*electronics*

MDPI

*Article*

# An Adaptive Atrous Spatial Pyramid Pooling Network for Hyperspectral Classification

**Tianxing Zhu [1,*], Qin Liu [1] and Lixiang Zhang [2]**

[1] School of Software Engineering, Tongji University, Shanghai 201804, China
[2] School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China
* Correspondence: txzhu@tongji.edu.cn

**Abstract:** Hyperspectral imaging (HSI) offers rich spectral and spatial data, beneficial for a variety of applications. However, challenges persist in HSI classification due to spectral variability, non-linearity, limited samples, and a dearth of spatial information in conventional spectral classifiers. While various spectral–spatial classifiers and dimension reduction techniques have been developed to mitigate these issues, they are often constrained by the utilization of handcrafted features. Deep learning has been introduced to HSI classification, with pixel- and patch-level deep learning (DL) classifiers gaining substantial attention. Yet, existing patch-level DL classifiers encounter difficulties in concentrating on long-distance dependencies and managing category areas of diverse sizes. The proposed Self-Adaptive 3D atrous spatial pyramid pooling (ASPP) Multi-Scale Feature Fusion Network (SAAFN) addresses these challenges by simultaneously preserving high-resolution spatial detail data and high-level semantic information. This method integrates a modified hyperspectral superpixel segmentation technique, a multi-scale 3D ASPP convolution block, and an end-to-end framework to extract and fuse multi-scale features at a self-adaptive rate for HSI classification. This method significantly enhances the classification accuracy of HSI with limited samples.

**Keywords:** hyperspectral image; atrous spatial pyramid pooling; superpixel segmentation; feature fusion



**Citation:** Zhu, T.; Liu, Q.; Zhang, L. An Adaptive Atrous Spatial Pyramid Pooling Network for Hyperspectral Classification. *Electronics* **2023**, *12*, 5013. https://doi.org/10.3390/electronics12245013

Academic Editor: Alberto Fernandez Hilario

Received: 24 October 2023
Revised: 6 December 2023
Accepted: 11 December 2023
Published: 15 December 2023



## 1. Introduction

A collected hyperspectral image captures intricate light distribution over several hundred spectral bands. This wealth of spectral and spatial information enhances discriminative capability compared to standard color images or multi-spectral images. Therefore, hyperspectral imaging has found utility in numerous applications, such as classification [1–3], object tracking [4–6], environmental monitoring [7,8], and object detection [9–11]. In recent years, the classification of hyperspectral images has emerged as a dynamic research topic. It is a potent approach for information extraction with extensive applications in geological prospecting, precision forestry, land resource surveys, and military defense [12–14]. However, the classification of hyperspectral images (HSI) presents a formidable challenge due to the significant interference caused by the substantial spectral variability and nonlinearity of HSI. Moreover, the limited availability of samples in practice exacerbates the complexity of interpreting high-dimensional data.

Motivated by the remarkable success of deep learning (DL), different works have utilized DL for hyperspectral image (HSI) classification. HSI classification based on DL can be grouped into subpixel, pixel, patch, and scene levels. The pixel-level method takes the spectral vector or the extracted one-dimensional feature of the selected sample as the input to the DL model. Typical pixel-level DL classifiers include one-dimensional convolutional neural networks (1D CNNs) [15], recurrent neural networks (RNNs) [16], and deep belief networks (DBNs) [17]. However, the pixel-level method cannot utilize neighborhood spatial information for HSI classification. To address this deficiency, patch-level DL classifiers

take the local cube within a certain neighborhood of the central pixel as input to the DL model, enabling better feature mining suitable for classification. Thus, patch-level DL classifiers have received extensive attention in HSI classification, with examples including 3D CNNs [18–21] and capsule networks [22] used to improve accuracy. Wei et al. [23] proposed a multi-scale principle of relevant information that leverages the multilayer structure for learning representations in a coarse-to-fine manner. This approach aims to learn discriminative spectral–spatial features. The Transformer, which incorporates a self-attention mechanism to capture long-range information, has gained popularity in the field of HSI processing and has demonstrated its advantages in handling sequential data. For instance, Hong et al. [24] proposed a backbone network called SpectralFormer for HSI classification. This network is capable of learning spectral local sequence information from neighboring bands of HSIs and generating group-wise spectral embeddings. He et al. [25] introduced a spatial–spectral Transformer classification network, which combines a well-designed CNN for extracting spatial features with a modified Transformer for capturing sequential spectra relationships. Selen et al. [26] developed a spectral-swin Transformer (SpectralSWIN) classification network that utilizes a swin-spectral module to process spatial and spectral features concurrently. However, some shortcomings remain in CNN-based methods. Convolution kernels have limited receptive fields, preventing a focus on long-distance dependencies. Moreover, for remote-sensing scene images, the size of different category areas varies greatly, sometimes not even in the same order of magnitude. Although traditional methods use several fixed convolution kernels of different scales in one layer to increase feature diversity, they cannot accommodate all category area situations. If a category area is too small, the receptive field contains more useless neighborhood information, adding useless calculations. In contrast, if an area is too large, the fixed receptive field may not cover it entirely, resulting in lost domain information [27,28].

To address previous challenges, we introduce the Self-Adaptive 3D ASPP Multi-Scale Feature Fusion Network (SAAFN) for the efficient preservation of both high-resolution spatial details and high-level semantic information in HSIs. The key contributions include:

- A newly modified hyperspectral superpixel segmentation method combining Euclidean spectral distance with Log-Euclidean distance to define homogeneous regions.
- A multi-scale 3D ASPP convolution block designed for consistent classification map output by refining spectral–spatial features.
- An end-to-end framework for HSI classification that self-adaptively extracts and fuses multi-scale features using global and local HSI information. This network preserves spatial and spectral details regardless of category area size, converting HSI data into discriminative features.

The paper structure is as follows: Section 2 describes the SAAFN algorithm details. Section 3 presents experiments and analysis. Section 4 discusses SAAFN's performance. The conclusion is summarized in Section 5.

## 2. Proposed Method

In this section, we propose a trainable end-to-end Self-Adaptive 3D ASPP Multi-Scale Feature Fusion Network (SAAFN) model for HSI classification. SAAFN integrates hierarchical convolutions and multi-scale feature fusion to fully exploit spatial and spectral discriminability, boosting classification accuracy based on spatial and spectral signatures of HSIs with limited samples. A self-adaptive dilation rate selection strategy based on hyperspectral superpixel segmentation solves large-scale differences in remote-sensing images. It adaptively adjusts the dilation rate and receptive field according to category area sizes in the HSI.

### *2.1. Hyperspectral Superpixel Segmentation*

Superpixels, which are perceptually meaningful connected regions grouping similarly colored pixels or other features, were first introduced by Ren and Malik [29]. Over the years, various algorithmic approaches have been proposed [30–32]. It has been proven that

using superpixels as adaptive regions [33] generates discriminative information for HSI classification. Cui et al. [34] demonstrated this by effectively optimizing a Support Vector Machine (SVM) probability map using a superpixel-based random walker. It was found that a superpixel spectrum is more stable and less influenced by noise than an individual pixel spectrum. Among clustering-based superpixel methods, Lloyd's algorithm [35], a modified version of the popular k-means clustering algorithm, is most commonly used. Yu et al. [36] proposed a label augmentation method to generate more labels for training based on the superpixel algorithm. In this context, we initially formalize the definition of superpixel segmentation.

