IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 9, 2019, accepted July 9, 2019, date of publication August 14, 2019, date of current version October 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2934893

Seismic Target Recognition Based on Parallel Recurrent Neural Network for Unattended Ground Sensor Systems

TONGHUA XU, NAN WANG, AND XIN XU, (Senior Member, IEEE)
College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China

Corresponding author: Nan Wang (acnwang@aliyun.com)

This work was supported in part by the National Research and Development Program of China under Grant 2018YFB1305105, and in part by the National Natural Science Foundation of China under Grant U1564214 and Grant 61751311.

ABSTRACT One of the most challenging problems in unattended ground sensor (UGS) systems is to distinguish human footstep signals from noise sources. As the distance between the sensor and the moving target increases, the signal-to-noise ratio (SNR) decreases rapidly. Many methods have been proposed to solve this problem, but most of them suffer from unacceptably high false alarm rate or omissive report rate in practical applications. In this paper, a novel approach based on parallel recurrent neural network (PRNN) is proposed to improve the seismic target recognition performance. The PRNN is composed of a time domain feature network and a frequency spectrum feature network. The time domain feature network is used to handle the running signals, and the frequency spectrum feature network is used to handle the walking signals. The output of the PRNN is a fusion of the two networks. Experimental results show that the proposed approach can improve the human recognition accuracy up to 98.3% and has a remarkable performance compared with other machine learning methods.

INDEX TERMS Data fusion, parallel recurrent neural network, signal-to-noise ratio, unattended ground sensor.

I. INTRODUCTION

Unattended Ground Sensor (UGS) was first used to monitor Viet Cong activity along the Ho Chi Minh Trail in the 1960s [1]. Nowadays UGS is widely used to monitor human activities, such as pedestrian movement and intruder detection in safe areas [2], [3]. The most commonly used sensor devices in UGS systems are micro-electro-mechanical-systems (MEMS) accelerometers [4], MEMS microphone array [5], quartz MEMS vibrating beam seismometer [6] and other seismic sensors [7]–[12]. The MEMS technology can integrate signal processing circuit and sensor element into a single chip, providing us with the control mechanism of sensor calibration and initial signal processing [4].

In UGS systems, the detection of human activities has been paid more and more attention [13]–[18]. Most methods for detecting human activities employ footstep seismic signal detection. Nevertheless, distinguishing human footstep signals from noise sources is a challenging problem, because the signal-to-noise ratio (SNR) decreases rapidly as the distance between the sensor and the moving target increases. In addition, the footstep signals have different features for different people, which makes the problem of target detection and classification more challenging [19]. Therefore, we need to find common features. The analysis of the human footstep seismic signals is an important part of UGS systems.

Current feature extraction methods for seismic signals can be divided into three categories: time domain [20], [21], frequency domain [4] and time-frequency domain [19], [22]. On the one hand, due to the interference noise, complex signal waveforms and changes in terrain, the time domain method may not be able to identify targets very accurately [22]. On the other hand, due to the potential non-stationary in the observed signal, the accuracy of frequency domain approach may be reduced [3]. Therefore, recent research has focused on time-frequency domain methods (e.g., wavelet transform) because of denoising and localization properties [3].

The associate editor coordinating the review of this article and approving it for publication was Anandakumar Haldorai.

The conventional seismic target recognition methods mainly include wavelet packet manifold (WPM) [23], symbolic dynamic filtering (SDF) [9], [19], multi-axial seismic fusion [24], dynamic data-driven application system (DDDAS) [25], multimodal sensor fusion [26], etc. WPM proposes the neighborhood preserving embedding algorithm of manifold learning on the wavelet packet node energy of seismic signals [23]. SDF contains wavelet-transformed preprocessing of signals to improve time-frequency localization and denoising [19]. In addition, BP neural network is used to extract the features of seismic signals, which has achieved better results [13]. However, the problem is that system performance is often limited by high false alarm rate or omissive report rate, which may be a lack of underlying algorithms for extracting features from sensory data and subsequent pattern classification.

In this paper, the parallel recurrent neural network (PRNN) is proposed for seismic target recognition. Specifically, the target recognition system is produced by the four following steps. First, the time-frequency analysis of the seismic signals of the human footsteps (including walking and running) obtained by seismic sensor device is conducted. The features of running signals in the time domain are more obvious, while the features of the walking signals in the frequency domain are more obvious. Second, there are some feature extractions need to be taken. The time domain signals use fast bandpass filter and peak value region extraction, while the frequency domain signals are obtained by Welch algorithm. Third, the PRNN for target recognition applications is designed, which consists of two Long Short-term Memory networks (LSTMs). The input to an LSTM is time series data and this LSTM is used to determine whether someone is running; the input to another LSTM is frequency spectrum data and this LSTM is used to determine whether someone is walking. Fourth, the network output data is used for data fusion, performing OR operation, thereby effectively identifying ground human activities. Experiments show that PRNN achieves high recognition accuracy. Because the proposed PRNN can improve the accuracy of classification up to 98.3%, it may be widely used in UGS.

This paper is organized as follows. In Section II, data acquisition and feature extractions such as fast bandpass filter, peak value region extraction, and Welch algorithm are presented in detail. In Section III, the PRNN for seismic target recognition and proof of accuracy formulas are elaborated. In Section IV, time-frequency analysis, experiments and results are presented in detail. Finally, Section V discusses conclusions and future research.

II. DATA ACQUISITION AND FEATURE EXTRACTIONS

A. DATA ACQUISITION

The seismic sensor device collects the continuous time signal $x(t)$ from the environment and samples according to the time interval $\Delta t = 1ms$, thereby obtaining a discrete time signal $x(n \circ \Delta t) \equiv x_n$, $n = 0, 1, 2, \cdots, N-1$, where

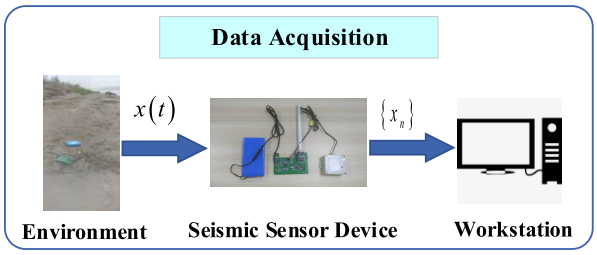


FIGURE 1. Diagram of data acquisition.

N is the length of the signal (The following is denoted as $\{x_n\}$). Finally, the sensor device sends the signal $\{x_n\}$ to the workstation (see Fig. 1). The environment and scene for data acquisition are detailed in Section IV.

B. FEATURE EXTRACTIONS

In the UGS system, the personnel target moves on the ground. For the ground soil layer, the target gives a certain incentive to the ground. From the perspective of geophysics, since the earth medium is non-rigid, this kind of incentive will definitely cause tiny deformation of the earth medium, which will propagate in the earth medium and form seismic waves. In order to effectively detect the seismic signal caused by the target motion, it is necessary to extract the signal features (see Fig. 2).

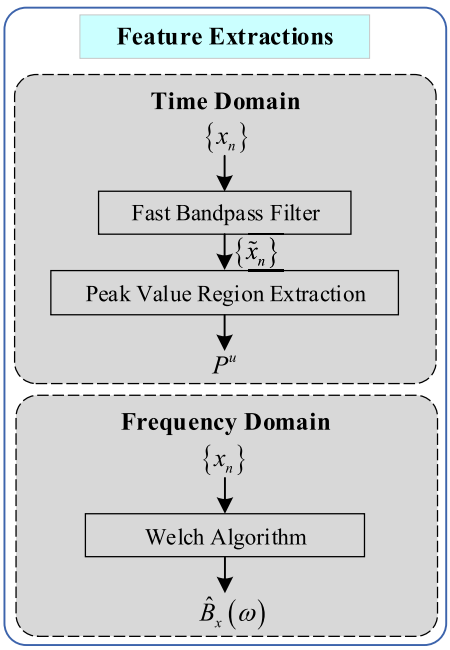


FIGURE 2. Diagram of feature extractions.

