applied sciences

MDPI

Article

Trunk Borer Identification Based on Convolutional Neural Networks

Xing Zhang [1,2], Haiyan Zhang [1,2,*], Zhibo Chen [1,2] and Juhu Li [1,2]

[1] School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China
[2] Engineering Research Center for Forestry-Oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China
* Correspondence: zhyzml@bjfu.edu.cn

Abstract: The trunk borer is a great danger to forests because of its strong concealment, long lag and great destructiveness. In order to improve the early monitoring ability of trunk borers, the representative *Agrilus planipennis* Fairmaire was selected as the research object. The convolutional neural network named TrunkNet was designed to identify the activity sounds of *Agrilus planipennis* Fairmaire larvae. The activity sounds were recorded as vibration signals in audio form. The detector was used to collect the activity sounds of *Agrilus planipennis* Fairmaire larvae in the wood segments and some typical outdoor noise. The vibration signal pulse duration is short, random and high energy. TrunkNet was designed to train and identify vibration signals of *Agrilus planipennis* Fairmaire. Over the course of the experiment, the test accuracy of TrunkNet was 96.89%, while MobileNet_V2, ResNet18 and VGGish showed 84.27%, 79.37% and 70.85% accuracy, respectively. TrunkNet based on the convolutional neural network can provide technical support for the automatic monitoring and early warning of the stealthy tree trunk borers. The work of this study is limited to a single pest. The experiment will further focus on the applicability of the network to other pests in the future.

Keywords: convolutional neural networks; trunk borer; vibration signal; voice recognition

check for updates

Citation: Zhang, X.; Zhang, H.; Chen, Z.; Li, J. Trunk Borer Identification Based on Convolutional Neural Networks. *Appl. Sci.* **2023**, *13*, 863. https://doi.org/10.3390/app13020863

Academic Editors: Phivos Mylonas, Katia Lida Kermanidis and Manolis Maragoudakis

Received: 8 December 2022
Revised: 29 December 2022
Accepted: 4 January 2023
Published: 8 January 2023

1. Introduction

Forestry resources are extremely important for the comprehensive development of China and the stability of the ecological environment. However, the forest is extremely vulnerable to the destruction of the trunk borer. Timely detection and early treatment of pests are undoubtedly the biggest difficulty [1], followed by continuous updates and iterations of monitoring and identification technology for pests, aiming to address this hidden danger. Over the years, trunk borers have caused serious damage to forests, not only causing water loss and nutrient loss of trees, but also endangering the growth of the main trunk of trees [2]. Trunk borers generally live in hosts in the early days, nibbling on branches to obtain nutrients and destroying the tissue structure of trees. Because it is difficult to find and hard to control, the challenge is posed to the detection of pests. Traditional pest detection methods make it difficult to find pests which are hidden in the trunk, and thus, foresters miss the best control period. Therefore, the harm is further aggravated, resulting in irreparable losses, especially in rare trees. In recent years, with the application of acoustic technology in pest control, new directions in thinking [3,4] for the early identification of pests have been provided.

Sound detection has gradually become a new type of pest detection method [5,6]. Compared with traditional methods such as spot detection and pheromone trapping, the method of detecting the activity sound of pests in trees has the advantages of easy convenience, early warning time, high efficiency and low destructiveness [7,8]. As early as the 1920s, people began to pay attention to the acoustic characteristics of insects, classifying them as characteristic information of insects and testing them [9]. With the development of technology, the use of electronic devices for food acoustic detection of fruit pests has proved for the first time the feasibility of detecting pests using acoustic characteristics. Through long-term technical progress and experiments, the recognition model has developed from a simple hidden Markov model and Gaussian mixture model to full connection depth neural network [10,11]. The accuracy of identification has been continuously improved, meeting the practical application requirements [12]. The popularity of deep neural networks has brought inspiration for researchers. The rise of computer technology and the development of electronic technology, as well as the innovation and progress of sound detection equipment and analysis methods, have promoted the development of acoustic testing [13,14]. At home and abroad pest monitoring and identification based on voice recognition technology is also in full swing [15]. Most of the research objects are the hidden activities of pests such as storage pests, wood quarantine pests, fruit pests and trunk borers. In the early days, the microphone was directly fixed on the surface of the trunk to record the sound transmitted into the air, and the effective feature information obtained was very one-sided and not enough to judge the presence of pests. The use of piezoelectric sensors to collect vibrational sound generated by pests' activity inside the trunk improves monitoring sensitivity and reduces the interference of environmental noise.

The vibration signals of the trunk borer are recorded by the sensor and saved in an audio format [16,17], where feature points are identified and classified using deep learning techniques. In the sound recognition of pests, the random short pulses in the audio are the key point of identification [18,19] because each pulse represents the activity of pests in the tree. Through the combination of three processes of feature extraction, the audio can be learned and the prediction probability can be given a corresponding confidence score to determine whether pests exist [20–24].

