Millimeter wave radar-based method for detecting typical dangerous driving manoeuvres

Zhanjun Hao (zhanjunhao@126.com)
Northwest Normal University

Zepei Li
Northwest Normal University

Xiaochao Dang
Northwest Normal University

Zhongyu Ma
Northwest Normal University

Research Article

Keywords: Dangerous driving warning, FMCW radar, eigenvector, micro Doppler spectrum, SlowFast CNN

Posted Date: March 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1375250/v1

Millimeter wave radar-based method for detecting typical dangerous driving manoeuvres

Zhanjun Hao,[1,2,*] Zepei Li,[1] Xiaochao Dang,[1,2] and Zhongyu Ma[1]
[1]College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu 730070, China
[2]Gansu Province Internet of 'ings Engineering Research Center, Northwest Normal University, Lanzhou, Gansu 730070, China
[*]Corresponding authord
Zhanjun Hao; zhanjunhao@126.com

***Abstract*—In response to the dangerous driving behaviour prevalent in current car drivers, there is a need to accurately warn and correct the driver's driving behaviour in real time in a small, mobile, enclosed space. This paper proposes a method of hazardous action detection based on Millimeter wave radar. A millimeter wave radar chip is installed in the cab and the collected FMCW differential frequency signal is Fourier transformed to obtain the IF signal, followed by the target action micro Doppler eigen spectrum as the eigen value. By parsing the matrix signal intensity of the signal echo model of the input channel, the synthesised waveform passing through the mixer is fed into a multi-layer filter for spurious filtering, and the lateral fusion method using the optimised deep learning SlowFast CNN is used to estimate and classify the feature vector against each action. In this paper, nine typical dangerous driving manoeuvres are experimentally set up and the accuracy of the classification results of the self-created dataset is analysed, verifying that the method is highly robust and the accuracy of the target manoeuvre recognition can reach up to 97%.**

***Index Terms*—Dangerous driving warning, FMCW radar, eigenvector, micro Doppler spectrum, SlowFast CNN**

I. INTRODUCTION

Motion recognition can improve the proactivity and safety of human-computer interaction depending on the depth of research, as it can respond to the state of the human being at a given moment in real time. Motion recognition in entertainment applications is manifested in immersive VR wearable following experiences that rely on individual body movements to control virtual space tasks 3D movement tasks; In terms of safety, this includes motion classification based fall detection for the elderly and recovery assistance for patients, while at the same time dangerous driving actions recognition has received attention from researchers for its specificity. Space constraints and in-vehicle conditions of the restrictions. Therefore,the research and development of action recognition technology has facilitated the development and application of wireless communication in cross-scene conditions.

The current popular motion recognition technology is a method of receiving signal strength [1]（RSS）and Based on channel status information[2]（CSI）etc，as well as wearing conventional wearable sensors on the human body to detect[3] local movements of the target can achieve superior recognition results. However, as wearable sensors can be a hindrance to the wearer's movement, Wi-Fi technology is making a splash in indoor sensing and positioning applications due to its universality and the advantage of not needing to be worn[4]. The action recognition in the cab is distinguished from traditional technology by the small space, the density of people and the greater accuracy of the conditions that limit it[5].

In order to accurately perceive the minute movements of indoor target personnel，there are currently techniques for signal typical identification of target movements of closed carriers，the first being the more common LIDAR, commonly known as the four-line LIDAR Ibeo and the 3D LIDAR Velodyne, Linear interpolation based on LIDAR using obstacle occupancy grid maps and search continuum algorithms A^*,Currently widely used in the traffic sector for vehicle obstacle avoidance systems[6]；The second is sensor sensing technology, such as IMU sensors to sense the surrounding environment and capture important information such as visual images of the target, but more expensive, where the capture of image information data is subject to diffuse reflection of the target shape, shadows, gullies and loss of information, and also

The date to submit your paper for review is December 13, 2021 This paper is supported by the National Nature Fund Project (61762079), Gansu Science and Technology support key Research and Development Program (20YF8GA048) and the 2019 "Light of the West" talent project of the Chinese Academy of Sciences.

Zhanjun Hao Auther is with the Northwest Normal University - China Northwest Normal Univ, Coll Comp Sci & Engn, Lanzhou 730070, Peoples R China. (e-mail: zhanjunhao@126.com).
Zepei Li Auther is with the Northwest Normal University - China Northwest Normal Univ, Coll Comp Sci & Engn, Lanzhou 730070, Peoples R China. (e-mail: 15503608775@163.com).
Xiaochao Dang Auther is with the Northwest Normal University - China Northwest Normal Univ, Coll Comp Sci & Engn, Lanzhou 730070, Peoples R China. (e-mail: dangxc@nwnu.edu.cn).
Zhongyu Ma Auther is with the Northwest Normal University - China Northwest Normal Univ, Coll Comp Sci & Engn, Lanzhou 730070, Peoples R China. (e-mail: mazybg@nwnu.edu.cn).

susceptible to interference from light intensity indices[7]; The third is the millimeter wave passive sensing technique used in this paper, characterised by sparse data, high sensing accuracy and a waveform signal that is minimally affected by the environment[8] -[15], Precise perception in sensing specific environments.

II. Related Work

Most of the in-vehicle radar is used for early warning parking radar outside the vehicle, which belongs to the category of LIDAR（Laser Radar）, usually installed on the exterior of the vehicle for sensing changes in the surrounding environment. Traditional LIDAR has a better angular resolution than millimeter wave, but requires a software platform and a back-end server, which is expensive making the commercial cost a challenge. In the literature [16] Hong, Feng et al. proposed a codec-based YOLOv3 vehicle detection algorithm based on the problem of vehicle target detection in complex scenes. To solve the multidimensional vehicle target detection problem, the new multi-level feature pyramid structure incorporates a codec module for detecting vehicle targets at different scales. In the literature [17] D'Antonio, Erika et al. introduced the open source library OpenPose, two webcams and a linear triangulation algorithm in their study of markerless capture of human movement, choosing the OpenPose skeletal keypoint detection algorithm, using the angular path provided by OpenPose with the angular path approximation method calculated by the inertial sensors at all camera locations markerless Marker detection of human running behaviour. With innovations in detection technology and a growing awareness of privacy, drivers do not want to capture RGB image video information in a private space and the visual approach is vulnerable to light conditions in the cab. In the literature [18] and [19]Snowden R J et al. perceive the similarities and differences between the motion states of elderly and young people by means of visual detection, for random point patterns and spatial gratings for threshold indexing.

The millimeter wave FMCW radar can obtain the radial distance of the target by transmitting an FM continuous signal, and the distance-Doppler motion detection is achieved through the Chirp composition matrix [20]。Denoising also provides detection thresholds, and the radar itself performs constant false alarm detection (CFAR) in signal processing, using cell values to compare whether the detection cell is a target or not [21]。Since Wi-Fi is more difficult to extract features in complex conditions of spatial environment, this experiment proposes WiGig[22] technology based on FMCW radar for in-vehicle hazardous driving action recognition, extracting micro-Doppler features as feature values through radar signal processing, dividing the self-built dataset into 2 fast and slow channels based on data sparsity into CNN network, and matching the collected actions with the training set for The result is an early warning effect[23][24].

