

文章编号: 2095-4980(2024)11-1304-08

基于深度学习的视频序列运动目标自适应跟踪

李嘉琪

(西北大学现代学院 电影学院, 陕西 西安 710130)

摘要: 针对视频序列中外观变化、背景杂波和严重遮挡等因素导致的目标跟踪精确度低的问题, 提出一种新型的双阶段自适应跟踪模型。该模型包含目标检测和边界框估计 2 个阶段: 在目标检测阶段, 模型对目标进行粗略定位; 在边界框估计阶段, 精确确定目标位置。为应对视频场景复杂性及小目标跟踪的挑战, 采用了多特征融合技术构建丰富的目标表示。实验结果表明, 与在线和实时跟踪(SORT)、Tracktor++、FairMOT、Transformer 等模型相比, 本模型表现出最优的综合性能, 有效平衡了计算速度与跟踪精确度之间的关系, 展现出良好的应用潜力。

关键词: 计算机视觉; 目标跟踪; 目标检测; 边界框估计; 判别相关滤波器

中图分类号: TP393

文献标志码: A

doi: 10.11805/TKYDA2024179

Adaptive tracking of moving targets in video sequences based on deep learning

LI Jiaqi

(School of Film, Modern College of Northwest University, Xi'an Shaanxi 710130, China)

Abstract: In response to the issues of low tracking accuracy in video sequences due to factors such as appearance changes, background clutter, and severe occlusions, a novel two-stage adaptive tracking model is proposed. This model includes two phases: target detection and bounding box estimation. In the target detection phase, the model roughly locates the target; in the bounding box estimation phase, the exact position of the target is determined. To address the complexity of video scenes and the challenges of tracking small targets, multi-feature fusion technology is employed to construct a rich target representation. Experimental results show that compared with models such as Simple Online and Realtime Tracking(SORT), Tracktor++, FairMOT, and Transformer, this model demonstrates the best overall performance, effectively balancing the relationship between computational speed and tracking accuracy, and showing good potential for application.

Keywords: computer vision; target tracking; object detection; bounding box estimation; Discriminant Correlation Filter(DCF)

近年来, 视频序列中的运动目标跟踪技术^[1-3]作为机器视觉的重要分支, 已引起了广泛的研究, 并在交通监控、无人机导航、智慧医疗等领域^[4-5]展现出巨大的应用潜力。根据不同的应用场景, 运动目标跟踪可细分为单目标跟踪和多目标跟踪两大类。单目标跟踪通常涉及在视频首帧中手动设定目标初始位置, 并在后续帧中持续预测其边界框位置; 多目标跟踪则侧重于对视频中出现的所有目标进行实时追踪。目前, 运动目标跟踪领域的研究已取得显著进展: 文献[6]提出一种考虑遮挡因素的自适应跟踪方法, 利用尺度不变特征变换(Scale Invariant Feature Transform, SIFT)提升跟踪能力; 文献[7]融合了神经过程和运动模型, 开发出一种新颖的机动目标跟踪算法。但面对外观变化、背景杂波以及严重遮挡等挑战, 实现鲁棒性跟踪仍是一个待解的难题。

传统的目标跟踪方法, 如在线实时跟踪(SORT)^[8]和FairMOT^[9], 已得到广泛应用。随着深度学习技术的蓬勃发展, 其强大的拟合能力为目标跟踪领域带来了新的突破: 文献[8]提出了RN-Tracktor++算法, 通过FRCNN-FPN(Faster Region-based Convolutional Networks with Feature Pyramid Networks)检测器和关系网络实现多目标跟踪的精准化; 文献[9]提出了一种基于时空一致性的多目标跟踪方法, 有效解决了空间、时间以及特征维度上的不一致性; 文献[10]采用卷积神经网络(Convolutional Neural Network, CNN)和Transformer构建双分支主干网络, 通过

特征提取和融合实现了多目标跟踪的高效性。

这些主流方法通过加深网络深度提高了跟踪精确度，但高昂的计算成本也限制了其在实际应用中的推广。为平衡跟踪精确度与计算成本，本文提出一种创新的双阶段运动目标自适应跟踪模型。该模型采用两阶段框架，首先通过目标检测阶段实现目标的粗略定位，随后利用边界框估计阶段对目标位置进行精确估计，在确保跟踪精确度的同时降低了计算成本。

1 双阶段运动目标跟踪模型

1.1 模型结构

所提出的双阶段运动目标跟踪模型结构如图 1 所示，该模型由 2 个阶段组成：目标检测阶段和边界框估计阶段。目标检测阶段基于判别相关滤波器(DCF)跟踪目标；边界框估计阶段基于实例感知注意网络和实例感知交并比网络(IoU-Net)估计目标的准确区域。

