



Article

AI-TFNet: Active Inference Transfer Convolutional Fusion Network for Hyperspectral Image Classification

Jianing Wang ^{1,2,*} , Linhao Li ², Yichen Liu ², Jinyu Hu ², Xiao Xiao ³ and Bo Liu ² ¹ School of Computer Science and Technology, Xidian University, No. 2 South TaiBai Road, Xi'an 710071, China² School of Artificial Intelligence, Xidian University, No. 2 South TaiBai Road, Xi'an 710071, China³ School of Telecommunications Engineering, Xidian University, No. 2 South TaiBai Road, Xi'an 710071, China

* Correspondence: jnwang@xidian.edu.cn

Abstract: The realization of efficient classification with limited labeled samples is a critical task in hyperspectral image classification (HSIC). Convolutional neural networks (CNNs) have achieved remarkable advances while considering spectral–spatial features simultaneously, while conventional patch-wise-based CNNs usually lead to redundant computations. Therefore, in this paper, we established a novel active inference transfer convolutional fusion network (AI-TFNet) for HSI classification. First, in order to reveal and merge the local low-level and global high-level spectral–spatial contextual features at different stages of extraction, an end-to-end fully hybrid multi-stage transfer fusion network (TFNet) was designed to improve classification performance and efficiency. Meanwhile, an active inference (AI) pseudo-label propagation algorithm for spatially homogeneous samples was constructed using the homogeneous pre-segmentation of the proposed TFNet. In addition, a confidence-augmented pseudo-label loss (CapLoss) was proposed in order to define the confidence of a pseudo-label with an adaptive threshold in homogeneous regions for acquiring pseudo-label samples; this can adaptively infer a pseudo-label by actively augmenting the homogeneous training samples based on their spatial homogeneity and spectral continuity. Experiments on three real HSI datasets proved that the proposed method had competitive performance and efficiency compared to several related state-of-the-art methods.

Keywords: hyperspectral image; classification; transfer convolutional neural networks; pseudo-label propagation



Citation: Wang, J.; Li, L.; Liu, Y.; Hu, J.; Xiao, X.; Liu, B. AI-TFNet: Active Inference Transfer Convolutional Fusion Network for Hyperspectral Image Classification. *Remote Sens.* **2023**, *15*, 1292. <https://doi.org/10.3390/rs15051292>

Academic Editors: Bo Du, Lan Du, Mingyi He and Jing Zhang

Received: 4 October 2022

Revised: 28 November 2022

Accepted: 6 December 2022

Published: 26 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In contrast to traditional panchromatic and multi-spectral images, hyperspectral images typically consist of dozens or even several hundred spectral bands in the visual and far-infrared spectra, and they can be effectively utilized to distinguish between different categories of land covers. In recent years, the analysis and processing of hyperspectral images have been used in many fields [1], such as in urban development and surveillance [2,3], environmental management [4], agriculture [5], etc.

Various supervised machine learning methods have been proposed and developed over time in order to improve the classification of HSIs, such as support vector machine (SVM) [6–8], *k*-nearest neighbor (K-NN) [9,10], and random forest [11–13]. These algorithms only consider the discriminant information of spectral signatures. Subsequently, spectral–spatial-based algorithms have been proposed that also consider spatial contextual features in order to improve classification accuracy and efficiency. A support vector machine with a composite kernel (SVMCK) is a representative patch-wise-based algorithm that simultaneously projects the spectral–spatial features in the reproducing kernel Hilbert space (RKHS) [14]. A joint sparse-representation-based approach involved simultaneously representing all pixels in the local patch, along with a group of common atoms in the training dictionary (JSRC) [15]. In ref. [16], a joint spectral–spatial derivative-aided kernel

sparse representation of patch-based kernels was proposed for HSI classification that considered the derivative features of the spectral variation simultaneously. Additionally, an adaptive non-local spectral-spatial kernel (ANSSK) was proposed in order to further exploit homogeneous spectral-spatial features in the embedded manifold feature space [17]. As for spatial filter feature extraction, various filter design algorithms, such as extended morphological profiles (EMPs) [18], edge-preserving features [19], and Gabor filters [20–23], have been proposed to improve classification performance. Most of the aforementioned classification algorithms adopted hand-crafted feature extractors and traditionally taught models; therefore, specialized field expertise is usually required for hand-crafted extraction.

Along with increased computational GPU resources, convolutional neural network (CNN)-based approaches have shown remarkable performances in visual tasks. For HSI classification, a 2D CNN [24] was proposed with differently designed convolutional operators. Thereafter, Song et al. designed a deep feature fusion network (DFFN) [25]. A spectral-spatial residual network (SSRN) was proposed by Zhong et al. in order to extract spectral-spatial features in an orderly fashion and classify HSIs according to joint spectral-spatial features [26]. Swalpa et al. designed a structure with a spectral-spatial 3D CNN to reduce the complexity of the model [27]. Mercedes et al. proposed a rotating variable model for HSI analysis, in which the conventional convolution kernel was substituted with circular harmonic filters (CHF) [28]. Wei et al. divided pixels into different clusters as a material map for extracting spatial features in order to achieve an effective classification [29]. Haokui et al. [30] proposed a method of HSI classification with a cross-sensor strategy and a cross-modal strategy based on transfer learning, and it utilized RGB image data and other HSI data collected by arbitrary sensors as pre-training datasets. Wang et al. proposed a network architecture search (NAS)-guided lightweight spectral-spatial attention feature fusion network (LMAFN) for HSI classification [31]. A novel multi-structure KELM with an attention fusion strategy (MSAF-KELM) was proposed in order to achieve the accurate fusion of multiple classifiers for effective HSI classification with ultra-small sample rates [32]. Yue et al. [33] enhanced the representation of learned features by reconstructing the spectral and spatial features of an HSI to achieve robust unknown detection. In addition, the graph convolutional network (GCN) [34,35] and fully convolutional neural network [36] have gradually attracted more and more attention due to the utilization of their inherent advantages. For instance, to explore the internal relationships of data for semi-supervised label propagation in few-shot image classification, an attention-weighted graph convolutional network (AwGCN) model was proposed [37]. L. Mou et al. constructed a graph-based end-to-end semi-supervised network, which was called the non-local GCN, that utilized both labeled and unlabeled data [38]. A spectral-spatial 3D fully convolutional network (SS3FCN) was designed for the simultaneous exploration of spectral-spatial and semantic information [39]. In Ref. [40], a fully convolutional neural network was introduced by including de-convolution layers and an optimized ELM for HSI classification. To augment the available features, Zhu et al. [41] first explored a generative adversarial network (GAN) for HSI classification, and it demonstrated better performance with limited training samples, as compared to some traditional CNNs. Nevertheless, the patch-wise-based GAN and CNN exposed the computational redundancy problem caused by the repetition of the patches of adjacent pixels during the training and testing processes.

In practical applications, high-dimensional spectral features and limited labeled samples have consistently challenged classification tasks. As a consequence, a number of unlabeled samples have been utilized to generate pseudo-labeled samples in order to increase the number of training samples and improve the performance of the classifier. Zhang et al. presented a semi-supervised classification algorithm that was based on simple linear iterative cluster (SLIC) splitting [42], and it was expected to improve the efficiency of an extended training set by selecting pseudo-labeled samples (PLSs). Considering the number of unlabeled samples has also provided abundant discriminant spectral-spatial features. Mingmin Chi et al. presented a continuation-method-based local optimization algorithm for global optimization, which was tuned with an iterative learning procedure

during the learning phase of the semi-supervised support vector machines (S3VMs) [43]. A non-parametric and kernel-based transductive support vector machine (TSVM) classification framework was proposed by L. Bruzzone to alleviate the Hughes phenomenon [44]. Meanwhile, a semi-supervised learning framework, based on spectral-spatial graph convolutional networks [36,45] and generative adversarial networks [46,47], was also exploited to increase the accuracy of the HSI classification by mitigating problems caused by limitations in the labeling samples.

In order to eliminate the computation redundancy caused by patch-wise-based algorithms and to fully utilize the abundance of unlabeled samples in an efficient way, we established a novel active inference transfer convolutional fusion network (AI-TFNet) for HSI classification. We have highlighted the notable outcomes of the proposed AI-TFNet as follows:

- In the proposed AI-TFNet, an active inference pseudo-label propagation algorithm for spatial homogeneity samples was constructed by utilizing the proposed TFNet to segment the homogeneous area, and the proposed spectral-spatial similarity metric learning function was constructed to select propagated pseudo-labels for spectral-spatial homogeneity and continuity. Meanwhile, an end-to-end, fully hybrid multi-stage transfer fusion network (TFNet) was designed for improving classification performance and efficiency.
- A metric confidence-augmented pseudo-label loss function (CapLoss) was designed to define the confidence of a pseudo-label by automatically assigning an adaptive threshold in homogeneous regions for acquiring homogeneous pseudo-label samples, which could actively infer the pseudo-label by augmenting the homogeneous training samples, based on spatial homogeneity and spectral continuity.
- In addition, to reveal and merge the local low-level and global high-level spectral-spatial contextual features during different feature extraction stages, a fully hybrid multi-stage transfer convolutional fusion network was designed to achieve end-to-end HSI classification and improve classification efficiency.

Experimental results demonstrated that, compared to other related algorithms, our proposed AI-TFNet achieved better results on several different HSI scenario datasets in terms of accuracy and efficiency.

The rest of this paper is organized as follows. In Section 2, we introduce our proposed algorithm in detail. In Section 3, the parameters' analysis and experimental results are illustrated and discussed. Finally, conclusions are drawn in Section 4.

2. Methodology

The proposed AI-TFNet classification framework was mainly categorized into the following parts: transfer fusion convolutional network (TFNet) for hyperspectral image classification; active inference for pseudo-label augmentation with adaptive threshold metric strategy (AI); and the proposed metric confidence augmented pseudo-label loss function (CapLoss). The whole flowchart of the proposed AI-TFNet is shown in Figure 1; we introduce the aforementioned parts in detail in this section.