This study focus on *Agrilus planipennis* Fairmaire [25]. The piezoelectric sensor was used to detect the vibrational sound emitted by the activity of the larvae in the wood segments, the sound recognition model was constructed and the neural network system was designed and trained to identify the *Agrilus planipennis* Fairmaire at an early stage. In this way, it provides technical support for the early monitoring, and offers new ideas for the identification of pests.

The rest of this study is organized as follows: Section 2 introduces the current related research work. Section 3 describes the method of data collection, data processing and data set division. In Section 4, a network model for identifying activity vibration signals of *Agrilus planipennis* Fairmaire is designed. Section 5 describes and analyzes the experimental results. Section 6 summarizes this paper.

2. Related Work

The research of pest vibration signal recognition can be divided into two stages. The first stage is to artificially analyze the pest vibration signals in the time-frequency domain, which depends on the professional level of the inspectors. The second stage is to automatically identify the pest signals through algorithms [26]. Machine learning is an important method of artificial intelligence, which aims to use machines to find relevant rules from the original data, and make responses and predictions to new samples. Neural network algorithm is the most widely used in machine learning. In the neural network, the structure and feature learning of models are very important. The neural network trains data by constructing more hidden layers, extracts more effective features of samples and determines their categories. Chestmore used time-domain signal processing and an artificial neural network algorithm to automatically recognize the sound of insects, and the experiment achieved a high recognition rate [27]. The Mankin research group of the US Department of Agriculture used Raven and least square methods to screen and classify pest sound audio. Km Sheetal Banga et al. introduced some methods for the early detection of pests, focusing on acoustic detection [28]. Tuo, Liu and others applied deep learning technology to pest signal recognition and achieved good results [16,17,29]. The pest vibration signal is transformed into audio after being processed by the sensor. Moreover, keyword detection can analyze active audio clips. Keyword detection is similar to pest vibration signal recognition, both of which take short pulses in audio as recognition objects [30].

At present, there is relatively limited research in the field of pest vibration signal recognition. With the continuous development of automation technology, pest vibration signal recognition technology has attracted more and more researchers' attention. Pest vibration signal recognition technology can not only provide a basis for pest classification and identification, but also play a huge role in pest prevention and control. At present, pest vibration signal recognition is still in the basic development period. The researchers have basically not explored the recognition efficiency of various methods. In this paper, we focus our research on the activity vibration signal of *Agrilus planipennis* Fairmaire. We use convolutional neural network to extract the characteristics of pests and identify them, so as to determine whether pests exist.

3. Data and Processing

3.1. Vibration Signal Collection of Pests

In the process of collecting the vibration signal of *Agrilus planipennis* Fairmaire larvae, it is necessary to embed the probe into the branches of the tree. The characteristics of the larvae are different from the worms which generally have vocal organs. The vibration signals generated when the pests act in the wood segments are collected by piezoelectric sensors, recorded as audio and stored. Coupling the collection probe with the tree trunk, not only reduces the difficulty of information acquisition, but also improves the purity of the signal.

In order to simulate the natural situation to a great extent, it is more realistic to approximate the actual application environment. In this study, two types of data were collected, one was the vibration sound of *Agrilus planipennis* Fairmaire larvae's activity in the wood segments of ash tree, and the other was the outdoor noise audio.

The collection of pest vibration signals was carried out in a soundproof room, and the collection equipment was SP-1L probe and self-developed sensor probe. The acquisition equipment is shown in Figure 1. SP-1L is a piezoelectric sensor probe with a sampling frequency of 44.1 kHz and a sampling accuracy of 16 bit. The front side is connected to a 6mm metal probe, and the vibration signal of the pest is recorded directly after being embedded in the trunk, as shown in Figure 2. The cylinder is the schematic diagram of the wood segment. According to the working principle of SP-1 L, the self-developed sensor was made of piezoelectric ceramics, drive amplification circuit, AC and DC amplification circuit. The recording effect of the self-developed sensor is the same as that of SP-1L, which reduces the cost of acquisition equipment and improves the autonomy of technology.

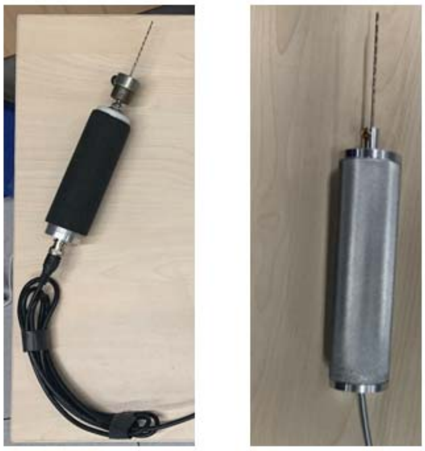

Figure 1. Acquisition equipment diagram.

