Movement Control with Vehicle-to-Vehicle Communication by using End-to-End Deep Learning for Autonomous Driving

Zelin Zhang and Jun Ohya
Department of Modern Mechanical Engineering, Waseda University, Tokyo, Japan

Keywords: Autonomous Driving, Deep Learning, End-to-End, Vehicle-to-Vehicle Communication.

Abstract: In recent years, autonomous driving through deep learning has gained more and more attention. This paper proposes a novel Vehicle-to-Vehicle (V2V) communication based autonomous vehicle driving system that takes advantage of both spatial and temporal information. The proposed system consists of a novel combination of CNN layers and LSTM layers for controlling steering angle and speed by taking advantage of the information from both the autonomous vehicle and cooperative vehicle. The CNN layers process the input sequential image frames, and the LSTM layers process historical data to predict the steering angle and speed of the autonomous vehicle. To confirm the validity of the proposed system, we conducted experiments for evaluating the MSE of the steering angle and vehicle speed using the Udacity dataset. Experimental results are summarized as follows. (1) "with a cooperative car" significantly works better than "without". (2) Among all the network, the Res-Net performs the best. (3) Utilizing the LSTM with Res-Net, which processes the historical motion data, performs better than "no LSTM". (4) As the number of inputted sequential frames, eight frames turn out to work best. (5) As the distance between the autonomous host and cooperative vehicle, ten to forty meters turn out to achieve the robust result on the autonomous driving movement control.

1 INTRODUCTION

During the past few years, autonomous self-driving cars have become more and more popular because of the development of sensor equipment and computer vision technology. Many technology companies and car manufactures have joined this industry, such as Google and General Motors. The purpose of autonomous driving is to let the vehicle perceive the surrounding environment and cruise with no human intervention. Therefore, the most important task for the autonomous driving system is to map the surrounding environment to the driving control.

Recently, deep convolutional networks have achieved great success in traditional computer vision tasks such as segmentation and object detection. It seems that the deep learning-based method is appropriate for autonomous driving because it can deal with more scenarios. Some state-of-the-art works divide the autonomous driving problem into several small tasks and fuse the results of each task to a final control decision. The rest of the state-of-the-art works provide an End-to-End solution that allows the autonomous system to learn the mapping from the raw image data to the steering control. Although all the recent state-of-the-art systems have achieved great successes, we still believe that those approaches lack the temporal information because of ignoring the relationship between sequential image frames. Therefore, we need a tool to capture and process temporal information. With the development of deep neural networks (DNN), Long Short-Term Memory (LSTM) has been designed to process sequential data in a time series. In recent years, it has gained more and more attention and has been popular in many fields such as human action recognition and natural language processing. It is a reasonable choice to apply the LSTM to deal with the temporal information in an autonomous driving problem.

One the other hand, the temporal information is not the only concern in autonomous driving. In our opinion, the interaction with the surrounding environment is also vital to autonomous driving because the autonomous vehicle is not isolated, especially in the urban scene. Considering the human manual driving, people would constantly check the movement of the vehicles that surround us. Based on the human driving habit, the autonomous vehicle should also consider the interaction with other surrounding vehicle movements. The interaction with two vehicles can be defined as Vehicle-to-Vehicle (V2V) communication. In our opinion, V2V

Zhang, Z. and Ohya, J.
Movement Control with Vehicle-to-Vehicle Communication by using End-to-End Deep Learning for Autonomous Driving.
DOI: 10.5220/0010235703770385
In *Proceedings of the 10th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2021)*, pages 377-385
ISBN: 978-989-758-486-2

communication can increase the level of certainty regarding a vehicle's surroundings and serves as an ability for autonomous driving.

In this paper, we propose an autonomous driving movement control system through V2V communication by applying End-to-End deep learning. Compared with the existing system, the method we propose has the following two main contributions, as follows.

1) We propose a novel method to predict the autonomous driving movement control through V2V communication. With the V2V communication, the autonomous vehicle can achieve a set of data including the motion state and the driver's first view images from the surrounding vehicle. The additional information through V2V communication can improve system performance.
2) We propose a novel network architecture for the autonomous driving movement control by combining CNN with LSTM. Using the current view and motion state from the past, the system can perform better, because it can capture more temporal information through the relationship between the sequence frames.

