Robust Probabilistic Discriminative Model Prediction Tracker via Improved Model Update Strategy

Shaolong Chen

Sun Yat-Sen University

Changzhen Qiu

Sun Yat-Sen University

Yurong Huang

Sun Yat-Sen University

Zhiyong Zhang (✉ zhangsysu2020@163.com)

Sun Yat-Sen University <https://orcid.org/0000-0003-0638-5434>

Research Article

Keywords: Object tracking, discriminative model prediction, model update strategy, feature fusion

Posted Date: December 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1138271/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Robust probabilistic discriminative model prediction tracker via improved model update strategy

Shaolong Chen · Changzhen Qiu · Yurong Huang ·
Zhiyong Zhang

Received: date / Accepted: date

Abstract In the visual object tracking, the tracking algorithm based on discriminative model prediction have shown favorable performance in recent years. Probabilistic discriminative model prediction (PrDiMP) is a typical tracker based on discriminative model prediction. The PrDiMP evaluates tracking results through output of the tracker to guide online update of the model. However, the tracker output is not always reliable, especially in the case of fast motion, occlusion or background clutter. Simply using the output of the tracker to guide the model update can easily lead to drift. In this paper, we present a robust model update strategy which can effectively integrate maximum response, multi-peaks and detector cues to guide model update of PrDiMP. Furthermore, we have analyzed the impact of different model update strategies on the performance of PrDiMP. Extensive experiments and comparisons with state-of-the-art trackers on the four benchmarks of VOT2018, VOT2019, NFS and OTB100 have proved the effectiveness and advancement of our algorithm.

Keywords Object tracking · discriminative model prediction · model update strategy · feature fusion

1 Introduction

With the widespread application and development of technologies such as video behavior analysis, autonomous driving, and human-computer interaction, visual object tracking technology has attracted people's attention. Many advanced object tracking methods have been proposed, such as optical flow [1], particle filter [2], Mean shift [3], MOSSE [4], CSK [5], KCF [6]. However, object tracking still faces many challenges. When complex situations such as changes in the appearance of the tracking object and background interference occur, it is easy to cause tracking failure. Therefore, it is still necessary to study in-depth methods with higher accuracy and robustness.

✉ Zhiyong Zhang
E-mail: zhangzhy99@mail.sysu.edu.cn

School of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou, Guangdong, China

In the past few years, deep learning have achieved milestones in computer vision field [7, 8]. Many object tracking algorithms based on deep learning have been proposed [9, 10, 11]. The tracking algorithm based on deep learning can be divided into offline model update [12, 13, 14, 15, 16] and online model update [17, 18, 19]. In general, compared to offline model update, the online model update method has higher accuracy and better robustness. Therefore, the online model update method has become a widely concerned part in recent researches [18, 19]. However, the online model update is a double-edged sword. It can adapt to the appearance changes of objects and background, but it is also easy to be contaminated by noise samples, which leads to tracking drift. PrDiMP [19] is a typical online model update tracker, which integrates the maximum response and second-maximum response ratio to establish evaluation criteria to evaluate the tracking results. PrDiMP evaluate the tracking results through the output of the tracker to guide the online update of the model. However, the tracker output is not always reliable, especially in the case of fast motion, occlusion or background clutter.

In this paper, we present a robust model update strategy which can effectively integrate maximum response, multi-peaks and detector cues to guide model update of PrDiMP tracker. Furthermore, we have analyzed the impact of different model update strategies on the performance of PrDiMP. Extensive experiments and comparisons with state-of-the-art trackers on the four benchmarks of VOT2018, VOT2019, NFS and OTB100 have proved the effectiveness and advancement of our algorithm. This not only significantly improves the tracking performance of PrDiMP, but also can be easy to be embedded into other online model update trackers.

Contributions:

1. A robust model update strategy is proposed, which can effectively integrate maximum response, multi-peaks and detector cues to guide model update of PrDiMP.
2. We analyzed the impact of different model update strategies on the performance of PrDiMP in detail.
3. Our method not only has better results compared with the corresponding baseline method, but also better than other excellent target tracking methods (on OTB100, NFS, VOT2018 and VOT2019).

