Adaptive Spatial Regularization Correlation Filters for UAV Tracking

Yulin Cao, Shihao Dong, Jiawei Zhang, Han Xu, Yan Zhang, and Yuhui Zheng

***Abstract*—As a tool for near-earth remote sensing, unmanned aerial vehicle (UAV) can be used to acquire images and data of the earth's surface. This provides a powerful support for Earth observation and resource management. Object tracking in UAV videos has been a topic of much interest in recent years. A large number of algorithms have been proposed. Among these algorithms, deep learning has achieved a high accuracy rate. However, it is difficult to carry hardware devices for UAV, which makes it difficult to be practically applied. The correlation filter does not require a graphic processing unit to accelerate the computation, but it uses only manual features, which makes it difficult to achieve satisfactory performance. In order to solve the above problems, we proposed adaptive spatial regularization correlation filters, called DTSRT. Specifically, we first introduce deep features in the correlation filter instead of the original manual features, which can greatly improve the discriminative ability of the model. At the same time, in order to prevent affecting the real-time performance of the algorithm, we use histogram of oriented gradients features to determine the target scale and deep features to determine the target location. In addition, considering the large inter-frame distance between targets in UAV videos, we use a saliency detection method to dynamically generate spatially constrained templates. The proposed DTSRT outperforms other state-of-the-art algorithms with area under curve of 0.481, 0.431, and 0.474 on UAV123@10FPS, UAVDT, and DTB70 datasets, respectively.**

***Index Terms*—Adaptive spatial regularization, correlation filter, deep features, UAV tracking.**

I. Introduction

Near-earth remote sensing refers to the acquisition of data and information from the near-Earth surface through the use of sensors. Unmanned aerial vehicle (UAV) can fly at low altitudes and carry various sensors, such as optical cameras, infrared cameras, light detection and ranging, and so on, which are used to acquire high-resolution images and data from the surface. Compared with satellite remote sensing, UAVs can control the flight altitude and trajectory more precisely when acquiring data and therefore can obtain higher resolution images. This gives UAVs an advantage in applications that require high precision and detailed information, such as agricultural monitoring, environmental monitoring, urban planning, and other fields. Therefore, UAVs are widely used in near-earth remote sensing applications [1], [2], [3], [4], [5]. One of the popular research directions is UAV tracking.

As the name suggests, UAV tracking is the continuous localization of prelabeled targets in UAV videos. Different from OTB dataset [6], UAV videos are all taken by UAVs. As the UAV shoots video, it is itself in constant motion, which makes the trajectory of the target irregular. In addition, the excessively large interframe distance makes some regularization methods play a detrimental role in the performance of the model. In addition to the above, low resolution and occlusion are also issues that have to be considered.

A robust model should be able to effectively deal with the above challenges and not lose track of the target throughout, which are undoubtedly difficult. Nevertheless, a large number of algorithms have been proposed for achieving accurate tracking. These algorithms can be broadly categorized into deep learning [7], [8], [9], [10], [11], [12] and correlation filters [13], [14], [15], [16], [17], [18], [19]. Deep learning extracts high-dimensional semantic information about the target mainly with the help of a convolutional neural network (CNN). Unlike handcrafted features such as HOG and color names (CN), the deep features extracted by CNN are more resistant to interference, and they are less susceptible to environmental influences. SwinTransformer neck and new data association method (STN-Track) [7] uses the swinTransformer as a backbone network for feature extraction, which significantly enhances the global interaction capability of the model. Similar to STN-Track, a hierarchical feature pooling transformer [8] also uses a Transformer in the tracking framework to enhance the representational information of small targets. An efficient aerial tracker (SmallTrack) [9] embeds graph neural network into the predictor header to improve the model's ability to recognize small targets. Aiming at the problem of difficult prediction of target motion trajectories in UAVs, meta twin delayed deep deterministic policy gradient (Meta-TD3) [10] proposes a reinforcement learning strategy based on meta-learning. Due to the frequent occlusion of targets in UAVs and the scarcity of such data, an attention-based mask generation network [11] designs a mask generation network to simulate the situation when occlusion and deformation occur. Global context embedding for vehicle tracking [12] is improved at the feature level by designing a novel feature fusion network

Manuscript received 21 January 2024; revised 26 February 2024; accepted 19 March 2024. Date of publication 25 March 2024; date of current version 12 April 2024. This work was supported in part by Natural Science Foundation of Jiangsu Province under Grant BK20211539, in part by the National Natural Science Foundation of China under Grant U20B2065 and Grant U22B2056, and in part by the Qing Lan Project. *(Corresponding author: Yuhui Zheng.)*

Yulin Cao and Yuhui Zheng are with the School of Computer, Qinghai Normal University, Xining 810000, China (e-mail: caoyulin@126.com; zhengyh@vip.126.com).

Shihao Dong, Jiawei Zhang, Han Xu, and Yan Zhang are with the School of Computer, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: dongshihao@nuist.edu.cn; zjwei_2020@163.com; 202212490766@nuist.edu.cn; 20211220042@nuist.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3380574

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

to make it more applicable to UAV scenarios. These algorithms achieve great performance, but they rely on the GPU that is difficult to deploy on UAVs to accelerate the computation.

In this case, correlation filter seems to be more competitive. It accelerates the computational efficiency of complex convolution operations in the frequency domain by fast Fourier transform. The tracking algorithms based on correlation filters have seen a large-scale development due to good accuracy as well as speed. Kernelized correlation filters (KCF) [13] proposed by Henriques et al. is the landmark in the development of correlation filters. KCF further adopts the multichannel HOG features. This allows the correlation filter to outperform the previously optimal algorithm and maintain a high speed. To make the correlation filtering algorithm more applicable to the field of remote sensing, some scholars have improved it. Given that most algorithms predict the likely future location of a target solely from the current frame, ignoring the role of past frames, Lin et al. [14] fully investigated the interframe information of target motion and find the reversibility, which allows the model to learn changes in the appearance of the target from past frames. In addition, targets in UAVs undergo drastic changes in appearance during movement, which can lead to contamination of the tracking template. The spatial regularization correlation filters with the Hilbert–Schmidt independence criterion (HSIC_SRCF) [15] utilizes peak-to-sidelobe ratio to adaptively update the templates, and the introduction of dynamic weights allows the templates to be updated over time, which is useful for improving the long-term tracking performance of the algorithm.

While the idea of cyclic translation of samples in the correlation filter greatly solves the problem of scarcity of training samples, the introduction of the boundary effect makes the tracking effect affected to some extent. Spatially regularized discriminative correlation filters (SRDCF) [16] is a classical method for target tracking, which effectively mitigates the boundary effect by introducing a spatial regularization term in the objective function. Considering the temporal order of the spatial regularization method, spatial-temporal regularized correlation filters (STRCF) [17] propose a regularization method on the spatio-temporal domain and adopts the alternating direction method of multipliers (ADMM) for the solution. Another class of methods to mitigate boundary effects are correlation filters with limited boundaries (CFLB) [18] and background-aware correlation filters (BACF) [19]. Such methods utilize the region of a larger background for target detection and filter learning, and unlike SRDCF, CFLB and BACF directly perform zero-completion operations on the filter edges through the clipping matrix to obtain correlation filters with smaller domains of action.