1) FAST BANDPASS FILTER

As the distance between the human target and the seismic sensor device increases, the SNR drops rapidly. In order to effectively remove the noise, the original signal $\{x_n\}$ is performed by a fast bandpass filter. Let $\Delta\omega$ be $\frac{2\pi}{N \circ \Delta t}$. According

to [27], the Discrete Fourier Transform (DFT) of $\{x_n\}$ is given as:

$$X_m = X(m \circ \Delta\omega) = \sum_{n=0}^{N-1} x_n e^{-i2\pi mn/N} \quad (m = 0, 1, 2, \cdots, N-1) \tag{1}$$

Discrete-time "analytic" signal $\xi_n = x_n + ix_n (n = 0, 1, 2, \cdots, N-1)$ of $\{x_n\}$ is introduced from [28]. Hence, the DFT of $\xi_n = x_n + ix_n$ is given as:

$$Z_m = \begin{cases} 2X_m & m_j \le m \le m_k \\ 0 & \text{others} \end{cases} \quad (m = 0, 1, 2, \cdots, N-1) \tag{2}$$

where $\{Z_m\}$ is the DFT of $\{\xi_n\}$. When subscript m is beyond the range $m_j \le m \le m_k$, the sequence $\{Z_m\}$ is equal to 0. The discrete fourier inverse transform of $\{Z_m\}$ is $\tilde{x}_n + i\tilde{x}_n$, where $\{\tilde{x}_n\}$ is a real part of $\{\xi_n\}$. Hence, $\{\tilde{x}_n\}$ is the filtered signal of the original signal $\{x_n\}$ passing the frequency band $m_j \le m \le m_k$.

2) PEAK VALUE REGION EXTRACTION

Seismic signals caused by personnel movements have periodicity feature, that is, peaks occur at intervals. In order to extract the peak value region signal, let the region width be W (take an even number), let the step length l, and let the peak region sequence be P^u, $u = 1, 2, \cdots, \coprod (N/l)$. In step u, the subscript corresponding to the maximum value in $\{\tilde{x}_n\}$ is $index\left(\max_{(u-1)l \le i < ul} (\tilde{x}_i)\right)$, where $index()$ is the subscript corresponding to value. The sequence of peak value region in step u is given as:

$$P^u = norm(\{\tilde{x}_i\}), \quad index\left(\max_{(u-1)l \le i < ul} (\tilde{x}_i)\right) - \frac{W}{2} \le i < index\left(\max_{(u-1)l \le i < ul} (\tilde{x}_i)\right) + \frac{W}{2} \tag{3}$$

where $norm()$ is the normalization of the sequence, and the normalization range is from 0 to 1000.

3) WELCH ALGORITHM

In order to effectively extract the frequency domain features, the original signal $\{x_n\}$ needs to be estimated by power spectral density (PSD). This paper employs the Welch algorithm, which is an improved method of the periodogram [29], [30]. The length N of $\{x_n\}$ is divided into P segments and each of them has M data. The correction periodogram in segment p is given as:

$$J_p(\omega) = \frac{1}{MU} \left| \sum_{n=0}^{M-1} x_p(n) w(n) e^{-j\omega n} \right|^2 \tag{4}$$

where $U(\omega) = \frac{1}{M} \sum_{n=0}^{M-1} w^2(n)$ is the normalization factor and $w(n)$ is the added window function. The periodograms of P segments are averaged to obtain an estimation $\hat{B}_x(\omega)$ of the PSD of the whole signal, as shown in Eq. (5):

$$\hat{B}_x(\omega) = \frac{1}{P} \sum_{p=0}^{P} J_p(\omega) \tag{5}$$

III. PARALLEL RECURRENT NEURAL NETWORK

Target recognition is a key factor in the UGS system. This paper proposes the PRNN for target detection and classification. The whole process of the system from acquiring data to building a target recognition system is shown in Fig. 3. The first step is data acquisition: seismic sensor device is used to obtain environmental information, and the acquired information is transmitted to the workstation. The second step is feature extractions: In the time domain, fast bandpass filter and peak value region extraction are implemented to extract time domain feature. In the frequency domain, Welch algorithm is implemented to extract frequency domain feature. The third step is target recognition: the time series data and the frequency spectrum data are respectively trained using different LSTMs. Outputs of the two LSTMs are fused by performing OR operation. The seismic sensor device and data acquisition scene are elaborated in the experiment. Feature extractions have been described in Section II. This section elaborates on the proposed PRNN for target recognition.

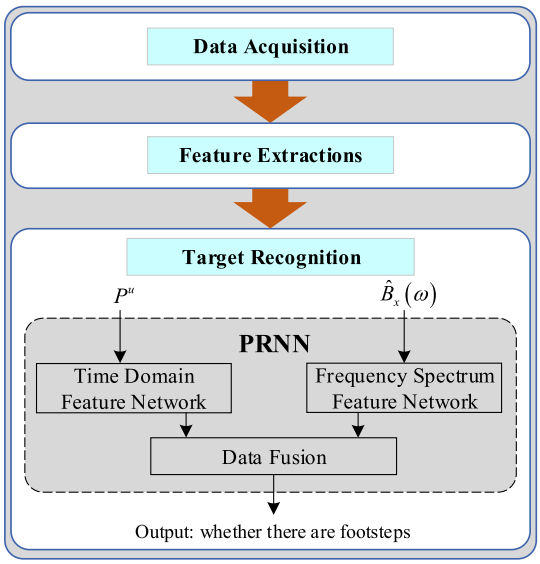


FIGURE 3. Overall structure of the proposed method.

A. PRNN BUILDING AND TRAINING

In this subsection, we propose the PRNN in target recognition system (see Fig. 4). The inputs to the PRNN are time series data P^u and frequency spectrum data $\hat{B}_x(\omega)$. The output is the result of the classification. The time series data P^u is obtained by fast bandpass filter and peak value region extraction, the frequency spectrum data $\hat{B}_x(\omega)$ is obtained by Welch algorithm.

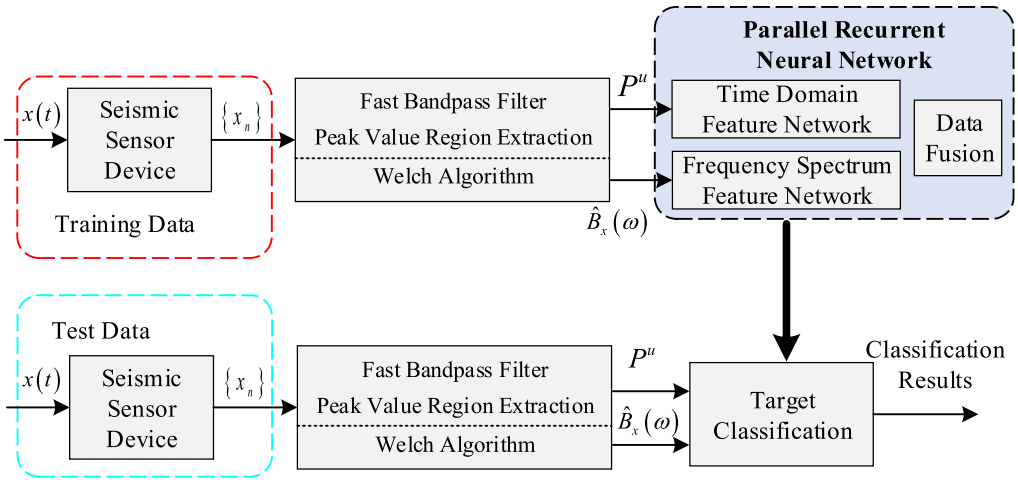


FIGURE 4. PRNN structure in the target recognition system.

