

基于 YOLOv8 的头盔检测模型设计与实现

陈邵杰, 郭昊, 蔡明珠, 刘政, 黄俊毅, 王晶昊, 杨双菁, 田睿朴, 苗馨月

Abstract—随着人工智能的发展和交通安全需求的增加, 基于深度学习的目标检测技术在行人头盔佩戴检测中发挥着重要作用。在本研究中, 针对行人检测场景采用 YOLOv8l 模型进行头盔检测, 并通过稀疏化和剪枝技术对模型进行优化, 在高精度的同时提升推理效率。利用稀疏化方法对 YOLOv8l 模型的权重进行优化, 增强模型的稀疏性, 减少计算冗余。通过剪枝技术有效去除冗余的网络参数和连接, 大幅减少模型的计算量和参数规模, 优化后的模型在嵌入式和边缘计算设备上运行效率显著提升。实验结果表明, 经过稀疏化和剪枝后的 YOLOv8l 模型参数量减少了约 30%, 推理速度提高了近 40%, 并且在行人头盔检测任务中维持了高达 98% 的准确率。我们还将优化后的模型与 YOLOv3、YOLOv5、YOLOv7 等多个版本进行了对比, 验证了 YOLOv8l 在行人检测精度和速度方面的优势。

Index Terms—交通管理, YOLOv8, 头盔检测, 深度学习, 目标检测

I. 引言

A. 选题背景与意义

Subsection text here.

1) 课题背景: subsection text here.

2) 研究意义: subsection text here.

B. 研究问题分析

Subsection text here.

II. 行人数据图像处理

The conclusion goes here.

A. 实验数据集

Subsection text here.A'

B. 图像预处理

Subsection text here.

C. yolo 标准数据集

Subsection text here.

感谢电子科技大学机器学习课程闫老师与张老师的教学。
2024 年 10 月

D. 数据集标注

Subsection text here.

III. 基于 YOLOv8 的检测系统实现

YOLOv8 是一种精准高效的目标检测解决方案, 系统采用 YOLOv8 作为核心检测算法, 识别电动车及其驾驶员, 并判断是否存在佩戴头盔行为。训练与测试使用的均为 540*355 的 YOLO 标准数据集, 图 3-1 为整个实验流程图:

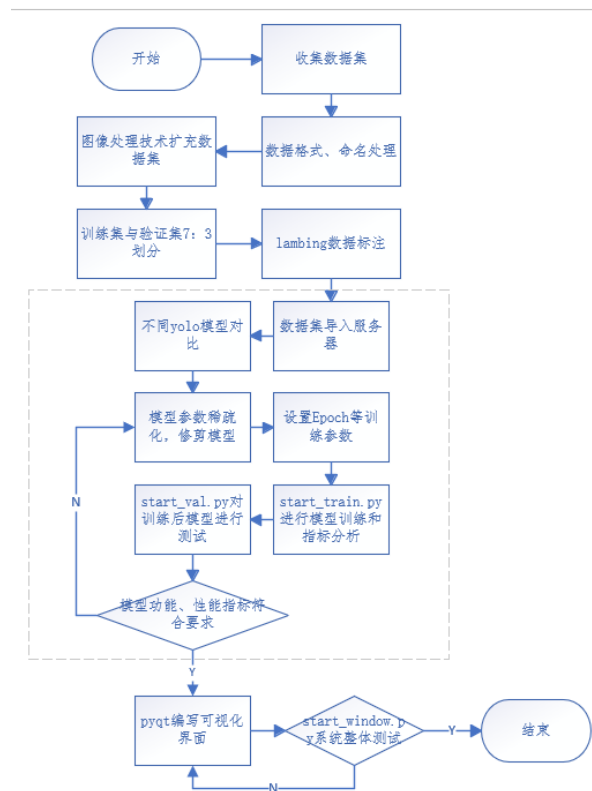


Fig. 3-1 实验流程图

A. 训练环境搭建

对于大量图片数据集与复杂模型, 通过 GPU 进行模型训练大加速, 通过 AutoDL 平台租借云服务器实现

YOLOv8 模型的搭建, 训练, 测试与部署工作。

本次实验使用的硬件环境为: CPU 16 vCPU Intel(R) Xeon(R) Platinum 8481C, GPU RTX 4090D, 显存大小 24G, 硬盘为 80G。