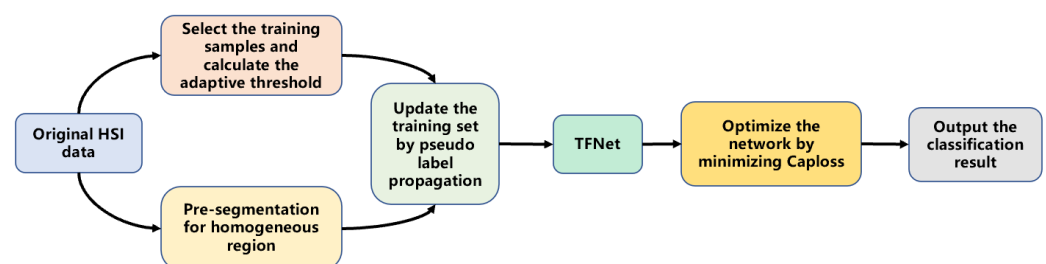


Figure 1. The overall flowchart of the proposed AI-TFNet.

2.1. Multi-Scale Transfer Fusion Convolutional Network

CNN-based algorithms have demonstrated satisfying feature extraction abilities in the computer vision field. However, several shortcomings have also been exposed, such as the loss of the location feature information by a fully connected layer in a CNN. Patch-wise-based CNN algorithms usually lead to computational redundancy, as the data in adjacent patches is calculated repeatedly. Therefore, in this paper, we constructed a hybrid multi-stage spectral–spatial fully convolutional transfer fusion network (TFNet) that captured the global spectral–spatial features during processing. We designed the proposed spatial convolutional layer and spectral convolutional layer tier by tier to augment and complement the spatial and spectral features identified by the proposed hybrid spectral–spatial (HSS) block. In the proposed structure, the multi-scale spectral information in different layers was exchanged to consolidate the discriminant information in the spectral features. The proposed HSS block is shown in Figure 2. Meanwhile, the local spatial features in shallow layers and the contextual features in deep layers were combined and exchanged in parallel to efficiently merge the spectral–spatial features in different stages. The proposed TFNet structure is shown in Figure 3. The key model HSS block in the proposed TFNet effectively exchanged and merged the spectrum and spatial feature information.

2.1.1. HSS Block

As shown in Figure 2, in each HSS block, the spectral C_{spe}^l and spatial feature maps C_{spa}^l in the upper layer were utilized as input. Meanwhile, merged features were simultaneously extracted by the combined spectral–spatial features in distinct stages. A 1×1 convolution kernel was exploited for the extraction of spectral information. Therefore, each channel of the convolution layer can be expressed as:

$$E_{spe}^{l+1,k} = \sum_j \left(w^k * \left(C_{spe}^{l,j} + C_{spa}^{l,j} \right) \right) + b^k \quad (1)$$

$$C_{spe}^{l+1,k} = \text{ReLu} \left(\text{BN} \left(E_{spe}^{l+1,k} \right) \right) \quad (2)$$

where $*$ represents the 2D convolution operation. E is an intermediate variable to simplify the interpretation. Furthermore, $C_{spe}^{l,k}$ is the k -th channel of the l -th spectral feature map, and $C_{spa}^{l,j}$ is the j -th channel of the feature map on the l -th spatial feature map. The variable w^k represents the k -th convolution kernel, b^k is the bias term of the k -th channel of the feature map, and $\text{ReLu}(x) = \max(0, x)$ is the linear rectification function where $\text{BN}(\cdot)$ represents the batch normalization function.

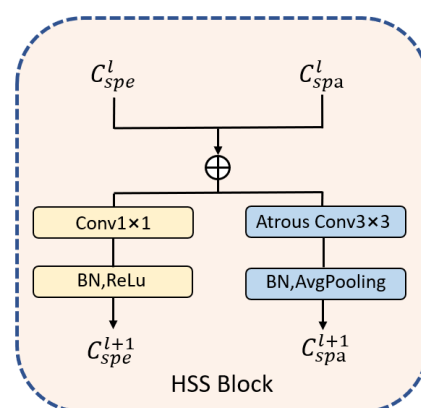


Figure 2. The structure of the HSS Block, where $C_{spe}^{l,k}$ and $C_{spa}^{l,j}$ are the k -th channel of the l -th spectral feature map and spatial feature map, respectively.

Similarly, atrous convolution was utilized as the basic operation in the multi-scale spatial feature extraction. The spatial feature map in each channel can be expressed as:

$$E_{spa}^{l+1,k} = \sum_j \left(w^k \otimes \left(C_{spa}^{l,j} + C_{spe}^{l,j} \right) \right) + b^k \quad (3)$$

$$C_{spa}^{l+1,k} = \text{AP} \left(\text{BN} \left(E_{spa}^{l+1,k} \right) \right) \quad (4)$$

where $\otimes$ represents the 2D atrous convolution operation. Furthermore, $C_{spa}^{l,k}$ is the k -th channel of the l -th spatial feature map, and $C_{spe}^{l,j}$ is the l -th channel of the feature map in the j -th spectral feature map. In addition, $\text{AP}(\cdot)$ is the 2D average pooling function. The application of atrous convolution enhanced the perceptual field significantly without increasing the computational costs, and it enhanced the spatial feature extraction performance.

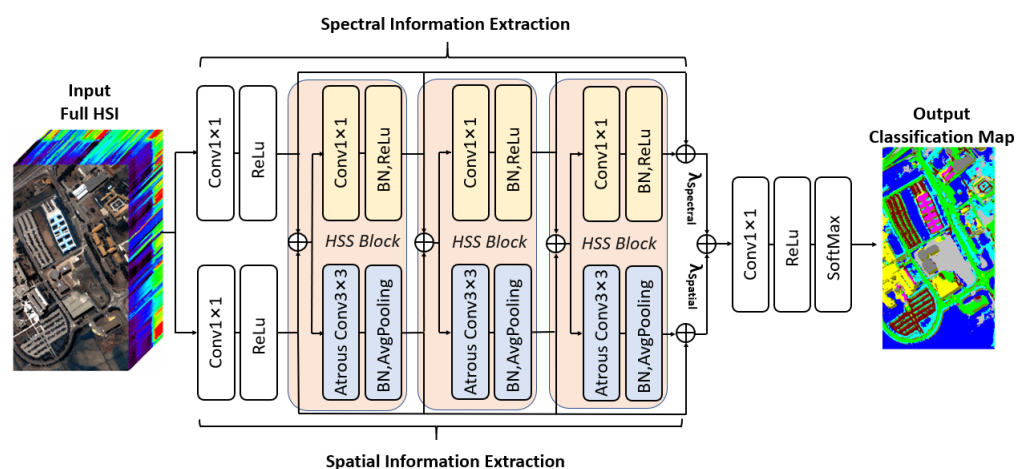


Figure 3. The architecture of the proposed TFNet for HSI classification. The information extracted by the spectral branch and spatial branch were fused by stacked HSS blocks. The merged feature map was combined into two feature maps with weighted edges.

2.1.2. TFNet

The complete TFNet structure is shown in Figure 3. In order to reduce the dimensions of the channel, the feature map of the first layer of spatial and spectral features was obtained by a 1×1 point-wise convolution. The low-level feature maps typically represent the detailed local contour features, and the high-level feature maps usually represent the contextual and semantic features. Thereby, by stacking HSS blocks, the discriminant information in the low-level feature map and the high-level contextual features can be efficiently enhanced and presented. Through this procedure, the omitted information can be supplemented and enhanced during the convolutional process. Therefore, the hybrid multi-stage spectral–spatial feature extraction not only revealed deep spatial and spectral contextual features but also augmented low-level pixel-wise spectral–spatial features. Furthermore, for preserving and merging more feature information, the feature maps extracted from each layer were fused to form an integrated layer, which can be expressed as:

$$\begin{aligned} C_{spa}^{merge,k} &= \sum_{l=1}^4 C_{spa}^{l,k} \\ C_{spe}^{merge,k} &= \sum_{l=1}^4 C_{spe}^{l,k} \end{aligned} \quad (5)$$

In the integrated layer, the spectral feature C_{spe}^{merge} and spatial feature C_{spa}^{merge} were extracted from HSS blocks with the corresponding weighting factors. The weights can

be learned automatically, allowing the model to adapt to different spectral and spatial conditions in HSIs. The integrated layer formed by weighted fusion can be represented as follows:

$$C_{unite} = \lambda_{spectral} C_{spectral} + \lambda_{spatial} C_{spatial} \quad (6)$$

The proposed TFNet classified a complete HSI image as input and ensured that only the labels from training samples were used for loss calculation and network optimization. When we only considered the labeled training samples, the loss function of TFNet can be expressed as follows:

$$L = \frac{1}{m} \sum_{i=1}^m Y_i \log(\hat{Y}_i) \quad (7)$$

where L is the cross-entropy loss function, and Y_i and $\hat{Y}_i$ are the labels and prediction labels from the training sample x_i , respectively.

2.2. Active Inference for Pseudo-Label Augmentation

The pseudo-label propagation algorithms assigned a pseudo-label, and the confidence of the pseudo-label was determined by calculating the distance between the labeled pixel and the unlabeled pixels, which were both located in the same homogeneous region obtained by clustering for augmenting the available labeled sample. Therefore, we exploited both spectral similarities and location metrics to measure the distance for labeling the pseudo-label probability of unlabeled samples as the given training samples for the homogeneous area. As shown in Figure 4, after considering that the smaller spectral distance of two pixels had a high probability of being in the same category, despite being located far from each other, we first calculated the spectral distance between the two pixels by the spectral feature metric for labeling the unlabeled samples. Then, we calculated the position relevance between the labeled and pseudo-labeled samples by the spatial location metric in order to assign confidence scores for the pseudo-labeled samples, which were based on the hypothesis that the pixels located closer to each other were more likely to belong to the same category.

