Journal of Computers Vol. 34 No. 2, April 2023, pp. 215-225
doi: 10.53106/199115992023043402016

Research on Dynamic Recognition and Tracking Technology for Complex Scenes of Automatic Driving

Shuai-Wu Zhang[1*], Yu-Mei Zhao[1], Xiang-Lian Yang[2]

[1] Tangshan Polytechnic College,
Tangshan City 063000, Hebei Province, China
{tgysb935223, sw2023030405}@126.com

[2] Wuxi Institute of Communications Technology,
Wuxi City 214000, Jiangsu Province, China
tgy511389410@163.com

Received 1 March 2023; Revised 15 March 2023; Accepted 16 March 2023

Abstract. With the development of automobile technology, intelligent vehicle and automatic driving technology will make due contributions to reducing traffic accidents. This paper aims to improve the dynamic identification and tracking technology in the current intelligent vehicle and automatic driving. First, it is improved based on the MobileNet V2 backbone network, and then a new tracking model framework is designed combining with the SiamRPN single target tracker. Secondly, it integrates space-time tracking clues to improve the stability and robustness of the algorithm. Finally, it constructs a pedestrian dynamic identification algorithm based on the dynamic pedestrian factors in the driving process. Through the training of data sets and video tracking experiments, the performance of the algorithm in this paper is proved quantitatively and qualitatively.

Keywords: autopilot, dynamic identification, image tracking, deep learning

1 Introduction

According to the statistics of traffic management department, by the end of 2022, China's car ownership has reached 323 million, ranking first in the world, followed by frequent traffic congestion and traffic accidents. Research shows that most traffic accidents are caused by improper operation of drivers themselves, such as fatigue driving, acceleration driving, and failure to avoid pedestrians in time

When a traffic accident occurs, if the driver can be reminded to take effective actions within the effective time, the loss caused by the traffic accident can be minimized. Intelligent vehicle and automatic driving technology is a comprehensive intelligent system that integrates environmental awareness, decision planning and motion control. It can help drivers avoid danger through multi-sensor fusion such as binocular camera, laser radar and millimeter wave radar. However, the existing intelligent vehicle and automatic driving technology are limited by the technical bottleneck of the camera, and can not better realize the identification of traffic signs and target tracking in the environment of insufficient light, bad weather, obstacle shelter, dynamic pedestrian interference, etc. Therefore, in order to improve the real-time and reliability of intelligent vehicles and automatic driving, the work done in this paper is as follows:

(1) The design process of dynamic recognition and tracking algorithm framework in intelligent driving process is described;

(2) Improve the recognition and tracking performance of the designed algorithm by integrating short-term and long-term space-time clues into the algorithm;

The algorithm preliminarily realizes the tracking and prediction of pedestrian dynamics, and the performance of the algorithm tracking is verified by experiments.

This paper is divided into six chapters in terms of structural arrangement. The second chapter integrates the relevant research results of relevant scholars to lay the foundation for the development of this paper; The third chapter describes the structure design and improvement process of the algorithm; The fourth chapter studies the recognition and tracking of dynamic pedestrians; The fifth chapter is the experimental data validation to prove the feasibility and performance improvement of the algorithm; The sixth chapter is the conclusion.

[*] Corresponding Author

2 Related Work

For dynamic target tracking, relevant scholars have done corresponding research and achieved certain research results. Vimal Kumar has developed a new algorithm for tracking vehicles in close range using low-density flash laser radar, and has realized the lightweight of the algorithm structure [1]. Priya Mariam Raju, Li Ming, the area method based on case segmentation, uses correlation filter to locate the final target, and improves the recognition and tracking ability. The experimental results show that the average accuracy gain is 2.47%

[2]. Castro proposed an improved dynamic optimization problem swarm intelligence algorithm, which uses the robust version of the double exponential smoothing (DES) model for outliers to predict the position of the target in the frame that delimits the solution space. The experiment shows that the average processing time of each frame is at least 10% faster than the competitive tracker [3]. Wael Farag proposed a method of data fusion based on lidar and radar measurement data installed on the vehicle, and used the customized unscented Kalman filter for data fusion. Unlike other detection and tracking methods, the balance of the accuracy and real-time performance of the pose estimator is its main contribution. The highly optimized mathematical and optimization library is used to achieve the best real-time performance and the simulation research is carried out [4]. Zilong Liu proposed a dual-mode weight self-updating twin network target tracking method, which effectively utilizes the complementary advantages of infrared and visible images in the field of target tracking. Experimental results show that the proposed target tracking algorithm has achieved good tracking results in various tracking scenarios [5]. Hongying Zhang proposed a real-time target tracking algorithm based on improved SiamFC, replacing the second layer of convolution in the original network structure with deep separable convolution, which improves the tracking speed and network recognition ability by reducing the amount of parameter calculation [6]. Junsong Zheng proposed a spatiotemporal continuous multi-feature fusion twin network algorithm, designed a robust feature that combines spatial information and semantic information from coarse to fine to express fast moving weak targets, and added feature attention mechanism. The algorithm idea has reference significance [7].

