remote sensing

Article

Fusion of Multidimensional CNN and Handcrafted Features for Small-Sample Hyperspectral Image Classification

Haojin Tang [1,2], Yanshan Li [1,2,*], Zhiquan Huang [1,2], Li Zhang [1,2] and Weixin Xie [1,2]

[1] ATR Key Laboratory, Shenzhen University, Shenzhen 518060, China
[2] Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China
* Correspondence: lys@szu.edu.cn

Abstract: Hyperspectral image (HSI) classification has attracted widespread concern in recent years. However, due to the complexity of the HSI gathering environment, it is difficult to obtain a great number of HSI labeled samples. Therefore, how to effectively extract the spatial–spectral feature with small-scale training samples is the crucial point of HSI classification. In this paper, a novel fusion framework for small-sample HSI classification is proposed to fully combine the advantages of multidimensional CNN and handcrafted features. Firstly, a 3D fuzzy histogram of oriented gradients (3D-FHOG) descriptor is proposed to fully extract the handcrafted spatial–spectral feature of HSI pixels, which is suggested to be more robust by overcoming the local spatial–spectral feature uncertainty. Secondly, a multidimensional Siamese network (MDSN), which is updated by minimizing both contrastive loss and classification loss, is designed to effectively exploit the CNN-based spatial–spectral features from multiple dimensions. Finally, the proposed MDSN combined with 3D-FHOG is utilized for small-sample HSI classification to verify the effectiveness of our proposed fusion framework. The experimental results on three public data sets indicate that the proposed MDSN combined with 3D-FHOG is significantly better than the representative handcrafted feature-based and CNN-based methods, which in turn demonstrates the superiority of the proposed fusion framework.

Keywords: small-sample hyperspectral image classification; spatial–spectral feature extraction; multidimensional CNN; handcrafted feature

Citation: Tang, H.; Li, Y.; Huang, Z.; Zhang, L.; Xie, W. Fusion of Multidimensional CNN and Handcrafted Features for Small-Sample Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 3796. https://doi.org/10.3390/rs14153796

Academic Editors: Saeid Homayouni, Pedram Ghamisi, Amin Zehtabian, Ali Mousivand and Fardin Mirzapour

Received: 30 June 2022
Accepted: 4 August 2022
Published: 6 August 2022

1. Introduction

Compared with gray-scale and RGB images, hyperspectral image (HSI) can provide a rich amount of spatial and spectral information of objects. Since the additional spectral information may help to overcome the existing difficulties of traditional image processing technology, HSI has attracted widespread concern in recent years. HSI classification [1–4], which is the focus of research in the field of HSI processing, has been widely applied in various areas, such as scene understanding [5,6], disease examination [7,8], face recognition [9,10] and city planning [11,12]. Note that the feature extracted from HSI is the basis of these applications, thus it is essential to obtain the more robust and effective spatial–spectral feature for HSI classification. However, due to the complexity and potential fatalness of the HSI acquisition environment, it is a laborious and time-consuming job to collect a huge amount of labeled samples. Therefore, how to effectively extract the spatial–spectral feature with small-scale training samples has become a hot spot of current research. Existing HSI feature extraction methods can be divided into two types: handcrafted feature extraction methods and deep feature extraction methods.

Before the development of deep learning, handcrafted feature extraction was the mainstream approach in the field of image processing, and its effectiveness has been verified in image matching and classification [13,14]. Lowe [13] designed an image feature extraction method called the scale-invariant feature transform (SIFT), which shows its robustness in object recognition. Local binary pattern (LBP) proposed by Ojala et al. [14] is a simple and efficient local feature descriptor, which is able to achieve gray-scale and rotation invariant texture classification. Meanwhile, since HSI can be represented as the 3D-structure data, many 3D handcrafted feature descriptors have been presented for the HSI feature extraction. Zhao et al. [15] designed the 3D-LBP operator to extract the dynamic feature from the spatial–temporal domain, which is an extension of LBP. Inspired by [15], Jia et al. [16] applied the 3D-LBP to the spatial–spectral domain of HSI, which exhibits excellent performance in HSI classification. He et al. [17] presented a 3D Gabor filter-based descriptor, which can be utilized to perform HSI classification in a computationally efficient way. The 3D discrete wavelet transform (3D-DWT) proposed by Cao et al. [18] can fully utilize the spatial–spectral information and improve the performance of HSI classification. However, due to the simple structure and fixed calculation pattern, handcrafted features are not robust when confronted with complex circumstances of HSI classification.

In recent years, with the development of hardware devices and the arrival of the big data era, deep learning technique has made great progress. Especially, convolutional neural network (CNN) is the most commonly used deep learning technique in the area of computer vision. Because of its local connection and nonlinear characteristic, making it able to extract the more discriminative feature, CNN is quite effective for image processing, including HSI classification. Sharma et al. [19] combined the band selection with 2D CNN-based features to enhance the performance of HSI classification, which outperforms the handcrafted feature-based methods. Meanwhile, 1D CNN and 1D recurrent neural network (RNN) features [20–22] were utilized to process HSI pixels as sequential data, which takes full advantage of spectral correlation and band-to-band variability. To fully exploit the spatial and spectral information, 3D CNN-based methods are also employed for HSI classification. Lee et al. [23] designed a deeper CNN architecture that uses 3D fully convolutional layers (FCN) to learn the more effective spatial–spectral feature. The semi-supervised 3D CNN-based algorithm proposed by Liu et al. [24] can simultaneously minimize the sum of supervised and unsupervised cost functions during training, which aims to solve the problem of limited labeled samples. Luo et al. [25] presented a 3D CNN framework for HSI classification, which exhibits a good trade-off between the number of training samples and the complexity of the network. Roy et al. [26] proposed a hybrid spectral CNN (HybridSN) for HSI classification, which combines advantages of both 3D CNN and 2D CNN. Although CNN-based methods can achieve state-of-the-art performance with sufficient labeled samples, they cannot provide a strict mathematical explanation for its decision making. In addition, HSI classification accuracy of CNN-based methods will significantly decrease in the scenery of small-scale training samples.

As can be seen from the literature, the handcrafted feature-based methods can provide a stricter mathematical explanation for the HSI feature extraction process, which makes it more reliable to be utilized in some high-sensitive areas, such as biomedicine and military. However, compared with CNN-based methods, the performance of handcrafted feature-based methods is not robust in some complicated HSI classification tasks. On the other hand, utilizing only the CNN-based feature causes the difficulty in achieving high accuracy with limited labeled samples, and lacks the strict mathematical explanation for its decision making. Therefore, it is essential to develop a HSI classification algorithm that combines the advantages of both handcrafted and CNN-based feature.

Small-sample classification has become an important research topic in the area of remote sensing. To tackle the challenge of small-scale training samples, the idea of transfer learning has been introduced in remote sensing scene classification. Rostami et al. [27] proposed a deep transfer learning-based algorithm for few-shot synthetic aperture radar (SAR) image classification, which is effective on the problem of ship classification in the SAR domain. Alajaji et al. [28] combined the prototypical network with pre-trained CNN for image embedding, which obtains excellent classification results on two remote sensing scene data set. To further extract the generalized features from source domain, attention mechanism and multi-scale feature fusion strategy [29,30] are introduced in remote sensing few-shot scene classification. However, most of the transfer learning-based algorithms utilize only the CNN-based feature, ignoring the superiority of handcrafted feature. Moreover, since there may be a mismatch between the source and target domain distributions, the performances of transfer learning-based algorithms are unpredictable. As a special type of remote sensing image, HSI can provide additional spectral information for feature extraction. It can be known that the handcrafted feature is more reliable and easier to carry out without training. Therefore, our proposed algorithm mainly focuses on how to utilize the handcrafted feature to enhance the performance of CNN-based models in the scenery of small-sample supervised learning.

It can be known that fusing different types of spatial–spectral features may cause the increase in computational cost. However, in terms of some special small-sample HSI classification tasks, more accurate classification results need to be achieved without considering the computational cost. To the best of our knowledge, there still lacks an in-depth study concentrated on utilizing the handcrafted feature to enhance the performance of CNN-based models in small-sample HSI classification. Therefore, we propose a fusion framework of multidimensional CNN and handcrafted features for small-sample HSI classification. Specifically, a multidimensional Siamese network (MDSN) combined with the 3D fuzzy histogram of oriented gradients (3D-FHOG) features is introduced to verify the effectiveness of our proposed fusion framework.

The main contribution of this paper includes the following three aspects.

(1) A 3D-FHOG descriptor is proposed to fully extract the handcrafted spatial–spectral feature of HSI pixels. It calculates the HOG features from three orthogonal planes to generate the final 3D-FHOG descriptor based on fuzzy fusion operation, which is able to overcome the local spatial–spectral feature uncertainty;
(2) An effective Siamese network, i.e., MDSN is designed for further exploiting the multidimensional CNN-based spatial–spectral feature in the scenery of small-scale labeled samples. It mainly utilizes the hybrid 3D-2D-1D CNN to learn the spatial–spectral feature from multiple dimensions and is updated by minimizing both contrastive loss and classification loss. Compared with the single-dimensional CNN-based networks, the performance of MDSN is significantly better in small-sample HSI classification;
(3) It provides a novel extensible fusion framework for the combination of hand- crafted and multidimensional CNN-based spatial–spectral features. More importantly, experimental results indicate that our proposed MDSN combined with 3D-FHOG features can achieve better performance than the handcrafted features-based and CNN-based algorithms, which in turn verifies the superiority of the proposed fusion framework.

The rest of this paper is organized as follows. Section 2 presents the related works of this study. The proposed methodology is presented in Section 3. Then, Section 4 reports the experimental results and discussions on three public data sets. Finally, a conclusion of this study is presented in Section 5.

2. Related Works

2.1. Histogram of Oriented Gradients

Histogram of oriented gradients (HOG) proposed by Dalal et al. [31] is a classical handcrafted feature descriptor, which is generated by computing the gradients of pixels in a local area. Additionally, it not only provides the rotation invariance and under- standability, but also has a strong capacity of shape feature expression, which makes it widely used in image recognition. Surasak et al. [32] applied the HOG algorithm to human detection in video, which is able to accurately obtain the number of people for each video frame. Mao et al. [33] utilized the HOG-based method and support vector machine (SVM) classifier to perform the preceding vehicle detection, which shows excellent performance in different traffic scenarios. Qi et al. [34] designed a ship histogram of oriented gradient (S-HOG) to characterize ship targets, which proved to be also effective when ship size varies. Since each HSI pixel corresponds to a spectral curve with different changing patterns, it is suggested that constructing the statistics histogram of local gradient change for HSI pixels is an effective solution for describing its local spatial–spectral features. Chen et al. [35] proposed a novel algorithm for hyperspectral face recognition by extracting the HOG feature, which outperforms several existing methods in the experiment. However, existing HOG-based algorithms ignore the characteristics of HSI, such as strong correlation between bands, vast amounts of redundant information and spatial–spectral feature uncertainty. Therefore, in this study, by introducing the fuzzy logic theory, we design a novel handcrafted feature descriptor named 3D-FHOG to fully exploit the spatial–spectral information and to overcome the spatial–spectral feature uncertainty.

2.2. Siamese Network

For the problem of small-scale labeled samples, it is suggested that a model named Siamese network [36,37] will be an effective solution for small-sample HSI classification. Specifically, Siamese network consists of two branches with the same architecture, and the image pairs are adopted as the input of Siamese network to minimize the contrastive loss. Early research of Siamese network mainly focuses on the application for target tracking. Tao et al. [38] first proposed to utilize the Siamese network in tracking tasks, which achieves state-of-the-art performance. Bertinetto et al. [39] designed a fully convolutional Siamese network to locate an exemplar image within a larger search image. Since the number of labeled samples will be augmented by generating the image pairs, Siamese network has been employed for few-shot classification tasks. Koch et al. [40] applied the Siamese network to the one-shot image recognition task, which obtains promising results. With respect to HSI classification, Zhao et al. [41] utilized the Siamese network to enlarge the training set and extract the effective spatial–spectral features, which is able to improve the classification performance. Liu et al. [42] proposed a Siamese network supervised with a margin ranking loss function for HSI classification, which can obtain better classification results than those of the conventional methods. Very recently, Cao et al. [43] designed a hybrid Siamese network called 3DCSN to perform HSI classification, which is suggested to be a robust and accurate classifier in the scenery of small-scale training samples. As described in [44], it can be known that the spatial–spectral features extracted from different CNN layers may contain the semantic information of objects with different scales. Therefore, in this paper, an effective Siamese network named MDSN is proposed to fully exploit the multidimensional CNN-based spatial–spectral feature. Moreover, we train the proposed MDSN by using both the contrastive loss function and classification loss function. Especially, our proposed MDSN is integrated with the idea of prototypical network in the testing phase, which is suggested to be more effective for small-sample HSI classification.

