



智能系统学报

CAA I TRANSACTIONS ON INTELLIGENT SYSTEMS

无人机视角下的多车辆跟踪算法研究

胡硕, 王洁, 孙妍, 周思恩, 姚美玉

引用本文:

胡硕,王洁,孙妍,周思恩,姚美玉. 无人机视角下的多车辆跟踪算法研究[J]. 智能系统学报, 2022, 17(4): 798–805.

HU Shuo,WANG Jie,SUN Yan,ZHOU Sien,YAO Meiyu. Research on multi-vehicle tracking algorithm from the perspective of UAV[J]. *CAA I Transactions on Intelligent Systems*, 2022, 17(4): 798–805.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202108014>

您可能感兴趣的其他文章

动态云台摄像机无人机检测与跟踪算法

Drone detection and tracking in dynamic pan-tilt-zoom cameras

智能系统学报. 2021, 16(5): 858–869 <https://dx.doi.org/10.11992/tis.202103032>

融合视觉显著性再检测的孪生网络无人机目标跟踪算法

Siamese network combined with visual saliency re-detection for UAV object tracking

智能系统学报. 2021, 16(3): 584–594 <https://dx.doi.org/10.11992/tis.202101035>

基于改进的Faster RCNN面部表情检测算法

Facial expression recognition based on improved Faster RCNN

智能系统学报. 2021, 16(2): 210–217 <https://dx.doi.org/10.11992/tis.201910020>

区域损失函数的孪生网络目标跟踪

Regional loss function based siamese network for object tracking

智能系统学报. 2020, 15(4): 722–731 <https://dx.doi.org/10.11992/tis.201910005>

高斯核函数卷积神经网络跟踪算法

Convolutional neural network tracking algorithm accelerated by Gaussian kernel function

智能系统学报. 2018, 13(3): 388–394 <https://dx.doi.org/10.11992/tis.201612040>



微信公众平台



期刊网址

DOI: 10.11992/tis.202108014

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20220530.2038.002.html>

无人机视角下的多车辆跟踪算法研究

胡硕, 王洁, 孙妍, 周思恩, 姚美玉

(燕山大学 电气工程学院, 河北 秦皇岛 066004)

摘要: 针对无人机视频中存在目标密集、运动噪声强而导致跟踪性能显著下降的问题, 提出了一种改进 YOLOv3 的车辆检测算法及一种基于深度度量学习的多车辆跟踪算法。针对车辆检测的精度与实时性问题, 采用深度可分离卷积网络 MobileNetv3 作为特征提取网络实现网络结构轻量化, 同时采用 CIoU Loss 作为边框损失函数对网络进行训练。为了在多目标跟踪过程中提取到更具判别力的深度特征, 提出了一种基于深度度量学习的多车辆跟踪算法, 实验证明, 本文提出的算法有效改善车辆 ID 跳变问题, 速度上满足无人机交通视频下车辆跟踪的实时性要求, 达到 17 f/s。

关键词: 车辆检测; 目标跟踪; 无人机视频; 特征提取; 轻量级网络; 深度特征; 损失函数; 深度度量学习

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2022)04-0798-08

中文引用格式: 胡硕, 王洁, 孙妍, 等. 无人机视角下的多车辆跟踪算法研究 [J]. 智能系统学报, 2022, 17(4): 798-805.

英文引用格式: HU Shuo, WANG Jie, SUN Yan, et al. Research on multi-vehicle tracking algorithm from the perspective of UAV[J]. CAAI transactions on intelligent systems, 2022, 17(4): 798-805.

Research on multi-vehicle tracking algorithm from the perspective of UAV

HU Shuo, WANG Jie, SUN Yan, ZHOU Sien, YAO Meiyu

(School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China)

Abstract: Aiming at the decline of tracking performance suffering from dense targets and strong motion noise in UAV video, we propose a vehicle detection algorithm based on improved YOLOv3 and a multi-vehicle tracking algorithm based on deep metric learning. To improve the vehicle detection system's accuracy and real-time performance, a deep separable convolution network, MobileNetv3, is adopted as the feature extraction network to realize a lightweight network structure, and the CIoU Loss is used as the frame loss function to train the network. A multi-vehicle tracking algorithm based on depth metric learning is proposed to extract more discriminative deep features during multi-target tracking. Experiments reveal that the algorithm proposed in this paper effectively improves the problem of target ID jump and meets the real-time requirement of vehicle tracking in UAV traffic video, reaching 17 FPS.

