

基于YOLOv10的事故调查报告版面分析方法研究

潘令宇¹, 周子翔¹, 张云雷^{2*}, 武文星³

¹华北科技大学计算机科学与工程学院, 北京

²河北省物联网监控技术创新中心, 北京

³青海师范大学计算机学院, 青海 西宁

收稿日期: 2025年12月16日; 录用日期: 2026年1月15日; 发布日期: 2026年1月23日

摘要

近年来, 随着自然和人为灾害的频发, 事故调查报告日益彰显其重要性。针对事故调查报告版面布局复杂、元素种类繁多且尺度差异大, 导致现有模型在关键信息提取中出现检测框重叠和小尺度元素识别困难等问题, 本文提出了一种基于改进YOLOv10的事故调查报告版面分析方法。首先, 通过收集整理各省应急管理部发布的文档, 构建了包含2500张图像的事故调查报告专题数据集, 并定义了23种细粒度布局元素标签以增强语义区分度。其次, 在YOLOv10主干网络中引入GL-CRM模块, 通过动态分配计算资源, 增强模型对不同尺度目标的特征提取能力。同时, 利用YOLOv10的无非极大值抑制(NMS-free)策略, 减少检测框重叠并提高推理速度。实验结果表明, 改进后的模型在专题数据集上的F1分数和mAP分别达到了87.72%和88.5%, 相较于基线模型分别提升了约4%和7%。验证了该方法在事故调查报告文档智能化处理中的有效性和优越性。

关键词

事故调查报告, 版面分析, YOLO, 文档智能

Research on Accident Investigation Report Layout Analysis Method Based on YOLOv10

Lingyu Pan¹, Zixiang Zhou¹, Yunlei Zhang^{2*}, Wenxing Wu³

¹School of Computer Science and Engineering, North China University of Science and Technology, Beijing

²Hebei Internet of Things Monitoring Technology Innovation Center, Beijing

³College of Computer Science, Qinghai Normal University, Xining Qinghai

Received: December 16, 2025; accepted: January 15, 2026; published: January 23, 2026

*通讯作者。

文章引用: 潘令宇, 周子翔, 张云雷, 武文星. 基于YOLOv10的事故调查报告版面分析方法研究[J]. 计算机科学与应用, 2026, 16(1): 305-316. DOI: 10.12677/csa.2026.161025

Abstract

In recent years, with the frequent occurrence of natural and man-made disasters, the importance of accident investigation reports has become increasingly apparent. To address the challenges of complex layouts, diverse element types, and significant scale variations in accident investigation reports—which lead to overlapping detection boxes and difficulties in recognizing small-scale elements with existing models during key information extraction—we propose a layout analysis method based on an improved YOLOv10. First, by collecting documents published by provincial emergency management departments, we construct a specialized dataset for accident investigation reports containing 2500 images, and 23 fine-grained layout element labels were defined to enhance semantic discrimination. Second, a GL-CRM module is introduced into the YOLOv10 backbone network. By dynamically allocating computational resources, this module enhances the model's feature extraction capabilities for targets of different scales, thereby resolving the issue of feature loss in fine-grained elements. Additionally, utilizing YOLOv10's NMS-free strategy, a dual-head architecture (comprising one-to-many and one-to-one detection heads) is adopted to reduce detection box overlap and improve inference speed. Experimental results indicate that the improved model achieved an F1 score of 87.72% and an mAP of 88.5% on the self-constructed dataset, representing improvements of approximately 0.04 and 0.7, respectively, compared to the baseline model. These results validate the effectiveness and superiority of the proposed method in the intelligent processing of accident investigation report documents.

Keywords

Accident Investigation Report, Layout Analysis, YOLO, Document Intelligence

Copyright © 2026 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

当事故发生之后，人们需要通过编写调查报告的方式去总结事故发生的原因、造成的损失等关键信息。尤其是在受到极端天气和高度工业化的影响下，自然和人为灾害频发的当今社会，从先前的事故调查报告提取到的关键信息对于寻找和修补安全隐患有着重大的参考意义[1]。然而，由于时间跨度长，事故调查报告内容的格式与载体变化幅度较大，尽管当下已经完成了事故调查报告载体的数字化，文档化的普及，在提取关键数据上，依然需要人工操作。伴随着人工智能领域上的技术突破，利用文档智能代替人工进行关键信息提取也成为了可能[2]。信息提取是文档智能任务的一部分，文档解析不仅涉及了基本的图像处理技术，还有复杂的版面分析。版面分析通常被视为对文档内容按照不同类型进行拆解和分类，在流程上被认为是信息提取任务的前置工作。高质量的版面识别不仅可以降低噪音对后续工作的影响，还可以降低传播时发生错误的概率，从而实现高质量的数据提取。因此，对文档版面分析方法的研究是实现事故调查报告智能化的重要步骤。

随着深度学习的发展，目标检测模型也随之不断的迭代革新。其中，YOLOv10 [3]就是其中的佼佼者。该模型通过OCR(光学识别)技术来检测和识别文档中的布局元素。现有的实验已经证明该模型在处理学术论文时具有不错的表现，然而在进行面向事故调查报告的版面分析任务时，使用基线模型会出现检测框重叠和小尺度元素无法正确识别的问题。究其原因主要是由于事故调查报告具有布局元素种类多和元

