科学技术与教育 2026年第10期
ISSN: 3079-4455

YOLO 视觉算法与多模态在雨污水管网视频检测应用研究

胥仁栋
华昕设计集团有限公司，江苏无锡，中国

【摘要】针对雨污水管网视频人工判读效率低、小目标缺陷易漏检、同一缺陷跨帧重复标记等问题，结合 YOLO 视觉算法与多模态大模型构建了一套两阶段管网缺陷检测系统。第一阶段采用 YOLOv12m 对视频抽帧进行 16 类缺陷纯识别并入库；第二阶段由 C#后台异步调用私有化部署的千问 VL 多模态模型对检测结果进行语义二次判别，进一步筛选缺陷内容。实验表明该系统 mAP50 达到 72.36%，支管暗接、残墙坝根、沉积等缺陷类别召回率均超过 0.85，同时通过阶段式解耦有效消除了千问复核对 YOLO 推理的串行拖累提升整体系统稳定性。系统已在真实管网检测场景部署应用，显著提升了检测效率与报告质量。

【关键词】YOLOv12；多模态大模型；管道缺陷检测；千问 VL；两阶段架构；雨污水管网

1.引言

城镇雨污水管网是城市基础设施的重要组成部分，随着城镇化进程加快，管网规模持续扩大、管网不断老化，破裂、变形、腐蚀、渗漏、错口、沉积等缺陷频发，直接影响排水系统安全运行与内涝防治能力，亟需对管网资产开展科学高效的检测与养护管理[1,2]。传统的管道检测多依赖 CCTV 机器人进入管内录像后由人工逐帧判读，劳动强度大、判读结果依赖人员经验、效率低下，难以满足大规模管网检测的需求[3,4]。

近年来，深度学习与机器视觉技术的快速发展为缺陷自动检测提供了新的途径。以 Faster R-CNN、SSD、YOLO 等为代表的深度学习目标检测算法被广泛应用于管道缺陷识别，国内外学者相继提出了多种面向管道缺陷的检测模型，在检测精度与速度上取得显著进步[5]。其中，针对渗漏、腐蚀、错口等典型缺陷的轻量化检测模型不断涌现，YOLOv5-Sewer 等针对污水管网场景优化的模型在保证精度的同时有效控制了参数量与计算量。

综合现有研究可以看出，当前管道缺陷检测方法大体沿两条路径发展：一是以 Faster R-CNN、YOLO 等为代表的目标检测路径，侧重缺陷的定位与分类；二是以多模态大模型为代表的语义理解路径，侧重对缺陷的距离、方位、等级等工程属性进行推理。然而，两条路径在实际部署中往往相互割裂：目标检测模型缺乏语义理解能力，多模态大模型直接处理视频帧则推理速度慢、训练成本高，难以满足大规模管网的实时检测需求。如何在工程系统中将二者有机协同，兼顾检测效率与语义理解深度，可作为一个应用研究的方向。

然而，现有方法在实际管网检测中仍面临三重挑战：其一，裂缝、渗漏等小目标缺陷在低分辨率影像中易漏检；其二，摄像机在管内匀速前移过程中，同一缺陷会被连续多帧重复标记，产生大量冗余结果；其三，单一检测模型难以有效的完成“检测”与“语义理解”双重任务[6]。

为此，本文结合 YOLO 视觉算法与多模态大模型，面向雨污水管网视频检测构建了一套两阶段智能检测系统：第一阶段由 YOLOv12m 完成缺陷快速纯识别并入库将缺陷尽可能入库；第二阶段由 C#后台异步调用私有化部署的千问 VL 多模态大模型对检测结果进行语义二次判别，针对 YOLO 模型前期数据集偏少召回率低的缺陷进行复核，同时基于物理里程去重消除跨帧重复标记[7,8]。该系统目前已在真实管网检测场景部署应用。

2.系统总体架构与相关技术

2.1YOLO 目标检测算法

YOLO 系列算法属于单阶段目标检测方法，将目标检测视为回归问题，直接在整幅图像上预测边界框与类别，兼顾了检测精度与实时性，相比两阶段的 Faster R-CNN 在速度上具有显著优势。随着残差网络与自注意力机制的引入，检测网络的特征提取能力不断增强。YOLOv12 采用了以注意力机制为核心的架构，具有更大的有效感受野，能够更好地利用图像全局上下文信息，对低分辨率影像中的小目标缺陷具有更强的检测能力；同时其模型规

