

基于YOLOv8的航拍车辆检测技术研究

冯 晶, 符丹丹, 刘 强, 朱克佳

广州软件学院电子系, 广东 广州

收稿日期: 2023年11月26日; 录用日期: 2023年12月21日; 发布日期: 2023年12月29日

摘 要

随着经济和科技的发展, 智慧交通逐渐被人们关注。作为智慧交通的关键技术, 航拍车辆检测技术也逐渐走入人们的视野。针对目前航拍车辆目标较小, 道路环境复杂造成的漏检率高的情况, 本文提出了一种基于YOLOv8的航拍车辆检测方法。在YOLOv8的骨干网络中增加注意力机制模块, 以增加网络间特征的联系, 同时将头部网络的部分C2f层修改为C3层。经实验表明, 改进后的网络在UCAS-AOD数据集上的准确率为93.4%, 达到了不错的效果。

关键词

深度学习, 车辆检测, YOLOv8

Research on Aerial Vehicle Detection Technology Based on YOLOv8

Jing Feng, Dandan Fu, Qiang Liu, Kejia Zhu

Department of Electronics, Software Engineering Institute of Guangzhou, Guangzhou Guangdong

Received: Nov. 26th, 2023; accepted: Dec. 21st, 2023; published: Dec. 29th, 2023

Abstract

With the development of economy and science and technology, intelligent transportation is gradually being paid attention to. As the key technology of intelligent transportation, aerial vehicle detection technology has also gradually come into people's view. Aiming at the current situation that aerial vehicle targets are small and the road environment is complex resulting in a high rate of missed detection, this paper proposes an aerial vehicle detection method based on YOLOv8. The attention mechanism module is added to the backbone network of YOLOv8 to increase the linkage of features between networks, and part of the C2f layer of the head network is modified to C3 layer. It is experimentally shown that the improved network achieves a good accuracy of 93.4% on

the UCAS-AOD dataset.

Keywords

Deep Learning, Vehicle Detection, YOLOv8

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着国民经济以及道路交通的飞速发展,全国人民汽车保有量激增[1]。随着汽车的普及,巨大的汽车数量以及复杂的道路环境导致道路交通面临着极大的挑战。近几年来,道路拥堵以及交通事故的出现频率大大增加。为了提高人们的出行效率,减少道路安全事故的发生,对于智慧交通的需求迫在眉睫。在智慧交通中使用的检测器多为地面检测器,固定位置进行安装,灵活性不高,并且对乡村道路的覆盖率较低[2]。在智慧交通的构建上,无人机航拍车辆检测拥有更灵活的应变性以及更广阔的视角,可以便捷、快速的获知目标区域的道路交通信息,是构建智慧交通的主要研究技术之一[3]。由于无人机高度较高,在航拍视角下,虽然视野较大,但车辆目标尺度较小,使得检测难度大大增加。并且城市道路环境复杂,部分道路车流较多,目标间的遮挡率也较高,为车辆检测的准确率提出了更高的挑战。

目前常用的航拍车辆检测算法,主要有传统的基于运动特征的方法、基于特征提取的方法以及深度学习的方法[4]。传统的方法通过帧间差距或者模板匹配等方式进行识别,检测准确率较低。基于特征提取的方法需手动提取特征,对特征的要求较高,特征提取的好坏直接影响到识别的效果。依赖于人工以及专家经验。近年来,深度学习技术在目标检测领域有较好的表现,不少专家学者将其运用到航拍车辆检测中达到了不错的效果[5]。基于深度学习的方法,采用卷积神经网络(CNN)自动进行特征提取,不依赖于专家经验手动提取特征,大大提高了网络的检测准确率于泛化性能。目前常用的深度学习网络大致可以分为两种,一种是基于区域提名的方法,其代表网络为 R-CNN 和 Faste R-CNN 等。该类方法大致可以分为两个步骤,首先提取特征,然后再通过分类器进行分类。这种方法采用非深度学习的方法提取特征,计算量大,检测速度慢。另一种是端到端的检测方法,将分类和回归一步完成,大大提高了检测时间,代表网络为 SSD, YOLO 等。在此基础上,结合航拍车辆检测的特点,有学者对其做了进一步研究。

