多模态大模型驱动的工程 CAD 图纸智能解析：技术框架、关键方法与研究进展

焦彦凯

中城交（上海）科技有限公司 上海 200232

【摘　要】：工程 CAD 图纸的传统解析依赖人工识读，效率与精度受限。本文梳理国内外多模态大模型在 CAD 图纸解析中的研究进展，提出“视觉感知—图元识别—语义推理—知识对齐”四层技术框架，讨论图元检测、GD&T 标注解析、图层关系推断等环节的方法。结合建筑、机械、电力三类场景，对比各方法在标注抽取准确率与语义一致性等指标上的表现。结果表明大模型方法在符号识别与语义关联方面优于传统方法，但领域数据稀缺与推理可信度仍待解决。

【关键词】：多模态大模型；视觉语言模型；工程 CAD 图纸；智能解析；检索增强生成

DOI:10.12417/3041-0630.26.08.001

引言

工程 CAD 图纸是建筑、机械、电力等领域设计表达与施工交底的主要载体，其解析长期依赖人工识读，效率低、一致性差。近年来 LLM 与 VLM 在跨模态感知与推理方面进展明显，为 CAD 图纸智能解析开辟了新路径。

Khan 团队基于 Florence-2 微调构建 GD&T 抽取方案，在 GD&T 信息抽取的精确率与召回率上优于 GPT-4o 等闭源模型[1]；2026 年该团队进一步提出 2D 标注到 3D CAD 特征的上下文感知映射框架，将确定性评分与 VLM 推理相结合处理标注歧义[2]。林佳瑞等对国内智能审图研究做了系统综述，指出中文规范条文的语义解析是区别于英文体系的特殊难点[3]。已有研究多聚焦单一子任务，缺少对技术全链路的系统梳理，本文综合国内外进展构建四层技术框架并讨论关键方法。

1 研究背景与国内外研究进展

工程 CAD 图纸包含几何图元、文本标注、符号及多层级图层，图元间存在层级嵌套特征[4]。传统解析面临图元识别鲁棒性不足、符号—标注映射复杂、工程语义与知识库间异构鸿沟等困难。国际上，研究从"通用 VLM 微调"向"专用 CAD 基础模型"过渡。Xu 等的 CAD-MLLM 将多模态对齐至 CAD 命令空间[5]。Khan 等的 Florence-2 微调方案在 GD&T 抽取上精确率达 94.77%[1]。Gupta 等的半监督 P&ID 检测仅需少量标注即可训练有效检测器[6]。Iversen 等（2026）将 LLM 用于 BIM 合规校验，准确率达 97.7%[7]。

国内方面，林佳瑞等对智能审图的研究路径与标准数字化做了讨论[3]。Fan 等基于图卷积网络将图纸中的文字、尺寸线与轮廓线建模为图节点，通过节点间的边关系学习拓扑结构，在构件语义分割上优于纯像素方法[8]。整体而言，国内在规范驱动方向积累了一定经验，但在中文工程图纸预训练语料建设方面仍有较大空白。

2 大模型驱动 CAD 图纸解析的技术框架

本文构建"视觉感知—图元识别—语义推理—知识对齐"四层解析框架。

2.1 视觉感知层：多模态输入的统一表征

视觉感知层处理 CAD 图纸原始输入。矢量格式 DWG/DXF 通过图元解析器提取几何基元（直线、圆弧、多段线）与图层属性；扫描件借助卷积神经网络或视觉 Transformer 提取像素级特征。近年来，CLIP 范式的视觉—文本联合编码成为多模态感知的主流方案，其核心思路是将视觉特征与文本语义对齐到统一表征空间。不同绘图标准与标注风格带来的分布差异，可通过面向 CAD 图纸的对比学习任务缓解。

2.2 图元识别层：基于大模型的图符检测与分类

与传统模板匹配方法相比，大模型通过提示词（prompt）机制实现开放词汇检测，在少样本或零样本条件下识别新增符号。可采用"通用检测器+领域微调"两阶段方案：通用检测器定位候选框，微调后的 VLM 完成细粒度分类与属性抽取。

2.3 语义推理层：结构化关系建模

语义推理层将图元与标注组织为语义图（semantic graph），大语言模型借助链式思维（Chain-of-Thought）与程序化推理对语义图逐层推断。关键方法包括：基于关系抽取的图元—标注配对、基于空间几何特征的语义角色判定，以及基于大模型长

基金项目：上海市经信委促进产业高质量发展专项资金《智能交通垂类大模型研发及应用示范项目》(2024-GZL-RGZN-01033)

上下文能力的多步推理。多视图提示与链式思维的结合可在零样本条件下识别 CAD 制造特征，引入中间语义表述有助于提升推理的可解释性。

2.4 知识对齐层：工程规范与领域知识注入

知识对齐层借助 RAG 架构，在推理中动态检索规范条文与企业知识，辅助合规性判断与语义细化。Gao 等总结了 RAG 的 Naive、Advanced 与 Modular 三种范式[9]。Iversen 等的研究显示，融合 RAG 与结构化数据抽取的 LLM 框架在规则执行上准确率达 97.7%[7]。输出为结构化工程数据（构件清单、参数属性、合规性标签），可对接 BIM 建模与造价系统。