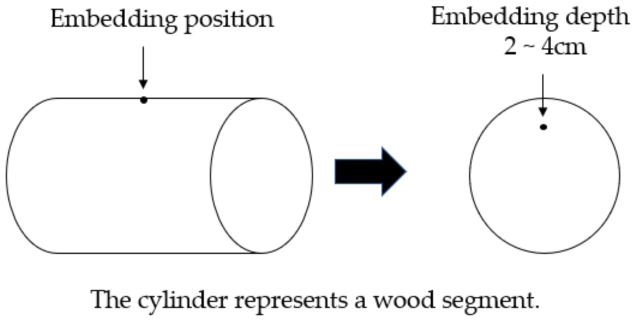

Figure 2. Schematic view of the sensor embedded in a wood segment.

Selecting 18 ash tree segments with a length of about 30 cm in the forest farm, we grouped them according to the growth state of the trees. In this case, we investigated trees in different damaged situations. We drilled holes in the middle of each wood section to facilitate insertion of the probes. During the period of frequent activity of the pest (around July), activity vibration signals were recorded using acquisition equipment, and the sound was recorded purely. The entire process took place in a closed and quiet environment. We stopped recording when the pests' activity was reduced to being undetectable. In this case, the effective vibration signal cannot be monitored. Keeping the consistency of the acquisition equipment, this research recorded the sound of the control group in an outdoor open environment, such as woods and roadsides. These include birdsong, pedestrian sounds and car sounds. All the sounds recorded were saved in .wav format.

3.2. Dataset Division

The specific information of the collected vibration signal data is shown in Table 1. The recorded audio signals were classified into two categories: *Agrilus planipennis* Fairmaire vibration signals and non-vibration signals. The duration time of *Agrilus planipennis* Fairmaire vibration signals was 32 h; the duration time of non-vibration signals was 10 h. In order to ensure the preciseness of the experiment, we divided the audio signals into groups according to trees. Among different types of signals, wood segments with high characteristic sounds, strong representation, and intense pest activity were selected as the test set, and the audio of the rest of the wood segments was used as the training set. In order to facilitate training, we split the data. The audio data were cut into audio slices. The cutting method is shown in Figure 3, and the cutting length wad 4 s. After data enhancement, *Agrilus planipennis* Fairmaire vibration signals were finally divided into 30,800 slices. After cutting and removing the unusable time period, non-vibration signals were divided into 7500 slices. The audio slices were divided into training set and test set in a ratio of 8:2.

Table 1. Signal details.

Categories	Duration (h)	Sample Rate (kHz)	Sample Depth (bit)
Agrilus planipennis Fairmaire vibration signals	32	44.1	16
Non-vibration signals	10	44.1	16

In order to fit the actual application environment as much as possible and improve the robustness of the identification network model, some of the training set data were processed in the training stage. Some noise slices were randomly mixed in the *Agrilus planipennis* Fairmaire vibration signals to simulate the forest environment, increase data complexity and improve model generalization ability. In the activity stage of the pests, the vibration sound was almost continuous. Therefore, the time for noise control in the experiment was less than the time for vibration control.

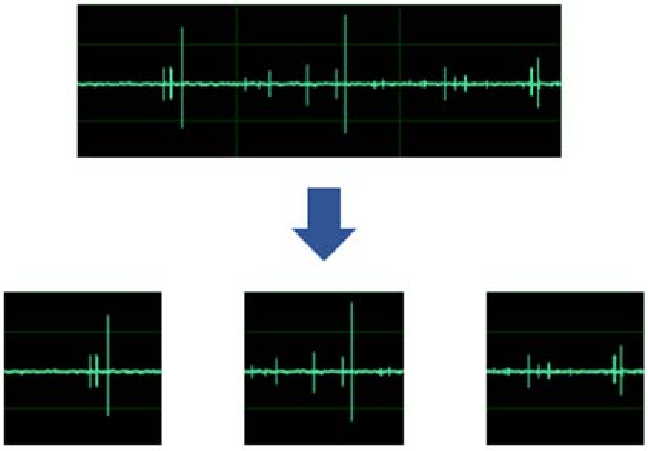

Figure 3. Schematic diagram of cutting.

