remote sensing

MDPI

Article

Few-Shot Shrub Identification via Hyperspectral Deep Learning: A Case Study of *Caragana microphylla* Lam in Shrub-Encroached Grasslands

Long Chen [1,2], Jie Wang [1,2,3], Bin Sun [1,2,*] and Zhihai Gao [1,2]

[1] Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China; chenlong@ifrit.ac.cn (L.C.); 2043206000178@ynnu.edu.cn (J.W.); zhgao@ifrit.ac.cn (Z.G.)
[2] Laboratory of Remote Sensing for Forestry and Grassland Monitoring and Evaluation, National Forestry and Grassland Administration, Beijing 100091, China
[3] Faculty of Geography, Yunnan Normal University, Kunming 650050, China
* Correspondence: sunbin@ifrit.ac.cn

Highlights

What are the main findings?

- An unsupervised band selection algorithm (US-BS-Net) incorporating contrastive learning was proposed, which effectively identifies sensitive spectral bands from UAV hyperspectral data and reduces data redundancy while preserving discriminative feature representations.
- The US-BS-P-Net framework, combining US-BS-Net with a prototypical network, achieved a high overall accuracy of 93.15% for shrub identification with 80 selected bands, outperforming full spectrum inputs and other deep learning models in small-sample scenarios.

What are the implications of the main findings?

- The deep synergy between hyperspectral technology and deep learning provides an effective solution for distinguishing between shrubs and herbs, such as spectral similarity and mixed spatial distribution, thereby enabling more accurate and scientific monitoring of shrub-encroached grasslands.
- The proposed framework offers a practical and efficient approach for fine-grained shrub distribution mapping, which can support informed decision-making in grassland management, ecological assessment, and targeted prevention or control strategies for shrub encroachment.

Abstract

Accurately extracting the spatial distribution of shrubs is an important basis for scientific diagnosis, rational prevention, and control of shrub-encroached grasslands (SGs). In SGs, the landscape exhibits a high degree of spatial intermingling among shrubs, herbaceous vegetation, and bare soil, resulting in a serious issue of mixed pixels. The combination of UAV hyperspectral imaging and deep learning provides the most promising technical approach for shrub–grass separation at present. However, issues such as data redundancy and algorithmic adaptability urgently need to be addressed. In this study, *Caragana microphylla* Lam, a typical shrub species found in the SGs of Inner Mongolia, is the subject for extraction. This work aims to extract sensitive spectral bands from UAV hyperspectral data, construct a deep learning-based framework for shrub identification, and map the distribution of *Caragana microphylla* Lam in the study area. First, a dataset of hyperspectral images in Xilinhot City, Inner Mongolia, China, was constructed by manual visual interpretation

Check for updates

Academic Editor: Nikolay S. Strigul

Received: 22 July 2026
Revised: 20 August 2026
Accepted: 25 August 2026
Published: 26 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

for shrub identification. Second, a novel unsupervised band selection algorithm, US-BS-Net, was proposed to optimize the proxy task of BS-Conv-Net by introducing contrastive learning. It can be used to obtain sensitive spectral bands with discriminative feature representations. Finally, the US-BS-P-Net framework was proposed by combining the US-BS-Net with the prototypical network to construct high-precision shrub identification models in small-sample scenarios. Taking 80 bands as input, the highest overall accuracy reaches 93.15%, which is better than the full spectrum and other deep learning models. The deep synergy between hyperspectral technology and deep learning effectively overcomes traditional challenges such as spectral similarity between shrubs and grasses and their mixed spatial distribution. In particular, the proposed US-BS-P-Net, which combines the advantages of US-BS-Net in sensitive band extraction and the prototypical network in few-shot model construction, provides excellent fundamental data for precise monitoring of SGs.

Keywords: shrub-encroached grasslands; shrub identification; contrastive learning; unsupervised band selection; prototypical network

1. Introduction

The shrub encroachment of global grassland ecosystems has significantly increased, mainly manifested by the expansion of shrubs [1,2]. Driven by multiple factors, including climate change, rising atmospheric CO_2 concentrations, overgrazing, and reduced fire disturbance, shrubs are gradually replacing herbs with their strong competitiveness, becoming the dominant species in communities or regions [3]. Statistics show that the global shrub-encroached grasslands (SGs) reached 5.52 × 106 km^2 in 2018 [4]. The SGs in China are mainly distributed in the Inner Mongolia Autonomous Region, and their dominant species is *Caragana microphylla* Lam, covering an area of over 5.1 × 104 km^2 [5,6]. Shrub encroachment, a structural change in grassland ecosystems, not only reshapes the spatial distribution of vegetation but also poses a serious challenge to grassland ecosystem services and the sustainable development of animal husbandry [7]. Therefore, accurately monitoring the extent of SGs and understanding their dynamic spatial patterns support scientific decision-making for early warning of grassland degradation and differentiated management.

A space–air–ground remote sensing system has been established, with multiple platforms, multi-modal, and three-dimensional observation capabilities, making it possible to monitor SGs at multiple scales [8,9]. Satellite-based remotely sensed data has become the main technique for identifying SGs due to its large observation range and low data acquisition cost [10,11]. However, it is difficult to accurately identify individual or young shrubs due to the limitations of spatial resolution and the availability of dense temporal data [12]. Unmanned aerial vehicle (UAV) remote sensing has the advantages of high spatial resolution, minimal impact from cloud cover, flexibility, low data acquisition cost, and rapid timeliness. UAVs can carry a variety of sensors, including multispectral, hyperspectral, and LiDAR, to acquire high-resolution, multimodal remotely sensed data. A previous study has shown that in scenes with simple shrub–grass distribution, such as desert steppes, high-precision identification of shrubs can be achieved using UAV multispectral images with visible and near-infrared spectra [13]. However, it is still difficult to effectively distinguish the spectral mixing features of shrubs and grasses in typical grasslands [14]. Considering this defect, the high-resolution diagnostic spatial–spectral features of UAV hyperspectral images can effectively distinguish between different ground

materials and vegetation types [15], potentially facilitating the precise identification of shrubs within different types of SGs.

The high spatial–spectral resolution of UAV hyperspectral data means large data volume and feature redundancy [16]. Traditional feature engineering methods based on prior knowledge are often targeted at specific scenarios, have weak generalization capabilities, and are inefficient in extracting features from high-dimensional remotely sensed data [17], hindering the construction of accurate shrub identification models. Dimensionality reduction in hyperspectral data to extract representative spatial–spectral features is an effective way to filter out redundant information [18,19]. Related dimensionality reduction methods can be divided into feature extraction and band selection. The former maps the high-dimensional features of hyperspectral data into a low-dimensional space according to a certain feature transformation principle [20], while band selection directly selects the most representative bands from the spectral information of the original hyperspectral image [21–23]. Feature extraction methods maximize the preservation of hyperspectral data information, but their destructive dimensionality reduction leads to weak interpretability. In contrast, band selection offers advantages in preserving the physical meaning of bands and enhancing scientific understanding of ground object characteristics [24]. But how to develop a scientific and reasonable criterion for assessing band importance is still a challenge. In the field of deep learning, band selection algorithms assist spectral reconstruction tasks with self-supervised learning paradigms to extract optimal bands [25,26]. For example, a previous study proposed a band selection algorithm based on deep convolutional neural network, i.e., BS-Net-Conv, whose core idea is to reconstruct complete spectral information using a few informative bands [27]. However, BS-Net-Conv relies on a global autoencoder for its proxy task and lacks the learning of discriminative spectral features, which may cause the omission of key information in important bands. Contrastive learning has significant advantages in learning discriminative feature representations by comparing the similarity of positive and negative samples in the embedding space [28]. Inspired by it, this study will fully utilize the innovation of contrastive learning in the expression of discriminative features to improve BS-Conv-Net and optimize the spectral band selection from UAV hyperspectral images.

Generally, the more parameters a deep learning network has, the more complex the training process will be and the larger the sample size required for model training. In particular, the foundation models in computer vision that have emerged in recent years have put forward higher requirements on computer performance and sample construction [29,30]. The strong generalization ability of deep learning models in various downstream tasks is often closely related to high-quality and large sample quantities. Currently, there is still a lack of large-scale datasets for identification of grassland shrubland using remotely sensed data, and manual visual interpretation of samples is inefficient and costly. The imbalance between the high-dimensional spatial–spectral features of hyperspectral data and the small number of labeled samples is the main reason for overfitting of identification models. Although unsupervised and semi-supervised algorithms can alleviate model overfitting caused by insufficient training samples to a certain extent, their results largely depend on the preset auxiliary training tasks [31]. In contrast, metric-based few-shot learning algorithms do not require complex auxiliary tasks and are trained by decomposing the dataset into different meta-tasks to adapt to new tasks unseen in training. Taking this strategy, metric learning learns how to cope with new small-sample problems by simulating and solving them in each training iteration [32–34]. For example, the prototypical network constructs a prototype for each class and employs nearest neighbor clustering to ensure that each sample in the feature space is close to its class prototype. It has been widely used in hyperspectral image classification tasks [35,36]. Metric learning has the advantages of

a simple network structure, no need for complex hyperparameter optimization, and the ability to leverage prior knowledge to guide new tasks.

To address the high-dimensional feature redundancy of hyperspectral data and the sample dependence of deep learning shrub identification models, a US-BS-P-Net framework is proposed by combining a novel unsupervised band selection algorithm with metric learning. By mining the spectral feature expression of shrub patches and extracting sensitive bands for shrub identification, the US-BS-P-Net leverages the advantages of metric learning in dealing with small-sample problems, constructing a high-precision shrub identification model in small-sample scenarios, and achieving accurate identification and mapping of shrub patches in SGs. The specific contributions are summarized as follows.

1. A novel band selection algorithm was constructed, i.e., US-BS-Net, which introduces unsupervised contrastive learning into BS-Net-Conv to help learn the discriminative feature representation between spectral bands of different ground objects and optimize the selection of sensitive bands.
2. A shrub identification framework in SGs was proposed, US-BS-P-Net, which combines the above-mentioned US-BS-Net and the metric-based prototypical network to achieve a high-precision shrub identification model in small-sample scenarios.

2. Materials and Methods

2.1. Study Area and Data Preprocessing

2.1.1. Study Area and Quadrat Survey

The study area is located in the northeast of Xilinhot City, Inner Mongolia, China (44°50′~44°51′N, 116°21′~116°22′E), which is a typical SG region of *Caragana microphylla* Lam (Figure 1). The northern region of Xilinhot City is dominated by gentle plains, with an altitude ranging from 939 to 952 m. Affected by the semi-arid continental climate in the mid-temperate zone, the study area is cold and dry in winter and warm and humid in summer, with an average annual precipitation of 309 mm, an average annual frost-free period of 144 days, and an average annual temperature of 0–3 °C. The growing season of vegetation usually lasts approximately 150 days per year from April to September. The dominant grass species include *Stipa krylovii* and *Leymus chinensis*. *Caragana microphylla* Lam is the only shrub species naturally growing in the study area.

A total of 40 quadrats were surveyed on 8 August 2024, including 17 grass quadrats, with 1 m^2 and 23 shrub quadrats (Figure 1b). To match the quadrats of *Caragana microphylla* Lam with unmanned aerial vehicle remotely sensed images, a real-time kinematic (RTK) device was used to conduct GNSS positioning on each shrub. As shown in Table 1, *Caragana microphylla* Lam in the study area has a mean height of 35 cm, a mean crown area of 212 × 162 cm^2, and a mean basal area of 117 × 91 cm^2. It has the characteristics of low growth, irregular and scattered distribution, which poses a challenge for accurately extracting shrubs in SGs.

