

文章编号: 1671-7449(2024)01-0001-11

深度特征目标感知交替方向乘子法优化多指标更新相关滤波跟踪算法

王国刚, 杨雨前, 李泽欣

(山西大学 物理电子工程学院, 山西 太原 030006)

摘要: 空间正则化相关滤波算法跟踪过程中仅采用手工特征表征目标, 高斯-赛德尔方法训练滤波器的复杂度高, 跟踪结果不可靠时仍逐帧更新模型, 导致跟踪效果不佳。针对空间正则化相关滤波算法存在的问题, 提出深度特征目标感知交替方向乘子法(Alternating Direction Method of Multipliers, ADMM)优化多指标更新相关滤波跟踪算法。该算法融入预训练网络提取的深度特征, 并依回归损失的梯度信息进行通道选择, 增强了对目标的表征能力; 采用交替方向乘子法训练相关滤波器, 降低算法复杂度, 提升跟踪速度; 根据多指标更新方法判断是否进行模型更新, 不但提升了算法运行效率, 而且避免了因学习到错误信息而导致的模型腐败。实验结果表明, 所提算法的成功率、精确度在数据集OTB2015上均优于其它8种对比算法, 且在复杂场景下具有更强的跟踪鲁棒性。

关键词: 相关滤波; 深度特征目标感知; 交替方向乘子法; 多指标更新方法

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.3969/j.issn.1671-7449.2024.01.001

引用格式: 王国刚, 杨雨前, 李泽欣. 深度特征目标感知交替方向乘子法优化多指标更新相关滤波跟踪算法[J]. 测试技术学报, 2024, 38(1): 1-11.

WANG Guogang, YANG Yuqian, LI Zexin. Correlation filtering algorithm based on target-aware deep tracking by admm optimization and multi-index update[J]. Journal of Test and Measurement Technology, 2024, 38(1): 1-11.

Correlation Filtering Algorithm Based on Target-Aware Deep Tracking by Alternating Direction Method of Multipliers Optimization and Multi-Index Update

WANG Guogang, YANG Yuqian, LI Zexin

(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: A novel algorithm named as correlation filtering algorithm based on target-aware deep tracking by alternating direction method of multipliers(ADMM) optimization and multi-index update is proposed to avoid the disadvantages of only using hand-crafted features to characterize the target, the higher complexity of Gauss-Seidel method, which is used to train the filter, and updating the model frame by frame even when the tracking results are unreliable in SRDCF algorithm. After fusing the deep features that are extracted by the pre-training network, the channel selection is performed according to the gradient information of regression loss, which enhances the characterization ability of the target. In order to reduce the computational complexity and accelerate the tracking speed, the proposed algorithm utilizes the alternating direction method of multipliers to train the

收稿日期: 2023-01-05

基金项目: 国家自然科学基金资助项目(11804209); 山西省自然科学基金资助项目(201901D111031, 201901D211173); 山西省高校科技创新计划资助项目(2019L0064, 2020L0051)

作者简介: 王国刚(1977-), 男, 副教授, 主要从事图像处理与视频分析研究。E-mail: kingguogang@sxu.edu.cn.

correlation filter. Furthermore, whether the model is updated or not depends on the multi-index update method, which not only improves the tracking efficiency, but also avoids the model corruption that is caused by learning wrong information. The experimental results show that the success rate and the precision of the presented algorithm outperform the compared algorithms on the OTB2015 dataset. Under the complex scenarios, the proposed approach has stronger tracking robustness compared with the state of the art approaches.

Key words: correlation filter; target-aware of deep features; alternating direction method of multipliers; multi-index update method

0 引言

在给定跟踪目标初始信息的前提下, 目标跟踪的任务是持续估计后续帧目标的位置、尺度等状态信息。作为计算机视觉领域的研究热点之一, 目标跟踪在军事侦察、智能机器人、手势识别等诸多领域具有重要的研究价值^[1-2]。

基于回归判别式模型, 传统相关滤波算法结合循环矩阵和快速傅里叶变换, 实现了对目标的跟踪^[3-4], 该类算法主要包括搜索区域、特征提取、目标定位和模型更新4个步骤。由于采用了数据量较小的手工特征, 传统相关滤波算法具有较快的跟踪速度。但是, 手工特征对旋转、形变等外观变化的鲁棒性较差, 因此该类算法在跟踪中难以取得较为鲁棒的跟踪效果。

得益于深度特征良好的表征能力和抗干扰能力, 结合深度学习的相关滤波算法^[5-7]相较于传统相关滤波算法具有较好的跟踪准确性和鲁棒性。该类算法主要分为两类: 一类是将预训练深度特征融入相关滤波器的HCF^[8], HDT^[9]等算法; 另一类是将相关滤波器集成到深度网络中的CFNet^[10], DCFNet^[11]和UDT^[12]等算法。这些结合深度学习的相关滤波算法虽然实现了精确和鲁棒的跟踪, 但是仍未解决因循环移位带来的边界效应, 导致跟踪算法的性能提升受到限制。

空间正则化相关滤波算法(Spatially Regularized Discriminative CF, SRDCF)^[13]按照滤波器系数的空间位置引入正则化项, 有效解决了边界效应。但是, 该算法跟踪中仅采用手工特征表征目标, Gauss-Seidel方法训练相关滤波器的时间复杂度高, 跟踪结果不可靠时仍逐帧更新模型。

针对SRDCF算法存在的问题, 提出深度特征目标感知ADMM优化多指标更新相关滤波跟踪算法(Target-Aware Deep Tracking by ADMM Optimization and Multi-Index Update, TACF)。该算法

融入预训练的深度特征, 并依回归损失的梯度信息进行通道选择, 提高对目标的表征能力; 采用交替方向乘子法(ADMM)^[14]训练相关滤波器, 降低算法复杂度, 加快跟踪速度; 根据多指标更新方法判断是否进行模型更新, 不但提升了算法运行效率, 还避免了因学习到大量干扰信息而导致的模型损坏。实验结果表明, TACF算法相较于其它8种现有算法实现了更精确、更鲁棒的跟踪。

1 TACF算法跟踪框架

针对SRDCF算法存在的缺陷, 提出TACF跟踪算法。该算法框图主要包含深度特征目标感知、TACF模型优化和多指标更新三部分, 如图1所示。

1.1 训练位置滤波器

相关滤波算法利用循环矩阵进行训练和检测, 尽管提高了计算效率, 但也带来了边界效应。为抑制边界效应, 引入空间正则化项, 构造如下目标函数, 训练TACF跟踪模型的位置滤波器。

$$E(f) = \min_f \frac{1}{2} \left\| \sum_{l=1}^L x^l \otimes f^l - y \right\|^2 + \frac{1}{2} \sum_{l=1}^L \|w \cdot f^l\|^2, \quad (1)$$

