Multi-Channel Fusion Human Activity Recognition Algorithm Based on Millimeter-Wave Radar

Junda Zhu*, Shisheng Guo*†‡, Longzhen Tang*, Cui Guolong*†
* School of Information and Communication Engineering,
University of Electronic Science and Technology of China, Chengdu 611731,China
† Yangtze Delta Region Institute, University of Electronic Science and Technology of China, Quzhou 324000, China
‡ Corresponding author: ssguo@uestc.edu.cn

***Abstract*—Traditional datasets for human activity recognition (HAR) are typically obtained from single-channel radar, which does not effectively utilize the target feature information. This paper introduces a multi-channel fusion HAR algorithm based on millimeter-wave radar. The algorithm leverages the multi-channel data acquisition capability of multi-input multi-output (MIMO) radar and employs three independent MobileNetV3 lightweight network models for classification and recognition. Subsequently, a hard classification algorithm is utilized to fuse the multi-channel information, enabling the identification of human behaviors. Experimental results indicate that the proposed algorithm achieves an excellent classification performance of 97.05% on the experimental dataset. Moreover, it demonstrates a marked improvement in accuracy and stability compared to traditional single-channel classification methods. Furthermore, when compared with commonly used classification networks for human behavior, MobileNetV3 used in the algorithm highlights its ability to achieve commendable classification performance with minimal computational cost. This validates the proposed method as a viable lightweight hardware-transplantation approach.**

***Index Terms*—multi-input multi-output (MIMO), human activity recognition (HAR), multi-channel fusion, lightweight neural network**

I. Introduction

With the continuous advancement in the field of artificial intelligence and the maturation of Internet of Things (IoT) technologies, human activity recognition (HAR) has emerged as a subject of significant interest for further research in various domains such as public safety, smart elderly care, interactive gaming, and traffic monitoring [1]–[4]. Among the various methods of HAR, the use of millimeter-wave radar technology has gained favor due to its advantages of continuous monitoring, non-intrusive nature, and lack of privacy concerns. The implementation of millimeter-wave radar for HAR currently involves several approaches.

The first kind of approach involves the application of logical judgments and threshold detection, which offers simplicity and rapid response processing. J. Baik *et al.* proposed a method that increases the number of judgment threshold features by adding two additional detection features, which are centroid distance and distance width, to the conventional speed and acceleration, achieving certain classification results using these four features [5]. However, the primary limitation of this method is the poor classification performance due to false negatives and false positives, and the number and accuracy of thresholds also affect the performance of algorithm.

The second kind of approach utilizes traditional machine learning methods, relying on features extracted from sliding windows and classic machine learning algorithms. A. Shrestha *et al.* processed continuous data using a sliding window approach and introduced Sequential Forward Selection (SFS) as a feature selection tool to further optimize the classification performance of the Support Vector Machine (SVM) [6]. F. J. Abdu *et al.* employed Canonical Correlation Analysis (CCA) algorithm in conjunction with the SVM classifier to create a special discriminant vector for activity recognition [7]. Nevertheless, traditional machine learning methods are plagued by the complexities of manual feature extraction and performance constraints.

Lastly, there is the increasingly popular method of deep learning, with an increasing number of high-performance networks being applied to HAR. This classification approach has gradually become mainstream [8]–[12]. J. Maitre *et al.* proposed a deep neural network (DNN) model consisting of a convolutional neural network (CNN), a long-short-term memory (LSTM) network, and a fully connected neural network for HAR [8]. H. Li *et al.* proposed a framework based on a bidirectional LSTM network that integrates various methods for multi-modal sensor fusion in HAR [9]. Deep learning eliminates the tedious step of feature extraction in traditional machine learning, allowing raw datasets to be trained directly. However, datasets are typically obtained from single-channel radar data, which fails to effectively utilize the target feature information captured by MIMO radar.

Building upon this research, this paper presents a human activity recognition algorithm based on MIMO millimeter-wave radar. Specifically, after necessary preprocessing of the multi-channel echo data acquired by the MIMO radar, the data are classified through three independent, lightweight MobileNetV3 network models. Subsequently, a hard classification algorithm is utilized to integrate the multi-channel information, yielding the final recognition outcomes. To validate recognition performance of the algorithm, a dataset comprising 600 groups of each of the six common human activities is collected for verification. Experimental results indicate that the algorithm significantly improves recognition accuracy compared to tra-

This work was supported by the Municipal Government of Quzhou under Grant 2023D032.

ditional single-channel HAR algorithms. Simultaneously, our comparative experiments demonstrate that the proposed algorithm is a viable lightweight hardware-transplantation method.

