

## 多特征感知的时空自适应相关滤波目标跟踪

孟庆姣, 姜文涛

引用本文

孟庆姣, 姜文涛. 多特征感知的时空自适应相关滤波目标跟踪[J]. 计算机科学, 2023, 50(11A): 230200096-9.

MENG Qingjiao, JIANG Wentao. [Multi-feature-aware Spatiotemporal Adaptive Correlation Filtering Target Tracking](#) [J]. Computer Science, 2023, 50(11A): 230200096-9.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[一种融合CNN和Swin Transformer的医学显微图像分割模型](#)

Medical Microscopic Image Segmentation Model Based on CNN Structure and Swin Transformer  
计算机科学, 2023, 50(11A): 230200119-8. <https://doi.org/10.11896/jsjcx.230200119>

[基于GRU与自注意力网络的声源到达方向估计](#)

Sound Source Arrival Direction Estimation Based on GRU and Self-attentive Network  
计算机科学, 2023, 50(11A): 220900135-7. <https://doi.org/10.11896/jsjcx.220900135>

[基于替代模型的批量零阶梯度符号算法](#)

Batch Zeroth Order Gradient Symbol Method Based on Substitution Model  
计算机科学, 2023, 50(11A): 230100036-6. <https://doi.org/10.11896/jsjcx.230100036>

[面向边缘计算的轻量级网络硬件加速设计](#)

Lightweight Network Hardware Acceleration Design for Edge Computing  
计算机科学, 2023, 50(11A): 220800045-7. <https://doi.org/10.11896/jsjcx.220800045>

[基于注意力机制和ConvLSTM的船舶交通流量预测算法](#)

Ship Traffic Flow Prediction Algorithm Based on Attention Mechanism and ConvLSTM  
计算机科学, 2023, 50(11A): 230800067-7. <https://doi.org/10.11896/jsjcx.230800067>

# 多特征感知的时空自适应相关滤波目标跟踪

孟庆姣 姜文涛

辽宁工程技术大学软件学院 辽宁 葫芦岛 125105

(3446628847@qq.com)

**摘要** 针对正则化滤波器预先定义正则化项,但无法实时抑制非目标区域学习的缺点,提出了一种多特征感知的时空自适应相关滤波目标跟踪的新方法。首先在目标函数中引入空间局部响应变化量实现空间正则化,使滤波器专注于学习对象中值得信任的部分,从而得到响应模型;其次根据全局响应变化决定滤波器的更新率;最后通过级联颜色直方图(Colour Name,CN)与降维后的梯度直方图(Fast Histogram of Oriented Gradient,fHOG)特征实现非卷积特征层面的融合,采用 ImageNet-VGG-2048 的 Conv1,Conv5 层提取目标的空间轮廓以及语义信息,并使用 ReLU 函数拟合训练数据,在保留主要信息的同时提高速率。在数据集 DTB70 上的精确率(0.747)和成功率(0.789)相较于 STRCF 算法的精确率(0.737)和成功率(0.760)分别提高了 1% 和 2.9%。大量实验证明该算法在复杂背景、物体遮挡、快速运动等多种场景下基本能满足实时性需求。

**关键词:** 目标跟踪;相关滤波;时空自适应;局部响应与全局响应;卷积神经网络;特征融合

**中图分类号** TP391

## Multi-feature-aware Spatiotemporal Adaptive Correlation Filtering Target Tracking

MENG Qingjiao and JIANG Wentao

School of Software,Liaoning Technical University,Huludao,Liaoning 125105,China

**Abstract** Aiming at the disadvantage that the regularization filter defines the regularization term in advance but cannot suppress the learning of non-target region in real time,a new method of multi-feature-aware spatiotemporal adaptive correlation filtering target tracking is proposed. Firstly,the spatial local response variation is introduced into the objective function to realize the spatial regularization,so that the filter can focus on the trustworthy part of the learning target,and then the response model is obtained. Secondly,the update rate of the filter is determined according to the change of the global response. Finally,the non-convolution feature level fusion is realized by cascading color histograms(CN) and dimensionally reduced gradient histograms(fHOG). The conv1 and conv5 layers of imagenet-vgg-2048 are used to extract the spatial contour and semantic information of the target. The ReLU function is used to fit and train the data to improve the speed while retaining the main information. Results:In this paper,we compared 8 algorithms of the same type,and used the defined baseline algorithm STRCF(2018) in the objective function,and KCF(2014),which introduces gauss kernels to increase computational speed and sample circularly using a cyclic matrix,and MOSSE\_CA(2021),which links context and scale filters,and DCF\_CA(2017),which increases the number of samples but reduces the search area Staple(2016) with temporal regularization;region constraint to reduce anomalous ARCF(2019);correlation filter HSTDCA(2021) with hierarchical spatiotemporal map regularization;and target segmentation into four blocks,the SAME\_CA(2020) of the scale factor is calculated by using the kernel correlation filter to find the maximum response position of each block. Compared with the accuracy(0.737) and success rate(0.760) of STRCF algorithm,the accuracy rate(0.747) and success rate(0.789) of DTB70 were increased by 1% and 2.9% respectively. Conclusion:The image information learned after multi-layer feature fusion is updated to obtain the overall contour,so as to adaptively track the target. A large number of experiments show that the algorithm basically meets the real-time requirements in complex background,object occlusion,fast motion and other scenarios.

**Keywords** Target tracking, Correlation filter, Spatio-Temporal adaptation, Local response and global response, Convolutional neural network, Feature fusion

## 1 引言

目标跟踪是计算机视觉领域的一个热点问题<sup>[1]</sup>,其因普

遍与实用性,在无人机驾驶、现代化军事<sup>[2]</sup>等领域具有广泛的应用。目标跟踪是利用样本的信息对目标进行特征提取从而训练滤波器进而达到定位跟踪的目的。但由于其利用的是在

基金项目:国家自然科学基金(61172144);辽宁省自然科学基金(20170540426);辽宁省教育厅基金(LJYL049)

This work was supported by the National Natural Science Foundation of China(61172144),Natural Science Foundation of Liaoning Province, China(20170540426) and Education Department Foundation of Liaoning Province(LJYL049)

通信作者:姜文涛(2645647850@qq.com)

线学习的方式,因此当目标物体存在不同程度的尺度、形状、速度变化时,会导致复杂场景下无法实现准确的跟踪,所以该领域仍面临着巨大的挑战。近年来,相关滤波因在速率和精度两方面均具有较强的鲁棒性而受到广大专家学者的认可,并在解决上述问题的过程中,取得了较好的成果。Bolme等<sup>[3]</sup>将相关滤波的概念引入目标跟踪过程,通过相关性衡量信号的相似度,将卷积算子由空域计算转换到频域计算,提出了最小输出平方和误差滤波跟踪算法,在特征提取阶段采用单通道的灰度特征。因为只能提取局部信息,所以跟踪结果并不理想。Henriques等<sup>[4]</sup>于2012年提出核循环结构CSK算法,将矩阵作用于整幅图像,随着样本数量的增多,精确率有所提高。Henriques等<sup>[5]</sup>又于2014年提出了核相关滤波算法,提取方式采用了HOG特征,并在目标函数中加入了高斯核,进行非线性到空间线性之间的转换,大大地减少了计算时间。