Table 1. Quadrat survey data of *Caragana microphylla* Lam.

No.	Number of Tillers	Crown Area (cm^2)	Basal Area (cm^2)	Height (cm)
1	17	167 × 137	39 × 21	45
2	77	256 × 229	178 × 156	47
3	87	275 × 193	206 × 154	18
4	111	245 × 178	158 × 150	33
5	32	186 × 138	62 × 77	48
6	39	194 × 143	69 × 52	54

Table 1. *Cont.*

No.	Number of Tillers	Crown Area (cm^2)	Basal Area (cm^2)	Height (cm)
7	37	189 × 152	75 × 76	28
8	56	202 × 144	113 × 61	31
9	39	147 × 98	99 × 54	36
10	88	259 × 212	168 × 151	40
11	47	324 × 266	123 × 123	58
12	114	288 × 162	192 × 79	57
13	35	141 × 115	39 × 50	18
14	64	160 × 138	95 × 69	40
15	83	207 × 152	104 × 82	25
16	41	176 × 167	83 × 85	46
17	67	214 × 132	128 × 57	23
18	56	171 × 147	101 × 62	30
19	31	191 × 132	92 × 73	18
20	75	193 × 158	136 × 119	26
21	90	222 × 163	134 × 112	13
22	79	227 × 183	182 × 112	24

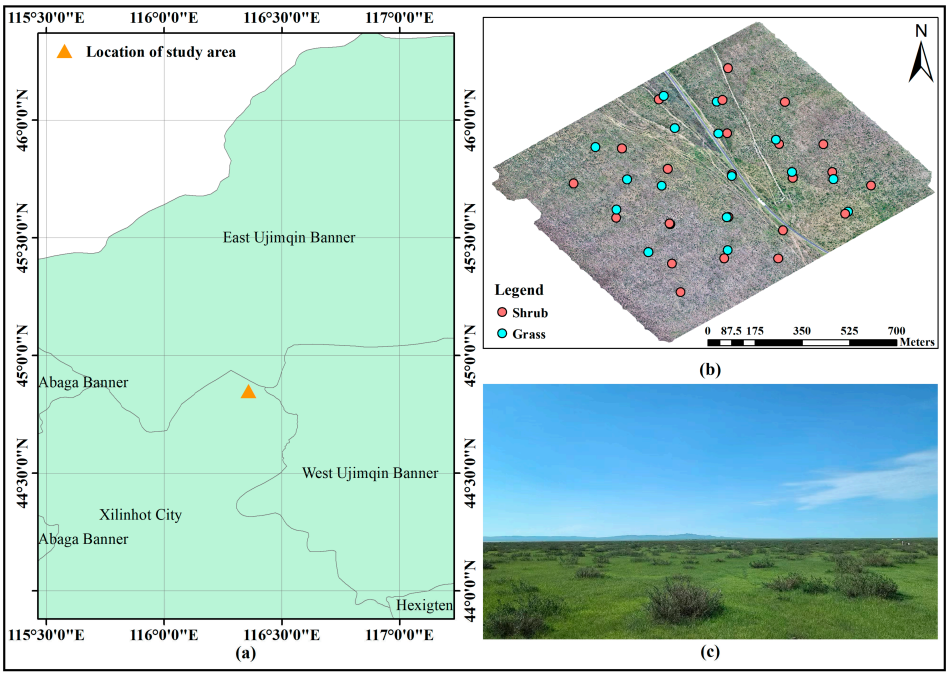


Figure 1. Overview of study area and spatial distribution of quadrats. (**a**) Location of study area; (**b**) digital orthophoto map derived from UAV LiDAR and location of quadrats; (**c**) shrub-encroached grassland landscape of *Caragana microphylla* Lam.

2.1.2. Hyperspectral Data Acquisition and Preprocessing

The UAV hyperspectral imagery was acquired using a DJI M300 Pro drone equipped with a ULTRIS X20 Plus imager produced by Cubert (Ulm, Germany). The spectral range is 350–1002 nm, including 164 bands, with a spectral resolution of 10 nm and a spectral sampling interval of 4 nm. Hyperspectral imagery acquisition and quadrat survey coincided with the vigorous growth period of *Caragana microphylla* Lam, allowing for effective differentiation between shrubs and non-shrubs. The drone flight mission took place between 11:00 AM and 1:00 PM in clear, cloudless weather with a light breeze. Pre-flight calibration was performed with a whiteboard, and the flight altitude was set to 100 m and the speed to 5 m/s. Forward and side overlaps were set to 80%, and the spatial resolution was 5.8 cm. The UAV hyperspectral imagery was stitched and radiometrically corrected to extract the

spectral reflectance of each band. To eliminate interference from light and atmosphere, a quick atmospheric correction method was applied to the hyperspectral imagery, ensuring it reflects the true reflectance of the ground materials. Furthermore, the Savitzky–Golay filter was used to process the hyperspectral data to smooth the noise. Finally, taking the ground control points obtained by RTK as a reference, the georeferencing function of ArcMap 10.8 software was adopted to align the hyperspectral image with the digital orthophoto map (DOM) derived from UAV LiDAR, with a total backward root mean square error < 1 pixel.

2.1.3. Dataset Construction

The shrub identification dataset, covering a sub-region with 340 × 230 m^2 (5811 × 4014 pixels) of the study area, consists of UAV hyperspectral images and the vector corresponding to their labeled pixels (Figure 2). Taking hyperspectral data as the base map, assisted by GNSS positioning data and DOM images, the pixels are labeled by manual visual interpretation. To avoid the influence of spatial autocorrelation of the samples on model validation, the shrub patches were labeled as much as possible throughout the area. Due to the enormous number of non-bush pixels, only 9 typical sub-regions of the hyperspectral images were selected for subtle vectorization, which meets the principles of uniform distribution and sample balance of different classes. First, the vectorized result is converted into a raster with the same resolution as the hyperspectral data. Then, taking the raster image as a mask, a sample was constructed by cropping a spatial–spectral cube with a fixed patch size, such as 21 × 21 × 164, on the hyperspectral image. Each sample adopts the label corresponding to its center pixel as the class. Finally, the shrub identification dataset is constructed from all samples with semantic information. The dataset of this study has a total of 8,982,979 samples, of which 4,407,170 are shrubs, and 4,575,809 are non-shrubs (Table 2). A ratio of 0.01:0.99:99 is used to divide the training set, validation set, and test set to simulate the shrub identification task under small-sample scenarios. Among them, the training set is used to construct the shrub identification models, which will be evaluated for accuracy in the validation set, while the test set will be used for the visualization analysis of the model's prediction results in the entire region.

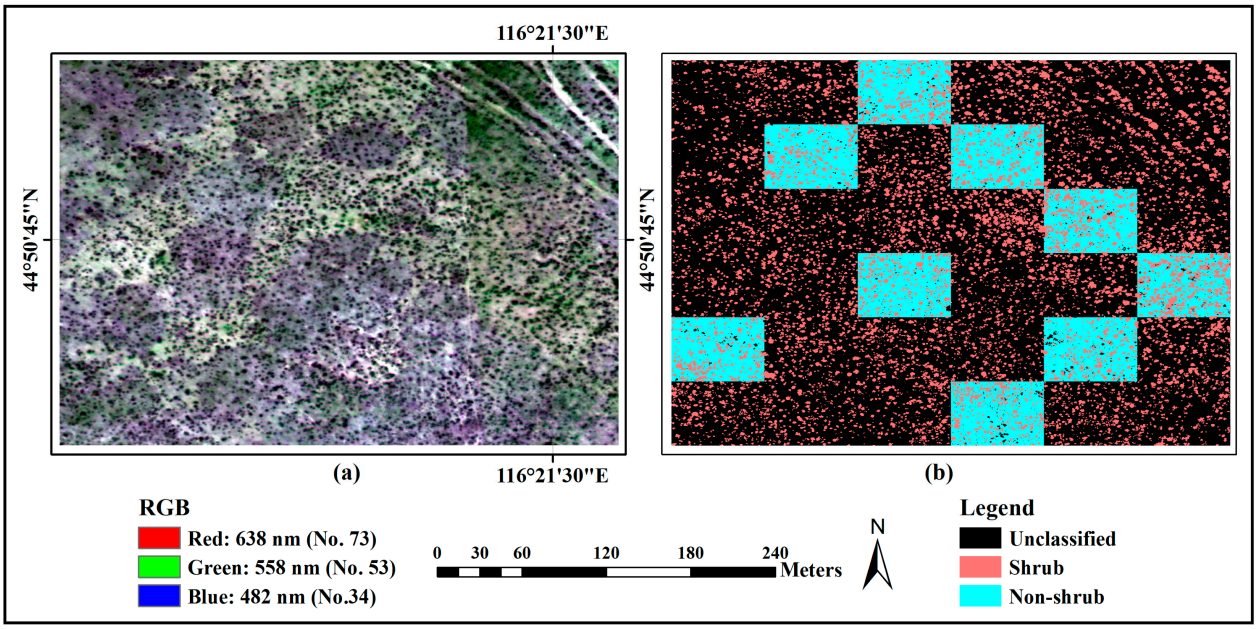


Figure 2. Description of the shrub identification dataset. The dataset includes UAV hyperspectral images (color synthesis, Red: 638 nm, Green: 558 nm, Blue: 482 nm) (**a**) and the vector corresponding to their labeled pixels (**b**).

Table 2. Sample division of the shrub identification dataset.

Class Name	Training Set	Validation Set	Test Set	Total
Ratio (%)	0.01	0.99	99.00	100.00
Shrub	441	43,630	4,363,099	4,407,170
Non-shrub	457	45,301	4,530,051	4,575,809
Total	898	88,931	8,893,150	8,982,979

2.2. Overview of the US-BS-P-Net

In this work, a novel US-BS-P-Net framework is proposed for shrub identification in SGs by combining an unsupervised deep learning band selection algorithm with a few-shot learning method. As shown in Figure 3, an unsupervised band selection algorithm, which includes a spectral attention module, L1 regularization, an auto-encoder, and contrastive learning, is constructed to extract informative spectral bands from raw UAV hyperspectral images. By introducing contrastive learning into BS-Net-Conv, the band selection model can not only learn the global interactions and nonlinear dependencies between spectral bands, but also extract discriminative feature representations that are beneficial to downstream shrub identification tasks. Then, the weight of each band is sorted to obtain the top-k sensitive bands as input for subsequent classifier construction. Finally, the prototypical network, a classic metric-based few-shot learning method, is adopted to build a shrub identification model, thereby achieving accurate extraction of shrubs in SGs under small-sample scenarios. For a detailed description of the proposed algorithm, please continue reading Sections 2.3 and 2.4.

