ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

A Study on Safety Driving of Intelligent Vehicles Based on Attention Mechanisms

Shukai Ding[1] and Jian Qu[2]

ABSTRACT

Intelligent vehicles attempt to improve the daily transportation of people, but their safety has been questioned after numerous traffic accidents. This paper proposes a Safety Driving Framework (SDF) to improve the ability of intelligent vehicles to avoid risks during emergencies. When performing autonomous driving tasks, the SDF interacts with the environment using a camera, makes decisions using convolutional neural networks, and can perform an emergency stop if it encounters an obstacle. In this paper, we examine the potential of attention mechanisms to enhance the performance of convolutional neural networks and construct four convolutional neural networks with attention mechanisms to use in experiments. Additionally, we extend a dataset and enhance the robustness of the model by implementing data augmentation (DA) techniques. We train the model using 10-fold cross-validation. In this article, we build an intelligent driving platform and a simulation track for simulation testing. The experimental results show that CNNs using data reinforcement and with an attention mechanism perform better than existing models. In particular, ENetb0-SE has an average recognition rate of 95.6% for obstacles and an accident rate of 2%, which is much better than existing models.

Article information:

Keywords: Attention Mechanisms, CNN, Jetson Nano, Data Augmentation

Article history:
Received: June 7, 2022
Revised: August 22, 2022
Accepted: September 17, 2022
Published: October 8, 2022 (Online)

DOI: 10.37936/ecti-cit.2022164.248674

1. INTRODUCTION

Autonomous driving is a critical future development direction for the automotive industry. In 1995, Pomerleau et al. at Carnegie Mellon University designed a simple self-driving car, the NavLab5 [1]. The NavLab5 is equipped with a camera that detects lane lines in the view of the vehicle and a vision system that controls the turning of the vehicle laterally, while the throttle and brakes of the vehicle are controlled manually. The NavLab5 was a huge success, which attracted a large number of research institutions and universities to work on autonomous driving projects. In 2004, DARPA, a U.S. military research institute, organized a global self-driving car race in the Mojave desert. Google was one of the first companies to enter the self-driving car market, setting up a self-driving car development project led by Thrun in 2009, and by 2014 it was demonstrating its first self-driving prototype. Simultaneously, Mercedes-Benz, BMW, Toyota, Ford, Bosch, Google, Baidu, Uber, and Tesla have all made significant progress toward developing and launching representative autonomous vehicles. However, the development of autonomous driving technology has also exposed the problems of intelligent vehicles in safe driving. Tesla and Google self-driving cars have previously experienced issues with out-of-control acceleration and an inability to detect obstacles in autopilot mode, causing numerous accidents. In March 2018, a self-driving car from Uber caused the first smart car death in the world.

The frequent occurrence of safety incidents has become a primary obstacle to the development of intelligent vehicles. Self-driving cars cause nearly five times as many traffic accidents as conventional vehicles, according to a 2015 report [2] from the Transportation Research Institute of the University of Michigan. An analysis of this report indicates that a primary contributing factor to the high rate of accidents in self-driving cars is that environmental awareness in autonomous vehicles remains woefully inadequate for responding to unexpected circumstances. Environmental awareness is the ability of a vehicle to recognize information about its surroundings, in partic-

[1,2]The authors are with Faculty of Engineering and Technology, Panyapiwat Institute of Management, Nonthaburi, Thailand 11120, Email: 6372100126@stu.pim.ac.th and jianqu@pim.ac.th

ular other vehicles, pedestrians, and obstacles with potentially dangerous characteristics. In intelligent vehicles, the ability to perceive is enhanced, reducing the possibility of accidents to some extent.

Many researchers have added multiple expensive sensors to their vehicles to solve this problem. Nanding and Jing [3] employ cameras and infrared sensors, while Rhe et al. [4] employ cameras and ultrasonic sensors to enable intelligent vehicles to perform road tracking and obstacle avoidance tasks. Intelligent vehicles detect obstacles by emitting infrared light and emitting ultrasonic waves. It is how they identify and avoid obstacles. Ranjbar and Vig [5] use cameras and color sensors to enable their intelligent vehicles to perform road tracking and steering tasks. By detecting the color of the turn mark, the color sensor transmits the information to the control platform, which then completes the turn operation.

These studies enhance the ability of intelligent vehicles to perceive their environment by diversifying the input data generated by sensors. However, this introduces additional problems: increased manufacturing costs for intelligent vehicles, complex systems, and increased development and maintenance costs.

Additionally, some studies [6-9] have utilized reinforcement learning as a strategy to improve the safety of self-driving vehicles. The principle of reinforcement learning states that when intelligent vehicles are stimulated by rewards or punishments provided by a particular environment, they gradually develop an expectation of stimulation and develop habitual behaviors that maximize benefits. Intelligent vehicles based on reinforcement learning learn the best strategies by improving their perception of a given environment through trial and error in that environment. Reinforcement learning, however, is strongly dependent on the training environment, and intelligent vehicles are unable to cope with environments they have not experienced before. Furthermore, intelligent vehicles using trial and error for reinforcement learning are prone to traffic accidents in the real world, and the cost is too high.

The approaches described above each attempt to improve the safety of autonomous vehicles in terms of adding multiple sensors and using reinforcement learning. But their approaches all increase the cost of autonomous vehicles. Therefore, it becomes crucial to improve vehicle perception and efficiently reduce accident rates. In human cognitive science, humans selectively focus on a portion of all information and thus ignore the rest of the visible information. Humans can focus their attention on what is important, increasing the efficiency of their perception of the environment.

In order to address these issues, we have focused our attention on neural networks. We discovered a 'component' that has a wide range of applications in the field of natural language processing: the Attention Mechanism. It shares similarities with human attention. The attention mechanism is a signal processing mechanism that was discovered in the 1990s by many scientists studying human vision. The attention mechanism is a technique in artificial neural networks that mimics cognitive attention. This mechanism focuses the attention of the network on a small part of the data that is most important by increasing the weight of some parts of the input data to the neural network and decreasing the weight of other parts.

Attention mechanisms are also known to have applications in image processing. Jie Hu, Li Shen, and Gang Sun [10] proposed Squeeze-and-Excitation Networks (SE) in 2018 and used them in conjunction with ResNet to win the classification project at ILSVRC 2017. Their method reduced the top-5 error to 2.251% on the ImageNet dataset, compared to the previous best of 2.991%. Subsequently, Sanghyun Woo et al. proposed the Macroeconomic Block Attention Module (CBAM) [11], a simple and effective attention module for convolutional neural networks. Their study experimented with ImageNet-1K, MS COCO detection, and VOC 2007 detection datasets. The results show that the model using CBAM modules has improved classification and detection performance compared with other models, including models using SE modules. Shukai Ding [12] and Jian Qu explored the effect of attentional mechanisms on autonomous driving in their study. However, they only studied CBAM and Resnet18 but did not do a more extensive study of other attentional mechanisms and CNNs.

