
A Novel Abnormal Driving Detection Method via Deep Learning in Wireless Sensor Network

Xi Liu¹, Mingyuan Luo², Wei Wang³, and Wei Huang^{*4}
{406130917318@email.ncu.edu.cn¹, 406130917323@email.ncu.edu.cn², wwang@chd.edu.cn³,
n060101@e.ntu.edu.sg⁴}

* indicates the corresponding author

School of Information Engineering, Nanchang University, Nanchang, China^{1,2,4}
School of Economics and Management, Chang'an University, Xi'an, China³

Abstract. In this study, the abnormal driving detection in the current research hotspot wireless sensor network (WSN) is emphatically discussed, and three improved fusion models based on Densely Connected Convolutional Network (DenseNet), which is named Wide Group Densely Network (WGD), Wide Group Residual Densely Network 1 (WGRD1), and Wide Group Residual Densely Network 2 (WGRD2) respectively, are proposed for the first time. WGD introduces two deep learning network indicators, width and cardinality, into DenseNet. WGRD1 and WGRD2, on the basis of WGD, use two different methods to introduce the important idea of ResNet into DenseNet, which is residual-block output and direct-connected streams are added by elements. These three models use end-to-end learning for training. The experimental analysis based on the abnormal driving image data set shows that the performance of our improved model for abnormal driving detection in the wireless sensor network is better than several excellent deep learning models and traditional deep learning models.

Keywords: Abnormal Driving Detection, Wide Group Densely Network, Wide Group Residual Densely Network.

1 Introduction

Wireless sensor network (WSN) has become an important field of contemporary multimedia research because of its great importance in both academic research and daily lives of human beings. Generally speaking, wireless sensor network techniques try to integrate diverse types of sensors, wireless communication techniques, computer algorithms, etc. to realize specific utilizations from the systematic perspective. In the field of automatic driving, multiple sensors are usually equipped both outside and inside the vehicle. These sensors are usually build up within a wireless network, thus the communication and collaboration among different sensors can be realized.

For automatic driving, the driving safety issue is often of the first priority. Also, it is generally accepted that, the driver's behavior needs to be well restricted, in order to avoid any unnecessary or serious incidents. Therefore, multiple sensors in front of the driver can be utilized to monitor the status of the driver in real-time. This network of sensors are of high importance, as abnormal driving behaviors are likely to be identified with the help of multiple sensors. Generally speaking, visual sensors (i.e., including visible / infrared cameras), audio

sensors, etc. are vastly employed to construct the network of sensors for monitoring abnormal behaviors of drivers [1].

According to the definition of the International Organization for Standardization (ISO), abnormal driving refers to the phenomenon that attention is directed to activities unrelated to normal driving, which leads to the decline of driving ability. From the source analysis of drivers' behaviors, abnormal driving behaviors can be divided into three categories. First, distracted driving behaviors that meet drivers' physical comfort needs, such as smoking, drinking water, eating and adjusting air conditioning. The second is to meet the driver's needs for pleasant mood distracted driving behavior, such as makeup, shaving, chatting, also including the use of mobile phone to make phone calls, receive and receive information; Third, distracted driving behaviors caused by the surrounding environment, such as taking care of children and paying attention to emergencies outside a car for a long time. With the rapid development of mobile Internet and the prevalence of various mobile phone applications and communication software, the use of mobile phones has become the most important form of abnormal driving. Scientists have shown in simulations that talking on the phone while driving causes drivers to lose 20 percent of their attention and 37 percent if the content of the call is important. Drivers who text while driving is 23 times more likely to have an accident than those who drive normally. [2] Aiming at this problem, abnormal driving detection technology has been paid attention to and popularized, and its purpose is to judge whether there is a potential danger in the process of driving a motor vehicle.

In this paper, the image captured by the single visible light camera in the wireless sensor network is used to carry out abnormal driving detection by combining several different traditional and classical deep learning models and our own improved models. The traditional and classical deep learning models we use include Convolutional Neural Network (CNN), Wide Convolutional Neural Network (Wide CNN), Group Convolutional Neural Network (Group CNN), Deep Residual Network (ResNet) [3], and Densely Connected Convolutional Network (DenseNet) [4]. Our improved model includes Wide Group Densely Network and two kinds of Wide Group Residual Densely Network. Therefore, the models are tested on the abnormal driving data set to analyze and evaluate their performance. This paper will introduce the work related to abnormal driving and deep learning in Section II, discuss the details of the traditional model and improved model in Section III, introduce the experiment we conducted and theoretical analysis on the experimental conclusion in Section IV, and give the research conclusion in Section V.