（Frequency Modulated Continuous Wave , FMCW）Based on the comparative analysis of the above methods, this paper uses a deep learning AlexNet framework to build a method to identify verification methods based on FM continuous wave echoes for waveform number-mode analysis, while traditional edgeless sensing techniques learn a priori from a large number of access points based on subcarriers and fingerprint libraries. The good angular resolution of the 77 GHz millimeter wave radar allows accurate Doppler mapping information to be obtained. The highly integrated nature of the radar allows for small antenna arrays and the concentrated arrangement of the antenna arrays allows for similar Doppler signatures at the same distance. Randomly distributed thermal noise features are eliminated by uncorrelated superposition of radar parallel channels to eliminate Doppler features of thermal noise at the same distance. The robustness of the driver's movements is significantly improved. And the high packaging saves development costs and time while saving space in the vehicle. WiGig is a key technology for high-bandwidth, high-rate communication based LANs and backbone networks, with WiGig messaging rates of up to gigabits. Real-time motion recognition features can be extracted efficiently, and features are extracted from micro-Doppler features rather than traditional signal amplitude features, thus improving feature accuracy [25].The results show that using a CNN two-channel deep neural network with micro-Doppler features as input after parameter estimation and image pre-processing of the target signal increases the network depth and is effective in classification [26][27].

III. Relevant Theory And Technology

A . FMCW radar-related theory

Millimeter wave radar as an integrated single chip processor based on FMCW（Frequency Modulated Continuous Wave）radar scanning technology with a high frequency linear continuous pulse of 4 GHz. The radar system includes the transmitter component TX and the receiver component RX RF component, in addition to important components such as the clock timing simulators TF-RF and RX-RF, the analogue-to-digital converter ADC, the microcontroller MCU and the digital signal processor DSP. The collected high frequency signals are connected to the user's experimental terminal through the port of the DCA1000EVM data capture card, and the collected signals are digitally resolved at the terminal. The parsing process is to store the collected echo information in ADC_data.bin file from 6 ADC channels to the internal memory with ECC, and reorganize it into data matrix according to the number of antennas of the received channels, and the data length of the received channels can be parsed out with radar. The basic radar parameters used to collect signals in this method are set to TX power 12.5dBm, maximum FM slope 100MHz/us, IF bandwidth 175KHz~5MHz, and maximum ADC sampling rate 12.5Msps (real)/6.25Msps (plural).

The FMCW radar transmit signal can be expressed as：

$$S_T(t) = A_T \cos\left[2\pi f_c t + 2\pi \int_0^t f_T(\tau) d\tau \right] \quad (1)$$

B is the maximum bandwidth, T is the period of the high-speed FM continuous wave, T_c is the pulse width of the linear FM wave, A_T is the amplitude of the transmitted signal, τ is the time delay between the transmitted signal and the return signal, S is the slope of the FM transmitted wave of the FMCW radar, c is the value of the speed of light and d is an unknown

number to indicate the distance of the radar from the target object.

The time node at which the FM continuous wave is transmitted to the moment at which the waveform is received produces a time delay Δt and a Doppler shift Δf_d due to the movement of the driver. With the model of the transmitted frequency signal, the received signal received by the receiving antenna is considered to change in the frequency domain and generate a time delay, after Fourier transform, so that the received signal can be derived as follows:

$$S_R(t) = A_R \cos\left[2\pi f_c(t-\Delta t) + 2\pi\int_0^t f_R(\tau)d\tau\right] \quad (2)$$

A_R indicates the amplitude of the received signal, Δt_f Indicates flight time，The first term in the expression is ignored because the time delay is very small due to the extremely fast propagation of electromagnetic waves. f_{IF} is only related to the number of sampling points x in the corresponding target FM period. The IWR1642's transmit system consists of two parallel transmit chains, each with a separate binary phase and amplitude control, supporting binary phase modulation. The receiver is made up of 4 parallel channels working in conjunction with each other, each with 5 sections: LAN, mixer, IF filtering, ADC and extraction. Each channel can be operated independently and can be powered down to disconnect a channel for a single line bus input quadrature mixer and direct IQ quadrature output. e.g. as shown in Fig.1, demonstrates the complete radar signal processing process.

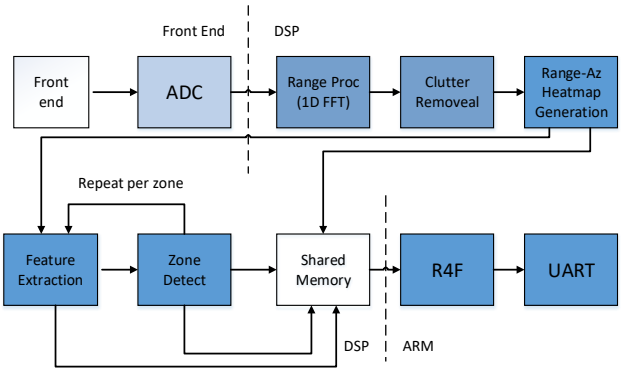

Fig. 1. Radar handling process.

Radar systems have a high recognition rate of radial distance resolution, the scattered signal is sparse and therefore the scattered centroid needs to be determined, the distance between the centroid and the radar antenna is also changing all the time when the target is moving, by collecting a large number of scattered centroids the dynamic trajectory of the target can be determined. Distance resolution is used to discern the spatial locality of multiple targets.

B. FM continuous echo correlation theory

The radar data cube is a three-dimensional dynamic feature,which includes distance, velocity and angle information. The required features can be extracted from them to form the radar data cube. One of the distance features uses the principle that the radar system's signal measurement of the distance to the target is based on the round-trip propagation time of the electromagnetic wave signal. Δt_d is the time delay of the signal propagation， c is the speed of light， The distance R between the radar and the target can be expressed as:

$$R = \frac{c \cdot \Delta t_d}{2} \quad (3)$$

In a millimeter wave radar system, Δt_d the time delay is so small that the frequency of the IF signal is used to represent the transmission time delay Δt_d for distance estimation to the target. The instantaneous frequency of the two signals synthesised into an IF signal is the difference in phase between the two input signal waveforms. The phase of the output signal X_{out} can therefore be expressed in the equation:

$$X_{out} = \sin\left[(\omega_1 - \omega_2)t + (\Phi_1 - \Phi_2)\right] \quad (4)$$

The linear FM pulses of TX and RX as a function of time when facing a single target can be represented as a visual graph showing the temporal frequency of the IF signal, where the transmitted and received linear FM pulses are parallel and the distance between the two lines is a nominal single tone signal value due to the characteristic principle of the device. The area where the waveform exists is restricted to the overlapping time series of TX transmit FM pulses and RX receive FM pulses, e.g. as shown in fig.2.

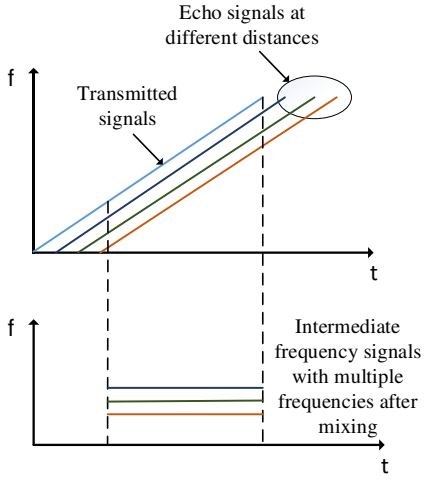

Fig. 2. IF signal generation process

The initial phase of the IF (Φ_0) is where the IF signal starts, and is also the difference between the phase of the TX transmit FM pulse and the RX receive FM pulse, expressed by the equation.

