A lightweight attention deep learning method for human-vehicle recognition based on wireless sensing technology

Mingxin Song[1,2], Rensheng Zhu[3], Xinquan Chen[1,2], Chunlei Zheng[4] and Liangliang Lou[1]*

[1]Institute of Intelligent Information Processing, Taizhou University, Taizhou, Zhejiang, China, [2]School of Science, Zhejiang University of Science and Technology, Hangzhou, Zhejiang, China, [3]China United Network Communications Co., Ltd., Taizhou Branch, Taizhou, Zhejiang, China, [4]Key Laboratory of Wireless Sensor Networks and Communications, Shanghai Institute of Microsystem and Information Technology, Shanghai, China

Wireless sensing-based human-vehicle recognition (WiHVR) methods have become a hot spot for research due to its non-invasiveness and cost-effective advantages. However, existing WiHVR methods shows limited performance and slow execution time on human-vehicle classification task. To address this issue, a lightweight wireless sensing attention-based deep learning model (LW-WADL) is proposed, which consists of a CBAM module and several depthwise separable convolution blocks in series. LW-WADL takes raw channel state information (CSI) as input, and extracts the advanced features of CSI by jointly using depthwise separable convolution and convolutional block attention mechanism (CBAM). Experimental results show that the proposed model achieves 96.26% accuracy on the constructed CSI-based dataset, and the model size is only 5.89% of the state of the art (SOTA) model. The results demonstrate that the proposed model achieves better performance on WiHVR tasks while reducing the model size compared to SOTA model.

KEYWORDS

human-vehicle recognition, channel status information, attention mechanism, depthwise separable convolution, wireless sensing

1. Introduction

INTELLIGENT Traffic Systems (ITS) is an important part of smart city (Santos et al., 2018; Jin and Ma, 2019; Zhao et al., 2019; Choy et al., 2020; de Oliveira et al., 2020), providing reliable, safe, and convenient services for road users (e.g., cars, motorcycle, pedestrians, etc.). As the number of road users continues to increase, a large number of existing ITSs are approaching their limits. In order to improve the performance of ITSs and relieve traffic pressure, the measurement of traffic parameters including road user behavior has become a research hotspot (Jiang et al., 2021; Park et al., 2021; Zhao and Huang, 2021). Generally, the behavior of road users includes human-vehicle recognition (HVR), traffic flow statistics, vehicle speed, and direction measurement, etc. As the foundation of road user behavior detection, the accuracy of human-vehicle recognition determines the performance of traffic parameter measurement (Won et al., 2017; Sliwa et al., 2020).

With the rapid development of artificial intelligence and deep learning techniques, image-based HVR methods (Huang et al., 2020; Du et al., 2021) have been widely used in ITSs. Such

HVR methods not only achieve excellent recognition performance, but also provide rich traffic image information for city managers. However, image-based HVR methods are susceptible to light so that their performance can be degraded rapidly in the case of low-light conditions such as night, cloudy, haze, etc. To alleviate the limitations of image-based HVR methods in low-light scenes, the microwave radar-based HVR method is proposed (Park et al., 2021; Singh et al., 2021; Tavanti et al., 2021). However, high-performance frequency-modulated continuous-wave (FMCW) radars come at a higher cost. In addition, the microwave radar has the problem of installation viewing angle, which leads to the high construction cost of the microwave radar-based HVR method.

Generally, the purpose of wireless sensing-based HVR (WiHVR) methods is to extract the energy information of surrounding wireless signals for target recognition. Since the propagation of wireless signals has no directionality, the WiHVR method does not have the problem with the above-mentioned viewing angle (Ma et al., 2019; Pan et al., 2019; Zhang et al., 2021). In recent years, the WiHVR methods based on extracted receiving signal strength (RSS) or channel state information (CSI) signatures from the wireless transceivers on 2.4 GHz band, such as Bluetooth (Sliwa et al., 2020; Wilby et al., 2020), ZigBee (Wang et al., 2017; Jiang et al., 2021), and WiFi (Chen et al., 2018; Wang F. et al., 2018), etc., have been widely employed to detect road users in ITSs.

1.1. CSI-based HVR methods

Wang W. et al. (2016) converted CSI signals to spectrograms, thereby describing human motion. Won et al. (2017) proposed a WiFi-based traffic monitoring system, in which the features of root mean square, median absolute deviation, mean, first quartile, and third quartile of the CSI signals were extracted, followed by a support vector machine for vehicle classification. Liu et al. (2017) showed a human motion detection method based on CSI phase difference. They discussed the situation of line of sight (LOS) and non-line of sight (NLOS). Arshad et al. (2018) proposed a WiFi-based device-free dangerous driving recognition system. This system extracted multi-domain features for both magnitude and phase of CSI signals. Wang J. et al. (2018) presented a new general device-free identification framework *via* empirical mode decomposition. They decomposed CSI signals into intrinsic mode functions (IMF) and extracted the time domain and frequency domain features from IMF components.

1.2. RSS-based HVR methods

Jiang et al. (2021) calculated the amplitude and mean information of RSS signals. They designed a HVR algorithm for WiHVR based on the calculated RSS features. Sliwa et al. (2020) provided a vehicle detection and classification method on the basis of the extracted RSS from transceivers on 2.4 GHz band. They used mean, minimum, standard deviation, and other characteristics of RSS signals to address the challenges of accuracy, robustness, and privacy. Abdelnasser et al. (2018) exploited a gesture recognition system in which the edge, frequency, and magnitude features of RSS signals were extracted for gesture recognition. Bhat et al. (2020) extracted the RSS power levels for human locomotion walking pattern recognition.

However, the above-mentioned WiHVR methods based on extracted RSS or CSI signatures from 2.4 GHz wireless transceivers like Bluetooth, ZigBee, and WiFi have the following drawbacks:

1) RSS is a coarse-grained signal, which leads to limited accuracy of HVR tasks based on RSS signals.
2) The effects of CSI or RSS on the performance of WiHVR in different application scenarios are not explored.

Recently, deep learning techniques (LeCun et al., 2015) consisting of a multi-layer network architecture have attracted much interest. One of the representative deep learning techniques is convolutional neural network (CNN) (Krizhevsky et al., 2012). Up to now, due to the powerful feature learning ability, CNNs have exhibited promising performance on various tasks such computer vision (Szegedy et al., 2016), speech signal processing (Zhang et al., 2017), natural language processing (Otter et al., 2020), and so on. However, few works have attempted to exploit the application of CNNs on WiHVR tasks.

To address the above-mentioned issues, this paper presents a novel WiHVR method based on the designed lightweight wireless sensing attention-based deep learning model (LW-WADL). Inspired by the recent-emerged convolutional block attention mechanism (Woo et al., 2018) (CBAM) and depthwise separable convolutions (Chollet, 2017), we propose a new deep model, which consists of a CBAM module and three depthwise separable convolution blocks in series to learn high-level features from preprocessing CSI signals for WiHVR. Compared with ordinary convolutions, depthwise separable convolutions have relatively low parameters and operations. Besides, we propose a novel CSI data enhancement method and a new subcarrier selection method. In particular, a new CSI-based dataset relates to road user behavior is constructed. In order to explore the effects of CSI on the performance of WiHVR in different application scenarios, the CSI dataset is divided into three taxonomies according to the number of categories, namely, two-category dataset, three-category dataset, and four-category dataset. Experimental results show that the accuracy of CSI-based methods decreases as the number of classification categories increases. For four-classification experiments, the proposed model achieves 96.26% accuracy and the model size is only 5.89% of the state of the art model.

To summarize, the main contributions of this paper are as follows:

1) This paper proposes a CSI data enhancement method, which preprocess the change trend of CSI data to one direction, thereby enhancing CSI data.
2) This paper provides a subcarrier selection method, which selects several subcarriers with large signal-to-noise ratios (SNR) as benchmarks and integrates them into a new CSI data.
3) This paper has proposed a lightweight wireless sensing attention-based deep learning model, and attempts to explore the effects of CSI on the performance of WiHVR in different application scenarios.

