

深度学习方法在工程图纸标识定位与识别中的应用研究

肖鑫¹, 陈青松¹, 吴思远², 胡瑞¹

(1. 中冶赛迪工程技术股份有限公司数字化中心, 重庆 401122; 2. 重庆邮电大学 计算机科学与技术学院, 重庆 400065)

摘要: 工程图纸关键标识的定位与识别一直是计算机视觉的关键应用之一。基于深度学习的文本检测方法相较于传统方法而言拥有更高的检测效率与准确性, 将现有的文本检测算法应用到工程图纸识别任务中是有必要的。提出了工程图纸关键标识定位与识别方法, 针对工程图纸中的索引标识和尺寸标识进行位置检测和内容识别。索引标识位置检测需要把图纸进行切割统一大小, 要用非极大值抑制剔除冗余候选框。尺寸标识位置检测通过对遮蔽后的图纸进行一次完整的检测, 对每一个检测框和索引标识位置计算面积交并比以剔除部分数据。实验结果表明, 该方法对工程图纸的索引标识和尺寸标识进行位置检测和内容识别具有较高的精确率和召回率。

关键词: 工程图纸; 目标检测; 字符识别; 深度学习

中图分类号: TP391; TN911

文献标志码: A

文章编号: 1673-825X(2025)05-0696-12

Research on the application of deep learning methods in identification localization and recognition of engineering drawings

XIAO Xin¹, CHEN Qingsong¹, WU Siyuan², HU Rui¹

(1. Digital Center, CISDI Engineering Co., Ltd., Chongqing 401122, P. R. China;

2. School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China)

Abstract: The localization and recognition of key symbols in engineering drawings have long been essential applications in computer vision. Compared with traditional methods, deep learning-based text detection approaches offer higher detection efficiency and accuracy. It is therefore necessary to apply existing text detection algorithms to engineering drawing recognition tasks. This paper proposes a deep learning-based method for the localization and recognition of key symbols in engineering drawings, focusing on the detection and recognition of index symbols and dimension symbols. For index symbol localization, the drawings are cropped to a uniform size, and non-maximum suppression is used to remove redundant candidate boxes. For dimension symbol localization, a complete detection is performed on the masked drawings, and the intersection-over-union between each detected box and index symbol location is calculated to filter out partial data. Experimental results demonstrate that the proposed method achieves high precision and recall in both the localization and recognition of index and dimension symbols in engineering drawings.

Keywords: engineering drawings; object detection; character recognition; deep learning

收稿日期: 2024-08-26 修订日期: 2025-06-20 通讯作者: 胡瑞 rui.hu@cisdi.com.cn

基金项目: 国家自然科学基金项目(62302074); 重庆市教委科学技术研究项目(KJQN202300631)

Foundation Items: National Natural Science Foundation of China (62302074); Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN202300631)

0 引言

作为描述建筑、结构、机械、电气等工程的设计、尺寸、材料和施工要求等细节一种技术文件,工程图纸是工程项目建设过程中不可或缺的重要参考文件,常用于指导施工、审查设计、材料采购、成本估算、质量控制和维护等方面。解释这些图纸需要拥有专业技能的人,在某些情况下需要长时间的工作。随着制图技术的迅速发展,工程图纸的数量日益增多,如何对大量的工程图纸进行检索查询成为了迫切需要解决的难题^[1]。通过人工对比进行查找的方式流程繁琐且效率低下,长时间重复性的工作还可能导致较高的错误率。工程图纸的数字化存储以及从这些图纸中提取信息的方式,在机械、电子系统等各方面的应用领域中都具有重要的实际意义,成为了学术界和工程界研究的热点。传统的识别方法效率较低且大多只适用某一类特定图纸,近年来,随着图形处理单元(graphics processing unit, GPU)的大规模并行处理能力大幅度增强,极大地加速了深度学习所需的数学计算过程,利用深度学习方法进行文本检测逐渐引起了学术界和工程界的注意。鉴于深度学习比传统方法有着更高的检测效率和准确率,逐渐成为文本检测的主流方法。

文本检测技术最先起源于光学字符识别,它使用光学方式将包含印刷体字符的纸质文档转换为图像文件,再使用计算机对输入的文本图像进行数字化处理,利用图像中梯度等特征信息检测并识别图像中的文本内容。文本检测是计算机视觉中一项关键而基础的技术,在许多领域都很重要,包括驾驶辅助、手写识别和帮助视障人士等^[2]。基于梯度的学习算法可以用于合成复杂的决策面,只需少量预处理,该决策面就可以对高维模式(如手写字符)进行分类。Lecun等^[3]综述了用于手写体字符识别的各种方法,并在一个标准手写体数字识别任务上进行了比较。Tian等^[4]提出了一种汉字检测算法,此算法可以有效地检测水平或略微倾斜的文本行,但对于一些旋转的文本行,其检测效果一般,并且文本线的构造也是局限在矩形,当文本出现倾斜时,文本线的构造就不够精准。Zhou等^[5]提出了一种适用于自然场景中的短文本的文本检测算法,对于旋转文本行的检测效果较好,但对长文本的检测不够准确。为了对一个句子进行建模,Yin等^[6]提出了基于注意力机制的识别算法,主要应用于不规则排列的文字识

别问题。Shi等^[7]提出端到端的不定长文字识别算法,主要应用于不定长的规则排列的文字识别问题。

目前主流的场景文本识别方法主要基于卷积神经网络和递归神经网络,由于卷积注意力网络可以很容易地堆叠注意层并处理字符序列的长期依赖关系,Xie等^[8]引入了一种基于卷积的注意力网络,用于无约束场景文本识别。传统的场景文本识别大多采用完全数据驱动的基于注意力的对齐,这种做法忽略了特定的字符几何信息。Hu等^[9]基于一组可学习的几何点,提出了一种新的形状驱动的注意对齐方法,该方法能够获得字符的几何特征。场景文本识别是计算机视觉对场景图像进行分析和识别的一项基本任务,将该任务应用在手持设备中,降低模型的参数量且减轻模型的架构是不可或缺的。Ghosh等^[10]针对场景文本检测和识别提出了一种轻量级模型,从而可以在资源受限的设备上使用。基于深度学习的方法在场景文本识别中表现出了优异的性能,为了将该方法应用在小语种识别中,Alshawi等^[11]收集了一个波斯语数字数据集并提出了一个基于卷积的模型,提出了一个端到端的波斯语数字识别模型。

直接将现有的文本检测技术应用于定位与识别工程图纸关键标识是不可行的。尽管现有的深度学习方法能够很好地识别自然场景中的各种文本,但想要尽可能完整地提取出工程图纸中的重要标识仍需克服许多难点与挑战。与自然场景相比,工程图纸更加复杂,图纸中会包含许多标线、图案等,这会干扰文本检测算法的准确性。工程图纸中的文本密度高,在图纸中的某些区域文本信息会很集中,甚至是出现符号重叠交错的情况,这会严重干扰到深度学习的识别效果。不同的工程图纸会包含不同的专业符号,有时还需要针对当前任务构建字符集。但无论如何,将有着高精度识别能力且在不断发展的深度学习方法应用于自动化识别工程图纸关键标识的定位与识别都是有意义的。