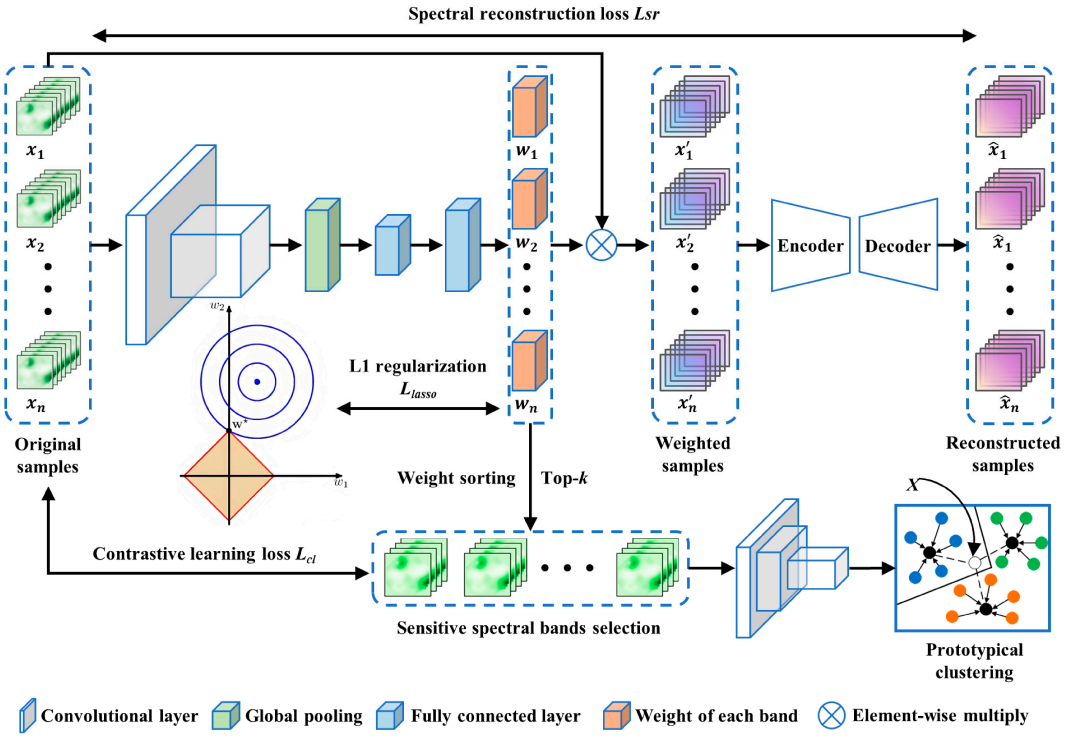


Figure 3. Overview of the US-BS-P-Net framework. US-BS-P-Net consists of an unsupervised deep learning band selection algorithm and a few-shot learning method.

Spectral attention module: The spectral attention module consists of a two-dimensional convolutional neural network, a global pooling layer, and a multi-layer perceptron network, which is used to extract informative spectral bands from the complete hyperspectral image cube. Taking a 164 × 5 × 5 image cube (with 164 bands and a spatial patch size of 5 × 5)

as input, the specific network structure is consistent with the literature [27], as shown in Table 3.

Table 3. The network structure of the spectral attention module.

Layer (Type)	Input Shape	Output Shape
BatchNorm2d	[164, 5, 5]	[164, 5, 5]
Conv2d	[164, 5, 5]	[64, 3, 3]
BatchNorm2d	[64, 3, 3]	[64, 3, 3]
ReLU	[64, 3, 3]	[64, 3, 3]
AdaptiveAvgPool2d	[64, 3, 3]	[64, 1, 1]
Flatten	[64, 1, 1]	[64]
Linear	[64]	[128]
ReLU	[128]	[128]
Linear	[128]	[164]
Sigmoid	[164]	[164]

L1 regularization: To make the band weights as sparse as possible to simulate the reconstruction and contrast tasks from a few key spectral information to global spectral bands, L1 regularization is used to constrain the band-weight output by the spectral attention module.

Auto-encoder: The basic architecture of the band selection algorithm proposed in this study is a convolutional autoencoder, whose goal is to extract the spatial–spectral information of the sample and restore the reweighted spectral bands to the original counterpart. The auto-encoder contains an encoder and a decoder, which are used to extract deep features and up-sample feature maps to the shape of the original sample. Assuming that the input size of samples is $164 \times 5 \times 5$, the specific network structure of the auto-encoder is shown in Table 4.

Table 4. The network structure of the auto-encoder.

Layer (Type)	Input Shape	Output Shape
Conv2d	[164, 5, 5]	[128, 3, 3]
BatchNorm2d	[128, 3, 3]	[128, 3, 3]
ReLU	[128, 3, 3]	[128, 3, 3]
Conv2d	[128, 3, 3]	[64, 1, 1]
BatchNorm2d	[64, 1, 1]	[64, 1, 1]
ReLU	[64, 1, 1]	[64, 1, 1]
ConvTranspose2d	[64, 1, 1]	[64, 3, 3]
BatchNorm2d	[64, 3, 3]	[64, 3, 3]
ReLU	[64, 3, 3]	[64, 3, 3]
ConvTranspose2d	[64, 3, 3]	[128, 5, 5]
BatchNorm2d	[128, 5, 5]	[128, 5, 5]
ReLU	[128, 5, 5]	[128, 5, 5]
Conv2d	[128, 5, 5]	[164, 5, 5]
Sigmoid	[164, 5, 5]	[164, 5, 5]

Contrastive learning: By introducing unsupervised contrastive learning in the spectral reconstruction process, the reweighted sample is similar to its corresponding original one and far away from others, thereby improving the ability of the model to learn discriminative feature representations. Please refer to Section 2.3 for details.

Prototypical network: It is a simple but effective metric-based few-shot learning method that constructs high-precision shrub identification models by learning the prototype of each class and clustering samples towards its prototype in deep feature space. The specific algorithm will be introduced in Section 2.4.

2.3. US-BS-Net for Spectral Band Selection

How to reduce the dimensionality of hyperspectral data with hundreds of bands and extract sensitive spectral bands for specific tasks is a key topic in the field of UAV hyperspectral image applications. To improve the ability of the unsupervised band selection model to learn discriminative spectral feature representations, a classic unsupervised contrastive learning algorithm, named SimCLR [37], is introduced into BS-Net-Conv to construct US-BS-Net in this part, thereby achieving discriminative spectral band extraction. Next, this paper will describe the components of the proposed US-BS-Net in detail.

Given a hyperspectral data cube $x \in R^{c \times p \times p}$ (c represents the number of bands and p is the spatial patch size of the sample) as input, the spectral attention module constructs a non-negative band-weight tensor $w \in R^{c \times 1 \times 1}$ (Equation (1)) by learning the global interaction and nonlinear dependency between spectral bands.

$$w = \sigma(g(x; \theta_b)), \tag{1}$$

where $g(\cdot)$ denotes the function of network weights, θ_b is the trainable parameters of the spectral attention module, and $\sigma(\cdot)$ represents a sigmoid function, which is used to map the weights of spectral bands to the range of 0–1.

Subsequently, according to the amount of spectral information contained in each band, a band-wise multiplication operator is adopted to assign weights to them, thereby constructing a reweighted sample $x' \in R^{c \times p \times p}$ can be defined by Formula (2).

$$x' = x \otimes w, \tag{2}$$

Here $\otimes$ denotes a band-wise multiplication operator.

To sparsify the above-mentioned weights tensor and improve the interpretability of the band selection model in extracting sensitive bands, L1 regularization is imposed as a constraint on band-weight learning. Assuming there is a batch of samples (set $X = \{x_i, i \in 1 \ldots n\}$ where n is the number of samples), the L1 regularization loss L_{lasso} can be defined as Formula (3).

$$L_{lasso} = \alpha \sum_{i=1}^{n} |w_i|, \tag{3}$$

where α denotes a regularization coefficient, which is used to trade off the L1 regularization loss with others, and $|\bullet|$ represents an absolute value operator.

Next, a convolutional auto-encoder network is employed to recover the original spectral bands from the reweighted ones so that the reconstructed sample $\hat{x} \in R^{c \times p \times p}$ is as close as possible to the original sample $x \in R^{c \times p \times p}$. Its objective function is the mean square error, which is used to evaluate the spectral reconstruction performance of the model. The spectral reconstruction loss L_{sr} is shown in Formula (4).

$$L_{sr} = \frac{1}{n} \sum_{i=1}^{n} \| f_{ae}(x_i) - f_{ae}(\hat{x}_i) \|_2^2, \tag{4}$$

where n is the number of samples in a training batch; $f_{ae}(\cdot)$ represents the embedding function of the auto-encoder network.

Different from BS-Net-Conv, a SimCLR algorithm is introduced to construct a contrastive learning loss L_{cl} in the band selection process to obtain discriminative spectral bands. This study randomly selects n original samples for a training task; then, there will be n reweighted samples constructed by the spectral attention module in the proposed US-BS-Net. Both together form a sample set $X = \{x_i, i \in 1 \ldots 2n\}$. Therefore, for original

sample x_i, its reweighted sample x_j is a positive sample, and the others are negative samples. The L_{cl} can be calculated according to Formula (5).

$$L_{cl} = -\log \frac{\exp(f(x_i) \bullet f(x_j)/\tau)}{\sum_{k=1}^{2n} 1_{[k \neq i]} \exp(f(x_i) \bullet f(x_k)/\tau)}, \quad (5)$$

where $f(\bullet)$ is the embedding function of the encoder network; τ represents a temperature coefficient; and $\bullet$ denotes an inner-product operator.

Finally, the overall objective function L_{bs} of US-BS-Net is composed of L1 regularization loss, spectral reconstruction loss, and contrastive learning loss, as shown in Formula (6). By minimizing it, the network parameters of US-BS-Net can be updated with gradients, and the discriminative spectral band can be obtained.

$$L_{bs} = L_{sr} + L_{lasso} + L_{cl}, \quad (6)$$

2.4. Metric-Based Prototypical Network for Shrub Identification

Prototypical network, a classic metric-based few-shot learning method, has been widely used in hyperspectral image classification applications in recent years due to its simple network structure, fast fitting, and excellent performance [35,38]. Although the research on shrub identification is still lacking, its potential for high-precision extraction of shrubs in SGs cannot be ignored.

Taking the top k bands ranked by spectral weights in Section 2.3 as input, the prototypical network is adopted to construct shrub identification models. The core of the prototypical network is to learn and build the cluster center of each type, namely the prototype. Different from the general classification network, the prototypical network adopts episodic sampling to obtain a batch of training samples. For example, C-way-K-shot episodic sampling means that C categories are randomly selected, where each category has N query samples and K support samples. For the c-th category, assuming that the set of support samples is $S_c = \{x_i, i \in 1 \ldots C \times K\}$, the prototype P_c can be calculated by averaging the deep features of the support samples. The formula is as follows:

$$P_c = \frac{1}{K} \sum_{x_i \in S_c} f(x_i), \quad (7)$$

where $f(\bullet)$ is the embedding function of the prototypical network.

The query samples are used to construct the loss function of the prototypical network and validate the model. Specifically, given a query sample x_j, the prototypical network calculates its loss L_{proto} using Euclidean distance and a softmax followed by a logarithmic function, as is shown in Formula (8). Using this mechanism, query samples are clustered towards their prototypes by minimizing L_{proto}.

$$L_{proto} = -\log \frac{\exp(-d(Enc(x_j), P_{x_j}))}{\sum_{c=1}^{C} \exp(-d(Enc(x_j), P_c))}, \quad (8)$$

where P_{x_j} is the prototype of the class to which x_j belongs, and $d(\bullet)$ is the squared Euclidean distance function.

2.5. Other Comparison Methods

The detailed description of each comparison method is as follows.

2D-CNN: It represents a common image identification algorithm based on a two-dimensional convolutional neural network (2D-CNN), which consists of six convolutional

layers. Three convolutional blocks are formed by combining every two convolutional layers, with filters of 32, 64, and 128, respectively [39].

Matching network: The matching network is a classic few-shot learning method, which constructs a classifier by measuring the similarity between query samples and support samples using the cosine function in the embedding space [40].

Prototypical network: It is the basic identification framework of US-BS-Net proposed in this study. The specific algorithm description can be found in Section 2.4.

Relation network: Its basic framework is similar to the prototypical network, but the Euclidean distance is replaced by a two-layer CNN in the distance measurement part [41].

US-BS-P-Net: The US-BS-P-Net framework is constructed by combining the proposed US-BS-Net band selection algorithm with the prototypical network.