综上,当滤波器所学习的目标出现非匀速运动和非平缓外观时,边界效应就此产生,迭代计算后导致跟踪失败。为此,Galoogahi等<sup>[6]</sup>提出有限边界的相关性滤波算法,因为所训练的图像块过大但样本有限,导致学习模型过度拟合,降低了准确率。Danelljan等<sup>[7-8]</sup>提出了基于空间正则化的相关滤波算法,记录每一帧样本信息,并根据不同的空间位置给予不同的权重比例,实时性与普适性较弱。Mueller<sup>[9]</sup>等提出上下文感知相关滤波器跟踪算法,重心放在改变相关滤波的框架上,通过学习目标中心的正样本信息再结合周围4个负样本的背景信息,得到新的图像样本信息。Li等<sup>[10]</sup>在SRDCF的基础上,提出了时空正则化相关滤波,该算法既提高了精确率又提高了速率。

随着相关滤波跟踪算法的迅速发展,特征提取方面也从单一的灰度特征、HOG特征到HOG与CN特征结合。Bertinetto等<sup>[12]</sup>提出背景直方图目标跟踪算法,在进行特征提取时结合HOG与CN两种特征,但高维的样本计算,导致其实用性较弱。为了更全面地利用深度特征,Ma等提出了一种结合3个不同卷积层的分层卷积特征(Hierarchical Convolutional Features, HCF),使用VGG(Visual Geometry Group)-19的Conv5-4, Conv4-4和Conv3-4的激活值为特征<sup>[13]</sup>,充分利用浅层和深层的信息,但其因速率较低,使用性较差,未得到广泛使用。Danelljan等提出用卷积神经网络中单层卷积层的深度特征代替SRDCF中传统的HOG特征<sup>[14]</sup>,实验表明跟踪性能得到大幅度提高。

针对STRCF相关滤波器中目标函数中预定义的正则化项无法适应新状况出现<sup>[15]</sup>,以及特征融合的方式导致部分特征信息丢失的问题,本文提出了多特征融合的时空自适应相关滤波跟踪算法:1)根据相邻两帧间的信息可以反映前一帧所学习到的外观模型和当前帧检测到的实际对象之间的相似性。根据局部变化的可信度判断滤波器学习的像素。2)结合空间惩罚缓解边界效应,既能保留较完整的边界,又能够突出学习区域,得到整体轮廓。实现利用局部和全局响应变化进行空间和时间同时正则化的目的。对自适应参数进行优化,避免不必要的学习,强化必要的学习,从而得到一个更加强壮的外观模型。3)基于STRCF相关滤波器的框架,采用fHOG特征、CN特征和ImageNet-VGG-2048的Conv1和Conv5层卷积特征进行跟踪。

本文算法提高了目标跟踪过程中的自适应能力,同时也提高了跟踪速度。在数据集OTB50和DTB70上与8种主流算法进行多属性的实验,结果证明本文算法的跟踪鲁棒性更强。

## 2 时空正则化相关滤波目标跟踪算法原理

通过最小化以下目标函数学习第 $t$ 帧中的最佳滤波器:

$$\epsilon(H_t) = \frac{1}{2} \left\| y - \sum_{k=1}^K x_k^t * h_k \right\|_2^2 + \frac{1}{2} \sum_{k=1}^K \| u \cdot h_k \|_2^2 + \frac{\theta}{2} \sum_{k=1}^K \| h_k - h_k^{-1} \|_2^2 \quad (1)$$

其中, $*$ 表示进行卷积运算。 $x_k^t \in R^{T \times 1}$  ( $k=1, 2, 3, \dots, K$ )代表在跟踪第 $t$ 帧时提取到的特征样本, $K$ 为跟踪过程中每帧图片的通道数量, $T$ 代表进行计算的矩阵维度。 $H = [h_1^t, h_2^t, \dots, h_k^t]$ ,  $h_k^t \in R^{T \times 1}$ 为第 $t$ 帧训练的第 $k$ 通道的滤波器。

STRCF算法引用SRDCF算法中的空间正则项 $\frac{1}{2} \sum_{k=1}^K \| u \cdot h_k^t \|_2^2$ ,其中 $u$ 为负高斯形状的空间权重向量,用于使正则化权重从目标区域平稳地过渡到背景区域。引入时间正则项,其中 $\theta$ 为系数, $h^t$ 和 $h^{t-1}$ 表示当前帧和上一帧训练的滤波器。

STRCF跟踪算法中引入了具有固定参数的时间正则项,虽然给予边缘区域较低的权重适当缓解了边界效应,但忽视了边缘样本信息的重要性。在逐帧学习的过程中很难适应跟踪场景出现的新情况,最终导致跟踪结果错误。所以本文充分利用局部和全局响应的变化,自适应地学习样本信息,训练具有时空正则化的滤波器,实现不断学习过程中自适应跟踪的目的。

## 3 自适应时空正则化相关滤波目标跟踪算法

在目标检测过程中,响应图中的信息至关重要<sup>[16]</sup>,因为其信息可以反映前一帧所学习到的外观模型和当前帧检测到的实际对象之间的相似性。但主流算法只考虑整体响应图变化,忽略了样本信息中处于不同位置的局部响应变化的可信度不同而进行无差别地跟踪,导致计算难度加大,从而降低了实用性。

### 3.1 自适应时空正则化

首先,本文定义局部响应变化量, $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_i)$ 用于对象边界框中的可视化。其第 $i$ 个元素 $\sigma_i$ 定义为:

$$\sigma_i = \frac{R_i'[\phi_\Delta] - R_i'^{-1}}{R_i'^{-1}} \quad (2)$$

其中, $\phi_\Delta$ 是在两个响应图 $R'$ 和 $R'^{-1}$ 产生的两个峰值的位置差<sup>[16]</sup>, $[\phi_\Delta]$ 表示相邻峰值为达到高度相似而移动的距离。峰值位置由滤波器中进行运算的逐个像素决定,所以局部响应变化显示当前帧搜索区域中每个像素的可信度。当外观发生较大变化时会导致 $[\phi_\Delta]$ 的值变大。因相似而移动的距离变大,故相似性低,像素可信度低。在学习过程中,应限制学习可信度低的像素,本文通过将局部变化 $\sigma$ 引入空间正则化参数 $\tilde{u}$ 来实现这一点:

$$\tilde{u} = P^T \delta \log(\sigma + 1) + u \quad (3)$$

其中, $P^T \in R^{T \times T}$ 用于裁剪对象所在滤波器的中心部分。目标区域的样本质量较高且计算复杂度较低,大大地提高了计算速率。 $\delta$ 是一个常数,用于调整局部响应变化的权重,而 $u$ 是从STRCF中继承来的权重向量,用于缓解边界效应<sup>[17]</sup>。局部响应变量反应了当前帧局部位置每个像素的可信度,

