Non-motorized lane target behavior classification based on millimeter wave radar with P-Mrca convolutional neural network

Jiaqing He, Yihan Zhu, Bing Hua, Zhihuo Xu, *Senior Member, IEEE*, Yongwei Zhang, *Member, IEEE*, Liu Chu, *Member, IEEE*, Quan Shi, *Member, IEEE*, Robin Braun, *Life Senior Member, IEEE*, Jiajia Shi, *Member, IEEE*

***Abstract*—In the fields of road regulation and road safety, the classification of target behaviors for non-motorized lanes is of great significance. However, due to the influence of adverse weather and lighting conditions on the recognition efficiency, we use radar to perform target recognition on non-motorized lanes to cope with the challenges caused by frequent traffic accidents on non-motorized lanes. In this paper, a classification and recognition method for non-motorized lane target behavior is proposed. Firstly, a radar data acquisition system is constructed to extract the micro-Doppler features of the target. Then, in view of the shortcomings of traditional deep learning networks, this paper proposes a multi-scale residual channel attention mechanism that can better perform multi-scale feature extraction and adds it to the convolutional neural network (CNN) model to construct a multi-scale residual channel attention network (MrcaNet), which can identify and classify target behaviors specific to non-motorized lanes. In order to better combine the feature information contained in the high-level features and the low-level features, MrcaNet was combined with the feature pyramid structure, and a more efficient network model feature pyramid-multi-scale residual channel attention network (P-MrcaNet) was designed. The results show that the model has the best scores on classification indexes such as accuracy, precision, recall rate, F1 value and Kappa coefficient, which are about 10% higher than traditional deep learning methods. The classification effect of this method not only performs well on this paper's dataset, but also has good adaptability on public datasets.**

***Index Terms*—attention mechanism, feature pyramid structure, gait recognition, micro-Doppler signatures, MrcaNet, multi-scale residual channel attention, P-MrcaNet.**

I. Introduction

NON-motorized lanes are an important part of the urban low-carbon transportation system and play an increasingly significant role in urban green travel. The classification and identification of non-motorized lane target behaviors can enable road supervision, improve traffic efficiency, and ensure road safety. This has become a research direction attracting much attention. The current urgent problem is the lack of effective monitoring methods on non-motorized lanes, making traffic management difficult. This situation can cause traffic congestion, increase the risk of accidents, and endanger the safety of people with reduced mobility [1]. Traditional methods are mostly based on video and infrared images, which have limited effectiveness under conditions such as low light or extremely strong light, and cannot work in hazy weather. Additionally, because video images are easily blocked, especially at turning intersections and non-motor vehicle intersection areas, the recognition accuracy is low, affecting the monitoring effect [2]. Radar has the advantage of not being affected by weather and occlusion and can stably measure and perceive the traffic environment all day long. It is an important supplement to the video image method and can achieve reliable monitoring of non-motorized lanes with a small increase in cost. As machine learning methods such as deep learning have made significant progress, there is increasing attention to the use of micro-Doppler [3],[4], sensor fusion [5], pattern recognition [6], and other technologies to achieve the identification and classification of road targets.

A. Motivation and Contribution

The above article does achieve a good effect on the multi-angle classification task. However, at different scales, the objects in the image may have different features, and the traditional method may only focus on the features on the fixed scale, while ignoring the information on other scales. This also limits the use of these methods in certain applications that require a high degree of interpretability. To solve these problems, this paper proposes a deep learning network architecture, P-MrcaNet, to better integrate multi-scale features. This can identify and classify Non-motorized lane specific target behaviors, strengthen the supervision of turning junctions and non-motor vehicle crossing areas, and thus reduce accident rates. Our contributions can be summarized as follows:

This work was supported in part by the National Natural Science Foundation of China under Grant 12102203, 62174091, 62201294, 61901235, the Natural Science Foundation of Jiangsu Province under Grant BK20231336, BK20200971, the Fire and Rescue Bureau Research Program under Grant 2019XFCX31, and the Postgraduate Research & Practice Innovation Program of School of Transportation and Civil Engineering, Nantong University NTUJTXYGI2301.

Jiaqing He, Yihan Zhu, Zhihuo Xu, Yongwei Zhang, Quan Shi, Jiajia Shi are with the School of Transportation and Civil Engineering, Nantong University, China.

Bing Hua is with the Nantong Fire Rescue Detachment, Nantong, Jiangsu, China.

Liu Chu is with the Center for Transformative Science, ShanghaiTech University, Shanghai, China

Robin Braun is with the School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW, Australia.

Corresponding Author: Jiajia Shi (e-mail: shijj@ntu.edu.cn).

1) This paper proposes a new network model, MrcaNet, which introduces a multi-scale residual attention mechanism into the model and searches for the optimal multi-scale attention mechanism through experiments.
2) In order to better combine the feature information contained in the high-level features and the low-level features in the operation of the network, MrcaNet is combined with the feature pyramid structure, and the network model P-MrcaNet with high accuracy is designed.
3) We have analyzed data from various perspectives by comparing the accuracy, precision, recall rate, F1 score, Kappa coefficient, and other classification performance measures of the data confusion matrix. Based on these measures, we conclude that the new model has the most effective classification performance.
4) The targets to be identified are not limited to simple gaits but also include more complex human activities, such as hobbling on crutches, hopping on one leg, riding a bicycle, etc. Compared with repetitive and single human activities, the complex activities identified in this article can better reflect people's intentions and have practical significance.

II. Related works

A. Explainable Gait Recognition

Pedestrian gait recognition technology can be used for pedestrian behavior analysis and detection [7],[8],[9]. Through the recognition and analysis of pedestrian gait, the pedestrian's walking speed, direction, and intention can be detected, which helps to improve the safety and convenience of urban traffic. As highlighted by Komura et al. [10], physiological effects can be attached to motion-captured data, providing deeper insights into human movement dynamics. Additionally, Chan et al. [11] developed a virtual reality dance training system using motion capture technology, showcasing the potential for immersive and interactive applications in this field. Moreover, Sandilands et al. [12] explored interaction capture using magnetic sensors, contributing valuable insights into capturing subtle nuances of human interaction through advanced sensor technologies. These works collectively underscore the significance of integrating motion capture techniques with gait analysis.

Research on Human Activity Recognition (HAR) has made significant progress in the past decade. Human activity classification methods based on radar micro-Doppler data usually extract physically interpretable features from the time-velocity domain, and use them for classification [13],[14]. Li et al. [15] elaborated on the application of radar systems and deep learning technology in detecting human behavior, and summarized corresponding deep learning algorithms for different types of radar echo forms. Ni et al. [16] used the micro-Doppler characteristics of human gait captured by millimeter-wave radar to build a micro-Doppler Gait (Mgait) system that can realize indoor multi-person identity recognition and intrusion detection with an accuracy of up to 88.59%. Bai et al. [17] proposed an effective method to generate radar echoes using infrared public motion capture datasets. Experiments show that the proposed method achieves higher recognition accuracy in fine human gait classification and can be easily It can be extended to the fine recognition of other human activities. Li et al. [18] proposed a high-precision and efficient human activity classification method based on radar micro-Doppler features, data enhancement, and deep neural networks. Experimental results show that the optimal parameters can be selected to classify different human micro-movements. The accuracy rate can reach more than 99%. Zhao et al. [19] proposed a new continuous human motion recognition (HMR) method using micro-Doppler features, which can work in scenarios with non-target micro-motion interference. At present, human activity classification methods based on radar micro-Doppler data have high accuracy, but compared with simple behaviors such as walking, running, squatting, etc., complex activities can better express people's intentions and have more practical significance.

