Research Article

Intelligent Control System of Unmanned Vehicle Based on CAN Controller

Jinhua Wu ⓘ, Yunfei Jiang ⓘ, and Fang Deng ⓘ

Jiangxi College of Applied Technology, Ganzhou 341000, China

Correspondence should be addressed to Jinhua Wu; hyperwuchn@cumt.edu.cn

In order to enable the driverless vehicle formation controller to automatically control the driving of the fleet, an intelligent driverless vehicle control system based on the CAN controller is proposed. Expansion convolution of different expansion rates is used to obtain multi-scale target information and to fuse the feature information at different scales during upsampling to enrich the semantic information. Finally, the driverless CAN bus communication platform was established, the driverless monitoring interface was developed, and the software program was written; experiments on the steering control, speed control, voltage, current, speed, and angular speed acquisition, respectively, were performed. The experimental results show that the average semantic segmentation accuracy of the obstacles in concentrated vehicles, pedestrians, and bicycles reached 84.6%, and the detection and segmentation accuracy of the models was good. Therefore, the unmanned intelligent vehicle control system designed in this paper can meet the performance requirements of the vehicle control. No matter whether the given desired path is a straight line or a curve, the unmanned car can complete the path tracking control quickly, stably, and accurately.

1. Introduction

In recent years, with the progress of hardware and software, unmanned driving technology has developed rapidly. For driverless vehicles, accurate detection of obstacles is crucial for safe driving [1]. With the development of deep convolutional neural network, semantic segmentation technology has been applied more and more widely in the field of unmanned driving environment perception, especially obstacle detection. For driverless vehicles, pedestrians, vehicles, and cyclists on the road can all be regarded as "obstacles," which have a huge impact on the safe driving of driverless vehicles. Semantic segmentation technology based on convolutional neural network can detect and identify these obstacles at pixel level, which can provide environmental information for the decision, planning, and control of unmanned vehicles [2]. The core problems to be solved in the detection and identification of real scenes by driverless vehicles are the improvement of detection speed of algorithms, the accurate classification of obstacles in driving scenes in various extreme environments, and the extraction

of road semantic information. Conventional visual algorithms based on image features are difficult to play a practical role in the highly dynamic environment of unmanned driving [3]. In recent years, visual detection and recognition algorithms based on neural network have become a new force and achieved good results in various visual detection and recognition applications. At present, using deep learning method to detect, track, and identify dynamic targets has become the mainstream. Deep learning methods include supervised learning, unsupervised learning, and semi-supervised learning, and its core problems mainly include: Construction of training and testing data sets, design of neural network structure, construction of loss functions for different application scenarios, and design of fast and efficient numerical optimization algorithms [4]. The only way for deep learning methods to be applied effectively to driverless vision is to comprehensively deal with all kinds of problems listed above. The convolutional neural network-based semantic segmentation subject frameworks mainly include VGGNet and ResNet VGGNet-based semantic segmentation models include FCN, SegNet, U-net, DeepLab,

etc., while resnet-based semantic segmentation models include PSPNet, ICNet, DeepLab V3++ [5], etc. [6]. In view of the structural characteristics of ResNet network, the semantic segmentation network based on ResNet framework has many layers and complex network structure, which requires high performance of hardware facilities [7]. The semantic segmentation network based on VGGNet framework is much simpler than that based on ResNet framework, but it often fails to meet the real-time requirements. For driverless cars, the real-time performance of the system is crucial, and the accuracy of segmentation results should also be taken into account [8].

Yamane et al. proposed an object detection model combining Bayes optimization with structured prediction, which can improve the positioning accuracy of the identified target through structured loss function [5]. Hua et al. proposed THE MR-CNN/S-CNN/LOC model to improve the accuracy of detection simultaneously through the candidate target region and depth feature map [9]. In this method, the candidate regions were first divided into multiple sub-regions with different classes, and then the corresponding features of these regions were extracted by MR-CNN. Yukun Zhu et al. proposed that segDeepM model can improve the results of target recognition and detection by dividing the environment information of target region [10].