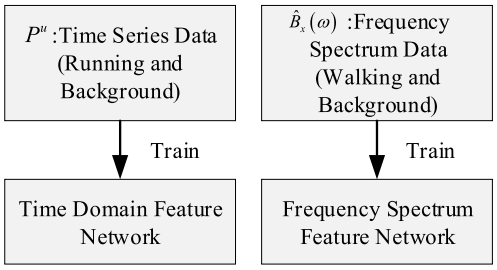


FIGURE 5. Diagram of PRNN training.

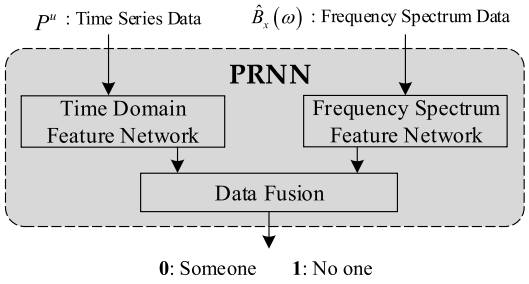


FIGURE 6. Target classification of PRNN.

In order to improve the generalization ability of the network, the PRNN needs to be trained by large amounts of data. The PRNN consists of two LSTMs [31], one is a time domain feature network and the other is a frequency spectrum feature network. Each network employs an LSTM architecture that uses purpose-built memory cells to store information and is better at discovering and exploiting long-term dependencies in the data [31]. The training data of the time domain feature network is time series data P^u, wherein the data is acquired only in the background and the running scenes; the training data of the frequency spectrum feature network is the frequency spectrum data $\hat{B}_x(\omega)$, wherein the data is acquired only in the background and the walking scenes. The training samples of both LSTMs contain labels, where label 0 represents no one (background signal) and label 1 represents someone (walking or running signal). Therefore, the network belongs to supervised learning. Diagram of PRNN training is shown in Fig. 5.

B. TARGET CLASSIFICATION

Target classification of PRNN is shown in Fig. 6, where PRNN has been trained. PRNN has two inputs: P^u and $\hat{B}_x(\omega)$. P^u includes three types of signals: background, walking and running. $\hat{B}_x(\omega)$ also includes three types of signals: background, walking and running. The time domain feature network is used to accurately identify the running signal, and the frequency spectrum feature network is used to accurately identify the walking signal. The output data of the two are fused. The data fusion is OR operation, which accurately identifies the moving personnel target (the relevant formula is proved below).

C. DERIVATION OF TARGET RECOGNITION ACCURACY FORMULA

According to the test accuracy of two LSTMs, the test accuracy range of the PRNN can be calculated. The derivation process is given below.

Assume that the test accuracy of the time domain feature network is P_1. Assume that the test accuracy of the frequency spectrum feature network is P_2. The number of samples used for training and testing is very large, and the test accuracy of the network converges in training. The generated network test results are basically consistent with the final results of the training network, so P_1 and P_2 are representative to some extent.

Assume that the number of test samples of the background signal after fast bandpass filter and peak value region extraction is x_1. Similarly, the number of test samples of the walking signal is x_2, and the number of test samples of the running signal is x_3. Assume that the number of the test samples of

the background signal after Welch algorithm process is $\hat{x}_1$. Similarly, the number of test samples of the walking signal is $\hat{x}_2$, and the number of test samples of the running signal is $\hat{x}_3$. Since the proposed PRNN judges the type of input data of different forms for the same signal, the following equation is given as:

$$x_1 = \hat{x}_1, \quad x_2 = \hat{x}_2, \; x_3 = \hat{x}_3 \tag{6}$$

In the time domain feature network, it is assumed that among the x_1 samples of the background signal, x_{11} samples are recognized as no one, and x_{13} samples are recognized as someone. It is assumed that among the x_2 samples of the walking signal, x_{21} samples are recognized as no one, and x_{23} samples are recognized as someone. It is assumed that among the x_3 samples of the running signal, x_{31} samples are recognized as no one, and x_{33} samples are recognized as someone. In the frequency spectrum feature network, it is assumed that among the $\hat{x}_1$ samples of the background signal, $\hat{x}_{11}$ samples are recognized as no one, and $\hat{x}_{12}$ samples are recognized as someone. It is assumed that among the $\hat{x}_2$ samples of the walking signal, $\hat{x}_{21}$ samples are recognized as no one, and $\hat{x}_{22}$ samples are recognized as someone. It is assumed that among the $\hat{x}_3$ samples of the running signal, $\hat{x}_{31}$ samples are recognized as no one, and $\hat{x}_{32}$ samples are recognized as someone. According to the above assumption, the following equation is given as:

$$\begin{cases} x_1 = x_{11} + x_{13} \\ x_2 = x_{21} + x_{23} \\ x_3 = x_{31} + x_{33} \\ \hat{x}_1 = \hat{x}_{11} + \hat{x}_{12} \\ \hat{x}_2 = \hat{x}_{21} + \hat{x}_{22} \\ \hat{x}_3 = \hat{x}_{31} + \hat{x}_{32} \end{cases} \tag{7}$$

Since the test accuracy P_1 of the time domain feature network is obtained by the background signal and the running signal test, and the test accuracy P_2 of the frequency spectrum feature network is obtained by the background signal and the walking signal test, so the following equations are given as:

$$P_1 = \frac{x_{11} + x_{33}}{x_1 + x_3} \tag{8}$$

$$P_2 = \frac{\hat{x}_{11} + \hat{x}_{22}}{\hat{x}_1 + \hat{x}_2} \tag{9}$$

In the PRNN, the outputs of the time domain feature network and the frequency spectrum feature network perform OR operation, and the result is the output of the PRNN. In two separate networks, the output 0 represents no one and the output 1 represents someone. The output of the PRNN has the same meaning as above. Their relationship is shown in Table 1.

As can be seen from Table 1, the PRNN output is recognized as no one if and only if both the time domain feature network and the frequency spectrum feature network are recognized as no one. Assume that the test accuracy of the PRNN is P, and the test error rate is P', so the following equation is given as:

TABLE 1. Output relationship between networks.

Cases	Time Domain Feature Network Output	Frequency Spectrum Feature Network Output	PRNN Output
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	1

$$P = 1 - P' \tag{10}$$

Three cases of test errors in the PRNN are given as:

- Any network in the time domain feature network and the frequency spectrum feature network recognizes samples of the background signal as someone. Suppose the set X_{13} contains all samples that misidentify the background signals as someone in the time domain feature network. Suppose the set $\hat{X}_{12}$ contains all samples that misidentify the background signals as someone in the frequency spectrum feature network. Therefore, the set $X_{13} \cup \hat{X}_{12}$ contains all samples that misidentify background signals as someone in any network.
- The time domain feature network and the frequency spectrum feature network simultaneously recognize samples of the walking signals as no one. Suppose the set X_{21} contains all samples that misidentify the walking signals as no one in the time domain feature network. Suppose the set $\hat{X}_{21}$ contains all samples that misidentify the walking signals as no one in the frequency spectrum feature network. Therefore, the set $X_{21} \cap \hat{X}_{21}$ contains all samples that simultaneously misidentify the walking signals as no one in two networks.
- The time domain feature network and the frequency spectrum feature network simultaneously recognize samples of the running signals as no one. Suppose the set X_{31} contains all samples that misidentify the running signals as no one in the time domain feature network. Suppose the set $\hat{X}_{31}$ contains all samples that misidentify the running signals as no one in the frequency spectrum feature network. Therefore, the set $X_{31} \cap \hat{X}_{31}$ contains all samples that simultaneously misidentify the running signals as no one in two networks.

