



Article

Attention-Guided Fusion and Classification for Hyperspectral and LiDAR Data

Jing Huang ¹, Yinghao Zhang ¹, Fang Yang ^{1,*} and Li Chai ²

¹ Engineering Research Center of Metallurgical Automation and Measurement Technology, Wuhan University of Science and Technology, Wuhan 430081, China; huangjing21@wust.edu.cn (J.H.); zyh981123@gmail.com (Y.Z.)

² State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China; chaili@zju.edu.cn

* Correspondence: yangfang.idif@wust.edu.cn

Abstract: The joint use of hyperspectral image (HSI) and Light Detection And Ranging (LiDAR) data has been widely applied for land cover classification because it can comprehensively represent the urban structures and land material properties. However, existing methods fail to combine the different image information effectively, which limits the semantic relevance of different data sources. To solve this problem, in this paper, an Attention-guided Fusion and Classification framework based on Convolutional Neural Network (AFC-CNN) is proposed to classify the land cover based on the joint use of HSI and LiDAR data. In the feature extraction module, AFC-CNN employs the three dimensional convolutional neural network (3D-CNN) combined with a multi-scale structure to extract the spatial-spectral features of HSI, and uses a 2D-CNN to extract the spatial features from LiDAR data. Simultaneously, the spectral attention mechanism is adopted to assign weights to the spectral channels, and the cross attention mechanism is introduced to impart significant spatial weights from LiDAR to HSI, which enhance the interaction between HSI and LiDAR data and leverage the fusion information. Then two feature branches are concatenated and transferred to the feature fusion module for higher-level feature extraction and fusion. In the fusion module, AFC-CNN adopts the depth separable convolution connected through the residual structures to obtain the advanced features, which can help reduce computational complexity and improve the fitting ability of the model. Finally, the fused features are sent into the linear classification module for final classification. Experimental results on three datasets, i.e., Houston, MUUFL and Trento datasets show that the proposed AFC-CNN framework achieves better classification accuracy compared with the state-of-the-art algorithms. The overall accuracy of AFC-CNN on Houston, MUUFL and Trento datasets are 94.2% , 95.3% and 99.5%, respectively.

Keywords: hyperspectral images; Light Detection And Ranging (LiDAR) data; fusion and classification; convolutional neural network; attention mechanism



Citation: Huang, J.; Zhang, Y.; Yang, F.; Chai, L. Attention-Guided Fusion and Classification for Hyperspectral and LiDAR Data. *Remote Sens.* **2024**, *16*, 94. <https://doi.org/10.3390/rs16010094>

Academic Editor: Kevin Tansey

Received: 27 November 2023

Revised: 22 December 2023

Accepted: 22 December 2023

Published: 25 December 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

Land cover classification of remote sensing images is a fundamental task for earth observation, which allows the accurate identification of materials on the surface. With the rapid advancement of geospatial information technology, remote sensing images are characterized with heterogeneous and multi-source data, which can provide complementary information for land-cover observation [1–3]. Among various multi-source data, hyperspectral images (HSIs) contain rich spectral information of the land objects because of their broad spectral dimension. However, the spatial resolution of HSIs is typically low, making them less effective in distinguishing materials with similar spectral responses [4,5]. Different from HSIs, Light Detection And Ranging (LiDAR) data offers detailed elevation information and is less affected by atmospheric interference and environmental changes. Nevertheless, LiDAR data can not distinguish between materials of the same height [6].

Given the different properties of HSI and LiDAR data, the joint use of HSI and LiDAR data can be more effective in differentiating objects and urban structures compared to using single-source data. This integration of HSI and LiDAR has found widespread application in the field of land-cover classification for remote sensing images [7–10].

Due to the heterogeneous properties of HSI and LiDAR data, extracting sufficient joint features and establishing complementary connections for accurate classification has become a considerable challenge [11–13]. In the last decades, many attempts have been made to improve the classification performance based on the fusion of HSI and LiDAR data. To fully explore the information of HSI and LiDAR data, various strategies for feature extraction have been adopted. A simple way is to stack elevation and intensity features of LiDAR as additional channels to the spectral bands of HSI, thus forming an extended feature vector for classification [14,15]. For example, in [16], Puttonena et al. fused LiDAR-derived and hyperspectral features and employed support vector machine (SVM) as a classifier for tree species classification. In [17], Pedergnana et al. employed two extended attribute profiles (EAPs) to extract features for multispectral and LiDAR images, respectively, realizing a fusion of the spectral, spatial and elevation information. Then two EAPs were concatenated into one stacked vector and classified by random forest and SVM classifiers. In [18], Ghamisi et al. utilized the extinction filter to extract spatial and contextual information of both hyperspectral and LiDAR features, and then adopted a random forest classifier to handle high feature redundancy. These approaches have demonstrated better classification performance compared to the methods using single-source datasets. However, the simple stacked-feature approach is not powerful enough to interpret the inherent features of image data since it may contain redundant information [8]. Additionally, considering the classification techniques, traditional SVM-based and decision tree-based classification techniques cannot handle complex multi-classification of terrain materials effectively. This is because the SVM is not good at weighting heterogeneous features since it regards all features equally and the decision tree is time-consuming [19]. These results limit the applicability of classification performance for HSI and LiDAR data.

Another feature extraction strategy is adopting a dual-branch architecture with a deep learning-based framework. Deep learning-based methods have shown superior performance in remote sensing image fusion and classification tasks due to the powerful feature learning ability of the convolutional neural networks (CNNs) [20–22]. For the dual-branch architecture, the features of HSI and LiDAR are first extracted in parallel of each branch, then these features are concatenated and used as a fusion representation for image data. Finally, the fused features are transferred to a classifier for further classification. For example, Xu et al. [20] proposed a two-branch deep CNN framework for multisource remote sensing data classification and achieved encouraging classification results. The network adopted 2D and 1D convolution operators to extract features from surrounding neighbors and to enhance the spectral information. Then a deep network including cascade blocks is designed to extract features from data. Huang et al. [23] used a two-branch CNN to extract both spatial and spectral features of high spatial resolution multispectral images for land-use classification. Feng et al. [24] utilized a residual block in HSI branch and a LiDAR branch to extract hierarchical, parallel, and multiscale features. Then an adaptive-feature fusion module based on the squeeze-and-excitation network was used to integrate the HSI and LiDAR features. However, the HSI feature extraction branch of these methods often use 2D-CNNs, which can only capture spatial features of HSI and ignore the spectral information from the original data, leading to poor feature fusion quality. Although 3D-CNN was also employed to extract spectral features from HSI [25–27], the classification results are still unsatisfactory. This is because the traditional 3D-CNNs can only extract features within a fixed-size range around the same pixel, limiting its feature extraction capability and hindering the classification performance [28]. Moreover, the dual-branch CNNs only realize fusion by concatenating features and do not consider the semantic relation between multisource data.

To tackle this issue, multi-branch networks combined with attention mechanisms for HSI-LiDAR data fusion have been developed to reduce the information loss in the fusion process. For example, in [29], Mohla et al. introduced an attention-based multimodal fusion network for land-cover classification to improve the classification performance, named as FusAtNet. The framework used a self-attention mechanism to extract spectral features in HSI and adopted a cross attention mechanism in LiDAR data to derive an attention mask, which can enhance the spatial features of HSI modality. In [30], the authors designed a dual-channel A^3 CLNN network which contains modules of spatial attention, spectral attention, multiscale residual attention and a three-level fusion strategy. In [31], Li et al. proposed a triplet semi-supervised deep network (TSDN) for fusion classification of HSI and LiDAR, where a 3D cross attention block is designed for extracting spatial complementarity of HSI and LiDAR. In [10], Wang et al. proposed a novel multi-attentive hierarchical fusion net (MAHiDFNet) to realize the feature-level fusion and classification. In this paper, a triple branch HSI-LiDAR CNN backbone was first developed to extract the spatial features, spectral features and elevation features of the land-cover objects, and then a modality attention module was designed for feature interaction and integration. The above fusion classification algorithms demonstrate that the spatial or spectral attention models can improve feature representation capability and classification performance. However, the above frameworks often require massive parameters and a high computational burden.

