TSCMDL: Multimodal Deep Learning Framework for Classifying Tree Species Using Fusion of 2-D and 3-D Features

Bingjie Liu, *Graduate Student Member, IEEE*, Yuanshuo Hao, Huaguo Huang, Shuxin Chen, Zengyuan Li, *Member, IEEE*, Erxue Chen, *Member, IEEE*, Xin Tian, and Min Ren

***Abstract*— Accurate tree species information is a prerequisite for forest resource management. Combining light detection and ranging (LiDAR) and image data is one main method of tree species classification. Traditional machine learning methods rely on expert knowledge to calculate a large number of feature parameters. Deep learning technology can directly use the original image and point cloud data to classify tree species. However, data with different patterns require the use of different types of deep learning methods. In this study, a tree species classification multimodal deep learning (TSCMDL) that fuses 2-D and 3-D features was constructed and then used to combine data from multiple sources for tree species classification. This framework uses an improved version of the PointMLP model as its backbone network and uses ResNet50 and PointMLP networks to extract the image features and point cloud features, respectively. The proposed framework was tested using unmanned aerial vehicle LiDAR (UAV LiDAR) data and red, green, blue (RGB) orthophotos. The results showed that the accuracy of the tree species classification using the TSCMDL framework was 98.52%, which was 4.02% higher than that based on point cloud features only. In addition, when the same hyperparameters were used for training the model, the efficiency of the model training was not significantly lower than for models based on point cloud features only. The proposed multimodal deep learning framework extracts features directly from the original data and integrates them effectively, thus avoiding manual feature screening and achieving more accurate classification. The feature extraction network used in the TSCMDL framework can be replaced by other suitable frameworks and has strong application potential.**

***Index Terms*— Deep learning, light detection and ranging (LiDAR), multimodal, point cloud, tree species classification.**

Manuscript received 18 January 2023; revised 9 March 2023 and 24 March 2023; accepted 6 April 2023. Date of publication 10 April 2023; date of current version 26 April 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFE0117700, in part by the National Science and Technology Major Project of China's High Resolution Earth Observation System under Grant 21-Y20B01-9001-19/22, and in part by the National Natural Science Foundation of China under Grant 42130111 and Grant 41871279. *(Corresponding author: Xin Tian.)*

Bingjie Liu and Huaguo Huang are with the Research Center of Forest Management Engineering of State Forestry and Grassland Administration, Beijing Forestry University, Beijing 100083, China (e-mail: liubj@geoe.ac.cn; huaguo_huang@bjfu.edu.cn).

Yuanshuo Hao is with the Key Laboratory of Sustainable Forest Ecosystem Management Ministry of Education, School of Forestry, Northeast Forestry University, Harbin, Heilongjiang 150040, China (e-mail: haoyuanshuo@nefu.edu.cn).

Shuxin Chen, Zengyuan Li, Erxue Chen, and Xin Tian are with the Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China (e-mail: chensx@ifrit.ac.cn; lizengyuan@ifrit.ac.cn; chenerx@ifrit.ac.cn; tianxin@ifrit.ac.cn).

Min Ren is with the China Mobile Group Shanxi Design Institute Company, Ltd., Taiyuan 030032, China (e-mail: renmin@cmdi.chinamobile.com).

Digital Object Identifier 10.1109/TGRS.2023.3266057

I. Introduction

Trees are the basic unit used in forest resource surveys, and accurate tree species information is a prerequisite for effective forest resource management. Tree species information is also a key factor in the calculation of various parameters related to forest carbon content. The identification and classification of tree species is mainly based on field surveys and the identification of trunks, branches, and leaves. Field surveys require a great deal of manpower and material resources, and a lot of time is needed to acquire the data. This makes it difficult to conduct highly detailed surveys within a short period of time. However, the development of remote sensing technology has made it easier to acquire tree species information over large areas more rapidly.

Light detection and ranging (LiDAR) technology can be used to obtain 3-D tree structure information quickly and accurately, and highly accurate tree classification results can be obtained based on terrestrial laser scanning (TLS) and mobile laser scanning data [1], [2], [3], [4], [5]. However, the collection of ground-based LiDAR data is time-consuming and labor-intensive and is difficult to obtain over large areas. Airborne laser scanning (ALS) and unmanned aerial vehicle LiDAR (UAV LiDAR) can be used to quickly obtain point cloud data for a wide range of forest scenes but cannot be used to collect detailed information on tree structure below the canopy [6], [7]. Therefore, when using ALS or UAV LiDAR data for forest species classification, additional data, such as orthophotos, multispectral imagery, or hyperspectral imagery, are required.

In most studies, features are extracted from ALS or UAV LiDAR data and red, green, blue (RGB), multispectral, or hyperspectral imagery. Feature screening is then used to select the appropriate variables after the LiDAR data have been segmented into individual trees. Finally, machine learning classification methods, such as the support vector machines [8], [9], [10], [11], [12], [13], random forest (RF) [14], [15], [16], neural networks [17], and k-nearest neighbor (k-NN) [11], are used to realize the tree species classification. In some studies, multispectral imagery and point clouds generated by digital photogrammetry have also been used for feature syndication [13], [18]. Data fusion-based approaches to tree species classification include the combination of features obtained from different data sources [9], [19]. The results of a series of studies show that when the high-level features extracted by ALS and UAV LiDAR point clouds are used for tree species classification, supplementing these features by the spectral and textural features provided by RGB, multispectral, or hyperspectral images further improves the classification accuracy achieved using machine learning methods [19].

When traditional machine learning methods are used for tree species classification, the aim is to extract and select a large number of key diagnostic features from a large amount of LiDAR data [20]. This feature extraction requires a large amount of expert knowledge, which limits the applicability of the method. In recent years, as deep learning technology has developed, the application of this technology to tree species classification has received increased attention, and a range of deep learning methods for tree species classification based on different types of data has been developed. The deep learning methods used for tree species classification can be divided into image deep learning models and point cloud deep learning models based on these different data types.

Deep learning technology uses deep neural networks to efficiently extract deep classification features from raw images, and this has been demonstrated to be a powerful method of classifying remote sensing imagery [21]. Deep learning models can also perform end-to-end classification tasks, reducing the need for expert knowledge. Regional-scale tree species classification and the mapping of RGB, multispectral, and hyperspectral images using image deep learning models have also been the subject of some studies. Zhang et al. [22] proposed an improved 3-D convolutional neural network (3-D-CNN) tree species classification model for use with airborne multispectral images. Using raw data as the input, this model could extract both spectral and spatial features. Fricker et al. [23] used airborne hyperspectral images and constructed a CNN-based tree species classification model. Torres et al. [24] used UAV RGB imagery to evaluate the effectiveness of five deep learning methods for segmenting individual tree species. Natesan et al. [25] used the ResNet model to identify two species of pine trees in complex forest areas. Natesan et al. [26] proposed a deep convolutional neural network that used multitemporal RGB images to identify tree species in orthophotos generated from images taken under different light conditions, at different viewing angles, and in different seasons.

Methods that use point cloud data for tree species classification can be divided into two types. The first type involves the conversion of point cloud data into raster imagery through various rotations and projections; the tree species classification is then performed by an image-based convolutional neural network [27]. Guan et al. [28] converted the point clouds corresponding to individual trees into a waveform map and used deep Boltzmann machines to extract tree features. Hamraz et al. [29] converted ALS data into digital surface model (DSM) data and then used CNN methods to extract features from DSM data for tree species classification. Marinelli et al. [30] decomposed the point cloud corresponding to individual trees into eight corner sectors along the vertical axis, thus generating a 2-D view of the vertical section of each sector to capture structural information. A multiview convolutional neural network was used to automatically extract the category features for each sector.

