

基于二阶池化特征融合的孪生网络目标跟踪算法

陈茂林^{1,2}, 侯志强^{1,2}, 余旺盛³, 马素刚^{1,2}, 蒲磊⁴

(1. 西安邮电大学计算机学院, 西安, 710121; 2. 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; 3. 空军工程大学信息与导航学院, 西安, 710077; 4. 火箭军工程大学作战保障学院, 西安, 710025)

摘要 为提升基于孪生网络目标跟踪算法的特征表达能力, 获得更好的跟踪性能, 提出了一种轻量级的基于二阶池化特征融合的孪生网络目标跟踪算法。首先, 使用孪生网络结构获取目标的深度特征; 然后, 在孪生网络结构的末端并行添加二阶池化网络和轻量级通道注意力, 以获取目标的二阶池化特征和通道注意力特征; 最后, 将目标的深度特征、二阶池化特征和通道注意力特征进行融合, 使用融合后的特征进行互相关操作, 得到地响应图能很好地区分目标和背景, 提高跟踪模型的判别能力, 改善目标定位的精度, 从而提升跟踪性能。所提算法使用 Got-10k 数据集进行端到端的训练, 并在数据集 OTB100 和 VOT2018 上进行验证。实验结果表明, 所提算法与基准算法相比, 跟踪性能取得了显著提升: 在 OTB100 数据集上, 精确度和成功率分别提高了 7.5% 和 5.2%; 在 VOT2018 数据集上, 预期平均重叠率(EAO)提高了 4.3%。

关键词 目标跟踪; 孪生网络; 二阶池化网络; 通道注意力

DOI 10.3969/j.issn.1009-3516.2022.03.011

中图分类号 TP391.4 **文献标志码** A **文章编号** 1009-3516(2022)03-0068-07

Siamese Network Target Tracking Algorithm Based on Second-Order Pooling Feature Fusion

CHEN Maolin^{1,2}, HOU Zhiqiang^{1,2}, YU Wangsheng³, MA Sugang^{1,2}, PU Lei⁴

(1. College of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 2. Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; 3. Information and Navigation School, Air Force Engineering University, Xi'an 710077, China; 4. School of Operational Support, Rocket Force Engineering University, Xi'an 710025, China)

Abstract In order to improve the feature expression ability of the target tracking algorithm based on Siamese network and obtain better tracking performance, a lightweight Siamese network target tracking algorithm based on second-order pooling feature fusion is proposed. First, the Siamese network architecture is used to obtain the deep features of the target; then, the second-order pooling network and the lightweight channel attention are added in parallel at the end of the Siamese network architecture to obtain the second-order pooling features and channel attention features of the target, respectively. Finally, the depth feature of the target, the second-order pooling feature and the channel attention feature are fused, and the fused

收稿日期: 2021-09-08

基金资助: 国家自然科学基金(62072370)

作者简介: 陈茂林(1996—), 男, 山东枣庄人, 硕士生, 研究方向为目标跟踪。E-mail: maolin_chen@126.com

引用格式: 陈茂林, 侯志强, 余旺盛, 等. 基于二阶池化特征融合的孪生网络目标跟踪算法[J]. 空军工程大学学报(自然科学版), 2022, 23(3): 68-74. CHEN Maolin, HOU Zhiqiang, YU Wangsheng, et al. Siamese Network Target Tracking Algorithm Based on Second-Order Pooling Feature Fusion [J]. Journal of Air Force Engineering University (Natural Science Edition), 2022, 23(3): 68-74.

feature is used for cross-correlation operation, and the obtained response graph can distinguish the target and the background well, and improve the discriminative ability of the model, and improve the accuracy of target positioning, thereby improving target tracking performance. The proposed algorithm only uses the Got-10k dataset for end-to-end training and is validated on the OTB100 and VOT2018 datasets. The experimental results show that the proposed algorithm achieves a significant improvement in tracking performance compared with the benchmark algorithm SiamFC: on the OTB100 dataset, the accuracy and success rate are increased by 7.5% and 5.2%, respectively; on the VOT2018 dataset, the expected average overlap rate increases by 4.3%.