In this paper, we discuss the impact of attention mechanisms on reducing the accident rate of intelligent vehicles. Using the SE and CBAM attention mechanisms, we construct four models: Efficientnet_b0-CBAM (ENetb0-CBAM), Efficientnet_b0-SE (ENetb0-SE), Resnet18-CBAM (RNet18-CBAM) and Resnet18-SE (RNet18-SE). Furthermore, we have developed the Safety Driving Framework (SDF), which is equipped with a convolutional neural network and is at the core of intelligent vehicles to ensure safe driving. It provides emergency stops in case of danger for intelligent vehicles. Data augmentation and 10-fold cross-validation are used to optimize the model, which improves its robustness and generalization, and reduces noise interference.

2. METHOD

This paper focuses on achieving safe driving of intelligent vehicles and reducing the number of accidents. First of all, we propose the Safety Driving Framework (SDF), which links hardware (camera, development board) and software (model and algorithm). Then, we employ convolutional neural networks based on attention mechanisms as predictive models for intelligent vehicles in order to improve environmental perception. Finally, data augmentation

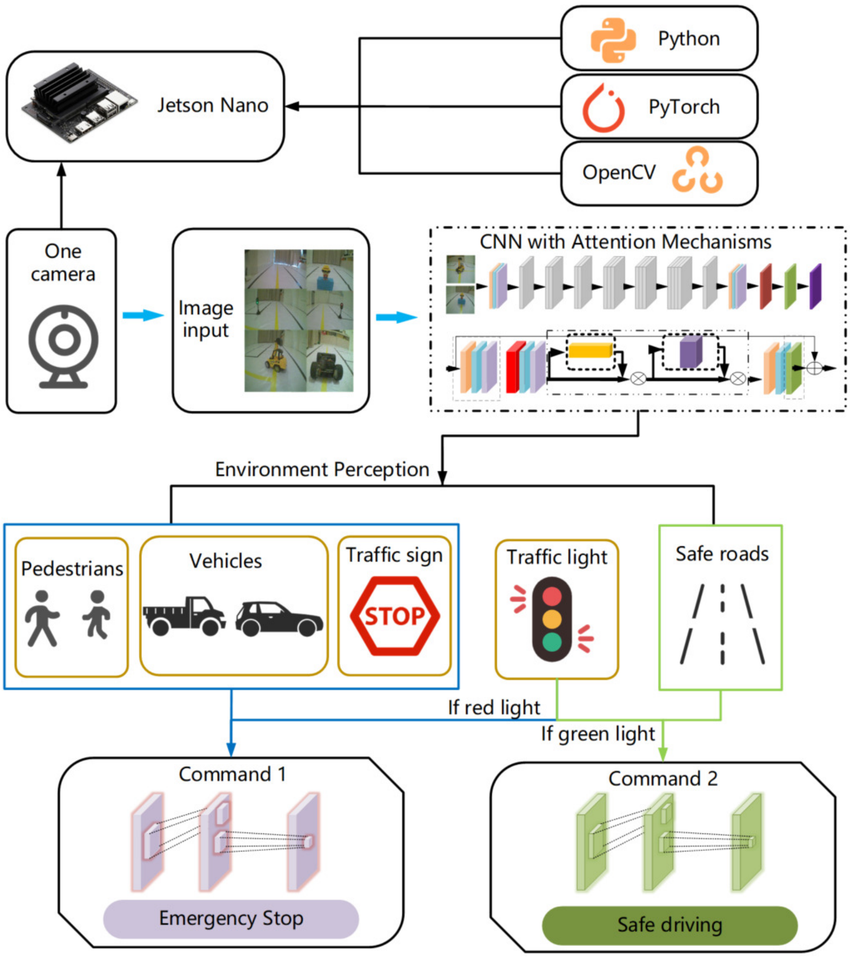

Fig.1: *Structure of The Safety Driving Framework.*

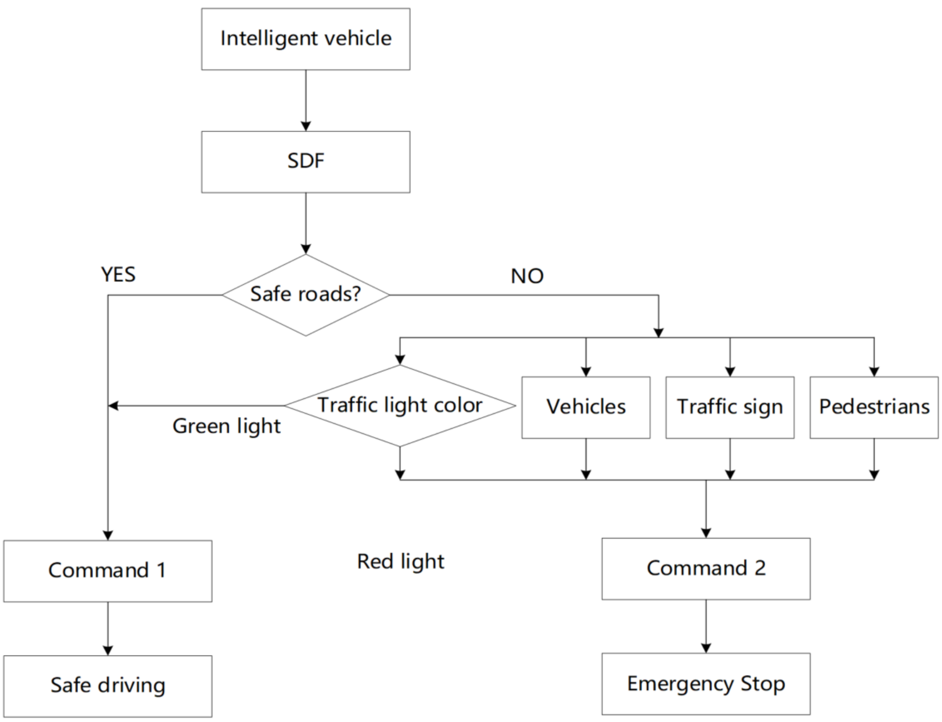

Fig.2: *The workflow of SDF.*

and 10-fold cross-validation are used to enhance the generalization capabilities and robustness of the model.

2.1 Safety Driving Framework

The Safety Driving Framework (SDF) is the basis for the safe operation of our autonomous vehicles. Using the Python programming language, SDF is built on a Jetson Nano development board. We do not choose the Raspberry Pi [13] as a development board because it is less capable of loading large neural networks than the Jetson Nano. The comparison of the Jetson Nano and Raspberry Pi is shown in Table 1. Additionally, we use OpenCV and PyTorch, which are commonly used computer vision libraries and application frameworks. The SDF framework is illustrated in Fig. 1, which shows how it connects hardware and software and performs emergency avoidance functions. In SDF, we use only one camera as the sensor, which is responsible for information input and environmental interaction. It transfers images to convolutional neural networks (predictive models) based on attention mechanisms. It is the core of our approach, which is described in detail in Section 2.2. SDF predicts road conditions based on the perceived images of the environment and provides appropriate instructions. The intelligent vehicles perform Command 1 (Emergency Stop) in situations where accidents are likely to occur, such as pedestrians and other vehicles near the intelligent vehicles. Intelligent vehicles will execute Command 2 (Continue driving) if the conditions are safe, such as pedestrians or vehicles passing safely. The SDF will make further judgments regarding dynamic sign traffic lights. If the light is red, the vehicle will come to an emergency stop. If the light is green, the vehicle will continue to drive. The flow diagram is shown in Fig. 2.