3 Design of Multi-Object Dynamic Recognition and Tracking Algorithm

Vehicle dynamic tracking is an indispensable part of intelligent vehicle automatic driving. Whether to effectively identify and track many targets and dynamic factors around the vehicle so as to realize the prediction of pedestrian movement and avoid the occurrence of accidents is the core issue to be solved by target tracking technology. Therefore, this paper adopts the improved MobileNet V2 [8] pedestrian detection framework model and SiamRPN [9] single target tracker, the Multi-Object tracking algorithm combined with spatiotemporal clue fusion can solve the false negative detection problem of the detector while increasing the matching computation, at the same time, cascade matching strategy is introduced into the similarity calculation. The overall process of the tracking framework is shown in Fig. 1.

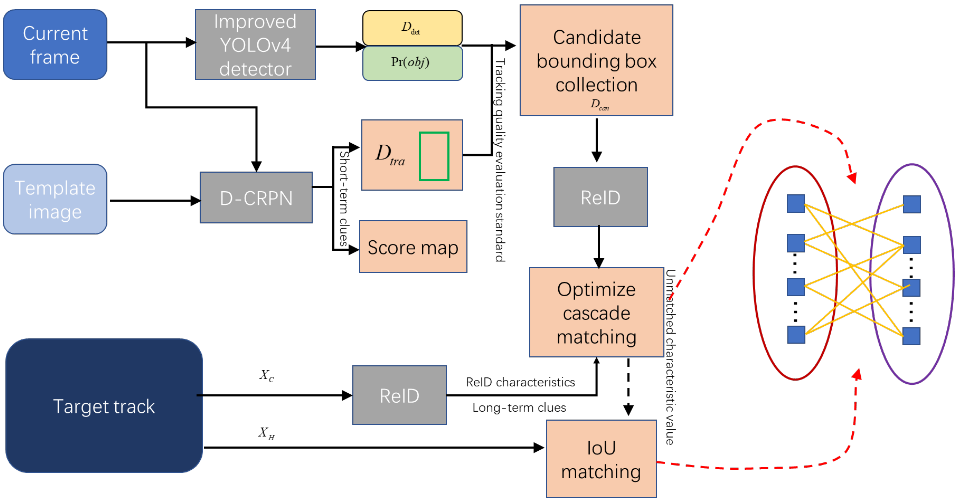

Fig. 1. Overall framework of target tracking and identification

3.1 Tracking Frame Design

Suppose the track of a tracked target is expressed as $T = \{T_t\}$ .The feature set saved in T is $F = \{F_t\}$. T_t , B_{test} and B_{tra} all contain the coordinates, width and height of the center point of the bounding box, so the online video tracking steps of the Multi-Object tracking framework are as follows, the name of each parameter is shown in Table 1:

Table 1. Tracking process parameter data sheet

Name	Character	Name	Character
Track collection	$T = \{T_t\}$	Detect bounding box	B_{test}
Feature set	$F = \{F_t\}$	Trace bounding box	B_{tra}
Video frame index	t	Detect bounding box collection	G
On the depth feature of X_t	F_t	Trace bounding box collection	N
Any element in H	D_{can}	Candidate bounding box collection	H
Unmatched set	P_I	Match pair set	P_c

Step 1: The improved MobileNet V2 detector is used to detect pedestrians in the current frame (time t), Filter all B_{test}'s by setting the minimum height condition and confidence threshold of the bounding box, gets the set of detection bounding boxes G about the current frame.

Step 2: SiamRPN tracker is used as the motion model of the visual multi-object tracking framework. Based on the multi-process strategy, target tracking results are predicted for the current frame, and the tracking boundary box set N is obtained, at the same time, the uncertainty factor of each SiamRPN tracker is expressed as $u_s + = 1$. The structure of SiamRPN is shown in Fig. 2.