II. Experimental Setup and Data Processing

A. Experimental Setup

The primary hardware components utilized in this experiment include the 60 GHz millimeter-wave radar sensor IWR6843, introduced by Texas Instruments (TI), and the DCA1000 data capture card, as illustrated in Fig. 1.

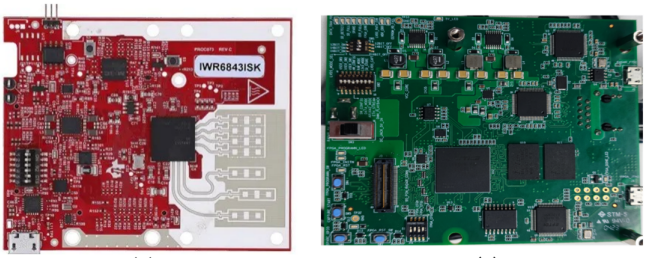

Fig. 1. Experimental instruments: (a) IWR6843 millimeter-wave radar and (b) DCA1000 data capture card

TABLE I presents the settings for the key parameters of the experimental radar system. Each frame of data lasts for 50 milliseconds, with a total of 100 frames collected in a single trial. The duration of data collection for each experiment is 5 seconds. The experiment employs a configuration of 3 transmitting and 4 receiving antennas, resulting in 12 data collection channels.

TABLE I
Experimental Radar System Parameter Settings

Start Frequency	60 GHz
Continuous Frequency Modulation Signal Bandwidth	1200 MHz
Continuous Frequency Modulation Signal Period	50 μs
Number of Transmitter and Receiver Antennas	3 transmitters 4 receivers
ADC Sampling Frequency	2 MHz
Radar Sampling Points	64
Continuous Frequency Modulation Slope	30.018 MHz/μs
Signal Rise Time	40 μs
Signal Sampling Gap	10 μs
Frame Sampling Periods	128 chirps per frame
Maximum Detectable Distance	10 m
Maximum Measurable Velocity	8.33 m/s

In this experiment, we focus on the multi-classification of six common human behaviors, which include waving hand, standing up, sitting down, falling, getting up and sleeping. Illustrations of these six behaviors are shown in Fig. 2. To enhance the generalization ability of the recognition results, all behavior data collection in this study is performed from different angles and distances. The specific experimental scenario model is depicted in Fig. 3, where the experiment is conducted in a relatively open space. All human behaviors are collected within a 5*5 meter square area, with the radar positioned at a height of 2.8 meters above the ground.

There are 10 participants, with heights ranging from 1.6 m to 1.85 m. Each participant performs behaviors facing the radar, but their orientation relative to the radar is random. The specific behavior dataset is introduced in TABLE II. While

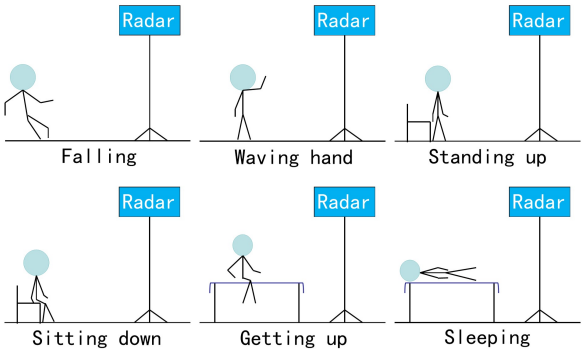

Fig. 2. Schematic diagram of six common human behaviors

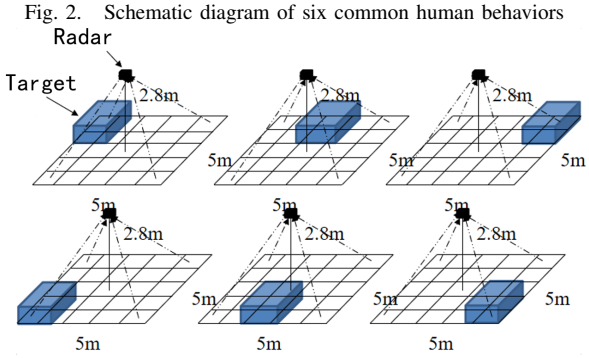

Fig. 3. Experimental scenario model

the radar detection range may have included interference from other static objects, the presence of rapidly moving dynamic objects is avoided. The actual experimental setup is depicted in Fig. 4.

