Research on Visual Perception Technology of Autonomous Driving Based on Improved Convolutional Neural Network

Yucheng Wang 1*, Zhihao Li 2, Hongye Hao 3, Hengyu Yang 4, Yixin Zheng 5
1,2,3,4 Chongqing University, Chongqing, Chongqing, 400000, China
5 Southwest University, Chongqing, Chongqing, 400700, China
*Corresponding author’s e-mail: jameswangyucheng@foxmail.com

Abstract. With the continuous development of the world's scientific and technological level and the continuous progress of society and economy, driverless cars have become the key research targets of the existing automotive industry. Based on the convolutional neural network theory, this paper redesigned the Faster R-CNN network framework to improve the target recognition accuracy and apply it to the unmanned visual perception system. At the same time, the improved algorithm is used to test the cars under different simulated travel conditions. The computer vision system toolbox of MATLAB was used to verify the effectiveness of the algorithm, and the driving scene designer was used to construct multi-sensor fusion and simulated road scenes, and the corresponding visual control system was designed to effectively realize the automatic code generation. The experimental results show that the average detection accuracy of the improved algorithm is higher, the algorithm is more feasible and robust, and it is helpful to promote the development of driverless cars.

1. Introduction
With the rapid development of artificial intelligence technology in recent years, breakthroughs have been made in visual perception technology centered on deep learning. At present, with the support of Baidu Apollo open source platform in China, a variety of visual perception systems have been mass-produced. Many startups have also implemented their own autonomous driving solutions. The Chinese Smart Car Future Challenge, which is jointly participated by many domestic universities and colleges, has also greatly promoted the theoretical research and popularization of related technologies. As an extremely critical part of driverless cars, environmental sensing requires deep learning-based target detection algorithms to be improved in recognition speed and accuracy [1-2]. Therefore, this paper improves the Faster R-CNN algorithm, and quickly verifies its rationality through the MATLAB computer vision toolbox, and then proposes a visual perception scene simulation scheme based on the MATLAB simulation system.

2. Traditional neural networks
Traditional neural networks use human-like neural networks to gradually build mathematical models, from perceptrons and multilayer perceptrons to the most classic fully connected BP neural networks. As shown in Figure 1. The BP neural network corrects the weight of each layer according to the error signal, and then adjusts the weight connection through the gradient descent method, and continues to alternate forward and backward propagation until the error reaches the minimum or close to the expected set value [3-4].
BP neural network reverse transfer process:
It is mainly based on the WIDROW-HOFF learning rules and gradient descent method to mathematically push the director and continuously update the weight and threshold to get the following formula:

\[ \beta_j = \beta_j - \eta \frac{\partial e(w, \beta)}{\partial \beta} \]  

\[ \beta_j = w_{ij} - \eta \frac{\partial e(w, \beta)}{\partial w_{ij}} \]

In the formula: \( i \) and \( j \) represent two different nodes respectively; \( w_{ij} \) is the weight between node \( i \) and node \( j \); \( \beta_j \) is the threshold value of node \( j \); \( w_{ij} \) and \( \beta_j \) are continuously updated by operations; \( e(w, \beta) \) is the error function, and the last term is the first-order partial derivative of the error function to the weight or threshold.

3. Convolutional neural network
A convolutional neural network consists of an input layer, a convolutional layer, an activation function layer, a pooling layer, and an output layer. Multi-level alternating matching is often used to achieve better optimization results. The principle of forward and back propagation of the convolution layer is analyzed below. as shown in picture 2.

Convolutional layer forward propagation [5-6]:

![Figure 1. BP neural network structure](image1)

![Figure 2. Convolutional neural network structure](image2)
\[ X'_l = F \left( \sum_{i \in M_l} X'_{l-1} \ast K'_j + O'_j \right) \]  \hspace{1cm} (3)

In the formula: \( X'_{l-1} \) is the \( i \)-th feature map of the \( l-1 \) layer; \( K'_j \) is the \( j \)-th convolution kernel of the \( l \)-th layer; \( O'_j \) is the bias parameter.

