IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 August 2022, accepted 28 September 2022, date of publication 4 October 2022, date of current version 12 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3211976

RESEARCH ARTICLE

An Intelligent Signal Processing Method for Motional Vital Signs Detection System Based on Deep Learning

SIYUN LIU, QINGJIE QI, HUIFENG CHENG, JINGWEN ZHANG, WENHAO XIAN, TIANFANG MA, YUE WANG, YINGJIE LIU, DAN LI, AND JIAMEI CHAI

Emergency Research Institute, Chinese Institute of Coal Science (CICS), Beijing 100013, China

Corresponding author: Qingjie Qi (qiqingjie_yz@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 52174188; and in part by the Project of China Coal Industry Group Company Ltd. under Grant 2022-TD-MS001, Grant 2022-2-MS004, Grant 2019-2-ZD004, and Grant 2019-2-ZD003.

ABSTRACT Detection of vital signs for motional human targets in complex environment has always been a major challenge in the field of remote detection, remote healthcare and emergency rescue, because polytropic and multimodal interferences make intelligent signal processing more difficult. In this paper, a systematic intelligent signal processing scheme which contains signal preprocessing, vital signs identification, motion trajectory estimation and respiratory signal and heartbeat signal extraction is established. Based on CNN (Convolutional Neural Networks) model, accurate identification of motional vital signs getting rid of the interference of harmonics and distortion can be realized. Then, the misidentified outliers are eliminated with K-means clustering algorithm. Next, the motion trajectory of human targets can be estimated with Kalman filtering algorithm. Finally, the SVD-EEMD algorithm is proposed for respiratory signal and heartbeat signal extraction of dynamic human targets. The introduction of deep learning algorithms makes the proposed method have good performance of high accuracy, good robustness, strong adaptability and high efficiency, which can be observed in actual detection tasks contrast experiments.

INDEX TERMS Intelligent signal processing, vital signs detection, deep learning, convolutional neural networks (CNN), Kalman filtering, SVD-EEMD, ultra-wideband (UWB) radar.

I. INTRODUCTION

Radar remote detection is an emerging technology, which is not only applied in military, but also in the fields of national economy and scientific research [1], [2]. With the development of electronic equipment and bioscience, the method of life detection with radar device breaks through the traditional detection technology, it emits electromagnetic waves and obtains information without directly contacting the target and can even detect across obstacles. The electromagnetic waves will modulate with human life characteristics such as breathing, heartbeat, gait movement and other motion information, and the modulated signal will be received and analyzed, so as to remote detect, locate and identify life targets. Among all kinds of radar detection devices, the performance of UWB radar is particularly outstanding [3]. It has the advantages of high range resolution, strong penetration ability, low power consumption and strong anti-jamming ability.

After years of research, radar remote detection technology has made great progress, but there are still some disadvantages in some aspects [4]. In particular, the disaster site environment is always very complex, at the same time, the trapped people may be injured, in which condition the vital signs are weaker than normal situation, so it also puts forward high requirements on the hardware and software of the system. In the aspect of UWB radar signal processing and life target recognition technology, people urgently need an algorithm with sensitivity, accuracy, robustness, and real-time performance to realize the accurate recognition and extraction of weak, disturbed and even distorted vital sign signals.

The associate editor coordinating the review of this manuscript and approving it for publication was Seifedine Kadry.

Among the preprocessing technologies of radar signal, mean subtraction (MS) [5] and linear trend subtraction (LTS) [6] are effective time-domain processing methods, which can effectively remove the static clutters and the linear trend term. Mean filtering [7] and median filtering [8] can also reduce the impact of static environment such as walls and obstacles on vital signs signals by filtering out low-frequency components. The auto gain control (AGC) [9] method realizes the enhancement and amplification of weak vital sign signals through the setting of adaptive compensation parameters. The above traditional processing methods can remove the interference of specific components in the signals, consequently, they have been widely used in the signal preprocessing steps.

Vital signs remote detection tasks are often carried out under the circumstances of very low signal-to-noise ratio. With the high complexity of the detection environment, the difficulty of effective extraction of vital information is also increasing [10]. In recent 20 years, the widely used methods for vital sign signal extraction include Fourier transform [11], wavelet transform (WT) [12] and empirical mode decomposition (EMD) [13]. In 2011, Tariq et al. adopted wavelet transform method to accurately extract respiratory signals [14]. He developed a method based on empirical wavelet transform to extract respiratory and heartbeat signals separately [15]. The University of Pennsylvania has effectively extracted the micro-Doppler features of human respiratory, heartbeat and gait through empirical mode decomposition (EMD) [16], [17]. Since then, the vital sign extraction methods evolved from EMD [18], [19], [20] has become an important branch in this field. Yang adopted ensemble empirical mode decomposition (EEMD) algorithm and effectively solved the mode aliasing problem of EMD [21], [22]. According to the number of targets in the detection scene, Yan used variational mode decomposition method to suppress mode aliasing and successfully extracted multi-target vital signs [23].

With the rise of artificial intelligence algorithm and big data analysis technology in recent years [24], remote sensing and detection technology has a new breakthrough. Currently, the widely applied deep learning models include convolutional neural network (CNN) [25], recurrent neural network (RNN), long short term memory (LSTM) [26], deep belief network (DBN) [27] and so on.

Deep learning methods greatly improve the performance of signal processing algorithm, making it possible to detect vital signs efficiently and accurately in complex emergency rescue scenes such as crossing obstacles [28], moving targets [29] and multiple life targets [30]. Based on the singular value decomposition (SVD) method, Liu decomposed the UWB matrix into eigen vectors in time-space dimension to realize the through-wall, multi-targets personnel positioning and vital signs extraction in complex scenes [31]. Bao proposed a novel people counting algorithm exploiting convolutional neural network (CNN) using IR-UWB radar [32]. Lv adopted an adaptive Kalman filter to fuse multisensory information, which solved the respiration detection problem caused by the human targets' body orientation [33]. Kwon proposed a hybrid model architecture that combines CNNs and LSTM for real-time Apnea-Hypopnea events detection based on IR-UWB radar [34].

However, above multitudinous studies seldom provide a good solution for the remote detection method of vital sign information of motional people in complex environment. A few applicable methods, such as CLEAN algorithm [29], CFAR algorithm [35], [36] and SVM [37], [38], are sometimes limited by their accuracy, effectiveness, stability, and real-time performance. In the tasks of moving target detection, vital signs signals are not only disturbed by a lot of noises and clutters, but also affected by human movement. The changing distance and dynamic characteristics bring great challenges to the accurate identification of vital signs information. In addition, the diverse movements bring a large number of harmonics and serious distortions to the signals [39], in this case, an intelligent signal processing algorithm with high resolution, accuracy, efficiency and robustness is urgently needed.

In order to solve the above problems, some scholars have developed a distributed radar system, which integrates multiple UWB signals to achieve high-precision detection function. In contrast, this paper proposes an algorithm framework based on single IR-UWB radar system to address these problems. The convolution neural network (CNN) model is constructed to realize the accurate detection of targets, and the motion information is extracted through clustering algorithm and Kalman filtering. Finally, the SVD-EEMD algorithm is proposed to realize the accurate extraction of vital signs under interference environment.

II. IR-UWB RADAR SIGNAL MODEL

UWB technology is a special technology for communication and detection by transmitting and receiving pulse signals ranging from nanoseconds to picoseconds. According to the definition formula of bandwidth:

$$\mu = \frac{\Delta f}{f_0} = \frac{2(f_H - f_L)}{f_H + f_L} \tag{1}$$

where, f_H and f_L are the maximum and minimum frequencies of the signal at -10 dB of the peak amplitude, respectively. When $\mu \geq 20\%$, the impulse signal is regarded as ultra-wideband (UWB) signal. IR-UWB radar achieves detection by transmitting and receiving continuous pulse sources, with good range resolution and resistance to clutter and multipath interference. The essence of radar wave is an electromagnetic wave, when propagating electromagnetic waves touch the human body, most of the electromagnetic waves are reflected by the human body and then received by the receiving antenna, and this part occupies the main energy of the radar echo signals. A small part of the electromagnetic waves will be absorbed by the environment and attenuated, and still a small amounts of radar waves will be received by the antenna after secondary reflection.

The signal models of IR-UWB radar when detecting static person and dynamic person have been shown in Figure 1.

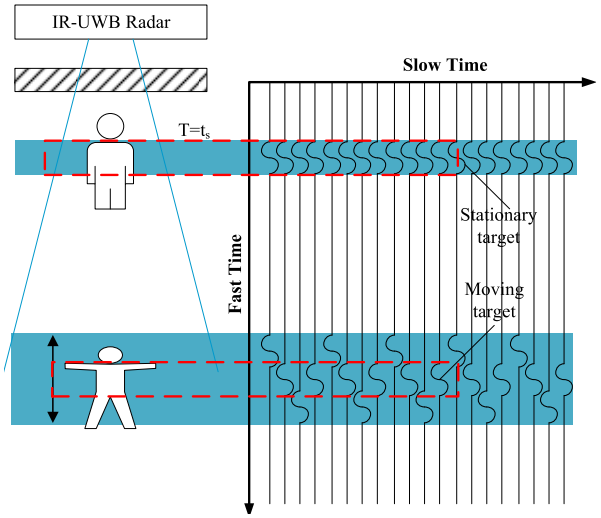


FIGURE 1. IR-UWB radar detection signal model.

When detecting the area, IR-UWB radar device sends out multiple continuous signals at a fixed pulse time interval and receives all the radar data to form a radar echo matrix which is characterized with two dimensions of fast time and slow time. The fast time characteristic reflects the change of radar signal with detection distance at a fixed time point. The slow time characteristic reflects the change of radar signal with time at a fixed distance point.

According to the characteristics of different target objects, the radar signal received by IR-UWB radar device can be expressed as follows:

$$R(t,\tau) = \alpha_s p(\tau - \tau_s) + \sum_i \alpha_d p(\tau - \tau_d(t)) \qquad (2)$$

where τ denotes the sampling time of the UWB radar along the propagation direction, namely, the fast time. And t denotes the accumulated time of multiple UWB pulses, namely, the slow time. The former term of equation (2) represents the static response generated in the environment, and the latter terms represent the dynamic signal response generated by the target persons, which includes the modulation effects of the person's movement, respiratory, heartbeat, etc. The dynamic terms vary not only regard with fast time, but also with slow time. The macro and micro movements of people cause the UWB signal showing dynamic characteristics. The above principled analysis lays a foundation for IR-UWB radar to detect vital signs in complex environment.

III. SYSTEM SETUP AND ALGORITHM

A. EXPERIMENTAL CONSTRUCTION

Based on the NVA6100 UWB radar system, this paper studies the UWB radar detection technology. The experimental system setting is shown in Figure 2. The IR-UWB radar system is equipped with two transceiver antennas based on Vivaldi antenna structure and transmits first-order Gaussian pulse signals aligned with the detection area. In order to ensure the efficient operation of deep learning algorithm, the processor should be equipped with GPU or FPGA to realize algorithm acceleration. The key parameter settings of NVA6100 UWB radar system are shown in Table 1. In particular, the fast time sampling frequency (F_s) is set to 5 GHz, the total sampling points number is set to 768, and the slow time radar pulse transmission frequency (f_s) is set to 8 Hz.

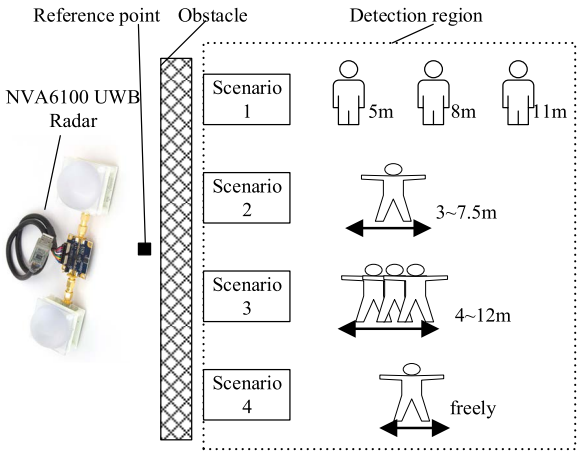


FIGURE 2. Experimental construction of vital signs detection system.

TABLE 1. Parameters configuration of system.