TABLE II
Experimental Radar System Parameter Settings

Behavior Types Collected	6
Number of Targets	10 individuals
Data Volume per Single Target and Single Behavior	60
Target Activity Range	5 m*5 m
Radar Height from Ground	2.8 m
Total Dataset Data Volume	3600

Fig. 4. Actual experimental scenes: (a) empty scene and (b) HAR scene

B. Experimental Data Processing

In the 3-transmit 4-receive mode used in this experiment, the radar RF transmission process is altered compared to the single-transmit mode. The radar divides each chirp into 3 time blocks, and during different time blocks, the 3 TX antennas sequentially transmit their signals, while the 4 RX antennas simultaneously receive the echo signals. This allows for radar

data collection under the 3-transmit 4-receive mode. After the data obtained from each channel is stored as a row vector in a specific order, we obtain the initial single-pulse sampling data matrix.

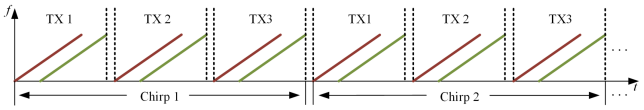

Fig. 5. Radar signal transmission model in 3-transmit 4-receive mode

After preprocessing the intermediate frequency signals and reshaping the data matrix, this paper obtains a radar three-dimensional data matrix. For the data from each channel, the Fast Fourier Transform (FFT) in the range dimension is performed, which is applied to each row of the data matrix. The specific expression for the FFT is as follows:

$$TR_{(m,k)} = \sum_{u=1}^{N_s} \|W_u S_{(n-u,m)} e^{-j2\pi ku/N}\|^2 \quad (1)$$

where $TR_{(m,k)}$ represents the amplitude of the mth chirp after discrete FFT at the digital frequency point $2\pi k/N$,where N is the number of points for the discrete FFT, W_u is a predefined Hamming window function, and $S_{(n-u,m)}$ denotes the data of the $n-u$ sample point of the mth chirp, where n represents the sample point, and u is the integration variable. The obtained Range-Time map is shown in Fig. 6.

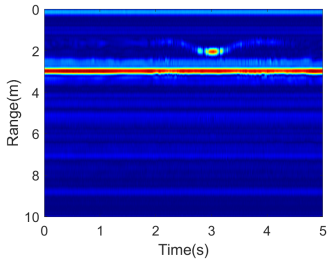

Fig. 6. Obtained Range-Time map

In practice, during the radar data collection process, in addition to human targets, many other objects can interfere with the echo signals, affecting the identification of the distance component. For these constant direct current (DC) components that are present in each collection, this paper employs intra-frame cancellation methods to remove them. The calculation formula is as follows:

$$MTI_{(m,k)} = TR_{(m,k)} - TR_{(m+j,k)} \quad (2)$$

where $MTI_{(m,k)}$ represents the Range-Time map after cancellation, and j denotes the span of continuous frequency modulation for each cancellation. After the aforementioned processing, the final Range-Time map can be obtained, as shown in Fig. 7.

After the application of range-FFT to the matrix data of each frame following clutter suppression, the sum of each row is computed. Subsequently, the maximum value is sought across each column, and the index of this maximum value is interpreted as the target position within the frame. A Doppler dimension FFT is then performed at the target position of each

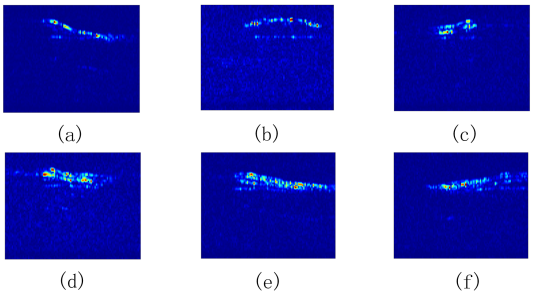

Fig. 7. Processed Range-Time map for 6 types of behaviors: (a) falling (b) waving hand (c) standing up (d) sitting down (e) sleeping (f) getting up

frame. Through a process analogous to the acquisition of the Range-Time map described previously, a Frequency-Time map is obtained.

