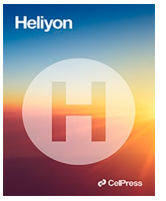
Heliyon 9 (2023) e20467

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



More appropriate DenseNetBL classifier for small sample tree species classification using UAV-based RGB imagery

Ni Wang [a,*], Tao Pu [a,b], Yali Zhang [a], Yuchan Liu [a,c], Zeyu Zhang [a]

[a] School of Geographic Information and Tourism, Chuzhou University, Chuzhou, 23900, China
[b] School of Geomatics, Anhui University of Science and Technology, Huainan, 232001, China
[c] School of Civil Engineering, Nanjing Forestry University, Nanjing, 210037, China

ARTICLE INFO

Keywords:
DenseNetBL
Forests tree species classification
Small sample classification
SLIC

ABSTRACT

To effectively classify tree species within datasets characterized by limited samples, we introduced a novel approach named DenseNetBL, founded upon the fusion of the DenseNet architecture and a pivotal bottleneck layer. This bottleneck layer, encompassing a compact convolutional component, played a central role in our methodology. The evaluation of DenseNetBL was conducted under varying conditions, encompassing small-sample tree species data, extensive remote sensing datasets, and state-of-the-art classifiers. Furthermore, a quantitative assessment was executed to extract tree species areas. This was achieved by quantifying pixel areas within manually delineated tree species maps and classifier-generated counterparts. The findings of our study indicated that, in scenarios devoid of pre-trained weights, DenseNetBL consistently outperformed its DenseNet counterpart with equivalent layer numbers. In the realm of small-sample situations, both the Swin Transformer and Vision Transformer exhibited inferior performance when juxtaposed with DenseNet and DenseNetBL. Remarkably, among the shallow architectures, DenseNet33BL showcased superior aptitude for small-sample tree species classification, culminating in the most commendable results (Overall Accuracy (OA) = 0.901, Kappa = 0.892). Conversely, the Vision Transformer yielded the least favorable classification outcomes (OA = 0.767, Kappa = 0.708). The amalgamation of DenseNet33BL and simple linear iterative clustering emerged as the optimal strategy for attaining robust tree species area extraction results across two prototypical forests. In contrast, DenseNet121 exhibited suboptimal performance in the same forests, attaining the least satisfactory tree species area extraction results. These comprehensive findings underscore the efficacy of our DenseNetBL approach in addressing the challenges associated with small-sample tree species classification and accurate tree species area extraction.

1. Introduction

Forests are an important component of the Earth's ecological environment, with significant roles in carbon sequestration, oxygen release, water conservation, and maintenance of biodiversity [1]. Trees are one of the important components of forests, and different tree species play different roles in the aforementioned functions. Therefore, it is necessary to understand the spatial distribution of tree species in forests. Generally, the spatial distribution of tree species in forests can be understood by drawing a map of their distribution,

* Corresponding author.School of Geographic Information and Tourism, Chuzhou University, Chuzhou, China.
E-mail address: wangni@chzu.edu.cn (N. Wang).

https://doi.org/10.1016/j.heliyon.2023.e20467
Received 27 June 2023; Received in revised form 26 September 2023; Accepted 26 September 2023
Available online 26 September 2023

and accurate classification of tree species is the first prerequisite for drawing high-precision maps. Traditional manual classification and mapping of tree species are difficult to balance accuracy, labor, and finances, and may lead to incorrect map information even under conditions of high labor and massive funds [2]. With the development of remote sensing technology, more and more forestry personnel are using remote sensing technology to classify tree species more efficiently and draw higher-precision maps [3]. Among them, low-cost drones can provide high spatial resolution remote sensing images. When combined with machine learning(ML) technology, relevant personnel can more effectively classify and map tree species in an area [4].

At present, tree species classification methods based on remote sensing technology can be roughly divided into two categories [5, 6]. The first category is traditional classifiers in ML. By building traditional classifiers such as support vector machines(SVM), decision trees(DT), and random forests(RF), tree species are manually selected and inputted based on their features [7]. For example, Zhang et al. conducted research on the identification of tree species in urban natural forests using longitudinal profiles of different tree species and DT based on knowledge-based and gain ratio criteria with very high resolution [8]. Sun et al. used RF to classify tree species in urban natural forests based on high-resolution aerial images [9]. Compared with the manual classification of tree species, this method eliminates the problem of low efficiency, improves the level of automation in tree species classification, but still relies on the manual selection of different features of tree species for classification, and the quality of the classification results is highly correlated with the quality of the manually selected features [10].

The second type is deep learning(DL), which uses classifiers that can automatically extract image features in DL to finely classify tree species [11]. For example, Schiefer et al. conducted segmentation research on tree species in high-resolution drone images using U-net [12]. Cao et al. classified tree species using the Res-Unet network, which combines the feature extraction architecture of ResNet and U-net [13]. This method overcomes the disadvantage of traditional classifiers requiring manual selection of features and further improves the level of automation, and has become the mainstream method for tree species classification [14].

The content shown above demonstrates that standard ML approaches have enhanced the automation level of classifying tree species, making it more practical than manually mapping out tree species. This method does involve more people, however, the accuracy of classifying a tree's species is closely tied to the manually chosen traits. In addition to becoming a popular research area, DL approaches have significantly increased the automation degree of tree species classification based on conventional ML methods. But unlike other classification domains, tree species classification research frequently employs a manual method to produce tree species datasets, making it challenging to produce huge datasets with a constrained picture quantity and scope. Deep, complicated, and advanced classifiers may exhibit underfitting on small sample datasets due to the complex structure and multiple hyperparameters, making them unable to identify unlabeled samples. Currently, there is minimal study on small sample tree species classification [15].

The methods for drawing tree species distribution maps can be broadly divided into two categories [16]. The first method is to use a classifier to classify the whole image by pixel to draw a tree species map (semantic segmentation). This method is relatively easy to implement, but it is prone to salt-and-pepper noise under extremely special conditions, and it is often used in large study areas [17]. The second category is to use a classifier to recognize the objects generated by segmentation algorithms to draw a distribution map [18]. Compared with semantic segmentation, this method increases the complexity of the drawing process but greatly reduces salt-and-pepper noise. It can build more accurate tree species distribution maps in small areas and is currently a popular method for drawing tree species maps.

DenseNet is composed of more dense connection mechanisms, which can build a deeper network compared to traditional DL classifiers. It can achieve more complex nonlinear feature extraction and has excellent classification results in large sample classification. However, there are complex and numerous dense connections in its dense blocks, which can achieve deep level feature extraction, but also make DenseNet parameters redundant and may not be suitable for small sample tree species classification fields [19]. To improve the accuracy of small sample tree species classification in small study areas and draw more accurate tree species distribution maps, this study proposes a new small sample tree species identification algorithm (DenseNetBL), which is based on DenseNet and inserts small convolutional layers as bottleneck layers between Dense Blocks to form Dense Block_BL. The bottleneck layer reduces the output parameters while maintaining the input feature map dimensions, allowing for the fusion of multi-dimensional feature information in a deep network architecture, further increasing the network's nonlinear expression. Four Dense Block_BLs are combined to form a new feature extraction architecture, followed by different optimizers and Softmax to build DenseNetBL. Two advanced and complex classifiers, Swin Transformer and Vision Transformer were also constructed, and the effects of different network depths on small sample classification ability and large sample image classification ability were quantitatively analyzed using small, medium, and large sample datasets. AdamW and SGD optimizers were also constructed, and the effects of batch size and learning rates on small sample tree species classification were tested. Additionally, we combined the best classifier with Simple Linear Iterative Clustering (SLIC) to draw a spatial distribution map of forest tree species, used manual drawing to draw a reference map, and quantitatively evaluated the effectiveness of the above classifier by calculating the difference in tree species area extracted by the two. This study can be used for small sample tree species classification, providing information support for obtaining region-al ecological information, improving urban environmental services, and promoting sustainable development of forestry.

2. Materials and methods

2.1. Study area and typical forest

Geographically, the research area is situated between 118°17′40″E and 118°18′00″E and 32°16′10″N and 32°16′30″N in Chuzhou City, Anhui Province, China's Langya Mountain, a 4 A-level scenic area (Fig. 1 (a)). The altitude ranges from 100 m to 320 m. It has an average annual temperature of 15.2 °C, an average annual precipitation of roughly 1050 mm, and a frost-free period of 217 days

throughout the year. It is part of the humid and semi-humid monsoon climate of the transition from the northern subtropical zone to the warm temperate zone.

Two typical forests (Fig. 1 (b,c)) were chosen as the experimental location because the original image included elements such as structures, lakes, roads, and grasslands that were not taken into account for the tree species categorization experiment. The typical woodland A, which is in the center of the image, is surrounded by trails, consists of planted forests, and is approximately 100 m × 100 m in size. Different tree species, such as Ligustrum lucidum Ait, Photinia beauverdiana, and Osmanthus fragrans, are distributed regularly and neatly throughout the woodland. The lower right corner of the photograph has typical woodland B, which measures 100 m by 100 m and has a winding road through it. There is a small amount of natural forest surrounding the planted forests, with clusters of tree species, primarily Photinia beauverdiana and Osmanthus fragrans, and the majority of the area is planted forest with a neat distribution of tree species, primarily including Cupressus, Ligustrum lucidum Ait, and Cinnamomum camphora. The remaining species make up roughly 10% of the total population and are evenly distributed.

2.2. High-resolution remote sensing image acquisition and preprocessing

The UAV high-resolution photos from Dajiang UAV (spirit p4r), with horizontal and vertical resolutions of 96 dpi, are employed in this investigation. The high-resolution image has a range of 680 m–1447 m with a spatial resolution of 0.05 m. On clear, sunny days, data were collected. With a greater solar altitude angle and steady illumination, the high spatial resolution image was captured between 9:00 and 12:00. Red, green, and blue bands make up the three bands that make up the imagery.

The noise that is produced during the storage and conversion of images is eliminated using Gaussian filtering. The basic concept is to create a Gaussian template using the two-dimensional Gaussian distribution function, and then utilize the template's weighted average of the domain's pixels as the pixel value of the template's center point in the new image.

