文章编号 1004-924X(2019)07-1621-11

引入再检测机制的孪生神经网络目标跟踪

梁 浩[1,2],刘克俭[3],刘 康[1,2],刘岩俊[1],陈小林[1*]

(1. 中国科学院 长春光学精密机械与物理研究所,吉林 长春 130033;
2. 中国科学院大学,北京 100049;
3. 中国人民公安大学,北京 100038)

摘要:针对全卷积孪生神经网络 SiamFC 在目标快速运动、相似干扰较多等复杂场景下跟踪能力不足的问题,本文引入 SINT 作为再检测网络对 SiamFC 进行了改进。本文算法在跟踪响应图出现较多波峰时,启用精确度更高的再检测网络对波峰位置进行重新判定。同时,本文采用了生成式模型构建模板来适应目标的各种变化,以及高置信度的模型更新策略来防止每帧更新可能对模板带来的污染。在 OTB2013 上对算法性能进行了测试,并选取了 9 个主流的目标跟踪算法进行对比,本文算法的跟踪精确度达到了 88.8%,排名第一,成功率达到了 63.2%,排名第二,相比 SiamFC 有很大地提升。对不同视频序列的分析结果表明,本文算法在目标快速运动、严重遮挡、背景杂波、光照变化和长期跟踪等场景下具有较强的准确性和鲁棒性。

关 键 词:目标跟踪;孪生神经网络;再检测;生成式模型;高置信度更新

中图分类号:TP391.4 **文献标识码**:A **doi**:10.3788/OPE.20192707.1621

Siamese network tracking with redetection mechanism

LIANG Hao[1,2], LIU Ke-jian[3], LIU Kang[1,2], LIU Yan-jun[1], CHEN Xiao-lin[1*]

(1. *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun* 130033, *China*;
2. *University of Chinese Academy of Sciences, Beijing* 100049, *China*;
3. *Remote Sensing Center, People's Public Security University of China, Beijing* 100038, *China*)
* *Corresponding author, E-mail*: 654040216@qq.com

Abstract: To solve the insufficient tracking capability problem for a fully convolutional Siamese network (SiamFC) in complex scenarios such as those involving fast motion and large similar interference, SINT was introduced as a redetection network to improve the SiamFC. When multiple peaks appeared in the tracking response map, the proposed algorithm enabled the redetection network to redetermine the target position with higher accuracy. At the same time, a generative model was adopted to construct a template to adapt to various appearance changes of the target, and a high-confidence model update strategy was used to avoid the model corruption problem. Our algorithm is tested on OTB2013, and nine state-of-the-art algorithms are selected for comparison. The tracking accuracy of

收稿日期:2018-09-21;**修订日期**:2018-12-23.
基金项目:国家重点研发计划资助项目(No. 2016YFC0803000);长春市科技发展计划资助项目(No. 17DY008)

our algorithm reaches 88.8%, the best among all the algorithms selectes for comparison, and the success rate reaches 63.2%, which is the second best. Both these properties offer considerable improvement over the SiamFC results. Analysis of several representative video sequences demonstrate that our algorithm has high accuracy and strong robustness in cases involving fast motion, severe occlusion, background clutter, illumination changes, and long-term tracking.

Key words: object tracking; siamese network; redetection; generative model; high-confidence update

1 引　言

目标跟踪是计算机视觉领域的一个基本问题，在诸如视频监控、人机交互等场景中被广泛应用[1]。在视频序列中的第一帧给定初始目标，如何在后续帧中找到此目标是目标跟踪的核心问题。由于遮挡、光照变化、运动模糊、外观变化等一系列因素，使得目标跟踪任务充满了挑战。

近年来，随着深度学习的引入，目标跟踪领域取得了很大进展。传统的判别式方法和 CNN 特征的结合[2-4]，大幅提高了跟踪的精度。但仅仅使用从计算机视觉其他领域预训练的神经网络提取的特征，难以充分利用神经网络端到端的强大的学习能力。MDNet[5]，ADNet[6]等使用端到端的方法来训练跟踪网络，并结合在线微调达到了很好的跟踪效果。但同时神经网络带来了计算量的大幅增加，导致跟踪速度的减慢，难以达到实时性的要求。

孪生神经网络是一类由两个或多个具有相同参数和权重的子网络组成的神经网络架构。孪生神经网络在涉及样本之间的相似性度量或两个可比较的事物之间的关系的任务中经常被使用。使用孪生神经网络的目标跟踪算法由于不进行网络的在线更新，在实时性方面有很大的优势。

使用孪生神经网络的代表性算法有 SiamFC[7]和 SINT[8]等，本文的工作正是基于这两个算法。SiamFC 训练了在一个较大的搜索区域搜索模板图片的孪生网络，此孪生网络是一个关于搜索区域的全卷积网络，而最后目标位置的估计通过计算两个输入的交叉相关，然后再进行插值得到，密集而且高效。该方法在公开的数据集上可以达到非常有竞争力的性能，并且在 GPU 上的运行帧率为 58 frame/s，远远超过实时性的要求。SINT 则使用了较为原始的孪生网络，通过生成多个候选的目标区域并通过前向传播获取评分，来估计目标的位置。这样的方式使得 SINT 在 GPU 上的运行帧率仅为 4 frame/s，但密集的候选框使其在跟踪精度比 SiamFC 通过交叉相关和插值得到的响应图更胜一筹。

SiamFC 由于拥有不错的实时性、有望应用到实际场景中而备受关注。尽管 SiamFC 同时拥有着相对不错的跟踪结果，但也存在着如下的缺点：SiamFC 仅使用第一帧标注的目标作为模板进行跟踪，这使得目标发生变化时难以更加准确的跟踪目标。如果目标发生了很严重的变化，将无法有效的识别目标。交叉相关及插值得到的跟踪精度略微粗糙，不能满足复杂场景的跟踪需求。对此，一些学者提出了改进措施。文献[9]提出了一种多模板融合策略，通过融合 3 种更新度不同的模板，来适应目标在运动过程中的变化。文献[10]通过融合不同卷积层的特征，保留了更多的空间位置信息，提升了网络的判别能力。

