

基于核相关滤波器和分层卷积特征的长时间目标跟踪

陈 威¹ 李决龙² 邢建春¹ 杨启亮¹ 周启臻¹

(陆军工程大学国防工程学院 南京 210007)¹ (海军海防工程研究中心 北京 100841)²

摘要 针对长时间目标跟踪中出现的目标形变、尺度变化、目标遮挡以及离开视野等问题,提出一种基于核相关滤波器和分层卷积特征的长时目标跟踪算法。首先,利用预训练的卷积神经网络模型提取分层卷积特征来训练核相关滤波器,进行位置估计。其次,构建目标尺度金字塔,进行尺度估计。最后,为了应对目标遮挡以及离开视野导致跟踪失败的情况,训练一个在线支持向量机进行目标再检测,从而实现长时间目标跟踪。在长时间目标跟踪数据集上的测试结果表明:所提算法的精度分别比其他几种主流跟踪算法 HCF, LCT, DSST, KCF 和 TLD 高出 7%, 15%, 17%, 21% 和 50%。

关键词 核相关滤波器, 分层卷积特征, 支持向量机, 长时目标跟踪

中图分类号 TP391.41 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.09.041

Long-term Object Tracking Based on Kernelized Correlation Filter and Hierarchical Convolution Features

CHEN Wei¹ LI Jue-long² XING Jian-chun¹ YANG Qi-liang¹ ZHOU Qi-zhen¹

(National Defense Engineering College, Army Engineering University of PLA, Nanjing 210007, China)¹

(Research Center of Coastal Defense Engineering, Beijing 100841, China)²

Abstract Aiming at the problems such as deformation, scale variation, target occlusion, and out of sight during long-term object tracking, this paper proposed a long-term object tracking algorithm based on kernelized correlation filter and hierarchical convolution feature. Firstly, the pre-trained convolution neural network is applied to extract the hierarchical convolution feature, so as to train correlation filter and estimate location. Then the target scale pyramid is constructed to estimate scale. In order to prevent tracking failure caused by target occlusion and target leaving the field of vision, an online support vector machine is trained for target re-detection to achieve long-term tracking. Experimental results on long-term object tracking dataset show that the accuracy of the proposed algorithm is 7%, 15%, 17%, 21% and 50% higher than that of HCF, LCT, DSST, KCF and TLD.

Keywords Kernelized correlation filter, Hierarchical convolution features, Support vector machine, Long-term object tracking

如今目标跟踪是计算机视觉领域的研究热点,在行为识别、安防监控和人机交互等方面都起到重要作用。尽管近年来目标跟踪领域的研究取得了很大的进展,但由于目标形变、尺度变化、遮挡以及离开视野等因素的影响,目标跟踪仍然是一个具有挑战性的研究课题。

根据外观统计模型视觉跟踪方法主要分为两大类:生成式跟踪方法^[1]和判别式跟踪方法^[2]。生成式跟踪方法首先学习目标外观模型,再以最小误差搜寻与该模型最相似的区域作为目标。对于判别式跟踪方法而言,跟踪任务变成一个二分类问题,首先利用训练样本学习分类器,然后将目标和背景进行区分。

近年来,已有学者将相关滤波引入到判别式跟踪方法中,取得了不错的效果。Bolme 等^[3]首先提出自适应相关滤波跟踪的概念,并在此基础上提出了基于最小输出平方误差和 (Minimum Output Sum of Squared Error, MOSSE) 的相关滤

波跟踪算法。该算法只需训练一个目标外观模型,通过离散傅里叶变换将复杂的相关运算转换成频域上的点积运算,大幅度提升了运算速度。Henriques 等^[4]提出基于循环结构的检测跟踪算法 (Circulant Structure of tracking-by-detection with Kernels, CSK)。该算法通过循环矩阵对训练样本循环移位构建样本集,利用循环矩阵在傅里叶变换中的性质进行高效采样。之后 Henriques 等^[5]在 CSK 跟踪器中加入梯度方向直方图 (Histogram of Gradient, HOG) 特征进行拓展,提出核相关滤波算法 (Kernelized Correlation Filter, KCF)。以上相关滤波跟踪算法未考虑目标出现尺度变化的情况,当跟踪边界框不能适应目标尺度的变化时,会带来除目标以外的背景信息,使得边界框中的图像块特征不能精确地表征目标。为了解决尺度变化的问题,Danelljan 等^[6]提出判别式多尺度跟踪算法 (Discriminative ScaleSpace Tracker, DSST),通过构建一个尺度金字塔来估计最佳尺度。

