

基于改进YOLOv8的火灾目标检测系统

马冉, 顾宏*

天津工业大学物理科学与技术学院, 天津

收稿日期: 2023年11月13日; 录用日期: 2024年2月19日; 发布日期: 2024年2月27日

摘要

火灾发生初期, 烟雾状态变化多端火焰的尺寸大小也非常小, 现有的目标检测算法面对这复杂情况下会出现检测速度慢、检测准确率低。针对类似这样的问题, 本文提出了基于改进YOLOv8的火灾目标检测系统。在YOLOv8的骨干网络末端添加BotNet结构, 用来增强网络对火灾的特征提取, 在YOLOv8的头部末端引入EMA注意力机制防止权重剧烈变化。改进的YOLOv8模型提高了目标检测的精确度。实验的结果表明, 改进的YOLOv8模型与YOLOv8模型对比, 改进的YOLOv8模型在mAP上提高了2.3%、火灾与烟雾的预测准确率也分别提高了1.4%和1%, 进一步说明改进的YOLOv8模型可以满足对火灾的目标检测。

关键词

目标检测, YOLOv8, BotNet, EMA

Fire Target Detection System Based on Improved YOLOv8

Ran Ma, Hong Gu*

School of Physical Science and Technology, Tiangong University, Tianjin

Received: Nov. 13th, 2023; accepted: Feb. 19th, 2024; published: Feb. 27th, 2024

Abstract

In the early stages of a fire, the smoke state changes frequently and the size of the flame is also very small. Faced with this complex situation, existing target detection algorithms will have slow detection speed and low detection accuracy. In response to problems like this, this article proposes a fire target detection system based on improved YOLOv8. The BotNet structure is added to the end of the backbone network of YOLOv8 to enhance the network's feature extraction of fires, and the EMA

*通讯作者。

attention mechanism is introduced at the head end of YOLOv8 to prevent drastic changes in weights. The improved YOLOv8 model improves the accuracy of target detection. The results of the experiment show that compared with the YOLOv8 model, the improved YOLOv8 model increased mAP by 2.3%, and the fire and smoke prediction accuracy also increased by 1.4% and 1% respectively, further demonstrating that the improved YOLOv8 model can Meet the target detection of fire.

Keywords

Target Detection, YOLOv8, BotNet, EMA

Copyright © 2024 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

火灾是一种具有高度破坏性和危险性的突发事件, 对人员生命安全和财产造成严重威胁。快速、准确地检测和响应火灾对于减小损失、保护人员安全和财产安全至关重要。近年来, 计算机视觉和人工智能技术的迅速发展为火灾监测和管理提供了新的可能性。在这一背景下, 火灾目标检测系统成为一个备受关注的研究领域, 其目标是自动地识别和定位火灾相关目标, 如火源、烟雾、人员和财产。

目前, 深度学习技术已经在目标检测领域取得了显著的成就, 特别是卷积神经网络(CNN)的应用。其中, YOLO (You Only Look Once)系列模型是一类广泛应用于目标检测的深度学习模型, 其独特的特点是端到端的检测, 能够在一次前向传播中同时完成目标检测和定位, 具有高效性和准确性的优势。然而, 传统 YOLO 模型在火灾目标检测中仍然存在一些挑战, 例如烟雾与火源的复杂背景、不同尺度目标的检测问题等。因此, 改进和优化 YOLO 模型以适应火灾目标检测的需求成为当前研究的热点之一。

传统的目标检测算法需要根据不同的具体任务手动设计不同的特征, 例如用于人脸识别的 Haar [1]特征和用于行人检测的 HOG [2]特征。这些功能并非普遍存在, 并且鲁棒性较差, 导致检测结果不佳。2012 年提出的 AlexNet 网络引发了深度学习的研究热潮, 而在 2014 年, R-CNN [3]真正将其引入目标检测领域。

目前来说, 目标检测可以分为两大类: 单阶段检测和双阶段检测, 两者的主要区别就是在检测过程中是否提前生成候选区。单阶段检测主要包括: R-CNN [3]、fast R-CNN [4]、faster R-CNN [5], 双阶段检测主要包括: SSD 系列、YOLO 系列。目标检测要求检测精度高、检测准确率高。双阶段检测根据检测算法提前生成候选区, 然后再对候选区进行分析, 这样在速度上与单阶段检测差别很大, 因此选用单阶段检测的方法用于道路目标检测会大大提高检测的实时性。