Heqian Qiu 等人基于 R-CNN 进行改进,提出了一种新颖的端到端自适应高宽比多尺度网络(A2RMNet),解决了由于物体的大小和长宽比不相同,导致的遗漏或错误定位物体的问题[6]。该网络由门融合模块、细化块和区域建议网络组成。设计了一个多尺度特征门融合网络,同时利用高宽比关注网络保持物体的高宽比,从而减轻了训练过程中因高宽比变化而导致的物体形状过度失真的问题。经实验表明,该网络可以自适应地整合物体的多尺度特征,达到了不错的检测效果。但特征提取网络模型较大,增加了计算量,运行速度较慢。李旭等人以 SSD 为基础,针对航拍车辆目标尺度小、特征点少的问题,提出了 DT-Net 网络。该网络主要分为车辆密度估计网络,掩膜生成模块,局部检测器,全局检测器以及 Soft-NMD 融合算法构成,在多数数据集上达到了不错的效果[7]。

以上方法都是从目标样本本身出发,解决了样本尺度问题,虽然检测准确率高,但检测时间相对较长,网络收敛速度较慢。2023年,单铭琦等针对网络参数冗余问题,提出了一种基于 YOLOv4 的改进算法[8]。该算法通过分析通道重要性,对重要通道进行利用,大大减少了网络参数量,同时采用知识蒸馏

算法, 弥补模型精度, 大大提高了网络运行速度, 但准确率较文献 6、7 有所下降。

本文针对航拍车辆目标检测, 城市道路环境复杂造成的错检, 以及车辆目标尺度较小、密度大、遮挡率高的问题, 本文在更轻量、快速的端到端的检测方法上进行研究, 提出了一种改进的 YOLOv8 网络模型。在 YOLOv8 网络的基础上, 对其进行改进, 在主干网络中加入注意力机制, 协助模型捕捉重要信息, 同时将头部网络的最后一个 C2f 结构修改为 C3 结构。采用 AdamW 优化器对网络进行优化, 提高网络收敛速度。最终在 UCAS-AOD 数据集上训练测试, 本文方法检测精度较高, 速度较快, 达到了不错的效果。

