IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 24, 2020, accepted June 30, 2020, date of publication July 3, 2020, date of current version July 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3006834

Multiscale Residual Attention Network for Distinguishing Stationary Humans and Common Animals Under Through-Wall Condition Using Ultra-Wideband Radar

YANGYANG MA, FUGUI QI, (Graduate Student Member, IEEE), PENGFEI WANG, FULAI LIANG, HAO LV, XIAO YU, ZHAO LI, HUIJUN XUE, JIANQI WANG, AND YANG ZHANG, (Member, IEEE)

Department of Medical Electronics, School of Biomedical Engineering, Fourth Military Medical University, Xi'an 710032, China

Corresponding authors: Yang Zhang (yangzhang@fmmu.edu.cn) and Jianqi Wang (wangjq@fmmu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0810201, in part by the National Natural Science Foundation of China under Grant 31800822, and in part by the Fourth Military Medical University (FMMU) Zhufeng Project under Grant 2018RCFC08.

ABSTRACT Distinguishing between humans and common animals through a wall is necessary for facilitating successful rescue of survivors and enhancing the confidence of rescuers in post-disaster search and rescue operations. However, few existing solutions are available with only dogs considered in this scenario. This poses an issue in ensuring the recognition accuracy involving different animal species. This work proposed a novel multiscale residual attention network for distinguishing between stationary humans and common animals under a through-wall condition based on ultra-wideband radar, which is yet to be performed by existing research using deep learning. Humans, dogs, cats, rabbits, and no target data are collected and distinguished. The overall architecture of the proposed method differed from conventional deep learning methods as it is constructed by parallel 3×3 and 5×5 kernels incorporated with the residual attention learning mechanism. The effect of the slow-time dimension on the classification performance is analyzed, thereby producing an optimal input size. The overall F1-score of the proposed network can reach a high value of 0.9064 and the recognition accuracy of human targets can reach 0.983 satisfying the requirements for post-disaster rescue. Then, the effectiveness and advancement of the three components of the overall network architecture are validated by ablation studies. Finally, the proposed method is compared with three state-of-the-art methods. Comparison results indicate that the proposed method achieve a better performance. The network and its results are envisioned to be applied in various practical situations, such as earthquake rescue and intelligent homecare.

INDEX TERMS Convolutional neural network (CNN), distinguishing between stationary humans and common animals, post-disaster rescue, residual attention learning mechanism, ultra-wideband (UWB) radar.

I. INTRODUCTION

The rapid development in ultra-wideband (UWB) radar life-detection technology [1]–[5] has attracted the interest of researchers in civilian and military applications mainly due to its advantage in penetrability of obstacles, robustness to weather conditions, and protection of visual privacy. Its applications include earthquake and hostage rescue operations [6], gesture recognitions [7], person identification [8], target imaging [9], human tracking [10], etc. In particularly, the distinction between humans and animals using UWB radar is garnering attention as it can obtain significant target information, thereby accurately guiding follow-up operations. W.D Van Eeden *et al.* [11] combined the Gaussian mixture model and hidden Markov model to distinguish slow-moving animals from human targets to detect potential livestock thieves and poachers in nature reserves and farmlands. Wang *et al.* [12] proposed a new parameter,

The associate editor coordinating the review of this manuscript and approving it for publication was Hasan S. Mir.

the respiratory and heartbeat energy ratio (RHER), to classify humans and animals through vital signs monitoring. Björklundet *et al.* [13] used a support vector machine classifier to distinguish between humans and animals for surveillance of approaching organisms at the perimeter of critical infrastructures.

However, these aforementioned works were mainly focused on the moving state or free space condition of a target. To date, there are no reports from other research groups focused on the distinction based on their stationary states and under a through-wall condition using UWB radar. It is crucial to distinguish stationary humans from stationary animals for post-disaster rescue operations where trapped targets are buried and unable to move. In the scenario, some common family pets may cause false alarms when detected as human targets. This will waste valuable research time and rescue resources and may miss opportunities to rescue more trapped survivors. The authors of this study combined twelve handcrafted features with a support vector machine (SVM) model in [14] to distinguish stationary humans and dogs in through-wall conditions. Aside from dogs, cats and rabbits are also common family pets that are likely to cause false alarms in actual post-disaster rescue operations. The proposed features of [14] may not function effectively in recognizing cats and rabbits. As priori and professional knowledge is essential in manual feature extraction, extracting of new handcrafted features to recognize more animal species is rather difficult and time-consuming. These disadvantages limit further development and more practical application of the SVM method. Thus, an automated feature extractor is desired.

Convolutional neural network (CNN) is a deep learning architecture adept at learning embedding from two-dimensional images. CNN is widely employed as an automatic feature extractor due to its accurate and robust performance [15]. With the development of graphic processing units (GPUs) and parallel computing techniques, vast quantities of data can be processed in a short time with CNN. Residual attention network is a new deep learning architecture that is proposed to guide more discriminative feature representations [16]. It is inspired by the brain signal processing mechanism considered unusual by the human vision, which only focuses on significant or interesting information. Similarly, with this architecture, more discriminative features are assigned with more attention.

In this study, a multiscale CNN model combined with residual attention learning is proposed for distinguishing between stationary humans and common animals under a through-wall condition using UWB radar. The radar data are preprocessed and then divided into windows with different widths and overlaps. The window width is first analyzed to obtain its optimal value. Then, the distinguishing performance with different parts of the combined model is compared and further discussed by ablation study. Finally, the comparison with a state-of-the-art method is conducted. To summarize, the main contributions of this study are as follows.

1) A data-set including stationary humans, dogs, cats, rabbits, and no targets collected under a through-wall condition using UWB radar is constructed.
2) A novel multiscale residual attention network is proposed to extract the discriminative features for distinguishing between stationary humans and common animals.
3) Intuitive in-depth analysis of the optimal input size and performance with different parts of the proposed network are performed.
4) The comparison results with the state-of-the-art method validate the effectiveness and advancement of the proposed method in addressing the issue.

The remainder of the paper is organized as follows. Section II illustrates the utilized UWB radar system and signal preprocessing steps. Section III describes the detailed architecture of the novel multiscale residual attention network. Section IV introduces evaluation indicators and implementation details for the network. Experimental results and their corresponding in-depth analysis are in Section V. Section VI provides suggestions for the directions of future research. Finally, the paper is concluded in Section VII.

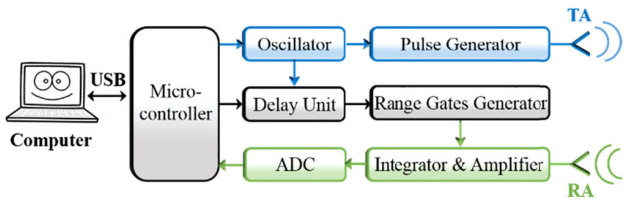


FIGURE 1. Block diagram of the utilized ultra-wideband (UWB) radar. Abbreviations: ADC, Analog-to-Digital Converter; TA, transmitter antenna; RA, receiving antenna; USB, universal serial bus.

II. UWB RADAR SYSTEM AND SIGNAL PREPROCESSING

A. UWB RADAR SYSTEM

The block diagram of the utilized UWB radar is presented in Fig. 1. Firstly, the trigger pulses with a center frequency of 500 MHz and pulse repetition frequency of 128 kHz are generated by the oscillator. The output of the oscillator is fed through an electromagnetic pulse generator to excite the transmitter antenna (TA). Then, the bowtie dipole TA transmits vertically polarized pulses with a peak power of approximately 5 W. Meanwhile, some pulses are sent into a delay unit and range gates generator, where a series of 300-ps wide software-controlled range gates are generated. Next, echoes are received by the receiving antenna (RA) and only the parts within the range gates are collected, integrated, and amplified through the integrator and amplifier. A series of waveforms which include reflected life and range information from media are sampled. Finally, the waveforms are converted to the computer through a high-speed Analog-to-Digital Converter (ADC) for further analysis [17]. The key parameters of the UWB radar are given in Table 1.

