



# Adaptive spatially regularized target attribute-aware background suppressed deep correlation filter for object tracking

Sathiyamoorthi Arthanari, Sathishkumar Moorthy, Jae Hoon Jeong, Young Hoon Joo \*

*School of IT Information and Control Engineering, Kunsan National University, 558 Daehak-ro, Gunsan-si, Jeonbuk 54150, Republic of Korea*



## ARTICLE INFO

**Keywords:**

Deep correlation filter  
Adaptive spatial regularization  
Background suppressed filter  
Adaptive attribute-aware  
Visual object tracking

## ABSTRACT

In recent years, deep feature-based correlation filters have attained impressive performance in robust object tracking. However, deep feature-based correlation filters are affected by undesired boundary effects, which reduce the tracking performance. Moreover, the tracker moves towards a region that is identical to the target due to the sudden variation in target appearance and complicated background areas. To overcome these issues, we propose an adaptive spatially regularized target attribute-aware background suppressed deep correlation filter (ASTABSCF). To do this, a novel adaptive spatially regularized technique is presented, which aims to learn an efficient spatial weight for a particular object and fast target appearance variations. Specifically, we present a target-aware background suppression method with dual regression approach, which utilizes a saliency detection technique to produce the target mask. In this technique, we employ the global and target features to get the dual filters known as the global and target filters. Accordingly, global and target response maps are produced by dual filters, which are integrated into the detection stage to optimize the target response. In addition, a novel adaptive attribute-aware approach is presented to emphasize channel-specific discriminative features, which implements a post-processing technique on the observed spatial patterns to reduce the influence of less prominent channels. Therefore, the learned adaptive spatial attention patterns significantly reduce the irrelevant information of multi-channel features and improve the tracker performance. Finally, we demonstrate the efficiency of the ASTABSCF approach against existing modern trackers using the OTB-2013, OTB-2015, TempleColor-128, UAV-123, LaSOT, and GOT-10K benchmark datasets.

## 1. Introduction

Visual object tracking (VOT) is one of the essential research topic for dynamic object tracking in computer vision that aims to accurately predict the target in video sequences. In recent years, VOT has gained significant research importance across various domains, including autonomous driving, video surveillance, mobile robot navigation, 3D pose estimation, and more. [1–5]. However, it is a quite difficult task to obtain efficient and robust object tracking under unrestrained scenarios, due to the background clutter, unpredictable object appearance variations, scale variations, and target object similarity. Because of this, it is important to optimize the tracker response and adapt to target variations in order to continuously track a moving object without disruption. To address these challenges, several tracking approaches, such as correlation filters (CF) and deep learning-based trackers have been developed in object tracking.

In the past few years, discriminative correlation filters (DCF) have obtained much popularity among researchers because of their impressive performance and computational efficiency [6–10]. Also, the

DCF-based trackers deal with the frequency domain transformed from the time domain for getting a remarkable tracking speed. Even though the frequency domain improves the computational efficiency of DCF-based trackers, the negative samples are generated through the circulant shift in the positive samples. Because of that, the trackers did not accurately represent the exact target location during the target appearance variations. Specifically, the negative samples acquire the boundary effects because of the circulant shift process, which considerably decreases the tracker performance. Recently, various tracking methods such as BACF [2], SRDCF [11], and LADCF [12] have been recommended to address the boundary effect issues. In this regard, the authors in [2] have presented a BACF tracker that employs a binary mask to recognize the tracking object from its background region. Also, the BACF tracker keeps the non-zero value within the target region, which considerably decreases the boundary effect issues. Specifically, the authors in [11] have presented the SRDCF tracker, which adopts spatial regularization in the SRDCF tracking approach as an effective way to alleviate boundary effects. Following that, the regularization

\* Corresponding author.

E-mail address: [yhjoo@kunsan.ac.kr](mailto:yhjoo@kunsan.ac.kr) (Y.H. Joo).

weights increase the robustness of the tracker by penalizing the filter values associated with features in the background region. Moreover, the authors in [12] have introduced the LADCF method that integrates spatial and temporal techniques to mitigate the boundary effect issues. Despite the CF tracker's improved tracking efficiency, the tracker may lose its precise object location as the tracking approach becomes more complex. For solving this problem, the spatio-temporal technique and context-aware approach have been taken into account in several analyses [13,14]. The authors in [13] have presented the STRCF tracking approach, which integrates both spatial and temporal regularization techniques into the DCF-based tracker and provides effective real-time visual tracking when the target region is more complex. Following that, the authors in [14] have proposed the CACF tracking technique, which suppresses the background region information around the target region to track the target accurately in more complex scenarios. Although the DCF-based trackers enhance the tracking performance using the hand-crafted feature, they still have unavoidable problems due to object rotation, object deformation, and occlusion. As a result of the above discussions, deep convolutional feature-based trackers can solve these issues by providing robust feature extraction ability.

On the other hand, deep learning-based trackers have obtained outstanding performance in object tracking due to their robust feature extraction capability [15–17]. Recently many researchers have focused on deep feature-based approaches such as SiameseNet [18], VGGNet [19], and ResNet [20]. Specifically, the SiameseNet has attained great success due to its simple architecture and excellent performance. In this regard, the authors in [18] have presented a method called SiameseNet for object tracking. This approach aims to learn a precise target appearance pattern that effectively distinguishes the target region from the background region. Moreover, the authors in [19] utilized a pre-trained VGGNet model for extracting the multiple convolutional layers that enable the recognition of the target from various angles. Also, the VGGNet model integrates the responses of the Conv-3, Conv-4, and Conv-5 layers to obtain accurate target locations and improve tracking performance. More specifically, the authors in [20] have exploited a pre-trained ResNet architecture, which is used to extract features more quickly than traditional methods and obtain adequate target representation. To leverage the rich target representation, we presented a ResNet model, which is used for robust feature extraction and accurate target prediction during the tracking process.

Specifically, the traditional methods for single object tracking often struggled with handling occlusions, scale variations, and abrupt object motions, necessitating robust and adaptable solutions. In recent years, the advent of deep learning has revolutionized the field of computer vision, leading to remarkable advancements in object tracking. Among the various deep learning architectures, transformer-based models have emerged as powerful contenders due to their unparalleled ability to capture long-range dependencies and contextual information within sequential data. In recent times, many researchers have focused on transformer-based methods [21–23]. The authors in [21] have proposed the TransTrack tracking method, which utilizes a transformer backbone to encode features from the current and previous frames. This allows the model to capture long-range temporal dependencies and context information crucial for accurate tracking. Moreover, the authors in [22] have introduced the TrackFormer approach, which demonstrates robustness to scale variations, allowing it to accurately track objects that undergo changes in size or scale across frames. This capability is particularly valuable in scenarios where objects may appear at different distances from the camera or undergo changes in perspective. Overall, transformer-based tracking methods showcase a promising trajectory towards achieving state-of-the-art performance in single object tracking tasks.

Inspired by the above discussions, we present a novel CF-based tracking approach for robust object tracking using an adaptive spatially regularized target attribute-aware background suppressed correlation filter (ASTABSCF). Specifically, the primary focus of this study

is to address the boundary effect problem through the utilization of the spatial regularization technique. Following this, the background suppression approach focuses on effectively dealing with background clutter to mitigate false detections. In addition, we incorporate an adaptive attribute-aware mechanism into the proposed methodology to boost tracking efficiency. By incorporating relevant attributes into the model update process, the filters can better adapt to changes in the target's appearance over time. Moreover, the major contribution of the proposed approach is described as follows:

1. We present an adaptive spatially regularized technique for learning an efficient spatial weight for a particular object, which effectively estimates the target and obtains a more reliable filter in the tracking phase.
2. A target-aware background suppression with dual regression approach is proposed that utilizes a saliency detection technique to produce the target mask, which helps to improve the target's response value.
3. An adaptive attribute-aware approach is presented, which performs post-processing operations during the filter learning stage to remove inappropriate and inconsistent channels and improve tracking performance.
4. A multi-future fusion strategy is presented that combines histogram-oriented gradient (HOG), ColorName, Intensity, and ResNet-50 features to improve target tracking performance and robustness.
5. Without bells and whistles, extensive experiments on the public datasets OTB-2013, OTB-2015, TempleColor-128, UAV-123, LaSOT, and GOT-10K demonstrate the superiority of our proposed approach over other state-of-the-art trackers.

## 2. Related works

### 2.1. Correlation filter-based tracking approaches

In recent years, the CF tracking techniques such as KCF [24], SRDCF [11], DSAR-CF [25], STRCF [13], LADCF [12], and RACF [26] have achieved impressive results in object tracking because of their great performance of the tracking and computational efficiency. In this regard, the authors in [24] have presented the KCF tracking approach, which integrates the multi-channel features (MCF) into the CF methods to enhance tracking efficiency. In particular, the CF-based trackers are mostly affected by undesired boundary effect issues, which impact the tracking efficiency of the tracker. To solve the boundary effect issue, the authors in [11] have presented the SRDCF tracking approach that focuses on the target region instead of the background region through spatial regularization techniques, and it significantly improves tracking performance. Despite that, the SRDCF tracker employs a fixed spatial regularization technique produced from a loose bounding box, which reduces tracking efficiency when the target or background areas express considerable variations, like object distortion and occlusion. To overcome these issues, the authors in [25] have proposed the DSAR-CF tracking technique, which uses the saliency technique to find more information about the target region. Further, the authors in [13] have presented the STRCF tracking technique to expand the search area and keep more information about the target, which effectively handles the target distraction and appearance variations during the tracking process. Specifically, the LADCF tracker [12] integrated spatial and temporal information in the filter learning process to provide an accurate target representation. Also, the LADCF tracking approach effectively identified the problem of spatial boundary effect and cluttered background. Moreover, CF-based trackers are generally affected by object rotation, which impacts tracking efficiency. To overcome this issue, the authors in [26] have presented the RACF tracking approach. Although these tracking methods have improved the tracking performance, the tracking process remains complex and challenging because of the presence of boundary effects. In this work, we present an adaptive spatially regularized CF approach to avoid boundary effect issues and distribute the robust appearance model (see Fig. 1).

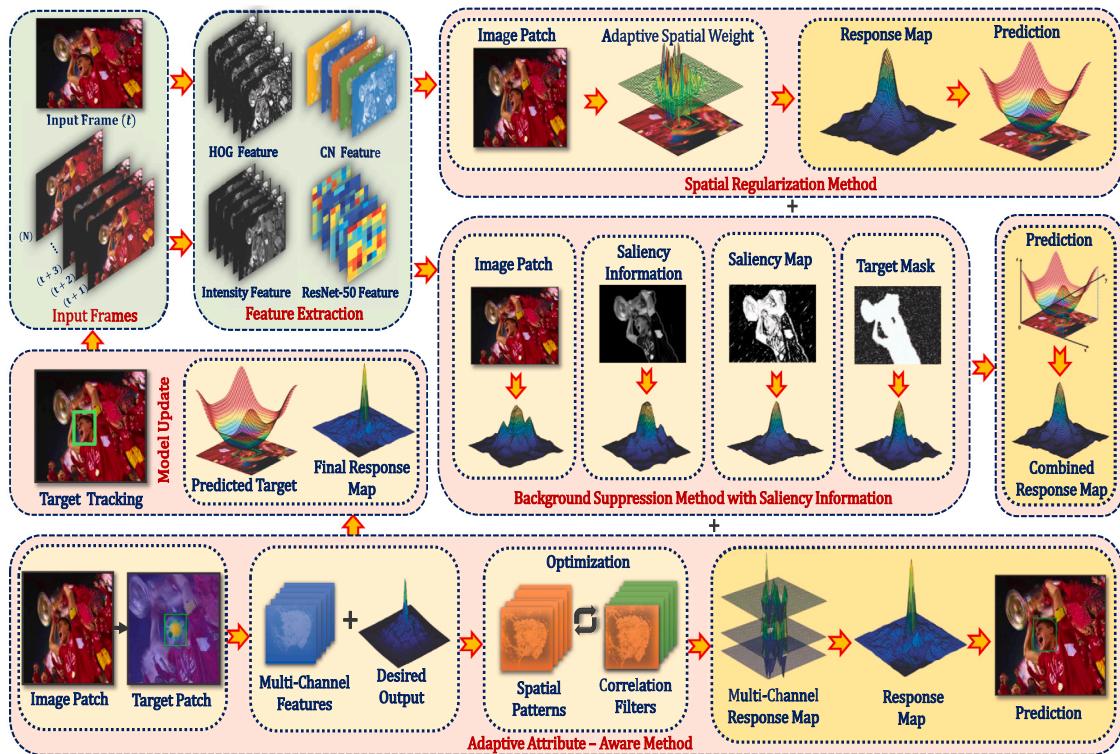


Fig. 1. The overall framework of our proposed ASTABSCF method.