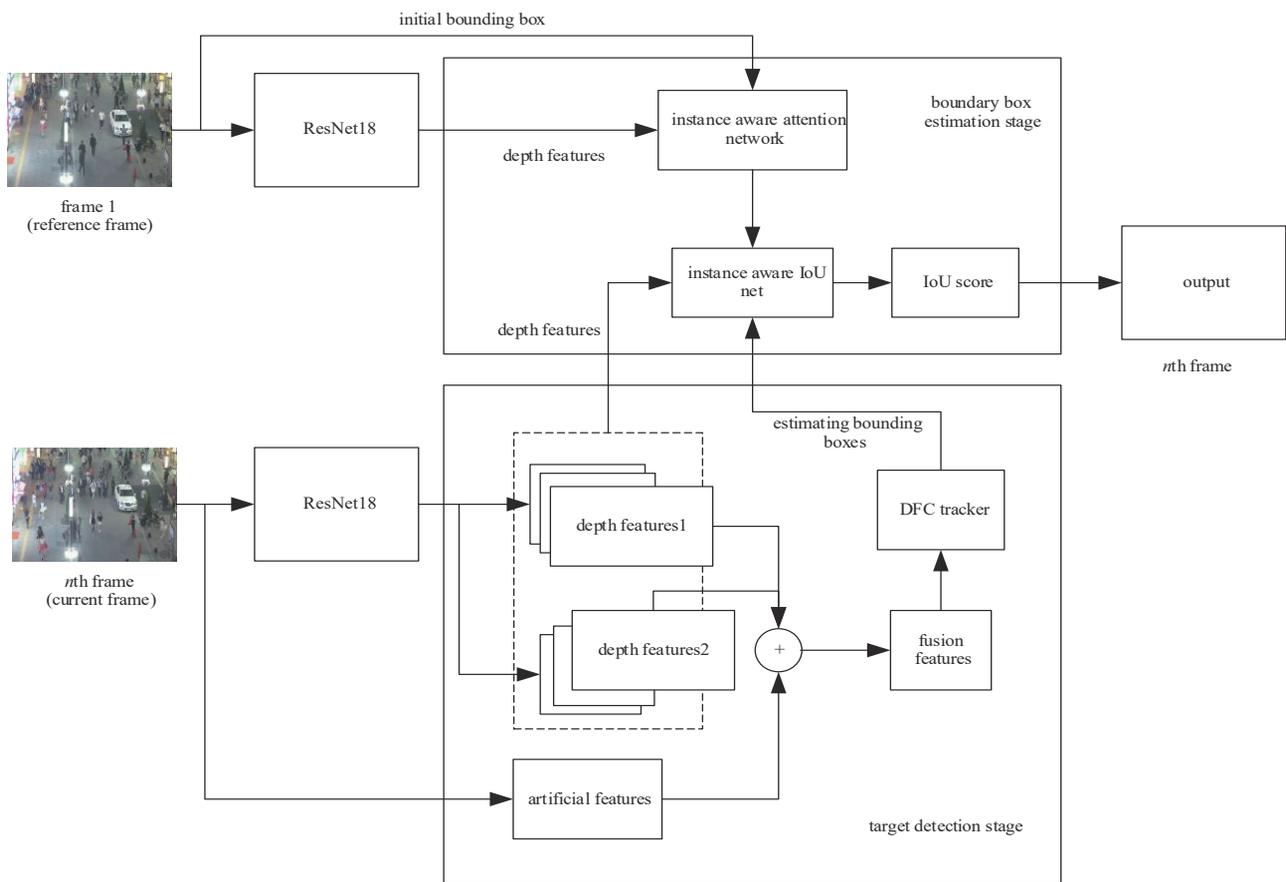


Fig.1 Structure of the proposed two-stage video sequence motion target tracking model
图 1 所提出的双阶段视频序列运动目标跟踪模型结构

目标检测阶段：基于 DCF 的跟踪器基本功能是粗略定位目标位置。为在视频场景中获得更好的性能和更强大的 DCF 跟踪器，本文设计了融合多级深度特征和人工特征的多特征融合方法。考虑到某些视频中目标的尺度相对较小，采用在 ImageNet 上训练的 ResNet-18 作为特征提取的骨干网络，避免在更深的网络中由于池化和卷积操作而导致的信息特征丢失。

边界框估计阶段：根据目标检测阶段给出的建议准确估计目标边界框。为提升视频序列中运动目标跟踪精确度，边界框估计阶段包含实例感知注意力和实例感知 IoU-Net 2 个子网络。实例感知注意力子网络用于对给定目标的基于实例的先验知识进行建模；实例感知 IoU-Net 子网络用于估计当前框架中不同建议的 IoU 分数。边界框估计阶段包含以下 4 个输入：当前视频序列帧的深度特征、从当前帧生成的初始边界框区域、参考帧的深度特征、参考帧中的估计目标边界框。在跟踪过程中，通过梯度上升最大化 IoU 获得最终的边界框。