2 Related work

2.1 Visual object tracking

In the past few years, deep learning have achieved milestones in computer vision field. Many object tracking algorithms based on deep learning have been proposed [9, 10, 11]. Siamese architecture [12, 13, 14, 15, 24, 25, 26] has end-to-end training capabilities and high efficiency. However, the method based on the siamese architecture can't integrate background information, and its discriminative ability is limited. Based on this, DiMP [18] and PrDiMP [19] develop an end-to-end tracking architecture, which can make full use of the appearance information of object and background for object model prediction. This framework is based on the object model prediction network, which is derived from a discriminative learning loss by applying an iterative optimization procedure. It can achieve effective end-to-end training while maximizing the discriminative ability of the prediction model.

2.2 Online update for visual object tracking

In the field of visual object tracking, online model update plays an important role, which enables models to adapt to the changes of object appearance and their surrounding background. The online model update is a double-edged sword. It can adapt to the appearance changes of objects and background, but it is also easy to be contaminated by noise samples, which leads to tracking drift. In order to enhance the model's ability to adapt to the changes of object appearance and their surrounding background, while not making the model contaminated. Some researchers have done a lot of work by designing some criteria to evaluate the reliability of current tracking results, delete unreliable samples or reject inappropriate updates, such as the confidence score [27], the maximum response [17], peak-to-sidelobe rate [17], average peak-to-correlation energy [28], and MAX-PSR [29]. These methods usually evaluate the tracking results through the output of the tracker to guide the online update of the model. However, the tracker output is not always reliable, especially in the case of fast motion, occlusion or background clutter [30, 31]. In this paper, we present a robust model update strategy which can effectively integrate maximum response, multi-peaks and detector cues to guide model update of PrDiMP. This not only significantly improves the tracking performance of PrDiMP, but also can be easy to be embedded into other online model update trackers.

3 Methods

In this section, we introduce our robust model update strategy which integrate maximum response, multi-peaks and detector cues to guide model online update of PrDiMP. First, we analyzed the problem of PrDiMP's model update strategy (Section 3.1). Then we introduce our robust model update strategy in detail (Section 3.2).

3.1 Analyze the problem in online model update strategy of PrDiMP

The PrDiMP [19] is a tracker that integrates the probabilistic regression method into the DiMP [18]. It consists of two parts: I) A object estimation module that is learned offline; II) A object classification module that is learned online. In this section, we mainly analyze the online model update strategy of PrDiMP. For more detailed information about the PrDiMP, see [18, 19]. The PrDiMP integrates maximum response (MAX) and second-maximum response ratio (SMR) to guide model update. *MAX* is defined as the maximum value of the classification network response map R_t ,

$$MAX = \text{Max}(R_t) \quad (1)$$

Here, the subscript t denotes the t -th frame.

In order to calculate *SMR*, the response map is divided into maximum and sidelobe (the remaining pixels of the $M * M$ window are not included around the maximum). Defines second response MAX^{sl} is the maximum response of the sidelobe. Then *SMR* is defined as

$$SMR = MAX^{sl} / MAX \quad (2)$$

Based on *MAX* and *SMR*, PrDiMP divides the tracking results into four types:

Algorithm 1

```

If  $MAX < threshold_1$ 
    The tracking result is 'Not found';
Else If  $SMR > threshold_2$ 
    Calculate the distance  $D_c$  between the  $MAX$  prediction position and the prediction position of the
    previous frame;
    Calculate the distance  $D_{sl}$  between the  $MAX^{sl}$  prediction position and the prediction position of the
    previous frame;
    If  $D_c < threshold_3$  and  $D_{sl} > threshold_3$ 
         $MAX$  prediction position as the prediction position of the current frame;
        The tracking result is 'Hard negative';
    Else If  $D_c > threshold_3$  and  $D_{sl} < threshold_3$ 
         $MAX^{sl}$  prediction position as the prediction position of the current frame;
        The tracking result is 'Hard negative';
    Else
         $MAX$  prediction position as the prediction position of the current frame;
        The tracking result is 'Uncertain';
    End
Else If  $SMR > threshold_4$  and  $MAX^{sl} > threshold_1$ 
     $MAX$  prediction position as the prediction position of the current frame;
    The tracking result is 'Hard negative';
Else
     $MAX$  prediction position as the prediction position of the current frame;
    The tracking result is 'Normal';
End

```

'Not found': The training samples is not update; the object size and position are not update;

'Uncertain': The training samples is not update; the object size and position are update;

'Normal': Add the current result to the training samples and update the model regularly; the object size and position are update;

'Hard negative': Add the current result to the training samples and update the model immediately; the object size and position are update.