2.3. Collection dataset and field surveying

Through SLIC, the fundamental unlabeled tree species data sets were obtained in two typical forests. 90 photos of bare land (Fig. 2 (f)) were also collected, bringing the total number of tree species data collected to 759. The Cinnamomum camphora (Fig. 2 (a)), Ligustrum lucidum Ait(Fig. 2 (b)), Photinia beauverdiana(Fig. 2 (c)), Cupressus(Fig. 2 (d)), and Osmanthus fragrans (Fig. 2 (e)) are among the five species of trees.

Field surveying refers to identifying the unlabeled samples obtained from SLIC segmentation in a typical sample [20]. Fieldwork was surveyed in July and August of 2021. Field surveying could only be done at the Langya forestry site if it was accessible due to its geographic location. Unconfirmed images are discarded, and visually interpreted images are reconfirmed by on-site measurements. The main causes for deleting photographs were the lake at the forestry site and the complicated geography. 52 photos in total were eliminated, leaving 759 photographs of tree species with identified categories.

2.4. Data enhancement and splitting

Training of a CNN is based on epochs, which is defined as when all tree species data sets are trained once in CNN [21]. The tree species data set must be separated into a training set and a verification set to assess whether a CNN is overfitted or under fitted throughout the whole training cycle. The model was trained using the training set, and a tree species classifier was built as a result. The accuracy of the classifier used to identify common forest tree species was examined using the verification set. 60% of the data is randomly chosen as classifier training data before classifier training. A further 40% of the data is used to verify the accuracy of the

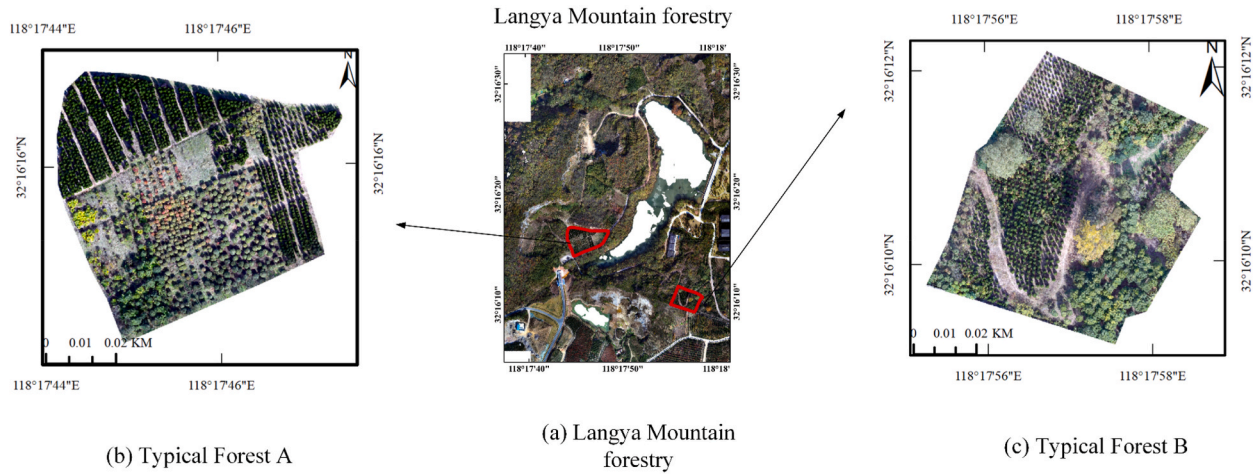


Fig. 1. The location of the study area in Langya Mountain, Chuzhou City in northern China.

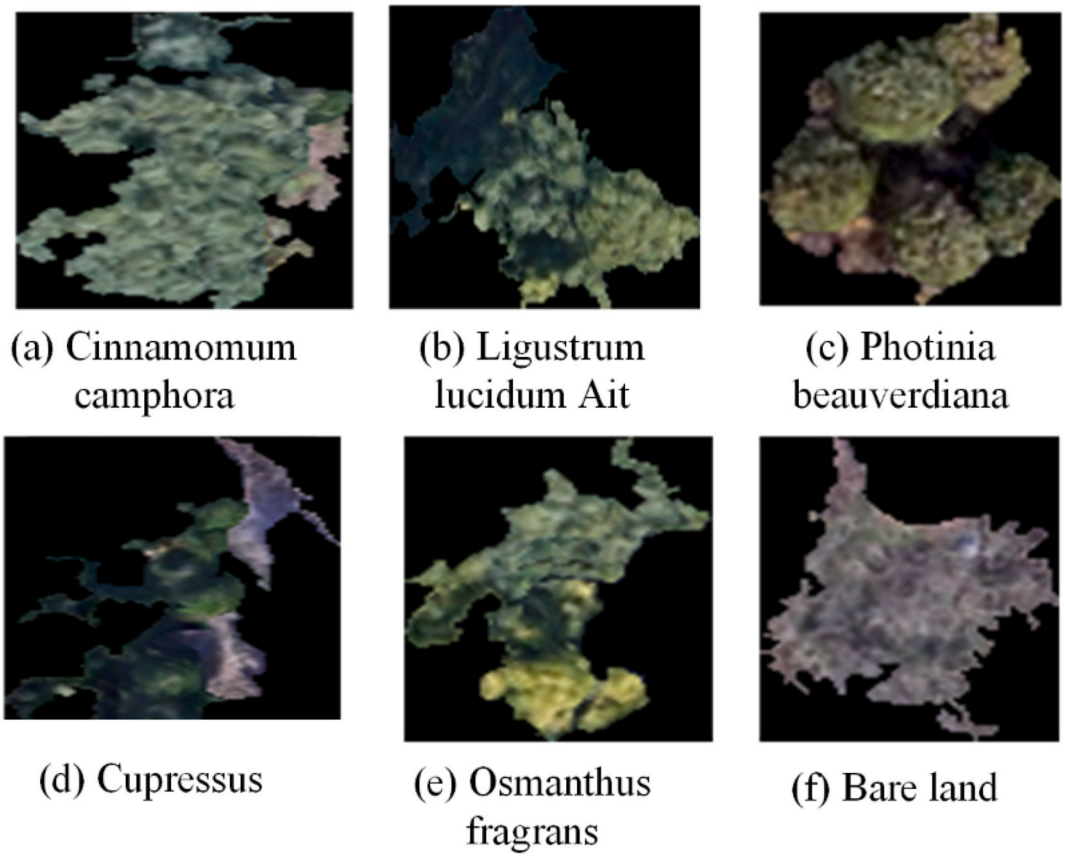


Fig. 2. Detailed overview of the occurring tree species and classes.

classifier (Table 1).

Due to the limitations of sensors, weather, and UAV power, obtaining images in various states as a training dataset is unreliable [22]. DL requires a large number of data distributions of different states to fit the distribution of tree species features of different objects in real scenes to increase the generalization ability and robustness of the classifier. Therefore, Random cropping, random up and down, and horizontal flipping were used to enhance the original tree species' spatial and geometric features [23]. In addition, the gamma transform is used to enhance the geometric and textural features of the original tree species, and the principle is shown in Eq. (1), where C is the pixel value of the original tree species and γ denotes the scaling factor, which is less than 1, the image grayscale is compressed and the image is brighter and greater than 1, the opposite [24]. Additionally, a new data enhancement method is proposed in this study. The idea is to convert the original three-band standard image into a hue, Saturation, and value standard image, and modify tree species' spatial and texture features by adaptive thresholding.

$$S = 255 \times \left(C/255.0 \right)^{\gamma} \tag{1}$$

Table 1
Training and verification sample details.

Tree species	Number of training samples	Number of Validation samples	Number of class
Ligustrum lucidum Ait	173	43	0
Photinia beauverdiana	104	25	1
Cupressus	132	33	2
Osmanthus fragrans	32	7	3
Cinnamomum camphora	96	24	4
Bare land	72	18	5
Total number	609	150	

2.5. Methods

Four steps make up the study's main method, as depicted in Fig. 3. The first step was to execute image preprocessing, which included cropping and enhancing. Second, a dataset of differentiating tree species was made. Third, maps of the spatial distribution of typical tree species in the study region were created after tree species classification. Finally, the results' correctness was assessed.

2.5.1. CNN-based tree species classification

Inspired by the skip connection in ResNet, DenseNet implements dense connections between different blocks, which alleviates the problem of gradient vanishing and enables deeper networks to have higher feature extraction capability (i.e., higher overall accuracy (OA) [25]. The skip connection is shown in Eq. (2), and it represents that the output of the (n+1)-th layer is obtained by concatenating the output of the n-th layer and a portion of the feature map from the other layer.

$$X_{n+1} = X_n + F(x_n + W_n) \tag{2}$$

A dense connection, as shown in Eq. (3), indicates that the input of the n-th layer comes from all the outputs of the previous n-1 layers. This method of directly connecting all layers ensures maximum information transmission [26].

$$X_{n+1} = X_n([X_0, X_1, X_2, \ldots X_n]) \tag{3}$$

Using Dense connection can construct deeper DenseNet, which can better fit the differences between different categories in a large annotated dataset, and achieve deeper semantic feature extraction of tree species [27]. However, building a large-scale dataset of tree species manually is challenging, and deep DenseNet is prone to underfitting on small datasets, leading to relatively poor performance in classifying tree species.

The bottleneck layer refers to a small convolution layer with a size of 1*1, which can increase the network's nonlinearity by adding an activation function, and better express the nonlinear features of tree species(Fig. 4) [28]. Adding bottleneck layers with a low number of channels after various structures can reduce the dimensionality of the classifier and reduce computational parameters [29]. The bottleneck layer also allows for the integration of channel information across multiple channels to achieve tree-species feature interaction [30]. In addition, the bottleneck layer is essentially a 1*1 small convolution layer, which can further achieve deep tree species feature extraction in the network.

2.5.2. DenseNetBL

This study implements Dense BlockBL based on the combination of bottleneck layer and Dense Block to further construct a small-sample tree species classifier, taking into account the capabilities of bottleneck layer dimensionality reduction, increased network nonlinearity expression, and further feature extraction. As seen in Fig. 5, the Dense BlockBL structure is made up of four bottleneck layers, four residual blocks, and a Relu activation function. The dotted line in Fig. 5 shows the usage of dense connections between the structures. Each layer structure is densely coupled to every layer structure that came before it, making maximum use of all feature information.