After signal acquisition, the raw echo data $D(m, n)$ are stored as waveforms. $m = 1, 2, \cdots, M$ denotes the sampling

TABLE 1. Key parameters of the UWB radar.

Parameters	Values
Center frequency	500 MHz
Bandwidth	500 MHz
Pulse repetition frequency	128 kHz
Sampling points	2048
Scanning speed	64 Hz
Detection distance range	0–3 m

point in propagation time and $n = 1, 2, \cdots, N$ denotes the sampling point in observation time. The time axis associated with range along each received waveform is addressed as "propagation time" and is in the order of nanoseconds. A total of $M = 2048$ sampling points were collected along with the recorded waveform $\tau = 20$ ns long corresponding to a 0–3-m detection range. The scanning speed of 64 waveforms per second along slow-time dimension satisfied the Nyquist sampling rate for respiration. The time axis along the measurement interval is termed as "observation time" or "slow-time" and is described by t in the order of seconds. Hence, the number of recorded waveforms along slow-time is $N = 64 \times T$, with T denoting the overall observation time. Fig. 2 shows the 2-D pseudo-color image of the raw echo data $D(m, n)$ when a male human target is 2.5 m behind a 28-cm thick brick wall with a radar attached on the other side of the wall.

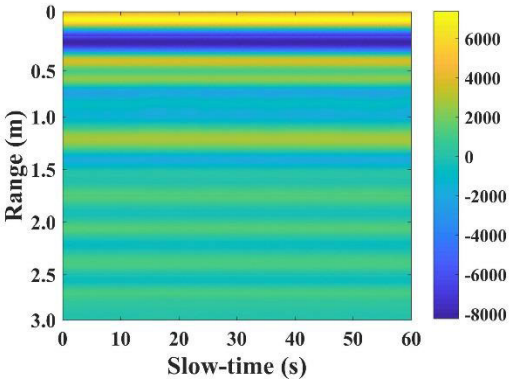


FIGURE 2. 2-D pseudo-color image of the raw echo data with a male human target 2.5 m behind a 28-cm thick brick wall.

B. SIGNAL PREPROCESSING

The signal preprocessing steps are illustrated in Fig. 3. A total of six steps, including range accumulation, normalization, direct current (DC) removal, 2-Hz low-pass (LP) filtering, adaptive filtering, and slow-time accumulation, are implemented. All subsequent model training and testing procedures are based on the preprocessed signals.

Range accumulation can help reduce the computational complexity by averaging the values within a range window along the propagation time index on the premise of

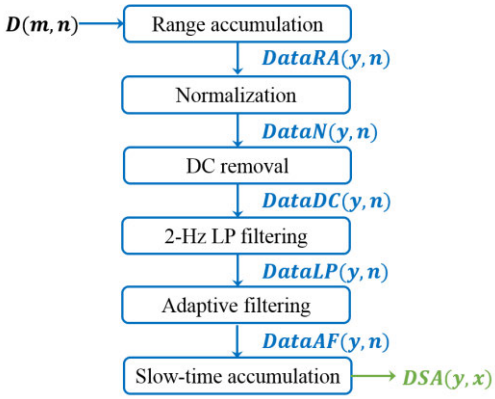


FIGURE 3. Signal preprocessing steps. Abbreviations: DC, direct current; LP, low-pass.

guaranteeing detailed information. It is defined in (1):

$$DataRA(y, n) = \frac{1}{W} \sum_{m=W(y-1)+1}^{Wy} D(m, n), \tag{1}$$

with $DataRA(y, n)$ as the echo data after range accumulation, W as the range window length along the propagation time index, and y as the point number on the propagation time dimension after accumulation. $W = 10$ was chosen here [14], thereby compressing the total number of points along the propagation time index to 200 from 2,048.

Then, normalization along the slow-time index is performed for echo data standardization to accelerate the convergence of the deep learning model as illustrated in (2).

$$DataN(y, n) = \frac{DataRA(y, n) - \min\limits_{1 \le n \le N} [DataRA(y, n)]}{\max\limits_{1 \le n \le N} [DataRA(y, n)] - \min\limits_{1 \le n \le N} [DataRA(y, n)]} \times 2 - 1, \tag{2}$$

with $DataN(y, n)$ denoting the echo signal after normalization.

Next, DC removal is implemented to remove the DC component and baseline drift. It is defined by:

$$DataDC(y, n) = DataN(y, n) - \frac{1}{100} \sum_{n}^{n+99} DataN(y, n). \tag{3}$$

2-Hz LP filtering is then performed to filter out high-frequency noise and retain respiratory signals along the observation time dimension of the echo data. It is determined as:

$$DataLP(y, n) = DataDC(y, n)^{*} h(t), \tag{4}$$

with $h(t)$ as the impulse function of finite impulse response (FIR) filter [18] and $*$ representing convolution operation.

Adaptive filtering is based on the least mean square (LMS) algorithm and is employed to suppress strong clutters. The detailed illustration and verification are provided in [19].

Finally, slow-time accumulation is performed to reduce the computational complexity along the slow-time dimension. It is illustrated by:

$$DSA(y,x) = \frac{1}{Q}\sum_{n=Q(x-1)+1}^{Qx} DataAF(y,n), \tag{5}$$

with $DSA(y,x)$ as the echo signal after slow-time accumulation and Q as the window along the slow-time dimension. $Q = 4$ was chosen, thereby suppressing the original 64-Hz scanning speed to 16 Hz. Fig. 4 shows a 2-D pseudo-color image of the preprocessed echo data when a male human target is 2.5 m behind the 28-cm thick brick wall.

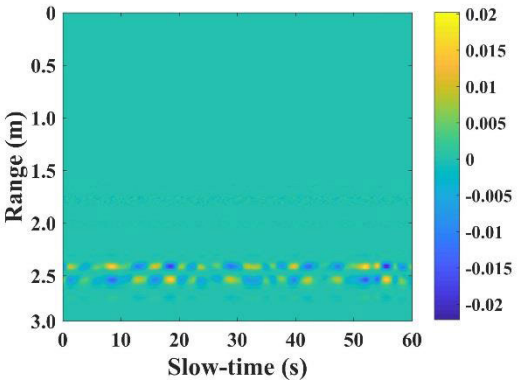


FIGURE 4. 2-D pseudo-color image of the preprocessed echo data when a male human target is 2.5 m behind a 28-cm thick brick wall.

It is noted that the data fed into the subsequent deep learning network have all been preprocessed. The range and slow-time accumulation reduced the data scale by 40 times. The original data scale of D is 2048 × 3712, and the final size after all preprocessing applied of DSA is 200 × 928. Thus, the training procedure of the deep learning model is accelerated. Meanwhile, the number of network layers can be increased with abstract and complex features extracted at higher layers [20].