## 2.2. Deep feature-based tracking approaches

Over the past few years, deep learning-based approaches have attained impressive performance, which has increased popularity among the tracking community due to their superior tracking performance and feature extraction ability in object tracking. In this regard, many researchers have sought to integrate the CF methods with deep features such as HDT [27], HCFM [28], Siamese-FC Network [29], and ResNet [20]. In this context, the authors in [27] have introduced the HDT tracker, which utilized the multi-layer convolutional features and adaptive hedged approach to find the target location more accurately. Specifically, the authors in [28] have presented the HCFM tracker, which decreases unwanted background region response weights by proposing the distractor-aware map (DAM). As a result, the HCFM tracker concentrates on the target region and improves its performance of the tracker. Following this, the fast motion and frequent changes in the target's appearance during the tracking process make it difficult to focus on the target. To fix this problem, the authors in [18] have presented Siamese-FC Network, which expands the search region and takes the background information into account to enhance the tracking efficiency. Finally, the authors in [20] have presented the ResNet model, which is utilized in object tracking for robust feature extraction and it helps to obtain the better tracking performance of the CF tracker. As a result, we adopt the ResNet model to demonstrate the excellent performance of the proposed tracker through comparison with existing methods.

## 2.3. Saliency detection technique

Recently, the saliency detection method has achieved impressive performance in traditional CF-based trackers, which can obtain saliency information of the target with greater efficiency. Presently, many researchers have focused on saliency detection methods to enhance DCF-based trackers such as DSAR-CF [25], MSLT [30], and BSCF [31]. In particular, the authors in [25] have presented a DSAR-CF tracker that integrates target information obtained through the saliency detection

approach into a spatial regularization technique, which is adaptively updated according to the target variations. In addition, a long-term tracking approach with a saliency detection technique has been introduced in [30], which employs tracking by detection techniques to identify the exact target location. Specifically, the authors in [31] have proposed the BSCF approach that effectively constructed the background suppression technique, which integrates the all-global background patches to enhance the tracking efficiency. Motivated by the above analysis, we employed the target mask through the efficient saliency detection technique to predict the target location. In this regard, we use the target filter for generating the target mask. Also, the target filter helps the global filter to predict the precise target during the detection phase.

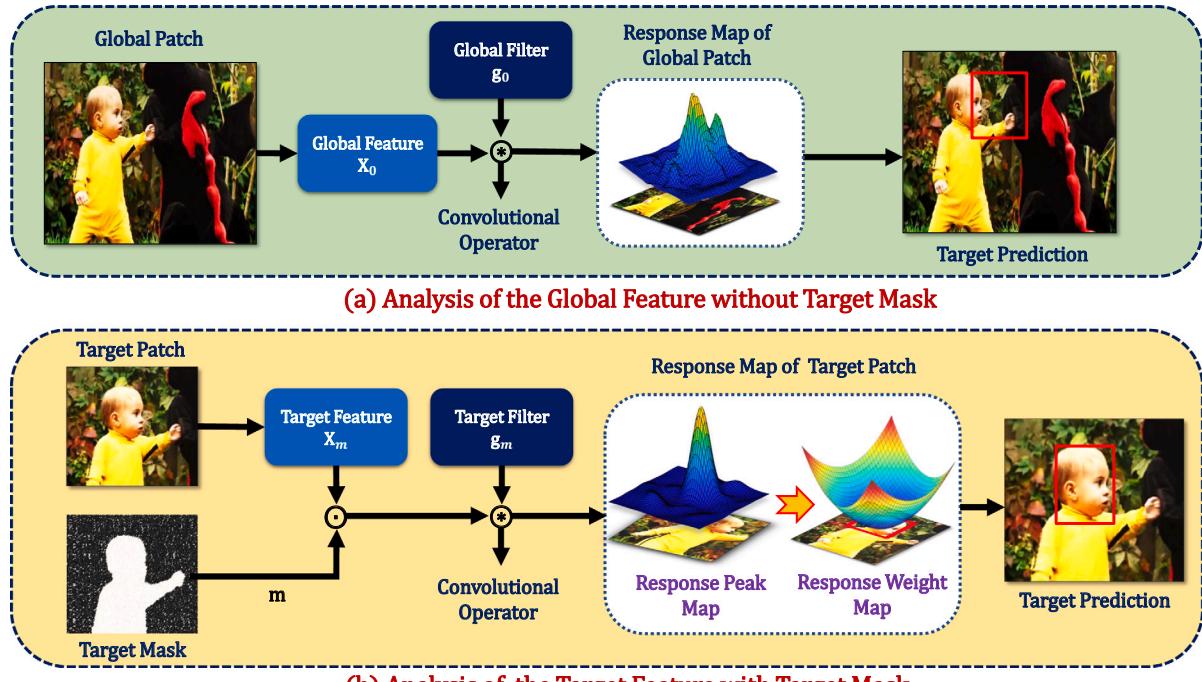
## 3. Proposed method and implementation

### 3.1. Baseline tracking method

In this work, we consider the BACF [2] tracker as our baseline tracker, which aims to maximize the correlation response of the target region while minimizing the response from the background region. Furthermore, the BACF tracker often utilizes FFT for efficient correlation calculation between the filter and the input image. The baseline tracker effectively handles the boundary effect problem by cropping more true negative samples using a binary matrix. Therefore, the formulation of our objective function is described in the following manner:

$$E(h_o) = \frac{1}{2} \sum_{j=1}^T \left\| y(j) - \sum_{k=1}^C X_o^k \otimes (h_o^k \odot P^k) \right\|_2^2 + \frac{\lambda_1}{2} \sum_{k=1}^C \| h_o^k \|^2, \quad (1)$$

where  $X_o^k \in \mathbb{R}^T$  is the training sample,  $y \in \mathbb{R}^T$  is the correlation output and  $h_o^k \in \mathbb{R}^D$  refers to the filter. Further, the  $P^k$  denotes the  $D \times T$  binary matrix and  $\otimes$  represents the circular convolution operation. The  $C$  indicates the number of the feature channels, and  $\lambda_1$  is a regularization parameter.



**Fig. 2.** The target-aware background suppressed technique consists of dual filters known as the global filter and target filter. Analysis of the global feature without a target mask is illustrated in Fig. 2(a). Analysis of the target feature with target mask is exhibited in Fig. 2(b).

### 3.2. Adaptive spatially-regularized correlation filter

The CF methods often face significant challenges due to unwanted boundary effect problems caused by the inclusion of circulant-shifted samples. These boundary effects lead to a decline in tracking efficiency, especially during the tracking process, as they introduce noise and reduce the discriminative power of the model. To address this limitation, an advanced adaptive spatially-regularized CF approach is proposed, which dynamically learns and applies spatial weights tailored to the specific object being tracked, even in scenarios involving rapid and significant target appearance variations. By effectively mitigating the impact of boundary effects and enhancing spatial adaptability, this approach improves the robustness and accuracy of object tracking under diverse and challenging conditions. Therefore, the objective function is calculated in the following way:

$$\begin{aligned} E(h_o, w) = & \frac{1}{2} \sum_{j=1}^T \left\| y(j) - \sum_{k=1}^C X_o^k \otimes (h_o^k \odot P^k) \right\|_2^2 \\ & + \frac{\lambda_1}{2} \sum_{k=1}^C \|w \odot h_o^k\|_2^2 + \frac{\lambda_2}{2} \|w - w_r\|_2^2, \end{aligned} \quad (2)$$

where  $w$  represents a spatial weight that requires to be optimized, and  $w_r$  denotes a reference weight. Specifically, the third term tries to make the adaptive spatial weight  $w$  similar to a reference weight  $w_r$ ,  $\odot$  indicates the element-wise multiplication, and  $\lambda_2$  is a regularization parameter, respectively.

### 3.3. Target-aware background suppressed correlation filter

#### 3.3.1. Target-aware method

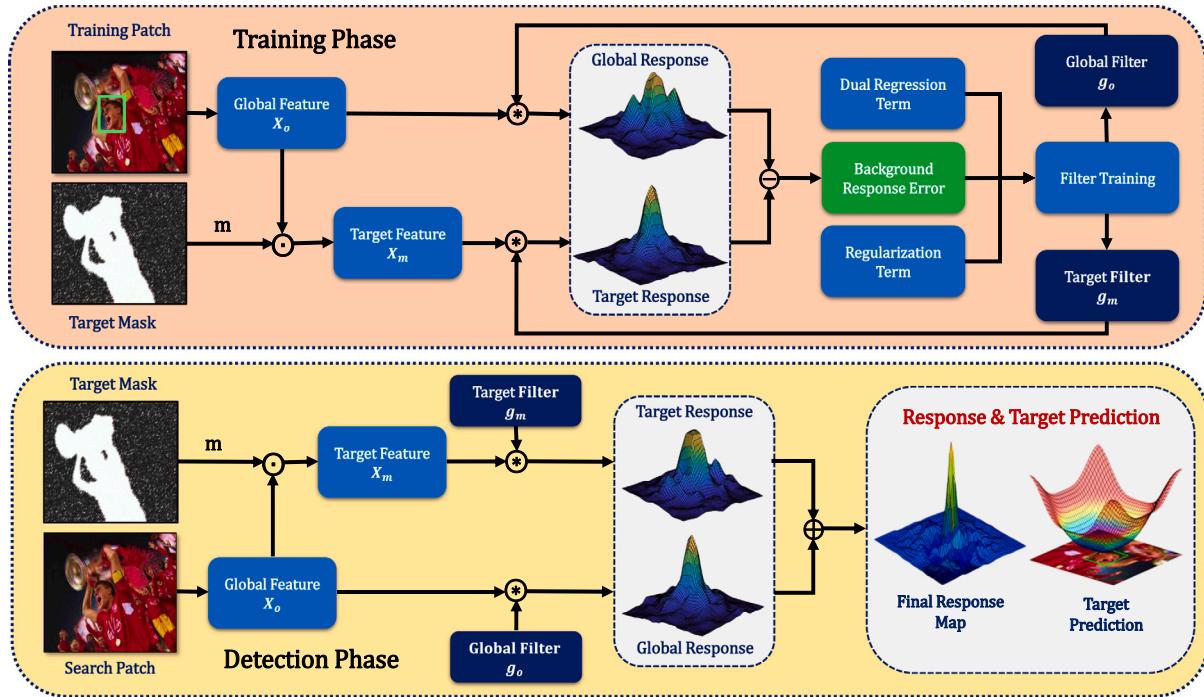
In this section, we discuss the target-aware method and background suppression technique, which is exhibited in Figs. 3 and 4. Also, an analysis of the global and target features is illustrated in Fig. 2. First, we train the global filter  $g_o$  and target filter  $g_m$  in the training stage to improve the response in the target region. Further, the global and target filters are obtained using the dual regression approach.

In particular, the global feature  $X_o$  and target feature  $X_m$  are used to train the global and target filters. Also, the training patch and target region are used to extract global and target features. By leveraging target-features, it optimizes the correlation filter to focus on the target while mitigating distractions from cluttered or dynamic backgrounds. This approach improves tracking accuracy and robustness, particularly in challenging scenarios involving occlusions or complex environments. Moreover, we utilize an effective saliency detection approach [32] to identify the target region, which is denoted by the target mask  $m \in \mathbb{R}^T$ , where  $X_m = X_o \odot m$ . Finally, the dual filters  $g_o$  and  $g_m$  response maps are generated in the detection phase, which is combined to obtain the final response map. As a result, the final response map improves the tracking performance during the detection phase.

#### 3.3.2. Background suppression method

The tracking performance is impacted in the response map, due to the cluttered backgrounds and similar target objects. So, the target filter produces a response map that contains only target information, whereas the background region response is usually zero because the target feature map only consists of target region information. In this regard, the distractors have an impact on the fused response map because of the global filter. Therefore, the problem of suppressing distractors becomes the problem of suppressing the background response produced by the global filter. To address the above-mentioned issue, a background suppression approach has been presented. In this approach, the target response can suppress the background region associated with a large number of disturbances in the global response, which ensures that the desired target response is highest in the final response map. In this regard, the target region of the global response is identical to the center region of the target response because the global and target response map to the same target. As a result, dual filters improve tracking performance by avoiding overfitting issues. The formulation of our objective function is described as follows:

$$E(h_o, w) = \frac{1}{2} \sum_{j=1}^T \left\| y(j) - \sum_{k=1}^C X_o^k \otimes (h_o^k \odot P^k) \right\|_2^2$$



**Fig. 3.** The overall framework of the background suppression method with dual regression technique is made up of two major blocks, which are referred to as the training phase and the detection phase.



**Fig. 4.** The proposed background suppression method with a target-aware strategy. The first row represents the input sample, and the second row represents the target mask extracted from the input samples using the saliency detection approach.

$$\begin{aligned}
 & + \frac{\lambda_1}{2} \sum_{k=1}^C \|w \odot h_o^k\|_2^2 + \frac{\lambda_2}{2} \|w - w_r\|_2^2 \\
 & + \frac{\gamma}{2} \left\| \sum_{k=1}^C X_o^k \otimes (h_o^k \odot P^k) - \sum_{k=1}^C X_m^k \otimes (h_m^k \odot P^k) \right\|_2^2,
 \end{aligned} \quad (3)$$

where  $h_o^k$  and  $h_m^k$  represent the filters and  $\gamma$  denote the regularization parameter. Further, the target response is summarized as  $R_m$  to simplify the subsequent derivation, where it is equal to  $\sum_{k=1}^C X_m^k \otimes (h_m^k \odot P^k)$ , respectively.

