

# Research on Autopilot Trajectory Planning Based on Deep Learning

Qiong Ren<sup>1, a</sup>, Junming Chang<sup>1, b</sup>

<sup>1</sup>Jiangnan University, Wuhan, 430056, China  
<sup>a</sup>qiongren@jhun.edu.cn, <sup>b</sup>JmChang@163.com

---

**Abstract:** Vehicle trajectory is usually used in driving behavior modeling and traffic safety. In order to obtain vehicle tracks in traffic video, an automatic acquisition method based on depth learning is proposed. In this paper, the YOLOV5 detector is used to detect vehicles in the region of interest in continuous frames, and the DEEPSORT tracker is used to track robust and fast vehicles. This tracker can effectively reduce the jump of target ID, improve tracking stability, obtain the real-world coordinates of vehicles through perspective transformation, and use a locally weighted regression algorithm to smooth the track. The experimental results show that the average multi-target tracking accuracy of the method in the expressway monitoring data set is 96.1%, and the average accuracy of trajectory acquisition is more than 93%. The smoothing algorithm can effectively remove the trajectory noise.

**Keywords:** Deep learning, Automatic driving, Trajectory planning.

---

## 1. Introduction

Traffic information is the basis and guarantee of an intelligent transportation system. Obtaining comprehensive, accurate, reliable, and real-time traffic information is vital in traffic management and safety, driving behavior modeling, and microscopic traffic flow analysis. The collection methods of traffic information include microwave radar, video, infrared and induction coil, etc. In these collection technologies, video images have the characteristics of high intuitiveness, rich information, and robust traceability. More and more researchers use image processing technology to extract traffic data from video. In these data, the driving track of vehicles fully reflects the driving behavior of vehicles, and microscopic parameters of traffic flow, such as road traffic volume, vehicle speed, headway, and so on, can be obtained by processing the vehicle track. The accurate extraction of vehicle trajectory is of great significance for analyzing traffic behavior and obtaining the running state of traffic flow. To extract vehicle trajectory based on a video, detecting and tracking vehicles is necessary[1].

The traditional vehicle detection and tracking methods are mainly divided into two categories: motion-based and feature based. The motion-based method is mainly realized by the frame difference method, background modeling, optical flow method[2]. Feature-based methods mainly detect vehicles by designing features such as artificial intelligence. This paper uses the background difference method based on symmetrical difference block modeling to detect vehicles in UAV video. A vehicle detection and tracking method based on multi-scale edge fusion and intelligent feature matching is proposed to overcome the noise and background interference of traditional edge feature detection methods. In this paper, the improved intelligent optical flow method is used to track vehicle feature points[3]. Then the same vehicle feature points are clustered and matched through the Gaussian mixture model and spectral clustering, and the mismatched tracks will be filtered. This paper uses an adaptive immune particle filter vehicle tracking algorithm. According to the characteristics of vehicle driving in the actual scene, adaptive

learning parameters are added to the corresponding part of the algorithm to ensure robustness of the algorithm[4]. In recent years, with the development of artificial intelligence technologies such as deep learning and computer vision, deep learning models have achieved high accuracy in the field of target detection and recognition by virtue of their powerful ability of feature extraction and abstract expression. In this paper, based on the depth learning method, the roadside video is used as the input to achieve fast and high-precision vehicle tracking. In this paper, we use an aggregate channel feature neural network based on human characteristics to extract vehicle trajectories under traffic congestion. Although the above Research has realized vehicle detection and tracking, most of them have limitations such as low extraction accuracy, weak track continuity, and can not cope with complex road environments and severe weather conditions. In this paper[5], A new method to extract vehicle trajectory, which mainly includes three steps: vehicle detection, vehicle tracking, and trajectory post-processing. The first step is to use an intelligent detector to automatically detect the vehicles in the region of interest on each lane in the video frame. After obtaining the vehicle position, the second step is to use the advanced tracker to achieve fast and accurate tracking. In the last step, the perspective transformation matrix is used to map the vehicle coordinate position from the pixel coordinate system to the world coordinate system, and then the locally weighted regression algorithm is used to eliminate the outliers and zigzag oscillations of the original acquisition track[6].