Back propagation of the convolutional layer:
\[ \psi'_j = \phi'^{l+1} \left[ F'(\mu'_j) \right] \circ up(\psi'^{l+1}) \]  \hspace{1cm} (4)

In the formula: \( \psi'_j \) represents the convolution layer of the first layer; \( \phi'^{l+1} \) represents the pooling layer of the \( l+1 \) layer; \( p(\psi'^{l+1}) \) is to expand the size of the \( l+1 \) layer to the \( l \) layer. The size of the feature map; \( F'(\mu'_j) \) is the derivative of the activation function, and \( \circ \) represents the multiplication of multiple elements.

4. Feature extraction and target detection classification methods

4.1. Feature extraction
In the process of detecting unmanned vehicles in front, traditional feature extraction algorithms require manual extraction and calibration, and then developed to extract edge features, shape features, color contrast features, texture features, and spatial relationship features. Or non-linear factors to improve extraction accuracy. The CNN's local connection and weight-sharing features enable different convolution kernels to extract different features, and its multi-layer feature extraction can effectively collect information and reduce image dimensions.

4.2. Comparison of target detection and classification algorithms

4.2.1 Object detection and classification of traditional models
Traditional object detection and classification methods include gradient histogram (HOG), HOG + support vector machine (SVM), HOG + cascade (Cascade), Haar features (Linear, Center, Diagonal Features) + Iterations (Adaboost), DPM, etc. They can quickly and accurately capture local features and have strong invariance in terms of geometry and optics. However, the above method is obviously unable to meet the requirements in the case of complicated road conditions, object occlusion and accurate extraction and comparison of multiple features.

4.2.2 Deep learning recognition and classification of the window of interest area
R-CNN, SPP-net, Fast R-CNN, Faster R-CNN, and R-FCN algorithms based on the evolution of convolutional neural networks detect and classify images by finding areas of interest. Faster R-CNN has been proposed since 2015, and it is considered as an algorithm with extremely high detection accuracy. In order to make it better applied to the required scene, the algorithm used in the experiments below is improved while retaining the high-precision characteristics of Faster R-CNN.

4.2.3 Detection and classification of deep learning regression models
This type of algorithm does not require nomination of candidate regions and can be classified as an end-to-end deep learning model. YOLO, SSD, and DenseBox are their typical representatives. In particular, YOLO has become the fastest target detection algorithm at present, and YOLO-based algorithms have been continuously proposed and evolved. YOLO is regarded as a regression model and does not require complicated channel connections. It can use full-picture information for prediction and quickly and accurately identify relevant information about the target.
5. Experimental process
In this experiment, MATLAB2018a was used to build the model and simulate the test. The specific operation process is shown in Figure 3.

5.1. Object detection of MATLAB computer vision toolbox combined with convolutional neural network
Import part of the self-built data set into the data set that comes with the computer vision toolbox, so that it can be easily called and implemented quickly. After the following improved algorithm framework is constructed, the algorithm is trained, and it is called directly after the model training is completed. The improved algorithm model is combined with the computer vision toolbox, automatic driving toolbox and neural network toolbox to realize the detection and recognition of multi-target vehicles in different scenarios. As shown in Figure 4. The convolutional neural network algorithm in this paper is improved based on Faster R-CNN as follows:

(a) Vehicle detection and recognition on highway
Vehicle Target Detection and Recognition at Crossroads with Complex Road Conditions

Figure 4. Vehicle target detection in three different scenarios

(1) Replace RPN with FPN for more accurate detection, reduce the number of Proposal, and fuse multi-feature information to improve accuracy;
(2) By increasing the ROI to avoid situations where valid information is blocked and identification is inaccurate when driving multiple vehicles
(3) Replace the basic network from VGG16 to ResNet residual network, so that the deep neural network's network performance is not easily weakened with the increase of the number of layers in the training process.

