基于双模板Siamese网络的鲁棒视觉跟踪算法

侯志强[①③] 陈立琳*[①③] 余旺盛[②] 马素刚[①③] 范九伦[①]

[①](西安邮电大学计算机学院 西安 710121)

[②](空军工程大学信息与导航学院 西安 710077)

[③](西安邮电大学陕西省网络数据分析与智能处理重点实验室 西安 710121)

摘 要：近年来，Siamese网络由于其良好的跟踪精度和较快的跟踪速度，在视觉跟踪领域引起极大关注，但大多数Siamese网络并未考虑模型更新，从而引起跟踪错误。针对这一不足，该文提出一种基于双模板Siamese网络的视觉跟踪算法。首先，保留响应图中响应值稳定的初始帧作为基准模板R，同时使用改进的APCEs模型更新策略确定动态模板T。然后，通过对候选目标区域与2个模板匹配度结果的综合分析，对结果响应图进行融合，以得到更加准确的跟踪结果。在OTB2013和OTB2015数据集上的实验结果表明，与当前5种主流跟踪算法相比，该文算法的跟踪精度和成功率具有明显优势，不仅在尺度变化、平面内旋转、平面外旋转、遮挡、光照变化情况下具有较好的跟踪效果，而且达到了46 帧/s的跟踪速度。

关键词：Siamese网络；目标跟踪；双模板；模板更新

中图分类号：TP391.4 **文献标识码**：A **文章编号**：1009-5896(2019)09-2247-09

DOI: 10.11999/JEIT181018

Robust Visual Tracking Algorithm Based on Siamese Network with Dual Templates

HOU Zhiqiang[①③] CHEN Lilin[①③] YU Wangsheng[②]

MA Sugang[①③] FAN Jiulun[①]

[①](*Institute of Computer*, *Xi'an University of Posts and Telecommunications*, *Xi'an* 710121, *China*)

[②](*Information and Navigation Institute*, *Air Force Engineering University*, *Xi'an* 710077, *China*)

[③](*Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing*, *Xi'an University of Posts and Telecommunications*, *Xi'an* 710121, *China*)

Abstract: In recent years, the Siamese networks has drawn great attention in visual tracking community due to its balanced accuracy and speed. However, most Siamese networks model are not updated, which causes tracking errors. In view of this deficiency, an algorithm based on the Siamese network with double templates is proposed. First, the base template R which is the initial frame target with stable response map score and the dynamic template T which is using the improved APCEs model update strategy to determine are kept. Then, the candidate targets region and the two template matching results are analyzed, meanwhile the result response maps are fused, which could ensure more accurate tracking results. The experimental results on the OTB2013 and OTB2015 datasets show that comparing with the 5 current mainstream tracking algorithms, the tracking accuracy and success rate of the proposed algorithm are superior. The proposed algorithm not only displays better tracking effects under the conditions of scale variation, in-plane rotation, out-of-plane rotation, occlusion, and illumination variation, but also achieves real-time tracking by a speed of 46 frames per second.

Key words: Siamese network; Object tracking; Dual templates; Template update

收稿日期：2018-11-06；改回日期：2019-05-29；网络出版：2019-06-12

*通信作者： 陈立琳 454525999@qq.com

基金项目：国家自然科学基金(61473309, 61703423)

Foundation Items: The National Natural Science Foundation of China (61473309, 61703423)

1 引言

视觉目标跟踪技术需要在视频序列中自动地定位目标，其在视频监控、军事侦察、自动驾驶和人体姿态估计等方面具有广泛的应用[1]。视觉目标跟踪的核心问题是如何在具有遮挡、运动模糊、复杂背景和目标形变等场景下准确检测和定位目标[2]。

近年来，基于外观相似性比较策略的Siamese网络由于其良好的跟踪性能，在视觉跟踪领域引起了极大的关注[3–12]。SINT[4], SiameseFC[5]和RASNet[6]以初始帧作为模板，学习先验深度Siamese网络的相似函数，通过相似性比较确定候选目标。虽然这些跟踪方法不仅获得了不错的跟踪精度，还具有较快的跟踪速度，但存在3个问题[4–6]：首先，大多数Siamese跟踪方法中使用的特征都是浅层外观特征，只能区分前景和非语义背景；其次，大多数的Siamese跟踪都无法进行模型更新，由于在跟踪过程中目标外观及场景视角发生变化，固定不变的模板会引起跟踪匹配误差及场景适应性下降，甚至会导致跟踪失败；最后，Siamese网络训练中存在样本不均衡，正样本种类不充足导致模型泛化性能不够，负样本过于简单大多不包含语义信息。针对缺乏深度特征这一不足，SA-Siam[3]使用了两组Siamese网络：语义分支和外观分支，浅层外观分支相关系数图和深层语义分支相关系数图按照一定比例加起来，得到最终的响应图，实现特征融合；FlowTrack[7]则在Siamese网络中利用光流运动信息来提高特征表示和跟踪精度。针对未考虑模型更新问题：CFNet[8]和Dsiam[9]始终以初始帧为输入，通过岭回归学习变换目标外观矩阵和背景矩阵达到适应目标时域变化和抑制背景变化的目的；EDCF[10]在SiameseFC特征层后加入反卷积网络，使得特征网络更多地关注目标的细节表述，并结合上下文感知的相关滤波抑制背景干扰，并且使得模型可以在线更新。针对训练样本不均衡的情况：SiameseRPN[11]和DaSiamRPN[12]在跟踪中引入检测方法，使用ILSVRC(Imagenet Large Scale Visual Recognition Challenge)[13]和YouTube-BB(YouTube-BoundingBoxes)[14]这两个超大数据集预训练模板分支，使用分支RPN网络检测目标。