式中: 大小为 $M \times N$ 的 x^l 和 f^l 分别为第 l 通道的目标特征和位置滤波器; L 为总通道数; \otimes 和 \cdot 分别为循环卷积操作和哈达玛积; y 为预定义的标签; 空间正则化项 $\frac{1}{2} \sum_{l=1}^L \|w \cdot f^l\|^2$, 抑制背景干扰, 凸显感兴趣的区域。

1.2 位置检测与模型更新

位置检测时, 从当前帧提取特征图 z^l , 与前一帧训练的位置滤波器 f^l 在频域的各通道上进行哈达玛求积运算, 再把各通道的哈达玛积的和经傅里叶逆变换得到跟踪响应图 S

$$S = F^{-1} \left(\sum_{l=1}^L \hat{z}^l \cdot \hat{f}^l \right), \quad (2)$$

式中： F^{-1} 表示傅里叶逆变换； $\hat{\cdot}$ 表示傅里叶变换。跟踪响应图 S 中最大响应值点即为预测的当前帧目标的位置。

为适应跟踪目标外观的不断变化，需要对跟踪模型进行如下更新

$$f^{\text{new}} = (1 - \eta) f^{\text{old}} + \eta f, \quad (3)$$

式中： η 为更新时的学习率。

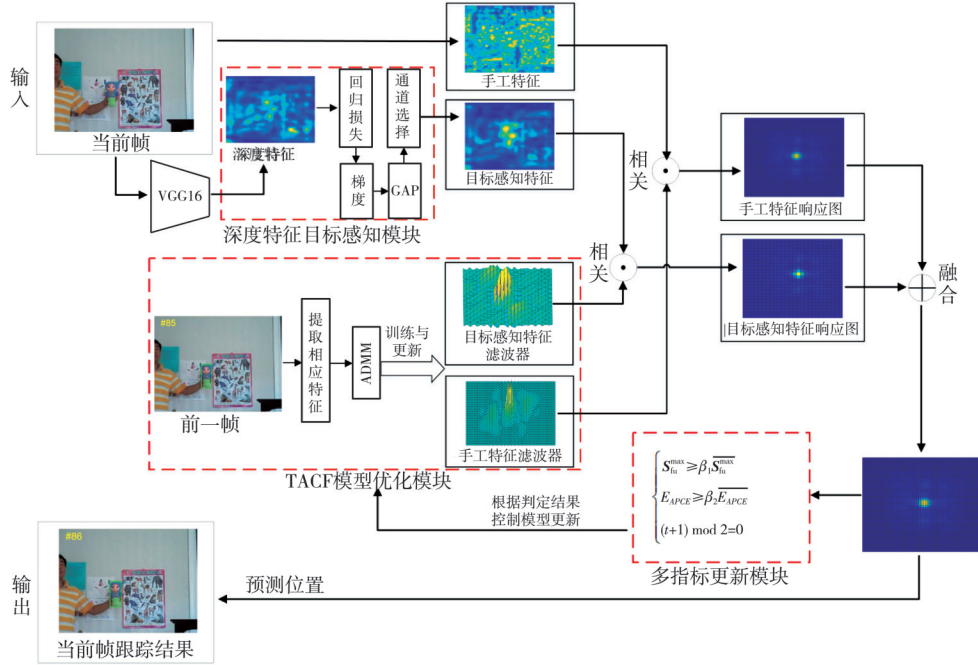


图 1 TACF 算法整体框架图

Fig. 1 The overall framework diagram of TACF algorithm

2 深度特征目标感知 ADMM 优化多指标更新相关滤波跟踪算法

2.1 深度特征目标感知

现有深度跟踪器用预训练的卷积神经网络 (Convolutional Neural Networks, CNN) 特征表征跟踪目标，而预训练的 CNN 特征先前一般是在

目标识别任务中用来描述目标外观特性的。与目标识别任务中的目标不同，跟踪目标可以是任意形式的任意对象类，所以，将深度特征直接用于跟踪不仅会带来大量的干扰信息，而且难以区分目标与背景，致使跟踪效果不佳。

为解决此问题，更好地提升对目标的表征能力，提出深度特征目标感知方法，对预训练的 CNN 特征图进行通道选择，具体流程如图 2 所示。

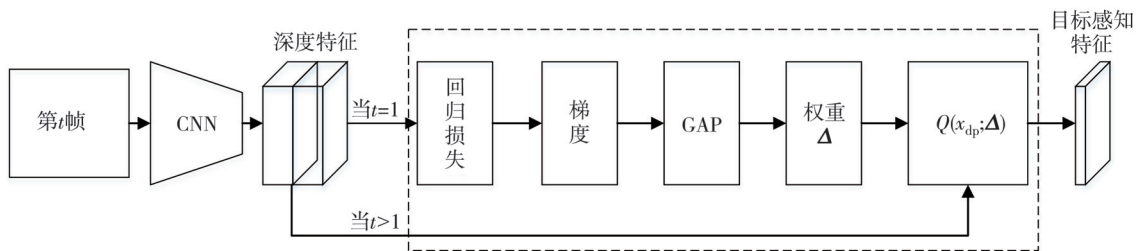


图 2 深度特征目标感知

Fig. 2 Target-aware of deep features

构建如下回归损失函数，得到适用于跟踪任务的目标感知特征。

$$R_{\text{reg}} = \|b(u, v) - h * x_{\text{dp}}\|^2 + \lambda \|h\|^2, \quad (4)$$

式中： h ， x_{dp} 分别为回归权重矩阵和预训练的深度特征图；标签 $b(u, v) = \exp\left(-\frac{u^2 + v^2}{2\sigma^2}\right)$ ， (u, v) 为

相对跟踪目标的偏移量； σ 为高斯内核的宽度。

使用式(5)计算回归损失梯度，以衡量各特征通道的重要性。

$$\frac{\partial R_{\text{reg}}}{\partial \mathbf{x}_{\text{dp}}} = \sum_{u,v} \frac{\partial R_{\text{reg}}}{\partial x_o(u,v)} \times \frac{\partial x_o(u,v)}{\partial \mathbf{x}_{\text{dp}}(u,v)} = \sum_{u,v} 2(\mathbf{b}(u,v) - x_o(u,v)) \times \mathbf{h}, \quad (5)$$

式中： $x_o(u,v)$ 为回归损失收敛后的权重矩阵 \mathbf{h} 与预训练的深度特征图 \mathbf{x}_{dp} 的卷积。

对回归损失梯度作全局平均池化（GAP）处理，可得特征通道重要性权重 $\mathbf{\Delta}$ ，如式(6)所示。

$$\mathbf{\Delta} = \text{GAP} \left(\frac{\partial R_{\text{reg}}}{\partial \mathbf{x}_{\text{dp}}^1}, \frac{\partial R_{\text{reg}}}{\partial \mathbf{x}_{\text{dp}}^2}, \dots, \frac{\partial R_{\text{reg}}}{\partial \mathbf{x}_{\text{dp}}^l}, \dots, \frac{\partial R_{\text{reg}}}{\partial \mathbf{x}_{\text{dp}}^L} \right). \quad (6)$$