Table 1: *The comparison of the Jetson Nano and Raspberry Pi.*

Name	Jetson Nano	Raspberry Pi 4B
CPU	Quad-Core ARM Cortex-A57 64-bit @ 1.42Ghz	64-bit 1.5GHz quad-core
GPU	NVIDIA Maxwell w/128 CUDA cores @ 921 Mhz	500MHz VideoCore IV
Memory	4GB LPDDR4	1-4GB DDR4
USB	4 * USB3.0, USB2.0 Micro-B	2*USB 3.0, *USB 2.0
Video Encode	H.264/H.265(4Kp30)	H.265(4Kp60 decode)
Networking	Gigabit Ethernet/M.2 Key E(for Wifi support)	Gigabit Ethernet
Display	HDMI2.0 and eDP1.4	2 micro HDMI

2.2 CNN with Attention Mechanism

2.2.1 SE and CBAM

The selective attention in human vision is a unique brain signal processing mechanism that enables humans to screen out high-value information rapidly from a large amount of information using limited attention resources. It is a survival mechanism formed by human beings in long-term evolution. This property greatly enhances the efficiency and accuracy of human visual information processing. The attention mechanism in deep learning is derived from the mindset of humans. In cognitive science, due to bottlenecks in information processing, humans selectively focus on a portion of all information while ignoring the rest of the visible information. Different parts of the human retina have different degrees of information processing capability, or acuity, with only the central concave part of the retina having the strongest acuity. In order to make good use of the limited visual information processing resources, humans need to select a specific part of the visual area and then focus on it. For example, when people read, usually only a small number of words will be attended to and processed. In summary, the attention mechanism has two main aspects: deciding which part of the input needs to be attended to; and allocating the limited information processing resources to the important parts.

In the attention mechanism, the neural network learns autonomously and then outputs a set of weighted coefficients. At the same time, the neural network uses a "dynamic weighting" approach to emphasize the areas that need attention and suppress irrelevant background areas. This capability is required for intelligent vehicles. As road conditions become more complex and unpredictable, attention resources become scarcer. Focusing most attention on dangerous scenarios and ignoring irrelevant information is crucial to reducing intelligent vehicle accidents. Therefore, this paper discusses the application of attentional mechanisms in intelligent vehicles.

There are three types of attention mechanisms, namely spatial, channel, and mixed. The specific structure of the SE is shown in Fig. 3, which is a channel attention mechanism. The SE performs the

