

基于全卷积对称网络的目标尺度自适应追踪

孙晓霞*, 庞春江

华北电力大学计算机系, 河北 保定 071000

摘要 针对目标追踪过程中由于目标快速运动及尺度变化导致追踪失败的问题,提出了一种基于全卷积对称网络的目标尺度自适应追踪算法。首先利用 MatConvNet 框架构建全卷积对称网络,使用训练好的网络得到实验图像与模板的多维特征图,两者通过互相关操作,选取置信分数最大的点为所追踪目标的中心位置;其次,对中心位置进行多尺度采样,将小于模板方差 1/2 的错误样本过滤掉;建立目标模板和样本概率直方图,计算模板与样本间的海林洛距离,选取合适的尺度作为目标追踪窗口的尺度。在 OTB-13 数据集上进行实验,与其他追踪算法性能比较,本文算法追踪成功率为 0.832,精度为 0.899,高于同类型深度学习追踪算法,平均追踪速度达到 42.3 frame/s,满足实时性的需求;挑选包含目标快速运动或尺度变化属性的追踪序列进一步进行测试,本文算法追踪性能仍高于其他算法。

关键词 机器视觉; 稳健跟踪; 全卷积对称网络; 深度学习; 尺度估计

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.011502

Object Scale Adaptation Tracking Based on Full-Convolutional Siamese Networks

Sun Xiaoxia*, Pang Chunjiang

Department of Computer Science, North China Electric Power University, Baoding, Hebei 071000, China

Abstract Aiming at the problem of tracking failure due to fast motion and scale variation during object tracking, an object scale adaptation tracking based on full-convolutional siamese networks is proposed. First, a full-convolutional symmetric network is constructed using MatConvNet framework, and the multidimensional feature maps of template images and experimental images are obtained by using the trained networks. Through the cross-correlation operation, the point with the highest confidence score is selected as the center of the tracked target. Then, through multi-scale sampling at the center, the error samples that are less than half the template variance are filtered out. The probability histograms of target templates and samples are built. The Hellinger distance between the template and the samples is calculated, and the appropriate scale is selected as the scale of the target tracking window. Experiments on the OTB-13 dataset are carried out. Compared with other tracking algorithms, the tracking success rate of proposed method is 0.832, and the precision is 0.899, which are higher than that of other algorithms, and the average tracking speed is achieved 42.3 frame/s, meeting the needs of real-time object tracking. Selecting the tracking sequences with fast motion or scale change attributes for further testing, the tracking performance of proposed method is still higher than other algorithms.

Key words machine vision; robust tracking; full-convolutional siamese networks; deep learning; scale estimation

OCIS codes 150.1135; 100.4996; 100.2960

1 引言

目标追踪技术^[1]广泛应用于智能交通、人机交互、军事等领域。然而,由于目标的快速移动、实际环境复杂多变、目标尺度变化等原因,在一定程度上

影响了追踪的准确性和稳健性。

近年来,利用深度学习方法解决目标追踪问题取得了良好的追踪效果。Wang 等^[2]首次将深度学习应用在单目标追踪问题上,采用离线预训练结合在线微调的方法,将候选图像输入到分类

收稿日期: 2018-06-07; 修回日期: 2018-07-06; 录用日期: 2018-07-18

* E-mail: 949625607@qq.com

网络中,确定跟踪目标,表现出了极佳的跟踪效果,引起了学者们的广泛关注。此后, Danelljan 等^[3]探究了不同卷积层特征对追踪结果的影响,使用第一层卷积特征训练相关滤波器,在损失函数中引入惩罚项抑制由循环样本产生边界效应。Wang 等^[4]稀疏化不同卷积层的结果特征图,构建特征筛选网络,通过分析得到更有分类性的特征用于追踪,实验结果表明,对目标形变追踪有更加稳健的效果。蔡玉柱等^[5]使用自适应降维技术对 VGG-NET 的 cov3-4 层降维处理,通过分类器获得置信度最大的位置,同时训练 SVM 分类器,在跟踪失败的情况下能够重新检测到目标。毛宁等^[6]利用卷积神经网络提取分层卷积特征,结合相关滤波算法获取卷积特征响应图,并通过响应图的加权融合估计目标位置。上述算法在追踪精度和成功率上相比传统方法都有很大的提升,但是追踪速度往往不能达到实时性的要求;由于没有目标尺度估计机制,往往引入过多的背景信息或不能覆盖目标导致追踪性能变差; Danelljan 等^[7]提出了一种基于一维相关滤波器的尺度评估方法,在目标尺寸变化的情况下仍能取得较好的追踪效果。王鑫等^[8]受此启发使用多层卷积特征进行目标的准确定位,通过一维尺度相关滤波器,确定目标的最佳尺度,取得了较好的追踪效果。

针对目标追踪过程中由于目标快速运动、尺度变化导致追踪失败以及追踪速度较慢的问题,结合全卷积对称网络结构与目标尺度估计方法,提出了一种基于全卷积对称网络的目标尺度自适应追踪算法。