An image of integer width $w$ and integer height $h$ is a function $I : \mathcal{X} \to \Omega$, where $\Omega$ is the image domain and $\mathcal{X} = [w] \times [h] \subset \mathbb{Z}^2$; then, a segmentation into superpixels is a partition $\{S_i\}_{i=1}^{n}$ of $\mathcal{X}$ such that, for each $1 \leq i \leq n$, we have:

$$S_i = \left\{ x : d((x, I(x)), F(S_i)) = \min_{1 \leq j \leq n} d\big((x, I(x)), F\big(S_j\big)\big) \right\} \tag{1}$$

where each $S_i$ is path-connected to $\mathbb{Z}^2$, $d$ is a metric on the space $\mathcal{X} \times \Omega$, and $F$ is the feature function on the set of all partitions.

The prevalent approach in hyperspectral image superpixel methodologies [37] involves inputting the first three principal components of HSI into RGB-based superpixel algorithms. However, this method may result in the loss of some higher-dimensional features. Therefore, when developing our hyperspectral superpixel map, it is crucial to design an algorithm capable of extracting more effective information from high-dimensional hyperspectral data. To start, we altered our algorithm from [31], as it takes image data with any number of bands $\mathcal{B}$. It can map the image $I$ to a 2-D manifold $\mathcal{M} \in \mathbb{R}^{\mathcal{B}+2}$ rather than the standard $\mathcal{M} \in \mathbb{R}^5$.

Let $\mathrm{I} = \{I_b\}, b = 1, \ldots, \mathcal{B}$ be an HSI with dimensions $\mathcal{W} \times \mathcal{H} \times \mathcal{B}$ representing the width, height, and number of bands, respectively, and $I_b : \mathcal{W} \times \mathcal{H}$. We started by performing dimensionality reduction via PCA [5] on $I$ for computational efficiency, to construct a dimensionally reduced HSI $\mathrm{I} = \left\{\hat{I}_b\right\}, a = 1, \ldots, \mathcal{A}$ where $\mathcal{A} \ll \mathcal{B}$.

Denoting an individual pixel as $p \in \hat{\mathrm{I}}$, we partitioned $\hat{\mathrm{I}}$ into superpixels by splitting $\hat{\mathrm{I}}$ into a family of disjoint sets, $\hat{\mathrm{I}} = \cup_{i=1}^{K} \mathcal{S}_i, \mathcal{S}_i \cap \mathcal{S}_j = \emptyset$, where $S_i$ corresponds to an individual superpixel and $K$ is the number of superpixels. Each superpixel $S_i$ is made up of a set of $n_i$ connected pixels, $\mathcal{S}_i = \left\{p_{i,1}, \ldots, p_{i,n_i}\right\}$. Our superpixel segmentations were produced via the minimization of:

$$Q(\{\mathcal{S}_1, \mathcal{S}_K\}) = \sum_{i=1}^{K} \sum_{p \in \hat{\mathrm{I}}} d\big((p, \hat{\mathrm{I}}(\mathrm{p})), F(\mathcal{S}_i)\big) \tag{2}$$

where $d$ is a distance function and $F(\mathcal{S}_i)$ is the average of $S_i$.

To combine the spatial and spectral data more effectively, a combination of the Euclidean spectral distance [31] and LED [38] of a covariance matrix representation [39] was proposed for clustering distance. For each pixel $p \in \hat{\mathrm{I}}$, we constructed a covariance matrix $C_p$ using the same methodology as Fang et al. [3] and used the LED metric to calculate the distances between these matrices. The distance between two pixels $p_x, p_y$ is given by:

$$d\big(p_x, p_y\big) = \| \log m\big(\mathrm{C}_{p_x}\big) - \log m\Big(\mathrm{C}_{p_y}\Big) \|_F + \| \hat{\mathrm{I}}(p_x) - \hat{\mathrm{I}}(p_y) \| + \frac{m}{S} \| p_x - p_y \| \tag{3}$$

where $m$ controls the compactness of superpixel and $S$ scales the spatial distance.

Finally, we merged superpixels with similar spectral properties; that is, we merged neighboring seeds which satisfy:

$$j = \underset{s_j \in \mathcal{N}}{\operatorname{argmin}} \| \mathcal{P}_i^m - \mathcal{P}_j^m \| \tag{4}$$

where $\mathcal{P}_i^m$ is the average spectral information of the seed $s_i$ and $\mathcal{N}$ is the set of neighboring seeds. The proposed changes allow the production of accurate superpixels for HSIs. The superpixel segmentation maps for the Indian Pines, KSC, and Pavia University experimental datasets are illustrated in Figure 1.

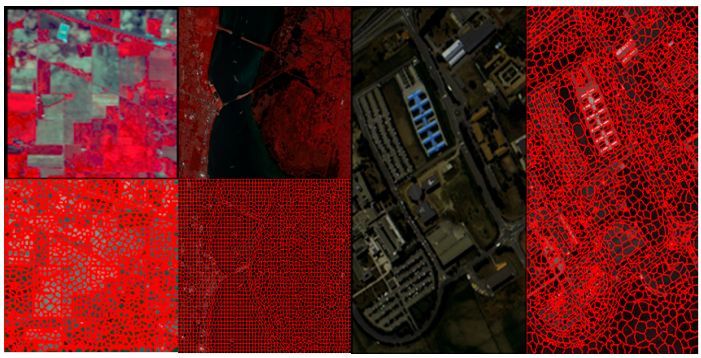



**Figure 1.** Visualization of (**a**) Indian Pines, (**b**) KSC, and (**c**) Pavia University datasets. For each dataset, the first and second images denote the false-color image and the homogeneous regions generated by the superpixel segmentation method.

### *2.2. 3D Atrous Spatial Pyramid Pooling Convolution Block*

In general convolutional neural networks, multi-scale information extraction has proven effective for the classification of related problems [40]. This is partly because multi-scale structures contain abundant context information. However, upsampling fits information through interpolation, so it cannot restore lost details. As a result, target object information cannot be fully reconstructed. Moreover, extracting features at different scales requires substantial computation to resize and aggregate feature maps. To address these issues, we propose a multi-scale 3D ASPP (atrous spatial pyramid pooling) convolution block, shown in Figure 2.