收稿日期:2018-08-15 返修日期:2018-12-15 本文受江苏省自然科学基金项目(BK20151451)资助。

陈 威(1994—),男,硕士生,主要研究方向为计算机视觉, E-mail: johnny185@126.com(通信作者);李决龙(1959—),男,硕士,教授,主要研究方向为复杂智能信息系统;邢建春(1964—),男,博士,教授,CCF 会员,主要研究方向为复杂智能信息系统;杨启亮(1975—),男,博士,副教授,CCF 会员,主要研究方向为复杂智能信息系统;周启臻(1993—),男,博士生,CCF 会员,主要研究方向为无线感知与图像处理。

然而,以上相关滤波跟踪算法都是针对短时间内的跟踪,只关注 100~500 帧内是否跟踪准确。而在长时间目标跟踪过程中,当目标遮挡或者离开视野时容易出现跟踪失败的情况。Kalal 等^[7]针对长时间目标跟踪问题提出了基于跟踪-学习-检测(Tracking-Learning-Detection, TLD)的跟踪算法。Ma 等^[8]提出长时间相关滤波算法(Long-term Correlation Tracking, LCT),在相关滤波器的基础上加入在线随机藏检测器,当跟踪失败时进行再检测以实现长时间目标跟踪。但是这些长时间跟踪算法都是采用 HOG 等传统图像特征,难以应对目标出现形状变化等因素带来的影响。

随着深度学习的发展,卷积神经网络已经在目标检测和图像识别领域取得了重大成功。经过大规模的图像识别数据集训练得到的卷积神经网络提取的 CNN 特征比传统的图像特征具有更好的表征能力。DLT(Deep Learning Tracker)算法^[9]在小型数据集上预训练一个神经网络,然后结合粒子滤波来定位目标。这种先预训练模型再离线微调的思想在之后的很多方法中得到应用。Wang 等^[10]提出基于全卷积网络的 FCNT 算法,从 VGG-Net 的 conv4-3 和 conv5-3 提取到的特征分别用来构建两个网络,以获得位置信息并区分干扰因素。Ma 等^[11]利用分层卷积特征从粗粒度到细粒度学习 3 个滤波器,然后生成 3 个响应图,最终响应图为 3 个子响应图的加权和,由最大响应值决定目标位置,该方法取得了不错的跟踪效果。

本文受核相关滤波算法运算速度快以及分层卷积特征在目标形变时仍具有较好的表征能力的启发,提出一种基于核相关滤波器和分层卷积特征的长时间目标跟踪算法。为了有效处理形状变化的情况,本文利用预训练得到的卷积神经网络 VGG-16 模型提取分层卷积特征来训练核相关滤波器,再进行位置估计。从不同层提取到的卷积特征既含空间信息又含语义信息,能够应对目标出现形状变化等因素的影响。为了适应目标尺度的变化,构建目标尺度金字塔来估计最佳尺度。为实现长时间目标跟踪,本文加入再检测环节,训练一个在线支持向量机在跟踪失败时对目标进行再检测。

1 核相关滤波器

核相关滤波器 KCF 的原理是先通过循环矩阵构建样本

集,再采用基于核函数的岭回归方法训练滤波器,最后使用训练好的滤波器对下一帧的候选图像块进行相关计算,响应值最大的位置即目标估计位置。利用循环矩阵在离散傅里叶变换中的性质提高训练速度。同时,核函数能将低维特征映射到高维特征空间,从而提高特征的表征能力。

为了得到核相关滤波器的滤波器参数 w ,首先利用循环矩阵进行密集采样,只需采集一张图片 $z \in N \times M$,并将 z 分别以 n, m 个像素为单位在垂直和水平方向上进行循环移位,从而得到具有循环结构的样本 $p_{n,m}$,其中 $n \in \{0, 1, \dots, N-1\}, m \in \{0, 1, \dots, M-1\}$ 。每个样本对应的回归值 $r(n, m)$ 服从二维高斯分布:

$$r(n, m) = \exp\left[-\frac{(n-N/2)^2 + (m-M/2)^2}{2\sigma^2}\right] \quad (1)$$

然后利用获得的样本对滤波器进行训练,目的是求最小化误差下的滤波器参数 w ,即:

$$w = \arg \min_w \sum_{n,m} |\phi(p_{n,m}) \cdot w - r(n, m)|^2 + \lambda |w|^2 \quad (2)$$

