

基于卷积神经网络的鲁棒高精度 目标跟踪算法

李 康, 李亚敏, 胡学敏, 邵 芳
(湖北大学计算机与信息工程学院, 湖北武汉 430062)

摘 要: 目标跟踪是计算机视觉中重要的研究领域之一. 为了跟踪复杂场景中外观变化剧烈的目标, 本文提出了一种基于卷积神经网络的目标跟踪算法. 算法中的网络模型结构包括预训练的特征提取层和自适应更新的分类器层. 在开始跟踪前, 首先训练全连接层和分类器层的参数, 以及目标的特征与位置之间的线性关系. 其次, 定义了评估跟踪结果可信度的标准. 如果得到的跟踪结果的可信度较高, 则根据跟踪结果的特征调整位置, 提高跟踪结果的精确度. 最后, 在训练网络时, 每次迭代都选择分类器得分的最高的负样本参与训练. 该策略可以提高模型的分辨能力. 在 OTB50 测试集中的实验结果表明, 我们的算法取得了良好的跟踪结果.

关键词: 目标跟踪; 神经网络; 计算机视觉; 机器学习

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2018)09-2087-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.09.007

A Robust and Accurate Object Tracking Algorithm Based on Convolutional Neural Network

LI Kang, LI Ya-min, HU Xue-min, SHAO Fang

(School of Computer Science and Information Engineering, Hubei University, Wuhan, Hubei 430062, China)

Abstract: Object tracking is one of the most important area of computer vision. In order to track the object whose appearance changes dramatically in complex scene, we propose a tracking algorithm based on the convolutional neural network. The network of our tracker has two parts: the feature extraction layer and the adaptive classifier layer. At the beginning, we train a fully-connected layer, a softmax layer and the linear relationship between feature and position of these samples. Next, we define a reliability of the tracking result. If the result is reliable, we will adjust the result location according to its features. Finally, in the network training process, we select the negative samples with max classifying scores in each iteration. The strategy could improve distinguishability of our tracker. Experiments on OTB50 show that our tracker could achieve state-of-the-art performance.

Key words: object tracking; neural network; computer vision; machine learning

1 引言

目标跟踪是计算机视觉中重要的研究领域之一. 它研究的是仅给出目标在场景中的初始位置, 设计算法跟踪场景中目标的问题^[1-3]. 目标跟踪在实时监控, 视频分析, 运动捕捉和游戏娱乐等领域有着广泛的应用, 是计算机视觉领域最为活跃的研究方向之一^[4-15]. 近年来, 有大量的跟踪算法被开发了出来, 但是由于跟踪目标信息的不足, 以及跟踪环境和目标变化的复杂

性, 开发出鲁棒的目标跟踪算法仍然是一项具有挑战性的研究课题.

基于检测的目标跟踪算法可以分为特征提取, 外观模型, 运动模型和模型更新步骤. 其中特征提取步骤最为重要, 文献[16]指出如果选择的特征具备很好的分辨能力, 那么仅仅使用简单的外观模型也能够取得较好的跟踪效果. 近年来, 使用卷积神经网络提取的抽象特征^[8]和基于 HOG 特征的目标跟踪算法^[10]取得了良好的跟踪效果.

为了应对跟踪中遇到的遮挡、形变等问题,本研究在基于卷积神经网络提取特征的基础上,设计了一种目标跟踪算法,并结合一系列策略进一步提高跟踪结果的鲁棒性.

2 相关工作

2.1 基于检测的目标跟踪框架

基于检测的目标跟踪是一种常用的目标跟踪框架,它描述了跟踪步骤中的运动模型.本算法在跟踪一帧目标时主要包括以下步骤:

(1) 假设已经得到第 t 帧的目标位置 x_t^* , 在目标位置周围采集正样本, 在远离目标的周围采集负样本, 将这些样本送入分类器进行训练.

(2) 进入第 $t+1$ 帧后, 在第 t 帧目标位置的附近区域内采集若干候选样本.

(3) 将候选样本输入训练的模型中, 模型输出每个候选样本成为目标的概率值, 最后选择概率最大候选样本作为第 $t+1$ 帧的目标. 跟踪系统重复以上步骤直到跟踪完成.

