



# Learning Attention Regularization Correlation Filter for Visual Tracking

Zhuling Qiu, Yufei Zha<sup>(✉)</sup>, Peng Zhu, and Fei Zhang

Aeronautics Engineering College, Air Force Engineering University, Xian, China  
zhayufei@126.com

**Abstract.** Spatial regularization can effectively solve the unwanted boundary effect of discriminative correlation filters (DCF). However, the predefined mask is independent of the feature, which limits the performance improvement. In this paper, we take the mask as a variable that plays the same role as the filter, and an attention regularization correlation filter (ARCF) is proposed for visual tracking. Especially, the mask is no longer a binary but a real value between 0 and 1, used as the weight of the corresponding feature. Additionally, the temporal coherence is also considered when the filter and the mask are simultaneously optimizing via ADMM algorithm, so the filter can fit the variation of the target in the temporal domain. Extensive experiments on the OTB100 database prove that our algorithm is much better than the traditional SRDCF algorithm both in the performance and speed.

**Keywords:** Object tracking · Regularization · Correlation filters · Attention

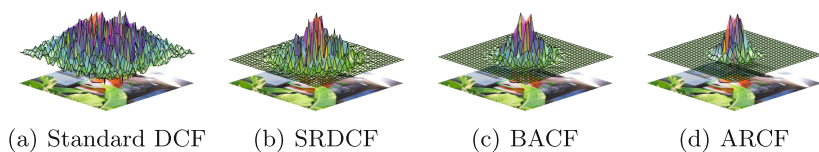
## 1 Introduction

Visual tracking is an important task in many computer vision topics. One of the main challenges of this task is to address the target's appearance changes over time. Recent years, discriminative correlation filters (DCF) [8] have shown state-of-the-art performance in the fashion tracking data set [17] and competitions [11]. The advantages of DCF [8] benefit from the periodic assumption of training samples. However, such an assumption leads to unwanted boundary effects since the examples including many unrealistic, wrapped-around circularly shifted versions of the target due to the circularity. Thus, the discriminative power of the learned filter shown in Fig. 1(a) is limited, so that the tracking performance is difficult to further improve.

The above problem was addressed in the recent works [6, 9, 12]. Danneljan *et al.* [6] introduced predefined Inverse Gaussian distribution matrix as a spatial regularization to penalize filter values outside the target boundaries, which

---

Supported by the National Natural Science Foundation of China (No. 61773397) and the Fundamental Research Funds for the Central Universities (No. 3102019ZY1003).



**Fig. 1.** Spatial Regularizations. Figure 1(a) shows filter of the standard discriminative correlation filters (DCF) [8], the filter regularized by the inverse Gaussian distribution matrix [6] and zero-padding mask [9] are shown in 1(b) and 1(c), respectively. The filter constrained by the proposed attention regularization is shown in 1(d).

is shown in Fig. 1(b). Different from the solution that was implemented by the Gauss-Seidel method with high computational complexity, STRCF tracker [12] trains the filter on the single sample via the alternating direction method of multipliers (ADMM) algorithm [3]. Galoogahi *et al.* [9] proposed zero-padding the filter shown in Fig. 1(c) to eliminate the background during training, then the optimization is also performed by the ADMM [3]. The ideas behind these methods are to design a predefined mask to overcome the boundary effects, however, there are some drawbacks: (1) The predefined regular shape of the mask can fit the appearance of the target (2) The value of the mask is binary that indicates whether this pixel is a target or not (3) The temporal coherence of the mask is not considered anymore.

To overcome these problems, an attention regularization correlation filter (ARCF) is proposed for visual tracking in this study. A spatial attention mask is learned with the filter and utilized to indicate the corresponding importance of each position in the filter. Unlike the existing methods that treat the mask as a hyper-parameter, we take the mask as a variable that plays the same role as the filter, then they are simultaneously optimized via ADMM algorithm. Here, the mask is no longer a binary but a real value between 0 and 1, used as the weight of the corresponding feature. Therefore, the position corresponding to the large weight forms the spatial attention of the image. Additionally, the temporal coherence is also considered when the filter and the mask are optimizing, so the filter can fit the variation of the target in the temporal domain. Figure 1(d) shows the learned filter by our spatial attention map. It can be seen that the discriminative ability of the features is enhanced by our method compared with the other methods. The contributions of this paper are summarized as follows:

- We propose a visual attention mechanism to regularize the correlation filter both in the spatial and temporal domain.
- The value of the spatial attention mask is released to  $[0, 1]$  replaced binary values  $\{0, 1\}$  to indicate the weights of the corresponding features.
- We propose to constrain the temporal coherence of the learned mask to adapt to the variation of the target in the temporal domain.

