

DOI:10.16136/j.joel.2022.09.0015

跟踪状态自适应的判别式行人单目标跟踪算法研究

丁明远, 蔡 靖, 周 冕, 薛彦兵*, 温显斌

(天津理工大学 计算机视觉与系统教育部重点实验室和天津市智能计算及软件新技术重点实验室, 天津 300384)

摘要: 本文针对在行人跟踪过程中遇到的背景相似物干扰、行人之间的相互遮挡和背景杂乱等导致跟踪状态不稳定的问题, 基于 DIMP (learning discriminative model prediction for tracking) 跟踪算法, 提出了一种跟踪状态自适应的判别式单目标行人跟踪算法。跟踪过程中由分类滤波器和搜索区域进行卷积操作得到响应图, 通过响应图判断跟踪状态, 跟踪状态分为弱响应状态、多峰强响应状态、单峰强响应状态。针对多峰强响应状态下的干扰物影响, 提出在线更新策略, 利用激励和抑制损失更新分类滤波器, 提高分类滤波器的判别能力。针对多峰强响应和弱响应状态下目标预测不准确的问题, 通过偏移量和增添候选框修正目标位置, 提高跟踪精度。实验验证提出的算法在行人视频序列上跟踪结果, 精度达到了 0.978, 成功达到了 0.740, 在 NVIDIA GTX 1650 显卡下有 30 fps 的实时速度。

关键词: 行人单目标跟踪; DIMP 算法; 在线更新; 分类滤波器; 跟踪状态

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1005-0086(2022)09-0940-08

Research on discriminative pedestrian single target tracking algorithm with adaptive tracking state

DING Mingyuan, CAI Jing, ZHOU Mian, XUE Yanbing*, WEN Xianbin

(Key Laboratory of Computer Vision and System, Ministry of Education of China, Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China)

Abstract: Based on tracking algorithm of learning discriminative model prediction for tracking (DIMP), a discriminative single target pedestrian tracking algorithm with adaptive tracking state is proposed to address the problems of unstable tracking state due to background similarities interference, mutual occlusion between pedestrians and background clutter encountered in the pedestrian tracking process. The response map is obtained by the convolution operation of the classification filter and the search region in the tracking process, and the tracking state is divided into weak response state, multi-peak strong response state, and single-peak strong response state by the response map. For the influence of disturbances in the multi-peak strong response state, an online update strategy is proposed to update the classification filter by using the excitation and suppression losses to improve the discriminative ability of the classification filter. For the problem of inaccurate target prediction in multi-peak strong response and weak response states, the target position is corrected by offset and adding candidate frames to improve the tracking accuracy. The proposed algorithm is experimentally verified, which achieves precision of 0.978 and a success rate of 0.740 on pedestrian video sequences with a real-time speed of 30 fps under NVIDIA GTX 1650.

Key words: pedestrian single target tracking; learning discriminative model prediction for tracking

* E-mail: xueyb0718@tjut.edu.cn

收稿日期: 2022-01-07 修订日期: 2022-01-24

基金项目: 天津市自然科学基金(18JCYBJC85500)、天津市人工智能专项(18ZXZNGX00150)和天津市科学技术局技术创新引导专项(21YDTPJC00250)资助项目

(DIMP) algorithm; online update; classification filter; tracking status

1 引言

目标跟踪是计算机视觉的一个重要的研究分支,在智能监控、人机交互、无人驾驶等方面得到了广泛地应用^[1]。行人单目标跟踪作为目标跟踪的一项子任务,在智能安防、智能交通等方面发挥了关键的作用。

根据跟踪算法的工作原理,跟踪算法通常分为生成式和判别式^[2]。早期跟踪算法主要集中于生成式算法研究,生成式算法首先通过目标特征建立目标模型,在后续跟踪过程中进行相似特征搜索,逐步迭代实现目标定位^[3]。早期生成式模型跟踪算法,有代表性的包括光流法、Kalman 滤波算法、MeanShift 算法等。CamShift 在 MeanShift 的基础上加入尺度自适应机制,同时引入 HSV 直方图提升表征能力。针对 CamShift 中的颜色特征信息较为单一的问题,后续研究者利用颜色、纹理、目标动作信息的全面特性结合 Kalman 滤波方法进行改进^[4]。