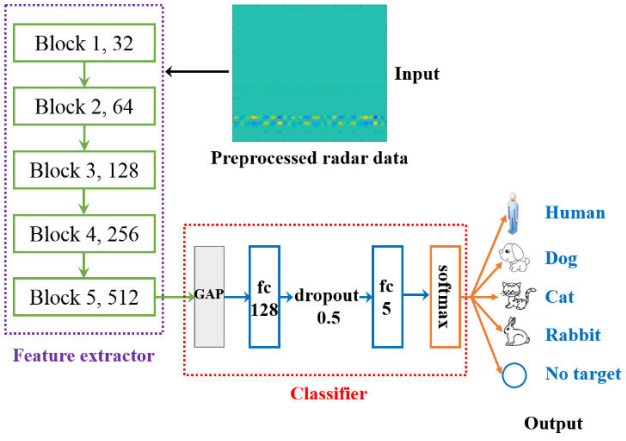


FIGURE 5. Overall structure of proposed multiscale residual attention network. Abbreviations: GAP, global average pooling layer; fc, full connected layer. Block 1, 32 represents that the outputs of block 1 have 32 channels.

III. MULTISCALE RESIDUAL ATTENTION NETWORK

A multiscale CNN model combined with residual attention learning is proposed to distinguish stationary humans from common animals. Its overall architecture is illustrated in Fig. 5. Firstly, the preprocessed radar data are fed into a feature extractor as the network input. The five blocks with specific channel numbers in the feature extractor are used to extract discriminative feature representations. Then, the extracted features are fed into a classifier consisting of a global average pooling (GAP) layer, two fully connected (FC) layers with different node numbers, a dropout layer, and a softmax classifier. Finally, classification results are obtained as the output.

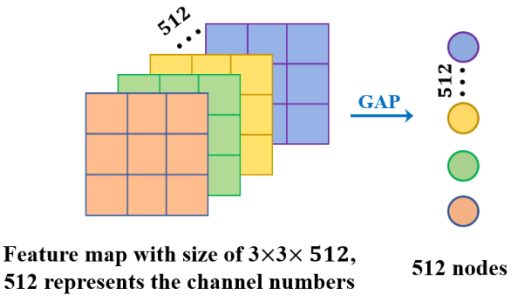


FIGURE 6. Global average pooling (GAP) structure. A 512-channel feature map can be suppressed to a 512-node array. Each node represents the average of a corresponding 3 × 3 channel.

The detailed structure and analysis of the feature extractor will be discussed in Section III.A. and III.B. The GAP in the classifier is a structural regularizer that prevents overfitting for the overall structure [21]. Meanwhile, it can greatly reduce the size of feature maps, which are the outputs of the feature extractor. Particularly, using GAP, a 512-channel feature map can be averaged to an array with only 512 nodes, thereby satisfying the size requirements of the input data for the subsequent FC layers. Each channel is in the form of a matrix, and a node is a specific value. Thus, this is a method for descending dimension of data scale. Fig. 6 illustrates the GAP structure. Given the size of input data fed into the network is (b, y, x, c), where b is the number of radar image within a mini batch, (y, x) is the data scale of each radar image, and c is the channel number. A mini batch contains data that are fed into the network for training or test each time. It is a way to reduce the computational complexity, different from training or testing all data in one time. Then, at the beginning, c is 1, namely the input data size is (b, y, x, 1). After processed by feature extractor, the data size varies to (b, y, x, 512). After GAP, it changes to (b, 1, 512). After the classifier, it changes to (b, 1, 5).

The dropout, which is between two FC layers, is a widely used method to prevent overfitting [22]. The nodes between two FC layers are unlikely to participate in the forward propagation nor the back propagation with a probability of 50%. At the output of the classifier, softmax regression is used to classify the five nodes of the second FC layer into five

categories, namely human, dog, cat, rabbit, and no target. Meanwhile, the corresponding probabilities of the five categories will be calculated and the category with the highest probability value will be regarded as the final classification result.

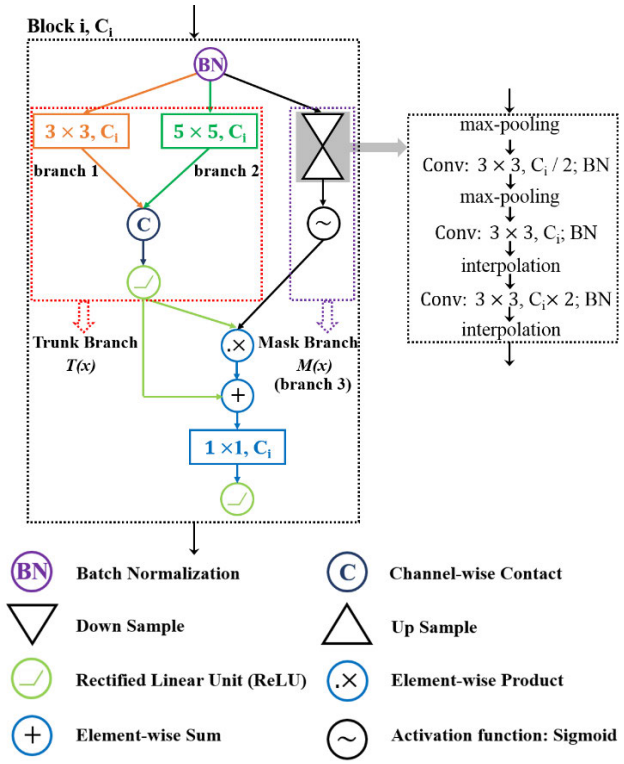


FIGURE 7. Detailed block structure of the multiscale residual attention network. Ci denotes the channel number of the output of block i. Abbreviations: Conv, convolution kernel; BN, batch normalization.

A. MULTISCALE LEARNING STRUCTURE

Multiscale learning structure is used to simultaneously extract features with different granularities. Two types of convolution kernels with different receptive fields are utilized in the network to achieve feature learning with different scales. As shown in Fig. 7, a 3×3 kernel and 5×5 kernel are designed in parallel in the trunk branch. Fine-scale features will be learned with the 3 × 3 kernel in branch 1, while the 5 × 5 kernel is utilized to extract coarse-scale features in branch 2. Each block in feature extractor of Fig. 5 has similar structure as depicted in Fig. 7, except the difference of the specific channel number. An example of the convolution operation of 3 × 3 and 5 × 5 kernels is illustrated in Fig. 8. The stride of the two kernels is both chosen as one (stride 1).

Ma *et al.* [14] proposed twelve handcrafted features belonging to four categories, which are compared when distinguishing stationary human and dog targets. Among the twelve features, some were used to reflect fine changes in radar echo signals, such as the change rate of correlation coefficient of micro vibration (CRCCMV) and mean of wavelet entropy (MWE) [23], while others were envisioned to reflect coarse changes, such as respiratory frequency (RF).

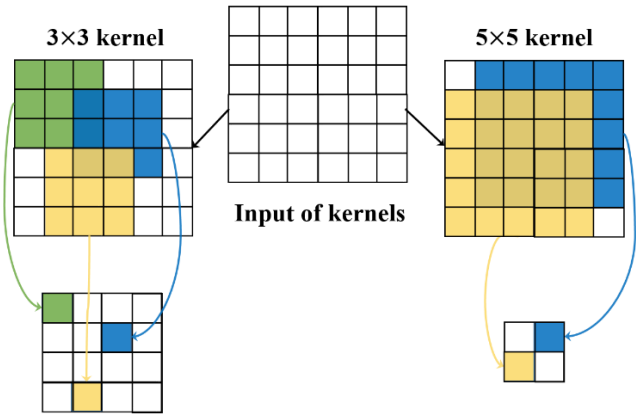


FIGURE 8. Sample convolution operation principles of 3 × 3 kernel and 5 × 5 kernels to learn fine-scale and coarse-scale features, respectively.

Thus, it is necessary to design a structure that can simultaneously learn fine-scale and coarse-scale representations in a CNN network.