#### 3.4. Adaptive attribute-aware correlation filter

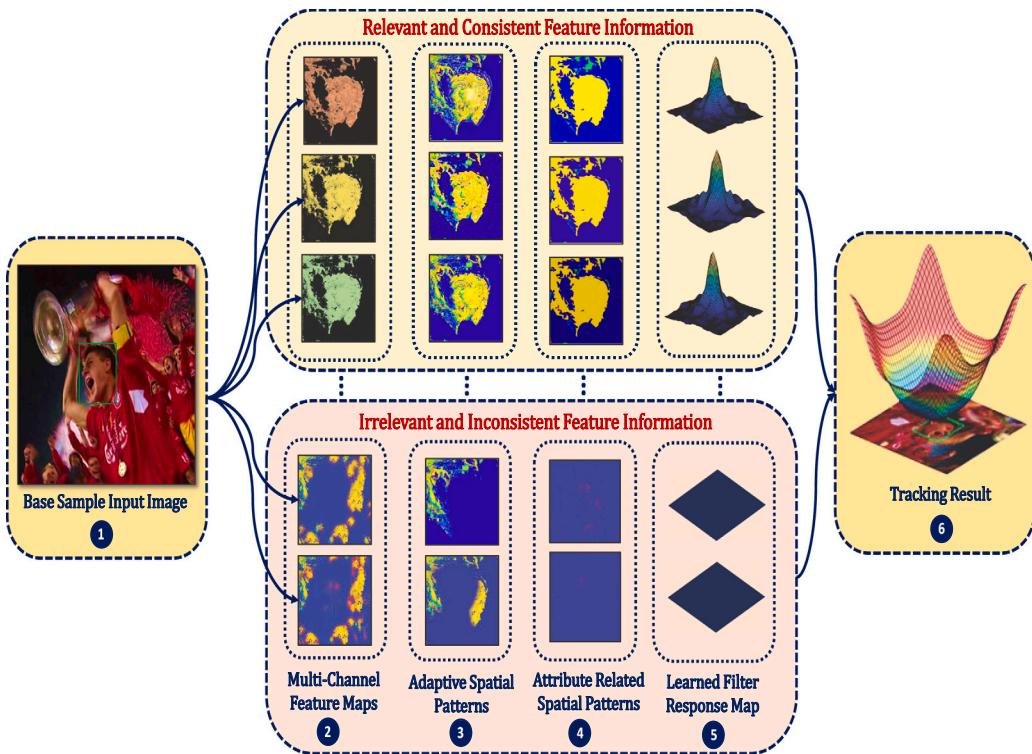
The filters are trained by using the MCF method, which is extracted from a search region around the target. Although, these MCF typically contain inappropriate and inconsistent information, which reduces the performance of the learned pattern. To overcome this problem, the adaptive attribute-aware approach is presented in the filter learning phase, which emphasizes the discriminating features of the specific channel. Therefore, the objective function is formulated as follows:

$$E(h_o, w, p) = \frac{1}{2} \sum_{j=1}^T \left\| y(j) - \sum_{k=1}^C X_o^k \otimes (h_o^k \odot P^k) \right\|_2^2$$

$$\begin{aligned}
 & + \frac{\lambda_1}{2} \sum_{k=1}^C \|w \odot h_o^k\|_2^2 + \frac{\lambda_2}{2} \|w - w_r\|_2^2 \\
 & + \frac{\gamma}{2} \left\| \sum_{k=1}^C X_o^k \otimes (h_o^k \odot P^k) - R_m \right\|_2^2 + \frac{\lambda_3}{2} \sum_{k=1}^C \|P^k - P_r^k\|_2^2,
 \end{aligned} \quad (4)$$

where  $P^k$  denotes the spatial attention model, which is represented as a binary matrix. Also,  $P_r^k$  denotes the predefined binary matrix, and  $\lambda_3$  indicates the penalty factor. In this regard, the first term in Eq. (4) is a minimization process in which the spatial attention model  $P^k$  and filter  $h_o^k$  are optimized simultaneously. In particular, the fifth term utilizes the previous information of the filters to concentrate their energy on the central position of the search region, which is represented in the form of binary mask  $P_r^k$ .

Moreover, the MCF maps and learned spatial attention patterns are illustrated in Fig. 5. From this figure, the first segment indicates the search region of the base sample, which aims to predict the central target position. The second segment represents the MCF maps of the search region, which utilizes the ColorName feature to extract more information in the search region. The third segment contains spatial information, which focuses on adaptively learned spatial information from the base sample. The fourth segment illustrates the learned attribute-related spatial patterns (ARSP) attained through the spatial



**Fig. 5.** The proposed adaptive attribute-aware approach consists of attribute-related spatial patterns. Further, this approach is divided into two major parts, the first part describes the relevant and consistent feature maps and the second part presents the irrelevant and inconsistent feature maps.

pattern post-processing technique. Particularly, the fourth segment is produced by imposing spatial patterns as a previous constraint on the third segment. Also, the fifth segment shows the response maps of the optimized correlation filters for the ARSP model. The final column exhibits the tracking result of the adaptive ARSP model. Therefore, the learned adaptive spatial attention patterns significantly reduce the irrelevant and inconsistent information of MCF and improve the tracker performance.

### 3.5. Frequency domain transformation

The Fourier transform is commonly used to learn the correlation filter and target detection in the frequency domain. To achieve the effective solution of Eq. (4), the spatial attention model is formalized as  $h_o^k \equiv h_o^k \odot P^k$ .

$$\begin{aligned} E(h_o, \hat{g}_o, w, p) = & \frac{1}{2} \left\| \hat{y} - \sum_{k=1}^C \hat{X}_o^k \odot \hat{g}_o^k \right\|_2^2 + \frac{\lambda_1}{2} \sum_{k=1}^C \|w \odot h_o^k\|_2^2 + \frac{\lambda_2}{2} \|w - w_r\|_2^2 \\ & + \frac{\gamma}{2} \left\| \sum_{k=1}^C \hat{X}_o^k \odot \hat{g}_o^k - R_m \right\|_2^2 + \frac{\lambda_3}{2} \sum_{k=1}^C \|P^k - P_r^k\|_2^2, \\ \text{s.t. } \hat{g}_o^k = & \sqrt{T}(FP^\top \otimes I^k)h_o^k, \end{aligned} \quad (5)$$

where  $\hat{g}_o^k$  denotes the auxiliary variable,  $\wedge$  represents the discrete Fourier transform,  $\otimes$  denotes the Kronecker product,  $I^k$  indicates the identity matrix and  $F$  denotes the  $P * P$  Fourier matrix.

### 3.6. Augmented Lagrangian method

We employ the ADMM method to boost the computational speed, which helps to guide the optimization process towards better convergence by gradually adjusting the penalty parameters. To solve the Eq. (5), we utilize the ALM technique in the frequency domain.

$$\mathcal{L}(h_o, \hat{g}_o, w, p, \hat{\zeta})$$

$$\begin{aligned} & = \frac{1}{2} \left\| \hat{y} - \sum_{k=1}^C \hat{X}_o^k \odot \hat{g}_o^k \right\|_2^2 + \frac{\lambda_1}{2} \sum_{k=1}^C \|w \odot h_o^k\|_2^2 + \frac{\lambda_2}{2} \|w - w_r\|_2^2 \\ & + \frac{\gamma}{2} \left\| \sum_{k=1}^C \hat{X}_o^k \odot \hat{g}_o^k - R_m \right\|_2^2 + \frac{\lambda_3}{2} \sum_{k=1}^C \|P^k - P_r^k\|_2^2 \\ & + \sum_{k=1}^C (\hat{\zeta}^k)^\top (\hat{g}_o^k - \sqrt{T}(FP^\top \otimes I^k)h_o^k) + \frac{\mu}{2} \sum_{k=1}^C \|\hat{g}_o^k - \sqrt{T}(FP^\top \otimes I^k)h_o^k\|_2^2, \end{aligned} \quad (6)$$

where  $\zeta$  denotes a Lagrange multiplier and  $\mu$  represents the penalty factor ( $\mu \geq 0$ ). Also, the Lagrangian term  $\mathcal{L}(h_o, \hat{g}_o, w, p, \hat{\zeta})$  can be minimized via ADMM technique by solving sub-problems  $h_{o+1}^*$ ,  $\hat{g}_{o+1}^*$ ,  $\hat{w}_{o+1}^*$  and  $\hat{P}_{o+1}^*$ .

#### 3.6.1. Sub-problem $h_{o+1}^*$

The sub-problem  $h_{o+1}^*$  is described as:

$$\begin{aligned} h_{o+1}^* = \underset{h_o}{\operatorname{argmin}} \Big\{ & \frac{\lambda_1}{2} \sum_{k=1}^C \|w \odot h_o^k\|_2^2 + \sum_{k=1}^C (\hat{\zeta}^k)^\top (\hat{g}_o^k - \sqrt{T}(FP^\top \otimes I^k)h_o^k) \\ & + \frac{\mu}{2} \sum_{k=1}^C \|\hat{g}_o^k - \sqrt{T}(FP^\top \otimes I^k)h_o^k\|_2^2 \Big\}. \end{aligned}$$

After solving Eq. (7), the closed-form solution of sub-problem  $h_{o+1}^*$  can be calculated as follows:

$$\begin{aligned} & = [\lambda_1 W^\top W + \mu T P^\top P]^{-1} T P (\zeta^k + \mu g_o^k), \\ h_{o+1}^* = & \frac{T p \odot (\zeta^k + \mu g_o^k)}{\lambda_1 (w \odot w) + \mu T p}, \end{aligned} \quad (7)$$

where  $W = \operatorname{diag}(w)$  denotes the diagonal matrix and  $p = [P_{11}, P_{22}, \dots, P_{TT}]^\top$  is the column vector formed by the diagonal elements of the cropping matrix  $P$ . Also, we have set the cropping matrix  $P = P^\top P$ .

### 3.6.2. Sub-problem $\hat{g}_{o+1}^*$

The sub-problem  $\hat{g}_{o+1}^*$  is described as:

$$\begin{aligned} g_{o+1}^* = \operatorname{argmin}_{g_o} & \left\{ \frac{1}{2} \left\| \hat{y} - \sum_{k=1}^C \hat{X}_o^k \odot \hat{g}_o^k \right\|_2^2 + \frac{\gamma}{2} \left\| \sum_{k=1}^C \hat{X}_o^k \odot \hat{g}_o^k - \hat{R}_m^k \right\|_2^2 \right. \\ & + \sum_{k=1}^C (\hat{\zeta}^k)^T (\hat{g}_o^k - \sqrt{T}(FP^T \otimes I^k)h_o^k) \\ & \left. + \frac{\mu}{2} \sum_{k=1}^C \left\| \hat{g}_o^k - \sqrt{T}(FP^T \otimes I^k)h_o^k \right\|_2^2 \right\}. \end{aligned} \quad (8)$$

After solving Eq. (8), we can get as follows:

$$\begin{aligned} \hat{g}_{o+1}^* = & \left( \frac{1}{1 + \gamma T} \right) \left( \hat{x}_o^k(t) (\hat{x}_o^k)^T(t) + \frac{\mu IT}{1 + \gamma T} \right)^{-1} \\ & \left( \hat{x}_o^k(t) \hat{y}(t) + \gamma T \hat{x}_o^k(t) \hat{R}_m^k(t) - T \hat{\zeta}(t) + \mu T \hat{h}_o^k(t) \right). \end{aligned} \quad (9)$$

To improve the computational efficiency and avoid the inverse operation, we employ the Sherman–Morrison method [33]. Therefore, Eq. (9) is redefined as follows:

$$\begin{aligned} \hat{g}_{o+1}^* = & a \left( \hat{x}_o^k(t) \hat{y}(t) + \gamma T \hat{x}_o^k(t) \hat{R}_m^k(t) - T \hat{\zeta}(t) + \mu T \hat{h}_o^k(t) \right) \\ & - a \frac{\hat{x}_o^k(t)}{b} \left( \hat{S}_x(t) \hat{y}(t) + \gamma T \hat{S}_x(t) \hat{R}_m^k(t) - T \hat{S}_\zeta(t) + \mu T \hat{S}_h(t) \right), \end{aligned} \quad (10)$$

where,  $a = \left( \frac{1 + \gamma T}{\mu T} \right)$ ,  $\hat{S}_x(t) = (\hat{x}_o^k)^T(t) \hat{x}_o^k(t)$ ,  $\hat{S}_\zeta(t) = (\hat{x}_o^k)^T(t) \hat{\zeta}(t)$ ,  $\hat{S}_h(t) = (\hat{x}_o^k)^T(t) \hat{h}_o^k(t)$  and  $b = (\hat{x}_o^k)^T(t) \hat{x}_o^k(t) + \frac{1}{a}$  are scalar.

### 3.6.3. Sub-problem $\hat{w}_{o+1}^*$

The sub-problem  $w_{o+1}^*$  is described as:

$$\begin{aligned} w_{o+1}^* = \operatorname{argmin}_{w_o} & \left\{ \frac{\lambda_1}{2} \sum_{k=1}^C \|N^k w\|_2^2 + \frac{\lambda_2}{2} \|w - w_r\|_2^2 \right\}, \\ & = \left( \lambda_1 \sum_{k=1}^C (N^k)^T N^k + \lambda_2 I \right)^{-1} \lambda_2 w_r, \\ w_{o+1}^* = & \frac{\lambda_2 w_r}{\lambda_1 \sum_{k=1}^C h_o^k \odot h_o^k + \lambda_2 I}, \end{aligned} \quad (11)$$

where  $N^k = \operatorname{diag}(h_o^k) \in \mathbb{R}^{T \times T}$  represents the diagonal matrix.