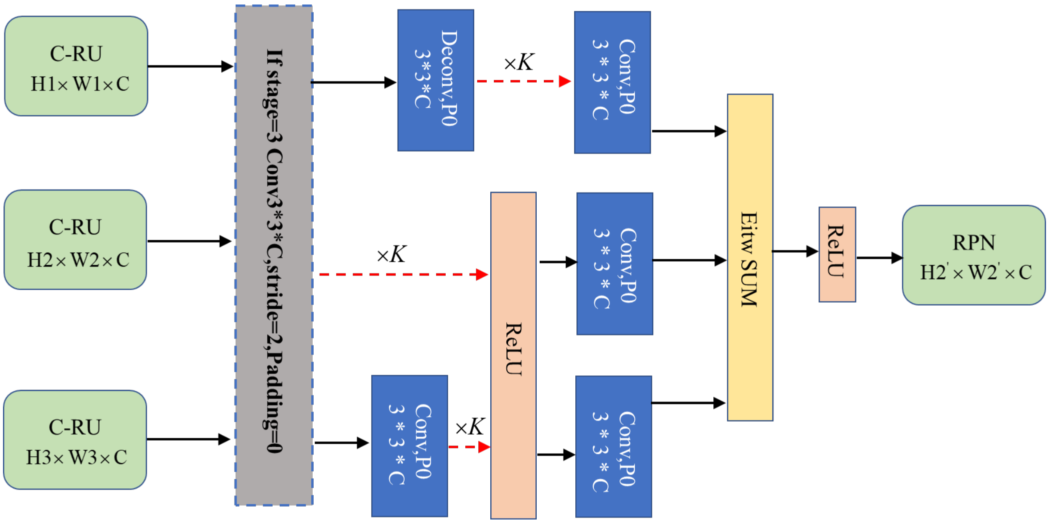

Fig. 2. SiamRPN structure diagram

Step 3: Use the spatiotemporal clue fusion strategy to filter the elements in N to get $N^{'}$, merge the sets G and $N^{'}$, and generate the candidate boundary box set $H = G \cup T^{'}$ about the tracking results.

Step 4: Divide the existing set of T into the subset of Confirmed tracks (If the match is successful for *1* or more consecutive times, it is recorded as T_c) and the subset of Unconfirmed tracks (Successful matching less than *1* times in a row, recorded as T_u) according to the number of consecutive matching successes, and perform the following operations:

For T_c, cascade and match it with all elements in *H*.

1) First, the depth feature of each candidate result in the current frame is extracted, and the scoring matrix S_c is generated by calculating the cosine distance between it and the feature set F of each track in a layer. Each element in the matrix is taken as the minimum cosine distance between the depth feature D_{can} and a track F_t.

2) Secondly, combining motion estimation and motion compensation algorithms, calculate the Mahalanobis distance between the final tracking result of T_c at time $t-1$ and all D_{can} at time t after the previous frame is successfully matched, obtain the index information with the result greater than the set threshold, and set the element value of the corresponding index to infinity in S_c.

3) Finally, apply matrix S_c to Hungarian algorithm to get the matching result of the current layer.

4) Cycle the above process and mark the matching set as P_c. For T_u and T_c that have not been matched successfully, B_{test} matching is performed for the output of the nearest moment and the remaining B_{tra} that does not have a matching relationship in the cascade matching to alleviate the interference caused by the sudden change of target apparent characteristics or local occlusion. If the obtained matching pair set is marked as P_I, the final matching pair set is $P = P_c \cup P_I$.

Step 5: Referring to the matching relationship in set p, save the depth feature of D_{can} to the feature set of matching T, and set the tracker parameter u_s corresponding to T to 0. In addition, if T_u has been successfully matched for 5 consecutive times, change it to T_c.

Step 6: For T that does not match successfully, if it belongs to T_u, it will be removed from the track set; If it belongs to T_c, and it has exceeded n frames and failed to match, it will also be removed from the track collection.

Step 7: Create a new tracker for B_{test} that does not match successfully, and output the tracking prediction result of T_c in the current frame.

Step 8: By setting $t = t + 1$, repeat steps 1 to 7 at the next frame until there is no new input frame.

