

基于空间加权对数似然比相关滤波与Deep Snake的目标轮廓跟踪

李 豪,袁广林,秦晓燕,琚长瑞,朱 虹
(中国人民解放军陆军炮兵防空兵学院信息工程系,安徽合肥 230031)

摘 要: 近年来,目标跟踪中目标的状态表示已由粗糙的矩形框转化为精细的目标掩膜.然而,现有方法利用区域分割得到目标掩膜,速度慢并且掩膜精度受限于目标跟踪框.针对以上问题,本文提出基于空间加权对数似然比相关滤波与Deep Snake的目标轮廓跟踪方法.该方法包括三个阶段:在第一阶段,利用提出的空间加权对数似然比相关滤波器估计目标的初始矩形框;在第二阶段,通过Deep Snake将初始矩形框变形为目标轮廓;在第三阶段,根据目标轮廓拟合出跟踪结果.对提出的方法在OTB(Object Tracking Benchmark)-2015和VOT(Visual Object Tracking)-2018数据集上进行了实验验证,结果表明:与现有先进的目标跟踪方法相比,本文提出的跟踪方法具有较优的性能.

关键词: 目标跟踪;深度主动轮廓;相关滤波;空间加权;对数似然比;视频目标分割

基金项目: 安徽省自然科学基金(No.2008085QF325)

中图分类号: TP391.41

文献标识码: A

文章编号: 0372-2112(2023)01-0105-12

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210676

Object Contour Tracking Based on Correlation Filters with Spatially-Weighted Logarithm Likelihood Ratio and Deep Snake

LI Hao, YUAN Guang-lin, QIN Xiao-yan, JU Chang-rui, ZHU Hong

(Department of Information Engineering, Army Academy of Artillery and Air Defense of PLA, Hefei, Anhui 230031, China)

Abstract: Recently, the state representation of the target in object tracking has been transformed from the coarse bounding-box to fine-grained segmentation map. However, the existing methods use pixel-based segmentation to obtain object mask, which is slow and the accuracy of mask is limited by the object bounding box of tracking. To solve the above problems, we propose an object contour tracking method based on correlation filters with spatially-weighted logarithm likelihood ratio and deep snake. The method consists of three stages: at the first stage, the initial bounding box of the object is estimated by the proposed correlation filters with spatially-weighted logarithm likelihood ratio; at the second stage, the initial bounding box is deformed into the object contour via deep snake; at the third stage, the tracking results are fitted with the object contour. Experimental results on OTB (Object Tracking Benchmark)-2015 and VOT (Visual Object Tracking)-2018 datasets show that the proposed method is superior to the state-of-the-art approaches.

Key words: object tracking; deep active contour; correlation filter; spatially-weighted; logarithm likelihood ratio; video object segmentation

Foundation Item(s): Natural Science Foundation of Anhui Province (No.2008085QF325)

1 引言

目标跟踪是计算机视觉领域的一个重要研究分支,其目的是利用视频数据估计目标状态.目标跟踪在智能监控、无人驾驶和人机交互等民用领域,以及视觉制导和防空预警等军事领域具有重要的应用价值.近年来,得益于人工智能技术的不断提高和发展,目标跟

踪有了较大进步,已经提出了诸多方法^[1-31].但是,目标跟踪是一项具有挑战性的任务,仍然面临着“复杂背景、目标变化和快速性要求”等难题.

目前,主流的目标跟踪方法有相关滤波跟踪^[1-16]和孪生网络跟踪^[17-23]两类.这些方法使用矩形框表示目标,利用多尺度搜索^[3]、锚框回归^[19-21]或无锚框回归^[22]

等方法估计目标状态.但是,使用矩形框表示目标有其自身的局限性:当目标发生形变或旋转时,矩形框表示目标会产生误差,导致跟踪精度降低.与矩形框相比,掩膜为目标提供了准确的形状描述.受视频目标分割^[24]的启发,目标跟踪领域的研究者将分割集成到跟踪中,提出了基于分割的目标跟踪方法^[25-29].SiamMask^[25]在SiamRPN(Siamese Region Proposal Network)^[19]的基础上增加掩膜分支实现目标分割,但它独立处理目标跟踪和分割,未能相互促进和优化.D3S(a Discriminative Single Shot Segmentation tracker)^[28]提出一种单阶段判别式分割跟踪器,通过优化网络将几何不变模型和几何约束欧几里得模型的输出相结合,得到目标掩膜并拟合出矩形框.AFOD(Adaptive Focused Discriminative Segmentation Tracker)^[29]利用交并比(Intersection over Union, IoU)掩膜损失和目标边界检测提升了D3S的掩膜分割精度.上述方法能同时得到目标矩形框和掩膜,提升了跟踪精度.但是,这些方法采用基于区域的目标分割方法,一方面,需要将目标区域特征图归一化为同一尺度,然后对目标区域内每一个像素进行逐点分类;另一方面,需要执行mask上采样等后处理操作.这两个方面导致其存

在对跟踪矩形框敏感和速度慢等问题.

为了克服这些问题,本文提出基于空间加权对数似然比相关滤波与Deep Snake^[32]的目标轮廓跟踪方法,称之为OCT-CFDS(Object Contour Tracking based on Correlation Filters and Deep Snake),其总体框架如图1所示.该方法的工作流程是:首先,利用提出的空间加权对数似然比相关滤波器估计目标的初始矩形框;其次,通过Deep Snake将初始矩形框变形为目标轮廓;最后,根据目标轮廓拟合出跟踪结果.

本文的主要贡献和创新点如下:

(1)利用Deep Snake提出一种基于轮廓分割的目标跟踪方法,提高了分割速度并增强了对目标变化的鲁棒性.

(2)提出空间加权对数似然比相关滤波器,自适应融合多特征估计目标的初始矩形框,提升了对复杂背景的鲁棒性.

(3)使用OTB-2015^[33]和VOT-2018^[34]数据集,对提出的方法进行了实验验证,结果表明:本文提出的目标跟踪方法达到了预期的效果,与现有先进的目标跟踪方法相比具有较高的跟踪性能.