3. Methodology

The fusion framework of multidimensional CNN-based and handcrafted features for small-sample HSI classification is shown in Figure 1. In this study, to verify the effectiveness of our proposed fusion framework, we design the 3D-FHOG and MDSN for the handcrafted and multidimensional CNN-based feature extraction, respectively. As shown in Figure 1, small-sample HSI classification of MDSN combined with 3D-FHOG features mainly consists of three parts: firstly, principal component analysis (PCA) [45] algorithm is implemented on HSI to extract the representative band data; next, 3D patches divided from HSI are utilized to perform the 3D-FHOG feature extraction and the hybrid 3D-2D-1D CNN feature extraction; finally, through the linear layers and the distance metric between labeled and unlabeled samples with the 3D-FHOG and MDSN features, three class-score vectors are obtained (i.e., P_1, P_2 and P_3), which are fused to compute the probability of a HSI pixel classified into a specific class.

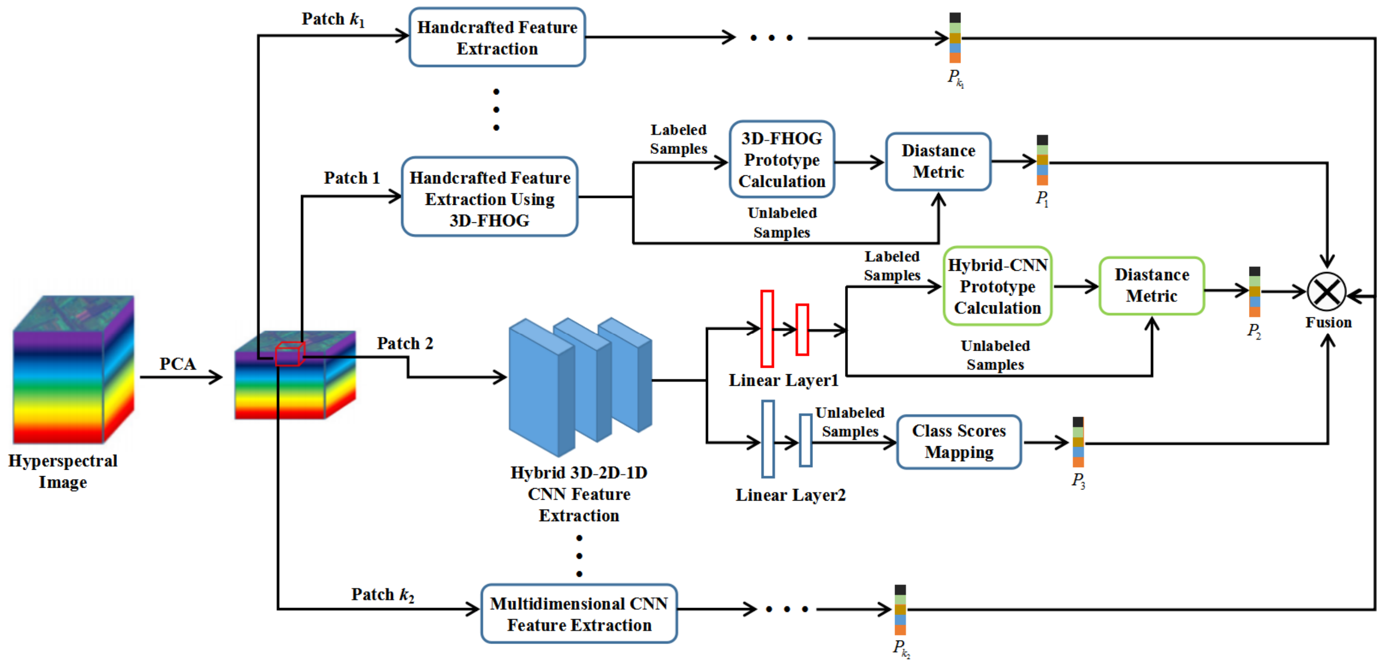


Figure 1. The fusion framework of multidimensional CNN and handcrafted features for small-sample HSI classification. Initially, 3D-FHOG is adopted as the handcrafted feature extraction method, and MDSN is utilized for multidimensional CNN-based feature extraction.

3.1. The Proposed 3D-FHOG

By introducing the fuzzy logic theory, the proposed 3D-FHOG is utilized to fully extract the handcrafted spatial–spectral feature and to overcome the spatial–spectral feature uncertainty. Figure 2 shows the schematic of 3D-FHOG feature extraction.

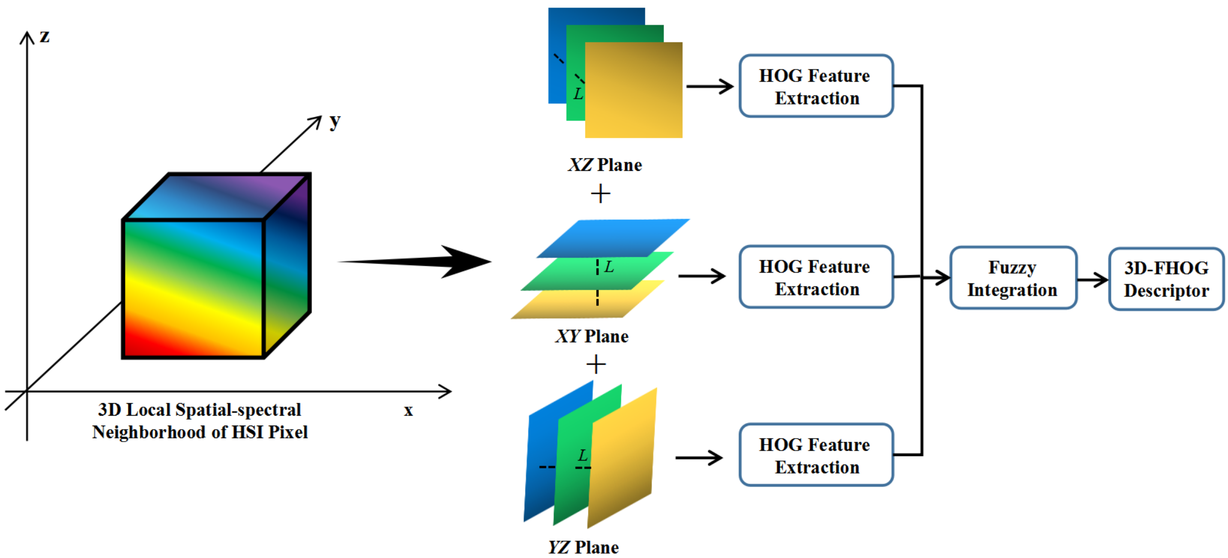


Figure 2. The schematic of 3D-FHOG feature extraction.

Let H be the HSI, thus the HSI pixel with a spatial coordinate of (x, y) and λ bands can be represented as a λ-dimensional vector, as follows:

$$\vec{H}_\lambda(x, y) = [H(x, y, z_1), H(x, y, z_2), \ldots, H(x, y, z_\lambda)] \tag{1}$$

where $H(x, y, z)$ denotes the spectral response of HSI pixel, and z represents the spectral domain coordinate. Then, the λ-dimensional vector of HSI pixel is converted into the 3D local spatial–spectral neighborhood for HOG feature extraction. It can be known that the 3-D local spatial–spectral neighborhood of HSI pixels can be expressed as a group of orthogonal planes, including XY, XZ and YZ planes. Therefore, HOG feature extraction is implemented on XY, XZ and YZ planes, respectively. For the XY planes, assuming that $H(x, y, k)$ denotes the spectral response of HSI pixel with spatial coordinate (x, y) at kth band, thus the x-axis oriented gradient and y-axis oriented gradient of $H(x, y, k)$ can be calculated as follows:

$$G_{xy1}(x, y, k) = H(x, y + 1, k) - H(x, y - 1, k) \tag{2}$$

$$G_{xy2}(x, y, k) = H(x + 1, y, k) - H(x - 1, y, k) \tag{3}$$

where $G_{xy1}(x, y, k)$ and $G_{xy2}(x, y, k)$ denote the y-axis and x-axis oriented gradient of $H(x, y, k)$, respectively. Therefore, the oriented gradient of HSI pixels in the kth XY plane of 3-D local spatial–spectral neighborhood can be expressed as follows:

$$G_{xy}(x, y, k) = \sqrt{G_{xy1}(x, y, k)^2 + G_{xy2}(x, y, k)^2} \tag{4}$$

$$\alpha_{xy}(x, y, k) = \tan^{-1}\left(\frac{G_{xy1}(x, y, k)}{G_{xy2}(x, y, k)}\right) \tag{5}$$

where $G_{xy}(x, y, k)$ represents the gradient magnitude of $H(x, y, k)$, $\alpha_{xy}(x, y, k)$ is the gradient direction of $H(x, y, k)$. Then, by setting the suitable block size and cell size of HOG descriptor in 3-D local spatial–spectral neighborhood, the final expression of HOG descriptor for XY planes is obtained. According to [31], the nine-bin histogram h^k_{xy} is obtained from each cell of HOG descriptor. Therefore, let the cell size of HOG descriptor be $M \times N$, the bin $h^k_{xy}(b)$ of h^k_{xy} can be expressed as follows:

$$h^k_{xy}(b) = \sum_{m=1}^{M} \sum_{n=1}^{N} s\big(\alpha_{xy}(m, n, k), b\big) \cdot G_{xy}(x, y, k) \tag{6}$$

where $s(a, b)$ is defined as:

$$s(x_1, x_2) = \left\{ \begin{array}{l} 0, x_1 - x_2 \geq x_2 + \beta \\ 1, 0 \leq x_1 - x_2 < x_2 + \beta \\ 0, x_1 - x_2 < 0 \end{array} \right\} \tag{7}$$

Since the histogram channels are spread over 0 to 180 degrees, thus β is normally set to be 20. In general, each block of HOG descriptor contains 2×2 cells, thus the HOG descriptor for XY planes can be represented as:

$$HOG^k_{xy} = \left[\left(h^k_{xy}\right)_1, \left(h^k_{xy}\right)_2, \ldots, \left(h^k_{xy}\right)_P\right] \tag{8}$$

where $(h_{xy})_i = \left[\left(h^k_{xy1}\right)_i, \left(h^k_{xy2}\right)_i, \left(h^k_{xy3}\right)_i, \left(h^k_{xy4}\right)_i\right]$, $\left(h^k_{xyj}\right)_i$ denotes the jth nine-bin histogram of the block, P denotes the number of blocks in the kth XY plane. Similarly, HOG descriptors for the kth XZ and YZ planes (i.e., HOG^k_{xz} and HOG^k_{yz}) can be obtained. Therefore, the kth 3D-HOG descriptor HOG^k_{3D} can be expressed as follows:

$$HOG^k_{3D} = \left[HOG^k_{xy}, HOG^k_{xz}, HOG^k_{yz}\right] \tag{9}$$

Because HSI has the characteristic of low spatial resolution and wide distribution of ground objects, thus the 3D local spatial–spectral neighborhood of a HSI pixel that belongs to a specific class may contain the spatial–spectral information of other classes. This leads to the problem of spatial–spectral feature uncertainty in the process of spatial–spectral feature extraction. Especially, when performing the feature fusion of three orthogonal planes, the confidence of HOG feature extracted from each plane is uncertain. Therefore, fusing the HOG feature of three orthogonal planes directly may result in the performance degradation of small-sample HSI classification.

Fuzzy logic proposed by Zadeh [46] is a significant approach for overcoming the uncertainties among the raw data. Inspired by this, we apply the theory of fuzzy integration to the process of 3D-HOG feature extraction. According to [47,48], let S_V represent the fuzzy integration function, thus it can be expressed as follows:

$$S_V[v_1, v_2, \ldots, v_n] = \left[\frac{1}{n}\sum_{i=1}^{n} v_i^q\right]^{\frac{1}{q}} \tag{10}$$

where q denotes the fuzzy factor. Hence, by performing the fuzzy integration for the HOG features extracted from three orthogonal planes, 3D-HOG descriptor is trans- formed into the 3D-FHOG descriptor with strong robustness, as below:

$$FHOG_{3D}^{k} = S_V\left[HOG_{xy}^{k}, HOG_{xz}^{k}, HOG_{yz}^{k}\right] \tag{11}$$

where $FHOG_{3D}^{k}$ represents the kth 3D-FHOG descriptor. Let L be the step size of 3D-FHOG feature, thus the final expression of 3D-FHOG descriptor for $\vec{H}_\lambda(x, y)$ can be formulated as below:

$$\begin{aligned} FHOG_{3D} \quad &= \Big\{S_V\left[HOG_{xy}^{1}, HOG_{xz}^{1}, HOG_{yz}^{1}\right] \\ &, S_V\left[HOG_{xy}^{1+L}, HOG_{xz}^{1+L}, HOG_{yz}^{1+L}\right], \\ &\cdots, S_V\left[HOG_{xy}^{\lambda-L}, HOG_{xz}^{\lambda-L}, HOG_{yz}^{\lambda-L}\right] \\ &, S_V\left[HOG_{xy}^{\lambda}, HOG_{xz}^{\lambda}, HOG_{yz}^{\lambda}\right]\Big\} \end{aligned} \tag{12}$$

As mentioned above, our proposed 3D-FHOG is able to provide a stricter mathematical explanation, which makes it more reliable to be utilized in the high sensitive areas. Moreover, it can not only fully extract the handcrafted spatial–spectral feature of HSI pixels, but also overcome the spatial–spectral feature uncertainty. Therefore, in this study, the proposed 3D-FHOG feature is used to enhance the performance of multidimensional CNN.