B. Target Recognition of Vulnerable Road Users

Non-motorized lanes contain not only pedestrians but also cyclists, such as bicycles and electric bicycles. In practical applications, pedestrians and non-motor vehicles are collectively called “Vulnerable Road Users (VRUs)”. By identifying different types of targets, traffic conditions on the road can be monitored more accurately to ensure the safety of pedestrians and vehicles [20],[21]. Du et al. [22] proposed a pedestrian and bicycle feature extraction method based on sparse coding of micro-Doppler features and verified it. The results showed that this method has higher recognition accuracy. Chipengo et al. [23] proposed a micro-Doppler feature difference classification method for disadvantaged road users based on a training machine learning algorithm. The research results show that the spectrogram obtained within a time window of 0.2s can achieve an accuracy of 95%. In the context of disadvantaged road users, Dubey et al. [24] proposed a new Bayesian-based deep metric learning method to learn feature embeddings corresponding to target micro-Doppler features. The results show that this method makes the target separable better. Gurbuz et al. [3] proposed a road target recognition scheme based on micro-Doppler, using joint radar and communication operations to achieve collaborative multi-static observation. The results show that the target recognition capability based on collaborative multi-sensors is significantly better than other methods.

Although the above machine learning methods are effective in certain scenarios, they cannot effectively train models in all cases and require a substantial amount of data as well as multi-sensor support. With the increasing complexity of tasks, whether these methods can maintain their performance advantage still necessitates further research and exploration.

C. Gait Recognition Based on Deep Learning

Many researchers use different human behaviors as inputs to neural networks to test the accuracy of their proposed algorithms for object recognition and prediction. Chen et al. [25] developed a Conditional Autoregressive Motion Diffusion Model (CAMDM) for character control, utilizing transformer-based architecture to generate diverse and high-quality character animations in real time. The model demonstrates robust performance across various dynamic user control signals, showcasing effective motion generation capabilities.

Jokanovic et al. [26] applied a 14-layer deep convolutional neural network (DCNN) on time-Doppler maps to classify human gait. Experimental results show that even in lower frequency or low signal-to-noise ratio environments, the DCNN structure can extract effective micro-Doppler features of human gait. R.P. et al. [27] used a superimposed auto-encoder to obtain the most significant features in radar echoes, and then used a softmax classifier to identify human motion. The results show that convolutional auto-encoder (CAE) is better than convolutional neural network (CNN) and ordinary auto-encoder in identifying human activities. W. Wang et al. [28] introduced NEURAL MARIONETTE, a transformer-based multi-action human motion synthesis system. Extensive experiments confirm the system's ability to generate motions with precision, naturalness, and fluidity. Furthermore, W. Wang et al. [29] proposed a technique used in the fields of computer graphics and artificial intelligence to generate realistic 3D human motion sequences based on descriptive tags or keywords.

As the above-mentioned deep learning models are increasingly used in radar target recognition, many target recognition and classification methods based on deep learning have been proposed. Therefore, in the field of radar target recognition, deep learning architecture has been achieved good results.

III. Data Construction

In this section, we describe the data collection and data processing process, as shown in Fig. 1.

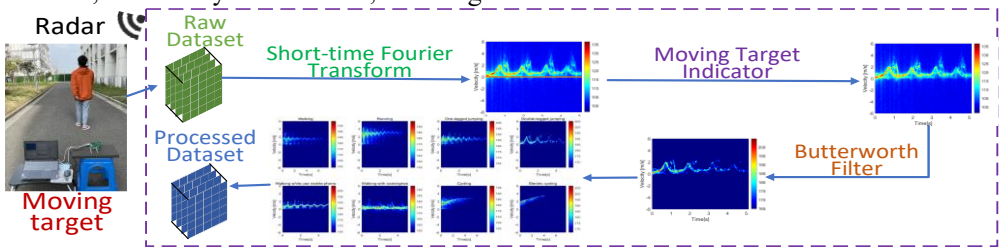

Fig. 1. Data collection and processing flow chart.

A. Data Collection

We collected eight different target behaviors on non-motorized lanes: walking; running; one-legged jumping; double-legged jumping; walking while using a mobile phone; walking with crutches; cycling; electric cycling. Each category collects 200 sets of data, with a total of 1,600 pieces of data in 8 categories. The outdoor pedestrian and non-motor vehicle behavior data collection system is shown in Fig. 2, in which Fig. 2(a) represents the moving target to be measured. Fig. 2(b) shows the data acquisition and processing platform of TI's millimeter wave radar chip, which is used to process the raw data obtained from the sensor. Fig. 2(c) represents the Radar system, which mainly transmits the target position, movement speed, and other information collected by AWR1642 and DCA1000EVM to the computer.

Fig. 2. Experimental environment diagram.

B. Data Preprocessing

Common features used for pedestrian and non-motor vehicle detection include speed, gait frequency, and stride length, which are extracted using short-time Fourier transform (STFT) [30], [31], [32] and radar reflection Doppler [33].

Due to the static interference of zero velocity in the actual measurement process, using a single delay line cancelling moving target indicator can filter out the static clutter returned by radar and keep only the moving target signal.

Due to the large amount of environmental noise present in the actual application scenario of this project, denoising algorithms can transform this data into a form more suitable for feature extraction [34] and neural network classification training. Therefore, this paper employs a Butterworth high-pass filter (BHPF) for processing, achieving filtering of the low-frequency part. Repeat the above steps for the eight categories respectively to obtain the speed-time micro-Doppler images of eight gaits as shown in Fig. 3.

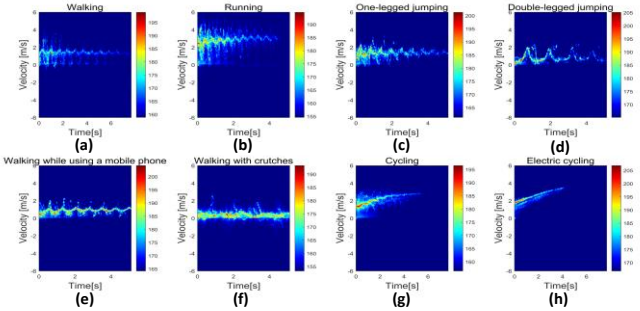

Fig. 3. Micro-Doppler images, (a)-(h) represent micro-Doppler images of different targets.

To remove the blank parts of each piece of data, each data is cropped and organized into $\boldsymbol{Dataset_C}$, with each sample size of $\mathbb{R}^{1\times500\times800}$. In order to expand the number of samples, data augmentation is applied to each sample. Each sample is horizontally cropped into 5 samples according to the following method, resulting in $\boldsymbol{Dataset_E}$ comprising 8000 samples:

$$\boldsymbol{Dataset_E}[i*5+j,0\sim500,0\sim400] = \boldsymbol{Dataset_C}[i,0\sim500,100*j\sim100*j+400] \quad (1)$$

where i represents the sample index in $\boldsymbol{Dataset_C}$, and j represents the corresponding index of the data cropped from each sample in $\boldsymbol{Dataset_E}$, with $j \in (0,1,2,3,4)$, the size of each sample in $\boldsymbol{Dataset_E}$ is $\mathbb{R}^{\mathbf{1\times500\times400}}$.