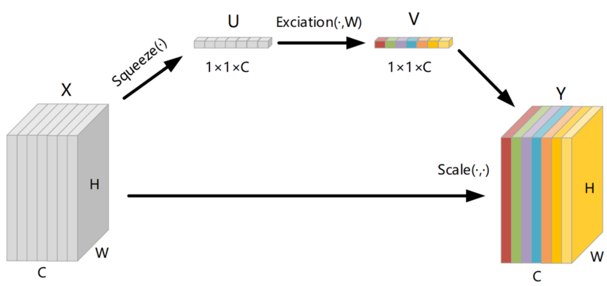

Fig.3: *Structure of the SE.*

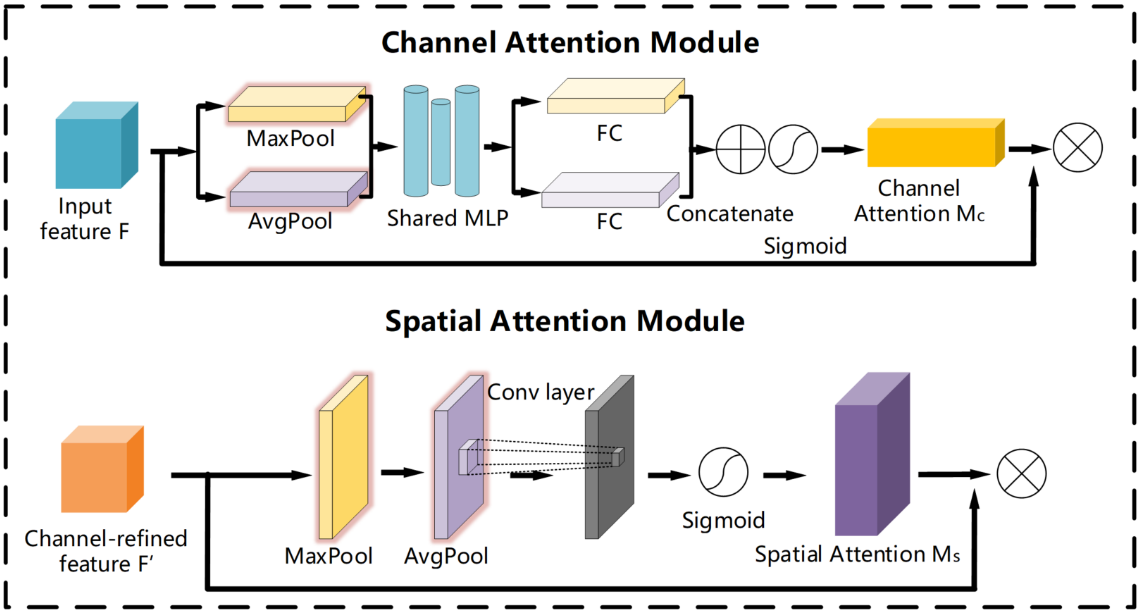

Fig.4: *Structure of The CBAM.*

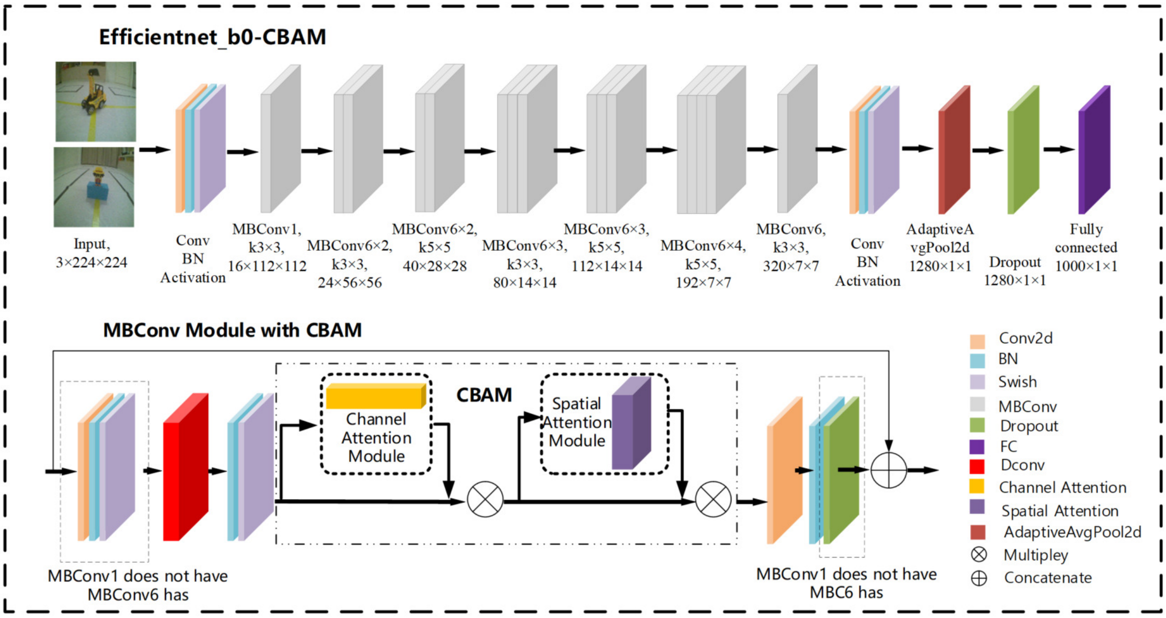

Fig.5: *Structure of The Efficientnet_b0-CBAM.*

Squeeze and Excitation modules to implement the "reduce dimension first, increase dimension later" strategy. It automatically obtains the importance of each feature channel using neural network learning. Finally, it assigns different weight factors to each channel to reinforce the important features and suppress the non-important ones. The working process of the SE involves two main operations: Squeeze and Excitation. First, the Squeeze compresses the spatial dimensions and pools each feature map globally. Second, the network outputs a feature map of 1×1×C size after the Squeeze, and the Excitation uses the weight W to learn the direct correlation of C channels.

CBAM is a mixed attention mechanism that includes both channel attention mechanisms and spatial attention mechanisms, as shown in Fig. 4. The channel attention mechanism of CBAM are similar to SE, but CBAM adds max pooling and shared MLP. For the channel attention mechanism, after the input feature F has been processed by global max pooling and global average pooling, its result is shared in the MLP. The MLP output feature is subjected to an elementwise summation operation and sigmoid activation to generate the final Channel Attention Mc. This Channel Attention Mc and input feature F are

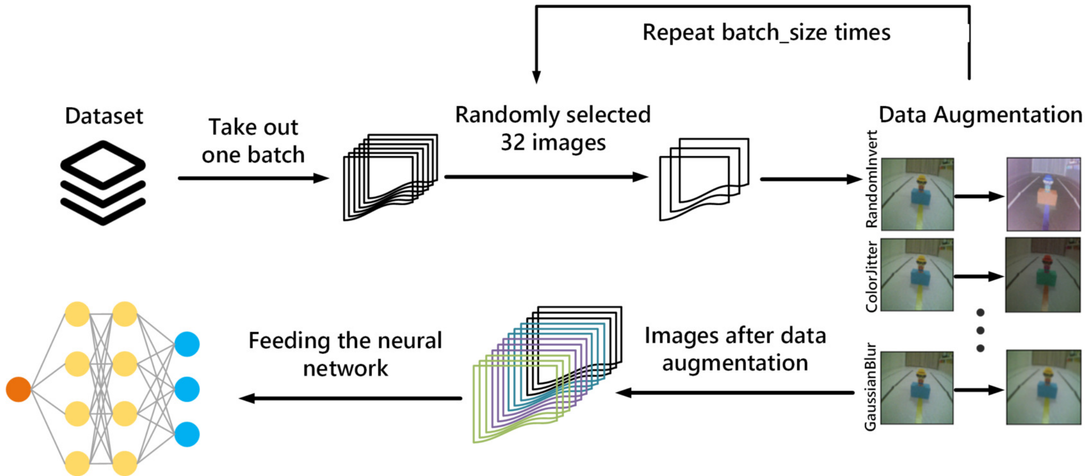

Fig.6: *The Working Process of Data Augmentation.*

subjected to an elementwise multiplication operation to generate Channel-refined feature F', which is also the input feature required by the Spatial attention mechanism. The Spatial attention mechanism first does global max pooling and global average pooling based on the channel, and then the two results are concatenated based on the Spatial attention mechanism. Then convolution and sigmoid operations are performed to generate spatial attention Ms. Finally, the feature is multiplied by the input feature of the module to obtain the final generated feature.

2.2.2 ENetb0-CBAM

Using Efficientnet_b0 [14] and Resnet18 [15] as the base framework and combining SE and CBAM, we construct four models, namely Efficientnet_b0-CBAM (ENetb0-CBAM), Efficientnet_b0-SE (ENetb0-SE), Resnet18-CBAM (RNet18-CBAM) and Resnet18-SE (RNet18-SE). Taking ENetb0-CBAM as an example, we show the structure of a convolutional neural network that incorporates attention mechanisms, as shown in Fig. 6.

We add CBAM to the MBConv module of the Efficientnet_b0. MBConv is the body module of the Efficientnet_b0, which is available in two versions, MBConv1 and MBConv6. MBConv6 has one more convolutional layer, one more batch normalization layer, one more activation layer, and one more dropout layer than MBConv1.

2.3 Data Augmentation

The development of deep learning cannot be separated from big data support. During training, the generalization ability of the model is directly influenced by the size and quality of the data. However, during the data collection, it is often not possible to fully cover the scene. In this case, data augmentation (DA) [16] is needed. Data augmentation is a way to augment the size and richness of data samples and

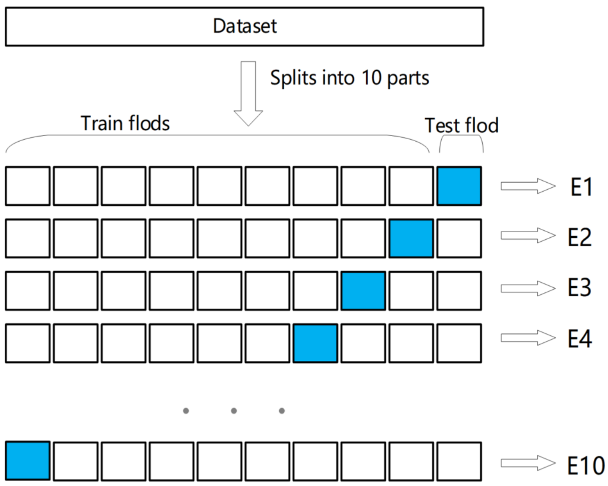

Fig.7: *Principles of 10-Fold Cross-Validation.*

Table 2: *Data Augmentation Methods.*

Data augmentation methods	Description
Resize	Uniform image size
ColorJitter	Randomly changing the color parameters of images
ToTensor	Convert images to tensor
Normalize	Normalizing data
RandomInvert	Inverted pixel values
RandomPerspectiv	Random perspective transformations
GaussianBlur	Image smoothing with Gaussian blur
RandomHorizontalFlip	Flip the image in the horizontal direction
RandAugment	Enhanced generalization performance
RandomCrop	Random cropping of a set size image

improve model robustness and generalization capabilities. For example, in a scene where intelligent vehicles collect traffic data, uncontrollable lighting conditions result in a relatively single dataset sample. Therefore, data augmentation on lighting variations can be added to the model training. Alternatively, when data collection and labeling are difficult, the data augmentation method can generate sufficient training data, lowering the cost of data collection and processing. Additionally, by performing data augmentation, the neural network can be prevented from learning irrelevant features, thereby significantly improving the overall performance of the model.

