软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2019,30(4):927−940 [doi: 10.13328/j.cnki.jos.005662]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于条件随机场的深度相关滤波目标跟踪算法*

黄树成[1], 张 瑜[2], 张天柱[3], 徐常胜[3], 王 直[1]

[1](江苏科技大学 计算机学院,江苏 镇江 212003)

[2](中国人民解放军 91917 部队,北京 100071)

[3](模式识别国家重点实验室(中国科学院 自动化研究所),北京 100190)

通讯作者: 徐常胜, E-mail: csxu@nlpr.ia.ac.cn

摘 要: 目标跟踪是计算机视觉领域众多应用中的重要组成部分之一.在实际环境中目标经常会因为形变、快速运动、背景杂波和遮挡而引起明显的表观变化,使得该问题具有一定的挑战性,因此如何对跟踪问题进行建模变得至关重要.基于深度卷积神经网络(convolutional neural network,简称 CNN)的判别式相关滤波(discriminative correlation filter,简称 DCF)跟踪方法自提出以来,就以兼顾准确率和速度的优势,吸引了大量研究者的关注,该方法通过相关滤波器获取目标候选区域的响应图,作为衡量目标位置的标准,理想响应图的最大值应该对应目标所在的位置.在此基础上,考虑到响应图中数值的连续性,对应的连续条件随机场(conditional random field,简称 CRF)模型中极大似然对数存在闭式解,因此对响应值的求解可以定义为一个连续 CRF 的学习问题.基于以上研究,提出了一种基于条件随机场的鲁棒性深度相关滤波目标跟踪算法,将 DCF 与 CRF 结合,设计了一个端到端的深度卷积神经网络,嵌入了 CRF 中的一元状态函数与二元转移函数,用来获取图片的响应.通过结合一元状态函数中的初始响应和二元转移函数中的相似度矩阵,优化后的算法可以得到一个更平滑、更精确的响应图,从而提高跟踪的鲁棒性.最后,在 OTB-2013 和 OTB-2015 这两个数据集上进行了大量的测试,并且与近年来 9 种在国际上具有代表性的相关算法进行对比分析,结果显示,在 OTB-2013 中,所提出的算法比基准方法的跟踪成功率高 3%,跟踪精度高 6.1%;在 OTB-2015 中,所提出的算法比基准方法的跟踪成功率高 3.5%,跟踪精度高 4.8%.

关键词: 目标跟踪;卷积神经网络;相关滤波;条件随机场;鲁棒性

中图法分类号: TP391

中文引用格式: 黄树成,张瑜,张天柱,徐常胜,王直.基于条件随机场的深度相关滤波跟踪算法.软件学报,2019,30(4):927−940. http://www.jos.org.cn/1000-9825/5662.htm

英文引用格式: Huang SC, Zhang Y, Zhang TZ, Xu CS, Wang Z. Improved deep correlation filters via conditional random field. Ruan Jian Xue Bao/Journal of Software, 2019,30(4):927−940 (in Chinese). http://www.jos.org.cn/1000-9825/5662.htm

Improved Deep Correlation Filters via Conditional Random Field

HUANG Shu-Cheng[1], ZHANG Yu[2], ZHANG Tian-Zhu[3], XU Chang-Sheng[3], WANG Zhi[1]

[1](School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

[2](Unit 91917 of People's Liberation Army of China, Beijing 100071, China)

[3](State Key Laboratory of Pattern Recognition (Institute of Automation, Chinese Academy of Sciences), Beijing 100190, China)

Abstract: Object tracking is one of the most important tasks in numerous applications of computer vision. It is challenging as target objects often undergo significant appearance changes caused by deformation, abrupt motion, background clutter and occlusion. Therefore, it is important to build a robust object appearance model for visual tracking. Discriminative correlation filters (DCF) with deep

* 基金项目: 国家自然科学基金(61772244)

Foundation item: National Natural Science Foundation of China (61772244)

本文由"多媒体数据的知识关联与理解专题"特约编辑蒋树强研究员、刘青山教授、孙立峰教授、李波教授推荐.

收稿时间: 2018-04-15; 修改时间: 2018-06-13, 2018-09-30; 采用时间: 2018-10-31

convolutional features have achieved favorable performance in recent tracking benchmarks. The object in each frame can be detected by corresponding response map, which means the desired response map should get a highest value at the location of the object. In this scenario, considering the continuous characteristics of the response values, it can be naturally formulated as a continuous conditional random field (CRF) learning problem. Moreover, the integral of the partition function can be calculated in a closed form so that the log-likelihood maximization can be exactly solved. Therefore, here a conditional random field based robust object tracking algorithm is proposed to improve deep correlation filters, and an end-to-end deep convolutional neural network is designed for estimating response maps from input images by integrating the unary and pairwise potentials of continuous CRF into a tracking model. With the combination between the initial response map and similarity matrix which are obtained through the unary and pairwise potentials respectively, a smoother and more accurate response map can be achieved, which improves the tracking robustness. The proposed approach against 9 state-of-the-art trackers on OTB-2013 and OTB-2015 benchmarks are evaluated. The extensive experiments demonstrate that the proposed algorithm is 3% and 3.5% higher than the baseline methods in success plot, and is 6.1% and 4.8% higher than the baseline ones in precision plot on OTB-2013 and OTB-2015 benchmarks respectively.

Key words: object tracking; convolutional neural network; correlation filters; conditional random field; robustness

目标跟踪是计算机视觉领域研究的热点之一,当前广泛应用于视频监控、人机交互等实际问题中,具有重要的研究价值.但是受限于实际环境的复杂性,例如遮挡、光照变化、目标形变以及背景相似干扰等,当前跟踪算法在准确性、鲁棒性以及实时性上还很难满足实际应用需求.因此它仍是一个极具挑战性的课题.

基于相关滤波的跟踪方法以其出色的性能和速度优势,在跟踪领域引起了很大的关注.近年来,涌现出了大量基于相关滤波的跟踪方法,如 KCF[1]、SAMF[2]、LCT[3]、MUSTer[4]和 CACF[5],这些方法多数采用手工特征,因此限制了算法的准确性和鲁棒性.随着 CNN 在目标识别领域中的成功应用,深度学习进入了目标跟踪领域,比较有代表性的基于 CNN 的跟踪方法有 DeepSRDCF[6]、HCF[7]、SiamFC[8]、CFNet[9]和 DCFNet[10]等,但这些算法仅考虑了对图像中目标表观特征的提取,特征响应存在多峰现象,跟踪结果易产生漂移.本文结合连续 CRF 模型提出了一种新的端到端的目标跟踪方法,充分考虑了图像相邻超像素块之间的相似性关系,利用该关系约束了初始响应值,抑制了跟踪过程中的漂移现象.

本文提出的基于条件随机场的深度相关滤波目标跟踪算法,在跟踪一段视频序列时,将其中每一帧图片的目标候选区域通过深度卷积神经网络提取特征,运用相关滤波计算初始响应图,与此同时,根据相邻超像素块之间的位置关系构建相似度矩阵.然后结合当前图片的初始响应与相似度矩阵更新响应图,最终确认目标的位置信息.如图 1 所示,初始响应图中目标位置附近多峰现象严重,本文加入了相邻超像素块之间的相似性关系,利用该关系去约束初始响应图,去除了多峰现象,抑制了跟踪过程中的漂移,使得优化后的响应图更加平滑,提高了判别目标位置的鲁棒性.