Recently, the integration of deep features into correlation-filtered target tracking algorithms has also become mainstream due to the wide application and superior results of deep convolutional features. Convolutional features for correlation filter (DeepSRDCF) [20] replaces HOG features with deep features in a single layer of the visual geometry group (VGG) network. The tracking performance of efficient convolution operators (ECO) [21] using a deeper network for feature extraction is not significantly improved, suggesting that correlation filters do not benefit from deeper convolutional network features. Unveiling the power of deep tracking [22] utilizes deep features to maintain robustness while shallow features are responsible for accuracy, balancing target localization accuracy and tracking robustness.

Inspired by the above algorithms, we proposed the adaptive spatial regularization correlation filters, called DTSRT. Given the low resolution of targets in UAV videos, it is difficult for manual features to efficiently represent targets. Therefore, we use deep features to replace the original manual features. Furthermore, it is known from [21] that deeper network layers do not lead to performance improvement. So we use VGG19 to complete the extraction of features and use HOG features to determine the target scale. Finally, the fixed spatial constraints in existing algorithms are difficult to apply in UAV videos. We use a detection algorithm to generate spatial constraints. Our main contributions can be summarized as follows.

1) We combine deep learning and correlation filters to strike a balance between performance and speed. The use of deep features can effectively improve the anti-interference ability of the model.
2) We use deep features to localize the target. Also to ensure that the speed of the algorithm is not seriously affected, we use HOG features to determine the target scale. Manual features are introduced to ensure speed without much loss of accuracy.
3) We impose a dynamic spatial constraint on the objective function, which can effectively mitigate the tracking drift caused by the large interframe distance.

The rest of this article is organized as follows. Section II describes the related works. Sections III and IV explain the proposed algorithm and the corresponding experimental results. We present future work in Section V. Section VI summarizes the proposed algorithm.

II. Related Works

A. Boundary Effects

The correlation filter can utilize the fast Fourier transform to accelerate the computation in the frequency domain, resulting in a substantial increase in tracking efficiency. However, when performing the Fourier transform, the cyclic shifting of samples will bring about boundary effects, which will lead to inaccurate representation of image blocks. For the problem of how to suppress the boundary effect, it is roughly divided into two categories: adding regularization terms and adding mask matrices.

The mask matrix acts directly on the circular shift samples, and it effectively weeds out inaccurate shift samples. CFLB [18] adds the mask matrix to the objective function for the first time, which can effectively reduce the number of samples with boundary effects by multiplying the resulting circularly shifted samples with the mask matrix. BACF [19] extends single-channel color features to multichannel HOG features for feature extraction and then each channel was solved according to the CFLB method, and finally the overall response map was obtained by summing.

SRDCF [16] introduces a spatial regularization matrix in the objective function to penalize the filter coefficients outside the target region of the sample, which can effectively suppress the influence of the background region. STRCF [17] employs both temporal regularization and spatial regularization to suppress boundary effects and improve the tracking efficiency. Introducing predefined regularization terms requires setting many human-defined hyperparameters in advance, which makes the tracking algorithm not generalizable. Automatic spatio-temporal regularization correlation filters (AutoTrack) [23] automatically adjusts both temporal regularization and spatial regularization, enabling adaptive changes in the relevant hyperparameters, thus saving the step of manually fine-tuning these hyperparameters. A novel object saliency-aware dual regularized correlation filter [24] simultaneously regularize the filters involving the correlation operation in ridge regression and the regularization term in the objective function, effectively suppressing the boundary effect.

Considering the importance of samples for filter training, it is necessary to alleviate the boundary effects caused by cyclic sampling. From the above, it is clear that regularization is an effective way to mitigate boundary effects. However, most algorithms only use the regularization term as a complementary term to the objective function. In this manuscript, a dynamic regularization operator is introduced, which can effectively utilize the spatial information of the object.

B. *Correlation Filters Combined With Deep Learning*

In recent years, the use of deep learning in the field of object tracking has led to a substantial increase in tracking accuracy. However, the computational complexity of deep features is too high, which makes the tracking real-time reduced. The real-time effect of target tracking algorithm based on correlation filters is very good. Thus, the integration of correlation filter and deep learning becomes an inevitable trend.

Discriminant correlation filters network (DCFNet) [25] replaces the similarity operation with a correlation filtering operation, thus enabling end-to-end tracking. Fast and robust online adaptation (MetaTracker) [26] uses real scenes from subsequent frames to learn a set of initialization parameters for the CNN network, allowing the deep network to model a specific target robustly and quickly. Wang et al. [27] proposed an unsupervised object tracking algorithm, which can utilize unlabeled datasets for unsupervised learning. In [28], knowledge distillation [29] is used to jointly compress and transmit off-the-shelf CNN teacher network models, resulting in a lightweight student network for feature extraction.

Combining deep neural network with correlation filters is undoubtedly a more advanced tracking method compared with traditional target tracking algorithms based on correlation filters. Neural networks have outstanding performance in target modeling, while correlation filters have unique advantages in tracking real-time performance, so the fusion of neural networks and correlation filters can achieve robust performance both in tracking accuracy and tracking real-time performance. In addition, considering the small target in the UAV video, it is difficult for the deep convolutional network to extract efficient semantic information, so it is necessary to select a lightweight and simple network.

C. *UAV Tracking*

Object tracking can be divided into traditional ground tracking, UAV tracking, and satellite video tracking based on application scenarios. Due to its compact and flexible characteristics, UAV tracking has always attracted the attention of scholars. However, unlike traditional tracking, targets in UAV videos often have lower resolution. Moreover, the relative motion between the UAV and the target makes tracking more difficult.

In order to accurately track targets in UAV videos, many tracking algorithms based on UAV videos have been proposed. Given that the aspect ratio of targets in UAV videos is more likely to change, compute a more accurate and robust correlation filter (ARTracker) [30] constructs a function similar to Gaussian labels for filter training, which effectively enhances the regression ability of the filter. An autoperceiving correlation filter (APCF) [31] models the target context using background features and proposes a new state estimation metric that predicts the target state by analyzing the spatial distribution of the response graph. In [32], the contextual information of the target is also used to enhance the recognition ability of the filter. Specifically, it enhances target information by suppressing incorrect information. Perceiving temporal environment for correlation filters (PTECF) [33] proposes a regularization term to learn the environmental differences between adjacent frames, thereby enhancing the sensitivity of the filter in different scenarios. In order to utilize the temporal information during the filter training process, a novel adaptive response reasoning approach (ReCF) [34] constructs auxiliary labels for the current sample to learn the general relationship between the current filter and the previous filter.