本文的主要研究工作:1) 针对目标快速运动的问题,使用全卷积对称网络确定追踪目标的中心位置,缓解目标的追踪漂移问题;2) 针对目标尺度变化过快问题,在目标中心位置处进行多尺度采样,构建目标尺度金字塔,利用图像方差阈值过滤错误样本,建立目标模型和样本概率直方图,通过计算两者的海林洛距离,选取合适的尺度作为目标追踪窗口的尺度;3) 为验证本文算法的追踪性能,在 OTB-13^[9]数据集上进行测试,并挑选包含尺度变化或快速运动属性的 42 组追踪序列进行实验,与近几年其他 9 种主流追踪算法对比,本文算法在追踪精度和成功率上均超越了其他算法。

2 基本原理

Siamese 网络^[10]接受成对样本的输入,度量样

本间的相似性,图 1 为 Siamese Network 结构图。对称的子网络可根据具体需要选择对应深度学习模型来实现,本文使用的是卷积神经网络。

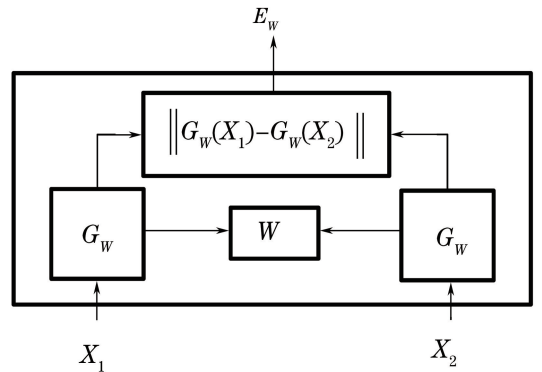


图 1 Siamese 网络结构图

Fig. 1 Siamese network architecture

Siamese 网络有两个结构相同的子网络,分别接收输入 X_1 与 X_2 ,采用权值共享的方式,将其转换为向量 $G_w(X_1)$ 与 $G_w(X_2)$,通过某种度量方式计算 X_1 与 X_2 之间的距离。

训练样本结构为 (X_1, X_2, y) ,其中 y 为该样本的标签, $y = -1$ 表示输入样本对不相似, $y = 1$ 表示相似;训练网络的目的是寻找一组权值 w ,最小化来自相同类别的一对样本的损失值,最大化来自不同类别的一对样本的损失值,假设损失函数与输入和权值有关,则损失函数定义为

$$\mathcal{L}(w) = \sum_{i=1}^p L [w, (X_1, X_2, y)^i], \quad (1)$$

$$L [w, (X_1, X_2, y)^i] = (1 - y)L_G [E_w (X_1, X_2)^i] + yL_1 [E_w (X_1, X_2)^i], \quad (2)$$

式中 $(X_1, X_2, y)^i$ 为第 i 个样本, L_G 是相似类别损失函数, L_1 是样本不相似的损失函数, p 为训练的样本数。

3 本文算法

利用 Siamese 网络能够度量样本间相似度的这一特性,设计了一种全连接对称网络完成目标中心定位;将 Siamese Network 架构中用于表征样本的子网络更换成结构为 5 层卷积 2 层池化层的卷积神经网络(CNN),如表 1 所示。通过最小化损失函数来获取最优模型,判断两张图片相似度最高的位置,实现目标中心定位;同时结合目标尺度自适应算法估计目标的尺寸,达到稳定追踪的目的,算法框架如图 2 所示。

表1 本文算法 CNN 体系结构
Table 1 Architecture of CNN of proposed algorithm

Layer	Support	Chan.map	Stride	For exemplar	For search	Chans
				127×127	255×127	×3
Conv1	11×11	96×3	2	59×59	123×127	×96
Pool1	3×3		2	29×29	61×61	×96
Conv2	5×5	256×48	1	25×25	57×57	×256
Pool2	3×3		2	12×12	28×28	×256
Conv3	3×3	384×256	1	10×10	26×26	×192
Conv4	3×3	384×192	1	8×8	24×24	×192
Conv5	3×3	256×192	1	6×6	22×22	×168