IV. Network Construction

In the actual scenes discussed in this article, scale changes of objects are common. The same object can appear at different scales, and different image scales contain varying levels of feature information, as shown in Fig. 4. The edge features of objects can usually be extracted in lower-level feature maps, while higher-level feature maps contain more abstract and feature-rich information. Therefore, multi-scale feature extraction is particularly important [35], [36].

To capture features of various scales and improve the network's perception ability, this paper proposes a channel attention mechanism for multi-scale residual networks based on CNN. It combines this mechanism with a pyramid network structure to build a model that can better classify and recognize the target behavior of non-motorized lanes.

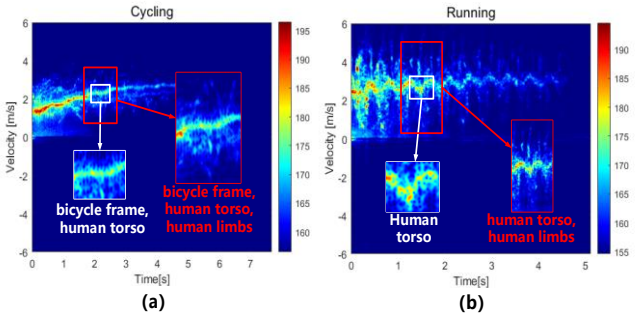

Fig. 4. Features of cycling (a) and running (b) at different scales.

A. Multi-scale Residual Channel Attention Mechanism

Traditional CNN models [37] face limitations in processing features of different scales due to fixed-size convolution and pooling kernels, leading to loss of details, especially in small targets or image boundaries. To capture the characteristics of various scales and improve the perceptual ability of the network, after introducing channel attention [38], a CNN-based method was proposed as the multi-scale residual channel attention (MRCA) mechanism, the MRCA structure is shown in Fig. 5a.

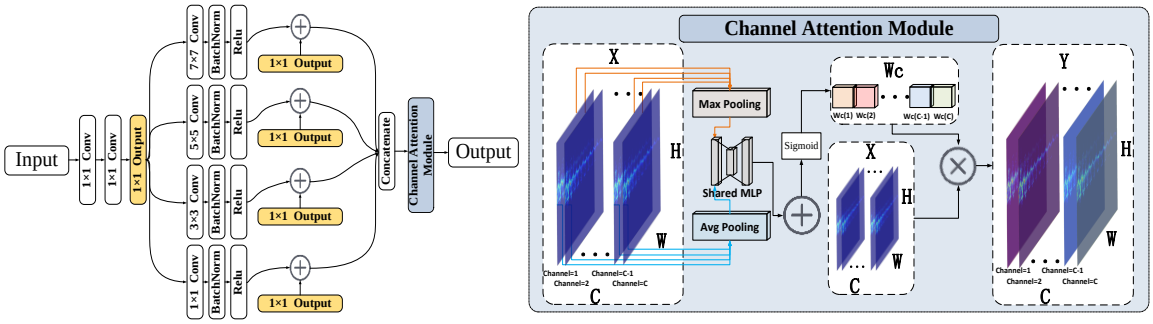

Fig. 5. Mrca block structure diagram, (a) Multi-scale residual channel attention mechanism, (b) Channel attention mechanism.

The feature extraction method in MRCA is as follows:

$$\boldsymbol{X_1} = Conv_{(1,1)}\left(Conv_{(1,1)}(\boldsymbol{X})\right) \tag{2}$$

$$\boldsymbol{X_i} = ReLU\left(BN\left(Conv_{(2i-1,2i-1)}(\boldsymbol{X_1})\right)\right) \tag{3}$$

where i represents the sequence number of feature extraction, $i \in (1,2,3,4)$, $Conv_{(2i-1,2i-1)}()$ represents a convolution operation whose convolution kernel size is $(2i-1, 2i-1)$. BN stands for batch normalization operation. The core idea of BN is to normalize the input of each feature channel into a distribution with mean 0 and variance 1, and then map it to the appropriate range through learnable scaling and translation parameters. In order to solve the problem of gradient disappearance and increase model flexibility, the residual connection is introduced to add $\boldsymbol{X_i}$ and $\boldsymbol{X_1}$.

As shown in Fig. 5b, the implementation of the channel attention mechanism is divided into global maximum pooling and global average pooling, and two feature vectors $\boldsymbol{F_{avg}}$ and $\boldsymbol{F_{max}}$ can be obtained. Then input $\boldsymbol{F_{avg}}$ and $\boldsymbol{F_{max}}$ into the fully connected layer to obtain two feature weight vectors:

$$\boldsymbol{W_{avg}} = \sigma\left(\boldsymbol{W_1 F_{avg}} + b_1\right), \boldsymbol{W_{max}} = \sigma(\boldsymbol{W_2 F_{max}} + b_2) \tag{4}$$

Among them, $\boldsymbol{W_1}$, b_1 and $\boldsymbol{W_2}$, b_2 are the weights and biases of the two fully connected layers respectively, σ representing the activation function (here we use Sigmoid). Then add the $\boldsymbol{W_{avg}}$ sum $\boldsymbol{W_{max}}$ element-wise to get the final weight vector:

$$\boldsymbol{W_c} = Sigmoid\left(\boldsymbol{W_{avg}} + \boldsymbol{W_{max}}\right) \tag{5}$$

Finally, the original feature map $X$and weights are weighted $\boldsymbol{W_c} \in [0,1]$and summed to obtain the enhanced feature map:

$$\boldsymbol{Y} = \sum_{i=1}^{C} \boldsymbol{W_{ci} X_{:,:,i}} \tag{6}$$

As can be seen from the above, after the feature map $\boldsymbol{X} \in \mathbb{R}^{\boldsymbol{H\times W\times C}}$ passes through a fully connected layer, $\boldsymbol{X}$ is compressed into a smaller feature map, and then activated by a Sigmoid function to obtain the weight vector $\boldsymbol{W_c}$ represents the importance of each channel, and is multiplied with the original feature map $\boldsymbol{X}$ to obtain $\boldsymbol{Y}$. Each channel of $\boldsymbol{Y}$ is assigned a different weight, indicating the importance of the channel.

The MRCA module considers multi-scale features and channel correlations using different-sized convolution kernels, capturing local and global context simultaneously, and adapting to multi-scale inputs. With channel attention, it weights feature channels, improving expressive ability and performance by focusing on meaningful scales, and accurately capturing key features for complex tasks. Adding MRCA to the convolutional neural network structure, the MRCA feature extraction structure (Features-Mrca) is obtained as shown in Fig. 6, which

represents the number of convolution and MRCA output channels. Four feature-MRCA blocks are connected in series with channel numbers 16, 64, 128, 512. Then fully connected layers with hidden layers of 256 and 512 are added at the end to construct the multi-scale residual channel attention mechanism network (MrcaNet).

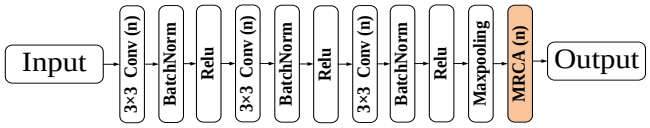

Fig. 6. MRCA feature extraction structure (Features-Mrca).

B. Feature Pyramid Multi-scale Residual Channel Attention Network

Feature pyramid structure (FPS) is a widely used technique for classification tasks. FPS constructs a multi-scale feature map pyramid to capture object features at various scales, enabling algorithms to handle objects of different sizes effectively. Feature Pyramid Network (FPN) [39],[40],[41] is a network architecture based on the FPS framework, which enables the generation of feature maps with varying resolutions across different levels.

