

引用格式:崔珂璠,熊淑华,陈洪刚,等. 基于非对称残差注意网络的目标跟踪算法[J]. 微电子学与计算机,2021,38(9):8-16. [CUI K F,XIONG S H,CHEN H G,et al. An object tracking algorithm based on asymmetric residual attention network [J]. Microelectronics & Computer,2021,38(9):8-16.]

基于非对称残差注意网络的目标跟踪算法

崔珂璠,熊淑华,陈洪刚,吴晓红,何小海
(四川大学 电子信息学院,四川 成都 610065)

摘 要: 针对目标跟踪算法在运动目标中存在的背景干扰和鲁棒性问题,提出一种基于 Siamese RPN++ 改进的非对称残差注意网络算法. 通过在模板分支对应的网络中添加非对称残差注意力结构,从而提取出采样图像的共同特征,形成较为稳定的目标轮廓,解决了目标运动背景发生变化的问题;采用自适应权值更新的方法融合不同区域候选网络模块输出的特征,得到更为鲁棒性的尺度变化特征表达,解决了目标形变的问题. 实验结果表明:提出的改进算法在具有挑战的跟踪测试视频上取得了良好的跟踪精度,且具有较好的鲁棒性,能够较好地应对运动背景变化、尺度变化等问题.

关键词: 目标跟踪;孪生网络;自适应权值更新;非对称残差注意力

中图分类号: TP391. 41

文献标识码: A

文章编号: 1000-7180(2021)09-0008-09

An object tracking algorithm based on asymmetric residual attention network

CUI Kefan,XIONG Shuhua,Chen Honggang,WU Xiaohong,HE Xiaohai
(College of Electronics and Information Engineering,Sichuan University,Chengdu 610065,China)

Abstract: In order to solve the background interference and robustness problems of object tracking algorithm in moving targets, an asymmetric residual attention network algorithm based on Siamese RPN++ was proposed. First, aAsymmetric residual attention structure is added to the network corresponding to the template branch, so as to extract the common features of the sampled images, form a relatively stable target contour, and solve the problem that the target moving background changes. The adaptive weight updating method is adopted to fuse the output features of candidate network modules in different regions to obtain more robust expression of scale change features and solve the problem of target deformation. The experimental results show that the proposed improved algorithm has good tracking accuracy and robustness in challenging tracking test videos, and can deal with the changes of moving background and scale.

Key words: object tracking;siamese network; adaptive weight update; asymmetric residual attention

1 引言

目标跟踪是计算机视觉领域一项非常重要且具有挑战性的工作,在很多实际应用中具有至关重要的地位,例如智能监控、自动驾驶、人机交互等. 经过

多年的研究,尽管诸多的目标跟踪算法及其衍生算法相继出现,但是在真实场景中,受到场景光照变化、遮挡、目标形变、尺度变化等各种因素的影响,给实现鲁棒的目标跟踪带来了严峻的挑战.

目标跟踪算法早期使用光流法、mean-shift 算

法、CamShift 算法等^[1],但是这类算法的精度和速度逐渐无法满足人们的需求.受到信号处理的启发,相关滤波被应用到计算机视觉领域^[2].随后,由于深度学习在视频领域中的深入应用,目标跟踪也使用深度学习进行跟踪. HCF^[3]跟踪方法使用 VGG 网络提取卷积特征,并结合了深层卷积特征语义信息和浅层卷积特征细节信息,但是深层网络限制了跟踪器的速度. Nam 等^[4]提出一个端到端的多域网络 MDNet,从多个标注的视频序列中学习物体共享的表示,在测试时将最后的多个全连接层替换为一个全连接层并在线微调,该方法虽然使用浅层网络,但得到了较好的效果. 这些方法利用离线预训练的深度网络进行特征提取,提升目标的特征表达能力从而提升精度,但是预训练的网络结构通常较大,影响了跟踪速度,多数算法无法达到实时.除了将特征提取应用到网络中,也有直接训练端到端的网络,例如基于孪生框架的算法.

SINT^[5]算法最早将基于相似性度量的孪生网络(Siamese network)结构引入目标跟踪中,它主要是 Siamese Network 学习视频中不同帧相似性的探索.基于 SINT, SiamFC^[6]提出端到端的全卷积跟踪网络并且获得了较快的跟踪速度,使得基于孪生网络的跟踪器迅速发展. SiamRPN^[7]创造性地引入了 Faster RCNN^[8]中的区域候选网络(Region Proposal Network, RPN)模块,该算法可以回归位置、形状,可以省掉多尺度测试,提高算法性能的同时算法的跟踪速度. Siamese RPN++^[9]的提出,解决了深层网络叠加破坏精度的问题,使得 Siamese 框架能够使用深层网络的同时保持良好的精度效果.