判别式模型同时考虑目标和背景信息,对目标和背景之间进行判别分类,将目标从背景中提取出来。基于判别式模型的相关滤波跟踪算法由于速度优势受到广泛关注。其中经典的跟踪算法有 MOSSE(minimum output sum of squared error)算法、KCF(kernelized correlation filter)算法等。针对模糊图像和尺度变换问题,林堪渺等^[5]在 KCF 基础上通过构建抗模糊特征检测与匹配模型,同时引入尺度缩放因子,估计目标最佳尺度。Staple 算法^[6]在 KCF 的基础上通过结合颜色直方图和 HOG(histogram of oriented gradient)特征提升目标的表征能力。针对遮挡、旋转问题,火元莲等研究者在 Staple 算法基础上提出了一种融合重检测机制的上下文感知目标跟踪算法(context-aware tracking fused with re-detection method, CA-RD),通过重检测增强了跟踪器的遮挡检测能力^[7]。

深度学习的出现,使得目标跟踪得到了进一步的发展。BERTINETO 等^[8]提出的 SiamFC(fully-convolutional siamese networks for object tracking)算法,确立了一种以全卷积神经网络为骨干网络的孪生架构,通过离线训练得到一个普适的相似性学习模型,在跟踪过程中预测目标的位置。为了应对跟踪过程中目标尺度的变化,REN 等^[9]引入 Faster R-CNN(faster region-convolutional neural network)中的 RPN(region proposal network)结构提出 SiamRPN(siamese with region

proposal network)^[10]。后续结合深层网络加强表征能力,如 SiamRPN++^[11]、SiamDW(deeper and wider siamese network)等^[12],采用 anchor-free 思想替代 RPN 结构如 SiamCAR(siamese fully convolutional classification and regression)^[13]、Ocean(object-aware anchor-free tracking)等^[14]算法,均在一定程度上提升跟踪模型的性能。

ATOM(accurate tracking by overlap maximization)算法^[15]结合深度学习和相关滤波二者的优势,通过将目标跟踪任务解耦为两个阶段,目标分类任务和目标估计任务来实现跟踪。目标分类任务用于目标粗略定位;在目标估计任务中,引入 IoU-Net(intersection over union network)^[16]结构预测目标状态。DIMP(learning discriminative model prediction for tracking)^[17]在 ATOM 的基础上进行改进,通过离线加在线的训练方式使得分类滤波器有着更强的判别能力,进一步提升了模型的性能。

在行人跟踪过程中,场景可能会复杂且多变,产生行人之间的相互遮挡、干扰等问题,DIMP 算法在上述情况下产生漂移甚至丢失,导致跟踪精度下降。跟踪过程中分类滤波器和当前测试帧图像搜索区域进行卷积得到当前场景的响应图,响应图反映出目标在当前场景下所处的状态,预测目标位置。针对上述问题,本文利用当前场景响应图的复杂性判断当前跟踪状态,提出一种跟踪状态自适应的判别式单目标行人跟踪算法。算法基于跟踪状态的更新策略,利用激励和抑制损失进行更新,加强分类滤波器的判别能力。处于多峰强响应状态和弱响应状态下,同时基于状态修正预测目标的位置。综合调优得出相对目标真实状态和位置误差最小的边界框,以此提升模型的成功率和精度。

2 DIMP 算法

DIMP 算法基于 ATOM 算法进行改进,同样地将目标跟踪任务解耦成目标分类任务和目标估计任务。

在目标分类任务中,通过初始化模块使初始帧初始化和优化器模块优化更新分类滤波器用于后续前、后背景分类,预测目标位置。选取第一帧作为参考帧并通过数据增强得到训练样本,同时根据其目标边界框作为 GT(ground truth),在初始化模块中通过 PrPool(precision roi pooling)对边界框内的特征进行池化提取得到初始化分类滤波器 f_0 。在优化

器模块中利用初始化分类滤波器 f_0 和训练样本特征进行相关即卷积操作得到训练样本响应图 s , 根据目标标注边界框得到训练样本的理想响应图 y_c , 即以目标中心生成的标签。通过损失函数计算响应图和标签之间的损失, 利用最速梯度下降法迭代优化更新分类滤波器, 通过迭代获得最终用于跟踪的分类滤波器 f 。

优化器模块的更新方式:

1) 通过初始化分类器 f_0 和训练样本特征 x 进行相关操作得到训练样本响应图 s , 如式(1)所示:

$$s = f_0 * x. \quad (1)$$

式中, $*$ 代表卷积操作。

2) 通过损失函数计算损失, DIMP 算法根据 SVM(support vector machine)中的 hinge-like 损失思想提出了一种新的损失函数, 如式(2)所示:

$$L(f) = \frac{1}{S_{\text{train}}} \cdot \sum_{(x,c) \in S_{\text{train}}} \|r(x * f, c)\|^2 + \|\lambda f\|^2, \quad (2)$$

式中, 公式转行时每行末尾 $\cdot$ 代表点乘符号, $S_{\text{train}} = \{(x_j, c_j)\}_{j=1}^{(n)}$ 其中 x_j 表示训练样本特征, c_j 表示其标注的目标边界框中心, (n) 表示样本数量; λ 表示惩罚项; $r(x * f, c)$ 表示预测响应图和真实标签的之间的残差, 计算方式如式(3)所示:

$$r(x * f, c) = v_c(m_c s + (1 - m_c) \max(0, s) - y_c), \quad (3)$$

式中, $s = x * f$ 表示预测响应图, v_c 为空间权重, m_c 为目标位置掩膜参数, 如式(5)所示, y_c 为样本标签如式(4)所示:

$$y_c(t) = \sum_{k=1}^{N-1} \phi_k^{(y)} \rho_k(\|t - c\|), \quad (4)$$

$$m_c(t) = \sum_{k=1}^{N-1} \phi_k^{(m)} \rho_k(\|t - c\|), \quad (5)$$

式中, t 表示样本标签中的每个位置, c 代表目标中心位置; $\phi_k^{(y)}$ 表示计算 y_c 中的空间系数, $\phi_k^{(m)}$ 表示计算 m_c 中的空间系数, 在离线训练时得到, ρ_k 代表距离计算函数, 如式(6)所示:

$$\rho_k(d) = \begin{cases} \max(0, 1 - \frac{|d - k\Delta|}{\Delta}), & k < N-1 \\ \max(0, \min(1, 1 + \frac{|d - k\Delta|}{\Delta}), & k = N-1 \end{cases}, \quad (6)$$

式中, d 表示位置到中心点 c 的距离, $k = N-1$ 表示所谓远离目标中心的位置, $N=9$, Δ 为结点位移。

3) 根据计算得到的损失 $L(f)$ 计算其梯度, 通过最速梯度下降法求得损失梯度更新分类滤波器 f 。

$$f_{(l+1)} = f_l + \alpha \nabla L(f_{(l)}), \quad (7)$$

式中, $f_{(l)}$ 代表第 i 次优化后的分类滤波器, α 代表学习率, ∇ 代表梯度计算。

将当前视频序列中的后续帧作为测试帧, 利用特征提取网络对测试帧进行特征提取, 得到测试帧的特征 x_{test} , 通过第一帧得到的分类滤波器 f 和测试帧特征进行互相关操作即卷积计算, 得到测试帧的响应图 s_{test} 。响应图最大响应得分位置即为预测的目标粗略位置。

在目标估计任务中, 以分类任务预测的目标粗略位置作为中心, 在其周围随机生成候选边界框集, 送入到 IoU-Net 网络架构中, 通过梯度上升最大化 IoU(intersection over union)得到最优边界框拟合目标状态和位置。通过分类任务确定目标位置, 将其作为中心点结合上一帧目标边界框尺寸生成候选边界框, 送入到 IoU-Net 网络架构中, 计算候选边界框的 IoU, 然后通过预测得到的 IoU 计算边界框梯度, 通过梯度上升最大化 IoU 的方式分别对候选边界框进行迭代优化, 经过迭代后即可得到优化后的候选边界框, 取 IoU 最大的 3 个候选边界框的坐标平均值, 作为预测边界框的坐标, 即为最终预测的目标边界框。

3 本文算法

3.1 判断跟踪状态

根据得到的测试帧的响应图 s_{test} 判断当前测试帧的跟踪状态, 跟踪状态分为单峰强响应状态(图 1 所示)、多峰强响应状态(图 2 所示)和弱响应状态(图 3)。

1) 弱响应状态(weak response state)。测试帧的响应图 s_{test} 中目标区域内响应得分相对杂乱, 同时最高响应得分也会陡然下降, 则说明此时待跟踪目标处于弱响应状态, 表征此时待跟踪目标可能遭遇遮挡、背景混淆、待跟踪目标和干扰物接近。如图 3 中所示。