2 Related work

2.1 Abnormal Driving Detection

Currently, the commonly used schemes to detect abnormal driving state of drivers include detection based on human physiological signals (i.e., encephalogram, electro-oculogram, blood flow and respiratory changes, respiratory airflow detection methods based on a variety of sensors), detection based on facial details (i.e., eye movement changes, mouth movement changes, head movement changes, hand features and other detection methods) and steering wheel motion characteristics. The detection of human physiological signals based on good real-time performance and high accuracy, but the disadvantage is that it will affect the normal driving of drivers, and human physiological signals vary greatly due to individual and

environmental differences, it is difficult to give a unified quantitative standard. Percentage of eyelid closure over the pupil over time (PERCLOS) [5] monitoring method is one of the more famous methods for face detection-based detection. The percentage of eye closure time in unit time monitored by this method can be expressed as equation (1).

$$PERCLOS = \frac{\text{Eye closing time}}{\text{Detection period}} \times 100\% \quad (1)$$

Abnormal driving is considered to occur when the percentage of closing time reaches 70 percent or 80 percent. PERCLOS detection method has been widely recognized as the most effective method to evaluate abnormal driving in real time. Although the PERCLOS principle can be used to monitor driver status well, there are many problems in practical application. For example, drivers with different physical conditions and living habits have very different eye states and some people don't close their eyes when they sleep, so the misjudgment rate is relatively high. PERCLOS is a multiplicity of non-rigorous object tracking. Changes in interior lighting conditions and head movements can lead to inaccurate or even unsuccessful predictions. When the driver's head is not facing the camera or his hair covers part of his eyes, he cannot identify eye features. These problems make it difficult to use the PERCLOS method. [6]

2.2 Deep Learning

Deep learning is developed from the artificial neural network in machine learning. For neural networks, depth refers to the number of non-linear operation combination levels in the functions obtained by network learning. The learning algorithms of traditional neural networks are mostly aimed at the lower level network structure, which is called as the shallow neural network, such as an input layer, a hidden layer and an output layer of neural network, as shown on the left side of **Figure 1**. On the contrary, the network with a higher level of non-linear operation combination is called deep neural network, such as one input layer, three hidden layers, and one output layer, as shown on the right side of **Figure 1**. Deep learning has many advantages over shallow learning, which indicates the necessity of introducing deep learning.

In terms of the ability of the network to express the complex objective function. Sometimes, the shallow neural network is unable to implement the representation of complex high-dimensional functions such as high-variable functions, while the deep neural network can better represent it.

In terms of the computational complexity of network structure. When a network structure with a depth of k can be used to express a function compactly, if a network structure with a depth less than k is used to express the function, it may need to increase the computational factors of exponential scale and number, which greatly increases the computational complexity. In addition, it is necessary to use training samples to adjust the parameter values of calculated factors. When the number of training samples of a network structure is limited and the number of calculated factors increases, its generalization ability will become very poor.

In terms of bionics, the deep learning network structure is the best simulation of the human cerebral cortex. Like the cerebral cortex, deep learning processes input data in a

hierarchical manner, using each layer of the neural network to extract different levels of characteristics of the original data.

In terms of information sharing, the multi-level extracted features obtained by deep learning can be reused in similar different tasks, which is equivalent to providing some unsupervised data for task solving and obtaining more useful information.

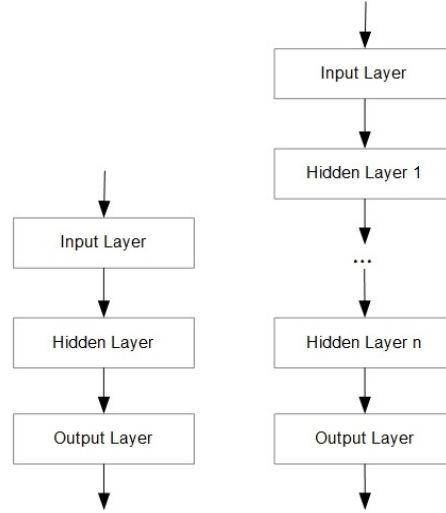


Fig. 1. Shallow neural network (left) and deep neural network (right).

Deep learning is more expressive than shallow learning and is able to learn potentially higher-level abstract features from data. In 2006, *Hinton* et al. proposed an unsupervised learning algorithm for Deep Belief Network (DBN) [7], which solves the problem of difficult optimization of the deep learning model. In 2012, *Hinton* research group used the designed deep convolutional neural network model AlexNet [8] to win the championship in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition [9], which made the deep convolutional neural network attract attention. Since then, deep learning has entered an explosive period. Various deep learning models (e.g., VGG [10], GoogleNet [11], ResNet, and DenseNet, etc.) have achieved better and better results in ImageNet competitions, which makes deep learning widely concerned. Deep learning model can achieve excellent results in image classification[12], object detection [13], and other fields.

3 Methodology

In this study, for abnormal driving detection, we introduced and analyzed five traditional and classical deep learning models in detail, including Convolutional Neural Network (CNN), Wide Convolutional Neural Network (Wide CNN), Group Convolutional Neural Network (Group CNN), Deep Residual Network (ResNet), and Densely Connected Convolutional Network (DenseNet). We designed three fusion deep models which are named Wide Group Densely Network (WGD), Wide Group Residual Densely Network 1 (WGRD1), and Wide

Group Residual Densely Network 2 (WGRD2) respectively. These three models were designed based on DenseNet with excellent characteristics of other models. Generally speaking, the main objective of abnormal driving detection is to detect the driver's driving state and judge whether the state is normal or not. That is to take an image as the input of the model and judge whether the state of the driver is normal according to the output label.