The detail strategy of the online model update is shown in **Algorithm 1**. The $threshold_1$, $threshold_2$, $threshold_3$, $threshold_4$ are manually set thresholds.

In PrDiMP, when $SMR > threshold_2$, the tracking result is evaluated according to the distance between the current double-peaks prediction position and the previous prediction position. The peak predicted position whose distance is less than threshold is selected as the current frame prediction position. However, the appearance of similar object and interference is usually random, and the distance between the interference and the predicted position of the previous frame is not necessarily greater than the distance between the object and the predicted position of the previous frame. Updating at this time may cause the model to be contaminated.

In addition, when $SMR > threshold_4$ and $MAX^{sl} > threshold_1$, the author regards the maximum response position as the current frame prediction position. Use it as a hard sample and increase its weight. However, in the case where the bimodal phase difference is not too large, what is considered to be interference is likely to be the real object. Updating at this time may cause the model to be contaminated.

Table 1 Compares $R_{distance}$ and $R_{disturb}$ of different update strategies in OTB100 dataset.

	$R_{distance}$	$R_{disturb}$
PrDiMP	0.321	0.219
Ours	0.309	0.063

Table 2 Compares $R_{distance}$ and $R_{disturb}$ of different update strategies in NFS dataset.

	$R_{distance}$	$R_{disturb}$
PrDiMP	0.655	0.465
Ours	0.529	0.221

To verify our analysis, we define proportion of error update $R_{distance}$ and $R_{disturb}$ respectively,

$$\begin{aligned} R_{distance} &= \frac{n_{distance_error}}{n_{distance}} \\ R_{disturb} &= \frac{n_{disturb_error}}{n_{disturb}} \end{aligned} \quad (3)$$

Here, $n_{distance}$ represent the total number of frames in which the current frame is used as a 'Hard negative' in the case of $SMR > threshold_2$, and $n_{distance_error}$ represent the number of frames in which the current predicted position and the actual object tracking distance are greater than 20 pixels. $n_{disturb}$ represent the total number of frames in which the current frame is used as a 'Hard negative' in the case of $SMR > threshold_4$ and $MAX^{sl} > threshold_1$. $n_{disturb_error}$ represent the number of frames in which the current predicted position and the actual object tracking distance are greater than 20 pixels. The results are shown in **Table 1** and **Table 2**. As shown in **Table 1** and **Table 2**, both $R_{distance}$ and $R_{disturb}$ of PrDiMP are higher. It shows that the proportion of error update is high, and the model is easy to be contaminated, which leads to tracking drift. Our methods $R_{distance}$ and $R_{disturb}$ are lower than PrDiMP, respectively. It shows that compared with PrDiMP, our method has lower proportion of error update, so our method is more robust than PrDiMP.

3.2 Robust model update strategy

Aiming at the problems of PrDiMP model update strategy, we integrate maximum response, multi-peaks and detector cues to guide the update of the tracker. The overall framework of the algorithm is shown in **Fig. 1**.

Maximum response cue. The goal of classification network is to distinguish the object from the surrounding background. MAX is defined as the maximum value of the classification network response graph R_t .

Multi-peaks cue. The MAX may be interfered by similar objects or certain noise leading to inaccurate detection. The inaccurate detection would further contaminate the model due to incorrect training samples. The peaks located at similar objects or background noise in the response map may approach, or even surpass the peak at the object. As above analysis, the object may locate at one of multiple peaks, all of them should be taken into consideration. The ratio of these peaks to the maximum peak PMR_i (The subscript i represent the the i -th peak) is calculated,

Fig. 1 The overall workflow of the proposed method in this paper.

$$PMR_i = \frac{s_t^i}{MAX} \quad (4)$$

Here, s_t^i represent the peak value of the i -th peak.

Detector cue. These peaks are verified by the detector to determine the object location. Specifically, we use the SiamBAN tracker [24] as a detector and select the peak closest to the predicted position of the detector as the object position.