Taking the CNN structure as an example, the input image is $\boldsymbol{I}$. Through a series of convolution and pooling operations, multiple feature maps are obtained, expressed as $\boldsymbol{F_i}, 2 \le i \le 5$, these maps correspond to different network levels. When $i = 2$, the feature map calculation process is as follows:

$$\boldsymbol{F_i} = ReLU(Conv(\boldsymbol{I}, \boldsymbol{W_i})) \tag{7}$$

Among them, $Conv()$ represents the convolution operation, $ReLU()$ represents the modified linear unit operation, and $\boldsymbol{W_i}$ is the convolution kernel parameter of the corresponding level. When $i > 2$, the feature map calculation process was as follows:

$$\boldsymbol{F_i} = ReLU(Conv(MaxPool(\boldsymbol{F_{i-1}}), \boldsymbol{W_i})) \tag{8}$$

Among them, $MaxPool()$ represents the maximum pooling operation. Due to the pooling operation, as the level increases, the resolution of the feature map gradually decreases, causing the network to lose detailed information when processing small objects. In order to solve this problem, FPN is introduced. First, perform an upsampling operation:

$$\boldsymbol{U_{F_i}} = Upsample(\boldsymbol{F_i}) \tag{9}$$

Among them, $Upsample()$ represents the upsampling operation, which uses bilinear interpolation to estimate the value of the new pixel using the weighted average of the surrounding four pixels. When the image is $\boldsymbol{I_U}$, interpolation is performed at the position of the image, set the four adjacent pixels respectively $\boldsymbol{I_U}(x_1, y_1)$, $\boldsymbol{I_U}(x_1, y_2)$, $\boldsymbol{I_U}(x_2, y_1)$, $\boldsymbol{I_U}(x_2, y_2)$, bilinear interpolation can be expressed as:

$$I_{interp}(x, y) = (1-\alpha)(1-\beta)\boldsymbol{I_U}(x_1, y_1) + \alpha(1-\beta)\boldsymbol{I_U}(x_2, y_1) + (1-\alpha)\beta\boldsymbol{I_U}(x_1, y_2) + \alpha\beta\boldsymbol{I_U}(x_2, y_2) \tag{10}$$

where α sum β is relative to the relative position (x, y) in the $x$and y direction:

$$\alpha = x - x_1, \beta = y - y_1 \tag{11}$$

where the value range of α and β is [0,1].

Then, Construct a feature pyramid through channel-dimensional connection and upsampling operations.

Combining MRCA with FPN creates the feature pyramid-multi-scale residual channel attention network (P-MrcaNet). P-MrcaNet leverages MRCA's multi-scale adaptability and FPN's multi-level feature fusion to enhance object detection across scales. FPN's top-down feature propagation complements MRCA's attention mechanism, improving feature relationships and contextual utilization. Integrating multi-scale features and rich contextual information enhances P-MrcaNet's detection accuracy and robustness, enabling precise target localization across various scales.

C. Improved Network Structure

To optimize Feature-MRCA combined with FPS, we placed multiple Feature-MRCAs before and after FPS in strategic locations and conducted experiments to analyze their impacts. Findings emphasized the importance of including a 16-channel Feature-MRCA before FPS for effective learning of low-level image features. Similarly, lacking a 32-channel Feature-MRCA before FPS led to degraded performance in recognizing medium-level features crucial for accurate recognition. Furthermore, the absence of a 64-channel Feature-MRCA impacted the network's understanding of the overall image structure, resulting in reduced performance when processing global information. Moreover, the absence of a 128-channel Feature-MRCA before FPS further hindered the network's ability to comprehend the overall image structure, especially in recognizing complex objects and scenes.

After conducting multiple experiments, we determined that incorporating Feature-MRCA with channel numbers 16, 32, 64, and 128 before FPS provided optimal results. Following FPS, we added another Feature-MRCA configuration with channel numbers 128, 256, 256, and 512. This refined setup effectively merged and processed features, assisting the network in comprehending the complex semantic and spatial information within images. The model structure is illustrated in Fig. 7, where the numbers displayed in the Features-MRCA block signify the number of output channels. The input to the network is an $X \in \mathbb{R}^{1\times500\times400}$ single-channel image.

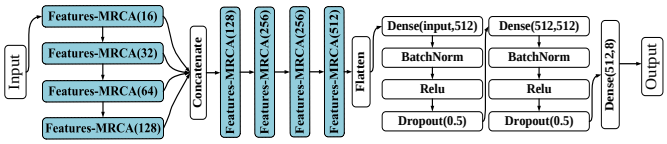

Fig. 7. Improved P- MrcaNet model structure.

The feature map after feature pyramid-multi-scale residual channel attention feature extraction is $\boldsymbol{X_i}$. When $i = 1$, the feature extraction process is as follows:

$$\boldsymbol{X_i} = Features\_MRCA(\boldsymbol{X}) \tag{12}$$

When $2 \le i \le 4$, the feature extraction process was as follows:

$$\boldsymbol{X_i} = Features\_MRCA(\boldsymbol{X_{i-1}}) \tag{13}$$

When $2 \le i \le 4$, the feature maps $\boldsymbol{X_i}$ are upsampled to the size of (250, 200) so that they have similar spatial resolution:

$$\boldsymbol{X_{i\_Upsample}} = Upsample(\boldsymbol{X_i}) \tag{14}$$

Then the feature map of the upper sample is concatenated with the output of the first feature extraction module to form the feature pyramid $\boldsymbol{Feature_{Pyramid}}$.

Taking $\boldsymbol{Feature_{Pyramid}} \in \mathbb{R}^{240\times250\times200}$ as input, the Feature graph $\boldsymbol{Feature_{Output}}$ is obtained by the 5th ~ 8th feature extraction. Flatten the feature map after feature extraction into a one-dimensional vector $\boldsymbol{Flatten_{Output}}$.

Finally, the final output is obtained by inputting the MLP structure $classifier$ consisting of three fully connected layers, each layer followed by batch normalization, $ReLU$ activation functions and $Dropout$ operations:

$$\boldsymbol{Output} = classifier\left(\boldsymbol{Flatten_{Output}}\right) \quad (15)$$

Among them, the output size of the last fully connected layer is 8, corresponding to 8 categories.

D. Comparison of Computational Resources

In this section, the computational resources required for the P-MrcaNet model and other models are compared, as shown in Table I, where the total params represent the total number of trainable parameters in the model, the pass size indicates the memory consumption size during forward/backward propagation, the params size indicates the memory size occupied by model parameters, and the total size indicates the total memory size expected to be occupied by the entire model.

Firstly, we compared the total parameter count of P-MrcaNet with several deep learning models, including LeNet [42], AlexNet [43], ZFNet [44], VGG13 [45], VGG16 [46], SENet [47], ShuffleNetV2 [48] and ViT [49]. The results indicated that P-MrcaNet has a total parameter count of 5.63e7, which falls within the moderate range. Next, we investigated the memory usage performance of P-MrcaNet. We calculated its memory size for forward/backward propagation to be 1057.29MB, with the parameter size being 214.84MB, totaling approximately 1272.89MB. Compared with the model with larger parameters, P-MrcaNet has lower memory requirements.