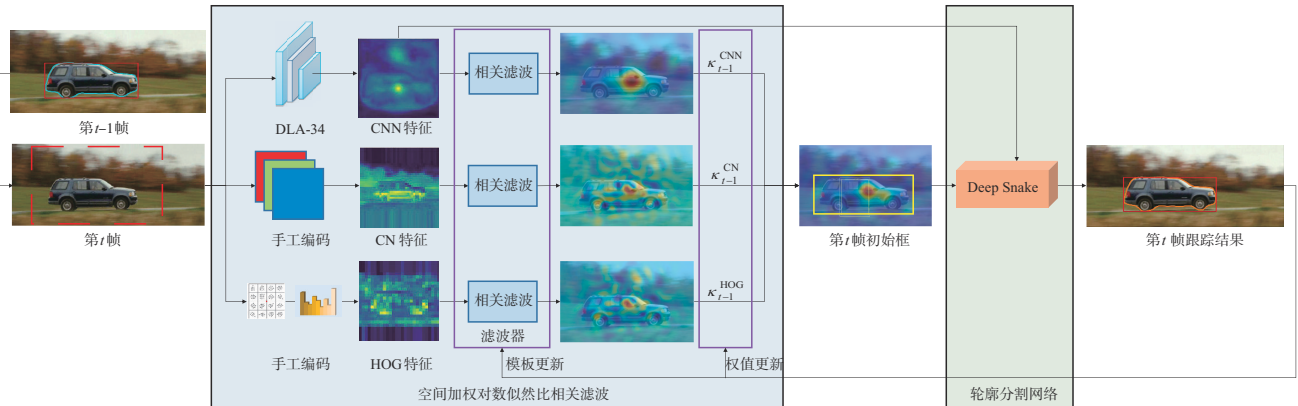


图1 本文方法总体框架和工作流程

2 相关工作

相关滤波跟踪由于速度快而受到了广泛关注.2010年,Bolme等人^[1]首次将相关滤波应用到目标跟踪中,提出基于单通道灰度特征的相关滤波跟踪方法MOSSE(Minimum Output Sum of Squared Error).为了提升目标跟踪的鲁棒性,2014年,Henriques等人^[2]提出基于多通道方向梯度直方图(Histogram of Oriented Gradient, HOG)特征的核相关滤波跟踪方法KCF(Kernelized Correlation Filter).早期的相关滤波跟踪虽然速度较快,但是存在尺度观测不准和边界效应等问题,针对这两个问题研究者提出了一些改进方法^[3].随着深度学习在目标跟踪中应用的深入,融合手工与深度特征

的相关滤波目标跟踪成为了研究热点^[4-6].2016年,Danelljan等人^[4]提出一种连续卷积相关滤波跟踪方法C-COT(Continuous Convolution Operator Tracker),融合手工特征与卷积神经网络(Convolutional Neural Network, CNN)特征跟踪目标.2017年,Danelljan等人^[5]提出有效卷积运算跟踪方法ECO(Efficient Convolution Operators),通过因式分解卷积操作对C-COT进行改进,在目标跟踪速度和精度上寻求平衡.为了实现浅层和深层CNN特征自适应融合跟踪目标,2018年,Bhat等人^[6]提出相关滤波目标跟踪方法UPDT(Unveiling the Power of Deep Tracking),该方法利用二次规划计算自适应融合权重,虽然提高了目标跟踪的鲁棒性和精度,

但是速度较慢。

孪生网络跟踪是深度跟踪的典型代表。2016年, Tao 等人^[17]提出的 SINT (Siamese Instance search Tracker)首次将孪生网络应用到目标跟踪中,把目标跟踪建模为深度特征模板匹配问题。2016年, Bertinetto 等人^[18]提出 SiamFC (Siamese Fully-Convolutional), 将目标跟踪建模为全卷积深度特征相关问题,不但精度高而且速度快。早期的孪生网络跟踪利用多尺度搜索估计目标的尺度,速度较慢。为此,2018年, Li 等人^[19]提出的 SiamRPN 将目标检测中的 RPN (Region Proposal Network) 模块应用到孪生网络跟踪中,把目标跟踪建模为锚框的回归和分类问题,提升了跟踪速度。2018年, Zhu 等人^[20]提出的 DaSiamRPN (Distractor-aware Siamese Region Proposal Networks) 利用数据增广和在线干扰物感知模块提高了对相似目标干扰的鲁棒性。2019年, Li 等人^[21]利用空间感知采样策略、多层融合和 Depthwise 相关提出孪生网络目标跟踪方法 SiamRPN++, 提升了跟踪性能。基于锚框的孪生网络跟踪^[19-21]难以回归与目标重叠较小的锚框,为此,2020年, Zhang 等人^[22]设计了无锚框的回归网络和特征对齐目标感知分类网络,在此基础上提出了目标跟踪方法 Ocean (Object-aware anchor-free networks), 提高了跟踪精度。

视频目标分割和视觉目标跟踪相互融合促进成为了研究热点。2018年, Cheng 等人^[24]提出一种基于部件跟踪的视频目标分割方法,提高了视频目标分割对目标形变和遮挡的鲁棒性。2019年, Wang 等人^[25]提出的 SiamMask 利用多任务学习实现了快速目标跟踪与分割,架起了视频目标分割和目标跟踪的桥梁。2020年, Chen 等人^[26]提出的状态感知跟踪器 SAT (State-Aware Tracker) 把视频目标分割建模为一个连续状态估计过程,利用双路反馈估计机制感知目标状态和适应目标变化,取得了较好的分割效果。2020年, Sun^[27]等人提出基于强化学习的模板匹配和更新机制实现视频目标分割,提高了分割速度。2020年, Chen 等人^[28]提出的 D3S 利用优化网络和基于目标掩膜的矩形框拟合方法,实现一种单阶段判别式分割跟踪器,缩短了视觉目标跟踪和视频目标分割的差距。2020年, Chen 等人^[29]提出的 AFOD 利用 IOU 掩膜损失和目标边界检测对 D3S 进行了改进,实现了精确鲁棒的目标跟踪和分割。

3 空间加权对数似然比相关滤波

由文献[32]可知,目标初始框的准确性与轮廓分割的精度成正相关关系,因此目标初始框的估计至关重要。为此,本文提出空间加权对数似然比相关滤波器估计目标初始框。

3.1 空间加权对数似然比

使用高质量的特征是实现鲁棒、准确跟踪的基础, KCF^[2]利用基于多通道的 HOG 特征捕获目标边缘或局部结构信息,但对目标形变鲁棒性较差。CN (Color Names)^[7]利用基于多通道的颜色特征取得了较快的计算速度,并且对目标形变和尺度变化具有鲁棒性,但是易受相似颜色目标的干扰。DeepSRDCF (Deep Spatially Regularized Discriminative Correlation Filters)^[8]基于相关滤波利用强语义 CNN 特征跟踪目标,对复杂背景比较鲁棒,但是计算复杂度较高。Staple (Sum of Template And Pixel-wise LEarners)^[9]根据 HOG 与 CN 特征的互补性实现目标跟踪,以相关滤波为框架利用得分图静态加权方法融合 HOG 特征和颜色直方图特征跟踪目标,有效提升了跟踪精度。C-COT^[4]利用累加得分图方法融合 CN、HOG 和 CNN 特征跟踪目标,并使用共轭梯度下降法求解最优目标位置。上述方法均采用静态加权特征融合方法,不能适应目标的变化。为此, UPDT^[6]利用基于二次规划的自适应特征加权融合相关滤波实现目标跟踪,提高了跟踪的鲁棒性和精度,但是二次规划导致其速度较慢。

由相关滤波跟踪的原理^[6]可知:特征的质量可以利用相关滤波得分图度量。图 2 给出了 OTB-2015 数据集中三个图像序列中一帧图像及其 CNN、CN 和 HOG 特征的相关滤波得分图。由图 2(a)~图 2(d)可以看出:对于 DragonBaby 序列, CNN 和 HOG 特征得分图的最大值在目标区域, CN 特征得分图的最大值在背景区域,即 CN 特征具有负面作用。由图 2(e)~图 2(h)可以看出:对于 Bolt2 序列, CN 特征得分图的最大值在目标区域, CNN 和 HOG 特征得分图的最大值在背景区域,因此, CNN 和 HOG 特征会导致目标跟踪失败。由图 2(i)~图 2(l)可以看出:对于 Human2 序列,虽然 CNN、CN 和 HOG 特征的最大得分均在目标区域,但是 CNN 和 CN 特征得分图的最大值偏移了目标中心,即这两个特征会降低目标跟踪精度。综合图 2 的分析可知:并非所有特征均对跟踪产生正向作用,在某些场景中个别特征会对跟踪产生负作用,简单的特征融合不一定能提高目标跟踪的鲁棒性,而且可能会降低目标跟踪的精度。