## 2. Essence and Challenge of Automatic Driving Trajectory Prediction

### 2.1. Track Prediction

In the actual model, it is mainly abstracted into two forms. One is to predict the dynamic model of the target at the next time according to the dynamic parameters of the target at this time (mainly the size and direction of the speed) and calculate the coordinate position on this basis. This modeling method is more common in traditional models. The other abstracts the trajectory as a set of coordinate points. At this time, the input

of the model is the trajectory coordinates of the target in the scene in the past period, and the output is the trajectory coordinates of the future period inferred by the model. In-depth learning model, the trajectory is often regarded as a sequence of coordinate points. In deep learning, such tasks, like natural language processing, belong to typical sequence processing problems, so we can usually see similar structures such as Long Short Term Memory (LSTM), Transformer[7].

The modeling of the trajectory prediction problem is not complex, but the diversity of factors affecting trajectory trend leads to the complexity of the problem, which can be divided into the following categories.

## 2.2. Intelligent Driving System for Trajectory Prediction

(1) Scenario factors. In different traffic scenarios, pedestrians and vehicles will be significantly affected by the specific environment. First, the objects of the scene

For example, pedestrians cannot pass through walls, and vehicles cannot drive through the center of a circle in a roundabout, which requires the algorithm to be able to identify obstacles. In addition, the scene will also have an implicit impact. For example, pedestrians crossing the street at the intersection will have a faster speed than students walking on the campus, and vehicles will be limited by signs and traffic rules, which also requires the algorithm to understand the type of scene.

(2) Individual differences. Due to individual differences in personality, age, mental state, and other aspects, their marching strategies will also be significantly different. For example, a pedestrian who is impatient or in a hurry may adopt a more aggressive strategy when avoiding obstacles; The young are often faster than the old; Some people's comfortable social distance is considerable, which leads to their trajectory always being far away from others.

(3) Complex interactions. The travel strategies of

pedestrians and vehicles are not independent of each other but have very complex interactions. Traffic participants in the same scene will react to their own strategies and adjust their own strategies. For example, three or five friends will walk together, and the driver will brake gently when there is a vehicle changing lanes ahead. To judge the interaction strategy of each traffic participant, its purpose and intention are crucial, which is entirely implicit in the model, so higher requirements are put forward for the algorithm. Therefore, the model should be able to infer the intentions of each traffic participant and, on this basis, understand the relationship between each traffic participant, which is one of the central issues of trajectory prediction.

## 3. Establishment of Automatic Driving Vehicle Cut in The Data Set

### 3.1. Automatic Driving Data

The data set used in this paper is the natural driving data collected by the "China Automotive Research Institute Natural Driving Data Research (CAERI-NDS)" project. The project has carried out large-scale driving data collection in many provinces and cities in China. Up to now, 100000 km of natural driving data have been collected by this technology. The data acquisition equipment configured for the test vehicle includes cameras, millimeter-wave radars, GPS, and other sensors, which can obtain self-driving data, surrounding target vehicle information, etc. The sampling frequency is 25Hz. The driving data of the self-owned vehicle is obtained by the CAN bus of the vehicle, including the speed, acceleration, steering wheel angle, etc. of the self-owned vehicle; The information of surrounding target vehicles is collected by cameras, radars, etc., including the speed, acceleration, and distance between the target vehicle and its own vehicle, See Figure 1.