为了对 SiamFC 的算法性能进行提升，并同时保证算法的实时性，本文采用了如下方法：引入再检测机制，跟踪过程中对 SiamFC 的响应图进行分析，在 SiamFC 结果不够好的时候使用 SINT 网络进行更加精确的目标定位；使用生成模型来构建模板，并采用高置信度的模型更新策略。

本文算法在 OTB2013 上取得了不错的效果。改进之后的算法相比 SiamFC 有很大提升，并且仍然能够保持较快的跟踪速度，在 GPU 上的运行帧率为 23 frame/s。

2 全卷积孪生神经网络 SiamFC

跟踪一个任意目标可以被当做一种相似性学习。SiamFC 旨在学习一个函数 $f(z,x)$，函数 f 比较模板 z 和搜索区域 x，然后返回一个得分图。得分图和搜索区域 x 尺寸相同。得分越高，则说明这个区域和模板越相似。要找到 z 在新一帧中的位置，只需要对所有的可能位置都计算相似度

即可。SiamFC 使用卷积神经网络来学习函数 f，而基于神经网络的相似性学习通常使用孪生网络。孪生网络的作用是一个特征提取器 φ，同时提取模板 z 和搜索区域 x 的特征，然后将提取到的特征送入另一个函数 g，所以该相似性度量函数可被描述为：

$$f(z,x)=g(\varphi(z),\varphi(x)), \tag{1}$$

其中函数 g 是一个距离度量或相似度度量。这种深度孪生网络早已被广泛应用于人脸确认，关键描述点学习，one-shot 字符识别等任务上。

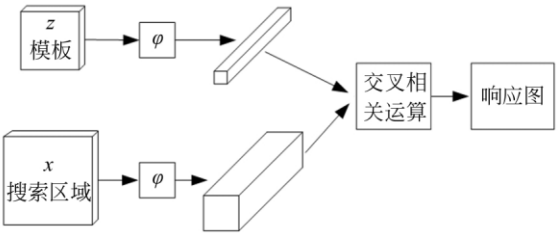


图 1　SiamFC 网络结构

Fig. 1　Network architecture of SiamFC

SiamFC 采用的神经网络为全卷积神经网络。全卷积神经网络应用于目标跟踪的优势在于，不必像目标检测一样生成很多候选区域，只需输入搜索区域的图像就可以转换成各个子窗口的相似性。该全卷积孪生网络可被定义为：

$$f(z,x)=\varphi(z)*\varphi(x)+b, \tag{2}$$

其中：x 是搜索区域图像，z 是模板图像，卷积内嵌函数 φ 使用了类似于 AlexNet[11] 结构的孪生网络，$*$ 表示交叉相关运算，b 为偏置项并且 $b\in R$。该函数的输出为一个数值为标量的得分响应图，其维度取决于搜索区域和模板图像的大小。把响应图得分最高的位置映射到搜索区域上，即得到了目标的位置。

训练过程使用模板图像和搜索区域组成的图像对，损失函数为：

$$L(y,v)=\frac{1}{|D|}\sum_{u\in D}l(y[u],v[u]), \tag{3}$$

其中：$y[u]\in\{+1,-1\}$ 和 $v[u]$ 是图像对的标签和响应图，$u\in D$ 对应响应图的每一个位置，l 是 logistic 损失函数：

$$l(y,v)=\log(1+\exp(-yv)). \tag{4}$$

卷积网络的参数 θ 通过随机梯度下降得到：

$$\underset{\theta}{\operatorname{argmin}}\ \underset{(z,x,y)}{E}\ L(y,f(z,x;\theta)). \tag{5}$$

3　孪生实例搜索跟踪网络 SINT

SINT 跟踪算法同样使用孪生神经网络来进行目标跟踪，该算法的流程为首先离线训练卷积网络得到匹配函数，然后在线跟踪，根据匹配函数选择与初始帧标定目标最为匹配的候选区域作为跟踪结果。网络结构如图 2 所示。

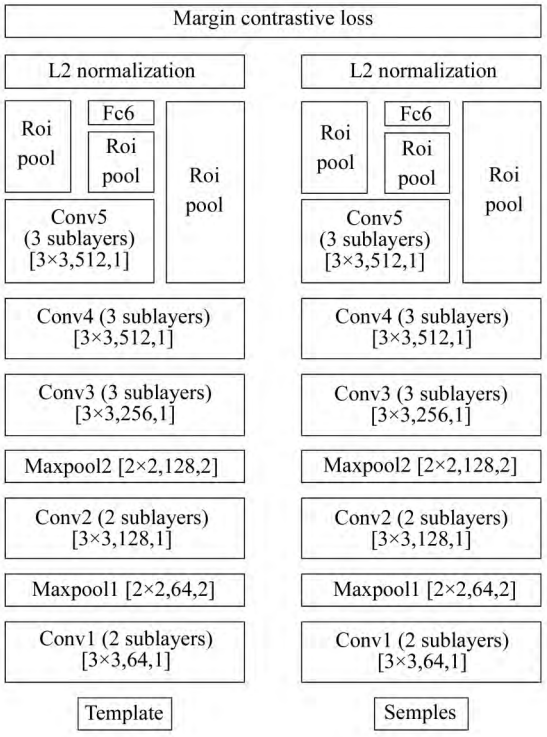


图 2　SINT 网络结构

Fig. 2　Network architecture of SINT

为了使定位更加精确，SINT 跟踪网络减少了最大池化层的使用。由于在每一帧的跟踪过程中，算法需要对几百个候选区域进行评估，带来了很大的计算量。因此，网络采用了感兴趣区域池化层（ROI Pooling）[12]，网络先对整张图片提取特征，然后通过感兴趣区域池化层将候选区域映射到特征图上，这样避免了对每个候选区域单独提取特征，大大减少了计算量。