首选判断是否为弱响应状态, 通过滑动方式计算测试帧之前的 p 帧的目标区域响应得分方差均值 $\bar{\sigma}$, 目标区域是利用上一帧目标边界框尺寸在测试帧的响应图最高响应得分位置周围框选的区域。记录测试帧之前的 p 帧的目标区域的响应得分方差 σ , 计算出其平均值, 即为响应得分方差均值 $\bar{\sigma}$; 测试帧最高响应得分 m_score 为测试帧的响应图 s_{test} 中响应得分的最大值, 记录测试帧之前的 p 帧各自的最高响应得分, 计算 p 帧平均值即为最高响应得分均值 $\overline{m_score}$, 当前测试帧之前的帧数量小于 p 帧时, 则计算到第二帧即可。利用式(8)计算目标区域响应得分方差 σ , 利用式(9)计算响应得分方差均值 $\bar{\sigma}$, 利用式(10)计算最高响应得分均值 $\overline{m_score}$ 。

$$\sigma = \frac{\sqrt{\sum_i^{w \times h} (score_i - \overline{score})^2}}{w \times h}, \quad (8)$$

式中, $score_i$ 为测试帧的响应图中对应目标区域中各个位置得分, 其中 i 代表响应图目标区域各个位置, $\overline{score}$ 为测试帧的响应图 s_{test} 中对应目标区域各个位置得分均值, $w \times h$ 为测试帧的响应图 s_{test} 中对应目标区域宽和高的尺寸乘积。

$$\bar{\sigma} = \frac{\sum_{j=1}^p \sigma_j}{p}, \quad (9)$$

$$\overline{m_score} = \frac{\sum_{j=1}^p m_score}{p}, \quad (10)$$

式中, p 代表当前测试帧之前的 p 帧。

若满足式(11), 则判断当前测试帧的跟踪状态

为弱响应状态。

$$(\sigma < z \times \bar{\sigma}) \& (m_score < z \times \overline{m_score}), \quad (11)$$

式中, $p=25$, $z=0.75$ 。若测试帧之前的帧数小于 25 帧, 则该计算过程从测试帧往前直到第二帧为止, 其中 p 和比例系数 z 取值是通过实验总结得到的最优选择。

2) 多峰强响应状态(multi-peak strong response state)。当测试帧搜索区域响应图 s_{test} 中有多个响应峰, 此时在搜索区域内遭遇背景干扰物, 背景干扰物和目标距离较远, 目标响应峰值和干扰物响应峰值同时有较高的响应得分。如图 2 所示。

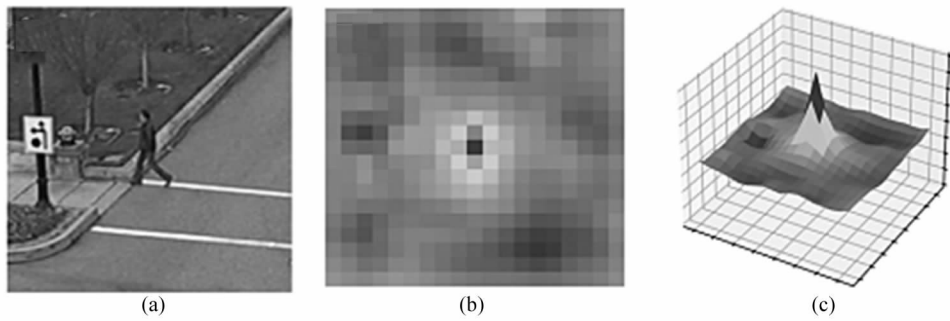


图 1 单峰强响应状态示意图:(a) 搜索区域图像;(b) 搜索区域响应热点图;(c) 搜索区域响应峰图

Fig. 1 Schematic diagram of the single-peak strong response state:(a) Image of the search area;

(b) Map of search area response hot spot; (c) Map of search area response peak

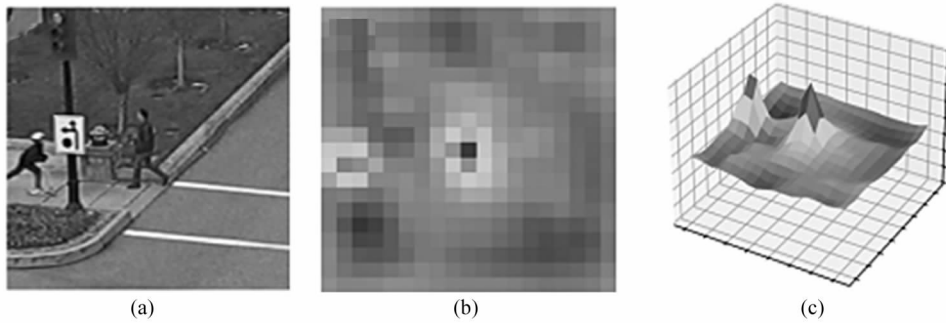


图 2 多峰强响应状态示意图:(a) 搜索区域图像;(b) 搜索区域响应热点图;(c) 搜索区域响应峰图

Fig. 2 Schematic diagram of the multi-peak strong response state:(a) Image of the search area;