本文在模型更新这一问题上对Siamese网络进行改进，提出一种基于双模板Siamese网络的方法进行跟踪：选择初始帧目标作为基准模板R，基准模板视为未被污染的模板，保证目标跟踪的平均性能；动态模板T学习目标在运动过程中表观上的变化，在平稳跟踪的基础上使跟踪结果更加准确，第1个动态模板同样选择初始帧目标，之后的动态模板使用改进的APCEs模型更新策略确定。双模板的方式一方面能够提高模板正确性和对跟踪的支持度，减少目标漂移；另一方面选择动态模板时采用的自适应更新策略能够减少动态模板的更新次数，从而保证跟踪速度。这一思想与Triplet网络有相似之处，但从算法本质上来讲，本文算法是两个Siamese网络跟踪结果的融合，以获得更好的跟踪性能，而Triplet网络[15]使用正负样本与目标进行网络训练，从而在建模时具有更好的细节效果。在OTB2013[16]和OTB2015[2]数据集上的实验结果表明，与当前5种主流的跟踪算法相比，本文算法在跟踪精度和成功率上均具有优势，同时在GPU上的运算速度达到了实时性。

2 SiameseFC网络

本文算法通过对2个SiameseFC网络[5]跟踪结果的融合，最终得到跟踪目标的位置和大小。SiameseFC网络去掉了原始Siamese网络[4]中的Padding层和全连接层，保留5层卷积，除第5层卷积外，每个卷积层后面都有1个ReLU非线性激活层，用于训练过程中降低过拟合的风险。全卷积网络让较大的搜索区域输入网络成为可能，可以详尽地测试目标在新图像中所有可能的位置，找到和目标外观相似度最高的候选区域，从而预测目标位置。

如图1所示，SiameseFC网络作为一个特征提取器，提取模板z和搜索区域x的特征，一起输入到相似度函数中计算相似度，并返回一个得分响应图。响应图反映模板z和搜索区域x中与模板z大小相同的候选模板x'相似性：得分高，x'与z相似，反之不相似。接下来，本文从相似性学习、损失函数训练和网络训练3个方面对SiameseFC网络做一个简单的说明。

2.1 相似性学习

SiameseFC网络在模板和搜索区域全卷积后做1次评估，计算相似度。网络中的深度相似性学习

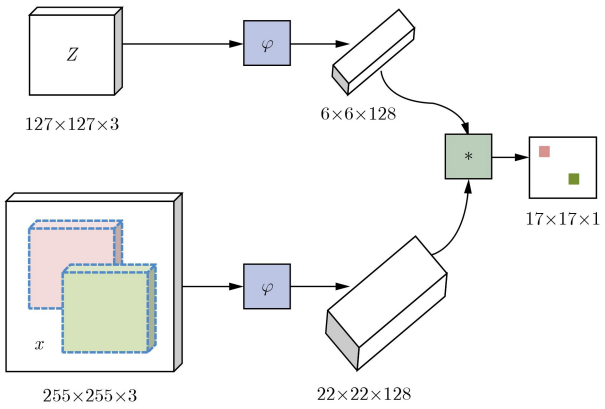


图 1 SiameseFC网络框架

函数可以是简单的距离度量或者相似度度量，在SiameseFC采用互相关函数$f(z;x)$作为相似度函数，计算经过φ特征提取后的特征图相似度

$$f(z;x)=\varphi(z)*\varphi(x)+bI \tag{1}$$

其中，*表示卷积，bI表示在响应图中每个位置的取值。

2.2 损失函数训练

SiameseFC网络用一种判别的方法，在正负样本对上采用最大似然估计进行损失函数训练。搜索区域x中的每一个候选子窗口，相当于1个样本，而它输出的得分就是它是正/负样本的概率。这是一个应用逻辑回归的典型二分类问题，逻辑损失就可以表示为

$$l(y,v)=\log_2(1+\exp(-yv)) \tag{2}$$

其中，v是候选位置的得分，y是它的真实类别,$y\in\{1,-1\}$。采用1个样本图像和1个较大的搜索区域图像来训练SiameseFC网络，定义1个响应图的损失函数为每一个损失的均值

