A research of target tracking algorithm based on deep learning and kernel correlation filter

Shengtao Sun
Department of Computer Science and Engineering
Yanshan University
Qinhuangdao 066004, China
xysst@ysu.edu.cn

Jibing Gong
Department of Computer Science and Engineering
Yanshan University
Qinhuangdao 066004, China
gongjibing@163.com

Yangyang Li
National Engineering Research Center for Risk Perception and Prevention
CAEIT
Beijing 100041, China
liyangyang@cetc.com.cn

Lizhe Wang
Department of Computer Science and Engineering
Yanshan University
Qinhuangdao 066004, China
lizhe.wang@gmail.com

Kaisheng Wang*
Department of Computer Science and Engineering
Yanshan University
Qinhuangdao 066004, China
ysusst@163.com

2022 International Joint Conference on Neural Networks (IJCNN) | 978-1-7281-8671-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/IJCNN55064.2022.9892579

***Abstract*—As a hot topic in computer vision, target tracking has a vital application in many scientific and technological fields. The tracking method based on correlation filtering transforms the target tracking from the time domain to the frequency domain through the Fourier transform, which can boost the accuracy and success rate. However, in the complex tracking environments, the target tracking process may be affected by deformation, occlusion, and other inferences, which make the traditional target tracking algorithms hardly accommodate the requirements of robustness. Aiming the target tracking in complex scenes, this paper tries to improve the feature extraction based on the Convolution Neural Network, which can learn deep features of the target from different convolution layers with more abstract characteristics. Then these multiple features are fused to enhance the robustness performance of the traditional Kernel Correlation Filter algorithm from the aspects of model characteristics. Furthermore, the accuracy and success rate of the proposed algorithm are verified based on comprehensive comparative experiments in the Object Tracking Benchmark with variant interferences.**

Keywords—target tracking, kernel correlation filter, convolution neural network, deep feature extraction, multiple feature fusion

I. INTRODUCTION

With the continuous progress of computer technology, computer vision has become a research field that attracts the attention of research scholars and technology companies [1]. As one of the hotspots in computer vision, target tracking is realized based on video target detection, which extracts the characteristic of the object to track the target region. During the tracking of a target in the video, the position, speed, and motion track of the target can obtain, while the target may deform or rotate in the movement [2]. To achieve target tracking accurately, target tracking algorithms with better robustness and real-time performance are necessary [3]. Target tracking algorithms have a wide range of application prospects and practical values in the fields of intelligent monitoring, automatic driving, human-computer interaction, and so on. However, in complex backgrounds, such as target scale variation, deformation, rotation, and motion blur, target tracking is difficult to meet the requirements of robustness.

Aiming the solution of the current challenge that most traditional target tracking algorithms use the single and manual setting feature for characteristic recognition, this paper proposes an improved Kernel Correlation Filter (KCF) algorithm based on deep feature fusion. This algorithm calculates the output response value of different deep features according to the advantages of each feature in different environments, considering that the tracking cannot perform accurately under the conditions of moving target size change, deformation, out of view and motion blur. To improve the robustness of the algorithm, the abstracted features are weighted by Bhattacharyya coefficients and then adopted to complete the feature fusion at the decision-making level. Moreover, this paper uses the OTB50 dataset [4] as the Object Tracking Benchmark to perform experimental verification of the proposed algorithm under different situations and evaluates the tracking accuracy and success rate with the comparative algorithms. Experimental results show that the robustness of the proposed algorithm for target tracking can improve to a certain extent.

II. RELATED WORKS

In recent years, the accuracy and speed of target tracking algorithms have improved significantly. Current tracking algorithms can mainly divide into two categories. One is deep learning-based target tracking algorithms using the features from convolutions, and another is tracking algorithms based on correlation filtering. Many effective and novel algorithms have proposed in target tracking technology, and the target tracking algorithm based on Deep Learning is more prominent in tracking performance [5]. Compared with these complex methods based on Deep Learning, some correlation filter algorithms are also competitive in efficiency [6]. By modeling the problem in the frequency domain, for the translation of the image sequence, the output of the linear classifier can be obtained by one single calculation [7]. It may use fewer computing resources to achieve high frame rate and performance, and multiple filters can be more widely used in target tracking.

Because of the rich hierarchical features in the convolutional neural network (CNN), the performance of deep learning-based target tracking algorithms can be superior. The CNN model has achieved great success in many computer vision tasks, especially image processing [8]. The high-level feature of the CNN model pays more attention to semantic information [9], which plays an important role in target classification, as well as, the low-level feature pays more attention to local details, which plays an essential role in target positioning [10]. However, due to the complexity of the model and the time-consuming training, the introduction of deep convolutional features leads to the tracking algorithm failing to meet real-time requirements in practice [11]. The difference between deep learning and traditional pattern recognition is that it can automatically learn features from large-scale data rather than hand-designed features [12]. Therefore, the CNN model can adopt as a feature extractor from different convolutional layers.

On the other hand, due to the increasing importance of real-time in target tracking algorithms, correlation filtering algorithm is widely used in target tracking, which achieved rapid development and remarkable results based on the double consideration of speed and performance. The color feature Color Name (CN) [13] was adopted to strengthen the traditional Circulant Structure Kernels (CSK) algorithm. Then to solve the problem of excessive computation of CN, the Kernel Correlation Filter [14] was proposed based on CSK, which introduced the characteristics of Histogram of Oriented Gradient (HOG) and reduced the computational strength by using the diagonalization property of Circulant Matrix. Moreover, based on the CSK algorithm, the Discriminative Scale Space Tracking (DSST) [15] was proposed with an improved scale strategy and scale filter. These research works can enhance the performance of the correlation filter tracking algorithm.