尽管基于孪生网络的目标跟踪方法近年来得到了很多的研究,但是依旧存在运动目标背景变化带来的干扰、不能有效应对目标跟踪过程尺度发生的变化、难以平衡精度和鲁棒性.针对这些问题,本文在 SiameseRPN++ 的基础上,在模板分支对应的网络中添加非对称残差注意力结构,从而提取出采样图像的共同特征,形成较为稳定的目标轮廓和定位;采用自适应权值更新的方法融合不同 RPN 模块输出的特征,得到更为鲁棒性的特征表达.最后将公开目标跟踪数据集 VOT2018 上的测试结果与其他主流算法进行对比,实验结果表明了该算法改进的有效性.

2 非对称残差孪生网络跟踪算法

尽管 SiameseRPN++ 算法已经达到良好的跟踪效果,但是网络层数的加深,也带来了时间成本和计算成本,为了减少网络中的参数量,提高计算速度,提高跟踪效率,同时增强网络对翻转图像的稳健性,提出了非对称残差注意力网络(Asymmetric Residual Attention Net, ARANet); SiameseRPN++ 中因为输入 RPN 模块中加权的方法,直接加权无法充分体现不同网络层的特征,故提出了自适应权值更新融合的方法来更好地学习特征.

本文基于 SiameseRPN++ 提出的非对称残差孪生目标跟踪算法(Asymmetric residual siamese network, SiamARA),如图 1 中(a)所示,首先将模板图像和搜索图像分别输入孪生网络的两端,通过共享 ResNet-50^[10]网络将原始图像映射到模板分支和搜索分支的特征空间中,将 Res3、Res4 和 Res5 特征层分别输入到改进的 Siamese RPN 网络中,对三个 RPN 的输出进行自适应加权融合,最后进行目标检测,输出搜索图像中被框出的目标信息.改进的 Siamese RPN 网络结构如图 1 中(b)所示,与原来的 SiameseRPN 相比,改进的网络将输入模板特征的分类分支和回归分支中的网络替换成我们提出的非对称残差注意力网络,同时采用自适应权值更新融合不同 RPN 模块的输出.在 SiameseRPN 网络中,首先分别输入由特征提取网络获得的模板分支和搜索分支的特征图 $F(z)$ 及 $F(x)$, $F(z)$ 输入至 ARANet 模块中以提取目标图像的共同特征, $F(x)$ 则连接一个基础的卷积层以提供一个非孪生的特征,然后将上述的输出分别进行相关操作,通过逐通道计算相关结果,以得到通道数非 1 的输出.多个 Siamese RPN 模块选择出多样化的样本或者具有判别性的样本块,第一个 Siamese RPN 可以去掉一些简单的样本块,而后面的网络进一步进行滤除,最终剩余一些复杂的负样本有利于提升网络的判别能力.通过自适应加权融合上述 Siamese RPN 模块,合理分配每个 RPN 特征图谱的权值.

2.1 非对称残差注意力网络

当 Res3、Res4 和 Res5 特征层分别输入到 Siamese RPN 网络中时,会分别通过一个卷积层再进行相关操作,由于需要学习的任务不同,所以卷积层的参数不共享,然后再通过一层卷积改变通道数来

获得分类和回归结果. 而使用 ResNet 作为特征提取网络,获得更多不同层次特征的同时网络深度得以加深,为了减少 RPN 层的参数量,减少网络深度

对跟踪效果的影响,从不同层次特征获得更稳定的特征,提出非对称残差注意力网络,从而有效提高跟踪的准确率.

