Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9127092, 10 pages
https://doi.org/10.1155/2021/9127092

Research Article

IOU-Guided Siamese Tracking

Jianjun Bao,[1,2] Haibo Wang,[1,2] Chen Lv,[3] Ke Luo,[1,2] and Xiaolin Shen[4]

[1]*Tiandi (Changzhou) Automation Co. Ltd., Changzhou 213015, China*
[2]*CCTEG Changzhou Research Institute, Changzhou 213015, China*
[3]*School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China*
[4]*Jiangsu Urban and Rural Construction College, Changzhou 213147, China*

Correspondence should be addressed to Chen Lv; chenlv@cumt.edu.cn

Received 18 July 2021; Revised 2 September 2021; Accepted 16 September 2021; Published 1 October 2021

Academic Editor: Xiao Chen

Target tracking is currently a hot research topic in machine vision. The traditional target tracking algorithm based on the generative model selects target features manually, which has a simple structure and fast running speed, but it cannot meet the requirements of algorithm accuracy in complex scenes. Compared with traditional algorithms, due to the good performance, the tracking method based on full convolutional network has become one of the important methods of target tracking. However, the RPN-based Siamese network lacks positional reliability when predicting the target area. Aiming at the low tracking accuracy of the RPN-based Siamese network, this paper proposes an improved framework model named IoU-guided SiamRPN (IG-SiamRPN). In the proposed IG-SiamRPN, the IoU-guided branch is first constructed and sample pairs are generated through data augmentation. Then, the Jittered RoI is constructed to train the network to realize the direct prediction of the localization confidence of the candidate area. Subsequently, a target selection method based on predicted IoU scores is proposed, which uses predicted IoU scores instead of classification scores to optimize the target decision strategy of the Siamese network. Finally, an optimization-based fine-tuning method for the Siamese network frame is proposed, which solves the problem of location degradation and improves the performance of the algorithm. Compared with other state-of-the-art target tracking algorithms, experimental results on popular databases demonstrate that the proposed IG-SiamRPN can achieve better performance in both tracking accuracy and robustness.

1. Introduction

In the field of computer vision, target tracking is one of the most challenging research topics. Target tracking aims to estimate the motion trajectory of the target of interest in the subsequent frames by the initial state of the video frame and finally realize the recognition, location, and tracking of the specified target. Currently, there are more and more target tracking algorithms, such as [1–4], which have good performance. And target tracking technology also plays an increasingly important role in work and life; meanwhile it is widely used, including autonomous driving [5–8], reality augmentation [9, 10], drones [11], sports competition [12], surgery [13], biology [14–16], and marine exploration [17]. Due to the excellent performance, the tracking method based on the fully convolutional Siamese networks has become one of the important methods of target tracking. However, the RPN-based Siamese network lacks localization confidence when predicting the target area, and there is a problem of location degradation after multiple bounding box regressions on the target.

For the above problems, based on the SiamRPN [18], this paper proposes a IoU-guided target tracking algorithm.

The Siamese network has attracted more and more attention from researchers because of its outstanding performance in target tracking. The Siamese network is a neural network with a special structure, including two symmetrical branches up and down [19]. It extracts features through a convolutional network with shared parameters, avoids increasing network parameters caused by multiple inputs, and solves the problem of inconsistent data distribution caused by spatial dimension difference

between target template and detection frame. The similarity measurement is performed on the features extracted from the upper and lower branches. The higher the score, the more similar the detection frame and the target template, and the target can be determined by the score.

The SiamRPN tracker introduces the Region Proposal Network (RPN) based on the Siamese network, which not only solves the problem of the scale change of the target but also can regress and fine-tune the predicted bounding box so that the accuracy of the algorithm is further improved.

Furthermore, the input of the SiamRPN consists of two parts: template frame and detection frame. The template frame is the tracking target, and the detection frame is the subsequent video frame that needs to be detected. Through the Siamese subnetwork with parameter sharing, convolution features of different sizes are obtained. After that, the convolution features are sent to the RPN module, and the RPN subnetwork first generates a region of interest (RoI) on the mAP and then performs operations on the region of interest of the convolution feature. The function of the classification branch of the RPN module is to distinguish between the foreground and the background of the detection frame's RoI, and the regression branch is used to fine-tune the bounding box to obtain the precise position of the target.

When SiamRPN algorithm tracks the target, the network divides the tracking task into two independent subtasks processed in parallel. On one hand, the classification branch outputs the classification score of the region of interest; on the other hand, the function of the regression branch is to fine-tune the position and size of the region of interest. However, the regression branch obtains the optimal transformation relationship between the region of interest and the real target and does not measure the positioning accuracy of the region of interest. That is, SiamRPN lacks localization confidence, which will bring two problems in the algorithm:

(1) The SiamRPN algorithm directly selects the region of interest with the highest classification score as the prediction target to output, while the classification score can only determine its category, and cannot be the basis for judging the positioning accuracy and positioning quality of RoI.

(2) Because of the lack of localization confidence, as the number of frame regression iterations increases, the intersection ratio between the region of interest and the real target becomes smaller and smaller, and thus the problem of positioning degradation will appear.

Aiming at the problem of the lack of positional reliability of the Siamese network, the IoU-guided Siamese network is proposed. The solutions to these two problems are as follows:

(1) The IoU-guided branch is introduced, then sample pairs are generated through data augmentation, and Jittered RoIs are constructed. The network is learned through samples to directly predict the localization confidence of the candidate area.

