Application of Deep Learning in Vehicle Driverless Technology

Ma chentian
Tianjin Chengjian University, Tianjin, China ,postcode: 300384
vivian.wang@cas-harbour.org
*Corresponding author's e-mail: 2519769784@qq.com

Abstract. In recent years, due to the rapid development of deep learning, the application of deep learning in the field of unmanned vehicle driving has achieved good results through a large number of experiments. First of all, this paper introduces the development of driverless technology at home and abroad, then it introduces the theoretical basis of deep learning. The aim is to give readers a quick look at the history and current status of driverless technology, and readers can understand the importance of deep learning in the field of driverless technology at the same time. Secondly, this paper introduced different intelligent scene recognition and compared them to find out which technology is more suitable for the current situation. Finally, the paper proposes the optimization scheme that can be improved or improved according to the shortcomings of the intelligent scene recognition technology and the problems and the development prospects of intelligent vehicles.

1. Introduction
With the development of today's society, there is a growing demand for deep learning in all walks of life. However, the problems that people encounter in urban traffic construction are also increasing, such as travel safety and road congestion, which slow down the development of urban traffic. The advent of driverless vehicle technology will effectively improve this situation, because it has abilities in different ways that humans don’t have, the first one is dealing with emergency traffic situation, next is analysis of road conditions, the last one is the choice of travel route under different circumstances. These advantages greatly prove the importance of deep learning to social problems. The development of driverless technology at home and abroad.

As early as 1950s, the research on driverless cars was started by The American Company Battery Electronics and the first autonomous navigation car was developed. Since then, more and more researches and experiments have been carried out in this field both in industry and academia. Abroad, as one of the representatives of driverless technology, Google company has profound technical accumulation in deep learning algorithm research, data acquisition and model construction, and is relatively advanced in the field of driverless cars. In recent years, it has launched many driverless cars. The first Google driverless car was adapted from the Toyota prius, the driverless vehicle will be able to sense and reconstruct its surroundings with the help of radar, which has been tested on the road in California.

The second-generation Google driverless car is a Lexus SUV, compared with the first generation, the second generation enhanced the image recognition ability of the unmanned vehicle and integrated the deep learning algorithm to improve the system environment perception ability. It can identify
nearby pedestrians, vehicles, traffic lights and other objects and visually output them in different colors and shapes. In addition, the vehicle can accurately determine the position between the unmanned vehicle and the surrounding objects. The third-generation Google unmanned vehicle maximizes the detection range of the sensor by designing the shape into a round shape. Driverless cars can tell where objects are by sensing their surroundings, and to predict the next moment of action of the surrounding objects, so as to determine their own trajectory and speed.

Figure 1. Google Driverless Car [9]

In China, Baidu's driverless technology has a representative level. With its accumulation in industrial intelligence and deep learning, Baidu has also made a series of good achievements in the field of unmanned driving in recent years. In December 2015, Baidu launched a modified BMW GT car, this unmanned vehicle uses sensors and map-assisted methods to identify road conditions, and combines the data collected in the background to make data analysis and decision through processor and cloud information, the final realization of unmanned lane change and car tracking and other functions.

2. Analysis
In the following part, a variety of target recognition research methods applied to the current driverless technology will be introduced.

2.1. CNN and RNN
CNN’s full name is convolution neural networks, which is one of the most widely used types of deep networks. Besides, Fully-connected nets treat far apart input pixels same as those close by – Hence spatial information must be inferred from the training data. In contrast, CNN proposes an architecture that inherently tries to take advantage of the spatial structure – Such an architecture makes convolutional networks fast to train – This, in turn, helps us train even deeper, many-layer networks.

Today, deep convolutional networks or some close variants are used in solving many interesting problems that go beyond image classification. People use image classification as a driving use case to explain the main concepts behind CNN. There are three main key ideas behind CNN, they are Local receptive fields, Shared weights and Pooling (you can also call it subsampling).[1]

The second basic theoretical of deep learning is RNN, which is the abbreviation of Recurrent neural networks.

