计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(7):284−292 [doi: 10.15888/j.cnki.csa.009139]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

融合坐标信息与模板更新的孪生网络目标跟踪①

侯艳丽, 魏义仑, 王鑫涛

(河北科技大学 信息科学与工程学院, 石家庄 050018)
通信作者: 魏义仑, E-mail: w18733103212@163.com

摘　要: 应对孪生网络单目标跟踪算法在跟踪中遇到背景杂乱、相似物影响、遮挡等复杂场景的问题导致跟踪系统精度和成功率下降的问题, 提出一种融合坐标注意力机制和模板更新的跟踪算法 MCUSiamRPN (MobileNet coordinate attention and updating of template SiamRPN). 在 SiamRPN 算法基础上, 采用改进的 MobileNetV3 为特征提取网络, 多层特征信息分别送入坐标注意力模块, 进行特征融合, 丰富语义信息; 设计了一种自适应模板更新模块, 结合初始模板和当前帧的模板用于估计下一帧的最佳模板更新模板信息. 在 OTB100 和 UAV123 两个数据集上进行测试, 结果显示: 相比于基准算法 SiamRPN, 精度分别提升了 5.3% 和 3.7%; 成功率分别提升了 3.7% 和 5.2%, 验证了该算法的有效性.

关键词: 单目标跟踪; 孪生网络; MobileNetV3; 坐标注意力; 模板更新

引用格式: 侯艳丽,魏义仑,王鑫涛.融合坐标信息与模板更新的孪生网络目标跟踪.计算机系统应用,2023,32(7):284–292. http://www.c-s-a.org.cn/1003-3254/9139.html

Siamese Network Target Tracking Fused with Coordinate Information and Template Update

HOU Yan-Li, WEI Yi-Lun, WANG Xin-Tao

(School of Information Science and Engineering, Hebei University Science and Technology, Shijiazhuang 050018, China)

Abstract: The single target tracking algorithm for siamese networks would encounter complex scenes such as background clutter, the influence of similar objects, and occlusion, which leads to a decrease in the accuracy and success rate of the tracking system. In response, this study proposes a tracking algorithm combining the coordinate attention mechanism and template update, i.e., MobileNet coordinate attention and updating of template SiamRPN (MCUSiamRPN). On the basis of the SiamRPN algorithm, the improved MobileNetV3 is used as the feature extraction network, and the multi-layer feature information is sent to the coordinate attention module to fuse features and enrich semantic information. An adaptive template updating module is designed, which combines the initial template and the template of the current frame to estimate the best template of the next frame for template information updating. The test results on OTB100 and UAV123 data sets show that compared with the benchmark algorithm SiamRPN, the proposed one has precision improved by 5.3% and 3.7% and achieves a success rate increased by 3.7% and 5.2%, respectively, which verifies the effectiveness of the developed algorithm.

Key words: single target tracking; Siamese network; MobileNetV3; coordinate attention; template update

目标跟踪[1] 作为计算机视觉的一个重要研究领域, 其具有一定挑战性, 一直以来都是计算机视觉领域中广大研究人员热切关注的对象. 目标跟踪技术在民用和军事方面都有着许多的应用前景. 在民用方面, 目前已经广泛应用于车辆跟踪导航、医学诊断、安防监控、虚拟现实等多个行业. 除了以上所述的民用之外,

① 基金项目: 河北省重点研发计划 (21355901D)
收稿时间: 2022-12-11; 修改时间: 2023-01-06; 采用时间: 2023-01-19; csa 在线出版时间: 2023-04-07
CNKI 网络首发时间: 2023-04-10

目标跟踪还在军事方面包括精准打击、无人机侦察、导航规划等领域具有同样的研究和应用价值. 尽管目标跟踪有着众多应用领域, 但随着不断变化的复杂场景和目标本身的变化作为影响跟踪效果的常见因素, 造成算法的性能下降, 设计一种在复杂场景下具备高精度和成功率的目标跟踪算法是目标跟踪领域的技术难题.

目标跟踪的方法按照模式主要分为生成式跟踪算法和判别式跟踪算法. 前者首先构建目标模型, 在后续帧中寻找相似特征, 逐步迭代完成目标匹配, 实现跟踪. 经典的生成式跟踪算法包括卡尔曼滤波算法 (KF)[2], 其应用在线性高斯问题的场景; 粒子滤波算法 (PF)[3] 可有效解决非线性非高斯问题, 但不适用于跟踪灵活性强的目标; 均值漂移算法[4] 通过将目标和候选目标的概率密度函数间的距离最小化来跟踪目标. 生成式跟踪算法没有利用背景信息, 仅通过简单目标模型来定义待跟踪目标, 具有很大局限性, 跟踪效果差. 判别式跟踪算法需要同时分析目标和背景信息, 通过比较二者不同, 以获得精确的目标模型, 进而确定当前帧中的目标位置实现跟踪, 背景信息的引入, 使得判别式跟踪算法优势增加. 在判别式跟踪算法中比较有代表性的为相关滤波算法, 具有较高的速度和准确度. 相关滤波的目标跟踪算法[5] 比较经典的包括核相关滤波器 (KCF)[6], 通过循环矩阵生成多个样本, 利用不同样本进行回归训练, 并采用点乘计算, 降低运算量, 使得算法满足实时性, 但难以应对物体的尺度变化; 空间正则化的判别式相关滤波器 (SRDCF)[7], 基于 DCF 算法, 填充操作取值为 4, 可检测区域增大, 同时在空间增加惩罚因子防止模型过拟合; 判别尺度空间跟踪算法 (DSST), 提出位置滤波器与尺度滤波器, 可以精准的估计目标尺度; 快速判别尺寸空间跟踪算法 (fDSST)[8] 在 DSST 的基础上进行特征压缩、降维和插值, 运算速度有很大提升; 以及 Staple 算法[9], 针对 HOG 特征应对形变严重情况抵抗力差和 COLOR 特征难以区分目标与背景的缺点, 结合两种特征因子优点学习目标.