The rest of this paper is organized as follows. Section 2 gives an overview of the state-of-the-art related work. The solution we propose is explained in Section 3. In Section 4, the experimental details for our system are given. Section 5 details the evaluation of the experiments and the corresponding analysis. The conclusion of our work is given in Section 6.

2 RELATED WORK

In the past few decades, great successes have been achieved in autonomous driving. With the development of basic knowledge in deep learning, research groups and companies have started to attempt a deep learning-based method to solve the autonomous driving problem. We analyze the great related work in the past few years and simply categorize them into two groups: rule-based methods and perception-based methods.

2.1 Rule-based Methods

Rule-based methods divide the autonomous driving problem into several small tasks, such as interaction with cars, lane following, pedestrian detection, and traffic light recognition. Rule-based methods tend to solve all the small tasks independently and fuse all the results obtained by each task to achieve the final movement control. The key point of the rule-based system is car detection and scene understanding. Some traditional classic methods use bounding boxes to detect cars. However, the size of the bounding box could influence the final results, and the margin of the bounding box seems a waste of spatial information. To solve these problems, semantic segmentation has become more and more popular. Semantic segmentation estimates the probability of every pixel and finally makes the whole scene understandable. Meanwhile, the lane detection also plays an important role in rule-based systems. It is a simple way to keep the vehicle in the lane and waiting for the next movement control.

Even though rule-based methods have achieved great success, the powerful sensor may intensively increase the budget. A sensor with the vision of the 360-degree field is five times as expensive as the one with the vision of the 120-degree field. It is very hard to build an autonomous vehicle in the budget unless the price of the sensor lowers. Besides, each result gained from the sensors could influence the final controls significantly. It may cause a significant problem even if one of the sensors is unfunctional. Although the rule-based system sounds reasonable, it is still a driver assistant system other than an autonomous driving system.

2.2 Perception-based Methods

Instead of dividing the large task into several small ones, the perception-based method simply learns the mapping from the images to the steering controls. ALVINN et al. proposed an idea first: they used a neural network to make the first attempt. Although the network is very simple and shallow, it can still be used in several certain situations. Based on that idea, LeCun et al. replaced the shallow network by six convolutional layers and called it an end-to-end system. The end-to-end system can map the raw pixels to the steering controls, and achieve great robustness to the various environment. With the development of convolutional neural networks in recent years, some traditional hardware companies have also joined this field. Recently, Nvidia collected the training datasets with three cameras from the left, right, and center, and trained a deep convolutional neural network to map the pixels to the steering controls. After the training procedure, one single image from the central camera can simply decide the steering control. However, all the methods mentioned above aim at processing data properly to achieve better performance.

In recent years, some companies and research groups such as DOCOMO and NISAN started to realize the importance of collecting data from other devices. However, none of them focus to use the data from other vehicles to improve the movement control of the autonomous vehicle. Olson et al. proposed a system that can collect the distance data from the other vehicles by scanning the QR code attached to the vehicle through the camera. Based on this idea, Eckelmann et al. improved the system by replacing the camera with a LiDAR sensor so that they can achieve the vehicle's position in the surrounding environment. Moreover, Khabbaz et al. proposed a system with an overlooking view. The system can collect from the camera attached to a drone in the sky. However, all the systems mentioned above aim at transportation planning macroscopically.

Our proposed work is based on the perception method. However, instead of directly using the images captured by the camera and simply mapping to the steering control, we take the V2V communication into account, which can provide more additional information to help the autonomous vehicle to make the movement decision. Besides, we consider the historical motion state as an important factor for making movement control decisions. Therefore, we design a network architecture using a combination of CNN and LSTM networks to fully use both spatial and the temporal information through the V2V communication.

3 METHODOLOGY

Our proposed work is based on the perception method. However, instead of directly using the images captured by the camera and simply mapping to the steering control, we take the V2V communication into account, which can provide more additional information. Besides, we consider the historical motion state as an important factor for making movement control decisions. Therefore, we design a network architecture using a combination of CNN and LSTM networks to fully use both spatial and temporal information through V2V communication.