软件环境: 操作系统 ubuntu22.04, 编程语言 python 版本 3.8.5, pytorch 版本 1.10.0, Cuda 版本 11.8。

B. 模型训练

本次实验采用 YOLOv8 模型进行目标检测训练, 整个训练过程依托于高性能服务器硬件及成熟的软件环境, 确保了模型的高效收敛和稳定性能。

训练流程从数据准备开始, 输入图像维度为像素 540*355, 首先对数据集进行了精细的标注、清洗和划分, 形成训练集、验证集和测试集, 为模型的训练和评估提供了坚实的数据基础。并且通过高性能硬件配置使得数据处理与模型计算可以高效并行执行, 显著提高了训练效率。

整个训练过程共进行了 100 个 epoch, 训练阶段首先加载数据集, 并进行数据增强操作, 提高模型的泛化能力。接着初始化 YOLOv8 模型, 设置包括学习率和批次大小等超参数, 以确保训练初期能够有效学习特征。每个 epoch 包含前向传播、损失计算、反向传播和权重更新等步骤。

在服务器上, GPU 的强大算力极大加速了这些运算, 在卷积层的特征提取和参数更新过程中尤为显著。每个 epoch 结束后, 模型会在验证集上进行评估, 以监控性能并避免过拟合的发生。训练日志详细记录了每个 epoch 的损失、精度和召回率等关键指标, 并且实时生成了损失曲线和精度曲线, 以便于监控训练过程中的动态表现和分析模型的改进方向。

在进行模型参数配置时, 为了确保对比实验的一致性和准确性, 我们需要核心参数固定不变, 以排除对训练效果产生的潜在影响。训练 epoch 为 100 轮, 使用随机梯度下降优化器, 初始学习率和周期学习率均设置为 0.01, 设置 batch 数量为 64。经训练, YOLOv8 模型最终达到了预期的检测效果, 能够精准地识别目标物体, 表现出良好的鲁棒性和检测性能。通过数据预处理、模型初始化、迭代训练和模型评估的完整闭环与高性能服务器硬件的支持, 使得模型在合理时间内完成了高效训练。

C. 模型部署

IV. 调整与改进

电动自行车头盔的检测可能由于骑手的高速移动或环境的复杂性而变得颇具挑战性。例如, 头盔可能因与摄像头距离较远而成为难以辨识的小目标。针对检测过程中低分辨率图像及小目标检测困难, 背景与检测目标相似等问题, 改进 Yolov8n 模型。传统目标检测网络容易忽略或误

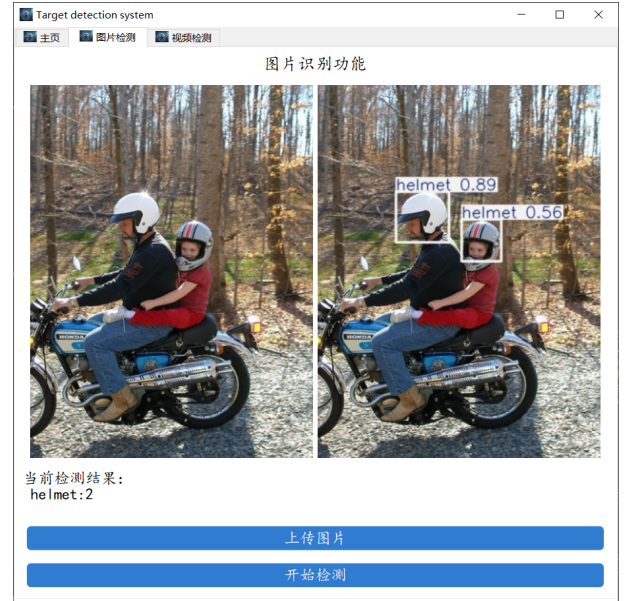


Fig. 3-2 部署界面

判电动车骑行过程中远处、边缘或比例较小的头盔, 进而导致漏检或误检。对此需要增强原始 YOLOv8n 模型以提升对小目标检测精度。

A. 卷积注意力模块

本研究模型加入了基于卷积注意力机制模块^[1](Convolutional Block Attention Module, CBAM)的小目标检测优化策略。CBAM 作为一种即插即用的注意力机制模块, 可直接嵌入至多种流行的卷积神经网络架构中。该模块通过对输入特征图的不同区域赋予不同的权重, 使网络能够更加聚焦于对特定任务具有重要性的信息, 同时忽略不相关的信息。因此, 在针对佩戴头盔的行人检测任务中, 由于目标与摄像头距离较远导致行人及头盔尺寸较小, 且在检测图像中所占比例较低, 引入 CBAM 能够使网络更加专注于前景中的小目标, 并有效定位感兴趣的区域。这进一步增强了网络的特征提取能力, 使得在最小感受野的特征图上能够更精确地识别小目标。CBAM 的总体架构如图 4-1 所示。