In addition to the aforementioned algorithms, many scholars [35], [36], [37] have made improvements to filters in other areas, such as using multiple features, introducing redetection mechanisms, and introducing regularization terms. Inspired by the above algorithm, we propose a novel UAV tracking algorithm. The use of deep features and dynamic regularization effectively enhances the tracking ability of the proposed algorithm on UAV videos.

III. Proposed Method

The tracking flow of the proposed DTSRT is shown in Fig. 1. Specifically, we extract the deep features of the target by VGG19. Also, in order to prevent the introduction of deep features from seriously affecting the speed of the algorithm, we use HOG features to determine the scale of the target. In addition, we dynamically generate spatial constraints on the target through the existing saliency detection method [38]. Next, we will provide a detailed introduction to the proposed algorithm.

A. *Overall Objective*

In order to train a robust filter, algorithms based on correlation filters usually use cyclic sampling to increase the training samples, which inevitably leads to boundary effects. STRCF [17] introduces a spatial regularization term in the objective function

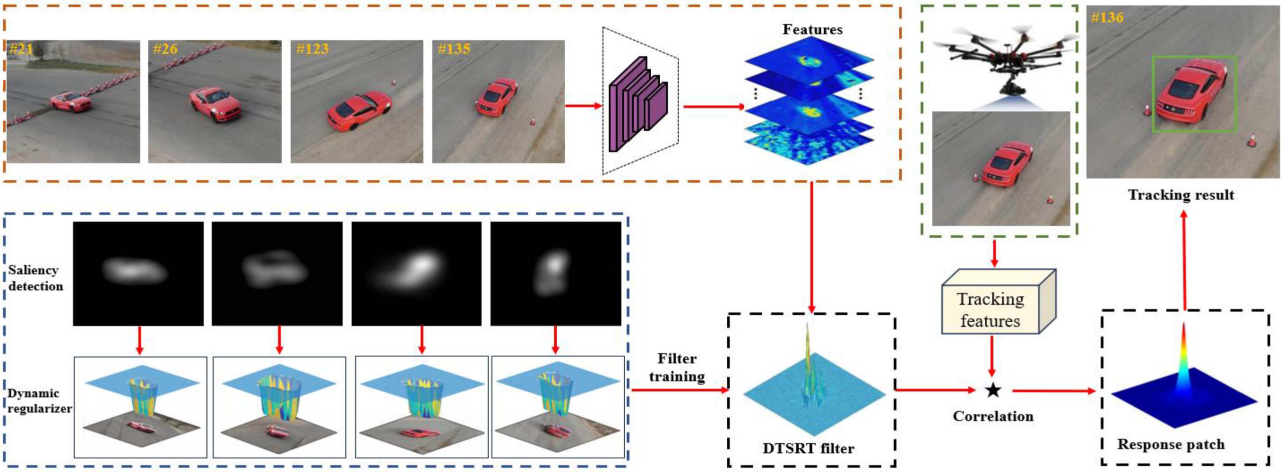

Fig. 1. Flowchart of the proposed DTSRT.

to mitigate boundary effects. Considering the large interframe distance between targets in UAV videos, it is difficult for the static regularization term to be fully effective. For this reason, we generate a dynamic constraint w_s through existing saliency detection method [38]. The objective function is shown below

$$E\ (H_t) = \frac{1}{2}\left\|\sum_{k=1}^{K} x_t^k * h_t^k - y\right\|_2^2 + \frac{1}{2}\sum_{k=1}^{K}\left\|w_s \odot h_t^k\right\|_2^2 + \frac{\mu}{2}\left\|h_t - h_{t-1}\right\|_2^2 \tag{1}$$

where y is the expected response consistent with a Gaussian distribution. t and k denote the tth frame and the kth channel, respectively. K is the total number of channels. x and h denote the feature map and filter, respectively.

Considering that (1) is a convex function, an augmented Lagrangian algorithm can be used to minimize it in order to obtain a globally optimal solution. In the optimization process, by introducing an auxiliary variable g, i.e., by making $h = g$, (1) can be rewritten as follows:

$$\begin{aligned} L\ (g_t, s) = &\frac{1}{2}\left\|\sum_{k=1}^{K} x_t^k * h_t^k - y\right\|_2^2 + \frac{1}{2}\sum_{k=1}^{K}\left\|w_s \odot g_t^k\right\|_2^2 \\ &+ \sum_{k=1}^{K}\left(h_t^k - g_t^k\right)^T s_t^k + \frac{\gamma}{2}\sum_{k=1}^{K}\left\|h_t^k - g_t^k\right\|_2^2 \\ &+ \frac{\mu}{2}\left\|h_t - h_{t-1}\right\|_2^2 \end{aligned} \tag{2}$$

where s is a Lagrange multiplier and γ is the control parameter. Let $f\ = \frac{s}{\gamma}$, the above equation can be rewritten as

$$\begin{aligned} L\ (g_t, f) = &\frac{1}{2}\left\|\sum_{k=1}^{K} x_t^k * h_t^k - y\right\|_2^2 + \frac{1}{2}\sum_{k=1}^{K}\left\|w_s \odot g_t^k\right\|_2^2 \\ &+ \frac{\gamma}{2}\sum_{k=1}^{K}\left\|h_t^k - g_t^k + f_t^k\right\|_2^2 + \frac{\mu}{2}\left\|h_t - h_{t-1}\right\|_2^2. \end{aligned} \tag{3}$$

Further, (3) can be solved and decomposed into the following three subproblems by the ADMM method:

$$\begin{aligned} h^{i+1} = \text{argmin}_h \Bigg\{ &\left\|\sum_{k=1}^{K} x_t^k * h_t^k - y\right\|_2^2 + \gamma\left\|h_t - g_t + f_t\right\|_2^2 \\ &+ \mu\left\|h_t - h_{t-1}\right\|_2^2 \Bigg\} \end{aligned} \tag{4}$$

$$g^{i+1} = \text{argmin}_g\ \left\{\sum_{k=1}^{K}\left\|w_s \odot g_t^k\right\|_2^2 + \gamma\|h_t - g_t + f\|_{t2}^{\ 2}\right\} \tag{5}$$

$$f^{i+1} = f^i\ + h^{i+1} - g^{i+1}. \tag{6}$$

1) Subproblem h: According to Parseval's theorem, this can be simplified to the following form:

$$\begin{aligned} \text{argmin}_h \Bigg\{ &\left\|\widehat{x_t^k} * \widehat{h_t^k} - \hat{y}\right\|_2^2 + \gamma\left\|\widehat{h_t} - \widehat{g_t} + \hat{f}\right\|_2^2 \\ &+ \mu\left\|\widehat{h_t} - \widehat{h_{t-1}}\right\|_2^2 \Bigg\} \end{aligned} \tag{7}$$

where ^ denotes the corresponding Fourier transform. For example, $\widehat{x_t^k}$ is the Fourier transform of x_t^k. It is too difficult to solve (7) directly, so the above equation can be solved after further simplification