To overcome the limitations of existing algorithms, in this paper, we propose a novel Attention-guided Fusion and Classification framework based on Convolutional Neural Network (AFC-CNN) for the land cover classification via the joint use of HSI and LiDAR data. Unlike previous 2D-CNN and traditional 3D-CNN approaches, our proposed AFC-CNN adopts an innovative 3D-CNN architecture that incorporates a multi-scale structure to extract multiscale spatial-spectral features from HSI. Moreover, the spectral attention mechanism and cross attention mechanism are introduced to enhance the inherent correlation between HSI and LiDAR. In the feature fusion extraction module, we design six layers of depth separable convolutional layers and connect them with residual structures to minimize data loss during backward propagation. The depth separable convolution can significantly reduce computational complexity compared with traditional 2D-CNNs, making it more efficient while maintaining reasonable performance.

Collectively, the main contributions of this paper can be summarized as follows:

- We design a dual-branch CNN fusion classification network named as AFC-CNN, which consists of a 2D-CNN for spatial features extraction in LiDAR data, and a novel 3D-CNN incorporated with a multi-scale structure for spatial-spectral features in HSI.
- AFC-CNN utilizes the spectral attention mechanism to strengthen the important features from spectral channels of HSI. Additionally, a cross attention mechanism module is introduced to enhance the inherent correlation between HSI and LiDAR.
- In the feature fusion module, AFC-CNN employs the depth separable convolutions connected through residual structure to extract the advanced features of the fusion information. Compared with the traditional 2D-CNN, the use of depth separable convolutions reduces computational complexity while maintaining high performance.
- Experimental results demonstrate that, the proposed algorithm AFC-CNN is more effective than the state-of-the-art methods in terms of the evaluation metrics and visual effects.

The remainder of the paper is organized as follows. Section 2 describes the details of the proposed classification framework. Section 3 shows the experimental results and analyzes the performance of the proposed method. Section 4 concludes this paper.

2. Proposed Fusion and Classification Framework

The framework of the proposed AFC-CNN algorithm is shown in Figure 1, which includes five main modules: HSI feature extraction module, LiDAR feature extraction module, attention mechanism module, feature fusion module, and linear classification module. In this section, we will introduce each module in detail.

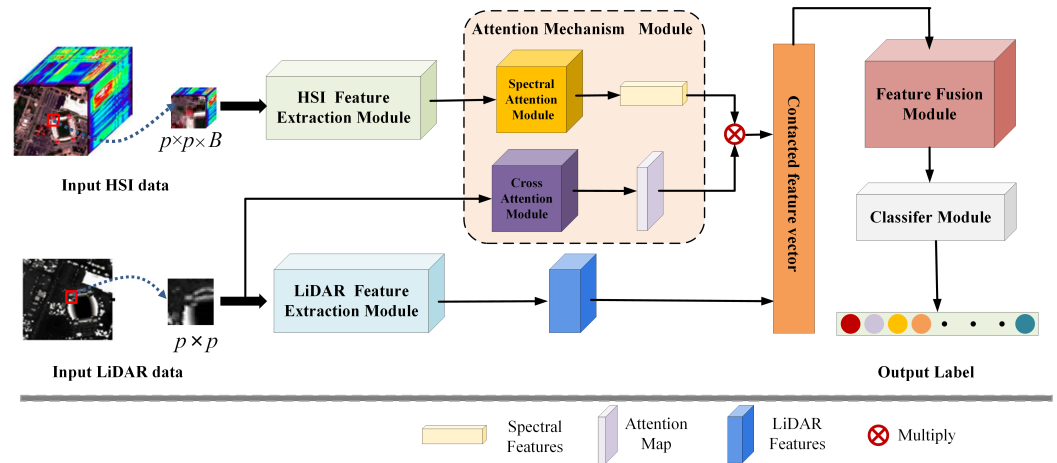


Figure 1. The framework of AFC-CNN.

2.1. HSI Feature Extraction Module

As previously stated, traditional 3D-CNN-based multimodal data fusion classification models can extract the spatial-spectral feature information of HSI. However, the conventional 3D-CNN can not extract enough global information of the image since it only extracts information within a fixed-size range. In this paper, we employ a multi-scale structure to improve the feature extraction capability of the 3D-CNN, which can help extract higher level semantic features.

Given an HSI $X_H \in \mathbb{R}^{S_h \times S_w \times B}$, where S_h and S_w represent the height and width of the image, respectively, and B is the number of spectral bands of the HSI. For a certain point (x, y) of the HSI, we first crop a small cube centered around the point, denoted as $x_H \in \mathbb{R}^{p \times p \times B}$. The cube size p is empirically set as 15 in this paper. After that, x_H are fed into a 3D convolutional layer to extract spatial-spectral features. Then, the extracted features are connected with two consecutive multi-scale feature extraction structures for more comprehensive feature extraction. As shown in Figure 2, the multi-scale structure includes three convolutional kernel sizes: $1 \times 1 \times 1$, $3 \times 3 \times 3$, and $5 \times 5 \times 5$, which can sufficiently capture higher level spatial-spectral features of pixels. For each convolutional layer, a Batch Normalization (BN) layer is incorporated to regularize and accelerate the training process, and a Rectified Linear Unit (ReLU) is followed to learn a nonlinear representation. Finally, the extracted spatial-spectral features are transferred into the spectral attention mechanism module.

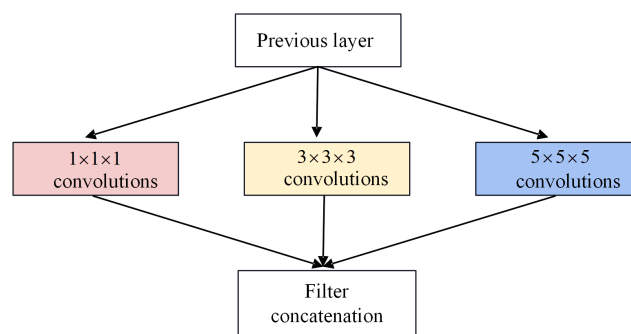


Figure 2. The framework of multi-scale structure.

2.2. LiDAR Feature Extraction Module

Given a LiDAR data $X_L \in \mathbb{R}^{S_h \times S_w}$, similarly, we crop X_L into image block $x_L \in \mathbb{R}^{p \times p}$ and transfer it to the LiDAR feature extraction module. In the LiDAR feature extraction module, three 2D-CNNs with a kernel size of 3×3 are employed for feature extraction. The extracted LiDAR features are then concatenated with an HSI feature extraction branch for

the feature fusion module. The overall parameter configurations of the feature extraction modules of HSI and LiDAR data are shown in Figure 3.