1.2 目标检测阶段

为使基于 DCF 的跟踪器目标识别的计算效率和鲁棒性之间达到更好的平衡, 本文基于 DCF 探索了一种用于视频运动目标跟踪场景的目标定位模型。DCF 推理过程可理解为从一组标记样本 $(x, y) = \{(x_k, y_k)\}_{k=1}^T$ 中学习多个滤波器(其中 T 为样本总数), 每个训练样本 x_k 包含从图像的感兴趣区域(Region Of Interest, ROI)提取的 c 维特征图。假设所有样本具有相同的空间大小 $W \times H$, 则在每个空间位置 $(w, h)(w \in W, h \in H)$ 可使用 c 维特征向量 \mathbf{x}_k^c 表示 x_k 的第 c 个特征层。所需滤波器 F 包含由每个特征层学习的 1 组 $W \times H$ 卷积滤波器, 滤波器 F 与大小为 $W \times H$ 的样本 x_k 卷积而获得的输出特征 $R(x_k)$ 可计算为:

$$R(x_k) = \sum_{c=1}^C \mathbf{x}_k^c \otimes f^c \quad (1)$$

式中: C 为从样本 x_k 中提取的特征的总数; f^c 为与第 c 个特征层相对应的大小为 $W \times H$ 的滤波器参数。

滤波器参数可通过最小化训练样本 x_k 的输出特征 $R(x_k)$ 和标签 y_k 之间的误差 $\varepsilon_T(f)$ 获取, 其表达式为:

$$\varepsilon_T(f) = \sum_{k=1}^T \alpha_k \|R(x_k) - y_k\|^2 + \lambda \sum_{c=1}^C \|f^c\|^2 \quad (2)$$

式中 α_k 和 λ 为超参数。其中 $\alpha_k \geq 0$, 决定了每个训练样本的影响; $\lambda \geq 0$, 为正则项的权重。

考虑到视频场景的复杂性和小目标带来的挑战, 采用多特征融合方法构建丰富的目标表示。具体采用了 2 个人工特征: 一个是对空间结构不敏感的颜色直方图分数; 另一个是对于光照变化具有鲁棒性的梯度直方图(Histogram Of Gradient, HOG)描述符。同时, 还使用了 ResNet-18 中第 1 和第 3 卷积层的深度特征, 以充分利用深度特征的代表能力。人工特征可作为深度特征的补充, 解决从深度神经网络中难以提取微小目标特征的难点。此外, 当目标和场景在相邻帧之间仅轻微变化时, 上述基于 DCF 的跟踪器可以准确地估计目标的新位置。但在真实的视频中, 视点、照明、背景和目标尺度也发生了巨大变化。因此, 该阶段的目标定位结果仅作为粗定位, 并在下一阶段进一步结合边界框估计值准确计算目标的位置。

1.3 边界框估计阶段

CNN 的低级特征包含更多的纹理信息, 高级特征包含更多的语义信息。本文将第 1 卷积层的特征与 ResNet-18 的高级特征相结合, 避免目标信息的丢失。边界框估计阶段网络结构如图 2 所示。为使网络更加关注目标, 在边界框估计阶段, 实例感知注意力网络为不同的特征通道赋予不同的权重。跟踪过程中, 实例感知 IoU-Net 以这些权重计算边界框的 IoU 分数, 使用梯度上升最大化 IoU 细化边界框, 并通过融合 IoU 得分前 3 的结果获得最终边界框。

实例感知注意力网络可描述为:

$$\omega(z, B_0) = v(P_{RP}(\beta(z), B_0)W_{\text{chn}}) \quad (3)$$

式中: $\beta(z)$ 为参考帧 z 的卷积特征; B_0 为参考帧中归一化的初始边界框; W_{chn} 为通过通道注意力获得的权重; $P_{RP}(\cdot)$ 为 ROI 池化操作; $v(\cdot)$ 为应用特征通道注意力后的卷积、连接和全连接层操作。

$P_{RP}(\cdot)$ 可以获取每个通道中基于 ROI 的特征, 该值主要通过任意给定区域 B 中特征 β 的平均池化操作计算:

$$P_{RP}(\beta, B) = \frac{\int_{y_1}^{y_2} \int_{x_1}^{x_2} g(x, y) dx dy}{(x_2 - x_1) \times (y_2 - y_1)} \quad (4)$$

式中: (x_1, y_1) 和 (x_2, y_2) 为边界框 B 的左上角和右下角的坐标; $g(x, y)$ 为对特征 β 执行双线性插值运算后的连续特征值。

