forests

MDPI

Article

An Improved Wood Recognition Method Based on the One-Class Algorithm

Jie He [1,†], Yongke Sun [1,*,†], Chunjiang Yu [1], Yong Cao [2,*], Youjie Zhao [1] and Guanben Du [3]

[1] School of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming 650224, China
[2] International Engineering and Technology Institute, Hongkong 999077, China
[3] Yunnan Provincial Key Laboratory of Wood Adhesives and Glued Products, Southwest Forestry University, Kunming 650224, China
* Correspondence: sunyongke@swfu.edu.cn (Y.S.); cn_caoyong@126.com (Y.C.)
† These authors contributed equally to this work.

Abstract: Wood recognition is necessary for work in the wood trade activities. The advantage of the one-class wood classification method is more generalization, and it only needs positive samples and does not need negative samples in the training phase, so it is suitable for rare wood species inspection. This paper proposed an improved method based on the one-class support vector machine (OCSVM) for wood species recognition. It uses cross-section images acquired with a magnifying glass, which uses a pre-trained VGG16 model for feature extraction, a normal distribution test for key features filtering, and OCSVM to determine the wood species. The results showed that the approach achieved a mean recall of 0.842 for both positive and negative samples, which indicates this method has good performance for wood recognition. In a negative public dataset, the negative recall reached as high as 0.989, which showed that this method has good generalization.

Keywords: wood recognition; transfer learning; one-class classification

Citation: He, J.; Sun, Y.; Yu, C.; Cao, Y.; Zhao, Y.; Du, G. An Improved Wood Recognition Method Based on the One-Class Algorithm. *Forests* **2022**, *13*, 1350. https://doi.org/10.3390/f13091350

Academic Editors: Veronica De Micco, Maks Merela and Angela Balzano

Received: 13 August 2022
Accepted: 22 August 2022
Published: 25 August 2022

1. Introduction

Wood identification is helpful for curbing illegal logging and protecting rare wood species, which play an essential role in timber trade activities. It can also help consumers protect their rights when buying rare and valuable wood products. China is the largest importer of timber in the world; as a result, wood species inspection is a heavy task. The traditional wood species identification methods depend on human expertise; it is inefficient because it is manually performed.

For the purpose of improving the efficiency of wood identification, researchers attempted to identify the wood species using computer technology. Because the cross-section images contain most of the identification features, most of them use the wood cross-section image to identify the species. The previous research works fell mainly into two kinds, the first is based on the traditional machine learning algorithm, and the second is based on the deep learning algorithm.

The first kind of method uses the traditional machine learning algorithm to identify the wood species. For example, Andrade employed support vector machines (SVM) to classify 21 wood species and the accuracy reached 97% [1]. Mujahid Mohamad proposed a method using e-nose with K-Nearest Neighbors (KNN) analysis to classify two kinds of agarwood in two mediums with 94.5% accuracy [2]. Xutai CUI proposed using machine learning classification methods including partial least squares-discrimination analysis (PLS-DA), random forest (RF), and other traditional methods to classify spectral data of eight wood species, with the highest correct classification rate (CCR) achieving 98.55% [3]. Hang-jun Wang proposed a new Gabor-based wood recognition method to classify 24 wood species, with the highest recognition rate of 97.3% [4].

The second kind of method is using a deep learning algorithm to identify the wood species. For example, Prabu Ravindran proposed a VGG16-based [5] wood cross-section images classification model trained by transfer learning to identify the wood species of 10 neotropical trees with the highest accuracy of 89.8% [6]. Fabijańska Anna adopted an approach of using a residual convolutional encoder network in a sliding window setting to classify 14 European tree species, and the correct recognition rate reached 93% [7]. Xinjie Tang proposed a method, the minified Squeeze-Network method, used for transfer learning, which is trained to identify 100 commonly trading wood types found in Malaysia with Top-1 accuracy of 78% [8]. Liu presented a split-shuffle-residual (SSR)-based convolutional neural network (CNN) that can extract features automatically from wood images for real-time classification of rubber wood boards and had an accuracy of 94.86% [9].

Although these two kinds of methods reduced the requirements of the operators and succeeded in some species, it is still hard to use in some rare and valuable wood species because the number of rare woods is stubbornly small. Furthermore, these methods cannot be used to identify the unknown wood species that do not take part in the model training. Any unknown species will be erroneously classified into one of the trained species. It affects the trust in the wood identification system seriously.

One-class classification (OCC) algorithms only require positive samples to train models and can filter unknown species with small training datasets. For example, Lukas Ruff proposed a new deep one-class classification method named deep support vector data description (Deep-SVDD). The experimental results on the MNIST and CIFAR-10 image datasets show the effectiveness of the performance of the Deep-SVDD method [10]. Paul Bergmann proposed an uninformed student-based one-class learning network and applied it to anomaly detection and one-class classification and improved over state-of-the-art methods on many datasets [11]. Wenpeng Hu proposed a new one-class classification method called HRN (H Regularization with 2-Norm instance-level normalization) and applied it to one-class classification. The experimental results show that HRN significantly outperforms the existing state-of-the-art deep or Non-deep learning models [12]. Although these methods have achieved good performance on public datasets, they do not have good generalization, and these public datasets are highly distinguishable and easy to identify [13].

In this article, we propose a new method that is based on the OCC method, which is able to distinguish the unknown species. It extracts the image features using the VGG16 model and recognizes the species with the OCC method. The object of this study is to classify cross-sectional images of wood. Not only are the macro-wood characteristics normally distributed, but also the distribution of many image features is normal [14–16]. Therefore, we also propose a normal distribution test method as a feature selector to improve the recognition performance. In the experiment on five rare wood species, the results showed that this method reached good accuracy. In the testing on a public wood image dataset, it displaced a good generalization performance.

The highlights of our article are as follows.

- We created a "Wood Image" dataset containing 585 images of rare and precious wood species from the Herbarium of Southwest Forestry University, and we have removed blurry and unclear images. The wood cross-section images in the dataset contain wood ray and tube hole distributions and color information for wood cross-sections. Annotation of these images is performed by experienced experts with specialized knowledge. This dataset does not just serve as a benchmark to evaluate the performance of the proposed method, but also provide a reference for follow-up research.
- In this paper, we first proposed a model combining a deep learning feature extractor based on transfer learning and a feature filter based on the normal distribution test with a one-class classification algorithm and applied it to the field of wood recognition. Our model can automatically extract features from wood cross-sectional images and quickly classify rare and valuable wood species.

- The classification performance on the "Wood Images" dataset shows that the proposed wood identification method based on one-class algorithm outperforms other traditional one-class classification methods and classic deep one-class classification methods. Tested on the public datasets, our model has a good generalization and can recognize the wood species which is untrained and put it into unknown wood species.

2. Materials

Five species of wood samples come from the wood herbarium of southwest forestry university, including *Dalbergia tucurensis*, *Dalbergia stevensonii*, *Diospyros crassiflora*, *Millettia stuhlmannii*, and *Cassia siamea*. All of them are valuable wood and are in demand for identification in commercial wood activity. Thirty blocks per species were collected, and each block takes 4∼5 images. The wood species and image number are shown in Table 1.

Table 1. Number of five wood species.

ID	Specie	Number
1	*Dalbergia tucurensis*	101
2	*Dalbergia stevensonii*	116
3	*Diospyros crassiflora*	109
4	*Millettia stuhlmannii*	124
5	*Cassia siamea*	135

Before acquiring the cross-section image, the cross-section was polished by sandpaper at 400 grits, 800 grits, and 1000 grits, respectively. Clearing the dust on the surface with a brush, and taking the picture using a camera(cell phone, OPPO Reno 5G) with a 20× lens in front of it. The cross-section image is shown in Figure 1. In total, 585 images were chosen and used in the experiments.