The Residual Block consists of two pathways, as shown in the schematic part of Fig. 5, specifically represented as the input x undergoing convolutional processing to obtain the output f(x). Additionally, the direct mapping of x added to the output after convolutional processing results in the final output f(x) + x [31].

Before introducing the bottleneck layer, a feature map of size x0*y0*z0 is forwarded to produce a feature map of size x0*y0*z0 and x1*y1*z1.

The computational complexity is given by Eq. (4).

$$A = X \times Y \times Z \times 3 \times 3 \times Z/2 \tag{4}$$

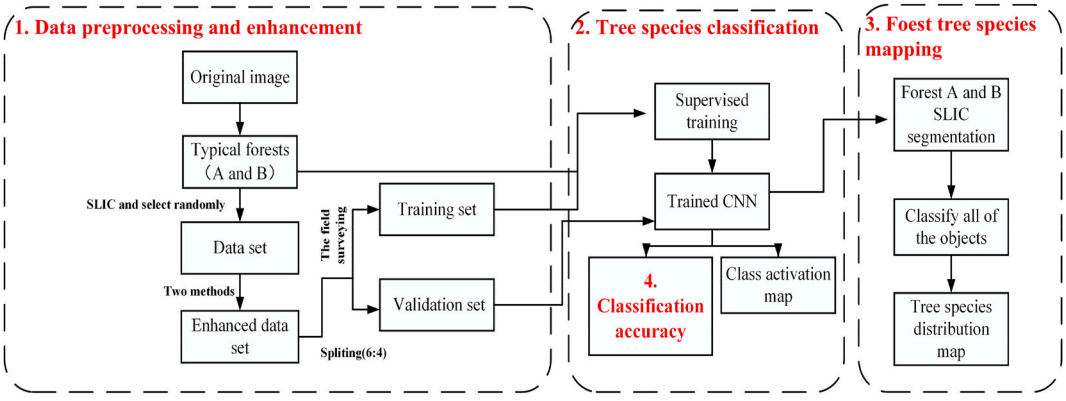


Fig. 3. The flowchart of tree species classification and mapping.

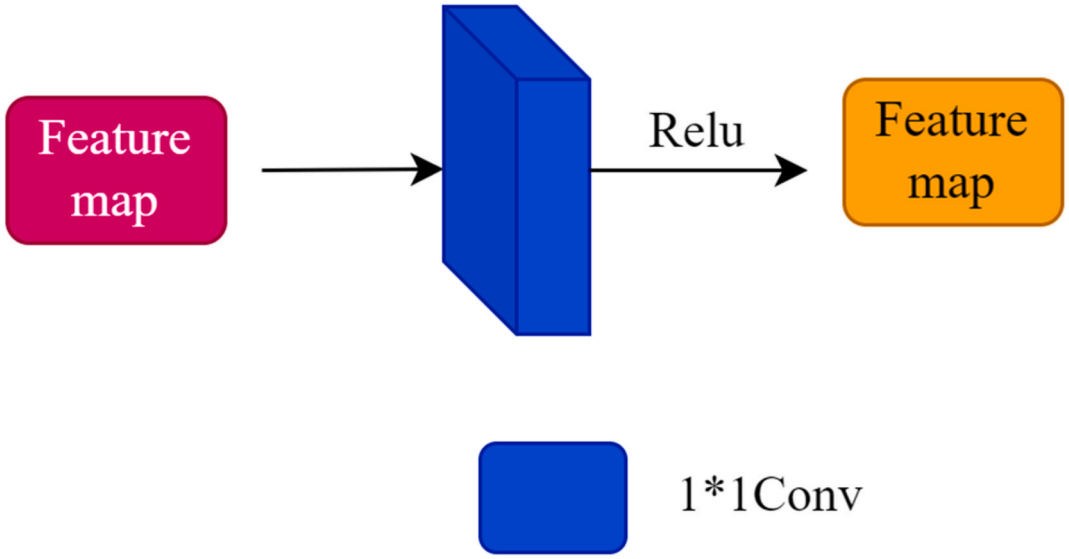


Fig. 4. A bottleneck layer structure.

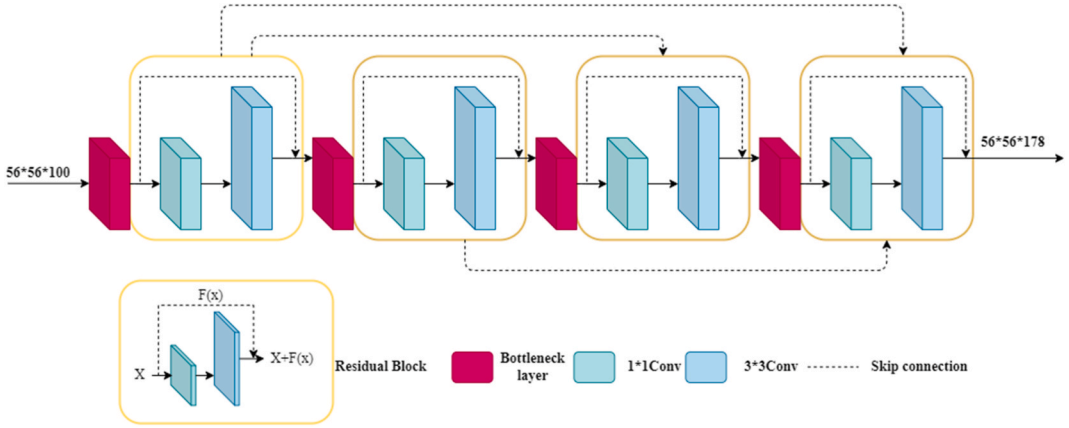


Fig. 5. A simple Dense BlockBL structure.

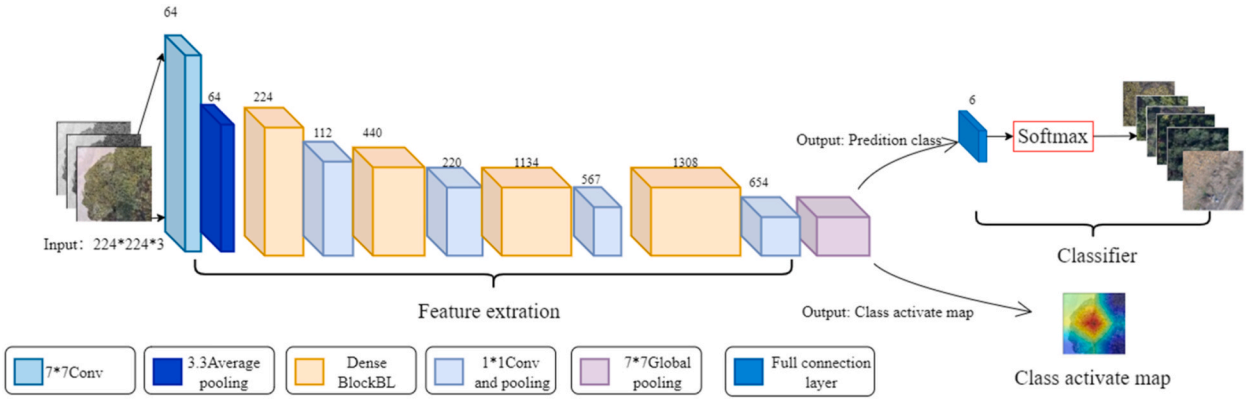


Fig. 6. A simple DenseNetBL structure.

The number of parameters is given by Eq. (5).

$$A = Z \times 3 \times 3 \times Z/2 \tag{5}$$

After adding a bottleneck layer, the incoming feature map is first reduced in dimension by the bottleneck layer, with computational complexity shown in Eq. (6) and parameter count shown in Eq. (7).

$$B = X \times Y \times Z \times 1 \times 1 \times Z/4 + X \times Y \times Z/4 \times 3 \times 3 \times Z/2 \tag{6}$$

$$B = Z \times 1 \times 1 \times Z/4 + Z/4 \times 3 \times 3 \times Z \tag{7}$$

After adding the bottleneck layer, both the parameter and calculation amounts are smaller than those without the bottleneck layer, indicating that the bottleneck layer effectively reduces the dimensionality of the feature map. Moreover, the activation function is integrated into the bottleneck layer, incorporating multiple-channel feature information and enhancing the nonlinear expression capability of the deep network. A new feature extraction architecture is formed by combining four Dense BlockBLs, which are subsequently connected to the AdamW optimizer and Softmax, and a fully connected layer classification tree with 6 connection nodes. Constructing DenseNetBL (Fig. 6) with different layers based on the stacking times of Residual Blocks in Dense BlockBL.

2.5.3. Gradient-weighted class activation map

This study designed the Gradient-Weighted Class Activation Map (Grad CAM) after the Feature extraction, which calculates the region of interest in the image predicted by the classifier in prediction mode through backpropagation to increase the interpretability of the classifier [32]. The principle of Grad CAM is shown in Equation (8), and the basic idea is to calculate the weight of each feature map in the last convolutional layer on the image category and calculate the weighted sum of the feature map to map it to the original image [33].

$$L^{c}_{Grad-CAM} = RELU\left(\sum_{k} \alpha^{c}_{k} A^{k}\right) \tag{8}$$

Eq. (8), A represents a certain feature layer, K is the first channel, and C is the prediction category. A^k represents the K channel in the feature map A. α^c_k is for the weight of A^k.

$$a^{c}_{k} = {}^{1}/_{Z} \sum_{i} \sum_{j} \partial y^{c} \Big/ \partial A^{k}_{ij} \tag{9}$$

The calculation method for α^c_k is shown in Eq. (9), where y^c represents the unprocessed score predicted by the network for the category, A^k_{ij} represents the data at ij in the K channel of the A feature map, and Z represents the width * height of the feature layer.