Methods that belong to the second type use a point-based deep learning model to extract features directly from the original point cloud. The development of point cloud deep learning networks reached a new stage with the PointNet [31] and PointNet++ [32] models. Since then, the field of computer vision has exploded, with more deep learning models for point cloud classification [33]. Over the past three years in particular, researchers have begun to focus on the use of point cloud deep learning models to identify tree species based on the direct classification of LiDAR point cloud data. Xi et al. [34] evaluated the use of seven deep learning models and six machine learning models for tree species classification based on TLS point clouds. Of the models tested, the PointNet++ model was found to produce the most accurate classification. This study provides a new idea for the direct use of LiDAR point clouds for tree species classification. Other studies have demonstrated that the PointNet model has good tree species classification capabilities using even small amounts of data [35]. Chen et al. [36] proposed a point cloud tree species classification network that could overcome some of the difficulties faced when using deep learning methods to classify tree species from laser data. The results showed that better classification results could be obtained using TLS data than UAV LiDAR data. Liu et al. [37] proposed a network structure similar to the PointNet model for the classification of birch and larch species. The classification accuracy of the LayerNet model proposed by Liu et al. [37] was found to be 88.8%, which is higher than that of machine learning methods, such as k-NN and RF. Lv et al. [38] proposed a convex hull-based descriptor to represent the characteristics of ALS tree point clouds. When the PointNet++ model was used to fuse the feature descriptor, the accuracy of the tree species classification was 82.0%. Seidel et al. [39] obtained classification accuracy higher than that obtained using the PointNet model by generating projected images of individual trees at ten different angles. However, in another study, an even higher accuracy was obtained with the same dataset using the PointNet++ model [40]. Liu et al. [41], [42] carried out tree species classification by applying multiple point cloud deep learning models based on a multilayer perceptron (MLP), convolution, and graph attention mechanisms to ground-based backpack LiDAR data and explored different methods of preprocessing point cloud data of individual trees.

From the results of previous studies on tree species classification using point cloud deep learning models, it is clear that reasonably accurate results can be achieved using ground-based LiDAR data. However, the accuracy of tree species classification based on the use of LiDAR data acquired by airborne platforms needs to be improved, and as explained above, classification based on this type of data requires the use of additional datasets. In some recent studies, ALS or UAV

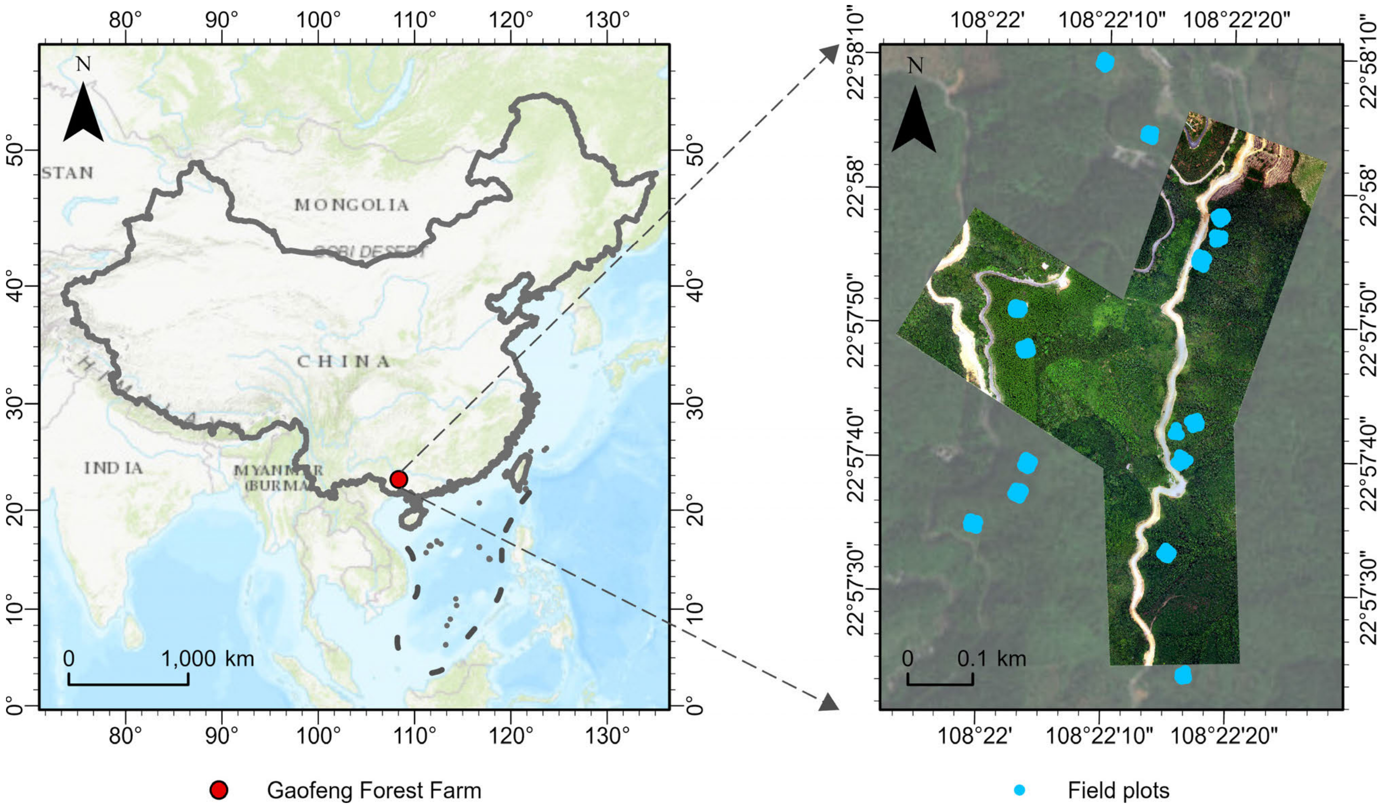

Fig. 1. Location of the study area and the distribution of study plots. The map on the right is based on an orthophoto acquired during the study.

LiDAR data have been combined with RGB, multispectral, or hyperspectral data to classify tree species using deep learning techniques. LiDAR data are used for the extraction of individual trees from canopy data by projecting LiDAR data into raster data and then combining them with image data; the classification of the tree species can then be performed using image-based deep learning methods [5], [43], [44], [45], [46]. Briechle et al. [47] proposed a dual-CNN model for tree species classification based on multisource data. The ALS point cloud data were projected into 12 raster images at different angles, which were then extracted using a separate ResNet18 network. Using these methods based on multiple views, the amount of detail in the representation of the point cloud is reduced because the projection causes a loss of information [48]. The classification of aggregations of features obtained using multiple views is also challenging [33].

In summary, the classification of tree species based on UAV LiDAR data alone produces less accurate results, and the current research using multiple data sources for tree species classification considers data fusion, which limits classification accuracy. The use of multimodal deep learning technology in computer vision is increasing [49]. However, studies on the construction of multimodal deep learning frameworks for tree species classification are limited. Therefore, it is necessary to study the construction of multimodal learning models for tree species classification. Herein, a tree species classification multimodal deep learning (TSCMDL) fused with 2-D and 3-D features is proposed for tree species classification. The framework has fused the feature extraction modules of ResNet50 [50] and PointMLP [51] models. Furthermore, the model can extract features from LiDAR and image data very conveniently and perform the classification task directly end-to-end. Hence, it avoids all kinds of complex processes that require raster conversion of point clouds in the existing studies. Moreover, the feature extraction module of this framework can be modified and further extended and has the advantage of high implementation efficiency and high classification accuracy for tree species classification tasks.

