Article

Encoding Spectral-Spatial Features for Hyperspectral Image Classification in the Satellite Internet of Things System

Ning Lv [1], Zhen Han [1], Chen Chen [2], Yijia Feng [3], Tao Su [4], Sotirios Goudos [5] and Shaohua Wan [6,*]

[1] The School of Electronic Engineering, Xidian University, Xi'an 710071, China; nlv@mail.xidian.edu.cn (N.L.); hanzhen@stu.xidian.edu.cn (Z.H.)

[2] State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China; cc2000@mail.xidian.edu.cn

[3] School of Electronics Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; fengyijia2015@sjtu.edu.cn

[4] National Laboratory of Radar Signal Processing of School of Electronic Engineering, Xidian University, Xi'an 710071, China; sutao@xidian.edu.cn

[5] Department of Physics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; sgoudo@physics.auth.gr

[6] School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China

* Correspondence: shwanhust@zuel.edu.cn

Abstract: Hyperspectral image classification is essential for satellite Internet of Things (IoT) to build a large scale land-cover surveillance system. After acquiring real-time land-cover information, the edge of the network transmits all the hyperspectral images by satellites with low-latency and high-efficiency to the cloud computing center, which are provided by satellite IoT. A gigantic amount of remote sensing data bring challenges to the storage and processing capacity of traditional satellite systems. When hyperspectral images are used in annotation of land-cover application, data dimension reduction for classifier efficiency often leads to the decrease of classifier accuracy, especially the region to be annotated consists of natural landform and artificial structure. This paper proposes encoding spectral-spatial features for hyperspectral image classification in the satellite Internet of Things system to extract features effectively, namely attribute profile stacked autoencoder (AP-SAE). Firstly, extended morphology attribute profiles EMAP is used to obtain spatial features of different attribute scales. Secondly, AP-SAE is used to extract spectral features with similar spatial attributes. In this stage the program can learn feature mappings, on which the pixels from the same land-cover class are mapped as closely as possible and the pixels from different land-cover categories are separated by a large margin. Finally, the program trains an effective classifier by using the network of the AP-SAE. Experimental results on three widely-used hyperspectral image (HSI) datasets and comprehensive comparisons with existing methods demonstrate that our proposed method can be used effectively in hyperspectral image classification.

Keywords: hyperspectral image; attribute profile; deep features

1. Introduction

The emergence of Satellite Internet of Things(IoT) system, which means combining various information sensor equipments with network into a huge network through satellite communication, has a profound impact on processing. Today, with the emergence of new acquisition platforms, smaller and more efficient sensors, and edge computing [1], remote sensing technology is once again on the edge of major technological innovation. Traditionally, remote sensing was a subject of aerial surveying and mapping, geographic information systems, and earth observation, but recent developments have shifted it to the direction of satellite Internet of Things. Ideally, the continuous streaming data from the interconnected devices on the aggregation platform will paint a vivid picture

of the world people live in. However, the real world is ever-changing with an enormous amount of details, but the capacity of the remote sensing system is limited. Facing the large amount of hyperspectral data and time-consuming data transmission, computing or caching the data at the edge can effectively reduce the amount of transmission [2], and the satellite Internet of Things can solve the latency and bandwidth issues in the data transmission process. The satellite IoT is shown in Figure 1. First, the hyperspectral data is collected through satellites, and the collected data is processed through multi-access edge computing, after which the results are sent to ground. Finally, the data is analyzed through data statistics and post-processing to realize data monitoring. As an aspect of the application of the research in this paper, the restrictions of on orbit satellite hyperspectral application can be resolved to a certain extent, and lay a foundation for the subsequent research on satellite IoT as well as other hyperspectral images applications.
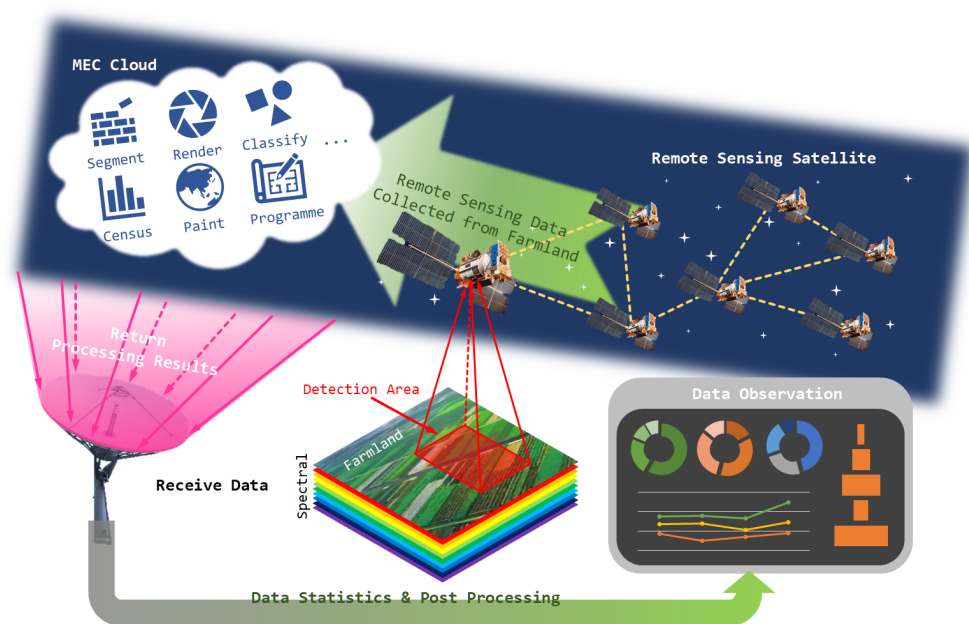

Figure 1. The satellite Internet of Things system.

With a prominent role in hyperspectral image classification, which is the core part of edge computing process, attribute profile (AP) [3] can use available attributes that can be calculated based on region to realize multi-scale analysis of images. AP is considered a multi-scale analysis tool, which can filter the connected components of gray image rather than single pixel executed by morphological attributes. In addition, in the case of limited sample numbers, high-quality samples for classifiers can also be generated by AP-based algorithms. Due to the high dimension of hyperspectral image data, the dimension reduction process before attribute filtering is common in hyperspectral images, which often leads to the loss of spectral information. Stacked filtered images are called extended attribute profiles (EAPs). As shown in [3–5], spatial information of connected region at different scales can be modeled by APs. Therefore, the multi-level spatial features of images can be created by applying APs in sequence, which make APs an effective spatial feature of hyperspectral data. As the [6,7] show, when EAPs combined with original spectral data is used as the input samples of the network, the extracted features of the network are better for classification, reflecting the great cooperation potential of combining EAP and deep learning. Also, seeing that images can be processed based on different attributes and thresholds which first can be calculated based on the connected components, AP can be used as a flexible tool. The traditional thresholds are set arbitrarily, but the tuning of the parameters of the attribute filter is rarely studied. In [8], an automatic feature selection method is proposed to tune the thresholds of attribute filters. Dalla [9] proves that the

two attributes (area and standard deviation) using automatic methods are separated from manual methods. Using the algorithm in this paper, it is simpler to obtain the threshold when the attribute area is considered.

There are many architectures to perform classification-related tasks, of which Autoencoder (AE) [10], as an unsupervised learning model, holds one of the most dominant positions. Chen et al. [11] introduce autoencoders into HSI data classification. Traditional research on AE in HSI classification tended to select raw spectral data combined with image patches as the input of the AE network to learn the spatial-spectral features. The extraction effect of depth features has a great impact on the classification accuracy [12]; effective feature representation can improve the efficiency of classifier [13]. Lauzon [14] and Lin [15] proposed that in those image patches the spatial information of the center pixel is represented by all the pixels in the region. Before getting the image patches, since the dimension of the raw HSI data is high, techniques reducing this dimensionality can be beneficial [16]. However, traditional methods of dimension reduction such as Principal Component Analysis (PCA)[17], Independent Component Analysis (ICA)[18], etc., tend to cause the loss of spectrum information, further leading to the decline of classification accuracy. Cavallaro [6] demonstrated that after encoding the raw spectral data, features can be classified more effectively.

The classification accuracy can be improved by the pre-training network obtained by the AE. In addition, the feature of its own dimension reduction coding contributes to reduce the dimension of hyperspectral images, which can further improve the performance of hyperspectral image classification. Besides, the selection of the parameters of the attribute filters is a major issue when the profiles are generated. There are related researches of this issue that can be found in [6,19], which are time-consuming and difficult to handle. This paper proposes to choose the strategy of selecting thresholds for attribute filters to construct area attribute profiles and then encoding APs with autoencoders for HSI classification. In this method, we focus on the parameter selection of attribute profiles to generate APs and process of encoding by autoencoder. The spatial-spectral features are extracted by EAPs, and in combination with the deep features learning by autoencoder, can acquire more effective features for classification. The innovative framework proposed in this paper can be introduced to other applications, such as the Internet of Vehicles [20].

Compared with state-of-the-art, the main contribution of this study can be summarized as follows:

(1) Spatial spectrum feature extraction. The space frequency characteristics of the joint spectral information and spatial information are used to solve the problems of "same spectral foreign matter" and "Homo object heterospectral" in hyperspectral data. The spatial information of hyperspectral data is extracted based on EMAP in this paper, leading to the full and comprehensive spatial features of hyperspectral images extraction.

(2) Multi-feature fusion. A multi feature hyperspectral image classification algorithm based on the fusion of depth feature and spatial spectrum feature is proposed. The stack autoencoder is selected to extract the depth feature from the training samples.