2.5.4. Tree crowns segmentation

Iteratively clustering and segmenting images based on color similarity and spatial distance correlations is what the SLIC superpixel segmentation algorithm does. Compact-like cells make up the tree crown objects that are formed, and the domain properties are clearly expressed [34]. The initial parameter setup for SLIC determines the size of the tree crown it produces, although the hyperparameter setting is straightforward, the algorithm runs quickly and differently than the watershed technique, and it can immediately achieve end-to-end multi-band image segmentation [35]. This study utilized SLIC to pre-segment individual tree crowns on images.

2.5.5. Experimental design

The fixed epoch is set to 30, and all classifier training and validation operations do not load pre-training weights, regardless of whether the succeeding classifiers converge or not. Build DenseNet based on whether to include bottleneck layers and use DenseNetBL to investigate how bottleneck layers affect the classification of a small sample of tree species. Based on the varying stacking durations of Residual Blocks in Dense Block and Dense BlockBL, DenseNet33, DenseNet33BL, DenseNet68, DenseNet68BL, DenseNet117, DenseNet117BL, DenseNet121, DenseNet121BL, DenseNet169, and DenseNet169BL were built. Comparative studies were done to examine how different classifier layers affected the classification of a small sample of tree species. Table 2 displays the stacking timings for Residual Blocks in DenseNet and DenseNetBL with various layers.

Swin transformer and Vision transformer are currently popular classifiers that can extract multi-region features from images. They have been proven to be suitable for most image classification fields, but there is currently no research on tree species classification [36, 37]. As additional comparison classifiers, this study also builds Swin small (Fig. 7) in the Swin transformer, ResNet16 in ResNet, and

Table 2
Construction parameters of different DenseNet and DenseNetBL.

	DenseNet33/33BL	DenseNet68/68BL	DenseNet117/117BL	DenseNet121/121BL	DenseNet169/169BL
Dense Block/BL(1)	3	6	6	6	6
Dense Block/BL(2)	4	6	6	12	12
Dense Block/BL(3)	4	6	12	24	32
Dense Block/BL(4)	3	16	32	16	32

Vision transformer base16 (Fig. 8) in the Vision transformer. Through the use of two sizable sample datasets in the field of remote sensing, the performance of the aforementioned classifier was quantitatively assessed.

Additionally, this study constructed two optimizers (SGD and AdamW) and tested their performance in the aforementioned classifiers through detailed batch size and learning rates experiments.

2.5.6. Accuracy assessment

The class accuracy (CA), Kappa coefficient, and OA of the trained classifier on the validation set are used as evaluation indicators [38,39]. The classifier's performance in terms of convergence is evaluated using the trend of loss values throughout the training phase. by manually drawing the canopy boundary to create a reference tree species distribution map, using the highest classifier in OA to predict the superpixels created by SLIC and assign labels, and combining the tree species map produced by the classifier to qualitatively analyze the accuracy of the tree species spatial distribution map [40]. Finally, based on the variance in tree area statistics based on the coverage area of pixel values, the classifier's effectiveness in extracting tree area is quantitatively assessed.

3. Results

3.1. Model training

Different classifiers' training and validation losses and OA were noted for each epoch, and the epoch-training loss and epoch-validation OA curves were shown. To smooth the training loss and validation OA curves, 100 fitting values of loss and OA were added between neighboring epochs using Gaussian interpolation (Fig. 9). The Vision transformer did not converge after 30 epochs, as shown in Fig. 9(a) and (b), which also exhibit a declining trend in loss and an uneven OA. The loss curves of the additional classifiers tended to settle after 25 epochs, and all classifiers converged. After 25 epochs, the other classifiers' OA on the validation set steadied, showing that the classifiers' performance was at its peak.

Moreover, as shown in Fig. 9(a), ResNet16 performs similarly to DenseNet, but its classification ability was not as good as that of DenseNet and DenseNetBL. Both DenseNet and DenseNetBL had better convergence and classification abilities. Notably, the shallow DenseNet33BL can converge in fewer epochs and achieve higher accuracy than the deep DenseNet33BL at 30 epochs, while the deep DenseNet121 and DenseNet121BL were not as good as DenseNet33BL.

3.2. Model results

Various classifiers were used with different hyperparameters, and accuracy metrics were calculated based on the confusion matrix of the validation set. Each set of hyperparameters was associated with a specific set of accuracy metrics. The optimal metric was shown in Table 3, while the other metrics associated with the other hyperparameters were discussed in the Discussion section. From Tables 3 and it was observed that both Swin Transformer and Vision Transformer had lower overall accuracy (OA) compared to other classifiers, with Vision Transformer performing the worst (OA = 0.767, Kappa = 0.708). DenseNet33BL had the best classification results (OA = 0.901, Kappa = 0.892) among the classifiers. It was also noted that the performance of different layers of DenseNetBL was superior to that of the original DenseNet, which indicated that the bottleneck layer had effectively improved the classifier's ability to recognize tree species using small samples. Furthermore, Table 3 indicates that shallow classifiers are better suited for classifying tree species in smaller sample sizes.

3.3. Class activation mapping

This research involved the calculation of class activation mapping using Eq. (8) and Eq. (9) to identify the specific areas that the

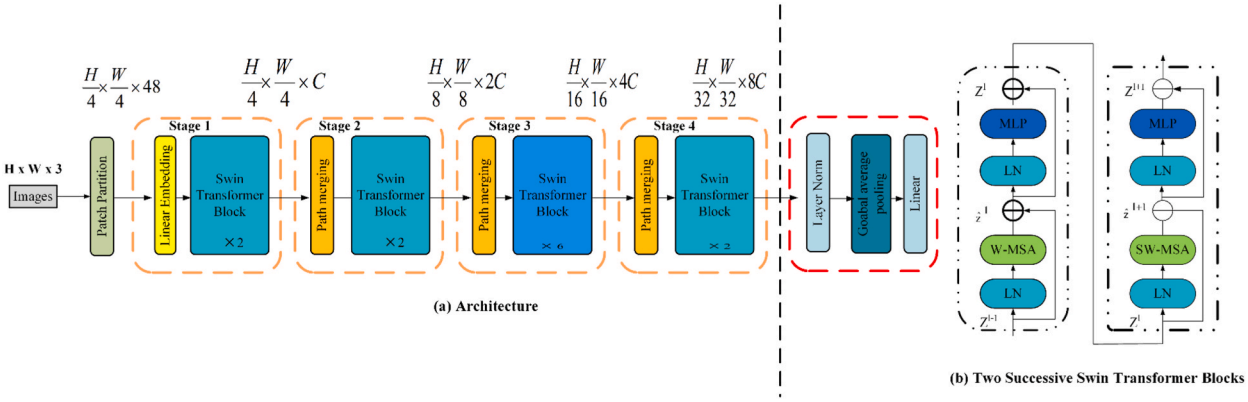


Fig. 7. Swin transformer structure.

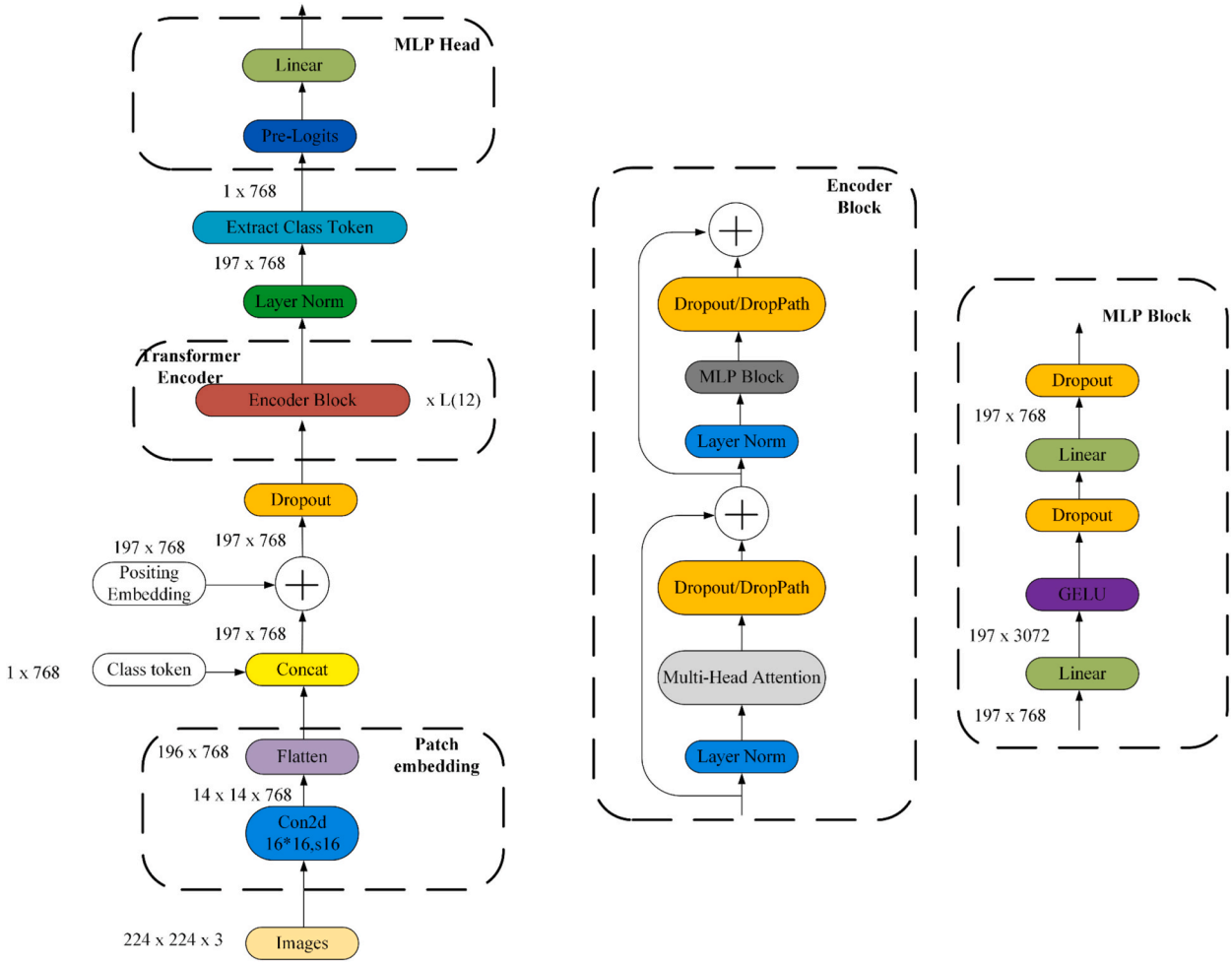


Fig. 8. Vision trasnformer structure.