II. Data Collection and Processing

A. Data Description

The study area was located at Gaofeng Forest Farm, near the city of Nanning, Guangxi, China (108°20′ E–108°32′ E; 22°56′ N–23°4′ N). This farm is the largest state-owned forest farm in Guangxi province. The farm is located in the subtropical monsoon climate zone; the average annual temperature at the location is 21 °C, and the average annual precipitation is in the range of 1200–1500 mm. The main tree species used for research are eucalyptus (*Eucalyptus robusta* Smith) and Chinese fir [*Cunninghamia lanceolata* (Lamb.) Hook.]. The location of the study area and the distribution of the ground survey plots are shown in Fig. 1.

A survey of the field plots was conducted in April 2021. The species, diameter at breast height, height, location, and other data were recorded for the individual trees in each plot. Most of the plots are single species, while some others are mixed. LiDAR data of the study area were collected in May

2020 using a GV2000 multirotor UAV equipped with a RIEGL VUX-1LR sensor. The drone was flown at an altitude of 250 m and a speed of 7 m/s; both the overlap rate in the flight direction and the side overlap rate were 90%. The data were acquired in strips with a width of 598.9 m. The LiDAR system had a scanning frequency of 200 kHz, a pulse emission frequency of 820 kHz, and an accuracy of 15 mm. At the same time, orthophotos were acquired using a DJI-PHANTOM 4-RTK drone, which was equipped with FC6310R sensors. The drone was flown at a height of 250 m at a speed of 10 m/s; in this case, the overlap rate in the flight direction and the side overlap rate were approximately 80%. The image sensor had a focal length of 9 mm, the exposure time was 1/320 s, the aperture was f/5.6, and the ground sampling distance was 6.85 cm.

The images acquired by the drone were aligned, stitched, and resampled to produce orthophotos with a spatial resolution of 0.1 m. One of these orthophotos is shown in Fig. 1.

B. Data Preprocessing

The coordinates of the raw LiDAR data were calculated, and the image strips were aligned to produce a dataset that covered the whole study area. The point cloud data were denoised using the height threshold method and the local distribution algorithm [52]. Ground points were filtered and removed from the original point cloud using the cloth simulation filtering algorithm [53]. Next, the hierarchical region-merging (HRM) algorithm [54] was used to complete the individual tree segmentation, and the results were manually checked and edited. Finally, precise point clouds of individual trees were obtained; these were used in the subsequent tree species identification. The canopy boundary for each tree was extracted using the LAS2shp tool provided by LAStools. These boundaries were used to crop the orthophotos to obtain image data that matched each individual tree. The original RGB bands were retained in the final orthophotos.

C. Create Training and Test Dataset

Considering the relatively low point count of UAV LiDAR-based partial individual tree point clouds, each of the point clouds for individual trees was downsampled to 1024 points. This was achieved using the non-uniform grid and farthest point sampling (NGFPS) point cloud downsampling strategy [42] for obtaining sample data that could be used as an input to the point cloud deep learning model. NGFPS is a point cloud downsampling strategy that combines nonuniform sampling with farthest point sampling (FPS). NGFPS retains the advantages of nonuniform sampling, thereby allowing the downsampled point cloud to preserve more accurate details of trees. The point clouds were then normalized, and the coordinates of each point were normalized into a unit sphere [41]. The purpose of the normalization was to improve the robustness of the model.

The LiDAR image and point cloud corresponding to each tree were named. The complete sample set consisted of UAV LiDAR images and orthophotos of 300 eucalyptus and 300 Chinese fir trees. In the subsequent experiment, 80% of the sample data were selected for the model training and 20% were used for testing the model.

D. Experimental Equipment

All deep learning models were developed based on the PyTorch (1.11.0) framework. The graphics processing unit (GPU) on the computer was an NVIDIA GeForce RTX 3070 (8-GB VRAM) and CUDA version 11.4. The CPU was Intel[1] Core[2] i7-10700KF (3.80 GHz). The memory of the device was 32 GB. The operating system of the computer was Windows 10 22H2.

III. Architecture of the Proposed Framework

To directly use raw information from the images and point clouds, we used early feature-fusion strategies to build multimodal deep learning classification networks. Early fusion refers to fusion that is performed at the feature level and is the most widely used fusion strategy. In early fusion, the features extracted from the data of different models are fused before any analysis is performed. The proposed TSCMDL framework combines LiDAR classification features with image data and is also based on feature-level fusion. PointMLP is used as the backbone for multimodal model building, and a PointMLP feature extraction module is used to extract the category features of the point cloud. The category features of the images are extracted using the ResNet50 feature extraction network. Fig. 2 shows the overall structure of the TSCMDL model, including the data input, feature extraction, feature fusion, and classification elements.

The PointMLP network framework used for the extraction of the LiDAR classification features is compact. This network uses pointwise MLP, which are appropriate for the characteristics of point cloud data displacement. By merging residual connections, the network structure can be extended to one with many layers to allow the extraction of the deep features of the point cloud.

Compared to other CNN architectures, ResNet50 has several advantages. One advantage is that it is able to train very deep networks and does not experience gradient vanishing problems. Another advantage is that it has a relatively small number of parameters compared to other models, which makes training faster and deployment easier. ResNet50 has been widely used in image classification tasks and has achieved the state-of-the-art results, as measured by several benchmarks. ResNet50 is also often used as a base model for transfer learning, where pretrained models are fine-tuned on new datasets before being applied to other tasks.

A. Point Cloud Feature Extraction Network

The main advantage of point-based deep learning models is that they can directly process point cloud data: the structural information and spatial relationships are preserved, thereby improving the accuracy of the model and making it more generally applicable. PointMLP is a simple and effective MLP-based point cloud analysis network that follows the

[1]Trademarked.
[2]Registered trademark.