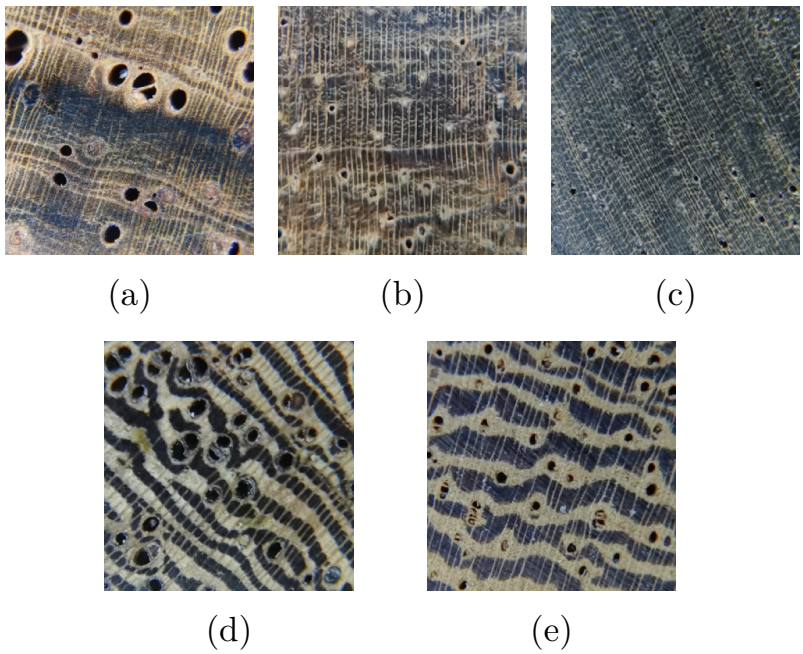

Figure 1. Cross-section image of wood samples. (**a**): *Dalbergia tucurensis*, (**b**): *Dalbergia stevensonii*, (**c**): *Diospyros crassiflora*, (**d**): *Millettia stuhlmannii*, (**e**): *Cassia siamea*.

3. Methods

The proposed method contains three steps as Figure 2 shows. The first step is image feature extracting, in which a pre-trained VGG16 model was used to extract the image

features, which amounts to 4096 data items. The second step is key feature picking, it filters the image features and pick up some of which in the special location as the key features. The key feature location trained by normal distribution test (NDT) method. The third step is classification, in which, a kind of OCC method was used to make a judgment on whether the image belongs to the species.

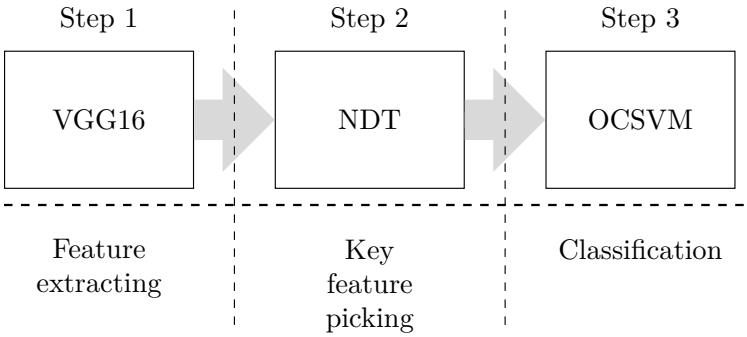

Figure 2. Training step of wood species classification model.

3.1. Feature Extracting

Feature extraction is the initial stage of the species recognition, and it affects the accuracy significantly. In the previous study, the gray-level co-occurrence matrix (GLCM)[17] and the histogram of oriented gradient (HOG)[18] were often used to extract the image features [19,20], but these features are weak in robustness. Later, the neural networks were adopted to extract the features, which improved the robustness and obtained better results. However, the extracted feature from neural networks depends on the size and quality of the training dataset. Collecting enough samples from rare and valuable wood for neural network training is tough. Transfer learning explored a new clue for small dataset training, which training model on a pre-trained model, and succeeded in many image classification fields [21]. The reason for success of transfer learning probably is that the pre-trained model is an organic structure system, and it can extract the same features from the same images.

The pre-trained neural network can be used to extract features directly for one-class classification. In experiments, we adopt a pre-trained VGG16 model to extract image features because it has good performance in previous wood classification work [22,23].

We removed the last classification layer of the VGG16 model, and kept the pre-trained weights of the convolutional layers, and fully connected (FC) layers. The modified structure is shown in Figure 3, the input of our modified model is a 224 × 224 pixel RGB image, and the output from fc7 layer is the image feature which contains 4096-dimension data.

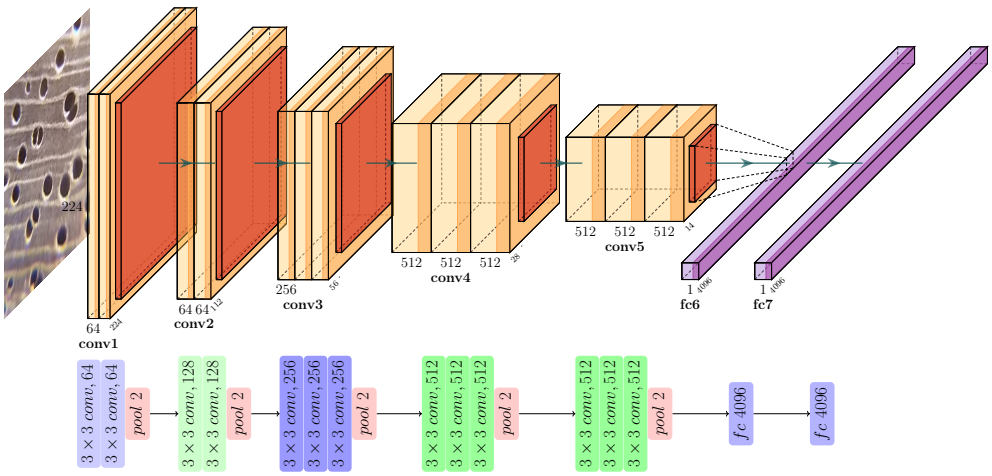

Figure 3. Modified VGG16 structure.

3.2. Key Feature Filtering

A feature of the VGG16 model is the 4096 dimension data items, some of which hold key features that are important for species classification, and some of which have no contribution to species classification even interference with the accuracy of classification [24,25]. Filtering the VGG16 output and picking up the key features for classification can decrease the interference and computational overhead.

The key feature filtering model is generated in the training, and the work process is shown in Figure 4. The origin input $X_{(m,n)}$ is the features of all training samples, which is an $m \times n$ dimension matrix, m is the rows indicating the number of samples, and n is columns indicating the index of the feature. The NDT model f is a filter that will be used in the inference process, which is trained with a normal distribution test.

$$X_{key} = f(X) = \{X_n \mid n \in N_{filter}\} \tag{1}$$

where N is a set of column indexes, and the data of each column, calculated by Formula (2), followed normal distribution.

$$N_{filter}(X_{(m,n)}) = \{n \mid p(X_{(:,n)}) > \tau\} \tag{2}$$

where p is the normal distribution test function. Using the filter function $p(X)$ tests each column and picks up the column indexes, where the features follow the normal distribution. Order and concatenate the picked data as the key features.