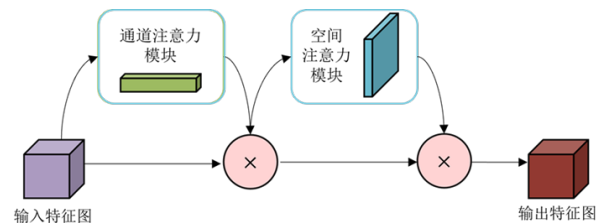


Fig. 4-3 CBAM 结构图

对于输入特征图, CBAM 先使用通道注意力模块, 在宽高维度上对输入进行平均池化和最大池化的操作, 得到两个大小的张量。然后, 将两个张量分别经过共享卷积层, 将输出进行逐元素相加, 再经过非线性激活操作, 得到通道注意力权值, 即大小为的张量。将该权值与输入特征图进行逐元素相乘, 得到通道注意力模块的输出特征图。接着, 对于通道注意力模块的输出特征图, 在通道维度上进行平均池化和最大池化操作, 得到两个大小的张量。然后, 将两个张量在通道维度上进行拼接, 并经过一个卷积操作, 将通道维度降为 1。最终, 经过非线性激活操作, 得到空间注意力权值, 即大小为的张量。将该权值与通道注意力模块的输出特征图进行逐元素相乘, 便得到经过卷积注意力模块处理后的输出特征图。根据论文 [1] 中的实验结果, 在不同的图像分类和目标检测数据集上, 将 CBAM 集成到不同的神经网络模型中后, 模型的性能得到很大的提高, 这证明了该模块的有效性。

针对 YOLOv8n 网络, 本文将卷积注意力 CBAM 模块应用在了特征金字塔网络下采样过程中的每一个跨阶段局部模块之后, 对不同层次的特征进行融合的 Neck 部分, 模型框架修改后如 2 所示。

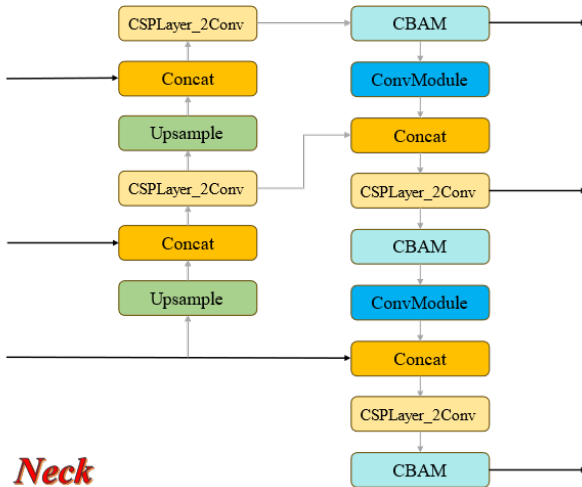


Fig. 4-4 应用 CBAM 卷积注意力模块网络 Neck 部分结构图

B. Transformer Block

Transformer^[2] 是一种专为处理序列数据而设计的深度学习架构。该模型利用自注意力机制实现了高效的并行计算, 并具备捕获长距离依赖关系的能力。

Transformer 的结构如图 3 所示, 主要由编码器和解码器两大部分构成。编码器负责将输入信息映射为中间的隐藏状态表示, 而解码器则将这些隐藏状态转换为最终的输出序列。编码器由若干相同的编码器层堆叠而成, 每一层均包含自注意力机制和前馈神经网络。解码器在结构上

与编码器相似, 但额外增加了一个注意力层, 该层用于关注编码器的输出。

多头自注意力机制是 Transformer 的核心, 它使得模型在处理输入序列时能够同时关注序列中不同位置的信息。查询、键和值通过不同的线性变换分成多组 (头), 然后对每组进行自注意力计算, 最后将所有头的输出拼接起来, 再次通过一个线性层得到最终的输出。

