A Method for Extracting Traffic Parameters from Drone Videos to Assist Car-Following Modeling

Xiangzhou Zhang, Zhongke Shi
School of Automation, Northwestern Polytechnical University, Xi'an, China

***Abstract*—A new method for extracting traffic parameters from UAV videos to assist in establishing a car-following model is proposed in this paper. The improved ShuffleNet network and GSConv module were introduced into the Yolov7-tiny neural network model as the target detection stage. HOG features and IOU motion metrics are introduced into the DeepSort multi-object tracking algorithm as the tracking matching stage. By building a self-built UAV aerial traffic data set, experiments were conducted to prove that the new method improved a few detection and tracking indicators. In addition, it improves the false detection, missed detection, wrong ID conversion and other phenomena of the previous algorithm, and improves the accuracy and lightweight of multi-target tracking. Finally, gray correlation was applied to analyze the traffic parameters extracted by the new method, and the driver's visual perception of collision was introduced into the car-following model. Through stability analysis, small disturbance simulation and collision risk assessment, the newly proposed traffic flow parameter extraction method has been proven to improve the dynamic characteristics and safety of the car-following model, and can be used to alleviate traffic congestion and improve driving safety.**

Keywords*—UAV; Yolov7-tiny; DeepSor; Car-following model; Stability analysis; Traffic congestion; safety assessment

I. INTRODUCTION

Traffic congestion leads to low transportation efficiency and air pollution. Frequent traffic accidents lead to casualties and economic losses. These problems are all important challenges facing the development of modern transportation. Researchers have developed numerous models based on traffic flow theory to explain traffic phenomena, thereby improving traffic efficiency and driving safety. Furthermore, the collection of data required for modeling has always been the basis and hot spot of research. UAV aerial video data is extremely informative both in content and time. The UAV is rapidly popularized due to its lightweight, easy operation, and low cost, making them increasingly important in the field of target detection and tracking [1-3]. It is often used in traffic law enforcement and monitoring in various countries. However, how to use drone aerial photography to assist in building a driver behavior model still needs to be explored.

The key to the application of UAV aerial photography data collection lies in the video vehicle detection and tracking algorithm, which extracts its speed, trajectory and other information through vehicle position information at different times. With the rapid development of deep learning technology in recent years, in terms of vehicle target detection, researchers have proposed a variety of improved target detection neural networks for different scenarios and tasks. Among them, the yolo series of single-stage multi-target detection algorithms is widely used due to its obvious advantages. Makarov et al. [4] used the yolo V2 network to realize the recognition of cars, large vehicles and other objects from the UAV perspective. Hoslain [5] and others migrated YOLO V3 and SSD to the edge-side onboard GPU Jetson TX2. Jetson Xavier implemented UAV detection of vehicles and provided accurate target locations and vehicle types. In addition, the problems introduced by the drone aerial photography perspective have been optimized. For example, to address the problem of an increase in small targets caused by UAVs. Zhang et al. [6] inserted three Spatial Pyramid Pooling modules between the fifth and sixth convolutional layers in front of the three detection heads of the YOLO V3 network to design the silm-yolo V3-SPP3 network. In order to solve the problem of low detection efficiency caused by the sparse and uneven distribution of target categories from the perspective of a drone. Li et al. [7] proposed DS YOLO V3, which added multiple detection heads connected to different layers of the backbone network to detect targets of different sizes. In addition, a multi-scale channel attention fusion module is designed to utilize complementary channel information.

In terms of UAV aerial photography target tracking, commonly used methods based on target trajectory mainly include Karman filtering and deepsort framework. Luo et al. [8] used yolov5 for feature extraction, Kalman filter to extract target motion information and update predictions, and Hungarian matching algorithm to obtain tracking results. Khalkhali et al. [9] proposed SAIKF (Situation Assessment Interactive Kalman Filter), which uses situation assessment information extracted from the traffic history of the same environment to improve tracking performance. The target trajectory prediction based on deepsort is as follows. Ning et al. [10] used yolov5 to obtain the real-time position of the target, and combined with the deepsort framework to achieve the speed measurement of the target. In addition to the above applications, many scholars have made various corresponding improvements to address the problems that arise in multi-target tracking from the UAV perspective. Du et al. [11] used OSNet to replace the simple feature extractor in Deep-SORT, used global clues to associate it with the trajectory, and proposed the EMA (Exponential Moving Average) strategy to achieve a more accurate association between small trajectories and detection results. Huang et al. [12] generated target bounding boxes through different prediction networks, performed cascade matching on all trajectories and detection results, performed unmatched tracking and detection through GIOU matching, and generated the final trajectory.

By establishing a mathematical model of the relative motion relationship between vehicles, car-following models are often suitable for the development of autonomous driving systems and traffic flow simulations. Therefore, it has been a hotspot in traffic flow research, and scholars have achieved rich achievements. Zhang [13] proposed a bi-directional visual angle car-following model considering collision sensitivity, which improved the dynamic characteristics and driving safety of car-following and lane-changing in the traffic flow. Zhang [14] proposed a small-radius curve following model that considers the driver's desired visual angle based on the impact of two-point preview steering decisions and parking sight distance on small-radius curve following behavior from the perspective of the driver's visual characteristics. Liu et al. [15] optimized traffic at signalized intersections by incorporating short-term driving memory on driver behavior. Ma et al. [16] introduced memory effect of headway changes into the driver behavior. Simulations demonstrate it has a significant effect on alleviating traffic congestion.

To sum up, the target detection and tracking technology of drone aerial videos is relatively mature. However, there are still some limitations in using UAV aerial photography to collect information to assist in driving behavior modeling: (1) Due to the size limitations of UAV equipment, there is still room for improvement in high-speed and high-precision traffic information collection. (2) Vehicle data that can quantify the driver's physiological and psychological behavior from a microscopic perspective is difficult to extract. (3) It is difficult to use traffic survey data to model driving behavior while optimizing the dynamic characteristics and safety of traffic flow. This study proposes a new method of assisted modeling of UAV aerial car-following images to solve the above problems. The method we proposed for the above problems has the following advantages: (1) The improved lightweight video detection network and improved target tracking method are conducive to improving the accuracy and speed of traffic information collection in UAV aerial videos. (2) A method that can collect the driver's psychological following behavior is proposed. (3) A method is proposed to use collected traffic data to model driving behavior to simultaneously improve dynamic characteristics and safety.

The remaining parts of this study are structured as follows: Our proposed method for extracting traffic parameters from UAV videos is described in detail in Section II. The improvements of the new traffic parameter extraction method in target detection, tracking and car-following behavior modeling are verified through experiments. The experiment results are described in the Section III. The collected traffic data is used to model the driver's psychological car-following behavior, and its dynamic characteristics and safety are improved in the Section IV. The work of this research is summarized and future work is prospected in Section V.

II. Matter And Methods

Fig. 1 shows a new method of extracting traffic parameters from UAV video to assist in establishing a car-following model. The method is divided into four stages: image processing, target detection, target tracking and traffic flow modeling. We introduce them separately below.