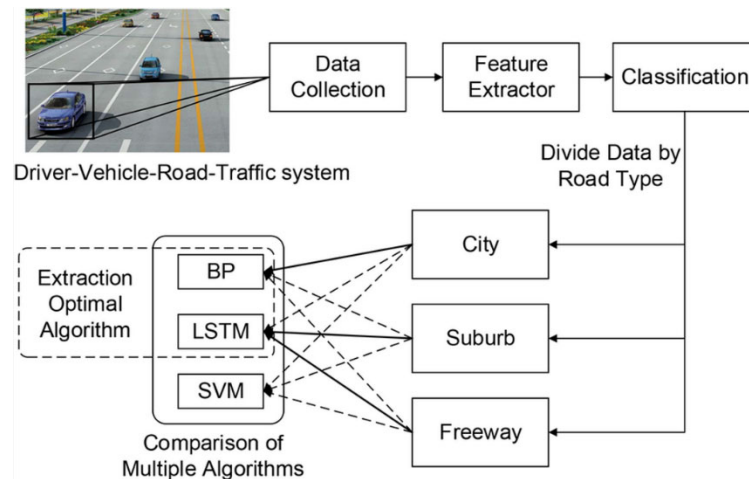


Figure 1. The model of automatic driving trajectory prediction system

## 3.2. Data Preprocessing

In order to reduce the subsequent calculation cost, the test data is first de-sampled, and the sampling frequency is reduced by one time. Driving data will inevitably be affected by some particular factors in the process of collection, such as light environment conditions and equipment conditions. These factors may lead to incomplete or inaccurate data collection, resulting in unpredictable consequences for the test. Therefore, before using the data, we must process the

missing values and noises in the data set. Due to the time characteristics of vehicle trajectory, the missing values are filled by combining the nearest neighbor filling method and the interpolation method. The noise in the driving data set is filtered. The algorithm in this paper is a method based on polynomial least square fitting in the time domain, which can ensure the shape and width of the signal are unchanged while filtering and is widely used in data smoothing and denoising. In this paper, the Savitzky Golay filter is used to filter the

noise of natural driving data. The expression is:

$$y(i) = \sum_{k=0}^p a_k t^k \quad (1)$$

$$E = \sum_{i=-M}^M [y(i) - x(i)]^2 \quad (2)$$

$$\frac{\partial E}{\partial a_k} = 0 \quad (3)$$

$Y_{(i)}$  is the processed data,  $i = -M, 0, M$ , is the total error square sum;  $X_{(i)}$  is unprocessed data,  $p < 2m$

### 3.3. Extraction of Vehicle Entry Scenarios

The timing and speed of vehicle entry are affected by the vehicle (primary vehicle) in the target lane. The vehicle cut-in scenario mainly includes a primary vehicle on the main lane and a target vehicle on the adjacent lane that wants to cut into the main lane. It shows the schematic diagram of a vehicle cut in a scenario under a human-machine mixed driving environment. Yellow AV represents driverless vehicles on the main lane, and blue HV represents manually driven vehicles that may cut into the main lane; At this time, the uncertainty of the future trajectory of the blue HV may cause the yellow AV to generate wrong decisions, resulting in vehicle collisions. Based on the cut-in scenario defined above, extract qualified cut-in scenario segment-related data from the natural driving dataset. First, define the coordinate system to determine the position of the entry target vehicle and the primary vehicle. Take the centerline of the lane where the leading vehicle is located as the reference, and take the direction perpendicular to the longitudinal axis  $x$  as the horizontal axis  $y$ . Make the vertical line perpendicular to the lane centerline through the track point of the primary vehicle in the first time step in the history observation sequence, take the vertical point as the origin of the coordinate system, and take the direction along which the vehicle moves forward as the positive direction of the longitudinal axis. Take the left side facing the forward direction of the vehicle as the positive direction of the transverse axis. The coordinates of the leading vehicle can be determined by collecting data, and then the coordinate position of the target cutting in the vehicle can be determined according to the relative distance between the primary vehicle and the target vehicle.

Deo et al. proposed that vehicle lane change includes the preparation stage and stability stage. If the vehicle is at the lane change point at time  $t$ , define the track on time interval  $t \pm 4s$  as lane change state; Inspired by this, it is set that if the last time node of the extracted track clip falls within the time interval  $t \pm 3.2s$ , the track clip is recorded as the cut in scene clip. Each 7.2s cut in track segment is selected as a training

sample, in which the first 4s data is used as historical track data, and the last 3.2s data is used as future track data. One time step of the sample is 0.08s. According to the established coordinate system, extract the cut-in scene clips. The extraction criteria are as follows.