The rectified linear unit (ReLU), a kind of activation function, in the trunk branch is a common method to achieve piecewise nonlinear operation in a CNN network [24]. It is defined by:

$$ReLu(x) = \max(0, x) = \begin{cases} x, & if\ x > 0 \\ 0, & if\ x < 0. \end{cases} \tag{6}$$

B. RESIDUAL ATTENTION LEARNING MECHANISM

Residual attention learning mechanism, which can be incorporated with other self-designed feed-forward network architecture in an end-to-end training fashion, is used to generate attention-aware representations [25]. It is inspired by the mechanism of the brain signal processing of unusual subjects of the human vision, in which only a focused location is selected and different representations of the objects at the selected location are enhanced. Four categories including twelve handcrafted features were compared in [14]. The results showed that the classification performance of different categories varied greatly. Thus, residual attention learning mechanism is vital in designing a CNN architecture to assign more attention to features with more discriminative representations.

Residual attention learning mechanism is often designed as a mask branch that uses a bottom-up top-down structure to softly weigh the features learned from the trunk branch [25], [26]. The trunk branch is utilized to implement feature extraction and can be extended to any state-of-the-art network structure. In this paper, the trunk branch is designed as a multiscale learning structure with 3 × 3 and 5 × 5 kernels in parallel as illustrated in Fig. 7. The residual attention mechanism in Fig. 7 is constructed by max-pooling for down-sampling, 3 × 3 convolutional kernels with different channel numbers, and interpolations for up-sampling. The channel number of mask branch output is equal to that of trunk brunch, and the numbers of max-pooling and interpolation modules are the same. Thus, data sizes of outputs of mask branch and

trunk branch are the same and the outputs can be calculated by element-wise sum and product operations. Mixed attention function which uses simple sigmoid for each channel and spatial position without additional restriction is employed in the output of mask branch. It is defined by:

$$f(-v_i, c) = \frac{1}{1 + \exp(-v_i, c)}, \tag{7}$$

with $f(-v_i, c)$ as the mixed attention function, v_i as the feature vector at the i-th spatial position, and c representing the channels.

Also, Fig.7 illustrates the principle of incorporating the trunk branch and mask branch. T(x) denotes the output of the trunk branch and M(x) denotes the output of the mask branch. After element-wise product and sum operations, the output of incorporating the trunk and mask branch can be defined by:

$$I(x) = T(x) \times (M(x) + 1). \tag{8}$$

IV. EVALUATION AND OTHER IMPLEMENTATION DETAILS

A. EVALUATION

Besides accuracy (Acc), recall (R), precision (P), and F1-score (F1) are also computed to evaluate the classification performance of the deep learning network more comprehensively [28], [29]. These parameters are respectively defined in (9)–(12).

$$Acc = \frac{Tp + Tn}{Tp + Tn + Fp + Fn}, \tag{9}$$

$$R = \frac{Tp}{Tp + Fn}, \tag{10}$$

$$P = \frac{Tp}{Tp + Fp}, \tag{11}$$

$$F1 = \frac{R \times P}{R + P} \times 2, \tag{12}$$

where Tp is the number of true positives, Tn is the number of true negatives, Fp is the number of false positives, and Fn is the number of false negatives. A more intuitive illustration of these variables is shown in Fig. 9.

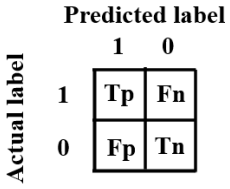


FIGURE 9. Illustration of true positives (*Tp*), true negatives (*Tn*), false positives (*Fp*), and false negatives (*Fn*).

B. OTHER IMPLEMENTATION DETAILS

The multiscale residual attention network used in this study was trained on a GeForce RTX 2080GPU with 11 GB of memory. The CUDA library was utilized for acceleration. An Adaptive Moment Estimation (Adam) optimizer was employed with mini batches of size 8 and L_2 regularization of 0.001. The network was trained for around 300 epochs. The property value of 'padding' is 'SAME' when performing kernel convolution, and this allows the output size of 3×3 and 5×5 kernels in the trunk branch to be the same and to be combined by channel-wise contact operation. The changed learning rate (LR), defined by (13), was then utilized.

$$LR = \begin{cases} 0.001, & if\ epoch \leq 200 \\ 0.0001, & if\ 200 < epoch \leq 300. \end{cases} \tag{13}$$

V. EXPERIMENTAL RESULTS

A. TARGET AND EXPERIMENTAL SETUP DESCRIPTION

This paper aims to distinguish stationary humans and common animals under a through-wall condition using UWB radar. The frequency of common pets in daily life may cause false alarms in post-disaster rescue applications. Thus, dogs, cats, and rabbits were chosen as the animal targets. The radar data of five healthy human targets aged 24 to 43 years old, five grown-up beagle dogs aged approximately 1 year, five healthy domestic cats aged about 28 months, and five New Zealand White rabbits aged approximately 5 months were collected. Table 2 illustrates the detailed information of the aforementioned targets. All animals, except cats, were from the Experimental Animal Center of Fourth Military Medical University. Five cats were voluntarily supplied by the members of our team. It is guaranteed that there was no harm inflicted upon all the animal targets involved during the experiments.

TABLE 2. Detailed information of collected targets.

	Human	Dog	Cat	Rabbit	Total
Target numbers	5	5	5	5	20
Male numbers	3	2	2	2	9
Female numbers	2	3	3	3	11
Average weight (kg)	65	9.43	2.95	2.74	
Average age (years)	29	1	2.4	0.4	
Data samples	5×10	5×10	5×10	5×10	20×10

In the experimental setup, a 28-cm thick brick wall is present between the target and radar. The raw radar echo data were collected by detecting each target approximately 2.5 m away from the brick wall ten times. Each detection was conducted for 58 s. The time interval of every two acquisitions, target distance from the wall, posture facing radar and environmental interference were not identical for each detection. Thus, the raw radar echo signals could be regarded as from different samples with smaller acquisition times, such as 10 times per target. Fig. 10 shows the geometries and photographs of the experimental scenarios.

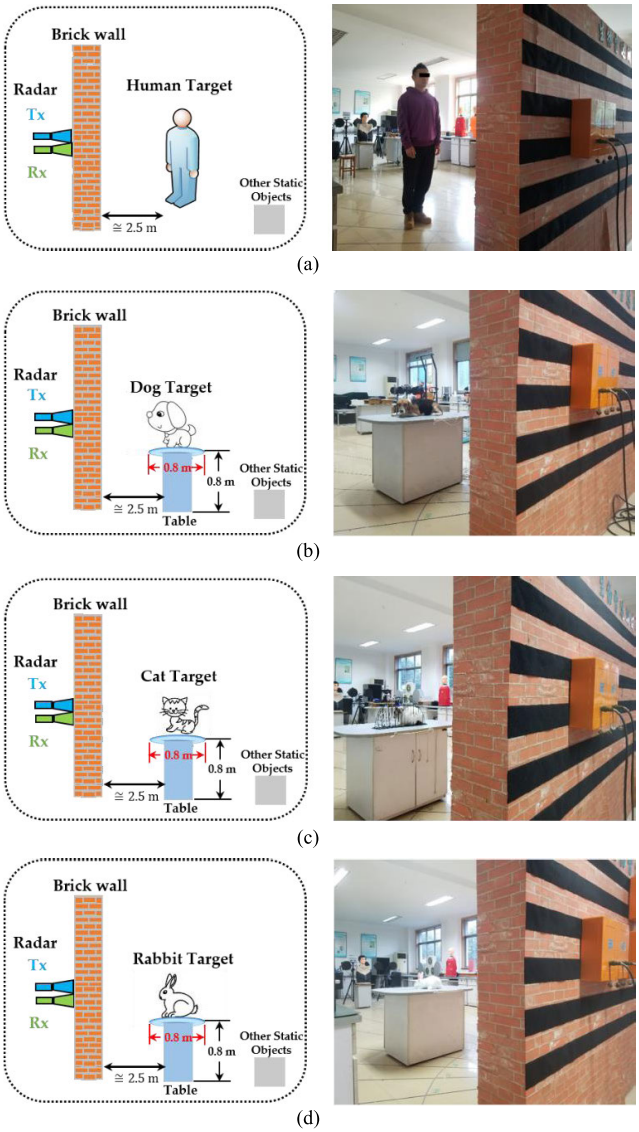


FIGURE 10. The sketch maps and actual measuring photographs of (a) human, (b) dog, (c) cat, and (d) rabbit targets.