Keywords: vehicle detection; object tracking; UAV video; feature extraction; lightweight network; deep feature; loss function; deep metric learning

多目标跟踪在现实生活中有很多应用, 例如视频监控、人机交互、智慧交通系统等^[1-2], 已经成为了计算机视觉领域内不可或缺的研究课题。在智能交通系统中, 车辆检测技术和跟踪算法为路况监测和行驶安全行为等高级任务提供技术支持。无人机具有体积小、飞行灵活等优点^[3], 若将

其用于对道路交通进行辅助监控, 若将其应用于辅助监控道路交通中的安全行为, 可以有效地监控固定摄像机无法拍摄到的盲区范围, 从而对实时路况中的车辆进行全方位的检测和跟踪。

对于目标检测, 早期的算法注重检测精度但检测速度较慢, 典型的算法有 R-CNN^[4]、Fast R-CNN^[5]、Faster R-CNN^[6]、Mask R-CNN^[7] 等。为了提升检测速度, 一系列基于回归的单阶段检测算法相继被提出, 如 YOLO 系列^[8-10]、SSD^[11] 系列

收稿日期: 2021-08-14. 网络出版日期: 2022-05-31.

基金项目: 国家自然科学基金项目 (62073279).

通信作者: 胡硕. E-mail: hus@ysu.edu.cn.

等。此外, Tian 等^[12]通过使用一种双神经网络目标检测方法对目标区域的二次特征进行分类,快速筛选出一阶段检测中遗漏的目标,从而实现对小目标的高质量检测。Wu 等^[13]提出了一种基于光流和深度卷积网络的无人机快速人体检测系统。Tijtgat 等^[14]提出一种新的多分支并行特征金字塔网络来提升对无人机图像中小目标的检测性能。Liu 等^[15]提出一个集成网络 SyNet 将单阶段和多阶段检测算法相结合从而实现对无人机图像的检测。

基于深度学习的跟踪算法在速度和精度两方面得到双重提升,行人重识别 (person re-identification, ReID) 技术^[16-17]的引入在一定程度推动其发展,其模型均为基于检测的跟踪方法。Sort 算法^[18]利用卡尔曼滤波器预测目标状态和匈牙利算法计算相似度从而实现预测轨迹与检测目标的数据关联。DeepSort 算法^[19]结构是检测和匹配两个独立过程,采用两个不同网络提取特征,引入 ReID 特征与运动信息计算相似度矩阵,数据关联采用级联匹配和 IoU(intersection over union ratio) 匹配相结合方式。Bewley 等^[20]针对无人机长时跟踪中的遮挡问题提出了一种基于局部特征匹配和密度聚类的跟踪算法。Xue 等^[21]对 DCF 跟踪算法进行了改进,提出了一个新的无人机视频语义感知实时相关跟踪框架。Li 等^[22]研究了无人机场景下进行实时跟踪的限制因素,提出了一种利用背景线索和畸变响应抑制机制进行跟踪的方法。Yu 等^[23]提出了一种针对无人机视频的基于个体和全局运动融合的条件生成对抗网络 (generative adversarial networks, GAN) 跟踪算法。当前的研究更多的注意力在于提高算法准确率,但是随着深度网络结构复杂,导致了模型参数量和计算量增加,虽然提升了精度但是也极大的牺牲了速度。

在无人机视角下拍摄的监控画面进行目标跟踪存在以下难点: 1) 无人机视角下车辆整体偏小且目标密集,无人机飞行灵活,在拍摄视频中易出现目标模糊和快速移动等问题; 2) 同一类别车辆仅有车型和颜色之分,俯视角度下极易造成相似车辆干扰。

本文以无人机拍摄的视频作为研究对象,提出一种基于无人机交通视频的多车辆跟踪算法。针对实时性要求,提出了一种改进 YOLOv3 的车辆检测算法; 为提高跟踪系统的特征表达能力,提出一种基于度量学习的特征提取网络 Resi-8; 在该算法中,卡尔曼滤波器用于预测车辆的状态,相邻帧的相似性矩阵由运动信息相似度和外

观特征相似度相结合方式计算,通过匈牙利算法实现相邻帧之间的数据关联,进而实现多目标的精准稳定跟踪。

1 多车辆跟踪算法

本文算法包含 2 个部分: 检测模块和跟踪模块。在检测模块,针对无人机视频的车辆检测特点,本文提出了一种改进 YOLOv3 的目标检测网络 YOMOV3-CIoU; 在跟踪模块,本文提出一种基于度量学习的特征提取网络 Resi-8,以提高系统的特征表达能力。