其中, λ 为正正则化系数,用来防止过拟合。 ϕ 表示把训练样本映射到高维特征空间的函数。

利用离散傅里叶变换^[3]来计算式(2),得到:

$$w = \sum_{n,m} \alpha(n, m) \phi(p_{n,m}) \quad (3)$$

其中,系数 α 通过式(4)进行定义:

$$A = F(\alpha) = \frac{F(r)}{F(\phi(p) \cdot \phi(p)) + \lambda} \quad (4)$$

其中, A 为相关滤波系数; F 和 F^{-1} 表示离散傅里叶变换及其逆变换; r 为 $N \times M$ 的矩阵,对应的元素是 $r(n, m)$ 。

训练上一帧目标图像块 x 得到参数 α ,再在当前帧中,于前一帧目标位置处采集相同大小的图像块,根据式(5)计算得到目标响应值。

$$\hat{y} = F^{-1}(F(\alpha) \odot F(\phi(z) \cdot \phi(\hat{x}))) \quad (5)$$

其中, \odot 表示对应元素点乘; \hat{x} 表示学习到的目标模型。根据式(4)和式(5)得到目标在当前帧的位置,即 \hat{y} 取最大响应值时所处的位置。

2 本文算法

本文算法的框架如图 1 所示。

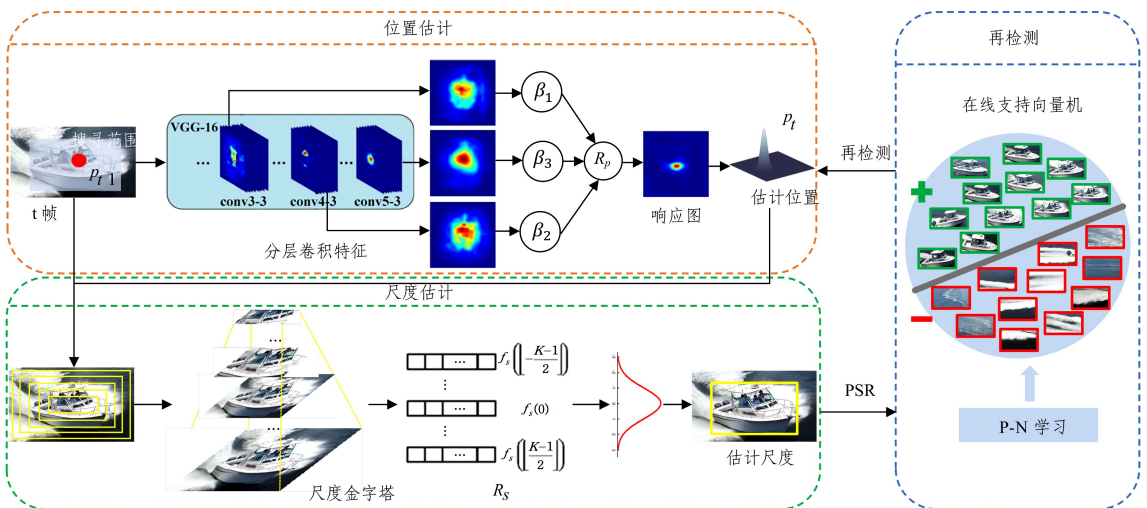


图 1 本文算法框架图

Fig. 1 Framework of the proposed algorithm

首先,通过搜索机制在上一帧中的目标位置附近得到测试样本,利用预训练得到的卷积神经网络 VGG-16 模型在样本中提取分层卷积特征训练核相关滤波器,得到位置模型 R_p 。再利用训练好的模型计算响应值,响应值最大的点即估计目标位置。然后,在估计位置处构建尺度金字塔,提取 HOG 特征训练核相关滤波器,得到尺度模型 R_s ,进行尺度估计。为了实现长时间目标跟踪,加入再检测环节,利用 P-N 学习方法训练一个在线支持向量机,采用峰值旁瓣比 (Peak to Sidelobe Ratio, PSR) 来检测目标估计位置的置信度。当峰值旁瓣比小于设定阈值时,激活在线支持向量机对目标进行再检测。

2.1 目标位置估计

文献[11]指出 VGG-16 卷积神经网络模型中的 3 个卷积层 conv3-4, conv4-4, conv5-4 分别进行卷积特征可视化,结果如图 2 所示。可以看出,当卷积层从原始图片中提取到卷积特征时,越高的卷积层提取到的特征越模糊,其分辨率越低。高层卷积特征含有更多语义信息^[11],可有效地应对目标形状变化的影响,能够实现范围定位。而低层卷积层的特征具有更高的分辨率,保留了更多的空间信息,能够实现更精确的定位。

