Article

A Ground Moving Target Detection Method for Seismic and Sound Sensor Based on Evolutionary Neural Networks

Kunsheng Xing ⬤, Nan Wang *⬤ and Wei Wang ⬤

College of Intelligent Science and Technology, National University of Defense Technology,
Changsha 410003, China
* Correspondence: acnwang@aliyun.com

Abstract: The accurate identification of moving target types in alert areas is a fundamental task for unattended ground sensors. Considering that the seismic and sound signals generated by ground moving targets in urban areas are easily affected by environmental noise and the power consumption of unattended ground sensors needs to be reduced to achieve low-power consumption, this paper proposes a ground moving target detection method based on evolutionary neural networks. The technique achieves the selection of feature extraction methods and the design of evolving neural network structures. The experimental results show that the improved model can achieve high recognition accuracy with a smaller feature vector and lower network complexity.

Keywords: moving target recognition; evolutionary neural networks; ambient intelligence; wireless sensor networks; unattended ground sensors

1. Introduction

Unattended ground sensor networks are wireless networks consisting of a large number of stationary, unattended ground sensors scattered randomly in a work area in a self-organizing and multi-hop manner, so as to detect abnormal events. Nodes in the networks are homogeneous, low cost, and small in size, and most can work for long periods of time. Unattended ground sensor networks are dedicated to detecting moving targets within alert areas (Figure 1). During the Vietnam War, the US Army first applied networks of unattended ground sensors to detect signs of Viet Cong activity along the Ho Chi Minh Trail. Unattended ground sensor systems were later used for border control [1], key facility protection [2,3], and intrusion prevention in nature reserves. The main detection means of unattended ground sensors are visible light, passive infrared, radar, magnetic field, seismic, and sound. Sound and seismic sensors are of particular research interest in the field of unattended ground sensor systems because of their long detection distance, low cost, light weight, and low power consumption. Targets identified by unattended ground sensors based on sound and seismic signals include persons and vehicles.

The work of ground moving target recognition based on sound and seismic signals usually includes: (1) the feature extraction of sound and seismic signals; and (2) the classification by an algorithm. Moving target feature extraction methods based on sound and seismic signals can be divided into the time, frequency, and time–frequency domains. Typical time-domain methods include root-mean-square (RMS), event isolation [2,3], STA/LT [4–6], footstep frequency [7,8], zero-crossing rate [9], skewness [10], and kurtosis. Frequency-domain methods include spectral center of gravity [11], IWF [12], and frequency entropy. Time- and frequency-domain features reflect the corresponding information of sound and seismic signals, respectively, and are usually one-dimensional. Time–frequency domain features are usually multidimensional, and associated feature extraction methods include MFCC [13] and LPCC. Moving target classification methods based on sound and seismic signals include threshold methods, traditional machine learning (TML), and deep learning

(DL). The threshold method is to set a fixed or variable feature threshold based on experience. When the feature value exceeds the set threshold, it is considered that there is a target intrusion. The threshold method has high computational efficiency, but its classification accuracy is easily affected by environmental noise and cannot be applied to the urban environment, which has a large amount of noise caused by human activity. Compared with the threshold method, the machine learning (ML) method has higher anti-interference ability and classification accuracy. TML classification methods include support vector machine (SVM) [13–15], K-nearest neighbor (KNN) [16,17], decision tree (DT) [18], naïve Bayes (NB) [7], artificial neural network (ANN) [10], and the Gaussian mixture model (GMM) [19]. DL methods for ground target recognition based on sound and seismic signals include CNN [20–23] and RNN [24,25]. Though the ML method improves accuracy compared with the threshold method, it also increases the complexity of the network and the hardware cost of the deployment model, making it difficult for unattended ground sensors to meet the requirements of long-term work. In order to overcome this problem, we propose an evolutionary neural network approach to ground target recognition, combining genetic algorithms and neural networks to select the best feature extraction methods, to optimize the neural network structure, and to achieve high recognition accuracy with minimum feature vector and network complexity. The machine learning methods for moving target recognition based on sound and seismic signals are summarized in Table 1.
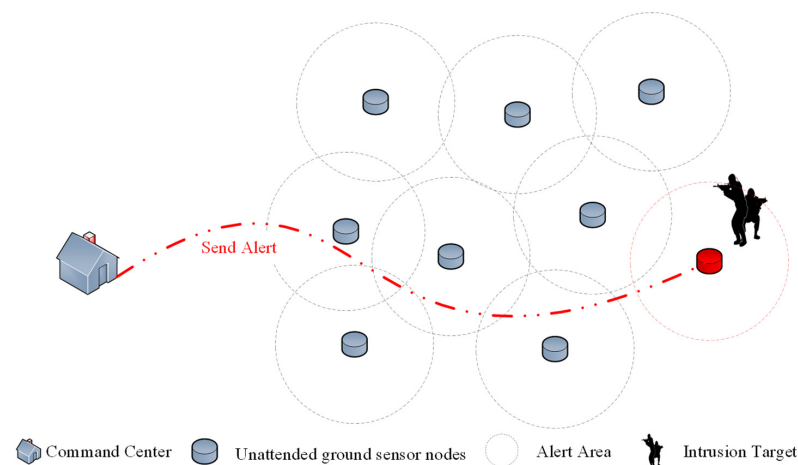

Figure 1. UGS principle of operation.