A. Transformation of Coordinate Systems

The real traffic parameters are obtained by converting the image coordinates into real coordinates to calculate the traffic parameters. Establishing a transformation matrix between image coordinates and world coordinates is the basis for traffic parameter extraction. Therefore, one frame of image in each video is selected as the reference frame N_0 of the video; several marker points are marked on the reference frame, their image coordinates are recorded, and the world coordinates of the same marker points are obtained. The same two sequences in the two coordinate systems have a corresponding relationship. The conversion between the two sets of coordinate systems can be achieved through the perspective projection matrix T of the reference frame image coordinate system and the world coordinate system. The coordinate relationship is as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{T} \times \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} = \begin{bmatrix} A & D & G \\ B & E & H \\ 0 & 0 & I \end{bmatrix} \times \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \quad (1)$$

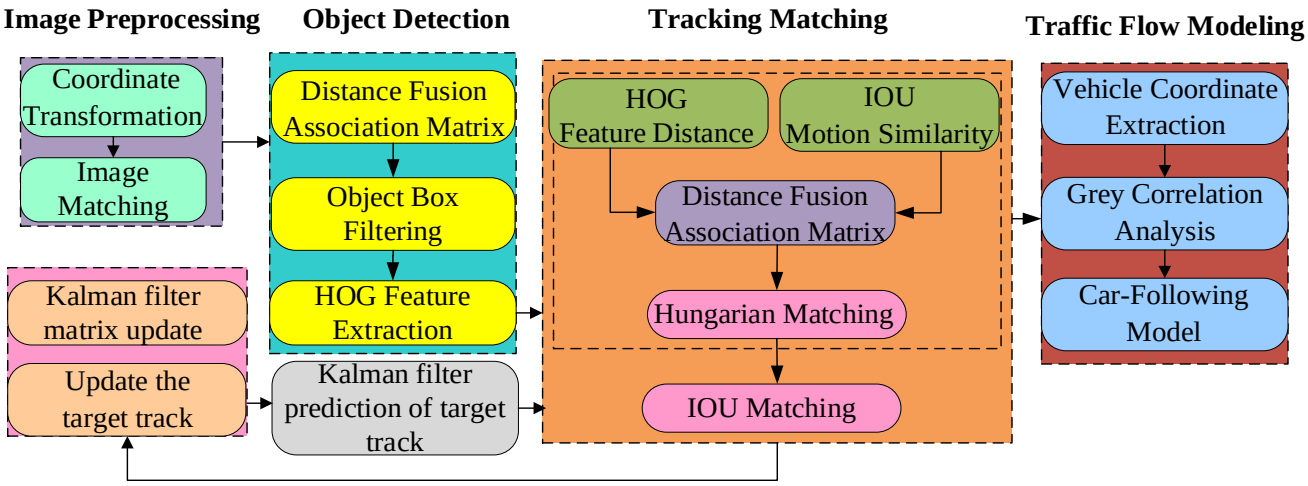


Fig. 1. UAV aerial images assist in establishing car-following models system.

where, $[u\ v\ 1]$ is the world coordinate, and $[x_0\ y_0\ 1]$ is the image coordinate of the reference frame. Since the image coordinates are two-dimensional coordinates, the coordinate system conversion problem is transformed into a single plane, so the elevation direction parameter in the three-dimensional coordinates is 1. Substituting the world coordinates and the image coordinates of the reference frame into the Eq. (1), the perspective projection matrix T can be obtained.

B. Image Matching

In actual scenes, the camera may be subject to external interference (such as breeze) and undergo slight displacements and changes in pitch angle. Especially for drones, it is impossible to be completely still during the process of collecting data, and the image coordinates of the same fixed point in different frames will change. Therefore, after completing the conversion of image coordinates and world coordinates, it is necessary to obtain the rotation displacement matrix w in the following formula to calibrate the correspondence between the n th frame and the reference frame w.

$$\begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix} = \boldsymbol{W} \times \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \tag{2}$$

where, (x_n, y_n) is the image coordinate of the nth frame; (x_0, y_0) is the image coordinate of the reference frame N_0,

C. Improved Target Detection Algorithm

In the target detection stage, the improved yolov7-tiny target detection model is first used to identify the input image of each frame of the video, and the detecting bounding box is obtained by screening. Then, the detecting bounding box is extracted through the convolutional neural network model to obtain depth features and manual HOG features. These two features are called the appearance features of the detecting bounding box. Finally, the appearance characteristics of the detecting bounding box and the position information characterizing its motion characteristics are input into the tracking and matching stage. This article considers using a more lightweight feature aggregation scheme, which requires fewer parameters and less calculations while ensuring rich features. Therefore, an improved yolov7-tiny target detection model is proposed. First, we use the idea of ShuffleNet, a lightweight network for image classification, to improve Backbone to reduce dense connections and increase network depth. Reducing dense connections can reduce the overall calculation amount, while appropriately increasing the network depth can obtain richer features. Secondly, in the Neck part of the model, the lightweight module GSConv is used for feature aggregation and ELAN is improved to further reduce the amount of model parameters, calculation amount and size while ensuring that rich features are not lost. In this way, the improved YOLOv7-tiny network is shown in Fig. 2.

1) Improved shuffleNet network: In the two basic modules of the ShuffleNet [17] network, after the input features enter the right branch, Grouped Convolution (GConv) is first performed, then channel shuffle is performed, and then Depthwise Convolution (DW Conv) with a convolution kernel size of 3 is performed. Finally, perform a Grouped Convolution. The entire process can significantly reduce the parameters of the network, but the resulting feature map will lose semantic information and also cause some loss in accuracy.

Based on the above analysis, the right branches of the two basic modules in Fig. 3(a) are improved. First, depthwise separable convolutions are replaced by group convolution modules. Through grouping, on the premise of increasing the amount of parameters, a certain amount of information can be exchanged between each channel of the feature map; secondly, the channel shuffling operation is changed to a standard convolution with a convolution kernel size of 1, and places it at the end of the branch. Standard convolution operations can play the same role, while also further enriching the semantic information of the feature map without adding parameters.

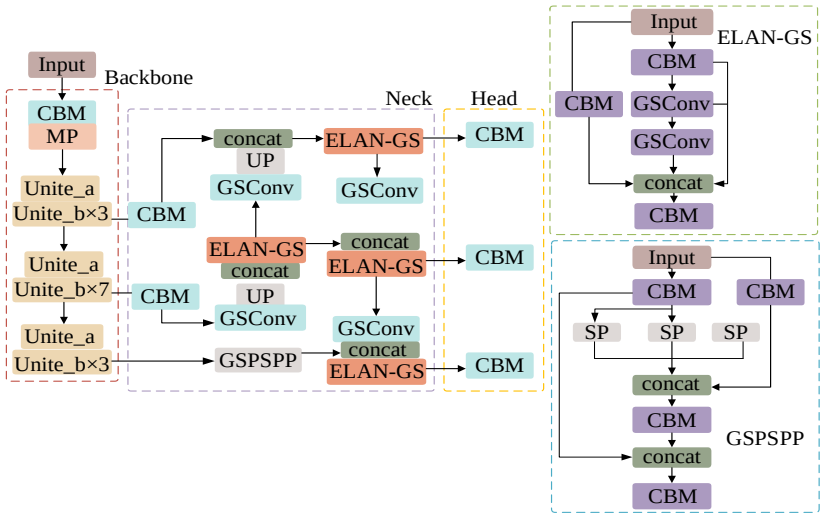


Fig. 2. Improved yolov7-tiny.

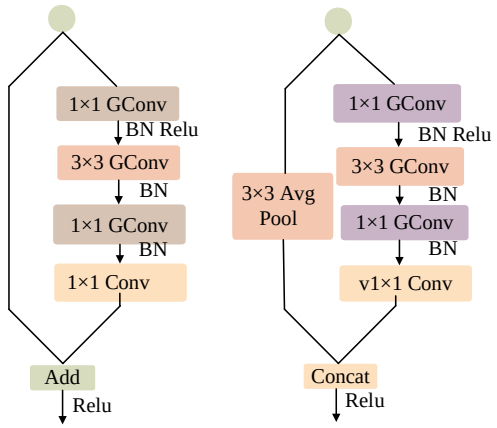


Fig. 3. Improved shufflenet.

2) GSConv module: The GSConv[18] structure is shown in Fig. 4. The input and output channel numbers are C1 and C2 respectively. First, the number of channels is reduced to C2/2 due to a standard convolution, and then it goes through a depth-separable convolution, and the number of channels remains unchanged. The channel information is evenly disrupted by shuffling, and the extracted semantic information is enhanced. Then the expressive ability of image features is improved and enhanced with the fusion of feature information.

When the network performs feature fusion at the Neck layer, semantic information will also be continuously transmitted downward. When the height, width and number of channels of the feature map are continuously compressed and expanded, the loss of part of the semantic information will affect the final prediction. This article introduces the GSConv module into the Neck layer of the network, using the GSConv module instead of standard convolution for upsampling and downsampling, reducing the amount of parameters and calculations of the model, and ensuring the sampling effect to the greatest extent. In addition, GSConv is also introduced into the ELAN module for improvement as shown in Fig. 4. The two convolutions before the Concat layer use the GSConv module to reduce the number of parameters of the model while ensuring detection accuracy and "slimming down" the ELAN module.