素之间尺度差异大的特性, YOLOv10 的 backbone(主干网络)无法有效地提取元素特征, 从而造成版面元素漏分, 错分的状况出现。为了解决上述问题, 提出了 GL-CRM [4] 用于改进主干网络的特征提取能力。GL-CRM 模块可以使主干网络在对不同尺度的目标进行特征提取时动态分配计算资源, 使得模型在进行特征提取和融合时对不同尺寸元素实现自适应增强, 从而提高模型执行细粒度作业时的准确性和鲁棒性。该方法可以有效地完成事故调查报告的版面分析任务, 对于文档智能化的实现具有实际的推动作用。

本文的贡献在于通过采集各省应急管理部门网站公布的事故调查报告, 构建事故调查报告专题数据集并使用 GL-CRM 模块改进后的 YOLOv10 模型进行训练。在此基础上用该模型对事故调查报告进行版面分析。该方法提高了 YOLOv10 的准确性和泛用性, 同时也为 DLA (文档版面分析) 提供一个新方法。

2. 相关工作

在 DLA 数据集方面, 如图 1 所示, PubLayNet [5] 包括了来自 PubMed centra 中的 10 万页开放学术论文, 标注了五种核心区域类型, 适用训练通用版面分割模型; DocLayNet [6] 涵盖六大类别的专业文档(金融报告、技术手册、科学文章、法律法规、专利和政府招标书)并且提供了 11 种布局元素; DocLayNet-v2 [7] 在 DocLayNet 基础上进行了精简并将布局元素增加至 17 种; ICDAR Table Competition [8] 是竞赛专业数据集, 扩大数据来源的范围, 包含扫描件、高清印刷件等, 目前已成为评估表格分析算法性能的重要基准; ReceiptHero [9] 和 ResumeNet [10] 的数据来源主要是简历文档, 用于执行特定的版面分析任务。总体而言, 当前的文档布局分析数据集在数量上可以满足绝大部分的训练需求, 但是在专业性和细粒度方面依然存在明显的局限性。

在基于 YOLO 框架的 DLA 模型方面, YOLO 将 DLA 任务视为一种特殊的目标检测问题, 通常使用通用检测器(YOLOv8 [11]、YOLOv9 [12])进行检测。在大部分 DLA 任务中, 现成的检测器可以通过 DLAs 数据集的训练来完成得到一个不错的结果。但在进行细粒度作业时, 受制于 YOLO 的检测方式, 会出现特征丢失从而影响后续识别的准确率[13]。

