

基于轻量化深度神经网络的监控视频中移动目标识别方法

高孟飞¹ 马 浩¹
GAO Mengfei MA Hao

摘 要

当前监控视频中移动目标的识别方法缺乏对特定区域和目标位置关系的关注,在目标移动范围较大或不规则时,识别误差累积导致失准。为此,文章提出了一种基于轻量化深度神经网络的监控视频中移动目标识别方法。该方法通过提取词袋特征,构建可视化特征库,实现对监控视频目标的初步分类。将初步分类的结果作为输入,利用轻量化深度神经网络对背景特征进行深入学习,准确地区分出背景与移动目标。通过比较差分图像的像素值与设定的动态阈值,对图像进行二值化处理,以简化图像信息并突出移动目标。采用中值滤波器对二值化后的图像进行噪声抑制,进一步提升图像质量。在处理后的图像上,通过设定敏感区域并依据质心与边界关系启动识别处理,使识别过程更具针对性。只有当移动目标质心超出预设的敏感边界时才启动目标精准识别,确保识别精度,减少误差累积。实验结果表明,该方法能够成功识别出所有移动目标,并且在识别效果上表现良好。

关键词

轻量化深度神经网络; 监控视频; 移动目标识别; 二值化; 噪声处理

doi: 10.3969/j.issn.1672-9528.2025.06.017

0 引言

随着监控技术的普及与发展,如何在海量监控视频中准确、高效地识别移动目标,已成为国内外学术界和工业界共同关注的热点问题。范建伟等人^[1]构建了基于改进YOLOv3的运动目标识别模型,通过残差模块深化网络、特征金字塔捕捉四尺度特征、K-means聚类优化候选框,并引入斥力损失函数,实现对不同帧视频图像中运动目标的精准识别。该模型未考虑目标与特定区域的位置关系,无法准确判断目标是否处于有效状态,识别效果差。赵佳冉等人^[2]通过工业摄像机采集图像,利用Otsu法进行阈值分割和轮廓逼近分类物体形状,结合灰度重心法与刚体变换估计运动轨迹,实现目标定位。由于该方法未识别目标与特定区域的关系,会按正常形状分类流程操作,致使后续运动目标识别精度受限。耿德新^[3]在YOLOv3框架上引入金字塔结构,结合深度残差网络与多尺度融合机制,构建高精度目标识别模型,全面识别大范围内小物体。多尺度融合机制未考虑目标与特定区域的关系,虽能提高识别精度,但无法确定目标是否处于特定区域内,影响监控中移动目标的识别效果。周献珍等人^[4]对运动视频图像进行中值滤波去噪,利用改进差分法检测运动目

标,并通过CamShift算法结合MeanShift迭代实现自动跟踪识别。此方法主要通过帧间差异检测运动目标,未考虑目标所在特定区域,无法正确记录和预警目标进入特定区域的情况,在动态环境下目标识别效果欠佳。针对上述不足,本文提出一种基于轻量化深度神经网络的监控视频中移动目标识别方法。该方法结合多种预处理和识别技术,能精确、高效地识别监控视频中的移动目标。

1 基于词袋模型的监控视频移动目标初步聚类方法

在监控视频中移动目标的识别过程中,由于视频内容复杂、背景多变、目标多样,直接对视频帧进行处理往往难以达到高效且准确的识别效果。为降低视频数据的复杂性并保留关键信息,本文采用了词袋模型。该模型能够提取出视频帧中的关键可视化特征,这些特征对于目标的识别至关重要。通过构建可视化特征库并进行目标聚类,将具有相似特征的目标进行准确归类,能够适应监控视频中目标的多样性,同时增强方法的适应性和鲁棒性。

利用词袋模型文档,进行特征矢量相似度聚类,将提取特征与特征字典比对后归类。若遇字典未涵盖特征则更新字典,纳入新特征以增强模型性能。词袋模型将整个监控视频帧视为一个可视文件,由多个可视单词构成。这些单词是通过无监督学习从视频帧中分解出来的,分解的依据是特征字典,每个部分都与字典中的单词相对应,整体