2.6. Experimental Setting

Overall accuracy (OA, Equation (9)), average accuracy (AA, Equation (11)), and kappa coefficient (Kappa × 100, Equation (12)) are adopted to conduct spatial consistency validation of the shrub identification models, and precision (PA, Equation (10)) is used to measure the performance of shrubs and non-shrubs separately. The main hyperparameters are set according to the literature [27,35] as shown in Table 5.

$$OA = \frac{\sum_{i=1}^{n} X_{ii}}{M}, \tag{9}$$

$$PA = \frac{X_{ii}}{X_{+i}}, \tag{10}$$

$$AA = \frac{\sum_{i=1}^{n} PA}{n}, \tag{11}$$

$$Kappa = \frac{M\sum_{i=1}^{n} X_{ii} - \sum_{i=1}^{n} X_{i+} \times X_{+i}}{M^2 - \sum_{i=1}^{n} X_{i+} \times X_{+i}}, \tag{12}$$

where X_{ii}, X_{+i}, and X_{i+} denote the number of correctly classified samples, reference samples, and total samples (including correctly classified samples and misclassified samples) for the i-th category ($i \in \{1...n\}$, n is the number of categories) in the error matrix, respectively, and M represents the number of samples in the dataset.

Table 5. Description of the main hyperparameter settings.

Name	Notation	Description
Learning rate	-	Each model uses a learning rate of 5 × 10−4, which is halved after every 2000 training iterations.
Temperature	$\tau \in (0, 1]$	The temperature of unsupervised contrastive learning is configured as 0.1.
C-way-*K*-shot	-	An episodic sampling of 2-way-5-shot is adopted, in which both shrubs and non-shrubs are randomly sampled in each training step, and the support set of each class contains five samples.
Alpha	$\alpha \in (0, 1]$	In the band selection, the trade-off parameter between L1 regularization loss and others is set to 0.01.

All algorithms are programmed in PyCharm 2024.3 using Python 3.11.7 and implemented with the PyTorch 2.5.1 deep learning framework. The experiments involved in this study are conducted on an operating system with the hardware configuration of Intel(R) Xeon(R) Silver 4110 CPU @ 2.10 GHz and NVIDIA GeForce RTX 2080Ti GPU.

3. Results

3.1. Performance of Prototypical Network on Samples with Different Patch Sizes

Fully utilizing the spatial information of hyperspectral images, such as texture, shape, neighborhood environment, and complementing the spectral features, is crucial for identifying naturally growing shrubs in this study. While retaining all 164 spectral bands, the prototypical network is applied to samples of different patch sizes ranging from 5 × 5 to 31 × 31, representing a spatial range of 0.29–1.55 m in the field, which is sufficient to identify shrubs at different growth stages in the study area. Then, the optimal patch size is obtained by calculating and comparing the performance of shrub identification for each spatial range. As shown in Figure 4, it can be found that the OA of shrub identification presents an upward trend from 5 × 5 to 21 × 21, and then decreases from 21 × 21 to 31 × 31. The performance of the prototypical network continuously improves until the patch size of 21 × 21, indicating that the joint effect of spatial–spectral features has a positive response to shrub identification. Simultaneously, the reduction trend in the amplitude of improvement during the above process reveals the peak value of 21 × 21 for the shrub identification task. At this point, the rich spatial–spectral features in the hyperspectral image are balanced with the feature extraction ability of the prototypical network. Subsequently, spatial information redundancy emerges as the patch size further expands, causing a decrease in identification performance. Although the shrub identification accuracy of 21 × 21 is only slightly better than that of 23 × 23, this study still chooses 21 × 21 as the optimal patch size for this study, considering its highest performance and the need to minimize the redundancy of spatial information as much as possible. Therefore, this study will train US-BS-P-Net with a fixed spatial size of 21 × 21 in subsequent experiments.

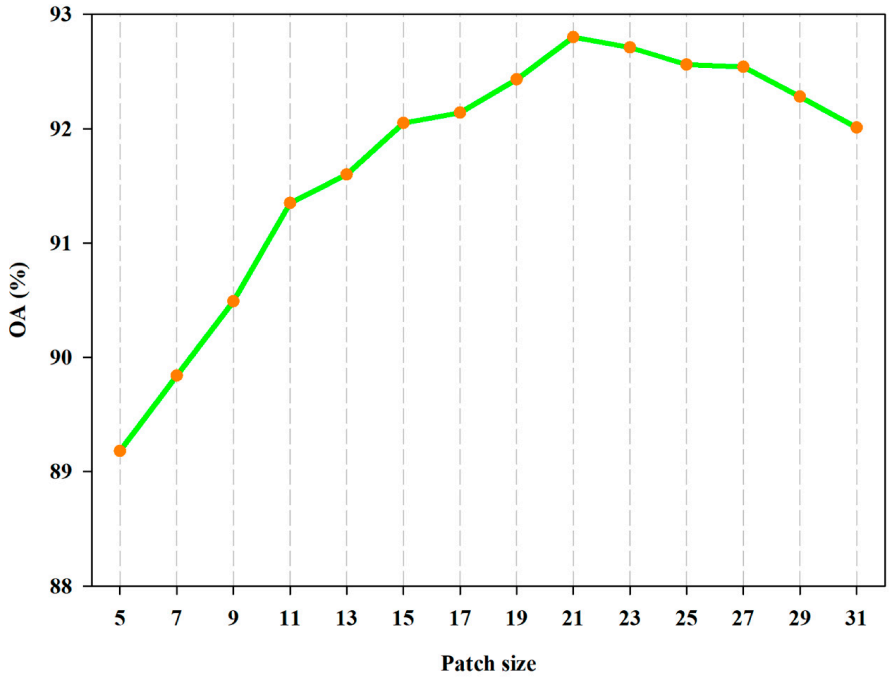


Figure 4. Overall accuracy of the prototypical network on samples with different patch sizes.

3.2. The Impact of Sensitive Spectral Bands Quantity on the Identification Accuracy

Effective data dimensionality reduction methods are the key to eliminating the spectral information redundancy of hyperspectral data with hundreds of dimensions. To verify the effectiveness of the sensitive spectral bands extracted by the proposed US-BS-P-Net for accurate identification of shrubs, an experimental analysis is conducted. Specifically, the spectral weights are sorted, and the top k bands of samples are selected for model training (k = 5–160, with an interval of 10 bands except for 5–10); the results are shown in Figure 5. It can be seen that using only a small amount of spectral information, for example, 5 bands, the OA of shrub identification is better than 92%. It demonstrates that high-precision shrub extraction in SGs can be achieved using only a few informative spectral bands of hyperspectral images, which is instructive for the application of multispectral sensors in shrub identification. As the number of sensitive spectral bands increases, the performance of the model also improves and gradually approaches that of the model trained on all 164 bands (baseline), and even slightly surpass when the quantity of bands reaches 50. The spectral band selection algorithm proposed in this study enables effective spectrum information compression, and its combination with clustering methods has a positive effect on task-specific band optimization. Afterwards, the shrub identification performance fluctuated around the baseline, and when the number of bands is 80, the OA reaches a peak value > 93%, exceeding that of 50. On the one hand, the excellent performance of the prototypical network without data dimensionality reduction highlights the effectiveness of the proposed algorithm in shrub identification. On the other hand, the practice of training high-precision shrub identification models based on a few spectral representations obtained by US-BS-P-Net shows the necessity of data dimensionality reduction.

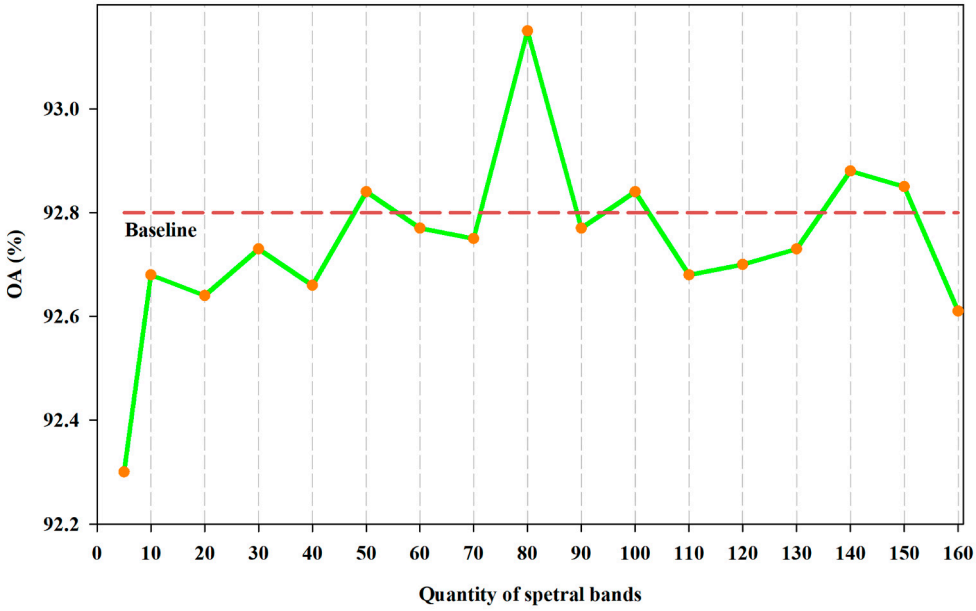


Figure 5. Overall accuracy of US-BS-P-Net with different quantities of sensitive spectral bands.

The band numbers for different quantities (5–80) of sensitive spectral bands selected by US-BS-P-Net are listed in Table 6. In addition, the spectral curves of shrub and non-shrub samples are plotted to visually demonstrate their spectral distinguishability (Figure 6). Since the non-shrub patches are mainly grasslands, which contain a large amount of herbaceous vegetation, their average spectral curves are similar to those of shrubs. From the perspective of spectral reflectance, there is a significant overlap in the reflectance distribution between shrubs and non-shrubs, but distinguishability is still observed in wavelength ranges of about 426–906 nm and 946–1002 nm, especially in 506–706 nm. It is

consistent with the sensitive band extraction results in Table 6, where the top 10 bands are all in 506–706 nm, and most of the top 80 bands are also within this range.

The selected spectral bands correspond to key physiological and structural features of vegetation spectral response. The 506–706 nm range encompasses the green reflectance peak (approximately 550 nm) and the onset of the red-edge region (approximately 680–706 nm). The former is associated with enhanced green light reflectance driven by chlorophyll content, while the latter is highly sensitive to chlorophyll density and canopy structure, serving as a crucial indicator for distinguishing between different vegetation types or varying growth stages of the same vegetation. The 426–906 nm range also includes the strong chlorophyll absorption bands in the blue (~480 nm) and red (~660 nm) regions, as well as the high-reflectance plateau in the near-infrared (750–900 nm), effectively reflecting vegetation biomass and cover. Furthermore, the 946–1002 nm region lies near a weak water absorption feature (approximately 970 nm) and responds to leaf water content, thereby enhancing the ability to differentiate between shrub and herbaceous vegetation.