近年来, 深度学习在计算机视觉领域取得了巨大的成果, 并且在道路目标检测方面也具有广阔的应用前景。本文将一种称为 YOLOv8 的单阶段深度学习方法应用于火灾目标领域。由于其高准确性、快速推理速度和灵活部署的优势, YOLOv8 自从其代码开源以来就受到了许多计算机视觉领域研究人员的密切关注和研究。为了提高 YOLOv8 在目标检测的性能, 我们引入了注意力机制的方法, 以使 YOLOv8 模型更加关注具有丰富信息的目标特征。

2. YOLOv8 模型

2.1. YOLOv8 模型概述

YOLOv8 是 Ultralytics 公司在 2023 年 1 月 10 号开源的 YOLOv5 的下一个重大更新版本, 目前支持

图像分类、物体检测和实例分割任务, 按照官方描述, YOLOv8 是一个 SOTA 模型, 它建立在以前 YOLO 版本的成功基础上, 并引入了新的功能和改进, 以进一步提升性能和灵活性。具体创新包括一个新的骨干网络、一个新的 Ancher-Free 检测头和一个新的损失函数, 可以在从 CPU 到 GPU 的各种硬件平台上运行, YOLOv8 对比其他版本的优点具体见图 1。

Ultralytics 开源库的两个主要优点是:

1. 融合众多当前 SOTA 技术于一体。
2. 未来将支持其他 YOLO 系列以及 YOLO 之外的更多算法。

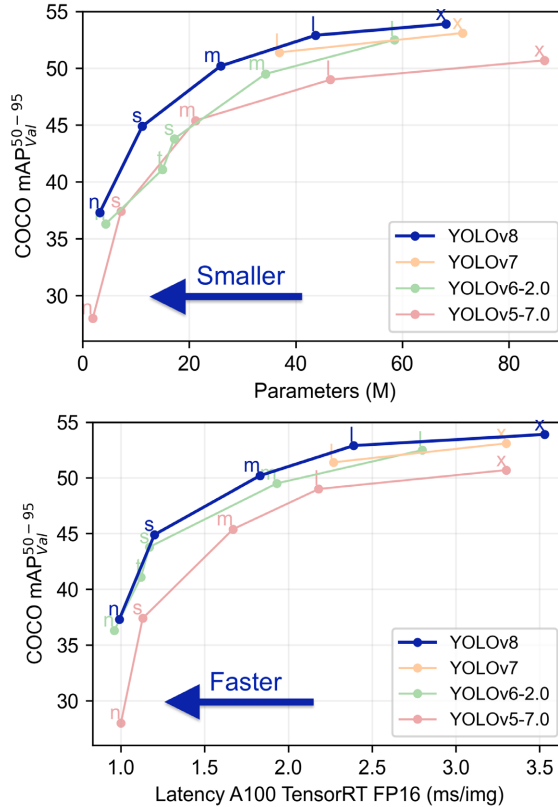


Figure 1. Advantages of YOLOv8 compared to other versions
图 1. YOLOv8 对比其他版本的优点

2.2. YOLOv8 结构分析

2.2.1. YOLOv8 骨干: 从输入图像提取特征信息

YOLOv8 的骨干网络是 Darknet-53。Darknet-53 是由 Joseph Redmon 提出的一个深度卷积神经网络模型, 它由 53 个卷积层和池化层组成用于提取图像特征。Darknet-53 在 YOLOv3 网络中被广泛使用, 并在 YOLOv4 和 YOLOv5 中也得到了应用。它具有较强的特征提取能力, 能够在目标检测任务中提供较好的性能。

2.2.2. YOLOv8 颈部: 为了更好的利用骨干提取的特征, 起着特征融合的作用

YOLOv8 中将 YOLOv5 的 C3 结构换成了梯度流更丰富的 C2f 结构, 并对不同尺度模型调整了不同的通道数 C3 模块的结构图, 然后再对比与 C2f 的具体的区别。针对 C3 模块(如图 2 所示), 其主要是借助 CSPNet 提取分流的思想, 同时结合残差结构的思想, 设计了 C3 Block, CSP 主分支梯度模块 BottleNeck [6]模块。同时堆叠的个数由参数 n 来进行控制, 也就是说不同规模的模型, n 的值是有变化的。