2.2 卷积神经网络

近年来, 卷积神经网络在计算机视觉的各个研究领域展现了优异的性能. 尤其是与目标跟踪紧密相关的目标检测和目标识别方向, 基于卷积神经网络的算法均取得优秀的结果^[17,18]. 文献[19]使用卷积神经网络离线训练模型, 取得了较好的跟踪效果. 但是由于它训练的样本只属于一个类别, 所以只能跟踪特定种类的目标, 例如跟踪行人. CNN SVM^[8]从大规模数据集中预训练出卷积神经网络模型用于提取特征, 然后使

用支持向量机作为分类器对跟踪中的目标和背景进行分类, 达到了先进的跟踪效果. 本研究在 CNN SVM 的基础上, 使用离线训练的卷积神经网络模型作为特征提取层, 在卷积层后加上 softmax 层作为分类器进行训练. 在跟踪过程中, 特征提取层参数保持不变, 而 softmax 层则根据跟踪到的目标自适应更新.

3 基于卷积神经网络的目标跟踪算法

3.1 网络结构

本研究所使用的卷积神经网络结构如图 1 所示, 其中 C1 ~ C4 表示卷积层, P1 ~ P2 表示池化层, 符号“@”前的数字表示该层特征图的层数, “@”后的数字表示特征图的尺寸, f_s^+ 表示网络预测输入样本为目标概率, f_s^- 表示网络预测输入样本为非目标(背景)的概率. 在网络中, C1 ~ C3 层的网络参数使用预训练的 VGG-M 网络层, 这些参数在跟踪过程中保持不变; C4 层和 softmax 层的参数使用跟踪序列的数据训练, 并且在跟踪过程中保持更新以适应目标在跟踪过程中的外观变化. 从实现的角度来看, C1 层至 C3 层可以看做是特征提取层, C4 层和 softmax 层为训练的分类器层. 如图 1 所示, C3 层输出的特征图大小为 3×3 个像素, 特征图的个数是 512 个. 将这些特征图拉伸成 $3 \times 3 \times 512$ 维的向量, 然后通过 ReLU (Rectified Linear Units) 激活函数后送入 C4 层. C4 层为全连接层, 共有 512 个神经元. C4 层到 softmax 层也是全连接的, 其中 softmax 层共有两个输出神经元, 它们输出的值为 $f_s^+(x)$ 和 $f_s^-(x)$. 在输入层中, 所有的样本被归一化为 107×107 大小, 单位为像素.

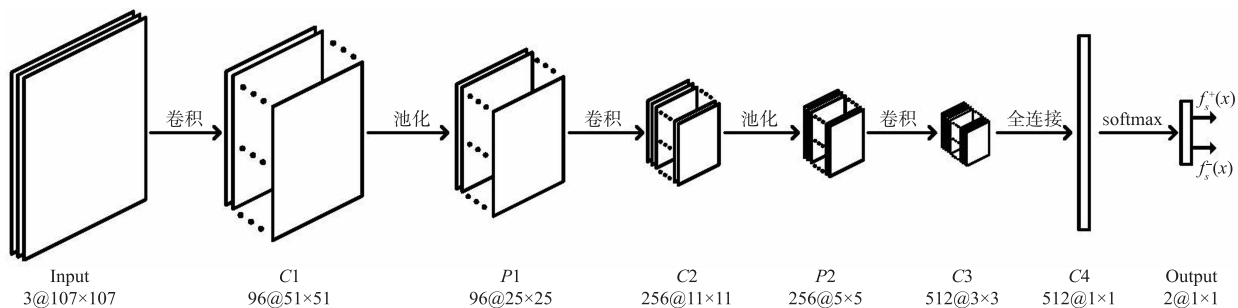


图1 卷积神经网络结构

假设已经完成了对 C4 层和 softmax 层的训练, 则跟踪新的一帧目标的步骤如下: 在跟踪过程进入到第 t 帧时, 根据目标在第 $t-1$ 帧的位置 x_{t-1}^* , 在第 t 帧图像周围采集候选样本集 S_t^o , 则目标在第 t 帧的位置 x_t^* 可以表示为

$$x_t^* = \operatorname{argmax}_{x \in S_t^o} H(x) \quad (1)$$

其中, $H(x) = f_s^+(x)$ 为 softmax 层中预测输入样本为目标概率, 上标 * 号表示该样本是被估计出的目标.