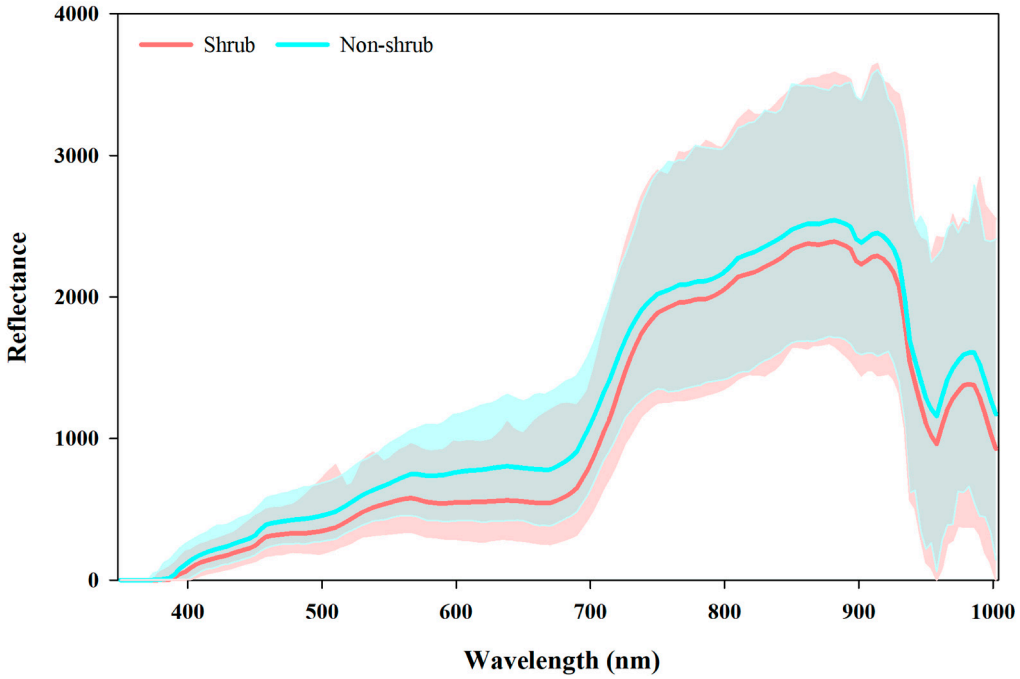


Figure 6. Spectral reflectance curves of shrub and non-shrub samples.

Table 6. The band numbers for different quantities of sensitive spectral bands.

Number of Bands	Wavelength (nm)
5	590, 530, 694, 510, 670
10	590, 530, 694, 510, 670, 646, 662, 514, 542, 626
20	590, 510, 530, 694, 670, 686, 514, 490, 622, 542, 646, 406, 626, 662, 658, 610, 698, 618, 478, 634
30	590, 510, 530, 670, 694, 686, 514, 490, 622, 542, 658, 646, 406, 626, 662, 610, 598, 618, 698, 422, 478, 394, 634, 398, 414, 562, 674, 482, 546, 470
40	590, 510, 530, 694, 670, 686, 514, 622, 490, 646, 542, 626, 406, 662, 698, 658, 610, 634, 618, 562, 478, 482, 398, 674, 414, 422, 394, 978, 458, 598, 546, 522, 446, 426, 502, 734, 558, 470, 430, 474

Table 6. *Cont.*

Number of Bands	Wavelength (nm)
50	590, 510, 530, 670, 694, 686, 514, 490, 622, 542, 646, 406, 626, 658, 662, 610, 698, 618, 598, 478, 422, 634, 398, 394, 414, 482, 562, 674, 546, 502, 470, 522, 458, 446, 978, 638, 706, 426, 682, 518, 614, 474, 558, 430, 594, 734, 450, 710, 606, 578
60	590, 510, 530, 670, 694, 686, 514, 490, 622, 542, 646, 406, 626, 658, 662, 610, 698, 618, 598, 478, 422, 634, 398, 394, 414, 482, 562, 674, 546, 502, 470, 522, 458, 446, 978, 638, 706, 426, 682, 518, 614, 474, 558, 430, 594, 734, 450, 710, 606, 578, 714, 650, 498, 494, 870, 586, 574, 538, 582, 770
70	590, 510, 530, 670, 694, 686, 514, 490, 622, 542, 658, 646, 406, 626, 662, 610, 598, 618, 698, 422, 478, 394, 634, 398, 414, 562, 674, 482, 546, 470, 502, 522, 446, 458, 638, 706, 978, 682, 614, 426, 518, 474, 594, 430, 558, 606, 710, 578, 450, 650, 714, 734, 494, 498, 586, 574, 870, 582, 538, 642, 678, 554, 770, 462, 506, 410, 722, 526, 570, 702
80	590, 670, 686, 694, 530, 510, 514, 490, 622, 658, 610, 542, 598, 646, 626, 406, 662, 422, 618, 698, 478, 394, 634, 398, 674, 562, 546, 470, 414, 482, 502, 638, 706, 446, 522, 614, 682, 458, 978, 426, 594, 518, 474, 558, 710, 430, 578, 606, 714, 650, 582, 494, 450, 574, 586, 498, 734, 538, 642, 870, 678, 554, 462, 722, 506, 526, 666, 654, 410, 770, 966, 570, 702, 718, 442, 566, 366, 350, 438, 602

3.3. Accuracy and Identification Map Production for the Optimal Model

Table 7 lists the identification results of the optimal model (US-BS-P-Net using only 80 bands) and its baseline (full bands). It demonstrates that the optimal model outperforms the full-band model in terms of OA, AA, and Kappa × 100, with improvements of 0.35%, 0.38%, and 0.72, respectively, thereby achieving effective dimensionality reduction while enhancing overall identification performance. Specifically, the identification accuracy for non-shrub improved by 1.11%, which is the primary driver of the gain in OA. Meanwhile, the identification accuracy for shrubs declined only slightly by 0.34%, which is an acceptable decrease in practical applications. More importantly, reducing the number of bands from full to 80 lowers computational and storage costs and may also alleviate the Hughes phenomenon in high-dimensional data, thus improving model generalization.

Table 7. Comparison of identification results for the optimal model (band number = 80) and its corresponding baseline (full bands).

Method	Baseline (Full Bands)	US-BS-P-Net (Band Number = 80)
OA (%)	92.80	**93.15**
Kappa × 100	85.59	**86.31**
AA (%)	92.80	**93.18**
Shrub (%)	**91.89**	91.55
Non-shrub (%)	93.70	**94.81**

The optimal model and its corresponding baseline model are adopted to perform pixel-wise prediction in the study area to generate shrub identification maps, as shown in Figure 7. Although the overall classification results are encouraging, confusion between shrub and herbaceous vegetation patches persists, resulting in indistinct boundaries for shrubs (e.g., the area within the black box in the top-right corner of Figure 7c,d). In comparison, the optimal model slightly alleviates this confusion due to its superior performance and effectively maps the distribution of shrubs generally in the study area. In addition, there are

some discrete points and isolated pixels in the identification map, which is primarily due to the fine-grained identification nature of the algorithms employed in this study, although model prediction errors may also contribute to this issue.

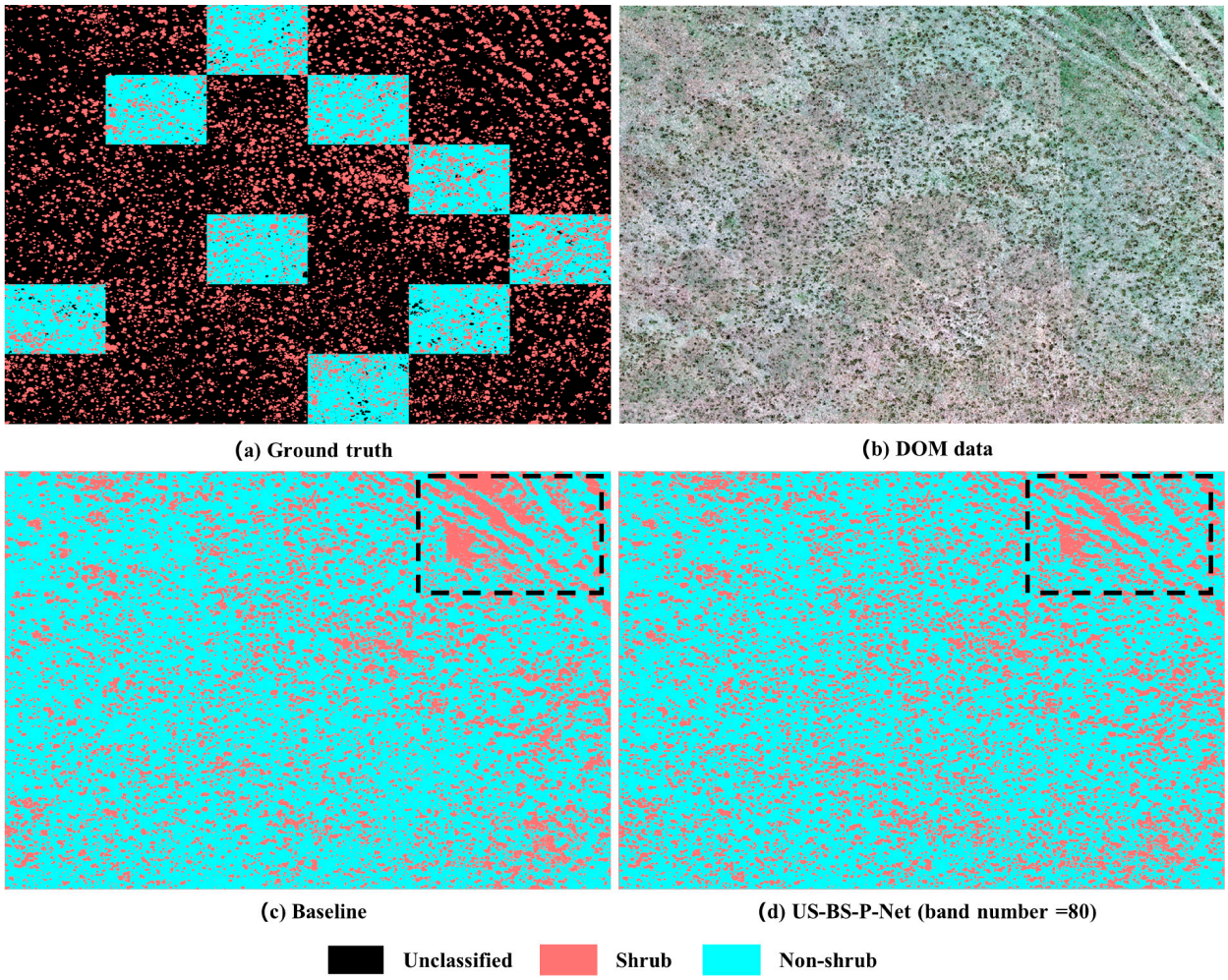


Figure 7. Identification maps generated by the optimal model (band number = 80) and its corresponding baseline (full bands). (**a**) Ground truth; (**b**) DOM image; (**c**) identification map generated by the optimal model; and (**d**) identification map generated by the baseline.

4. Discussion

4.1. Hyperparameter Sensitivity Analysis

To evaluate the hyperparameter sensitivity of the proposed US-BS-P-Net framework, a comprehensive parameter tuning experiment is conducted on the key parameters, i.e., temperature and alpha; the trends in OA are illustrated in Figure 8. Among them, temperature is a core hyperparameter in unsupervised contrastive learning, reflecting the difficulty of model optimization in the training task. In Figure 8a, the optimal model was obtained at a lower temperature (0.1), indicating that it tends to enhance sample discriminability to extract discriminative feature representations during the unsupervised band selection. Subsequently, the temperature is fixed at 0.1, and alpha is adjusted, which is the trade-off parameter between the L1 regularization loss and the others. In Figure 8a, the optimal alpha value is found to be 0.01, indicating that only a mild degree of band-wise sparsity regularization strength is required to obtain the optimal model. Overall, the fluctuation in OA of temperature parameter adjustments is greater than that of alpha, which means the US-BS-P-Net framework proposed in this study is more sensitive to temperature.