3.2. The Proposed MDSN

The structure of the MDSN is presented in Figure 3. As shown in Figure 3, the spatial–spectral features are first extracted by the 3D convolutional blocks from the input patches. Secondly, the 2D convolutional block is performed to further enhance the spatial feature. Then, spectral features are further extracted by the 1D convolutional block. Finally, through the linear layer, the obtained MDSN feature is adopted to compute contrastive loss, and the classification loss is calculated based on the class probability output from the linear classifier.

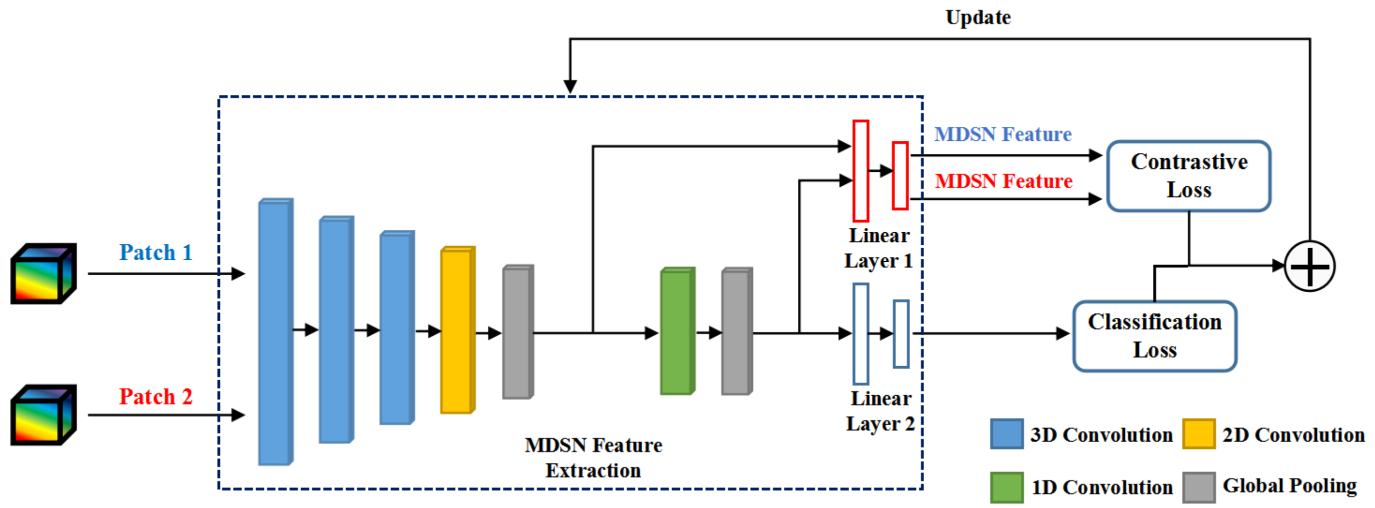


Figure 3. The structure of MDSN. Each convolutional block contains a convolutional layer, a batch normalization layer and a ReLU nonlinearity corresponding to its convolutional dimension. When performing the contrastive learning, two different patches are adopted at the same time. In the classification phase, only one patch is used.

Let Ψ be the training set of N labeled samples, as follows:

$$\Psi = \{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_N, y_N)\} \tag{13}$$

where $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}\left(x_i \in \mathbb{R}^{W_1 \times W_2 \times K}\right)$ represents the 3D patches of HSI pixels divided from HSI, $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$ denotes the corresponding label. Next, a pair of 3D patches $\left(x_i, x_j\right)$ is randomly selected from $\mathbf{x}$, which is adopted as the input of MDSN. Let $y_{i,j}$ be the label of $\left(x_i, x_j\right)$, the value of which is defined as below:

$$y_{i,j} = \begin{cases} 1, y_i = y_j \\ 0, y_i \neq y_j \end{cases} \tag{14}$$

During the training process, our proposed MDSN is updated by minimizing both contrastive loss and classification loss. When performing the contrastive learning, let θ be the nonlinear parameters in MDSN. Thus, the formula of θ updated by contrastive loss can be expressed as follows:

$$\theta = \underset{\theta}{\operatorname{argmin}}\left\{L_c\left[g_1(x_i), g_1\left(x_j\right), y_{i,j}; \theta\right]\right\} \tag{15}$$

where $g_1(\cdot)$ denotes the encoder function in the branch of MDSN utilized for computing the contrastive loss, L_c represents the contrastive loss function, as below:

$$L_c = \frac{1}{2}\left[y_{i,j}d_{i,j}^2 + (1 - y_{i,j})\max\left(\text{margin} - d_{i,j}, 0\right)^2\right] \tag{16}$$

$$d_{i,j} = \left\|g_1(x_i) - g_2\left(x_j\right)\right\|_2 \tag{17}$$

where margin is a constant, and its typical value is 1.25. In the classification phase, we adopt the cross-entropy loss function L_s to compute the classification loss. Besides, only one patch x_i is used at a time. Hence, the formula of θ updated by classification loss can be represented as follows:

$$L_s = -\sum_{i=1}^{N} y_i * \log \hat{y}_i \tag{18}$$

$$\theta = \underset{\theta}{\operatorname{argmin}}\{L_s[h(g_2(x_i)), y_i; \theta]\} \tag{19}$$

where $g_2(\cdot)$ represents the encoder function in the branch of MDSN utilized for computing the classification loss, $h(\cdot)$ denotes the class-score mapping function of linear layers, $\hat{y}_i$ is the predicted label of x_i. In summary, by training with contrastive loss and classification loss, our proposed MDSN can effectively exploit the multi-dimensional CNN-based spatial–spectral feature. Compared with the single-dimensional CNN-based models, MDSN is able to achieve better performance in small-sample HSI classification.

3.3. MDSN Combined with 3D-FHOG for Small-Sample HSI Classification

As described early, in some special HSI classification tasks with only a few labeled samples, we need to achieve higher classification accuracy without considering the computational cost. However, the scarcity of labeled samples makes it difficult to train an effective CNN-based classifier. In terms of the handcrafted feature, it can provide a stricter mathematical explanation for its feature extraction process, but it is difficult to be applied in some complex data processing tasks. Therefore, in this paper, we design a fusion framework of multidimensional CNN and handcrafted features for small-sample HSI classification. Especially, our proposed MDSN combined with 3D-FHOG is utilized to verify the effectiveness of our proposed fusion framework.

According to Equations (12) and (13), let Ψ_k be the training set of N_1 labeled samples labeled with class k. After 3D-FHOG feature extraction, Ψ_k can be expressed as below:

$$\Psi_k = \left\{ (F(x_1), y_1), \ldots, (F(x_i), y_i), \ldots, \left(F\left(x_{N_1}\right), y_{N_1}\right) \right\} \tag{20}$$

where $F(\cdot)$ represents the 3D-FHOG feature embedding function. According to the idea of prototypical network [49,50], 3D-FHOG prototype can be calculated through the mean method, as follows:

$$c_k = \frac{1}{|\Psi_k|} \sum_{(x_i, y_i) \in \Psi_k} F(x_i) \tag{21}$$

Then, based on the distance metric with 3D-FHOG prototypes, the probability of pixel x classified as class k can be formulated as below:

$$P_1(y = k|x) = \frac{-d(F(x_i), c_k)}{\sum\limits_{j=1}^{G} (-d(F(x_i), c_j))} \tag{22}$$

where $d(\cdot)$ denotes the distance function, G is the total number of classes. As mentioned above, MDSN feature vector is extracted from the branch of MDSN utilized for computing the contrastive loss. Therefore, class probability based on the distance metric with hybrid-CNN prototypes can be formulated as follows:

$$P_2(y = k|x) = \frac{-d(g_1(x_i), c'_k)}{\sum\limits_{j=1}^{G} (-d(g_1(x_i), c'_j))} \tag{23}$$

where c'_k denotes the hybrid-CNN prototype labeled with class k. We assume that P_3 denotes the class-score vector output from the linear layers. Hence, by performing the fusion operation on P_1, P_2 and P_3, the final class probability of HSI pixels, which is obtained from the MDSN combined with 3D-FHOG features, can be expressed as below:

$$P = \delta(P_1, P_2, P_3) \tag{24}$$

where $\delta(\cdot)$ denotes the fusion function. Specifically, the fusion method for P_1, P_2 and P_3, such as concatenation and average, can be designed according to the needs of computational cost. In our experiment, P_1, P_2 and P_3 are fused by average. To sum up, by integrating with the idea of prototype calculation, MDSN and 3D-FHOG features are effectively fused to calculate the class probability of HSI pixels, which is able to obtain more accurate results in small-sample HSI classification.

The detailed process of MDSN combined with 3D-FHOG for small-sample HSI classification is described as follows (Algorithm 1).

Algorithm 1. MDSN combined with 3D-FHOG for small-sample HSI classification

Input: HSI pixels x.
Output: The fused class probability P.
Step 1. Generating the 3D patches $x_1 \in \mathbb{R}^{W_1 \times W_1 \times K_1}$ and $x_2 \in \mathbb{R}^{W_2 \times W_2 \times K_2}$ from the local spatial–spectral neighborhood of x.
Step 2. Performing the 3D-FHOG feature extraction and MDSN feature extraction on x_1 and x_2, respectively, by using $F(\cdot)$ and $g_1(\cdot)$.
Step 3. Computing the distance metric between $F(x_1)$ and c_k to obtain the class probability P_1.
Step 4. Calculating the distance metric between $g_1(x_2)$ and c'_k to obtain the class probability P_2.
Step 5. Getting the class probability P_3 output from $h(g_2(x_2))$.
Step 6. Fusing the class probability P_1, P_2 and P_3 by using Equation (24).

4. Experiments and Results

4.1. Data Sets

The Indian Pine (IP) data set which contains 16 classes and 10,249 samples was captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in north-western Indiana. Its spatial size and resolution are 145 × 145 pixels and 20-m per pixel. The number of bands is reduced to 200 by removing bands covering the region of water absorption. The false-color image and corresponding ground-truth map of the IP data set are shown in Figure 4. Table 1 lists the samples of the IP data set.

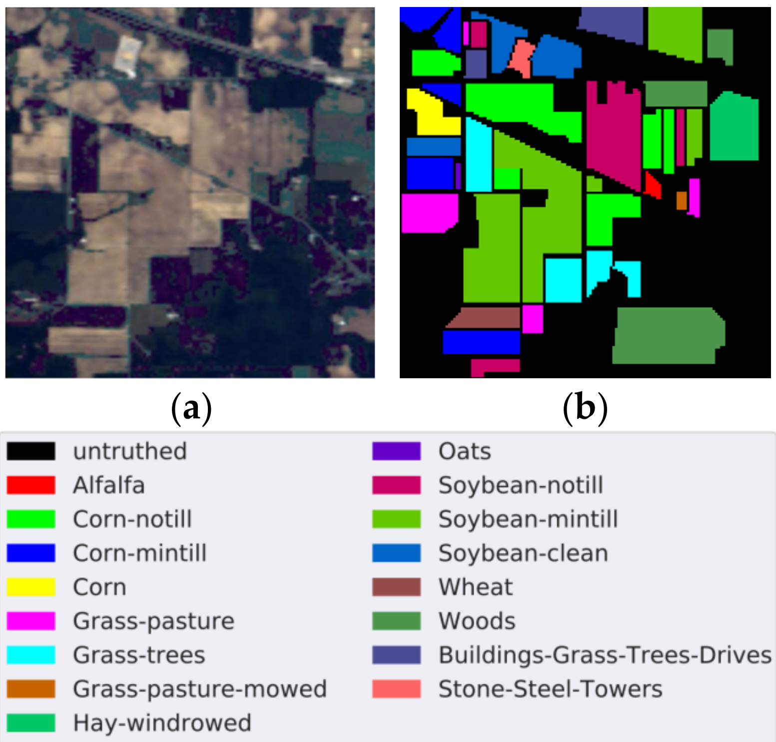


Figure 4. IP data set. (**a**) False-color image. (**b**) Ground-truth map.

Table 1. Land cover classes and the numbers of samples in the IP data set.

Class	Name	Samples	Training Samples	Testing Samples
1	Alfalfa	46	3	43
2	Corn-notill	1428	3	1425
3	Corn-mintill	830	3	827
4	Corn	237	3	234
5	Grass-pasture	483	3	480
6	Grass-trees	730	3	727
7	Grass-pasture-mowed	28	3	25
8	Hay-windrowed	478	3	475
9	Oats	20	3	17
10	Soybean-notill	972	3	969
11	Soybean-mintill	2455	3	2452
12	Soybean-clean	593	3	590
13	Wheat	205	3	202
14	Woods	1265	3	1262
15	Buildings-Grass-Trees-Drives	386	3	383
16	Stone-Steel-Towers	93	3	90
	Total	10,249	48	10,201

The Pavia University (PU) data set contains 610 $\times$ 340 pixels. It has 103 spectral bands that range from 430 to 860 nm. Its spatial resolution is 1.3-m per pixel. It was collected by the Reflective Optics Spectrographic Image System (ROSIS). It contains nine categories representing different types of land cover. Figure 5 shows the false-color image and corresponding ground-truth map of the PU data set. The samples of the PU data set are listed in Table 2.