Data augmentation is mainly divided into two categories:

(1) Offline augmentation: Direct processing of existing data sets, the size of the data will become the product of the augmentation factor, and the size of the original data set, suitable for small data sets.

(2). Online augmentation: The dataset is expanded during the model training process. Specifically, after obtaining the data of one batch, the data under the match is enhanced. In this work, we choose the online augmentation approach. The working process of data augmentation is shown in Fig. 5.

There are many methods of data augmentation, including, but not limited to, flipping, scaling, clipping, and Gaussian noise. The data augmentation methods below have been selected in this paper, as shown in Table 2.

2.4 10-Fold Cross-Validation

The 10-fold cross-validation [17] has been widely adopted as a standard for evaluating model performance. This paper aims to achieve safe driving of intelligent vehicles and has high requirements for the stability and generalization ability of the model. As a result, this paper adopts the 10-fold cross-validation method to train the model and evaluate its stability and generalization ability. It can be used to evaluate model performance, and it can also be used to compensate for the insufficient dataset.

The structure of the 10-fold cross-validation is shown in Fig. 7. The dataset is divided into ten mutex subsets of similar size, with each of them preserving the consistency of data distribution to the greatest extent possible. Each time, we select one subset as the test set, and the union of the remaining subsets as the training set. It can obtain ten training/test sets, allowing for ten training sessions and ten tests. The accuracy (or error value) is generated for each test.

3. EXPERIMENT

3.1 Data Collection and Model Training

The simulated track we have constructed is shown in Fig. 8. It is a complicated track, and it can design different routes as needed at will. For example, the rectangular route of the outer ring of the track is circular, and the inner circle is a complex route containing multiple turns. Simulated tracks provide various driving routes that simulate complex real-world road conditions. In addition, in order to explore the impact of attention mechanisms on the safe driving of intelligent vehicles, we add risk factors (traffic accidents that are likely to cause traffic accidents) to the simulation tracks, referring to pedestrians, vehicles, traffic lights, and traffic signs.

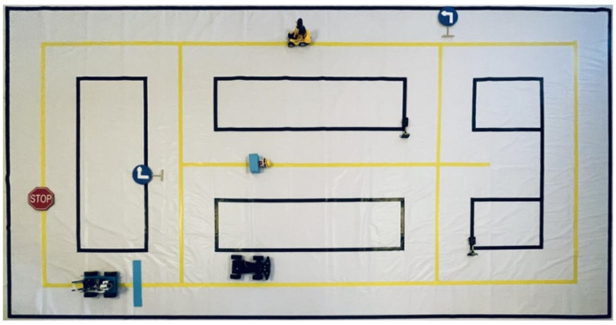

Fig.8: *The Simulation Track.*

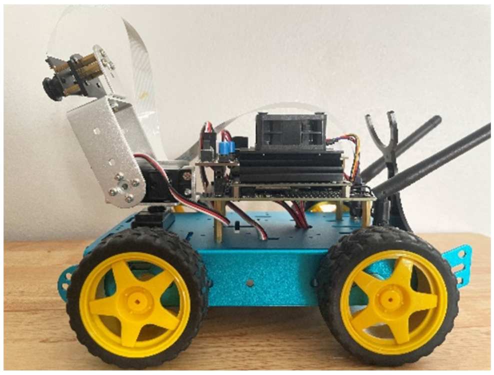

Fig.9: *Intelligent Driving Platform.*

As shown in Fig. 9, we built an intelligent driving platform. Its development board is Jetson Nano, and it is equipped with only a camera that serves as a sensor for environmental interaction. As the experimental subject, the intelligent driving platform collects images on the simulated track. Since Resnet18 and Efficientnet_b0 (the two baseline models we have chosen) require 224*224 input images due to their model structure, we collect images of 224*224. If we use high-resolution images, it will slow down the speed of image processing. In addition, we are using an embedded development board, the Jetson Nano, which has limited computing power, making it difficult to deploy models that use high resolution smoothly.

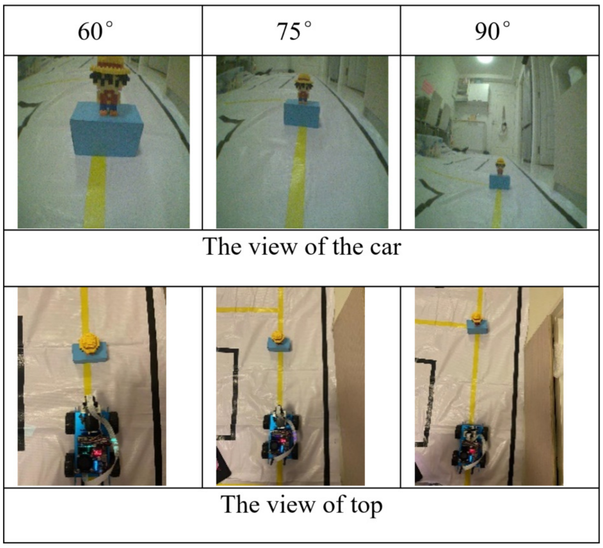

Fig.10: *Comparison of camera angles.*

In addition, the tilt angle of the camera is important for image collection. Different angles will collect different images, and the quality of the images will also determine how good the model is. The tilt angle of our camera is 75°. Our track and the experimental environment determined it. As shown in Fig. 10, we compared the view of the car at 60°, 75° and 90° and the distance when the car stopped against an obstacle. We found that the greater the angle of the camera, the wider the field of view of the car, but also the more noisy environment is collected by the camera. The smaller the angle of the camera, the narrower the field of view of the car and the less likely it is to recognise an obstacle and stop, so the closer the point at which the car stops is to the obstacle. When the camera angle is less than 60°, the car's field of view is insufficient to see the obstacle's full extent. So we did not compare angles below 60°. After testing, we found that the best camera angle was 75°. The image collection is divided into five categories: pedestrians, vehicles, traffic lights, traffic signs, and safety tracks. We collect 1000 images of safe tracks and 500 images of each of the remaining four categories.

We train the model in Colab and optimize some of the hyperparameters. The results show that choosing the Smooth L1 loss function and the Adam optimizer works best.