重要性权值越大，特征通道对目标的表征能力越强^[15]。因此，深度特征目标感知策略保留权值为正的特征通道，剔除权值为负的特征通道，得到如式(7)所示的适用于跟踪任务的目标感知特征 \mathbf{x}_{ta} 。

$$\mathbf{x}_{\text{ta}} = Q(\mathbf{x}_{\text{dp}}; \mathbf{\Delta}), \quad (7)$$

式中： $Q(\cdot)$ 为根据权重 $\mathbf{\Delta}$ 选择目标感知特征的函数。

绘制特征热力图，比较原始深度特征和目标感知特征的不同，以验证深度特征目标感知策略的有效性。图3为绘制的热力图，从左到右分别为视频帧、原始深度特征和目标感知特征。

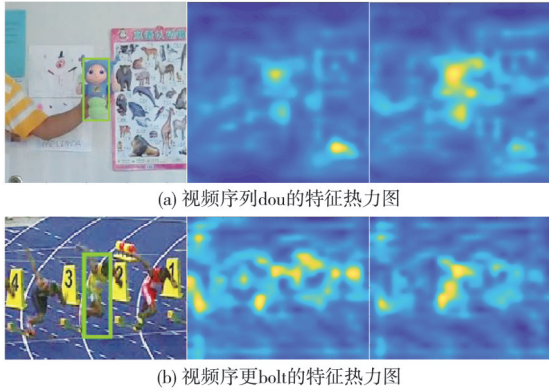


图3 原始深度特征和目标感知特征热力图

Fig. 3 Thermal maps of original deep features and target-aware features

如图3(a)所示，原始深度特征热力图焦点没有聚集在目标区域，未能有效区分目标与背景；而目标感知特征热力图焦点全部聚集在目标区域，剔除了右下角的干扰信息，能将目标与背景区分开来。图3(b)中，与跟踪目标相似的干扰物较多，原始深度特征热力图在干扰物上有大量焦点聚集，包含较多干扰信息；而在目标感知特

征热力图中，大量干扰信息被剔除，焦点主要聚集在目标区域，凸显了跟踪目标的外观特征。

2.2 TACF模型优化

式(1)为TACF跟踪模型的目标函数，为优化模型，引入约束条件 $\mathbf{f} = \mathbf{g}$ ，式(1)的增广拉格朗日方程可以表示为

$$\zeta(\mathbf{f}, \mathbf{g}, \mathbf{s}) = \frac{1}{2} \left\| \sum_{l=1}^L \mathbf{x}^l \otimes \mathbf{f}^l - \mathbf{y} \right\|^2 + \frac{1}{2} \sum_{l=1}^L \|\mathbf{w} \cdot \mathbf{g}^l\|^2 + \sum_{l=1}^L (\mathbf{s}^l)^\top (\mathbf{f}^l - \mathbf{g}^l) + \frac{\rho}{2} \sum_{l=1}^L \|\mathbf{f}^l - \mathbf{g}^l\|^2, \quad (8)$$

式中： \mathbf{s} 为拉格朗日乘子； ρ 为步长参数。

令 $\mathbf{d} = \frac{1}{\rho} \mathbf{s}$ ，式(8)可以简化为

$$\zeta(\mathbf{f}, \mathbf{g}, \mathbf{d}) = \frac{1}{2} \left\| \sum_{l=1}^L \mathbf{x}^l \otimes \mathbf{f}^l - \mathbf{y} \right\|^2 + \frac{1}{2} \sum_{l=1}^L \|\mathbf{w} \cdot \mathbf{g}^l\|^2 + \sum_{l=1}^L \|\mathbf{f}^l - \mathbf{g}^l + \mathbf{d}^l\|^2. \quad (9)$$

将TACF模型优化问题拆分为如式(10)所示的3个局部子问题，采用ADMM算法交替求解。

$$\begin{cases} \mathbf{f}^{(j+1)} = \arg \min_f \frac{1}{2} \left\| \sum_{l=1}^L \mathbf{x}^l \otimes \mathbf{f}^l - \mathbf{y} \right\|^2 + \frac{\rho}{2} \|\mathbf{f} - \mathbf{g} + \mathbf{d}\|^2, \\ \mathbf{g}^{(j+1)} = \arg \min_g \frac{1}{2} \sum_{l=1}^L \|\mathbf{w} \cdot \mathbf{g}^l\|^2 + \frac{\rho}{2} \|\mathbf{f} - \mathbf{g} + \mathbf{d}\|^2, \\ \mathbf{d}^{(j+1)} = \mathbf{d}^{(j)} + \mathbf{f}^{(j+1)} - \mathbf{g}^{(j+1)}. \end{cases} \quad (10)$$

1) 求解子问题 \mathbf{f}

$$\arg \min_f \frac{1}{2} \left\| \sum_{l=1}^L \mathbf{x}^l \otimes \mathbf{f}^l - \mathbf{y} \right\|^2 + \frac{\rho}{2} \|\mathbf{f} - \mathbf{g} + \mathbf{d}\|^2. \quad (11)$$

依帕塞瓦尔定理，式(11)在频域中可化为

$$\arg \min_f \frac{1}{2} \left\| \sum_{l=1}^L \hat{\mathbf{x}}^l \cdot \hat{\mathbf{f}}^l - \hat{\mathbf{y}} \right\|^2 + \frac{\rho}{2} \|\hat{\mathbf{f}} - \hat{\mathbf{g}} + \hat{\mathbf{d}}\|^2. \quad (12)$$

由式(12)可知，标签 $\hat{\mathbf{y}}$ 的第 i 个元素只取决于目标特征 $\hat{\mathbf{x}}$ 和相关滤波器 $\hat{\mathbf{f}}$ 在所有通道中的第 i 个元素。令 $\mathbf{K}_i(\cdot)$ 表示各个通道的第 i 个元素组成的列向量，所以可将子问题 \mathbf{f} 进一步分解为 MN 个

子问题, 每个子问题如式(13)所示。

$$\arg \min_{\kappa_i(\hat{f})} \frac{1}{2} \left\| \mathbf{K}_i(\hat{x})^\top \mathbf{K}_i(\hat{f}) - \hat{y}_i \right\|^2 + \frac{\rho}{2} \left\| \mathbf{K}_i(\hat{f}) - \mathbf{K}_i(\hat{g}) + \mathbf{K}_i(\hat{d}) \right\|^2. \quad (13)$$

