

Article

Interval Spatio-Temporal Constraints and Pixel-Spatial Hierarchy Region Proposals for Abrupt Motion Tracking

Daxiang Suo * and Xueling Lv

College of Management and Economics, Tianjin University, Tianjin 300072, China

* Correspondence: suodaxiang@tju.edu.cn

Abstract: The RPN-based Siamese tracker has achieved remarkable performance with real-time speed but suffers from a lack of robustness in complex motion tracking. Especially when the target comes into an abrupt motion scenario, the assumption of motion smoothness may be broken, which will further compromise the reliability of tracking results. Therefore, it is important to develop an adaptive tracker that can maintain robustness in complex motion scenarios. This paper proposes a novel tracking method based on the interval spatio-temporal constraints and a region proposal method over a pixel-spatial hierarchy. Firstly, to cope with the limitations of a fixed-constraint strategy for abrupt motion tracking, we propose a question-guided interval spatio-temporal constraint strategy. Based on the consideration of tracking status and the degree of penalty expansion, it enables the dynamic adjustment of the constraint weights, which ensures a match between response scores and true confidence values. Secondly, to guarantee the coverage of a target using candidate proposals in extreme motion scenarios, we propose a region proposal method over the pixel-spatial hierarchy. By combining visual common sense with reciprocal target-distractor information, our method implements a careful refinement of the primary proposals. Moreover, we introduce a discriminative-enhanced memory updater designed to ensure effective model adaptation. Comprehensive evaluations on five benchmark datasets: OTB100, UAV123, LaSOT, VOT2016, and VOT2018 demonstrate the superior performance of our proposed method in comparison to several state-of-the-art approaches.

Keywords: visual object tracking; memory model; spatio-temporal constraint; region proposal



Citation: Suo, D.; Lv, X. Interval Spatio-Temporal Constraints and Pixel-Spatial Hierarchy Region Proposals for Abrupt Motion Tracking. *Electronics* **2024**, *13*, 4084. <https://doi.org/10.3390/electronics13204084>

Academic Editor: Stefanos Kollias

Received: 21 August 2024

Revised: 23 September 2024

Accepted: 26 September 2024

Published: 17 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Visual tracking is a crucial area within computer vision, focused on locating the position of a given target throughout a video sequence. Visual tracking is essential across various domains, including video surveillance, vehicle navigation, et al. Along with the development of tracking methods such as discriminant correlation filter (DCF)-based solutions [1,2] and convolution regression-based solutions [3,4], the trackers have achieved satisfactory performance. However, most existing trackers work upon the assumption of spatio-temporal continuity, assuming that the position of a target changes smoothly. However, these trackers may perform poorly when the target is in abrupt motion, as shown in Figure 1.

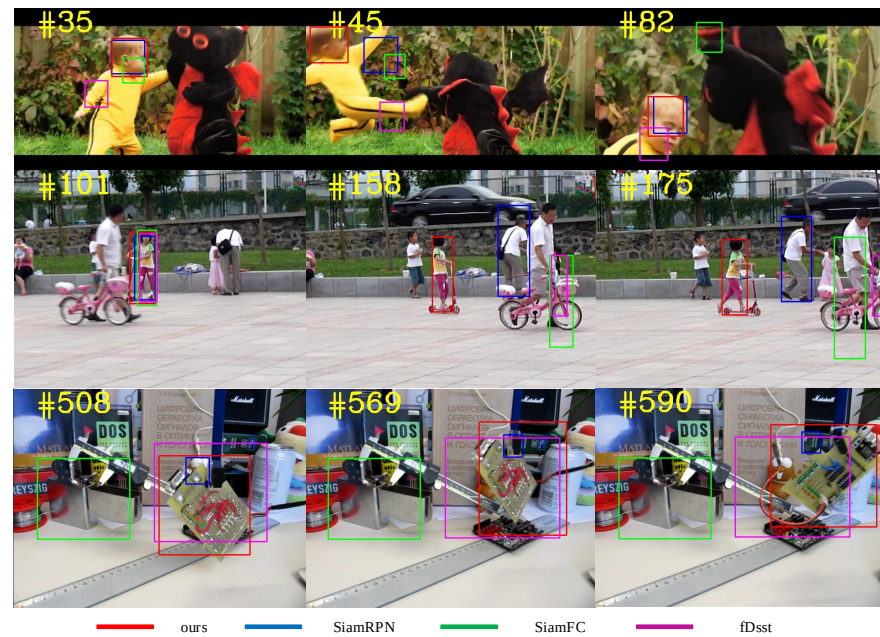


Figure 1. Compared with fDsst [5], SiamFC [6], and SiamRPN [7], our tracking algorithms perform favorably under abrupt motion.

When the tracking target is under the abrupt motion scenario, the spatio-temporal constraint strategy founded on the smoothness of motion is broken. Thus, the probability of the target occurring at each spot is not proportional to the distance of this spot from the target location in the previous frame. However, the conventional spatial-temporal constraints tend to impose an incorrect penalty term on the target confidence value, making it difficult for memory models to recover tracking. One intuitive solution to address this problem is to reconfigure the spatio-temporal constraint strategy. For instance, Zhang et al. [8] introduce a multi-object tracker leveraging spatio-temporal topological (STAT) constraints aimed at addressing the aforementioned challenges. They developed the feature adaptive association module (FAAM), which regionally associates motion with appearance, effectively merging these features. Furthermore, Zhang et al. [9] employ a spatio-temporal relationship learning module to extract relational features from historical trajectories, enabling the modeling of object correlations and the dynamic construction of online group structures. Recently, Xu et al. [10] proposed an online updating method that models appearance within an affine subspace, facilitating joint discriminative learning, which enhances the discrimination and interpretability of the filters.

An appropriate spatio-temporal constraint strategy can reduce the possibility of tracking failures occurring, but it may fail when dealing with the randomness of target position variations due to the working principle, which is based on imposing penalty terms on the response map to suppress the scores of similar objects. The randomness of target position changes is enhanced in abrupt motion scenarios, so the candidate proposals generated by traditional region proposal methods are difficult in terms of ensuring the coverage of the target, which is the underlying reason for the reduced accuracy of the tracking results. In order to ensure robust tracking results in complex motion scenarios, it is meaningful to develop an effective region proposal method. With the advancement of deep learning, region proposal networks (RPNs) [7] have been introduced to predict target boundaries and scores. However, these methods often generate numerous low-quality region proposals in challenging backgrounds, which hampers the tracker's ability to quickly and accurately capture the target.

Overall, significant efforts have been made to enhance tracking algorithm performance, leading to the development of many high-performing methods. However, the algorithm discussed still has several limitations: (1) The spatio-temporal constraint strategy may

incorrectly suppress the true target response, especially when the assumption of motion smoothness is not satisfied; (2) Many trackers struggle with target localization due to ineffective region proposal methods, particularly when the target exhibits highly erratic motion; (3) Current updating techniques face challenges in efficiently handling abrupt motion, often due to inadequate training data and suboptimal loss functions.