由上述分析可知:特征的选择对目标跟踪的精度和速度有着重要的影响,在目标跟踪中 CN、HOG 和 CNN 特征可以互补,融合这些特征可以提升相关滤波目标跟踪的鲁棒性和精度; Staple^[9]和 C-COT^[4]采用得分图静态加权融合特征跟踪目标,不能适应目标的变化,虽然 UPDT^[6]实现了自适应特征融合目标跟踪,但是它利用约束二次规划求解融合权重和目标位置,导致其速度较慢。由文献[6]可知:特征的质量取决于其准确性和鲁棒性。特征的准确性与其最大得分位置和



图2 相关滤波得分图

当前时刻目标的中心位置相关,最大得分位置与目标中心位置的距离越近准确性越好,即特征的准确性与最大得分位置和目标中心位置之间的距离反相关.特征的鲁棒性与目标区域最大得分以及背景区域最大得分相关,目标区域最大得分与背景区域最大得分的比值越大其精度越高,如果背景区域的最大得分大于目标区域的最大得分则会导致跟踪失败.基于上述分析,本文提出基于空间加权对数似然比的特征融合权值计算方法,如式(1)所示:

$$\kappa_t^f = \exp\left(-\|p_t^f - c_t^o\|_2^2\right) \max\left\{0, \log\left(\frac{\max\{s_t^f(\mathbf{y}_o), \delta\}}{\max\{s_t^f(\mathbf{y}_b), \delta\}}\right)\right\} \quad (1)$$

其中 $f \in \{\text{CN}, \text{HOG}, \text{CNN}\}$, p_t^f 是 t 时刻特征 f 相关滤波得分图的最大得分位置, c_t^o 为 t 时刻目标的中心位置, $s_t^f(\mathbf{y}_o)$ 为 t 时刻特征 f 在目标区域 \mathbf{y}_o 处的相关滤波得分值, $s_t^f(\mathbf{y}_b)$ 为 t 时刻特征 f 在背景区域 \mathbf{y}_b 处的相关滤波得分值, δ 是为防止分子分母出现0的极小变量(本文设置为0.001), 实验中 κ_t^f 在特征 f 中做归一化处理, 保证 κ_t^f 在0和1之间.

式(1)中指数函数部分是空间加权函数, 表征了特征精度; 另一部分是对数似然比, 度量了特征鲁棒性. 因此, 通过式(1)可以选择用于跟踪的最佳特征、排除不相关或者冗余的特征, 从而提高跟踪的性能.

3.2 空间加权对数似然比相关滤波方法

文献[3]提出多通道特征相关滤波目标跟踪方法, 其模型如下:

$$\min_{\{\mathbf{w}_k^f\}_{k=1}^{K_f}} \left(\left\| \sum_{k=1}^{K_f} \mathbf{w}_k^f \circ \mathbf{x}_k^f - \mathbf{g} \right\|_2^2 + \lambda \sum_{k=1}^{K_f} \|\mathbf{w}_k^f\|_2^2 \right) \quad (2)$$

其中 $f \in \{\text{CN}, \text{HOG}, \text{CNN}\}$, K_f 和 $\{\mathbf{w}_k^f\}_{k=1}^{K_f}$ 分别是特征 f 的通道数和滤波器, $\mathbf{x}_k^f \in \mathbb{R}^D$ 和 $\mathbf{w}_k^f \in \mathbb{R}^D$ 分别是特征 f 的第 k 个通道特征和滤波器, $\mathbf{g} \in \mathbb{R}^D$ 是由高斯函数生成的期望得分, λ 是正则化参数(本文 λ 取0.01), “ \circ ”表示相关操作. 由文献[3]可知, 式(2)在傅里叶变换域的解如下:

$$\mathbf{W}_k^f = \frac{\bar{\mathbf{G}} \mathbf{X}_k^f}{\sum_{l=1}^{K_f} \bar{\mathbf{X}}_l^f \mathbf{X}_l^f + \lambda} \quad (3)$$

其中, \mathbf{W}_k^f 、 \mathbf{G} 、 \mathbf{X}_k^f 和 \mathbf{X}_l^f 分别是 \mathbf{w}_k^f 、 \mathbf{g} 、 \mathbf{x}_k^f 和 \mathbf{x}_l^f 的傅里叶变

换, \bar{G} 和 \bar{X}_t^f 分别是 G 和 X_t^f 的复共轭.

将式(1)与式(3)相结合, 本文提出空间加权对数似然比相关滤波方法. 假设 t 时刻目标框 $b_t = (x_t, y_t, w_t, h_t)$, 其中 (x_t, y_t) 是目标中心位置, w_t 和 h_t 分别是目标的宽和高, \bar{W}_k^f 是 W_k^f 的复共轭, 则该方法描述见算法 1.

算法 1 空间加权对数似然比相关滤波方法

输入: $t-1$ 时刻目标框 $b_{t-1} = (x_{t-1}, y_{t-1}, w_{t-1}, h_{t-1})$ 、相关滤波器

$\{W_{k,t-1}^f\}_{k=1}^{K_f}$ 、特征融合权重 κ_{t-1}^f 和 t 时刻目标图像 I_t .

输出: t 时刻目标初始框 $b_t^{\text{init}} = (x_t^{\text{init}}, y_t^{\text{init}}, w_t^{\text{init}}, h_t^{\text{init}})$.

1: 计算搜索区域: $r_t = (x_{t-1}, y_{t-1}, 4w_{t-1}, 4h_{t-1})$;

2: 提取搜索区域 r_t 的 CNN 特征并计算其傅里叶变换 $\{X_{k,t}^{\text{CNN}}\}_{k=1}^{K_{\text{CNN}}}$ 、

CN 特征并计算其傅里叶变换 $\{X_{k,t}^{\text{CN}}\}_{k=1}^{K_{\text{CN}}}$ 和 HOG 特征并计算其傅里

叶变换 $\{X_{k,t}^{\text{HOG}}\}_{k=1}^{K_{\text{HOG}}}$;

3: 计算 CNN 特征相关滤波得分图: $\hat{s}_t^{\text{CNN}} = F^{-1} \left(\sum_{k=1}^{K_{\text{CNN}}} \bar{W}_{k,t-1}^{\text{CNN}} X_{k,t}^{\text{CNN}} \right)$;

4: 计算 CN 特征相关滤波得分图: $\hat{s}_t^{\text{CN}} = F^{-1} \left(\sum_{k=1}^{K_{\text{CN}}} \bar{W}_{k,t-1}^{\text{CN}} X_{k,t}^{\text{CN}} \right)$;