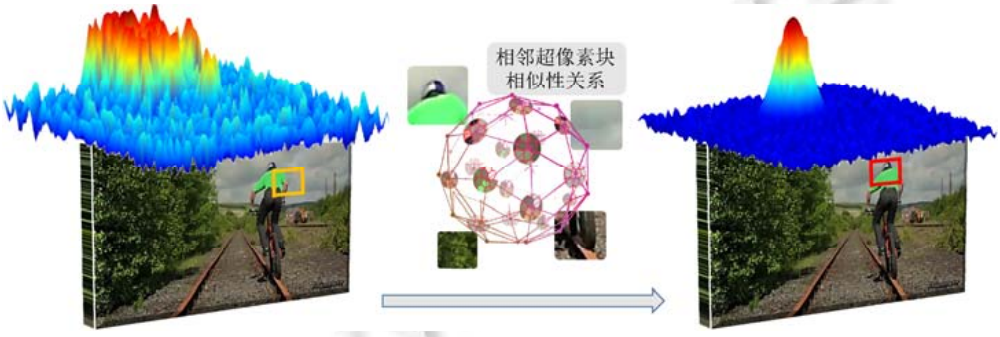


Fig.1 Process of model optimization

图 1 算法模型优化过程

本文的贡献及创新点总结如下.

(1) 将连续 CRF 模型运用到目标跟踪领域中,通过对 CRF 模型中单个势节点和相邻势节点之间的关系建

模,优化了基于深度学习的 DCF 跟踪算法,有效缓解了不同情况下的跟踪漂移问题,特别是在目标尺度变化时,明显提高了跟踪的成功率;

(2) 设计了一个端到端的框架,将 CRF 模型嵌入深度卷积神经网络中,在保持算法实时性的前提下,提升了算法的跟踪精度;

(3) 构建了一个基于图像相邻超像素块之间位置关系的相似度矩阵,在响应图中校正了目标的相对位置,减少了目标周边背景噪声的影响;

(4) 将本文算法在 OTB-2013 的 50 个视频序列和 OTB-2015 的 100 个视频序列[11]上进行了大量的测试,并与 9 种近年来在国际上具有代表性的相关算法进行了对比分析,显著性实验结果验证了算法的有效性、准确性和鲁棒性.

1 相关工作

本文的主要贡献是将CRF模型运用到目标跟踪领域,优化基于深度学习的DCF跟踪算法.涉及到本文相关工作的两个方面主要包括基于 DCF 的跟踪方法以及 CRF 的应用情况.

1.1 基于DCF的目标跟踪方法

基于 DCF 的目标跟踪算法利用快速傅里叶变换进行滤波器的训练和响应图的计算,极大地提高了跟踪效率,具有很好的扩展性.传统的基于 DCF 的算法有 KCF、MOSSE[12]和 CSK[13],随后出现了各种改进算法,包括优化尺度变换的跟踪器 SAMF 和 fDSST[14]、将颜色信息考虑在内的 Staple[15]和 CN[16]、希望跟踪器能长期跟踪的 LCT 和 MUSTer、考虑缓和边界效应的 SRDCF[17]和 CACF,但这些方法多数使用手工特征,因此限制了算法的鲁棒性.

随着 CNN 在图片分类[18,19]、目标检测[20]以及图像分割[21]工作中的迅速发展,目标跟踪领域也开始将深度卷积神经网络作为解决问题的工具之一.越来越多的算法将 DCF 框架和 CNN 结合在一起,例如 HCF 和 HDT[22],提出使用分层卷积特征对跟踪目标进行学习和表达,代替了原来的 HOG 特征.SRDCF 和 DeepSRDCF 改善了 DCF 中存在的边界效应,后者在前者基础上将手工特征替换为 CNN 特征,并说明了在解决跟踪问题时,采取 CNN 的底层特征效果较好,得出解决跟踪问题并不需要太高语义信息的结论.CFNet 和 DCFNet 在 SiamFC 的结构上加入了 CF 层,实现了网络的端到端训练,用实验表明这种网络结构可以用较少的卷积层而不会降低精度.

1.2 CRF的应用

CRF 由 Lafferty 等人[23]提出,结合了最大熵模型和隐马尔可夫模型的特点,是一种无向图模型,近年来在分词、词性标注和命名实体识别等序列标注任务中取得了良好的效果.CRF 模型很少应用在处理连续的回归问题上,最早采用连续 CRF 模型的工作之一是 Qin 等人[24]提出的,应用在文献检索中解决全球排名问题,在一定的约束条件下,可以用确定的规范化因子来优化极大似然函数.在这之后,连续 CRF 模型被成功开发应用在各种结构化回归问题中,例如图像降噪[25]和遥感领域[26],值得一提的是,随着 CNN 的普及应用,Liu 等人[27]成功地将连续 CRF 模型用于图像深度估计,结合深度值的连续性,学习连续 CRF 在 CNN 框架中的势能函数.

到目前为止,还没有将连续 CRF 模型应用到目标跟踪领域的鲁棒性算法,本文提出的目标跟踪模型建立在连续图像响应值上,用连续 CRF 来估计目标候选区域的响应,旨在共同探索 CRF 模型结合 CNN 的跟踪方法在学习目标特征表示时的能力和潜力.

2 基于 CRF 的相关滤波跟踪方法

本节首先对基于 CRF 的相关滤波目标跟踪方法进行简要概述,然后详细介绍本文算法中各个模块的原理和实现方法,最后,介绍改进后的算法流程以及网络的优化过程.

2.1 基于CRF的深度相关滤波目标跟踪概述

目标跟踪的主要目的在于确定目标在视频帧中的位置信息,需要通过相应的目标表观特征描述方法将其中相对稳定的统计特征或某些不变的特征提取出来,一般通过相关滤波器来获取目标候选区域的响应,作为判断目标位置的标准,与背景加以区分.由于视频中的每一帧图像都由若干个像素组成,假设可以将每一张图像分割成若干个超像素块,并且认为图像模型是由很多超像素块所构成.

如图2所示,输入 $x\in\mathbb{R}^{W_0\times H_0}$ 表示图像的目标候选区域,用 $w\times h$ 的核 $(w,h\in\mathbb{Z})$ 对 x 进行池化操作(步长分别为 *stride_w* 和 *stride_h*),得到 $x'\in\mathbb{R}^{W\times H}$, 且 $W=(W_0-w)/stride\_w+1$,$H=(H_0-h)/stride\_h+1$;再经过全连接操作将图片分为 $n=W\times H$ 个超像素块.其中,对应的响应值表示为向量的格式,即 $y=[y_1,...,y_n]^T\in\mathbb{R}^n$.

