


## Article

# Semi-Supervised Tree Species Classification for Multi-Source Remote Sensing Images Based on a Graph Convolutional Neural Network

Xueliang Wang <sup>1,\*</sup> , Jian Wang <sup>2</sup>, Zuozheng Lian <sup>3</sup> and Nan Yang <sup>4</sup><sup>1</sup> Network Information Center, Qiqihar University, Qiqihar 161006, China<sup>2</sup> National Asset Management Office, Qiqihar University, Qiqihar 161006, China<sup>3</sup> School of Computer and Control Engineering, Qiqihar University, Qiqihar 161006, China<sup>4</sup> School of Architecture and Civil Engineering, Qiqihar University, Qiqihar 161006, China

\* Correspondence: wangxueliang@qqhru.edu.cn

**Abstract:** As a current research hotspot, graph convolution networks (GCNs) have provided new opportunities for tree species classification in multi-source remote sensing images. To solve the challenge of limited label information, a new tree species classification model was proposed by using the semi-supervised graph convolution fusion method for hyperspectral images (HSIs) and multispectral images (MSIs). In the model, the graph-based attribute features and pixel-based features are fused to deepen the correlation of multi-source images to improve accuracy. Firstly, the model employs the canonical correlation analysis (CCA) method to maximize the correlation of multi-source images, which explores the relationship between information from various sources further and offers more valuable insights. Secondly, convolution calculations were made to extract features and then map graph node fusion, which not only reduces redundancy features but also enhances compelling features. Finally, the relationship between representative descriptors is captured through the use of hyperedge convolution in the training process, and the dominant features on the graph are fully mined. The tree species are classified through two fusion feature operations, leading to improved classification performance compared to state-of-the-art methods. The fusion strategy can produce a complete classification map of the study areas.



**Citation:** Wang, X.; Wang, J.; Lian, Z.; Yang, N. Semi-Supervised Tree Species Classification for Multi-Source Remote Sensing Images Based on a Graph Convolutional Neural Network. *Forests* **2023**, *14*, 1211. <https://doi.org/10.3390/f14061211>

Academic Editor: Feng Zhao

Received: 19 April 2023

Revised: 25 May 2023

Accepted: 9 June 2023

Published: 11 June 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/>).

**Keywords:** hypergraph convolution; data fusion; classification of tree species

## 1. Introduction

Recently, achieving accurate and reliable tree species classification from a large number of trees has gained more attention. Multi-source products typically provide more trustworthy information than a single product of ground surface covering [1,2]. Hyperspectral images (HSIs) are an essential part of multi-source data learning and can reflect the spectral characteristics of forest mapping, which is crucial for understanding forest cover [3]. Multispectral data images (MSIs) contain high-resolution spatial information, which is also helpful when analyzing forest tree species. By integrating multi-source data, data fusion can overcome the limitations of a single data source [4]. Current fusion methods for HSIs and MSIs rely on feature extraction and feature fusion, respectively, to leverage the correlation between the two data sources [5]. To leverage diverse information from multiple sources of data, it is necessary to implement strategies that enable the effective extraction, integration, and analysis of data [6–8].

Deep learning has been applied to feature fusion to improve its performance in terms of feature fusion and has achieved satisfactory results [9–11]. Li et al. proposed an effective CNN (PPF-CNN) based on pixel features [12] in combination with a small number of existing samples, which enabled data enhancement to optimize the classification results. A

multi-region CNN (MRCNN) [13] algorithm is proposed to mine spectral–spatial information, which improves the mining performance. However, with the further development of the network, limited labeled samples may lead to overfitting or performance degradation. Fortunately, a semi-supervised approach can be used to mitigate this shortcoming. The insufficiency of labels in remote sensing image classification can now be addressed with the recent emergence of graph convolution networks. The recent advancement of GCNs has provided a promising solution to the insufficient label problem in hyperspectral or multispectral image (HSI/MSI) classification. Unlike traditional methods, GCNs operate on a graph and require only a small amount of labeled data to establish the relationships between multi-source nodes. By effectively aggregating and transforming features from a node’s neighborhood, GCNs provide an efficient pathway for multi-source image classification [14]. GCNs are particularly suitable for handling non-Euclidean data, which refers to datasets that do not adhere to the principles and assumptions of Euclidean geometry, and by learning node features through hidden layers, they better capture the local features, resolving the issue of missing class boundary information. The Chebyshev polynomial [15] parameterized differentiable graph convolution algorithm is used by GCNs to transport the node information after using the feature construction of all samples to create the topological structure (G-Conv). The whole learning process of this method does not need manual intervention. By excavating the structural information of many unlabeled samples in the feature space, the deviation of learning trained with labeled samples is corrected. The potential value of unlabeled samples is fully utilized, and the ‘small sample’ problem in terms of classification is effectively solved. Not only is this approach applicable to non-Euclidean statistics, but it also has broad applicability to standard domains [16]. In reality, there is much research on the use of GCNs in relation to remote sensing images. For example, Qin developed a spectral–spatial GCN (S2GCN) by employing current pixel spatial information [17], which has made significant improvement to the original GCN. However, at the end of the above network, the SoftMax function is usually used to analyze the extracted features, which generates a probability vector that reflects the category of the pixel. This method lacks intraclass compactness, which reduces classification performance [18]. Spectral and spatial information were extracted to construct adjacent matrices, and an innovative prototype layer was designed. This prototype layer contains distance-based cross-entropy loss function and novel temporal entropy-based regularization, which can not only generate more low-level features, such as separable between species and compact within species, but also represent the prototypes belonging to each species [19].

Most methods extract features and then combine them using various techniques. Additionally, low-rank model methods are used to convert multiple sources of features into a common space through low-rank sparse representation. Feature fusion strategies are used to convert multi-source features into a unified fused feature, but the process of feature extraction and fusion are separate, which may result in changes to the original information contained in the features. In graph representation learning, taking into account both the global and local structure of the data can make the graph representation model more robust against the effects of noisy and sparse data. However, there have been only a limited number of GCN models that have prioritized preserving both the local and global structures of the data concurrently.

Existing graph/hypergraph-based neural networks suffer from a significant limitation in that they only make use of the initial graph/hypergraph structures and do not account for dynamic modifications that may occur in the feature embedding process. This limitation hinders the network’s ability to adapt to changing input data.

To address this issue, it is crucial to develop approaches that can account for the modifications of graph/hypergraph structures and ensure that the original information in the features remains intact throughout the fusion process. A semi-supervised graph model is proposed based on an extraction fusion network for HSIs and MSIs, to fully use the correlation of multi-source data. The feature extraction method is directed by the model via feature fusion. The model directly outputs unified fusion features from multi-source