Inspired by the above semantic segmentation models, a lightweight semantic segmentation model is proposed for obstacle detection to meet the real-time and accuracy requirements of unmanned vehicles. The model can greatly reduce the calculation in the next sampling stage, and can effectively extract the image features. In the upsampling stage, the lightweight semantic segmentation model proposed in this paper incorporates the multi-stage original subsampling features, which improves the accuracy of the segmentation model, and the extracted image features are restored to the original image size, and the pixel level segmentation is carried out according to the image semantics.

2. Related Work

Most common deep convolutional neural networks directly use 3×3 size convolution kernel to extract convolution features, which leads to a large number of model parameters and affects the execution speed and memory overhead of the model. At the same time, with the increase of the depth of the network, the problem of gradient disappearance will appear in the training model, which makes it difficult to update the weight of the shallow network in the back propagation, resulting in the performance of the model reduced. Drawing on ResNet residual module and Inception module in GoogleNet, this paper proposes feature extraction block to extract image features. The structure of feature extraction block is shown in Figure 1.

The feature extraction block has two branches. One branch directly uses 1×1 convolution to extract features, the number of convolution kernels is $2n$, and the step size $s = 2$. In the other branch, detachable convolution operation is
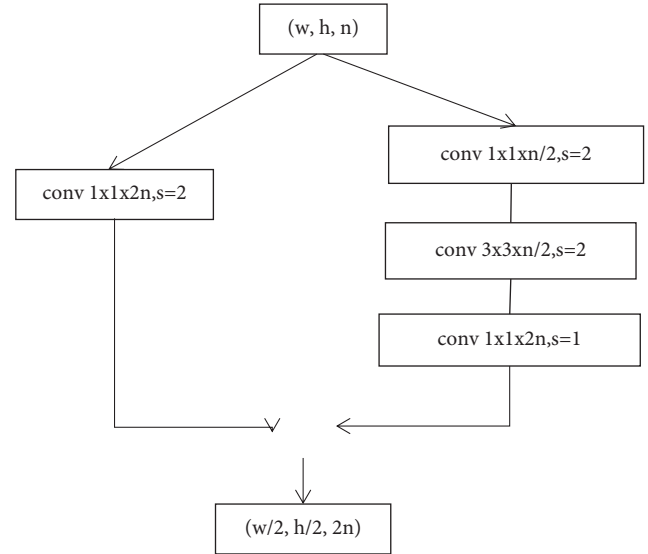

FIGURE 1: Feature extraction block structure.

adopted. First, 1×1 convolution is used to extract features. The number of convolution kernels is $n/2$, the convolution step $s = 2$. After that, 3×3 convolution was used to extract features. The number of convolution kernels was n/2, and the convolution step size was $s = 1$. Finally, 1×1 convolution is used for dimension raising operation of feature channel. The number of convolution kernels is 2n, and the number of channels of the output feature map becomes $2n$ with convolution step $s = 1$. Finally, the corresponding positions of the output of the two branches are superposed, and the size of the output feature graph is ($h/2$, $W/2$, $2n$). The proposed feature extraction block refers to the ResNet residual module and fuses the low-level features with the advanced features, which can refine the boundary and improve the edge segmentation accuracy.

2.1. Fast Target Detection and Recognition Algorithm Analysis. Deep neural network-based detection and recognition algorithms have become the mainstream method in this field, and they are far beyond the traditional visual algorithms in most visual perception research directions. The most representative example is that in 2012, the team led by Professor Hinton constructed AlexNet using deep convolutional neural network and trained it on the ImageNet data set, successfully beating all the traditional target recognition methods at that time. According to existing research, deep neural network-based target detection and classification algorithms can be divided into the following three categories: 1. Target detection and classification based on region prediction represented by R–CNN, Fast R–CNN, and Faster R–CNN; 2. Detection and classification algorithm based on regression analysis represented by POLO and SSD; 3. Search-based detection and classification algorithms represented by reinforcement learning and AttentionNet.

For deep neural network-based target detection and classification algorithms, large training and detection databases are needed. As such data sets are generally large, the

existing algorithm verification is generally based on the data sets publicly available on the network. Among these data sets, the most representative and most widely used are ImageNet data set, COCO data set, and PASCAL VOC data set. The images in these data sets are RGB images.