$$\begin{aligned} \text{argmin}_{\Gamma_j\left(\widehat{h_t}\right)} \Bigg\{ &\left\|\Gamma_j \widehat{x_t}^T \Gamma_j\left(\widehat{h_t}\right)\right\|_2^2 + \gamma\left\|\Gamma_j\left(\widehat{h_t}\right) - \Gamma_j\left(\widehat{g_t}\right)\right. \\ &\left. + \Gamma_j\left(\widehat{f_t}\right)\right\|_2^2 + \mu\left\|\Gamma_j\left(\widehat{h_t}\right) - \Gamma_j\left(\widehat{h_{t-1}}\right)\right\|_2^2 \Bigg\} \end{aligned} \tag{8}$$

where $\Gamma_j(\widehat{x_t})$ denotes the vector representation of the jth pixel over all channels. The following solution can be obtained by the Sherman–Morrison formula:

$$\Gamma_j\left(\widehat{h_t}\right) = \frac{1}{\gamma+\mu}\left(I - \frac{\Gamma_j\left(\widehat{x_t}\right)\Gamma_j\left(\widehat{x_t}\right)^T}{\gamma+\mu+\Gamma_j\left(\widehat{x_t}\right)^T\Gamma_j\left(\widehat{x_t}\right)}\right) q \quad (9)$$

where $q = \Gamma_j\left(\widehat{h_t}\right)\widehat{y_j} + \gamma\Gamma_j\left(\widehat{g_t}\right) - \gamma\Gamma_j\left(\widehat{f_t}\right) + \mu\Gamma_j\left(\widehat{h_{t-1}}\right)$.

2) Subproblem g: Equation (5) can be solved directly in the time domain. It is solved as follows:

$$g = \left(D^T D + \gamma I\right)^{-1}\left(\gamma f_t + \gamma h_t\right) \quad (10)$$

where D is a diagonal matrix consisting of w_s and γ is the step size

$$\gamma^{i+1} = \min\left(r^{\max},\ \rho r^i\right) \quad (11)$$

where γ and ρ are the maximum step size and scale factor, respectively.

After the filter h_t is solved by the above method, the response map is obtained by correlating it with the search image. The location where the maximum response is located is the location of the target

$$r_t = \mathcal{F}^{-1}\sum_{k=1}^{K}\left(\widehat{x_t}^k \odot \widehat{h_t}^k\right) \quad (12)$$

where r_t is the response map and $\mathcal{F}^{-1}$ is inverse Fourier transform.

B. Adaptive Spatial Regularization

Due to the rapid movement of the target or the relative motion of the UAV, the interframe distance of the target may be much larger than expected. This situation makes it difficult for static spatial regularization template w to be fully effective. In this manuscript, a dynamic spatial regularization term is proposed.

Although deep learning-based detection algorithms now have high accuracy, their large number of parameters can seriously affect the real-time performance of the algorithms. In addition, in the field of object tracking, the algorithms usually crop out only an image patch of 2.5 times the size of the target and predict the target position in the image patch instead of the whole image. Therefore, traditional algorithms can still achieve more robust performance.

Specifically, we use the saliency detection method [38] to predict the approximate location of the target to generate the original saliency map. After multiplying by a cos window, the original map is then resized to the corresponding size in the appearance model used in filter training, and its coefficients are remapped to the regulation weights by a threshold. In this article, the threshold is set to 0.1. Then, the static template is weighted by the final saliency map s. As shown in Fig. 2, the green region is where the target is located, and we crop out a region twice the size of the target for saliency detection, i.e., the red region. With the above operation, the target information can be fully utilized while the background information is penalized. Finally, the adaptive spatial constraint w_s is obtained by s, i.e., $w_s = w \odot s$.

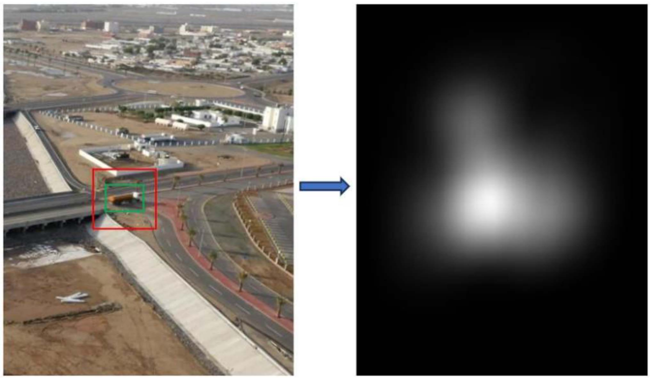

Fig. 2. Process of saliency detection.

C. Deep Features

Most of the correlation filters use manual features such as HOG features, CN features, and so on. But these features are only used to describe a single attribute. When the tracking scene changes dramatically, these features are often difficult to describe the target efficiently. This challenge is even more pronounced in UAV videos. On the one hand, it is the small size of the targets in UAV videos, whose features are difficult to extract adequately. On the other hand, the scenes in UAV videos change more frequently.

For this reason, the use of deep features has become an inevitable trend. It is clear that shallow features have a distinct appearance or outline and contain a great deal of detail about the target. Then, the deep features contain rich semantic information and are extremely resistant to interference. Therefore, deep features are more advantageous than manual features in target characterization.

IV. Experiments

A. Experimental Setup

In the experiments, the experimental equipment is a server equipped with NVIDIA GTX 2080ti GPUs and an Intel I9-9900X CPU. Scale factor p, the maximum value of the step parameter $r^{\max}$, and the initial step parameter r^0 were set to 1.2, 100, and 10, respectively. The threshold for saliency detection and the coefficient of the temporal regularity term μ were set to 0.1 and 7, respectively. To comparatively analyze the performance of the trackers, the threshold of the localization error in the experiment was set to 20 for the precision rate.

B. Datasets and Compared Algorithms

In order to fully validate the performance of the proposed DTSRT, we have selected more classical UAV datasets such as UAV123@10FPS [39], UAV20L [39], UAVDT [40], and DTB70 [41]. The UAV123@10FPS dataset consists of 123 image sequences captured by UAVs, which contain some virtual image sequences in addition to real images. The targets to be tracked in it contain various categories such as pedestrians, vehicles, buildings, etc. The long image sequences in the UAV123@10FPS

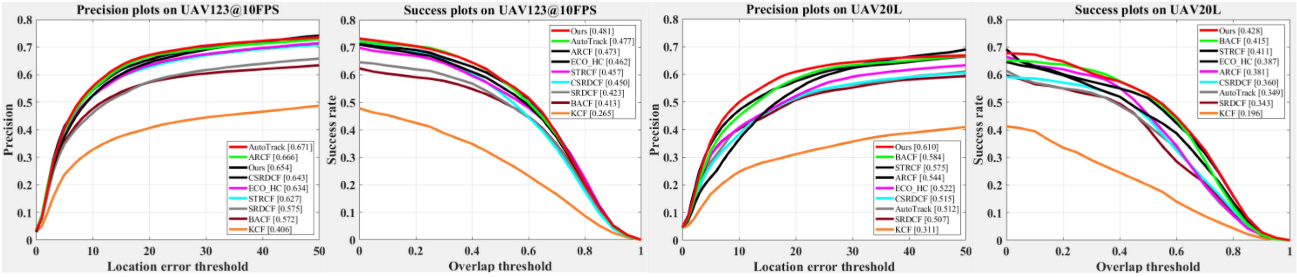

Fig. 3. Precision plots and success plots. The precision rates of precision plots and AUC of success plots are given in brackets, respectively.

dataset are selected to form the UAV20L dataset. It consists of 20 long sequences and each of these image sequences has a total frame count of 1000 or more. This dataset is generally used to test whether a model can track a target over time. The UAVDT dataset is a versatile dataset, which can be used for object tracking or detection. Most of the targets in it are vehicles. The DTB70 dataset gives 70 short image sequences, which can all be used to test the performance for UAV tracking algorithms.