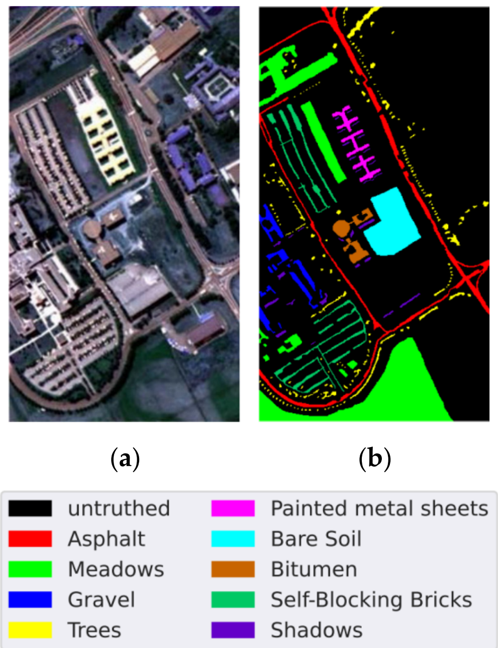


Figure 5. PU data set. (**a**) False-color image. (**b**) Ground-truth map.

Table 2. Land cover classes and the numbers of samples in the PU data set.

Class	Name	Samples	Training Samples	Testing Samples
1	Asphalt	6631	3	6628
2	Meadows	18,649	3	18,646
3	Gravel	2099	3	2096
4	Trees	3064	3	3061
5	Sheets	1345	3	1342
6	Bare soil	5029	3	5026
7	Bitumen	1330	3	1327
8	Bricks	3682	3	3679
9	Shadow	947	3	944
	Total	42,776	27	42,749

The Salinas Scene (SA) data set was gathered by the AVIRIS sensor over Salinas Valley, California. It consists of 145 $\times$ 145 pixels with 204 spectral bands that range from 400 to 2500 nm. The data set contains 16 categories of objects, with a total of 54,129 samples. The false-color image and corresponding ground-truth map of the SA data set are shown in Figure 6. Table 3 lists the samples of the SA data set.

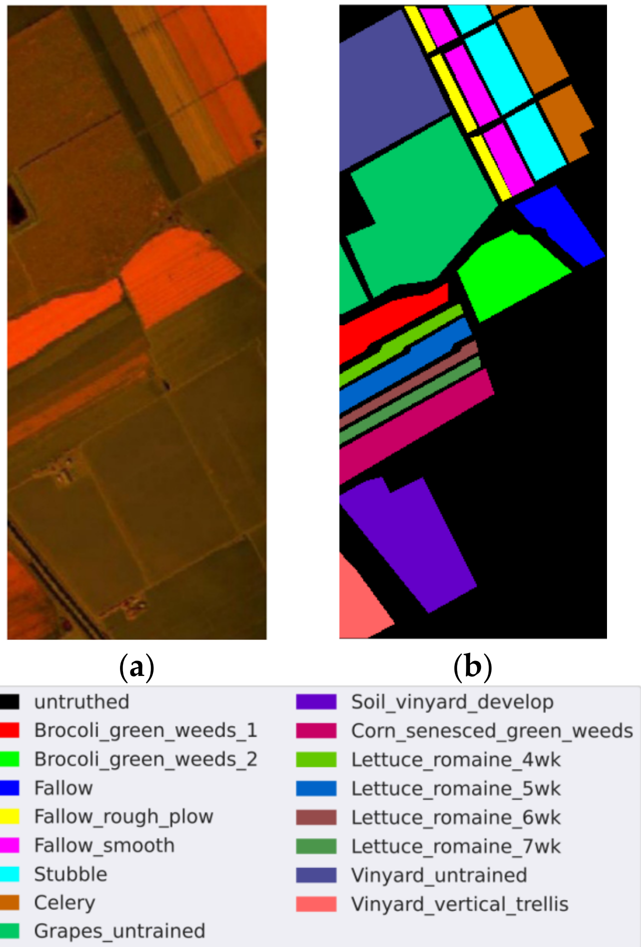


Figure 6. SA data set. (**a**) False-color image. (**b**) Ground-truth map.

Table 3. Land cover classes and the numbers of samples in the SA data set.

Class	Name	Samples	Training Samples	Testing Samples
1	Brocoli_green_weeds_1	2009	3	2006
2	Brocoli_green_weeds_2	3726	3	3723
3	Fallow	1976	3	1973
4	Fallow_rough_plow	1394	3	1391
5	Fallow_smooth	2678	3	2675
6	Stubble	3959	3	3956
7	Celery	3579	3	3576
8	Grapes_untrained	11,271	3	11,268
9	Soil_vinyard_develop	6203	3	6200
10	Corn_senesced_green_weeds	3278	3	3275
11	Lettuce_romaine_4wk	1068	3	1065
12	Lettuce_romaine_5wk	1927	3	1924
13	Lettuce_romaine_6wk	916	3	913
14	Lettuce_romaine_7wk	1070	3	1067
15	Vinyard_untrained	7268	3	7265
16	Vinyard_vertical_trellis	1807	3	1804
	Total	54,129	48	54,081

4.2. Experimental Setup

In our experiment, we mainly perform the small-sample HSI classification to demonstrate the effectiveness and robustness of the proposed fusion framework. Since HSI contains a rich amount of redundant information, we need to preprocess the HSI data to improve the efficiency of subsequent feature extraction. Accroding to [26], PCA algorithm, which is a commonly used strategy for preprocessing the HSI data, is first employed for the dimensionality reduction in HSI to extract the representative band data. After preprocessing, different handcrafted feature-based and CNN-based methods are utilized for the feature extraction of HSI pixels in the experiment.

At the first set of experiments, our proposed 3D-FHOG is mainly compared with the handcrafted feature-based methods including the original spectral feature, extended multi-attribute profile (EMAP) [51], HOG [31], SIFT [13], 3D-LBP [14], 3D-Gabor [17] and 3D-DWT [18], respectively. In our experiment, SVM [52] is applied to classify the feature vectors of HSI pixels. Then, seven CNN-based methods are considered, i.e., semi-1D CNN [21], 3D FCN [23], semi-3D CNN [24], 3D CNN [25], 1D RNN [22], HybridSN [26] and 3DCSN [43], which are utilized to compare with MDSN and the fusion of 3D-FHOG and MDSN (3D-FHOG + MDSN).

In the training process of MDSN, two phases are included. Firstly, the parameter of our proposed model is updated by Adam optimizer [53] and contrastive loss when performing the contrastive learning. Besides, we use an initial learning rate of 5×10^{-3}, and the weight decay is set to be 0 in contrastive learning phase. In the classification training phase, we use another Adam optimizer and cross-entropy loss to update the parameters of our model. Moreover, the learning rate is set to be 1×10^{-3}, and the weight decay is set to be 5×10^{-5} in this phase. According to [43], the input patch size of MDSN is empirically set to be $25 \times 25 \times 30$ for IP and $25 \times 25 \times 15$ for PU and SA, respectively.

In order to compare the classification performance of the above different methods, overall accuracy (OA), average accuracy (AA) and kappa coefficient (κ) are adopted as the evaluation metric. Quantitatively, the greater the values of OA, AA, and κ are, the better the classification result is. Moreover, the classification experiments of each method are repeated 5 times to avoid the accidental phenomenon. Classification accuracy mean and variance of each method are shown in the experimental statistical table.

4.3. Experimental Result and Analysis

In this section, the classification results of different feature extraction methods on three public HSI data sets are analyzed visually and quantitatively.

4.3.1. Influence of the Input Patch Size for 3D-FHOG

First, influence of the input patch size for our proposed 3D-FHOG is examined. Specifically, the size of input patch for 3D-FHOG is set to be $7 \times 7 \times 7$, $9 \times 9 \times 9$, $11 \times 11 \times 11$, $13 \times 13 \times 13$, $15 \times 15 \times 15$ and $17 \times 17 \times 17$ for analysis. Table 4 shows the classification perfor- mance of 3D-FHOG with different input patch sizes on IP data set. It can be concluded that when the size of input patch increases, the values of OA, AA and κ of 3D-FHOG show a trend of increasing first and then decreasing. When input patch size is $15 \times 15 \times 15$, the classification performance of 3D-FHOG is the best. The reason for the decrease in classification performance is the redundant information contained in the local spatial–spectral neighborhood. Therefore, in the following experiment, the input patch size is fixedly set to be $15 \times 15 \times 15$ for 3D-FHOG.

Table 4. Classification performance of 3D-FHOG with different input patch sizes on IP data set.

Evaluation Metric \ Input Patch Size	$7 \times 7 \times 7$	$9 \times 9 \times 9$	$11 \times 11 \times 11$	$13 \times 13 \times 13$	$15 \times 15 \times 15$	$17 \times 17 \times 17$
OA	36.30 ± 0.15	39.70 ± 0.04	42.66 ± 0.03	48.71 ± 0.02	**51.89 ± 0.01**	50.97 ± 0.01
AA	42.27 ± 0.07	46.56 ± 0.05	49.23 ± 0.09	54.23 ± 0.09	**57.61 ± 0.05**	57.12 ± 0.06
κ	28.49 ± 0.17	32.39 ± 0.05	35.70 ± 0.04	42.02 ± 0.03	**45.71 ± 0.01**	44.81 ± 0.01

4.3.2. Compared with Handcrafted Feature-Based Methods

In order to verify the effectiveness of 3D-FHOG in small-sample HSI classification, the proposed 3D-FHOG is compared with seven handcrafted feature-based methods. When three labeled samples per class are adopted as the training set, detail classification results with different handcrafted feature-based methods are listed in Tables 5–7 for IP, PU and SA data sets, respectively. As observed from Tables 5–7, we can make four observations, which are described as follows.

Table 5. Classification results (%) for Spectral, EMAP, HOG, SIFT, 3D-LBP, 3D-Gabor, 3D-DWT, and 3D-FHOG on the test set of IP data set, with three labeled samples per class as training set.

Class	Spectral		EMAP		HOG		SIFT		3D-LBP		3D-Gabor		3D-DWT		3D-FHOG	
	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var	Mean	Var
1	78.60	2.04	73.02	4.29	67.44	4.82	99.53	0.01	100.00	0.00	36.28	2.91	59.53	6.70	73.95	3.19
2	24.60	0.59	11.84	2.15	20.39	0.31	31.90	2.00	27.37	2.07	39.26	3.59	23.24	2.24	28.69	0.40
3	21.69	2.59	34.32	3.42	17.34	0.68	25.92	1.22	23.70	0.85	24.93	1.54	23.43	3.30	23.53	1.15
4	20.85	2.70	33.85	0.91	23.42	0.16	55.47	1.57	44.87	8.25	37.26	5.11	25.30	0.38	31.54	0.95
5	40.92	4.54	34.08	4.10	36.54	0.67	36.75	1.72	15.12	0.50	19.46	5.71	12.79	8.18	58.42	0.30
6	27.54	1.41	69.52	1.36	36.73	1.79	54.22	4.07	49.57	0.92	29.32	2.91	49.90	2.52	67.98	0.94
7	93.60	0.05	89.60	0.37	66.40	3.11	100.00	0.00	100.00	0.00	100.00	0.00	76.00	0.88	72.00	4.35
8	49.39	2.33	38.19	1.13	28.21	1.05	64.21	2.87	71.62	2.74	91.96	0.13	66.90	4.38	63.49	3.59
9	71.76	0.76	64.70	0.69	87.38	1.34	100.00	0.00	100.00	0.00	75.29	1.80	100.00	0.00	94.44	1.23
10	37.21	1.12	36.62	4.73	21.63	0.21	32.57	4.61	34.08	3.93	17.87	7.04	50.18	2.16	46.81	0.92
11	32.28	1.46	37.75	4.18	14.69	0.21	17.35	0.94	26.66	5.48	26.32	10.73	18.67	3.28	67.11	0.16
12	12.27	2.31	11.97	0.38	19.56	0.24	24.54	1.96	31.93	1.12	14.81	1.46	22.85	2.53	28.95	1.45
13	94.36	0.18	92.97	0.30	67.03	1.38	73.66	3.38	96.44	0.09	75.74	0.51	53.17	11.96	74.26	1.71
14	59.30	5.30	59.49	9.23	18.49	0.29	45.91	8.80	62.98	1.61	67.84	8.33	88.73	2.74	63.72	2.45
15	12.64	0.63	28.41	1.36	28.82	0.75	57.18	2.32	36.55	1.91	27.36	0.61	12.74	0.49	46.42	1.74
16	82.00	1.16	91.11	0.07	50.67	3.11	86.67	2.03	96.44	0.39	93.33	0.44	81.33	0.49	80.44	2.62
OA	34.95	0.08	38.50	0.51	22.90	0.02	35.98	0.15	38.60	0.21	36.80	0.32	37.41	0.16	**51.89**	0.01
AA	47.44	0.05	50.46	0.07	37.80	0.05	56.62	0.14	57.33	0.05	48.57	0.05	47.80	0.09	**57.61**	0.05
κ	27.48	0.07	31.79	0.46	15.82	0.01	29.28	0.17	33.77	0.20	29.89	0.24	30.68	0.18	**45.71**	0.01

Table 6. Classification results (%) for Spectral, EMAP, HOG, SIFT, 3D-LBP, 3D-Gabor, 3D-DWT, and 3D-FHOG on the test set of PU data set, with three labeled samples per class as training set.