5: 计算 HOG 特征相关滤波得分图: $\hat{s}_t^{\text{HOG}} = F^{-1} \left(\sum_{k=1}^{K_{\text{HOG}}} \bar{W}_{k,t-1}^{\text{HOG}} X_{k,t}^{\text{HOG}} \right)$;

6: 融合相关滤波得分图: $\hat{s}_t = \kappa_{t-1}^{\text{CN}} \hat{s}_t^{\text{CN}} + \kappa_{t-1}^{\text{HOG}} \hat{s}_t^{\text{HOG}} + \kappa_{t-1}^{\text{CNN}} \hat{s}_t^{\text{CNN}}$;

7: 计算 t 时刻目标中心位置: $(x_t^{\text{init}}, y_t^{\text{init}}) = \arg \max_{x,y} (\hat{s}_t(x,y))$;

8: 估计 t 时刻目标尺度: $w_t^{\text{init}} = w_{t-1}, h_t^{\text{init}} = h_{t-1}$;

9: 输出 t 时刻目标初始框: $b_t^{\text{init}} = (x_t^{\text{init}}, y_t^{\text{init}}, w_t^{\text{init}}, h_t^{\text{init}})$.

4 Deep Snake 轮廓分割网络

4.1 轮廓表示与 1D 循环卷积

目标轮廓是图像上一组有序点 $V = \{y_i\}_{i=1}^N$ 的集合, 其中 $y_i = (x_i, y_i)$ 是点的坐标, N 是点的个数. 点 y_i 的特征表示为 $\hat{f}_i = [F(y_i); y_i']$, 其中 $F(y_i)$ 是点 y_i 在 F 上的双线性插值, F 是原始输入图像通过 CNN 得到的特征图, $y_i' = (x_i - x_{\min}, y_i - y_{\min})$ 是 y_i 的平移不变坐标, 其中 $x_{\min} = \min \{x_i\}_{i=1}^N, y_{\min} = \min \{y_i\}_{i=1}^N$. 如图 3 所示, Deep Snake^[34] 将轮廓上点的特征向量看作周期信号 $(\hat{f}_N)_i \triangleq \hat{f}_{i \bmod N}$, 并定义其 1D 循环卷积如下:

$$(\hat{f}_N * \hat{k})_i = \sum_{j=-r}^r (\hat{f}_N)_{i+j} \hat{k}_j \quad (4)$$

其中 \hat{k} 是 1D 卷积核, “*” 表示 1D 卷积操作. 图 3 中蓝色点表示轮廓上点的特征, 黄色点表示核函数, 绿色点表示卷积结果. 高亮绿色点表示黄色卷积核与对应的轮廓点进行 1D 循环卷积得到的特征.

4.2 网络结构与轮廓分割

Deep Snake 由骨干网络、融合模块和预测模块三部分组成. 骨干网络由 8 个 “CirConv-Bn-ReLU” 层组成, 所

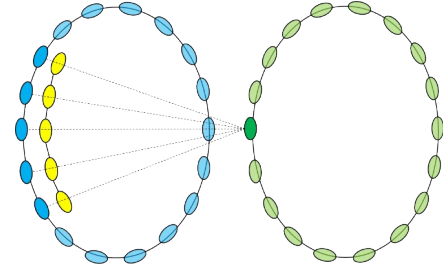


图 3 1D 循环卷积示意图

有层均使用残差连接, 其中 “CirConv” 表示 1D 循环卷积, “Bn” 表示批量归一化, “ReLU” 表示激活函数. 融合模块首先通过 1×1 卷积层融合骨干网络中所有层的特征, 其次进行最大池化, 最后将融合特征与骨干网络中每层特征进行级联得到最终特征. 预测模块通过三个 1×1 卷积层输出轮廓点的偏移值.

Deep Snake 轮廓分割网络将矩形框分割为目标轮廓的过程是: 首先, 连接矩形框各边的中点得到菱形轮廓; 其次, 菱形轮廓通过 Deep Snake 得到极值点, 进而得到八边形轮廓; 最后, 八边形轮廓通过 Deep Snake 迭代变形为目标轮廓 (本文迭代次数为 3).

4.3 损失函数

Deep Snake 采用极值点预测损失函数和轮廓变形损失函数对网络进行监督学习. 极值点预测损失函数为 $L_{\text{ex}} = \frac{1}{4} \sum_{i=1}^4 \text{Smooth}_{L_1}(\tilde{y}_i^{\text{ex}} - y_i^{\text{ex}})$, 其中 \tilde{y}_i^{ex} 是极值点的预测坐标, y_i^{ex} 是极值点的标注坐标. 轮廓变形损失函数为 $L_{\text{iter}} = \frac{1}{N} \sum_{i=1}^N \text{Smooth}_{L_1}(\tilde{y}_i - y_i^{\text{st}})$, 其中 \tilde{y}_i 是轮廓点的预测坐标, y_i^{st} 是轮廓点的标注坐标, N 是点的个数 (本文实验中 N 取 128).

5 跟踪方法

跟踪方法见算法 2, 其中 N_s 和 η 分别是模板和滤波器更新频率参数和权重参数, 本文实验中 N_s 取 5, η 取 0.025. 需要说明的是: 对于 VOT-2018 数据集, 方法 2 中步骤 6 不仅得到最小外接矩形框 b_t , 而且根据旋转矩形框拟合方法^[28] 得到并输出旋转矩形框 b_t^r .

6 实验及分析

6.1 实验环境、数据集及评价指标

在 SYS-7048GR-TR 台式机 (CPU 型号为 Intel Xeon (R) ES-2630v4@2.20 GHz×20, 内存为 64 GB, GPU 为 RTX2080Ti 11 G) 上实现了本文方法, 软件环境是: Ubuntu 18.04、Python 3.7、torch 1.1.0、cuda10.0 和 cudnn7.5. 利用 OTB-2015^[33] 和 VOT-2018^[34] 数据集对提出的目标跟踪方法进行实验验证. OTB-2015 数据集包含 100 个视频序列, 具有光照变化 (Illumination Varia-

算法 2 基于空间加权对数似然比相关滤波与 Deep Snake 的目标轮廓跟踪方法

输入: 第一帧目标框 $\mathbf{b}_1=(x_1, y_1, w_1, h_1)$ 、目标图像序列 $\mathbf{I}=\{\mathbf{I}_t\}_{t=1}^T$ 和

特征 $f \in \{\text{CN}, \text{HOG}, \text{CNN}\}$

输出: t 时刻目标框 $\mathbf{b}_t=(x_t, y_t, w_t, h_t)$ 和目标轮廓 $V_t = \{y_i^t\}_{i=1}^N$

1: 根据第一帧目标框 \mathbf{b}_1 初始化目标特征模板 $\{\mathbf{X}_{k,1}^f\}_{k=1}^{K_f}$;

2: 利用目标特征模板 $\{\mathbf{X}_{k,1}^f\}_{k=1}^{K_f}$ 和式(3)初始化相关滤波器 $\{\mathbf{W}_{k,1}^f\}_{k=1}^{K_f}$;

3: 初始化特征融合权重 $\kappa_1^f = 1/N$ (N 为特征个数);