$\log(\sigma+1)$ 作为单调递增函数,与 $\sigma$ 的增减保持相同趋势,放大了 $\sigma$ 的变化。通过式(3),将其引入到空间正则化参数中,则位于具有剧烈响应变化的像素处的滤波器将避免学习部分新外观,达到空间自适应的目的。在STRCF中,相邻两帧滤波器的变化率由一个固定参数 $\theta$ 决定,本文尝试通过联合优化其数值和滤波器,自适应确定参数的值,定义一个参考 $\tilde{\theta}$ 作为目标函数的全局响应:

$$\tilde{\theta} = \frac{1}{\|\sigma\|_2}, \tilde{\theta} \leq \varphi \quad (4)$$

引入欧几里德范数<sup>[18]</sup>来比较 $R^t$ 和 $R^{t-1}$ 两个响应图的差异水平。发生畸变时,相似性会突然下降,则 $\sigma_t$ 的值会升高。当全局响应变量高于阈值 $\varphi$ 时,相关滤波器停止学习;反之,低于阈值时,响应变化越剧烈, $\tilde{\theta}$ 值越小,从而可以放宽对相关滤波器时间变化的限制,达到时间自适应的目的。

### 3.2 目标函数模型

$$\begin{aligned} \epsilon(H_t, \theta_{\text{score}}^t) = & \frac{1}{2} \left\| y - \sum_{k=1}^K x_k^t * h_k^t \right\|_2^2 + \frac{1}{2} \sum_{k=1}^K \left\| \tilde{u} \cdot h_k^{t-1} \right\|_2^2 + \\ & \frac{\theta_{\text{score}}^t}{2} \sum_{k=1}^K \left\| h_k^t - h_k^{t-1} \right\|_2^2 + \frac{1}{2} \left\| \theta_{\text{score}}^t - \tilde{\theta} \right\|_2^2 \end{aligned} \quad (5)$$

其中, $\theta$ 和 $\theta_{\text{score}}^t$ 分别表示参考和优化的时间正则化参数, $\tilde{u}$ 表示由式(3)计算得到的自适应空间正则化项。有效缓解边界效应的同时,确保空间权重随提取样本信息的变化而变化,从而实现空间自适应的目的。根据帕塞瓦尔(Parseval)定理<sup>[17]</sup>,信号在时域和频域上的功率相等,引入辅助变量, $\hat{g}_k \in R^{TK \times 1}$ , $\hat{G} = [\hat{g}_1, \hat{g}_2, \dots, \hat{g}_K]$ ,符号 $\wedge$ 表示信号的离散傅里叶变换(DFT),可将式(5)转化到频域:

$$\begin{aligned} \epsilon(H_k, \theta_{\text{score}}^t, \hat{G}_k) = & \frac{1}{2} \left\| y - \sum_{k=1}^K \hat{x}_{k-1}^t \cdot \hat{h}_k^{t-1} \right\|_2^2 + \\ & \frac{1}{2} \sum_{k=1}^K \left\| \tilde{u} \cdot \hat{h}_k^{t-1} \right\|_2^2 + \frac{\theta_{\text{score}}^t}{2} \sum_{k=1}^K \left\| \hat{g}_k^t - \hat{g}_k^{t-1} \right\|_2^2 + \frac{\theta}{2} \left\| \theta_{\text{score}}^t - \tilde{\theta} \right\|_2^2 \end{aligned} \quad (6)$$

其中, $\hat{y} = \sqrt{TF}y$ , $\hat{x}_k = \sqrt{TF}x_k$ , $\hat{g}_k = \sqrt{TF}g_k$ 为傅里叶变换, $F \in R^{T \times T}$ 为正交矩阵。

### 3.3 目标函数优化

式(6)是凸光滑可微函数,本文采取交替方向乘法(ADMM)迭代求解<sup>[18]</sup>,将原问题的目标函数等价分解成若干个可求解的子问题。并行求解的方式,提高了跟踪速度。方程如下:

$$\begin{aligned} L_k(H_k, \theta_{\text{score}}^t, \hat{G}_k, \hat{M}_k) = & \epsilon(H_k, \theta_{\text{score}}^t, \hat{G}_k) + \frac{\gamma}{2} \sum_{k=1}^K \left\| \hat{g}_k^t - \sqrt{TF}h_k^t \right\|_2^2 + \sum_{k=1}^K \left( \hat{g}_k^t - \sqrt{TF}h_k^t \right)^T \hat{m}_k^t \end{aligned} \quad (7)$$

其中, $\hat{M} = [\hat{m}_1, \hat{m}_2, \dots, \hat{m}_K] \in R^{T \times K}$ 为拉格朗日乘子; $\gamma$ 为步长参数。上角标 $T'$ 表示矩阵的共轭转置。

引入 $v_k^t = \frac{1}{\gamma} \hat{m}_k^t$ , $V = [v_1, v_2, \dots, v_K]$ ,式(7)可转化为:

$$\begin{aligned} L'(H_k, \theta_{\text{score}}^t, \hat{G}_k, \hat{M}_k) = & \epsilon(H_k, \theta_{\text{score}}^t, \hat{G}_k) + \frac{\gamma}{2} \sum_{k=1}^K \left\| \hat{g}_k^t - \sqrt{TF}h_k^t + v_k^t \right\|_2^2 \end{aligned} \quad (8)$$

通过使用ADMM算法,等式可被分解为 $\hat{G}$ , $\theta_{\text{score}}^t$ , $H$ 3个

子问题,式(5)可通过对这3个子问题进行迭代解出最优解。

1)求解 $\hat{G}$

假设给定 $H_k, \theta_{\text{score}}^t, \hat{V}_k, \hat{G} = [\hat{g}_1, \hat{g}_2, \dots, \hat{g}_K]$ ,目标函数为:

$$\hat{G}^* = \arg \min \left\{ \frac{1}{2} \left\| y - \sum_{k=1}^K \hat{x}_k^t \cdot \hat{g}_k^t \right\|_2^2 + \frac{\theta_{\text{score}}^t}{2} \sum_{k=1}^K \left\| \hat{g}_k^t - \hat{g}_k^{t-1} \right\|_2^2 + \frac{\gamma}{2} \sum_{k=1}^K \left\| \hat{g}_k^t - \sqrt{TF}h_k^t + v_k^t \right\|_2^2 \right\} \quad (9)$$