With a good balance between tracking accuracy and efficiency, the KCF-based target tracking has become a dominant approach in online object tracking. However, the features used by the KCF algorithm have limited capability to express the target with complex background. The emergence of these problems makes target tracking more difficult. To improve the precision and robustness of the target tracking algorithm based on the correlation filter, this paper abstracts deep features of the tracked target complementary between different characteristics, and fuses the features at the decision-making level to obtain a new tracking response diagram through the weighted processing of variant features.

III. Target Tracking Algorithm based on Deep Feature Fusion

In the target tracking algorithm, after the tracking region is determined, it is necessary to select the appearance characteristics of the target to determine the location of the target. Different appearance features have various effects on the performance of the target tracking algorithm. The Haar-like feature can cope with low image resolution, but it is ineffective in scenes with similar background colors. The HOG feature has good accuracy in the case of diverse background colors, but it does not adapt to image blur. The CN feature has high tracking accuracy but shows poor real-time performance. Local Binary Pattern (LBP) has a fast feature computation speed and high real-time performance, but the tracking accuracy is weak. The traditional algorithm often uses single and manual selected features for tracking to model the target, which may perform well under specific factors, but the overall accuracy is not ideal in complex situations. It expects that the performance of the target tracking algorithm can improve effectively by extracting deep features of tracking targets automatically and combing multiple features for the robust tracking of the targets in complex situations.

A. Deep Feature Extraction based on Convolutional Neural Network

The Kernel Correlation Filter algorithm uses the directional gradient histogram features to describe the target. Although it performs well in the tracking effect when motion blur and illumination change, it is not ideal for the tracking effect when the target is rapidly deformed and occluded. In the process of KCF, the first key puzzle is feature extraction, which is also the most crucial step in KCF. The manually selected features represent almost the shallow features of the image, which have limited capability to express the target with complex background.

Based on the advantage of graduated feature extraction in the CNN model, this paper adopts the off-line training method to pre-train the CNN model under the sample images with variant inferences such as rotation, deformation, fast motion, illumination variation, and so on. As we know, the convolutional features extracted at the lower level have more details, which can help locate the target accurately. Although the convolutional features of the deeper level contain more semantic information, this information with more fuzzy visualization is not suitable for accurate target tracking [16]. Therefore, the outputs of convolutional features from the first three layers selected by experimental comparison can train the correlation filter, and the deep features of different convolutional layers can combine in a weighted way. In the third layer, the maximum value of the response map can obtain, and after the supplement of regression weight, it is transferred layer by layer to the response diagram of the lower layer.

After the pre-training and fine-tuning of the CNN model (this paper adopts a network architecture consistent with the VGG model and fine-tunes the parameters of the pre-trained CNN model by using the target images), the final output features can gradually abstract from local details to high-level semantic information []. Based on this type of off-line training of the CNN model, we can obtain the most suitable features Haar-like feature, HOG feature, CN feature, and LBP feature, which can represent the detailed characteristic of target tracking in the complex environments with variant inferences. Based on the deep feature extraction results, the next step is how to make full use of the advantages of these deep features, and improve the accuracy of target tracking. This paper adopts the target tracking method based on multiple feature fusion and tries to fuse the deep features of the target image at the decision layer.

B. The Design of Multiple Feature Fusion Strategy

There are two kinds of multiple feature fusion methods: feature-level fusion and decision-level fusion. The feature-level fusion fuses different features into one feature processing, and the combined feature vector is represented as a single feature kernel. The decision-level fusion extracts multiple features from the appearance model and trains classifiers independently on the corresponding features. Then it combines different classifiers to form a model with complementary features after training classification. Based on the same tracking results of all features, the appearance model of the tracked target extracted by multiple different features can effectively increase the anti-interference capability of the combined model. Therefore, the fusion method at the decision level based on the KCF algorithm is adopted.

The fusion of multiple feature weights assigns different values to each feature to represent the proportion of the features in the description target. In the training process of the classifier, the value in the range of [0,1] uses to mark the samples. The closer the sample is to the target, the higher the tag weight is, and the closer the tag value obtained is to 1. When the sample target is far away, the smaller the weight of this tag is, the closer the tag value is to 0.

The candidate model compares with the reference target region tracked in the previous frame. Then the similarity between the candidate model and the target model can evaluate by the distance function based on the Bhattacharyya coefficient.

$$\rho\left(q_u(y_0), p_u\right) = \sum_{u=1}^{m} \sqrt{q_u p_u} \tag{1}$$

In Equation 1, the value of ρ is in the range [0,1], p_u is the candidate model under different features represented by the histogram of the current frame, and q_u is the target reference model of the previous frame. The larger the Bhattacharyya coefficient ρ is, the more similar the candidate model is to the target model, and the closer it is to the actual location of the target.

When detecting the tracked target, the fusion features need to follow the principles of additive fusion and multiplicative fusion. This paper adopts an additive fusion strategy to measure the reliability of tracking results in each frame through the Bhattacharyya coefficient. The reliability of the tracking result is measured by the Bhattacharyya distance between the peak value of the response graph and the target model to achieve dynamic weight adjustment. According to the weight coefficient corresponding to different features, the confidence matrix is multiplied by all features. Then all features are added to obtain the weighted confidence matrix, which is the final detection result.

$$f(x) = r_{Haar} f_{Haar}(x) + r_{HOG} f_{HOG}(x) + r_{CN} f_{CN}(x) + r_{LBP} f_{LBP}(x) \tag{2}$$

In Equation 2, r_{Haar}, r_{HOG}, r_{CN}, and r_{LBP} represent the weights of different features during fusion, where $r_{Haar}+r_{HOG}+r_{CN}+r_{LBP}=1$. And $f_{Haar}(x)$, $f_{HOG}(x)$, $f_{CN}(x)$, and $f_{LBP}(x)$ present the target positions of tracking based on Haar-like feature, HOG feature, CN feature, and LBP feature, which range from 0 to 1. Since the OTB dataset includes a 25% gray image sequence, which is difficult to extract the CN feature, the gray image needs to be fused with the other three features to represent the target. That is to say, the value of r_{CN} is 0 for the gray image, and all the features extracted above can use to represent the target for the color image.