III. Multi-Channel Fusion Algorithm for Human Behavior Information Based on MobileNetV3

A. Network Classification Architecture Based on MobileNetV3

To enhance the generalization ability of the algorithm, this paper proposes to use the MobileNetV3 network, which has shown superior performance in the field of lightweight classification, as the basic architecture for the classification network.

The core idea of the MobileNetV3 is to use depthwise separable convolutions to replace the traditional pointwise convolution methods, thereby significantly reducing the computational complexity and achieving network lightweighting. Depthwise separable convolutions consist of two steps: depthwise convolution and pointwise convolution, as illustrated in Fig. 8. Depthwise convolution refers to the application of convolution to each input channel separately, resulting in multiple output channels. Pointwise convolution involves convolving each output channel, thus merging the output channels from multiple depthwise convolutions. This method of depthwise separable convolution reduces the amount of computation and the number of parameters while maintaining good accuracy and generalization capabilities.

The core structure of MobileNetV3 is the bottleneck (bneck) structure, as shown in Fig. 9. The structure begins with an expanded convolution, which increases the number of channels through a 1*1 convolution. This is followed by a depthwise separable convolution. MobileNetV3 then introduces a Squeeze-and-Excitation (SE) channel attention mechanism. After passing through the residual structure of the linear bottleneck layer, the h-swish activation function is used [13]. This maintains performance while improving computational efficiency.

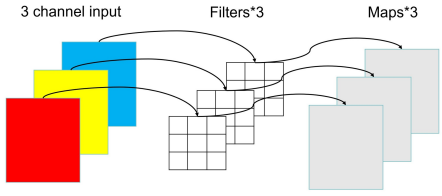

Fig. 8. Depthwise convolution in depthwise separable convolution

Compared to traditional CNN represented by ResNet, MobileNetV3, through techniques such as depthwise separable

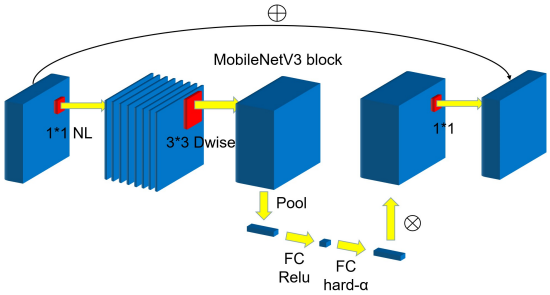

Fig. 9. MobileNetV3 bneck structure

convolutions and the Squeeze-and-Excitation (SE) channel attention mechanism, can significantly reduce computational complexity and the number of parameters while maintaining good accuracy and generalization capabilities. This has made it one of the important models for efficiently performing visual tasks such as image classification and object detection on mobile devices.

B. Multi-Channel Information Fusion Algorithm Based on Hard Classification

In typical HAR tasks, limited by radar hardware or other reasons, other algorithms often use single-channel radar data for HAR, which can lead to two issues. Firstly, single-channel radar systems can only provide information from a single angle or direction, which is insufficient for acquiring a comprehensive spatial understanding of target behaviors. Secondly, datasets acquired using a single radar are prone to environmental noise and interference, leading to significant fluctuations in training and testing results, particularly on more lightweight network architectures. By integrating multi-channel data for HAR, we can not only extract target behavior features from different orientations and at multiple levels but also enhance the generalization ability of the HAR task through certain channel information fusion algorithms, thereby improving task performance.

For the 3-transmit 4-receive MIMO radar used in this experiment, as shown in Fig. 10, in order to fully obtain information from different angles, the experiment plans to use the data from channels 1, 8, and 11 to synthesize target information. As illustrated, these three channels can fully capture target behavior information from multiple angles for HAR. After appropriate preprocessing, the information from the three channels is processed through three independent MobileNetV3 network classification architectures for classification training.