Based on maximum response, multi-peaks and detector cues, our method divides the tracking results into three types:

'Not found': The training samples is not update; the object size and position are not update;

'Uncertain': The training samples is not update; the object size and position are update;

'Hard negative': Add the current result to the training samples and update the model immediately; the object size and position are update.

The detail strategy of model update is shown in **Algorithm 2**.

4 Results and Discussion

Extensive experiments and comparisons with state-of-the-art trackers on the four benchmarks of OTB100, NFS, VOT2018 and VOT2019.

4.1 Implementation details

Our algorithm is implemented in Python with PyTorch and run on an RTX 3090 GPU. In order to make a fair comparison, we use the same parameters as in PrDiMP [19]. ATOM [17], DiMP [18], PrDiMP [19] and our method were run 5 times in OTB100, NFS, and 15 times on VOT2018 and VOT2019.

Algorithm 2

```

Input: Initial object position  $L_1$ ; frame  $F$ ;
Output: Object position of each frame  $L_t$ ;
Initialize the tracker network and the detector network;
While current frame is valid do
    Run tracker network and return the response map  $R_t$ , maximum response  $MAX$ ;
    Run detector network and return the response map  $R_{td}$ , maximum response  $MAX_d$ ;
    Calculate the distance  $D_{ci}$  between the  $PMR_i$  prediction position and the prediction position of the detector network;
    Calculate the distance  $D_c$  between the  $MAX$  prediction position and the prediction position of the detector network;
    Calculate the average prediction object scale of previous frame  $S$ ;
    If  $MAX < threshold_1$ 
        The tracking result is 'Not found';
    Else If  $MAX(PMR_i) > threshold_2$ 
        If  $D_c < S$  and  $Min(D_{ci}) > S$ 
             $MAX$  prediction position as the prediction position of the current frame;
            The tracking result is 'Hard negative';
        Else If  $D_c > S$  and  $Min(D_{ci}) < S$ 
             $Min(D_{ci})$  prediction position as the prediction position of the current frame;
            The tracking result is 'Hard negative';
        Else
             $MAX$  prediction position as the prediction position of the current frame;
            The tracking result is 'Uncertain';
        End
    Else If  $D_c < S$ 
         $MAX$  prediction position as the prediction position of the current frame;
        The tracking result is 'Hard negative';
    Else
         $MAX$  prediction position as the prediction position of the current frame;
        The tracking result is 'Uncertain';
End

```

Table 3 Comparison of different model update strategies on the combined OTB100, NFS datasets.

	AUC
No update	0.652
Model averaging	0.653
PrDiMP	0.669
Detector	0.670
Ours	0.683

4.2 Model update strategy analysis

We analyze the impact of model online update and compare different model update strategies. I) No update: The model is not updated during tracking. II) Model averaging: In each frame, the model is updated using the linear combination of the current and newly predicted model. III) PrDiMP: See [19]. IV) Detector: Use SiamBAN [24] as detector to guide model update. V) Ours: See 3.2. The results are shown in **Table 3**. Compared with other methods, the AUC of our method has been significantly improved. These results indicate that our method can effectively adapt the object model online.

Fig. 2 Comparison with state-of-the-art trackers on the OTB100 dataset.

Table 4 Comparison with state-of-the-art trackers on the NFS dataset.

	Success
ATOM	0.584
C-COT	0.488
UPDT	0.537
SiamBAN	0.594
ECO	0.466
Siam R-CNN	0.639
DiMP	0.620
PrDiMP	0.635
Ours	0.651

4.3 State-of-the-art comparison

OTB100 [20]: OTB100 is a widely used public tracking benchmark consisting of 100 sequences. **Fig. 2** compares our results to eight state-of-the-art (SOTA) trackers: PrDiMP [19], DiMP [18], ATOM [17], Siam R-CNN [26], Staple [35], SiamBAN [24], SiamCAR [25], SiamRPN++ [15]. The results show that, our approach not only achieves better results compared with the corresponding baseline method, but also better than other excellent object tracking methods.