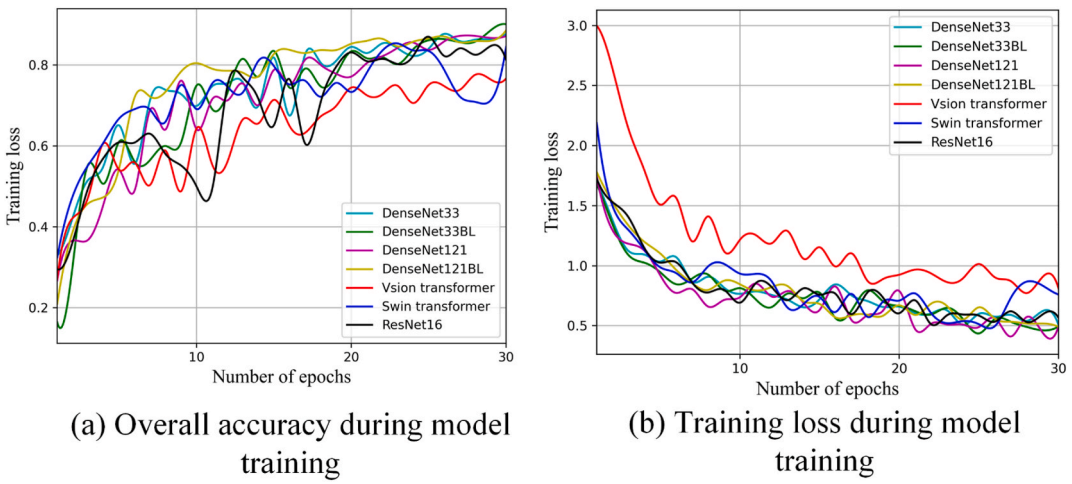


Fig. 9. (a)Training loss and (b) validation OA during model training. Curves were smoothed for better visualization.

Table 3
All accuracy metrics on the validation set.

Inputdata	Swintransformer	Visiontransformer	DenseNet33	DenseNet68	DenseNet117	DenseNet121	DenseNet169	DenseNet33BL	DenseNet68BL	DenseNet117BL	DenseNet121BL	DenseNet169BL	ResNet16
Bare land	0.805	0.774	0.889	0.903	0.871	0.832	0.856	0.926	0.854	0.862	0.853	0.879	0.868
Cinnamomum camphora	0.866	0.843	0.963	0.954	0.958	0.963	0.871	0.972	0.985	0.975	0.949	0.981	0.925
Ligustrum lucidum Ait	0.972	0.851	0.982	0.979	0.949	0.961	0.897	0.985	0.981	0.956	0.987	0.978	0.984
Photinia beauverdiana	0.815	0.668	0.845	0.862	0.836	0.888	0.798	0.897	0.886	0.895	0.892	0.885	0.835
Cupressus	0.879	0.848	0.896	0.861	0.879	0.848	0.859	0.933	0.928	0.924	0.865	0.931	0 .889
Osmanthus fragrans	0.143	0.043	0.029	0.039	0.049	0.029	0.015	0.057	0.082	0.079	0.079	0.071	0.056
OA	0.845	0.767	0.877	0.870	0.864	0.872	0.859	0.901	0.892	0.891	0.895	0.892	0.865
Kappa	0.805	0.708	0.845	0.821	0.819	0.825	0.792	0.892	0.835	0.831	0.886	0.829	0.813

classifier focused on when predicting tree species. The darker regions in the mapping indicate where the classifier paid more attention, and these regions can be interpreted as the features used by the classifier to distinguish between classes. Fig. 10 displays the class activation mapping of DenseNet33 and DenseNet33BL (with the best OA) for tree species classification. As shown in Fig. 10, as the classifier's depth increases, it becomes more focused on the tree canopy. Due to the increased non-linear feature extraction in the bottleneck layer, DenseNetBL pays more attention to the regions (as highlighted by the red boxes in Fig. 10(a, b, a1, b1)). The blue boxes (Fig. 10(c, c1)) in the upper left corner of DenseNet33(−1) and DenseNet33BL(-1) in Fig. 10 indicate that DenseNet33BL focuses more on the tree canopy during the final feature extraction, while DenseNet33 emphasizes the bare land more.

3.4. Typical forests tree species spatial distribution mapping

In Fig. 11, the segmentation boundaries of typical forests generated by SLIC are displayed. SLIC uses initial clustering parameters to create segmentation boundaries of similar sizes. The results in Fig. 11 (a) demonstrate that SLIC is successful in segmenting individual tree crowns in typical forest A, where the tree crown ranges for different species are consistent. However, in typical forest B (Fig. 11 (b)), SLIC struggles to segment regions with uneven tree crown sizes, such as the Cinnamomum camphora.

Spatial distribution maps of tree species were drawn using the DenseNet33, DenseNet33BL, DenseNet121, and DenseNet121BL models with the highest OA values as predicted by SLIC segmentation results. A reference map for forest tree species was also generated by manually delineating the tree crown boundaries and assigning categories. Fig. 12 (a,b,c,d) shows the map with uneven tree crown sizes reflected by different tree species in typical forest B. The segmentation effect of tree crowns was greatly affected, resulting in low accuracy of the tree species map. In contrast, typical forest A had a more regular distribution of tree species and consistent tree crown sizes, making the mapping effect better. Fig. 12 (a,b,c,d) also showed that the mapping effect of DenseNetBL was superior to that of the original DenseNet, given the same segmentation effect, because DenseNetBL had a higher OA. Moreover, the mapping effect of shallow networks DenseNet33 and DenseNet33BL was better than that of corresponding DenseNet121 and DenseNet121BL, as the OA of shallow classifiers was higher than that of deep classifiers.

Specifically, we selected and enlarged the drawing results of the same area in all tree species maps, and compared them with the manually drawn reference map. Specifically, we selected and enlarged the drawing results of the same area in all tree species maps, and compared them with the manually drawn reference map. The results showed that DenseNet33BL identified tree species in the selected

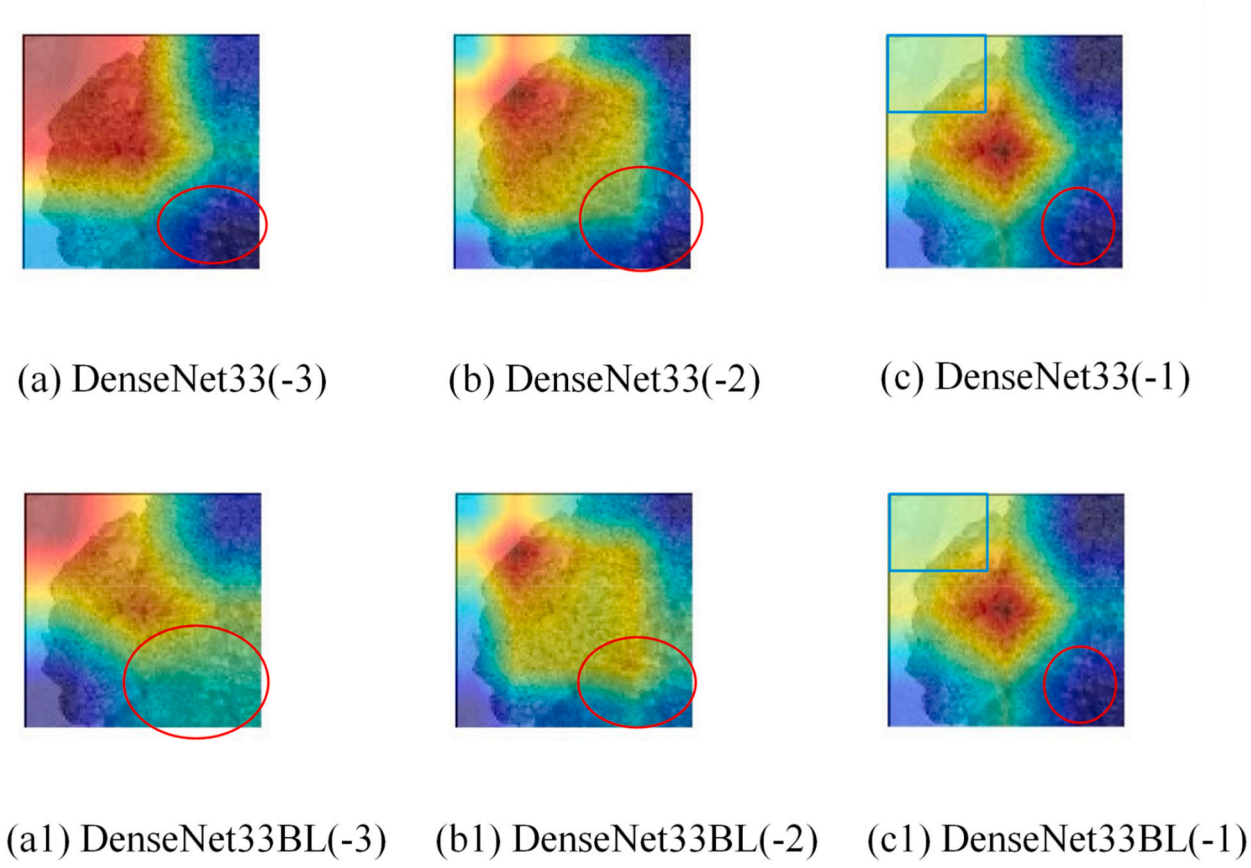


Fig. 10. Class activation mapping of DenseNet and DenseNetBL (-i denotes the inverse i-th extracted feature map).