In view of the tracking difficulty in complex motion scenarios, we devise a tracking algorithm with three components: the question-guided interval spatio-temporal constraint strategy, a region proposal method over the pixel-spatial hierarchy, and the discriminant enhanced memory updater. Firstly, we propose a question-guided interval spatio-temporal constraint strategy. Based on the consideration of tracking status and the degree of penalty expansion, it enables the dynamic adjustment of the constraint weights during tracking. Secondly, to guarantee the coverage of the target by candidate proposals in extreme motion scenarios, we propose a region proposal method over the pixel-spatial hierarchy. By combining visual common sense with reciprocal target-distractor information, it implements a careful refinement of the primary proposals, guaranteeing the reliability of tracking results. Moreover, we introduce a discriminative-enhanced memory updater designed to ensure effective model adaptation. By employing a meticulously crafted Matthew effect loss function [11] alongside a multi-peak evaluation strategy for historical frames, the updater effectively addresses the imbalance between positive and negative samples. This enhances the tracker's robustness through efficient online fine-tuning.

Our main contributions can be summarized as follows:

- We propose a question-guided interval spatio-temporal constraint strategy to adjust the penalty weight according to the tracking state. This constraint can eliminate boundary discontinuity during correct tracking and avoid tracking drift by using similar objects when tracking fails;
- We propose a region proposal method over the pixel-spatial hierarchy. This method organically integrates visual common sense and reciprocal target-distractor information to maximize the accuracy of region suggestions. In this way, abrupt motion issues can be alleviated;
- We design a discriminant enhanced memory updater based on a multi-peak sample evaluation strategy and a Matthew effect loss function. The proposed updater achieves efficient model updating.

2. Related Work

In this section, we review the related work on the spatio-temporal constraints and region proposal method and describe the updating mechanisms.

2.1. Spatio-Temporal Constraints

The spatio-temporal constraint strategy can suppress distractors based on relative shifts in time and spatial location, which, in turn, prevents the loss of targets. Early MOSSE and KCF methods [12,13] introduce the cosine window as a spatio-temporal constraint to avoid boundary effects in the tracking process. In the subsequent tracking algorithm, besides the cosine window, many different window functions were also introduced, such as the rectangle window, Hamming window, and Blackman window. However, earlier tracking algorithms performed poorly when faced with abrupt motion. One of the reasons is that when the spatial-temporal continuity of the target is broken, the conventional window function based on motion smoothness tends to impose an inappropriate penalty term. Therefore, some tracking algorithms eliminate this incorrect weight by using a special spatio-temporal constraint. Li et al. [14] introduced an innovative window function tailored specifically for visual tracking, which adaptively mitigates variable noise by leveraging similarity map observations. Furthermore, to address the detrimental impact of static spatio-temporal constraints, other researchers have explored the removal of the window function during tracking. This approach aims to better accommodate abrupt target movements.

Li et al. [15] employed a spatial regularization technique to eliminate the cosine window from the CF tracker and introduced both binary and Gaussian-shaped mask functions to effectively address issues related to boundary discontinuities and sample contamination. Despite these advancements, many spatio-temporal constraint methods fail to adapt penalty weights to varying tracking conditions. The rigid penalization method often encounters difficulties across various tracking scenarios, resulting in a discrepancy between high confidence scores on response maps and the actual dependability of tracking results. In order to address this issue, we introduce a new constraint strategy that adaptively modifies penalty weights according to tracking conditions, thereby enhancing the reliability of the response map.

2.2. Region Proposal Method

Efficient region proposal methods are crucial for the swift tracking of targets, as they provide initial proposals for predicting the target's subsequent locations. Although the sliding window search method covers the target comprehensively, it imposes a substantial computational load. In order to mitigate this, researchers have developed global search-based trackers with region proposal networks (RPNs) [16,17] to enhance robustness in tracking performance. Zhang et al. [18] proposed a cascaded RPN fusion framework featuring two RPN blocks that process features from different layers. This method integrates the results from both classification and regression tasks through weighted fusion to pinpoint the target's position accurately. Huang et al. [19] introduced a tracking method that relies exclusively on global search, demonstrating robust performance in long-term tracking benchmarks. Nonetheless, this approach faces challenges with short-term tracking, especially in scenes cluttered with numerous distracting objects. Song et al. [20] presented DTNet, a novel framework that integrates object detection and tracking within an ensemble system. DTNet employs hierarchical reinforcement learning to generate effective region proposals. Additionally, Hui et al. [21] developed the template-bridged search region interaction (TBSI) module, which uses templates to enhance interactions between RGB and thermal infrared (TIR) search regions. This method improves the collection and distribution of relevant objects and environmental context for tracking. However, the sliding window technique used here remains computationally demanding, making real-time tracking challenging. In order to overcome these limitations, we propose a new region proposal method that utilizes spatial-pixel information. This approach improves proposal accuracy by leveraging detailed visual data, achieving effective target coverage with fewer proposals and enhancing tracker robustness.

2.3. Trackers with Updater

The updater component facilitates dynamic model adjustments, allowing the tracker to adapt to intricate motion scenarios. Efficient model updates enable the tracker to adjust dynamically, ensuring effective performance in complex motion environments. Dai et al. [22] introduced a novel meta-updater that integrates geometric and appearance cues in sequence, optimizing update timing based on this combined information.

Liu et al. [23] proposed a template update strategy to improve tracking accuracy. In the presence of background clutter, the original template is retained, and both the original and updated templates are used at the location predicted by optical flow, with the more precise template being selected.

Additionally, Dai et al. [22] developed an offline-trained meta-updater to tackle the challenge of determining when to update the tracker in each frame. This meta-updater uses a cascaded LSTM module to extract sequential data and is trained to produce a binary signal to guide the update process, making it adaptable to various tracking systems.

Although these advanced update mechanisms enhance tracker robustness to some extent, their effectiveness tends to decline over time due to the lack of a robust update strategy and efficient hard negative sample mining. In order to address these issues, we propose a Matthew effect loss function to balance positive and negative samples during

updates, combined with a multi-peak sample evaluation strategy to identify hard negative samples effectively, thus ensuring the reliability of model updates.

3. Proposed Method

In this work, we develop an adaptive tracker for abrupt motion. The network architecture consists of three components: the question-guided interval spatio-temporal constraint strategy, a region proposal method over the pixel-spatial hierarchy, and a discriminant enhanced memory updater.

3.1. Overall Framework

The whole framework of our method is shown in Figure 2. Suppose the base tracker has successfully tracked the target for $t - 1$ frames in the video sequence, and the tracking result is unsatisfactory at frame t . We first activate the question-guided interval spatio-temporal constraint strategy, which integrates the inflation rate of penalty weights and the target loss time to assign appropriate penalty weights to the response map so as to ensure the reliability of the confidence scores in abrupt motion scenarios. If the tracking results are still unfavorable after adjusting the spatio-temporal constraint strategy, it is reasonable to assume that this is caused by the absence of high-quality region proposals that could cover the target even in a complex motion scenario. Therefore, the region proposal method over the pixel-spatial hierarchy would be activated, which refines the initial proposals at multiple levels by combining common sense and reciprocal target-distractor information to generate region proposals adapted to complex motion scenarios. Specifically, the question-guided interval spatio-temporal constraint strategy and the region proposal method over the pixel-spatial hierarchy will be silent if the tracking is recovered, which ensures the flexibility of the tracker and reduces the computational burden.