$$\Phi_0 = 2\pi f_c \tau = \frac{4\pi d}{\lambda} \quad (5)$$

The velocity characteristics are expressed in the FMCW radar system, which must transmit more than two linear FM continuous pulses with an interval of T_c. The transmitting linear FM pulses are processed with an FFT, which means that the target velocity is measured with a distance Fourier (distance FFT). The phase difference is derived from the characteristics of the signal, and the speed of the target's movement can be

deduced using the known phase difference, using the following formula:

$$v = \frac{\lambda \Delta \phi}{4\pi T_c} \quad (6)$$

where $\Delta\phi$ is the phase difference, T_c is the period of each chirp signal. A Doppler FFT was used to perform an FFT operation on the second vector of the N interval signal, which was used to strip two target objects. The target velocity interval Δv has a correlation with the frequency interval Δw of the target angular velocity. T_c is the interval period of the linear FM pulse, e.g. as shown in Fig.3. the interval of the angular velocity can be expressed by the equation:

$$\Delta w = \frac{4\pi \Delta v T_c}{\lambda} \quad (7)$$

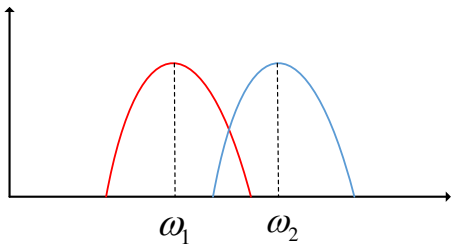

Fig. 3. Different target angular velocity intervals

The AOA angular signature is based on the observation that small changes in object distance lead to phase changes in the peak of the distance-Fast Fourier Transform or Doppler-Fast Fourier Transform. The differential Δd distance from the target to each antenna results in a phase change in the FFT peak, which is used to estimate the angle of arrival. Suppose the separation distance between the receiving antennas is d, which is known by the Pythagorean theorem $\Delta d = l\sin(\theta)$, the angle of arrival can be calculated as:

$$\theta = \sin^{-1}\left(\frac{\lambda \Delta \phi}{2\pi d}\right) \quad (8)$$

Velocity has a range interval to avoid velocity blur, the angle likewise has an interval to avoid field of view, the maximum angular field of view of the radar being determined by the maximum AOA that can be estimated. When $\Delta\phi > \pi$, this creates an angular ambiguity to the extent that it is impossible to determine the orientation of the target.

$$\Delta\phi = \frac{2\pi d \sin(\theta)}{\lambda} < \pi \Rightarrow \theta_{\max} = \sin^{-1}\left(\frac{\lambda}{2d}\right) \quad (9)$$

IV. Action Classification Models And Recognition Methods

A. Overall model flow

To address the needs and limitations of in-vehicle driver hazard action analysis, this paper proposes a two-channel deep convolutional neural network model based on micro-Doppler features. The antenna is packaged on a single chip with both transceiver and transmitter functions. The transmitted and reflected signals are mixed and then the A/D signals are processed and analysed in baseband by digital processing devices such as low-pass filtering. The Doppler feature spectrum is obtained according to the Fast Fourier Transform and input into this method using the model two-channel CNN for the classification of hazardous actions by micro-Doppler features, The overall process of the model is shown in Fig.4.

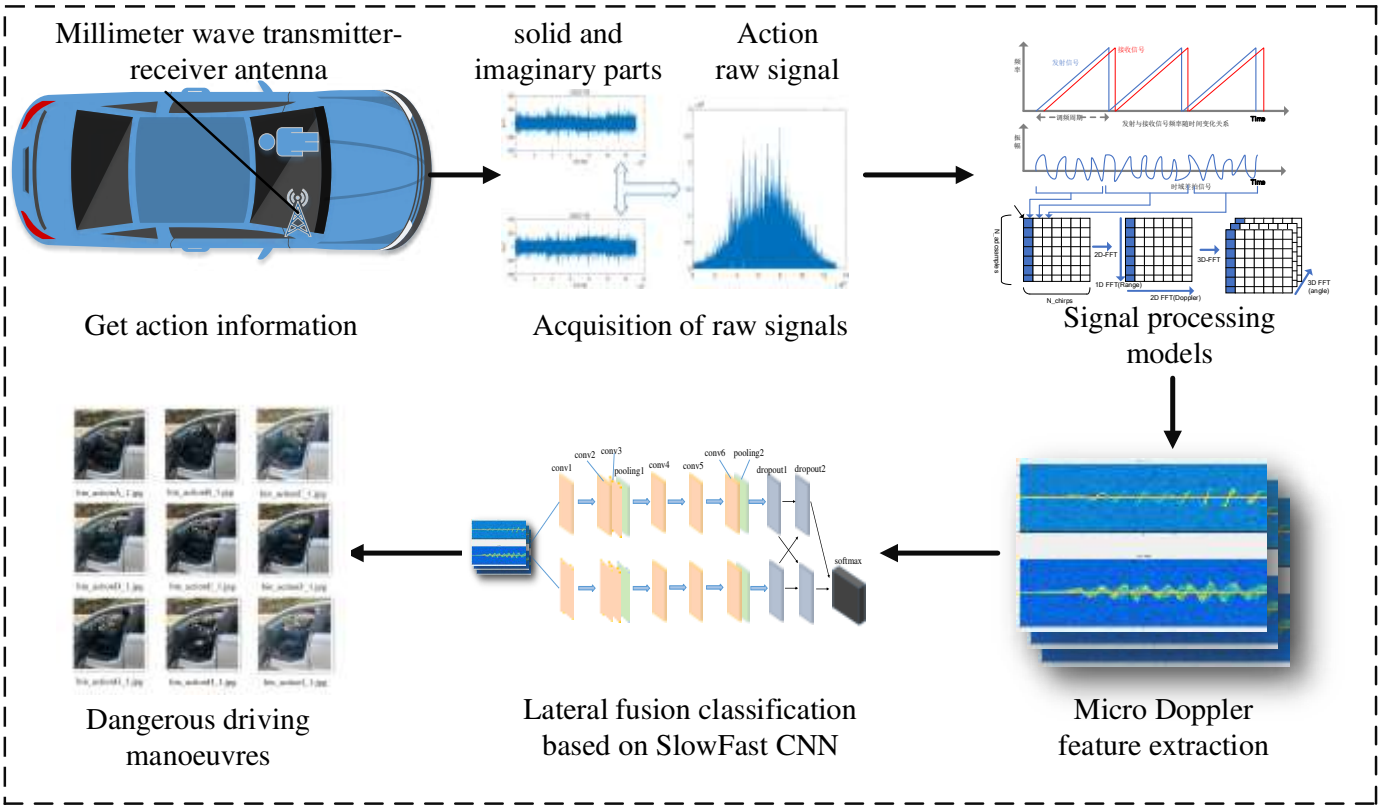

Fig. 4. Overall flow diagram of the model

B. *Doppler feature vector extraction*

In this research experiment, each frame contains 128 chirp signals, for which the chirp signal can be reflected and mixed by a mixer to obtain the intermediate frequency signal (IF), and each chirp signal can be manually set for the experiment with 128 sampling points. When using the chirp signal to represent each frame, a 128*128 matrix is used to represent it, and the frame format is composed in row*column numbering according to the row numbering of the IF signal. A Doppler FFT of the data sampling points in the intercepted frequency band of the chirp signal within the frame period of the acquired target action signal results in a distance Doppler distribution based on driving hazard manoeuvres. A one-dimensional chirp signal can be obtained as a distance Doppler spectrum by fast discrete Fourier transform, while a two-dimensional chirp signal can be obtained as a distance-velocity Doppler spectrum by fast discrete Fourier transform. The collected distance Doppler spectral distribution map will do constant false alarm rate (CFAR) on not only for a single target detection, can be in the signal reception of the separation of irrelevant target reflections and the processing of various intensities of thermal noise, and a single target velocity indexing, to get the results of 2DFFT, x-axis is the distance, y-axis is the velocity, z-axis is the signal intensity, from the projection to the horizontal plane corresponding to the peak in Fig.5, the distance and speed information of the target detection in Fig.6 can be obtained . For multiple targets, CFAR detection processing is required.