随着深度学习方法的发展, 在判别式跟踪算法中出现了基于深度学习建立的跟踪框架, 它通过训练端到端的卷积神经网络模型[10], 实现目标跟踪. 与相关滤波的跟踪算法相比, 通过神经网络学习的特征辨别能力更强且跟踪效果更稳定, 因此逐渐在目标跟踪领域脱颖而出. 基于孪生网络的目标跟踪算法由于具有领先的精度和准确率, 在各类跟踪算法中脱颖而出. 全卷积孪生网络 (SiamFC)[11] 将孪生网络应用到目标跟踪中, 将 AlexNet 作为主干网络进行特征提取, 模板分支和搜索分支输出特征图, 进行互卷积操作, 得到目标响应结果再反向映射到原图, 得到当前帧位置. 孪生候选区域生成网络 (SiamRPN)[12] 除孪生网络部分, 新增了候选区域部分, 其包含分类分支与回归分支, 并将 anchor 引入到跟踪领域, 计算位置差得到回归标签, 极大地提高了定位精度. 分散注意力的孪生候选区域生成网络 (DaSiamRPN)[13] 对训练数据进行处理, 产生正负样本, 大大提高网络判别能力, 并提出了干扰物感知模型和 local-to-global 的策略来更新模板、进行长时间跟踪.

以上方法在当时都取得了不错效果, 但相关滤波建立的模型无有效预测机制, 训练样本不会保存, 以往样本逐渐失效, 且模型不会区别对待异常样本, 若出现目标形状变化、背景干扰、遮挡等情况, 会导致跟踪失效. 此外, 基于深度学习的目标跟踪方法不断改进, 但包括孪生网络在内的深度学习跟踪算法仍有一定不足, 具体包括: 1) 空间位置信息及背景信息易忽略, 在相似物干扰及背景杂乱情况识别目标能力差; 2) 多数网络采用浅层网络提取特征, 不能充分提取丰富特征信息; 3) 模板图像采用视频第 1 帧作为模板或者加权平均更新模板, 导致目标被遮挡或者发生较大形变等情况后跟踪框漂移.

针对上述问题, 本文优化了应用浅层网络 AlexNet[14] 提取目标特征的 SiamRPN, 提出一种融合坐标信息及模板更新的孪生网络跟踪算法 MCUSiamRPN. 特征提取网络采用改进使之更适用于跟踪的 MobileNetV3 网络 (MobileNetV3S)[15], 提取更丰富的特征信息; 融入坐标注意力模块[16], 关注目标位置空间信息; 引入模板自适应更新模块[17], 根据初始模板和当前帧, 自适应性地更新下一帧模板. 算法分别在 OTB100 与 UAV123 数据集上进行跟踪实验, 本文所提算法的跟踪精度与成功率数据均最高, 并做了消融实验对比, 各个模块的引入均对算法的跟踪性能有效提升.

1 MCUSiamRPN 算法描述

本文算法 MCUSiamRPN 的总体框架如图 1 所示, 主要分为 3 个模块: 特征提取网络模块、RPN 回归分类模块、模板更新模块. 在 SiamRPN 算法的基础上,

选取 MobileNetV3S 进行特征提取, 模板和搜索两分支的权值共享、结构相同, 截取跟踪视频的第 1 帧作为模板帧 Z、后续帧为检测帧 X, 127×127×3 的 Z 与 255×225×3 的 X 作为输入分别通过特征提取网络, 在第 3、7 和 11 个 block 融入坐标注意力模块, 将位置信息引入到通道, 分别提取特征并融合特征信息, 将融合后的特征信息与经过完整特征提取网络的特征信息结合生成特征图, 输入到 RPN 部分, 模板分支输出的特征图作为卷积核与搜索分支输出的特征图进行互相关操作, 分类分支最终得到 Z 中各区域的目标和背景信息, 回归分支得到预测框与真实框的相对位置差, 使得回归后的跟踪框确定目标真实位置. 另外, 引入自适应模板模块, 将视频首帧信息与模板更新模块的输出信息融合并应用到当前帧的跟踪, 以此来进行跟踪过程的模板更新. 本文框架的设计, 有效地提取了丰富的特征信息及位置信息, 并解决了模板单一的问题.