3.1 Deep learning models

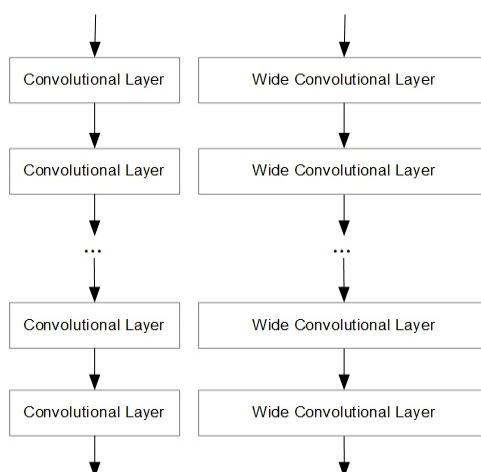


Fig. 2. Convolutional Neural Network (left) and Wide Convolutional Neural Network (right).

Convolutional Neural Network (CNN). The most original convolutional neural network LeNet-5 [14] was born in 1998, proposed by *Yan LeCun*, one of the three giants of deep learning, who is also known as the father of convolutional neural network. LeNet-5 is mainly used for recognition and classification of handwritten characters, with an accuracy of 98%. LeNet-5 has been put into use in American Banks and used to read about 10% of North American cheques. LeNet-5 has laid the foundation of the modern convolutional neural network, whose structure consists of the convolutional layer, pooling layer, and full convolutional layer. Convolutional layer and pooling layer are combined to form multiple convolution groups to extract features layer by layer, and finally, classification is completed through a number of fully connected layers. The AlexNet model proposed in 2012 is another breakthrough of the convolutional neural network. AlexNet won the championship in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition in that year, which proved the effectiveness of convolutional layer under the complex model. The deep convolutional network was able to learn the potential high-level characteristics of data, which made the convolutional neural network attracted attention.

Wide Convolutional Neural Network (Wide CNN). The idea of wide convolution neural network comes from Wide Residual Network (WRN) [15]. Wide residual network is on the basis of the Deep Residual Network (ResNet) to increase the number of layer convolution kernels. **Figure 2** shows the difference between the traditional convolution network and the wide convolution network. As the depth of the model deepens, the weight of each residual module cannot be guaranteed to flow through when the gradient is propagated backward, so it

is difficult for it to learn something. Therefore, in the whole training process, only a few residual modules can learn useful expressions, and most of the residual modules play a small role. The wide residual network hopes to use a shallow model with wider width (i.e. more convolution cores) to improve the performance of the model more effectively. Relevant experiments have also proved that the performance of the shallow but wide residual network is comparable to that of the deep but narrow residual network. On this basis, we directly focus on the feature of the wide convolution layer as the wide convolution neural network.

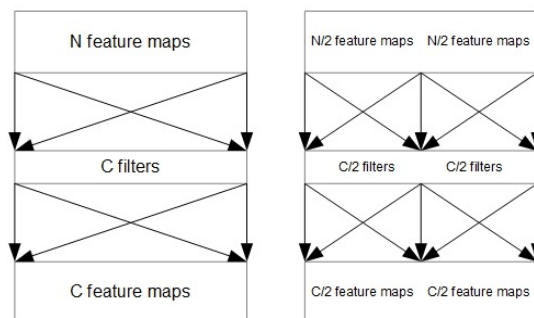


Fig. 3. Traditional convolutional layer (left) and group convolutional layer (right) with the group is 2.

Group Convolutional Neural Network (Group CNN). The difference between group convolution and traditional convolution is that in traditional convolution, each filter operates on all channels, while each filter in group convolution does not operate on all channels, but operate on only one set of channels., as shown on the left side of **Figure 3**. Group CNN first appeared in AlexNet. Due to the limited hardware resources at that time, convolution operations could not be all processed in the same GPU when AlexNet was trained. Therefore, the author divided feature maps into multiple GPUs for processing respectively and finally fused the results of multiple GPUs. The ResNeXt model [16] improved based on the residual network also uses the idea of group convolution. ResNeXt adopts the idea of VGG stack and Inception's "split-transform-merge" strategy [17]. It has strong scalability, improves the accuracy rate without increasing the complexity of parameters, and at the same time reduces the number of super parameters. ResNeXt presents the dimension of "cardinality" different from depth and width, and shows that increasing cardinality is more effective than increasing depth and width. At the same time, the aggregated residual structure of ResNeXt model is completely equivalent to group convolution.

Deep Residual Network (ResNet). Deep residual network is the latest technical concept in the field of deep learning proposed by *Kaiming He* from 2015 to 2016. It can be used to solve degradation problems and accelerate the convergence speed of the deep network. In the contest of ImageNet and COCO2015, 152-layer deep residual network achieved the best results in image classification, target detection and semantic segmentation, and related papers were awarded the best CVPR paper twice in a row. The residual structure diagram is shown in **Figure 4**. The core idea of this method is to add a parallel identity mapping layer to the original network layer to form a residual learning structure. To be specific, assuming that the potential mapping that needs to be learned is $H(x)$, then a nonlinear stacked network layer is constructed to correspond to another residual mapping $F(x):=H(x)-x$. In this way, the potential

mapping can be written as $F(x)+x$. However, it is easier to optimize the residual mapping of the original network layer than to optimize potential mapping.