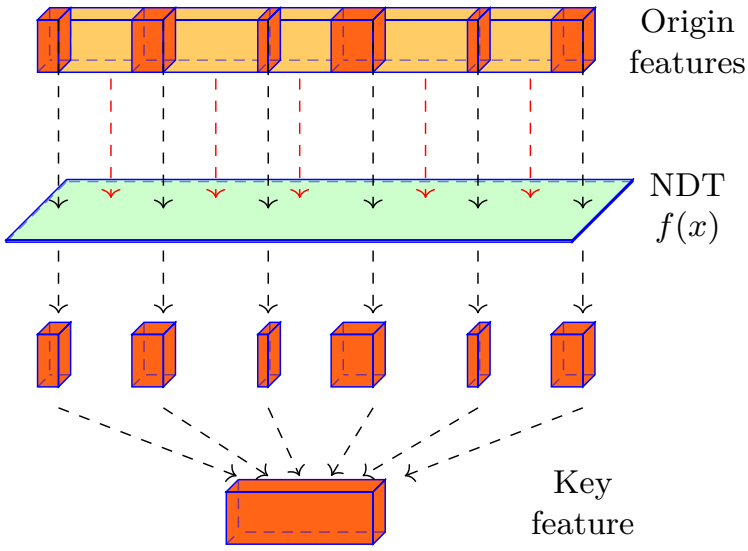

Figure 4. Key feature picking.

In experiments, we use Formula (3) to test each column and find out the index position that the same specie response follows the normal distribution, and this formula is recommended and has good accuracy in previous work [26].

$$p(X) = \frac{T}{n^2 S} \tag{3}$$

where

$$T = \sum (i - \frac{n+1}{2}) x_{(i,ord)} \tag{4}$$

$$S = \frac{\sum (x_i - \bar{x})^2}{n} \tag{5}$$

where $x_i \in X$, and $x_{(i,ord)}$ is the ordered X.

The key feature indexes were calculated as Algorithm 1 in the training phase and the key features were filtered using Algorithm 2 in the inference phase.

Algorithm 1: Key indexes finding.

Input: X: The feature matrix of the training dataset
Output: X_key: column indexes
/* Obtain the size of the matrix */
rows, cols=X.shape $N\_filted\_col = [\,]$ **for** $c \in cols$ **do**
  data=X[:,c] // Obtain the $c-th$ column of X
  _data_ord=order(data) $l = len(data)$ T=0 **for** $i \in l$ **do**
    $T+ = (i - \frac{len+1}{2}) \times \_data\_ord[i]$
  end
  $S = 0$ **for** $i \in l$ **do**
    $S+ = (data[i] - \bar{x})^2$
  end
  $S = S/l$ $p = T/(n^2 S)$ **if** $p \geqslant \tau$ **then**
    $N\_filter\_col.append(c)$
  end
end
return N_filter_col

Algorithm 2: Key feature Filtering.

Input: X: features of one sample.
Input: N_filter: indexes of features follow the normal distribution.
Output: Key features
$X\_key = [\,]$ **for** $i \in range(len(N\_filter))$ **do**
  X_key.append($X[i]$)
end
return X_key

3.3. Classification

The OCC algorithm detects a boundary of samples which was used to judge whether a data belongs to a group. Figure 5 shows a principle of the OCC, in which, the red points are training data, and the blue points are other data. The OCC method is used to find a boundary that can separate these two kinds of points. This method only needs positive samples in the training process and does not need negative samples. Therefore, it is suitable for small datasets, especially for rare wood species, because it is hard to collect enough images of the rare and valuable wood to train a traditional neural network. OCSVM [27], IF [28], and LOF [29] is usually adopted the one-class method in recent years, and more articles showed that the OCSVM has the best robustness among them [30–32].

OCSVM constructs a hyper-plane that was used to classify the data. The principle as Formula (6) shows, in which, w is a normal vector of the hyper-plane, $\Phi(x)$ is a function maps points on the sample space to the feature space, and b is the compensation vector.

$$f(x) = sign(w \cdot \Phi(x) - b) \tag{6}$$

The objective function of OCSVM is finding a minimum hyper-space that surrounds the positive samples in each dimension, as shown in Formula (7). The constraints of the objective function are represented in Formula (8).

$$y = min(\frac{1}{2}||w||^2 + \lambda \cdot \sum_{i=1}^{n}(\xi_i - b)) \tag{7}$$

$$s.t : w \cdot \Phi(x_i) \geqslant b - \xi_i,\ \xi_i > 0 \tag{8}$$

where, λ is Lagrange Multiplier.

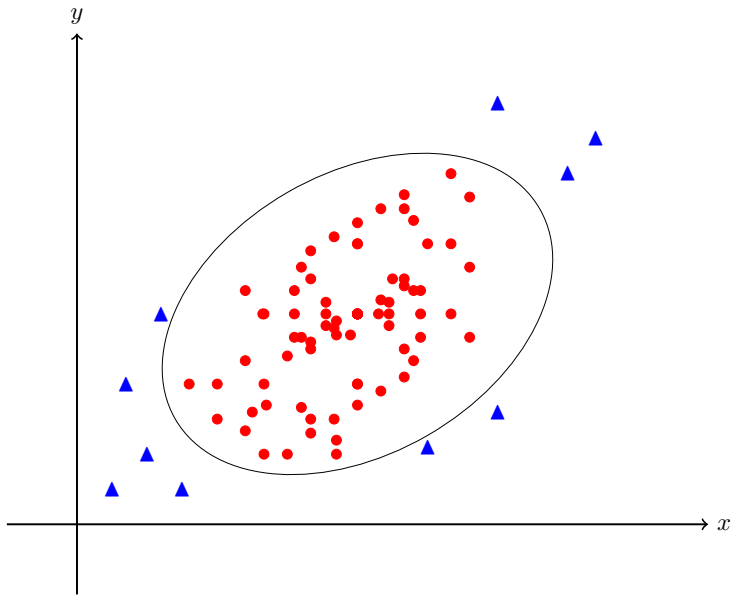

Figure 5. One-class classification. The red points are training sample, and blue points are other data.

According to the Lagrange function, the optimized function $f(x)$ can be represented as Formula (9). α is a Lagrange multiplier vector.

$$\begin{aligned} f(x) &= sign(\sum_{i=1}^{n} \alpha_i \cdot K(x_i, x_j) - \rho) \\ \rho &= \sum_{i=1}^{n} \alpha_i \cdot K(x_i, x_j) \end{aligned} \tag{9}$$

where, K is the Gaussian kernel function as Formula (10) shows.

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \tag{10}$$

where, σ is standard deviation of the x.

3.4. Evaluating

Cross-validation (CV) is an effective method used to evaluate the model performance in a limited dataset [33]. Five-fold CV was used in the experiments to split the dataset into five groups, using 4 of the 5 to train the model and using the remaining 1 to test the model each time. Finally, using the mean of the five validations as the model performance.

Accuracy, precision, recall, and F1-score are four classic measurements often used to describe the classification model [34–36]. They are defined as Formulas (11)–(14), in which, TP indicates the number of true positives, TN indicates the number of true negatives, FP indicates the number of false positives, and FN indicates the number of false negative [37].

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

$$F1 - score = \frac{2 * Precision * Recall}{Recall + Precision} \tag{14}$$

4. Results and Discussions

The proposed method was used to build models for each wood species, and the models were tested separately. For each test, we split the samples into two parts, the first part was marked as the positive sample, which contains the specified wood species. Furthermore, the second part was marked as the negative sample, which contains the other wood species. For each model, the number of the positive sample around is 100, and the number of the negative sample approximate is 485. Table 2 is the recognition results of five wood species using our method, and it showed that our proposed method is a feasible one-class wood species classification. The mean accuracy of models is 0.848, the mean of recall of models is 0.848, the mean precision of models is 0.896, and the mean of F1-score of models is 0.856.

Table 2. Precision, Recall, and F1-score of five wood species.

