International Journal of Computational Intelligence Systems and Applications. Volume 8, Issue 1, (2021) 1-11
http://www.cayley-nielson.com/ijcisa/

A Survey of Multi-Target Detection and Tracking Algorithms

Mengxin Li, Jiali Hao, Songang Li, Yu Han, Yiying Li
School of Information and Control Engineering, Shenyang Jianzhu University, Shenyang 110168, China
E-mail: *846007599@qq.com*

Abstract

Multi-target detection and tracking is the technical basis for vehicle behavior dialysis and identification [1]. The advantages and disadvantages of multi-target detection technology and multi-target tracking technology are described in this paper in recent years and researchers presented many methods in order to overcome their shortcomings. The performance of multi-target detection and tracking technologies are compared. Finally, the future development direction of these two technologies are discussed.

Keywords: Multi-Target Detection, Multi-Target Tracking, SSD Algorithm, RetinaNet Algorithm, YOLO Algorithms.

1. Introduction

The two technologies of detection and tracking are critical to the entire collection and tracking process. Generally speaking, the process of searching and tracking can be roughly divided into four categories: tracking the area of interest, tracking the target, completing the route tracking in the area of interest, and taking the target as the target [2]. Tracking goals can help identification by providing appropriate goals and providing a solid foundation for subsequent behavioral analysis. If the detection is unsuccessful, you can complete the lost target through tracking, and be strong for a cycle. The simple flow chart of multi-target detection and tracking is shown in Figure 1[3].

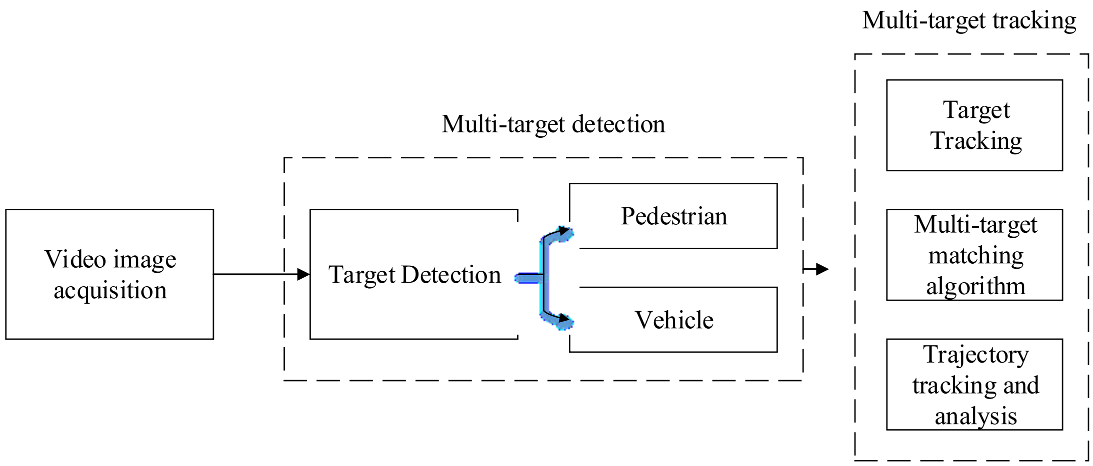

Figure 1 Block diagram of multi-target detection and tracking

2. Multi-target detection algorithms

The function of multi-target detection is to extract the object of interest from the video, and to understand and study it in the follow-up. Traditional multi-target detection is the process of preprocessing

ISSN 2331-1924 /
May, 2021

the target of interest, then feature extraction and classification, and finally the detection process; and the multi-target detection algorithm introduced next is based on deep learning, which emerged in recent years. Multi-target detection algorithms are the Single Shot MultiBox Detector (SSD) algorithm, RetinaNet algorithm, and You Only Look Once (YOLO) series of algorithms.

2.1. SSD algorithm

2.1.1. SSD algorithm introduction

In the SSD algorithm, the image can be compiled, the image is allowed to use the neural network to extract annotation features, so that the feature map is created and generated, and then 6-layer feature maps are extracted, a default box is created for each feature map, and finally all default boxes is created. Everything is merged, everything is thrown to non max suppression（NMS）, the default box after filtering is the product, and the detection function has been completed.

In the SSD algorithm, optimized views are used together with multi-level mapping to partition and resize. First, the original image is scaled to a fixed size, and then the map of the target product is extracted through the network SSD, and a map with the number of pixels equal to many fixed-size pixels is obtained. Compared with the text box, the first set of boxes and pre-established boxes composed of different numbers are compared with the "Intersection Point" (IOU) through length and width, and finally the bounding box is adjusted under extreme suppression, and the final positioning is completed.

Its loss function is mainly divided into two parts: one is to calculate the corresponding default box; the other is the confidence loss of the target category and the corresponding position regression. The source of the loss function is shown in Figure 2.