2. Related Work

The introduction of AP aims to make full use of the spatial information in hyperspectral images, but spatial features have limited ability to represent hyperspectral images, so it is necessary to select various features fused to improve the classification accuracy of hyperspectral images, such as AE, the structure adopted by this paper.

2.1. Attribute Profile

In order to alleviate the problems of "same object with different spectrum" and "same spectrum foreign object" in hyperspectral image classification, and reduce the probability of misclassification of edge pixels, spatial features are introduced into hyperspectral image classification features. In order to make full use of the spatial information in hyper-

spectral images, AP is used in this paper to extract the spatial information in multi-scale. The concept of AP is based on morphology profiles (MP), which is constructed based on the repeated use of openings and closings by reconstruction with a structuring element (SE)[21]. The MP has some limitations because of SE's properties; to overcome these limitations, morphological AP has been proposed. AP can analyze many geometric attributes such as area, standard deviation, and the diagonal of the box bounding the regions, and in this way various spatial information can be obtained according to different attributes.

More specifically, APs rely on morphological attribute filters (AFs), since an AP is achieved by using AFs with a set of thresholds [3]. AFs process connected components either by keeping or merging them. The decision on the AFs to be performed on each region is given by the result of the threshold that evaluates if a given attribute which is computed on a connected component is greater/lower than the arbitrary reference value [22]. If the comparsion result is not verified, then the region is merged to the adjacent region having a closer gray-level value (either greater or equal to the one of the evaluated region). In general, features of the connected component on which the AFs are applied are compared to the given threshold.

The set of thresholds can be set manually or predicted by algorithm. The thresholds are calculated manually based on the statistics and selected in a trial-and-error way [3,19], while the predicates are calculated automatically according to value of attributes [8]. The classification accuracy obtained from automaic prediction may be slightly lost, but automatic method was chosen as its universality in satellite applications. The *predicates* are set to represent a set of thresholds predicated by values of image attribute in this paper. More formally, given a set of predicates of length L in order ($P_{\lambda_j} \subseteq P_{\lambda_k}, j \leq k$), $P_\lambda : \{P_{\lambda_i}\}(i = 1, ..., L, \lambda_i = 0, ..., n)$, let $\phi^{P_{\lambda_i}}$ and $\gamma^{P_{\lambda_i}}$ denote the attribute thinning and thickening operation respectively.

An AP of a gray image is defined as in (1),

$$AP(f) = \{\underbrace{\phi^{P_{\lambda_L}}(f), \phi^{P_{\lambda_{L-1}}}(f), ..., \phi^{P_{\lambda_1}}(f)}_{\text{thickening profile}}, f,$$
$$\underbrace{\gamma^{P_{\lambda_1}}(f), ..., \gamma^{P_{\lambda_{L-1}}}(f), \gamma^{P_{\lambda_L}}(f)}_{\text{thinning profile}}\} \tag{1}$$

where the f represents the original gray image, P_{λ_L} represents different predicates, $\phi^{P_{\lambda_L}}(f)$ represents image after thickening operation with the predicate of P_{λ_L}, $\gamma^{P_{\lambda_L}}(f)$ represents image after thinning operation with the predicate of P_{λ_L}, respectively. It is possible to note how the sequence of thinning transformations is taken considering the sequence of predicates in increasing order, while thickening transformations refer to decreasing order. That is to say, progressively strict criteria leading to progressively coarse images. When λ_i = 0, $\phi^{P_{\lambda_i}}(f) = \gamma^{P_{\lambda_i}}(f) = P_{\lambda_i}(f) = f$.

Figure 2 shows an example of an AP formed by attribute filtering on one of the principal components (PC) after PCA operation on hyperspectral data. Different images can be obtained by using different predicates on the original PC. Therefore, an AP is a stack of thickening and thinning profiles. The original image f can be regarded as the level zero of both the thickening and thinning profiles. It's obvious that given the original image f as input, after the attribute filtering, there are $2L + 1$ output images as APs. In order to expand AP to the hyperspectral image spatial information extraction field, people proposed the concept of EAP. EAP is extracted on the first m principle components (PCs) transformed from HSI data. The attribute filtered PCs construct extended APs (EAPs). More formally, let g indicates m PCs, the process of generating an EAP can be formalized as in (2).
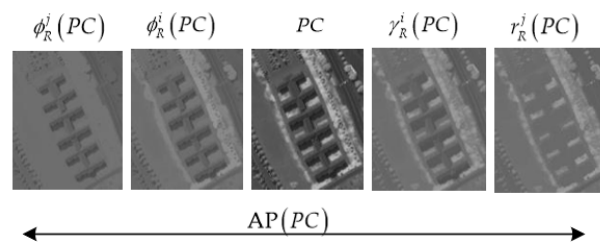
$\phi_R^j(PC)$ $\phi_R^i(PC)$ PC $\gamma_R^i(PC)$ $r_R^j(PC)$

$AP(PC)$

Figure 2. Using attribute filtering to build AP samples for a single PC.

$$EAP(g) = \{AP(g_1), AP(g_2), ..., AP(g_m)\} \tag{2}$$

When we use two or more attributes, we can get EMAPs. Assuming that k attributes are selected, EMAP can be expressed as formula (3).

$$EMAP(g) = \{EAP_{A_1}(g), EAP'_{A_2}(g), ..., EAP'_{A_k}(g)\} \tag{3}$$

where EAP_{A_i} is an EAP built with a set of predicates evaluating the attribute A_i and $EAP' = EAP\backslash\{g_i\}_{i=1,...,m}$. In order to avoid redundancy information, the original component $\{g_i\}$ of EAP is removed.

2.2. Autoencoder and Classifier

There are many methods proposed in remote sensing images classification, but regarding the lack of labeled samples, the supervised methods and semi-supervised methods are not suitable for hyperspectral images classification. Therefore, the unsupervised methods are adopted in this paper, among which SAE performs well, which is a deep learning network structure for hyperspectral image classification in common use. The most commonly reported paradigm for classification of autoencoders consists of unsupervised pre-training, followed by supervised fine-tune and ends with its' classification often by logistic classifier or softmax classifier. The typical autoencoder is a three-layered network, consist of an input layer, a hidden layer and an output layer, it aims to minimize the reconstruction error and then learn a network which can learn deep features of the input data. For this reason, it encodes the input data to get the feature data, next decodes the feature data to obtain the reconstruction data, then defines the loss function and optimizes the function until the network training finishes.

The encoding process from the input layer to the hidden layer is a linear combination with a nonlinear activation function. Similarity, the decoding process from hidden layer to the output layer is still a linear combination with a nonlinear activation function. Let x, h, z represent the input data, the output data of encoding, and the output data of decoding, respectively; these processes can be formalized as shown in (4) and (5) below.

$$h = f(W_h x) + b_h \tag{4}$$

$$z = f(W_z x) + b_z \tag{5}$$

where W_h and b_h are the encoding weight matrix and bias, W_z and b_z are the decoding weight matrix and bias, $f(\cdot)$ indicates the nonlinear activation function. To expand the unsaturated region of sigmoid activation function, we use the parametric sigmoid which allows some flexibility in network training in this paper. The parametric sigmoid function is defined as (6) [23]:

$$\mathcal{F}_{PSigmoid}(x) = \frac{\alpha}{1 + e^{-\beta(x-\gamma)}} \tag{6}$$

where x is the input, α, β and γ are the parameters and/or hyper-parameters which have been kept either trainable or fixed under different setting scenarios. Keeping α equal to 1, $F_{PSigmoid}(x) \in [0,1]$. The α, β and γ in this paper are hyper-parameters. As a improved function of Sigmoid, the introduction of parametric sigmoid function makes it easier for the

model to learn the training dataset irrespective of easy or hard examples. Besides, in order to simplify the training processing of autoencoder, tied weights strategy is employed.

There are many distance metrics to evaluate the performance of the reconstruction from z to x, such as mean squared error (MSE) cross entropy. In this paper, MSE is chosen as cost function. Our goal is to minimize the cost function defined as:

$$J(W,b) = \frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} \|x - z\|^2 \tag{7}$$

where N_{tr} indicates the number of the training samples. Equation (7) can be solved by minibatch stochastic gradient descent (MSGD) method.

The parameter matrix of the autoencoder has been optimized to minimize reconstruction errors.There are many distance measurement functions to evaluate the reconstruction performance of, such as Mean Squared Error (MSE) function. In the autoencoder, MSE is generally selected as the loss function. The goal of training the autoencoder is to minimize the loss function defined as Formula (7). N represents the number of input samples.

After pre-training, the output layer of autoencoder will be replaced by a logistic regression(LR) layer. Since LR works in a supervised manner, the input of the network should be the input data and its label information, and the label is the output of the network. In more detail, the sigmoid function is still the activation function in LR layer, h is the encoding result and the input data of LR layer, the probability of h belongs to *cth* class can be defined as:

$$P(h = c|h, W_h, b_h) = softmax(W_h h + b_h) \tag{8}$$

The output of LR is between [0,1]. And the cost function is:

$$J(l,h) = \frac{1}{N} \sum_{i=1}^{N} \|l - h\|^2 \tag{9}$$

where N is the number of input samples and l is the number of true label. Dalla Equation (9) can also be solved by stochastic gradient descent (MSGD) method.