(2) A target selection method based on predicted IoU scores is proposed. And the predicted IoU scores were used instead of classification scores to optimize the target decision strategy of the Siamese network. Finally, an optimized Siamese network frame fine-tuning method is proposed, the precision RoI pooling layer is introduced, continuous gradient of the frame of the RoI is obtained through integration, and hence the quantization errors caused by discretization is avoided. On the basis of it, the predicted IoU score is used as the evaluation criterion and basis for frame regression, which solves the degradation problem of localization and improves the performance of the algorithm.

2. Proposed Method

In order to solve the above problems, this section first introduces the IoU-guided module, increases the measurement index of localization confidence, and uses the localization confidence predicted by IoUNet to replace the classification confidence to output the prediction results. Also, the precision pooling layer was introduced. The localization confidence was used as the measurement index of the border regression, and the bounding box fine-tuning method based on optimization was used for calculation. The proposed network is called IoU-Guided SiamRPN (IG-SiamRPN).

The network structure of IG-SiamRPN is shown in Figure 1. The blue area is the main framework for IoU-guided network. The Siamese network is still used for feature extraction, the backbone network is AlexNet, and the RPN is used to generate region of interest, simultaneously; the IoU subnetwork performs localization confidence prediction on the generated region of interest. The CNN module in the IoU subnetwork includes three 3×3 convolution kernels, which perform the convolution operation on the features after the cross-correlation operation again to deepen the network depth of the IoU subnetwork, so that the network has better fitting and nonlinear expression ability.

IG-SiamRPN uses the predicted localization confidence obtained by the IoU subnetwork instead of the classification confidence used by SiamRPN. It makes the region of interest with high positioning accuracy preserved, which solves the problem that SiamRPN uses classification score to measure the accuracy of positioning of the bounding box.

When training the IoU network, sample pairs and labels are randomly generated through data augmentation. It makes the IoU network more robust. The training of predicting IoU is independent of the SiamRPN tracking algorithm, so whatever happens to the distribution of the inputs to the tracking algorithm, it will not cause too much interference in the prediction of the localization confidence. Candidate region pairs are generated by randomly manually adjusting the bounding boxes of the template box and the detection box in the training set, and then the samples with IoU <5 are regarded as invalid sample pairs and eliminated, and finally Jittered RoIs are randomly sampled from the

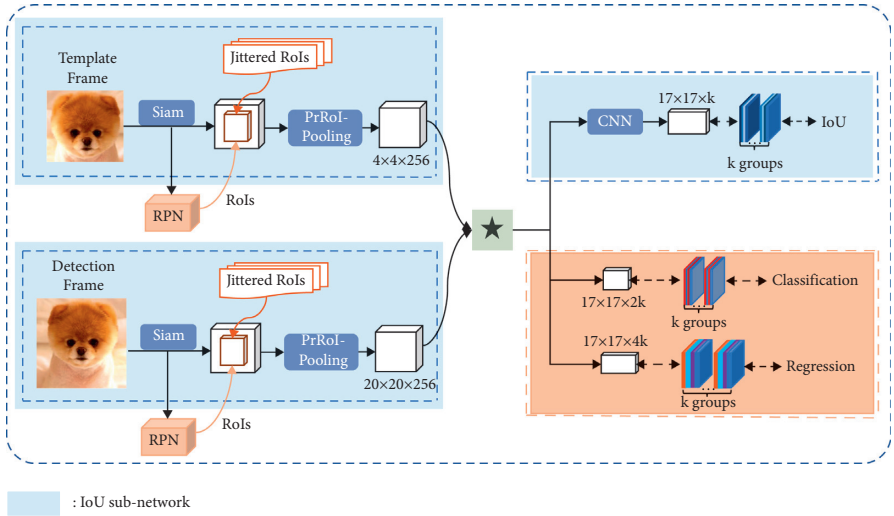

Figure 1: Full architecture of IG-SiamRPN.

remaining candidate region pairs to train the IoU-guided network.

In the phase of prediction, the feature extraction is carried out through the Siamese subnetwork and the interesting area is generated by RPN. Then, the PrRoI Pooling layer is used to extract features again. The convolution features with the size of $4 \times 4 \times 256$ and $20 \times 20 \times 256$ are generated in the template frame and the detection frame, respectively, and the cross-correlation operation is performed. After CNN operation, the localization confidence corresponding to the output region of interest of IoU branch is predicted, which is the predicted value of the cross-comparison with the real target.

This research introduces the precise RoI pooling (PrRoI pooling) layer in the network structure. The continuous gradient of the RoI frame is obtained through integration, which avoids the quantization error caused by discretization, and we propose an optimization-based bounding box fine-tuning method based on the PrRoI pooling layer and the IoU-guided network. Fine-tuning the predicted target to obtain precise positioning is a process of finding the optimal solution, which can be expressed by

$$c^* = \arg\min_c \operatorname{crit}\left(\operatorname{transform}\left(\operatorname{box}_{\text{det}}, c\right), \operatorname{box}_{\text{gt}}\right). \tag{1}$$

In formula (1), $\operatorname{box}_{\text{det}}$ represents the bounding box of the predicted target, $\operatorname{box}_{\text{gt}}$ represents the bounding box of the real target, $\operatorname{transform}(\cdot)$ represents the conversion function, which is used to convert the predicted target frame into a result closer to the real target frame, and $\operatorname{crit}(\cdot)$ is the evaluation criterion of the conversion result. In this paper, the IoU-guided score is used as the criterion for bounding box regression and the final optimization goal, and the optimal solution c^* is obtained through gradient descent, which solves the problem of lacking theoretical basis for bounding box regression and the inability to perform multiple iterations.

Based on the optimized IG-SiamRPN bounding box fine-tuning method, after predicting the localization confidence of the region of interest, the frame of the region of interest is iterated through the localization confidence gradient to maximize the localization confidence of the frame. Because of the introduction of localization confidence, the bounding box regression has a scientific evaluation index, and the process of the border regression is more reasonable.