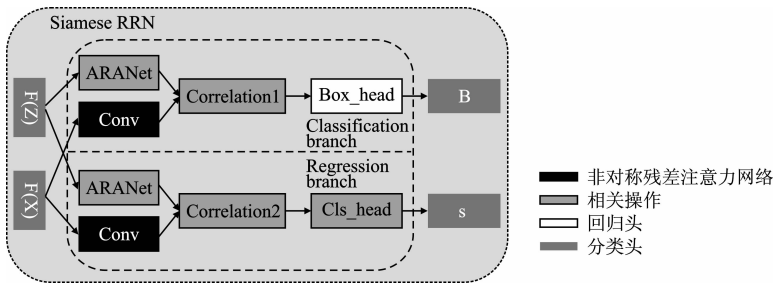
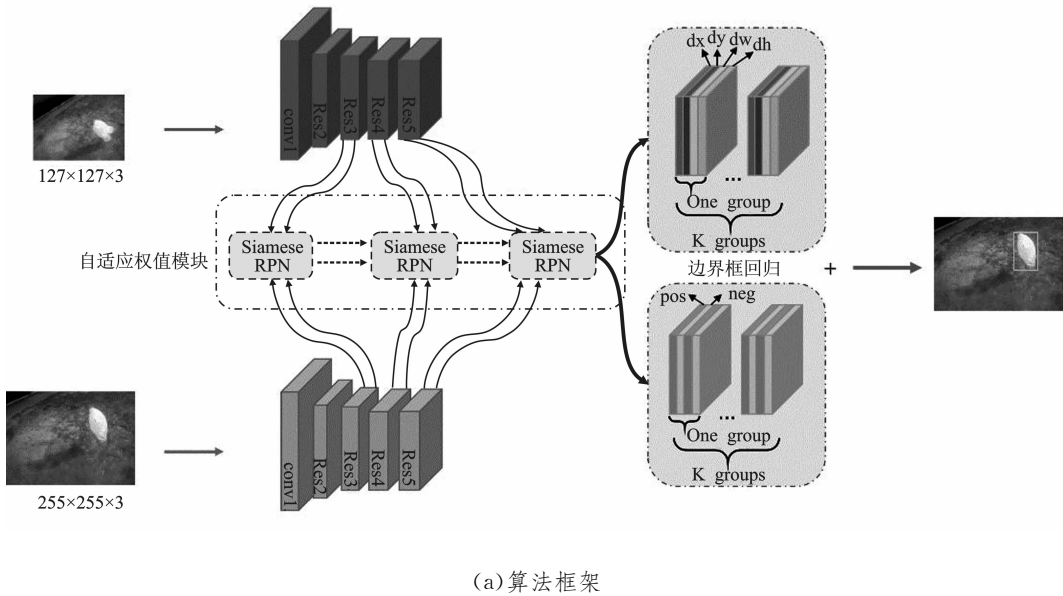
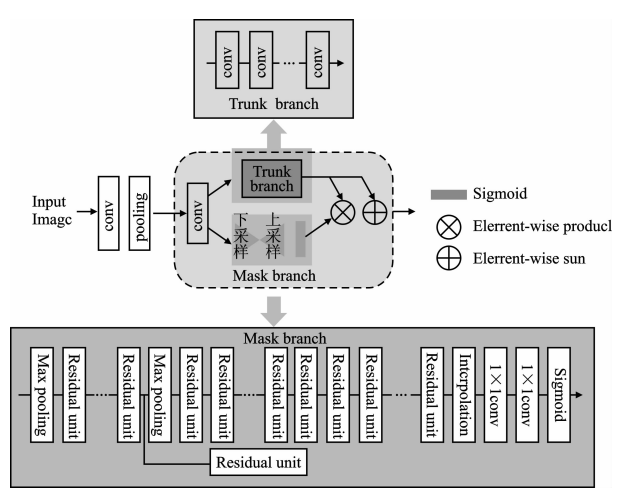


图 1 算法框架

计算机视觉中的注意力机制的基本思想是让模型学会专注,把注意力集中在重要的信息上而忽视不重要的信息. 如果只使用单个注意力模型,则可能会存在图片背景混乱、复杂前景、形态多变等问题导致分类失误,然而简单地堆积注意力模型会对特征的表达有负面的影响. 为了能学到更多不同层次的图像信息,选择了残差注意力网络. 然而注意力网络的加深会导致整体网络性能的降低,同时由于注意力的堆叠导致跟踪网络整体加深,对图像翻转和旋转的鲁棒性降低,其结果则会导致跟踪算法对旋转视频等跟踪精度降低. 为了降低模型的时间成本和计算成本,有效的减少参数量,将非对称卷积网络应用到残差注意力网络中,非对称残差注意力网络结构如图 2 所示.

注意力模块分为两部分: 掩膜分支 (mask



branch)和主干分支(trunk branch),掩膜分支会为了输出权重归一的特征图,用 Sigmoid 作为激活函

数,将输出归一化到0至1之间后与主干分支进行点乘会使得特征图的输出响应变弱,多层叠加会使得最终输出的特征图每个点上的值变得很小;掩膜分支输出的特征图可能会破坏主干分支的优点,会使得深层网络的梯度不能很好地回传.残差注意力网络能够解决这个问题,残差注意力网络包含数个堆积的注意力模块,每个模块专注于不同类型的注意力信息,增加更多的注意力模块可以提升网络的分类性能,基于不同深度的特征图可以提取额外的注意力模型.残差注意力模型可以结合到目前的大部分深层网络中,做到端到端训练结果,因为残差结构的存在,可以很容易将网络扩展到百数层.将得到的注意力特征图与主干特征图进行 element-wised add,注意力输出为:

$$H_{i,c}(x) = (1 + M_{i,c}(x)) * F_{i,c}(x) \quad (1)$$

式中, x 为输入,主干分支输出特征图为 $H_{i,c}(x)$,掩膜 $M_{i,c}(x)$ 取值在 $[0,1]$ 区间内, $F_{i,c}(x)$ 是一个深层卷积神经网络输出的特征.

为了降低残差注意力网络加深导致的计算量,在残差注意力中引入非对称卷积函数.标准的 $d \times d$ 对称卷积层可分解为包含 $d \times 1$ 和 $1 \times d$ 卷积核的连续两层,这样能够减少参数和所需的计算,并且对于翻转图像仍能正确提取特征.

通过滑动窗口形式的卷积计算,可以验证 3×1 和 1×3 是和 3×3 兼容的,如图3所示.

