

基于视觉大模型的城轨机房 智能巡检方法与技术研究

杨连报, 苗壮壮, 谢 欣

(北京和利时系统工程有限公司, 北京 100176)

摘 要: 针对城轨机房智能巡检算法在复杂环境中特征提取能力弱、多源数据适应性差的问题, 本文提出一种基于视觉大模型的城轨机房智能巡检创新方案。该方案利用视觉大模型强大的深层语义建模能力, 对机房监控图像进行多粒度特征提取, 同时引入自适应多尺度融合模块, 以优化特征表征能力。为兼顾计算效率与识别精度, 本方案设计云-边协同架构。云端部署完整视觉大模型, 负责复杂场景下的高精度推理与模型迭代, 边缘端则部署轻量化推理模块, 实现巡检图像的预处理、实时检测和故障预警, 保障机房运维的即时性与安全性。为验证所提方案的有效性, 基于大连地铁1号线大连北站机房真实运营数据开展测试验证, 涵盖多种设备类型。实验结果表明, 该算法在设备状态识别任务中 mAP@50 达 0.923, F1 分数为 0.915, 综合检测性能优于主流巡检算法。经济效益分析显示, 方案年均运营成本约 44 万元, 较传统人工及定期维护方案年均节约 46 万元, 具备良好的经济可行性。此外, 本方案采用通用工业硬件采集架构、标准化数据接入协议与模块化算法设计, 场景适配性强, 不局限于城轨机房, 还可拓展应用于地铁隧道、车站等轨道交通场景。

关键词: 轨道交通; 视觉大模型; 城轨机房; 智能巡检; 目标检测; 故障预警

中图分类号: TP391.4

文献标志码: A

文章编号: 1672-6073(2026)04-0100-07

Research on Intelligent Inspection Methods and Technologies for Urban Rail Transit Equipment Rooms Based on Large Vision Models

Yang Lianbao, Miao Zhuangzhuang, Xie Xin

(Beijing Hollysys Co., Ltd., Beijing 100176)

Abstract: To address the challenges of weak feature extraction capability and poor adaptability to multi-source data encountered by intelligent inspection algorithms in urban rail transit equipment rooms under complex environments, this paper proposes an innovative inspection scheme based on large vision models. This scheme leverages the powerful deep semantic modeling capability of large vision models to perform multi-granularity feature extraction equipment-room monitoring images, while simultaneously introducing an adaptive multi-scale fusion module to optimize feature representation capability. To balance computational efficiency and detection accuracy, a cloud-edge collaborative architecture is designed. The cloud hosts the complete large vision model and is responsible for high-precision inference in complex scenarios and model iteration, while the edge side deploys a lightweight inference module for preprocessing, real-time detection, and fault early warning of inspection images, ensuring the timeliness and safety of equipment room operation and maintenance. To validate the effectiveness of the proposed scheme, tests are conducted using real operational data from the equipment room of Dalian North Station on Dalian Metro Line 1, covering various equipment types. Experimental results show that the algorithm achieves an mAP@50 of 0.923 and an F1-score of 0.915 in equipment state detection tasks, outperforming mainstream inspection algorithms in overall detection performance. Economic benefit analysis indicates that the scheme incurs an average annual operating cost of approximately RMB 440,000 yuan, representing an annual saving of about RMB 460,000 yuan per year compared with traditional manual and

收稿日期: 2025-08-27 修回日期: 2026-01-05

第一作者: 杨连报, 男, 博士, 高级工程师, 研究方向为大数据与人工智能等新技术应用, yanglianbao@hollysys.com

基金项目: 国家铁路局装备技术中心科研项目(2026-4); 国家能源集团科技创新项目(SHTL-24-52)

引用格式: 杨连报, 苗壮壮, 谢欣. 基于视觉大模型的城轨机房智能巡检方法与技术研究[J]. 都市轨道交通, 2026, 39(4): 100-106.

periodic maintenance approaches, demonstrating sound economic feasibility. Furthermore, the scheme adopts a general-purpose industrial hardware acquisition architecture, standardized data access protocols, and a modular algorithm design, offering strong scenario adaptability. It is not limited to urban rail transit equipment rooms and can be extended to other rail transit scenarios such as subway tunnels and stations.