Class	Spectral		EMAP		HOG		SIFT		3D-LBP		3D-Gabor		3D-DWT		3D-FHOG	
	Mean	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**
1	38.97	10.24	49.72	0.90	24.34	2.03	56.90	1.91	35.26	13.03	17.06	1.93	34.06	3.05	58.95	0.83
2	30.82	10.07	44.82	4.46	21.17	0.18	24.07	4.58	44.38	11.18	29.50	2.82	46.16	3.51	44.84	3.00
3	31.14	0.64	73.39	0.76	38.27	1.60	29.37	0.95	50.13	6.87	25.53	5.45	76.81	1.25	71.87	1.47
4	75.48	2.30	76.67	0.74	54.13	0.66	28.35	3.07	49.63	7.55	56.37	4.45	96.49	0.01	75.51	0.76
5	71.47	0.08	99.15	0.00	55.75	2.11	64.59	2.94	93.02	0.34	99.40	0.01	100.00	0.00	77.03	1.48
6	40.17	11.70	57.73	4.51	26.43	0.79	33.97	1.48	54.91	2.95	77.59	2.68	44.16	9.67	49.71	2.76
7	88.45	0.81	69.15	3.63	30.25	1.02	25.64	0.62	82.14	0.56	80.47	1.02	77.27	10.03	79.74	3.67
8	73.11	0.39	32.06	2.06	57.73	0.47	31.22	2.64	85.67	1.12	50.43	3.94	22.23	0.69	79.36	1.28
9	99.89	0.00	95.05	0.05	85.93	0.48	43.28	2.27	30.95	3.66	33.11	6.88	65.42	0.93	76.52	0.94
OA	44.63	0.63	53.25	0.80	31.42	0.02	33.25	0.43	50.82	3.49	40.61	0.64	50.18	0.42	**56.89**	0.57
AA	61.06	0.18	66.41	0.14	43.78	0.03	37.49	0.08	58.46	0.42	52.16	0.11	62.51	0.09	**68.17**	0.20
κ	35.29	0.42	44.5	0.73	21.64	0.02	21.20	0.11	42.52	3.12	31.16	0.51	40.26	0.33	**48.13**	0.56

Table 7. Classification results (%) for Spectral, EMAP, HOG, SIFT, 3D-LBP, 3D-Gabor, 3D-DWT, and 3D-FHOG on the test set of SA data set, with three labeled samples per class as training set.

Class	Spectral		EMAP		HOG		SIFT		3D-LBP		3D-Gabor		3D-DWT		3D-FHOG	
	Mean	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**
1	98.28	0.02	97.90	0.05	49.29	0.84	41.40	0.78	57.43	10.06	68.32	11.18	95.65	0.05	96.96	0.08
2	70.20	4.32	96.10	0.14	36.37	0.76	37.82	4.48	63.12	0.64	76.93	7.91	85.01	0.12	92.46	0.34
3	49.98	0.40	76.34	2.26	31.42	1.21	26.34	2.72	49.49	4.18	21.28	1.00	50.85	7.75	93.36	0.68
4	98.13	0.02	92.38	0.34	69.89	1.89	85.74	0.73	95.87	0.01	77.94	6.14	97.44	0.01	84.02	2.35
5	97.70	0.00	81.14	4.44	50.31	2.77	31.50	2.25	93.81	0.05	60.26	5.94	79.58	10.32	78.65	8.17
6	96.67	0.02	99.59	0.00	50.69	1.76	51.69	2.30	56.15	10.42	90.36	0.02	77.50	17.58	97.56	0.04
7	97.75	0.05	99.61	0.00	40.75	1.07	18.79	0.23	85.70	0.05	32.38	6.32	64.80	1.68	95.22	0.05
8	45.40	1.86	44.12	4.26	25.96	1.05	28.69	10.42	39.31	10.00	32.26	11.49	46.18	13.01	68.29	0.41
9	75.46	7.87	90.75	1.64	35.71	1.04	14.99	0.71	75.41	3.07	97.64	0.01	75.34	17.42	100.00	0.01
10	29.34	6.40	63.81	3.48	36.71	4.02	18.04	1.09	64.05	1.06	22.00	4.32	25.95	6.86	64.18	5.06
11	77.35	0.18	80.02	2.14	63.83	3.27	89.56	0.06	92.60	0.00	61.67	7.09	78.33	4.84	76.73	2.54
12	73.79	0.40	75.55	1.79	62.85	3.87	60.57	9.16	61.23	13.09	54.07	4.06	65.97	1.20	81.05	0.43
13	95.64	0.36	76.12	3.00	64.82	2.13	39.01	6.72	36.17	16.42	95.44	0.06	89.18	0.02	89.35	0.90
14	76.94	3.91	75.35	6.23	75.00	2.57	77.56	0.76	74.13	1.31	62.38	10.54	66.64	1.66	82.19	2.28
15	66.34	2.49	73.11	3.68	18.79	0.43	45.66	10.26	77.79	1.12	63.14	8.10	40.54	8.41	38.67	1.68
16	20.45	0.95	72.66	1.57	26.53	1.46	24.70	0.98	41.35	4.88	52.93	0.37	21.44	4.00	57.35	0.66
OA	67.96	0.19	76.03	0.14	37.37	0.02	35.74	0.17	63.78	0.18	57.82	0.10	60.35	0.71	**76.95**	0.03
AA	73.09	0.05	80.91	0.16	46.18	0.09	43.25	0.06	66.48	0.16	60.56	0.12	66.27	0.54	**80.94**	0.08
κ	64.59	0.23	73.51	0.16	31.49	0.04	29.70	0.14	60.14	0.19	53.64	0.10	56.35	0.84	**74.46**	0.04

Firstly, compared with the 2D handcrafted feature descriptors, spectral feature and 3D handcrafted feature descriptors can achieve better classification performance. For instance, the OA of 3D-LBP is 15.70% higher than that of HOG on the IP data set. This indicates that only extracting the spatial information will destroy the correlation between spectral bands.

Secondly, the classification performance of 3D handcrafted feature descriptors is always superior to the spectral feature. Because 3D handcrafted feature descriptors can exploit both spatial and spectral information, which makes them more effective in HSI classification.

Thirdly, EMAP feature descriptor shows more excellent classification performance on PU and SA data sets. Since EMAP is based on morphological attribute filters and multi-level analysis, it is suggested that combining the features with different dimensions or scales is an effective solution for small-sample HSI classification.

Finally, by comparing with spectral, 2D handcrafted and 3D handcrafted feature descriptors, we can observe that our proposed 3D-FHOG feature descriptor obtains the best classification results in three public data sets. By integrating with fuzzy logic, 3D-FHOG is able to overcome the local spatial–spectral feature uncertainty and extract more discriminative spatial–spectral features.

4.3.3. Compared with CNN-Based Methods

To further demonstrate the effectiveness and robustness of the proposed fusion framework, 3D-FHOG + MDSN method is compared with eight representative CNN- based methods. These CNN-based methods include Semi-1D CNN, 3D FCN, Semi-3D CNN, 3D CNN, 1D RNN, HybridSN, 3DCSN and MDSN. Table 8 reports the OA, AA, κ and the classification accuracy of each class for HSI classification with three labeled samples per class on the IP data set. The statistical results suggest that HybridSN, 3DCSN and MDSN methods, which incorporate the multidimensional CNN feature, are superior to the single-dimensional CNN-based methods. Additionally, both Semi-3D CNN and 1D RNN can achieve excellent classification performance. This verifies that 3D CNN combined with semi-supervised learning is effective for small-sample HSI classification, and 1D RNN is able to take full advantage of spectral correlation and band-to-band variability. Moreover, MDSN and 3DCSN, which are based on the learning mechanism of Siamese network, can obtain more accurate results. Especially, 3D-FHOG + MDSN method has the best classification results on the IP data set, which indicates that our proposed fusion framework can fully combine the advantage of multidimensional CNN and handcrafted features.

Table 8. Classification results (%) for eight different CNN-based methods and 3D-FHOG + MDSN on the test set of IP data set, with three labeled samples per class as training set.

Class	Semi-1D CNN		3D FCN		Semi-3D CNN		3D CNN		1D RNN		HybridSN		3DCSN		MDSN		3D-FHOG +MDSN	
	Mean	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**
1	3.22	0.09	4.74	0.22	42.14	2.70	7.80	0.57	24.69	1.00	73.49	1.35	100.00	0.00	100.00	0.00	99.53	0.01
2	6.36	0.68	4.96	0.55	28.22	1.21	10.85	1.98	24.71	0.22	19.87	1.14	46.29	0.03	53.38	0.22	51.21	0.10
3	8.33	0.95	10.88	1.58	20.13	0.48	5.09	0.40	23.73	0.52	27.64	0.65	53.01	1.49	59.54	0.19	63.41	0.22
4	3.81	0.11	12.17	0.79	22.93	0.86	12.16	1.00	18.54	0.70	17.61	0.80	50.60	0.54	57.52	0.38	55.13	0.52
5	11.33	2.96	6.91	1.20	42.58	0.33	15.32	1.15	46.47	0.15	58.46	1.55	55.50	0.04	54.58	0.02	57.54	0.02
6	7.14	2.04	3.03	0.29	69.21	0.94	15.00	4.01	55.14	2.00	70.84	2.34	88.09	0.21	90.76	0.02	91.77	0.02
7	7.99	0.49	4.00	0.05	40.29	2.39	0.75	0.02	29.69	1.08	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00
8	8.80	0.69	7.49	2.24	78.91	2.95	15.47	9.57	63.07	3.37	89.39	0.19	99.12	0.02	98.36	0.01	99.49	0.00
9	2.14	0.03	4.66	0.06	25.82	0.29	4.67	0.34	23.85	3.75	95.29	0.33	100.00	0.00	100.00	0.00	100.00	0.00
10	5.87	0.52	11.50	1.42	32.52	0.13	8.41	2.34	29.13	0.44	24.99	3.59	58.39	0.13	58.02	0.10	59.86	0.09
11	25.22	2.83	10.57	4.47	38.19	3.38	20.45	3.52	26.45	1.59	17.19	1.99	49.35	0.63	45.53	0.11	53.60	0.08
12	9.83	0.71	5.14	0.31	20.00	0.07	7.79	1.45	17.44	0.18	30.61	1.00	60.10	0.31	55.05	0.27	62.14	0.02
13	24.44	2.92	18.83	1.93	65.11	0.26	25.52	7.23	65.46	2.49	81.88	7.62	81.78	0.87	87.23	0.11	83.56	0.16
14	7.16	1.59	34.84	13.17	74.91	1.32	46.28	14.50	66.19	0.34	65.55	0.38	83.00	0.35	85.48	0.22	85.86	0.25
15	9.03	0.46	7.72	0.65	24.92	0.23	8.04	0.68	27.61	0.45	50.39	2.92	76.60	0.00	76.66	0.00	76.76	0.00
16	11.64	4.40	46.29	6.65	74.71	3.11	7.09	0.23	52.40	2.57	59.56	0.92	72.67	0.05	72.00	0.16	55.56	0.05
OA	13.98	0.25	15.22	1.68	42.95	0.12	20.80	1.02	35.67	0.10	38.52	0.50	62.55	0.06	63.51	0.01	**66.08**	0.00
AA	9.52	0.06	12.11	0.58	43.79	0.04	13.17	0.69	37.16	0.15	55.17	0.25	73.40	0.02	74.63	0.00	**74.72**	0.01
κ	7.62	0.12	10.62	1.07	36.23	0.08	14.92	0.84	29.13	0.08	33.76	0.45	58.09	0.07	59.32	0.01	**62.07**	0.00

As for the PU data set, detail classification results with nine different methods are listed in Table 9. As observed in Table 9, the performance of MDSN, 3DCSN and HybridSN is significantly higher than that of the Semi-1D CNN, 3D FCN, Semi-3D CNN, 3D CNN and 1D RNN on the PU data set, which further demonstrates the superiority of multidimensional CNN features and learning mechanism of Siamese network in small- sample HSI classification. Furthermore, 3D-FHOG + MDSN consistently provides the best classification results on PU data set.

Table 9. Classification results (%) for eight different CNN-based methods and 3D-FHOG + MDSN on the test set of PU data set, with three labeled samples per class as training set.