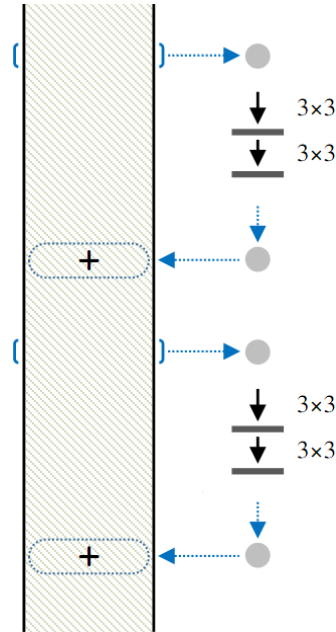


Fig. 4. The structure of Deep Residual Network (ResNet).

Densely Connected Convolutional Network (DenseNet). The DenseNet mainly refers to ResNet, Highway Network [18] and GoogleNet. The construction of DenseNet is based on the following. If the convolution layer is closer to the input layer or the output layer, then it is easier to converge. Therefore, the practice of DenseNet is more violent than ResNet. The ResNet only adds the output between the front and rear layers by elements, while the DenseNet requires that all layers are related, as shown in **Figure 5**. That is, if the ResNet has L layers, there will be $L-1$ connection, while the DenseNet has one connection between every two layers, that is $L*(L+1)/2$ connections. When DenseNet changes the traditional network reverse transmission, gradient (or information) transmission mode changes from linear to treelike reverse, which has the advantage of reducing the possibility of gradient disappearance and accelerating training, which is conducive to the training of deeper network.

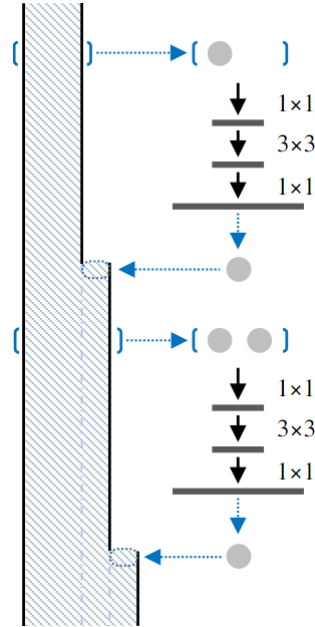


Fig. 5. The structure of Densely Connected Convolutional Network (DenseNet).

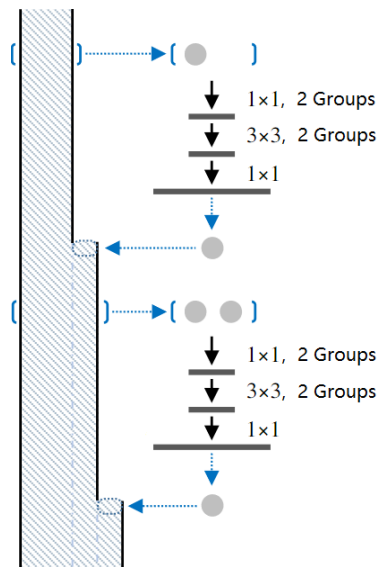


Fig. 6. The structure of Wide Group Densely Network (WGD).

3.2 Fusion deep models

Wide Group Densely Network (WGD). On the basis of DenseNet, convolution is replaced by the group and wide convolution. Grouping makes convolution contain fewer parameters, but the network generalization ability decreases at the same time. Therefore, the network generalization ability can be guaranteed by appropriately increasing the number of feature graphs, so that the model has fewer parameters and better performance, as shown in **Figure 6**.

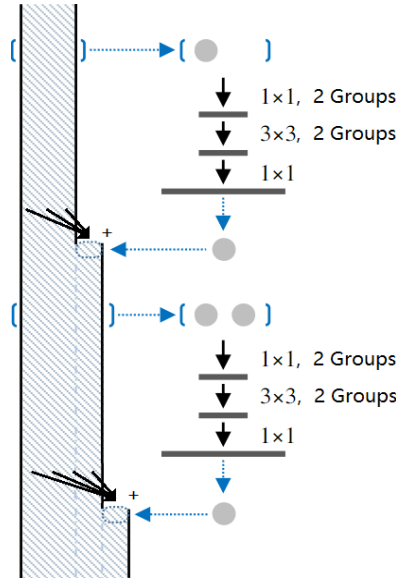


Fig. 7. The structure of Wide Group Residual Densely Network 1 (WGRD1).

Wide Group Residual Densely Network 1 (WGRD1). On the basis of WGD, the operation of parallel output is changed. First, the directly connected data streams are divided into parallel channels in each parallel connection, and then the divided channels are added as elements. Next, the output of the current dense block is added as an element, and the result of the addition is parallel to the directly connected data stream, as shown in **Figure 7**. We know that adding the output of the directly connected data stream and the residuals block by elements is the most important step of the residuals network. This allows the direct connect data stream to add up the output of all the previous residual blocks. This step in DenseNet is as follows. First, add the data flow inside the direct connection, then add the output, and finally parallel to the direct connection data flow. This is the main idea of WGRD1. The purpose of this design is to make each part of the directly connected data stream is the superposition of the output of different dense blocks, so as to generate more complex features and strengthen the learning ability of the whole network. To be specific, for the input image x_0 defines an L -layer network, its l -layer is a non-linear transformation (e.g., BN, ReLU, Conv, etc.), set as $H_l(\cdot)$, and the output is x_l , so we generally have $x_l = H_l(x_{l-1})$. DenseNet and WGD can be expressed by equation (2).

$$x_i = H_i[x_0, x_1, \dots, x_{i-1}] \quad (2)$$

while WGRD1 can be expressed by equation (3).

$$x_i = H_i[x_0, x_0 + x_1, \dots, \sum_{i=0}^{i-1} x_i] \quad (3)$$

where $\sum_{i=0}^{i-1} x_i$ means $x_i(i=0,1,\dots,i-1)$ is added to each of the elements and the operator $[\cdot]$ represents parallel operations.