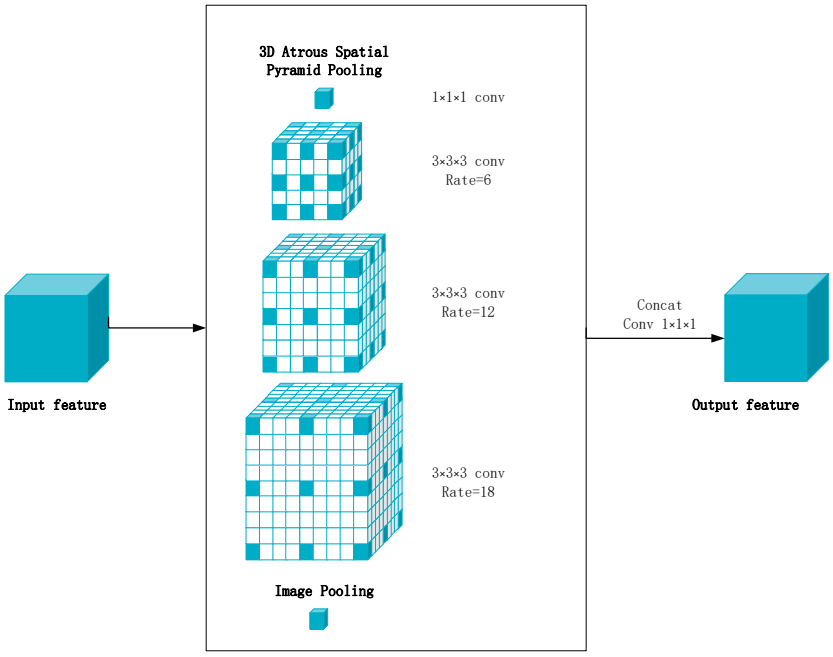



**Figure 2.** Illustration of the 3D ASPP Convolution Block with dilation rates of {6, 12, 18}.

Based on atrous spatial pyramid pooling (ASPP) [41], the multi-scale dilated convolution enables combining features extracted at variable scales. It realizes a larger receptive field without increasing convolution parameters or sacrificing resolution. The proposed block is designed to fuse different levels of feature outputs to concurrently preserve spatial details and spectral information. This converts the HSI's spatial and spectral data into discriminative features. The joint abstract features are acquired by the same dilated convolution with different dilation rates, then refined as spectral–spatial features. The block takes full advantage of hierarchical complementarity without sacrificing time cost. As such, it could be utilized as a basic structure to construct more powerful CNN models for HSI detection and classification.

### *2.3. SAAFN Model*

Our HSI superpixel segmentation block and 3D ASPP convolution block form the components of an SAAFN model, as depicted in Figure 3. The SAAFN model comprises a superpixel segmentation block, a 3D ASPP convolution block, and a fully connected layer. Initially, an HSI undergoes dimensional reduction prior to superpixel segmentation. Given that each superpixel block signifies the smallest homogeneous region, we strived to capture as much of the entire category areas as possible. This was achieved by extending the receptive field, denoted as $f_r$, by one pixel on the largest superpixel edge length.

$$f_r = max(edge_{horizontal}, edge_{vertical}) + 1 \tag{5}$$

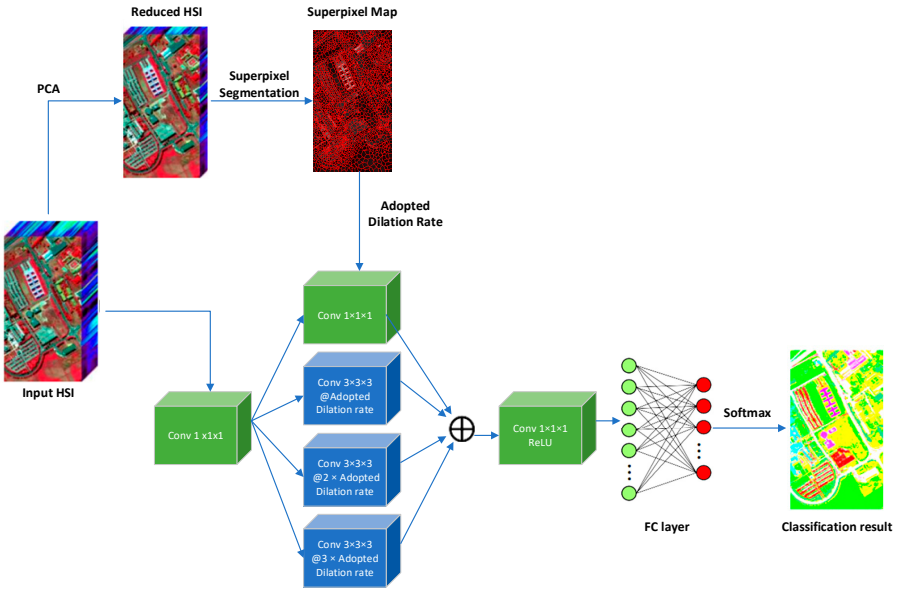



**Figure 3.** The overall architecture of the SAAFN framework.

Based on [42], the adaptive dilation rate $r_{dilation}$ can be calculated as

$$r_{dilation} = \frac{max(edge_{horizontal}, edge_{vertical}) + k - 1}{k - 1} \tag{6}$$

where $k$ is the size of the dilated convolution kernel.

Given the constraints of labeled data in the hyperspectral imaging (HSI) field, along with the spatial resolution of the data and the target sizes for each class to be classified, a relatively small kernel size in both the spatial and spectral dimensions was suitable for our experiments. The 3D atrous spatial pyramid pooling (ASPP) convolution block employs three dilated convolutions with adaptive dilation rates, allowing different scale levels on the feature map. Global average pooling is used to capture comprehensive image context data. Despite the use of a relatively small kernel size, a larger receptive field is achievable

without an increase in convolution parameters or sacrificing spatial and spectral resolution. The same model setting was applied to all datasets; hyperparameters are not intentionally adjusted to achieve higher performance. Our network was trained using multinomial logistic loss.

$$E = -\frac{1}{N}\sum_{n=1}^{N} log(p_l^n) \tag{7}$$

$p$ is the output of the softmax layer:

$$p_i = \frac{\exp x_i}{\sum_{i'=1}^{m} \exp x_{i'}} \tag{8}$$

where $N$ is the number of training samples, $l$ the correspondent label of sample $n$, $m$ is the number of classes, and $x$ is the input of the softmax layer.

## 3. Experimental Results and Analysis

### *3.1. Experimental Settings*

The performance of the SAAFN algorithm was evaluated on three commonly used hyperspectral datasets: Indian Pines, Pavia University, and Kennedy Space Center (KSC). Six representative hyperspectral feature extraction classifiers were selected as benchmarks, including SVM with RBF kernel, 2D+1D CNN [43], 3D CNN [44], SSRN (spectral–spatial residual network) [45], and HybridSN (3D+2D CNN) [46], comprising both shallow and deep classifiers. All CNN networks were trained using 3 × 3-pixel patches. As expected, we obtained lower classification accuracy than results reported in the literature, since prior works used slightly different setups:

- Some articles consider only a subset of the classes, excluding classes with fewer than 100 samples in Indian Pines, for example.
- The number of training samples varies, with some articles using a fixed percentage of the full training set and others using a fixed amount per class.
- Some authors further divide the training set into proper training and validation sets for hyperparameter tuning.

The experiments were run on a computer with an Intel i7-11700K 3.60 GHz CPU and NVIDIA RTX 3080Ti 12 GB GPU using PyTorch 1.13.0. For all experiments, only 10% of the labeled samples were randomly selected for training, with the rest used for testing. Both qualitative maps and quantitative evaluations comprehensively analyzed performance using four common metrics: producer's accuracy, overall accuracy, average accuracy, and kappa. The experiments were repeated 10 times independently, reporting average precision and standard deviation.