由于采用了具有无界正范围的整流线性单元，在卷积神经网络中，网络的输出是无界的，并且在不同的尺度范围内会发生变化。不在同一尺度范围的特征，会导致被跟踪对象提取特征时效果不好。因此，网络在损失函数层之前加入了 l_2

正则化层，使得不同尺度的特征被映射到同一尺度范围内。

网络的最后，两个孪生网络分支连接到同一个损失函数层。在跟踪过程中，希望网络生成特征表示，这些特征表示使得正样本对之间的距离足够小，负样本对之间的距离足够大。因此采用了如下的边际对比损失函数：

$$L(x_j,x_k,y_{j,k})=\frac{1}{2}y_{j,k}D^2+\frac{1}{2}(1-y_{j,k})\max(0,\varepsilon-D^2),\quad(6)$$

其中：$D=\|f(x_j)-f(x_k)\|_2$ 是两个 l_2 正则化特征之间的欧氏距离，$y_{j,k}\in\{0,1\}$ 用来表示 x_j 和 x_k 是否为相同的目标，ε 是不同目标特征之间的最小距离。

跟踪时，网络不再进行在线更新。由于第一帧的目标没有受到污染，后面每一帧的候选区域都与第一帧进行对比，得分最高的位置即为最终的目标位置：

$$\hat{x}=\underset{x\in x_{j,t}}{\operatorname{argmax}}M(x_{to},x),\quad(7)$$

其中：$x_{j,t}$ 是第 t 帧中的所有候选区域，x_{to} 是第一帧中的目标，M 是匹配函数。匹配函数的结果是 x_{to} 和 x 两个特征向量之间的内积。样本的生成采用了半径采样(Radius Sampling)策略[13]，即在前一帧目标的中心位置周围以不同的半径尺度采样。在确定最终目标位置之后，使用边框回归对目标边框进行精修，回归器通过对第一帧采样得到的正样本进行训练得到。

4　本文算法实现

4.1　算法流程

本文算法在 SiamFC 的基础上，引入了再检测机制，集成了 SINT 跟踪算法作检测网络，取得了超过 SINT 算法的跟踪结果，同时仍然保证了算法的实时性。基于 SiamFC 的跟踪网络负责主要的跟踪部分，得到响应图之后进行置信度判定，在跟踪结果不理想的情况下，启用基于 SINT 的检测网络重新对目标进行判定，之后对模板进行更新，进行下一帧的跟踪。算法流程图如图 3 所示。

跟踪过程中，当新一帧图像到来时，SiamFC 跟踪网络通过对比模板和搜索区域的相似度得到响应图，理想情况下，响应最大的位置即为目标的位置。但是，由于遮挡、相似物干扰等复杂情况的存在，目标可能会出现在非极大响应值的位置上。SiamFC 原算法中采用了余弦窗来抑制边缘响应，越靠近中心的响应赋予越高的权重，给算法性能的稳定性带来了很大提升。但是同时也带来了风险，在目标运动较快的时候容易跟丢目标。因此，本文去除了加在响应图上的余弦窗，在次级波峰较多的情况下，使用检测网络对多个波峰位置进行重新判定，来确定目标的最终位置。

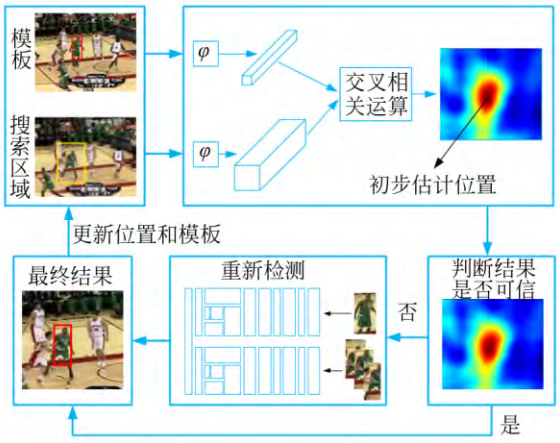


图 3　跟踪流程图

Fig. 3　Flowchart of proposed method

图 4 为 OTB 数据集中 Walking2 视频序列第 215 帧的再检测示意图(彩图见期刊电子版)。其中，图 4(a)为 SiamFC 网络得到的响应图，存在两个明显的波峰。其中一个波峰为真实目标，对应图 4(b)红色框区域。另外一个波峰为干扰，对应图 4(b)黄色框区域。此时干扰位置的响应强度已经超过了真实的目标，如果没有再检测算法将跟丢目标。在本文算法中，存在大于 0.75 倍主波峰的次级波峰时，就会启动再检测网络。为了充分利用响应图的信息，在响应图的波峰位置进行随机采样得到候选区域，如图 4(c)。得到候选区域后，由 SINT 网络进行判定，得到最终的目标位置，结果如图 4(d)所示。

4.2　模板构建与更新

不同于 SiamFC 仅使用第一帧的目标作为目标，本文使用了生成式模型来构建模板，同时采用了高置信度的模型更新策略。我们使用混合高斯模型来生成模板，模型 x 的概率分布为：

$$p(x)=\sum_{l=1}^{L}\pi_l N(x;\mu_l;\boldsymbol{I}),\quad(8)$$

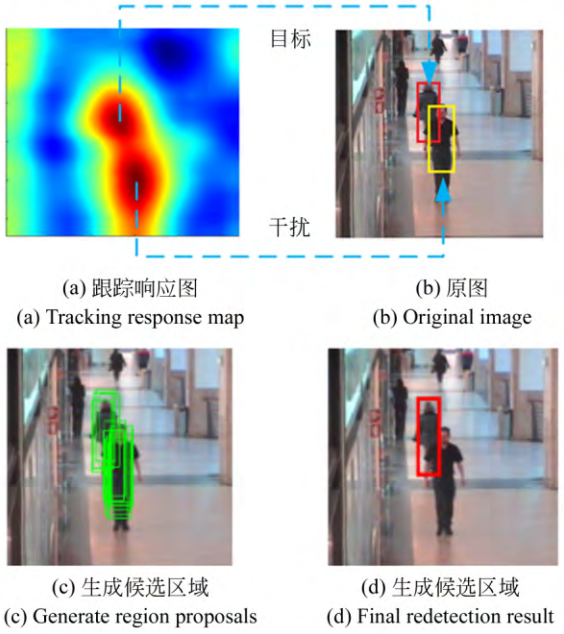


图 4　再检测流程图

Fig. 4　Theillusion of redetection

其中：L 是混合高斯模型组件 $N(x;\mu_l;\boldsymbol{I})$ 的数量，π_l 是组件 l 的先验权重，μ_l 是组件 l 的均值，将协方差矩阵设置为单位矩阵 $\boldsymbol{I}$，以避免在高维样本空间中推断计算量过大。