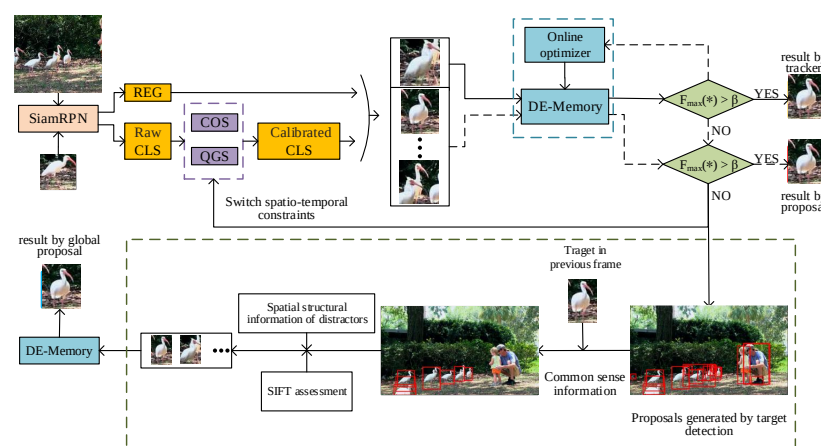


Figure 2. The proposed tracking framework. DE-memory represents the discriminant enhanced memory model, QGS represents the question-guided interval spatio-temporal constraint, and COS represents the cosine window. The bottom dotted box represents the region proposal method over the pixel-spatial hierarchy that integrates common sense and target-distractor reciprocal information.

3.2. Question-Guided Interval Spatio-Temporal Constraint Strategy

Both the target and distractor are restricted to the one-shot learning modality and can be assigned high confidence scores in the response map.

Therefore, it is worthwhile to suppress the response scores of distractors by assigning penalty weights through a spatio-temporal constraint strategy [12,13]. However, if the constraints are still implemented simply based on the motion smoothness assumption when the target movement is highly unstable, it might bring a mismatch between the response value and the true reliability. In order to address this issue, we propose the question-guided interval spatio-temporal constraint strategy to accommodate tracking tasks in various

scenarios. The dynamic constraint strategy enables the adaptive adjustment of the penalty term in accordance with the target motion, thus ensuring robustness to multiple scenarios.

We designed the average expansion rate α and average expansion step γ to monitor penalty weights in real time; the former measures the average increasing speed of penalty weights from the start of losing a target to the current time step; the latter integrates the average growth value in each time unit with its ratio compared to the form of the window function penalty. The calculation formula is defined as follows:

$$\alpha = \frac{1}{m} \times \ln \left(\frac{P_i}{P_f} \right), \quad (1)$$

$$\gamma = -\frac{1}{\alpha} \times \ln \left(\frac{P_i}{P_o} \right), \quad (2)$$

where m is the expansion time that is counted from losing a target, P_i represents the penalty value of the current point in the previous frame, P_f is the maximum allowed to be reached, and P_o is the penalty value calculated based on the window function. With α and γ , the penalty value P_i^{new} can be calculated as follows:

$$P_i^{new} = P_o \times e^{-\alpha(t+\gamma)}. \quad (3)$$

where t is counted from the beginning of the video sequence. The confidence score S_c^{new} is defined as follows:

$$S_c^{new} = (1 - w_c) \times S_c + P_i^{new} \times w_c. \quad (4)$$

where w_c is the hyperparameter, which is used to adjust the weight of the original response value S_c with the specially designed penalty term P_i^{new} .

With such careful design, our constraint strategy has the following attributes:

- (1) It can consider both the tracking status and expansion rate, adjusting penalty weights dynamically to maintain fitness between the response value and true confidence score.
- (2) By calculating the expansion rate α and translation margin γ , we can monitor the application of the interval spatial-temporal constraint strategy in real time, ensuring the penalty term stays within a reasonable range. More details are shown in Figure 3.

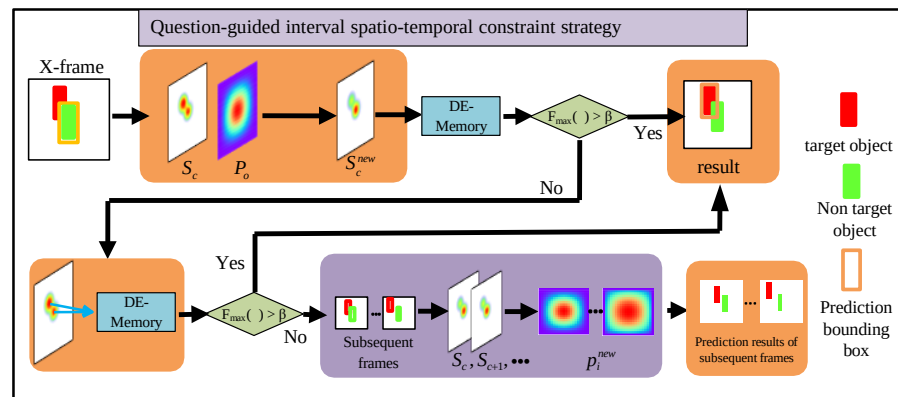


Figure 3. The question-guided interval spatio-temporal constraint strategy is continuously enabled when the highest confidence score F_{max} of all proposals is less than β , and it is dormant when F_{max} is greater than β . In order to avoid the adverse effects of the assumptions about motion smoothness being broken, the penalty weights on the response map are not assigned depending on the distance alone but integrate the expansion rate and the loss time of the target.

3.3. Region Proposal Method over the Pixel-Spatial Hierarchy

If the tracking results are still undesirable after the application of interval spatio-temporal constraints, it is reasonable to suppose that the conventional region proposal method is not sufficient for complex motion scenarios. On this basis, we propose the

region proposal method over the pixel-spatial hierarchy, which refines the proposals by fully mining the information in the video sequences. The generated region proposals can guarantee the coverage of a target with drastic position changes, which further ensures the robustness of the tracking results.

We took advantage of RefineDet [24] to perform a global search and revise the initial output to accommodate the tracking task. The set of primary proposals can be expressed as

$$B_t^{RD} = \{(B_t^\Omega, C_b) \mid c_t^* \geq \text{threshold}, c_t^* \in C_b\}, \quad (5)$$

where c_t^* is the score of each proposal in a certain category. $B_t^\Omega = \{b_t^1, b_t^2, \dots, b_t^k\}$ denotes the set of the top k proposals regarding the target category, and C_b is the set of the corresponding scores. t represents the index of the frame. B_t^{RD} means the set of region proposals applied to the tracking task for subsequent processing.

Figure 4 illustrates the overview of the information-augmented region proposal approach. Our region proposal approach integrates common sense with reciprocal target-distractor constraints; the former maintains immunity from the background, and the latter enables the refinement of proposals with sift feature matching and distractor pairing in adjacent frames. In this way, it is capable of addressing the drastic shifts of target motion in complex tracking scenarios, ensuring robustness on the basis of maintaining the coverage of the target.