FOR $t = 2:T$

4: 调用算法 1 得到 t 时刻目标初始框 $\mathbf{b}_t^{\text{init}}=(x_t^{\text{init}}, y_t^{\text{init}}, w_t^{\text{init}}, h_t^{\text{init}})$;

5: 将 $\mathbf{b}_t^{\text{init}}$ 输入 Deep Snake 轮廓分割网络, 得到 t 时刻目标轮廓 $V_t =$

$\{y_i^t\}_{i=1}^N$;

6: 根据目标轮廓 V_t 拟合得到 t 时刻的最小外接矩形框 $\mathbf{b}_t =$

(x_t, y_t, w_t, h_t) ;

IF $t \text{ MOD } N_s == 0$

7: 提取目标框 \mathbf{b}_t 的特征, 并计算其傅里叶变换 $\{\mathbf{X}_{k,t}^f\}_{k=1}^{K_f}$;

8: 更新目标特征模板: $\mathbf{X}_{k,t}^f = (1-\eta)\mathbf{X}_{k,t-1}^f + \eta\mathbf{X}_{k,t}^f, k=1, \dots, K_f$;

9: 利用目标特征模板 $\{\mathbf{X}_{k,t}^f\}_{k=1}^{K_f}$ 和式(3)训练相关滤波器 $\{\mathbf{W}_{k,t}^f\}_{k=1}^{K_f}$;

10: 更新相关滤波器: $\mathbf{W}_{k,t}^f = (1-\eta)\mathbf{W}_{k,t-1}^f + \eta\mathbf{W}_{k,t}^f, k=1, \dots, K_f$;

11: 利用式(1)计算特征融合权重 κ_t^f , 并对其更新: $\kappa_t^f = (1-\eta)\kappa_{t-1}^f + \eta\kappa_t^f$;

END IF

12: 输出 t 时刻目标框 $\mathbf{b}_t=(x_t, y_t, w_t, h_t)$ 和目标轮廓 $V_t = \{y_i^t\}_{i=1}^N$.

END FOR

tion, IV)、尺度变化 (Scale Variation, SV)、遮挡 (Occlusion, OCC)、形变 (Deformation, DEF)、运动模糊 (Motion Blur, MB)、快速运动 (Fast Motion, FM)、平面内旋转 (In-Plane Rotation, IPR)、平面外旋转 (Out-of-Plane Rotation, OPR)、目标出界 (Out-of-View, OV)、背景混杂 (Background Clutters, BC) 和低分辨率 (Low Resolution, LR) 11 个挑战属性. 该数据集利用 OPE (One-Pass Evaluation) 方式计算 Precision 和 Success 指标评价跟踪器的精度和鲁棒性. 由于 OTB-2015 数据集的标注矩形框为正矩形框, 因此, 本文方法利用目标轮廓的最小外接矩形框计算 Precision 和 Success. VOT-2018 数据集包含 60 个视频序列, 其中部分视频序列与 OTB-2015 数据集重合, 但是总体难度比 OTB-2015 数据集高. 该数据集的评价指标包含 EAO (Expected Average Overlap)、

Accuracy 和 Robustness, 其中 EAO 和 Accuracy 用于评价跟踪器的精度, 跟踪器的精度与这两个指标正相关, Robustness 用于评价跟踪器的鲁棒性, 跟踪器的鲁棒性与该指标反相关. 目标跟踪速度用 FPS (Frames Per Second) 指标评价. 由于 VOT-2018 数据集的标注矩形框为旋转矩形框, 因此, 本文方法利用轮廓拟合得到旋转矩形框计算 EAO、Accuracy 和 Robustness. 对于掩膜分割的误差, 本文利用标注掩膜和预测掩膜的平均交并比 (mIOU) 进行度量.

6.2 实现细节

算法 1 中使用的 CNN 特征是 DLA-34 网络^[35] 的最后一层特征, 其分辨率为原图的 1/4 倍, 通道数为 64. CN 特征^[7] 的通道是 2, HOG 特征^[2] 的通道数是 31. DLA-34 网络利用 COCO 数据集训练, Deep Snake 轮廓分割网络利用 Sbd 数据集训练. 在训练阶段, 输入图像分辨率归一化为 512×512 . 在测试阶段, 输入图像分辨率仿射变换为 32 的倍数.

6.3 实验结果与分析

6.3.1 OTB-2015 数据集

表 1 给出了本文方法 OCT-CFDS、DaSiamRPN^[20]、SiamRPN++^[21]、ECO^[5]、MDNet (Multi-Domain Network)^[30]、C-COT^[4]、MCPF (Multi-task Correlation Particle Filter)^[10]、TADT (Target-Aware Deep Tracking)^[31]、DeepSRDCF^[8] 和 SRDCF (Spatially Regularized Discriminative Correlation Filters)^[11] 共 10 种目标跟踪方法在 OTB-2015 数据集上的 Precision, Success 与 FPS 指标, 图 4 为 10 种目标跟踪方法在 OTB-2015 数据集上的 Precision 和 Success 指标曲线. 由表 1 和图 4 可以看出: OCT-CFDS 的 Success 指标最优, 对于 Precision 指标, OCT-CFDS 与 DaSiamRPN 相近且优于其它目标跟踪方法; 在跟踪速度方面, OCT-CFDS 比 ECO, MDNet, C-COT, MCPF, DeepSRDCF 和 SRDCF 等目标跟踪方法快, 与 DaSiamRPN, SiamRPN++ 和 TADT 等目标跟踪方法相比, OCT-CFDS

表 1 10 种目标跟踪方法在 OTB-2015 数据集上的性能指标

跟踪方法	Precision (pixels)	Success (AUC)	速度 (FPS)
OCT-CFDS	0.922	0.700	31
DaSiamRPN	0.923	0.665	160
SiamRPN++	0.915	0.696	35
ECO	0.910	0.691	8
MDNet	0.909	0.678	1
C-COT	0.903	0.673	0.3
MCPF	0.873	0.628	3.2
TADT	0.866	0.660	33.7
DeepSRDCF	0.851	0.635	1
SRDCF	0.789	0.598	8.2