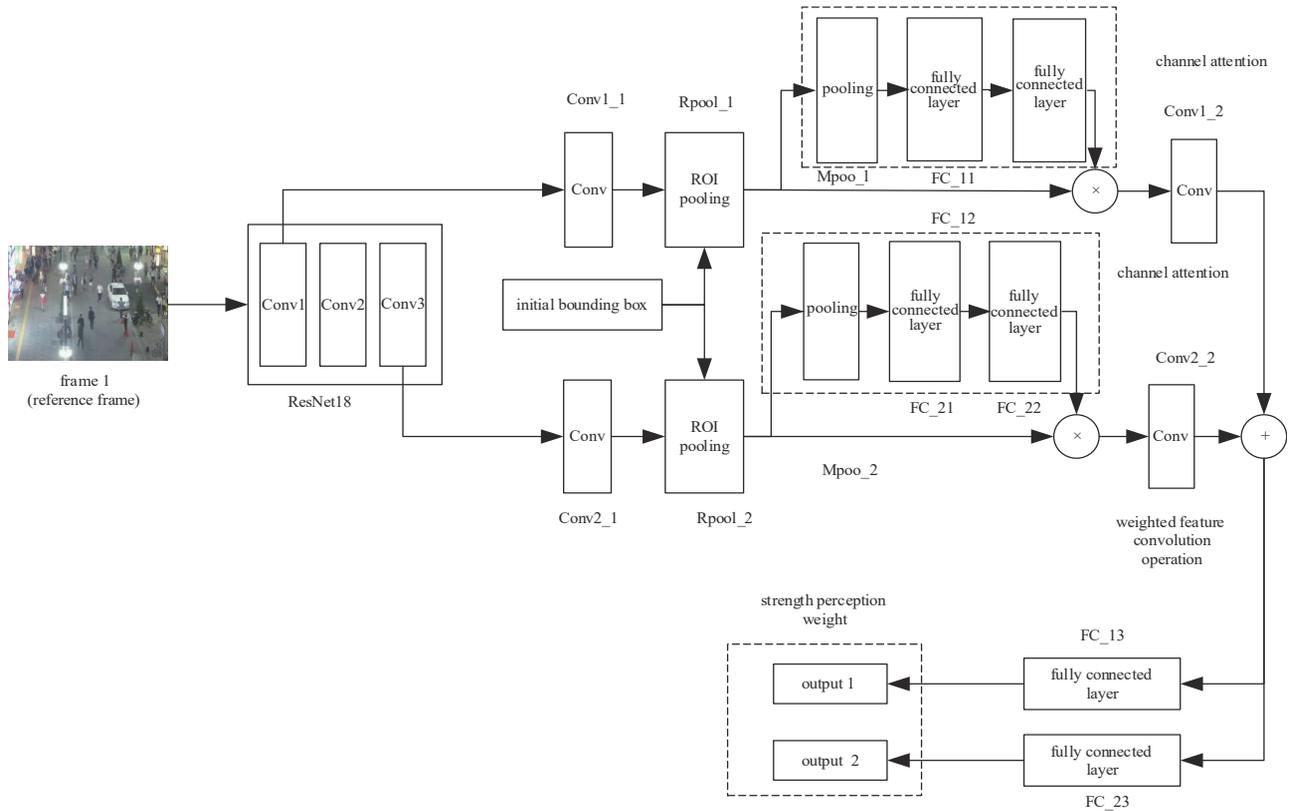
通道注意力模块通过计算 W_{chn} 向特征通道分配权重, 抑制与目标无关的特征通道。该通道注意力模块执行 3 个操作: a) 每个通道执行全局池化操作, 使结果包含全局感受野表征全局响应; b) 考虑到特征的灵活性和简单性, 使用 2 个级联的全连接层(第 1 层压缩通道数量, 第 2 层将通道数量恢复到原始值), 以进一步学习得到一组参数, 对特征通道之间的相关性进行建模; c) 执行加权特征卷积运算, 使卷积核的大小与每个通道的特征维数一致。

实例感知 IoU-Net 用于评估当前帧 x_i 中第 i 个边界框 B_i 的 IoU 得分, 其公式为:

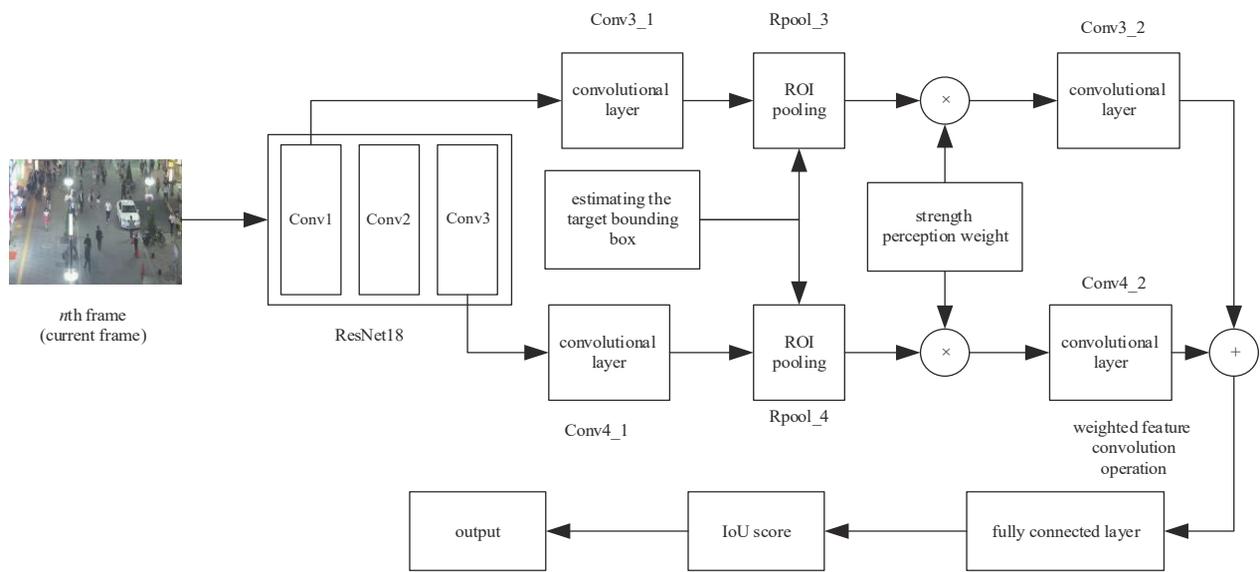
$$s_{IoU}(x_t, B_t^i) = p(\omega P_{RP}(\beta(x_t), B_t^i)) \tag{5}$$