The flow of the optimized IG-SiamRPN bounding box regression algorithm is shown in Algorithm 1. In Step 6 of Algorithm 1, after PrRoI pooling of the region of interest, the IoU of the region and the real target is predicted, namely, the localization confidence, and the partial derivative of the region border coordinates is calculated.

This study selected a total of 181,402 video sequences from the ILSVRC2015-VID dataset [20] and the YouTube-Bounding Boxes dataset (YT-BB) [21] to train the IG-SiamRPN tracker. The number of videos in the training set is 179,587, and the number of videos in the verification set is 11,815.

IG-SiamRPN uses the pretrained AlexNet weights of ImageNet to initialize the network, and for the newly added layers in the network; it uses a Gaussian distribution with a mean value of 0 and a variance of 0.01 to initialize the weights randomly. When training the IoU predictor, the Smooth Loss is used as the loss function for predicting the IoU branch. The expression of the loss function is shown in

$$\operatorname{smooth}_{L1}(x, \sigma) = \begin{cases} 0.5|x|\sigma^2, & \text{if} |x| < \dfrac{1}{\sigma^2} \\ |x| - \dfrac{1}{2\sigma^2}, & \text{if} |x| \geq \dfrac{1}{\sigma^2} \end{cases}. \tag{2}$$

Inputs x and σ of the loss function are the predicted IOU score and the real IOU score, respectively.

In the process of prediction, firstly, the region of interest is generated by the RPN subnetwork, and their categories are determined. The region of interest with a classification score greater than 0.7 is regarded as prospects and retained. The

Input: $\mathscr{B} = \{b_1, \ldots b_n\}$, $\mathscr{F}$, λ, $\mathscr{T}$, Ω_1, Ω_2
$\mathscr{B}$ represents a set of detected bounding boxes in the form of (x_1, y_1, x_2, y_2), where (x_1, y_1) and (x_2, y_2), respectively, represent the continuous coordinates of the upper left corner and the lower right corner of the area.
$\mathscr{F}$ is the feature map of the grid area before PrRoI pooling.
$\mathscr{T}$ is the number of iteration steps to fine-tune the bounding box, λ is the step size, Ω_1 is the threshold for stopping the iteration early, and Ω_2 is the tolerance of positioning degradation.
Output: predicted target bounding box after fine-tuning

(1) **begin**
(2) $\mathscr{A} \longleftarrow \varnothing$
(3) **for** $i = 1$ to $\mathscr{T}$ **do**
(4) **for** $b_j \in \mathscr{B}$ and $b_j \notin \mathscr{A}$ **do**
(5) PrevScore $\longleftarrow$ IoU(PrPool$(b_j, \mathscr{F})$)
(6) grad $\longleftarrow \nabla_{bj}$IoU(PrPool$(b_j, \mathscr{F})$)
(7) $b_j \longleftarrow b_j + \lambda \times$ scale(grad, b_j)
(8) NewScore $\longleftarrow$ IoU(PrPool$(b_j, \mathscr{F})$)
(9) **if** |PrevScore − NewScore| $< \Omega_1$ **or** PrevScore − NewScore $< \Omega_2$ **then**
(10) $\mathscr{A} \longleftarrow \mathscr{A} \cup \{b_j\}$
(11) **end if**
(12) **end for**
(13) **end for**
(14) **return** $\mathscr{B}$

Algorithm 1: The optimized IG-SiamRPN bounding box regression algorithm.

regression branches are used for bounding box regression, and the localization confidence of this region of interest is generated by the IoU-guided branch. After the regression, the optimized bounding box fine-tuning method is used to adjust the bounding box again. The parameters used are as follows: iteration steps, step size, the threshold for stopping the iteration early, and the tolerance of positioning degradation. Finally, the region of interest with the highest localization confidence is output, which is the final result predicted by the IG-SiamRPN tracker.

3. Experiment and Analysis

This section will make a comprehensive longitudinal comparison of IG-SiamRPN and the current mainstream tracking algorithms in the OTB100 and VOT dataset. They are SimaDWfc [22], GradNet [23], DaSiamRPN [24], SRDCF [25], DeepSRDCF [26], SiamRPN, CFNet [27], SiamFC [28], and SimaDW-RPN [29], respectively. Moreover, the experimental results will be analyzed quantitatively, and the performance of the algorithm will be analyzed in the face of various disturbances.

3.1. Performance Evaluation of Two Improved Methods. The experiments in this section mainly include three aspects. After introducing the IoU-guided module, this article first conducts ablation experiments to verify the optimization-based bounding box fine-tuning method and the target selection method based on the predicted IoU score, and test the actual performance of these two methods. Finally, the overall performance of IG-SiamRPN is verified. For the final IG-SiamRPN, after the bounding box regression, the optimized bounding box fine-tuning method is used to fine-tune the prediction again, and the target selection method based on the predicted IoU score is used to output the final predicted target.

The algorithm in this paper and SiamRPN algorithm are tested and compared on OTB100, VOT2017, and VOT2018, respectively. Tables 1–3, respectively, show the evaluation results of the above experiments. By the way, the bounding box fine-tuning method of SiamRPN algorithm uses regression-based strategy, and the target selection method uses classification score.