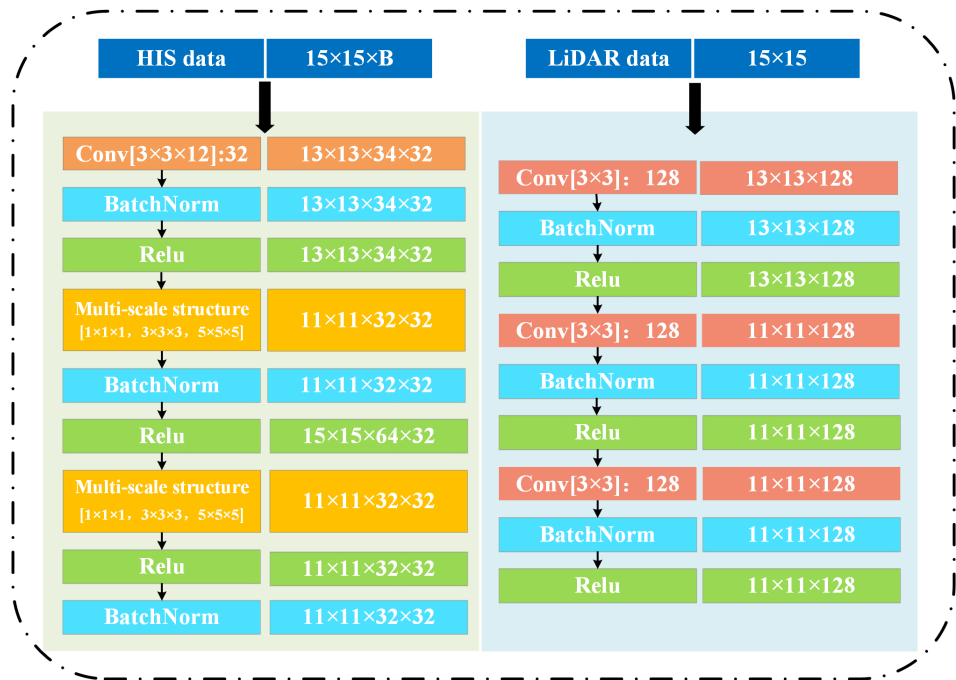


Figure 3. Overall parameter configurations of feature extraction modules of HSI and LiDAR data in the designed AFC-CNN network.

2.3. Attention Mechanism Module

In previous multimodal data fusion classification networks, feature extraction was performed independently on each data branch. These networks did not facilitate the interactions to share high-level features between different modalities, leading to a low quality of feature fusion and classification performance. Therefore, in this paper, we first introduce the spectral attention mechanism module to allocate weights of spectral channels in HSI. Then the cross attention mechanism is incorporated to enhance the correlation between HSI and LiDAR data, which extracts attention masks from LiDAR data and subsequently employs them to enhance the representation of HSI spatial features. The whole framework of the attention mechanism module is depicted in Figure 4.

In the spectral attention mechanism module, the input feature $F \in \mathbb{R}^{H \times W \times D}$ is the spatial-spectral feature extracted from the HSI feature extraction module. We first perform the global maximum pooling and the global average pooling operations to generate two spatial context features: F_{avg}^c and F_{max}^c . Then the two feature vectors are transferred into a shared multi-layer perceptron (MLP) layer, which conducts deep-level feature extraction through a two-layer CNN. The outputs are aggregated and directed into a sigmoid classification unit to produce the spectral attention features $M_s \in \mathbb{R}^{1 \times 1 \times D}$. The calculation process of this module can be represented as follows [32]:

$$\begin{aligned}
 M_s(F) &= \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \\
 &= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c)))
 \end{aligned} \quad (1)$$

where F is the input spatial-spectral feature and σ denotes the sigmoid function: $\sigma(x) = \frac{1}{1+\exp(-x)}$, F_{avg}^c and F_{max}^c denote the data after global average pooling and global maximum pooling, $W_0 \in \mathbb{R}^{D/r \times D}$ and $W_1 \in \mathbb{R}^{D \times D/r}$ represent the weights of the two convolutional layers in the shared MLP layer, where r is the reduction ratio. $M_s(\cdot)$ represents spectral attention results.

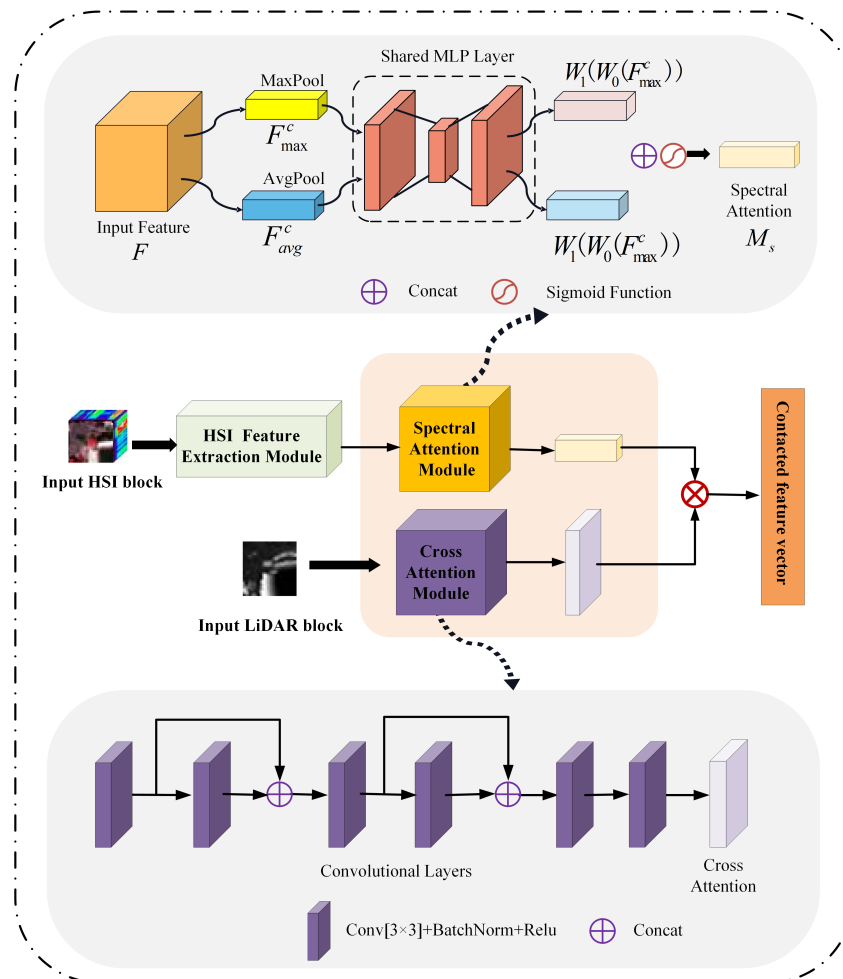


Figure 4. The framework of the attention mechanism module.

In the cross attention mechanism module, the input data is LiDAR image block x_L . As illustrated in Figure 4, this module consists of six 2D convolutional layers, of which the convolutional layers are sequentially followed by a BN layer and a ReLU layer. These layers adopt a dense connection strategy for backward propagation. The input LiDAR block is transformed into weighted attention map and then multiplied with weighted HSI spectral channels. Thus, the cross attention mechanism module can reinforce the interaction ability of HSI and LiDAR and share the high-level features between them.

2.4. Feature Fusion Module

The feature fusion module can extract higher-level features from the fused data, and enhance the correlation between different modalities. In this module, we replace the traditional 2D-CNN with depthwise separable convolutions to reduce the computational complexity and improve the nonlinear fitting ability of the model. Moreover, we design six convolutional layers in the feature fusion module with residual connections between them, which can reduce data loss via backward propagation and improve the quality of data fusion. The framework of the feature fusion module is illustrated in Figure 5.

The sketch maps of standard convolution and depthwise separable convolution are shown in Figure 6. In Figure 6b, we can observe that depthwise separable convolutions consist of two key components: depthwise (DW) convolution and pointwise (PW) convolution. The DW convolution is employed for each channel of the input feature map, with each channel only convolved by one specific convolutional kernel. Subsequently, the outputs of DW convolutional kernels go through the PW convolution to obtain the final output, which

allows for flexibility in adjusting the number of output channels and facilitates the fusion of channel results.

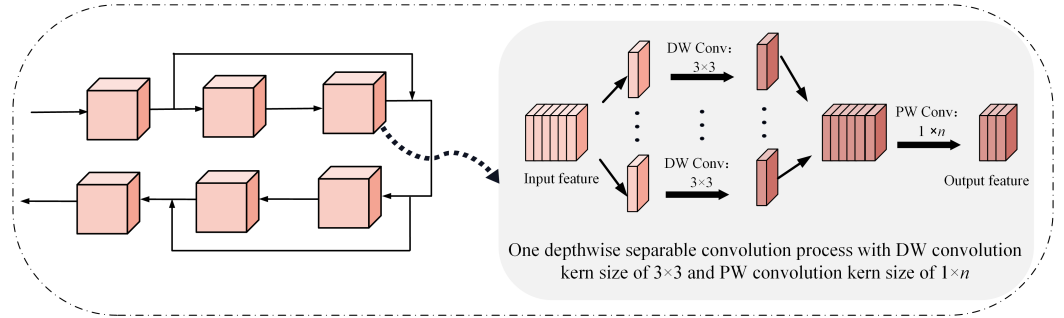


Figure 5. The framework of the feature fusion module.

