remote sensing

MDPI

Article

Self-Supervised Assisted Semi-Supervised Residual Network for Hyperspectral Image Classification

Liangliang Song [1,2], Zhixi Feng [1,3,*], Shuyuan Yang [1,3], Xinyu Zhang [1] and Licheng Jiao [1]

[1] School of Artifical Intelligentce, Xidian University, Xi'an 710071, China; songll@stu.xidian.edu.cn (L.S.); syyang@xidian.edu.cn (S.Y.); xinyu_zhang@stu.xidian.edu.cn (X.Z.); lchjiao@mail.xidian.edu.cn (L.J.)
[2] Guangzhou Institute of Technology, Xidian University, Xi'an 710071, China
[3] Intelligent Decision and Cognitive Innovation Center, State Administration of Science, Technology and Industry for National Defense, Beijing 100048, China
* Correspondence: zxfeng@xidian.edu.cn

Abstract: Due to the scarcity and high cost of labeled hyperspectral image (HSI) samples, many deep learning methods driven by massive data cannot achieve the intended expectations. Semi-supervised and self-supervised algorithms have advantages in coping with this phenomenon. This paper primarily concentrates on applying self-supervised strategies to make strides in semi-supervised HSI classification. Notably, we design an effective and a unified self-supervised assisted semi-supervised residual network (SSRNet) framework for HSI classification. The SSRNet contains two branches, i.e., a semi-supervised and a self-supervised branch. The semi-supervised branch improves performance by introducing HSI data perturbation via a spectral feature shift. The self-supervised branch characterizes two auxiliary tasks, including masked bands reconstruction and spectral order forecast, to memorize the discriminative features of HSI. SSRNet can better explore unlabeled HSI samples and improve classification performance. Extensive experiments on four benchmarks datasets, including Indian Pines, Pavia University, Salinas, and Houston2013, yield an average overall classification accuracy of 81.65%, 89.38%, 93.47% and 83.93%, which sufficiently demonstrate that SSRNet can exceed expectations compared to state-of-the-art methods.

Keywords: semi-supervised learning; self-supervised learning; auxiliary tasks; hyperspectral image classification

check for updates

Citation: Song, L.; Feng, Z.; Yang, S.; Zhang, X.; Jiao, L. Self-Supervised Assisted Semi-Supervised Residual Network for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 2997. https://doi.org/10.3390/rs14132997

Academic Editors: Adrian Stern and Filiberto Pla

Received: 12 May 2022
Accepted: 21 June 2022
Published: 23 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the development of remote sensing and information processing technology, HSI technology has become the focus in the remote sensing community. The hyperspectral image is a precise remote sensing means that contains rich spatial texture information and spectral reflectance information [1], which has unique advantages in subtle recognition and detection missions. e.g., vegetation cover monitoring [2], atmospheric environmental research [3], and marine monitoring [4].

Based on the recent literature reviewed in [5], HSI classification is the most vibrant field of research in the hyperspectral community (i.e., each pixel in hyperspectral images is appointed a unique category). Within the initial stage of HSI classification, most methods take spectral features, such as independent component analysis (ICA) [6] and support vector machines (SVM) [7], as the primary classification basis. However, the HSI classification results obtained by these methods are unsatisfactory since the spatial features are not well exploited. Due to the characteristics of spatial homogeneity and heterogeneity and mixed pixels of HSI, it is difficult to fully utilize the features of HSI by spectral feature extraction alone. The spatial features can improve the classification performance of HSI [8]; increasingly, classification strategies that combine spatial features have been proposed. For example, a hyperspectral data preprocessing method based on mathematical morphol-

ogy was proposed in [9] in which extended morphological profiles (EMPs) were utilized to exact spatial construction information through morphological manipulation.

With the rapid development of deep learning (DL) [10], DL has been presented in numerous computer vision tasks and has made worthwhile breakthroughs. As a typical deep learning(DL) model, a convolutional neural network (CNN) has been considered to create full utilization of the spatial and spectral features of HSI [11], and researchers have proposed a series of HSI classification methods based on CNN. Hu et al. [12] used a deep CNN model for HSI classification and achieved good performance. Chen et al. [13] proposed a 3D-CNN model for HSI classification, which performs superior to a 1D-CNN and a 2D-CNN. Cheng et al. [14] designed a spatial–spectral random patch network, which made adequate utilization of the spatial and spectral information and achieved satisfactory performance. However, due to the scarcity of labeled samples in HSI, DL-based strategies cannot obtain satisfactory accuracy. The collection and labeling of HSI are complicated, time-consuming, and high-cost. Therefore, the number of labeled training samples is greatly limited, and the shortage of training samples is one of the main obstacles to productive HSI classification methods.

Some recent works have begun to explore self-supervised or semi-supervised strategies for HSI classification to solve this problem. Semi-supervised learning methods aim to improve performance by simultaneously using a few labeled samples and a large number of unlabeled samples. Several semi-supervised methods have been used for HSI classification, which are roughly categorized into three classes: (1) self-training [15–17]; e.g., Li et al. [16] iteratively enlarged the training sample set and retrained the classifier, and the selection of training samples was based on the region information, so the risk of assigning wrong labels was primarily reduced. (2) generative models [18–20]; e.g., Feng et al. [19] proposed a semi-supervised dual-branch convolutional autoencoder with self-attention. (3) graph-based methods [21–23]; e.g., Ding et al. [23] proposed a semi-supervised locality-preserving dense graph neural network (GNN) for HSI classification in which autoregressive moving average filters and context-aware learning are integrated. Moreover, the self-supervised learning methods are also applied to the few-shot HSI classification [24–27]. In [26], a self-supervised contrastive fruitful disymmetrical expanded network is presented for HSI classification. In [27], a self-supervised learning strategy with flexible distillation is proposed for HSI classification. Nevertheless, these approaches suffer from some limitations. Self-training requires high-confidence samples with their "pseudo-labels" to update the training set, and the performance will get worse once the "pseudo-labels" are incorrect. Methods based on a graph should construct a structural graph, but it is troublesome due to the fact that the latent spatial–spectral structural information is not easy to learn.

In view of these limitations, we attempt to incorporate a self-supervised strategy into a semi-supervised framework and propose a unified SSRNet. The proposed SSRNet is designed as two branches: a semi-supervised and a self-supervised branch. The semi-supervised component consists of a residual feature extraction network (RNet) that extracts discriminative spectral–spatial features from HSI cubes. Since random perturbation is proved to be an effective way for robust classification [28–31], we implement perturbation by spectral feature shift in this framework. The self-supervised component consists of two auxiliary tasks: a spectral order forecast and a masked bands reconstruction, which can learn the discriminative features of HSI.

To summarize, the contributions of the proposed methods are threefold as follows:

(1) Self-supervised learning is integrated into a semi-supervised framework for HSI classification by designing a unified multi-task SSRNet. SSRNet has competitive performance, especially under few labeled samples conditions.

(2) A semi-supervised data random perturbation strategy is proposed. This perturbation strategy is the bidirectional movement of some randomly selected spectral segments along the spatial dimension, respectively, on the HSI feature maps.

(3) Two types of self-supervised auxiliary tasks are presented for SSRNet. The two auxiliary tasks, i.e., masked bands reconstruction and spectral order forecast, can help the network learn the discriminative features.

2. Methodology

This section will present the proposed SSRNet, including the semi-supervised and self-supervised branches. In the semi-supervised branch, we amplify the mean-teacher [29] framework with one kind of random perturbation, i.e., spectral feature shift. Furthermore, we design a residual feature extraction network (RNet) for learning spectral–spatial features. In the self-supervised branch, two auxiliary tasks: masked bands reconstruction and spectral order forecast, are explored to help in training the proposed SSRNet. Figure 1 illustrates the outline of the SSRNet.

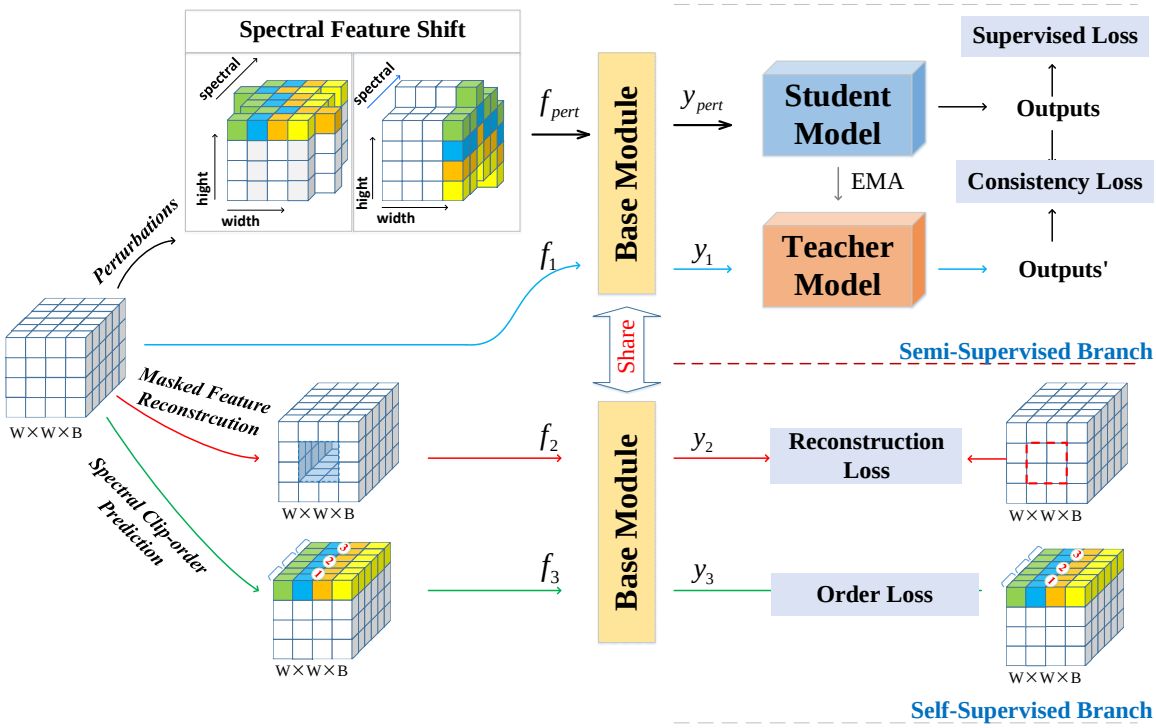


Figure 1. Overview of the proposed SSRNet.