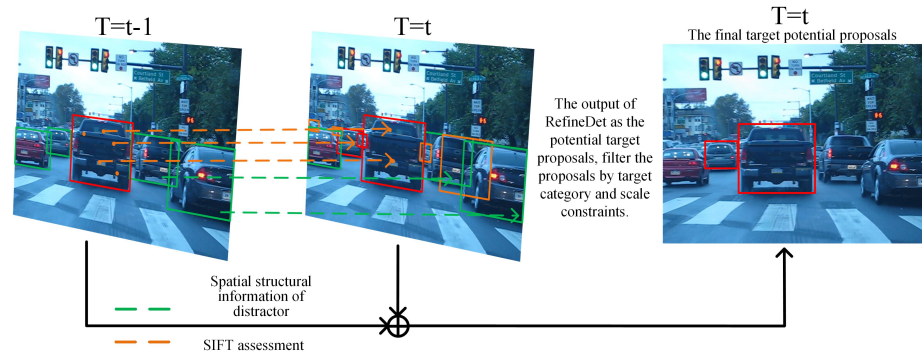


Figure 4. Overview of information-augmented region proposal approach. Our network performs the following two operations simultaneously: constrain the similarity of proposals to the distractor to obtain the set B_t^{EA} and restrict the sift feature match of proposals to the target to obtain the set B_t^{OF} . The intersection of these two sets is then performed to obtain the final set of regional proposals B_t^F .

3.3.1. Refining Proposals with Common Sense

The refinement of proposals is restricted by the limited visual information within a specific video sequence, meaning it is always hard to meet the accuracy requirements. Therefore, it is significant to introduce more generalized common sense as additional information in the tracking process. We found that the dimensional characteristics in adjacent frames would hold some stability despite the target tending to change position drastically in scenes of abrupt motion. Inspired by this idea, we constrain the aspect ratio and size shift variation of proposals to a certain range. The formula is expressed as

$$r_1 \frac{W_{t-1}^*}{H_{t-1}^*} \leq \frac{W_t^i}{H_t^i} \leq r_2 \frac{W_{t-1}^*}{H_{t-1}^*}, \quad (6)$$

$$r_1 W_{t-1}^* H_{t-1}^* \leq W_t^i H_t^i \leq r_2 W_{t-1}^* H_{t-1}^*, \quad (7)$$

where the subscripts $t-1$ and t denote the current frame and the previous frame, respectively. W and H represent the width and height of proposal b_t^i . r_1 and r_2 are the hyperparameters that are used to constrain the dimensions to a reasonable boundary.

The potential target proposals satisfying the dimensional constraints are formed into a candidate set B_t^O :

$$B_t^O = \bigcup_{i=1}^n \{b_t^i\}. \quad (8)$$

where b_t^i is the region proposal that satisfies the scale constraints. Through such a simple and portable design, we augment the visual information available in a specific video sequence, providing additional metrics for proposal refinement.

3.3.2. Reciprocal Target-Distractor Constraints

During the tracking process, the distractor tends to share spatial information during a period of time, and the intrinsic characteristics of the target would keep solidity to a certain extent. Therefore, how to maximize the information embedded in the distractor and target to refine the proposal is highly significant. On this basis, the reciprocal target-distractor constraints method is proposed, which integrates the information in the target and distractor to complement each other for the further refinement of the region proposal. Our target-distractor constraints method could make the preliminary proposals adaptable to tracking tasks in complex motion scenarios. The whole process contains two parts: the spatial structural information of the distractor and SIFT assessment.

I. Spatial structural information of the distractor: By mining spatial information from distractors and applying it as a criterion for assessing regional proposals, the inadequacy of target characteristics can be alleviated. Our network would store the spatial information of the distractor, measure the similarity between the distractor and the current candidate box, and drop the proposals with an intersection ratio above the threshold in the candidate set. The formula is expressed as

$$B_t^{EA} = \{(B_t^O, s_t^i) \mid s_t^i \leq \text{threshold}, s_t^i \in S_b\}. \quad (9)$$

where s_t^i represents the IoU score between the potential proposal and the distractor. The proposals in B_t^{EA} calculate the IoU score, with the distractor extracted from previous frames one by one, and the proposals with a greater score than the set threshold are rejected.

II. SIFT assessment Under specific tracking challenges (e.g., camera switching and focal length shifts), the dimensional continuity regarding distractors may be compromised. In order to doubly guarantee the reliability of the regional proposals, we took advantage of SIFT [25] as a complement assessment. The formula can be expressed as

$$R_{t-1} = \{r_1, r_2, \dots, r_{128}\}, \quad (10)$$

$$G_t = \{g_1, g_2, \dots, g_{128}\}, \quad (11)$$

$$d(R_{t-1}, G_t) = \sqrt{\sum_{j=1}^{128} (r_j - g_j)^2}, \quad (12)$$

$$\frac{d_{first}}{d_{second}} < \text{Threshold}. \quad (13)$$

where R_{t-1} and G_t are the 128-dimensional vector representations of a sift feature point of the last frame and the current frame, respectively. d_{first} denotes the nearest point of R_{t-1} in the current frame, and d_{second} means the next closest point of R_{t-1} . By restricting the ratio of the distance between the closet neighbor and the second-closest neighbor, we can obtain the candidate proposal set B_t^{OF} .

In order to avoid the prejudice associated with relying on one criterion alone, B_t^{EA} and B_t^{OF} are integrated into the final region proposals set B_t^F by performing OR operations:

$$B_t^F = \{B_t^{EA} \cup B_t^{OF}\}. \quad (14)$$

Finally, the global region proposals will be fed to the discriminant enhanced memory model one by one for assessment.

3.4. Discriminant Enhanced Memory Updater

Given the inherent uncertainty of target motion during practical tracking, it is essential to evaluate the tracking state accurately and update the model effectively. In order to address this challenge, we introduce a Matthew effect loss function combined with a multi-peak sample evaluation method to manage the imbalance between positive and negative samples in training. This loss function focuses on difficult negative samples by redistributing loss, and the evaluation method expands the negative sample set by identifying high-interference objects throughout the video sequence. This strategy ensures that our memory updater remains robust and addresses sample imbalance efficiently without excessive computational costs.

The network structure of the discriminant enhanced memory updater is presented in Figure 5. We took advantage of the light structure of MDNet [26] as an online memory network and changed the *fc6* layer in MDNet to a score-evaluation layer.

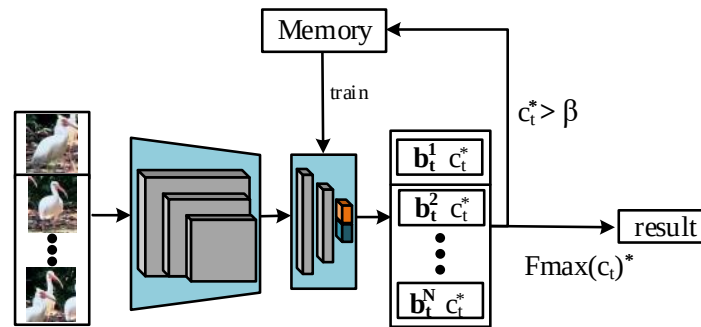


Figure 5. The network architecture of the discriminant enhanced memory updater. The memory updater evaluates all candidate proposals, b_t^i , one by one, and c_t^* is the score of each proposal evaluated. If the highest score is greater than the set threshold, the current tracking strategy will be considered for adaptation to the target's motion. Otherwise, the spatio-temporal constraint strategy and region proposal method over the pixel-spatial hierarchy will be gradually activated.