1.1 基于改进 YOLOv3 的车辆检测

YOLOv3 算法作为一个基于回归的单阶段检测算法,在检测精度上有着优异的表现。为使其进一步适用于无人机拍摄视频中的车辆检测,提出以下几点改进策略。

1.1.1 先验框聚类算法

YOLOv3 算法在聚类先验框时使用的是 k-means 算法^[24],该算法是采用随机地选取 k 个聚类点作为初始聚类中心的策略,初始聚类中心的选取会对于聚类结果具有较大的影响。为了应对初始点敏感的问题,我们引入 k-means++ 算法^[25]完成先验框的聚类。

采用控制变量法,将 k-means 和 k-means++ 两种算法聚类出的 anchor box 分别用于同一检测算法 YOLOv3,在数据集上分别对模型进行训练,实验结果如表 1 所示。通过 k-means++ 算法得到的先验框 (anchor box) 在车辆检测算法中获得更高的检测结果,召回率、mAP(mean average precision) 和 F_1 分数均得到显著提升。

表 1 聚类算法的检测结果对比
Table 1 Comparison of detection results of different clustering algorithms %

聚类方法	精确度	召回率	mAP	F_1
k-means	50.1	49.5	52.7	51.3
k-means++	47.7	64.1	56.8	54.6

1.1.2 网络结构的改进

视频序列下的车辆所处的环境相对是固定的,通常位于停车场或道路上,检测类别单一,同时算法应尽可能满足实时性要求。为了达到上述目的,需要考虑模型的计算量和网络的参数量,尽可能的降低复杂度进而满足实时性需求,于是选择轻量级网络 MobileNetv3^[26]作为 YOLOv3 的特征提取网络,本网络结合通道注意力模块,逆残差模块和深度可分离卷积结构,具体结构细节

如图 1 所示。

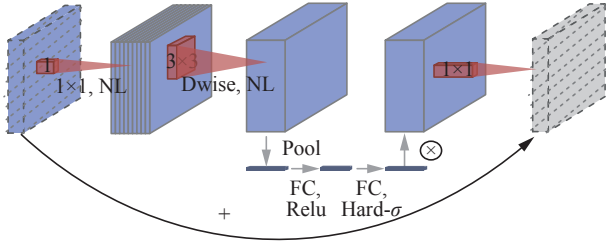


图 1 MobileNetV3 模块
Fig. 1 MobileNetV3 Module

1.1.3 损失函数优化

IoU 通常是衡量预测框与真实框的交并比, 其作为评价标准具有尺度不变性特点, 基于 IoU 的损失函数^[27]被定义为

$$L = 1 - \text{IOU} + R(b, b_{gt})$$

式中: $R(b, b_{gt})$ 是添加的惩罚项, b 是预测框中心点, b_{gt} 是目标框中心点。但是当两个框并没有重合时, 本损失函数无法确定预测框的位置信息, 并不能有效表达预测框和目标框的位置信息。

综上损失函数分析, 本文引入 CIoU Loss^[26] 作为本文算法的边框回归损失函数。CIoU 损失函数克服上述损失函数的缺点, 引入两个惩罚项如式 (1) 所示, 第 1 项中 c 表示包含两个框最小矩形区域的对角线距离, $\rho(\cdot)$ 表示两个中心点的欧式距离; 第 2 项是考虑到形状的差异, 即引入宽高比, 通过式 (2) 计算参数 v , α 是平衡参数, 由式 (3) 进行计算。

$$R_{\text{CIoU}} = \frac{\rho(b, b_{gt})}{c^2} + \alpha v \quad (1)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{\omega_{gt}}{h_{gt}} - \arctan \frac{\omega}{h} \right)^2 \quad (2)$$

$$\alpha = \frac{v}{(1 - \text{IOU}) + v} \quad (3)$$

CIoU Loss 可以定义为

$$L_{\text{CIoU}} = 1 - \text{IOU} + \frac{\rho^2(b, b_{gt})}{c^2} + \alpha v$$

参数 v 在反向传播时的梯度由式 (4) 计算, 为避免训练时发生梯度爆炸, 一般将 $1/(w^2 + h^2)$ 用 1 代替计算。

$$\frac{\partial v}{\partial \omega} = \frac{8}{\pi^2} \left(\arctan \frac{\omega_{gt}}{h_{gt}} - \arctan \frac{\omega}{h} \right) \times \frac{h}{\omega^2 + h^2} \quad (4)$$

1.2 基于深度度量学习的多车辆跟踪

现有的多目标跟踪算法, 无法有效地适用于无人机视角下的车辆检测, 交通视频中目标尺寸偏小且相似度高, 在车辆跟踪中精度无法得到保证。因此我们在 1.1 节提出的目标检测算法的基础上, 提出一种基于深度度量学习的多车辆跟踪算法。