为了高效地对混合高斯模型进行更新，采用了文献[14]中的在线更新算法的简化版。给定一个新的样本 x_j，对新的组件 m 进行初始化，$\pi_m=\gamma$（γ 为学习率），$\mu_m=x_j$。如果组件的数量超过了 L，如果原有某一组件的权重低于设定的阈值，抛弃该组件加入新组件。否则，对原有组件中最接近的两个组件 k 和 l 合并为组件 n，并加入新的组件。

$$\pi_n=\pi_k+\pi_l,\mu_n=\frac{\pi_k\mu_k+\pi_l\mu_l}{\pi_k+\pi_l}. \tag{9}$$

不同于多数算法每一帧或者隔固定帧更新模板，本文采取了高置信度的模型更新策略。置信度的高低，取决于响应图的质量。文献[15]中基于相关滤波的响应图，提出了 APCE（Average Peak-to-Correlation Energy）指标来衡量响应图的质量。

$$APCE=\frac{|F_{\max}-F_{\min}|^2}{mean\left(\sum_{w,h}(F_{w,h}-F_{\min})^2\right)}, \tag{10}$$

其中：$F_{\max}$，$F_{\min}$，$F_{w,h}$ 分别表示响应图 F 的最大值、最小值和第 w 行、第 h 列的值。APCE 表示响应图的波动程度和检测目标的置信度。峰值越高，噪声越小，目标被检测到的概率越大，APCE 值越大。反之，如果出现遮挡或者相似物干扰等情况，APCE 值将明显地下降。在 APCE 大于某一比例的历史均值时，对模板进行更新，可以在一定程度上防止模板被污染。

5　实验结果与分析

本文在公开数据集 OTB2013 上测试了算法的性能，该数据集包含了 51 个视频序列。这些序列在每一帧都标注了目标的位置，并且每个序列标注了 11 个属性，涵盖了各种挑战因素：尺度变化（SV），遮挡（OCC），光照变化（IV），运动模糊（MB），变形（DEF），快速运动（FM），平面外旋转（OPR），背景杂波（BC），出视野（OV），平面内旋转（IPR）和低分辨率（LR）。采用的测试方法为 OPE（One-Pass Evaluation），主要采用了两个评价指标：精确度图（Precision plot）和成功率图（Success plot）。精确度得分表示与真实位置相比，估计位置在 20 像素以内的帧的百分比。成功率（Success）定义为每个成功率图的曲线下面积（AUC），即采样重叠阈值对应的成功率平均值。

同时我们将本文的算法和 9 个主流的跟踪算法进行了比较，包括：基于孪生网络跟踪的 SiamFC，SINT，CFNet[16]，DCFNet[17]，引入了检测机制的长期跟踪算法 LCT[18]，引入跟踪置信度的 LMCF，以及基于相关滤波的实时性算法 Staple[19]，DSST[20]，KCF[21]。

本文算法的实验平台为 MATLAB R2016a，使用的神经网络框架包括 MatConvNet[22] 和 MatCaffe[23]。使用的实验设备 CPU 为 Intel i7-6850k 3.60 GHz，运行内存 64 GB，GPU 为 Titan X。

5.1　整体性能评估

图 5 为所有算法的精确度图和成功率图，从图中可以看出，本文算法相对于 SiamFC 有很大的提升，精确度达到了 88.8%，成功率达到了 63.2%。并且，算法在保证实时性的情况下，精确度超过了用来重新检测的速度仅为 4 frame/s 的算法 SINT。表 1 中列出了本文算法和 SiamFC，SINT 在精确度、成功率和帧率上的对比。