3.2 Multi-Object Tracking with Space-time Clues

SiamRPN target tracker is used to extract short-term clues for Multi-Object Tracking. First, MobileNet V2 target detector is applied to the initial video frame to obtain the regression boundary box information about all human targets in the scene, and then the center of the boundary box of each target area is expanded. If the boundary box size is (w, h), the original template area size is expressed as:

$$w_z = w + (w + h) / 2 . \tag{1}$$

$$h_z = h + (w + h) / 2 . \tag{2}$$

Under symmetric expansion, the expansion distance is expressed as:

$$p = (w + h) / 4 . \tag{3}$$

If the branch image resolution of the actual input twin network model is A, the scaling factor D from the square area with B as the side length to C is obtained by formula 4 and 5:

$$c_z = \sqrt{(w+2p)\times(h+2p)}\ . \tag{4}$$

$$A = s(w+2p)\times s(h+2p)\ . \tag{5}$$

The short-term clues extraction network based on SiamRPN is shown in Fig. 3:

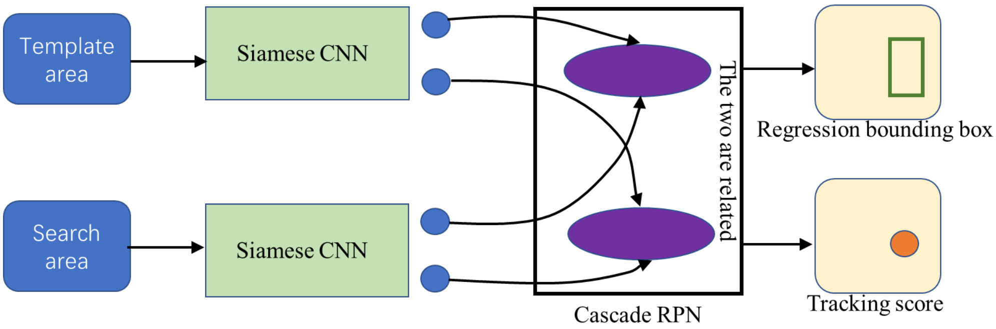

Fig. 3. SiamRPN structure diagram

The spatiotemporal clues fusion strategy of the short-term clues extracted by the target tracker and the historical appearance characteristics of the target track is as follows:

Step 1: In this paper, B_{test} and B_{tra} are merged to expand the set of candidate boundary boxes participating in the matching, and part B_{tra} is added to the matching sequence. Through the final track update result, the tracking drift phenomenon is suppressed.

Step 2: Through the quantitative analysis of the output quality at time t and time $t-1$, the tracking quality evaluation criteria shown in formula 6 are formulated to achieve the screening of B_{tra}.

$$q_t = p_t \cdot \exp(\max_{i=1,2,\cdots,n} IoU(B_{tra}^{t}, B_{test}^{t_i})) + \lambda \cdot p_{t-1} \cdot IoU(B_{tra}^{t-1}, D_X^{t-1})\ . \tag{6}$$

Where, p_t and p_{t-1} represent the target tracking output results B_{tra}^{t} and B_{tra}^{t-} scores at t and $t-1$, respectively, $B_{test}^{t_i}$ represents the ith detection result in frame t, D_X^{t-1} represents the final tracking update result of track X at $t-1$, and λ is the weight coefficient. If q_t is greater than the threshold ξ_R, the corresponding B_{tra}^{t} will be added to the matching sequence separately.

Step 3: After the matching of frame t fails, the subsequent output result will calculate the cosine distance s with the international feature set F. If $s < \xi_F$, and the tracking output B_{tra}^{t+f} meets the conditions set in formula 6, B_{tra}^{t+f} will be added to the matching sequence again. If the result of single target tracking is not output for 6 consecutive frames, the status will stop updating.

4 Implementation of Pedestrian Dynamic Recognition Algorithm

Dynamic pedestrian recognition is to identify the target pedestrian in the video sequence of the existing non-overlapping camera field of view, which belongs to the sub-problem of image retrieval [10]. The content of this chapter is based on the deep learning algorithm, which is described from three parts: the construction of the algorithm model, the establishment of the track scoring mechanism, and the motion matching and data association algorithm. Fig. 4 shows the ReID system diagram.