3.3. Preprocessing of Vibration Signal

Before inputting audio to the neural network for learning, it needs to perform feature extraction on the audio. We analyzed the collected vibration signal, which has the characteristics of short-term high energy, interval, short and sharp. The signal is composed of several discrete pulses. We observed the time domain signal in detail and found that the waveforms are all saw-toothed, and the interval time between pulses is random. The pulse waveform is large in the front and small in the back, reaching the maximum amplitude quickly and then gradually smoothing out. The reason for this may be that the sound propagation process in the wood segment is attenuated, and the waveform changes are irregular. There are various indications that the sound of activity is a composite wave formed by the superposition of multiple sine waves. Using Adobe Audition software to check the sound in detail, it was found that the energy distribution frequency of the sound data was mostly concentrated around the 8.0 kHz frequency band. This research cut the sound clip to 4.0 kHz~12.0 kHz. On the one hand, it can be more focused on learning the changes of the signal spectrum, and on the other hand, it can effectively reduce the interference of equipment errors.

Like most recognition networks, the study used the Mel Frequency Cepstral Coefficient method to simulate the human ear’s perception of sound, and extract features from audio signal feature segments. In sound recognition, feature extraction is the key to determine the effect of audio recognition. First, the frame length was 1024 and the step size was 256 to process the input signal into frames. Then the frame signal was separated through the Hanning window. This step is to ensure that the spectrum does not lose. Then, short-time Fourier transform was performed on the signal of each frame, and the square of the absolute value of the result was obtained. Finally, it was filtered through 128 sets of Mel filters, and discrete cosine transform was performed. After obtaining the features, the method of image recognition in deep learning was referred to, and the features were processed into 16×8 image data as the input of the network. As shown in Figure 4, 0 represents the vibration signal containing the *Agrilus planipennis* Fairmaire vibration signal, and 1 represents non-vibration signal.

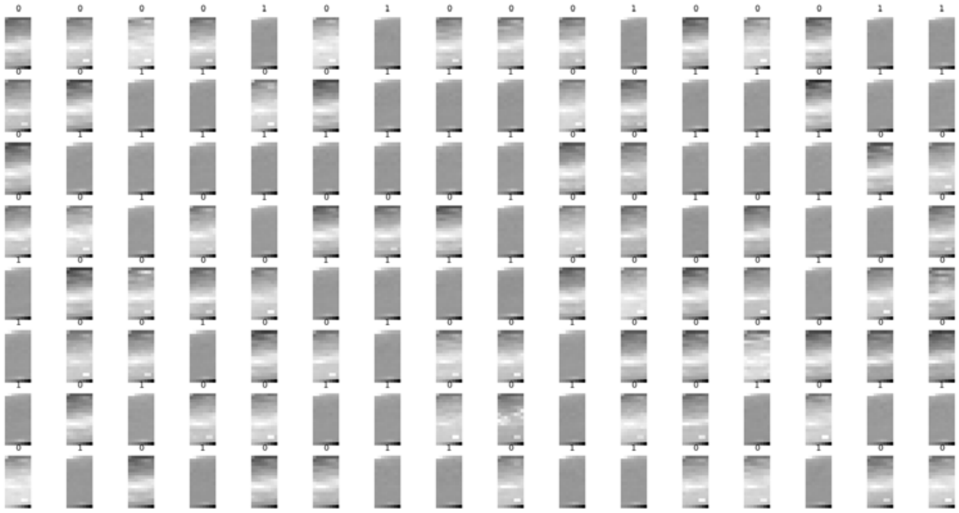

Figure 4. Signal characteristic spectrogram.

4. Vibration Signal Recognition

The focus of this study is on the identification of activity vibrational signals from pests. The accuracy of audio recognition is often determined by the Mel spectrogram of audio information, and a recognition model is built using a convolutional neural network. Local connection and weight sharing are the characteristics of convolutional neural networks, which are different from ordinary sound recognition. The pest vibration signals collected in the study are irregular and have uncertain energy. The feature positions of each piece of audio are different, and statistics of all feature information are the key to research. When the neural network obtains the local information of each position of the Mel spectrogram, various redundant information will inevitably be generated. This redundant information will not improve the recognition effect of the model, but will waste resources. Aiming at the special features of the research, this paper designs a neural network for recognition.

4.1. Local Feature Learning

Feature learning is the basis of recognition, and this paper constructs a local feature learning block based on convolution operations. The local characteristic learning block is composed of a convolutional layer, a batch normalization layer, an exponential linear unit activation function and an averaged pooling layer, and the convolutional layer learns the characteristics of the input Mel spectrogram. The batch normalization layer normalizes the output of each batch of convolutional layers to improve the performance and stability of the deep neural network; The exponential linear unit activation function defines the output of a batch normalization layer. Exponential linear units have negative values, making the average value of the activations closer to zero, thus speeding up the convergence of the network and achieving higher recognition accuracy. The formula for the exponential linear unit activation function is as follows:

$$\sigma(x) = \begin{cases} x & ,x \geq 0 \\ \rho(e^x - 1) & ,x < 0 \end{cases} \tag{1}$$

where the $\sigma(x)$ is exponential linear cell activation function, x is the input value, and the ρ hyperparameter can control when the negative part of the output of the activation function is saturated, which is assumed as 1.0 in this study.