3.2 Image Augmentation

The images collected in the simulated environment have certain limitations. Light may be ignored because it is difficult to collect images throughout the day. Therefore, the collection of images is single in type and not very wide. In addition, we collected only 1500 images, which is a small dataset. Therefore, we used a variety of data augmentation methods from Table 2 to extend the dataset. The effect of image augmentation is shown in Fig. 11. One original image was generated with several different images after different data augmentation methods, which greatly enriched the dataset. We use this approach to improve the robustness and generalization of the model.

3.3 Test

We fused attention mechanisms to construct and train four CNNs: ENetb0-CBAM (Efficientnet_b0-CBAM), ENetb0-SE, RNet18-CBAM (Resnet18-CBAM) and RNet18-SE. In addition, our approach was compared with existing studies. We compare the four CNNs mentioned above with two existing models: Efficientnet_b0 [14] (ENetb0) and Resnet18 [15] (RNet18).

Abenezer Girma et al [18] used deep bidirectional long and short-term memory (LSTM) and an attention mechanism model based on the hybrid state system (HSS) framework in their study to solve the problem of estimating driver behavior near road intersections. However, they only worked on image datasets. They used the transverse swing rate and speed data of vehicles approaching the intersection at each time step was recorded and discrete into a series of symbols for training and testing the proposed model. They did not test the model in the real world. Instead of them, we have built a simulation track and an intelligent driving platform. We can dynamically test our proposed method and model in real-time in the real world. This dynamic testing approach fully reflects the effects of environmental noise and allows for a more realistic test of the performance of the models.

Pedestrians, vehicles, traffic lights, and traffic signs are randomly placed on the simulated track. In addition, we explore the impact of data augmentation (DA) on model stability. The experiments were conducted in two environments: a normal environment (the environment in which the model was trained) and a noisy environment. The reason for the noisy environment is that noise immunity is an assessment metric for evaluating the stability of the model. The higher the noise, the higher the demands on the model.

The noisy environment is where light noise and other invisible objects are added to the normal environment. For light noise, we use lights of different light intensities acting on the track to simulate an environment containing light noise. Noise-disturbing environments are environments where lighting and other unseen objects are added to a normal environment. The setting of the noise environment has important practical significance for intelligent vehicles. In the real world, intelligent vehicles will encounter numerous unexpected accidents, such as animals that appear suddenly in front of them, or a clear sky suddenly covered in dark clouds. Intelligent vehicles must possess superior environmental perception,

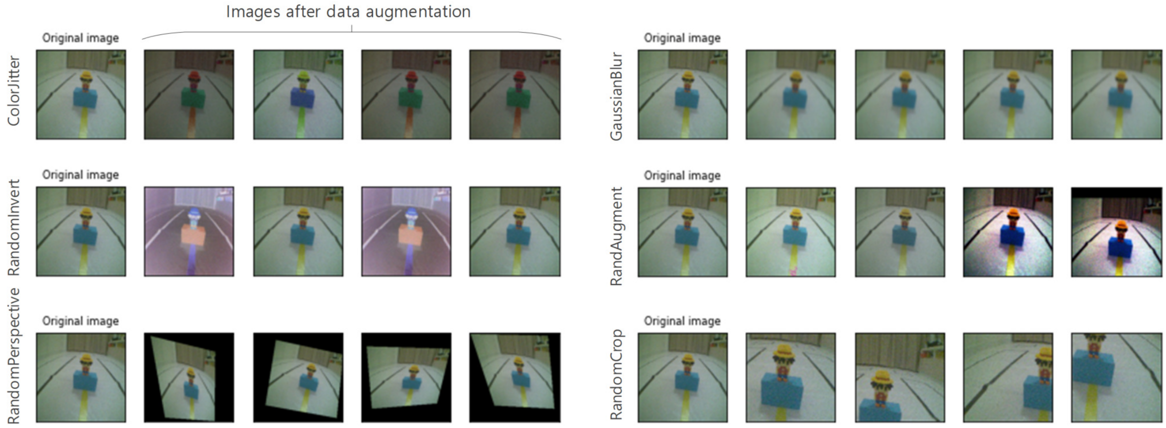

Fig.11: *Variation of Spacing with Suspension Preload.*

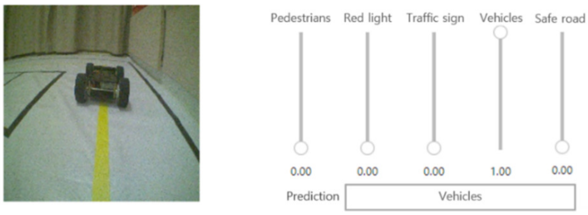

Fig.12: *An example of the recognition rate of an intelligent driving platform.*

image processing capabilities, and extreme robustness in order to deal with unexpected situations.

The average recognition rate and accident rate are used as evaluation indicators to assess the quality of the model. The recognition rate refers to the predicted probability of the obstacle by the intelligent driving platform. The testing process of an intelligent driving platform is variable, so its recognition rate is also a variable value. As shown in Fig. 12, it is an example of the recognition rate of a smart driving platform. This example has a recognition rate of 100% for vehicles, with no vehicles misidentified as pedestrians or other signs. It means that the intelligent driving platform predicts that the obstacle in the middle of the road ahead is the vehicle, which belongs to the dangerous area, and then the intelligent driving platform will stop driving until the obstacle is cleared. The intelligent driving platform drives on a simulation track, so recognising obstacles is a dynamic process. The speed of the intelligent driving platform and its distance from the obstacle greatly influence the recognition rate. The recognition rate will be low if the distance between the car and the obstacle is far and close; if the car is fast, the recognition rate will be low; if the car is too slow, the car will not start. After experiments, we found that the optimal speed of the car is 0.6. As the intelligent driving platform gradually approaches the obstacle at the optimal speed, its recognition rate gradually increases from low to high. When the intelligent driving platform stops, its recognition rate reaches its maximum. Each obstacle is randomly placed ten times on the simulated track for each experiment. The average recognition rate of obstacles in each experiment will be calculated and used as one of the final evaluation indicators. Another evaluation indicator is the accident rate of the intelligent driving platform. Every time the intelligent driving platform encounters an obstacle, it is considered a test. If the intelligent driving platform collides with an obstacle or goes off the track, it is considered a violation of the operation. The accident rate refers to the percentage of the number of violations of the intelligent driving platform in the simulated track to the total number of tests. There are five main types of accident, as shown in Table 3.

Table 3: *Types of accidents.*

Accident type	Reason
Hitting pedestrians	No pedestrian recognition
Crash	No vehicle recognition
Running a red light	Unrecognisable traffic light
Hitting the STOP sign	STOP sign not recognised, unable to stop
Driving off the track	No road tracking

4. RESULTS AND ANALYSIS

Our experiments discuss four questions:

1. The comparison between CNNs with and without attentional mechanisms. Results as shown in Fig. 13.

2. The comparison of models for noise resistance. Results as shown in Fig. 14.

3. The comparison between CNNs with and without data augmentation. Results as shown in Fig. 15.

4. Comparison of our approach with existing models. Results as shown in Fig. 16. The existing models are the original unprocessed Resnet18 and Efficientnet_b0.