We compare the proposed model with some classical algorithms on the above datasets. The algorithms compared are AutoTrack [23], aberrance repressed correlation filters (ARCF) [42], ECO_HC [21], STRCF [17], SRDCF [16], BACF [19], KCF [13], DCF [13], mutation sensitive correlation filter (MSCF) [43], multicue correlation filters (MCCT_H) [44], and discriminative correlation filter with channel and spatial reliability (CSRDCF) [45].

C. Comparison With Other Algorithms

1) Results on UAV123@10FPS: The results on the UAV123@10FPS dataset are shown in Fig. 3. We utilize the precision plots and success plots to compare the performance of each algorithm. It can be seen that the precision rate and AUC of the proposed model are 0.654 and 0.481, respectively. Our algorithm is ranked third on precision plots and first on success plots. The proposed DTSRT, AutoTrack, and STRCF all add regularization terms to mitigate boundary effects. Considering that precision rate is limited by preset threshold, so AUC is generally used to assess the overall performance of the algorithm. By comparison, our model achieves the best performance. It can be seen that the use of deep features has a great effect on the improvement of accuracy. AutoTrack automatically adjusts both temporal regularization and spatial regularization, enabling adaptive changes in the relevant hyperparameters, thus saving the step of manually fine-tuning these hyperparameters, and it has the second highest AUC at 0.477. KCF uses only manual features and does not have any effective treatment for boundary effects, making it difficult to achieve satisfactory performance in UAV videos. In conclusion, the proposed DTSRT achieves satisfactory performance on this dataset.

2) Results on UAV20L: The last two plots in Fig. 3 show the experimental results on UAV20L dataset. Since this dataset has only 20 image sequences and the challenge is homogenous, i.e., long-term motion. So this dataset is generally used to test whether the model is able to track the target consistently. In the precision plots, the proposed DTSRT has an accuracy of 0.610, which is 0.026 higher than the second place. In the success plots, the proposed DTSRT has a success rate of 0.428, which is 0.013 higher than the second place. The proposed models all achieve the optimal performance. This implies that our model has robust long-term tracking capability. BACF uses real samples to train the filter and thus also achieves better performance, with an AUC of 0.415, ranking second. ECO_HC though improves the discriminative ability of the classifier by fusing HOG features and CN features. But its AUC only reaches 0.387, which shows that the use of multiple features is not very useful in UAV videos.

3) Results on UAVDT: The targets to be tracked in the UAVDT dataset are mostly vehicles. In addition, some of the image sequences in this dataset were taken at night or in fog, resulting in blurred targets. As a result, the results on this dataset are not as good as the results on the other datasets. As can be seen in Table I, our model still performs well with an AUC of 0.431, second only to BACF. But in terms of precision rate, our model is 0.022 higher than BACF. It can be seen that our model can accurately localize the position of the target, but cannot efficiently fit the scale of the target. This is mainly because the proposed DTSRT still uses HOG features to predict the scale of the target. In order to prevent the model speed from being seriously affected, we only use deep features when predicting the target location. Overall, the proposed model significantly outperforms other algorithms on the UAVDT dataset.

4) Results on DTB70: To further compare the advantages and disadvantages between the algorithms, we conducted additional experiments on the DTB70 dataset. The results of each algorithm can be seen in Table II. Considering that the DTB70 dataset consists of 70 short image sequences, the overall performances are all satisfactory. AUC is generally used to rank individual algorithms. It can be seen that the proposed DTSRT has an AUC of 0.474 and is ranked first. The AUCs of MSCF and ECO_HC are 0.458 and 0.453, ranking second and third, respectively. Compared to our model, their AUCs are reduced by 0.016 and 0.021, respectively. In addition, our algorithm still outperforms other algorithms in terms of precision rate, with an accuracy of 0.676. Overall, the proposed DTSRT achieves outstanding performance on the DTB70 dataset.

5) Visualization Results: In order to visualize the tracking performance of the proposed DTSRT, we show the visualization

TABLE I
RESULTS ON UAVDT DATASET

	Ours	MCCT_H	ECO_HC	STRCF	CSRDCF	SRDCF	BACF	DCF	KCF
Precision	0.708	0.667	0.681	0.629	0.674	0.658	0.686	0.559	0.571
AUC	0.431	0.402	0.410	0.411	0.389	0.419	0.433	0.288	0.290

TABLE II
RESULTS ON DTB70 DATASET

	Ours	MSCF	MCCT_H	ECO_HC	STRCF	CSRDCF	SRDCF	BACF	KCF
Precision	0.676	0.664	0.604	0.643	0.649	0.646	0.512	0.590	0.468
AUC	0.474	0.458	0.405	0.453	0.437	0.438	0.363	0.402	0.280