### *3.2. Experiment on Indian Pines Dataset*

The Indian Pines image covers an agricultural area in Indiana, USA, obtained by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor [47]. The 145 × 145-pixel image has 200 bands from 0.4 to 2.5 µm after removing water absorption bands. It contains 16 different classes, including many similar crop subclasses. The scene has an imbalanced sample distribution, with single-digit training samples for several classes. This makes classification very challenging with limited samples.

The classification maps and quantitative evaluations are shown in Figure 4 and Table 1. The optimal and sub-optimal values are in bold and underlined, respectively, for each method. Overall, the deeper classifiers performed better, yielding higher accuracies. SSRN obtained results comparable to SVM, with 55.51% and 56.38% OA. The 3D CNN surpassed the 2D+1D CNN, achieving 78.09% OA and 0.86 kappa versus 69.94% and 0.66. The proposed SAAFN algorithm further improved performance, achieving the highest accuracy with 87.45% OA and 0.85 kappa. It obtained optimal or sub-optimal precision for most classes, especially grass-pasture and woods. Overall, SAAFN outperformed the other classifiers, achieving the best results for this scene in both visual and quantitative evaluations.

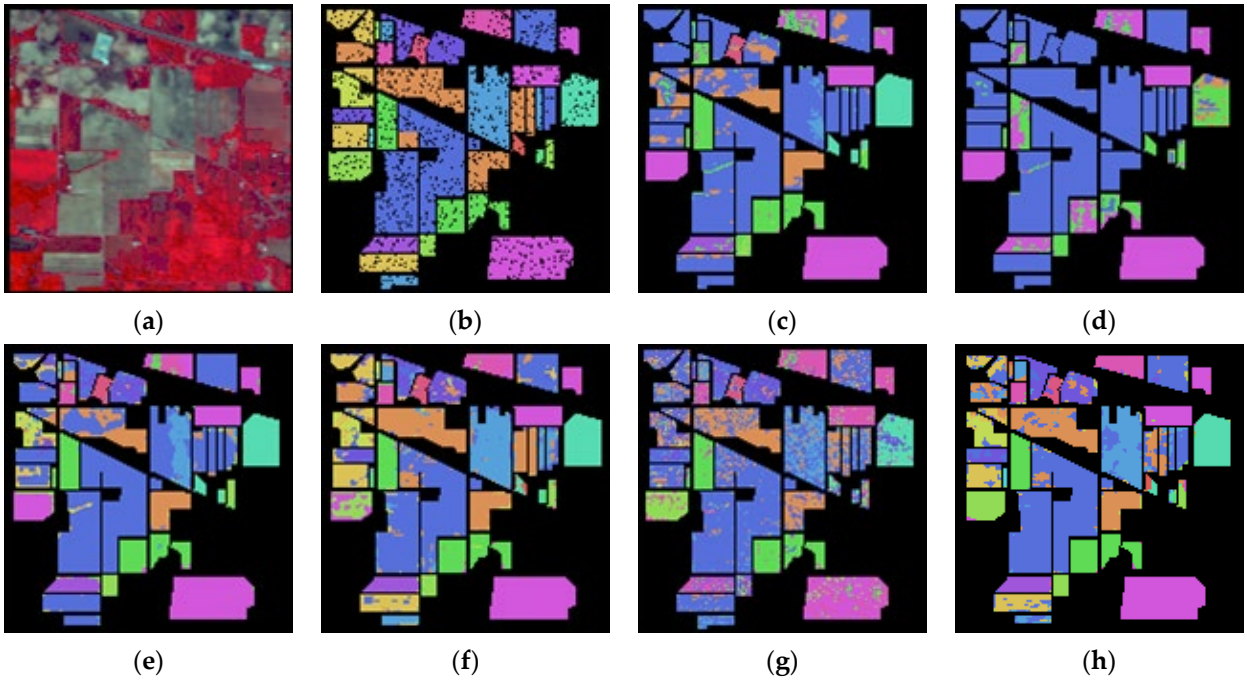



**Figure 4.** The results for Indian Pines experiment: (**a**) false-color image; (**b**) ground truth; (**c**) SVM; (**d**) 1D CNN; (**e**) 2D+1D CNN; (**f**) 3D CNN; (**g**) RNN; (**h**) SAAFN.

**Table 1.** Quantitative evaluations for Indian Pines experiment (%). The optimal and sub-optimal values are **bolded** and underlined, respectively.

| Class | SVM | 2D+1D CNN | 3D CNN | SSRN | HybridSN | SAAFN |
|---|---|---|---|---|---|---|
| Alfalfa | 43.42 | 28.4 | 62.7 | 19 | 69.5 | **83.2** |
| Corn-notill | 45.3 | 66 | 79 | 51 | **90.0** | 86 |
| Corn-mintill | 8.4 | 46.8 | 69.1 | 28 | 77.1 | **80.7** |
| Corn | 4.6 | 51.4 | 63.9 | 24.5 | 75.7 | **83.3** |
| Grass-pasture | 14.4 | 74.4 | 81.2 | 54.2 | 85.4 | **90.2** |
| Grass-trees | 77.8 | 93 | 86.3 | 78.6 | 93.3 | **96.7** |
| Grass-pasture-mowed | **94.91** | 10.2 | 61.4 | 22.8 | 64.4 | 91.1 |
| Hay-windrowed | 89.8 | 93.5 | 87.4 | 84.3 | 88.1 | **98.7** |
| Oats | 31.67 | 4.8 | 60.4 | 4.4 | **71.0** | 67.5 |
| Soybean-notill | 20.9 | 60.5 | 76 | 44.5 | **90.4** | 82.4 |
| Soybean-mintill | 60.1 | 55.8 | 78.9 | 45.3 | 81.8 | **86.7** |
| Soybean-clean | 3.9 | 59.6 | 72 | 34.4 | 77.5 | **78.2** |
| Wheat | 78.3 | 94.5 | 88.4 | 76.2 | 88.8 | **98** |
| Woods | 83.8 | 91.9 | 86.1 | 79.5 | 87.2 | **96.4** |
| Buildings-Grass-Trees-Drives | 9 | 62.4 | 67.6 | 46.2 | 69.8 | **76.2** |
| Stone-Steel-Towers | 91 | **94.2** | 85.2 | 61.2 | 90.7 | 82.6 |
| OA | 56.38 | 69.94 | 78.09 | 55.51 | 86.69 | **87.45** |
| | ±0.96 | ±7.38 | ±1.28 | ±8.84 | ±1.38 | ±1.07 |
| AA | 47.33 | 61.71 | 75.35 | 47.13 | 81.29 | **86.2** |
| | ±1.01 | ±11.48 | ±4.87 | ±14.59 | ±3.75 | ±3.86 |
| Kappa | 0.48 | 0.66 | **0.86** | 0.5 | 0.81 | 0.85 |
| | ±0.01 | ±0.08 | **±0.02** | ±0.10 | ±0.01 | ±0.01 |

### *3.3. Experiment on Kennedy Space Center Dataset*

The second experiment used the Kennedy Space Center (KSC) image obtained by AVIRIS over Kennedy Space Center, FL, USA [47]. This 512 × 614-pixel scene has 176 bands after removing water absorption and noisy bands, with an 18 m spatial resolution. It contains 13 classes in total.

