

融合图元检测和领域先验的变电站 端子排图纸解析

吴勇¹, 葛奕雯², 洪文谦², 高正霄², 戴挈军¹, 李自然¹, 陈蕾^{2,3*}

(1. 江苏省送变电有限公司, 江苏 南京 211106;

2. 南京邮电大学 计算机学院, 江苏 南京 210023;

3. 江苏省大数据安全与智能处理重点实验室, 江苏 南京 210023)

摘要:端子排图纸是变电站二次回路施工、检修和运行维护的重要依据。应用计算机视觉技术对其进行电气语义解析存在复杂性和专业性上的双重挑战,使其不可避免地面临难例样本被误解析的窘境,从而阻碍该类方法在真实场景的落地应用。针对端子排图纸电气语义解析存在的不足,提出一类融合图元检测和领域先验的变电站端子排图纸电气语义解析方案。该方案涉及图元检测、文本识别及线帽语义标签解析三个步骤,具体由基于滑动窗冗余切割和改进YOLOv8的多尺度图元检测,以及基于文本识别与领域先验的图纸拓扑匹配与纠错拒识两个算法组成。该方案在基于真实场景构建的图元检测数据集与线帽语义标签解析数据集上分别进行了性能验证,实验结果表明,图元检测算法优于当前主流检测模型,图纸拓扑匹配与纠错拒识算法则在拒识率为10.8%的基础上使得线帽语义标签解析的拟召回率和拟精确率都达到了100%。

关键词:变电站图纸识别;端子排图纸解析;图元检测;文本识别;目标检测;领域先验

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2025)01-0184-08

doi:10.20165/j.cnki.ISSN1673-629X.2024.0274

Analysis of Substation Terminal Strip Drawings Integrating Symbol Detection and Domain Priors

WU Yong¹, GE Yi-wen², HONG Wen-qian², GAO Zheng-xiao², DAI Qie-jun¹,

LI Zi-ran¹, CHEN Lei^{2,3*}

(1. Jiangsu Power Transmission & Transformation Co., Ltd., Nanjing 211106, China;

2. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China;

3. Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing 210023, China)

Abstract: Terminal strip drawings are essential for the construction, maintenance, and operation of substation secondary circuits. The application of computer vision technology for their electrical semantic analysis presents dual challenges of complexity and expertise, inevitably leading to the misinterpretation of difficult samples, which hinders the practical application of these methods in real-world scenarios. To address these shortcomings in the electrical semantic analysis of terminal strip drawings, we propose a substation terminal strip drawings electrical semantic analysis scheme that integrates symbol detection and domain knowledge. This scheme involves three steps: symbol detection, text recognition, and wire marker semantic label analysis. It includes multi-scale drawing symbol detection based on sliding window redundancy cutting and an improved YOLOv8, as well as drawing topology matching and error correction rejection algorithms based on text recognition and domain knowledge. Performance validation on a symbol detection dataset and a wire marker semantic label analysis dataset constructed from real-world scenarios shows that the proposed symbol detection algorithm outperforms current mainstream detection models, and the proposed topology matching and error correction rejection algorithm achieves a rejection rate of 10.8%, with both the quasi recall rate and quasi precision rate for wire marker semantic label analysis reaching 100%.

Key words: substation drawing recognition; terminal strip drawing analysis; symbol detection; text recognition; object detection; domain priors

收稿日期:2024-06-03

修回日期:2024-10-09

基金项目:国家重点研发计划项目(2022YFB3303800);江苏省送变电有限公司科技项目(2023外306)

作者简介:吴勇(1987-),男,高级工程师,研究方向为电力自动化与电力工程建设;通信作者:陈蕾(1975-),男,教授,博导,CCF高级会员(16224S),研究方向为计算机视觉、人工智能与机器学习。

0 引言

变电站二次接线图中的端子排图纸是变电站二次回路施工、检修和运行维护的重要依据,记录了变电站机柜内部线缆的接线方式、设备布局及其相互之间的连接关系,对电力系统的设计、建设、运维和故障分析等方面具有至关重要的作用,贯穿了变电站新建、扩建、技改、运维的全部生命周期。