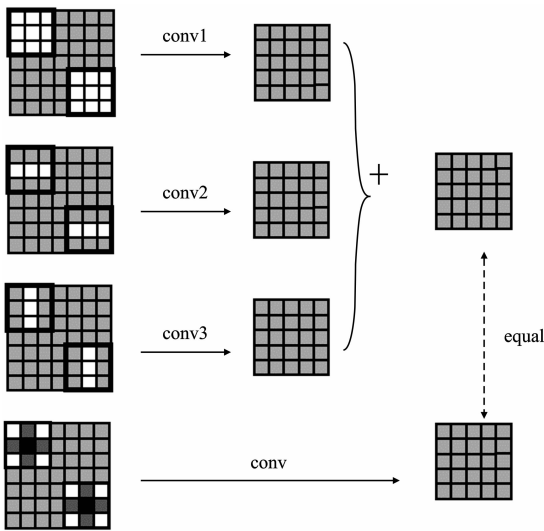


图3 不同核大小的二维卷积的可加性

非对称卷积使用一些与架构无关的结构来增强现有的模型,能够强化特征提取,实现效果提升的同时不增加计算量,提升模型对图像翻转的鲁棒性.

对于非对称残差注意力网络,给定输入特征图谱 x ,对其进行两次下采样,两次上采样,得到特征图谱 $M(x)$,经过函数得到输出的掩模:

$$S(x) = \text{sigmoid}(M(x)) = \frac{1}{1 + e^{-M(x)}} \quad (2)$$

输出的掩模能够自适应地学习有效特征同时抑制住噪声因素.输出得到的非对称残差注意力模块 $P(x)$ 公式如式(3)所示:

$$P_{i,c}(x) = (1 + S_{i,c}(x)) * x_{i,c} \quad (3)$$

式中, i 代表神经元的空间位置, c 代表神经元的通道位置.同残差网络中的恒等映射类似, $S(x)$ 的输出范围为 $[0,1]$,当 $S(x)$ 接近0时, $P(x)$ 的输出接近原始的特征 x .

在非对称残差注意力模块中,注意力掩模在向前传播中可看作特征选择器,在反向传播中进行梯度更新操作.因此通过残差注意力机制可以使得整个网络根据目标的变化而自适应地进行调整.

2.2 自适应权值更新

由于SiameseRPN++模型中三个RPN模块的输出尺寸具有相同的空间分辨率,因此可以直接对RPN输出进行加权求和,加权融合层结合了所有的输出,SiameseRPN++使用的是线性加权融合的方式,直接加权融合得到的结果并不能充分地表达特征,故提出了自适应权值更新融合的方法来更好的学习特征,以突显出视频帧中更具鲁棒性的特征.

网络中使用ResNet中的第三、四、五层的输出特征来衡量模板分支和搜索分支的相似度,最后进行跟踪操作.

对于不同RPN模块输出的特征,获得的损失值越小则表明该特征下学习的滤波器性能越好,则应对该特征图提供更高的权重.

任一RPN输出的特征 f 对应的归一化的权重为:

$$w_{\text{reg}}^f = \frac{\text{Loss}_{\text{reg}}^{(F-f)}}{\text{Loss}_{\text{reg}}^{\text{rpn1}} + \text{Loss}_{\text{reg}}^{\text{rpn2}} + \text{Loss}_{\text{reg}}^{\text{rpn3}}} \quad (4)$$

$F = \{\text{rpn1}, \text{rpn2}, \text{rpn3}\}$ 表示特征的集合,其中: $\{F-f\}$ 表示 F 中不同于 f 的另一特征.采用各自的RPN网络得到的特征图分别表示为: $R_{\text{reg}}^{\text{rpn1}}, R_{\text{reg}}^{\text{rpn2}}, R_{\text{reg}}^{\text{rpn3}}$,则特征图层面的特征融合采用如下的自适应加权方式实现:

$$R_{\text{reg}} = w_{\text{reg}}^{\text{rpn1}} \times R_{\text{reg}}^{\text{rpn1}} + w_{\text{reg}}^{\text{rpn2}} \times R_{\text{reg}}^{\text{rpn2}} + w_{\text{reg}}^{\text{rpn3}} \times R_{\text{reg}}^{\text{rpn3}} \quad (5)$$

网络损失函数为:

$$\text{loss} = L_{\text{cls}} + \lambda L_{\text{reg}} \quad (6)$$

分类损失 L_{cls} 采用交叉熵损失, 回归损失 L_{reg} 采用 smooth_{L_1} 损失, 计算如下:

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (7)$$

式中, $L_1(x)$ 为

$$L_1(x) = |x| \quad (8)$$

3 实验与分析

实验环境如下: Intel(R)Core(TM)i7 CPU 处理器, Ubuntu18.04 64 位操作系统, 显卡为 NVIDIA GTX 2080Ti, 深度学习框架选用 pytorch. 目标跟踪从客观评价和主观视觉效果两个方面对算法进行综合评价, 其中, 客观评价采用准确率 (Accuracy)^[11]、鲁棒性 (Robustness)^[11] 和平均重叠期望 (Expect Average Overlaprate, EAO)^[12] 三个客观评价指标.