1. 郑州西亚斯学院计算机与软件工程学院 河南郑州 451100
[基金项目] 河南省2021年民办普通高等学校学科专业建设资助项目(软件工程专业)

则构成词组,代表原视频帧的不同部分^[5]。监控视频目标初步聚类步骤如下:

先对监控视频中的各个样本进行预处理,以提高图像质量。再采用超像素分割方法,将视频帧划分为多个部分,并将这些部分与“单词”相对应,完成词袋特征的提取。在图像分割处理阶段,计算目标点与超平面的距离,以进一步确定分割的准确性。计算公式为:

$$\gamma = \frac{1}{\|\mathbf{w}\|} |\mathbf{w}^T \mathbf{x} + b| \quad (1)$$

式中: b 表示判定阈值; T 表示分割时长; $\mathbf{w}$ 表示法向量。

通过最大化间隔的转换值 $\frac{1}{\|\mathbf{w}\|}$,并基于范数等效原理,将分类问题转化为求解 $\min \|\mathbf{w}\|^2$ 最小值问题,以确保分类间隔最大化。随后,在各类型的移动目标上提取可视化特征并进行人工分类标注,构建初步特征库。

基于初步可视化特征,用相似策略整合语义相近特征,完善可视化特征库。利用误差平方和准则函数计算各聚类子集数据均值,迭代归类优化。误差平方函数表达式为:

$$E = \sum_{i=1}^m |U - O_i|^2 \quad (2)$$

式中: m 表示划分个数; O_i 表示数据簇中心; U 表示样本数据; i 表示聚类次数。

在给定样本集中随机均匀选取初始聚类中心,计算各数据点与这些中心距离并归类,重新计算各类样本均值定位聚类中心 k ,重复直至聚类中心稳定或达到收敛条件,完成目标聚类。

通过上述过程能够将视频中的目标按照其视觉特征进行归类,为后续的目标识别奠定基础。初步分类的结果可能并不完全准确,但能够提供一个大致目标类别范围,帮助缩小后续识别的搜索空间,有助于提高目标识别的效率和准确性。

2 基于轻量化深度神经网络的监控视频中移动目标识别

2.1 基于轻量化深度神经网络的背景与目标区分

在监控视频中移动目标的识别过程中,背景信息对于准确区分移动目标和静态背景至关重要。传统背景建模方法容易受到光线变化、动态背景(如摇曳的树叶、流动的水面)等因素的干扰,导致背景与目标的区分度降低。为此,本研究提出基于轻量化深度神经网络的背景识别方法,以初步分类结果为部分输入,将动态背景信息作为关键识别依据。利用深度特征学习技术从真实背景像素提取更精细、鲁棒的特征,提升背景特征在识别中的置信度与区分能力,精确区分背景与移动目标^[6]。

为实现这一目标,采用基于轻量化深度神经网络的加权

特征比率来进行背景与目标的识别。特征提取阶段,以初步分类结果作为部分输入,利用轻量化神经网络从视频帧中提取特征,并对这些特征的权重进行标准化处理。标准化处理的计算公式为:

$$B = \arg \min \left(\sum_{i=1} w_i(x, y) \geq T_h \right) \quad (3)$$

式中: T_h 表示设定的阈值; $w_i(x, y)$ 表示权重标准化结果。通过该公式,可滤除不符合背景特征高斯分布或深度特征的部分,仅保留前 D 个最符合背景特征的模型用于背景识别。

为精确估计并识别背景,将当前视频帧中的每个像素点与通过轻量化深度神经网络学习到的背景特征模型进行匹配^[7]。设某像素点的值为 $I_i(x, y)$,判断其特征与背景特征模型均值是否满足匹配标准,判断公式为:

$$|I_i(x, y) - \bar{z}_{j-1}| < \mu \sigma_{i-1}(x, y) \quad (4)$$

式中: $\bar{z}_{j-1}$ 表示从视频开始到第 j 帧为止,通过轻量化深度神经网络学习到的背景特征均值模型; μ 表示基于深度学习设定的匹配常数; $\sigma_{i-1}(x, y)$ 表示从视频开始到第 j 帧为止,通过轻量化深度神经网络学习到的各像素点特征的方差。