Wide Group Residual Densely Network 2 (WGRD2). There is a second way to introduce adding by elements which are the important step of ResNet into DenseNet, similar to WGRD1, is different to the operation of the dense blocks of output and input. The output of the dense block is first parallel to the directly connected data stream, which is the same as DenseNet and WGD. When entering the next dense block, the directly connected data stream at this time is not taken as the input directly. The directly connected data flow is divided into parallel channels in each parallel connection, and the divided channels are added as input elements, as shown in **Figure 8**. WGRD2 can be expressed by equation 4.

$$x_i = H_i(\sum_{i=0}^{i-1} x_i) \quad (4)$$

This is similar to ResNet, if the ReLU after each residual block in ResNet is removed, it will be the same. It is worth noting that this is very different from WGRD1, where the directly connected data stream of WGRD1 is the output of different dense blocks in parallel after the addition of elements, while the directly connected data stream of WGRD2 is the output of multiple dense blocks in parallel.

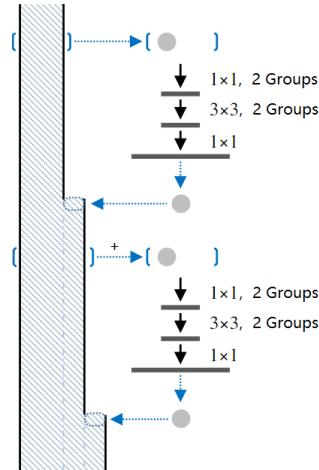


Fig. 8. The structure of Wide Group Residual Densely Network 2 (WGRD2).

3.3 Loss function

Since it is a classification problem, cross entropy loss is used to evaluate the performance of different models. It is given that the driver's image data set is defined as the i th driver image in the image data set, and y_i is defined as the corresponding one-hot label of x_i (i.e., y_i is a c -dimensional row vector with a total of c categories). Given that the prediction function of the model is $P(\cdot)$, $y_i' = P(x_i)$ is the probability that the model predicts that x_i belongs to each category (i.e., y_i' is also a c -dimensional row vector). Therefore, the cross-entropy loss can be expressed as equation 5.

$$L = -\sum_{i=1}^m \sum_{j=1}^c y_{ij} \log y_{ij}' = -\sum_{i=1}^m \sum_{j=1}^c y_{ij} P(x_{ij}) \quad (5)$$

where m represents the number of driver images. It is easy to see that the purpose of this loss function is to minimize the difference between the predicted probability distribution (i.e., y_i') and the real probability distribution (i.e., y_i). Make the predicted probability distribution close to the real probability distribution and improve the classification accuracy. We use stochastic gradient descent algorithm to optimize cross entropy loss.

4 Experiments

4.1 Database and Experimental Settings

To verify the effect of deep learning model and fusion depth model on abnormal driving detection, Kaggle competition platform was used to obtain relevant competition data sets. The data set consists of 22,424 RGB driver images of 640*480 in 10 different categories. These 10 categories include safe driving, texting (right), talking on the phone (right), texting (left), talking on the phone (left), operating the radio, drinking, reaching behind, hair and makeup, and talking to passenger, as shown in **Figure 9**.

On this dataset, we compare the deep learning model with the fusion deep model. The deep learning model included Convolutional Neural Network (CNN), Wide Convolutional Neural Network (Wide CNN), Group Convolutional Neural Network (Group CNN), Deep Residual Network (ResNet) and Densely Connected Convolutional Network (DenseNet). The improved fusion deep models included Wide Group Densely Network (WGD), Wide Group Residual Densely Network 1 (WGRD1), and Wide Group Residual Densely Network 2 (WGRD2).

We preprocessed each image in the dataset, including adding noise, changing the intensity of light, changing color, rotating the image, and scaling the image numerically from 0 to 1. The purpose of such preprocessing is to make the data distribution more random and diversified, to strengthen the learning ability of each model and improve the generalization ability. After preprocessing, we randomly divide images into a training set and test set in each category of data set. The training set and the test set respectively contain a considerable number of images for each category. After the partition, we trained the model on the training set. After the training, the category of each image was predicted on the test set and compared with the real category. The performance of each model is evaluated by this method.



Fig. 9. Abnormal driving image dataset. The images show safe driving, texting (right), talking on the phone (right), texting (left), talking on the phone (left), operating the radio, drinking, reaching behind, hair and makeup, and talking to passenger from top to bottom and left to right.

Repeat the process to reduce the uncertainty caused by randomness. Experimental results are based on an average estimate of all experimental results. The batch size of all models is set to 32, the number of training epochs is set to 10, and the learning rate is set to 0.0001. In order to obtain the optimal comprehensive performance of each model, the errors of these super-parameters must be compared. It is worth pointing out that WGD has a total of 1.54m parameters, WGRD1 has a total of 1.58m parameters, and WGRD2 has a total of 1.42m parameters. For other comparison models, CNN, Wide CNN, Group CNN, ResNet, and DenseNet have parameters of 9.953m, 19.356m, 5.252m, 11.181m and 0.41m respectively. Therefore, WGD, WGRD1, WGRD2, and DenseNet have less training time. All implementations are implemented through CentOS 7 operating system, based on PyTorch 1.0.0, using workstations equipped with Intel Xeon Silver 4110 CPU, 128G RAM, and Nvidia Titan V GPU card.