According to the above three cases, the following equation is given as:

$$P' = \frac{num\left(X_{13} \cup \hat{X}_{12}\right) + num\left(X_{21} \cap \hat{X}_{21}\right) + num\left(X_{31} \cap \hat{X}_{31}\right)}{x_1 + x_2 + x_3} \tag{11}$$

where $num\,()$ is the number of the set elements. According to the relationship between the sets, the following equation is

given as:

$$\begin{cases} num\left(X_{13} \cup \hat{X}_{12}\right) \leq num\left(X_{13}\right) + num\left(\hat{X}_{12}\right) = x_{13} + \hat{x}_{12} \\ num\left(X_{21} \cap \hat{X}_{21}\right) \leq num\left(\hat{X}_{21}\right) = \hat{x}_{21} \\ num\left(X_{31} \cap \hat{X}_{31}\right) \leq num\left(X_{31}\right) = x_{31} \end{cases} \tag{12}$$

According to Eqs. (11) and (12), the following equation is given as:

$$P' \leq \frac{x_{13} + \hat{x}_{12} + \hat{x}_{21} + x_{31}}{x_1 + x_2 + x_3} \tag{13}$$

The transformation is done as follows:

$$\frac{x_{13} + \hat{x}_{12} + \hat{x}_{21} + x_{31}}{x_1 + x_2 + x_3} = \frac{x_{13} + x_{31}}{x_1 + x_2 + x_3} + \frac{\hat{x}_{12} + \hat{x}_{21}}{x_1 + x_2 + x_3} \tag{14}$$

According to Eqs. (6), (13) and (14), the following equation is given as:

$$P' \leq \frac{x_{13} + x_{31}}{x_1 + x_2 + x_3} + \frac{\hat{x}_{12} + \hat{x}_{21}}{x_1 + x_2 + x_3} \leq \frac{x_{13} + x_{31}}{x_1 + x_3} + \frac{\hat{x}_{12} + \hat{x}_{21}}{\hat{x}_1 + \hat{x}_2} \tag{15}$$

According to Eqs. (7), (8), (9) and (15), the following equation is given as:

$$\begin{aligned} P' &\leq 1 - \frac{x_{11} + x_{33}}{x_1 + x_3} + 1 - \frac{\hat{x}_{11} + \hat{x}_{22}}{\hat{x}_1 + \hat{x}_2} = 1 - P_1 + 1 - P_2 \\ &= 2 - P_1 - P_2 \end{aligned} \tag{16}$$

According to Eqs. (10) and (16), the test accuracy of the PRNN is given as:

$$P = 1 - P' \geq 1 - (2 - P_1 - P_2) = P_1 + P_2 - 1 \tag{17}$$

In the experiment, the values of P_1 and P_2 are close to 1, so the test accuracy P of the PRNN is also close to 1.

IV. EXPERIMENTS AND RESULTS

A. DATA COLLECTING SYSTEM

As can be seen from Fig. 7(a), the seismic sensor device contains a power supply, a printed circuit board (PCB), and a Micro-Electro-Mechanical-Systems (MEMS) sensor. The device acquires ground seismic information through the MEMS sensor, and then wirelessly transmits it to the workstation after simple processing. The seismic sensor device uploads data packet every 5 milliseconds through the wireless network, and 1000 data can be collected per second.

The environment selected in this paper is on hard soil, as shown in Fig. 8(a). As can be seen from Fig. 8(a), the MEMS sensor is buried underground and the depth is about 10 cm. Someone on the ground moves within 30 meters of the seismic sensor device. In order to enable the seismic sensor device to collect seismic data at different distances, we adopt the diameter method. The illustration of the data collecting scene is shown in Fig. 8(b).

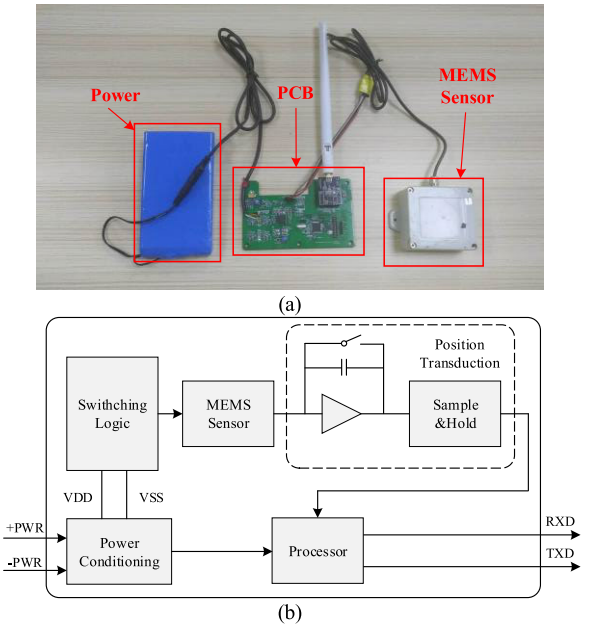


FIGURE 7. (a) Instance diagram of the seismic sensor device. (b) Schematic diagram of circuit design principle.

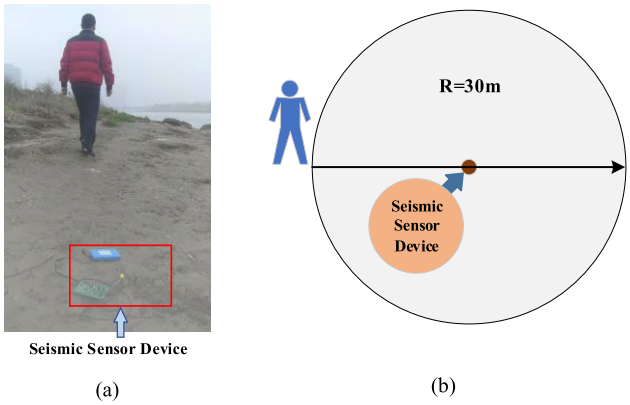


FIGURE 8. (a) Diagram of the data acquisition environment. (b) Illustration of data acquisition scene.

The seismic sensor device transmits the collected data to the workstation in real time. The CPU of the workstation is an Intel Core i7-5960X. The GPUs are two NVIDIA GeForce GTX 1080Ti 11GB. The RAM size is 32GB. We implemented PRNN using Tensorflow 1.2.0 in Ubuntu 14.04 LTS and experimented with the open-source neural network library written in Python 2.7.13.

B. FEATURE EXTRACTION RESULTS

1) ORIGINAL SIGNALS

The time domain diagram of original signal is shown in Fig. 9. The signal in Fig. 9(a) is the data acquired without humans, i.e. the background signal. The signal in Fig. 9(b) is the data collected in a walking scene, i.e. the walking signal. The signal in Fig. 9(c) is the data collected in a running scene, i.e. the running signal.

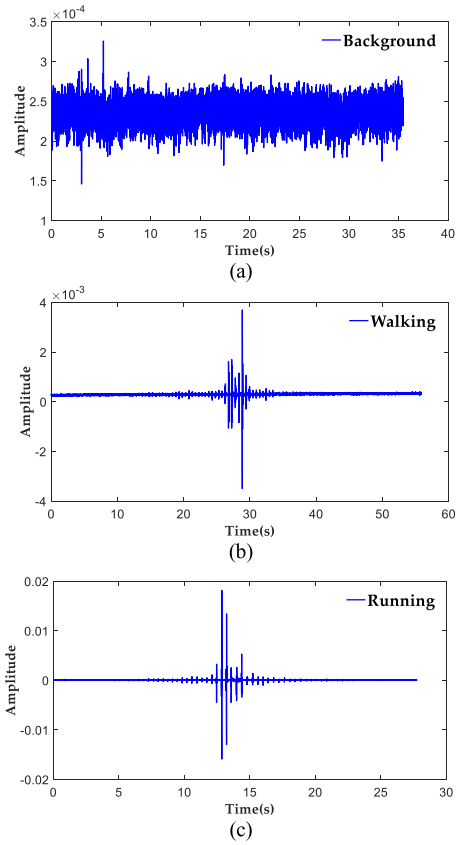


FIGURE 9. (a) Time domain diagram of the background signal. (b) Time domain diagram of the walking signal. (c) Time domain diagram of the running signal.