随着人工智能迅速发展,“智能电网”和“数字电网”^[1]的概念日益普及,深度学习在工业应用中取得了重大进展^[2],YOLO^[3](You Only Look Once)系列模型已被广泛应用于解决图纸识别^[4]、行人检测^[5]、无人机检测^[6]等任务。然而,在变电站端子排图纸电气语义解析方面仍面临一系列挑战。首先,图纸尺度差异大导致缺乏统一的缩放标准。由于电气图纸特性且各设计院缺乏统一的制图规范,图纸差异较大,包括图纸尺度和目标大小、清晰度以及需识别的内容占比等方面,基于传统的卷积神经网络的图纸检测方案^[7]难以有效处理这些问题。若统一缩放尺寸^[8],会严重影响检测器精度。YOLT^[9](You Only Look Twice)是一种常用的处理大尺度差异的目标检测方法,通过将大尺度图像分割成小切片,分别检测各个切片后再合并结果。然而,由于图纸差异较大,难以确定单一的缩放尺度,切割时大目标易被切割,从而产生误识别。其次,感兴趣目标(该文在第 1 节将其命名为图元)类间差异小,导致判别特征难以提取。图纸中大部分目标远小于 COCO 数据集^[10]对小目标的定义,即目标尺寸为原图的 1/10。目标极小,且目标特征极相似,类间差异小,受图纸线框和文本框干扰严重,因此检测模型难以准确提取各类别的判别特征,影响检测的准确

性^[11]。常用的解决方法是采用双层-分块的 YOLOv5^[12]目标检测模型,先识别图纸的关键区域,再以高精度识别目标。尽管这种方法提高了小目标相对比例并减少信息丢失,但滑窗和特殊区域的设置可能导致小目标被切割,增加特征干扰,使识别更加复杂化。最后,由于图纸内容复杂,解析中的错误和遗漏不可避免。因此,系统需要提高算法的使用效率,确保解析结果的可信性。

针对上述挑战,该文提出了一种融合图元检测和领域先验的变电站端子排图纸电气语义解析方法。首先,设计了一套多尺度滑窗机来解决图纸尺度差异的问题,通过采用双尺度模型实现对不同尺度图纸的检测,并引入和改进 WBF^[13](Weighted Box Fusion)算法,提升检测模型性能。其次,在目标检测模型中融合了两种注意力机制,显著提升图纸中小目标的特征提取效果。最后,基于领域先验,设计了一套智能解析与识别算法,通过对识别结果进行修正与拒识,提高了算法的识别精度,并能智能提示错误信息,实现了图纸的高精度解析。

1 问题描述

1.1 变电站端子排图纸电气语义解析

图纸电气语义解析需要提取出图纸中所有接线线缆的信息,如图 1 所示,在最终解析文档中,需要提取两根线缆的所有信息,包括单元标牌(42YD)、端子号(4、5)、回路编号(B811、B812)以及线缆名称(1B-131)。这些信息构成了完整的线缆信息,其中一根则表示为(B811/42YD:4 1B-131)。

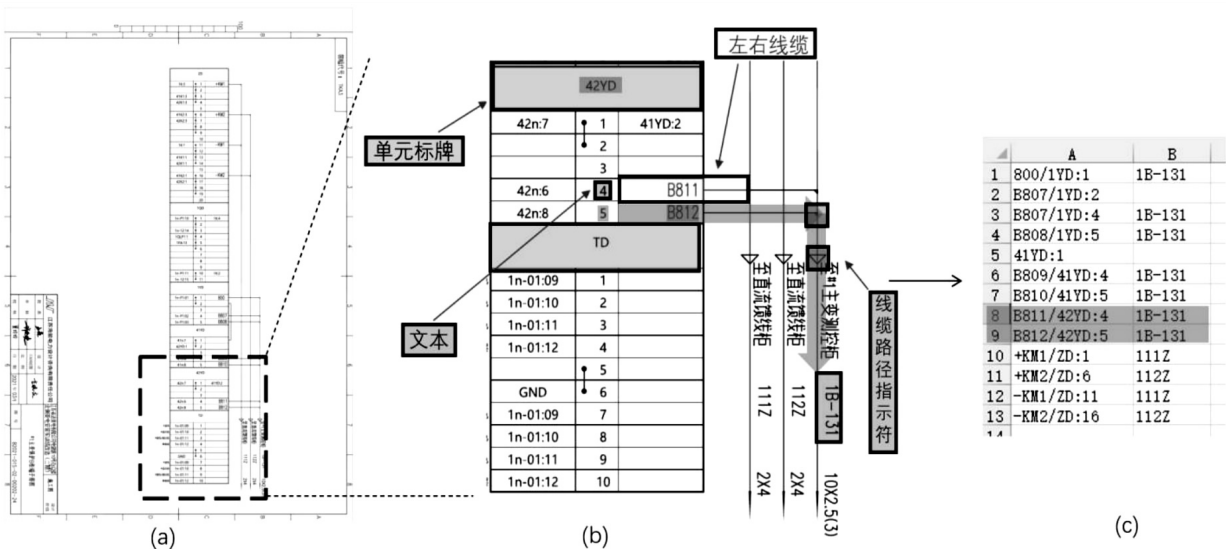


图 1 图纸检测需求示意图

1.2 变电站端子排图纸图元目标检测

图元检测是图纸智能解析的基础任务,为识别线

缆信息,该文将图元定义为电气专业图元、文本图元两大类,如图 1(b)所示,共 11 种图元目标(10 类电气专

业图元和 1 类文本图元),具体包含单元标牌、左右线缆(两类)、线缆路径指示符(7 类,包含 6 类线缆弯折符号、1 类线缆结束符号),以及文本。通过对左右线缆及线缆路径指示符进行拓扑匹配,获取线缆名称等信息。