在估计出当前的目标位置后, 我们定义了一个“可信度”的概念来判断该估计出的目标结果是否可信. 由式(1)可以看出, 由于判断目标的标准是根据 softmax 层预测候选样本为目标概率值, 值越大则候选样本为目标概率越高. 假设某一帧估计出的目标为 x^* , 可以推断出 $H(x^*)$ 越大, 则该帧的跟踪结果越可信, 否则跟踪结果很可能在该帧中发生了漂移, 即估计出的目标 x^* 可能是不准确的. 在接下来论文中, 我们定义如

果 $H(x^*) > T_H$, 则认为该帧的目标跟踪结果是可信的, 该帧则被称为可信帧。

3.2 跟踪位置调整

在目标检测中, 在得到检测到的目标位置后, 通常调整估计的目标包围框的大小, 使其更接近真实的目标位置^[18]. 在本研究中, 将目标检测中的包围框调整方法应用到目标跟踪中来, 在得到当前帧的跟踪结果后, 若结果是可信的, 则对其进行微调以获得更精确的目标包围框. 假设模型估计的目标的包围框位置为 P , 真实的目标位置为 G , 当 P 和 G 的空间位置相近时, 它们之间的位置关系与它们的特征之间存在近似的线性关系:

$$T_*(G, P) = w_*^T (\Phi(G) - \Phi(P)) \quad (2)$$

其中, $T_*(G, P)$ 表示 G 与 P 之间的位置关系; w_*^T 为 $T_*(G, P)$ 与 G, P 特征之间的线性关系矩阵; $\Phi(G)$ 和 $\Phi(P)$ 表示从包围框 G 和 P 中提取的特征向量, 下标 $*$ 是 r, v, w 和 h (即包围框的纵坐标, 横坐标, 宽度和高度) 中的一种. 在本研究中, 定义

$$T(G, P) = \begin{cases} T_r(G, P) = (G_r - P_r)/P_w \\ T_v(G, P) = (G_v - P_v)/P_h \\ T_w(G, P) = \log(G_w/P_w) \\ T_h(G, P) = \log(G_h/P_h) \end{cases} \quad (3)$$

为了计算矩阵 w_*^T , 在 G 周围采集 N 个训练样本 $P^i (i=1:N)$, 然后通过最小化等式(4)求得 w_* .

$$w_* = \underset{w_*}{\operatorname{argmin}} \sum_{i=1}^N \left\| T_*^{(i)}(G, P^i) - \hat{w}_*^T (\Phi(G) - \Phi(P^i)) \right\|^2 + \lambda \|\hat{w}_*\|^2 \quad (4)$$

最后, 根据式(5)将预测的结果 P 调整为 \hat{G}

$$\begin{cases} \hat{G}_r = P_w T_r(G, P) + P_r \\ \hat{G}_v = P_h T_v(G, P) + P_v \\ \hat{G}_w = P_w (\exp(T_w(G, P))) \\ \hat{G}_h = P_h (\exp(T_h(G, P))) \end{cases} \quad (5)$$

在本研究中, 根据以下原则使用位置调整策略:

(1) 提取特征的函数 $\Phi(x)$ 使用从样本 x 中提取的 $C3$ 层特征. 这种深度特征在目标检测中常被用来训练位置调整模型.

(2) 仅使用第 1 帧的目标和样本训练 w_*^T , 即令 $G = x_1^*$, 然后根据式(4)训练出 w_*^T . 这是因为第 1 帧给出的目标外观最为完整, 最适合用于训练目标特征与目标位置之间的线性关系.

(3) 假设已估计出目标在第 t 帧的位置 x_t^* , 如该帧的跟踪结果是可信的, 则说明 x_t^* 与目标的真实位置很相近, 只有在这种情况下才使用跟踪位置调整策略.