From Fig. 9, it can be seen that the background time series data is a random signal. The time series signals of walking and running have large vibration amplitudes when the person is close to the seismic sensor device. On the contrary, the vibration amplitudes are small when the person is away from the sensor device. From the longitudinal axis of the three figures in Fig. 9, it can be seen that the vibration amplitude of the background signal is small overall, while that of the walking and running signals is generally large. From the time axis, it can be found that the walking and running signals have periodicity feature, i.e., peaks occur at intervals. From the point of view of noise interference, SNR of the walking and running signals is obviously small when the human target is far away from the seismic sensor device.

2) PRELIMINARY EXTRACTION OF TIME DOMAIN FEATURE

According to Section II, the frequency band of fast bandpass filter is $[m_j, m_k]$, where m_j is less than m_k. As can be seen in Fig. 10(b), rhythm features of the running signal mainly concentrate in the frequency band from 25 Hz to 50 Hz. So here the frequency m_j is 25 Hz and the frequency m_k is 50 Hz.

Fig. 11(a) is the original time series running signal, and Fig. 11(b) is an enlarged view of the first 3 seconds of the

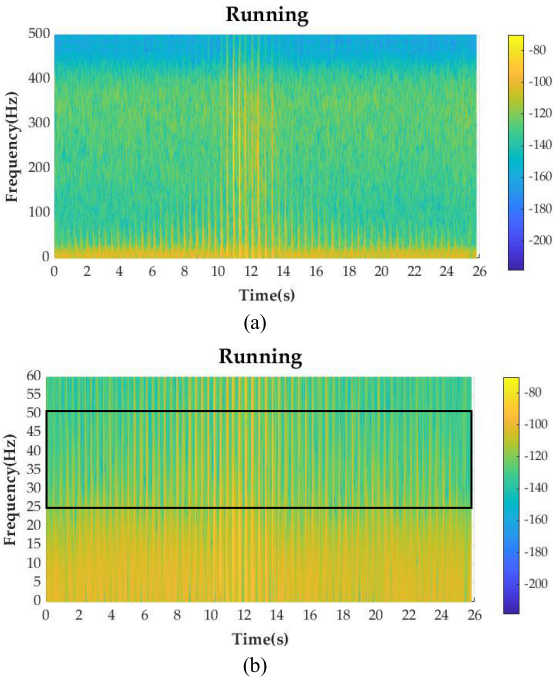


FIGURE 10. (a) Diagram of the spectrogram of the running signal. (b) Enlarged view of the red area of (a).

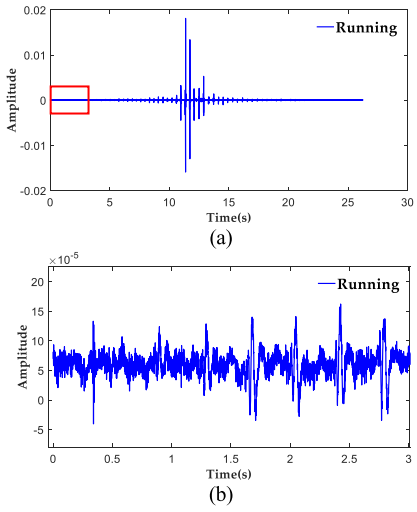


FIGURE 11. (a) Original time series running signal. (b) Enlarged view of the red area of (a).

signal. Fig. 12(a) is the fast bandpass filtered time series running signal, and Fig. 12(b) is an enlarged view of the first 3 seconds of the signal. As can be seen from Fig. 11 and Fig. 12, the rhythm features of the fast bandpass filtered running signal are very obvious. Meanwhile, the SNR increases significantly in Fig. 12(b). The peak of the running signal appears approximately every 0.375 seconds in Fig. 12(b).

3) FURTHER EXTRACTION OF TIME DOMAIN FEATURE

Region width W is 500, and step length l is 400. The peak region sequence P^u is shown in Fig. 13. The difference in

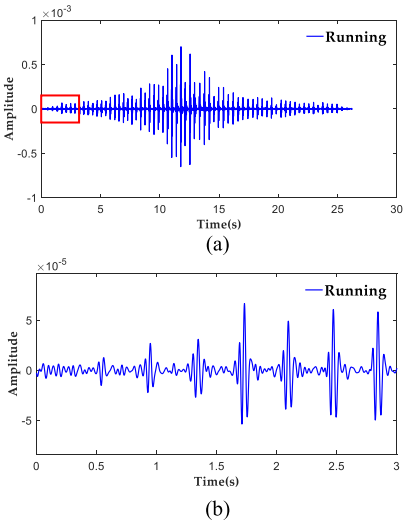


FIGURE 12. (a) Fast bandpass filtered time series running signal. (b) Enlarged view of the red area of (a).

features between the running signal and the background signal is very obvious.

4) EXTRACTION OF FREQUENCY DOMAIN FEATURE

The frequency domain diagrams of the background, walking and running signals using Welch algorithm are shown in Fig. 14.

As can be seen from Fig. 14(a) and Fig. 14(b), the frequency domain features of the background signal and the walking signal are significantly different. Especially in the 150-350 Hz band, the PSD curve of the walking signal is relatively smooth.

C. TARGET CLASSIFICATION RESULTS

The PRNN consists of two LSTMs, whose training samples are different. The training samples for the time domain feature network are time series data for the background and running signals. The training samples for the frequency spectrum feature network are the frequency spectrum data for the background and walking signals. When all networks are training, the training samples are divided into 80% training set and 20% test set. In order to make the network more general, the experiment collected the footstep seismic data of 5 people. In order to effectively test the generalization ability of the network, only 4 people's data are used for training, and 5 people's data are used for testing.

The sample grouping of the time domain feature network is shown in Table 2. The human activity data in Table 2 were collected in the running scene.

From the data in Table 2, the total number of training samples of the time domain feature networks is 5389, and the number of test samples is 931. The loss function graph and the test accuracy graph of the time domain feature network training are shown in Fig. 15. It can be seen from Fig. 15 that when the time domain feature network training is iterated

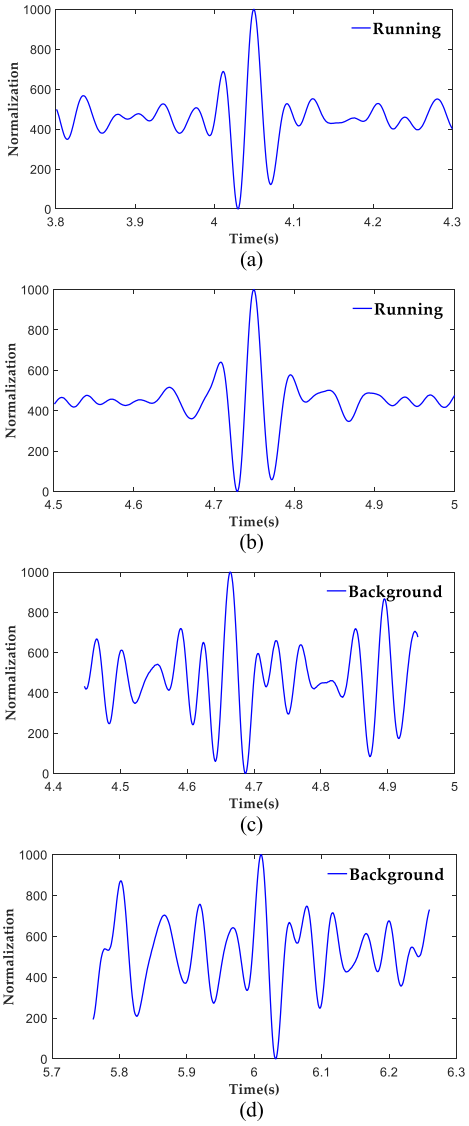


FIGURE 13. Peak value region extraction of the signals at different moments. (a) The effect diagram of the running signal in 3.8-4.3s. (b) The effect diagram of the running signal in 4.5-5.0s. (c) The effect diagram of the background signal in 4.45-4.95s. (d) The effect diagram of the background signal in 5.76-6.26s.

450 times, both the accuracy and the loss function value of the test set converge. The accuracy of the test set of 800 training iterations is 98.52%, and the loss function value is 0.3274.