Table 1. List of references on machine learning methods for moving target recognition based on sound and seismic signals.

Ref.	Feature Extraction Method	Classification Method
2017, Kucukbay et al. [13]	MFCC	SVM
2021, Bin et al. [14]	Fractal dimension	SVM
2019, Zhong et al. [15]	Wavelet energy ratio, zero-crossing rate, mean value, peak index, SD, and waveform index	SVM
2011 Jin et al. [16]	Probabilistic finite state automata	KNN
2013 Huang et al. [17]	Wavelet packet node energy	KNN
2010, Narayanaswami et al. [18]	Wavelet statistics, spectral statistics, cadence, and kurtosis	DT
2011, Damarla et al. [7]	Energy of voice spectra, cadence, and frequency rhythm	NB
2017, Choudhary et al. [10]	Automatic extraction	ANN
2010, Park et al. [19]	Gait frequency and temporal pattern	GMM
2021, Bin et al. [20,22]	Compression sensing	CNN
2019, Wang et al. [21]	MFCC	CNN
2019, Jin et al. [23]	Automatic extraction	CNN
2019, Xu et al. [24]	Automatic extraction	RNN
2021, Li et al. [25]	Automatic extraction	RNN

The rest of this paper is structured as follows. Section 2 introduces the necessity of signal preprocessing and the methods of signal preprocessing. Section 3 focuses on the principle of the evolutionary neural network, optimal feature extraction methods, and neural network structure. Section 4 presents a moving target sound and seismic signal dataset, and compares the experimental results of the proposed method with those of TML and DL methods. Section 5 provides our conclusions.

2. Signal Preprocessing

Due to differences in sensor sensitivity and other hardware circuits, there are differences in the mean and amplitude of the signals collected by different signal acquisition devices at the same event. To ensure that the initial energy of each seismic and sound signal segment is at the same level, preprocessing of the original signal is required. Signal preprocessing has two parts: the normalization of signal units and the removal of DC components.

DC component removal is the removal of the DC bias of the signals generated by the acquisition instrument. A fast Fourier transform is applied to the signal to remove the zero-frequency signal. Then, a fast inverse Fourier transform is used to obtain the acoustic and seismic signal after removing the DC component.

The signal unit of the seismic geophone after acquisition by the AD acquisition card is V. The goal is to convert the signal unit to mm/s. The velocity signal conversion formula of ground vibration can be obtained according to its sensitivity as

$$SN = SR/SU_{seismic},\qquad(1)$$

where $SU_{seismic}$ is the sensitivity of the seismic sensor, SR is the raw signal collected by the seismic sensor, and SN is the seismic signal after normalization. After the sound sensor is acquired by the AD acquisition card, the signal unit is V. The goal is to convert the sound signal to a voltage signal collected by a standard microphone. The microphone sensitivity is

$$SU_{sound} = 20\log_{10}\frac{SR}{SS},\qquad(2)$$

where SU_{sound} is the sensitivity of the sound sensor, SR is the raw signal collected by the sound sensor, and SS is the sound signal collected by a standard microphone. According to the amplification of the amplifier circuit, the standard microphone output signal is

$$SS = (SR/10^{Mag/20})/10^{SU_{sound}/20},\qquad(3)$$

where SU_{sound} is the sensitivity of the sound sensor, Mag is the magnification of the sound sensor, SR is the raw signal collected by the sound sensor, and SS is the sound signal after normalization.

3. Proposed Method

The target recognition method deployed in unattended sensors should have both high recognition accuracy and low algorithmic complexity, which brings great challenges to the structural design of neural networks and the optimal selection of feature extraction methods. In order to overcome this problem, we propose an evolutionary neural network-based target recognition method to overcome this challenge. The evolutionary neural network is a new network model that integrates evolutionary computation and neural networks, using the principle of biological evolution to search for neural networks that perform well for a given task in a feasible domain space. The method uses a genetic algorithm as the optimization method, the selection of feature and fully connected neural network structure design as optimization objects, and the highest network classification accuracy and lowest network complexity as optimization goals. This method is named evolutionary neural networks for feature selection and network design (FSND-ENN).