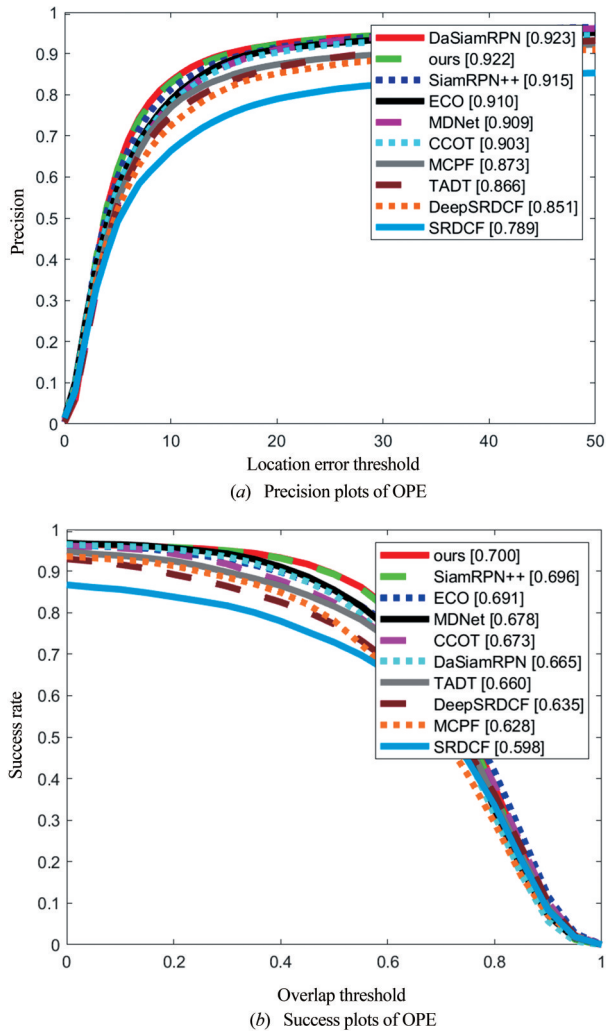


图 4 10 种目标跟踪方法在 OTB-2015 数据集上的性能指标曲线

表 2 10 种目标跟踪方法在 11 个不同属性视频上的 Precision 指标

跟踪方法	IV	SV	OCC	DEF	MB	FM	IPR	OPR	OV	BC	LR
OCT-CFDS	0.923	0.903	0.892	0.941	0.876	0.927	0.915	0.919	0.911	0.943	0.907
DaSiamRPN	0.917	0.874	0.885	0.940	0.870	0.894	0.901	0.914	0.867	0.924	0.900
SiamRPN++	0.925	0.905	0.883	0.913	0.878	0.929	0.916	0.920	0.838	0.903	0.911
ECO	0.914	0.870	0.889	0.922	0.859	0.888	0.880	0.904	0.906	0.942	0.888
MDNet	0.915	0.867	0.891	0.907	0.874	0.882	0.887	0.902	0.812	0.925	0.854
C-COT	0.884	0.869	0.888	0.916	0.872	0.884	0.875	0.894	0.889	0.882	0.885
MCPF	0.881	0.868	0.844	0.859	0.829	0.864	0.855	0.862	0.746	0.823	0.918
TADT	0.864	0.840	0.849	0.866	0.822	0.814	0.850	0.862	0.802	0.805	0.890
DeepSRDCF	0.791	0.805	0.808	0.836	0.811	0.818	0.811	0.834	0.794	0.841	0.702
SRDCF	0.792	0.695	0.750	0.816	0.781	0.728	0.702	0.773	0.619	0.775	0.631

6.3.2 VOT-2018 数据集

表 4 给出了本文方法 OCT-CFDS 和其它 10 种高性能目标跟踪方法在 VOT-2018 数据集上的跟踪性能指标. AFOD^[29], D3S^[28] 和 SiamMask^[25] 是基于区域分割的目标跟踪方法, 其它高性能目标跟踪方法是

速度较慢, 但仍能够实时跟踪目标. 与 DaSiamRPN, SiamRPN++ 和 TADT 相比, 本文方法速度慢的原因是: 本文方法在得到目标初始框后需要利用 Deep Snake 将其变形为目标轮廓, 并通过轮廓拟合出矩形框.

表 2 给出了 10 种目标跟踪方法对 OTB-2015 数据集中 11 个不同挑战属性视频目标跟踪的 Precision 指标, 由表 2 可以看出: 本文方法 OCT-CFDS 在遮挡 (OCC)、形变 (DEF)、目标出界 (OV) 和背景混杂 (BC) 等挑战性视频中 Precision 均有提升, 这一结果表明 OCT-CFDS 提升了目标跟踪的精度.

表 3 给出了 10 种目标跟踪方法对 OTB-2015 数据集中 11 个不同挑战属性视频目标跟踪的 Success 指标, 由表 3 可以看出: 本文方法 OCT-CFDS 在光照变化 (IV)、尺度变化 (SV)、遮挡 (OCC)、快速运动 (FM)、平面内旋转 (IPR)、平面外旋转 (OPR) 和背景混杂 (BC) 等挑战性视频中 Success 均有提升, 这一结果表明 OCT-CFDS 提升了目标跟踪的成功率.

图 5 给出了 10 种目标跟踪方法在 4 个挑战性视频上的跟踪结果. Biker 视频的主要挑战是运动模糊和快速运动, Bolt2 视频的主要挑战是目标形变和背景混杂, CarScale 视频的主要挑战是尺度变化和平面内旋转, Ironman 视频的主要挑战是光照变化、背景混杂和低分辨率. 由图 6 可以看出, 本文方法 OCT-CFDS 在应对目标快速移动、尺度变化和背景混杂等挑战时表现较优.

图 6 给出了本文方法 OCT-CFDS 在 OTB-2015 数据集上的轮廓跟踪结果, 由图 7 可以看出, 本文方法 OCT-CFDS 得到了平滑、准确的目标轮廓.

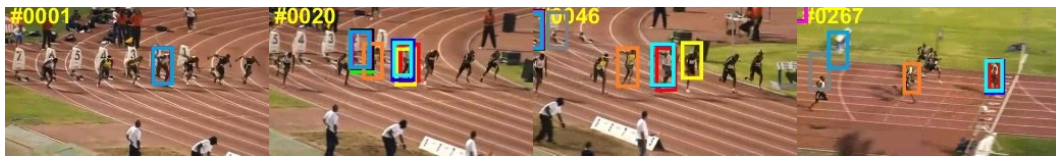
LADCF (Learning Adaptive Discriminative Correlation Filters)^[12], SiamRPN++^[21], DaSiamRPN^[20], SPM (Series-Parallel Matching)^[23], ASRCF (Adaptive Spatially-Regularized Correlation Filters)^[13], STRCF (Spatial-Temporal Regularized Correlation Filters)^[14] 和 ATOM

表 3 10种目标跟踪方法在11个不同属性视频上的Success 指标

跟踪方法	IV	SV	OCC	DEF	MB	FM	IPR	OPR	OV	BC	LR
OCT-CFDS	0.716	0.693	0.663	0.717	0.681	0.694	0.685	0.696	0.678	0.703	0.660
DaSiamRPN	0.642	0.637	0.619	0.691	0.640	0.643	0.650	0.658	0.639	0.660	0.623
SiamRPN++	0.713	0.690	0.655	0.710	0.674	0.691	0.682	0.692	0.648	0.691	0.658
ECO	0.713	0.664	0.657	0.716	0.680	0.652	0.652	0.677	0.681	0.700	0.617
MDNet	0.689	0.651	0.652	0.699	0.673	0.637	0.647	0.665	0.631	0.676	0.591
C-COT	0.682	0.651	0.647	0.698	0.680	0.639	0.632	0.658	0.665	0.652	0.619
MCPF	0.628	0.620	0.595	0.620	0.573	0.611	0.598	0.607	0.558	0.601	0.598
TADT	0.676	0.641	0.630	0.678	0.656	0.614	0.630	0.650	0.631	0.622	0.646
DeepSRDCF	0.621	0.601	0.581	0.647	0.632	0.595	0.584	0.612	0.583	0.627	0.474
SRDCF	0.613	0.535	0.556	0.625	0.608	0.540	0.524	0.573	0.487	0.583	0.480



(a) Biker



(b) Bolt2



(c) CarScale



(d) Ironman



图 5 10种高性能跟踪方法在4个挑战性视频上的跟踪结果

(Accurate Tracking by Overlap Maximization)^[15]. 表 4 中的结果表明:OCT-CFDS 相较于其它高性能方法提升了精度,增强了目标跟踪的鲁棒性. 图 7 为 OCT-CFDS 在 VOT-2018 数据集上的跟踪结果,由图 7 可以看出,OCT-CFDS 能够准确地跟踪目标轮廓.