3.1 数据集

目标跟踪算法常用的数据集包括两种, 分别是 OTB(Online Object Tracking Benchmark) 和 VOT (Visual Object Tracking), VOT 数据集存在更多的挑战且一直在更新, 包含 60 个公开彩色序列, 涵盖了不同场景不同条件下的跟踪视频, 故使用 VOT2018^[11] 用于算法的验证实验.

3.2 评价指标

3.2.1 Accuracy

Accuracy 用来评价 tracker 跟踪目标的准确度, 是指跟踪器在单个测试序列的平均重叠率. 它借用了 IoU (Intersection-over-Union, 交并比) 定义, 具体为预测区域和真值区域的交集与并集之比, 数值越大, 准确度越高.

某序列第 t 帧的 accuracy 定义为:

$$\varphi_t = \frac{A_t^G \cap A_t^T}{A_t^G \cup A_t^T} \quad (9)$$

式中, A_t^G 代表第 t 帧 ground truth 对应的 bounding box, A_t^T 代表第 t 帧 tracker 预测的 bounding box. 由于 tracker 会在一个序列上重复运行多次, 故定义 $\Phi_i(i, k)$ 为第 i 个 tracker 在第 k 次重复序列中在第 t 帧上的准确率, 设重复次数为 N_{rep} 定义为:

$$\Phi_i(i) = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} \Phi_i(i, k) \quad (10)$$

第 i 个 tracker 的 average accuracy 定义为:

$$\rho_A(i) = \frac{1}{N_{\text{valid}}} \sum_{t=1}^{N_{\text{valid}}} \Phi_i(i) \quad (11)$$

3.2.2 Robustness

Robustness^[11] 用来评估 tracker 跟踪目标的稳定性, 是指单个测试序列中跟踪器失败的次数, 数值越大表示跟踪器稳定性越差. 定义 $F(i, k)$ 为第 i 个 tracker 在第 k 次重复中失败的次数. 所以第 i 个 tracker 的 average robustness 定义为:

$$\rho_R(i) = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} F(i, k) \quad (12)$$

3.2.3 EAO

EAO 定义为在一个短时图像序列中的每个 tracker 的非重置重叠期望值, 为 VOT 评估跟踪精度的最重要指标. 首先将所有序列按照长度进行分类, 令 tracker 在长度为 N_s 的视频序列上测试, 然后获得每一帧的准确率 $\Phi(t)$, 之后进行平均, 得到该序列上的准确率:

$$\Phi_{N_s} = \frac{1}{N_s} \sum_{t=1}^{N_s} \Phi(t) \quad (13)$$

最后 tracker 在所有长度为 N_s 的序列上全部运行一遍, 再求平均, 得到了该 tracker 在长度 N_s 序列上的 EAO 值:

$$\bar{\Phi}_{N_s} = \frac{1}{N_s} \sum_{i=1}^{N_s} \Phi_{N_s} \quad (14)$$

3.3 定量分析

为了验证文中所提模型有效性, 选用 VOT2018 数据集进行测试. 所有网络结构使用的训练参数相同, 采用 SGD 优化算法, 每次读取的样本数为 28, 总共迭代 20 个 epoch, 前 5 个 epoch 使用 0.001 的 warmup 学习率来训练 RPN 分支, 后 15 个 epoch 中, 整个网络都是端到端的训练, 学习率从 0.005 指数衰减到 0.000 5. 权重衰减设为 0.000 5, 动量设为 0.9.

3.3.1 消融实验结果

使用 VOT2018 作为验证集和复现的 Siamese RPN++ 作为基准算法来分析本文提出的每个分量的有效性. 表 1 列出了所提模型在 VOT2018 数据集的消融实验. ARANet 模型为所提的非对称残差注意力结构改进的模型, selfWeight 模型为自适应权值更新方法改进的模型.

表 1 所提模型在 VOT2018 数据集的消融实验

模型	EAO	Accuracy	Robustness	fps
Base	0.414	0.600	0.234	71.0
Base+ARANet	0.410	0.602	0.253	66.9
Base+ARANet+ selfWeight	0.414	0.611	0.215	63.8

从表中可以看出,与基准模型相比,所提 ARANet 模型 Accuracy 指标提高了 0.002,跟踪准确率有所提升,但是跟踪器的失败次数上升,EAO 相对减小,跟踪速度下降;在上述模型的基础上,添加自适应权值更新模块,Accuracy 指标整体提升 0.009,跟踪失败次数下降了,即鲁棒性也得到提升,EAO 相对增加.由于模型额外增加了网络分支,计算量有所增加,故速度相比基准模型有所下降,为 63.8 fps,但仍能达到实时.相关研究表明,在 VOT2018 挑战赛中,平均重叠期望、准确率和鲁棒性这三个指标难以同时提升,因此综合此数据集中 Accuracy 和 Robustness 两个指标,所提模型能够有效提升跟踪性能.