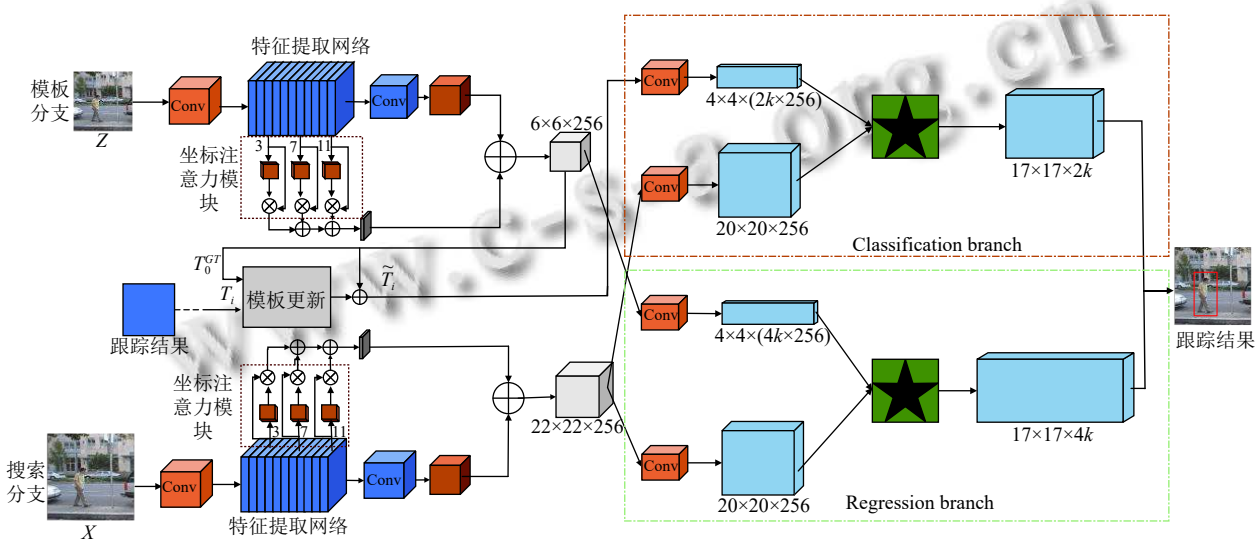


图 1 MCUSiamRPN 算法框架图

1.1 特征提取网络 MobileNetV3S

本文基准算法 SiamRPN 采用 5 层卷积的 AlexNet 进行特征提取, 不能提取丰富语义信息, 跟踪模型鲁棒性较差. 目标跟踪特征提取网络排名靠前的深层网络 ResNet50[18], 精度提升显著但网络结构复杂、参数过多导致计算量过大并不适用于配置低的设备. 为了提取浅层与深层特征的同时减少参数, 本文选取 MobileNetV3 作为 MCUSiamRPN 的特征提取网络. MobileNetV3 采用深度可分离卷积[19], 与标准卷积相比, 卷积层数加深的同时参数减少, 计算量降低. 另外 MobileNetV3 中采用逆残差结构[20], 与标准残差结构不同, 其调换了降维和升维的顺序, 先使用 1×1 卷积实现升维, 再通过 3×3 的逐通道卷积提取特征, 最后使用 1×1 卷积实现降维, 就能够提取更多的信息.

为了使 MobileNetV3 在复杂场景下有效的处理单目标跟踪问题, 本文对网络结构进行了重新改进, 得到 MobileNetV3S 网络结构检测分支特征提取网络结构见表 1.

表 1 MobileNetV3S 网络结构

输入	卷积层	输出维度	SE	最大值池化	非线性激活函数	步长
255×255×3	conv2d, 3×3	16	—	√	H-Swish	1
112×112×16	block, 3×3	16	√	—	ReLU	1
110×110×16	block, 3×3	24	—	—	ReLU	1
108×108×24	block, 3×3	40	—	√	ReLU	1
54×54×40	block, 3×3	40	√	—	ReLU	1
52×52×40	block, 3×3	48	√	—	H-Swish	1
50×50×48	block, 3×3	48	√	—	H-Swish	1
48×48×48	block, 3×3	96	√	√	H-Swish	1
24×24×96	block, 3×3	96	√	—	H-Swish	1
22×22×96	block, 3×3	96	√	—	H-Swish	1
22×22×96	conv2d, 1×1	256	—	—	H-Swish	1
255×255×3	conv2d, 3×3	16	—	√	H-Swish	1

MobileNetV3S 具体改进如下.

(1) 应对孪生网络输入模板图像尺寸为 127×127×3 尺寸较小的问题, 将 MobileNetV3 中卷积层和 block 步长缩短为 1, 用于保存深层特征图中丰富的有效信息, 更有利于跟踪.

(2) 优化现有的主干网络结构, 删减网络末端平均

池化层和 1×1 卷积层, 使用常规卷积层与坐标注意力模块结合提取目标的特征信息, 丰富预测特征, 弥补删除末端卷积层的不足.

(3) 由于 MobileNetV3 需进行 padding 操作维持特征尺寸, 当目标移动到搜索范围边界附近时, 原始特征周围的填充会诱发潜在的位置偏差, 预测精度会降低, 为了减少 padding 操作对跟踪效果的影响, 将原网络中 5×5 的卷积核缩小为 3×3 的卷积核, 产生最外层无效信息通过 Crop 操作裁剪.

1.2 坐标注意力模块

注意力模块可从全局信息中重点关注对当前跟踪目标更关键的细节信息. 目前常用的卷积注意力模块[21]只引入局部位置信息, 不能捕获特征图相距较远两个像素之间的相关性, 即远程依赖. 坐标注意力模块可将位置信息引入到通道, 其将通道注意力分解为两个 1 维特征编码过程, 分别沿 X 方向和 Y 方向聚合特征, 分别捕获远程依赖关系和保留精确的位置信息, 生成一对方向和位置敏感的特征图. 具体操作如图 2 所示.

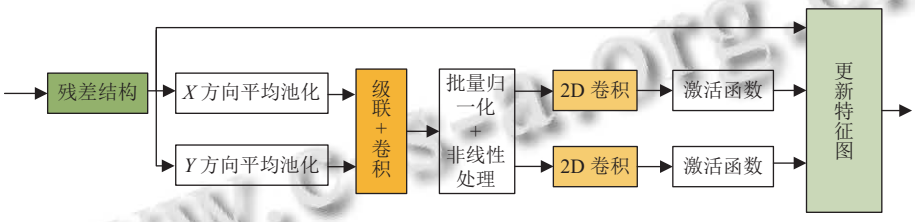


图 2 坐标注意力模块

(1) 坐标信息嵌入

输入的特征信息经过残差结构, 使用尺寸 $(H, 1)$ 和 $(1, W)$ 的池化核对各通道沿 X 方向和 Y 方向进行编码, X 方向得到的一维特征为:

$$z_c^h(h)=\frac{1}{W}\sum_{0\leqslant i<W}x_c(h,i) \tag{1}$$

Y 方向得到的一维特征为:

$$z_c^w(w)=\frac{1}{H}\sum_{0\leqslant j<H}x_c(j,w) \tag{2}$$

其中, h 与 w 为输入特征图的高度和宽度; c 为所在通道数.