1.2.1 运动特征相似度

马氏距离是计算两种未知样本之间相似度的方法, 它区别于欧氏距离更加考虑各种特性之间的联系, 有效的解决维度尺度相关且不一致的问题, 本文将其用来度量预测状态与当前帧的检测距离, 计算如式 (5) 所示。

$$d^{(1)}(i, j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i) \quad (5)$$

式中: d_j 表示第 j 个预测框对应的位置状态 (u, v, γ, h) , y_i 表示第 i 个跟踪轨迹的预测位置, S_i 是由卡尔曼滤波预测此刻位置的协方差矩阵。

在计算马氏距离时通过添加一个门控函数, 使用卡方分布 0.95 作为置信度, $t^{(1)} = 9.4877$, 可以有效剔除不合适的数据关联, 具体表达式如式 (6), 当马氏距离低于 $t^{(1)}$ 数值时, 表示匹配成功。

$$b_{i,j}^{(1)} = 1[d^{(1)}(i, j) \leq t^{(1)}] \quad (6)$$

式中: $[]$ 中是条件, 满足条件则保留该轨迹, 即匹配成功。

1.2.2 外观特征相似度

在路况复杂的环境下, 检测跟踪目标车辆只用运动特征相似度进行判断会导致跟踪算法的 ID 跳变严重, 我们通过使用运动特征和外观特征共同加权计算相似度矩阵的方式来避免因检测跟踪复杂环境造成问题, 即

$$d^{(2)}(i, j) = \min \{ 1 - r_j^T r_k^{(i)} | r_k^{(i)} \in R_i \} \quad (7)$$

式中: $r_j^T r_k^{(i)}$ 是余弦相似性, 余弦距离 $= 1 - r_j^T r_k^{(i)}$, 采用最小余弦距离度量相似性可以更准确地进行匹配判断是否检测跟踪的目标一致。如式 (8) 所示, 我们设置一个门控函数, 如果余弦距离小于 $t^{(2)}$, 则认为匹配成功。

$$b_{i,j}^{(2)} = 1[d^{(2)}(i, j) \leq t^{(2)}] \quad (8)$$

如式 (9) 所示, 将运动特征相似度和外观特征相似度作为综合相似性矩阵 $C = [c_{i,j}]$, 添加门控函数如式 (10) 所示, 只有当 $b_{i,j} = 1$ 时为匹配成功。

$$c_{i,j} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}(i, j) \quad (9)$$

$$b_{i,j} = \prod_{m=1}^2 b_{i,j}^{(m)} \quad (10)$$

1.2.3 基于度量学习的外观特征提取

度量学习^[28] (metric learning) 即相似度学习。两个样本数据之间的深度度量学习的欧氏距离可以表达为

$$D(f(x_i), f(x_j)) = \|f(x_i) - f(x_j)\|_2^2$$

深度度量学习中关键的一环是损失函数。本文将使用 Triplet Loss^[29] 损失函数进行网络训练, 将其称为 Triplet Network。Triplet Loss 具体表达如图 2 所示。该损失函数旨在通过训练拉近正样

本, 远离负样本, 计算相似度损失, 表示为式 (11)。

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 \quad (11)$$

式中: $f(\cdot)$ 表示卷积神经网络操作, α 为边界阈值。Triplet Loss 表达式如式 (12), 其中“ $[\cdot]_+$ ”表示当数值大于零的常数时, 优化锚样本与正样本距离, 远离与负样本距离, 否则不需要优化:

$$L_{\text{tri}} = \sum_i^N [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha]_+ \quad (12)$$

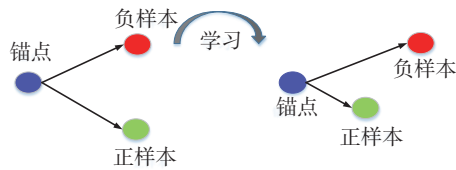


图2 Triplet Loss 示意图

Fig. 2 Triplet Loss Schematic diagram

1.3 多车辆跟踪算法整体框架

本文算法整体跟踪框图如图3所示, 具体流程如下:

- 1) 目标检测网络 YOMOV3-CIoU 对图像帧进行车辆检测, 获取车辆位置信息 d_j 。
- 2) 车辆状态是包括目标候选框坐标信息和相

应速度信息的8维向量。利用卡尔曼滤波预测前一帧中车辆轨迹的状态, 得到当前帧中车辆坐标信息 y_i 。

3) 通过公式 (5) 计算车辆运动特征相似度 C_m , 采用平方马氏距离来度量预测状态与当前帧的检测距离。

4) 对上述检测得到的车辆图像通过度量学习网络 Resi-8 提取外观特征 r_j ; 将距离最近 $L_k=100$ 个跟踪轨迹存放在外观描述集合 R_k 中。