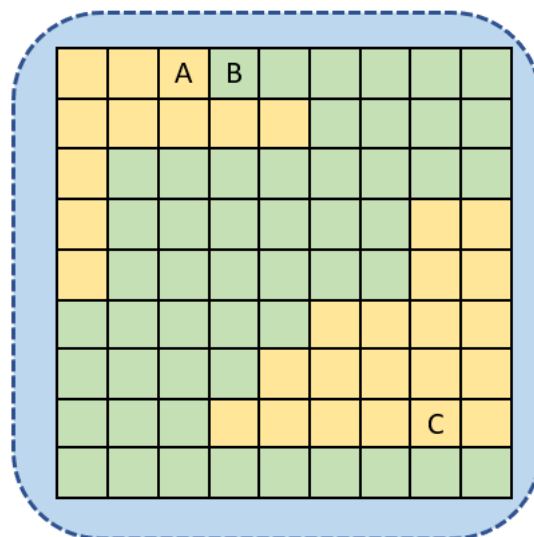


Figure 4. The ground distribution of an HSI is very complex. Pixel A and pixel B are very close in spatial location, but the categories are different. Pixel A and pixel C are far from each other but have the same category.

2.2.1. Pre-Classification of HSI

The accuracy of the hyperspectral image classification task is limited by the number of training samples, so there have been many methods described in recent years for increasing training samples [28]. Existing information in hyperspectral data, even without the label information, has been used to increase training samples. For supervised classification tasks, the domain is accessible as $\{\mathbf{x}_i, y_i\}_{i \in [N^s]}$ with N^s data points $\mathbf{x}_i$ and the corresponding labels y_i from a discrete $y_i \in \mathcal{Y} = \{1, \dots, Y\}$. For an unsupervised pre-segmentation task, the accessible domain includes N^u data points in $\{\mathbf{x}_i\}_{i \in [N^u]}$. Obviously, these two domains have the same distribution: $\mathbf{x}_i \in \mathcal{X}$. Considering a situation where we had two eigenfunctions $\Phi_s, \Phi_u : \mathcal{X} \rightarrow \mathcal{R}^d$, we used the eigenfunction to map the original distribution $\mathcal{X}$ to $\mathcal{R}^d$, as applicable for classification and pre-classification, respectively. We emulated the common features through a subset of parameters that were shared among the feature functions as $\Phi_s = \Phi_{\theta_c, \theta_s}$ and $\Phi_u = \Phi_{\theta_c, \theta_u}$ by implementing TFNet as a feature extraction function. The variable Φ_{θ_c} corresponded to the parameters in the first few layers of the TFNet network, and θ_s, θ_u corresponded to the relevant last layers. We selected k -means unsupervised clustering for the original hyperspectral images to obtain the clustering label matrix $L_{clu} \in \mathbb{R}^{H \times W}$, where H and W are the width and height of the label matrix. TFNet and clustered labels were utilized to pre-classify hyperspectral images in order to obtain the pre-segmentation label matrix $L_{pre} \in \mathbb{R}^{H \times W}$ and the pre-segmentation network weight Φ_{θ_c} .

2.2.2. Spectral–Spatial Adaptive Threshold

In order to actively augment the unlabeled samples by the pre-segmentation label matrix, we assigned an adaptive threshold that provided an adaptive circumstance-changing strategy. For each pixel $\mathbf{x}_i$, the spectral distance D_{spe} between any pixel $\mathbf{x}_i$ and the specified pixel $\mathbf{x}_j$ was defined by the following:

$$D_{spe} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad (8)$$

We defined the training set as $\mathbf{X}_{train} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$, and its label set is $Y = \{Y_1, Y_2, \dots, Y_M\}$, where N and M are the number of classes and the number of labeled samples, respectively. $\mathbf{X}_i = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_{n_i}^i\}$ represents the i th training set. In order to adaptively evaluate the similarity of the unlabeled samples in the same pre-classification region, we first calculated the average spectral vector $\bar{\mathbf{x}}_i$ of the i th training set $\mathbf{X}_i$, according to the following formula.

$$\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_j^i \quad (9)$$

In order to adaptively retain homogeneous region samples that were similar to the target-labeled samples while eliminating dissimilar samples, we defined an adaptive threshold β as the minimum inter-class distances by calculating the D_{spe} of all the mean vectors for N classes comprehensively. The details are described in Algorithm 1. Thereby, we used the adaptive threshold β to adaptively and actively expand the available training set. Since the pre-segmentation area containing the labeled sample $\mathbf{x}_i$ consisted of r_i unlabeled sample set $S_i^u = \{\mathbf{x}_1^u, \mathbf{x}_2^u, \dots, \mathbf{x}_{r_i}^u\}$, the distances between the labeled sample $\mathbf{x}_i$ and all r_i unlabeled samples were calculated by Equation (8), where the distance reflects the level of similarity between the unlabeled and labeled target samples in the same pre-segmentation area. By calculating all of the distances between r_i distinct unlabeled samples, we could select all of the samples for which the distances were smaller than the threshold β to propagate the pseudo-labels as the same as the labeled sample and augment the available pseudo-label sets $\mathbf{X}_i^p = \{\mathbf{x}_{i,1}^p, \mathbf{x}_{i,2}^p, \dots, \mathbf{x}_{i,p_i}^p\}$ by the following function:

$$\mathbf{x}_{i,j}^p = \begin{cases} \mathbf{x}_j^u & D_{spe}(\mathbf{x}_i, \mathbf{x}_j^u) < \beta \\ \text{deleted} & \text{otherwise} \end{cases} \quad (10)$$

$$Y_{i,j}^P = Y_i \quad (11)$$

2.2.3. Spectral–Spatial Confidence Metric

From Figure 4, we observed two phenomena: (1) the pixels located at the boundary of the different classes belonged in different categories even when they were close to each other (such as pixels A and B in Figure 4), and (2) the pixels that were far apart could also belong to the same category based on having similar spectral signatures (such as pixels A and C in Figure 4). Therefore, a compromise between the spatial location metric and the spectral feature similarity metric had to be designed to address the differences in categories between those with a minimal spatial distance between them, as well as for those in the same category despite a minimal spectral distance between them. Since the pseudo-labels' assignments were based on the spectral metric, we proposed that pseudo-label confidence weights be used to enforce the spatial relations through the spatial metric.

Algorithm 1: Label propagation based on pre-segmentation map with spectral and spatial metrics.

Input: Training set $\mathbf{X}_{train} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ and its label set $Y = \{Y_1, Y_2, \dots, Y_M\}$; pre-segmentation region of each training sample $S_i^u = \{\mathbf{x}_1^u, \mathbf{x}_2^u, \dots, \mathbf{x}_{r_i}^u\}$; the adaptive threshold β .

- 1 **for** $1 \leq i \leq M$ **do**
- 2 Calculate all the r_i spectral distances between $\mathbf{x}_i$ and all unlabeled samples in S_i^u by (8).
- 3 **for** $1 \leq j \leq r_i$ **do**
- 4 Assign the $\mathbf{X}_i^p = \{\mathbf{x}_{i,1}^p, \mathbf{x}_{i,2}^p, \dots, \mathbf{x}_{i,p_i}^p\}$ as the same label with $\mathbf{x}_i$ to generate $Y_{i,j}^P$ by (10) and (11).
- 5 Assign the confidence weight $Conf_j$ by (13).
- 6 Augment the sample, pseudo-label, and confidence weight to the pseudo-sample set $\mathbf{X}_{pseudo}$, pseudo-label set Y_{pseudo} , and confidence weight set $Conf_{pseudo}$, respectively.
- 7 $\mathbf{X}_{pseudo} \leftarrow \mathbf{X}_i^p$
- 8 $Y_{pseudo} \leftarrow Y_{i,j}^P$
- 9 $Conf_{pseudo} \leftarrow Conf_j$
- 10 **end**
- 11 **end**

Output: Pseudo-label sample set $\mathbf{X}_{pseudo}$, pseudo-label set Y_{pseudo} , and confidence weight set $Conf_{pseudo}$.

For each pseudo-labeled sample $\mathbf{x}_{i,k}^p$ belonging to $\mathbf{X}_i^p = \{\mathbf{x}_{i,1}^p, \mathbf{x}_{i,2}^p, \dots, \mathbf{x}_{i,p_i}^p\}$, we defined and calculated the spatial location distance D_{spa} by (12). Furthermore, the confidence weighting function $Conf_k$ indicated the possibility that $\mathbf{x}_{i,k}^p$ and $\mathbf{x}_i$ could belong to the same class, which was defined as follows.

$$D_{spa}^k = \sqrt{(h_i - h_k)^2 + (v_i - v_k)^2} \quad (12)$$

$$Conf_k = 1 - \frac{D_{spa}^k}{D_{spa}^{\max}} \quad (13)$$

where (h_i, v_i) and (h_k, v_k) are the spatial coordinates of $\mathbf{x}_i$ and $\mathbf{x}_{i,k}^p$. Furthermore, $D_{spa}^{\max}$ is the maximum spatial distance between all pseudo-labeled samples $\mathbf{x}_{i,k}^p$ and target-labeled samples $\mathbf{x}_i$. The confidence weighting function was based on the hypothesis that pixels located closer to each other were more likely to belong to the same category.

Therefore, the proposed adaptive homogeneous label propagation strategy derived homogeneous samples in the same pre-segmentation area. The complete pseudo-label propagation process is illustrated in Figure 5. The complete procedure is summarized in Algorithm 1.

2.3. The Proposed CapLoss Function

The procedure of the proposed AI-TFNet classification framework is as follows. First, through this active inferential pseudo-label propagation strategy, we obtained the pseudo-label samples, pseudo-label confidence, and the training weight of our pre-segmentation. Next, the pseudo-label samples were added to the original training set, and then, the TFNet was trained after the initialized pre-training weight.