3.1 System Overview

In this paper, we propose a novel method to learn the mapping from the images to the movement controls by taking advantage of both spatial and temporal information through V2V communication. We define the cooperative vehicle as to the one that can interact with the autonomous vehicle through V2V communication. The cooperative vehicle is a vehicle under human manual driving in front of the autonomous vehicle. We consider that a camera is attached to the center of the cooperative vehicle that can collect image data with the driver's view. A cooperative vehicle can collect data and help the autonomous vehicle make the movement control by sending the data back to the autonomous vehicle. Due to the low latency and high reliability of the 5G system, the data from the cooperative vehicle can not only enhance the safety of transportation but also manage the traffic intelligently. Through the V2V communication, additional data captured by cooperative vehicle could constantly send back to the autonomous vehicle. On the other hand, the temporal information is provided by the time-series images from both autonomous vehicles and cooperative vehicle. The system combines two sets of data to decide the driving motion control for the autonomous vehicle.

We believe that the perception-based method should imitate human manual driving. The only three concerns that arise when people drive manually are the steering angle and the speed of the current car, which can be interpreted as acceleration or brake, as well as the surrounding environment. In case of V2V communications, the surrounding environment is the state of the cooperative vehicle. Therefore, the vehicle movement control can be described as follows:

$$MC_V = \{\text{steering angle}, \text{speed}\} \quad (1)$$

where the MC denotes movement control. V represents vehicle. Based on the description (2), we can formulate our model as follow:

$$MC_{auto}^{t+\Delta t} = \{MC_{auto}^{t}, MC_{coo}^{t}\} \quad (2)$$

where the auto stands for the autonomous vehicle, and Coor stands for the cooperative vehicle. The movement control of the autonomous vehicle at time $t + \Delta t$ is determined by two sets of data that are both the movement control of the autonomous vehicle at time t, and the movement control of the autonomous vehicle at time t. Each set of the movement control can be described by two parameters that are the steering angle and the speed of the current vehicle.

Figure 1 shows an overview of our proposed method. Our approach tries to minimize the difference between the human manual driving and autonomous driving, and make the system more stable and accurate. Our approach takes the two sequential image frames and historical motion state as

the input and predicts the future motion state which is the steering angle and the speed for the autonomous vehicle.

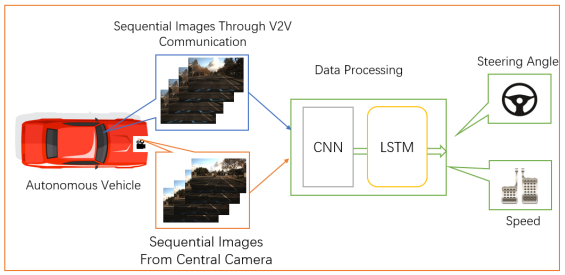

Figure 1: Overview of the V2V communication Autonomous movement planning system. The red one is the autonomous host vehicle with a front central camera. The sequential images in the blue rectangle is collected through V2V communication. The data processing part combines two sets of data to control the autonomous vehicle movement that are the steering angle and the speed.

3.2 Vehicle-to-Vehicle Communication

In recent years, the Vehicle-to-Vehicle (V2V) communication has gained more and more attention. With the development of the Fifth Generation (5G) mobile communication system, we believe that the V2V communication is a key technology to achieve high-level autonomous driving. In order to make the autonomous driving system work under 5G in the near future, we try to establish an autonomous driving system through V2V communication. Most of the existing autonomous driving systems rely on the sensor equipment to monitor the surrounding environment. However, the powerful sensor equipment would cost too much, and exceed the budget eventually. More importantly, the autonomous vehicle would never be isolated in the environment especially in the urban scene. The interaction with other vehicles should be also taken into consideration.

In Figure 2, the host car is an autonomous vehicle with a camera in the center. The central camera captures the sequential images while driving.

Meantime, the cooperative vehicle is ahead of the host autonomous vehicle, and also keeps capturing the images through its own center camera. Under the V2V communication, we assume that the cooperative vehicle communicates with the host autonomous vehicle by sending the driving state, which is speed, steering angle, and the sequential images from the central camera. The autonomous vehicle receives the data from the cooperative vehicle and combine the received data with the autonomous vehicle's data, and start to process the two sets of the data through a deep neural network (DNN). In this paper, we assume the autonomous vehicle and cooperative vehicle are away from each other. The movement control using the model indicated in (3) through V2V communication.