When the target tracks the current frame, different maximum response values of each feature in the previous frame are used to calculate the weights of different features. According to the difference of each feature, the maximum tracking response value obtained compares with the target. The larger the response value is, the better the description effect of the feature is for the target. To evaluate the similarity of modified features according to the Bhattacharyya coefficient, higher weights should be set to the feature that best represents the tracked target and distinguishes the target from the background.

$$\begin{cases} r_{Haar} = \dfrac{\rho_{Haar}}{\rho_{Haar}+\rho_{HOG}+\rho_{CN}+\rho_{LBP}} \\ r_{HOG} = \dfrac{\rho_{HOG}}{\rho_{Haar}+\rho_{HOG}+\rho_{CN}+\rho_{LBP}} \\ r_{CN} = \dfrac{\rho_{CN}}{\rho_{Haar}+\rho_{HOG}+\rho_{CN}+\rho_{LBP}} \\ r_{LBP} = \dfrac{{}_{LBP}}{\rho_{Haar}+\rho_{HOG}+\rho_{CN}+\rho_{LBP}} \end{cases} \tag{3}$$

In Equation 3, ρ_{Haar}, ρ_{HOG}, ρ_{CN}, and ρ_{LBP} present the similar degrees obtained by the Bhattacharyya coefficient of the target tracking based on the Haar-like feature, HOG feature, CN feature, and LBP feature separately.

This multiple feature fusion strategy has two advantages. Firstly, when the tracked target is in a complex environment, the multiple feature description at the decision level can describe the tracked target specifically. Secondly, in the whole process of target tracking, the algorithms used to track different features of the target can slightly adjust according to the specific situation to predict the next target position.

C. The Design of Target Tracking Algorithm based on Multiple Feature Fusion

When the trained model is used to search the target, the classifier can test the samples in the region and calculate the response value of the candidate block. For the newly input image block z, the classifier response output of the KCF tracking algorithm is as below.

$$f(z) = w^T z = \sum_{i}^{n} \alpha_i k(z, x_i) \tag{4}$$

The calculation speed can greatly increase when all samples are tested, because the candidate blocks are also tested by cyclic shift. K^Z is used to represent the kernel matrix between all training and test samples. The training samples are obtained by the base sample x cyclic shift, and the test samples are also obtained by the z cyclic shift. The following unitary invariant kernel function is adopted in this algorithm.

$$\begin{gathered} f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n cos\frac{n\pi x}{L} + b_n sin\frac{n\pi x}{L}\right) \\ K^z = C(k^x z) \end{gathered} \tag{5}$$

In Equation 5, $k^x z$ represents the kernel correlation between x and z. Based on the linear relationship of discrete Fourier transform, the sum of the results of each channel in the Fourier domain can be further calculated, and then the classifier response of all candidate image blocks can be calculated as below.

$$f(z) = \alpha (K^z)^T \tag{6}$$

In Equation 6, f(z) is the output vector that contains all test samples (all test results). Based on the above designs, the target tracking algorithm based on multiple feature fusion can implement, as shown in Table 1.

TABLE I. THE DESCRIPTION OF TARGET TRACKING ALGORITHM BASED ON MULTIPLE FEATURE FUSION

Target tracking algorithm based on multiple feature fusion
Input: The image sequence and position of the target in the first frame
Output: Target location corresponding to the current frame

Step1: Based on the extraction of the positive and negative samples, the target tracking frame information is determined in the first frame, and the target model q_1 can be obtained.
Step2: When reading the u-th frame, its features are extracted, and each feature is used to search the target position separately to evaluate the similarity ρ between multiple feature candidate models and target model q_{u-1} based on Equation 1.
Step3: The corresponding weight is set to the obtained feature similarity based on Equation 3, and then weighted calculation is executed based on Equation 2.
Step4: The position of the tracking target in the current frame is estimated based on the weighted response diagram.
Step5: The appearance model in the current frame of the target is updated based on Equation 5 and Equation 6.
Step6: If current frame is the last one, then target tracking is completed. Otherwise, next frame will be read, and the above operations will repeated until the end of the data.

IV. THE EXPERIMENTAL VERIFICATION AND ANALYSIS

With the rapid progress of computer vision technology, to evaluate the performance and effect of target tracking algorithms more objectively, many researchers have provided image sequence datasets (Object Tracking Benchmark) for comparative tests respectively, such as dataset OTB and VOT. The difference between OTB and VOT lies in that OTB includes a 25% gray sequence, while VOT is an all-color sequence. Therefore, different datasets lead to differences in some color feature algorithms, and the algorithm evaluation criteria of OTB and VOT are also different. To evaluate the performance of the proposed target tracking algorithm and evaluate the performance under different interferences, such as Illumination Variation (IV), Scale Variation (SV), Deformation (DEF), Out-Of-View (OV), Motion Blur (MB), and Fast Motion (FM), this paper adopts the OTB50 dataset for comparative experiment and performance verification.

A. Experimental Dataset and Evaluation Indicators

The OTB50 consists of 50 image sequences, and it was proposed in 2013 (also known as OTB-2013). Each group of image sequences in this dataset is manually marked with the specific positions of the tracking target. In the comparative experiment, we compare the experimental tracking location with the manually marked locations, and we can obtain the tracking accuracy and success rate of the target tracking algorithm. The OTB50 dataset consists of Basketball, Biker, Bird1, and other image sequences. Each image sequence in the OTB dataset is composed of several images as the frames in the corresponding image sequence. The final tracking result can obtain by tracking each frame with the algorithm. In general, precision degree and success rate are two kinds of evaluation indicators adopted to evaluate the performance of the target tracking algorithm.