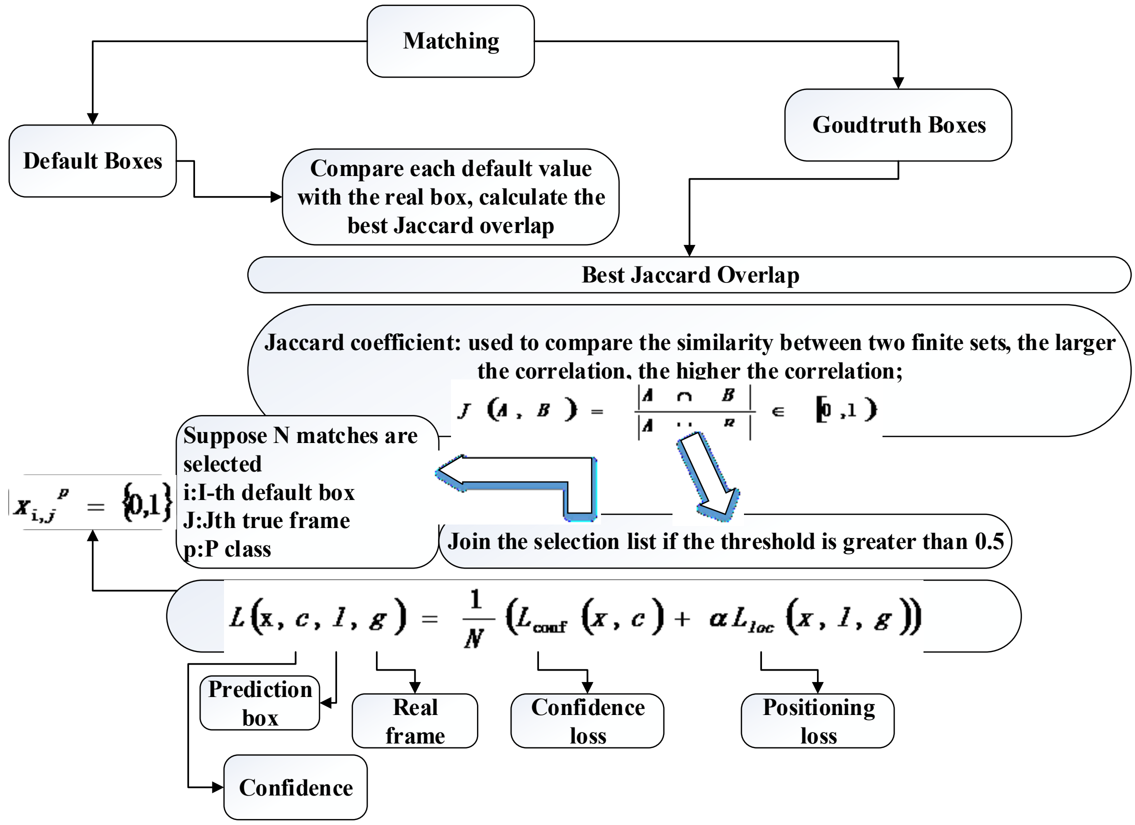

Figure 2 source graph of loss function

Among them, The default and real frames of the original image are represented by A and B; the i-th default value is represented by i, the j-th real frame is represented by j, and the p-th class is represented by p; the confidence is represented by c, and the predicted frame is represented by i, the ground truth box is represented by g, the default number of boxes from matching to the ground truth value is represented by N, the confidence loss is represented by conf, and the location loss is represented by loc.

2.1.2. Advantages and disadvantages of SSD algorithm

The advantage of the SSD algorithm is its speed. Only one step is required in the whole process. After the image is recruited and sampled, the extracted features can be used for direct classification and regression. However, its biggest drawback is that small targets are missed or undetected, which leads to errors, and due to its dense sampling, uneven positive and negative sampling is generated, which makes training difficult and reduces the accuracy of the model.

Some people have made some improvements on the shortcomings of the SSD algorithm. Li G.J. and others [4] have proposed a new and improved Inception module for the missed detection of the SSD algorithm, and replaced it with the Conv8, Conv9 and Conv10 layers in the SSD network; Liu D. and others [5] have proposed to improve the a priori frame ratio parameters and network structure of SSD for its poor obstacle detection effect; Zhou Q. and others [6] have proposed the use of Mobilenetv2 to lighten its real-time performance and accuracy. The network is used as the SSD backbone network, and the network model is scaled and the activation function is improved; Ruan S.F. and others [7] have proposed an integrated network model of set feature extraction and detection and classification on the issue of pose correction, feature extraction, and classifier. An improved non-maximum suppression processing layer is added at the end; Dong Y.C. and others [8] have proposed DenseNet as the basic network of SSD for its difficult training and low detection accuracy, and then added four layers of convolutional layers to build in the new network, in order to make full use of the information of different depth convolutional layers, the last four layers of the new network and the last two DenseBlocks in DenseNet are taken to extract the target frame.