2.1. The Overall Framework of the Proposed SSRNet

The HSI data cube is denoted by $D \in R^{M \times N \times L}$. M and N signify the width and height of the HSI individually. L is the number of spectral bands. The corresponding category label set of each pixel in D is $Y \in R^{1 \times 1 \times C}$ in which C means the amount of the land cover categories. First, principal component analysis (PCA) is utilized to decrease the spectral dimension of HSI, while keeping the same spatial size. We signify the data cube after PCA by $X \in R^{M \times N \times B}$ in which X is the input after PCA and B is the number of spectral bands after PCA, i.e., B andis set as 30 in our framework. Then, the HSI is split into overlapping 3D-patches centering on each pixel, which is represented by $I \in R^{W \times W \times B}$. I is the input data for the SSRNet, and the label of each 3D-patch is determined by the label of the center pixel. Furthermore, the W, i.e., patch size, is set as 11.

Figure 1 illustrates the schematic diagrams of the proposed SSRNet. First, we utilized PCA to reduce the spectral dimensionality of HSI. Then, adjacent cubes of each pixel were taken as the center to form a new data representation. There is a random perturbation of HSI data in the semi-supervised branch: spectral-feature shift. The base module takes the perturbed data and the unperturbed data as inputs. The student and teacher models have a unified model framework and distinctive weight updating strategies. RNet is made up of a base module, a student model and a teacher model. The self-supervised branch consists

of two additional tasks: masked bands reconstruction and spectral order forecast. Lastly, a multi-task framework is explored for optimization.

2.2. Semi-Supervised Learning Branch

This section will introduce the semi-supervised branch of the SSRNet and provide a brief description of the mean-teacher framework. Then, we present the RNet, which includes the base module (BM) and residual feature extraction module (REM). Afterward, we give one type of data random perturbation method called spectral-feature shift.

2.2.1. Mean-Teacher Framework

The mean-teacher framework is extended from a supervised learning paradigm with two models: a student model f_θ and a teacher model $f_{\theta'}$. For the student model, the weights θ are optimized by the HSI supervised losses in the same way as that for supervised learning. The student model is RNet in our proposed SSRNet. The teacher model shares the unified model architecture with the student, but its weights θ' are updated with an exponential moving average (EMA) of the consequences from a sequence of student models of different training iterations. EMA can be formulated as follows:

$$\theta'_T = \beta\theta'_{T-1} + (1-\alpha)\theta_T \tag{1}$$

where T represents the iteration of the training process, and β is a smoothing coefficient. Its default option is 0.999.

2.2.2. The RNet Overview

To verify our semi-supervised framework and better clarify our method, we design a residual network (RNet). The RNet is shown in Figure 2. The RNet comprises two modules: a BM and a residual feature extraction module (REM). The BM copes with the input feature α and outputs feature α' shared by the following REM. The details of the BM are shown in Figure 3. The REM copes with the input feature α' and outputs feature ϕ. Finally, ϕ is sent into the classifier for HSI classification.

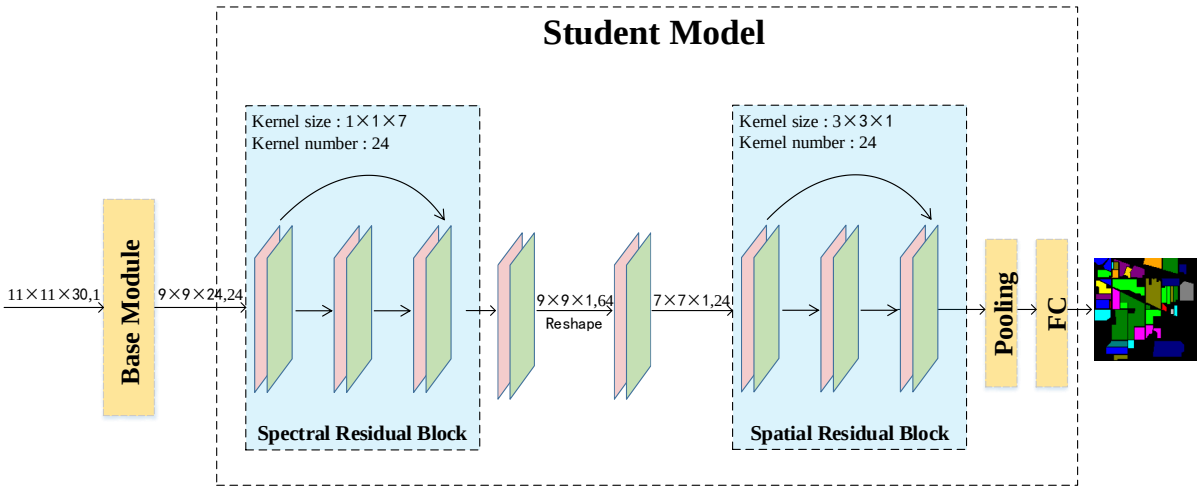


Figure 2. The details structure of the RNet.

The classical CNN model has been applied to hyperspectral classification and has achieved advanced results. However, the classification precision diminishes with the increase in the convolution layers [32]. This problem can be successfully eased by attaching shortcut connections between other layers to make residual blocks [33]. According to the spatial correlation and spectral characteristics of HSI, we designed a residual feature extraction module (REM). The residual structure is shown in Figure 2. We created two kinds of residual feature extraction modules: spectral residual and spatial residual modules.

For the spectral residual feature extraction module, the size of the input feature is $p \times p \times k$ and has n channels. The kernel size of $1 \times 1 \times d$ is applied to the two convolution layers. Meanwhile, the input feature is kept at $p \times p$ unaltered through a padding method. The spectral residual module is formulated as follows:

$$X^{i+2} = X^i + F(X^i; \theta) \tag{2}$$

where X^i means the input feature of the i-th convolution layer, $F(X^i; \theta)$ represents the output feature of the $(i+2)$-th convolution layer, and θ is the weight parameter of the convolution layers.

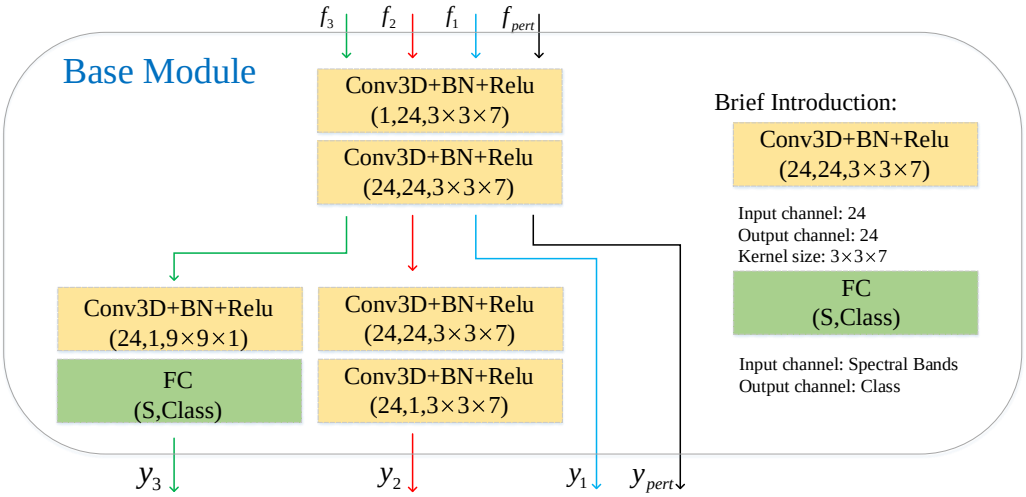


Figure 3. The details structure of the base module.

2.2.3. Data Random Perturbation

Random perturbation is proved to be effective for robust semi-supervised learning models [28–31,34]. The work in [29] adds Gaussian noise to intermediate feature maps of the mean-teacher framework. For HSI data, spatial and spectral information are both essential. We propose a primary data random perturbation in our work: spectral feature shift.

Spectral feature shift is the bidirectional movement of some randomly selected spectral segments along the horizontal and vertical dimensions on the feature maps, respectively. The schematic diagrams of the spectral feature shift are shown in Figure 1. Therefore, spectral feature shift can significantly boost the multifariousness of the input features and make the semi-supervised learning model more robust. First, we randomly select μ spectral bands. Then, $\mu/2$ spectral bands are bidirectionally offset in the horizontal spatial dimension, and the other $\mu/2$ spectral bands are bidirectionally offset in the vertical spatial dimension. We use μ to mean the rank of the spectral shift. Moreover, we will discuss the influence of μ size on HSI classification accuracy in Section 3.