由上式可以看出,如果直接对 $\hat{g}_k^t$ 求导,第一项的求和和运算会增加较大的计算复杂度。由于 $\hat{x}_k$ 为稀疏带状变量,并且 $\hat{y}$ 只依赖于 $\hat{x}_k = [\hat{x}_k^1, \hat{x}_k^2, \dots, \hat{x}_k^K]^T$ 和 $\hat{g}_k = [\text{conj}(\hat{g}_k^1), \text{conj}(\hat{g}_k^2), \dots, \text{conj}(\hat{g}_k^K)]^T$ 中的元素,其中 $\text{conj}(\cdot)$ 表示复共轭运算,所以可以考虑通过单独计算 $\hat{y}$ 中的元素即 $\hat{y}_k (k=1, 2, \dots, K)$ 来简化计算。记 $\zeta_j(\hat{G}) \in R^{K \times 1}$ ,表示在像素 $j$ 上 $\hat{x}$ 的所有 $K$ 通道值的向量,解:

$$\begin{aligned} \zeta_j^*(\hat{G}^t) = & \arg \min \left\{ \left\| \hat{y}_j - \zeta_j(\hat{X}^t)^T \zeta_j(\hat{G}^t) \right\|_2^2 + \left\| \zeta_j(\hat{G}^t) + \zeta_j(\hat{V}^t) - \zeta_j(\sqrt{TF}H^t) \right\|_2^2 + \theta_{\text{score}}^t \left\| \zeta_j(\hat{G}^t) - \zeta_j(\hat{G}^{t-1}) \right\|_2^2 \right\} \end{aligned} \quad (10)$$

上式时间复杂度为 $O(TK^3)$ ,使用Sherman-Morrison公式推导后,其解可以转化为:

$$\zeta_j^*(\hat{G}^t) = \frac{1}{\gamma T + \theta_{\text{score}}^t} \left( I - \frac{\zeta_j(\hat{X}^t) \zeta_j(\hat{X}^t)^T}{\theta_{\text{score}}^t + \gamma T + \zeta_j(\hat{X}^t)^T v_j(\hat{X}^t)} \right) \rho \quad (11)$$

其中, $\rho = \zeta_j(\hat{X}^t) y_j + \theta_{\text{score}}^t \zeta_j(\hat{G}^{t-1}) - \gamma \zeta_j(\hat{V}^t) + \gamma \zeta_j(\sqrt{TF}H^t)$ 。此时的时间复杂度为 $O(TK)$ ,显著地缩短了程序计算的时间。

2)求解 $H$

假设给定 $\theta_{\text{score}}^t, \hat{G}^t, \hat{V}^t, H \in R^{T \times 1}$ ,其优化 $h_k$ 可以表示为:

$$\arg \min \left\{ \frac{1}{2} \left\| \tilde{u} \cdot h_k^t \right\|_2^2 + \frac{\gamma}{2} \left\| \hat{g}_k^t - \sqrt{TF}h_k^t + v_k^t \right\|_2^2 \right\} \quad (12)$$

$h_k$ 的封闭解为:

$$\begin{aligned} h_k^* = & [\tilde{U}^T \tilde{U} + \gamma T]^{-1} \gamma (\sqrt{TF})^T (v_k^t + \hat{g}_k^t) \\ = & \frac{\gamma T (v_k^t + \hat{g}_k^t)}{(\tilde{u} \cdot \tilde{u}) + \gamma T} \end{aligned} \quad (13)$$

其中, $\tilde{U} = \text{diag}(\tilde{u}) \in R^{T \times T}$ ,由式(13)可以看出对 $h_k$ 的求解过程仅需逐元相乘和快速傅里叶<sup>[19]</sup>逆变换(IFFT)即可得到,即 $v_k = \frac{1}{\sqrt{T}} F^T v_k^t, g_k = \frac{1}{\sqrt{T}} F^T g_k^t$ 。所以,对 $H$ 进行求解的时间复杂度为 $O(KT \lg T)$ 。

3)求解优化的时间正则化参数 $\theta_{\text{score}}^t$

给定式(8)中的其他变量 $H, \hat{G}, \hat{V}$ ,对 $\theta_{\text{score}}^t$ 采用ADMM算法求 $\theta_{\text{score}}^t$ 的最优解可确定为:

$$\begin{aligned} \theta_{\text{score}}^t = & \arg \min \left\{ \frac{\theta}{2} \sum_{k=1}^K \left\| \hat{g}_k^t - \hat{g}_k^{t-1} \right\|_2^2 + \frac{1}{2} \left\| \theta_{\text{score}}^t - \tilde{\theta} \right\|_2^2 \right\} \\ = & \tilde{\theta} - \frac{\sum_{k=1}^K \left\| \hat{g}_k^t - \hat{g}_k^{t-1} \right\|_2^2}{2} \end{aligned} \quad (14)$$

4) 拉格朗日乘子更新

$$\hat{V}^{i+1} = \hat{V}^i + \tau(\hat{G}^{i+1} - \hat{H}^{i+1}) \quad (15)$$

其中,  $i$  和  $i+1$  表示迭代次数;  $\tau$  为更新参数;  $\hat{G}^{i+1}$  和  $\hat{H}^{i+1}$  表示在第  $(i+1)$  次迭代中, 式(6)子问题  $\hat{G}$  和式(12)子问题  $\hat{H}$  的解。

$$\tau^{i+1} = \min(\tau_{\max}, \beta\tau^i) \quad (16)$$

其中,  $\beta=10, \tau^{\max}=10000$ 。

3.4 目标定位

$$R^t = F^{-1} \sum_{k=1}^K (\hat{Z}_k \cdot \hat{g}_k^{-1}) \quad (17)$$

其中,  $R^t$  为  $t$  帧响应图,  $Z_k$  为  $t$  帧提取的特征映射的傅里叶形式,  $g_k^{-1}$  表示第  $t-1$  帧中训练好的相关滤波器的傅里叶形式。响应值数值最大处即跟踪目标的中心位置, 从而实现目标的定位。

4 特征提取

特征提取过程中, 高维数据将占用大量的存储空间, 因此, 本文利用特征自身相关性<sup>[20]</sup>, 考虑结合卷积神经网络, 以在获取更多图像信息的同时提升计算速率。

4.1 特征融合

因为梯度主要存在于图像边缘, 作用于局部, 对光照变化具有良好的鲁棒性<sup>[21]</sup>, 所以使用 HOG 特征描述局部目标的表象和形状, 但由于 HOG 特征维数较高, 导致迭代过程中增加了计算复杂度, 所以本文采用维数更低的 fHOG 特征。CN 特征作为全局特征, 对图像的像素进行遍历, 所以对图像区域的形变以及运动变化不敏感。在目标发生形变时, 单个特征读取信息会发生变化, 导致目标丢失。考虑到 CN 特征与 fHOG 特征各自的优势, 采用 cat 级联函数从而实现在非卷积特征层面进行融合的目的。