1) Precision degree

The evaluation standard based on center position tracking accuracy is adopted to represent the overall performance of the tracking sequence by using the average center error of all frames. The average pixel error value is calculated on the pixel distance between the predicted center position and actual position of the target. The greater the error value is, the lower the precision degree is, as shown below.

$$CE_2 = (x_t^g - x_t^0)^2 + (y_t^g - y_t^0)^2 \tag{7}$$

In Equation 7, (x_t^0 , y_t^0) presents the center point coordinates of the target position estimated by the algorithm for one frame, and (x_t^g, y_t^g) presents the center coordinates of the position occupied by the target in the same frame.

When the algorithm cannot track the target, it will generate a random output position, resulting in the average error obtained, which cannot accurately evaluate the tracking performance. Therefore, the accuracy curve is usually adopted to evaluate the overall tracking performance of the algorithm, which can predict the frame proportion in the specific threshold interval between the predicted position and actual position, and the threshold is set to 20 pixels generally. The percentages often change under different thresholds, and the comparative curve can obtain. Nevertheless, this accuracy evaluation method cannot reflect the change in object size and scale.

2) Success rate

The overlap rate is defined as the expected overlap area between the reference interface and tracking interface, and the proportion of frames is higher than the fixed threshold in the total number of frames in the sequence. When the coincidence rate is greater than the given threshold, the frame is regarded as successful, and the percentage of the total number of successful frames to all frames is the success rate. The higher the rate is, the more frames are consistent with the condition, and the more significant the tracking effect is achieved.

$$\rho = \frac{|r_t \cap r_a|}{|r_t \cup r_a|} \tag{8}$$

In Equation 8, r_t presents the tracking box in a frame evaluated by the algorithm, and r_a presents the actual tracking box of the tracked target in this frame.

Usually, success rate represents the proportion of the number of frames matching the condition in all frames during the period of increasing overlap rate from 0 to 1. If only one threshold rate is adopted, it is not representative to evaluate the overall performance of the tracking algorithm. Therefore, the success rate of the tracking algorithm can evaluate based on the area size below the success curve.

This paper adopts the One-Pass Evaluation (OPE) method to compare the tracking performance of different target tracking algorithms and evaluate the efficiency of multiple feature fusion in target tracking through experimental comparison. OPE gives the exact position of the tracked target tracking frame from the first frame, which compares the target frame obtained by the target tracking algorithm with the given tracking frame in the subsequent frames of the image sequence, and calculates the accuracy and success rate of the target tracking.

B. Experimental Results and Analysis

This paper adopts the same experimental environment for the target tracking algorithm experiments, with the hardware configuration of Intel Core i7-9700k CPU (main frequency 3.6GHz) and 16GB memory, and the software platform of MATLAB 2020. The algorithm code is written in MATLAB and C language. The default parameters in the traditional KCF algorithm are followed, and the original experimental parameters of the comparison algorithms are adopted.

The typical correlation filter algorithms of target tracking involved in this comparative experiment include the CN, CSK, KCF, DSST, SAMF, fDSST, Staple, and SRDCF. The CN, CSK, and KCF algorithms are all classical algorithms of correlation filtering in the field of target tracking. The DSST proposes an accurate target scale estimation strategy, and fDSST is the accelerated optimization of DSST based on dimension reduction and interpolation. The SAMF [17] algorithm integrates HOG features and CN features based on a correlation filtering framework and uses an adaptive target size strategy for target tracking. The Staple [18] algorithm uses histogram information to assist tracking based on correlation filtering. And SRDCF (Spatially Regularized Discriminative Correlation Filters) [19] is an improved DCF (Discriminative Correlation Filter) algorithm with a more discriminative appearance model.

1) The overall performance evaluation of target tracking algorithms

Through the comparative experiments of the proposed algorithm (ours) and other comparison algorithms on the OTB50 dataset, the tracking accuracy of different algorithms in various error thresholds and the tracking success rate based on multiple overlapping thresholds can obtain, as shown in Fig. 1.

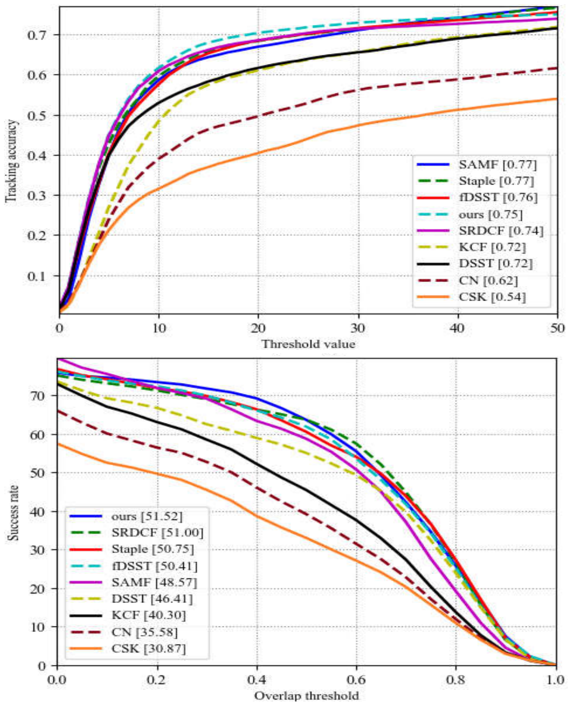


Fig. 1. The overall performance evaluation of different algorithms based on the OTB50 dataset

As shown in Fig. 1, the tracking accuracy increases with the rising of the center error threshold, while the target tracking success rate decreases with the rising of the overlap threshold. Based on the experiment results, the tracking accuracy of the proposed algorithm is 75%, and the tracking success rate is 51.52%, which are both better than the KCF algorithm. The overall tracking accuracy is improved by 3%, and the tracking success rate also improved by 11.22%, which can verify the improvement effect of the traditional KCF algorithm in this paper. Moreover, compared with CN and CSK algorithms, the tracking accuracy and success rate are significantly improved, which increases the tracking accuracy by 13% and 21%, and the tracking success rate by 15.94% and 20.65%, respectively. And compared with DSST and SRDCF algorithms, the tracking accuracy and tracking success rate are also improved to a certain degree by 3%, 5.11%, and 1%, 1.11%, respectively. Furthermore, the proposed algorithm can achieve almost the same performance as SAMF, fDSST, and Staple algorithms, especially the success rate.