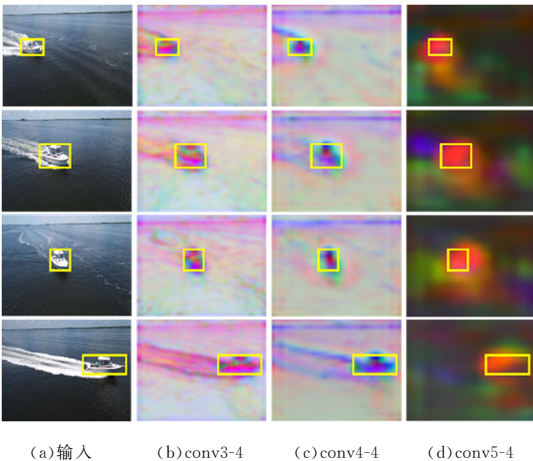


图 2 卷积特征可视化

Fig. 2 Convolution feature visualization

本文利用在大规模图像识别数据集 ImageNet^[12]上训练得到的卷积神经网络 VGG-16 模型^[13],在目标的初始位置提取卷积特征,这些特征分别来自 conv3-4, conv4-4 和 conv5-4 3 个卷积层。为了平滑响应图边界的不连续性,本文引入余弦窗对卷积特征通道进行处理。由于卷积神经网络中存在池化操作,卷积层的分辨率会随着层数的增加而减小,因此利用双线性内插方法把全部特征图都缩放到同一尺寸,再利用缩放后的卷积特征图训练 3 个相关滤波器,得到位置模型 R_p 。然后通过式(5)得到目标响应图,最终的目标响应图 \hat{y}_p 为 3 个子响应图的加权融合,如式(6)所示:

$$\hat{y}_p = \sum_{i=1}^3 \beta_i \hat{y}_i \quad (6)$$

其中, \hat{y}_i 为每个目标响应图, β_i 为每个目标响应图的权重系数。 \hat{y}_p 的最大值所处的位置即对应目标的估计位置。

2.2 目标尺度估计

为了得到尺度响应图,本文在估计位置附近构建目标尺度金字塔。测试帧中目标大小调整为 $W \times H$, K 表示尺度数

量 $s \in S$ 。图 3 为金字塔尺度搜索流程,对于尺度空间 $S = \{s^k | k = -(K-1)/2, -(K-3)/2, \dots, (K-1)/2\}$ 中的每一个尺度,在预测位置中心提取大小为 $sW \times sH$ 的图像块 J_s ,再利用双线性插值方法将 J_s 尺寸调整为 $W \times H$ 的图像块 J_s^* 。然后对 J_s^* 提取 HOG 特征训练核相关滤波器,得到尺度模型 R_s 。最后利用 R_s 计算核相关滤波器对每个图像块 J_s^* 的响应 \hat{y}_j ,则当前目标的最佳尺度为:

$$\hat{S} = \arg \max_j (\max(\hat{y}_1), \max(\hat{y}_2), \dots, \max(\hat{y}_j)) \quad (7)$$

其中, \hat{S} 表示尺度相关响应, $j \in 1, 2, \dots, K$ 表示不同尺度水平。

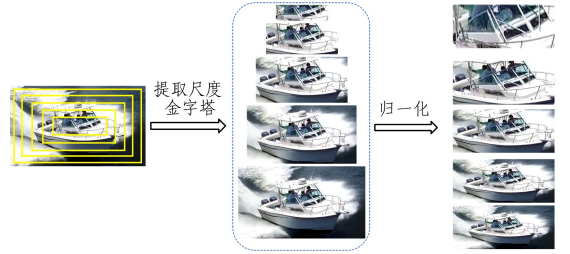


图 3 金字塔尺度搜索流程

Fig. 3 Scale pyramid search process

2.3 再检测环节

当出现目标遮挡或离开视野的情况时,核相关滤波跟踪算法仍旧在原来的位置附近搜索目标。这种方法会使得结果产生较大的误差,而且在训练滤波器时会导致滤波器被背景因素所影响,最终使得分类器的检测能力降低甚至导致目标跟踪失败。为了解决长时间目标跟踪过程中出现目标遮挡或离开视野等情况导致跟踪失败的问题,本文引入再检测环节来重新跟踪目标。

考虑到算法跟踪实时性的要求,本文未使用在每帧图像上都进行训练的核相关滤波器,而是利用由 HOG 特征训练的在线 SVM 分类器重新检测目标。若图像块的 PSR 值低于该激活分类器阈值 T_r ,则激活在线 SVM 分类器生成高分的检测建议区域,然后比较置信度并判断是否采用再检测结果。若图像块的 PSR 值超过训练分类器阈值 T_a ,则在此图像块附近选取样本来训练分类器。对于 SVM 分类器,所有样本图像块大小调整为 $W \times H$ 。