## 2 Related Work

### 2.1 Spatial Regularization

Unwanted boundary effects in correlation filter based tracking lead to inaccurate representation and insufficient discrimination of targets, especially in the cluttering background. Some works [6, 12] wanted to solve this limitation by investigating the scale relationship between the training samples and filters. That is to say, the filter coefficients are penalized in terms of spatial locations [6] or temporal rank [12] to achieve more robust appearance modeling suitable for large variations. But the introduced noise of background became inevitable [10].

Different from those methods that perform regularization and filtering in a separated process with auxiliary features, our method is only required the features for visual tracking and simultaneously optimized the filter and spatial map. This is the motivation of this study.

### 2.2 Visual Attention

The visual attention derived from the cognitive neuroscience has been applied to some computer vision tasks, such as image classification [15] and image caption. It is so popular because the attention mechanism gives the model the discriminative ability between objects. The spatial weights, such as the cosine window map [2] and the Gaussian window map [8], are used as an attentional mechanism to be integrated into the correlation filter for visual tracking tasks.

However, these approaches emphasized attentive features and resort to additional attention modules to generate feature weights. In contrast to that, our method is self-attention, which exploited the attention map as a regularization term coupled with the standard correlation filter. And the attention map and the filter can be optimized simultaneously by the ADMM [3] algorithm for robust trackers.

## 3 Method

### 3.1 Learning Attention Regularization

Recently, the correlation filter received much attention with its ability to use circular matrix for dense sampling. But, the unwanted boundary effects derived from the periodic assumption of training samples limits the performance improvement further. To address this problem, an inverse Gaussian distribution matrix [6] is as a spatial regularization to penalize filter values outside the target boundaries. The spatial regularization correlation filter is rewritten with  $T$  training samples as:

$$\mathcal{L}(\Phi) = \sum_{k=1}^T \epsilon_k \left\| \sum_{l=1}^D \Phi^l * x_k^l - y_k \right\|^2 + \sum_{l=1}^D \|w \odot \Phi\|^2, \quad (1)$$

where  $\Phi \in \mathbb{R}^D$  denotes the filter, the symbol  $*$  represents correlation operation,  $y$  is the regression values of the feature  $x \in \mathbb{R}^D$  and  $\epsilon_k$  is the regularization term of the  $k$ th sample  $x_k$ . The size of feature  $x$ , filter  $\Phi$  and regression  $y$  is  $M \times N$ .  $w$  is the spatial regularization matrix, which is the weight of the location in the filter  $\Phi$ .

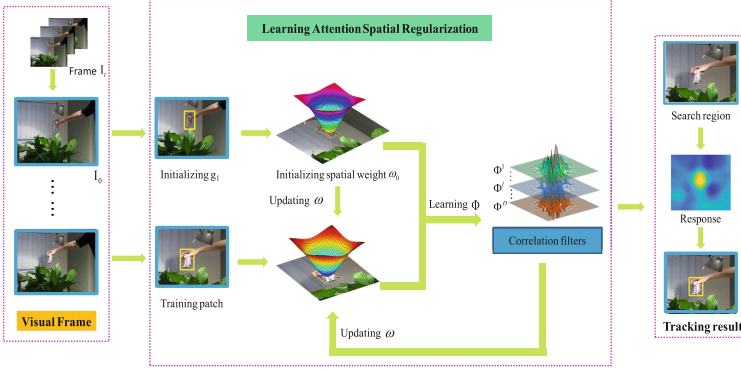
In this study, we introduce a attention mechanism, which makes the filter pay more attention to the target and the desired response lower at the background. Additionally, the temporal coherence is also considered constraining the regularization term  $w$  learning. We learn the spatial attention map correlation filter with the loss function:

$$\mathcal{L}(\Phi, w) = \underbrace{\frac{1}{2} \left\| \sum_{l=1}^D \Phi^l * x^l - y \right\|^2}_{\text{Regression Term}} + \underbrace{\frac{1}{2} \sum_{l=1}^D \|w \odot \Phi^l\|^2}_{\text{Spatial Term}} + \underbrace{\frac{\mu}{2} \|w - w_0\|^2}_{\text{Temporal Term}}, \quad (2)$$

where  $\mu$  is temporal regularization coefficients, respectively. Unlike the existing works, the  $w$  is a variable to learn, not a hyper-parameter. Here,  $w_0$  is an initial prior distribution which is predefined as an invert Gaussian distribution similar to the work [6].

