

4-12-2023

Adaptive Correction Tracking Algorithm Based on Detector and Locator Fusion

Yecai Guo

School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; guo-yecai@163.com

Cheng Liu

School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Adaptive Correction Tracking Algorithm Based on Detector and Locator Fusion

Abstract

Abstract: In order to avoid tracking failure caused by occlusion, rotation and other factors in complex dynamic scenes, *an adaptive correction tracking algorithm based on detector and locator fusion is proposed. The locator trains a convolutional neural network (CNN) filter for location estimation by extracting the deep features of target. The CNN filter adds two layers of shallow features to the three layers of the convolution features of original CF2 algorithm, which enhances the extraction of target texture information. The detector calculates the confidence score by extracting histogram of oriented gradient(HOG) feature of target and combining the context information. The average peak-to-correlation energy (APCE) and maximum response value of current frame are compared separately with the historical average to comprehensively judge whether the tracking fails are due to occlusion and other factors. If the tracking fails, combine the detector to relocate the target, otherwise estimate the scale of the target. Update the model when the model has high confidence.* The experimental results show that the distance accuracy and overlap accuracy of the algorithm are good.

Keywords

convolution features, detector, tracking, adaptive correction, high confidence

Recommended Citation

Yecai Guo, Cheng Liu. Adaptive Correction Tracking Algorithm Based on Detector and Locator Fusion[J]. Journal of System Simulation, 2023, 35(4): 709-720.

基于检测器与定位器融合的自适应校正跟踪算法

郭业才, 刘程

(南京信息工程大学 电子与信息工程学院, 江苏 南京 210044)

摘要: 针对传统目标跟踪算法在复杂动态场景中因目标发生遮挡、旋转等多种因素而导致的跟踪失败问题, 提出了一种基于检测器与定位器融合的自适应校正跟踪算法。定位器通过提取目标的深度特征训练 CNN(convolutional neural network) 滤波器进行位置估计, CNN 滤波器在原 CF2(hierarchical convolutional features for visual tracking) 算法的 3 层卷积特征的基础上加入了 2 层浅层特征, 增强了对目标纹理信息的提取。检测器通过提取目标的 HOG(histogram of oriented gradient) 特征, 结合上下文信息计算置信度评分, 用当前帧的平均峰值能量和响应最大值分别与历史均值比较, 综合判断是否因为遮挡等因素导致跟踪失败, 如果跟踪失败, 结合检测器进行目标的重定位, 否则对目标进行尺度估计。当模型具有高置信度时, 更新模型。实验结果表明: 算法距离精度和重叠精度均取得不错的效果。

关键词: 卷积特征; 检测器; 跟踪; 自适应校正; 高置信度

中图分类号: TP391.41

文献标志码: A

文章编号: 1004-731X(2023)04-0709-12

DOI: 10.16182/j.issn1004731x.joss.21-1273

引用格式: 郭业才, 刘程. 基于检测器与定位器融合的自适应校正跟踪算法[J]. 系统仿真学报, 2023, 35(4): 709-720.

Reference format: Guo Yecai, Liu Cheng. Adaptive Correction Tracking Algorithm Based on Detector and Locator Fusion [J]. Journal of System Simulation, 2023, 35(4): 709-720.

Adaptive Correction Tracking Algorithm Based on Detector and Locator Fusion

Guo Yecai, Liu Cheng

(School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: In order to avoid tracking failure caused by occlusion, rotation and other factors in complex dynamic scenes, an adaptive correction tracking algorithm based on detector and locator fusion is proposed. The locator trains a convolutional neural network (CNN) filter for location estimation by extracting the deep features of target. The CNN filter adds two layers of shallow features to the three layers of the convolution features of original CF2 algorithm, which enhances the extraction of target texture information. The detector calculates the confidence score by extracting histogram of oriented gradient(HOG) feature of target and combining the context information. The average peak-to-correlation energy (APCE) and maximum response value of current frame are compared separately with the historical average to comprehensively judge whether the tracking fails are due to occlusion and other factors. If the tracking fails, combine the detector to relocate the target, otherwise estimate the scale of the target. Update the model when the model has high confidence. The experimental results show that the distance accuracy and overlap accuracy of the algorithm are good.

Keywords: convolution features; detector; tracking; adaptive correction; high confidence

收稿日期: 2021-12-22 修回日期: 2022-03-16

基金项目: 国家自然科学基金(61673222)

第一作者: 郭业才(1962-), 男, 教授, 博士, 研究方向为通信信号处理、水声信号处理等。E-mail: guo-yecai@163.com

0 引言

目标跟踪算法主要分为基于生成式的目标跟踪算法、基于判别式的目标跟踪算法^[1]和基于深度学习的目标跟踪算法^[2]。传统的基于生成式的目标跟踪算法由于跟踪模型的精度与成功率较差,已经不能满足人们的日常需求。基于深度学习的目标跟踪算法尽管精度与成功率都很高^[3],但其训练时间越来越长^[4],网络结构越来越复杂^[5],对设备的要求也越来越高^[6]。而判别式跟踪算法由于能较好地复杂场景下进行跟踪,且结合卷积神经网络后可以通过训练降低训练模型时间,对设备要求也不高,得到了更多国内外学者的重视^[7-8]。

传统的结合卷积神经网络的判别式跟踪算法在遇到遮挡或旋转等复杂场景时,会发生不同程度的跟踪漂移甚至出现跟踪失败。以CF2(hierarchical convolutional features for visual tracking)^[9]算法为代表,由于没有尺度变换,且缺少对目标区域的有效检测,单纯依靠提取的深度特征结合相关滤波器进行由粗到细的定位跟踪,当遇到目标被遮挡或旋转时,跟踪器无法准确定位目标,从而导致跟踪失败。

面对目标跟踪中复杂场景的变化,不少跟踪器都做出了相关的探索与改进。TLD(tracking-learning-detection)^[10]算法将传统的跟踪算法与检测算法相结合,通过一种改进的在线学习机制不断更新检测模块的目标模型与参数来解决目标被遮挡的问题。其主要将目标跟踪分为3个模块:第1个模块为跟踪模块,加入了跟踪失败检测。第2个模块为检测模块,对每一帧图片进行全局搜索以定位目标。第3个模块为学习模块,通过跟踪模块的结果对检测模块的错误进行评估。根据评估结果能更新检测模块的模型,能有效缓解因部分遮挡引起的跟踪失败问题,但是其检测模块扫描大量窗口会导致检测时间过长,并且在跟踪过程中当目标发生严重遮挡、形变时,算法会出现跟踪失败。MUSTer(multi-store tracker)^[11]算法提