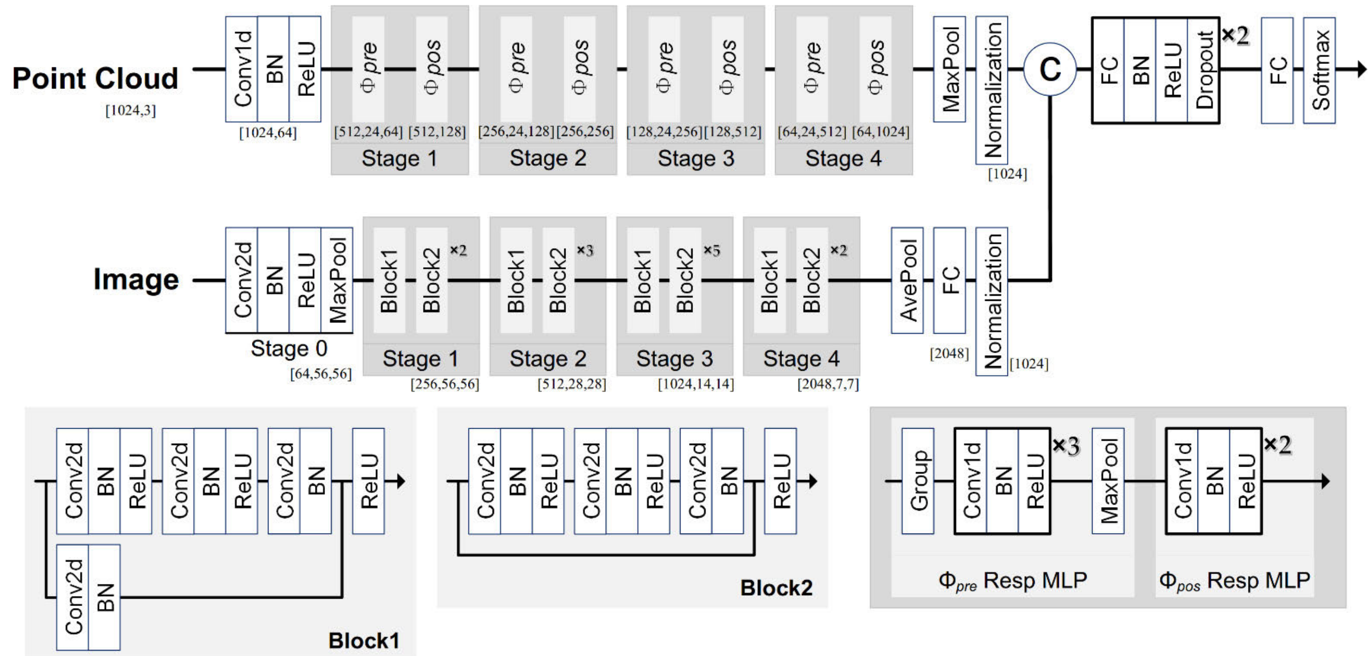

Fig. 2. Diagram showing the structure of the TSCMDL framework.

design philosophy of the PointNet++ model; no complex or heavy operations are included in the network. The "group" layer is used to group points together to reduce model complexity. In each stage, the FPS method is used to select $N/2$ points from the N points; then, the k-NN algorithm is used to calculate K ($K = 24$) nearest points for each selected point from the input point set N. Each of the 24 nearest points constitutes a group. The input data dimension of Stage1 is [1024, 64], the grouped data dimension is [512, 24, 64], and the output data dimension is [512, 128] after a series of convolution operations. Fig. 2 shows the corresponding data dimension for each stage. The core structure of the network can be expressed as follows:

$$g_i = \Phi_{\text{pos}}\big(\mathcal{A}\big(\Phi_{\text{pre}}\big(f_{i,j}\big), \mid j = 1, \ldots, K\big)\big) \tag{1}$$

where Φ_{pre} and Φ_{pos} represent the residual blocks of the MLP, $\Phi_{\text{pre}}(\cdot)$ is used to learn the shared weights from a local region, and $\Phi_{\text{pos}}(\cdot)$ is used to aggregate features. The aggregate function $\mathcal{A}(\cdot)$ represents the maximum pooling. Unlike PointNet++, PointMLP uses the k-NN algorithm to aggregate adjacent points, and K is set to 24 by default. $f_{i,j}$ represents a block used for a geometric affine operation

$$\{f_{i,j}\} = \alpha \odot \frac{\{f_{i,j}\} - f_i}{\sigma + \epsilon} + \beta \tag{2}$$

$$\sigma = \sqrt{\frac{1}{k \times n \times d} \sum_{i=1}^{n} \sum_{j=1}^{k} \big(f_{i,j} - f_i\big)^2} \tag{3}$$

where α and β are the learnable parameters, $\odot$ indicates the Hadamard product, and $\epsilon = 1e^{-5}$. The set of local points processed by (2) satisfies the normal distribution while maintaining the original geometric properties; this improves the robustness of the model. Equation (1) represents Stage 1 of the point cloud feature extraction network illustrated in Fig. 2; in the complete PointMLP feature extraction network, needs to repeat (1) four times.

In this study, before the feature extraction was conducted, the following three methods were used to augment the point cloud data: 0.0%–87.5% of the points were randomly deleted, points were randomly moved by −0.2 to 0.2 units, and the point cloud was jittered by a Gaussian noise with zero mean and 0.02 standard deviation.

B. Image Feature Extraction Network

The ResNet50 network can be considered to consist of five stages. Stage 0 consists of a set of convolutions, batch normalization, rectified linear unit (ReLU) activation functions, and max pooling. The remaining four stages are each made up of different residual blocks. These residual blocks can be categorized as either Block1 or Block2 (see Fig. 2). Block1 and Block2 represent the convolutional and identity blocks, respectively. As can be seen from Fig. 2, the short-cut path in Block1 has one more convolutional operation than that of Block2, thus achieving the effect of dimension modification. The input and output dimensions of the Block2 are the same, so different numbers of Block2 are used to increase the depth of the network. The residual blocks can be described by the following equation:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x} \tag{4}$$

where $\mathbf{x}$ represents the network input, $\mathbf{y}$ represents the output, and $\mathcal{F}(\cdot)$ represents the residual mapping that needs to be learned. $F(\cdot)$ in both Block1 and Block2 represents three-layer convolution operations. $\mathbf{x}$ in Block1 represents a layer of convolution operations. Stages 1–4 form the backbone of the ResNet50 model used for feature extraction; ResNet models belonging to different series differ in the way these stages are constructed.