The sample grouping of the frequency spectrum feature network is shown in Table 3. The human activity data in Table 3 were collected in the walking scene.

From the data in Table 3, the total number of training samples of the frequency spectrum feature network is 9228, and the number of test samples is 1225. The loss function graph and the test accuracy graph of the frequency spectrum feature network training are shown in Fig. 16. It can be seen from Fig. 16 that when the frequency spectrum feature network training is iterated 1400 times, both the accuracy and

TABLE 2. The sample grouping of the time domain feature network.

Scenes	Personnel 1	Personnel 2	Personnel 3	Personnel 4	Personnel 5	Background Noise
Number of training samples	1051	1003	1089	1123	0	1023
Number of test samples	134	162	151	167	133	184

TABLE 3. The sample grouping of the frequency spectrum feature network.

Scenes	Personnel 1	Personnel 2	Personnel 3	Personnel 4	Personnel 5	Background Noise
Number of training samples	1875	1912	1892	1895	0	1654
Number of test samples	194	198	206	187	216	224

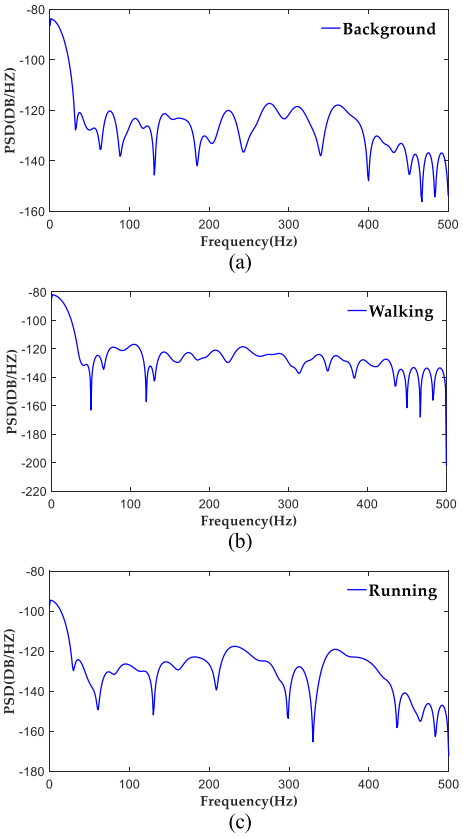


FIGURE 14. (a) Frequency domain diagram of the background signal. (b) Frequency domain diagram of the walking signal. (c) Frequency domain diagram of the running signal.

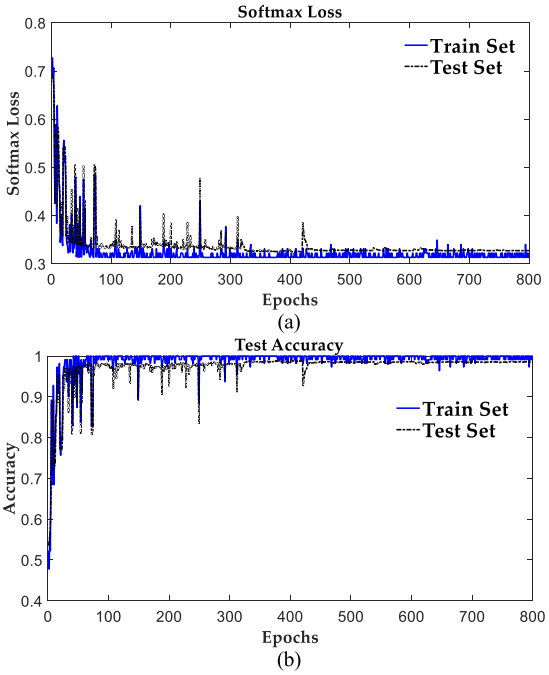


FIGURE 15. Time domain feature network training results. (a) Loss function graph of training set (blue line) and test set (black dot drawing line). (b) Test accuracy graph of training set (blue line) and test set (black dot drawing line).

the loss function value of the test set converge. The accuracy of the test set of 1800 training iterations is 98.65%, and the loss function value is 0.3268.

A large number of samples are tested on the two trained LSTMs. The test accuracy P_1 of the time domain feature network is 98.06%, and the test accuracy P_2 of the frequency spectrum feature network is 98.41%. According to the mathematical derivation in Section III, the theoretical accuracy of the PRNN is not less than 96.47%.

D. COMPARATIVE EXPERIMENTS

In this paper, two groups of comparative experiments were conducted. One group is the proposed PRNN target recognition experiment, and the other group is the single LSTM target recognition experiment. Both groups of experimental methods use LSTM for target recognition, but the difference between the network structure and the data set processing method determines the final different recognition effects.

1) LSTM EXPERIMENTS

Before putting forward the PRNN, we have done a lot of experiments, trying to train an LSTM. The training samples of an LSTM contain three types of signals: background, walking and running. We found that it was difficult to train an LSTM

TABLE 4. The sample grouping of a LSTM with time series data as training data.

Scenes	Personnel 1	Personnel 2	Personnel 3	Personnel 4	Personnel 5	Background Noise
Number of training samples	1041	1112	1065	1124	0	954
Number of test samples	642	687	654	663	601	732

TABLE 5. The sample grouping of a LSTM with frequency spectrum data as training data.

Scenes	Personnel 1	Personnel 2	Personnel 3	Personnel 4	Personnel 5	Background Noise
Number of training samples	2222	2322	2367	2297	0	2024
Number of test samples	478	502	481	472	508	512

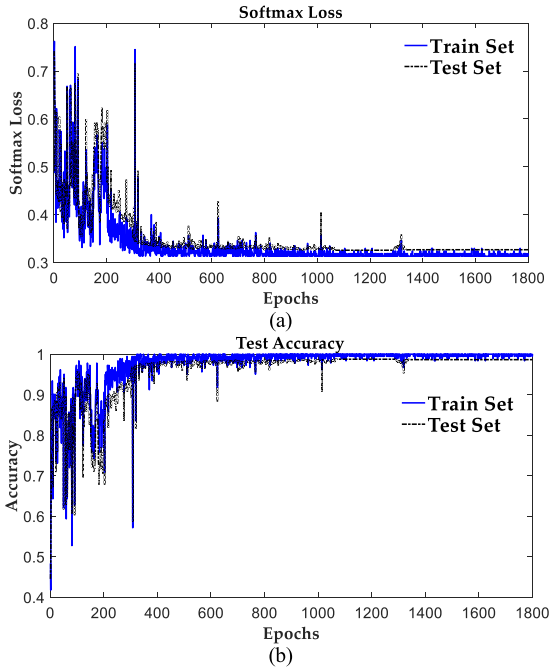


FIGURE 16. Frequency spectrum feature network training results. (a) Loss function graph of training set (blue line) and test set (black dot drawing line). (b) Test accuracy graph of training set (blue line) and test set (black dot drawing line).

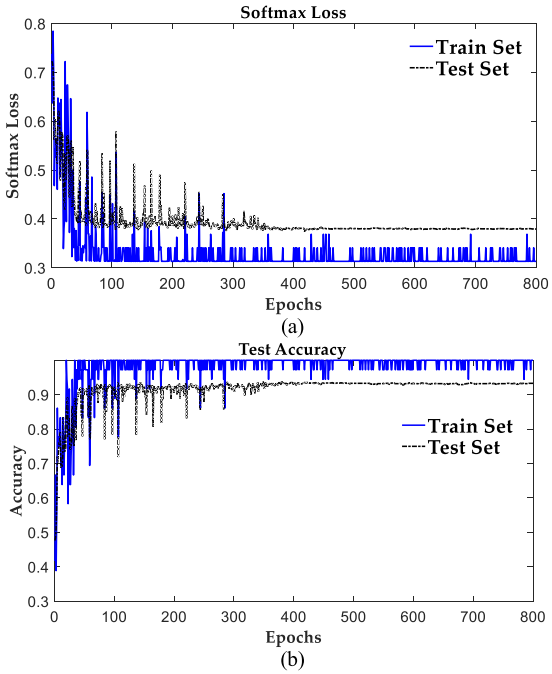


FIGURE 17. LSTM training results with time series data as training data. (a) Loss function graph of training set (blue line) and test set (black dot drawing line). (b) Test accuracy graph of training set (blue line) and test set (black dot drawing line).

with high recognition rate whether the training samples were time series data or frequency spectrum data. The LSTM training results of time series data and frequency spectrum data are given below. The sample grouping of an LSTM with time series data as training data is shown in Table 4. The human activity data in Table 4 were collected in walking and running scenes.