The experimental results are shown in Table 4. The recognition rate is positively correlated with the accident rate. If the recognition rate is high, the accident rate will be low. However, there is no specific functional relationship between the two sides. The intelligent driving platform is moving, and the process of obstacle prediction is a dynamic one. The variable position and environment of the intelligent driving platform make obstacle prediction more difficult and increase the probability of an accident.

Table 4: *The Experimental Results*

Model	Average accuracy rate	Accident rate
ENetb0 (Existing)	13.3%	91%
ENetb0+DA	82.6%	19%
ENetb0+DA+noise	23.1%	88%
ENetb0-CBAM+DA	90.5%	13%
ENetb0-CBAM+DA+noise	65.6%	43%
ENetb0-SE+DA	95.6%	2%
ENetb0-SE+DA+noise	73.5%	19%
RNet18 (Existing)	23.7%	84%
RNet18+DA	79.4%	22%
RNet18+DA+noise	22.6%	90%
RNet18-CBAM+DA	92.8%	6%
RNet18-CBAM+DA+noise	69.8%	39%
RNet18-SE+DA	87.9%	15%
RNet18-SE+DA+noise	67.8%	42%

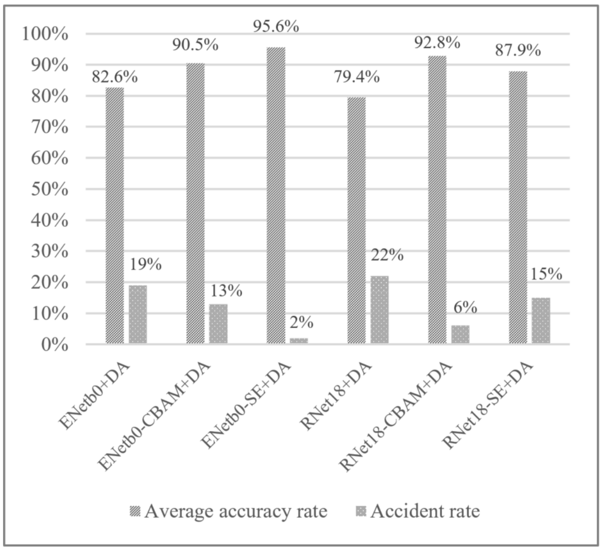

Fig.13: *The comparison between CNNs with and without attentional mechanisms.*

In Fig. 13, six CNNs are tested in a normal environment. According to our findings, CNNs with SE and CBAM attention mechanisms are superior to the existing network in terms of recognition rates and error rates. The most effective CNN is ENetb0-SE. The recognition rate is 95.6%, and the accident rate is only 2%, which is attributed to the attention mechanism. SE and CBAM attention mechanisms improve the ability of CNNs to extract image information to some extent. It is also evidenced by the results in Fig. 14. In Fig. 14, six models have been tested in a noisy environment. Because of the impact of environmental noise, the recognition rate of ENetb0 and RNet18 is greatly reduced, and the accident rate is greatly increased. As compared to them, the recognition and incident rates of the four models ENetb0-CBAM, ENetb0-SE, RNet18-CBAM, and RNet18-SE with attention mechanisms were not much weakened. It also demonstrates the high robustness of models with attention mechanisms to noisy environments. The combination of Fig. 13 and 14 shows that neither SE nor CBAM is superior or inferior to each other. In comparison to CBAM, SE has better test results in ENetb0 than RNet18. In contrast, CBAM performed best in RNet18. It implies that there is no optimal attention mechanism, only the most appropriate attention mechanism for each CNN.

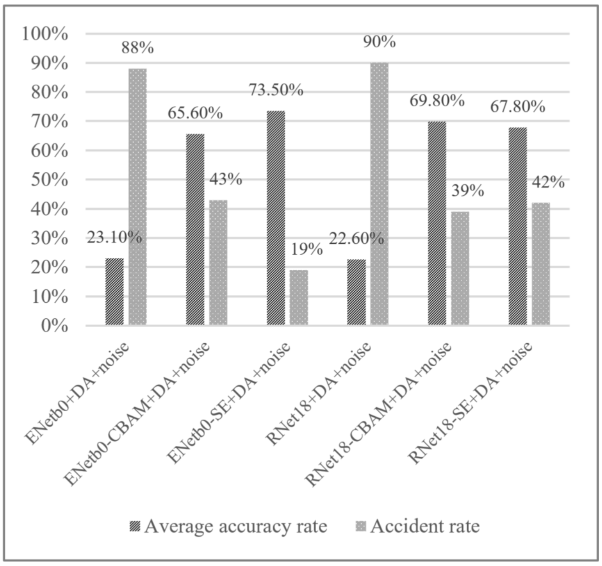

Fig.14: *Experimental results in the noise environment.*

The results of the tests on data augmentation are shown in Fig. 15. We compare ENetb0 and RNet18 tests with and without data augmentation. The test results show that the recognition rate of ENetb0 and RNet18 without data augmentation is very low, and the accident rate is high, far lower than the results with data augmentation. For tests without data augmentation, the reason for these results may be that it is too small a dataset and the images are of a single type. On the other hand, the use of data augmentation extends the dataset and improves the robustness of the model. For intelligent driving platforms, the safety of autonomous driving can be improved.

The above comparison illustrates the important role of attentional mechanisms and data augmentation in improving the robustness and generalisation

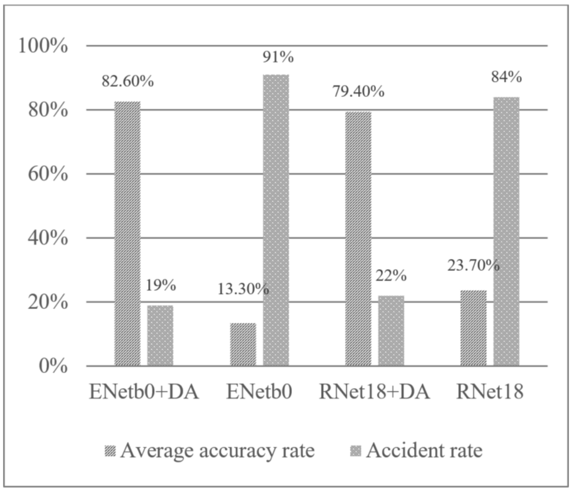

Fig.15: *Experimental results of data augmentation.*

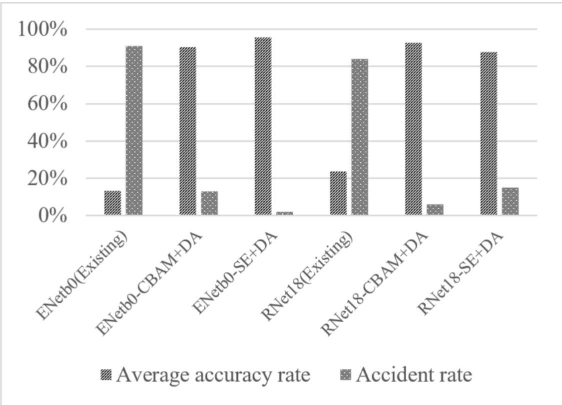

Fig.16: *Experimental results of our approach compared to existing models.*

of models. They can even achieve good results in noisy environments. The core part of our proposed Safety Driving Framework (SDF) is to train CNNs that have incorporated the attention mechanism with a data augmentation dataset. We used the SDF and obtained four models (ENetb0-CBAM+DA, ENetb0-SE+DA, RNet18-CBAM+DA and RNet18-SE+DA). We compared the four models with the existing ENetb0 (Efficientnet_b0) and RNet18 (Resnet18). The experimental results are shown in Fig. 16. All four models achieved good results compared to the two existing models. We attribute this to CNNs that incorporate attention mechanisms and data reinforcement. The attention mechanism can help CNN focus more efficiently on the important parts. Data reinforcement can help CNN to expand the dataset and improve robustness.