Keywords: rail transit; large vision model; urban rail transit equipment room; intelligent inspection; object detection; fault early warning

随着城市化进程加速和轨道交通网络不断扩张,城轨机房作为保障地铁、轻轨等城市轨道交通系统正常运行的核心基础设施,其稳定性和安全性日益受到关注。城轨机房承担着信号传输、电力供应、环境控制等关键功能。其内部部署了服务器集群、通信设备、电力控制装置、空调系统等多种精密仪器。这些设备通常需要 7×24 h 不间断运行,任何微小故障若不能及时发现和处理,都可能引发连锁反应,轻则导致列车晚点、运营效率下降,重则造成安全事故和重大经济损失^[1]。因此,针对城轨机房的常态化检查是保障设备安全高效运行的必要举措。与此同时,庞大的设备规模和运维压力,对检查工作的全面性、及时性和准确性提出了极高要求。

城轨机房巡检的核心目的在于实时监测各个设备的工况,及时发现并消除安全隐患,保障系统稳定运行^[2]。在信息化建设初期,由于技术条件有限,相关部门主要采用人工巡检方式实现对机房设备的日常监控与管理。这种模式通常需要组建专业的运维团队,根据机房规模、设备重要性和风险等级,制定详细的分时排班计划。运维人员需按照排班表,每隔一段时间对机房环境、设备运行状态等关键指标进行逐项检查,并在纸质巡检表上手工记录设备各项工况参数。这种人工巡检方式效率较低,异常响应滞后,且消耗较多人力物力,存在明显局限性。随着人工智能技术的快速发展,人工巡检模式正在被更加智能化的监控系统所取代^[3-4]。这种智能监控系统基于先进的边缘计算设备和深度学习目标检测算法^[5-6],通过在边缘侧部署专用智能分析硬件,并结合目标检测模型,从而实现智能监测。相比人工巡检方式,该智能化方案大幅降低人力成本和管理开销。然而,在工业环境复杂、设备类型多样的实际场景下,传统的目标检测算法面临特征提取能力不足、多源数据适应性差等挑战,难以满足城轨机房对低误报率、高可靠性的巡检要求。

面对上述挑战,需探索更具泛化能力与鲁棒性的智能感知技术路径。目前国家正大力推动人工智能与重点行业深度融合,明确提出要“建设一批综合性重大场景、行业领域集成式场景、高价值小切口场景”,

强调以真实业务痛点为牵引,加速技术成果转化^[7]。城轨机房智能巡检精准聚焦运维效率低、人工巡检风险高等核心瓶颈,具备技术适配性强、实施路径清晰、安全与经济效益显著等优势,是轨道交通领域典型的高价值小切口场景,完全契合政策鼓励方向。

相较传统的目标检测算法,视觉大模型通过海量数据预训练,能够学习到更具通用性和鲁棒性的视觉特征表征^[8-9]。因此,本文以视觉大模型为特征提取主干网络,并设计一种基于注意力机制的多尺度特征融合模块,深入探索其在城轨机房智能巡检中的算法应用。该方案不仅响应国家场景培育战略,也为提升轨道交通基础设施的智能化运维水平提供可落地、可扩展的新一代人工智能技术路径。

1 方法

本文整体架构设计分为两级部署:视觉大模型部署在云服务器上,传统小模型部署在边缘服务器上。端侧摄像头采集的视频帧同时发送至边缘与云端,边缘侧实现实时快速检测,云端则精度更高但推理耗时。当边缘检测置信度低于某一阈值时,将结合云端的检测结果进行复核,以避免系统误报。本章聚焦视觉大模型检测算法,其框架如图 1 所示,由特征提取主干网络^[10]、多尺度特征融合模块及回归头构成。