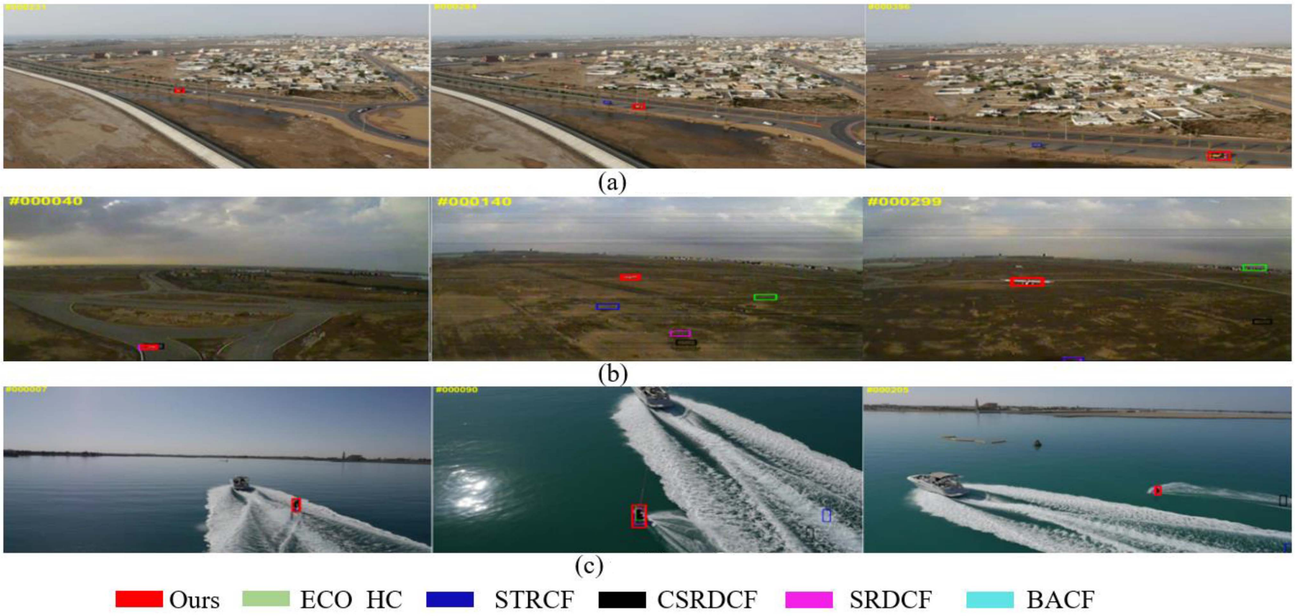

Fig. 4. Visualization results. (a) truck2. (b) UAV1. (c) walkboard2.

results of some algorithms on truck2, uav1, and walkboard2 image sequences. As illustrated in Fig. 4, we have selected a few challenging image sequences to better visualize and compare the differences between the algorithms. In truck2, the tracked target is small, occupying just a few tens or hundreds of pixels. In this case, it is difficult for the algorithm to extract efficient features. As a result, all algorithms except ours fail to track. Given that our algorithm uses deep features to model the target, it is highly discriminative. Despite the low resolution of the target, the proposed DTSRT accurately localizes the target. In uav1, the target being tracked is a drone. The target has been moving rapidly, which poses a huge challenge to the algorithm. In addition, clutter occurs frequently in this image sequence. It can be seen that only our algorithm can accurately localize the target, and all other algorithms fail to track. On the one hand, deep features enhance the representational information of the target. On the other hand, the dynamic regularization term mitigates the interference of clutter. Although STRCF also introduces a regularization term, it is clear that the static regularization term is difficult to deal with clutter interference effectively. In walkboard2, the target has been moving erratically and with occasional nonrigid deformations. When deformation occurs, interference from background noise is inevitably introduced into the original bounding box. Our model introduces a regularization term to penalize the background region, so it can keep tracking the target successfully. BACF, however, suffers from background interference and eventually loses the target.

6) Results in Different Attributes: In order to verify the generalization ability of the proposed algorithm, we test its performance in different challenge attributes, and the results are shown in Fig. 5. Since scale variation and aspect ratio variation occur frequently in UAV videos, we first conduct experiments in these two scenarios. It can be seen that our algorithm achieves the best performance with AUCs of 0.447 and 0.421, respectively. Due to the use of deep features, the discriminative ability of the algorithm is greatly improved, so when the target is more similar to the background, our algorithm can still accurately localize the target. In addition, our algorithms are still applicable to other scenarios, such as camera motion, fast motion, etc. However, when occlusion occurs, our algorithm is unable to continue the tracking because no redetection mechanism is introduced, which

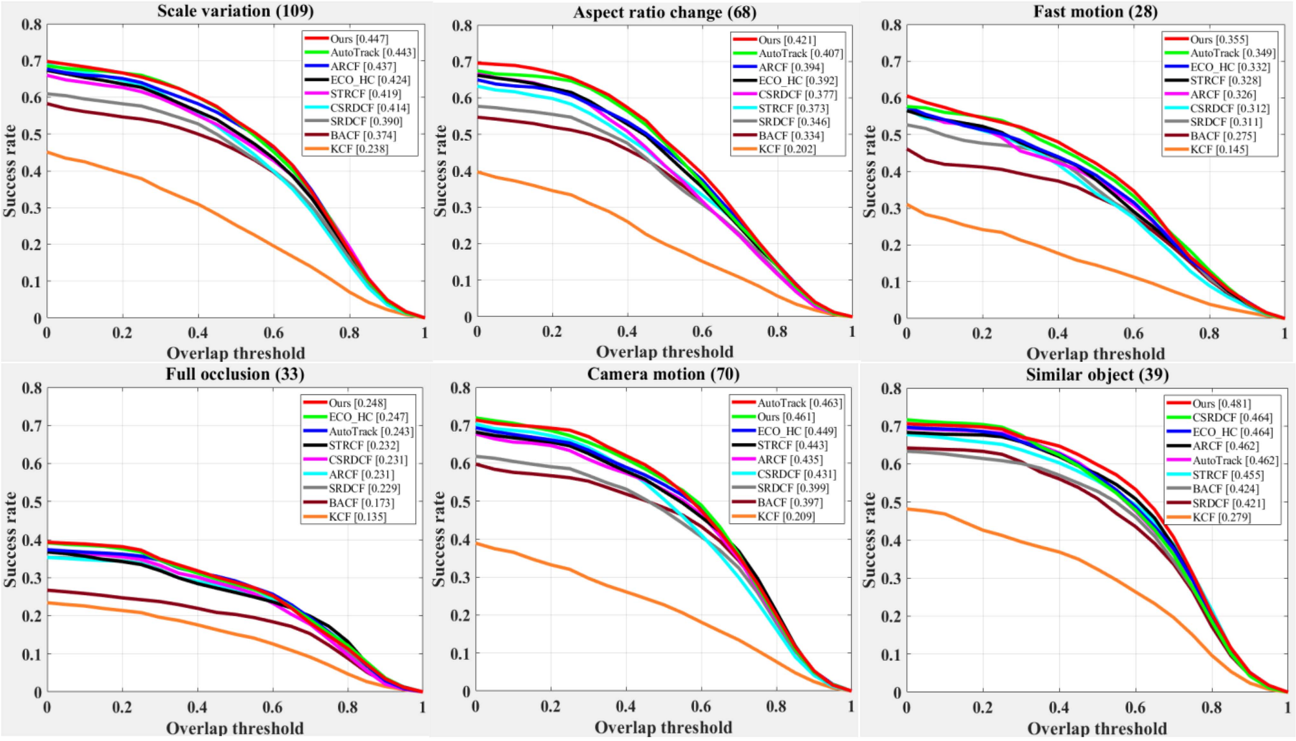

Fig. 5. Results in different attributes. The values in legends are the AUC.

TABLE III
RESULTS OF ABLATION EXPERIMENTS ON UAV123@10FPS DATASET

	Baseline	+ w_s	+ w_s + CNN	DTSRT
Precision	0.627	0.649	0.661	0.654
AUC	0.457	0.467	0.487	0.481
FPS	32.3	28.6	5.2	9.7

ultimately leads to tracking failure. It can be seen that the AUC is only 0.248 in the occlusion scenario.