(1) Calculate the vehicle displacement in each time step, and round off the vehicle track with displacement more significant than 3m in one time step; Since the Research is to cut into the scene and focus on the lateral displacement of the vehicle, the vehicle track with a lateral displacement more significant than 0.2m in a time step is discarded.

(2) To predict the future trajectory according to the historical trajectory information, the trajectory clip information of sufficient length is required. The trajectory clip with a duration of less than 10s is omitted here.

(3) Find the intersection point of the cut-in vehicle track and lane line and determine it as the cut-in point. Then, sample 40 time steps along the time axis direction and 129-time steps along the opposite direction of the time axis (at least 50-time steps must be met to meet the definition of the cut in scene segment), and determine the vehicle track in this period as a cut in scene segment.

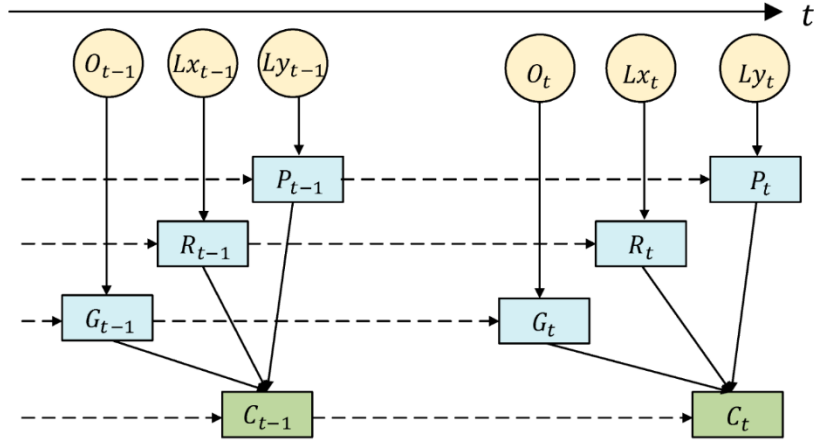
## 4. Cut in A Trajectory Prediction Model

### 4.1. Input Characteristic Vehicle

The selection of input characteristics of cut in trajectory prediction model is crucial to the accuracy of prediction. For the short-term future trajectory prediction of the target vehicle, its historical trajectory, especially the part near the current time point, has more effective information characteristics. On the other hand, the interaction between vehicles and the behavior of surrounding vehicles will also affect the future trajectory of the target vehicle. Therefore, the selected input feature contains the following two parts.

(1) Information of target vehicle: longitudinal coordinate, transverse coordinate, speed and acceleration of target vehicle, and direction angle and direction angular velocity of vehicle relative to the self vehicle.

(2) Interaction information: relative longitudinal distance, relative transverse distance, relative longitudinal velocity, relative transverse velocity, and relative longitudinal acceleration between the target vehicle and the owner vehicle. Therefore, the information of the input feature can be expressed Information of the target vehicle; Is the information about the interaction between the target vehicle and its vehicle The longitudinal and transverse coordinates of the target vehicle; Is the speed of the target vehicle; Is the acceleration of the target vehicle, see Fig. 2.



**Figure 2.** Research on automatic driving vehicle cut in data set algorithm system

And are respectively the direction angle and direction angular velocity of the target vehicle relative to the self vehicle; And are the relative longitudinal distance and transverse distance between the target vehicle and the owner vehicle respectively; And are the relative longitudinal speed and relative transverse speed of the target vehicle and the owner vehicle respectively; Is the relative longitudinal acceleration of the target vehicle and the owner vehicle. 3 Verification and Analysis 3.1 Test Setup This paper uses average displacement error (ADE) and final displacement error (FDE) to evaluate the trajectory prediction model. ADE represents the average Euclidean distance (MED) between all predicted positions and the actual position in the prediction window, and FDE represents the average Euclidean distance between the final predicted position at the end of the prediction track and the corresponding absolute position. The smaller the values of ADE and FDE, the better the prediction effect. The formula is:

$$ADE = \frac{1}{y} \sum_{i=1}^N \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2} \quad (4)$$

$$FDE = \sqrt{(x_e - \hat{x}_e)^2 + (y_e - \hat{y}_e)^2} \quad (5)$$

Where:  $N$  is the total amount of samples;  $X_i$  and  $Y_i$  are the actual values of the vertical and horizontal coordinates, respectively;  $\hat{x}$ ,  $\hat{y}$  is the predicted value of the longitudinal coordinate and transverse coordinate, respectively; The subscript  $e$  represents the final predicted position. In order to verify the effect of the prediction model proposed in this paper, four kinds of networks are used to compare the performance of each model.