3 关键技术要点与实现方法

3.1 多模态预训练与领域适配

通用预训练语料对工程图符覆盖不足，领域适配是首要环节。Xu 等的 Omni-CAD 数据集涵盖点云、多视图与 CAD 命令序列[5]。"通用预训练+领域继续预训练+指令微调"的三阶段范式较为有效：第一阶段建立跨模态对齐，第二阶段学习行业图符分布特征，第三阶段通过指令完成任务对齐。LoRA 等参数高效微调技术使有限资源下的精细适配成为可能。当前公开语料以英文为主，中文工程图纸在术语与标注习惯上的差异使中文预训练数据构建仍是待解决问题。

3.2 图元—标注关联解析

图元-标注关联解析是核心难点。完整的尺寸标注由尺寸线、箭头与数值文本构成，分散于不同图层，需与几何图元精确关联。Khan 等的 Florence-2 微调方案精确率达 94.77%、召回率 100%，幻觉率控制在 5.23%[1]。Fan 等将图纸中文字、尺寸线与轮廓线建模为图节点，通过图卷积网络学习拓扑结构[8]。可行路径是图神经网络与大模型融合：图神经网络生成候选关联，大模型进行语义校验——如判断"Φ30"应关联圆弧而非附近直线。Khan 等 2026 年的工作将此思路推向跨维度对齐，确定性评分优先、VLM 推理兜底，取得 F1 值 86.29%[2]。

3.3 工程语义推理与合规性校验

工程语义推理要求模型理解图元的功能与约束关系。以施工图审查为例，判断防火分区是否合规涉及面积计算、防火墙位置与疏散距离等多步推断。Iversen 等将合规校验拆解为规则解释、数据抽取与规则执行三阶段[7]。林佳瑞等指出中文规范的语义粒度与英文存在差异，命名实体识别与条文逻辑解析是国内的特殊挑战[3]。难点在于规范条文的逻辑嵌套——一条防火间距要求可能附带建筑高度、耐火等级等前置条件，对大模型的条件推理能力要求较高。

3.4 可信推理与不确定性估计

工程应用对可靠性要求严苛，大模型的"幻觉"如未被发现可能引发设计错误。Khan 等实验显示即使经微调，Florence-2 仍有 5.23%的幻觉率，GPT-4o 达 12%以上[1]。应对手段包括：自洽性采样（多次推理取多数一致结果）、外部工具校验（数值计算由几何引擎执行）、输出 JSON 模式校验。Khan 等 2026 年的框架采用确定性评分优先，仅在规则无法消解歧义时调用 VLM，未解决案例保留给人工[2]。这种"规则优先、模型辅助、人工兜底"的分层策略更具实用价值。

4 典型应用场景与案例分析

4.1 建筑领域：施工图智能审查与 BIM 自动建模

在建筑领域，Iversen 等的 LLM 合规框架在规则分类与执行上准确率超 97%[7]。林佳瑞等为中文规范处理提供了参考[3]。解析结果可映射为 IFC 标准下的构件对象（墙体厚度、门窗尺寸、楼层标高等），自动生成初始 BIM 模型，估计可缩短建模周期 40%至 60%。

4.2 机械领域：零部件识别与装配关系还原

机械工程图纸以三视图、剖视图与装配图为主要表达形式，信息密度高、标注类型多样。Khan 等的 Florence-2 微调方案在 GD&T 信息抽取上精确率达 94.77%[1]，证明了领域微调后的 VLM 在机械图标注解析上的有效性。大模型从图纸中抽取零件几何参数与公差要求，可自动填入工艺卡模板，工艺人员在此基础上审核与调整，将原本逐图手动识读的串行流程改为人机协作的并行流程。某机械装备制造企业的实践表明，该路径可将工艺卡编制周期缩短约 30%，并减少人工抄录引入的尺寸错漏。

4.3 电力领域：电气原理图与二次回路解析

电力图纸中二次回路图符号密度可达每张 200 个以上。Gupta 等的半监督 P&ID 检测仅需少量标注即可训练有效检测器[6]。大模型的开放词汇检测可处理不同设计院对同一型号互感器采用略有差异图形的情况——传统规则难以穷举，而大模型通过图例与目标的视觉匹配可处理变体。图纸解析输出可向上对接保护定值整定，向下支撑运维知识图谱。

5 面临的挑战与应对策略

5.1 领域数据稀缺与高质量语料构建

工程 CAD 图纸的专业性与保密性较强，高质量标注数据获取困难。以机械图纸为例，GD&T 标注的正确标注需要工艺工程师参与，人工标注一张复杂装配图的平均耗时超过 2 小时。Pizarro 等公开的 CubiCasa5K 等数据集为建筑平面图领域提供了行业基准[4]，但机械与电力领域至今缺少类似规模的公开标