3.1. FSND-ENN Principle

The evolutionary neural network includes initialization, evaluation, evolution, and termination condition discrimination, as shown in Figure 2a. Initially, the FSND-ENN method will generate a random population. As seen in Figure 2b, a population is an aggregate of individuals, and a population contains a certain number of individuals. Each individual carries genetic information, which is a binary code that contains information about feature selection and neural network structure. Secondly, the fitness of each individual in the population will be calculated. This process is called evaluation. Third, each individual in the parent population passes on genetic information to individuals in the offspring population through crossover and mutation. This process is called evolution. As seen in Figure 2c, a parent is an individual of the previous generation that passes genetic information to the next generation, and a child is an individual of the next generation. Finally, it is determined whether the requirement of evolutionary generations is satisfied.
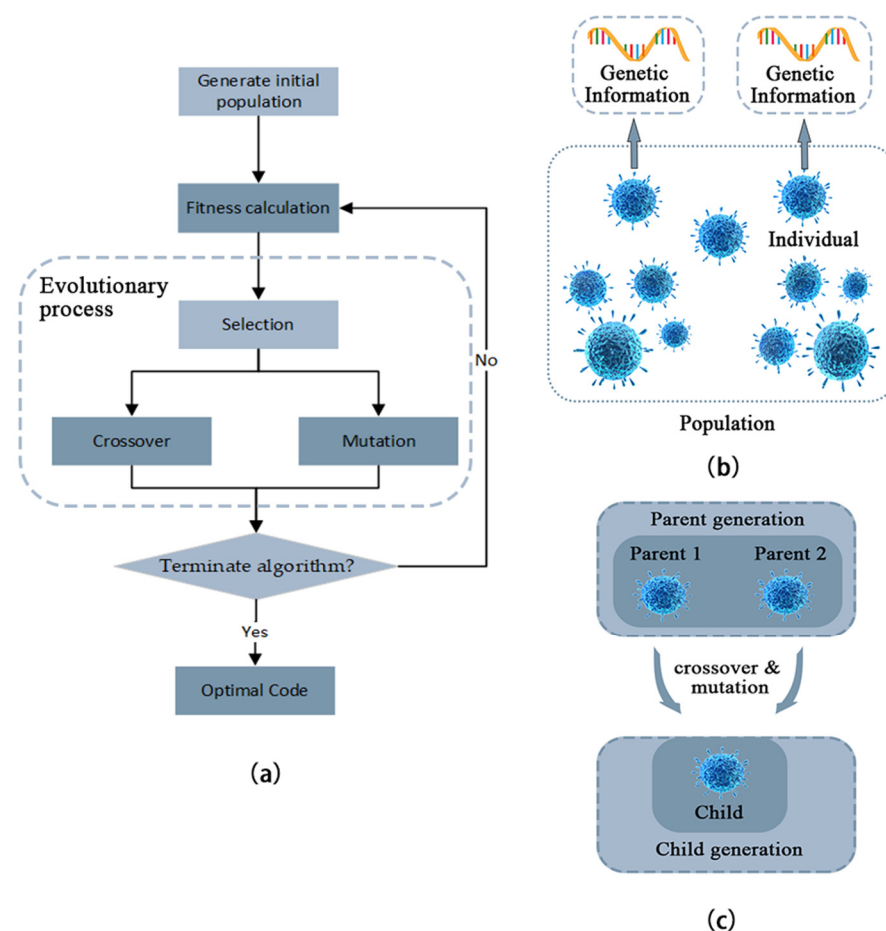

Figure 2. FSND-ENN principle diagram. (**a**) Flowchart of genetic algorithm. (**b**) Relationship between populations, individuals, and genotypes. (**c**) Relationship between parent generation, parent, child generation, and child.

3.1.1. Initialization Process

To encode the optimization problem, the optimizable object is transformed to a binary genetic code, and a unique binary code exists for any solution to the problem. The genetic information encoding of a subject includes feature selection, the number of fully connected network layers, and the number of nodes in each layer (Figure 3). The genetic information encoding of an evolutionary neural network consists of a string of binary codes of length $d + M_1 + M_2 + M_3$. M_1 and M_2 are parameters to be set, which are related to the maximum number of layers, $2^{M_1+1} - 1$, and the maximum number of nodes, $2^{M_2+1} - 1$ per layer

of the fully connected neural network. These parameters depend on the complexity of the training data of the fully connected neural network; the higher the complexity, the higher the maximum number of layers and maximum number of nodes per layer of the evolutionary neural network. M_3 is the number of layers of the fully connected network.
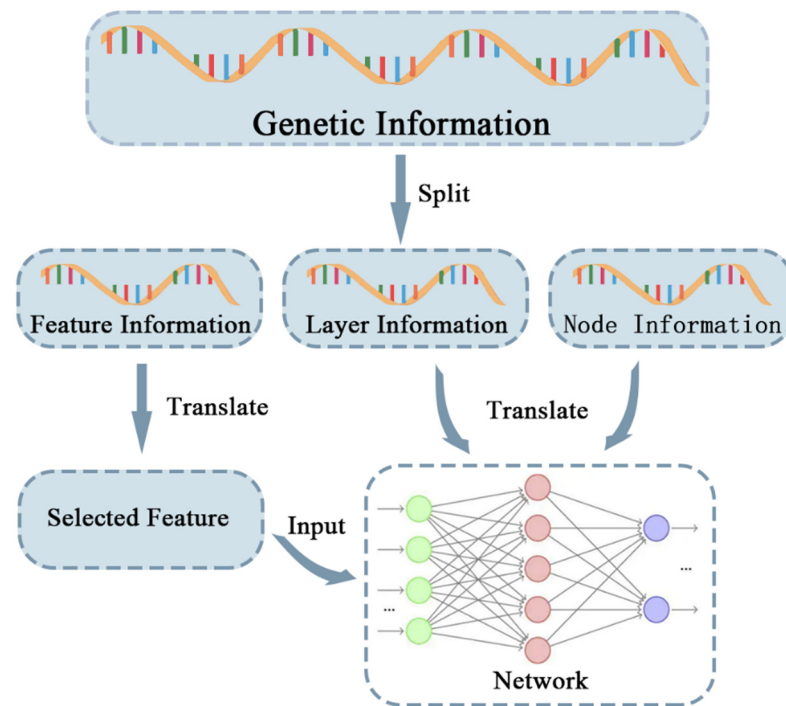

Figure 3. Genetic information.