The remainder of this paper is organized as follows. Section "2. Preliminaries" introduces the CSI extraction and the theoretical analysis of WiHVR. Section "3. Proposed method" elaborates the proposed LW-WADL for WiHVR. Section "4. Experiment study" shows experimental results and analysis. Section "5. Conclusion and future work" gives the conclusions and future work.

2. Preliminaries

This paper aims to establish a lightweight and efficient WiHVR method to explore the effects of CSI on the performance of WiHVR in different application scenarios. The system architecture of the proposed WiHVR method is shown in **Figure 1**.

From **Figure 1**, it can be found that wireless transceiver prototype (WTP) is built and placed on both sides of the road. The WTP is mastered by the ESP32 chip for generating and receiving wireless signals. Once a road user appears in the WTP sensing area, the CSI signal collected by the WTP will be attenuated due to the road user. Therefore, WTP can extract CSI signals related to the road user information.

2.1. CSI extraction

This paper uses the designed WTP to extract CSI data related to road users. CSI represents the fine-grained channel features of wireless communication links between transmitters and receivers based on orthogonal frequency division multiplexing (OFDM) technology. Besides, CSI describes the changes of phase and amplitude caused by multipath effect and transmission loss in wireless signal transmission. The CSI channel gain matrix is expressed as:

$$Mcsi = \begin{pmatrix} h_{11} & \dots & h_{1n} \\ \vdots & \ddots & \vdots \\ h_{m1} & \cdots & h_{mn} \end{pmatrix} \quad (1)$$

where h_{mn} represents the different subcarriers. m and n represent the transmitting and receiving antennas, respectively. Each sub-element h_{mn} represents:

$$h_{mn} = ||h_{mn}||e^{j\eta_{mn}} \quad (2)$$

where $||h_{mn}||$ is the amplitude of the sub-carrier h_{mn}, and $e^{j\eta_{mn}}$ represent the phase of h_{mn}. From Eqs 1, 2, it can be known that CSI is not a supersession of all subcarrier signals, it describes a multipath signal with more characteristics. In this case, the CSI extracted by WTP contains multiple subcarrier information. These subcarriers have different sensitivities to road users, so it is necessary to filter out the subcarriers with lower sensitivity. The specific method will be elaborated in Section "3. Proposed method."

The specific process of CSI signal extraction is shown in **Figure 2**. The acquisition of CSI signal needs to be operated by inverse OFDM. In order to eliminate inter-symbol interference and inter-channel interference, OFDM will use cyclic prefix (C/P), but this part is not real data, so this part needs to be removed in inverse OFDM. After that, it is necessary to convert the series signal to the parallel signal (S/P), and perform discrete Fourier transform (DFT) or fast Fourier transform (FFT) to obtain the required CSI signal.

2.2. Theoretical analysis of WiHVR

The idea of WiHVR is based on the fact that road users of existence and movement affect the wireless propagation paths. To understand the relation of road users movement with received CSI, the wireless propagation model should be first studied. In a typical wireless environment, there is one main path line-of-sight (LOS) and several reflected paths by the surroundings. As shown in **Figure 1**, if a road user is present in the WTP sensing area, it will cause multipath propagation of the wireless signal. In this case, according to the free space model, the received power by a receiver antenna which is separated from a radiating transmitter antenna by a distance $_d$, is given by the Friis free space equation,

$$P_{\mathrm{r}} = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} \quad (3)$$

where P_r and P_t are the receiving and transmitting power, respectively. G_r and G_t are the receiving and transmitting antenna gains, respectively. λ is the wavelength in meters. d is the distance between the transmitter and receiver in meters, that is, the propagation path length. When a road user exists in the wireless environment, several scattered paths are produced by road user. Those scattered power should also be added in the final received power.

$$P_{\mathrm{r}} = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 (d^2 + \delta^2)} \quad (4)$$

where δ is a brief representation of path length caused by road user. If a road user is static in the environment, P_r is almost stable. However,
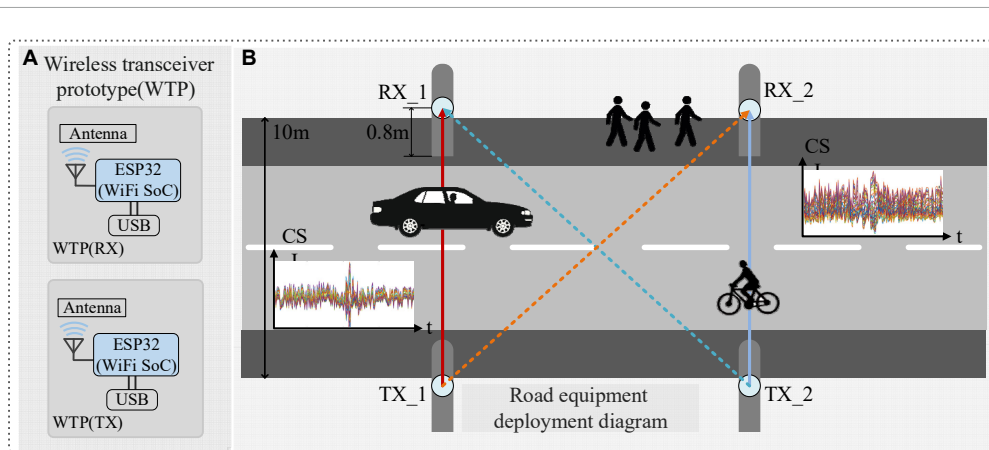

FIGURE 1
System architecture of the proposed WiHVR method. **(A)** Wireless transceiver prototype (WTP). **(B)** Road equipment deployment diagram.

along with the move of a road user, the scattered paths change in a fast speed, resulting in the variance in received signal power.

According to Eq. 4, the differences in size and speed of road users lead to different attenuation of wireless signals. Hence, the CSI readings measured by the WTP prototype are various.

3. Proposed method

According to the above analysis, since the differences in the size and speed of humans or vehicles moving on the road, the attributes of energy attenuation caused by two targets are various. In this case, it is feasible to design a WiHVR method. To this end, a deep learning-based WiHVR method is proposed in purpose of analyzing the effects of CSI on the performance of WiHVR in different application scenarios.

3.1. System overview

The overview of the proposed WiHVR based on a lightweight wireless sensing attention-based deep learning model (LW-WADL) is shown in **Figure 3**. The proposed WiHVR contains three key modules: Data collection, CSI preprocessing, and Deep feature extraction and classification. The data collection module consists of a

pair of WTPs, both of which are made up of an ESP32 module, so as to collect CSI data of different road users in WTPs sensing area. The CSI preprocessing module includes CSI filtering, CSI augmentation, CSI subcarriers selection, and CSI segmentation. The core deep feature extraction and classification module, i.e., the proposed LW-WADL method consisting of a CBAM module and three depthwise separable convolution blocks, followed by a global average-pooling (GAP) layer for reducing computational complexity. In addition, GAP essentially is an average pooling operation which is intended to replace fully connected layers in classical CNNs. Thus, GAP is a special kind of average pooling where the sliding window of the average operation expands to the entire feature maps. Besides, after completing the final feature representations of the GAP layer, a C-class vector (C is the number of categories) is output through the Softmax function.