(2) 坐标注意力生成

将池化操作生成的两个特征图 $z_c^h(h)$ 与 $z_c^w(w)$ 通过 cascade 级联, 再利用 1×1 卷积进行 F_1 转换操作, 获得注意力图:

$$f=\delta(F_1([z^h,z^w])) \tag{3}$$

其中, δ 为非线性激活函数 ReLU; F_1 是将 X 方向与 Y 方向的池化结果进行级联操作.

然后经过批量归一化和 H-Swish 函数的非线性操作, 将 f 分为 f^h 和 f^w 两个独立张量. 利用两个 1×1 卷积 F_h 和 F_w 将特征图 f^h 和 f^w 变换为和输入同样的通道数, 得到两个方向的权重:

$$g^h=\sigma(F_h(f^h)) \tag{4}$$

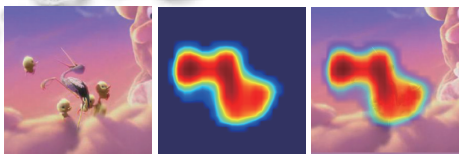
$$g^w=\sigma(F_w(f^w)) \tag{5}$$

其中, σ 为 Sigmoid 激活函数.

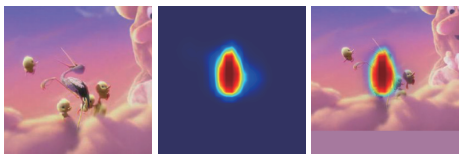
最后将 g^h 和 g^w 拓展为与输入特征具有相同维度, 与输入特征 $x_c(i,j)$ 按位相乘得到坐标注意力模块最终输出为:

$$y_c(i,j)=x_c(i,j)\times g_c^h(i)\times g_c^w(j) \tag{6}$$

采用 AlexNet 和本文结合坐标注意力模块的 MobileNetV3S 特征提取网络分别在 OTB100 数据集的 bird2 序列场景中提取特征信息, 特征图如图 3 所示.



(a) AlexNet 提取特征图



(b) MobileNetV3S 提取特征图

图 3 提取特征图分析

由图 3 可以看出, 经过 MobileNetV3S 提取多层特征信息并融合坐标注意力的输出特征图不容易受周围环境及相似物干扰, 响应得分图高的区域与所跟踪目标外观几乎重合.

1.3 自适应模板更新模块

SiamRPN 选择视频的首帧作为模板图像, 不能适应后期目标的形变、遮挡等场景, 因此引入自适应模板更新模块来解决这个问题, 将首帧模板信息与模板更新的输出融合并应用于当前帧的跟踪, 以提升应对复杂场景跟踪框漂移的稳健性, 其结构如图 4 所示.

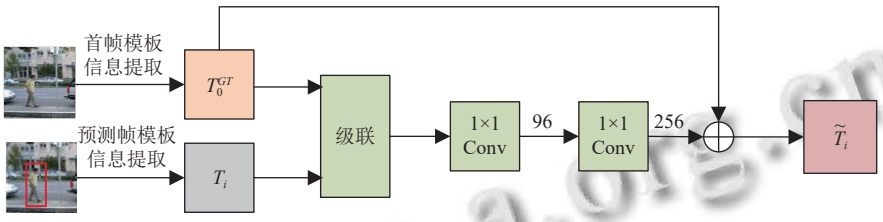


图 4 自适应模板更新模块

模板更新的计算过程如下.

(1) 将视频首帧模板信息T_0^{GT}和当前帧预测的模板信息 T_i 级联;

(2) 级联的特征经过 1×1×2×C×96 的卷积层、ReLU 层和 1×1×96 的卷积层, 与首帧模板信息融合, 输出更新模板. 其中, C 代表T_0^{GT}与 T_i 的通道数.

模板更新函数为:

$$\tilde{T}_i = \phi(T_0^{\mathrm{GT}}, T_i) + T_0^{\mathrm{GT}} \tag{7}$$

其中, $\tilde{T}_i$为自适应模板更新模块输出, T_0^{GT}为初始第 1 帧给定的目标模板信息 (GT 代表 ground truth), T_i 为当前帧预测到模板信息, ϕ为级联后经过模板更新的卷积层.

2 实验结果与分析

2.1 实验环境

实验处理器采用 CPU Intel(R) Xeon(R) CPU E5-26304 @ 2.20 GHz 2.20 GHz (2 处理器), RAM 为 64 GB, 编程语言使用 Python 3.7, 在 PyTorch 1.7.1 框架下进行跟踪操作. 在 PyCharm 2018.2.2.0 中进行程序的搭建及调试.

2.2 数据集

本实验训练集采用了 GOT-10k 数据集[22], 训练后选用了 100 个包含多种挑战属性视频序列的 OTB100 数据集[23] 和包含 123 个由低空无人机捕获的视频序列的 UAV123 数据集[24] 进行测试, 采用精度 (precision) 和成功率 (success rate) 作为性能评价指标.