Class	Semi-1D CNN		3D FCN		Semi-3D CNN		3D CNN		1D RNN		HybridSN		3DCSN		MDSN		3D-FHOG +MDSN	
	Mean	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**
1	1.00	0.02	33.05	4.85	61.66	1.80	75.39	0.31	42.33	9.42	53.59	1.17	52.58	0.50	63.63	0.35	67.44	0.24
2	11.90	0.86	39.22	5.49	41.14	4.97	55.24	1.37	44.65	1.35	67.38	2.70	65.48	0.27	63.51	0.33	64.01	0.23
3	11.07	1.05	18.40	1.35	33.01	0.27	17.02	1.86	19.47	5.48	76.12	1.22	87.97	0.78	87.36	0.27	89.83	0.11
4	11.93	1.04	49.64	1.66	41.05	0.87	48.31	5.91	58.79	0.54	18.63	0.11	34.45	1.93	37.84	0.56	41.80	0.58
5	23.68	8.97	80.47	6.07	73.99	2.68	99.39	0.00	73.57	2.84	100.00	0.00	99.73	0.00	99.97	0.00	99.79	0.00
6	10.74	1.74	34.57	0.06	34.79	0.17	36.20	0.09	25.79	0.03	77.96	0.92	74.46	2.07	78.12	0.86	79.32	0.81
7	13.05	2.12	26.46	1.90	33.64	0.78	22.89	3.03	21.71	0.85	74.35	1.33	94.06	0.52	91.94	0.54	97.66	0.08
8	10.35	3.32	32.58	0.56	50.49	0.74	30.97	8.60	38.20	5.41	47.12	1.78	61.66	0.20	69.22	0.34	70.55	0.20
9	1.55	0.04	61.04	11.43	62.02	1.84	97.49	0.08	54.68	1.19	46.46	1.26	71.48	0.18	69.56	0.06	74.02	0.11
OA	13.88	0.32	39.43	0.60	44.47	0.42	53.76	0.54	41.66	0.72	62.46	0.59	65.18	0.09	67.23	0.11	**68.97**	0.09
AA	10.58	0.04	41.71	0.30	47.98	0.13	53.66	0.39	42.13	0.37	62.40	0.10	71.32	0.08	73.46	0.03	**76.05**	0.04
κ	5.63	0.06	29.30	0.46	35.81	0.24	43.91	0.57	32.86	0.62	53.73	0.61	56.97	0.12	59.58	0.13	**61.62**	0.11

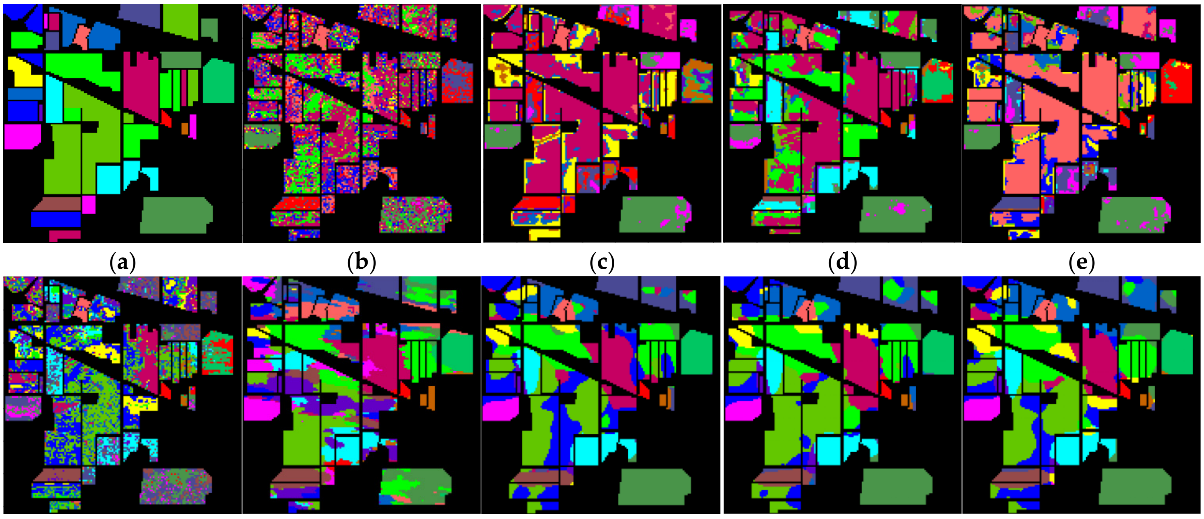
With respect to the SA data set, for three labeled samples per class, the OA, AA and κ measure for each class using different approaches are shown in Table 10. From Table 10, it is found that OA and κ of the proposed 3D-FHOG + MDSN are 92.06% and 91.16, respectively, in comparison with the OA and κ of 24.31% and 19.30, 32.38% and 28.11, 64.79% and 60.91, 32.34% and 27.07, 67.42% and 63.99, 84.07% and 82.29, 91.69% and 90.74, 91.72% and 90.78 for Semi-1D CNN, 3D FCN, Semi-3D CNN, 3D CNN, 1D RNN, HybridSN, 3DCSN and MDSN, respectively. The same conclusion can be drawn that classification accuracies obtained by 3D-FHOG + MDSN are better than others, which in turn demonstrates the effectiveness of our proposed fusion framework.

Table 10. Classification results (%) for eight different CNN-based methods and 3D-FHOG + MDSN on the test set of SA data set, with three labeled samples per class as training set.

Class	Semi-1D CNN		3D FCN		Semi-3D CNN		3D CNN		1D RNN		HybridSN		3DCSN		MDSN		3D-FHOG +MDSN	
	Mean	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**	**Mean**	**Var**
1	0.00	0.00	27.60	5.49	29.38	5.77	19.72	2.59	44.33	5.10	98.99	0.01	99.97	0.00	99.79	0.00	99.78	0.00
2	21.46	3.46	34.09	4.30	32.76	12.28	21.44	6.90	38.63	13.56	98.30	0.08	99.40	0.00	99.30	0.01	99.54	0.00
3	13.73	1.40	2.15	0.11	51.79	1.21	8.15	0.69	46.95	6.78	81.88	1.01	99.83	0.00	99.91	0.01	99.94	0.00
4	30.45	10.00	60.99	12.46	92.24	0.34	32.08	10.37	91.74	0.12	66.77	7.51	99.15	0.00	99.08	0.01	99.61	0.00
5	14.05	7.90	16.52	3.68	68.87	3.20	28.87	11.02	58.39	15.68	57.70	4.99	94.88	0.34	94.49	0.15	95.24	0.16
6	30.00	8.63	65.66	9.94	98.60	0.00	76.07	3.42	98.96	0.00	99.42	0.00	98.59	0.02	99.14	0.01	99.67	0.00
7	26.15	7.92	8.12	2.64	93.31	0.03	0.03	0.00	93.07	0.22	96.26	0.10	98.90	0.01	99.08	0.01	99.50	0.00
8	12.54	5.76	11.91	2.25	60.37	3.67	30.07	7.50	59.36	3.60	76.19	0.85	85.97	0.01	84.86	0.03	84.47	0.02
9	33.27	6.81	29.99	7.45	81.87	0.34	41.37	13.06	72.43	13.37	95.95	0.09	96.61	0.21	97.12	0.26	99.30	0.01
10	10.73	0.90	8.20	0.75	38.28	0.39	15.78	4.98	58.19	5.03	87.66	0.13	95.47	0.02	94.54	0.02	94.52	0.01
11	2.13	0.18	27.21	5.42	37.03	0.68	10.97	2.49	36.12	6.11	97.39	0.01	99.66	0.00	99.59	0.00	99.91	0.00
12	30.11	6.40	31.57	2.50	60.37	11.17	23.35	5.02	72.69	7.13	80.83	0.40	98.34	0.02	98.40	0.02	98.11	0.01
13	15.41	1.88	17.41	4.81	80.12	1.26	42.65	8.33	89.82	0.54	95.60	0.67	99.87	0.00	99.96	0.00	99.96	0.00
14	41.57	13.93	13.52	0.13	70.73	4.34	26.72	10.96	86.54	0.38	87.52	1.69	98.67	0.00	98.74	0.01	99.02	0.00
15	6.36	0.96	34.87	3.76	44.37	2.95	2.28	0.20	39.59	3.99	65.97	2.80	71.67	0.29	72.73	0.52	72.82	0.45
16	13.11	1.00	43.96	0.21	49.94	6.54	11.81	4.48	79.97	0.15	97.55	0.10	90.30	0.42	93.10	0.22	93.72	0.11
OA	24.31	0.40	32.38	0.45	64.79	0.08	32.34	2.52	67.42	0.34	84.07	0.01	91.69	0.00	91.72	0.01	**92.06**	0.01
AA	18.82	0.36	27.11	0.08	61.88	0.23	24.46	2.03	66.67	0.48	86.50	0.02	95.45	0.01	95.61	0.00	**95.94**	0.01
κ	19.30	0.32	28.11	0.34	60.91	0.10	27.07	2.48	63.99	0.38	82.29	0.01	90.74	0.00	90.78	0.01	**91.16**	0.01

4.3.4. Classification Maps

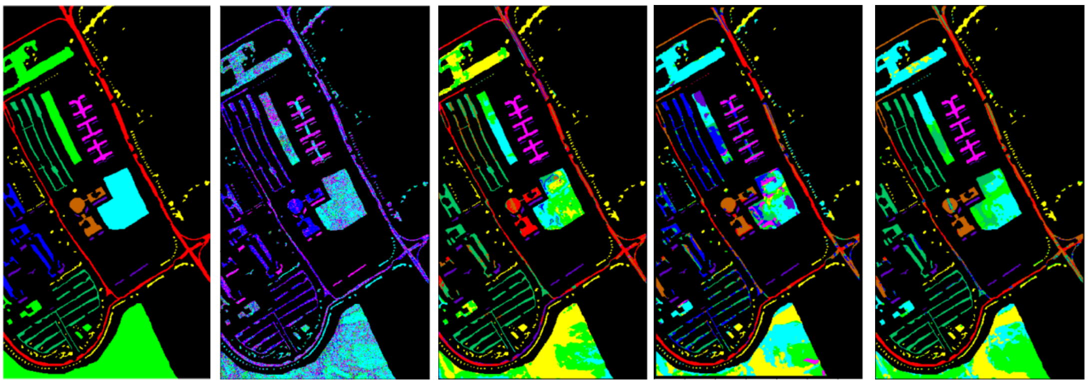
Furthermore, classification performances of nine different methods are visually investigated on three public HSI data sets. Figure 7 shows the classification maps of Semi-1D CNN, 3D FCN, Semi-3D CNN, 3D CNN, 1D RNN, HybridSN, 3DCSN, MDSN and 3D-FHOG + MDSN on the IP data set with three labeled samples per class. Comparing the classification maps of each method in Figure 7, the maps of HybridSN, 3DCSN, MDSN and 3D-FHOG + MDSN are obviously more similar to the ground map than those of other methods. Especially, the map of 3D-FHOG + MDSN, which effectively combines the multidimensional CNN and handcrafted features, is the most similar. Additionally, the classification maps on the PU data set using nine different methods with three labeled samples per class are shown in Figure 8. It can be seen from Figure 8 that more query samples are obviously assigned to the correct class on the maps of multidimensional CNN-based methods than others, and the map of 3D-FHOG + MDSN is more consistent with the ground-truth map, which indicates that the performance of MDSN is effectively enhanced with the 3D-FHOG feature. In terms of the SA data set, Figure 9 displays the classification maps resulting from nine different methods for the SA data set with three labeled samples per class. The same conclusion can be drawn that the map of 3D-FHOG + MDSN is more similar to the ground-truth map than other methods, which further shows its robustness in small-sample HSI classification.



(**a**) (**b**) (**c**) (**d**) (**e**)

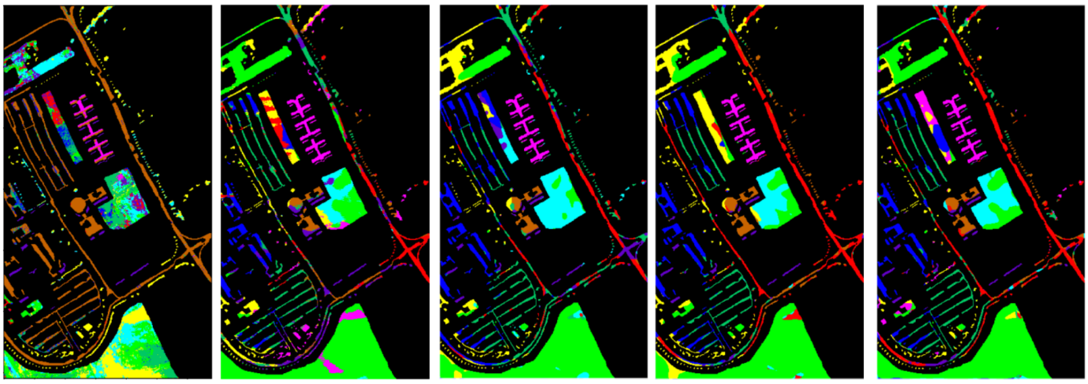
(**f**) (**g**) (**h**) (**i**) (**j**)

Figure 7. Classification maps resulting from nine different methods for the IP data set with three labeled samples per class. (**a**) Ground-truth. (**b**) Semi-1D CNN (16.32%). (**c**) 3D FCN (21.79%). (**d**) Semi-3D CNN (41.64%). (**e**) 3D CNN (17.78%). (**f**) 1D RNN (34.37%). (**g**) HybridSN (47.51%). (**h**) 3DCSN (61.20%). (**i**) MDSN (63.51%). (**j**) 3D-FHOG + MDSN (66.42%).