3.2. Data collection

As shown in **Figure 3**, this paper captures the CSI data in space through the developed WTP. To extract CSI data, a threshold-based road user detection algorithm is exploited in this paper. The purpose of road user detection is to find out whether there are dynamic targets in the sensing area. According to the analysis in Section "2. Preliminaries," it can be found that when there are no road users in the wireless environment, the CSI patterns stabilize around a reference value. Once a road user passes through the wireless
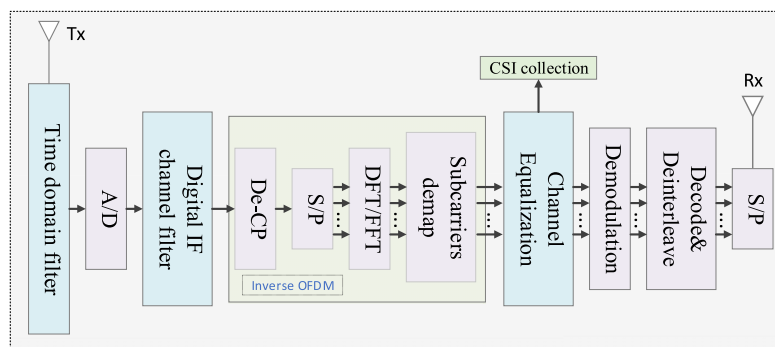

FIGURE 2
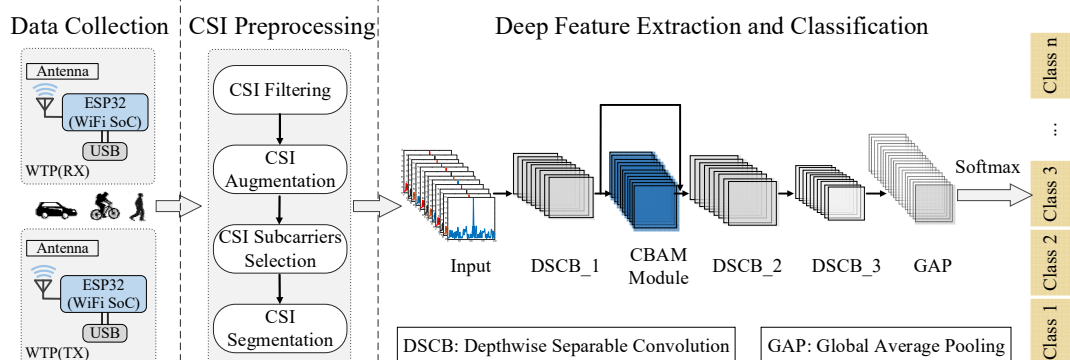The specific process of CSI extraction.

FIGURE 3
Overview of the proposed WiHVR based on a lightweight wireless sensing attention-based deep learning model (LW-WADL).

environment, the amplitude of CSI patterns will drop sharply. Therefore, the presence of road users in the region of interest can be detected by the following threshold-based algorithm:

$$X_{\det ection}[k+1] = \qquad (5)$$

$$\begin{cases} Static, \ X_{\det ection}[k] = Detected \ and \\ \quad \prod_{n=k-W+1}^{k} sign(|x[n] - x_{static}[n]| < T_{object}) > 0 \\ Detected, \ X_{\det ection}[k] = Static \ and \\ \quad \prod_{n=k-W+1}^{k} sign(|x[n] - x_{static}[n]| \geq T_{object}) > 0 \end{cases}$$

where $x[n]$ ($n > 0$) represents the n-th CSI reading. $x_{static}[n]$ is the average of CSI readings when there are no road users in the wireless environment, namely, CSI baseline. T_{object} is the decision threshold to determine whether there is a road user (dBm). Here, T_{object} is set to 4 dBm in this work. W is the size of the judgment window, and is set to 50 when the sampling rate is 50 Hz. $X_{\det ection}[]$ is the object detection result. "Detected" indicates that there are road users in the range of interest, and otherwise "Static" denotes no road users.

Moreover, the environmental factors such as rain, fog, temperature, etc., can affect the CSI baseline $x_{static}[]$. Thus, to improve the performance of the above-mentioned fluctuation detection algorithm, an adaptive baseline adjustment method is proposed, which can be calculated by:

$$x_{static}[n+1] = \begin{cases} \beta \cdot x_{static}[n] + (1-\beta) \cdot x[n+1], \\ \quad X_{\det ection}[n] = X_{\det ection}[n+1] = Static \quad (6) \\ x_{static}[n], \ others \end{cases}$$

where β is a correction factor with a value of 0.96 in this paper. It can be seen from Eq. 6 that the CSI baseline will be updated as long as there are no road users in the wireless environment, otherwise it will not be updated. Hence, the problem of CSI baseline drift caused by environmental factors can be solved efficiently, as well as the robustness of the fluctuation detection algorithm can be improved.

Finally, the CSI data extracted by WTP contains 52 subcarriers, and each subcarrier contains amplitude and phase information. In order to improve execution efficiency of LW-WADL, this paper converts the raw two-dimensional CSI data containing amplitude information into one-dimensional data. Then, one-dimensional CSI data containing road user behavior information will be sent to the second stage for data preprocessing.

3.3. CSI preprocessing

The CSI preprocessing module includes the following four steps: CSI filtering, CSI augmentation, CSI subcarriers selection, and CSI segmentation.

3.3.1. CSI filtering

To guarantee the robustness of road users recognition, smoothing filtering is used to remove noise from the raw CSI data, as defined by:

$$X_{filter}(n) = \frac{1}{N} \sum_{j=1}^{N-1} X_{raw}(n-j) \qquad (7)$$

where X_{raw} represent the raw CSI data, and $X_{filter}(n)$ is the average processed data and then the filter shift window size used is N, where is set to five. The raw CSI waveform vs. filtered waveform is shown in **Figure 4**. As can be seen from **Figure 4**, by applying moving

average filter, the high-frequency noise has been removed from the CSI waveform without changing the trends of the waveform. The waveform changes of the filtered data (**Figures 4C, D**) are more pronounced than before filtering, thereby improving the efficiency and accuracy of road user detection.

3.3.2. CSI augmentation

Channel state information augmentation aims to find a way to enhance the CSI features without changing the raw CSI features. According to the characteristics of the raw CSI signal waveform, this paper proposes a novel CSI data enhancement method. This method first calculates the average value of a set of CSI amplitude, and then takes the absolute value of the CSI amplitude which is smaller than the average value. In this way, the decay of the CSI amplitude is amplified, thereby enhancing CSI features. First, the baseline X_{base} of a set of CSI data needs to be calculated, which can be expressed as:

$$X_{base} = \frac{1}{I} \cdot \frac{1}{T} \sum_{i=1}^{I} \sum_{n=1}^{T} X_{csi}(i, n) \qquad (8)$$

where i represents the i-th CSI subcarrier, n represents the n-th sampling point of the i-th subcarrier. I represents the number of CSI subcarriers, which is 52 in this paper. T is the number of sampling points of a group of CSI data. The enhanced CSI data $X_{csi\_aug}(i, n)$ can be obtained according to the CSI baseline X_{csi_base}:

$$X_{csi\_aug}(i, n) = |X_{csi}(i, n) - X_{csi\_base}| \qquad (9)$$

where $| |$ denotes the absolute value operation. According to the Eqs 8, 9, the enhanced CSI data can be obtained.

3.3.3. CSI subcarriers selection

Although CSI augmentation have enhanced CSI features related to road users. In practical applications, different subcarriers of CSI have different sensitivities to road users, e.g., some subcarriers fluctuate greatly when encountering road users, while other subcarriers fluctuate less. Therefore, to further enhance CSI data, we design a raw CSI subcarrier selection method to remove subcarriers with low sensitivity in CSI data. In order to evaluate the sensitivity of CSI subcarriers, this paper calculates the SNR of the CSI data amplitude, as expressed by:

$$SNR = 10 \lg \left| \frac{x_{peak} - x_{static}}{n_{noise} - x_{static}} \right| \qquad (10)$$

where x_{peak} is the peak value of CSI with respect to a road user. x_{static} is the average of CSI readings when there are no road users within a wireless environment. n_{noise} is the peak value of noise. According to Eq. 10, the SNR $X_{csi\_SNR}(n)$ of all subcarriers in a set of CSI data is obtained:

$$X_{csi\_SNR}(n) = \{x_{SNR}(1), x_{SNR}(2), ..., x_{SNR}(m), ..., x_{SNR}(n)\} \qquad (11)$$

where $x_{SNR}(m)$ represents the SNR value of the m-th subcarrier. For the convenience of calculation, it is assumed that $x_{SNR}(n)$ has been arranged in descending order of SNR, that is, $\{x_{SNR}(1) > x_{SNR}(2) > ... > x_{SNR}(m) > ... > x_{SNR}(n)\}$. According to Eq. 11, "m" subcarriers with larger SNR are selected, where "m" is defined as the CSI factor. The selection of the CSI factor "m" is discussed in detail in Section "4. Experiment study." The mean of "m" subcarriers is calculated, which can be expressed as:

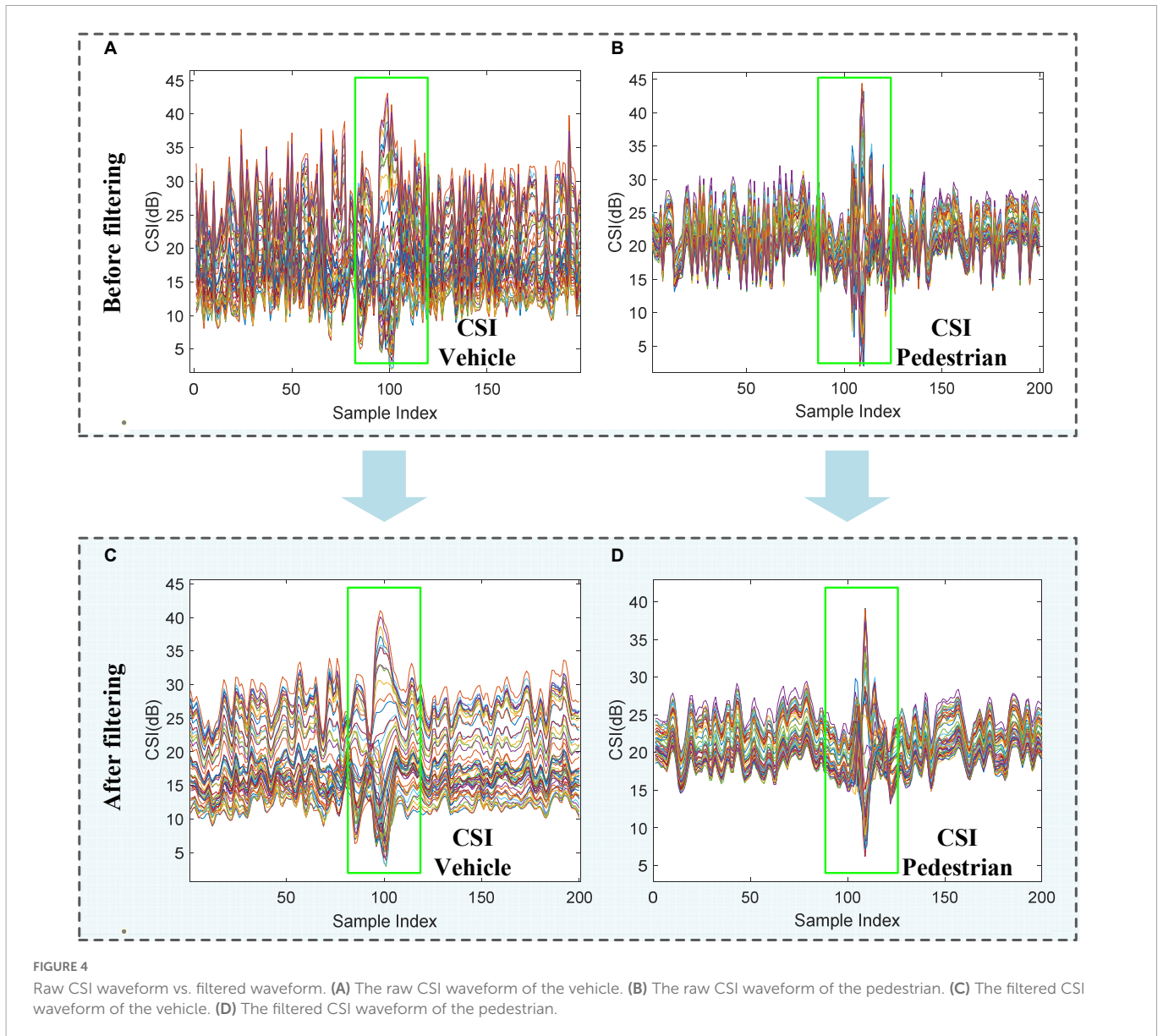$$\bar{X}_{csi\_aug}(n) = \frac{1}{m} \sum_{n=1}^{m} X_{csi\_aug}(n) \qquad (12)$$

To demonstrate the validity of Eq. 12, we compare our proposed method with k-subcarriers weight fusion (Kong et al., 2019) and average-subcarriers (Wang Y. et al., 2016), as shown in **Figure 5**. It can be seen that on a pedestrian and a vehicle CSI sample, our CSI subcarrier selection method performs best, the SNR of the CSI amplitude is 3.5 and 6.1 dB, respectively.

3.3.4. CSI segmentation

The selected CSI data $\bar{X}_{csi\_aug}(n)$ containing multiple CSI features is split into certain segment-level sub-samples, each of which consists of one complete CSI feature of a road user. Specifically, the single CSI feature can be divided in terms of the local minimum in $\bar{X}_{csi\_aug}(n)$. These local minimums are defined as decision points. Specially, $x_d[i]$ represents the i-th decision point, $x_d[i]$ can be calculated as:

$$x_d[i] = \min_{s \cdot i - w + 1 \leq n \leq s \cdot i} \bar{X}_{csi\_aug}(n), \ i = 1, 2, 3, ..., L \quad (13)$$

where $x_d[i]$ is the minimize value within the value range of $\bar{X}_{csi\_aug}(n)$. L is the number of decision points. w is the size of sliding window, and is set to 50, s is the step size of the window, and is

set to 200. Additionally, the index of decision points in $\bar{X}_{csi\_aug}(n)$ is represented by P_i. According to Eq. 13 and P_i, $\bar{X}_{csi\_aug}(n)$ can be divided into L segments. $\tilde{x}_i[n]$ is the i-th segment, which can be defined as:

$$\tilde{x}_i[n] = \left\{ \bar{X}_{csi\_aug}[P_i - c_0], ..., \bar{X}_{csi\_aug}[P_i], ..., \bar{X}_{csi\_aug}[P_i + c_0] \right\} \quad (14)$$

where c_0 is the slicing factor, and is set to 100. In this case, a new CSI dataset is developed. About 500 samples of four categories are included in the dataset: pedestrian, bicycle, motorcycle, and car.

3.4. Deep feature extraction and classification

According to the features of CSI signals containing time series features, as shown in **Figure 3**, a lightweight wireless sensing attention-based recognition algorithm, namely LW-WADL is proposed for deep feature learning from CSI features on HVR tasks. The proposed LW-WADL contains a CBAM module and three
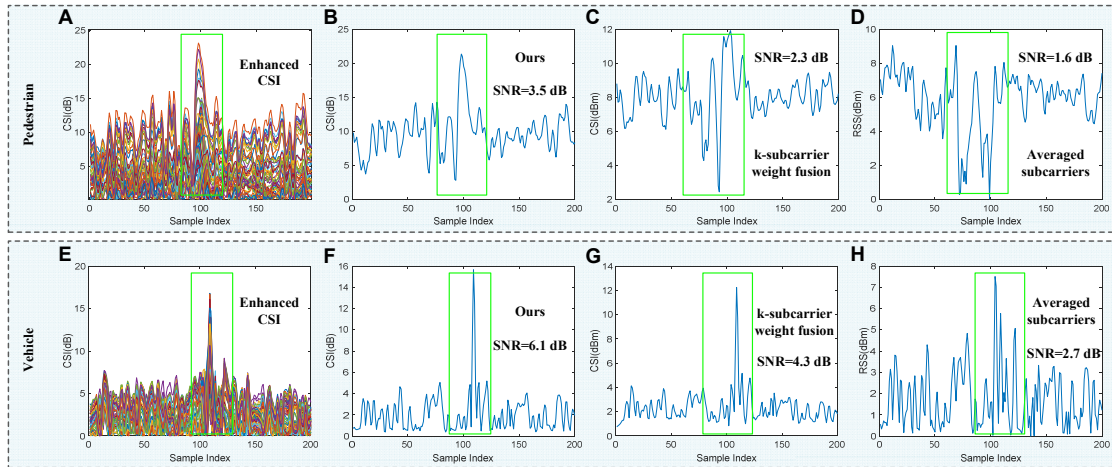
FIGURE 5
Different CSI subcarrier selection method. **(A,E)** Represent the enhanced raw CSI waveforms of the vehicle and pedestrian, respectively. **(B–D,F–H)** Represent the CSI waveforms of the vehicle and pedestrian generated by the three subcarrier selection methods, respectively.

depthwise separable convolution modules, followed by a GAP layer, as described below.