文本的主要贡献在于:利用基于光学字符识别技术和图像检测技术对图纸中的索引标识以及尺寸标识进行位置和内容提取。本文在企业提供的数据集上的实验结果表明,该方法对工程图纸的索引标识和尺寸标识进行位置检测和内容识别具有较高的精确率和召回率。

1 相关工作

传统的文本检测方法需要先根据像素点的特征

获取图像中的文本候选区域,再利用手工设计的特征对候选区域进行筛选和融合,以此获得最终的文本区域^[12]。1996年,陈勇等^[13]针对二值数字化机械工程图中字符的提取和分割问题,提出了一种新的图文自动分离算法。已有的字符分离方法需要首先生成全图轮廓链码,再对每一闭合轮廓进行内、外边缘判别,以提取字符的外轮廓。为了避免重新计算各轮廓点的重复劳动、提高运行效率,胡友兰等^[14]基于规则从扫描图像全图轮廓中分离字符轮廓,提出了同步边缘特性检测的轮廓跟踪算法。使用连通域中心线定向的方法在字体形状变化较大时会产生较大的方向误差,并且不能处理和识别字串粘连、字线粘连及字线相交的情况。为了解决上述问题,邹荣金等^[15]描述了工程图纸矢量化中多向粘连字符及字线相交的分割算法与识别方法,从连通群的特征矢量出发,逐次逼近解求字串的方向,并采用纵向投影技术实施字符的分割,对字线粘连、字线相交的分割算法作了详细描述。现有的字符串提取算法对矢量化内容的特定环境考虑得较少,没有利用被矢量化对象本身的背景知识,往往很少考虑工程图纸的一些特性,仅仅从光学字符识别的角度出发去解决问题。范帆等^[16]考虑到字符串与字符串、图元与标注信息之间的逻辑联系,提出了一种基于工程图纸知识的预分割字符串及标注信息提取方法。

传统的基于特征的检测方法效率较低,同时标注信息的提取是字符与图元相结合的结果,基于深度学习的方法可以取得更好的效果。随着计算机硬件的快速发展,将深度学习应用于工程图纸文本检测成为了新的研究热点。依靠强大的学习和特征表达能力,基于深度学习的文本检测方法不需要依赖大量拥有专业知识的人进行手动处理;在需要快速响应的环境中,基于深度学习的方法有着压倒性的优势;相较于发展速度缓慢的传统方法而言,深度学习迭代更新的速度很快,可以利用当下发展的新技术进行调整维护。基于深度学习的文本检测方法在检测精度和鲁棒性上都远胜传统方法。杜建强等^[17]根据轮廓研究出一种对图文进行分离的方法,能够从工程图纸上分离识别出字符信息,但是并不能适应于字符图形文字互相粘连的情况。1997年,蔡良伟等^[18]根据工程图纸中字符的特点,定义了字符轮廓的长宽比、形状复杂度、内孔数和区域密集度,提出了基于这些轮廓特征的字符分离方法,且将

神经网络应用于工程图纸的字符识别。由于神经网络分类器可以任意精度逼近任意形状边缘,杨万山等^[19]利用反向传播神经网络(back propagation neural network, BPNN),提出了一种识别工程图中图形符号的方法,但该方法学习速度慢,对样本有依赖性,且容易过拟合。为了正确识别工程图纸中拥有着各种空间特征的符号,董海卫等^[20]提出了一种统计识别和结构识别相结合的符号识别方法,以提高识别正确率。为了克服BPNN容易陷入局部极小值的缺点,翟传敏等^[21]基于径向基概率神经网络提出一种扫描工程图纸图像分割后的图形符号识别方法。宋晓宇等^[22]构建工程图纸中常用图形对象的知识库,在后续图形对象识别过程中利用该图库中存储的图形对象的特征信息标记其他待识别工程图纸中相似度较高的图形对象。实验表明,基于拓扑结构的工程图纸识别方法在前期图元矢量化的基础上能够准确高效地进行二维图形对象的识别。单世哲等^[23]采用Matlab函数库训练了BPNN结构模型对电气图形符号进行分类,实验存在的问题是图像的噪声会影响模型的识别准确度。

受限于硬件能力,上述模型达成的效果并不算理想。随着计算机技术高速发展,设计神经网络训练模型进行工程图纸文本检测再次引起了研究者的注意。为了提升银行、邮局等场合的实时数字识别准确率,代贺等^[24]以Lenet-5模型为基础改进了卷积神经网络结构并推导了改进后的前向和反向传播算法,在不同数据集上的测试结果表明,该模型识别性能明显高于传统网络结构。在工程图纸中,一个典型的图表通常包含大量属于不同类别的不同类型的符号,某些类型的符号在数据集中占主导地位,而其他类型的符号在数据集中几乎没有表示。为了解决上述问题,Elyan等^[25]提出了一种用于工程图中符号定位和识别的边界盒检测方法,并利用深度生成对抗神经网络的类不平衡,成功改善了工程图中符号的分类问题。

现有的工程图纸图像识别方法大多采用BPNN来实现对图纸上信息的提取,具有受限于工程图纸种类多、符号多样性高、复杂度高等特点,特征提取效果较差。为了更好地进行特征提取,陈玉辉^[26]采用卷积神经网络来完成工程图纸图像识别任务,并将实验结果与BPNN算法进行对比。实验结果表明,深度卷积神经网络比浅层BPNN具有更好的综合性能。朱莉等^[27]提出一种适用于任意方向的自

然场景文本识别算法。该算法使用高分辨分割网络作为基础框架提取文本的空间信息,利用卷积长短时记忆网络提取文本的时空序列信息。测试结果表明,无论是规则文本还是不规则文本都取得了不错的效果。为了在多标签分类任务中考虑标签之间以及标签与文本各个部分之间的相关性,王进等^[28]提出了编码器图嵌入和区域注意力机制解码器模型。现有的文本检测方法大多针对自然场景,工程图中的文本有着密度分布不均,文本方向多样、语种多样的特点,在进行序号检测的时候,工程图中的零件参数、机件图案和引出线与序号具有相似的纹理特征,会极大干扰工程图序号的检测精度。为避免将现有的针对自然场景的文本检测方法直接应用于工程图序号检测而出现的漏检、误检等问题。胡洪^[29]将注意力机制应用到工程图序号检测方法中以提升网络的检测性能。