Table 1 shows the test results of the IG-SiamRPN and SiamRPN algorithms on the OTB100 dataset. It can be observed that IG-SiamRPN performs better than the SiamRPN algorithm whether it uses the optimization-based bounding box fine-tuning method or the target selection method based on the predicted IoU. Compared with the SiamRPN algorithm, IG-SiamRPN, which only uses the optimization-based bounding box fine-tuning algorithm and does not improve the target selection strategy, improves the precision and success by 2.21% and 2.74%, respectively. At the same time, IG-SiamRPN improves the precision and success by 2.82% and 3.02%, respectively, which only uses the target selection method based on the predicted IoU score and does not improve the bounding box fine-tuning method. And compared with SiamRPN, IG-SiamRPN has improved precision by 5.9%, and its success has increased by 5.71%, which uses optimization-based methods to fine-tune the bounding box and chooses a target selection strategy based on the predicted IoU score after performing traditional bounding box regression. When only the bounding box fine-tuning method is improved, the algorithm based on the optimized bounding box regression has improved accuracy, robustness, and EAO by 0.058%, 1.29%, and 0.095% compared with the SiamRPN algorithm.

Table 1: Test results on OTB100 of the proposed methods.

Algorithm	Bounding box fine-tuning method		Target selection strategy		Test dataset	
	Based on regression	Based on optimization	Based on classification score	Based on predicted IoU score	OTB100	
					Precision	Success
SiamRPN	√		√		0.847	0.629
IG-SiamRPN		√	√		0.866	0.646
	√			√	0.871	0.648
	√	√		√	**0.897**	**0.665**

Table 2: Test results on VOT2017 of the proposed methods.

Algorithm	Bounding box fine-tuning method		Target selection strategy		Test dataset		
	Based on regression	Based on optimization	Based on classification score	Based on predicted IoU score	VOT2017		
					Accuracy	Robustness	EAO
SiamRPN	√		√		0.519	0.462	0.316
IG-SiamRPN		√	√		0.522	0.456	0.319
	√			√	0.527	0.451	0.321
	√	√		√	**0.529**	**0.435**	**0.326**

Table 3: Test results on VOT2018 of the proposed methods.

Algorithm	Bounding box fine-tuning method		Target selection strategy		Test dataset		
	Based on regression	Based on optimization	Based on classification score	Based on predicted IoU score	VOT2018		
					Accuracy	Robustness	EAO
SiamRPN	√		√		0.585	0.376	0.323
IG-SiamRPN		√	√		0.593	0.371	0.328
	√			√	0.590	0.368	0.329
	√	√		√	**0.594**	**0.361**	**0.334**

Table 2 shows the test results of the proposed method on the VOT2017 dataset. When only the bounding box fine-tuning method is improved, the algorithm based on the optimized bounding box regression has improved accuracy, robustness, and EAO by 0.058%, 1.29%, and 0.095% compared with the SiamRPN algorithm. And when only the target selection strategy is optimized, the accuracy of the target selection method based on the predicted IoU score is improved by 1.54%, the robustness is improved by 2.37%, and the EAO is improved by 1.58%. And the final IG-SiamRPN algorithm has improved accuracy, robustness, and EAO by 1.90%, 5.84%, and 3.16%, respectively.

Table 3 shows the test result on the VOT2018 dataset. It can be seen that the optimized bounding box regression algorithm has improved accuracy, robustness, and EAO by 1.44%, 1.45%, and 1.69%, respectively. And the accuracy of the target selection method based on the predicted IoU score increased by 0.94%, the robustness increased by 2.11%, and EAO increased by 1.73%. The final IG-SiamRPN algorithm has increased by 1.53%, 3.99%, and 3.41% in these three areas, respectively.

In terms of algorithm running speed, the average running speed of SiamRPN on this experimental platform is 108FPS, and the running speed of IG-SiamRPN is 87FPS. While the speed is slightly reduced, it does not affect the real-time tracking of the algorithm.

3.2. Quantitative Analysis. Generally, the test results of the tracker in the OTB100 dataset can be drawn into the corresponding precision plot and success plot for visual display, as shown in Figures 2 and 3.

From Figures 2 and 3, it can be seen that the IG-SiamRPN algorithm proposed in this study is superior to the SiamRPN algorithm and the improved DaSiamRPN algorithm based on SiamRPN. IG-SiamRPN has achieved the best results in terms of both precision and success rate. It can be observed from the trend of the curve that the performance of IG-SiamRPN is greatly improved mainly in the interval with higher threshold setting requirements, which is related to the training method of the predictive IoU module. After data enhancement, only positive samples with IoU greater than 0.5 are used for training. Therefore, when the IoU of the predicted frame and the real target is low, the predicted IoU score obtained is not accurate, resulting in an insignificant improvement in the performance of the algorithm. In terms of tracking speed, the IG-SiamRPN algorithm can reach 87FPS, while the comparison algorithm only has a real-time tracking speed of about 30FPS. Therefore, the overall performance of the IG-SiamRPN tracking algorithm is better.

In addition, this section also compares IG-SiamRPN with the main algorithms on the VOT2017 and VOT2018 datasets. The comparison results are shown in Table 4.

In the VOT2017 dataset, compared with other algorithms, IG-SiamRPN has achieved the best results in

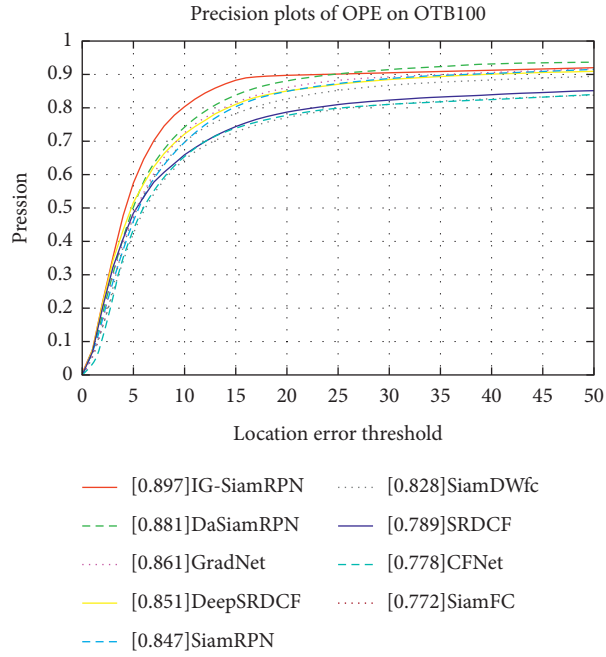

Figure 2: Precision plots of OPE on OTB100.