4.2 Experimental Results and Statistical Analysis

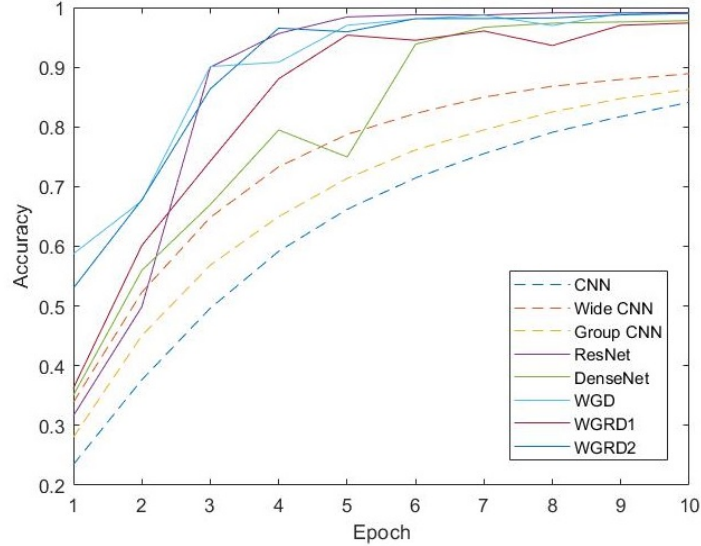


Fig. 10. The accuracy curve of all models.

Figure 10 shows the variation curve of classification accuracy of all models in the test set with the increase of training epochs. As can be seen from the **Figure 10**, the accuracy of classification prediction of all 8 models increases with the increase of training epochs and tends to be stable at a higher accuracy, which is good evidence of convergence of all models. Specifically, WGD and WGRD2 achieved high accuracy in the first epoch of training, and then the accuracy growth rate was at the average level of all models, and finally reached nearly 100% accuracy. ResNet, DenseNet, and WGRD1 achieved low accuracy in the first round of training, but after that, the growth rate of accuracy was at a fast speed, and finally reached nearly 100% accuracy. Finally, CNN, Wide CNN, and Group CNN had low accuracy at the beginning, and then their accuracy grew at a relatively steady rate, and the accuracy was between 80% and 90% at the end.

We used the Precision-Recall (P-R) curve to qualitatively describe the 8 models involved in the comparison. Since the abnormal driving image data set is a multi-classification dataset, we use the following method to calculate the P-R curve values. Suppose the number of test samples is m and the number of categories is n (i.e., suppose the category labels are $0, 2, \dots, n-1$). After the training is completed, the probability or confidence of each test sample in each category is calculated, and an $[m, n]$ shaped matrix P is obtained. Each row represents the probability value of a test sample in each category (sorted by category label). Accordingly, the tags of each test sample are converted into a binary one-hot form, and each position is used to mark whether it belongs to the corresponding category (also sorted by tags, to correspond to the probability matrix P), so that an $[m, n]$ tag matrix L can be obtained. For a test sample, the label consists only of 0 and 1, the position of 1 indicates its category (which can correspond to

positive examples in the dichotomous problem), and 0 indicates other categories (which correspond to negative examples in the dichotomous problem). If the classifier classifies the test sample correctly, the probability value of the position corresponding to 1 in the sample label in the probability matrix P is greater than the probability value of the position corresponding to 0. Based on these two points, the label matrix L and the probability matrix P are expanded by row respectively, and the transpose results in two columns, which results in a binary classification result. Therefore, the final P-R curve values can be obtained after calculation.

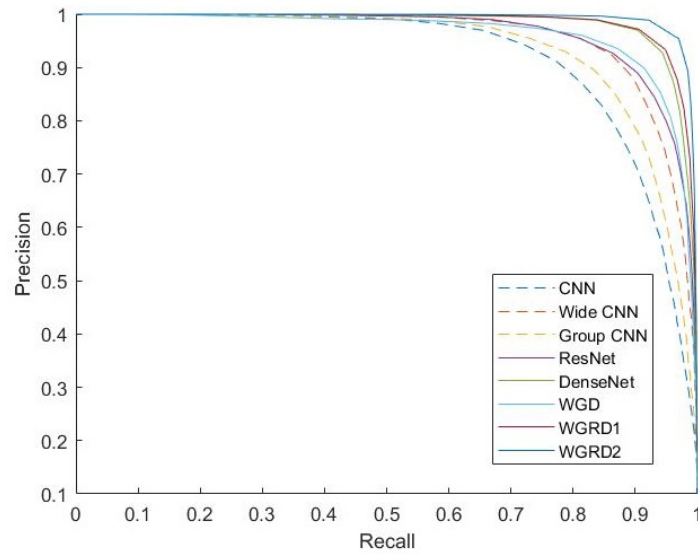


Fig. 11. The Precision-Recall (P-R) curve of all models.