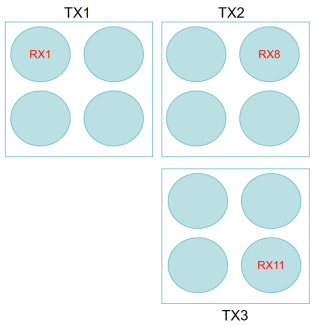

Fig. 10. Data channels used in 3-transmit 4-receive MIMO radar

Hard classification is a fundamental classification method in the fields of machine learning and pattern recognition, where each sample is assigned a clear category label. For a given input sample, a hard classifier outputs a category, usually the index or name of the category. Compared to the soft classification used in neural network training, hard classification has a stronger ability to resist overfitting in small datasets. Moreover, in the recognition of easily confused human behaviors, hard classification can better integrate the target feature differences from different channels, thus increasing the recognition accuracy of easily confused behaviors.

The experimental multi-channel information fusion algorithm based on hard classification decision is illustrated in Fig. 11. Initially, the range echo data from different channels are classified using independent network architectures to obtain decision vectors. Subsequently, a non-probabilistic model is applied to the decision vectors of each channel, selecting the decision vector that appears most frequently among those representing different target behaviors as the final decision vector, which is then used for the multi-class classification task of behavioral targets.

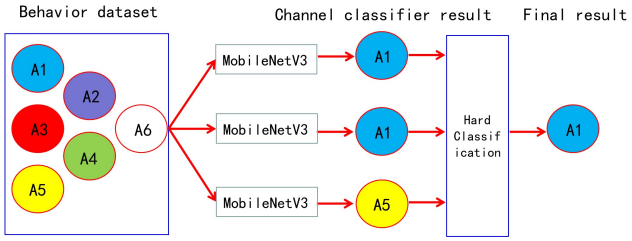

Fig. 11. Multi-channel information fusion model based on hard classification

In cases where discrimination is not possible, a further hard classification is performed on the frequency-time maps from different channels to obtain classification results. If discrimination still fails, a reclassification based on the independent lightweight network model for different channels is conducted for the easily confused behaviors recorded in the aforementioned two rounds of decision-making. Unlike the previous two rounds of discrimination, in this round, the test set categories are only labeled with the recorded easily confused behaviors, allowing for a targeted hard classification to obtain the final decision label. The specific multi-round hard classification process is depicted in Fig. 12. The aforementioned method not only comprehensively integrates multi-channel information for more accurate judgments but also synthesizes multi-domain information and addresses easily confused human behaviors, thereby effectively achieving the task of HAR.

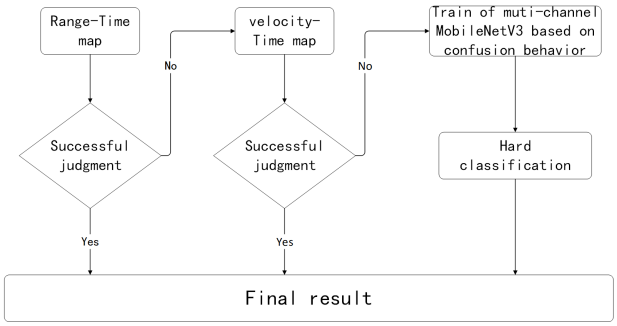

Fig. 12. Multi-round hard classification

IV. Experimental Results and Analysis

A. Network Training Details

Regarding the division of the training set, for the 6 types of behaviors, each behavior has a dataset of 600. The experiment uses a 7:3 ratio to divide the training and test sets, meaning each behavior has a training set of 420 and a test set of 180. The experiment samples using a uniform sampling scheme to ensure the generalization ability of the experiment.

As for the setting of experimental hyperparameters, the experiment uses the Adam optimizer and cross-entropy loss function for training. The training learning rate is set to 0.001, with a total of 100 epochs for training. The GPU device used for training is the RTX4070Ti.

B. Ablation Experiment Results of Single-Channel and Multi-Channel

In the ablation experiments, we conduct classification experiments using the MobileNetV3-s network architecture for a single channel. To mitigate the potential impact of insufficient data volume on the results of the single-channel ablation experiment, we integrate the data from three channels as the training and test sets for the single-channel experiment, thereby ensuring that the training data volume for the multi-channel and single-channel experiments is identical. The confusion matrices for multi-channel classification based on hard classification algorithm decisions and single-channel classification are depicted in Fig. 13(a) and (b) respectively. The average accuracy comparison is presented in TABLE III.