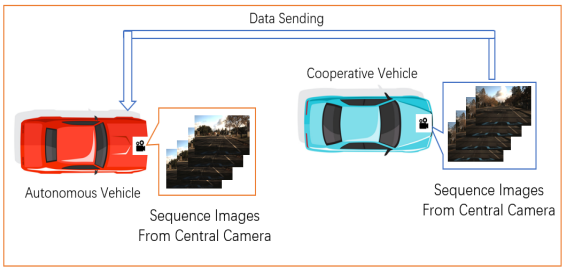

Figure 2: Overview of the V2V communication. The red one is the autonomous host vehicle with a front central camera. The blue one is the cooperative vehicle that also collects the sequence images from the central camera. The V2V communication allows the cooperative vehicle to send the data back to the autonomous vehicle to help the autonomous driving system make the movement control.

3.3 CNN-LSTM Architecture

The goal of the system is to learn the mapping from two sets of image data to the autonomous vehicle movement control. Both spatial and temporal information would be used through the End-to-End structure. The spatial information is the features extracted by CNN from the input frames that are obtained by the autonomous vehicle and the cooperative vehicle. The temporal information comes from the relationship between the sequential image frames. In this paper, we propose a novel network architecture that combines the CNN and LSTM networks, as shown in Figure 3 We use the CNN network to capture the feature from the input frames, and use the LSTM network to fuse these features. The recursive structure of LSTM can keep past information for regression. The whole network can take advantage of both spatial and temporal information. We detail the CNN and LSTM networks as follows.

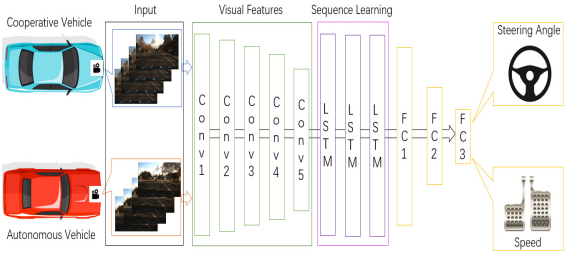

Figure 3: The data processing flow of the system proposed in this paper.

CNN: CNN has been approved as the most efficient and suitable way to solve traditional problems in the computer vision field such as segmentation and scene understanding. The key to train a network is the input data, annotation, and network architecture. Our goal is to find the best network architecture for feature extraction. We want to keep as much as spatial information through the convolution processing. To the best of our knowledge, ResNet has been demonstrated in many fields. In this work, we apply the transfer learning. The whole CNN network is pre-trained on the ImageNet dataset.

LSTM: To the best of our knowledge, the exiting work only focuses on one input frame, which may result in lack of temporal information of the whole system. We believe that human manual driving should also contain the drive motion history. Therefore, we think the motion history state is also important to train the network and make it stable. Based on this idea, we apply the LSTM network to capture the temporal information. LSTM is a well-improved recurrent neural network by introducing memory gates. LSTM avoids gradient vanishing and is capable of learning long-term dependencies. In recent years, LSTM has been demonstrated as a good way in the prediction field. LSTM is widely used also in the classification field.

As shown in Fig. 3, we place the LSTM network after CNN. This placement allows the LSTM network to fuse all the spatial features and the motion history state into the current state. The final movement control for autonomous vehicle is decided by its own motion history state and the cooperative vehicle's. Table 1 shows the network architecture proposed in this paper, where Conv represents the convolutional layer.

Table 1: Proposed Network Architecture.

Layer	Type	Size	Stride	Activation
1	Conv1	5*5*24	5,4	ReLu
2	Conv2	5*5*32	3,2	ReLu
3	Conv3	5*5*48	5,4	ReLu
4	Conv4	5*5*64	1,1	ReLu
5	Conv5	5*5*128	1,2	ReLu
6	LSTM	64	-	Tanh
7	LSTM	64	-	Tanh
8	LSTM	64	-	Tanh
9	FC	100	-	ReLu
10	FC	50	-	ReLu
11	FC	10	-	ReLu
12	FC	2	-	Linear

4 EXPERIMENTS

In this section, we explain the data setup and details of our experiments.