From the data in Table 4, the total number of training samples of the LSTM (time series data) is 5296, and the number of test samples is 3979. The loss function graph and the test accuracy graph of the LSTM training are shown in Fig. 17. It can be seen from Fig. 17 that when the LSTM training is iterated 500 times, both the accuracy and the loss function value of the test set converge. The accuracy of the test set of 800 training iterations is 93.3%, and the loss function value is 0.3794.

The sample grouping of an LSTM with frequency spectrum data as training data is shown in Table 5. The human activity data in Table 5 were collected in walking and running scenes.

From the data in Table 5, the total number of training samples of the LSTM (frequency spectrum data) is 11232, and the number of test samples is 2953. The loss function graph and the test accuracy graph of the LSTM training are shown in Fig. 18. It can be seen from Fig. 18 that when the LSTM training is iterated 100 times, both the accuracy and the loss function value of the test set converge. The accuracy of the test set of 800 training iterations is 85.36%, and the loss function value is 0.4597.

Two trained LSTMs are tested with large samples. The test accuracy of LSTM (time series data) is 91.68%, and that of LSTM (frequency spectrum data) is 73.78%.

TABLE 6. The test sample sources.

Test samples	Personnel 1	Personnel 2	Personnel 3	Personnel 4	Personnel 5	Background Noise
Test sample 1	274	261	255	268	265	273
Test sample 2	325	354	342	336	346	365
Test sample 3	521	561	544	552	574	586

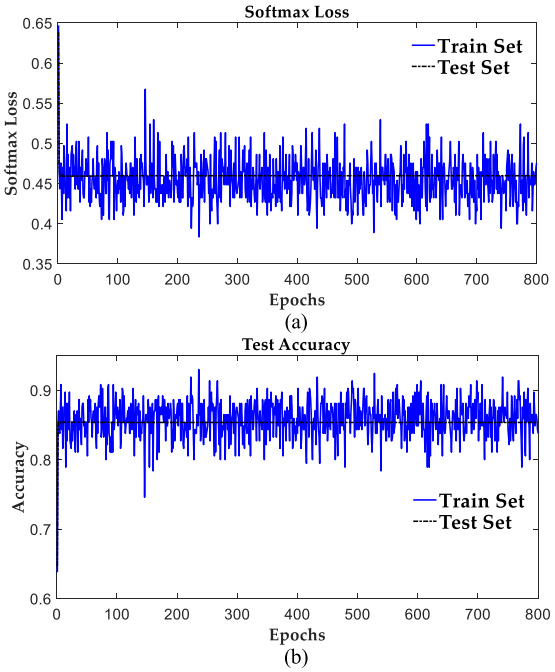


FIGURE 18. LSTM training results with frequency spectrum data as training data. (a) Loss function graph of training set (blue line) and test set (black dot drawing line). (b) Test accuracy graph of training set (blue line) and test set (black dot drawing line).

TABLE 7. False alarm rate.

Test samples	PRNN	LSTM (time series data)	LSTM (frequency spectrum data)
Test sample 1	1.02%	5.83%	19.23%
Test sample 2	0.93%	5.17%	17.16%
Test sample 3	0.82%	5.12%	16.87%

2) COMPARISON OF TEST RESULTS

In order to compare the recognition accuracy of the PRNN and the two LSTMs, we use the same test sample for detection. The test sample sources are shown in Table 6. The human activity data in Table 6 were collected in walking and running scenes.

The PRNN and the two LSTMs are tested with three test samples in Table 6. The false alarm rate and omissive report rate are shown in Table 7 and Table 8, and test accuracy results are shown in Fig. 19.

As shown in Table 7 and Table 8, the false alarm rate and the omissive report rate of PRNN are very low. It can be seen from Fig. 19 that the PRNN has the highest accuracy in all three tests. The average test accuracy of the PRNN

TABLE 8. Omissive report rate.

Test samples	PRNN	LSTM (time series data)	LSTM (frequency spectrum data)
Test sample 1	0.86%	4.95%	17.16%
Test sample 2	0.84%	4.71%	15.57%
Test sample 3	0.62%	4.63%	15.51%

FIGURE 19. The tests results of the PRNN and the LSTMs.

obtained by experiments is 98.3%, which obviously satisfies Eq. (17). Based on the experimental results, the LSTM is not suitable for target recognition in UGS systems. On the contrary, the proposed PRNN proves that its high recognition rate will have an absolute advantage in the application of UGS systems.

V. CONCLUSION

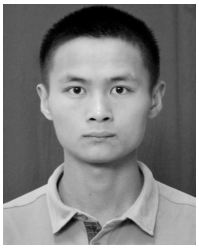
Accurate recognition of human target is especially important in UGS systems. The PRNN proposed in this paper has a very high accuracy of target recognition, low false alarm rate, and low omissive report rate. The PRNN consists of two LSTMs with different training samples. The training samples of the time domain feature network are time series data, and the training samples of the frequency spectrum feature network are frequency spectrum data. Meanwhile, the test accuracy of each network is above 98%. According to Eq. (17), the theoretical test accuracy of the PRNN is above 96.47%. This paper also illustrates the shortcoming of LSTM in the target recognition of UGS systems. The innovation of PRNN is to connect one LSTM in parallel with another LSTM and recognize human target effectively through data fusion. The PRNN can be applied not only to UGS systems but also to other pattern recognition systems.

REFERENCES

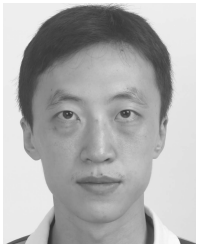
[1] G. Prado and R. A. Johnson, "Changing requirements and solutions for unattended ground sensors," *Proc. SPIE*, vol. 6736, Oct. 2007, Art. no. 67360X.

[2] P. Karthik. B. S. Kumar, and M. S. Khan, "Trespasser detection and categorization using unattended ground sensors (UGS)," in *Proc. Int. J. Sci. Eng. Technol. Res.*, Nov. 2014, pp. 7877–7886.