式中： s_{IoU} 为 IoU 得分； $\beta(x_t)$ 为样本 x_t 的卷积特征； ω 为包含实例的先验信息，即参考图像中的实例感知值。

实例感知 IoU-Net 首先对当前帧的卷积特征进行融合运算，以适应特定的跟踪任务；然后，将特征与边界框执行加权特征卷积运算，并最大化 IoU 分数，获得更准确的跟踪结果。



(a) instance-aware attention network structure



(b) instance-aware IoU-net network structure

Fig.2 Network structure of bounding box estimation stage

图2 边界框估计阶段网络结构

1.4 网络训练

网络训练前，设定了图像块对最小化预测值 s_{IoU} 和真值之间的误差。为提升训练性能，首先使用大型跟踪数

据集 TrackingNet 和 VisDrone2018 进行预训练。实例感知注意力网络配置及参数设置如表 1 所示。其中, Conv 表示卷积层; Rpool 表示 ROI 池化层; Mpool 表示最大池化层; FC 为全连接层。后接数字用于区分不同层, 如 Conv1_1 表示第 1 组第 1 个卷积层。从表 1 可以看出, 实例感知注意力网络由 2 组对称结构组成, 所有卷积层的卷积核大小设置为 3×3 , 不同组池化层的 2 个超参数分别设置为 5×5 和 3×3 。

实例感知 IoU-Net 的配置和参数设置详见表 2。该网络主要由 4 个 3×3 卷积核的卷积层、2 个超参数分别为 7×7 和 5×5 的 ROI 池化层以及 1 个 512×1 的全连接层组成。所有卷积层和全连接层均接入归一化和 ReLU 激活函数。在训练过程中, 会从视频不同帧中采样一对图像块, 帧间隔设为 40。具体而言, 为获得实例感知向量值, 以目标为中心, 在原始图像上采样一个正方形的图像块, 其大小为目标的 5.2 倍。对于 IoU 分数预测分支, 使用相同的策略对区域进行采样, 但给位置和尺度添加了一些扰动, 模拟目标的运动。对于每对图像块, 将高斯噪声添加到真值以生成 16 个候选图像, 并且这些候选图像与真值之间的最小 IoU 小于 0.1。整个训练过程中, 本文使用图像翻转、平移、色差抖动增强数据。目标检测阶段, 根据前一帧的目标位置裁剪出一个 150×150 的图像块作为目标搜索区域, 提取该区域中 4 个单元的 HOG 特征和颜色直方图特征; 然后基于前一帧的目标位置, 对正方形图像块进行采样, 其面积约为目标面积的 5 倍, 输入到 ResNet-18 以获得目标的深度特征, 对每个特征图执行相关性滤波操作, 获得深度特征图; 最后, 将所有特征图线性插值到相同分辨率 (75×75) 并采用线性加权方式进行融合, 获得目标的最终位置。

边界框估计阶段, 通过在第一帧计算实例感知向量, 其维度为 $1 \times 1 \times 256$ 。对于随后的帧, 通过实例感知 IoU-Net 评估目标边界框。该过程采用梯度上升最大化 IoU 得分, 并通过融合前 3 个结果获得最终跟踪结果。

2 实验与分析

2.1 数据集和实验设置

实验时使用的数据集包括 PathTrack 和 MOT16 数据集。PathTrack 数据集是一个用于人物追踪的影视视频数据集, 包含 720 个视频序列和 15 000 多条人物轨迹。MOT16 包含 14 个视频序列, 分为训练和测试集, 分别用静态和动态摄像机拍摄。PathTrack 和 MOT16 数据集包含不同复杂背景、不同尺度情况下行人、车辆等检测场景, 可有效验证不同模型跟踪性能。

实验时将所有数据划分为训练和测试两部分。对于 PathTrack 数据集, 每个视频划分为固定百分比的帧, 即将 80% 的帧分为训练集, 其余 20% 用于验证。训练时跟踪器使用在 ImageNet 上预训练的 ResNet-18 作为骨干网络。实验参数设置如下: 初始学习率设置为 0.05, 学习率衰减因子为 0.1, 衰减周期为 100, 最大迭代次数为 10 000。

