperspectral image classification," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 9, p. 338, Aug. 2018.
- [36] S. S. Hashjin, A. D. Boloorani, S. Khazai, and A. A. Kakroodi, "Selecting optimal bands for sub-pixel target detection in hyperspectral images based on implanting synthetic targets," *IET Image Process.*, vol. 13, no. 2, pp. 323–331, Feb. 2019.
- [37] S. Li, Z. Zheng, Y. Wang, C. Chang, and Y. Yu, "A new hyperspectral band selection and classification framework based on combining multiple classifiers," *Pattern Recognit. Lett.*, vol. 83, pp. 152–159, Nov. 2016.
- [38] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral-spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2565–2574, May 2014.
- [39] X. Cao, C. Wei, J. Han, and L. Jiao, "Hyperspectral band selection using improved classification map," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2147–2151, Nov. 2017.
- [40] S. Patra, P. Modi, and L. Bruzzone, "Hyperspectral band selection based on rough set," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5495–5503, Oct. 2015.
- [41] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.
- [42] A. Zhu, C. Xu, Z. Li, J. Wu, and Z. Liu, "Hybridizing grey Wolf optimization with differential evolution for global optimization and test scheduling for 3D stacked SoC," *J. Syst. Eng. Electron.*, vol. 26, no. 2, pp. 317–328, Apr. 2015.
- [43] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, Dec. 2019.
- [44] Y. Xu, Q. Du, and N. H. Younan, "Particle swarm optimization-based band selection for hyperspectral target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 554–558, Apr. 2017.



**Yulei Wang** (Member, IEEE) was born in Yantai, Shandong, China, in 1986. She received the B.S. and Ph.D. degrees in signal and information processing from Harbin Engineering University, Harbin, China, in 2009 and 2015, respectively.

She has been with Dalian Maritime University, Dalian, China, since 2016. She is currently an Associate Professor with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian. She used to be an Algorithm Engineer of advanced driver assistance systems in Beijing Jingwei Hirain Technologies Company Inc., Beijing, China. Her research interests include hyperspectral image processing, vital signs' signal processing, and recognition in vision systems of auto-driving.

Dr. Wang was awarded by the China Scholarship Council in 2011 as a joint Ph.D. student to study in Remote Sensing Signal and Image Processing Laboratory, University of Maryland at Baltimore, Baltimore, MD, USA, for two years.



**Qingyu Zhu** was born in Suzhou, Anhui, China, in 1997. She received the B.S. degree in electronic information science and technology from the Information Science and Technology College, Dalian Maritime University, Dalian, Liaoning, China, in 2019, where she is currently pursuing the M.S. degree in information and communication engineering.

Her research interest includes hyperspectral image processing techniques, especially hyperspectral band selection and classification.



**Haipeng Ma** was born in Suzhou, Anhui, China, in 1997. He received the B.S. degree in electronic information engineering from the Information Science and Technology College, Dalian Maritime University, Dalian, China, in 2019, where he is currently pursuing the M.S. degree in information and communication engineering.

His research interest includes hyperspectral image processing techniques, especially deep learning for hyperspectral band selection and classification.



**Haoyang Yu** (Member, IEEE) received the B.S. degree in information and computing science from Northeastern University, Shenyang, China, in 2013, and the Ph.D. degree in cartography and geographic information system from the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), Beijing, China, 2019.

He is currently the Xing Hai Associate Professor with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. His research interests include models and algorithms for hyperspectral image processing, analysis, and applications.