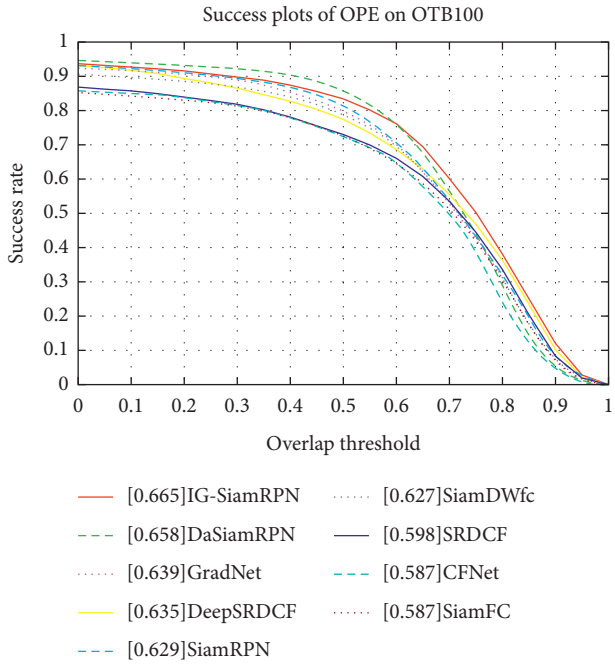

Figure 3: Success plots of OPE on OTB100.

robustness and EAO. In the VOT2018 dataset, the robustness of IG-SiamRPN is not as good as the DaSiamRPN algorithm and the SiamDW-RPN, but the main performance indicators EAO and accuracy are higher than the comparison algorithm.

3.3. Comparative Analysis under Various Interference Situations. In order to study the tracking performance of the improved algorithm in the face of specific interference, this paper selects six kinds of tracking difficulties encountered in actual scenes in the OTB100 dataset for testing. They are scale variation, deformation, fast motion, illumination variation, occasion, and background clutters.

The success rate index of the OTB100 dataset reflects the performance of the algorithm more than accuracy, so the success rate graph of each algorithm is selected for display.

Figure 4 shows a comparison chart of the success rate under the interference. The six pictures correspond to the above six types of interference. In summary, the IG-SiamRPN proposed in this study has better robustness under the interference and is the best among all the Siamese network algorithms under the conditions of scale change, deformation, fast target movement, occlusion, and background clutter. Except when the illumination changes and occlusion appears, the performance of IG-SiamRPN is not optimal. In other cases, IG-SiamRPN achieved the best results.

3.4. Qualitative Analysis. In order to demonstrate the tracking effect of the IG-SiamRPN algorithm clearly, this study selects 5 video sequences containing different tracking interference items in the VOT2018 dataset to compare and verify with 4 representative tracking algorithms based on the Siamese network. Figure 5 shows part of the visualization results of different algorithms. The figure corresponds to the five test video sequences of singer2, rabbit, tiger, girl, and graduate in the VOT dataset from top to bottom. As shown in Figure 5, the IG-SiamRPN algorithm can achieve better results in different tracking scenarios and has good performance under various challenges.

The interference of Singer2 video sequence is mainly the change of illumination. It can be observed that at the beginning of the tracking stage all the five algorithms are

Table 4: Test results on VOT datasets.

Algorithm	VOT2017			VOT2018		
	Accuracy	Robustness	EAO	Accuracy	Robustness	EAO
SRDCF	0.491	0.973	0.120	—	—	—
SiamFC	0.501	0.586	0.188	0.504	0.583	0.186
CFNet	0.488	0.595	0.195	0.466	0.531	0.194
GradNet	0.507	0.375	0.247	—	—	—
DeepSRDCF	0.492	0.363	0.256	—	—	—
SiamRPN	0.519	0.462	0.316	0.585	0.376	0.323
SiamDWfc	0.518	0.496	0.324	0.558	0.374	0.239
DaSiamRPN	0.548	0.441	0.326	0.569	0.337	0.323
SiamDW-RPN	0.521	0.458	0.312	0.587	0.263	0.332
IG-SiamRPN	**0.529**	**0.435**	**0.326**	**0.594**	**0.361**	**0.334**