不同层的卷积特征在视觉跟踪中的表现不同<sup>[22-23]</sup>, Conv1 层卷积神经网络可以学习目标的空间信息, 便于较准确地预测目标位置, 但面对复杂背景的情况, 其表现出非适应性。Conv5 层卷积神经网络可以学习丰富的语义信息<sup>[24-25]</sup>, 便于区分背景和目标, 噪声的干扰较少。为避免过多卷积层增加计算复杂度, 从而降低速率, 本文采用卷积神经网络 ImageNet-VGG-2048 的 Conv1 和 Conv5 层实现卷积特征层面的融合, 在保留空间轮廓和语义信息的前提下, 增强目标的表现能力。

表 1 不同特征的消融实验

Table 1 Ablation experiments with different characteristics

HOG	fHOG	CN	VGG-2048-Conv1+Conv5	Average(accuracy and success rate)
✓				73.70
	✓			74.70
✓		✓		74.50
✓			✓	75.40
	✓	✓		76.00
		✓	✓	81.11
	✓		✓	81.18
	✓	✓	✓	82.00

为了验证多特征融合方法的有效性, 本文进行了消融实验, 结果如表 1 所列, 并在 OTB-2013 数据集集中的 50 个视频

序列上进行对比实验, 结合图 1 可以看出, 本文算法的 fHOG 特征在成功率相差不大的情况下, 精确率较 HOG 特征提高了 9%, 级联后的 HOG\_CN 算法相比于 HOG 算法在精确率和成功率上均有较高鲁棒性, 说明结合 CN 和 HOG 特征能更好地提取目标特征。如图 1 所示, 本文融合算法的精确率和成功率均最优, 说明了采用多特征融合方法对跟踪算法性能提升的可行性。

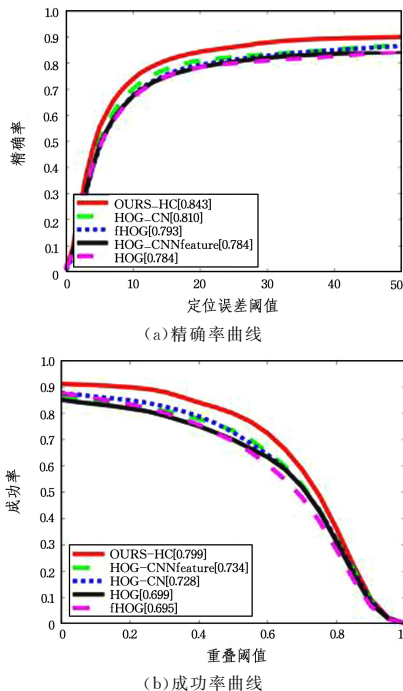


图 1 不同特征融合的结果

Fig. 1 Results of different feature fusion

4.2 激活函数

在神经网络中, 网络深度即卷积层数增多容易造成梯度消失的现象, 当梯度小于 1 时, 网络权值更新不稳定, 在卷积的不断作用下, 预测值所在网络层与真实值所在网络层相比, 梯度变化更小, 权值变化更加缓慢, 导致模型收敛速度逐渐下降, 最终为零。为了缓解上述现象, 本文将修正线性单元(ReLU)引入到神经网络中, ReLU 函数的表达式为:  $ReLU(x) = \begin{cases} x \rightarrow, & \text{if } x > 0 \\ 0 \rightarrow, & \text{if } x \leq 0 \end{cases}$ , 函数图像如图 2 所示。

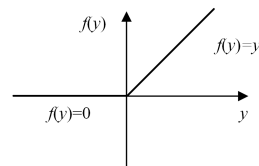


图 2 ReLU 函数

Fig. 2 ReLU function

从图 2 可以看出, ReLU 函数作为分段函数, 可以抑制输入样本为负数的情况, 而保留输入样本为正数的情况, 在对特征进行提取的过程中, 只需关注与目标相关的特征信息, 而这种单方面的抑制作用可以确保卷积神经网络(CNN)模型的收敛速度不随层数增加而变化, 激活后的模型可更好地挖掘目标相关特征, 拟合训练数据。

5 算法步骤

1) 裁剪得到初始帧目标区域, 使用基线算法中的空间

正则项权重向量对本文算法的权重进行初始化。使用轻量化的卷积神经网络进行特征提取,结合降维处理后的非卷积层,训练相关滤波器,得到初始帧的目标模型和响应图模型。

2)将上一帧经过训练的相关滤波器,作用于当前帧的目标区域,得到训练后的响应图,目标中心位置即响应峰值。

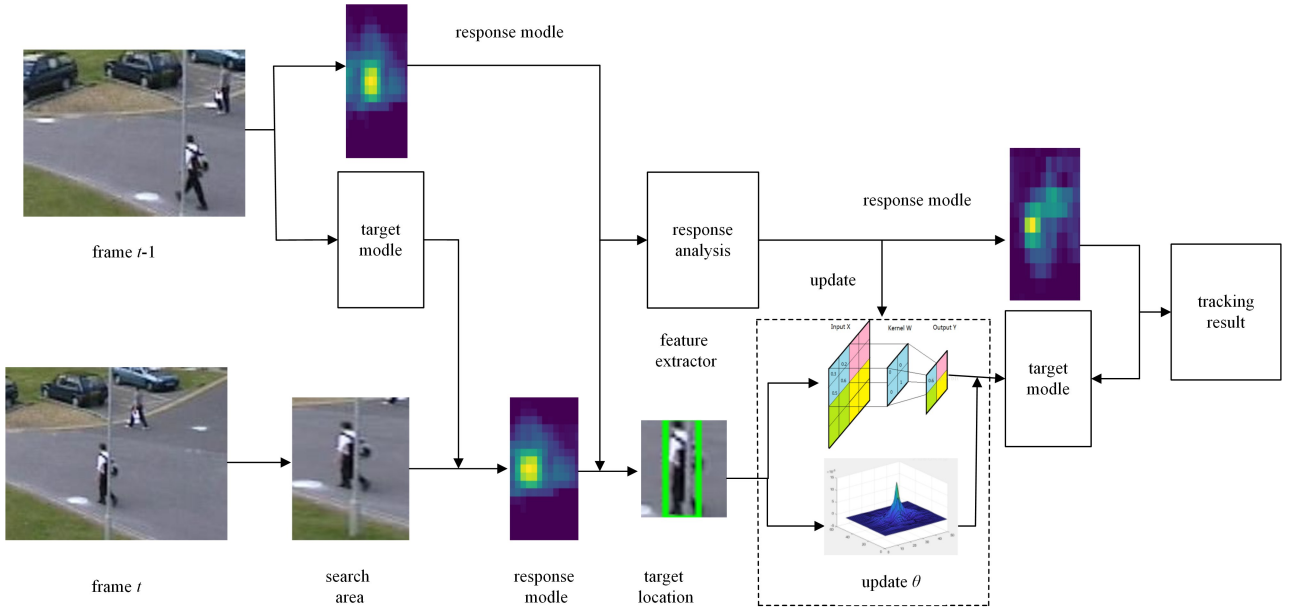


图3 本文算法总体框架示意图

Fig. 3 Overall schematic diagram of the proposed algorithm

## 6 实验验证与分析

### 6.1 实验环境及参数配置

本文所选用的编程软件为 MATLAB R2018b,实验操作系统为 mac OS Mojave 10.14.6,处理器为 2.9 GHz Intel Core i7 CPU,内存为 8 GB。