4.2. Identification Model

Generally speaking, audio recognition refers to the process of analyzing, recognizing and classifying sound signals. The audio signal has the characteristics of time series and frequency domain. Convolutional neural network can deal with it properly [29,31]. A convolution neural network can provide translation invariant convolution in time and space. When the idea of convolutional neural network is applied to the modeling of pest vibration signal recognition, the invariance of convolution can be used to overcome the diversity of the vibration signal itself. From the practical point of view, convolutional neural network is also easy to realize for a large-scale parallel operation.

As mentioned above, this paper designs the convolutional neural network Trunk-Net. The network input is an audio signal in wav format, and the main structure is a convolutional layer with five layers in the series. After each layer of convolution, connect the activation function and the mean pooling operation. Finally, the linear classifier is connected to output the recognition result of the network.

The key role of the first layer is to ensure that the size of the input layer is consistent with the size of the extracted Mel spectrum. The main part of the network is mainly composed of convolutional layers and connection layers. The core idea of convolutional layers is based on the mathematical operation of convolution, using convolutional kernels and parts of the eigengram for convolutional calculations. The same convolutional kernel is continuously convolved with the next region by sliding steps, eventually producing a new feature map and serving as input to the next layer. The general understanding of neural networks is that shallow networks extract edge and local features, and deep networks extract semantic features. After experimental exploration and continuous correction, deep convolutional networks did not improve the effect of the network. The design of the convolution layer is finally determined, the size of the convolution kernel is 3×3 and the stride is 1.

After the end of the convolution operation, the mean pooling layer is introduced. On the one hand, it reduces the dimension of the feature, and on the other hand, it can effectively retain the information in the feature After the convolution calculation is over, the network reaches 13,376-dimensional features. The fully connected layer integrates the feature information and calculates it according to the recognition probability of the audio. The weights are connected to complete features, and neurons are activated to realize the function of recognition and prediction. In the output layer, the number of neurons is set to 2, and the recognition and prediction results of the two kinds of sound signals are obtained.

The overall network structure diagram is shown in Figure 5. The ConvBlock, which consists of a convolution part, activation functions and a pooling part, is local feature learning block constructed in this paper.It plays the role of dimensional transformation. The changes in the convolution layer input and output are shown in Table 2. The dimensional change that was previously entered can achieve a good recognition effect. We use averaging for the dimensionality reduction of feature maps. The mean pooling operation is different from the max pooling operation. The operation method of max pooling is to take the maximum value of the feature map corresponding to the convolution kernel as the representative value of this part, and the average value pooling is to take the average value as the representative value. This operation reduces the amount of parameters while maintaining valid features, and preserves more texture features in the map. To prevent overfitting, we set the dropout layer. At the same time, the activation function is introduced to achieve nonlinear changes in the network, which facilitates model convergence.

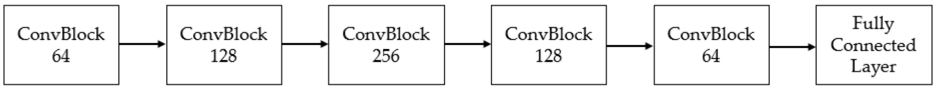

Figure 5. Network structure diagram.

Table 2. Network structure.

Layer	Feature Map Change
ConvBlock 1	(1, 64)
ConvBlock 2	(64, 128)
ConvBlock 3	(128, 256)
ConvBlock 4	(256, 128)
ConvBlock 5	(128, 64)

4.3. Identification Process

The overall recognition process is shown in Figure 6.

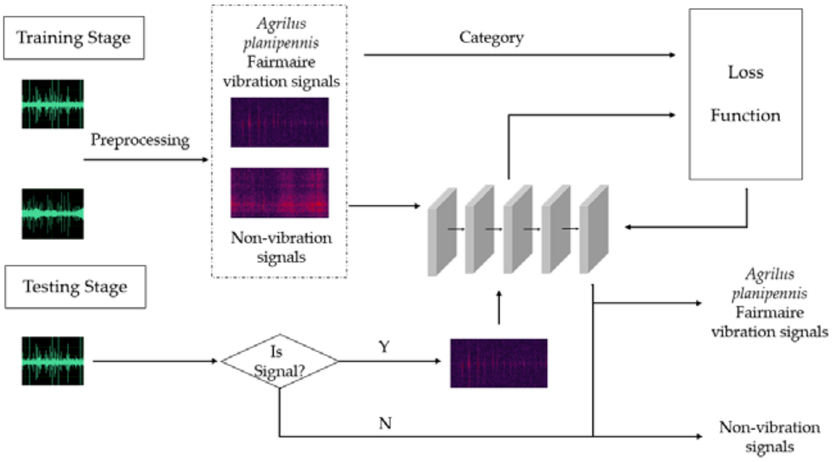

Figure 6. Flowchart of identification.