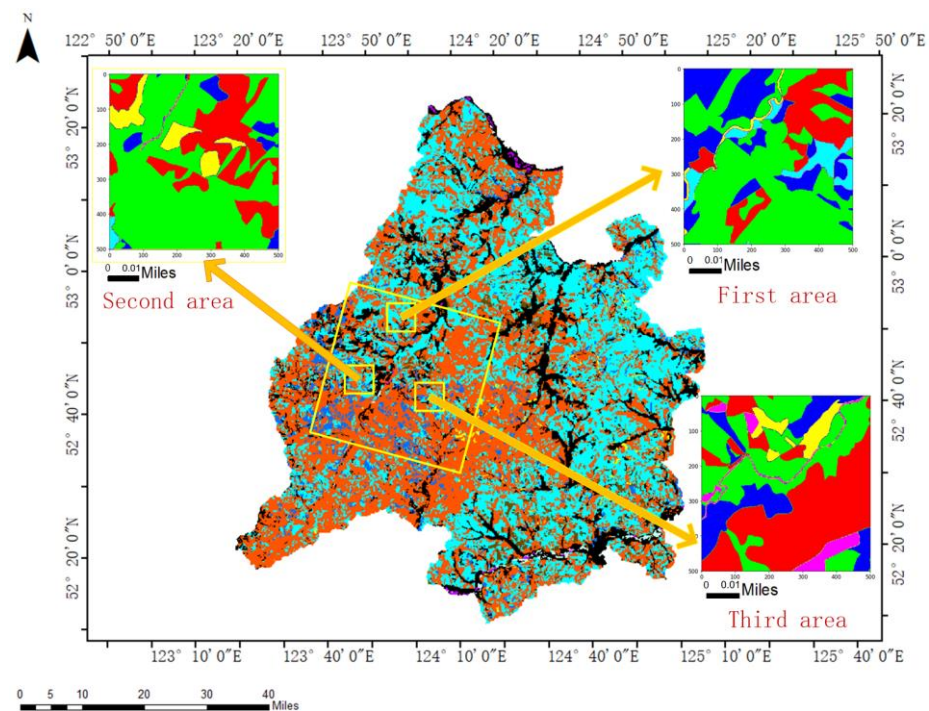
data as input. For feature fusion, a multimodal graph is built, and feature extraction is constrained using the graph-based loss function. The innovation points of the paper are as follows:

- (1) To extract the discriminative features, a common subspace is explored and found by CCA operations on HSI and MSI, and the correlation is maximized between HSI and MSI inputs.
- (2) For the information fusion between HSI and MSI, both the node features and hypergraph features are integrated to improve the ability of global information extraction, and the ability to express the relationship between all vertices becomes more robust. During the initialization of hypergraph convolution, feature fusion is performed on the nodes, and the hyperedge features are fused in the process of hypergraph convolution learning.
- (3) Compared with other state-of-the-art converged networks, it is more efficient and achieves better classification results.

## 2. Materials and Methods

### 2.1. Study Area

The areas were studied in the Tahe Forestry Bureau (Figure 1), which is located in the Daxing'an Mountains, northwest of Heilongjiang Province, China ( $123^{\circ}$  to  $125^{\circ}$  E and  $52^{\circ}$  to  $53^{\circ}$  N). The studied areas have a borderline of 173 km and a total area of 14,420 km<sup>2</sup>. The climate of the studied areas is a cold-temperate continental climate and experiences severe climatic changes, with short hot, humid summers and long, dry, cold winters. The annual average temperature of the area is  $-2.4^{\circ}\text{C}$ , and the average yearly precipitation is 463.2 mm, occurring mainly in July and August. The forest, with a storage capacity of 53.4 million m<sup>3</sup>, covers 81% of the total area. Dominant tree species include Birch, Larch, Spruce, Mongolica Pine, Willow, and Poplar [20].



**Figure 1.** Map of the study area.

### 2.2. Data

To classify the tree species, we used data taken from HJ-1A and Sentinel-2. Figure 1 displays the HSI data for HJ-1A and the MSI data for Sentinel-2A, collected from the China Center for Resources Satellite Data and Application and the USGS, respectively. The

HJ-1A satellite has a high-speed imaging system with 115 bands and a spatial resolution of 100 m [13], while Sentinel-2A offers 13 spectral bands with a spatial resolution of 10 m, providing rich data for coastal and land remote sensing [20]. We used ENVI 5.1 software to enhance the resolution of the HJ-1A/HSI images (collected on 20 August 2016) to match the MSI spatial resolution and fill the gap concerning the relatively low HSI resolution. The interpolation method was used to resample the experimental HSI data. The Tahe Forestry Bureau conducted a survey in 2018 and used the results to classify major forest species in the research region. The study areas were  $500 \times 500 \times 115$  pixels and  $500 \times 500 \times 13$  pixels for HSI and MSI data, respectively. We selected the area with the most species as the research object, which included Birch, Larch, Spruce, Mongolia, Willow, and Poplar. Table 1 lists the three study areas used in this work, where the training samples comprise approximately one-third of the total samples.

**Table 1.** List of 6 tree species samples of the three study areas.

	Birch	Larch	Mongolia	Poplar	Spruce	Willow
First area	130,124	39,216	57,620	3019	15,330	3492
Second area	150,771	58,829	11,412	2175	17,048	1067
Third area	99,082	82,746	38,114	1013	13,460	1,515,486

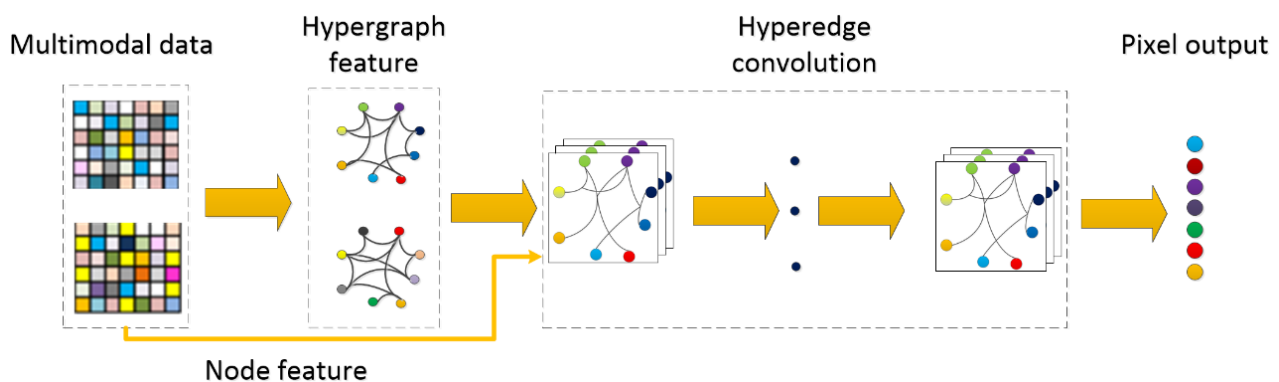
### 2.3. Classification Method

#### 2.3.1. Hypergraph

The Hypergraph Neural Network, which is commonly referred to as HGNN [21], has been visually depicted in Figure 2. Each dataset in the multimodal dataset contains numerous nodes with features. Then, using the complex correlation of multimodal data sets, several parts of hyperedge features are constructed. The hypergraph adjacency matrix and node features are input into HGNN to output the pixel features classification map [22]. Hyperedge convolution is computed as follows:

$$X^{k+1} = \sigma \left( D_v^{-\frac{1}{2}} H W D_e^{-1} H^T D_v^{-\frac{1}{2}} X^k \Theta^k \right) \quad (1)$$