Species	Accuracy	Precision	Recall	F1-score
1	0.93	0.93	0.93	0.92
2	0.80	0.88	0.80	0.82
3	0.95	0.95	0.95	0.95
4	0.81	0.87	0.81	0.82
5	0.75	0.85	0.75	0.77
Mean	0.848	0.896	0.848	0.856

1: *Dalbergia tucurensis*; 2: *Dalbergia stevensonii*; 3: *Diospyros crassiflora*; 4: *Millettia stuhlmannii*; 5: *Cassia siamea*.

The recall is an appropriate evaluation criterion for recognizing special species, and the positive recall indicates the ability of the model picked the positive sample from the amassed and mixed samples set. Table 3 is the recall of the positive sample, in which, the mean of recall reached 0.842, and the recall of four in five species reached 0.90. It implied that our method has a good ability to pick up the assigned wood species from mixed samples.

Table 3. Our model hyper parameters and recall of five wood species classifier.

Species	Kernel	Gamma	Nu	Positive Recall	Negative Recall
Dalbergia tucurensis	rbf	0.01	0.1	0.61	0.99
Dalbergia stevensonii	rbf	auto	0.1	0.90	0.78
Diospyros crassiflora	rbf	auto	0.1	0.90	0.96
Millettia stuhlmannii	rbf	auto	0.1	0.90	0.78
Cassia siamea	rbf	auto	0.1	0.90	0.70
Mean				0.842	0.842

The advantage of one-class classification is that it has the ability to reject negative samples. For the purpose of evaluating the model performance of rejecting the negative samples, around 585 images were tested in the experiments. The negative recall is the best measurement for rejecting negative samples, as Table 3 shows. The results show that the mean of negative recall is 0.842, for *Dalbergia tucurensis*, the negative recall is as high as 0.99, which means the *Dalbergia tucurensis* model has good accuracy in rejecting the sample faking as *Dalbergia tucurensis*. These results provide evidence that our proposed method has a good ability to reject negative samples.

Considering both the positive recall and the negative recall, it supports the idea that the pre-trained VGG16 neural network has the ability to extract similar features from similar wood images. The notable finding is that the model was trained only by one species. It not only recognized the positive samples but also rejected the negative samples. This study represented a new approach that recognizes the wood species only use the positive

samples and do not need to collect a mass of negative samples. It saves time and labor because the species of the tree are huge, and it is hard to collect all species samples.

4.1. The Comparsion of Classifier

Three different classifiers were compared in the experiments, and the results showed that our method has the best performance in accuracy, precision, recall, and F1-score. Figures 6–9 illustrated the difference between them. It is easy to find that our method performance is higher than OCSVM and IF and LOF, and the previous study also displayed this phenomenon [38,39].

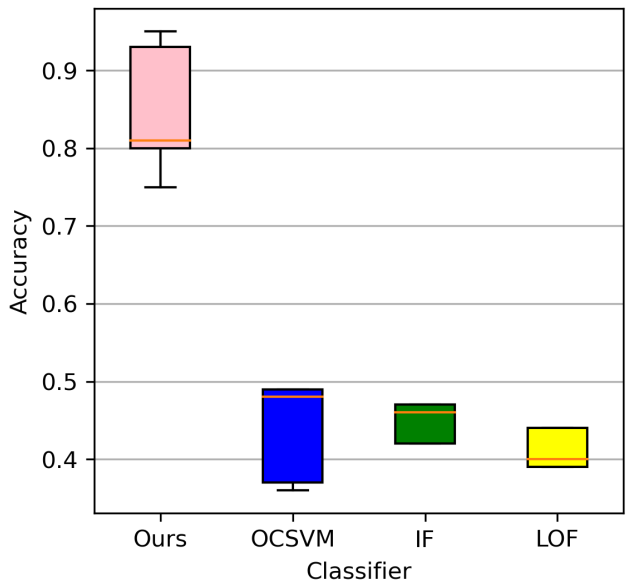

Figure 6. Comparison of the accuracy of our method and three different classifiers.

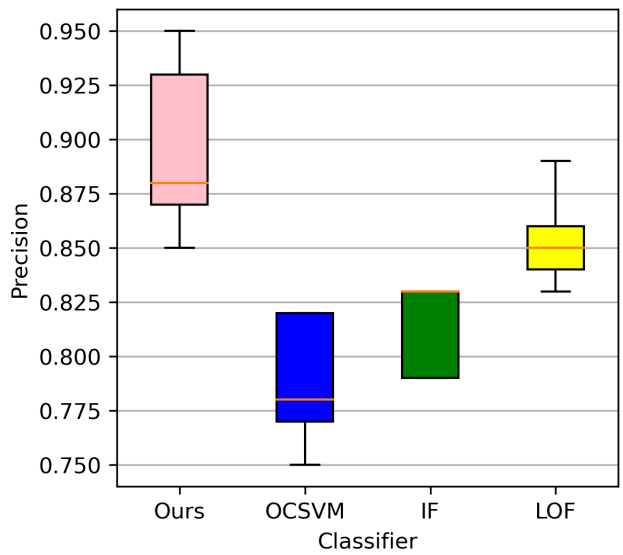

Figure 7. Comparison of the precision of our method and three different classifiers.

In previous work, Local Binary Pattern (LBP), Fourier descriptor, Gabor filter and the Wavelet descriptor were used to extract the image feature, and the F1-score reached 0.81 [40,41]. Our method used a pre-trained neural network (VGG16) instead of these extractors and obtained a similar F1-score of 0.87. It was shown that the pre-trained neural network is an organic structure that can find common features from similar images.

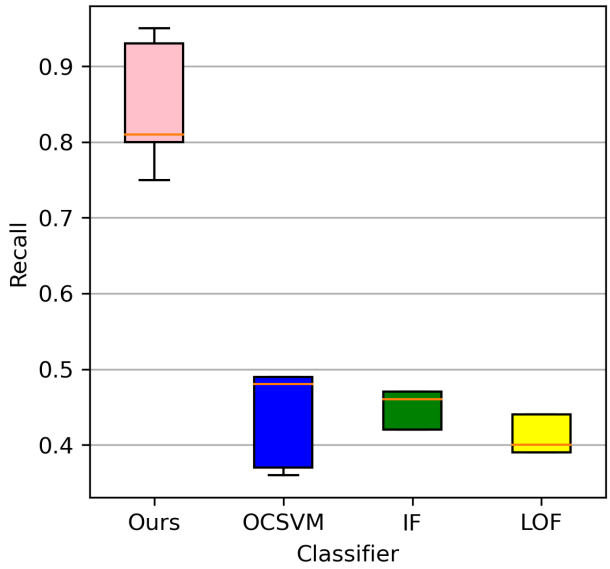

Figure 8. Comparison of the recall of our method and three different classifiers.

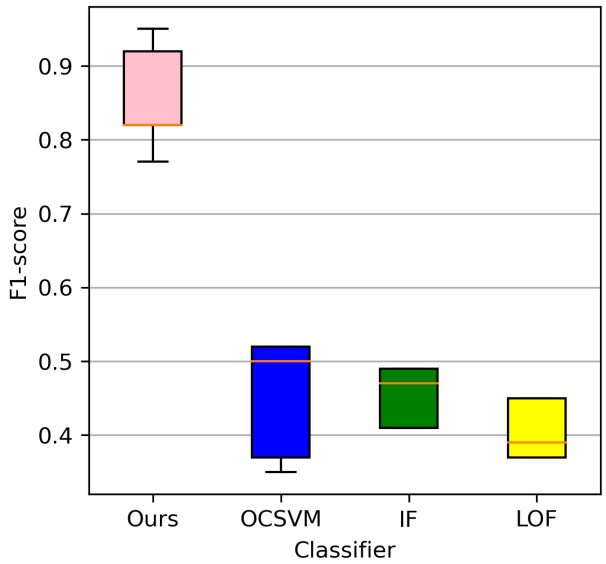

Figure 9. Comparison of the F1-score of our method and three different classifiers.