At the beginning of network training, the logarithmic Mel spectrum extracted in the preprocessing stage is used as input into the network to start training as the same as the general sound recognition network. In the research, 128 Mel spectral features of audio data are extracted and processed uniformly into 16 × 8 image data as the input data of the network. First, we predict the category through forward propagation, then use the loss function to calculate the error between the predicted category and the true category, and finally update the parameters through the back propagation of the network.

To verify the accuracy of the model, the testing stage uses completely different audio data from the training set. In the specific test process, in order to reduce the use of computing resources and improve the speed of recognition, the input audio energy is first judged. Audio that does not pass endpoint detection (i.e., audio with lower energy) will be directly determined as non-vibration signals, which can save the tedious feature extraction. The audio signal through endpoint detection is fed into a well-tuned convolutional neural network, which predicts the class based on the results of the neural network.

5. Experimental Results and Analysis

5.1. Experimental Environment

The institute uses environmental devices such as Intel i7-6700K CPU (32 GB of memory) and GeForce GTX 1080Ti (12 GB of video memory), based on the PyTorch deep learning framework implementation.

Methods for setting hyperparameters include manual tuning and automatic tuning. Based on the related experiments of audio recognition and the workload of this research, we adjusted the hyperparameters manually. After a substantial amount of experimentation, we found the best performing parameter size and used it. The initial learning rate was set to 0.0003 in the whole process, using the cross-entropy function as the loss function. Meanwhile, the batch size was 64. The model training ended after 1100 rounds.

5.2. Experimental Results

The identification of pests is carried out in units of a single audio. The research uses the audio recognition accuracy rate and F1-score as the final model evaluation index. When comparing the recognition effects of each network, the audio to be tested needs to be preprocessed before being inputted into the network model and converted into a logarithmic Mel sound feature spectrum. This process is the basic work of audio recognition and has nothing to do with the choice of recognition method.

To verify the recognition accuracy of TruckNet, we used some mature network structures for comparative experiments. In convolutional neural networks, as the number of layers increases, so does its ability to fit more complex functions. Neural networks use a backpropagation algorithm for weight updates, but increasing the number of layers will eventually cause the gradient to vanish. The short-circuit connection of ResNet is very useful for how to avoid the problem of gradient disappearance and explosion. The depthwise separable convolution of MobileNet_V2 expands the feature map channel and enriches the number of features by introducing residual blocks. This approach significantly improves the recognition accuracy. In this study, in order to test the proposed network effect, we introduced ResNet, MobileNet and other network comparisons through transfer learning. We used the same dataset and data processing. We modified the last layer of the network model and set its last layer as a new classification layer. In this way, the network can ensure that the output dimension is the same as the number of types in the dataset, and then the network is evaluated. So, we used VGGish, ResNet and MobileNet as the compared network to recognize *Agrilus planipennis* Fairmaire vibration signals.

The experimental results are shown in Table 3. The recognition accuracy rate of TrunkNet was 96.89%, while that of MobileNet_V2, ResNet18 and VGGish were 84.27%, 79.37% and 70.85%, respectively. The F1-score of TrunkNet was 0.92, while that of MobileNet_V2, ResNet18 and VGGish were 0.86, 0.69 and 0.58, respectively. It can be seen from the experimental results that the effect of the mature recognition network directly applied to this experiment is not ideal. In the recognition and classification experiments, the performance of the network is often related to the original dataset. The recognition model needs to be flexibly adjusted according to the characteristics of the dataset of the collection target to achieve the desired effect.

Table 3. Recognition result.

Recognition Nets	Accuracy Rate (%)	F1-Score
TrunkNet	96.89	0.92
MobileNet_V2	84.27	0.86
ResNet18	79.37	0.69
VGGish	70.85	0.58

Nowadays, in the research of pest identification, the acquisition of original data is always a big problem because of the particularity of the living environment of the borer pests. The lack of a unified collection standard results in very few public datasets, which seriously affects the progress and development of the identification technology of the trunk borer. In this study, the self-developed piezoelectric ceramic sensor, on the one hand, solves the problem of being unable to capture the activities of pests inside the trunk, and on the other hand, can record the characteristic information of pests in the form of audio, which is reflected in the audio frequency spectrum. Using sensors can monitor the activity of stem-boring pests in trees and save the signal without damaging the tree. The simple collection method provides reference opinions for other researchers and lays the foundation for subsequent research.