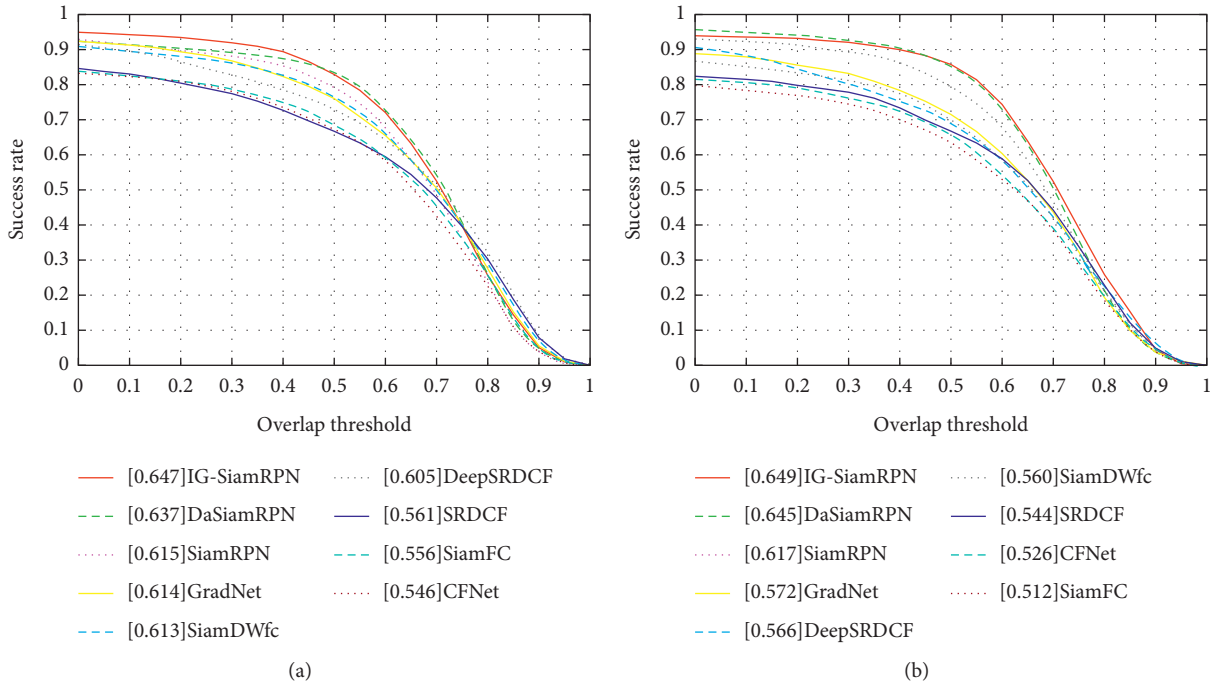

Figure 4: Continued.

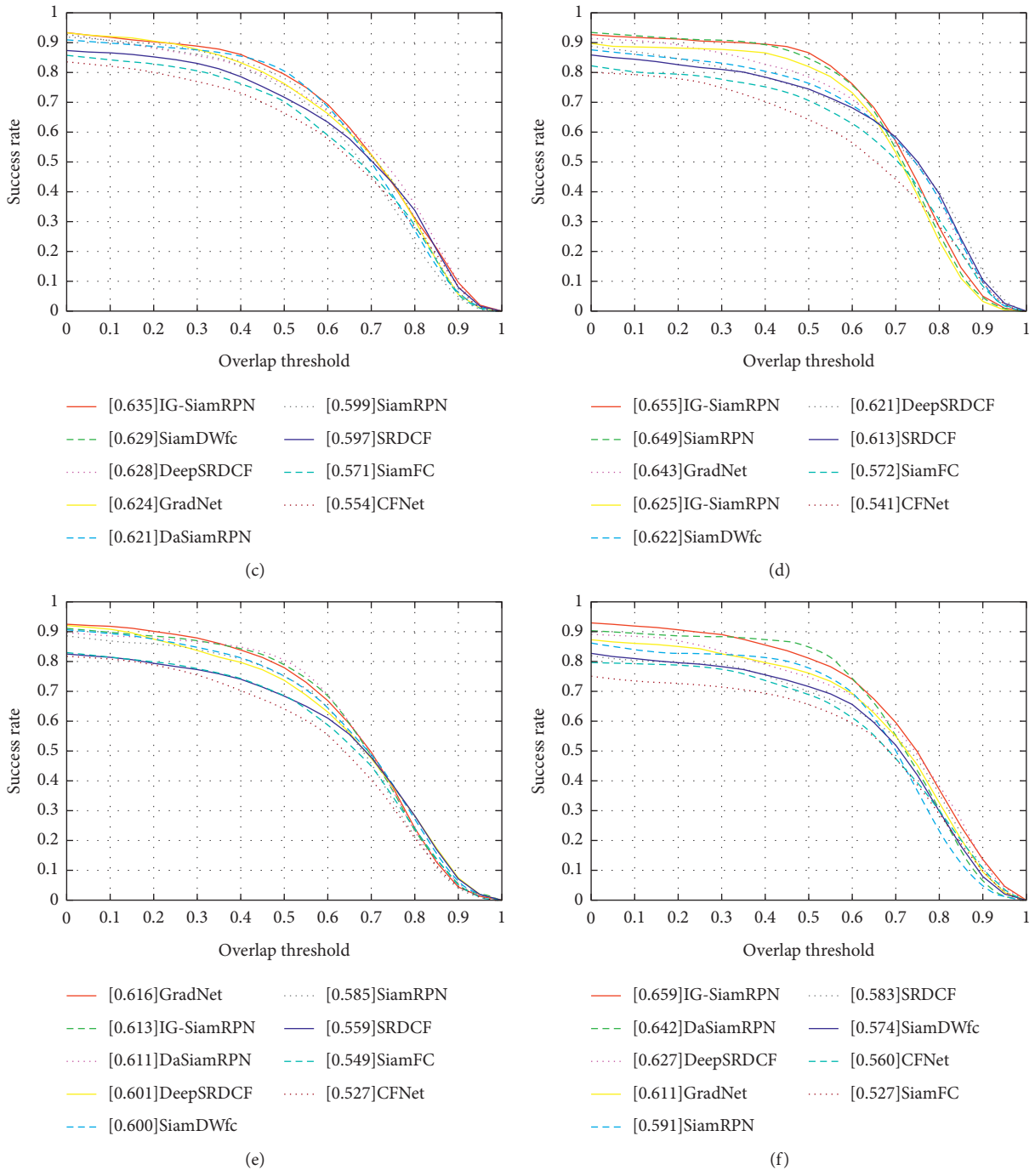

Figure 4: Comparison chart of the success rate of each algorithm when interference occurs. (a) Success plots of OPE—scale variation, (b) success plots of OPE—deformation, (c) success plots of OPE—fast motion, (d) success plots of OPE—illumination variation, (e) success plots of OPE—occlusion, and (f) success plots of OPE—background clutters.