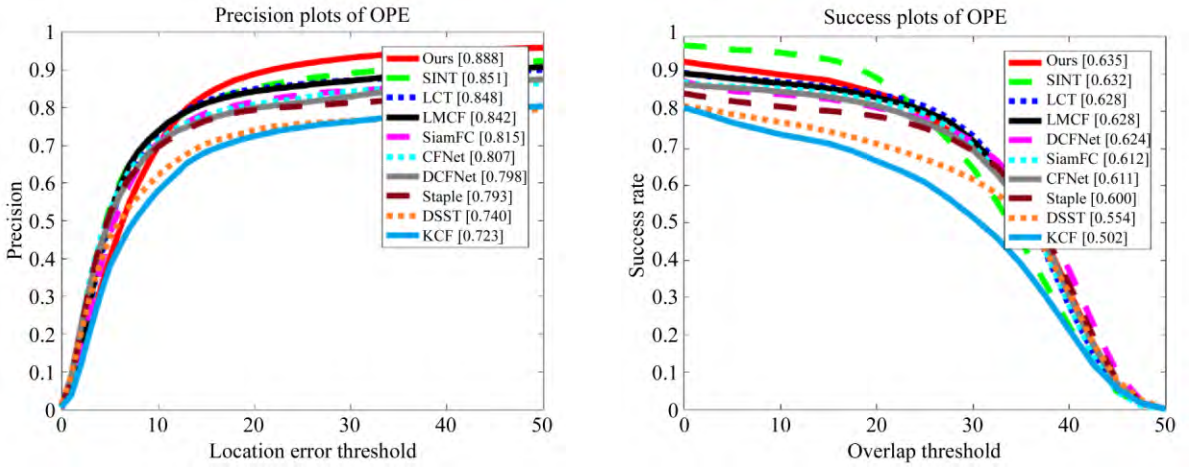


图 5　在 OTB2013 使用 OPE 的精确度图和成功率图

Fig. 5　Precision and success plots using the OPE on OTB2013

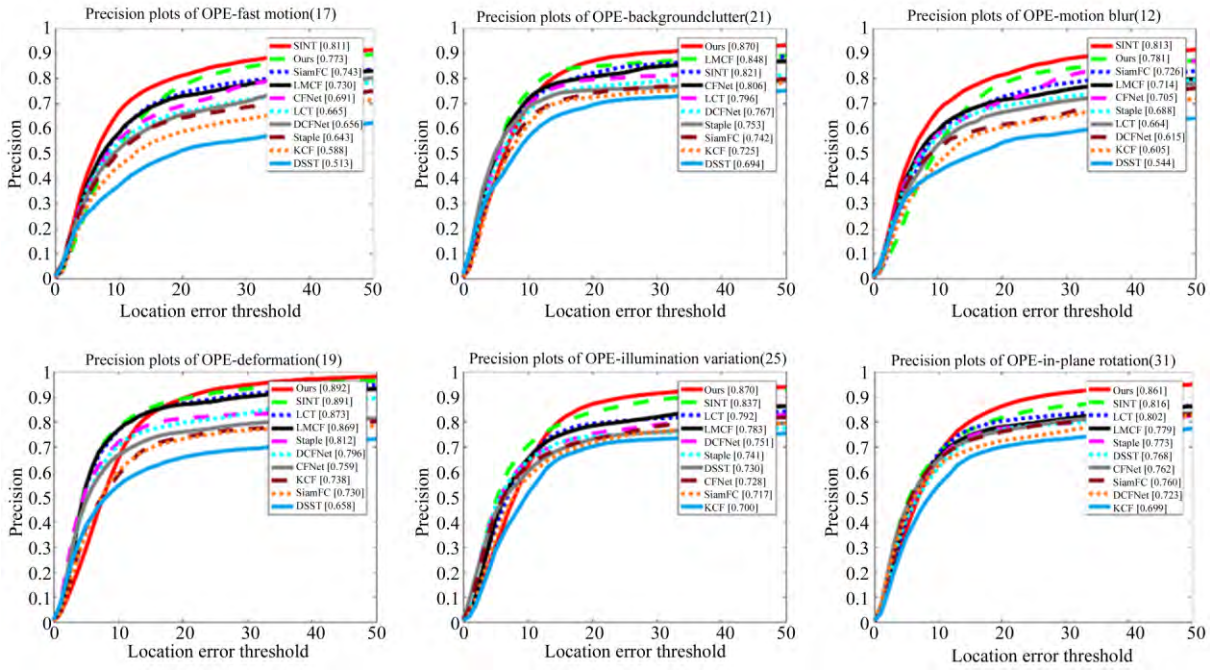
表 1　OTB2013 上的实验总结

Tab. 1　Summary of experiments on OTB-100 on OTB2013

	Precision	AUC	FPS
SiamFC	0.815	0.612	58
SINT	0.851	0.635	4
Ours	0.888	0.632	23

5.2　基于属性的性能分析

OTB2013 数据集标注了与跟踪场景相关的 11 个属性，这些属性影响着跟踪器的性能，可以用来评估跟踪器在不同场景下的表现。图 6 为 10 个算法在 OTB2013 的 11 个属性上的精确度图。从图 6 可以看出，本文算法在 8 个属性上排名第一，尤其在低分辨率属性上远远超过其他算法。在快速运动和运动模糊两个属性，仅次于 SINT 排名第二。在出视野的情况下，算法未能超过基准算法 SiamFC。相比较于 SiamFC，本文算法在 11 个属性上分别提高了 3%，12.8%，5.5%，16.2%，16.3%，10.1%，19.3%，6.3%，6.7%，10%，−0.07%。



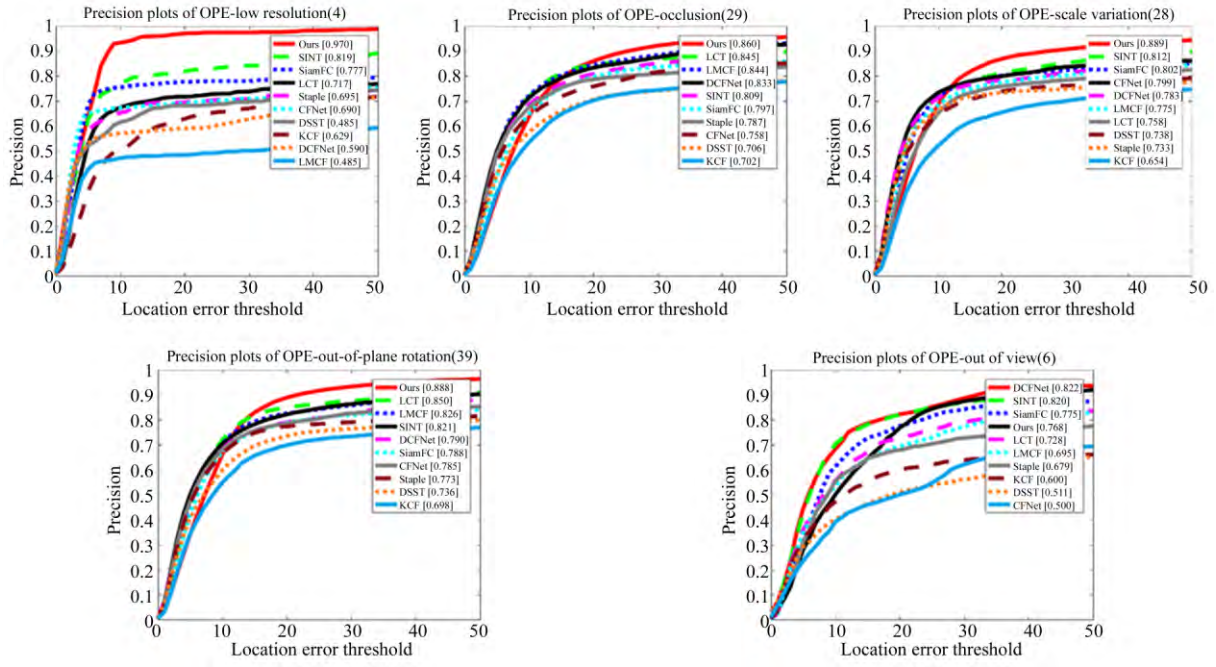


图 6　OTB2013 上 11 个属性的精确度图(标题上的数字表示此属性的数据集数量)