During the radar data acquisition of the animals, each target freely lied at a comfortable position until they could maintain a fixed position quietly for over one minute on the experimental table, which is 0.8-m wide, 0.8-m high, and 2.5 m away from the brick wall. Thus, animals may lay with random postures and face different directions in each data acquisition. Similarly, the human target faced different directions during each data acquisition trials. A cat cage was provided during the measuring experiments for the cats as it was infeasible to wait for them to lay quietly for over one minute. The cage fence is spaced at approximately 3.5 cm, thereby still permitting cats to move freely in the fixed space. This would allow cats to reach their stationary state faster to satisfy the requirements of the experimental setup. The measuring intervals for each human target ranged from 1 to 30 min while those of each animal ranged from 1 min to 5 days. Therefore, there were large differences among the radar data of similar target due to significant variations of the target states and environmental interference during each trial.

No-target scenarios are also important and often appear in real rescue situations, thereby possibly causing false alarm in discerning human targets in post-disaster rescue operations. Thus, similar to target trials, 50 no target samples were collected. The corresponding sketch map and actual measuring photograph of these trials are shown in Fig. 11.

FIGURE 11. Geometry and photograph of radar data acquisition with no targets.

B. ANALYSIS OF SLOW-TIME DIMENSION

It is of vital importance to scientifically design relevant input dimensions during the feature extracting process in deep learning network [27]. An optimal range window (ORW) along the propagation time dimension, where the target signal is in the middle, was set to be 31-point long in [14]. The corresponding classification results in [14] showed ORW length yielded great performance. Thus, considering the even input size must be in the residual attention learning mechanism, a range window with a length of 32 points along the propagation time dimension with one point at the window bottom was chosen for this study. Then, the key turns to choosing the appropriate time window (TW) along the slow-time dimension with the best multi-classification performance.

When comparing the classification performance with TWs of different widths, it is crucial to contain at least one respiratory period per TW. If TW width is less than the length of one respiratory period, the information and representations in the TW will be insufficient. Thus, TW widths ranging from 4 to 16 s with 2-s intervals step were chosen. After signal preprocessing, the corresponding sampling frequency becomes 16 Hz (waveforms per second). Hence, in a TW, there are 64 waveforms when the width is 4 s and 256 waveforms when the width is 16 s. To take full advantage of the 58-second per continuous radar echo signal with duration, overlaps of consecutive waveforms between adjoining TWs are commonly used [16], [27]. The relationship between TW width and overlaps in this research is defined in (14).

$$O_w = \frac{TW \times 20 - 58 \times 16}{20 - 1}, \tag{14}$$

where TW is the width of the time window and O_w is the number of waveforms per overlap. A total of 20 TWs are involved in each preprocessed radar echo signal. Thus, there will be a total of 5000 samples belonging to 250 preprocessed radar echo signals of five target types. The distinguishing

results can be higher than those in the case of non-overlapping windows of adjoining cases due to the overlap of training and testing windows. The schematic overview of the analysis is depicted in Fig. 12.

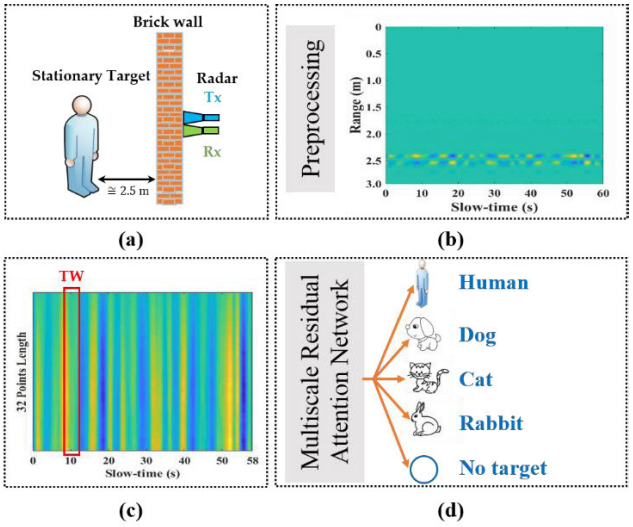


FIGURE 12. Schematic overview for analysis of slow-time dimension. (a) Raw radar data acquisition for each target. (b) Preprocessing of raw radar data. (c) Radar data after preprocessing are reduced to fragments with different TW widths along the slow-time dimension. Each fragment has 32 points length along the propagation time dimension. (d) The fragments are fed into a multiscale residual attention network, which predicts probabilities for each target type.

To prevent overfitting, five-fold cross-validation was employed. It is implemented by firstly dividing the 5000 samples randomly into five copies with equal number. Then, four of them are used as training set and the remaining copy is used as the test set per time, until each data copy is used as test data. The comparison results of different TWs using five-fold cross-validation are shown in Fig. 13 where a 12-s wide TW has been found to have the optimal classification performance with an F1-score of 0.9064.

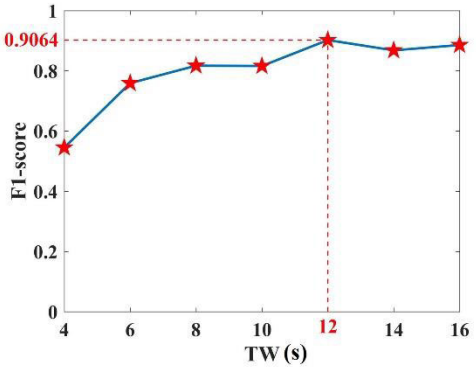


FIGURE 13. Comparison results with different TWs. Abbreviations: TW, time window (TW) along the slow-time dimension. A 12-s wide TW has the optimal classification performance.

When the TW width is small, such as 4 s, the TW period is short with much less corresponding information. Thus, the difference between TWs is insignificant producing much lower F1-scores. Relatively, when the TW width is large, such as 16 s, of course the TW period is long enough with more corresponding information. However, the width of the overlaps and TWs will be much closer. When TW is 16-s wide, the calculated overlaps obtained from (13) will be 220 waveforms, which is equivalent to 13 s. Therefore, the difference between the contiguous TWs will be much smaller, thereby having much lower F1-scores. When the TW is 12-s wide, a compromise will be reached, thereby obtaining the highest F1-score.

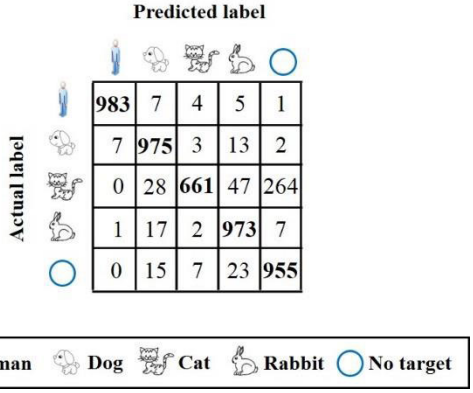


FIGURE 14. Confusion matrix of distinguishing result when the width of TW is 12 s.