2.1.1. Data Set ImageNet.
ImageNet contains more than 14 million color images and a thousand target categories, which can be widely used for image target detection, location, and classification. There are more than 100 images in the data set that clearly marked the image pose and category information, including SIFT features, target attribute values, and target borders. The ImageNet data set is the largest image Recognition data set in the world, and the LargeScale Visual Recognition Challenge, a computer vision competition based on it, has greatly advanced the field. In 2015, researchers for the first time used software to surpass human task processing capabilities on the ImageNet data set. However, Olga Russakovsky points out that the program is only capable of sorting a thousand categories, whereas humans have far more than 1,000 categories.

2.1.2. Data Set VOC (Computational and Learning Visual Object Classes).
PASCAL VOC (Pattern Analysis Statistical Modeling, Computational and Learning Visual Object Classes) Data set is a kind of standardized data set for image target classification and recognition. An image recognition challenge based on this data set was held every year from 2005 to 2012. The data set contains five folders: Annotations, ImageSets, JPEGImages, SegmentationClass, and Segmentation0.ect. There are more than 10,000 images and 20 target categories, these categories include: people, animals (cats, dogs, cows, sheep, horses, and birds), means of transportation (bicycles, motorcycles, cars, trains, and boats) and some indoor objects (chairs, tables, sofas, televisions, potted plants, and bottles). VOC data set consists of two parts, namely, training set and test set, and its label information is saved in XML format. Of the five folders described above, the JPEGImage folder holds all the images in the data set, these images are named "year_number" JPg and are approximately 500×375 in size for landscape and 375×500 in size for portrait.

2.2. Road Semantic Segmentation Based on VGG16-FCN8 Network.
The output of each layer of the convolutional neural network is a three-dimensional tensor denoted by $H \times W \times d$, where H and W are spatial dimensions, and D represents the dimension of features or channels. The first layer of the network is h by W images. The higher level corresponds to the position in the image to which their path connects, which is called their local receptive field. Convolutional neural networks have inherent translation invariance. Their basic operations (convolution, pooling, and activation) operate on the local input region, relying only on the relevant spatial coordinates. For the data vector x_{ij} at the position (I, J) of a specific layer, the input corresponding to the next layer can be expressed as:

$$y_{ij} = f_{ks}\left(\left\{X_{si+\delta i+\delta j}\right\}_{0 \le \delta i, \delta j < k}\right). \tag{1}$$

Here k represents the dimension of the convolution kernel, S represents the sampling step, and f_{ks} represents the type of layer. The functional form of (2) satisfies the following transformation rule ,

$$f_{ks} \circ g_{k's'} = (f \circ g)_{k'+(k-1)s',ss'}\bullet \tag{2}$$

A nonlinear filter based only on the above form is called a full convolution network. A full convolutional neural network can input images of any dimension and output images of the same dimension.

Generally, CNN network will be connected with several fully connected layers behind the convolution layer, which will transform the feature map generated by the convolution layer into a feature vector with a fixed dimension. FCN is a classification network at pixel level, which can be used for image segmentation at semantic level. Different from CNN, FCN can input images of any size and perform upsampling of feature map through reverse pen stacking, and can restore the output to the original image size to achieve the classification of each I pixel. In short, the difference between FCN and CNN is that CNN's full connection layer is replaced by volumes and layers, which output marked images. We know that CNN can automatically learn different levels of features. The shallow perception area is small and the learning is local features. Deeper layers can learn more abstract features. The knowledge area is small, and the learning is local features. Deeper layers can learn more abstract features. These abstract features are less sensitive to the size, direction, and position of information in the image, which helps to improve the recognition performance. However, because these features lose some details of the object in the picture, it is difficult to give the specific outline of the object, and the object to which each pixel belongs, so the precise segmentation of the object cannot be satisfied. The full convolutional network FCN recovers the category of each pixel from the abstract features.