模适中，可在 24GB 显存级别的 GPU 上完成训练与推理，适合工程化部署。考虑到管网缺陷中裂缝、渗漏等小目标占比较高，本文选用 YOLOv12m 作为基础检测模型[9]，在检测精度与计算开销之间取得平衡。

2.2 多模态大模型

多模态大模型能够联合理解图像与语言，具备强大的视觉-语义推理能力。CLIP 通过大规模图文对比学习实现了视觉与语言的对齐，GPT 系列大语言模型展现出少样本条件下的强大推理能力，以 VisionLLM 为代表的通用多模态大模型可端到端处理多种视觉-语言任务。多模态大模型在缺陷检测与评估中显示出良好潜力，在缺陷描述、等级评估、场景语义理解等方面可对传统检测结果进行有效补充。

2.3 系统总体架构

本文构建的雨污水管网缺陷检测系统采用两阶段总体架构，如图 1 所示。第一阶段由 YOLOv12m 完成视频抽帧图像的缺陷纯识别并入库，检测结果仅含缺陷类别与置信度等基础信息；第二阶段由 C#后台异步调用私有化部署的千问 VL 多模态模型对置信度低的检测结果进行语义二次判别，并识别缺陷类别、物理距离、时钟表示、是否管道内及缺陷等级等语义信息，最后按物理里程去重。

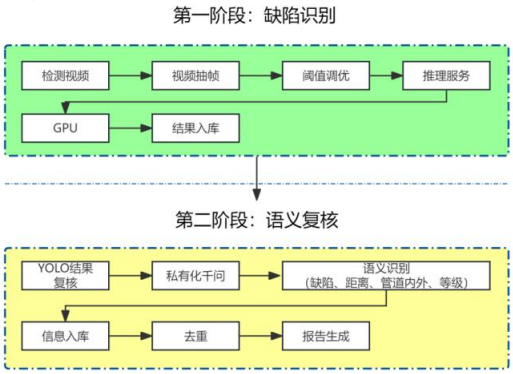


图 1.雨污水管网缺陷检测系统两阶段总体架构

3.YOLO 检测模型设计

3.1 数据与标注

系统采集了雨污水管道 CCTV 巡检视频，按固定帧间隔抽取图像构建训练样本，并依据城镇排水管道检测与评估相关规范对缺陷进行标注，覆盖破裂、变形、腐蚀、错口、渗漏、残墙坝根、沉积、障碍物等 16 类缺陷。由于各类缺陷在管网中出现的频次差异较大，样本存在明显的类别不平衡问题，部分缺陷类别样本数量较少，个别类别（如浮渣）样本极少且特征混杂，在训练中被剔除以避免干扰。标注过程中，对裂纹、渗漏等小目标缺陷采用放大复核方式保证边界框质量。

3.2 模型选型与训练策略

综合检测精度、小目标适应性与部署成本，本文选用 YOLOv12m 作为检测模型。针对类别不平衡，在训练中采用 copy-paste、mixup、mosaic、close_mosaic、label_smoothing 等数据增强策略，并对少样本类别进行针对性的样本增强补偿，而不对多数类别做过采样平衡——因为过度过采样会导致类别分布偏移，实验中曾出现过采样后 mAP50 明显退步的情况。训练采用 300 轮、输入尺寸 1024×1024，训练及推理在 2 块 RTX 4090 上完成。单卡批量设为 8 以保证显存安全。

3.3 类别级阈值与推理部署

针对不同缺陷类别检测置信度差异较大的问题，系统支持按类别独立配置置信度阈值，通过任务级配置与全局兜底相结合的方式下发，使低置信度高价值的缺陷（如渗漏）不易被漏检，同时抑制高置信度误报。视频推理按可调的帧间隔抽帧，兼顾检测覆盖度与处理效率。训练与推理的关键配置参数如表 1 所示。

表 1.训练与推理关键配置

配置项	参数值
基础模型	YOLOv12m（注意力架构）
训练轮数/输入尺寸	300 轮 / 1024×1024
单卡批量（安全上限）	8（12 会触发 OOM）
训练设备	2×RTX 4090
数据增强	copy_paste/mixup/mosaic/close_mosaic/label_smoothing
检测类别	16 类管道缺陷
帧抽帧间隔	frame_interval=30（可调）
当前最佳检测指标	mAP50 = 62.36%（177 轮双卡）

