基于空间和通道注意力机制的目标跟踪方法

刘嘉敏* 谢文杰 黄 鸿 汤一明

(重庆大学光电技术及系统教育部重点实验室 重庆 400044)

摘 要：目标跟踪是计算机视觉中重要的研究领域之一，大多跟踪算法不能有效学习适合于跟踪场景的特征限制了跟踪算法性能的提升。该文提出了一种基于空间和通道注意力机制的目标跟踪算法(CNNSCAM)。该方法包括离线训练的表观模型和自适应更新的分类器层。在离线训练时，引入空间和通道注意力机制模块对原始特征进行重新标定，分别获得空间和通道权重，通过将权重归一化后加权到对应的原始特征上，以此挑选关键特征。在线跟踪时，首先训练全连接层和分类器层的网络参数，以及边界框回归。其次根据设定的阈值采集样本，每次迭代都选择分类器得分最高的负样本来微调网络层参数。在OTB2015数据集上的实验结果表明：相比其他主流的跟踪算法，该文所提算法获得了更好的跟踪精度，重叠成功率和误差成功率分别为67.6%，91.2%。

关键词：目标跟踪；深度学习；空间注意力；通道注意力；在线学习

中图分类号：TN911.73; TP391.4 **文献标识码**：A **文章编号**：1009-5896(2021)09-2569-08

DOI: 10.11999/JEIT200687

Spatial and Channel Attention Mechanism Method for Object Tracking

LIU Jiamin XIE Wenjie HUANG Hong TANG Yiming

(*Key Laboratory of Optoelectronic Technique System of the Ministry of Education, Chongqing University, Chongqing* 400044, *China*)

Abstract: Object tracking is one of the important research fields in computer vision. However, most tracking algorithm can not effectively learn the features suitable for tracking scene, which limits the performance improvement of tracking algorithm. To overcome this problem, this paper proposes a target tracking algorithm based on CNN Spatial and Channel Attention Mechanisms (CNNSCAM). The method consists of an off-line training apparent model and an adaptive updating classifier layer. In the offline training, the spatial and channel attention mechanism module is introduced to recalibrate the original features, and the space and channel weights are obtained respectively. The key features are selected by normalizing the weights to the corresponding original features. In online tracking, the network parameters of the full connection layer and classifier layer are trained, and the boundary box regression is used. Secondly, samples are collected according to the set threshold, and the negative sample with the highest classifier score is selected for each iteration to fine tune the network layer parameters. The experimental results on OTB2015 dataset show that compared with other mainstream tracking algorithms, the proposed method achieves better tracking accuracy. The overlap success rate and error success rate are 67.6% and 91.2% respectively.

Key words: Object tracking; Deep learning; Spatial attention; Channel attention; Online learning

1 引言

视觉目标跟踪是计算机视觉的重要研究课题，目的是估计目标在各种场景下的位置，被广泛用于智能视频监控、自动驾驶、机器人导航、人机交互等领域[1,2]。目标跟踪算法流程主要包含目标初始化、目标表观建模、运动预测和目标定位，其中目标表观建模是算法的关键。在跟踪过程中，目标遮挡、旋转以及尺度变化等因素的影响，导致目标外观表示模型发生较大的变化，使得对运动目标的跟踪变得很困难[3]。因此提高模型对复杂背景的自适

收稿日期：2020-08-05；改回日期：2021-03-20；网络出版：2021-04-16
*通信作者： 刘嘉敏 liujm@cqu.edu.cn
基金项目：国家自然科学基金(41371338)、重庆市基础与前沿研究计划(cstc2018jcyjAX0093)、重庆市留学人员回国创业创新支持计划(cx2019144)、重庆市研究生科研创新项目(CYB19039, CYB18048)
Foundation Items: The National Natural Science Foundation of China (41371338), Chongqing Basic and Frontier Research Program (cstc2018jcyjAX0093), Chongqing Returned Overseas Students' Entrepreneurship and Innovation Support Program (cx2019144), Chongqing Graduate Research and Innovation Project (CYB19039, CYB18048)

应性和可分性是实现鲁棒跟踪的关键。

深度学习由于其强大的特征提取和表达能力，得到了学者的重视，从而在计算机视觉领域得到广泛使用。Hong等人[4]将预训练的CNN提取深度特征，再利用SVM方法进行跟踪。由于正负样本不均衡，Zhu等人[5]在DaSiamRPN算法中引入了detection的数据，模型的泛化性能得到了提升。Li等人[6]提出了回归loss和rank loss来最有效地表示当前目标的特征，并且将这些target-aware的特征与Siamese的框架相结合，减少了跟踪时使用的特征，加快了速度。Wang等人[7]提出了一种无监督跟踪算法，利用一个consistency loss来衡量forward和backward之间的差异来对网络进行训练，实现了无需标注的视频数据训练。上述方法对CNN提取目标深度特征进行了积极的探索，但大多只利用了原有的CNN网络提取特征。事实上，目标特征也有不同的重要程度，因此，为了有效地提取具有鉴别性的特征，需要利用注意力机制来关注目标中的重要特征。