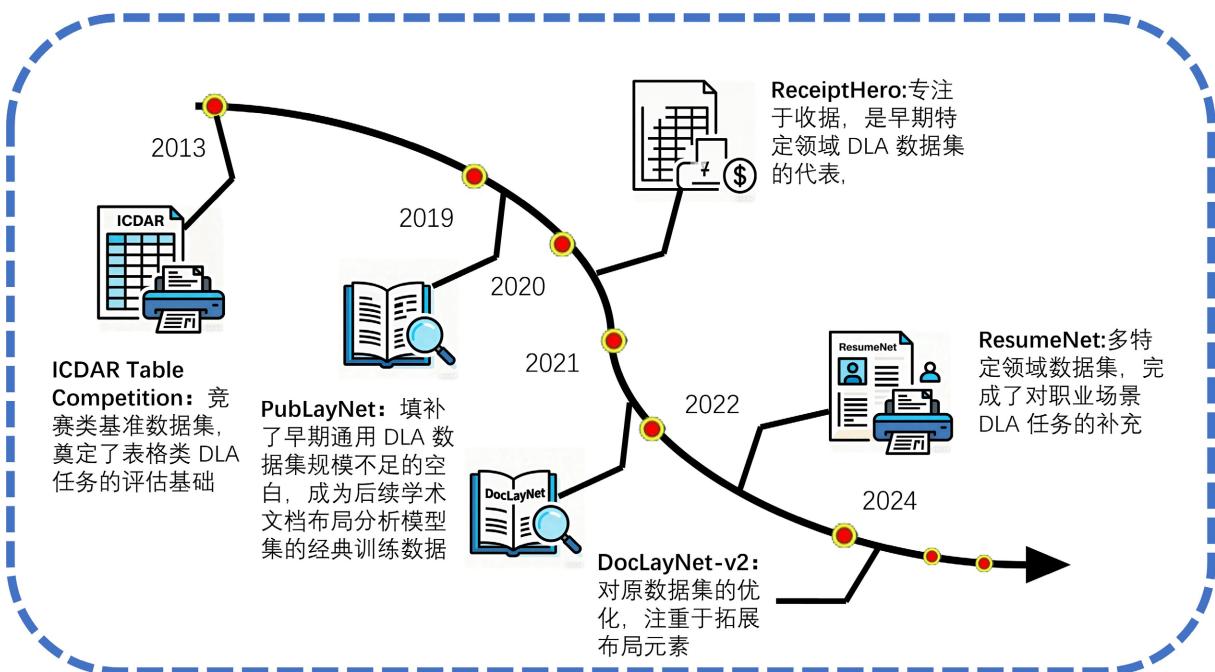


Figure 1. The development of datasets

图 1. 数据集的发展

3. 数据预处理

本文依托于中国安全生产大数据平台,从应急管理部、各地应急管理部门网站收集的6000余份灾害事故调查评估相关文档,涵盖20余类灾害事故类型与5类文档格式。随后按照真实性、完整性、多样性、典型性这四大原则,我们对这些文档进行了筛选。此外,本文还从技术层面进行了要求,其一,要求筛选的文档必须尽量包含复杂版面,这是构建专用数据集的核心需求,其二,要求尽量选取页面数量不少于15页的文档,先前的数据集构建工作表明这样的文档具有更高的信息价值[14],符合本文的研究目的。最后,对筛选后的文档进行预处理工作,通过去重、统一格式和分辨率、降噪和裁剪页边空白,最终获得2500张用于标注的灾害事故调查评估数据图像样本,如图2所示,数据集包含标题区4216个、正文区18,952个、表格区5836个、图片区4568个、公式区1236个、签名盖章区1528个、附件说明区2298个。

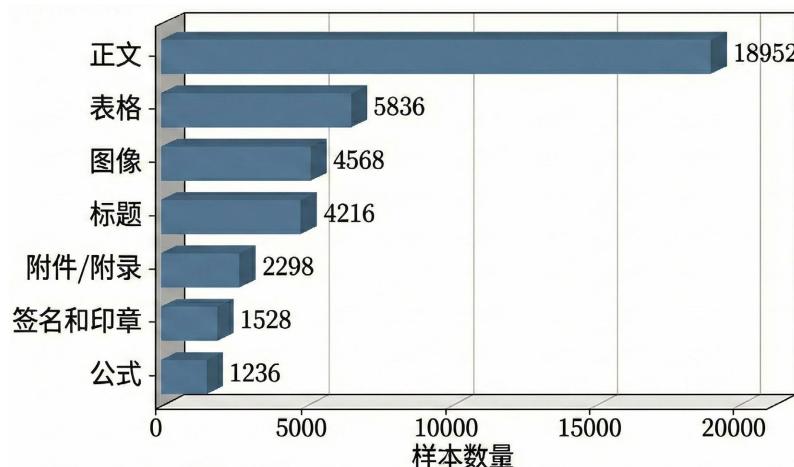


Figure 2. Statistical distribution of annotated semantic regions in the dataset
图2. 数据集中标注语义区域的统计分布

3.1. 版面元素分类

在标签选择上,在参考了先前的工作的基础上,本文定义7大类、16小类核心信息区域。这23中类别标签的定义表1所示。

Table 1. Definition of dataset layout elements

表1. 数据集布局元素定义

编号	类型名称	类型定义
1	Formula	公式, 通常为数学公式或科学表达式。
2	Figure & Figure Caption	图及其说明文字(图和图注一起)。
3	Image	图片、插图等视觉内容。
4	Image Caption	图片的说明文字(图片下方的描述)。
5	Cover Page	封面页, 文档的首页或封皮。
6	Document Information	文档信息, 如作者、日期、机构等元数据。
7	File Name	文件名。
8	Heading 1	一级标题, 章节或主要部分的标题。

续表

9	Heading 2	二级标题, 章节下的子标题。
10	Heading 3	三级标题, 更细一层的子标题。
11	Body Paragraph - Full	正文段落(完整)。
12	Body Paragraph - No Header	正文段落(无头)。
13	Body Paragraph - With Whitespace	正文段落(带有空白区域, 可能有分隔)。
14	Body Paragraph - Unfinished	未完成的正文段落(如被截断或遗漏部分)。
15	Table of Contents	目录, 列出文档结构和页码。
16	Signature	签名, 通常为手写签名或电子签名。
17	Footnote Content	页脚内容, 补充说明或参考信息。
18	Footnote Symbol	页脚符号(如*、†等, 标识脚注位置)。
19	Table & Table Caption	表格及其标题(表和表注一起)。
20	Table	表格内容。
21	Table Caption	表格标题或说明文字。
22	General Title	通用标题(如文档总标题)。
23	Page Number	页码。

粗粒度分类会导致标注信息泛化过度, 使得不同功能的元素被归为一类, 影响模型学习到元素的语义差异。本文将布局元素进行了更为细致的分类, 提高标注数据能被模型识别到的特征差异, 从而减少模型无法精确识别正确布局元素的可能性; 同时, 细粒度的元素划分还能增强模型的泛化能力和场景适配性, 减少模型对非常规布局的鲁棒性不足问题。在实际模型训练时, 细粒度分类可以让标注错误得到更好的排查, 减少资源和时间的浪费。

总之, 本文定义并选择这样的布局元素, 可以体现灾害事故图文数据表达特征并实现将灾害事故图文数据进行结构化拆分。

3.2. 标注方式选择

在图文数据标注技术领域, 已形成人工标注、半自动化标注、全自动标注三类主流方式[15]。考虑到算法自动标注目前还无法胜任细粒度标注任务, 因此本文的数据集构建采取人工标注方式。人工标注, 尽管在效率上存在短板, 但在准确性、适应性与对复杂数据处理能力上的优势, 完全匹配对高精度的核心需求。在标注工具的选择上, 本文选择 labelImg[16]作为该项任务的标注工具, 其核心优势在于操作极简, 支持边界框快速绘制与自定义类别预设, 在核心标注功能、操作效率、格式兼容性上的表现完全适配任务需求的同时, 还可以自动导出 YOLO 模型可直接识别的 PASCAL VOC XML 格式, 方便后续进行模型训练任务。