为了解决针对特定特征多标签学习算法未能在聚类以及分类阶段考虑标签相关性的问题,王进等^[30]提出了一种基于标签相关性的特定特征多标签学习算法。该算法将标签空间加入特征空间进行聚类构建分类模型,通过树集成表达标签间多种不同结构的强相关性。结果表明,该算法较其他对比多标签算法,在整体上能取得较好的效果。针对图神经网络中存在的过度平滑问题,王进等^[31]提出了一种基于多状态图神经网络的文本分类算法。该算法是利用网络层的多个历史状态信息对图神经网络进行强化,构建合理的文本图结构数据作为模型输入。在缓解网络层过度平滑问题的同时,结合2种改进后的不同类型的图神经网络来增强模型的特征提取与特征聚合能力。

目前基于图结构的文本分类方法存在边噪声和节点噪声干扰、缺乏文本层次信息和位置信息等问题。甘玲等^[32]提出了一种基于正则约束的分层仿射图神经网络文本分类模型(hierarchical affine regularized-constraint graph network for text classification, Text-HARC)。该模型融合了图注意力网络与门控图神经网络,消融实验结果也验证了该方法的有效性。随着金融业务快速发展,盖章文书图像的数量急剧增加,传统的人工审核方式显现出工作效率低、审核错误率高等特点。为了解决这一问题,基于光学字符识别和图像检测技术,曹菁^[33]设计并实现了一个自动审核方法以解决盖章文书图像审核耗时、低效、准确率无保障等问题。针对电力工程图纸识

别中存在无法整体标记粘连字符以及标记效率较低的问题,樊冬梅等^[34]提出了基于数学形态膨胀处理及简化决策树的连通域标记算法。通过分析相邻像素的各种可能性,设计了一种基于简化决策树的高效连通域标记算法,进而提升了电力工程图纸的识别效率。为了解决工程图纸的标题栏结构和内容格式不统一导致的图纸定位不精准问题,章喻龙等^[35]利用YOLOX(you look only once X)深度学习框架,通过自动化标注标题栏的方法,实现了非标准化标题栏的高准确性识别。为了实现自动评审电力施工图,张华等^[36]结合改进的加速分段测试特征检测(features from accelerated segment test, FAST)算法和视觉几何组网络(visual geometry group network, VGG)网络完成了电力工程图纸符号智能识别与电力工程量的智能校核。周文举等^[37]针对航空电气线路互联系统线束模板图纸在更改时的识别与分类问题,设计了一种基于迁移学习的飞机线束模板图纸内容分类方法。为了改变传统电网工程评审依赖于人力的现状,董智等^[38]将图像识别技术应用于电网工程辅助评审领域,提出一种基于改进非极大值抑制的快速区域卷积神经网络(modified non-maximum suppression faster region-based convolutional neural network, MNMS Faster R-CNN)的电网工程图纸字符识别与检测算法。为了更有效地利用标签间的依赖关系并整合标签特征与文本特征,王旭阳等^[39]提出了一种名为融合自注意力机制与基于关联的鲁棒检测网络的组合式图神经网络(compositional graph neural network with transformer and correlation-based robust detection network, CGTCN)的多标签文本分类模型,该模型能够在多标签文本分类任务中取得更好的分类效果。

然而上述方法要么着重考虑工程图纸中表格的识别与提取,要么只适用于特定情况,例如:盖章文书的审核或者仅考虑长度型的标注尺寸等,并没有很好地考虑到图纸内部的标识识别,这些内部标识通常具有重要的意义,有助于工程师、设计师和施工人员查看、审核和分析工程图的全貌。考虑到对于不同的标识而言,识别完整性要求并不同,本文将工程图纸关键标识分为两类:索引标识和尺寸标识。其中,索引标识需要识别得更精确同时具有相对明显的特征,而尺寸标识的识别要求相对较低且通常更为复杂。本文在不同标识的识别效果和识别效率之间做出权衡,采用不同的识别策略,利用深度学习

方法对工程图纸进行信息提取。

2 方法总体介绍

本文提出的工程图纸关键标识定位与识别方法主要针对工程图纸中的索引标识(包括引线标识、剖面标识和横截面标识)和尺寸标识进行位置检测和识别。图 1 展示了对应的样本,方法的流程如图 2 所示。图 1 中方形边框的内容为索引标识,三角形边框的内容为剖面标识,菱形边框的内容为横截面标识,数字代表尺寸标识。

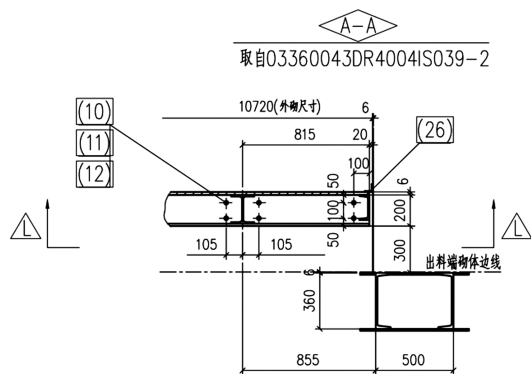


图 1 工程图纸零件样本

Fig.1 Sample of engineering drawings

具体方法步骤如下。

步骤 1 将输入的原始图纸图片按照检测训练时输入的尺寸大小的窗口进行滑动切割,为避免完整标识被窗口截断,在窗口滑动时进行一半的冗余处理,即每次横向和纵向滑动都只移动窗口宽和高的一半的距离。

步骤 2 将每张切割子图输入给训练好的“索引标识检测模型”(YOLOv8),定位索引标识在子图中的位置。

步骤 3 根据每张子图基于原图的偏移位置还原出索引标识在原图的位置,由于窗口滑动的冗余处理,可能会存在同一标识在不同切割子图中被检测到多次,从而还原到原图中时该标识会有多个位置候选框,因此,还需进行非极大值抑制剔除冗余的候选框。

步骤 4 使用飞桨光学字符识别工具(paddle optical character recognition, PaddleOCR)中通用的文字检测,对原始图纸中所有文字信息进行定位。

步骤 5 使用训练好的“非尺寸标识检测模型”(YOLOv8),对图纸中含有大块文字且并非尺寸标识的信息(主要包括表格和文字说明)进行定位,非尺寸标识区域如图 3 所示。