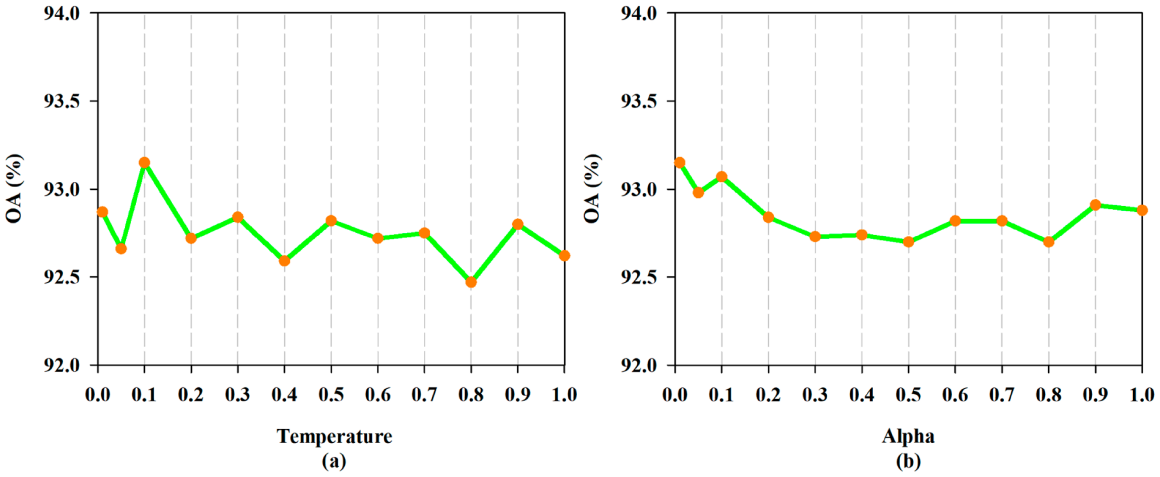


Figure 8. Trends in OA via the US-BS-P-Net model with different temperature (**a**) and alpha (**b**) on shrub identification.

4.2. Ablation Analysis

A comprehensive ablation study is conducted by detaching components such as L1 regularization, contrastive learning, and prototypical network separately to evaluate the contribution of each factor in the proposed US-BS-P-Net framework. It is worth noting that when all the aforementioned components are removed, a 2D-CNN is employed, serving as the baseline model for this section. In addition, the spectral attention module is excluded from the ablation analysis since it is an essential component of unsupervised band selection. Table 8 presents all ablation performance results, with the best accuracies highlighted in bold. It is evident that the addition of each component improves the model's identification performance. In US-BS-Net, the introduction of L1 regularization and contrastive learning achieved OA improvements of 3.28% and 2.45%, respectively. Furthermore, their combination achieves a higher accuracy gain (an OA increase of 4.21%) than either module alone. In contrast, the prototypical network makes the most significant contribution (an OA increase of 6.16%) to the proposed framework, specifically through its superior metric-based few-shot identification algorithm. By combining it with L1 regularization and contrastive learning to construct the US-BS-P-Net framework, achieving an optimal identification model with an OA improvement of 6.51% over the baseline.

Table 8. Ablation analysis of each component in the proposed US-BS-P-Net framework.

L1 Regularization	Contrastive Learning	Prototypical Network	OA (%)	Kappa × 100	AA (%)
✗	✗	✗	86.64	73.33	86.99
✓	✗	✗	89.92	79.85	89.92
✗	✓	✗	89.09	78.21	89.39
✓	✓	✗	90.85	81.68	91.03
✗	✗	✓	92.80	85.59	92.80
✓	✓	✓	**93.15**	**86.31**	**93.18**

4.3. Comparison of Various Dimensionality Reduction Methods

In this section, our proposed US-BS-Net is compared with principal component analysis (PCA) and BS-Net-Conv. Among them, PCA is a classic feature transformation-based dimensionality reduction method, whose core idea is to map high-dimensional data to a low-dimensional space while retaining the original information as much as possible. In contrast, BS-Net-Conv, a representative dimensionality reduction method based on

band selection, works by utilizing several informative spectral bands to reconstruct the complete spectral information. Obviously, the latter is more interpretable in extracting sensitive bands. The US-BS-Net proposed in this study is an improved version based on BS-Net-Conv, which mainly takes into account the lack of learning discriminative spectral features in the proxy task of BS-Net-Conv.

In Section 3.2, high-precision shrub identification in the study area can be achieved regardless of whether a small number of bands or sufficient spectral information is used. Therefore, this part only compares the performance of extracting five bands (a small number of bands) and 80 bands (sufficient spectral information) using various dimensionality reduction methods. The experimental results are shown in Table 9. When the spectral dimension is reduced to 5, it can be clearly seen that the deep learning-based band selection methods, i.e., US-BS-Net and BS-Net-Conv, perform better than PCA. This proves that the spectral bands extracted by the band selection method not only retain their physical meaning but also maintain strong sensitivity to specific tasks. PCA-based dimensionality reduction method shows that the OA of shrub identification decreased by more than 5% when the number of bands increased to 80. The reason for the decline in PCA performance can be attributed to the fact that the noise brought by the less informative bands seriously reduces the feature extraction efficiency of the identification model. Although the performance difference between US-BS-Net and BS-Net-Conv is negligible when there are a small number of bands, the former outperforms the latter when there are 80 bands, achieving the highest accuracy (OA of 93.15%, Kappa $\times$ 100 of 86.31, and AA of 93.18%). It demonstrates that discriminative feature extraction via contrastive learning is beneficial to the construction of high-precision shrub identification models. In addition, the number of trainable parameters and the training time per epoch for both algorithms are counted, finding that US-BS-Net only required an additional 8320 parameters and an extra 8 s/epoch compared to BS-Net-Conv, which indicates that the loss in computational efficiency is within an acceptable range.

Table 9. Identification results of various dimensionality reduction methods.

Method	**PCA**		**BS-Net-Conv**		**US-BS-Net**	
Number of Bands	**5**	**80**	**5**	**80**	**5**	**80**
OA (%)	90.78	85.51	**92.34**	92.56	92.30	**93.15**
Kappa $\times$ 100	81.57	70.95	**84.69**	85.13	84.59	**86.31**
AA (%)	90.78	85.86	**92.36**	92.58	92.29	**93.18**
Shrub (%)	89.94	89.00	90.93	91.02	**91.48**	**91.55**
Non-shrub (%)	91.62	82.73	**93.79**	94.15	93.10	**94.81**
Trainable parameters	-		520,016		528,336	
Training times (s/epoch)	-		22		31	

To intuitively demonstrate the difference between BS-Net-Conv and the proposed US-BS-Net in spectral band learning, the average spectral weight curve of the training samples is plotted, as shown in Figure 9. Obviously, there are significant differences in the weight distribution of spectral bands between US-BS-Net and BS-Net-Conv. From a spectral wavelength perspective, the top-ranked sensitive bands obtained by BS-Net-Conv are widely distributed in the near-ultraviolet, visible light, and near-infrared ranges. In contrast, wavelengths with weight > 0.5 based on US-BS-Net are mainly distributed in the visible light range, and only several bands are in the near-ultraviolet and near-infrared spectrum. This phenomenon indicates that the feature differences between the samples in this study are mainly reflected in the visible light spectrum range. Since the objective of this study is to identify shrubs in SGs, the classification system does not distinguish between

vegetation and non-vegetation in non-shrub types, which may cause the US-BS-Net model to be less sensitive to the near-infrared bands that characterize vegetation growth conditions such as cover and biomass.

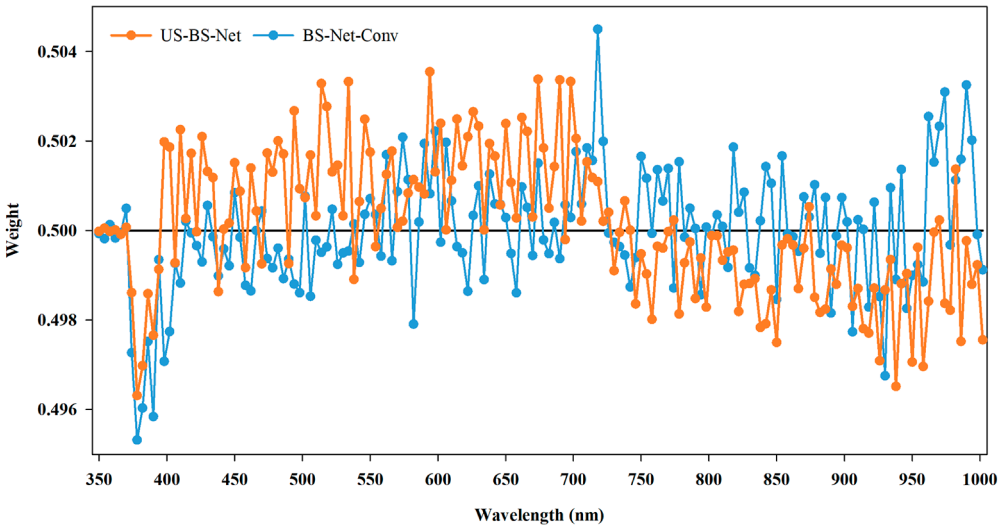


Figure 9. Average spectral weight curves of the training samples using US-BS-Net and BS-Net-Conv.

4.4. Comparison of Different Identification Models

As for identification models, this study also compares the proposed US-BS-P-Net framework with 2D-CNN, matching network, relation network, and prototypical network. It is worth noting that the number of optimal sensitive spectral bands extracted by US-BS-Net is 80, as illustrated in Section 3.2. As shown in Table 10, the traditional 2D-CNN algorithm has the lowest shrub identification accuracy in the study area, with an OA of only 86.64%. In contrast, the identification performance of the metric-based few-shot learning methods, i.e., matching network, relation network, and prototypical network, is significantly superior to that of the 2D-CNN. This proves the effectiveness of few-shot learning algorithms in identifying shrubs in SGs under small-sample scenarios. Among them, the OA of the prototypical network is slightly better than that of the matching network and the relational network. By comparing with others, the shrub identification performance of the proposed US-BS-P-Net is highlighted. The US-BS-P-Net constructed using the optimal sensitive bands (80 bands) as input achieved the highest OA of 93.15% (please refer to Table 7). This indicates that the combination of spectral band optimization and few-shot learning enables a positive effect on shrub identification in this study.

Table 10. Identification results of different deep learning models.

Method	2D-CNN	Matching Network	Relation Network	Prototypical Network	US-BS-P-Net
OA (%)	86.64	92.64	92.32	92.80	**93.15**
Kappa × 100	73.33	85.28	84.63	85.59	**86.31**
AA (%)	86.99	92.65	92.32	92.80	**93.18**
Shrub (%)	82.92	91.36	**92.43**	91.89	91.55
Non-shrub (%)	91.05	93.94	92.21	93.70	**94.81**

The models constructed by the above methods are adopted to perform pixel-wise prediction in the study area to generate shrub identification maps, as shown in Figure 10. In addition, to verify the robustness of the proposed algorithm in identifying shrubs in other areas, this study compares and analyzes these results with the test set (ground

truth) in several typical regions (Figure 11) by local zooming in. Referring to the high-resolution DOM data (the first row) and ground truth (the second row) of the dataset, it can be observed that the shrub identification map generated by the 2D-CNN obviously misidentifies non-shrubs as shrubs, resulting in an overestimation of the shrub cover in the study area. In contrast, the shrub identification models constructed by metric-based few-shot learning significantly improve the above phenomenon, but their results present some fragmented patches since the identification error still exists in the small-sample scenarios. In general, with the improvement in identification accuracy, the map of US-BS-P-Net is closer to the ground truth than that of others. However, limited by the number of training samples (898 in total, accounting for 0.01% of the total samples in the dataset), there is further room for improvement in the accuracy of shrub recognition and the elimination of fragmented patches and noise with a point or island shape.