3.3.2 其他主流算法对比

从基于相关滤波器和孪生网络类的方法中选取一些代表性的跟踪方法作为对比.表中的算法表示使用 ARANet+selfWeight 模型进行测试对比.表 2 列出了上述算法在 VOT2018 数据集的跟踪结果,同表中其他算法进行比较,本文算法 EAO 指标和 Accuracy 指标最高,分别达到了 0.414 和 0.611,可以看出本文算法跟踪准确率较高,体现了跟踪框的高重叠率.但是由于基于深度学习算法缺乏对模板的自适应性,鲁棒性仍然与依赖于在线更新的相关滤波方法存在差距,但是精度和 EAO 明显优于相关滤波方法,达到了较好的跟踪效果.图 4 为表 2 获得的不同算法的期望重叠曲线和 A-R 图,图 4(a)为期望重叠曲线,Rank 越小 EAO 越大表示算法跟踪越好;图 4(b)为根据 Accuracy 和 Robustness 获得的 A-R 图,坐标越大跟踪性能越好.

为了更直观全面地评价所提算法的性能,在 VOT2018 数据集中,分别对比不同算法在具有不同属性的视频序列上的效果,各算法分别在具有不同视频属性的序列上进行跟踪,获得各算法在不同属性视频上的精度结果如表 3 所示.

表 2 不同算法在 VOT2018 测试集的精度

算法	EAO	Accuracy	Robustness
SiamRPN ^[13]	0.384	0.588	0.276
MFT ^[13]	0.386	0.505	0.140
DaSiamRPN ^[13]	0.326	0.569	0.337
LADCF ^[13]	0.389	0.503	0.159
SiamAN ^[13]	0.22	0.54	0.181
STRCF ^[13]	0.345	0.523	0.215
CCOT ^[13]	0.267	0.49	0.32
CFNet ^[13]	0.1875	0.498	0.602
DCFNet ^[13]	0.182	0.464	0.352
UPDT ^[13]	0.379	0.536	0.184
DRT ^[13]	0.355	0.518	0.201
SiamARA	0.414	0.611	0.215

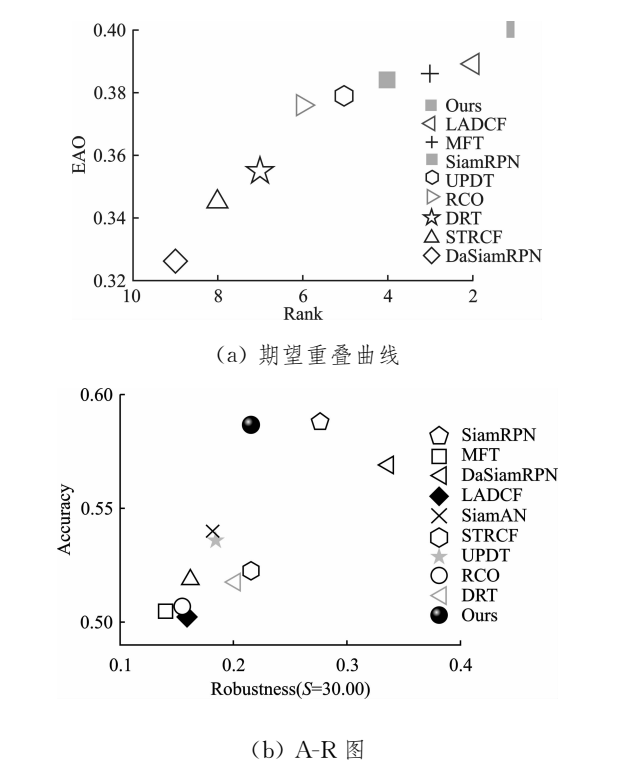


图 4 VOT2018 上的期望重叠曲线和 A-R 图

表 3 不同算法对不同属性视频序列的精度

算法	IV	SV	OCC	FM	OV	LR
SiamRPN	0.454	0.454	0.415	0.51	0.525	0.453
DaSiamRPN	0.551	0.552	0.453	0.554	0.553	0.554
CFNet	0.548	0.46	0.42	0.51	0.515	0.510
CCOT	0.446	0.450	0.391	0.483	0.515	0.499
DCFNet	0.530	0.448	0.379	0.465	0.487	0.469
ECO	0.501	0.45	0.358	0.468	0.52	0.47
SiamARA	0.612	0.58	0.52	0.601	0.62	0.601

表 3 中,各视频属性分别为:光照变化(Illumination Variation, IV)、尺度变化(Scale Variation, SV)、遮挡(Occlusion, OCC)、快速运动(Fast Motion, FM)、视野之外(Out-of-View, OV)、低分辨率(Low Resolution, LR)。