2.3 OTB100 数据集实验结果分析

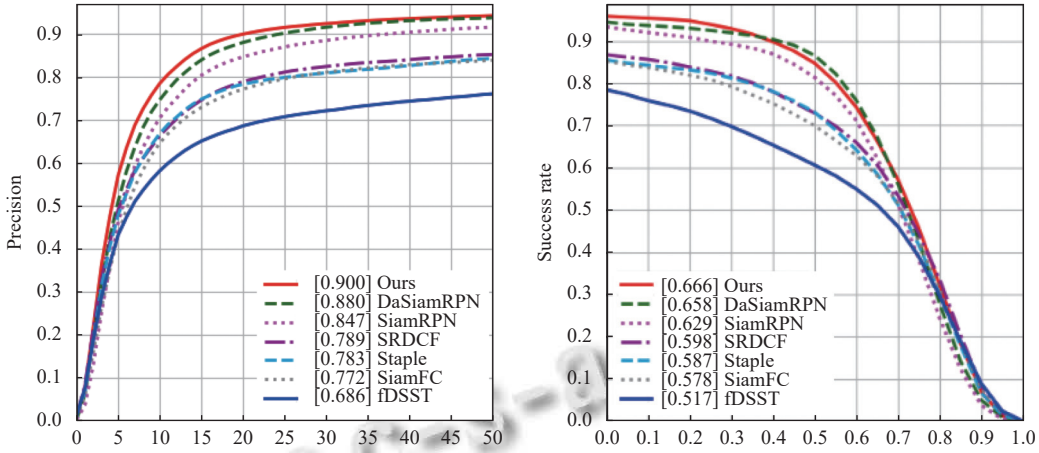
2.3.1 定性分析

OTB100 数据集包括了离开视野、光照变化、尺度变化、遮挡、形变等多种接近现实情况并具有挑战性复杂场景, 为了充分验证本文算法的跟踪能力, 将本文提出的算法 MCUSiamRPN (使用 Ours 表示) 与具代表性的相关滤波及孪生网络算法进行对比, 对比算法包括: fDSST、Staple、SRDCF、SiamFC、SiamRPN、DaSiamRPN, 实验结果如图 5 所示. 由图 5 可看出, 相关滤波类的跟踪算法精度和成功率都比较低, 其中跟踪效果相对较好的 SRDCF 的跟踪精度仅达到了 0.789, 成功率达到了 0.598, 跟踪效果最差的 fDSST, 跟踪精度为 0.666, 成功率为 0.517. 孪生网络算法跟踪效果除了 SiamFC 外, 其他算法较高, SiamRPN 与 DaSiamRPN 的跟踪精度都在 0.840 以上, 成功率也都在 0.620 以上. 本文算法在精度和成功率上都优于 6 种对比算法, 与基准算法 SiamRPN 相比有很大提升, 精度提升了 5.3%, 达到了 0.900; 成功率提升了 3.7%, 达到了 0.666.

2.3.2 定量分析

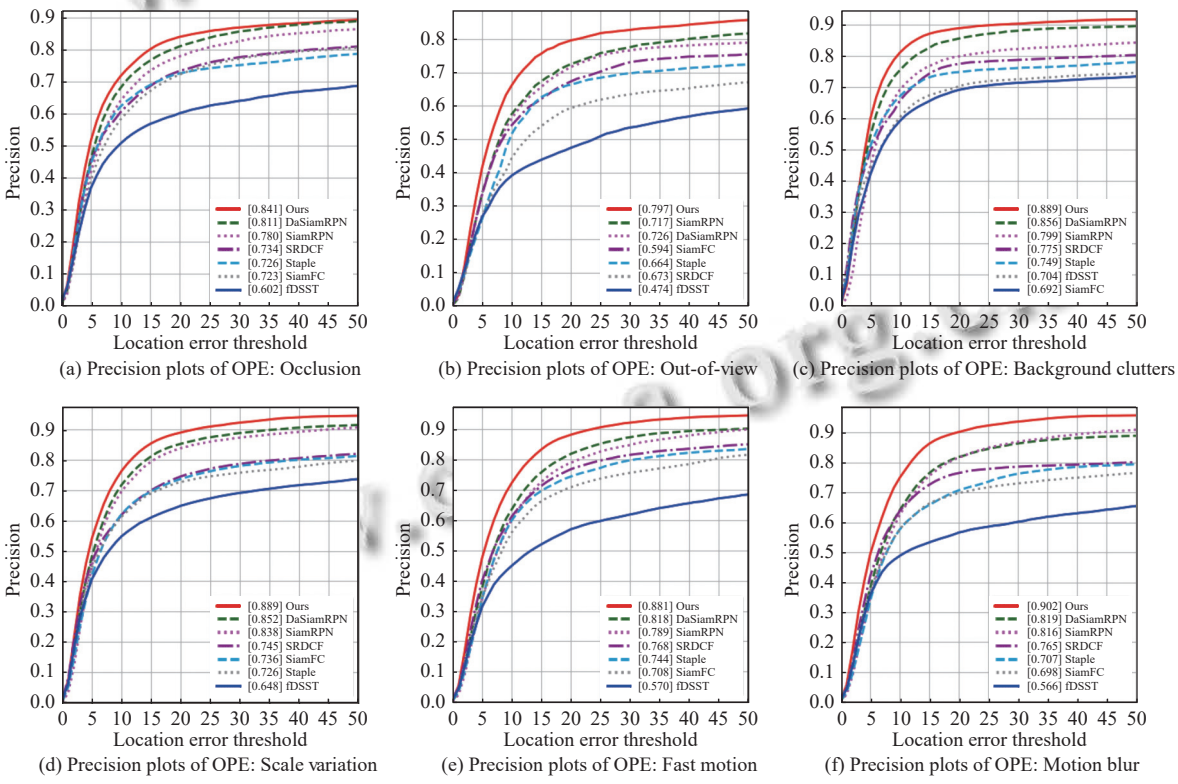
为了深入验证本文算法应对不同复杂场景的跟踪性能, 在遮挡 (occlusion)、离开视野 (out-of-view)、背景干扰 (background clutters)、尺度变化 (scale variation)、快速移动 (fast motion)、运动模糊 (motion blur) 这 6 个常见的复杂场景进行跟踪测试. 跟踪结果的精度如图 6 所示, 成功率如图 7 所示. 由图 6、图 7 可以看出, 在这 6 个场景中, 基于孪生网络的跟踪算法整体跟踪性能要优于相关滤波的跟踪算法, 尤其是应对离开视野的场景, 优越性更加明显. MCUSiamRPN 的精度和成功率都高于其他算法, 尤其应对离开视野、快速移动和运动模糊 3 个场景, 优势更加突出, 跟踪精度达到了 0.797、0.881、0.902, 成功率达到了 0.616、0.650、0.661; 而 DaSiamRPN 在这 3 个场景的跟踪精度分别为 0.717、0.818、0.819, 成功率为 0.537、