2.3. Design of CAN Controller for Unmanned Intelligent Vehicle

2.3.1. CAN Controller.
Considering the diversification of sensor output information formats, such as the rotation angle value detected by the multi-turn absolute encoder, the inclination angle output by the inclination sensor, the pose value output by the GPS in RS232 format, and angular acceleration output in RS485 format, the number of buses increases. The simple computer serial port is not enough for engineering use, and the reliability, flexibility, and real-time performance will be deteriorated. At this time, these buses must be "merged" into one category, so the RS232/RS485 to CAN controller is designed.

2.3.2. CAN Bussing Technique.
CAN bus technology itself is a serial data communication method, and its transmission rate can reach up to 1 Mbps. According to the communication,
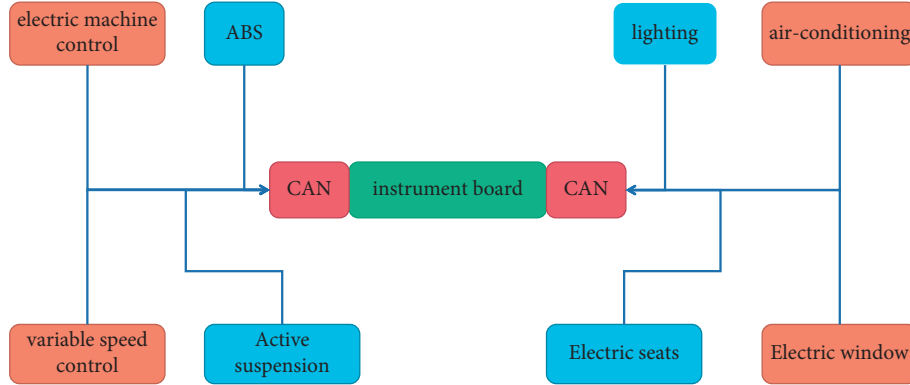
Figure 2: Application of automobile CAN bus.

CAN bus has outstanding reliability, flexibility, and real-time performance, and it is widely used as a very effective formation of split stone J-type, real-time detection, and control system.

The Controller Area Network bus was originally a digital signal communication protocol designed by German Bosch Company to solve many complex technologies and problems in the automotive monitoring system. It belongs to a bus-type serial communication network. In 1991, Bosch Company formulated and released the technical specifications of CAN2.OA and CAN2.OB. Among them, CAN2.OA gave the CAN standard message format, and CAN2.OB gave the standard frame and extended frame format. With the formation of the international standard of CAN bus, its application scope is not only limited to the automotive industry but also has been extended to a wide range of fields such as robots, CNC machine tools, machinery industry, medical machinery, textile machinery, and household appliances. The current application of CAN in automobiles is shown in Figure 2.

3. Experimental Analysis

3.1. Experimental Environment and Data Sets. The semantic segmentation model was built using the TensorFlow deep learning framework. The hardware and software configurations of the machine are shown in Table 1.

The selection of data sets has a crucial impact on the training of the model. In order to make the trained semantic segmentation model adapt to a variety of complex realistic environments and realize the accurate detection of obstacles such as vehicles and pedestrians, apolloScape data set launched by Baidu Apollo in China and Cityscapes data set promoted by Mercedes-Benz abroad were selected to train the model.

3.2. Model Training. In order to give full play to the advantages of the selected data set, the training model is divided into two stages. In the first stage, training is carried out on foreign Cityscapes data sets, mainly to obtain the weight parameter data and semantic feature information of the preliminary model fitting. In the second stage, training was conducted on China ApolloScape data set to adjust and optimize the weight parameters fitted in the first stage to make the model more

Table 1: Experimental configuration.

Project	Content
CPU	Intel(R)Core(TM) i5-7500CPU@ 3. 4 GHz
GPU	NVIDIA GeForce GTX 1070TI
Internal storage	8 GB
Operating system	Ubuntu 16. 04 LTS
CUDA	CUDA9. 0 with CUDNN7. 0
TensorFlow	1. 7 with Python2. 7

adaptable to China's traffic environment and obtain more accurate semantic segmentation results.