2 文中方法

2.1 变电站端子排图纸电气语义解析方案总体架构

针对变电站二次接线图纸智能识别场景,提出了一种融合图元检测和领域先验的变电站端子排图纸电气语义解析方法。如图 2 所示,该方法由图元检测和图纸智能解析与拒识两个模块组成,完成对输入图纸的自动识别,最终输出解析文档。

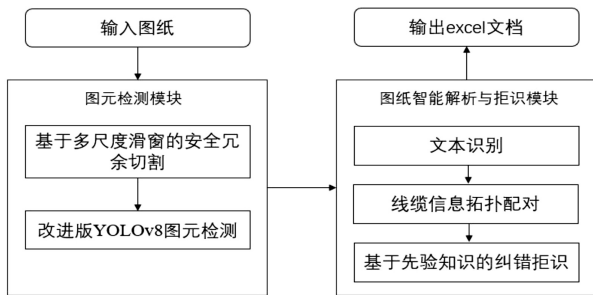


图 2 变电站图纸识别总体架构

其中,图元检测模块主要用于检测输入图像中的感兴趣目标,为后续图纸智能解析提供支持。智能解析与拒识模块涵盖拓扑匹配线缆信息、基于领域先验的自动纠正和拒识。拓扑匹配通过分析位置关系来确定接线线缆位置、接线名称等重要信息。

2.2 基于滑动窗冗余切割和改进 YOLOv8 的多尺度图元检测

2.2.1 基于多尺度滑动窗的图纸安全冗余切割

变电站图纸由不同设计院的专业人员绘制,缺乏统一的标准和规范,导致图纸差异显著,常规目标检测方法难以有效处理。此外,小目标特征难以提取,且与无需检测的表格、线缆等目标之间存在严重干扰,不同图纸面积比差异达到 1 : 16,即使将图纸切割为相同尺寸的切片,小目标依然小于切片尺寸的 1/10,容易受到干扰。这些因素直接导致了原始模型性能表现不佳。

该文对 YOLT 算法进行了改进,提出了基于多尺度滑动窗口的图纸安全冗余切割策略。该策略根据图纸和目标的尺寸预设了两个滑动窗口尺寸,以优化检测性能。滑动窗口尺寸的设置基于四个原则:(1)滑动窗口宽高比与原图相似;(2)大窗口能完整检测目标;(3)小窗口避免漏检小目标;(4)冗余区域可以防止切片断裂导致的图元缺失。根据上述规则,设定的大小尺度滑动窗口分别为 $2\ 560 * 3\ 488$ 和 $1\ 280 *$

$1\ 760$ 像素,冗余区域为 500 和 300 像素。通过这两种尺度滑动窗口机制,可以适应不同尺寸的图纸输入,增强模型对多尺度差异图纸的检测能力。

在模型训练与推理阶段,采用了两个结构一致但参数不同的改进版 YOLOv8^[14] 模型,分别处理经过多尺度滑窗切割的不同尺度的切片,获得目标检测结果。随后,利用后续章节介绍的改进自适应模型融合策略,对这两种尺度下的预测结果进行综合分析融合。算法流程见图 3。

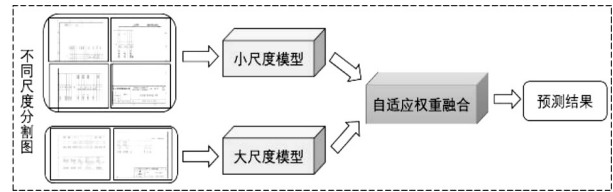


图 3 多尺度滑动窗口算法流程

依据图纸特性设定窗口尺寸,实现了不同尺度情况下对图纸的高效精准解析并缓解了图纸小目标识别难题。然而,这两种尺度实际上构成了两个独立的处理分支,尽管多尺度滑窗在一定程度上缓解了尺度差异带来的问题,但这一问题仍然存在。