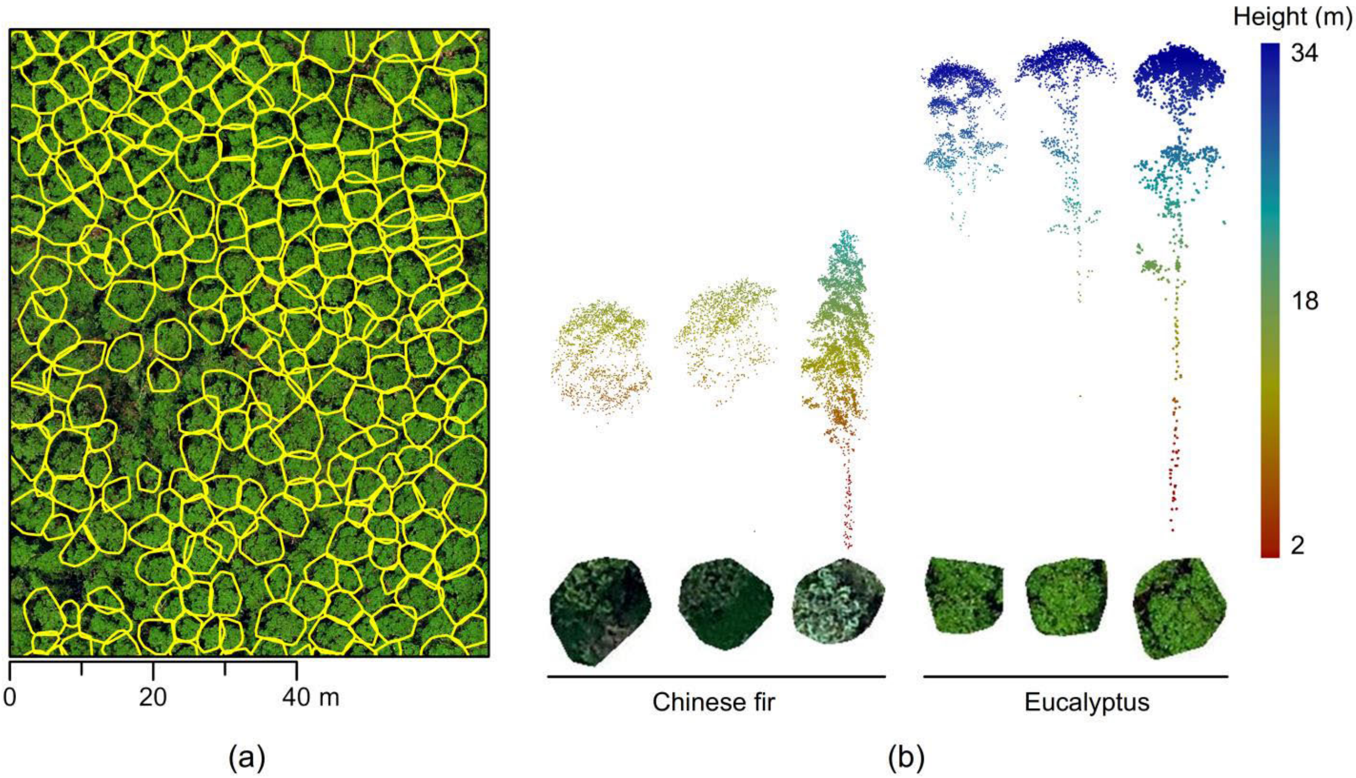

Fig. 3. Results of the individual tree segmentation. (a) Canopy boundaries. (b) LiDAR point cloud and DOM image data for individual trees of two species.

Methods of image data augmentation include random cropping (224 × 224), random horizontal flipping, and normalization. In this study, the mean values of the normalized treatment were 0.485, 0.456, and 0.406, and the standard deviations were 0.229, 0.224, and 0.225, respectively. Using the transfer learning approach, the training results obtained by applying the ResNet50 model to the ImageNet dataset are used as pretraining parameters for the model input when using the ResNet50 network for feature extraction.

C. Multimodal Feature Fusion

The key to multimodal deep learning models is to match and fuse category features from different modalities. The point cloud features extracted by PointMLP were used for max pooling processing to obtain classification features with a dimension of 1024. The category features extracted from the images were first processed using the average pooling function, and the resulting feature dimension was 2048. This was reduced to 1024 using a fully connected layer. Since the classification features of the different modes were quite different, a normalization layer was used to normalize the features of the two modes separately. The normalized features are concatenated to generate fused features. Concatenation of features includes combining the features into a single vector. This allowed the model to capture the differences between the features more effectively, since each feature was retained in its entirety rather than being concatenated. Additionally, this approach enabled the model to learn the most important information from each feature separately, leading to better performance. The fused features were fed into the classification network; finally, the predicted object class was obtained. The classifier was mainly composed of three fully connected networks. Dropout processing was used to avoid overfitting of the model. The multimodal deep learning tree species classification model was trained using model hyperparameters consistent with the PointMLP module (see Table I).

TABLE I
HYPERPARAMETER SETTINGS USED FOR THE MULTIMODAL MODEL

Hyperparameter	Value
Batch Size	8
Number of Points	1024
Number of Categories	2
Epochs	300
Optimizer	SGD
Learning Rate	0.1
Learning Rate Scheduler	CosineAnnealingLR
Weight Decay	0.0002
Momentum	0.9
Loss Function	CrossEntropyLoss
Activation Function	ReLU; AdaptiveMaxPool1d

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Results of the Individual Tree Segmentation and Analysis

The UAV LiDAR data of the study area were accurately segmented and extracted using the HRM method [54]. Fig. 3(a) shows the canopy boundaries of individual trees generated from the UAV LiDAR data in part of the plot in the study area. The maximum values of the number of individual trees were 21 132 and 5658 for eucalyptus and Chinese fir, respectively, and their average values were 3150 and 2155, respectively. Fig. 3(b) shows some of the individual tree samples used in this experiment, including the point cloud of the individual tree and the canopy image of the corresponding individual tree. Generally, the trees in the study area were evenly distributed. The range of canopy area for eucalyptus was 5.79–46.56 m^2 with a mean value of 20.20 m^2, while that for Chinese fir was 7.28–49.09 m^2 with a mean value of 20.81 m^2.

Using the HRM method, the point cloud oversegmentation was first processed according to the local density of the point cloud of the forest scene; the oversegmented area was then merged hierarchically using the global compactness of the complete point cloud as the merging condition until the stopping condition was met. In this way, the final single-tree segmentation results were obtained. Using this method, the search for local "peak" features, which is a feature of traditional algorithms, is replaced by the use of the global compactness; this improves the segmentation in situations, where the tree tops are not obvious or the canopy occlusion is severe and results in greater segmentation where the stand density is higher [54]. By comparing and analyzing the point clouds of individual trees and the canopy imagery after segmentation, we found that the individual tree point clouds extracted from the UAV LiDAR point cloud data by the HRM method were accurate.

B. Model Training

To verify the performance of the proposed TSCMDL model for tree classification, we compared the accuracy of tree classification using RGB images and point cloud data, respectively. The ResNet50 model was used to classify tree species from RGB image data. The PointMLP model has been used for tree species classification from point cloud data, and the training parameters of the optimal model were stored.

The TSCMDL framework is a multimodal deep learning framework for tree species classification based on LiDAR and orthophoto data. The model consists of three main parts: image feature extraction, LiDAR feature extraction, and feature fusion and classification. During the training of the deep learning model, we used the ResNet50 network to extract the image features and the transfer learning technique to pretrain the model parameters on the ImageNet dataset. The PointMLP architecture does not officially provide pretrained parameters, so to train the TSCMDL model, we designed two sets of comparative experiments. The first set of experiments trained the TSCMDL network architecture from scratch, and pretrained parameters were not provided for the point cloud feature extraction. In the second set of experiments, the parameters stored by the PointMLP model in the point cloud data training experiments were used as pretraining parameters for the TSCMDL model.

After the model training had been completed, it was found that in the experiments that did not provide pretrained parameters for the point cloud features, 120 epochs were required for the model to achieve a stable classification accuracy. In contrast, when pretrained parameters obtained using the PointMLP network were used, the optimal classification accuracy was achieved after ten epochs. This demonstrates that, for the TSCMDL network to achieve the optimal classification accuracy quickly, appropriate pretrained parameters based on LiDAR point cloud features should be provided.

C. Operational Efficiency Analysis

The time required to train the two deep learning models, TSCMDL and PointMLP, was recorded. The same training strategy was used to train both models and both were trained over 300 epochs. The time per epoch calculated in this article was obtained by dividing the total time spent for experiment by 300. It was calculated that 0.47 min was required to train each epoch of the TSCMDL model; the occupancy of the GPU memory was 98%. When training with the PointMLP model alone, there was no need to normalize the features of the classifier input, which results in large feature dimensions and large feature values, thereby increasing the classification time. The PointMLP model had a training time of approximately 0.50 min/epoch, and the occupancy of the GPU memory was 92%. Another set of comparative experiments was then designed. The batch size used for the training of the PointMLP model was increased to 12. After doing this, it was found that the training time for each epoch was still 0.50 min, but the GPU memory occupancy had increased to 96%.

These results show that if the same hyperparameters are used, a deep learning model with multimodal features requires more memory than a model that does not; however, the inclusion of multimodal features has only a small impact on the training time. The PointMLP model does not include complex feature extractors and mainly relies on highly optimized feed forward MLPs [51]. This means that even if this model is used for a four-stage feature extraction process, it can still run efficiently. This is one of the reasons why the PointMLP model is used as the backbone of multimodal deep learning models.