为了验证掩膜分割的准确性,在 VOT-2018 数据集上,利用标注掩膜和预测掩膜的 mIOU^[28]对 OCT-CFDS、

D3S 和 SiamMask 进行比较. 实验结果如表 5 所示,由表 5 可以看出,OCT-CFDS 的 mIOU 优于其他方法.

6.3.3 计算复杂度分析

对现有基于区域分割的目标跟踪方法和本文方法的计算复杂度进行了分析. 由于 OCT-CFDS、SiamMask、D3S 和 AFOD 等方法的计算复杂度主要取决于其网络的计算复杂度,因此本文以网络的计算复杂度作

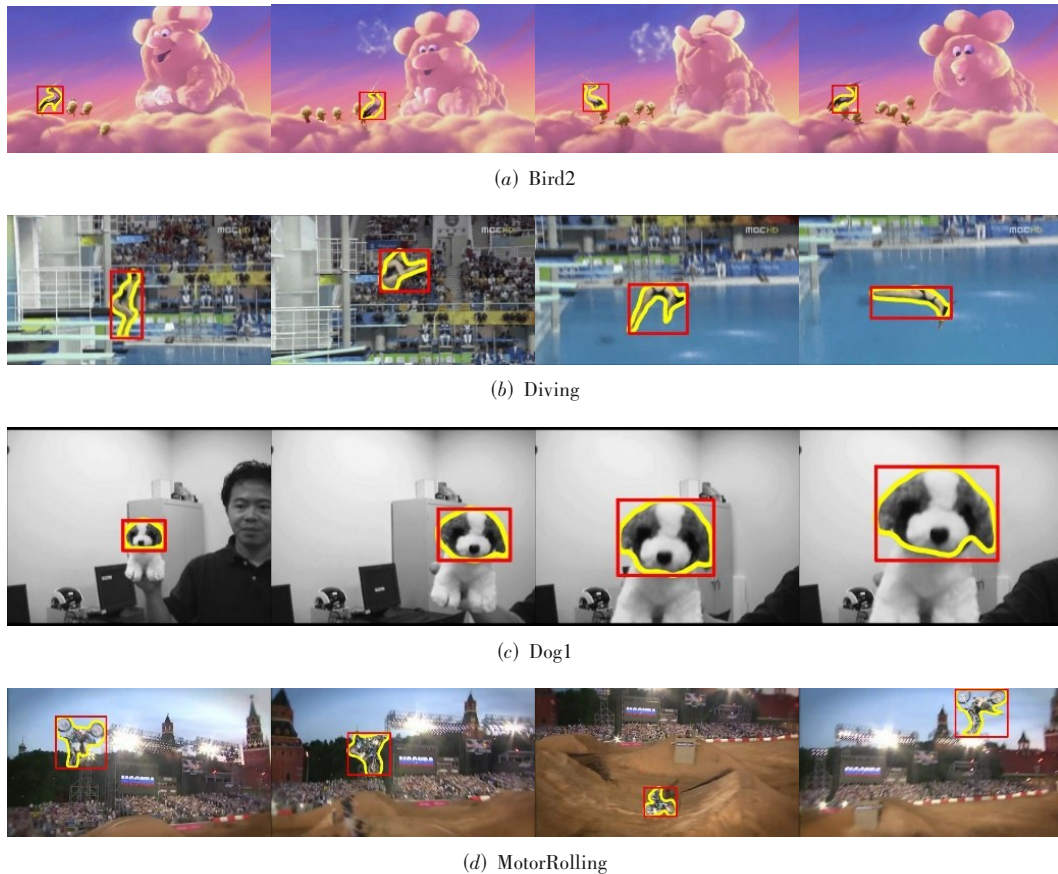


图6 OCT-CFDS在OTB-2015数据集上的轮廓跟踪结果

表4 11种目标跟踪方法在VOT-2018数据集上的性能指标

	OCT-CFDS	AFOD	D3S	SiamRPN++	ATOM	DaSiamRPN	LADCF	SiamMask	SPM	ASRCF	STRCF
EAO ↑	0.506	0.491	0.489	0.414	0.401	0.383	0.389	0.380	0.338	0.328	0.345
Accuracy ↑	0.67	0.634	0.64	0.60	0.59	0.59	0.51	0.61	0.58	0.49	0.523
Robustness ↓	0.113	0.117	0.150	0.234	0.204	0.276	0.159	0.276	0.300	0.234	0.215
速度(FPS) ↑	31	20	25	35	30	160	10.8	55	120	28	2.9

表5 3种目标跟踪方法在VOT-2018数据集上的mIOU指标

	OCT-CFDS	D3S	SiamMask
mIOU	0.68	0.66	0.63

表6 4种目标跟踪方法的计算复杂度

跟踪方法	骨干网络	输入图像尺寸	计算复杂度 (FLOPs)
OCT-CFDS	DLA-34	255×255	5.64×10 ⁹
SiamMask	Resnet-50	255×255	7.23×10 ⁹
D3S	Resnet-50	255×255	6.81×10 ⁹
AFOD	Resnet-50	255×255	6.97×10 ⁹

为其计算复杂度进行分析. 分析方法采用文献[36]中的方法, 计算复杂度计算公式为: $O\left(\sum_{l=1}^D M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l\right)$, 其中 D 是网络深度, l 代表网络第 l 层, M_l 是网络第 l

层的输出特征图大小, K_l 是网络第 l 层的卷积核大小, C_{l-1} 是网络第 l 层的输入通道数, C_l 是网络第 l 层的输出通道数. 利用该方法得到本文方法 OCT-CFDS, SiamMask, D3S 和 AFOD 的计算复杂度如表6所示, 由表6可以看出, SiamMask, D3S 和 AFOD 的计算复杂度高, 本文方法 OCT-CFDS 的计算复杂度较小.

6.3.4 讨论

综合跟踪精度、鲁棒性和速度三个性能指标, 由上述实验可以看出: 与现有目标跟踪方法相比, 本文方法具有较高的性能. 其主要原因在于: 一方面, 现有方法^[25, 28, 29]采用基于区域的分割方法, 计算复杂度高且分割精度依赖于跟踪框的准确性, 而本文采用基于轮廓的分割方法, 计算量小且对跟踪框具有

鲁棒性^[34];另一方面,现有方法^[3,18-21]利用多尺度搜索和锚框回归估计目标的尺度,精度低且速度慢,而本文方法利用轮廓上多点回归目标的尺度,精度高且分割速度快.本文方法对目标姿态变化、几何形变、目标运动、光照变化、视角变化和目标遮挡等因素引起目标表观的变化具有适应性.其原因有两个

方面:一是空间加权对数似然比相关滤波器的自适应性;二是Deep Snake对目标边缘的回归能力,虽然目标表观发生了变化,但是其轮廓是存在的.另外,本文方法利用空间加权对数似然比相关滤波器,自适应融合多特征估计目标的初始矩形框,提升了对复杂背景的鲁棒性.