4.基于多模态的两阶段二次判别架构

4.1 问题分析

传统方案在同一推理进程中同步调用千问模型完成缺陷识别与语义理解，由于千问推理速度慢且易超时，会严重拖累 YOLO 检测的整体吞吐，导致检测任务长时间无法完成。同时，摄像机在管内匀速前移时，同一缺陷会在连续多帧中重复出现并被重复标记，产生大量无效冗余结果，给后续报告编制带来困扰。

4.2 两阶段解耦设计

为解决上述问题，系统将“检测”与“语义理解”解耦为两个独立阶段。第一阶段由 YOLOv12m 完成缺陷纯识别并入库，推理过程中完全不调用千问，保证检测吞吐；第二阶

段由 C#后台在推理完成后异步触发复核循环，逐批取出未复核的检测结果，调用千问 VL 进行语义二次判别。该设计使 YOLO 推理与千问复核并行进行，消除了相互拖累。

4.3 千问 VL 语义识别

复核阶段读取检测结果的原始图像，编码为 base64 后发送至私有化部署的千问 VL 模型，通过结构化提示词引导其输出缺陷类别、物理距离（QWJL）、时钟表示（SZBS）、是否处于管道内（Gdn）及缺陷等级等工程属性。模型返回结果经解析后回写至检测结果记录，并以千问识别的缺陷类别为准完成缺陷代码映射，从而修正 YOLO 在复杂场景下的类别误判。对于千问解析成功但未识别出明确缺陷的结果，系统将其缺陷类别标记为“无”，避免该类样本反复进入复核队列。为抑制模型推理失败导致的复核中断，系统设置了超时与重试机制，并将复核失败的样本暂时跳过、等待后续重试，避免单条异常阻塞整体复核流程。

4.4 物理里程去重与筛选

针对跨帧重复标记问题，系统不以跨帧 IOU 作为主要去重依据（镜头旋转会导致同一缺陷的 IOU 剧烈变化而失效），而是以缺陷的物理里程距离（QWJL）作为主要去重依据：对同一缺陷类别、物理里程距离接近的检测结果进行合并去重，仅保留最具代表性的结果，并通过 IsShow 标记控制缺陷的显示与隐藏，供人工复核时确认。

4.5 并发与性能优化

为提升复核吞吐并避免多任务并发导致千问服务排队超时，系统采用“单次取 4 条、2 路并发调用千问”的复核策略，并以任务为单位进行并发控制，保证同一任务内部串行复核、不同任务之间并行推进，从而在保证稳定性的前提下显著提升复核速度。同时，通过对千问服务端的并行度进行配置调优，配合任务的限流策略，可在多个检测任务同时进行时避免推理服务因排队超时而失败，确保复核流程的稳定可靠。

5.实验与结果分析

5.1 实验环境

系统后端采用 C# ASP.NET Core（YoloWebApi），前端为 HTML 页面，推理与训练由 Python 多服务实现。推理与训练服务器配置 2 块 RTX 4090 显卡，Python 环境基于 conda，SQL Server 存储检测结果，千问 VL 模型以私有化方式部署于内网，保障数据安全。

5.2 检测性能

在真实管网数据集上，YOLOv12m 模型的最优检测指标 mAP50 达到 72.36%（177 轮双卡训练），为当前最优结果。各主要缺陷类别的召回率如图 2 所示。

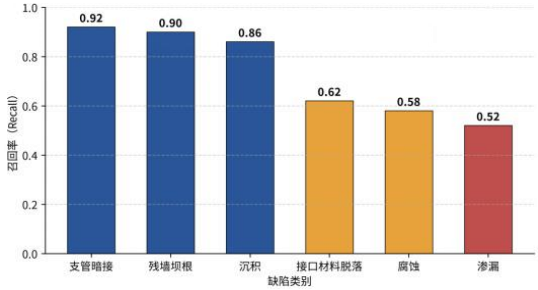


图 2.部分缺陷类别召回率

支管暗接、残墙坝根、沉积等纹理与形态特征明显的缺陷类别召回率分别达到 0.92、0.90 与 0.86，识别效果良好；渗漏、腐蚀等对比度较低、形态不规则的缺陷类别召回率也提升至 0.52 与 0.58，接口材料脱落类别召回率为 0.62。这类难检缺陷对管道安全影响大，后续仍可通过补充针对性样本、类别级阈值下调以及增强策略改进来进一步提升召回率，部分缺陷类别召回率如表 2 所示。