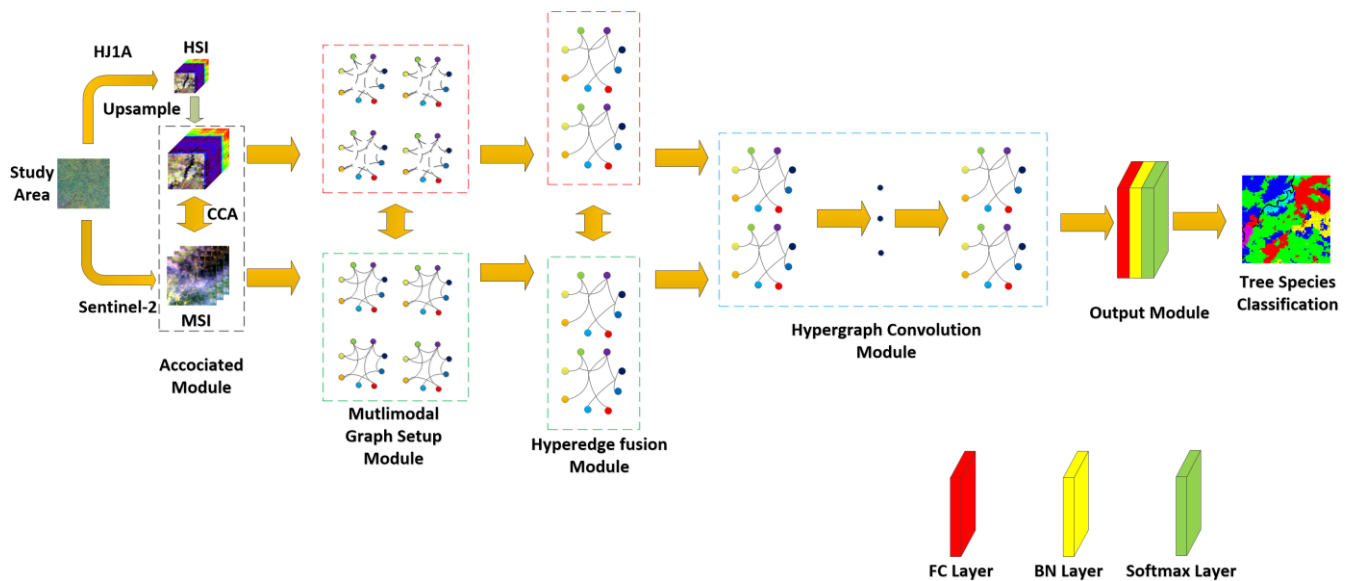
where  $X^k \in \mathbb{R}^{N \times C}$  is the feature of the  $l$  th layer.  $X^0 = X$ , and  $\sigma$  is the nonlinear activation function. The initial node features  $X^k$  are learned through filtering matrix  $\Theta^k$  to extract the dimensional feature  $C_2$ . Then, according to the node features of the hyperedge  $\mathbb{R}^{E \times C_2}$ , the hyperedge feature is realized via  $H^T \in \mathbb{R}^{E \times N}$ . The output node feature is then produced by multiplying the hyperedge features that are associated with it, and the hyperedge feature is produced from the matrix  $H$ .  $D_v$  and  $D_e$  in the Hyperedge convolution play the role of normalization [21]. Therefore, through hyperedge convolution, the HGNN layer can successfully extract the high-order correlation of the hypergraph.



**Figure 2.** Hypergraph neural network (HGNN).

### 2.3.2. Overall Architecture

The graph neural network employs an undirected graph to model the data and utilizes graph convolution for feature extraction by calculating various data relationships. Building on this method, we propose a tree species classification model that leverages the distinct framework of hyperspectral and multispectral data modules for feature fusion. The model takes in a multi-source remote sensing image as input and produces unified fusion features as output, as illustrated in Figure 3. The framework encompasses association feature extraction, hypergraph convolution learning, and classifier classification.



**Figure 3.** Flowchart of multi-source fusion hypergraph convolution network.

The fusion module is designed to extract and merge features from both HSI and MSI data. The weight matrices of HSI and MSI are merged to generate the incidence matrix of the multimodal graph, which accounts for complementary information and correlations between the two data sources. The feature extraction and fusion network is trained using a loss function that incorporates graph embedding, enabling the network to effectively capture the features of interest. Finally, the SoftMax classifier is used to categorize the tree species map at the pixel level.

To lower the dimensionality of the HSI data from 115 to 12, we first employ the KPCA approach. This generates a vector that represents each pixel in the data collection. The complete image's vector is then fed as input to the network. In this setup,  $X^H$  and  $X^L$  correspond to the HSI and MSI data, respectively,

$$X^H = \{X_1^H, X_2^H, \dots, X_n^H\}, \quad X_i^H \in \mathbb{R}^h \quad (2)$$

$$X^L = \{X_1^L, X_2^L, \dots, X_n^L\}, \quad X_i^L \in \mathbb{R}^m \quad (3)$$

where  $h$  and  $m$  are the numbers of spectral channels for the HSI and MSI, respectively, and  $X$  is the vector representing the  $i$ -th pixel. Therefore, the input of the network is as follows:

$$X = \{X_1, X_2, \dots, X_n\}, \quad X_i \in \mathbb{R}^{h+m} \quad (4)$$

where  $X_i = \text{CAT}(X_i^H, X_i^L)$ , and  $\text{CAT}()$  represents concatenate operation. Next, we feed  $X$  into the network for feature extraction and fusion. Although the network structure is not the primary focus of our research, the multimodal graph and graph loss are crucial for feature extraction and fusion. We employ the Smish method [23] as the activation

function in this study. In practice, the foundation of the feature extraction network can be substituted with other networks, such as a convolution layer, because the network's input consists of multimodal images. The network outputs unified fused features using the loss function based on the multimodal graph, which are then provided to the classifier for pixel recognition. We adopt the FC layer and Softmax layer as the output layer of the proposed network to demonstrate the potential of multimodal and loss-based graphs function.

### 2.3.3. Associated Feature Module

The primary objective of multi-source learning is to establish the connection between various data, which is crucial for comprehending the relationship depicted in multi-source remote sensing images. By exploiting the relationship between different viewpoints, we can improve the final interpretation performance [24,25]. This research area has received increasing attention in the field of data mining over the past decade [26,27]. In this part, we focus on multi-perspective learning from the perspective of feature fusion and classification methods. We use the common subspace approach, which maximizes the correlation between two inputs, as explored via the CCA method. This standard two-view subspace learning approach is employed to achieve our research objectives.

For a multi-source learning problem, hyperspectral and multispectral images are represented as  $\alpha \in \mathbb{R}^{L \times W \times H}$ ,  $\beta \in \mathbb{R}^{L \times W \times M}$ , respectively, where  $L$  represents the length,  $W$  represents the width, and  $H$  and  $M$  represent the number of bands in the two data sources, respectively. Then,  $\alpha$  and  $\beta$  are transformed into  $V^{v \times H}$  and  $V^{v \times M}$ , respectively,  $v = L \times W$ . We assume that the linear representation of  $\alpha$  and  $\beta$  are represented as follows,

$$U_H = r_1(\alpha) \quad (5)$$

$$U_M = r_2(\beta) \quad (6)$$

$r_1, r_2$  represent the projection directions of HSI and MSI, respectively. CCA is obtained by maximizing the correlation between  $\alpha$  and  $\beta$ . The first projection direction can be obtained by optimizing the following equation, and  $r_1, r_2$  represent the HSI and MSI projection axes, respectively. By maximizing the correlation between  $\alpha$  and  $\beta$  with the vector generated by CCA, the following equation can be optimized to yield the initial projection direction,

$$\begin{aligned} \max \rho(r_1, r_2) &= r_1^T S_{HM} r_2 \\ \text{s.t. } r_1^T S_{HH} r_1 &= 1, r_2^T S_{MM} r_2 = 1 \end{aligned} \quad (7)$$

$S_{HM}$  is the covariance matrix of the HSI and MSI among them. The Lagrangian multiplier operator can be used to maximize the objective function and find the optimal solution sum of  $r_1^*$  and  $r_2^*$  for the problem.

$$U_H^* = r_1^*(\alpha) \quad (8)$$

$$U_M^* = r_2^*(\beta) \quad (9)$$

Multi-source image categorization involves assigning the same space to data from various sources, as based on Equations (8) and (9). The use of a sum representation enhances the relevance of the data and features, which is highly beneficial for the multi-source classification of tree species. This approach not only processes the initial input but also reduces its redundancy and complexity. However, the rate of convergence for deep learning is slow [28]. By providing HSI- and MSI-related features, this approach enables the development of the depth model, which can lead to further improvements in classification performance.

### 2.3.4. Multi-Source Hypergraph Fusion

To efficiently integrate information across multimodal images, input pixels are represented using a graph structure. Compared to CNNs, the graph structure offers a higher capacity to capture the relationship between all of the vertices, as the size of the convolution kernel in a CNN limits the extraction of global information.