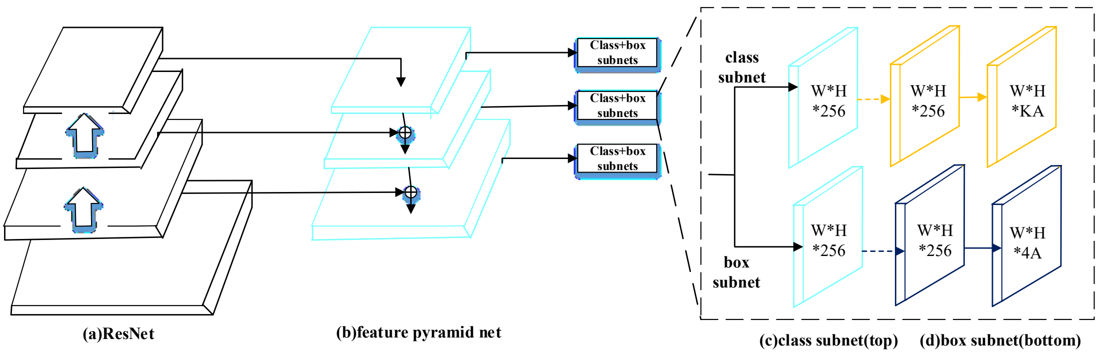

Figure 3 RetinaNet algorithm network diagram

2.2. RetinaNet algorithm

2.2.1. Introduction to RetinaNet algorithm

The RetinaNet algorithm is derived from the 2018 Facebook AI Research Paper Focal Loss for Dense Object Detection, which is presented by Ross G G, Kaiming G G, and Piotr G. G [9]. The biggest contribution of this paper is that the focus loss (Focal Loss) is proposed to solve the problem of category imbalance, thus RetinaNet (first-level target detection algorithm) is created, and the accuracy of the classic second-level Faster-RCNN target detection network is exceeded.

The RetinaNet algorithm has a backbone network and two subnets. The feature pyramid is largely dependent on the backbone network, and the image attributes are extracted by the backbone network. The first subnet is used for zoning, and the second subnet is used for boundary adjustment. The entire network is much simpler than the fast RCNN, which is mainly created by the ResNet + FPN + 2xFCN network. The network diagram is shown in Figure 3.

2.2.2. The advantages and disadvantages of the RetinaNet algorithm

One of the advantages using the RetinaNet algorithm is the structure of multi-layer pyramid. With the existence of this multi-layered pyramid, the problem of sample-level inequality can be solved and the detection accuracy can be improved, and at the same time, multiple predictions at the entire network level can be improved. But its biggest disadvantage is that many useless bounding boxes are deeply discovered by it, so the burden of counting and storage will be increased.

Regarding the shortcomings of the RetinaNet algorithm, some people have proposed a lot of improvement measures. For example, Qiu L. and others[9] have proposed an improved RetinaNet network that uses a feature pyramid network to generate multi-scale feature maps and appropriately reduces the convolutional layer to improve its stability and accuracy; Ming H.Y. and others [10] have proposed the use of Gaussian mixture clustering method based on EM algorithm to solve the ambiguity detection to improve the counting accuracy on the problem that the original RetinaNet framework is not ideal for the detection of dense targets; Wu H.Y. and others[11] have proposed to recombine the standard ResNet50 network and introduce the expansion convolution module to expand the receptive field of the feature map on the problem of improving the detection accuracy of the algorithm; Wang L.L. and other[12] shave proposed the accuracy problem. Through the dimensional clustering algorithm, the optimal size of the anchor point is found, and the appropriate feature map is found for prediction.

2.3. YOLO algorithms

2.3.1. Introduction to YOLO Algorithms

In the field of deep learning, YOLO algorithms are also one of the most popular multi-target detection algorithms. The following is an introduction to YOLO algorithms.

In YOLOv1, when performing detection, the image is first split into a suitable S×S grid [13]. If the center of the object is on a specific grid, that grid is responsible for finding the object. For each grid, in YOLO, B bound boxes are estimated. In YOLOv1, 2 bound boxes are estimated, and for each bound box, 5 values are estimated, where the position of the bound box is represented by 4, and the confidence of the bound box is represented by one.

The detection process of YOLOv2 is basically the same as that of YOLOv1. In YOLOv2, the following aspects are mainly improved, Structurally, the new Darknet-19 feature network is used, a regular layer is added to all convolutional layers, the fully aligned YOLOv1 is deleted, and an integrated layer connecting the top and bottom feature maps is added. With the improvement of accuracy, the adjustment of the estimated value is the previous frame obtained by the multi-scale training method.