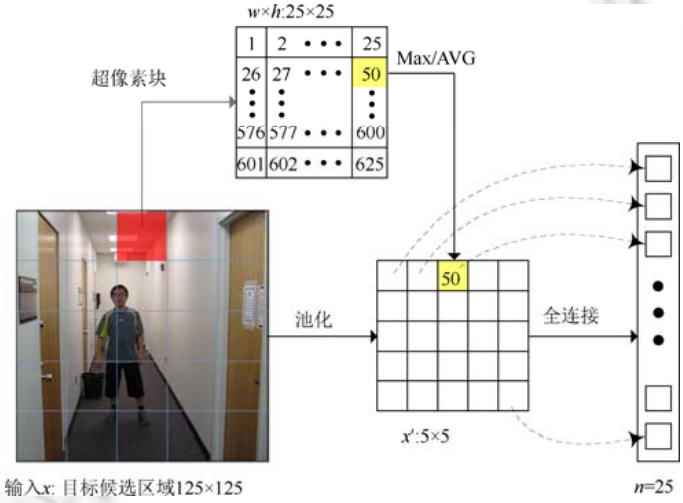


Fig.2 An illustration of the superpixel pooling method

图2 图像超像素化过程

根据传统的CRF参数化模型[28],在随机变量 X 取值为 x 的条件下,随机变量 Y 取值为 y 的条件概率密度函数可以表示为

$$\Pr(y|x)=\frac{1}{Z(x)}\exp\left[\sum_{i,l}\mu_l s_l(y_i,x,i)+\sum_{i,j,k}\lambda_k t_k(y_i,y_j,x,i,j)\right] \tag{1}$$

式中,$i,j=1,2,...,n$,s_l 和 t_k 是特征函数,μ_l 和 λ_k 是对应的权值,求和是在所有可能的输出序列上进行的,$Z(x)$表示规范化因子,本文将特征表示部分用能量函数 $G(y,x)$来表示,因此条件概率密度函数可以写为

$$\Pr(y|x)=\frac{1}{Z(x)}\exp\left(-G(y,x)\right) \tag{2}$$

式(2)中,规范化因子 $Z(x)$表示为

$$Z(x)=\int_y\exp\left\{\sum_{i,l}\mu_l s_l(y_i,x,i)+\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)\right\}dy \tag{3}$$

由于这里的 y 是连续值,与离散情况不同,因此用积分来替换原来的求和运算.为了预测目标位置,可用模型的最大后验概率(maximum a posteriori,简称 MAP)来推断视频帧的响应值,具体表示为

$$y^*=\arg\max_{y\in\mathbb{R}^n}\Pr(y|x) \tag{4}$$

能量函数 $G(y,x)$由一元状态函数 V 和二元转移函数 E 构成,这里,V 对应于图像中 n 个超像素块,旨在回归每个超像素块对应的响应;E 依赖于图像中 S 对相邻超像素块之间的关系,该关系的相似性会起到约束响应值的作用,超像素块间的相似度越高,对应的响应值越接近.这里的函数 V 和函数 E 可以表示为

$$\sum_{i\in n} V(y_i,x) = -\sum_{i,l} \mu_l s_l(y_i,x,i) \tag{5}$$

$$\sum_{(i,j)\in S} E(y_i,y_j,x) = -\sum_{i,j,k} \lambda_k t_k(y_i,y_j,x,i,j) \tag{6}$$

因此,能量函数 $G(y,x)$可以表示为

$$G(y,x) = \sum_{i\in n} V(y_i,x) + \sum_{(i,j)\in S} E(y_i,y_j,x) \tag{7}$$

本文将 V 和 E 嵌入到统一的 CNN 框架中,结合相关滤波建立一个深度网络来得到图片目标候选区的响应值,从而准确地预测目标的位置信息.

2.2 基于CRF和DCF的深度目标跟踪模型

图 3 展示了本文提出的基于 CRF 的深度网络框架,整个网络架构包括一元特征模块、二元关系模块和学习更新模块 3 个部分.一元特征模块实现了在深度网络中对图片目标候选区的特征提取,并且通过相关滤波输出初始响应值;二元关系模块通过网络输出一组一维向量,该向量建立了相邻超像素块之间的相似性关系,用来约束一元特征模块中的初始图片响应;学习更新模块结合一元特征模块中的初始响应值和二元关系模块中的相似度矩阵更新响应图,最终确认目标的位置信息.

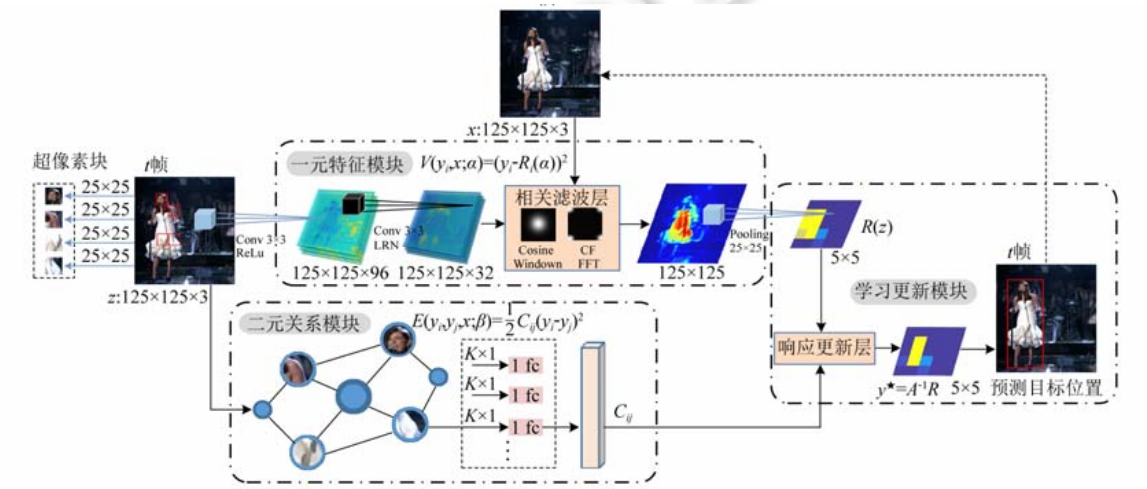


Fig.3 Improved deep correlation filters via conditional random field

图 3 基于 CRF 和 DCF 的深度目标跟踪模型

2.2.1 一元特征模块

本文使用一元状态函数 V 对一元特征模块进行建模,目的是通过训练深度网络获取图像目标候选区域的特征,得到理想的滤波器,输出初始响应.

$$V(y_i,x;\alpha) = \left(y_i - R_i(\alpha)\right)^2, \quad \forall i=1,2,\ldots,n \tag{8}$$

这里,y_i 表示理想状态下的响应,$R_i(\alpha)$表示网络输出的响应,α为网络中的参数.

如第 2.1 节所述,将图像分为 n 个超像素块,每个超像素块的区域 $v_i=\{a_i,b_i,w_i,h_i\},\forall i=1,2,...,n$ 对应 CRF 中的节点信息,其中,(a_i,b_i)表示超像素块的中心位置,(w_i,h_i)表示超像素块的宽度和高度,考虑到对每一个超像素块单独进行卷积运算会存在大量的内存消耗,导致计算效率降低,本文利用 Fast R-CNN[29]的思想,首先对图像整体进行卷积运算,再对结果进行分割,从而提高网络训练的速度和准确率.该模块中的特征提取部分如图 3 所示,主要由两个卷积层、一个修正线性单元(rectified linear unit,简称 ReLU)和一个局部响应归一化(local response normalization,简称 LRN)层构成.假设模块的输入为图片块 $x^0\in R^{125\times125\times3}$. 在第 1 个卷积层中,为了能够更好地利用位置对应信息,本文在 x^0 的外侧进行填补,填补的尺寸为 1.x^0 通过一个权重矩阵 W^1 得到了 96 个特征映射,W^1