根据上述理论, 在本研究中, 根据以下步骤训练 w_*^T .

首先, 令 $G = x_1^*$; 其次, 在 G 周围随机采集与 G 重合度大于 0.7 的训练样本 $P^i (i=1:N)$; 最后通过式(4)求解 w_* .

在得到第 t 帧估计出的目标位置 x_t^* 后, 如果该帧的跟踪结果是可信的, 那么则根据式(5)求得 \hat{G} , 最后将 \hat{G} 赋值给 x_t^* 来微调目标位置 \hat{G} .

3.3 在线学习

为了使得训练的模型能够适应目标外观的变化, 需要根据当前跟踪到的目标状态对已训练的模型进行更新. 在更新过程中, 我们遵循以下 3 个原则:

(1) 在更新时, 为了避免过拟合并且保证训练效率, 仅更新 $C4$ 层和 softmax 层参数, $C1$ 至 $C3$ 层参数保持不变.

(2) 在更新过程时, 仅仅对可信的结果进行更新, 这样可以避免使用跟踪不准确的目标训练分类器而造成分类器能力退化的问题. 这种策略提高了跟踪的鲁棒性. 当跟踪目标发生巨大形变时, 跟踪精度可能会降低, 此时不使用这些发生形变的目标更新分类器, 从而保留了分类器的纯洁性. 因此, 在目标从巨大形变中恢复后本算法能够重新定位目标.

(3) 在更新过程中, 采用批量更新法, 以提高更新速度.

3.3.1 训练样本采集

在跟踪开始时, 初始化可信帧集 C 为空集, 初始化正样本集 S^+ 和 S^- 为空集. 在跟踪过程中, 如果 t 时刻的跟踪结果是可信的, 那么就将该帧加入到训练集中: $C \leftarrow C \cup t$. 当 C 中可信帧的数量超过 T_s 帧时, 则采集正样本和负样本对模型进行更新. 当进行更新时, 对于任意 $c \in C$, 首先在 x_c^* 周围随机采集 M_+ 个与 x_c^* 重合度大于 0.7 的样本作为正样本集 S_c^+ , 其次在 x_c^* 周围随机采集 M_- 个与 x_c^* 重合度小于 0.3 的样本作为负样本集 S_c^- , 最后这些正样本集和负样本集加入到训练集中 $S^+ \leftarrow S^+ \cup S_c^+$, $S^- \leftarrow S^- \cup S_c^-$. 则在更新时, S^+ 中包含 $T_s \times M_+$ 个正样本, S^- 中包含 $T_s \times M_-$ 个负样本.

3.3.2 模型更新

为了使模型能够适应目标外观的变化, 需要根据当前跟踪的结果对模型进行更新. 在本研究中, 使用梯度下降法对更新 $C4$ 层和 softmax 层的参数. 在实验中, 每次更新需进行 K 次迭代, 每次迭代从 S^+ 中随机选择 m_+ 个正样本以及从 S^- 中选择当前分类器响应值最高的 m_- 个负样本作为训练样本. 这种选择负样本的方式是基于如下原因: 在跟踪过程中, 若场景中存在与目标外观相似的物体(干扰物), 算法经常漂移到干扰物上. 这是由于干扰物的分类器响应值比目标样本高, 所以在训练过程中, 优先选择干扰物作为负样本进行训练. 这样能够在后续的跟踪中排除干扰物的影响, 提高了

模型的判别能力. 将选择出的正样本和负样本送入模型后,使用梯度下降法更新全连接层(C4层)和 softmax 层的参数,在一次更新完成后,将集合 C, S^+, S^- 置为空集,准备下次训练. 具体的更新步骤如算法 1 所示.