The classification maps and quantitative evaluations are presented in Figure 5 and Table 2. Consistent with the Indian Pines experiment, SSRN performed worst, while the proposed SAAFN algorithm again achieved the best performance. SAAFN obtained optimal or sub-optimal producer's accuracies for most classes, demonstrating its effectiveness. Notably, it yielded significant improvements for the scrub and graminoid marsh classes. Overall, SAAFN achieved the best visual map and highest accuracies, including 88.40% overall accuracy, 82.02% average accuracy, and 0.87 kappa.

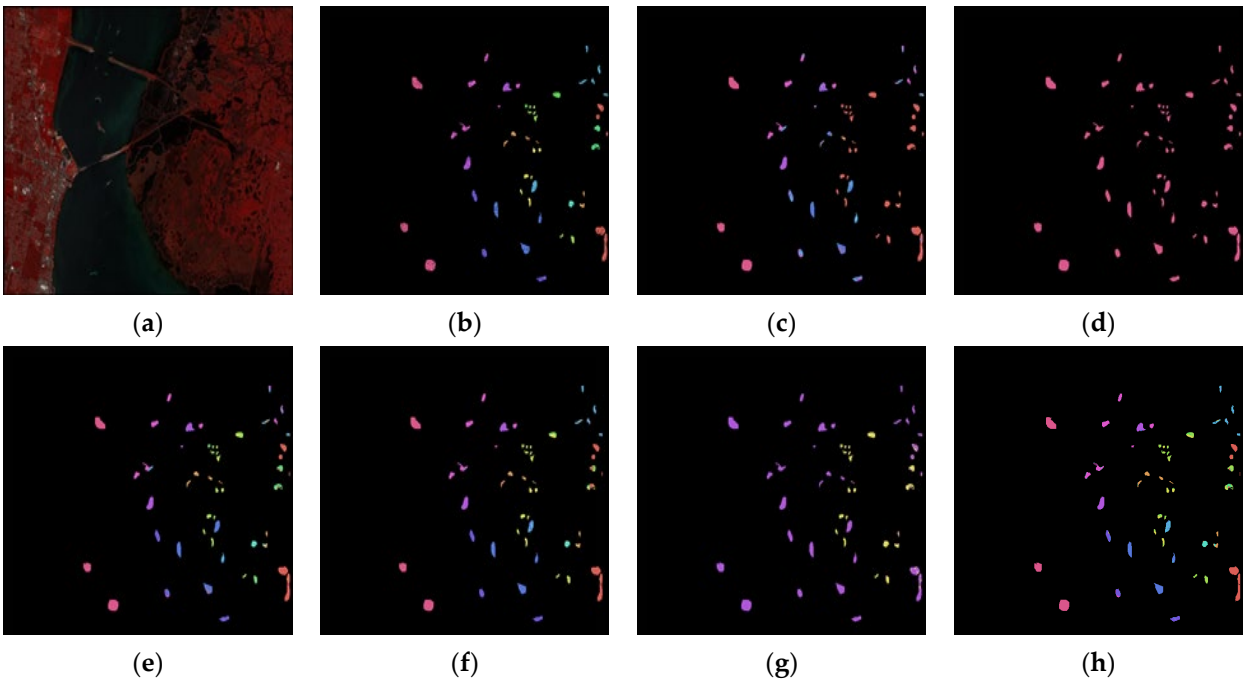


**Figure 5.** The results for KSC experiment: (**a**) false-color image; (**b**) ground truth; (**c**) SVM; (**d**) 1D CNN; (**e**) 2D+1D CNN; (**f**) 3D CNN; (**g**) RNN; (**h**) SAAFN.

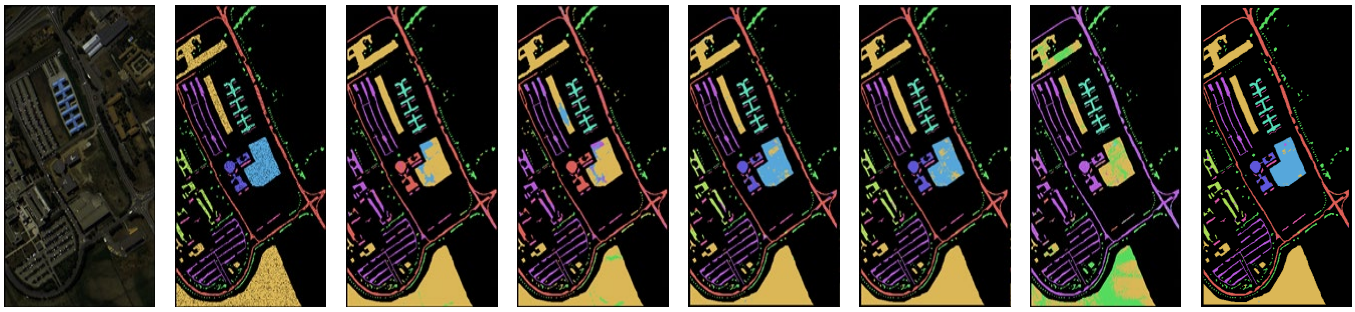
**Table 2.** Quantitative evaluations for KSC experiment (%). The optimal and sub-optimal values are **bolded** and underlined, respectively.

| Class | SVM | 2D+1D CNN | 3D CNN | SSRN | HybridSN | SAAFN |
|---|---|---|---|---|---|---|
| Scrub | 56.7 | 69.2 | 93.1 | 39.4 | 93.4 | **94.7** |
| Willow swamp | 42.8 | 53.1 | 87.3 | 39.0 | 87.4 | **88.6** |
| CP hammock | 70.8 | 10.9 | 47.8 | 55.2 | 68.7 | **70.9** |
| CP/Oak | 28.7 | 18.7 | 34.8 | 21.3 | 24.1 | **48.3** |
| Slash pine | 25.6 | **46.4** | 23.3 | 23.0 | 36.5 | 42.5 |
| Oak/Broadleaf | 28.0 | 12.4 | 53.6 | 30.7 | 46.7 | **74.2** |
| Hardwood swamp | 55.9 | 15.7 | 76.1 | 26.1 | 70.8 | **80.6** |
| Graminoid marsh | 32.4 | 47.7 | 79.6 | 31.0 | 78.8 | **84.8** |
| Spartina marsh | 59.5 | 62.4 | 91.1 | 39.2 | 92.2 | **93.2** |
| Cattail marsh | 23.9 | 69.0 | **97.5** | 27.6 | 91.0 | 94.5 |
| Salt marsh | 91.4 | 91.7 | 98.1 | 42.4 | 94.7 | **99.1** |
| Mud flats | 72.2 | 81.2 | 91.9 | 24.6 | 87.8 | **94.9** |
| Water | 98.5 | 99.7 | 99.9 | 46.2 | **100.0** | **100.0** |
| OA | 59.74 | 66.42 | 85.2 | 44.03 | 82.7 | **88.4** |
| | ±1.35 | ±7.56 | ±3.23 | ±18.77 | ±0.8 | ±0.49 |
| AA | 52.79 | 52.16 | 74.93 | 34.28 | 75.35 | **82.02** |
| | ±2.82 | ±15.31 | ±10.42 | ±20.22 | ±3.44 | ±3.25 |
| Kappa | 0.54 | 0.62 | 0.84 | 0.18 | 0.83 | **0.87** |
| | ±0.02 | ±0.09 | ±0.04 | ±0.20 | ±0.02 | ±0.01 |

### *3.4. Experiment on Pavia University Dataset*

The third dataset was the Pavia University image, obtained by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over Pavia University, Northern Italy [47]. This urban scene covers 610 × 340 pixels with 103 bands from 0.43 to 0.86 µm. It contains nine different classes in total with a complex distribution.