(b) Map of search area response hot spot; (c) Map of search area response peak

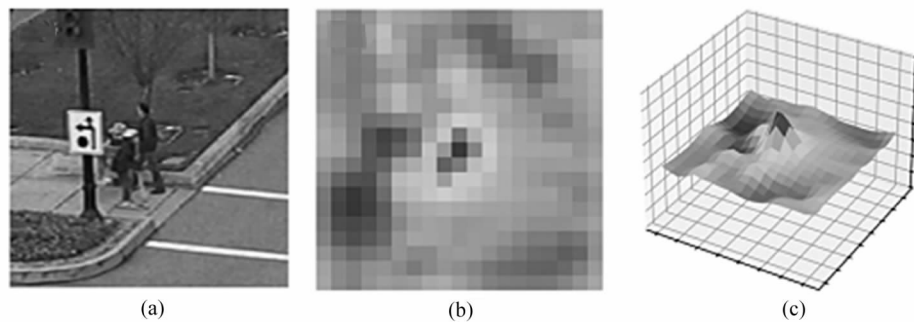


图 3 弱响应状态示意图:(a) 搜索区域图像;(b) 搜索区域响应热点图;(c) 搜索区域响应峰图

Fig. 3 Schematic diagram of the weak response state:(a) Image of the search area;

(b) Map of search area response hot spot; (c) Map of search area response peak

若当前跟踪状态不是弱响应状态的情况下,记录响应图目标区域外的响应得分最大值为 dis_score 。若满足式(12)则判断当前跟踪状态为多峰强响应状态。

$$dis_score > 0.8 \times m_score. \quad (12)$$

3) 单峰强响应状态(single-peak strong response state)。若不满足式(12)和式(11)则为单峰强响应状态。

不满足多峰强响应状态,表示测试帧搜索区域的响应图只有单峰强响应,且响应峰值拥有较高的响应得分,目标区域内响应得分相对简单,说明此时只有待跟踪目标或者待跟踪目标和背景有明确分界,代表测试帧的场景简单,通过分类任务可以简单地识别待跟踪目标和背景。如图1所示。

3.2 基于跟踪状态的在线更新策略

在跟踪过程中遭遇遮挡、相似物干扰等问题时,DIMP算法的预测结果可能不够精确,若将其作为更新样本存放到更新样本池更新分类滤波器,在更新的过程中由于预测不准确可能会引入过多的背景信息,影响分类滤波器的判别能力。针对上述问题,提出基于不同的跟踪状态采用激励和抑制损失对分类滤波器进行更新,提升分类滤波器的判别能力。

单峰强响应状态下将当前测试帧特征和其预测边界框作为更新样本放到更新样本池中,在弱响应状态下不对更新样本池中的样本添加或替换,在单峰强响应状态下采用每隔20帧更新分类滤波器,或者多峰强响应状态下更新分类滤波器。

针对多峰强响应状态下干扰物的影响,提出了一种激励和抑制损失来对分类滤波器进行优化更新,旨在提高分类器对目标和干扰物之间判别能力,以增强模型的鲁棒性。针对样本响应图和样本标签的残差计算,当遭遇干扰物时在样本标签 y_c 中对目标干扰物区域赋予较大的负值得到新的样本标签 y_{new} ,可以抑制背景中干扰物的影响和激励目标的响应。如式(13)所示,对单峰强响应状态(single-peak)和多峰强响应状态(multi-peak)采取不同标签计算方式来达到激励和抑制的效果。其中响应图和样本标签之间的残差计算如式(14)所示。

$$y_{new} = \begin{cases} y_c, \text{single-peak} \\ y_c + y_{disc}, \text{multi-peak} \end{cases}, \quad (13)$$

$$r(x \times f, c) = v_c(m_c s + (1 - m_c) \max(0, s) - y_{new}), \quad (14)$$

$$y_c(t) = \sum_{k=1}^{N-1} \Phi_k^y \rho_k(\|t - c\|), \quad (15)$$

$$y_{disc}(t) = - \sum_{k=1}^{N-1} \phi_k^y \rho_k(\|t - disc\|), \quad (16)$$

$$y_{disc}(t) = \begin{cases} y_{disc}(t), y_{disc}(t) \leq 0 \\ 0, y_{disc}(t) > 0 \end{cases}, \quad (17)$$