1.1 特征提取主干网络

特征提取主干网络如图 2 所示。整体来看,该主干网络由 1 个压缩模块、3 个下采样模块和 4 个特征编码模块组成。压缩模块用于对原始图像进行初步特征提取,将特征图的分辨率降低为原来的 1/4。该模块包含 2 个二维卷积层(2D Convolution, 2DConv)、2 个归一化层(Layer Normalization, LN)和 1 个 GELU 激活函数。其中,2DConv 的卷积核尺寸均为 3×3,步长为 2,填充大小为 1。3 个下采样模块穿插在 4 个特征编码模块之间,用于将特征图的空间尺寸逐次缩小一半。每个下采样模块由一个步长为 2、填充为 1 的 3×3 卷积层后接一个 LN 层构成。特征编码模块作为主干网络的核心,主要由可变形卷积、LN 层以及前馈神经网络(Feed Forward Network, FFN)构成。该卷积具备

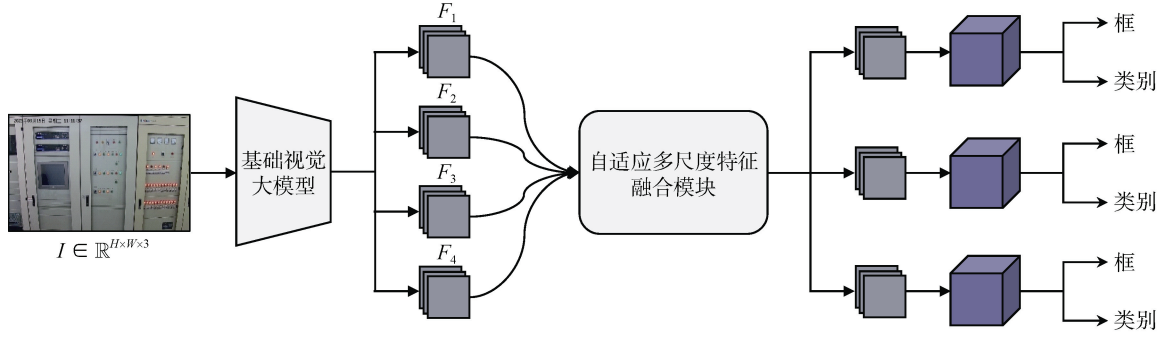


图 1 基于视觉大模型的城轨机房智能巡检方法

Figure 1 Intelligent inspection method for urban rail transit equipment rooms based on large vision models

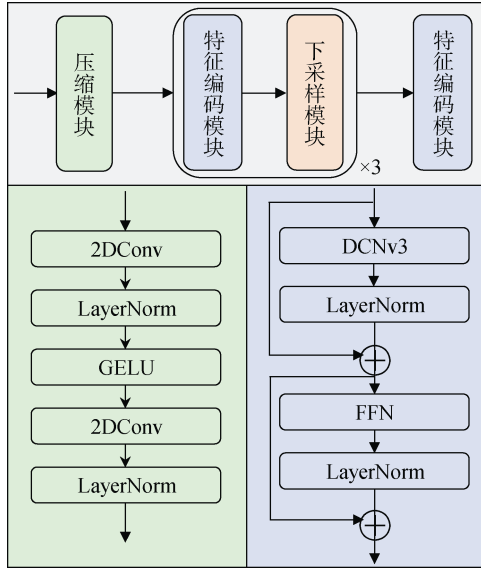


图 2 特征提取主干网络

Figure 2 Feature extraction backbone network

与多头自注意力机制相当的全局特征聚合能力，同时摒弃了传统卷积的静态权重与强归纳偏置，从而在大规模模型构建中兼具高效性与灵活性。这里采用的可变形卷积为 DCNv3，DCNv3 是 DCNv2 的扩展版本。通常情况下，DCNv2 被用作常规卷积的扩展，通过加载常规卷积的预训练权重并进行微调以提升性能，但这并不完全适用于需要从头开始训练的大规模视觉基础模型。为了解决这个问题，DCNv3^[11]在卷积神经元之间共享权重、引入分组机制，并沿采样点对调制标量进行归一化。基于以上扩张后，DCNv3 的公式可以表示为

$$y(p_0) = \sum_{g=1}^G \sum_{k=1}^K w_g m_{gk} x_g(p_0 + p_k + \Delta p_{gk}) \quad (1)$$