Each mini-batch incorporates both labeled HSI data and unlabeled HSI data during the training process. Moreover, we take a dropout method to avoid overfitting. The dropout strategy is a simple and effective technique that prevents overfitting by discarding a certain percentage of units during the training process [35]. In the mean-teacher framework, the labeled samples are trained using supervised loss. Unlabeled samples have no ground truth labels, so their supervised loss is undefined. Consistency regularization utilizes unlabeled HSI data based on the assumption that the model should output similar forecasts when fed perturbed forms of the same input. The consistency loss is utilized to the labeled HSI data and unlabeled HSI data in the semi-supervised branch. Therefore, the total loss in the semi-supervised branch is:

$$L_{supervised} = -\frac{1}{N_c}\sum_{i=1}^{N_c} y_i \log y_i' \tag{3}$$

$$L_{spectral_shift} = \frac{1}{N}\sum_{i=1}^{N} \|stu(x_{si}) - tea(x_i)\|_2 \tag{4}$$

$$L_{semi} = L_{supervised} + \lambda_1 L_{spectral_shift} \tag{5}$$

where y_i means the label of i-th training sample and y_i' indicates the predictive label, N_c denotes the training samples with labels in each mini-batch, and x_{si} means the i-th training sample of spectral feature shift and $stu(x_{si})$ is the output of the student model. Here, x_i represents the i-th training sample and $tea(x_i)$ is the output of the teacher model. It should be noted that only the training samples of the student network will be perturbed by the spectral feature shift. The hyper-parameters λ_1 are set to 1. $L_{spectral_shift}$ is consistency loss for spectral feature shift random perturbation, and $L_{spectral_shift}$ is $L2$-loss. $L_{supervised}$ is supervised loss for the labeled data, and $L_{supervised}$ is typical cross entropy loss.

2.3. Self-Supervised Learning Branch

Motivated by recent advances in self-supervised learning of HSI classification [32,36–39], we hypothesize that the semi-supervised HSI classification method could significantly benefit from self-supervised learning strategies. Furthermore, based on this motivation, we propose two auxiliary tasks in the self-supervised branch: masked bands reconstruction and spectral order forecast.

2.3.1. Masked Bands Reconstruction

As shown in Figure 1, the critical thought of this self-supervised auxiliary task is to generate the feature f_2 by stochastically masking the HSI feature f_1 at a few areas on the spatial dimensions. Then the BM utilizes f_2 to reconstruct f_1. The schematic diagrams of the BM are shown in Figure 3. Masked bands reconstruction generates self-supervised signs from the original HSI feature f_1, which could learn discriminative representations simply and effectively. The loss formula for the masked features auxiliary reconstruction task is:

$$L_{masked_recons} = \frac{1}{N}\sum_{i}^{N} \|m(x_i) - y_{ri}\|_2 \tag{6}$$

where $m(x_i)$ represents the i-th training sample of randomly masking the feature at some area along the spatial dimension, y_{ri} means the reconstruction of the i-th training sample, and N represents the number of mini-batches.

At the same time, the SSRNet is trained in a multi-task pattern.In the pretext task of reconstruction of masked bands, the BM will be driven to notice and aggregate features from the context to predict the discarded areas. In this way, the learned features spontaneously conduct semi-supervised HSI classification. We use κ to mean the rank of the mask. Moreover, we will discuss the influence of κ size on HSI classification accuracy in Section 3.

2.3.2. Spectral Order Forecast

As shown in Figure 1, this auxiliary task needs to predict spectral feature sequences corrected in stochastically scrambled feature maps. The spectral order forecast is formulated as a classification task. The input is an HSI patch of clip spectral order, and the output is a probability distribution of the spectral order. The loss presentation for the spectral order forecast auxiliary task is:

$$L_{spectral_order} = -\frac{1}{N}\sum_{i=1}^{N} s_i \log s_i' \tag{7}$$

where s_i means the i-th sample label for the correct spectral order, s_i' represents the i-th sample label for the predicted spectral order. Spectral order forecast can utilize the spectral order of features to learn discriminative spectral representations.

2.4. Overall Loss

The overall loss is made up of the losses from Sections 2.2 and 2.3 The total loss function is:

$$L_{total} = L_{semi} + \lambda_2 L_{masked_recons} + \lambda_3 L_{spectral_order} \quad (8)$$

Loss function L_{masked_recons} is designed for masked bands reconstruction, and L_{masked_recons} is $L2$-loss, while $L_{spectral_order}$ is cross entropy loss. Finally, the total loss function L_{total} is made up of the L_{semi}, L_{masked_recons} and the $L_{spectral_order}$. Hyper-parameters λ_2 and λ_3 are set to 0.0001 and 0.001. A multi-task framework is exploited for optimization.

3. Experiments

In this section, we describe the experiment we performed to demonstrate the effectiveness of the proposed SSRNet for HSI classification. First, a brief introduction of the datasets used in the experiment is given, and then the experimental setup and comparison with other advanced methods follow. Lastly, we analyze the running time of the SSRNet and perform an ablation study to affirm the effectiveness of each component.

3.1. Dataset Description

Four widely used HSI datasets were used in the experiments, including Indian Pines, University of Pavia (PaviaU), Salinas and Houston 2013 datasets. We will give a short introduction of datasets as follows:

(1) Indian Pines: The dataset was collected by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over Northwestern Indiana. The data contain 200 spectral bands in the wavelength within range of 0.4 µm to 2.5 µm and 16 land cover categories. The spatial size is 145 × 145 pixels with resolution of 20 m/pixel. The total number of valid samples is 10,249 excluding background samples.

(2) University of Pavia: The dataset was captured by the ROSIS-03 sensor in the University of Pavia, Italy. The dataset consists of 103 spectral bands with the wavelength range of 0.43 µm to 0.86 µm and nine land cover classes, and its spatial size is 610 × 340 pixels, with a total of 42,776 labeled samples excluding background classes.

(3) Salinas: The dataset was collected by the AVIRIS sensor over Salinas Valley, California, USA. and consists of 204 spectral bands in the wavelength range of 0.4 µm to 2.5 µm and 16 land cover classes. The spatial size is 512 × 217 with the spatial resolution is 3.7 m/pixel. The total number of valid samples is 54,129, excluding background classes.

(4) Houston 2013: The dataset was acquired by the ITRES CASI-1500 sensor at the Houston University campus and its surrounding area, which contains 144 spectral bands in the wavelength range of 0.38 µm to 1.05 µm and 15 land cover classes, and its spatial size is 349 × 1905, with a total of 15,029 labeled samples, excluding background classes. The spatial resolution is 2.5 m/pixel.

3.2. Experiment Setup

To assess the effectiveness of the SSRNet, we contrasted the SSRNet with other advanced HSI classification methods, including the traditional feature extraction method, SVM with RBF kernel [7], and other deep-learning-based classifiers: SSRN [40], SSLSTM [41], DBMA [42], HybridSN [43], CDCNN [44] and 3D-CAE [37]. Our proposed SSRNet and all DL-based methods were executed with PyTorch, and SVM was executed with sklearn. To make full utilization of unlabeled samples to improve learning performance, we randomly selected 10 labeled samples and 20% unlabeled samples for each class as training samples, and the rest as testing samples. For other supervised methods, we selected 10 labeled samples for each class as training samples. Details of training samples and testing samples of the dataset are listed in Tables 1–4. The batch size was 16, the optimizer was Adam [45]

with a learning rate of 0.0005, and the number of epochs was 80. Moreover, all experimTents at the same computing platform were configured with NVIDIA GeForce GTX1660 SUPER GPU and with 8 GB of memory. Next, the above methods will be briefly introduced.

Table 1. The number of training, testing samples in the Indian Pines dataset.

Class No.	Land Cover Type	Training	Testing
1	Alfalfa	10	27
2	Corn-notill	10	1133
3	Corn-mintill	10	655
4	Corn	10	180
5	Grass-pasture	10	377
6	Grass-tree	10	575
7	Grass-pasture-mowed	10	13
8	Hay-windrowed	10	373
9	Oats	10	7
10	Soybean-notill	10	768
11	Soybean-mintill	10	1955
12	Soybean-clean	10	465
13	Wheat	10	155
14	Woods	10	1003
15	Buildings-Grass-Trees	10	299
16	Stone-Steel-Towers	10	65
	Total	160	8050

Table 2. The number of training, testing samples in the University of Pavia dataset.

Class No.	Land Cover Type	Training	Testing
1	Asphalt	10	5295
2	Meadows	10	14,910
3	Gravel	10	1670
4	Trees	10	2442
5	Metal Sheets	10	1067
6	Bare Soil	10	4014
7	Bitumen	10	1055
8	Bricks	10	2936
9	Shadows	10	748
	Total	90	34,137

Table 3. The number of training, testing samples in the Salinas dataset.

Class No.	Land Cover Type	Training	Testing
1	Brocoli-green-weeds-1	10	1598
2	Brocoli-green-weeds-2	10	2971
3	Fallow	10	1571
4	Fallow-rough-plow	10	1106
5	Fallow-smooth	10	2133
6	Stubble	10	3158
7	Celery	10	2854
8	Grapes-untrained	10	9007
9	Soil-vinyard-develop	10	9007
10	Corn-senesced-green-weeds	10	4953
11	Lettuce-romaine-4wk	10	2613
12	Lettuce-romaine-5wk	10	845
13	Lettuce-romaine-6wk	10	1532
14	Lettuce-romaine-7wk	10	723
15	Vinyard-untrained	10	847
16	Vinyard-vertical-trellis	10	5805
	Total	160	43,152

Table 4. The number of training and testing samples in the Houston 2013 dataset.

Class No.	Land Cover Type	Training	Testing
1	Healthy grass	10	991
2	Stressed grass	10	994
3	Synthetic grass	10	548
4	Trees	10	986
5	Soil	10	984
6	Water	10	251
7	Residential	10	1005
8	Commercial	10	986
9	Road	10	992
10	Highway	10	972
11	Railway	10	979
12	Parking Lot 1	10	977
13	Parking Lot 2	10	366
14	Tennis Court	10	333
15	Running Track	10	519
	Total	150	11,883

(1) SVM [7] is a traditional classification method, and all spectral bands are taken as the input of the SVM with a radial basis function (RBF) kernel.

(2) SSRN [40] is a fully supervised method based on ResNet and 3D CNN in which the patch size is set to be 7×7.

(3) SSLSTM [41] is a method using spectral–spatial long short-term memory (LSTM) networks.

(4) DBMA [42] is a supervised method based on a 3D CNN, attention mechanism and DenseNet in which the patch size is set to be 7×7.