The experimental results show that based on the abstraction of deep features by the CNN and the fusion of multiple features by the complementary characteristics, we can combine the advantages to achieve a better target tracking algorithm. In terms of overall tracking success rate, the proposed algorithm in this paper has good performance, which means that it is not easy to lose the target in the target tracking process based on multiple features fusion. On the other hand, in terms of overall tracking accuracy, the tracking performance of the proposed algorithm does not achieve a better degree than the SAMF, fDSST, and Staple algorithms. The Staple algorithm uses color histogram information to assist its tracking, and the SAMF algorithm combines CN and HOG features with the strategy of adaptive target size, which makes the tracking accuracy of these algorithms still have some advantages.

2) Quantitative comparison of tracking performance under different interferences

To evaluate the tracking accuracy and success rate of the multiple feature fusion algorithms more comprehensively, this paper performs further experiments with six different interferences in the OTB50 dataset respectively and obtains the comparison diagram of tracking accuracy and success rate, as shown in Fig. 2 to Fig. 7 successively.

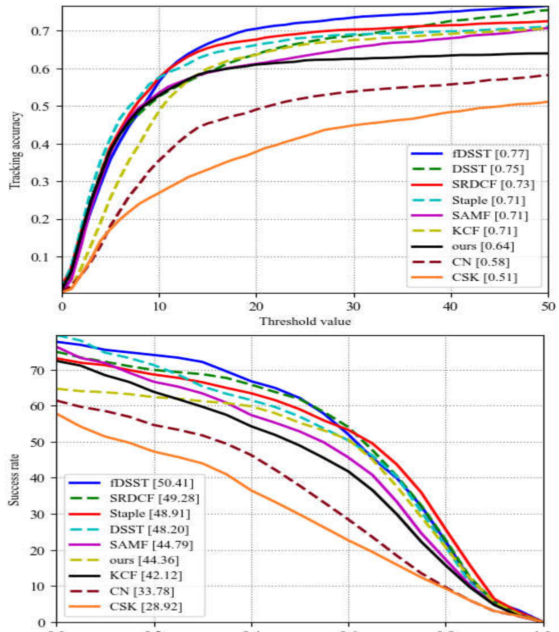


Fig. 2. The comparison diagram of tracking accuracy and success rate with illumination variation

Fig. 2 shows the comparison results with illumination variation. The proposed algorithm in this paper does not have strong adaptability to IV. The success rate is only 2.24% higher than the KCF algorithm, and the tracking accuracy is even slightly lower than it.

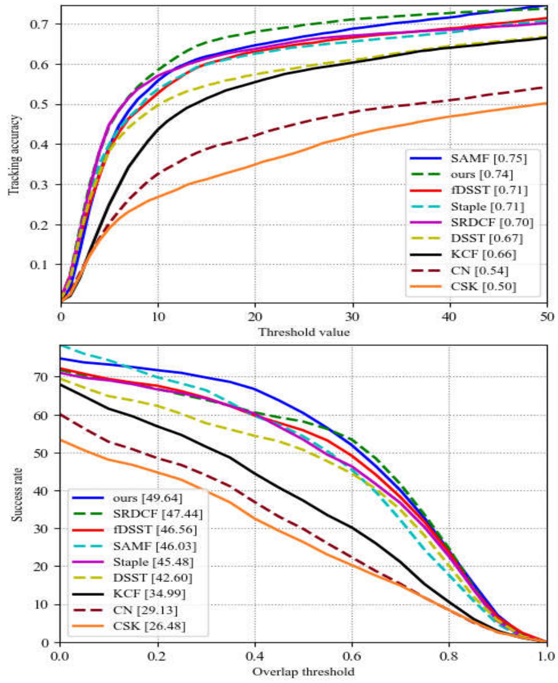


Fig. 3. The comparison diagram of tracking accuracy and success rate with scale variation

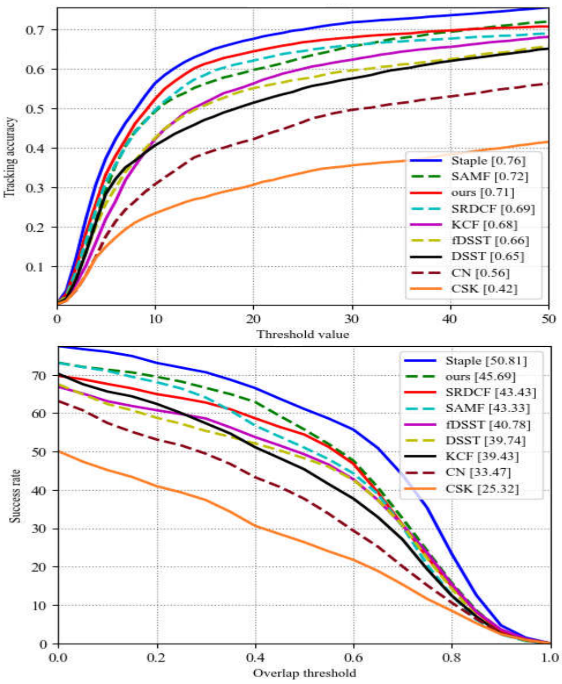


Fig. 4. The comparison diagram of tracking accuracy and success rate with deformation

Fig. 3 shows the comparison results with scale variation. The proposed algorithm shows good tracking performance under SV, and both tracking accuracy and success rate are better than other algorithms. The tracking accuracy and the success rate are 8% and 14.65% higher than the KCF algorithm, and the tracking success rate is higher than other comparative algorithms. Moreover, it also shows that the fused features are not affected by the change in target scale, and the success rate of tracking is significantly improved.