如图 4 所示,本文利用 P-N 学习方法^[7]来训练在线 SVM 分类器,以提高再检测环节的检测性能,具体流程如下:

(1)构建两个训练集作为在线 SVM 分类器的训练样本:标记集 $U_u = \{[x_r, y_r]\}$ 和未标记集 V_v ,其中 $u \ll v$, x_r 为训练样本, y_r 为样本标签。如式(8)所示, SVM 训练样本的标签 y_r 由 IOU 值决定,其中 IOU 值表示样本和目标回归框之间的重叠率。重叠率大于 δ_p 的样本定义为正样本,标签为 1;重叠率小于 δ_n 的样本定义为负样本,标签为 -1。

$$y_i = \begin{cases} +1, & \text{if } IOU > \delta_p \\ -1, & \text{if } IOU < \delta_n \end{cases} \quad (8)$$

(2)利用标记集 U_u 训练 SVM 分类器得到初始化参数 ω_r^0 , SVM 分类器由 $f(x) = \omega_r \cdot x_i + b$ 表示,通过带有 $y \in \{-1, +1\}$ 标签的样本进行训练。目标函数如式(9)所示:

$$\min_{\omega_r, b, \xi} \frac{1}{2} \|\omega_r\|^2 + C \sum_{i=1}^N \xi_i \quad (9)$$

其中, ξ_i 表示松弛因子, C 为常量。

限制条件为:

$$y_i(\omega_r^T \cdot x_i + b) - 1 + \xi_i \geq 0, \xi_i \geq 0, \forall i \in \{1, \dots, N\} \quad (10)$$

其中, 上标 T 表示转置, b 为偏置项。

(3) 将经过上述初始化训练后的在线 SVM 分类器在未标记集 V_u 上进行迭代训练, 迭代 ρ 次后得到分类器参数 ω_r^{ρ} 。利用 P-N 学习方法中的结构性约束, 判断在线 SVM 分类器检测未标记样本的结果是否与结构化约束条件冲突。

(4) P-N 专家找到被错误分类的样本, 并依此对训练样本集进行修正, 使得下一次迭代训练之后得到的分类器的性能有所改善。P 专家根据约束条件将那些被分类器错误标记为“负”的样本赋予“正”的标签, 并添加到训练样本集中; 而 N 专家则根据约束条件将那些被分类器错误标记为“正”的样本赋予“负”的标签, 并添加到训练样本集中。P 专家增加了分类器的鲁棒性, 而 N 专家增加了分类器的判别能力。

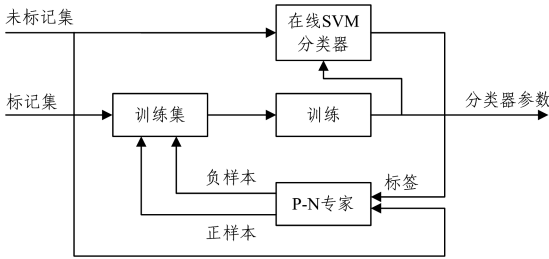


图4 P-N学习流程

Fig. 4 P-N learning process

2.4 模型更新

由于在跟踪过程中, 目标的形状和尺度都会改变, 因此需要对构建的目标模型进行更新。在经典的长时间目标跟踪算法 LCT 中, 当计算新一帧中跟踪结果的置信度时仅考虑当前帧的样本, 会减少整个跟踪结果之间的联系, 出现遮挡时会影响跟踪性能。为了解决这个问题, 我们利用可靠历史样本学习相关滤波器模型, 通过 \hat{y} 的最大值计算自适应在线学习率 δ , 从而更新目标模型。

$$\delta = \delta_0 \times \frac{\max(\hat{y}_t)}{\max(\hat{y}_{t-1})} \quad (11)$$

其中, δ_0 表示初始学习率, t 表示当前帧, $\max(\cdot)$ 表示最大响应值。

为了确保跟踪器的稳定性, 当出现目标遮挡和形状变化等情况时, 在每帧将目标模型和相关滤波系数 A 按照下式更新:

$$\hat{x}_t = (1 - \delta)\hat{x}_{t-1} + x_t \quad (12)$$

$$\hat{A}_t = (1 - \delta)\hat{A}_{t-1} + \hat{A}_t \quad (13)$$

其中, t 表示当前帧, δ 表示自适应学习率。

当目标出现遮挡情况时, 容易出现模型更新错误, 导致跟踪漂移。为了解决这一问题, 本文引入峰值旁瓣比 (Peak to Sidelobe Ratio, PSR)^[14] 来检测目标受目标遮挡等因素影响的程度, PSR 值越大说明目标估计位置的置信度越高, 其定义如式 (14) 所示:

$$PSR(y) = \frac{y_{\max} - \mu_{st}}{\sigma_{st}} \quad (14)$$

其中, y_{\max} , μ_{st} 和 σ_{st} 分别表示响应图的峰值、均值和标准差。

因此本文使用预定义的阈值 $T_{\text{update}} = 10$ 来决定是否进行模型更新, 即当 $PSR(y) < T_{\text{update}}$ 时停止外观模型更新。