The first through d-th codes indicate the feature selection result, where d is the number of features to be selected. Each value in the feature selection code corresponds to a feature, where a value of 1 indicates that a feature is selected, and 0 indicates rejection. The $(d + 1)$-th through $(M_1 + d)$-th encoding denote the number of network layers $\sum_{i=1}^{M_1} 2^i \times x_i$ of the fully connected neural network, where x_i is the i-th value in this binary encoding.

Codes $d + M_1 + (n - 1) \times M_2$ to $d + M_1 + n \times M_2$ represent the number of nodes $\sum_{i=1}^{M_1} 2^i \times x_i$ in the n-th fully connected layer, where x_i is the i-th value in this binary code. For example, if the parameters are set as $d = 5$, $M_1 = 4$, $M_2 = 4$ and the total length of the encoding is 69, the genetic information is converted as shown in Figure 4. The first to fifth encodings are for feature selection, the sixth to the ninth are the network layer encodings, and the $(10 + 4(n - 1))$-th to $(10 + 4n)$-th are the node encodings of the n-th layer network. Feature selection code 01001 indicates that the second and fifth features are selected, and the rest are discarded. The network layer code 0011 can be translated to a three-layer network. A first layer network node encoding of 1011 indicates that the first layer network has 11 nodes.

3.1.2. Evaluation Process

The fitness function is applied to each individual in the population, and fitness is the criterion for the optimal solution. Evolutionary neural networks aim for the highest network classification accuracy and lowest network complexity. Therefore, the fitness function includes classification accuracy and complexity scores, with different weight coefficients. The fitness is calculated as

$$f(x_i) = 0.7 \times acc(x_i) + 0.3 \times (1 - \frac{N(x_i)}{AN}), \tag{4}$$

where $acc(x_i)$ is the classification performance of individual x_i, expressed as the classification performance of the neural network classifier. AN is the maximum number of nodes of the optimized fully connected layer network, and $N(x_i)$ is the total number of nodes of the neural network for individual x_i.
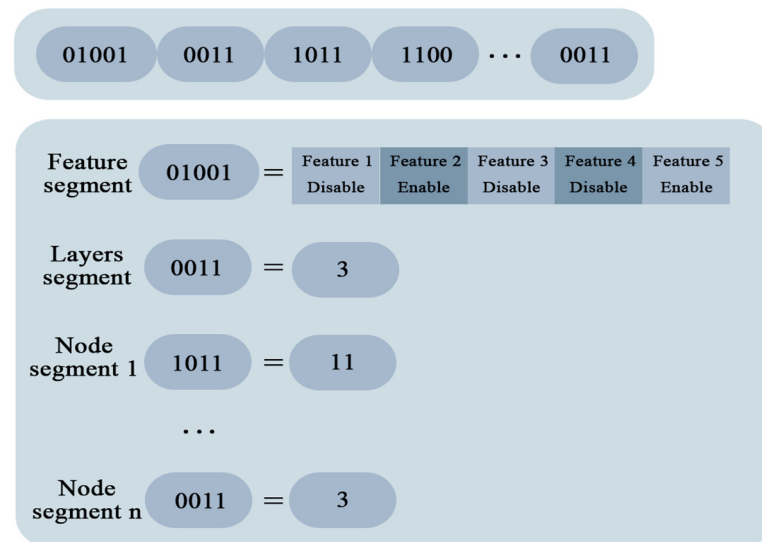

Figure 4. Gene translation.

3.1.3. Evolutionary Process

The evolutionary process includes selection, crossover, and variation. Selection eliminates individuals that do not perform well. During the selection process, individuals in the population are randomly and repeatedly selected N times, and the genes of the selected individuals are used as genotypes that can be passed on to the next generation (the same genotype may be selected multiple times), where N is the number of individuals in the population. The probability that an individual is selected is determined by the proportion of its fitness value in the total fitness value of the population. The population size is M, and the probability that individual i with fitness f_i is selected is

$$P_i = \frac{f_i}{\sum_{i=1}^{M} f_i}. \tag{5}$$

Obviously, the higher the fitness of an individual, the greater its probability of passing a gene to the next generation. The selected genotypes are used as individuals of the parent generation for the next step, crossover, and mutation.

In the crossover process, parent individuals are traversed, and one is randomly selected to form N crossover combinations to generate offspring, where N is the parent population size. The crossover combination randomly generates a binary mask whose length is that of the individual. If the mask is 1 (0), the offspring will inherit the genetic information of parent 1 (2) in the crossover combination. During the mutation process, the genetic information of the offspring is coded with a small probability to select the position of an individual for variable transformation, 0 becomes 1, and 1 becomes 0. The offspring of these three genetic processes constitute the next-generation population, and genetic evolution cycles until the optimization criterion is satisfied. In Figure 5, a segment of genetic code of length 8 is used as an example. If the mask is 000101010, then bits 4, 6, and 7 in parent 1 and bits 1, 2, 3, 5, and 8 in parent 2 are selected to form the next generation of genes. If position 5 is selected as the variant, the final gene code of the offspring is 111011110. The black part of the offspring code in Figure 5 is inherited from parent 1, the white part is inherited from parent 2, and the red part is the variant.
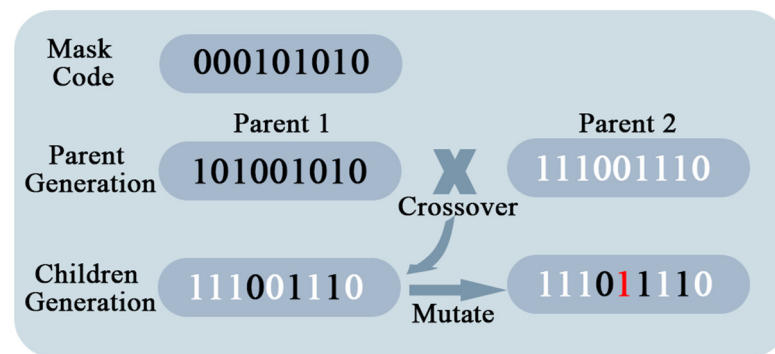
Figure 5. Crossover and variation.