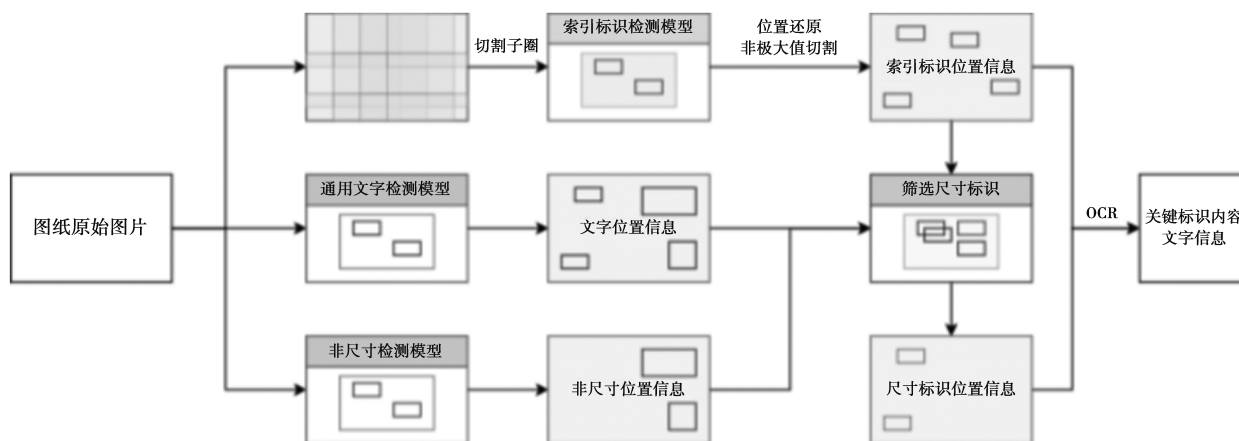


图 2 工程图纸标识检测与识别流程图

Fig.2 Process diagram for locating and identifying key markings on engineering drawings

步骤 6 根据上述检测到的索引标识位置和非尺寸标识位置,从图纸中所有文字候选框中筛选出尺寸标识。主要通过计算两者的候选框面积交并比来判断文字候选框是否与索引候选框表示同一标识位置,以及是否属于非尺寸区域内部。

步骤 7 使用 PaddleOCR 对所有索引标识和尺寸标识区域进行文字识别。

3 核心算法详细介绍

在整个工程图纸关键标识的定位与识别任务中,输入一张工程图纸,需要达成的效果:①寻找到图纸中的所有索引标识,对每一个索引标识都要具体划分为引线标识、剖面标识或是横截面标识;②尽可能完整地找到尺寸标识;③识别出找到的索

引标识和尺寸标识。其中,主要的任务可以分为 2 类:①目标检测;②文字识别。需要用到 2 个深度学习框架去完成这 2 个任务。

3.1 目标检测

作为一种计算机视觉技术,目标检测用于在图像或视频中检测出物体,并给出具体的位置和类别。在工程图纸中,本文将利用目标检测技术寻找索引标识和尺寸标识。YOLO 是一种流行的物体检测和图像分割模型,由华盛顿大学的 J. Redmon 和 A. Farhadi 开发。自 2015 年推出,YOLO 以其高速度和高精度迅速受到欢迎^[40-41]。随着多年来的发展,Ultralytics 提供了 YOLO 的最新版本 YOLOv8。YOLOv8 支持全方位的视觉 AI 任务,这种多功能性

使用户能够在各种应用和领域中利用 YOLOv8 的功能。2023 年,针对人工目测法检测汽车齿轮端面缺陷效率低、精度低的问题,Wang 等^[42]提出了一种基于 YOLOv8 算法的汽车齿轮端面缺陷检测方法。该方法用于真实场景检测任务中的汽车齿轮缺陷检测,有效降低了漏检率,表现出良好的性能。满足真实工业场景的检测需求,在汽车齿轮缺陷检测中具有先进性和实用性。中药片是一类生产工艺复杂、品种繁多的中药制剂,鉴别中药片是一项具有挑战性的任务。基于 YOLOv8,Su 等^[43]提出了一种中药片目标检测方法。在本文中,YOLOv8 将被应用于寻找索引标识和尺寸标识。

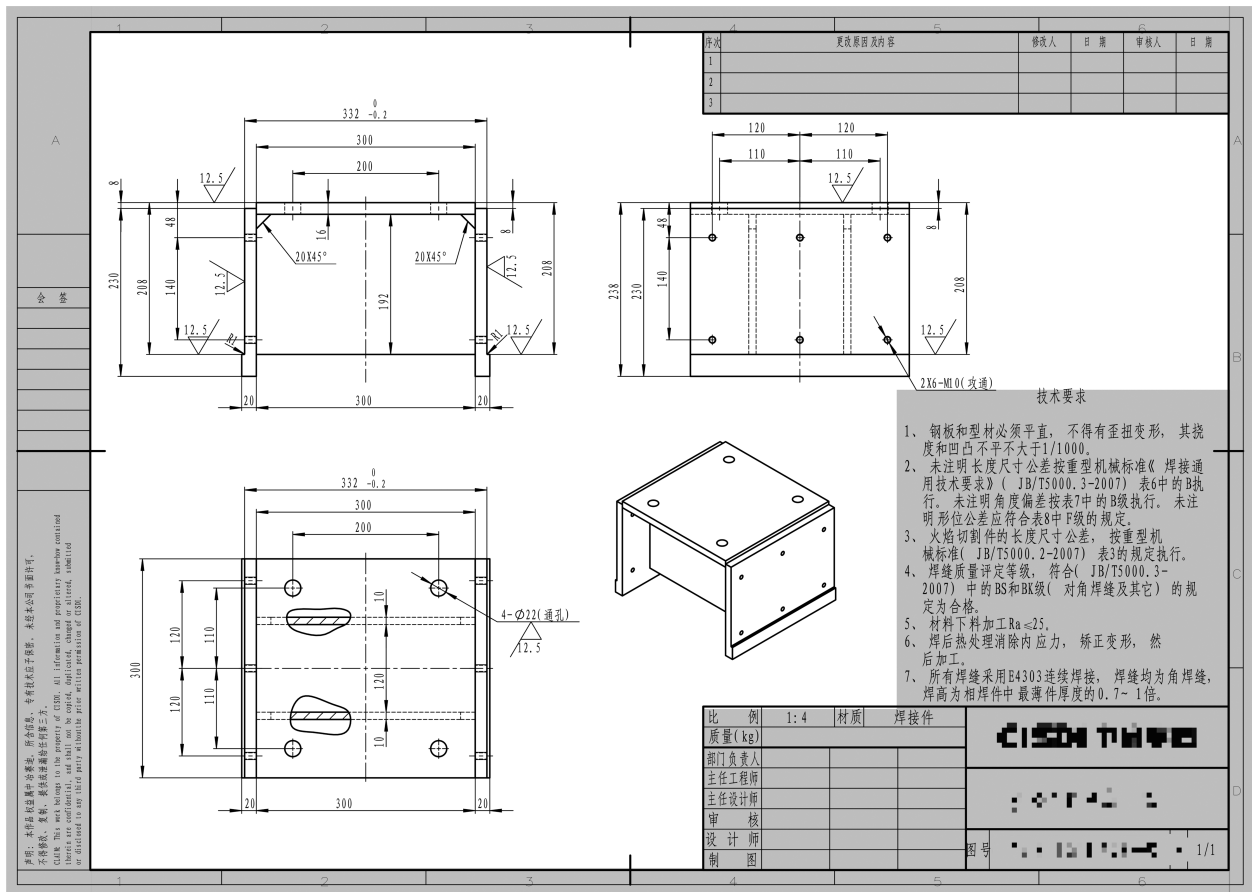


图 3 非尺寸标识区域

Fig.3 Area containing non-size identifiers

3.2 文本识别

文字识别通过对图像中的文本进行识别和提取,将文本信息转换为可读的数字文本格式。在工程图纸中,将利用文字识别技术识别寻找到的索引标识和尺寸标识。PaddlePaddle 是百度飞桨推出的