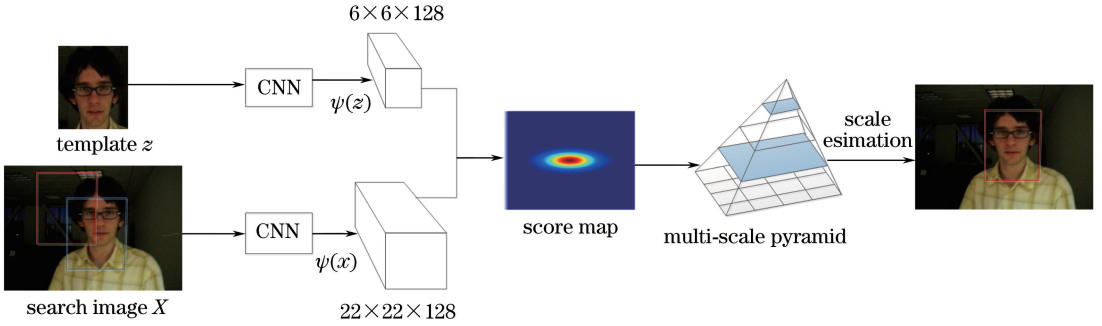


图2 本文算法框架图

Fig. 2 Framework of proposed algorithm

3.1 基于全卷积对称网络的追踪目标中心定位

Siamese 的主要任务就是学习一个相似度匹配函数 $f(x, z)$, 计算图像 x 和模板图像 z 之间的相似度, 模板图像为第一帧追踪目标。若返回的分数值较大, 则两个图像块表示的是相同的目标; 若返回的分数值较小, 则两个图像块表示的是不同的目标。在新的图像中, 通过在整张图片中寻找最大相似度的候选块, 确定目标的中心位置。训练模型时, 通过最小化损失函数获取最优模型, 本文算法为了构造有效的损失函数对搜索区域的位置点进行了正负样本的区分, 即目标一定范围内的点为正样本, 范围外的点定义为负样本, 损失函数定义为

$$l(y, v) = \log_2 [1 + \exp(-yv)], \quad (3)$$

式中 v 为图像 x 的置信分数, 数值越大则表示与模板相似度越高, $y \in \{+1, -1\}$ 为真实标签。通过对称的全卷积网络比较模板与样本图像块的相似程度, 形成一张置信分数图, 将置信分数图的损失定义为整张图像 D 损失的平均值, 即

$$L(y, v) = \frac{1}{|D|} \sum_{u \in D} l(y[u], v[u]), \quad (4)$$

然后利用随机梯度下降 (SGD) 最小化损失函数, 训练网络使用的训练数据集对 (z, x, y) , 训练出最优的网络参数 θ :

$$\arg \min_{\theta} L[y, f(z, x; \theta)], \quad (5)$$

式中 f 为置信分数。

将首帧图像的目标图像块输入到网络中, 以目标图像块大小为基准遍历下一帧图像, 得到候选图像块输入到网络中, 经过子网络的卷积层和池化层, 将原始图像映射到特定的特征空间, 使用卷积嵌入函数 ϕ 和互相关层融合得到每个点的置信分数, 即

$$f(z, x) = \phi(z) * \phi(x), \quad (6)$$

式中 $*$ 为互相关操作, 选取分数最大的点为追踪目标的中心位置 p_i 。需要注意的是在追踪过程中, 使用前一帧目标位置为中心构建搜索区域, 得到 17×17 的置信分数图对其进行双线性插值扩展到 272×272 , 这样响应最大的点才能准确对应到原图像位置。

3.2 目标尺度估计

确定追踪目标中心后, 需要对目标尺度进行估计, 若选取目标尺度不合理, 会影响追踪效果, 严重的情况下会导致追踪失败。所以本文结合多尺度采样方法和相似度匹配原理, 估计目标尺度。

首先在目标中心 p_i 位置处进行多尺度采样, 构建目标尺度金字塔, 如图 3 所示, 尺度选取原则参考文献[7]方法, 即

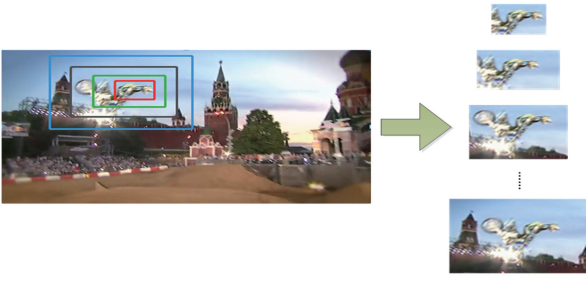


图3 多尺度采样构造金字塔

Fig. 3 Multi-scale sampling construction pyramid

$$a^n W \times a^n H, n \in \left\{ \left[-\frac{s-1}{2} \right], \dots, \left[\frac{s-1}{2} \right] \right\}, \quad (7)$$

式中 W, H 分别为目标模板的宽度和高度, 尺度因子 $a=1.02, n$ 为尺度等级, $s=33$ 。目标尺度变化并不是单纯的线性变化, 这样选取尺寸可以使采样结果在原尺度附近分布比较稠密, 远离原尺度采样分布比较稀疏, 符合在目标追踪过程中目标尺度的变化规律。

其次, 计算图像像素灰度值的方差, 比较每层金字塔图像块的方差与目标模板方差, 保留满足条件的金字塔图像块。公式为

$$s_i^2 = \frac{1}{n} \sum_{j=1}^n \|x_j - m_i\|^2, \quad (8)$$

$$m_i = \frac{1}{n} \sum_{j=1}^n x_j, \quad (9)$$

式中 i 表示某幅图像, n 表示像素数, x_j 表示第 j 个像素的灰度值, m_i 表示第 i 幅图像的灰度平均值,

$$\begin{cases} s_i^2 > \vartheta, & \text{reserve} \\ s_i^2 < \vartheta, & \text{discard} \end{cases}, \quad (10)$$

构建方差分类器, 取模板方差的 $1/2$ 作为阈值 ϑ , 可以最大可能的过滤掉不符合的图像块, 又能极大地保留最符合的图像块; 若不满足条件就舍弃, 通过实验证明至少 30% 的错误图像块可以成功地过滤掉。