近年来，视觉注意力机制被广泛应用于图像分类、语义分割等计算机视觉领域。Hu等人[8]提出的挤压和激励模块(Squeeze-and-Excitation，SE)可以使得网络关注通道之间的关系，利用网络自动学习到不同通道特征的重要程度，关注于重要特征通道，提高了图像分类的精度。在此基础上，Woo等人[9]提出了卷积注意力模块(Convolutional Block Attention Module, CBAM)，该模块在SE模块的基础上，引入了空间注意力机制，并且在考虑到max-pooling使得网络关注到重要通道特征的基础上，关注目标空间区域，使得网络在图像分类上的错分率更低，分类更加稳定。随着视觉注意力机制在其他领域取得了良好的结果，相关学者将注意力机制引入到了跟踪领域，Wang等人[10]基于孪生网络结构引入注意力机制，通过将CNN结构与残差注意力结构和通道注意力结构连接，并且与通道注意力叠加获得最后的注意力热图，从而提高了基于孪生网络追踪器的性能。Chen等人[11]提出的MAM(Multi-Attention Module)算法将多个注意力机制与LSTM(Long Short-Term Memory)模型结合，突出复杂背景中的目标信息。上述的实验成果证明了注意力机制在特征提取、抑制背景信息干扰方面有着良好的效果。

基于此，本文结合空间、通道注意力机制，提出了一种新的基于注意力机制的网络结构(CNN + Spatial and Channel Attention Modules, CNNSCAM)。本文在CNN网络第1层嵌入空间注意力模块，该模块能对特征图中每个位置的空间依赖性进行聚合，形成空间注意力图。在第2层和第3层之间引入通道注意力模块，帮助网络关注重要特征通道。实验证明，本文通过引入空间、通道注意力机制能使网络有效地抑制背景噪声，突出目标区域，更好地提取目标特征，提高了算法的跟踪效果。

2 模型框架

2.1 总体框架

本文研究所使用的神经网络结构如图1所示，其中Conv1～Conv3表示卷积单元，前两个卷积单元由卷积层、ReLU层、批归一化层和最大池化层组成，第3个卷积单元仅由卷积层和ReLU层组成，卷积层之间引入空间注意力模块(Spatial Attention Module, SAM)和通道注意力模块(Channel Attention Module, CAM)，Fc4和Fc5表示全连接层，f_x^+表示网络预测候选样本为目标的概率，f_x^-表示网络预测候选样本为背景的概率。

2.2 空间注意力机制

为了能从复杂的背景下区分出跟踪目标，需要对目标中的特征进行聚焦，增加特征之间的判别性。由此，本文引入空间注意力机制模块，如图2所示，用来赋予特征图不同位置的重要性，增强重要区域，抑制不重要的区域。

在本文的模型中，将Conv1输出的特征图作为空间注意力机制模块的输入特征图F。SAM通过全局最大池化和全局平均池化对输入特征的通道域特征进行了压缩，接着通过卷积将多通道特征压缩为

