

# 计算机视觉中的目标检测与跟踪算法研究

刘 聪

广州城市理工学院 广东广州 510800

**摘要：**在机器视觉中，目标检测和跟踪是实现智能感知的核心技术，在无人驾驶、安全监控和机器人导航等领域有着广泛的应用。在复杂动态环境中，由于目标遮挡、姿态多变、光照变化等因素的影响，使得算法的实时性和鲁棒性受到了极大的挑战。本文以目标检测和跟踪算法为研究对象，从场景特性和目标属性两个方面入手，构建融合多源信息的目标检测算法框架，包括检测模型选择、跟踪策略设计和动态自适应机制。以智能监控场景为研究对象，通过实验验证，该方法在有遮挡情况下跟踪准确率达到92%，帧速维持在30帧以上，为复杂场景下视觉感知任务提供技术支持。

**关键词：**计算机视觉；目标检测；跟踪算法

随着深度学习和传感技术的快速发展，计算机视觉已经由静态图像分析过渡到动态场景理解，其中目标检测和跟踪是智能系统感知能力的核心环节。在实际应用中，复杂背景下的目标往往面临快速运动、部分/全遮挡、表观突变等干扰，传统算法很容易出现检测漂移或无法跟踪的问题。如无人驾驶场景中的行人突入，安全监控中突然出现的灯光变化等，对算法的实时响应和动态调整能力提出了更高的要求。因此，研究同时具备高精度和强自适应能力的目标检测与跟踪算法，对于推进智能交通系统的实用化具有重要的理论和现实意义。

## 一、目标检测与跟踪的技术挑战分析

### （一）复杂场景的干扰因素

复杂场景下的动态干扰是导致目标检测跟踪性能瓶颈的重要原因，目标表观的非刚体变化，如人体姿态的伸屈、车辆转向等，造成目标特征分布的剧烈波动，传统的模板匹配方法很难适应<sup>[1]</sup>。环境噪声具有动态耦合性，如车辆进出隧道等气象因素，如车辆进出隧道等，会导致像元值畸变，影响特征提取的稳定性。由于空间交互而产生的遮挡，将其划分为多个目标之间的相互遮挡（如人群中的相互遮挡）和背景上的遮挡（如树木对行人的遮挡），当遮挡比例大于50%时，大多数算法的跟踪精度都会降低30%。

### （二）算法性能的核心指标

对检测和跟踪算法进行评价时，需要综合考虑多个维度的指标。在探测任务中，平均精度（mAP）反映对不同尺度物体的识别能力，特别是小目标（ $<32 \times 32$ ）的检测率，直接影响到安防监控中对远程目标的感知效果。

中心位置误差（CLE）和跟踪成功率（SOT）是衡量跟踪精度的两个重要指标<sup>[2]</sup>。

### （三）技术演进的约束条件

在技术演化过程中，算法性能受到硬件计算能力和场景复杂性的双重制约。虽然深度学习模型提高检测精度，但是参数的急剧增加也导致了计算开销成倍增加，如基于Transformer的单GPU检测模型，其帧频只有传统CNN模型的1/5。同时，动态场景的不可预测性（如突然闯入目标等）对算法的在线学习能力提出更高的要求，同时模型参数的实时更新又需要额外的计算资源，形成“精度-速度-适应性”的三角矛盾。

## 二、目标检测算法的设计要点

### （一）特征提取架构的选型

目标检测的基础是特征提取，框架选择必须与场景特征分布相匹配，采用自顶向下的特征金字塔（FPN）结构，融合高层语义和底层细节信息，提高弱小目标的检测能力<sup>[3]</sup>。如在交通监测场景下，与单一尺度网络相比，该结构对远距离车辆（ $32 \times 32$ 个像素）的检测率提高20%。注意机制嵌入技术针对复杂场景，利用空间注意力抑制非相关区域（如广告牌等），增强关键特征（如车辆轮廓边缘），使特征响应信噪比提高15-25%。

### （二）检测框优化策略

检测框的定位精度直接影响后续跟踪的效果，需要采用双策略对其进行优化。锚框架自适应调整根据目标先验分布动态生成锚框架（如长宽比1:2.5）和车辆前方锚框架（1:1.2）等，将初始框架-目标的交并比IoU提高到0.7以上。改进非最大值抑制（NMS）方法，解决

重叠目标的误检测问题<sup>[4]</sup>。

### （三）轻量化设计方法

为满足实时性的要求，在保证精度损失的前提下，轻量化设计需要对模型进行压缩。深度可分卷积将标准卷积分解成深度卷积和逐点卷积，在保证识别准确率超过90%的前提下，可用于边缘设备（如嵌入式相机）。动态信道修剪技术可根据场景复杂度实时调整网络结构宽度，在简单背景情况下关闭30%冗余信道，使推理速度提高1.5倍，准确率低于3%。