Key words target tracking; siamese network; second-order pooling network; channel attention

目标跟踪是计算机视觉领域中最重要研究方向之一^[1],是指在视频初始帧中给出目标,在后续帧中对目标进行状态估计和精准定位的过程^[3],在安防、监控和巡检等领域有着广泛的应用^[4-5]。

目前,要实现鲁棒的目标跟踪器依然面临着诸多困难,主要包括相似背景、尺度变化和目标遮挡等^[6]。Wang 等^[7]通过实验表明:良好的特征表达是实现鲁棒的目标跟踪系统的关键因素之一。众多学者也先后设计了灰度、颜色和方向梯度直方图等手工特征,在简单场景中具有良好的性能表现,但是在复杂场景中性能不佳。近年来,卷积神经网络(convolutional neural networks, CNN)^[8]在计算机视觉领域中得到了广泛应用,并在图像分类、目标检测和图像分割等任务中获得大幅度的性能提升。Bertinetto 等^[9]提出全卷积的孪生网络目标跟踪算法 SiamFC,将目标跟踪问题转化为相似性匹配问题,取得了良好的跟踪性能。Li 等人^[10]提出 SiamRPN 算法,将孪生网络与区域建议网络结合,实现了灵活应对目标尺度的变化。Zhu 等^[11]提出 Da-SiamRPN 算法,通过引入干扰感知模块,增加训练时的负样本,使模型有效地捕捉更多上下文信息以适应目标外观变化。

但是,孪生网络对同类目标中特定目标的特征表达能力不足,不能很好地将特定目标从同类目标中区分出来,同时,由于基于孪生网络的目标跟踪算法大多数都只使用了一阶图像信息表示,这严重限

制了孪生网络的非线性建模能力,因此需要学习高阶统计信息表示来增强非线性建模能力^[12]。高阶统计信息能够很好地描述同类目标中特定目标的信息,并在图像分类和细粒度图像分类等任务中得到应用。

本文针对基于孪生网络目标跟踪算法对相似目标表达能力不足的问题,提出一种轻量级的基于二阶池化特征融合的孪生网络目标跟踪算法。

1 本文算法

1.1 总体框架

本文算法框架见图 1。该算法使用了结构相同并且权值共享的孪生网络结构,其主干网络是 AlexNet,该算法将视频序列的第 1 帧作为模板帧,后续帧作为搜索帧^[13]。首先,使用孪生网络结构获取模板帧目标的特征和搜索帧中候选样本的深度特征;然后,在孪生网络结构末端并行添加二阶池化网络和轻量级通道注意力,获取目标的二阶池化特征和通道注意力特征,其中二阶池化特征是对深度特征的细化和增强,可以提高对相似目标的判别能力;通道注意力特征更关注目标的语义信息,最后,将目标的深度特征、二阶池化特征和通道注意力特征以 Add 方式进行融合,使用融合后的特征进行互相关操作,得到的响应图能很好地区分目标和背景,提高模型的判别能力,改善目标定位的精度,从而提升跟踪性能。

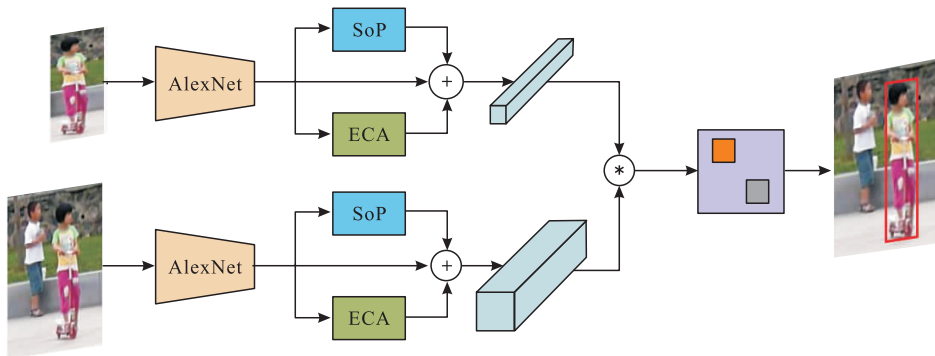


图 1 算法框架图

1.2 本文基准算法

SiamFC^[9]分为离线训练和在线评估两个阶段。

1) 离线训练阶段: 首先, 对训练集中的视频帧进行裁剪缩放, 得到以目标为中心, 大小为 127×127 的模板图像 z 和 255×255 的搜索图像 x ; 然后, 使用 AlexNet 端到端的训练相似度匹配函数 $f(z, x)$:

$$f(z, x) = \varphi(z) * \varphi(x) + b_1 \quad (1)$$

式中: φ 表示卷积嵌入函数; $*$ 表示卷积互相关操作; b_1 表示得分图中各个位置的不同偏置项。

在线评估阶段: 以目标的初始外观作为模板, 使用相似度匹配函数 $f(z, x)$ 评估模板帧与当前帧待

搜索图像的相似性, 经过深度互相关操作计算相似性得分图, 分数越高, 二者相似度越高; 从得分图中寻找分数最高的候选区域并乘以相应的步长, 得到目标在当前帧与模板帧之间的偏移量, 从而确定目标在当前帧中的位置。

1.3 二阶池化网络

Gao 等人^[14]指出二阶池化网络能够显著增强卷积神经网络模型的非线性拟合能力, 受此工作启发, 本文将二阶池化网络应用到 SiamFC 算法中, 所使用的二阶池化网络如图 2 所示, 包括挤压模块(squeeze module)和激励模块(excitation module)。

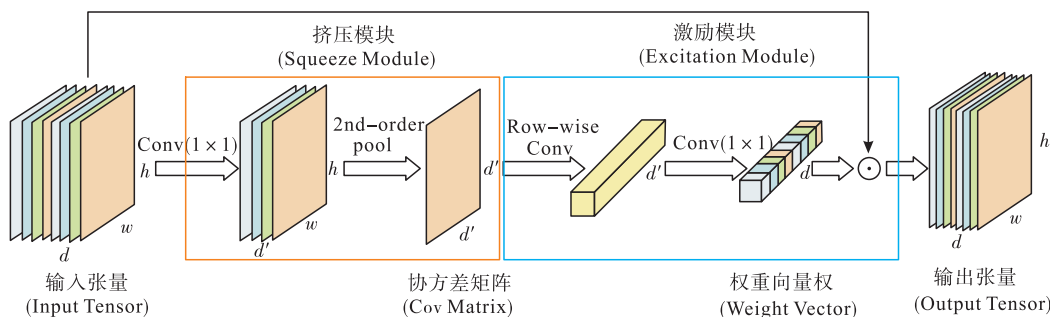


图 2 SoP 模型

挤压模块的目的是在通道维度上对输入张量的信息进行二阶统计建模, 获取具有通道相关性的协方差矩阵。假设特征提取网络最后一层的输出张量 $\mathbf{X}' \in \mathbf{R}^{w \times h \times d}$, 其中 d 代表特征通道数, w 和 h 为特征图的宽和高。为了降低后续计算开销, 采用卷积进行降维处理, 得到张量 $\mathbf{X}'' \in \mathbf{R}^{w \times h \times d'}$, 并进行二阶池化操作, 其过程如图 3 所示。

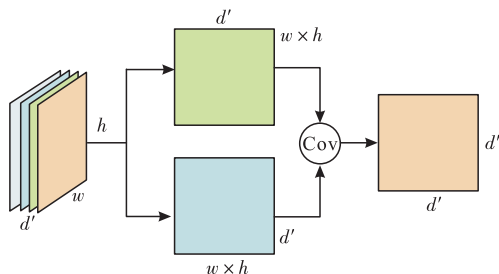


图 3 二阶池化操作

首先将 \mathbf{X}' 进行两次 Reshape 操作, 得到 $\mathbf{X} \in \mathbf{R}^{d' \times (w \times h)}$ 和 $\mathbf{X}^T \in \mathbf{R}^{(w \times h) \times d'}$ 的特征矩阵, 然后计算特征图 \mathbf{X}' 的协方差矩阵:

$$\mathbf{P} = \mathbf{X} \bar{\mathbf{I}} \mathbf{X}^T \quad (2)$$

其中: $\bar{\mathbf{I}} = \frac{1}{N} (\mathbf{I} - \frac{1}{N} \mathbf{I} \mathbf{I}^T)$, $\mathbf{I} = [1, \dots, 1]$ 为一个 N 维的向量, 接着采用特征值分解的方式对得到的协方差矩阵进行处理, 获得其特征值和特征向量:

$$\mathbf{P} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T \quad (3)$$

式中: $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 为对角矩阵; λ_i 为特征值; $\mathbf{U} = [u_1, u_2, \dots, u_n]$, 为对应的特征向量。通过上述的特征值分解可以将对矩阵的幂运算转换为对特