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

对于每一个位置u，需要1个真正的标签$y[u]\in\{1,-1\}$。当位置u与响应图中心位置p的欧氏距离在某一阈值R内时，认定其为正样本，否则为负样本。

2.3 网络训练

目标跟踪使用的3个传统的数据集VOT[17], ALOV[18], OTB[2,16]总共有不到600个视频序列，并且数据集中的视频还有一些重叠，因此SiameseFC网络没有采用这3个传统数据集对网络进行训练，而是采用包含4500个视频序列的ILSVRC[13]数据库进行了训练。SiameseFC网络采用离线训练，在离线阶段解决相似度学习问题，网络训练中通过随机梯度下降方法调整网络参数。

3 本文算法

3.1 总体框架

在Siamese网络中[3–10]，使用初始帧目标作为模板，后续跟踪都和初始帧的目标模板(后文称为基准模板R)进行匹配度计算，以此来估计两帧之间区域特征相似性。然而，当目标快速运动、目标形变、光照变化等情况下，不更新模板往往会造成跟踪失败；但只使用动态模板跟踪也是有风险的：当跟踪出现模型漂移时，现有的跟踪算法没有有效的应对措施对漂移进行校正，就会造成跟踪结果出现进一步漂移，一直到跟踪失败。因此，本文在SiameseFC框架的基础上提出采用双模板的方式进行跟踪：保留未被污染的初始帧目标作为基准模板R，使用改进的APECs更新策略进行模板更新得到动态模板T，两个模板相辅相成，分别与搜索区域进行相似度匹配，得到各自的响应图，对2个响应图加权得到最终响应。本文算法框架如图2所示。

3.1.1 模板、搜索区域的获取

模板区域的获取如图3所示。

(1) 模板的获取

首先，以被选为模板的目标中心位置P和目标大小(w,h)裁剪一个正方形区域，该正方形的边长$s_z=(w+2p)\times(h+2p)$，其中p为上下文余量

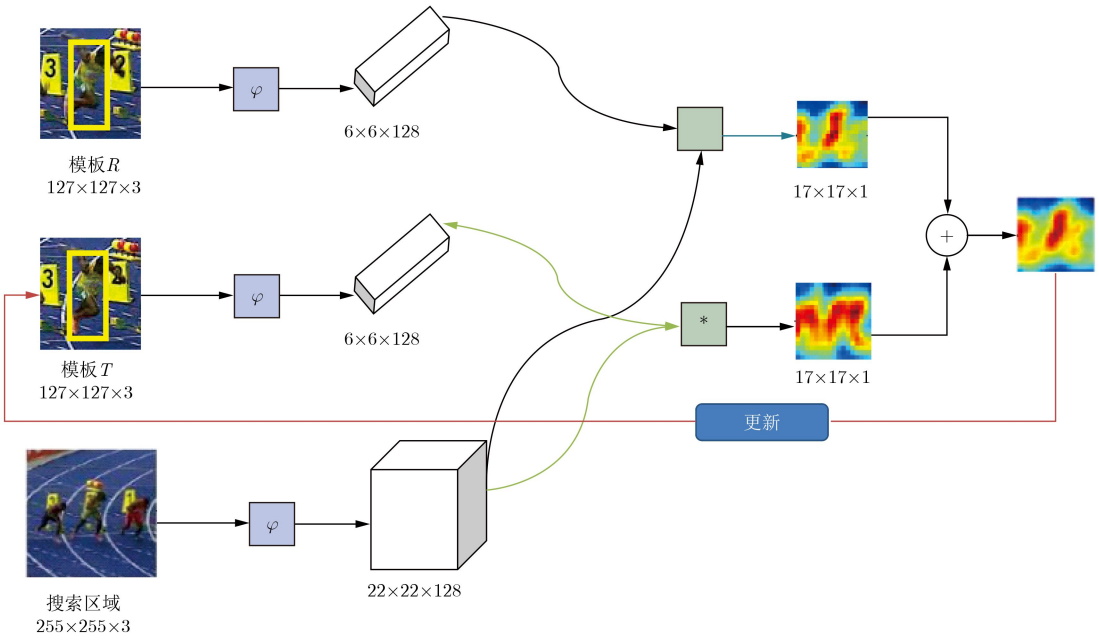


图 2 基于Siamese网络下的双模板跟踪

$p=(w+h)/4$。然后，以RGB 3通道均值填充，乘以尺度因子s，放缩为$127\times127\times3$的模板图像。s满足$s(w+2p)\times s(h+2p)=127\times127$。