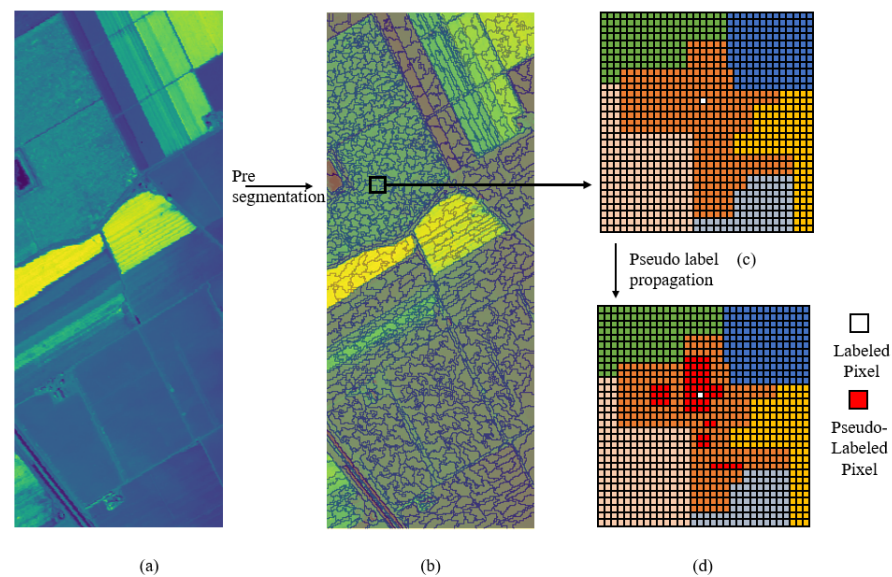


Figure 5. The pseudo-label propagation of HSI data. (a) is the HSI after the dimension reduction algorithm (e.g., PCA). (b) is the image pre-classification by TFNet. Different colored regions represent the different score areas on the table. (c) is the small portions of the hyperspectral image after pre-classification. In (d), the increased number of red squares indicates the pseudo-label is the same as the original label propagation in the pre-segmentation map.

In addition, AI-TFNet was a more efficient classification strategy due to exploiting the pseudo-label propagation strategy when the number of labeled samples was limited. Therefore, we speculated that the original training samples would have a greater impact on the loss reduction, and the pseudo-label samples would adaptively participate in the loss calculation. Therefore, the final objective function was mainly composed of two items: the loss of the labeled samples and the loss of the pseudo-labeled samples with the confidence factor Cof_i to balance the two items. A metric known as the "confidence augmented score" pseudo-label loss (CapLoss) function L_{Cap} for AI-TFNet was defined as follows.

$$L_{Cap} = L + L_{pseudo} \quad (14)$$

$$L_{pseudo} = \frac{1}{\tilde{p}} \sum_i^{n_p} \left\{ Cof_i \times Y_{pseudo}^i \log(\hat{Y}_{pseudo}^i) \right\} \quad (15)$$

$$\tilde{p} = \sum_{i=1}^m p_i \quad (16)$$

where $\tilde{p}$ is the number of pseudo-labeled samples and Y_{pseudo}^i , $\hat{Y}_{pseudo}^i$, and $Conf_i$ are the labels, prediction labels, and confidence weights of the pseudo-labeled sample x_p^i , respectively. The proposed CapLoss function efficiently utilized the labeled training sample set and also exploited the augmented pseudo-sample set according to the confidence weights, which efficiently improved the classification accuracy and performance even with a limited training sample size. The complete procedure for AI-TFNet is summarized in Algorithm 2.

Algorithm 2: The procedure of AI-TFNet.

Input: The original HSI $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$; training set

$\mathbf{X}_{train} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N\} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$ and its label set
 $Y = \{Y_1, Y_2, \dots, Y_M\}$.

- 1 Reduce the dimensionality of $\mathbf{X}$ with PCA.
- 2 K -means is used to cluster the HSI after dimensionality reduction and use TFNet to acquire pre-segmentation map enclosed areas $S_i^u = \{\mathbf{x}_1^u, \mathbf{x}_2^u, \dots, \mathbf{x}_{r_i}^u\}$ of each training sample and the weight W of the network.
- 3 Generate the adaptive threshold β using Algorithm 1.
- 4 Generate pseudo-sample set $\mathbf{X}_{pseudo}$, pseudo-label set Y_{pseudo} , and confidence weight set $Conf_{pseudo}$ using Algorithm 2.
- 5 Initialize AI-TFNet with weights $\mathbf{W}$ and use the training set $\mathbf{X}_{train}$ and pseudo-label sample set $\mathbf{X}_{pseudo}$ to train the AI-TFNet and update the parameters of the network with the CapLoss by (14)–(16).
- 6 The original HSI is used as the model input to obtain the classification map.

Output: The classification map.

3. Experiment

In this section, we evaluate the proposed AI-TFNet on several commonly used HSI datasets, such as the University of Pavia dataset, the Salinas dataset, and the Houston dataset. The parameters' analysis of the proposed algorithm, the comparison of classification accuracy, and the classification performance results with several existing relative algorithms are illustrated and analyzed in this section.

3.1. Hyperspectral Datasets

The first dataset was obtained from the Reflective Optical System Imaging Spectrometer (ROSIS) in the sky over the University of Pavia in northern Italy. The original image consisted of 610×340 pixels with 103 spectral bands that covered from 430 nm to 860 nm with a spatial resolution of 1.3 m. The training and test sample sizes and the nine labeled classes are shown in Table 1 and Figure 6.

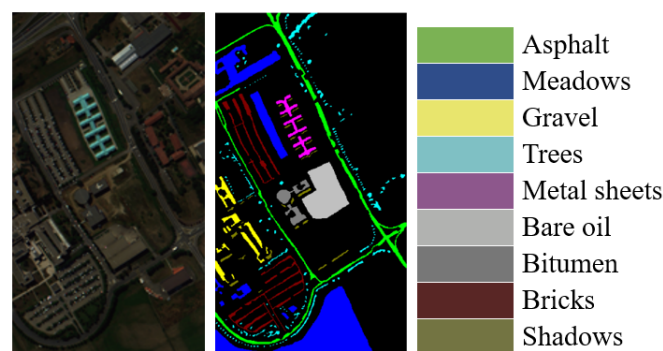


Figure 6. (Left) False color image and (Right) ground truth map of the University of Pavia dataset.

The second dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in the sky over Salinas Valley, California. The original image was composed

of 512×217 pixels and had a high spatial resolution of 3.7 m/pixel; it consisted of 204 spectral bands. The training and test sample sizes and 16 labeled classes are shown in Table 2 and Figure 7.

Table 1. The numbers of the training and testing sampling for the University of Pavia dataset.

Class	Class Name	Total	Train	Test
1	Asphalt	6631	20	6611
2	Meadows	18,649	20	17,136
3	Gravel	2099	20	2079
4	Trees	3064	20	3044
5	Metal sheets	1345	20	1325
6	Bare Soil	5029	20	5009
7	Bitumen	1330	20	1310
8	Bricks	3682	20	3662
9	Shadows	947	20	927
Total		42,776	180	42,596



Figure 7. (Left) False color image and (Right) ground truth map of the Salinas dataset.

Table 2. The numbers of the training and testing sampling for the Salinas dataset.

Class	Class Name	Total	Train	Test
1	Broccoli green weeds 1	1977	10	1967
2	Broccoli green weeds 2	3726	10	3716
3	Fallow	1976	10	1966
4	Fallow rough plow	1394	10	1384
5	Fallow smooth	2678	10	2668
6	Stubble	3959	10	3949
7	Celery	3579	10	3569
8	Grapes untrained	11,213	10	11,203
9	Soil vineyard develop	6197	10	6187
10	Corn senescent green weeds	3249	10	3239
11	Lettuce romaine 4wk	1058	10	1048
12	Lettuce romaine 5wk	1908	10	1898
13	Lettuce romaine 6wk	909	10	898
14	Lettuce romaine 7wk	1061	10	1051
15	Vineyard untrained	7164	10	7154
16	Vineyard vertical trellis	1737	10	1727
Total		53,785	160	53,625

The last dataset was captured by the ITRES-CASI 1500 sensor over the University of Houston campus and its neighboring urban areas in Texas. The original image consisted of 1905×349 pixels with 144 spectral bands that covered from 380 nm to 1050 nm with a

spatial resolution of 2.5 m. The training and test sample sizes and the 15 labeled classes are shown in Table 3 and Figure 8.

Table 3. The numbers of the training and testing sampling for the Houston dataset.

Class	Class Name	Total	Train	Test
1	Grass healthy	1251	10	1241
2	Grass stressed	1254	10	1244
3	Grass synthetic	697	10	687
4	Trees	1244	10	1234
5	Soil	1242	10	1232
6	Water	325	10	315
7	Residential	1268	10	1258
8	Commercial	1244	10	1234
9	Road	1252	10	1242
10	Highway	1227	10	1217
11	Railway	1235	10	1225
12	Parking lot1	1233	10	1223
13	Parking lot2	469	10	459
14	Tennis court	428	10	418
15	Running track	660	10	650
Total		15,029	150	14,879

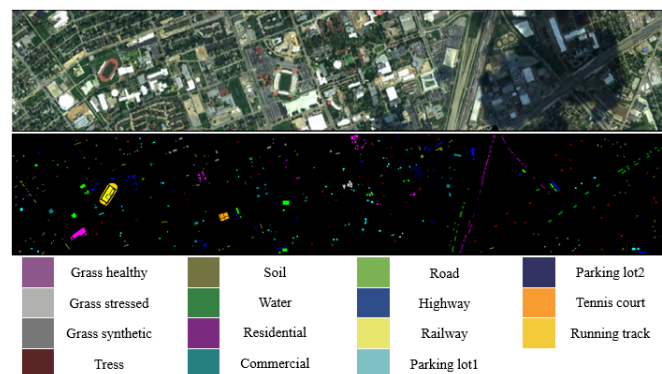


Figure 8. (Up) False color image and (Down) ground truth map of the Houston dataset.

3.2. Parameter Effect Analysis

In this section, we analyze the impact of the parameters of the proposed TFNet and AI-TFNet on the different datasets. Dilated convolution was used in TFNet as a basic operation of spatial feature extraction. The dilation rate K of the dilated convolution and the numbers of the trained samples in each class were the main parameters analyzed in this section. The experimental results were evaluated based on their overall accuracy (OA). During the training process, we used the Adam optimization algorithm as an optimizer, with a learning rate of 0.000025.