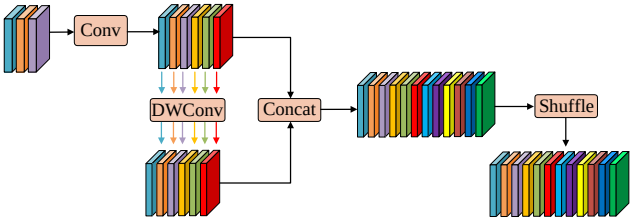


Fig. 4. GSConv module.

D. Improved Target Tracking Algorithm

The tracking matching stage obtains motion features and appearance features based on the target detection stage, and measures the similarity with the predicted motion features and appearance features of the existing tracked target. Then the correlation matrix is fused and cascade matching is performed.

The detecting bounding box and tracking target that failed to complete the matching are matched again by IOU. Through two matching processes, the pairing of the current frame detecting bounding box and the existing tracking target is completed. Finally, the status flag of the tracking target in the tracker is updated, and the update and prediction of the Kalman filter are completed.

1) Motion feature measurement based on IOU: In order to better measure the motion characteristics in multi-target tracking, this paper proposes to use the IOU measurement of the detecting bounding box and the tracking target prediction frame to replace the Mahalanobis distance measurement method in the DeepSORT algorithm.

In DeepSort, the degree of matching between the detecting bounding box and the motion characteristics of the previously tracked target's predicted position is characterized by the Mahalanobis distance between the two. The Mahalanobis distance $d'(i,j)$ between the tracking target i and the detection target j is expressed as,

$$d'(i,j) = \left(x_j - y_i\right)^{\mathrm{T}} S_i^{-1} \left(x_j - y_i\right) \tag{3}$$

where, x_j and y_i respectively are the observed quantities for detecting target j and the predicted quantities for tracking the motion of target i, and S_i represents the state covariance matrix of the Kalman filter. The larger the Mahalanobis distance, the greater the difference in motion characteristics between the two. False correlations can be excluded by setting a threshold for the Mahalanobis distance.

However, this Mahalanobis distance measure only uses the distance relationship between the detecting bounding box and the tracking target, and cannot accurately describe the motion information of the two. When two targets with similar appearance characteristics are close to each other, it is easy to cause ID switching problems. At the same time, when the tracking target is After occlusion for a period, the Mahalanobis distance is affected by the increased uncertainty of Kalman filter prediction, making it difficult to reliably measure the motion state. The motion feature measurement based on IOU can more accurately describe the positional relationship between the detecting bounding box and the tracking target. The formula is:

$$IOU(i,j) = \frac{Area\left(P_i \cap Q_j\right)}{Area\left(P_i \cup Q_j\right)} \tag{4}$$

The value range of $IOU(i,j)$ is [0,1], and the IOU distance between the tracking target i and the detection target j is

$$d_{\mathrm{IOU}}(i,j) = 1 - IOU(i,j) \tag{5}$$

Through the IOU measurement method, the difference in motion features between the detecting bounding box and the tracking target can be more accurately measured, and the

motion feature value threshold matrix expression can be obtained:

$$b_{\mathrm{IOU}}(i,j) = I\left[d_{\mathrm{IOU}}(i,j),, t_{\mathrm{IOU}}\right] \quad (6)$$

In the formula, I is the indicator function, which takes 1 when the conditions are met, and 0 otherwise. After experiments, t_{IOU} is set to 0.9, which can eliminate most of the erroneous correlations between the detecting bounding box and the tracking target prediction frame, improve tracking accuracy, and reduce the number of ID switching times.

2) Appearance feature measurement based on HOG feature: In order to more accurately associate the matching detecting bounding box and the tracking target in the tracking and matching stage, this paper proposes to fuse the HOG feature distance in the cascade matching view feature measurement. The HOG feature is characterized by statistics and calculation of the gradient direction histogram of the local area of the image. It is often used to characterize the edge information of objects and is widely used in image recognition. It can maintain good invariance to geometric deformation and illumination changes of the detection frame. Moreover, extracting the HOG features of the calculated image requires a small amount of calculation, fast operation, and has little impact on speed performance. By introducing HOG feature distance fusion, it can better reflect the shallow image features of the target and improve the accuracy and robustness of appearance similarity measurement.

In the target detection stage, the size of the obtained detection frame is adjusted to 256x128, the pre-trained ResNet-18 network is input, and the 512-dimensional depth feature D_j is output. Then the HOG features of the detection frame are extracted and the 8505-dimensional HOG feature H_j is output.

The HOG feature distance uses the minimum cosine distance as the measurement criterion, and are only calculated for the features in the N frames closest to the tracking target i. They can be expressed by the following formulas respectively.

$$d_{\mathrm{HOG}}(i,j) = \min\left\{1 - H_j^{\mathrm{T}} H_k^i \mid H_k^i \in R_{\mathrm{HOG}}^i\right\}, R_{\mathrm{HOG}}^i = \left\{H_k^i\right\}_{k=1}^{N} \quad (7)$$

$$R_{\mathrm{HOG}}^i = \left\{H_k^i\right\}_{k=1}^{N} \quad (8)$$

Similarly, the appearance feature distance measurement also has a threshold t_F to avoid false matching. Its expression is:

$$b_F(i,j) = I\left[d_F(i,j),, t_F\right] \quad (9)$$

where, t_F is usually set to 0.2. Combining the two thresholds of Eq. (5) and Eq. (7), the threshold function $b(i,j)$ is obtained, and its formula is as follows:

$$b(i,j) = b_{\mathrm{IOU}}(i,j) \circ b_F(i,j) \quad (10)$$

In the formula, $\circ$ represents the Hadamard product of the matrix. The fusion coupling matrix $C_{i,j}$ can be obtained from the fusion appearance distance, and the expression is:

$$C_{i,j} = d_F(i,j) \circ b(i,j) \quad (11)$$

The fusion correlation matrix is matched using the Hungarian matching rule to obtain the correct correspondence between the tracking target and the detecting bounding box.

E. Gray Correlation Calculation

We use gray correlation to analyze the correlation between the driver's visual factors and acceleration decisions during car-following. Gray relational analysis is well suited for studies of small sample size data sets and fleeting microscopic driving behaviors. The calculation steps are as follows.

1) Determine the analysis sequence: We select the acceleration of the following vehicle as the reference sequence, $A = \{A(j) \mid j = 1,2\cdots,n\}$; The newly introduced visual angle related parameters are selected as the comparison sequence, $\theta_i = \left\{\theta_i(j) \mid j = 1,2,\cdots,n\right\}$.

2) Dimensionless variables

$$x_i(j) = X_i(j) \Big/ \sqrt{\sum_{i=1}^{m} X_i^2(j)}, j = 1,2,\ldots,n; i = 0,1,2,\ldots,m \quad (12)$$

3) Calculate gray correlation coefficient

The gray correlation coefficient between $a_i(j)$ and $\theta_i(j)$ is

$$\xi_i(j) = \frac{\min\limits_i \min\limits_j \left|a(j) - \theta_i(j)\right| + \rho \max\limits_i \max\limits_j \left|a(j) - \theta_i(j)\right|}{\left|a(j) - \theta_i(j)\right| + \rho \max\limits_i \max\limits_j \left|a(j) - \theta_i(j)\right|} \quad (13)$$

Let $\Delta_i(j) = \left|y(j) - x_i(j)\right|$, the following can be obtained.

$$\xi_i(j) = \frac{\min\limits_i \min\limits_j \Delta_i(j) + \rho \max\limits_i \max\limits_j \Delta_i(j)}{\Delta_i(j) + \rho \max\limits_i \max\limits_j \Delta_i(j)} \quad (14)$$

where, $\rho \in (0,\infty)$ is the resolution coefficient. Generally, the value range of ρ is $(0,1)$. Here we set $\rho = 0.5$.