0.621、0.625; 跟踪效果较好的相关滤波算法 SRDCF 跟踪精度与成功率仅为 0.594、0.768、0.765 与 0.460、0.597、0.594. 在这 6 种场景中与基准算法 SiamRPN 相比, 精度分别提升了 6.1%、7.1%、9.0%、5.1%、9.2%、8.6%, 成功率分别提升了 4.3%、7.4%、6.3%、3.2%、5.1%、3.9%.



(a) Precision plots of OPE on OTB100 (b) Success plots of OPE on OTB100

图 5 OTB100 数据集精度与成功率对比



(a) Precision plots of OPE: Occlusion (b) Precision plots of OPE: Out-of-view (c) Precision plots of OPE: Background clutters

(d) Precision plots of OPE: Scale variation (e) Precision plots of OPE: Fast motion (f) Precision plots of OPE: Motion blur

图 6 部分复杂场景精度图

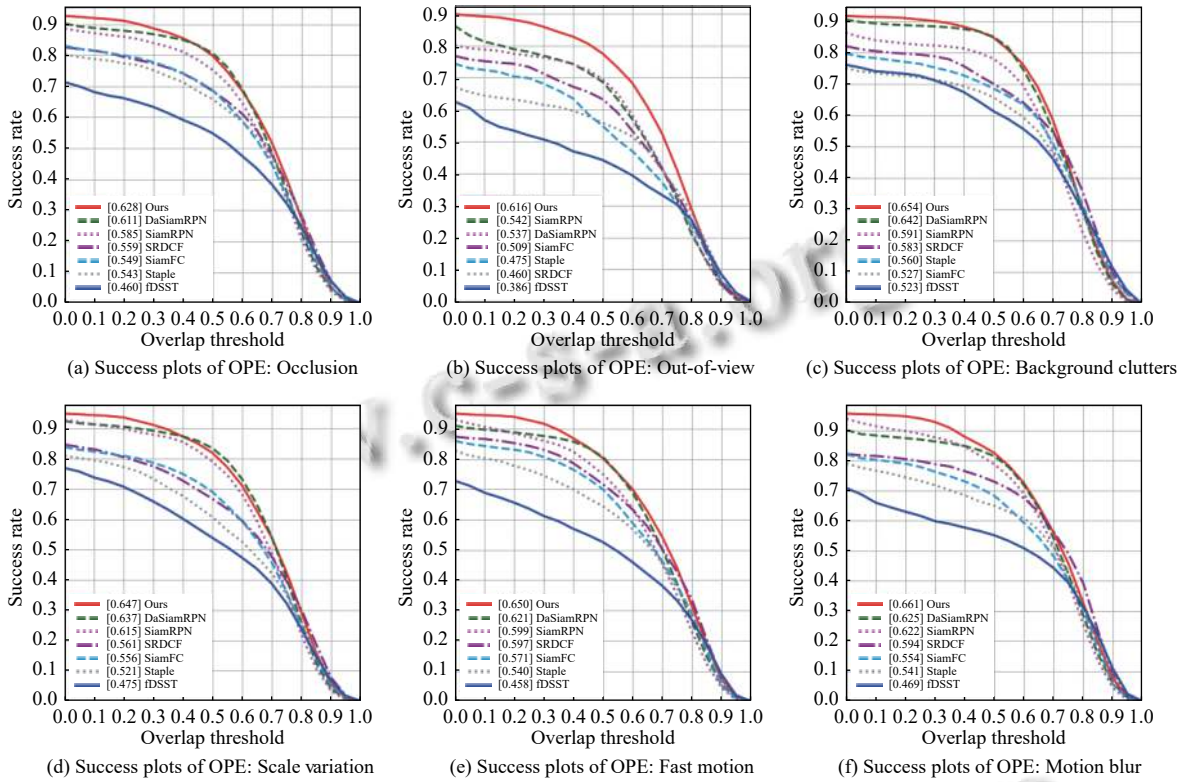
2.3.3 不同场景的可视化结果分析

采用本文算法和 6 种比较算法在 OTB100 数据集 4 个视频序列进行可视化结果分析, 如图 8 所示.

Bolt2 视频序列中, 存在目标快速运动、相似物过多以及短时遮挡的影响, 相关滤波算法中, SRDCF 表现良好, 但 Staple 和 fDSST 跟踪性能较差; 孪生网络算法中, SiamRPN 在 196 帧由于相似物过多, 跟踪框发生轻微飘逸; SiamFC 由于其网络结构简单, 跟踪飘逸严

重. 本文算法和 DaSiamRPN 可以较好地应对这个场景.

Biker 视频序列中, 跟踪目标快速移动导致运动模糊和目标尺度变化, SiamRPN 完成了跟踪任务, 但 84 帧尺度不能适应快速移动后的目标; DaSiamRPN 在目标瞬时掉头后产生了跟踪框飘逸, 84 帧飘逸出目标. 其他相关滤波算法和孪生网络算法都在目标掉头 70 帧后跟踪性能下降, 跟踪框漂移到目标外; 本文算法可以较好地适应尺度的变化并完成跟踪.