2.2.2 融合多重注意力与领域先验的 YOLOv8 图元目标检测

引入多尺度空洞注意力(MSDA^[15])从模型角度缓解图纸差异,引入卷积注意力(CBAM^[16])增强模型对图纸小目标的关注,提升小目标的检测性能。

MSDA 聚焦不同尺度特征,提高模型对细节的捕捉能力,尤其在目标大小变化较大时表现突出。它同时考虑局部和全局信息,增强了目标检测的准确性和鲁棒性,使模型更适应多尺度目标的检测,缓解了多尺度图纸差异和小目标问题。MSDA 采用多头设计,利用滑动窗口膨胀注意力(Sliding Window Dilated Attention, SWDA)在不同尺度上聚合语义信息,有效减少自注意力机制的冗余。CBAM 提升了对小目标的识别能力,减少背景干扰。它结合通道注意力模块(CAM)和空间注意力模块(SAM),优化了对图纸上小目标的识别,通过串联这两个模块,显著提升了模型对细节的捕捉能力。这种方法优先处理关键特征,显著提高了检测精度和整体性能。

该文将 MSDA 和 CBAM 集成至 YOLOv8 模型中。MSDA 整合在 Backbone 网络的 C2f 模块,以适应并提取不同尺度特征;CBAM 融入 Neck 网络的 C2f 模块,以强化关键特征并减少复杂背景和表格对小目标检测的干扰。此架构中,Backbone 网络负责全局特征提取,利用空洞卷积的广泛感受野处理不同尺度的输入;Neck 网络负责特征聚合,确保有效提取和突出重要特征。模型结构见图 4。

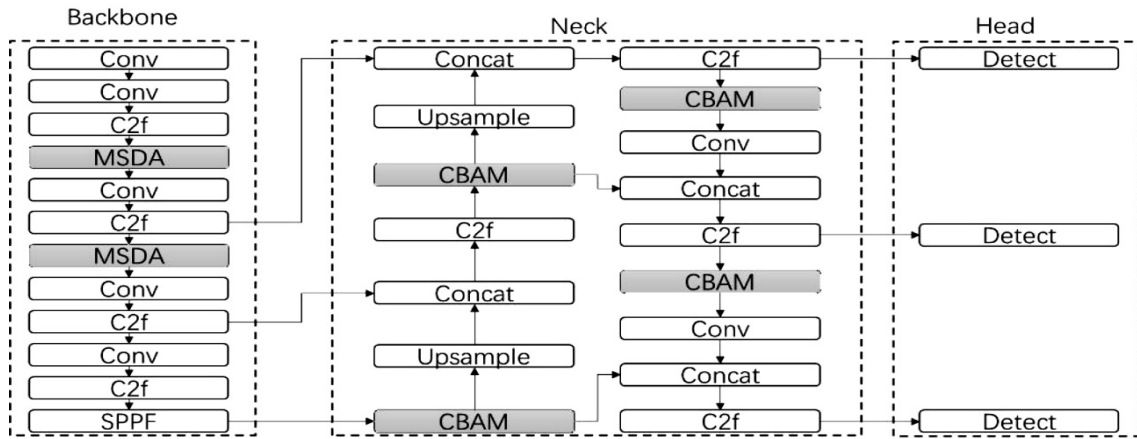


图 4 融合多重注意力机制的改进 YOLOv8 网络结构

该文提出的滑动窗口冗余切割及双尺度模型方法,通过多尺度滑动窗口机制对原始图像进行冗余切割处理,生成两种尺度的子图像,分别输入对应的尺度模型。该方法融合冗余区域内的单一模型结果和不同模型之间的结果。常规的融合算法如非极大值抑制(NMS^[17])在处理多模型和多切片检测结果时效果不佳,且固定切割方式可能导致长目标断裂。此外,图纸各个类别具有非常明显的领域先验,但常规算法未能加以利用。为此,提出了基于 WBF 权重框融合与领域先验的自适应融合算法,综合考虑了置信度和重叠程度,通过加权平均保留高质量的预测框。该算法通过置信度指标确定聚类中心和各预测框在融合中的贡献程度,以及聚类指标确定哪些预测框属于同一簇。该文根据先验为预测结果融合添加预处理部分,并动态计算各类别的置信度和聚类指标,从而优化 WBF 的表现。

该文总结了图纸中的领域先验,如目标尺寸、密集程度和类别特点,具体为:子图边缘的预测结果易断裂;单元标牌尺寸较大,密度低,易断裂;左右线缆尺寸大;密度多变,易聚集误断;线缆路径指示符尺寸小,密度多变,易误断。

基于上述领域先验,将图纸预测结果融合分为预处理和分类融合两步。预处理中,根据图纸切割特性和类别特性处理预测框。在分类融合中,根据各类别自适应计算的指标后进行融合。具体指标计算如表 1 所示。

表 1 分类融合指标

类别	置信度指标	聚类指标
单元标牌	$W_{\text{预}} \times H_{\text{预}}$	0.15
线缆	预测框得分	0.3
路径指示符	$\frac{H_{\text{线缆}}}{\frac{W_{\text{预}} + H_{\text{预}}}{2} - H_{\text{线缆}}}$	0.2

算法细节如下所示:

算法 1:基于领域先验的自适应融合策略算法

输入:子图上预测结果集合 P,子图外圈区域 O,线缆预测平均高度阈值 T_h ,表 1 中的置信度指标 C,聚类指标 M

输出:最终融合的预测结果集合 F

#预处理预测结果

1: Initialize: C, C_i, M_i

2: for p in P do:

3: if not intersects(p, O) then:

4: if $0.7 * T_h \leq \text{height}(p) \leq 1.3 * T_h$ then:

5: $\text{map_to_original_size}(p)$

6: add p to C

#排序

7: sort C by confidence indices C_i

#初始化列表

8: Initialize G = [], F = []

#聚类

9: while C is not empty do:

10: $\text{cluster_center} = C[0]$

11: $S = [\text{cluster_center}]$

12: remove cluster_center from C

13: for candidate in C do:

14: if is_similar(cluster_center, candidate, M)

then:

15: add candidate to S

16: remove candidate from C

17: add S to G

#生成最终预测框

18: for S in G do:

19: $x1 = \text{average}(x1 \text{ of all predictions in } S)$

20: $y1 = \text{average}(y1 \text{ of all predictions in } S)$

21: $x2 = \text{average}(x2 \text{ of all predictions in } S)$

22: $y2 = \text{average}(y2 \text{ of all predictions in } S)$

23: add (x1, y1, x2, y2) to F

24: return F

2.3 基于文本识别与领域先验的图纸拓扑匹配与纠错拒识

2.3.1 文本识别

图纸文本识别是后续实现图纸智能解析和拒识的关键步骤之一。为了精确识别文本,该文舍弃了以 Attention OCR^[18]方法为代表的单阶段端到端文本识别网络,选择了 CRNN^[19](Convolutional Recurrent Neural Network)为主要识别网络的双阶段方法。双阶段文本识别网络可以按顺序分为文本区域检测、文本识别两个步骤。

由于该文任务中文本区域方向类别较少,重要识别对象多为水平垂直,因此文本区域检测网络使用了和图元检测相同的改进后的 YOLOv8 模型替代传统的 DBNet 等场景文字检测模型并将文本区域检测任务并入前文所述图元检测任务中。文本识别网络使用现有 CRNN+CTC^[20]识别框架。

2.3.2 基于领域先验的图纸线帽语义标签拓扑匹配

为完成对图纸的智能识别,该文设计了一套基于领域先验的图纸智能拓扑匹配算法,通过识别线缆路径上的关键目标从而将线缆抽象识别为各条线缆路径以获取所需的各项线缆信息,具体算法分为:(1)匹配关联端子号-单元标牌。根据单元标牌及端子号的纵向位置关系形成“单元标牌-端子号”集合;(2)线缆横向拓扑配对。从线缆出发水平寻找可达的最近端子号与回路编号,形成“单元标牌-端子号-线缆-回路编号”集合;(3)线缆纵向拓扑配对,其具体算法如下所示:

算法 2:线缆纵向拓扑匹配算法

输入:线缆位置集合 C_{pos} , 线缆方向集合 C_{direct} , 线缆路径指示符集合 Sign, 文本集合 Text

输出:线缆名称列表 N

#初始化线缆名称列表 N

1: Initialize: N

#遍历线缆,查询线缆路径指示符集合,直到找到结束标记前不断更新当前位置和方向

2: for (p_i, d_i) in (C_{pos}, C_{direct}) do:

#初始化是否找到结束指示符或下一个指示符的标记

3: Initialize findEnd=False, findNext=True

4: while not findEnd and findNext do:

5: findEnd, findNext, $p_i, d_i \leftarrow \text{findNearestSign}(p_i, d_i, \text{Sign})$

6: end while

#若找到结束标记则根据结束位置与方向寻找线缆名称,若没有则线缆名称为空

7: if findEnd do:

8: $N_i \leftarrow \text{findEndText}(p_i, d_i, \text{Text})$

9: else do:

10: $N_i \leftarrow \text{None}$

11: end for

12: return N

2.3.3 基于领域先验的图纸线帽语义标签纠错拒识算法

完成拓扑匹配后,得到初步的线缆信息可能存在错误,需要对结果进行纠正与拒识。

领域先验可总结为:线缆宽度基本相同,相邻线缆顺序排列;线缆名称与规格在图纸中相邻,文本识别错误多因错误粘连;回路编号常含冒号等干扰信息;线缆名称由大写字母和数字组成;同一张图纸上,端子号由相同规律字符组成;单元号在出现特定数字字母组合时易产生文本识别错误,如 I 识别成 1。