3.4.1. Matthew Effect Loss

We introduce a Matthew effect loss function aimed at optimizing the training process of the memory updater by addressing the imbalance between positive and negative samples. Unlike conventional loss functions, such as BCE loss and L2 loss, which uniformly apply loss calculations to all samples, the Matthew effect loss function is specifically designed to reduce the influence of easy samples in the background while compelling the memory model to concentrate on challenging negative samples. The proposed Matthew effect loss is expressed as follows:

$$L_M = \begin{cases} -(1 - S_d)^{\min(r(t), r_{\max})} \log(S_d), & \text{if } S = 1 \\ -S_d^{\min(r(t), r_{\max})} \log(1 - S_d), & \text{otherwise} \end{cases} \quad (15)$$

where $S \in \{0, 1\}$ represents the true class label, while S_d denotes the predicted output of the memory model for the corresponding class S , ranging from either 0 or 1. The epoch index is denoted by t , and the function $r(t) = r_0 + \sigma t$ increases over time, where r_0 is the initial focusing factor and σ represents the rate of increment. $r_{\max}$ serves as the upper limit

for the focusing factor. As t grows, $r(t)$ escalates, leading to a reduction in losses associated with easy examples, thereby diminishing their influence on the online learning process. Essentially, the Matthew effect loss intrinsically reduces the contribution of easy examples over time, directing focus towards harder negative samples as training progresses.

We conducted training on the memory updater using both BCE loss and the Matthew effect loss separately and subsequently analyzed the impact of each loss function on the performance of the memory model. During the video sequence with occlusion, BCE loss is accompanied by large fluctuations, while the Matthew effect loss function maintain stability in the face of a challenge, as shown in Figure 6.

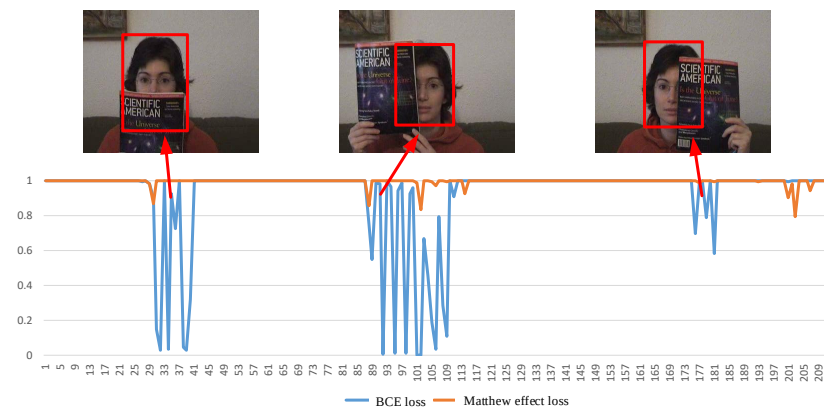


Figure 6. The results of quality assessment after the online fine-tuning of the memory module using BCE loss and Matthew effect loss. The horizontal axis represents the frames of the video, and each frame feeds the bounding box of the target into the memory model. The vertical axis represents the quality assessment score of the memory model for the target.

3.4.2. Multi-Peak Sample Evaluation

The response map contains several high-scoring points after the basic calculation, which can be denoted as multiple peaks. These multiple peaks represent objects with high interference that may lead to tracking drift on the one hand but also negative samples with considerable training value on the other hand. Based on the idea of making full use of highly interfering objects, we propose the multi-peak sample evaluation strategy. Instead of relying entirely on random sampling to collect training data, the hard-negative sample pool is expanded by selecting samples with high response values from the previous framework. The formula can be expressed as

$$N := \{\forall n_i \in N, F(n_i) > h \cap n_i \neq z_t, \} \quad (16)$$

where $F()$ represents the discriminant enhanced memory model, n_i denotes the proposals that do not belong to the response peak. h is the predefined threshold, z_t is the selected target in frame t , and N is the set of proposals corresponding to the peak. Specifically, the multi-peak sample evaluation strategy will be silent before the basic tracker fails; thus, it would not increase the computational burden for online applications.

4. Experiment

In this section, we evaluate our tracker's performance on several benchmark datasets and compare it against leading tracking algorithms. We also perform ablation studies to examine the impact of each model component on its overall performance.

4.1. Datasets and Implementation Details

Our tracker SiamIPA was evaluated on five datasets: OTB-100 [27], UAV123 [28], LaSOT [29], VOT2016 [30], and VOT2018 [31]. The experiment was conducted on a PC with Intel (R) Core (TM) i7-9700K CPU 3.60GHZ and NVIDIA Quadro RTX 4000.

We evaluated performance by using public metrics on the benchmark datasets. For the OTB-100 and LaSOT datasets, we applied one-pass evaluation (OPE) metrics, focusing on precision and success rates. This method involves tracking throughout the entire video sequence based solely on information from the initial frame. Precision is determined by the proportion of frames where the tracked location is within a specified distance from the ground truth, while the success rate measures the overlap between the predicted and actual bounding boxes. In assessing the UAV123 dataset, we use the average ratio of errors (AREs) and average overlap ratio (AOR). ARE quantifies tracking accuracy by calculating the ratio of error between the estimated and true positions, whereas AOR evaluates the overlap between the estimated and actual bounding boxes. In order to compare tracker performance more generally, we employed the distance precision (DP) score with a 20-pixel threshold. For the VOT2016 and VOT2018 datasets, we analyzed performance based on accuracy, robustness, and expected average overlap (EAO). Accuracy indicates the average overlap during successful tracking, while EAO offers a holistic measure of tracker performance. Robustness is assessed by the stability of the tracker, with lower values indicating fewer tracking failures.

4.2. Comparison with the State-of-the-Art Method

We compare our tracking algorithm with other state-of-the-art methods on five publicly available datasets.

4.2.1. Result on OTB-100

The OTB-100 dataset consists of 100 sequences from widely used tracking benchmarks and covers 11 distinct challenging factors, including *illumination*, *scale*, *occlusion*, *deformation*, *motion blur*, *fast movement*, *in-plane*, and *out-of-plane rotation*, among other difficulties. The evaluation was conducted using two metrics: center location error and bounding box overlap ratio. Our algorithm was compared with nine representative trackers, including DaSiamRPN [32], ECO [33], GradNet [34], DeepSRDCF [35], SiamRPN [7], SiamDW [36], CFNet [37], SiamFC [6], and fDSST [5]. Figure 7 shows that our method surpasses correlation filter-based trackers, such as ECO, DeepSRDCF, and fDSST, in tracking performance. Moreover, it achieves the highest SUC score among Siamese network-based trackers, demonstrating superior robustness in handling abrupt target motion.