前馈网络是注意力机制之后的一个关键组件。对注意力层的输出进行进一步的非线性变换, 以捕获更复杂的特征和表示。

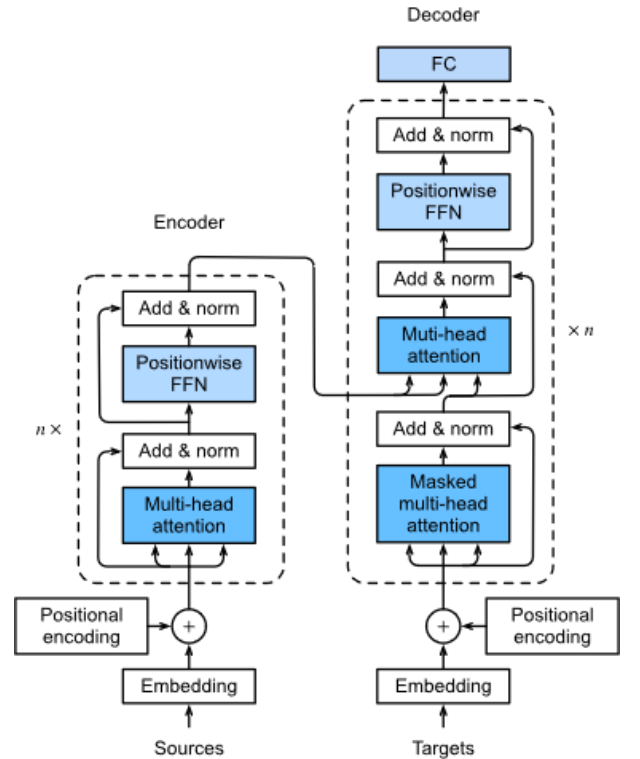


Fig. 4-5 Transformer 结构图

针对 YOLOv8n 网络, 本文在特征处理的中间层 C2f 中加入 Transformer 模块, 修改后的 C2f 结构如图 4 所示。通过引入自注意力机制, 修改后的模型能够更好地捕捉长距离依赖关系, 这有助于模型在处理图像时能够更好地理解全局上下文信息, 从而提升特征的代表能力。且由于 Transformer block 能够有效捕捉全局信息, 从而帮助模型在不同的尺度上捕捉到小目标的特征, 能够更好地区分头盔与复杂的背景。此外, 在 YOLOv8n 的架构中, 特征融合是一个关键环节。Transformer block 可以作为一种有效的特征融合工具, 通过自注意力机制融合不同层次的特征信息, 提高模型在各种环境下的鲁棒性, 减少误报和漏报, 进一步提升检测性能。

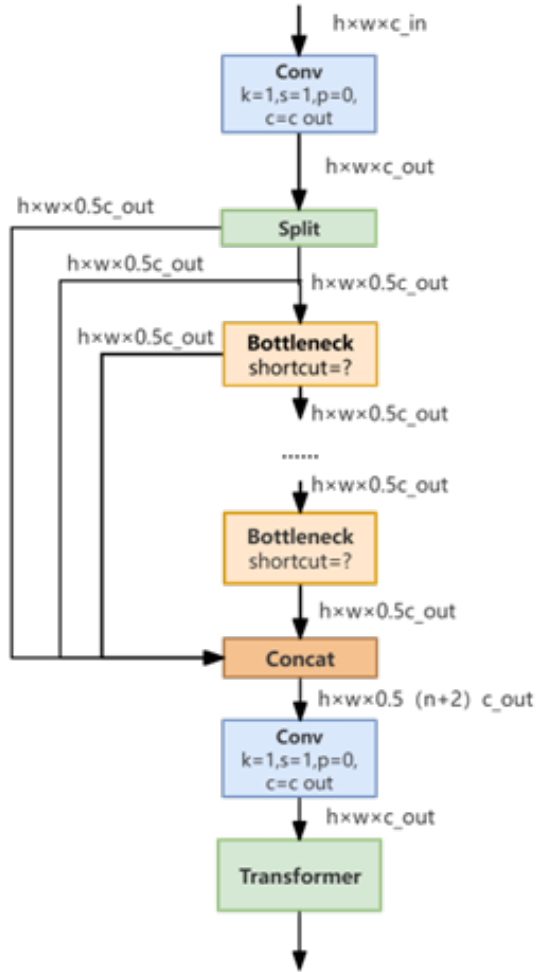


Fig. 4-6 添加 Transformer 后的 C2f 结构图

C. 创新点 3

Subsection text here.

V. 实验结果

The conclusion goes here.

VI. 总结与展望

The conclusion goes here.

致谢

The authors would like to thank...

REFERENCES

- [1] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [2] Vaswani A. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017.