式中, t 表示样本标签中的每个位置, c 代表目标中心位置; $disc$ 为干扰物中心位置; y_{new} 表示新定义的样本标签。

3.3 基于跟踪状态修正预测的目标位置

3.3.1 多峰强响应状态下修正预测的目标位置

在分类任务中,通过分类器和测试帧搜索区域特征进行卷积操作得到的测试帧响应图 s_{test} ,其中响应得分最大值所对应的位置坐标作为第一响应点,即 m_score 在响应图的位置。测试帧的响应图中划分的区域外为目标区域外,将目标区域外的最高响应得分所对应的位置视为第二响应点,即 dis_score 在响应图的位置。

假设当前第一响应点位置为 $c_1[x_1, y_1]$,第二响应点位置为 $c_2[x_2, y_2]$,当前测试帧的上一帧的目标中心点在测试帧的响应图的位置为 $c_0[x_0, y_0]$,即响应图中目标区域中心点,则 c_1 和 c_2 相对于 c_0 的位置偏移量,如式(18)和式(19)所示:

$$d_1 = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}, \quad (18)$$

$$d_2 = \sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2}. \quad (19)$$

判断当前目标的真实位置:

根据式(18)和式(19)计算得到的偏移量在不同的取值范围内,则返回不同的响应点位置作为预测的目标中心位置,如式(20)所示:

$$\begin{cases} c_1, (d_1 < \Omega \& d_2 > \Omega) / \\ (d_1 > \Omega \& d_2 > \Omega) / \\ (d_1 < d_2 \& d_1 < \Omega \& d_2 < \Omega), \\ c_2, (d_1 > \Omega \& d_2 < \Omega) / \\ (d_1 > d_2 \& d_1 < \Omega \& d_2 < \Omega) \end{cases}, \quad (20)$$

式中,公式转行时每行末尾||代表逻辑“或”符号, Ω 代表取值范围中的阈值。

3.3.2 弱响应状态下修正预测的目标位置

当目标处于弱响应状态下,可能遇到行人相互遮挡、背景相融等问题,此时通过分类任务中利用响应图最高响应确定的目标位置不够准确。

因此结合行人目标在运动过程中的固有特性,即行人目标在单峰强响应状态下目标跟踪正常,行人目标在运动过程中没有剧烈的尺度变换,行人运动相对平缓,所以记录单峰强响应状态下目标中心的偏移量,通过滑动帧的方式记录 v 帧中每两帧之间的目标中心相对偏移量,计算其均值作为预测偏

移量,在弱响应状态下通过预测偏移量对分类任务中预测的目标位置进行修正。

在回归任务中,针对弱响应状态下边界框预测不准确的问题,我们提出了新的候选边界框生成方案。通过上一帧的目标边界框作为初始候选边界框,若当前测试帧为第二帧,就将第一帧标注的目标边界框作为初始候选边界框,若为后续其他帧,则将预测的边界框作为初始候选边界框,在最终预测的目标中心位置周围,随机生成10个不同比例的候选边界框组成初始候选边界框集。结合行人目标在运动过程中的固有特性,行人目标在运动过程中尺度变化相对平缓,结合参考帧(即第一帧)人工标注的目标边界框,将其作为参考候选边界框,并在最终预测的目标中心位置周围,随机生成4个不同比例的参考候选边界框组成参考候选边界框集。将初始候选边界框集和参考候选边界框集进行组合,得到14个候选边界框作为候选边界框集,同样通过估计任务中边界框预测模块对最终目标边界框进行预测。参考候选边界框的选择首先以第一帧标注的边界框作为标准,在后续跟踪过程中处于单峰强响应状态,且响应图中最高响应得分处于较高水平时,将当前帧预测目标边界框作为新的参考边界框。

4 实验及数据分析

本文的实验是使用的 Python 3.7, CUDA 10.1 版本, Pytorch 1.4.0 框架作为基础编程环境, 设备系统为 Ubuntu 16.04, Intel(R) Core(TM) i5-10400F CPU @ 2.90 GHz、RAM 16 GB、GTX 1650 4 G 显存。