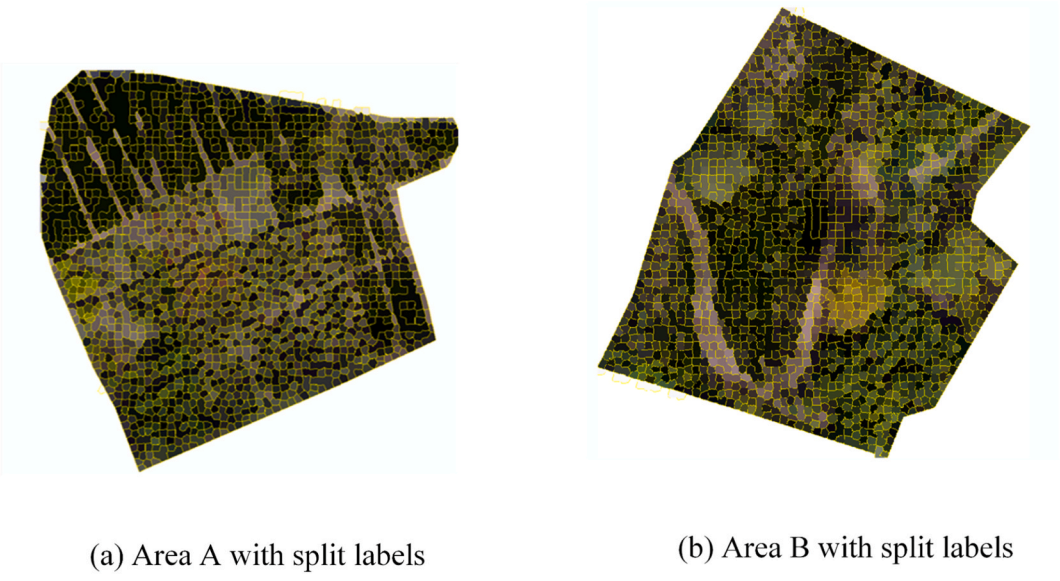


Fig. 11. Individual tree crown segmentation of typical forests.

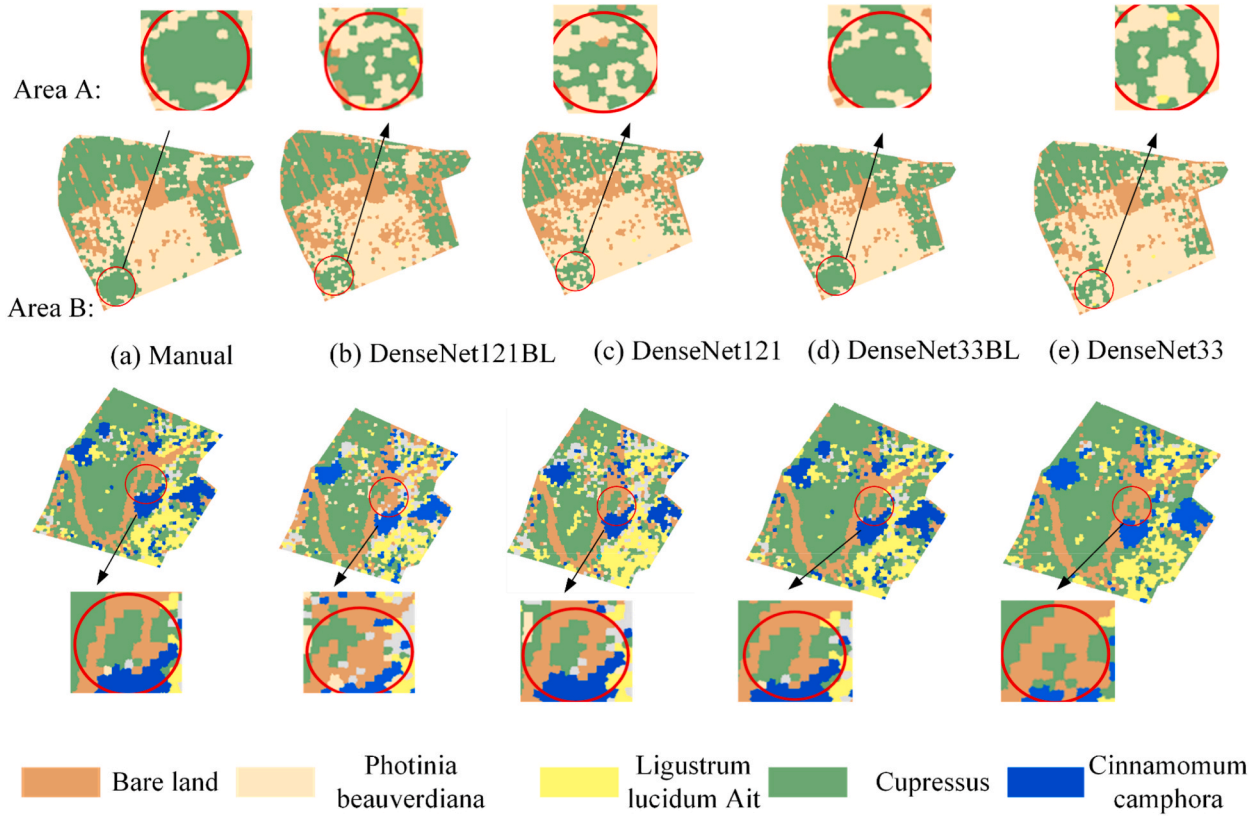


Fig. 12. Maps of typical forests tree species and enlarged detail drew by different classifiers and manual method.

area in both typical forest areas in all enlarged images, while DenseNet did not recognize them. In addition, the tree species identified by shallow DenseNet are significantly superior to deep DenseNet (Fig. 12 (d, e)). The above results show that the classification ability of shallow DenseNet is significantly stronger than that of deep DenseNet, while the ability of DenseNetBL inserted into the bottleneck layer to identify tree species has been further improved.

We utilized spatial distribution maps to calculate the area of tree species in two study areas and compiled the results in Tables 4 and 5, with the area unit being ha. The "Difference" column represents the variation between the area extracted via manual and classifier methods, while the "Total" column shows the sum of the differences. In Area A, DenseNet33BL exhibited the smallest difference with manual extraction at 0.08ha, while DenseNet121 had the largest difference at 0.18ha. Similarly, in Area B, DenseNet33BL showed the smallest difference with manual extraction at 0.17ha, while DenseNet121 had the largest difference at 0.22ha. Notably, due to the more complex forest environment in Area B, the performance of SLIC segmentation was inferior, leading to a larger variation between the area of tree species extracted by the classifier and manual methods than in Area A.

4. Discussion

Creating a tree species dataset through manual combination with field surveying is challenging due to limited images and range. Different classifiers struggle to fit the tree species feature distribution and identify unlabeled tree species in small samples. Current research on tree species classification focuses on developing new, complex, and advanced classifiers to improve performance. Advanced classifiers rely on effective feature extraction methods like Swin transformer's multi-region feature extraction to achieve higher accuracy compared to outdated and simple classifiers when trained with large samples and sufficient epochs. However, few studies have explored their ability to classify small sample tree species without using transfer learning and under small epochs. Previous research mainly focused on the final tree species classification accuracy without considering the impact of different optimizers and hyperparameters on classification.

In this study, we analyzed the performance of classifiers using two different optimizers and their respective hyperparameters (batch size and learning rates). We also tested the classification abilities of various layers from DenseNet, DenseNetBL, Vision transformer, and Swin transformer on two large datasets without pre-training weights. Moreover, we explored how classifier layers affect classification results.

4.1. Different hyperparameter

4.1.1. Different optimizer and learning rates

We conducted experiments by changing the batch size and learning rates of two optimizers, AdamW and SGD, and recorded the corresponding OA on the validation set. Firstly, we fixed the batch size of all adopted classifiers to 16 and divided the learning rates into {0.01, 0.001, 0.0001} to test the impact of these three learning rates on the performance of the two optimizers.

Fig. 13 (a, b) displays the OA variation curves for distinct learning rates and batch sizes. As observed in Fig. 13 (a), the optimization performance of AdamW in shallow networks is minimally impacted by various learning rates. On the other hand, Fig. 13 (b) reveals that the optimization effect of SGD is significantly influenced by the learning rate size. Fig. 13 (a, b) illustrates that the OA of deep classifiers (DenseNet169) and complex networks (Vision transformer and Swin transformer) is greatly affected by learning rates. Specifically, at a learning rate of 0.0001, the OA is remarkably lower compared to learning rates of 0.01 and 0.001 due to having more parameters to optimize. With smaller learning rates, the parameters update at a slower pace in small epochs, resulting in slower convergence of the classifiers. In section 4.2, different hyperparameters are explored.

4.1.2. Different batch sizes and the influence number of network layers

This study tested the effect of batch size on small sample tree species classification under fixed epochs using the optimal learning rates under 4.1.1, as shown in Fig. 14. In this study, we examined the impact of different batch sizes (2, 4, 8, 6, 32) on the accuracy of classifying small sample tree species. Our findings, as depicted in Fig. 14 (a, b), indicate that the accuracy improves as batch size increases and reaches its maximum at a batch size of 32 when learning rates and batch size are fixed.

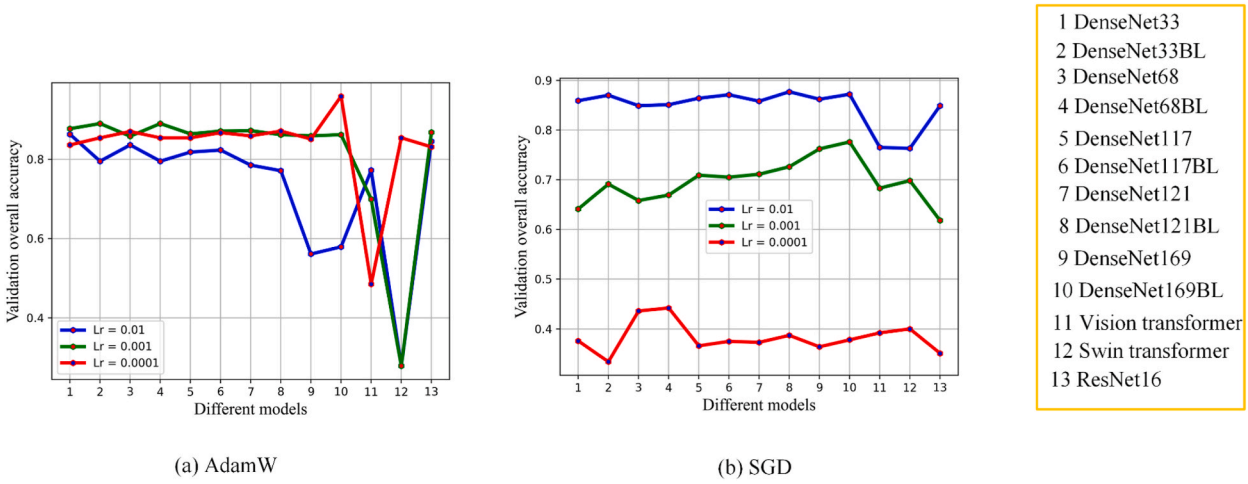
On the other hand, Figs. 13 (a) and Fig. 14 (a) show that increasing the number of layers in DenseNet and Dense-NetBL leads to a decrease in accuracy. While deeper networks can extract high-level tree species features, they struggle to recognize basic-level features and even high-level features cannot be detected. Furthermore, CNNs can be seen as a complex nonlinear fitting function, and in small datasets, deeper networks may overfit, causing a decrease in accuracy. Shallow networks, on the other hand, can better demonstrate the performance of classifiers on large datasets and achieve higher accuracy.