(5) HybridSN [43] is a supervised method of mixing a 3D CNN and a 2D CNN in which the patch size is set to be 25×25.

(6) CDCNN [44] is a supervised method based on a 2D CNN and ResNet in which the patch size is set to be 5×5.

(7) 3D-CAE [37] is a self-supervised method based on a 3D convolutional autoencoder in which the patch size is set to be 5×5.

3.3. Experimental Results

The results of overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa) are demonstrated in Tables 5–8. Each experiment was repeated 10 times in which the mean and the standard deviation of each index were reported. Figure 4–7 illustrates the classification maps of our SSRNet and other compared methods. The proposed methods outperform the compared methods on four datasets.

For instance, in the Indian Pines dataset, the SSRNet achieved the best OA of 81.65%, our SSRNet yielded over 20% higher than SVM (54.22%) and CDCNN (58.50%), and over 4% higher than other advanced deep-learning-based methods SSRN (77.48%) and DBMA (70.73%). In the University of Pavia, our proposed SSRNet was over 20% higher than the traditional SVM (64.32%), which was higher than the counterpart of SSRN (85.24%), CDCNN (74.39%) and DBMA (85.66%). Especially in the Salinas dataset, the SSRNet achieved the best OA of 93.47%, over 4% higher than advanced DBMA, but the OA of the other methods did not reach above 90%. Compared with self-supervised methods, the SSRNet was superior to the 3D-CAE in all four datasets, especially in the Indian Pines dataset, where the OA of SSRNet yielded over 20% higher than the 3D-CAE. We also made a visual comparison utilizing classification maps obtained by those methods. Other methods have classification errors in almost every land cover, and SSLSTM, CDCNN and 3D-CAE have obvious classification errors in the class of Broccoli-green-weeds-2 in the Salinas dataset. It can be observed that the SSRNet restored the distribution of surface

objects well and maintained the best boundary region for these four datasets, which further validates the outstanding performance of the SSRNet.

Table 5. Classification results on the Indian Pines dataset with 10 labeled samples, including OA (%), AA (%) and Kappa×100 in which the best results are denoted in bold.

Class No.	SVM [7]	SSLSTM [41]	CDCNN [44]	3DCAE [37]	SSRN [40]	HybridSN [43]	DBMA [42]	Proposed
1	20.59 ± 4.55	30.55 ± 10.8	35.84 ± 13.8	52.98 ± 28.9	71.49 ± 19.3	34.37 ± 25.6	74.96 ± 15.2	**98.76** ± 1.74
2	42.21 ± 4.31	55.35 ± 7.89	55.86 ± 13.3	50.15 ± 19.4	**76.25** ± 7.47	58.12 ± 8.74	65.09 ± 9.11	66.66 ± 11.7
3	35.40 ± 10.0	37.91 ± 11.1	39.87 ± 8.63	48.59 ± 14.1	**69.08** ± 16.1	45.15 ± 15.0	57.39 ± 14.8	64.98 ± 7.70
4	23.26 ± 3.96	37.88 ± 12.2	34.71 ± 9.83	31.33 ± 10.6	57.53 ± 15.3	35.23 ± 15.7	54.05 ± 15.9	**95.36** ± 4.72
5	63.52 ± 8.94	64.48 ± 19.2	65.62 ± 15.5	78.46 ± 12.0	**93.90** ± 7.17	70.62 ± 18.0	92.35 ± 4.55	85.76 ± 5.11
6	87.13 ± 2.95	78.57 ± 10.9	85.22 ± 6.22	72.74 ± 17.8	96.64 ± 3.54	82.52 ± 10.9	97.43 ± 2.97	**98.60** ± 0.37
7	26.52 ± 12.4	18.81 ± 6.41	23.40 ± 0.11	35.70 ± 23.1	45.81 ± 21.9	27.74 ± 24.3	25.73 ± 9.42	**100.0** ± 0.00
8	95.51 ± 1.08	95.07 ± 4.30	96.42 ± 11.9	82.92 ± 28.1	98.63 ± 3.22	77.09 ± 31.8	**99.86** ± 0.22	99.55 ± 0.63
9	13.41 ± 5.06	14.39 ± 6.69	17.52 ± 3.09	20.37 ± 20.3	46.15 ± 18.9	20.35 ± 21.0	09.61 ± 4.89	**100.0** ± 0.00
10	46.77 ± 7.51	46.61 ± 7.86	17.52 ± 11.6	49.53 ± 21.4	67.42 ± 12.3	57.34 ± 9.61	67.32 ± 12.5	**79.16** ± 3.40
11	62.14 ± 4.69	63.90 ± 6.00	64.98 ± 10.1	64.19 ± 21.8	**79.76** ± 6.15	69.35 ± 9.09	78.54 ± 8.25	75.91 ± 4.42
12	28.09 ± 2.84	31.50 ± 6.74	31.74 ± 7.28	42.24 ± 17.1	59.93 ± 14.6	35.25 ± 14.6	52.58 ± 18.6	**79.92** ± 3.78
13	82.81 ± 5.26	76.39 ± 8.12	83.83 ± 7.09	73.12 ± 19.3	94.87 ± 4.78	77.85 ± 16.0	91.57 ± 8.62	**99.13** ± 1.22
14	89.44 ± 4.23	83.15 ± 6.27	87.19 ± 11.0	86.55 ± 9.08	**97.12** ± 1.75	86.37 ± 10.1	94.90 ± 4.42	94.94 ± 0.44
15	42.56 ± 6.75	47.70 ± 9.20	53.70 ± 4.00	48.42 ± 17.2	75.73 ± 10.2	41.17 ± 9.81	60.55 ± 9.66	**92.52** ± 5.06
16	91.39 ± 8.22	67.35 ± 11.6	68.15 ± 18.9	61.69 ± 12.3	84.12 ± 9.81	50.83 ± 6.68	81.72 ± 9.11	**100.0** ± 0.00
OA (%)	54.22 ± 2.47	56.57 ± 3.49	58.50 ± 3.36	57.24 ± 4.66	77.48 ± 3.96	57.31 ± 5.83	70.73 ± 4.83	**81.65** ± 1.71
AA (%)	55.09 ± 1.74	53.10 ± 2.25	55.89 ± 3.53	56.19 ± 6.37	75.90 ± 3.71	54.34 ± 4.87	68.98 ± 2.67	**86.46** ± 1.17
Kappa×100	48.88 ± 2.64	51.08 ± 3.62	53.20 ± 3.56	52.75 ± 4.71	74.61 ± 4.24	52.66 ± 6.05	67.09 ± 5.25	**79.21** ± 1.92

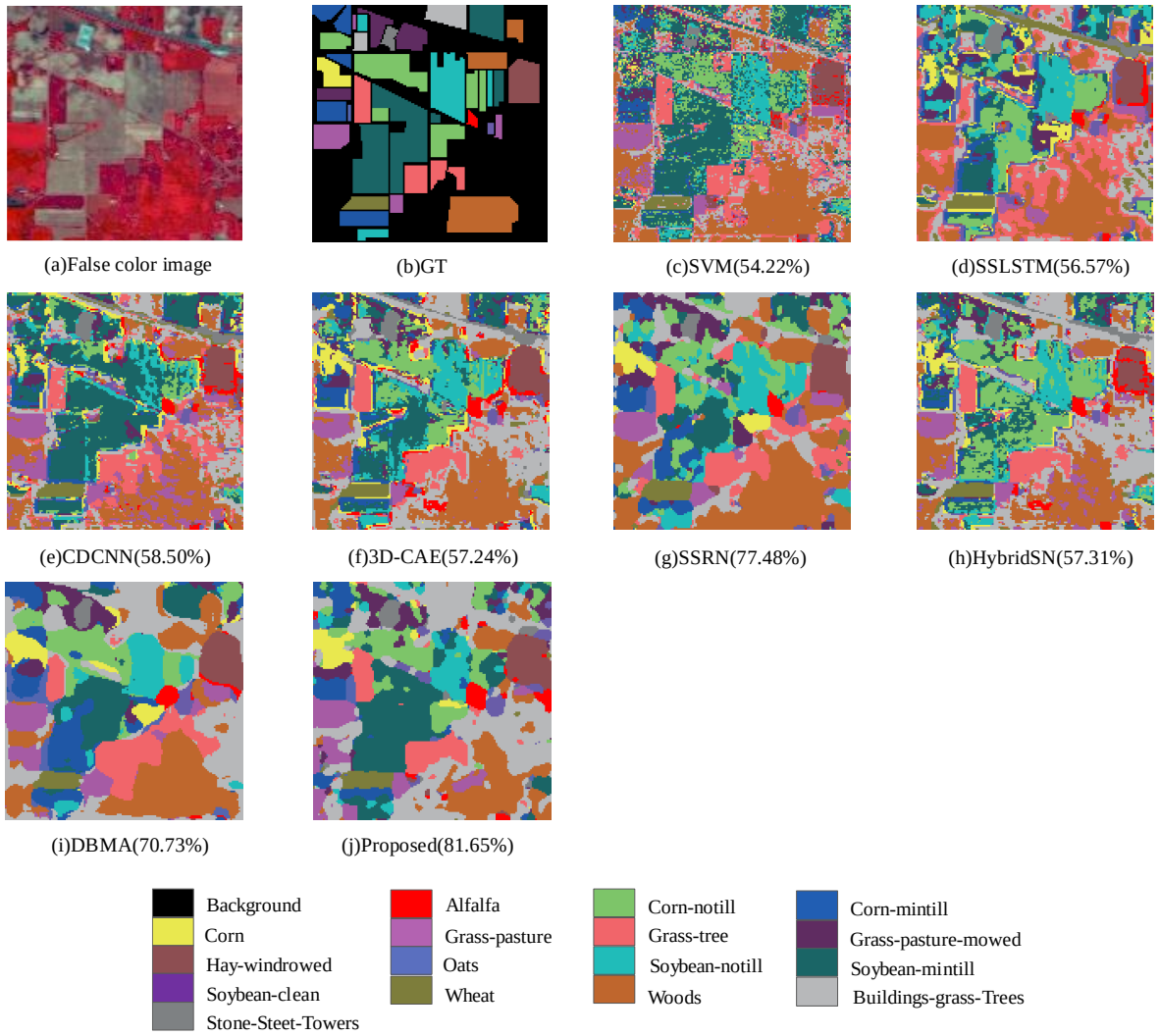


Figure 4. The classification maps of the Indian Pines with 10 labeled samples.