3. Proposed Method for Spectral-Spatial Features Encoding

Our proposed framework is shown in Figure 3. It contains two learning stages, which are optimized step by step for different objectives: the former is the training of feature extractors, and the latter is the joint training for hyperspectral image classifiers. At the first stage, the program imposes a similarity regularization on each hidden layer of SAE to learn a discriminative feature space in which homogeneous pixels are mapped closely and inhomogeneous pixels are mapped further separately. At the second stage, the program acquires an effective classifier by replacing the reconstruction layer with softmax layer. The output is class labels of pixels in HSI.

There is only one hidden layer in AE, while the hyperspectral data in this paper contains many bands. If AE is chosen to transfer high dimensions data as input through neurons directly, the difficulty of network training will be increased, making the network difficult to converge, and reducing the accuracy of feature learning. The stacked autoencoder(SAE) increases the number of hidden layers on the basis of AE, whose effect is equivalent to superposition of several AEs. SAE can fit the nonlinear relationship in the spectral information of hyperspectral image well, so as to achieve efficient representation of background image, and the parameters can be self-adaptive by learning the image information, It is a deep learning network structure commonly used in hyperspectral image classification.
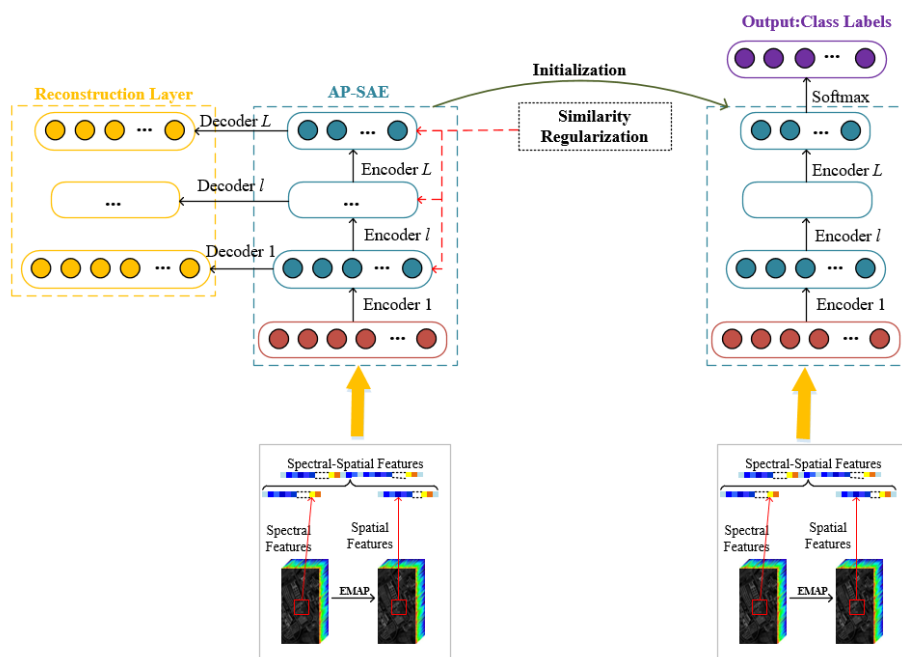
Figure 3. Outline of the proposed pixel classification strategy.

The most commonly reported paradigm for classification of autoencoders consists of unsupervised pre-training, followed by supervised fine-tune and ends with its classification often by logistic classifier or softmax classifier. The typical autoencoder is a three-layered network, consist of an input layer, a hidden layer and an output layer, it aims to minimize the reconstruction error and then learn a network which can learn deep features of the input data. For this reason, it encodes the input data to get the feature data, next decodes the feature data to obtain the reconstruction data, then defines the loss function and optimizes the function until the network training finished. The outline of the proposed classification strategy is shown in Figure 4.
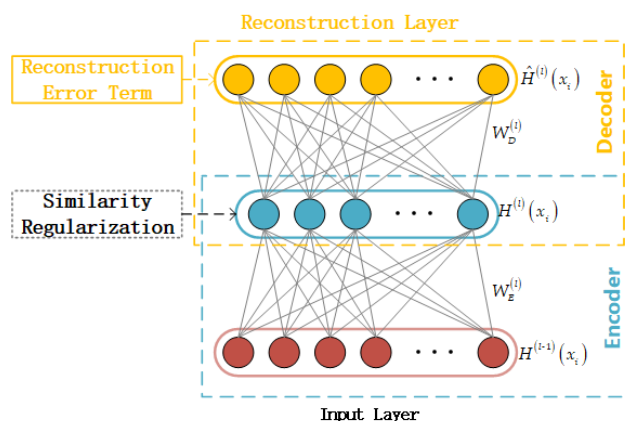

Figure 4. Illustration of the *l*th AE training in stage 1. The AP-SAE can be optimized layer-wisely by minimizing the reconstruction error of AE with similarity regularization.

The principle of AP-SAE proposed is shown as Figure 5. Suppose that the proposed AP-SAE consists of L stacked AE, and the hidden layer dimension of lth AP-SAE is $d^{(l)}$, where $l = 1, 2, ..., L$. Let $\hat{X} = (x_i)_{i=1}^{N} \in R^{d \times N}$ denote the training set, where $x_i \in R^d$

is the spectral-spatial feature of the *i*th training sample and *N* is the total number of training samples. So the *l*th AP-MAE has two parts: one part is an encoder to learn the feature mapping matrix, another part is a decoder to restore the input of AP-SAE with Sam constraint. For the *l*th AP-SAE, let $H^{(l)}(x_i)$ be the weights of the hidden layer, $I^{(l)}(x_i)$ be the data fed into the *l*th AP-SAE which is equivalent to $Y^{(l-1)}(x_i)$ ($I^{(l)}(x_i) = x_i$ when $l = 1$), and $Y^{(l)}(x_i)$ be the restore of input $I^{(l)}(x_i)$. The process is formulated as

$$H^{(l)}(x_i) = f(W_E^{(l)} Y^{(l-1)}(x_i) + biase_E^{(l)}) \tag{10}$$

$$Y^{(l)}(x_i) = f(W_D^{(l)} Y^{(l)}(x_i) + biase_D^{(l)}) \tag{11}$$

where $W_E^{(l)} \in R^{d^{(l)} \times d^{(l-1)}}$ is the weights matrix vector and $biase_E^{(l)} \in R^{d^{(l)}}$ is the bias vector of the encoder to be learned in *l*th AP-SAE. The $W_D^{(l)} \in R^{d^{(l-1)} \times d^{(l)}}$ is the weights matrix vector and $biase_D^{(l)} \in R^{d^{(l-1)}}$ is the bias vector of the decoder to be learned in *l*th AE. $f(\cdot)$ is the activation function, which uses parameter sigmoid in this method. Besides, in the case of difficult in the training of the stacked autoencoder, tied weights strategy is employed.
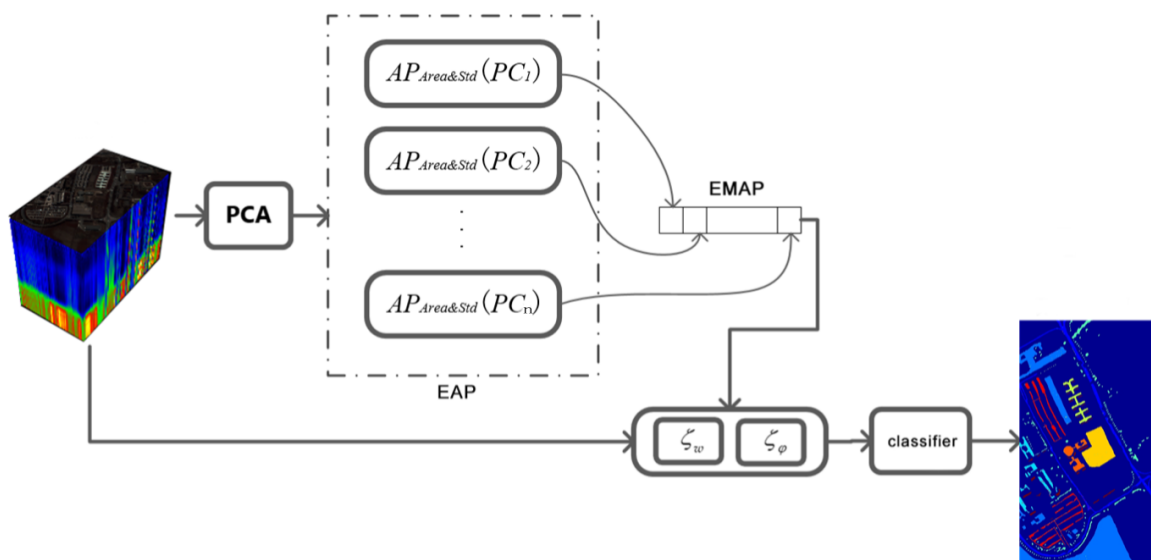

Figure 5. The principle of AP-SAE.

In varied image classification and annotation applications, there are many index or criteria to evaluate the performance of the approximation from input of encoder to output of decoder, such as mean squared error (MSE) or cross-entropy. To achieve fast convergence, each of the *l*th AP-SAE sub network is trained by using the following objective function:

$$J(W_E^{(l)}, biase_E^{(l)}, W_D^{(l)}, biase_D^{(l)}) = \min(L(I^{(l)}, Y^{(l)}) + \lambda \Psi(I^{(l)}, Y^{(l)})) \tag{12}$$

where λ is trade off parameters, $L(I^{(l)}, Y^{(l)})$ represents the reconstruction error term, and $\Psi(I^{(l)}, Y^{(l)})$ represents the discriminant regularization term.