出了长短时记忆思想,在短时跟踪中利用聚合通道相关滤波储存目标信息,当目标跟踪失败时,启动长时跟踪模块,同时进行目标重检测,并使用关键点匹配来预测跟踪成功,提高了模型在复杂场景下的跟踪成功率,但是精度不高。LCT(long-term correlation tracking)^[12]算法结合了模型上下文之间的学习,而不是简单的使用二分类模型,将跟踪问题进一步分解为目标的平移估计和尺度估计,其中平移估计采用上下文之间的学习;尺度估计使用相关滤波器,同时增加了一个高置信度的检测器,通过检测器与固定阈值的比较来判断是否跟踪失败,尽管能在目标被遮挡时重新检测目标位置,但是只通过当前帧响应最大值与固定阈值的比较来判断当前目标是否被遮挡,存在一定误差。后改进的LCT+^[13]算法则将LCT中的随机阈分类器换成了SVM(support vector machine)分类器,并引入了深度特征,大幅提高了跟踪器的性能。在CF2算法的基础上进行改进的HCFtstar(robust visual tracking via hierarchical convolutional features)^[14]算法,通过学习另一个具有保守学习率的相关滤波器,以保持目标外观作为区分分类器的长期记忆,同时加入了重检测模块,重检测模块通过学习另外一个具有保守学习率的相关滤波器,用来判断目标跟踪失败后的目标位置,但是其需要通过计算当前帧的响应分数是否小于设定的固定阈值来判断是否跟踪失败,不能够自适应地判断当前帧是否跟踪失败。

本文在CF2算法上进一步改进,在原始提取的3层深度特征上,加入了2层纹理信息较重的底层特征进一步准确估计目标位置,同时提取目标的HOG(histogram of oriented gradient)特征训练检测器,结合上下文的信息,利用当前帧的平均峰值能量与响应最大值分别与历史帧比较来自适应判断当前帧是否跟踪失败,如果跟踪失败则自适应开启目标的重定位,能够在目标发生遮挡时,及时搜寻目标的新位置,从而大幅提高模型的精度。

1 相关滤波器的训练

1.1 HOG 检测器训练

受 TLD 算法^[10]的启发, 将检测器与追踪器结合起来进行目标的定位, 使算法的精度大幅提升, 本文提取目标的 HOG 特征训练检测器。给定一个图像补丁 z 和学习相关滤波器 w , 相关响应为 $f(z) = w^T z$ 。学习相关滤波器 w 可以用来解决岭回归问题:

$$\min_w \sum_{k=1}^T \left\| \sum_{d=1}^D x_k^d * w^d - y_k \right\|^2 + \lambda \sum_{d=1}^D \|w^d\|^2 \quad (1)$$

式中: $*$ 为循环相关; x_k^d 为 d 维通道 x 中的 k 次样本, 对于每个移位样本 x_k , $k \in \{1, 2, \dots, T\}$, 高斯函数 $y_k \in y$; w^d 为 w 的第 d 次通道; λ 为正则化参数且 $\lambda > 0$, 则其频域解为

$$w = \frac{x^T y}{x^T x + \lambda} \quad (2)$$

结合 KCF^[15] 算法中对高斯内核进行重新处理, 将 w 表达为样本的线性组合:

$$w = \sum_i \alpha_i \phi(x_i) \quad (3)$$

通过映射优化 α 而不是 w 通过映射到非线性特征空间 $\phi(x)$ 。内核功能满足 $\phi(x)^T \phi(x') = k(x, x')$ 和内核矩阵 $K_{i,j} = k(x_i, x_j)$ 。

制定快速核回归的解决方案为

$$\alpha = (K + \lambda I)^{-1} y \quad (4)$$

通过离散傅里叶变换对角化为

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda} \quad (5)$$

式中: $\hat{\cdot}$ 为离散傅里叶变换。检测图像块 z 内的目标位置为

$$\hat{f}(z) = \hat{k}^{xz} \hat{\alpha} \quad (6)$$

式中: \hat{k}^{xz} 是 k^{xz} 的离散傅里叶变换, 这表示 x 和 z 的内核相关性。高斯内核可以表示为

$$k^{xz} = \exp \left(-\frac{1}{\sigma^2} \left(\|x\|^2 + \|z\|^2 - 2F^{-1} \left(\sum_c \hat{x}_c^* \cdot \hat{z}_c \right) \right) \right) \quad (7)$$

式中: σ 为高斯内核的带宽; F^{-1} 为离散傅里叶逆变换; $*$ 为复合缀合; c 为图像特征的通道。在跟踪过程中, 首先裁剪前一帧中位置的图像块 z , 然后在傅里叶域中使用目标模板 x 计算响应。

$$f(z) = F^{-1}(\hat{f}(z)) = F^{-1}(\hat{k}^{xz} \hat{\alpha}) \quad (8)$$

对于每个指定的输入图像块 z , 设置其置信度评分^[16]为

$$p(z) = \text{sum}(k^{xz} \cdot \alpha) \quad (9)$$

式中: 输出 $p(z)$ 为矩阵中的所有元素的求和, 是标量值而不是矩阵。

1.2 CNN 滤波器训练

本文 CNN 滤波器的训练选择的卷积神经网络为 VGGNet19, 其结构如图 1 所示。

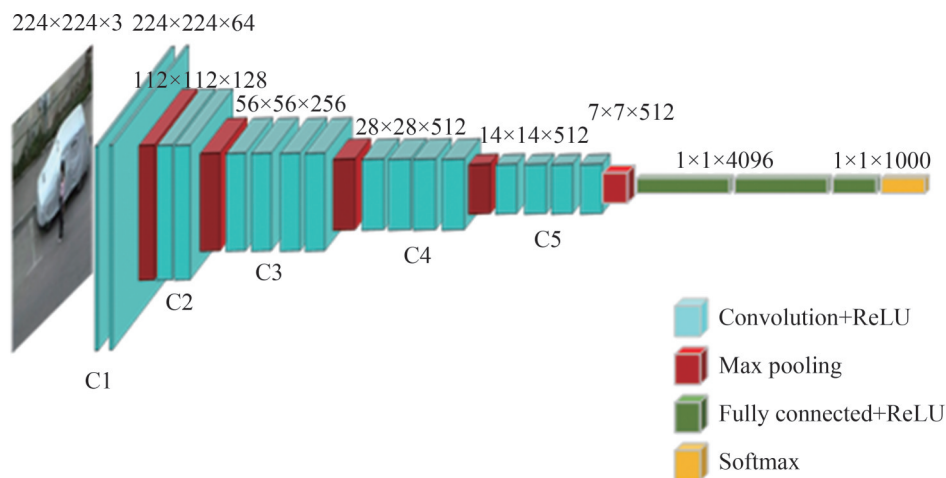


图1 VGGNet19 的结构
Fig. 1 Structure of VGGNet19

<http://www.china-simulation.com>

由图1可知, VGGNet19网络共19层, 包含16个卷积层和3个全连接层, 其中16个卷积层被分成了5个卷积块C1~C5。C1由2个 $3 \times 3 \times 64$ 的卷积层, C2由2个 $3 \times 3 \times 128$ 的卷积层组成^[17], C3由4个 $3 \times 3 \times 256$ 的卷积层组成, C4由4个 $3 \times 3 \times 512$ 的卷积层组成, C5由4个 $3 \times 3 \times 512$ 的卷积层组成, 其中每一个卷积层后有一个激活函数ReLU, 激活函数公式为 $\text{ReLU}(x) = \max(0, x)$, 每一个卷积块后有一个最大池化层, 第5个卷积块后接3个全连接层, 最后经过softmax层得到1000分类结果输出^[18]。本文VGGNet19已在大型ImageNet数据集上进行了预训练。