Table 6. Classification results on the University of Pavia dataset with 10 labeled samples, including OA (%), AA (%) and Kappa×100 in which the best results are denoted in bold.

Class No.	SVM [7]	SSLSTM [41]	CDCNN [44]	3DCAE [37]	SSRN [40]	HybridSN [43]	DBMA [42]	Proposed
1	92.38 ± 2.05	92.66 ± 1.65	90.29 ± 3.17	80.00 ± 13.1	**98.18** ± 1.16	54.58 ± 30.3	95.64 ± 2.06	86.09 ± 9.03
2	81.67 ± 8.25	88.39 ± 1.30	92.00 ± 2.28	93.54 ± 1.32	96.29 ± 2.13	73.87 ± 37.2	**96.78** ± 2.40	90.80 ± 2.75
3	40.65 ± 4.72	52.57 ± 5.11	51.01 ± 14.8	58.49 ± 5.70	64.42 ± 11.9	34.77 ± 20.9	77.21 ± 12.1	**91.01** ± 7.41
4	60.97 ± 11.8	70.93 ± 11.8	75.75 ± 16.1	66.91 ± 10.1	79.92 ± 16.9	60.83 ± 23.3	85.22 ± 18.0	**93.51** ± 5.09
5	90.77 ± 6.92	92.17 ± 3.92	93.47 ± 5.77	63.32 ± 44.8	99.10 ± 1.48	94.49 ± 4.96	98.86 ± 1.21	**99.75** ± 0.35
6	34.24 ± 5.09	45.63 ± 6.35	51.85 ± 16.1	78.07 ± 4.04	73.38 ± 15.3	58.08 ± 19.7	65.15 ± 13.5	**96.51** ± 3.04
7	44.62 ± 5.84	50.90 ± 6.05	52.51 ± 14.5	65.12 ± 11.9	66.81 ± 16.1	53.73 ± 17.8	85.52 ± 15.6	**97.75** ± 1.51
8	70.31 ± 6.73	79.22 ± 2.64	73.13 ± 6.80	49.34 ± 1.70	79.65 ± 6.81	46.58 ± 11.6	**81.79** ± 7.17	64.77 ± 31.3
9	99.88 ± 0.10	**99.89** ± 0.09	61.63 ± 24.7	65.24 ± 19.5	99.55 ± 0.86	57.31 ± 20.7	92.01 ± 4.25	98.74 ± 1.24
OA (%)	64.32 ± 6.29	74.29 ± 2.72	74.39 ± 6.50	76.54 ± 4.68	85.24 ± 4.32	63.05 ± 12.5	85.66 ± 4.55	**89.38** ± 1.14
AA (%)	68.39 ± 1.85	74.71 ± 2.08	71.29 ± 5.36	68.89 ± 5.49	84.14 ± 2.90	59.36 ± 9.42	86.46 ± 3.74	**90.99** ± 1.88
Kappa×100	55.47 ± 6.63	67.36 ± 3.13	67.68 ± 7.60	70.13 ± 5.45	81.08 ± 5.25	55.59 ± 12.4	81.72 ± 5.49	**86.18** ± 1.52

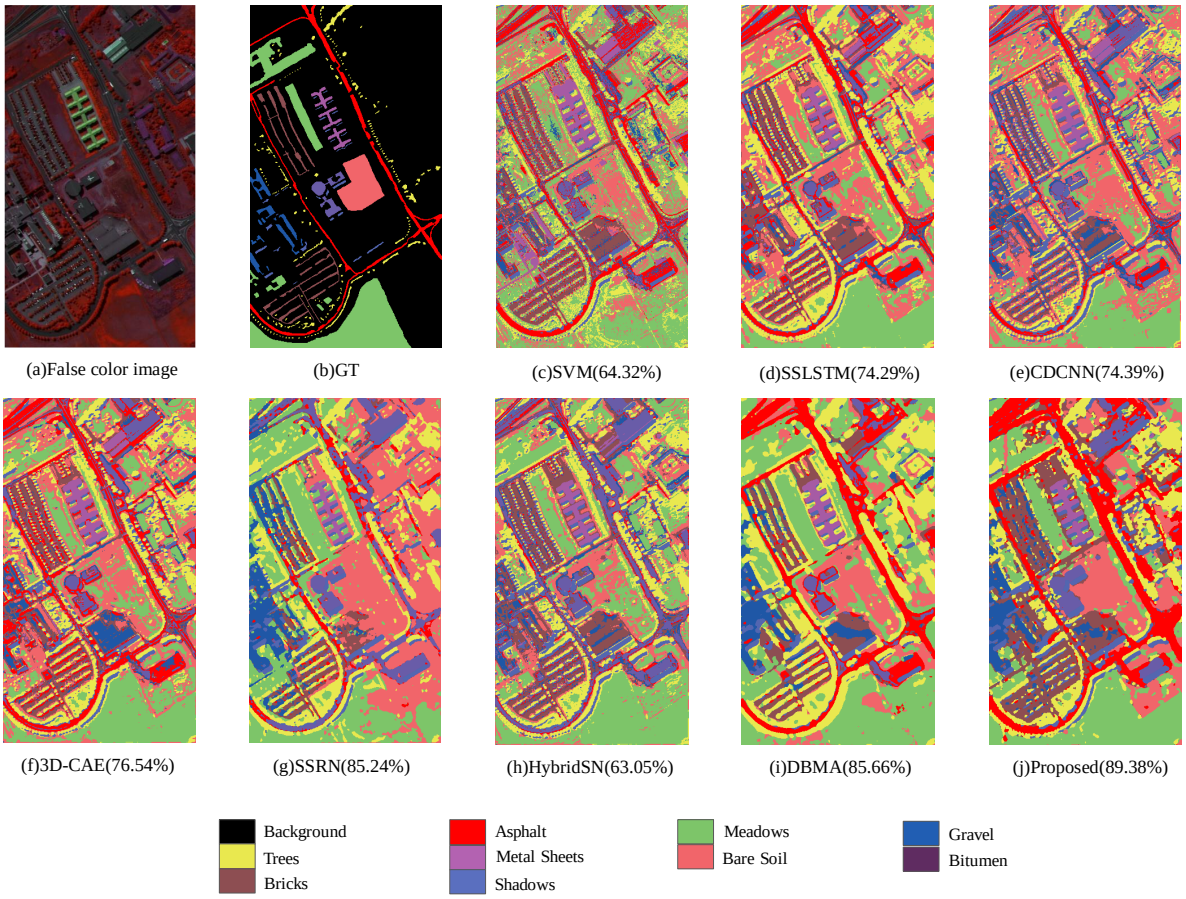


Figure 5. The classification maps of the University of Pavia with 10 labeled samples.

Table 7. Classification results on the Salinas dataset with 10 labeled samples, including OA (%), AA (%) and Kappa×100 in which the best results are denoted in bold.

Class No.	SVM [7]	SSLSTM [41]	CDCNN [44]	3DCAE [37]	SSRN [40]	HybridSN [43]	DBMA [42]	Proposed
1	98.49 ± 1.37	81.94 ± 21.2	85.02 ± 20.2	81.68 ± 16.2	87.74 ± 30.0	95.46 ± 6.99	99.10 ± 2.68	**99.70** ± 0.23
2	98.95 ± 0.41	85.61 ± 15.4	96.29 ± 6.37	89.38 ± 4.85	99.96 ± 0.07	94.02 ± 6.43	**99.99** ± 0.02	97.76 ± 2.87
3	86.03 ± 5.29	93.98 ± 4.91	91.71 ± 8.09	64.88 ± 45.8	92.56 ± 4.19	95.29 ± 4.58	97.60 ± 1.10	**99.95** ± 0.06
4	97.30 ± 1.06	96.62 ± 2.38	93.23 ± 8.56	74.28 ± 23.9	92.98 ± 13.2	92.07 ± 8.11	90.69 ± 2.57	**98.79** ± 1.28
5	97.14 ± 1.70	99.09 ± 0.48	96.47 ± 3.94	93.84 ± 4.28	98.47 ± 2.32	93.55 ± 4.99	**98.75** ± 1.64	97.24 ± 0.05
6	99.94 ± 0.06	98.93 ± 0.66	97.01 ± 2.17	95.86 ± 1.90	**99.94** ± 0.06	97.65 ± 3.46	99.58 ± 0.53	99.61 ± 0.47
7	95.38 ± 2.65	98.67 ± 1.05	97.25 ± 2.91	94.80 ± 5.35	96.50 ± 6.35	97.62 ± 2.17	97.85 ± 2.71	**99.45** ± 0.77
8	70.82 ± 2.56	80.39 ± 7.40	68.33 ± 21.7	80.99 ± 3.23	82.33 ± 4.75	83.57 ± 5.78	**89.42** ± 5.00	84.84 ± 4.97
9	98.83 ± 1.15	98.27 ± 1.54	99.41 ± 0.44	92.80 ± 1.82	97.27 ± 6.38	95.79 ± 4.68	99.36 ± 0.41	**99.86** ± 0.16
10	78.67 ± 9.37	87.99 ± 1.93	85.23 ± 5.82	90.92 ± 6.56	**94.13** ± 4.28	86.57 ± 10.2	90.55 ± 4.97	89.34 ± 6.07
11	79.57 ± 7.85	81.93 ± 8.79	72.03 ± 11.2	74.02 ± 13.6	95.38 ± 2.02	83.11 ± 16.6	91.55 ± 7.81	**99.84** ± 0.22
12	93.88 ± 3.65	96.57 ± 1.31	95.91 ± 3.22	96.74 ± 2.85	99.34 ± 0.65	86.96 ± 29.1	**99.40** ± 0.95	98.86 ± 1.41
13	91.47 ± 5.19	92.15 ± 3.22	88.81 ± 7.92	54.86 ± 8.18	96.05 ± 8.74	33.20 ± 42.2	91.69 ± 6.83	**98.42** ± 1.43
14	83.71 ± 9.73	95.53 ± 3.42	92.89 ± 4.44	71.57 ± 8.54	88.55 ± 22.8	53.57 ± 24.5	92.71 ± 7.94	**98.97** ± 0.54
15	54.96 ± 5.77	44.81 ± 3.07	52.23 ± 7.04	70.13 ± 14.7	66.87 ± 9.12	78.20 ± 7.75	64.27 ± 12.8	**83.68** ± 3.80
16	90.53 ± 5.75	92.04 ± 9.08	94.44 ± 3.91	83.88 ± 7.25	99.37 ± 0.82	89.07 ± 9.26	98.35 ± 1.97	**99.44** ± 0.46
OA (%)	83.53 ± 1.81	79.39 ± 2.82	81.00 ± 3.91	83.27 ± 5.72	88.32 ± 5.76	87.35 ± 3.30	88.84 ± 4.03	**93.47** ± 1.04
AA (%)	88.48 ± 1.22	89.03 ± 2.76	87.89 ± 2.37	81.91 ± 6.86	92.96 ± 5.18	84.73 ± 6.41	93.80 ± 1.33	**96.61** ± 0.05
Kappa×100	81.73 ± 1.97	77.31 ± 3.09	78.97 ± 4.24	81.52 ± 6.25	87.03 ± 6.35	85.97 ± 3.64	87.67 ± 4.40	**92.75** ± 1.15