包含 96 个子矩阵,即 $W^1=[W_1^1;W_2^1;...;W_{96}^1],\forall W_i^1\in R^{3\times3\times3}$,其中,$W_i^1$ 表示第 1 个卷积层中每个卷积核对应的参数,卷积核的通道数和尺寸分别为 3 和 3×3,采样间隔为 1.因此,输出的 96 个特征映射 $\{x_i^1\}_{i=1}^{96}$ 是通过对 x^0 卷积响应和经过激励函数计算得到的.其中,$x_i^1=f(W_i^1\times x^0+b_i^1)$,选用修正线性单元作为激活函数 $f(\cdot)=\max(0,\cdot)$,“*”代表卷积运算,W_i^1 和 b_i^1 分别表示权重矩阵与偏置项.通过将所有的 x_i^1 连接在一起,可以得到一个完整的特征映射 $x^1\in R^{125\times125\times96}$.在第 2 个卷积层中,为了能更充分地利用输入信息 x_i^1 的外侧进行填补,填补的尺寸为 1.之后,将 x^1 与 32 个卷积核依次进行卷积,对应的权重参数为 $W^2=[W_1^2;W_2^2;...;W_{32}^2],\forall W_i^2\in R^{3\times3\times96}$ 每个卷积核的通道数和尺寸分别为 96 和 3×3,采样间隔为 1,可得 $x_i^2=s(W_i^2\times x^1+b_i^2)$,其中,$s(\cdot)$ 选用局部响应归一化层来实现,使得其中响应比较大的值变得相对更大,并抑制其他反馈较小的神经元,增强了模型的泛化能力.将所有的特征映射 x_i^2 组成 $x^2\in R^{125\times125\times32}$.每跟踪一个新的视频帧,网络就会输出该帧的特征 $\varphi(z)$,再输入到相关滤波层,根据上一帧的特征 $\varphi(x)$来更新一个新的滤波器 w,从而得到当前帧目标候选区域的初始响应图 $R(z)$:

$$R(z)=\sum_{l=1}^{D}\varphi^l(z)\otimes w^l \tag{9}$$

这里,$\varphi^l(z)$ 表示CNN中提取特征的第 l 个通道,“⊗”代表循环矩阵的卷积运算,w^l 表示第 l 个理想的滤波器,可以表示为[30]

$$w^l=\frac{\hat{\varphi}^l(x)\odot\hat{y}^*}{\sum_{i=1}^{D}\hat{\varphi}^i(x)\odot\left(\hat{\varphi}^i(x)\right)^*+\lambda} \tag{10}$$

这里,y 表示目标位置的真实值,$\hat{y}$ 表示离散傅里叶变换 $\hat{y}=F(y)$,*表示变量的复共轭,⊙表示矩阵的哈达玛积.

2.2.2 二元关系模块

本文使用二元转移函数 E 对二元关系模块进行建模,目的是为了通过相邻超像素块之间的相似性关系来平滑视频帧的输出响应.

$$E(y_i,y_j,x;\beta)=\frac{1}{2}C_{ij}(y_i-y_j)^2,\forall i,j=1,2,...,n \tag{11}$$

其中,C_{ij} 表示相邻超像素块 v_i 和 v_j 之间的依赖关系,可以用全连接层来表示:

$$C_{ij}=\beta^T[S_{ij}^{(1)},...,S_{ij}^{(K)}]^T=\sum_{k=1}^{K}\beta_kS_{ij}^{(k)} \tag{12}$$

这里,β是网络参数,$S^{(k)}$表示相邻超像素块之间的第 k 种相似性关系矩阵.可以用位置信息、HOG 特征等建立相邻超像素块之间的相似性关系,本文将这 k 种相似性关系的模型表示为

$$S_{ij}^{(k)}=\gamma\left\|S_i^{(k)}-S_j^{(k)}\right\|_2,k=1,2,...,K \tag{13}$$

其中,$S_i^{(k)}$ 和 $S_j^{(k)}$ 表示相邻超像素块 v_i 和 v_j 对应的特征值,γ是常数,用来调节关系的弹性.

2.2.3 学习更新模块

根据式(8)给出的一元状态函数 V 和式(11)给出的二元转移函数 E 的定义,能量函数 $G(y,x)$可以表示为

$$G(y,x)=\sum_{i\in n}\left(y_i-R_i\right)^2+\sum_{(i,j)\in S}\frac{1}{2}C_{ij}(y_i-y_j)^2 \tag{14}$$

为了便于函数表达和计算,这里定义矩阵 A:

$$A=I+D-C \tag{15}$$

其中,I 表示 $n\times n$ 的单位阵,D 表示由 $D_{jj}=\sum_iC_{ij}$ 组成的度矩阵,它是一个对角阵,C 表示由 C_{ij} 组成的邻接矩阵,$D-C$ 是一个图拉普拉斯矩阵,因此,这里的矩阵 A 是一个正则化拉普拉斯矩阵,能量函数 $G(y,x)$可以替换为

$$\begin{aligned} G(y,x) &= (y^T Iy - 2R^T y + R^T R) + \frac{1}{2}\sum_i \sum_j C_{ij}(y_i^2 - 2y_i y_j + y_j^2) \\ &= y^T Iy - 2R^T y + R^T R + \frac{1}{2}\sum_i D(y_i^2 + y_j^2) - \sum_i \sum_j C_{ij} y_i y_j \\ &= y^T Ay - 2R^T y + R^T R \end{aligned} \tag{16}$$

由于上式中关于y的二次项系数是矩阵A,这里的A是正定阵,结合均值θ服从先验norm分布,且θ~$N(\mu_0,\Sigma_0)$的多维高斯分布的公式为

$$P(\theta) = (2\pi)^{-\frac{n}{2}} \cdot |\Sigma_0|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(\theta-\mu_0)^T \Sigma_0^{\ T}(\theta-\mu_0)\right\} \tag{17}$$

规范化因子$Z(x)$的积分可推算得到:

$$\begin{aligned} Z(x) &= \int_y \exp\{-G(y,x)\}\mathrm{d}y \\ &= \exp(-R^T R)\int_y \exp\{-y^T Ay + 2R^T y\}\mathrm{d}y \\ &= \frac{(\pi)^{\frac{n}{2}}}{|A|^{\frac{1}{2}}}\exp(R^T A^{-1}R - R^T R) \end{aligned} \tag{18}$$

根据式(1)、式(16)和式(18),推出 CRF 的先验概率模型可以表示为

$$\begin{aligned} \Pr(y|x) &= \frac{1}{Z(x)}\exp\{-G(y,x)\} \\ &= \frac{\exp(-y^T Ay + 2R^T y - R^T R)}{\frac{(\pi)^{\frac{n}{2}}}{|A|^{\frac{1}{2}}}\exp(R^T A^{-1}R - R^T R)} \\ &= \frac{|A|^{\frac{1}{2}}}{(\pi)^{\frac{n}{2}}}\exp(-y^T Ay + 2R^T y - R^T A^{-1}R) \end{aligned} \tag{19}$$

这里,$R=[R_1,...,R_n]^T$,是在初始响应的基础上做池化后得到的矩阵,$|\cdot|$表示矩阵的行列式,A^{-1}是A的逆矩阵.因此,根据式(4)可以得到图像中目标候选区域的响应值为

$$\begin{aligned} y^{\bigstar} &= \arg\max_y \log \Pr(y|x) \\ &= \arg\max_y -y^T Ay + 2R^T y \end{aligned} \tag{20}$$

根据式(15)中对矩阵A的定义,矩阵A是对称阵,即$A^T=A$,令:

$$\left.\begin{aligned} &\frac{\partial(-y^T Ay + 2R^T y)}{\partial y} = 0 \\ &\Rightarrow -(A+A^T)y + 2R = 0 \\ &\Rightarrow y = A^{-1}R \end{aligned}\right\} \tag{21}$$

因此,式(20)的闭式解为

$$y^{\bigstar}=A^{-1}R \tag{22}$$

当不考虑二元关系模块中相邻超像素块之间的关系时,即$C_{ij}=0$时,上式可以表示为$y^{\bigstar}=R$,这是一般的 CNN 回归模型,本文将其作为比较的基准算法,在第 3 节会对比这种模型的实验结果.