对于每一层的输出特征尺寸为 $M \times N \times D$, M , N 代表图像特征的宽度和高度^[19], D 为特征的通道数, 结合CF2算法训练CNN滤波器, 通过最小化问题学习具有循环移位样本 \mathbf{x} 的相关滤波器 \mathbf{w} 为

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \sum_k \|\mathbf{w} \cdot \mathbf{x} - y_k\|^2 + \lambda \|\mathbf{w}\|_2^2 \quad (10)$$

用傅里叶变换转换到频域, 解为

$$\mathbf{W}^d = \frac{\mathbf{Y} \odot \bar{\mathbf{X}}^d}{\sum_{d=1}^D \mathbf{X}^d \odot \bar{\mathbf{X}}^d + \lambda} \quad (11)$$

式中: \mathbf{Y} 为 y_k 的傅里叶变换; \odot 为乘积。

当给定下一帧的第 l 层图像 \mathbf{z} 后, 计算当前层的响应图为

$$f_l = \mathbf{F}^{-1} \left(\sum_{d=1}^D \mathbf{W}^d \odot \bar{\mathbf{Z}}^d \right) \quad (12)$$

本文在综合比较VGGNet19的16个卷积层后, 发现VGGNet19不同卷积层具有不同的表达能力。随着VGGNet19的网络深度不断加强, 特征图的语义信息得到不断增强, 但由于池化的作用, 特征图的空间分辨率逐渐减少, 纹理信息不断减弱。底层特征空间分辨率较高, 包含了大量的纹理信息, 而高层特征拥有大量的语义信息, 在具有形变和平面外旋转的目标场景下, 表现出强大的不变性。传统的目标跟踪算法, 只提取了VGGNet19的3层特征作卷积融合提取, 如CF2算

法只提取了conv3-4、conv4-4、conv5-4, 这导致其对目标的纹理信息掌握不够, 本文在此基础上加以改进, 加入纹理信息丰富的conv1_1和分辨率良好的conv2_2, 共5层卷积特征作加权融合。同时采用由粗到细的跟踪策略, 对每一个卷积层的特征通过式(12)输出一个独立的相关滤波响应, 设 $(m_q, n_q) = \arg \max_{m, n} f_l(m, n)$ 表示最大值在第 l 层的位置, 则目标在第 $l-1$ 层的位置估计为

$$\arg \max_{m, n} f_{l-1}(m, n) + \mu f_l(m, n) \quad (13)$$

$$\text{s.t. } |m - m_p| + |n - n_p| \leq r \quad (14)$$

式中: μ 为正则项加权值。式(14)为约束项, 表示第 $l-1$ 层相关响应图中只搜索以 (m_q, n_q) 和半径为 r 的区域。

2 基于检测器与定位器融合的自适应校正跟踪算法

2.1 自适应校正策略

传统的CF2算法在面对遮挡或旋转的复杂环境时, 会发生跟踪失败。本文将HOG滤波器与CNN滤波器结合起来进行目标的精确定位, HOG滤波器主要用来作为检测器, CNN滤波器主要用来进行位置估计。大多数现有的跟踪器不考虑检测是否准确, 一旦目标在当前帧中被错误检测、严重遮挡或完全丢失, 可能会导致后续帧的跟踪失败。本文引入了一种度量机制来确定当前目标对象是否因为遮挡等环境因素引起跟踪失败。LMCF^[20]提出了APCE的概念来解决相似物体干扰, 发现在目标周围有特征相似的干扰物体时, 响应图会有多个峰值, 且最高的一个有可能是干扰物体的, 这时候可能就会引起误判, 于是该算法提出将APCE与响应最大值结合用来进行模型更新的判断, 效果很好。受此启发, 当目标面临遮挡时, 是否可以将APCE与响应最大值结合进行对当前帧目标是否因为遮挡等环境因素引起的跟踪失败的检测。同时本文在实验时发现, 如果只用当前帧的APCE与响应最大值对当前帧是否

跟踪失败进行判断, 效果并不理想, 本文在此基础上进行改进, 计算包含当前帧的所有历史帧的 $APCE$ 与响应最大值, 用当前帧的 $APCE$ 与响应最大值分别与历史均值进行比较, 发现效果很好, 并且当满足跟踪失败的条件时, 算法模型能自适应地开启目标位置的重新检测, 大幅提高了算法的精度。

$$APCE = \frac{|f_{\max} - f_{\min}|^2}{\text{mean}\left(\sum_{w,h} (f_{w,h} - f_{\min})^2\right)} \quad (15)$$

式中: f_{\max} 、 f_{\min} 和 $f_{w,h}$ 分别为响应图的最大, 最小响应分数和 f 的第 w 行第 h 列元素。

$$\overline{APCE} = \frac{\sum_{k=1}^n APCE_k}{n} \quad (16)$$

$$\bar{f}_{\max} = \frac{\sum_{k=1}^n f_{\max_k}}{n} \quad (17)$$

式中: \overline{APCE} 和 \bar{f}_{\max} 分别为当前所有帧的 $APCE$ 均值与 f_{\max} 响应均值。

用 $APCE_{\text{current}}$ 表示当前帧 $APCE$ 值, $f_{\max, \text{current}}$

表示当前帧的 f_{\max} 值, 如果 $APCE_{\text{current}} < \beta_1 \cdot \overline{APCE}$ 且 $f_{\max, \text{current}} < \beta_2 \cdot \bar{f}_{\max}$, 则判断目标跟踪失败, 其中 β_1 与 β_2 为判断是否自适应校正系数, 此时需要使用检测器重新检测目标位置, 同时添加一个在线的 SVM 线性分类器, 该分类器主要目的是能够快速地区分前景与背景, 如图 2 所示。

重新检测的过程参考了 HCFTstar 算法, 首先通过 Edgebox 算法^[21]生成多尺度区域候选框, 然后计算当前帧(第 t 帧)生成的 k 个矩形框 b_t^k 与前一帧的目标矩形框 b_{t-1} 之间的中心位置距离:

$$L(b_t^k, b_{t-1}) = \exp\left(-\frac{1}{2\delta^2} \|b_t^k - b_{t-1}\|^2\right) \quad (18)$$