In order to improve the position and accuracy of target recognition, in YOLOv3, a deep neural network is designed, such as the backbone network of YOLOv3, which has 53 types of convolutional layers, called Darknet-53[14]. The feature pyramid scale multi-scale feature network design function is used to improve the detection results of small targets.

In YOLOv4, the improvement direction of the target detection model is divided into Bag of freebies (Bo F) and Bag of Specials (Bo S) [15]. In Bo F, its training plan has only been changed or training costs have been increased. Improving data is one of the common methods for the goals defined by Bo F. In Bo S, the reasoning cost of a small number of networks will only be increased, but the modules and post-processing methods that can make the algorithm accuracy are significantly improved. In target detection, the commonly used methods that meet the Bo S definition include expanding the receptive field (such as SPP), Introduce attention mechanism, enhance communication between features (such as FPN), etc. In YOLOv4, CSP Darknet53 is used as its backbone network. And SPP and PAN are introduced into the spine network to achieve signal compatibility, and a set of Bo F and Bo S functions are added.

In the YOLOv5 algorithm, the input terminal, backbone, neck and prediction are mainly included. In the input terminal, data enhancement and adaptive anchor frame calculation are mainly used. In the Backbone part, the Focus structure and the CSP structure are mainly included, and in the Neck part, the FPN and PAN structure is mainly included. In the Prediction part, the GIOU-Loss function is mainly used.

2.3.2. The advantages and disadvantages of YOLO algorithms

The YOLO algorithm is gradually developed. Their advantage is that through data enhancement, a better path can be provided for detecting small target objects. Through adaptive anchor frame calculation, the prediction result can be made more reasonable. Through this structure, the image can be sliced better, the problem of large amount of calculation in network design can be solved, and the integration of network functions can be strengthened. But its loss function is one of their biggest flaws, faster convergence speed and better performance cannot be obtained.

For these shortcomings, people have proposed improvements. Jiang W.Z. and others [16] have proposed that YOLOv3 error detection and missed detection problems are based on the original network structure of YOLOv3. An additional output scale is derived from the backbone network, and it is compared with the above. The feature information in an output scale is used for feature splicing; Zhang W. and others[17] have proposed the use of a new feature fusion method in the YOLOv4 model and the use of loss function weighting methods to constrain the update of weights and biases for the low detection efficiency and time-consuming issues of YOLOv4 ; Zhang Q.L. and others[18] have proposed to improve the target detection effect by introducing data enhancement and label smoothing methods, improving the loss function to DIOU and adding a network processing layer for small targets on the problem of YOLOv5s low detection accuracy.

3. Multi-target tracking algorithm

In multi-target tracking, the target position is estimated based on the current detected target information and the subsequent continuous image frames. Multi-target tracking is well reflected in many timeliness issues. At the same time, it also has a lot of target tracking based on its mainstream framework, such as the method based on iterative prediction and the framework based on feature point matching. These are all based on Target tracking is done on a traditional basis. But the following three types of multi-target tracking algorithms with their unique points are the DeepSORT tracking algorithm, the STC tracking algorithm and the TLD tracking algorithm.