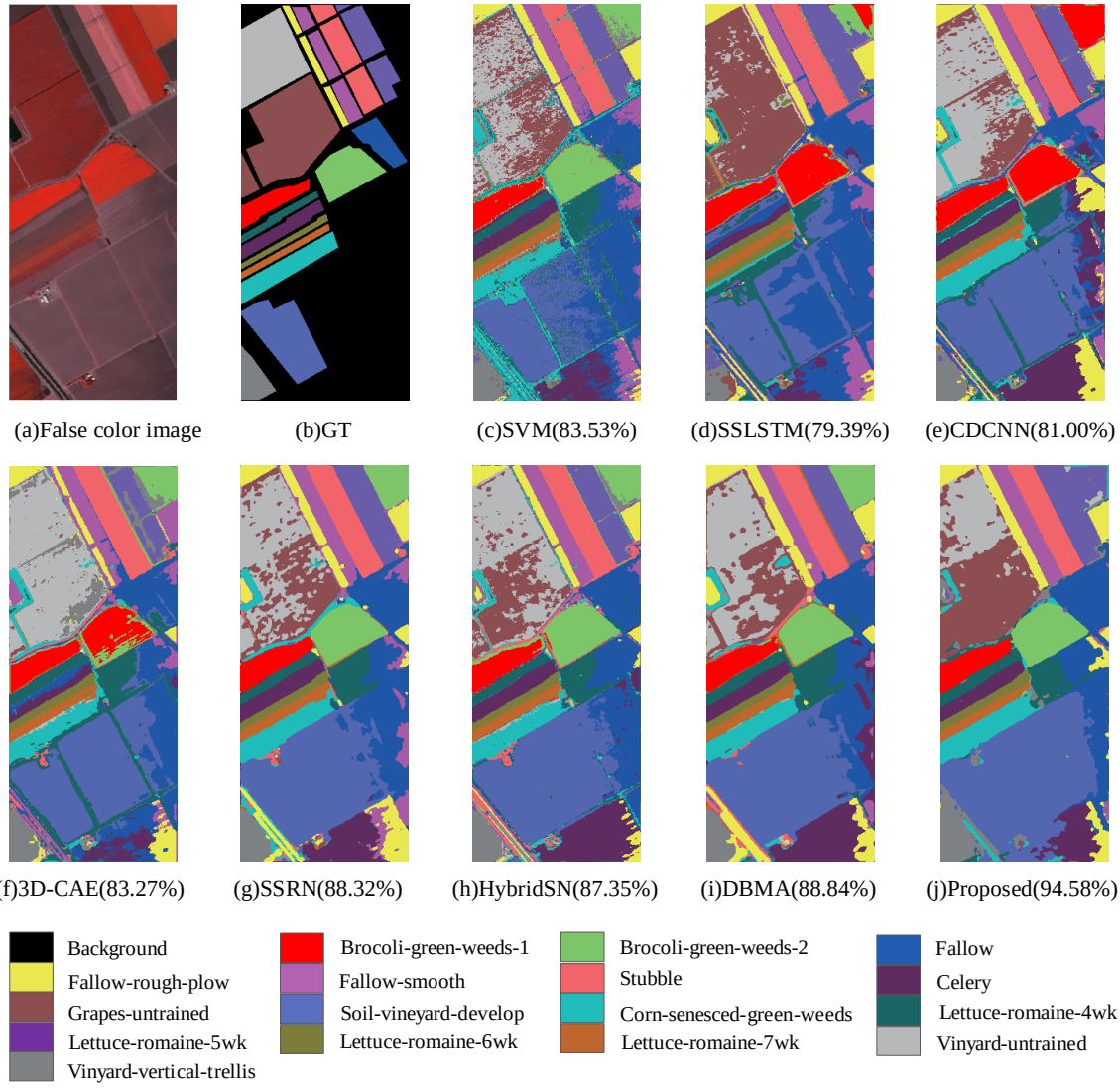


Figure 6. The classification maps of the Salinas with 10 labeled samples.

Table 8. Classification results on the Houston 2013 dataset with 10 labeled samples, including OA (%), AA (%) and Kappa×100 in which the best results are denoted in bold.

Class No.	SVM [7]	SSLSTM [41]	CDCNN [44]	3DCAE [37]	SSRN [40]	HybridSN [43]	DBMA [42]	Proposed
1	**88.65** ± 4.32	79.20 ± 10.8	81.77 ± 9.50	87.56 ± 6.25	83.84 ± 5.78	75.18 ± 26.5	88.33 ± 4.24	83.18 ± 4.24
2	91.38 ± 6.54	94.13 ± 7.35	89.77 ± 10.7	93.06 ± 3.65	**94.26** ± 5.24	77.59 ± 12.4	92.36 ± 4.33	87.58 ± 9.24
3	87.88 ± 11.2	75.87 ± 23.5	88.05 ± 15.7	87.45 ± 10.6	99.64 ± 0.43	92.56 ± 6.99	**99.94** ± 0.11	98.17 ± 2.06
4	96.09 ± 3.53	91.05 ± 10.6	94.96 ± 5.69	91.28 ± 3.33	**97.07** ± 2.25	81.88 ± 12.0	94.66 ± 8.99	91.74 ± 4.71
5	90.91 ± 2.62	92.36 ± 2.71	95.36 ± 2.75	88.39 ± 2.51	93.15 ± 4.83	84.62 ± 8.58	93.91 ± 3.11	**99.96** ± 0.05
6	93.99 ± 5.88	95.70 ± 3.31	85.34 ± 6.97	93.55 ± 5.95	94.97 ± 8.88	79.39 ± 18.9	95.22 ± 3.26	**99.46** ± 0.75
7	67.72 ± 9.40	**79.74** ± 7.14	74.40 ± 4.42	70.47 ± 14.3	75.90 ± 10.1	63.50 ± 13.0	78.17 ± 12.6	75.28 ± 2.72
8	66.75 ± 10.9	84.54 ± 4.33	81.29 ± 9.67	56.72 ± 5.72	80.49 ± 21.4	49.88 ± 30.8	**94.07** ± 6.12	64.56 ± 13.0
9	62.88 ± 9.74	72.80 ± 7.18	70.61 ± 9.83	61.88 ± 14.9	66.63 ± 7.93	64.06 ± 12.3	**75.42** ± 8.31	70.46 ± 7.82
10	59.57 ± 7.49	67.72 ± 9.71	61.68 ± 9.19	39.36 ± 28.2	69.39 ± 12.3	57.21 ± 24.7	70.02 ± 10.7	**78.22** ± 4.01
11	58.80 ± 6.83	64.71 ± 7.71	65.33 ± 9.44	83.24 ± 3.79	77.09 ± 6.22	74.47 ± 9.72	61.67 ± 11.1	**85.52** ± 6.27
12	59.63 ± 4.29	75.72 ± 7.38	74.23 ± 5.62	57.91 ± 8.05	73.57 ± 9.88	62.53 ± 10.5	75.75 ± 8.49	**82.01** ± 8.35
13	31.88 ± 10.3	83.53 ± 8.37	80.28 ± 8.39	64.04 ± 19.4	**93.11** ± 2.99	71.87 ± 11.2	77.75 ± 13.8	90.97 ± 5.78
14	79.28 ± 8.88	76.69 ± 12.1	73.72 ± 11.9	89.26 ± 6.98	85.64 ± 16.7	70.68 ± 30.2	**98.46** ± 3.06	74.97 ± 31.6
15	99.26 ± 0.53	91.15 ± 4.17	88.54 ± 6.66	85.45 ± 2.88	95.90 ± 1.96	66.02 ± 25.4	92.70 ± 4.77	**100.0** ± 0.00
OA (%)	74.36 ± 0.02	79.05 ± 3.59	78.32 ± 1.74	75.08 ± 1.76	81.79 ± 4.77	71.37 ± 3.86	82.16 ± 3.27	**83.93** ± 0.88
AA (%)	75.64 ± 0.01	81.66 ± 3.18	80.36 ± 1.63	76.64 ± 3.69	85.38 ± 3.84	71.43 ± 5.31	85.90 ± 2.11	**85.54** ± 0.43
Kappa×100	72.29 ± 0.02	77.38 ± 3.86	76.58 ± 1.87	73.10 ± 1.88	80.31 ± 5.15	69.10 ± 4.14	80.72 ± 3.53	**82.65** ± 0.94