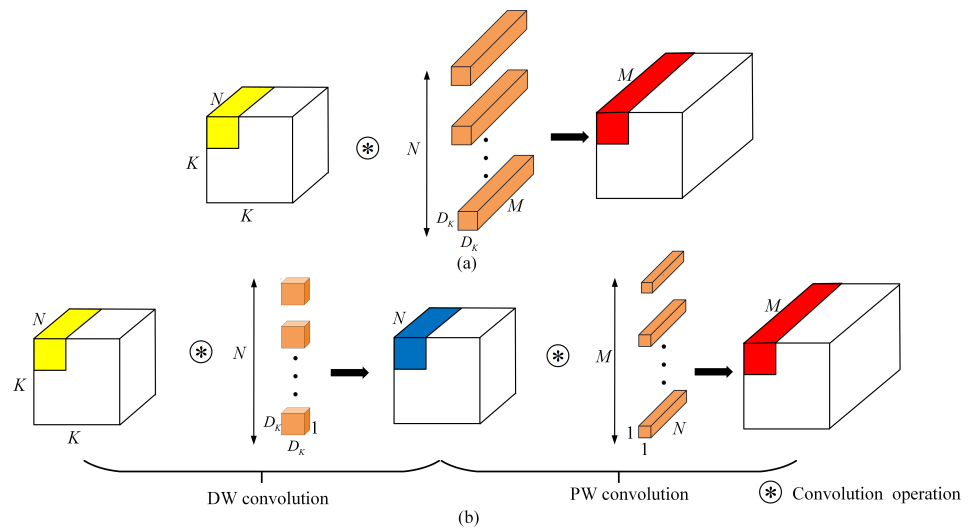


Figure 6. The sketch map of standard convolutions and depthwise separable convolutions. (a) Standard convolution; (b) depthwise separable convolution.

Assuming that the size of input feature is $K \times K \times N$, the size of the convolution kernel is $D_K \times D_K \times N$, the number of kernels is M and the stride is 1. So the number of the floating point operations (FLOPs) of standard convolution is:

$$K \times K \times D_K \times D_K \times N \times M. \quad (2)$$

For depthwise separable convolution, the number of FLOPs is:

$$K \times K \times D_K \times D_K \times N + K \times K \times N \times M. \quad (3)$$

The ratio of computational complexity between standard convolution and depthwise separable convolution is:

$$\frac{K \times K \times D_K \times D_K \times N + K \times K \times N \times M}{K \times K \times D_K \times D_K \times N \times M} = \frac{1}{M} + \frac{1}{D_K^2}. \quad (4)$$

Thus, the computational complexity of the depthwise separable convolution is significantly lower compared to the standard convolution.

2.5. Linear Classification Module

After the feature fusion module, the extracted high-level features are vectorized to a one-dimensional vector and transferred to the linear classification module for the final

classification. The framework of the linear classification module and the corresponding parameters are shown in Figure 7. It can be observed that, the linear classification module comprises three fully connected layers, which establishes the association between the input data and the corresponding labels through feature mapping. To mitigate overfitting, a dropout layer is employed with a dropout rate set to 0.5 in this module.

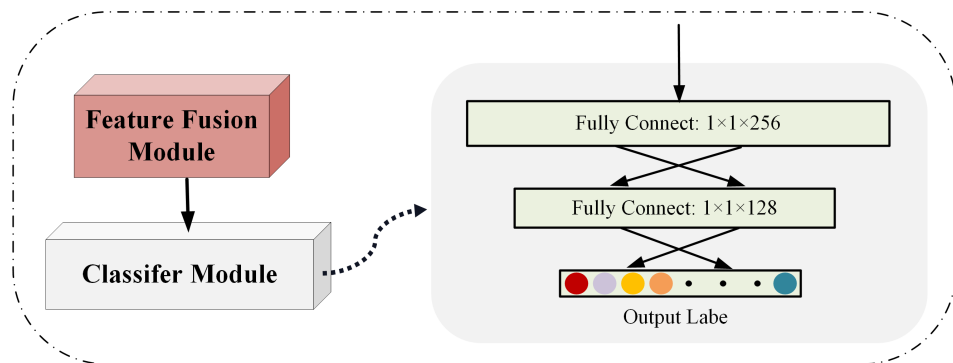


Figure 7. The framework of the linear classification module.

3. Experimental Results and Analysis

In this section, we conduct experiments on a variety of datasets to test the performance of the proposed AFC-CNN algorithm for HSI and LiDAR data fusion and classification. We first employ an ablation experiment to verify the effectiveness of the proposed AFC-CNN framework. Then, we compare the AFC-CNN algorithm with the traditional SVM method, an advanced CNN model FusAtNet [29] along with their corresponding model based on HSI alone. The experimental datasets, parameters and results are presented in detail.

3.1. Datasets

In the experiment, three public datasets serve as benchmarks for evaluating the performance: Houston2013, MUUFL, and Trento. The detailed information of these datasets are listed in Table 1.

Table 1. Description of datasets used in the experiments.

Dataset	Pixel No.	Sample No.	Class No.	Sensor Type	Wavelength	Spatial Resolution	Band No.
Houston2013	664845	15209	15	HSI	0.38 μm –1.05 μm	2.5 m	144
				LiDAR	/	2.5 m	1
MUUFL	71500	53687	12	HSI	0.38 μm –1.05 μm	0.54 m \times 1 m	64
				LiDAR	1.06	0.6 m \times 0.78 m	2
Trento	99600	30214	6	HSI	0.42 μm –0.99 μm	1 m	63
				LiDAR	/	1 m	1

(1) Houston Dataset: this dataset was acquired by ITRES CASI-1500 sensor over the University of Houston campus and its neighboring urban area. The Houston2013 dataset was provided to participants in the 2013 IEEE GRSS Data Fusion Contest [33]. It consists of HSI and LiDAR data, both of which have a spatial size of 349×1905 (total 664,845) pixels with a spatial resolution of 2.5 m. The HSI data contains a total of 144 spectral bands. In total, 15,029 pixels of this dataset have been labeled as samples, including 15 land cover classes, such as grassland, artificial turf, residential areas, commercial areas, etc. The dataset presents significant challenges for classification tasks due to the large number of unlabeled data points and the scattered distribution of labeled sample points throughout the spatial domain.

(2) MUUFL Dataset: this dataset was captured by an airborne hyperspectral sensor in the Gulfport National Park, located in the southern Mississippi Gulf Coast [34]. The

spatial size of HSI and LiDAR data is 325×220 (total 71,500) pixels. The spatial resolution of the HSI data and the LiDAR data are $0.54 \text{ m} \times 1 \text{ m}$, and $0.6 \text{ m} \times 0.78 \text{ m}$, respectively. The initial HSI data consists of 72 spectral bands, but the first four and last four bands are with noise contamination. So, in the experiment, the remaining 64 spectral bands of HSI data are considered for the classification. The dataset consists of 12 land cover classes, including trees, roads, sidewalks, etc.

(3) Trento Dataset: this dataset was collected over a rural area in the south of city of Trento, Italy [35]. The LiDAR data was acquired by the Optech ALTM 3100EA sensor, and the HSI data was acquired by the AISA Eagle sensor. Both of them have the spatial size of 166×600 (total 99,600) pixels with a spatial resolution of 1m. The HSI data consists of 63 bands covering the range from $0.42 \mu\text{m}$ to $0.99 \mu\text{m}$. The dataset has 30,214 pixels that are labeled as samples, including six land cover classes, such as apple trees, buildings, roads, etc.