一款开源深度学习平台^[44]。在进行文字识别时,PaddlePaddle 提供了多种基于深度学习的文字识别工具和模型,如 PaddleOCR 等,这些工具和模型具有高准确率、快速识别和易用性等优势,可以满足不同用户的需求。作为一个完善的深度学习平台,

PaddlePaddle 仍然在不断地更新迭代^[45]。在本文中,PaddlePaddle 将被应用于识别寻找到的索引标识和尺寸标识。

4 图纸各类标识检测与识别

4.1 索引标识检测训练

图纸中的索引标识一共有 3 种,分别是引线标识、剖断面标识和横截面标识。本文主要使用 YOLOv8 训练对这 3 种目标进行分类检测。在权衡任务的复杂度和不同规模模型的性能之后,具体地,选择了 s 级规模的模型(即 YOLOv8s 模型)来处理该项任务,该模型参数量是 11.2M。从最后的结果来看,选择该规模较小的模型不仅在此类较为简单的任务上有不错的精度表现,而且在训练周期和推理速度上也有较好的表现。

训练过程中,首先使用 labelme 标注工具对数据集中的上述 3 类目标进行分类标注;然后在训练前需将 labelme 生成的标注信息 json 文件通过脚本程序转换为 Ultralytics YOLO 格式的 txt 文件;最后将对应的数据集和标注信息放在 Ultralytics 项目下指定的文件夹中,并附上相应的数据集配置说明文件后开始训练模型,由于数据集都是 1 280×720 尺寸的图片,因此,需要将参数 imgsz(image size)由默认的 640 修改为 1 280。

4.2 尺寸标识检测策略

在工程图纸中存在大量的零件尺寸数据,它们数量众多,位置不定,且格式多样,如果使用人工去标记尺寸数据并制作数据集会耗费大量的时间和人力。PaddleOCR 提供了端到端的文字识别解决方案,从原始图像直接输出检测框位置和文本内容,本文利用该功能检测和识别出工程图纸中的所有文本,并排除掉工程图纸中的标识数据和其他无关文本,从而过滤出尺寸数据,经验证该方法具有较快的速度和良好的准确性。

具体上,图纸中含有文字且不属于尺寸标识的区域(除上述索引标识以外)主要包括图纸中的各类表格(包括零件明细表、图纸信息表、负责人员表等)和关于图纸内容的技术说明等大块文字描述。本文主要使用 YOLOv8 训练对这些非尺寸目标进行分类检测,与上述索引标识检测训练不同的是,由于索引标识通常较小,可以保证在一定尺寸窗口中完全包含该标识,而图纸中的表格和文字描述的包围盒大小较大,且难以用固定大小尺寸的窗口保证包

含该内容,因此,先对图纸进行等比例缩小至 640×640 分辨率以内后,再将其作为数据集。推理使用时,先记录好缩小的比例再将模型输出的非尺寸目标位置信息根据该比例还原回基于原图纸大小的位置。并且由于非尺寸目标足够大,经验证后在图纸缩小后仍能保留足以使得模型学习辨识的关键特征。

过滤策略上,尺寸候选样本集便是通过 PaddleOCR 检测到的所有文字框,负例样本包括索引标识框、表格框和文字描述框。对于每个候选框都需要和每个负例样本计算面积交并比 r_1 ,表示为

$$r_1 = \frac{S_{\text{交集}}}{S_{\text{并集}}} \quad (1)$$

式(1)中: $S_{\text{交集}}$ 表示候选框和负例样本交集的面积; $S_{\text{并集}}$ 表示候选框和负例样本并集的面积。

若 r_1 超过阈值 $t_1 = 0.6$ 则判定该候选框属于非尺寸标识,应被剔除;此外由于表格和文字描述区域较大,一部分候选框会包含在其中,候选框包含在非尺寸目标内部样例图如图 4 所示。

同时,还需计算两者相交的面积与候选框面积之比 r_2 ,表示为

$$r_2 = \frac{S_{\text{交集}}}{S_{\text{候选框}}} \quad (2)$$

式(2)中, $S_{\text{候选框}}$ 表示候选框的面积。

若其超过阈值 $t_2 = 0.9$,则判定该候选框被非尺寸目标包含,也属于非尺寸标识,应被剔除。最终保留下来的候选框便是尺寸标识。在实际工程识别任务中,可以结合人工处理的方式,对不正确的地方进行矫正。通过 labelme 直接标注非尺寸标识,取消这部分区域的非尺寸标识识别任务,从而得到更好的尺寸标识识别结果。

4.3 标识内容文字识别

PaddleOCR 提供了多种轻量级的文本识别模型以及丰富的文本识别工具,本文使用了 PP-OCRv4 模型用作工程图纸的文本识别。在使用 YOLO 检测出工程图纸中的各类索引标识后,可以得到每个标识在图纸上的位置信息,用 4 个点坐标表示,分别代表了标识检测框的 4 个顶点。首先利用这 4 个顶点从图纸中裁剪出仅包含标识文本的图片,然后使用 PaddleOCR 的文本识别工具对裁剪图片中的文本进行识别,最后输出每一个标识的位置以及对应的识别内容。