(a) Success plots of OPE: Occlusion (b) Success plots of OPE: Out-of-view (c) Success plots of OPE: Background clutters

(d) Success plots of OPE: Scale variation (e) Success plots of OPE: Fast motion (f) Success plots of OPE: Motion blur

图 7 部分复杂场景成功率图

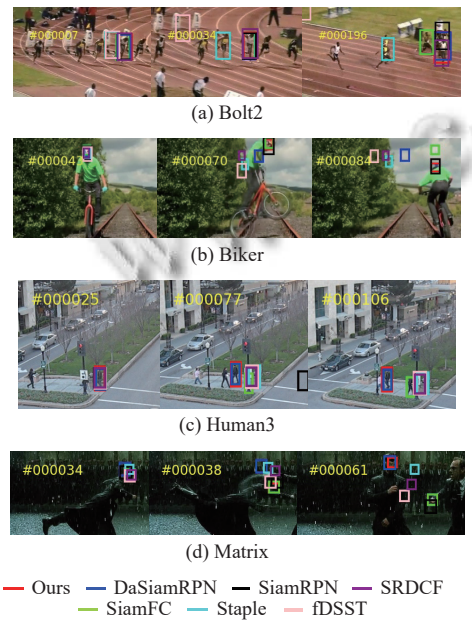


图 8 部分视频序列跟踪结果

Human3 视频序列中, 目标存在着运动模糊、遮挡及相似物等干扰, 目标穿过人群, 其他算法都发生了跟踪错误, 跟踪框转移到附近相似物; 本文算法和 DaSiamRPN 在整个序列能够准确跟踪目标.

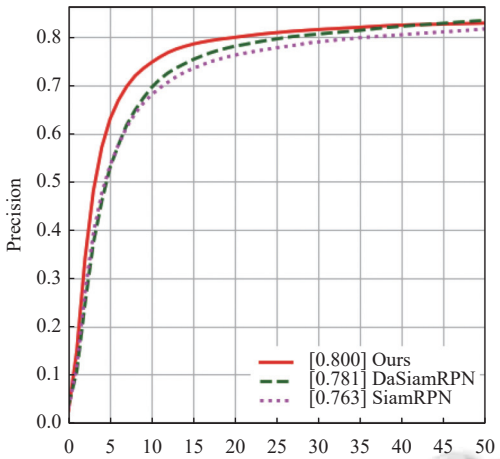
Matrix 视频序列中, 涉及运动模糊、遮挡、背景干扰等因素影响, 本文算法全程实现了更加准确的跟踪目标; DaSiamRPN 的跟踪框在 38 帧发生短暂飘逸, 由于其具有从局部到全局的搜索策略, 在 61 帧跟踪框又回到目标周围; 除 DaSiamRPN 的其他算法在经过 38 帧后, 由于场景复杂程度过高, 导致目标跟丢.

2.4 UAV23 数据集实验结果分析

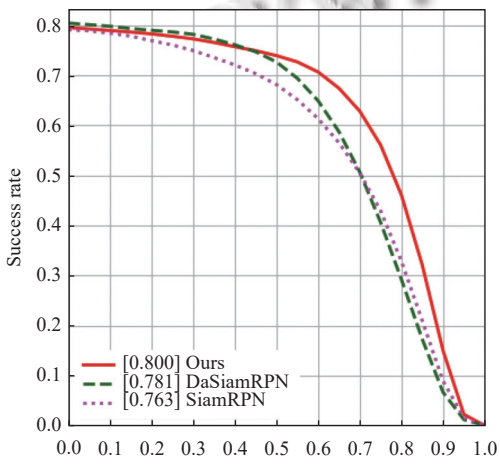
为了确保算法适应不同类型场景, 本文将 SiamRPN、DaSiamRPN、和 MCUSiamRPN (使用 Ours 表示) 在 UAV123 数据集上进行测试, 实验结果如图 9 所示.

由图 9 可得, MCUSiamRPN 算法的精度达到 0.800, 高出 SiamRPN 算法 3.7%; 成功率达到 0.608, 高

出 SiamRPN 算法 5.2%, 也都高于采用模板更新策略的 DaSianRPN, 可见本文算法在同类算法中优势明显.



(a) Precision plots of OPE on UAV123



(b) Success plots of OPE on UAV123

图 9 UAV123 数据集精度与成功率对比

将 MCUSiamRPN 与 SiamRPN、DaSiamRPN 这 3 种算法在 UAV123 数据集中 5 种复杂场景进行试验, 对比算法的跟踪效果. 跟踪精度和成功率结果如表 2 所示.

表 2 3 种算法在 5 种场景跟踪效果

算法	SiamRPN		DaSiamRPN		MCUSiamRPN	
	精度	成功率	精度	成功率	精度	成功率
相机运动	0.784	0.561	0.786	0.581	0.810	0.610
尺度变化	0.733	0.529	0.754	0.544	0.775	0.585
快速运动	0.710	0.479	0.737	0.520	0.756	0.537
离开视野	0.695	0.484	0.693	0.509	0.775	0.567
纵横比改变	0.725	0.505	0.756	0.537	0.760	0.562

由表 2 可得, 在所选相机运动、尺度变化、快速运动、离开视野与纵横比改变 5 种不同复杂场景下, 本文算法相较于 DaSiamRPN 与 SiamRPN 在各类场景的跟踪精度和成功率都有一定领先, 特别是在离开视野场景中, 本文算法和 SiamRPN 算法比较, 精度和成功率分别超出了 8.0% 和 8.3%, 达到了 0.775 与 0.567. 由此可见, 本文算法可较好的应对 UAV123 中的这些复杂场景, 具有较强的鲁棒性.