The first item in (12) is the reconstruction cost between the input data and its corresponding reconstruction data, which is calculated by

$$\begin{aligned} L\left(I^{(l)}, Y^{(l)}\right) &= \frac{1}{2} \sum_{i=1}^{N} \left\| I^{(l)}(x_i) - Y^{(l)}(x_i) \right\|^2 \\ &= \frac{1}{2} \sum_{i=1}^{N} \left\| H^{(l-1)}(x_i) - f(W_D^{(l)} H^{(l)}(x_i) + biase_D^{(l)}) \right\|^2 \end{aligned} \tag{13}$$

The second item in (12) is the reconstruction cost between input data and its corresponding reconstructed data, which is calculated by

$$
\begin{aligned}
\Psi\left(I^{(l)}, Y^{(l)}\right) &= \arccos \frac{(<I^{(l)}(x_i), Y^{(l)}(x_i)>}{\|I^{(l)}(x_i)\| \times \|Y^{(l)}(x_i)\|)} \\
&= \arccos \left(\frac{\sum\limits_{j=1}^{d^{(l-1)}} I^{(l)}(x_i)^j \cdot Y^{(l)}(x_i)^j}{\left[\sum\limits_{j=1}^{d^{(l-1)}} \left[I^{(l)}(x_i)^j\right]^2\right]^{1/2} \times \left[\sum\limits_{j=1}^{d^{(l-1)}} \left[Y^{(l)}(x_i)^j\right]^2\right]^{1/2}} \right)
\end{aligned}
\tag{14}
$$

We integrate (13), (14) into (12) to obtain the following objective function of AP-SAE:

$$
\begin{aligned}
& J(W_E^{(l)}, bias e_E^{(l)}, W_D^{(l)}, bias e_D^{(l)}) \\
&= \min \left(\begin{array}{l} \frac{1}{2}\sum\limits_{i=1}^{N} \left\| H^{(l-1)}(x_i) - f(W_D^{(l)} H^{(l)}(x_i) + bias e_D^{(l)}) \right\|^2, \\ +\lambda \arccos(< I^{(l)}(x_i), Y^{(l)}(x_i) > / \left\| I^{(l)}(x_i) \right\| \\ \times \left\| Y^{(l)}(x_i) \right\|) \end{array} \right)
\end{aligned}
\tag{15}
$$

By optimizing the objective function in (15), a compact and distinctive low dimensional feature space is obtained to cover the similar spatial context in HSI. The stochastic gradient descent method is used to resolve the Equation (15).

After the pre-training, the output layer of the auto encoder will be replaced by the logistic regression (LR) layer for the purpose of classification. Once all hidden layers of AP-SAE are pre-trained, the network will converted into the second stage of multi class classifier training. The method first integrate a C-way softmax classification layer at the top of the AP-SAE network, and then train the network by minimizing the classification error, where C is the number of land cover classifications. The softmax classifier is characterized by $\left\{ W_E^{(L+1)} \in R^{C \times d^{(L)}}, bias e_E^{(L+1)} \in R^C \right\}$.

For a training sample x_i, let $Y^{L+1}(x_i)$ be the output of softmax classifier and $Y^L(x_i)$ be its input, where $Y^L(x_i)$ is the Lth hidden layer of AP-SAE. The softmax classifier is formulated as

$$
Y^{(L+1)}(x_i) = \varphi\left(W_E^{(L+1)} Y^{(L)}(x_i) + bias e_E^{(L+1)} \right)
\tag{16}
$$

where $\phi(\cdot)$ is the softmax activation function. The objective function is the softmax cross-entropy loss, which is formulated as follows:

$$
J(X, Y) = -\frac{1}{N}\sum\limits_{i=1}^{N} \left\langle y_i, \log\left(Y^{(L+1)}(x_i)\right) \right\rangle
\tag{17}
$$

where $Y = \{y_i\}_{i=1}^{N} \in R^{C \times N}$ is the label set of training set X. $y_i \in R^C$ is the label vector of the ith training sample x_i, in which only one element is 1 and others are zeros.

4. Experiments and Classification Results

4.1. Dataset Description

In this experiment, the performance of the proposed algorithm is evaluated by three hyperspectral images. The first one is Pavia University (610 × 340 pixels) and it has been acquired by the ROSIS-03 sensor with 1.3 m spatial resolution over the city of Pavia, Italy. They possess 115 bands with a spectral coverage ranging from 0.43 μm to 0.86 μm. After the elimination of 12 noisy bands, 103 bands have been left. here are 9 representative categories. Figure 6 shows the false-color image and ground-truth map. The Salinas is the second data set used for HSI classification. It contains 204 spectral bands (removing 20 water absorption bands). The data set contains 512 × 214 pixels with the spatial resolution of 3.7 m. There

are 16 representative categories. Figure 7 shows the false-color image and ground-truth map.The third data set was acquired by the airborne visible/infrared imaging spectrometer. This data set is obtained from an airplane. The size of the hyperspectral image of each channel is 145 pixels × 145 pixels, and it has 220 spectral bands in the wavelength range of 0.4 μm–2.5 μm. The spectral bands are removed by removing the water absorption band. The number is reduced to 200. Figure 8 shows the false-color image and ground-truth map.
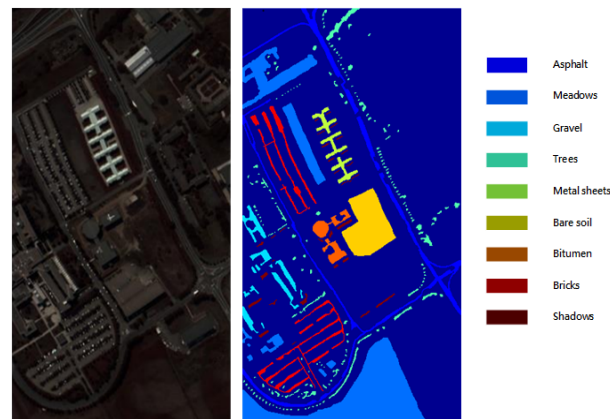
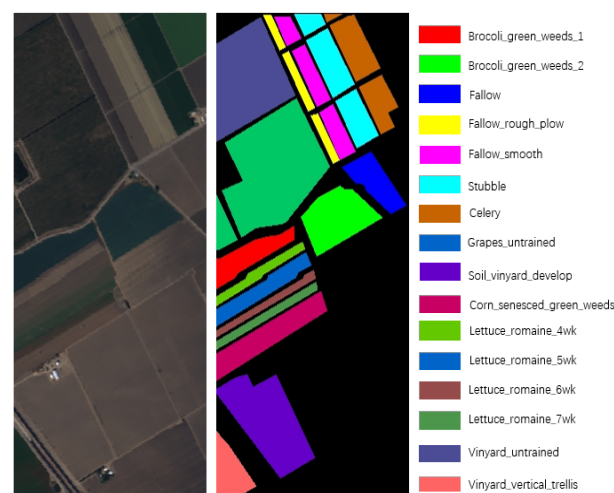**Figure 6.** University of Pavia: false-color image and ground-truth map.

Figure 7. Salinas: false-color image and ground-truth map.
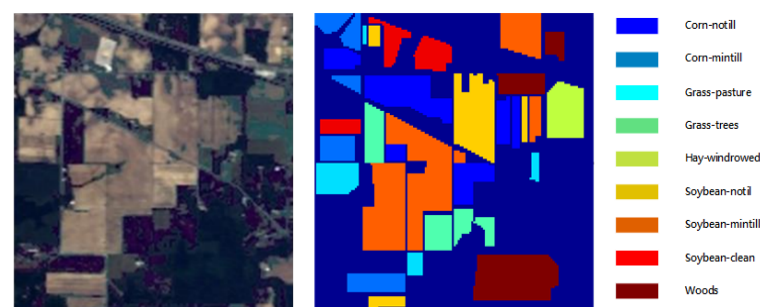

Figure 8. Indian Pines: false-color image and ground-truth map.

4.2. Parameter Setting

In order to achieve the classification of HSI of n bands, we reduce the spectral dimension from n to $r \ll n$; firstly, there are various dimension reduction techniques can be used,

in this paper, principal component analysis (PCA) is chosen due to its widespread use with APs. Secondly, we construct the maxtree for each PCs. Next step is to apply each of r PCs with attribute filters possessing L thresholds which are listed in Table 1, so that the attribute profile of length $c = (2 \times L + 1) \times r$ can be obtained. And then the pixels in the APs are the samples of the rest network.

Parameter of the thresholds is available from Table 1 with area and standard deviation as the attributes. To obtain EMAP features, the principle components containing approximately 99% of the total variance for dataset should be preserved. Then the whole data in EMAPs is normalized, and all of the available labeled set is randomly partitioned into the training set, validation set and test set with a ratio of 5:2:3. As for the number of neurons of the hidden, there is an experiment to analyze the behavior of network and obtain this parameter. The number of hidden layers of AP-SAE is set to 2 (i.e., L = 2) for each data set and each kind of feature. In addition, we set the number of neurons in the first hidden layer to be about 50% of the dimensions of the original input features. Thus, the dimensionality of spectral-spatial feature is 171 for University of Pavia, 255 for Salinas and 285 for Indian Pines, respectively. The number of neurons in the first hidden layer is set to 100 for University of Pavia , 120 for Salinas and 150 for Indian University. The number of neurons in the second hidden layer is set varying from the set of {40,60,80,100} for three data sets.To optimize parameters λ, we set its value varying from the set of {0.001,0.01,0.1}.The classification results measured in terms of OA by using different neuron nodes in the second hidden layer and different input features (spectral feature versus spectral-spatial feature) on three widely-used HSI classification data sets . Finally we obtain the option parameters as listed in Table 2. These parameters are fixed in the following experiments.