Parameter	Value
Transmitted waveform	First order Gaussian pulse
Output power	-14 dBm
Vivaldi antenna gain	6 dBi
Pulse width	2 ns
Offset distance from reference	2.5 m
Sampling frequency (F_s)	5 GHz
Total sampling points number	768
Radar pulse frequency (f_s)	8 Hz
Equivalent distance resolution	30 mm

Under emergency rescue conditions, UWB radar signal often needs to penetrate the ruins to find survivors. In order to simulate the complex scenes, indoor and outdoor experimental sites are built up respectively. The obstacle set in the indoor experimental site is a 4.2 cm thick wooden board with a density of 750 kg/m^3, which can be moved along the chute. The obstacle set in the outdoor site is a concrete wall with a thickness of 28 cm and a density of 2500 kg/m^3, the UWB radar device was placed on one side of the wall, and the person targets were arranged in the outdoor environment on the other side of the wall. During the experiments, the positions or motion states of the targets were recorded according to the scale marks and timer.

In addition, as shown in Figure 2, this project has also set up four experimental scenarios. In scenario 1, the three detected persons were standing still at positions of 5 m, 8 m, and 11 m from the reference point, respectively. In scenario 2,

one detected person was performing a uniform reciprocating movement from 3 m to 7.5 m from the reference point. In scenario 3, three detected persons were forming in a row along the direction of UWB signal and performing a constant speed reciprocating movement from 4 m to 12 m simultaneously. In scenario 4, one detected person was walking freely in the detection area along the detection direction of IR-UWB radar. Experimental scenarios 1 and 2 were tested in the indoor experimental environment shown in Figure 3 (b), and experimental scenarios 3 and 4 were tested in the outdoor experimental environment shown in Figure 3 (c). The relevant personnel in the experiments were randomly selected from 5 volunteers, and the information of the 5 volunteers is shown in Table 2.

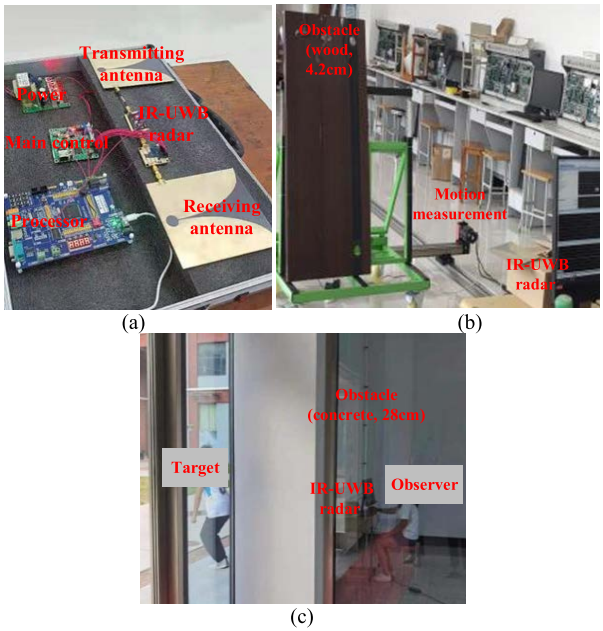


FIGURE 3. Physical maps of experimental sites setting: (a) IR-UWB radar detection device, (b) Indoor experimental site setup, (c) Outdoor experimental site setup.

TABLE 2. Brief physical information of the five volunteers.

Personnel	Gender	Age	Height (m)	Weight (kg)
A	Male	26	1.76	72
B	Male	30	1.83	96
C	Male	38	1.74	63
D	Male	42	1.71	86
E	Female	23	1.60	48

B. ALGORITHM ARCHITECTURE

Figure 4 shows the algorithm architecture proposed in this paper. In order to realize the vital signs intelligent detection function of multitargets and dynamic targets in complex emergency rescue scenes, the algorithm includes the following modules: signal preprocessing module, CNN intelligent recognition module, motion estimation module and vital signs extraction module.

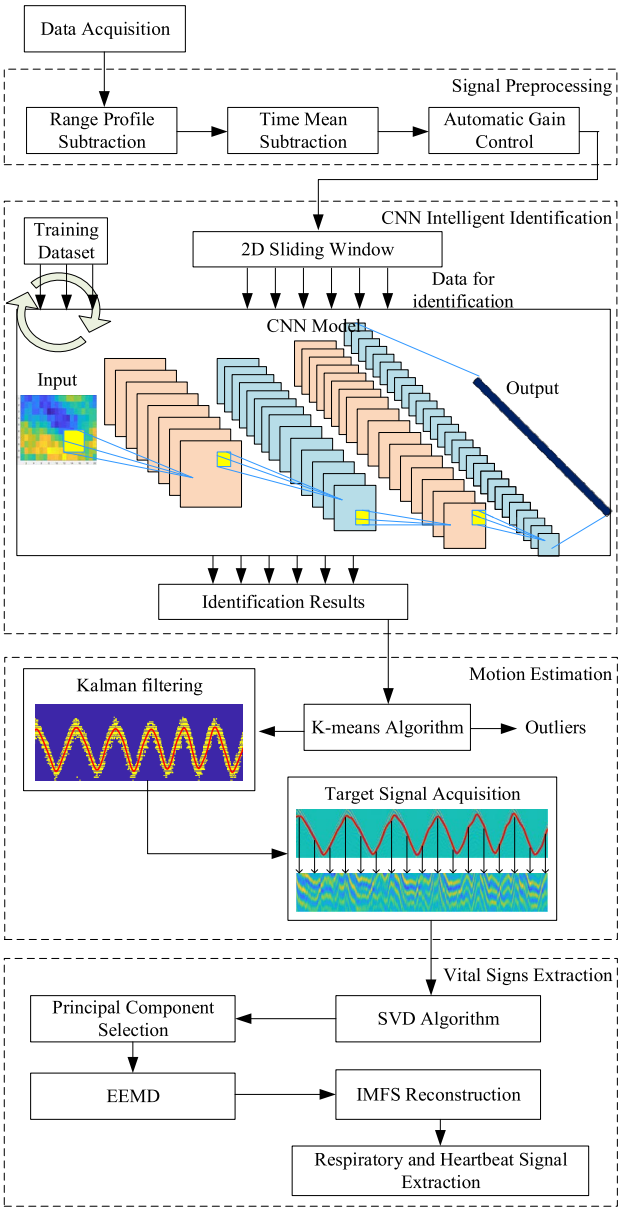


FIGURE 4. Specific flowchart of proposed algorithm.

In the signal preprocessing module, the range profile subtraction (RPS) and time mean subtraction (TMS) methods are adopted to remove a large number of clutters and noises, and the automatic gain control (AGC) method is adopted to enhance the signal-to-noise ratio. In the CNN intelligent identification module, an intelligent identification model based on convolutional neural network is built. After being trained by numerous samples, the CNN model has been endowed with the ability to identify the life information, environmental information, and interference information. When the system performing the detection task, the two-dimensional sliding window will traverse the received UWB radar signals and input the data matrices into CNN model to obtain the identification results of life information in the detection time-space range. In the motion estimation module, K-means clustering

algorithm is used to classify the identified data information firstly, and the misidentified outliers are eliminated at the same time. Then the Kalman filtering algorithm is adopted to estimate the trajectories of the dynamic targets, and then the shaped life signals are obtained from a large number of data points. Finally, the shaped signals are input into the vital signs extraction module, the principal component characteristic vectors can be extracted based on singular value decomposition (SVD), and then several IMFs components are obtained by EEMD algorithm, the respiratory and heartbeat signals are extracted according to energy threshold.

IV. SIGNAL PREPROCESSING

In the actual detection tasks, the original UWB echo signal contains not only vital signs information, but also many other interference components. Therefore, the UWB echo matrix in the real measurement environment is modeled as:

$$R[m,n]=r[m,n]+c[n]+w[m,n]+d[m]+l[m,n]+t[m,n] \tag{3}$$

where $R[m,n]$ indicates the original echo matrix obtained in the detection environment, $r[m,n]$ indicates the vital signs signal, $c[n]$ indicates static background clutter, $w[m,n]$ indicates additive white noise, $d[m]$ indicates the unstable fast time DC component, $l[m,n]$ indicates the linear trend along slow time axis due to the unstable amplitude of the radar system, $t[m,n]$ indicates the harmonic and signal distortion caused by dynamic target. Some clutters and noises interferences could be removed by preprocessing method.

Firstly, range profile subtraction (RPS) is adopted to remove static background clutter $c[n]$, and the method is as follows:

$$\begin{aligned} R'[m,1] &= R[m,1] \\ R'[m,n] &= R[m,n]-R[m,n-1] \\ &\quad \times (m=1,\ldots,M, n=2,\ldots,N) \end{aligned} \tag{4}$$

Then, time mean subtraction (TMS) is adopted to filter the DC component $d[m]$. The method is as follows:

$$\begin{aligned} R''[m,n] = R'[m,n] - \frac{1}{N}\sum_{i=1}^{N} R'[m,i] \\ \times (m=1,\ldots,M, n=1,\ldots,N) \end{aligned} \tag{5}$$

In order to improve the signal-to-noise ratio, automatic gain control (AGC) is adopted to enhance weak vital signs in the fast time direction. The specific algorithm steps are as follows:

$$g(m,i) = \frac{2d_s+1}{\sqrt{\sum_{k=i-d}^{i+d} R''(m,i)^2}} \tag{6}$$

$$g_{mask}(m,i) = \begin{cases} g_{\max}, & g(m,i) > g_{\max} \\ g(m,i), & g(m,i) \le g_{\max} \end{cases} \tag{7}$$

$$X(m,i) = g_{mask}(m,i)R''(m,i) \tag{8}$$

where d_s indicates the set sliding window length and g_{max} indicates the maximum gain threshold.

V. CNN INTELLIGENT RECOGNITION

Convolutional neural network (CNN) is a common deep learning method in the field of image recognition. Based on CNN, an intelligent model which can automatically distinguish vital signs information, environmental information and interference information is designed in this paper. The establishment and application of CNN model include the following steps: data set construction, CNN model construction, model training and testing, and vital signs identification.

A. DATA SET CONSTRUCTION

The input data of CNN is often in the form of matrix. In order to realize the precise and refined perception, various experimental environments (indoor and outdoor, with and without obstacles, static and dynamic targets, single target and multi-targets, different detection distances, etc.) are settled to acquire multifarious UWB radar echo matrices. According to the position of the tested personnel, the positions of vital signs information and environmental information in UWB echo matrix can be deduced. In the data acquisition experiments, two typical groups of preprocessed UWB radar echo matrices are shown in Figure 5, in which Figure 5 (a) shows that the tested person is stationary at 4 m and Figure 5 (b) shows that the tested person is reciprocating from 3 to 12 m.

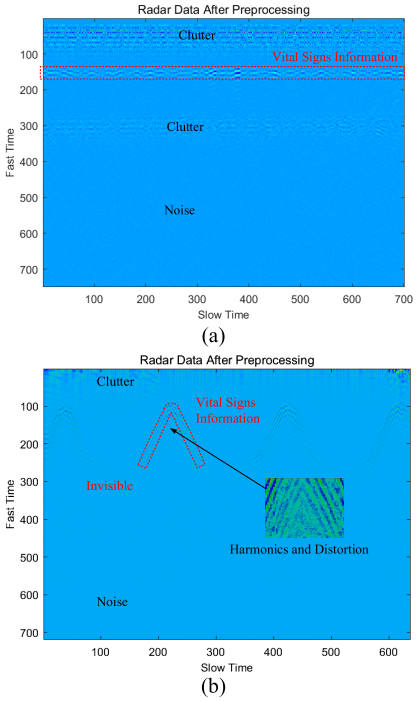


FIGURE 5. Preprocessed UWB radar echo matrices in data acquisition experiments: (a) Group 1: Tested person stationary at 4 m, (b) Group 2: Tested person reciprocating from 3 to 12 m.