The calculation formula of Softmax function used by the model is as follows:

$$P_K = \frac{\exp(a_k(x))}{k = \sum_{k=1}^{k} \exp(a_k(x))},$$ (3)

where x is the pixel position on the feature map, $a_k(x)$ represents the value of the KTH channel corresponding to pixel x in the last output layer of the network, and $p_k(x)$ represents the probability that pixel X belongs to class k.

The loss function uses negative class cross entropy:

$$L = \sum_x w(x)\log(p_{l(x)}(x)),$$ (4)

where $p_{l(x)}(x)$ represents the output probability of pixel X on the channel where the real label is located, $w(x)$ represents the category probability that pixel X belongs to in the real label, and the value is 1 or 0.

The training model adopts momentum gradient descent method, the initial momentum is set as 9, the learning rate is set as 0.001, the weight attenuation coefficient is set as 0.0005, the batch data amount is set as 1, and the whole training sample data is learned for 10 rounds. The curve of the relation between the loss value and the number of iterations during the model training is shown in Figure 3. The training sample data was iterated for 10 rounds, and the weight parameters were updated for 100,000 times. In Figure 3, the red curve represents the pre-training on the foreign Cityscapes data set in the first stage, while the black curve represents the retraining on the Chinese ApolloScape data set in the second stage. As can be seen from Figure 3, after the pre-training, the loss function converges faster and
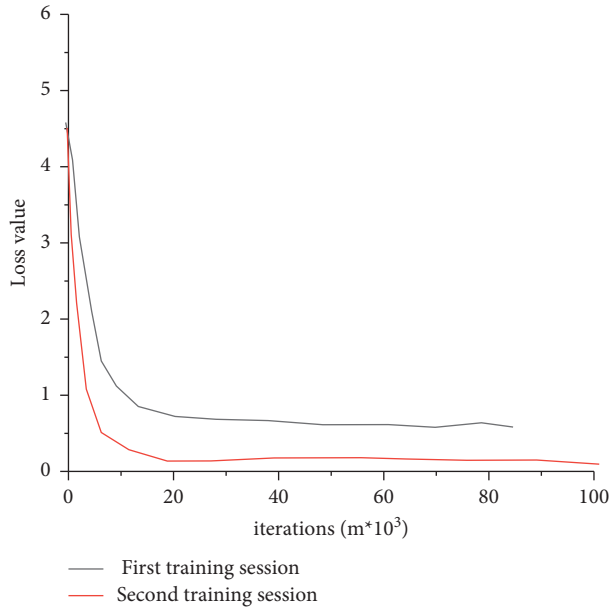
FIGURE 3: Loss function comparison curve.

TABLE 2: Performance of semantic segmentation model.

mIoU (%)	Model size (Mb)	Processing speed (FPS)
85.7	11.3	33

the loss value is close to 0, showing a better effect on the test set and higher detection accuracy of obstacles.

3.3. Analysis of Experimental Results. Based on the proposed semantic segmentation model, pedestrians, cyclists, cars, and two wheelers (bicycle, electric car, and motorcycle) are segmented and identified, respectively, represented by pink, red, and blue on the picture. This model can effectively detect the target obstacles such as vehicles and pedestrians. The autonomous driving research platform is used to collect actual road images, the trained language segmentation model is used to detect obstacles, and the trained semantic segmentation model has more accurate detection results for obstacles such as vehicles and cyclists in the real scene. The average pixel overlap rate (mIoU) is often used to evaluate the accuracy of the semantic segmentation model, which can effectively measure the segmentation performance of the model for target obstacles. 1000 images were selected in the test data set, including 500 Chinese road images and 500 foreign road images, the Chinese road images were from the Chinese ApolloScape data set and the foreign road images were from the foreign Cityscapes data set. The performance of the constructed semantic segmentation model is shown in Table 2.

In the test set, the average accuracy of semantic segmentation of obstacles such as vehicles, pedestrians, and cyclists is 84.6%. In terms of the overall effect of detection and segmentation, the designed and built model has a good detection and segmentation accuracy. In the target detection and segmentation of a single image, the time cost is about 30 ms, and the average is 33FPS. In terms of the time consumed by detection and recognition, the model has met the requirement of real-time segmentation of target objects. The model size is 11.7 Mb and the memory usage is low, which meets the requirements of the on-board computing model.