Both the association features of HSIs and MSIs processed by the CCA algorithm are  $Q_h^{L \times W \times H} = U_H^*$ ,  $Q_m^{L \times W \times M} = U_M^*$ , respectively. Each pixel was represented as a vertex of the hypergraph, and their dimensions are transformed into  $X^{n \times H}$  and  $Y^{n \times M}$ , where  $n = L \times W = |V|$  is the number of hypergraph vertices, and  $H$  and  $M$  represent the spectral dimensions of the HSIs and MSIs, respectively. Their features are extracted as  $X_i^{n \times H_i}$  and  $Y_i^{n \times M_i}$ . For each vertex  $v \in V$  and the hyperedge  $e \in E$ , the incidence matrix generated from the selected  $k$  nearest neighbors is  $H^{|V| \times |E|}$ , where,  $|V| = |E| = n$ .

$$h(i, j) = \begin{cases} e^{-\frac{n\sigma\|x_i-x_j\|^2}{\sum_{i=1}^n d(x_i, x_j)}}, & x_i \in N_k(x_j) \\ 0 & \end{cases} \quad (10)$$

where  $\sigma$  is an adjustable hyper-parameter,  $d(X_i, X_j)$  is the Euclidean distance between the two vertices  $X_i$  and  $X_j$ . The mean value is used to regulate the multimodal distance and simplify the process of adjusting the hyperparameters.

It was assumed that  $[f_1, f_2, \dots, f_n]$  is a multimodal feature vector. According to Equation (10), the incidence matrix  $[H_1^h, H_2^h, \dots, H_n^h]$  and  $[H_1^m, H_2^m, \dots, H_n^m]$  of HSIs and MSIs are calculated, respectively. Then, the fused features are obtained as  $H_f^h = \text{CAT}(H_1^h, H_2^h, \dots, H_n^h)$ ,  $H_f^m = \text{CAT}(H_1^m, H_2^m, \dots, H_n^m)$ , where  $\text{CAT}()$  represents the multi vector connection operation. Then, the obtained hyperedge features are further studied.

### 2.3.5. Hyperedge Learning

To obtain fused hyperedges from multimodal features, we connect their incidence matrices. This process enables the hypergraph convolution in Equation (1) to be applied, which becomes

$$X^{l+1} = \sigma\left(D_v^{-\frac{1}{2}} H_f W_f D_e^{-1} H_f^T D_v^{-\frac{1}{2}} X^l \Theta^l\right) \quad (11)$$

In the case of without considering regularization [29], the equation is simplified as follows:

$$X^{l+1} = \sigma\left(H_f W_f H_f^T X^l \Theta^l\right) \quad (12)$$

Since  $H$  and  $W$  are diagonal matrices, the equation becomes

$$X^{l+1} = \sigma\left(\left(H_1 W_1 H_1^T + \dots H_n W_n H_n^T\right) X^l \Theta^l\right) \quad (13)$$

For multi-source remote sensing images, each node has many characteristics [30]. The hyperedges of its hypergraph are first learned, respectively, then they are integrated. The objective function in backpropagation is calculated via cross-entropy loss function and the final feature map outputs with pixel-level SoftMax function. The Algorithm 1 is presented as follows:

**Algorithm 1** Pseudo code of hypergraph feature fusion for HSI and MSI

---

Input: HSI associated feature  $X_H$ , MSI associated feature  $X_M$ , neighbor node number  $k$ , iteration number of layer  $n$ , number of graph convolution layer  $g$ .

- 1: Generate  $X'_H$  and  $X'_M$  by flattening  $X_H$  and  $X_M$ , respectively
- 2: Generate  $X$  by connecting  $X'_H$  and  $X'_M$  horizontally
- 3: Generate the fusion incidence matrix of HSI and MSI as  $H$ , according to Equations (8) and (9)
- 4: Calculate the degree diagonal matrix  $De$  of the hyperedge and the degree diagonal matrix  $Dv$  of the vertex
- 5: Initialization parameters  $W$  and  $\Theta$
- 6: for  $i = 1$  to  $n$
- 7:   for  $j = 1$  to  $g$
- 8:     Calculate characteristic  $X$  according to Equation (10)
- 9:      $X_{pre} = \text{SoftMax}(\text{BN}(\text{FC}(\text{Hconv}(X))))$
- 10:     Calculate losses  $L$ , update  $W$  and  $\Theta$
- 11:     Gradient back propagation
- 12:   end for
- 13: end for
- 14: Output tree species classification map based pixel node

---

## 2.3.6. Evaluation Indicators

To test the tree species classification accuracy of the proposed method, the OA, average accuracy (AA), and Kappa coefficient (kappa), were determined using Equations (14)–(16), respectively.

$$OA = \frac{\sum_{i=1}^k C(i, i)}{M}, \quad (14)$$

$$AA = \frac{\sum_{i=1}^k OA}{K}, \quad (15)$$

$$kappa = \frac{M \sum_{i=1}^k C(i, i) - \sum_{i=1}^k (C(i, +)C(+, i))}{M^2 - \sum_{i=1}^k (C(i, +)C(+, i))}, \quad (16)$$

where  $i$  and  $k$  represent  $i$ -th tree species and the size, respectively. OA represents the proportion of correctly classified samples in the whole test sample, AA denotes the average accuracy of every tree species, and kappa is a statistical measure that reflects the consistency between the ground truth and classified ground maps.

## 3. Results

## 3.1. Experimental Setup

The experiment uses HJ-1A and Sentinel2A images as datasets, which were introduced in Section 2.2. Several compared models are as follows:

**SpectralNET [31]:** A deep learning method for spectral clustering by embedding input data points into the eigenspace of their associated graph Laplacian matrix and subsequently clusters them.

**FuNet [32]:** A new minibatch GCN was proposed by training large-scale GCNs in a mini-batch mode. The method has the ability to predict data that is not part of the training set without the need to retrain the networks.

**MFDF [33]:** A classification model based on decision fusion between multiple features and super-pixel segmentation, which integrated 2D and 3D Gabor features of multi-source datasets.

**DMULN [6]:** end-to-end pattern model which integrates the multi-view features, and the view union pool was proposed by associating with the feature extractor, and the fused features are input into the classifier.

The proposed model and other compared methods were evaluated using 10%, 20%, and 30% of the samples as randomly chosen training sets. For the other samples, we



allocated 30% of the samples as validation sets and the rest as test sets randomly. Parameter settings have a great impact on performance. Although the resolution of the datasets is different, the resolution of the geomap is fixed. The experiment is implemented in Python 3. The parameters of graph convolution are set in Table 2, where ‘Hconv’ refers to the hypergraph convolution layer.

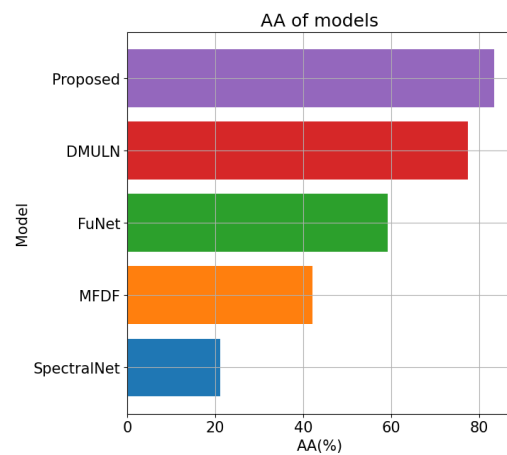
**Table 2.** Detailed layers and shape in multi-source fusion hypergraph convolution model.