本文算法将矩阵A定为位置关系矩阵,根据第 2.2.1 节中对超像素块区域的定义,利用相邻超像素块$v_i=\{a_i, b_i,w_i,h_i\}$和$v_j=\{a_j,b_j,w_j,h_j\}$之间的欧式距离建立的相似性关系C_{ij}和度矩阵D可以表示为

$$C_{ij} = \mathrm{d}(S_i,S_j) = \sqrt{(a_i-a_j)^2 + (b_i-b_j)^2} \tag{23}$$

$$D_{jj}=\sum_i\sqrt{(a_i-a_j)^2+(b_i-b_j)^2} \tag{24}$$

根据式(15)将矩阵 A 带入式(22)可以更新初始响应矩阵,校正目标的位置信息,得到校正后的响应 $y^{\bigstar}$,从而确定目标的位置,具体流程详见算法 1.

算法 1. 基于条件随机场的鲁棒性深度相关滤波目标跟踪算法.

输入:视频帧序列 $F_n=\{I_1,I_2,...,I_n\}$;初始帧目标位置$\{x_1,y_1,w_1,h_1\}$;

输出:预测目标的位置$\{x_t,y_t,w_t,h_t\}$.

01: **FOR** t **from** 1 **to** n **DO**

02: **IF** t=1 **THEN**

03: 根据$\{x_1,y_1,w_1,h_1\}$确定目标位置;

04: **ELSE**

05: 以 I_{t-1} 的目标位置为中心,在 I_t 中剪取目标候选区域图像块 z,利用网络提取该区域的特征 $\varphi(z)$;

06: 根据式(9)使用相关滤波方法计算 z 的响应 $R(z)$;

07: 利用如图 2 所示的方法对 z 做超像素化;

08: 根据式(15)对 z 中相邻超像素块间的位置信息构建相似度矩阵 A;

09: 根据式(22)更新 z 的响应图 $y^{\bigstar}$;

10: 根据 $y^{\bigstar}$更新 I_t 中目标的位置信息;

11: **END IF**

12: **END FOR**

2.3 网络优化

本文将CRF和基于深度学习的DCF跟踪网络相结合,设计了一个新的端到端的网络框架.为了使模型的跟踪效果更优,算法鲁棒性更强,本文使用了基于随机梯度下降的反向传播来优化网络参数.

假设通过网络获取的当前帧响应为 $y^{\bigstar}$,理想的响应值为 $\tilde{y}$,定义损失函数为

$$L(\theta)=\left\|y^{\bigstar}-\tilde{y}\right\|^2+\gamma\|\theta\|^2 \tag{25}$$

其中,θ表示网络中的所有参数,γ用来约束正则化项.

$$\left.\begin{aligned} y^{\bigstar}&=A^{-1}R,\\ R&=\mathcal{F}^{-1}\left(\sum_{l=1}^{D}\hat{w}^{l*}\odot\hat{\varphi}^{l*}(z)\right),\\ w^l&=\frac{\hat{\varphi}^l(x)\odot\hat{y}^*}{\sum_{i=1}^{D}\hat{\varphi}^i(x)\odot\left(\hat{\varphi}^i(x)\right)^*+\lambda}\end{aligned}\right\} \tag{26}$$

根据文献[31],离散傅里叶变换与离散傅里叶逆变换的梯度可用下述公式计算:

$$\hat{y}^{\bigstar}=\mathcal{F}(y^{\bigstar}),\frac{\partial L}{\partial\hat{y}^{\bigstar *}}=\mathcal{F}\left(\frac{\partial L}{\partial y^{\bigstar}}\right),\frac{\partial L}{\partial y^{\bigstar}}=\mathcal{F}^{-1}\left(\frac{\partial L}{\partial\hat{y}^{\bigstar *}}\right) \tag{27}$$

在学习更新模块中,前向传播过程只包含一般矩阵乘法,因此可以计算矩阵的导数:

$$\frac{\partial L}{\partial\hat{R}^*}=\mathcal{F}\left(\frac{\partial L}{\partial R}\right) \tag{28}$$

在二元转移模块中,由于关系矩阵 A 可以通过位置信息直接构建,因此这支网络无需从响应更新层反向传播更新参数.在一元特征模块中,损失函数对响应更新层的偏导数可以表示为

$$\frac{\partial L}{\partial R}=\frac{\partial L}{\partial y^{\bigstar}}\frac{\partial y^{\bigstar}}{\partial R}=(A^{-1})^T\frac{\partial L}{\partial y^{\bigstar}} \tag{29}$$

在该模块中,网络的输入分别为当前帧图片 z 和前一帧图片 x 对应网络的检测分支与学习分支. $\frac{\partial L}{\partial\varphi^l(z)}$ 表示

损失函数对检测分支的偏导数,具体计算如下:

$$\frac{\partial L}{\partial \varphi^l(z)}=\mathcal{F}^{-1}\left(\frac{\partial L}{\partial\left(\hat{\varphi}^l(z)\right)^*}\right) \tag{30}$$

$$\frac{\partial L}{\partial\left(\hat{\varphi}^l(z)\right)^*}=\frac{\partial L}{\partial \hat{R}^*}\frac{\partial \hat{R}^*}{\partial\left(\hat{\varphi}^l(z)\right)^*}=\frac{\partial L}{\partial \hat{R}^*}(\hat{w}^l) \tag{31}$$

$\frac{\partial L}{\partial \varphi^l(x)}$ 表示损失函数对学习分支的偏导数,计算过程如下(其中,$\hat{\varphi}^l(x)$ 和 $\left(\hat{\varphi}^l(x)\right)^*$ 看作两个独立的变量):

$$\frac{\partial L}{\partial \varphi^l(x)}=\mathcal{F}^{-1}\left(\frac{\partial L}{\partial\left(\hat{\varphi}^l(x)\right)^*}+\left(\frac{\partial L}{\partial \hat{\varphi}^l(x)}\right)^*\right) \tag{32}$$

$$\frac{\partial L}{\partial \hat{\varphi}^l(x)}=\frac{\partial L}{\partial \hat{R}^*}\frac{\partial \hat{R}^*}{\partial \hat{\varphi}^l(x)}=\frac{\partial L}{\partial \hat{R}^*}\frac{\left(\hat{\varphi}^l(z)\right)^*\hat{y}^*-\hat{R}^*\left(\hat{\varphi}^l(x)\right)^*}{\sum_{k=1}^{D}\hat{\varphi}^k(x)\left(\hat{\varphi}^k(x)\right)^*+\lambda} \tag{33}$$

$$\frac{\partial L}{\partial\left(\hat{\varphi}^l(x)\right)^*}=\frac{\partial L}{\partial \hat{R}^*}\frac{\partial \hat{R}^*}{\partial\left(\hat{\varphi}^l(x)\right)^*}=\frac{\partial L}{\partial \hat{R}^*}\frac{-\hat{R}^*\hat{\varphi}^l(x)}{\sum_{k=1}^{D}\hat{\varphi}^k(x)\left(\hat{\varphi}^k(x)\right)^*+\lambda} \tag{34}$$

误差经过反向传播到有实值的特征图后,余下的传播过程可以看作是传统的卷积神经网络优化问题,这里不再赘述.由于本文算法中反向传播涉及的运算只是复频域中的哈达玛积以及一般的矩阵乘法,因此可以在大量数据集中进行离线训练,再通过网络模型进行在线跟踪.