2.5 消融实验

为了验证 MCUSiamRPN 在使用 MobileNetV3S 进行多层特征融合、坐标注意力及模板更新这 3 个改进点的有效性, 本文在 OTB100 数据集进行消融实验, 在 SiamRPN 算法的基础上, 分别引入各个改进点进行对比, 其中多层特征融、坐标注意力模块与模板更新都依赖于 MobileNetV3S, 因此无法直接单独导入. 本实验通过控制变量所做 7 个对比实验得到的结果如表 3 所示.

实验结果表明, 本文提出 4 个部分的改进, 分别对跟踪精度和成功率起到了一定作用. 实验 2 结果, 表明使用 MobileNetV3S 进行特征提取后效果提升较好, 实验 3 表明, 特别是多层特征融合之后, 效果明显. 实验 4 和实验 6 表明, 坐标注意力模块的效果稍优于模板更新模块. 实验 7 表明, 4 个部分共同改进的结果最优.

表 3 消融实验结果

对比实验	SiamRPN	MobileNetV3S	多层特征融合	坐标注意力模块	模板更新	精度	成功率
实验1	√	—	—	—	—	0.847	0.629
实验2	√	√	—	—	—	0.863	0.641
实验3	√	√	√	—	—	0.875	0.649
实验4	√	√	√	√	—	0.887	0.655
实验5	√	√	—	—	√	0.873	0.645
实验6	√	√	√	—	√	0.884	0.651
实验7	√	√	√	√	√	0.900	0.666

注: √表示所选模块

3 结束语

本文提出融合坐标信息及模板更新的孪生网络跟踪算法, 通过 MobileNetV3S 提取各层特征并分别送入坐标注意力模块进行多层融合, 丰富语义信息; 为了解决目标跟丢及形变问题, 使用改进的自适应模块更新模板, 整合首帧与当前帧模板信息, 自适应更新的目标模板信息, 有效的应对目标因前景背景发生变化产生与初始目标帧差别过大的问题. 由实验可分析得出, 本文所提算法 MCUSiamRPN 相比于 SiamRPN 精度和成功率都有了很大提升, 能够较好地应对多种复杂场景, 特别对在目标离开视野、快速运动等场景具有良好的

稳性, 证明了本文算法的有效性.

参考文献

1 韩瑞泽, 冯伟, 郭青, 等. 视频单目标跟踪研究进展综述. 计算机学报, 2022, 45(9): 1877–1907.

2 王广玉, 窦磊, 窦杰. 基于自适应卡尔曼滤波的多目标跟踪算法. 计算机应用, 2022, 42(S1): 271–275.

3 Zhang LP, Nie JH, Zhang SL, *et al*. Research on the particle filter single-station target tracking algorithm based on particle number optimization. Journal of Electrical and Computer Engineering, 2021, 2021: 2838971.

4 Yu HQ, Sharma A, Sharma P. Adaptive strategy for sports video moving target detection and tracking technology based on mean shift algorithm. International Journal of System Assurance Engineering and Management, 2021. 1–11.

5 虞跃洋, 史泽林, 刘云鹏. 基于前景感知的时空相关滤波跟踪算法. 激光与光电子学进展, 2019, 56(22): 221503.

6 Henriques JF, Caseiro R, Martins P, *et al*. High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583–596. [doi: 10.1109/TPAMI.2014.2345390]

7 Danelljan M, Häger G, Khan FS, *et al*. Learning spatially regularized correlation filters for visual tracking. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 4310–4318.

8 孟琭, 李诚新. 近年目标跟踪算法短评-相关滤波与深度学习. 中国图象图形学报, 2019, 24(7): 1011–1016. [doi: 10.11834/jig.190111]

9 Bertinetto L, Valmadre J, Golodetz S, *et al*. Staple: Complementary learners for real-time tracking. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 1401–1409.

10 盖荣丽, 蔡建荣, 王诗宇, 等. 卷积神经网络在图像识别中的应用研究综述. 小型微型计算机系统, 2021, 42(9): 1980–1984. [doi: 10.3969/j.issn.1000-1220.2021.09.030]

11 Bertinetto L, Valmadre J, Henriques JF, *et al*. Fully-convolutional Siamese networks for object tracking. Proceedings of the Computer Vision Workshops. Amsterdam: Springer, 2016. 850–865.

12 Li B, Yan JJ, Wu W, *et al*. High performance visual tracking with siamese region proposal network. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8971–8980.

13 Zhu Z, Wang Q, Li B, *et al*. Distractor-aware siamese networks for visual object tracking. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 103–119.

14 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe: ACM, 2012. 1097–1105.

15 Howard A, Sandler M, Chen B, *et al*. Searching for MobileNetV3. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 1314–1324.

16 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13713–13722.

17 Zhang LC, Gonzalez-Garcia A, van de Weijer J, *et al*. Learning the model update for siamese trackers. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 4009–4018.

18 Li B, Wu W, Wang Q, *et al*. SiamRPN++: Evolution of siamese visual tracking with very deep networks. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 4277–4286.

19 Howard AG, Zhu ML, Chen B, *et al*. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.

20 Sandler M, Howard A, Zhu ML, *et al*. MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520.

21 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.

22 Huang LH, Zhao X, Huang KQ. GoT-10k: A large high-diversity benchmark for generic object tracking in the wild. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1562–1577. [doi: 10.1109/TPAMI.2019.2957464]

23 Wu Y, Lin J, Yang MH. Object tracking benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834–1848. [doi: 10.1109/TPAMI.2014.2388226]

24 Mueller M, Smith N, Ghanem B. A benchmark and simulator for UAV tracking. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 445–461.

(校对责编: 牛欣悦)