令式(13)的导数为0, 可得

$$\mathbf{K}_i(\hat{f}) = (\mathbf{K}_i(\hat{x})\mathbf{K}_i(\hat{x})^\top + \rho\mathbf{I})^{-1} \cdot (\mathbf{K}_i(\hat{x})\hat{y}_i + \rho\mathbf{K}_i(\hat{g}) - \rho\mathbf{K}_i(\hat{d})). \quad (14)$$

由式(14)和 Sherman-Morrison 公式, 可得子问题 f 在频域中的封闭解

$$\mathbf{K}_i(\hat{f}) = \frac{1}{\rho} \left(\mathbf{I} - \frac{\mathbf{K}_i(\hat{x})\mathbf{K}_i(\hat{x})^\top}{\rho + \mathbf{K}_i(\hat{x})^\top \mathbf{K}_i(\hat{x})} \right) (\mathbf{K}_i(\hat{x})\hat{y}_i + \rho\mathbf{K}_i(\hat{g}) - \rho\mathbf{K}_i(\hat{d})). \quad (15)$$

子问题 f 的时域解可由频域解的傅里叶逆变换得到。

2) 求解子问题 g

$$\arg \min_g \frac{1}{2} \sum_{l=1}^L \left\| \mathbf{w} \cdot \mathbf{g}^l \right\|^2 + \frac{\rho}{2} \left\| \mathbf{f} - \mathbf{g} + \mathbf{d} \right\|^2. \quad (16)$$

令式(16)的导数为0, 可得子问题 g 的封闭解

$$\mathbf{g} = (\mathbf{W}^\top \mathbf{W} + \rho\mathbf{I})^{-1} (\rho\mathbf{f} + \rho\mathbf{d}), \quad (17)$$

式中: \mathbf{W} 为 1 个 $LMN \times LMN$ 的对角矩阵。

3) 更新步长参数

$$\rho^{(j+1)} = \min(\rho_{\max}, \gamma\rho^j), \quad (18)$$

式中: ρ_{\max} 为 ρ 的最大值; γ 为尺度参数。

2.3 多指标更新

因为深度特征包含丰富的语义信息, 所以, 基于目标感知特征的跟踪模型对目标的外观变化具有较强的鲁棒性, 但目标感知特征的空间分辨率较低, 目标定位准确性较差。另一方面, 手工特征空间分辨率较高, 有利于目标的精确定位, 但包含颜色、纹理等信息的跟踪器鲁棒性较差。因此, TACF 采用目标感知特征和手工特征训练相应的跟踪模型, 分别利用式(2)得到响应图 \mathbf{S}_{ta} 和 \mathbf{S}_{hc} , 再依式(19)获得融合后的响应图 \mathbf{S}_{fu} 。

$$\mathbf{S}_{\text{fu}} = \theta_1 \times \mathbf{S}_{\text{ta}} + \theta_2 \times \mathbf{S}_{\text{hc}}, \quad (19)$$

式中: \mathbf{S}_{ta} , \mathbf{S}_{hc} 分别为目标感知特征响应图和手工特征响应图; θ_1 和 θ_2 为融合系数。

得到响应图 \mathbf{S}_{fu} 后, 可根据最大响应分数 $\mathbf{S}_{\text{fu}}^{\max}$ 判断当前帧跟踪结果的好坏。如图 4 所示,

第 2 帧的跟踪目标没有受到干扰, 最大响应分数较大, 跟踪结果良好; 第 73 帧, 跟踪目标被相似物干扰, 最大响应分数减小, 跟踪结果开始出现偏差; 第 110 帧, 跟踪目标被严重遮挡, 最大响应分数很小, 跟踪结果越来越差。如果把响应分数最大值点预测为目标位置, 那么在第 110 帧, 由于遮挡和相似物干扰, 导致目标位置出现错判。因此, 仅采用最大响应分数作为模型更新的判据会因判断不准确使跟踪发生偏移。

跟踪结果置信度包含最大响应分数和平均峰值相关能量 (Average Peak Correlation Energy, APCE)^[16] 两个指标。APCE 描述响应图波动程度, 定义为

$$E_{\text{APCE}} = \frac{|\mathbf{S}_{\text{fu}}^{\max} - \mathbf{S}_{\text{fu}}^{\min}|^2}{\text{mean} \left(\sum_{m,n} (\mathbf{S}_{\text{fu}}^{m,n} - \mathbf{S}_{\text{fu}}^{\min})^2 \right)}, \quad (20)$$

式中: $\mathbf{S}_{\text{fu}}^{\max}$, $\mathbf{S}_{\text{fu}}^{\min}$ 分别为响应图最大值和最小值; $\mathbf{S}_{\text{fu}}^{m,n}$ 为响应图第 m 行第 n 列的值。

如图 5 所示, 第 2 帧目标周围无干扰物, $\mathbf{S}_{\text{fu}}^{\max}$, E_{APCE} 分别为 0.139 3 和 191.029 1, 最大响应分数较大, 响应图较为平坦, 跟踪置信度较高; 第 15 帧, 目标被相似物干扰, $\mathbf{S}_{\text{fu}}^{\max}$, E_{APCE} 分别为 0.081 6 和 99.773 3, 最大响应分数减小, 响应图轻微波动, 跟踪置信度降低; 第 41 帧, 目标被严重遮挡, $\mathbf{S}_{\text{fu}}^{\max}$, E_{APCE} 急剧下降为 0.039 3 和 43.793 4, 响应图剧烈波动, 跟踪置信度很低。最大响应分数和 APCE 从局部响应值情况和整体波动程度这两方面反映了跟踪置信度。因此, 可采用这两个指标作为模型更新的重要判据。

深度特征的引入使模型参数量急剧增加。由于跟踪过程中训练数据的稀缺性, 训练数据量小于模型参数量, 可能会导致模型过拟合。而且, 逐帧更新的传统策略虽然能使模型较好地适应不断变化的目标外观, 但过度更新也会使模型对最新几帧样本过拟合。此外, 相邻两帧的目标通常不会有较大变化, 采用前一帧的模型仍可精确定位当前帧目标。因此, TACF 采用隔帧更新方法以提高算法的运行速度, 防止模型过拟合。

综合最大相应分数、APCE 和隔帧更新策略, 提出了多指标更新方法, 如式(21)所示。

$$\begin{cases} \mathbf{S}_{\text{fu}}^{\max} \geq \beta_1 \overline{\mathbf{S}_{\text{fu}}^{\max}}, \\ E_{\text{APCE}} \geq \beta_2 \overline{E_{\text{APCE}}}, \\ (t+1) \bmod 2 = 0, \end{cases} \quad (21)$$