The effects of the dilation rate K of the atrous convolution for the different datasets are illustrated in Figure 9. The atrous convolutions with a varying dilation rate K efficiently exploited the different perceptions of the spatial regions. We observed that the OA value increased slowly when K increased to four and then decreased slowly when K increased to eight, which indicated the small dilation rate was likely to overlook contextual information, while a larger dilation rate was likely to overwhelm the network when attempting to capture detailed local features. Therefore, we selected an optimal dilation rate K of four in the following experiments.

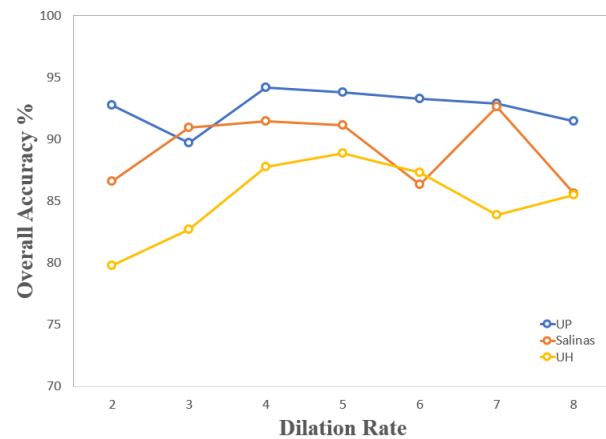


Figure 9. OA curve of two datasets with different dilation rates.

In the second set of experiments, we compared our proposed method with the following algorithms: (1) a spectral-based SVM (SVM) [6]; (2) an SVM with CK (SVMCK) [14]; (3) a dual-channel capsule GAN (DcCapsGAN) [48]; (4) a spectral-spatial based LMAFN (LMAFN) [31]; (5) a spectral-spatial residual network (SSRN) [26]; and (6) a spectral-spatial fully convolutional network (SSFCN). These were carried out on three HSI datasets, used as benchmarks, for different training sampling rates [49]. As shown in Figure 10, it was evident that the OA improved when the training size increased. Deep-learning-based algorithms (DcCapsGAN, LMAFN, SSRN, SSFCN, TFNet), when compared with conventional machine-learning methods (SVM, SVMCK), presented more requirements for the labeled training set, further proving that deep-learning models require a large dataset for achieving better performance. The proposed pseudo-label propagation strategy demonstrated that AI-TFNet yielded the most robust results across all sampling rates, especially on training samples that were limited in size. Furthermore, AI-TFNet yielded a considerable improvement in classification accuracy, as compared to TFNet, due to active pseudo-label propagation learning.

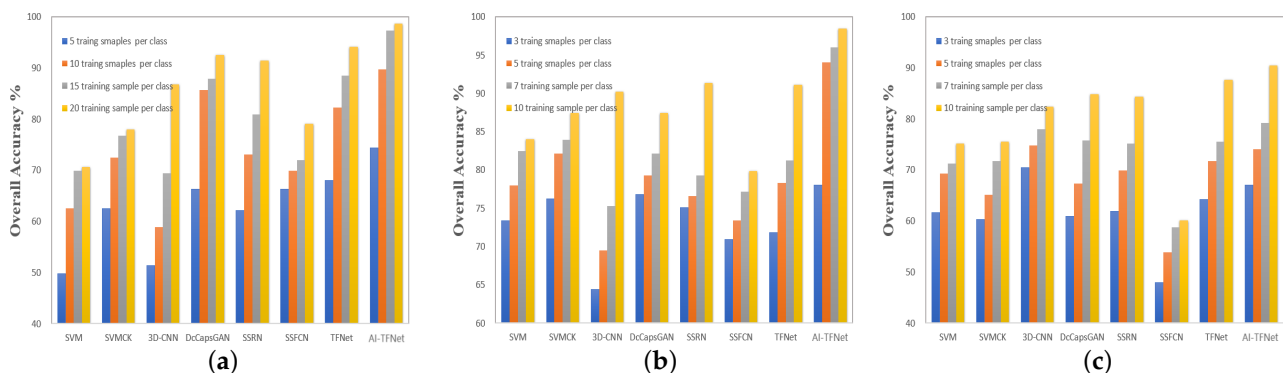


Figure 10. The overall accuracy of different methods on three datasets with different training sampling rate: (a) the University of Pavia dataset; (b) the Salinas dataset; and (c) the Houston dataset.

3.3. Ablation Experiment

At this point, we conducted an ablation test and verification to demonstrate the advantage of the proposed active inference. Inaccurate pseudo-label samples could have a negative impact on the classification results, which could then reduce the spatial and spectral constraints of our pseudo-inference propagation strategy. The spatial constraint was defined as our pre-segmentation results having produced adequate and rational homogeneous regions. The spectral constraint was defined as when the pseudo-label samples would be introduced, that is, only when the spectral distance was less than

the minimum inter-class distance. These two constraints were imposed in tandem to administrate the pseudo-label propagation. The complete propagation results are shown in Tables 4 and 5. These results demonstrated that a large number of pseudo-samples were involved in this procedure and only a few incorrect labels were identified.

In the second set of experiments, we conducted an ablation test to demonstrate the advantage of the proposed CapLoss approach. The results in Table 6 indicated that AI-TFNet combined with CapLoss yielded better OA accuracy than the original cross-entropy losses for the three datasets. This ablation experiment further demonstrated that CapLoss had extracted information from the pseudo-labeled samples based on the generated confidence score, indicating that it could efficiently provide additional useful information for optimizing the whole network.

Table 4. Number of original samples, pseudo-label samples, and incorrect pseudo-label samples of each class for the University of Pavia dataset.

Class	Original	Pseudo	Incorrect Pseudo Labels
1	20	2736	0
2	20	3704	0
4	20	3913	0
5	20	240	0
6	20	1498	0
7	20	1072	0
8	20	4319	0
9	20	2417	0
Total	180	17,899	0

Table 5. Number of original samples, pseudo-label samples, and incorrect pseudo-label samples of each class for the Houston dataset.

Class	Original	Pseudo	Incorrect Pseudo Labels
1	10	4094	0
2	10	1412	0
3	10	7809	0
4	10	1582	0
5	10	1865	0
6	10	1608	0
7	10	1067	0
8	10	3364	12
9	10	1164	0
10	10	8684	0
11	10	2856	0
12	10	3201	0
13	10	983	0
14	10	1599	0
15	10	1273	0
Total	150	42,561	12

Table 6. OA(%) for AI-TFNet using cross-entropy loss or CapLoss for different datasets.

Loss Function	UP	Salinas	Houston
Cross-entropy loss	98.57	98.22	89.89
CapLoss	98.73	98.59	90.74

Furthermore, for verifying the effectiveness of the active inference on the parameter migration and the sample expansion in the proposed pseudo-label propagation, we

observed in Table 7 that the active inference transfer parameter had improved the classification accuracy, which ensured a more precise and efficient initialization for TFNet. As a result, TFNet could then provide better convergence results. The active sample expansion increased the diversity of the training set samples and improved the classification capacity of the network. The experiments on multiple datasets further confirmed the efficiency and suitability of the proposed AI-TFNet.

Table 7. Classification accuracy of ablation experiments for three datasets. (The best results are represented in bold).

TFNet	Pre-Seg Transfer	AI Sample Expansion	UP	Salinas	Houston
✓			94.16	91.09	87.58
✓	✓		94.57	91.50	88.69
✓		✓	96.05	97.47	88.10
✓	✓	✓	98.64	98.56	90.50

3.4. Classification Result and Analysis

In this section, we compare our proposed method with the aforementioned algorithms on the three different datasets. A total of 20, 10, and 10 samples from each class were selected as the labeled training samples for the University of Pavia, Salinas, and Houston datasets, respectively. The RGB image segmentation model Deeplabv3 was also evaluated in our experiments, as its loss function design was similar to that used in our proposed TFNet. The training and test sets of three datasets are listed in Tables 1–3. The classification results with the mean and the standard deviation values of the different algorithms are summarized in Tables 8–10, and ten random iterations were performed in order to reduce any potential bias. The optimal results are indicated in bold.

Table 8. Classification accuracy of different methods on the University of Pavia dataset with 20 labeled samples for 10 random iterations. (The best results in each row are represented in bold).

Class	SVM	SVMCK	DcCapsGAN	LMAFN	SSRN	SSFCN	TFNet	AI-TFNet
1	62.80 ± 4.04	71.24 ± 6.06	93.77 ± 0.05	93.63 ± 5.21	99.74 ± 0.81	58.10 ± 3.12	88.71 ± 2.71	97.81 ± 0.41
2	65.35 ± 1.89	67.30 ± 7.50	90.04 ± 0.01	88.88 ± 7.31	99.10 ± 0.62	84.91 ± 3.29	92.55 ± 2.12	99.90 ± 0.13
3	72.15 ± 5.68	90.73 ± 2.28	81.99 ± 0.09	99.39 ± 0.70	79.45 ± 21.15	79.36 ± 6.78	95.33 ± 3.14	100 ± 0
4	92.54 ± 1.28	95.79 ± 1.08	97.16 ± 0.03	97.54 ± 1.15	78.89 ± 4.57	89.09 ± 2.11	98.42 ± 0.84	98.52 ± 0.18
5	99.14 ± 0.41	99.62 ± 0.32	99.92 ± 0.06	99.88 ± 0.14	99.92 ± 0.13	95.54 ± 0.12	100 ± 0	100 ± 0
6	65.03 ± 10.46	92.40 ± 3.01	95.93 ± 0.01	97.51 ± 2.48	85.93 ± 18.10	81.69 ± 7.32	98.96 ± 0.72	94.17 ± 1.24
7	86.20 ± 1.51	92.29 ± 4.94	98.44 ± 0.03	100 ± 0	79.25 ± 8.77	86.33 ± 5.75	99.84 ± 1.31	100 ± 0
8	77.39 ± 2.54	85.29 ± 8.07	93.46 ± 0.05	86.82 ± 19.77	83.82 ± 7.85	61.82 ± 3.84	95.84 ± 3.14	98.97 ± 0.19
9	96.83 ± 1.28	95.07 ± 2.83	99.85 ± 0.13	99.01 ± 1.07	98.88 ± 1.92	99.78 ± 1.65	99.89 ± 1.78	99.65 ± 0.12
OA (%)	70.60 ± 2.49	77.98 ± 3.02	92.50 ± 0.16	92.49 ± 5.68	91.36 ± 0.03	79.11 ± 3.84	94.16 ± 2.45	98.73 ± 0.20
AA (%)	79.71 ± 2.03	87.76 ± 0.89	94.51 ± 0.02	95.85 ± 3.65	89.44 ± 2.02	73.15 ± 2.14	96.62 ± 2.74	98.81 ± 0.19
KAPPA (%)	63.13 ± 3.10	72.50 ± 3.42	90.22 ± 0.02	90.38 ± 7.06	88.82 ± 4.26	72.72 ± 3.61	92.39 ± 1.57	98.32 ± 1.82
Test Time (s)	2.89	45.10	33.62	4.23	28.54	0.13	0.07	0.06

Table 9. Classification accuracy of different methods on the Salinas dataset with 10 labeled samples for 10 random iterations. (The best results in each row are represented in bold).