### 6.2 多特征感知的时空自适应相关滤波目标跟踪算法的对比实验

为了客观并且有针对性验证本文算法的有效性,我们将其与 8 种同类型算法进行了对比实验,分别是:本文的基线算法 STRCF(2018);利用循环矩阵进行采样的 KCF 算法(2014);引入联系上下文和尺度滤波器 MOSSE\_CA(2015);提高样本数量但减小搜索区域得到 DCF\_CA(2017);引入时间正则项的 Staple(2016);对采样区域进行约束,减少异常学习的 ARCF(2019);采用分层时空图正则化的相关滤波器 HSTDCF\_CA(2021);通过将目标分成 4 个块,利用核相关滤波器找到每个块的最大响应位置来计算尺度因子的 SAMF\_CA(2020)。本文算法采用一次性通过评价(one-time evaluation,OPE)标准来分析各算法的跟踪精度、成功率和跟踪速度等各项性能指标。

实验中,采用正方形作为搜索框,使用 ADMM 算法进行迭代时的次数为 4;在对自适应空间正则项权重初始化时,参数  $\delta=10$ ;在模型更新时,目标模型的学习参数  $\alpha$  为 0.02,响应图模型的阈值  $\varphi=3000$ 。其余参数与本文基线算法 STRCF 的参数设置相同。

### 6.3 定性分析

图 4 为数据集 OTB100 中快速运动(FM)、复杂背景

3)利用相邻响应图之间信息的相似性,采用式(3)自适应地更新空间正则化项,并进行滤波器训练。

4)根据每一帧,采用式(4)判定目标的状态,选择是否学习,采用式(6)更新目标模型和响应图模型。

5)输出跟踪结果,执行步骤 2),使用滤波器对下一帧的图像进行训练并实现模型的更新。

(BC)、非刚性形变(DEF)、光照变化(IV)、超出视野(OV)、运动模糊(MB)、平面内/外旋转(in-plane/out-of-plane rotation)等 9 种不同属性的视频序列,从图 4 中可以看出,与其他对比算法相比,本文算法的跟踪性能较优。

1)遮挡和形变。从图 4(a)可以看到,在视频序列 Girl2\_1 中,在第 105 帧之前因为未出现任何遮挡所以各算法跟踪性能保持良好。但在第 107 帧到第 298 帧开始出现遮挡并逐渐消失的过程中,SAME\_CA, MOSSE\_CA, Staple, DCF\_CA, KCF 和 ARCF 因为没有时空正则化项,导致背景与前景区分不明显,出现不同程度的漂移。到第 358 帧时,HSTDCF\_CA 和 STRCF 虽具有空间正则化项,但也出现了漂移现象。因为本文算法具有自适应时空正则化项,深度提取的背景特征的权重会随目标变化而降低,所以到第 717 帧只有本文算法仍然可以进行有效跟踪。

2)背景杂乱。从图 4(b)可以看到,在视频序列 Box\_1 中,到第 365 帧时, MOSSE\_CA, SAME\_CA 和 KCF 采用核滤波器,没有区分背景和目标的判别器,导致跟踪失败。在第 458 帧到第 933 帧由于目标和背景信息相似度高且光照持续,导致 Staple, DCF\_CA, STRCF, HSTDCF\_CA 和 ARCF 算法陆续跟踪失败。在第 933 帧到第 1323 帧时,只有本文算法可以继续跟踪,通过空间局部响应和全局响应的变化确定局部响应的可信度,学习过程中,限制了学习可信度低的区域,从而有效地区分目标和背景信息,有较高的精确率。

3)快速运动。从图 4(c)序列 Dragon-baby 中可以看到,从第 10 帧开始,使用单一的灰度特征的特征的 KCF 和 Staple 算法

就出现了跟踪漂移现象。在 38 到 82 帧时,目标快速移动, MOSSE\_CA, DCF\_CA, SAME\_CA, ARCF 和 STRCF 虽使用了 CN 特征,但快速移动导致模型尺度产生变化,学习次数累加导致偏差增大,最终不能及时跟踪到目标。从初始帧至最后一帧,目标发生多次旋转和快速运动,只有本文算法和 HSTDCF\_CA 算法从始至终完全跟踪到目标,且本文跟踪结果有更好的鲁棒性。本文使用多特征融合的方式能够更好地使用图像信息,并利用相邻两个峰值响应的相似性,判断学习区域,从而实现自适应更新目标模型的目的,使得该算法在快速运动的场景下具备较好的鲁棒性。

4) 运动模糊。从图 4(d)可以看出,在视频序列 Board\_1 中,在初始帧到第 363 帧时,各算法均可跟踪到目标,但 KCF 和 Staple 算法由于没有处理边界效应,不能较好地适应目标外观的变化,所以当摄像机与目标存在相对运动时,图像开始模糊;HSTDCF\_CA, DCF\_CA, MOSSE\_CA 算法未加入深度特征提取;在 363 帧到第 561 帧时, ARCF, SAMF\_CA, STRCF 算法无法完整提取并利用图像位置信息,导致不能准

确包含目标信息;在 561 帧到第 581 帧时,因为本文算法在提取样本信息时能够通过深度学习提取到更准确的样本特征,并通过 ReLU 函数减少负采样,即只需关注与目标相关的特征信息,在减少计算复杂度的同时提高了准确性,避免过拟合从而增加了计算速率,所以能以较高速率和准确率跟踪到目标。

5) 非刚性形变、平面内/外旋转。从图 4(e)可以看出,视频序列 Rider 中,在第 60 帧之前,所有算法均可准确跟踪到目标。从第 67 帧到第 89 帧,目标发生了 1 次平面外旋转,1 次平面内旋转。在这期间, KCF, Staple, DCF\_CA, MOSSE\_CA 因为采用手动的方式裁剪图像,且只是单纯地使用常规的特征提取方式,导致其对图像深度特征提取不完善,不能较好地更新目标的外观模型。由于 ARCF, HSTDCF\_CA, SAMF\_CA 和 STRCF 没有自适应的时空正则项,对于背景和前景的区分不够明显,对于权重的更新没有明确定义,所以在第 142 帧时,只有本文算法在消除运动影响的情况下,进行了深度的特征提取,准确跟踪到目标。