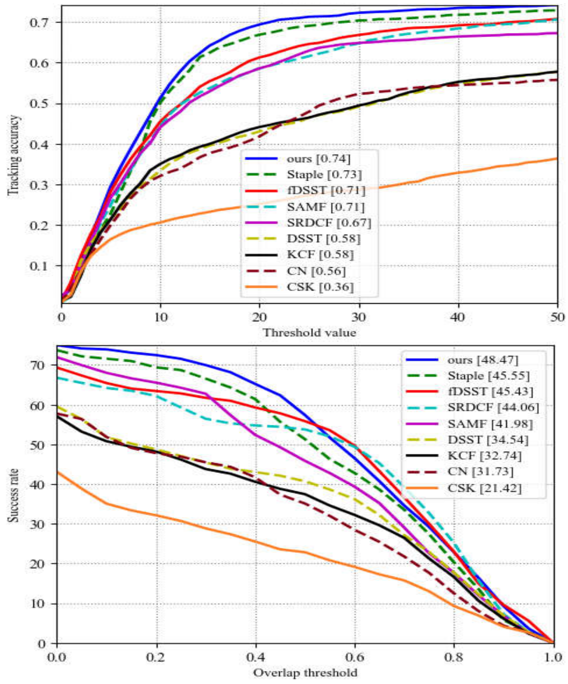


Fig. 5. The comparison diagram of tracking accuracy and success rate with out-of-view

Fig. 4 shows the comparison results with deformation, and Fig. 5 shows the comparison results with out-of-view. The DEF and OV of the target are common situations in actual environments, and the performance of the algorithm is very important under similar circumstances. Through the experimental results, it can be found that the proposed algorithm can still show good tracking performance under the interference of DEF and OV, especially the out-of-view situation of targets, both in tracking accuracy and success rate.

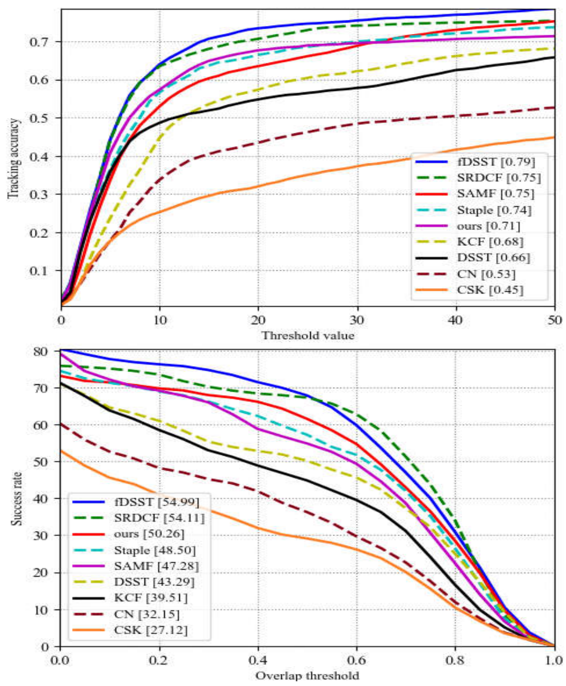


Fig. 6. The comparison diagram of tracking accuracy and success rate with motion blur

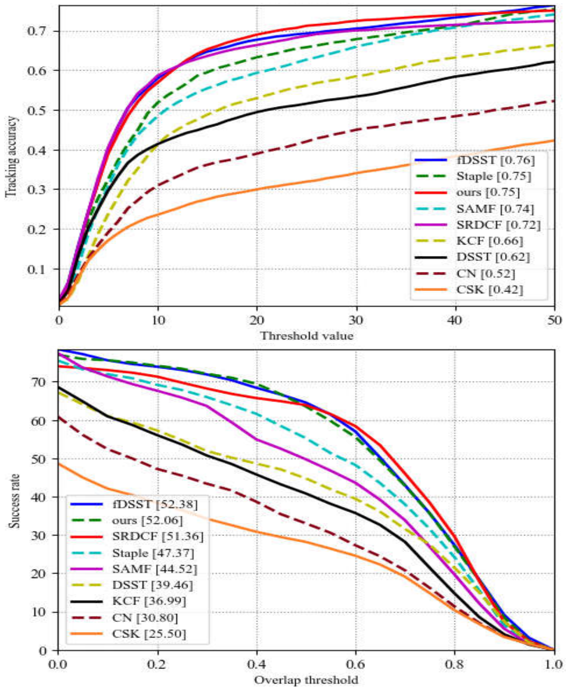


Fig. 7. The comparison diagram of tracking accuracy and success rate with fast motion

Fig. 6 and Fig. 7 show the comparison results with motion blur and fast motion. The tracking accuracy and the success rate of the proposed algorithm in the case of MB are 71% and 50.26% respectively. Compared with the KCF algorithm, the target tracking accuracy of the proposed algorithm only improves by 3%, which may be due to the appearance model of the tracked target in the case of MB, and the difference between the extracted feature points is too large to model the target accurately. However, the tracking success rate is significantly improved compared with the KCF algorithm. In addition, the proposed algorithm has improved both tracking accuracy and tracking success rate under FM, which shows that FM has little impact on it.

In long-term target tracking, there are often various interference situations. Therefore, ensuring tracking accuracy and success rate are crucial standards to measure its performance in a complex environment. The performance comparison of quantitative analysis based on tracking accuracy and success rate are shown in Table 2 and Table 3 more concretely.