3.2. Step of Feature Selection and Network Design

The optimization results of the evolutionary neural network include the feature selection and the structure of the fully connected neural network. Using the above datasets as a measure for the calculation of fitness in the evaluation process of evolutionary neural networks, the feature selection and network design method is implemented as follows.

(1) Calculate the 120 features of the training and testing data and form them into feature vectors;

(2) Design the parameters of the evolutionary neural network, where the parameters include population size, crossover probability, variation probability, and the number of evolutionary generations, and randomly generate a primitive population;

(3) For the genetic information in the population, generate the feature vector corresponding to each individual with the neural network, and calculate the fitness of each individual. In the process of fitness calculation, the neural network is trained with 1500 epochs using the training set mentioned above, and the classification score of the trained network for the test set is the score of the classification performance in the fitness calculation of this network;

(4) Generate offspring populations through election, crossover, and mutation of parent populations;

(5) Determine whether the requirement of evolutionary generations is satisfied. If so, the individual with the highest adaptation degree is output; if not, repeat step 3.

After the process of evolution, an individual with optimal adaptation is obtained, and is used in a fully connected neural network with one seismic signal feature, one sound signal feature, and two hidden layers. The selected features are

$$kurtosis = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} - 3, \tag{6}$$

$$Vibration\ energy = 10\log_{10}(\frac{1}{n}\sum_{i=1}^{n}x_i^2), \tag{7}$$

where x_i is the $i-$th data point in x and n is the length of the signal x. The optimized network has only two hidden layers, where the number of nodes in the first layer is 12 and the number of nodes in the second layer is 8.

The network uses cross-entropy loss as the objective function. Hence, the classifier has loss function

$$L(\Theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C}t_{ij} \times \log(p(y_j|x_i, \Theta)), \tag{8}$$

where N is the number of samples, C is the number of target types, t_{ij} is the probability that the i-th sample belongs to category j, and Θ denotes the set of all parameters.

4. Experimental Results

4.1. Experimental Hardware

The signal acquisition unit consisted of a sensor module, AD acquisition module, and a laptop computer (Figure 6). The sensor module included seismic and acoustic sensors. The seismic sensor had a natural frequency of 4.5 Hz and a sensitivity of 100 mv/m/s. The sound sensor consisted of an analog MEMS microphone with a sensitivity of −42 dBV, and a 110 dB amplifier circuit. The AD acquisition module had an AD acquisition card with 24-bit acquisition accuracy, and the sampling frequency was set to 10,000 Hz. The laptop computer was a Lenovo Saver R7000K with an i710875H CPU and a GeForce RTX2070 Super GPU.
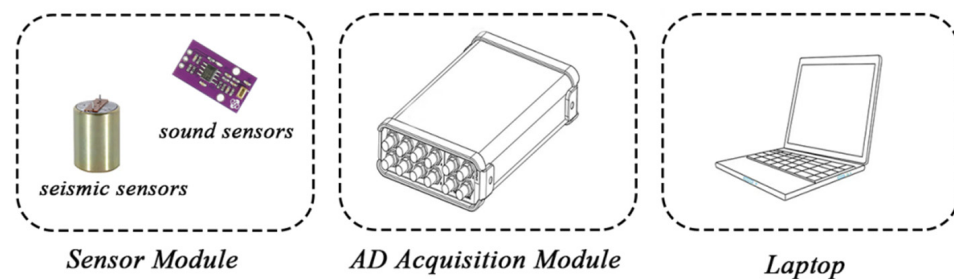

Figure 6. Experimental hardware.

4.2. TRESS01 Datasets

The dataset used in this paper was collected in the first target recognition experiment of sound and seismic signals (TRESS01), which was organized by our team. The experimental site was on the campus of Hunan Normal University in Changsha, Hunan Province. The experiments were conducted over a period of two weeks. A signal acquisition device was placed on the east side of a concrete road on the campus. As people and vehicles moved along the road, people moved within 15 m from the UGS, and vehicles within 30 m.