3.3. 数据集的构建

考虑到人工标注的效率和协作上与自动化标注相比存在短板, 将标注人员进行合理分组分工和采用统一归档办法可以有效提高效率, 减少标注错误。本文标注流程如图 3 所示: 工作人员共计 8 人, 每人分配 300 余页样本进行第一次标注。之后每两人一组, 通过交叉核对, 标记异议的方式进行检验, 由专门负责人进行最终判定并修改标注文件。最后对所有标注文件进行格式效验和随机抽查, 检验合格后分类归档。这样的流程可以最大限度减少数据标注这一环节的问题, 保证数据集的质量。



Figure 3. Flowchart of dataset construction

图 3. 数据集构建流程图

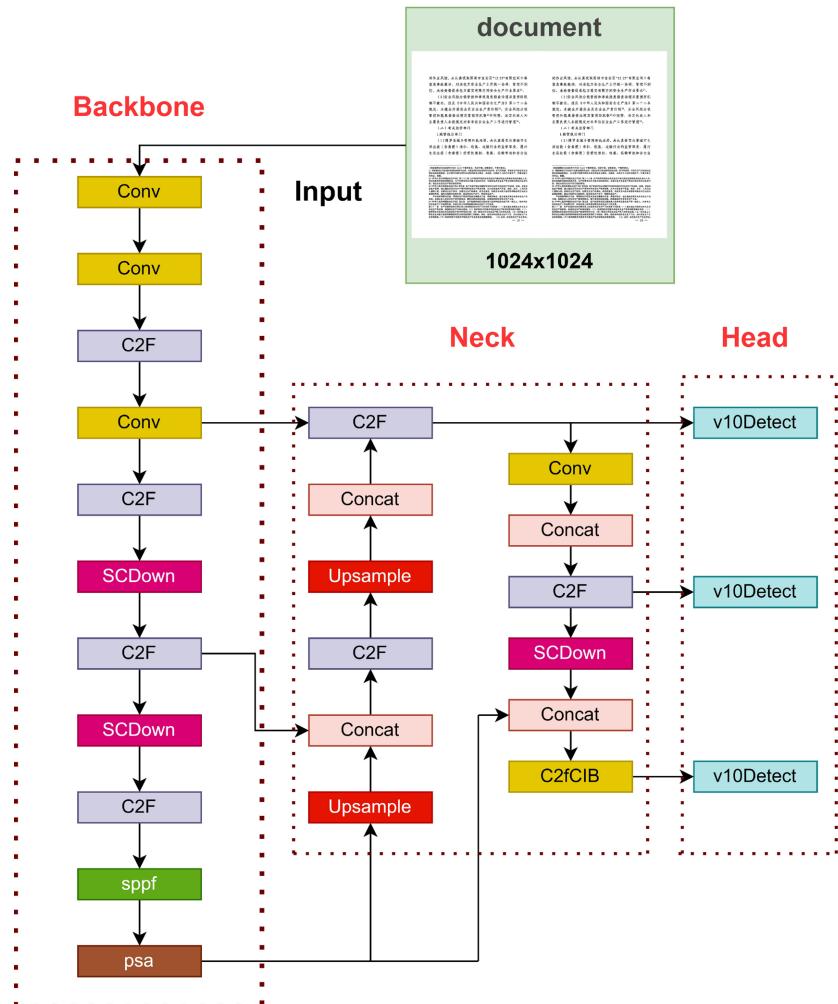


Figure 4. YOLOv10 process structure diagram

图 4. YOLOv10 流程结构图

4. 模型构建

在模型的选择上, 为应对 DLA 任务中存在的挑战, 本文选择 YOLOv10 作为模型基底。该模型是在 YOLOv8 的基础上进行创新的。如图 4 所示, 在整体工作流程上, YOLOv10 与该系列先前模型类似, 主要由 backbone (主干网络) 进行特征提取、neck (颈部) 进行特征融合和 head (检测头) 进行后处理三个主要步骤组成。其主要在 backbone 使用了新的卷积模块来实现轻量化, 以及在后处理阶段采取无 NMS 的策略用于减少模型在后处理过程中的资源使用。YOLOv10 相对轻量级的 backbone 可以进行模块的增加, 为此本文引入了 GL-CRM 模块, 本文使用 GL_CRM 模块来替换 backbone 中 C2F 模块, 该模块在继承原先 CSP [17] 模块作用的同时针对文档图像进行了改进, 使模块可以获得更加全面的上下文感知, 从而在进行特征提取时避免检测细粒度元素时出现特征丢失。