## 三、目标跟踪算法的关键设计

### （一）跟踪初始化机制

在跟踪过程中，需要将检测结果和历史信息有机地结合起来。多帧融合算法利用前3-5帧的检测框拟合目标运动趋势，去除偶然出现的错误帧等噪声，将初始跟踪框架的定位误差降低到5个像素以内<sup>[5]</sup>。特征模板更新通过对目标稳定状态下的表现特征（如无遮挡情况下的衣服纹理等）进行再识别，当目标再次出现时，识别准确率达到90%以上，有效解决短时间遮挡后的跟踪连续性问题。

### （二）运动模型构建

运动模型对目标位置的变化进行预测，并对不同的运动状态进行自适应。针对匀速运动（如直行车辆）和变速运动（如行人突然加速），自适应卡尔曼滤波动态调整过程噪声协方差，减少20-30%的定位误差。利用光流辅助跟踪技术，针对快速运动（如球类等）目标，利用稀疏光流进行像素点位移估计，弥补检测帧之间的信息缺口，使跟踪中断率降低15%。在高速运动场景下，与单一模型相比，该方法的跟踪成功率可明显提高。

### （三）遮挡处理策略

遮挡问题是目标跟踪面临的一大难题，需要分层处理。当部分遮挡（<50%）被遮挡的情况下，利用视觉区域匹配技术提取出无遮挡的特征（如遮挡行人头部等），并结合运动预测进行跟踪。当完全遮挡（>50%）时，根据历史轨迹（如直线、曲线等）对目标进行外推，并在遮挡区域设置搜索窗口，当目标再现时触发再识别。在有车辆遮挡的情况下，提出的方法可以将跟踪的连续性提高40%，并且将遮挡时间由原来的5秒减少到2秒。

## 四、检测与跟踪的协同机制

### （一）时空信息融合

要实现检测和跟踪的协同，就必须打破模块之间的壁垒，使信息在两个方向上流动。在时间维度上，将跟踪结果反馈给探测模块，根据历史轨迹对当前探测区域

进行约束（例如缩小搜索区域），将检测速度提高30%。在空间维度上，检测框修正跟踪漂移，当跟踪框和检测框的IoU<0.5时，根据检测结果对跟踪状态进行复位，避免错误积累。对于动态场景（如购物中心等），时空融合算法可使系统整体鲁棒性提高25%。

### （二）动态适应策略

为适应环境的变化，算法需要具有在线自适应能力。在光照变化等环境噪声影响下，特征权值自适应调整，当光照强度较大时，增加边缘特征权值（60%），纹理特征权值降到40%，特征匹配稳定性提高20%。模型参数微调采用增量式学习方法对分类器进行更新，每隔100帧就增加一个新样本（如目标新姿态等），既避免过拟合，又保持模型自适应能力，使其在长时间跟踪任务（如24小时监测）下的精度下降率不超过5%。

### （三）多模态信息互补

单一视觉模式容易受到外界环境的约束，而多模式融合则可以提高视觉感知的可靠性。在夜间场景下，红外图像提供目标轮廓信息，可见光图像补充纹理细节，将检测率从65%提高到85%。该方法利用深度相机获取目标距离信息，并能准确识别出前景和背景干扰（如壁画中的人和真实行人），对复杂背景下的跟踪准确率提高30%以上。多模态数据融合特别适合于雾天、夜间等极端环境，拓展了算法的适用范围。

## 五、案例分析

### （一）应用场景概况

以城市快速路智能监控系统为研究对象，针对早高峰（7:00-9:00）、平峰（10:00-17:00）、晚高峰（18:00-20:00）和夜间（21:00-次日6点）四个时段，对机动车（轿车，货车，公交车）、非机动车（自行车、电动车）和行人三类目标进行同步监测。场景中多种干扰因素：早高峰时段车辆密度高达80辆/分，容易出现交叉遮挡；中午的太阳直接照射在路面上，造成路面反射，亮度变化幅度在30%左右；夜间道路照明不均匀，部分地区照度不足20 lux；树木的影子随阳光而动，形成了一条每小时5公尺左右的动态阴影地带。系统硬件采用边缘计算设备（NVIDIA Jetson AGX Xavier），对检测精度 $\geq 90\%$ ，跟踪持续性 $\geq 85\%$ ，实时帧速率 $\geq 25$  fps，为交通流量统计、违规行为识别（如闯红灯、逆行等）提供支持。

### （二）算法方案设计

针对场景特性，设计融合多模态信息的检测跟踪协同框架，核心模块包括：

### 1. 检测模块优化

选择基于轻量级的特征金字塔网络 (MobileNetV3-FPN) 结构, 将图像的输入尺寸设置为 640x480 像素。在特征提取方面, 采用高斯核过滤路面反射区 (加权衰减系数 0.3)、通道注意层 (加权衰减系数为 1.2) 和行人姿态特征 (权重提高到 1.1) 的双重注意机制。采用动态锚框策略对检测框进行优化, 对车辆产生 1: 1.3 的矩形锚框架 (覆盖 92%)、非机动车和行人分别为 1: 2.8 和 1: 3.2 (匹配直立), 初始交并比 (IoU) 阈值设为 0.75。采用 Soft-NMS 对非最大值抑制 (NMS) 进行改进, 在 0.5-1.0 范围内根据 IoU 的变化动态调整置信权系数, 从而有效解决车辆密集干扰下的误抑问题。