Fig. 6　Precision plots of the attributes on OTB2013. The values appearing in the title denote the number of videos associated with the respective attribute

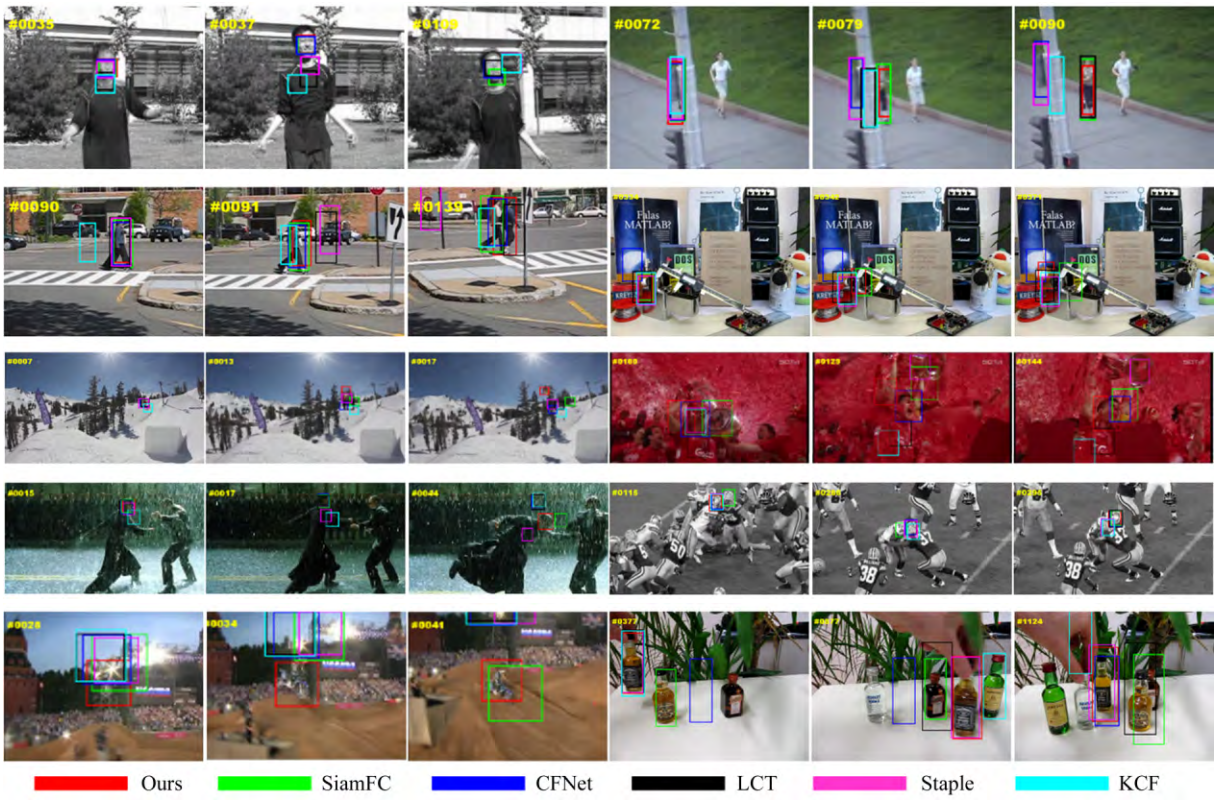


图 7　6 个跟踪算法在 10 个视频序列上的结果展示

Fig. 7　Visualization of tracking results of six representative trackers on ten challenging sequences

5.3　定性比较

为了更好地说明跟踪算法的性能，本文选择了一些具有挑战性的视频序列来进一步对比。这些视频序列代表着不同的复杂场景，同时为了更方便的展示结果，这一部分选择了 6 个跟踪算法来进行对比。在 10 个视频序列上的结果如图 7 所示，图左上角数字为帧数，视频序列依次为 Jump，jogging1，Couple，Lemming，Skiing，Soccer，Matrix，Football，MotorRolling and Liquor(彩图见期刊电子版)。

5.3.1　快速运动

为了在目标快速运动时捕捉到目标，跟踪算法不宜选择较小的搜索区域，而较大的搜索区域增加了目标被相似物干扰的风险。传统相关滤波类算法 KCF 和 Staple 由于在训练阶段加入了余弦窗，对远处目标的判别能力较弱。在 Jumping 视频序列中，目标从 33 帧开始向上跳起，35 帧时 LCT 和 KCF 最先丢失目标，37 帧时 Staple 也丢失了目标。SiamFC 由于加在响应图上的余弦窗，抑制了目标在搜索区域边缘处的响应，在后续帧中经常需要两帧来跟上快速运动的目标。在 Skiing 视频序列中，目标为小目标，而且运动速度较快。KCF 一开始就就丢失了目标，在 13 帧的时候，目标和背景中的树木融为一体，本文算法出现了定位失误，但在下一帧又重新检测到目标。在后面几帧的快速运动中，其他算法均丢失了目标。在 Couple 视频序列中，91 帧时镜头突然移动带来了目标在图像中的位置突变，相关滤波类算法 LCT，KCF 和 Staple 跟踪结果此时仍然停留在上一帧的位置。