4) Calculate gray correlation

$$r_i = \sum_{j=1}^{n} \xi_i(j) \Big/ n, j = 1,2,\cdots,n \quad (15)$$

III. Experimental Results and Discussions

This experiment collected and annotated a video data set containing 12 groups of traffic scenes named UVACAR-MOT. The video data is 30 frames/s, and one frame is extracted every three frames to form a new video sequence. We selected eight groups as training sets and four groups as validation sets. The experimental setup employs the AutoDL cloud computing platform with a Linux system, 32GB of memory, and PyTorch 1.7 for deep learning. The graphics card is NVDIA Quadro

V100 with 32G of video memory. In the target tracking experiment parameter settings, the initial frame is three frames; the maximum threshold distance for IOU matching is 0.7; and the maximum retained frame for lost tracking is 30. The number of frames for calculating appearance characteristics is N=100. In the target detection experiment parameter settings, the detecting bounding box confidence threshold of yolov7-tiny is 0.3, and the IOU threshold of NMS non-maximum filtering in yolov7-tiny is 0.5.

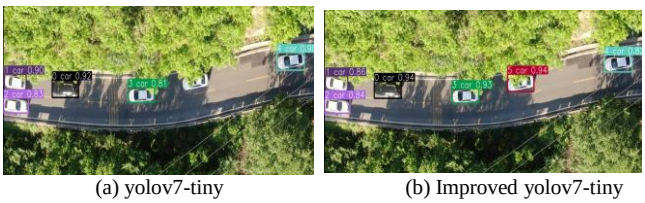
A. Target Detection

Table I shows that compared with the SSD algorithm and the yolov7-tiny algorithm, the improved yolov7-tiny algorithm has an increase in average accuracy and a decrease in parameter scale. The new algorithm has the characteristics of better detection accuracy, low parameters and low computational load. Compared with the mainstream yolo5s, it is more suitable for deployment on drones. Although the accuracy of the new algorithm is slightly lower than that of yolox-s, the parameters and model size are lower.

TABLE I. OTHER IMAGE DETECTION ALGORITHMS ON UVACAR-MOT

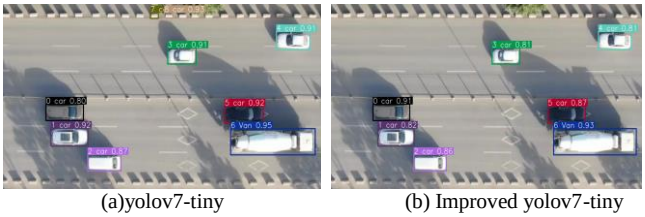
	mAP_0.5%	mAP_0.95/%	Params/10^6	Flops/10^9	Size/MB
SSD	49.1	29.2	40.3	371.2	267.6
YoloV7-tiny	54.1	36.2	5.8	14.3	11.9
Yolo5s	55.3	35.8	6.6	15.2	13.5
Yolox-s	58.3	39.9	8.9	25.7	67.5
ours	58.1	39.5	4.7	13.2	10.9

Fig. 5 and Fig. 6 respectively show the improvement effect of the improved yolov7-tiny on missed detections and false detections. Fig. 5(a) shows the missed detection of car no. 5 under occlusion for the yolov7-tiny algorithm. Fig. 5(b) shows that car no.5 is still detected under the improved yolov7-tiny algorithm. This shows that GSConv in the new algorithm enhances feature richness and makes detection more accurate. Fig. 6(a) shows that yolov7-tiny false detection curbstones as vehicles. The detection accuracy of the improved yolov7-tiny in the same detecting bounding box is shown in Fig. 6(b). This shows that the improved shuffleNet in the new algorithm effectively enhances the feature extraction capability and improves the accuracy of vehicle detection.



(a) yolov7-tiny (b) Improved yolov7-tiny

Fig. 5. Improved target detection algorithm improves missed detections.



(a)yolov7-tiny (b) Improved yolov7-tiny

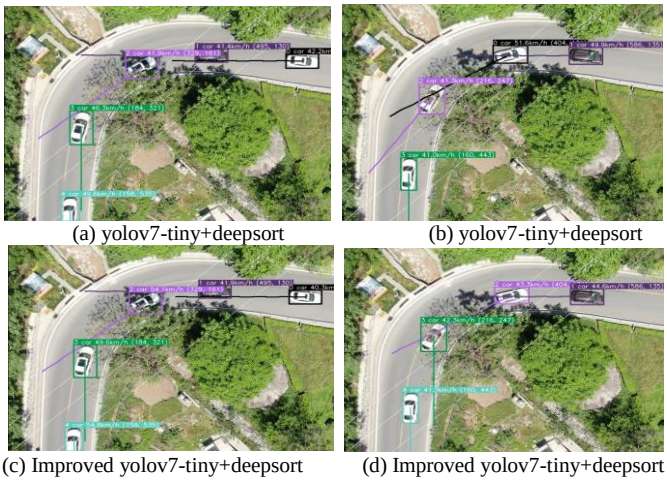
Fig. 6. Improved target detection algorithm improves false detections.

B. Target Tracking

Table II shows that most of the indicators of improved yolov7-tiny+deepsort in the UAVCAR-MOT data set are better than the deepsort tracking algorithm. The improvement of specific MOTA and MOTP indicators shows that the tracking accuracy has been greatly improved. IDsw is reduced to two times, which proves that the number of ID switching times for tracking the same target is very small and the tracking retention ability is strong.

TABLE II. OTHER IMAGE TRACKING ALGORITHMS ON UVACAR-MOT

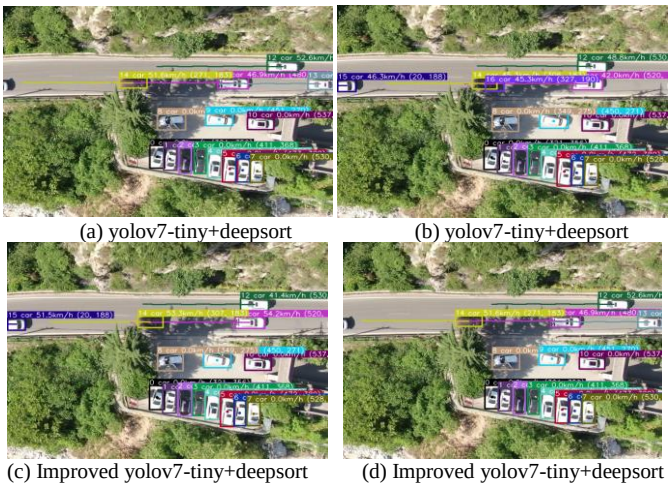
	MOTA/%	MOTP/%	IDsw	MT/%	FPS
EAMTT	51.88	73.81	45	71	16.77
POI	64.34	70.56	43	72	17.25
Sort	58.45	74.79	39	70	29.92
Deepsort	61.41	74.73	34	72	18.53
Ours	62.02	76.51	2	74	28.83



(a) yolov7-tiny+deepsort (b) yolov7-tiny+deepsort

(c) Improved yolov7-tiny+deepsort (d) Improved yolov7-tiny+deepsort

Fig. 7. Improved target tracking algorithm to improve ID hopping.



(a) yolov7-tiny+deepsort (b) yolov7-tiny+deepsort

(c) Improved yolov7-tiny+deepsort (d) Improved yolov7-tiny+deepsort

Fig. 8. Improved target tracking algorithm improves false detection of shadow.

Fig. 7 shows the different tracking effects of the improved front and rear tracking algorithms when blocked by trees. In Fig 7(a), the ID of the same vehicle in the two bounding box s

is switched from 3 to 2 when using the improved yolov7-tiny+deepsort algorithm. However, no such switching occurs after using the yolov7-tiny+deepsort algorithm in Fig. 7(b). This reflects that the improved yolov7-tiny+deepsort algorithm has better tracking capabilities. Fig. 8 shows the impact when light and shadow changes occur. Fig. 8(a) shows the red arrow vehicle being given two tracking IDs and boxes of different sizes. However, such an error did not occur in Fig. 8(b). This shows that the fusion of appearance feature measurement with HOG feature distance can improve the accuracy of appearance feature measurement and tracking accuracy.

C. Target Tracking

In the traffic phenomenon investigation experiment, we chose Jinye Road, Xi'an City, China, and its satellite map is shown in Fig. 9. The aerial photography scene is shown in Fig. 10.We chose clear and windless weather at 3 pm. Furthermore, there are more types and numbers of vehicles on this road, but there is no congestion. The image collection equipment uses a zoom drone, and the video quality is 1080p, 30fps.In order to avoid errors caused by image distortion in the later stage as much as possible, we hovered the drone at 90° directly above the road to shoot and kept the height at 145m. Table III shows one of our multiple sets of aerial traffic data. We selected the vehicle with ID 20 as the following car (width=1.6m) and extracted relevant image and driving data.