OCSVM divides the dataset with a hyperplane, it is suitable for the sparse features with robustness [42]. LOF calculates the class boundary using a distance, it is also based on the hypersphere to determine whether a data belongs to a group. Previous one-class classification research has succeeded in many fields; for instance, it was used to detect the anomaly signal in the diffusion process of semiconductor manufacturing [43], was used to classify condition monitoring of marine machinery systems [44], and has good accuracy. However, another study reported that the LOF is not very good for classifying the cyanobacterial fluorescence signals. Furthermore, the author said the performance of the LOF is low due to the small distance between the normal data and the outlier [45]. This implied that the features of wood are sparse.

4.2. Comparison with Deep-SVDD

The Deep-SVDD is a one-class method based on the deep neural network, and it had good performance in open datasets [10]. For comparison with Deep-SVDD, we add negative images in each species around 10%, and split the dataset into a training sub-

dataset and a testing sub-dataset with a ratio of 8:2. The results of the testing dataset are shown in Figure 10; it shows that the accuracy of our method is higher than Deep-SVDD in majority species, especially in the fifth species, our method is far higher than Deep-SVDD. Furthermore, it shows that the Deep-SVDD is not suitable for all wood species, indicating that Deep-SVDD's generalization performance is not as good as our method.

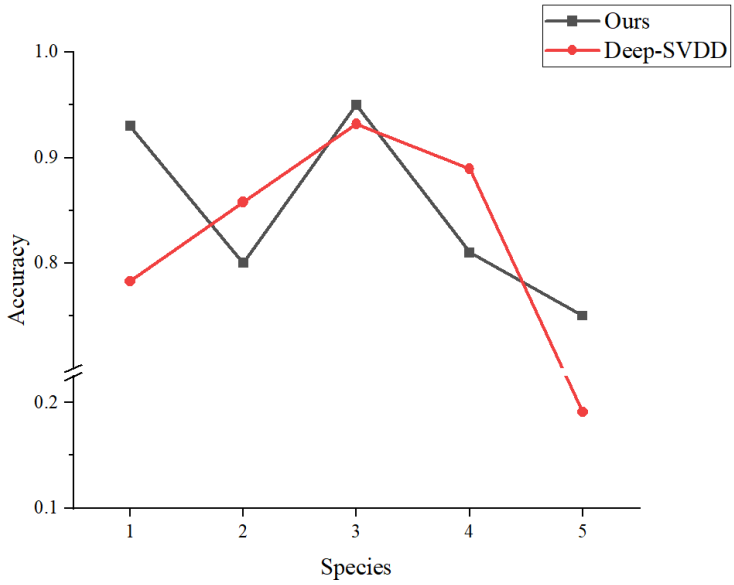

Figure 10. The difference of our methd with the Deep-SVDD.

We split the training dataset and testing dataset randomly 10 times, and obtained the mean results in Table 4. It not only presented the difference in the accuracy but also presented the difference in the precision, recall, and F1-score. A significant difference is that the precision, recall, and F1-score is 0 in the fifth wood species. Formulas (12)–(14) indicate that the variable TP is 0, which means the Deep-SVDD model does not have the ability to recognize the fifth wood species.

Table 4. Precision, Recall, and F1-score of five wood species with the Deep-SVDD.

Species	Accuracy		Precision		Recall		F1-Score	
	Ours	Deep-SVDD	Ours	Deep-SVDD	Ours	Deep-SVDD	Ours	Deep-SVDD
1	0.93	0.78	0.93	0.90	0.93	0.83	0.92	0.86
2	0.80	0.85	0.88	0.95	0.8	0.87	0.82	0.91
3	0.95	0.93	0.95	0.98	0.95	0.94	0.95	0.96
4	0.81	0.88	0.87	0.91	0.81	0.97	0.82	0.94
5	0.75	0.19	0.85	0	0.75	0	0.77	0

1: *Dalbergia tucurensis*; 2: *Dalbergia stevensonii*; 3 :*Diospyros crassiflora*; 4: *Millettia stuhlmannii*; 5: *Cassia siamea*.

This comparison shows that our method is better than the Deep-SVDD for three reasons. First, our method does not need the negative samples, but it is necessary for the Deep-SVDD. Second, the measurements of our method are higher than the Deep-SVDD in the majority of species. Third, the generalization performance of our method is better than the Deep-SVDD.

4.3. Features of Wood

Because we acquired images with 20X magnification, a cross-section image cannot contain all anatomical features that are used to distinguish the wood species. We can easily find the difference between the images in Figure 11, in which the distributions of the pore are significantly different. For example, Figure 11a has a three-pore feature, Figure 11b has

a two-pore feature, and Figure 11c has a fewer pore feature in it. Figure 11d–f are *Dalbergia stevensonii* and Figure 11g–i are *Diospyros crassiflora*, Figure 11j–l are *Millettia stuhlmannii*, Figure 11m–o are *Cassia siamea*, and they all have different features in different image, especially the feature of pore distribution as above.

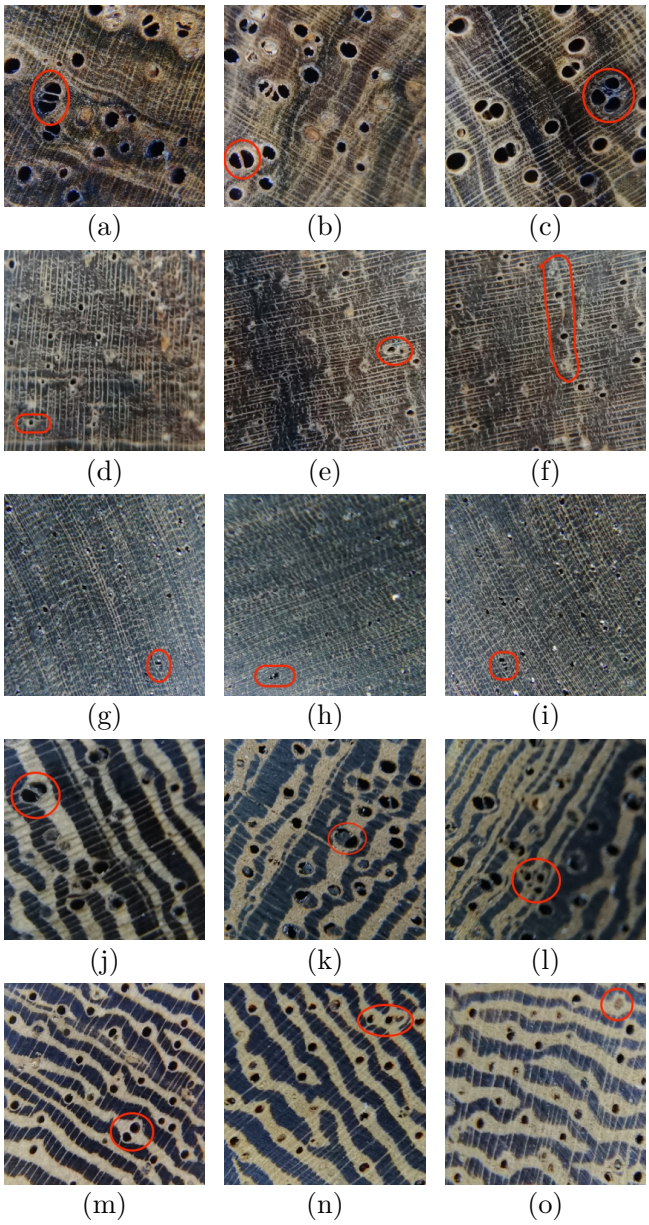

Figure 11. Different pore distribution in different wood species. (**a**–**c**): *Dalbergias tucurensis*, (**d**–**f**): *Dalbergia stevensonii*, (**g**–**i**): *Diospyros crassiflora*, (**j**–**l**): *Millettia stuhlmannii*, (**m**–**o**): *Cassia siamea*.