所提双阶段目标跟踪框架由 PyTorch 0.12.1+cu11.3 搭建, 并基于 PyCharm 编译。训练时硬件参数: 操作系统为 Ubuntu 18.04; CPU 为 Intel® Xeon® Processor E7-4870、64 G RAM; 显卡为 Nvidia RTX 3060, 共 2 块。

2.2 消融研究

为验证所提双阶段目标跟踪框架的有效性, 设计了 2 个消融实验: a) 对不同策略下网络配置进行测试; b) 对多特征融合方法进行测试。

第 1 个消融实验中对比了 3 组网络配置策略: 策略 1, 目标检测阶段仅使用 DCF 定位目标; 策略 2, 使用目标检测及边界框估计, 首先使用 DCF 定位目标并生成 16 个候选区域, 然后使用 IoU-Net (不添加实例感知权重) 预测每个候选区域的 IoU 分数。在这个过程中, 通过 5 个梯度优化连续修改候选区域的坐标, 最后通过融合前 3 个

表 1 实例感知注意力网络配置及参数设置

Table 1 Configuration and parameter settings of instance-aware attention network

name	convolutional kernel	stride
Conv1_1	3×3	1
Rpool_1	5×5	-
Mpool_1	5×5	1
FC_11	256×16	-
FC_12	16×256	-
Conv1_2	3×3	1
FC_13	512×256	-
Conv2_1	3×3	1
Rpool_2	3×3	-
Mpool_2	3×3	1
FC_21	256×16	-
FC_22	16×256	-
Conv2_2	3×3	1
FC_23	512×256	-

表 2 实例感知 IoU-Net 配置及参数设置

Table 2 Configuration and parameter settings of instance-aware IoU-Net

name	convolutional kernel	stride
Conv3_1	3×3	1
Rpool_3	7×7	-
Conv3_2	3×3	1
Conv4_1	3×3	1
Rpool_4	5×5	-
Conv4_2	3×3	1
FC	512×1	-

建议结果获得最终边界框；策略 3，采用所提双阶段框架，与策略 2 唯一不同之处为该框架使用实例感知 IoU-Net 预测目标的规模，并获得最终的边界框。消融实验时选取 ROC 曲线下面积(Area Under the ROC Curve, AUC)和精确度作为测试指标。

不同策略下网络配置消融实验对比结果如表 3 所示。可以看出，策略 1 中仅使用目标检测网络及 DCF，性能最低，AUC 和精确度指标分别为 0.474 8 和 0.723 5；策略 2 和策略 3 中使用了双阶段框架，性能明显提升。策略 3 中，IoU-Net 添加实例感知权重后，AUC 指标提高了 0.064 3。实验结果验证了双阶段框架的有效性，并表明 IoU-Net 通过在特征图上引入实例感知权重，有助于降低网络对各种干扰的敏感性，提升网络性能。

表3 不同策略下网络配置消融实验对比结果

Table3 Comparison results of network configuration ablation experiments under different strategies

model	AUC	precision
object detection+DCF	0.474 8	0.723 5
object detection+bounding box estimation+DCF+IoU-net(without instance-aware weights)	0.534 5	0.787 9
object detection+bounding box estimation+DCF+IoU-net(with instance-aware weights)	0.598 8	0.843 7

在 IoU-Net 使用实例感知权重情况下，分别对多特征融合方法进行消融实验，实验结果如表 4 所示。其中，策略 1 为仅使用人工特征方法；策略 2 为仅使用深度特征方法；策略 3 为使用所提多特征融合方法。可以看出，策略 1 在所有属性视频序列中的 AUC 分数最低。策略 3 与策略 1 相比，AUC 和精确度指标分别提升 0.103 6 和 0.098 5；与策略 2 相比，AUC 和精确度指标分别提升 0.021 6 和 0.025 1。

表4 多特征融合方法消融实验结果

Table4 Results of ablation experiments using multiple feature fusion methods

method	AUC	precision
artificial features	0.476 9	0.714 8
depth features	0.558 9	0.788 2
multi-feature fusion	0.580 5	0.813 3

2.3 综合对比

选取 IDF1、HOTA 和每秒帧数(Frames Per Second, FPS)指标评价不同模型综合性能。其中，IDF1 指标用于测量目标关联性能；HOTA 指标旨在平衡测量检测和目标关联综合性能；FPS 表征模型推理速度。实验时将所提双阶段跟踪框架与主流方法进行了比较，包括 SORT^[8]、Trackor++^[10]、FairMOT^[11]、Transformer^[12]。