Table 1. Attribute and thresholds.

Attribute	Pavia University	Salinas	Indian Pines
Area	55,879, 93,720, 131,561, 169,402	24,174, 45,732, 67,290, 88,848	4660, 8743, 12,827, 16,910
Standard Deviation	14, 26, 39, 52	12, 22, 32, 42	10, 20, 31, 41

Table 2. Optimal parameter setting for different datasets.

dataset	λ	Number of Hidden Layer Neurons	AP-SAE Structure
Pavia University	0.1	80	171-100-80-9
Salinas	0.001	80	255-120-80-16
Indian pines	0.001	40	285-150-40-9

4.3. Ablation Studies

The Table 3 shows the comparison results of hyperspectral image classification using spectral features and spatial spectral features in three datasets. From the comparison results, for the three datasets used in this paper, the classification accuracy of using spatial features is higher than that of using spectral features in OA, AA and kappa coefficients. That is to say, the combination of spatial features and spectral features can effectively improve the classification accuracy of hyperspectral images and get better classification results.

Table 3. Classification results of spectral and spatial features.

Evaluation Criterion	Pavia University		Salinas		Indian Pines	
	Spectral	Spatial	Spectral	Spatial	Spectral	Spatial
OA	0.88	0.982	0.87	0.95	0.77	0.92
AA	0.83	0.98	0.88	0.95	0.71	0.96
Kappa	0.85	0.98	0.85	0.95	0.74	0.92

Figures 9–11 show the classification result map of three datasets, including the real mark map of ground features, the classification result map based on spectral features and the classification result map based on spatial spectral features.
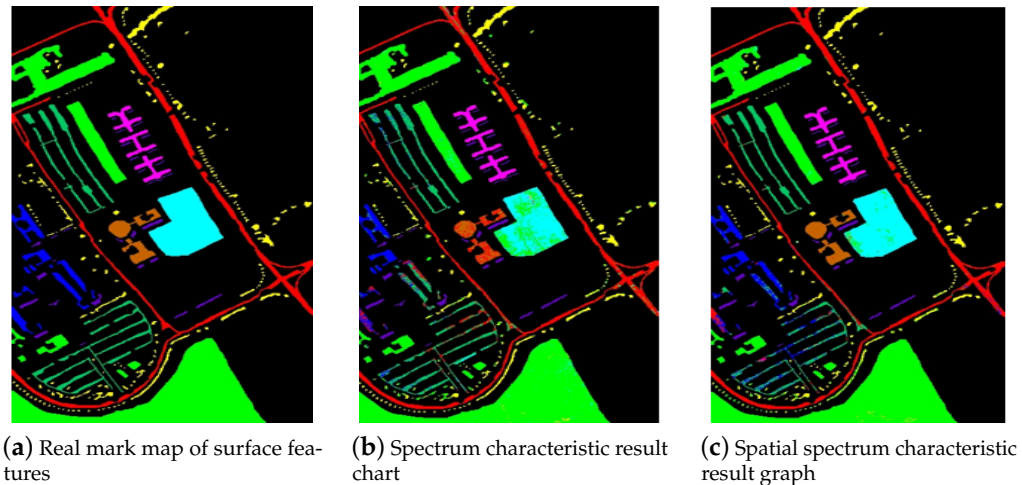

(**a**) Real mark map of surface features

(**b**) Spectrum characteristic result chart

(**c**) Spatial spectrum characteristic result graph

Figure 9. Classification results of Pavia University dataset.

(**a**) Real mark map of surface features

(**b**) Spectrum characteristic result chart

(**c**) Spatial spectrum characteristic result graph
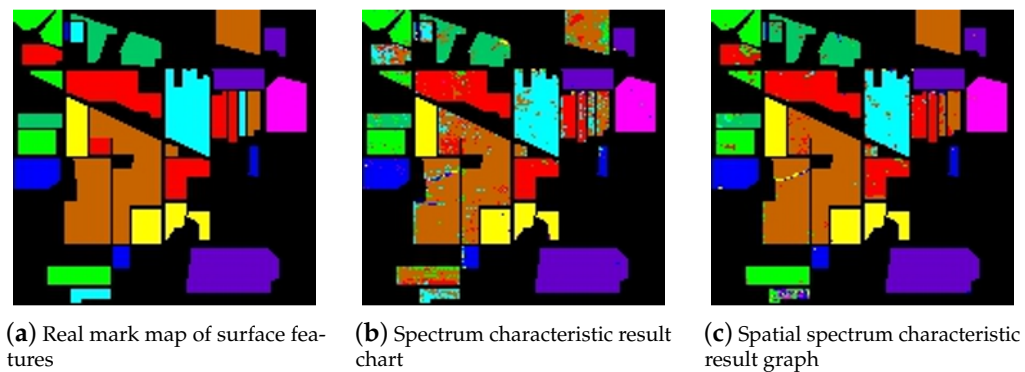
Figure 10. Classification results of Indian Pines dataset.

(**a**) Real mark map of surface features

(**b**) Spectrum characteristic result chart

(**c**) Spatial spectrum characteristic result graph
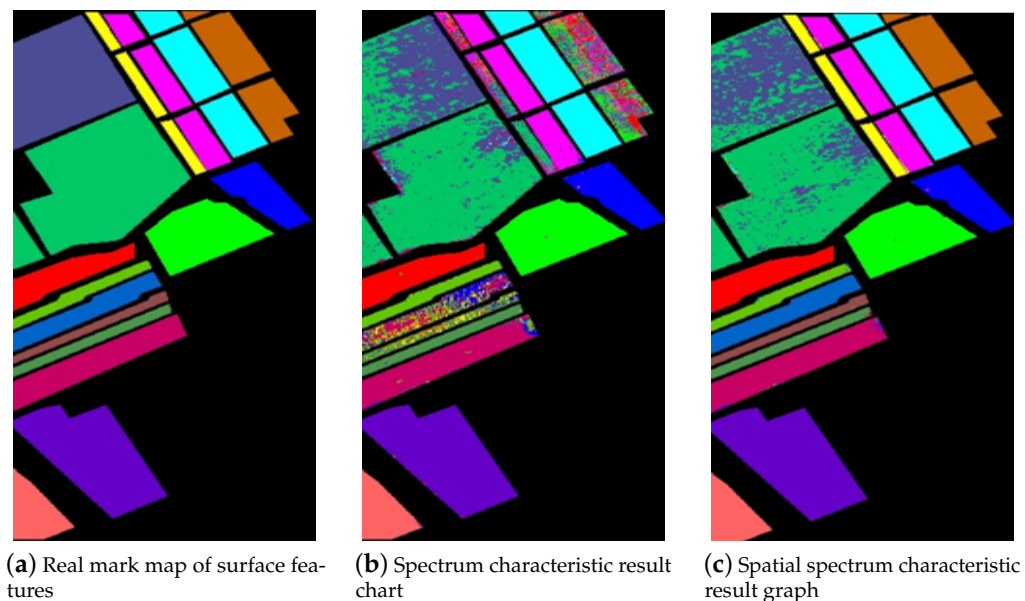
Figure 11. Classification results of Salinas dataset.

In order to evaluate the effect of adding similarity constraints on the classification results, this paper gives a comparison between SAE network structure without similarity constraints and AP-SAE network structure with similarity constraints. Table 4 lists the classification results of this network models under three different datasets under different numbers of neurons. As the experimental results show, adding similarity constraints has a significant effect on improving the classification accuracy.

Table 4. Different neuron numbers use AP-SAE model classification results on different data sets.

Evaluation Criterion	Neurons Number	Pavia University	Salinas	Indian Pines
OA	40	99.22	98.00	96.45
	60	99.11	98.22	94.86
	80	99.28	98.32	95.57
	100	99.14	98.27	95.17
AA	40	98.91	98.83	96.10
	60	98.90	98.88	95.20
	80	99.01	98.91	94.65
	100	98.79	98.80	93.48
Kappa	40	98.97	97.77	95.83
	60	98.82	98.01	93.41
	80	99.05	98.13	94.29
	100	98.86	98.07	93.81

This paper compares the running efficiency of the classification algorithm framework based on SAE and AP-SAE, and the required schedule is shown in Table 5.

Table 5. Running Efficiency of AP-SAE Model on Different Datasets.

Dataset	Model Training Duration(s)		Model Running Duration(s)	
	SAM	AP-SAM	SAM	AP-SAM
Pavia University	2979.4	2953.5	3.6	3.7
Salinas	716.3	751.4	2.2	2.3
Indian pines	4485.8	4395.6	5.9	6.1

4.4. Comparison with State of the Arts

In order to quantitatively evaluate the effectiveness of the AP-SAE framework proposed in this paper, a comparison between some of the latest hyperspectral image classification methods and the method proposed in this paper is essential, with which the effectiveness of the method proposed in this paper to integrate depth features and spatial spectrum features can be verified. The latest hyperspectral image classification methods are Compact and Discriminative Stacked Autoencoder (CDASAE), local binary pattern (LBP)-ELM, 1-D CNN, SVM-random feature selection (RFS) and CNN-pixel-pair feature (PPF) methods. Among them, CDASAE [24] adds discriminant condition constraints and regularized diversity constraints to the SAE structure. LBP-ELM [25] uses LBP to extract local hyperspectral data features, and uses ELM to classify the extracted features. SVM-RFS [26] is to use SVM-based system and RFS to achieve hyperspectral image classification. 1-D CNN [27] is to use CNN to directly extract the spectral information of hyperspectral data to achieve classification. CNN-PPF [28] is to use CNN to learn PPF features of hyperspectral images, where PPF features are obtained based on the information of pixels and their neighboring pixels.