以式(1)为依据,当像素点与背景特征模型中的均值之差满足式(1)时,说明该像素点与背景特征匹配成功,被识别为背景;否则,该像素点被识别为移动目标。

2.2 差分图像二值化处理

在移动目标识别时,利用轻量化深度神经网络区分背景与移动目标后,图像中的目标和背景已经有一个相对明确的划分。此时再进行二值化处理,就可以更加有针对性地对已经区分出的移动目标部分进行处理,避免将背景像素误判为目标像素或者反之的情况。这样既能消除差分运算造成的背景残留痕迹,又能凸显移动目标动态特性,为后续目标识别、跟踪等步骤奠定高效精确的基础。

针对监控视频中的每个像素点 (x, y) ,采用二值化处理,计算公式为:

$$f_i(x, y) = \begin{cases} f_{\text{background}}(x, y) < T \\ f_{\text{object}}(x, y) \geq T \end{cases} \quad (5)$$

式中: $f_{\text{background}}(x, y)$ 表示背景像素值; T 表示基于设定的动态阈值; $f_{\text{object}}(x, y)$ 表示移动目标像素值^[8]。实际二值化处理中,将差分图像各像素值与动态阈值比较,小于阈值 T 判定为背景像素,大于阈值 T 则判定为移动目标像素,确保目标与背景精确彻底分割,为后续处理分析提供可靠基础。

2.3 中值滤波提升二值化图像质量

在监控视频中移动目标识别过程中,二值化处理虽能将图像像素点有效划分为前景(移动目标)和背景两类,但在目标边缘或背景复杂时可能引入噪声,干扰后续识别结果导

致精度下降。因此，二值化处理后，需用中值滤波器抑制图像噪声，使图像更平滑，提升识别准确性与稳定性。

二值化处理之后，借助阈值判别策略来判断像素点是否归属于运动目标，计算公式为：

$$\Delta G_t(x, y) = |G_t(x, y) - G_{t-1}(x, y)| \quad (6)$$

式中： $G_t(x, y)$ 表示当前 t 时刻像素点的灰度级； $G_{t-1}(x, y)$ 表示前一时刻 $t-1$ 像素点的灰度级。若背景保持静态不变，且目标也处于静止状态，此时式（6）的计算结果将显示为 0；若目标正在运动，则计算结果会变为 1，由此可以判定该像素点属于移动目标。

移动目标在监控视频中常伴随着拖影现象，为了更精细地处理这些细节，引入了中值滤波器。中值滤波器是一种非线性窗口滤波技术，能够在去噪同时保留图像细微结构。其机制是对图像每个像素点，以相邻区域像素值排序后取中间值替代中心像素值，可消除相邻区域离散点（如噪声）干扰。

2.4 基于敏感区域和目标质心计算的移动目标智能识别

监控视频中并非所有移动目标都重要，缺乏筛选机制会使系统识别效率低，易误报和漏报。为此，本文引入敏感区域和目标质心计算的方法实现精确识别。通过这种方法，能够实现对移动目标的智能监控：目标进入预设敏感区域时，系统才进一步处理识别，实现监控资源有效利用。目标质心计算提供准确坐标点，帮助系统精确判断目标位置和移动轨迹，降低误报漏报率，提升识别准确性。

监控操作员需预先在监控场景内设定一个敏感区域，此区域通常为预设的矩形范围。设该矩形敏感区域为 R_e ，其左上角坐标为 (x_1, y_1) ，右下角坐标为 (x_2, y_2) 。若监控过程中捕获的移动目标的质心坐标位于敏感区域 R_e 内（即质心所在位置未超出 R_e 的边界），则视为系统已成功识别到该移动目标，并随即采取相应的处理措施；反之，若质心坐标不在 R_e 内，则系统不进行特别处理，继续维持监控状态。此判定逻辑可表述为：

$$W = \begin{cases} \text{true, if Area}_{R_e} \\ \text{false, else} \end{cases} \quad (7)$$