3.2. Evaluation Criteria

We analyze the classification performance of the proposed method and the comparison algorithms according to four metrics: the overall accuracy (OA), the average accuracy (AA), the Kappa coefficient (Kappa) and the per-class accuracy [36]. OA shows a global perspective on accuracy, and it is derived by the ratio of correctly classified pixels to the total number of pixels as follows:

$$OA = \frac{\sum_{i=1}^C x_{i,i}}{N_s}, \quad (5)$$

where C is the number of samples classified correctly, N_s is the number of total samples, and $x_{i,i}$ denotes diagonal elements of the confusion matrix, which are the correctly predicted samples. AA denotes the average accuracy across all classes, and it can be calculated as:

$$AA = \frac{1}{C} \sum_{i=1}^C \frac{x_{i,i}}{N_i}, \quad (6)$$

where N_i represents the total number of each class samples. Kappa measures the percentage of agreement beyond chance, and ranges from 0 for agreement consistent with chance to 1 for perfect agreement with ground truth. The Kappa is computed as:

$$Kappa = \frac{N_s \sum_{i=1}^C x_{i,i} - \sum_{i=1}^C (x_{i,+} \times x_{+,i})}{N_s^2 - \sum_{i=1}^C (x_{+,i} \times x_{i,+})}, \quad (7)$$

where $x_{i,+}$ denotes the number of ground truth samples for each class, and $x_{+,i}$ means the number of predicted samples for each class. The per-class accuracy assesses the performance on individual classes. The higher OA, AA and Kappa value indicates a better classification performance.

3.3. Experimental Setting

In the experiment, we select 10% of the data from the dataset as the training set, and the remaining data as the test set to verify the effectiveness of the proposed AFC-CNN. For FusAtNet, we also select 10% of the samples for training and the remaining samples for testing. For SVM, we choose 30% of the data as the training set empirically. This is because SVM is a traditional machine learning algorithm that requires a large number of training samples to achieve a good classification accuracy.

In the training process, a different learning rate controls the convergence rate of the objective function and the network performance. The learning process is implemented by using the Adam optimizer for 100 epochs, and the initial learning rate is set as 0.005. The parameters of the comparison methods are default. The detailed parameter settings of each method are summarized in Table 2.

Table 2. Detailed parameters setting of each methods.

Method	Parameters Setting
AFC-CNN	Training set: 10% of the datasets, Test set: 90% of the datasets, Patch size: 15×15 , learning rate: 0.005, training epochs: 100
SVM	Training set: 30% of the datasets, Test set: 70% of the datasets
FusAtNet	Training set: 10% of the datasets, Test set: 90% of the datasets, Patch size: 11×11 , learning rate: 0.000005, training epochs: 1000

The experimental platform is based on Python 3.8 with the TensorFlow-GPU framework. In the training process, the Adam [37] optimizer is utilized to optimize the model training. All experiments are implemented on an Intel(R) Core(TM) i9-11900 @2.50 GHz and an NVIDIA GeForce RTX 3090 with 128 GB memory.

3.4. Ablation Experiment

In this section, we conduct the ablation experiment to verify the effect of the multi-scale structure in the HSI feature extraction module, the cross attention mechanism module, the depthwise separable convolution and residual structure in the feature fusion module. The experimental comparison results are shown in Table 3.

For multi-scale structure in the HSI feature extraction module, we compare the classification results by replacing the multi-scale structure with traditional 3D convolutional layers. The kernel size of the traditional 3D convolution is set as $3 \times 3 \times 3$. The kernel sizes of the multi-scale structure are set as $1 \times 1 \times 1$, $3 \times 3 \times 3$ and $5 \times 5 \times 5$, respectively. As seen in Table 3, on the Houston dataset, the OA and Kappa values have increased, but the OA metric has slightly decreased. This is because the Houston dataset is more complex due to the smaller sample size and scattered distribution, limiting the ability for multi-scale feature extraction structure to extract features. The OA, AA and Kappa metrics have been improved on both the MUUFL and the Trento datasets which have more land cover categories. These results demonstrate that the multi-scale structure can extract richer features from HSI, enhancing the quality of multi-modal feature fusion and consequently improving classification accuracy.

For the cross attention mechanism module, we can observe that, the introduction of the cross attention mechanism contributes to an improvement in classification accuracy for all three datasets. The cross attention mechanism combines the features of both HSI and LiDAR, generating more enriched fusion features for the classification task. This validates the effectiveness of the cross attention mechanism in the fusion and classification of HSI and LiDAR data.

In the feature fusion module, we replace the traditional 2D-CNN by the depthwise separable convolutions. As observed in Table 3, the classification results are degraded when employing traditional convolutions across all three datasets. Conversely, through the incorporation of depthwise separable convolutions, the proposed model can efficiently learn spatial and channel-wise interactions independently, enabling the network to capture complex hierarchical features of the data. Moreover, depthwise separable convolutions can reduce the number of parameters and complexity compared to the traditional convolutions. This result can mitigate the risk of overfitting and improve generalization performance of the model, especially in situations with limited training data. A detailed complexity comparison is discussed in Section 3.5.

Table 3. The classification results of the ablation experiment. Bold values represent highest value of three metrics in each column.

Method	Metric	Classification Accuracy		
		Houston	MUUFL	Trento
No multiscale extraction module	OA	0.943	0.934	0.988
	AA	0.926	0.839	0.976
	Kappa	0.936	0.912	0.982
No cross attention mechanism	OA	0.927	0.939	0.992
	AA	0.916	0.854	0.983
	Kappa	0.924	0.923	0.986
No residual structure	OA	0.912	0.946	0.995
	AA	0.907	0.852	0.986
	Kappa	0.909	0.932	0.994
No depthwise separable convolutions	OA	0.924	0.938	0.989
	AA	0.937	0.825	0.972
	Kappa	0.918	0.918	0.986
The proposed framework	OA	0.942	0.953	0.995
	AA	0.938	0.862	0.987
	Kappa	0.938	0.938	0.994

Additionally, in the fusion module, the CNN adopts the residual structure in the process of backward propagation. To test the effect of the residual structure, we compare the experimental results of the proposed model and the model without adding the residual structure. As shown in Table 3, it is evident that the use of residual structures can help improve classification accuracy for all three datasets. The enhancement effect is more obvious on the Houston dataset since residual structures can better reduce data loss on smaller feature samples. Thus, the residual structure can help improve feature fusion quality and obtain higher classification accuracy.

In summary, each module in the architecture of our proposed AFC-CNN presents necessity and effectiveness.

3.5. Complexity Analysis

In this section, we analyze the computational complexity of the proposed method. Table 4 shows the parameters of six convolutional layers by using traditional 2D-CNN and depthwise separable convolutions in the feature fusion module. It is obvious that the parameter number of each convolutional layer employing 2D convolutions are significantly larger than that utilizing depthwise separable convolutions. Specifically, the parameter count for traditional 2D convolutions is 5,675,506 and for depthwise separable convolutions is 478,976, yielding approximately a 10.8-times reduction in the number of parameters. Thus, the proposed AFC-CNN can enhance training efficiency with less computation complexity with depthwise separable convolutions.

Table 4. The parameter comparison of feature fusion module with traditional 2D-CNN and depthwise separable convolutions.

Module	Type/Stride	Convolutional Kernel Size	Input Size	Parameters No.
Feature fusion module	Conv 2D/1	3 × 3	1152 × 11 × 11	3981746
	Conv 2D/1	3 × 3	384 × 9 × 9	221248
	(Conv 2D/1) × 4	3 × 3	64 × 7 × 7	368128

Table 4. Cont.