基于此,该文提出纠正与拒识策略,包括:(1)横向纠正:若端子号未被正确识别,同一单元下,算法将搜索相邻已知端子号的线缆并计算距离,推算当前的端子号。(2)纵向修正:匹配识别的线缆名称,匹配成功则删除线缆规格。(3)回路编号拒识:检查回路编号是否匹配正则表达式,匹配则拒识,标记异常字符或同单元下重复编号。(4)线缆名称拒识:标记包含异常字符或无法识别的线缆名称。(5)线缆单元号、端子号拒识:标记不符合规律的端子号。

通过拓扑配对线缆信息的初步结果与基于领域先验的纠正与拒识,完成了对图纸的智能识别,且确保了结果的可信度。

3 实验设置与结果分析

3.1 图元检测性能评价

3.1.1 图元检测数据集构建

变电站端子排图纸多尺度目标检测数据集由南京邮电大学模式识别与机器学习实验室联合江苏省送变电有限公司构建,均来源于真实场景,包括宿迁闻涛 220 千伏变电站新建工程、江苏扬州-镇江直流输电工程少游±200 kV 换流站工程、虞城(姑苏)±800 kV 换流站工程、江苏省送变电有限公司电建路 108 号办公楼北侧变电站安装实训场改造工程。该文从送变电公司提供的数据中选取了 354 张图纸构建数据集,最大图纸尺寸为 7 063 * 17 421 像素,最小图纸尺寸为 2 532 * 1 805 像素,平均尺寸 3 420 * 5 458 像素。并基于领域先验将图中检测实例划分为 10 类电气专业图元(不含文本图元),使用 LabelImg 标注软件进行水平框标注,所有图纸均由专业人员按照标注规范进行严格标注。

数据集包含 15 633 个实例,图像均未经切割。该数据集共 354 张图像,按 8 : 2 的比例划分成训练集和测试集。

3.1.2 实验设置与评价指标

实验在 Pytorch2.0.0 环境下进行,系统环境为

Windows 10, CUDA 版本为 12, 使用了 NVIDIA GeForce RTX 3090 GPU, 其显存大小为 24 GB。大尺度和小尺度目标检测模型分别使用经过多尺度滑动窗口切割的数据集的大、小尺寸切片进行训练。每个模型训练进行 300 轮, 早停策略为 50 轮。初始学习率为 0.01, 训练过程中衰减至最终学习率 0.000 1, 动量参数为 0.937, 权值衰退系数为 0.000 5。数据增强方面主要采用了 Mosaic^[21]方法, 关闭了左右翻转数据增强。模型的最大预测数量设为 1 000, 并在训练的最后十轮关闭了数据增强。所有模型平均训练至收敛, 平均收敛轮数为 120 轮。

实验采用常见的目标检测指标, 包括召回率 (R)、精确率 (P) 和平均精度 (AP)。其中, AP 表示精确率-召回率 (PR) 曲线下的面积, 平均精度均值 (mAP) 则是各类别 AP 的平均值。mAP 的 IoU (交并比) 阈值设置为 0.5。

3.1.3 实验结果分析

首先探讨图元检测步骤中所使用的 YOLOv8 目标检测模型的有效性, 横向对比目前较为主流的目标检测模型。进行对比实验时, 所有模型均使用前文提及的多尺度滑动窗口冗余切割与图元切片自适应融合算法。实验结果如表 2 所示。

表 2 图元检测基准模型对比实验 %

模型	P	R	mAP@0.5
Faster R-CNN	83.6	86.0	86.7
DETR	91.2	91.6	92.1
YOLOv6	88.9	90.6	89.8
YOLOv7	91.3	90.0	93.3
YOLOv8	89.5	90.4	92.6
YOLOv8+ 多重注意力	90.7	92.5	94.3

不难看出, YOLO 系列为代表的单阶段目标检测方法全面领先于两阶段目标检测方法, 与以 Transformer 架构为基础的 DETR 表现较为接近。与目前主流的检测模型相比, 结合多重注意力的 YOLOv8 在召回率及平均精度均值上取得了最佳性能, 且图纸图元检测对于此性能表现要求较高, 故选取结合多重注意力的 YOLOv8 作为文中模型。

该文继续验证了 MSDA 和 CBAM 注意力机制的有效性。具体实验中, 分别讨论了注意力机制对不同尺度模型的影响效果。注意力机制对比实验的结果如表 3 所示, 相较于不使用注意力机制或仅单独采用一种注意力机制, 将两种注意力机制叠加使用能够达到最佳的检测性能。观察 MSDA 与 CBAM 的具体检测数据可以发现, MSDA 的整体表现优于 CBAM。通过

多尺度空洞卷积, MSDA 能够更好地捕捉更多的空间信息, 改善前文提及的多尺度图像差异, 提高召回率。相反, CBAM 的整体表现略低, 但它更能够捕捉细节信息, 优化了前文提到的小目标问题, 提高了精确率。