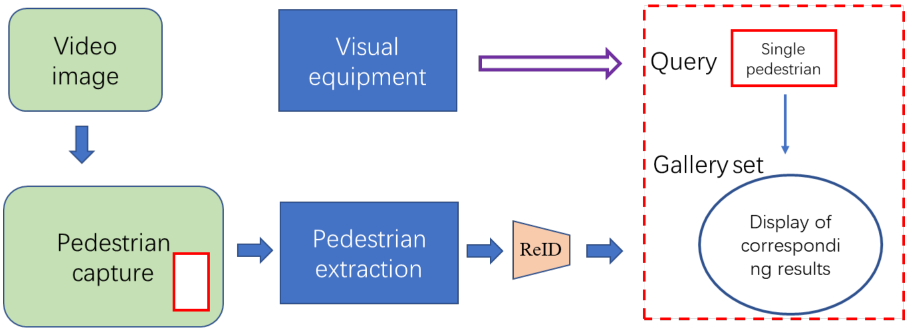

Fig. 4. ReID structure diagram

4.1 Construction of Algorithm Model

This paper uses the improved SSD [11] network as the backbone of the pedestrian dynamic identification network, as shown in Table 2. The inception-A module is integrated in the network, as shown in Table 3. Take all pedestrian image blocks belonging to the same identity as a category, and add a linear classifier $soft\max$ after the backbone network to calculate the probability score of the sample for each category.

Table 2. The improved SSD structure diagram

Type	Kernel/Stride	Output size	Padding
Input	/	$3\times224\times224$	-
Convolution	$7\times7/2$	$29\times112\times112$	3
Pool	$3\times3/2$	$29\times56\times56$	-
Convolution	$1\times1/1$	$27\times56\times56$	-
Convolution	$3\times3/1$	$142\times56\times56$	1
Pool	$3\times3/2$	$142\times28\times28$	-
Inception -A	/	$379\times28\times28$	-
Inception -A	/	$679\times28\times28$	-
Pool	$3\times3/2$	$679\times14\times14$	-
Inception -A	/	$1037\times18\times18$	-
Inception -A	/	$1002\times18\times18$	-
Inception -A	/	$938\times18\times18$	-
Inception -A	/	$861\times18\times18$	-
Pool	$14\times14/1$	$861\times1\times1$	-
Fc	/	256	-

Table 3. The inception -A structure diagram

Step	Branch A	Branch B	Branch C	Branch D
1	$Conv1\times1$	$Conv3\times3$	$Conv3\times3$	$Conv1\times1$
2		$Conv1\times1$	$Conv3\times3$	$Conv3\times3$
3			$Conv1\times1$	

The cross entropy loss function used in the optimization task is shown in formula 7, where M represents the number of categories, x and y represent the output of the network and the true value of the sample.

$$L_{reid}(x,y) = -\sum_{i=0}^{M-1} y[i]\log(x[i]) . \tag{7}$$

The cross entropy loss function used in the optimization task is shown in formula 7, where A represents the number of categories, B and C represent the output of the network and the true value of the sample.

Set the index set of frame K selected as the appearance of track history as $J = \{t_1, t_2, ..., t_K\}$, and the filtering rule is as shown in formula 8.

$$t_i = \underset{t-i\delta<t\leq t-(i-1)\delta}{\arg\max}\ Q(I_t^X), i = 1,2,...,K . \tag{8}$$

Q is the optimal measure used to calculate the detection results [12], and the output value range is [0, 1]. The loss function corresponding to its network parameters is shown in Formula 9. The training positive sample and the pedestrian target true value IoU must be greater than 0.5, otherwise it is set as negative sample. I_t^X represents the image area of track X at time t , and δ is the super parameter that determines the selection interval. The selected K images and the candidate results to be matched will be sent to the pedestrian dynamic recognition network to output their appearance features. At this time, the saved feature set in X is the long-term clue of the system. In order to save the calculation cost, the network only extracts the appearance features of each image in the track once, and saves the output for further use.

$$L_{cls}(x,y) = -(1-y)\log(1-x) - y\log(x) . \tag{9}$$

4.2 Establishment of Track Scoring Mechanism

In this paper, a concise and effective track scoring mechanism is proposed from the two dimensions of time and space to solve the problem of track priority evaluation in the cascade matching process. The track of the same pedestrian may be interrupted to form several sub-tracks. Each time the target is retrieved from the lost state, and the tracker under the process has been removed, the target tracker will be reinitialized, so the object participating in the evaluation is actually a set of sub-tracks.