Module	Type/Stride	Convolutional Kernel Size	Input Size	Parameters No.
Feature fusion module	Conv DW/1	3 × 3	1152 × 11 × 11	10368
	Conv PW/1	1 × 1152	1152 × 9 × 9	442368
	Conv DW/1	3 × 3	384 × 9 × 9	3456
	Conv PW/1	1 × 64	384 × 7 × 7	4096
	(Conv DW/1, Conv PW/1) × 4	3 × 3 1 × 64	64 × 7 × 7 64 × 7 × 7	576 4096

3.6. Comparative Experiment

We compare the experimental results of AFC-CNN with the SVM and FusAtNet on the fusion of HSI and LiDAR data and single HSI, respectively. HSI is expressed as H, the LiDAR data is denoted as L, and H + L represents the HSI and LiDAR data are concatenated together for classification.

Table 5 lists the OA, AA and Kappa results on the three datasets. The proposed AFC-CNN is obviously superior to other methods in both single H and fusion H + L classification framework. The AFC-CNN (H + L) achieves the highest classification accuracy in terms of OA and Kappa metrics. For example, on the Houston dataset, the AFC-CNN (H + L) achieves an OA of 0.942, yielding approximately 4.6% and 10.6% improvement than FusAtNet (H + L) and SVM (H + L), respectively. The Kappa of the proposed AFC-CNN (H + L) on the three datasets are 0.938, 0.938 and 0.9, which is the best among all the classifiers. These results prove the effectiveness of the proposed algorithm.

Table 5. Classification performance on the three datasets. Bold values represent highest value of three metrics in each column.

Dataset	Metric	SVM		FusAtNet		AFC-CNN	
		H	H + L	H	H + L	H	H + L
Houston	OA	0.802	0.842	0.857	0.899	0.922	0.942
	AA	0.842	0.868	0.886	0.947	0.909	0.938
	Kappa	0.783	0.829	0.845	0.891	0.915	0.938
NUUFL	OA	0.873	0.884	0.894	0.915	0.937	0.953
	AA	0.585	0.603	0.707	0.786	0.828	0.862
	Kappa	0.819	0.837	0.858	0.887	0.916	0.938
Trento	OA	0.906	0.924	0.985	0.991	0.987	0.995
	AA	0.718	0.873	0.976	0.985	0.982	0.987
	Kappa	0.861	0.876	0.979	0.988	0.982	0.994

Compared with single H branch framework, the two-branch H + L framework can obtain better classification accuracy. The single H branch framework lacks detailed elevation information of LiDAR data, resulting in poor classification quality. So it can be concluded that the joint use of HSI and LiDAR can greatly improve the performance of classification. The classification accuracy of SVM (H) is the lowest, which indicates that the traditional machine learning methods still have deficiencies in classification tasks compared with deep learning algorithms. This is primarily due to the fact that the performance of the traditional machine learning algorithms depends highly on the handcrafted feature design and prior information, the representation ability is limited. The FusAtNet performs better than SVM since the deep learning methods utilize 2D-CNN to learn the high-level features that have advantages over traditional machine learning algorithms. However, the 2D-CNN used for HSI feature extraction can not effectively extract the spectral features of HSI and establish complementary connections between HSI and LiDAR data. Thus, the classification performance of FusAtNet (H) and FusAtNet (H + L) decreases slightly compared with AFC-CNN (H) and AFC-CNN (H + L).

The good classification performance of AFC-CNN benefits from three main reasons: (1) the 3D-CNN and a multi-scale structure in AFC-CNN can extract both spatial and spectral feature information of HSI. (2) AFC-CNN utilizes the cross attention mechanism module to enhance the inherent correlation between HSI and LiDAR by increasing the variance between different categories of samples. (3) In the feature fusion module, AFC-CNN adopts the residual structure to reduce data loss and fully extract advanced features from the data. Moreover, it replaces the traditional 2D convolution with depth-separable convolution, improving the fitting ability of the model. This enhances its feature representation ability of fusion data and better classification accuracy.

Tables 6–8 illustrate the per-class accuracy results on the Houston, MUUFL and Trento datasets, respectively. We can see that the proposed method shows better classification performance compared with other methods on almost all land-cover class. For example, in Table 7, the classification accuracy of AFC-CNN on grass, mixed ground surface and dirt and sand are 0.93, 0.95 and 0.91, yielding approximately 22.1%, 8.4% and 5.5% improvements over FusAtNet and 22.3%, 16.8% and 7.69% over SVM. SVM even fails to distinguish the terrain materials of water, building shadow and yellow curb in the MUUFL dataset.

Table 6. Classification accuracy of different methods on the Houston dataset. Bold values represent highest value of three metrics in each column.

Class Name	SVM		FusAtNet		AFC-CNN	
	H	H + L	H	H + L	H	H + L
Healthy grass	0.88	0.85	0.83	0.83	0.92	0.98
Stressed grass	0.86	0.92	0.85	0.96	0.98	0.99
Synthetic grass	0.99	0.99	1	1	0.97	1
Trees	0.98	0.99	0.92	0.93	0.95	0.98
Soil	0.98	0.96	0.97	0.99	0.98	0.99
Water	0.99	0.98	1	1	0.6	1
Residential	0.76	0.88	0.94	0.94	0.84	0.91
Commercial	0.62	0.72	0.76	0.92	0.94	0.97
Road	0.6	0.74	0.85	0.84	0.89	0.77
Highway	0.62	0.92	0.63	0.64	0.96	0.98
Railway	0.91	0.88	0.72	0.9	0.84	0.92
Parking Lot1	0.58	0.62	0.89	0.92	0.98	0.92
Parking Lot2	0.87	0.59	0.93	0.88	0.94	0.81
Tennis Court	0.99	0.99	1	1	0.97	1
Runing Track	1	0.99	1	0.99	0.98	1

Table 7. Classification accuracy of different methods on the MUUFL dataset. Bold values represent highest value of three metrics in each column.

Class Name	SVM		FusAtNet		AFC-CNN	
	H	H + L	H	H + L	H	H + L
Mostly grass	0.68	0.72	0.64	0.72	0.92	0.93
Mixed ground surface	0.77	0.79	0.86	0.87	0.87	0.95
Dirt and sand	0.81	0.84	0.87	0.86	0.86	0.91
Road	0.89	0.89	0.93	0.95	0.94	0.93
Water	0	0	0.25	0.91	0.87	0.88
Building shadow	0	0	0.73	0.74	0.91	0.93
Buildings	0.85	0.86	0.96	0.98	0.96	0.96
Sidewalk	0.56	0.62	0.56	0.6	0.82	0.85
Yellow curb	0	0	0.07	0.09	0.26	0.37
Cloth panels	0.91	0.94	0.92	0.93	0.87	0.74

Table 8. Classification accuracy of different methods on the Trento dataset. Bold values represent highest value of three metrics in each column.

Class Name	SVM		FusAtNet		AFC-CNN	
	H	H + L	H	H + L	H	H + L
Buildings	0.8	0.81	0.97	0.98	0.98	0.99
Ground	0	1	1	0.99	0.99	1
Woods	0.99	0.99	1	1	1	1
Vineyard	0.87	0.76	0.99	0.99	0.98	1
Roads	0.82	0.86	0.89	0.93	0.94	0.97

Figures 8–10 depict the classification map obtained by different classification methods on three datasets. It can be clearly seen that the proposed method achieves the most accurate and noiseless classification maps, such as, the grass and mixed ground surface on the MUUFL dataset, the vineyard class on Trento dataset. These visual results are consistent with the per-class accuracy results shown in Tables 6–8.