式中: δ 为初始目标矩形框的对角线长度。

结合检测器构造目标函数为

$$\arg \min_k \{p(b_t^k) + \tau L(b_t^k, b_{t-1}) | p(b_t^k) > S\} \quad (19)$$

式中: τ 和 S 分别为折衡系数和阈值。

同时带有自适应校正策略的改进算法与原始 CF2 算法在具体的含有遮挡与旋转的 6 组视频上进行了中心像素误差的对比, 中心像素误差为真实中心和预测中心的欧式距离, 具体如图 3 所示。

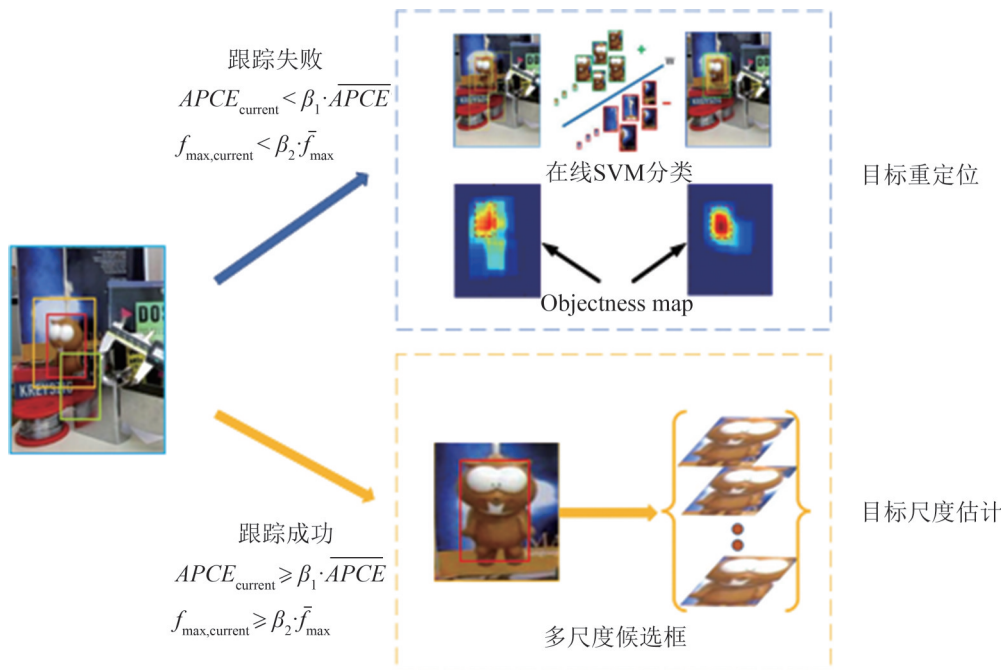


图2 自适应校正策略图
Fig. 2 Adaptive correction strategy diagram

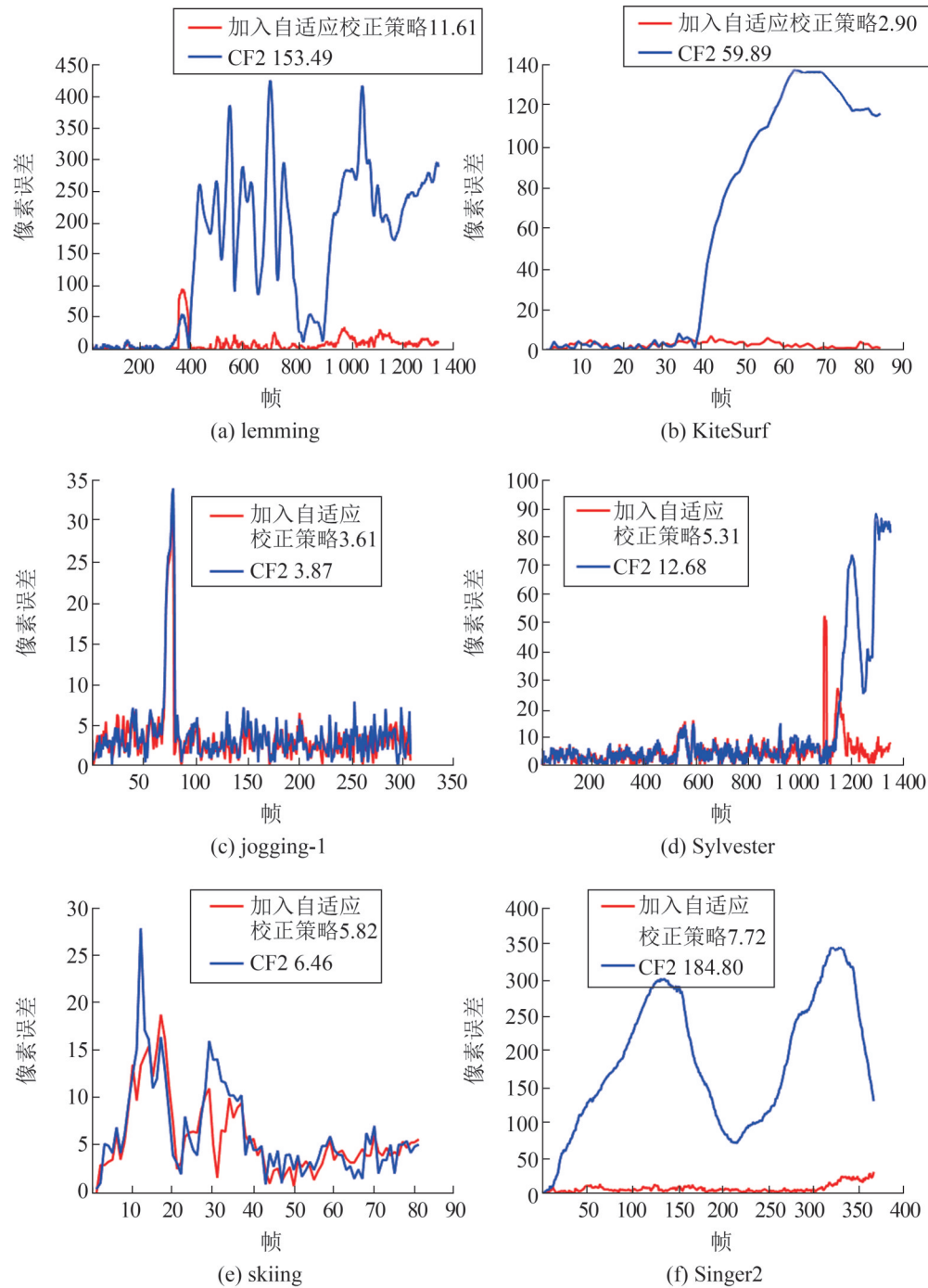


图3 中心像素误差的对比
Fig. 3 Comparison of center pixel errors

2.2 目标尺度估计

当目标没有发生跟踪失败, 即当 $APCE_{\text{current}} \geq \beta_1 \cdot \overline{APCE}$ 且 $f_{\text{max, current}} \geq \beta_2 \cdot \bar{f}_{\text{max}}$ 时, 在当前位置下计算尺度得分为 f_T , 然后以当前位置使用 EdgeBox