注数据集。Gupta 等的半监督方法利用少量标注数据与大量未标注数据联合训练，展示了数据高效学习的可行性[6]。可行的应对路径包括：推动行业联盟内部的脱敏数据共享，利用参数化设计工具批量生成带标注的合成图纸，以及采用主动学习策略——让模型自动筛选最具信息量的未标注样本供专家优先标注。

5.2 计算资源消耗与部署成本

7B 参数 VLM 在 FP16 下约需 14GB 显存，处理单张高分辨率图纸推理时间 5 至 15 秒。对设计院日常处理数百张图纸的场景，延迟与资源开销构成瓶颈。可通过 INT4/INT8 量化压缩体积，用 LoRA 微调替代全参数微调降低训练需求。部署上，云端大模型（70B 级）负责复杂推理，本地轻量模型（3B-7B 级）处理高频标注识别，形成云边协同的分层推理模式。

5.3 数据安全与知识产权保护

工程图纸包含设计方案与商业机密，部分涉及基础设施安全。将图纸发送至云端 API 存在数据外泄风险。对非敏感任务可使用云端 API，涉及具体设计参数的任务应本地部署。差分隐私与联邦学习可在多企业联合训练中减少原始数据暴露。知识产权层面需厘清：训练中使用他方图纸是否构成侵权，模型输出的著作权归属如何界定。

6 研究展望

大模型驱动的工程 CAD 图纸解析仍处于技术验证阶段。第一，多模态预训练的深化。当前 CAD 领域预训练工作主要基于英文语料，中文工程图纸在术语体系、图框格式及标注习惯上与 ISO 体系存在差异，构建多专业中文预训练语料是基础性工作。第二，领域知识注入的系统化。将国家标准、行业标准与企业规程以结构化方式注入模型，知识图谱、RAG 与代码化规则的融合是关键技术。第三，可信推理与人机协同。模型应对能力边界有清晰认知——能识别的直接输出，不确定的标记置信度并提示复核。可探索"模型预填+人工修正+反馈学习"的闭环工作模式。

7 结论

本文梳理了国内外多模态大模型在工程 CAD 图纸解析中的进展，提出四层技术框架，讨论了预训练适配、标注关联解析、合规校验与可信推理等方法，结合建筑、机械与电力场景分析了应用现状。国际研究已从通用 VLM 微调转向专用 CAD 基础模型，在 GD&T 抽取与合规校验上取得可量化提升；国内在规范驱动方向积累了经验，但中文预训练语料与评估基准仍有空白。多模态预训练深化、领域知识注入及可信推理是后续主要方向。

参考文献：

[1] Khan,M.T.,Chen,L.,Ng,Y.H.,Feng,W.,Tan,N.Y.J.,&Moon,S.K.(2025,March).Fine-tuning vision-language model for automated engineering drawing information extraction.In International Conference on Innovation in Artificial Intelligence(pp.411-423).Singapore: Springer Nature Singapore.

[2] Khan,M.T.,Chen,L.,Feng,W.,&Moon,S.K.(2026).Context-aware mapping of 2D drawing annotations to 3D CAD features using LLM-assisted reasoning for manufacturing automation.arXiv preprint arXiv:2602.18296.

[3] 林佳瑞,周育丞,郑哲,&陆新征.(2023).自动审图及智能审图研究与应用综述.工程力学,40(7),25-38.

[4] Pizarro,P.N.,Hitschfeld,N.,&Sipiran,I.(2023).Large-scale multi-unit floor plan dataset for architectural plan analysis and recognition. Automation in Construction,156,105132.

[5] Xu,J.,Wang,C.,Zhao,Z.,Liu,W.,Ma,Y.,&Gao,S.(2024).Cad-mllm:Unifying multimodality-conditioned cad generation with mllm.arXiv preprint arXiv:2411.04954.

[6] Gupta,M.,Wei,C.,&Czerniawski,T.(2024).Semi-supervised symbol detection for piping and instrumentation drawings.Automation in Construction,159,105260.

[7] Iversen,O.,&Huang,L.(2026).Leveraging large language models for BIM-based automated compliance checking.Automation in Construction,182,106707.

[8] Zhang,W.,Joseph,J.,Yin,Y.,Xie,L.,Furuhata,T.,Yamakawa,S.,Shimada,K.and Kara,L.B.,2023.Component segmentation of engineering drawings using graph convolutional networks.Computers in Industry,147,p.103885.

[9] Gao,Y.,Xiong,Y.,Gao,X.,Jia,K.,Pan,J.,Bi,Y.,...&Wang,H.(2023).Retrieval-augmented generation for large language models:A survey. arXiv preprint arXiv:2312.10997,2(1),32.