The detection distance of an unattended sensor is closely related to the noise intensity of its deployment area. Its alert distance is shorter in a noisy urban area than in a quiet rural area [26]. Vibration intensity (*VI*) and sound pressure level (*SPL*) are used to describe the respective noise levels of seismic and sound signals:

$$VI = 20\log_{10}\frac{S_{RMS}}{S_{pre}}, \tag{9}$$

$$SPL = 20\log_{10}\frac{P_{RMS}}{P_{ref}}, \tag{10}$$

where S_{RMS} is the *RMS* of the seismic signal, S_{ref} and P_{ref} are the relative reference values of the *VI* and *SPL*, both taken as 10×10^{-5}, and P_{RMS} is the *RMS* of the sound signal.

Table 2 shows the *VI* and *SPI* indicators of ambient noise at each data acquisition, the temperature and humidity of the environment at the time of acquisition, and the data of the acquisition experiment. The datasets include a training set for model fitting, and a test set to evaluate the generalization ability of the final model. To ensure a more effective assessment, the training and test sets are composed of sound and seismic signals collected at different times. The target acoustic and seismic signals collected from the first to the fifth data acquisitions constitute the training set, and the target acoustic and seismic signals from the fifth data acquisition constitute the test set. Table 3 shows the size of the processed dataset.

Table 2. Recording environments.

	VI/dB	*SPL*/dB	Temperature	Date
First recording	36.88	25.71	26 °C	30 May 2022
Second recording	38.82	26.32	25 °C	31 May 2022
Third recording	39.03	24.49	23 °C	1 June 2022
Fourth recording	41.99	28.44	23 °C	11 June 2022
Fifth recording	40.83	26.69	27 °C	15 June 2022

Table 3. Size of datasets.

	Background Noise/Second	Person/Second	Vehicle/Second
Train datasets	2372	2483	1954
Test datasets	403	371	452

4.3. Evaluation Metrics

The accuracy [27], false alarm rate, and underreporting rate are usually used to evaluate the model in an UGS system. Accuracy is the proportion of correctly predicted samples,

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

where *TP* (true positive) is the number of correctly classified intrusion events, *TN* (true negative) is the number of correctly classified noise samples, *FN* (false negative) is the number of missed intrusion events, and *FP* (false positive) is the number of intrusion events misjudged as noise. The false alarm rate is the proportion of noise samples misjudged as intrusion events,

$$FAR = \frac{FP}{TP + TN},\tag{12}$$

where *TP*, *TN*, and *FP* are defined by Equation (11). The underreporting rate is the proportion of samples that misjudge an intrusion event as noise,

$$UR = \frac{FN}{TP + TN},\tag{13}$$

where *TP*, *TN*, and *FN* are defined by Equation (11).

4.4. Experimental Results on TRESS01 Dataset

The training set of the TRESS01 datasets was used as the input of the model, which was trained for 200 epochs (Figure 7). The model basically reached a steady state after the 50th epoch. Then, the classification ability of the trained model was evaluated on the test set. The median results of 11 replicate trials are selected for discussion. Figure 7 shows the accuracy and loss curves of the model on the training and testing datasets. Figure 8 shows the confusion matrix of the model on the test set. The overall correct recognition rate for these three categories is 98.49% (Figure 8). For the results of single target types, the accuracy values of tracked vehicles and noise are high, both exceeding 98%. For people, the classification accuracy is 97.66%. The false alarm rate is 0.43%, and the leakage rate is 1.08%. The experimental results show that the algorithm performs well. Among the misclassified samples are several of people misclassified as ambient noise. A few samples of tracked vehicles are classified as pedestrians and noise. Misclassification is mainly due to ambiguous identification of samples with low signal-to-noise ratios.
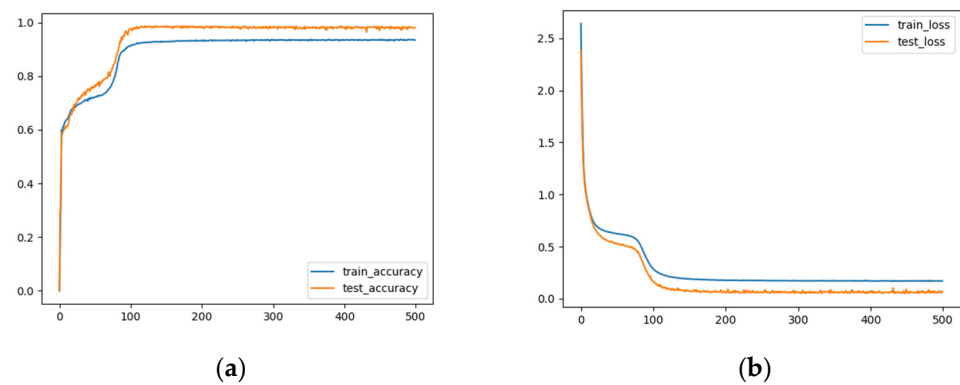
Figure 7. Experimental results of model on the train dataset and validation set. (**a**) Accuracy. (**b**) Loss.
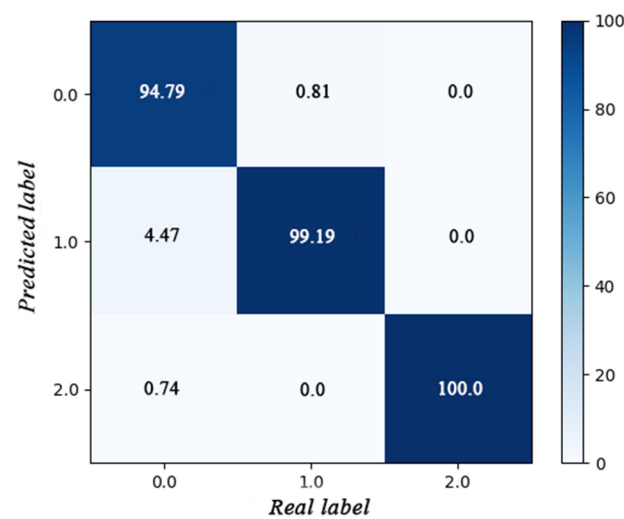

Figure 8. Confusion matrix for the model results on the test set.