[3] X. Jin, S. Sarkar, A. Ray, S. Gupta, and T. Damarla, "Target detection and classification using seismic and PIR sensors," *IEEE sensors J.*, vol. 12, no. 6, pp. 1709–1718, Jun. 2012.
[4] R. Berešík, J. Puttera, and F. Nebus, "Seismic sensor system for security applications based on MEMS accelerometer," in *Proc. Int. Conf. Appl. Electron.*, Sep. 2014, pp. 31–36.
[5] F. Guo, J. Huang, X. Zhang, Y. Cheng, H. Liu, and B. Li, "A two-stage detection method for moving targets in the wild based on microphone array," *IEEE Sensors J.*, vol. 15, no. 10, pp. 5795–5803, Oct. 2015.
[6] R. Levy, J. Moras, and B. Pannetier, "Vibrating beam MEMS seismometer for footstep and vehicle detection," *IEEE Sensors J.*, vol. 17, no. 22, pp. 7306–7310, Nov. 2017.
[7] Y. H. Kim and K. S Chung, "An efficient intruder detection using the seismic sensor," *J. Korea Inst. Military Sci. Technol.*, vol. 14, no. 6, pp. 1129–1137, 2011.
[8] U. Antao, J. Choma, A. Dibazar, and T. Berger, "40nW subthreshold event detector chip for seismic sensors," in *Proc. IEEE Int. Symp. Technol. Homeland Secur. (HST)*, Apr. 2015, pp. 1–6.
[9] X. Jin, S. Gupta, A. Ray, and T. Damarla, "Symbolic dynamic filtering of seismic sensors for target detection and classification," in *Proc. Amer. Control Conf.*, Jun./Jul. 2011, pp. 5151–5156.
[10] S. G. Iyengar, P. K. Varshney, and T. Damarla, "On the detection of footsteps based on acoustic and seismic sensing," in *Proc. Conf. Rec. 40st Asilomar Conf. Signals, Syst. Comput.*, Nov. 2007, pp. 2248–2252.
[11] K. M. Houston, and D. P. McGaffigan, "Spectrum analysis techniques for personnel detection using seismic sensors," *Proc. SPIE*, vol. 5090, pp. 162–174, Sep. 2003.
[12] M. S. Richman, D. S. Deadrick, R. J. Nation, and S. Whitney, "Personnel tracking using seismic sensors," *Proc. SPIE*, vol. 4393, pp. 14–22, Sep. 2001.
[13] B. W. Larsen, H. Chung, A. Dominguez, J. Sciacca, N. Kovvali, A. Papandreou-Suppappola, and D. R. Allee, "Robust discrimination of human footsteps using seismic signals," *Proc. SPIE*, vol. 8046, May 2011, Art. no. 80460D.
[14] B. W. Larsen, H. Chung, A. Dominguez, J. Sciacca, N. Kovvali, A. Papandreou-Suppappola, and D. R. Allee, "Applying matching pursuit decomposition time-frequency processing to UGS footstep classification," *Proc. SPIE*, vol. 8711, Jun. 2013, Art. no. 871104.
[15] S. Sarkar, T. Damarla, and A. Ray, "Real-time activity recognition from seismic signature via multi-scale symbolic time series analysis (MSTSA)," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2015, pp. 5818–5823.
[16] R. S. Shah, R. S. Shah, A. L. Kumar, S. Abhijith, and S. D. Badiger, "Footstep detection using Laplacian distribution," in *Proc. Int. Conf. Energy, Commun., Data Anal. Soft Comput. (ICECDS)*, Aug. 2017, pp. 821–824.
[17] L. Peck, J. Lacombe, T. S. Anderson, J. Gagnon, D. Fisk, C. Williams, and K. Wesson, "Seismic detection of personnel: Field trials and signatures database," in *Proc. Human, Light Vehicle Tunnel Detection Workshop*, University, MS, USA: NCPA, 2009, pp. 1–8.
[18] G. P. Mazarakis, and J. N. Avaritsiotis, "A prototype sensor node for footstep detection," in *Proc. 2nd Eur. Workshop Wireless Sensor Netw.*, Feb. 2005, pp. 415–418.
[19] S. Bahrampour, A. Ray, S. Sarkar, T. Damarla, N. M. Nasrabadi, "Performance comparison of feature extraction algorithms for target detection and classification," *Pattern Recognit. Lett.*, vol. 34, no. 16, pp. 2126–2134, Dec. 2013.
[20] G. P. Succi, D. Clapp, R. Gampert, and G. Prado, "Footstep detection and tracking," *proc. SPIE*, vol. 4393, pp. 22–30, Sep. 2001.
[21] Z. Liang, J. Wei, J. Zhao, H. Liu, B. Li, J. Shei, and C. Zeng, "The statistical meaning of kurtosis and its new application to identification of persons based on seismic signals," *Sensors*, vol. 8, no. 8, pp. 5106–5119, Aug. 2008.
[22] Y. Tian, H. Qi, and X. Wang, "Target detection and classification using seismic signal processing in unattended ground sensor systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 4, May 2002, p. 4172.
[23] J. Huang, Q. Zhou, X. Zhang, E. Song, B. Li, and X. Yuan, "Seismic target classification using a wavelet packet manifold in unattended ground sensors systems," *Sensors*, vol. 13, no. 7, pp. 8534–8550, Jul. 2013.
[24] S. Schumer, "Analysis of human footsteps utilizing multi-axial seismic fusion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 697–700.
[25] N. Virani, S. Marcks, S. Sarkar, K. Mukherjee, A. Ray, and S. Phoha, "Dynamic data driven sensor array fusion for target detection and classification," *Procedia Comput. Sci.*, vol. 18, pp. 2046–2055, Jan. 2013.
[26] X. Jin, S. Gupta, A. Ray, and T. Damarla, "Multimodal sensor fusion for personnel detection," in *Proc. 14th Int. Conf. Inf. Fusion*, Jul. 2011, pp. 1–8.
[27] M. A. De Jesus, M. Teixeira, L. Vicente, and Y. Rodriguez, "Nonuniform discrete short-time Fourier transform a goertzel filter bank versus a FIR filtering approach," in *Proc. 49th IEEE Int. Midwest Symp. Circuits Syst.*, vol. 2, Aug. 2006, pp. 188–192.
[28] L. Marple, "Computing the discrete-time 'analytic' signal via FFT," *IEEE Trans. signal Process.*, vol. 47, no. 9, pp. 2600–2603, Sep. 1999.
[29] P.K. Rahi, and R. Mehra, "Analysis of power spectrum estimation using welch method for various window techniques," *Int. J. Emerg. Technol. Eng.*, vol. 2, no. 6, pp. 106–109, 2014.
[30] F. Zhang, XF. Shi, XZ. Zhang, "Study and simulation of welch power spectrum estimation method," in *Proc. J. Xi'an Technol. Univ.*, vol. 4, Sep. 2009, p. 013.
[31] A. Graves, "Generating sequences with recurrent neural networks," Aug. 2013, *arXiv:1308.0850*. [Online]. Available: https://arxiv.org/abs/1308.0850



TONGHUA XU received the B.S. degree in automation from the Department of Engineering, Ocean University of China, Qingdao, China, in 2017. He is currently pursuing the M.S. degree in control science and engineering with the College of Intelligence Science and Technology, National University of Defense Technology, Changsha, China. His research interests include machine learning, pattern recognition, reinforcement learning, and multiagent systems.



NAN WANG received the B.S. degree in automatic control from the College of Mechatronics and Automation, National University of Defense Technology (NUDT), Changsha, China, in 2003, and the M.S. and Ph.D. degrees in control science and engineering from the College of Mechatronics and Automation, NUDT, in 2005 and 2012, respectively, where he is currently an Associate Professor with the Institute of Unmanned Systems, College of Intelligence Science and Technology. He has published more than ten articles in international journals and conferences and has coauthored one book. His research interests include mission planning, autonomous control, and collaborative architecture of unmanned systems. He has served as an Academic Editor for the *Tactical Missile Technology*.

XIN XU (M'07–SM'12) received the B.S. degree in electrical engineering from the Department of Automatic Control, National University of Defense Technology (NUDT), Changsha, China, in 1996, and the Ph.D. degree in control science and engineering from the College of Mechatronics and Automation, NUDT, in 2002. He has been a Visiting Professor with The Hong Kong Polytechnic University, University of Alberta, University of Guelph, and the University of Strathclyde, U.K. He is currently a Full Professor with the Institute of Unmanned Systems, College of Intelligence Science and Technology, NUDT. He has coauthored more than 160 articles in international journals and conferences and four books. His research interests include intelligent control, reinforcement learning, approximate dynamic programming, machine learning, robotics, and autonomous vehicles. He is a member of the IEEE CIS Technical Committee on Approximate Dynamic Programming and Reinforcement Learning and the IEEE RAS Technical Committee on Robot Learning. He received the National Science Fund for Outstanding Youth in China and the second-class National Natural Science Award of China. He has served as an Associate Editor or a Guest Editor for the *Information Sciences*, the *International Journal of Robotics and Robotics*, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, INTELLIGENT AUTOMATION AND SOFT COMPUTING, the *International Journal of Adaptive Control and Signal Processing*, and the *Acta Automatica Sinica*.