3.4.1. LW-WADL network structure

The overall network structure of the proposed LW-WADL is presented in **Figure 3**. LW-WADL involves of Three depthwise separable convolution blocks (DSCB_1, DSCB_2 and DSCB_3) in series. Then, in order to focus on learning the relevant information of feature maps while suppressing the irrelevant information, a CBAM module is concatenated after DSCB_1. The CBAM module can further improve the discriminating power of feature representations learned by DSCB_1. Finally, output features in DSCB_3 are achieved through a GAP layer.

3.4.2. CBAM attention module

The attention mechanism makes the model tend to pay attention to some information about the auxiliary classification in the feature map, while suppressing other useless information, thereby improving the classification ability of the model. The CBAM module consists of a channel attention module and a spatial attention module. The detailed structure is shown in **Figure 6**.

The channel attention module first performs maxpooling and average pooling based on the height and width of the DSCB_1 feature map to obtain two one-dimensional vectors. Then, it is input into the shared multi-layer perceptron (Shared MLP), and the corresponding elements of the output features of the MLP are summed point by point. The result is input into the Sigmoid activation function, and then the inner product operation is performed with the initial input feature map. The final output feature map is used as the input of the spatial attention module.

The spatial attention module performs maxpooling and average pooling based on the channel, and then uses the convolution (abbreviated as Conv) operation to merge the output features on the channel dimension. The merged features are input into a sigmoid activation function, then an inner product operation is performed on the obtained output features and the input of the spatial attention module. Finally, the output of the inner product operation is combined with the output of the DSCB_1 module to form the input features of the DSCB_2 module.

3.4.3. Softmax classifier output

WiHVR is fundamentally a multi-classification task, so we choose the Softmax function to produce final classification results. Through the Softmax function, the output values of classifier can be converted into a probability distribution in the range [0, 1].

The cross-entropy loss function is implemented as the training objective function for LW-WADL:

$$L_{loss} = -\sum_i \widehat{y}_i \log(y_i) \qquad (15)$$

where $\hat{y}_i = 1$ if the class is i, otherwise $\hat{y}_i = 0$. y_i represents the output of the LW-WADL model, the probability that the class is i. L_{loss} is a loss measure of the difference between two probability distributions.

4. Experiment study

4.1. Experiment setup

As can be seen from **Figure 7**, the proposed WTP prototype contains two main components: antenna, ESP32. ESP32 is a WiFi SoC working at a frequency of 2.4 GHz. In the experiment, the two WTP prototypes were installed on both sides of a road with a width of 10 m, and antenna heights is set to 1 m. In addition, for training LW-WADL models, the Adam optimizer with a learning rate of 0.0001 is used. The batch-size is 16 and the maximum of epochs is 200. Besides, to explore the effects of CSI on the performance of WiHVR in different application scenarios, the developed CSI dataset is divided into three taxonomies according to the number of categories, namely, two-category dataset, three-category dataset and four-category dataset. Finally, 80% of the data in the dataset is used as the training set, while the rest is used for testing.

4.2. Evaluation indicators

The performance of the designed LW-WADL is evaluated by three typical metrics such as "Accuracy," "Recall," and "Precision."
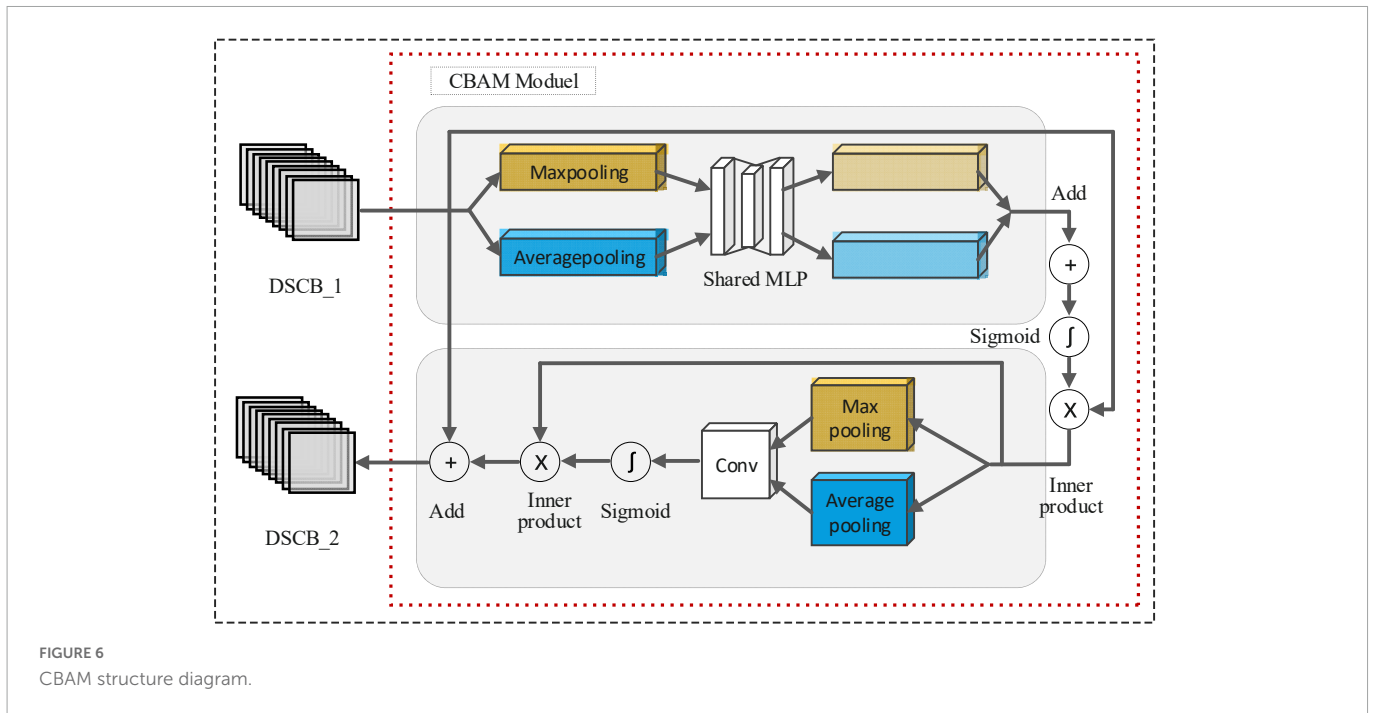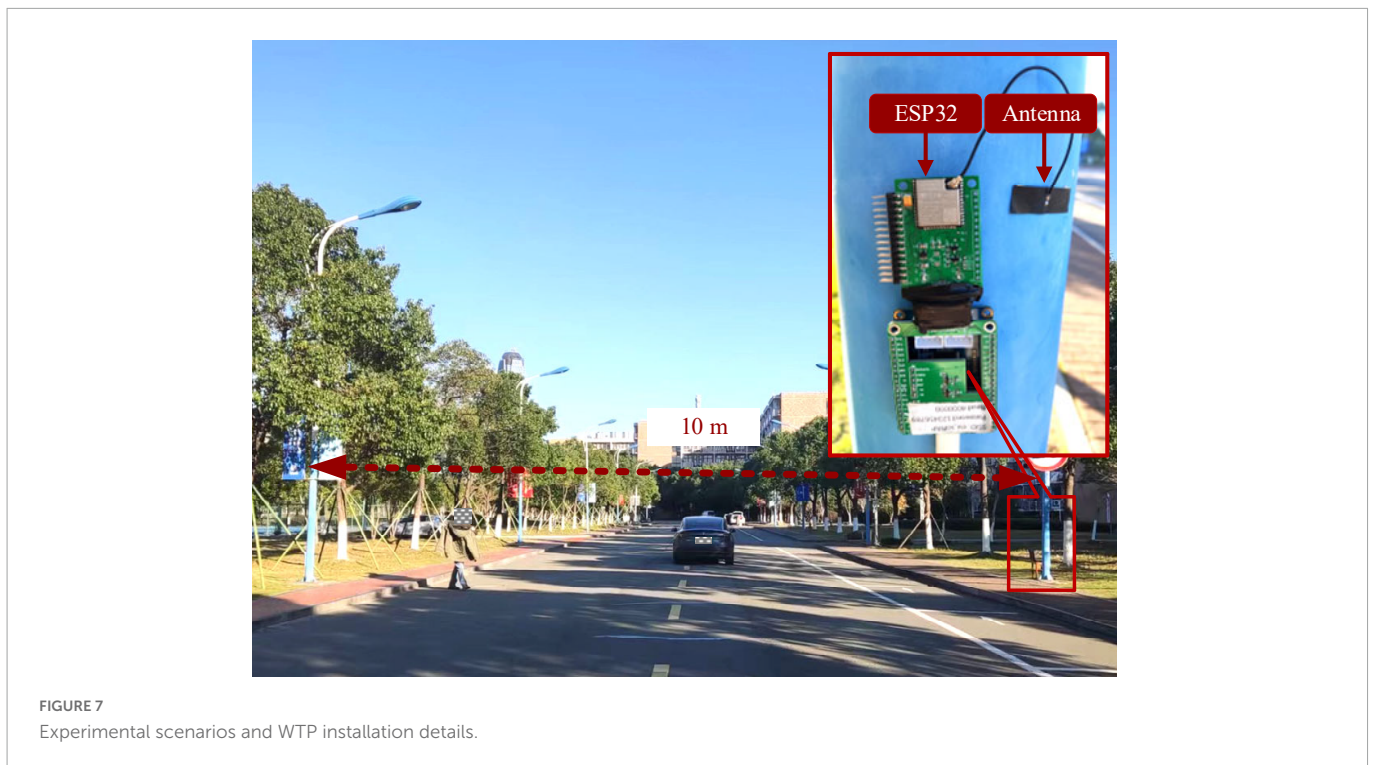
FIGURE 6
CBAM structure diagram.