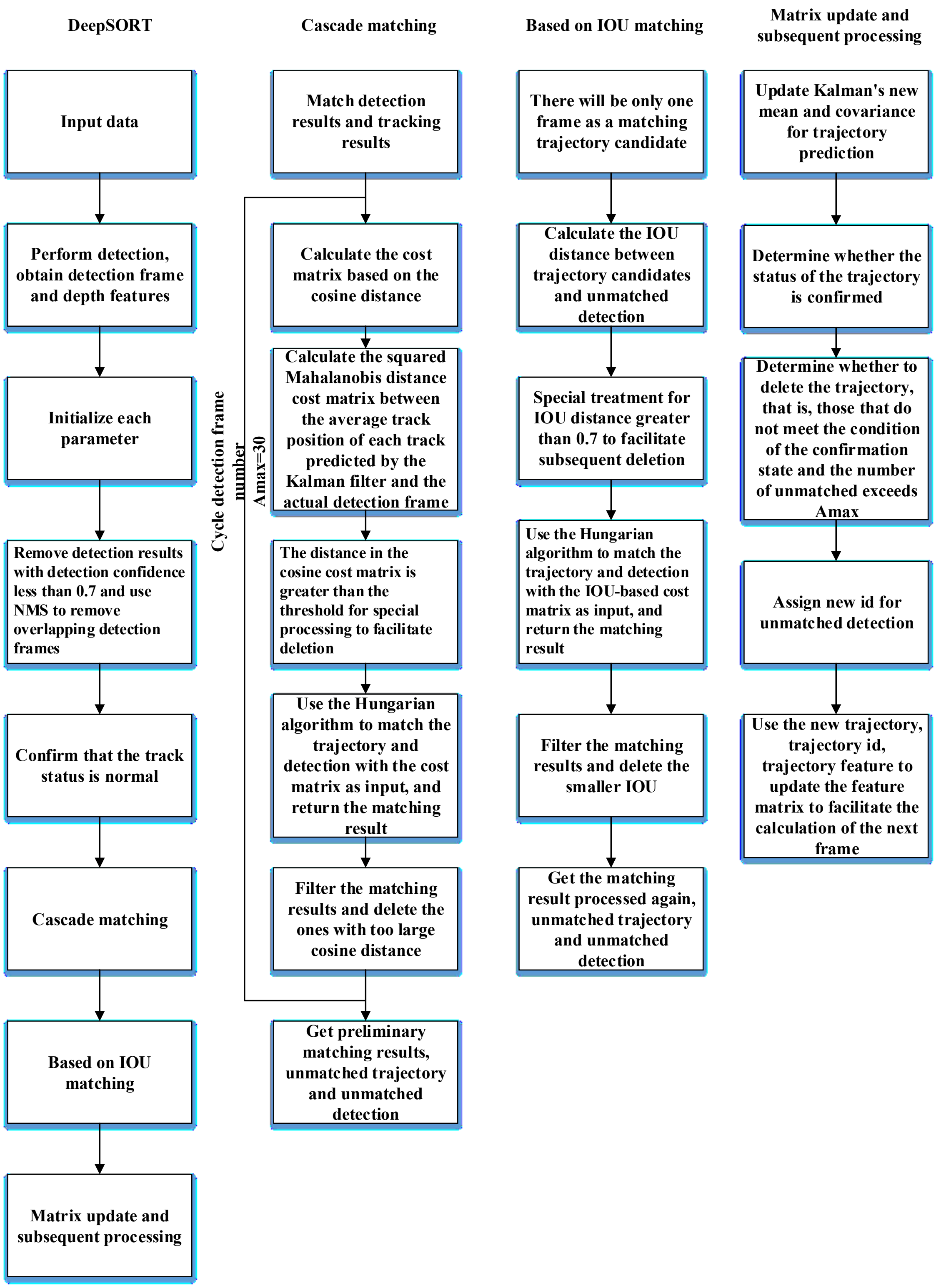

Figure 4 DeepSORT tracking algorithm

3.1. DeepSORT tracking algorithm

3.1.1. Introduction to DeepSORT tracking algorithm

The main direction of target tracking methods has gradually tended to target tracking based on detection, and then single-camera multi-target tracking is introduced through multi-hypothesis tracking and joint probabilistic data association filters. These methods carry out frame-by-frame data association, but they are complicated. The degree is relatively large and not easy to achieve. Subsequently, simple online and real-time tracking is proposed. The Hungarian algorithm is used to measure the relationship between the predicted trajectory and the tracked target, and the degree of association is used as a measure of bounding box overlap. Compared with the previous algorithm, performance has been achieved. Great improvement. However, the SORT algorithm has a large number of identity conversions. For this reason, later generations improved Deep-Sort on its basis. In Deep-Sort, the idea of deep learning is introduced, mainly because the features are extracted in the convolutional neural network, and in the tracking process, the recursive Kalman filter is used to predict the Markov between the newly arrived state and the state Distance and cosine distance. The allocation problem is solved by jointly measuring the relationship between the combined index and the threshold, and the sub-problems in the global allocation are solved by using cascade matching.

In the DeepSORT algorithm, first the data is passed in, and then it is found to find the detection frame and depth feature, and then the initialization of each parameter is started, the detection result is deleted with a confidence of less than 0.7, and the detection frame is covered by a non-maximum suppression algorithm to verify its normal movement. Then it is measured, its IOU comparison is performed, and finally the matrix analysis and traceability are executed to complete the trace operation [19].

The main process of the DeepSORT algorithm includes cascade matching, IOU-based matching, matrix update and subsequent processing, as shown in Figure 4.

3.1.2. Advantages and disadvantages of DeepSORT tracking algorithm

The advantage of the DeepSORT algorithm is that in addition to the use of dynamic information, the robustness of losses and obstacles has also been increased, and the obvious similarity and whether the similarity can be used as the standard for the same target is calculated. However, its disadvantage is that its robustness is affected by other factors. For example, If there is an error corresponding to the first tracking framework, it is more complicated, and in the operation process of the algorithm, inconsistencies often appear, which will affect the robustness of the tracking.

Regarding the defects of the DeepSORT algorithm, some people have proposed many improved methods. For example, Zhang M.H. and others [19] have proposed a method based on multi-feature and unscented Kalman filter (Unscented Kalman Filter, UKF.) Fusion of the DeepSORT algorithm; Li Z.X. and others [20] have proposed the introduction of LSTM motion model based on the DeepSORT tracking algorithm for its large prediction error; Wu M.Q. and others [21] have proposed the use of the DeepSORT single hypothesis tracking matching framework composed of Kalman filtering and Hungary algorithm to predict and preliminary match the target trajectory on the problem of trajectory matching confusion caused by occlusion between targets.