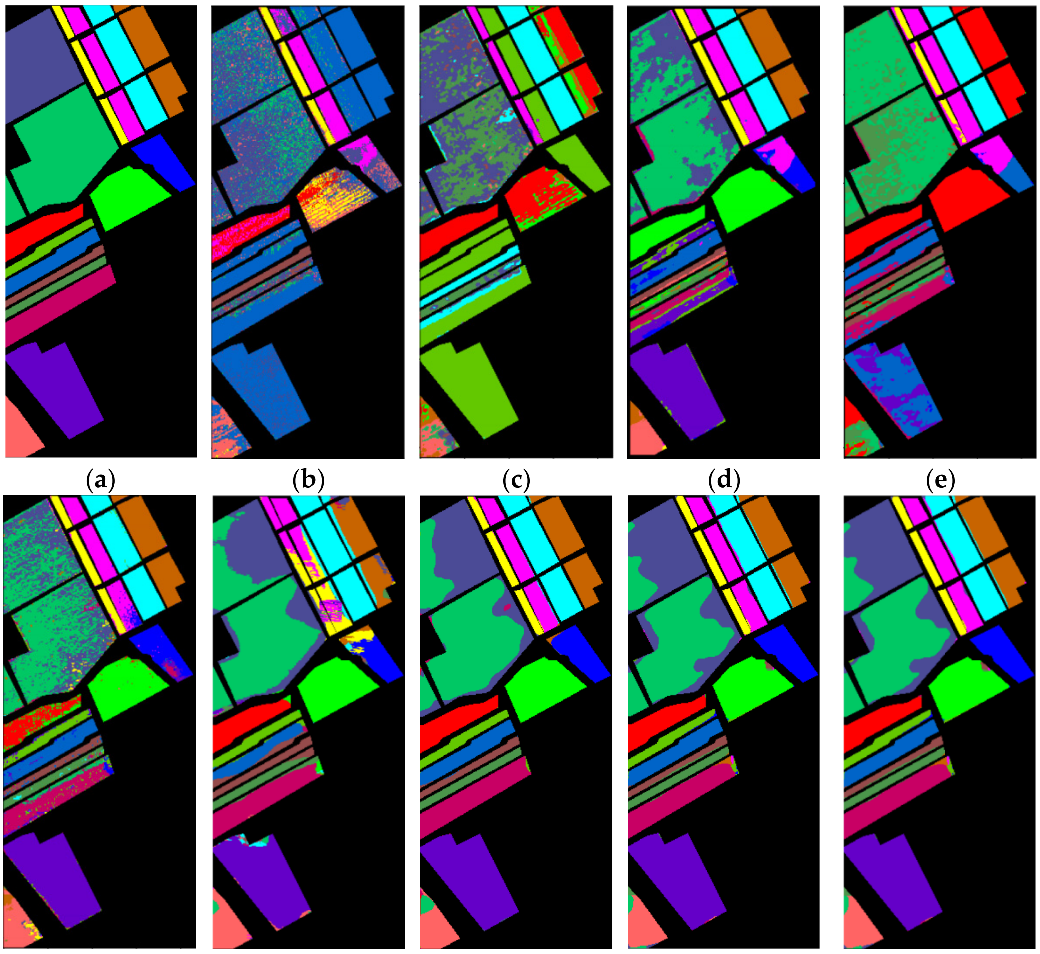
(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 8. *Cont.*



(**f**) (**g**) (**h**) (**i**) (**j**)

Figure 8. Classification maps resulting from nine different methods for the PU data set with three labeled samples per class. (**a**) Ground-truth. (**b**) Semi-1D CNN (18.25%). (**c**) 3D FCN (50.48%). (**d**) Semi-3D CNN (35.54%). (**e**) 3D CNN (52.10%). (**f**) 1D RNN (28.85%). (**g**) HybridSN (64.06%). (**h**) 3DCSN (65.08%). (**i**) MDSN (68.39%). (**j**) 3D-FHOG + MDSN (73.29%).



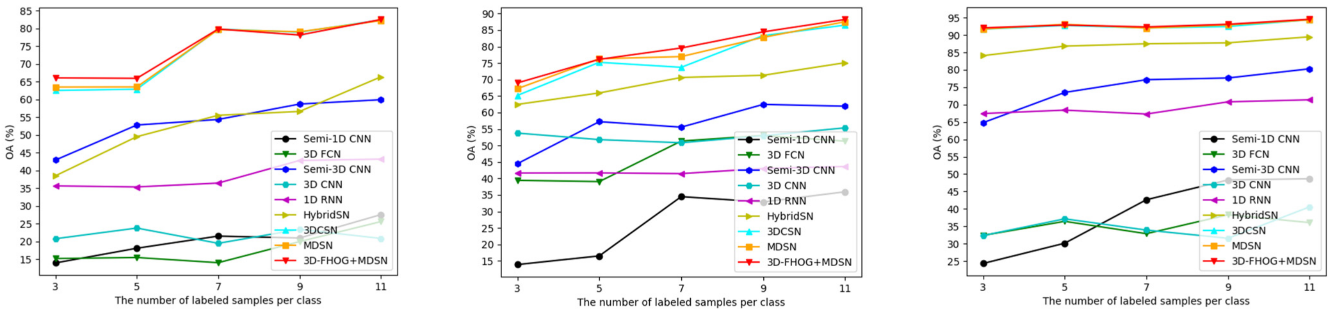
(**a**) (**b**) (**c**) (**d**) (**e**)

(**f**) (**g**) (**h**) (**i**) (**j**)

Figure 9. Classification maps resulting from nine different methods for the SA data set with three labeled samples per class. (**a**) Ground-truth. (**b**) Semi-1D CNN (26.94%). (**c**) 3D FCN (28.18%). (**d**) Semi-3D CNN (72.80%). (**e**) 3D CNN (42.56%). (**f**) 1D RNN (80.14%). (**g**) HybridSN (82.34%). (**h**) 3DCSN (91.30%). (**i**) MDSN (91.93%). (**j**) 3D-FHOG + MDSN (93.25%).

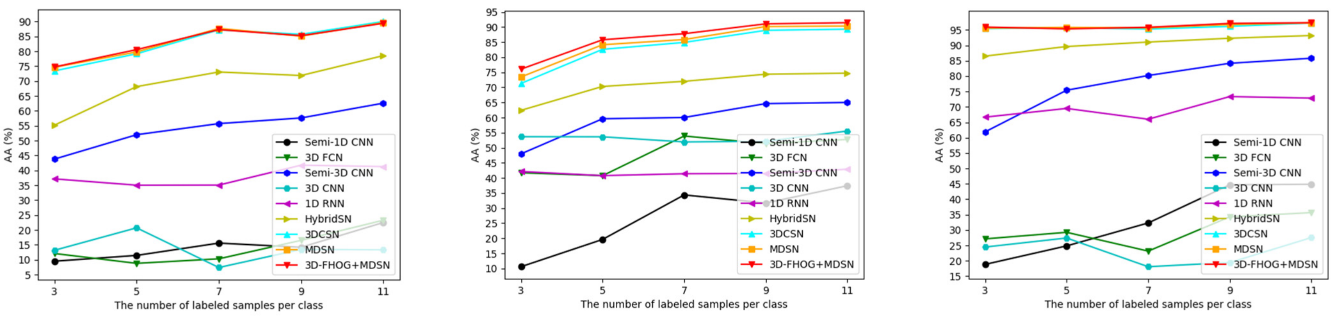
4.3.5. Influence of Training Sample Size

To further illustrate the superiority of 3D-FHOG + MDSN with different numbers of labeled samples, we take 3, 5, 7, 9, and 11 labeled samples for each class to build the training data set. Specifically, we have conducted three groups of experiment on three public HSI data sets. Then, the curve change of nine methods is obtained. It can be seen from Figure 10 that the OA of each method generally rises as the number of labeled samples increases. However, single-dimensional CNN-based methods are unstable in the scenery of small-scale training samples, which may result in the sharp decrease in classification accuracy when the number of labeled samples increases. Especially, the proposed 3D-FHOG + MDSN method outperforms other methods in most cases, which demonstrates its adaptability to the variance of the number of labeled samples. Additionally, Figures 11 and 12 display the AA and κ measure as functions of the number of labeled samples per class. The same conclusion can be drawn that the AA and κ of our proposed 3D-FHOG + MDSN method is always the best in terms of different training sample sizes. Besides, we also find that the gap between the classification accuracy of various methods increases when the number of training samples is fewer. This indicates that the performance of CNN-based method will be enhanced with the increase in labeled samples, which in turns minimizes the contribution of handcrafted features. Meanwhile, incorporating the handcrafted feature with the CNN-based method will not cause a decrease in classification accuracy, but will improve the reliability of classification result. Hence, it is suggested that our proposed 3D-FHOG + MDSN method is more robust and reliable in the scenery of small-scale training samples.



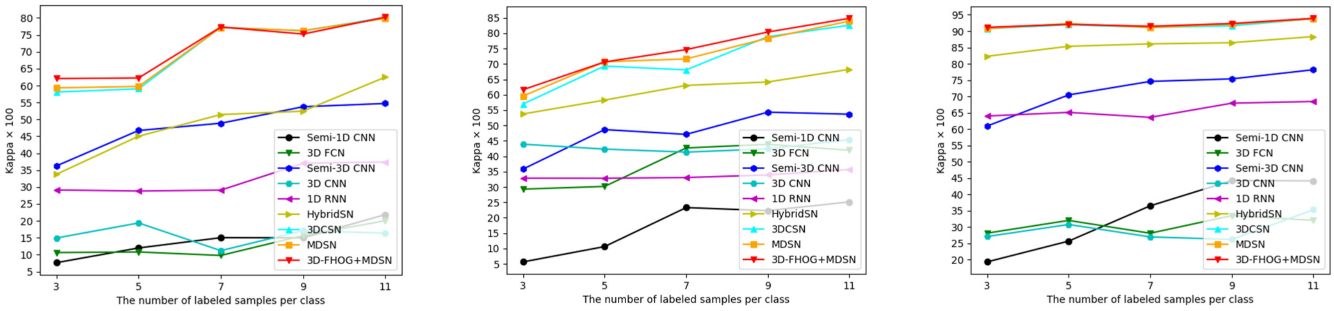
(**a**) (**b**) (**c**)

Figure 10. OA as functions of the number of labeled samples per class on three test data sets: (**a**) IP data set; (**b**) PU data set; (**c**) SA data set.



(**a**) (**b**) (**c**)

Figure 11. AA as functions of the number of labeled samples per class on three test data sets: (**a**) IP data set; (**b**) PU data set; (**c**) SA data set.



(**a**) (**b**) (**c**)

Figure 12. Kappa coefficient as functions of the number of labeled samples per class on three test data sets: (**a**) IP data set; (**b**) PU data set; (**c**) SA data set.

In summary, the performance of MDSN enhanced with 3D-FHOG feature in small-sample HSI classification is better than those of representative handcrafted and CNN-based spatial–spectral feature extraction methods, especially when the training sample size is smaller. This in turn verifies the effectiveness of the proposed fusion framework.

4.3.6. Time Consumption

In this section, the running time of different methods is analyzed to evaluate their computational efficiency. Table 11 reports the running time of different methods on the three HSI data sets with three labeled samples per class. All the experiments are conducted on a computer with an Intel Core i3-4160 processor with 3.6 GHz, 8 GB of DDR3 RAM, an NVIDIA Geforce RTX 1060 graphical processing unit (GPU). For the higher computational cost methods including Semi-1D CNN, 3D FCN and Semi-3D CNN, the processing time is long. In terms of 3D CNN, it has a relatively short training time, but achieves poor classification performance. In addition, for the lightweight networks (i.e., 1D RNN and HybridSN), the running time is short, and these methods can obtain better classification results. Additionally, for the 3DCSN, MDSN and 3D-FHOG+ MDSN, since these methods are based on the idea of Siamese network and composed of CNN blocks with multiple dimensions, more time is consumed by learning the multi- dimensional CNN features from the HSI pixel pairs. Meanwhile, the classification performances of these methods are effectively improved.

Table 11. Running time of different methods on the three HSI data sets with three labeled samples per class.