5.3.2　遮挡

遮挡是目标跟踪中一大难点，如果目标被遮挡时未采取有效的更新策略，滤波器或者模板将错把遮挡物学习为目标。在 Jogging1 视频序列中，目标从 72 帧开始几乎完全被遮挡，79 帧目标出现在视野中时，本文算法和 SiamFC 最先定位到目标，SiamFC 不更新模板在此情况下获得了优势，而本文算法的高置信度 模型更新也避免了模板被污染。具有再检测机制的 LCT 在 82 帧时也重新定位到目标，KCF，Staple 则已经将遮挡区域确认为目标，从此丢失了目标。在 Soccer 视频序列中，从 100 帧开始，目标出现了严重的遮挡，目标几乎已经不可见，各个跟踪算法的结果都处于随机游走状态。到 129 帧时，目标较为清晰的出现在视野中，CFNet 最先定位到了目标。144 帧时本文算法依靠再检测网络定位到了目标，LCT.，Staple 和 KCF 则再也没有定位到目标。在 Lemming 视频序列中，玩具熊从 334 帧开始被严重遮挡，在 342 帧时本文算法和 SiamFC 跟踪到了目标旁边的游标卡尺上。371 帧本文算法重新检测到目标，CFNet 和 LCT 也分别在 374 帧和 376 帧重新定位到目标。此时，经历了几十帧的严重遮挡，KCF 和 Staple 已经将遮挡区域判定为目标。CFNet 则在视频一开始就丢失了目标。

5.3.3　背景杂波和光照变化

背景杂波是指目标附近的背景颜色或纹理与目标相似，这对算法的判别能力提出了很高的要求。使用了 CNN 特征的本文算法、SiamFC 和 CFNet，在对相似物的辨别能力上要比其他算法表现更好。光照变化在一些不复杂的场景下比较容易克服，但和背景杂波结合在一起时，跟踪结果就很容易漂移到周围的相似干扰物上。在 Matrix 视频序列中，人物的头部和衣服及周围的背景灰度都很接近。同时在 15 帧，光照也发生了剧烈变化，KCF 最先丢失目标。17 帧光照再次发生变化，Staple 也丢失了目标。由于目标和背景太相似，目标运动到 37 帧时，所有算法都已丢失目标。44 帧时光照再次变亮，本文算法重新检测到了目标。在 Football 视频序列中，多个运动员的头部极为相似。115 帧时 SiamFC 将定位到了目标旁边的运动员头部，295 帧时 KCF 和 Staple 也定位失误。在 MotorRolling 视频序列中，摩托车在第 28 帧出现了快速运动，此时背景也发生了剧烈变化，并且背景中多处地方与摩托车颜色相似，形成了干扰。依靠再检测网络，本文算法及时跟住了目标，其余算法均已丢失目标。

5.3.4　长期跟踪

OTB2013 数据集中有 3 个超过 1 000 帧的视频序列，分别为 Doll，Lemming 和 Liquor。表 1 为 6 个算法在 3 个平均 CLE(中心位置误差)值。其中 Doll 视频序列场景较为简单，除了 SiamFC 未能适应目标的尺度变化，在最后的几百帧丢失了目标，其他跟踪算法都取得了不错的效果。而 Lemming 视频序列目标存在长时间遮挡这一严重问题，LCT 在 1 271 帧开始丢失目标，只有本文算法和 SiamFC 一直到视频序列结束都没有丢失

目标。Liquor 序列中瓶子多次出现平面外旋转，并且周围存在相似瓶子的干扰。除了本文算法和 Staple，其他算法均出现了较长时间的丢失目标。尽管本文算法在瓶子被遮挡时也出现了几次短暂的丢失目标，但依靠检测算法又迅速的进行了重新定位。高置信度的模型更新策略和重新检测模块，使得本文算法在长期跟踪中能够更好地适应目标的变形和遮挡。

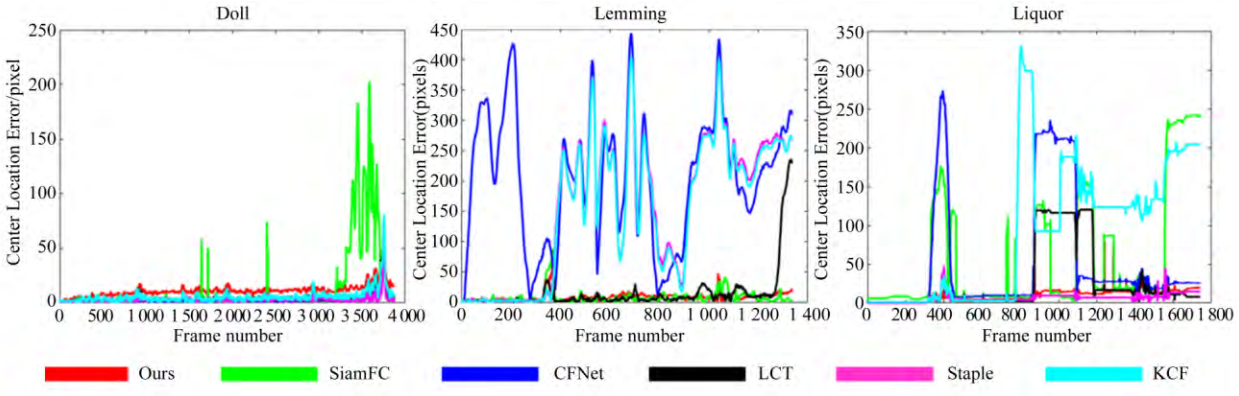


图 8　3 个长时间序列每帧的中心位置误差对比

Fig. 8　Fame-by-frame comparison of center location errors (in pixels) on 3 long-term sequences

6　结　论

针对 SiamFC 跟踪算法在目标快速运动、相似目标干扰、光照变化、背景杂波等情况下跟踪效果较差的问题，本文引入了 SINT 算法中所使用的孪生神经网络进行二次检测，在 SiamFC 表现不佳的场景下对目标重新定位。同时，针对 SiamFC 不进行模型更新、在长期跟踪中难以适应目标本身变化的问题，本文使用生成式模型对目标进行建模，同时引入置信度评估，采用了高置信度的模型更新策略。在 OTB2013 上 51 个视频测试序列上的评估结果表明，本文算法相对于 SiamFC 在跟踪精度上有了很大的提升，精确度达到了 88.8%，成功率达到了 63.2%，同时运行帧率仍然能保持在 23 frame/s。和目前的一些能达到实时性的主流算法的对比也表明，在很多复杂场景中本文算法都有着不错的表现，在低分辨率的情况下表现尤为突出。本文的下一步工作将针对再检测网络进行改进，以进一步提升算法的性能和实时性。