4.5. Comparison with Benchmark Methods

Table 4 presents the classification performance of state-of-the-art ML algorithms in the field of ground-based moving target recognition based on seismic and sound signals, for the comparison of classification performance with the proposed method. These algorithms include genetic algorithm optimized support vector machine (GA-SVM) [15], improved BP neural network (BPNN) [28], and Vib-CNN [21]. GA-SVM and BPNN are TML methods, and Vib-CNN is an end-to-end DL method. GA-SVM uses the wavelet decomposition energy ratio, zero-crossing rate, mean, peak, and waveform factor as feature extraction methods, and improved BPNN uses kurtosis, zero-crossing rate, and ratio of high- and low-frequency energy. To fairly evaluate these methods, a model was trained for 11 experiments, each with 200 epochs. The median classification accuracy of the test set was taken as the classification effectiveness, and the computation time on the test set was recorded. All methods were run on a desktop computer with an Intel Core i710 CPU and a GeForce RTX 2070S GPU.

The classification results of the models on the test set were similar, but their running times differed significantly. Vib-CNN had a much greater time than the other models, but its accuracy for target recognition was not much different from that of the other models. The classification performance and operation time of GA-SVM and improved BPNN were similar, and the classification accuracy and operation time of GA-SVM was slightly higher than BPNN. GA-SVM is slightly higher than the proposed method in this paper in terms of classification accuracy, but the performance time of the GA-SVM method is five times higher than the proposed method in this paper. Compared with TML and other DL methods, the

proposed method has lower algorithmic complexity and similar recognition accuracy, and can be deployed at a lower cost and with lower power consumption.

Table 4. Performance comparison between the proposed method and the benchmark methods.

Model	Accuracy	False Alarm Rate	Underreporting Rate	Time
GA-SVM	98.56%	0.96%	0.48%	20.27 s
Improved BPNN	97.47%	0.49%	2.04%	18.37 s
Vib-CNN	98.16%	0.43%	1.41%	200.83 s
ENN (Proposed in this paper)	98.21%	1.31%	0.49%	4.54 s

5. Conclusions

We proposed a ground moving target recognition method based on an evolutionary neural network by combining a genetic algorithm and neural network, which achieves high recognition accuracy with simple feature extraction methods and neural networks. By evolving the neural network, kurtosis and vibration energy were determined as feature extraction methods for seismic and sound signals, respectively, and a fully connected neural network with two hidden layers was obtained. The accuracy of the method for target classification reached 98.21% in the experiments. Compared with other methods, this method can be deployed at a lower cost and with less power consumption.

Author Contributions: Conceptualization, K.X. and N.W.; methodology, K.X.; software, K.X.; validation, K.X. and W.W.; formal analysis, K.X. and N.W.; investigation, K.X.; resources, K.X.; data curation, K.X.; writing—original draft preparation, K.X.; writing—review and editing, N.W. and W.W.; supervision, N.W.; project administration, N.W.; funding acquisition, N.W. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the author.

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

References

1. Bojor, L.; Cîrdei, I.A. Unattended Ground Sensor Borders—The Forgotten Solution for Afghanistan. In Proceedings of the International Conference Knowledge-Based Organization, Sibiu, Romania, 11 June 2021; Volume 27, pp. 8–13.
2. Li, F.; Clemente, J.; Valero, M.; Tse, Z.; Song, W.Z. Smart Home Monitoring System via Footstep Induced Vibrations. *IEEE Syst. J.* **2019**, *14*, 3383–3389. [CrossRef]
3. Clemente, J.; Li, F.; Valero, M.; Song, W.Z. Smart Seismic Sensing for Indoor Fall Detection, Location, and Notification. *IEEE J. Biomed. Health* **2019**, *24*, 524–532. [CrossRef]
4. Mukhopadhyay, B.; Anchal, S.; Kar, S. Detection of an Intruder and Prediction of His State of Motion by Using Seismic Sensor. *IEEE Sens. J.* **2017**, *18*, 703–712. [CrossRef]
5. Trnkoczy, A. Understanding and parameter setting of STA/LTA trigger algorithm. In *IASPEI New Manual of Seismological Observatory Practice (NMSOP)*; Deutsches GeoForschungsZentrum GFZ: Potsdam, Germany, 2002; Volume 2, pp. 1–19.
6. Anchal, S.; Mukhopadhyay, B.; Kar, S. Predicting gender from footfalls using a seismic sensor. In Proceedings of the 2017 9th International Conference on Communication Systems and Networks (COMSNETS), Bengaluru, India, 4–8 January 2017; pp. 47–54.
7. Damarla, T.; Mehmood, A.; Sabatier, J. Detection of people and animals using non-imaging sensors. In Proceedings of the 14th International Conference on Information Fusion, Chicago, IL, USA, 5–8 July 2011; pp. 1–8.
8. Sabatier, J.M.; Ekimov, A.E. A Review of Human Signatures in Urban Environments Using Seismic and Acoustic Methods. In Proceedings of the IEEE Conference on Technologies for Homeland Security, Waltham, MA, USA, 12–13 May 2008.
9. Evans, N. Automated Vehicle Detection and Classification Using Acoustic and Seismic Signals. Ph.D. Thesis, University of York, York, UK, 2010.