NFS [21]: NFS has two versions: 240FPS version rate and 30FPS version, and we evaluate our tracker in the 30FPS version of the dataset, **Table 4** compares our results to eight state-of-the-art (SOTA) trackers: PrDiMP [19], DiMP [18], ATOM [17], C-COT [11], UPDT [32], SiamBAN [24], Siam R-CNN [26], ECO [40]. The results show that, our approach not only achieves better results compared with the corresponding baseline method, but also better than other excellent object tracking methods.

VOT2018 [22]: **Table 5** lists results obtained by PrDiMP [19], DiMP [18], ATOM [17], STMTrack [37], SiamBAN [24], DasiamRPN [39], SiamRPN++ [15], LADCF [33]. The results show that, our approach not only achieves better results compared with the corresponding baseline method, but also better than other excellent object tracking methods.

VOT2019 [23]: **Table 6** lists results obtained by PrDiMP [19], DiMP [18], LightTrack [41], ResPUL [36], SiamDW-ST [34], SiamBAN [24], SiamRN [38], ARTCS [23]. The

Table 5 Comparison with state-of-the-art trackers on the VOT2018 dataset.

	Accuracy	Robustness	EAO
ATOM	0.590	0.203	0.401
STMTrack	0.590	0.159	0.447
SiamRPN++	0.604	0.234	0.417
SiamBAN	0.597	0.178	0.452
DasiamRPN	0.586	0.276	0.383
LADCF	0.503	0.159	0.389
DiMP	0.597	0.153	0.440
PrDiMP	0.618	0.165	0.442
Ours	0.619	0.144	0.465

Table 6 Comparison with state-of-the-art trackers on the VOT2019 dataset.

	Accuracy	Robustness	EAO
LightTrack	0.552	0.310	0.357
ResPUL	0.515	0.828	0.198
SiamRN	0.593	0.306	0.341
SiamBAN	0.602	0.396	0.327
SiamDW-ST	0.600	0.467	0.299
ARTCS	0.602	0.482	0.287
DiMP	0.582	0.371	0.321
PrDiMP	0.614	0.323	0.340
Ours	0.615	0.296	0.353

results show that, our approach not only achieves better results compared with the corresponding baseline method, but also better than other excellent object tracking methods.

5 Conclusions

We improved PrDiMP, and integrate maximum response, multi-peaks and detector cues to guide model update of PrDiMP. This method greatly reduces the risk of online model update. We comparisons with state-of-the art trackers on four benchmarks: OTB100, NFS, VOT2018 and VOT2019. The results show that, our approach not only achieves better results compared with the corresponding baseline method, but also better than other excellent object tracking methods. This not only significantly improves the tracking performance of PrDiMP, but also can be easy to be embedded into other online model update trackers.

List of abbreviations

PrDiMP: probabilistic discriminative model prediction; MAX: maximum response; SMR: second-maximum response ratio.

Competing interests

The authors declare that they have no competing interests.

Funding

This work is not supported by funding.

Authors' contributions

Shaolong Chen, Changzhen Qiu and Yurong Huang conceived the method and developed the algorithm. Zhiyong Zhang oversaw the project. Shaolong Chen assembled formulations and drafted the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

Data and source code are available from the corresponding author upon request.

Acknowledgements

Not applicable.

Ethics approval and consent to participate

Not applicable.

References

1. Sun D Q, Roth S, Black M J. Secrets of optical flow estimation and their principles. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
2. Nummiaro K, Koller-Meier E, Van Gool L. An adaptive color-based particle filter. *Image and Vision Computing*, 2003, 21(1): 99-110.
3. Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5): 603-619.
4. Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
5. Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels. *European Conference on Computer Vision (ECCV)*, 2012.
6. Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596.
7. Zhu H Y, Wang H. Robust visual tracking via samples ranking. *EURASIP Journal on Advances in Signal Processing*, 2019, 1: 41.
8. Junos M H, Khairuddin, A S M, Thannirmalai S, et al. An optimized YOLO-based object detection model for crop harvesting system. *IET Image Processing*, 2021, Early Access.
9. Chen S L, Qiu C Z, Zhang Z Y. An efficient method for tracking failure detection using parallel correlation filtering and Siamese network. *Applied Intelligence*, 2021, Early Access.
10. Danelljan M, Hager G, Khan F S, et al. Convolutional features for correlation filter based visual tracking. *IEEE International Conference on Computer Vision (ICCV)*, 2016.
11. Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: Learning continuous convolution operators for visual tracking. *European Conference on Computer Vision (ECCV)*, 2016.
12. Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking. *European Conference on Computer Vision (ECCV)*, 2016.
13. Tao R, Gavves E, Smeulders A W M. Siamese instance search for tracking. *IEEE International Conference on Computer Vision (ICCV)*, 2016.