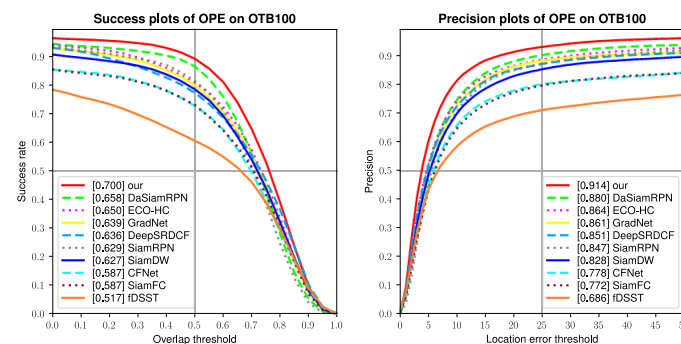


Figure 7. Comparison of the experimental results in terms of success and precision plots using the OTB-100 dataset.

4.2.2. Result on UAV123

The UAV123 dataset, which includes 123 videos recorded by a low-altitude drone, is notable for its frequent viewpoint variations. In Table 1, we assess our algorithm using the same evaluation metrics as the OTB-100 dataset. Our results were compared with several other tracking algorithms, including SiamCAR, SiamRPN++, DaSiamRPN, SiamRPN, ECO, SiamFC, DeepSRDCF, Staple, and MEEM. As indicated in Table 1, our method achieves the highest rankings in both ARE and AOR metrics. These results demonstrate that our tracker effectively addresses challenges related to target deformation and occlusion.

Table 1. Comparison of the results of the state-of-the-art trackers in UAV123. The best, second-best, and third-best values are highlighted in red, blue, and green.

Tracker	ARE	AOR
SiamCAR [4]	0.760	0.614
SiamRPN++ [38]	0.752	0.610
DaSiamRPN [32]	0.724	0.569
SiamRPN [7]	0.710	0.577
ECO [33]	0.688	0.525
SiamFC [6]	0.648	0.485
DeepSRDCF [35]	0.627	0.463
Staple [39]	0.614	0.450
MEEM [40]	0.570	0.412
SiamIPA (Ours)	0.815	0.620

4.2.3. Results When Using LaSOT

The LaSOT benchmark comprises 280 videos in the test set, covering 70 tracking categories, with an average of approximately 2500 frames per video. As a recent long-term dataset with high-quality annotations, it demands effective online model adaptation. Figure 8 illustrates the success and precision plots for our proposed framework alongside other leading trackers, such as SiamCAR, SiamBAN [41], SiamRPN++ [38], MDNet, SiamFC, DSiam [42], Staple [39], and KCF [13]. Our tracker achieves an SUC score of 0.511, surpassing the score of 0.495 for SiamRPN++. Compared to SiamFC, our tracker shows significant improvement, with an increase of 0.336 points for the SUC score, yielding a comparable performance to the SOTA SiamCAR. For accuracy, our tracker scores 0.509, markedly better than the 0.373 score of MDNet.

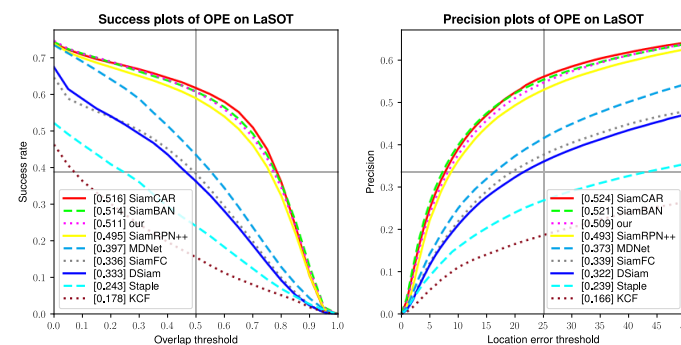


Figure 8. Comparison of the experimental results in terms of success and precision plots when using the LaSOT dataset.

4.2.4. Results When Using VOT2016

VOT2016 includes 60 video sequences and employs a range of metrics to assess tracking performance, including expected average overlap (EAO), accuracy, and robustness, which measure overall accuracy and the frequency of tracking failures per frame. The challenge protocol requires re-initializing the target each time tracking fails, and it reports both accuracy (assessed through bounding box overlap) and robustness (determined by the number of failures). Table 2 shows a comparison of our method with C-COT, ECO-HC, Staple, EBT, MDNet, SiamRN, SiamAN, and SiamRPN. Our approach outperforms these methods in all evaluated metrics: EAO, accuracy, and robustness, which demonstrate its effectiveness in maintaining stability and achieving high overlap with ground truth.

Table 2. Comparison of the results of the state-of-the-art trackers when using VOT2016. The best, second-best, and third-best values are highlighted in red, blue, and green.

Tracker	EAO	Accuracy	Robustness
SiamRPN [7]	0.344	0.56	1.08
C-COT [43]	0.331	0.53	0.85
ECO-HC [33]	0.322	0.53	1.08
Staple [39]	0.295	0.54	1.35
EBT [44]	0.291	0.47	0.9
MDNet [26]	0.257	0.54	1.2
SiamRN [45]	0.277	0.55	1.37
SiamAN [31]	0.235	0.53	1.65
SiamIPA (Ours)	0.407	0.619	0.228

4.2.5. Results When Using VOT2018

In order to assess the performance of our proposed algorithm, we conducted experiments using the VOT2018 dataset and benchmarked our tracker against several established methods, including C-COT, UpdateNet, SiamFC, Staple, DAT, SiamLM, MEEM, and KCF, as summarized in Table 3. The comparison of the results demonstrates that our method surpasses the others, achieving the highest EAO score of 0.309 and an accuracy score of 0.578. Conversely, SiamLM achieved the highest robustness score of 0.297.

Overall, our method achieves superior performance by taking advantage of spatio-temporal constraints and pixel-spatial hierarchy proposals.

Table 3. Details about the state-of-the-art trackers when using VOT2018. The best, second-best, and third-best values are highlighted in red, blue, and green.

Tracker	EAO	Accuracy	Robustness
C-COT [43]	0.267	0.494	0.318
DAT [46]	0.144	0.435	0.721
UpdateNet [47]	0.244	0.519	0.454
Staple [39]	0.169	0.530	0.688
SiamFC [6]	0.187	0.503	0.585
SiamLM [31]	0.230	0.50	0.297
MEEM [48]	0.193	0.463	0.534
KCF [13]	0.134	0.447	0.773
SiamIPA (Ours)	0.309	0.578	0.314

4.3. Ablation Study

In this section, we perform an ablation study to assess the effects of the discriminant-enhanced memory model, the question-guided interval spatio-temporal constraint approach, and the pixel-spatial hierarchy proposals. In order to illustrate the impact of these components, we evaluated our algorithm using the OTB100 dataset and a one-pass evaluation method. The outcomes, which include detection precision (DP) and overlap precision (OP), are detailed in Table 4. In concrete, ‘Siambase’ refers to the basic tracker utilized in our model. ‘SiamM’ represents the tracker enhanced with the discriminant memory updater. ‘SiamMQ’ indicates the tracker incorporating both the discriminant memory updater and the question-guided interval spatio-temporal constraint strategy. Finally, ‘SiamIPA’ denotes the tracker that integrates all three components.