4.1. GL-CRM 模块

如图 5 所示, 基线模型的 C2F 模块是将输入的特征张量分为两条支路: 直连分支和主分支。直连分支感受野小, 用于负责提取纹理这样的细节特征, 记为 Y_1 。主分支中通过 bottleneck 模块扩大感受野, 检测范围可以覆盖更多的区域, 从而提取目标的语义特征和全局特征。经过的 bottleneck 模块越多, 感受野覆盖的范围也越大, 记为 $[Y_2, Y_3, \dots, Y_{n-1}]$ 其中 n 为经过的 bottleneck 模块的个数。最后 concat 进行特征的拼接, 形成一组不同粒度的特征 $Y = [Y_1, Y_1, Y_2, \dots, Y_{n-1}]$ 。使用这种方法使得 YOLO 在自然图像的检测中可以结合全局关系来提取多尺度特征, 由此来实现自然图像的精确检测。

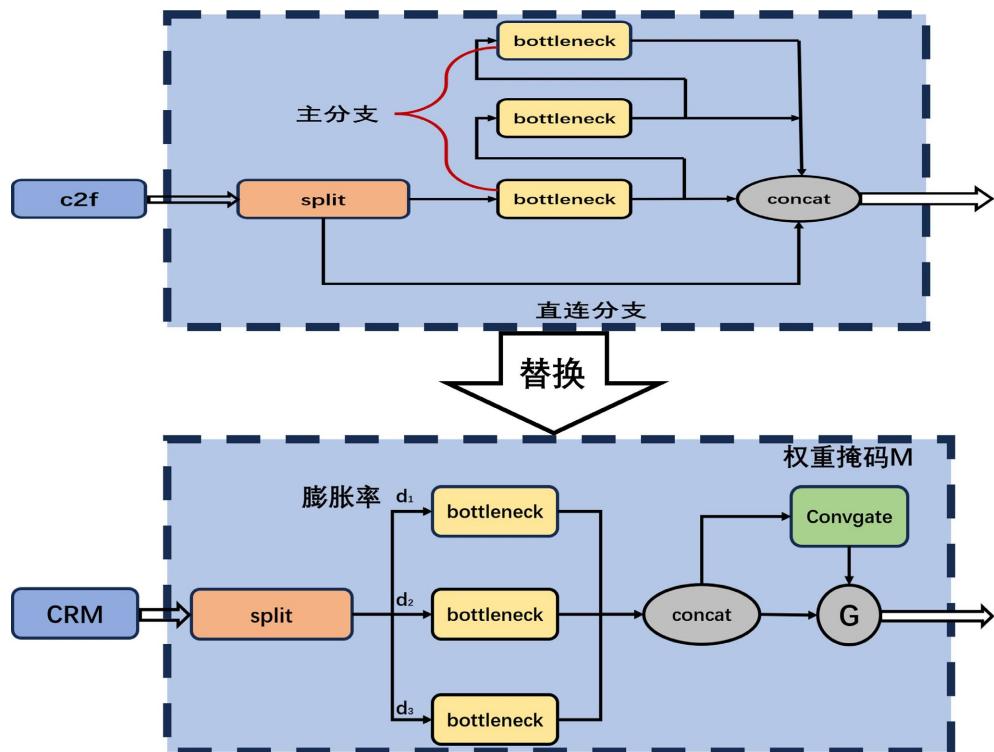


Figure 5. Comparison of workflow structures between C2F and CRM modules
图 5. C2F 和 CRM 模块工作流程结构对比

然而在 DLA 任务中, 文档与自然图像的不同之处在于布局元素, 不同布局元素尺度差异过大。例如在检测页码和占据一半页面的大型图表时, 检测页码这样的小目标特征时会更加依赖注重细节的局部层;

在检测图表特征时会更依赖注重语义结构的全局层。在传统的 C2F 的支持下只能让模型进行多尺度特征提取时采取固定且有限度的感受野适配。此外，在检测大型图表时，由于元素的尺寸依旧占据一半页面的尺寸，需要堆叠庞大数量的 bottleneck 模块才能满足需要。因此使用 C2F 模块进行多尺度特征提取时远远无法满足需要的。

本文引入 GL_CRM 来解决这项问题，如图所示，相较于 C2F 模块通过堆叠 bottleneck 模块来实现感受野的扩大，CRM 使用了一组超参数 $d = [d_1, d_2, \dots, d_n]$ 和距离参数 k 来控制提取不同尺度元素时内核元素间距和内核尺寸，这使得感受野覆盖范围能符合布局元素的同时，

精简了原先的链式架构。表示为 $F = [F_1, F_2, \dots, F_n]$ ，同时考虑到不同的尺寸的元素对于不同尺度特征的偏重不同，本文使用一个 $1*1$ 轻量级卷积层和一组 nC 来提取一组值在 0 到 1 之间的掩码 M，用来反应不同尺度的特征对于布局元素的重要程度。

4.2. 无 NMS 策略

先前的 YOLO 通常是利用 NMS (非极大抑制) [18] 来进行后处理的，该方法通过对目标置信度进行排序来选出最优检测框。这在目标检测中可以提高模型检测的精度并且降低后处理的成本，然而在 DLA 任务中，由于检测的布局元素数量相对于常规目标检测有所增加，且理想效果是实现版面元素和检测框对应输出。NMS 的处理模式会使得输出结果出现检测框重叠的现象，从而造成模型在进行 DLA 任务时训练推理结果不理想。虽然先前的工作通过对 NMS 进行加权或者更改 IOU (交并比) 的方式来解决这个问题，但这样的解决方式会使得模型增加格外算力用于后处理阶段。增大模型的体积也会影响在 backbone 阶段的特征提取能力，使得模型在训练和实际运用环境受到限制。