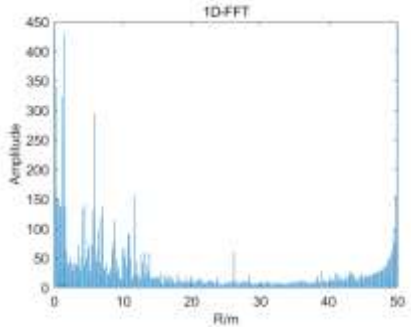

Fig. 5. Distance Doppler spectrum

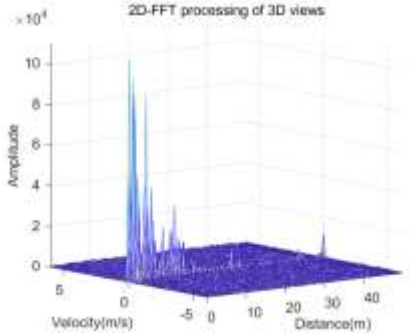

Fig. 6. Distance-velocity Doppler spectrum

The detection part of the target object is reflecting and inevitably scattering the signal emitted by the transmitter, so the receiver receives a relatively small quantitative change in the frequency of the echo information, which arises due to the different offset of the target velocity, such a process is called micro-Doppler information. The Doppler shift is invariant within a fixed FM period of the signal and the Δf_d is related to

the number of indexes y within the fixed FM period. The whole acquisition process takes place in the free time between the effective linear FM time and the end of the frame, i.e. the "inter-frame" period. From the above synthesis, it can be seen that the IF signal mixed with the FM signal in one frame period can be expressed in a discrete way as:

$$s_{IF}(x,y)=A_{IF}(x,y)\cdot\exp\left\{j2\pi\left[f_{IF}(x)-\Delta f_d(y)\right]x/F_s\right\} \quad (10)$$

Where $A_{IF}(x,y)$ denotes the amplitude of the IF signal. F_s indicates the sampling rate of the ADC analog to electrical conversion。Fourier one-dimensional transformation of the sampled chirp signal distributed according to the time domain characteristics in any FM period y of the acquired frame，Obtain the frequency of the transmitted signal in relation to the return IF signal $f_{IF}(x)(x=1,2,3,\mathrm{L}\ ,N_{adc})$ or velocity-dependent Doppler shift of the Fourier 2D transform of any chirp signal period sequence in the frame $\Delta f_d(y)\left(y=1,2,3,\mathrm{L}\ ,N_{chirp}\right)$， where N_{adc} is the number of samples per FM cycle， N_{chirp} is the FM signal period per frame. The feature vectors were extracted from the micro Doppler spectrograms for each manoeuvre, the fig.7 shows the left turn 90° micro Doppler features for the nine hazard manoeuvres in the dataset.

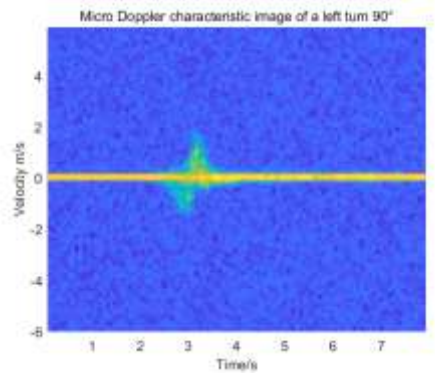

Fig. 7. 90° micro Doppler characteristic of the left rotating body

After the original amplitude and phase data is collected and mixed by the mixer, the resulting synthesized waveform information is fed into the filter, and the signal output from the filter can be expressed as $x(0),x(1),x(2),\mathrm{L}\ ,x(k)$ ， Representing the $x(k)$ expression as the set of filtered signals, with k representing the number of samples in the period of the exponential linear FM continuous wave signal, the filter expression is expressed as follows:

$$x(0)=f_B(0)+f_B(1)+f_B(2)+\mathrm{L}\ +f_B(N-1) \quad (11)$$

$$x(1)=f_B(0)+f_B(1)e^{-j\frac{2\pi}{N}}+\mathrm{L}\ +f_B(N-1)e^{-j\frac{2\pi}{N}(N-1)} \quad (12)$$

$$x(k)=f_B(0)+f_B(1)e^{-j\frac{2\pi}{N}k}+\mathrm{L}\ +f_B(N-1)e^{-j\frac{2\pi}{N}(N-1)} \quad (13)$$

C. Movement analysis and classification

1) *Rectangular action area characterisation detection*

In the data cube，Detection area defined by both distance and azimuth data，The range of values for the detection area is independent of the personnel, but is related to the installation position and orientation attitude of the radar signal transmitter receiver. Defining the driver's driving seat position as the target area for signal acquisition，Defining the target area in terms of range azimuth expressions can be expressed as：

$$Z_i=(n,m):r_l^j\le r_n<r_u^j \quad (14)$$

$$\theta_l^i\le\theta_m<\theta_u^i \quad (15)$$

Z_i is the area number in the set of all range-azimuths within the rectangular boundary of the target area,According to the definition of the four parameter variables r_l^j、r_u^j、θ_l^i、θ_u^i in Eq. Fig.8 shows locale dependencies.

Fig. 8. Experimental equipment area

To determine if there is a valid active human target in the detection target area， judgement by the energy-dense movement characteristics of the area for a given frame. It is clear from the target detection that the presence or absence of an effective target figure has a relatively large effect on the energy dense power of the area. The average zone-power energy (the average zone-power) of the i-th zone at frame t can be expressed as:

$$Q_i[t]=\frac{1}{|Z_i|}\sum S_{n,m}[t] \quad (16)$$

The quantities characterised in relation to the regional average power energy are: moving average power energy,average power ratioand regional power correlation coefficient. The moving average power energy is related to the length of the moving average window of the previous frame；The average power ratio indicates the ratio of a region to the total energy； The regional power correlation coefficient indicates the correlation between the different regions.

2) *Personnel saturation detection*

The radar transmitting component does not require dangerous behaviour detection of persons at vehicle stops, and detection in specific areas requires time and temporal constraints. Although the target area is a single detection location and there is no problem of exchanging arrangements of people, there is still a decision result, with two detection saturation states, occupied and unoccupied, indicated by "1" for occupied and "0" for empty. the distribution state of the drivers can be expressed by the expression:

$$O_i[t] \in 0,1 \qquad (17)$$

The definition of the spatial location probability distribution with driver and passenger at the target location can be expressed by the expression:

$$\hat{p}_{(O_1,O_{2\perp},O_N)}[t] = g\sum_{i=0}^{N} W_{(O_1,O_{2\perp},O_N)} i^{x[t]} \qquad (18)$$

$g(\cdot)$ is a non-linear function，logistic regression functions is $g(z) = \frac{1}{1+e-z}$.