参考文献：

[1] 程帅，孙俊喜，曹永刚，等. 增量深度学习目标跟踪[J]. 光学 精密工程，2015，23(4)：1161-1170.
CHENG SH，SUN J X，CAO Y G，*et al*.. Target tracking based on incremental deep learning [J]. *Opt. Precision Eng.*，2015，23(4)：1161-1170. (in Chinese)

[2] MA C，HUANG J B，YANG X K，*et al*.. Hierarchical convolutional features for visual tracking [C]//2015 *IEEE International Conference on Computer Vision (ICCV)*，*December* 7-13，2015. *Santiago*，*Chile*. *New York*，*USA*：*IEEE*，2015：3074-3082.

[3] WANG N Y，LI S Y，GUPTA A，*et al*.. Transferring rich feature hierarchies for robust visual tracking [J]. *arXiv preprint arXiv*：1501.04587，2015

[4] DANELLJAN M，ROBINSON A，SHAHBAZ KHAN F，*et al*.. *Beyond Correlation Filters*：*Learning Continuous Convolution Operators for Visual Tracking* [M]//Computer Vision-ECCV 2016. Cham：Springer International Publishing，2016：472-488.

[5] NAM H，HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]//2016

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016. *Las Vegas, NV, USA. New York, USA: IEEE*, 2016: 4293-4302.

[6] YUN S, CHOI J, YOO Y, *et al*.. Action-decision networks for visual tracking with deep reinforcement learning[C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July* 21-26, 2017. *Honolulu, HI. New York, USA: IEEE*, 2017.

[7] BERTINETTO L, VALMADRE J, HENRIQUES J F, *et al*.. *Fully-convolutional Siamese networks for object tracking*[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016: 850-865.

[8] TAO R, GAVVES E, SMEULDERS A W M. Siamese instance search for tracking[C]//2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June* 27-30, 2016. *Las Vegas, NV, USA. New York, USA: IEEE*, 2016: 1420-1429.

[9] DAI K H, WANG Y H, YAN X Y. Long-term object tracking based on Siamese network[C]//2017 *IEEE International Conference on Image Processing (ICIP), September* 17-20, 2017. *Beijing. New York, USA: IEEE*, 2017: 3640-3644.

[10] KUAI Y L, WEN G J, LI D D. Hyper-feature based tracking with the fully-convolutional Siamese network[C]//2017 *International Conference on Digital Image Computing: Techniques and Applications (DICTA), November* 29-December 1, 2017. *Sydney, NSW. New York, USA: IEEE*, 2017: 1-7.

[11] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.

[12] GIRSHICK R. Fast r-cnn [C]. *Proceedings of the IEEE International Conference on Computer Vision*, 2015: 1440-1448.

[13] HARE S, SAFFARI A, TORR P H S. Struck: structured output tracking with kernels[C]//2011 *International Conference on Computer Vision, November* 6-13, 2011. *Barcelona, Spain. New York, USA: IEEE*, 2011: 263-270.

[14] DECLERCQ A, PIATER J H. Online learning of Gaussian mixture models-a two-level approach [C]. *VISAPP*, 2008: 605-611.

[15] WANG M M, LIU Y, HUANG Z Y. Large margin object tracking with circulant feature maps [C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July* 21-26, 2017. *Honolulu, HI. New York, USA: IEEE*, 2017: 21-26.

[16] VALMADRE J, BERTINETTO L, HENRIQUES J, *et al*.. End-to-end representation learning for correlation filter based tracking [C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July* 21-26, 2017. *Honolulu, HI. New York, USA: IEEE*, 2017: 5000-5008.

[17] WANG Q, GAO J, XING J L, *et al*.. Dcfnet: Discriminant correlation filters network for visual tracking [J]. *arXiv preprint arXiv*: 1704.04057, 2017

[18] MA C, YANG X K, ZHANG C Y, *et al*.. Long-term correlation tracking[C]//2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June* 7-12, 2015. *Boston, MA, USA. New York, USA: IEEE*, 2015: 5388-5396.

[19] BERTINETTO L, VALMADRE J, GOLODETZ S, *et al*.. Staple: complementary learners for real-time tracking[C]//2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June* 27-30, 2016. *Las Vegas, NV, USA. New York, USA: IEEE*, 2016: 1401-1409.

[20] DANELLJAN M, H？GER G, SHAHBAZ KHAN F, *et al*.. Accurate scale estimation for robust visual tracking[C]//*Proceedings of the British Machine Vision Conference* 2014, *Nottingham. British Machine Vision Association*, 2014: 1-5.

[21] 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.

[22] VEDALDI A, LENC K. Matconvnet: Convolutional neural networks for matlab [C]. *Proceedings of the 23rd ACM International Conference*

on Multimedia, 2015: 689-692.

[23] JIA Y Q, SHELHAMER E, DONAHUE J, *et al*.. Caffe: Convolutional architecture for fast feature embedding [C]. *Proceedings of the 22nd ACM International Conference on Multimedia*, 2014: 675-678.

作者简介：

梁　浩(1994－)，男，河南驻马店人，硕士研究生，2016 年于华中科技大学获得学士学位，主要从事计算机视觉及机器学习方面的研究。E-mail: hawkeye.liang@foxmail.com

导师简介：

陈小林(1980－)，男，吉林长春人，副研究员，硕士生导师，2003 年于吉林大学获得学士学位，2010 年于中国科学院长春光学精密机械与物理研究所获得博士学位，研究方向为：实时视频图像处理，嵌入式处理器设计，景象模拟仿真嵌入式实时处理器设计。E-mail: 654040216@qq.com