10. Choudhary, P.; Goel, N.; Saini, M. A Seismic Sensor based Human Activity Recognition Framework using Deep. In Proceedings of the 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 16–19 November 2021; IEEE: Washington, DC, USA, 2021; pp. 1–8.

11. Wellman, M.C.; Srour, N.; Hillis, D.B. Feature extraction and fusion of acoustic and seismic sensors for target identification. In Proceedings of the Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors, Orlando, FL, USA, 22–23 April 1997; SPIE: Orlando, FL, USA, 1997; pp. 139–145.

12. Cheng, Z.; Gao, M.; Liang, X.; Liu, L. Incipient fault detection for the planetary gearbox in rotorcraft based on a statistical metric of the analog tachometer signal. *Measurement* **2019**, *151*, 107069. [CrossRef]

13. Küçükbay, S.E.; Sert, M.; Yazici, A. Use of Acoustic and Vibration Sensor Data to Detect Objects in Surveillance Wireless Sensor Networks. In Proceedings of the International Conference on Control Systems & Computer Science, Bucharest, Romania, 29–31 May 2017.

14. Bin, K.; Long, Y.; Tong, X.; Lin, J. Ground Moving Target Detection with Seismic Fractal Features. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]

15. Zhong, Z.; Li, H. Recognition and prediction of ground vibration signal based on machine learning algorithm. *Neural Comput. Appl.* **2020**, *32*, 1937–1947. [CrossRef]

16. Jin, X.; Gupta, S.; Ray, A.; Damarla, T. Multimodal sensor fusion for personnel detection. In Proceedings of the International Conference on Information Fusion, Chicago, IL, USA, 5–8 July 2011.

17. Huang, J.; Zhou, Q.; Zhang, X.; Song, E.; Li, B.; Yuan, X. Seismic Target Classification Using a Wavelet Packet Manifold in Unattended Ground Sensors Systems. *Sensors* **2013**, *13*, 8534–8550. [CrossRef] [PubMed]

18. Narayanaswami, R.; Gandhe, A.; Tyurina, A.; Mehra, R.K. Sensor fusion and feature-based human/animal classification for Unattended Ground Sensors. In Proceedings of the 2010 IEEE International Conference on Technologies for Homeland Security (HST), Waltham, MA, USA, 8–10 November 2010.

19. Dibazar, A.A.; Yousefi, A.; Park, H.O.; Bing, L.; George, S.; Berger, T.W. Intelligent acoustic and vibration recognition/alert systems for security breaching detection, close proximity danger identification, and perimeter protection. In Proceedings of the 2010 IEEE International Conference on Technologies for Homeland Security (HST), Waltham, MA, USA, 8–10 November 2010.

20. Bin, K.; Luo, S.; Zhang, X.; Lin, J.; Tong, X. Compressive Data Gathering with Generative Adversarial Networks for Wireless Geophone Networks. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 558–562. [CrossRef]

21. Wang, Y.; Cheng, X.; Zhou, P.; Li, B.; Yuan, X. Convolutional neural network-based moving ground target classification using raw seismic waveforms as input. *IEEE Sens. J.* **2019**, *19*, 5751–5759. [CrossRef]

22. Bin, K.; Lin, J.; Tong, X. Edge Intelligence-Based Moving Target Classification Using Compressed Seismic Measurements and Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]

23. Jin, G.; Ye, B.; Wu, Y.; Qu, F. Vehicle Classification Based on Seismic Signatures Using Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 628–632. [CrossRef]

24. Xu, T.; Wang, N.; Xu, X. Seismic Target Recognition Based on Parallel Recurrent Neural Network for Unattended Ground Sensor Systems. *IEEE Access* **2019**, *7*, 137823–137834. [CrossRef]

25. Li, X.; Wang, N.; Ding, Y.; Xing, K. Research on Moving Ground Target Classification Method for Seismic and Acoustic Sensors. In Proceedings of the International Conference on Autonomous Unmanned Systems, Changsha, China, 24–26 September 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 852–859.

26. Sabatier, J.M.; Ekimov, A.E. Range limitation for seismic footstep detection. In Proceedings of the SPIE Defense and Security Symposium, Orlando, FL, USA, 16–20 March 2008; SPIE: Bellingham, WA, USA, 2008; Volume 6963, pp. 247–258.

27. Liu, S.; Jiang, W.; Wu, L.; Wen, H.; Wang, Y. Real-Time Classification of Rubber Wood Boards Using an SSR-Based CNN. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8725–8734. [CrossRef]

28. Peng, Z.-Q.; Cao, C.; Huang, J.-Y.; Liu, Q.-S. Seismic signal recognition using improved BP neural network and combined feature extraction method. *J. Cent. South Univ.* **2014**, *21*, 1898–1906.