The correlation between the driver's visual factors and vehicle acceleration in Table IV calculated from Table III are all above 0.65. This shows that they are strongly related. In addition, we found that when the driver makes the acceleration and deceleration decision, the consideration of the possibility of collision is more important than the headway and velocity.

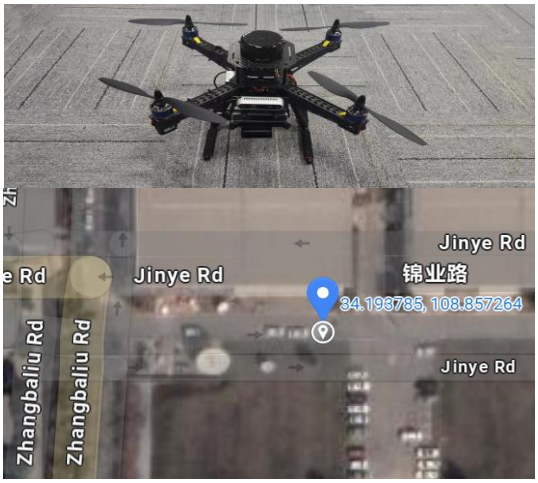


Fig. 9. Google maps for car-following.

Fig. 10. Target tracking for car-following.

TABLE III. AERIAL TRAFFIC DATASET

Pixel coordinates	$a_n(t)$	$v_n(t)$	$\theta_n(t)$	$\dot{\theta}_n(t)$	$\varphi_m(t)$	$\dot{\varphi}_m(t)$	$\dot{\theta}_m(t)/\theta_m(t)-\dot{\beta}_m(t)/\beta_m(t)$	$\dot{\varphi}_m(t)/\varphi_m(t)$
(723,351)	0.0500	12.1040	-0.0019	-0.0010	0.1294	0.0045	0.0348	-0.5498
(724,351)	0.1750	12.2500	-0.0021	-0.0012	0.1303	0.0062	0.0479	-0.5852
(726,352)	-0.0125	12.3850	-0.0024	-0.0015	0.1316	0.0084	0.0635	-0.6620
(728,353)	0.1000	12.5082	-0.0027	-0.0021	0.1332	0.0086	0.0644	-0.8065
(730,353)	-0.0875	12.4025	-0.0031	-0.0028	0.1350	0.0115	0.0854	-0.9518
(732,354)	-0.0250	12.1850	-0.0036	-0.0042	0.1373	0.0089	0.0652	-1.1941
(735,356)	-0.1000	11.7400	-0.0045	-0.0068	0.1391	0.0085	0.0614	-1.5910
(737,357)	-0.2000	11.3600	-0.0058	-0.0088	0.1408	0.0065	0.0459	-1.5991
(739,356)	-0.1750	10.9200	-0.0076	-0.0144	0.1421	0.0048	0.0335	-2.0249
(743,357)	-0.1875	10.6650	-0.0105	-0.1000	0.1430	0.0012	0.0081	-10.2236

TABLE IV. GRAY RELATIONAL DEGREE

factor	$V_n(t)$	$\theta_n(t)$	$\dot{\theta}_n(t)$	$\varphi_m(t)$	$\dot{\varphi}_m(t)$	$\frac{\dot{\theta}_m(t)}{\theta_m(t)}-\frac{\dot{\beta}_m(t)}{\beta_m(t)}$	$\frac{\dot{\varphi}_m(t)}{\varphi_m(t)}$
Correlation	0.807	0.820	0.802	0.711	0.790	0.794	0.873

IV. APPLICATIONS OF MODELING

A. Baseline Model

Based on the optimized speed following model, Jiang [19] proposed a classic full speed difference model with over a thousand references.

$$a_n(t) = \alpha\left\{V\left[\Delta x_n(t)\right] - v_n(t)\right\} - \lambda \Delta v_n(t) \tag{16}$$

where, α and λ are the sensitivity coefficient. $\Delta x_n(t)$ and $\Delta v_n(t)$ are respectively the relative distance and velocity between the front and rear vehicles.

However, from the perspective of driver psychology, the most important perceptual information in car following behavior may be visual information. so Jin [20] replaces the traditional headway with the visual angle in following behavior.

$$a_n(t) = \alpha\left\{V\left[\theta_n(t)\right] - v_n(t)\right\} - \lambda\, \mathrm{d}\theta_n(t)/\mathrm{d}t \tag{17}$$

$$\theta_n(t) = w/\left(\Delta x_n(t) - l\right) \tag{18}$$

where, $\theta_n(t)$ is the visual angle of drivers. w and l are the width and length of leading vehicles. $\mathrm{d}\theta_n(t)/\mathrm{d}t$ represents change rate of visual angles. $V(\theta_n)$ is the optimized velocity.

B. New Model Derivation

The TTC is the time until vehicles crash assuming the collision path and velocity differential are maintained [21]. The driver of the following car will take control measures such as acceleration or deceleration according to the change in TTC with the leading vehicle [22]. The commonly used traffic accident alternative evaluation index TTC is used as our reference. We combine the correlation between the driver's visual psychology and driving operations in Table 4 to try to establish a more realistic and accurate car-following model to quantify the risk perception of collision accidents. Therefore, driver sensitivity to lateral collision is incorporated into the visual angle model to improve the stability and safety of traffic flow. The TTC expression for a single lane can be written as

$$\mathrm{TTC}_m(t) = \frac{x_{m+1}(t) - x_m(t)}{v_m(t) - v_{m+1}(t)} = \frac{\varphi_m}{\dot{\varphi}_m} \tag{19}$$

where, $x_{m+1}(t)$, $x_m(t)$ and $v_{m+1}(t)$, $v_{m+1}(t)$ are the position and velocity of vehicles. However, the car-following considering lateral influence has the following geometric relationship as shown in Fig. 11.

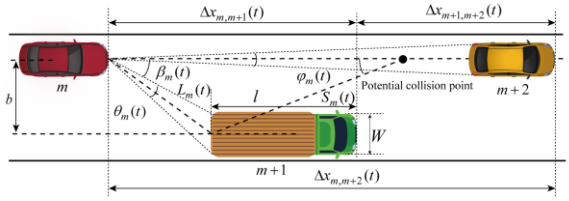


Fig. 11. Car-following behavior with visual angle.

$$\mathrm{S}_m(t) = \sin\beta_m(t) \cdot \frac{\mathrm{L}_m(t)}{\sin\gamma_m(t)} \tag{20}$$

$$\dot{\mathrm{S}}_m(t) = \frac{\left(\dot{\mathrm{L}}_m(t)\sin\beta_m(t) + \mathrm{L}_m(t)\cos\beta_m(t)\dot{\beta}_m(t)\right)}{\sin\gamma_m(t)} \tag{21}$$

where, $\mathrm{L}_m(t)$ is the distance between the midpoint of tail of preceding vehicles and the midpoint of head of the following vehicles, $\mathrm{S}_m(t)$ is the distance between the head of preceding vehicles and the conflict point. The angle $\beta_m(t)$ defined by $\mathrm{L}_m(t)$ and the distance to the collision point. We assume that the leading car maintains constant velocity and $\varphi_m(t)$ for the current brief time. we roughly think that $\tan\beta_m(t) = \beta_m(t)$. Substituting Eq. (20) into Eq.(21) and eliminating the $\sin\gamma_m(t)$.

$$\frac{\dot{\mathrm{S}}_m(t)}{\mathrm{S}_m(t)} = \frac{\dot{L}_m(t)}{L_m(t)} + \frac{\dot{\beta}_m(t)}{\beta_m(t)} \tag{22}$$

Then the potential collision with side vehicle can be expressed.

$$\frac{1}{\mathrm{TTC}} = \frac{\dot{\theta}_m(t)}{\theta_m(t)} - \frac{\dot{\beta}_m(t)}{\beta_m(t)} \tag{23}$$

Therefore, we extend the TTC indicator in car-following behavior to car-following behavior, and can obtain a new visual angle model considering collision visual sensitivity:

$$\begin{aligned} a_m(t) = \alpha\left\{V\left[\theta_m(t), \varphi_m(t)\right] - v_m(t)\right\} - \lambda\Big[\left(1-\xi_m\right)\dot{\theta}_m(t) \\ + \xi_m\dot{\varphi}_m(t)\Big] - \kappa\left[\left(1-\xi_m\right)\left(\frac{\dot{\theta}_m(t)}{\theta_m(t)} - \frac{\dot{\beta}_m(t)}{\beta_m(t)}\right) + \xi_m\frac{\dot{\varphi}_m(t)}{\varphi_m(t)}\right] \end{aligned} \tag{24}$$

The comprehensive optimization velocity is as follows:

$$\begin{aligned} V\left[\theta_m(t), \varphi_m(t)\right] = V_1 + V_2\tanh\Big\{C_1\Big[\left(1-\xi_m\right)w/2\tan\left(\theta_m(t)/2\right) \\ + \xi_m\, w/2\tan\left(\varphi_m(t)/2\right)\Big] - C_2\Big\} \end{aligned} \tag{24}$$

These parameters V_1, V_2, C_1, C_2 were verified by Zhang [14].