Figure 6 and Table 3 present the classification results. Overall accuracy was better than the Indian Pines experiment, likely due to more training samples and higher spatial resolution (20 m/pixel vs. 1.3 m/pixel). As before, SSRN performed worst. The deeper classifiers yielded higher accuracies, especially for asphalt and metal sheets classes, with over 80% OA. The 3D CNN slightly outperformed SAAFN, achieving 94.67% OA and 0.93 kappa compared to 93.89% and 0.92 for SAAFN. Both 3D CNN and SAAFN achieved the best performance, with 3D CNN having the highest average OA, while SAAFN had the best single run but larger deviation due to 10% random sampling.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**) (**g**) (**h**)

**Figure 6.** The results for Pavia University experiment: (**a**) false-color image; (**b**) ground truth; (**c**) SVM; (**d**) 1D CNN; (**e**) 2D+1D CNN; (**f**) 3D CNN; (**g**) RNN; (**h**) SAAFN.

**Table 3.** Quantitative evaluations for Pavia University experiment (%). The optimal and sub-optimal values are **bolded** and underlined, respectively.

| Class | SVM | 2D+1D CNN | 3D CNN | SSRN | HybridSN | SAAFN |
|---|---|---|---|---|---|---|
| Asphalt | 83.8 | 93.3 | 96.3 | 93.1 | **97.0** | 95.8 |
| Meadows | 88.8 | 93.6 | 96.2 | 45.0 | **96.5** | 95.3 |
| Gravel | 4.9 | 86.0 | 91.5 | 42.3 | 91.7 | **92.0** |
| Trees | 89.9 | 93.6 | 96.7 | 62.7 | 97.5 | **97.9** |
| Metal sheets | 99.1 | 99.7 | 99.6 | 86.7 | 99.3 | **99.9** |
| Bare soil | 33.8 | 88.6 | 97.2 | 31.4 | **97.3** | 97.1 |
| Bitumen | 83.1 | 86.7 | 93.1 | 21.4 | **93.9** | 91.5 |
| Bricks | 76.1 | 91.8 | 94.3 | 13.0 | 94.9 | **95.7** |
| Shadows | 99.9 | 99.5 | 99.6 | 43.1 | **100.0** | 99.7 |
| OA | 79.76 | 91.24 | **94.67** | 44.12 | 90.10 | 93.89 |
| | ±0.09 | ±1.61 | ±0.42 | ±18.88 | ±0.43 | ±0.74 |
| AA | 73.27 | 92.53 | 96.06 | 48.74 | 95.83 | **96.10** |
| | ±0.48 | ±2.43 | ±1.28 | ±22.80 | ±1.52 | ±2.39 |
| Kappa | 0.72 | 0.87 | **0.93** | 0.36 | 0.90 | 0.92 |
| | ±0.01 | ±0.02 | ±0.01 | ±0.17 | ±0.01 | ±0.01 |

### *3.5. Discussion*

Analysis of the results indicates the traditional SVM classifier generally failed to achieve high performance due to limited shallow feature discriminability, resulting in numerous misclassifications. Especially for Indian Pines and KSC, SVM obtained relatively low average accuracies (under 60%), insufficient for practical applications. Although a deep classifier, SSRN suffered from overfitting and had the lowest accuracy. The remaining deep classifiers showed clear superiority by learning high-level non-linear spatial–spectral representations hierarchically.

The 3D CNN significantly boosted accuracy by 37.93%, 3.18%, and 21.21% over 2D + 1D CNN, demonstrating the efficient fusion of spatial and spectral patterns. Compared to other classifiers, SAAFN further improved performance, achieving the best single OA consistently. It had the best average OA for Indian Pines and KSC and the second best for Pavia University. The optimal maps and metrics demonstrated effectiveness in mining discriminative spatial–spectral features. Specifically, SAAFN integrated complementary hierarchical information and cross-channel dependencies to overcome limitations.

SAAFN successfully achieved its objectives in classes where HybridSN failed. For instance, it significantly enhanced PAs for various classes in Indian Pines as well as in KSC. This serves as further evidence of the effectiveness of SAAFN. In general, SAAFN outperformed alternative spatial–spectral classifiers, clearly demonstrating its competitiveness in HSI classification.

## 4. Ablation Study and Comparative Experiments

We designed an ablation study to confirm that the effectiveness of the proposed Self-Adaptive Atrous Filter Network (SAAFN) is primarily due to the following factors:

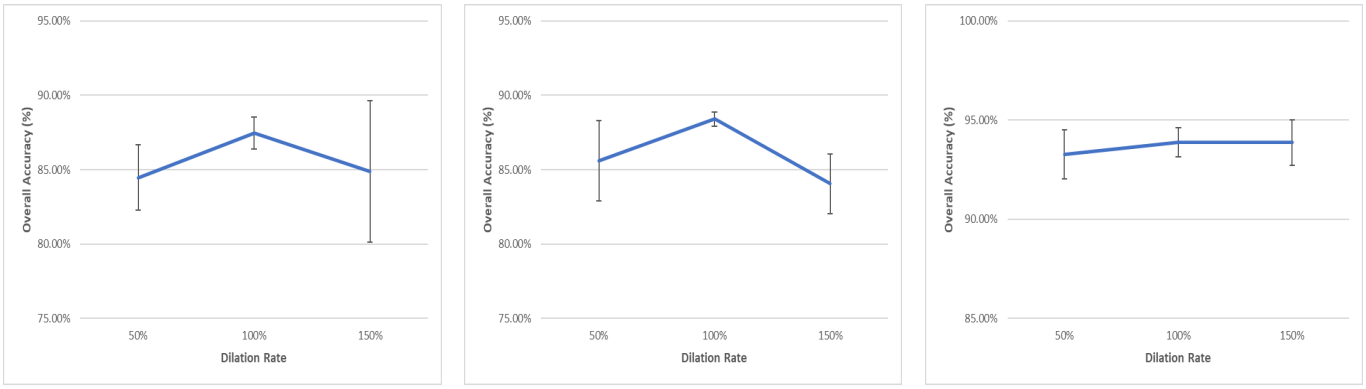
The efficacy of the self-adaptive dilation rate derived from the modified hyperspectral superpixel segmentation method. This method should provide an optimal receptive field size that covers the entire area without the loss of domain-specific information.