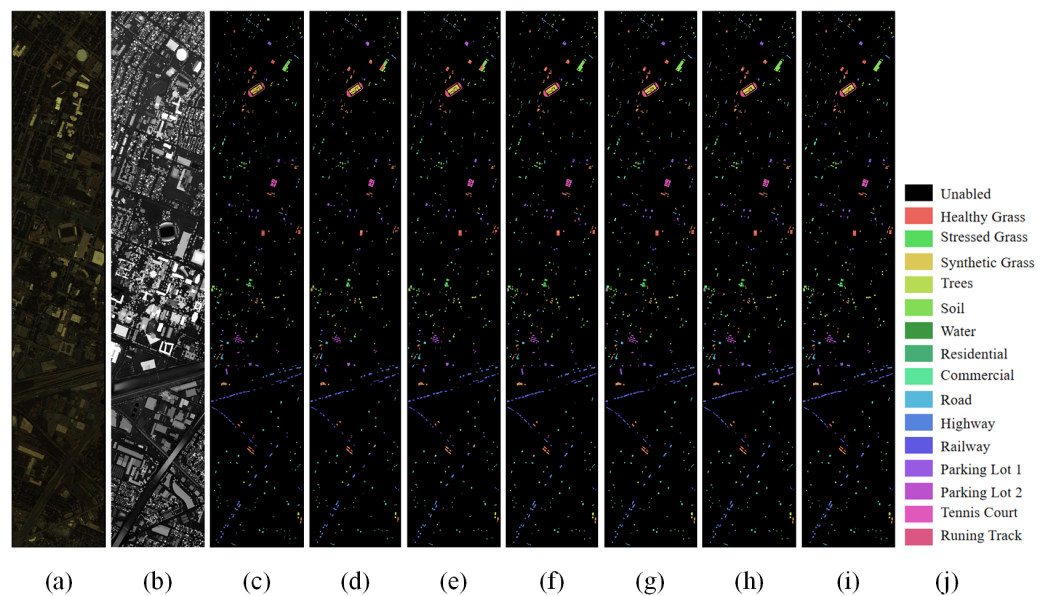


Figure 8. Dataset visualization and classification maps of the Houston dataset obtained with different models. From left to right: (a) HSI, (b) gray image for LiDAR, (c) label, (d) SVM (H), (e) SVM (H + L), (f) FusAtNet (H), (g) FusAtNet (H + L), (h) ours (H), (i) ours (H + L), (j) legend.

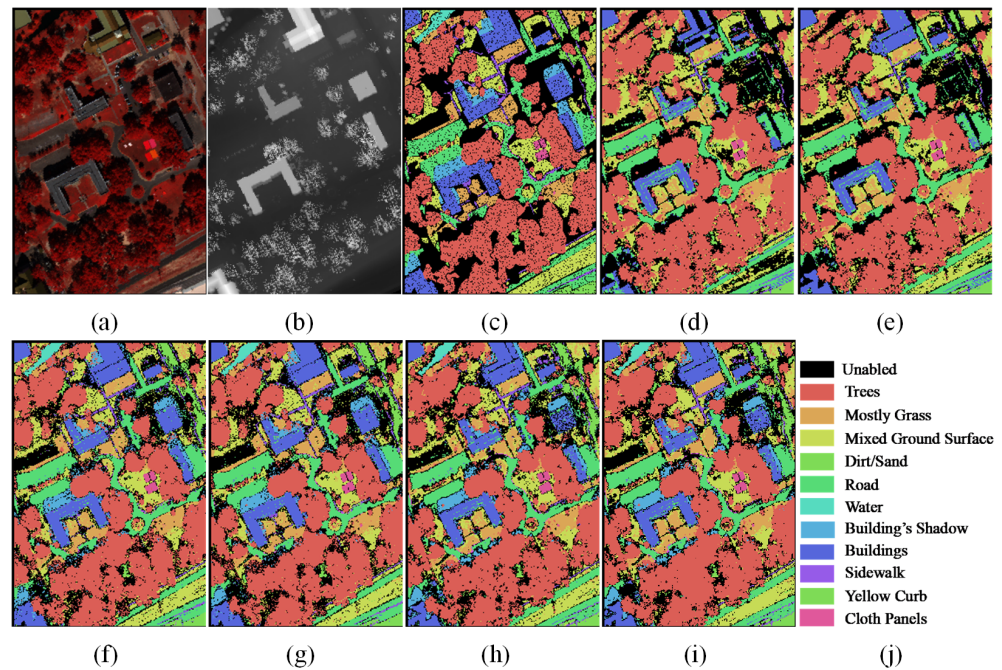


Figure 9. Dataset visualization and classification maps of the MUUFL dataset obtained with different models. From left to right: (a) HSI, (b) gray image for LiDAR, (c) label, (d) SVM (H), (e) SVM (H + L), (f) FusAtNet(H), (g) FusAtNet (H + L), (h) ours (H), (i) ours (H + L), (j) legend.

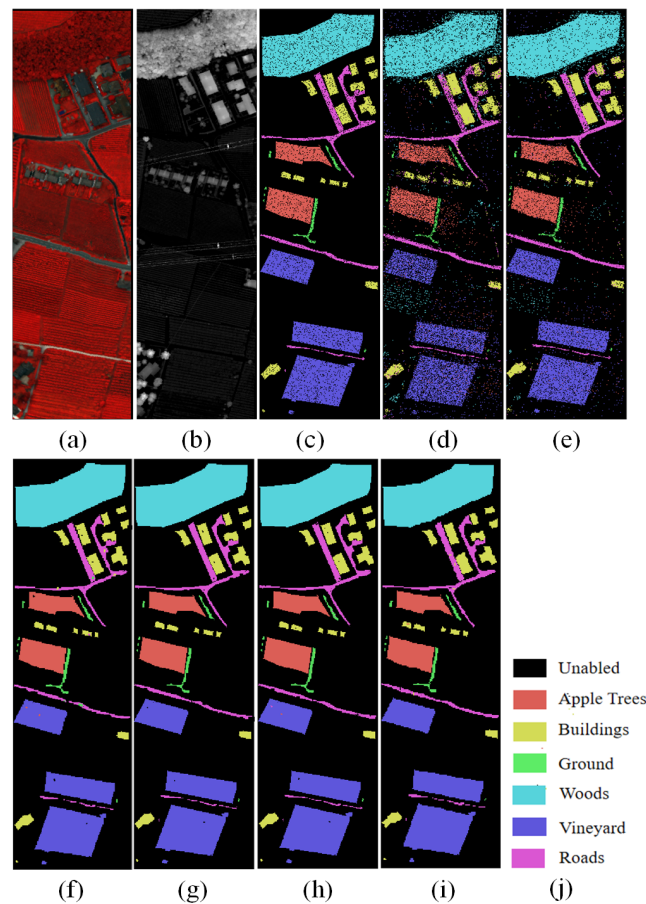


Figure 10. Dataset visualization and classification maps of the Trento dataset obtained with different models. From left to right: (a) HSI, (b) gray image for LiDAR, (c) label, (d) SVM (H), (e) SVM (H + L), (f) FusAtNet (H), (g) FusAtNet (H + L), (h) ours (H), (i) ours (H + L), (j) legend.

4. Conclusions

In this paper, we propose a new fusion and classification framework named AFC-CNN based on the joint use of HSI and LiDAR data. AFC-CNN applies 3D-CNN and a multi-scale structure to extract more spatial-spectral features in the HSI feature extraction branch. Moreover, the spectral attention mechanism is adopted to strengthen more important features from the spectral channels while reducing interference from less relevant features. To enhance the interaction between HSI and LiDAR data, AFC-CNN utilizes a cross attention mechanism module to impart the spatial significance weights from LiDAR to HSI. This integration harmoniously combines both data sources, leveraging more advanced feature representation. In the feature fusion module, AFC-CNN adopts a deeper network layer to augment the extraction of features from the fused data. The convolutional layers of this module are connected through residual structures to minimize data loss during backward propagation. Additionally, the depth-wise separable convolutions are used to reduce computational complexity and improve the fitting ability.

Through ablation experiments, we demonstrate the effectiveness of each module in our proposed framework. We evaluate the classification performance with traditional machine learning algorithms, namely SVM, and an advanced dual-branch classification network, namely FusAtNet. Extensive experimental results on three benchmark remote sensing datasets show that the proposed AFC-CNN can significantly outperform the compared methods in both single HSI branch and concatenated HSI and LiDAR branch frameworks.

Author Contributions: Conceptualization, Methodology, F.Y.; Writing—original draft, J.H.; writing—review and editing, J.H. and F.Y.; Software, Validation, Y.Z.; Funding acquisition, F.Y.; Supervision, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (62101392).