When constructing the data set, in order to ensure that the CNN model can effectively identify vital signs on a small

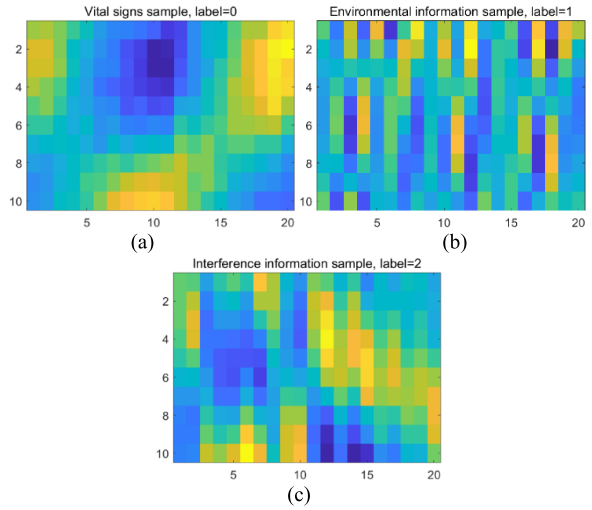


FIGURE 6. Input matrices of samples: (a) Vital signs sample, (b) Environmental information sample, (c) Interference information sample.

scale, the input data size $u \times v$ of the CNN model is determined according to the characteristics of human vital signs in the time-space dimension:

$$u = [\frac{2\alpha_1 F_s D_m}{c}], \quad v = [\alpha_2 D_n f_s] \tag{9}$$

where, c indicates the speed of light, sampling frequency (F_s) equals to 5 GHz, radar pulse frequency (f_s) equals to 8 Hz, D_m indicates the fast time characteristic resolution, and D_n indicates the slow time characteristic resolution, α_1 indicates the fast time scale coefficient, α_2 indicates the slow time scale coefficient. According to the dynamic characteristics of human vital signs, taken $D_m = 0.3$ m, $D_n = 2.5$ s, $\alpha_1 = \alpha_2 = 1$, then one can obtain that the size of input matrix is (1,10, 20).

As shown in Figure 5, despite the preprocessing, the vital signs in the UWB radar echo are still not obvious, and it is easy to be submerged in various clutters and noises. And with the increase of detection distance, the vital signs signal becomes weaker and even difficult to be observed directly. In particular, when detecting dynamic targets, human motion always produces harmonics, which cause the distortions of UWB radar echo signal behind it. These distortions often have the same phase with human vital signs, so it is difficult to distinguish them effectively by traditional signal processing methods and machine learning methods. Based on the above situation, after acquisition of the vital sign's information in various experimental environments, one can obtain abundant $u \times v$ input matrices and form the vital signs samples, the classification labels of vital signs samples are set to 0. Similarly, the $u \times v$ data matrices of noises, clutters, and various empty environments are obtained to form the environmental information samples, and their classification labels are set as 1. In particular, in order to ensure the effectiveness of the algorithm when detecting dynamic targets, the $u \times v$ matrices of harmonic and distortion data behind moving targets are obtained to form interference information samples, and their classification labels are set as 2. The input matrices of the above three types of samples are shown in Figure 6.

Deep learning usually requires a large amount of data as support. In order to prevent over fitting problems and ensure that the algorithm has a high detection success rate and stability in various complex environments, the following data augmentation technologies are randomly used to process the acquired samples: flip, translation, rotation, scale, crop, Gaussian noise, centralized brightness adjustment, etc. After the above augmentations, the irrelevant accidental features in the samples are greatly reduced, the numbers and types of samples are increased, and the generalization ability of CNN model is guaranteed.

B. CNN MODEL CONSTRUCTION

The CNN model architecture constructed in this paper has been shown in Figure 7. Its overall architecture is a five-layers neural network, including four convolution layers (''Conv1'', ''Conv2'', ''Conv3'' and ''Conv4'') and a full connection layer (''FC''). In addition, a residual layer is set to ensure the optimal mechanism of training effect. In CNN model, the key parameters of each convolution layer include: numbers of filters (N_C), kernel size ($F_H \times F_W$), stripe length (S_L) and padding (P_n). Assuming that the input feature map size of each convolution layer is (ch, H_{in}, W_{in}), its output feature map size is (N_C, H_{out}, W_{out}). In this paper, channel of input data is equal to 1, and other parameters can be calculated as follows:

$$H_{out} = \frac{H_{in} - F_H + 2P_n}{S_L} + 1, \quad W_{out} = \frac{W_{in} - F_W + 2P_n}{S_L} + 1 \tag{10}$$

The input of the whole CNN model is a three-dimensional array with a size of $1 \times 10 \times 20$. In order to ensure that the size of the output feature map remains unchanged after each convolution layer feature extraction, namely $H_{out} = H_{in}$, the key parameters of each convolution layer should be specially designed, and the parameters values have been shown in Table 3. After each convolution layer, the Batch normalization layer and a non-linear activation layer ''ReLU'' are set to standardize the weight parameters, so as to improve the performance of CNN model and solve the problem of gradient disappearance. Different with other three convolution layers, a Max-pooling layer is set at the output of the ''Conv4'' for down-sampling of the feature map, it has a 2×2 pool size each with stride 2. After the Max-pooling layer, a full connection layer (''FC'') with 3 units (outputs) are included to generate predicted value for each category. Finally, the ''softmax'' layer outputs the most possible prediction label as the final result according to the maximum of the three prediction values.

In particular, in order to solve the degradation problem in deep network, this paper designs the residual layer with reference to ''ResNet'' neural network model [40]. When the original data has been trained by two convolution layers (''Conv1'', ''Conv2''), most essential features have been

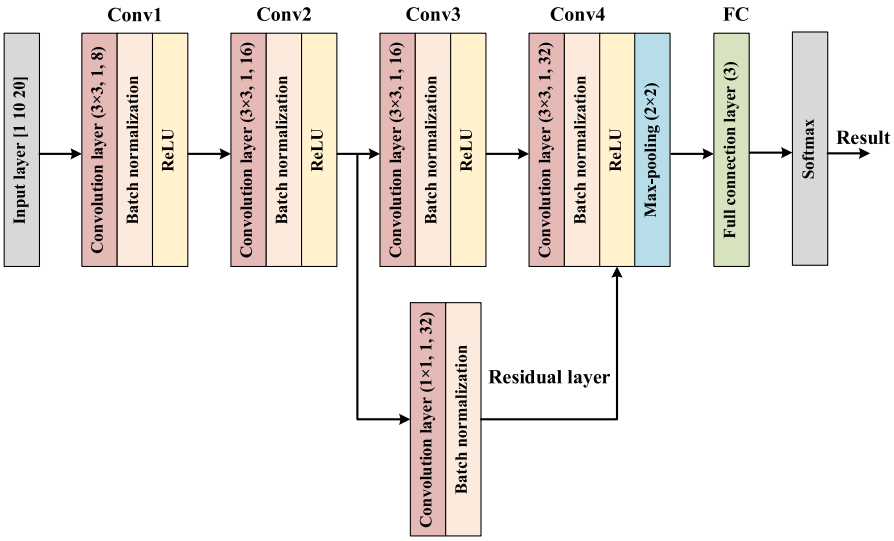


FIGURE 7. Network architectures for proposed CNN model.

TABLE 3. Key parameters values of the constructed CNN model.

Layer name	Input size (ch, H_{in}, W_{in})	Output size (N_C, H_{out}, W_{out})	Filters numbers (N_C)	Kernel size ($F_H \times F_W$)	Stride (S_L)	Padding (P_n)
Conv1	(1,10,20)	(8,10,20)	8	3 × 3	1	1
Conv2	(8,10,20)	(16,10,20)	16	3 × 3	1	1
Conv3	(16,10,20)	(16,10,20)	16	3 × 3	1	1
Conv4	(16,10,20)	(32,10,20)	32	3 × 3	1	1
Residual layer	(16,10,20)	(32,10,20)	32	1 × 1	1	0
Max-pooling	(32,10,20)	(32,5,10)	32	2 × 2	2	0
FC	32×5×10=1600	3	/	/	/	/

extracted. In order to ensure that there is no gradient dispersion / explosion in the deeper network layers behind and further reduce the recognition rate, a shortcut connection mechanism of residual layer is set up. According to Figure 8, the design principle is as follows:

$$F(x) = W^{[4]}\,\mathrm{ReLU}(W^{[3]}x + b^{[3]}) + b^{[4]} \tag{11}$$

$$R(x) = W^{[r]}x \tag{12}$$

$$\begin{aligned} H(x) &= F(x) + R(x) \\ &= W^{[4]}\,\mathrm{ReLU}(W^{[3]}x + b^{[3]}) + b^{[4]} + W^{[r]}x \end{aligned} \tag{13}$$

where $W^{[3]}$ and $b^{[3]}$ indicates the weight and bias parameters of "Conv3", $W^{[4]}$ and $b^{[4]}$ indicates the weight and bias parameters of "Conv4", $W^{[r]}$ indicates the weight parameters of residual layer. Whereas the calculation formula of nonlinear activation function "ReLU "is as follows:

$$\mathrm{ReLU}(x) = \max(\alpha x, x) \tag{14}$$

In order to ensure that $R(x)$ and $F(x)$ have the same dimension, in the residual layer, 32 convolution filters with kernel size of 1 × 1 and stripe length of 1 are adopted for linear transformation, and the dimension of the feature map is transformed from 16 to 32. The parameters values of residual layer have also been shown in Table 3.

When CNN model is iterative updating, back propagation step needs to be carried out to calculate the weight parameter gradients of each convolution layer. In the network structure of residual layer, the training goal of model optimization is to approximate the residual $F(x) = H(x) - x$ to 0. In this way, the identity transformation of residual layer can ensure that the CNN model will not degenerate as the depth increases, and the deep convolutional layers ("Conv3", "Conv4") will continue looking for a further optimization mechanism based

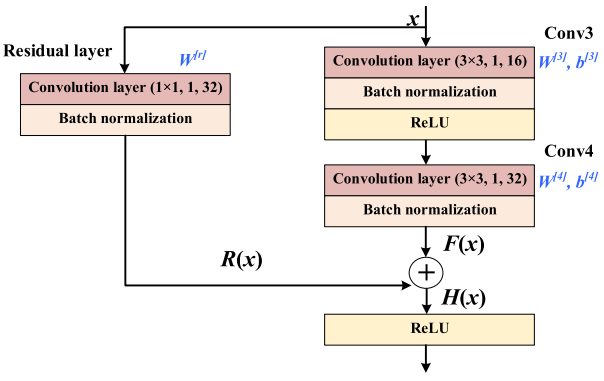


FIGURE 8. Principle of residual layer in CNN model.

on the original performance, that is to optimize the residual error of the identity transformation with small disturbances.

C. MODEL TRAINING AND TESTING

The algorithm is trained based on the CNN model established above. For all data samples, the vital signs samples, environmental information samples and interference information samples are randomly mixed and divided into several batches. The data of one batch at a time are taken for model training. For each input data x_i, after CNN model operation, its output is a K-dimensional vector containing the predicted values of all classification labels, which can be represented as:

$$\vec{q}(x_i) = \left[q_{i,0}, q_{i,1} \ldots, q_{i,K-1}\right] \tag{15}$$

In this paper, there are three types of labels for all data samples, namely, $K = 3$. After each updating iteration of CNN model, the loss function needs to be calculated. According to the characteristics of multi classification tasks in this paper, the loss function is defined as:

$$loss = -\frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{K-1} y_{i,k}\ln(q_{i,k}) + \lambda \sum_l \sum_n (W_n^{[l]})^2 \tag{16}$$

In the loss function, the former is the cross entropy of the prediction results, and the latter is the regularization penalty term. Where N is the batch size, K is the number of classification labels, and $q_{i,k}$ represents the output value for the i^{th} sample being predicted to be the k^{th} label. λ is a regularization parameter, which is a very small constant. $W_n^{[l]}$ represents the weight parameter of each layer in CNN model, l represents the number of layers, and n represents the number of weights in this layer. The introduction of regularization penalty term makes CNN take into account the universality of weight parameters and further reduce the risk of overfitting while training and optimizing continuously. The key parameters and options for model training have been shown in Table 4.

TABLE 4. CNN model training parameters and options.