D. Ablation Studies

In order to verify the effectiveness of the proposed adaptive spatial regularization w_s and deep features, we conducted a series of ablation experiments. Specifically, we introduce w_s on top of baseline to verify the performance gain that adaptive spatial regularization brings to the model. As can be seen in Table III, the precision rate of the algorithm has improved by 0.022 and the AUC has improved by 0.01. It can be seen that the introduction of adaptive spatio-temporal regularization can effectively mitigate the interference of the background, and its plays an active role in UAV videos. We then replaced the handmade features in that experiment with deep features, the proposed DTSRT. It should be noted in particular that at this point we only use deep features when predicting the target location, and we still use HOG features when predicting the scale. It can be seen that the AUC improved by 0.014 and the accuracy improved by 0.005. We can conclude that deep features have a stronger characterization ability compared to handcrafted features. However, it is easy to see that while deep features bring performance gains they also lead to a decrease in the speed of the algorithm. FPS dropped from 28.6 to 9.7. Finally, in order to further exploit the performance of deep features, we also use deep features when predicting the scale, i.e., baseline + w_s + CNN. Not surprisingly, the precision rate and AUC are 0.661 and 0.487, respectively, achieving the optimal performance among all the experiments. However, its drawback is also obvious, with a speed of only 5.2. To summarize, we ended up using deep features only for predicting position, and still used HOG features for predicting size.

V. CONCLUSION

In this article, we propose adaptive spatial regularization correlation filters, called DTSRT, for UAV tracking. Given the large interframe distances between targets in UAV videos, it is difficult for the original static spatial regularization to be fully effective. So we generate a dynamic spatial regularization constraint with the help of the existing saliency detection method. Dynamic constraints can update the regularization template in real time according to the target position and it is more suitable for UAV scenarios. In addition, we use deep features rather than handcrafted features to model the target due to the strong representational and anti-interference capabilities of deep features. As can be seen from Table III, the deep features improve the performance while reducing the real-time performance of the model. Therefore, we still use HOG features in predicting the target scale. Experiments on some classical UAV datasets show that the proposed DTSRT has robust tracking performance.

The proposed DTSRT has high accuracy but it lacks occlusion detection mechanism. This also means that it is difficult for our model to continue tracking when occlusion occurs. In future work, we plan to focus on exploring redetection algorithms applicable to UAV videos. In this case, the long-term tracking capability of the algorithm can be greatly improved.

VI. Discussion

Remote sensing touches all aspects of our lives [46], and UAVs can be used to collect near-earth remote sensing data. Based on this, object tracking in UAV videos has been attracting the attention of many scholars. However, unlike traditional object tracking, targets in UAV videos tend to have low resolution, which makes it difficult for algorithms to efficiently model targets. To enhance the representational information of the target, many algorithms [47] fuse multiple features to improve their accuracy in multiple scenarios. Although the use of multiple features improves the algorithm's discriminative ability to some extent, manual features are only applicable to a single scenario. In complex scenarios, these algorithms often perform poorly.

Due to the powerful modeling capabilities of deep learning, we use deep features to describe the target, which greatly improves the robustness of the algorithm. As can be seen in Fig. 5, the tracked targets are not only of low resolution, but are also accompanied by challenges such as fast motion and deformation. Among all the algorithms, ours is the only one that can track the target accurately. In contrast, ECO_HC uses a variety of features, but it still ends up failing to track. Therefore, deep features have better modeling ability for small targets in UAV videos. However, the use of deep features often comes with a huge time overhead, so in this article, deep features are only used to predict the location of the target. For the scale of the target, we still use HOG features. From Table III, we can see that the use of HOG features enhances the speed of the algorithm.

In addition to this, boundary effects have to be taken into account, especially in UAV videos, which are very likely to lead directly to tracking failures. BACF [19] mitigates the boundary effect by using real samples to train the filter and extend the search region. STRCF [17] suppresses boundary effects by introducing a regularization term. However, in UAV videos, the target moves faster and the relative motion of the target and the camera can result in larger distances between neighboring frames. Therefore, static regularization methods are difficult to be fully effective. To solve the above problem, we introduce a dynamic regularization term that generates constraints in real time based on the location of the target. SRDCF [16], STRCF [17], CSRDCF [45] all use static regularization terms, while we use dynamic regularization terms. As can be seen from Fig. 5, our algorithm significantly outperforms other algorithms in a variety of challenging scenarios, such as similar object, fast motion, and so on.

Although the proposed algorithm, DTSRT, achieves excellent performance on several UAV datasets, it is difficult to apply it to occlusion scenarios. As shown in Fig. 5, the AUC of our algorithm is only 0.248 in occlusion scenario. Since occlusion occurs frequently in UAV videos, it is necessary to introduce a redetection mechanism to mitigate the adverse effects caused by occlusion. In future work, we intend to apply the object detection technique in the proposed algorithm. Specifically, when occlusion occurs, we use the detection algorithm to redetect the target location as a way to improve the long-term tracking capability of the algorithm.

References

[1] L. Shi, Q. Zhang, B. Pan, J. Zhang, and Y. Su, "Global-local and occlusion awareness network for object tracking in UAVs," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8834–8844, 2023.

[2] X. Xu et al., "STN-track: Multiobject tracking of unmanned aerial vehicles by swin transformer neck and new data association method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8734–8743, 2022.

[3] J. Cui, M. Liu, Z. Zhang, S. Yang, and J. Ning, "Robust UAV thermal infrared remote sensing images stitching via overlap-prior-based global similarity prior model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 270–282, 2021.

[4] J. Zhu et al., "Urban traffic density estimation based on ultrahigh-resolution UAV video and deep neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4968–4981, Dec. 2018.

[5] S. Liu, J. Cheng, L. Liang, H. Bai, and W. Dang, "Light-weight semantic segmentation network for UAV remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8287–8296, 2021.

[6] Y. Wu, J. Lim, and M. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.

[7] X. Xu et al., "STN-Track: Multiobject tracking of unmanned aerial vehicles by swin transformer neck and new data association method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8734–8743, 2022.

[8] H. Wang, W. Ma, S. Zhang, and W. Hao, "Hierarchical feature pooling transformer for efficient UAV object tracking," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 6010405.

[9] Y. Xue et al., "SmallTrack: Wavelet pooling and Graph enhanced classification for UAV small object tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5618815.

[10] B. Li, Z. Gan, D. Chen, and D. S. Aleksandrovich, "UAV maneuvering target tracking in uncertain environments based on deep reinforcement learning and meta-learning," *Remote Sens.*, vol. 12, no. 22, pp. 3789–3808, Nov. 2020.

[11] Y. Bai et al., "Occlusion and deformation handling visual tracking for UAV via attention-based mask generative network," *Remote Sens.*, vol. 14, no. 19, pp. 4756–4776, Oct. 2022.

[12] H. Wu, Z. He, and M. Gao, "GCEVT: Learning global context embedding for vehicle tracking in unmanned aerial vehicles videos," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 6000705.

[13] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[14] F. Lin, C. Fu, Y. He, F. Guo, and Q. Tang, "Learning temporary blocked-based bidirectional incongruity-aware correlation filters for efficient UAV object tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 6, pp. 2160–2174, Jun. 2021.