Class	SVM	SVMCK	DcCapsGAN	LMAFN	SSRN	SSFCN	TFNet	AI-TFNet
1	97.09 ± 0.60	96.74 ± 3.99	99.98 ± 0.02	99.68 ± 0.28	99.95 ± 0.09	92.59 ± 0.23	100 ± 0	100 ± 0
2	98.48 ± 0.59	94.29 ± 5.52	99.97 ± 0.02	99.00 ± 1.32	98.37 ± 2.59	88.83 ± 2.12	98.54 ± 1.21	99.94 ± 0.05
3	88.82 ± 4.52	85.33 ± 10.27	99.98 ± 0.02	96.02 ± 1.87	94.05 ± 7.52	96.94 ± 3.75	100 ± 0	100 ± 0
4	99.42 ± 0.37	99.08 ± 0.04	99.90 ± 0.02	98.60 ± 2.80	97.04 ± 3.41	99.71 ± 1.02	100 ± 0	97.22 ± 2.49
5	96.80 ± 1.15	95.51 ± 5.43	96.28 ± 0.03	96.96 ± 2.62	98.72 ± 1.00	87.29 ± 5.85	95.61 ± 2.14	98.61 ± 0.78
6	99.29 ± 0.20	97.72 ± 3.21	99.96 ± 0.02	99.97 ± 0.05	99.70 ± 0.44	99.29 ± 1.13	98.75 ± 2.14	100 ± 0
7	99.36 ± 0.09	90.06 ± 9.00	99.95 ± 0.03	99.96 ± 0.04	98.65 ± 2.30	91.11 ± 3.78	99.91 ± 0.51	99.97 ± 0.03
8	70.76 ± 13.07	73.38 ± 1.07	53.60 ± 0.02	85.04 ± 6.35	88.44 ± 4.16	56.46 ± 17.52	75.57 ± 3.85	96.56 ± 0.94
9	97.03 ± 0.62	95.74 ± 2.38	99.96 ± 0.01	99.93 ± 0.13	99.36 ± 0.58	87.22 ± 4.86	99.62 ± 1.21	99.80 ± 0.16
10	79.43 ± 8.15	89.57 ± 2.67	93.51 ± 0.03	94.39 ± 1.05	95.03 ± 3.13	87.08 ± 5.78	84.36 ± 3.75	99.08 ± 0.03
11	97.19 ± 2.13	99.74 ± 0.43	99.81 ± 0.01	99.85 ± 0.19	95.45 ± 3.55	94.8 ± 3.21	91.68 ± 2.85	99.90 ± 0.09
12	99.86 ± 0.13	95.93 ± 5.92	99.98 ± 0.03	99.87 ± 0.08	99.81 ± 0.18	97.02 ± 1.78	99.42 ± 0.77	100 ± 0
13	97.16 ± 0.16	97.57 ± 1.12	99.88 ± 0.01	99.98 ± 0.04	95.43 ± 3.71	94.92 ± 2.18	99.77 ± 0.41	98.78 ± 0.22
14	94.08 ± 0.89	97.29 ± 0.84	99.55 ± 0.05	99.87 ± 0.14	92.24 ± 9.58	80.84 ± 7.59	99.90 ± 0.37	99.34 ± 0.19
15	47.60 ± 23.66	77.48 ± 7.66	83.28 ± 0.01	78.54 ± 18.11	69.50 ± 5.22	66.25 ± 16.37	86.81 ± 5.12	96.72 ± 0.25
16	94.10 ± 3.03	84.90 ± 7.23	99.25 ± 0.03	93.46 ± 7.52	100 ± 0	64.49 ± 7.32	96.6 ± 2.14	99.47 ± 0.14
OA (%)	83.96 ± 1.95	87.43 ± 2.16	87.43 ± 0.01	93.00 ± 2.02	91.38 ± 0.87	79.84 ± 3.12	91.09 ± 2.14	98.56 ± 0.12
AA (%)	91.03 ± 1.04	91.90 ± 1.88	95.30 ± 0.02	96.32 ± 1.45	95.11 ± 0.63	86.55 ± 2.18	89.85 ± 1.75	99.08 ± 0.11
KAPPA (%)	82.16 ± 2.14	81.70 ± 2.58	86.10 ± 0.01	92.20 ± 2.27	90.43 ± 0.95	77.76 ± 2.75	90.09 ± 0.97	98.40 ± 0.13
Test Time (s)	5.46	86.60	89.29	5.35	51.98	0.16	0.06	0.06

Table 10. Classification accuracy of different methods on the Houston dataset with 10 labeled samples for 10 random iterations. (The best results in each row are represented in Bold).

Class	SVM	SVMCK	DcCapsGAN	LMAFN	SSRN	SSFCN	TFNet	AI-TFNet
1	89.77 ± 5.82	75.46 ± 5.99	85.03 ± 0.03	93.94 ± 5.34	87.65 ± 1.45	73.05 ± 0.73	95.81 ± 3.42	92.05 ± 1.99
2	87.56 ± 8.74	93.62 ± 2.80	99.86 ± 0.07	92.83 ± 4.74	97.57 ± 3.55	76.49 ± 0.67	83.31 ± 1.41	78.21 ± 0.88
3	99.71 ± 0.13	99.24 ± 0.95	98.68 ± 0.11	98.75 ± 0.73	100 ± 0	74.45 ± 0.16	95.78 ± 2.99	100 ± 0
4	89.30 ± 2.33	77.29 ± 2.80	92.78 ± 0.22	90.66 ± 2.36	98.72 ± 2.22	58.67 ± 0.25	93.65 ± 0.37	96.74 ± 0.02
5	98.62 ± 0.79	93.18 ± 2.54	98.91 ± 0.07	88.85 ± 22.06	94.93 ± 7.69	86.47 ± 0.99	99.54 ± 0.45	100 ± 0
6	89.52 ± 6.65	93.84 ± 4.52	88.46 ± 0.39	91.94 ± 5.38	100 ± 0	43.73 ± 1.44	76.93 ± 3.65	87.30 ± 0.03
7	63.67 ± 10.54	67.30 ± 18.77	78.21 ± 0.11	88.47 ± 5.40	91.67 ± 3.05	55.94 ± 0.52	86.27 ± 0.95	89.88 ± 3.39
8	55.62 ± 9.05	50.31 ± 5.76	68.09 ± 1.79	72.35 ± 4.67	98.03 ± 3.41	50.20 ± 0.27	66.67 ± 1.78	73.5 ± 1.62
9	71.58 ± 7.26	70 ± 8.70	48.95 ± 0.39	76.09 ± 6.55	75.31 ± 7.49	35.51 ± 0.49	61.84 ± 0.45	76.35 ± 4.53
10	75.02 ± 5.63	70.09 ± 4.94	93.50 ± 0.03	95.22 ± 4.70	82.14 ± 18.99	73.40 ± 0.71	92.14 ± 0.53	100 ± 0
11	55.66 ± 6.87	61.08 ± 6.14	77.49 ± 0.33	72.42 ± 14.78	54.50 ± 5.08	45.16 ± 0.07	93.36 ± 1.02	97.81 ± 0.22
12	41.03 ± 9.45	57.25 ± 7.79	88.93 ± 0.32	86.15 ± 8.77	75.67 ± 14.79	50.72 ± 0.53	91.99 ± 3.96	95.05 ± 0.36
13	37.30 ± 2.47	82.57 ± 10.59	70.73 ± 1.87	86.15 ± 8.77	93.42 ± 3.14	24.24 ± 24.24	88.45 ± 0.79	72.44 ± 0.54
14	97.89 ± 1.63	94.55 ± 4.04	99.60 ± 0.11	99.62 ± 0.77	98.82 ± 1.25	88.94 ± 88.94	100 ± 0	100 ± 0
15	98.25 ± 0.88	99.75 ± 0.42	99.94 ± 0.07	100 ± 0	95.97 ± 0.67	50.88 ± 50.88	97.49 ± 4.35	98.84 ± 0.15
OA (%)	75.13 ± 1.42	75.55 ± 0.59	84.79 ± 0.13	87.79 ± 1.34	84.35 ± 0.97	60.09 ± 0.10	87.58 ± 0.66	90.50 ± 0.23
AA (%)	76.70 ± 1.06	79.04 ± 0.68	85.95 ± 0.04	89.58 ± 1.15	89.63 ± 0.83	59.19 ± 0.10	88.21 ± 0.68	90.61 ± 0.16
KAPPA (%)	73.12 ± 1.53	73.60 ± 0.62	83.54 ± 0.14	86.79 ± 1.45	83.08 ± 1.05	56.94 ± 0.10	86.58 ± 0.72	89.73 ± 0.25
Test Time (s)	1.20	20.19	21.04	1.89	11.55	0.37	0.10	0.12

In Tables 8–10, we observed that the deep-learning-based algorithms (DcCapsGAN, LMAFN, SSRN, and TFNet) outperformed SVM in terms of their strong feature extraction abilities through convolution and nonlinear activation functions. Compared to SVM, which only utilized spectral information, the spectral–spatial-based algorithms greatly improved the classification performance due to the combination of spatial and spectral information. While the sequential spectral and spectral–spatial feature extraction method (SSRN) and spectral–spatial feature extraction method with two branches (SSFCN) fused the spectral and spatial information in their last steps, the proposed TFNet performed a fusion operation at different hybrid stages in order to exploit the spectral and spatial features at both low and high levels, which led to more representative and discriminant features for the HSI classification task. Compared to the proposed TFNet, the AI-TFNet improved the classification efficiency by adaptively propagating the pseudo-labels in the pre-segmentation regions with the proposed adaptive spectral–spatial metric threshold for

augmenting the available training datasets. Therefore, the proposed AI-TFNet achieved the best accuracy on all three datasets. In the case of 20 training samples for each class in the Pavia University dataset, the results were more than 98% accurate. For the Salinas datasets, the classification accuracy was better than 98% with only 10 labeled samples, indicating that the proposed AI-TFNet also achieved the best classification results for most categories. Therefore, the accuracy performance verified the superiority of the proposed TFNet and AI-TFNet, which further demonstrated the effectiveness of the proposed multi-stage hybrid structure with an adaptive active pseudo-label propagation learning strategy.