Parameter/Option	Value
Batch size (N)	64
Total samples number	8400
Training set proportion	80%
Testing set proportion	20%
Maximum epochs	80
Solver	Adam
Initial learn rate	0.003
Learn rate drop period	10
Learn rate drop factor	0.8
Shuffle	every epoch
Drop out	False
Validation frequency (Iterations)	20

As shown in Table 4, 80% of all data samples are used for CNN model training, and the remaining 20% are used for CNN model testing, and the maximum number of training epochs is set to 80. The Adam optimization strategy

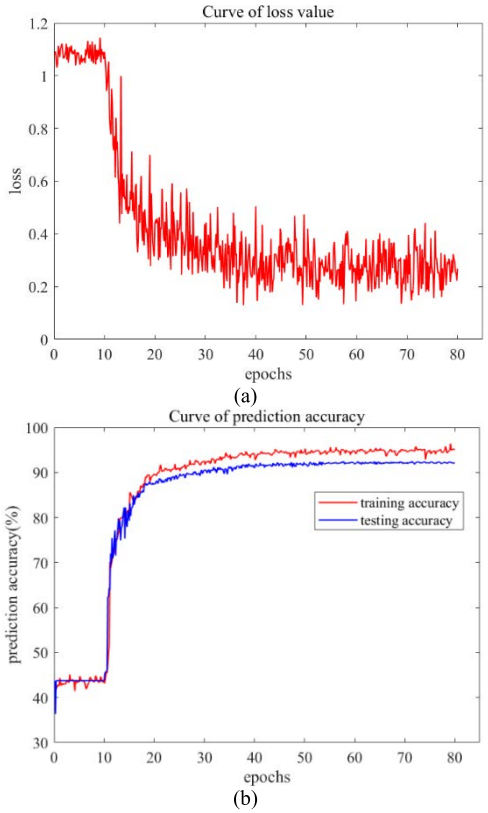


FIGURE 9. Training effect of CNN model: (a) Curve of loss value, (b) Curve of prediction accuracy.

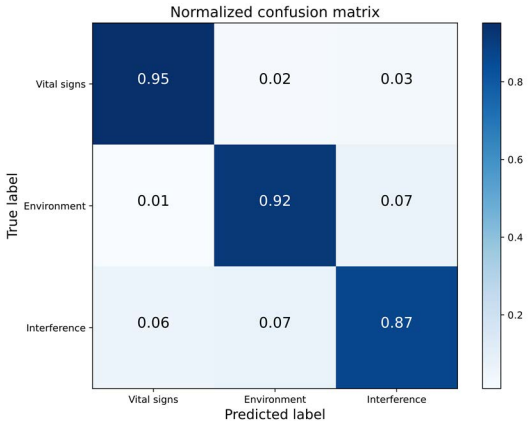


FIGURE 10. Normalized confusion matrix of constructed CNN model.

is selected, and the initial learning rate is set to 0.003. After every 10 times updating, the learning rate will be attenuated with the drop factor of 0.8, so as to slow down the weight parameters updating gradient and optimize the CNN model more finely. With the continuous optimization of CNN model, the curves of loss value and prediction accuracy of training data samples and testing data samples have been shown in Figure 9.

From Figure 9, one can observe that the CNN model finally obtains a very low loss value and high prediction accuracy

for both training data and testing data. In the last few epochs, the performance of CNN model is almost stable, and there is no big fluctuation in both loss and accuracy. It is worth noting that the prediction accuracy of testing data is slightly lower than that of training data, however, the gap is small and acceptable, which means that CNN model has not produced fitting phenomenon. During the CNN model training, the lowest loss value is 0.1292, the highest prediction accuracy for training data is 96.41%, and the highest prediction accuracy for testing data is 92.47%. The CNN model parameters of the best testing prediction accuracy are chosen as the preferred final selection. In addition, the normalized confusion matrix shown in Figure 10 reveals the specific correct and incorrect prediction situation. The prediction success rate for vital signs is the highest among three types of data samples, and most of the failure predictions happened when the CNN model confusing the environment data and the interference data, which have no impact on the vital signs identification purpose for this paper.

For better illustration the performance of the proposed CNN model in our paper, Table 5 records some important properties of the constructed CNN model compared with some standard common CNN models. The basic CNN structure is composed of convolution layer and full connection layer. When the number of convolution layers is too small, the required accuracy cannot be achieved, on the contrary, if the number of convolution layers is too large, the real-time performance of the algorithm will be greatly reduced. After many tests, the 4-layer convolution layer CNN has been proved to be the optimal structure, and the residual layer effectively further improved the identification accuracy. In Table 4, some typical CNN frameworks for image recognition, such as VGG and ResNet-50 have also been verified, due to the differences between this project and image recognition tasks, some important properties, especially the real-time performance, are not applicable for the vital signs identification. The CNN model constructed in this paper not only has a high identification accuracy which satisfy the life detection requirements, but also has the good characteristic of fast detection to ensure the timeliness of rescue.

TABLE 5. Properties comparison of different CNN models.

CNN Structure	Loss	Training Acc.	Testing Acc.	Training Time (s)	Running Time for 1000 data (s)
2 Conv + 1 FC	0.4324	81.96%	80.33%	306	12
3 Conv + 1 FC	0.2879	89.04%	86.78%	379	13
4 Conv + 1 FC	0.1565	92.87%	87.65%	491	16
5 Conv + 1 FC	0.1472	94.86%	89.31%	783	25
VGG	0.1968	88.84%	85.73%	1662	69
ResNet-50	0.1247	96.89%	94.21%	4130	318
Proposed	0.1292	96.41%	92.47%	523	15

D. VITAL SIGNS IDENTIFICATION

The UWB echo matrix is obtained through IR-UWB radar device and processed with techniques above mentioned. Then, a two-dimensional sliding window is adopted to traverse the matrix, convert the data into $u \times v$ form and input it into the trained CNN model. For the four scenarios shown in Figure 2, the vital signs identification effects have been shown in Figure 11. To reveal the superior performance of proposed algorithm, another model based on multi feature fusion and SVM [37], [38] is selected as the comparison group.

From Figure 11, one can conclude that the constructed CNN model has an obvious more superior performance compared with the reference method. For complex tasks such as static multiple targets, regular moving target, multiple simultaneous moving targets and irregular moving targets, the effective recognition of vital signs' information can be realized. The trajectories of the detected targets can be clearly distinguished, and there are only a few false recognition phenomena. In contrast, the reference method can successfully identify human in the static targets' detection task of scenario 1, but it can almost be considered as failures for the dynamic target detection tasks. This is because traditional machine learning algorithms such as SVM cannot extract the potential depth features of data, however, the harmonics and signal distortion caused by moving objects are similar to the vital signs' information in many appearance features, which are difficult to be distinguished.

VI. MOTION ESTIMATION

A. MISIDENTIFIED OUTLIERS REMOVAL

The results of vital signs identified by CNN model still contain a small amount of false identification information (It is not necessarily that CNN model is inaccurate, but also that there are maybe interference factors in the on-site environment.) Interference cancellation of the misidentified outliers based on K-means clustering algorithm [41] is necessary according to the following steps:

According to the Euclidean distance, the k nearest neighbor points matrices as well as the corresponding indexes of each data x_i $(i = 1, 2, \ldots, m)$ in the recognition results can be obtained. The Euclidean distance can be calculated as:

$$d_{xy} = \sqrt{\sum_{k=1}^{n} |x_k - y_k|} \tag{17}$$

where n indicates the data space dimension. According to Euclidean distance, one should find k points closest to each data x_i $(i = 1, 2, \ldots, m)$, and record the data set represented by these k points as $N_k(x)$.

In $N_k(x)$, the cluster category y_i for each x_i $(i = 1, 2, \ldots, m)$ is determined according to the following classification rules:

$$y = \arg\max_j \sum_{x_i \in N_k(x)} I\{y_i = C_j\}, \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots s \tag{18}$$

s is the total number of clusters, and then a k-MST structure is established for identification result data. Calculate the

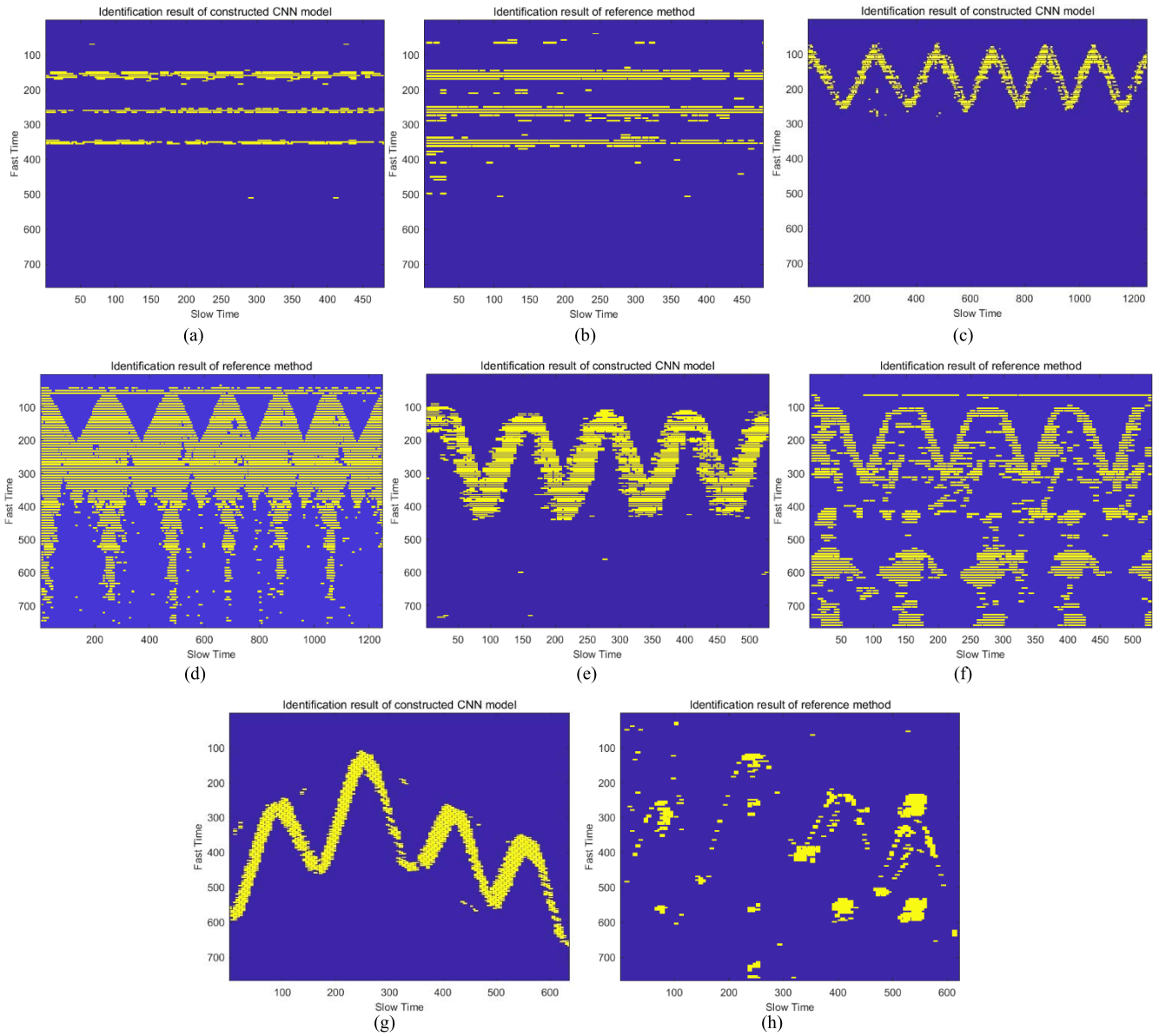


FIGURE 11. Vital signs identification effects of proposed algorithm and reference method: (a) Identification result of constructed CNN model for scenario 1, (b) Identification result of reference method for scenario 1, (c) Identification result of constructed CNN model for scenario 2, (d) Identification result of reference method for scenario 2, (e) Identification result of constructed CNN model for scenario 3, (f) Identification result of reference method for scenario 3, (g) Identification result of constructed CNN model for scenario 4, (h) Identification result of reference method for scenario 4.

maximum Euclidean distance and average Euclidean distance for each category y_i, one can realize the removal of misidentified data. The misidentified outliers' removal effects for all four scenarios have been shown in Figure 12. The k-MST structures are plotted with black lines, and the red cross markers indicate the points considered as the misidentified outliers which should be removed from the identification result data.