由表 3 可以看出,对于尺度变化下的视频能达到较高的精度,在快速运动下相较于其他算法,也能够达到更好的效果,快速运动下,背景会发生改变,

由此对比可以看出所提算法在尺度变化及背景变化上有较好的提升。

3.4 定性分析

选择具有不同属性的视频序列 Skating (FM、SV)、Matrix (FM、SV)、motorRolling (FM、SV)、girl (SV)、CarScale (SV) 进行分析,比较的算法分别为:CFNet、CCOT、DCFNet、ECO 以及本文提出的算法 SiamARA。结果如图 5 所示。

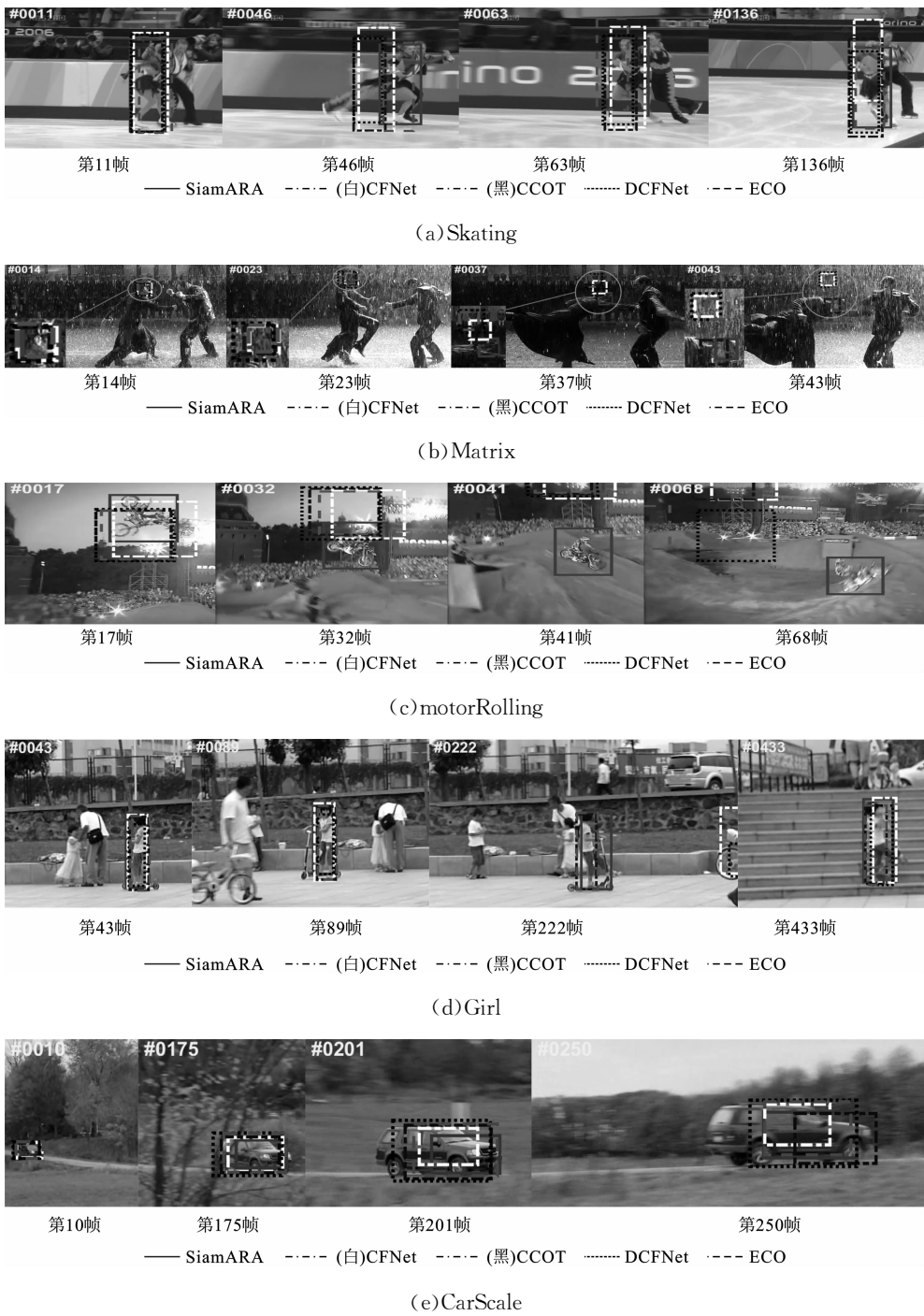


图 5 5 种算法对各视频序列跟踪结果对比

选取每个视频中跟踪效果明显的四帧图像作为结果显示,五种算法的跟踪结果用不同颜色在每帧框出。

由图 5(a)中可以看出从第 46 帧开始背景以及人物的体型发生极大的变化,只有本算法和 ECO 算法能准确跟踪到目标,CFNet、CCOT、DCFNet 算法发生目标漂移现象。