为验证本文所提出方法的可行性,选取具有挑战性的公开目标跟踪测试数据集(OTB100、VOT2018、TC128),对所有的行人目标序列进行提取,得到26个行人视频序列,作为行人视频序列测试数据集。并和DIMP、ATOM、SiamRPN++等优秀算法对比,下面将在行人视频序列测试集上进行定量分析消融实验来验证所提出方法的有效性。

4.1 定量分析

为了确定本文改进算法的性能,本小节采用 OTB100 数据集的评估方式,即精确度(Precision)和成功率(Success)对整个算法模型进行定量分析。

精确度(Precision):通过计算算法预测的目标边界框的中心点和真实目标边界框的中心点的位置偏差,设定阈值为20个像素位,若位置偏差小于一定的阈值,则认为预测正确,大于阈值,则代表预测错

误。通过对整个测试集中所有帧的预测结果和真实目标位置做位置偏差计算,预测正确的帧数除以整个测试集的帧数作为精度。

成功率(Success):通过计算算法预测的目标边界框和真实目标边界框的IoU,设定阈值为0.5,若IoU大于阈值,则认为预测正确,小于阈值,则代表预测错误。通过对整个测试集中所有帧的预测结果和真实目标位置作位置偏差计算,预测正确的帧数除以整个测试集的帧数作为成功率。

表1 各模型的 Success 指标和 Precision 指标

Tab.1 Success metrics and Precision metrics for each model

Tracker	Success	Precision
SiamRPN++	0.633	0.856
SiamDW	0.650	0.879
ATOM	0.696	0.908
DIMP	0.681	0.893
SiamCAR	0.655	0.868
Ocean	0.617	0.862
Our	0.740	0.978

通过实验评估,将本文算法与当前比较优秀、具有代表性的算法进行对比,包括 SiamRPN++、Ocean、SiamCAR、SiamDW、DIMP 和 ATOM。评估跟踪器的传统方法是在整个测试序列中运行跟踪器,利用一次评估(one-pass evaluation, OPE)进行评估,从第一帧中的目标位置开始初始化,并报告平均精度或成功率。评估结果如表1所示,从中可以看到本文的算法表现出最佳性能,本文的算法相较于 DIMP 在精度和成功率上获得了 8.5% 和 5.9% 的显著增益,同时本文的算法较其他算法都有着不错提升。结果证明本文的方法可以实现稳定准确的跟踪。

4.2 消融实验

本文提出的基于 DIMP 算法自适应跟踪状态行人的行人单目标跟踪算法,通过设立消融实验来验证我们提出方法的有效性。

从基于跟踪状态的更新策略(update strategies based on tracking states, US-TS)、基于多峰强响应状态修正预测的目标位置(correct predicted target position based on multi-peak strong response state, CPTP-MSR)以及基于弱响应状态修正预测的目标位置(correct predicted target position based on weak response state, CPTP-WR)3个方面来提升整

个模型的精度和成功率。从表 2 可以看到,本文的算法都较DIMP算法都有一定的提升,通过融合后

表 2 消融实验

Tab. 2 Ablation experiment

Baseline	USBS-TS	CPTP-MSR	CPTP-WR	Success	Precision
✓				0.663	0.870
✓	✓			0.690	0.917
✓		✓		0.698	0.923
✓			✓	0.720	0.950
✓	✓	✓	✓	0.740	0.978

使得新的算法得到了显著的提升。

5 结 论

针对行人跟踪过程中存在的行人之间遮挡、背景相似物干扰等问题,本文提出一种跟踪状态自适应的判别式单目标行人跟踪算法。首先根据场景变化,通过计算确立不同的跟踪状态,跟踪状态分为弱响应状态、多峰强响应状态、单峰强响应状态。提出一种基于跟踪状态的更新策略,利用激励和抑制损失对分类滤波器进行优化更新,以此来增加分类滤波器对目标和背景相似物的判别能力,处理相似物干扰问题。接着在针对弱跟踪状态和多峰响应跟踪状态目标预测位置不准确的问题,提出基于跟踪状态修正预测目标位置,在多峰强响应状态下利用位置偏移量确定目标位置,在弱响应状态下通过位置偏差修正目标位置,在回归任务中增添候选边界框获得更拟合真实目标的边界框,提升模型跟踪精度和成功率。