### 3.6.4. Sub-problem $\hat{P}_{o+1}^*$

The sub-problem  $P^k$  is a binary optimization, which is embedded into filter  $h_o^k$ . In this regard, a variable is summarized  $J$  as  $J^k = h_o^k \odot P^k$  and  $J_r$  as  $J_r^k = h_o^k \odot P_r^k$ . The sub-problem  $\hat{P}_{o+1}^*$  is introduced as sub-problem  $\hat{J}_{o+1}^*$ . Hence, the sub-problem  $J_{o+1}^*$  is calculated as follows:

$$\begin{aligned} \hat{J}_{o+1}^* = \operatorname{argmin}_{J_o} & \left\{ \frac{1}{2} \left\| \hat{y} - \sum_{k=1}^C \hat{X}_o^k \odot J_o^k \right\|_2^2 + \frac{\lambda_3}{2} \sum_{k=1}^C \left\| \hat{J}^k - \hat{J}_r^k \right\|_2^2 \right\}, \\ \hat{J}_{o+1}^* = & \frac{\lambda_3 J_r^k + \hat{y}^T \hat{X}_o^k}{(\hat{X}_o^k)^T \hat{X}_o^k + \lambda_3 I}. \end{aligned} \quad (12)$$

In Eq. (12), the  $J^k$  value is directly calculated by predefined value  $P_r^k$ . Further, if the  $J^k$  value is identified in the spatial domain operation using the inverse discrete Fourier transform (DFT) of  $\hat{J}^k$ , we include it binarization process to generate the binary matrix  $P^k$ . In this regard, we apply the binarization process to  $J_t^k$ , where  $J_t^k$  defined as  $J_t^k = (1 - \gamma) J_{t-1}^k + \gamma J_t^k$  and  $\gamma$  represents predefined constant. Also,  $t$  indicates the current frame of the video sequence, and  $P^k$  is denoted as the initial ARSP model. Specifically, we utilized the identified spatial model in a post-processing operation to obtain the final ARSP model, which is defined as the prior constraint  $P_r^k$ . Finally, the previous constraint is executed to update the  $P^k$ .

## Post processing technique

After getting the primary ARSP model  $P^k$ , the post-processing operation is employed. In particular, we upgrade  $P^k$  by applying a previous constraint:

$$P^k = P^k \odot P_r^k. \quad (13)$$

In Eq. (13), the  $P^k$  is considered as the final ARSP model, which focuses on the more central region of the target. Therefore, the learned adaptive spatial attention models significantly reduce the irrelevant information of multi-channel features.

### Update filter $h_o^k$

After updating the spatial attention model  $P^k$ , we update the filter  $h_o^k$  by applying the spatial attention model. Therefore, the filter updating process is defined as follows:

$$h_o^k = h_o^k \odot P^k. \quad (14)$$

In Eq. (14), the filter  $h_o^k$  is updated in the spatial domain. Also, after updating the filter, the frequency domain is used to upgrade the Lagrangian multipliers.

### Lagrangian multiplier update

We update the Lagrangian multiplier as follows:

$$\begin{aligned} \hat{\zeta}_{t+1} & \leftarrow \hat{\zeta}_t + \mu (\hat{g}_{o+1}^* - \hat{h}_{o+1}^*), \\ \mu_{t+1} & = \min(\mu_{\max}, \beta \mu_t), \end{aligned} \quad (15)$$

where  $\hat{\zeta}_t$  represents the Lagrangian parameter,  $\mu_{\max}$  denotes the predefined maximum value and  $\beta$  indicates the scale factor.

### Model update

This section mainly focuses on the online model update technique, which is an essential component of object tracking. In addition, the target appearance of an object frequently varies during the tracking process due to background clutter, scale variations, occlusion, rotation, object deformation, and so on. Specifically, we improve the tracking efficiency and robustness of the proposed approach by training the filter through an online update technique similar to existing CF-based trackers such as [2,24]. Hence, the online model update technique is calculated in the following way:

$$\hat{X}_t^{\text{model}} = (1 - \gamma) \hat{X}_{t-1}^{\text{model}} + \gamma \hat{X}_t, \quad (16)$$

where  $\hat{X}_t^{\text{model}}$  represents the newly updated model and  $\hat{X}_{t-1}^{\text{model}}$  denotes the old model. Moreover,  $\hat{X}_t$  indicates the current learned filter and  $\gamma$  represents the learning rate of the target model. Finally,  $t$  and  $t-1$  denote the current frame and learned model of the previous frame, respectively.

## 4. Experimental results and analysis

### 4.1. Experimental setup

In this section, we discuss our experimental setup and feature representation details. The proposed tracking approach is implemented in MATLAB and the experiments are carried out on a PC with an Intel(R) Core(TM) i5-6500 CPU at 3.20 GHz and 16 GB RAM. Also, the proposed approach utilizes deep and hand-crafted features such as HOG, Intensity, ColorName, and ResNet-50 that are used to extract the features in the image sequences. Furthermore, these features exhibit robust feature representations, which play a crucial role in improving tracking efficiency.

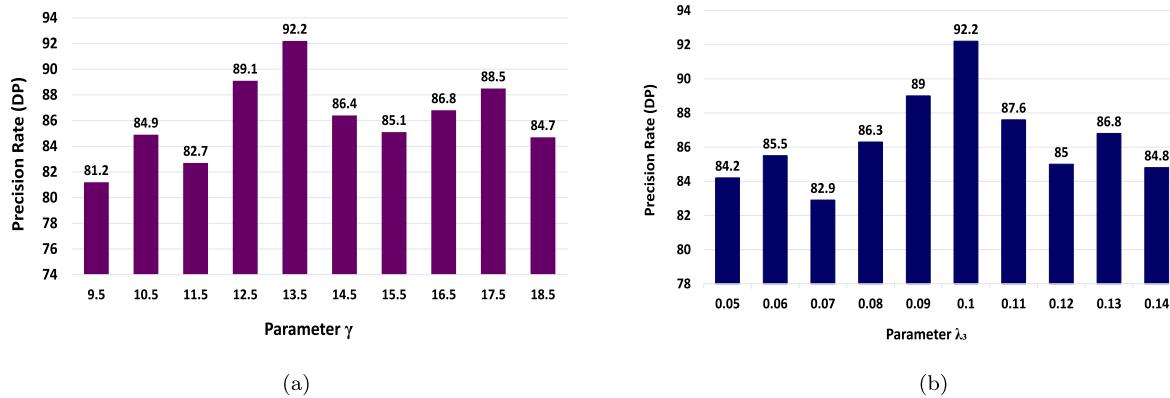


Fig. 6. Parameter analysis of  $\gamma$  and  $\lambda_3$  on the OTB-2015 dataset.

#### 4.2. Evaluation metrics

In this study, we present the experimental results on six standard challenging datasets, such as OTB-2013, OTB-2015, TempleColor-128, UAV-123, LaSOT, and GOT-10K datasets. Moreover, the performance of the proposed approach is calculated through one-pass evaluation (OPE) and center location error (CLE). Finally, the video sequences in all datasets are classified with 11 different challenging attributes as follows: low resolution (LR), out-of-view (OV), illumination variation (IV), occlusion (OCC), fast motion (FM), in-plane rotation (IPR), deformation (DEF), scale variation (SV), out-of-plane rotation (OPR), motion blur (MB), and background clutter (BC).

#### 4.3. Parameter analysis

##### 4.3.1. Parameter analysis on the OTB-2015 dataset

In our experiments, we conduct the parameter analysis on the OTB-2015 dataset. Initially, we analyze the  $\gamma$  parameter in Eq. (8) and set the value of  $\gamma = 13.5$ . Moreover, the experimental results are exhibited in Fig. 6(a). Further, we investigate the regularization parameters  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  in Eqs. (7), (11) and (12). The parameters  $\lambda_1$  and  $\lambda_2$  fixed value, such that,  $\lambda_1 = 100$  and  $\lambda_2 = 0.2$ , respectively. When  $\lambda_3$  changes from 0.05 to 0.14, the precision score improves consistently and obtains the highest value at  $\lambda_3 = 0.1$ . Also, the  $\lambda_3$  parameter analysis with different values is illustrated in Fig. 6(b). Finally, we set the learning rate ( $\eta$ ) parameters 0.01 and 0.008 for hand-crafted and deep features, respectively.

##### 4.3.2. Parameter analysis on the TC-128 dataset

We conduct a parameter analysis on the TC-128 dataset, focusing on the  $\gamma$  and  $\lambda$  parameters, which are crucial for enhancing tracker performance. As shown in Fig. 7(a), when increasing the  $\gamma$  value from 10.1 to 20.9 yields the best result of 81.8% at  $\gamma = 13.5$ . Additionally, we analyze the parameter  $\lambda_3$ . When  $\lambda_3$  increases from 0.02 to 0.18, the precision score consistently improves, reaching its highest value at  $\lambda_3 = 0.1$  as shown in Fig. 7(b).

#### 4.4. Ablation studies

To evaluate the proposed tracker performance variations of each component, we perform an ablation analysis on the OTB-2015 dataset. To do this, we compared our proposed approach with the baseline tracker and different components of the proposed technique, such as spatial regularization (SR), background suppression (BS), and adaptive attribute-aware (AAA). Further, the ablation analysis results are illustrated in Fig. 8. As shown in Fig. 8, we can confirm that the baseline tracker (*Baseline*) obtained the distance precision (DP) score of (80.2%) and area under curve (AUC) score of (61.0%). We can

see that by adding the deep feature (ResNet-50) to our base tracker (*Baseline + Deep Feature*), the tracking performance is significantly improved in DP and AUC scores (5.4% / 1.8%). In addition, we observed that DP/AUC scores are greatly increased (7.5% / 3.0%) after adding the spatial regularization term with our baseline tracker (*Baseline + Deep Feature + SR*). Also, by incorporating the background suppression method with our baseline tracker (*Baseline + Deep Feature + SR + BS*), we ensure that the tracking efficiency is increased in DP/AUC scores of (10.3% / 5.7%). Specifically, we have proven that the proposed tracker (*Baseline + Deep Feature + SR + BS + AAA*) attains the best results in terms of precision and success scores (12.0% / 8.2%) by adding an adaptive attribute-aware approach. Finally, compared to the baseline method, we ensure that the overall presented technique demonstrates a better result in DP/AUC scores (92.2% / 69.2%) (see Table 1).

#### 4.5. Comparative analysis with state-of-the-art trackers

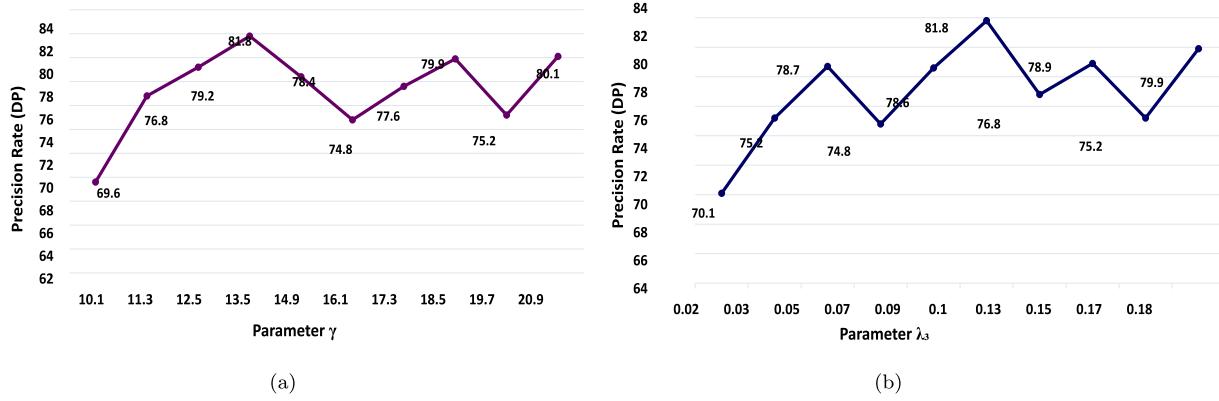
In this division, we evaluate our proposed tracker with other 34 various trackers such as BACF [2], SRDCF [11], LADCF [12], STRCF [13], MEVT [19], KCF [24], DSAR-CF [25], RACF [26], HDT [27], HCFM [28], CSK [34], MCCT [35], A3DCF [36], AutoTrack [37], MDNet [38], SASR [39], ARCF [40], BSTCF [41], STAR [42], AMCF [43], SAMF [44], VITAL [45], SiamFC [46], StructSiam [47], DSiam [48], CFNet [49], CSRDCF [50], DSST [51], AFSN [52], ATOM [53], DaSi-amRPN [54], SiamRPN [55], FDSiamFC [56], and GOTRUN [57].

#### 4.6. Experimental evaluation

##### 4.6.1. Evaluation on OTB-2013 dataset

We analyze the proposed ASTABSCF approach with the other 19 modern trackers on the OTB-2013 dataset. Table 2 presents the comparison results with the conventional and proposed trackers DP and AUC scores. As illustrated in Table 2 and Fig. 9, the presented approach obtains great DP/AUC scores (94.5%) and (71.8%) when compared to all other trackers.