由图 5(b)中可以看出,由于人物的快速移动和光线原因,从第 23 帧开始,部分算法跟踪开始发生漂移,本算法依旧能跟踪到目标,在第 37 帧中,由于目标的快速运动及遮挡,本算法和 CCOT 算法能够跟踪到目标,但是 CCOT 算法的跟踪框中包含了部分背景信息,其他算法均跟踪失败。

在图 5(c)中,由于背景的快速变化,从第 32 帧开始,大部分算法出现跟踪漂移,后续帧中其他算法跟踪出现明显误差,本算法能够成功跟踪到目标。

在图 5(d)中,从 89 帧开始出现行人,经过并遮挡住目标后,有本算法、CCOT、ECO 三个算法能够成功跟踪到目标,其中本算法能够框出整个目标。

在图 5(e)中,在目标没有明显尺寸变化时五种算法均能正确跟踪到目标,但是从第 175 帧开始,目标尺寸与最初的尺寸明显不同时,本算法和 CCOT 算法能够框出目标,CFNet 和 ECO 无法框出变大的目标,DCFNet 发生漂移现象,从第 201 帧到第 250 帧,本算法能够框出变大的目标,但是 CFNet 算法和 ECO 算法只能框出目标的一部分,CCOT 算法则出现了漂移现象。

4 结束语

本文研究了一种基于非对称残差注意力机制的孪生网络结构的目标跟踪方法。在模板分支对应的网络中添加非对称残差注意力结构,提取出采样图像的共同特征,形成较为稳定的形状轮廓;采用自适应权值更新的方法融合不同 RPN 模块输出的特征,从而学习更好的特征。最后在 VOT2018 数据集的实验结果表明,改进模型在准确率和鲁棒性方面都具有较好的性能,实现了较好的目标跟踪效果。

参考文献:

[1] 文武,姜涛. 融合 SIFT 和尺度方向自适应的 Mean shift 目标跟踪算法[J]. 微电子学与计算机, 2015, 32(10): 93-97. DOI: 10.19304/j.cnki.issn1000-7180.

2015. 10. 020.

WEN W, JIANG T. Object tracking algorithm fusing SIFT and scale-orientation adaptive Mean shift[J]. Microelectronics&Computer, 2015, 32(10): 93-97. DOI: 10.19304/j.cnki.issn1000-7180. 2015. 10. 020.

[2] 孙志旻,张秀彬,张峰,等. 复杂背景中移动目标的快速识别与跟踪[J]. 微电子学与计算机, 2003, 20(11): 48-51. DOI: 10.3969/j.issn.1000-7180. 2003. 11. 014.

SUN Z M, ZHANG X B, ZHANG F, et al. The fast recognition and trace of moving targets in a complicated background [J]. Microelectronics&Computer, 2003, 20(11): 48-51. DOI: 10.3969/j.issn.1000-7180. 2003. 11. 014.

[3] MA C, HUANG JB, YANG X K, et al. Hierarchical convolutional features for visual tracking [C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 3074-3082. DOI: 10.1109/ICCV. 2015. 352.

[4] NAM H, HAN B. Learning multi-domain convolutional neural networks for visual tracking [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 4293-4302. DOI: 10.1109/CVPR. 2016. 465.

[5] TAO R, GAVVES E, SMEULDERS A W M. Siamese instance search for tracking [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1420-1429. DOI: 10.1109/CVPR. 2016. 158.

[6] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking [C]//European Conference on Computer Vision. Amsterdam: Springer, 2016: 850-865. DOI: 10.1007/978-3-319-48881-3_56.

[7] LI B, YAN J J, WU W, et al. High performance visual tracking with siamese region proposal network [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8971-8980. DOI: 10.1109/CVPR. 2018. 00935.

[8] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149. DOI: 10.1109/TPAMI. 2016. 2577031.

[9] LI B, WU W, WANG Q, et al. Siamrpn++: evolu-

tion of siamese visual tracking with very deep networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 4282-4291. DOI: 10.1109/CVPR.2019.00441.

[10] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778. DOI: 10.1109/CVPR.2016.90.

[11] 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: Springer, 2018: 3-53. DOI: 10.1007/978-3-030-11009-3_1.

[12] KRISTAN M, MATAS J, LEONARDIS A, et al. The visual object tracking VOT2015 challenge results [C]//Proceedings of the IEEE International Conference on Computer Vision Workshop. Santiago: IEEE, 2015: 564-586. DOI: 10.1109/ICCVW.2015.79.

[13] 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, 2019. DOI: 10.1109/TITS.2020.3046478.

作者简介:

崔珂璿 女,(1997-),硕士研究生. 研究方向为计算机视觉、图像处理.

熊淑华 女,(1969-),博士,副教授. 研究方向为多媒体通信.

陈洪刚(通讯作者) 男,(1991-),博士,助理研究员. 研究方向为图像处理. E-mail:honggang_chen@scu.edu.cn.

吴晓红 女,(1970-),博士,副教授. 研究方向为图像处理与模式识别、计算机视觉、无线通信.

何小海 男,(1964-),博士,教授. 研究方向为图像处理、模式识别、图像通信.