式中： $\overline{S_{fu}^{\max}}$ ， $\overline{E_{APCE}}$ 分别为 S_{fu}^{\max} 和 E_{APCE} 的历史均值。

当 S_{fu}^{\max} ， E_{APCE} 分别大于等于 $\beta_1 \overline{S_{fu}^{\max}}$ 和 $\beta_2 \overline{E_{APCE}}$ ，且帧数 $(t+1)$ 为2的倍数时，TACF算法依

式(3)进行模型更新；若不满足条件，则不进行模型更新，避免了因学习到大量干扰信息而导致的模型损坏。

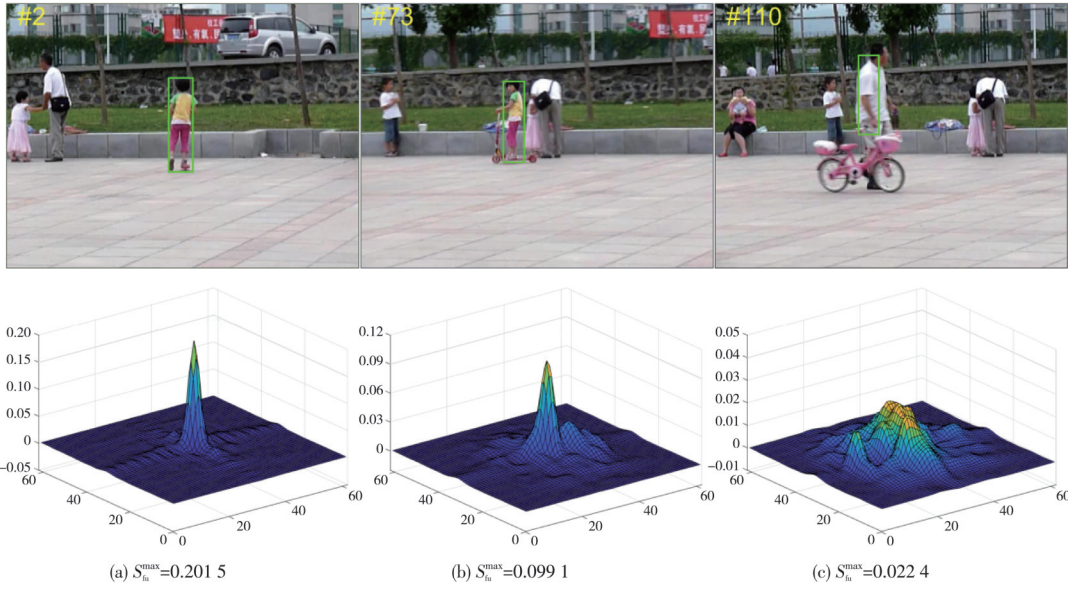


图4 不同跟踪场景下的最大响应分数

Fig. 4 Maximum response score under different tracing scenarios

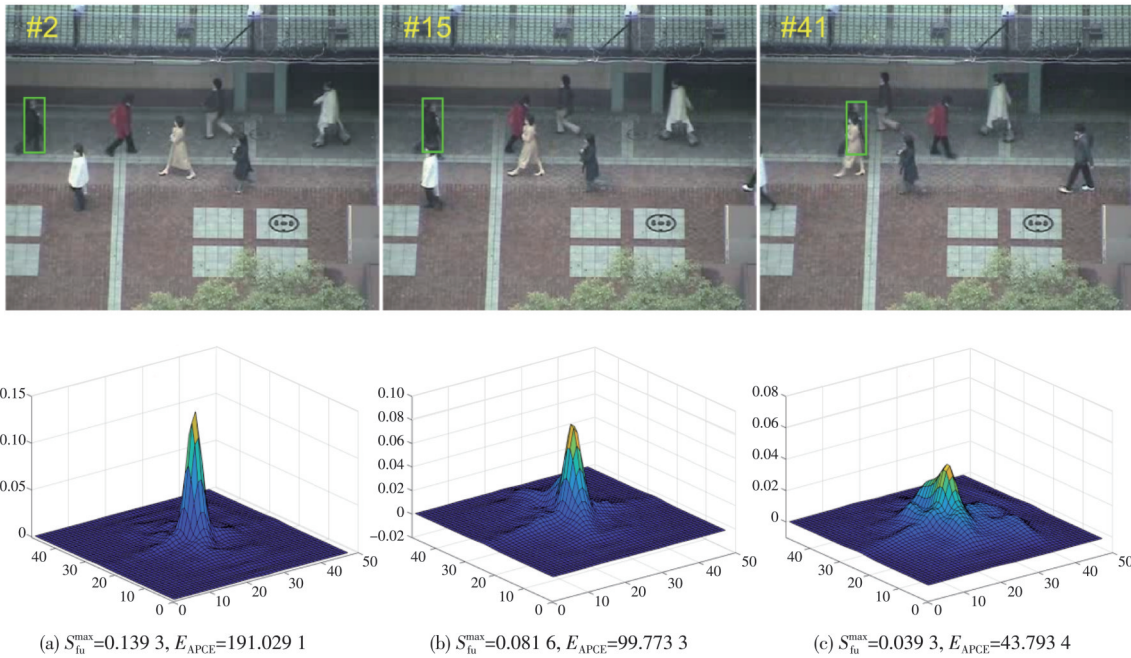


图5 不同跟踪场景下的跟踪置信度

Fig. 5 Tracking confidence under different tracing scenarios

2.4 算法流程

TACF算法流程如图6所示。

跟踪阶段，TACF算法对预训练网络提取的深度特征进行通道选择，得到适合跟踪任务的目

标感知特征，以提升对目标的表征能力；同时，采用ADMM算法优化跟踪模型，加快相关滤波器的训练速度；最后，根据最大响应分数、APCE和隔帧更新策略判断是否更新模型，避免模型损坏。

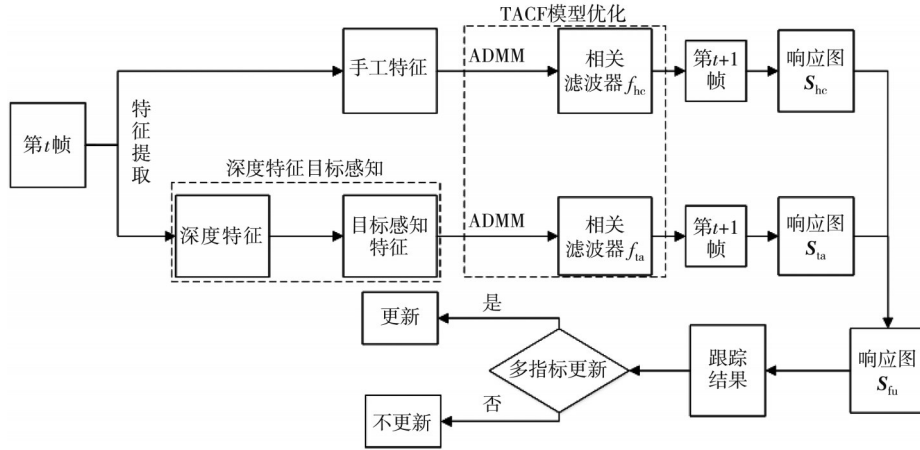


图 6 TACF 算法流程图

Fig. 6 The flow chart of the TACF algorithm

3 实验结果及分析

3.1 实验环境与配置

实验采用 Windows10 操作系统下的 Matlab2019a 对 TACF 算法进行实现, 并通过 Mat-ConvNet 工具包实现预训练网络 VGG-16^[17] 的前向传播。硬件平台配置为: 16 GB 内存, AMD R5 4600H 6 核 3.00 GHz CPU, GTX1650 GPU。