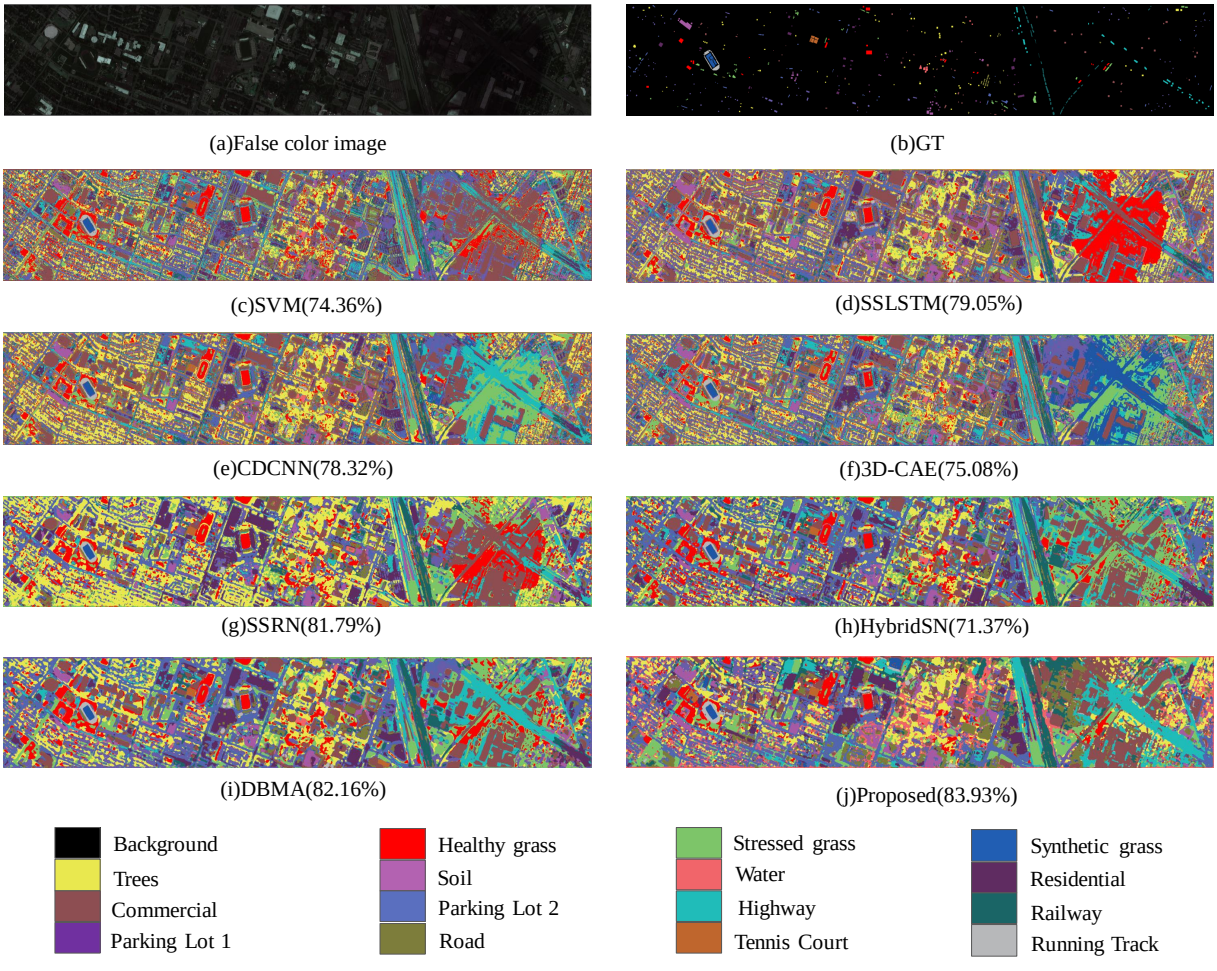


Figure 7. The classification maps of the Houston 2013 with 10 labeled samples.

3.4. Ablation Study

3.4.1. Complementarity between Components

The proposed SSRNet consists of two branches: semi-supervised and self-supervised. Spectral feature shift (S) proposes a random data perturbation strategy in the semi-supervised branch. Two auxiliary tasks are proposed in the self-supervised branch: masked bands reconstruction (R) and spectral order forecast (O). To prove the validity of our proposed

three components, we made exhaustive ablation studies to evaluate different components of the SSRNet on the Indian Pines dataset and the Houston 2013 dataset. Ablation studies include the following:

- SSRNet-S-R-O: The spectral feature shift in the semi-supervised branch is discarded, and two self-supervised auxiliary tasks are discarded;
- SSRNet-S: Only the spectral feature shift in the semi-supervised branch is discarded;
- SSRNet-R: Only the masked bands reconstruction in the self-supervised branch is discarded;
- SSRNet-O: Only the spectral order forecast in the self-supervised branch is discarded;
- SSRNet (ALL): No components are discarded.

We designed two data selection methods, one was 20% unlabeled samples and 10 labeled samples for each class, and the other was 20% unlabeled samples and 20 labeled samples for each class. Table 9 demonstrates that the three components are complementary. Moreover, perfect precision is achieved when integrated with three components (i.e., SSRNet (ALL)).

Table 9. Ablation study of the influence of components in our SSRNet on Indian Pines and Houston 2013.

Datasets	Methods	OA (L = 10) (%)	OA (L = 20) (%)
Indian Pines	SSRNet-S-R-O	72.19	79.26
	SSRNet-S	73.74	82.40
	SSRNet-R	76.37	83.51
	SSRNet-O	81.13	86.86
	SSRNet (ALL)	83.96	87.34
Houston 2013	SSRNet-S-R-O	77.39	86.49
	SSRNet-S	79.33	89.72
	SSRNet-R	81.67	87.88
	SSRNet-O	83.30	91.63
	SSRNet (ALL)	85.54	92.79

3.4.2. Choice of Hyper-Parameters

Figure 8 shows the OA of the spectral feature shift perturbation and the masked bands reconstruction auxiliary task under different hyper-parameter choices on the Indian Pines dataset in which the coefficient μ is used to denote the level of spectral feature shift, and the coefficient κ was used to mean the rank of mask. The adjustment of hyper-parameter has a certain effect of HSI classification precision. Moreover, $\mu = 6$ and $\kappa = 7$ seem to be the best parameters.

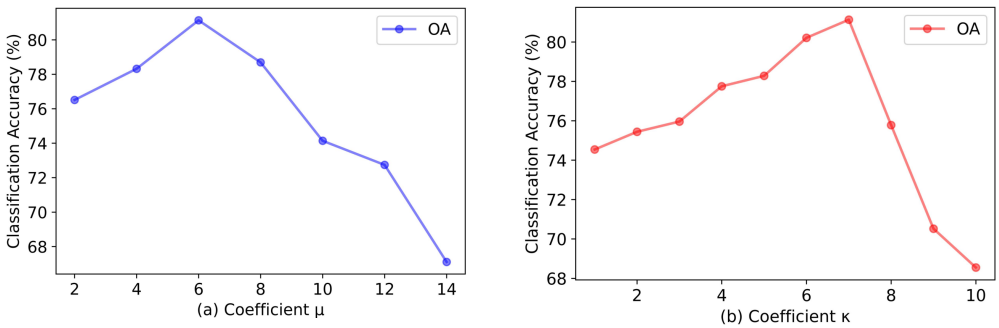


Figure 8. The effects of spectral feature shift and masked bands reconstruction auxiliary task under different hyper-parameter choices on the Indian Pines dataset.

3.4.3. Choice of Patch Size

Table 10 shows the OA of the SSRNet with various patch sizes, which varied from 7×7 to 13×13 with an interval of 2. As the patch size increased, the OA of the Salinas dataset kept increasing. For the other three datasets, the OA began to decline after 11×11,

where they reached the highest OA of 83.96%, 90.99% and 85.54%, respectively. It has been found that the 11 × 11 patch size is most suitable.

Table 10. OA(%) of the SSRNet with various patch sizes.

Patch Size	Indian Pines	PaviaU	Salinas	Houston 2013
7 × 7	81.13	86.45	93.78	83.30
9 × 9	82.50	88.56	94.57	83.32
11 × 11	83.96	90.99	94.58	85.54
13 × 13	83.34	90.55	94.91	85.18

3.5. Investigation on Running Time

The whole training time and testing time of our SSRNet and other methods are reported in Tables 11–14. The SVM consumed less training and testing time than the deep-learning-based methods because the deep-learning-based methods have more parameters and larger input feature maps generally. Moreover, since our proposed SSRNet aims to improve learning performance by using fewer labeled samples and a large number of unlabeled samples, it requires more training time, but testing time still has one advantage compared with other methods. Considering the classification accuracy, our proposed SSRNet is competitive.

Table 11. Training and testing time on the Indian Pines dataset.

Dataset	Method	Training Times (s)	Test Times (s)
Indian Pines	SVM [7]	3.12	0.88
	CDCNN [44]	13.28	1.89
	3DCAE [37]	15.29	1.79
	SSRN [40]	56.84	11.09
	SSLSTM [41]	65.35	6.55
	HybridSN [43]	4.48	0.85
	DBMA [42]	85.38	13.42
	Proposed	211.84	4.33

Table 12. Training and testing time on the University of Pavia dataset.