TABLE II. THE COMPARISON OF TRACKING ACCURACY OF DIFFERENT TARGET TRACKING ALGORITHMS IN DIFFERENT SITUATIONS

	CSK	KCF	DSST	SAMF	Staple	SRDCF	fDSST	ours
IV	0.51	0.71	0.75	0.71	0.71	0.73	**0.77**	0.64
SV	0.50	0.66	0.67	**0.74**	0.71	0.70	0.71	**0.74**
DEF	0.42	0.68	0.65	0.72	**0.76**	0.69	0.66	0.71
OV	0.36	0.58	0.58	0.71	0.73	0.67	0.71	**0.74**
MB	0.45	0.68	0.66	0.75	0.74	0.75	**0.79**	0.71
FM	0.42	0.66	0.62	0.74	0.75	0.72	**0.76**	0.75

TABLE III. THE COMPARISON OF TRACKING SUCCESS RATE OF DIFFERENT TARGET TRACKING ALGORITHMS IN DIFFERENT SITUATIONS

%	CSK	KCF	DSST	SAMF	Staple	SRDCF	fDSST	ours
IV	28.92	42.12	48.20	44.79	48.91	49.28	**50.41**	44.36
SV	26.48	34.99	42.60	46.03	45.48	47.44	46.56	**49.64**
DEF	25.32	39.43	39.74	43.33	**50.81**	43.43	40.78	45.69
OV	21.42	32.74	34.54	41.98	45.55	44.06	45.43	**48.47**
MB	27.12	39.51	43.29	47.28	48.50	54.11	**54.99**	50.26
FM	25.50	36.99	39.46	44.52	47.37	51.36	**52.38**	52.06

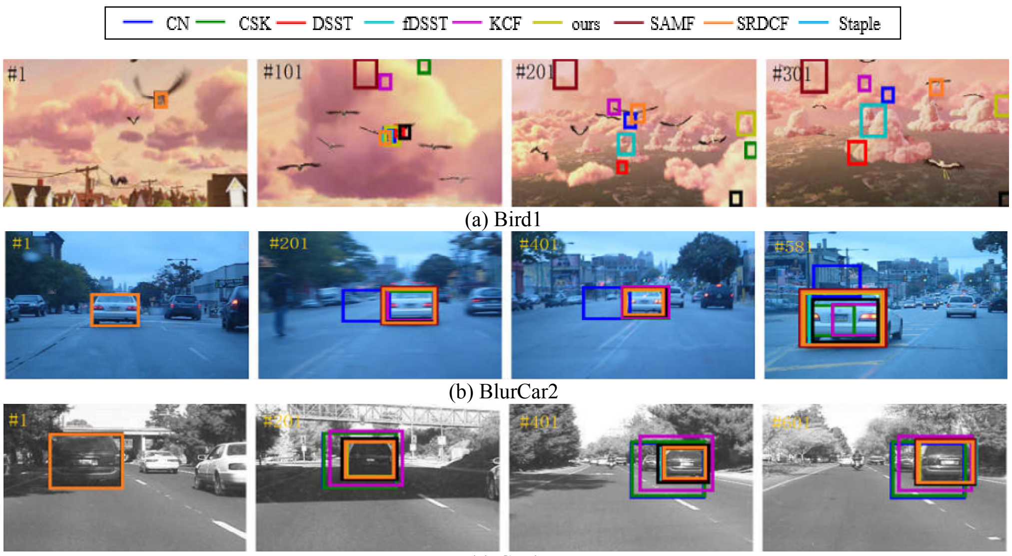
From the comparative results under different situations with inferences, the tracking accuracy and success rate are almost better than traditional KCF tracking algorithm and can also show certain performance advantages with other comparative algorithms.

3) Qualitative comparison of tracking performance based on different datasets

In this experiment, three typical datasets (Bird1, Blurcar2, and Car4) are selected from the OTB50 dataset with variant inferences to compare and analyze the tracking performances qualitatively. The information of each dataset is listed in Table 4, and Fig. 8 shows the tracking result of different algorithms on some frames of the datasets.

TABLE IV. THE INFORMATION OF THE SELECTED EXPERIMENT DATASETS IN THE OTB50

Image Dataset	Frames	Inferences
Bird1	408	DEF, FM, OV
BlurCar2	585	SV, MB, FM
Car4	659	IV, SV



(a) Bird1

(b) BlurCar2

(c) Car4

Fig. 8. The comparison diagram of different algorithm tracking results based on selected datasets

As shown in Fig. 8, it appears a great impact on the target tracking algorithm with the interference of DEF, FM, and OV in dataset Bird1. However, the proposed algorithm in this paper has better performance than the KCF algorithm. And in dataset BluCar2 with the interference of SV, MB, and FM, it almost does no influence the proposed algorithm, and the overall tracking effect is good. Moreover, in dataset Car4 with the situation of IV and SV, it has not much impact on the target tracking effect of the proposed algorithm. Based on the integration of multiple features at the decision-making level, the performance of the proposed algorithm is better than the KCF and CSK algorithms.

The KCF target tracking algorithm based on deep feature and multiple feature fusion proposed in this paper tracks the target position respectively through each feature of the target. Based on the advantages of each deep feature in different environments, the results are weighted by the Bhattacharyya coefficient to complete the feature fusion at the decision-making level. Through the target tracking performance comparison experiments with typical tracking algorithms under different image datasets with variant interference conditions, we can see that the proposed algorithm has a better performance and higher robustness.