Table 4
Planting Area (ha) of tree species in Area B extracted by different methods.

Tree species	Manual	DenseNe-t121BL	Difference	Dense-Net121	Difference	DenseN-et33BL	Difference	Dense-Net33	Difference
Photinia beauverdiana	34.64	34.69	0.05	34.57	0.06	34.63	0.00	34.65	0.02
Cupressus	36.84	36.87	0.03	36.78	0.06	36.86	0.02	36.88	0.04
Ligustrum lucidum Ait	0.92	0.90	0.02	0.98	0.06	0.90	0.02	0.90	0.02
Cinnamomum camphora	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Osmanthus fragrans	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total			0.11		0.18		0.05		0.08

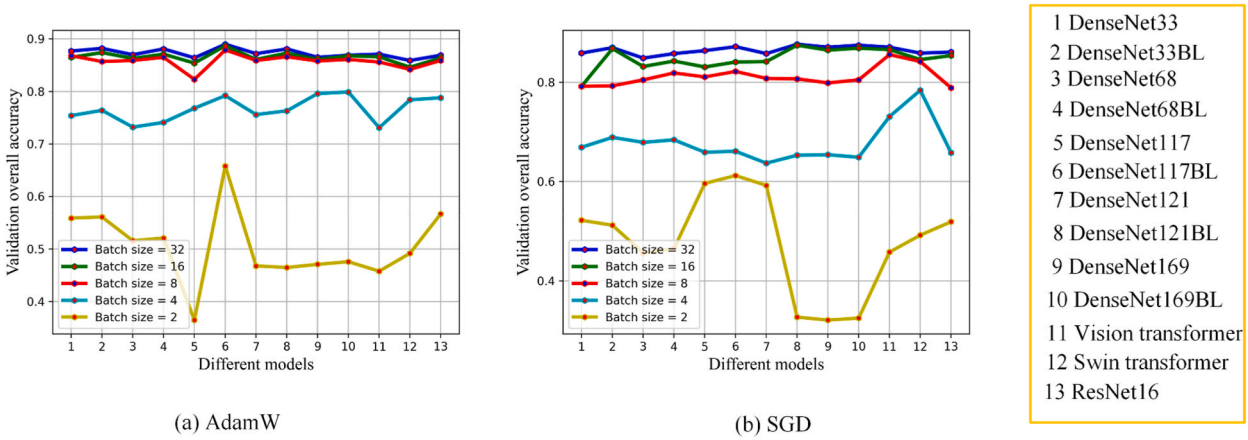
Table 5
Planting Area (ha) of tree species in Area B extracted by different methods.

Tree species	Manual	DenseNet121BL	Difference	DenseNet121	Difference	DenseNet33BL	Difference	DenseNet33	Difference
Photinia beauverdiana	1.07	1.17	0.10	1.16	0.09	0.99	0.08	0.97	0.10
Cupressus	18.09	18.05	0.04	18.06	0.03	18.07	0.02	18.05	0.04
Ligustrum lucidum Ait	13.54	13.55	0.02	13.59	0.05	13.51	0.03	13.52	0.02
Cinnamomum camphora	6.85	6.92	0.07	6.79	0.06	6.89	0.04	6.82	0.03
Osmanthus fragrans	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total			0.22		0.24		0.17		0.19



(a) AdamW (b) SGD

Fig. 13. OA curves corresponding to the different optimizers, learning rates, and number of networks layers.



(a) AdamW (b) SGD

Fig. 14. OA curves of DenseNet, DenseNetBL, Vision transformer, and Swintrasns-former corresponding to different batch sizes of different optimizers.

4.2. Large sample comparison experiment

There are two benchmark datasets for remote sensing scene classification: NWPU-RESISC45 and AID. The AID dataset, released in 2017 by Huazhong University of Science and Technology and Wuhan University, consists of 10,000 scene images across 30 categories.

Table 6
Ablation experiment.

Classifiers	AID	NWPU
DenseNet33	0.856	0.876
DenseNet33BL	0.889	0.914
Densenet68	0.878	0.892
Densenet68BL	0.896	0.917
Densenet117	0.876	0.898
Densenet117BL	0.886	0.906
Densenet121	0.89	0.91
DenseNet121BL	0.897	0.915
Densenet169	0.912	0.928
Densenet169BL	0.938	0.946
Swin transformer	0.785	0.764
Vision transformer	0.433	0.368
ResNet16	0.845	0.839

Each category contains approximately 220–420 images, and each image has a pixel size of about 600*600 [41]. The NWPU-RESISC45 dataset was released in 2017 by Northwestern Polytechnical University. It comprises 31,500 scene images divided into 45 categories, with roughly 700 images per category. Each pixel in the images is approximately 256*256 in size [42]. This study tested the classification abilities of DenseNetBL, DenseNet, Resnet, Vision transformer, and Swin transformer on two large-scale datasets. The classifiers were trained and validated using the best learning rate and batch size obtained, without pre-trained weights. Table 6 shows the obtained overall accuracy (OA). It also reveals that the bottleneck layer enhances the OA of large-scale classification due to two reasons. Firstly, the activation function added after the bottleneck layer improves the non-linear expression of the classifier network, allowing it to express more complex features. Secondly, the bottleneck layer acts as a convolutional layer that integrates multi-dimensional feature information and reduces the dimensionality of the feature map, enabling higher-level feature extraction.

According to Table 6, classifiers with more complex network architectures tend to perform better when applied to large datasets. This is because larger datasets contain a greater number of images and diverse image categories, which makes it harder for shallow classifiers to accurately determine categories. In contrast, deep classifiers can extract intricate high-level features from the extensive layers within their network architectures, allowing them to more accurately discern image categories. This is because deep classifiers can extract more nuanced representations, which results in more precise classification outcomes when dealing with the complex variations presented by diverse image categories within extensive datasets.

5. Conclusion

This study presents the development of a DenseNetBL classifier tailored for the classification of tree species with limited samples. This approach leverages bottleneck layers and the DenseNet architecture. A comprehensive evaluation was conducted to compare the classification performance between models trained on small and large datasets. The effectiveness of DenseNet, Swin Transformer, and Vision Transformer was examined in generating forest tree species maps. This investigation encompassed two optimization techniques: AdamW and SGD. The outcomes of this study revealed that, among all classifiers, DenseNet33BL exhibited the most promising results for small sample tree species classification (OA = 0.901, Kappa = 0.892) in instances where no pre-trained weights were employed. Notably, the shallower classifier demonstrated a notably superior capacity to classify tree species with limited samples compared to its deeper counterpart. Furthermore, a preference for AdamW emerged over SGD in the context of small sample classification, given its propensity to yield reduced variations in accuracy across diverse learning rates. In terms of larger datasets, both Vision Transformer and Swin Transformer exhibited significantly lower OA in comparison to DenseNet and DenseNetBL. Moreover, shallow variants of DenseNet and DenseNetBL yielded OA lower than those achieved by their deeper counterparts. Incorporating DenseNet33BL in conjunction with SLIC led to the extraction of tree species areas that closely approximated manually delineated regions. Remarkably, the disparity in Area A amounted to a mere 0.08 ha.

Looking ahead, our future research endeavors encompass the design of a semi-supervised algorithm featuring multiple classifiers. This will be complemented by the utilization of multi-temporal high spatial resolution images to enhance the accuracy of small sample tree species classification. Additionally, object-based methodologies will be harnessed to generate more expansive spatial distribution maps of forest tree species. These maps will serve as vital references for resource surveys, statistical analyses, and comprehensive management strategies related to forest tree species.

Funding

This work was supported in part by the Key Projects of Natural Science Projects in the Colleges and Universities in Anhui Province (KJ2021A1075), the General Items of Support Plan for Outstanding Young Talents in Colleges and Universities (gxyq2021217), the National Natural Science Foundation of China (41601455), the Graduate Innovation Fund of Anhui University of Science and Technology (2022CX2163), the Anhui Province University Natural Science Foundation (KJ2020A0717), the Chuzhou University Talent Foundation Project (2020qd31), and the Open Fund of State Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing of Wuhan University (21R04).

Data availability statement

Data will be made available on request.
Declaration of interest's statement:
The authors declare no conflicts of interest.

CRediT authorship contribution statement

Ni Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Tao Pu:** Writing – original draft, Validation, Software, Investigation, Data curation. **Yali Zhang:** Validation, Resources, Conceptualization. **Yuchan Liu:** Supervision, Resources. **Zeyu Zhang:** Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. Song, F. He, Y. Liu, A non-invasive learning branch to capture leaf-image attention for tree species classification, Multimed. Tool. Appl. 81 (2022) 13961–13978, https://doi.org/10.1007/s11042-022-12036-6.

[2] A. Ghosh, F.E. Fassnacht, P.K. Joshi, B. Koch, A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales, Int. J. Appl. Earth Obs. Geoinf. 26 (2014) 49–63, https://doi.org/10.1016/j.jag.2013.05.017.