B. MOTION TRAJECTORY ESTIMATION AND SHAPED SIGNAL ACQUISITION

After eliminating the misidentified outliers, the real-time motion state of the target can be estimated through appropriate technology. For the static targets in scenario 1, the personnel's position does not change with time, therefore the trajectory presents a straight line in the UWB radar echo matrix. The static target trajectory can be estimated by calculating the centroid position of each cluster data points. The centroid of each cluster in the data clusters $\{C_1, C_2, \ldots, C_k\}$ can be calculated as follows:

$$\mu_j = \frac{1}{|C_j|} \sum_{x \in C_j} x \tag{19}$$

The centroid positions and trajectory estimation results of detected targets in scenario 1 have been shown in Figure 13 (a). In contrast, the motion trajectory estimation

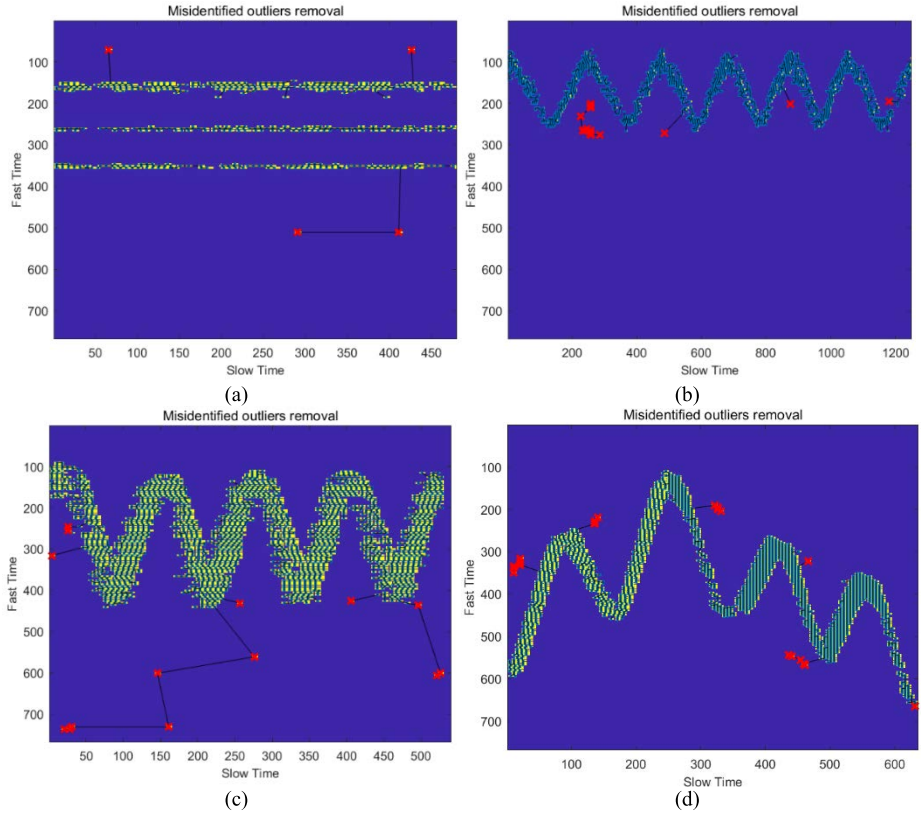


FIGURE 12. The misidentified outliers' removal effects: (a) Result for scenario 1, (b) Result for scenario 2, (c) Result for scenario 3, (d) Result for scenario 4.

method of dynamic targets in scenario 2, 3 and 4 is more complex. In this paper, Kalman filtering is adopted for optimally estimation of the motion state from the observation data which is identified by CNN model.

In the UWB radar detection system of this paper, with the change of slow time, the state equation and observation equation of the system are as follows:

$$X_k = A_k X_{k-1} + B_k U_k + w_k \tag{20}$$

$$Z_k = H_k X_k + v_k \tag{21}$$

In the state equation, X_k is the estimated value of the system state at time k, U_k represents the control quantity applied to the system at time k, A_k is the state transition matrix at time k-1, B_k is the control matrix on U_k, which describes the influence of control quantity on system state, w_k indicates the process noise, which follows the independent normal distribution with mean value of 0 and covariance of Q_k. In the observation equation, Z_k is the measured value of the system state at time k, and H_k is the observation matrix, which describes the mapping relationship from the actual state to observation state; v_k indicates the observation noise, which follows the independent normal distribution with mean value of 0 and covariance of R_k.

Kalman filtering algorithm includes two stages: prediction and updating. In the prediction stage, the algorithm calculates the current state according to the previous records of system state. In the updating stage, the algorithm comprehensively considers the current measured value and the predicted value of previous time and gives the optimal estimation of the current system state.

In the prediction stage, the calculation method is as follows:

$$\hat{X}_k^- = A_k \hat{X}_{k-1} + B_k U_k \tag{22}$$

$$P_k^- = \mathrm{cov}(X_k - \hat{X}_k^-) = A_k P_{k-1} A_k^T + Q_k \times (Q_k = \mathrm{cov}(w_k)) \tag{23}$$

Firstly, according to the state estimation value at k-1 time, the system prior prediction at k time $\hat{X}_k^-$ can be obtained, where $\hat{X}_{k-1}$ represents the optimal estimation of the system state at k-1 time. Then, the estimation error P_k^- between the predicted value and the actual value of the system at time k is calculated, where Q_k indicates the covariance matrix of the process noise w_k.

In the updating stage, the calculation method is as follows:

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \times (R_k = \mathrm{cov}(v_k)) \tag{24}$$

$$\hat{X}_k = \hat{X}_k^- + K_k (Z_k - H_k \hat{X}_k^-) \tag{25}$$

$$P_k = (I - K_k H_k) P_k^- \tag{26}$$

Firstly, the Kalman gain K_k is calculated, where R_k indicates the covariance matrix of the observation noise v_k.

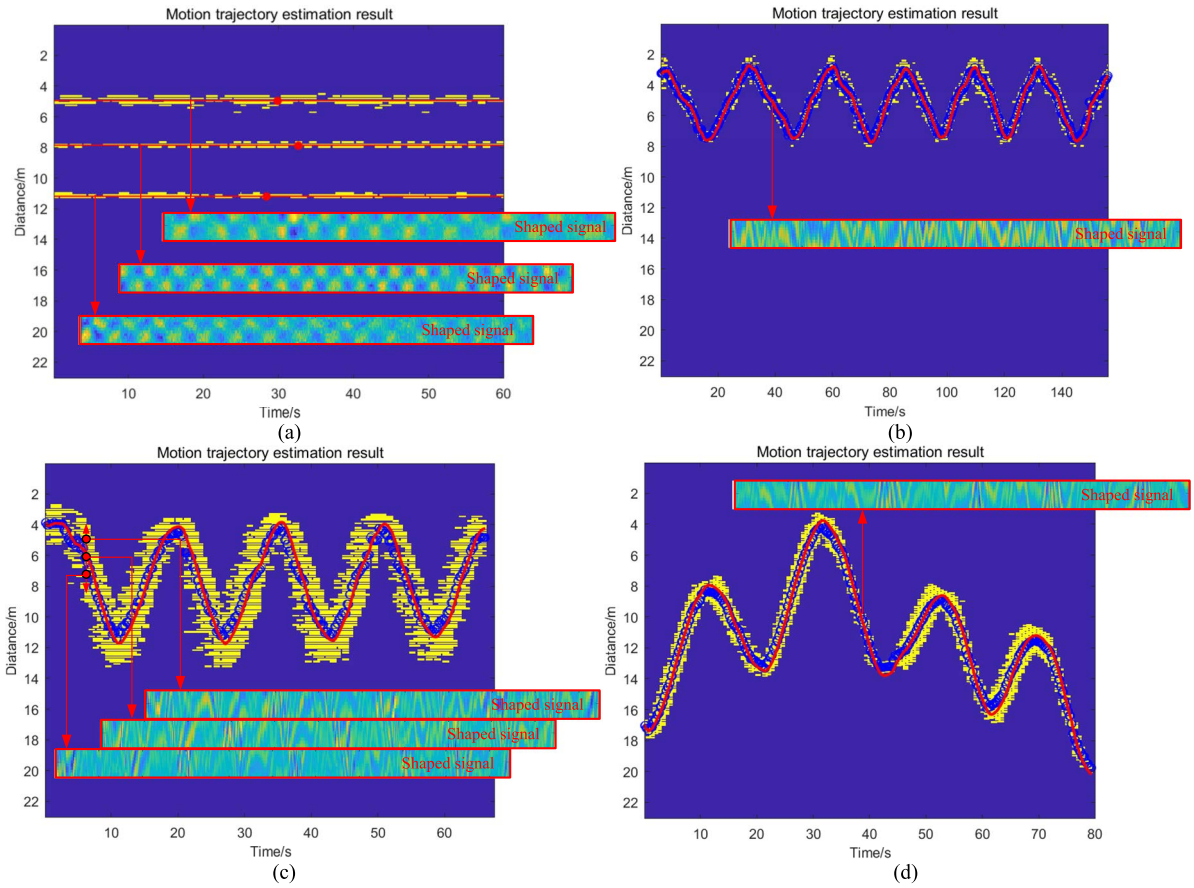


FIGURE 13. The motion trajectory estimation and shaped signal acquisition results: (a) Result for scenario 1, (b) Result for scenario 2, (c) Result for scenario 3, (d) Result for scenario 4.

The Kalman gain describes the relationship between the predicted value and the measured value of the system state at time k, and its optimization makes the estimated state close to the real state to the greatest extent. Then, the optimal estimation $\hat{X}_k$ of the system state is calculated by integrating the prior prediction $\hat{X}_k^-$, the measured value Z_k and the Kalman gain K_k. Finally, the posteriori error covariance matrix P_k is updated.

As in the above steps, the Kalman filtering algorithm is continuously updating and iterating to obtain the motion trajectory estimation results of the dynamic targets.

In order to simplify the task, the constant velocity model is adopted to describe the motion properties of the dynamic targets. When estimating the system state with Kalman filtering algorithm, the state of the system X_k includes two two-dimensional parameters of position and velocity. The measured value Z_k of the system can be obtained from the average position of the identification results by CNN model in each slow time, that is:

$$X_k = \begin{bmatrix} x_k \\ y_k \\ v_k^x \\ v_k^y \end{bmatrix} \quad Z_k = \begin{bmatrix} \bar{N}_x \,|t = k \\ \bar{N}_y \,|t = k \end{bmatrix} \tag{27}$$

The state transition matrix A_k, the control matrix B_k and the observation matrix H_k can be defined as:

$$A_k = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad B_k = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$
$$H_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{28}$$

The process noise w_k and the observation noise v_k are regarded as invariant parameters, and according to the UWB system characteristic, their covariance matrices can be evaluated as:

$$Q_k = 10^{-4} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad R_k = 10^{-2} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{29}$$

After the above Kalman filtering method calculation, the motion trajectory estimation results of detection targets in all scenarios can be obtained. Simultaneously, according to the estimation trajectories, the original curve-shaped life signals are shaped into straight lines to lay the foundation for

the follow-up vital signs extraction. The results have been shown in Figure 13. In particular, the identified signal band of vital signs in Figure 13 (c) is significantly wider, because it contains vital signs of three tested personnel. According to the motion trajectory extracted by Kalman filter, the whole signal band can be divided into three parts for extraction of respiratory and heartbeat signals of three people respectively.

VII. VITAL SIGNS EXTRACTION

In the detection tasks of this paper, due to the motion of the objects, a large number of dynamic interference clutters are introduced, and the originally regular vital signs are also distorted. Therefore, traditional vital signs extraction methods are hardly applicable. Based on EEMD method, this paper develops an improved SVD-EEMD algorithm to achieve dynamic vital signs extraction under interference and distortion. Due to the length limitation of this paper, the following texts will take the detection task in scenario 2 as an example to illustrate the SVD-EEMD algorithm, while for scenarios 1, 3 and 4, only the final results will be shown, and the intermediate process will not be described in detail.

Firstly, Singular Value Decomposition (SVD) is performed on the shaped life signal X_{NxM}. The process is as follows:

$$X_{N\times M} = U_{N\times N}\Sigma_{N\times M}V^T_{M\times M} = \sum_{i=1}^{k} u_i\sigma_i v_i^T \tag{30}$$

$$\Sigma_{N\times M} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \end{bmatrix}_{N\times M} \tag{31}$$

$$U_{N\times N} = \left[\vec{u}_1, \vec{u}_2, \vec{u}_3, \cdots \vec{u}_N\right] \tag{32}$$

$$V_{M\times M} = \left[\vec{v}_1, \vec{v}_2, \vec{v}_3, \cdots \vec{v}_M\right] \tag{33}$$

where Σ_{NxM} is the singular value matrix, the main diagonal elements $\sigma_1, \sigma_2, \ldots, \sigma_k$ are k singular values of X_{NxM} ($\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_k$). u_i represents the i^{th} column vector of the matrix U_{NxN}, which is called the i^{th} order left singular vector. v_i represents the i^{th} column vector of the matrix V_{MxM} which is called the i^{th} order right singular vector.