ADMM 参数设置为: 初始步长 $\rho = 1$, 最大步长 $\rho_{\max} = 10\,000$, 尺度 $\gamma = 10$ 。多指标更新参数设置为: $\beta_1 = 0.6$, $\beta_2 = 0.5$, 学习率 $\eta = 0.0185$ 。响应图融合系数设置为: $\theta_1 = 1$, $\theta_2 = 1$ 。深度特征目标感知参数设置为: 回归损失收敛阈值 = 0.02, 最大迭代次数 = 200, 原始深度特征为 VGG-16 中 Conv4_3 层的输出。

3.2 数据集与评价指标

实验数据集为 OTB2015^[18] (object tracking benchmark), 该数据集含目标部分离开视野、遮挡、平面内旋转、背景杂乱等 11 种场景属性。

实验采用一次通过评估 (One Pass Evaluation, OPE)、空间鲁棒性评估 (Spatial Robustness Evaluation, SRE)、时间鲁棒性评估 (Temporal Robustness Evaluation, TRE)、11 种复杂场景下的精确度和成功率作为评价指标。OPE, SRE, TRE、精确度和成功率的定义如下:

1) OPE: 在目标初始信息为理想的情况下, 算法维持其跟踪性能的特性。

2) SRE: 在首帧的真实跟踪框出现位置平移或尺度缩放的情况下, 算法维持其跟踪性能的特性^[21]。

3) TRE: 在起始帧发生改变 (不为第 1 帧) 的情况下, 算法维持其跟踪性能的特性^[18]。

4) 精确度: 预测中心位置与真实中心位置的欧氏距离小于 20 pixel 的帧数占总帧数的比值。

5) 成功率: 预测框与真实框的交并比大于 0.5 的帧数占总帧数的比值。

OPE 模式下的精确度和成功率主要反映跟踪算法在理想情况下的跟踪性能, 是目标跟踪领域常用的评价指标。而 SRE 和 TRE 模式下的精确度和成功率能在一定程度上评价算法在真实情况下的跟踪性能, 是更加重要的评价指标。

3.3 定量分析

为评估所提算法性能, 将 TACF 算法与 SRDCF, AutoTrack^[19], TADT^[20], CFNet, DCFNet, UDT, STRCF^[21], LADCF^[22] 8 种主流算法进行比较, 实验结果如表 1 所示, OPE 精确度、成功率如图 7 所示。

表 1 OTB2015 数据集上的实验结果

Tab. 1 The experimental results on OTB2015 dataset

算法	精确度	成功率
Ours	0.902	0.691
AutoTrack	0.787	0.587
STRCF	0.855	0.651
TADT	<u>0.864</u>	<u>0.658</u>
LADCF	0.837	0.646
DCFNet	0.818	0.626
UDT	0.770	0.595
SRDCF	0.763	0.583
CFNet	0.749	0.570

注: 加粗、下划线数字分别表示最优和次优结果。

由表 1 和图 7 可知, TACF 算法的 OPE 精确度、成功率分别为 0.902 和 0.691, 相较于 Auto-

Track, STRCF, TADT, LADCF, DCFNet, UDT, SRDCF和CFNet, 精确度分别提高11.5%, 4.7%, 3.8%, 6.5%, 8.4%, 13.2%, 13.9%和

15.3%, 成功率分别提高10.4%, 4.0%, 3.3%, 4.5%, 6.5%, 9.6%, 10.8%和12.1%。

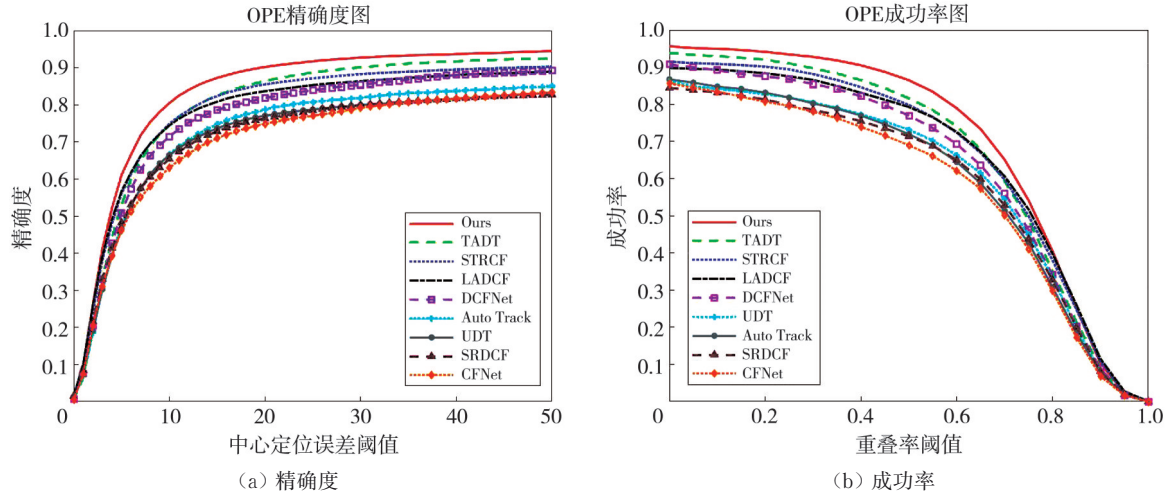


图7 OPE精确度和成功率

Fig.7 Precision and success rate of OPE

在初始帧数改变、第一帧真实框平移或尺度变化的情况下, 跟踪算法性能可能有较大变化。因此, 采用TRE、SRE对TACF和其它8种对比算法作进一步分析, 实验结果如表2、表3所示。

表2 9种算法的时间鲁棒性评估

Tab.2 TRE for the nine algorithms

	Ours	AutoTrack	STRCF	TADT	LADCF	DCFNet	UDT	SRDCF	CFNet
精确度	0.898	0.827	0.842	<u>0.867</u>	0.840	0.804	0.777	0.786	0.772
成功率	0.695	0.625	0.652	<u>0.673</u>	0.656	0.620	0.606	0.615	0.599