4.1 Data Setup

In this paper, we apply the Udacity dataset for experiments. The Udacity dataset has 223GB of image frames and logs data from 70 minutes during driving with the annotation of latitude, longitude, gear, brake, throttle, steering angles, and speed. The data were collected on two separate days, where one day was sunny, and the other was overcast. The FPS of each video data is 20Hz with a resolution of 640*480 pixels.

In this paper, the maximul speed of the autonomous vehicle and the cooperative vehicle is 100km/h. The steering angle of the autonomous vehicle and the cooperative vehicle ranges $[-\pi/2, \pi/2]$. The range in $[-\pi/2, 0]$ is defined as to go left, and $[0, \pi/2]$ is defined as go right. We set a fault tolerance of autonomous driving. The fault tolerance for the steering angle is 4 degrees, and the fault tolerance for acceleration is 4km/h.

4.2 Network Setup

After the data setup, we obtain the images data with the corresponding annotation. For training the autonomous driving model, we use a subset of the data acquired from the Udacity dataset. Based on the 80/20 split policy, we split the data into 80-20 training-testing for our experiments. In the training procedure, we set the learning rate as 0.0001 with stochastic gradient descent (SGD). The momentum is set to 0.99, with a batch size 16. We also apply the dropout layer and batch normalization layer to let the network avoid the overfitting problem. The input to the CNN is images with the annotation. The output from the CNN and the motion history state are the input to the LSTM. In order to avoid the gradient explosion of the LSTM, we set the gradient clip to 10. The hidden unit number of the LSTM is set to 64. The whole network outputs the next autonomous driving movement control, which is the steering angle and speed in continuous values.

5 RESULTS AND DISCUSSION

In this paper, we conducted the following five experiments to evaluate our work.

(1) We conduct experiments for evaluating the validity of the V2V communication by comparing the two cases: with and without a cooperative car.
(2) We conduct experiments for finding the best combination of the CNN-LSTM architecture by applying three CNN architectures which is Nvidia, Inveption-v3, and the ResNet-152.
(3) We conduct experiments for evaluating the proposed CNN-LSTM system by comparing the network with and without LSTM.
(4) We change the values for the parameter x, which is for controlling the number of input frames so that the x value that achieves the best performance is found.
(5) We change the distance between the autonomous and cooperative vehicles so that we can evaluate how the distance influences the movement control of the autonomous vehicle.

All the experiments use the mean absolute error for the steering angle and speed (Eq. (3)) for evaluation:

$$MSE = \frac{1}{n}\sum_{k=1}^{n} |p_i - g_i|. \quad (3)$$

where p_i stands for the prediction, and the g_i stands for the ground truth. The angle is in degree, and the speed is in km/h. In the experiments, we try to minimize the difference between the output and the ground truth. However, we do not need the output to be exactly the same as the ground truth. In other words, the system can still perform well if there is only a small error between the prediction and the ground truth. Therefore, we set this small error as a threshold ts to evaluate driving motion DM. Predictions whose MSE are below the ts should be considered as correct driving motions; otherwise, wrong driving motions. We set parameter CDM to represent the correct driving motion, and WDM to represent the wrong driving motion. Then parameter CDM and WDM can be calculated in Eq.(4) as follows:

$$\begin{cases} |p_i - g_i| \geq ts, \ DM = CMD \\ |p_i - g_i| < ts, \ DM = WMD \end{cases} \quad (4)$$

Then, we define a system performance score as $PESC$ in Eq.(5) to evaluate the system:

$$PESC = CDM/(CDM + WDM). \quad (5)$$

As mentioned earlier, we set the threshold of the steering angle as 4 degrees, and the 4km/h for the speed. If the driving motion control is decided every half second, the physical error in the vehicle moving direction is 0.399m and the vertical direction is 0.055m. It would cause no physical damage in the real world.