5.2. Data analysis
It can be seen from the analysis in Figure 4 that the improved convolution neural network algorithm in this paper can obtain high vehicle target detection accuracy in high-speed road, cross complex intersection and urban street scene, with the minimum value of 78.42%, the maximum value of 98.84% and the average detection accuracy of more than 85%, which proves that the target detection algorithm in this paper has good robustness and feasibility.

(1) Vehicle detection and recognition on highway
(2) Detection and recognition of vehicles at complex intersections
(3) Urban street vehicle detection and recognition

5.3. Automatic driving toolbox
Visual perception model design is implemented in Simulink's Automated Driving System Toolbox (automatic driving system toolbox). The improved convolution neural network and the sensor model built in the driving scene designer are used to interact with the visual perception model built by Simulink to verify, so as to obtain a safe and reliable visual perception scheme.
5.4. Road scene simulation
For complex urban overpasses, viaducts, roundabouts, roadways, culverts, tunnels and other special scenes, a lot of experiments and simulations are needed to achieve the best safety perception effect. The driving scenario designer of MATLAB is convenient to build road scene model quickly, and load moving objects, obstacles and visual sensors for real-time simulation.

5.5. Model generation and automatic code generation
(1) After the road scene, vehicle, pedestrian, obstacle and moving object path planning are configured through the driving scene designer, if the operation results meet the requirements, export can be performed. After output, the matlab code and relevant data parameters of the design scenario can be automatically generated.

(2) After the model design of the autopilot sensing system is completed by Simulink, the relevant parameters are configured. After configuration, embedded coder quick star based on C / C++ can be compiled, and C / C++ code can be generated automatically after relevant compilation options are completed. The automatically generated code can be transplanted to the required engineering file after fine tuning and optimization, which is convenient for use on the embedded controller, and then the rationality of the above design scheme is verified.

6. Conclusion
This paper introduces a design scheme of vision perception system for driverless vehicle based on MATLAB. The experiment shows that the improved fasterrcnn algorithm can improve the detection accuracy, and the multi-target detection for different scenes can reach more than 80%, which proves the excellence and robustness of the algorithm. The combination of automatic driving toolbox, neural network toolbox, Computer Vision Toolbox, Simulink simulation control and other technologies in matlab2018a is convenient for the rapid verification of target detection technology, which greatly accelerates the research and development of automatic driving perception system in simulation design and simulation control, and greatly reduces the development cycle of researchers.

References
[1] Zhang, Q., Wei, R. (2019) Unmanned Aerial Vehicle Perception System Following Visual Cognition Invariance Mechanism. IEEE Access., 7: 45951–45960.
[2] Yu, B., Bao, S., Chen Y. (2019) Using 3D Mobile Mapping to Evaluate Intersection Design Through Drivers’ Visual Perception. IEEE Access., 7: 19222–19231.
[3] Yang, A., Zhuansun, Y., Liu, C., et al. (2019) Design of Intrusion Detection System for Internet of Things Based on Improved BP Neural Network. IEEE Access., 7: 106043–106052.
[4] Yei, W., Feng, G., Li, Z., et al. (2005) Deterministic convergence of an online gradient method for BP neural networks. IEEE Transactions on Neural Networks., 16: 533-540.
[5] Zhang, P., Wang, X., Zhang, W., et al. (2019) Learning Spatial–Spectral–Temporal EEG Features With Recurrent 3D Convolutional Neural Networks for Cross-Task Mental Workload Assessment. IEEE Transactions on Neural Systems and Rehabilitation Engineering., 27: 31-42.
[6] Passalis, N., Tefas, A. (2019) Training Lightweight Deep Convolutional Neural Networks Using Bag-of-Features Pooling. IEEE Transactions on Neural Networks and Learning Systems., 30: 1705-1715.