There are some features about RNN. The first one is it has notion of dynamic change over time (time-varying behavior). Some neurons can fire for some limited duration of time and then go quiet. That firing stimulate other neurons, which fire a little while later for a limited duration. Over time we get a cascade of neurons firing. The second feature is that feedback loops are possible since a neuron's output only affects its input at some later time, not instantaneously. Next one is there are many kinds of ways to mathematically formalize our informal description of recurrent nets. The last one is learning algorithms for recurrent nets are (at least to date) less powerful.[2]
2.2. Alexnet

The first driverless technology introduced is Alexnet based traffic sign recognition. This type of technology is specific to static research objects, specifically traffic signs. The technology is different from the traditional static target recognition technology. In the past, the traditional one recognized the target by studying the objects’ shape, colour and other very obvious characteristics, and got the result finally. In addition, the traditional static target recognition algorithm relies on low-order features, and its scalability is poor. In the field of unmanned driving, the problems it faces in the real situation is more complex and changeable, and the traditional method cannot adapt to the changes, which is a disadvantage that it cannot be applied in real life. Therefore, a new static target recognition method in driverless technology came into being—CNNS.

Compared with the old methods, CNNS has strong feature fusion ability, feature extraction ability and information expression ability. AlexNet's emergence is a milestone in the field of deep learning. In 2012, AlexNet won the laurel of ImageNet LSVRC. AlexNet won the champion and its accuracy rate was much higher than the second place. Up until now, deep learning was not a hot field, in other words. However, after the appearance of AlexNet, to some extent, it makes deep learning become a hot algorithm in the field of computer target recognition. AlexNet is so popular because it includes the following technical features which were relatively new at that time. AlexNet is also the first successful integration of ReLU, Dropout, LRN and other key factors in CNN. AlexNet network USES GPU for computing acceleration.[5]

2.3. C3D-BR

The previous technology, AlexNet, is mainly used to study the network of static target recognition. In the actual traffic situation, not only the static target should be recognized and judged, but also the dynamic target is a very important part. This section is to introduce the C3D-BR network structure of a technology that studies the dynamic target recognition. In the field of unmanned driving technology, the main research object of C3D-BR network structure is human movement, because human movement recognition involves the acquisition of motion features, therefore, in addition to spatial features, the network also needs to carry out time-domain feature extraction of the input.

2.4. 3D-CNNs

In recent years, the performance level of 3D CNNs in the field of motion recognition has been significantly improved. What’s more, 3D CNNs has two outstanding advantages: continuous multi-frame image input and spatiotemporal feature extraction. Therefore, in the later development, 3D CNNs is more widely used in the field of video sequence processing. In comparison with 2D CNNs, the convolution kernel and pooling layer of 3D CNNs increase the time step channel, which can be clearly recognized in the following formula.[3]

\[
V_{ij}^{xy} = f \left( b_{ij} + \sum_{m} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} w_{pq}^{ijkl} v_{(i-1)m}^{x+p,y+q} \right) \quad 2D \quad (1)
\]

\[
V_{ij}^{xyz} = f \left( b_{ij} + \sum_{m} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} w_{pq}^{i,j,m} v_{(i-1)m}^{x+p,y+q,z+r} \right) \quad 3D \quad (2)
\]

Respectively the first and the second formula is 2D convolution kernels, and the calculation process of 3D convolution kernels, V represents in the formula for the layer first j a figure on the characteristics of the pixel coordinates the characteristic value of (x, y), the first k W represents the connection characteristic diagram pixel coordinates (p, q) for convolution kernel weight value, in addition, p and q are respectively the convolution kernel value of the high and wide. Similarly, in the second formula, V represents the eigenvalue whose pixel coordinates on the JTH eigengraph of the ith layer are (x, y, z), R represents the step length of the 3D convolution kernel, and W represents the pixel coordinates of the connecting MTH eigengraph are (P, Q, R) the convolution kernel weight value. F() is the activation function.
This chapter mainly studies the most important human movement in dynamic target recognition. Single frame input does not have the function to quickly and accurately determine the category of action, mainly because of the continuous feature of dynamic target in time. Therefore, to establish a new representation method, the time dimension should be combined with the actual situation.

The C3D-BR motion recognition network is a new network structure which was improved in a graduate student's graduation thesis and named as C3D-BR.

![Figure 2. Schematic of 2D and 3D convolution kernel output](image)

2.5. SceneNet Scene identifies the network model structure
The SceneNet Network model has two parts: the traditional CNN network and the SceneNetLoss layer in the latest research. The purpose of traditional technology CNN network is to obtain feature graphs in scene pictures that need to be recognized. This part is basically composed of VGG-NET, GooglNet and ResNet commonly seen in traditional convolutional neural network technology. Therefore, SceneNet network model is very strong in scalability.[4]

The second component in the SceneNet Network model, SceneNetLoss, is a new research result, which can be understood as dividing the similarity between features learned by CNN into two levels: coarse and fine-gained. To explain these two similarities, Coarse similarity refers to the large precision. If two features are extracted from different scenes that are in the same road structure attribute, they are from the same scene attribute and have certain similarity. But since they are not in the same scene, it is considered to have coarse granularity similarity. The fine-gained Similarity is the one with small precision. If the two features are not only extracted from the same scene attribute, but also from the same scene, it indicates that they have fine-grained similarity.