5) 计算车辆外观特征相似度矩阵 C_n , 通过式 (7) 计算 r_j 和 $r_k^{(j)}$ 之间的最小余弦距离为 C_n 。

6) 计算综合相似度矩阵 C_{mn} , 利用式 (9) 得到 C_m 和 C_n 的线性加权结果作为 C_{mn} 。

7) 使用匈牙利算法实现前后帧之间的数据关联, 并分配相应的 ID。

8) 利用匈牙利算法匹配程度判断是否参数更新。对于运动轨迹和目标检测匹配一致, 则更新卡尔曼滤波器的参数; 若没有对应的轨迹, 则不更新参数但继续在 70 帧之内对轨迹进行匹配, 否则删除该轨迹; 若没有匹配的检测结果, 将其作为新目标分配新轨迹。

9) 重复上述过程, 完成整个序列的检测跟踪。

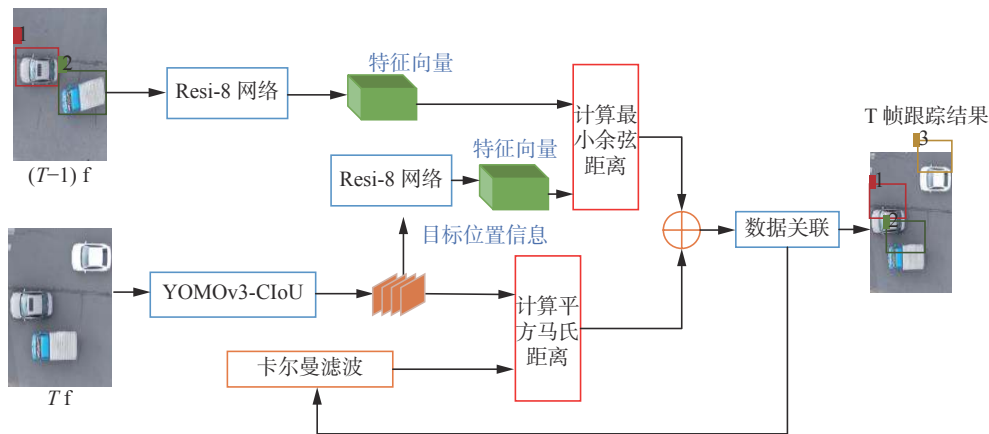


图3 本文算法整体流程

Fig. 3 Flow of the algorithm

本文引入运动特征和外观特征共同决定相似性矩阵, 有效的解决在跟踪遮挡等复杂环境下造成的 ID 跳变等问题, 算法鲁棒性更强。

2 实验结果分析

为了验证本文算法性能, 本文进行了实验并对实验结果进行了分析。使用 Python 作为编程语言, 深度学习框架基于 Pytorch0.4.0, 实验环境如表2:

表2 实验平台参数
Table 2 Experimental platform parameters

运行系统	处理器	显卡	CUDA
Linux	Intel i7-7700K	GTX1080	Cuda 10.0

2.1 车辆检测算法对比实验

将 1.1 中仅修改特征提取网络后的模型记为 YOMOV3, 将在 YOMOV3 基础上修改损失函数为 Ciou Loss 的检测模型记为 YOMOV3-CIoU。

本节车辆检测实验所使用的数据集为 Vis-

Drone2019^[31]。将其中 4 种类别车辆的数据集称为 Vis-4。

2.1.1 模型参数量对比

表 3 中对 YOLOv3-tiny、YOLOv3 和 YOMOV3, 3 个模型进行参数量统计, Mobilenetv3 网络由于引入深度可分离卷积在一定程度上减少参数量, 仅为 Darknet-53 的 2/5, 有效地减小了网络的计算复杂度, 为检测速度的实时性提供了可能。

表 3 不同模型参数量对比

Table 3 Comparison of the number of different model parameters

模型	骨干网络	参数
YOLOv3-tiny	6.00M	8.45M
YOLOv3	38.74M	59.13M
YOMOV3	2.84M	23.09M

2.1.2 阈值选择对比实验

在网络检测层输出结果中会存在多个预测框, 为筛选最终合适的目标框, 我们通过引入非极大值抑制^[32](non-maximum suppression, NMS)方法进行筛选。本节通过选择不同的 NMS 阈值来