D. Classification Accuracy Results and Analysis

The tree classification results that were obtained using the multimodal deep learning strategy after feature fusion and those obtained using the point cloud deep learning method without feature fusion were compared. The results were evaluated based on the overall accuracy (OA), precision, recall, and F1 score. These metrics are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{8}$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives,

TABLE II
VALUES OF INDICATORS USED TO EVALUATE THE TREE SPECIES CLASSIFICATION RESULTS

	PointMLP Point Cloud Feature-Based			TSCMDL Multimodal Feature-Based		
	Precision	*Recall*	*F1*	*Precision*	*Recall*	*F1*
Chinese fir	0.9406	0.9500	0.9453	0.9755	0.9950	0.9851
Eucalyptus	0.9495	0.9400	0.9447	0.9949	0.9750	0.9848
Average	0.9450	0.9450	0.9450	0.9852	0.9850	0.9850

respectively. The precision is thus equal to the user's accuracy, and the recall corresponds to the producer's accuracy. OA represents the proportion of correctly classified results. When evaluating classification results, higher values of these four metrics indicate a better model performance.

The overall accuracies of the tree species classification obtained using the TSCMDL and PointMLP models were 0.9450 and 0.9850, respectively. When using only RGB images for tree species classification, the accuracy (recall) for eucalyptus and Chinese fir was 0.8500 and 0.8167, respectively. The values of the different evaluation metrics for the two models are summarized in Table II. It can be seen that, for all the evaluation metrics, the values for the classification based on multimodal feature fusion are higher than those for the classification based on point cloud features without feature fusion. Thus, it can be concluded that the fusion of 2-D and 3-D features can improve the accuracy of tree species classification. A comparison of the F1 scores for the two classification methods shows that the models' ability to identify tree species is approximately the same for different species.

From the above results, it can be seen that the accuracy of the classification obtained using the PointMLP method is also good. In reported studies on the use of point cloud deep learning methods for tree species classification based on an MLP, the PointMLP method has been shown to perform very well. In one study [40] in which the PointMLP model was used to classify four different tree species, an accuracy of 0.9474 was achieved. In another study [42], four different point cloud deep learning tree classification methods were compared, and the PointMLP model achieved an accuracy of 0.9827 in identifying seven different tree species. In a recent study [55], the PointMLP network was also incorporated into other multimodal object classification models.

Although good results can be obtained by using deep learning models to classify trees based on point clouds corresponding to individual trees [42], until now, this type of research has mainly relied on the use of ground-based LiDAR data [35], [36], [37]. From the results of this study, it can be seen that, because UAV LiDAR cannot completely detect the tree structure in the lower part of the canopy, some point clouds of individual trees included only the canopy and it was difficult to directly identify the tree species using the point cloud features. However, after the fused image features were included, the classification accuracy improved. Therefore, it is not enough to use only the characteristics of point clouds for tree species classification, especially when the point clouds of individual trees do not depict the tree structure completely. Using a range of data types, information about the properties of many types of objects can be obtained, and more accurate classification of tree species can be achieved if the features contained in different types of data are fused.

Until now, in research on deep learning methods that use multisource data for tree species classification, point cloud data are usually converted into image data that can be used as the input to deep learning models. However, choosing the right data-conversion method for the conversion of the point cloud data requires expert knowledge, and different experimental data may require the use of different data conversion and processing techniques. Finding an efficient and optimal way to aggregate the features of the converted imagery is still challenging [56].

How to effectively extract category features directly from original point clouds is a focus of current research. By using multimodal deep learning methods based on raw point clouds and image data, complex point cloud data-conversion processes can be avoided, and highly accurate classification of tree species can be achieved. Getting the right "ground truth" is critical to ensure the quality of the data produced [57]. The results discussed above show that the proposed TSCMDL framework is a valuable and practical approach that can be applied to the investigation of forest resources and identification of tree species.

V. CONCLUSION

In this study, image data and LiDAR data were combined to provide a larger number of features for use in tree species classification. The proposed multimodal deep learning tree species classification (TSCMDL) framework was used to directly extract features from the imagery and original point clouds, thus avoiding the need for manual feature screening. The accuracy of the tree classification that was achieved had a high accuracy. The inclusion of a multimodal learning framework helped to resolve the problems caused by the shortcomings of image conversion of LiDAR data in traditional multisource data classification methods and provided richer, more accurate point cloud classification features. By combining UAV LiDAR data with image data, the amount of feature information used in the tree species classification was expanded; hence, better results were obtained than those using acquired LiDAR data using drones or other airborne platforms because these data cannot depict the complete structure of trees. The proposed TSCMDL framework can thus be used to modify or replace other feature extraction networks and has wide application potential.

Acknowledgment

The authors would like to thank Chao Ma for his help in field data collection. They would also like to thank Enago (www.enago.cn) for providing English editing.

References

[1] Y. Lin and M. Herold, "Tree species classification based on explicit tree structure feature parameters derived from static terrestrial laser scanning data," *Agricult. Forest Meteorol.*, vol. 216, pp. 105–114, Jan. 2016, doi: 10.1016/j.agrformet.2015.10.008.

[2] L. Terryn et al., "Tree species classification using structural features derived from terrestrial laser scanning," *ISPRS J. Photogramm. Remote Sens.*, vol. 168, pp. 170–181, Oct. 2020, doi: 10.1016/j.isprsjprs.2020.08.009.

[3] M. Wang, M. S. Wong, and S. Abbas, "Tropical species classification with structural traits using handheld laser scanning data," *Remote Sens.*, vol. 14, no. 8, p. 1948, Apr. 2022, doi: 10.3390/rs14081948.

[4] Y. Chen et al., "Rapid urban roadside tree inventory using a mobile laser scanning system," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3690–3700, Sep. 2019, doi: 10.1109/JSTARS.2019.2929546.

[5] Y. Li, G. Chai, Y. Wang, L. Lei, and X. Zhang, "ACE R-CNN: An attention complementary and edge detection-based instance segmentation algorithm for individual tree species identification using UAV RGB images and LiDAR data," *Remote Sens.*, vol. 14, no. 13, p. 3035, Jun. 2022, doi: 10.3390/rs14133035.

[6] J. J. Donager, A. J. S. Meador, and R. C. Blackburn, "Adjudicating perspectives on forest structure: How do airborne, terrestrial, and mobile LiDAR-derived estimates compare?" *Remote Sens.*, vol. 13, no. 12, p. 2297, Jun. 2021, doi: 10.3390/rs13122297.

[7] E. Hyyppä et al., "Accurate derivation of stem curve and volume using backpack mobile laser scanning," *ISPRS J. Photogramm. Remote Sens.*, vol. 161, pp. 246–262, Mar. 2020, doi: 10.1016/j.isprsjprs.2020.01.018.

[8] Y. Wu and X. Zhang, "Object-based tree species classification using airborne hyperspectral images and LiDAR data," *Forests*, vol. 11, no. 1, p. 32, 2019, doi: 10.3390/f11010032.

[9] H. Zhong et al., "Identification of tree species based on the fusion of UAV hyperspectral image and LiDAR data in a coniferous and broad-leaved mixed forest in northeast China," *Frontiers Plant Sci.*, vol. 13, Sep. 2022, Art. no. 964769, doi: 10.3389/fpls.2022.964769.

[10] D. Zhao, Y. Pang, L. Liu, and Z. Li, "Individual tree classification using airborne LiDAR and hyperspectral data in a natural mixed forest of Northeast China," *Forests*, vol. 11, no. 3, p. 303, Mar. 2020, doi: 10.3390/f11030303.