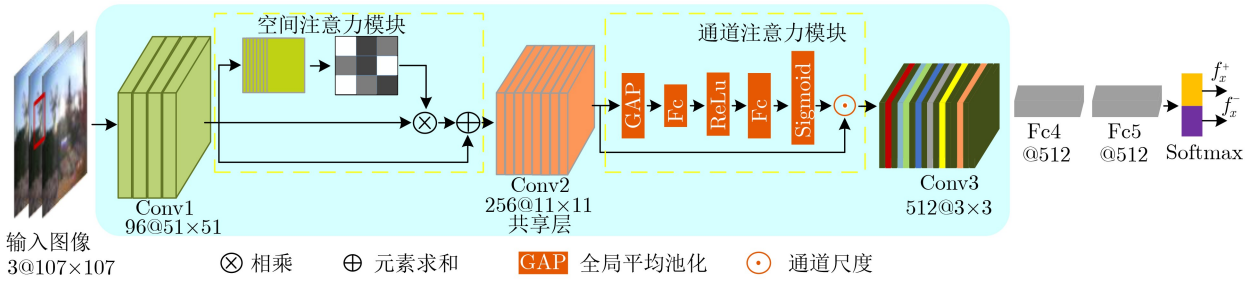

图 1 算法模型

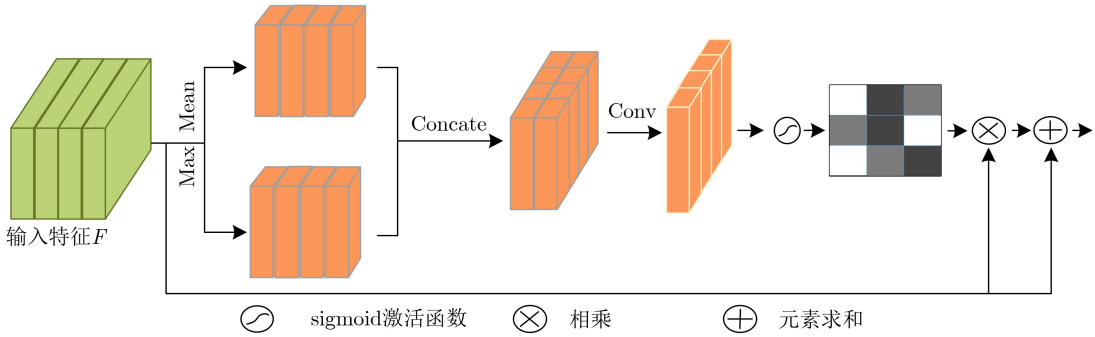

图 2 空间注意力机制

单通道，消除通道间信息分布对空间注意力机制的影响，然后通过激化函数归一化空间权重信息，最终将空间权重信息和输入特征图对应元素相乘，生成不同权重的特征图。

空间注意力模块的运算过程如式(1)

$$M_s(F)=\delta\left(f^{3\times3}[\text{AvgPool}(F);\text{MaxPool}(F)]\right) \quad (1)$$

其中，F为输入的特征图，δ表示sigmoid激活函数，f表示卷积层，卷积核大小为3×3, $[\text{AvgPool}(F);\text{MaxPool}(F)]$表示池化后的特征图，$M_s(F)$是一个空间注意力参数矩阵。

2.3 通道注意力机制

CNN卷积核可以被视为模式检测器，有些卷积核对颜色信息敏感，其他可能对物体的边缘响应高[11]。因此，在跟踪的过程中，需要选择与当前跟踪效果较好的卷积核，能更好地提取有用的通道特征，提高模型特征提取的能力。如图3所示，通道注意力模块包含3个部分：压缩模块、激励模块和注意力模块[8]。

压缩模块通过使用一个池化层，把每个通道内的全局空间特征信息进行求和压缩，形成各自的通道特征，该特征能够体现全局的通道特征信息，相当于扩大了网络的感受野。

激励模块是为了降低模块的参数数量同时增强模块的迁移能力，模块采用两个全连接层得到各自层的权重参数W_0和W_1，在模块的训练过程中可以学习得到每个通道域的特征权重和通道之间的相关性。

注意力模块在每个通道域上对得到的特征权重与原卷积相应的通道特征值进行加权融合，可以使得卷积通道特征表现出不同的权重，从而提取出表征目标中的关键信息，具体如式(2)所示

$$\begin{aligned}M_c(F)&=\sigma\left(\text{MLP}\left(\text{AvgPool}(F)\right)\right)\\&=\sigma\left(W_1\left(W_0(F^c_{\text{avg}})\right)\right)\end{aligned} \quad (2)$$

其中，σ为sigmoid激活函数，MLP表示3层感知机，F^c_{avg}表示平均池化特征，$W_0\in\mathbb{R}^{C/r\times C}$，

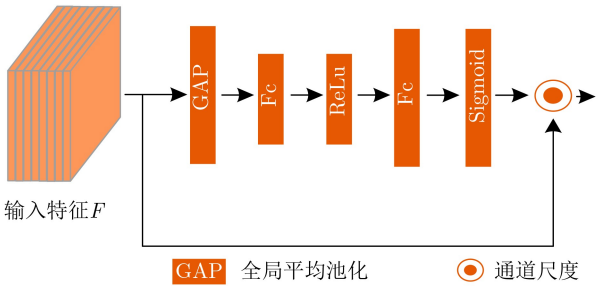

图 3 通道注意力机制

$W_1\in\mathbb{R}^{C\times C/r}$, r表示减少率，C是特征通道维度的大小，$M_c(F)$是一个通道注意力参数矩阵。

2.4 边界回归

在目标检测中，边界回归方法得到了广泛的运用，目的是在输出检测目标位置的基础上利用一种回归的方式调整检测框位置，提高检测的效果[12]。本文出于同样的目的，引入边界回归对当前网络预测的目标位置进行微调来增加目标跟踪框的精度。在本文实验中，利用视频序列给定的第1帧目标的位置信息，在其周围采集一定量的样本和给定的目标位置训练一个线性回归模型。

2.5 离线训练

本文提出的模型需要进行离线训练，其中Conv1～Conv3中的网络参数利用预训练的VGG-M网络参数进行初始化，空间和通道注意力机制模块、Fc4-Fc5层和Softmax层参数进行随机初始化。在训练的过程中，利用SGD优化的方法进行训练，采用了K个视频序列作为数据集，在第$k\in K$个序列中随机选择一定量的视频帧，并从每一帧中的目标周围采集一定量的正样本和负样本组成一个批量送入模型进行训练。通过不断的重复训练直到网络已经收敛或达到设定的训练次数。