3 实验结果及分析

本文算法在 Matlab 2015b 上实现,网络层使用 MatConvNet 工具[32]训练.计算机配置为 Intel i7®Core™-4770CPU@3.40GHz×8,内存为 32GB RAM,显卡为 NVIDIA GeForce GTX Titan X.本文的训练视频来源于 NUS-PRO[33]、TempleColor128[34]和 UAV123[35],大约共有 166 643 帧.对于每个视频,本文选取相邻 t–1 和 t 两帧进行配对,然后将每帧裁剪出以目标位置为中心、1.5 倍 padding 大小的图片块,并统一成 125×125 个像素点.学习率 γ 设为 0.072,目标候选区域 padding 为 2.3.本文利用随机梯度下降的方法更新网络参数,大约训练了 20 epoch.网络测试所选用的数据集为 OTB-2013 和 OTB-2015,其中,训练集与测试集无交叉,包含了各种具有挑战性的场景,如:目标遮挡、光照变化以及目标快速运动等.本文选取了 9 种当前国际上具有代表性的相关算法,分别是 DCFNet 算法、SRDCF 算法、CFNet 算法、KCF 算法、LCT 算法、MEEM 算法、Staple 算法、SAMF 算法、DSST 算法,并将本文算法与这 9 种算法进行了对比实验.其中,DCFNet 算法是基于 CNN 直接回归图像中目标候选区域响应的一般回归算法,本文将其作为对比的基准算法.

3.1 定性评估实验

图 4 给出了 5 种跟踪算法在数据集中 5 个视频序列上的部分跟踪结果,视频按照从左到右、从上到下的顺序分别是 carScale(第 110、167、183、194 帧)、bird2(第 49、60、70、75 帧)、tiger1(第 34、37、39、93 帧)、deer(第 11、26、28、33 帧)、trans(第 37、43、50、53 帧).其中,不同的跟踪算法用不同的颜色表示,红色为本文算法,左上角的数字为当前图像帧数.通过它们在具体视频序列中的表现,对结果进行比对和分析后可以发现,本文提出的算法对目标位置的预测结果是比较理想的.

(1) 快速尺度变化:以“CarScale”为例,目标在跟踪过程中出现了剧烈的尺度变化,虽然给出的 5 种算法都能始终跟踪目标,但是只有本文算法能够很好地适应目标的尺度,随着目标尺度的变化实现理想跟踪.

(2) 目标平面内/外旋转、目标形变:以“bird2”为例,目标在跟踪过程中出现了内外旋转变化,对算法的高度旋转不变性提出了要求,这里,DCFNet、SRDCF 以及 KCF 跟踪结果都有偏差,只有本文提出的算法和 CFNet 方

法能够较好地跟踪目标.

(3) 遮挡、光照变化:以“tiger1”为例,在第 34、37 和 39 帧时,目标被树叶遮挡,跟踪结果中其他算法都出现了不同程度的跟踪漂移,只有本文算法对目标遮挡问题具有较好的鲁棒性,能够始终准确地跟踪目标,在第 93 帧时,背景光照出现了剧烈变化,除了本文提出的算法和 DCFNet 方法能够较好地跟踪目标外,其他算法都偏离了目标.

(4) 运动模糊、低分辨率:以“deer”为例,这段视频跟踪目标的分辨率较低,而且小鹿在跳跃中运动出现了模糊,考验了算法在复杂条件下对目标特征的提取,在跟踪结果中,DCFNet 和 KCF 都出现了跟踪失败,本文算法能够始终鲁棒地跟踪目标,对低分辨率的模糊运动目标具有较好的处理能力.

(5) 快速运动、相似背景:以“trans”为例,目标在快速运动过程中伴随着与目标相似的背景,对算法跟踪的准确性具有很大的挑战,只有本文提出的算法和 DCFNet 方法能够始终准确地跟踪目标,而其他算法都出现了不同程度的漂移现象.

Fig.4 Qualitative comparision of our approach with other four state-of-the-art trackers

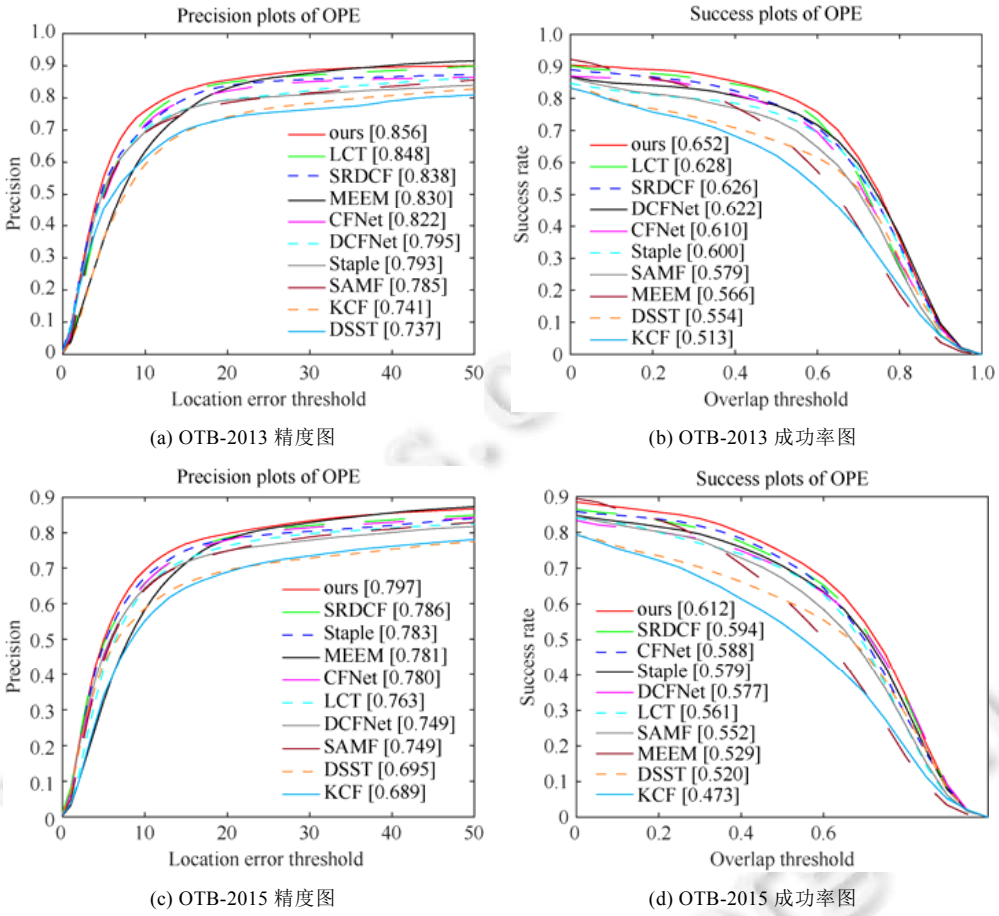
图 4 本文算法与其他 4 种跟踪算法的部分跟踪结果对比

3.2 定量分析结果

为了综合评价算法的跟踪性能,本文采用跟踪精度和跟踪成功率这两个通用的评价指标来进行定量分析.其中,跟踪精度是指当平均中心位置误差小于 20 像素时,算法成功跟踪的帧数与视频总帧数的比值;跟踪成功率是指当覆盖率 *overlap*>0.5 时,算法成功跟踪的帧数与视频总帧数的比值.