D. CNN classification models

1) *The overall process of classification*

The CNNs used in this paper all use the underlying system structure of AlexNet, using ReLU as the activation function for this structure， The model consists of a two-channel convolutional network with six convolutional layers in one channel, two pooling layers, two fully connected layers, and finally a softmax layer to output the classification results. First the redundant and space-intensive LRN layer is removed, since the role of this layer is to make the larger values of the response larger，as small data parameters also take up storage and system learning time, the effect on classification appears to be very small and the operation is made to improve the efficiency of the system. The effect of blurring due to the traditional average pooling of convolutional neural networks is discarded and the reduced dimensional convolutional features are all represented by the maximum pooling operation. In the convolution kernel size setting, since the data set is the raw waveform signal of human action rather than image information，The size of the convolutional kernel is set to 3*3 to make the training process light and fast, as the classification area of the feature distribution is relatively small. As shown in Fig.9,the convolutional layers are both filled with SAME and the pooling layers are both VALID. Due to the small parameters of the convolutional kernel, in order to avoid overfitting, the

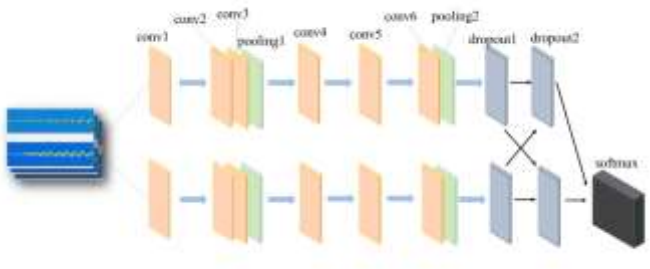

Fig. 9. CNN classification flow chart

features are extracted and the data structure is reconstructed and enhanced by adding a convolutional layer in front of each pooling layer to increase the network depth. raining data after feature extraction using multidimensional convolutional neural networks for inter-frame continuity of target actions, allowing for layered lateral fusion of data according to time series，the accuracy of the type of action and the convergence of the action timing can be substantially improved.

2) *Lateral fusion method based on SlowFast CNN*

In order to improve the recognition rate of action recognition and reduce feature data redundancy, this paper uses a two-channel lateral fusion deep convolutional neural network. The downscaled collected feature action information is fed into two parallel convolutional neural networks in a simultaneous sequence. According to the time speed drive can be divided into two channels, low frequency sampling channel to capture the slow target signal, the filtering of reflected clutter; fast sampling network channel to capture the target waveform larger waveform, so the channel is very lightweight, the acquisition of the main limb movements of the target, fast sampling channel in the process of running without the use of time pooling and step convolution, in order to maintain the accuracy of time details. The two channels are different time-speed driven rates, fusing the two fast and slow paths laterally, whether fast or low speed acquisition signal networks refer to 2D-B-AlexNet for the extraction of information in multi-dimensional space. The expression for the convolution kernel can be expressed as $\{T \times S^2, C\}$， where T 、 S 、 C denote the size of the temporal, spatial and channel respectively. The proposed lightweight model requires the convolution width of the fast target channel in the dual channel acquisition to be 1/2 of the low slow sampling rate, without compromising recognition accuracy, and sets the data in the sampling rate to 120 frames per second, with a setting of 120 indicating 6 frames per second acquisition.

processing of training samples with convolutional neural networks. The CNN lateral fusion process shown in Fig.10. **Step 1** is to set a uniform weight and a threshold value that best fits the classification, with the threshold value being artificially specified in the range (-1, 1). Low frequency sampling channel with slow frame refresh captures carrier information for channel sparse frames, high frequency sampling channel operates at fast refresh and high temporal

resolution，High frequency channels run in tiny steps of $\frac{\tau}{\alpha}$，$\alpha>1$ is the ratio of the frame rate between the fast and slow paths, Therefore, the channel density of the high speed path is α times higher than that of the low speed path，Set the value of α for this experiment to 8. Based on the channel capacity set in the laboratory, the fast path channel convolution kernel expression is $\{8T,\ S^2,C\}$, The slow path channel convolution kernel expression is $\{T,\ S^2,4C\}$，channel lateral fusion unifies the format of the two paths，two-channel fusion operation with the convolution kernel of $\{5\times1^2,2C\}$.

Step 2 The sample set is denoted as (x_i, y_i) , and the output bandwidth of the system network can be expressed as an expression based on the sample set as:

$$\hat{y}_j = f(\beta_j - \theta_j) \tag{19}$$

The $f(x)$ in the expression represents the ReLU activation function; θ_j denotes the threshold of the jth neuron in the layer that passes in the neural network.，β_j denotes the real-time input value of the same neuron in the transfer layer of the neural network. The expression for the input value can be expressed as:

$$\beta_j = \sum_{i=1}^{n} \omega_{ij} x_i \tag{20}$$

ω_{ij} represents the weighting of the weights between the i-th and j-th neurons of the output and hidden layers of the neural network.

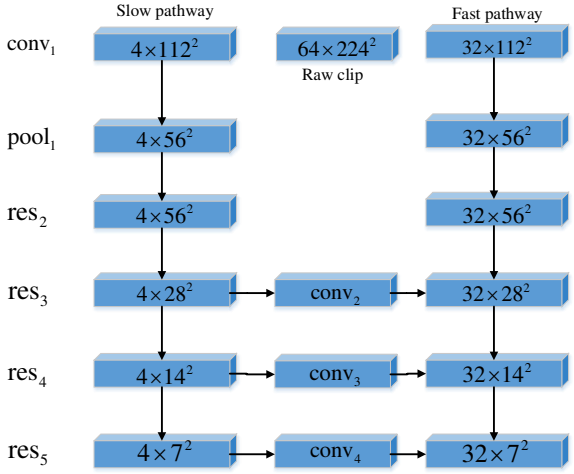

Fig. 10. Flowchart of SlowFast CNN for lateral fusion

Step 3 The mean squared error in the neural network matrix (x_i, y_i) is evaluated, and the specific calculation procedure is expressed by the expression：$E = \frac{1}{2}\sum_{j=1}^{m}\left(\hat{y}_j - y_j^i\right)^2$， Since the output of a convolutional neural network is similar to multiple output edges due to errors, $\hat{y}_j$ denotes the actual output and y_j^i denotes the desired output value.

Step 4 The input neural network is conditioned on the minimum value of the error to be bounded by the conditions allowed for learning, and on the minimum value of the threshold for the number of learnings. The condition is satisfied for normal learning, otherwise the weights of the CNN are updated in the direction of the bias of the target threshold. Let the learning rate of the CNN be η , the CNN weight update expression.：

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta \hat{y}_j \left(1 - \hat{y}_j\right)\left(y_j^i - \hat{y}_j\right) x_i \tag{21}$$

Step 5 Once the initialisation is complete, loop steps two, three and four and keep looping until the end if the conditions are not met. At the end of the loop the threshold and weights of the neural network must be determined.

V. EXPERIMENTAL DESIGN AND ANALYSIS

A. Experimental design

The experimental equipment in this paper is a Texas Instruments millimeter wave radar clocked at 77-81 GHz (corresponding to a wavelength of approximately 4 mm and a bandwidth of 4 GHz), which has a superior UHF bandwidth to conventional wireless sensing devices in terms of hardware, ensuring highly accurate recognition of movements. The equipment for the experiment consisted mainly of the IWR1642 radar sensor piece and DCA1000EVM data capture card, sensor mount, Lenovo R7000 i5-7300HQ GTX1050 802.11AC wireless network card laptop and two compartment car. Experimental scenario with the radar sensor fixed to the B-pillar on the driver's side of the vehicle in clear weather，height is 70 cm from the vertical height of the driver's seat，The radar transmitter and receiver are placed spatially perpendicular to each other in the direction of the antenna facing the driver's chest, placing the human body position exactly squarely within the sector of the radar scan for real-time acquisition of its overall behaviour. The experiment consisted of six persons of different attributes, with an age range distributed between 22 to 29 years and a range of 1.55 to 1.90 metres in height. The sample set for this experiment was a test set of 9 movements, the time taken for one acquisition was 6 seconds, and a person's particular target movement was repeatedly tested and executed 50 times, resulting in a total of 2700 sets of data information tested. The data sample for each set of data collected is 255*255*3. The Doppler spectrogram is generated by setting the Doppler sensing distance to 0.66 and 0.96 m, which results in two 2D FFT velocity profiles at two different distances，Accurate detection of the velocity of the target point by two fixed distances allows the effective manner and physical characteristics of the target's movements to be fixed.