Layer	Shape	Layer	Shape
Input	$(500 \times 500 \times 115)$	Input	$(500 \times 500 \times 12)$
	CCA		$(500 \times 500 \times 17)$
	Calculate $W_h$		Calculate $W_m$
	Normalization		Normalization
Hconv	128	Hconv	128
	Smish		Smish
	Fusion hypergraph		
	Hconv		
	FC Layer		
	BN Layer		
	Softmax		

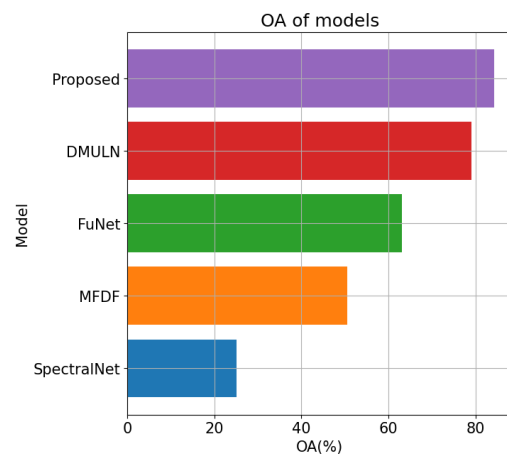
After hypergraph fusion, the proposed model consisted of two FC-BN layers and two active layers. The patch size was set to 7, and we set both the learning rate and weight decay to 0.005. We used the KNN method ( $k = 10$ ) to construct the initial graph for the datasets, with  $k$  values set to [5,10,15,20] and the number of convolution layers set to 15. We initialized the weights of all methods using the Glorot method. Adam was utilized as an optimizer, with a maximum of 1000 epochs. To ensure the optimal performance of other comparative models, we consulted the relevant literature. The method was repeated 100 times, with the average outcome for 10 iterations and the corresponding standard deviation used as the result. The training procedure was terminated if the loss did not decrease for 100 consecutive epochs.

### 3.2. Classification Performance Comparison

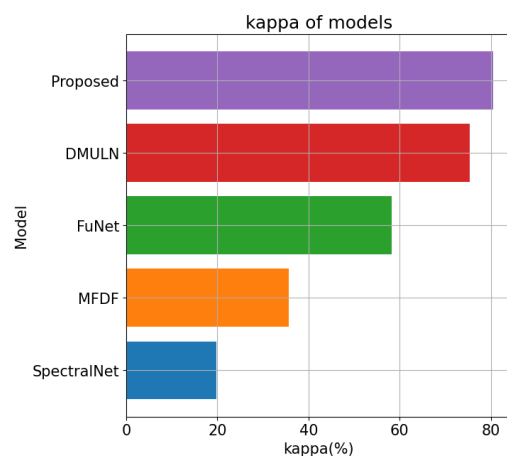
The average accuracy (AA) of tree species classification in terms of the three multi-source datasets is shown in Figure 4. The proposed method achieves the highest performance, followed by MFDF, FuNet, DMULN, and SpectralNet, as shown. The proposed model outperforms MFDF, FuNet, DMULN, and SpectralNet by 0.67, 0.46, 0.3, and 0.7, respectively, in terms of OA, as shown in Figure 5. Figure 6 shows the KAPPA values of the proposed model are 0.38, 0.23, 0.16, and 0.69 higher than those of MFDF, FuNet, DMULN, and SpectralNET, respectively. These results indicate that the proposed method is superior to the other methods. The performance of the proposed model is further demonstrated in the three figures.



**Figure 4.** AA of all methods across the tree species dataset.



**Figure 5.** OA of all methods across three datasets.



**Figure 6.** Kappa of all methods across three datasets.

Tables 3–6 illustrate the confusion matrix of five models used for tree species classification, and MFDF, FuNet, DMULN, and SpectralNET are unsatisfactory compared to the proposed model. Spruce is particularly challenging to classify, but the proposed model has a higher recognition rate for Spruce than that in the other models, leading to an overall increase in OA. The classification effect of the SpectralNET model is inadequate, as it identifies almost no other tree species except for Larch and Birch. Other models also lack the advanced ability to identify multiple tree species. The DMULN model mistakenly

classified almost 30% of Spruce trees as Larch, while the recognition rate of Poplar was 0. The OA of MFDF for Poplar is 0.61, which is better than DMULN. The classification performance of FuNet for other tree species is slightly better than DMULN, except for Spruce. As shown in Figure 5 and Table 7, the recognition rate of the proposed model for Spruce, Mongolian, and Willow is better than that of MFDF, but the recognition rate for Spruce is not significantly improved.

**Table 3.** Confusion matrix of tree species classification using the SpectralNet method. Tree species code, column (ground truth code), and row (prediction code).

Tree Species	Code	0	1	2	3	4	5
Birch	0	462	513	101	99	2	0
Larch	1	428	2535	701	311	0	0
Spruce	2	433	1411	525	103	0	0
Mongolica	3	81	153	52	49	0	0
Willow	4	4	100	0	0	0	0
Poplar	5	2	88	0	0	0	0
Precision		32.76	52.81	38.07	8.71	0	0

**Table 4.** Confusion matrix of tree species classification when using the DMULN method.

Tree Species	Tree Species Code	0	1	2	3	4	5
Birch	0	805	342	4	25	0	0
Larch	1	297	3600	16	51	1	0
Spruce	2	186	11.21	507	44	0	0
Mongolica	3	70	103	14	146	0	0
Willow	4	1	68	0	1	32	0
Poplar	5	15	67	1	5	0	0
Precision		68.30	90.01	27.29	43.30	31.69	0

**Table 5.** Confusion matrix of tree species classification when using the FuNet method.

Tree Species	Tree Species Code	0	1	2	3	4	5
Birch	0	859	240	42	34	0	1
Larch	1	314	35.66	31	47	0	3
Spruce	2	63	850	930	14	0	1
Mongolica	3	56	126	0	151	0	0
Willow	4	0	70	1	2	30	0
Poplar	5	1	33	0	0	0	55
Precision		73.03	89.98	50.0	45.19	29.29	61.14

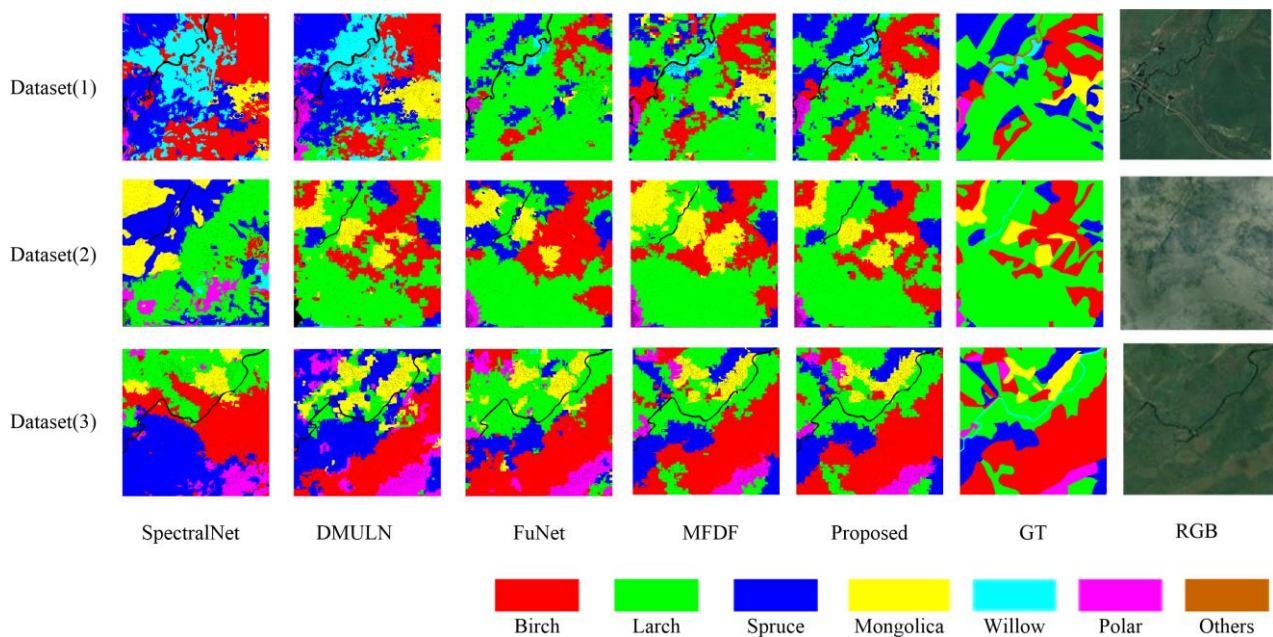
**Table 6.** Confusion matrix of tree species classification when using the MFDF method.