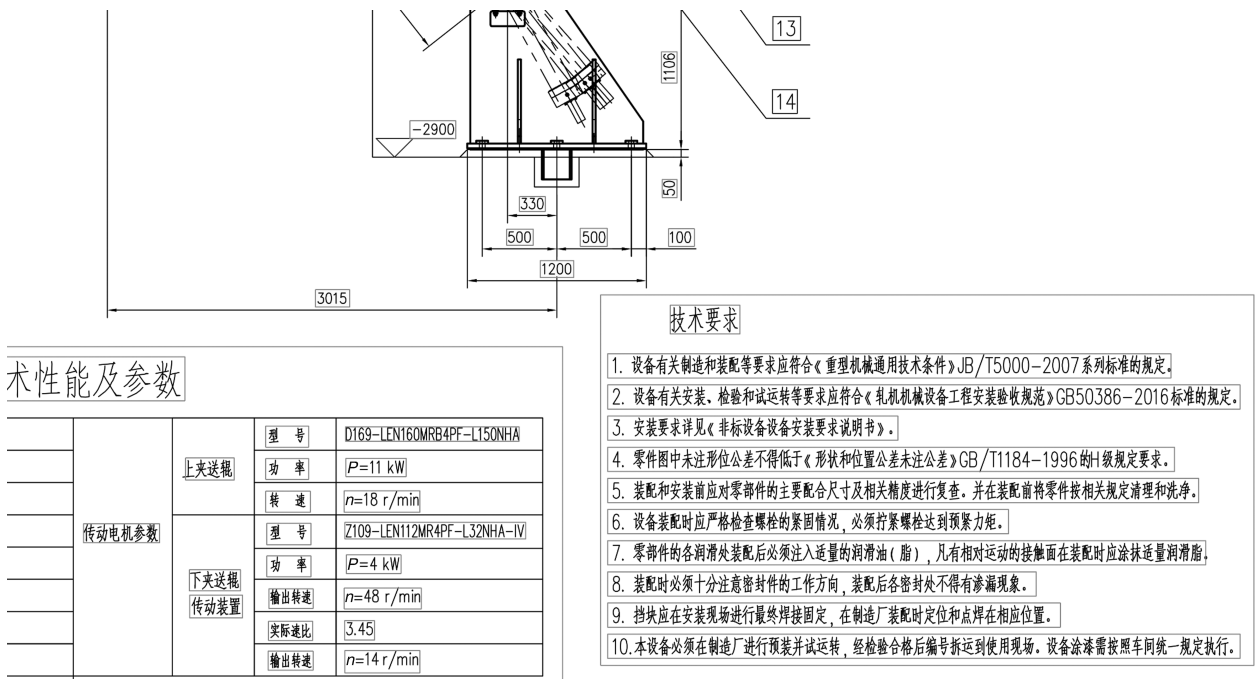


图 4 候选框包含在非尺寸目标内部样例图

Fig.4 Sample of candidate boxes are included within non-size identifiers

5 实验结果与分析

5.1 工程图纸关键标识数据集

由于建筑行业的工程图纸保密级别较高,数据量较少,本文使用的工程图纸为企业提供的专用图纸,共 470 张 PDF 文件,将其转换为 png 图片格式用于制作数据集。由于工程图纸的分辨率较高,部分图纸可能达到上万的分辨率,为了后续将数据集用于训练,将图纸分割成了 1 280×720 分辨率的小图片,并以此为基础制作数据集。在分割图片时,每张图片间会留有冗余,这样做的目的是防止在分割线处破坏图纸中的文本信息,同时也会增加分割后的图片数量,相当于进行了数据增广,解决了工程图纸数量少的问题。

在得到分割的图片后,人工标注工程图纸中的索引标识信息,本文将工程图纸中的索引标识分为 3 类:引线标识、剖断面标识和横截面标识。使用 labelme 分别标注图纸中的 3 类标识信息,并将得到的标注数据转换为 Ultralytics YOLO 格式的数据集,最终得到 8 924 张数据集,并将其用于之后的标识检测训练。

5.2 标识检测网络训练与测试

划分工程图纸中的每一类标识都有各自的特

点,例如引线标识会有指向零件的折线以及对应的编号,剖断面标识会有成对的大写字母以及箭头,基于以上的特点,可以利用现有的目标检测网络训练适用于工程图纸标识检测的模型,本文使用了 YOLOv8 来训练标识检测模型,将 8 924 张数据集按 0.7/0.15/0.15 的比例划分为训练集、验证集和测试集。网络模型的训练与测试均在服务器上进行,服务器的配置为 CPU: Intel(R) Core(TM) i7-9700K/GPU: RTX 2080 Ti,使用的操作系统为 Ubuntu 20.04.2 LTS,Pytorch 版本为 1.12。经实验证明,该检测模型具有优良的准确性和泛化能力。

表 1 数据集分布

Tab.1 Dataset distribution

类型	训练集	验证集	测试集
含有索引标识	2 461	525	527
不含索引标识	3 791	810	810

5.3 标识检测实验结果

本文使用精确率 (precision) 和召回率 (recall) 以及综合指数 F_1 值作为工程图纸标识检测模型的评估标准。精确率的计算式为

$$P = \frac{T_{TP}}{T_{TP} + F_{FP}} \quad (3)$$

召回率的计算式为

$$R = \frac{T_{TP}}{T_{TP} + F_{FN}} \quad (4)$$

F_1 值的计算式为

$$F_1 = \frac{2T_{TP}}{2T_{TP} + F_{FP} + F_{FN}} \quad (5)$$

式(3)~(5)中: T_{TP} 为正类判定为正类的样本数量; F_{FN} 为正类判定为负类的样本数量; F_{FP} 为负类判定为正类的样本数量。精确率 P 能够反映模型发现正样本相关性的能力,关注的是模型预测的准确性;而召回率 R 能够反映模型在整个数据集中发现正样本的能力,关注的是模型能够识别出多少真正的正样本; F_1 值是精确率和召回率的调和平均数,是一个综合考虑精确率和召回率性能的指标。这 3 个指标共同作用,对本文模型的性能进行全面评估。本文同样使用 YOLOv8 进行模型的测试,实验结果如表 2 所示。可以看到,该模型具有优良的精确率和召回率,证明本文的方法能够有效地检测工程图纸中多种类别的标识。鉴于尺寸标识的复杂性和多样性、尺寸标识容易与图纸中的其他部分重叠,以及工程界关于尺寸标识的识别要求不明确,上述结果没有统计尺寸标识数据的结果。目前的识别目标是需要精确地识别定位索引标识,以及尽可能完整地得到尺寸标识,本文也是根据这一目标确定的识别策略。鉴于工程图纸的保密性,目前所用的工程图纸都是由企业提供的专用图纸,还没有在其他公司的工程图纸上测试过,如果其他工程图纸中使用的规范差距较大,需要根据任务目标重新定制识别策略或重新训练模型。

表 2 工程图纸标识检测结果

Tab.2 Result of detecting the identification in the engineering drawings

索引标识类别	精确率/%	召回率/%	F_1 值
引线标识	97.74	99.20	0.984 6
剖断面标识	97.06	97.06	0.970 6
横截面标识	96.09	96.85	0.964 7

6 总结与展望

如何利用计算机去认知、查找工程图纸以替代繁重的人工查找一直都是计算机视觉领域的研究热点。传统的检测方法效率较低,且通常只适用于某一类图纸,缺乏泛化性。基于深度学习的方法依靠