FIGURE 7
Experimental scenarios and WTP installation details.

For the computational complexity analysis of deep learning methods, two well-known computational indicators, the network parameters (abbreviated as param.) and floating-point operations (FLOPs) are employed. Specifically, "Accuracy" is the ratio of all correct predictions to the whole number of predictions. "Precision" is the ratio of correct predictions with positive values to total predictions with positive values. "Recall" is the ratio of predicted positives to the total number of actual positives. They are defined as:

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

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

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

where TP denotes the number of true positive samples classified as positive. FP denotes the number of true negative samples classified as positive. FN denotes the number of true positive samples classified as negative. TN denotes the number of true negative samples classified as negative.

4.3. Comparison of different methods

To verify the effectiveness of our LW-WADL for WiHVR, we adopt the developed CSI four-category dataset, so as to compare the performance of various deep learning models on WiHVRs tasks. Table 1 present a performance comparison of four methods, including full convolutional network (FCN) (Long et al., 2015), DeepWiTraffic (Won et al., 2019), and deep residual network (ResNet) (He et al., 2016). Among them, FCN and ResNet are used as baseline methods to provide benchmarking performance, whereas DeepWiTraffic is used as comparing work. Besides, FCN is composed of three convolutional layers. ResNet contains three residual blocks. DeepWiTraffic contains two convolutional layers and two max pooling layers. Experimental results are shown in Table 1.

From Table 1, it can be found that the designed deep learning model has the highest classification performance with an accuracy of 96.26%, a precision of 96.23%, and a recall of 96.16%. Compared with DeepWitraffic, our model not only makes an improvement of 1.67%, but also exhibits much lower computational complexity in which 93.25% of the parameters (Param, FLOPs) can be reduced. Moreover, the test time is just 0.0575 s, which is much less than DeepWitraffic. This shows that our model is a lightweight model. Additionally, compared with FCN and ResNet, our method yields an accuracy improvement of 2.8 and 1.87%.

4.4. Selection of CSI factor m

A set of experiments are designed to investigate the effect of CSI factor "m" on the accuracy of WiHVR. Figure 8 shows the accuracy of the WiHVR for different "m," where "all" represents the maximum "m." Experiments are performed on the CSI four-category dataset.

Each CSI factor "m" corresponds to a CSI four-category dataset, and these datasets are identical except for the CSI factor "m." To make the results more reliable, the ResNet, DeepWiTraffic, FCN, and our model are used. The experimental results are shown in Figure 8.

As shown in Figure 8, with the increase of "m," both the ResNet, DeepWiTraffic, FCN, and ours model show a trend of increasing first and then decreasing, reaching the highest accuracy of 94.39, 94.59, 93.46, and 96.26% when "m" is 4, respectively. It can be found that only one subcarrier with the highest SNR or the average of all subcarriers cannot obtain the best HVR performance. This is because the sensitivity of different CSI subcarriers varies greatly. Some subcarriers are less sensitive, while some subcarriers with higher sensitivity are too sensitive to environmental changes, resulting in reduced recognition ability.

4.5. Comparison of CSI subcarrier selection methods

To verify the performance of the proposed CSI subcarrier selection method, we conduct comparative experiments on the CSI four-category dataset. Recently-merged CSI subcarrier selection methods such as k-subcarrier weight fusion (Kong et al., 2019) and averaged subcarriers (Wang Y. et al., 2016) are used for comparative experiments. The recognition accuracy of FCN, DeepWiTraffic, ResNet, and our LW-WADL under different CSI subcarrier selection methods are shown in Table 2.

As shown in Table 2, the four used models perform best under our CSI subcarrier selection method, and the accuracy of HVR is 93.46, 94.59, 94.39, and 96.26%, respectively. The results show the superiority of our CSI subcarrier selection method, which is consistent with the conclusion drawn in Figure 5. In addition, the

TABLE 1 Performance comparison of different methods on four-category dataset.

Method	Accuracy (%)	Precision (%)	Recall (%)	Param (M)	FLOPs (M)	Testing time(s)
FCN	93.46%	93.34%	93.46%	1.6451	3.2888	0.2019
DeepWiTraffic	94.59%	94.45%	94.59%	0.2090	0.4176	0.1398
ResNet	94.39%	95.28%	94.39%	0.7427	1.5047	0.2982
Ours	**96.26%**	**96.23%**	**96.16%**	**0.0123**	**0.0282**	**0.0575**

The bold values represent the indicator with the best performance, that is, the highest Accuracy, Precision, and Recall as well as the lowest Param, FLOPs, and Testing time.
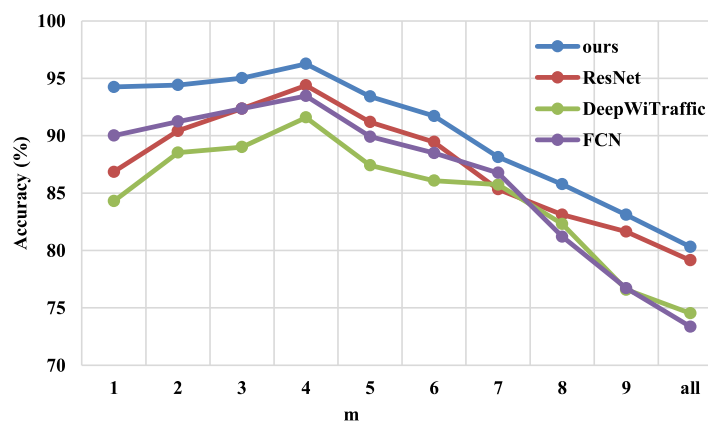

FIGURE 8
The classification accuracy of different CSI factors "m" on the CSI four-category dataset.