以初始帧目标中心位置p_{xr}和目标大小(w_{xr},h_{xr})裁剪、放缩得到$127\times127\times3$大小的基准模板R；以模板更新帧的目标中心位置p_{xt}和目标大小(w_{xt},h_{xt})裁剪、放缩得到$127\times127\times3$大小的动态模板T，如图3所示。

(2) 搜索区域的获取

获取当前帧的搜索区域，先以上一帧中心位置p_{t-1}确定感兴趣区域为ROI的中心位置，ROI边长为$s_x=(s_z+2\times\mathrm{pd})\times s$，考虑背景填充区域pd和尺度因子$s$,$\mathrm{pd}=(255-127)/2s$，ROI图像块作为候选样本，放缩得到一个$255\times255\times3$的搜索区域，如图3所示。

3.1.2 相似度估计

模板R、模板T和搜索区域通过特征映射操作φ分别得到$6\times6\times128$的特征图$\varphi(z_r)$、特征图$\varphi(z_t)$和$22\times22\times128$的特征图$\varphi(x)$。$\varphi(z_r)$,$\varphi(z_t)$分别与搜索区域提取的特征图$\varphi(x)$进行相似度计算。本文改进了SiameseFC网络和互相关函数，将模板R、模板T得到的特征图$\varphi(z_r)$,$\varphi(z_t)$合并为1个4维矩阵与搜索区域特征图$\varphi(x)$一起输入互相关层，达到同时计算相似性的目的，有效地提高了跟踪速度。

对2个响应图进行加权处理得到1个新的响应图

$$\mathrm{F}(z_r,z_t;x)=\omega_1\times(f(z_r;x))+\omega_2\times(f(z_t;x)) \tag{4}$$

其中，权重参数的确定，非常重要。ω_1太小会导致模型漂移，ω_2太小又会使模型更新作用甚微。因此，在本文算法中，赋予基准模板较大权重，动态模板较小权重，即$\omega_1\in[0.5,1]$, $\omega_2\in[0.2,0.5]$。在这一原则下，对2个响应值进行综合分析，确定权重参数具体数值。经过加权处理后图上响应值最高的位置为相似度最高的位置，响应值最高的位置相对于中心的偏移再乘以步长，就是目标在下一帧的真实位置。

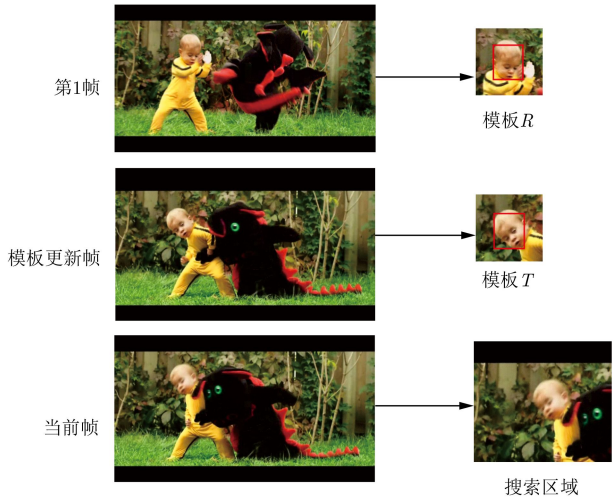


图 3 模板与搜索区域

3.2 更新策略

在跟踪过程中，模型更新过慢会造成模板无法跟上目标的变化；更新过快又会导致跟踪速度的下降。因此，模板更新需要根据造成模板变化的不同情况而定，在目标发生变化时更新，在目标被遮挡时停止更新。

本文算法使用LCMF[19]中跟踪置信度APCE来判断模板的更新时机，在OTB2015数据集上成功率达到0.609，平均跟踪速度44 fPS。实验过程发现响应图中最高响应值变化更为剧烈，使用最高响应值减去(w,h)位置上的响应值作为分母置信度变化更为剧烈和明显。因此，本文对APCE改进为

$$\mathrm{APCEs}=\frac{|F_{\max}-F_{\min}|^2}{\mathrm{mean}\left(\sum_{w,h}(F_{\max}-F_{w,h})^2\right)} \tag{5}$$