着更高的检测精度以及更好的鲁棒性成为了图纸识别的研究热点。本文采纳了深度学习与光学字符识别技术,提出了一套面向工程图纸中关键标识的高精度定位与解析的方法。该方法能有效识别图纸中的索引标识和尺寸标识,并通过一系列实验证明了其出色的准确度和适应性。接下来将考虑如何利用深度学习方法定位、识别工程图纸中的表格:标题栏和设备明细表。目前的表格识别算法提取得到的表格结构都会有一些差异,这种差异在工程图纸表格识别任务中是致命的,这代表无法用一个统一的方法识别表格单元,如果出现串行等现象,将会得到完全错误的识别结果。工程图纸中的标题栏拥有完全的表格结构,设备明细表拥有统一的表头,鉴于图像识别具有较高的准确率可以将表格定位任务转化为图像识别任务,利用表格结构的统一性可以将表格识别任务转变为关键位置文本识别任务,从而得到较高的准确性,直接将结果输入到预先准备的表格结构中,避免了后续识别表格单元的步骤。

参考文献:

- [1] 宋小鸿.工程图纸的关键信息智能提取技术[D].北京:北京交通大学,2019.
SONG X H. Intelligent extraction technology of key information in engineering drawings [D]. Beijing: Beijing Jiaotong University, 2019.
- [2] PRATHWIN I, RODRIGUES A P, VIJAYA P, et al. Tu-lu language text recognition and translation[J]. IEEE Access, 2024(12): 12734.
- [3] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [4] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist text proposal network [C]//LEIBE B, MATAS J, SEBE N. Computer Vision-ECCV 2016, Lecture Notes in Computer Science. Cham, Germany: Springer, 2016: 56-72.
- [5] ZHOU X, YAO C, WEN H, et al. EAST: An efficient and accurate scene text detector [C]//2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2017: 2642-2651.
- [6] YIN W, SCHUTZE H, XIANG B, et al. ABCNN: Attention-based convolutional neural network for modeling sentence pairs[J]. Transactions of the Association for Computational Linguistics, 2016(4): 259-272.
- [7] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its ap-

- plication to scene text recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017(3): 2298-2304.
- [8] XIE H, FANG S, ZHA Z J, et al. Convolutional attention networks for scene text recognition[J]. ACM Trans Multimedia Comput Commun Appl, 2019(15): 3-10.
- [9] HU Y, DONG B, HUANG K, et al. Scene text recognition via dual-path network with shape-driven attention alignment[J]. ACM Trans Multimedia Comput Commun Appl, 2024, 20(4): 107-116.
- [10] GHOSH J, TALUKDAR A K, SARMA K K. A lightweight natural scene text detection and recognition system [J]. Multimedia Tools and Applications, 2024 (83): 6651-6683.
- [11] ALSHAWI A A A, TANHA J, BALAFAR M A. An attention-based convolutional recurrent neural networks for scene text recognition [J]. IEEE Access, 2024 (12): 8123-8134.
- [12] 张习文,欧宗瑛.机械图纸扫描图像尺寸信息提取[J].机械科学与技术, 2001, 20(2): 291-293.
ZHANG X W, OU Z Y. An algorithm to extract information of dimensions from vectorized scanned image of mechanical drawings[J]. Mechanical Science and Technology, 2001, 20(2): 291-293.
- [13] 陈勇,朱林,常明.工程图中粘连字符的提取与分割[J].华中理工大学学报, 1996(24): 23-26.
CHEN Y, ZHU L, CHANG M. The extraction and segmentation of connected characters in engineering drawings [J]. Journal of Huazhong University of Science and Technology, 1996(24): 23-26.
- [14] 胡友兰,黄树槐,常明.工程图中字符分离和标注字符串生成技术[J].华中理工大学学报,1997(25):30-33.
HU Y L, HUANG S H, CHANG M. Character separation and text strings generation in engineering drawings [J]. Journal of Huazhong University of Science and Technology, 1997(25): 30-33.
- [15] 邹荣金,蔡士杰,张福炎,等.字符粘连及字线相交的分割与识别方法[J].软件学报, 1999(10): 241-247.
ZHOU R J, CAI S J, ZHANG F Y, et al. Segmentation and recognition methods of adhesion and intersection character string with the line [J]. Journal of Software, 1999(10): 241-247.
- [16] 范帆,关佑红.工程图纸字符串及标注信息提取[J].计算机工程与应用, 2012, 48(7): 161-164.
FAN F, GUAN G H. Recognition of characters and annotations on engineering drawings [J]. Computer Engineering and Applications, 2012, 48(7): 161-164.
- [17] 杜建强,陈月林,刘少媚,等.工程图纸上的字符提取和识别系统[J].计算机工程, 1995(21): 62-65.
DU J Q, CHEN Y L, LIU S M, et al. System of character extraction and recognition in engineering drawings [J]. Computer Engineering, 1995(21): 62-65.
- [18] 蔡良伟,韩大庆.工程图纸中字符的分离和识别[J].深圳大学学报, 1997(14): 21-26.
CAI L W, HAN D Q. Separation and recognition of characters in engineering drawings [J]. Journal of Shenzhen University (Science and Engineering), 1997 (14): 21-26.
- [19] 杨万山,陈松乔,唐连章.基于BP神经网络的工程图纸图形符号识别[J].微型电脑应用, 2000(16): 22-23.
YANG W S, CHEN Q S, TANG L Z. Recognition of graphic symbols in engineering drawings based on BP-Neural Network [J]. Microcomputer Applications, 2000 (16): 22-23.
- [20] 董海卫,江早,王永军.统计与结构结合的工程图纸符号识别方法[J].计算机辅助设计与图形学学报, 2000 (12): 375-379.
DONG H W, JIANG Z, WANG Y J. Statistic and structure integrated approach of symbol recognition for engineering drawings [J]. Journal of Computer-Aided Design and Computer Graphics, 2000(12): 375-379.
- [21] 翟传敏,杜吉祥,黄飞.基于径向基概率神经网络的工程图纸图形符号识别[J].智能系统学报, 2006(1): 88-91.
ZHAI C M, DU J X, HUANG F. Graphic symbol recognition of engineering drawings based on radial basis probabilistic neural networks [J]. CAAI Transactions on Intelligent Systems, 2006(1): 88-91.
- [22] 宋晓宇,李玉冲,刘继飞.基于拓扑结构的工程图纸识别方法[J].沈阳建筑大学学报(自然科学版), 2011, 27(4): 776-781.
SONG X Y, LI Y C, LIU J F. The research on engineering drawing recognition based on topology [J]. Journal of Shenyang Jianzhu University (Natural Science), 2011, 27 (4): 776-781.
- [23] 单世哲,许伯强,韩刚.电气工程图纸矢量化技术滤波算法的研究分析[J].农业科技与装备, 2012(215): 30-31, 34.
SHAN S Z, XU B Q, HAN G. Research and analysis on filtering algorithm of vectorization technology for electrical engineering drawings [J]. Agricultural Science and Technology and Equipment, 2012(215): 30-31, 34.
- [24] 代贺,陈洪密,李志申.基于卷积神经网络的数字识别[J].贵州师范大学学报(自然科学版), 2017, 35(5): 96-101.
DAI H, CHEN H M, LI Z S. Digital recognition based on