Fig. 14 illustrates the confusion matrix of the distinguishing result when the TW is 12-s wide. The corresponding key evaluation indicators are shown in Tables 3 and 4.

TABLE 3. Overall F1-score and accuracy when the width of TW is 12 s.

Indicators	Values
F1-score	0.9064
Accuracy	0.9094

TABLE 4. Recall and precision of targets when the width of TW is 12 s.

	Human	Dog	Cat	Rabbit	No target
Recall	**0.9830**	0.9750	0.6610	0.9730	0.9550
Precision	**0.9919**	0.9357	0.9764	0.9171	0.7771

The overall F1-score and accuracy obtained high values of 0.9064 and 0.9094, respectively. The recall and precision of human targets reached 0.983 and 0.9919, respectively, satisfying the requirements of post-disaster rescue applications. A significantly high distinguishing rate of human targets will be beneficial for improving the success rate of rescuing survivors and enhancing the confidence of rescuers. The recall of cat was relatively low with a value of 0.661 due to oversight of 264 of 1,000 samples, which were distinguished as no targets, as presented in Fig. 14. Moreover, the iron material of the specific cat cage used in the cat's experiments to limit their movement in a fixed space and accelerate the

experiments affected the propagation of the UWB radar signals, thereby weakening the signals. Therefore, the signals of cats and no targets exhibited a degree of similarity that lead to the misjudgment. This reduced the overall F1-score and accuracy of the multi-classification task. However, it did not affect the identification of human targets, which still conformed to the original goal for the post-disaster rescue of survivors.

C. PERFORMANCE VALIDATION OF THE NETWORK ARCHITECTURE

Ablation studies are commonly chosen for understanding the contribution of various components when different structures are combined into a network [16], [30]. It is typically performed by removing some structure of the overall network and then assessing how the performance is affected. 3×3 kernels, 5×5 kernels, and residual attention learning mechanism are integrated to construct a novel multiscale residual attention network in this study. Thus, ablation studies are utilized to analyze and validate the performance of these incorporated architectures. The inputs are the 5000 samples with 32-point length in the propagation time dimension and 12-s wide in slow-time dimension, which has been validated to yield the optimal performance. Five-fold cross-validation is also used here. The results of the corresponding ablation studies of the novel multiscale residual attention network are shown in Table 5. F1-score is chosen to be the evaluation indicator.

TABLE 5. Results of the ablation studies of the multiscale residual attention network.

3×3 Kernels	✓	✓		✓	✓	
5×5 Kernels	✓	✓	✓			✓
Residual Attention Learning Mechanism	✓		✓	✓		
F1-score	**0.906**	0.812	0.874	0.830	0.723	0.621

The network combining all three structures has the optimal performance with the highest F1-score in Table 5. The F1-score of the multiscale structure with parallel 3×3 and 5×5 kernels parallel is 0.812. Thus, by solely implementing a multiscale structure, the network performance cannot be significantly improved. When the network has no residual attention learning mechanism, a larger decrease in the F1-score is observed than the absence of 3×3 kernels or 5×5 kernels, thereby illustrating its more important role in the overall architecture. When there are only independent 3×3 kernels or 5×5 kernels, the F1-score decreased to the lowest. The above results validated that a network simultaneously combining the three structures in parallel best satisfied the circumstance in distinguishing stationary humans from common animals under a through-wall condition using UWB radar. Moreover, the architecture effectiveness and advancement of the multiscale residual attention network is verified.

D. COMPARISON WITH STATE-OF-THE-ART METHOD

1) COMPARISON WITH THE SVM METHOD

To the best of our knowledge, only the team of [14] is dedicated to research on novel and practical methods of distinguishing between stationary humans and animals under a through-wall condition using UWB radar for post-disaster rescue applications. Thus, the method in [14] is chosen as the baseline for the comparison with a state-of-the-art method. In [14], twelve handcrafted features belonging to four categories were combined by the SVM method. These features are listed in Table 6.

TABLE 6. Twelve handcrafted features in [14].

Categories	Handcrafted Features
Energy-corresponding features	Standard deviation change rate of micro vibration (StdCRMV)
	Energy ratio of the reference frequency band (ERRFB)
Correlation coefficient-corresponding features	Optimal Correlation Coefficient of Micro vibration (OCCMV)
	Change rate of correlation coefficient of micro vibration (CRCCMV)
Wavelet entropy-corresponding features	Mean of wavelet entropy of target signal (MWE)
	Standard deviation of wavelet entropy of target signal (StdWE)
	Mean of MWE in the OW window (MMWEOW)
	Ratio of wavelet entropy (RWE)
Frequency-corresponding features	$f_{1/4}$
	$f_{3/4}$
	Width between$f_{1/4}$ and $f_{3/4}$ (WOHMS)
	Respiratory Frequency (RF)

According to the requirements of the feature extraction procedures in [14], $DataAF(y, n)$ are the outputs for the signal preprocessing procedure, as shown in Fig. 3. In order to maintain the correspondence with the data dividing method in the comparison, five-fold cross-validation is used in dividing the 250 samples of $DataAF(y, n)$. This is performed by firstly randomly dividing all samples into five copies with equal numbers. Then, four of them are used as the training set while the remaining copy is used as the test set each time until each data copy is used as the test set. The comparison results are shown in Table 7 and Fig. 15.

TABLE 7. Performance comparison results.

	F1-score	Accuracy
SVM with 12 Handcrafted Features	0.8140	0.8080
Multiscale Residual Attention Network	0.9064	0.9094

The F1-score and accuracy of the multiscale residual attention network were 0.0924 and 0.1014, respectively, higher

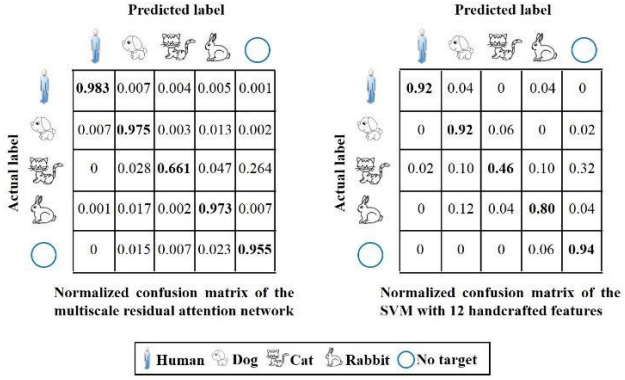


FIGURE 15. Comparison results of normalized confusion matrix between the multiscale residual attention network and SVM method with 12 handcrafted features.

than that of the SVM with 12 handcrafted. Thus, this validated the ability of the network to significantly enhance multi-classification performance. The lower performance of the SVM may be attributed to its handcrafted features, a majority of which were designed for distinguishing between stationary humans and dogs, thereby making them unfit for distinguishing between humans, dogs, cats, rabbits, and no target.

From the normalized comparison results of the confusion matrix in Fig. 15, the proposed network has better distinguishing rates for all kinds of targets than those of SVM. However, for both methods, the distinguishing rate of cats is very low (0.661 and 0.46 respectively), which can be due to the iron material of the cat cage that affected the propagation of the electromagnetic waves.