14. Li B, Yan J, Wu W, et al. High performance visual tracking with siamese region proposal network. *IEEE International Conference on Computer Vision (ICCV)*, 2018.
15. Li B, Wu W, Wang Q, et al. SiamRPN plus plus: Evolution of siamese visual tracking with very deep networks. *IEEE International Conference on Computer Vision (ICCV)*, 2019.
16. Wang Q, Zhang L, Bertinetto L, et al. Fast online object tracking and segmentation: A unifying approach. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
17. Danelljan M, Bhat G, Khan F S, et al. ATOM: Accurate tracking by overlap maximization. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
18. Bhat G, Danelljan M, Van Gool L, et al. Learning discriminative model prediction for tracking. *IEEE International Conference on Computer Vision (ICCV)*, 2019.
19. Danelljan M, Van Gool L, Timofte R. Probabilistic regression for visual tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
20. Wu Y, Lim J, Yang M H. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1834-1848.
21. Galoogahi H K, Fagg A, Huang C, et al. Need for speed: A benchmark for higher frame rate object tracking. *IEEE International Conference on Computer Vision (ICCV)*, 2017.
22. Kristan M, Leonardis A, Matas J, et al. The sixth visual object tracking VOT2018 challenge results. *European Conference on Computer Vision (ECCV)*, 2018.
23. Kristan M, Matas J, Leonardis A, et al. The seventh visual object tracking VOT2019 challenge results. *IEEE International Conference on Computer Vision (CVPR)*, 2019.
24. Chen Z D, Zhong B N, Li G R, et al. Siamese box adaptive network for visual tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
25. Guo D Y, Wang J, Cui Y, et al. SiamCAR: Siamese fully convolutional classification and regression for visual tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
26. Voigtlaender P, Luiten J, Torr P H S, et al. Siam R-CNN: Visual tracking by re-detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
27. Hyeonseob N, Bohyung H. Learning multi-domain convolutional neural networks for visual tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
28. Wang M M, Liu Y, Huang Z Y. Large margin object tracking with circulant feature maps. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
29. Lukezic A, Zajc L C, Vojir T, et al. FuCoLoT - A fully-correlational long-term tracker. *Asian Conference on Computer Vision (ACCV)*, 2018.
30. Gao P, Zhang Q Q, Wang F, et al. Learning reinforced attentional representation for end-to-end visual tracking. *Information Sciences*, 2020, 517: pp 52-67.
31. Gao P, Yuan R Y, Wang F, et al. Siamese attentional keypoint network for high performance visual tracking. *Knowledge-based Systems*, 2020, 193: Article Number: 105448.
32. Bhat G, Johnander J, Danelljan M, et al. Unveiling the power of deep tracking. *European Conference on Computer Vision (ECCV)*, 2018.
33. Xu T Y, Feng Z H, Wu X J, et al. Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking. *IEEE Transactions on Image Processing*, 2019, 28(11): 5596-5609.
34. Zhang Z P, Peng H W. Deeper and wider siamese networks for real-time visual tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
35. Bertinetto L, Valmadre J, Golodetz S, et al. Staple: Complementary learners for real-time tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
36. Wu Q Q, Wan J, Chan A B. Progressive unsupervised learning for visual object tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
37. Fu Z H, Liu Q J, Fu Z H, et al. STMTrack: Template-free visual tracking with space-time memory networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
38. Cheng S Y, Zhong B N, Li G R, et al. Learning to filter: Siamese relation network for robust tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
39. Zhu Z, Wang Q, Li B, et al. Distractor-aware siamese networks for visual object tracking. *European Conference on Computer Vision (ECCV)*, 2018.
40. Danelljan M, Bhat G, Khan F S, et al. ECO: Efficient convolution operators for tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
41. Yan B, Peng H W, Wu K, et al. LightTrack: Finding lightweight neural networks for object tracking via one-shot architecture search. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.