图 4 9 种算法在图像序列上的跟踪情况对比

Fig. 4 Comparison of tracking performance of nine algorithms on image sequences

#### 6.4 定量比较

本文算法在数据集 OTB50 中精确率为 0.816, 成功率为 0.744, 高于 STRCF 算法的精确率 0.811 和成功率 0.733; 在数据集 DTB70 中的精确率为 0.747, 成功率为 0.789, 均高于 STRCF 算法的精确率(0.737)和成功率(0.760)。

从图 5 中可以看出,在数据集 OTB50 和 DTB70 中,相较于基准框架 STRCF,本文算法的成功率和精确率均较高。为了更清晰全面地体现本文算法在不同数据集上优越的跟踪性能,表 2、表 3 分别列出了本文算法与其他主流算法在数据集 OTB100 上各种属性的跟踪成功率与跟踪精确率数据。

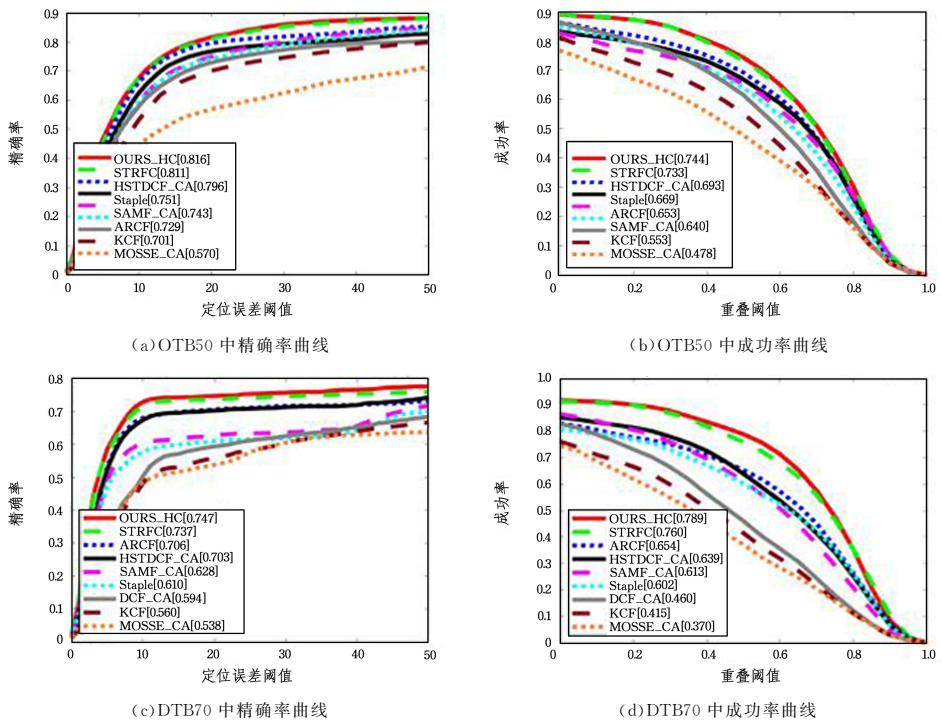


图 5 算法在不同数据集上的准确率和成功率曲线对比图

Fig. 5 Comparison curve of accuracy and success rate of various algorithms on dataset

表 2 9 种算法在数据集 OTB100 上跟踪属性的精确率

Table 2 Accuracy of nine algorithms in tracking attributes on dataset OTB100

Properties	Ours	STRCF	SAMF_CA	ARCF	Staple	HSTDCE_CA	DCF_CA	KCF	MOSSE_CA
BC	0.790	0.766	0.689	0.671	0.668	0.666	0.578	0.501	0.349
DEF	0.831	0.810	0.745	0.688	0.724	0.735	0.669	0.632	0.515
IPR	0.833	0.826	0.741	0.734	0.699	0.749	0.711	0.601	0.547
MB	0.747	0.737	0.703	0.706	0.610	0.628	0.594	0.560	0.538
OCC	0.847	0.840	0.755	0.732	0.724	0.759	0.694	0.635	0.504
OV	0.816	0.811	0.743	0.729	0.768	0.796	0.751	0.701	0.570
OPR	0.850	0.850	0.771	0.727	0.738	0.764	0.692	0.677	0.525
FM	0.808	0.802	0.723	0.752	0.710	0.757	0.727	0.435	0.546
LR	0.747	0.737	0.703	0.706	0.610	0.628	0.594	0.560	0.538

表 3 9 种算法在数据集 OTB100 上跟踪属性的成功率

Table 3 Success rate of nine algorithms in tracking attributes on dataset OTB100

Properties	Ours	STRCF	SAMF_CA	ARCF	Staple	HSTDCE_CA	DCF_CA	KCF	MOSSE_CA
BC	0.716	0.693	0.641	0.587	0.576	0.548	0.517	0.457	0.311
DEF	0.791	0.750	0.664	0.658	0.653	0.680	0.529	0.512	0.429
IPR	0.823	0.794	0.703	0.718	0.650	0.708	0.634	0.550	0.530
MB	0.663	0.652	0.534	0.633	0.472	0.494	0.317	0.295	0.319
OCC	0.789	0.760	0.613	0.654	0.602	0.639	0.460	0.415	0.370
OV	0.744	0.733	0.640	0.653	0.669	0.693	0.616	0.553	0.478
OPR	0.790	0.768	0.687	0.657	0.649	0.672	0.556	0.527	0.425
FM	0.774	0.757	0.636	0.716	0.645	0.701	0.610	0.526	0.506
LR	0.663	0.652	0.534	0.633	0.472	0.494	0.317	0.295	0.319

结合表 2 和表 3 可以看出本文算法每种属性的精确率和成功率得分均高于 STRCF 算法。本文算法由于具有空间自适应性,能够分辨背景进而控制权重,所以在处理背景杂乱等问题时具有较高精确率;通过空间局部响应和全局响应的变化确定局部响应的可信度,从而控制模型的更新速率,继而在处理形变和快速运动的问题上保持较高成功率。

由于本文算法选择多层卷积特征融合的方式,使得算法在处理光照变化、遮挡时有良好的效果。其在处理超视距属性时,作用于多层卷积的浅层表象特征以及卷积神经网络的

深层语义特征,共同作用使得本文算法的整体跟踪效果与 8 种对比算法相比,在确保实时跟踪的前提下,达到了较好的跟踪效果。

### 6.5 跟踪速度分析

本文算法随机选取数据集 OTB100 中 20 组序列进行跟踪比较,计算方法为总帧数除以总时间,结果如表 4 所列,结合表 2、表 3 可以看出,在跟踪实验序列的过程中,本文算法的跟踪成功率与跟踪准确率较高,且满足实时性的需求,跟踪性能良好。

表 4 9种算法在图像序列上的平均跟踪速度

Table 4 Average tracking speed of nine algorithms on image sequences

(单位: f/s)								
Ours	STRCF	SAMF_CA	Staple	HSTDCF_CA	KCF	ARCF	DCF_CA	MOSSE_CA
16.9	22.6	16.2	16.8	13.8	131.5	12.5	14.9	178.9