Dataset	Method	Training Times (s)	Test Times (s)
PaviaU	SVM [7]	1.24	2.95
	CDCNN [44]	9.53	9.27
	3DCAE [37]	21.69	7.52
	SSRN [40]	43.87	23.77
	SSLSTM [41]	41.26	31.60
	HybridSN [43]	4.26	3.67
	DBMA [42]	48.31	32.06
	Proposed	624.84	17.48

Table 13. Training and testing time on the Salinas dataset.

Dataset	Method	Training Times (s)	Test Times (s)
Salinas	SVM [7]	2.78	4.53
	CDCNN [44]	13.12	10.41
	3DCAE [37]	15.24	9.51
	SSRN [40]	73.67	57.02
	SSLSTM [41]	58.41	14.11
	HybridSN [43]	6.15	4.57
	DBMA [42]	83.04	75.16
	Proposed	795.61	24.08

Table 14. Training and testing time of on the Houston 2013 dataset.

Dataset	Method	Training Times (s)	Test Times (s)
Houston 2013	SVM [7]	2.36	1.16
	CDCNN [44]	15.10	3.04
	3DCAE [37]	28.57	2.66
	SSRN [40]	120.44	11.41
	SSLSTM [41]	41.38	7.58
	HybridSN [43]	19.46	1.34
	DBMA [42]	45.65	11.28
	Proposed	257.09	6.06

4. Conclusions

This paper incorporates self-supervised learning in the semi-supervised HSI classification and proposes a unified multi-task SSRNet framework. Especially in the semi-supervised branch, we have designed one type of random data perturbation and a residual feature extraction network to capture the spectral–spatial feature of HSI. We also presented two types of self-supervised auxiliary tasks for the SSRNet. The experiment results demonstrated that the SSRNet performs consistently with satisfactory performance for all four HSI datasets, especially when training samples are limited. As many unlabeled samples are utilized to improve learning performance, we will explore how to decrease training time in future work.

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

Funding: This work was supported by the National Natural Science Foundation of China (Nos. 61906145, 61771376, 61771380); the Science and Technology Innovation Team in Shaanxi Province of China (Nos. 2020TD-017); the 111 Project, the Foundation of Key Laboratory of Aerospace Science and Industry Group of CASIC, China; the Key Project of Hubei Provincial Natural Science Foundation under Grant 2020CFA001, China.

Data Availability Statement: The Indiana Pines, University of Pavia, and Salinas Valley datasets are available online at https://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (accessed on 3 July 2021). The Houston 2013 dataset is available online at https://hyperspectral.ee.uh.edu/?page_id=459 (accessed on 3 July 2021).

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

References

1. Li, Z.; Huang, L.; He, J. A multiscale deep middle-level feature fusion network for hyperspectral classification. *Remote Sens.* **2019**, *11*, 695. [CrossRef]
2. Awad, M.; Jomaa, I.; Arab, F. Improved capability in stone pine forest mapping and management in Lebanon using hyperspectral CHRIS-Proba data relative to Landsat ETM+. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 725–731. [CrossRef]
3. Ibrahim, A.; Franz, B.; Ahmad, Z.; Healy, R.; Knobelspiesse, K.; Gao, B.C.; Proctor, C.; Zhai, P.W. Atmospheric correction for hyperspectral ocean color retrieval with application to the Hyperspectral Imager for the Coastal Ocean (HICO). *Remote Sens. Environ.* **2018**, *204*, 60–75. [CrossRef]
4. Foglini, F.; Angeletti, L.; Bracchi, V.; Chimienti, G.; Grande, V.; Hansen, I.M.; Meroni, A.N.; Marchese, F.; Mercorella, A.; Prampolini, M.; et al. Underwater Hyperspectral Imaging for seafloor and benthic habitat mapping. In Proceedings of the 2018 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea), Bari, Italy, 8–10 October 2018; pp. 201–205.
5. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 37–78. [CrossRef]
6. Villa, A.; Benediktsson, J.A.; Chanussot, J.; Jutten, C. Hyperspectral image classification with independent component discriminant analysis. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4865–4876. [CrossRef]
7. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]

8. Ghamisi, P.; Maggiori, E.; Li, S.; Souza, R.; Tarablaka, Y.; Moser, G.; De Giorgi, A.; Fang, L.; Chen, Y.; Chi, M.; et al. New frontiers in spectral–spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning. *IEEE Geosci. Remote Sens. Mag.* **2018**, *6*, 10–43. [CrossRef]
9. Benediktsson, J.A.; Palmason, J.A.; Sveinsson, J.R. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 480–491. [CrossRef]
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*. [CrossRef]
11. Makantasis, K.; Karantzalos, K.; Doulamis, A.; Doulamis, N. Deep supervised learning for hyperspectral data classification through convolutional neural networks. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 4959–4962.
12. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep convolutional neural networks for hyperspectral image classification. *J. Sens.* **2015**, *2015*, *12*. [CrossRef]
13. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [CrossRef]
14. Cheng, C.; Li, H.; Peng, J.; Cui, W.; Zhang, L. Hyperspectral image classification via spectral–spatial random patches network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4753–4764. [CrossRef]
15. Dópido, I.; Li, J.; Marpu, P.R.; Plaza, A.; Dias, J.M.B.; Benediktsson, J.A. Semisupervised self-learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4032–4044. [CrossRef]
16. Li, F.; Clausi, D.A.; Xu, L.; Wong, A. ST-IRGS: A region-based self-training algorithm applied to hyperspectral image classification and segmentation. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 3–16. [CrossRef]
17. Wu, Y.; Mu, G.; Qin, C.; Miao, Q.; Ma, W.; Zhang, X. Semi-supervised hyperspectral image classification via spatial-regulated self-training. *Remote Sens.* **2020**, *12*, 159. [CrossRef]
18. He, Z.; Liu, H.; Wang, Y.; Hu, J. Generative adversarial networks-based semi-supervised learning for hyperspectral image classification. *Remote Sens.* **2017**, *9*, 1042. [CrossRef]
19. Feng, J.; Ye, Z.; Li, D.; Liang, Y.; Tang, X.; Zhang, X. Hyperspectral Image Classification Based on Semi-Supervised Dual-Branch Convolutional Autoencoder with Self-Attention. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 1267–1270.
20. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. *IEEE Geosci. Remote Sens. Lett.* **2012**, *10*, 318–322.
21. Camps-Valls, G.; Marsheva, T.V.B.; Zhou, D. Semi-supervised graph-based hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3044–3054. [CrossRef]
22. De Morsier, F.; Borgeaud, M.; Gass, V.; Thiran, J.P.; Tuia, D. Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3410–3420. [CrossRef]
23. Ding, Y.; Zhao, X.; Zhang, Z.; Cai, W.; Yang, N.; Zhan, Y. Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–12. [CrossRef]
24. Sun, Q.; Liu, X.; Bourennane, S. Unsupervised Multi-Level Feature Extraction for Improvement of Hyperspectral Classification. *Remote Sens.* **2021**, *13*, 1602. [CrossRef]
25. Zhao, B.; Ulfarsson, M.O.; Sveinsson, J.R.; Chanussot, J. Unsupervised and supervised feature extraction methods for hyperspectral images based on mixtures of factor analyzers. *Remote Sens.* **2020**, *12*, 1179. [CrossRef]
26. Zhu, M.; Fan, J.; Yang, Q.; Chen, T. SC-EADNet: A Self-supervised Contrastive Efficient Asymmetric Dilated Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–17. [CrossRef]
27. Yue, J.; Fang, L.; Rahmani, H.; Ghamisi, P. Self-supervised learning with adaptive distillation for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–13. [CrossRef]
28. Miyato, T.; Maeda, S.i.; Koyama, M.; Ishii, S. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 1979–1993. [CrossRef] [PubMed]
29. Tarvainen, A.; Valpola, H. Mean-teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Adv. Neural Inf. Process. Syst.* **2017**, *30* , 1195–1204.
30. Wang, X.; Kihara, D.; Luo, J.; Qi, G.J. Enaet: Self-trained ensemble autoencoding transformations for semi-supervised learning. *arXiv* **2019**, arXiv:1911.09265.
31. Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; Raffel, C.A. Mixmatch: A holistic approach to semi-supervised learning. *Adv. Neural Inf. Process. Syst.* **2019**, *32* , 5050–5060.
32. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107. [CrossRef]
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
34. Laine, S.; Aila, T. Temporal ensembling for semi-supervised learning.*arXiv* **2016**, arXiv:1610.02242.
35. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

36. Tao, C.; Pan, H.; Li, Y.; Zou, Z. Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2438–2442.
37. Mei, S.; Ji, J.; Geng, Y.; Zhang, Z.; Li, X.; Du, Q. Unsupervised spatial–spectral feature learning by 3D convolutional autoencoder for hyperspectral classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6808–6820. [CrossRef]
38. Liu, L.; Wang, Y.; Peng, J.; Zhang, L.; Zhang, B.; Cao, Y. Latent relationship guided stacked sparse autoencoder for hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 3711–3725. [CrossRef]
39. Liu, B.; Yu, A.; Yu, X.; Wang, R.; Gao, K.; Guo, W. Deep multiview learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 7758–7772. [CrossRef]
40. Zhong, Z.; Li, J.; Luo, Z.; Chapman, M. Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 847–858. [CrossRef]
41. Zhou, F.; Hang, R.; Liu, Q.; Yuan, X. Hyperspectral image classification using spectral–spatial LSTMs. *Neurocomputing* **2019**, *328*, 39–47. [CrossRef]
42. Ma, W.; Yang, Q.; Wu, Y.; Zhao, W.; Zhang, X. Double-branch multi-attention mechanism network for hyperspectral image classification. *Remote Sens.* **2019**, *11*, 1307. [CrossRef]
43. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 277–281. [CrossRef]
44. Lee, H.; Kwon, H. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [CrossRef]
45. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.