(1) LSTM: Long and short-term memory networks are used for trajectory prediction.

(2) Gated current unit (GRU): A variant of RNN is proposed to solve long-term memory problems and gradient in backpropagation. Structurally, GRU has only two gates, and its performance is similar to LSTM. GRU has fewer parameters and lower calculation costs.

(3) Bi LSTM: bidirectional LSTM network.

(4) Bi GRU: bidirectional GRU network.

## 4.2. NGSIM Dataset Verification

In order to test the robustness of the proposed model, this paper also uses the vehicle trajectory data of the NGSIM project of the data set for comparative testing. Aerial cameras

collect the data of the detected road section with a sampling frequency of 10Hz. The data set provides the vehicle speed, acceleration, location coordinates, lane information, etc. The lane change track and lane keeping segment are extracted from the data set to generate the test data set, divided into a training set, verification set, and test set according to the ratio of 8:1:1.

**Table 3.** Evaluation Indicators of Each Model

Model	evaluation index	
GRU	2.236	2.486
LSTM	2.707	3.158
Bi-GRU	1.814	2.884
Bi-LSTM	1.741	2.929

To sum up, the improved Bi LSTM proposed in this paper can better adapt to the impact of data set differences and has advantages and robustness in NGSIM data sets.

## 5. Conclusion

In this paper, the extraction criteria of vehicle cut-in scenes are designed. The eligible vehicle cut-in scene fragments are extracted from the natural driving data. The sliding time window method filters the above-extracted cut-in-scene fragments and generates a data set, which is used to train and test the proposed model. Considering the influence of self-driving vehicles on the future state of the front entry vehicle, combining the advantages of Bi LSTM network that is good at learning and using context information and the ability of in out fast connection to effectively reduce gradient disappearance and network degradation, a modified Bi LSTM model based on deep learning is proposed to predict the entry trajectory of vehicles in a man-machine hybrid driving environment. Through the comparison and analysis of the test results, it can be proved that the proposed method has greatly improved the prediction effect of the cut in vehicle trajectory and has a high application value.

## Acknowledgment

This paper is sponsored by Scientific research project of Hubei Provincial Department of Education (Project No. 17G046).

## References

- [1] Zhang Xinyu, Gao Hongbo, Zhao Jianhui, et al. Overview of autonomous driving technology based on deep learning Journal of Tsinghua University, Vol. 16 (2020) No. 06, p.312-316.
- [2] Leo. Target detection using deep learning China Equipment Engineering, Vol. 14 (2021) No. 10, p.98-99.
- [3] Ren Bohan. Image Segmentation Using Deep Learning. China Equipment Engineering, Vol. 12 (2022) No. 14, p.124-126.
- [4] Liu Zhiqiang, Zhang Zhongyun, Ni Jie, et al. Research on multisensor fusion vehicle detection method under dual dynamic conditions. Mechanical Design and Manufacturing, Vol. 18 (2020) No. 16, p.85-86.
- [5] Wang Kejun, Zhao Yandong, Xing Xianglei. Research progress in the application of deep learning in the field of driverless vehicles Journal of Intelligent Systems, Vol. 10 (2021) No. 10, p.136-138.
- [6] Huang Liangji, Tang Tao. Analysis and design of auto drive system for metro trains Journal of Beijing Jiaotong University, Vol. 19 (2021) No. 12, p.26-29.
- [7] Liu Jun, Yuan Jun, Cai Junyu, et al. Agricultural machinery auto drive system based on GPS / INS and steering by wire Journal of Agricultural Engineering, Vol. 21 (2020) No. 08, p.112-115.