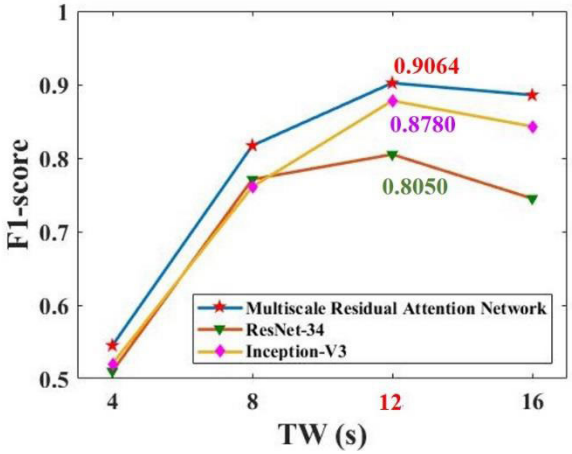


FIGURE 16. Performance comparison with two CNN models on test set.

2) COMPARISON WITH OTHER CONVOLUTIONAL NEURAL NETWORKS

To further validate the novelty and advancement of the multiscale residual attention network, two CNN models are chosen for comparison. They are Inception-v3 [31], [32], and ResNet-34 [33], respectively. The above two models are relatively new and have been demonstrate to have excellent performance toward to many kinds of issues. The data for comparison are the same as those used in section V.B when analyzing impact of slow-time dimension. The performance comparison results are shown in Fig.16. TW widths of 4 s, 8 s, 12 s, and 16 s are analyzed. Performance of all the three models have similar trends as the TW width changes. They all obtain the highest F1-score when the TW is 12-s wide, because a compromise will be reached at this point as illustrated in Section V.B. The multiscale residual attention network can get a highest F1-score value of 90.64%, 2.84% higher than that of Inception-v3 model, and 10.14% higher than that of ResNet-34. These results indicate that the proposed multiscale residual attention network has better performance and novelty.

VI. DISCUSSION

The proposed multiscale residual attention network provided an outstanding performance in distinguishing stationary humans and common animals under a through-wall condition using UWB radar. In particular, it outperformed the state-of-the-art method using SVM with 12 handcrafted features in [14]. Notably, there were no reports from other research groups that employed deep learning in distinguishing stationary humans and animals under a through-wall condition. There are still experimental setup-corresponding issues that may be further considered for more practical applications. For instance, in the following in-depth research, more realistic scenario of search and rescue will be researched and more investigation of the effect of changing locations, objects, clutter, between the wall and the target will be implemented.

VII. CONCLUSION

This study addressed the issue of distinguishing stationary humans, dogs, cats, rabbits, and no targets under a through-wall condition using UWB radar by proposing a novel multiscale attention network. This study is the first to use a deep learning method in this application. 3 × 3 kernels, 5 × 5 kernels, and a residual attention learning mechanism are integrated parallel to construct the overall architecture. In the experimental setup, all targets are about 2.5 m from a 28-cm thick brick wall. Then, an optimal input size is analyzed and chosen to feed into the network. Input with a 32-point long range window along the propagation time dimension and 12-s wide TW along the slow-time dimension are validated to exhibit the best multi-classification performance with an F1-score and accuracy of 0.9064 and 0.9094, respectively. Particularly, the recall of human targets can reach a high value of 0.983, which satisfies the requirement for rescuing survivors in post-disaster scenarios. Next, the performances of the three parallel components of the overall network architecture are analyzed by ablation studies, thereby validating its effectiveness and superiority. Finally, the multiscale residual attention network is compared with three start-of-the-art methods, including an SVM method that extracted twelve handcrafted features, two new CNN models of Inception-v3 and ResNet-34. The F1-score of the proposed network is 9.24% higher than that of the SVM, 2.84% higher than

that of the Inception-v3 model, and 10.14% higher than that of the ResNet-34 model. These comparison results indicate that the proposed method is more suitable for the scientific problem of this paper. Simultaneously, the effectiveness and advancement of the multiscale residual attention network are validated. In summary, the proposed multiscale residual attention network satisfied the requirement for rescuing applications more practically with the reduced need for advanced professional knowledge owing to its automated and exceptional feature extractor. We envision the implementation of our proposed method in practical post-disaster rescue applications, such as earthquake rescue missions

ACKNOWLEDGMENT

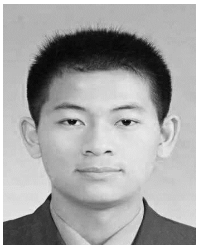
The authors would like to thank Editage (www.editage.cn) for English language editing. *(Yangyang Ma and Fugui Qi are co-first authors.)*

REFERENCES

[1] C. Li and J. Lin, *Microwave Noncontact Motion Sensing and Analysis*. Hoboken, NJ, USA: Wiley, 2013.
[2] Y. Zhang, F. Qi, H. Lv, F. Liang, and J. Wang, "Bioradar technology: Recent research and advancements," *IEEE Microw. Mag.*, vol. 20, no. 8, pp. 58–73, Aug. 2019.
[3] J. D. Taylor, *Ultra-Wideband Radar Technology*. Boca Raton, FL, USA: CRC Press, 2000.
[4] H. Hong, L. Zhang, H. Zhao, H. Chu, C. Gu, M. Brown, X. Zhu, and C. Li, "Microwave sensing and sleep: Noncontact sleep-monitoring technology with microwave biomedical radar," *IEEE Microw. Mag.*, vol. 20, no. 8, pp. 18–29, Aug. 2019.
[5] G. Gennarelli, G. Ludeno, and F. Soldovieri, "Real-time through-wall situation awareness using a microwave Doppler radar sensor," *Remote Sens.*, vol. 8, no. 8, p. 621, Jul. 2016.
[6] S. Wu, K. Tan, J. Chen, Z. Xia, F. Guangyou, and S. Meng, "Improved human respiration detection method via ultra-wideband radar in through-wall or other similar conditions," *IET Radar, Sonar Navigat.*, vol. 10, no. 3, pp. 468–476, Mar. 2016.
[7] S.-J. Ryu, J.-S. Suh, S.-H. Baek, S. Hong, and J.-H. Kim, "Feature-based hand gesture recognition using an FMCW radar and its temporal feature analysis," *IEEE Sensors J.*, vol. 18, no. 18, pp. 7593–7602, Sep. 2018.
[8] Y. Lang, Q. Wang, Y. Yang, C. Hou, H. Liu, and Y. He, "Joint motion classification and person identification via multitask learning for smart homes," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9596–9605, Dec. 2019.
[9] Y. Sasaki, F. Shang, S. Kidera, T. Kirimoto, K. Saho, and T. Sato, "Three-dimensional imaging method incorporating range points migration and Doppler velocity estimation for UWB millimeter-wave radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 122–126, Jan. 2017.
[10] S. Chang, M. Wolf, and J. W. Burdick, "Human detection and tracking via ultra-wideband (UWB) radar," in *Proc. IEEE Int. Conf. Robot. Automat.*, Anchorage, AK, USA, May 2010, pp. 452–457.
[11] W. D. van Eeden, J. P. de Villiers, R. J. Berndt, W. A. J. Nel, and E. Blasch, "Micro-Doppler radar classification of humans and animals in an operational environment," *Expert Syst. Appl.*, vol. 102, pp. 1–11, Jul. 2018.
[12] P. Wang, Y. Zhang, Y. Ma, F. Liang, Q. An, H. Xue, X. Yu, H. Lv, and J. Wang, "Method for distinguishing humans and animals in vital signs monitoring using IR-UWB radar," *Int. J. Environ. Res. Public Health*, vol. 16, no. 22, p. 4462, Nov. 2019.
[13] S. Bjorklund, T. Johansson, and H. Petersson, "Target classification in perimeter protection with a micro-Doppler radar," in *Proc. 17th Int. Radar Symp. (IRS)*, Kraków, Poland, May 2016, pp. 1–5.
[14] Y. Ma, F. Liang, P. Wang, H. Lv, X. Yu, Y. Zhang, and J. Wang, "An accurate method to distinguish between stationary human and dog targets under through-wall condition using UWB radar," *Remote Sens.*, vol. 11, no. 21, p. 2571, Nov. 2019.
[15] X. Li, Y. He, and X. Jing, "A survey of deep learning-based human activity recognition in radar," *Remote Sens.*, vol. 11, no. 9, p. 1068, May 2019.
[16] Y. He, X. Li, and X. Jing, "A mutiscale residual attention network for multitask learning of human activity using radar micro-Doppler signatures," *Remote Sens.*, vol. 11, no. 21, p. 2584, Nov. 2019.
[17] H. Lv, G. H. Lu, X. J. Jing, and J. Q. Wang, "A new ultra-wideband radar for detecting survivors buried under earthquake rubbles," *Microw. Opt.*, vol. 52, no. 11, pp. 2621–2642, Aug. 2010.
[18] Y. Wang, "Study on the technology of distinguishing between humans and animals via UWB bio-radar," Ph.D. dissertation, Dept. Biomed. Eng., Fourth Mil. Med. Univ., Xi'an, China, 2014.
[19] Z. Li, W. Li, H. Lv, Y. Zhang, X. Jing, and J. Wang, "A novel method for respiration-like clutter cancellation in life detection by dual-frequency IR-UWB radar," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 5, pp. 2086–2092, May 2013.
[20] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
[21] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*. [Online]. Available: http://arxiv.org/abs/1312.4400
[22] P. Cao, W. Xia, M. Ye, J. Zhang, and J. Zhou, "Radar-ID: Human identification based on radar micro-Doppler signatures using deep convolutional neural networks," *IET Radar, Sonar Navigat.*, vol. 12, no. 7, pp. 729–734, Jul. 2018.
[23] Y. Wang, X. Yu, Y. Zhang, H. Lv, T. Jiao, G. Lu, W. Z. Li, Z. Li, X. Jing, and J. Wang, "Using wavelet entropy to distinguish between humans and dogs detected by UWB radar," *Prog. Electromagn. Res.*, vol. 139, pp. 335–352, Jan. 2013.
[24] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
[25] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6450–6458.
[26] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 640–651.
[27] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, "Indoor person identification using a low-power FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3941–3952, Jul. 2018.
[28] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Advances in Information Retrieval*. Berlin, Germany: Springer, 2005, pp. 345–359.
[29] H. Chen, D. Ni, J. Qin, S. Li, X. Yang, T. Wang, and P. A. Heng, "Standard plane localization in fetal ultrasound via domain transferred deep neural networks," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 5, pp. 1627–1636, Sep. 2015.
[30] M. Hessel, J. Modayil, H. van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar, and D. Silver, "Rainbow: Combining improvements in deep reinforcement learning," in *Proc. AAAI*, San Francisco, CA, USA, Feb. 2017, pp. 1–14.
[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014, *arXiv:1409.4842*. [Online]. Available: https://arxiv.org/abs/1409.4842
[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. CVPR*, Jun. 2016, pp. 2818–2826.
[33] 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.