Moreover, when compared to the baseline [2] approach, we ensure that our proposed technique enhances a significant gain in DP score by (10.2%) and AUC score by (7.2%). Compared to hand-crafted feature-based approaches such as STRCF [13], LADCF [12], SRDCF [11], and ARCF [40] the proposed method outperforms (5.5%/4.0%), (8.1%/4.3%), (10.7%/9.2%), and (11.7%/9.2%) in DP/AUC scores, respectively. In addition, when compared to the deep learning-based methods such as DaSi-amRPN [54], HCFM [28], BSTCF [41], SiamRPN [55] STAR [42], and SASR [39], we demonstrate that our presented tracker achieves better improvements of (2.1%/1.4%), (2.4%/2.0%), (3.4%/3.5%), (3.7%/2.6%), (5.3%/3.0%), and (13.1%/23.8%) in DP and AUC scores on the OTB-2013 benchmark dataset.

Fig. 7. Parameter analysis of  $\gamma$  and  $\lambda_3$  on the TC-128 dataset.

**Table 1**  
Ablation analysis for each proposed technique with speed on the OTB-2015 dataset.

| Methods  | DP (%) | AUC (%) | FPS  |
|--|--------|---------|------|
| Baseline + Deep Feature + SR + BS + AAA [Ours] | 92.2   | 69.2    | 10.5 |
| Baseline + Deep Feature + SR + BS              | 90.5   | 66.7    | 11.3 |
| Baseline + Deep Feature + SR                   | 87.7   | 64.0    | 18.5 |
| Baseline + Deep Feature                        | 85.6   | 62.8    | 23.2 |
| Baseline                                       | 80.2   | 61.0    | 26.7 |

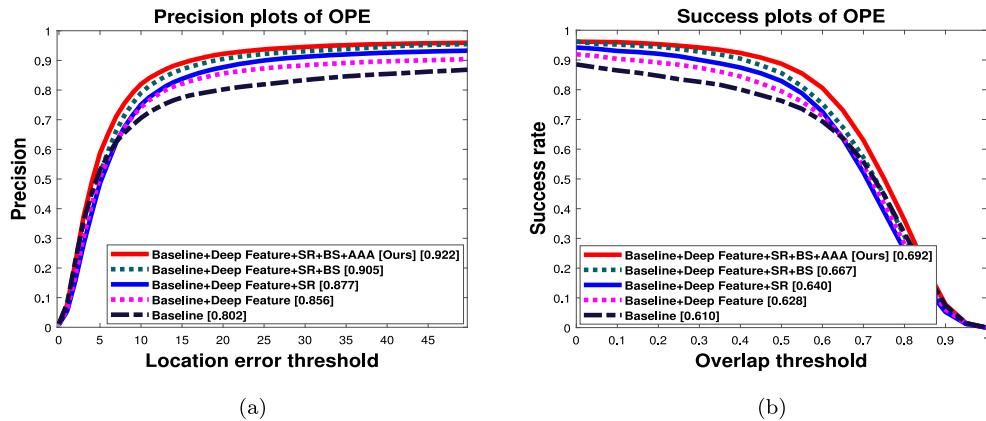


Fig. 8. The precision and success plots of the proposed method on the OTB-2015 dataset are illustrated in Fig. 8(a) and Fig. 8(b).

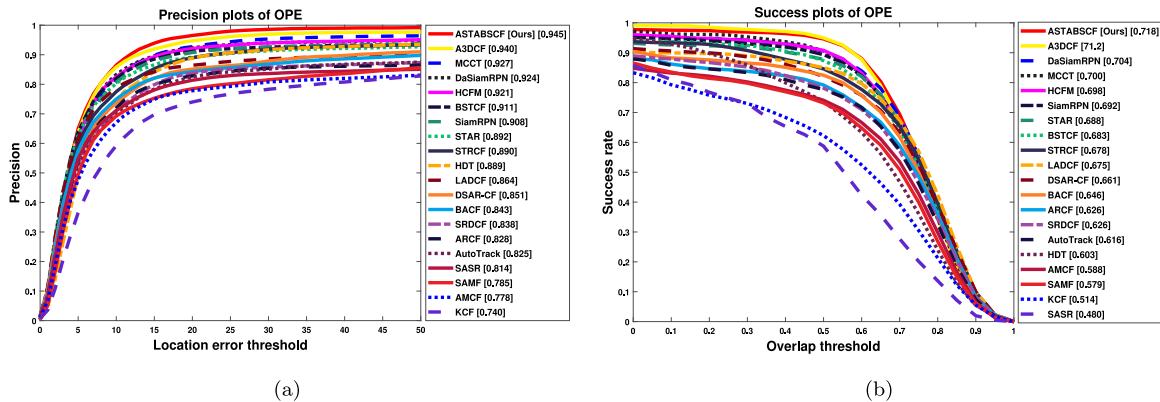


Fig. 9. The distance precision and overlap success results of the OTB-2013 dataset are exhibited in Fig. 9(a) and Fig. 9(b).

#### 4.6.2. Statistical analysis

In this section, we performed the statistical analysis of the proposed method with 5 cutting-edge approaches on the OTB-2013 dataset, which is illustrated in Table 3. The first column represents the precision

score for each sequence, while the second column shows the tracker ranking position. As shown in Table 3, our proposed approach achieved the best result in the precision score and ranking position compared to the cutting-edge methods such as ARCF [40], LADCF [12], BACF [2],

**Table 2**

| Comparative results of conventional trackers with the proposed tracker on the OTB-2013 dataset with 50 sequences. The results of the top 15 trackers are presented as follows: |         |                   |                    |                   |                        |                   |                    |                      |            |            |                  |            |             |             |       |             |
|--|---------|-------------------|--------------------|-------------------|------------------------|-------------------|--------------------|----------------------|------------|------------|------------------|------------|-------------|-------------|-------|-------------|
| Methods  | Metrics | Ours <sup>a</sup> | A3DCF <sup>a</sup> | MCCT <sup>a</sup> | DaSiamRPN <sup>a</sup> | HCFM <sup>a</sup> | BSTCF <sup>a</sup> | SiamRPN <sup>a</sup> | STAR       | STRCF      | HDT <sup>a</sup> | LADCF      | DSAR-CF     | BACF        | SRDCF | ARCF        |
| OTB-2013   | DP      | <b>94.5</b>       | 94.0               | 92.7              | 92.4                   | 92.1              | 91.1               | 90.8                 | 89.2       | 89.0       | 88.9             | 86.4       | 85.1        | 84.3        | 83.8  | 82.8        |
|  | AUC     | <b>71.8</b>       | 71.2               | 70.0              | 70.4                   | 69.8              | 68.3               | 69.2                 | 68.8       | 67.8       | 60.3             | 67.5       | 66.1        | 64.6        | 62.6  | 62.6        |
| FPS  |         | <b>10.8</b>       | —                  | <b>7.8</b>        | <b>160.0</b>           | <b>2.7</b>        | <b>19.0</b>        | <b>160.0</b>         | <b>5.5</b> | <b>5.8</b> | <b>10.0</b>      | <b>1.5</b> | <b>16.0</b> | <b>25.4</b> | —     | <b>26.2</b> |

<sup>a</sup> Denotes the hand-crafted and deep features-based trackers.

**Table 3**

The left column under each tracker name contains the precision value for each video sequence on the OTB-2013 dataset, while the right column shows the position in the ranking with respect to the rest of the trackers. The top and second best precision score and ranking position are highlighted in red and blue color fonts.

| Dataset      | Ours        | Rank        | ARCF | Rank | LADCF       | Rank        | BACF | Rank | SASR | Rank | STRCF       | Rank        |
|--------------|-------------|-------------|------|------|-------------|-------------|------|------|------|------|-------------|-------------|
| Basketball   | 0.98        | 2           | 1    | 1    | 1           | 1           | 0.98 | 2    | 0.26 | 3    | 1           | 1           |
| Bolt         | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| Boy          | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| Car4         | 1           | 1           | 0.97 | 3    | 1           | 1           | 0.98 | 2    | 1    | 1    | 1           | 1           |
| CarDark      | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| CarScale     | 0.71        | 2           | 0.73 | 1    | 0.68        | 3           | 0.71 | 2    | 0.67 | 4    | 0.68        | 3           |
| Coke         | 0.96        | 2           | 0.88 | 5    | 0.97        | 1           | 0.84 | 6    | 0.95 | 3    | 0.93        | 4           |
| Couple       | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 0.97 | 2    | 1           | 1           |
| Crossing     | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| David        | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| David2       | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| David3       | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| Deer         | 1           | 1           | 1    | 1    | 1           | 1           | 0.97 | 2    | 1    | 1    | 1           | 1           |
| Dog1         | 0.90        | 4           | 0.93 | 3    | 1           | 1           | 1    | 1    | 0.85 | 5    | 0.98        | 2           |
| Doll         | 0.99        | 1           | 0.99 | 1    | 0.99        | 1           | 0.99 | 1    | 0.87 | 2    | 0.99        | 1           |
| Dudek        | 0.95        | 1           | 0.61 | 5    | 0.86        | 3           | 0.80 | 4    | 0.58 | 6    | 0.88        | 2           |
| FaceOcc1     | 0.34        | 4           | 0.40 | 3    | 0.62        | 1           | 0.30 | 5    | 0.34 | 4    | 0.48        | 2           |
| FaceOcc2     | 0.90        | 1           | 0.68 | 6    | 0.76        | 3           | 0.71 | 5    | 0.72 | 4    | 0.77        | 2           |
| Fish         | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| FleetFace    | 0.76        | 1           | 0.64 | 4    | 0.69        | 2           | 0.53 | 6    | 0.57 | 5    | 0.68        | 3           |
| Football     | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| Football1    | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| Freeman1     | 1           | 1           | 1    | 1    | 1           | 1           | 0.37 | 2    | 1    | 1    | 1           | 1           |
| Freeman3     | 1           | 1           | 0.80 | 4    | 1           | 1           | 0.95 | 2    | 0.90 | 3    | 1           | 1           |
| Freeman4     | 0.93        | 3           | 0.99 | 1    | 0.98        | 2           | 0.18 | 5    | 0.17 | 6    | 0.80        | 4           |
| Girl         | 0.99        | 2           | 0.98 | 3    | 1           | 1           | 0.95 | 4    | 0.89 | 5    | 1           | 1           |
| Ironman      | 0.80        | 1           | 0.18 | 4    | 0.18        | 4           | 0.16 | 5    | 0.61 | 2    | 0.21        | 3           |
| Jogging1     | 0.97        | 2           | 0.98 | 1    | 0.98        | 1           | 0.98 | 1    | 0.98 | 1    | 0.98        | 1           |
| Jogging2     | 0.99        | 2           | 0.17 | 5    | 0.95        | 3           | 0.78 | 4    | 0.95 | 3    | 1           | 1           |
| Jumping      | 1           | 1           | 0.99 | 2    | 1           | 1           | 0.98 | 3    | 0.11 | 4    | 1           | 1           |
| Lemming      | 1           | 1           | 0.28 | 5    | 0.92        | 3           | 0.78 | 4    | 0.28 | 5    | 0.98        | 2           |
| Liquor       | 0.63        | 4           | 0.69 | 3    | 0.96        | 1           | 0.85 | 2    | 0.25 | 5    | 0.96        | 1           |
| Matrix       | 0.96        | 1           | 0.01 | 5    | 0.30        | 4           | 0.31 | 3    | 0.77 | 2    | 0.29        | 5           |
| Mhyang       | 1           | 1           | 1    | 1    | 1           | 1           | 0.93 | 3    | 0.97 | 2    | 1           | 1           |
| MotorRolling | 0.96        | 1           | 0.09 | 5    | 0.30        | 4           | 0.43 | 2    | 0.37 | 3    | 0.30        | 4           |
| MountainBike | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| Shaking      | 1           | 1           | 0.03 | 4    | 0.98        | 3           | 0.98 | 3    | 0.99 | 2    | 0.99        | 2           |
| Singer1      | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 0.99 | 2    | 1           | 1           |
| Singer2      | 0.99        | 2           | 1    | 1    | 0.33        | 4           | 1    | 1    | 0.36 | 3    | 0.36        | 3           |
| Skating1     | 0.69        | 2           | 0.73 | 1    | 0.60        | 3           | 0.21 | 6    | 0.52 | 5    | 0.57        | 4           |
| Skiing       | 1           | 1           | 0.07 | 5    | 0.12        | 4           | 0.74 | 2    | 0.14 | 3    | 0.14        | 3           |
| Soccer       | 0.90        | 1           | 0.27 | 2    | 0.22        | 5           | 0.24 | 4    | 0.24 | 4    | 0.25        | 3           |
| Subway       | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| Suv          | 0.89        | 3           | 0.98 | 1    | 0.98        | 1           | 0.98 | 1    | 0.97 | 2    | 0.97        | 2           |
| Sylvester    | 0.98        | 3           | 1    | 1    | 0.99        | 2           | 0.89 | 4    | 0.84 | 5    | 0.98        | 3           |
| Tiger1       | 0.96        | 1           | 0.89 | 4    | 0.95        | 2           | 0.61 | 6    | 0.94 | 3    | 0.84        | 5           |
| Tiger2       | 0.95        | 1           | 0.90 | 3    | 0.45        | 6           | 0.55 | 5    | 0.91 | 2    | 0.72        | 4           |
| Trellis      | 1           | 1           | 1    | 1    | 0.99        | 2           | 1    | 1    | 1    | 1    | 1           | 1           |
| Walking      | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 1    | 1    | 1           | 1           |
| Walking2     | 1           | 1           | 1    | 1    | 1           | 1           | 1    | 1    | 0.39 | 2    | 1           | 1           |
| Woman        | 1           | 1           | 0.94 | 4    | 0.94        | 4           | 0.99 | 2    | 0.93 | 5    | 0.98        | 3           |
| Average      | <b>0.94</b> | <b>1.45</b> | 0.79 | 2.27 | <b>0.85</b> | <b>1.90</b> | 0.81 | 2.49 | 0.76 | 2.62 | <b>0.85</b> | <b>1.90</b> |

SASR [39], and STRCF [13], which is highlighted in the red color font. In addition, the LADCF and STRCF trackers achieved the second-best results in terms of precision score and ranking position, which is indicated in the blue color font. Moreover, the average precision score and ranking position of each tracker are displayed in the last row of the table. Based on the above discussions, our proposed method demonstrates superior performance in terms of the precision score and ranking position.

