

Research on Improved YOLOv8 Document Image Text Detection

Chengxiang Yu

School of Automation, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, China

Abstract

This paper proposes a text detection algorithm based on YOLOv8 with CA attention mechanism. In order to address the problem of difficulty in distinguishing handwritten and printed text in the digital detection and recognition process of document images, this paper analyzes the YOLOv8 object detection algorithm and points out the errors that existing object detection algorithms may produce in the classification of printed and handwritten text in paper document text detection. In the application of automatic recording of actual digital archive images, the recognition difficulty of handwritten text is greater than that of printed text, and it needs to be distinguished separately. This paper proposes an improved detection algorithm with CA attention mechanism. This paper mainly uses text detection algorithm for table type document images. The experimental results show that this text detection algorithm can effectively improve the classification accuracy of printed and handwritten text in document images, meeting the requirements of engineering applications.

Keywords

digitalization of archives; text detection; YOLOv8; CA attention mechanism

基于改进的 YOLOv8 文档图像文字检测研究

余承翔

南京理工大学自动化学院, 中国 · 江苏 南京 210094

摘 要

论文提出了一种基于YOLOv8添加CA注意力机制的文字检测算法。为了应对文档图像的档案数字化检测识别过程中手写体和印刷体无法较好区分的问题, 论文分析了YOLOv8目标检测算法, 指出了现有的目标检测算法在纸质文档文字检测的过程中可能产生的印刷体与手写体类别的误差, 在实际档案图像数字化的自动著录的应用中, 手写体的识别难度相较印刷体大, 需要区别单独识别, 论文提出了一种添加CA注意力机制的改进的检测算法, 论文主要针对表格类型的文档图像采用文字检测算法, 实验结果表明, 该文字检测算法能够有效提高文档图像的印刷体和手写体的分类精度, 满足工程应用的要求。

关键词

档案数字化; 文本检测; YOLOv8; CA注意力机制

1 引言

文字作为传递信息和沟通的基本的媒介和工具, 在人们的日常生活、学术研究和职业活动中扮演着至关重要的角色。伴随着科技的进步, 尤其是深度学习技术的突破, 已经在文字检测领域取得了显著成就, 成为该领域的主流算法。在深度学习技术广泛应用之前, 传统文字检测方法主要通过三个步骤进行: 图像预处理、识别和后处理。首先对文字区域进行定位, 然后对倾斜文字进行矫正, 在分割出单字符后, 并利用人工提取的特征 (如 HOG) 与分类模型结合, 以识别单个文字^[1]。尽管这些传统方法在简单背景和单一文字场景下效果良好, 但在复杂场景中, 如手写体和印刷体混合出

现时, 其准确性和可靠性会大打折扣。这时, 深度学习算法的优势就显得尤为突出, 它们在提高文字检测和识别的速度及准确性方面发挥了重要作用。

随着深度学习技术的持续发展, 越来越多的档案机构和研究人员开始利用机器学习技术来优化档案管理流程, 这包括档案的分类、手写文档的识别、档案的数字化处理以及档案图像的自动分析等方面^[2]。在档案图像处理中, 手写体和印刷体汉字是两种常见的字体类型, 由于它们的生成机制不同, 各自展现出独有的特征, 这意味着我们不能将这两种字体混为一谈, 而是应该分别对它们进行识别训练。论文提出了一种基于目标检测的方法, 将档案图像中的印刷体和手写体视为两种不同的目标, 旨在识别文档图像中的字体类型。通过改进 YOLOv8 网络, 论文的方法能够有效地检测档案图像中的印刷体和手写体文字, 识别精度得到一定提高。

【作者简介】余承翔 (2001-), 男, 中国安徽宣城人, 硕士, 从事图像处理研究。

2 YOLOv8 算法及改进

2.1 YOLOv8 算法

YOLO (You Only Look Once) 是一种实时目标检测算法,它能够在图像中快速准确地识别出多个物体的位置和类别。YOLO 算法采用的是端到端的训练,即整个网络可以直接从原始图像到最终的目标检测结果进行训练,而无需复杂的预处理或后处理步骤,使得其在实时应用中具有很大的优势。

YOLOv8 算法由 Glenn-Jocher 提出,是跟 YOLOv3 算法、YOLOv5 算法一脉相承的^[3],YOLOv8 算法,它可以支持多种导出格式,而且该模型可以在 CPU 和 GPU 上运行。其网络结构由输入端、Backbone、Neck 和输出端四部分构成。

YOLOv8 改进主要涉及到:backbone 使用 C2f 模块,检测头使用了 anchor-free+Decoupled-head,损失函数使用了分类 BCE、回归 CIOU+VFL (新增项目)的组合,框匹配策略由静态匹配改为了 Task-Aligned Assigner 匹配方式、最后 10 个 epoch 关闭 Mosaic 的操作、训练总 epoch 数从 300 提升到了 500^[4]。