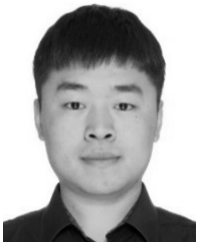
YANGYANG MA received the B.E. degree from Fourth Military Medical University (FMMU), Xi'an, Shaanxi, China, in 2018, where he is currently pursuing the M.E. degree in machine learning and bio-radar technology.

He is also with the Bio-radar and Signal Processing Laboratory, Department of Bio-medical Engineering, FMMU. His research interests include classification of humans and animals under through-wall condition and bio-radar-based gesture recognition.



FUGUI QI (Graduate Student Member, IEEE) received the B.E. degree from Fourth Military Medical University (FMMU), Xi'an, China, in 2014, where he is currently pursuing the Ph.D. degree in BME.

From 2019 to 2020, he was a Visiting Scholar with the University of Tennessee. He is currently with the Bio-radar and Signal Processing Laboratory, Institute of Bio-medical engineering, FMMU. His research interests include through-wall detection and classification of human motion.



PENGFEI WANG received the B.E. degree from the Harbin Institute of Technology, Harbin, Heilongjiang, in 2013, and the M.E. degree from the PLA University of Science and Technology, Nanjing, Jiangsu, in 2016. He is currently pursuing the Ph.D. degree with the Department of Biomedical Engineering, Fourth Military Medical University, Xi'an, Shaanxi, China.

His research interest includes modeling, extraction, and discrimination of physiological signals.

FULAI LIANG received the B.S., M.S., and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha, Hunan, China, in 2006, 2008, and 2013, respectively.

He is currently an Associate Professor with Fourth Military Medical University. His research interests include modeling, signals processing, multiple human targets imaging, and detection and positioning.

HAO LV received the B.E., M.S., and Ph.D. degrees from Fourth Military Medical University (FMMU), Xi'an, Shaanxi, China, in 2004, 2007, and 2010, respectively.

Since 2010, he has been teaching with the School of Biomedical Engineering, FMMU, where he has been an Associate Professor, since 2016. His research interests include biomedical signal processing and bio-radar based imaging.

XIAO YU received the B.S., M.S., and Ph.D. degrees from the Department of Bio-medical Engineering, Fourth Military Medical University (FMMU), Xi'an, Shaanxi, China, in 2003, 2006, and 2012, respectively.

He is currently a Lecturer with FMMU. His research interests include distinguishing between humans and animals-based on bio-radar and signal processing.

ZHAO LI received the B.E., M.S., and Ph.D. degrees from the Department of Electronics, School of Biomedical Engineering, Fourth Military Medical University, Xi'an, Shaanxi, China, in 2007, 2010, and 2013, respectively.

Since 2013, he has been working as a Lecturer with the Department of Electronics, School of Biomedical Engineering, Fourth Military Medical University. His research interests include biomedical signal processing, ultra-wideband-based human detection, and detection of weak vital signs under complex environment.

HUIJUN XUE received the B.S. and M.S. degrees from Shaanxi Normal University, in 2007 and 2010, respectively, and the Ph.D. degree in biophysics from Fourth Military Medical University (FMMU), Xi'an, Shaanxi, China, in 2018.

Since 2014, she has been a Lecturer with FMMU. Her research interest includes detection and processing of non-contact biomedical signals.

JIANQI WANG was born in Xi'an, Shaanxi, China, in 1962. He received the B.E. degree from Xi'an Jiaotong University, Xi'an, in 1984, the M.E. degree from the National University of Defense Technology, Changsha, China, in 1990, and the Ph.D. degree from the Key Laboratory of the Ministry of Education of China, Xi'an Jiaotong University, in 2006.

Since 1990, he has been teaching with the School of Biomedical Engineering, Fourth Military Medical University (FMMU), Xi'an. He is currently a Professor and the Director of the Department of Medical Electronics, School of Biomedical Engineering, FMMU. He pioneered in radar-based human targets detection in China, in 1998. He has published more than 100 articles corresponding to the technologies. His research interests include bio-radar-based signal processing, human targets detection, imaging, and machine learning.

YANG ZHANG (Member, IEEE) received the B.E., M.E., and Ph.D. degrees from the Department of Biomedical Engineering, Fourth Military Medical University, Xi'an, Shaanxi, China, in 2001, 2006, and 2009, respectively.

From 2013 to 2014, he was a Visiting Scholar with the Department of Computer Science, Texas Tech University, Lubbock, TX, USA. He is currently an Associate Professor with the Bio-radar and Signal Processing Laboratory, Fourth Military Medical University. His research interests include RF sensing, radar biomedical signal processing, target detection and positioning, and machine learning.