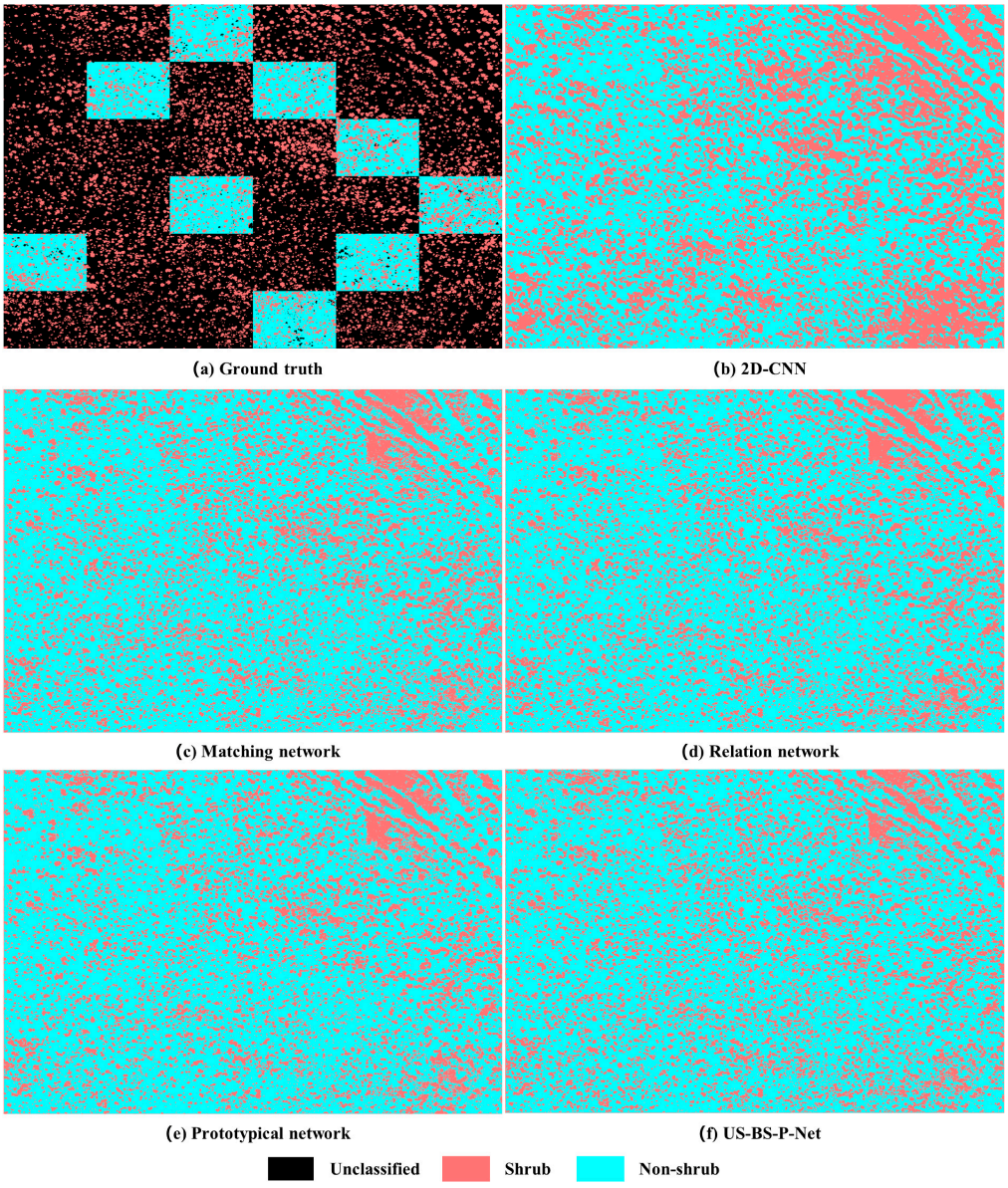


Figure 10. Identification maps generated by different models on the shrub identification dataset.

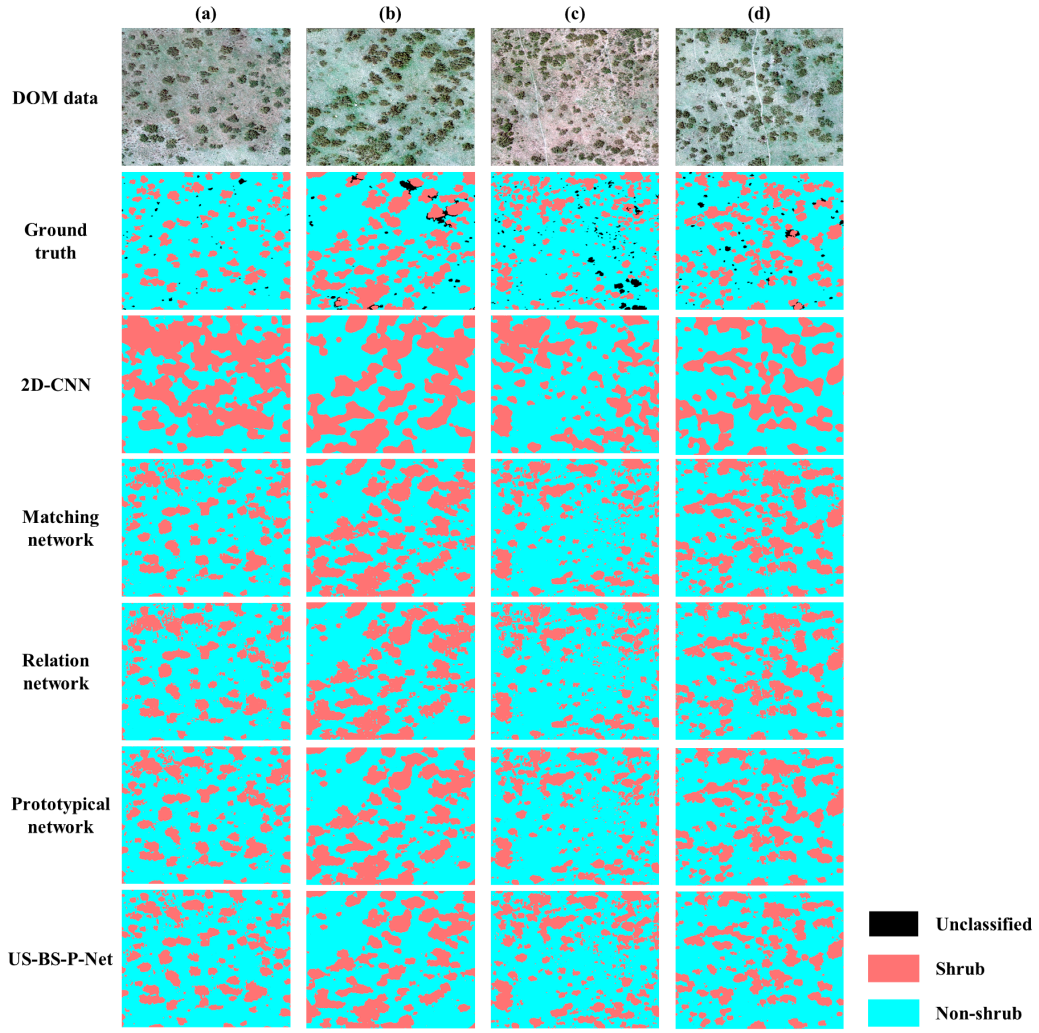


Figure 11. Identification maps using different models in several typical regions and their reference ground truth. (**a**–**d**) represent different sub-regions.

4.5. Limitations

The dataset and the US-BS-P-Net framework proposed in this study represent a preliminary exploration of deep learning algorithms for shrub identification in SGs. By combining unsupervised band selection with few-shot learning, the framework achieves high-precision shrub identification using hyperspectral imagery with limited training samples, thereby validating the effectiveness of this approach. However, we also clearly recognize that the current framework faces several key limitations. First, a single hyperspectral image can only obtain the texture and spectrum of the shrub canopy surface, but lacks vertical structural information, which limits shrub identification at the individual scale. Therefore, we plan to fuse hyperspectral and LiDAR data to construct high-precision individual shrub identification models by combining the biochemical information of hyperspectral data and the physical properties of LiDAR. Furthermore, this study only focuses on a single shrub-encroached grassland region. We expect that the novel algorithm proposed in this work can be verified on hyperspectral datasets from other study areas, and potentially even in fields such as crops, tree species, buildings, and medicine. Subsequently, vegetation spectra exhibit strong temporal dynamics, i.e., seasonal changes drive periodic fluctuations in leaf chlorophyll, water, and canopy structure, resulting in distinct spectral differences for the same species across different growing seasons. However, the shrub identification in this study relied on data from a single time point; the performance of the model may de-

cline when extrapolated across seasons. Incorporating multi-temporal shrub identification samples or time-series modules may be an effective technical means to address this issue. Finally, the goal of fine-grained identification of shrubs is to provide reliable samples for the classification of shrub-encroached grasslands. Therefore, future research should focus on how to scale up to satellite observations.

5. Conclusions

In this study, a US-BS-P-Net framework consisting of a band selection algorithm called US-BS-Net and a metric-based prototypical network is proposed for shrub identification in SGs. It extracts spectral signals sensitive to shrub identification from UAV hyperspectral images in an unsupervised band selection manner and inputs them into the prototypical network to build a high-precision shrub identification model in small-sample scenarios.

First, a benchmark for identifying *Caragana microphylla* Lam shrubs in SGs was constructed using UAV hyperspectral images by means of visual interpretation. When the training set accounts for 0.01% of all samples, the experimental results show that the optimal spatial patch size of the sample is 21 × 21, and the OA exceeds 92%. It illustrates the potential of the prototypical network for high-precision shrub identification in small-sample scenarios.

Certainly, our innovation mainly focuses on introducing contrastive learning into the original BS-Conv-Net to construct a novel US-BS-Net for optimizing the selection of spectral bands sensitive to shrub identification. Using only five bands obtained by the proposed US-BS-Net can achieve over 92% OA for shrub identification. When the number of bands is 80, the identification performance reaches the highest OA of 93.15%, which is better than 92.56% of BS-Net-Conv and far exceeds 85.51% of PCA. It is sufficient to demonstrate the contribution of discriminative feature representation learning in the proposed US-BS-Net to the extraction of sensitive bands for shrub identification.

Finally, our US-BS-P-Net was compared with several deep learning algorithms, and its outstanding performance verified the superiority of US-BS-P-Net in shrub identification with extremely few training samples. Meanwhile, the optimal model via US-BS-P-Net was used to generate shrub distribution maps, providing excellent basic data for SG monitoring.

Overall, our proposed dataset and US-BS-P-Net framework are preliminary studies on the application of deep learning algorithms in the field of shrub identification in SGs. The combination of an unsupervised band selection algorithm with few-shot learning is an effective way to achieve high-precision shrub identification using hyperspectral images, especially in small-sample scenarios.

Author Contributions: Conceptualization, L.C. and B.S.; methodology, L.C.; software, L.C.; validation, L.C. and B.S.; formal analysis, L.C.; investigation, L.C.; resources, L.C. and B.S.; data curation, L.C.; writing—original draft preparation, L.C., J.W., and B.S.; writing—review and editing, L.C., J.W., B.S., and Z.G.; visualization, L.C. and J.W.; supervision, B.S. and Z.G.; project administration, B.S. and Z.G.; funding acquisition, L.C. and B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fundamental Research Funds of CAF, grant number CAFYBB2025QF002, and the National Key R&D Program of China, grant number 2024YFF1308104.