2.2 YOLOv8+CA 注意力机制

Coordinate Attention (CA) 是一种用于移动网络的轻量级注意力机制,旨在在不增加计算成本的前提下增强特征表达能力。与传统的通道注意力机制(如 Squeeze-and-Excitation)不同,CA 在通道注意力中嵌入了位置信息,从而能更好地在空间上进行关注。CA 注意力机制通过将二维的通道注意力拆分为两个一维的编码步骤,分别沿着垂直和水平方向来整合特征。使模型不仅能捕获长距离的空间依赖关系,还能保留细粒度的位置信息,帮助模型在需要定位物体结构的任务中表现更优^[5]。

具体的模块添加是在 YOLOv8 的 ultralytics 包中的 nn 包的 modules 里添加 CA 注意力模块,这里选择在 conv.py 文件中添加 CA 注意力机制,分别在 ultralytics/nn/modules/_init_.py 文件和 ultralytics/nn/tasks.py 文件夹中实现 CA 注意力机制的注册和引用。

3 实验过程介绍

3.1 实验数据集的制作

鉴于档案包含敏感个人信息,目前对于档案图像的研究缺乏公开的公共数据集。论文所用数据集由开发档案 AI 项目期间收集的山西省部分干部档案数据(大多为公务员干部等入职所需填写的表单相关的表格文档)组成,这里的档案主要是通过扫描仪对填写完成的档案图片采集的结果。为了实验数据的丰富性,在档案信息数据源文件中选取了复杂程度各不相同的多种类型档案素材,具体涉及到封闭表格线所划分的区域个数、印刷体和手写体的位置、印刷体和手写体的数量等。本实验总计选取了 8 种不同类型的档案图像,这些档案图像中印刷体和手写体位置的分布也都不是确定的。

原始数据中包含了很多纯打印体文本以及纯印刷体文本,在经过筛选总共选取 1000 张档案表格图片,档案图像的整体结构通常包括顶部的标题栏,标题栏下方是由水平线和垂直线构成的表格区域,具体的档案文本中包含的信息有姓名、年龄、籍贯、家庭成员情况等。所选取有代表性的两类图如图 1、图 2 所示,其中图 1 是左右布局的表格档案,这里档案图像通常所需填写的信息较少,检测的难度会容易一些,图四属于典型的上下布局的表格图像,这类图像的特点是对应的文本内容一行印刷体对应着需要填写的多行的手写体类型,这类图像通常需要检测的对象很多,容易产生干扰的信息也相应的多一些。

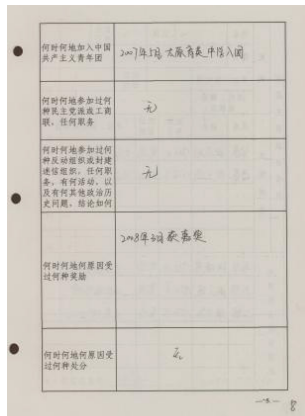


图 1 左右布局表格档案

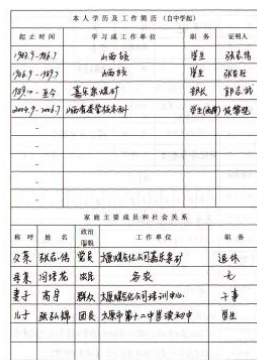


图 2 上下布局表格档案

选取完成数据集档案图片,由于要检测档案图像中的手写体以及印刷体这两种类型的目标对象,接下来的步骤是对档案图片中的印刷体和手写体文字进行标记。这里使用图像标注工具 labeling,将图片中的手写体汉字标记为 HWC (代表手写字符),而将印刷体汉字标记为 PC (代表印刷字符)。接着便可以着手研究档案图像中手写体和印刷体的检测问题。

3.2 实验环境及结果

3.2.1 实验环境

本实验训练平台采用的是 NVIDIA RTX4060, Intel (R) Xeon (R) Silver 4110 CPU、12G 的显存、GPU 为技嘉 GeForce GTX2080Ti。本实验在训练模型前需进行参数调整,实验 YOLOv8 模型 YOLOv8+CA 注意力机制模型参数值是不同的,这里展示 YOLO+CA 注意力机制的模型,其对应的训练的参数配置如表 1 所示。

表 1 YOLOv8+CA 注意力机制训练参数配置

参数	作用	设置的值
weights	存放网络的初始权重	'weights/best.pt'
workers	载入数据的线程数	8
data	训练数据集的配置文件	'datasets/mydata.yaml'
epochs	规定训练轮次	100
batch	规定训练批次	64
model	用于构建网络和初始化	'v8/yolov8_CA.yaml'

3.2.2 评价指标分析

本实验针对表格类档案图像中所包含的手写体以及印刷体采用 YOLOv8 模型以及 YOLOv8 添加 CA 注意力机制的模型分别进行了检测实验,图 3 对应的是 YOLOv8 在新增加了 CA 注意力机制之后的训练指标图,包含了锚框损失 (box_loss)、分类损失 (cls_loss)、分配焦点损失 (dfl_loss) 等参数指标,相应地反映实验的准确性。