Set the number of Chirp samples per unit to 128, the number of Chirp signals per unit frame to 128 and a fixed frame period of 30ms, With 200 frames captured per target driving

manoeuvre setting, each target manoeuvre has a capture time of 6.0s based on the frame period, with a speed resolution of 0.070m/s. The presence of thermal noise in an everyday air medium gives a value of ± 4.564m/s for the maximum ambiguity-free velocity based on the Boltzmann constant and the effective noise bandwidth.

The data collection part of the experiments in this paper was conducted entirely outdoors to classify the typical driving hazard manoeuvres of drivers, so the target for identification was the driver's single driving behaviour，According to the "Road Traffic Safety Law of the People's Republic of China" on the standardization of driver behavior requirements and in the daily bad driving habits of the questionnaire survey through the collation and summary, the main screening of bad driving behavior of a single person include: Operating the steering wheel with one hand (A), picking up objects with your head down (B), violent movements (C), drinking water while driving (D), answering the phone while driving (E), turning right horizontally and back (F), turning 90 degrees to the right (G), turning left horizontally and back (H), turning 90 degrees to the left (I) 9 main actions. In the detection of the driver's movements, the identity of the subject is not unique, including the human nature of men and women of different heights and sizes，Variable as a change in the identity label of the person under test, which must be set constant and quantitative in order to capture a specific waveform signal under different actions ，Invariants include device type、data acquisition time、number of frames、in the same scene, etc. During the recognition process, 50 samples of the same person are tested for each of the same movements, as shown in Fig.11.Due to the limited space in the vehicle, some movements of the limbs are obscured or not recognized due to interference.Recognition rate of 91.67% can be achieved，n situations where the background is not very complex, more standard actions can be recognized accurately in CNN，Experiments show that smaller data volumes can be effectively identified in CNNs with stronger action targeting and better generalisation.

Fig. 11. Example graph of driving behaviour in a dataset

In order to match the pictures to the pulse signal data, the naming of the pictures was standardised, e.g. the first set of data for single-handed driving is named bin_actionA_1.jpg, The photographs taken of driving movements are represented in the data acquisition transfer format bin_ and the letter of the movement. The advantage of this setup is that the data set can be collated later to avoid disordering the data image information and to reduce the hassle of manually labelling the information.

B. Experimental analysis

1) *Model performance analysis*

The network framework used in the collation of this experimental dataset uses the PyTorth model to build the overall system. Compilation software with Python 2.7 and Anaconda, PyTorth can support Python 2.7 environment configuration ， PackageManager selects conda by experimental science and feasibility, requiring the installation of the powerful data science toolkit Anaconda. Ten samples from each group in the training set of the motion detection experiment were selected as test samples for the training set for the data network framework update，this experiment uses a self-built dataset for experimental validation. The two-channel convolutional neural network mentioned above (including the low-speed sampling channel and the high-speed sampling channel) is shown Fig. 12 and Fig. 13 in the table below，In the experiments to verify the effect of the separated channels, it can be seen that both separated channels have very regular characteristic change curves by way of fusion collaboration.

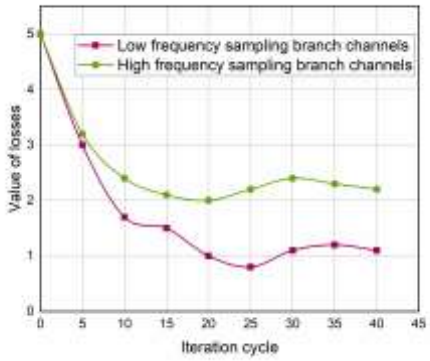

Fig. 12. Variation in sample loss rate

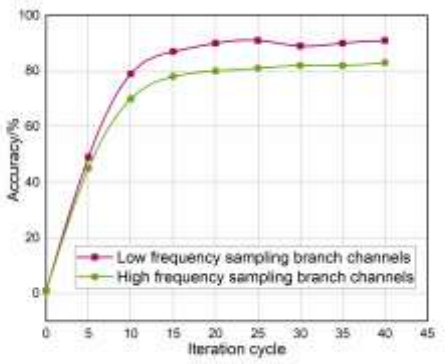

Fig. 13. Variation in sample accuracy rate

As can be seen from the graph, the low frequency sampling channel has a lower loss rate than the high frequency sampling channel when it picks up slow-motion micro-motions during the sampling signal，suggesting that a low sampling channel in the classification process will produce a lower number of failed samples， However, the low frequency sampling channel is not as accurate in sampling the signal as the high frequency sampling rate，It shows that the dense and fast set of data samples are more distinctive and have good accurate recognition rates. And as the iteration cycle progresses both the loss rate and the accuracy rate level off. Therefore changing the input channel to a dual GPU dual channel can improve the performance of the experimental model.

2) *Combined eigenvector analysis*

Six eigenvectors can be extracted from the collected micro-Doppler spectrogram, and the spectrogram consists of these six eigenvectors as the most important part， the different micro-Doppler spectrograms can be represented by a classification of the different values of the eigenvectors.Due to the different types of movements, eigenvectors exist that have a greater and lesser impact on the classification of the micro-Doppler spectrum. Very low impact feature vectors are removed based on feedback from the classification effector, thus avoiding overfitting in the classification process. 6 feature vectors can be freely combined to produce up to 720 feature combinations， The analysis of the accuracy of the classification effect of the feature vector and the combination of the features led to，Feature vector torso micro Doppler vector v_{torso} 、Shoulder micro Doppler vector $v_{shoulder}$ 、Arm micro Doppler vector $v_{\lim b}$ has a greater impact on classification results，Feature vectors U_1 、U_2 and U_3 are not as effective as the first three in terms of results, with U_2 and U_3 in particular having almost no effect on the classification results. Based on the above evaluation of the feature vectors, a good combination of feature vectors is selected. This includes a combination of 3 feature vectors: trunk micro Doppler vector v_{torso}, shoulder micro Doppler vector $v_{shoulder}$ and arm micro Doppler vector $v_{\lim b}$.

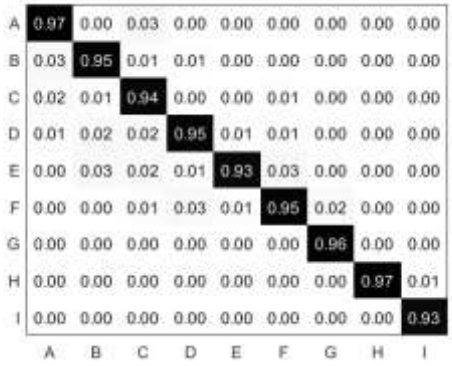

Fig. 14. Confusion matrix for each action test of the test for known persons

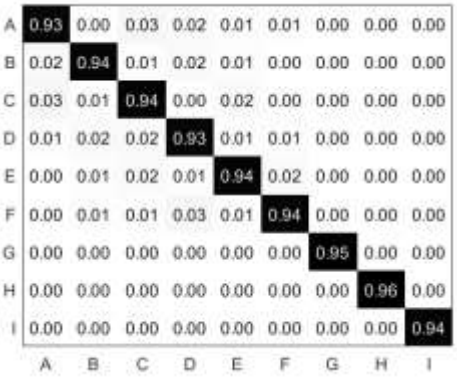

Fig. 15. Confusion matrix for each action test of the test for unknown persons

By combining good feature vectors, including the feature vector torso micro Doppler vector v_{torso} , shoulder micro Doppler vector $v_{shoulder}$, arm micro Doppler vector $v_{\lim b}$ Free vector combination. e.g. as shown in Fig.14.and 15,The nine dangerous actions in the dataset were identified for accuracy, and the classification effect accuracy was broken down into testing accuracy for known persons and testing accuracy for unknown persons. The confusion matrix is characterised by the fact that the values are simply concentrated on the main diagonal, with the horizontal and vertical axes representing the classification results and the true data type of the test dataset respectively. So the confusion matrix provides a visual representation of the distribution of the accuracy of each feature. The confusion matrix for each manoeuvre tested by known personnel shows that manoeuvre A (one-handed operation of the steering wheel) can achieve a maximum accuracy of 97%, and the confusion matrix for each manoeuvre tested by unknown persons shows that manoeuvre H (left horizontal and turn back) is accurate up to 96%. The minimum test accuracy is not less than 93%, so the individual algorithms in this paper are able to achieve the accuracy required for the detection of dangerous movements, both for known and unknown persons.