Data Availability Statement: The datasets used during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank Wei Yue, Ziyu Yan, Hanwen Cui, Hao Liu, Jiamin Wu, and Rongfen Xia from the Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing, China, for their assistance with the quadrat survey. We also thank AZUP Scientific (www.azup.cn) for their help in UAV hyperspectral data acquisition and preprocessing.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SGs	Shrub-encroached grasslands
UAV	Unmanned aerial vehicle
LiDAR	Light detection and ranging
GNSS	Global navigation satellite system
RTK	Real-time kinematic
DOM	Digital orthophoto map
SimCLR	Simple contrastive learning of representations
OA	Overall accuracy
AA	Average accuracy
Kappa	Kappa coefficient
US-BS-Net	Unsupervised band selection network
PCA	Principal component analysis
2D-CNN	Two-dimensional convolutional neural network

References

1. Eldridge, D.J.; Bowker, M.A.; Maestre, F.T.; Roger, E.; Reynolds, J.F.; Whitford, W.G. Impacts of Shrub Encroachment on Ecosystem Structure and Functioning: Towards a Global Synthesis. *Ecol. Lett.* **2011**, *14*, 709–722. [CrossRef] [PubMed]
2. Applestein, C.; Germino, M.J. Patterns of Post-Fire Invasion of Semiarid Shrub-Steppe Reveals a Diversity of Invasion Niches within an Exotic Annual Grass Community. *Biol. Invasions* **2022**, *24*, 741–759. [CrossRef]
3. Chen, L.; Li, H.; Zhang, P.; Zhao, X.; Zhou, L.; Liu, T.; Hu, H.; Bai, Y.; Shen, H.; Fang, J. Climate and Native Grassland Vegetation as Drivers of the Community Structures of Shrub-Encroached Grasslands in Inner Mongolia, China. *Landsc. Ecol.* **2015**, *30*, 1627–1641. [CrossRef]
4. Deng, Y.; Li, X.; Shi, F.; Hu, X. Woody Plant Encroachment Enhanced Global Vegetation Greening and Ecosystem Water-Use Efficiency. *Glob. Ecol. Biogeogr.* **2021**, *30*, 2337–2353. [CrossRef]
5. Peng, H.Y.; Li, X.Y.; Li, G.Y.; Zhang, Z.H.; Zhang, S.Y.; Li, L.; Zhao, G.Q.; Jiang, Z.Y.; Ma, Y.J. Shrub Encroachment with Increasing Anthropogenic Disturbance in the Semiarid Inner Mongolian Grasslands of China. *Catena* **2013**, *109*, 39–48. [CrossRef]
6. Chen, M.; Zhao, X.Y.; Yue, P.; Guo, X.X.; Qiao, J.J.; Li, X.Y. Effect of Grazing Disturbance on Floral Display, Pollen Limitation and Plant Pollination Efficiency in the Desert Steppe. *BMC Plant Biol.* **2022**, *22*, 514. [CrossRef] [PubMed]
7. Zhou, L.; Shen, H.; Chen, L.; Li, H.; Zhang, P.; Zhao, X.; Liu, T.; Liu, S.; Xing, A.; Hu, H.; et al. Ecological Consequences of Shrub Encroachment in the Grasslands of Northern China. *Landsc. Ecol.* **2019**, *34*, 119–130. [CrossRef]
8. Chen, A.; Xu, C.; Zhang, M.; Guo, J.; Xing, X.; Yang, D.; Xu, B.; Yang, X. Cross-Scale Mapping of above-Ground Biomass and Shrub Dominance by Integrating UAV and Satellite Data in Temperate Grassland. *Remote Sens. Environ.* **2024**, *304*, 114024. [CrossRef]
9. Zhao, Y.; Liu, X.; Wang, Y.; Zheng, Z.; Zheng, S.; Zhao, D.; Bai, Y. UAV-Based Individual Shrub Aboveground Biomass Estimation Calibrated against Terrestrial LiDAR in a Shrub-Encroached Grassland. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *101*, 102358. [CrossRef]
10. McManus, K.M.; Morton, D.C.; Masek, J.G.; Wang, D.; Sexton, J.O.; Nagol, J.R.; Ropars, P.; Boudreau, S. Satellite-Based Evidence for Shrub and Graminoid Tundra Expansion in Northern Quebec from 1986 to 2010. *Glob. Chang. Biol.* **2012**, *18*, 2313–2323. [CrossRef]
11. Brandt, J.S.; Haynes, M.A.; Kuemmerle, T.; Waller, D.M.; Radeloff, V.C. Regime Shift on the Roof of the World: Alpine Meadows Converting to Shrublands in the Southern Himalayas. *Biol. Conserv.* **2013**, *158*, 116–127. [CrossRef]
12. Yang, Y.; Zhao, X.; Shen, H.; Yan, Z.; Shi, Y.; Yang, L.; Fang, J. Persistently-Enhanced Shrub Encroachment in Inner Mongolian Grassland from 1985 to 2023: Insights from Integrated Drone and Landsat Imagery. *Ecol. Indic.* **2025**, *181*, 114404. [CrossRef]
13. Chen, W.; Sun, W.; Wu, Z.; Wang, Y.; Wang, Y.; Bai, Y.; Zhao, Y. Microtopography-Guided Precision Restoration of Sandy Lands through UAV: A Case Study in Hunshandake Sandy Land, China. *Catena* **2024**, *247*, 108489. [CrossRef]
14. Dong, Y.; Yan, H.; Wang, N.; Huang, M.; Hu, Y. Automatic Identification of Shrub-Encroached Grassland in the Mongolian Plateau Based on UAS Remote Sensing. *Remote Sens.* **2019**, *11*, 1623. [CrossRef]
15. Ghiyamat, A.; Shafri, H.Z.M.; Mahdiraji, G.A.; Shariff, A.R.M.; Mansor, S. Hyperspectral Discrimination of Tree Species with Different Classifications Using Single- and Multiple-Endmember. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 177–191. [CrossRef]
16. Zhang, X.Y.; Liu, C.L.; Suen, C.Y. Towards Robust Pattern Recognition: A Review. *Proc. IEEE* **2020**, *108*, 894–922. [CrossRef]
17. Ghosh, A.; Fassnacht, F.E.; Joshi, P.K.; Kochb, B. A Framework for Mapping Tree Species Combining Hyperspectral and LiDAR Data: Role of Selected Classifiers and Sensor across Three Spatial Scales. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 49–63. [CrossRef]

18. Feng, J.; Jiao, L.; Liu, F.; Sun, T.; Zhang, X. Mutual-Information-Based Semi-Supervised Hyperspectral Band Selection with High Discrimination, High Information, and Low Redundancy. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2956–2969. [CrossRef]
19. Jiang, X.; Song, X.; Zhang, Y.; Jiang, J.; Gao, J.; Cai, Z. Laplacian Regularized Spatial-Aware Collaborative Graph for Discriminant Analysis of Hyperspectral Imagery. *Remote Sens.* **2019**, *11*, 29. [CrossRef]
20. Patro, R.N.; Subudhi, S.; Biswal, P.K.; Dell'Acqua, F. A Review of Unsupervised Band Selection Techniques: Land Cover Classification for Hyperspectral Earth Observation Data. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 72–111. [CrossRef]
21. Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A Review of Supervised Object-Based Land-Cover Image Classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 277–293. [CrossRef]
22. Shi, Y.; Skidmore, A.K.; Wang, T.; Holzwarth, S.; Heiden, U.; Pinnel, N.; Zhu, X.; Heurich, M. Tree Species Classification Using Plant Functional Traits from LiDAR and Hyperspectral Data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 207–219. [CrossRef]
23. Sun, X.; Shen, X.; Pang, H.; Fu, X. Multiple Band Prioritization Criteria-Based Band Selection for Hyperspectral Imagery. *Remote Sens.* **2022**, *14*, 5679. [CrossRef]
24. Gong, M.; Zhang, M.; Yuan, Y. Unsupervised Band Selection Based on Evolutionary Multiobjective Optimization for Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 544–557. [CrossRef]
25. Yu, C.; Zhou, S.; Song, M.; Gong, B.; Zhao, E.; Chang, C.I. Unsupervised Hyperspectral Band Selection via Hybrid Graph Convolutional Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [CrossRef]
26. Dou, Z.; Gao, K.; Zhang, X.; Wang, H.; Han, L. Band Selection of Hyperspectral Images Using Attention-Based Autoencoders. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 147–151. [CrossRef]
27. Cai, Y.; Liu, X.; Cai, Z. BS-Nets: An End-to-End Framework for Band Selection of Hyperspectral Image. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 1969–1984. [CrossRef]
28. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum Contrast for Unsupervised Visual Representation Learning. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
29. Sun, X.; Wang, P.; Lu, W.; Zhu, Z.; Lu, X.; He, Q.; Li, J.; Rong, X.; Yang, Z.; Chang, H.; et al. RingMo: A Remote Sensing Foundation Model With Masked Image Modeling. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–22. [CrossRef]
30. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything. In Proceedings of the IEEE International Conference on Computer Vision, Paris, France, 1–6 October 2023.
31. Li, J.; Zhou, P.; Xiong, C.; Hoi, S.C.H. Prototypical Contrastive Learning of Unsupervised Representations. *arXiv* **2020**, arXiv:2005.04966.
32. Zhu, K.; Qiu, S.; Cao, J.; You, Y.; Qiao, Y. IPM-Net: An Interpretable Partial-Level Matching Network for Few-Shot Fine-Grained Classification in Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 1–19. [CrossRef]
33. Fedele, A.; Ferrigno, B.; Guidotti, R. Explaining Relation Networks for Few-Shot Learning. In Proceedings of the 2025 IEEE International Conference on Big Data (BigData), Macau, China, 8–11 December 2025; pp. 3950–3959.
34. Ye, X.; Mei, Y.; Zhou, K.; Wu, Z.; Zhao, J. DDSF: Robust Few-Shot Learning via Disentangled Subspaces with Determinantal Point Process. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Denver, CO, USA, 3–7 June 2026; pp. 19591–19601.
35. Chen, L.; Wu, J.; Xie, Y.; Chen, E.; Zhang, X. Discriminative Feature Constraints via Supervised Contrastive Learning for Few-Shot Forest Tree Species Classification Using Airborne Hyperspectral Images. *Remote Sens. Environ.* **2023**, *295*, 113710. [CrossRef]
36. Gao, K.; Guo, W.; Yu, X.; Liu, B.; Yu, A.; Wei, X. Deep Induction Network for Small Samples Classification of Hyperspectral Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3462–3477. [CrossRef]
37. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the 37th International Conference on Machine Learning, ICML 2020, Virtual, 13–18 July 2020.
38. Wu, J.; Chen, L.; Wang, J.; Li, Y.; Chen, E.; Zhang, X. A Novel Framework Combining Band Selection Algorithm and Improved 3D Prototypical Network for Tree Species Classification Using Airborne Hyperspectral Images. *Comput. Electron. Agric.* **2024**, *219*, 108813. [CrossRef]
39. Mou, L.; Zhu, X.X. Learning to Pay Attention on Spectral Domain: A Spectral Attention Module-Based Convolutional Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 110–122. [CrossRef]
40. Vinyals, O.; Blundell, C.; Lillicrap, T.; Kavukcuoglu, K.; Wierstra, D. Matching Networks for One Shot Learning. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016.
41. Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P.H.S.; Hospedales, T.M. Learning to Compare: Relation Network for Few-Shot Learning. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.