C. Stability Analysis

We assume that same size vehicles traveling on a ring road with a uniform flow as the initial state. Each vehicle maintains the same headway h with adjacent vehicles at a uniform velocity $V(h)$ .So, the initial moment can be considered as,

$$x_m^0(t) = hn + V(\theta_0, \varphi_0)t \tag{25}$$

where $\theta_0 = 2\arctan\left[w/2(h-l)\right]$, $\varphi_0 = 2\arctan\left[w/(4h-2l)\right]$.

When the perturbation $y_m(t)$ appear, we can obtain

$$x_m(t) = x_m^0(t) + y_m(t) \tag{26}$$

Visual angles formed by the driver when he observes himself and the side vehicle, and angles formed by side vehicle and the collision position are respectively expressed as,

$$\theta_m(t) = 2\arctan\left[w/2\left(\Delta x_{m,m+1}(t) - l\right)\right] \tag{27}$$

$$\varphi_m(t) = 2\arctan\left[w/2\left(\Delta x_{m,m+2}(t) - l\right)\right] \tag{28}$$

$$\beta_m(t) = \arctan\left[d\tan\gamma/\left(\Delta x_{m,m+1}(t) - l\right) - \tan\gamma\right] \tag{29}$$

To facilitate calculation of the visual angle expression is linearized using Taylor expansion Substituting Eq. (27) into Eq. (27), Eq. (28) and Eq. (29) higher-order terms of $\Delta y_m(t)$ can be rounded off. Drivers' visual angles are expressed as follows

$$\theta_m(t) = 2\arctan\frac{w}{2(h-l)} - \frac{w}{(h-l)^2 + (w/2)^2}\Delta y_{m,m+1}(t) \tag{30}$$

$$\beta_m(t) = \arctan\left(\frac{d\tan\gamma}{h-l} - \tan\gamma\right) - \frac{d\tan\gamma}{(h-l)^2 + \tan^2\gamma(d-h+l)^2}\Delta y_{m,m+1}(t) \tag{31}$$

$$\varphi_m(t) = 2\arctan\frac{w}{2(h-l)} - \frac{w}{(h-l)^2 + (w/2)^2}\Delta y_{m,m+2}(t) \tag{32}$$

Substituting (31)~ (33) into (24), the following is obtained

$$\begin{aligned}\ddot{y}_m(t) = {} & \alpha A\dot{V}(\theta_0,\varphi_0)\left\{(1-\xi_m)\Delta y_{m,m+1}(t) - \xi_m\Delta y_{m,m+2}(t)\right. \\ & \left. -\dot{y}(t)\right\} + \lambda A\left((1-\xi_m)\Delta\dot{y}_{m,m+1}(t) + \xi_m\Delta\dot{y}_{m,m+2}(t)\right) \\ & + \kappa\left((1-\xi_m)\left(\frac{-\Delta\dot{y}_{m,m+1}(t)}{B-\Delta y_{m,m+1}(t)} + \frac{\Delta\dot{y}_{m,m+1}(t)}{C-\Delta y_{m,m+1}(t)}\right)\right. \\ & \left. + \xi_m\frac{-\Delta\dot{y}_{m,m+2}(t)}{B-\Delta y_{m,m+2}(t)}\right)\end{aligned} \tag{33}$$

where,

$$\begin{cases} A = -\dfrac{w}{(h-l)^2 + (w/2)^2} \\ B = \dfrac{(h-l)^2 + (w/2)^2}{w}2\arctan\dfrac{w}{2(h-l)} \\ C = \dfrac{(h-l)^2 + \tan^2\gamma(d-h+l)^2}{d\tan\gamma}\arctan\left(\dfrac{d\tan\gamma}{h-l} - \tan\gamma\right) \end{cases} \tag{34}$$

where, $\dot{V}(\theta_0,\varphi_0) = dV(\theta,\varphi)/d(\theta)\big|_{\theta=\theta_0,\varphi=\varphi_0}$,We expand $\Delta y_m(t) = B\exp(ikm + zt)$ by Fourier series and substitute it into Eq. (34).

$$\begin{aligned} z^2 = {} & \alpha A\dot{V}(\theta_0,\varphi_0)\left\{(1-\xi_m)\left(e^{ik}-1\right) - \xi_m\left(e^{2ik}-1\right)\right. \\ & \left. -\dot{y}_n(t)\right\} + \lambda A\left((1-\xi_{m+1})z\left(e^{ik}-1\right) + \xi_m z\left(e^{2ik}-1\right)\right) \\ & -\kappa\left((1-\xi_m)\left(\frac{-z\left(e^{ik}-1\right)}{B-\left(e^{ik}-1\right)} + \frac{z\left(e^{ik}-1\right)}{C-\left(e^{ik}-1\right)}\right)\right. \\ & \left. +\xi_m\frac{-z\left(e^{2ik}-1\right)}{B-\left(e^{2ik}-1\right)}\right)\end{aligned} \tag{35}$$

Expanding z according to $z = z_1(ik) + z_2(ik)^2 + \cdots$,we can obtain as follows:

$$\begin{cases} z_1 = (1-\xi_m)A\dot{V} \\ z_2 = \dfrac{(1+3\xi_m)A\dot{V}}{2} + \dfrac{1}{\alpha}\left[\dfrac{\kappa(1-\xi_m)(B-C) + 2\kappa\xi_m BC}{BC}\right. \\ \qquad \left. + A\lambda(1+\xi_m)\right]z_1 - \dfrac{1}{\alpha}z_1^2 \end{cases} \tag{36}$$

According to the hypothesis of long-wave expansion, if z_2 is positive, traffic flow is still steady under small disturbances. Therefore, the stability curve is obtained.

$$\alpha = \frac{2A(1-\xi_m)^2(\dot{V}+\lambda)}{1+3\xi_m} - \frac{2\kappa(1-\xi_m)^2(B-C) + 4\kappa(1-\xi_m)}{BC} \tag{39}$$

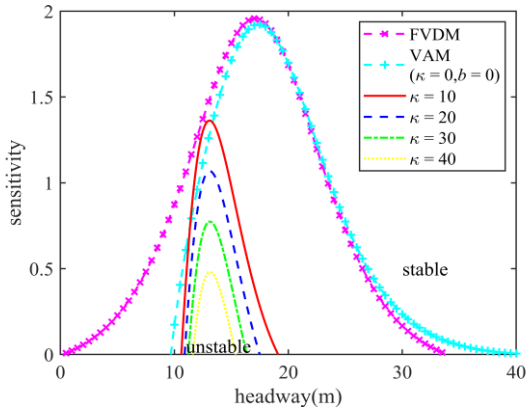


Fig. 12. Stability curves at different κ.