TABLE I
PERFORMANCE METRICS OF MODELS

Model	Total params	Pass size (MB)	Params size (MB)	Total size (MB)
ShuffleNetV2[48]	1.16e6	637.77	4.43	642.97
SENet[47]	2.18e6	793.67	8.31	802.74
LeNet[42]	2.40e7	19.08	91.60	111.44
MrcaNet	**2.60e7**	**1024.16**	**99.23**	**1124.15**
ViT[49]	4.35e7	255.38	3.16	259.3
P-MrcaNet	**5.63e7**	**1057.29**	**214.84**	**1272.89**
AlexNet[43]	5.70e7	33.70	217.52	251.98
ZFNet[44]	5.83e7	106.88	222.34	329.98
VGG13[45]	4.04e8	790.73	1540.03	2331.53
VGG16[46]	4.13e8	869.84	2446.39	2446.39

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, in order to train and optimize the model effectively, the Adam optimizer with a learning rate of 1e-4 is used for updating the model parameters. The batch size is set to 32 to balance training speed and memory consumption. The training model is run for 300 epochs, as verified by several experiments to ensure that the model achieves sufficient convergence and generalization ability during the training process. For the loss calculation, we use the cross-entropy loss function. The dataset is divided into the training set, validation set, and test set according to the ratio of 5:3:2. The training set is used to train the model parameters, the hyperparameters are adjusted, and the performance of the model is evaluated using the validation set. Finally, the test set is used for the final evaluation of the model to ensure the generalization and reliability of the model.

A. T-SNE Visualization

In order to intuitively show the classification performance of each network model, we visualize the t-Distributed Stochastic Neighbor Embedding (t-SNE) plots for each network model, using data that is not involved in model training.

Through t-SNE, high-dimensional features are converted into points on a two-dimensional plane, and the classification effect is displayed using graphics. Fig. 8 shows the t-SNE visualization results of some network models.

The t-SNE visualization results show that P-MrcaNet is superior to other network models in classification.

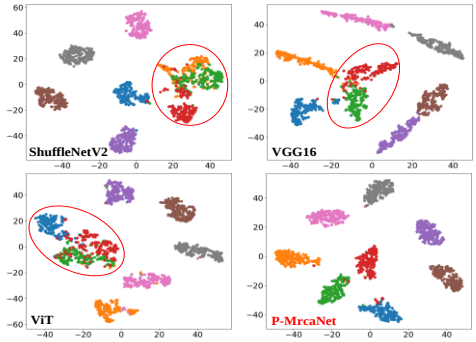

Fig. 8. t-SNE visualization of some models.

B. Accuracy Comparison

Use the optimal parameter combination built above to train the network model and verify its classification and recognition effect. Then, the classification results of various target samples are visualized through the confusion matrix. The confusion matrix and accuracy of each model are shown in Fig. 9.

The accuracy of all models in classifying target behavior on non-motor vehicle lanes ranges from 80.563% to 97.500%.

LeNet's shallow structure limited its ability to process complex images, resulting in reduced performance in capturing intricate features. AlexNet, while an improvement over LeNet, still faced structural limitations affecting its effectiveness in handling diverse target behaviors. VGG13 and VGG16 improved performance by deepening the network but increased computational complexity, potentially leading to overfitting and higher costs. ZFNet, despite architectural optimizations, struggled with local information extraction and multi-scale feature fusion, impacting its accuracy.

ShuffleNetV2, a lightweight model, achieved 91.125% accuracy, demonstrating promising performance and computational advantages. However, for diverse target behaviors, more robust feature extraction capabilities may be necessary. ViT excelled in capturing global information but had 89.563% accuracy, indicating weakness in extracting critical local information for non-motorized lane target behavior classification.

SENet with attention mechanisms reached 90.375% accuracy, highlighting the utility of attention mechanisms but suggesting room for improvement in handling complex scenarios.

In contrast, MrcaNet significantly improved accuracy to 95.625%, with P-MrcaNet further increasing accuracy to 97.500%. These models comprehensively captured image information, making them leaders in classifying target behaviors on non-motorized lanes.

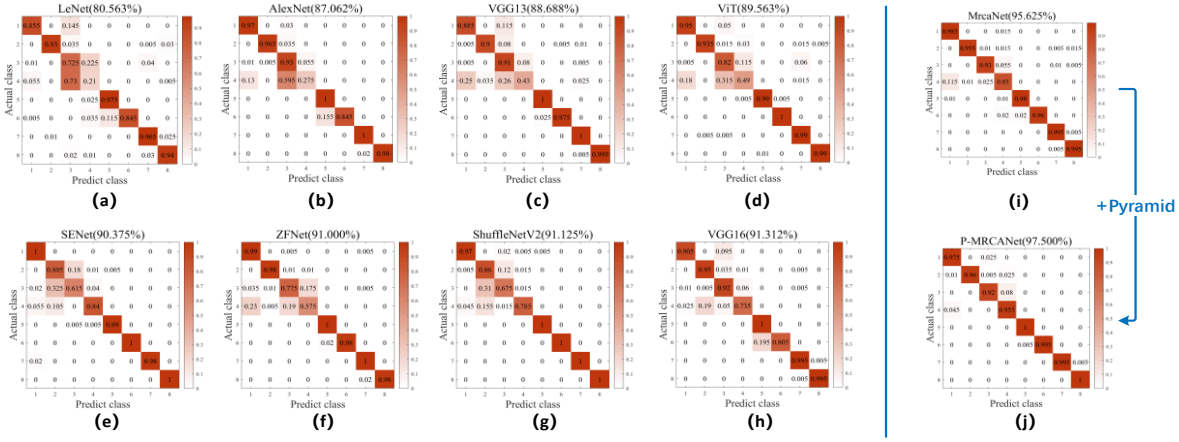

Fig. 9. Confusion matrix diagrams for different networks, where (a)-(h) are confusion matrix diagrams for other networks, and (i), (j) are confusion matrix diagrams for the network proposed in this paper.

Combined with Table I and Fig. 9, the classification accuracy and required resources of all models were graded. Models with accuracy below 85% are rated as "Inadequate", those with accuracy between 85% and 95% are rated as "Adequate", and those with accuracy above 95% are rated as "Good". Models with total params less than 1e7 are classified as "Small", those with params between 1e7 and 1e8 are classified as "Medium", and those with params above 1e8 are classified as "Large". Models with a total size less than 500MB are labeled as "Small", those between 500MB and 1500MB are labeled as "Medium", and those above 1500MB are labeled as "Large". The summary of these classifications is provided in Table II.

As seen from Table II, P-MrcaNet achieves "Good" level accuracy with a medium number of parameters and model size compared to other models. Specifically, P-MrcaNet avoids the massive parameter count and the large-scale structure of large models like VGG13 and VGG16, while providing better accuracy than early models such as LeNet and AlexNet. Compared to models like SENet, ZFNet, and ShuffleNetV2, P-MrcaNet offers higher accuracy while maintaining a similar scale. Furthermore, although ViT also has "Adequate" accuracy and a medium-sized parameter count, P-MrcaNet still holds an advantage in terms of accuracy.