征值的幂运算进行求解:

$$\mathbf{P} \triangleq \mathbf{U} \mathbf{F}(\mathbf{\Lambda}) \mathbf{U}^T \quad (4)$$

在本文中 $\alpha = 0.5$, $\mathbf{F}(\mathbf{\Lambda}) = \text{diag}(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n))$, 其中 $f(\lambda_i)$ 表示对特征值进行幂运算:

$$f(\lambda_i) = \lambda_i^\alpha \quad (5)$$

协方差矩阵是为获得各通道中像素间的相关性, 描述如下:

$$\begin{bmatrix} \text{cov}(\mathbf{X}_1, \mathbf{X}_1) & \text{cov}(\mathbf{X}_1, \mathbf{X}_2) & \cdots & \text{cov}(\mathbf{X}_1, \mathbf{X}_{d'}) \\ \text{cov}(\mathbf{X}_2, \mathbf{X}_1) & \text{cov}(\mathbf{X}_2, \mathbf{X}_2) & \cdots & \text{cov}(\mathbf{X}_2, \mathbf{X}_{d'}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\mathbf{X}_{d'}, \mathbf{X}_1) & \text{cov}(\mathbf{X}_{d'}, \mathbf{X}_2) & \cdots & \text{cov}(\mathbf{X}_{d'}, \mathbf{X}_{d'}) \end{bmatrix} \quad (6)$$

式中: \mathbf{X}_i 表示第 i 个通道的特征图, $i \in [1, d']$ 。

激励模块的目的是在通道维度上对输入张量的信息进行重新标定, 获取具有通道相关性的输出张量。为保持输入张量的原始结构信息, 首先对 $d' \times d'$ 协方差矩阵进行逐行卷积操作, 得到一个 $d' \times 1$ 的向量; 然后使用卷积和非线性激活函数 Sigmoid, 得到一个 $d' \times 1$ 的权重向量; 最后执行权重向量和通道之间的点积操作, 得到具有通道相关性的输出张量, 该张量是对原始深度特征的细化, 对相似目标具有良好的判别能力。

2 实验

本文算法使用的实验平台为 Ubuntu 16.04,

GPU 为 1080 Ti,采用 PyTorch1.6 实现。所提算法仅使用 Got-10k^[15]数据集进行端到端的训练,在训练过程中,batch_size 设置为 8,学习率从 0.01 衰减到 0.000 01,动量设置为 0.9,使用平衡损失函数提高跟踪精度,使用随机梯度下降优化器拟合收敛速度,测练周期为 50 个批次。在两个具有代表性的目标跟踪测试集 OTB100^[16]和 VOT2018^[17]来验证本文算法的有效性。

2.1 消融实验

在 OTB100 上对本文算法进行消融实验来验证所提算法各个模块的有效性和对目标跟踪的实时性影响。实验结果如表 1 所示(最优性能加粗表示),首先,复现基准算法 SiamFC,精确度和成功率分别为 0.777 和 0.580,速度为 37FPS;然后,在 SiamFC 中添加 ECA 模块,精确度和成功率分别为 0.807 和 0.604,速度为 35FPS,表明 ECA 模块在增加少量运算的同时可以获取更多目标的语义信息,优化跟踪模型,使跟踪性能得到一定的提升;其次,在 SiamFC 中添加 SoP 模块,精确度和成功率分别为 0.828 和 0.616,速度为 32FPS,表明 SoP 模块在增加一定运算的基础上可以获取目标的二阶信息,能够更好地优化跟踪模型增强对目标的判别能力,使跟踪性能得到显著的提升;最后,在 SiamFC 中添加 ECA 模块和 SoP 模块,精确度和成功率分别为 0.852 和 0.632,速度为 31FPS,表明融合目标的原始特征、二阶池化特征和通道注意力特征可以更好地优化跟踪模型,提高目标定位的精确性,从而获得良好的跟踪性能。

表 1 ECA 模块和 SoP 模块在 OTB100 上的消融实验

SiamFC	ECA	SoP	Precision	Success	FPS
✓			0.777	0.580	37
✓	✓		0.807	0.604	35
✓		✓	0.828	0.616	32
✓	✓	✓	0.852	0.632	31