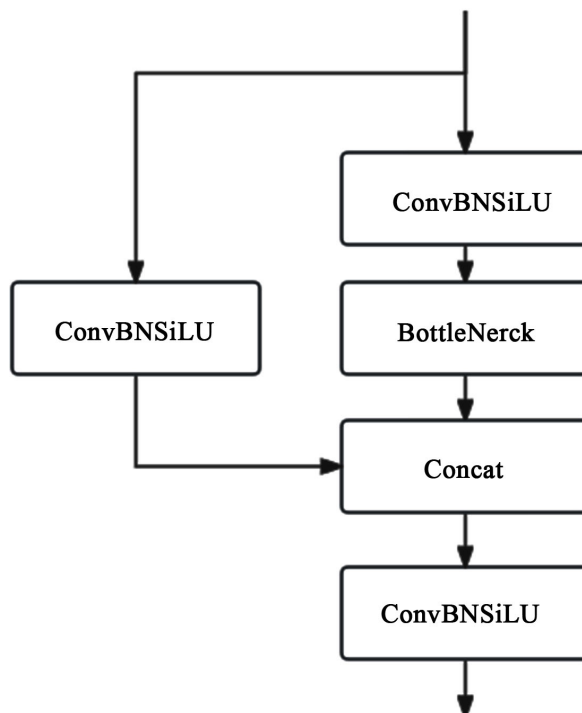


Figure 2. C3 module structure diagram
图 2. C3 模块结构图

C2f 模块(如图 3 所示)就是参考了 C3 模块以及 ELAN 的思想进行的设计,让 YOLOv8 可以在保证轻量化的同时获得更加丰富的梯度流信息。

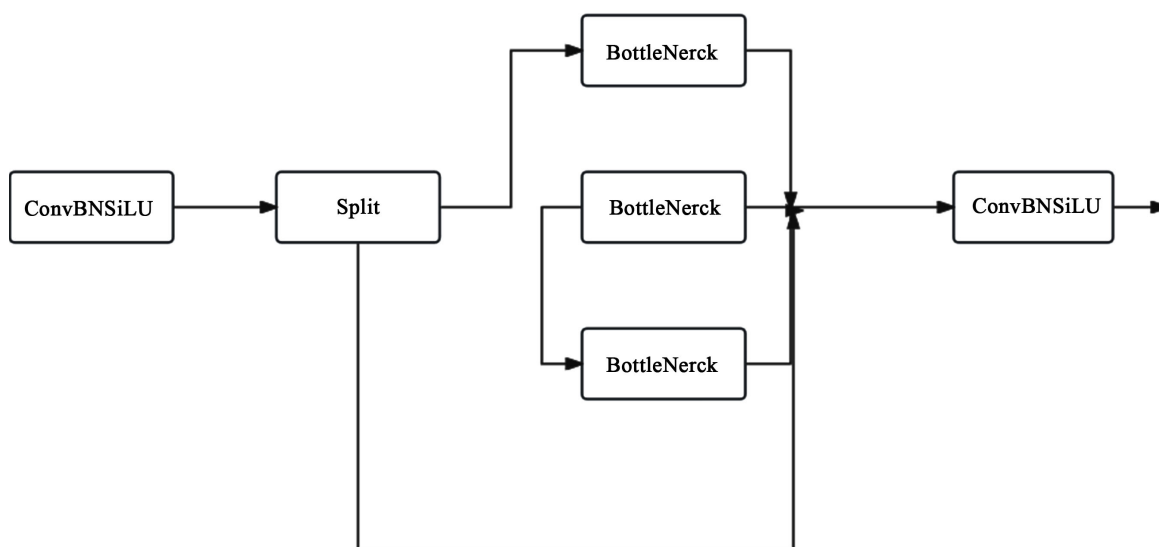


Figure 3. C2F module structure diagram
图 3. C2F 结构图

2.2.3. YOLOv8 头部：输出边界框和预测的类别

“head”的主要作用是将特征图(来自特征提取网络)映射到目标检测结果。它通常包括一个或多个卷积层、全连接层和激活函数,以及一些参数化的操作。具体作用如下:

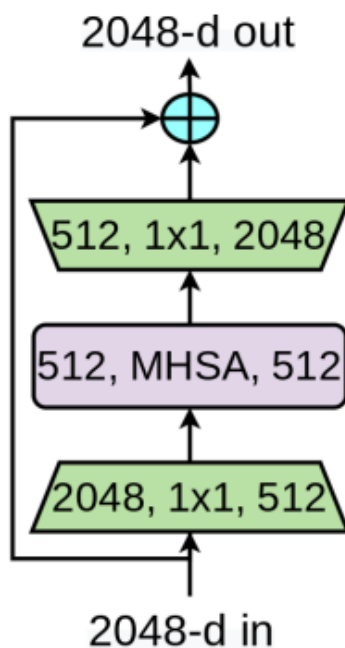
1. 目标框坐标预测：“head”会生成目标框的坐标信息，通常以边界框的左上角坐标和右下角坐标(或中心点坐标、宽度和高度)的形式输出。这些坐标信息用于确定检测到的目标的位置。
2. 目标类别预测：“head”也负责为每个检测到的目标框分配一个类别标签。这通常涉及到使用 Softmax 激活函数，使得每个框的类别概率总和为 1，以确定目标属于哪个类别。
3. 目标得分(置信度)预测：每个目标框都伴随一个置信度分数，该分数表示该框中是否包含目标物体。“head”生成这些置信度分数，通常使用 sigmoid 激活函数，值范围在 0 到 1 之间。
4. 多尺度输出：YOLO 通常在不同尺度下进行目标检测，因此对于每个尺度，都会有一个相应的“head”来生成目标检测结果。这有助于检测不同尺寸的目标物体。

2.3. YOLOv8 算法的改进

2.3.1. 加入 BotNet 结构

BotNet [7]一种简单但功能强大的 backbone 架构(见图 4)，该架构将自注意力纳入了多种计算机视觉任务，包括图像分类，目标检测和实例分割。该方法在实例分割和目标检测方面显著改善了基线，同时还减少了参数，从而使延迟最小化。

BoTNet 使用 Mask R-CNN 框架在 COCO 实例分割基准测试中取得了 44.4%的 Mask AP 和 49.7%的 Box AP，超过了 ResNeSt 在 COCO 验证集上以单模型和单尺度获得的最佳结果。



Bottleneck Transformer

Figure 4. BotNet module structure diagram
图 4. BotNet 结构图

2.3.2. 引入 EMA 注意力机制

EMA [8] (Exponential Moving Average)注意力机制是一种用于改进深度学习模型性能的技术，特别是在自然语言处理领域中的注意力机制(见图 5)。它的名称源自指数移动平均，这种方法用于平滑时间序列数据。EMA 注意力机制的核心思想是使用指数移动平均来平滑模型中的注意力权重，以减少噪声并提高

模型的稳定性和泛化能力。EMA 注意力机制通过引入指数移动平均来解决权重变化剧烈问题。具体来说, 它在每个训练步骤中计算一个新的 EMA 权重, 并将其与当前的注意力权重相结合。这使得 EMA 权重相对平滑, 减少了权重的剧烈变化, 从而提高了模型。

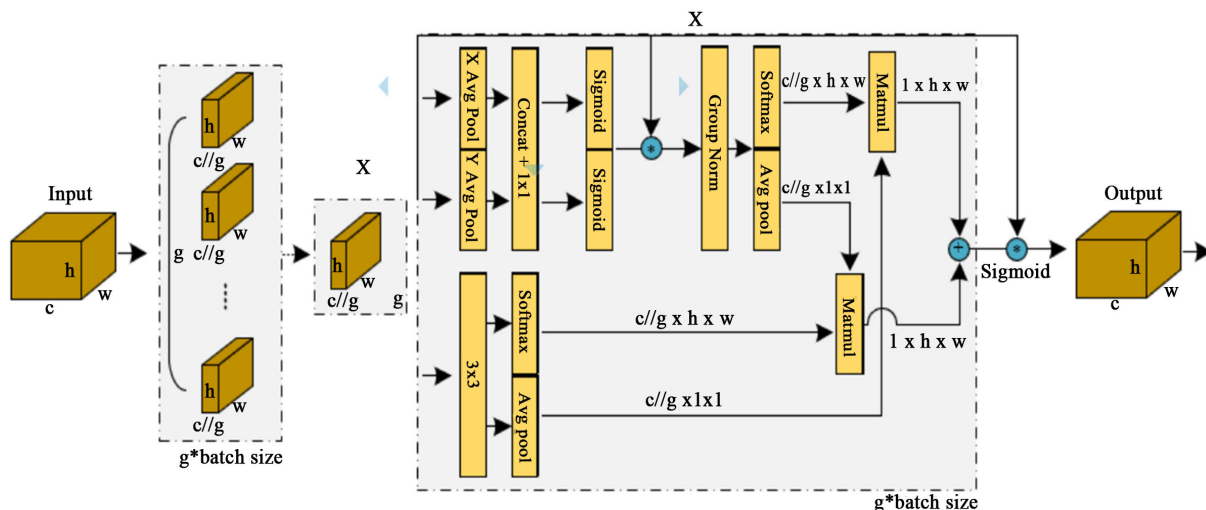


Figure 5. EMA attention mechanism structure diagram

图 5. EMA 注意力机制结构图

2.3.3. 算法改进说明

1. 在 YOLOv8 的骨干网络末端引入 BotNet 架构, 其作用是用来优化骨干网络对输入图像特征的提取。
2. 在 YOLOv8 的头部网络末端引入 EMA 注意力机制, 其作用是用来防止权重的剧烈变化, 进而提高模型的鲁棒性。