SVD algorithm can perform macro analysis on the acquired shaped vital signals and extract the universal characteristic components of vital signs from interference, clutters, and distortion. After SVD, the first-order principal component of the signal can be regarded as the most important component, and the singular value of the first-order principal component σ_1 represents the specific gravity of the principal component. The first order left singular vector u_1 represents the change of the principal component signal in spatial position (fast time), which is related to the distortion, deformation, and dynamic characteristics of the signal. The first order right singular vector v_1 represents the change of the principal component signal in the time dimension (slow time), which includes the change characteristics of vital signs such as heartbeat and respiration. In the detection task of scenario 2, the first order right singular vector of acquired life signal is shown in Figure 14.

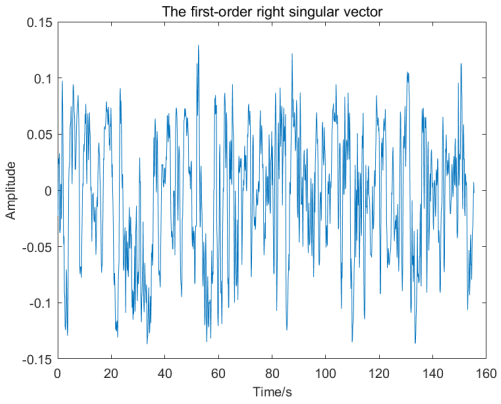


FIGURE 14. The first order right singular vector of the acquired life signal in scenario 2.

The first order right singular vector of the obtained life signal is performed with EEMD to extract vital signs of respiration and heartbeat. The specific steps are as follows:

Firstly, Gaussian white noise is added to first order right singular vector $v(t)$:

$$s_i(t) = v(t) + n_i(t) \tag{34}$$

where, $n_i(t)$ represents the Gaussian white noise signal, and $s_i(t)$ represents the signal after the i^{th} time addition process of white noise. Then, perform empirical mode decomposition (EMD) on $s_i(t)$ according to the steps described in [22] to obtain all IMF components:

$$s_i(t) = \sum_{k=1}^{n} c_{i,k}(t) + r_{i,k}(t) \tag{35}$$

where, n is the number of IMF decomposed by EMD, $c_{i,k}(t)$ are IMF components (IMFs), $r_{i,k}(t)$ is the residual component. Repeat the above process for M cycles, add white noise with different amplitude in each cycle, and a series of IMFs can be obtained:

$$[\{c_{1,n}(t)\}, \{c_{2,n}(t)\}, \ldots, \{c_{M,n}(t)\}] \quad n = 1, 2, \ldots, N; \quad i = 1, 2, \ldots, M \tag{36}$$

Then average all IMF components, and take it as the final target IMF component of EEMD $c_n(t)$:

$$c_n(t) = \frac{1}{M}\sum_{i=1}^{M} c_{i,n}(t) \tag{37}$$

where $c_{i,n}(t)$ is the IMF component in the i^{th} cycle. The time domain waveform of each IMF component after EEMD for scenario 2 has been shown in Figure 15.

Important components are concentrated in the low frequency range, in which the distribution range of heartbeat frequency is 1∼2.2Hz and the distribution range of respiratory frequency is 0.2∼0.8Hz. Therefore, some IMFs within

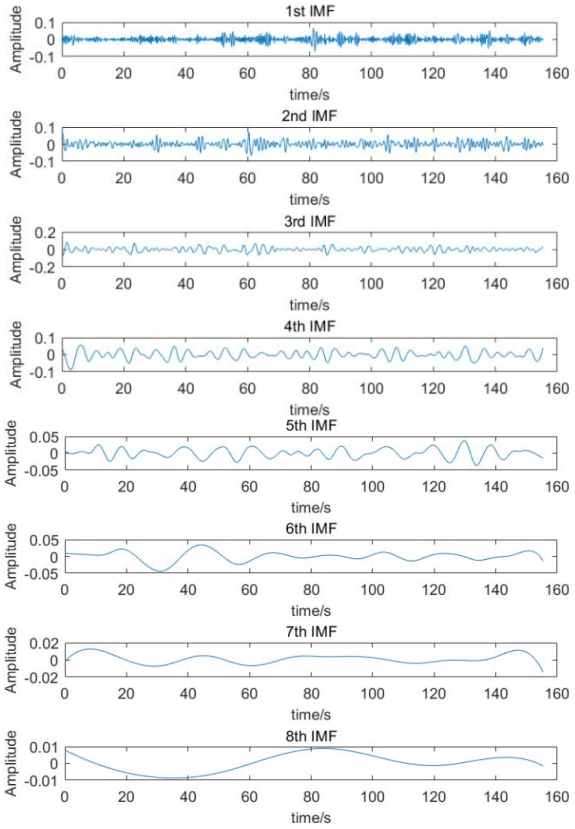


FIGURE 15. Waveform of IMF components after EEMD for scenario 2.

the spectrum of vital signs can be used to reconstruct respiratory signals and heartbeat signals. Preform Fourier transformation on each IMF component, then calculate the total energy $E(f)$ in frequency domain, the energy $Er(f)$ within the respiratory frequency distribution range (0.2∼0.8Hz), and the energy $Eh(f)$ within the heartbeat frequency distribution range (1∼2.5Hz). The Energy proportion of respiratory and heartbeat range are calculated as $Er(f)/E(f)$ and $Eh(f)/E(f)$ respectively, and the IMF selection threshold for respiratory signal and heartbeat signal are set as 0.8 and 0.5 respectively.

Therefore, the energy percentage within respiratory and heartbeat bands of each IMF component for scenario 2 has been shown in Figure 16. Finally, the first IMF component is selected to reconstruct the heartbeat signal and the third IMF component is selected to reconstruct the respiratory signal.

In order to further extract the heartbeat signal, a Butterworth filter is established, the upper and lower boundary frequencies of the passband are set to 1.0 Hz and 2.2 Hz respectively, the passband attenuation is set to 0.1 dB. The upper and lower boundary frequencies of the stopband are set to 0.7 Hz and 2.5 Hz, respectively, the stopband attenuation is set to 1 dB. The order of the Butterworth filter equals 2. Finally, the vital signs extraction of respiratory signal and heartbeat signal can be realized. Perform the above procedures for the tasks in scenarios 1, 2, 3, and 4, one can obtain the final vital signs extraction results, the waveforms and

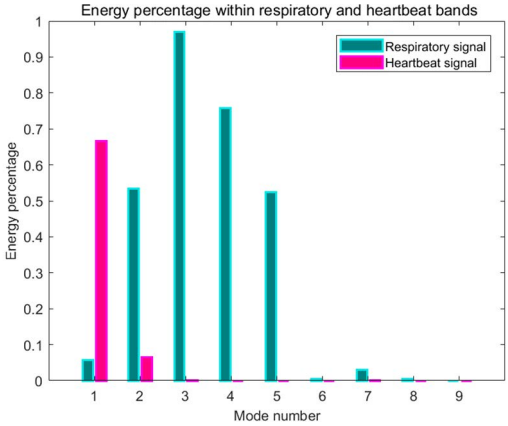


FIGURE 16. The energy percentage within respiratory and heartbeat bands of each IMF component for scenario 2.

power spectrums of respiratory and heartbeat signals for all objects in scenarios 1, 2, 3, and 4 are shown in Figure 17.

VIII. RESULTS AND DISCUSSION

In order to verify the progressiveness of the proposed method, 20 sets of repeated experiments for the proposed algorithm and some existing methods in the above four scenarios with different setting parameters were carried out, some important indicators were calculated during each detection task in four scenarios, the results have been shown in Table 6. Where, the indicator "Accuracy" indicates the vital signs identification accuracy of specific algorithm, the indicator "RMSE" indicates root mean square error of the estimation trajectory relative to the real motion trajectory, the indicator "RF" indicates respiratory frequency, the indicator "R-SNR" indicates signal-to-noise ratio of extracted respiratory signal, the indicator "HF" indicates heartbeat frequency, the indicator "H-SNR" indicates signal-to-noise ratio of extracted heartbeat frequency, the indicator "Time" indicates algorithm running time in specific device configuration for each detection task, and the indicator "Success rate" indicates the task success rate of 20 groups of repeated experiments after changing the settings of different personnel, distance, movement speed, obstacles and so on. In addition, most of the exiting methods have only achieved dynamic target trajectory detection, and not extracted vital signs information, therefor the "CLEAN-KF" method [29], "SVM" method [37] and OS-CFAR [35] are chosen as the control group for indicators "Accuracy", "RMSE", "Time" and "Success rate" to illuminate the improvement of proposed method, other indicators are advanced functions that have not been implemented before.

The measurement results of key indicators in Table 6 show that the proposed algorithm has high accuracy, real-time performance, robustness and stability. In multiple static targets detection task, most of the algorithms can ensure the successful detection, however the CNN model recognition accuracy and position detection accuracy of the algorithm in this paper are the highest. In the dynamic target detection tasks, the