successful in tracking without changing the illumination. When the illumination changes, IG-SiamRPN, DaSiamRPN, and SiamRPN can locate the target, and IG-SiamRPN can accurately identify the target; in other words, the tracking effect is the best. SiamDWfc and SiamFC have a large deviation, while SiamFC is basically unable to locate the target. Subsequently, the interference caused by illumination changes in frames 41 and 180 is further strengthened, except that IG-SiamRPN and DaSiamRPN can track the target; the other algorithms all lose the target.

The tracking challenge of the rabbit video sequence with a lot of noise is that it is not degenerated, and all five trackers perform poorly. In the initial stage of tracking, there is a certain offset, and finally, all the targets are lost. This shows that the

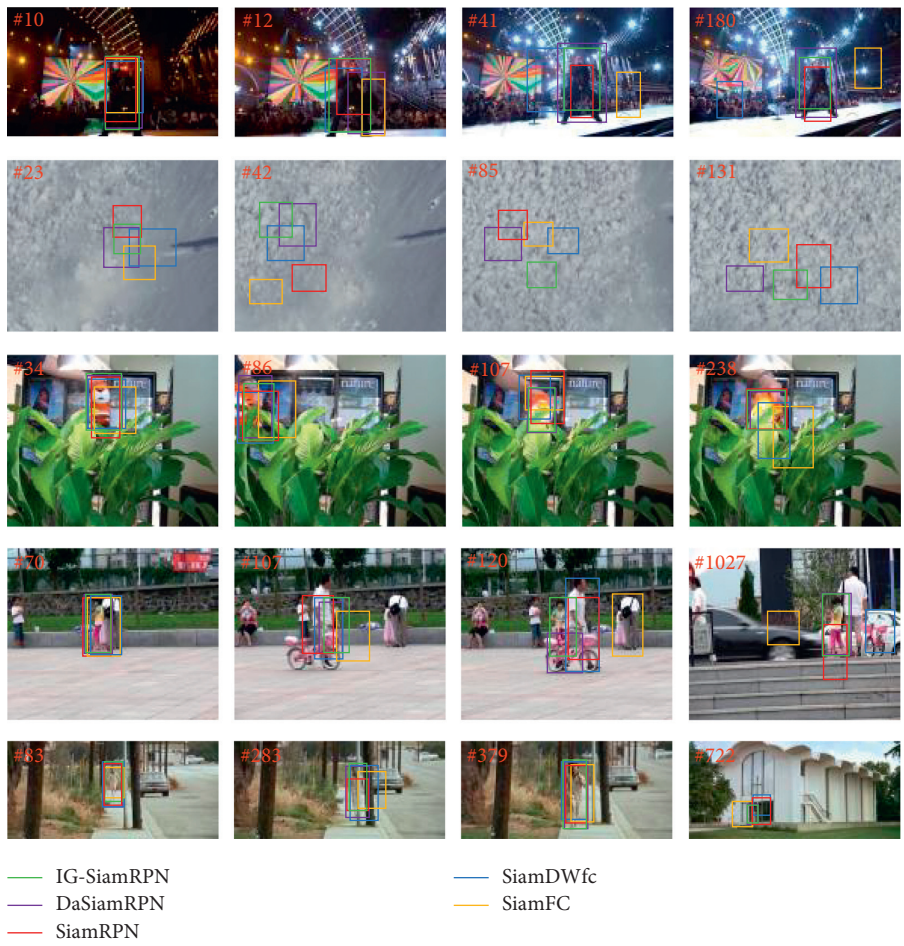

Figure 5: Comparison of partial video visualization results of different algorithms.

complex background interference is still an important factor affecting the accuracy of the algorithm.

The interference in the tiger and girl video sequence is mainly target occlusion. When the target is occluded, the other algorithms show a large offset or even loses the target. Although IG-SiamRPN also has a certain offset, it can still track the target.

The interference of the graduate video sequence is occlusion and the change of target scale. When the target is blocked, IG-SiamRPN and DaSiamRPN have the best tracking effect. When the target changes on a large scale, SiamFC loses the target due to the lack of corresponding coping strategies. Although SiamRPN and SiamDWfc can track the target, a large deviation appears. The IG-SiamRPN and DaSiamRPN have the best tracking effect.

4. Conclusions

In target tracking, the Siamese network based on the RPN module directly selects the region of interest with the highest classification score as the predicted target output, but there is no positive correlation between the classification score and the positioning accuracy, which leads to a decrease in the tracking accuracy of the algorithm. This research proposes a Siamese network algorithm based on IoU-guided to solve the lack of positional reliability. Through training, the IoU branch is directly used to predict the localization confidence. The bounding box regression strategy was improved on this basis. IOU score was used as the basis of bounding box regression, and a bounding box regression method based on optimization was proposed to solve the problem of location degradation. Finally, the IOU score was used as the evaluation standard of the final target output. This research mainly focuses on the SiamRPN\, and the proposed algorithm has achieved good results on the OTB dataset and the VOT dataset.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Science and Technology Innovation Special Fund Key Project of China Coal

Technology and Engineering Group under Grant No. 2018-TD-ZD005.