In the field of traditional voice recognition, some networks have shown excellent performance. From data-driven clustering to residual structure optimization, they solved the degradation problem in neural networks. However, there are relatively few audio signal features and the discontinuity of feature information in this study. The existing models are not very good at completing the monitoring task. According to the experimental results of the above research, it can be analyzed that, different from human voice or other sounds with obvious characteristics, the vibration signal of pests has strong characteristics, and the intermittent and irregular audio frequency is difficult to identify. The identification network method in this study provides new research ideas and technical support for forestry borer monitoring.

6. Conclusions

In this study, piezoelectric ceramic sensors were used to collect the eating sound of Agrilus planipennis Fairmaire, and at the same time, the noise audio outdoor was collected. The network model proposed by the research has been adjusted according to the characteristics of pest sounds. In comparison with other networks' identification results, TrunkNet can efficiently and accurately identify pests. Therefore, the method proposed in the study can adapt to the monitoring task of forest tree borer dry pests, and provide technical support for the automatic monitoring and early warning identification of forest borer pests.

In the future, we will collect vibration signals from different forest areas to further verify the feasibility of convolutional neural networks. We will pay more attention to the population density information of pests to determine the specific damage of trees.

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

Funding: This research was funded by National Natural Science Foundation of China (Grant number 32071775).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset and code of the article are in the following link: https://pan.baidu.com/s/1ITikKnKhJ0xn9_HEQKDTfA?pwd=wmws (accessed on 29 December 2022).

Acknowledgments: We would like to thank the authors who jointly completed the article and the peer reviewers for their valuable contributions to this research.

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

References

1. Li, J.; Chang, G.; Qu, T.; Cui, Y.; Yan, J.; Song, Y. Hazard assessment of forest pests in China. *China For. Dis. Pests* **2019**, *38*, 11–17. (In Chinese)
2. Bu, Y.; Qi, X.; Wen, J.; Xu, Z. Analysis of sound characteristics of 7 species of tree stem borers. *J. Nanjing For. Univ.* **2016**, *40*, 179–184. (In Chinese)
3. Mankin, R.W.; Hagstrum, D.W.; Smith, M.T.; Roda, A.L.; Kairo, M.T. Perspective and promise: A century of insect acoustic detection and monitoring. *Am. Entomol.* **2011**, *57*, 30–44. [CrossRef]
4. Yazga, B.G.; Mürvet, K.; Müjgan, K. Detection of sunn pests using sound signal processing methods. In Proceedings of the 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Tianjin, China, 18–20 July 2016; IEEE: Piscataway, NJ, USA, 2016.
5. Jalinas, J.; Güerri-Agulló, B.; Dosunmu, O.G.; Haseeb, M.; Lopez-Llorca, L.V.; Mankin, R.W. Acoustic Signal Applications in Detection and Management of Rhynchophorus spp. in Fruit-Crops and Ornamental Palms. *Fla. Entomol.* **2019**, *102*, 475–479. [CrossRef]
6. Mankin, R.W.; Al-Ayedh, H.Y.; Aldryhim, Y.; Rohde, B. Acoustic detection of *Rhynchophorus ferrugineus* (Coleoptera: Dryophthoridae) and *Oryctes elegans* (Coleoptera: Scarabaeidae) in *Phoenix dactylifera* (Arecales: Arecacae) trees and offshoots in Saudi Arabian orchards. *J. Econ. Entomol.* **2016**, *109*, 622–628. [CrossRef] [PubMed]
7. Hetzroni, A.; Soroker, V.; Cohen, Y. Toward practical acoustic red palm weevil detection. *Comput. Electron. Agric.* **2016**, *124*, 100–106. [CrossRef]