2. YOLOv8 网络

2.1. 网络结构

YOLOv8 网络是由 Ultralytics 公司在 2023 提出的, 其主要由头部网络和骨干网络构成[9]。其网络结构图如图 1 所示。

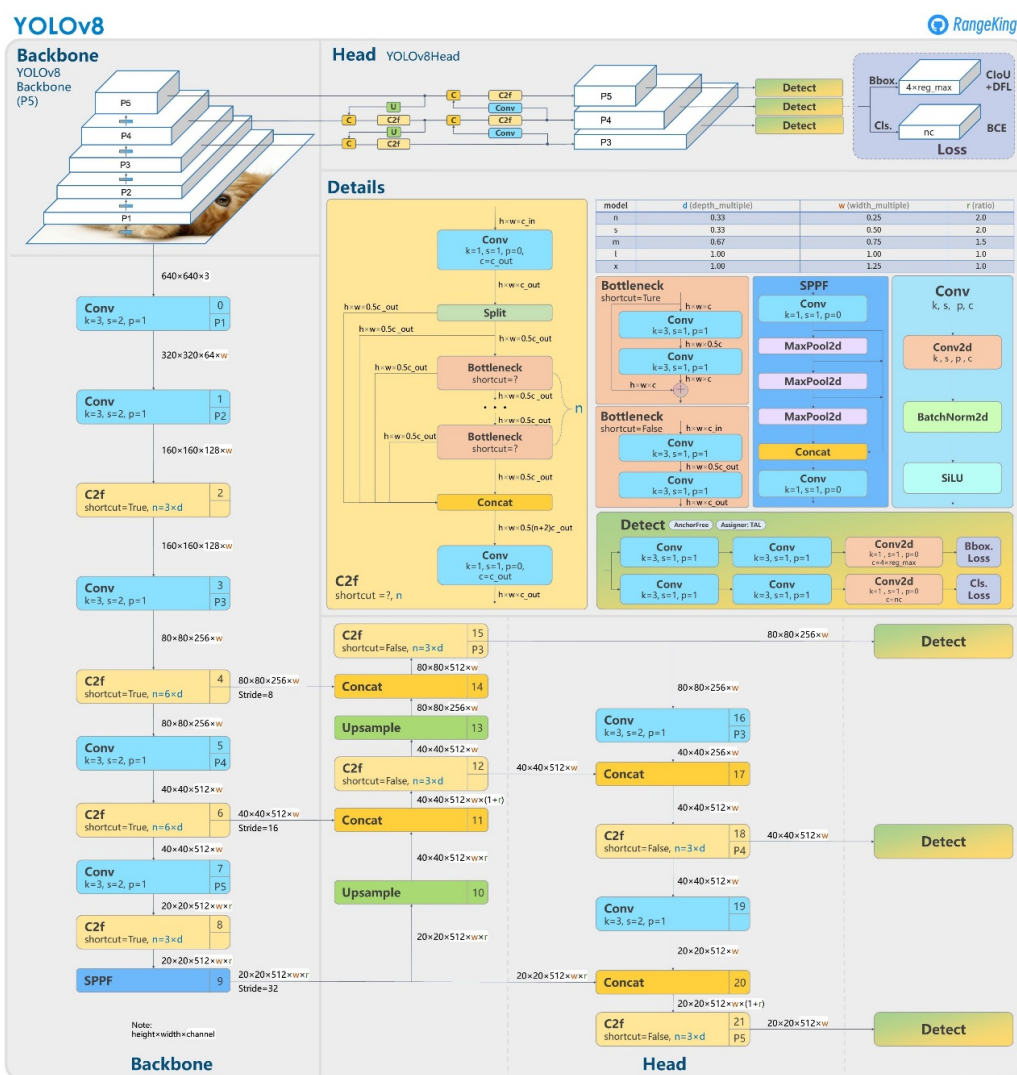


Figure 1. YOLOv8 network diagram

图 1. YOLOv8 网络图

其中 YOLOv8 的骨干网络主要通过不同通道的卷积层进行叠加, 采用 C2f 结构来提取图像特征; 颈部网络主要将骨干网络提取出来的特征进一步处理、融合后向头部网络输出; 头部网络主要做预测和边界框的微调, 将分类和检测相分离, 达到了更好的检测效果。同时采用 Anchor-Free, 对目标有更好的适应性。

在 YOLOv8 中采用 C2f 结构, 该结构将不同层的特征图跨层进行融合, 丰富模型的梯度流; 并且添加了一些 Split 操作, 采用深度可分离卷积, 减少参数量使网络更加轻量化, 运行速度更快。图 2 为 C2f 结构图。

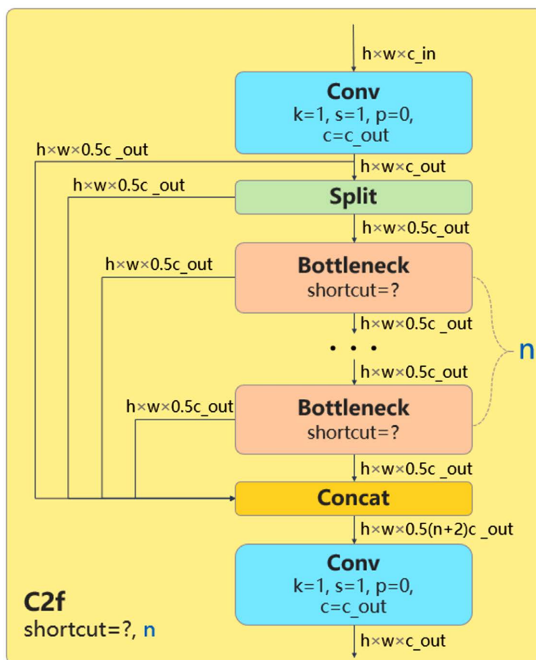


Figure 2. C2f structure

图 2. C2f 结构图

2.2. 损失函数

YOLOv8 的损失函数采用了 Task Aligned Assigner 正样本分配策略, 定义为分类损失(BCE Loss), 和回归损失(CIOU Loss + DFL)。

其中 BCE 损失定义为:

$$L = 1/N \sum_i -[y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (1)$$

CIOU Loss 综合考虑了宽高比和面积可以使结果框更加接近真实框, 为了提高泛化性, 加快训练速度, 还增加了 DFL 损失, 定义如下:

$$\text{DFL}(S_i, S_{i+1}) = -((y_{i+1} - y) \log(S_i) + (y - y_i) \log(S_{i+1})) \quad (2)$$