Tree Species	Tree Species Code	0	1	2	3	4	5
Birch	0	10.81	71	1	21	0	1
Larch	1	390	35.25	7	35	2	3
Spruce	2	39	496	13.00	24	0	0
Mongolica	3	39	32	1	261	0	0
Willow	4	0	13	1	7	82	0
Poplar	5	3	5	0	0	0	81
Precision		91.90	88.95	69.85	78.13	78.72	90.28

**Table 7.** Confusion matrix of tree species classification when using the proposed method.

Tree Species	Tree Species Code	0	1	2	3	4	5
Birch	0	10.82	71	1	21	0	1
Larch	1	286	3601	74	2	1	0
Spruce	2	48	79	1714	6	5	5
Mongolica	3	12	13	14	290	2	2
Willow	4	7	2	5	0	90	0
Poplar	5	1	3	4	1	0	80
	Precision	92.07	90.82	92.14	86.78	85.87	91.05

The results of the various methods used to generate tree species classification maps in three regions are presented in Figure 7. The proposed method employs the fusion map convolution method using HJ-1A and Sentinel-2 data, achieving an OA of 0.88, an AA of 0.85, and a Kappa of 0.82 in the consistent areas. The proposed model outperforms other methods in identifying Spruce and Larch, which have commercial value due to their rarity. The SpectralNet method performs poorly, followed by DMULN, FuNet, and MFDF. Other methods have blurred edges, low recognition rates, and a high rate of misclassification and fragmentation. Overall, the proposed method accurately identifies all of the tree species and yields favorable results, surpassing the compared methods. The superior performance is attributed to the proposed strategy based on depth hypergraph convolution fusion and hyperedge convolution fusion.

**Figure 7.** Tree species classification map of multi-source datasets in three datasets. (1) SpectralNet, (2) DMULN, (3) FuNet.

### 3.3. Parameter Analysis

In this section, we utilize three tree species datasets from Section 2.2 to analyze the key parameters that affect classification performance. These parameters include the labeling ratio (partial labeling of the total datasets), K value, and depth. We conduct tests and analyses to examine the impact of these parameters on classification performance. Figure 8 displays the tree species classification accuracy of the five models with varying label rates. The classification accuracy of all five models increases as the label rates increase. However, the proposed model achieves desirable accuracy and outperforms the other methods significantly.

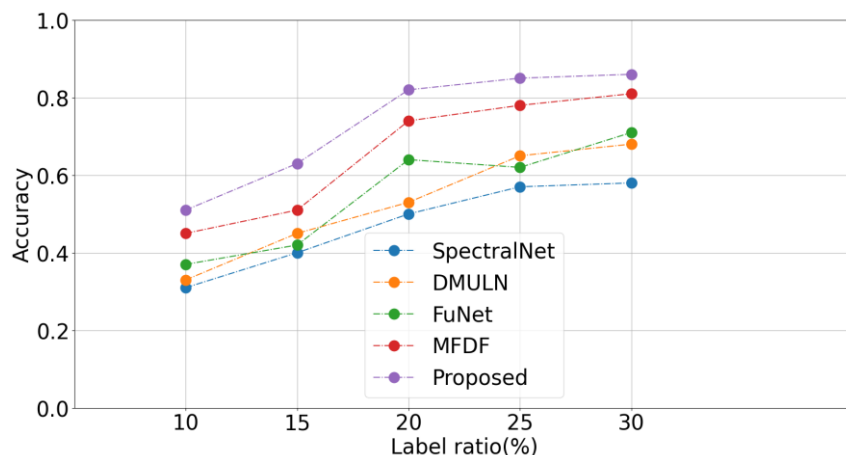


Figure 8. The accuracy of tree species classification with different label rates.

To verify the robustness of the method, it is advisable to strive for consistency in the selection of  $k$  across different modal features while minimizing any potential impact on accuracy. This approach allows for a thorough assessment of the method’s resilience, particularly in terms of its ability to handle variations across modalities. By maintaining a consistent value of  $K$ , the performance of the method can be effectively evaluated, and the robustness can be determined in terms of achieving accurate results while considering the unique characteristics of each modality. Figure 9 illustrates the classification accuracy of three models for different  $K$  values ( $K \in \{5, 10, 15, 20, 25, 30\}$ ). As only three of the compared models have a  $K$  value parameter, the results show that the accuracy of the three methods varies with  $K$ . The accuracy tends to increase as the  $K$  value is set between 5 and 15. However, when  $K$  is set between 15 and 30, the accuracy starts to decrease.

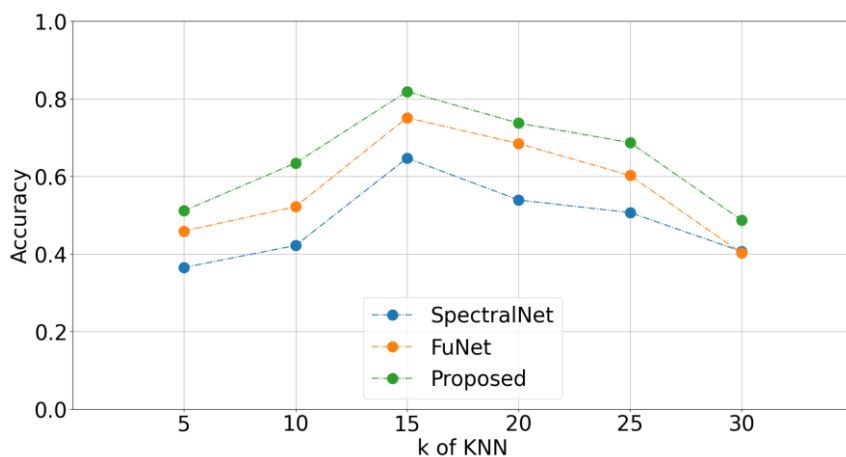
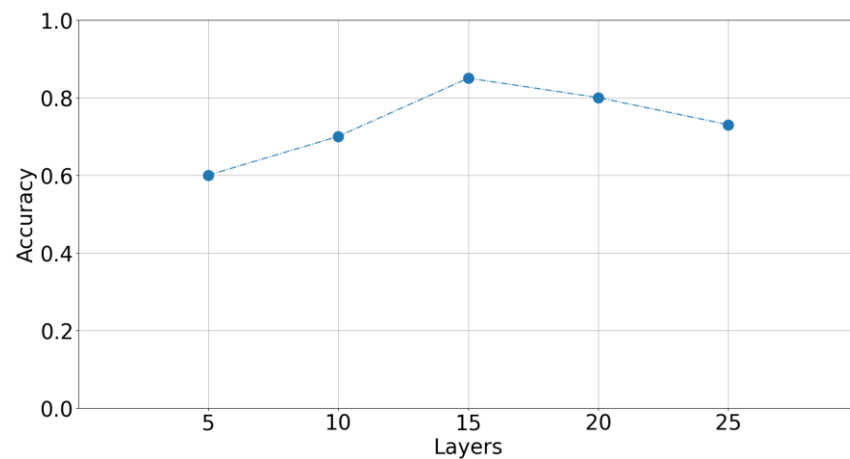


Figure 9. The accuracy of tree species classification with  $k$  Values ( $k$  value in KNN).