不同模型综合对比结果如表 5 所示，可以看出，SORT 的 IDF1、HOTA 指标表现最差，得分明显低于其他主流方法。这主要是因为 SORT 使用卡尔曼滤波器基于估计的运动关联轨迹，模型鲁棒性较差，跟踪性能欠佳。与 SORT、Tracktor++、FairMOT、Transformer 等主流方法相比，除 MOT16 数据集中 HOTA 指标略低于 Transformer，所提模型 IDF1 和 HOTA 指标均表现出较优的跟踪性能。与 SORT 相比，所提模型在 PathTrack 数据集中 IDF1 和 HOTA 指标分别提升 18.5% 和 18.4%；MOT16 数据集中 IDF1 和 HOTA 指标分别提升 54.4% 和 41.9%。

表5 不同模型综合对比结果

Table5 Comprehensive comparison results of different models

dataset	model	IDF1	HOTA	FPS
Path Track	SORT	0.448 5	0.360 5	58.7
	Tracktor++	0.527 4	0.404 1	14.2
	FairMOT	0.529 5	0.414 3	18.3
	Transformer	0.530 7	0.421 3	7.9
	the proposed model	0.531 5	0.426 8	24.5
MOT16	SORT	0.398 1	0.338 5	60.3
	Tracktor++	0.545 9	0.448 6	18.9
	FairMOT	0.548 5	0.424 6	23.6
	Transformer	0.596 3	0.482 5	12.4
	the proposed model	0.605 1	0.480 3	26.8

与 Tracktor++、FairMOT、Transformer 及所提模型相比，SORT 模型 FPS 最高，约为 58.7。主要是由于 SORT 方法不需要大量的模型计算以及复杂的网络资源消耗，因此计算速度快，FPS 性能较优。Transformer 模型的 FPS 最低，尤其是在 PathTrack 数据集中已低于 10，表明 Transformer 模型计算更为复杂，需要消耗更多的计算资源。所提模型 FPS 约为 25，可以在通用图形处理单元(Graphic Processing Unit, GPU)计算平台上进行近实时处理，与 Transformer 相比，速度提升 1.5 倍。主要是由于所提模型采用了相关滤波操作跟踪目标，并在边界框估计阶段基于实例感知注意网络和实例感知 IoU-Net 估计目标的准确区域。

MOT16 数据集中存在街道等复杂背景，且行人由远及近尺度差异变化较大。以 MOT16 数据集为例，对所提模型识别效果及不同模型跟踪误差进行验证分析。图 3 为识别效果展示图，可以看出，在复杂的街道背景及不同

尺度行人条件下, 所提模型均能较好地跟踪行人。图 4 为跟踪误差曲线对比结果, 可以看出, 与 SORT、Tractor++、FairMOT、Transformer 等主流模型相比, 所提模型跟踪精确度较高, 具有较优的跟踪性能。



Fig.3 Display of recognition effect
图3 识别效果展示

综上所述, 所提模型表现出最优的综合性能, 可在复杂背景、多尺度场景下有效跟踪目标; 同时, 可平衡计算速度与跟踪精确度之间的关系, 具备实际工程应用价值。

3 结论

本文研究了视频序列运动目标跟踪方法, 提出了一种双阶段的运动目标跟踪模型。该模型包括目标检测段和边界框估计 2 个阶段, 可有效处理视频中背景杂波、尺度变化和视场变化等应用场景。通过实验验证, 所提出的双阶段模型具备较高的检测精确度和较优的处理速度, 该方法可为视频序列运动目标跟踪领域的发展提供一定的借鉴, 具有较好的应用前景。

参考文献:

- [1] 李长江, 肖文显, 王俊阁. 基于相似度优化的混合式视觉跟踪方法[J]. 太赫兹科学与电子信息学报, 2022, 20(11): 1198–1204. (LI Changjiang, XIAO Wenxian, WANG Junge. Hybrid visual tracking method based on similarity optimization[J]. Journal of Terahertz Science and Electronic Information Technology, 2022, 20(11): 1198–1204.) doi:10.11805/TKYDA2020400.
- [2] 吴宏涛. 基于改进 Camshift 的道路运动车辆跟踪算法[J]. 山西交通科技, 2023(6): 103–107. (WU Hongtao. Road moving vehicle tracking algorithm based on improved Camshift[J]. Shanxi Science & Technology of Transportation, 2023(6): 103–107.)
- [3] 李思腾, 刘梦佳, 刘晓阳. 基于深度学习网络的运动目标跟踪研究[J]. 智能城市, 2023, 9(11): 9–11. (LI Siteng, LIU Mengjia, LIU Xiaoyang. Research on motion target tracking based on deep learning networks[J]. Smart City, 2023, 9(11): 9–11.) doi:10.19301/j.cnki.zncs.2023.11.003.
- [4] 梅建华, 杨彦辰, 云利军, 等. 多视角人体红外步态信息采集与处理[J]. 云南师范大学学报(自然科学版), 2022, 42(4): 36–40. (MEI Jianhua, YANG Yanchen, YUN Lijun, et al. Multi-view infrared gait information acquisition and processing[J]. Journal of Yunnan Normal University(Natural Science Edition), 2022, 42(4): 36–40.) doi:10.7699/j.ynnu.ns-2022-048.
- [5] 王金磊, 魏同权, 邓亮, 等. 车载手部小目标运动跟踪算法研究[J]. 传感器与微系统, 2023, 42(8): 65–68, 77. (WANG Jinlei, WEI Tongquan, DENG Liang, et al. Motion tracking algorithm research on vehicle-mounted small target of hands[J]. Sensors and Microsystems, 2023, 42(8): 65–68, 77.) doi:10.13873/J.1000-978(2023)08-0065-04.
- [6] 窦海波. 考虑遮挡因素的视频人体运动目标自适应跟踪[J]. 吉林大学学报(信息科学版), 2023, 41(3): 566–573. (DOU Haibo. Adaptive tracking algorithm for video human moving objects considering occlusion[J]. Journal of Jilin University(Information Science Edition), 2023, 41(3): 566–573.) doi:10.19292/j.cnki.jdxp.2023.03.006.
- [7] 朱洪峰, 熊伟, 崔亚奇, 等. 神经过程和运动模型混合驱动的机动目标跟踪算法[J]. 西安交通大学学报, 2023, 57(4): 152–161. (ZHU Hongfeng, XIONG Wei, CUI Yaqi, et al. Hybrid-driven maneuvering target tracking algorithm based on neural process[J]. Journal of Xi'an Jiaotong University, 2023, 57(4): 152–161.) doi:10.7652/xjtxxb202304016.