算法 1 模型更新

输入:可信帧的集合 C

1. if ($|C| > T_s$)
2. 对于所有 C 中的帧,采集正样本集 S^+ 和负样本集 S^-
3. for $k=1$ to K
4. 从 S^+ 中随机选择出 m_+ 个正样本
5. 计算负样本集 S^- 中每个元素的分类器响应值 $H(x), x \in S^-$
6. 选择出分类器响应值最高的 m_- 个负样本
7. 根据选出的 m_+ 个正样本和 m_- 个负样本,使用梯度下降法训练 C4 层和 softmax 层的参数
8. 将 $C \leftarrow \emptyset, S^+ \leftarrow \emptyset, S^- \leftarrow \emptyset$

4 实验

4.1 实验配置

本研究在 OTB 标准目标跟踪测试集(共 50 个测试序列)上进行测试,测试视频数据以及人工标定的目标位置可以从文献[3]获得. 本研究实验的基本参数如下:可信度的阈值 $T_u = 0.7$;训练样本 $N = 200$,其中包含 50 个正样本和 150 个负样本;当 C 中元素的数量超过 $T_s = 0.7$ 帧时进行训练;迭代训练次数 $K = 30$;学习率为 0.001. 实验平台为:操作系统为 64 位 Windows10,内存为 16GB,CPU 为 4 核 4.2GHz Intel I7,显卡为 NIV-1070,使用 Matlab 2016b 在 Matconvnet 深度学习库上编程,跟踪速度为每秒 1.2 帧.

为了全面展示本算法的跟踪效果,选取 9 个相关

算法在同样的数据集中进行比较. 这些算法是:CT^[20], MIL^[21], IVT^[22], STRUCK^[9], KCF^[10], CCOT^[5], CNNSVM^[8], FCNT^[6], HDT^[7], 其中 CT, MIL, IVT, STRUCK 是经典算法, CNNSVM, FCNT 是最新的基于深度学习的目标跟踪算法, KCF, CCOT, HDT 是最新的基于相关滤波的算法. 其中 CT, MIL, IVT, STRUCK 的实验数据从文献[3]中获得; KCF, CCOT, CNNSVM, FCNT, HDT 的实验结果数据从它们各自的项目网页中获得.

为了全面地评价算法的性能,对跟踪结果使用两种方式进行评估:

(1) 距离误差成功率. 如果在某一帧,跟踪算法得出的结果与人工标定的跟踪结果距离误差小于某个阈值,那么该帧被认定为跟踪成功.

(2) 重合度成功率. 如果在某一帧,跟踪算法得出的结果与人工标定的跟踪结果重合度大于某个阈值,那么该帧被认定为跟踪成功.

图 2 是跟踪 OTB50 测试视频的成功率示意图,其中,图 2(a)为距离误差成功率示意图,图 2(b)为重合度成功率示意图. 图 2(a)横轴表示距离误差的阈值,纵轴表示距离误差小于该阈值的帧数与总帧数的比值;标题后的数值表示测试视频中包含该跟踪特性的视频数量;算法后面的数字表示曲线与坐标轴围成的面积(Area Under Curve, AUC);OPE(One-Pass Evaluation)为一次性跟踪完整段视频. 距离误差成功率能够反映出跟踪位置的准确度. 图 2(b)表示重合度成功率,其中横轴表示重合度的阈值,纵轴表示重合度大于该阈值的帧数与总帧数的比值. 重合度成功率能够反映算法整体跟踪的精确度. 图 3 列出了部分跟踪结果.