3.2. STC tracking algorithm

3.2.1. STC tracking algorithm introduction

In visual tracking, the location of the location includes the origin of the target and the specific location around it. This is because, in fact, there is a strong temporary relationship between consecutive frames and the position around the target. For example, the focus of the image above is very close, causing a major change in the appearance of the target. However, since only a small part of the enclosed area is enclosed, the top area and the query area remain the same, and the target situation does not change much.

Therefore, the position of the current frame helps to estimate the target area in the next frame.

In the STC tracking algorithm, a Bayesian framework is used to relax the temporal and spatial relationship between the target and the local content, so as to obtain a lower-level statistical correlation between the feature and the surrounding area. Then the spatio-temporal relationship and focus of attention function are integrated into the biological vision system to evaluate the confidence map of the position that appears in the new frame. The most confidence is the target position of the new frame.

The process of algorithm implementation [22]:

t frame: according to the frame image l and the target image obtained X^*.

(1) Learn space context model.

(2) Update the spatiotemporal context model needed to track the target in the next frame.

(3) Update parameters such as scale.

t+1 frame:

(1) Calculate the confidence map.

(2) Find the maximum value, the position of this maximum value is the target position we require.

3.2.2. Advantages and disadvantages of STC tracking algorithm

In the STC tracking algorithm, its advantage is that it is realized by Fourier transform, and the learning and detection speed is faster, but the biggest problem is that its target window cannot adapt to the change of target ratio, which leads to errors. target setting.

Regarding the shortcomings of the STC tracking algorithm, some people have proposed other improvement measures. Zheng H.L. and others[22] have proposed a spatiotemporal context tracking algorithm that incorporates color histogram responses for problems such as sudden target deformation and partial occlusion; Wu D.H. and others[23] have proposed a spatiotemporal context target tracking algorithm based on scale filter on tracking drift and other issues; Chen F.L. and others[24] have proposed an anti-occlusion real-time target tracking algorithm using spatiotemporal context for tracking stability degradation and failure.

3.3. TLD tracking algorithm

3.3.1. Introduction to TLD tracking algorithm

An algorithm for long-term tracking of objects in video is proposed by Zdenek Kalal [25], In this algorithm, first the tracking target is manually selected in the initial frame, and then the tracker and the detector are tracked separately, and then the results of the two are synthesized through the synthesis module Output. And the tracker and the detector are calibrated by applying a learning algorithm.

In the TLD tracking algorithm, it is a long-term tracking algorithm, which includes three parts, one is the tracker, one is the learning module, and the other is the detector. The previous trajectory and the next moment of the current point of the tracked current point are predicted, and then the backtracking of the next moment is completed by using the same principle, the backward trajectory from the next moment to the previous moment is predicted, and the error is calculated, and the search Points and discarded points, and the position of the target frame is predicted by successfully tracking the coordinate distribution of the points. Its basic block diagram is shown as in Fig. 5.

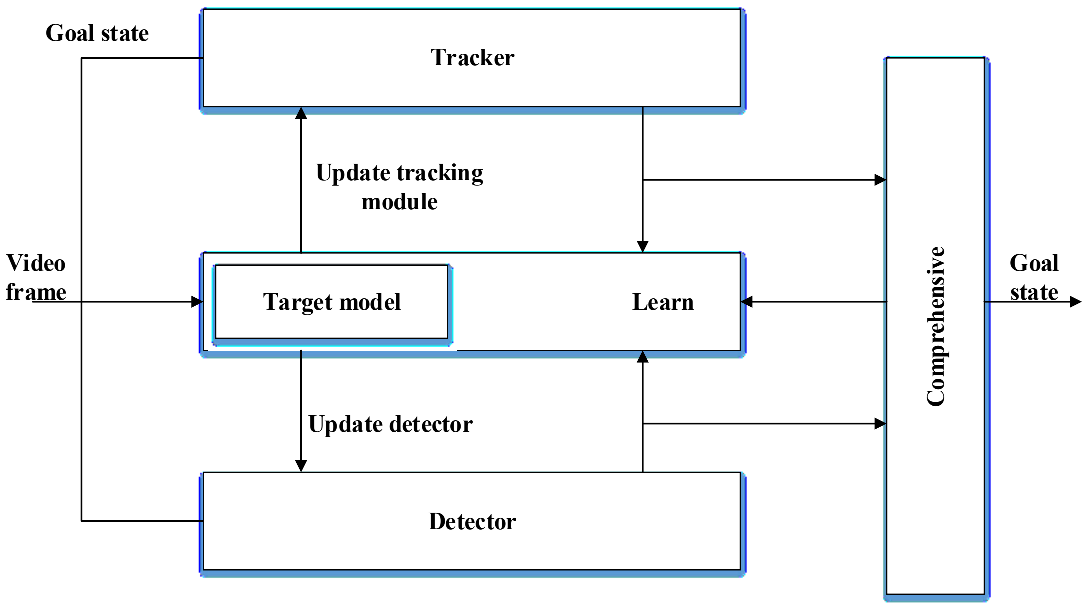

Figure 5 TLD tracking algorithm

3.3.2. Advantages and disadvantages of TLD tracking algorithm

The advantage of the TLD tracking algorithm is that it has a small amount of calculation, good real-time performance, and the target stably can be tracked in a simple scene, but its disadvantage is that if the tracking target disappears from the scene and reappears, it will not be tracked.