2.2 OTB100 实验结果

OTB100^[16]由 100 个视频序列组成,是目标跟踪领域中最流行的数据集之一,该数据集的两个重要评价指标是精确度(Precision)和成功率(Success)。将本文算法与 KCF^[18]、SRDCF^[19]、SiamFC^[9]、Staple^[20]、BACF^[21]、SiamRPN^[10]、SiamDW-FC^[22]、ATOM^[23]、UniTrack^[24]、SiamSE^[25]等具有代表性的目标跟踪算法进行比较,其中 SiamFC 作为所提算法的基准算法。在 OTB100 数据集上进行定性和定量分析。

2.2.1 定性分析

为了定性分析本文算法的性能,在 OTB100 数

据集中挑选 6 组视频进行比较。图 4 给出了本文算法和另外 5 种对比算法的部分跟踪结果,下面对跟踪中存在的 3 个挑战属性进行定性分析。

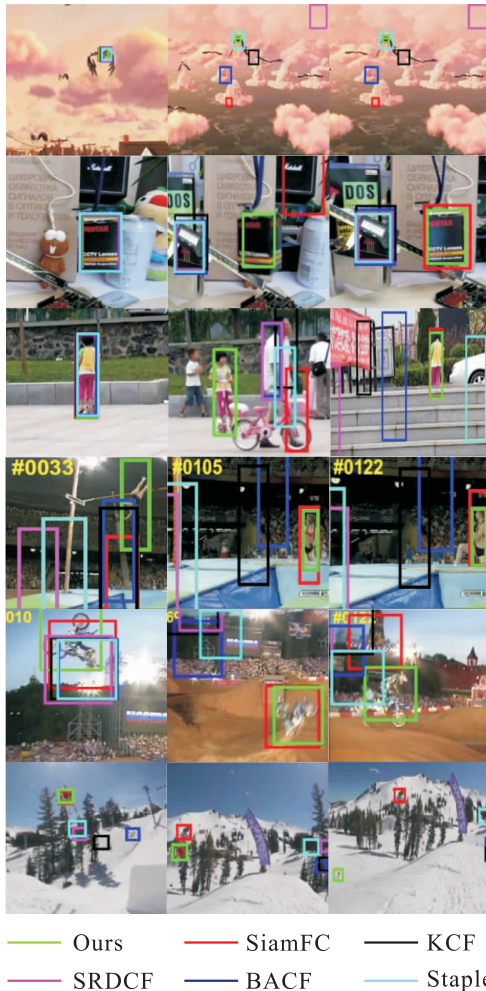


图 4 选定视频序列跟踪效果

(注:第 1 行为 Bird1,第 2 行为 Box,第 3 行为 Girl2,

第 4 行为 Jump,第 5 行为 Motor Rolling,第 6 行为 Skilling)

1)快速运动:以视频 Bird1 和 Skilling 为例,快速运动导致目标表观信息发生变化。由于 KCF 和 SRDCF 等相关滤波类算法训练模型的判别性不足,导致跟踪漂移;快速运动对互相关类跟踪算法也会产生较大的影响,SiamFC 在快速运动后丢失目标,后续过程不能恢复跟踪;而本文算法使用二阶池化网络和轻量级通道注意力可以提高模型的判别能力,在快速运动过程中能够持续稳定地进行跟踪。

2)目标遮挡:以视频 Box 和 Girl2 为例,在目标被遮挡时,AlexNet 提取特征的能力有限,导致 SiamFC 跟踪漂移;虽然 Staple 结合了 HOG 和 CN 特征能很好地解决目标尺度变化问题,但是由于没有采用 CNN 提取深度特征,在目标被遮挡时,提取的特征能力有限,导致跟踪漂移;而本文算法使用二阶池化网络和通道注意力,在目标特征提取上更能关注到目标的显著特征,提取的特征更具有鲁棒性。

3) 目标形变:以视频 MotorRolling 和 Jump 为例,目标视角和姿态的变换,导致目标发生形变和旋转,从而会引入一定的背景干扰信息。由于 Siam-FC 判别能力不足,导致跟踪漂移;然而 BACF 没有考虑目标形变问题,导致提取的特征不能完全表示目标整体特征信息;而本文算法使用二阶池化网络和通道注意力可以提高表观模型的抗干扰能力,提取的特征更具有泛化性。