YOLOv10 的后处理采用的是无 NMS 方法，如图所示，具体来说设计了两种检测头：一对一检测头和一对多检测头。它们采取相同的 TAL [19] 分配原则，即

$$m(\alpha, \beta) = s \cdot p^\alpha \cdot \text{IoU}(\hat{\beta}, b)^\beta \quad (1)$$

其中 p 为分类分数， b 和 \hat{b} 分别表示预测和实例的边界框。 s 表示空间先验，表示预测的锚点是否在实例内 [19]。 α 和 β 是两个重要的超参数，可以平衡语义预测任务和位置回归任务的影响。

如图 6 所示，在训练阶段，两个检测头共同参与，利用一对多检测头提供的丰富监督信号提高收敛速度。在推理阶段，则单独启用一对一检测头进行结果的输出。这减少了模型在后处理阶段的资源使用，同时提高了推理速度和准确率。

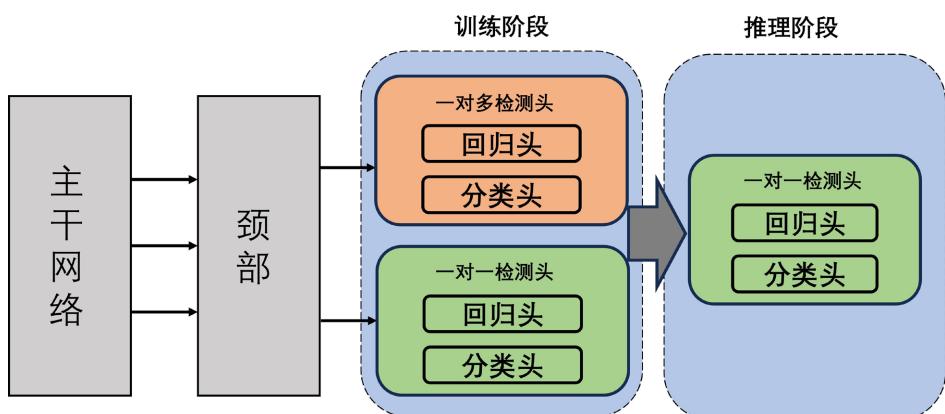


Figure 6. NMS-free detection strategy process

图 6. 无 NMS 检测策略流程

5. 实验

为了保证实验结果可信，本文选用了 YOLOv10m, YOLOv10x 和本文的模型在数据集上进行性能对比测试，为了保证对比的公平性，我们在这些模型的预训练基准上进行了微调，来保证在版面分割任务中获得最有效的结果。

5.1. 评估指标

实验使用 precision, map 和 fps 来验证数据集的有效性，使用 F1 分数和 map 来测量模型该数据集上进行版面分析任务的性能。其中，precision 反应模型识别区域类别的正确率；fps 反应模型的推理速度；F1 分数反应模型进行分类任务的总和性能；mAP 数值反映了模型在多类别目标的定位精度、类别识别准确性及应对复杂场景能力的综合表现。这些数值的计算公式如图所示，其中 TP、FP、FN、K、AP_k 分别代表真阳性，假阳性，假阴性，类别总数和第 k 类的平均精度。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{MAP} = \frac{1}{K} \sum_{k=1}^K \text{AP}_k \quad (5)$$

5.2. 模型训练

所有模型均在相同的超参数设置下进行训练：初始学习率为 0.01，权重衰减为 0.0005，采用自动选择的优化器。训练过程最长持续 500 个 epoch，并设置了早停机制，当连续 100 个 epoch 验证集损失未提升时终止训练过程。为降低收敛到局部最优的风险，训练初期增加了缓冲，确保模型至少完成 100 轮迭代。训练批量大小设为 32，输入图片尺寸为 1024，训练过程中每轮均在验证集上评估模型性能。为应对数据中类别分布不均的问题，模型采用类别加权的损失函数，每一类别的权重与其在数据集中的频率成反比，从而引导模型更多关注样本较少的类别，减缓类别不平衡对检测性能的影响。所有实验均在 8 线程环境下进行，训练过程保存全部模型权重及相关指标，用于后续分析和比较。数据集的划分为 7:2:1。其中 70% 为训练集，20% 为验证集，10% 为测试集。

5.3. 主要结果

Table 2. Performance comparison of various models on the dataset

表 2. 各模型在数据集上性能比较

模型	F1	map
yolo10m (our)	0.8772	0.885
yolov10x	0.8332	0.884
yolov10m	0.8324	0.791
yolov13n	0.8136	0.859

经过测试实验结果如表 2 所示，发现改进后的模型在相较于基线模型在 F1 和 MAP 这两个指标均有提升，在 MAP 和 F1 上提高了约 0.07 和 0.04，这表明本文提出的模型在综合性和检测小目标上优于基线