It can be seen that most of the features marked on the image are sparse, so we choose the classification method that is suitable for sparse features. Figures 6–9 show the experimental results of our method compared with three conventional methods. It is obvious that the proposed method outperformed the others in accuracy, precision, recall, and F1-score. The experimental results proved that our method is suitable for sparse features and works best.

For other wood species images, they all have different anatomy and texture features in different images. In the majority, the features of the wood species are sparse. In the real laboratory, wood species identification often needs multiple slice images at the same magnification because one slice image has not concluded all anatomy features. Thus,

using multiple images to classify and using voting to determine the wood species is a feasible solution.

4.4. Feature Filtering

Feature filtering is necessary before the one classification due to the features from the neural network being a vast dataset. Some of them are species features and some of them are noise that has interference when training the one classifier model. A normal distribution test method was used in experiments, which picked responses from the locations which have a similar output for species from fully connected layers.

The pre-trained VGG16 network was trained by the ImageNet dataset, which contains 1000 categories and generates similar outputs for the same category, and it indicates that giving the VGG16 model similar input will obtain similar output. Wood images from the same species often are similar and theoretically, the outputs also are similar. Table 5 demonstrated the feature number of VGG16 and the number of filtered $t \geqslant 0.05$, it shows that the feature was shrunk significantly.

Table 5. Feature dimensions of five trees after dimension reduction.

Specie	Origin	Filtered
Dalbergia tucurensis	4096	236
Dalbergia stevensonii	4096	1022
Diospyros crassiflora	4096	1866
Millettia stuhlmannii	4096	1052
Cassia siamea	4096	593

The resolve is $n \geqslant 5$ will obtain $R \geqslant 0.99$ when $r = 0.75$ which is the worst recall in experiments.

Figures 12–15 show that the filtered feature data obtains better experimental performance than the original feature data in wood species of the 2–5. When identifying tree species 1, the filtered feature dimension is reduced a lot, which greatly reduces the amount of computation with almost no loss of accuracy.

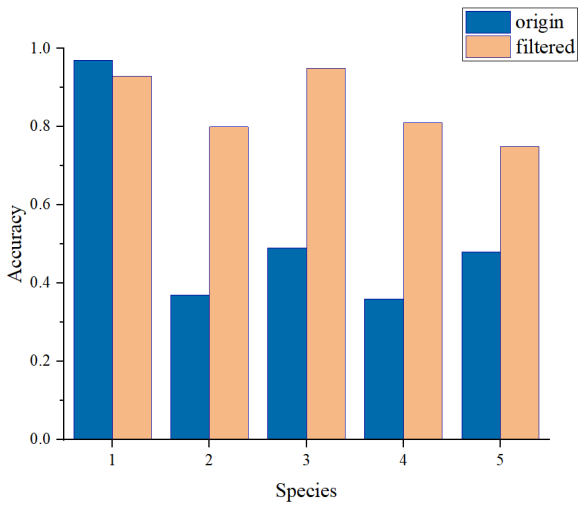

Figure 12. Comparison of experimental results of classification accuracy between original feature data and filtered data.

The *Dalbergia tucurensis* has more filtered features than others, it implied that the majority of features of this species are centered in a small area, which can easily be acquired by a single magnification. Other species have more features after being filtered, which means their features spread over a wide area and are often distributed in different magnified images.

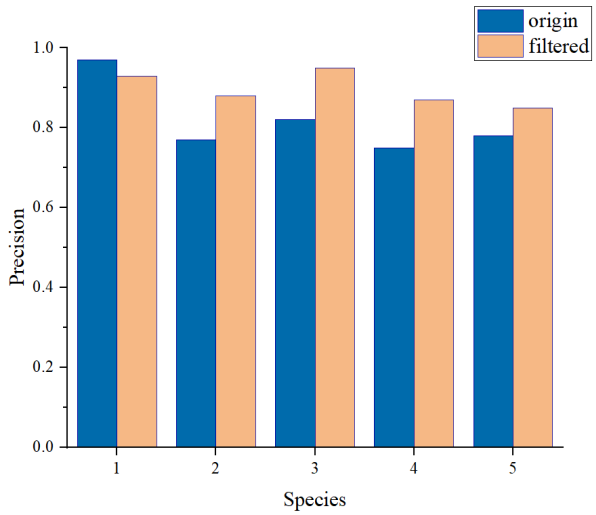

Figure 13. Comparison of experimental results of classification precision between original feature data and filtered data.

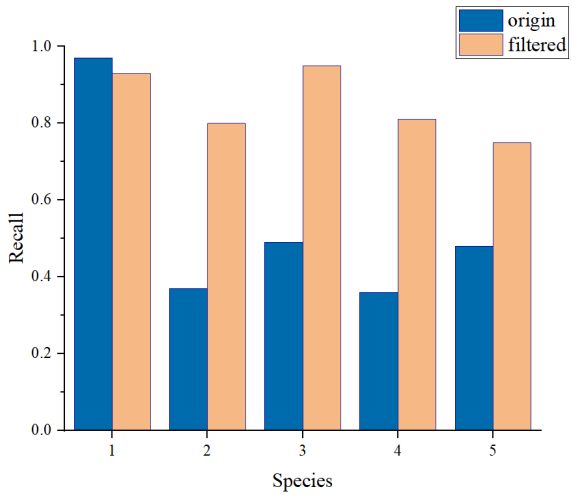

Figure 14. Comparison of experimental results of classification recall between original feature data and filtered data.

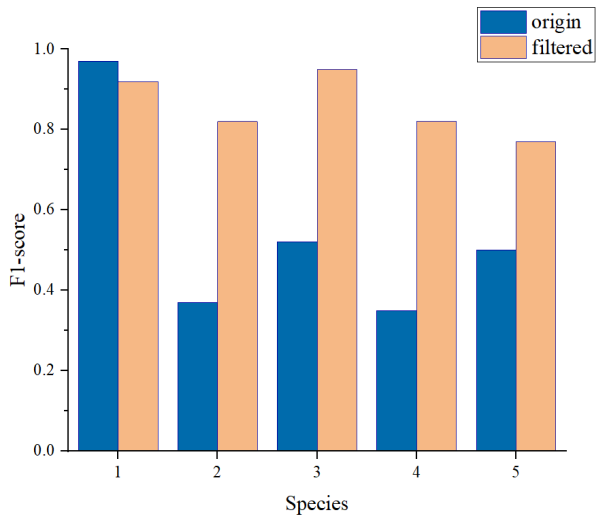

Figure 15. Comparison of experimental results of classification F1-score between original feature data and filtered data.

For the purpose of obtaining a certain result, a feasible approach is to use nine images from different positions of the same sample to identify wood species and treat the sample as a specific species when more than five images are classified as the same species. Define the recall is r and image number is n, then final recall R can be represented as Formula (15).

$$R = 1 - (1 - r)^n \tag{15}$$

4.5. Generalization for Negative Samples

We tested our built models on a public dataset [40], which contains 440 images from 11 wood species. Figure 16 is the sample images of the public dataset. All images also acquired by magnifying glass.