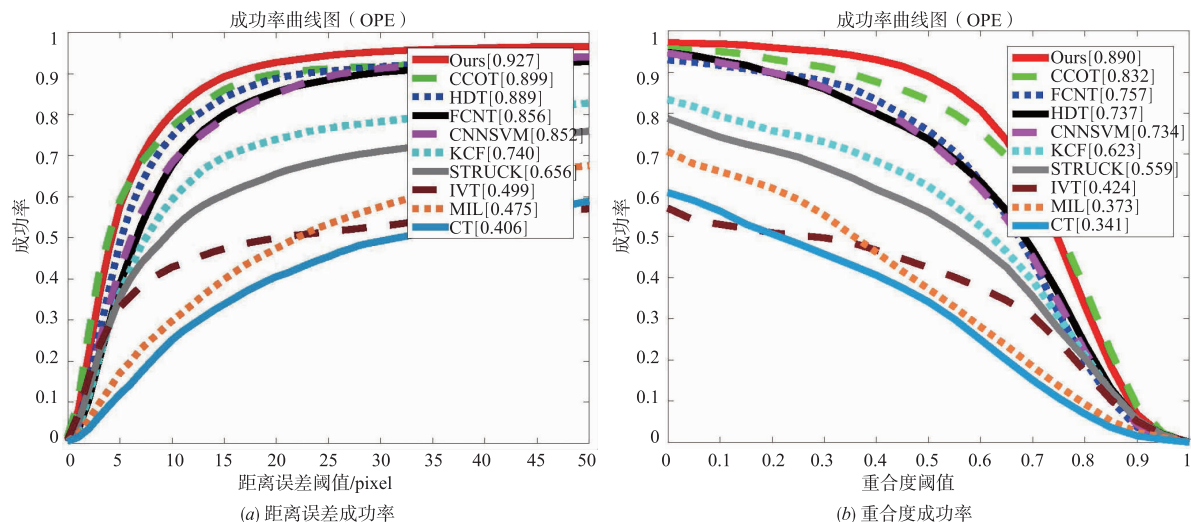


图2 总体成功率

4.2 实验结果分析

从图 2 中可以看出,在整体表现上,本算法(Ours)无论在距离误差成功率上还是在重合度成功率上的 AUC 值都超过了经典跟踪算法,也达到了世界先进水平.下面从 OTB 数据库中取出 4 个视频进行详细分析,它们分别是:Basketball, Bolt, Jumping, Tiger2.

Basketball:本视频中需要跟踪的目标是一名篮球队员.目标的为非刚性物体,并且跟踪区域是非凸的,例如,目标的包围框中包含了较多的背景(头肩部以及两腿之间).这些非目标区域随着目标的移动而不断变化,使得算法在确定目标位置时产生混淆.并且,该测试视频中的目标周围存在多个与目标相似物体,并且在跟踪过程中,目标遭受了其他队员的遮挡.从图 3 中可以看出,在第 401 和 504 帧,CT,STRUCK 和 MIL 发生了漂移.在测试序列中,本算法能够全程准确地跟踪到目标,这主要是因为本文的训练算法能够有效的排除干扰物的影响.

Bolt:本视频中跟踪的是一名短跑运动员,与测试视频 Basketball 类似,该视频同样具有目标非凸,背景干扰等特点.并且在本视频中,目标的移动速度和形变程度都相对较大,并且周围相似的运动员对跟踪造成

了较大的干扰,跟踪难度很大.从图 3 中可以看出,在刚开始的第 15 帧,CT,STRUCK, MIL, IVT 就发生了较大程度的漂移,而 FCNT 虽然全程都能跟踪到目标,但是对目标的尺度跟踪却不准确.在此视频中,我们的算法能够准确的跟踪到目标,说明了本算法在应对剧烈形变,背景干扰等跟踪特性时有较好的效果.

Jumping:本视频中跟踪的目标正在跳绳.跟踪的难点主要在于视频的分辨率较低,并且由于目标剧烈的运动,其外观一直处于模糊状态.如图 3 所示,CT, IVT 和 KCF 在跟踪过程中发生了漂移,其他算法都能够较好的跟踪到目标.这说明本算法能够有效地对跟踪中的运动模糊.

Tiger2:本视频的目标是一只玩具老虎.本视频的跟踪难点在于:第一,目标在测试视频中被树叶多次遮挡(如第 106 帧);第二,目标在移动时,周围的灯光环境发生了变化(第 238 帧);第三,目标发生了巨大的形变(第 301 帧).如图 3 所示,IVT 在第 106 帧发生了漂移,这主要是因为 IVT 使用的底层特征不能够应对目标的光照变化.在第 256 帧,STRUCK 飘到了树叶上.本算法在整个视频中都能够准确的跟踪到目标,说明了本算法应对遮挡和光照变化时的有效性.