3) *Analysis of the impact of unrelated targets in the vehicle*

This experiment is independent of directionality as the driver's seat is fixed in direction to the radar transmitter end.

The effect of personnel irrelevance is produced in the car, and the passenger seat is analysed in relation to the effect of the irrelevant target person in the back row on the target. Four sets of comparison experiments were set up: no interference, interference from the passenger side, interference from the rear, and interference from both the passenger side and the rear. as can be seen from the graph below, the passenger seat and the rear row do not have a significant impact on the classification of dangerous movements of the target, as the millimeter wave radar produces different spectral maps for movements at different distances and has a strong directionality. The distance for target recognition is set from 1.1 m to 1.75 m. The distance for the passenger side is less than the distance for detecting the target and the distance for the rear row is greater than the distance for detecting the target, so the interference from irrelevant targets can be effectively removed. The effect of extraneous interference is shown in Fig.16.

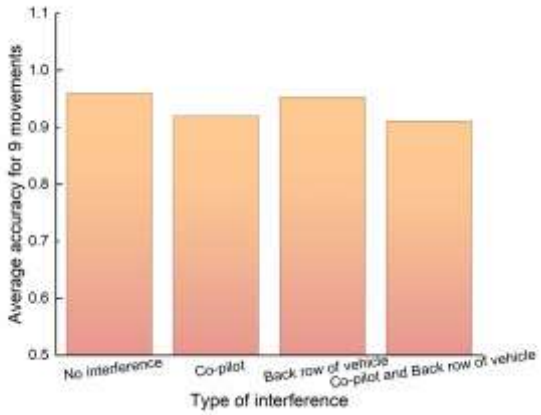

Fig. 16. Histogram of the accuracy of unrelated target interference

4) *Analysis of the diversity of personnel*

In the experimental set-up of six subjects with large differences in age, height and weight, the problem of individual driving habits and weight on the effective recognition distance at the radar receiver resulted in weak differences in the micro-Doppler spectrum of the input training set for the same manoeuvre for each subject. Fig.17 chart shows the box plots for the six subjects.

With the two-channel convolutional neural network model, each subject showed small variability for each of the nine actions in the experimental setup. Subject 2 had a low error for the same accuracy, with a maximum accuracy of 96.8% and a minimum accuracy of 95.8%. The highest accuracy rate of 87% could be achieved by Subject 3, and the lowest was 85.9%. Both can meet the accurate classification of dangerous driving manoeuvres, and the difference in accuracy rate for the same manoeuvre is very small, indicating the robustness of the method in this paper.

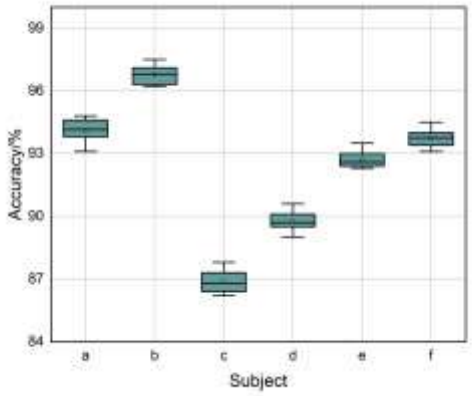

Fig. 17. Box line plot of 6 subjects for the same action

5) *Sample size optimisation analysis*

To determine the effect of sample size on movement accuracy, 50 sample test sets for each movement were tested in experiments with 10 and 100 sets of oversampled data under the same environmental conditions and personnel. The accuracy of the oversampled 100-group sample was found to be higher than the 10-group and 50-group training. Since 10

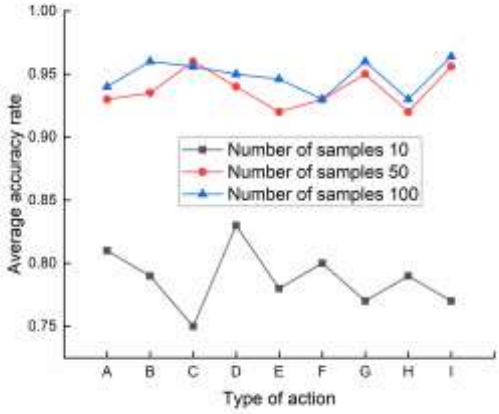

Fig. 18. Sample size analysis line graph

sets of training samples are not enough, 100 sets of oversampling can achieve the expected results, so an appropriate increase in the training Doppler spectrum can effectively improve the recognition accuracy. Fig.18 can reflect the relationship between the number of samples and the recognition accuracy.

6) *Analysis of identification results*

The hazardous action recognition method in this paper is a two-channel convolutional neural network with micro-Doppler features. In terms of vector features, it is possible to extract a spectral map of minute action features of the target action, and it is possible to carry out hierarchical and optimised classification of coarse and fine granularity based on the micro-Doppler spectral map, which can meet the expected requirements for driving warnings. The literature [28] uses CNN single-feature time-distance feature classification, and there is a large loss in classification accuracy when the micro-Doppler feature vector dataset of this

method is placed into this method, and the classifier is unable to validate the classification of head motion and shoulder micro-Doppler vectors for individual movements due to the small difference in distance. The literature10 uses ST-GCN to classify key points of the human skeleton, and due to the large variation in the physical signs of the experimenters, the micro-Doppler data of this method can only achieve a maximum of 89.4% classification of the graphical convolutional network. According to the following figure, it can be concluded that the method of this paper has a significant effect on the classification of dangerous driving manoeuvres. The following Fig.19.shows the experimental results of the different movement methods compared on the data set of this paper.

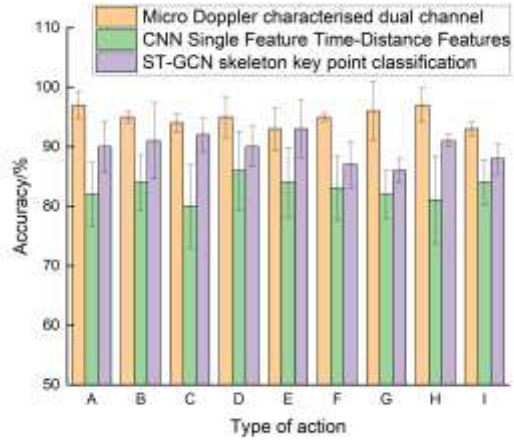

Fig. 19. Histogram for analysis of experimental results

VI. Conclusion

This paper presents a method for the recognition of dangerous driving manoeuvres based on FWCW millimeter waves, proposing a new integrated recognition technique by adjusting the radar parameters and updating the classification method. Using convolutional neural networks in signal image feature vector representation, a convolutional layer is added before each pooling layer, allowing the network depth to increase to achieve more accurate classification results. Inputting signals from neurons into synapses for selective channel processing of different feature vectors according to the SlowFast CNN dual channel, The high frequency sampling channel is lateralised to the low frequency sampling channel to fuse the data. Training and testing of the input linear FM continuous wave signal after feature extraction. The experimental results show that the robustness and accuracy of the dual-channel micro-Doppler features in this paper with CNN single-feature time-distance features and ST-GCN for key points of the human skeleton are better. An upper body recognition accuracy of 97% can be achieved for dangerous driving behaviour for six training subjects and 85.6% for one test subject. The model has been shown to have strong generalisation to the general population. The method will be improved in the future to increase the accuracy and identify the movements of multiple people in the vehicle.