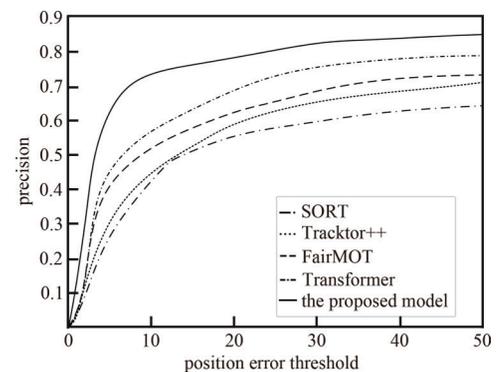


Fig.4 Comparison results of tracking error curves
图4 跟踪误差曲线对比结果

- [8] BEWLEY A, GE Z Y, OTT L, et al. Simple online and realtime tracking[C]// 2016 IEEE International Conference on Image Processing(ICIP). Phoenix,AZ,USA:IEEE, 2016:3464–3468.doi:10.1109/ICIP.2016.7533003.
- [9] ZHANG Yifu,WANG Chunyu,WANG Xinggang, et al. FairMOT:on the fairness of detection and re-identification in multiple object tracking[J]. International Journal of Computer Vision, 2021(129):3069–3087. doi:10.1007/s11263-021-01513-4.
- [10] 姬张建,薛冰心. 融合关系网络的Tracktor++多目标跟踪算法[J]. 山西大学学报(自然科学版), 2023,46(5):1076–1084. (JI Zhangjian,XUE Bingxin. Combining relation network in Tracktor++ for multi-target tracking[J]. Journal of Shanxi University (Natural Science Edition), 2023,46(5):1076–1084.) doi:10.13451/j.sxu.ns.2023015.
- [11] 彭嘉淇,王涛,陈柯安,等. 结合时空一致性的FairMOT跟踪算法优化[J]. 中国图象图形学报, 2022,27(9):2749–2760. (PENG Jiaqi,WANG Tao,CHEN Kean, et al. Spatio-temporal consistency based FairMOT tracking algorithm optimization[J]. Journal of Image and Graphics, 2022,27(9):2749–2760.) doi:10.11834/jig.220116.
- [12] 张英俊,白小辉,谢斌红. CNN-Transformer特征融合多目标跟踪算法[J]. 计算机工程与应用, 2024,60(2):180–190. (ZHANG Yingjun,BAI Xiaohui,XIE Binhong. Multi-object tracking algorithm based on CNN-Transformer feature fusion[J]. Computer Engineering and Applications, 2024,60(2):180–190.) doi:10.3778/j.issn.1002-8331.2211-0028.

作者简介:

李嘉琪(1990-), 女, 硕士, 讲师, 主要研究方向为影视传媒 .email:egolge401@163.com.

(上接第 1282 页)

- [11] CAROSI D,CESAROTTI F,LASAPONARA L,et al. Delay control in wideband front-end components for multi-role phased-array systems[C]// 2011 IEEE MTT-S International Microwave Symposium. Baltimore,MD,USA:IEEE, 2011:1. doi:10.1109/MWSYM.2011.5973446.

作者简介:

陈月盈(1987-), 女, 硕士, 高级工程师, 主要研究方向为微波毫米波单片集成电路及器件模型设计. email:chyy0225@163.com.

刘会东(1982-), 男, 硕士, 研究员级高级工程师, 主要研究方向为微波毫米波单片电路设计.

杨 柳(1987-), 女, 硕士, 高级工程师, 主要研究方向为微波毫米波单片集成电路设计.

赵子润(1984-), 男, 硕士, 高级工程师, 主要研究方向为 III-V 族化合物半导体集成电路设计.