Tables 6–8 record the comparison between the proposed method and the latest hyperspectral classification results for the three datasetS. The results in the table illustrate

that the method proposed in this paper achieves the best results on the three evaluation criteria of OA, AA and Kappa coefficients, that is to say, the model proposed in this paper can effectively improve the classification accuracy of hyperspectral images, which further confirms the effectiveness and superiority of the proposed method.

Table 6. Performance comparion of different thresholds for pavia university.

Evaluation Criterion	AP-SAE	CDA-SAE	SVM-RFS	1-D CNN	CNN-PPF	LBP-ELM
OA	99.28	97.59	96.48	97.59	91.10	92.27
AA	99.01	97.66	91.81	92.92	93.30	96.98
kappa	99.05	96.86	95.48	96.90	88.53	89.89

Table 7. Performance Comparison of Different Thresholds for Salinas.

Evaluation Criterion	AP-SAE	CDA-SAE	SVM-RFS	1-D CNN	CNN-PPF	LBP-ELM
OA	98.32	96.07	93.15	89.28	94.80	92.42
AA	98.91	97.56	96.87	94.83	97.73	96.31
kappa	98.13	96.78	92.35	88.13	94.17	91.55

Table 8. Performance Comparison of Different Thresholds for Indian Pines.

Evaluation Criterion	AP-SAE	CDA-SAE	SVM-RFS	1-D CNN	CNN-PPF	LBP-ELM
OA	96.51	95.81	97.33	89.83	86.44	94.34
AA	96.74	97.38	90.59	93.36	91.58	96.78
kappa	95.90	95.30	85.94	88.65	84.88	93.63

In order to more intuitively show the classification of various methods on different feature categories, Table 9 shows the value of the classification accuracy of the six methods compared in this article on the Pavia University dataset. There are six methods compared in this paper for the Pavia University dataset, which has a total of nine feature categories. As can be seen in the table, the classification accuracy of the most categories is pretty high, while the classification effect of the second category (Asphalt), the third category (Meadows) and the eighth category (Bare soil) perform worse than the other six categories in most methods listed.

Figure 12 shows the line chart of the classification accuracy of each of the six methods compared in this paper for the Pavia University dataset. In the Figure 12, the abscissa represents different feature category numbers. The ordinate represents the classification accuracy of each category, and the accuracy value is displayed as a percentage. As can be seen from the histogram, in terms of the classification accuracy of the second category (Asphalt), the third category (Meadows) and the eighth category(Bare soil), while the other methods underperform, the method proposed in this paper all achieve the maximum classification accuracy.

Table 9. Different methods of various types of accuracy comparison results for the Pavia University dataset.

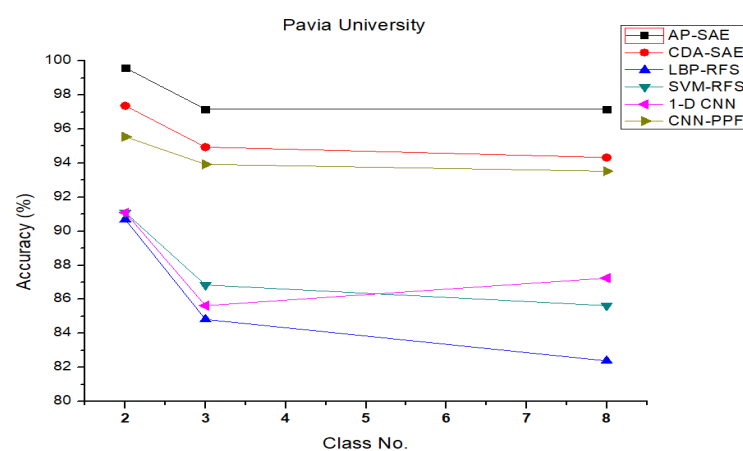| Datasets	AP-SAE	CDA-SAE	SVM-RFS	1-D CNN	CNN-PPF	LBP-ELM
1	98.99	99.39	99.39	99.39	99.39	99.39
2	99.60	97.37	90.69	91.10	91.10	95.55
3	97.17	94.94	84.82	86.84	85.63	93.93
4	98.79	98.18	96.36	95.34	96.96	96.76
5	99.39	99.80	99.39	99.60	99.60	99.80
6	99.39	99.39	94.13	94.13	96.15	98.99
7	98.79	96.96	95.75	94.53	93.52	95.95
8	97.17	94.33	82.39	85.63	87.25	93.52
9	99.60	98.79	99.60	99.60	99.40	99.40

Figure 12. Different methods of various types of accuracy comparison results for the Pavia University dataset.

Table 10 shows the value of the classification accuracy of the six methods compared in this article on the Salinas dataset. There are six methods compared in this paper for the Salinas dataset. As can be seen in the table, the classification accuracy of the most categories is over 95%, but the classification effect of the eighth category (Grapes) and fifteenth category (Vinyard untrained) is below 85% in most methods listed. Especially,the vegetation characteristics of Lettuce romaine are obvious, where contains more samples of Lettuce romaine 5 wk, so the accuracy of the six methods all reach 100.

Figure 13 shows the classification accuracy histogram for the eighth category (Grapes) and fifteenth category (Vinyard untrained) comparing the effect among the six methods. In the figure, the abscissa represents different feature category numbers. The ordinate represents the classification accuracy of each category, which value is displayed as a percentage. As is shown in the histogram, in the 8th category (Grapes) and the 15th category (Vinyard untrained) classification accuracy, compared with other models, the classification accuracy of the proposed AP-SAE model is significantly improved.

Table 10. Different methods of various types of accuracy comparison results for the Salinas dataset.

| Datasets	AP-SAE	CDA-SAE	SVM-RFS	1-D CNN	CNN-PPF	LBP-ELM
1	100.00	100.00	99.58	99.39	100.00	99.39
2	100.00	100.00	100.00	99.37	100.00	100.00
3	100.00	98.95	99.58	96.65	99.58	99.58
4	99.58	99.79	99.79	99.79	99.58	99.58
5	99.58	98.74	98.11	97.07	98.32	98.95
6	100.00	99.58	99.79	99.58	100.00	99.79
7	100.00	99.58	99.79	99.58	100.00	99.79
8	96.86	93.92	84.91	72.33	88.68	84.07
9	100.00	99.16	99.58	99.60	98.32	100.00
10	98.74	98.74	96.44	91.41	98.74	94.97
11	100.00	98.95	98.74	97.69	99.58	96.86
12	100.00	100.00	100.00	100.00	100.00	100.00
13	98.95	93.29	99.17	98.95	99.58	98.32
14	95.81	95.81	98.95	95.18	98.95	97.90
15	94.34	85.11	76.52	76.94	83.65	72.96
16	99.58	99.58	99.58	99.95	99.37	99.16
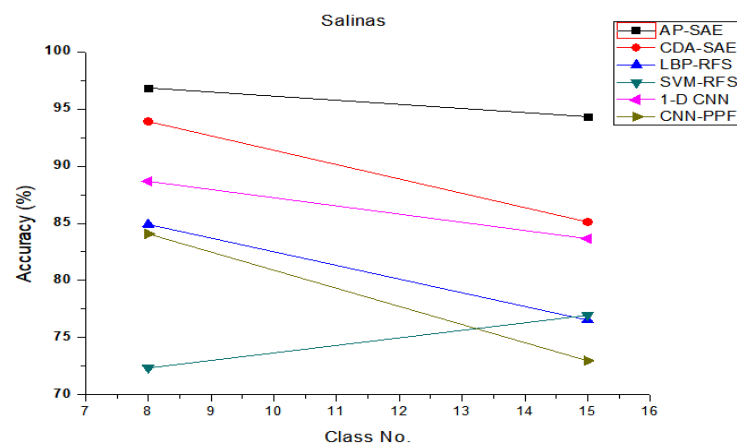
Figure 13. Different methods of various types of accuracy comparison results for the Salinas dataset.

Table 11 shows the value of the classification accuracy of the six methods compared in this article on the Indian Pines dataset. There are nine feature categories of the Indian Pines dataset are selected to compare the effect of the methods listed. As can be seen in the table, in the first category(Corn-notill), the second category(Corn-mintill), the sixth category (Soybean-notill) and the seventh category(Soybean-mintill) the accuracy value of most methods is below 90%.

In particular, the phenomenon where the value of accuracy of several experimental results reaches 100% in Tables 10 and 11, which also appears in some references, such as [24–28] does not mean that the problem is from the experiment itself. As special characteristics of several kinds of surface features listed in the Salinas dataset and Indian Pines dataset, the results reach 100% because of their stable performance in experiments.

Table 11. Different methods of various types of accuracy comparison results for the Indian Pines dataset.