TABLE 2   Accuracy of different subcarrier selection methods.

CSI subcarrier selection methods	FCN	DeepWiTraffic	ResNet	LW-WADL
k-subcarrier weight fusion	89.17%	90.86%	90.77%	93.01%
Averaged subcarriers	82.09%	87.89%	83.36%	90.53%
Ours	**93.46%**	**94.59%**	**94.39%**	**96.26%**

The bold values represent that our subcarrier selection method achieves the highest accuracy among the four compared deep learning models.

TABLE 3   Performance evaluation of different classification tasks.

Tasks	FCN	DeepWiTraffic	ResNet	LW-WADL
2	100%	100%	100%	**100%**
3	95.12%	96.86%	96.21%	**97.96%**
4	93.46%	94.59%	94.39%	**96.26%**

The bold values indicate that the designed deep learning model (LW-WADL) outperforms on different classification tasks.

k-subcarrier weight fusion and averaged subcarriers methods will not remove those CSI subcarriers with too low or too high sensitivity, which may have a negative impact on CSI waveform. In this case, the accuracy of the above two methods in Table 2 is lower than that of our proposed method.

4.6. Performance evaluation of different classification tasks

To explore the performance of CSI signals on different classification tasks, three groups of experiments are set up, namely two-classification tasks, three-classification tasks, and four-classification tasks. Each group of experiments selects four methods to test, FCN, DeepWiTraffic, ResNet, and our LW-WADL, respectively, so that the results are more credible. The experimental results are shown in Table 3.

The results of Table 3 shows that the compared methods perform best and the same on the two-classification task, and the accuracy of HVR reaches 100%. However, with the increase of road user categories, the classification accuracy of the used four methods decrease. For three-classification task, the classification accuracy of four methods are 95.12, 96.86, 96.21, and 97.96%, respectively. For four-classification task, the classification accuracy of four methods is 1.66, 2.27, 1.82, and 1.73% lower than three-classification task. Overall, our method achieves more than 96% accuracy in different classification tasks.

4.7. CSI confusion matrices

To further display the recognition accuracy for each class of road users, Figure 9 shows confusion matrices of the classification
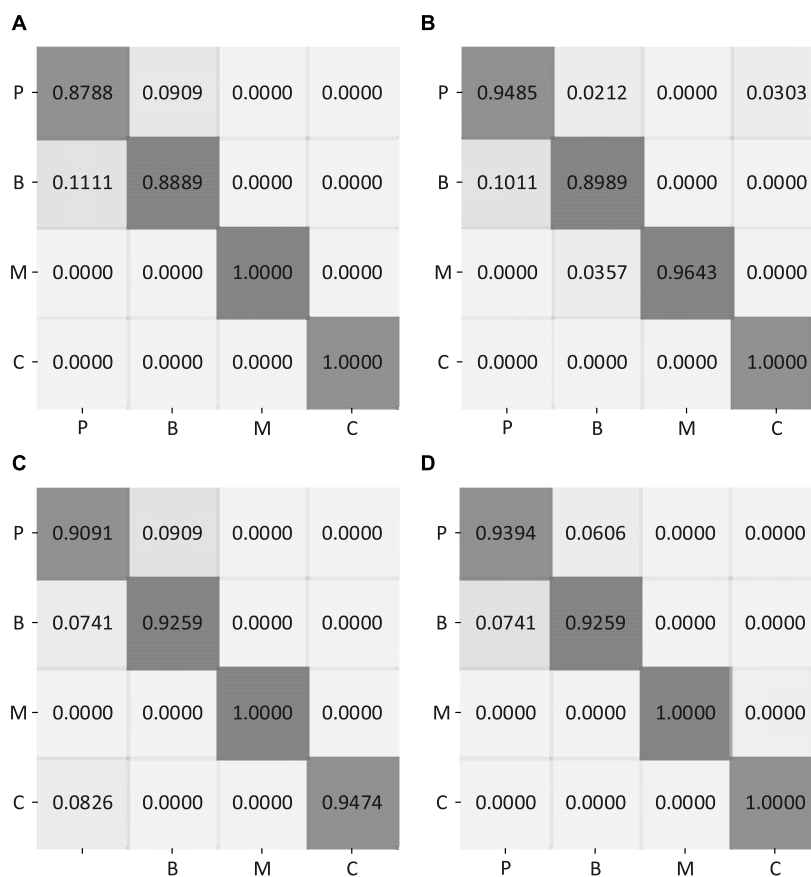

FIGURE 9
Confusion matrices of four methods on CSI four-classification dataset. (A) CSI confusion matrix of FCN. (B) CSI confusion matrix of DeepWiTraffic. (C) CSI confusion matrix of ResNet. (D) CSI confusion matrix of ours. P, pedestrian; B, bicycle; M, motorcycle; C: car.

results, when FCN, DeepWiTraffic, ResNet, and our LW-WADL methods obtain 93.46, 94.59, 94.39, and 96.26% accuracy. As shown in **Figure 9**, the "car" category is well-recognized for most of used methods. Among all the models, only ResNet incorrectly classify the "car" as the "pedestrian." This may be attributed to the fact that a car usually has a much larger volume than other road users. As a result, the attenuation of CSI readings caused by cars is quite different from those cases caused by other road users. For all used models, a major part of the error arises from misclassifying "pedestrian" as "bicycle." This reveals that the group, i.e., "pedestrian" vs. "bicycle" is easily confused with each other. This phenomenon is hinted by the overlap among some real-world road user shapes.

5. Conclusion and future work

This paper has proposed a lightweight wireless sensing attention-based deep learning model (LW-WADL). In order to evaluate the classification ability of LW-WADL, three CSI-based datasets are established, namely two-category dataset, three-category dataset. and four-category dataset. The experimental results on the developed dataset show that the classification accuracy of LW-WADL decreases with the increase of road user categories, but it is higher than 96%. In addition, this paper provides a novel CSI subcarrier selection method, which calculates the SNR of all subcarriers and selects the first four subcarriers with larger SNR for fusion. Besides, a new CSI data enhancement method is exploited to preprocess the change trend of CSI data to one direction, thereby enhancing CSI data.

In future, the performance of other advanced deep learning-based WiHVR methods will be investigated. It is also significant to explore the human-vehicle recognition task based on multiple sets of WTPs. Additionally, it is meaningful to explore the applications of the proposed methods in real scenarios such as multi-lane roads.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

MS was responsible for the writing of the manuscript and some experiments. RZ was responsible for some experiments of the manuscript. XC was responsible for the data collection and processing of the manuscript. CZ was responsible for the revision of the manuscript. LL provided the fund support for this project. All authors contributed to the article and approved the submitted version.

Conflict of interest

RZ was employed by China United Network Communications Co., Ltd., Taizhou Branch.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdelnasser, H., Harras, K., and Youssef, M. (2018). A ubiquitous WiFi-based fine-grained gesture recognition system. *IEEE Trans. Mobile Comput.* 18, 2474–2487. doi: 10.1109/TMC.2018.2879075

Arshad, S., Feng, C., Elujide, I., Zhou, S., and Liu, Y. (2018). "SafeDrive-Fi: A multimodal and device free dangerous driving recognition system using WiFi," in *Proceedings of the 2018 IEEE international conference on communications (ICC)* (Kansas City, MO: IEEE), 1–6. doi: 10.1109/ICC.2018.8422431

Bhat, S. A., Mehbodniya, A., Alwakeel, A. E., Webber, J., and Al-Begain, K. (2020). "Human motion patterns recognition based on rss and support vector machines," in *Proceedings of the 2020 IEEE wireless communications and networking conference (WCNC)* (Seoul: IEEE), 1–6. doi: 10.1109/WCNC45663.2020.9120797

Chen, Z., Zhang, L., Jiang, C., Cao, Z., and Cui, W. (2018). WiFi CSI based passive human activity recognition using attention based BLSTM. *IEEE Trans. Mobile Comput.* 18, 2714–2724. doi: 10.1109/TMC.2018.2878233