式中， G 表示聚合组的总数；对于第 g 组， $w_g \in \mathbb{R}^{C \times C'}$ 表示该组的位置无关投影权重，其中 $C' = C/G$ 表示组

的维度； $m_{gk} \in \mathbb{R}$ 表示第 g 组中第 k 个采样点的调制标量，通过 Softmax 函数在维度 K 上进行归一化； x_g 表示切片后的输入特征图； p_0 表示当前像素点； p_k 表示预设网格的第 k 个位置； Δp_{gk} 是第 g 组中网格采样位置 p_g 的偏移量。

如图 1 所示，输入图像 $I \in \mathbb{R}^{H \times W \times 3}$ 经特征提取主干网络后，输出 4 个不同尺度的特征图 $\{F_i\}_{i=1}^4$ ，满足

$F_i \in \mathbb{R}^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i}$ ，其中 H 、 W 为输入图像的高和宽， C_i 为第 i 级特征的通道数。

1.2 自适应多尺度特征融合模块

在机房智能巡检任务中，待检设备与关键部件在图像中存在显著尺度差异，跨度可达数十倍。深层特征具备强大的语义理解能力，浅层特征具备更好的细节信息。通过引入多尺度融合模块构建特征金字塔，能够协同利用低层空间信息与高层语义特征，使得图像特征能够具备多尺度的特征信息。因此本文设计一种自适应多尺度特征融合模块。

该模块整体采用自上而下的金字塔结构，首先对高层特征图进行上采样，再逐步与低层特征图融合，最终输出包含多尺度信息的特征表示。以 F_4 和 F_3 的融合为例，首先对 F_4 进行双线性插值上采样，使其空间分辨率与 F_3 一致(即 $H/16 \times W/16$)，并引入一个 2D 卷积层以缓解上采样可能引起的棋盘效应。为有效融合不同尺度的特征，本文设计了一种自适应特征融合单元。该单元通过注 2D 注意力机制自适应地学习融合权重。具体而言，首先沿通道维度拼接上采样后的 F_4 与 F_3 ，将两种特征放置到一个空间中，得到拼接特征 $[F_4 \parallel F_3] \in \mathbb{R}^{H/16 \times W/16 \times 2C_3}$ ，并使用通道标准化(Channel-Norm)消除来自不同子空间的图像特征差异性。然后，通过一个可学习的嵌入向量 $W \in \mathbb{R}^{2C_3 \times 2}$ 对该拼接特

征进行线性变换, 将其映射至二维空间, 以建模跨尺度特征间的依赖关系。最后, 利用 Sigmoid 激活函数生成(0, 1)范围内的融合权重。整体的表达式为

$$A = \text{Sigmoid}(\text{ChannelNorm}([F_4 | F_3])W) \quad (2)$$

式中, $A \in \mathbb{R}^{H \times W \times 2}$ 表示各项大于 0 的注意力系数矩阵, 且 $A_{i,j,1}$ 和 $A_{i,j,2}$ 分别表示在特征中 (i, j) 位置 F_4 和 F_3 对应的融合权重值。因此, 对于 F_4 和 F_3 的自适应融合可以表示为

$$Z_{ij} = 2\text{DConv}(A_{i,j,1} \otimes F_4 + A_{i,j,2} \otimes F_3) \quad (3)$$

$$Z = [Z_{ij}]_{i=1, j=1}^{H, W} \in \mathbb{R}^{H \times W \times C_3} \quad (4)$$

式中, $Z_{ij} \in \mathbb{R}^{1 \times 1 \times C_3}$ 表示 (i, j) 位置融合后的通道特征; Z 表示两种尺度特征融合后的最终特征表示。

为了进一步增强模型对不同尺度目标的检测能力, 在特征融合阶段对顶层特征采用了多尺度全局池化策略。具体来说, 将融合后的多尺度图像特征分别池化至 3 种具有代表性的尺度: $H/8 \times W/8$ 、 $H/16 \times W/16$ 和 $H/32 \times W/32$ 。通过这种多尺度特征表示, 模型能够自适应地处理实际应用场景中存在的各种尺度变化, 显著提升对不同尺寸目标检测的鲁棒性。