当完成了对整个模型的离线训练时，Conv1～Conv3卷积层和两个注意力机制模块参数在跟踪的过程中作为共享层，参数需要保持不变，而Fc4-Fc5和Softmax层需要进行训练，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。

2.6 在线跟踪样本采集

为了保证在跟踪过程中分类器的性能，需要在跟踪的过程对分类器输出概率较好视频帧进行采集，便于后序模型参数的微调。在所提算法中设置采集阈值T_H=0.7，在开始跟踪时，设置集合B用于存储输出概率大于T_H的视频帧，当B中的数量超过T_s帧时，利用B中的视频帧采集正负样本分别存放于正样本集S^+和负样本集S^-。

2.7 模型更新

为了保证模型能适应复杂环境下目标的变化，在跟踪过程中需要对模型参数进行更新。在本文中，使用SGD的方法来更新Fc4-Fc5和Softmax层的参数。当触发模型更新时，设定K'次迭代，在迭代中需要从正样本集S^+和负样本集S^-中随机选择n^+个正样本和分类器响应值最高的n^-个负样本组成一个输入批次。将输入批次送入模型后，利用SGD的方法反向更新Fc4-Fc5和Softmax层的参数，在迭代循环结束之后，清除集合B, S^+, S^-内存，为下次采集更新样本做准备。

3 实验与分析

3.1 实验配置

本文在OTB数据集上对本文的算法进行了评估，并将该算法与近年来的一些主流跟踪算法进行了比较。本文算法模型使用Python3.6和Pytorch深度训练框架进行编程，在配备Intel(R) Core(TM) i5-9400F(2.9 GHz), NVIDIA RTX2060的机器上能以2.9 fps的速度运行。

3.2 数据集和评价标准

Object Tracking Benchmark(OTB)[13]数据集是单目标跟踪领域的视频基准库，包含OTB50, OTB2013, OTB2015。OTB数据集包含了目标跟踪中常见的难点，包括光照变化、尺度变化、遮挡、形变、运动模糊等11个方面。OTB2015含有100个跟踪视频，包含了OTB50和OTB2013，因此，后序的实验只在OTB2015上进行。

为了全面地评价算法的性能，对跟踪结果使用以下两种方法进行评估：

(1)距离误差成功率：在跟踪过程中算法预测的目标位置与标签位置之间的距离误差小于设置阈值的数量与视频序列的总数量之比。

(2)重合度成功率：在跟踪过程中算法预测的目标位置与标签位置之间的重合度小于设置阈值的数量与视频序列的总数量之比。

3.3 实验参数

(1)候选样本采集：为了在每一帧采集候选样本，利用高斯分布$p(m_t|m_{t-1})=N(m_t;m_{t-1},\Sigma)$，以前一帧$m_{t-1}$的坐标为中心，在$m_t$帧采集256个候选样本。协方差$\sum$=(0.09$r^2$, 0.09$r^2$, 0.25)，其中$r$表示$x_{t-1}$宽度和高度的平均值。

(2)训练数据：本文利用多个场景下的视频序列来微调CNN网络和训练CAM, SAM，离线训练的过程中，输入的样本大小为107×107×3，首先利用候选样本采集的策略，随机在每个序列目标周围选择4个正样本，此时正样本与目标的重叠率大于0.7，总共选取8帧，组成32个正样本，同理，保证负样本与目标的重叠率小于0.5，生成96个负样本。生成的正样本和负样本组成一个批次送入网络中进行训练。在线更新中，T_H=0.7, T_s=10，在每一帧中会采集50个正样本(重叠率大于0.7)，200个负样本(重叠率小于0.3)。为了适应不同的跟踪序列，本文在第一帧会采集500个正样本和5000个负样本，训练Fc4-Fc5和Softmax层的参数，同理采集1000个正样本训练边界回归模型。

(3)网络训练参数：本文中的离线训练迭代次数为10^5，卷积层的学习率为0.0001，SAM和CAM、全连接层的学习率为0.001。在跟踪的时候，在第1帧采集数据迭代30次，微调Fc4-Fc5和Softmax层参数。而其他帧进行更新的时候迭代次数K'=15。

3.4 模型分析实验

为了评估本文所嵌入的空间和通道模块对跟踪精度的贡献，本文进行了对比实验，如图4所示，其中SG(Single Method)表示没有加任何模块、SG-SAM表示了在Conv1和Conv2之间嵌入了空间注意力机制模块、SG-CAM表示在SG的基础上的Conv2和Conv3之间嵌入了通道注意力机制模块、CNNSCAM表示本文算法在SG的基础上同时引入了SAM和CAM模块。