The classification results of the different algorithms are shown in Figures 11–13. The spectral-based SVM presented less spatial continuity in the classification map due to the loss of spatial information. Meanwhile, we observed that the only spatial information included in the algorithms was likely to omit discrete or tiny objects or misclassify the pixels around the boundary of different categories. We observe that the classification maps in Figure 12i,j contained fewer misclassified pixels than those in Figure 12c–h. Specific categories, such as the untrained grape and the untrained grape vineyard, had better connectivity and smoothness in their classification results using TFNet and AI-TFNet. Therefore, as shown in Table 10, the accuracy of these two categories was higher than for the other approaches, which further proved that the merging of the spatial–spectral features of different layers had augmented the distinctions for different categories. In addition, in Tables 8–10 and the classification maps in Figures 11–13, we noted that the active pseudo-sample propagation learning utilized in AI-TFNet accuracy and performance, even with limited training samples. This further illustrated the efficiency of the proposed active inference pseudo-sample propagation and CapLoss functions. In Figure 11, we noted that in the red rectangular region, our proposed AI-TFNet indicated smoother classification results than other conventional algorithms. In Figure 12, in the black rectangular region, our proposed AI-TFNet provided more distinguishable details for two related land-cover categories.

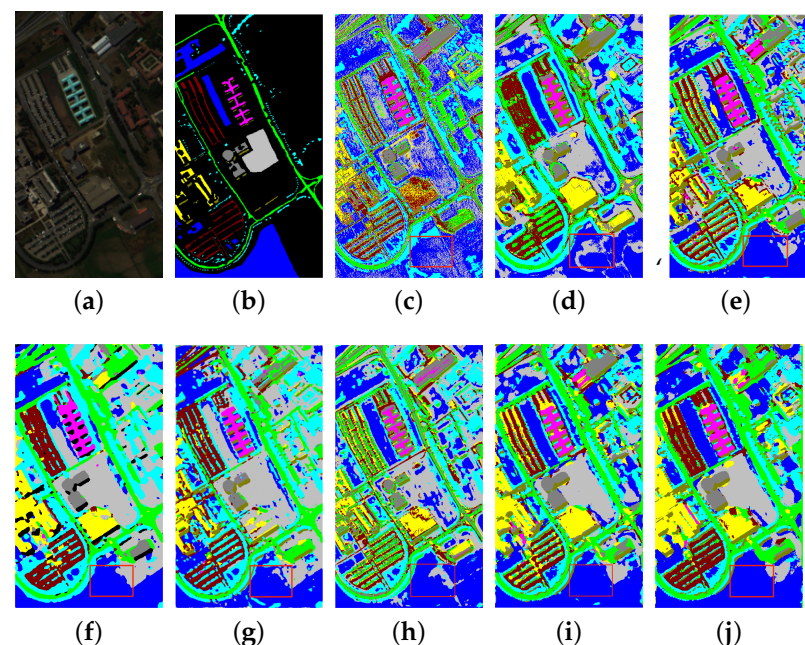


Figure 11. Classification maps for the University of Pavia dataset with 20 labeled training samples per class. (a) False color image; (b) ground truth map; (c) SVM; (d) SVMCK; (e) DcCapsGAN; (f) LMAFN; (g) SSRN; (h) SSFCN; (i) TFNet; (j) AI-TFNet.

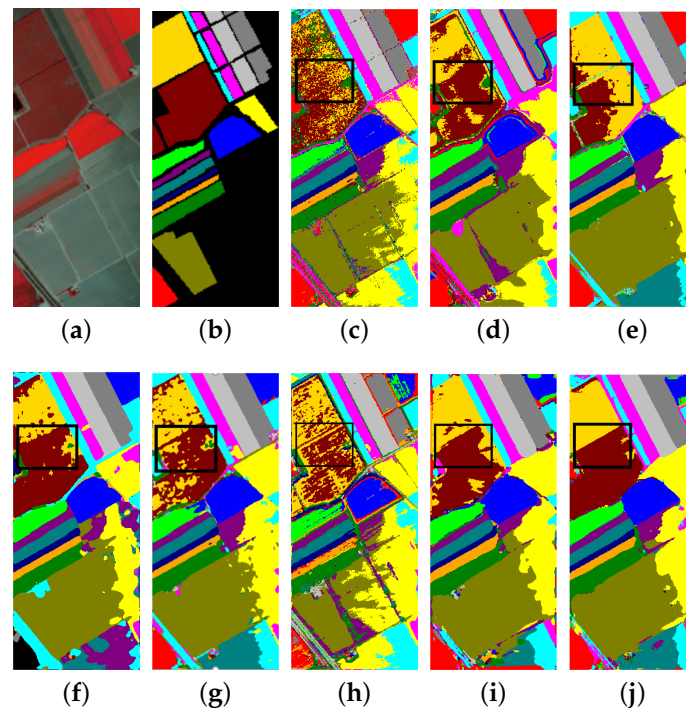


Figure 12. Classification maps for the Salinas dataset with 10 labeled training samples per class. (a) False color image; (b) ground truth map; (c) SVM; (d) SVMCK; (e) DcCapsGAN; (f) LMAFN; (g) SSRN; (h) SSFCN; (i) TFNet; (j) AI-TFNet.

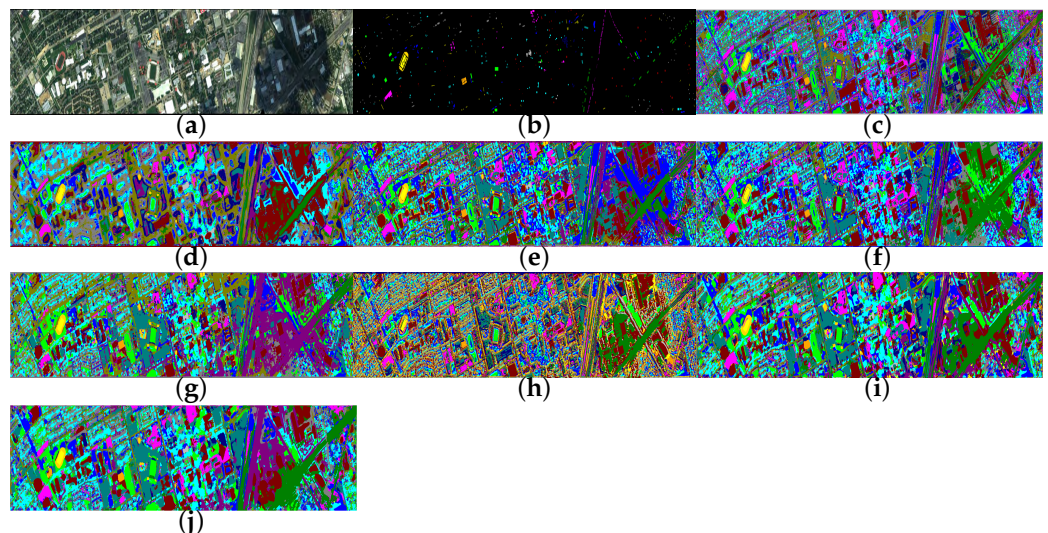


Figure 13. Classification maps for the Houston dataset with 10 labeled training samples per class. (a) False color image; (b) ground truth map; (c) SVM; (d) SVMCK; (e) DcCapsGAN; (f) LMAFN; (g) SSRN; (h) SSFCN; (i) TFNet; (j) AI-TFNet.

Furthermore, to prove the efficiency of the proposed TFNet and AI-TFNet, we also listed the operation time of all the algorithms on each dataset in Tables 8–10. In practical remote-sensing applications, the training process can be executed by an offline model; therefore, only the testing time was reported in Tables 8–10. We observed that the proposed TFNet and AI-TFNet only cost 0.07 seconds and 0.06 seconds, 0.06 seconds and 0.06 seconds, and 0.1 seconds and 0.12 seconds when tested on the three different datasets, respectively. Therefore, the experimental results demonstrated that TFNet and AI-TFNet had better performance and efficiency than other methods. This can be attributed to the end-to-end

structure of TFNet, which was adopted to overcome the challenge introduced in patch-wise-based repetitive computations. Meanwhile, the proposed active inference pseudo-sample propagation strategy with a CapLoss function further mitigated the requirement of a high quantity of labeled training samples for deep-learning-based algorithms.

4. Conclusions

In this paper, we proposed a novel active inference transfer convolutional fusion network (AI-TFNet) to improve the accuracy and efficiency of HSI classification, especially when training samples were limited in quantity. First, the proposed multi-stage hybrid spectral-spatial fully convolutional fusion structure (TFNet) overcame the computational repetition caused by patch-wise-based deep-learning algorithms. In addition, the multi-stage hybrid structure was able to merge low-level spectral-spatial features (detailed information) with high-level spectral-spatial features (contextual information), which not only avoided the redundant path-wise computations but also revealed local and high-level contextual features. In addition, a confidence score and a correct CapLoss function were designed and utilized to augment the training sample sets for active inferential pseudo-labeled samples and supported the backpropagation in the training stage, even with small sample sets. The experimental results on three HSI datasets further demonstrated that the proposed TFNet and AI-TFNet had better outcomes in accuracy, efficiency, and classification performance, regardless of sample size.