#### 4.6.3. Experiments on OTB-2015 dataset

We conduct an experiment on the OTB-2015 dataset, which consists of 100 challenging video sequences. Specifically, the proposed approach obtained better efficiency on the OTB-2015 dataset, which is exhibited in Fig. 10 and Table 4. As illustrated in Table 4, we observe that our presented approach has obtained excellent efficiency with a DP score (92.2%) and AUC score (69.2%), which is better than

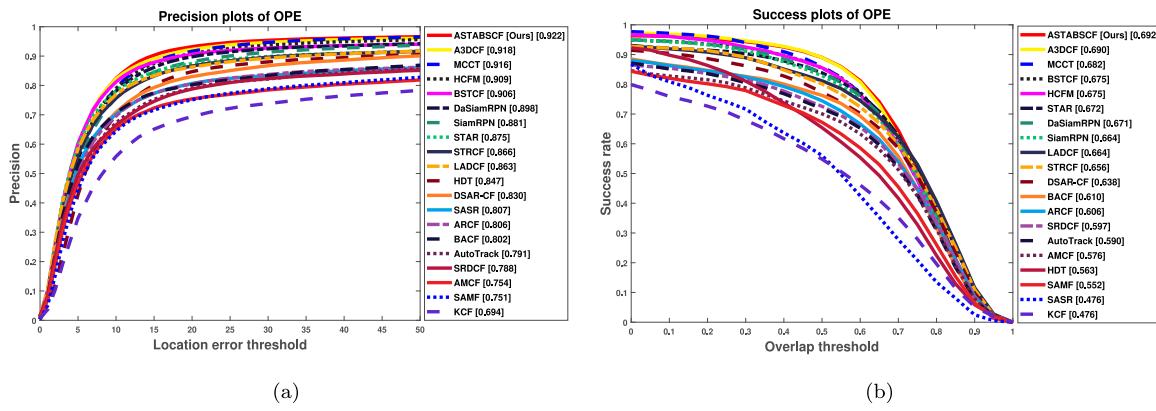


Fig. 10. The distance precision and overlap success results of the OTB-2015 dataset are exhibited in Fig. 10(a) and Fig. 10(b).

Table 4

| Comparative results of conventional trackers with the proposed tracker on the OTB-2015 dataset with 100 sequences. The results of the top 15 trackers are presented as follows:: |         |                   |                    |                   |                   |                    |                        |                      |            |             |            |                  |            |                   |             |             |
|--|---------|-------------------|--------------------|-------------------|-------------------|--------------------|------------------------|----------------------|------------|-------------|------------|------------------|------------|-------------------|-------------|-------------|
| Methods  | Metrics | Ours <sup>a</sup> | A3DCF <sup>a</sup> | MCCT <sup>a</sup> | HCFM <sup>a</sup> | BSTCF <sup>a</sup> | DaSiamRPN <sup>a</sup> | SiamRPN <sup>a</sup> | STAR       | STRCF       | LADCF      | HDT <sup>a</sup> | DSAR-CF    | SASR <sup>a</sup> | ARCF        | BACF        |
| OTB-2015   | DP      | <b>92.2</b>       | 91.8               | 91.6              | 90.9              | 90.6               | 89.8                   | 88.1                 | 87.5       | 86.6        | 86.3       | 84.7             | 83.0       | 80.7              | 80.6        | 80.2        |
|  | AUC     | <b>69.2</b>       | 69.0               | 68.2              | 67.5              | 67.5               | 67.1                   | 66.4                 | 67.2       | 65.6        | 66.4       | 56.3             | 63.8       | 47.6              | 60.6        | 61.0        |
|  | FPS     | <b>10.2</b>       | <b>4.2</b>         | <b>7.8</b>        | <b>2.3</b>        | <b>15.6</b>        | <b>160.0</b>           | <b>160.0</b>         | <b>2.5</b> | <b>24.3</b> | <b>0.9</b> | <b>2.7</b>       | <b>6.1</b> | —                 | <b>10.2</b> | <b>26.7</b> |

<sup>a</sup> Denotes the hand-crafted and deep features-based trackers.

Table 5

Attribute-based evaluation of the proposed approach on the OTB-2015 dataset (DP & AUC Scores). The best results are denoted in red, blue, and green fonts.

| Methods | Ours             | HCFM             | BSTCF            | STAR              | STRCF     | DSAR-CF   | BACF      | HDT       | LADCF     | SASR      | SRDCF     | ARCF      | AMCF      | KCF       | SAMF      |
|---------|------------------|------------------|------------------|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| IV      | <b>93.8/71.7</b> | <b>89.1/69.0</b> | 88.4/67.6        | 86.0/ <b>68.2</b> | 83.8/65.2 | 81.3/64.8 | 80.9/63.0 | 80.5/64.6 | 79.1/48.6 | 78.8/61.0 | 76.0/59.6 | 75.1/58.6 | 81.7/53.2 | 71.5/47.5 | 71.3/53.1 |
| SV      | <b>91.3/67.1</b> | 87.6/63.6        | <b>88.5/65.2</b> | 84.0/ <b>63.9</b> | 84.3/63.6 | 82.4/62.7 | 74.3/55.9 | 83.5/63.8 | 77.5/42.5 | 74.4/56.2 | 76.8/56.0 | 69.6/53.6 | 80.6/48.6 | 63.3/39.3 | 70.5/49.7 |
| OCC     | <b>88.5/66.4</b> | 86.4/64.0        | <b>88.3/66.6</b> | 83.6/ <b>64.5</b> | 82.6/62.3 | 77.1/60.3 | 69.7/54.8 | 82.4/63.9 | 73.2/38.7 | 72.5/54.9 | 73.1/55.1 | 68.4/52.4 | 76.6/51.7 | 61.7/42.9 | 71.9/53.0 |
| DEF     | <b>89.6/65.7</b> | <b>89.5/63.8</b> | 87.5/63.7        | 83.8/62.8         | 84.1/60.5 | 78.9/59.0 | 76.3/57.7 | 81.3/59.7 | 76.7/44.1 | 72.9/54.0 | 76.5/58.0 | 68.5/52.2 | 81.9/54.0 | 61.1/43.1 | 68.4/50.6 |
| MB      | <b>91.5/69.9</b> | <b>88.0/68.0</b> | 86.1/67.2        | 82.3/66.8         | 84.1/67.0 | 81.3/64.4 | 73.1/57.6 | 81.4/65.9 | 76.3/45.6 | 77.6/60.7 | 76.8/61.5 | 72.6/57.6 | 78.9/56.0 | 61.0/45.0 | 67.4/53.0 |
| FM      | <b>89.9/67.9</b> | 87.5/64.9        | 84.1/64.6        | 81.3/ <b>65.1</b> | 79.5/62.9 | 78.4/61.4 | 75.8/58.7 | 77.8/62.0 | 77.9/46.8 | 75.7/59.2 | 75.9/58.8 | 70.0/57.0 | 79.8/54.7 | 61.4/44.4 | 65.4/50.1 |
| OPR     | <b>92.7/68.1</b> | <b>89.9/65.6</b> | 89.8/65.5        | 87.3/65.3         | 85.5/62.7 | 81.0/61.1 | 77.1/57.7 | 83.4/63.0 | 79.3/43.7 | 74.1/54.6 | 77.1/55.7 | 71.3/52.8 | 80.6/52.9 | 67.5/44.9 | 73.7/53.2 |
| IPR     | <b>94.2/67.3</b> | <b>91.8/65.3</b> | 87.2/63.1        | 83.8/62.1         | 80.9/59.8 | 77.9/57.8 | 75.7/56.3 | 80.1/59.6 | 78.8/44.8 | 73.5/53.5 | 77.8/55.4 | 74.9/55.0 | 84.0/54.7 | 69.2/46.0 | 71.4/51.0 |
| OV      | 84.4/61.8        | <b>87.8/65.6</b> | <b>87.5/65.4</b> | 77.0/60.0         | 75.7/57.8 | 72.7/57.5 | 70.9/54.0 | 82.8/63.0 | 72.2/35.4 | 62.4/48.9 | 69.6/52.9 | 67.3/53.5 | 68.6/50.0 | 53.4/42.9 | 65.2/50.9 |
| BC      | <b>89.6/67.2</b> | <b>91.4/68.7</b> | 88.0/65.7        | 85.9/ <b>66.0</b> | 87.3/64.8 | 81.5/63.7 | 77.8/60.0 | 84.4/64.9 | 79.1/48.1 | 77.5/58.3 | 76.0/58.9 | 73.8/57.5 | 84.4/57.8 | 71.3/49.8 | 68.9/52.5 |
| LR      | <b>94.4/66.1</b> | <b>89.4/62.6</b> | 87.3/63.3        | 81.2/60.3         | 75.6/55.8 | 75.8/57.1 | 69.0/50.6 | 75.4/56.5 | 63.1/48.0 | 71.4/49.4 | 56.6/42.6 | 76.6/42.0 | 54.6/30.7 | 71.0/45.9 |           |

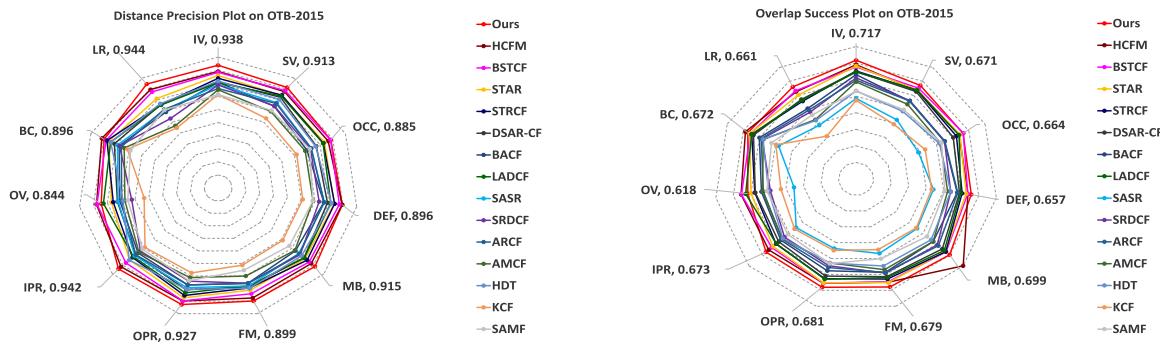


Fig. 11. Attribute-based distance precision and overlap success score plots on OTB-2015 dataset.

the other conventional trackers. Moreover, the efficiency of the presented approach is compared with 11 Attributes as shown in Fig. 11, and comparisons of the precision and success scores are illustrated in Table 5.

In addition, the comparative performance of the ASTABSCF tracker with the other five modern trackers is exhibited in Fig. 16. From these results, we ensure that the presented approach obtained the best results by achieving DP and AUC scores (12.0%/8.2%) compared to the baseline [2] tracker. Furthermore, when compared to hand-crafted feature-based approaches such as STRCF [13], LADCF [12], and SRDCF [11], and ARCF [40], we ensure that our presented method improves the better results in DP and AUC scores (5.6%/3.6%), (5.9%/2.8%),

(13.4%/9.5%), and (11.6%/8.6%), respectively. Likewise, when compared to the deep learning-based approaches like DaSiamRPN [54], SiamRPN [55] HCFM [28], BSTCF [41], STAR [42], and SASR [39], we observe that our presented approach attains better enhancement in precision and success scores (2.4%/2.1%), (4.1%/2.8%), (1.3%/1.7%), (1.6%/1.7%), (4.7%/2.0%), and (11.5%/21.6%), on the OTB-2015 dataset, respectively. Finally, we can conclude that from this deep analysis, our proposed tracker obtains better tracking efficiency when compared to handcrafted and deep feature-based approaches.