实验结果表明，CNNSCAM引入了空间和通道注意力机制能有效地提高算法的跟踪精度。这是因为通道注意力机制能给目标区域赋予更高的权重，增加了网络的表征能力，同时引入的通道注意力机制可以根据跟踪的场景选择适合跟踪场景的通道特征，提高了网络提取特征的鲁棒性。因此，结合空间和通道注意力机制能有效地提高算法跟踪的精度。

3.5 性能对比实验

为了验证本文算法的性能，选取了5个主流的算法在OTB2015数据集上进行了对比。对比算法分别是：DaSiamRPN[5], TADT[6], MCPF[14], CNN-SVM[4], BACF[15]。整体的实验精度如图5所示，本文算法取得了较好的跟踪效果。

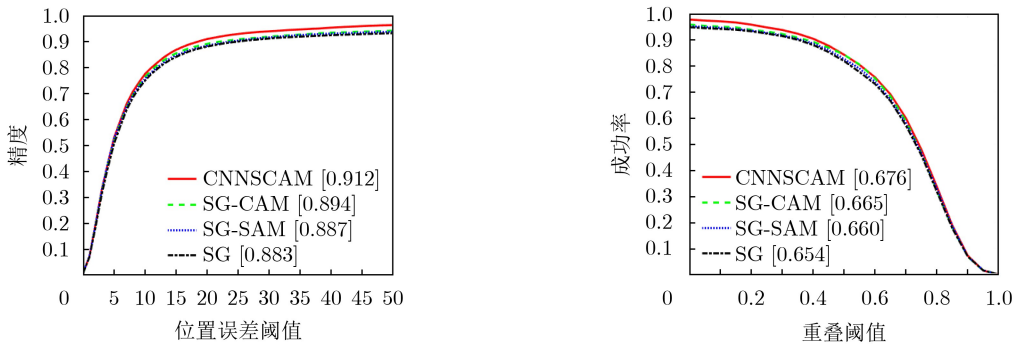

图 4 在OTB2015数据集上网络嵌入CAM, SAM的精度和重合度成功率

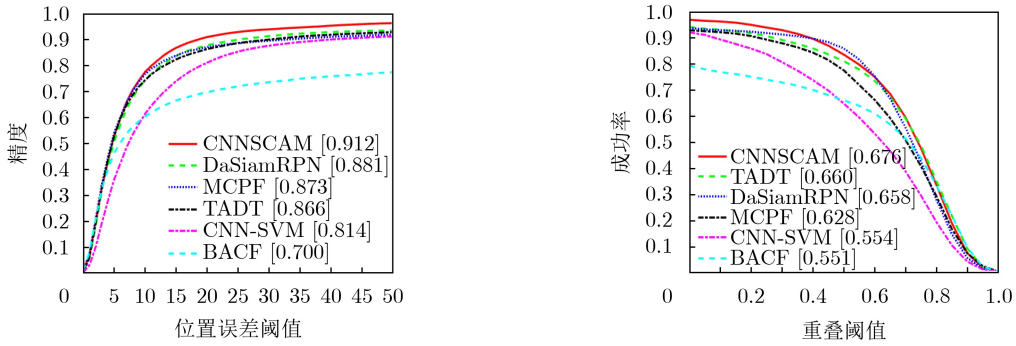

图 5 算法在OTB2015数据集上的整体精度和成功率

由图5可知，MCPF, BACF会导致训练模型的判别性不足，产生跟踪漂移，这是因为这几种算法归类于传统算法，而传统算法多采用手工语义特征，如像素、颜色、HOG和融合特征等特征来构建目标整体的表观模型，在目标被遮挡、发生形变时，训练样本引入背景像素，影响模型性能。DaSiamRPN, TADT, CNN-SVM属于深度跟踪算法，即利用CNN搭建的模型实现图像特征的提取，但是原始的CNN模型提取的特征含有较多的空间和通道特征冗余度，当进行特征对比分类的时候不能保证较好的特征鉴别性和区分性，而本文在传统的卷积网络中嵌入了空间和通道注意力机制，使得离线训练的表观模型在复杂场景下具有更好的鲁棒性，在提取的特征上更加有鉴别性，保证了更好的跟踪精度。

为进一步评估所提算法的跟踪效果，本文还对OTB2015中的11个跟踪场景进行了对比，分别是光照变化(IV)、平面外旋转(OPR)、尺度变化(SV)、遮挡(OCC)、运动模糊(MD)、快速移动(FM)、平面内旋转(IPR)、视野外(OV)、形变(DEF)、背景集群(BC)和低分辨率(LR)。表1和表2分别总结了各个算法的重叠成功率和距离误差成功率(其中加粗的分数表示该算法在挑战因素中排名第一)。

通过分析表1和表2，本文算法在多个跟踪场景中都取得了良好的跟踪效果，特别是在背景干扰(BC)，尺度变化(SV)，视野外(OV)等挑战因素下跟踪精度更高。这是因为本文算法在目标缺失、背景干扰的情况下，通过注意力机制可以分配网络中的参数权重，关注目标中重要的信息，提高了模型获取特征的能力，保证了跟踪效果。