[15] Z. An, X. Wang, B. Li, and J. Fu, "Learning spatial regularization correlation filters with the Hilbert–Schmidt independence criterion in RKHS for UAV tracking," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 5011612.

[16] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4310–4318.

[17] F. Li, C. Tian, W. Zuo, L. Zhang, and M.-H. Yang, "Learning spatial-temporal regularized correlation filters for visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4904–4913.

[18] H. K. Galoogahi, T. Sim, and S. Lucey, "Correlation filters with limited boundaries," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4630–4638.

[19] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning background-aware correlation filters for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1144–1152.

[20] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop*, 2015, pp. 621–629.

[21] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ECO: Efficient convolution operators for tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6931–6939.

[22] G. Bhat, J. Johnander, M. Danelljan, F. S. Khan, and M. Felsberg, "Unveiling the power of deep tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 493–509.

[23] Y. Li, C. Fu, F. Ding, Z. Huang, and G. Lu, "AutoTrack: Towards highperformance visual tracking for UAV with automatic spatio-temporal regularization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11923–11932.

[24] C. Fu, J. Xu, F. Lin, F. Guo, T. Liu, and Z. Zhang, "Object saliency aware dual regularized correlation filter for real-time aerial tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8940–8951, Dec. 2020.

[25] Q. Wang, J. Gao, J. Xing, M. Zhang, and W. Hu, "DCFNet: Discriminant correlation filters network for visual tracking," 2017, *arXiv:1704.04057*.

[26] E. Park and A. C. Berg, "Meta-tracker: Fast and robust online adaptation for visual object trackers," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 569–585.

[27] N. Wang, Y. Song, C. Ma, W. Zhou, W. Liu, and H. Li, "Unsupervised deep tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1308–1317.

[28] N. Wang, W. Zhou, Y. Song, C. Ma, and H. Li, "Real-time correlation tracking via joint model compression and transfer," *IEEE Trans. Image Process.*, vol. 29, pp. 6123–6135, 2020.

[29] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 6, pp. 1789–1819, 2021.

[30] J. Chen, T. Xu, B. Huang, Y. Wang, and J. Li, "ARTracker: Compute a more accurate and robust correlation filter for UAV tracking," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6514605.

[31] L. Wang et al., "Auto-Perceiving correlation filter for UAV tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5748–5761, Sep. 2022.

[32] Y. Li, C. Fu, Z. Huang, Y. Zhang, and J. Pan, "Intermittent contextual learning for keyfilter-aware UAV object tracking using deep convolutional feature," *IEEE Trans. Multimedia*, vol. 23, pp. 810–822, 2021.

[33] F. Zhang, S. Ma, Y. Zhang, and Z. Qiu, "Perceiving temporal environment for correlation filters in real-time UAV tracking," *IEEE Signal Process. Lett.*, vol. 29, pp. 6–10, 2022.

[34] F. Lin, C. Fu, Y. He, W. Xiong, and F. Li, "ReCF: Exploiting response reasoning for correlation filters in real-time UAV tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 10469–10480, Aug. 2022.

[35] Y. Li, H. Zhang, Y. Yang, H. Liu, and D. Yuan, "RISTrack: Learning response interference suppression correlation filters for UAV tracking," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 8000705.

[36] X. Xue, Y. Li, X. Yin, C. Shang, T. Peng, and Q. Shen, "Semantic-aware real-time correlation tracking framework for UAV videos," *IEEE Trans. Cybern.*, vol. 52, no. 4, pp. 2418–2429, Apr. 2022.

[37] H. Zhang et al., "UAV tracking based on correlation filters with dynamic aberrance-repressed temporal regularizations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7749–7762, 2023.

[38] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.

[39] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 445–461.

[40] D. Du et al., "The unmanned aerial vehicle benchmark: Object detection and tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 370–386.

[41] S. Li and D.-Y. Yeung, "Visual object tracking for unmanned aerial vehicles: A benchmark and new motion models," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4140–4146.

[42] Z. Huang, C. Fu, Y. Li, F. Lin, and P. Lu, "Learning aberrance repressed correlation filters for real-time UAV tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 2891–2900.

[43] G. Zheng, C. Fu, J. Ye, F. Lin, and F. Ding, "Mutation sensitive correlation filter for real-time UAV tracking with adaptive hybrid label," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2021, pp. 503–509.

[44] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue correlation filters for robust visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4844–4853.

[45] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6309–6318.

[46] L. Yan, B. Fan, H. Liu, C. Huo, S. Xiang, and C. Pan, "Triplet adversarial domain adaptation for pixel-level classification of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3558–3573, May 2020.

[47] C. Fu, F. Lin, Y. Li, and G. Chen, "Correlation filter-based visual tracking for UAV with online multi-feature learning," *Remote Sens.*, vol. 11, no. 5, pp. 549–571, 2019.

Yulin Cao was born in Qinghai, China, in 1969. He received the B.S. degree in mathematics from the Department of Qinghai Normal University, Xining, China, in 1994, and the Ph.D. degree in computer science from the Shanxi Normal University, Xian, China, in 2019.

His research areas include mobile ad-hoc networks, social computing, and intelligent information processing.

Shihao Dong was born in 1995. He received the B.S. degree in medical information engineering from Xuzhou Medical University, Xuzhou, China, in 2018, and the M.S. degree in computer science from Zhejiang Normal University, Jinhua, China, in 2021. He is currently working toward the Ph.D. degree in computer science from Nanjing University of Information Science and Technology, Nanjing, China.

His research interests include cluster analysis and deep learning.

Jiawei Zhang was born in 1996. He received the B.S. degree in 2019, software engineering from Binjiang College, Nanjing University of Information Science and Technology, Nanjing, China, where he is currently working toward the Ph.D. degree in information security.

His research interests include multimedia security and deep learning security.

Han Xu was born in 1998. He received the B.E. degree in computer science and technology from Zhengzhou University of Aeronautics, Zhengzhou, China, in 2020. He is currently working toward the M.E. degree in computer technology with Nanjing University of Information Science and Technology, Nanjing, China.

His current research interests include computer vision and object tracking.

Yan Zhang was born in 1998. He received the B.E. degree in computer science and technology from Nanjing University of Information Science and Technology, Nanjing, China, in 2021. He is currently working toward the M.E. degree in computer science and technology with Nanjing University of Information Science and Technology, Nanjing.

His current research interests include multimedia processing and object tracking.

Yuhui Zheng was born in Shanxi, China, in 1982. He received the B.S. degree in chemistry and the Ph.D. degree in computer science from Nanjing University of Science and Technology, Nanjing, China, in 2004 and 2009, respectively.

His main research areas include image and video analysis, scene understanding, visual tracking, and pattern recognition. From 2014 to 2015, he was a Visiting Professor with the Digital Media Laboratory, School of Electronic and Electrical Engineering, Sungkyunkwan University, Seoul, South Korea. He is currently a Full Professor with the School of Computer, Nanjing University of Information Science and Technology, Nanjing, China.