在数据集 OTB-2013 下,由图 5(a)和图 5(b)可以得到所列 10 种算法的跟踪精度和跟踪成功率曲线 AUC(area

under curve)值.由图 5(a)可以看出,本文算法的精度最高,达到了 0.856,相比 DCFNet 算法,提高了 6.1%,相比 CFNet 算法,提高了 3.4%;由图 5(b)中的曲线可以看出,本文算法的跟踪成功率最高,AUC 值达到了 0.652,相比 DCFNet 算法,提高了 3%,相比 CFNet 算法,提高了 4.2%.这种跟踪精度和跟踪成功率的明显提升,是因为本文引入了相邻超像素块之间的相似性关系,使得响应图更加平滑,校正了目标位置.



(a) OTB-2013 精度图 (b) OTB-2013 成功率图

(c) OTB-2015 精度图 (d) OTB-2015 成功率图

Fig.5 Precision and success plots showing comparisions with state-of-the-art methods on OTB-2013 and OTB-2015

图 5 在 OTB-2013 和 OTB-2015 数据集下,测试序列的成功率曲线和精度曲线

在数据集 OTB-2015 下,由图 5(c)和图 5(d)中的曲线可以看出,与所列的 9 种其他跟踪算法相比,本文算法的跟踪精度最高,达到了 0.797,相比 DCFNet 算法,提高了 4.8%,相比 CFNet 算法,提高了 1.7%;跟踪成功率也最高,AUC 值达到了 0.612,相比 DCFNet 算法,提高了 3.5%,相比 CFNet 算法,提高了 2.4%,验证了算法的有效性和鲁棒性.

为了进一步分析跟踪算法在不同跟踪条件下的跟踪性能,表 1 和图 6 分别给出了 10 种算法在数据集 OTB-2013 下 11 种不同属性的跟踪结果,包括算法的成功率和跟踪精度.表 1 中红色加粗的数字表示最优结果,蓝色加粗的数字表示次优结果,黑色加粗的数字表示排名第 3 的结果,其中的字母缩写分别表示不同的跟踪条件,分别是:LR(低分辨率)、BC(背景杂波)、OV(目标超出视野)、IPR(平面内旋转)、FM(快速运动)、MB(运动模糊)、DEF(目标形变)、OCC(目标遮挡)、SV(尺度变化)、OPR(平面外旋转)、IV(光照变化).

由表 1 和图 6 可以看出,在 11 种不同属性的跟踪条件中,除了 DEF 属性,本文算法的成功率和跟踪精度在

其他属性中均处于最优或次优的位置,尤其是在 SV 属性下,成功率达到了 0.664,比第 2 名的 DCFNet 高了 4.5%,比第 3 名的 SRDCF 高了 7.7%,展现了本文算法在目标尺度变化上的跟踪优势,同时也表明了对于其他复杂条件下的跟踪,本文算法也具有较好的鲁棒性.

Table 1 Success score of average AUC for each attribute on OTB-2013

表 1 不同属性下算法在 OTB-2013 中的跟踪成功率对比结果

	LR	BC	OV	IPR	FM	MB	DEF	OCC	SV	OPR	IV
ours	**0.474**	**0.613**	**0.68**	**0.606**	**0.597**	**0.582**	0.591	**0.628**	**0.664**	**0.63**	**0.61**
DCFNet	**0.496**	**0.579**	**0.69**	0.572	0.534	0.515	0.606	**0.645**	**0.619**	**0.612**	**0.596**
CFNet	0.434	0.568	0.423	0.565	0.52	0.535	0.581	0.566	0.584	0.583	0.531
SRDCF	0.426	**0.587**	0.555	0.566	**0.569**	**0.601**	**0.635**	**0.627**	**0.587**	0.599	0.576
Staple	**0.438**	0.576	0.547	**0.58**	0.508	**0.541**	0.618	0.593	0.551	0.575	0.568
DSST	0.409	0.517	0.459	0.56	0.435	0.464	0.51	0.534	0.541	0.535	0.563
SAMF	0.388	0.52	0.555	0.525	0.483	0.461	**0.625**	0.612	0.507	0.559	0.513
MEEM	0.36	0.569	**0.606**	0.535	**0.553**	**0.541**	0.56	0.552	0.498	0.558	0.533
KCF	0.31	0.533	0.55	0.497	0.461	0.499	0.533	0.513	0.427	0.496	0.494
LCT	0.286	**0.587**	0.594	**0.592**	0.534	0.524	**0.668**	**0.627**	0.553	**0.624**	**0.588**

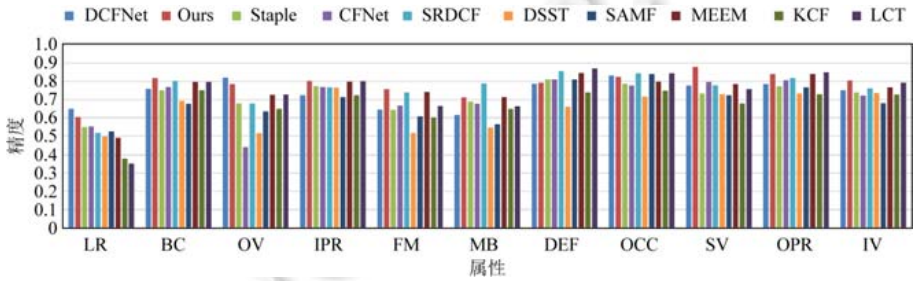


Fig.6 Precision score at 20 pixels for 11 attributes on OTB-2013

图 6 OTB-2013 中 11 种属性下算法的跟踪精度对比结果

3.3 算法跟踪速率评估

表 2 给出了各种算法在 OTB-2013 和 OTB-2015 这两个数据集上的平均视频跟踪速率(单位为 fps).可以发现,本文算法的跟踪速度与 DCFNet 相近,比 CFNet 高约 14 帧/s 左右,这是因为在滤波器的在线更新过程中,通过向量的傅里叶变换和点积运算取代了时域的卷积运算,同时避开了矩阵求逆,将原先矩阵相乘 $O(n^3)$的计算量转换为了傅里叶变换 $n\log(n)$和向量点乘 n,极大地提高了滤波器的训练速度.

总体来说,本文算法在基于深度学习的跟踪算法中,跟踪速率较快,可以实现跟踪的实时性要求.

Table 2 Tracking speed for OTB-2013 and OTB-2015 compared with baseline methods

表 2 各种算法在 OTB-2013 和 OTB-2015 数据集上的平均视频跟踪速度

	Ours	DCFNet	CFNet	SRDCF	KCF	DSST	LCT	MEEM	Staple	SAMF
OTB-2013	89.9	89	75	3.6	245.9	60.5	21.6	20.8	44.90	18.6
OTB-2015	88.3	88	75	3.5	243.4	53.2	20.8	20.7	42.88	16.9

4 结 论

本文将连续 CRF 模型运用到目标跟踪领域中,将一元状态函数与二元转移函数嵌入到深度卷积神经网络中,设计了一个端到端的框架.该算法通过结合一元状态函数得到的初始响应图和二元转移函数得到的相似度矩阵对目标位置进行校正,从而得到了一个更平滑、更精确的响应图,提高了跟踪的精度.本文在 OTB-2013 和 OTB-2015 这两个数据集上进行了大量的测试,在复杂的跟踪条件下,与近年来 9 种在国际上具有代表性的相关算法进行了对比分析,实验结果表明,优化后的算法不仅得到了精度高、鲁棒性好的跟踪结果,同时也较好地解决了跟踪过程中的各类复杂状况,有效地提高了目标跟踪的成功率.