Although the proposed TFNet and AI-TFNet had robust results for classification accuracy, expanding their application with more adaptive, automatic training samples via online inference and contextual analysis is a challenging direction to be addressed in future research.

Author Contributions: Conceptualization, J.W.; formal analysis, L.L. and J.H.; funding acquisition, J.W.; methodology, J.W.; project administration, Y.L. and J.H.; software, L.L.; visualization, Y.L.; writing—original draft, J.W. and L.L.; writing—review & editing, J.W., X.X., and B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported, in part, by the National Natural Science Foundation of China, under grant numbers 61801353, 61977052, and 61971273; in part, by GHfund B, under grant numbers 202107020822 and 202202022633; and in part, by the project supported by the China Postdoctoral Science Foundation funded project, under grant number 2018M633474.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the editors and anonymous reviewers for their valuable comments and suggestions, which have greatly improved the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

References

1. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Deep learning for hyperspectral image classification: An overview. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6690–6709. [\[CrossRef\]](#)
2. Ghamisi, P.; Mura, M.D.; Benediktsson, J.A. A survey on spectral-spatial classification techniques based on attribute profiles. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 2335–2353. [\[CrossRef\]](#)
3. Uzket, B.; Rangnekar, A.; Hoffman, M. Aerial vehicle tracking by adaptive fusion of hyperspectral likelihood maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 39–48.
4. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [\[CrossRef\]](#)
5. Lacar, F.M.; Lewis, M.M.; Grierson, I.T. Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia. In Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217), Sydney, Australia, 9–13 July 2001; pp. 2875–2877.
6. Tan, K.; Zhang, J.; Du, Q.; Wang, X. GPU parallel implementation of support vector machines for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4647–4656. [\[CrossRef\]](#)

7. Gao, L.; Li, J.; Khodadadzadeh, M.; Plaza, A.; Zhang, B.; He, Z.; Yan, H. Subspace-based support vector machines for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2014**, *12*, 349–353.
8. Liu, L.; Huang, W.; Liu, B.; Shen, L.; Wang, C. Semisupervised hyperspectral image classification via Laplacian least squares support vector machine in sum space and random sampling. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *8*, 4086–4100. [\[CrossRef\]](#)
9. Tu, B.; Huang, S.; Fang, L.; Zhang, G.; Wang, J.; Zheng, B. Hyperspectral image classification via weighted joint nearest neighbor and sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4063–4075. [\[CrossRef\]](#)
10. Blanzieri, E.; Melgani, F. Nearest neighbor classification of remote sensing images with the maximal margin principle. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1804–1811. [\[CrossRef\]](#)
11. Ham, J.; Chen, Y.; Crawford, M.M.; Ghosh, J. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 492–501. [\[CrossRef\]](#)
12. Zhang, Y.; Cao, G.; Li, X.; Wang, B. Cascaded random forest for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1082–1094. [\[CrossRef\]](#)
13. Peerbhay, K.Y.; Mutanga, O.; Ismail, R. Random forests unsupervised classification: The detection and mapping of solanum mauritanium infestations in plantation forestry using hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3107–3122. [\[CrossRef\]](#)
14. Camps-Valls, G.; Gomez-Chova, L.; Munoz-Mari, J.; Vila-Frances, J.; Calpe-Maravilla, J. Composite kernels for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 93–97. [\[CrossRef\]](#)
15. He, Z.; Liu, L.; Zhu, Y.; Zhou, S. Anisotropically foveated nonlocal weights for joint sparse representation-based hyperspectral classification. In Proceedings of the 2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Tokyo, Japan, 2–5 June 2015; pp. 1–4.
16. Wang, J.; Jiao, L.; Liu, H.; Yang, S.; Liu, F. Hyperspectral image classification by spatial-spectral derivative-aided kernel joint sparse representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *8*, 2485–2500. [\[CrossRef\]](#)
17. Wang, J.; Jiao, L.; Wang, S.; Hou, B.; Liu, F. Adaptive nonlocal spatial-spectral kernel for hyperspectral imagery classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4086–4101. [\[CrossRef\]](#)
18. Wang, J.; Zhang, G.; Cao, M.; Jiang, N. Semi-supervised classification of hyperspectral image based on spectral and extended morphological profiles. In Proceedings of the 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA, 21–24 August 2016; pp. 1–4.
19. Kang, X.; Xiang, X.; Li, S.; Benediktsson, J.A. PCA-based edge-preserving features for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 7140–7151. [\[CrossRef\]](#)
20. Shen, L.; Jia, S. Three-dimensional Gabor wavelets for pixel-based hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 5039–5046. [\[CrossRef\]](#)
21. Jia, S.; Xie, Y.; Shen, L.; Deng, L. Hyperspectral image classification using Fisher criterion-based Gabor cube selection and multi-task joint sparse representation. In Proceedings of the 2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Tokyo, Japan, 2–5 June 2015; pp. 1–4.
22. Ye, Z.; Bai, L.; Tan, L. Hyperspectral image classification based on gabor features and decision fusion. In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017; pp. 478–482.
23. Li, W.; Du, Q. Gabor-filtering-based nearest regularized subspace for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1012–1022. [\[CrossRef\]](#)
24. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [\[CrossRef\]](#)
25. Song, W.; Li, S.; Fang, L.; Lu, T. Hyperspectral image classification with deep feature fusion network *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3173–3184. [\[CrossRef\]](#)
26. Zhong, Z.; Li, J.; Luo, Z.; Chapman, M. Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 847–858. [\[CrossRef\]](#)
27. Roy, S.K.; Krishna, G.; Dubey, S.R.; Chaudhuri, B.B. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 277–281. [\[CrossRef\]](#)
28. Paoletti, M.E.; Haut, J.M.; Roy, S.K.; Hendrix, E.M. Rotation equivariant convolutional neural networks for hyperspectral image classification. *IEEE Access* **2020**, *8*, 179575–179591. [\[CrossRef\]](#)
29. Yao, W.; Lian, C.; Bruzzone, L. ClusterCNN: Clustering-based feature learning for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1991–1995. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#)
31. Wang, J.; Huang, R.; Guo, S.; Li, L.; Zhu, M.; Yang, S.; Jiao, L. NAS-guided lightweight multi-scale attention fusion network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 8754–8767. [\[CrossRef\]](#)
32. Sun, L.; Fang, Y.; Chen, Y.; Huang, W.; Wu, Z.; Jeon, B. Multi-Structure KELM With Attention Fusion Strategy for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [\[CrossRef\]](#)
33. Yue, J.; Fang, L.; He, M. Spectral–Spatial Latent Reconstruction for Open-Set Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2022**, *31*, 5227–5241. [\[CrossRef\]](#)

34. Wan, S.; Gong, C.; Zhong, P.; Du, B.; Zhang, L.; Yang, J. Multiscale dynamic graph convolutional network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *58*, 3162–3177. [\[CrossRef\]](#)
35. Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Plaza, A.; Chanussot, J. Graph convolutional networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5966–5978. [\[CrossRef\]](#)
36. Zheng, Z.; Zhong, Y. S3NET: Towards real-time hyperspectral imagery classification. In Proceedings of the IEEE 2019 International Geoscience and Remote Sensing Symposium (IGARSS), Yokohama, Japan, 28 July–02 August 2019; pp. 3293–3296.
37. Tong, X.; Yin, J.; Han, B.; Qv, H. Few-shot learning with attention-weighted graph convolutional networks for hyperspectral image classification. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 1686–1690.
38. Mou, L.; Lu, X.; Li, X.; Zhu, X.X. Nonlocal graph convolutional networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8246–8257. [\[CrossRef\]](#)
39. Zou, L.; Zhu, X.; Wu, C.; Liu, Y.; Qu, L. Spectral-spatial exploration for hyperspectral image classification via the fusion of fully convolutional networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 659–674. [\[CrossRef\]](#)
40. Li, J.; Zhao, X.; Li, Y.; Du, Q.; Xi, B.; Hu, J. Classification of hyperspectral imagery using a new fully convolutional neural network. *IEEE Geosci. Remote. Sens. Lett.* **2018**, *15*, 292–296. [\[CrossRef\]](#)
41. Zhu, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Generative adversarial networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5046–5063. [\[CrossRef\]](#)
42. 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. [\[CrossRef\]](#)
43. Chi, M.; Bruzzone, L. Classification of hyperspectral data by continuation semi-supervised SVM. In Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium, Barcelona, Spain, 23–28 July 2007; pp. 3794–3797.
44. Bruzzone, L.; Chi, M.; Marconcini, M. Transductive SVMs for semi-supervised classification of hyperspectral data. In Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium, Seoul, Republic of Korea, 29 July 2005; p. 4.
45. Qin, A.; Shang, Z.; Tian, J.; Wang, Y.; Zhang, T.; Tang, Y.Y. Spectral-spatial graph convolutional networks for semi-supervised hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 241–245. [\[CrossRef\]](#)
46. Zhan, Y.; Medjadba, Y.; Wang, G.; Yu, X.; Qin, J.; Huang, T.; Wu, K.; Hu, D.; Zhao, Z.; Wang, Y.; et al. Hyperspectral image classification based on generative adversarial networks with feature fusing and dynamic neighborhood voting mechanism. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 811–814.
47. Zhan, Y.; Hu, D.; Wang, Y.; Yu, X. Semisupervised hyperspectral image classification based on generative adversarial networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *15*, 212–216. [\[CrossRef\]](#)
48. Wang, J.; Guo, S.; Huang, R.; Li, L.; Zhang, X.; Jiao, L. Dual-channel capsule generation adversarial network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–16. [\[CrossRef\]](#)
49. Xu, Y.; Du, B.; Zhang, L. Beyond the patch-wise classification: Spectral-spatial fully convolutional networks for hyperspectral image classification. *IEEE Trans. Big Data.* **2019**, *6*, 492–506. [\[CrossRef\]](#)

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