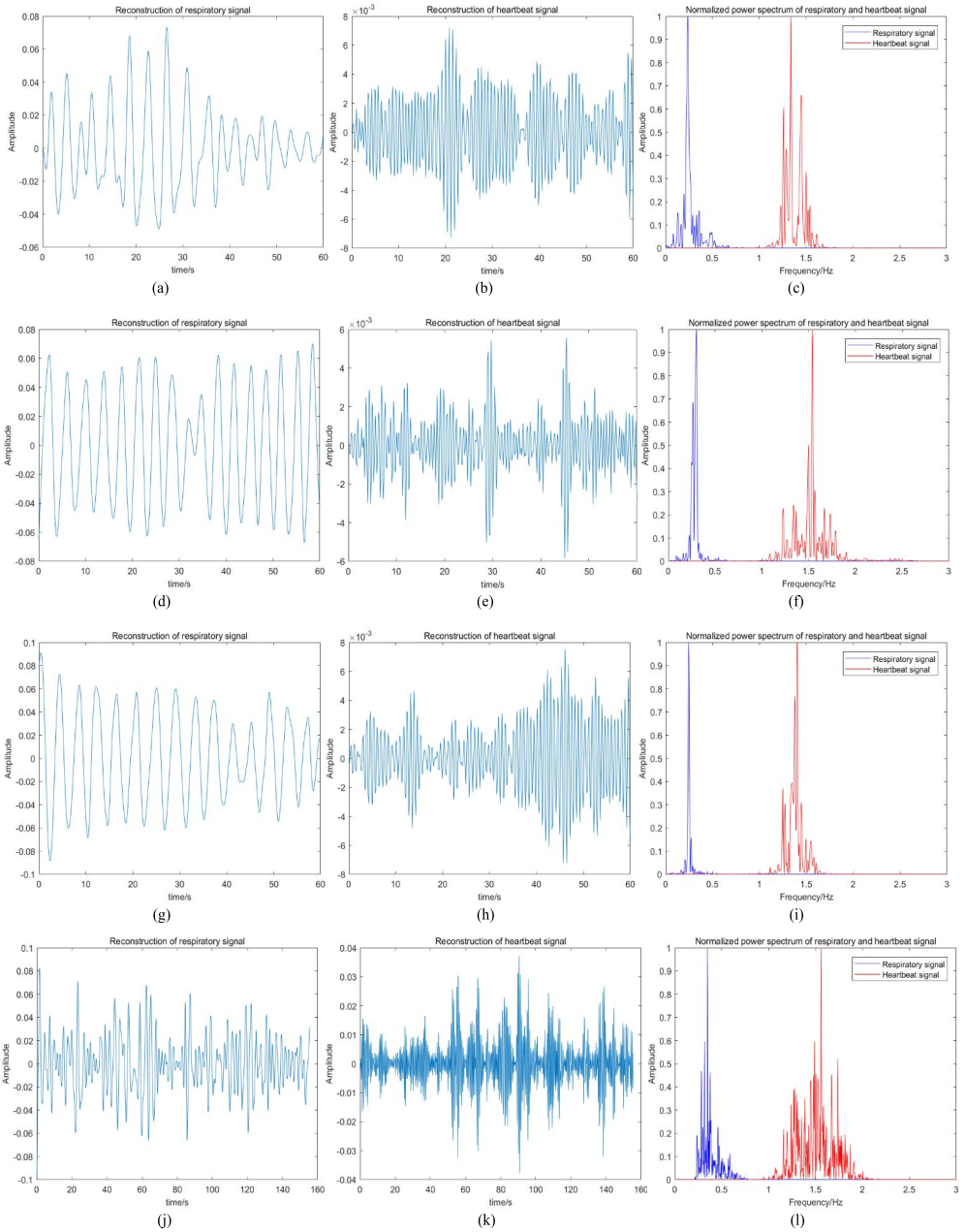


FIGURE 17. The waveforms and power spectrums of respiratory and heartbeat signals for all detected person in scenarios 1, 2, 3, and 4: (a) Waveform of respiratory signal of personnel 1 in scenario 1, (b) Waveform of heartbeat signal of personnel 1 in scenario 1, (c) Normalized spectrum of respiratory and heartbeat signals of personnel 1 in scenario 1, (d) Waveform of respiratory signal of personnel 2 in scenario 1, (e) Waveform of heartbeat signal of personnel 2 in scenario 1, (f) Normalized spectrum of respiratory and heartbeat signals of personnel 2 in scenario 1, (g) Waveform of respiratory signal of personnel 3 in scenario 1, (h) Waveform of heartbeat signal of personnel 3 in scenario 1, (i) Normalized spectrum of respiratory and heartbeat signals of personnel 3 in scenario 1, (j) Waveform of respiratory signal of personnel in scenario 2, (k) Waveform of heartbeat signal of personnel in scenario 2, (l) Normalized spectrum of respiratory and heartbeat signals of personnel in scenario 2, (m) Waveform of respiratory signal of personnel 1 in scenario 3, (n) Wave-form of heartbeat signal of personnel 1 in scenario 3, (o) Normalized spectrum of respiratory and heartbeat signals of personnel 1 in scenario 3, (p) Waveform of respiratory signal of personnel 2 in scenario 3, (q) Waveform of heartbeat signal of personnel 2 in scenario 3, (r) Normalized spectrum of respiratory and heartbeat signals of personnel 2 in scenario 3, (s) Waveform of respiratory signal of personnel 3 in scenario 3, (t) Waveform of heartbeat signal of personnel 3 in scenario 3, (u) Normalized spectrum of respiratory and heartbeat signals of personnel 3 in scenario 3, (v) Waveform of respiratory signal of personnel in scenario 4, (w) Waveform of heartbeat signal of personnel in scenario 4, (x) Normalized spectrum of respiratory and heartbeat signals of personnel in scenario 4.

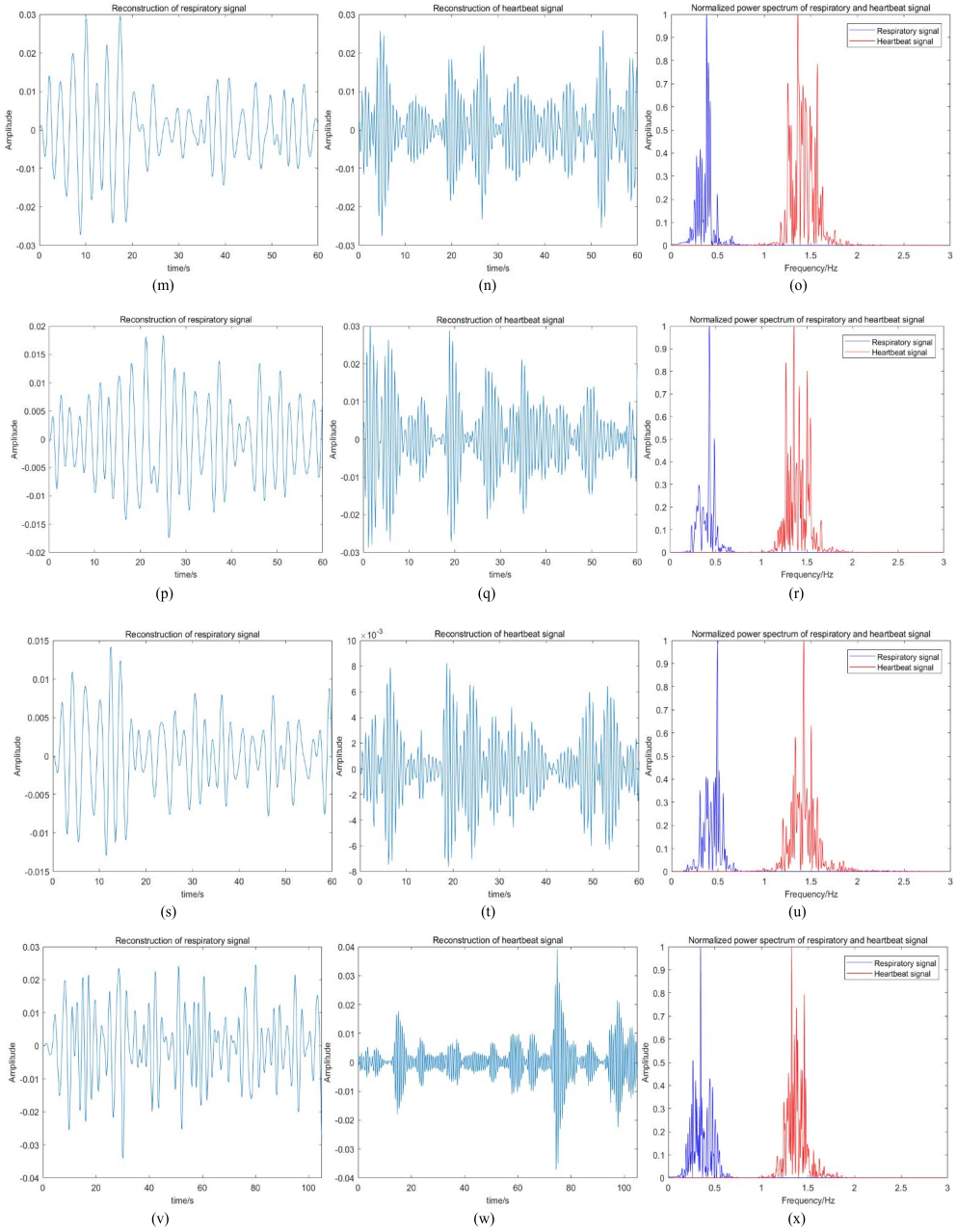


FIGURE 17. *(Continued.)* **The waveforms and power spectrums of respiratory and heartbeat signals for all detected person in scenarios 1, 2, 3, and 4: (a) Waveform of respiratory signal of personnel 1 in scenario 1, (b) Waveform of heartbeat signal of personnel 1 in scenario 1, (c) Normalized spectrum of respiratory and heartbeat signals of personnel 1 in scenario 1, (d) Waveform of respiratory signal of personnel 2 in scenario 1, (e) Waveform of heartbeat signal of personnel 2 in scenario 1, (f) Normalized spectrum of respiratory and heartbeat signals of personnel 2 in scenario 1, (g) Waveform of respiratory signal of personnel 3 in scenario 1, (h) Waveform of heartbeat signal of personnel 3 in scenario 1, (i) Normalized spectrum of respiratory and heartbeat signals of personnel 3 in scenario 1, (j) Waveform of respiratory signal of personnel in scenario 2, (k) Waveform of heartbeat signal of personnel in scenario 2, (l) Normalized spectrum of respiratory and heartbeat signals of personnel in scenario 2, (m) Waveform of respiratory signal of personnel 1 in scenario 3, (n) Wave-form of heartbeat signal of personnel 1 in scenario 3, (o) Normalized spectrum of respiratory and heartbeat signals of personnel 1 in scenario 3, (p) Waveform of respiratory signal of personnel 2 in scenario 3, (q) Waveform of heartbeat signal of personnel 2 in scenario 3, (r) Normalized spectrum of respiratory and heartbeat signals of personnel 2 in scenario 3, (s) Waveform of respiratory signal of personnel 3 in scenario 3, (t) Waveform of heartbeat signal of personnel 3 in scenario 3, (u) Normalized spectrum of respiratory and heartbeat signals of personnel 3 in scenario 3, (v) Waveform of respiratory signal of personnel in scenario 4, (w) Waveform of heartbeat signal of personnel in scenario 4, (x) Normalized spectrum of respiratory and heartbeat signals of personnel in scenario 4.**

TABLE 6. Results and performance comparison with the referenced algorithm.

Scenario	Value	Method								
		CLEAN-KF	SVM	OS-CFAR	Proposed			Measured		
1	Accuracy (%)	94.5%	91.8%	93.7%	99.3%			/		
	RMSE (m)	0.158	0.278	0.192	0.123			/		
	RF (Hz)	/	/	/	0.238	0.300	0.256	0.250	0.283	0.267
	R-SNR (dB)	/	/	/	1.399	2.028	5.074		/	
	HF (Hz)	/	/	/	1.338	1.540	1.404	1.317	1.467	1.367
	H-SNR (dB)	/	/	/	-3.795	-2.441	-1.184		/	
	Time (s)	39	107	38	48				/	
	Success rate (%)	100%	100%	100%	100%				/	
2	Accuracy (%)	92.2%	62.4%	88.2%	98.7%				/	
	RMSE (m)	0.234	fail	0.317	0.142				/	
	RF (Hz)	/	/	/	0.348				0.383	
	R-SNR (dB)	/	/	/	-1.603				/	
	HF (Hz)	/	/	/	1.556				1.367	
	H-SNR (dB)	/	/	/	-4.875				/	
	Time (s)	82	232	86	104				/	
	Success rate (%)	100%	10%	90%	100%				/	
3	Accuracy (%)	90.7%	71.8%	88.4%	95.2%				/	
	RMSE (m)	0.274	fail	0.386	0.207				/	
	RF (Hz)	/	/	/	0.382	0.432	0.488	0.367	0.400	0.417
	R-SNR (dB)	/	/	/	-0.374	-1.673	-2.048		/	
	HF (Hz)	/	/	/	1.362	1.358	1.422	1.333	1.316	1.383
	H-SNR (dB)	/	/	/	-5.049	-4.552	-4.297		/	
	Time (s)	48	168	45	61				/	
	Success rate (%)	95%	0%	80%	100%				/	
4	Accuracy (%)	91.2%	64.8%	89.27%	96.9%				/	
	RMSE (m)	0.279	fail	0.301	0.146				/	
	RF (Hz)	/	/	/	0.346				0.350	
	R-SNR (dB)	/	/	/	-2.871				/	
	HF (Hz)	/	/	/	1.324				1.333	
	H-SNR (dB)	/	/	/	-3.097				/	
	Time (s)	40	139	39	47				/	
	Success rate (%)	90%	0%	80%	100%				/	

SVM method is difficult to achieve successful detection, the CLEAN-KF method and OS-CFAR method have good performance, and proposed method has a higher identification accuracy, a smaller motion trajectory estimation error and only a little bit longer running time relative to CLEAN-KF method and OS-CFAR method. At the same time, when the experimental settings are changed, the CLEAN-KF method and OS-CFAR method fail in some highly complex scenarios such as multi-target interference and long-range detection, however the proposed method still has a high detection success rate, which illustrates the high sensitivity and stability of it. Furthermore, a more complete scheme which can extract the vital signs information of respiratory and heartbeat signals is proposed. For most detection tasks, the extracted respiratory signal is relatively pure and accurate, but the signal-to-noise ratio of heartbeat signal is relatively low, this is because the human heartbeat signal is inherently unstable and vulnerable to interference during motion.

IX. CONCLUSION

In this paper, an intelligent signal processing method for remote vital signs detection which is suitable not only for stationary targets but also for dynamic targets is proposed. The detection experiments are carried out based on NVA6100 pulse radar system and four different simulation scenarios under complex rescue conditions are established. A CNN model which contains four convolution layers, one full connection layer and a residual layer structure is established for vital signs identification, the identification results show that the CNN model can achieve precise recognition of vital signs for both stationary targets and dynamic targets getting rid of the interference of harmonics and distortion. The motion trajectory is estimated by misidentified outliers' elimination with K-means clustering algorithm and Kalman filtering algorithm, then the curved life signals can be shaped into straight lines according to the motion trajectory. Finally, the SVD-EEMD algorithm is proposed for vital signs extraction of dynamic targets, the feature vector of principal component is extracted after SVD to obtain the important characteristics of vital signs, then EEMD algorithm is adopted for respiratory signal and heartbeat signal extraction. The experiments results and some important indicators compared with reference methods illustrates that the proposed algorithm not only has high accuracy, real-time performance, robustness and stability for dynamic targets identification and motion trajectory estimation, but also build a more systematic scheme which can extract the vital signs information of respiratory and heartbeat signals.