5.1 Evaluation of V2V Communication

Here, we evaluate the influence of V2V communication. In our opinion, the autonomous vehicle is never isolated in the environment, and the interaction with other vehicles would constantly happen. We believe that the autonomous driving system could perform better if it can get the data from the cooperative vehicle. In order to confirm the validity of our proposed system, we test the system with and without the V2V communication and keep the rest of the setting the same in the comparison experiment. The results are shown in Table 2, which lists the MSE and system performance score $PESC$ for the steering angle and speed. The MSE of angle is in degree, and the speed is in km/h.

The results indicate that the system performs better with V2V communication. The MSE and system performance score $PESC$ are both improved by taking advantage of the data from the cooperative car. This result implies that the V2V communication can make the system more stable and accurate, which is similar to humans' manual driving: i.e., humans keep attention to the frontal car.

Table 2: Performance of with and without V2V communications.

Method	Item	MSE	PESC
Without V2V	Angle	10.73	60.8%
	Speed	11.35	56.3%
With V2V	Angle	**3.69**	**84.2%**
	Speed	**3.16**	**78.9%**

Table 3: Performance of different CNN-LSTM architectures.

CNN	Item	MSE	PESC
Nvidia	Angle	7.28	67.3%
	Speed	10.89	61.1%
Inveption-v3	Angle	7.11	71.7%
	Speed	4.68	73.5%
ResNet-152	Angle	**3.69**	**84.2%**
	Speed	**3.16**	**78.9%**

5.2 Evaluation with Different CNN Architectures

In our proposed CNN-LSTM architecture, experiments compare different CNN's. This experiment implements three different CNN-LSTM combination architectures: NVIDIA, Inception-v3,

and ResNet-152 CNN architecture. Table 3 shows the results, where MSE and system performance score $PESC$ for the angle and speed for the three CNN's are listed.

Obviously, ResNet-LSTM architecture performs the best among all the three architectures.

5.3 Evaluation of LSTM Architecture

This experiment evaluates how the LSTM architecture, which processes the motion history state, influences the autonomous driving. The experiments compare the following two networks: ResNet-LSTM, and the ResNet only (without LSTM). The results are shown in Table 4, in which MSE and system performance score $PESC$ for the angle and speed for ResNet only and ResNet-LSTM are listed.

Obviously, the CNN-LSTM architecture improves both MSE and system performance score $PeSc$, which demonstrates that the motion history state information does influence the future motion state. Although the MSE for speed is not significantly improved, the CNN-LSTM structure does improve the system performance score $PESC$, which means that the future motion state prediction with the historical information makes the whole system more accurate and stable.

Table 4: Performance of architecture with or without LSTM.

Architecture	Item	MSE	PESC
ResNet only	Angle	15.63	63.8%
	Speed	9.89	74.1.%
ResNet-LSTM	Angle	**3.69**	**84.2%**
	Speed	**9.27**	**78.9%**

5.4 Evaluation of Sequential Image Frame Inputs

In this experiment, we formulate the V2V communication as follows. The autonomous vehicle and cooperative vehicle are away from each other. At time t, we assume the cooperative vehicle starts to acquire sequential image and send them to the autonomous vehicle. During this procedure, the autonomous vehicle can collect x frames in total. Combining the data from the autonomous vehicle, it can obtain 2x frames as the input to the DNN.

The primary task is finding the optimal x for autonomous driving through V2V communication. In this experiment, we change the value of x, keeping the other parameters constant. Since the FPS of the Udacity dataset is 20, it means that the movement control decision is made in every x/20 second. Therefore,we set the x as {2,4,6,8,10,12,14,16,18,20}, and find the x that achieves the best performance among the above-mentioned x values. The results are shown in Table 5, in which MSE for the angle and speed for the different x values are listed.

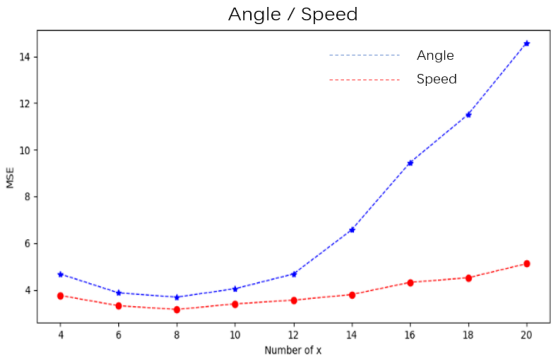

Figure 4: Performance with different parameter x.