References

[1] X. Li, Q. Liu, N. Fan, Z. Zhou, Z. He, and X.-y. Jing, "Dual-regression model for visual tracking," *Neural Networks*, vol. 132, pp. 364–374, 2020.

[2] Z. Zhou, N. Fan, K. Yang, W. Hongpeng, and H. Zhenyu, "Adaptive ensemble perception tracking," *Neural Networks*, vol. 142, pp. 316–328, 2021.

[3] D. Yuan, X. Chang, P.-Y. Huang, Q. Liu, and Z. He, "Self-supervised deep correlation tracking," *IEEE Transactions on Image Processing*, vol. 30, pp. 976–985, 2021.

[4] Q. Liu, X. Li, and Z. Y. He, "Learning deep multi-level similarity for thermal infrared object tracking," *IEEE Transactions on Multimedia*, vol. 23, pp. 2114–2126, 2021.

[5] M. F. Chang, J. Lambert, P. Sangkloy et al., "Argoverse: 3D tracking and forecasting with rich maps," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8748–8757, Long Beach, CA, USA, June 2019.

[6] W. Luo, B. Yang, and R. Urtasun, "Fast and furious: real time end-to-end 3D detection, tracking and motion forecasting with a single convolutional net," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3569–3577, Salt Lake City, UT, USA, June 2018.

[7] P. Giro, A. Asvadi, P. Peixoto, and U. Nunes, "3D object tracking in driving environment: a short review and a benchmark dataset," in *Proceedings of the PPNIV16 Workshop, IEEE 19th International Conference on Intelligent Transportation Systems (ITSC 2016)*, pp. 7–12, Janeiro, Brazil, November 2016.

[8] C. Li, X. Liang, Y. Lu, N. Zhao, and J. Tang, "RGB-T object tracking: benchmark and baseline," *Pattern Recognition*, vol. 96, pp. 8–15, 2019.

[9] M. Klopschitz, G. Schall, D. Schmalstieg, and G. Reitmayr, "Visual tracking for augmented reality," in *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 19–28, Zurich, Switzerland, September 2010.

[10] F. Ababsa, M. Maidi, J.-Y. Didier, and M. Mallem, "Vision-based tracking for mobile augmented reality," *Studies in Computational Intelligence*, Springer Berlin Heidelberg, Berlin, Germany, pp. 297–326, 2008.

[11] J. Hao, Y. Zhou, G. Zhang, Q. Lv, and Q. Wu, "A review of target tracking algorithm based on UAV," in *Proceedings of the 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, pp. 328–333, Shenzhen, China, October 2018.

[12] M. Manafifard, H. Ebadi, and A. Moghaddam, "A survey on player tracking in soccer videos," *Computer Vision and Image Understanding*, vol. 159, pp. 19–46, 2017.

[13] D. Bouget, M. Allan, D. Stoyanov, and P. Jannin, "Vision-based and marker-less surgical tool detection and tracking: a review of the literature," *Medical Image Analysis*, vol. 35, pp. 633–654, 2017.

[14] V. Ulman, M. Maška, K. E. G. Magnusson et al., "An objective comparison of cell-tracking algorithms," *Nature Methods*, vol. 14, no. 12, pp. 1141–1152, 2017.

[15] T. He, H. Mao, J. Guo, and Z. Yi, "Cell tracking using deep neural networks with multi-task learning," *Image and Vision Computing*, vol. 60, pp. 142–153, 2016.

[16] D. E. Hernandez, S. W. Chen, E. E. Hunter, E. B. Steager, and V. Kumar, "Cell tracking with deep learning and the viterbi algorithm," in *Proceedings of the 2018 International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS)*, pp. 1–6, Nagoya, Japan, July 2018.

[17] J. Luo, Y. Han, and L. Fan, "Underwater acoustic target tracking: a review," *Sensors*, vol. 18, no. 2, p. 112, 2018.

[18] B. Li, J. J. Yan, W. Wu, Z. Zhu, and X. Hu, "High performance visual tracking with siamese region proposal network," in *Proceedings of the 2018 Ieee/Cvf Conference on Computer Vision and Pattern Recognition*, pp. 8971–8980, Salt Lake City, UT, USA, June 2018.

[19] J. Zhang and X. Tan, "Visual object tracking based on siamese network with a deeper and more powerful backbone," in *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pp. 785–788, Dalian, China, June 2020.

[20] O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[21] E. Real, J. Shlens, S. Mazzocchi, X. Pan, and V. Vanhoucke, "You Tube-Bounding Boxes: a large high-precision human-annotated data set for object detection in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7464–7473, Honolulu, HI, USA, July 2017.

[22] Z. P. Zhang and H. W. Peng, "Deeper and wider siamese networks for real-time visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4586–4595, Long Beach, CA, USA, June 2019.

[23] P. Li, B. Chen, W. Ouyang, D. Wang, X. Yang, and H. Lu, "GradNet: gradient-guided network for visual object tracking," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6161–6170, Seoul, South Korea, October 2020.

[24] Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, and W. Hu, "Distractor-aware siamese networks for visual object tracking," in *Proceedings of the European Conference on Computer Vision*, pp. 103–119, Munich, Germany, September 2018.

[25] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4310–4318, Santiago, Chile, December 2015.

[26] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4310–4318, Santiago, Chile, December 2015.

[27] J. Valmadre, L. Bertinetto, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "End-to-End representation learning for correlation filter based tracking," in *Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5000–5008, Honolulu, Hawaii, July 2016.

[28] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," *Lecture Notes in Computer Science*, Springer, Basel, Switzerland, pp. 850–865, 2016.

[29] Z. Zhang and H. Peng, "Deeper and wider siamese networks for real-time visual tracking," in *Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4591–4600, Long Beach, CA, USA, June 2019.