REFERENCES

[1] Z. H. Chai, "View on foreign radar technology development," *Mod. Defense Technol.*, 2004.
[2] W. Hu, W.-Q. Wu, X.-Z. Zheng, and J. Guo, "Application and development trend of radar detection technology in mine rescue," in *Proc. 11th Int. Mine Ventilation Congr.*, 2019, pp. 942–950.
[3] O. Sisma, A. Gaugue, and C. Liebe, *UWB Radar: Vision Through a Wall*. Springer, 2007.
[4] H.-B. Li and R. Miura, "Fundamental study on UWB radar for respiration and heartbeat detection," *IEICE Trans. Commun.*, vol. E97.B, no. 3, pp. 594–601, 2014.
[5] S. C. Gao, Y. Su, C. L. Huang, and Y. Li, "Using of mean-subtraction method to suppress the direct wave in instantaneous signal receiving," *Syst. Eng. Electron.*, vol. 26, no. 1, pp. 21–25, 2004.
[6] A. Nezirovic, "Stationary clutter- and linear-trend suppression in impulse-radar-based respiratory motion detection," in *Proc. IEEE Int. Conf. Ultra-Wideband (ICUWB)*, Sep. 2011, pp. 331–335.
[7] K. J. Wang, X. Y. Xiong, and Z. Ren, "Highly efficient mean filtering algorithm," *Appl. Res. Comput.*, vol. 27, no. 2, pp. 434–438, 2010.
[8] B. J. Justusson, *Median Filtering: Statistical Properties*. Berlin, Germany: Springer, 1981.
[9] H. Kurumatani and S. Katsura, "Auto gain control technique realizing wide-range and precise A/D conversion for dynamic signal sensing," *SICE J. Control, Meas., Syst. Integr.*, vol. 12, no. 1, pp. 20–28, Jan. 2019.
[10] D. Yang, Z. Zhu, J. Zhang, and B. Liang, "The overview of human localization and vital sign signal measurement using handheld IR-UWB through-wall radar," *Sensors*, vol. 21, no. 2, p. 402, 2021.
[11] B. G. Osgood, *Lectures on the Fourier Transform and Its Applications*. Providence, RI, USA: American Mathematical Society, 2019.
[12] M. Rhif, A. B. Abbes, I. R. Farah, B. Martínez, and Y. Sang, "Wavelet transform application for/in non-stationary time-series analysis: A review," *Appl. Sci.*, vol. 9, no. 7, p. 1345, 2019.
[13] S. Maheshwari and A. Kumar, "Empirical mode decomposition: Theory & applications," *Int. J. Electron. Eng.*, vol. 7, no. 8, pp. 873–878, 2014.
[14] A. Tariq and H. Ghafouri-Shiraz, "Vital signs detection using Doppler radar and continuous wavelet transform," in *Proc. 5th Eur. Conf. Antennas Propag. (EUCAP)*, 2011, pp. 285–288.
[15] M. He, Y. Nian, L. Xu, L. Qiao, and W. Wang, "Adaptive separation of respiratory and heartbeat signals among multiple people based on empirical wavelet transform using UWB radar," *Sensors*, vol. 20, no. 17, p. 4913, 2020.
[16] C. P. Lai, Q. Ruan, and R. M. Narayanan, "Hilbert–Huang transform (HHT) analysis of human activities using through-wall noise radar," in *Proc. Int. Symp. Signals, Syst. Electron.*, 2007, pp. 115–118.
[17] C. P. Lai, R. M. Narayanan, and Q. Ruan, "Hilbert–Huang transform analysis of human activities using through-wall noise and noise-like radar," *IET Radar, Sonar Navigat.*, vol. 2, no. 4, pp. 244–255, 2008.
[18] J. Feng and S. Pan, "Extraction algorithm of vital signals based on empirical mode decomposition," *J. South China Uni. Techn.*, vol. 38, pp. 1–6, Oct. 2010.
[19] K.-K. Shyu, L.-J. Chiu, P.-L. Lee, T.-H. Tung, and S.-H. Yang, "Detection of breathing and heart rates in UWB radar sensor data using FVPIEF-based two-layer EEMD," *IEEE Sensors J.*, vol. 19, no. 2, pp. 774–784, Jan. 2018.
[20] L. Liu, Z. Liu, and B. E. Barrowes, "Through-wall bio-radiolocation with UWB impulse radar: Observation, simulation and signal extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 4, pp. 791–798, Dec. 2011.
[21] D. Yang, Z. Zhu, and B. Liang, "Vital sign signal extraction method based on permutation entropy and EEMD algorithm for ultra-wideband radar," *IEEE Access*, vol. 7, pp. 178879–178890, 2019.
[22] Z. Zhu, D. Yang, R. Zhao, and B. Liang, "Vital sign signal extraction method based on permutation entropy and EMD algorithm for ultra-wideband radar," in *Proc. 3rd Int. Conf. Electron. Inf. Technol. Comput. Eng. (EITCE)*, Oct. 2019, pp. 1268–1273.
[23] J. Yan, H. Hong, H. Zhao, Y. Li, C. Gu, and X. Zhu, "Through-wall multiple targets vital signs tracking based on VMD algorithm," *Sensors*, vol. 16, no. vol. 8, p. 1293, Aug. 2016.
[24] V. Kolisetty and D. Rajput, "A review on the significance of machine learning for data analysis in big data," *Jordanian J. Comput. Inf. Technol.*, vol. 6, no. 1, pp. 155–171, 2020.
[25] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.
[26] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.
[27] I. Sohn, "Deep belief network based intrusion detection techniques: A survey," *Exp. Syst. Appl.*, vol. 167, Apr. 2021, Art. no. 114170.
[28] Y. Song, J. Hu, N. Chu, T. Jin, J. Zhang, and Z. Zhou, "Building layout reconstruction in concealed human target sensing via UWB MIMO through-wall imaging radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1199–1203, Aug. 2018.
[29] V.-H. Nguyen and J.-Y. Pyun, "Location detection and tracking of moving targets by a 2D IR-UWB radar system," *Sensors*, vol. 15, no. 3, pp. 6740–6762, 2015.
[30] X. Yang, W. Yin, L. Li, and L. Zhang, "Dense people counting using IR-UWB radar with a hybrid feature extraction method," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 30–34, Jan. 2019.
[31] S. Liu, Q. Qi, H. Cheng, L. Sun, Y. Zhao, and J. Chai, "A vital signs fast detection and extraction method of UWB impulse radar based on SVD," *Sensors*, vol. 22, no. 3, p. 1177, Feb. 2022.
[32] R. Bao and Z. Yang, "CNN-based regional people counting algorithm exploiting multi-scale range-time maps with an IR-UWB radar," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13704–13713, Jun. 2021.
[33] H. Lv, F. Qi, Y. Zhang, T. Jiao, F. Liang, Z. Li, and J. Wang, "Improved detection of human respiration using data fusion basedon a multistatic UWB radar," *Remote Sens.*, vol. 8, no. 9, p. 773, Sep. 2016.
[34] H. B. Kwon, D. Son, D. Lee, H. Yoon, M. H. Lee, Y. J. Lee, S. H. Choi, and K. S. Park, "Hybrid CNN-LSTM network for real-time apnea-hypopnea event detection based on IR-UWB radar," *IEEE Access*, vol. 10, pp. 17556–17564, 2022.
[35] D. Urdzik and D. Kocur, "CFAR detectors for through wall tracking of moving targets by M-sequence UWB radar," in *Proc. 20th Int. Conf. Radioelektronika*, Apr. 2010, pp. 1–4.
[36] Z. Cao, J. Li, C. Song, Z. Xu, and X. Wang, "Compressed sensing-based multitarget CFAR detection algorithm for FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9160–9172, Nov. 2021.
[37] J. Bryan and Y. Kim, "Classification of human activities on UWB radar using a support vector machine," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, Jul. 2010, pp. 1–4.
[38] S. P. Rana, M. Dey, H. U. Siddiqui, G. Tiberi, M. Ghavami, and S. Dudley, "UWB localization employing supervised learning method," in *Proc. IEEE 17th Int. Conf. Ubiquitous Wireless Broadband (ICUWB)*, Sep. 2017, pp. 1–5.
[39] X. Yang, X. Zhang, Y. Ding, and L. Zhang, "Indoor activity and vital sign monitoring for moving people with multiple radar data fusion," *Remote Sens.*, vol. 13, no. 18, p. 3791, 2021.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[41] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.

SIYUN LIU received the B.S. degree from Huazhong University of Science and Technology, Wuhan, China, in 2017, and the M.S. degree from Tsinghua University, Beijing, China, in 2020. He is currently a Researcher with the Emergency Research Institute, China Coal Research Institute, Beijing. His research interests include machine learning, deep learning, robotics, UWB radar signal processing, SLAM, intelligent algorithm, and artificial intelligence.

QINGJIE QI received the Ph.D. degree from the College of Energy Engineering, Zhejiang University, Zhejiang, China, in 2002. He is currently a Chief Scientist with the Emergency Research Institute, China Coal Research Institute CCRI, Beijing, China. His current research interests include mining engineering, UWB radar rescue detection, signal processing, and safety engineering.

HUIFENG CHENG received the B.S. degree from Dalian Jiaotong University, Dalian, China, in 2005, and the M.S. degree from the Dalian University of Technology, Dalian, in 2009. He is currently a Senior Engineer with the Emergency Research Institute, China Coal Research Institute, Beijing, China. His research interests include control engineering, UWB radar signal processing, deep learning, emergency rescue equipment, emergency detection, and advanced rescue technology.

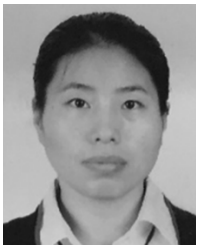
JINGWEN ZHANG received the B.S. degree from Xi'an Jiaotong University, Xi'an, China, in 2018, and the M.S. degree from Tsinghua University, Beijing, China, in 2020. She is currently a Researcher with the Emergency Research Institute, China Coal Research Institute, Beijing. Her research interests include combustion and pollution control, UWB radar signal processing, and deep learning.

WENHAO XIAN received the B.S. degree from Beijing Jiaotong University, Beijing, China, in 2017, and the M.S. degree from Tsinghua University, Beijing, in 2020. He is currently a Researcher with the Emergency Research Institute, China Coal Research Institute, Beijing. His research interests include communication technology, tribology, machine learning, robotics, UWB radar signal processing, and intelligent algorithm.

TIANFANG MA received the B.S. degree from Beihang University, Beijing, China, in 2013, and the M.S. degree from Northeastern University, Shenyang, China, in 2017. He is currently a Researcher with the Emergency Research Institute, China Coal Research Institute, Beijing. His research interests include UWB radar signal processing, FDTD Numerical modeling, and CFD.

YUE WANG received the B.S. degree from the University of Electronic Science and Technology, Chengdu, China, in 2017, and the M.S. degree from the University of Science and Technology Beijing, Beijing, China, in 2021. She is currently a Researcher with the Emergency Research Institute, China Coal Research Institute, Beijing. Her research interests include UWB radar signal processing, machine vision, deep learning, and advanced rescue technology.

YINGJIE LIU received the Ph.D. degree from the College of Safety Science and Engineering, Chang'an University, Xi'an, China, in 2018. He is a Postdoctoral Researcher with the Emergency Research Institute, China Coal Research Institute (CCRI), Beijing, China. His current research interests include safety engineering, robotics, emergency rescue equipment, exoskeleton, artificial intelligence, and advanced rescue technology.



DAN LI received the B.S. degree from Jilin University, in 2005, the M.S. degree from Southwest Forestry University, in 2012, and the Ph.D. degree from the University of Chinese Academy of Sciences, in 2018. She is currently an Associate Researcher with Emergency Research Institute, China Coal Research Institute, Beijing, China. Her research interests include deep learning, UAV remote sensing, vegetation remote sensing, hydraulic remote sensing, and disaster remote sensing.

JIAMEI CHAI received the B.S. and the M.S. degrees from Liaoning Technical University, in 2006 and 2009, respectively. She is currently an Associate Researcher with the Emergency Research Institute, China Coal Research Institute, Beijing, China. Her research interests include emergency science and engineering, scientific research management, emergency rescue technology, and machine learning.