Clearly, the number of input frames influence the system performance significantly. In Fig. 3, we plot the results shown in Table 5. According to Table 5 and Fig. 4, it turns out that x=8 gives the best performance for both angle and speed which means that the movement decision made in every 8/20 = 0.4s can achieve the best performanve overall.

Table 5: The Performance with different parameter x.

X	4	6	8	10
Angle	4.68	3.87	3.69	4.05
Speed	3.76	3.32	3.16	3.40
X	12	14	16	18
Angle	4.68	6.57	9.45	11.52
Speed	3.56	3.80	4.32	4.52

5.5 Evaluation of Distance

As can be seen in the model described in Fig. 1 in Section 3.2, there is a distance between the autonomous vehicle and the cooperative vehicle in the vehicle moving direction. Although V2V communication could improve the performance of the autonomous driving system, not all cooperative vehicles could influence the autonomous vehicle movement control. Same as human manual driving, the interaction with the nearby vehicle has a larger influence on the movement control. To verify the distance influence, we set a distance gap parameter Δd. In this experiment, we find out how the distance between the autonomous vehicle and the cooperative vehicle could influence the autonomous vehicle movement control.

Similar to Section 5.4, we change the parameter Δd, keeping the other parameters constant. In this experiment, as described in Section 5.4, we set x as 8, and try to find out how the Δt would influence the results. This experiment sets Δd value every 10 between 0 and 90. We use the MSE of the angle and speed to evaluate the influence of Δd. Table 6 shows the results and Fig. 5 plots all the results.

Table 6: Performance of different Δd values.

Δd	10	20	30	40
Angle	3.64	3.53	3.69	4.22
Speed	3.94	3.83	3.76	3.90
Δd	50	60	70	80
Angle	5.11	6.97	8.35	10.19
Speed	4.12	5.86	6.39	8.13

In Fig. 5 we can see that the best performance is achieved at Δd =20. Moreover, Fig. 4 also shows that MSE significantly increases when Δd exceeds 40. Therefore, we assume that Δd between 10 and 40 has the strongest influence on the autonomous vehicle. Considering the current speed is nearly 20m/s, the V2V communication works best when the distance from the autonomous host vehicle to the cooperative vehicle is between 10m to 40m.

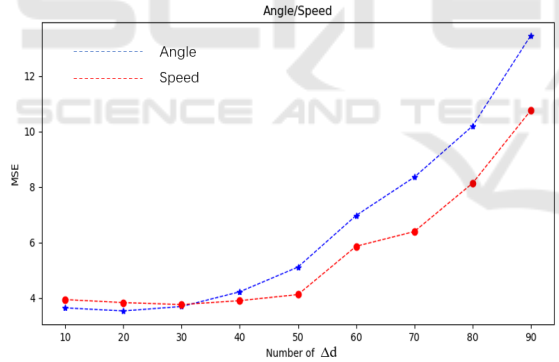

Figure 5: Performance with different distance values.

6 CONCLUSIONS

In this paper, we have proposed a novel V2V communication based autonomous vehicle driving system that takes advantage of both spatial and temporal information. The proposed system consists of a novel combination of the CNN-LSTM networks for controlling steering and speed based on the spatial and temporal information obtained by the cameras attached to the autonomous host vehicle and cooperative vehicle.

To confirm the validity of the proposed system, we conducted experiments for evaluating the MSE and accuracy of the steering angle and car speed using the Udacity dataset. Experimental results are summarized as follows.

(1) By comparing the two cases: with and without a cooperative vehicle, "with a cooperative vehicle" outperforms "without", which clarifies the validity of the V2V communication.

(2) By comparing the networks with and without the LSTM in the proposed system, it turns out that "with" works better than "without", which means that historical motion information, which is processed by the LSTM, is useful.

(3) To find the best CNN architecture in the proposed system, three different CNN's are compared. Res-Net turns out to be the best among the three.

(4) As a result of exploring the number of input sequential frames, eight frames turn out to achieve the best performance.

(5) As a result of exploring the distance between the host and cooperative cars, ten to forty meters turn out to achieve a robust result on the autonomous driving movement control.