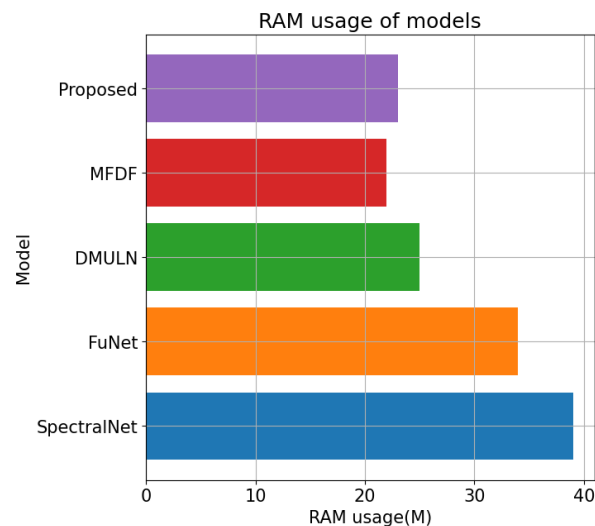
The experimental results demonstrate that the proposed method achieves the best performance when  $K$  is set to 15. These findings lead to two main conclusions: (1) A small  $K$  value may fail to capture the neighborhood of the data, while an increasing  $K$  value could result in incorrect neighborhood samples that render the relationship between samples less discriminative. (2) The proposed fusion learning method is sensitive to the choice of the  $K$  value.

To investigate the influence of the depth of the proposed model, we set the range of the DHCN layers to  $\{5, 10, 15, 20, 25\}$ . Figure 10 demonstrates that the accuracy of the classification result with DHCN is highest when the  $K$  value is set to 15, and the method is not extremely sensitive to the number of layers. However, as the number of layers increases beyond 15, performance slightly degrades with excessive smooth curve.

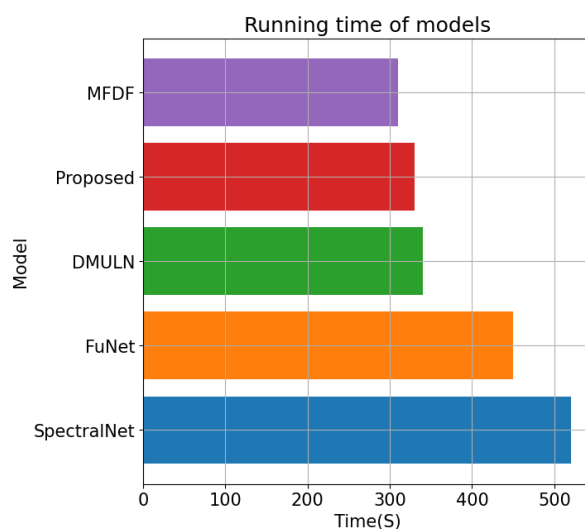


**Figure 10.** The accuracy of tree species classification with different depths of the proposed model.

As shown in Figures 9 and 10, the main factors that affect the computation time are the complexity of the datasets, the number of categories, the number of spectral channels, and the image size. The computation time is influenced by various factors, including image size, data complexity, and the model parameters. Figures 11 and 12 illustrate the RAM usage and running time of different models in the classification experiments. Larger images and more complex datasets necessitate increased memory and computation time. Comparatively, the GCN-based methods require more memory and time compared to the CNN-based method, primarily due to the time-consuming computation of the adjacency matrix. However, the proposed model, with its fusion graph structure, outperforms other GCN methods in terms of speed. This is achieved by eliminating the utilization of ineffective features, resulting in improved overall system efficiency. By removing irrelevant or redundant features from the data, the model can concentrate on the most informative aspects of the input, leading to enhanced performance and faster computation times.



**Figure 11.** RAM usage of different models.



**Figure 12.** Running time of different models.

#### 4. Discussion

The proposed model first performs typical association analysis on the two data sources used as the input, maximizes the correlation of multi-source data, performs convolution calculation on the generated vector to extract features, and then fuses the nodes of its graph structure. This process not only reduces redundant information but also strengthens the effective features. Finally, hyperedge convolution is introduced into the graph convolution training process to adaptively mine the relationship between the representative descriptors and fully integrate the node and attribute features.

The SpectralNet model exhibits significant statistical advantages when using the spectral clustering method as it overcomes the scalability and generalization of the spectral embedding. However, in our experiment, the SpectralNet method displayed severe shortcomings in terms of coniferous forest species classification, with almost no Willow identified and other tree species misrecognized as Larch. The DMULN method [32] utilizes an encoder–decoder network to input the features related to the two data sources separately. Its recognition ability in terms of Mongolian, Poplar, and Willow is better than the other three methods, owing to the benefits of deep multi-view learning and view pooling. The DMULN method proved to be superior to SpectralNet in terms of tree species classification performance as it can learn both spectral and spatial modal features simultaneously during the experiment. However, it is inferior to FuNet and MFDF. FuNet utilizes mini batches of non-European features in graph convolution processing as well as European features CNN processing, which are then fused together. This approach has demonstrated impressive performance in a single hyperspectral data source. However, in the present experiment, FuNet did not perform as well when using multi-source tree species datasets. MFDF, which is based on Gabor wavelet feature representation, utilizes a two-dimensional Gabor filter [31], making it more suitable for feature extraction when using multi-source datasets. As we use Sentinel-2 data as the multispectral data, which is more effective when extracting spatial features, the Gabor extraction of spatial features is slightly worse, resulting in a lower OA in terms of tree species classification compared to the proposed method [34]. However, its recognition rate for Spruce is significantly lower than the proposed method, which uses a graph structure to represent the higher-order features. Hyperedge learning also integrates the features of the graph structure from the two data sources, thus improving the recognition of Spruce and the overall recognition rate more accurately [33]. The proposed hypergraph fusion structure can transfer the complex high-order correlation between HSIs and MSIs, and better represent the underlying data interrelation between them than the basic graph structure. Additionally, the proposed method has the advantage of fusing multimodal information into the same data structure with flexible hyperedges, owing to

the existence of multimodal features. Through hypergraph fusion and hyperedge convolution fusion, the multi-source graph convolution model proposed significantly reduces computation time while improving learning efficiency.

Our model outperforms the compared models in the classification of six tree species, yielding a higher AA. The proposed model has several advantages:

- (1) The model utilizes multiple graph learning and multi-source fusion, where each graph provides complementary information that is unique from the other graphs. By removing the noise hyperedges present in tiny graphs, the model improves tree species classification performance.
- (2) Multi-graph learning is proven to be feasible in tree species classification, and our model considers both the global and local features of multi-source data simultaneously with regularization.
- (3) Compared to other models, our proposed method is more effective in classifying tree species by using the fusion of multi-source data. The utilization of multimodal graph learning enhances the effectiveness of the classification process.

## 5. Conclusions

In this paper, we proposed a novel model for tree species classification by designing a multi-source fusion graph neural network. The proposed model first calculates the pixel-based correlation between HSIs and MSIs, generating two types of hypergraph structures. Both the HSI graph structure and MSI graph structure are saved in each initial graph and fused with each other in the hyperedge learning process. The proposed model fuses the two data sources twice, capturing the global graph from the low-dimensional space of the original high-dimensional data. We propose a new fusion method that combines complementary and common information to correctly capture the graph structure inherent in the data. We evaluated our method using a tree species dataset and compared it with state-of-the-art approaches. The experimental results show that the proposed method is effective in improving the accuracy of tree species classification.