式中：**true** 表示监控视频成功识别到移动目标；**false** 表示未识别到移动目标； Area_{R_e} 表示上述定义的矩形敏感区域。移动目标质心坐标落入预设的敏感区域 R_e 时，视频监控会启动目标识别处理程序；未进入则继续监控，实现移动目标智能识别与高效处理。

3 实验分析

为证实基于轻量化深度神经网络的监控视频中移动目标识别方法的有效性，进行实验验证与分析。

3.1 实验指标

为了从数值上对监控视频移动目标识别效果进行评估，采用了 TA、TP 两种指标，计算公式为：

$$\begin{aligned} \text{TA} &= 1 - \frac{P_F + P_T}{\sum_t L_t} \\ \text{TP} &= \frac{\sum_i d_{t,i}}{\sum_t u_t} \end{aligned} \quad (8)$$

式中：TA 表示移动目标识别的误差积累程度； P_F 表示漏报率； L_t 表示在 t 时刻实际目标数量； P_T 表示误报率；TP 表示衡量移动目标匹配平均精准度； u_t 表示成功匹配的目标数量； $d_{t,i}$ 表示 t 时刻第 i 个目标匹配误差。TA 计算结果越小，说明监控视频移动目标识别效果越好。TP 计算结果越大，说明监控视频移动目标识别效果越好。

3.2 确定实验对象

为验证本文方法在视频序列中对运动目标识别的性能优势，采用高清摄像机收集交通监控视频信息，从而构建了实验所需的样本数据集。在此数据集中，选取的交通场景中的车辆作为多目标识别的对象，具体示例如图 1 所示。

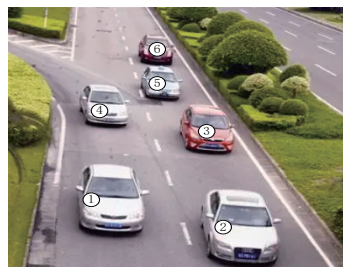


图 1 实验对象

根据图 1 所示，选定该目标后，进行了相应的实验验证与分析。

3.3 实验结果与分析

分别运用基于改进神经网络的视频序列运动目标识别方法（方法 1）、基于机器视觉的运动目标识别与定位（方法 2）和本文所提出的基于轻量化深度神经网络的监控视频中移动目标识别方法，对监控结果进行对比分析，分析结果如图 2 所示。



图 2 三种方法监控视频结果对比分析

根据图 2 所示，在方法 1 的识别结果中，最后一辆车被

错误地识别为背景；方法2的情况更为严重，最后两辆车均被判定为背景，这两种方法都未能成功识别出所有的移动车辆。相比之下，本文方法借助深度学习网络对背景特征进行学习，使系统能够更精准地区分背景与移动目标，从而成功对所有移动车辆实现了准确识别。