1.3 检测头

检测头采用三头检测结构, 分别对应不同尺度的特征图($H/8 \times W/8$, $H/16 \times W/16$, $H/32 \times W/32$)。这种多尺度检测设计能够有效捕捉不同大小的目标, 其中较大尺度的特征图($H/8 \times W/8$)适合检测小目标, 中等尺度的特征图($H/16 \times W/16$)适合检测中等大小的目标, 而较小尺度的特征图($H/32 \times W/32$)则更适合检测大目标。

每个检测头由若干卷积模块堆叠而成, 每个模块依次包含卷积层、批归一化(batch normalization, BN)和 SiLU 激活函数。具体而言, 检测头首先通过一个 1×1 卷积层调整输出通道数, 再经由一个 3×3 卷积层进一步提取特征, 最终分别通过分类分支和回归分支输出预测结果: 分类分支预测目标的类别概率, 回归分支则输出边界框的位置参数(包括中心点坐标、宽度和高度)。

此外, 检测头还采用了自适应特征融合策略, 通过跨尺度特征交互增强不同层级之间的信息流动, 从而提升检测性能。这种结构设计在保持较高检测精度的同时, 能够有效平衡计算效率与模型复杂度。

1.4 损失函数

为高效训练网络, 本文采用复合损失函数。回归分支结合分布式焦点损失(distributed focal loss, DFL)

与完全交并比损失(complete intersection over union, CIoU), 其中 DFL 用于建模边界框 4 个坐标(如到左、上、右、下边界的距离)的离散分布, CIoU 则用于度量预测框与真实框之间的几何对齐程度。分类分支采用二元交叉熵损失(binary cross-entropy, BCE)。因此, 损失函数公式表示为

$$\text{Loss} = \lambda_1 \cdot L_{\text{BCE}} + \lambda_2 \cdot L_{\text{CIoU}} + \lambda_3 \cdot L_{\text{DFL}} \quad (5)$$

式中, L_{BCE} 、 L_{CIoU} 、 L_{DFL} 分别为分类分支的 BCE 损失与回归分支的 CIoU 损失、DFL 损失; $\{\lambda_1, \lambda_2, \lambda_3\}$ 为各损失项的权重系数, 用于平衡不同任务分支的训练贡献。

2 实验

基于大连地铁 1 号线大连北站机房构建了多样本数据集, 本章通过系统化的实验设计和模块化消融分析, 验证所提出的视觉大模型算法的有效性。

2.1 数据集与实验指标

数据集来源于大连地铁实际运营场景, 通过现场部署的高清摄像头采集不同工况下的机房设备图像数据。采集对象涵盖高压开关柜、直流开关柜、低压开关柜、电力变压器、整流器、轨电位排流柜、交直流屏以及控制信号屏等 8 类核心电气设备。为确保数据多样性, 采集了在不同温湿度和光照条件下的外部机房环境下的数据。针对样本数量不足的问题, 采用数据增强技术进行有效扩充。在传统方法层面, 应用图像几何变换(包括随机翻转、裁剪和平移等操作); 同时结合大模型进行高质量样本生成。通过这种混合增强策略, 最终实现原始数据集规模数倍以上的有效扩充。共获取 10 790 张图像, 包含大部分场景。使用 labelme 软件^[12]对数据进行标注, 其中训练集含 8 632 张图像, 测试集含 2 158 张图像。在训练阶段, 将所有图像自适应缩放至 640 像素 $\times$ 640 像素, 以便于模型训练。使用 mAP@50、F1-Score 作为评价模型性能的主要指标。

2.2 实验环境

为评估基于视觉大模型的方法在城轨机房智能巡检任务中的应用效果, 本文搭建了基于深度学习的目标检测实验环境。实验模型基于深度学习的 Pytorch 框架构建, 在模型训练过程中核心的参数项和参数值基于经验法、文献调研以及多次实验测试确定。最终设置采用 AdamW 优化器完成。设置学习率为 $2e^{-5}$, 权重衰减为 0.05, 层衰减率为 0.9。综合考虑计算资源、