其中，$F_{\max}$，$F_{\min}$，$F_{w,h}$代表响应图上的最高响应值、最低响应值和(w,h)位置上的响应值。本文把原APCE分母$\mathrm{mean}\left(\sum_{w,h}(F_{w,h}-F_{\min})^2\right)$改进为$\mathrm{mean}\left(\sum_{w,h}(F_{\max}-F_{w,h})^2\right)$。当APCEs突然减小时，一般是目标被遮挡或者目标丢失的情况，不进行模型更新，避免模型漂移。这种更新方式有效地区分了目标表观变化和目标遮挡、目标丢失对跟踪的不同影响，提高了算法的鲁棒性。只有当APCEs和$F_{\max}$都以一定比例大于各自的历史均值mAPCEs, mFmax的时候，模型才进行更新，这样一方面大大减少了模型漂移的情况，另一方面减少了模型更新的次数，达到了加速的效果。

λ值的确定是本文算法的关键，参数太小会造成更新过于频繁，容易出现模型过更新；参数太大会使更新速度滞后于目标表观变化，导致跟踪性能下降。通过大量实验，确定了阈值λ为0.85，如表1所示。

3.3 算法具体流程

本文主要算法流程如表2所示。

4 实验

本文采用MATLAB2017a和Visual Studio 2013编程来验证本文算法的性能，在Intel(R) Core(TM)i7-6850k 3.6 GHz处理器上进行测试，并采用GPU(NVIDIA GTX 1080Ti)进行加速。分别在2个流行的跟踪数据集上做了实验：包含51个视

频序列(36个彩色序列)的OTB2013[16]，包含100个视频序列(77个彩色序列)的OTB2015[2]。本文算法与5个主流的跟踪算法进行比较：SiameseFC和SiameseFC_3S[5], Staple[20], SRDCF[21]和MEEM[22]。其中，SiameseFC, SiameseFC_3S为基于深度学习的跟踪算法，也是本文算法的基准算法；Staple, SRDCF为基于相关滤波的跟踪算法；MEEM考虑了模型更新问题。

4.1 定性分析

图4给出了本文算法和另外5种算法在OTB2015数据集上的部分跟踪结果，从以下5个方面对算法进行定性分析：

(1) 尺度变化：以视频Doll和Dog1为例，跟踪过程中出现了明显的尺度变化，尺度变化使目标外观发生了变化，虽然6种算法都能始终跟上目标，但只有本文算法和SRDCF算法、SiameseFC算法

表 1　λ取值对精度、成功率的影响(OTB2015)

λ	0.50	0.60	0.70	0.80	0.850	0.90	1.00	1.10
成功率	0.447	0.513	0.587	0.603	**0.614**	0.605	0.585	0.591
精度	0.642	0.697	0.742	0.779	**0.793**	0.761	0.761	0.774

表 2　基于Siamese网络下的双模版跟踪算法

输入: 图像序列: I_1, I_2, I_n; 初始目标位置: $P_0=(x_0, y_0)$, 初始目标大小: $s_0=(w_0, h_0)$

输出: 预估目标位置: $P_e=(x_e, y_e)$, 预估目标大小: $s_e=(w_e, h_e)$.

for t=1, 2,⋯,n, do:

步骤1　跟踪目标

(1) 以上一帧中心位置P_{t-1}裁剪第t帧中的感兴趣区域ROI，放大为搜索区域；

(2) 提取基准模板R，动态模板T和搜索区域的特征；

(3) 使用式(4)计算两个模板特征与搜索区域特征的相似性，得到结果响应图,响应图中最高响应点即为预估目标位置。

步骤2　模型更新

(1) 使用式(5)计算跟踪置信度APCEs；

(2) 计算$F_{\max}$和APCEs的平均值mF$_{\max}$和mAPCEs；

(3) 如果$F_{\max}>\lambda$mF$_{\max}$且APCEs $>\lambda$mAPCEs，更新动态模板T；

Until图像序列的结束。

图 4 本文和5种算法的部分跟踪结果对比

能够较好地适应目标的尺度变化。SRDCF[22]算法是在时域上对滤波器进行正则化，从而对背景区域的响应达到有效抑制，在尺度变化、背景杂波下获得了更好的性能。SiameseFC算法和本文算法都采用了5个尺度，并通过线性插值方法更新尺度，在跟踪速率上有一定优势；

(2) 目标旋转：以视频Board, MotorRolling和Diving为例，在跟踪过程中目标出现了明显的旋转变化，要求算法具有高度旋转不变性。在MotorRolling视频中可以明显看出，大部分的算法都出现了跟踪漂移或跟踪失败，但本文算法和SRDCF算法能够较好地跟踪目标；

(3) 目标遮挡：以视频Coke和Walking2为例，目标在跟踪过程中被遮挡，目标遮挡导致跟踪偏移或最终导致跟踪失败。在目标被遮挡和重新出现的情况下，本文算法、SRDCF算法、SiameseFC算法和Staple算法能够很好地跟踪目标。Staple[21]算法使用HOG特征和COLOR特征两种互补的特征因子对目标进行学习，融合跟踪结果，实现互补，在对跟踪速度无较大影响的情况下跟踪效果得到了提升。本文算法使用跟踪置信度来判断更新模板的时机，避免在目标被遮挡时更新模板，从而有效避免了模型漂移；