方法生成多尺度区域候选框, 使用非最大抑制阈值过滤掉与目标边界框区域重叠度交并比(以最后帧的比例为中心)小于 0.9 或大于 0.6 的候选框。通过使用非极大值抑制进行排序, 选取得分最高的前 200 个候选帧, 并计算每一个候选框的相关性得

分, 如果候选框的相关性得分大于 f_T , 则意味着目标发生了显著的尺度变化, 尺度更新方法如下:

$$(w_t, h_t) = (w_{t-1}, h_{t-1}) + \psi[(w_{p,t}, h_{p,t}) - (w_{t-1}, h_{t-1})] \quad (20)$$

式中: w_{t-1}, h_{t-1} 为目标第 $t-1$ 帧中候选框的宽度和高度; $w_{p,t}, h_{p,t}$ 为目标第 t 帧中候选框的宽度和高度; ψ 为尺度更新。

2.3 模型更新策略

相关过滤器必须随时间更新以适应目标外观的变化。传统的目标跟踪算法采用逐帧更新模型, 导致当目标遇到遮挡等复杂场景时, 会学习得到大量的干扰信息, 导致后续帧中模型的跟踪失败。本文采用高置信度模型更新策略, 只有当检测置信度 $p(z) > T$ 时, 更新检测器模型:

$$\hat{\alpha}^t = \eta \hat{\alpha}_N + (1 - \eta) \hat{\alpha}^{t-1} \quad (21)$$

$$\hat{x}^t = \eta \hat{x}^t + (1 - \eta) \hat{x}^{t-1} \quad (22)$$

式中: η 为学习率, 同时 CNN 滤波器的更新参照原 CF2 算法。

2.4 本文目标跟踪流程图

本文算法流程如图 4 所示。

3 实验结果及其分析

实验环境基于 Window10 系统和 Matlab 2020b, CPU 为 Inter(R) Core(TM) I7-8700 @ 3.20 Ghz, GPU 为 GTX1080。具体实验参数为正则化参数 $\lambda=0.0001$, 模型更新率 $\eta=0.01$, 模型更新的置信度 $T=0.3$, 高斯内核带宽 $\sigma=0.1$, 尺度更新 $\psi=0.5$, 折衡系数 τ 和阈值 S 分别设置为 0.1 与 0.15, 开启自适应校正参数 β_1 与 β_2 分别设置为 0.3 和 0.46。本文算法采用定量分析与定性分析 2 种评价方式。采用基于距离精度图和重叠精度图 2 个评价指标, 距离精度图是指目标估计位置的中心点与真实标定位置的中心点之间的距离小于给定阈值的视频帧的百分比。重叠精度图是指视频序列中成功跟踪的帧数占总帧数的百分比。

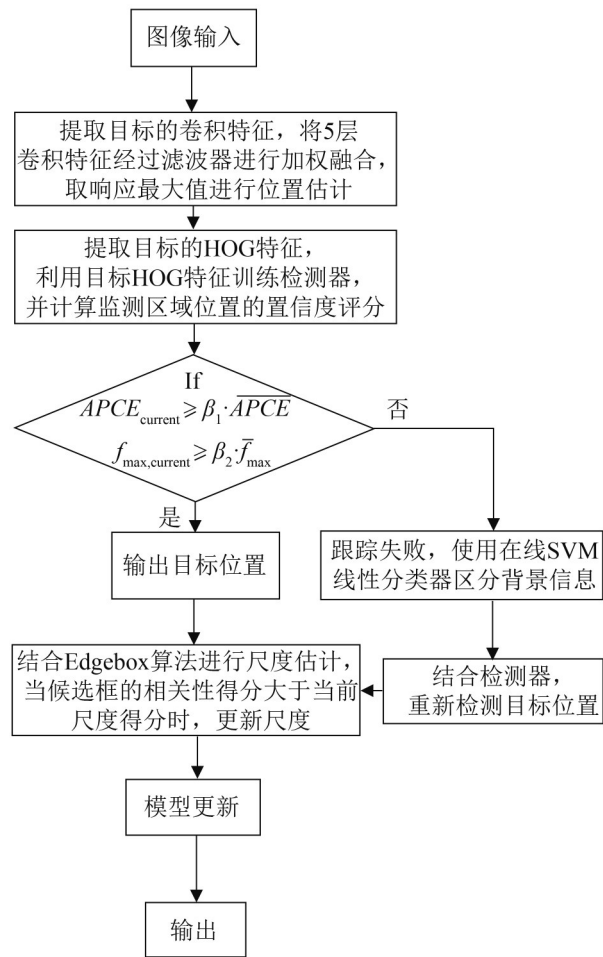


图 4 本文算法流程
Fig. 4 Algorithm flow chart of this article

3.1 定量分析

3.1.1 卷积层的消融实验

本文在 OTB2013 数据集的 51 个视频上进行卷积层的消融实验, 具体如图 5 所示, 其中, c_5, c_4, c_3, c_2, c_1 分别表示为 conv5-4、conv4-4、conv3-4、conv2-2、conv1-1。

3.1.2 开启自适应校正阈值的调试

能否准确判断当前帧因为遮挡等因素导致的跟踪失败是本文算法的关键。针对开启自适应校正参数 β_1 与 β_2 在 OTB2013 数据集的 51 个视频上进行了大量的具体参数调试, 保持其余参数不变, 选取了具有代表性的 9 组调试数值, 具体如图 6 所示。

由图6可知,当 β_1 与 β_2 分别取0.3和0.46时,效果最好。

3.1.3 OTB数据集测试

本文算法与CF2算法、HDT算法^[22]、fDSST算法^[23]、siamFC算法^[24]、samf算法^[25]、DSST算法^[26]和KCF算法进行比较,分别在OTB2013数据集与OTB2015数据集上进行距离精度与重叠精度的对比。本文算法距离精度与重叠精度在OTB2013数据集与OTB2015数据下均排名第一,具体如图7~8所示。

3.1.4 OTB2015数据集下11种属性测试

OTB数据集下的11个挑战属性分别为遮挡(OCC)、形变(DEF)、快速移动(FM)、视野外(OV)、光照变化(IV)、尺度变化(SV)、运动模糊(MB)、低分辨率(LR)、面内旋转(IPR)、背景杂乱(BC)和面外旋转(OPR)。经过测试,在OTB2015数据集的11种挑战属性测试结果对比,除形变(DEF)与低分辨率(LR)外,本文算法性能均排名第一,具体如表1所示。