V. CONCLUSIONS

Based on the investigation of typical target tracking algorithms, this paper points out that current target tracking algorithms can hardly track accurately under complex interference environment. Therefore, the traditional algorithms should improve the model characteristics and updating mechanism to ensure tracking performance. To address the limitations of target tracking algorithms in complex environments, this paper proposes an improved KCF algorithm from the perspective of multiple feature fusion, which uses the complementary characteristics of deep features to perform a weighted fusion of the target at the decision-making level. The deep features of the target are abstracted by the CNN model, and these features are weighted and fused by the Bhattacharyya coefficient to improve the target tracking accuracy and success rate. Through the comparative experiment on the OTB50 dataset, the overall accuracy and success rate can be improved to 0.75% and 51.52% respectively. Moreover, based on further experiments under six complex situations with different interferences, the proposed algorithm shows good performance and high robustness. In this paper, the improvement of the target tracking algorithm based on KCF has achieved some promising results, but there are still some shortcomings and defects. The weighted fusion and added update strategy may increase the amount of calculation and affect the real-time performance, and the tracking performance in situations of occlusion, rotation, background clutter, and low resolution needs to be improved further.

ACKNOWLEDGMENT

We sincerely acknowledge the support from the Natural Science Foundation of Hebei Province (D2022203002) and the Central Guidance on Local Science and Technology Development Fund of Hebei Province (202260201013242).

REFERENCES

[1] A. Brunetti, D. Buongiorno, G. F. Trotta and V. Bevilacqua, “Computer vision and deep learning techniques for pedestrian detection and tracking,” Neurocomputing, vol. 300, pp. 17-33, 2018.

[2] X. Yang, S. Li, J. Yu, K. Zhang, J. Yang and J. Yan, “GF-KCF: Aerial infrared target tracking algorithm based on kernel correlation filters under complex interference environment,” Infrared Physics and Technology, vol. 119, pp. 1-15, 2021.

[3] C. Kwan, B. Chou and L. M. Kwan, “A Comparative Study of Conventional and Deep Learning Target Tracking Algorithms for Low Quality Videos,” 15th International Symposium on Neural Networks, Minsk, Belarus, pp. 25-28, June 2018.

[4] Visual Tracker Benchmark. Available online: http://www.visual-tracking.net/ (accessed on 20 May 2020)

[5] C. Kwan, B. Chou, J. Yang and T. Tran, “Deep Learning Based Target Tracking and Classification for Infrared Videos Using Compressive Measurements,” Journal of Signal and Information Processing, vol. 10, no. 4, pp. 167-199, 2019.

[6] H. Peng, J. Li, Y. He, Y. Liu, M. Bao, L. Wang, Y. Song and Q. Yang, “Large-Scale Hierarchical Text Classification with Recursively Regularized Deep Graph-CNN,” 26th International World Wide Web Conference, Lyon, France, pp.1063-1072, April 2018.

[7] L. Zhang and X. Wang, “Improved Multiple Hypothesis Tracker for Joint Multiple Target Tracking and Feature Extraction,” IEEE Transactions on Aerospace and Electronic Systems, vol. 55, no. 6, pp. 3080-3089, 2019.

[8] X. Hu, W. Shi, Y. Zhou, H. Tang and S. Duan, “Quantized and adaptive memristor based CNN (QA-mCNN) for image processing,” Science China Information Sciences, vol.65, pp. 119104, 2022.

[9] J. Gong, S. Wang, J. Wang, W. Feng, H. Peng, J. Tang and P. S. Yu, “Attentional Graph Convolutional Networks for Knowledge Concept Recommendation in MOOCs in a Heterogeneous View,” International ACM SIGIR Conference on Research and Development in Information Retrieval, Xiamen, China, pp. 79-88, July 2020.

[10] H. Peng, J. Li, Q. Gong, Y. Ning, S. Wang and L. He, “Motif-matching based Subgraph-level Attentional Convolutional Network for Graph Classification”, 34th AAAI Conference on Artifical Intelligence, New York, USA, pp. 5387-5394, February 2020.

[11] N. Mahmoudi, S. M. Ahadi and Mohammad Rahmati, “Multi-target tracking using CNN-based features: CNNMTT,” Multimedia Tools and Applications, vol. 78, no. 6, pp. 7077-7096, 2019.

[12] H. Peng, R. Yang, Z. Wang, J. Li, L. He, P. S. Yu, A. Y. Zomaya and R. Ranjan, LIME: Low-Cost and Incremental Learning for Dynamic Heterogeneous Information Networks, IEEE Transactions on Computers, vol. 71, no. 3, pp. 628-642, 2021.

[13] T. Zhang, C. Xu and M. Yang, “Robust structural sparse tracking,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 473-486, 2019.

[14] W. Chen, K. Zhang and Q. Liu, “Robust visual tracking via patch based kernel correlation filters with adaptive multiple feature ensemble,” Neurocomputing, vol. 214, pp. 607-617, 2016.

[15] M. Danelljan, G. Häger, F. S. Khan and M. Felsberg, “Discriminative Scale Space Tracking,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 8, pp. 1561-1575 , 2017.

[16] H. Peng, J. Li, S. Wang, L. Wang, Q. Gong, R. Yang, B. Li, L. He and P. S. Yu, “Hierarchical Taxonomy-Aware and Attentional Graph Capsule RCNNs for Large-Scale Multi-Label Text Classification,” IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 6, pp. 2505-2519, 2021.

[17] H. Peng, J. Li, Y. Song, R. Yang, R. Ranjan, P. S. Yu and L. He, “Streaming Social Event detection and evolution discovery in heterogeneous information networks,” ACM Transactions on Knowledge Discovery from Data, vol. 15, no. 5, pp. 1-33, 2021.

[18] J. Gong, Y. Mei and Y. Zhou, “Research on an Improved KCF Target Tracking Algorithm Based on CNN Feature Extraction,” IEEE International Conference on Artificial Intelligence and Computer Applications, Dalian, China, pp. 538-543, June 2020.

[19] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik and P. H. S. Torr, “Staple: Complementary Learners for Real-Time Tracking,” IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, pp. 1401-1409, June 2016.

[20] X. Meng and J. Duan, “An object tracking algorithm based on SRDCF and deformable diversity similarity,” IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, Vancouver, Canada, pp. 1-3, November 2018.