Figure 11 shows the P-R curves of 8 different models. It is worth pointing out that the actual area under the P-R curve of each model reflects the accuracy of the model in predicting classification on the data set. As can be seen from the **Figure 11**, our improved fusion model WGRD2 achieved the best performance, followed by WGRD1, DenseNet, WGD, ResNet, Wide CNN, Group CNN, and finally is CNN. There is a clear trend here that compared with the CNN series models (i.e., CNN, Wide CNN, and Group CNN), other models involved in the comparison have more complex structures but relatively fewer parameters and all have higher P-R curves and better performance. The reason for this conclusion can be considered as that complex network structure can learn complex and diverse high-level features, which are more widely distributed, can strengthen network learning and reduce the degree of overfitting, so that the network generalization performance is higher.

In terms of the CNN series models, after convolution is replaced by group convolution, the number of parameters decreases and the performance improves (i.e., the P-R curve of Group CNN is above the P-R curve of CNN). This result verifies the validity of the concept of cardinality. As the third dimension different from width and depth, increasing the cardinality can improve the network performance. Wide CNN has more parameters than CNN and Group CNN, so it has a larger capacity and a higher performance.

The performance of WGRD2 and WGRD1 in our improved fusion models exceed that of DenseNet, which is enough to prove that our improvement is effective. Introducing the idea of adding elements of ResNet into DenseNet can improve the performance of DenseNet. At the same time, the performance of the improved fusion model WGD is lower than that of DenseNet but exceeds ResNet. The possible reason for this result is that the widening of convolution in the WGD network structure increases the number of feature maps output by each dense blocks, but the number has been greatly reduced after the transition layers. The proportion of reduction is much higher than that of the original DenseNet, which makes the WGD network structure have a bottleneck and limits the performance of WGD.

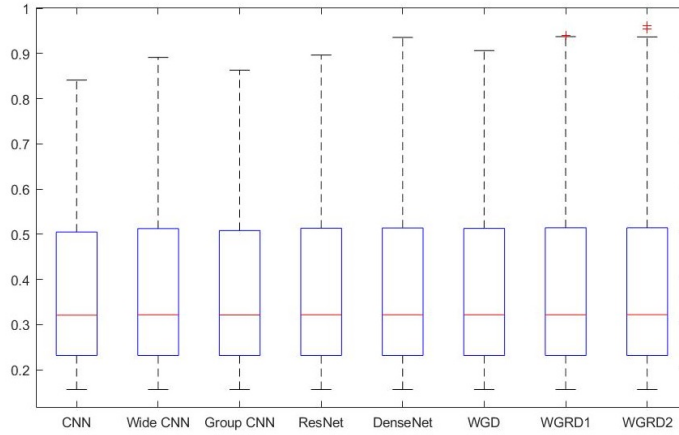


Fig. 12. The F-measure boxplot of all models.

Then, we use a statistically significant quantitative comparison method, F-measure boxplot, to conduct a quantitative analysis of the 8 models, as shown in **Figure 12**. F-measure is calculated according to the unbiased definition of equation 6.

$$F - \text{measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

In **Figure 12**, the red horizontal line on each box represents the median value of F-measure, and the upper and lower quartile of F-measure is represented by the blue line above and below the median value of each box. Draw a vertical dotted line from the upper quartile and the lower quartile to their most extreme data points, which are within the 1.5 quartile range (IQR). Each data point over 1.5 IQR is marked with a red plus sign. As can be seen from **Figure 12**, the improved statistical box of WGRD2 model in this paper is higher than that of other comparable models, which indicates that the result of F-measure of WGRD2 model is better than that of other models. Meanwhile, our improved WGRD1 ranked second and WGD ranked fourth. This result is exactly the same as that of the P-R curves, which supports the conclusion of the P-R curves.

Table 1. Point estimation and interval estimation between WGD and other 5 models.

Model 1	Model 2	Point estimation	Interval estimation
WGD	CNN-12	0.00099	[0, 0.06529]
WGD	Wide CNN-12	0.00012	[0, 0.01491]
WGD	Group CNN-12	0.00052	[0, 0.04313]
WGD	ResNet-18	0.00012	[0, 0.00992]
WGD	DenseNet-29	0.00005	[-0.02922, 0]

Table 2. Point estimation and interval estimation between WGRD1 and other 5 models.

Model 1	Model 2	Point estimation	Interval estimation
WGRD1	CNN-12	0.00099	[0, 0.09960]
WGRD1	Wide CNN-12	0.00012	[0, 0.04922]
WGRD1	Group CNN-12	0.00052	[0, 0.07744]
WGRD1	ResNet-18	0.00012	[0, 0.04423]
WGRD1	DenseNet-29	0.00005	[0, 0.00509]

Table 3. Point estimation and interval estimation between WGRD2 and other 5 models.