The aim of minimizing the loss of Eq. (2) is to learn the filter  $\Phi$  and the attention map  $w$  simultaneously. The first term is the regression term to learn the filter  $\Phi$  with the feature  $x$  and the expected response  $y$ , which is same as the standard correlation filter. The spatial and temporal regularization terms are shown in the second term and third term to learn the attention regularization. According to the importance of the spatial position to learn the attention map  $w$ , the feature of the target are attached with the smaller spatial weights, and the background feature gives a bigger spatial constraint weight, which makes the learned filter more discriminative than that learned by the fixed spatial regularization. This can enhance the distinction between goals and background. Additionally, in order to deal with the variation of the target, the attention regularization is also constrained in the temporal domain which is represented in the third term of the Eq. (2). Temporal regularization terms make the filter change not too severe in the case of target occlusion, which can guarantee the performance of tracking.

According to the above theory, the flow chart of the algorithm is as shown in Fig. 2. By using the first frame  $I_0$  information, the target frame is initialized and the spatial constraint weight  $w_0$  in the first frame is assigned to the inverse Gaussian distribution, and the training is performed to obtain the filter  $\Phi$ . The target position is predicted in the next frame by using the trained filter. At the same time, using the information of target position in the current frame can update the filter, and the weight map  $w$  is updated in the time domain and the frequency domain according to the position feature weight map in the current frame and the initial frame  $w_0$ , which can achieve more robust tracking.



**Fig. 2.** Pipeline of learning attention regularization for correlation filter tracking.  $g_1$  is the object bounding box in the first frame and  $w_0$  is the spatial weight in the first frame.  $w$  is updated according to the spatial attention map in each subsequent frame during the learning process.

### 3.2 Model Optimization

In this subsection, we will introduce how to optimize the loss function Eq. (2), which is convex, and the optimal solution can be solved by iterative the alternating direction method of multipliers (ADMM) algorithm [3]. Therefore, through introducing the constraint condition  $\Phi = \Theta$ , Lagrangian equation of the Eq. (2) can be rewritten as:

$$\begin{aligned} \mathcal{L}(\Phi, w, \Theta, \beta) = & \frac{1}{2} \left\| \sum_{l=1}^D \Phi^l * x^l - y \right\|^2 + \frac{1}{2} \sum_{l=1}^D \|w \odot \Theta^l\|^2 \\ & + \frac{\mu}{2} \|w - w_0\|^2 + \beta^T \sum_{l=1}^D (\Phi^l - \Theta^l) + \frac{\alpha}{2} \sum_{l=1}^D \|\Phi^l - \Theta^l\|^2, \end{aligned} \tag{3}$$

where  $\beta$  is the Lagrange multiplier and  $\alpha$  is the penalty parameter.

When  $\delta = \frac{\beta}{\alpha}$ , the augmented Lagrangian equation can be written as:

$$\begin{aligned} \mathcal{L}(\Phi, w, \Theta, \delta) = & \frac{1}{2} \left\| \sum_{l=1}^D \Phi^l * x^l - y \right\|^2 + \frac{1}{2} \sum_{l=1}^D \|w \odot \Theta^l\|^2 \\ & + \frac{\mu}{2} \|w - w_0\|^2 + \frac{\alpha}{2} \sum_{l=1}^D \|\Phi^l - \Theta^l + \delta^l\|^2. \end{aligned} \tag{4}$$

The ADMM algorithm is used to solve the following subproblems:

$$\begin{cases} \Phi^{(i+1)} = \arg \min_{\Phi} \left\| \sum_{l=1}^D \Phi^l * x^l - y \right\|^2 + \alpha \|\Phi - \Theta + \delta\|^2, \\ \Theta^{(i+1)} = \arg \min_{\Theta} \sum_{l=1}^D \|w \odot \Theta^l\|^2 + \alpha \|\Phi - \Theta + \delta\|^2, \\ \delta^{(i+1)} = \delta^{(i)} + \Phi^{(i+1)} - \Theta^{(i+1)}. \end{cases} \tag{5}$$