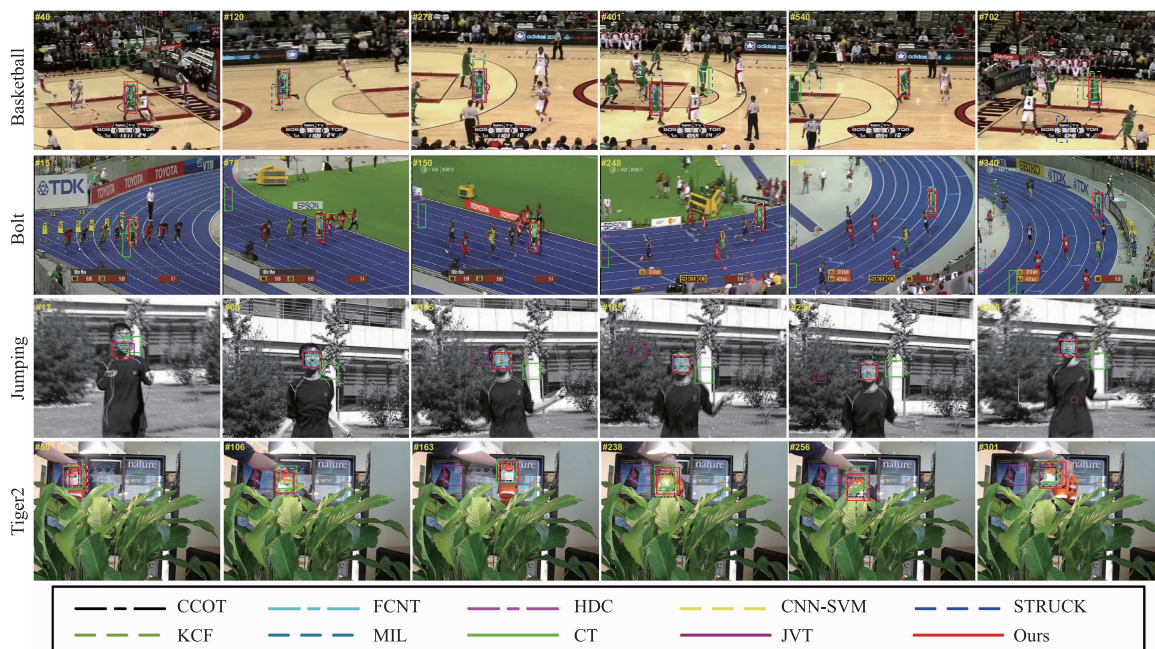


图3 部分跟踪结果

5 总结

针对复杂环境中外观变化较大的目标难以跟踪的问题,本文提出了一种基于卷积神经网络的判别式目标跟踪算法.该算法在跟踪场景中外观变化较大的目标时能够得到鲁棒精确的跟踪结果.实验表

明本文提出的算法与 FCNT 相比在距离误差成功率上提高了 7%,在重合度成功率上提高了 10%.在未来的研究中,我们拟通过提高候选样本覆盖目标的可能性来减少采集的候选样本的数量,从而提高跟踪速度.