[11] G. Yang et al., "Tree species classification by employing multiple features acquired from integrated sensors," *J. Sensors*, vol. 2019, pp. 1–12, Mar. 2019, doi: 10.1155/2019/3247946.

[12] X. Tian, X. Zhang, and Y. Wu, "Classification of planted forest species in southern China with airborne hyperspectral and LiDAR data," *J. Forest Res.*, vol. 25, no. 6, pp. 369–378, Nov. 2020, doi: 10.1080/13416979.2020.1786898.

[13] C. Sothe et al., "Tree species classification in a highly diverse subtropical forest integrating UAV-based photogrammetric point cloud and hyperspectral data," *Remote Sens.*, vol. 11, no. 11, p. 1338, Jun. 2019, doi: 10.3390/rs11111338.

[14] Y. Shi, T. Wang, A. K. Skidmore, and M. Heurich, "Improving LiDAR-based tree species mapping in central European mixed forests using multi-temporal digital aerial colour-infrared photographs," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 84, Feb. 2020, Art. no. 101970, doi: 10.1016/j.jag.2019.101970.

[15] H. Qin, W. Zhou, Y. Yao, and W. Wang, "Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data," *Remote Sens. Environ.*, vol. 280, Oct. 2022, Art. no. 113143, doi: 10.1016/j.rse.2022.113143.

[16] P. Krzystek, A. Serebryanyk, C. Schnörr, J. Červenka, and M. Heurich, "Large-scale mapping of tree species and dead trees in Šumava National Park and Bavarian Forest National Park using LiDAR and multispectral imagery," *Remote Sens.*, vol. 12, no. 4, p. 661, Feb. 2020, doi: 10.3390/rs12040661.

[17] J. Marrs and W. Ni-Meister, "Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data," *Remote Sens.*, vol. 11, no. 7, p. 819, Apr. 2019, doi: 10.3390/rs11070819.

[18] Z. Xu et al., "Tree species classification using UAS-based digital aerial photogrammetry point clouds and multispectral imageries in subtropical natural forests," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 92, Oct. 2020, Art. no. 102173, doi: 10.1016/j.jag.2020.102173.

[19] K. Wang, T. Wang, and X. Liu, "A review: Individual tree species classification using integrated airborne LiDAR and optical imagery with a focus on the urban environment," *Forests*, vol. 10, no. 1, p. 1, Dec. 2018, doi: 10.3390/f10010001.

[20] H. Li, B. Hu, Q. Li, and L. Jing, "CNN-based individual tree species classification using high-resolution satellite imagery and airborne LiDAR data," *Forests*, vol. 12, no. 12, p. 1697, Dec. 2021, doi: 10.3390/f12121697.

[21] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review," *J. Imag.*, vol. 5, no. 5, p. 52, May 2019, doi: 10.3390/jimaging5050052.

[22] B. Zhang, L. Zhao, and X. Zhang, "Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images," *Remote Sens. Environ.*, vol. 247, Sep. 2020, Art. no. 111938, doi: 10.1016/j.rse.2020.111938.

[23] G. A. Fricker, J. D. Ventura, J. A. Wolf, M. P. North, F. W. Davis, and J. Franklin, "A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery," *Remote Sens.*, vol. 11, no. 19, p. 2326, Oct. 2019, doi: 10.3390/rs11192326.

[24] D. L. Torres et al., "Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery," *Sensors*, vol. 20, no. 2, p. 563, Jan. 2020, doi: 10.3390/s20020563.

[25] S. Natesan, C. Armenakis, and U. Vepakomma, "ResNet-based tree species classification using UAV images," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W13, pp. 475–481, Jun. 2019, doi: 10.5194/isprs-archives-XLII-2-W13-475-2019.

[26] S. Natesan, C. Armenakis, and U. Vepakomma, "Individual tree species identification using dense convolutional network (DenseNet) on multitemporal RGB images from UAV," *J. Unmanned Vehicle Syst.*, vol. 8, no. 4, pp. 310–333, Dec. 2020, doi: 10.1139/juvs-2020-0014.

[27] X. Zou, M. Cheng, C. Wang, Y. Xia, and J. Li, "Tree classification in complex forest point clouds based on deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2360–2364, Dec. 2017, doi: 10.1109/LGRS.2017.2764938.

[28] H. Guan, Y. Yu, Z. Ji, J. Li, and Q. Zhang, "Deep learning-based tree classification using mobile LiDAR data," *Remote Sens. Lett.*, vol. 6, no. 11, pp. 864–873, Sep. 2015, doi: 10.1080/2150704x.2015.1088668.

[29] H. Hamraz, N. B. Jacobs, M. A. Contreras, and C. H. Clark, "Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 219–230, Dec. 2019, doi: 10.1016/j.isprsjprs.2019.10.011.

[30] D. Marinelli, C. Paris, and L. Bruzzone, "An approach based on deep learning for tree species classification in LiDAR data acquired in mixed forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3181680.

[31] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 652–660.

[32] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.

[33] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3D point clouds: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4338–4364, Dec. 2021, doi: 10.1109/TPAMI.2020.3005434.

[34] Z. Xi, C. Hopkinson, S. B. Rood, and D. R. Peddle, "See the forest and the trees: Effective machine and deep learning algorithms for wood filtering and tree species classification from terrestrial laser scanning," *ISPRS J. Photogramm. Remote Sens.*, vol. 168, pp. 1–16, Oct. 2020, doi: 10.1016/j.isprsjprs.2020.08.001.

[35] E. K. Sakharova, D. D. Nurlyeva, A. A. Fedorova, A. R. Yakubov, and A. I. Kanev, "Issues of tree species classification from LiDAR data using deep learning model," in *Advances in Neural Computation, Machine Learning, and Cognitive Research V* (Studies in Computational Intelligence). Cham, Switzerland: Springer, 2022, ch. 42, pp. 319–324, doi: 10.1007/978-3-030-91581-0_42.

[36] J. Chen, Y. Chen, and Z. Liu, "Classification of typical tree species in laser point cloud based on deep learning," *Remote Sens.*, vol. 13, no. 23, p. 4750, Nov. 2021, doi: 10.3390/rs13234750.

[37] M. Liu, Z. Han, Y. Chen, Z. Liu, and Y. Han, "Tree species classification of LiDAR data based on 3D deep learning," *Measurement*, vol. 177, Jun. 2021, Art. no. 109301, doi: 10.1016/j.measurement.2021.109301.

[38] Y. Lv et al., "A convex hull-based feature descriptor for learning tree species classification from ALS point clouds," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3055773.

[39] D. Seidel et al., "Predicting tree species from 3D laser scanning point clouds using deep learning," *Frontiers Plant Sci.*, vol. 12, Feb. 2021, Art. no. 635440, doi: 10.3389/fpls.2021.635440.

[40] B. Liu, H. Huang, X. Tian, and M. Ren, "Individual tree species classification using the pointwise MLP-based point cloud deep learning method," *Environ. Sci. Proc.*, vol. 22, no. 1, p. 19, 2022, doi: 10.3390/IECF2022-13049.

[41] B. Liu, S. Chen, H. Huang, and X. Tian, "Tree species classification of backpack laser scanning data using the PointNet++ point cloud deep learning method," *Remote Sens.*, vol. 14, no. 15, p. 3809, Aug. 2022, doi: 10.3390/rs14153809.

[42] B. Liu et al., "Tree species classification using ground-based LiDAR data by various point cloud deep learning methods," *Remote Sens.*, vol. 14, no. 22, p. 5733, Nov. 2022, doi: 10.3390/rs14225733.