Regarding the shortcomings of the TLD tracking algorithm, some people has improved it. Zhong X. and others[25] have proposed an improved SIFT feature matching algorithm combined with the optical flow method in the TLD algorithm for its low accuracy; Zhang J. and others[26] have proposed a TLD real-time target tracking algorithm that is improved by fusion CN tracking algorithm for the problem of tracking drift in scenarios such as target non-rigid deformation, rotation, and background clutter; Fan M. and others[27] have proposed an improved TLD method for the problem of low tracking recognition rate. In the tracking module, the pyramid optical flow method is introduced to construct the pyramid image.

4. Development trend

4.1. Multi-target detection algorithm

In the traditional and deep learning frameworks, the multi-target detection algorithm has problems such as error detection, missed detection, and low accuracy and stability [28]. People thought of combining the target detection algorithm with other algorithms or improving the detection algorithm. Part of the link is to complete these problems in the multi-target detection algorithm. Multi-target detection algorithms have achieved many important results in many fields. For example, in the field of car assisted driving, which is widely concerned, multi-target detection can accurately detect people, vehicles, road signs and other information around the body, real-time alarms, etc [29]. Improvement can bring more applications, more perfect perfection brings more accurate detection, challenges and difficulties will exist, we need to solve and discover.

4.2. Multi-target tracking algorithm

There are still many problems in multi-target tracking, such as tracking occlusion, target deformation, and target drift [30]. Therefore, people have proposed many improved methods of multi-target tracking. They improved the multi-target tracking model or one of them. Loss function and other aspects to improve the shortcomings of multi-target tracking [31-32]. Multi-target tracking is becoming more and more novel with the development of the times. It is applied in many advanced fields. The derivation of new technologies has brought convenience to the society, but there are still many aspects of multi-target tracking that need to be better embodied and improved. There are many challenges and difficulties that require us to continue to study.

5. Conclusion

Moving target detection and tracking is an important part of computer vision technology. It is currently used in many aspects and is closely related to people's daily life. When studying target detection and tracking algorithms, the convergence and robustness of target detection and tracking are still difficult problems to be solved. After years of research, many algorithms based on deep learning have slowly developed. At present, many emerging algorithms are sprouting, and there are many development prospects for multi-target detection and tracking.

References

[1] Pietro S, Federico P, Marco C, Giuseppe L, Alberto D.B, Vittorio M. Non-myopic information theoretic sensor management of a single pan–tilt–zoom camera for multiple object detection and tracking[J]. Computer Vision and Image Understanding,2015, 134:

[2] Carlos R. del-Blanco, Fernando J, Narciso G. An efficient multiple object detection and tracking framework for automatic counting and video surveillance applications[J]. IEEE Transactions on Consumer Electronics,2012,58(3):

[3] Sualeh M, Kim G.W. Dynamic Multi-LiDAR Based Multiple Object Detection and Tracking. [J]. Sensors (Basel, Switzerland),2019,19(6):

[4] Li G.J, Hu J, Ai J.Y. Vehicle detection based on improved SSD algorithm[J]. Computer Engineering:1-15.

[5] Liu D, Tang W, Chen Z.H, Zhao C. Improved SSD algorithm for multi-scale obstacle detection of blind people in outdoor travel[J]. Journal of Jiangxi University of Science and Technology, 2021, 42(01): 87-97.

[6] Zhou Q, Xia M.Y. Research on multi-UAV detection based on improved SSD algorithm[J]. Information Technology, 2020, 44(12): 71-76.

[7] Ruan S.F. Research on Pedestrian Wearing Mask Detection Based on Improved SSD Algorithm[J]. Journal of Science & Technology Economics Guide, 2020, 28(35): 9-13.

[8] Dong Y.C, Shan Y.G, Yuan J. Pedestrian detection method based on improved SSD algorithm[J]. Computer Engineering and Design, 2020, 41(10): 2921-2926.

[9] Qiu L, Ren D.J, Gao M, Fu L, Wu H.Y, Hu Y.Q. Medical plastic bottle packing and counting algorithm based on improved RetinaNet[J]. Computer and Modernization, 2020, (12): 99-103+111.