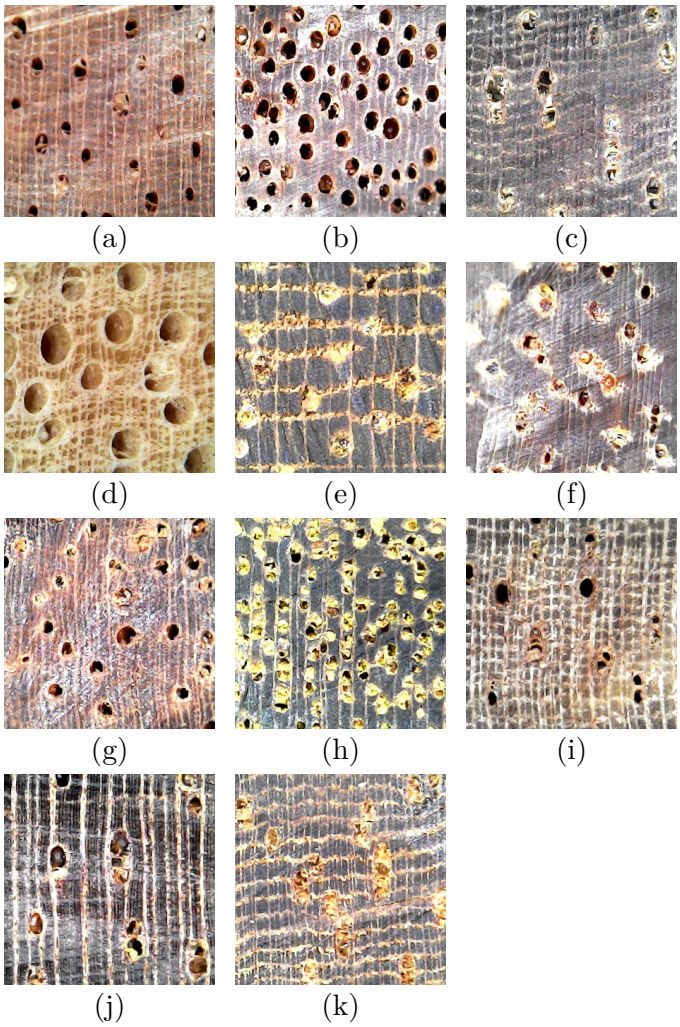

Figure 16. Sample of public dataset. (**a**): *Allantoma decandra*, (**b**): *Caraipa densifolia*, (**c**): *Cariniana micrantha*, (**d**): *Caryocar villosum*, (**e**): *Clarisia racemosa*, (**f**): *Dipteryx odorata*, (**g**): *Goupia glabra*, (**h**): *Handroanthus incanus*, (**i**): *Lueheopsis duckeana*, (**j**): *Osteophleum playspermum*, (**k**): *Pouteria caimito*.

The wood recognition models are not trained from these species, they all are negative samples for the models. The test results showed that our models have a high negative recall to pick them out, as shown in Table 6. There is strong evidence that our method has good generalization performance for unknown species, and has a good recognition ability to pick up the assigned species in the complex environment. It is inferred that this method has a better ability to extract common features from the wood species, and it is a feasible approach for wood recognition in real application.

Table 6. Recall of species in public dataset.

Species	Negative Recall				
	$Model_1$	$Model_2$	$Model_3$	$Model_4$	$Model_5$
(a)	0.97	0.98	0.99	1	1
(b)	1	0.98	1	1	0.99
(c)	1	0.98	0.99	1	1
(d)	1	0.95	1	1	1
(e)	1	0.97	1	0.97	0.94
(f)	1	0.98	1	1	1
(g)	1	1	1	1	1
(h)	1	0.98	1	1	1
(i)	0.97	0.97	1	0.99	0.98
(j)	1	0.99	1	1	1
(k)	1	0.86	0.98	1	1
Mean	0.995	0.967	0.996	0.996	0.992

$model_1$: for *Dalbergia tucurensis*; $model_2$: for *Dalbergia stevensonii*; $model_3$: for *Diospyros crassiflora*; $model_4$: for *Millettia stuhlmannii*; $model_5$: for *Cassia siamea*.

5. Conclusions

We proposed a new method that can improve the accuracy of one-class classification. Using the VGG16 pre-trained model to extract the wood features and reduce the calculating cost, using a normal distribution test method to pick the key features and decrease the interference, using the OCSVM to classify the species and building a model which can recognize the assigned wood species from the complex environment that contains other unknown species. We evaluated the performance of the models of five wood species, in our datasets, the mean of recall reached 0.848, and it inferred that this method has good accuracy. In a public dataset, the negative recall was as high as 0.989 on average, which implied that this method has good generalization performance. Even though this method only needs positive images for training the classifier, it has good generalization performance in recognizing the negative samples. It affords a convenient method for assigned wood inspection in timber trading that contains various wood species.

Author Contributions: Conceptualization, J.H. and Y.S.; methodology, Y.C.; software, J.H.; validation, J.H., Y.S. and Y.C.; formal analysis, Y.S.; investigation, Y.Z.; resources, Y.S.; data curation, Y.S.; writing—original draft preparation, J.H.; writing—review and editing, Y.S., C.Y. and G.D.; visualization, J.H.; supervision, Y.C.; project administration, G.D.; funding acquisition, Y.C., Y.Z. and G.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (31960142, 32071688 and 61962055) and the Major Project of Science and Technology of Yunnan Province under Grant number 202002AD080002.

Data Availability Statement: We have uploaded image dataset to GitHub (https://github.com/JieHeswfu/woodimages, (accessed on)).

Acknowledgments: Thanks to Qiujian etc. from the Herbarium of Southwest Forestry University for providing the wood samples.

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

References

1. de Andrade, B.G.; Basso, V.M.; de Figueiredo Latorraca, J.V. Machine vision for field-level wood identification. *IAWA J.* **2020**, *41*, 681–698. [CrossRef]
2. Mohamad, M.; Najib, M.S.; Tajuddin, S.N.; Daud, S.M.; Majid, N.F.H.; Zaib, S.; Zahari, M.F. kNN: Classification of Agarwood Types in Oil and Wooden Using E-nose. In *Proceedings of the 6th International Conference on Electrical, Control and Computer Engineering*; Springer: Singapore, 2022; pp. 575–586.
3. Cui, X.; Wang, Q.; Wei, K.; Teng, G.; Xu, X. Laser-induced breakdown spectroscopy for the classification of wood materials using machine learning methods combined with feature selection. *Plasma Sci. Technol.* **2021**, *23*, 055505. [CrossRef]