TABLE II
MODEL PERFORMANCE AND SCALE SUMMARY

Model	Accuracy	Total params	Total size
LeNet[42]	Inadequate	Medium	Small
AlexNet[43]	Adequate	Medium	Small
VGG13[45]	Adequate	Large	Large
ViT[49]	Adequate	Medium	Small
SENet[47]	Adequate	Small	Medium
ZFNet[44]	Adequate	Medium	Small
ShuffleNetV2[48]	Adequate	Small	Medium
VGG16[46]	Adequate	Large	Large
MrcaNet	**Good**	**Medium**	**Medium**
P-MrcaNet	**Good**	**Medium**	**Medium**

C. Comparison of Different Performance Indicators

Based on the above research, the classification performance of different models is measured by comparing the accuracy, precision, recall, F1 value, Kappa coefficient, and other classification performance measures of the data confusion matrix [50]. Accuracy is the correct rate:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{16}$$

where the true number (TP) is the correct prediction number that belongs to the class and is consistent with the actual class; False positives (FP) are counts of predictions made for categories that do not match the actual category; False negatives (FN) are instances of the actual class present, but the prediction does not include the class count; True negative (TN) is the count of instances where the prediction does not include a class and the actual class does not exist.

Precision represents the proportion of samples in the study area that were correctly predicted:

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

Recall is also called recall rate. It represents the proportion of samples with correct predictions in the study area:

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

The F1 value is the arithmetic mean divided by the geometric mean, and the larger the value, the better. It is derived by combining Precision and Recall:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{19}$$

The Kappa coefficient is used for consistency testing and can also be used to measure classification accuracy. Its calculation is based on a confusion matrix, which represents the ratio of error reduction between classification and completely random classification.

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \tag{20}$$

where the p_0 is the sum of the number of correctly classified samples of each category divided by the total number of samples, which is the overall classification accuracy. p_e is

calculated by the following formula:

$$p_e = \frac{a1 * b1 + a2 * b2 + \cdots + ac * bc}{n * n} \quad (21)$$

As can be seen from the results of different classification indicators in Fig. 10, each classification indicator of P-MrcaNet is optimal and maintained at around 97.5%.

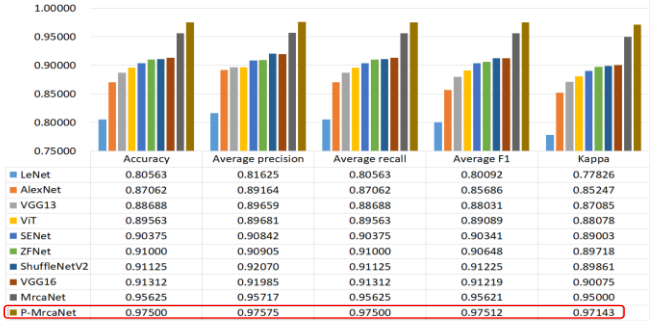

Fig. 10. Results chart of different classification indicators.

D. Ablation Experiments

In order to better analyze the role and importance of each component in P-MrcaNet, we conducted ablation experiments on the model. The details of the experimental parameters are consistent with our previous experimental settings. The ablation models are: (a) the model that removes the multi-scale residual channel attention block, (b) the model that removes the multi-scale residual feature block, (c) the model that removes the channel attention mechanism block, (d) the feature pyramid block is deleted model, and (e) P-MrcaNet.

The results of the ablation experiments are shown in Table III. The results of ablation experiments show that multi-scale residual channel attention and feature pyramid structure play significant roles in this classification task.

TABLE III
RESULTS OF THE ABLATION EXPERIMENTS

Model	Multi-scale residual	Channel attention	Feature pyramid	Accuracy
a	No	No	Yes	92.500%
b	No	Yes	Yes	93.125%
c	Yes	No	Yes	94.063%
d	Yes	Yes	No	95.625%
e	Yes	Yes	Yes	97.500%

E. Validation on Public Datasets

In this paper, to enhance the credibility and performance of the P-MrcaNet model, we conducted further validation using a public dataset. We selected the Defence Institute of Advanced Technology micro-Doppler Human Activity Recognition (DIAT-μRadHAR) dataset [51] as the experimental foundation.

The DIAT-μRadHAR dataset was acquired using an X-band continuous wave (CW) 10 GHz radar system. This dataset consists of records documenting suspicious human activities, such as army crawling, army jogging, gun jumping, army marching, boxing, rock/grenade throwing, etc. Throughout the data acquisition process, individuals with varying heights, weights and genders were instructed to perform these suspicious activities at different target angles (0°, ±15°, ±30° and ±45°), within a range spanning from 10 m to 0.5 km from the radar. Subsequently, micro-Doppler features were extracted from the acquired data and compiled into this dataset.

To simulate a more complex real-world environment, we introduced random noise from a normal distribution with a mean of 0 and a standard deviation equal to the standard deviation of the input samples into the DIAT-μRadHAR dataset. The same experimental methods used for the dataset collected in this study were applied to this public dataset, and the experimental results are shown in Table IV.

The results of the experiments clearly demonstrate the excellent performance of the P-MrcaNet model on the DIAT-μRadHAR dataset compared to traditional deep learning models. This experiment further solidifies the scientific foundation of our proposed model, providing a robust theoretical basis for its effectiveness in addressing diverse environments and challenges.

TABLE IV
SUMMARY OF RECOGNITION ACCURACY OF DIFFERENT NETWORKS ON PUBLIC DATASETS

Network	Accuracy
LeNet[42]	82.940%
ShuffleNetV2[48]	85.185%
ViT[49]	90.608%
AlexNet[43]	90.741%
ZFNet[44]	91.931%
SENet[47]	92.989%
VGG16[46]	93.254%
VGG13[45]	93.386%
MrcaNet	**95.635%**
P-MrcaNet	**96.296%**

VI. CONCLUSION

In order to solve the problem of low efficiency in identifying vulnerable road users, this paper proposes a new non-motorized lane user behavior classification and identification method based on the multi-scale residual attention mechanism and the feature pyramid model P-MrcaNet and uses actual measurements. The data generates a confusion matrix to compare the accuracy and performance measures of different networks. The results show that the convolutional neural network based on P-MrcaNet can effectively distinguish eight behavioral categories such as walking, running, walking with crutches, bicycles, and electric vehicles. The overall accuracy of the model on the test set is 97.500%, which is better than the classification accuracy of ZFNet, VGG16, ShuffleNetV2, ViT, and other networks.

Finally, the accuracy, precision, recall, F1 value, Kappa coefficient, and other classification performance measures of the data confusion matrix are compared to show that the new model has the optimal classification effect. In the future, the branches of the P-MrcaNet model can be expanded according to the input features to meet more feature inputs, allowing the model to better adapt to different types of data, such as images, text, and audio. The model can also be adaptively adjusted according to the input features to achieve the best classification effect. In addition, it can also be planned to expand the recognition of all target types on urban roads, such as cars, trucks, tricycles, etc., to better solve urban road congestion and safety problems.

REFERENCES

[1] M. H. Zaki and T. Sayed, "Exploring walking gait features for the automated recognition of distracted pedestrians," *IET Intelligent Transport Systems*, vol. 10, no. 2, pp. 106–113, Mar. 2016.

[2] M. Kumar, N. Singh, R. Kumar, S. Goel, and K. Kumar, "Gait recognition based on vision systems: A systematic survey," *Journal of Visual Communication and Image Representation*, vol. 75, p. 103052, Feb. 2021.

[3] S. Z. Gurbuz and M. G. Amin, "Radar-Based Human-Motion Recognition With Deep Learning," *IEEE SIGNAL PROCESSING MAGAZINE*, 2019.

[4] D. Lee, H. Park, T. Moon, and Y. Kim, "Continual Learning of Micro-Doppler Signature-Based Human Activity Classification," *IEEE Geosci. Remote Sensing Lett.*, vol. 19, pp. 1–5, 2022.