On the time level, let L_{tra} represent the number of frames between the track matching failure and the current time, L_{can} represent the number of tracking results associated with the track, and α and β are the adjustment factors, so the scoring function is expressed as:

$$T_{tim} = \max(1-\log(1+\alpha \cdot L_{tra}),0) \cdot \min(\log(1+\beta \cdot L_{can}),1) . \tag{10}$$

At the spatial level, the scoring function considering the track spatial information is shown in Formula 11:

$$S_{qa} = \frac{\omega}{m} \cdot \sum\nolimits_{i=1}^{m} S_{ca}^{i} \cdot p_{can}^{i} + \frac{\theta}{n} \sum\nolimits_{i-1}^{n} S_{iou}^{i} \cdot p_{\det}^{i} . \tag{11}$$

Where, p_{can}^i represents the ith classifier score in T , $p_{\det}^i$ represents the detection classifier score corresponding to the ith result from IoU matching in T , S_{ca}^i and S_{iou}^i represent the cosine distance value and IoU matching value of the trajectory result at the matching time respectively, ω and θ are coefficients used to adjust the matching result and the importance of the matching result, and finally the integrated trajectory scoring equation is obtained through the weight coefficient ρ , from which the T_c priority order participating in the matching can be calculated, The formula is as follows:

$$V = S_{tim} + \rho \cdot S_{qa} . \tag{12}$$

4.3 Motion Matching and Data Association Algorithm

In this paper, Mahalanobi [13] distance is used to represent the distance between borders, and the matching degree between border boxes is calculated using formula 13:

$$d_{mov}(i, j) = (d_j - y_i)^T S_i^{-1}(d_j - y_i) . \tag{13}$$

Where, $d_{mov}\ (i, j)$ is the matching degree between the candidate result j and the track i , S_i is the covariance matrix of the observation space at the current time based on the existing results in T_c , and any element S_{ij} is obtained by calculating the formula, which is expressed as follows:

$$S_{ij} = \mathrm{cov}(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] . \tag{14}$$

Where, $\mu = E(x_i)$ represents the expected value of the i element in the state vector, y_i is the final tracking result of T_c at moment $t-1$ after the previous frame is successfully matched, and d_j is the candidate result of j .

The sparse optical flow is used to estimate the motion between frames, and the solved rotation translation matrix is used for the motion compensation of video frames. First, it is assumed that there is a homography relationship between two consecutive frames. Secondly, two strategies are mixed to ensure that enough matching point pairs can be obtained:

1) Features that are robust to motion blur are extracted from video frames, and feature points are matched using the nearest neighbor rule.

2) Based on the good-features-to-track criterion [14], the minimum eigenvalues of the autocorrelation matrix of the sampling points are thresholded to screen out significant feature points.

3) The homography transformation matrix between adjacent frames is estimated to correct the image offset caused by camera motion.

Use formula 15 to calculate the appearance matching between the depth feature r_j of the candidate result and the feature set R_k of the track k . The formula is as follows:

$$d_{app}(i, j) = \min\{1 - r_j^T r_k^{(i)} \,\big|\, r_k^{(i)} \in R_k\} . \tag{15}$$

Use the improved Hungarian algorithm [15] to solve the data association problem under Multi-Object Tracking, and make full use of the motion, appearance and spatial information contained in the scoring matrix generated by cascade matching or IoU matching to improve the data association quality. The algorithm is expressed as follows:

$$f_x[u] + f_y[v] >= \sigma \cdot e(u, v) . \tag{16}$$

Where, $f_x[i]$ represents the top mark of the left point and $f_y[i]$ represents the top mark of the right point. σ represents the weight coefficient, $e(u, v)$ represents any edge of the recognition frame. The Hungarian algorithm flow is as follows:

Step 1: Initialize the value of feasible top mark, that is, set the initial value of $f_x[i]$ and $f_y[i]$;

Step 2: The Hungarian algorithm is used to find complete matching of equal subgraphs;

Step 3: If the augmentation value is not found, modify the value of the feasible top mark;

Step 4: Repeat steps 2 and 3 until the result is matched, and complete the data association.

5 Experimental Results and Analysis

Software environment: Ubuntu 16.04 64-bit operating system, PyTorch in-depth learning framework, CUDA 9.1, cuDNN 7.1, Python 3.8.0.

Hardware environment: Intel (R) Core (TM) i7-7700 CPU @ 3.60GHz processor, 32GB memory; NVIDIA GeForce GTX 1080Ti GPU, 11GB.

Use the YouTube-BoundingBoxes data set as the training set, which is composed of 380000 video clips of 15 to 20 seconds, and contains 23 categories of 5 million manually labeled single target bounding boxes [15].