VII. Declarations

A. Funding and/or Conflicts of interests/Competing interests

(1)This work was supported by the National Nature Fund Project (Grant numbers 61762079). Author Hao has received research support from the National Nature Fund Project.
(2)This work was supported by Gansu Science and Technology support key Research and Development Program (Grant numbers 20YF8GA048). Author Hao has received research support from Gansu Provincial Department of Science and Technology.
(3) This work was supported by the 2019 "Light of the West" talent project of the Chinese Academy of Sciences. Author Hao has received research support from the Chinese Academy of Sciences.

B. Data availability" / "Availability of Data and Materials

The datasets generated and/or analysed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

C. For experiments involving human participants

Demonstrate the following: all subjects and/or their legal guardians gave informed consent to participate in the study and to have identifying information/images published publicly online.

Reference

[1] Gao Q, Wang J, Ma X, et al. CSI-based device-free wireless localization and activity recognition using radio image features[J]. IEEE Transactions on Vehicular Technology, 2017, 66(11): 10346-10356.
[2] Rao X, Lau V K N. Interference alignment with partial CSI feedback in MIMO cellular networks[J]. IEEE Transactions on Signal Processing, 2014, 62(8): 2100-2110.
[3] Yao S, Swetha P, Zhu Y. Nanomaterial-enabled wearable sensors for healthcare[J]. Advanced healthcare materials, 2018, 7(1): 1700889.
[4] Camps-Mur D, Garcia-Villegas E, Lopez-Aguilera E, et al. Enabling always on service discovery: Wifi neighbor awareness networking[J]. IEEE Wireless Communications, 2015, 22(2): 118-125.
[5] Chen L, Ma N, Wang P, et al. Survey of pedestrian action recognition techniques for autonomous driving[J]. Tsinghua Science and Technology, 2020, 25(4): 458-470..
[6] Chen J, Tian S, Xu H, et al. Architecture of vehicle trajectories extraction with roadside LiDAR serving connected vehicles[J]. Ieee Access, 2019, 7: 100406-100415.
[7] Fan W, Min H J. Accurate Recognition and Simulation of 3D Visual Image of Aerobics Movement[J]. Complexity, 2020, 2020.
[8] El Ayach O, Rajagopal S, Abu-Surra S, et al. Spatially sparse precoding in millimeter wave MIMO systems[J].

IEEE transactions on wireless communications, 2014, 13(3): 1499-1513.
[9] Cardillo E, Sapienza G, Li C, et al. Head Motion and Eyes Blinking Detection: a mm-Wave Radar for Assisting People with Neurodegenerative Disorders[C]//2020 50th European Microwave Conference (EuMC). IEEE, 2021: 925-928.
[10] Zheng C, Chen H, Wang A. High Angular Resolution for 77GHz FMCW Radar via a Sparse Weighted Quadratic Minimization[J]. IEEE Sensors Journal, 2021, 21(9): 10637-10646.
[11] Chipengo U, Sligar A, Carpenter S. High Fidelity Physics Simulation of 128 Channel MIMO Sensor for 77GHz Automotive Radar[J]. IEEE Access, 2020, 8: 160643-160652.
[12] Alkhateeb A, El Ayach O, Leus G, et al. Channel estimation and hybrid precoding for millimeter wave cellular systems[J]. IEEE journal of selected topics in signal processing, 2014, 8(5): 831-846.
[13] Koenig S, Lopez-Diaz D, Antes J, et al. Wireless sub-THz communication system with high data rate[J]. Nature photonics, 2013, 7(12): 977-981.
[14] Yan Y, Xie G, Lavery M P J, et al. High-capacity millimetre-wave communications with orbital angular momentum multiplexing[J]. Nature communications, 2014, 5(1): 1-9.
[15] Gao X, Dai L, Han S, et al. Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(4): 998-1009.
[16] Hong F, Lu C H, Liu C, et al. A traffic surveillance multi-scale vehicle detection object method base on encoder-decoder[J]. IEEE Access, 2020, 8: 47664-47674.
[17] D'Antonio E, Taborri J, Mileti I, et al. Validation of a 3D Markerless System for Gait Analysis based on OpenPose and Two RGB Webcams[J]. IEEE Sensors Journal, 2021.
[18] Snowden R J, Kavanagh E. Motion perception in the ageing visual system: Minimum motion, motion coherence, and speed discrimination thresholds[J]. Perception, 2006, 35(1): 9-24
[19] LUO Hui-lan, TONG Kang, KONG Fan-sheng. The Progress of Human Action Recognition in Videos Based on Deep Learning: A Review[J]. Acta Electronica Sinica, 2019, 47(5): 1162-1173. 2019.
[20] Li X, Tao X, Zhu B, et al. Research on a simulation method of the millimeter wave radar virtual test environment for intelligent driving[J]. Sensors, 2020, 20(7): 1929.
[21] Jiang T, Wu H, Wu K, et al. Threshold design method of CFAR for millimeter-wave collision warning radar[J]. JOURNAL OF INFRARED AND MILLIMETER WAVES-CHINESE EDITION-, 2005, 24(3): 217.
[22] Kim S, Yun JH. Motion-Aware Interplay between WiGig and WiFi for Wireless Virtual Reality. Sensors (Basel). 220 Nov 27;20(23):6782. doi: 10.3390/s20236782. PMID: 33261082; PMCID: PMC7729580.
[23] Ng W, Minasny B, Montazerolghaem M, et al. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra[J]. Geoderma, 2019, 352: 251-267.
[24] Zhao A, Qi L, Li J, et al. A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data[J]. Neurocomputing, 2018, 315: 1-8.
[25] Balal Y, Balal N, Richter Y, et al. Time-Frequency Spectral Signature of Limb Movements and Height Estimation Using Micro-Doppler Millimeter-Wave Radar[J]. Sensors, 2020, 20(17): 4660.
[26] Jiang X, Liu W, Zhang Y, et al. Spectral–Spatial Hyperspectral Image Classification Using Dual-Channel Capsule Networks[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(6): 1094-1098.
[27] Kamnitsas K, Ledig C, Newcombe V F J, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation[J]. Medical image analysis, 2017, 36: 61-78.
[28] Cheng W L, Liu Q N, Zhao R, et al. Experimental investigation of parameters effect on heat transfer of spray cooling[J]. Heat and mass transfer, 2010, 46(8):

Zhanjun Hao was born in Xingtai, Hebei Province in 1979, Doctor of Lanzhou Jiaotong University， He is current a professor with College of Computer Science and Engineering at Northwest Normal University (NWNU), China. The main research directions are location service, wireless positioning technology and wireless sensor networks.

Zepei Li , born in Baoji, Shaanxi Province. He is currently a graduate student in the School of Computer Science and Engineering at Northwest Normal University (NWNU), China.His main research interests are millimeter wave perception.

Xiaochao Dang, born in 1963. Master,Professor of Northwest Normal University, Lanzhou, Gansu, China. The main research directions are Internet of things, sensor network, wireless sensing technology and so on.

ZhongYu Ma,He is current an associate professor with College of Computer Science and Engineering at Northwest Normal University (NWNU), China. His research interests are in the area of wireless networking and communications, including millimeter wave communication, wireless network optimization and resource allocation, designment and simulation for wireless communication protocol.