图6展示了本文算法与其他4个跟踪算法在5个视频序列定性评估的跟踪结果。

表 1　在OTB2015数据集中的11个跟踪场景下算法的重叠成功率

	IV	OPR	SV	OCC	MD	FM	IPR	OV	DEF	BC	LR
CNNSCAM	0.680	**0.657**	**0.663**	**0.644**	**0.671**	**0.658**	**0.660**	**0.651**	0.631	**0.675**	0.622
DaSiamRPN	0.662	0.644	0.641	0.617	0.625	0.621	0.652	0.537	**0.652**	0.642	0.588
TADT	**0.681**	0.646	0.655	0.643	**0.671**	0.657	0.621	0.625	0.607	0.622	**0.634**
MCPF	0.629	0.619	0.604	0.620	0.599	0.597	0.620	0.553	0.569	0.601	0.581
CNN-SVM	0.537	0.548	0.489	0.514	0.578	0.546	0.548	0.488	0.547	0.548	0.403
BACF	0.547	0.506	0.532	0.475	0.541	0.511	0.497	0.483	0.499	0.552	0.502

表 2 在OTB2015数据集中的11个跟踪场景下算法的距离误差成功率

Attribute	IV	OPR	SV	OCC	MD	FM	IPR	OV	DEF	BC	LR
CNNSCAM	**0.905**	**0.901**	**0.910**	**0.862**	**0.862**	**0.869**	**0.910**	**0.864**	0.880	**0.927**	0.889
DaSiamRPN	0.878	0.878	0.858	0.818	0.820	0.819	0.889	0.720	**0.887**	0.856	0.814
TADT	0.865	0.872	0.863	0.842	0.833	0.834	0.832	0.816	0.822	0.805	0.881
MCPF	0.882	0.816	0.862	**0.862**	0.840	0.845	0.888	0.764	0.815	0.823	**0.911**
CNN-SVM	0.792	0.798	0.785	0.727	0.751	0.747	0.813	0.650	0.791	0.776	0.811
BACF	0.665	0.650	0.673	0.590	0.649	0.627	0.645	0.613	0.655	0.700	0.665