5.1 Performance Evaluation Index

1) Multi-target tracking accuracy (MOTA):

$$\text{MOTA} = 1 - \frac{\sum_t FP_t + FN_t + IDS_t}{\sum_t G_t} . \tag{17}$$

Where, G_t is the number of tracks marked in frame t, FN_t is the number of missed detections, FP_t is the number of false alarms, and IDS_t is the number of times target identity switching IDS occurs in frame t.

2) Target identification precision (IDF1):

$$\text{IDF}_1 = \frac{\text{2IDTP}}{\text{2IDTP+IDFP+IDFN}} . \tag{18}$$

Where, TP_t is the number of correct matches and d_t^i is the measure between the i correct matches.

3) The number of successfully tracked targets (MT): the number of correctly matched targets with more than 80% of the real labeled tracks.

5.2 Quantitative Analysis

In order to facilitate comparison, the algorithm proposed in this paper is later called MN-SRPN, and different algorithms are used to compare on the YouTube BoundingBoxes dataset. The comparison results are shown in Fig. 5.

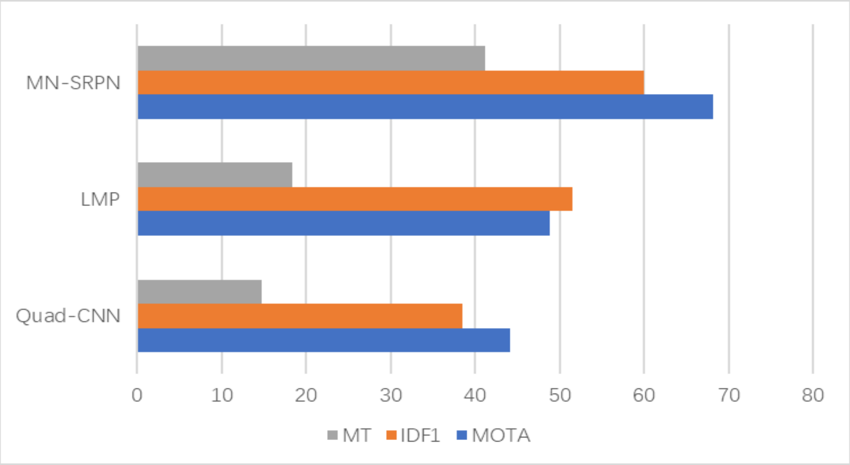

Fig. 5. Algorithm data comparison

Through comparison, we can see that the algorithm proposed in this paper is the best of all algorithms, and the accuracy of multi-target tracking has reached a high level, 20 percentage points higher than the second algorithm. The target identification accuracy is at least 9 percentage points higher than other algorithms. Finally, the number of targets that can be successfully tracked is not the highest, but it is also at a high level. Therefore, the performance of the algorithm proposed in this paper has been verified.

5.3 Qualitative Analysis

Qualitatively analyze the tracking results of MN-SRPN on the YouTube-BoundingBoxes dataset, as shown in Fig. 6, the visualization effect of the model at different times in three test video clips.