The integration of the self-adaptive dilation rate with a multi-scale 3D atrous spatial pyramid pooling (ASPP) convolution block, which is engineered to merge different levels of feature outputs. This integration utilizes the relationship between homogeneous regions to determine dilation rates according to the range of distinct category areas. Regardless of the category areas' size, the network can preserve both spatial details and spectral information simultaneously, effectively transforming the spatial and spectral information of hyperspectral imaging (HSI) into discriminative features.

To conduct the ablation study and validate the effectiveness of the self-adaptive dilation rate, the classification performance of various adapted dilation rates was compared using three datasets: Indian Pines, Kennedy Space Center (KSC), and Pavia University. To analyze the impact of dilation rate at varying scales, we compared the derived adapted dilation rates at 100%, 50%, and 150%. To further assess the effectiveness and robustness of the proposed SAAFN and its ability to preserve spatial details and spectral information, different methods were compared using various sizes of training samples (i.e., 10%, 20%, 30%, 40%, 50%) for the Indian Pines, KSC, and Pavia University datasets.

### *4.1. Dilation Rate vs. Classification Performance Comparison*

The effectiveness of the self-adaptive dilation rate was evaluated by comparing the classification performance of the adapted dilation rate, 50% of the adapted dilation rate, and 150% of the adapted dilation rate across all three datasets. The adapted dilation rates for the Indian Pines, KSC, and Pavia University datasets were {6, 10, 15}, and the corresponding 50% and 150% dilation rates were {3, 5, 8} and {9, 15, 23}, respectively. The results, as depicted in Figure 7, show that the adapted dilation rate significantly outperforms the others in terms of overall accuracy and standard deviations across all datasets. This suggests that the optimized dilation rate effectively translates the spatial and spectral information of hyperspectral imaging (HSI) into discriminative features regardless of the size of the category areas.

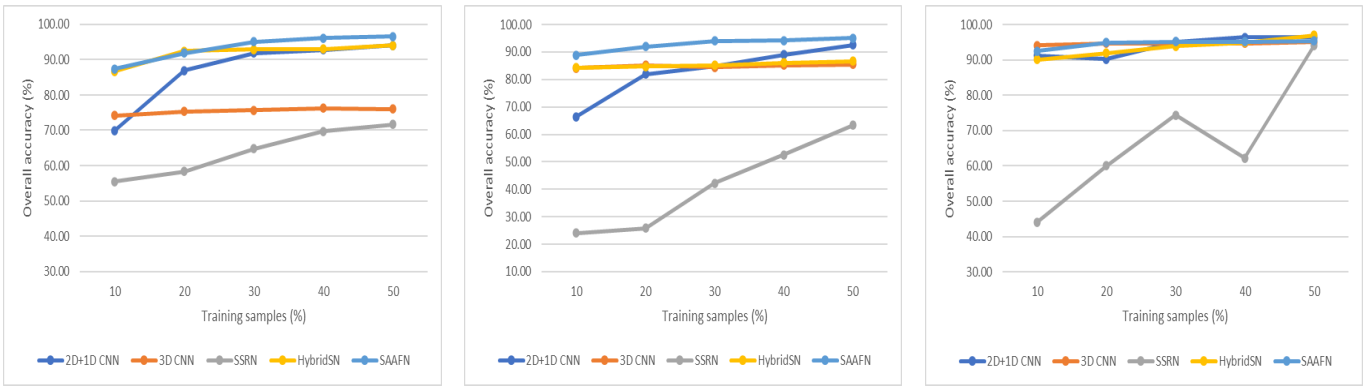



(**a**) (**b**) (**c**)

**Figure 7.** The classification performance using various dilation rates: (**a**) Indian Pines image; (**b**) KSC image; (**c**) Pavia University image.

### *4.2. Training Sample Size vs. Classification Performance Comparison*

The effectiveness and robustness of the classification performance of differing methods were evaluated by comparing them using various training sample sizes, namely, 10%, 20%, 30%, 40%, and 50%. The results of this comparison are presented in Figure 8. It is evident from the data that the overall accuracy of all methods escalated in direct proportion to the increase in training samples. On a general note, SRN demonstrated the most significant improvement, primarily due to the reduction in overfitting as the sample size expanded. However, SAAFN consistently outperformed all other methods in all scenarios. More specifically, SAAFN displayed a pronounced superiority when dealing with small- to medium-sized samples.




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

**Figure 8.** The classification performance of different methods using various sizes of training samples: (**a**) Indian Pines image; (**b**) KSC image; (**c**) Pavia University image.

## 5. Conclusions

In this study, we introduce a novel end-to-end self-adaptive multi-scale feature fusion-based SAAFN for hyperspectral image (HSI) classification. Initially, we propose a modified hyperspectral superpixel segmentation method that combines Euclidean spectral distance with the Log-Euclidean distance (LED) of a covariance matrix representation as a clustering distance. This approach enables us to define more meaningful homogeneous regions to establish an optimized receptive field size that considers all category areas. Furthermore, we propose a multi-scale 3D ASPP convolution block to merge different levels of feature output. This technology leverages hierarchical complementarity without sacrificing time efficiency, making it suitable as a foundational structure for constructing more robust CNN models

for HSI detection and classification. The final SAAFN framework utilizes the relationship between homogeneous regions to determine optimized dilation rates according to the range of different category areas for the 3D ASPP network. Regardless of the category areas' size, our proposed framework can simultaneously preserve spatial details and spectral information, converting the HSI's spatial and spectral data into discriminative features. This is achieved by fully utilizing the global and local information of the HSI. We tested the proposed SAAFN on three commonly used hyperspectral image datasets. The results demonstrate that it surpasses other typical spectral classifiers and can be a competitive method for practical applications.

**Author Contributions:** Conceptualization, T.Z.; methodology, T.Z.; software, T.Z.; writing—original draft preparation, T.Z.; writing—review and editing, Q.L. and L.Z.; supervision, Q.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data can be found here: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (accessed on 10 December 2023).