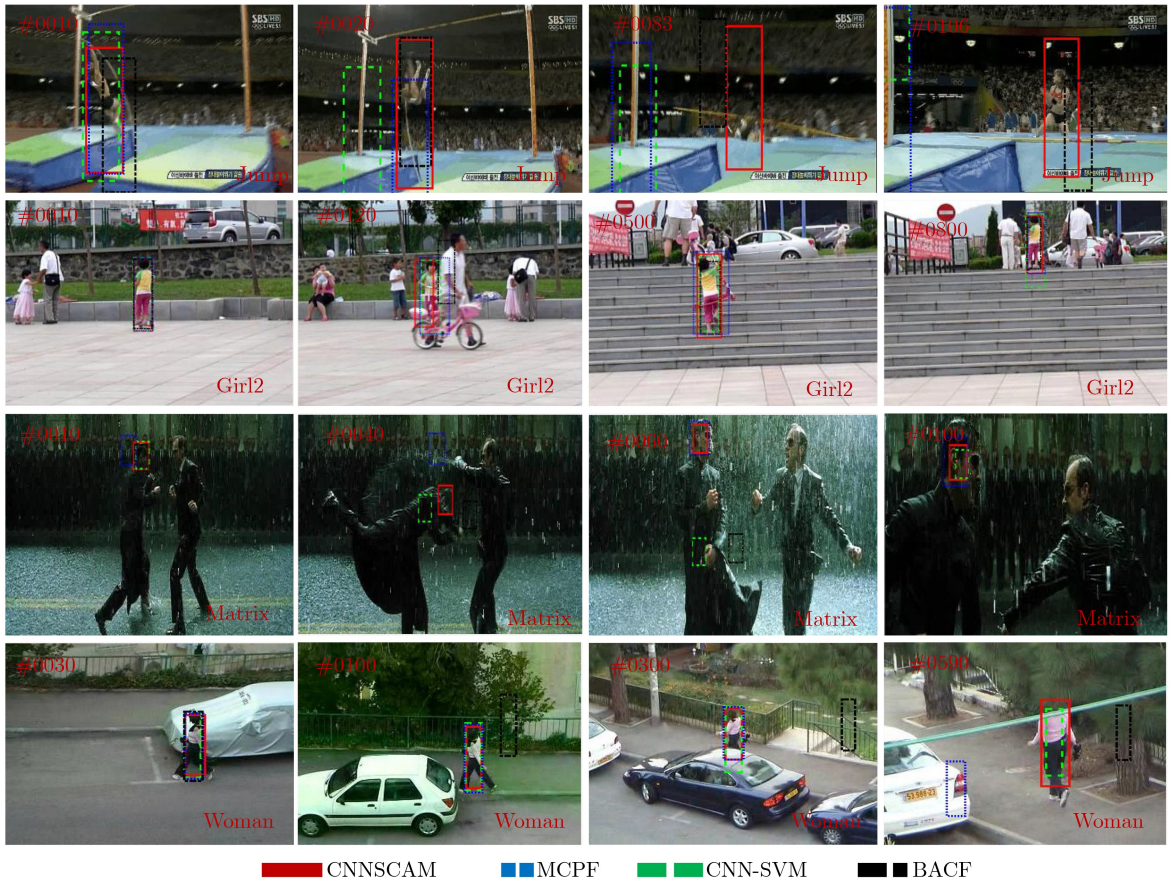

图 6 多个序列中部分跟踪结果

(1) Jump：目标的视角和姿态的变换，导致目标发生尺度和旋转变化，使得目标引入了一定的背景干扰信息，导致模型的判别性不足，产生跟踪漂移，而本文算法利用空间和通道注意力机制可以提高表观模型的抗干扰能力和提取特征的能力，提取的特征更加具有鲁棒性。

(2) Girl2和Women：主要是目标发生了完全遮挡或部分遮挡和尺度变化，其中CNN-SVM相对于本文算法不能提取有效的特征，没有在线更新也导致跟踪效果较差；BACF没有考虑多尺度问题，导致算法提取到的特征并不能完全表示目标的整体特征；MCPF利用粒子滤波能较好地解决目标尺度变化的问题，但是没有采用CNN提取特征，在目标被遮挡时，提取的特征能力有限，导致跟踪漂移，而本文的算法结合空间和通道注意力机制，在图像特征提取上更能关注到目标显著特征，并且采用在线微调网络层参数保证在线分类器的性能，因此，本文算法的跟踪效果更好。

(3) Matrix：其他算法都发生了不同程度上的跟踪漂移，特别是在第40帧，人物的头部发生快速移动和遮挡的时候，只有本文算法跟踪正确，实验证明本文算法在复杂的场景下有着更好的鲁棒性。

3.6 影响目标跟踪定位性能的主要参数

在本文算法中，在采集候选样本时，利用高斯分布 $p(m_t|m_{t-1})=N(m_t;m_{t-1},\Sigma)$，以前一帧

m_{t-1}的坐标为中心，在m_t帧采集候选样本。其具体高斯函数为$f(x_t,y_t)=A\exp\left(-\left(\frac{(x_t-x_{t-1})^2}{2\sigma_{x_t}^2}+\frac{(y_t-y_{t-1})^2}{2\sigma_{y_t}^2}\right)\right)$，其中$(x_t, y_t)$表示第$m_t$帧的中心位置；$(x_{t-1}, y_{t-1})$表示第$m_{t-1}$帧的中心位置，$A$表示幅值，$\sigma_{x_t}$, σ_{y_t}表示方差。在本算法中高斯函数的方差σ_{x_t}, σ_{y_t}取同一个值并记为v。幅值和方差决定生成样本的位置尺度范围，是影响跟踪定位的主要参数。以下就两者的取值进行了实验分析。

首先固定v=1.00, A从0.1到1进行取值，从表3可知，当A=0.6的时候，距离误差成功率Prec取得最优值0.912。

接着，也对候选样本生成的边框大小进行讨论，此时固定A=0.6, v从1.00到1.10进行取值。分析表4可知，当v=1.05的时候，重合度成功率取得最优值0.676。由实验可知，当A=0.6, v=1.05的时候，所提算法的跟踪精度最好。

表 3　在OTB2015数据集中固定v=1.00时，不同A取值的距离误差成功率

A取值	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.00
Prec	0.686	0.770	0.834	0.850	0.877	**0.912**	0.876	0.886	0.875	0.858

表 4　在OTB2015数据集中固定A=0.6时，不同v取值的距离误差成功率

v取值	1.00	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09	1.10
Suc	0.599	0.621	0.641	0.651	0.661	**0.676**	0.671	0.666	0.657	0.643	0.622

4　总结

针对复杂环境中目标外观变化大难以跟踪的问题，本文提出了一种空间和通道双注意力机制的跟踪方法，通过引入空间和通道注意力机制离线训练得到算法的表观模型。实验结果表明，该表观模型能有效地抑制背景噪声、突出目标区域，具有更好的鲁棒性。在跟踪时，采用了在线分类的方式定位目标，同时根据跟踪的结果在线微调全连接层参数保证了网络的泛化性能，在OTB2015数据集上与主流的算法进行了对比，实验结果表明：本文算法相对其他算法跟踪性能更稳定，在多种复杂场景下(包括旋转、尺度变化及视野外)有着更好的跟踪精度。