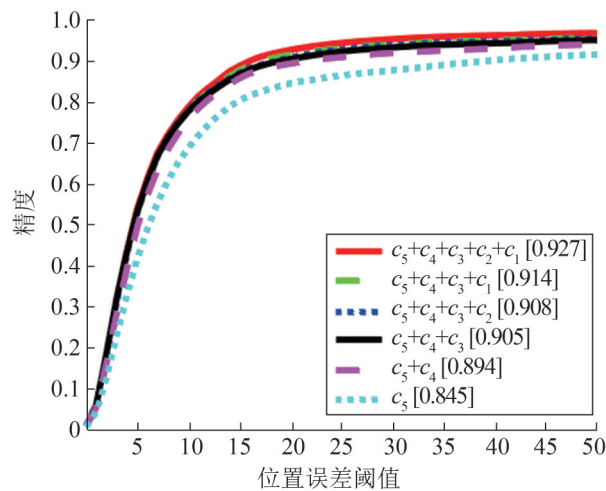


图5 卷积层的消融实验

Fig. 5 Ablation experiments of convolutional layers

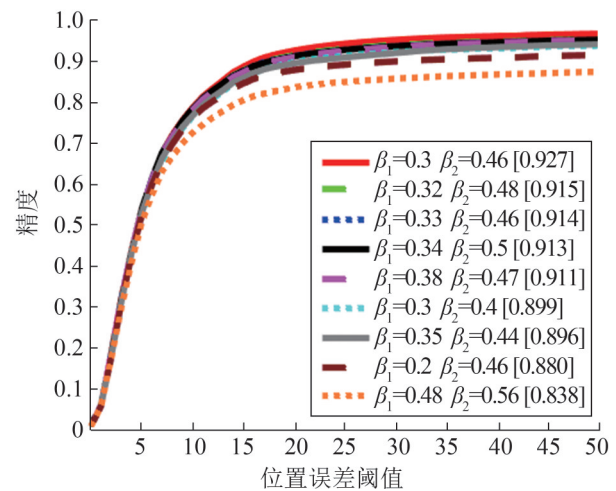
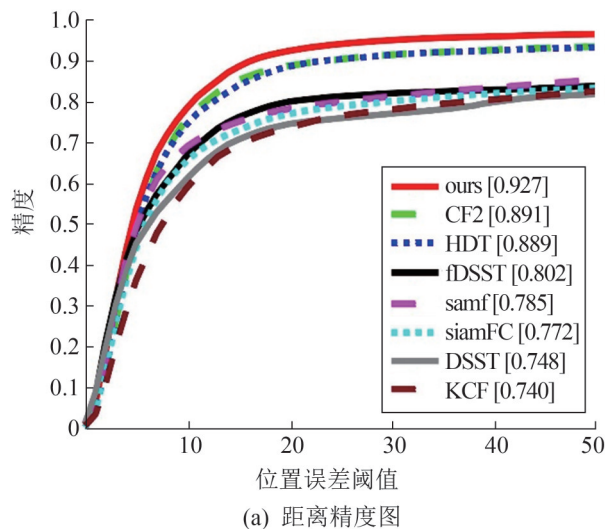
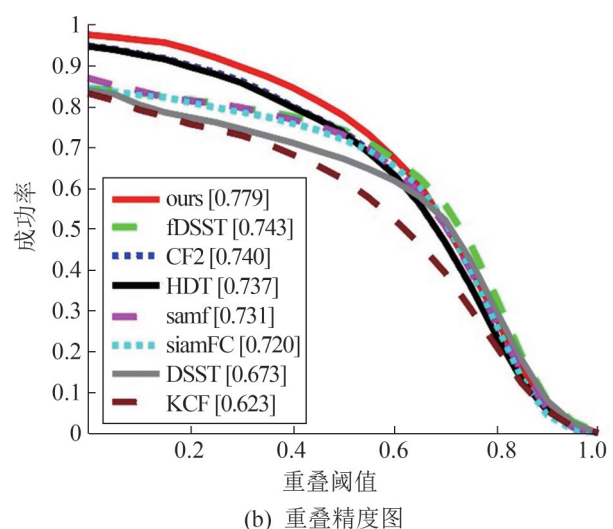


图6 β_1 与 β_2 阈值调试

Fig. 6 Threshold debug for β_1 and β_2



(a) 距离精度图



(b) 重叠精度图

图7 各算法在OTB2013数据集的距离精度与重叠精度对比

Fig. 7 Comparison of accuracy and success rate of each algorithm in OTB2013 dataset

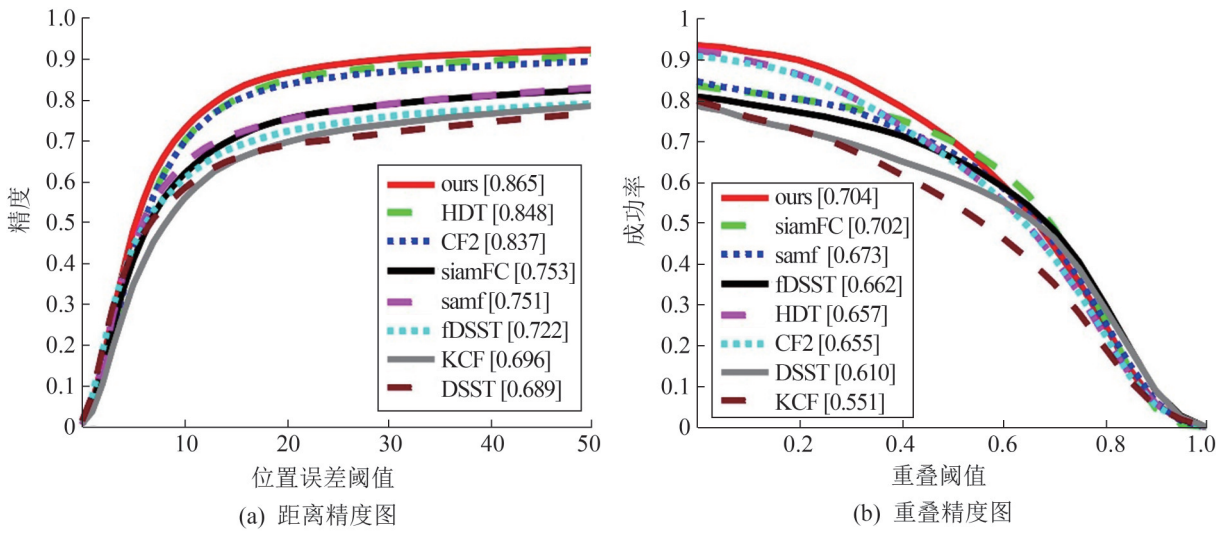


图 8 各算法在 OTB2015 数据集的距离精度与重叠精度对比
Fig. 8 Comparison of accuracy and success rate of each algorithm in OTB2015 dataset