数据集大小及模型大小，确定最大迭代次数为 100，Batchsize 为 16。

在实际部署过程中边缘服务器的算法采用轻量化 YoloV5 模型，并基于 TensorRT 实现推理加速。边缘服务器采用了和利时自研边缘计算单元 EIC5801，具备 40TOPs 算力。经过测试，算法在主流边缘服务器上同样具备适配性，包括 Jetson AGX Orin、Jetson Orin

Nano 及 Jetson Xavier NX 系列。

2.3 对比实验

为了验证算法有效性，利用相同的数据集，使用 YoloV8^[13]、YoloV5、Faster R-CNN 模型进行对比实验。以 mAP@50 和 F1-Score 进行评价，其中 mAP@50 定义为交并比阈值设定为 0.5 时的平均精度均值，实验定量结果如表 1 所示。

表 1 模型试验结果对比
Table 1 Comparison of model testing results

类别	本文方法		YoloV8		YoloV5		Faster R-CNN	
	mAP@50	F1-Score	mAP@50	F1-Score	mAP@50	F1-Score	mAP@50	F1-Score
高压开关柜	0.942	0.911	0.924	0.888	0.911	0.871	0.896	0.858
直流开关柜	0.954	0.946	0.929	0.902	0.918	0.900	0.898	0.885
低压开关柜	0.961	0.951	0.917	0.896	0.929	0.915	0.899	0.881
变压器	0.878	0.873	0.866	0.861	0.851	0.847	0.839	0.835
整流器	0.917	0.911	0.878	0.873	0.859	0.865	0.842	0.850
轨电位排流柜	0.932	0.935	0.894	0.888	0.905	0.904	0.880	0.874
交直流屏	0.912	0.911	0.897	0.897	0.881	0.884	0.865	0.868
控制信号屏	0.892	0.882	0.855	0.843	0.841	0.825	0.830	0.814
平均	0.923	0.915	0.895	0.881	0.887	0.876	0.869	0.858

本文方法在城轨机房智能巡检任务中展现出卓越的性能优势，在各类关键电力设备检测指标上全面超越现有主流算法。由表 1 可知：针对城轨机房特有的复杂环境和多样化设备，本文方法在检测精度和鲁棒性方面都具有显著提升。在高压开关柜检测任务中，实现了 94.2%的 mAP@50 和 91.1%的 F1-Score，相比次优方法 YoloV8 分别提升 1.8%和 2.3%，这一改进对于确保城轨供电系统安全运行具有重要意义；在直流开关柜检测方面，达到 95.4%的 mAP@50 和 94.6%的 F1-Score，较次优模型提升 2.5%和 4.4%，这种性能提升能够有效降低机房巡检中的误检和漏检风险；在城轨机房中普遍存在的低压开关柜检测上，本文方法表现尤为突出，以 96.1%的 mAP@50 和 95.1%的 F1-Score 取得最优效果，这主要得益于针对机房环境特别优化的多尺度特征融合策略；对于整流器、轨电位排流柜等城轨特有设备，本文提出的方法展现出 2%~3%的稳定优势，这对保障城轨电力系统稳定运行至关重要；在控制信号屏等小目标检测任务中，相比 YoloV8 提升 3.7%的 mAP@50，显著改善了传统算法对小目标检测效果不佳的问题。从整体性能来看，本文提出的方法平均 mAP@50 达到 92.3%，较 YoloV8 提升 2.8%，这种全面提升源于在视觉大模型强大的特征提取能力及

适应不同设备尺寸的多尺度检测机制等方面的创新。这些技术创新使提出的方法能够更好地应对城轨机房巡检中面临的设备种类多、环境复杂等实际挑战，为城轨智能运维提供了更可靠的技术支持。