**Subproblem  $\Phi$ .** According to the iterative equation of ADMM algorithm, the solution of subproblem  $\Phi$  can be converted to Fourier domain for solving,

$$\mathcal{L}(\hat{\Phi}) = \left\| \sum_{l=1}^D \hat{\Phi}^l \odot \hat{x}^l - \hat{y} \right\|^2 + \alpha \left\| \hat{\Phi} - \hat{\Theta} + \hat{\delta} \right\|^2, \quad (6)$$

where  $\hat{\Phi}$  is the discrete Fourier transform of the filter  $\Phi$ . By taking the derivative of  $\hat{\Phi}$  be zero, the equation can be obtained as follows:

$$2(\hat{\Phi} \odot \hat{x} - \hat{y}) \odot \hat{x}^T + 2\alpha(\hat{\Phi} - \hat{\Theta} + \delta) = 0. \quad (7)$$

So, we have a closed-form solution of  $\hat{\Phi}$ :

$$\hat{\Phi} = \frac{\alpha \hat{\Theta} - \alpha \hat{\delta} + \hat{y} \odot \hat{x}^T}{\hat{x} \odot \hat{x}^T + \alpha I}. \quad (8)$$

**Subproblem  $\Theta$ .** For the solution of subproblem  $\Theta$ , we can take the derivative of  $\Theta$  be zero in the time domain directly,

$$2w^T \odot w\Theta + 2\alpha(\Phi - \Theta + \delta) = 0. \quad (9)$$

And we can also get a closed-form solution for  $\Theta$ :

$$\Theta = \frac{\alpha\Phi + \alpha\delta}{w^T \odot w + \alpha}. \quad (10)$$

**Subproblem  $w$ .** For updating the spatial weight  $w$  temporally, we can utilize Eq. (4) to take the derivative of  $w$  directly,

$$\frac{\partial \mathcal{L}(\Phi, \Theta, \delta)}{\partial w} = w \odot \sum_{l=1}^D (\Theta^l)^2 + \mu(w - w_0). \quad (11)$$

By solving  $\frac{\partial \mathcal{L}(\Phi, \Theta, \delta)}{\partial w} = 0$  we get the closed-form solution

$$w = \frac{\mu w_0}{Q + \mu}, \quad (12)$$

where  $Q$  is  $\sum_{l=1}^D (\Theta^l)^2$ . By Eq. (12), we can update  $w$  which includes information about the target in the current frame.

**Updating Penalty Parameter  $\alpha$ .** The stepsize parameter  $\alpha$  is updated as:

$$\alpha^{(i+1)} = \min(\alpha^{max}, \rho\alpha^{(i)}), \quad (13)$$

where  $\alpha^{max}$  is the maximum value of  $\alpha$  and the scale factor  $\rho$ .

### 3.3 Object Tracking

---

**Algorithm 1.** Learning Attention Regularization Correlation Filter for Visual Tracking

---

**Input:**

1: Frames:  $I_t, t = 0, 1, 2, \dots$ ; initial object bounding box:  $g_1 = (x_1, y_1, w_1, h_1)$ ;

**Output:**

2: Prediction :  $r_{t-1} = (x_{t-1}, y_{t-1}, w_{t-1}, h_{t-1})$  ;

**3: Initialization:**

4: initialize the correlation filters, initialize the spatial weight map:  $w_0$  is an invert Gaussian distribution;

5: Learn  $\Phi$  by minimizing Eq. (2), update  $w_1$  by the solution Eq. (12) via the first frame with given bounding box,  $t = 2$  ;

**6: while  $t \geq 2$  do**

7: Crop an image patch  $R_t$  from  $I_t$  at the last bounding  $g_t$  and extract its feature map  $x_t$ ;

8: Detect the object location by calculating the response via  $x_t$  and  $\Phi$  and the estimate the scale of the target, thus get  $g_t$ ;

9: Update  $\Phi_t$  by Eq. (2) using iterative ADMM algorithm via  $x_t$  and  $w_{t-1}$ ;

10: Learn  $w_t$  by the closed-form solution Eq. (12) ;

11:  $t = t + 1$ ;

12: **return**  $r_t$

---

In this subsection, we describe the proposed tracking framework based on learning attention regularization. The overview of the proposed tracker is shown in Algorithm 1.