注: 加粗、下划线数字分别表示最优和次优结果。

表3 9种算法的空间鲁棒性评估

Tab.3 SRE for the nine algorithms

	Ours	AutoTrack	STRCF	TADT	LADCF	DCFNet	UDT	SRDCF	CFNet
精确度	0.873	0.760	0.823	<u>0.831</u>	0.813	0.752	0.725	0.761	0.734
成功率	0.646	0.553	0.610	<u>0.622</u>	0.609	0.557	0.543	0.562	0.545

注: 加粗、下划线数字分别表示最优和次优结果。

表4 9种算法在不同场景下的成功率

Tab.4 The success rate of the 9 methods under various scenarios

	成功率								
	Ours	AutoTrack	STRCF	TADT	LADCF	DCFNet	UDT	SRDCF	CFNet
SV	0.667	0.537	0.627	<u>0.652</u>	0.613	0.598	0.566	0.544	0.568
IV	0.708	0.602	0.656	<u>0.680</u>	0.647	0.613	0.570	0.584	0.582
DEF	0.651	0.555	<u>0.604</u>	0.602	0.577	0.537	0.518	0.532	0.489
OCC	0.663	0.549	0.602	<u>0.641</u>	0.620	0.608	0.574	0.538	0.518
MB	0.693	0.580	0.642	<u>0.670</u>	0.621	0.610	0.581	0.559	0.520
FM	0.692	0.578	0.620	<u>0.658</u>	0.615	0.635	0.599	0.573	0.531
IPR	0.667	0.550	0.598	0.618	0.587	<u>0.620</u>	0.590	0.519	0.537
OPR	0.678	0.549	0.618	<u>0.646</u>	0.614	0.606	0.580	0.536	0.538
OV	0.652	0.532	0.592	<u>0.624</u>	0.598	0.574	0.525	0.445	0.505
LR	<u>0.624</u>	0.534	0.543	0.633	0.541	0.532	0.531	0.492	0.501
BC	0.681	0.565	<u>0.655</u>	0.622	0.636	0.597	0.579	0.549	0.561

注: 加粗、下划线数字分别表示最优和次优结果。

由表 2 和表 3 可知, TACF 算法的 TRE 精确度、成功率分别为 0.898 和 0.695, 相较于 SRDCF 算法分别提高 11.2% 和 8.0%; SRE 精确度、成功率分别为 0.873 和 0.646, 相较于 SRDCF 算法分别提高 11.2% 和 8.4%。由表 2 和

表 3 还可知, TACF 算法的 TRE 和 SRE 均好于其它 8 种算法。

为分析跟踪算法在复杂场景下的性能, 实验对 9 种算法在 11 种场景下的 OPE 精确度和成功率进行比较, 结果如表 4、表 5 所示。

表 5 9 种算法在不同场景下的精确度

Tab. 5 The precision of the 9 methods under various scenarios

	精确度								
	Ours	AutoTrack	STRCF	TADT	LADCF	DCFNet	UDT	SRDCF	CFNet
SV	0.874	0.734	0.831	<u>0.860</u>	0.800	0.785	0.726	0.719	0.742
IV	0.887	0.787	0.825	<u>0.865</u>	0.804	0.788	0.727	0.742	0.734
DEF	0.869	0.731	<u>0.828</u>	0.818	0.774	0.715	0.676	0.706	0.662
OCC	0.864	0.731	0.787	<u>0.841</u>	0.793	0.793	0.745	0.691	0.673
MB	0.868	0.735	0.809	<u>0.833</u>	0.780	0.744	0.715	0.707	0.630
FM	0.890	0.746	0.794	<u>0.834</u>	0.774	0.806	0.747	0.730	0.667
IPR	0.892	0.771	0.805	0.830	0.786	<u>0.837</u>	0.789	0.705	0.718
OPR	0.901	0.761	0.831	<u>0.872</u>	0.814	0.799	0.769	0.717	0.723
OV	0.890	0.709	0.781	<u>0.816</u>	0.788	0.739	0.668	0.566	0.663
LR	<u>0.848</u>	0.768	0.744	0.881	0.734	0.719	0.722	0.658	0.742
BC	0.874	0.760	<u>0.860</u>	0.805	0.830	0.787	0.757	0.703	0.730

注: 加粗、下划线数字分别表示最优和次优结果。

由表 4、表 5 可知, TACF 算法的成功率和精确度在低分辨率(Low Resolution, LR)场景下仅比 TADT 算法略低, 在其它 10 种场景下均为最优。与 SRDCF 算法相比, TACF 算法的跟踪性能在所有场景下均有较大提升, 尤其在超出视野(Out-of-View, OV)、运动模糊(Motion Blur, MB)和遮挡场景(Occlusion, OCC)下, 成功率分别提升 20.7%、13.4% 和 12.5%, 精确度分别提升 32.4%、16.1% 和 17.3%。

3.4 定性分析

实验挑选 OTB2015 的 5 个典型视频序列以定性分析 TACF 和 8 种主流对比算法。所选视频序列含背景杂乱、遮挡、快速运动和尺度变化等多种场景属性。实验结果如图 8 所示。

视频序列 bolt2 跟踪过程中, 主要存在的挑战因素是背景杂乱。从第 10 帧到第 292 帧, 跟踪目标被视频背景中的相似物干扰, SRDCF, DCFNet, UDT 和 AutoTrack 相继出现严重的跟踪漂移, TADT 和 LADCF 跟踪到相似物上, 仅 TACF、STRCF 和 CFNet 能准确定位目标, 如图 8(a) 所示。目标被相似物干扰时, TACF 的跟踪效果优于 SRDCF, 是因为 TACF 采用的深度特征目标感知策略有效区分了目标与背景。

视频序列 girl2 跟踪过程中, 主要存在的挑战因素是遮挡。如图 8(b) 所示, 从第 109 帧到第 120 帧, 目标被干扰物严重遮挡, 仅 TACF 算法能对目标实

现精确和鲁棒的跟踪, 而其它对比算法均出现不同程度的跟踪漂移; 到第 170 帧, TACF 仍可精确定位目标, 而其它对比算法均完全丢失了跟踪目标。遮挡场景下, TACF 的跟踪效果优于 SRDCF, 是因为 TACF 采用的多指标更新方法避免了由于学习到过多干扰信息而造成的模型腐败。