在此基础上通过海林格距离度量剩余图像块与目标模板直方图特征的相似度, 选择相似度最大的图像块尺度作为目标最佳尺度。目标模板与剩余图像块直方图概率分布表示为

$$\begin{cases} q = C_0 \sum_{j=1}^m \delta [b(x_j) - u] \\ p_i = C_i \sum_{j=1}^m \delta [b(x_j) - u] \end{cases}, \quad (11)$$

式中 q 表示模板图像的直方图概率分布, p_i 表示第 i 块图像直方图概率分布, $b(x_j)$ 为第 j 个像素对应的灰度直方图映射函数, $u \in \{1, 2, \dots, 255\}$ 为灰度直方图索引, $\delta(\cdot)$ 是一个克罗内克脉冲函数, C_i 为

归一化常数, 则 q 与 p_i 的海林洛距离为

$$D(q, p_i) = \sqrt{1 - \rho[q, p_i]}, \quad (12)$$

式中 $\rho[q, p_i]$ 为两者的巴氏系数, 寻找与模板图像最相似的图像块 i^* 等价于最大化巴氏系数目标函数, 即

$$i^* = \arg \max_i \rho[q, p_i]. \quad (13)$$

4 分析与讨论

实现算法的编码环境 MATLAB R2015, 使用图形处理器(GPU)进行加速运算, 计算机的配置为 windows 操作系统 64 位、Inter Xeon 2.4 GHz、NVIDIA GTX-TITAN X、内存 32 GB。将本算法在 tracker_benchmark_v1.0 下运行以便和其他算法比较。

使用 OTB-13 数据集进行测试分析, 综合评判本文算法的跟踪性能, 同时为了进一步分析算法在面临目标快速运动以及尺度变化问题的跟踪性能, 选取了 OTB-13 中具有目标快速运动或者尺度变化属性的 42 组追踪序列进行实验, 图 4 为部分实验追踪序列, 每组追踪序列都标注了该序列所属属性: 尺度变化(SV)、遮挡(OCC)、形变(DEF)、快速运动(FM)、平面向外旋转(OPR)、超出视野(OV)、光照变化(IV)、背景杂斑(BC)、低分辨率(LR)、平面内旋转(IPR)、运动模糊(MB)等, 并与当前主流的判别相关滤波网络(DCFNet)^[11]、Siamfc^[12]、CF2^[13]、对冲深度跟踪(HDT)、CSK^[14]、核相关滤波(KCF)算法^[15]、SRDCF、加尺度估计的相关滤波器(DSST)^[8]、互补学习跟踪算法(Staple)^[16]、跟踪学习检测(TLD)算法^[17]进行比较; 其中 DCFNet、Siamfc、CF2、HDT 是基于深度学习的跟踪器, CSK、KCF、SRDCF 是基于相关滤波的跟踪器, DSST 是基于尺度评估的跟踪器, Staple 算法结合梯度方向直方图(HOG)特征与颜色特征进行目标追

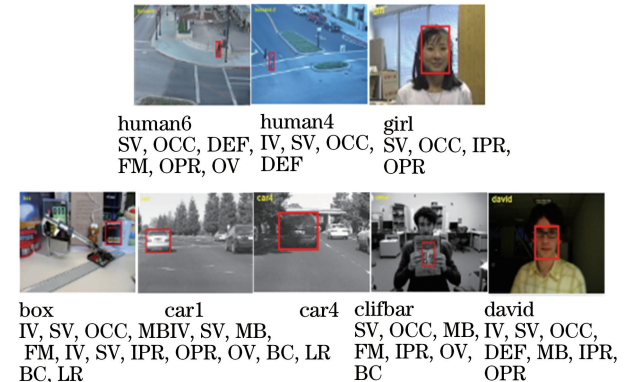


图4 部分实验追踪序列

Fig. 4 Part of the experiment tracking sequences

踪, TLD 是长时间追踪器。

4.1 定性分析

从 FM、SV、OCC、DEF、IV 等属性下分析追踪算法性能, 图 5 展示了 9 组追踪序列结果, 每组追踪序列下方标注了序列名称与属性, 最后一张图片的结果均为最终追踪结果, 红色线条表示本文算法跟踪结果。

1) FM: 以 5(a)、(b) 和 (d) 为例, 目标在水平或垂直方向快速运动过程中本文算法始终能够稳定地追踪, 追踪窗口没有发生偏离。图 5(d) 序列在 #0075 由于快速运动导致目标模糊不清, DCFNet 算法追踪窗口漂移, 最后的追踪结果只有本文追踪算法结果与目标的真实位置与尺度最接近。