在今后的工作中,我们将会继续研究由相邻超像素块之间依赖关系构建的相似度矩阵,并且进一步优化各类参数,从而校正目标位置的响应,提高算法的鲁棒性.

References:

[1] Henriques JF, Rui C, Martins P, *et al*. High-speed tracking with kernelized correlation filters. IEEE Trans. on Pattern Analysis & Machine Intelligence, 2014,37(3):583–596.

[2] Li Y, Zhu J. A scale adaptive kernel correlation filter tracker with feature integration. In: Proc. of the European Conf. on Computer Vision Workshops. 2014,8926:254–265.

[3] Ma C, Yang X, Zhang C, *et al*. Long-term correlation tracking. In: Proc. of the Computer Vision and Pattern Recognition. 2015. 5388–5396.

[4] Hong Z, Chen Z, Wang C, Mei X, Prokhorov D, Tao D. MUlti-store tracker (MUSTer): A cognitive psychology inspired approach to object tracking. In: Proc. of the 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2015. 749–758. [doi: 10.1109/CVPR.2015.7298675]

[5] Mueller M, Smith N, Ghanem B. Context-aware correlation filter tracking. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. 1387–1395. [doi: 10.1109/CVPR.2017.152]

[6] Danelljan M, Häger G, Khan FS, Felsberg M. Convolutional features for correlation filter based visual tracking. In: Proc. of the IEEE Int'l Conf. on Computer Vision Workshop. IEEE Computer Society, 2015. 621–629.

[7] Ma C, Huang JB, Yang X, Yang MH. Hierarchical convolutional features for visual tracking. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2015. 3074–3082.

[8] Bertinetto L, Valmadre J, Henriques JF, Vedaldi A, Torr PHS. Fully-convolutional siamese networks for object tracking. In: Proc. of the European Conf. on Computer Vision. 2016. 850–865.

[9] Valmadre J, Bertinetto L, Henriques J, Vedaldi A, Torr PHS. End-to-end representation learning for correlation filter based tracking. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. 5000–5008.

[10] Wang Q, Gao J, Xing J, Zhang M, Hu W. DCFNet: Discriminant correlation filters network for visual tracking. 2017. https://www.researchgate.net/publication/316098189_DCFNet_Discriminant_Correlation_Filters_Network_for_Visual_Tracking

[11] Wu Y, Lim J, Yang MH. Object tracking benchmark. IEEE Trans. on Pattern Analysis & Machine Intelligence, 2015,37(9):1834–1848.

[12] Bolme DS, Beveridge JR, Draper BA, Lui YM. Visual object tracking using adaptive correlation filters. In: Proc. of the 23rd IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2010. 2010. 2544–2550.

[13] Henriques JF, Rui C, Martins P, *et al*. Exploiting the circulant structure of tracking-by-detection with kernels. In: Proc. of the European Conf. on Computer Vision. 2012. 702–715.

[14] Danelljan M, Hager G, Khan FS, Felsberg M. Discriminative scale space tracking. IEEE Trans. on Pattern Analysis & Machine Intelligence, 2016,39(8):1561–1575.

[15] Bertinetto L, Valmadre J, Golodetz S, Miksik O, Torr PHS. Staple: Complementary learners for real-time tracking. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition (CVPR). 2016. 1401–1409.

[16] Danelljan M, Khan FS, Felsberg M, Weijer J. Adaptive color attributes for real-time visual tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 1090–1097.

[17] Danelljan M, Häger G, Khan FS, Felsberg M. Learning spatially regularized correlation filters for visual tracking. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2015. 4310–4318.

[18] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proc. of the Int'l Conf. on Neural Information Processing Systems. Curran Associates Inc., 2012. 1097–1105.

[19] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 770–778.

[20] Ren S, Girshick R, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. on Pattern Analysis & Machine Intelligence, 2017,39(6):1137–1149.

[21] Sun W, Wang R. Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM. IEEE Geoscience & Remote Sensing Letters, 2018,PP(99):1–5.

[22] Qi Y, Zhang S, Qin L, Lim J, Yang MH. Hedged deep tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 4303–4311.

[23] Lafferty J, McCallum A, Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proc. of the 18th Int'l Conf. on Machine Learning. 2001. 282–289.

[24] Qin T, Liu TY, Zhang XD, Wang DS, Li H. Global ranking using continuous conditional random fields. In: Proc. of the Conf. on Neural Information Processing Systems. Vancouver, 2008. 1281–1288.

[25] Ristovski K, Radosavljevic V, Vucetic S, Obradovic Z. Continuous conditional random fields for efficient regression in large fully connected graphs. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2013. 840–846.

[26] Radosavljevic V, Vucetic S, Obradovic Z. Continuous conditional random fields for regression in remote sensing. In: Proc. of the European Conf. on Artificial Intelligence. 2010. 809–814.

[27] Liu F, Shen C, Lin G. Deep convolutional neural fields for depth estimation from a single image. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 5162–5170.

[28] Li H. The Method of Statistical Learning. Beijing: Tsinghua University Press, 2012 (in Chinese).

[29] Wang X, Shrivastava A, Gupta A. A-fast-RCNN: Hard positive generation via adversary for object detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 3039–3048.

[30] Danelljan M, Hager G, Khan FS, Felsberg M. Accurate scale estimation for robust visual tracking. In: Proc. of the British Machine Vision Conf. 2014.

[31] Boeddeker C, Hanebrink P, Drude L, Heymann J, Haeb-Umbach R. On the computation of complex-valued gradients with application to statistically optimum beamforming. Technical Report, Department of Communications Engineering, Institute for Electrical Engineering and Information Technology, Paderborn University, 2017.

[32] Hong S, You T, Kwak S, Han B. Online tracking by learning discriminative saliency map with convolutional neural network. In: Proc. of the Int'l Conf. on Machine Learning. 2015. 597–606.

[33] Li A, Lin M, Wu Y, Yang MH, Yan S. NUS-PRO: A new visual tracking challenge. IEEE Trans. on Pattern Analysis & Machine Intelligence, 2016,38(2):335–349.

[34] Liang P, Blasch E, Ling H. Encoding color information for visual tracking: Algorithms and benchmark. IEEE Trans. on Image Processing, 2015,24(12):5630–5644.

[35] Mueller M, Smith N, Ghanem B. A benchmark and simulator for UAV tracking. Far East Journal of Mathematical Sciences, 2013, 2(2):445–461.

附中文参考文献:

[28] 李航.统计学习方法.北京:清华大学出版社,2012.

黄树成(1969－),男,江苏徐州人,博士,教授,主要研究领域为机器学习,多媒体数据分析.

徐常胜(1969－),男,博士,研究员,博士生导师,CCF 杰出会员,主要研究领域为多媒体分析与检索,计算机视觉,模式识别.

张瑜(1991－),女,助理工程师,主要研究领域为计算机视觉,多媒体数据分析,机器学习.

王直(1964－),男,教授,主要研究领域为工业控制,导航控制.

张天柱(1982－),男,博士,副研究员,CCF 专业会员,主要研究领域为模式识别,计算机视觉,多媒体计算,机器学习.