We use the information of the first frame to initialize the target frame and filter. The spatial regularization weight  $w_0$  in the first frame is assigned to the inverse Gaussian distribution. During the tracking process, the filters obtained by training in the previous frame are used to detect the position of the target in the search area of the next frame. After determining the target position, the training region centered on the target position of the current frame is extracted to update filter model. According to the spatial attention map, the spatial constraint weight  $w$  is adjusted.

## 4 Experiments

In this section, we present comprehensive experimental evaluations of the proposed algorithm using OTB100 [17] data set. First, we describe the implementation details and the evaluation protocols. Next, we demonstrate the effectiveness of each component in the proposed tracker in the form of experiment. Finally, the algorithm proposed in this paper is compared with other representative algorithms to obtain comprehensive experimental results.

## 4.1 Implementation Details

**Tracker Parameters.** Our filter is based on a regularized filter, but the proposed algorithm has a certain change in the parameter setting because the filter weight parameter is no longer a super parameter, but a real number from 0 to 1. Through many experiments, we set the hyperparameter in Eq. (2) to  $\mu = 18$ . Initial constraint parameters  $\alpha^{(0)}$ , maximum constraint parameters  $\alpha^{max}$ , and scale factor  $\rho$  are set to 10, 100 and 1.2.

**Evaluation Protocols.** In this paper, the algorithm is evaluated by the success rate and precision rate curve. The AUC is area under the curve for success rate. The DP is the value in the precision rate curve when the threshold is 20. Based on the benchmark library settings, we compare the proposed tracker with the state-of-the-art trackers using one-pass evaluation (OPE) (each tracker evaluates in the initial frame with the ground-truth box until the end of each sequence).

## 4.2 Overall Performance

The table below shows the algorithm presented in this paper performs significantly better than most of the competing trackers that use different tracking methods.

**Table 1.** The algorithm of this paper is compared with the regularization-based algorithm on the OTB100 [17] data set. The AUC, DP and the operation speed are used as evaluation criteria.

Tracker	AUC (%)	DP (%)	Speed (FPS)
OURS	<b>65.7</b>	<b>87.0</b>	27
SRDCF [6]	59.8	78.9	5
STRCF [12]	61.4	86.3	<b>36</b>
CSR-DCF [13]	59.8	73.3	15
BACF [9]	63.0	81.6	35

**Comparison with the Trackers Based on Spatial Regularization.** We evaluated the proposals for the four recently released trackers: STRCF [12], CSR-DCF [13], BACF [9], SRDCF [6]. The Table 1 and Fig. 3 shows that the tracker proposed in this paper achieves excellent results under two test criteria. As the benchmark algorithm SRDCF [6] uses Gauss-Seidel iterative method in the algorithm operation, its tracking speed is slower. Meanwhile, because the temporal regularization isn't introduced to SRDCF [6], its performance is poor when facing videos such as occlusion. Therefore, the proposed algorithm has a larger improvement compared with the benchmark algorithm SRDCF [6]. And we can see the success plot has increased by about 8%, and the precision plot has increased by about 7%.

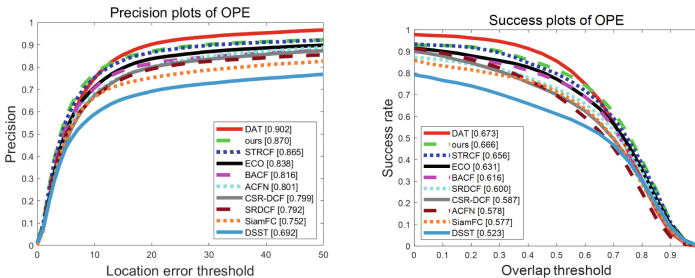


**Table 2.** The algorithm of this paper is compared with the algorithm based on neural network attention mechanism on the OTB100 [17] data set. The AUC, DP and the operation speed are used as evaluation criteria.

Tracker	AUC (%)	DP (%)	Speed (FPS)
OURS	65.7	87.0	27
DAT [14]	<b>66.3</b>	<b>90.2</b>	1
ACFN [4]	57.5	80.2	15
RASNet [16]	64.2	—	<b>83</b>

### Comparison with the Trackers Based on Neural Network Attention Mechanisms.

We evaluate the trackers proposed in this paper compared with state-of-the-art neural network attention-based trackers, including DAT [14], RASNet [16], and ACFN [4]. The algorithm improves the tracking effect by using a more flexible filter weight coefficient, which can improve filter response to the target and reduce background interference to the target. As are shown in Table 2 and Fig. 3, the algorithm performs better than RASNet [16] and ACFN [4]. When compared with DAT [14], although the proposed algorithm is different from DAT [14] by about 3% in performance, it has an obvious performance in terms of tracking speed as this paper uses ADMM [3] (Alternating Direction Method of Multipliers) iterative algorithm. The improvement of the algorithm proposed in this paper is 27 times faster than DAT [14].