[3] X. Feng, P. Li, A tree species mapping method from UAV images over urban area using similarity in tree-crown object histograms, Rem. Sens. 11 (2019), https://doi.org/10.3390/rs11171982, 1982.

[4] H. Qin, W. Zhou, Y. Yao, W. Wang, Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data, Rem. Sens. Environ. 280 (2022), 113143, https://doi.org/10.1016/j.rse.2022.113143.

[5] B. Aygunes, R.G. Cinbis, S. Aksoy, Weakly supervised instance attention for multisource fine-grained object recognition with an application to tree species classification, ISPRS J. Photogramm. Rem. Sens. 176 (2021) 262–274, https://doi.org/10.1016/j.isprsjprs.2021.03.021.

[6] M. Lechner, A. Dostálová, M. Hollaus, C. Atzberger, M. Immitzer, Combination of sentinel-1 and sentinel-2 data for tree species classification in a central European biosphere reserve, Rem. Sens. 14 (2022) 2687, https://doi.org/10.3390/rs14112687.

[7] S. Illarionova, A. Trekin, V. Ignatiev, I. Oseledets, Neural-based hierarchical approach for detailed dominant forest species classification by multispectral satellite imagery, IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 14 (2021) 1810–1820, https://doi.org/10.1109/JSTARS.2020.3048372.

[8] K. Wang, T. Wang, X. Liu, A review: individual tree species classification using integrated airborne LiDAR and optical imagery with a focus on the urban environment, Forests 10 (2019) 1, https://doi.org/10.3390/f10010001.

[9] Y. Sun, Q. Xin, J. Huang, B. Huang, 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. Top. Appl. Earth Obs. Rem. Sens. 12 (2019) 4415–4425, https://doi.org/10.1109/JSTARS.2019.2950721.

[10] Z. He, D. He, Bilinear squeeze-and-excitation network for fine-grained classification of tree species, Geosci. Rem. Sens. Lett. IEEE 18 (2021) 1139–1143, https://doi.org/10.1109/LGRS.2020.2994952.

[11] X. Guo, H. Li, L. Jing, P. Wang, Individual tree species classification based on convolutional neural networks and multitemporal high-resolution remote sensing images, Sensors 22 (2022) 3157, https://doi.org/10.3390/s22093157.

[12] F. Schiefer, T. Kattenborn, A. Frick, J. Frey, P. Schall, B. Koch, S. Schmidtlein, Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks, ISPRS J. Photogramm. Rem. Sens. 170 (2020) 205–215, https://doi.org/10.1016/j.isprsjprs.2020.10.015.

[13] K. Cao, X. Zhang, An improved res-UNet model for tree species classification using airborne high-resolution images, Rem. Sens. 12 (2020) 1128, https://doi.org/10.3390/rs12071128.

[14] R. Pu, Mapping tree species using advanced remote sensing technologies: a state-of-the-art review and perspective, J. Rem. Sens. 2021 (2021) 1–26, https://doi.org/10.34133/2021/9812624.

[15] L.E.C. La Rosa, C. Sothe, R.Q. Feitosa, C.M. de Almeida, M.B. Schimalski, D.A.B. Oliveira, Multi-task fully convolutional network for tree species mapping in dense forests using small training hyperspectral data, ISPRS J. Photogramm. Rem. Sens. 179 (2021) 35–49, https://doi.org/10.1016/j.isprsjprs.2021.07.001.

[16] F.E. Fassnacht, H. Latifi, K. Stereńczak, A. Modzelewska, M. Lefsky, L.T. Waser, C. Straub, A. Ghosh, Review of studies on tree species classification from remotely sensed data, Rem. Sens. Environ. 186 (2016) 64–87, https://doi.org/10.1016/j.rse.2016.08.013.

[17] M. Deur, M. Gašparović, I. Balenović, An evaluation of pixel- and object-based tree species classification in mixed deciduous forests using pansharpened very high spatial resolution satellite imagery, Rem. Sens. 13 (2021) 1868, https://doi.org/10.3390/rs13101868.

[18] J.A. Correa Martins, G. Menezes, W. Gonçalves, D.A. Sant'Ana, L.P. Osco, V. Liesenberg, J. Li, L. Ma, P.T. Oliveira, G. Astolfi, H. Pistori, J.M. Junior, Machine learning and SLIC for Tree Canopies segmentation in urban areas, Ecol. Inf. 66 (2021), 101465, https://doi.org/10.1016/j.ecoinf.2021.101465.

[19] A.M. Roy, J. Bhaduri, Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4, Comput. Electron. Agric. 193 (2022), 106694, https://doi.org/10.1016/j.compag.2022.106694.

[20] D. Bandyopadhyay, S. Mukherjee, Tree Species Classification from Hyperspectral Data Using Graph-Regularized Neural Networks, 2022, https://doi.org/10.48550/arXiv.2208.08675.

[21] L. Gu, F. He, S. Yang, Crop classification based on deep learning in northeast China using SAR and optical imagery, 2019 SAR in Big Data Era (BIGSARDATA) (2019) 1–4, https://doi.org/10.1109/BIGSARDATA.2019.8858437.

[22] L. Chen, Y. Wei, Z. Yao, E. Chen, X. Zhang, Data augmentation in prototypical networks for forest tree species classification using airborne hyperspectral images, IEEE Trans. Geosci. Rem. Sens. 60 (2022) 1–16, https://doi.org/10.1109/TGRS.2022.3168054.

[23] J. Ni, Z. Shao, Z. Zhang, M. Hou, J. Zhou, L. Fang, Y. Zhang, Ldp-Net, An unsupervised pansharpening network based on learnable degradation processes, IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 15 (2022) 5468–5479, https://doi.org/10.1109/JSTARS.2022.3188181.

[24] W. Wang, X. Yuan, Z. Chen, X. Wu, Z. Gao, Weak-light image enhancement method based on adaptive local gamma transform and color compensation, J. Sens. 2021 (2021), e5563698, https://doi.org/10.1155/2021/5563698.

[25] H.S. Elsayed, O.M. Saad, M.S. Soliman, Y. Chen, H.A. Youness, Attention-based fully convolutional DenseNet for earthquake detection, IEEE Trans. Geosci. Rem. Sens. 60 (2022) 1–10, https://doi.org/10.1109/TGRS.2022.3194196.

[26] L. Shan, W. Wang, DenseNet-based land cover classification network with deep fusion, Geosci. Rem. Sens. Lett. IEEE 19 (2022) 1–5, https://doi.org/10.1109/LGRS.2020.3042199.

[27] X. Gao, T. Chen, R. Niu, A. Plaza, Recognition and mapping of landslide using a fully convolutional DenseNet and influencing factors, IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 14 (2021) 7881–7894, https://doi.org/10.1109/JSTARS.2021.3101203.

[28] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual Dense Network for Image Super-resolution, 2018, pp. 2472–2481. https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Residual_Dense_Network_CVPR_2018_paper.html. (Accessed 24 January 2023).

[29] Z. Wu, C. Shen, A. van den Hengel, Wider or deeper: revisiting the ResNet model for visual recognition, Pattern Recogn. 90 (2019) 119–133, https://doi.org/10.1016/j.patcog.2019.01.006.

[30] S. Targ, D. Almeida, K. Lyman, Resnet in Resnet: Generalizing Residual Architectures, 2016, https://doi.org/10.48550/arXiv.1603.08029.

[31] W. Zhang, G. Wang, J. Qi, G. Wang, T. Zhang, Research on the extraction of wind turbine all over the China based on domestic satellite remote sensing data, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, IEEE, Brussels, Belgium, 2021, pp. 4167–4170, https://doi.org/10.1109/IGARSS47720.2021.9553559.

[32] W. Yang, H. Huang, Z. Zhang, X. Chen, K. Huang, S. Zhang, Towards rich feature discovery with class activation maps augmentation for person Re-identification, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, CA, USA, 2019, pp. 1389–1398, https://doi.org/10.1109/CVPR.2019.00148.

[33] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-Cam, Visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2020) 336–359, https://doi.org/10.1007/s11263-019-01228-7.

[34] Y. Zhang, K. Liu, Y. Dong, K. Wu, X. Hu, Semisupervised classification based on SLIC segmentation for hyperspectral image, Geosci. Rem. Sens. Lett. IEEE 17 (2020) 1440–1444, https://doi.org/10.1109/LGRS.2019.2945546.

[35] J. Yin, T. Wang, Y. Du, X. Liu, L. Zhou, J. Yang, SLIC superpixel segmentation for polarimetric SAR images, IEEE Trans. Geosci. Rem. Sens. 60 (2022) 1–17, https://doi.org/10.1109/TGRS.2020.3047126.
[36] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lučić, C. Schmid, ViViT: A Video Vision Transformer (2021) 6836–6846. (Accessed 5 September 2022).
[37] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin Transformer, Hierarchical Vision Transformer Using Shifted Windows, 2021, pp. 10012–10022. https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.html. (Accessed 5 September 2022).
[38] H. Su, B. Zhao, Q. Du, P. Du, Kernel collaborative representation with local correlation features for hyperspectral image classification, IEEE Trans. Geosci. Rem. Sens. 57 (2019) 1230–1241, https://doi.org/10.1109/TGRS.2018.2866190.
[39] X. Wang, Y. Wang, C. Zhou, L. Yin, X. Feng, Urban forest monitoring based on multiple features at the single tree scale by UAV, Urban For. Urban Green. 58 (2021), 126958, https://doi.org/10.1016/j.ufug.2020.126958.
[40] N. Wang, T. Pu, T. Chen, Simple linear iterative clustering and ConvNeXt for mapping vectorize tree species, JARS 17 (2023), 038502, https://doi.org/10.1117/1.JRS.17.038502.
[41] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, X. Lu, AID: a benchmark data set for performance evaluation of aerial scene classification, IEEE Trans. Geosci. Rem. Sens. 55 (2017) 3965–3981, https://doi.org/10.1109/TGRS.2017.2685945.
[42] G. Cheng, J. Han, X. Lu, Remote sensing image scene classification: benchmark and state of the art, Proc. IEEE 105 (2017) 1865–1883, https://doi.org/10.1109/JPROC.2017.2675998.