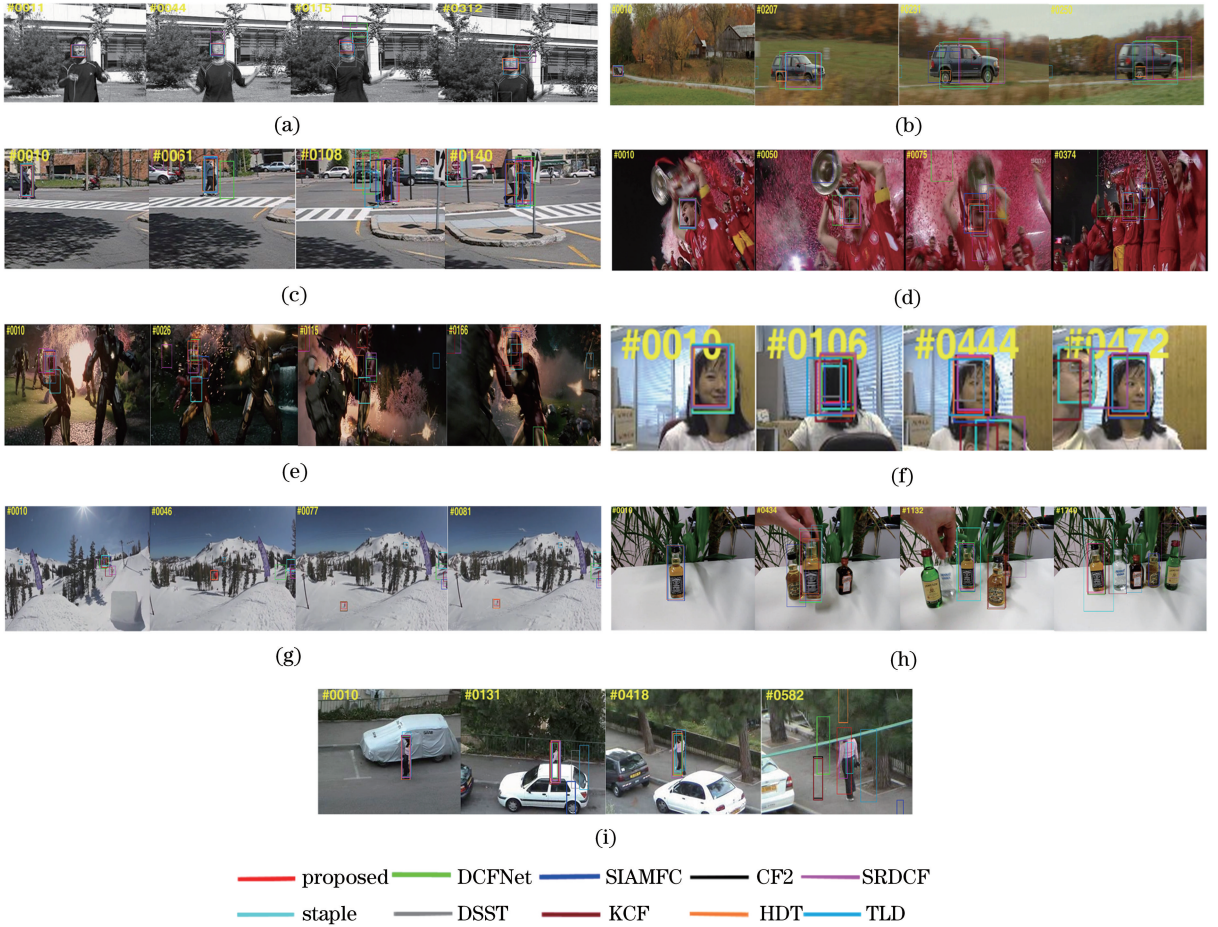


图 5 10 种算法的部分追踪结果对比图。(a) Jumping (MB, FM); (b) CarScale (SV, OCC, FM, IPR, OPR); (c) Couple (SV, DEF, FM, BC, OPR); (d) Soccer (IV, SV, OCC, MB, FM, IPR, OPR, BC); (e) Ironman (IV, SV, OCC, MB, FM, IPR, OPR, OV, BC, LR); (f) Girl (SV, OCC, IPR, OPR); (g) Skiing (IV, SV, DEF, IPR, OPR); (h) Liquor (IV, SV, OCC, FM, IPR, OPR, OV, BC); (i) Woman (IV, SV, OCC, DEF, MB, FM, OPR)

Fig. 5 Comparison of tracking results of 10 algorithm. (a) Jumping (MB, FM); (b) CarScale (SV, OCC, FM, IPR, OPR); (c) Couple (SV, DEF, FM, BC, OPR); (d) Soccer (IV, SV, OCC, MB, FM, IPR, OPR, BC); (e) Ironman (IV, SV, OCC, MB, FM, IPR, OPR, OV, BC, LR); (f) Girl (SV, OCC, IPR, OPR); (g) Skiing (IV, SV, DEF, IPR, OPR); (h) Liquor (IV, SV, OCC, FM, IPR, OPR, OV, BC); (i) Woman (IV, SV, OCC, DEF, MB, FM, OPR)

2) SV: 以图 5(b)、(i)、(d)、(e) 和 (f) 为例, 目标尺度在运动过程中不断变化。图 5(b) 中汽车尺度持续变化, HDT 和 KCF 固定目标尺度追踪, 出现较大误差性能下降, SRDCF、DCFNet、Staples 虽然也能追踪到汽车, 但是算法只追踪到车身, 对后续追踪效果不利, 本文算法能依据汽车的尺度变化追踪到车头位置。