3. 基于 YOLOv8 的航拍车辆检测

3.1. 注意力机制

注意力机制可以帮助模型更好的关注数据的关键信息, 提高模型的性能。其基本思想为, 为输入数

据的每个部分分配一个权重，表示对当前任务的重要程度[10]。CBAM 为混合注意力机制，其作为一种空间与通道相结合的注意力机制模块，通过减少通道数进而使用大尺寸卷积来利用位置信息。CBAM 注意力模型实现混合注意力机制具体的实现过程如下：

首先是通道注意力模块，通过计算每个通道的平均值和最大值，并将这些值作为输入传递到两个全连接层中，根据通道重要性自适应分配权重向量，对有用或者需要的特征通道进行选择，减少对无用或者不重要的通道的依赖；其次，经过处理的数据信息在通过空间注意力模块时模型对特征图像的不同位置进行卷积操作，自适应地挑选出最重要的空间位置，并对有用或者重要的区域的特征表示进行加强；随后，将通道和空间注意力的结果相乘，最终取得加权后的特征。CBAM 混合注意力模型与单独使用通道或空间注意力机制的模型相比较，模型的表现大部分场景更胜一筹，在处理捕捉输入图像的特征方面更具优势。

通过添加全局注意力机制，可以保留空间和通道之间的关联信息，更好的区分目标，提高模型性能。本文在骨干网络中添加一层 CBAM 层，提高数据关联信息。改进后的主干网络如图 3 所示。

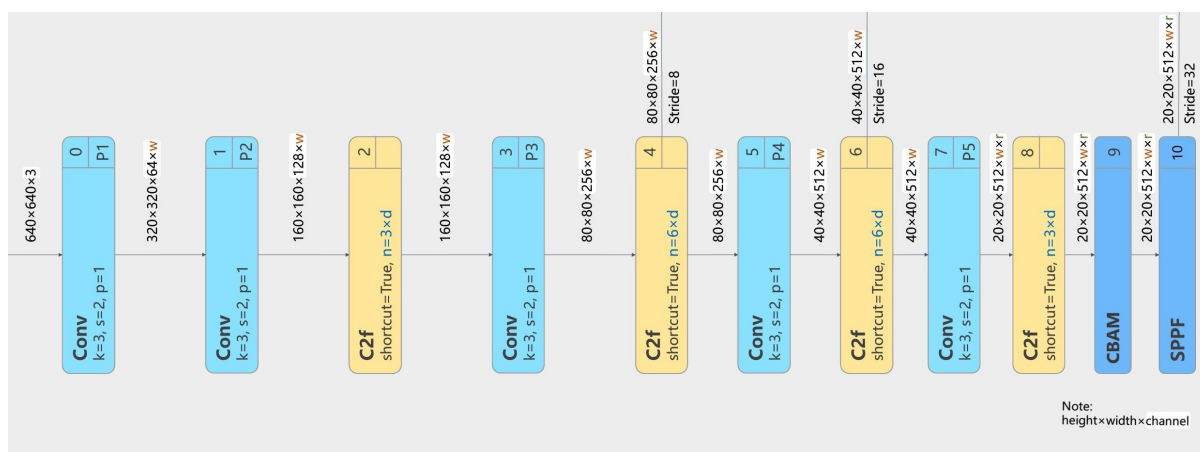


Figure 3. Backbone network

图 3. 骨干网络

3.2. Adam Weight Decay Optimizer 优化器

2014 年 12 月，Kingma 和 Lei Ba 两位学者提出了 Adam (Adaptive Moment Estimation) 优化器，结合 AdaGrad 和 RMSProp 两种优化算法的优点，另外，经偏置校正后，每一次迭代学习率都有个确定范围，使得参数比较平稳。对梯度的一阶矩估计(First Moment Estimation，即梯度的均值)和二阶矩估计(Second Moment Estimation，即梯度的未中心化的方差)进行综合考虑，动态调整针对于每个参数的学习速率[11]。为了更快地收敛以及防止发生局部最优的情况，AdamW 在 Adam 的基础上添加权重衰减，用于减少权重幅度，减少对偏置参数的影响，提高梯度下降速度。本文采用 AdamW 优化算法进行优化。