The stability curves of the car-following considering visual sensitivity to collision at different κ are depicted in Fig. 12. The upper and lower sides of the stability curve respectively are stable and unstable regions. Stability curves of different collision sensitivity coefficients κ are depicted in Fig. 12. We set $\lambda = 1, \xi = 0.1, b = 3.6, w = 1.6$ .When the parameters are $\kappa = 0, b = 0, \xi = 0$, our model degenerates into the VAM and

the following vehicle only optimizes its own velocity according to leading vehicles on the current lane. The area formed by the x-axis and the stability curves of the new model shrinks as κ increases, and is smaller than the area of FVDM and VAM in the Fig. 12. The order of areas formed by curves and x-axis in Fig. 12 is FVDM> VAM> SCVAM, and they gradually increase as κ decreases. The findings indicate that the platoon is more stable if the visual collision factor is introduced into the car-following model. In addition, the stability gradually improves with the increase of visual collision sensitivity, and the dynamic performance is better than the full velocity difference (FVDM) and visual angle models (VAM).

D. Simulation

Within this part of the research, we performed a series of simulations on the evolution of small perturbations for the new visual model Eq. (24) to analyze its dynamic performance. To validate the theoretical results obtained above, numerical simulations of Eq. (15) with periodic boundary conditions is given. The time step is 0.1s. The initial setting was adopted.

$$\begin{cases} \Delta x_i(0) = \Delta x_i(1) = 5.0, & (i \neq 30, 31) \\ \Delta x_i(0) = \Delta x_i(1) = 5.0 - 0.1, & (i = 30) \\ \Delta x_i(0) = \Delta x_i(1) = 5.0 + 0.1, & (i = 31) \end{cases} \tag{37}$$

where 100 vehicles are traveling on a 1500m ring road.

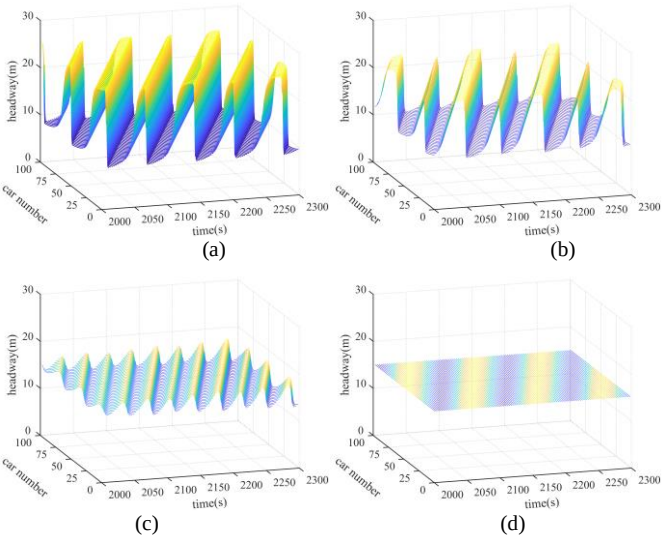


Fig. 13. Space-time graphs of headway at different κ.

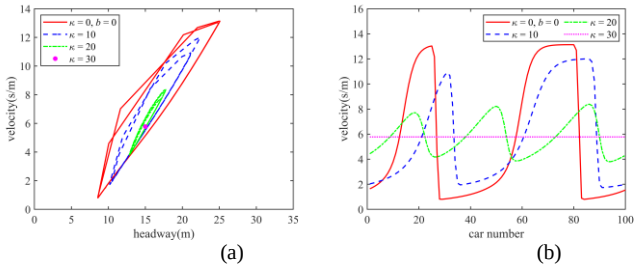


Fig. 14. (a) Hysteresis loops with different κ (b) Velocity images of all vehicles with different κ.

Fig. 13(a) to Fig. 13(d) reflect the temporal and spatial evolution of the car-following model considering the visual sensitivity of collision with different κ from 2000s to 2300s. The parameters are $\alpha = 0.41, b = 3.6, \lambda = 20, w = 2$. The vehicle platoon is unstable at $\kappa = 0, 10, 20, 30$ by stability condition Eq.(39). Small disturbances in the system will gradually amplify over time and cause vehicle platoon congestion, which can be clearly seen in Fig. 11(a) to Fig. 11(c). Especially when the lateral distance b and collision sensitivity κ are not considered, the newly proposed visual angle car-following model Eq. (24) degenerates to VAM ($\kappa = 0, b = 0$), the greatest fluctuation is in headway as shown in Fig. 11(a). When $\kappa = 30$, the stability condition is satisfied, and the local disturbance of the distance between the vehicles in the platoon will gradually return to stable state in Fig. 11(d). Similar, the areas of hysteresis loops decrease gradually when the visual sensitivity of collision κ increase in Fig. 14(a). As the visual sensitivity of collision κ increases, the velocity fluctuation decreases are depicted in Fig. 14(b).

Fig. 13 and Fig. 14 show that the stability of the vehicle platoon is well maintained when taking into account the collision visual sensitivity to the lateral and longitudinal leading vehicles. As the sensitivity κ increases, the stability gradually increases.

E. Safety Assessment

The traffic conflict theory can observe a large amount of non-accident data before the accident, and estimate the dangerous action highly related to the accident by analyzing the vehicle. Numerous indicators have been developed for alternative safety measures to evaluate collisions risk. TTC is the most used indicator among them. However, traditional traffic conflict alternative indicators have two shortcomings: 1) They do not start from the driver's real visual perspective. The vehicle is regarded as a point, and the impact of its size on collision is not considered. 2) The potential risk of collision with the two vehicles in front is not considered when following a car. Therefore, we propose a new car-following collision risk indicator STTC, which is expressed as follows.

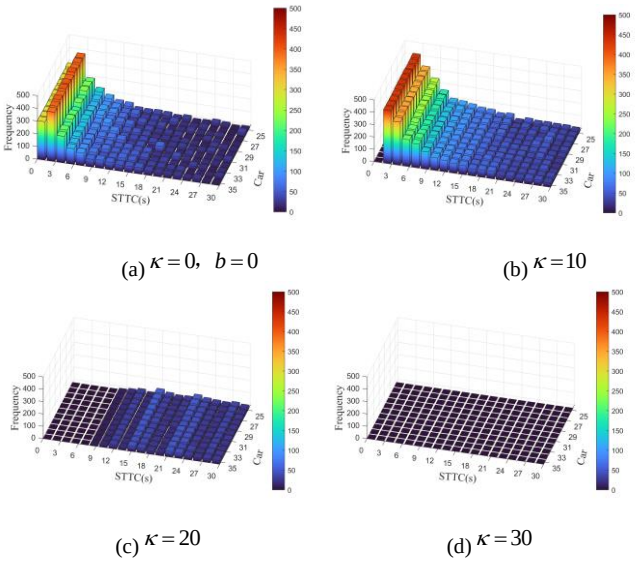
$$\text{STTC} = (1 - \xi_m)\left(\frac{\dot{\theta}_m(t)}{\theta_m(t)} - \frac{\dot{\beta}_m(t)}{\beta_m(t)}\right) + \xi_m \frac{\dot{\varphi}_m(t)}{\varphi_m(t)} \tag{38}$$

We also use the small perturbation evolution scenario in previous section to study potential collision risks. When the 30th vehicle in the uniform traffic flow suddenly suffered a small disturbance, we selected it and five vehicles at the front and rear, a total of eleven vehicles, for potential collision risk assessment. Since the simulation step in car-following behavior is small and the velocity difference between nearby vehicles is small, we use a larger range of STTC to evaluate the risk. Similar to the common threshold standard of the TTC index, we define $\text{Frequency}_{\text{STTC<30s}}$ as the traffic environment with potential collisions. With reference to the common international standards of TTC, $\text{Frequency}_{\text{STTC<3s}}$ and $\text{Frequency}_{\text{STTC<5s}}$ are defined as the frequency of serious traffic conflicts and relatively serious traffic conflicts.

The safety of the entire platoon is evaluated by the dispersion of velocity and headway. The standard deviation of velocity and headway can be calculated by following formulas.

$$SD_V = \sum_{t=1}^{T}\sum_{m=1}^{M}\left(v_{t,m}-\bar{v}\right) \Big/ (M \cdot T - 1), \quad SD_h = \sum_{t=1}^{T}\sum_{m=1}^{M}\left(h_{t,m}-\bar{h}\right) \Big/ (M \cdot T - 1) \quad (39)$$

where $\bar{v}$ and $\bar{h}$ are the average velocity and headway.