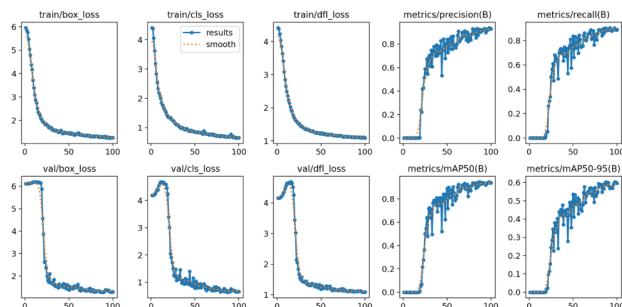


图 3 YOLOv8+CA 注意力机制模型训练指标图

分析指标图得出档案图像的训练集和验证集的锚框损失 (box_loss)、分类损失 (cls_loss)、分配焦点损失 (dfl_loss) 随着训练批次的增加,损失值明显下降,在训练即将结束时,损失值一直在某一固定值徘徊,损失值越小,训练精度越高。在上述这几个指标上添加 CA 注意力机制前后实验结果差别不大,都是最后趋于稳定,但是对应精准率 (precision) 和召回率 (recall) 参数的结果相比存在较大差异,其中在训练的前 50 个批次中,精确度 (precision) 的波动较为显著,50 次以后逐渐平缓,在训练接近尾声时,YOLO+CA 注意力机制的模型一直在 0.8 以上徘徊,相比较 YOLOv8 模型的 0.6 以上徘徊准确性得到提高; recall 参数也都是在训练批次达到 50 之前变化幅度较大,之后的变化趋势相对平缓,通过比较可以发现 YOLOv8 在增加了 CA 注意力机制之后最终的召回率也是在原模型基础上有提高的。

3.2.3 实验效果图

图 4 和图 5 分别指使用 YOLOv8 和 YOLOv8+CA 注意力机制预测的锚框,图片中的字母代表着数据的类别,数字代表置信度,置信度的高低可以作为衡量模型训练成效的指标,值越高表示模型训练的越有效。

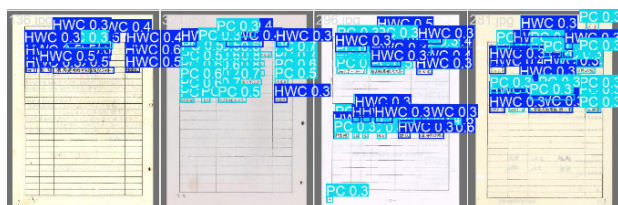


图 4 YOLOv8 模型预测锚框

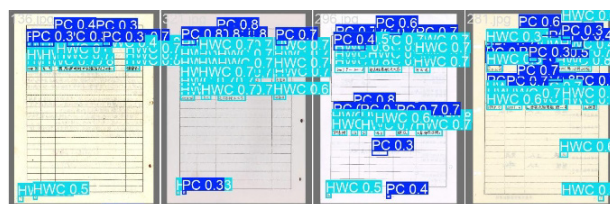


图 5 YOLOv8+CA 注意力机制模型预测锚框

通过图 4 和图 5 预测结果可以看出,使用 YOLOv8 模型检测后绝大多数区域的原始锚框中的标签都能正确检测框选出来,只是有些图片清晰度很低的地方未能检测出来,说明 YOLOv8 模型作为经典的目标检测模型的在文字检测处的应用的可行性还是比较高的。对比图 4 和图 5 可以看出增加了 CA 注意力机制之后的 YOLOv8 模型的预测锚框 PC (Printing Characters) 印刷体和 HWC (Handwriting Writing Characters) 的置信度都有提高,同时可以发现有些档案图像中在表格边缘区域的文字使用 YOLOv8 模型未能检测出来,添加了 CA 注意力机制之后检测出来了,相应的识别准确度和精度都有所提高。

4 结语

综上所述,论文提出的 YOLOv8+CA 通道注意力机制的模型对一千张档案图像数据集中的文字进行检测,首先对档案图像进行数据集的标签标注,然后再调整相关的训练图像的模型代码以及参数。最后在训练中调整到合适的模型并加以验证,这种方法在数据集上的 mAP 达到了 0.93,相比于基于 YOLOv8 模型的文字检测提高了 4%~6%,实验表明,使用 YOLOv8+CA 注意力机制模型进行文字检测,在对于不同类别的文字时达到了实用的同时也提升准确度。以此为基础,增加数据的数量以及复杂度,可进一步研究在纯文本行文档图像以及包含印章的复杂图像的印刷体与手写体检测。

参考文献

- [1] 冯少贤.基于深度学习的OCR检测与识别技术研究[D].北京:北方工业大学,2023.
- [2] 牛旭.面向手写档案数字化的汉字分类与识别研究[D].郑州:中原工学院,2022.
- [3] 韩强.面向小目标检测的改进YOLOv8算法研究[D].吉林:吉林大学,2023.
- [4] 马明扬,宋淑彩,赵一航,等.基于YOLOv8的车辆类型识别研究[J].河北建筑工程学院学报,2024,42(1):247-252.
- [5] Yu X, Liu J, Lu Y, et al. Early diagnosis of Alzheimer's disease using a group self-calibrated coordinate attention network based on multimodal MRI[J]. Scientific Reports,2024,14(1):24210-24213.