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

## References

1. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [CrossRef]
2. Fang, L.Y.; Li, S.T.; Kang, X.D.; Benediktsson, J.A. Spectral-spatial classification of hyperspectral images with a super-pixel-based discriminative sparse model. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4186–4201. [CrossRef]
3. Fang, L.; He, N.; Li, S.; Plaza, A.J.; Plaza, J. A New Spatial–Spectral Feature Extraction Method for Hyperspectral Images Using Local Covariance Matrix Representation. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3534–3546. [CrossRef]
4. Wang, T.; Zhu, Z.; Blasch, E. Bio-Inspired Adaptive Hyperspectral Imaging for Real-Time Target Tracking. *IEEE Sensors J.* **2010**, *10*, 647–654. [CrossRef]
5. Uzkent, B.; Hoffman, M.J.; Vodacek, A. Real–time vehicle tracking in aerial video using hyperspectral features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 27–30 June 2016; pp. 36–44.
6. Uzkent, B.; Rangnekar, A.; Hoffman, M.J. Aerial Vehicle Tracking by Adaptive Fusion of Hyperspectral Likelihood Maps. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 233–242.
7. Ellis, R.J.; Scott, P.W. Evaluation of hyperspectral remote sensing as a means of environmental monitoring in the St. Austell China clay (kaolin) region Cornwall UK. *Remote Sens. Environ.* **2004**, *93*, 118–130. [CrossRef]
8. Manfreda, S.; McCabe, M.F.; Miller, P.E.; Lucas, R.; Madrigal, V.P.; Mallinis, G.; Ben Dor, E.; Helman, D.; Estes, L.; Ciraolo, G.; et al. On the Use of Unmanned Aerial Systems for Environmental Monitoring. *Remote Sens.* **2018**, *10*, 641. [CrossRef]
9. Pan, Z.; Healey, G.; Prasad, M.; Tromberg, B. Face recognition in hyperspectral images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1552–1560.
10. Tian, C.; Zheng, M.; Zuo, W.; Zhang, B.; Zhang, Y.; Zhang, D. Multi-stage image denoising with the wavelet transform. *Pattern Recognit.* **2023**, *134*, 109050. [CrossRef]
11. Tian, C.; Zheng, M.; Zuo, W.; Zhang, S.; Zhang, Y.; Lin, C.-W. A cross Transformer for image denoising. *Inf. Fusion* **2023**, *102*, 102043. [CrossRef]
12. Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Plaza, A.; Chanussot, J. Graph convolutional networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5966–5978. [CrossRef]
13. Haut, J.M.; Paoletti, M.E.; Plaza, J.; Plaza, A.; Li, J. Visual Attention-Driven Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8065–8080. [CrossRef]
14. Zhai, H.; Zhang, H.; Zhang, L.; Li, P. Nonlocal Means Regularized Sketched Reweighted Sparse and Low-Rank Subspace Clustering for Large Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4164–4178. [CrossRef]
15. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep convolutional neural networks for hyperspectral image classification. *J. Sens.* **2015**, *2015 Pt 3*, 1–12. [CrossRef]
16. Mou, L.; Ghamisi, P.; Zhu, X.X. Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3639–3655. [CrossRef]
17. Chen, Y.; Zhao, X.; Jia, X. Spectral–Spatial Classification of Hyperspectral Data Based on Deep Belief Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2381–2392. [CrossRef]
18. Zhang, H.; Li, Y.; Jiang, Y.; Wang, P.; Shen, Q.; Shen, C. Hyperspectral Classification Based on Lightweight 3-D-CNN With Transfer Learning. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5813–5828. [CrossRef]

19. Tian, C.; Xu, Y.; Zuo, W.; Zhang, B.; Fei, L.; Lin, C.-W. Coarse-to-Fine CNN for Image Super-Resolution. *IEEE Trans. Multimedia* **2020**, *23*, 1489–1502. [CrossRef]
20. Tian, C.; Xu, Y.; Zuo, W.; Lin, C.-W.; Zhang, D. Asymmetric CNN for Image Superresolution. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *52*, 3718–3730. [CrossRef]
21. Tian, C.; Yuan, Y.; Zhang, S.; Lin, C.W.; Zuo, W.; Zhang, D. Image super-resolution with an enhanced group convolutional neural network. *Neural Netw.* **2022**, *153*, 373–385. [CrossRef]
22. Ding, X.; Li, Y.; Yang, J.; Li, H.; Liu, L.; Liu, Y.; Zhang, C. An Adaptive Capsule Network for Hyperspectral Remote Sensing Classification. *Remote Sens.* **2021**, *13*, 2445. [CrossRef]
23. Wei, Y.; Yu, S.; Giraldo, L.S.; Príncipe, J.C. Multiscale principle of relevant information for hyperspectral image classification. *Mach. Learn.* **2021**, *112*, 1227–1252. [CrossRef]
24. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.; Chanussot, J. SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [CrossRef]
25. He, X.; Chen, Y.; Lin, Z. Spectral-Spatial transformer for hyperspectral image classification. *Remote Sens.* **2021**, *13*, 498. [CrossRef]
26. Selen, A.; Esra, T.-G. SpectralSWIN: A spectral-swin transformer network for hyperspectral image classification. *Int. J. Remote Sens.* **2022**, *43*, 4025–4044.
27. Paoletti, M.E.; Haut, J.M.; Plaza, J.; Plaza, A. A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 120–147. [CrossRef]
28. Zhu, H.; Zeng, H.; Liu, J.; Zhang, X. Logish: A new nonlinear nonmonotonic activation function for convolutional neural network. *Neurocomputing* **2021**, *458*, 490–499. [CrossRef]
29. Ren, X.; Malik, J. Learning a classification model for segmentation. *Proc. Int. Conf. Comput. Vis.* **2003**, 10–17. [CrossRef]
30. Liu, Y.-J.; Yu, C.-C.; Yu, M.-J.; He, Y. Manifold SLIC: A Fast Method to Compute Content-Sensitive Superpixels. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 651–659.
31. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [CrossRef]
32. Maierhofer, G.; Heydecker, D.; Aviles-Rivero, A.I.; Alsaleh, S.M.; Schonlieb, C.-B. Peekaboo-Where are the Objects? Structure Adjusting Superpixels. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Countrydate of Conference, Athens, Greece, 7–10 October 2018; pp. 3693–3697.
33. Jia, S.; Deng, B.; Zhu, J.; Jia, X.; Li, Q. Superpixel–based multitask learning framework for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2575–2588. [CrossRef]
34. Cui, B.; Xie, X.; Ma, X.; Ren, G.; Ma, Y. Superpixel-Based Extended Random Walker for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3233–3243. [CrossRef]
35. Lloyd, S.P. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [CrossRef]
36. Yu, X.; Ma, Y.; Farrington, S.; Reed, J.; Ouyang, B.; Principe, J.C. Fast segmentation for large and sparsely labeled coral images. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–6.
37. Fang, L.; Li, S.; Duan, W.; Ren, J.; Benediktsson, J.A. Classification of Hyperspectral Images by Exploiting Spectral–Spatial Information of Superpixel via Multiple Kernels. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6663–6674. [CrossRef]
38. Arsigny, V.; Fillard, P.; Pennec, X.; Ayache, N. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **2007**, *29*, 328–347. [CrossRef]
39. Tuzel, O.; Porikli, F.; Meer, P. Region covariance: A fast descriptor for detection and classification. In Proceedings of the Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 589–600.
40. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
41. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding Convolution for Semantic Segmentation. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460.
42. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1–14.
43. Hamida, A.B.; Benoit, A.; Lambert, P.; Ben, A.C. Deep Learning Approach for Remote Sensing Image Analysis. In Proceedings of the Big Data from Space (BiDS'16), Santa Cruz de Tenerife, Spain, 15–17 March 2016; p. 133.
44. Li, Y.; Zhang, H.; Shen, Q. Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Net-work. *Remote Sens.* **2017**, *9*, 67. [CrossRef]
45. Mou, L.; Ghamisi, P.; Zhu, X.X. Unsupervised Spectral–Spatial Feature Learning via Deep Residual Conv–Deconv Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 391–406. [CrossRef]

46. Roy, S.; Krishna, G.; Dubey, S. HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 277–281. [CrossRef]
47. Hyperspectral Remote Sensing Scenes. Available online: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes (accessed on 10 March 2022).