(4) 快速运动：以视频Diving和Skating2为例，由于快速运动目标表观发生了明显变化，导致模板与搜索区域匹配度降低，增加了跟踪难度。对于Diving视频，SiameseFC_3S算法和SRDCF算法在46帧时就完全丢失了目标；对于Skating2视频，MEEM算法和SiameseFC_3S算法均出现了跟踪漂移，而本文算法由于及时更新了模板能够跟踪到目标；

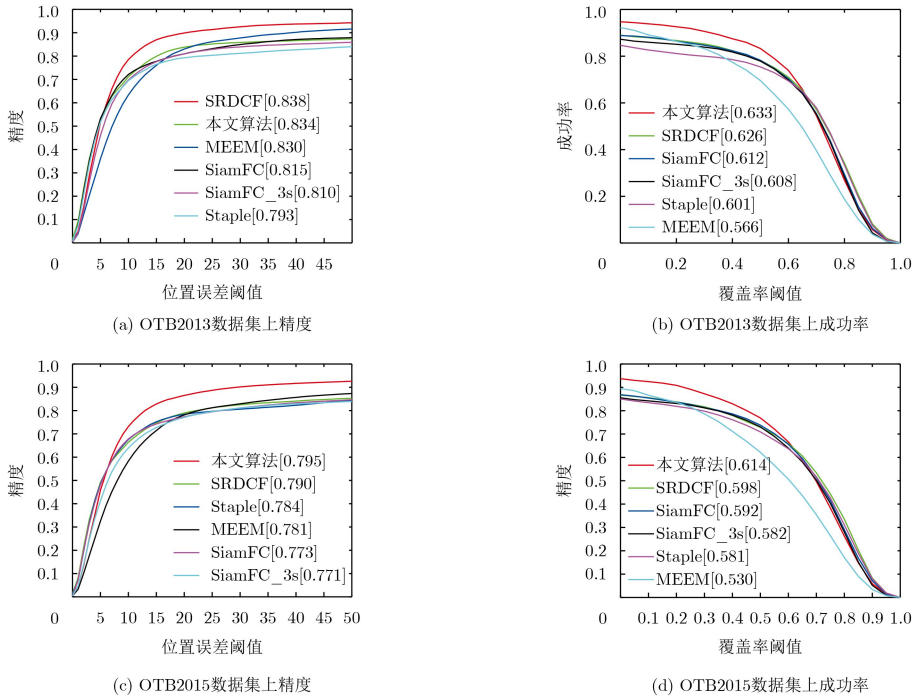
(5) 光照变化：以视频MotorRolling为例，跟踪过程中背景光照条件出现了剧烈的变化，要求算法对光照变化具有较好的稳健性。只有基于分层卷积特征的本文算法和SRDCF算法能够始终跟踪目标。

4.2 定量分析

对跟踪算法进行评估的方法主要体现在中心位置误差和覆盖率两个评价指标上：覆盖率指的是跟踪结果与真实目标的重叠率，如果当前覆盖率超过某个阈值，就判定帧中的目标被成功跟踪；中心位置误差指跟踪结果与真实目标的中心位置的欧式距离，如果中心位置误差低于给定的阈值，就判定目标跟踪成功。覆盖率和中心位置误差分别在成功率图和精度图中体现。

图5分别表示5种跟踪算法在OTB2013和OTB2015数据集上的整体成功率曲线和精度曲线。由图5可以看出，本文算法的成功率高于其他对比算法，和SiameseFC算法比较：在OTB2013数据集上成功率提升了2.1%，精度提升了1.9%；在OTB2015数据集上成功率提升了1.6%，精度提升了2.0%，获得了46 帧/s的速度。

为了进一步分析该算法的优缺点，本文提供了基于属性的性能分析来说明本文的跟踪算法在关键属性上的优势。OTB中的所有视频序列都被手动标注了几个具有挑战性的属性，包括尺度变化(SV)、



(a) OTB2013数据集上精度　(b) OTB2013数据集上成功率

(c) OTB2015数据集上精度　(d) OTB2015数据集上成功率

图 5 OTB2013和OTB2015成功率和精度

遮挡(OCC)、光照变化(IV)、形变(DEF)、运动模糊(MB)、快速运动(FM)、平面内旋转(IPR)、平面外旋转(OPR)、超出视野(OV)、背景杂波(BC)和低分辨率(LR)[2]。表3和表4分别列出了上述11种属性的跟踪条件下6种跟踪算法的成功率和精度，其中最优结果加粗显示，次优结果加下划线表示，第3优结果加下划虚线表示。表中的字母缩写分别表示不同的属性，括号内的数字表示包含的视频数目。由表3和表4可以看出，在11种不同属性的跟踪条件中，本文算法除背景杂波、光照变化和低分辨率属性外跟踪成功率均处于最优位置，同时，跟踪成功率和跟踪精度均处于前三。由此表明，本文算法对于目标发生变化具有良好的跟踪性能，而且对于其他复杂条件下的跟踪也具有较好的鲁棒性。