In the future, we may expand the V2V communication working scenarios to combine it with signals like traffic. Meanwhile, we would like to improve the network learning ability and accuracy and try to make the system work for more application scenarios.

REFERENCES

Gibbs, S. (2017). Google sibling waymo launches fully autonomous ride-hailing service. *The Guardian*, *7*.

Wehner, M., Truby, R. L., Fitzgerald, D. J., Mosadegh, B., Whitesides, G. M., Lewis, J. A., & Wood, R. J. (2016). An integrated design and fabrication strategy for entirely soft, autonomous robots. *Nature*, *536*(7617), 451-455.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

Toderici, G., Vincent, D., Johnston, N., Jin Hwang, S., Minnen, D., Shor, J., & Covell, M. (2017). Full resolution image compression with recurrent neural networks. In *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition (pp. 5306-5314).

Liu, J., Shahroudy, A., Xu, D., & Wang, G. (2016, October). Spatio-temporal lstm with trust gates for 3d human action recognition. In *European conference on computer vision* (pp. 816-833). Springer, Cham.

Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016, November). Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 606-615).

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2). Cambridge: MIT press.

Wang, X., Jiang, R., Li, L., Lin, Y., Zheng, X., & Wang, F. Y. (2017). Capturing car-following behaviors by deep learning. *IEEE Transactions on Intelligent Transportation Systems*, *19*(3), 910-920.

Zhang, L., Lin, L., Liang, X., & He, K. (2016, October). Is faster R-CNN doing well for pedestrian detection?. In *European conference on computer vision* (pp. 443-457). Springer, Cham.

Li, X., Ma, H., Wang, X., & Zhang, X. (2017). Traffic light recognition for complex scene with fusion detections. *IEEE Transactions on Intelligent Transportation Systems*, *19*(1), 199-208.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3213-3223).

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., & Fei-Fei, L. (2017). Fine-grained car detection for visual census estimation. *arXiv preprint arXiv:1709.02480*.

Pomerleau, D. A. (1989). Alvinn: An autonomous land vehicle in a neural network. In *Advances in neural information processing systems* (pp. 305-313).

LeCun, Y., Bengio, Y., & Hinton, G. Deep learning." nature 521.7553 (2015): 436. *DOI: https://doi.org/10.1038/nature14539*

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., ... & Zhang, X. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345-1359.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.

Carvalho, A., Lefévre, S., Schildbach, G., Kong, J., & Borrelli, F. (2015). Automated driving: The role of forecasts and uncertainty—A control perspective. *European Journal of Control*, *24*, 14-32.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, *28*(10), 2222-2232.

Cho, H., Seo, Y. W., Kumar, B. V., & Rajkumar, R. R. (2014, May). A multi-sensor fusion system for moving object detection and tracking in urban driving environments. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1836-1843). IEEE.

Wei, J., Snider, J. M., Kim, J., Dolan, J. M., Rajkumar, R., & Litkouhi, B. (2013, June). Towards a viable autonomous driving research platform. In *2013 IEEE Intelligent Vehicles Symposium (IV)* (pp. 763-770). IEEE.

Mei, J., Zheng, K., Zhao, L., Teng, Y., & Wang, X. (2018). A latency and reliability guaranteed resource allocation scheme for LTE V2V communication systems. *IEEE Transactions on Wireless Communications*, *17*(6), 3850-3860.

Andrews, J. G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A. C., & Zhang, J. C. (2014). What will 5G be?. *IEEE Journal on selected areas in communications*, *32*(6), 1065-1082.

Olson, E., Strom, J., Goeddel, R., Morton, R., Ranganathan, P., & Richardson, A. (2013). Exploration and mapping with autonomous robot teams. *Communications of the ACM*, *56*(3), 62-70.

Eckelmann, S., Trautmann, T., Ußler, H., Reichelt, B., & Michler, O. (2017). V2v-communication, lidar system and positioning sensors for future fusion algorithms in connected vehicles. *Transportation Research Procedia*, *27*, 69-76.

Khabbaz, M., Hasna, M., Assi, C. M., & Ghrayeb, A. (2014). Modeling and analysis of an infrastructure service request queue in multichannel v2i communications. *IEEE Transactions on Intelligent Transportation Systems*, *15*(3), 1155-1167.