为了进一步验证基于轻量化深度神经网络的监控视频中移动目标识别方法的有效性，对3种方法的TA、TP计算结果进行对比，结果如图3所示。

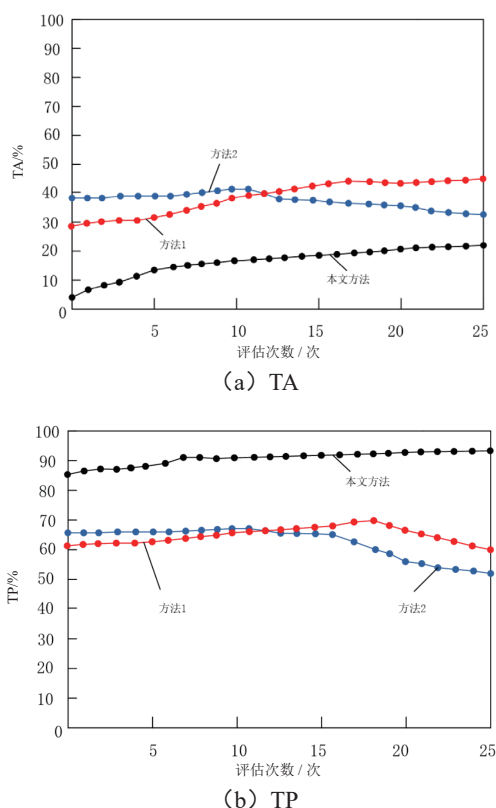


图3 三种方法TA、TP计算结果对比分析

从图3可知，方法1的TA最大值为48%，TP最大值为69%；方法2的TA最大值为42%，TP最大值为68%；本文方法的TA最大值为23%，TP最大值为94%。分析结果表明，虽然本文方法在TA指标上最大值低于前两种方法，但在TP指标上达到94%，远高于前两种方法。综合而言，基于轻量化深度神经网络的监控视频中移动目标识别方法整体效果更佳。本文方法的优势在于运用轻量化深度学习网络学习背景特征以识别背景和移动目标，这使得TA和TP指标较高，为后续处理提供准确的基础数据，进而实现对监控视频中移动目标的精准识别。

4 结语

本文提出的基于轻量化深度神经网络的监控视频中移动目标识别方法，依托词袋特征构建可视化特征库以完成初步

目标聚类，随后运用轻量化深度神经网络深度解析背景特征，精确区分背景与移动目标。通过差分图像的二值化处理及中值滤波降噪，该方法有效简化了图像信息并突出了移动目标。此外，通过设定敏感区域并依据质心位置启动识别，使得识别过程更加精准且高效，显著减少了误差累积。尽管该方法在实验中展现了良好的识别效果，但在实际应用中仍面临复杂场景下目标遮挡和光照变化等挑战，这些因素可能对识别性能产生一定影响。展望未来，将持续优化算法，增强对目标遮挡和光照变化的鲁棒性，并探索更为高效的轻量化神经网络架构，旨在进一步提升识别速度和精度，为监控视频中的移动目标识别领域提供更加坚实的技术支撑。

参考文献:

- [1] 范建伟, 李琳, 靳志鑫, 等. 基于改进神经网络的视频序列运动目标识别方法[J]. 现代电子技术, 2024, 47(20): 118-122.
- [2] 赵佳冉, 张爱华, 姜超群, 等. 基于机器视觉的运动目标识别与定位[J]. 智能计算机与应用, 2022, 12(11): 117-121.
- [3] 耿德新. 基于特征选取的监控运动目标识别模型研究[J]. 信息记录材料, 2023, 24(2): 22-24.
- [4] 周献珍, 张抒. 基于机器学习的动态环境运动目标自动识别研究[J]. 微型电脑应用, 2022, 38(9): 34-37.
- [5] 许鲲. 基于多目标识别的智能安防视频监控研究[J]. 自动化技术与应用, 2024, 43(7): 168-171.
- [6] 王健, 王建, 薛念明, 等. 基于超参数优化轻量化深度神经网络的安全围栏图像分割技术[J]. 现代信息科技, 2024, 8(16): 123-126.
- [7] 夏玉果, 董天天, 丁晟. 基于轻量化深度迁移神经网络的电子元器件识别[J]. 电子器件, 2023, 46(6): 1673-1679.
- [8] 刘洪宁, 李容露, 凌龙, 等. 基于深度学习的轻量化遥感图像目标检测模型设计研究[J]. 通信与信息技术, 2023(3): 19-22.

【作者简介】

高孟飞(1997—), 男, 河南郑州人, 硕士, 助教, 研究方向: 物联网技术、深度学习。

马浩(1989—), 男, 河南新郑人, 硕士, 助教, 研究方向: 计算机视觉、深度学习。

(收稿日期: 2024-12-12 修回日期: 2025-04-18)