表 2.部分缺陷类别召回率

缺陷类别	召回率
支管暗接	0.92
残墙坝根	0.90
沉积	0.86
接口材料脱落	0.62
腐蚀	0.58
渗漏	0.52

5.3 两阶段架构效果

通过两阶段解耦，YOLO 推理与千问复核由串行改为并行，消除了千问模型对 YOLO 检测吞吐的拖累，检测任务完成时间大幅缩短；同时千问复核在后台异步进行，推理工作台可实时展示复核进度，用户体验明显提升。基于物理里程的去重机制有效抑制了跨帧重复标记，报告中的缺陷数量更加精简、可读性更强，减轻了人工筛选的工作量。与现有 YOLO 类管道缺陷检测方法相比，本文方法在引入多模态语义理解能力的同时，保持了 YOLO 推理的高效性，实现了“检测”与“理解”的解耦协同，并在多 GPU 调度、类别级阈值、物理里程去重等方面形成了较为完整的工程化方案。

6.经济与社会效益分析

从经济效益看，该系统以自动化检测替代人工逐帧判读，显著降低了管网检测的人力成

本与时间成本。传统人工判读一名技术人员需长时间逐帧浏览视频，劳动强度大且效率低，同时极度依赖个人经验，不同技术人员对同一形态的缺陷判读存在差异；系统实现缺陷自动识别与报告自动生成后，单条管段的检测与报告编制时间大幅缩短，同时通过跨帧去重避免了无效缺陷的重复统计，提高了结果的可信度与可用性，进而缩短了管网养护决策周期，提升了管网资产管理的经济性。系统已在真实管网检测场景部署应用，缺陷检测与报告生成的全流程数字化为相关业务提供了可复用的技术积累。

从社会效益看，管网缺陷的及时、准确检测有助于预防管道渗漏、错口等安全隐患，降低内涝与水体污染风险，保障城市排水系统安全运行，对城市水环境治理与基础设施安全具有积极意义。同时，多模态大模型的引入也推动了人工智能技术在市政工程检测领域的落地应用，为智慧城市建设提供了示范[10]。

7.结论

本文结合 YOLO 视觉算法与多模态大模型，构建了一套两阶段雨污水管网缺陷检测系统。通过 YOLOv12m 完成缺陷纯识别、千问 VL 完成语义二次判别，实现了“检测”与“理解”的解耦协同；通过物理里程去重有效消除了跨帧重复标记，提升了报告质量。系统 mAP50 达到 72.36%，并在真实管网检测场景中部署应用。

未来工作将从两方面展开：一是针对渗漏、腐蚀等难检类别，通过补充样本与针对性增强进一步提升召回率；二是优化千问复核的调度策略与提示词，提高语义判别的准确性与稳定性，并探索更多缺陷语义维度的自动提取。

参考文献

[1] 邵延华，张铎，楚红雨，等.基于深度学习的 YOLO 目标检测综述[J].电子与信息学报，2022，44（10）：3697-3708.
[2] 王俊岭，邓玉莲，李英，等.排水管道检测与缺陷识别技术综述[J].科学技术与工程，2020，20（33）：13520-13528.
[3] 胡泽文.AI 算法在供热管网动态调控中的低碳应用[J].工程学研究与实用，2026.
[4] 郭城.AI 视觉分析算法和模型在火电厂智能巡检中运用研究[J].工程施工新技术，2025.
[5] 罗来鸣，王婷.基于优化 YOLO 算法的管道裂缝识别研究[J].给水排水，2023.
[6] 完颜健飞，江雅馨，徐晓龙，等.基于改进 YOLOv5 的城市地下管网缺陷识别算法[J].计算机测量与控制，2024，32（11）：258-264.DOI：10.16526/j.cnki.11-4762/tp.2024.11.036.
[7] 卢守东..NET 中基于回调的 Web 服务异步调用技术[J].计算机时代，2010，（1）：44-45.
[8] 高颖，张强，朱锁玲.融合 LoRA 与 Qwen3-VL 的古农书多模态 RAG 智能问答系统构建研究[J/OL].图书馆杂志，1-15[2026-09-16].https//link.cnki.net/urlid/31.1108.g2.20260109.1235.010.
[9] 刘雨青，隋佳蓉，魏星，等.基于轻量级 YOLOv4 的小目标实时检测[J].激光与光电子学进展，2023.
[10] 谷俊鹏.市政排水管网运维智能化管理策略与实践[J].新城建科技，2024，33（8）：31-33.