2.2.2 定量分析

图 5 给出了本文算法与其它 10 种对比算法在 OTB100 数据集上的整体精确曲线和成功曲线,本文算法的精确度和成功率分别为 85.2% 和 63.2%。精确度高于现主流算法 UniTrack、SiamDW-FC 和 SiamRPN;成功率高于 UniTrack,与 SiamDW-FC 和 SiamRPN 相当。由于 ATOM 和 SiamSE 算法比较复杂,都使用了性能更优的网络和更复杂的优化策略,故本文算法的性能略低于 ATOM 和 SiamSE,但本文算法性能相较于现主流算法依然有一定的竞争力。与基准算法 SiamFC 相比,跟踪性能取得了显著提升,精确度提升了 7.5%,成功率提升了 5.2%。

为进一步分析本文算法的优缺点,提供了基于属性的性能分析来说明本文算法在关键属性上的优势。OTB100 中的视频序列标注了 11 个具有挑战性的属性,包括:光照变化(illumination variation, IV)、尺度变化(scale variation, SV)、遮挡(occlusion, OCC)、形变(deformation, DEF)、运动模糊(motion blur, MB)、快速移动(fast motion, FM)、平面内旋转(in-plane rotation, IPR)、平面外旋转(out-of-plane rotation, OPR)、离开视野(out-of-view, OV)、相似背景(background clutters, BC)、低分辨率(low resolution, LR)。

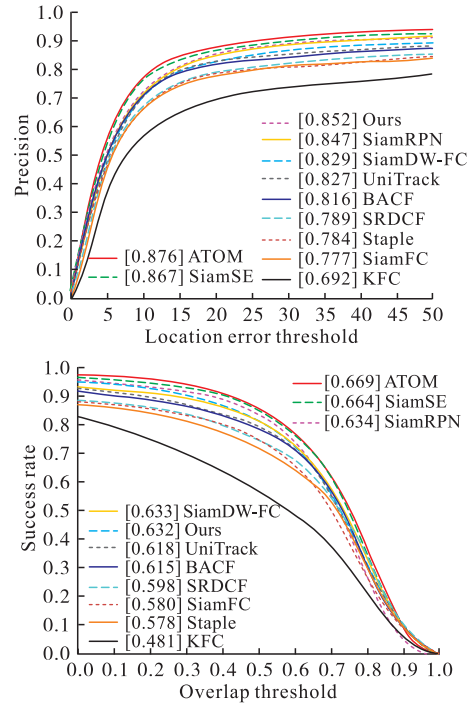


图 5 OTB100 的定量对比结果

表 2 列出了 11 种属性下的本文算法和对比算法的精确度(结果最优加粗表示,次优加下划线表示,第三优加点式下划线表示)。通过分析表 2,本文算法在多个跟踪场景中均取得了较好的跟踪效果,尤其是在相似背景(BC)、尺度变化(SV)、快速运动(FM)等挑战属性下精确度处于次优和较优位置,同时,在其它属性下精确度均处于较优位置。这是因为本文算法在相似背景、尺度变化、快速运动等属性下,通过二阶池化网络获取目标特征信息的全局依赖关系,通过通道注意力分配网络中的权重参数,更好地关注目标特征中的重要信息,改善了特征提取网络的表达能力,提高了跟踪模型的判别能力,从而取得了良好的跟踪性能。