[43] Y. Sun, Q. Xin, J. Huang, B. Huang, and H. Zhang, "Characterizing tree species of a tropical wetland in southern China at the individual tree level based on convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4415–4425, Nov. 2019, doi: 10.1109/JSTARS.2019.2950721.

[44] Ø. D. Trier et al., "Tree species classification in Norway from airborne hyperspectral and airborne laser scanning data," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 336–351, Jan. 2018, doi: 10.1080/22797254.2018.1434424.

[45] V. M. Scholl, J. McGlinchy, T. Price-Broncucia, J. K. Balch, and M. B. Joseph, "Fusion neural networks for plant classification: Learning to combine RGB, hyperspectral, and LiDAR data," *PeerJ*, vol. 9, Jul. 2021, Art. no. e11790, doi: 10.7717/peerj.11790.

[46] J. Mäyrä et al., "Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks," *Remote Sens. Environ.*, vol. 256, Apr. 2021, Art. no. 112322, doi: 10.1016/j.rse.2021.112322.

[47] S. Briechle, P. Krzystek, and G. Vosselman, "Silvi-Net—A dual-CNN approach for combined classification of tree species and standing dead trees from remote sensing data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 98, Jun. 2021, Art. no. 102292, doi: 10.1016/j.jag.2020.102292.

[48] Z. Yang, Y. Sun, S. Liu, X. Shen, and J. Jia, "STD: Sparse-to-dense 3D object detector for point cloud," Presented at the IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019.

[49] S. Jabeen, X. Li, M. S. Amin, O. Bourahla, S. Li, and A. Jabbar, "A review on methods and applications in multimodal deep learning," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 2, pp. 1–41, Jun. 2023, doi: 10.1145/3545572.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Presented at the IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016.

[51] X. Ma, C. Qin, H. You, H. Ran, and Y. Fu, "Rethinking network design and local geometry in point cloud: A simple residual MLP framework," in *Proc. Int. Conf. Learn. Represent.*, Apr. 2022, pp. 25–29.

[52] X.-F. Han, J. S. Jin, M.-J. Wang, W. Jiang, L. Gao, and L. Xiao, "A review of algorithms for filtering the 3D point cloud," *Signal Process., Image Commun.*, vol. 57, pp. 103–112, Sep. 2017, doi: 10.1016/j.image.2017.05.009.

[53] W. Zhang et al., "An easy-to-use airborne LiDAR data filtering method based on cloth simulation," *Remote Sens.*, vol. 8, no. 6, p. 501, Jun. 2016, doi: 10.3390/rs8060501.

[54] Y. Hao, F. R. A. Widagdo, X. Liu, Y. Liu, L. Dong, and F. Li, "A hierarchical region-merging algorithm for 3-D segmentation of individual trees using UAV-LiDAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5701416, doi: 10.1109/TGRS.2021.3121419.

[55] L. Xue et al., "ULIP: Learning a unified representation of language, images, and point clouds for 3D understanding," 2022, *arXiv:2212.05171*.

[56] D. Wu, Z. Liang, and G. Chen, "Deep learning for LiDAR-only and LiDAR-fusion 3D perception: A survey," *Intell. Robot.*, vol. 2, no. 2, pp. 105–129, 2022, doi: 10.20517/ir.2021.20.

[57] J. Wen et al., "Characterizing the effect of spatial heterogeneity and the deployment of sampled plots on the uncertainty of ground 'truth' on a coarse grid scale: Case study for near-infrared (NIR) surface reflectance," *J. Geophys. Res., Atmos.*, vol. 127, no. 11, Jun. 2022, Art. no. e2022JD036779, doi: 10.1029/2022jd036779.

Bingjie Liu (Graduate Student Member, IEEE) received the B.S. degree in geographic information science and the M.S. degree in forest management from Northeast Forestry University, Harbin, China, in 2016 and 2019, respectively, He is currently pursuing the Ph.D. degree with Beijing Forestry University, Beijing, China.

His research interests include deep learning and LiDAR applications in forest ecology, vegetation remote sensing and geoscience technology, and spatiotemporal data analysis and modeling.

Yuanshuo Hao received the B.S. degree in geographic information science and the Ph.D. degree in forest management from Northeast Forestry University, Harbin, China, in 2016 and 2022, respectively.

Since 2022, he has been with the Key Laboratory of Sustainable Forest Ecosystem Management-Ministry of Education, School of Forestry, Northeast Forestry University, where he is currently an Associate Professor. His research interests include point clouds processing and remote sensing applications in forestry resources inventory and modeling.

Huaguo Huang received the B.Sc. degree in forestry and the M.Sc. degree in forest management from Beijing Forestry University, Beijing, China, in 2001 and 2004, respectively, and the Ph.D. degree in quantitative remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, in 2007.

Since 2007, he has been with Forestry College, Beijing Forestry University, where he is currently a Full Professor. His research interests include 3-D modeling of reflectance, thermal emission, LiDAR and microwave scattering of land surfaces (vegetation in particular), and remote sensing applications in agriculture and forest environment.

Shuxin Chen received the B.S. degree in geographic information science from Northeast Forestry University, Harbin, China, in 2019, where he is currently pursuing the Ph.D. degree with the Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing, China.

His research interests include stereo photogrammetry technology, especially based on UAV tilt photogrammetry and GF-7.

Zengyuan Li (Member, IEEE) received the B.S. degree from Inner Mongolia Agriculture University, Hohhot, China, in 1982, and the M.S. and Ph.D. degrees from China Agricultural University, Beijing, China, in 1987 and 1990, respectively.

From 1994 to 1996, he was a Post-Doctoral Researcher of tropical forest classification and a Visiting Scholar of paddy field dynamic monitoring using the ERS SAR with the Laboratory of Remote Sensing, Joint Research Center, European Union. From 1990 to 1996, he was a Research Assistant and an Associate Professor with the Research Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing. Since 1996, he has been a Professor and the Deputy Director of the Institute of Forest Resource Information Techniques, Chinese Academy of Forestry. His research interests include digital forestry, remote sensing of ecosystem environment, and vegetation information extraction from microwave and LiDAR technology.

Erxue Chen (Member, IEEE) received the B.S. and M.Agr. degrees in forest management and statistics forestry from Shandong Agricultural University, Tainan, China, in 1992 and 1995, respectively, and the Ph.D. degree in SAR image geocoding and topography radiometric correction algorithm from the Chinese Academy of Forestry, Beijing, China, in 2004.

He is currently a Professor with the Institute of Forest Resource Information Techniques, Chinese Academy of Forestry. He has taught courses including radar remote sensing of forest resources and radar remote sensing of vegetation. His research interests include surface height and vegetation spatial structure from PolInSAR and SAR data processing and forest applications.

Xin Tian received the B.S. and M.S. degrees from Beijing Forestry University, Beijing, China, in 2001 and 2004, respectively, and the Ph.D. degree from the University of Twente, Enschede, the Netherlands, in 2015.

Since 2011 and 2021, he has been a Professor and Principle Investigator of the Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing, respectively. His research interests include quantitative inversion of forest parameters using multisource remote sensing.

Min Ren received the B.S. degree in electronic information science and technology and the M.S. degree in communication and information system from the Xi'an University of Science and Technology, Xi'an, China, in 2008 and 2011, respectively.

She is currently a Senior Consulting Designer with the China Mobile Communications Group Design Institute Company, Ltd., Taiyuan, China.