Model 1	Model 2	Point estimation	Interval estimation
WGRD2	CNN-12	0.00110	[0, 0.12057]
WGRD2	Wide CNN-12	0.00023	[0, 0.07019]
WGRD2	Group CNN-12	0.00063	[0, 0.09841]
WGRD2	ResNet-18	0.00023	[0, 0.06520]
WGRD2	DenseNet-29	0.00016	[0, 0.02606]

In addition, Table 1, 2, and 3 respectively quantitatively describe the pairwise comparison details between WGD, WGRD1 and WGRD2 and the other 5 models. Point estimation and interval estimation are used in the quantitative description. Among them, point estimation refers to the direct difference of F-measure value between the two models, that is the value of Model 1 minus the value of Model 2, and interval estimation refers to the interval with 95% confidence that the direct difference of F-measure between the two models may fall into. As can be seen from Table 2 and 3, when WGRD1 and WGRD2 are compared with the other 5 models, point estimation is positive, which indicates that from the perspective of point estimation, the F-measure of WGRD1 and WGRD2 is higher than that of other methods. From the perspective of interval estimation, the upper bound of these intervals is always positive and the lower bound is always 0, which indicates that the difference between WGRD1 and WGRD2 and the F-measure of the other 5 models is always positive from the perspective of interval estimation (i.e., the F-measure of WGRD1 and WGRD2 is higher). This shows that WGRD1 and WGRD2 perform better than other models. As can be seen from Table 1, the difference estimation of F-measure between WGD and the other 4 models (i.e., except DenseNet) is positive, and the interval estimation is positive. This indicates that WGD performs better than the other four models (i.e., except DenseNet). These two conclusions are consistent with the results of P-R curves and F-measure boxplots.

4.3 Discussions

In this section, more details on the use of WGD, WGRD1, and WGRD2 in abnormal driving detection are presented. On the dataset of our experiment, WGD, WGRD1, and WGRD2 obtained more accurate classification results than ResNet and DenseNet (WGD was only more accurate than ResNet classification). From the specific images below, **Figure 13-17** respectively show driver images that WGD, WGRD1, and WGRD2 can correctly classify, but ResNet and DenseNet cannot correctly classify (WGD only shows images that are more accurate than ResNet classification). As can be seen from these examples, when the light in the cab is very dark, even if the driver is driving normally, it is easy to identify as abnormal driving. However, for some abnormal driving conditions that are very similar to normal driving, for example, the driver's eyes are not looking at the road but slightly inclined to look at other places, such abnormal driving conditions are easy to be missed, which is very challenging. The above examples can be summarized as follows. WGRD1 and WGRD2 can solve the problem of detecting abnormal driving well in extremely challenging situations (e.g., the cab is dark and abnormal driving is very similar to normal driving).



Fig. 13. The examples where WGD classifies correctly but ResNet does not.



Fig. 14. The examples where WGRD1 classifies correctly but ResNet does not.



Fig. 15. The examples where WGRD1 classifies correctly but DenseNet does not.



Fig. 16. The examples where WGRD2 classifies correctly but ResNet does not.



Fig. 17. The examples where WGRD2 classifies correctly but DenseNet does not.

5 Conclusion

In this study, based on the design of DenseNet, we proposed improved fusion models WGD, WGRD1 and WGRD2, and applied them to abnormal driving detection in wireless

sensor network for the first time. We have two main ideas for improvement. First, the idea of width from AlexNet and Wide CNN and the idea of cardinality from ResNeXt are introduced into DenseNet. Second, the core operation of adding the output of ResNet's residuals block to the directly connected data stream as elements are introduced into DenseNet. The improved model was tested and evaluated on an abnormal driving data set along with other classic superior models. All experimental results show that all the models can solve the problem of abnormal driving detection very well. At the same time, our improved fusion models can achieve the highest detection accuracy among all models. In the future, we will consider using models with fewer parameters but better results for abnormal driving detection.

Acknowledgments. The authors would like to acknowledge the grant 61862043 approved by National Natural Science Foundation of China, key grants 20181ACB20006 and 20171ACB21017 approved by Natural Science Foundation of Jiangxi Province for supporting this study.

References

- [1] Cordell J P, Welland R V, McKelvie S J, et al. Method and Apparatus for Providing Driver Information Via Audio and Video Metadata Extraction: U.S. Patent Application 15/672,832[P]. 2018-2-15.
- [2] Olson R L, Hanowski R J, Hickman J S, et al. Driver distraction in commercial vehicle operations[R]. United States. Federal Motor Carrier Safety Administration, 2009.
- [3] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [4] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//CVPR. 2017, 1(2): 3.
- [5] Knipling R F, Wierwille W W. Vehicle-based drowsy driver detection: Current status and future prospects[M]. National Highway Traffic Safety Administration, Office of Crash Avoidance Research, 1994.
- [6] Acioğlu A, Erçelebi E. Real time eye detection algorithm for PERCLOS calculation[C]//Signal Processing and Communication Application Conference (SIU), 2016 24th. IEEE, 2016: 1641-1644.
- [7] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. *Neural computation*, 2006, 18(7): 1527-1554.
- [8] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [9] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [10] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv preprint arXiv:1409.1556*, 2014.
- [11] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [12] Huang Y, Cheng Y, Chen D, et al. GPipe: Efficient Training of Giant Neural Networks using Pipeline Parallelism[J]. *arXiv preprint arXiv:1811.06965*, 2018.
- [13] Singh B, Najibi M, Davis L S. SNIPER: Efficient Multi-Scale Training[J]. *arXiv preprint arXiv:1805.09300*, 2018.

-
- [14] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [15] Zagoruyko S, Komodakis N. Wide Residual Networks[J]. 2016.
- [16] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. IEEE, 2017: 5987-5995.
- [17] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//AAAI. 2017, 4: 12.
- [18] Srivastava R K, Greff K, Schmidhuber J. Highway Networks[J]. Computer Science, 2015.