参考文献

- [1] Yilmaz A, Javed O, Shah M. Object tracking: A survey [J]. *ACM Computing Surveys*, 2006, 38(4):13.
- [2] Yang H, Shao L, Zheng F, et al. Recent advances and trends in visual tracking: A review [J]. *Neurocomputing*, 2011, 74(18):3823–3831.
- [3] Wu Y, Lim J, Yang M H. Online object tracking: A benchmark [A]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* [C]. Portland: IEEE Press, 2013. 2411–2418.
- [4] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking [A]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* [C]. Las Vegas: IEEE Press, 2016. 4293–4302.
- [5] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: Learning continuous convolution operators for visual tracking [A]. *European Conference on Computer Vision* [C]. Amsterdam: Springer International Publishing, 2016. 472–488.
- [6] Wang L, Ouyang W, Wang X, et al. Visual tracking with fully convolutional networks [A]. *Proceedings of the IEEE International Conference on Computer Vision* [C]. Santiago: IEEE Press, 2015. 3119–3127.
- [7] Qi Y, Zhang S, Qin L, et al. Hedged deep tracking [A]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* [C]. Zurich: IEEE Press, 2016. 4303–4311.
- [8] Hong S, You T, Kwak S, et al. Online tracking by learning discriminative saliency map with convolutional neural network [A]. *International Conference on Machine Learning* [C]. Lille, France: JMLR, 2015. 597–606.
- [9] Hare S, Golodetz S, Saffari A, et al. Struck: Structured output tracking with kernels [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(10):2096–2109.
- [10] 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.
- [11] 李康, 何发智, 潘一腾, 等. 基于簇相似的多分类器目标跟踪算法 [J]. *电子学报*, 2016, 44(4):821–825.
LI Kang, HE Fa-zhi, PAN Yi-teng, SUN Hang. Multi-classifier object tracking based on cluster similarity [J]. *Acta Electronica Sinica*, 2016, 44(4):821–825. (in Chinese)
- [12] 李康, 何发智, 陈晓. 基于 MAP 多子空间增量学习的目标跟踪算法 [J]. *中国科学: 信息科学*, 2016, 4:005.
LI Kang, HE Fa-zhi, CHEN Xiao. Visual tracking algorithm based on MAP multi-subspace incremental learning [J]. *Scientia Sinica Informationis*, 2016, 4:005. (in Chinese)
- [13] Li K, He F, Chen X. Real-time object tracking via compressive feature selection [J]. *Frontiers of Computer Science*, 2016, 10(4):689–701.
- [14] 王治丹, 蒋建国, 齐美彬. 基于最大池图匹配的形变目标跟踪方法 [J]. *电子学报*, 2017, 45(3):704–711.
WANG Zhi-dan, JIANG Jian-guo, QI Mei-bin. Deformable object tracking based on max-pooling graph matching [J]. *Acta Electronica Sinica*, 2017, 45(3):704–711. (in Chinese)
- [15] 宋涛, 李鸥, 刘广怡, 等. 基于改进协作目标外观模型的在线视觉跟踪 [J]. *电子学报*, 2017, 45(2):384–393.
SONG Tao, LI Ou, LIU Guang-yi, CUI Hong-liang. Online visual tracking based on improved collaborative appearance model [J]. *Acta Electronica Sinica*, 2017, 45(2):384–393. (in Chinese)
- [16] Wang N, Shi J, Yeung D Y, et al. Understanding and diagnosing visual tracking systems [A]. *Proceedings of the IEEE International Conference on Computer Vision* [C]. Santiago: IEEE Press, 2015. 3101–3109.
- [17] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [A]. *Advances in Neural Information Processing Systems* [C]. Lake Tahoe: Bartlett, 2012. 1097–1105.
- [18] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [A]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* [C]. Columbus: IEEE Press, 2014. 580–587.
- [19] Fan J, Xu W, Wu Y, et al. Human tracking using convolutional neural networks [J]. *IEEE Transactions on Neural Networks*, 2010, 21(10):1610–1623.
- [20] Zhang K, Zhang L, Yang M H. Real-time compressive tracking [A]. *European Conference on Computer Vision* [C]. Firenze: Springer Berlin Heidelberg, 2012. 864–877.
- [21] Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning [A]. *IEEE Conference on Computer Vision and Pattern Recognition* [C]. Miami: IEEE Press, 2009. 983–990.
- [22] Ross D A, Lim J, Lin R S, et al. Incremental learning for robust visual tracking [J]. *International Journal of Computer Vision*, 2008, 77(1):125–141.

作者简介



李 康 男,1986 年生于安徽亳州,现为湖北大学计算机与信息工程学院讲师,主要研究方法为目标跟踪和机器学习.

E-mail:likang@hubu.edu.cn



李亚敏(通信作者) 女,1990 年生于河北邢台,现为湖北大学计算机与信息工程学院讲师,主要研究方法为图像处理和物联网技术.

E-mail:yamin.li@hubu.edu.cn



胡学敏 男,1985 年出生于湖南岳阳,现为湖北大学计算机与信息工程学院讲师,主要研究方向为图像处理,模式识别和计算机视觉.

E-mail:huxuebin2003@163.com



邵 芳 女,1995 年生于湖北荆门. 现为湖北大学计算机与信息工程学院本科生,主要研究方向为图像处理.

E-mail:notheve@163.com