表 3 注意力机制对比实验结果 %

模型	注意力机制	P	R	mAP@0.5
小尺度	无	84.9	86.8	90.4
小尺度	MSDA	86.0	91.9	91.8
小尺度	CBAM	86.1	91.6	90.9
小尺度	MSDA+CBAM	87.5	92.3	92.1
大尺度	无	87.4	79.4	85.3
大尺度	MSDA	87.7	79.6	86.0
大尺度	CBAM	87.5	79.5	85.7
大尺度	MSDA+CBAM	87.6	80.0	86.4

为了验证基于领域先验的自适应融合算法的有效性, 采用前文训练得到的最佳双尺度模型。在原数据集上, 使用双尺度模型进行双尺度滑动窗口切割预测, 然后对预测结果进行不同算法的融合, 并统计相应的指标。实验结果如表 4 所示。相较于单尺度模型, NMS 算法对双尺度模型的简单融合略微提升了性能。但由于图纸目标具有领域先验特性和切割算法的影响, 断裂目标对原图纸的预测产生干扰, 提升幅度有限。同理, WBF 也难以提高模型表现, 甚至降低了召回率。该文提出的自适应融合算法考虑了领域先验, 并对多尺度预测进行了特殊处理, 因此表现出良好的性能。引入领域先验自适应融合后, 由于置信度指标调整为与线缆高度的相对误差及绝对大小, 两种类型的目标预测得到了有效修正。

表 4 多尺度模型融合算法实验结果 %

融合算法	P	R	mAP@0.5
NMS	89.9	90.5	92.6
WBF	90.1	90.0	92.7
自适应融合算法	90.7	92.5	94.3

3.2 线帽语义标签解析性能评价

3.2.1 线帽语义标签解析数据集构建

线帽语义标签解析数据集来源于图元检测数据集相同。为确保实验的公平性, 该文选取了与图元检测无交集的 400 张典型图纸, 共包含 8 677 根线缆实例, 平均每张图纸包含 22 根线缆, 并按 8 : 2 的比例划分为训练集和测试集。为验证智能解析精度, 按照要求的最终输出文档, 人工识别图纸后以单根线缆信息为单位制作最终解析文档。数据集中每个文档包含该图纸中每根线缆的所有信息, 包括端子号、单元标牌号、回路编号及线缆名称。

3.2.2 实验设置与评价指标

实验使用常见的算法指标,即召回率(R)、精确率(P)。由于引入纠错与拒识机制,故新增指标:拟召回率(Q_r)、拟精确率(Q_p)、拒识率(R_i)。五种算法指标具体公式如下:

$$P = TP / (TP + FP + RS) \quad (1)$$

$$R = TP / (TP + NP) \quad (2)$$

$$R_j = RS / (TP + FP + RS) \quad (3)$$

$$Q_p = (TP + RP) / (TP + FP) \quad (4)$$

$$Q_r = (TP + RP) / (TP + NP) \quad (5)$$

其中,TP 表示析出的真正例数(解析结果正确的线帽标识样例数),FP 表示析出的假正例数(解析结果错误的线帽标识样例数),RS 表示拒识的样例数(解析结果需交由人工复核的线帽样例数),NP 表示未析出的正例数(未能有解析结果的线帽标识样例数),RP

表示拒识的正例数。

3.2.3 实验结果分析

为了验证图纸智能解析及纠错拒识算法的有效性,在实验过程中,对比了使用纠错与拒识算法与常规拓扑匹配的性能差异。消融实验结果如表 5 所示。常规拓扑匹配算法的召回率高达 98.9%,但精确率仅 88.5%,表明漏检现象较少,但由于涉及多种处理流程,导致一整行信息完全正确的情况较少。引入纠错算法后,文本粘连和错框现象大幅度减少,极大增强了算法的可用性。引入拒识算法后,通过领域先验规避大部分错误,拟精确率和拟召回率均达到 100%,这表明在引入人工复检的情况下,算法的识别精度与可用性非常优异。拒识算法整体拒识率为 10.8%,在平均每张图纸包含 22 根线缆的情况下,每张图纸需人工复检两根线缆,达到了较为理想的智能解析效果。

表 5 纠错与拒识算法消融实验 %

常规拓扑匹配	纠错算法	拒识算法	P	R	Q_p	Q_r	R_j
√	-	-	88.5	98.9	-	-	-
√	√	-	91.6	99.7	-	-	-
√	√	√	91.6	99.7	100.0	100.0	10.8

最后探讨了改进图元检测对算法整体性能的影响,通过消融实验对比了各项改进对最终解析精度的影响,结果如表 6 所示。数据表明,图纸解析任务受尺度影响极大。不使用多尺度策略时,精确率及召回率均不理想;引入多尺度滑动窗切割算法后召回率显著提升,能检测出更合理的图元。多重注意力机制通