3) OCC: 以图 5(f) 和 (i) 为例, 在目标运动过程中发生了不同程度的遮挡。图 5(f) 从第 40 帧开始, 干扰目标出现遮挡追踪目标最后完全遮挡, 干扰目标离开后, KCF、TLD 算法识别干扰目标继续追踪, 本文算法和其他算法没收到干扰项的影响继续追踪。

4) DEF: 以图 5(c) 视频序列为例, 开始阶段目标与干扰目标并肩而走如第 10 帧, 随着移动速度不

一样目标出现分离,DCFNet 算法出现追踪漂移现象,等两人完全分开时,只有本文算法和 SRDCF 算法能够稳健地追踪。

5) IV:以图 5(g)、(h)和(i)为例,追踪过程中目标所在区域都发生了剧烈光照变化。以 Woman 追踪序列为例,目标还出现了遮挡、尺度变化等情况,最后一帧中只有本文算法能够准确追踪,其他算法在中途发生追踪漂移,Staple 算法虽然追踪位置也在目标位置但是追踪窗口偏小。

4.2 定量分析

使用 OTB-13^[9]评价标准的追踪精度和追踪成功率来评估本算法的追踪性能;中心位置误差是衡量追踪精度的标准,通过比较追踪位置的中心位置与手工定位的准确中心的平均欧氏距离来确定中心位置误差,使用 20 个像素为评价阈值;边界框重叠

率 S 是追踪成功率的重要指标,设人工标注边框为 Y_a ,追踪结果边框为 Y_t ,则 $S = |Y_t \cap Y_a| / |Y_t \cup Y_a|$,当 S 大于阈值 t_0 时,认为追踪成功,成功率图给出了 t_0 从 0~1 变化时成功帧所占的比例,使用每一个成功率图的曲线下面积(AUC)给追踪算法进行排序。图 6 为 50 个追踪序列的精度和成功率曲线,图标按照算法性能由高到低排序,可以看出,不管追踪精度还是成功率,本文算法性能最好,在精度方面,比同样基于全连接网络的 CF2 算法提高了 0.8%,追踪成功率比基于对称网络的 Siamfc 算法提高了 4.1%,主要是因为本文算法有准确的目标尺度估计机制,使追踪目标能够较准确地覆盖,从而提升了追踪成功率。

为了进一步分析算法在面临目标 FM 以及 SV 问题的性能,图7为在这两种属性下42组追踪序列

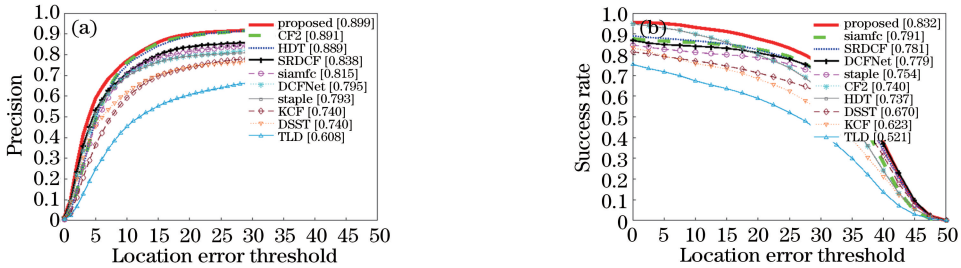


图 6 50 个追踪序列的精度和成功率曲线。(a)精度曲线;(b)成功率曲线

Fig. 6 Precision and success rate of 50 sequences. (a) Precision; (b) success rate

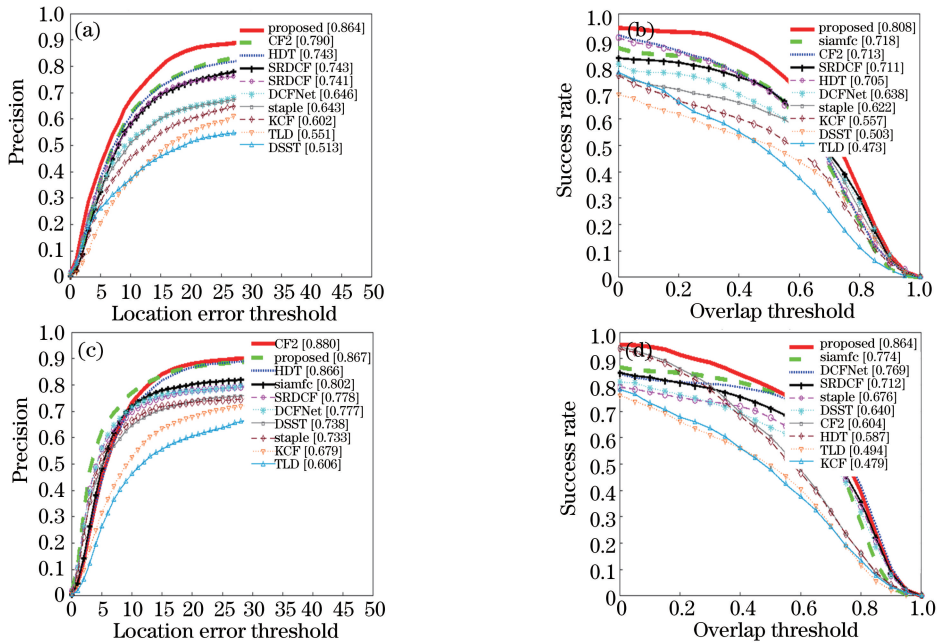


图 7 FM 和 SV 属性下追踪精度和准确率曲线。(a) FM 下精度曲线;(b) FM 下成功率曲线;
(c) SV 下精度曲线;(d) SV 下成功率曲线

Fig. 7 Precision and success rate on FM and SV attributes. (a) Precision on FM attribute; (b) success rate on FM attribute; (c) precision plot on SV attribute; (d) success rate on SV attribute