图 8 不同算法的跟踪结果

Fig. 8 Tracking results of the different methods

视频序列 biker 跟踪过程中, 存在运动模糊、平面外旋转和尺度变化等多种挑战因素。从第 65 帧到第 75 帧, 目标跳跃、旋转, 出现运动模糊, SRDCF 完全丢失目标, TADT, STRCF, CFNet 和 UDT 均发生一定程度的跟踪漂移; 从第 105 帧到第 130 帧, AutoTrack、LADCF 和 DCFNet 也完全丢失了目标, 而 TACF 一直可实现较为精确和鲁棒的跟踪, 如图 8(c) 所示。

视频序列 matrix 跟踪过程中, 存在光照变化、快速运动以及背景杂乱等多种挑战因素。第 13 帧, 目标被相似物干扰, CFNet、LADCF 和 AutoTrack 出现跟踪漂移; 从第 46 帧到第 100 帧, 目标快速运动, 出现光照变化, TADT, STRCF, DCFNet、UDT 和 SRDCF 均发生不同程度的跟踪漂移, 仅 TACF 能一直精确定位目标, 如图 8(d) 所示。

视频序列 DragonBaby 跟踪过程中, 存在超出视野、快速运动和平面内旋转等多种挑战因素。从第 24 帧到第 113 帧, 目标剧烈运动, AutoTrack, TADT, DCFNet, UDT, SRDCF 和 CFNet 均出现不同程度的跟踪漂移, 而 TACF 自始至终能准确跟踪目标, 如图 8(e) 所示。

3.5 消融实验

针对所提深度特征目标感知 (Target-Aware, TA)、交替方向乘子法 (ADMM) 和多指标更新 (Multi-Index Update, MIU) 方法, 实验在 OTB2015 上对 TACF 算法作消融分析, 结果如表 6 所示。

由表 6 第 1~2 行可知, SRDCF-ADMM 算法的精确度、成功率虽比 SRDCF 算法分别下降 0.6% 和 7.3%, 但是跟踪速度提升了 375%。这说明采用 ADMM 方法求解相关滤波器, 降低了算法时间复杂度, 加快了跟踪速度。由表 6 第 2~3 行可看出, 在 SRDCF-ADMM 基础上, 融入深度特征目标感知方法, 精确度、成功率分别提升 15.5% 和 23.3%, 这说明深度特征目标感知方法有效提升了算法对目标的表征能力, 提高了跟踪准确性。由表 6 第 3~4 行还可看出, 在采用了深度特征目标感知方法 SRDCF-ADMM 基础上, 融入多指标更新方法, 精确度、成功率、跟踪速度分别提升 2.9%, 3.7% 和 24.5%, 这说明多指标更新方法不仅提升了算法运算效率, 还避免了因学习到错误信息而导致的模型损坏。

表 6 消融实验
Tab. 6 Ablation experiment

SRDCF	SRDCF-ADMM	TA	MIU	精确度	成功率	跟踪速度/FPS
√				0.763	0.583	8.9
	√			0.758	0.54	42.3
	√	√		0.876	0.666	4.9
	√	√	√	0.902	0.691	6.1

4 结论

本文提出了深度特征目标感知 ADMM 优化多指标更新相关滤波跟踪算法。该算法融入预训练网络提取的深度特征, 并根据回归损失的梯度信息进行通道选择, 提高了对目标的表征能力; 采用 ADMM 方法训练相关滤波器, 降低计算复杂度, 加快跟踪速度; 根据多指标更新方法控制模型更新, 不仅提升了算法运行效率, 而且避免了由于学习到大量干扰信息而导致的模型损坏。实验结果表明, TACF 算法的成功率、精确度在数据集 OTB2015 上均优于其它 8 种对比算法, 且在复杂场景下实现了更鲁棒的跟踪。

参考文献:

- [1] 张长弓, 杨海涛, 王晋宇, 等. 基于深度学习的视觉单目标跟踪综述[J]. 计算机应用研究, 2021, 38(10): 2888-2895.
ZHANG Changgong, YANG Haitao, WANG Jinyu, et al. Survey on visual single object tracking based on deep learning[J]. Application Research of Computers, 2021, 38(10): 2888-2895. (in Chinese)
- [2] 韩瑞泽, 冯伟, 郭青, 等. 视频单目标跟踪研究进展综述[J]. 计算机学报, 2022, 45(9): 1877-1907.
HAN Ruize, FENG Wei, GUO Qing, et al. Single object tracking research: a survey[J]. Chinese Journal of Computers, 2022, 45(9): 1877-1907. (in Chinese)
- [3] 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-Volume Part IV, 2012: 702-715.
- [4] 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, 2015, 37(3): 583-596.
- [5] LIANG Y, WU Q, LIU Y, et al. Deep correlation filter tracking with shepherded instance-aware proposals[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(8): 11408-11421.
- [6] SUN S, GONG J, LI Y, et al. A research of target tracking algorithm based on deep learning and kernel correlation filter [C]//International Joint Conference on Neural Networks (IJCNN), 2022: 1-8.
- [7] HOU Y, LIN X, LI J. Correlation filter and deep siamese network hybrid algorithm for visual object tracking[C]//IEEE 6th International Conference on Intelligent Computing and Signal Processing (ICSP), 2021: 73-76.
- [8] CHAO M, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking [C]//IEEE International Conference on Computer Vision, 2015: 3074-3082.
- [9] QI Y, ZHANG S, QIN L, et al. Hedged deep tracking [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4303-4311.
- [10] VALMADRE J, BERTINETTO L, HENRIQUES J F, et al. End-to-end representation learning for correlation filter based tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5000-5008.
- [11] WANG Q, GAO J, XING J, et al. DCFNet: discriminant correlation filters network for visual tracking [J]. arXiv Preprint arXiv: 1704.04057, 2017.
- [12] WANG N, SONG Y, MA C, et al. Unsupervised deep tracking [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1308-1317.
- [13] DANELLJAN M, HAGER G, KHAN F S, et al. Learning spatially regularized correlation filters for visual tracking[C]//IEEE International Conference on Computer Vision, 2015: 4310-4318.
- [14] BOYD S, PARIKH N, CHU E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. Foundations and Trends in Machine Learning, 2010, 3(1): 1-122.
- [15] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//IEEE International Conference on Computer Vision, 2017: 618-626.
- [16] WANG M, LIU Y, HUANG Z. Large margin object tracking with circulant feature maps [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4800-4808.
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv Preprint arXiv: 1409.1556, 2014.
- [18] WU Y, LIM J, YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [19] LI Y, FU C, DING F, et al. AutoTrack: towards high-performance visual tracking for UAV with automatic spatio-temporal regularization [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11920-11929.
- [20] LI X, MA C, WU B, et al. Target-aware deep tracking [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1369-1378.
- [21] LI F, TIAN C, ZUO W, et al. Learning spatial-temporal regularized correlation filters for visual tracking [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4904-4913.
- [22] 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 tracking [J]. IEEE Transactions on Image Processing, 2019, 28(11): 5596-5609.