Datasets	AP-SAE	CDA-SAE	SVM-RFS	1-D CNN	CNN-PPF	LBP-ELM
1	93.60	93.39	86.16	88.84	78.72	93.18
2	97.93	96.28	88.43	91.32	85.33	96.90
3	96.90	98.76	96.28	97.73	95.87	98.76
4	98.76	100.00	99.79	100.00	100.00	100.00
5	100.00	100.00	100.00	100.00	99.79	100.00
6	90.29	97.31	90.08	91.74	89.88	96.49
7	97.31	92.36	71.07	78.93	81.61	88.02
8	97.11	99.17	85.74	94.01	95.66	99.17
9	100.00	100.00	98.76	98.97	98.76	100.00

Figure 14 shows the line chart of the classification accuracy of the six methods compared in this article on the first category(Corn-notill), the second category(Corn-mintill), the sixth category (Soybean-notill) and the seventh category(Soybean-mintill) of the Indian Pines dataset. The abscissa in the figure represents the number of different feature categories. The ordinate represents the classification accuracy of each category, and the accuracy value is displayed as a percentage. As can be seen from the histogram, the AP-SAE model proposed in this paper in the first category (Corn-notill), the second category (Corn-mintill), the sixth category (Soybean-notill) and the seventh category (Soybean-mintill) is inferior to the classification effect of other methods in terms of the classification accuracy. While the classification accuracy of the AP-SAE model is lower than others in the sixth category, for the reason that the surface of Soybean-notill or Corn-notill is bare soil, and the characteristics of surface properties are not obvious. On the contrary, there are texture features left by human operation on the surface of Soybean-mintill or Corn-mintill area, which make it easier to extract the corresponding spatial features.
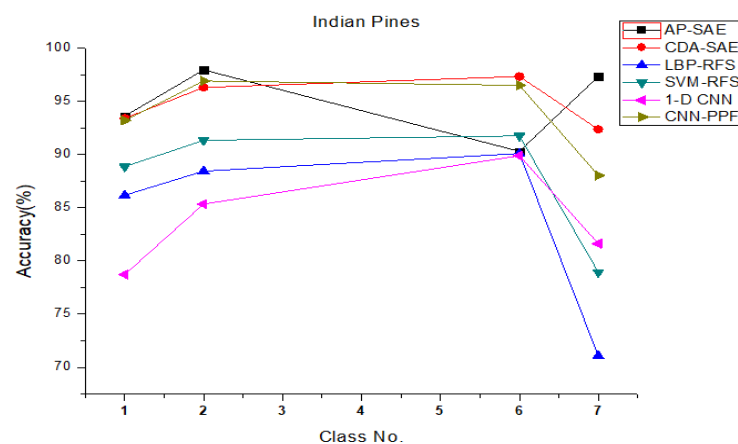
Figure 14. Different methods of various types of accuracy comparison results for the Indian Pines dataset.

Figures 15–17 show the thematic maps. We produced groundcover maps of entire image scenes (including unlabeled pixels). However, to facilitate comparison between methods, with ground truth are shown in these maps. Some areas in the classification maps produced by the proposed AP-SAE are obviously less noisy than those of SVM, ELM and CNN-PPF, e.g., the regions of Bare soil in Figure 15.
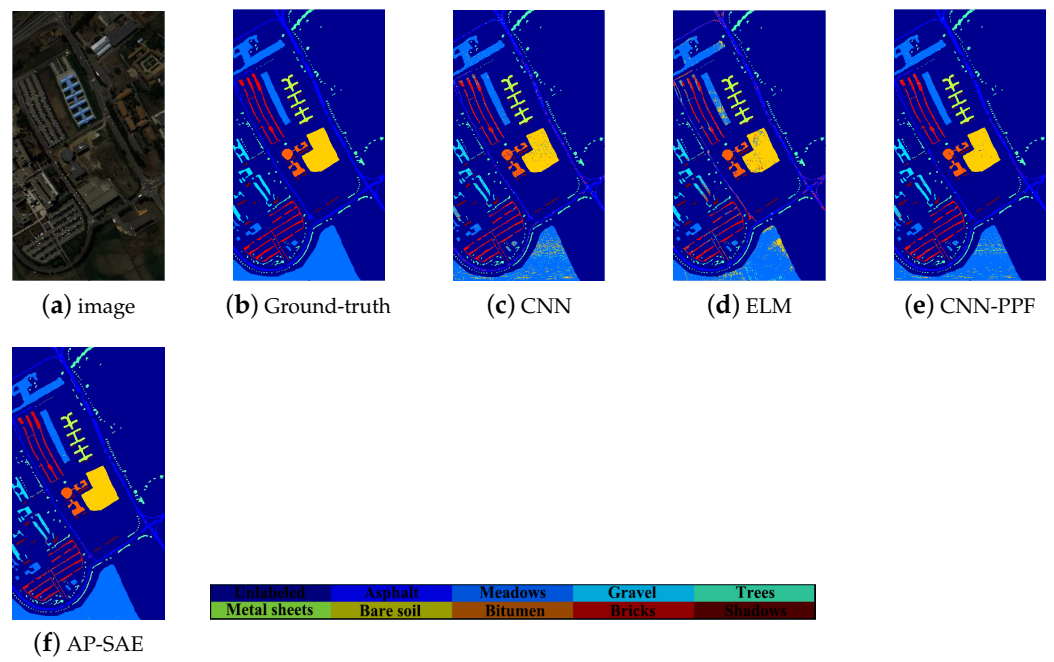
(**a**) image   (**b**) Ground-truth   (**c**) CNN   (**d**) ELM   (**e**) CNN-PPF

(**f**) AP-SAE

Figure 15. Classification result with nine classes for the University of Pavia data set,as thematic maps.

(**a**) image   (**b**) Ground-truth   (**c**) CNN   (**d**) ELM   (**e**) CNN-PPF

(**f**) AP-SAE
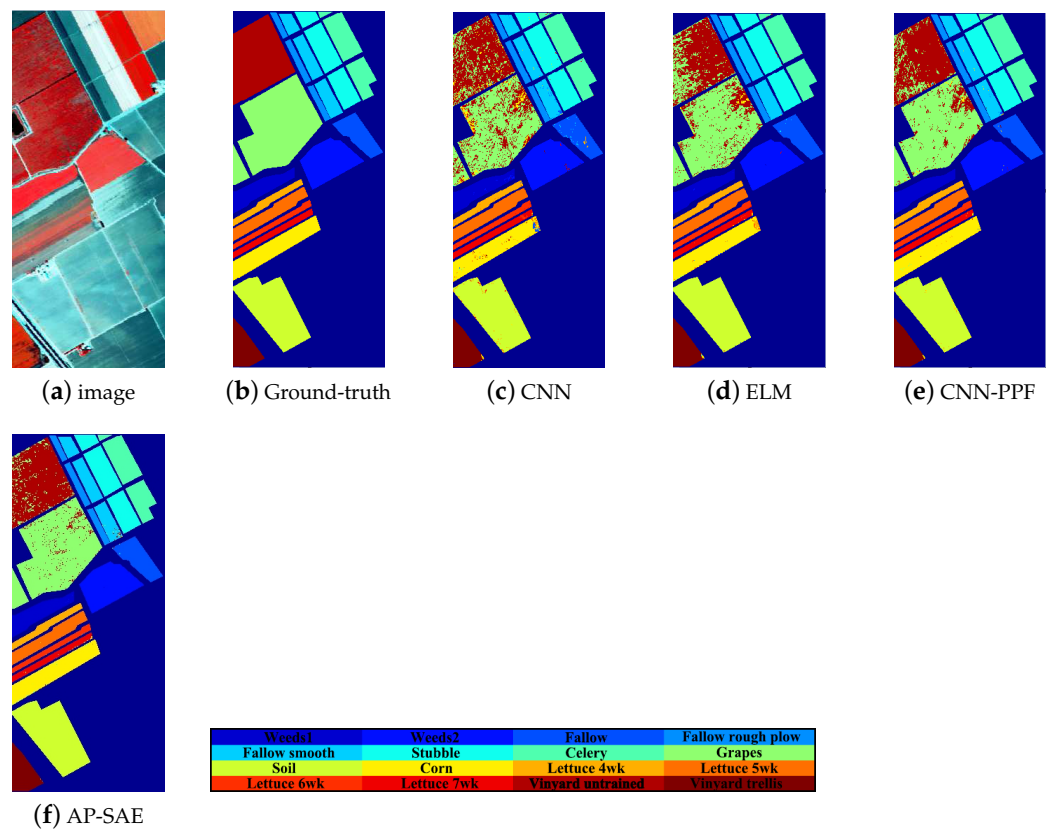
Figure 16. Classification result with 16 classes for the Salinas data set, as thematic maps.

(**a**) image   (**b**) Ground-truth   (**c**) CNN   (**d**) ELM   (**e**) CNN-PPF

(**f**) AP-SAE

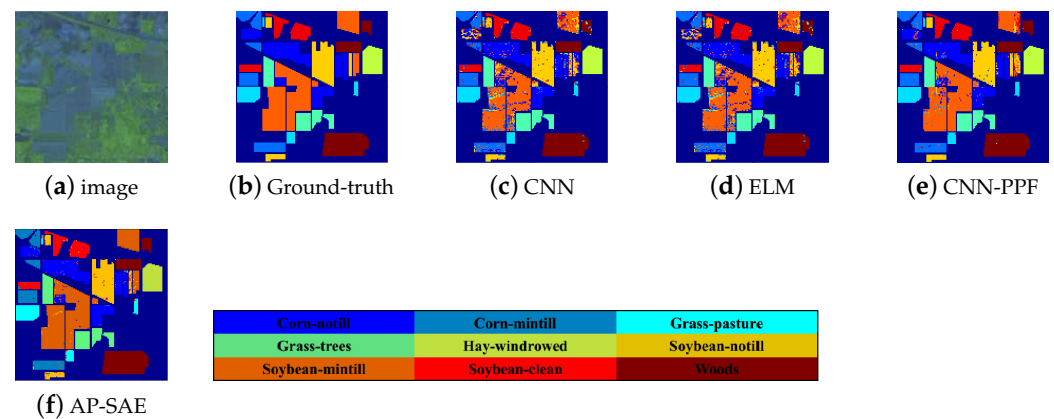| Corn-notill	Corn-mintill	Grass-pasture
Grass-trees	Hay-windrowed	Soybean-notill
Soybean-mintill	Soybean-clean	Woods

Figure 17. Classification result with 9 classes for the Indian Pines data set, as thematic maps.