[5] P. Striano, C. V. Ilioudis, C. Clemente, and J. J. Soraghan, "Assessment of Micro-Doppler based road targets recognition based on co-operative multi-sensor automotive radar applications," *IEEE Radar Conference*, 2020.

[6] J. K. T. Tang, H. Leung, T. Komura, and H. P. H. Shum, "Emulating human perception of motion similarity," *Computer Animation & Virtual*, vol. 19, no. 3–4, pp. 211–221, Jan. 2008.

[7] Y. Kim and T. Moon, "Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks," *IEEE Geosci. Remote Sensing Lett.*, vol. 13, no. 1, pp. 8–12, Jan. 2016.

[8] X. Jiang, Y. Zhang, Q. Yang, B. Deng, and H. Wang, "Millimeter-Wave Array Radar-Based Human Gait Recognition Using Multi-Channel Three-Dimensional Convolutional Neural Network," *Sensors*, vol. 20, no. 19, p. 5466, Sep. 2020.

[9] Y. Shi, X. Liao, Z. Yu, Z. Li, C. Wang, and S. Xue, "Robust Gait Recognition Based on Deep CNNs With Camera and Radar Sensor Fusion," *IEEE INTERNET OF THINGS JOURNAL*, vol. 10, no. 12, 2023.

[10] T. Komura, Y. Shinagawa, and T. L. Kunii, "Attaching Physiological Effects to Motion-Captured Data," *Graphics interface*, pp. 27–36, 2001.

[11] J. C. P. Chan, H. Leung, J. K. T. Tang, and T. Komura, "A Virtual Reality Dance Training System Using Motion Capture Technology," *IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES*, vol. 4, no. 2, 2011.

[12] P. Sandilands, M. G. Choi, and T. Komura, "Interaction capture using magnetic sensors," *Computer Animation & Virtual*, vol. 24, no. 6, pp. 527–538, Nov. 2013.

[13] S. Björklund, H. Petersson, and G. Hendeby, "Features for micro-Doppler based activity classification," *IET Radar, Sonar & Navigation*, vol. 9, no. 9, pp. 1181–1187, Dec. 2015.

[14] X. Li and F. Fioranelli, "Semisupervised Human Activity Recognition With Radar Micro-Doppler Signatures," *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, vol. 60, 2022.

[15] X. Li, Y. He, and X. Jing, "A Survey of Deep Learning-Based Human Activity Recognition in Radar," *Remote Sensing*, vol. 11, no. 9, 2019.

[16] Z. Ni and B. Huang, "Gait-Based Person Identification and Intruder Detection Using mm-Wave Sensing in Multi-Person Scenario," *IEEE Sensors J.*, vol. 22, no. 10, pp. 9713–9723, May 2022.

[17] X. Bai, Y. Hui, L. Wang, and F. Zhou, "Radar-Based Human Gait Recognition Using Dual-Channel Deep Convolutional Neural Network," *IEEE Trans. Geosci. Remote Sensing*, vol. 57, no. 12, pp. 9767–9778, Dec. 2019.

[18] J. Li, X. Chen, G. Yu, X. Wu, and J. Guan, "High-precision human activity classification via radar micro-doppler signatures based on deep neural network," in *IET International Radar Conference (IET IRC 2020)*, , Online Conference: Institution of Engineering and Technology, 2021, pp. 1124–1129.

[19] R. Zhao, X. Ma, X. Liu, and F. Li, "Continuous Human Motion Recognition Using Micro-Doppler Signatures in the Scenario With Micro Motion Interference," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5022–5034, Feb. 2021.

[20] T. Lavrenko, T. Gessler, T. Walter, H. Mantz, and M. Schlick, "Radar Based Detection and Classification of Vulnerable Road Users," in *The 8th International Symposium on Sensor Science*, MDPI, May 2021, p. 67.

[21] P. Rippl, J. Iberle, and T. Walter, "Classification of Vulnerable Road Users Based on Spectrogram Autocorrelation Features," in *2021 18th European Radar Conference (EuRAD)*, London, United Kingdom: IEEE, Apr. 2022, pp. 293–296.

[22] R. Du, Y. Fan, and J. Wang, "Pedestrian and Bicyclist Identification Through Micro Doppler Signature With Different Approaching Aspect Angles," *IEEE Sensors J.*, vol. 18, no. 9, pp. 3827–3835, May 2018.

[23] U. Chipengo, A. P. Sligar, S. M. Canta, M. Goldgruber, H. Leibovich, and S. Carpenter, "High Fidelity Physics Simulation-Based Convolutional Neural Network for Automotive Radar Target Classification Using Micro-Doppler," *IEEE Access*, vol. 9, pp. 82597–82617, 2021.

[24] A. Dubey, A. Santra, J. Fuchs, M. Lubke, R. Weigel, and F. Lurz, "A Bayesian Framework for Integrated Deep Metric Learning and Tracking of Vulnerable Road Users Using Automotive Radars," *IEEE Access*, vol. 9, pp. 68758–68777, 2021.

[25] R. Chen, M. Shi, S. Huang, P. Tan, T. Komura, and X. Chen, "Taming Diffusion Probabilistic Models for Character Control." arXiv, Apr. 23, 2024. Accessed: May 18, 2024. [Online].

[26] B. Jokanovic, M. Amin, and F. Ahmad, "Radar fall motion detection using deep learning," in *2016 IEEE Radar Conference (RadarConf)*, Philadelphia, PA, USA: IEEE, May 2016, pp. 1–6.

[27] R. P. Trommel, R. I. A. Harmanny, L. Cifola, and J. N. Driessen, "Multi-target human gait classification using deep convolutional neural networks on micro-doppler spectrograms," in *Proc. 13th European Radar Conference*, London, UK, 5–7 October 2016, pp. 81–84.

[28] W. Wang *et al.*, "NEURAL MARIONETTE: A Transformer-based Multi-action Human Motion Synthesis System." arXiv, Nov. 27, 2023. Accessed: Mar. 27, 2024. [Online].

[29] W. Wang and R. Chen, "Tag-driven 3D human motion generation with transformer VAE," in *International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2023)*, A. Bhattacharjya and X. Feng, Eds., Hangzhou, China: SPIE, May 2023, p. 144.

[30] Z. Ni and B. Huang, "Robust Person Gait Identification Based on Limited Radar Measurements Using Set-Based Discriminative Subspaces Learning," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[31] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel Recognition and Gait Classification Based on Multistatic Micro-Doppler Signatures Using Deep Convolutional Neural Networks," *IEEE Geosci. Remote Sensing Lett.*, vol. 15, no. 5, pp. 669–673, May 2018.

[32] Y. Lang, Q. Wang, Y. Yang, C. Hou, Y. He, and J. Xu, "Person identification with limited training data using radar micro-Doppler signatures," *Micro & Optical Tech Letters*, vol. 62, no. 3, pp. 1060–1068, Mar. 2020.

[33] S. M. M. Islam, O. Borić-Lubecke, Y. Zheng, and V. M. Lubecke, "Radar-Based Non-Contact Continuous Identity Authentication," *Remote Sensing*, vol. 12, no. 14, p. 2279, Jul. 2020.

[34] S. Zhang, Y. Wang, and A. Li, "Gait Energy Image-Based Human Attribute Recognition Using Two-Branch Deep Convolutional Neural Network," *IEEE Trans. Biom. Behav. Identity Sci.*, vol. 5, no. 1, pp. 53–63, Jan. 2023.