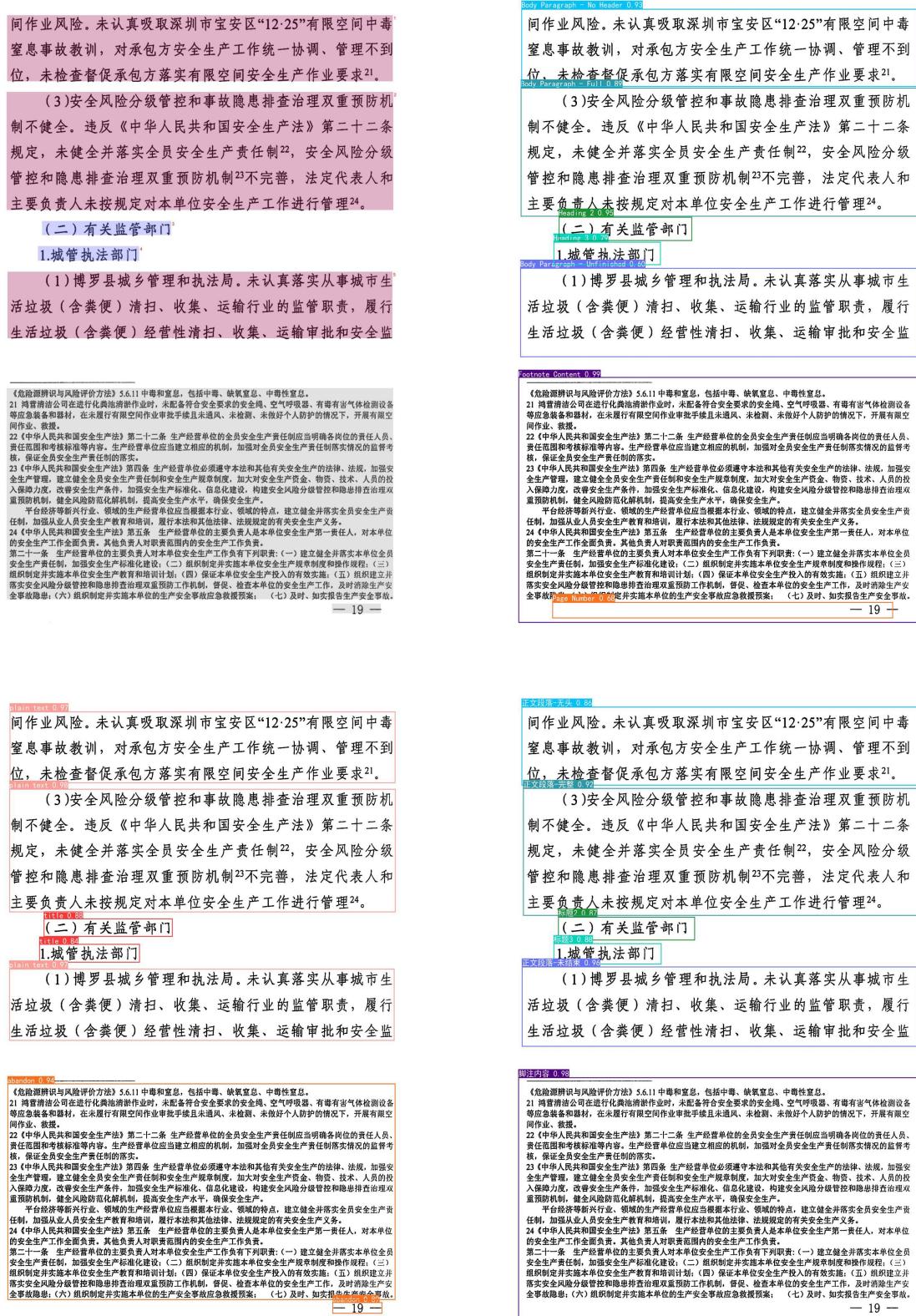


Figure 7. Visualization results of model inference (in the top left, top right, bottom left, and bottom right respectively YOLO10m (our), YOLOv10m, YOLOv10x)

图 7. 各模型推理可视化效果图(左上,右上,左下,右下分别为原版、YOLO10m(our)、YOLOv10m、YOLOv10x)

模型，并且也略高于目前最新框架的 YOLOv13。同时也注意到改进模型的性能也略高于具有更大参数的 YOLOv10x，这可能是因为数据集包含的样本数量无法充分的训练 YOLOv10x 的 backbone 所导致的性能下降。尽管如此，本次实验依然证明了模型改进的有效性，能够更好的在复杂版面的文档上进行 DLA 任务。图 7 展示测试模型在同一文档页面中的检测效果。

5.4. 消融实验

最后对本研究所提出的 GL-CRM 模块进行消融实验，使用相同参数的 YOLOv10m 进行，在本文提出的数据集上进行测试。如表 3 所示，实验结果表明，增加了 GL-CRM 模块的模型在宽松条件下检测精度相比较于基线模型有显著的提高。实验验证了该模型设计的有效性。

Table 3. Performance comparison of various models on the dataset
表 3. 各模型在数据集上的对比表现

模型	mAP50
YOLOv10-m (our)	0.885
YOLOv10-m (baseline)	0.791

6. 总结和未来工作

本文介绍一项专门用于事故调查报告的数据集，它有效解决了事故调查报告数据集不足的问题，同时也对该类型的数据进行更专业的构建，具体表现为重新划分了布局元素类型，已满足实际问题的需要，这两方面都是至关重要的。