(a) $\kappa = 0,\ b = 0$ (b) $\kappa = 10$

(c) $\kappa = 20$ (d) $\kappa = 30$

Fig. 15. Statistical distributions for STTC with different κ

TABLE V. TOTAL OCCURRENCES OF STTC FOR DIFFERENT κ

κ	$0(\xi = 0)$	10	20	30
$Frequency_{STTC<3s}$	720	441	0	0
$Frequency_{STTC<5s}$	952	875	0	0
$Frequency_{STTC<30s}$	2189	1646	442	0

TABLE VI. SD OF HEADWAY AND VELOCITY WITH DIFFERENT κ

κ	$0\ (\xi = 0)$	10	20	30
Velocity	4.61	3.03	0.49	0.05
Headway	6.15	3.59	0.52	0.07

Fig. 15 shows that when the collision sensitivity is not considered $(\kappa = 0, b = 0)$, the frequency of $VT < 3s$ is the highest, and the risk of serious collision conflict is the highest. In addition, the frequency of dangerous driving within 5s threshold and the potential collision conflict within 30s threshold are also the highest. However, as the driver's sensitivity κ to the collision of two vehicles intersecting in front increases, the frequency of the three safety evaluation indicators within the threshold gradually decreases until it completely enters a safe driving state.

The STTC values of different collision sensitivities at different risk thresholds are counted in Table V. The standard deviation of the headway and velocity at different collision sensitivities κ are descriptive statistics in Table VI. This shows that increasing the sensitivity to the collision of horizontal and longitudinal leading vehicles can reduce dispersion of headway and velocity, thereby improving driving safety.

Through the above simulation experiments and statistical analysis, we found that after analyzing the traffic behavior data collected by flexible and convenient drone images, the newly established car-following model can describe the collision risk of driving behavior. And by enhancing the sensitivity of the newly introduced visual collision factor in the model as in Eq.(24), the driver's safety in car following behavior will be improved. This puts forward new ideas for safety modeling and evaluation of driving behavior.

V. CONCLUSION

To sum up, the method in this article for extracting traffic parameters from UAV video and applying it to assist in establishing a car-following model can improve the accuracy and lightweight of multi-target detection and tracking. The improved ShuffleNet network and GSConv module introduce the Yolov7-tiny target detection stage to reduce the number of parameters and calculations of the model and ensure accuracy. HOG features and IOU motion metrics are introduced into the DeepSort multi-target tracking algorithm to improve the target appearance representation capabilities and the accuracy of tracking targets. The traffic parameters extracted by the new method can be used to analyze driving psychology and car-following behavior, and the analysis results can be used to model car-following behavior with the aim of enhancing both the safety and stability of traffic flow. This will in turn ease traffic congestion and reduce driver collision risks. In the future, we will further improve vehicle tracking in more complex traffic scenarios, and verify the accuracy of traffic parameter extraction through on-board comparison.

ACKNOWLEDGMENT

This study was supported by the National Natural Science Foundation of China (Grant No. 11772264) and the Natural Science Basic Research Program of Shaanxi Province (Grant No. 2020JM119).

REFERENCES

[1] A. Ammar, A. Koubaa, M. Ahmed, A. Saad and B. Benjdira, "Vehicle Detection from Aerial Images Using Deep Learning: A Comparative Study." Electronics. vol. 10, no. 7. 2021.

[2] H. Yao, R. Qin and X. Chen, "Unmanned Aerial Vehicle for Remote Sensing Applications-A Review." Remote Sensing. vol. 11, no. 12. 2019.

[3] R. Ravindran, M. J. Santora and M. M. Jamali, "Multi-Object Detection and Tracking, Based on DNN, for Autonomous Vehicles: A Review." Ieee Sensors Journal. vol. 21, no. 5, pp. 5668-5677. 2021.

[4] S. B. Makarov, V. A. Pavlov, A. K. Bezborodov, A. I. Bobrovskiy and D. Ge, "Multiple Object Tracking Using Convolutional Neural Network on Aerial Imagery Sequences." 2021.

[5] S. Hossain and D.-J. Lee, "Deep Learning-Based Real-Time Multiple-Object Detection and Tracking from Aerial Imagery via a Flying Robot with GPU-Based Embedded Devices." Sensors. vol. 19, no. 15. 2019.

[6] P. Zhang, Y. Zhong and X. J. I. Li, "SlimYOLOv3: Narrower, Faster and Better for Real-Time UAV Applications." 2019.

[7] Z. Li, X. Liu, Y. Zhao, B. Liu, Z. Huang and R. Hong, "A lightweight multi-scale aggregated model for detecting aerial images captured by

UAVs." Journal of Visual Communication and Image Representation. vol. 77. 2021.

[8] X. Luo, R. Zhao and X. Gao, "Research on UAV Multi-Object Tracking Based on Deep Learning," 2021 IEEE International Conference on Networking, Sensing and Control (ICNSC), Xiamen, China, pp. 1-6, 2021.

[9] M. B. Khalkhali, A. Vahedian and H. S. Yazdi, "Situation Assessment-Augmented Interactive Kalman Filter for Multi-Vehicle Tracking." Ieee Transactions on Intelligent Transportation Systems. vol. 23, no. 4, pp. 3766-3776. 2022.

[10] M. Ning, X. Ma, Y. Lu, S. Calderara and R. Cucchiara, SeeFar: Vehicle Speed Estimation and Flow Analysis from a Moving UAV, 21st International Conference on Image Analysis and Processing (ICIAP), Lecce, ITALY, 2022, pp. 278-289.

[11] Y. Du, J. Wan, Y. Zhao, B. Zhang, Z. Tong, J. Dong and I. C. Soc, GIAOTracker: A comprehensive framework for MCMOT with global information and optimizing strategies in VisDrone 2021, 18th IEEE/CVF International Conference on Computer Vision (ICCV), Electr Network, 2021, pp. 2809-2819.

[12] W. Huang, X. Zhou, M. Dong and H. Xu, "Multiple objects tracking in the UAV system based on hierarchical deep high-resolution network." Multimedia Tools and Applications. vol. 80, no. 9, pp. 13911-13929. 2021.

[13] X. Z. Zhang, Z. K. Shi, J. Z. Chen and L. J. Ma, "A bi-directional visual angle car-following model considering collision sensitivity." Physica a-Statistical Mechanics and Its Applications. vol. 609. 2023.

[14] X. Zhang, Z. Shi, S. Yu and L. Ma, "A new car-following model considering driver's desired visual angle on sharp curves." Physica a-Statistical Mechanics and Its Applications. vol. 615. 2023.

[15] D.-W. Liu, Z.-K. Shi and W.-H. Ai, "Enhanced stability of car-following model upon incorporation of short-term driving memory." Communications in Nonlinear Science and Numerical Simulation. vol. 47, pp. 139-150. 2017.

[16] G. Ma, M. Ma, S. Liang, Y. Wang and Y. Zhang, "An improved car-following model accounting for the time -delayed velocity difference and backward looking effect." Communications in Nonlinear Science and Numerical Simulation. vol. 85. 2020.

[17] W. Sun, B. Fu and Z. Zhang, "Maize Nitrogen Grading Estimation Method Based on UAV Images and an Improved Shufflenet Network." Agronomy-Basel. vol. 13, no. 8. 2023.

[18] X. Zhao and Y. Song, "Improved Ship Detection with YOLOv8 Enhanced with MobileViT and GSConv." Electronics. vol. 12, no. 22. 2023.

[19] R. Jiang, Q. S. Wu and Z. J. Zhu, "Full velocity difference model for a car-following theory." Physical Review E. vol. 64, no. 1. 2001.

[20] S. Jin, D.-H. Wang, Z.-Y. Huang and P.-F. Tao, "Visual angle model for car-following theory." Physica a-Statistical Mechanics and Its Applications. vol. 390, no. 11, pp. 1931-1940. 2011.

[21] M. M. Minderhoud and P. H. Bovy, "Extended time-to-collision measures for road traffic safety assessment." Accident analysis and prevention. vol. 33, no. 1, pp. 89-97. 2001.

[22] D. N. Lee, "A theory of visual control of braking based on information about time-to-collision." Perception. vol. 38, no. 6, pp. A43-A65. 2009.