首先用式(3)来设定梯度。

$$g_t = \nabla_{\theta} J(\theta_{t-1}) \quad (3)$$

其中 g_t 为小批量随机梯度，表示第 t 时刻参数 θ_{t-1} 的梯度。

采用 AdamW 优化器更新规则如下：

(1) 用式(4)来计算一阶矩估计即梯度的指数移动平均数 m_t ， m_0 初始化为 0。其中 β_1 为指数衰减率，默认为 0.9。

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (4)$$

(2) 用式(5)计算二阶矩估计即梯度平方的指数移动平均数 v_t ， v_0 初始化为 0。其中 β_2 为指数衰减率，对梯度平方进行加权均值，默认为 0.999。

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (5)$$

(3) 用式(6)来对梯度平均值 m_t 进行偏差纠正，降低 m_0 初始化为 0 对训练的影响，避免 m_t 向 0 偏置。

$$\hat{m}_t = m_t / (1 - \beta_1^t) \quad (6)$$

(4) 原理同上，用式(7)对梯度平方的平均值 v_t 进行偏差纠正，实现无偏估计。

$$\hat{v}_t = v_t / (1 - \beta_2^t) \quad (7)$$

(5) 设定当前的梯度为一阶矩除以二阶距的平方根。对于波动大的梯度可以减少它在纵轴上的步长，增加横轴的步长。而 m_t 和 v_t 包含历史梯度信息，能较好的处理梯度稀疏的情况。所以由式(8)来更新参数 θ_t ，达到对参数 θ_t 的优化，其中 λ 为权重衰减参数，设为 0.01。

$$\theta_t = \theta_{t-1} - \alpha * \hat{m}_t / (\sqrt{\hat{v}_t} + \varepsilon) + \lambda \theta_{t-1} \quad (8)$$

其中为了避免除数为 0，超参数默认值为 $\alpha = 0.001$ ， $\varepsilon = 10^{-8}$ 。

4. 检测结果与分析

4.1. 数据集

本文采用由中国科学院大学模式识别与智能系统开发实验室制作的 UCAS-AOD 数据集，该数据集使用 Google Earth 软件在全球部分区域中截取的图像，包含航拍视角下的汽车和飞机。本文采用其中的汽车数据集，共计 510 幅图像，其中有 7114 辆汽车样本。然后将其转化为 YOLO 的数据集格式，最后采用 7:2:1 的比例将其划分为训练集、测试集和验证集。最终图像数见表 1。

Table 1. Number of images in the dataset

表 1. 数据集图像数

数据集	训练集	测试集	验证集	总计
图像数(幅)	367	102	41	510

4.2. 实验结果与分析

本文采用 AdamW 优化器，所有图片分批次投入模型训练，提升了梯度下降时更新所有参数的速度，尽可能获得参数解的最优值。本文设定每次输入模型训练的图片数量为 8，设置 epoch 为 85，最终训练集准确率为 94.15%，测试集准确率为 93.4%。

图 4 是训练过程曲线图，横坐标为 epoch，纵坐标为其相关数值，第一行从左到右依次为：训练集边框损失曲线、训练集分类损失曲线、训练集 DFL 损失曲线、精确率曲线、召回率曲线。第二行为验证集相关损失曲线以及 IOU 阈值 > 50 和 IOU 阈值在 0.5 到 0.95 的平均准确率曲线。从图中可以看出不论是训练集和验证集的准确率还是损失曲线是有波动的，并不是平滑的，这是由于本文采用的是 AdamW 下降法，每次训练的只是所有数据集中的一批图像，对这一批图像进行计算得到的损失和准确率，从上图可以看到曲线虽然有所起伏，但是准确率的整体趋势是上升的，而损失是下降的，通过不断迭代计算，在 epoch 为 0~50 时模型处于学习阶段，曲线变化很明显，在 epoch 为 50 之后 loss 曲线下降较平缓，逐

