Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 2, No. 3, 2021

Research on Target Tracking Algorithm of Twin Networks Integrating Attention Mechanism

Relizha Yeerlanbieke[1], Huazhang Wang[1]

[1]Southwest Minzu University college of Electrical Engineering, Chengdu, Sichuan, 610041, China

Abstract: Aiming at the current stage of the twin network target tracking algorithm, the tracking target is occluded, the tracking is affected by illumination, and the target's scale change from far to near or from near to far causes tracking failure. This article will optimize and improve from two directions. The twin neural network first uses an adaptive detailed feature extraction, adds a residual network to the twin network, and embeds a detailed feature retention module in each layer, amplifies the changes in the target feature, and retains the important structure of the original target feature Details: Secondly, the introduction of a spatial attention mechanism allows the main branch to pay more attention to the area to be matched, improves the ability to distinguish features, and makes the tracking effect better. In order to verify the effectiveness of this experiment, this experiment was tested on the data set OTB2015. The experiment proved that the proposed algorithm performs better in accuracy and success rate.

Keywords: Deep learning, Twin neural network, Target tracking, Attention mechanism.

1. Preface

Target tracking technology is one of the important research parts of artificial intelligence research. Target tracking refers to determining the coordinates and size of the target in the initial frame, and letting the tracking frame describe the moving trajectory of the tracking target in the subsequent frames. Now its technology is applied to many fields such as safety monitoring, competitive sports, human-computer interaction, etc. [1]. Due to the fast speed of the twin network target tracking algorithm, it has attracted the interest of many researchers. The target tracking algorithm based on this algorithm has also developed rapidly. The algorithm extracts features from the input image and the image to be measured, and then performs calculations and evaluates the results of the calculation. Their similarity. Bertinetto et al. [2] proposed a fully convolutional twin network algorithm. The network has only two layers, a convolutional layer and a pooling layer. It is characterized by an online non-update strategy, which makes the tracking speed faster. Guo et al. [3] proposed a dynamic twin network tracking algorithm, which effectively suppressed the background interference information and the interference caused by deformation in a complex environment, and effectively improved the tracking effect. Li et al. [4] proposed an efficient twin area recommendation network tracking algorithm, which merged the area recommendation network into the twin neural network structure. RPN is composed of classification branches and regression branches. The classification branch is used to identify the target and background; the regression branch is used to adjust the candidate area. Zhu et al. [5] proposed a twin network tracking algorithm based on jammer perception. Compared with SiamRPN, this algorithm can not only get a better tracking frame in the tracking process, but also improve the generalization ability of the tracker [6] The role of. Wang et al. [7] proposed a fast online target tracking and segmentation algorithm [8], mainly replacing AlexNet with ResNet [9] and adding Mask to RPN in parallel to further improve the accuracy of tracking.

It can be concluded that the existing tracking algorithm based on the twin network performs very well in real-time, but in the tracking process, the extraction of the target feature is often based on the preprocessed first frame of picture, and no further updates will be performed afterwards. Therefore, when the target's shape changes from far to near or from near to far, it cannot better adapt to the changes of the target; when the background and foreground in the template image are similar in appearance, the background image may get a larger similarity score. Lead to tracking drift and other problems.

Aiming at the above-mentioned problems based on the twin neural network tracking algorithm, this article made the following improvements on the twin network algorithm: 1) Using an adaptive detailed feature extraction, adding a residual network to the twin network, and The detail feature retention module is embedded in each layer to amplify the changes of target features and retain the important structural details of the original target features. It plays a certain role in the search and discrimination of details in target tracking, which solves the problem of the original tracking network in the parameter scale. The problem of the loss of detailed features of the target when it is lowered. 2) The spatial attention mechanism is used to improve the network model's attention to effective features during feature extraction.

2. Related Work

2.1. Target Tracking Algorithm Based on Twin Neural Network

The SiamFC network is mainly used to calculate the similarity of the features extracted by the two branches in the twin network, and estimate the position of the target through the highest point of the similarity response. The target tracking process can be regarded as a method to solve similar learning problems [10]. If the target image z and the candidate image x are the same target, then the mapping function will return a higher similarity score, on the contrary, it will return a lower similarity score. To find the new position of the target in the new video frame image, we have to perform a complete test on all possible positions and select the candidate target with the highest similarity. The mapping function f(z, x) is shown in the following formula 1, where b_1 represents the offset signal of the position fetching point.

$$f(z,x) = g(\varphi(z),(x)) = \varphi_z * \varphi_x + b_1 \quad (1)$$

The structure diagram of the tracking algorithm based on the twin neural network is shown in Figure 1 below:

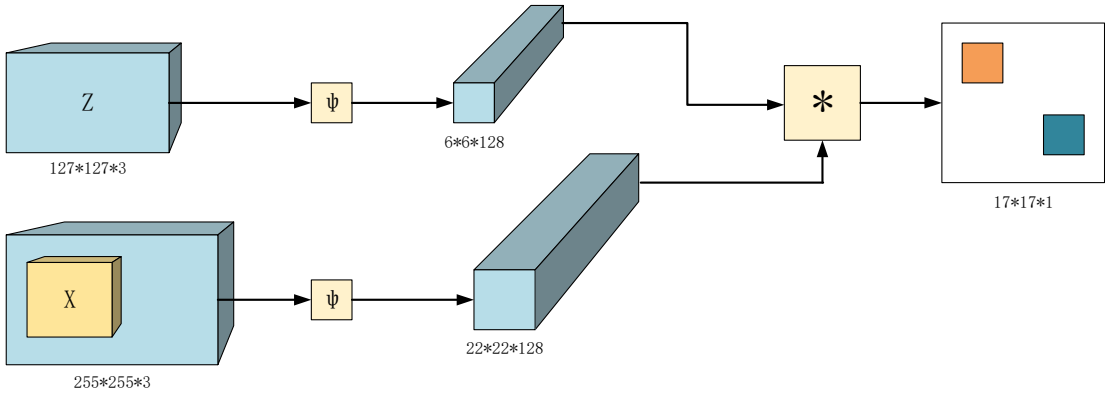

Figure 1. Structural diagram of tracking algorithm based on Siamese network

As shown in Figure 1, z represents the template image (the real target in the first frame), x represents the search image (the search area acquired at the center is the target center of the previous frame), and the template image and the search image are the input network. φ refers to the depth features extracted from the corresponding images of each branch, and * represents the cross-correlation operation.

2.2. Attention Mechanism

In the field of image processing, the attention mechanism draws on the human visual mechanism, while the network focus is limited to a sub-region of the image. The emphasis on the target region improves the feature expression ability of the model. It is similar to the human cognition process, which can selectively focus on part of all information while ignoring other information. In the neural network structure, the attention model achieves attention to specific information by assigning network weights. In recent years, with the continuous research of attention model, many effective attention model network structures have appeared. The attention model is widely used in many aspects such as target detection, target classification, and target tracking. Training can improve the model's ability to model the space, channel, background and other information of the neural network, which helps to improve the representation performance of the convolutional neural network

2.3. Spatial Attention Mechanism

Each picture contains a wealth of data information, and pixels and pixels contain a certain special connection, which can be used to obtain features with more semantic information. This paper introduces the spatial attention mechanism, which is to give different weights to different positions in different spatial positions during feature extraction, so as to achieve the extraction of different feature positions on the feature map, taking the same position between the feature maps as The unit, by comparing the similarity between the feature maps, pays attention to the importance of a certain position in the spatial position. For each feature in the same position in the two feature maps, it can be updated by the weighted sum method to strengthen or suppress The weights of different spatial location features, the greater the weight, the more similar the features. By introducing the spatial attention mechanism, it is possible to increase the spatial weight of important features, effectively extract important features, and effectively improve the feature expression ability of the algorithm model without affecting the computational complexity and speed of the algorithm. Figure 2 shows the structure of this module.

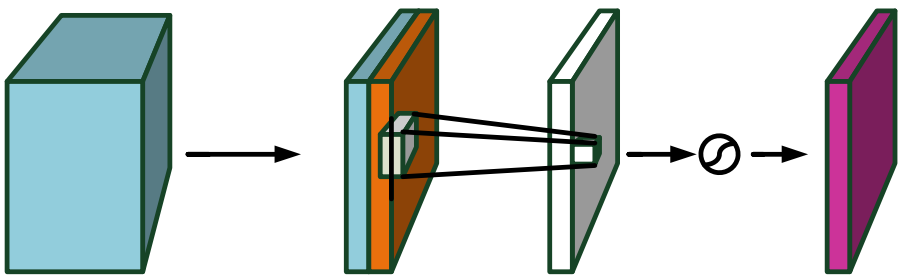

Figure 2. Spatial Attention Module

2.4. Residual Network

The increase in model depth in the network caused a series of gradient-related problems. Scholars found in the development of convolutional neural networks that the performance of neural networks is affected by the depth of the network. As the number of network layers increases, the better the model performance, but also The problem of gradient disappearance and explosion will eventually hinder the convergence of the model due to the failure of network parameter update. In order to solve this problem, He et al. proposed a convolutional neural network called residual net in 2015, which trains CNN to a depth of 152 layers under the action of the internal residual module, and performs multiple detection, segmentation, and First place in the positioning competition. The network introduces the input characteristics to the output, does not increase the complexity of the network model, and reduces the error rate at the same time.

3. The Algorithm of This Paper

The SiamFC algorithm (siamFC) transforms the tracking problem into solving the similarity problem in the target tracking process. Although the algorithm has achieved good results in the tracking process, when the background changes many times during the movement, the algorithm It may be tracked to a target similar to the background in the template, resulting in tracking drift. In order to better solve this problem, this paper adopts an adaptive detailed feature extraction, adding a residual network to the twin network, and adding it to each layer They are all embedded in the detail feature retention module, amplify the changes of target features, retain the important structural details of the original target features, play a certain role in the search and discrimination of details in target tracking, and solve the problem of the original tracking network when the parameter scale is reduced. The problem of the loss of detailed features, and the use of a spatial attention mechanism, is used to improve the network model's attention to effective features during feature extraction.