虽然本文算法在OTB2015数据集上取得较好的性能，但是距离实现实时跟踪还有很大的差距，因此考虑在下一步工作中，把注意力机制运用到孪生网络中，采用一种端对端的训练方式在多个数据集上训练，提高算法的跟踪精度和跟踪速度。

参 考 文 献

[1] 蒲磊, 冯新喜, 侯志强, 等. 基于自适应背景选择和多检测区域的相关滤波算法[J]. 电子与信息学报, 2020, 42(12): 3061–3067. doi: 10.11999/JEIT190931.
PU Lei, FENG Xinxi, HOU Zhiqiang, *et al*. Correlation filter algorithm based on adaptive context selection and multiple detection areas[J]. *Journal of Electronics & Information Technology*, 2020, 42(12): 3061–3067. doi: 10.11999/JEIT190931.

[2] 李康, 李亚敏, 胡学敏, 等. 基于卷积神经网络的鲁棒高精度目标跟踪算法[J]. 电子学报, 2018, 46(9): 2087–2093. doi: 10.3969/j.issn.0372-2112.2018.09.007.
LI Kang, LI Yamin, HU Xuemin, *et al*. A robust and accurate object tracking algorithm based on convolutional neural network[J]. *Acta Electronica Sinica*, 2018, 46(9): 2087–2093. doi: 10.3969/j.issn.0372-2112.2018.09.007.

[3] 王鹏, 孙梦宇, 王海燕, 等. 一种目标响应自适应的通道可靠性跟踪算法[J]. 电子与信息学报, 2020, 42(8): 1950–1958. doi: 10.11999/JEIT190569.
WANG Peng, SUN Mengyu, WANG Haiyan, *et al*. An object tracking algorithm with channel reliability and target response adaptation[J]. *Journal of Electronics & Information Technology*, 2020, 42(8): 1950–1958. doi: 10.11999/JEIT190569.

[4] HONG S, YOU T, KWAK S, *et al*. Online tracking by learning discriminative saliency map with convolutional neural network[C]. Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015: 597–606.

[5] ZHU Zheng, WANG Qiang, LI Bo, *et al*. Distractor-aware Siamese networks for visual object tracking[C]. Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 2018: 104–119. doi: 10.1007/978-3-030-01240-3_7.

[6] LI Xin, MA Chao, WU Baoyuan, *et al*. Target-aware deep tracking[C]. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 1369–1378. doi: 10.1109/CVPR.2019.00146.

[7] WANG Ning, SONG Yibing, MA Chao, *et al*. Unsupervised deep tracking[C]. Proceedings of 2019 IEEE/CVF

Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 1308–1317. doi: 10.1109/CVPR.2019.00140.

[8] HU Jie, SHEN Li, and SUN Gang. Squeeze-and-excitation networks[C]. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132–7141. doi: 10.1109/CVPR.2018.00745.

[9] WOO S, PARK J, LEE J Y, *et al*. CBAM: Convolutional block attention module[C]. Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 2018: 1352–1368. doi: 10.1007/978-3-030-01234-2_1.

[10] WANG Qiang, TENG Zhu, XING Junliang, *et al*. Learning attentions: Residual attentional Siamese network for high performance online visual tracking[C]. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4854–4863. doi: 10.1109/CVPR.2018.00510.

[11] CHEN Boyu, LI Peixia, CHONG Sun, *et al*. Multi attention module for visual tracking[J]. *Pattern Recognition*, 2019, 87: 80–93. doi: 10.1016/j.patcog.2018.10.005.

[12] 张文明, 姚振飞, 高雅昆, 等. 一种平衡准确性以及高效性的显著性目标检测深度卷积网络模型[J]. 电子与信息学报, 2020, 42(5): 1201–1208. doi: 10.11999/JEIT190229.
ZHANG Wenming, YAO Zhenfei, GAO Yakun, *et al*. A deep convolutional network for saliency object detection with balanced accuracy and high efficiency[J]. *Journal of Electronics & Information Technology*, 2020, 42(5): 1201–1208. doi: 10.11999/JEIT190229.

[13] 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.

[14] ZHANG Tianzhu, XU Changsheng, and YANG M H. Multi-task correlation particle filter for robust object tracking[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 4819–4827. doi: 10.1109/CVPR.2017.512.

[15] GALOOGAHI H K, FAGG A, and LUCEY S. Learning background-aware correlation filters for visual tracking[C]. Proceedings of 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 1144–1152. doi: 10.1109/ICCV.2017.129.

刘嘉敏：男，1973年生，副教授，研究方向为图像处理、模式识别.

谢文杰：男，1995年生，硕士生，研究方向为图像处理、视频跟踪.

黄 鸿：男，1980年生，教授，研究方向为流形学习、模式识别和遥感影像智能化处理.

汤一明：男，1993年生，博士生，研究方向为模式识别、图像处理、深度学习和视觉跟踪.

责任编辑：陈 倩