表 2 不同属性下算法的跟踪精确度结果

算法	IV	OPR	SV	OV	DEF	LR	FM	OCC	BC	IPR	MB
Ours	0.815	0.844	<u>0.840</u>	<u>0.796</u>	0.802	<u>0.978</u>	<u>0.809</u>	<u>0.801</u>	<u>0.815</u>	<u>0.835</u>	<u>0.831</u>
SiamSE	<u>0.872</u>	<u>0.847</u>	<u>0.880</u>	<u>0.825</u>	0.823	<u>0.988</u>	<u>0.852</u>	<u>0.818</u>	<u>0.812</u>	<u>0.865</u>	0.826
UniTrack	0.812	0.817	0.810	0.683	<u>0.831</u>	0.974	0.792	0.726	0.741	0.851	0.791
ATOM	0.884	0.861	0.886	0.826	0.843	0.989	0.853	0.836	0.827	0.866	<u>0.840</u>
SiamDW-FC	0.795	0.830	0.818	0.780	0.765	0.902	0.808	0.800	0.762	0.823	0.841
SiamRPN	<u>0.860</u>	<u>0.852</u>	0.838	0.726	<u>0.826</u>	0.977	0.789	0.782	0.799	0.834	0.818
BACF	0.807	0.741	0.771	0.752	0.768	0.760	0.788	0.733	0.805	0.792	0.743
Staple	0.782	0.737	0.726	0.664	0.751	0.690	0.709	0.727	0.749	0.768	0.698
SiamFC	0.711	0.757	0.758	0.645	0.746	0.845	0.754	0.715	0.699	0.770	0.758
SRDCF	0.792	0.780	0.754	0.594	0.734	0.838	0.768	0.734	0.775	0.745	0.765
KCF	0.711	0.657	0.624	0.478	0.638	0.595	0.629	0.609	0.716	0.671	0.602

2.3 VOT2018 实验结果

VOT2018^[17]数据集包含 60 个具有挑战性的彩色视频序列,是目标跟踪领域中最流行的数据集之一,该数据集的主要评价指标是预期平均重叠率(expected average overlap, EAO)、准确率(accuracy)和鲁棒性(robustness)。

表 3 展示了本文算法与具有影响力的跟踪器 SiamFC^[9]、DCFNet^[26]、DSiam^[27]、DensSiam^[28]、SiamRPN^[10]、SiamDW-FC^[22]、SiamSE^[25] 的各项指标结果(性能最优结果加粗表示,次优结果加下划线表示,第三优结果加点式下划线表示),本文算法的

预期平均重叠率为 0.231,准确度为 0.516,鲁棒性为 0.503。在预期平均重叠率指标上略高于主流跟踪器 SiamDW-FC,在准确度指标上高于主流跟踪器 SiamDW-FC 和 SiamRPN,在鲁棒性指标上比 SiamDW-FC、SiamPRN、SiamSE 略低。相较于基准算法 SiamFC,预期平均重叠率提高了 4.3%,准确度提升了 1.6%,鲁棒性提升了 5.6%,其原因是本文算法使用二阶池化网络和通道注意力可以提高跟踪模型的判别能力,在目标特征提取上更能关注到目标的显著特征,提取的特征具有鲁棒性和泛化性。

表 3 VOT2018 实验结果

Tracker	Ours	SiamSE	SiamDW-FC	SiamRPN	DensSiam	DSiam	DCFNet	SiamFC
EAO ↑	<u>0.231</u>	0.270	0.230	0.243	0.174	0.196	0.182	0.188
Accuracy ↑	<u>0.516</u>	0.538	0.500	0.490	0.462	0.512	0.470	0.500
Robustness ↓	0.503	0.432	<u>0.490</u>	<u>0.460</u>	0.688	0.646	0.543	0.590

3 结语

本文提出一种轻量级的基于二阶池化特征融合的孪生网络目标跟踪算法。首先,使用孪生网络结构获取目标的深度特征;然后,在孪生网络结构的末端并行添加二阶池化网络和轻量级通道注意力,获取目标的二阶池化特征和通道注意力特征;最后,将目标的深度特征、二阶池化特征和通道注意力特征进行融合,使用融合后的特征进行互相关,得到的响应图能很好地区分目标和背景,提高模型的判别能力,改善目标定位的精度,从而提升跟踪性能。

在未来的研究中,将进一步挖掘二阶池化网络的潜力,使用性能更优的主干网络,获取中间层特征与图像全局的依赖关系,充分利用中间层特征潜在的信息,增强特征提取网络的表达能力,提高模型的判别能力,从而更有效地提升目标跟踪的性能。

参考文献

[1] 李玺,查宇飞,张天柱,等.深度学习的视觉跟踪算法综述[J].中国图象图形学报,2019,24(12):2057-2080.

[2] MARVASTI-ZADEH S M, CHENG L, GHANEI-YAKHDAN H, et al. Deep Learning for Visual Tracking: A Comprehensive Survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(5):3943-3968

[3] 柏罗,张宏立,王聪.基于高效注意力和上下文感知的目标跟踪算法[J/OL].北京航空航天大学学报:1-12.(2021-03-10).http://10.13700/j.bh.1001-5965.2021.0013.