在目标跟踪的迭代过程中,时间复杂度尤为重要,所以本文算法在利用深度特征的优势下,采用轻量形的卷积神经网络。在 ReLU 激活函数的作用下,减少过拟合现象,从而进一步减少时间复杂度。确保本文算法在更好地利用图像样本信息的基础上,具备较高的实时性。

**结束语** 本文提出了多特征感知的时空自适应相关滤波目标跟踪算法。通过初始化空间正则化项,引入欧几里德范数,结合样本信息,训练相关滤波器;利用响应图之间的相似性判断样本可信度,进而限制局部相关滤波器的学习,动态地更新学习率;在非卷积层采用级联降维处理的 fHOG 特征和 CN 特征,在卷积层选择采用 ImageNet-VGG-2048 的 Conv1 和 Conv5 层卷积特征进行跟踪并使用 ReLU 激活函数来处理提取的 CNN 特征,确保图像信息的准确性和高效性。本文算法的优势在于:1)在 STRCF 算法的框架下引入自适应空间正则项,局部响应变化显示当前帧搜索区域中每个像素的可信度。在保证跟踪速度的同时,判断目标与背景,从而分配不同的权重,实现空间自适应性。2)利用前后两帧响应图的相似性,判断局部可信度,在得到更强大的外观模型的同时控制模型的更新速率,实现时间自适应性。3)考虑到 HOG 特征维度过高导致速率降低,选择 fHOG 特征。结合 ReLU 函数来处理 ImageNet-VGG-2048 的 Conv1 和 Conv5 层卷积提取的特征,既提高了速率又保留了深度图像的特征。通过在数据集 OTB50,DTB70 和 OTB100 上与主流的 8 种目标跟踪算法进行大量对比实验,可以得出,本文算法能够有效处理背景杂乱、快速运动、遮挡等一些复杂问题;且相比其他主流算法,均保持着高度鲁棒性。在本文的基础上,下一步将围绕如何在更完善地利用图像信息的同时提高跟踪速度而展开深入研究。

## 参考文献

- [1] PU L, WEI Z H, HOU Z Q, et al. Twin network visual tracking algorithm based on asymmetric convolution[J]. Journal of Electronics and Information, 2022, 44(8): 2957-2965.
- [2] JAVED S, MAHMOOD A, DIAS J, et al. Hierarchical Spatio-temporal Graph Regularized Discriminative Correlation Filter for Visual Object Tracking[J]. IEEE Transactions on Cybernetics, 2022, 52(11): 12259-12274.
- [3] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]// Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA; IEEE, 2010: 2544-2550.
- [4] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C]// Proceedings of the 12th European Conference on Computer Vision. Florence, Italy; Springer, 2012: 702-715.
- [5] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(3): 583-596.
- [6] GALOOGAHI H K, SIM T, LUCEY S. Correlation filters with limited boundaries[C]// Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA; IEEE, 2014: 4630-4638.
- [7] DANELLJAN M, HÄGER G, KHAN F S, et al. Accurate scale estimation for robust visual tracking[C]// Proceedings of 2014 British Machine Vision Conference. Nottingham, UK; BMVA Press, 2014: 1-11.
- [8] DANELLJAN M, HÄGER G, KHAN F S, et al. Adaptive decontamination of the training set: a unified formulation for discrimination visual tracking[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE, 2016: 1430-1438.
- [9] MUELLER M, SMITH N, GHANEM B. Context-aware correlation filter tracking[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu; IEEE, 2017: 1387-1395.
- [10] LI Y, FU C, F DING, et al. AutoTrack: Towards High-Performance Visual Tracking for UAV with Automatic Spatio-Temporal Regularization[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2020.
- [11] QING L B, WU M F, LIU G, et al. Image restoration algorithm based on wavelet domain ADMM depth network[J]. Engineering Science and Technology, 2022, 54(5): 257-267.
- [12] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: complementary learners for real-time tracking[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE, 2016: 1401-1409.
- [13] MA C, HUANG J B, YANG X K, et al. Hierarchical convolutional features for visual tracking[C]// Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE, 2015: 3074-3082.
- [14] DANELLJAN M, ROBINSON A, KHAN F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[C]// Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands; Springer, 2016: 472-488.
- [15] LI F, TIAN C, ZUO W M, et al. Learning spatial-temporal regularized correlation filters for visual tracking[C]// CVPR. 2018: 4904-4913.
- [16] HOU Z Q, GUO H, MA S G, et al. An anchor free target detection algorithm based on dual branch feature fusion[J]. Journal of Electronics and Information, 2022, 44(6): 2175-2183.
- [17] LIU Y, DU Y M, GUAN Y, et al. An improved algorithm of acoustic signal weighting based on frequency domain[J]. Journal of Chengdu University of Information Engineering, 2021, 36(1): 68-72.
- [18] JIANG W T, LIU X X, TU C, et al. Target tracking of adaptive spatial anomaly[J]. Journal of Electronics and Information, 2022, 44(2): 523-533.

[19] XU J P,WANG F. Image classification method based on improved S-ReLU activation function[J]. Science, Technology and Engineering, 2022, 22(29): 963-968.

[20] JIANG W T, MENG Q J. Adaptive spatiotemporal regularization based correlation filtering for target tracking[J]. Journal of Intelligent Systems, 2023, 18(4): 754-763.

[21] XU T, FENG Z H, WU X J, et al. Learning Adaptive Discriminative Correlation Filters via Temporal Consistency Preserving Spatial Feature Selection for Robust Visual Object Tracking[J]. IEEE Transactions on Image Processing, 2019, 28(11): 5596-5609.

[22] YE X Y, GUO W F, ZENG M S, et al. Image steganography detection based on multi-layer perceptual convolution and channel weighting[J]. Journal of Electronics and Information, 2022, 44(8): 2949-2956.

[23] SHI C Y, SUN Q, LU J P, et al. Image edge detection based on deep fusion convolution neural network[J]. Modern Electronic Technology, 2022, 45(24): 141-144.

[24] WIDYANINGRUM E, BAI Q, FAJARI M K, et al. Air-borne

laser scanning point cloud classification using the DGCNN deep learning method[J]. Remote Sensing, 2021, 13(5): 859.

[25] ZHANG Y L, QIAN X Y, GE H J, et al. Adaptive multi feature fusion correlation filtering target tracking[J]. Chinese Journal of Image Graphics, 2020, 25(6): 1160-1170.



**MENG Qingjiao**, born in 1998, master candidate, is a member of China Computer Federation. Her main research interests include image and visual computing, pattern recognition and artificial intelligence.



**JIANG Wentao**, born in 1986, Ph. D, master supervisor, is a member of China Computer Federation. His main research interests include image and visual computing, pattern recognition and artificial intelligence.