图7 OCT-CFDS在VOT-2018数据集上的跟踪结果

7 结论

针对基于区域分割的目标跟踪方法存在的速度慢和掩膜精度受限目标跟踪框的问题,本文提出基于空间加权对数似然比相关滤波与Deep Snake的目标轮廓跟踪方法.一方面,本文利用Deep Snake提出一种基于轮廓分割的目标跟踪方法,提高了分割速度,增强了对目标变化的鲁棒性;另一方面,本文提出空间加权对数似然比相关滤波器,自适应融合多特征估计目标的初始矩形框,提升了对复杂背景的鲁棒性.利用OTB-2015和VOT-2018数据集对提出的方法进行了实验验证,结果表明:与现有先进的目标跟踪方法相比,本文提出的跟踪方法取得了较好的跟踪性能.

致谢 感谢李从利教授给本文提出的参考意见.

参考文献

- [1] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2010: 2544-2550.
- [2] 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.

- [3] DANELLJAN M, HÄGER G, SHAHBAZ KHAN F, et al. Accurate scale estimation for robust visual tracking[C]// Proceedings of the British Machine Vision Conference 2014. Nottingham: BMVA Press, 2014: 1-11.
- [4] DANELLJAN M, ROBINSON A, KHAN F S, et al. Beyond correlation filters: Learning continuous convolution operators for visual tracking[C]//European Conference on Computer Vision. Cham: Springer, 2016: 472-488.
- [5] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: Efficient convolution operators for tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6931-6939.
- [6] BHAT G, JOHNANDER J, DANELLJAN M, et al. Unveiling the power of deep tracking[C]//European Conference on Computer Vision. Cham: Springer, 2018: 493-509.
- [7] DANELLJAN M, KHAN F S, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 1090-1097.
- [8] DANELLJAN M, HÄGER G, KHAN F S, et al. Convolutional features for correlation filter based visual tracking [C]//2015 IEEE International Conference on Computer Vision Workshop. Piscataway: IEEE, 2015: 621-629.
- [9] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: Complementary learners for real-time tracking[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1401-1409.
- [10] ZHANG T Z, XU C S, YANG M H. Multi-task correlation particle filter for robust object tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 4819-4827.
- [11] DANELLJAN M, HÄGER G, KHAN F S, et al. Learning spatially regularized correlation filters for visual tracking [C]//2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 4310-4318.
- [12] XU T Y, 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: A Publication of the IEEE Signal Processing Society, 2019, 28(11): 5596-5609.
- [13] DAI K N, WANG D, LU H C, et al. Visual tracking via adaptive spatially-regularized correlation filters[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 4665-4674.
- [14] LI F, TIAN C, ZUO W M, et al. Learning spatial-temporal regularized correlation filters for visual tracking[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4904-4913.
- [15] DANELLJAN M, BHAT G, KHAN F S, et al. ATOM: Accurate tracking by overlap maximization[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 4655-4664.
- [16] 蒲磊, 冯新喜, 侯志强, 等. 基于深度空间正则化的相关滤波跟踪算法[J]. 电子学报, 2020, 48(10): 2025-2032.
- PU L, FENG X X, HOU Z Q, et al. Correlation filter tracking based on deep spatial regularization[J]. Acta Electronica Sinica, 2020, 48(10): 2025-2032. (in Chinese)
- [17] TAO R, GAVVES E, SMEULDERS A W M. Siamese instance search for tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1420-1429.
- [18] 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.
- [19] LI B, YAN J J, WU W, et al. High performance visual tracking with Siamese region proposal network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8971-8980.
- [20] ZHU Z, WANG Q, LI B, et al. Distractor-aware siamese networks for visual object tracking[C]//European Conference on Computer Vision. Cham: Springer, 2018: 103-119.
- [21] LI B, WU W, WANG Q, et al. SiamRPN: Evolution of Siamese visual tracking with very deep networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 4277-4286.
- [22] ZHANG Z P, PENG H W, FU J L, et al. Ocean: Object-aware anchor-free tracking[C]//European Conference on Computer Vision. Cham: Springer, 2020: 771-787.
- [23] WANG G T, LUO C, XIONG Z W, et al. SPM-tracker: Series-parallel matching for real-time visual object tracking[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 3638-3647.
- [24] CHENG J C, TSAI Y H, HUNG W C, et al. Fast and accurate online video object segmentation via tracking parts [C]//2018 IEEE/CVF Conference on Computer Vision

- and Pattern Recognition. Piscataway: IEEE, 2018: 7415-7424.
- [25] WANG Q, ZHANG L, BERTINETTO L, et al. Fast online object tracking and segmentation: A unifying approach[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 1328-1338.
- [26] CHEN X, LI Z X, YUAN Y, et al. State-aware tracker for real-time video object segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9381-9390.
- [27] SUN M J, XIAO J M, LIM E G, et al. Fast template matching and update for video object tracking and segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 10788-10796.
- [28] LUKEŽIČ A, MATAS J, KRISTAN M. D3S—A discriminative single shot segmentation tracker[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 7131-7140.
- [29] CHEN Y W, XU J T, YU J Q, et al. AFOD: Adaptive focused discriminative segmentation tracker[C]//European Conference on Computer Vision. Cham: Springer, 2020: 666-682.
- [30] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4293-4302.
- [31] LI X, MA C, WU B Y, et al. Target-aware deep tracking [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 1369-1378.
- [32] PENG S D, JIANG W, PI H J, et al. Deep snake for real-time instance segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 8530-8539.
- [33] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [34] KRISTAN M, LEONARDIS A, MATAS J, et al. The sixth visual object tracking VOT 2018 challenge results [C]//European Conference on Computer Vision. Cham: Springer, 2019: 3-53.
- [35] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points [EB/OL]. (2019-04-16)[2022-05-26]. <https://arxiv.org/abs/1904.07850>.
- [36] XIE E Z, SUN P Z, SONG X G, et al. PolarMask: Single shot instance segmentation with polar representation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 12190-12199.

作者简介



李豪男, 1994年生, 硕士研究生. 主要从事实例分割、目标跟踪及目标检测方面的研究.

E-mail: HaoLi086@163.com



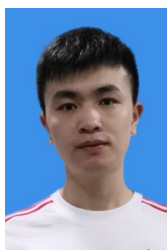
袁广林(通讯作者) 男, 1973年生, 博士, 副教授. 主要从事计算机视觉、机器学习及其应用方面的研究.

E-mail: 1183212999@qq.com



秦晓燕 女, 1980年生, 硕士, 副教授. 主要从事目标检测、机器学习及其应用方面的研究.

E-mail: 70853559@qq.com



锯长瑞 男, 1994年生, 硕士研究生. 主要从事计算机视觉方面的研究.

E-mail: juchangrui1994@163.com



朱虹 女, 1987年生, 硕士, 讲师. 主要从事图像处理、计算机视觉方面的研究.

E-mail: 729039126@qq.com