的精度与成功率曲线图。图 7(a)和(b)表示在目标 FM 属性下本算法与其他算法相比性能最好；图 7(c)和(d)表明在 SV 属性下追踪的精确度略低于 CF2 算法,但在追踪成功率上仍排在第一位。

同时还比较了该算法在视频序列中所包含的

其他属性(IV、OCC、LR、DEF、BC 和 OV)的性能,由图 8 可知,除了 BC 属性下图 8(i)、(j)和 LR 属性下图 8(e)、(f)情况,本文算法追踪性能不太理想,在其余属性下,本文算法精度和成功率都排在了第一位。

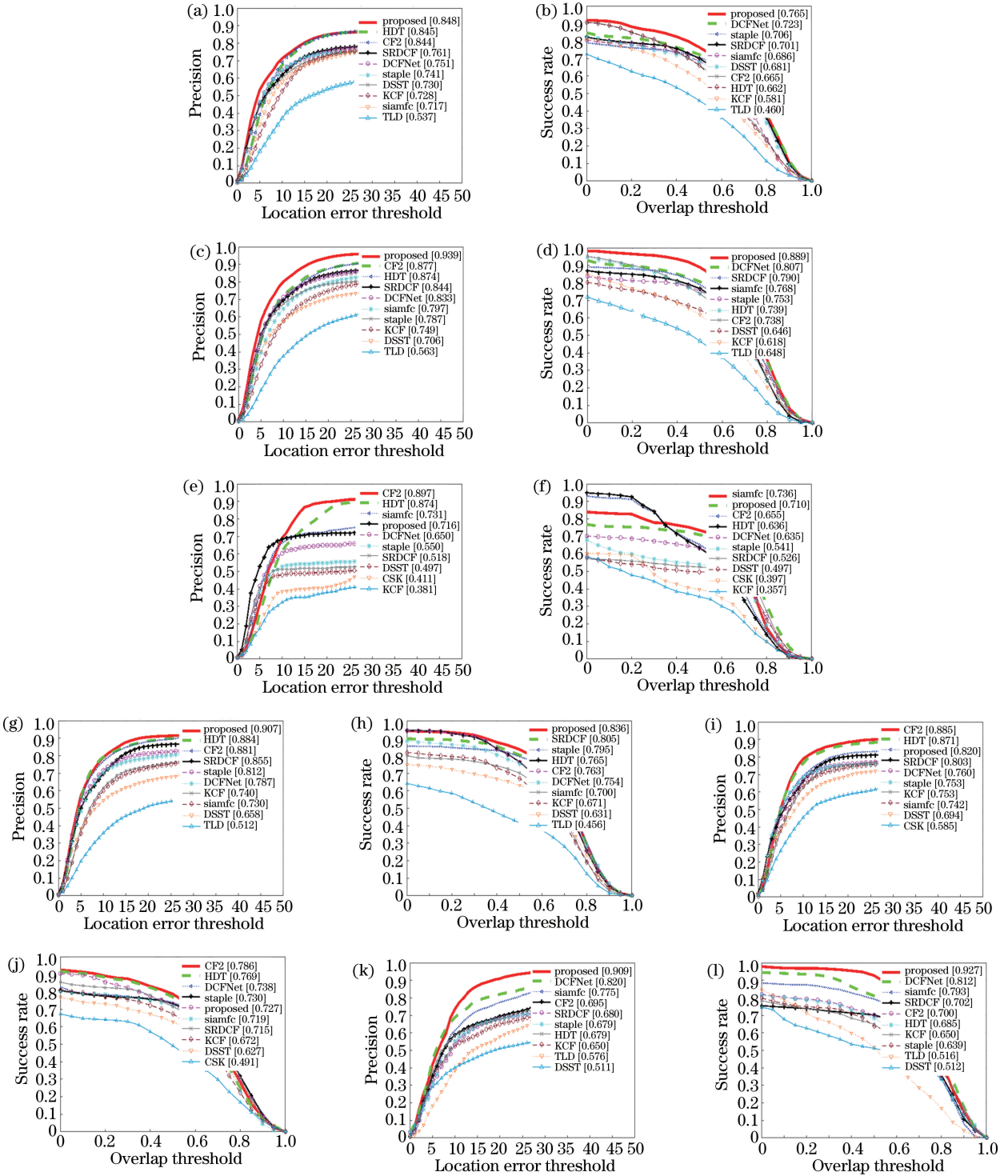


图 8 (a)(b) IV、(c)(d) OCC、(e)(f) LR、(g)(h) DEF、(i)(j) BC、(k)(l) OV 属性下追踪精度和成功率曲线

Fig. 8 Tracking precision and success rate on (a)(b) IV, (c)(d) OCC, (e)(f) LR, (g)(h) DEF, (i)(j) BC, and (k)(l) OV attributes