4. Experiment and Analysis

4.1. Experimental Environment

The algorithm in this paper is implemented based on CUDA 10.2 and pytorch1.2.0 programming language, using five NVIDIA Geforce GTX 1080ti 11G for training, and the model is trained offline on the GOT-10K labeled data set; the initial value of the learning rate is 0.01; the cross-entropy loss function is used; Train for 200 epochs. In order to verify whether the effect of the algorithm in this paper is improved, the OTB2015 data set will be compared, and a representative video sequence will be selected for quantitative analysis.

4.2. Evaluation Index

This article judges the comprehensive performance of each algorithm based on the accuracy and success rate of target tracking.

(2) As shown in formula (3): the calculation of the target tracking success rate refers to the area of the overlap rate (IoU) curve of the target's prediction block diagram and the true boundary block diagram.

$$\mathrm{IoU} = \frac{|R_t \cap R_a|}{|R_t \cup R_a|} \tag{2}$$

The greater the value of IoU, the higher the success rate. When IoU>0.5, it can be considered that the target has been successfully located.

(2) The formula for calculating target tracking accuracy is shown in Figure (3) below, which means calculating the Euclidean metric between the center positions (xt, yt) and (xa, ya) of Rt and Ra.

$$\varepsilon = \sqrt{(x_t - x_a)^2 + (y_t - y_a)^2} \tag{3}$$

4.3. Quantitative Molecules

In order to comprehensively evaluate the algorithm proposed in this paper, tests are performed on the OTB data set to evaluate the distance and overlap accuracy of the algorithm in this paper. In view of the various illumination changes and scale changes in the OTB data set, Table 1 also shows the accuracy and tracking success rate of the algorithm in this paper in dealing with these complex tracking factors.

Table 1. Quantitative comparrisn of different target trackers on OTB2015

	Success rate	Precision
SA-ResDPP-SiamFC	0.512	0.522
ResDPP-SiamFC	0.484	0.496
DPP-SiamFC	0.479	0.488
SiamFC	0.456	0.478

5. Summary and Outlook

In this paper, the SiamFC algorithm is improved and optimized. First, the detailed feature extraction is introduced into the Vgg network, and the residual network is used, so that the detailed feature retention module is embedded in each layer, and the change of the target feature is enlarged and retained. The important structural details of the original target feature play a certain role in the search and discrimination of the details in the target tracking, and solve the problem of the original tracking network losing the target details when the parameter scale is reduced. Finally, the spatial attention mechanism is used. To improve the network model's attention to effective features during feature extraction. Although the accuracy and success rate in target tracking have been improved, there are still some shortcomings in real-time performance, and the real-time problem in target tracking needs to be solved.

References

[1] PENG J Y, CHEN X B. Novel models for one-sided hysteresis of piezoelectric actuator [J]. Mechatronics, 2012, 22(6):757-765.

[2] Yang Kang, Song Huihui, Zhang Kaihua. Real-time visual tracking based on dual attention siamese network [J]. Journal of Computer Applications, 2019, 39(6):1652-1656. (in Chinese)

[3] Kuai Y,WenG,Li D.Hyper-Siamese network forrobust visual tracking[J]. Signal, Image and Video Processing, 2019, 1 3(1):35-42.

[4] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and MachineIntelli -gence, 2015, 37(3): 583-596.

[5] XU T, FENG Z H,WU X,etal. Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking [J]. IEEE Transactions on Image Processing, 2019, 28(11):5596-5609.

[6] Abbas Manuel, Somme Dominique, Le Bouquin Jeannès Régine. D-SORM: A digital solution for remote monitoring based on the attitude of wearable devices[J]. Computer Methods and Programs in Biomedicine,2021,208:

[7] [7] Bai Xingzhen, et al."Target tracking for wireless localization systems using set-membership filtering: A component-based event-triggered mechanism." Automatica 132.(2021): doi:10.1016/J.AUTOMATICA.2021.109795.

[8] Bednarz Bryan P.,Jupitz Sydney,Lee Warren,Mills David,Chan Heather,Fiorillo Timothy,Sabitini James,Shoudy David, Patel Aqsa, Mitra Jhimli,Sarcar Shourya,Wang Bo, Shepard Andrew, Matrosic Charles,Holmes James, Culberson Wesley, Bassetti Michael, Hill Patrick, McMillan

Alan,Zagzebski James,Smith L. Scott,Foo Thomas K.. First-in-human imaging using a MR-compatible e4D ultrasound probe for motion management of radiotherapy[J]. Physica Medica, 2021, 88:

[9] Elgamoudi Abulasad,Benzerrouk Hamza,Elango G. Arul, Landry René. A Survey for Recent Techniques and Algorithms of Geolocation and Target Tracking in Wireless and Satellite Systems[J]. Applied Sciences,2021,11(13):

[10] Sun Peng,Zhu Bing,Zuo Zongyu,Basin Michael V.. Vision-based finite-time uncooperative target tracking for UAV subject to actuator saturation[J]. Automatica,2021,130.