3. 实验的结果与分析

3.1. 实验的软件版本与硬件配置

本文所有的实验都基于同一个实验环境以保证实验数据的真实性与可靠性。本文实验所用的操作系统为 Windows11 版本, 深度学习的框架为 Pytorch1.8.1, 语言版本为 Python3.8, GPU 为 NVIDIA RTX 2080 Ti, CPU 为 Intel Core i7-4790, Cuda 版本为 11.1。

3.2. 数据集介绍

Kaggle 是一个数据科学和机器学习社区, 它提供了大量的开放数据集供数据科学家、机器学习工程师和研究者使用。这些数据集涵盖了各种不同的主题和领域, 可以用于各种分析、建模和可视化项目。本文的数据集选自 Kaggle 平台名为 Fire_Eye 的数据集, 其中训练集有 3871 张图片, 验证集有 431 张图片, 每张图片的大小尺寸都是 640×640 , 总共分为 Fire 与 Smoke 两个分类。

3.3. 实验结果

准确率(Precision)、召回率(Recall)和平均准确度(mAP)被用作评估指标来评估我们改进模型的性能。准确率用于评估模型检测的准确性, 而召回率用于评估模型找到所有正样本的能力。AP 是一个综合评价指标, 结合了准确率和召回率, 它使用积分方法计算由准确率 - 召回率曲线和坐标轴所围成的面积。整

个数据集上的 mAP 可以通过对所有类别的 AP 值进行平均得到。

在目标检测中通常有四个检测评估指标：

TP (True Positives): 实际正类预测为正类的数量。

TN (True Negative): 实际负类预测为负类的数量。

FP (False Positives): 实际负类预测为正类的数量。

FN (False Negative): 实际正类预测为负类的数量。

准确率和召回率可以计算公式为：

$$precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

将 P 表示为准确率, r 表示为召回率, N 表示为种类数, 就可以得到 mAP 值的计算公式：

$$mAP = \frac{AP}{N} = \frac{\int_0^1 p(r)dr}{N}$$

本篇文章是关于火灾的目标检测, 只有 Fire 与 Smoke 两个分类, 所以 N 的取值为 2。

本次实验对比 YOLOv8 模型与引入 BotNet 和 EMA 注意力机制的 YOLOv8 模型(下文称 BotNet-EMA-YOLOv8 模型)做了深刻的对比, 得出 BotNet-EMA-YOLOv8 模型相比于 YOLOv8 模型 mAP 值提高了 2.3%, Fire 与 Smoke 的预测准确率也分别提高了 1.4% 和 1%。具体结果如下图 6~9 所示。由此可见, BotNet-EMA-YOLOv8 模型在在火灾检测方面性能更佳。