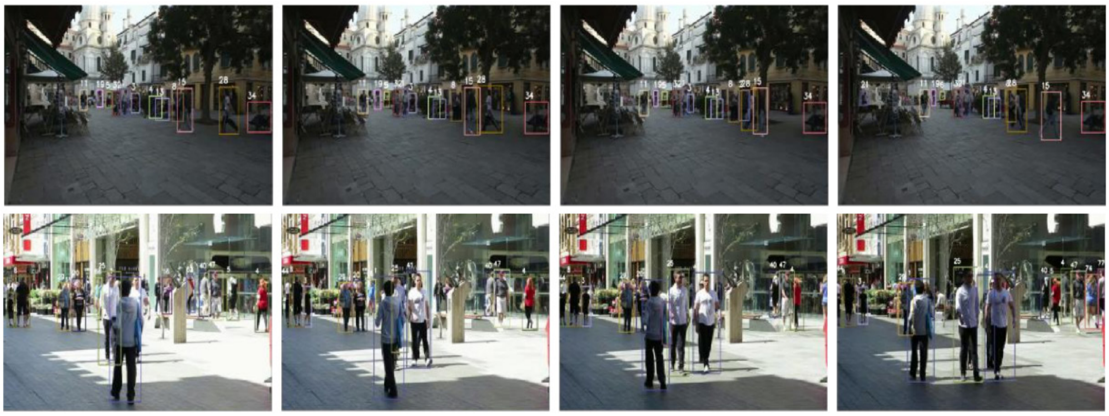

Fig. 6. Experimental result

The results show that although there are different levels of target interleaving in the two video sequences, the proposed algorithm has strong robustness for this situation, and the algorithm can still maintain high pedestrian recognition accuracy for the objects that reappear after occlusion. The lighting conditions in different video sequences differ significantly, involving indoor and outdoor scenes respectively. The visualization effect of the proposed Multi-Object Tracking tracker also verifies its tracking stability under different lighting conditions.

6 Conclusion

This paper discusses an algorithm of dynamic recognition and tracking technology in the automatic driving scene, and designs the algorithm structure. The algorithm can realize accurate recognition and tracking of specific targets. With dynamic pedestrians as the target object, it can realize the recognition and tracking of dynamic pedestrians, and can realize the prediction of pedestrian behavior. In order to verify the performance of the algorithm proposed in this paper, the algorithm is tested in this paper. Through comparison with other algorithms and tracking of tasks in video, the comprehensive performance of the algorithm is due to other algorithms.

In the next step, this paper will take dynamic vehicles and other dynamic factors into account, further improve the recognition and tracking of intelligent vehicles and automatic driving for more dynamic factors, and improve the ability of recognition and tracking.

References

[1] V. Kumar, S.C. Subramanian, R. Rajamani, A novel algorithm to track closely spaced road vehicles using a low density flash lidar, Signal Processing 191(2022) 108360.

[2] P. M. Raju, D. Mishra, P. Mukherjee, DA-SACOT: Domain adaptive-segmentation guided attention for correlation

based object tracking, Image and Vision Computing 112(2021) 104215.

[3] E.C. de Castro, E.O.T. Salles, P.M. Ciarelli, A New Approach to Enhanced Swarm Intelligence Applied to Video Target Tracking, Sensors 21(5)(2021) 1903.

[4] W. Farag, Real-time lidar and radar fusion for road-objects detection and tracking, international Journal of Computational Science and Engineering 24(5)(2021) 517-529.

[5] Z.-L. Liu, C. Wang, Target tracking algorithm in Siamese network based on bimodal input, Application Research of Computers 38(12)(2021) 3796-3800.

[6] H.-Y. Zhang, P.-Y. He, H.-S. Wang, A real-time target-tracking algorithm based on improved siamFC, Laser & Optoelectronics Progress 58(6)(2021) 0615003.

[7] J.-S. Zheng, H. Guo, A.-B. Li, J.-B. An, Real-Time Tracking of Fast Moving Weak Object Based on Siamese Network, Laser & Optoelectronics Progress 59(4)(2022) 0140011-1-0410011-9.

[8] Y.-J. Hama W.-C. Ren, C. Zhang, H. Zhang, Defect recognition of diode glass shells based on lightweight network MobileNet V2*, Transducer and Microsystem Technologies 41(4)(2022) 153-155.

[9] H. Han, J.-F. Lu, An Improved SiamRPN Method Based on Actor-Critic Sequential Pre-positioning Method, Computer & Digital Engineering 49(11)(2021) 2222-2228.

[10] F.-L. Zheng, Research on Pedestrian Re-identification Algorithm Based on Dynamic Image, China Computer & Communication 3(2019) 52-53.

[11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: Proc. European conference on computer vision, 2016.

[12] S. Hare, S. Golodetz, A. Saffari, V. Vineet, M.-M. Cheng, S.L. Hicks, P.H.S. Torr, Struck: Structured Output Tracking with Kernels, IEEE Transactions on Pattern Analysis and Machine Intelligence 38(10)(2016) 2096-2109.

[13] K.-Y. Wang, H.-W. Wang, Q. Zhao, S.-S. Wang, A modified Mahalanobis distance discriminant method, Journal of Beijing University of Aeronautics and Astronautics 48(5)(2022) 824-830.

[14] H. Zheng, S.-A. Chula. Tomasi, Algorithm of multi-feature track association based on topology, Command Information System and Technology 11(5)(2020) 83-88.

[15] Y. Song, C. Ma, X. Wu, L. Gong, L. Bao, W. Zuo, C. Shen, R.W.H. Lau, M.-H. Yang, VITAL: visual tracking via adversarial learning, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2018.