5. Conclusions

Hyperspectral image classification is of significant value in remote sensing analysis, including the latest trend of satellite IoT, which can be applied in various scenarios, such as crop supervision, forest management, urban development and risk management. At the same time, continuity of the data as well as extrapolation among several scales, temporal, spatial and spectral, are key components of hyperspectral image classification [29].

Unfortunately, the traditional satellite system is facing the issue of latency and efficiency caused by gigantic amount of data collected by remote sensors. The remote sensing data are transmitted back to the ground for processing in traditional satellite system, the transparent forwarding of data was implemented without any processing on the satellite. The latency caused by transmission and processing on the ground can be decreased greatly if on-board computing can be introduced. Besides, with the development of spacecraft, issues related to performing on-board and automatic data computing and analysis as well as decision planning and scheduling will figure among the most important requirements. The method proposed in this paper can be adapted to other hyperspectral data with similar wavelength range and spectral channel number, so it can be extended to the satellite IoT application. Also, due to the influence of vegetation spectral similarity and the loss of spectrum information during the process of dimension reduction, high classification accuracy of some geomorphic types is hard to be obtained. This paper proposed an effective HSI classification model named AP-SAE at the edge of satellite IoT, and the classification accuracy can be significantly improved by our method without obvious efficiency degradation.

Experiments are made in this paper to prove the superiority of the method proposed, but there are also some deficiencies. For example, the determination of the number of middle layer neurons in AE lacks of generalization ability. At present, the determination of the number of middle layer neurons of AE is obtained by artificial experiments. It is still a problem to propose the algorithm framework to determine the neuron format in mathematics and formula level. In future research, it is worthwhile to try this innovative framework in various settings to test its applicability such as intelligent transportation networks [30]. Possibly, datasets with wide variations in volume, velocity, variety and veracity may lead to different performance of this framework. Moreover, with the upgrade of sensors, processors and transmitters on satellites, the division of work between the edge processing and the ground processing should be adjusted intelligently to reach optimal whole-system performance.

Author Contributions: Conceptualization: N.L.; methodology, N.L. and Z.H.; validation, C.C.; data curation, C.C. and Y.F.; writing—original draft preparation, N.L. and C.C.; formal analysis, T.S. and S.W.; writing—review and editing, N.L. and Z.H.; visualization, T.S. and S.G.; investigation, S.W.; supervision, S.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Chen, C.; Liu, B.; Wan, S.; Qiao, P.; Pei, Q. An Edge Traffic Flow Detection Scheme Based on Deep Learning in An Intelligent Transportation System. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1840–1852. [CrossRef]
2. Chen, C.; Wang, C.; Qiu, T.; Atiquzzaman, M.; Wu, D. Caching in Vehicular Named Data Networking: Architecture, Schemes and Future Directions. *IEEE Commu. Sury. Tutor.* **2020**, *22*, 2378–2407. [CrossRef]
3. Dalla Mura, M.; Atli Benediktsson, J.; Waske, B.; Bruzzone, L. Morphological Attribute Profiles for the Analysis of Very High Resolution Images. *IEEE Trans. Intell. Transp. Syst.* **2010**, *20*, 3747–3762. [CrossRef]
4. Ghamisi, P.; Dalla Mura, M.; Benediktsson, J. A Survey on Spectral–Spatial Classification Techniques Based on Attribute Profiles. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2335–2353. [CrossRef]
5. MURA, M.D.; Benediktsson, J.A.; WASKE, B.; Bruzzone, L. Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *Int. J. Remote Sens.* **2010**, *31*, 5975–5991.
6. Cavallaro, G.; Dalla Mura, M.; Benediktsson, J.A.; Bruzzone, L. Extended Self-Dual Attribute Profiles for the Classification of Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1690–1694. [CrossRef]
7. Aptoula, E.; Ozdemir, M.C.; Yanikoglu, B. Deep Learning With Attribute Profiles for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1970–1974. [CrossRef]
8. Cavallaro, G.; Falco, N.; Mura, M.; Bruzzone, L.; Benediktsson, J.A. Automatic Threshold Selection for Profiles of Attribute Filters Based on Granulometric Characteristic Functions. In Proceedings of the 12th International Symposium on Mathematical Morphology, Reykjavik, Iceland, 27–29 May 2015; Springer: Cham, Switzerland, 2015.
9. Dalla Mura, M.; Bruzzone, L.; Notarnicola, C.; Benediktsson, J.A.; Bruzzone, L.; Posa, F. SPIE Proceedings [SPIE SPIE Europe Remote Sensing—Berlin, Germany (Monday 31 August 2009)] Image and Signal Processing for Remote Sensing XV—Modeling structural information for building extraction with morphological attribute filters. *Proc. SPIE Int. Soc. Opt. Eng.* **2009**, *7477*, 747703.
10. Zhang, X.; Liang, Y.; Li, C.; Huyan, N.; Jiao, L.; Zhou, H. Recursive Autoencoders-Based Unsupervised Feature Learning for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1928–1932. [CrossRef]
11. 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]
12. Lv, N.; Chen, C.; Qiu, T.; Sangaiah, A.K. Deep Learning and Superpixel Feature Extraction Based on Contractive Autoencoder for Change Detection in SAR Images. *IEEE Trans. Ind. Inf.* **2018**, *14*, 5530–5538. [CrossRef]
13. Chen, C.; Liu, L.; Qiu, T.; Yang, K.; Gong, F.; Song, H. ASGR: An Artificial Spider-Web-Based Geographic Routing in Heterogeneous Vehicular Networks. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 1604–1620. [CrossRef]
14. Lauzon, F.Q. IEEEexample: An introduction to deep learning. In Proceedings of the 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), Montreal, QC, Canada, 2–5 July 2012; pp. 1438–1439.
15. Lin, Z.; Chen, Y.; Zhao, X.; Wang, G. Spectral-spatial classification of hyperspectral image using autoencoders. In Proceedings of the 2013 9th International Conference on Information, Communications & Signal Processing, Tainan, Taiwan, 10–13 December 2013; pp. 1–5.
16. Sun, X.; Zhou, F.; Dong, J.; Gao, F.; Mu, Q.; Wang, X. Encoding Spectral and Spatial Context Information for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2250–2254. [CrossRef]
17. Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **1901**, *2*, 559–572. [CrossRef]
18. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **1999**, *10*, 626–634. [CrossRef]

19. Marpu, P.R.; Pedergnana, M.; Dalla Mura, M.; Benediktsson, J.A.; Bruzzone, L. Automatic Generation of Standard Deviation Attribute Profiles for Spectral–Spatial Classification of Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 293–297. [CrossRef]

20. Chen, C.; Liu, Z.; Wan, S.; Luan, J.; Pei, Q. Traffic Flow Prediction Based on Deep Learning in Internet of Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 3776–3789. [CrossRef]

21. Surhone, L.M.; Tennoe, M.T.; Henssonow, S.F. *Structuring Element*; Betascript Publishing: Warszawa, Poland, 2011.

22. Pedergnana, M.; Marpu, P.R.; Mura, M.; Benediktsson, J.A.; Bruzzone, L. A Novel Technique for Optimal Feature Selection in Attribute Profiles Based on Genetic Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3514–3528. [CrossRef]

23. Srivastava, Y.; Murali, V.; Dubey, S.R. PSNet: Parametric Sigmoid Norm Based CNN for Face Recognition. In Proceedings of the IEEE CICT 2019 Conference, Prayagraj, India, 6–8 December 2019; pp. 1–5.

24. Zhou, P.; Han, J.; Cheng, G.; Zhang, B. Learning Compact and Discriminative Stacked Autoencoder for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4823–4833. [CrossRef]

25. Li, W.; Chen, C.; Su, H.; Du, Q. Local Binary Patterns and Extreme Learning Machine for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3681–3693. [CrossRef]

26. Waske, B.; van der Linden, S.; Benediktsson, J.A.; Rabe, A.; Hostert, P. Sensitivity of Support Vector Machines to Random Feature Selection in Classification of Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2880–2889. [CrossRef]

27. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep Convolutional Neural Networks for Hyperspectral Image Classification. *J. Sens.* **2015**, *2015*, 258619. [CrossRef]

28. Jiao, L.; Liang, M.; Chen, H.; Yang, S.; Liu, H.; Cao, X. Deep Fully Convolutional Network-Based Spatial Distribution Prediction for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5585–5599. [CrossRef]

29. Moigne, J.L. Multi-Sensor Image Registration, Fusion and Dimension Reduction. *Online J. Space Commun.* **2003**, *2*, 15.

30. Chen, C.; Hu, J.; Qiu, T.; Atiquzzaman, M.; Ren, Z. CVCG: Cooperative V2V-aided transmission scheme based on coalitional game for popular content distribution in vehicular ad-hoc networks. *IEEE Trans. Mob. Comput.* **2018**, *18*, 2811–2828. [CrossRef]