**Fig. 3.** The success plot and precision plot on the OTB100 [17] data set are quantitatively evaluated by the OPE method. The legend is the AUC and DP scores for each algorithm.

**Compare with the Most Advanced and Classic Algorithms.** SiamFC [1] and ECO [5] are currently advanced trackers, which uses different ways. SiamFC [1] classify the target using the method of joining the Alexnet network to improve the extraction accuracy of the target feature. However, due to the classification nature of the network, the problem of similar background interference cannot be solved, which makes tracking effect worse. The proposed algorithm solves the

problem of similar interference by introducing temporal regularization, so it is far ahead of SiamFC [1] in performance. ECO [5] is due to the sparse update strategy, which makes the calculation process complicated and the operation speed slow down. The algorithm can be slightly weaker than the ECO [5], but the tracking speed is more than 4 times that of ECO [5]. DSST [7] is a relatively classic algorithm proposed in 2014. It adopts the method of feature fusion, which enables the algorithm to have a better adaptive process for scale variation of the target. As shown in Fig. 3, the proposed algorithm performs far better than DSST [7].

### 4.3 Ablation Study

The core idea of this paper mainly includes the real valued (learning attention) between the filter weight coefficient from fixed super-parameters to variable 0 to 1, and a regularization method in the time domain. In order to prove that each component improves the performance of the algorithm, an assessment of each part of the algorithm will be performed.

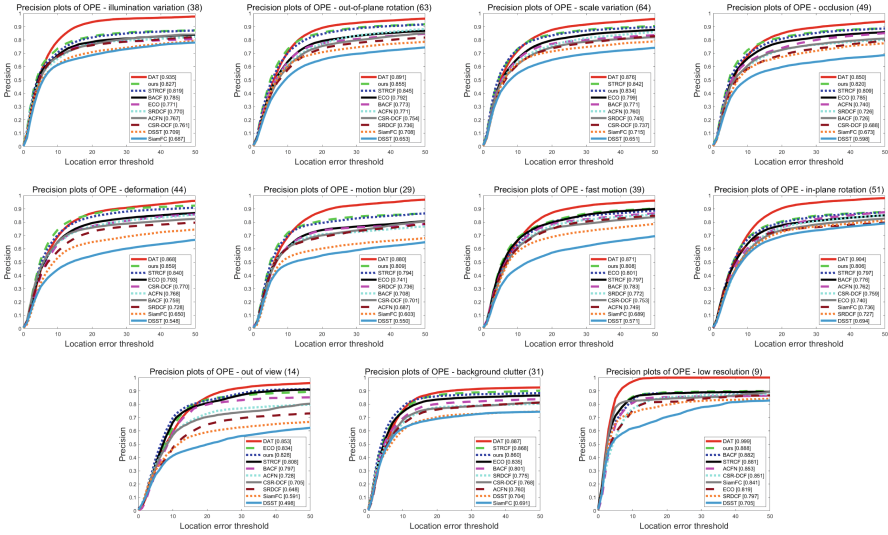
Regularization in the time domain can effectively solve the problem of occlusion of the target. We will remove the algorithm of time domain regularization with ARCFp. The change of the filter weight parameter can make the response value of the target larger and reduce the background interference. We will remove the learning attention algorithm by ARCFq. As is shown in Table 3, the results are compared.

**Table 3.** In this evaluation, the OTB100 [17] data set is used as the test sequence. Meanwhile, DP and AUC are used as the evaluation criteria, where DP is the value with a threshold of 20 and AUC is the area under the curve of success plot curve.