渐向 0 靠近；mAP 曲线也逐渐趋于稳定，模型是在逐渐收敛，训练速度较快。

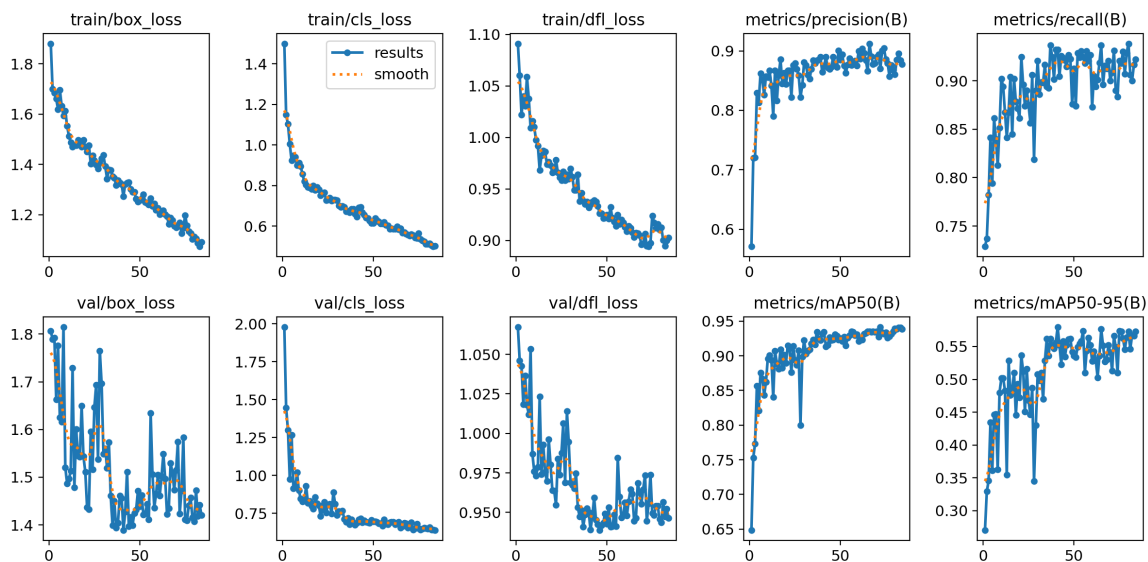


Figure 4. Training process graph

图 4. 训练过程曲线图

图 5 为测试集 P-R 曲线图，表示了精确率 P 和召回率 R 的关系。横坐标为召回率，纵坐标为精确率，P-R 曲线下围成的面积表示准确率 AP。从图中可以看出最终汽车类别即本文所有分类类别的准确率为 93.4%。

图 6 为测试集随机选取的 16 幅图像检测的结果，图中红色框为模型检测为 car 类别的预测结果，右上角为该预测框对应的置信度的值。从图中可看出本文网络对图像中车辆基本都可以识别成功，对相互遮挡以及环境复杂的情况也有较好的识别效果。