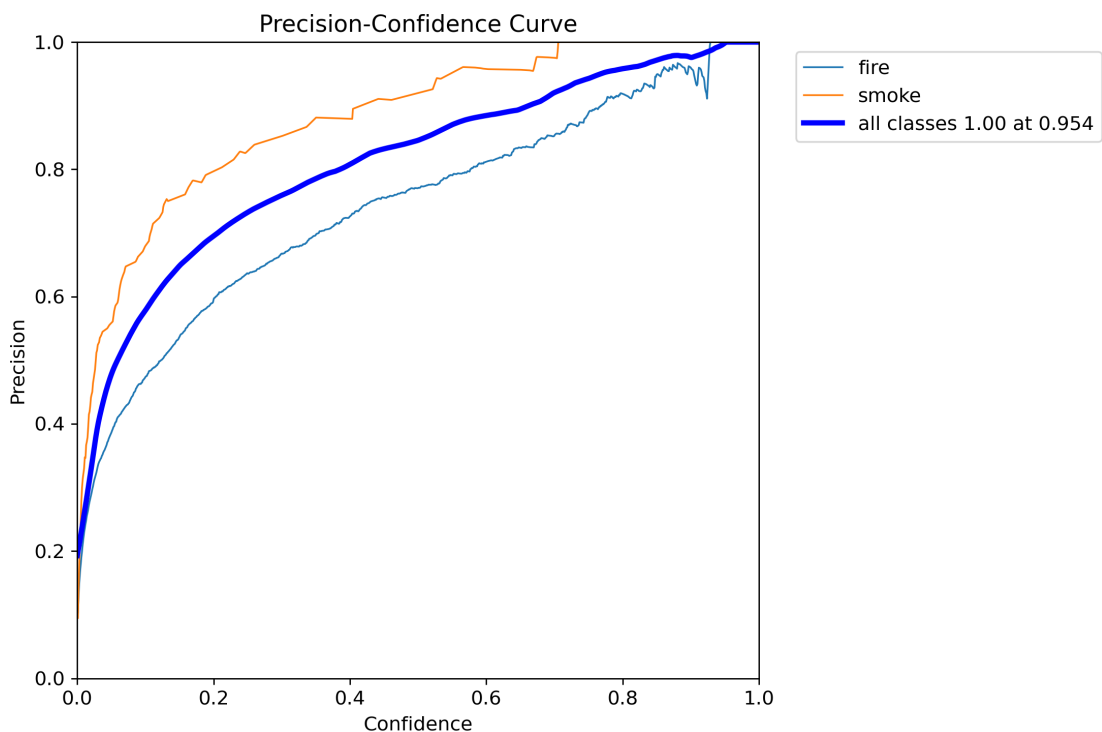


Figure 6. BotNet-EMA-YOLOv8 model precision-confidence curve

图 6. BotNet-EMA-YOLOv8 模型精度置信曲线

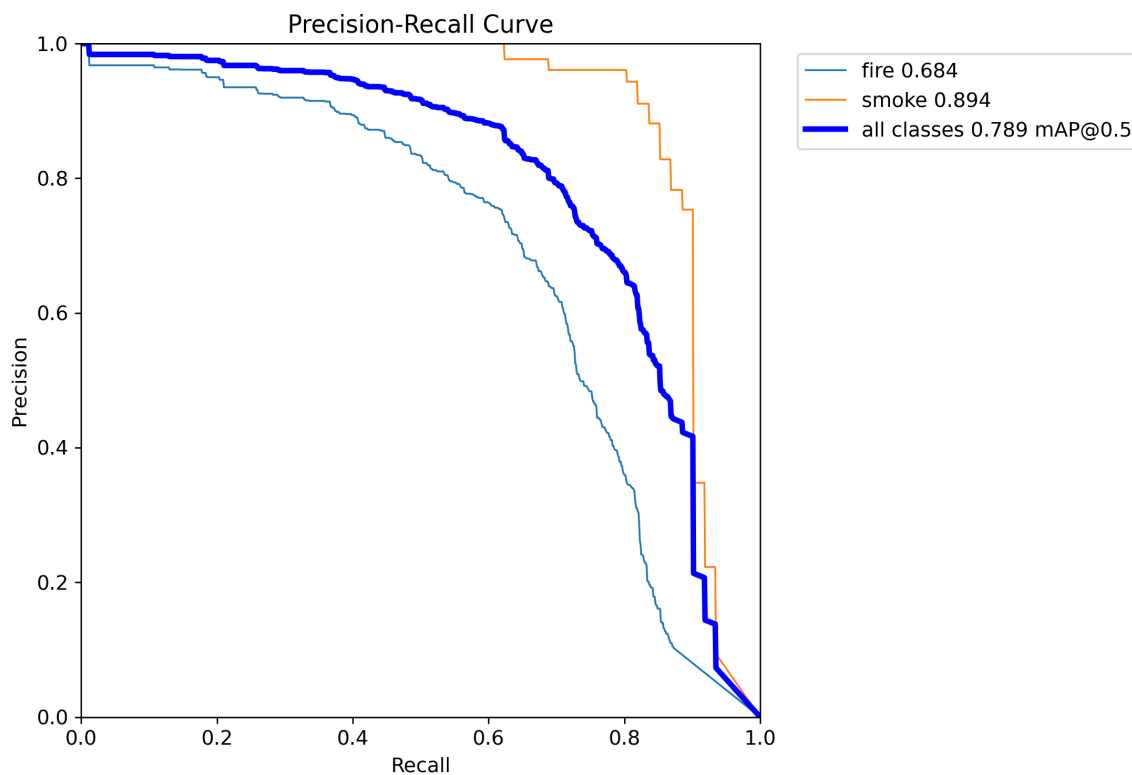


Figure 7. BotNet-EMA-YOLOv8 model precision-confidence curve

图 7. BotNet-EMA-YOLOv8 模型精度召回曲线

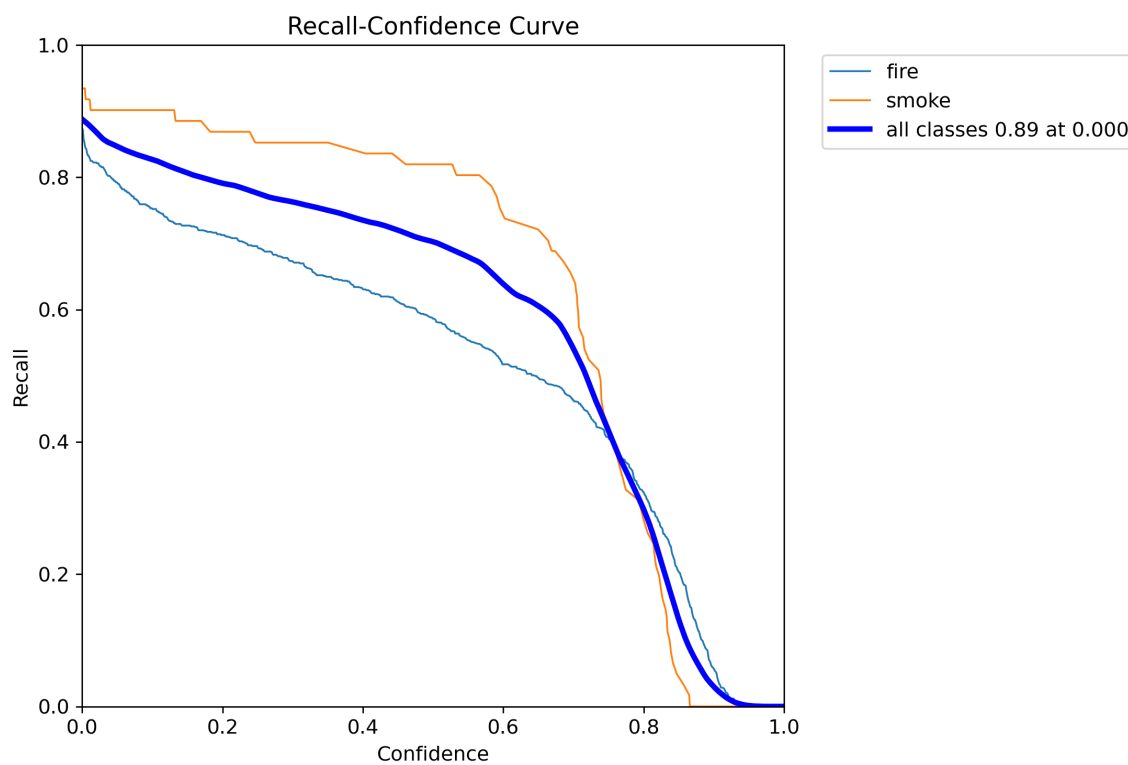


Figure 8. BotNet-EMA-YOLOv8 model recall-confidence curve

图 8. BotNet-EMA-YOLOv8 模型召回置信曲线

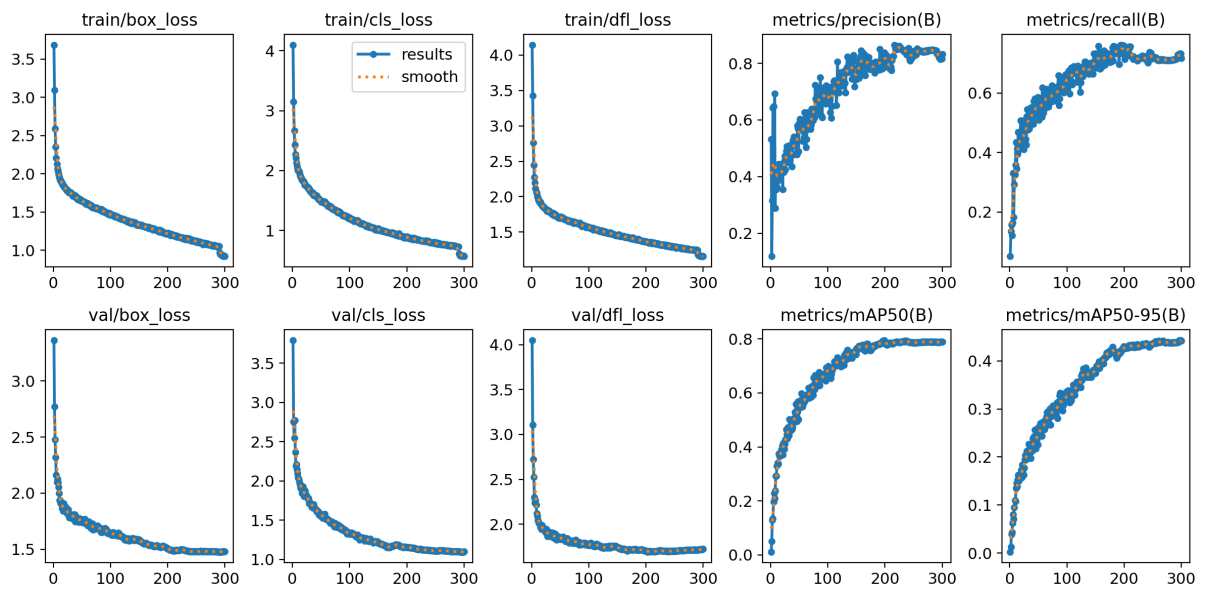


Figure 9. BotNet-EMA-YOLOv8 model other related results
图 9. BotNet-EMA-YOLOv8 模型其他相关结果图

火灾检测结果如图 10 所示, 左侧的一列为 BotNet-EMA-YOLOv8 模型所检测的结果, 右侧的一列为 YOLOv8 模型所检测的结果, 经过对比可以得出 BotNet-EMA-YOLOv8 模型对火灾的小目标检测比 YOLOv8 更加好, 对火灾的大目标检测比 YOLOv8 模型准确率更高。



Figure 10. BotNet-EMA-YOLOv8 model fire detection pictures
图 10. BotNet-EMA-YOLOv8 模型火灾检测图

4. 结论