We can observe that, under the condition of equal training data volume, the multi-channel classification based on hard classification algorithm decisions not only comprehensively integrates target behavior information but also demonstrates a significant improvement in HAR accuracy. Moreover, it effectively enhances the generalization ability of recognition.

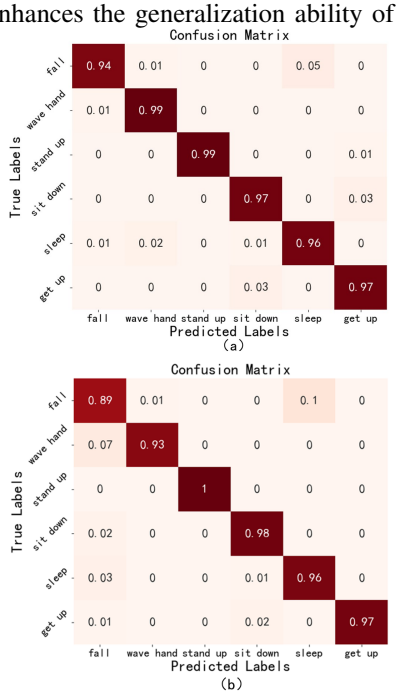

Fig. 13. Ablation experiment classification confusion matrices: (a) multi-channel classification and (b) single-channel classification

TABLE III
Ablation Experiment Average Training Accuracy Comparison

Different Channel MobileNetV3 Model	Average Accuracy
Three-Channel Fusion	97.05%
Channel 1	94.04%
Channel 8	94.52%
Channel 11	93.25%

C. Comparative Experiment Results with MobileNetV3 and Different Network

For the comparative experiment, we select several network structures that are popular in lightweight networks, including GhostNet and EfficientNetv2-s, as well as traditional CNN networks such as ResNet18, and Vision Transformer (ViT) networks with transformer structures. Fig. 14 display the classification confusion matrices for MobileNetV3-s, GhostNet, EfficientNetv2-s, ResNet18, and ViT networks. TABLE IV presents the comparison of average training accuracy, the number of parameters and floating point operations (FLOPs) for each network model. It can be seen that MobileNetV3 achieves a high classification accuracy with extremely small computational costs, demonstrating unique advantages in scenarios where radar device integration and small data processing are required.

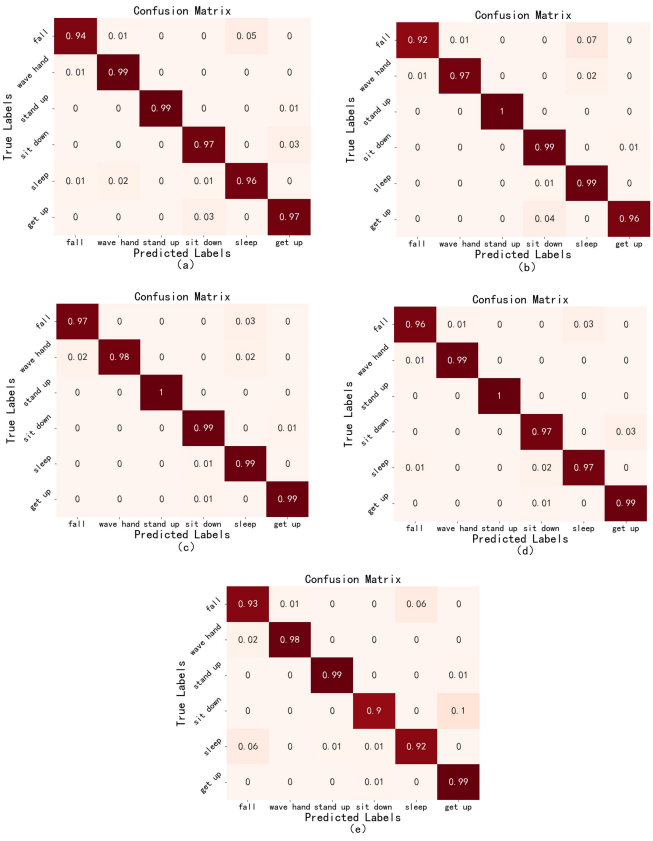

Fig. 14. Comparative experiment classification confusion matrices: (a) MobileNetV3-s (b) GhostNet (c) EfficientNetv2-s (d) ResNet18 (e) ViT