Data Availability Statement: The Houston dataset used in this study is available at: https://hyperspectral.ee.uh.edu/?page_id=1075 (accessed on 26 November 2023); the MUUFL dataset is available from: <https://github.com/GatorSense/MUUFLGulfport/> (accessed on 26 November 2023); the Trento dataset is provided by Lorenzo Bruzzone of the University of Trento and is available at: https://github.com/Ding-Kexin/IF_CALC?tab=readme-ov-file (accessed on 26 November 2023).

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

References

1. Jin, H.; Mountrakis, G. Fusion of optical, radar and waveform LiDAR observations for land cover classification. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 171–190. [[CrossRef](#)]
2. Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C. Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes. *Remote Sens. Environ.* **2022**, *268*, 112780. [[CrossRef](#)]
3. Taiwo, B.E.; Kafy, A.A.; Samuel, A.A.; Rahaman, Z.A.; Ayowole, O.E.; Shahrier, M.; Duti, B.M.; Rahman, M.T.; Peter, O.T.; Abosede, O.O. Monitoring and predicting the influences of land use/land cover change on cropland characteristics and drought severity using remote sensing techniques. *Environ. Sustain. Indic.* **2023**, *18*, 100248. [[CrossRef](#)]
4. Dian, R.; Li, S.; Sun, B.; Guo, A. Recent advances and new guidelines on hyperspectral and multispectral image fusion. *Inf. Fusion* **2021**, *69*, 40–51. [[CrossRef](#)]
5. Liu, Y.; Hu, J.; Kang, X.; Luo, J.; Fan, S. Interactformer: Interactive transformer and CNN for hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [[CrossRef](#)]
6. Ghamisi, P.; Höfle, B.; Zhu, X.X. Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *10*, 3011–3024. [[CrossRef](#)]
7. Xu, Y.; Du, B.; Zhang, L.; Cerra, D.; Pato, M.; Carmona, E.; Prasad, S.; Yokoya, N.; Hänsch, R.; Le Saux, B. Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 IEEE GRSS data fusion contest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1709–1724. [[CrossRef](#)]
8. Zhang, M.; Li, W.; Tao, R.; Li, H.; Du, Q. Information fusion for classification of hyperspectral and LiDAR data using IP-CNN. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–12.

9. Roy, S.K.; Deria, A.; Hong, D.; Ahmad, M.; Plaza, A.; Chanussot, J. Hyperspectral and LiDAR data classification using joint CNNs and morphological feature learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–16. [[CrossRef](#)]
10. Wang, X.; Feng, Y.; Song, R.; Mu, Z.; Song, C. Multi-attentive hierarchical dense fusion net for fusion classification of hyperspectral and LiDAR data. *Inf. Fusion* **2022**, *82*, 1–18. [[CrossRef](#)]
11. 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](#)]
12. Imani, M.; Ghassemian, H. An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges. *Inf. Fusion* **2020**, *59*, 59–83. [[CrossRef](#)]
13. Wu, H.; Dai, S.; Liu, C.; Wang, A.; Iwahori, Y. A novel dual-encoder model for hyperspectral and LiDAR joint classification via contrastive learning. *Remote Sens.* **2023**, *15*, 924. [[CrossRef](#)]
14. Sugumaran, R.; Voss, M. Object-oriented classification of LiDAR-fused hyperspectral imagery for tree species identification in an urban environment. In Proceedings of the 2007 Urban Remote Sensing Joint Event, Paris, France, 11–13 April 2007; pp. 1–6.
15. Dalponte, M.; Bruzzone, L.; Gianelle, D. Fusion of hyperspectral and LiDAR remote sensing data for classification of complex forest areas. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1416–1427. [[CrossRef](#)]
16. Puttonen, E.; Jaakkola, A.; Litkey, P.; Hyypää, J. Tree classification with fused mobile laser scanning and hyperspectral data. *Sensors* **2011**, *11*, 5158–5182. [[CrossRef](#)]
17. Pedergrana, M.; Marpu, P.R.; Dalla Mura, M.; Benediktsson, J.A.; Bruzzone, L. Classification of remote sensing optical and LiDAR data using extended attribute profiles. *IEEE J. Sel. Top. Signal Process.* **2012**, *6*, 856–865. [[CrossRef](#)]
18. Ghamisi, P.; Souza, R.; Benediktsson, J.A.; Zhu, X.X.; Rittner, L.; Lotufo, R.A. Extinction profiles for the classification of remote sensing data. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5631–5645. [[CrossRef](#)]
19. Gu, Y.; Wang, Q.; Jia, X.; Benediktsson, J.A. A novel MKL model of integrating LiDAR data and MSI for urban area classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5312–5326.
20. Xu, X.; Li, W.; Ran, Q.; Du, Q.; Gao, L.; Zhang, B. Multisource remote sensing data classification based on convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 937–949. [[CrossRef](#)]
21. Li, Y.; Zhang, H.; Shen, Q. Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sens.* **2017**, *9*, 67. [[CrossRef](#)]
22. Zhang, M.; Li, W.; Du, Q.; Gao, L.; Zhang, B. Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN. *IEEE Trans. Cybern.* **2018**, *50*, 100–111. [[CrossRef](#)]
23. Huang, B.; Zhao, B.; Song, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens. Environ.* **2018**, *214*, 73–86. [[CrossRef](#)]
24. Feng, Q.; Zhu, D.; Yang, J.; Li, B. Multisource hyperspectral and LiDAR data fusion for urban land-use mapping based on a modified two-branch convolutional neural network. *ISPRS Int. J.-Geo-Inf.* **2019**, *8*, 28. [[CrossRef](#)]
25. Chen, Y.; Li, C.; Ghamisi, P.; Jia, X.; Gu, Y. Deep fusion of remote sensing data for accurate classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1253–1257. [[CrossRef](#)]
26. Wang, J.; Zhang, J.; Guo, Q.; Li, T. Fusion of hyperspectral and lidar data based on dual-branch convolutional neural network. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3388–3391.
27. Ge, Z.; Cao, G.; Li, X.; Fu, P. Hyperspectral image classification method based on 2D–3D CNN and multibranch feature fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5776–5788. [[CrossRef](#)]
28. Keceli, A.S.; Kaya, A. Violent activity classification with transferred deep features and 3D-CNN. *Signal Image Video Process.* **2023**, *17*, 139–146. [[CrossRef](#)]
29. Mohla, S.; Pande, S.; Banerjee, B.; Chaudhuri, S. FusAtNet: Dual attention based spectrospatial multimodal fusion network for hyperspectral and LiDAR classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 92–93.
30. Li, H.C.; Hu, W.S.; Li, W.; Li, J.; Du, Q.; Plaza, A. A³CLNN: Spatial, spectral and multiscale attention convLSTM neural network for multisource remote sensing data classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *33*, 747–761. [[CrossRef](#)]
31. Li, J.; Ma, Y.; Song, R.; Xi, B.; Hong, D.; Du, Q. A triplet semisupervised deep network for fusion classification of hyperspectral and LiDAR data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [[CrossRef](#)]
32. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
33. Xiu, D.; Pan, Z.; Wu, Y.; Hu, Y. MAGE: Multisource attention network with discriminative graph and informative entities for classification of hyperspectral and LiDAR data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–14. [[CrossRef](#)]
34. Gader, P.; Zare, A.; Close, R.; Aitken, J.; Tuell, G. MUUFL Gulfport Hyperspectral and LiDAR Airborne Data Set. Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570. 2013. Available online: <https://github.com/GatorSense/MUUFLGulfport/> (accessed on 7 January 2023).
35. Rasti, B.; Ghamisi, P.; Gloaguen, R. Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3997–4007. [[CrossRef](#)]

-
36. Hong, D.; Gao, L.; Yokoya, N.; Yao, J.; Chanussot, J.; Du, Q.; Zhang, B. More diverse means better: Multimodal deep learning meets remote-sensing imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4340–4354. [[CrossRef](#)]
 37. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

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.