判断对模型结果的影响, 在实验中通常将 mAP 作为阈值选择标准, 实验对比结果如表 4 所示。

表 4 阈值选择对比结果

Table 4 Threshold selection comparison results

阈值	0.3	0.4	0.5	0.6	0.7	0.9
mAP/%	52.2	52.9	53.5	53.0	52.3	43.3

在实验结果中, 当阈值为 0.5 时, mAP 达到了 53.5%, 在对比实验中结果最高, 因此将阈值设为 0.5。

2.1.3 优化后的检测算法实验结果对比

表 5 为进行不同模型的检测结果, YOLOv3 算法是 4 种算法里检测精度最高的一个, 但 YOLOv3 算法参数量多且复杂程度高, 速度仅为 21 f/s, 对于实时性要求有待提高; YOLOv3-tiny 的 mAP 只有 29.6%, 但其速度达到了 188 f/s, 属于典型的牺牲检测精度换取检测速度的模型, 对于精度要求高的情况无法满足。本文所提出的 YOMOV3 算法通过轻量化特征提取网络, 在保证检测精度的情况下检测速度达到 31 f/s; 经过损失函数优化后, mAP 提升 0.9%。

表 5 不同模型检测结果对比

Table 5 Comparison of testing results of different models

模型	骨干网络	损失函数	精确度/%	Recall/%	mAP/%	F_1 score/%	速度/(f·s ⁻¹)
YOLOv3-tiny	Tiny	GIoU	35.8	37.5	29.6	36.3	188
YOLOv3	Darknet-53	GIoU	47.7	64.1	56.8	54.6	21
YOMOV3	Mobilenetv3	GIoU	47.6	60.9	53.5	53.3	31
YOMOV3-CIoU	Mobilenetv3	CIoU	41.8	65.1	54.4	53.8	31

表 6 是检测算法 YOMOV3 以及损失函数优化后算法 YOMOV3-CIoU 与各项检测算法的实验对比结果。本文所提的两个算法作为一阶段检测算法, 在精度相当的情况下, 速度上提高了近两倍。

表 6 YOMOV3 与其他检测算法性能对比

Table 6 Performance comparison between YOMOV3 and other detection algorithms

模型	骨干网络	mAP/%	速度/(f·s ⁻¹)
Faster R-CNN	ResNet-50	63.1	10
	ResNet-101	63.0	10
RetinaNet	ResNet-50	56.1	12
	ResNet-101	56.1	12
ATSS	ResNet-50	62.7	11
	ResNet-101	61.9	11
YOMOV3	MobileNetv3	53.5	31
YOMOV3-CIoU	MobileNetv3	54.4	31

2.1.4 实际场景下的检测结果

我们在 Vis-4 数据集中选取几个较为典型的图像序列进行检测结果对比, 可以有效直观的分析跟踪效果。图 4 是本文算法在夜间情况下实验结果。从图 4 可知, YOLOv3-tiny 中出现漏检和误检的情况, YOMOV3-CIoU 成功将其检测出且置信度较高。图 5 是模糊情况下对车辆检测的效果, 以下检测算法均在一定程度上对车辆出现漏检误检的情况, YOLOv3-tiny 和 YOLOv3 算法均出现漏检和误检情况, 且定位出现漂移现象, 抗遮挡能力差; 本文所提出的算法 YOMOV3 和 YOMOV3-CIoU 没有出现误检情况且定位精准。



(a) YOLOv3-tiny



(b) YOLOv3



图4 夜间情况检测结果

Fig. 4 Detection results of nocturnal environment

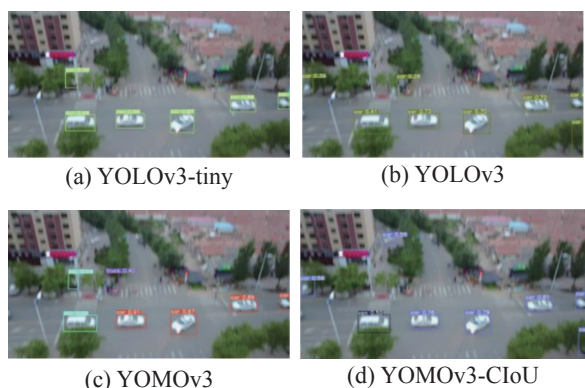


图5 模糊情况下的检测结果

Fig. 5 Detection results in the case of fuzziness

2.2 多车辆跟踪算法实验

为有效证明提出的跟踪算法进行评估,在数据集中选择6个具有代表性的视频序列进行分析。从表7中可以看出对于不同的视频序列,本文所提车辆跟踪算法的MOTP (multiple object tracking precision) 的值均在54%左右,MOTA (multiple object tracking accuracy) 均值为62.0%,在所有视频序列中最高达到了71.8%,最低也保持在58.7%。MOTA是衡量确定目标数目、相关属性准确度以及定位误差累计的指标。MOTP主要体现在确定目标定位上的精确度,比如预测框的定位。这两个数值越大则跟踪算法越好。二者共同衡量算法连续跟踪目标的能力。实验数据表明本文算法的有效性。