In addition, when collecting the data, we did not collect images of scenes that included a mixture of multiple signs. We only collected images of scenes with a single sign. For example, we collected images of only obstacles or only traffic lights. However, when we tested the intelligent driving platform by placing both green lights and pedestrians on the simulation track, it would stop rather than continue driving. We thought this situation might be because pedestrians are larger than traffic lights, so pedestrians are more easily recognised.

Shukai Ding conducted the experiment and drafted the manuscript. Jian Qu guided and advised the experiment and co-drafted the manuscript. Both authors each contributed equally to this work. Jian Qu is the corresponding author.

5. CONCLUSION

In this work, we present the Safety Driving Framework (SDF), which is the basis for intelligent driving platforms to enable safe driving. We also explore the positive impact of attention mechanisms on self-driving cars and build several models of ENetb0-SE, ENetb0-CBAM, RNet18-SE, and RNet18-CBAM to test. In addition, we optimize the models using the data augmentation method and 10-fold cross-validation in order to increase their robustness and generalizability. Experiments demonstrated that CNNs with attention mechanisms perform better than their existing models (ENetb0 and RNet18). Our method improves the environmental perception of the intelligent driving platform and reduces the probability of colliding with obstacles. ENetb0-SE, in particular, has an average recognition rate of 95.6% for obstacles, an accident rate of only 2%, and the intelligent driving platform has basically no violations. As shown through noise interference experiments, the attention mechanism also positively improves the generalization ability and anti-interference ability of CNN models. Our data augmentation experiments revealed that data augmentation techniques could be extremely beneficial for smaller datasets. It is capable of expanding the data set and enriching the variety of images through color transformation, cutting, and rotation operations, thereby resolving the issue of insufficient data collection. In the future, we intend to investigate additional methods for enhancing the stability and safety of autonomous driving beyond attention mechanisms in order to reduce further the accident rate caused by intelligent vehicles.

ACKNOWLEDGMENT

The first author received scholarship support from CPALL for conducting this research in PIM.

References

[1] D. Pomerleau and T. Jochem, "Rapidly adapting machine vision for automated vehicle steering," *IEEE expert*, vol. 11, no.2, pp. 19-27, 1996.

[2] B. Schoettle and M. Sivak, "A preliminary analysis of real-world crashes involving self-driving

vehicles," *University of Michigan Transportation Research Institute*, 2015.

[3] C. Nanding and C. Jing, "Design of intelligent tracing car based on infrared sensor," *IOP Conference Series: Materials Science and Engineering*, vol.563, no. 4, pp. 1-5, 2019.

[4] J. H. Rhee and J. J. S. Seo, "Low-cost curb detection and localization system using multiple ultrasonic sensors," *Sensors*, vol. 19, no.6, p. 1839, 2019.

[5] A. Ranjbar and S. Vig, "A Prototype on Design and Development of Self-Driving Vehicle," *IRJET Journal*, vol. 8, no.6, pp. 20-25, 2021.

[6] S. Yuenyong and Q. Jian, "Generating Synthetic Training Images for Deep Reinforcement Learning of a Mobile Robot," *Journal of Intelligent Informatics and Smart Technology*, vol. 2, pp. 16-20, 2017.

[7] R. R. O. Al-Nima, T. Han and T. Chen, "Road Tracking Using Deep Reinforcement Learning for Self-driving Car Applications," *International Conference on Computer Recognition Systems*, 2019.

[8] X. Zong, G. Xu, G. Yu, H. Su and C. Hu, "Obstacle Avoidance for Self-Driving Vehicle with Reinforcement Learning," *SAE International Journal of Passenger Cars - Electronic and Electrical Systems*, vol. 11, no. 1, 2018.

[9] R. Chopra and S. S. Ro, "End-to-End Reinforcement Learning for Self-driving Car," *Advanced Computing and Intelligent Engineering*, pp. 23-61, 2020.

[10] J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132-7141, 2018.

[11] S. Woo, J. Park, J. Y. Lee and I. S. Kweon, "CBAM: Convolutional block attention module," *European conference on computer vision (ECCV)*, pp. 3-19, 2018.

[12] S. Ding and J. Qu, "Smart car with road tracking and obstacle avoidance based on Resnet18-CBAM," *2022 7th International Conference on Business and Industrial Research (ICBIR2022)*, pp. 582-585, 2022.

[13] Y. Li and J. Qu, "Intelligent road tracking and real-time acceleration-deceleration for autonomous driving using modified convolutional neural networks," *CURRENT APPLIED SCIENCE AND TECHNOLOGY*, vol. 22, no.6, 2022.

[14] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *Proceedings of the 36th International Conference on Machine Learning*, pp. 6105-14, 2019.

[15] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

[16] C. Shorten, T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no.1, pp. 1-48, 2019.

[17] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statistics and Computing*, vol. 21, no.2, pp. 137-146, 2011.

[18] A. Girma, S. Amsalu, A. Workineh, M. Khan, and A. Homaifar, "Deep learning with attention mechanism for predicting driver intention at intersection," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1183-1188, 2020.

Shukai Ding is currently studying for the Master of Engineering Technology, Faculty of Engineering and Technology, Panyapiwat Institute of Management, Thailand. He received B.E from Nanjing Tech University Pujiang Institute, China, in 2020. His research interests are artificial intelligence, image processing, and autonomous driving.

Jian Qu is an Assistant professor at the Faculty of Engineering and Technology, Panyapiwat Institute of Management. He received Ph.D. with Outstanding Performance award from Japan Advanced Institute of Science and Technology, Japan, in 2013. He received B.B.A with Summa Cum Laude honors from Institute of International Studies of Ramkhamhaeng University, Thailand, in 2006, and M.S.I.T from Sirindhorn International Institute of Technology, Thammast University, Thailand, in 2010. He has been a house committee for Thai SuperAI since 2020. His research interests are natural language processing, intelligent algorithms, machine learning, machine translation, information retrieval, image processing, and automatic driving.