4.3 算法跟踪速率

模板更新频率对跟踪速度有着较大的影响，更新频率越高，算法的跟踪速度越慢。本文算法对模板、搜索区域特征提取和进行相关性计算，更新频率越高，特征提取次数越多，因此，跟踪速度随模板更新频率存在差异。在GPU条件下，本文算法在OTB2015的100组视频序列中的平均跟踪速度为46 fPS。表5列出了本文算法与5种算法的跟踪速度对比，分别列出了各个算法的编程方式和实验平台，M代表MATLAB, C代表C++, Y代表实时跟踪算法，N代表非实时跟踪算法。与传统算法和深度学习等比较新的算法相比，本文算法在速度上占有较大优势，但由于加入了模板更新，速度要慢于SiameseFC算法和Staple算法。

5 结束语

本文在SiameseFC基础上提出一种双模板的跟踪算法，通过对候选目标与基准模板和动态模板相似性结果融合，并依据跟踪置信度对动态模板进行更新，使模板更新速度与目标表观变化相适应的同时抑制模板过更新。在两个流行的跟踪数据集OTB2013, OTB2015上进行评估，结果表明该算法有效提高了跟踪的成功率和准确率，并且达到了46帧/s的实时跟踪速度。

实验中发现，在背景杂波和运动模糊情况下，本文算法跟踪效果还有待提高。这是由于本文算法仍然使用比较浅层的AlexNet[23]，获得浅层外观特征，只能区分前景和非语义背景。这就需要修改网络结构或者融合运动特征。修改网络结构必须满足两个条件：(1)网络需要满足严格的平移不变性；(2)网络需要具有对称性。近期发布的SiamRPN++[24]

表 3　不同属性下算法的跟踪成功率对比结果

算法	SV(64)	OPR(63)	IPR(51)	OCC(49)	DEF(44)	FM(39)	IV(38)	BC(31)	MB(29)	OV(14)	LR(9)
本文算法	**0.577**	**0.596**	**0.595**	**0.613**	**0.573**	**0.607**	0.605	0.577	**0.633**	**0.538**	0.460
SiameseFC	0.553	0.549	0.579	0.564	0.510	0.569	0.550	0.572	0.525	0.467	0.584
SiameseFC_3S	0.552	0.558	0.557	0.567	0.506	0.568	0.568	0.523	0.550	0.506	**0.618**
SRDCF	0.561	0.550	0.544	0.569	0.544	0.597	**0.613**	**0.583**	0.595	0.460	0.514
Staple	0.525	0.535	0.552	0.561	0.554	0.537	0.598	0.574	0.546	0.481	0.459
MEEM	0.470	0.526	0.529	0.495	0.489	0.542	0.517	0.519	0.557	0.488	0.382

表 4　不同属性下算法的跟踪精度对比结果

算法	SV(64)	OPR(63)	IPR(51)	OCC(49)	DEF(44)	FM(39)	IV(38)	BC(31)	MB(29)	OV(14)	LR(9)
本文算法	**0.781**	**0.796**	**0.815**	**0.811**	**0.804**	**0.816**	**0.801**	0.770	0.749	**0.717**	0.878
SiameseFC	0.732	0.744	0.780	0.720	0.690	0.735	0.711	0.748	0.654	0.615	0.805
SiameseFC_3S	0.735	0.757	0.742	0.722	0.690	0.743	0.736	0.690	0.705	0.669	**0.900**
SRDCF	0.745	0.571	0.745	0.735	0.734	0.769	0.792	**0.775**	**0.767**	0.597	0.765
Staple	0.727	0.738	0.770	0.726	0.748	0.697	0.792	0.766	0.708	0.661	0.695
MEEM	0.736	0.795	0.794	0.741	0.754	0.752	0.740	0.746	0.731	0.685	0.808

表 5　本文算法与5种算法跟踪速度对比

	本文算法	SiameseFC	SiameseFC_3S	SRDCF	Staple	MEEM
Code	M+C	M+C	M+C	M+C	M+C	M+C
PlatformFPS	GPU46(Y)	GPU58(Y)	GPU86(Y)	GPU5(N)	CPU80(Y)	CPU10(N)

以均匀分布的采样方式让目标在中心点附近进行偏移，缓解了深度网络破坏严格平移不变性带来的影响，让深度网络应用于Siamse中成为可能。如何将本文算法应用于深度网络中，如何在跟踪中引入更鲁棒的特征，将是下一步工作研究的重点。

参 考 文 献

[1] 侯志强, 韩崇昭. 视觉跟踪技术综述[J]. 自动化学报, 2006, 32(4): 603–617.
HOU Zhiqiang and HAN Chongzhao. A survey of visual tracking[J]. *Acta Automatica Sinica*, 2006, 32(4): 603–617.