[10] Ming H.Y, Chen C.M, Liu G.H, Deng H. An improved algorithm for dense rebar counting based on RetinaNet[J]. Sensors and Microsystems, 2020, 39(12): 115-118.

[11] Wu H.Y, Ren D.J, Lu Y.Z, Hu B, Fu L, Qiu L. Based on improved RetinaNet medical empty bottle surface bubble detection[J]. Journal of Sichuan University (Natural Science Edition), 2020, 57(06): 1090 -1095.

[12] Wang L.L, Zhang W, Sun Q.L. An improved RetinaNet indoor occupant detection algorithm[J]. Journal of Xidian University, 2019, 46(05): 69-74+104.

[13] Moechammad S, Cahya R, Berkah A N A. Detecting body parts from natural disaster victims using You Only Look Once (YOLO)[J]. IOP Conference Series: Materials Science and Engineering,2021,1073(1).

[14] Kamaru Z.F.H, Che A.S.A, Razak N. A, Johari J, Pasya I, Abu K.K.A. Visual-Based Motorcycle Detection using You Only Look Once (YOLO) Deep Network[J]. IOP Conference Series: Materials Science and Engineering, 2021, 1051(1).

[15] Bochkovskiy A, Wang C, Liao H. YOLOv4: Optimal Speed and Accuracy of Object Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 198-215.

[16] Jiang W.Z, Li B.Z, Gu J.J, Liu K. Ship target detection algorithm based on improved YOLO V3 [J]. Electro-Optics and Control:1-5.

[17] Zhang W, Wang M.S, Li Y. Server component detection algorithm based on improved Yolov4 [J]. Information Technology, 2021, (02): 98-102+108.

[18] Zhang Q.L, Lin Q.P, Xiao L. Improved aerial image recognition algorithm of YOLOv5 [J]. Yangtze River Information and Communication, 2021,34(03):73-76.

[19] Zhang M.H. Research on Highly Robust Multi-Pedestrian Tracking Algorithm Based on DeepSort Framework [D]. Harbin Institute of Technology, 2020.

[20] Li Z.X, Sun Wei, Liu M.M, Zheng L.L, Chen S.Y. Research on vehicle detection and tracking algorithms in traffic monitoring scenes[J]. Computer Engineering and Applications, 2021, 57(08): 103-111.

[21] Wu M.Q, Liu J.Q. Research on the problem of single hypothesis multi-target tracking method based on DeepSORT[J]. Information and Communication, 2020, (11): 40-42.

[22] Zheng H.L, Lin B, Wang H.T, Wang Z.Q, Wu W.C. Spatio-temporal context tracking algorithm fused with color histogram response[J]. Journal of Terahertz Science and Electronic Information, 2020, 18(03): 509-514.

[23] Wu D.H, Zhu E.C, Xu L. Spatial-temporal context target tracking algorithm based on scale filter[J]. China Water Transport (second half of the month), 2020, 20(06): 72-74.

[24] Chen F.L, Ding Q.H, Luo H.B, Hui B, Chang Z, Liu Y.P. Anti-occlusion real-time target tracking using spatio-temporal context[J]. Infrared and Laser Engineering, 2021, 50(01): 333-343.

[25] Zhong X, Li D.X, Chang J.J, Wang Q.N, Du W.H. Research on Robust Tracking Algorithm of Fighter in Complex Scenes[J]. Electro-Optics and Control:1-6.

[26] Zhang J, Huang H.M, Wang J.M, Bao J.R. Improved TLD real-time target tracking algorithm fused with CN tracking algorithm[J]. Computer Engineering and Science, 2020, 42(07): 1215-1225.

[27] Fan M, Fan Y.S, Ren F.H. Research on moving target tracking algorithm based on improved TLD[J]. Progress in Laser and Optoelectronics, 2020, 57(12): 253-258.

[28] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE C-on-ference on Computer Vision and Pattern Recognition, Seat-tle, Jun 27-30, 2016. Piscataway: 2016, 779-788.

[29] Leal-Taixe L, Milan A, Reid I, etc. Motchallenge 2015: Towards a benchmark for multi-target tracking [EB/OL]. ArXiv:1504.01942[cs], 2015.

[30] Jamal J.K, Munirah K.L, Noori H.W, Zakaria A, Muhammad M.S.Z.S. Multi-Target Detection and Tracking (MTDT) Algorithm Based on Probabilistic Model for Smart Cities[J]. Journal of Physics: Conference Series,2021,1755(1):

[31] Multimedia; New Findings from Wuhan University Update Understanding of Multimedia (Real-time Detection Tracking and Recognition Algorithm Based On Multi-Target Faces) [J]. Journal of Engineering, 2020.

[32] Chang J.L, Sang K.P, Myo T. L. Multi-target Tracking and Track Management Algorithm Based on UFIR Filter With Imperfect Detection Probability[J]. International Journal of Control, Automation and Systems,2019,17(9)