4. Conclusions

This paper studies the control system of the unmanned intelligent vehicle. The core is to design the lateral control unit and the longitudinal control unit for the unmanned vehicle, and combine the characteristics of the vehicle and the safety of the driving vehicle to complete the lateral control system of the unmanned intelligent vehicle and Design with longitudinal control system. In the lateral control system, after setting the desired path for the unmanned vehicle, the unmanned vehicle can quickly and stably follow the given path and can drive at a certain speed. At present, the main completed work:

To meet the requirements of real-time performance and accuracy for vehicle terminal obstacle detection, a lightweight semantic segmentation model is constructed using feature extraction blocks, depthwise separable convolutions, and dilated convolutions. The feature extraction block draws lessons from the structure of residual modules, and the skip-layer structure can refine boundary information by combining low-level features with high-level features. The depthwise separable convolution can effectively reduce the number of parameters and computation of the model, and the dilated convolution operations with different dilation rates can extract multi-scale target information and enrich the semantic information. The unmanned vehicle bus control system is completed by designing a CAN converter, and a stable experimental platform is built for the unmanned vehicle control system experiment. And the monitoring of the performance states of the unmanned vehicle is carried out. From the monitoring data, it can be concluded that the experimental platform established in this experiment can meet the safety requirements of the unmanned vehicle. To meet the requirements of full travel, using the unmanned test vehicle as a platform combined with longitudinal control and lateral control algorithms, it is verified that the unmanned vehicle can accurately track the path and smoothly adjust the vehicle running speed when the unmanned vehicle has a given path in a straight line or a curve.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] L. Ye, T. Duan, and J. Zhu, "Neural network-based semantic segmentation model for robot perception of driverless vision," *IET Cyber-Systems and Robotics*, vol. 2, no. 4, pp. 190–196, 2020.

[2] A. Shen, H. Dong, K. Wang, Y. Kong, and H. Shu, "Automatic extraction of blur regions on a single image based on semantic segmentation," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.

[3] N. Audebert, B. Le Saux, and S. Lefèvre, "Segment-before-detect: vehicle detection and classification through semantic segmentation of aerial images," *Remote Sensing*, vol. 9, no. 4, p. 368, 2017.

[4] L. Zhiyi, S. Shaoyuan, R. Zhengyun et al., "Semantic segmentation of nocturnal infrared images of unmanned vehicles based on improved deeplabv3+," *Journal of Applied Optics*, vol. 41, no. 1, pp. 180–185, 2020.

[5] T. Yamane and P.-j. Chun, "Crack detection from a concrete surface image based on semantic segmentation using deep learning," *Journal of Advanced Concrete Technology*, vol. 18, no. 9, pp. 493–504, 2020.

[6] H. Liu, H. Du, D. Zeng, and Q. Tian, "Cloud detection using super pixel classification and semantic segmentation," *Journal of Computer Science and Technology*, vol. 34, no. 3, pp. 622–633, 2019.

[7] S. Hwang, S. Cho, and H. Shim, "Aircraft detection using deep convolutional neural networkbased semantic segmentation," *Journal of Institute of Control, Robotics and Systems*, vol. 23, no. 8, pp. 625–634, 2017.

[8] K. Gao, S. Sun, G. Yao, and H. Zhao, "Semantic segmentation of night vision images for unmanned vehicles based on deep learning," *Journal of Applied Optics*, vol. 38, no. 3, pp. 421–428, 2017.

[9] Y. Hua, D. Marcos, L. Mou, X. X. Zhu, and D. Tuia, "Semantic segmentation of remote sensing images with sparse annotations," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, no. 99, pp. 1–5, 2022.

[10] A. d. F. Lemos, L. A. R. da Silva, and B. V. Nagy, "Correction to: automatic monitoring of steel strip positioning error based on semantic segmentation," *International Journal of Advanced Manufacturing Technology*, vol. 110, no. 11-12, pp. 2861-2862, 2020.

[11] H. Gao, L. Cao, D. Yu, X. Xiong, and M. Cao, "Semantic segmentation of marine remote sensing based on a cross direction attention mechanism," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.