[2] WU Yi, LIM J, and YANG M H. Object tracking benchmark[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1834–1848. doi: 10.1109/TPAMI.2014.2388226.

[3] HE Anfeng, LUO Chong, TIAN Xinmei, *et al*. A twofold Siamese network for real-time object tracking[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4834–4843.

[4] TAO Ran, GAVVES E, and SMEULDERS A W M. Siamese instance search for tracking[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1420–1429.

[5] BERTINETTO L, VALMADRE J, HENRIQUES J F, *et al*. Fully-convolutional Siamese networks for object tracking[C]. 2016 European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 850–865.

[6] WANG Qiang, TENG Zhu, XING Junliang, *et al*. Learning attentions: Residual attentional Siamese network for high performance online visual tracking[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4854–4863.

[7] ZHU Zheng, WU Wei, ZOU Wei, *et al*. End-to-end flow correlation tracking with spatial-temporal attention[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 548–557.

[8] 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, Honolulu, USA, 2017: 5000–5008.

[9] GUO Qing, FENG Wei, ZHOU Ce, *et al*. Learning dynamic Siamese network for visual object tracking[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 1781–1789.

[10] WANG Qiang, ZHANG Mengdan, XING Junliang, *et al*. Do not lose the details: Reinforced representation learning for high performance visual tracking[C]. 2018 International Joint Conferences on Artificial Intelligence, Stockholm, Swedish, 2018.

[11] LI Bo, YAN Junjie, WU Wei, *et al*. High performance visual tracking with Siamese region proposal network[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8971–8980.

[12] ZHU Zheng, WANG Qiang, LI Bo, *et al*. Distractor-aware Siamese networks for visual object tracking[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 103–119.

[13] RUSSAKOVSKY O, DENG Jia, SU Hao, *et al*. ImageNet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211–252. doi: 10.1007/s11263-015-0816-y.

[14] REAL E, SHLENS J, MAZZOCCHI S, *et al*. YouTube-boundingboxes: A large high-precision human-annotated data set for object detection in video[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 7464–7473.

[15] HERMANS A, BEYER L, and LEIBE B. In defense of the triplet loss for person re-identification[EB/OL]. https://arxiv.org/abs/1703.07737, 2017.

[16] WU Yi, LIM J, and YANG M H. Online object tracking: A benchmark[C]. 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013: 2411–2418. doi: 10.1109/CVPR.2013.312.

[17] KRISTAN M, MATAS J, LEONARDIS A, *et al*. The visual object tracking VOT2015 challenge results[J]. 2015 IEEE International Conference on Computer Vision Workshop, Santiago, Chile, 2015: 564–586. doi: 10.1109/ICCVW.2015.79.

[18] SMEULDERS A W M, CHU D M, CUCCHIARA R, *et al*. Visual tracking: An experimental survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(7): 1442–1468. doi: 10.1109/TPAMI.2013.230.

[19] WANG Mengmeng, LIU Yong, and HUANG Zeyi. Large margin object tracking with circulant feature maps[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 4800–4808.

[20] ZHANG Jianming, MA Shugao, and SCLAROFF S. MEEM: Robust tracking via multiple experts using entropy minimization[C]. The 13th European Conference on Computer Vision, Zurich, Switzerland, 2014: 188–203.

[21] 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, Las Vegas, USA, 2016: 1401–1409.

[22] DANELLJAN M, HÄGER G, KHAN F S, *et al*. Learning spatially regularized correlation filters for visual tracking[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 4310–4318. doi: 10.1109/ICCV.2015.490.

[23] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84–90. doi: 10.1145/3065386.

[24] LI Bo, WU Wei, WANG Qiang, *et al*. SiamRPN++: Evolution of Siamese visual tracking with very deep networks[EB/OL]. https://arxiv.org/pdf/1812.11703.pdf, 2018.

侯志强：男，1973年生，教授，博士生导师，研究方向为图像处理、计算机视觉.

陈立琳：女，1989年生，硕士生，研究方向为计算机视觉、目标跟踪和深度学习.

余旺盛：男，1985年生，博士，研究方向为计算机视觉、图像处理，模式识别.

马素刚：男，1982年生，博士生，研究方向为计算机视觉、机器学习.

范九伦：男，1964年生，教授，博士生导师，研究方向为模式识别、图像处理.