In the future, our research aims to investigate multi-source feature fusion algorithms based on self-supervised learning methods. Additionally, we intend to explore tree species classification in higher-resolution remote sensing images. These endeavors will further enhance the accuracy and capabilities of our classification models, enabling us to tackle more complex and detailed datasets in the field of tree species classification.

**Author Contributions:** X.W.: conceptualization, methodology, software, data curation, funding acquisition, writing—original draft—writing—review and editing; J.W.: writing—review and editing—supervision; Z.L.: supervision; N.Y.: writing—review and editing—supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Basic Scientific Research Project of Heilongjiang Provincial Universities (Grant number 145109219).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Hao, S.; Zhou, Y.; Guo, Y. A brief survey on semantic segmentation with deep learning. *Neurocomputing* **2020**, *406*, 302–321. [[CrossRef](#)]
2. Ran, Y.H.; Li, X.; Lu, L.; Li, Z.Y. Large-scale land cover mapping with the integration of multi-source information based on the Dempster–Shafer theory. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 169–191. [[CrossRef](#)]
3. Hua, T.; Zhao, W.; Liu, Y.; Wang, S.; Yang, S. Spatial consistency assessments for global land-cover datasets: A comparison among GLC2000, CCI LC, MCD12, GLOBCOVER and GLCNMO. *Remote Sens.* **2018**, *10*, 1846. [[CrossRef](#)]



4. Hou, W.; Hou, X. Data fusion and accuracy analysis of multi-source land use/land cover datasets along coastal areas of the Maritime Silk Road. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 557. [CrossRef]
5. Liu, K.; Xu, E. Fusion and Correction of Multi-Source Land Cover Products Based on Spatial Detection and Uncertainty Reasoning Methods in Central Asia. *Remote Sens.* **2021**, *13*, 244. [CrossRef]
6. Liu, X.; Jiao, L.; Li, L.; Cheng, L.; Liu, F.; Yang, S.; Hou, B. Deep multiview union learning network for multisource image classification. *IEEE Trans. Cybern.* **2020**, *52*, 4534–4546. [CrossRef]
7. Ma, Z.; Jiang, Z.; Zhang, H. Hyperspectral Image Classification Using Feature Fusion Hypergraph Convolution Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5517314. [CrossRef]
8. Mu, C.; Dong, Z.; Liu, Y. A Two-Branch Convolutional Neural Network Based on Multi-Spectral Entropy Rate Superpixel Segmentation for Hyperspectral Image Classification. *Remote Sens.* **2022**, *14*, 1569. [CrossRef]
9. See, L.M.; Fritz, S. A method to compare and improve land cover datasets: Application to the GLC-2000 and MODIS land cover products. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1740–1746. [CrossRef]
10. Yu, Y.; Liang, S.; Samali, B.; Nguyen, T.N.; Zhai, C.; Li, J.; Xie, X. Torsional capacity evaluation of RC beams using an improved bird swarm algorithm optimised 2D convolutional neural network. *Eng. Struct.* **2022**, *273*, 115066. [CrossRef]
11. Yu, Y.; Li, J.; Li, J.; Xia, Y.; Ding, Z.; Samali, B. Automated damage diagnosis of concrete jack arch beam using optimized deep stacked autoencoders and multi-sensor fusion. *Dev. Built Environ.* **2023**, *14*, 100128. [CrossRef]
12. Chao, G.; Sun, S. Semi-supervised multi-view maximum entropy discrimination with expectation laplacian regularization. *Inf. Fusion* **2019**, *45*, 296–306. [CrossRef]
13. Zhu, P.; Hu, Q.; Hu, Q.; Zhang, C.; Feng, Z. Multi-view label embedding. *Pattern Recognit.* **2018**, *84*, 126–135. [CrossRef]
14. Qin, A.; Shang, Z.; Tian, J.; Wang, Y.; Zhang, T.; Tang, Y.Y. Spectral-spatial graph convolutional networks for semisupervised hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 241–245. [CrossRef]
15. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional Neural Networks on Graphs with Gated Localized Spectral Filtering. In Proceedings of the Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 3844–3852.
16. Tang, C.; Chen, J.; Liu, X.; Li, M.; Wang, P.; Wang, M.; Lu, P. Consensus learning guided multi-view unsupervised feature selection. *Knowl.-Based Syst.* **2018**, *160*, 49–60. [CrossRef]
17. Ding, S.; Cong, L.; Hu, Q.; Jia, H.; Shi, Z. A multiway p-spectral clustering algorithm. *Knowl.-Based Syst.* **2019**, *164*, 371–377. [CrossRef]
18. Li, S.; Liu, H.; Tao, Z.; Fu, Y. Multi-view graph learning with adaptive label propagation, in: Big Data (Big Data). In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 110–115.
19. Xi, B.; Li, J.; Li, Y.; Song, R.; Xiao, Y.; Du, Q.; Chanussot, J. Semisupervised cross-scale graph prototypical network for hyperspectral image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**. Available online: <https://ieeexplore.ieee.org/abstract/document/9740412> (accessed on 17 April 2023). [CrossRef] [PubMed]
20. Wang, L.; Fan, W.Y. Identification of forest dominant tree species group based on hyperspectral remote sensing data. *J. Northeast. For. Univ.* **2015**, *43*, 134–137.
21. Feng, Y.; You, H.; Zhang, Z.; Ji, R.; Gao, Y. Hypergraph neural networks. In Proceedings of the AAAI conference on artificial intelligence, Hilton, HI, USA, 27 January–1 February 2019; Volume 33, pp. 3558–3565.
22. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
23. Wang, X.; Ren, H.; Wang, A. Smish: A novel activation function for deep learning methods. *Electronics* **2022**, *11*, 540. [CrossRef]
24. Zhou, D.; Huang, J.; Schölkopf, B. Learning with Hypergraphs: Clustering, Classification, and Embedding. In Proceedings of the Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 1601–1608.
25. Wang, X.; Ren, H. DBMF: A Novel Method for Tree Species Fusion Classification Based on Multi-Source Images. *Forests* **2021**, *13*, 33. [CrossRef]
26. Xue, X.; Nie, F.; Li, Z.; Wang, S.; Li, X.; Yao, M. A multiview learning framework with a linear computational cost. *IEEE Trans. Cybern.* **2017**, *48*, 2416–2425.
27. Xie, G.S.; Zhang, X.Y.; Yan, S.; Liu, C.L. Hybrid CNN and dictionary-based models for scene recognition and domain adaptation. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *27*, 1263–1274. [CrossRef]
28. Xie, G.S.; Zhang, X.Y.; Shu, X.; Yan, S.; Liu, C.L. Task-driven feature pooling for image classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1179–1187.
29. Pei, J.; Huang, Y.; Sun, Z.; Zhang, Y.; Yang, J.; Yeo, T.S. Multiview synthetic aperture radar automatic target recognition optimization: Modeling and implementation. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6425–6439. [CrossRef]
30. Hardoon, D.R.; Szedmak, S.; Shawe-Taylor, J. Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.* **2004**, *16*, 2639–2664. [CrossRef] [PubMed]
31. Shaham, U.; Stanton, K.; Li, H.; Nadler, B.; Basri, R.; Kluger, Y. Spectralnet: Spectral clustering using deep neural networks. *arXiv* **2018**, arXiv:1801.01587.
32. 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]

33. Jia, S.; Zhan, Z.; Zhang, M.; Xu, M.; Huang, Q.; Zhou, J.; Jia, X. Multiple feature-based superpixel-level decision fusion for hyperspectral and LiDAR data classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 1437–1452. [[CrossRef](#)]
34. Wang, X.; Yang, N.; Liu, E.; Gu, W.; Zhang, J.; Zhao, S.; Sun, G.; Wang, J. Tree Species Classification Based on Self-Supervised Learning with Multisource Remote Sensing Images. *Appl. Sci.* **2023**, *13*, 1928. [[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.