	ARCFp	ARCDq	ARCF
AUC (%)	61.5	47.7	65.7
DP (%)	82.6	69.3	87.0

### 4.4 Qualitative Analysis

We analyze the tracker performance using 11 annotation attributes in the OTB100 [17] data set: illumination variation, out-of-plane rotation, scale variation, occlusion, deformation, motion blur, etc. Figure 4 shows the results of a one-pass evaluation of these challenging attributes for visual object tracking. From the results, the tracker proposed in this paper in the illumination variation, out-of-plane rotation, scale change, occlusion, deformation, motion blur, fast motion, in-plane rotation, background clutter and low resolution can perform well and score at the top. Due to the fixed weight coefficient of the filter, other algorithms using similar methods have problems in poor ability of discriminating target and background and uneven mask distribution, resulting in poor



**Fig. 4.** The precision plot curve under each difficulty attribute, where the value after the curve is the value when the threshold is 20. (This evaluation method is the current mainstream qualitative analysis method)

overall tracking performance. However, the filter weight coefficient of the proposed algorithm is no longer a fixed weight or an inverse Gaussian distribution, but can vary from 0 to 1 depending on the target and background, so that the filter constraint weights at the background are gradually increasing as the target response increases. This can improve the tracking effect.

## 5 Conclusion

In this paper, we proposed an attention regularization correlation filter (ARCF) for visual tracking. The mask is treated as a variable that plays the same role as the filter, then they are simultaneously optimized via ADMM algorithm. Here, the greater the weight is, the more important the corresponding feature is. Additionally, the temporal coherence is also considered when the filter and the mask are optimizing, so the filter can fit the variation of the target in the temporal domain. Extensive experiments show that our method is much better than the traditional SRDCF tracker both in the performance and speed.

In the future, we want to investigate how to generally apply the proposed method with the CNN features that are powerful ability to describe the object in the semantic domain. This is helpful to distinguish the background, even distractors.

## References

1. Bertinetto, L., Valmadre, J., Henriques, J.F., Vedaldi, A., Torr, P.H.S.: Fully-convolutional siamese networks for object tracking. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9914, pp. 850–865. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-48881-3\\_56](https://doi.org/10.1007/978-3-319-48881-3_56)
2. Bolme, D.S., Beveridge, J.R., Draper, B.A., Lui, Y.M.: Visual object tracking using adaptive correlation filters. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 2544–2550 (2010)
3. Boyd, S., Boyd, S., Vandenberghe, L., Press, C.U.: Convex Optimization. Cambridge University Press, Cambridge (2004)
4. Choi, J., Chang, H.J., Yun, S., Fischer, T., Demiris, Y., Choi, J.Y.: Attentional correlation filter network for adaptive visual tracking. In: CVPR, pp. 4828–4837. IEEE Computer Society (2017)
5. Danelljan, M., Bhat, G., Khan, F.S., Felsberg, M.: ECO: efficient convolution operators for tracking. In: Conference on Computer Vision and Pattern Recognition, CVPR, pp. 6931–6939 (2017)
6. Danelljan, M., Hager, G., Khan, F.S., Felsberg, M.: Learning spatially regularized correlation filters for visual tracking. In: International Conference on Computer Vision, (ICCV) (2015)
7. Danelljan, M., Häger, G., Khan, F.S., Felsberg, M.: Discriminative scale space tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(8), 1561–1575 (2017)
8. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 583–596 (2015)
9. Kiani Galoogahi, H., Fagg, A., Lucey, S.: Learning background-aware correlation filters for visual tracking. In: The IEEE International Conference on Computer Vision (ICCV), October 2017
10. Kiani Galoogahi, H., Sim, T., Lucey, S.: Correlation filters with limited boundaries. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015
11. Kristan, M., et al.: A novel performance evaluation methodology for single-target trackers. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(11), 2137–2155 (2016)
12. Li, F., Tian, C., Zuo, W., Zhang, L., Yang, M.H.: Learning spatial-temporal regularized correlation filters for visual tracking. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018
13. Lukezic, A., Vojár, T., Zajc, L.C., Matas, J., Kristan, M.: Discriminative correlation filter tracker with channel and spatial reliability. *Int. J. Comput. Vision* **126**(7), 671–688 (2018)
14. Pu, S., Song, Y., Ma, C., Zhang, H., Yang, M.H.: Deep attentive tracking via reciprocative learning. In: Neural Information Processing Systems (2018)
15. Wang, F., et al.: Residual attention network for image classification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017
16. Wang, Q., Teng, Z., Xing, J., Gao, J., Hu, W., Maybank, S.J.: Learning attentions: residual attentional siamese network for high performance online visual tracking. In: CVPR, pp. 4854–4863. IEEE Computer Society (2018)
17. Wu, Y., Lim, J., Yang, M.: Object tracking benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(9), 1834–1848 (2015)