- convolutional neural network[J]. Journal of Guizhou Normal University (Natural Sciences), 2017, 35(5): 96-101.
- [25] ELYAN E, JAMIESON L, ALI-GOMBE A. Deep learning for symbols detection and classification in engineering drawings[J]. Neural Networks, 2020(129): 91-102.
- [26] 陈玉辉. 基于卷积神经网络的工程图纸图像识别算法研究[D]. 湖南: 南华大学, 2020.
- CHEN Y H. Engineering drawing recognition technology based on convolutional neural network[D]. Hunan: Nanhua University, 2020.
- [27] 朱莉, 陈宏, 景小荣. 任意方向自然场景文本识别[J]. 重庆邮电大学学报(自然科学版), 2022, 34(1): 125-133.
- ZHU L, CHEN H, JIN X R. Text recognition of natural scenes in any direction[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2022, 34(1): 125-133.
- [28] 王进, 徐巍, 丁一, 等. 基于图嵌入和区域注意力的多标签文本分类[J]. 江苏大学学报(自然科学版), 2022, 43(3): 310-318.
- WANG J, XU W, DING Y, et al. Multi-label text classification based on graph embedding and region attention [J]. Journal of Jiangsu University (Natural Science Edition), 2022, 43(3): 310-318.
- [29] 胡洪. 基于注意力机制的工程图序号检测方法研究[D]. 重庆: 重庆邮电大学, 2023.
- HU H. Research on serial number detection of engineering drawing based on attention mechanism[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2023.
- [30] 王进, 梁晨, 孙开伟, 等. 基于标签相关性的标签特定特征多标签学习[J]. 江苏大学学报(自然科学版), 2023, 44(5): 554-563.
- WANG J, LIANG C, SUN K W, et al. Multi label learning with label specific features via label correlations[J]. Journal of Jiangsu University (Natural Science Edition), 2023, 44(5): 554-563.
- [31] 王进, 陈重元, 邓欣, 等. 多状态图神经网络文本分类算法[J]. 重庆邮电大学学报(自然科学版), 2023, 35(2): 193-201.
- WANG J, CHEN C Y, DENG X, et al. Multi-state graph neural network for text classification [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2023, 35(2): 193-201.
- [32] 甘玲, 刘菊. 基于正则约束的分层仿射图神经网络文本分类模型[J]. 重庆邮电大学学报(自然科学版), 2023, 35(4): 715-721.
- GAN L, LIU J. Hierarchical affine graph neural network text classification model based on regular constraints[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2023, 35(4): 715-721.
- [33] 曹菁, 陈康, 齐宁, 等. 基于 OCR 和图像检测的盖章文书图像自动审核方法[J]. 应用科学学报, 2023, 41(6): 1058-1067.
- CAO J, CHEN K, QI N, et al. Auto-checking stamped document image based on OCR and image detection[J]. Journal of Applied Sciences, 2023, 41(6): 1058-1067.
- [34] 樊冬梅, 严文豆, 梁毅, 等. 基于改进连通域检测算法的电力工程图纸识别技术[J]. 沈阳工业大学学报, 2024, 46(4): 361-366.
- FAN D M, YANG W D, LIANG Y, et al. Technology based on improved connected component labeling algorithm for power engineering drawings recognition[J]. Journal of Shenyang University of Technology, 2024, 46(4): 361-366.
- [35] 章喻龙, 罗壮强, 石凌峰. 基于 YOLOX 算法的工程图纸标题栏识别[J]. 数字技术与应用, 2024, 42(3): 198-202.
- ZHANG Y L, LUO Z Q, SHI L F. Title block recognition of engineering drawings based on YOLOX algorithm[J]. Digital Technology and Application, 2024, 42(3): 198-202.
- [36] 张华, 肖莹, 孙海森. 基于改进 FAST 的电力工程图纸识别与校核技术[J]. 电子设计工程, 2024, 32(10): 186-190.
- ZHANG H, XIAO Y, SUN H S. Identification and verification technology of power engineering drawings based on improved FAST [J]. Electronic Design Engineering, 2024, 32(10): 186-190.
- [37] 周文举, 朱一丁, 杨庆华, 等. 基于迁移学习的飞机线束模板图纸识别模型研究[J]. 信阳师范学院学报(自然科学版), 2024, 37(2): 240-245.
- ZHOU W J, ZHU Y D, YANG Q H, et al. Research on recognition model of aircraft wiring harness template drawings based on transfer learning [J]. Journal of Xinyang Normal University (Natural Science Edition), 2024, 37(2): 240-245.
- [38] 董智, 赵阳, 田福丰. 基于改进卷积神经网络的电网工程图纸字符识别与检测算法[J]. 电子设计工程, 2023, 31(13): 27-31.
- DONG Z, ZHAO Y, TIAN F F. Character recognition and detection algorithm of power grid engineering drawings-based on improved convolutional neural network[J]. Electronic Design Engineering, 2023, 31(13): 27-31.

- [39] 王旭阳,卢世红.基于标签构建与特征融合的多标签文本分类研究方法[J].贵州师范大学学报(自然科学版), 2025, 43(1): 105-114.
WANG X Y, LU S H. Research methodology of multi-label text classification based on label construction and feature fusion [J]. Journal of Guizhou Normal University (Natural Sciences), 2025, 43(1): 105-114.
- [40] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 779-788.
- [41] REDMON J, FARHADI A. Yolo9000: Better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2017: 6517-6525.
- [42] WANG H, YANG G, HAO X, et al. Automotive gear defect detection method based on YOLOv8 algorithm[C]//Asia Symposium on Image Processing (ASIP). Tianjin, China: IEEE, 2023: 19-23.
- [43] SU Y, CHENG B, CAI Y. Detection and recognition of traditionalchinese medicine slice based on YOLOv8 [C]//IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT). Qingdao, China: IEEE, 2023: 214-217.
- [44] BI R, XU T, XU M, et al. Paddlepaddle: A production-oriented deep learning platform facilitating the competency of enterprises [C]//2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys). Hainan, China: IEEE, 2022: 92-99.
- [45] LI X, XIONG H, LI X, et al. InterpretDL: Explaining deep models in PaddlePaddle [J]. J Mach Learn Res, 2022(23): 197-208.

作者简介:

肖鑫,高级工程师,博士,主要研究方向为工程数字化、信息化。E-mail:xin.xiao@cisdi.com.cn。

陈青松,高级工程师,硕士,主要研究方向为工程数字化、信息化。E-mail:qingsong.chen@cisdi.com.cn。

吴思远,副教授,硕士,主要研究方向为数据库系统技术、计算机应用。E-mail:wusy@cqupt.edu.cn。

胡瑞,工程师,学士,主要研究方向为工程数字化、信息化。E-mail:rui.hu@cisdi.com.cn。

(编辑:王敏琦)