8. Qi, X.; Bu, Y.; Xu, Z.; Wen, J. Acoustic detection of the number of larvae of A. chinensis in the Yangshu segment. *Entomol. Environ.* **2016**, *38*, 529–534. (In Chinese)
9. Mankin, R.W.; Burman, H.; Menocal, O.; Carrillo, D. Acoustic detection of *Mallodon dasystomus* (Coleoptera: Cerambycidae) in *Persea americana* (Laurales: Lauraceae) branch stumps. *Fla. Entomol.* **2018**, *101*, 321–323. [CrossRef]
10. Asadolahzade Kermanshahi, M.; Homayounpour, M.M. Improving Phoneme Sequence Recognition using Phoneme Duration Information in DNN-HSMM. *J. AI Data Min.* **2019**, *7*, 137–147.
11. Wu, X.; Zhu, M.; Wu, R.; Zhu, X. A Self-Adapting GMM Based Voice Activity Detection. In Proceedings of the IEEE 23rd International Conference on Digital Signal Processing (DSP), Shanghai, China, 19–21 November 2018; pp. 1–5.
12. Liu, H.L.; Cheng, T. Research on Multi-Physical Domain Information Fusion Method of Intelligent Processing Machine Based on GMM-HMM. In *Applied Mechanics and Materials*; Trans Tech Publications Ltd.: Zurich, Switzerland, 2017; Volume 864, pp. 184–191.
13. Zhang, Y.; Chen, D.; Ye, C. *Toward Deep Neural Networks: WASD Neuronet Models, Algorithms, and Applications*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2019.
14. Yu, H.; Zhao, J.; Yang, S.; Wu, Z.; Nie, Y.; Zhang, W.-Q. Language Recognition Based on Unsupervised Pretrained Models. In Proceedings of the Interspeech 2021, Brno, Czech, 30 August–3 September 2021; pp. 3271–3275.
15. Kahl, S.; Wilhelm-Stein, T.; Klinck, H.; Kowerko, D.; Eibl, M. A Baseline for Large-Scale Bird Species Identification in Field Recordings. In Proceedings of the CLEF (Working Notes), Avignon, France, 10–14 September 2018; p. 2125.
16. Sun, Y.; Tuo, X.; Jiang, Q.; Zhang, H.; Chen, Z.; Zong, S.; Luo, Y. Vibration identification method of two kinds of pests based on lightweight neural network. *For. Sci.* **2020**, *56*, 100–108. (In Chinese)
17. Liu, X.; Sun, Y.; Cui, J.; Jiang, Q.; Chen, Z.; Luo, Y. Early artificial intelligence recognition of feeding sounds of borer pests. *For. Sci.* **2021**, *57*, 93–101. (In Chinese)
18. Zhang, H.; Yuan, M.; Jiang, Q.; Sun, Y.; Cui, J.; Ren, L.; Luo, Y. Deep Learning Model Compression for Real-Time Monitoring of Drilling Vibration. *J. Beijing For. Univ.* **2021**, *43*, 92–100. (In Chinese)
19. Tu, W.; Yang, Y.; Du, B.; Yang, W.; Zhang, X.; Zheng, J. RNN-based signal classification for hybrid audio data compression. *Computing* **2020**, *102*, 813–827. [CrossRef]
20. Zhang, Q.; Lu, H.; Sak, H.; Tripathi, A.; McDermott, E.; Koo, S.; Kumar, S. Transformer Transducer: A Streamable Speech Recognition Model with Transformer Encoders and RNN-t Loss. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 7829–7833.
21. Chuanyu, G.G.; Yimin, S. Discussion on optimization algorithm of speech recognition decoding based on BLSTM Heilongjiang. *Sci. Technol. Inf.* **2020**, *18*, 86–87. (In Chinese)
22. Xie, X.; Zhang, L.; Wang, J. Application of Residual Network in Infant Cry Recognition. *J. Electron. Inf.* **2019**, *41*, 233–239. (In Chinese)
23. Bhanja, C.; Bisharad, D.; Laskar, R.H. Deep residual networks for pre-classification based Indian language identification. *J. Intell. Fuzzy Syst.* **2019**, *36*, 2207–2218. [CrossRef]
24. Zeng, M.; Xiao, N. Effective combination of DenseNet and BiLSTM for keyword spotting. *IEEE Access* **2019**, *7*, 10767–10775. [CrossRef]
25. Tejashri, M.; Srijita, B. Birds Voice Classification using ResNet. *Int. J. Eng. Dev. Res.* **2018**, *6*, 2321–9939.
26. Du, D. Research on Acoustic Information Characteristics and Automatic Identification of Blueberry Typical Pests. Master's Thesis, Guizhou University, Guiyang, China, 2019. (In Chinese).
27. Chesmore, D. Automated bioacoustic identification of species. *An. Acad. Bras. Cienc.* **2004**, *76*, 436–440. [CrossRef] [PubMed]
28. Banga, K.S.; Kotwaliwale, N.; Mohapatra, D.; Giri, S.K. Techniques for insect detection in stored food grains: An overview. *Food Control* **2018**, *94*, 167–176. [CrossRef]
29. Tuo, X. Sound Recognition of Borers Based on Deep Learning. Master's Thesis, Beijing Forestry University, Beijing, China, 2020.
30. Chen, G.; Parada, C.; Heigold, G. Small-Footprint Keyword Spotting Using Deep Neural Networks. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014.
31. Erdogan, H.; Hershey, J.R.; Watanabe, S.; Roux, J.L. Phase-Sensitive and Recognition-Boosted Speech Separation Using Deep Recurrent Neural Networks. In Proceedings of the ICASSP 2015—2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, Australia, 19–24 April 2015.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.