TABLE IV
COMPARISON OF AVERAGE TRAINING ACCURACY AND NUMBER OF PARAMETERS AND FLOPS FOR EACH NETWORK MODEL

Network Model	Average Accuracy	Parameters	FLOPs
MobileNetV3-s	97.05%	1.62M	0.061G
GhostNet	96.75%	4.40M	0.22G
EfficientNetv2-s	97.42%	20.19M	2.89G
ResNet18	97.13%	11.2M	1.82G
ViT	95.23%	86.5M	16.86G

In summary, through comparative and ablation experiments, it can be observed that the multi-channel classification algorithm based on MobileNetV3 proposed in this paper achieves superior HAR performance with a smaller computational cost in application scenarios involving embedded devices and small datasets. Compared to other traditional HAR algorithms, it exhibits strong advantages.

V. CONCLUSIONS

This paper proposes an algorithm for HAR using MIMO millimeter-wave radar, which is based on the MobileNetV3 network and multi-channel information fusion. Through experiments, it can be observed that the proposed multi-channel information fusion algorithm outperforms traditional single-channel algorithms in terms of recognition accuracy and stability and the recognition accuracy reaches 97.05%. For comparative experiments, the MobileNetV3 network achieves a relatively high recognition accuracy with smaller parameter count and FLOPs compared with other popular HAR classification networks, indicating that the MobileNetV3 satisfies the condition of achieving superior HAR classification with minimal computational cost, and represents a viable lightweight hardware-transplantation recognition approach.

REFERENCES

[1] Y.Qiao, J.Luo, T.Cui, *et al,* "Soft Electronics for Health Monitoring Assisted by Machine Learning,"*Nano-Micro Letters*,2023,15(5): 83-168.

[2] Yussof, Salleh, Tokhi, *et al,* "Towards understanding on the development of wearable fall detection an experimental approach,"*IEEE Sensors Journal*,2022,12(2):345-358.

[3] M.Vito, A.Federica, P.Luca, *et al,* "Biomechanical Measures for Fall Risk Assessment and Fall Detection in People with Transfemoral Amputations for the Next-Generation Prostheses: A Scoping Review,"*Journal of prosthetics and orthotics: JPO*,2022,34(3):144-162.

[4] A.B.H.Khaled, A.Khalfallah, M.S.Bouhlel, "Systematic review of indoor fall detection systems for the elderly using Kinect,"*International journal of telemedicine and clinical practices*, 2022,3(4):276-301.

[5] J.Baik, C.Jung, A.Nam, *et al, Fall detection and reducing detection error using FMCW radar*.2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). IEEE, 2022, pp.553-555.

[6] H.Li, A.Shrestha, H.Heidari, *et al,Activities recognition and fall detection in continuous data streams using radar sensor*. 2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC). IEEE, 2019,1, pp.1-4.

[7] F.J.Abdu, Y.Zhang, Z.Deng, "Activity classification based on feature fusion of FMCW radar human motion micro-Doppler signatures,"*IEEE Sensors Journal*, 2022, 22(9): 8648-8662.

[8] J.Maitre, K.Bouchard, S.Gaboury, "Fall detection with UWB radars and CNN-LSTM architecture," *IEEE Journal of Biomedical and Health Informatics*, 2020, 25(4): 1273-1283.

[9] H.Li, A.Shrestha, H.Heidari, *et al,* "Bi-LSTM network for multimodal continuous human activity recognition and fall detection,"*IEEE Sensors Journal*,2019, 20(3): 1191-1201.

[10] T.Liang, H.Xu, *A posture recognition-based fall detection system using a 24GHz CMOS FMCW radar SoC*. 2021 IEEE MTT-S International Wireless Symposium (IWS). IEEE,2021, pp. 1-3.

[11] A.S.Shah, A.Tahir, J.Le Kernec, *et al,* "Data portability for activities of daily living and fall detection in different environments using radar micro-doppler,"*Neural Computing and Applications*, 2022, 34(10): 7933-7953.

[12] X.Yu, C.Feng, L.Yang, *et al,* "Human Fall Detection by FMCW Radar Based on Time-Varying Range-Doppler Features,"*International Journal of Computer and Systems Engineering*,2022, pp. 318-323.

[13] A.Howard, M.Sandler, G.Chu, *et al, Searching for mobilenetv3*. Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.