[4] 刘嘉敏,谢文杰,黄鸿,等.基于空间和通道注意力机制的目标跟踪方法[J].电子与信息学报,2021,43(9):2569-2576.

[5] 刘如浩,张家想,金辰曦,等.基于可变形卷积的孪生网络目标跟踪算法[J].控制与决策,2022,37(8):2049-2055.

[6] 宋鹏,杨德东,李畅,等.整体特征通道识别的自适应孪生网络跟踪算法[J].浙江大学学报(工学版),2021,55(5):966-975.

[7] WANG N, SHI J, YEUNG D Y, et al. Understanding and Diagnosing Visual Tracking Systems[C]//Proceedings of the IEEE International Conference on Computer Vision. Chili: IEEE, 2015: 3101-3109.

[8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.

[9] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-Convolutional Siamese Networks for Object Tracking [C]// European Conference on Computer Vision. [S. l.]: Springer, 2016: 850-865.

[10] LI B, YAN J, WU W, et al. High Performance Visual Tracking with Siamese Region Proposal Network [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8971-8980.

[11] ZHU Z, WANG Q, Li B, et al. Distractor-Aware Siamese Networks for Visual Object Tracking[C]// Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 101-117.

[12] 蒲磊,冯新喜,侯志强,等.基于二阶池化网络的鲁棒视觉跟踪算法[J].电子学报,2020,48(8):1472-1478.

- [13] 姜珊,底晓强,韩成. 融合时空特性的孪生网络视觉跟踪[J]. 兵工学报, 2021, 42(9): 1940-1950.
- [14] GAO Z, XIE J, WANG Q, et al. Global Second-Order Pooling Convolutional Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE, 2019: 3024-3033.
- [15] HUANG L, ZHAO X, HUANG K. Got-10k: A Large High-Diversity Benchmark for Generic Object Tracking in the Wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(5): 1562-1577.
- [16] WU Y, LIM J, YANG M H. Object Tracking Benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [17] KRISTAN M, LEONARDIS A, MATAS J, et al. The Sixth Visual Object Tracking Vot2018 Challenge results[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany; ECCV, 2018: 3-53.
- [18] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-Speed Tracking with Kernelized Correlation Filters[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(3): 583-596.
- [19] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Learning Spatially Regularized Correlation Filters for Visual Tracking [C]// Proceedings of the IEEE International Conference on Computer Vision. [S. l.]:IEEE, 2015: 4310-4318.
- [20] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: Complementary Learners for Real-Time Tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE. 2016: 1401-1409.
- [21] KIANI GALOOGAHI H, FAGG A, LUCEY S. Learning Background-Aware Correlation Filters for Visual Tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l.]: IEEE, 2017: 1135-1143.
- [22] ZHANG Z, PENG H. Deeper and Wider Siamese Networks for Real-Time Visual Tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE, 2019: 4591-4600.
- [23] DANELLJAN M, BHAT G, KHAN F S, et al. Atom: Accurate Tracking by Overlap Maximization [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2019: 4660-4669.
- [24] WANG Z D, ZHAO H S, YA L L, et al. Do Different Tracking Tasks Require Different Appearance Models [EB/OL]. (2021-7-5). <http://doi.org/10.48550/arxiv.2107.02156>.
- [25] SOSNOVIK I, MOSKALEY A, SMEULDERS A. Scale Equivariance Improves Siamese Tracking[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S. l.]: IEEE, 2021: 2765-2774.
- [26] WANG Q, GAO J, XING J L, et al. Dcfnet: Discriminant Correlation Filters Network for Visual Tracking[EB/OL]. (2017-04-13) ArXiv Preprint ArXiv:1704.04057, 2017.
- [27] GUO Q, FENG W, ZHOU C, et al. Learning Dynamic Siamese Network for Visual Object Tracking [C]//Proceedings of the IEEE International Conference on Computer Vision. [S. l.]: IEEE, 2017: 1763-1771.
- [28] ABDELPAKEY M H, SHEHATA M S, MOHAMED M M. Denssiam: End-to-End Densely-Siamese Network with Self-Attention Model for Object Tracking[C]//International Symposium on Visual Computing. [S. l.]: Springer, 2018: 463-473.

(编辑:徐敏)