### 2. 跟踪模块设计

跟踪融合前 5 帧的检测结果, 采用最小二乘法拟合目标运动轨迹 (线性度误差  $\leq 3$  个像素), 剔除偏离轨迹的误差 (比例  $\leq 5\%$ )。利用自适应卡尔曼滤波方法建立了运动模型, 对行人、非机动车等速度  $\leq 10$  米/秒的过程噪声协方差设为 0.02; 当车辆行驶在 10 米/s 以上时, 将协方差调整到 0.05, 提高对高速运动物体的预报精度。在轻微遮挡 (20%-50%) 的情况下, 提取可见区域特征 (如车窗、行人头部等), 并对其进行匹配; 在严重遮挡情况下 (50%-80%), 结合历史轨迹和光流量 (稀疏光流点密度  $\geq 15$  个/点) 进行定位预测。在完全遮挡 ( $>80\%$ ) 的情况下, 启动跟踪外推模式, 根据遮挡时间的长短, 动态扩展搜索窗口 (初始 200\*200 像素, 每帧放大 5%)。

### 3. 协同机制构建

时空融合采用“检测-追踪”双向反馈机制: 在时间维度上, 每隔 3 帧就会反馈一次跟踪轨迹到探测模块, 使目标搜索范围缩小到原来的 60%, 从而减少无效运算; 在空间维度上, 追踪框和检测框 IoU  $< 0.4$  时, 触发检测复位机制, 根据最新的检测结果对追踪状态进行修正。动态自适应策略包括特征权值调整 (夜间模式 60% 的红外特征权重) 和模型在线更新 (每 200 帧引入新样本, 学习率 0.001), 以保证长期跟踪的稳定性。

### (三) 实施效果

检测模块在全时段的平均精度 (mAP) 为 92.3%, 其中机动车检测准确率 94.7% (货车因遮挡略低, 为 91.2%), 非机动车 89.5% (电动车与自行车区分准确率 87.3%), 行人 90.8% (小目标行人  $< 32 \times 32$  像素检测率 85.6%)。跟踪模块的中心位置误差 (CLE) 平均 8.7 像

素, 跟踪成功率 (SOT) 89.4%, 在车辆交叉遮挡场景中, 跟踪续接准确率达 83.6% (传统算法为 62.1%)。实时性方面, 边缘设备处理帧率稳定在 32fps, 单帧处理延迟 28ms, 满足实时监控要求 (数据表格分析各干扰因素对算法的影响如表 1)。

表 1 数据表格分析各干扰因素对算法的影响

| 场景类型    | 检测准确率 (%) | 跟踪成功率 (%) | 帧率 (fps) | 主要影响因素    |
|---------|-----------|-----------|----------|-----------|
| 早高峰密集车流 | 90.2      | 86.5      | 31       | 车辆交叉遮挡    |
| 正午强光    | 89.7      | 87.3      | 32       | 路面反光、光照突变 |
| 夜间低光照   | 88.5      | 85.7      | 30       | 光线不足、对比度低 |
| 树荫动态遮挡  | 91.8      | 88.9      | 32       | 背景干扰、局部遮挡 |

### 结束语

综上, 目标检测与跟踪是计算机视觉领域一个非常有意义的研究课题。随着深度学习技术及硬件计算能力的不断提高, 目标检测与跟踪算法在性能提升的同时, 也面临着新的挑战与机遇。为适应日益复杂的应用需求, 应进一步关注算法的实时性、鲁棒性以及可扩展性。另外, 多模数据融合、跨域应用、与人工智能等其他领域深度融合是未来目标检测与跟踪的重要发展方向。通过研究与创新, 可望在目标检测与跟踪领域取得突破, 为促进人工智能技术的发展与应用提供强有力的支撑。

### 参考文献

[1] 沈金锐, 焦健. 计算机视觉在海洋石油支持船智能视频危险识别与预警系统中的应用 [J]. 大数据时代, 2024, (11): 48-52.

[2] 罗茜茜. 计算机视觉技术在无人驾驶汽车中的应用 [J]. 汽车测试报告, 2024, (22): 26-28.

[3] 李甫, 戴成元, 李微雨, 梁邦勋, 刘其舟. 基于计算机视觉的施工现场多目标检测方法 [J]. 湖北工程学院学报, 2024, 44 (06): 99-107.

[4] 闫鹏飞, 肖莎, 张志勇, 赵铖, 刘聪, 米凯臣, 刘静峰. 计算机视觉技术在现代畜牧业中的应用与研究进展 [J]. 中国猪业, 2024, 19 (05): 83-89.

[5] 雷佩莲. 计算机视觉技术在网络视频编辑中的应用 [J]. 信息与电脑 (理论版), 2024, 36 (16): 8-11.