火灾目标检测对于保护人民的财产安全以及生命安全有着重要的研究价值, 有利于防患于未然, 具有一定的经济意义。本文应用火灾与烟的数据集在模型上训练, 在 YOLOv8 的基础上引入 BotNet 架构和 EMA 注意力机制, BotNet-EMA-YOLOv8 模型比 YOLOv8 模型在小目标检测上检测的更加精准, 该模型不仅提升了小目标的检测能力, 也提升了包括 mAP 值在内的检测评估指标, 具有非常高的实用价值。

参考文献

- [1] 胡宇晨, 李秋生. 基于 Haar-Like T 和 LBP 特征的人脸识别方法[J]. 自动化与仪表, 2023, 38(10): 52-56+61. <https://doi.org/10.19557/j.cnki.1001-9944.2023.10.012>
- [2] 李文书, 韩洋, 阮梦慧, 等. 改进的基于增强型 HOG 的行人检测算法[J]. 计算机系统应用, 2020, 29(10): 199-204. <https://doi.org/10.15888/j.cnki.csa.007587>
- [3] Wang, F., Fu, X., Duan, W., et al. (2023) Visual Detection of Lost Ear Tags in Breeding Pigs in a Production Environment Using the Enhanced Cascade Mask R-CNN. *Agriculture*, **13**, Article 2011. <https://doi.org/10.3390/agriculture13102011>
- [4] Cao, Z., Yi, Z., Pan, J., Su, H., et al. (2023) L-Dwarf Detection from SDSS Images Using Improved Faster R-CNN. *The Astronomical Journal*, **165**, Article 184. <https://doi.org/10.3847/1538-3881/acc108>
- [5] Sivakumari, T. and Vani, R. (2023) Deep Learning-Based Automated Knee Joint Localization in Radiographic Images Using Faster R-CNN. *Current Medical Imaging*, **20**, Article ID: e15734056262464. <https://doi.org/10.2174/0115734056262464230922112606>
- [6] Hu, Y., Huang, X., Luo, X., et al. (2024) Learning Foreground Information Bottleneck for Few-Shot Semantic Segmentation. *Pattern Recognition*, **146**, Article ID: 109993. <https://doi.org/10.1016/j.patcog.2023.109993>
- [7] Antonia, A., Stefania, Z., Giovanni, S., et al. (2023) The Evolution of Mirai Botnet Scans over a Six-Year Period. *Journal of Information Security and Applications*, **79**, Article ID: 103629. <https://doi.org/10.1016/j.jisa.2023.103629>
- [8] 彭朝琴, 李奇聪, 陈娟, 等. 基于 GRU 和改进注意力机制的多信息融合的 EMA 故障诊断方法[J/OL]. 北京航空航天大学学报: 1-16. <https://doi.org/10.13700/j.bh.1001-5965.2023.0584>, 2023-11-04.