4. Wang, H.j.; Qi, H.n.; Wang, X.F. A new Gabor based approach for wood recognition. *Neurocomputing* **2013**, *116*, 192–200. [CrossRef]
5. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
6. Ravindran, P.; Costa, A.; Soares, R.; Wiedenhoeft, A.C. Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. *Plant Methods* **2018**, *14*, 25. [CrossRef] [PubMed]
7. Fabijańska, A.; Danek, M.; Barniak, J. Wood species automatic identification from wood core images with a residual convolutional neural network. *Comput. Electron. Agric.* **2021**, *181*, 105941. [CrossRef]
8. Tang, X.J.; Tay, Y.H.; Siam, N.A.; Lim, S.C. MyWood-ID: Automated macroscopic wood identification system using smartphone and macro-lens. In Proceedings of the 2018 International Conference on Computational Intelligence and Intelligent Systems, Phuket, Thailanda, 17–19 November 2018; pp. 37–43. [CrossRef]
9. Liu, S.; Jiang, W.; Wu, L.; Wen, H.; Liu, M.; Wang, Y. Real-Time Classification of Rubber Wood Boards Using an SSR-Based CNN. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8725–8734. [CrossRef]
10. Ruff, L.; Vandermeulen, R.; Goernitz, N.; Deecke, L.; Siddiqui, S.A.; Binder, A.; Müller, E.; Kloft, M. Deep one-class classification. In Proceedings of the International Conference on Machine Learning (PMLR), Stockholm, Sweden, 10–15 July 2018; pp. 4393–4402.
11. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. Uninformed Students: Student-Teacher Anomaly Detection with Discriminative Latent Embeddings. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 4182–4191.
12. Hu, W.; Wang, M.; Qin, Q.; Ma, J.; Liu, B. HRN: A Holistic Approach to One Class Learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 19111–19124.
13. Perera, P.; Oza, P.; Patel, V.M. One-Class Classification: A Survey. *arXiv* **2021**, arXiv:2101.03064.
14. Comaniciu, D.; Meer, P. Robust analysis of feature spaces: Color image segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 1997; pp. 750–755.
15. Cofer, R.H.; Kozaitis, S.P. Image chain assessment for feature extraction. In *Visual Information Processing XII*; SPIE: Bellingham, WA, USA, 2003; Volume 5108, pp. 287–294.
16. Thi, K.M. Face Recognition for Human Identification using BRISK Feature and Normal Distribution Model. *Int. J. Trend Sci. Res. Dev.* **2019**, *3*, 1139-1143.
17. Mokji, M.; Bakar, S.A. Gray Level Co-Occurrence Matrix Computation Based On Haar Wavelet. In Proceedings of the Computer Graphics, Imaging and Visualisation (CGIV 2007), Bangkok, Thailand, 14–16 August 2007; pp. 273–279. [CrossRef]
18. Kim, S.; Cho, K. Trade-off between accuracy and speed for pedestrian detection using HOG feature. In Proceedings of the 2013 IEEE Third International Conference on Consumer Electronics ¿ Berlin (ICCE-Berlin), Berlin, Germany, 9–11 September 2013; pp. 207–209. [CrossRef]
19. Sethy, P.K.; Barpanda, N.K.; Rath, A.K.; Behera, S.K. Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* **2020**, *175*, 105527. [CrossRef]
20. Xu, J.L.; Gowen, A.A. Spatial: Pectral analysis method using texture features combined with PCA for information extraction in hyperspectral images. *J. Chemom.* **2020**, *34*, e3132. [CrossRef]
21. Setiawan, W.; Utoyo, M.I.; Rulaningtyas, R. Transfer learning with multiple pre-trained network for fundus classification. *Telkomnika* **2020**, *18*, 1382–1388. [CrossRef]
22. Zhao, Z.; Yang, X.; Ge, Z.; Guo, H.; Zhou, Y. Wood Microscopic Image Identification Method Based on Convolution Neural Network. *BioResources* **2021**, *16*, 4986–4999. [CrossRef]
23. Shustrov, D.; Eerola, T.; Lensu, L.; Kälviäinen, H.; Haario, H. Fine-grained wood species identification using convolutional neural networks. In *Scandinavian Conference on Image Analysis*; Springer: Cham, Switzerland, 2019; pp. 67–77. [CrossRef]
24. Amin, J.; Sharif, A.; Gul, N.; Anjum, M.A.; Bukhari, S. Integrated Design of Deep Features Fusion for Localization and Classification of Skin Cancer. *Pattern Recognit. Lett.* **2020**, *131*, 63–70. [CrossRef]
25. Huan, E.Y.; Wen, G.H. Multilevel and Multiscale Feature Aggregation in Deep Networks for Facial Constitution Classification. *Comput. Math. Methods Med.* **2019** , *2019* , 1258782. [CrossRef] [PubMed]
26. D'AGOSTINO, R.B. An omnibus test of normality for moderate and large size samples. *Biometrika* **1971**, *58*, 341–348. [CrossRef]
27. Schölkopf, B.; Williamson, R.C.; Smola, A.; Shawe-Taylor, J.; Platt, J. Support vector method for novelty detection. *Adv. Neural Inf. Process. Syst.* **1999**, *12*, 583–588.
28. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 413–422.
29. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas TX, USA, 15–18 May 2000; pp. 93–104.
30. Todkar, S.; Baltazart, V.; Ihamouten, A.; Dérobert, X.; Bastard, C.L. One-class SVM based outlier detection strategy to detect thin interlayer debondings within pavement structures using Ground Penetrating Radar data. *J. Appl. Geophys.* **2021**, *192*, 104392. [CrossRef]
31. Guo, W.; Wang, Z.; Hong, S.; Li, D.; Du, W. Multi-kernel Support Vector Data Description with boundary information. *Eng. Appl. Artif. Intell.* **2021**, *102*, 104254. [CrossRef]
32. Wang, S.; Liu, Q.; Zhu, E.; Porikli, F.; Yin, J. Hyperparameter selection of one-class support vector machine by self-adaptive data shifting. *Pattern Recognit.* **2018**, *74*, 198–211. [CrossRef]

33. Wong, T.T.; Yeh, P.Y. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans. Knowl. Data Eng.* **2019**, *32*, 1586–1594. [CrossRef]
34. Berriri, M.; Sofiane Djema, G.R.; Dartiguespallez, C. Multi-Class Assessment Based on Random Forests. *Educ. Sci.* **2021**, *11*, 92. [CrossRef]
35. Grimwood, A.; Thomas, K.; Kember, S.; Aldis, G.; Lawes, R.; Brigden, B.; Francis, J.; Henegan, E.; Kerner, M.; Delacroix, L.; et al. Factors affecting accuracy and precision in ultrasound guided radiotherapy. *Phys. Imaging Radiat. Oncol.* **2021**, *18*, 68–77. [CrossRef]
36. Anwar, S.; Nouman, A.; Bahar, A.; Muhammad, T.K.; JingTao, Y. A Three-way Clustering Approach for Novelty Detection. *Inf. Sci.* **2021**, *569*, 650–668. [CrossRef]
37. Sangeetha, M.; Kumaran, M.S. Deep learning-based data imputation on time-variant data using recurrent neural network. *Soft Comput.* **2020**, *24*, 13369–13380. [CrossRef]
38. Maboudou-Tchao, E.M. Change detection using least squares one-class classification control chart. *Qual. Technol. Quant. Manag.* **2020**, *17*, 609–626. [CrossRef]
39. Velazquez-Pupo, R.; Sierra-Romero, A.; Torres-Roman, D.; Shkvarko, Y.; Santiago-Paz, J.; Gómez-Gutiérrez, D.; Robles-Valdez, D.; Hermosillo-Reynoso, F.; Romero-Delgado, M. Vehicle Detection with Occlusion Handling, Tracking, and OC-SVM Classification: A High Performance Vision-Based System. *Sensors* **2018**, *18*, 374. [CrossRef] [PubMed]
40. de Geus, A.R.; Backes, A.R.; Gontijo, A.B.; Albuquerque, G.H.Q.; Souza, J.R. Amazon wood species classification: a comparison between deep learning and pre-designed features. *Wood Sci. Technol.* **2021**, *55*, 1282. [CrossRef]
41. Souza, D.V.; Santos, J.X.; Vieira, H.C.; Naide, T.L.; Oliveira, L.E.S. An automatic recognition system of Brazilian flora species based on textural features of macroscopic images of wood. *Wood Sci. Technol.* **2020**, *54*, 1196. [CrossRef]
42. Xing, H.J.; Li, L.F. Robust least squares one-class support vector machine. *Pattern Recognit. Lett.* **2020**, *138*, 571–578. [CrossRef]
43. Chang, K.; Yoo, Y.; Baek, J.G. Anomaly Detection Using Signal Segmentation and One-Class Classification in Diffusion Process of Semiconductor Manufacturing. *Sensors* **2021**, *21*, 3880. [CrossRef] [PubMed]
44. Tan, Y.; Tian, H.; Jiang, R.; Lin, Y.; Zhang, J. A comparative investigation of data-driven approaches based on one-class classifiers for condition monitoring of marine machinery system. *Ocean Eng.* **2020**, *201*, 107174. [CrossRef]
45. Almuhtaram, H.; Zamyadi, A.; Hofmann, R. Machine learning for anomaly detection in cyanobacterial fluorescence signals. *Water Res.* **2021**, *197*, 117073. [CrossRef]