2.5 算法流程

本文算法的具体流程如算法 1 所示。

算法 1

输入: 初始目标框 x_0

输出: 估计目标位置 $x_t = (x_t, y_t, \hat{S})$, 位置模型 R_p , 尺度模型 R_s , 在线 SVM 分类器

过程:

Repeat

// 位置估计

根据 $(\hat{x}_{t-1}, \hat{y}_{t-1})$ 在 t 帧时裁剪搜索窗, 利用 VGG-16 的 conv3-4、conv4-4 和 conv5-4 提取分层卷积特征;

利用 R_p 和式 (6) 计算相关图 y_p , 估计目标新位置 (x_t, y_t) ;

// 尺度估计

在 (x_t, y_t) 周围提取 HOG 特征, 建立目标尺度金字塔并利用 R_s 和式 (5) 计算相关图 y_s ;

利用式 (7) 估计最优尺度 \hat{S} , $x_t = (x_t, y_t, \hat{S})$;

// 目标再检测

if $\max(\hat{y}) < T_r$ then

 利用在线 SVM 分类器进行再检测并找到置信度最高的候选框 X ;

 for X 中的 x_i' do

 利用 R_p 和式 (5) 计算置信度 y_i'

 end

 if $\max(\hat{y}_i') > T_a$ then

$x_t = x_i'$, 其中 $i = \text{argmax}_i y_i'$

 end

end

// 模型更新

if $PSR(y) > T_{\text{update}}$ then

 利用式 (9)、式 (10) 更新 R_s 和 R_p ;

end

更新在线 SVM 分类器;

Until 视频序列结束;

3 实验与仿真

本节对所提方法的性能进行全面、综合的评估。本文测试数据集选用长时间目标跟踪数据集 TLP Dataset^[15], 其包含了 50 个视频序列, 每个序列均在 2000 帧以上。数据集涵盖了视觉跟踪中常见的几大挑战: 形状变化、尺度变化、目标遮挡、离开视野。然后在数据集上将本文算法与其他几种主流的算法进行比较, 主流算法包括 HCF^[11], LCT^[8], DSST^[16], KCF^[5], TLD^[6]。其中 HCF 为深度学习跟踪领域的经典算法, LCT 和 TLD 为长时间跟踪领域的经典算法, DSST 和 KCF 为核相关滤波跟踪领域的经典算法。

3.1 实验环境与参数设置

实验平台配置: Matlab2015b, Intel i7-8700k, 3.7 GHz, 16GB 内存, 使用 MatConvNet toolbox 进行 CNN 卷积特征提取, 通过 NVIDIA 1080 GPU 加速。参数设置: 核相关滤波器中的正则化系数 λ 为 10^{-4} , 高斯核的宽度 σ 为 0.1。分层卷积特征提取中各层的权重系数 $\beta_1, \beta_2, \beta_3$ 分别为 0.5, 1, 0.02。

尺度估计中尺度数量 K 为 21。再检测环节中设定的阈值 T_r 和 T_a 分别为 0.25 和 0.5, δ_p 和 δ_n 分别为 0.9 和 0.3。模型更新中的初始学习率 δ_0 为 0.01, T_{update} 为 10。

3.2 评估方法

为了进行定量分析,本文引入以下指标:成功率图、距离精确度图、LSM(Longest Subsequence Measure)指标和速度。其中重叠率定义为 $S = |R_T \cap R_G| / |R_T \cup R_G|$, R_T 表示跟踪区域, R_G 表示人工标注真实区域。成功率定义为重叠率大于阈值的帧占总帧数的比例。中心距离偏差通过目标中心位置与人工标注真实值的欧氏距离进行计算。距离精确度定义为为中心距离偏差小于某个阈值的帧所占的百分比。LSM 指标计算最长成功跟踪子序列所占帧数占序列总帧数的比例,用来评估长时间跟踪的性能。若子序列 95% 的帧跟踪结果的重叠率均大于 0.5,则该子序列为成功跟踪。