Chollet, F. (2017). "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, 1251–1258. doi: 10.1109/CVPR.2017.195

Choy, J. L. C., Wu, J., Long, C., and Lin, Y.-B. (2020). Ubiquitous and low power vehicles speed monitoring for intelligent transport systems. *IEEE Sens. J.* 11, 5656–5665.

de Oliveira, L. F. P., Manera, L. T., and Da Luz, P. D. G. (2020). Development of a smart traffic light control system with real-time monitoring. *IEEE Internet Things J.* 8, 3384–3393. doi: 10.1109/JIOT.2020.3022392

Du, Y., Qin, B., Zhao, C., Zhu, Y., Cao, J., and Ji, Y. (2021). A novel spatio-temporal synchronization method of roadside asynchronous MMW radar-camera for sensor fusion. *IEEE Trans. Intell. Transp. Syst.* 23, 22278–22289. doi: 10.1109/TITS.2021.3119079

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, 770–778. doi: 10.1109/CVPR.2016.90

Huang, Q., Cai, Z., and Lan, T. (2020). A new approach for character recognition of multi-style vehicle license plates. *IEEE Trans. Multimed.* 23, 3768–3777. doi: 10.1109/TMM.2020.3031074

Jiang, Y., Shuai, Y., He, X., Wen, X., and Lou, L. (2021). An energy-efficient street lighting approach based on traffic parameters measured by wireless sensing technology. *IEEE Sens. J.* 21, 19134–19143. doi: 10.1109/JSEN.2021.3089208

Jin, J., and Ma, X. (2019). A multi-objective agent-based control approach with application in intelligent traffic signal system. *IEEE Trans. Intell. Transp. Syst.* 20, 3900–3912. doi: 10.1109/TITS.2019.2906260

Kong, H., Lu, L., Yu, J., Chen, Y., Kong, L., and Li, M. (2019). "Fingerpass: Finger gesture-based continuous user authentication for smart homes using commodity wifi," in *Proceedings of the twentieth ACM international symposium on mobile ad hoc networking and computing*, Catania, 201–210. doi: 10.1145/3323679.3326518

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 25, 1097–1105.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Liu, J., Wang, L., Guo, L., Fang, J., Lu, B., and Zhou, W. (2017). "A research on CSI-based human motion detection in complex scenarios," in *Proceedings of the 2017 IEEE 19th international conference on e-health networking, applications and services (Healthcom)* (Dalian: IEEE), 1–6. doi: 10.1109/HealthCom.2017.8210800

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Boston, MA, 3431–3440. doi: 10.1109/CVPR.2015.7298965

Ma, X., Zhao, Y., Zhang, L., Gao, Q., Pan, M., and Wang, J. (2019). Practical device-free gesture recognition using WiFi signals based on metalearning. *IEEE Trans. Industr. Inform.* 16, 228–237. doi: 10.1109/TII.2019.2909877

Otter, D. W., Medina, J. R., and Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 604–624. doi: 10.1109/TNNLS.2020.2979670

Pan, X., Jiang, T., Li, X., Ding, X., Wang, Y., and Li, Y. (2019). "Dynamic hand gesture detection and recognition with WiFi signal based on 1d-CNN," in *Proceedings of the 2019 IEEE international conference on communications workshops (ICC Workshops)* (Shanghai: IEEE), 1–6. doi: 10.1109/ICCW.2019.8756690

Park, J.-K., Choi, I.-O., and Kim, K.-T. (2021). Length prediction of moving vehicles using a commercial FMCW radar. *IEEE Trans. Intell. Transp. Syst.* 23, 14833–14845. doi: 10.1109/TITS.2021.3134408

Santos, P. M., Rodrigues, J. G. P., Cruz, S. B., Lourenço, T., d'Orey, P. M., Luis, Y., et al. (2018). PortoLivingLab: An IoT-based sensing platform for smart cities. *IEEE Internet Things J.* 5, 523–532. doi: 10.1109/JIOT.2018.2791522

Singh, R., Saluja, D., and Kumar, S. (2021). R-comm: A traffic based approach for joint vehicular radar-communication. *IEEE Trans. Intell. Vehicles* 7, 83–92. doi: 10.1109/TIV.2021.3074389

Sliwa, B., Piatkowski, N., and Wietfeld, C. (2020). The channel as a traffic sensor: Vehicle detection and classification based on radio fingerprinting. *IEEE Internet Things J.* 7, 7392–7406. doi: 10.1109/JIOT.2020.2983207

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, 2818–2826. doi: 10.1109/CVPR.2016.308

Tavanti, E., Rizik, A., Fedeli, A., Caviglia, D. D., and Randazzo, A. (2021). A short-range FMCW radar-based approach for multi-target human-vehicle detection. *IEEE Trans. Geosci. Remote Sens.* 60, 1–16. doi: 10.1109/TGRS.2021.3138687

Wang, F., Gong, W., and Liu, J. (2018). On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach. *IEEE Internet Things J.* 6, 2035–2047. doi: 10.1109/JIOT.2018.2871445

Wang, J., Zhao, Y., Fan, X., Gao, Q., Ma, X., and Wang, H. (2018). Device-free identification using intrinsic CSI features. *IEEE Trans. Vehicular Technol.* 67, 8571–8581. doi: 10.1109/TVT.2018.2853185

Wang, Q., Zheng, J., Xu, H., Xu, B., and Chen, R. (2017). Roadside magnetic sensor system for vehicle detection in urban environments. *IEEE Trans. Intell. Transp. Syst.* 19, 1365–1374. doi: 10.1109/TITS.2017.2723908

Wang, W., Liu, A. X., and Shahzad, M. (2016). "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*, Heidelberg, 363–373. doi: 10.1145/2971648.2971670

Wang, Y., Wu, K., and Ni, L. M. (2016). Wifall: Device-free fall detection by wireless networks. *IEEE Trans. Mobile Comput.* 16, 581–594. doi: 10.1109/TMC.2016.2557792

Wilby, M. R., González, A. B. R., Pozo, R. F., and Díaz, J. J. V. (2020). Short-term prediction of level of service in highways based on bluetooth identification. *IEEE Trans. Intell. Transp. Syst.* 23, 142–151. doi: 10.1109/TITS.2020.3008408

Won, M., Sahu, S., and Park, K.-J. (2019). "DeepWiTraffic: Low cost WiFi-based traffic monitoring system using deep learning," in *Proceedings of the 2019 IEEE 16th international conference on mobile ad hoc and sensor systems (MASS)* (Monterey, CA: IEEE), 476–484. doi: 10.1109/MASS.2019.00062

Won, M., Zhang, S., and Son, S. H. (2017). "WiTraffic: Low-cost and non-intrusive traffic monitoring system using WiFi," in *Proceedings of the 2017 26th international conference on computer communication and networks (ICCCN)* (Vancouver, BC: IEEE), 1–9. doi: 10.1109/ICCCN.2017.8038380

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in *Proceedings of the european conference on computer vision (ECCV)*, Munich, 3–19. doi: 10.1007/978-3-030-01234-2_1

Zhang, S., Zhang, S., Huang, T., and Gao, W. (2017). Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching. *IEEE Trans. Multimed.* 20, 1576–1590. doi: 10.1109/TMM.2017.2766843

Zhang, X., Tang, C., Yin, K., and Ni, Q. (2021). WiFi-based cross-domain gesture recognition via modified prototypical networks. *IEEE Internet Things J.* 9, 8584–8596. doi: 10.1109/JIOT.2021.3114309

Zhao, L., and Huang, Z. (2021). A moving object detection method using deep learning-based wireless sensor networks. *Complexity* 2021:5518196. doi: 10.1155/2021/5518196

Zhao, L., Wang, J., Liu, J., and Kato, N. (2019). Routing for crowd management in smart cities: A deep reinforcement learning perspective. *IEEE Commun. Mag.* 57, 88–93. doi: 10.1109/MCOM.2019.1800603