#### 4.6.4. Evaluation on TC-128 dataset

We perform an experimental evaluation on the TC-128 dataset to evaluate the tracking effectiveness of our proposed technique. The

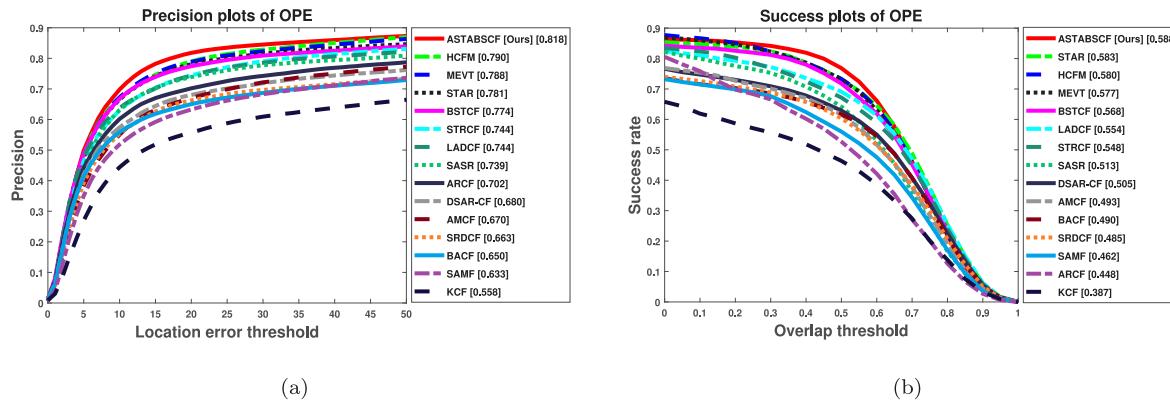


Fig. 12. The distance precision and overlap success results of the TC-128 dataset are exhibited in Fig. 12(a) and Fig. 12(b).

Table 6

Comparative results of conventional trackers with the proposed approach on the TC-128 dataset.

| Methods | Metrics | Ours        | HCFM | MEVT | STAR | BSTCF | STRCF | LADCF | SASR | ARCF | DSAR-CF | AMCF | SRDCF | BACF | SAMF | KCF  |
|---------|---------|-------------|------|------|------|-------|-------|-------|------|------|---------|------|-------|------|------|------|
| TC-128  | DP      | <b>81.8</b> | 79.0 | 78.8 | 78.1 | 77.4  | 74.4  | 74.4  | 73.9 | 70.2 | 68.0    | 67.0 | 66.3  | 65.0 | 63.3 | 55.8 |
|         | AUC     | <b>58.8</b> | 58.0 | 57.7 | 58.3 | 56.8  | 54.8  | 55.4  | 51.3 | 44.8 | 50.5    | 49.3 | 48.5  | 49.0 | 46.2 | 38.7 |

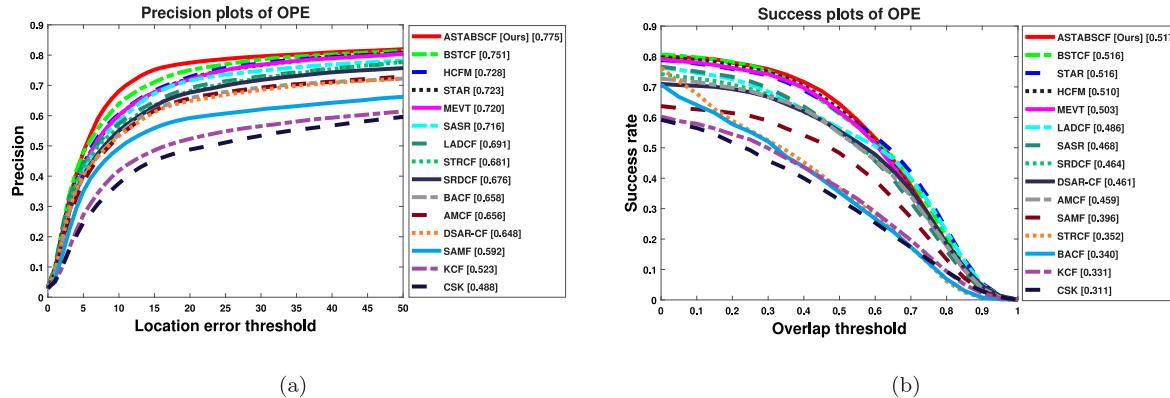


Fig. 13. The distance precision and overlap success results of the UAV-123 dataset are exhibited in Fig. 13(a) and Fig. 13(b).

experimental results are illustrated in Fig. 12 and Table 6. From these results, we ensure that the presented approach achieved better results in the DP and AUC scores (81.8%/58.8%), which shows a better tracking performance than the other conventional methods.

Particularly, when compared to the baseline [2] method, we observe that the presented ASTABSCF method enhances the best improvement in the DP and AUC scores (16.8%/9.8%). Specifically, compared to hand-crafted feature-based approaches like SRDCF [11], DSAR-CF [25], STRCF [13], and LADCF [12], our proposed tracker obtains excellent improvement in DP and AUC scores (15.5%/10.3%), (13.8%/8.3%), (7.4%/4.0%) and (7.4%/3.4%), respectively. Moreover, we show that our presented method exceeds the other deep learning-based approaches like SASR [39], BSTCF [41], STAR [42], MEVT [19], and HCFM [28], obtains great results in DP/AUC scores (7.9%/7.5%), (4.4%/2.0%), (3.7%/0.5%), (3.0%/1.1%), and (2.8%/0.8%), respectively. In the end, we confirm that the presented method outperforms well compared to hand-crafted and deep learning-based trackers.

#### 4.6.5. Evaluation on UAV-123 dataset

We conduct an experimental evaluation of the UAV-123 dataset to estimate the tracking performance of our presented technique. Specifically, the proposed ASTABSCF approach comparison results are illustrated in Fig. 13 and Table 7. Further, we observed from Table 7 that our proposed approach has obtained better performance in DP and AUC scores compared to other modern trackers.

From these analyses, we observe that our presented method increases the better results in precision and success scores of (77.5%/51.7%). Specifically, compared to the baseline [2] method, we ensure that the presented ASTABSCF technique enhances the tracking performance with DP and AUC scores of (11.7%/17.7%). Furthermore, compared to hand-crafted feature-based methods such as DSAR-CF [25], SRDCF [11], STRCF [13], and LADCF [12], we ensure that our presented technique achieves the better improvement in the DP and AUC scores (12.7%/5.6%), (9.9%/5.3%), (9.4%/16.5%), and (8.4%/3.1%), respectively. Furthermore, we confirm that our presented method outperforms well compared to deep learning-based approaches such as SASR [39], MEVT [19], STAR [42], HCFM [28], and BSTCF [41] obtains best improvements in DP/AUC scores (5.9%/4.9%), (5.5%/1.4%), (5.2%/0.1%), (4.7%/0.7%), and (2.4%/0.1%).

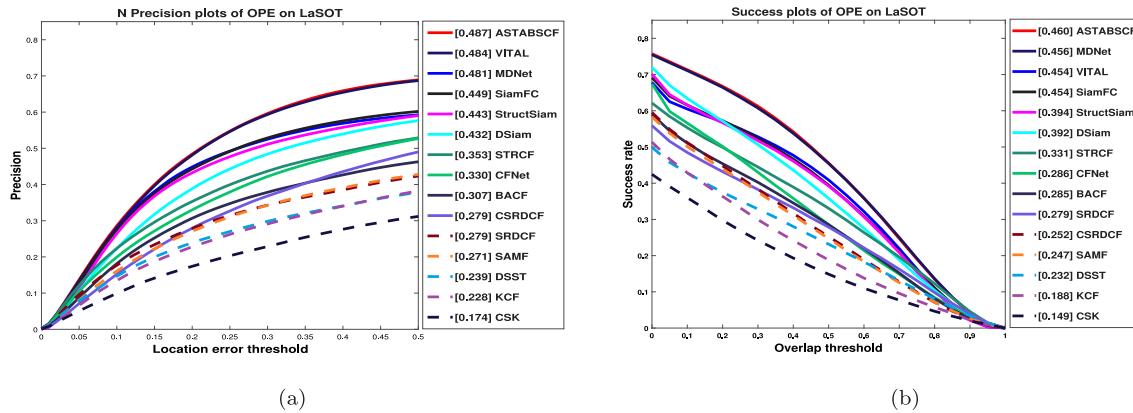
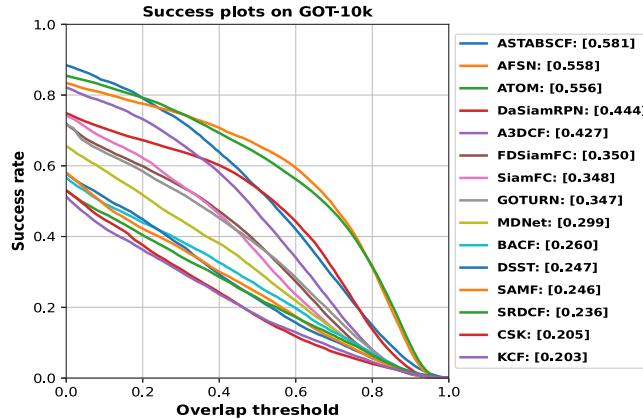
#### 4.6.6. Evaluation on LaSOT dataset

We conduct experiments on the LaSOT dataset to evaluate the proposed tracker performance. For comparison purposes, we used 15 modern trackers with the presented approach on the LaSOT dataset. Also, the LaSOT dataset is a large-scale dataset, which contains 1400 video sequences with 70 different object categories and each category has 20 sequences. In this work, we used the 1400 sequences for our tracker evaluation. Moreover, the experimental results are exhibited in Fig. 14. As illustrated in Fig. 14(a) and Fig. 14(b), we can see that the

**Table 7**

Comparative results of conventional trackers with the proposed approach on the UAV-123 dataset.

| Methods | Metrics | Ours        | BSTCF | HCFM | STAR | MEVT | SASR | LADCF | STRCF | SRDCF | BACF | AMCF | DSAR-CF | SAMF | KCF  | CSK  |
|---------|---------|-------------|-------|------|------|------|------|-------|-------|-------|------|------|---------|------|------|------|
| UAV-123 | DP      | <b>77.5</b> | 75.1  | 72.8 | 72.3 | 72.0 | 71.6 | 69.1  | 68.1  | 67.6  | 65.8 | 65.6 | 64.8    | 59.2 | 52.3 | 48.8 |
|         | AUC     | <b>51.7</b> | 51.6  | 51.0 | 51.6 | 50.3 | 46.8 | 48.6  | 35.2  | 46.4  | 34.0 | 45.9 | 46.1    | 39.6 | 33.1 | 31.1 |

**Fig. 14.** The precision and success plots of the proposed method on the LaSOT dataset are exhibited in Fig. 14(a) and Fig. 14(b), respectively.**Fig. 15.** The success plots of the state-of-the-art trackers on the GOT-10K dataset.

presented approach obtains remarkable tracking efficiency in terms of a precision score of 48.7% and success score of 46.0%, respectively.

Additionally, when compared to the baseline method, we ensure that our presented method enhances the better performance in DP/AUC scores of (18.0% / 17.5%). Furthermore, compared to hand-crafted feature-based trackers such as STRCF [13], CSRDCF [50], SRDCF [11], SAMF [44], and DSST [51], we ensure that our proposed technique improves the significant tracking performance in terms of DP and AUC scores of (13.4%/12.9%), (20.8%/20.8%), (20.8%/18.1%), (21.6%/21.3%), and (24.8%/22.8%), respectively. Specifically, compared to deep learning-based approaches such as VITAL [45], MDNet [38], SiamFC [46], StructSiam [47], and CFNet [49], we confirm that our presented tracking approach increases the superior performance in DP/AUC scores of (0.3%/0.6%), (0.6%/0.4%), (3.8%/0.6%), (4.4%/6.6%), and (15.7%/17.4%) respectively. From these analyses, we confirm that our presented method outperforms well compared to the hand-crafted and deep learning-based approaches.

#### 4.6.7. Experiments on GOT-10K dataset

We also estimate our proposed approach on the GOT-10K large-scale benchmark dataset, which contains more than 10000 video sequences.

Following that, we employ the 180 test sequence with 84 different objects to evaluate our proposed tracker performance. Moreover, we evaluate the success rate metric on the GOT-10K dataset, which achieves significant performance in the AUC score of 58.1%, as illustrated in Fig. 15.

Besides, when compared to the baseline method [2], we ensure that our tracking approach increases better tracking efficiency in the AUC score of 32.1%. Specifically, we confirm that when compared to the hand-crafted feature-based methods such as DSST [51], SAMF [44], CSK [34], and KCF [24] our presented approach improves remarkable tracking performance in the success rate 33.4%, 33.5%, 37.6%, and 37.8%, respectively. Furthermore, when compared to deep feature-based trackers such as DaSiamRPN [54], ATOM [53], FDSiamFC [56], SiamFC [46], and GOTURN [57], we can see that our tracker obtains significant results in AUC scores 13.7%, 2.5%, 23.1%, 34.8%, and 23.4%, respectively. From these hand-crafted and deep feature analyses, we confirmed that our proposed ASTABSCF tracker demonstrates superior efficiency on the GOT-10K benchmark dataset.