3.3 定量分析

3.3.1 整体性能分析

表 1 展示了本文算法与其他几种算法在测试数据集 TLP Dataset 上的 LSM 指标比较。从实验结果可以看出,本文算法的 LSM 值最高,分别比 HCF, LCT, DSST, KCF 和 TLD 高出 8%, 6%, 10%, 13% 和 11%。本文算法成功跟踪的视频帧序列帧数占总帧数的比例最高,说明本文算法的长时间跟踪性能优于其他几种算法。

表 1 6 种算法的成功率比较

Table 1 Success rate comparison of six algorithms

算法	本文	HCF	LCT	DSST	KCF	TLD
LSM	20	12	14	10	7	9

(单位: %)

此外, TLP Dataset 参考目标跟踪领域主流数据集 OTB100 benchmark 选取每个序列的前 600 帧构建 TinyTLP Dataset, 总计 20s, 更加突出挑战特性对算法的影响。图 5、图 6 分别展示了 6 种算法在测试数据集 TinyTLP Dataset 的不同阈值下的距离精确度和成功率。

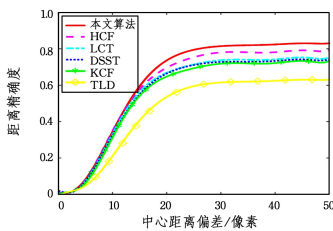


图 5 6 种算法在不同阈值下的距离精确度

Fig. 5 Distance precision of six algorithms under different thresholds

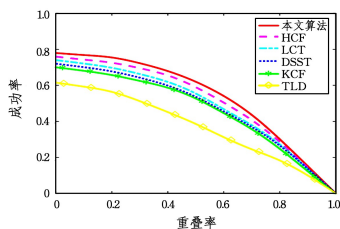


图 6 6 种算法在不同阈值下的成功率

Fig. 6 Success rate of six algorithms under different thresholds

在中心距离偏差阈值为 20 时,所提算法的距离精确度分

别比其他几种主流跟踪算法 HCF, LCT, DSST, KCF 和 TLD 高出 7%, 15%, 17%, 21% 和 50%。在重叠率阈值为 0.4 时,所提算法的成功率分别比其他几种主流跟踪算法 HCF, LCT, DSST, KCF 和 TLD 高出 3%, 4%, 11%, 12.5% 和 31%。

3.3.2 不同挑战特性下的性能分析

TLP Dataset 数据集还针对每一个特性分别挑选出 15 段视频短序列, 每一个视频短序列主要含有其代表的挑战特性, 以此构建 TLPAttrdataset。表 3 展示了在含有不同挑战特性的测试数据集 TLPAttrdataset 上, 本文算法与其他几种算法的成功率比较。从实验结果可以看出, 本文算法在 4 种挑战特性下的成功率均为最高, 说明本文的鲁棒性较好, 能应对不同挑战特性的影响。

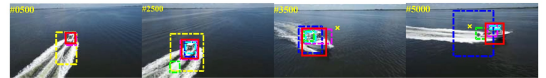
表 2 不同挑战特性下 6 种算法的成功率比较

Table 2 Success rate comparison of six algorithms in different challenge characteristics

成功率 特性	算法					
	本文	HCF	LCT	DSST	KCF	TLD
目标形变	60	57	37	36	35	24
尺度变化	65	43	27	26	24	24
目标遮挡	62	36	55	47	30	51
离开视野	54	35	40	35	32	36

3.4 定性分析

下面结合图 7, 针对视频序列的不同挑战特性对本文算法进行定性分析。



(a) Boat



(b) Drone1



(c) Violinist



(d) Mohiniyattam



(e) Carchasel



(f) Drone2



(g) Hideway

— 本文算法 — — HCF - - - LCT - · - · DSST - - - KCF - - - TLD

图 7 6 种算法的跟踪结果

Fig. 7 Tracking results of six algorithms

3.4.1 形状变化或尺度变化情况下性能分析

从图7可以看出,在 Boat, Violinist 和 Hideway 序列中,目标均出现了不同程度的尺度变化。本文算法、LCT 算法和 DSST 算法采用目标尺度金字塔方法估计尺度,能较好地处理尺度变化的情况。KCF 算法由于使用固定窗口大小的样本训练分类器,导致跟踪边界框大小一直不变,当目标出现尺度变化时,其跟踪效果不佳。