经过实验评估,本文的算法在行人单目标的跟踪上展现出较好的性能,在 GTX1650 上达到 30 fps 的实时速度,且其跟踪成功率更高,具有一定的实用价值。

参考文献:

- [1] LU H C, LI P X, WANG D. Visual object tracking: a survey [J]. Pattern Recognition and Artificial Intelligence, 2018, 31(1): 61-76.
卢湖川, 李佩霞, 王栋. 目标跟踪算法综述[J]. 模式识别与人工智能, 2018, 31(1): 61-76.
- [2] LI X, ZHA Y F, ZHANG T Z, et al. Survey of visual object tracking algorithms based on deep learning[J]. Journal of Image and Graphics, 2019, 24(12): 2057-2080.
李玺, 查宇飞, 张天柱, 等. 深度学习的目标跟踪算法综述[J]. 中国图象图形学报, 2019, 24(12): 2057-2080.
- [3] MENG L, YANG X. A survey of object tracking algorithms [J]. Acta Automatica Sinica, 2019, 45(7): 1244-1260.
孟录, 杨旭. 目标跟踪算法综述[J]. 自动化学报, 2019, 45(7): 1244-1260.
- [4] LI Q S, ZHAO L J, ZHANG Z F. Research on target tracking technology considering tracking real-time and tracking precision in complex background[J]. Journal of Optoelectronics • Laser, 2020, 31(2): 117-124.
李庆生, 赵丽君, 张志锋. 复杂背景下兼顾跟踪实时性和跟踪精度的目标跟踪技术研究[J]. 光电子 • 激光, 2020, 31(2): 117-124.
- [5] LIN Z X, ZHENG X N, WU C M, et al. A adaptive scale KCF tracking algorithm with anti-fuzzy properties [J]. Journal of Optoelectronics • Laser, 2020, 31(11): 1166-1174.
林湛黔, 郑兴宁, 吴成茂, 等. 自适应尺度抗模糊核相关滤波跟踪法[J]. 光电子 • 激光, 2020, 31(11): 1166-1174.
- [6] BERTINETTO L, VALMADRE J, GOLLODETZ S, et al. Staple: complementary learners for real-time tracking [C]//IEEE Computer Vision & Pattern Recognition, June 26-July 1, 2016, Las Vegas, Nevada, USA. New York: IEEE, 2016: 1420-1429.
- [7] HUO Y L, LI M, ZHENG H L. Context-aware target tracking algorithm fused with redetection mechanism[J]. Journal of Optoelectronics • Laser, 2021, 32(9): 992-999.
火元莲, 李明, 郑海亮, 等. 融合重检测机制的上下文感知目标跟踪算法[J]. 光电子 • 激光, 2021, 32(9): 992-999.
- [8] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking [C]//European Conference on Computer Vision (ECCV), October 8-16, 2016, Amsterdam, The Netherlands. Berlin, Heidelberg: Springer-Verlag, 2016: 850-865.
- [9] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [10] LI B, YAN J, WU W, et al. High performance visual track-

- ing with siamese region proposal network [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 19-21, 2018, Salt Lake City, Utah, USA. New York: IEEE, 2018: 8971-8980.
- [11] 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 (CVPR), June 16-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 4282-4291.
- [12] ZHANG Z, PENG H. Deeper and wider siamese networks for real-time visual tracking [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 16-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 4591-4600.
- [13] GUO D, WANG J, CUI Y, et al. SiamCAR: siamese fully convolutional classification and regression for visual tracking [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 16-18, 2020, Virtual. New York: IEEE, 2020: 6269-6277.
- [14] ZHANG Z, PENG H, FU J, et al. Ocean: object-aware anchor-free tracking [C]//Computer Vision-ECCV 2020: 16th European Conference, August 23-28, 2020, Glasgow, UK. Berlin, Heidelberg: Springer-Verlag, 2020: 771-787.
- [15] DANELLJAN M, BHAT G, KHAN F S, et al. Atom: accurate tracking by overlap maximization [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 16-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 4660-4669.
- [16] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection [C]//European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. Berlin, Heidelberg: Springer-Verlag, 2018: 784-799.
- [17] BHAT G, DANELLJAN M, GOOL L V, et al. Learning discriminative model prediction for tracking [C]//IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea. New York: IEEE, 2019: 6182-6191.

作者简介:

薛彦兵 (1979—), 男, 硕士, 副研究员, 硕士生导师, 主要从事计算机视觉、机器学习方面的研究。