[35] B. Lin, S. Zhang, Y. Liu, and S. Qin, "Multi-Scale Temporal Information Extractor For Gait Recognition," in *2021 IEEE International Conference on Image Processing (ICIP)*, Anchorage, AK, USA: IEEE, Sep. 2021, pp. 2998–3002.

[36] A. Sepas-Moghaddam, S. Ghorbani, N. F. Troje, and A. Etemad, "Gait Recognition using Multi-Scale Partial Representation Transformation with Capsules," in *2020 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy: IEEE, Jan. 2021, pp. 8045–8052.

[37] C. Filipi Gonçalves Dos Santos et al., "Gait Recognition Based on Deep Learning: A Survey," ACM Comput. Surv., vol. 55, no. 2, pp. 1–34, Feb. 2023.

[38] A. Huang, L. Jiang, J. Zhang, and Q. Wang, "Attention-VGG16-UNet: a novel deep learning approach for automatic segmentation of the median nerve in ultrasound images," *Quant Imaging Med Surg*, vol. 12, no. 6, pp. 3138–3150, Jun. 2022.

[39] S. Hou, X. Liu, C. Cao, and Y. Huang, "Set Residual Network for Silhouette-Based Gait Recognition," *IEEE Trans. Biom. Behav. Identity Sci.*, vol. 3, no. 3, pp. 384–393, Jul. 2021.

[40] J. Chen, Z. Wang, P. Yi, K. Zeng, Z. He, and Q. Zou, "Gait Pyramid Attention Network: Toward Silhouette Semantic Relation Learning for Gait Recognition," *IEEE Trans. Biom. Behav. Identity Sci.*, vol. 4, no. 4, pp. 582–595, Oct. 2022.

[41] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 936–944.

[42] M. Kayed, A. Anter, and H. Mohamed, "Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture," in *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*, Aswan, Egypt: IEEE, Feb. 2020, pp. 238–243.

[43] A. Nainwal, G. Sharma, V. Kansal, S. Bhatla and B. Pant, "Comparative Study of VGG-13, AlexNet, MobileNet and Modified-DarkCovidNet for Chest X-Ray Classification," in *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2023, pp. 413-417.

[44] K. Singh and N. Singh, "Multi-level authentication model with Political Dingo Optimizer-enabled ZFNet," in *2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*, Greater Noida, India: IEEE, Jan. 2023, pp. 1022–1026.

[45] S. Qasim Gilani, T. Syed, M. Umair, and O. Marques, "Skin Cancer Classification Using Deep Spiking Neural Network," *J Digit Imaging*, vol. 36, no. 3, pp. 1137–1147, Jan. 2023.

[46] J. N. Mogan, C. P. Lee, K. M. Lim, and K. S. Muthu, "VGG16-MLP: Gait Recognition with Fine-Tuned VGG-16 and Multilayer Perceptron," *Applied Sciences*, vol. 12, no. 15, p. 7639, Jul. 2022.

[47] L. Runyu, "Pedestrian detection based on SENet with attention mechanism," in *2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC)*, Chongqing, China: IEEE, Sep. 2023, pp. 616–619.

[48] S. N. P. S, J. V. B. Benifa, A. K. P, and A. Vijayakumar, "Human Activity Recognition using ShuffleNetV2 Model," in *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, Chennai, India: IEEE, Dec. 2023, pp. 1–5.

[49] J. N. Mogan, C. P. Lee, K. M. Lim, and K. S. Muthu, "Gait-ViT: Gait Recognition with Vision Transformer," *Sensors*, vol. 22, no. 19, p. 7362, Sep. 2022.

[50] Md. M. Islam, S. Nooruddin, F. Karray, and G. Muhammad, "Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects," *Computers in Biology and Medicine*, vol. 149, p. 106060, Oct. 2022.

[51] M. Chakraborty, H. C. Kumawat, S. V. Dhavale, and A. A. B. Raj, "DIAT-μ RadHAR (Micro-Doppler Signature Dataset) & μ RadNet (A Lightweight DCNN)—For Human Suspicious Activity Recognition," *IEEE Sensors Journal*, vol. 22, no. 7, pp. 6851–6858, Apr. 2022.

Jiaqing He received the B.S. degree from the School of Computer and Software Engineering, Anhui Institute of Information Technology, Wuhu, China, in 2021. he is currently pursuing the M.S. degree from the School of Transportation and Civil Engineering, Nantong University, Nantong, China. His current research interests include radar signal processing and human behavior recognition.

Yihan Zhu received the B.S. degree from the School of Transportation and Civil Engineering, Nantong University, Nantong, China, in 2022, where she is currently pursuing the M.S. degree. Her current research interests include radar signal processing and Deep learning.

Hua Bing obtained a master's degree from the Party School of the Jiangsu Provincial Committee of the Communist Party of China in 2009 and obtained the intermediate professional technical qualification for fire fighting and rescue in 2012. He is currently the captain of the Nantong Fire Rescue Detachment. His research interests include fire fighting and rescue and the development of smart firefighting technology.

Zhihuo Xu (Senior Member, IEEE) received the Ph.D. degree in communication and information system from the University of Chinese Academy of Sciences (UCAS), Beijing, China, in 2016. He founded Radar Remote Sensing Group, Nantong University, China, in 2016. From 2017 to 2018, he was an Academic Visitor with the University of Birmingham, Birmingham, U.K. His current research interests include radar system design, radar signal, and image processing.

Yongwei Zhang (Member, IEEE) received the B.S. degree in communication engineering from Jilin University, Changchun, China, in 1996, and the Ph.D. degree from the School of Electrical and Electronic Engineering, The University of Manchester, Manchester, U.K., in 2007. Dr. Zhang was awarded as an Outstanding Member of the PHS Project Team in 2000 and received the 2001 Chairman Award from Lucent Technologies (China) Company Ltd.

Liu Chu (Member, IEEE) received the B.E. degree in material science and engineering and the M.E. degree in mechanics from Dalian Maritime University, Dalian, China, in 2010 and 2012, respectively, and the Ph.D. degree in mechanics from the Institut National des Sciences Appliquées de Rouen (INSA Rouen), Rouen, France, in 2017. She is currently a Research Associate with the Center for Transformative Science, ShanghaiTech University, Shanghai, China. Her research interests include artificial material microstructure optimization.

Quan Shi (Member, IEEE) received the M.S. and Ph.D. degrees in management information systems from the University of Shanghai for Science and Technology, Shanghai, China, in 2005 and 2011, respectively. He is currently a Professor with the School of Transportation and Civil Engineering, Nantong University, Nantong, China. His research interests include the development of signal and image processing and big data techniques.

Robin Braun (Life Senior Member, IEEE) received the B.Sc. degree (Hons.) from Brighton University, Brighton, U.K., in 1980, and the M.Sc.(Eng.) and Ph.D. degrees from the University of Cape Town, Cape Town, South Africa, in 1982 and 1986, respectively. He started his academic career at the University of Cape Town, in 1986. In 1998, he moved to the University of Technology Sydney, Australia, where he occupied the Chair of Telecommunications Engineering. His recent work has been in network protocols and the management of complex next-generation networks.

Jiajia Shi (Member, IEEE) received the B.Sc. and M.E. degrees from Central South University, Changsha, China, in 2007 and 2010, respectively, and the Ph.D. degree from the University of Technology at Sydney, Sydney, NSW, Australia, in 2015. He is currently an Associate Professor with the School of Transportation and Civil Engineering, Nantong University, Nantong, China. His research interests include signal processing.