Table 4 demonstrates that the performance metrics (DP and OP) improve progressively as each of the three additional components is integrated into the base tracker. In order to provide a clearer understanding of how the tracker performs under various challenging conditions, Figure 9 illustrates the tracking performance under out-of-view scenarios, out-of-plane rotations, motion blur, background clutter, illumination changes, rapid movements, scale variations, and occlusion cases. We can see that our method behaves well under different scenarios.

Table 4. Ablation studies when using the OTB100 dataset. The best, second-best, and third-best values are highlighted in red, blue, and green.

Attributes	DP	OP
Siambase	0.847	0.629
SiamM	0.872	0.662
SiamMQ	0.883	0.672
SiamIPA (Ours)	0.914	0.700

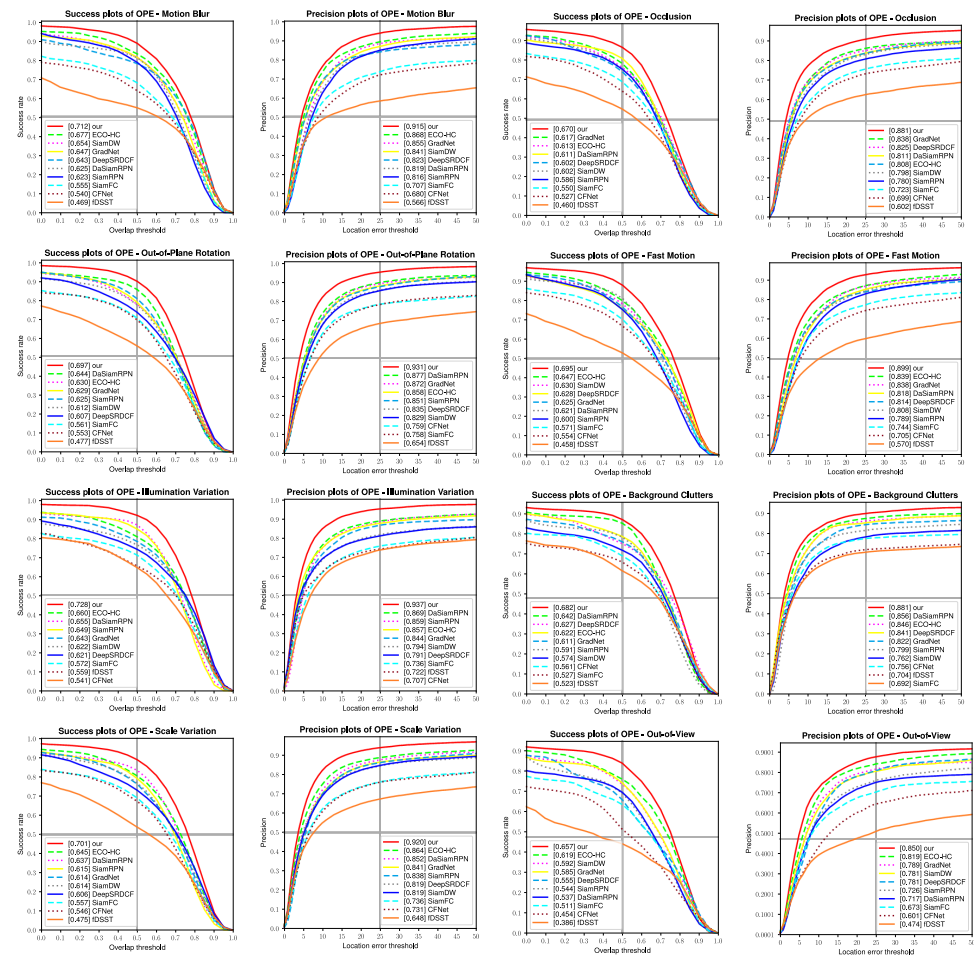


Figure 9. Comparison of experiments in terms of precision plots and success plots for challenging attributes, including out of view, out-of-plane rotation, motion blur, background clutters, illumination variation, fast motion, scale variation, and occlusion.

4.4. Limitation and Future Work

Considering the importance of feature representation for target appearance, we would employ a more powerful pretrained backbone to replace VGGnet in the future, such as a transformer-based backbone (e.g., ViT [49]) to enhance the feature representation ability.

Moreover, the proposed model is built upon the Siamese network and runs on the NVIDIA RTX 4000. In order to adapt the proposed model for more general and lower-performance hardware, we exploited applying some lightweight strategies, such as knowledge distillation or binary quantization.

5. Conclusions

This paper proposes a novel tracking method based on the interval spatio-temporal constraints and pixel-spatial hierarchy proposals for abrupt motion tracking. The question-guided interval spatio-temporal constraint strategy could assign penalty weights based on

the expansion rate and the target loss time, ensuring the match between the response score and true confidence level in abrupt motion scenarios. Then, the pixel-spatial hierarchy region proposal method integrates the information of visual common sense and reciprocal target-distractor information, maximizing the visual information to achieve the elaborate refinement of the region proposal. On this basis, our region proposal method can guarantee the coverage of a target with a limited number of high-quality candidate proposals, even when the motion of a target is highly unstable. Extensive experiments across various benchmark datasets demonstrate that our proposed tracker outperforms existing state-of-the-art methods.