4.3 算法追踪速率

表 2 为各算法追踪速率对比结果, Tracker 表示追踪器, Platform 表示实验平台, FPS 表示追踪速率, Real 为是否达到实时性的要求, 从表 2 可以看出, 本文算法速度远大于 CF2 和 HDT 算法, 达到了实时性的要求; 但由于其追踪过程中要进行尺度估计, 因此速度稍慢于 Siamfc 算法; 构建的卷积网络比 DCFNet 复杂, 所以慢于 DCFNet 算法。其他算法的追踪速度都是原文给出。

表 2 本文算法与其他算法追踪速率对比

Table 2 Tracking speed comparison of proposed algorithm and other algorithms

Tracker	Proposed	CF2	DCFNet	Siamfc	HDT
Platform	GPU	GPU	GPU	GPU	GPU
FPS / (frame · s ⁻¹)	42.3	11	65.94	58	10
Real	Y	N	Y	Y	N

5 结 论

提出了一种基于全卷积对称网络的目标尺度自适应追踪算法, 首先构建全卷积对称的 CNN 网络确定目标的中心位置; 然后对目标候选区域进行多尺度采样, 通过比较模板与样本图像的方差阈值和海林洛距离, 确定目标追踪窗口。大量实验结果表明, 与其他追踪算法相比, 本文算法追踪精度和成功率都排第一名; 比同样使用对称网络的 Siamfc 算法在精度和成功率上分别提高了 8.4% 和 4.1%。综上所述, 本文算法较好地解决了目标 FM 及 CV 导致追踪失败的问题, 有效地提高了追踪精度和成功率, 满足实时追踪的要求。

在实验过程中发现, 当追踪序列出现分辨率较低、背景模糊的情况时, 本文算法容易跟踪失败, 因此, 如何结合视频预处理方法, 使深度学习网络学习到更加稳健的特征是下一步研究的重点。

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] Ma C, Yang X K, Zhang C Y, *et al.* Learning a temporally invariant representation for visual tracking [C] // IEEE International Conference on Image Processing (ICIP), 2015: 857-861.

- [3] Danelljan M, Häger G, Khan F S, *et al.* Convolutional features for correlation filter based visual tracking [C] // IEEE International Conference on Computer Vision Workshop (ICCVW), 2015: 621-629.
- [4] Wang L J, Wanli O Y, Wang X G, *et al.* Visual tracking with fully convolutional networks [C] // IEEE International Conference on Computer Vision (ICCV), 2015: 3119-3127.
- [5] Cai Y Z, Yang D D, Mao N, *et al.* Visual tracking algorithm based on adaptive convolutional features [J]. Acta Optica Sinica, 2017, 37(3): 0315002.
蔡玉柱, 杨德东, 毛宁, 等. 基于自适应卷积特征的目标跟踪算法 [J]. 光学学报, 2017, 37(3): 0315002.
- [6] Mao N, Yang D D, Yang F C, *et al.* Adaptive object tracking based on hierarchical convolution features [J]. Laser & Optoelectronics Progress, 2016, 53(12): 121502.
毛宁, 杨德东, 杨福才, 等. 基于分层卷积特征的自适应目标跟踪 [J]. 激光与光电子学进展, 2016, 53(12): 121502.
- [7] Danelljan M, Hager G, Khan F S, *et al.* Discriminative scale space tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [8] Wang X, Hou Z Q, Yu W S, *et al.* Target scale adaptive robust tracking based on fusion of multilayer convolutional features [J]. Acta Optica Sinica, 2017, 37(11): 1115005.
王鑫, 侯志强, 余旺盛, 等. 基于多层卷积特征融合的目标尺度自适应稳健跟踪 [J]. 光学学报, 2017, 37(11): 1115005.
- [9] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2411-2418.
- [10] Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05), 2005: 539-546.
- [11] Zhang M D, Wang Q, Xing J L, *et al.* Visual tracking via spatially aligned correlation filters network [M] // Cham: Springer International Publishing, 2018: 484-500.
- [12] Bertinetto L, Valmadre J, Henriques J F, *et al.* Fully-convolutional siamese networks for object tracking [C] // European Conference on Computer

- Vision, 2016:850-865.
- [13] Valmadre J, Bertinetto L, Henriques J, *et al.* End-to-end representation learning for correlation filter based tracking [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, 2017: 5000-5008.
- [14] 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, 2012:702-715.
- [15] 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.
- [16] Bertinetto L, Valmadre J, Golodetz S, *et al.* Staple: complementary learners for real-time tracking [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 1401-1409.
- [17] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
- [18] Li S S, Zhao G P, Wang J N. Distractor-aware object tracking based on multi-feature fusion and scale-adaption [J]. Acta Optica Sinica, 2017, 37(5): 0515005.
- 李双双, 赵高鹏, 王建宇. 基于特征融合和尺度自适应的干扰感知目标跟踪 [J]. 光学学报, 2017, 37(5): 0515005.