## 5. Discussion

The experimental results demonstrate the efficacy of our proposed tracking method, which integrates adaptive spatially regularized filtering, target-aware background suppression, and an attribute-aware approach. The improved tracking accuracy can be attributed to the robust feature fusion strategy, which leverages both deep learning-based and handcrafted features such as HOG, ColorName, and ResNet-50. This combination enhances the discriminative power of the model, ensuring the effective separation of the target object from the background. The performance gains observed in precision and success scores indicate that the proposed framework effectively mitigates challenges such as occlusion, scale variations, and background clutter. Moreover, a detailed analysis of the tracking performance suggests that the background suppression technique plays a significant role in reducing false positives, particularly in complex environments. The spatial regularization technique further refines the correlation filter by penalizing irrelevant background features, thereby improving robustness. Additionally, the adaptive attribute-aware approach dynamically optimizes filter responses by suppressing inconsistent or redundant features during training. These findings collectively highlight the importance



**Fig. 16.** Comparison of ASTABSCF tracker with other five modern trackers such as BACF [2], BSTCF [41], SRDCF [11], SASR [39], and STRCF [13] in seven various challenging sequences in OTB-2015 dataset. The main sequence are from top to bottom Bird1, Biker, Liquor, DragonBaby, Shaking, FaceOcc1, Soccer.

of integrating spatial awareness, background suppression, and feature adaptation to achieve superior tracking performance compared to conventional methods.

## 6. Conclusion

In this study, a novel adaptive spatially regularized background-suppressed target attribute-aware deep correlation filter has been presented. In this regard, the spatially regularized approach has been proposed that learns spatial weight for a specific object to effectively find the estimated target during the tracking process. Following that, a target-aware background-suppressed regularization tracking strategy has been presented to address the tracking issues caused by distractors in the response map. In addition, the target-aware background-suppressed technique incorporated dual regression, which produced the target and global filters to obtain a more accurate performance. Specifically, an adaptive attribute-aware approach has been introduced to recognize the feature information in each channel, which performs post-processing operations in the filter training stage to remove inappropriate and inconsistent feature channels. Finally, the experimental results on the benchmark datasets OTB-2013, OTB-2015, TempleColor-128, UAV-123, LaSOT, and GOT-10K have demonstrated the superior performance of our proposed approach over other state-of-the-art-trackers.

Although the ASTABSCF tracker improves tracking performance, it suffers from severe occlusion and fast-moving objects. These difficulties will be taken into account in future work to improve tracking performance.

## CRediT authorship contribution statement

**Sathiyamoorthi Arthanari:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Conceptualization. **Sathishkumar Moorthy:** Investigation, Supervision, Validation, Visualization, Writing – review & editing. **Jae Hoon Jeong:** Writing – review & editing, Validation, Supervision, Investigation. **Young Hoon Joo:** Validation, Supervision, Investigation, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported in part by the Basic Science Research Program under Grant NRF-2016R1A6A1A03013567 and Grant NRF-2021R1A2B5B01001484 and by the framework of the International Cooperation Program under Grant NRF-2022K2A9A2A06045121 through the National Research Foundation of Korea (NRF) funded by the Ministry of Education.

## Data availability

Data will be made available on request.

## References

- [1] D. Elayaperumal, Y.H. Joo, Aberrance suppressed spatio-temporal correlation filters for visual object tracking, *Pattern Recognit.* 115 (2021) 107922.
- [2] H. Kiani Galoogahi, A. Fagg, S. Lucey, Learning background-aware correlation filters for visual tracking, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1135–1143.
- [3] X. Yang, Y. Song, Z. Zhang, Y. Zhao, L. Li, W. Li, Robust correlation filter tracking based on response map analysis network, *Signal Process., Image Commun.* 108 (2022) 116768.
- [4] S. Arthanari, J.H. Jeong, Y.H. Joo, Exploring multi-level transformers with feature frame padding network for 3D human pose estimation, *Multimedia Syst.* 30 (5) (2024) 243.
- [5] S. Arthanari, J.H. Jeong, Y.H. Joo, Exploiting multi-transformer encoder with multiple-hypothesis aggregation via diffusion model for 3D human pose estimation, *Multimedia Tools Appl.* (2024) 1–29.
- [6] X. Yang, S. Li, J. Ma, J.-y. Yang, J. Yan, Co-saliency-regularized correlation filter for object tracking, *Signal Process., Image Commun.* 103 (2022) 116655.
- [7] Z. Peng, X. Lu, Learning region sparse constraint correlation filter for tracking, *Signal Process., Image Commun.* 90 (2021) 116042.

- [8] A. Sezer, A. Altan, Detection of solder paste defects with an optimization-based deep learning model using image processing techniques, *Solder. Surf. Mount Technol.* 33 (5) (2021) 291–298.
- [9] İ. Yağ, A. Altan, Artificial intelligence-based robust hybrid algorithm design and implementation for real-time detection of plant diseases in agricultural environments, *Biology* 11 (12) (2022) 1732.
- [10] Y.B. Özçelik, A. Altan, Overcoming nonlinear dynamics in diabetic retinopathy classification: a robust AI-based model with chaotic swarm intelligence optimization and recurrent long short-term memory, *Fractal Fract.* 7 (8) (2023) 598.
- [11] M. Danelljan, G. Hager, F. Shahbaz Khan, M. Felsberg, Learning spatially regularized correlation filters for visual tracking, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 4310–4318.
- [12] T. Xu, Z.-H. Feng, X.-J. Wu, J. Kittler, Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking, *IEEE Trans. Image Process.* 28 (11) (2019) 5596–5609.
- [13] F. Li, C. Tian, W. Zuo, L. Zhang, M.-H. Yang, Learning spatial-temporal regularized correlation filters for visual tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4904–4913.
- [14] M. Mueller, N. Smith, B. Ghanem, Context-aware correlation filter tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1396–1404.
- [15] D. Elayaperumal, Y.H. Joo, Learning spatial variance-key surrounding-aware tracking via multi-expert deep feature fusion, *Inform. Sci.* 629 (2023) 502–519.
- [16] S.S. Kuppusami Sakthivel, S. Moorthy, S. Arthanari, J.H. Jeong, Y.H. Joo, Learning a context-aware environmental residual correlation filter via deep convolution features for visual object tracking, *Mathematics* 12 (14) (2024) 2279.
- [17] S.S. KS, Y.H. Joo, J.H. Jeong, Keypoint prediction enhanced siamese networks with attention for accurate visual object tracking, *Expert Syst. Appl.* 268 (2025) 126237.
- [18] X. Sun, G. Han, L. Guo, H. Yang, X. Wu, Q. Li, Two-stage aware attentional siamese network for visual tracking, *Pattern Recognit.* 124 (2022) 108502.
- [19] S. Moorthy, Y.H. Joo, Multi-expert visual tracking using hierarchical convolutional feature fusion via contextual information, *Inform. Sci.* 546 (2021) 996–1013.
- [20] K. Nai, Z. Li, H. Wang, Dynamic feature fusion with spatial-temporal context for robust object tracking, *Pattern Recognit.* 130 (2022) 108775.
- [21] P. Sun, J. Cao, Y. Jiang, R. Zhang, E. Xie, Z. Yuan, C. Wang, P. Luo, Trantrack: Multiple object tracking with transformer, 2020, arXiv preprint [arXiv:2012.15460](https://arxiv.org/abs/2012.15460).
- [22] T. Meinhartd, A. Kirillov, L. Leal-Taixe, C. Feichtenhofer, Trackformer: Multi-object tracking with transformers, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8844–8854.
- [23] S. Moorthy, S.S. KS, S. Arthanari, J.H. Jeong, Y.H. Joo, Hybrid multi-attention transformer for robust video object detection, *Eng. Appl. Artif. Intell.* 139 (2025) 109606.
- [24] J.F. Henriques, R. Caseiro, P. Martins, J. Batista, High-speed tracking with kernelized correlation filters, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (3) (2014) 583–596.
- [25] W. Feng, R. Han, Q. Guo, J. Zhu, S. Wang, Dynamic saliency-aware regularization for correlation filter-based object tracking, *IEEE Trans. Image Process.* 28 (7) (2019) 3232–3245.
- [26] S. Li, Y. Liu, Q. Zhao, Z. Feng, Learning residue-aware correlation filters and refining scale for real-time UAV tracking, *Pattern Recognit.* 127 (2022) 108614.
- [27] Y. Qi, S. Zhang, L. Qin, H. Yao, Q. Huang, J. Lim, M.-H. Yang, Hedged deep tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4303–4311.
- [28] J. Zhang, Y. Liu, H. Liu, J. Wang, Y. Zhang, Distractor-aware visual tracking using hierarchical correlation filters adaptive selection, *Appl. Intell.* 52 (6) (2022) 6129–6147.
- [29] L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi, P.H. Torr, Fully-convolutional siamese networks for object tracking, in: European Conference on Computer Vision, Springer, 2016, pp. 850–865.
- [30] L. Liu, T. Feng, Y. Fu, Learning multifeature correlation filter and saliency redetection for long-term object tracking, *Symmetry* 14 (5) (2022) 911.
- [31] B. Huang, T. Xu, Z. Shen, S. Jiang, J. Li, BSCF: Learning background suppressed correlation filter tracker for wireless multimedia sensor networks, *Ad Hoc Netw.* 111 (2021) 102340.
- [32] C. Guo, Q. Ma, L. Zhang, Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8.
- [33] J. Sherman, W.J. Morrison, Adjustment of an inverse matrix corresponding to a change in one element of a given matrix, *Ann. Math. Stat.* 21 (1) (1950) 124–127.
- [34] J.F. Henriques, R. Caseiro, P. Martins, J. Batista, Exploiting the circulant structure of tracking-by-detection with kernels, in: Proceedings of the European Conference on Computer Vision, ECCV, Springer, 2012, pp. 702–715.
- [35] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, H. Li, Multi-cue correlation filters for robust visual tracking, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4844–4853.
- [36] X.-F. Zhu, X.-J. Wu, T. Xu, Z.-H. Feng, J. Kittler, Robust visual object tracking via adaptive attribute-aware discriminative correlation filters, *IEEE Trans. Multimed.* 24 (2021) 301–312.
- [37] Y. Li, C. Fu, F. Ding, Z. Huang, G. Lu, AutoTrack: Towards high-performance visual tracking for UAV with automatic spatio-temporal regularization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11923–11932.
- [38] H. Nam, B. Han, Learning multi-domain convolutional neural networks for visual tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4293–4302.
- [39] C. Fu, W. Xiong, F. Lin, Y. Yue, Surrounding-aware correlation filter for UAV tracking with selective spatial regularization, *Signal Process.* 167 (2020) 107324.
- [40] Z. Huang, C. Fu, Y. Li, F. Lin, P. Lu, Learning aberrance repressed correlation filters for real-time UAV tracking, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 2891–2900.
- [41] J. Zhang, Y. He, W. Feng, J. Wang, N.N. Xiong, Learning background-aware and spatial-temporal regularized correlation filters for visual tracking, *Appl. Intell.* (2022) 1–16.
- [42] L. Xu, P. Kim, M. Wang, J. Pan, X. Yang, M. Gao, Spatio-temporal joint aberrance suppressed correlation filter for visual tracking, *Complex Intell. Syst.* (2021) 1–13.
- [43] Y. Li, C. Fu, F. Ding, Z. Huang, J. Pan, Augmented memory for correlation filters in real-time UAV tracking, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2020, pp. 1559–1566.
- [44] Y. Li, J. Zhu, A scale adaptive kernel correlation filter tracker with feature integration, in: Proceedings of the European Conference on Computer Vision, ECCV, Springer, 2014, pp. 254–265.
- [45] Y. Song, C. Ma, X. Wu, L. Gong, L. Bao, W. Zuo, C. Shen, R.W. Lau, M.-H. Yang, Vital: Visual tracking via adversarial learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8990–8999.
- [46] L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi, P.H. Torr, Fully-convolutional siamese networks for object tracking, in: Computer Vision-ECCV 2016 Workshops: Amsterdam, the Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part II 14, Springer, 2016, pp. 850–865.
- [47] Y. Zhang, L. Wang, J. Qi, D. Wang, M. Feng, H. Lu, Structured siamese network for real-time visual tracking, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 351–366.
- [48] Q. Guo, W. Feng, C. Zhou, R. Huang, L. Wan, S. Wang, Learning dynamic siamese network for visual object tracking, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1763–1771.
- [49] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, P.H. Torr, End-to-end representation learning for correlation filter-based tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2805–2813.
- [50] A. Lukezic, T. Vojir, L. Čehovin Zajc, J. Matas, M. Kristan, Discriminative correlation filter with channel and spatial reliability, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6309–6318.
- [51] M. Danelljan, G. Häger, F.S. Khan, M. Felsberg, Discriminative scale space tracking, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (8) (2016) 1561–1575.
- [52] S. Peng, Y. Yu, K. Wang, L. He, Accurate anchor free tracking, 2020, arXiv preprint [arXiv:2006.07560](https://arxiv.org/abs/2006.07560).
- [53] M. Danelljan, G. Bhat, F.S. Khan, M. Felsberg, Atom: Accurate tracking by overlap maximization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4660–4669.
- [54] Z. Zhu, Q. Wang, B. Li, W. Wu, J. Yan, W. Hu, Distractor-aware siamese networks for visual object tracking, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 101–117.
- [55] B. Li, J. Yan, W. Wu, Z. Zhu, X. Hu, High performance visual tracking with siamese region proposal network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8971–8980.
- [56] H. Huang, G. Liu, Y. Zhang, R. Xiong, Feature distillation siamese networks for object tracking, *Appl. Soft Comput.* 132 (2023) 109912.
- [57] D. Held, S. Thrun, S. Savarese, Learning to track at 100 fps with deep regression networks, in: Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer, 2016, pp. 749–765.