通过对尺度及细节的关注,同样在一定程度上提升了模型表现。自适应融合算法相比传统 NMS,能显著提高解析线缆的数量。尽管添加了这些预处理、模型及后处理改进,由于图纸本身的复杂特性,指标依旧不能达到完全理想的状态,因此纠错与拒识算法至关重要。

表 6 不同图元检测策略对整体解析结果影响 %

多尺度滑动窗冗余切割	多重注意力	自适应融合	P	R
-	-	-	84.6	90.1
√	-	-	86.9	96.8
√	√	-	87.3	96.9
√	√	√	87.8	97.7

4 结束语

针对图纸差异大、小目标识别难度高以及图纸解析精度挑战等问题,提出了一套融合图元检测和领域先验的变电站端子排图纸电气语义解析方法,通过真实数据验证了算法的有效性。在图元检测阶段,采用滑动窗口机制切割大尺度图片,但带来了割裂目标和难以设定通用尺度等问题。因此,设计了多尺度滑动窗口机制,并引入基于领域先验的融合方法和注意力机制解决小目标问题及切片融合难题。在智能解析与拒识阶段,结合图纸领域先验和正则表达式、纠正与拒

识算法提升了识别结果的精确率与召回率,使模型具备智能提示、拒识并标记的能力。为进一步提升算法的精确率及可用性,可关注数据中的样本不平衡,通过损失函数对各类别进行区分,提升模型在少见目标上的表现。

参考文献:

[1] 白浩,周长城,袁智勇,等. 基于数字孪生的数字电网展望和思考[J]. 南方电网技术,2020,14(8):18-24.
 [2] 罗东亮,蔡雨萱,杨子豪,等. 工业缺陷检测深度学习方法综述[J]. 中国科学:信息科学,2022,52(6):1002-1039.
 [3] REDMON J, DIVVALA S K, GIRSHICK R B, et al. You only

- look once: unified, real-time object detection [J]. CoRR, 2015, abs/1506.02640.
- [4] 戴威, 陆小锋, 钟宝燕, 等. 一种基于视觉分析的指针式仪表智能抄读方法[J]. 计算机技术与发展, 2023, 33(1): 200-205.
- [5] HSU W Y, LIN W Y. Ratio-and-scale-aware YOLO for pedestrian detection [J]. IEEE Transactions on Image Processing, 2020, 30: 934-947.
- [6] 李利霞, 王鑫, 王军, 等. 基于特征融合与注意力机制的无人机图像小目标检测算法[J]. 图学学报, 2023, 44(4): 658-666.
- [7] 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.
- [8] 秦辞海, 顾万里. 基于深度学习的变电站钢结构图纸标题栏文字检测与识别[J]. 土木工程信息技术, 2022, 14(2): 110-115.
- [9] VAN ETEN A. You only look twice: rapid multi-scale object detection in satellite imagery [J]. arXiv:1805.09512, 2018.
- [10] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu: IEEE, 2017: 2117-2125.
- [11] 褚雪汝, 陈中, 吴聪颖, 等. 基于深度学习的电气二次图纸语义识别方法[J]. 浙江电力, 2023, 42(8): 1-11.
- [12] 程鑫, 褚雪汝, 邓旭晖, 等. 基于双层-分块检测网络的厂站接线图纸图符检测方法[J]. 东南大学学报: 自然科学版, 2022, 52(6): 1137-1144.
- [13] SOLOVYEV R, WANG W, GABRUSEVA T. Weighted boxes fusion: ensembling boxes from different object detection models [J]. Image and Vision Computing, 2021, 107: 104117.
- [14] JOCHER G, CHAURASIA A, QIU J. Ultralytics YOLO (Version 8.0.0) [EB/OL]. 2023. <https://github.com/ultralytics/ultralytics>.
- [15] JIAO J, TANG Y M, LIN K Y, et al. Dilateformer: multi-scale dilated transformer for visual recognition [J]. IEEE Transactions on Multimedia, 2023, 25: 8906-8919.
- [16] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module [C]//Proceedings of the European conference on computer vision. Munich: ECCV, 2018: 3-19.
- [17] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression [C]//18th international conference on pattern recognition (ICPR'06). Hong Kong, China: IEEE, 2006: 850-855.
- [18] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(11): 2298-2304.
- [19] ZHANG J, WANG W, HUANG D, et al. A feasible framework for arbitrary-shaped scene text recognition [J]. arXiv:1912.04561, 2019.
- [20] YAN H, HE Q, XIE W. CRNN-CTC based mandarin keywords spotting [C]//ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). Barcelona: IEEE, 2020: 7489-7493.
- [21] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [J]. arXiv:2004.10934, 2020.