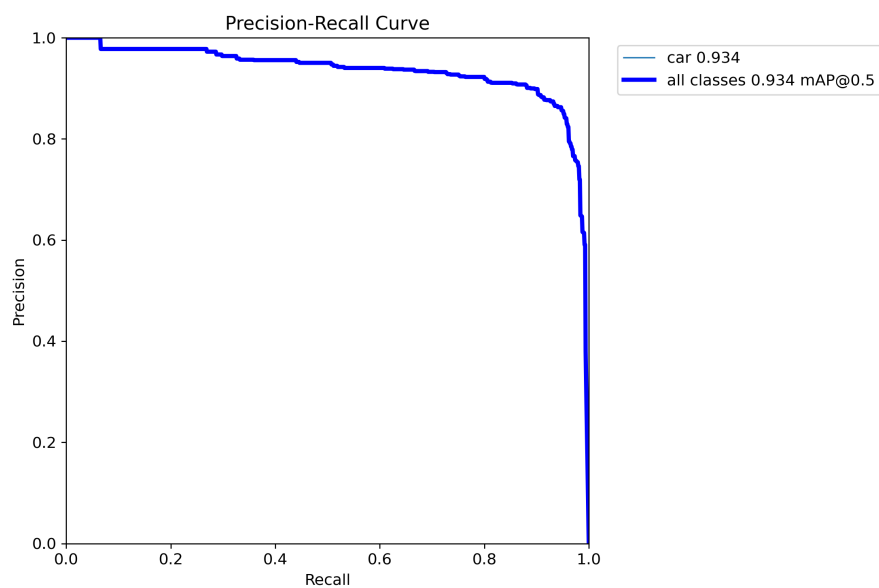


Figure 5. P-R curve

图 5. P-R 曲线



Figure 6. Chart of predicted results
图 6. 预测结果图

4.3. 实验结果对比

在表 2 中比较了本文算法与 YOLOv2、SSD、Faster R-CNN 以及 Deformable R-FCN 网络模型在 UCAS-AOD 数据集测试集上的 AP 以及单幅图片检测时间。表 2 中可看出本文提出的方法在测试集上 AP 为 93.4%，较其它方法准确率较高；单幅图像检测时间为 0.021 s，检测时间快，满足对航拍车辆检测的实时性的要求，效果优于其它网络模型。

Table 2. Model accuracy vs. detection time
表 2. 模型准确率与检测时间对比

网络模型	准确率(%)	检测时间(s)
YOLOv2	55.82	0.024
SSD	65.58	0.023
Faster R-CNN	83.0	0.092
Deformable R-FCN	91.7	0.100
本文算法	93.4	0.021

5. 总结

航拍车辆识别具有广阔的应用前景，但是传统方法需要进行繁杂的特征设计和特征提取。针于航拍

车辆目标检测, 城市道路环境复杂造成的错检, 以及车辆目标尺度较小、密度大、遮挡率高的问题, 本文提出一种基于改进的 YOLOv8 的目标检测方法。在 YOLOv8 网络的基础上, 对其进行改进, 在骨干网络中加入注意力机制, 协助模型捕捉重要信息, 同时将头部网络的最后一个 C2f 结构修改为 C3 结构。采用 AdamW 优化器进行梯度下降, 加快网络收敛。经实验本文方法在检测精度较高, 速度较快, 能很好地解决航拍车辆目标检测问题。

基金项目

在此特别感谢广州软件学院校级科研项目对本文的支持: ky202201、ky202207、KY202305、KY202306。

参考文献

- [1] 肖雨晴, 杨慧敏. 目标检测算法在交通场景中应用综述[J]. 计算机工程与应用, 2021, 57(6): 30-41.
- [2] 梁业成. 基于特征聚焦的航拍图像旋转目标检测算法研究[D]: [硕士学位论文]. 长春: 吉林大学, 2023.
- [3] 南晓虎, 丁雷. 深度学习的典型目标检测算法综述[J]. 计算机应用研究, 2020, 37(S2): 15-21.
- [4] 赵爽, 黄怀玉, 胡一鸣, 等. 基于深度学习的无人机航拍车辆检测[J]. 计算机应用, 2019, 39(2): 91-96.
- [5] 姜淙文, 金立左. Vehicle-YOLO——一种基于航拍影像的车辆检测模型[J]. 微型电脑应用, 2023, 39(9): 134-137.
- [6] Qiu, H., Li, H., Wu, Q., *et al.* (2019) A2RMNet: Adaptively Aspect Ratio Multi-Scale Network for Object Detection in Remote Sensing Images. *Remote Sensing*, **11**, Article ID: 1594. <https://doi.org/10.3390/rs11131594>
- [7] 李旭, 宋世奇, 殷晓晴. 基于目标空间分布特征的无人机航拍车辆实时检测技术研究[J]. 中国公路学报, 2022, 12(35): 193-204.
- [8] 单铭琦, 文峰, 高文印. 基于模型压缩算法改进 YOLOv4-tiny 模型的车辆检测技术研究[J]. 沈阳理工大学学报, 2023, 42(2): 36-42.
- [9] Reis, D., Kupec, J., Hong, J., *et al.* (2023) Real-Time Flying Object Detection with YOLOv8. arXiv: 2305.09972v1.
- [10] 李耀. 基于深度卷积神经网络的航拍图像车辆目标检测算法研究[D]: [硕士学位论文]. 开封: 河南大学, 2023.
- [11] 杨波, 何金平, 张立娜. 基于改进 SPP-x 的 YOLOv5 神经网络水稻叶片病害识别检测[J]. 中国农机化学报, 2023, 44(9): 190-197.