Author Contributions: Methodology, X.L.; Software, D.S.; Writing—original draft, D.S.; Writing—review & editing, X.L.; Supervision, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, S.; Liu, D.; Srivastava, G.; Polap, D.; Woźniak, M. Overview and methods of correlation filter algorithms in object tracking. *Complex Intell. Syst.* **2021**, *7*, 1895–1917. [\[CrossRef\]](#)
2. Hu, W.M.; Wang, Q.; Gao, J.; Li, B.; Maybank, S. Dcfnet: Discriminant correlation filters network for visual tracking. *J. Comput. Sci. Technol.* **2024**, *39*, 691–714. [\[CrossRef\]](#)
3. Cui, Y.; Jiang, C.; Wang, L.; Wu, G. Fully convolutional online tracking. *Comput. Vis. Image Underst.* **2022**, *224*, 103547. [\[CrossRef\]](#)
4. Guo, D.; Wang, J.; Cui, Y.; Wang, Z.; Chen, S. SiamCAR: Siamese fully convolutional classification and regression for visual tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6269–6277.
5. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Discriminative scale space tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1561–1575. [\[CrossRef\]](#)
6. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-convolutional siamese networks for object tracking. In Proceedings of the European Conference Computer Vision Workshops, Amsterdam, The Netherlands, 11–14 October 2016; pp. 850–865.
7. Li, B.; Yan, J.; Wu, W.; Zhu, Z.; Hu, X. High performance visual tracking with siamese region proposal network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8971–8980.
8. Zhang, J.; Wang, M.; Jiang, H.; Zhang, X.; Yan, C.; Zeng, D. STAT: Multi-object tracking based on spatio-temporal topological constraints. *IEEE Trans. Multimed.* **2023**, *26*, 4445–4457. [\[CrossRef\]](#)
9. Zhang, Y.; Liang, Y.; Leng, J.; Wang, Z. SCGTracker: Spatio-temporal correlation and graph neural networks for multiple object tracking. *Pattern Recognit.* **2024**, *149*, 110249. [\[CrossRef\]](#)
10. Xu, T.; Zhu, X.F.; Wu, X.J. Learning spatio-temporal discriminative model for affine subspace based visual object tracking. *Vis. Intell.* **2023**, *1*, 4. [\[CrossRef\]](#)
11. Rigney, D. *The Matthew Effect: How Advantage Begets Further Advantage*; Columbia University Press: New York, NY, USA, 2010.
12. Bergen, S.W.; Antoniou, A. Design of ultraspherical window functions with prescribed spectral characteristics. *EURASIP J. Adv. Signal Process.* **2004**, *2004*, 196503. [\[CrossRef\]](#)
13. Kaiser, J.; Schafer, R. On the use of the I_0 -sinh window for spectrum analysis. *IEEE Trans. Acoust. Speech Signal Process.* **1980**, *28*, 105–107. [\[CrossRef\]](#)
14. Li, S.; Zhao, S.; Cheng, B.; Chen, J. Noise-aware framework for robust visual tracking. *IEEE Trans. Cybern.* **2020**, *52*, 1179–1192. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Li, F.; Wu, X.; Zuo, W.; Zhang, D.; Zhang, L. Remove cosine window from correlation filter-based visual trackers: When and how. *IEEE Trans. Image Process.* **2020**, *29*, 7045–7060. [\[CrossRef\]](#)
16. Dai, P.; Weng, R.; Choi, W.; Zhang, C.; He, Z.; Ding, W. Learning a proposal classifier for multiple object tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2443–2452.
17. Voigtlaender, P.; Luiten, J.; Torr, P.H.; Leibe, B. Siam r-cnn: Visual tracking by re-detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6578–6588.
18. Zhang, J.; Wang, K.; He, Y.; Kuang, L. Visual Object Tracking via Cascaded RPN Fusion and Coordinate Attention. *CMES-Comput. Model. Eng. Sci.* **2022**, *132*. [\[CrossRef\]](#)
19. Huang, L.; Zhao, X.; Huang, K. Globaltrack: A simple and strong baseline for long-term tracking. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11037–11044.

20. Zhang, W.; Song, R.; Li, Y. Online decision based visual tracking via reinforcement learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 11778–11788.
21. Hui, T.; Xun, Z.; Peng, F.; Huang, J.; Wei, X.; Wei, X.; Dai, J.; Han, J.; Liu, S. Bridging search region interaction with template for rgb-t tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 13630–13639.
22. Dai, K.; Zhang, Y.; Wang, D.; Li, J.; Lu, H.; Yang, X. High-performance long-term tracking with meta-updater. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6298–6307.
23. Liu, S.; Liu, D.; Muhammad, K.; Ding, W. Effective template update mechanism in visual tracking with background clutter. *Neurocomputing* **2021**, *458*, 615–625. [[CrossRef](#)]
24. Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; Li, S.Z. Single-shot refinement neural network for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4203–4212.
25. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
26. Nam, H.; Han, B. Learning multi-domain convolutional neural networks for visual tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4293–4302.
27. Wu, Y.; Lim, J.; Yang, M.H. Online object tracking: A benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2411–2418.
28. Mueller, M.; Smith, N.; Ghanem, B. A benchmark and simulator for uav tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 445–461.
29. Fan, H.; Bai, H.; Lin, L.; Yang, F.; Chu, P.; Deng, G.; Yu, S.; Huang, M.; Liu, J.; Xu, Y.; et al. Lasot: A high-quality large-scale single object tracking benchmark. *Int. J. Comput. Vis.* **2021**, *129*, 439–461. [[CrossRef](#)]
30. Gündoğdu, E.; Alatan, A.A. The Visual Object Tracking VOT2016 challenge results. In Proceedings of the Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, 8–10.15–16 October 2016.
31. Kristan, M.; Leonardis, A.; Matas, J.; Felsberg, M.; Pflugfelder, R.; Čehovin Zajc, L.; Vojir, T.; Bhat, G.; Lukežić, A.; Eldesokey, A.; et al. The sixth visual object tracking vot2018 challenge results. In Proceedings of the European Conference on Computer Vision Workshops, Munich, Germany, 8–14 September 2018.
32. Zhu, Z.; Wang, Q.; Li, B.; Wu, W.; Yan, J.; Hu, W. Distractor-aware siamese networks for visual object tracking. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 101–117.
33. Danelljan, M.; Bhat, G.; Shahbaz Khan, F.; Felsberg, M. Eco: Efficient convolution operators for tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6638–6646.
34. Li, P.; Chen, B.; Ouyang, W.; Wang, D.; Yang, X.; Lu, H. GradNet: Gradient-guided network for visual object tracking. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6162–6171.
35. Danelljan, M.; Hager, G.; Shahbaz Khan, F.; Felsberg, M. Learning spatially regularized correlation filters for visual tracking. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4310–4318.
36. Zhang, Z.; Peng, H. Deeper and wider siamese networks for real-time visual tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4591–4600.
37. Shen, Z.; Dai, Y.; Rao, Z. Cfnet: Cascade and fused cost volume for robust stereo matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13906–13915.
38. Li, B.; Wu, W.; Wang, Q.; Zhang, F.; Xing, J.; Yan, J. Siamrpn++: Evolution of siamese visual tracking with very deep networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4282–4291.
39. Bertinetto, L.; Valmadre, J.; Golodetz, S.; Miksik, O.; Torr, P.H. Staple: Complementary learners for real-time tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1401–1409.
40. Zhang, J.; Ma, S.; Sclaroff, S. MEEM: Robust tracking via multiple experts using entropy minimization. In Proceedings of the European Conference Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 188–203.
41. Chen, Z.; Zhong, B.; Li, G.; Zhang, S.; Ji, R. Siamese box adaptive network for visual tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2014; Springer: Berlin/Heidelberg, Germany, 2020; pp. 6668–6677.
42. Guo, Q.; Feng, W.; Zhou, C.; Huang, R.; Wan, L.; Wang, S. Learning dynamic siamese network for visual object tracking. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1763–1771. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1401–1409.
43. Danelljan, M.; Robinson, A.; Shahbaz Khan, F.; Felsberg, M. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In Proceedings of the European Conference Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 472–488.
44. Zhu, G.; Porikli, F.; Li, H. Beyond local search: Tracking objects everywhere with instance-specific proposals. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 943–951.

45. Cheng, S.; Zhong, B.; Li, G.; Liu, X.; Tang, Z.; Li, X.; Wang, J. Learning to filter: Siamese relation network for robust tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, US, 20–25 June 2021; pp. 4421–4431.
46. Pu, S.; Song, Y.; Ma, C.; Zhang, H.; Yang, M.H. Deep attentive tracking via reciprocative learning. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 1935–1945.
47. Zhang, L.; Gonzalez-Garcia, A.; Weijer, J.v.d.; Danelljan, M.; Khan, F.S. Learning the model update for siamese trackers. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4010–4019.
48. Hadfield, S.; Bowden, R.; Lebeda, K. The visual object tracking VOT2016 challenge results. *Lect. Notes Comput. Sci.* **2016**, *9914*, 777–823.
49. Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.