表 1 各算法在 OTB2015 数据集的 11 种挑战属性对比
Table 1 Comparison of 11 challenge attributes of each algorithm in OTB2015 dataset

算法	BC	DEF	FM	IPR	IV	LR	MB	OCC	OV	OPR	SV
本文	0.883	0.810	0.822	0.884	0.884	0.760	0.817	0.787	0.707	0.841	0.816
HDT	0.844	0.821	0.802	0.844	0.820	0.740	0.794	0.774	0.663	0.805	0.811
CF2	0.843	0.791	0.792	0.854	0.817	0.763	0.797	0.767	0.677	0.807	0.802
siamFC	0.705	0.661	0.722	0.742	0.703	0.749	0.676	0.661	0.664	0.733	0.731
samf	0.689	0.686	0.667	0.721	0.714	0.678	0.677	0.728	0.627	0.738	0.710
fDSST	0.779	0.611	0.690	0.727	0.746	0.586	0.700	0.640	0.578	0.666	0.667
KCF	0.713	0.617	0.620	0.701	0.719	0.505	0.618	0.630	0.501	0.677	0.639
DSST	0.715	0.561	0.569	0.708	0.738	0.568	0.608	0.614	0.480	0.658	0.648

3.2 定性分析

为了相对直观地看出本文算法的优劣, 从 OTB2015 数据集中选取了 5 组视频图像进行对比分析, 视频名称分别为 jogging-1、ClifBar、Bird2、coke、skiing, 其中 jogging-1 具有 OCC、DEF 和 OPR 3 种挑战属性, ClifBar 具有 OCC、FM、OV、SV、MB、IPR 和 BC 7 种挑战属性, Bird2 具有 OCC、DEF、FM、IPR 和 OPR 5 种挑战属性, coke 具有 OCC、FM、IV、IPR、BC 和 OPR 6 种挑战属性, skiing 具有 DEF、IV、SV、IPR 和 OPR 5 种挑战属性, 具体如图 9 所示。

从图 9 可以看出, 在 jogging-1 视频序列上, 第 79 帧时目标发生遮挡且目标从遮挡处首次出现

时, 只有本文算法能准确定位目标, 其余算法均发生不同程度的定位失败, 第 80 帧与 306 帧时, fDSST 算法与 KCF 算法依旧无法准确定位目标。

在 ClifBar 视频序列上, 第 82 帧时, KCF 算法与 HDT 算法已经发生了严重的跟踪失败, 且这种跟踪失败现象一直持续到视频最后, 只有本文算法、CF2 算法与 fDSST 算法能一直持续地精确定位目标。

在 Bird2 视频序列上, 第 50 帧时除本文算法能精确定位目标位置外, 其余算法均发生不同程度的偏移, 第 56 帧与 65 帧时, KCF 算法与 fDSST 算法定位发生了严重的漂移, 第 99 帧时, KCF 算法与 fDSST 算法已经完全跟踪失败。

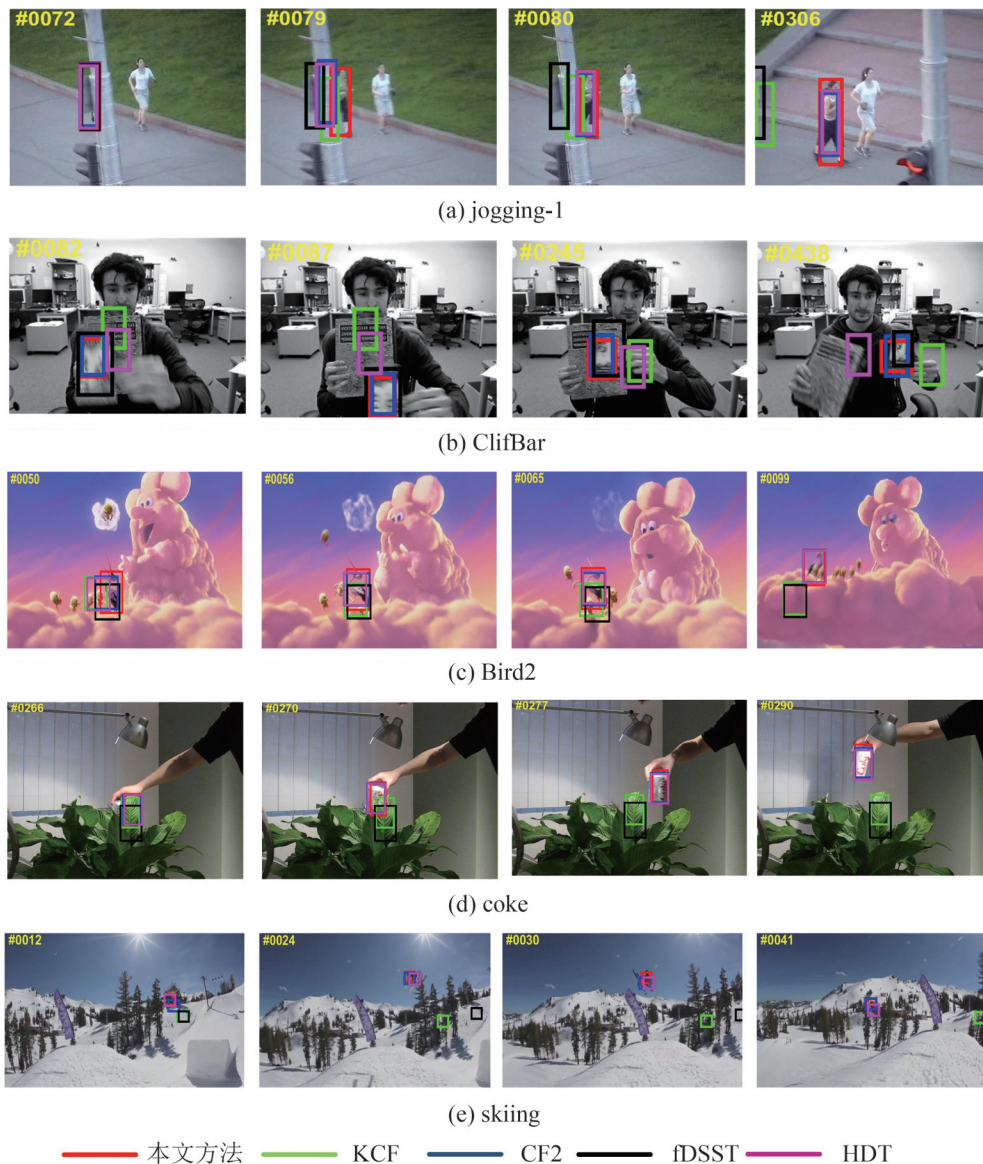


图9 各算法在5组视频序列上的可视化图
Fig. 9 Visualization of each algorithm on 5 groups of video sequences