2.6. Build the SYSU-DS data set
In addition to the three scene recognition technologies introduced above, a data set with two important characteristics of density and difference is also indispensable for the field of unmanned driving and even deep learning. Therefore, this section will introduce the first large-scale data set of unmanned driving scene recognition, SYSU-DS.
The SYSU-DS data set is the world's first large-scale data set for driverless scene recognition. So far, SYSU-DS contains a total of 282,700 marked images of driverless scenes, 4 subsets of scene categories, 57 kinds of scene category labels, and the total number of scene labels exceeds 950,000. It is the largest and most complete set of driverless scene identification data in the world.

The above is a brief introduction to three driverless target recognition direction technologies and the first large-scale data set of driverless scene recognition. It can be seen that with the continuous development of science and technology, more and more network models and data sets will appear in the field of unmanned driving. The progress of network models and data sets is complementary to each other and indispensable. However, despite the rapid pace of innovation, there will be more problems in the future.[6]

3. Conclusion
This paper mainly introduces the specific application of deep learning in the field of unmanned driving. Firstly, this paper introduces the history and current situation of driverless vehicles, which is divided into the development of driverless vehicles abroad and the specific situation in China. In addition, the specific introduction and research of the basic knowledge of deep learning are carried out. This chapter is very important, because only after laying a good foundation can we have a clearer understanding of the following chapters. Next, the static target recognition technology is studied. Aiming at the unique characteristics of static targets and considering the limitations of traditional methods, 2D CNNs is used to improve the advantages of feature extraction, so as to make the static target recognition technology more mature and more suitable for the current situation. After studying the static target, the research object becomes the dynamic target recognition. Like the static target recognition, the C3D-BR network is proposed after considering the characteristics of the dynamic target. Since dynamic is more complex than static and its features are more complicated, RNNs is combined on the basis of using 3D CNNs for the purpose of improving research capability. Finally, sysu-DS, the world's first large-scale data set for driverless scene recognition, is introduced in detail.

With the progress of science and technology, the application of deep learning in the field of unmanned driving is becoming more and more mature, such as the technology described above, but this is still the beginning of this technology, and there are still some shortcomings.

A large amount of data is needed for training in CNNs, but manual input alone will not only consume time and energy, but also have a great impact on training results due to manual errors. There
is still a big gap between target recognition of driverless vehicles (either static target recognition or dynamic target recognition) and human brain's understanding of target recognition. The amount of data in the scene identification data set is still not rich enough, and the development speed of network model construction is also very slow.

In view of the above problems, deep learning still needs to be developed in the field of vehicle unmanned driving. In the future, target recognition for driverless cars combined with deep information and semantic understanding could make it easy to solve the first and second problems mentioned above. To solve the third problem, it is necessary to combine the data set with the Internet to crawl more picture data related to scene recognition to narrow the gap.

In addition to the above problems related to the content of this paper, the impact of network simplification and compression and deep learning on the safety performance of unmanned vehicles should also become the research focus in the future.

Acknowledgments
Writing to the last place, it means that this research is coming to an end. Even though I experienced the difficulties in scientific research at the beginning, I began to adapt to the new field of study in the middle of the course, and finally I successfully finished the writing of the paper. I was mixed with excitement and reluctance in my heart. Looking back on this summer vacation, the scientific research activities were very happy and smooth, which could not have been more than the teachers and students' help to me. I would like to express my heartfelt thanks to them here.

References
[1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional
[2] LeC Y, Botton L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
[3] Yankun Z Y Z, Hong C H C, Weyrich N. A single camera based rear obstacle detectionsystem[C]// Intelligent Vehicles Symposium IEEE, 2011
[4] Morales-Gonzalez, Annette, Garcia-Reyes, et al. Simple object recognition based on spatial relations and visual features; represented using irregular pyramids[J]. Multimedia Tools & Applications, 2013, 63(3): 875-897.
[5] Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks[J]. 2013.
[6] Mahendran A, Vedaldi A. Understanding Deep Image Representations by Inverting Them[J]. 2014.
[7] Han Xinfai. Driverless scene recognition based on deep learning[D]. Sun yat-sen university, 2017.
[8] Song Jiarong. Research on target Recognition of Unmanned vehicle based on deep learning[D]. Nanjing University of Aeronautics and Astronautics, 2019.
[9] Google Company. https://www.egouz.com/topics/16023.html