	Model	Semi-1D CNN	3D FCN	Semi-3D CNN	3D CNN	1D RNN	HybridSN	3D CSN	MDSN	3D-FHOG + MDSN
IP	Training Time (s)	449.72	197.29	500.10	21.56	9.76	4.09	295.81	314.16	314.17
	Testing time (s)	0.44	15.04	3.91	4.81	1.46	10.36	10.58	12.58	15.50
PU	Training Time (s)	3659.20	735.33	3710.15	37.50	23.03	1.77	34.64	38.66	38.58
	Testing time (s)	3.53	140.01	23.22	19.21	7.88	19.98	20.46	26.58	41.78
SA	Training Time (s)	2443.70	990.52	2893.45	95.27	39.32	2.18	102.17	114.46	115.28
	Testing time (s)	2.37	83.36	29.18	27.51	7.78	24.91	25.54	34.93	64.84

Especially, our proposed 3D-FHOG + MDSN contains the additional time of handcrafted feature extraction (HFE). Table 12 reports the HFE time of 3D-FHOG + MDSN on the three HSI data sets with three labeled samples per class. Note that the HFE time represents the time of HFE for all HSI samples in the data set. Hence, the more samples the data set contains, the more time it takes for 3D-FHOG feature extraction. In some high-sensitive areas, we can spend more time to obtain the more reliable and accurate results. Note that HFE time for each pixel is 0.07s. In real military applications, a military target contained in the HSI is generally composed of about 100 pixels, which only takes 7s for HFE. Therefore, the increase in time is acceptable. To sum up, in terms of some special small-sample HSI classification tasks without considering the computational cost, our proposed method will be an effective solution to achieve more accurate and reliable classification results.

Table 12. Handcrafted feature extraction (HFE) time of 3D-FHOG + MDSN on the three HSI data sets with three labeled samples per class.

	Model	3D-FHOG + MDSN
IP	HFE time (s)	722.57
	HFE time for each pixel (s)	0.07
PU	HFE time (s)	3040.99
	HFE time for each pixel (s)	0.07
SA	HFE time (s)	3815.43
	HFE time for each pixel (s)	0.07

5. Conclusions

In this paper, a fusion framework of multidimensional CNN and handcrafted features is proposed for small-sample HSI classification. Specifically, we design the 3D-FHOG descriptor to extract the handcrafted spatial–spectral feature, which is suggested to be more robust by overcoming the local spatial–spectral feature uncertainty. Then, to further extract the CNN-based spatial–spectral feature, an effective Siamese network, i.e., MDSN is proposed, which can effectively achieve the integration of CNN-based spatial–spectral features from multiple dimensions. Finally, our proposed MDSN combined with 3D-FHOG is employed for small-sample HSI classification to verify the effectiveness of our proposed fusion framework. Experiment results on three public HSI data sets indicate that our proposed MDSN combined with 3D-FHOG is superior to the representative handcrafted and CNN-based spatial–spectral feature extraction methods, which in turn demonstrates the effectiveness of the proposed fusion framework. More importantly, our proposed fusion framework has the advantage of expandability. In the future work, we will continue to explore the more discriminative and efficient spatial–spectral feature extraction methods, and integrate them into our proposed fusion framework, which helps to improve the small-sample HSI classification accuracy.

Author Contributions: Methodology, H.T. and Y.L.; investigation, Z.H., L.Z. and W.X.; resources, Y.L., L.Z. and W.X.; writing—original draft preparation, H.T.; writing—review and editing, H.T.,Y.L. and Z.H.; visualization, H.T. and Z.H.; supervision, Y.L., L.Z. and W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by National Natural Science Foundation of China (Nos. 61771319, 62076165, 61871154), Natural Science Foundation of Guangdong Province (No. 2019A1515011307), Shenzhen Science and Technology Project (No. JCYJ20180507182259896, No. 20200826154022001) and the Other Project (Nos. 2020KCXTD004, WDZC20195500201).

Data Availability Statement: The IP, PU and SA data sets can be obtained from http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (accessed on 28 June 2022).

Acknowledgments: We would like to thank the authors for providing the data used in this study. We would also like to thank all the professionals for kindly providing the codes associated with the experiments.

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

References

1. Chakraborty, T.; Trehan, U. Spectralnet: Exploring spatial-spectral waveletcnn for hyperspectral image classification. *arXiv* **2021**, arXiv:2104.00341.
2. Zhang, Y.; Liu, K.; Dong, Y.; Wu, K.; Hu, X. Semisupervised classification based on SLIC segmentation for hyperspectral image. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1440–1444. [CrossRef]
3. Li, Y.; Tang, H.; Xie, W.; Luo, W. Multidimensional local binary pattern for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–13. [CrossRef]

4. Luo, F.; Zou, Z.; Liu, J.; Lin, Z. Dimensionality reduction and classification of hyperspectral image via multistructure unified discriminative embedding. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [CrossRef]
5. Zhou, Y.; Chen, P.; Liu, N.; Yin, Q.; Zhang, F. Graph-Embedding Balanced Transfer Subspace Learning for Hyperspectral Cross-Scene Classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2022**, *15*, 2944–2955. [CrossRef]
6. Ye, M.; Qian, Y.; Zhou, J.; Tang, Y. Dictionary learning-based feature-level domain adaptation for cross-scene hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1544–1562. [CrossRef]
7. Mustafa, G.; Zheng, H.; Khan, I.H.; Tian, L.; Jia, H.; Li, G.; Cheng, T.; Tian, Y.; Cao, W.; Zhu, Y.; et al. Hyperspectral Reflectance Proxies to Diagnose In-Field Fusarium Head Blight in Wheat with Machine Learning. *Remote Sens.* **2022**, *14*, 2784. [CrossRef]
8. Zhang, N.; Yang, G.; Pan, Y.; Yang, X.; Chen, L.; Zhao, C. A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. *Remote Sens.* **2020**, *12*, 3188. [CrossRef]
9. Uzair, M.; Mahmood, A.; Mian, A. Hyperspectral Face Recognition with Spatiospectral Information Fusion and PLS Regression. *IEEE Trans. Image Process.* **2015**, *24*, 1127–1137. [CrossRef]
10. Zhang, X.; Zhao, H. Hyperspectral-cube-based mobile face recognition: A comprehensive review. *Inf. Fusion* **2021**, *74*, 132–150. [CrossRef]
11. Dobler, G.; Ghandehari, M.; Koonin, S.E.; Sharma, M.S. A hyperspectral survey of New York City lighting technology. *Sensors* **2016**, *16*, 2047. [CrossRef] [PubMed]
12. Baur, J.; Dobler, G.; Bianco, F.; Sharma, M.; Karpf, A. Persistent hyperspectral observations of the urban lightscape. In Proceedings of the 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Anaheim, CA, USA, 26–29 November 2018; pp. 983–987.
13. Lowe, D. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
14. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [CrossRef]
15. Zhao, G.; Pietikainen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 915–928. [CrossRef]
16. Jia, S.; Hu, J.; Zhu, J.; Jia, X.; Li, Q. Three-dimensional local binary patterns for hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2399–2413. [CrossRef]
17. He, L.; Li, J.; Plaza, A.; Li, Y. Discriminative Low-Rank Gabor Filtering for Spectral-Spatial Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1381–1395. [CrossRef]
18. Cao, X.; Xu, L.; Meng, D.; Zhao, Q.; Xu, Z. Integration of 3-dimensional discrete wavelet transform and Markov random field for hyperspectral image classification. *Neurocomputing* **2017**, *226*, 90–100. [CrossRef]
19. Sharma, V.; Diba, A.; Tuytelaars, T.; Gool, L. *Hyperspectral CNN for Image Classification & Band Selection, with Application to Face Recognition*; Tech. Rep. KUL/ESAT/PSI/1604; KU Leuven: Leuven, Belgium, 2016.
20. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep convolutional neural networks for hyperspectral image classification. *J. Sens.* **2015**, *2015*, 1–12. [CrossRef]
21. Boulch, A.; Audebert, N.; Dubucq, D. Autoencodeurs pour la visualisation d'images hyperspectrales. In Proceedings of the 25th Colloque Gretsi, Juan-les-Pins, France, 5–8 September 2017; pp. 1–4.
22. Mou, L.; Ghamisi, P.; Zhu, X. Deep recurrent neural networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3639–3655. [CrossRef]
23. Lee, H.; Kwon, H. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [CrossRef]
24. Liu, B.; Yu, X.; Zhang, P.; Tan, X.; Yu, A.; Xue, Z. A semi-supervised convolutional neural network for hyperspectral image classification. *Remote Sens. Lett.* **2017**, *8*, 839–848. [CrossRef]
25. Luo, Y.; Zou, J.; Yao, C.; Zhao, X.; Li, T.; Bai, G. HSI-CNN: A novel convolution neural network for hyperspectral image. In Proceedings of the 2018 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 16–17 July 2018; pp. 464–469.
26. Roy, S.; Krishna, G.; Dubey, S.; Bidyut, B. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 277–281. [CrossRef]
27. Rostami, M.; Kolouri, S.; Eaton, E.; Kim, k. Deep transfer learning for few-shot SAR image classification. *Remote Sens.* **2019**, *11*, 1374. [CrossRef]
28. Alajaji, D.; Alhichri, H.S.; Ammour, N.; Alajlan, N. Few-shot learning for remote sensing scene classification. In Proceedings of the 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), Tunis, Tunisia, 9–11 March 2020; pp. 81–84.
29. Yuan, Z.; Huang, W. Multi-attention DeepEMD for few-shot learning in remote sensing. In Proceedings of the 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 11–13 December 2020; Volume 9, pp. 1097–1102.
30. Kim, J.; Chi, M. SAFFNet: Self-attention-based feature fusion network for remote sensing few-shot scene classification. *Remote Sens.* **2021**, *13*, 2532. [CrossRef]
31. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.

32. Surasak, T.; Takahiro, I.; Cheng, C.; Wang, C.; Sheng, P. Histogram of oriented gradients for human detection in video. In Proceedings of the 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, 17–18 May 2018; pp. 172–176.
33. Mao, L.; Xie, M.; Huang, Y.; Zhang, Y. Preceding vehicle detection using histograms of oriented gradients. In Proceedings of the 2010 International Conference on Communications, Circuits and Systems (ICCCAS), Chengdu, China, 28–30 July 2010; pp. 354–358.
34. Qi, S.; Ma, J.; Lin, J.; Li, Y.; Tian, J. Unsupervised Ship Detection Based on Saliency and S-HOG Descriptor from Optical Satellite Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1451–1455.
35. Chen, G.; Krzyzak, A.; Xie, W. Hyperspectral face recognition with histogram of oriented gradient features and collaborative representation-based classifier. *Multimed. Tools. Appl.* **2022**, *81*, 2299–2310. [CrossRef]
36. Chopra, S.; Hadsell, R.; LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 539–546.
37. Melekhov, I.; Kannala, J.; Rahtu, E. Siamese network features for image matching. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 378–383.
38. Tao, R.; Gavves, E.; Smeulders, A. Siamese instance search for tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Cancun, Mexico, 4–8 December 2016; pp. 1420–1429.
39. Bertinetto, L.; Valmadre, J.; Henriques, J.; Vedaldi, A.; Torr, P. Fully-convolutional siamese networks for object tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, the Netherlands, 8–16 October 2016; pp. 850–865.
40. Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese neural networks for one-shot image recognition. *ICML Deep Learn. Workshop* **2015**, *2*, 1–30.
41. Zhao, S.; Li, W.; Du, Q.; Ran, Q. Hyperspectral classification based on siamese neural network using spectral-spatial feature. In Proceeding of 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 2567–2570.
42. Liu, B.; Yu, X.; Zhang, P.; Yu, A.; Fu, Q.; Wei, X. Supervised deep feature extraction for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 1909–1921. [CrossRef]
43. Cao, Z.; Li, X.; Jiang, J.; Zhao, L. 3D convolutional siamese network for few-shot hyperspectral classification. *J. Appl. Remote Sens.* **2020**, *14*, 048504. [CrossRef]
44. Zeiler, M.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 5–12 September 2014; pp. 818–833.
45. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
46. Zadeh, L. Fuzzy set theory. *Inf. Control.* **1965**, *8*, 338–353. [CrossRef]
47. Guo, H.; Zhang, X. A Track-to-Track Association Algorithm Based on Fuzzy Synthetical Function and Its Application. *Syst. Eng. Electron.* **2003**, *25*, 1401–1403.
48. Liu, J.; Li, R.; Liu, Y.; Zhang, Z. Multi-sensor data fusion based on correlation function and fuzzy integration function. *Syst. Eng. Electron.* **2006**, *28*, 1006–1009.
49. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Adv. Neural Inf. Processing Syst.* **2017**, *30*, 1–11.
50. Tang, H.; Li, Y.; Han, X.; Huang, Q.; Xie, W. A spatial–spectral prototypical network for hyperspectral remote sensing image. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 167–171. [CrossRef]
51. Mura, M.; Benediktsson, J.; Waske, B.; Bruzzone, L. Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *Int. J. Remote Sens.* **2010**, *31*, 5975–5991. [CrossRef]
52. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
53. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.