从图7可以看出,在 Boat, Drone1, Violinist 和 Mohiniyattam 序列中,目标均出现了不同程度的形状变化,DSST 算法、KCF 算法和 TLD 算法的跟踪效果均不佳。由于 DSST 算法、KCF 算法和 TLD 算法使用的人工设计的特征不具有语义信息,对目标的表征能力较差,因此当目标出现形状变化时,不能有效地跟踪目标。而本文算法和 HCF 算法引入分层卷积特征,其含有较多语义信息,对目标形状变化的适应性较强,因此在形状变化下仍能取得不错的跟踪效果。

3.4.2 目标遮挡或离开视野情况下的性能分析

从图7还可以看出,Carchase1, Drone2 和 Hideway 序列均出现了目标遮挡或离开视野的情况。在 Carchase1 序列中,第2438帧时目标小车被大卡车遮挡,第2484帧时目标小车重新出现在视野中,本文算法、LCT 算法和 TLD 算法能重新检测到目标,而 HCF 算法、DSST 算法和 KCF 算法的跟踪结果出现漂移。在 Drone2 序列中,第5140帧之后目标将要离开视野,在第5490帧目标重新回到视野时,本文算法能够重新跟踪到目标,其余算法均跟踪失败。在 Hideway 序列中,目标在第2345帧即将受到遮挡,在第2390帧目标重新回到视野时,同样仅有本文算法能够重新跟踪到目标,HCF 算法、DSST 算法和 KCF 算法由于未加入再检测环节不具备抗遮挡能力。而 LCT 算法和 TLD 算法虽然具备再检测环节,但是 LCT 算法再检测环节中的在线随机滤波检测器的性能较弱;TLD 算法的跟踪器基于简单的光流法跟踪目标,性能较差,而检测器根据跟踪器提供的运动信息进行跟踪,因此容易导致跟踪失败。而本文利用 P-N 学习方法训练的在线支持向量机具有较强的检测性能,当跟踪失败时能够重新检测目标。

结束语 针对目标跟踪中出现的形状变化、尺度变化、目标遮挡以及离开视野等问题,本文提出一种基于核相关滤波器和分层卷积特征的长时间目标跟踪方法。首先针对位置估计,本文利用预训练得到的卷积神经网络 VGG-16 模型提取分层卷积特征训练核相关滤波器,然后分别计算不同卷积层的目标响应图,最终响应图为子响应图的加权融合,最大响应值所处位置即估计目标的最终位置。从不同层提取到的卷积特征既含空间信息又含语义信息,能够应对目标形状变化等因素的影响。针对尺度估计,构建目标尺度金字塔来估计目标尺度的变化。为实现长时间目标跟踪,本文加入再检测环节,训练一个在线支持向量机在跟踪失败时对目标进行再检测。在长时间目标跟踪数据集上测试的实验结果表明:与其他几种主流跟踪算法对比,本文算法具有更高的精度和鲁棒性。

参考文献

[1] ORON S, BAR-HILLEL A, LEVI D, et al. Locally orderless

tracking[J]. International Journal of Computer Vision, 2015, 111(2):213-228.

- [2] BABENKO B, YANG M H, BELONGIE S. Robust object tracking with online multiple instance learning[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(8): 1619-1632.
- [3] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2010:2544-2550.
- [4] HENRIQUES J F, RUI C, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]// European Conference on Computer Vision. Springer-Verlag, 2012:702-715.
- [5] HENRIQUES J F, RUI C, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(3): 583-596.
- [6] DANELLJAN M, HÄGER G, KHAN F S. Accurate scale estimation for robust visual tracking[C]// British Machine Vision Conference. 2014:65. 1-65. 11.
- [7] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-Learning-Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(7):1409-22.
- [8] MA C, YANG X, ZHANG C, et al. Long-term correlation tracking[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015:5388-5396.
- [9] WANG N, YEUNG D Y. Learning a deep compact image representation for visual tracking[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc., 2013:809-817.
- [10] WANG L, OUYANG W, WANG X, et al. Visual tracking with fully convolutional networks[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2015:3119-3127.
- [11] MA C, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking[C]// IEEE International Conference on Computer Vision. IEEE, 2016:3074-3082.
- [12] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009:248-255.
- [13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409. 1556, 2014.
- [14] KUMAR B V K V, MAHALANOBIS A, JUDAY R D. Correlation Pattern Recognition [M]. Cambridge University Press, 2005.
- [15] MOUDGIL A, GANDHI V. Long-Term Visual Object Tracking Benchmark[J]. arXiv:1712. 01358, 2017.
- [16] DANELLJAN M, HÄGER G, KHAN F, et al. Accurate scale estimation for robust visual tracking[C]// British Machine Vision Conference. Nottingham: BMVA Press, 2014.