表7 多车辆跟踪算法实验结果

Table 7 Experimental results of multi-vehicle tracking algorithm

视频	ID	MOTA/%	MOTP/%
uav_3240	39	64.4	51.8
uav_5448	4	71.8	56.2
uav_0001	191	59.6	54.4
uav_1380	8	58.7	52.7
uav_1288	10	59.8	51.3
uav_0920	155	64.2	56.8

在6个视频序列中将本文算法与Sort和DeepSort算法进行性能评测,对比结果如表8所示。在Sort算法的对比下,本文算法的ID大幅度降

低,有效证明算法中引入深度度量学习的特征实现相似性度量具有更好的判别效果。与DeepSort算法相比,实时性得到更好的满足,提升到17 f/s。综上,本文多车辆跟踪算法有效改善目标ID跳变问题,速度上满足实时性。

表8 不同跟踪算法实验对比结果

Table 8 Experimental comparison of different tracking algorithms

算法	ID	MOTA/%	MOTP/%	速度/(f·s ⁻¹)
Sort	435	61.1	53.6	19
Deep Sort	346	62.3	56.4	10
Ours	363	62.0	53.8	17

我们在视频序列测试了不同影响因素下的车辆跟踪。图6中主要是针对车辆之间的局部或全部遮挡情况,且目标外观相似,对车辆跟踪有直接的干扰;图7中出现了外部遮挡,路灯对车辆进行了遮挡。本算法中使用基于度量学习提取车辆外观特征和运动特征共同计算相似度矩阵,使用级联匹配策略将目标轨迹进行关联,保证目标ID不发生跳变和车辆跟踪的稳定性;图8中是车辆发生不同尺度下测试效果,本算法对网络的3个检测层输出特征图进行融合以及引入k-means++完成先验框的聚类,可以很好地应对尺度变化问题,算法依然保持稳定的跟踪和相同的ID。



图6 车辆间遮挡

Fig. 6 Occlusion of a vehicle

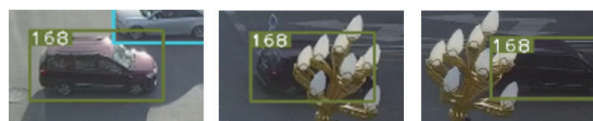


图7 外部遮挡

Fig. 7 External occlusion



图8 尺度变化

Fig. 8 Scale change

3 结束语

为有效适用于无人机视角下车辆视频的检测与跟踪,本文提出了一种改进YOLOv3的车辆检测算法,通过k-means++算法聚类出合适的候选

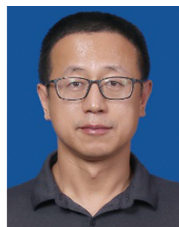
框,采用 MobileNetV3 网络作为算法中的特征提取网络;选择 CIoU Loss 作为预测框的损失函数,有效得到预测框与真实框之间的位置信息;针对车辆跟踪过程中 ID 易跳变的问题,和进一步提升对车辆目标的特征表达,本文提出了一种基于深度度量学习的多车辆跟踪算法,将运动信息和外观信息相结合确定相似性矩阵,通过匈牙利算法实现数据关联。实验结果表明,本文所提车辆跟踪算法改善了 ID 跳变问题,在保证跟踪精度的前提下提升了跟踪速度,达到 17 f/s。

参考文献:

- [1] 肖克来提. 无人机目标跟踪综述 [J]. 信息通信, 2020, 33(7): 130–131.
XIAO K. Summary of UAV target tracking[J]. Information & communications, 2020, 33(7): 130–131.
- [2] 孟磊, 杨旭. 目标跟踪算法综述 [J]. 自动化学报, 2019, 45(7): 1244–1260.
MENG Lu, YANG Xu. A survey of object tracking algorithms[J]. Acta automatica sinica, 2019, 45(7): 1244–1260.
- [3] NOORWALI A, AWAIS JAVED M, ZUBAIR KHAN M. Efficient UAV communications: recent trends and challenges[J]. Computers, materials & continua, 2021, 67(1): 463–476.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580–587.
- [5] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440–1448.
- [6] REN Shaoqing, HE Kaiming, 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.
- [7] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980–2988.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779–788.
- [9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6517–6525.
- [10] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. New York: arXiv, 2018. (2018–04–08)[2021–08–14].<https://arxiv.org/abs/1804.02767>.
- [11] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21–37.
- [12] TIAN Gangyi, LIU Jianran, YANG Wenyuan. A dual neural network for object detection in UAV images[J]. Neurocomputing, 2021, 443: 292–301.
- [13] WU Qingtian, ZHOU Yimin, WU Xinyu, et al. Real-time running detection system for UAV imagery based on optical flow and deep convolutional networks[J]. IET intelligent transport systems, 2020, 14(5): 278–287.
- [14] TIJTGAT N, VAN RANST W, VOLCKAERT B, et al. Embedded real-time object detection for a UAV warning system[C]//2017 IEEE International Conference on Computer Vision Workshops. Venice: IEEE, 2017: 2110–2118.
- [15] LIU Yingjie, YANG Fengbao, HU Peng. Small-object detection in UAV-captured images via multi-branch parallel feature pyramid networks[J]. IEEE access, 8: 145740–145750.
- [16] 尹梓睿, 张索非, 张磊, 等. 适于行人重识别的二分支 EfficientNet 网络设计 [J]. 信号处理, 2020, 36(9): 1481–1488.
YIN Zirui, ZHANG Suofei, ZHANG Lei, et al. Design of a two-branch efficientnet for person re-identification[J]. Journal of signal processing, 2020, 36(9): 1481–1488.
- [17] 冯霞, 杜佳浩, 段仪浓, 等. 基于深度学习的行人重识别研究综述 [J]. 计算机应用研究, 2020, 37(11): 3220–3226, 3240.
FENG Xia, DU Jiahao, DUAN Yinong, et al. Research on person re-identification based on deep learning[J]. Application research of computers, 2020, 37(11): 3220–3226, 3240.
- [18] ALBABA B M, OZER S. SyNet: an ensemble network for object detection in UAV images[C]//2020 25th International Conference on Pattern Recognition. Milan: IEEE, 2021: 10227–10234.
- [19] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on Image Processing. Beijing: IEEE, 2017: 3645–3649.
- [20] BEWLEY A, GE Zongyuan, OTT L, et al. Simple online and realtime tracking[C]//2016 IEEE International Conference on Image Processing. Phoenix: IEEE, 2016: 3464–3468.

- [21] XUE Xizhe, LI Ying, YIN Xiaoyue, et al. Semantic-aware real-time correlation tracking framework for UAV videos[J]. *IEEE transactions on cybernetics*, 2022, 52(4): 2418–2429.
- [22] LI Tian, DING Feifei, YANG Wenyuan. UAV object tracking by background cues and aberrances response suppression mechanism[J]. *Neural computing and applications*, 2021, 33(8): 3347–3361.
- [23] YU Hongyang, LI Guorong, SU Li, et al. Conditional GAN based individual and global motion fusion for multiple object tracking in UAV videos[J]. *Pattern recognition letters*, 2020, 131: 219–226.
- [24] 周爱武, 于亚飞. K-Means 聚类算法的研究 [J]. *计算机技术与发展*, 2011, 21(2): 62–65.
ZHOU Aiwu, YU Yafei. The research about clustering algorithm of K-means[J]. *Computer technology and development*, 2011, 21(2): 62–65.
- [25] ARTHUR D, VASSILVITSKII S. K-Means++: The advantages of careful seeding[C]//Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics. Philadelphia: ACM, 2007, 11(6): 1027–1035.
- [26] HOWARD A, SANDLER M, CHEN Bo, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 1314–1324.
- [27] ZHENG Zhaohui, WANG Ping, LIU Wei, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12993–13000.
- [28] XIE Pengtao, XING Eric P. Multi-modal distance metric learning[C]//Proceedings of the 23rd International Joint Conference on Artificial Intelligence. Beijing: IJCAI/AAAI, 2013: 1806–1812.
- [29] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 815–823.
- [30] KALMAN R E. A new approach to linear filtering and prediction problems[J]. *Journal of basic engineering*, 1960, 82(1): 35–45.
- [31] ZHU Pengfei, WEN Longyin, DU Dawei, et al. Vis-Drone-DET2018: the vision meets drone object detection in image challenge results[C]//European Conference on Computer Vision. Cham: Springer, 2019: 437–468.
- [32] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression[C]//ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition - Volume 03. New York: ACM, 2006: 850–855.

作者简介:



胡硕, 讲师, 主要研究方向为模式识别、机器学习、机器视觉。



王洁, 硕士研究生, 主要研究方向为机器学习、目标跟踪。



孙妍, 硕士研究生, 主要研究方向为机器视觉、目标跟踪。