本文提出的 DLA 模型在专题数据集上展现了良好的性能，对比基线模型有了较大的提升，即使和拥有较大参数量的 YOLOv10x 相比依然展现出微弱的性能优势。作为未来工作的一部分，我们计划通过引入更多不同来源的事故调查报告来拓展数据集，以进一步扩大数据集的异质性和泛用性。可能的想法是补全各省市的事故调查报告文档和公司内部调查文档。在模型层面，将优化模型参数和结构，以提高模型的检测能力和性能。

基金项目

本研究得到了国家重点研发计划(项目编号：2024YFC3016801、2024YFC301805)和中央高校基本科研业务费专项资金(批准号：3142021009)的资助。

参考文献

- [1] Zhang, Y., Wu, B., Ning, N., Song, C. and Lv, J. (2019) Dynamic Topical Community Detection in Social Network: A Generative Model Approach. *IEEE Access*, **7**, 74528-74541. <https://doi.org/10.1109/access.2019.2921824>
- [2] Ke, W., Zheng, Y., Li, Y., Xu, H., Nie, D., Wang, P., et al. (2025) Large Language Models in Document Intelligence: A Comprehensive Survey, Recent Advances, Challenges, and Future Trends. *ACM Transactions on Information Systems*, **44**, 1-64. <https://doi.org/10.1145/3768156>
- [3] Wang, A., et al. (2024) YOLOv10: Real-Time End-to-End Object Detection. *Conference on Neural Information Processing Systems*, Vancouver, 9-15 December 2024, 107984-108011.
- [4] Zhao, Z.Y., Kang, H.R., Wang, B. and He, C.H. (2024) DocLayout-YOLO: Enhancing Document Layout Analysis through Diverse Synthetic Data and Global-to-Local Adaptive Perception. *Computing Research Repository*.
- [5] Zhong, X., Tang, J. and Jimeno Yepes, A. (2019) PubLayoutNet: Largest Dataset Ever for Document Layout Analysis. *2019 International Conference on Document Analysis and Recognition (ICDAR)*, Sydney, 20-25 September 2019, 1015-1022. <https://doi.org/10.1109/icdar.2019.00166>

-
- [6] Pfitzmann, B., Auer, C., Dolfi, M., Nassar, A.S. and Staar, P. (2022) DocLayNet: A Large Human-Annotated Dataset for Document-Layout Segmentation. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Washington DC, 14-18 August 2022, 3743-3751. <https://doi.org/10.1145/3534678.3539043>
 - [7] Livathinos, N., *et al.* (2025) Docling: An Efficient Open-Source Toolkit for AI-driven Document Conversion. Computing Research Repository.
 - [8] Gao, L.C., *et al.* (2019) ICDAR 2019 Competition on Table Detection and Recognition (cTDAr). *IEEE International Conference on Document Analysis and Recognition*, Sydney, 20-25 September 2019, 1510-1515.
 - [9] Yu, J.-M., Ma, H.-J. and Kong, J.-L. (2025) Receipt Recognition Technology Driven by Multimodal Alignment and Lightweight Sequence Modeling. *Electronics*, **14**, Article No. 1717. <https://doi.org/10.3390/electronics14091717>
 - [10] Luo, Y., Zhang, H., Wang, Y., Wen, Y. and Zhang, X. (2018) ResumeNet: A Learning-Based Framework for Automatic Resume Quality Assessment. 2018 *IEEE International Conference on Data Mining (ICDM)*, Singapore, 17-20 November 2018, 307-316. <https://doi.org/10.1109/icdm.2018.00046>
 - [11] Varghese, R. and M., S. (2024) YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. 2024 *International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, 18-19 April 2024, 1-6. <https://doi.org/10.1109/adics58448.2024.10533619>
 - [12] Wang, C.-Y., Yeh, I.-H. and Mark Liao, H.-Y. (2024) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. Computing Research Repository.
 - [13] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/cvpr.2016.91>
 - [14] Peña, A., Morales, A., Fierrez, J., Ortega-Garcia, J., Puente, I., Cordova, J., *et al.* (2024) Continuous Document Layout Analysis: Human-in-the-Loop AI-Based Data Curation, Database, and Evaluation in the Domain of Public Affairs. *Information Fusion*, **108**, Article ID: 102398. <https://doi.org/10.1016/j.inffus.2024.102398>
 - [15] Zottin, S., *et al.* (2024) U-DIADS-Bib: A Full and Few-Shot Pixel-Precise Dataset for Document Layout Analysis of Ancient Manuscripts. *Neural Computing and Applications*, **36**, 11777-11789.
 - [16] Ilani, M.A. and Banad, Y.M. (2025) LabelImg: CNN-Based Surface Defect Detection.
 - [17] Wang, C., Mark Liao, H., Wu, Y., Chen, P., Hsieh, J. and Yeh, I. (2020) CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, 14-19 June 2020, 1571-1580. <https://doi.org/10.1109/cvprw50498.2020.00203>
 - [18] Hosang, J., Benenson, R. and Schiele, B. (2017) Learning Non-Maximum Suppression. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6469-6477. <https://doi.org/10.1109/cvpr.2017.685>
 - [19] Esser, P., Rombach, R. and Ommer, B. (2021) Taming Transformers for High-Resolution Image Synthesis. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 12868-12878. <https://doi.org/10.1109/cvpr46437.2021.01268>