在 coke 视频序列上, 从第 266 帧目标发生遮挡到 270 帧目标首次出现, 本文算法、CF2 算法与 HDT 算法均能准确捕捉目标, 其余算法定位均发生不同程度的偏移, 在第 277 帧与 290 帧时, KCF 算法与 fDSST 算法已经发生跟踪失败。

在 skiing 视频序列上, 当目标发生严重的旋转与遮挡时, 只有本文算法能全程准确定位目标, 其余算法均发生不同程度的定位漂移或跟踪失败。

4 结论

本文提出了一种基于检测器与定位器融合的自适应校正跟踪算法。定位器运用 VGGNet19 的强大特征表达能力, 融合目标的 5 层深度特征进行位置估计, 检测器通过提取目标的 HOG 特征, 对当前位置的目标设置置信度评分, 自适应校正策略采用当前帧的平均峰值能量和响应最大值与历史均值的比较来综合判断目标在该帧是否出现跟踪失败, 如果跟踪失败则结合检测器重新检测

目标, 否则对目标进行尺度估计。在OTB2013数据集与OTB2015数据集的距离精度和重叠精度上, 相比另外7种跟踪算法, 本文算法均排名第一。但是本文设计的算法在OTB2015数据集的形变(DEF)与低分辨率(LR)2个挑战属性上分别不及HDT与CF2算法, 今后也将努力寻求该问题的解决方案。

参考文献:

- [1] 李琪, 墨瀚林, 王向东, 等. 短道速滑场景下的多目标跟踪与运动仿真[J]. 系统仿真学报, 2021, 33(5): 1039-1050.
Li Qi, Mo Hanlin, Wang Xiangdong, et al. Multiple Object Tracking and Kinematic Simulation for Short Track Speed Skating[J]. Journal of System Simulation, 2021, 33(5): 1039-1050.
- [2] 侯志强, 陈立琳, 余旺盛, 等. 基于双模板Siamese网络的鲁棒视觉跟踪算法[J]. 电子与信息学报, 2019, 41(9): 2247-2255.
Hou Zhiqiang, Chen Lilin, Yu Wangsheng, et al. Robust Visual Tracking Algorithm Based on Siamese Network with Dual Templates[J]. Journal of Electronics & Information Technology, 2019, 41(9): 2247-2255.
- [3] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-Convolutional Siamese Networks for Object Tracking [C]//European Conference On Computer Vision. Cham: Springer, 2016: 850-865.
- [4] Li B, Wu W, Wang Q, et al. Siamrpn++: Evolution of Siamese Visual Tracking with Very Deep Networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4282-4291.
- [5] Liu Y, Li R, Cheng Y, et al. Object Tracking Using Spatio-Temporal Networks for Future Prediction Location [C]//European Conference on Computer Vision. Cham: Springer, 2020: 1-17.
- [6] Bhat G, Danelljan M, Van Gool L, et al. Know Your Surroundings: Exploiting Scene Information for Object Tracking[C]//European Conference on Computer Vision. Cham: Springer, 2020: 205-221.
- [7] Danelljan M, Bhat G, Shahbaz Khan F, et al. Eco: Efficient Convolution Operators for TRacking[C]//IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6638-6646.
- [8] Dai K, Wang D, Lu H, et al. Visual Tracking via Adaptive Spatially-Regularized Correlation Filters[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4670-4679.
- [9] Ma C, Huang J B, Yang X K, et al. Hierarchical Convolutional Features for Visual Tracking[C]//IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 3074-3082.
- [10] Kalal Z, Mikolajczyk K, Matas J. Tracking-Learning-Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence(S0162-8828), 2011, 34(7): 1409-1422.
- [11] Hong Z, Chen Z, Wang C, et al. Multi-store Tracker (muster): A Cognitive Psychology Inspired Approach to Object Tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 749-758.
- [12] Ma C, Yang X, Zhang C, et al. Long-term Correlation Tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 5388-5396.
- [13] Ma C, Huang J B, Yang X, et al. Adaptive Correlation Filters with Long-term and Short-term Memory for Object Tracking[J]. International Journal of Computer Vision(S0920-5691), 2018, 126(8): 771-796.
- [14] Ma C, Huang J, Yang X, et al. Robust Visual Tracking via Hierarchical Convolutional Features[J]. Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence(S0162-8828), 2018, 41(11): 2709-2723.
- [15] 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(S0162-8828), 2014, 37(3): 583-596.
- [16] Lu X, Ma C, Ni B, et al. Adaptive Region Proposal with Channel Regularization for Robust Object Tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology(S10518215), 2019, 31(4): 1268-1282.
- [17] 卫鑫, 武淑红, 王耀力. 基于深度卷积长短期记忆网络的森林火灾烟雾检测模型[J]. 计算机应用, 2019, 39(10): 2883-2887.
Wei Xin, Wu Shuhong, Wang Yaoli. Forest Fire Smoke Detection Model Based on Deep Convolution Long Short-term Memory Network[J]. Journal of Computer Applications, 2019, 39(10): 2883-2887.
- [18] 徐亮, 张江, 张晶, 等. 基于VGG网络的鲁棒目标跟踪算法[J]. 计算机工程与科学, 2020, 42(8): 1406-1413.
Xu Liang, Zhang Jiang, Zhang Jing, et al. A Robust Target Tracking Algorithm Based on VGG Network[J]. Computer Engineering & Science, 2020, 42(8): 1406-1413.
- [19] 魏永强, 杨小军. 深度特征的核相关滤波视觉跟踪[J]. 计算机工程与应用, 2020, 56(15): 147-152.
Wei Yongqiang, Yang Xiaojun. Kernel Correlation Filtering Visual Tracking of Deep Feature[J]. Computer Engineering and Applications, 2020, 56(15): 147-152.

- [20] Wang M, Liu Y, Huang Z. Large Margin Object Tracking with Circulant Feature Maps[C]//IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 4021-4029.
- [21] Zitnick C L, Dollar P. Edge Boxes: Locating Object Proposals from Edges[C]//European Conference on Computer Vision. Cham: Springer, 2014: 391-405.
- [22] Qi Y, Zhang S, Qin L, et al. Hedged Deep Tracking[C]//IEEE Conference On Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 4303-4311.
- [23] Danelljan M, Häger G, Khan F S, et al. Discriminative Scale Space Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence(S0162-8828), 2016, 39(8): 1561-1575.
- [24] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese Networks for Object Tracking[C]//European Conference on Computer Vision. Cham: Springer, 2016: 850-865.
- [25] Li Y, Zhu J. A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration[C]//European Conference on Computer Vision. Cham: Springer, 2014: 254-265.
- [26] Danelljan M, Häger G, Khan F, et al. Accurate Scale Estimation for Robust Visual Tracking[C]//British Machine Vision Conference, September 1-5, 2014. Nottingham: Bmva Press, 2014.