2.4 消融实验

为了验证本文提出的多尺度特征融合模块的有效性，将其与主流的特征金字塔网络(feature pyramid network, FPN)^[14]和自适应空间特征融合技术(adaptive spatial feature fusion, ASFF)^[15]等方法进行了对比，实验效果如表 2 所示。由表 2 可知：本文提出的模块在高压开关柜、整流器和控制信号屏 3 类目标检测任务中均展现出最优性能，全面超越传统 FPN 和 ASFF 方法。在高压开关柜检测中，本文模块的 mAP@50(94.2%)和 F1-Score(91.1%)分别较 FPN 提升 0.7%和 0.6%，较 ASFF 提升 0.5%和 0.1%；在整流器检测中，mAP@50 (91.7%)和 F1-Score(91.1%)分别较 FPN 提升 0.3%和 0.8%，较 ASFF 提升 0.4%和 0.4%；在控制信号屏检测中，mAP@50(89.2%)和 F1-Score(88.2%)分别较 FPN 提升 0.5%和 0.2%，较 ASFF 提升 0.8%和 0.3%。实验结果表明，本文模块通过改进多尺度特征融合策略，在保持高检测精度的同时显著提升了综合性能，尤其对高压开关柜和整流器等关键设备具有更强的检测鲁棒性。

表 2 消融实验

Table 2 Ablation study

类别	本文提出的模块		FPN		ASFF	
	mAP@50	F1-Score	mAP@50	F1-Score	mAP@50	F1-Score
高压开关柜	0.942	0.911	0.935	0.905	0.937	0.910
整流器	0.917	0.911	0.914	0.903	0.913	0.907
控制信号屏	0.892	0.882	0.887	0.880	0.884	0.879

3 实际应用探讨

目标检测技术与城轨机房巡检流程的衔接是算法实际应用落地的关键问题。本文通过实地调研分析,将目标检测模型嵌入到巡检计划制定、现场检查、异常确认等关键环节,构建了“智能识别-人工复核-闭环管理”的协同工作机制。具体而言,模型自动识别各个设备的工况,并实时推送结构化检测结果,该监测结果可存至现有系统数据库中,使得系统可以调用查询,完成无缝系统集成。也可以通过系统独立的方式进行实现,通过开发独立的智能巡检系统完成前端展示。基于该结构化结果,巡检人员可快速定位问题并完成复核确认,系统自动生成标准化巡检报告并触发后续维修工单。这种设计不仅保留了人工复核的关键质量控制环节,同时显著提升了异常识别的效率和标准化程度。此外,通过建立误报反馈机制,持续优化模型性能,确保技术方案在实际运维环境中的适用性和可靠性。

本文通过实证分析与综合成本效益测算,系统评估了所提出方法的技术可行性与经济实用性。以典型城轨机房运维场景为例,传统人工巡检方案需配置 40 个检测点,每班次依赖 2 名持证运维人员执行现场检查,单次作业耗时约 2 h,按每日 3 班次运行计算,年均人力投入达 6 人/d,对应直接人力成本约为 90 万元/年(按人均综合成本 15 万元/年估算)。此外,该模式存在漏检率高、响应延迟大等问题,间接导致设备故障发现滞后,增加非计划停机风险。

相比之下,本文提出的智能视觉巡检方案基于最优摄像头布局策略,充分利用相邻检测点的视野重叠特性,在保障全覆盖的前提下将物理摄像头数量优化至 32 台。系统架构由前端感知层(32 路高清摄像头)、边缘计算层(和利时自研的 EIC5801 边缘计算单元)及中心云服务器(配备 2 块 GPU $\geq$ 64GB、内存 $\geq$ 256GB)构成,采用跳帧检测机制平衡实时性与算力消耗。

尽管该方案在初期涉及较高的硬件采购与系统集

成投入(预计一次性资本支出约 95 万元,含设备购置、网络改造与部署调试),且伴随一系列持续性的运营开销,具体包括:

数据管理成本:视频数据的存储、归档与定期清理,年均约 8 万元;

模型维护成本:应对环境变化导致的数据漂移,需每季度开展模型再训练与验证,折合人力与算力成本约 6 万元/年;

系统运维成本:软硬件监控、安全补丁更新、故障排查等,约 6 万元/年;

技术支持成本:由 AI 算法供应商或内部技术团队提供远程诊断、功能优化与应急响应支持,约 5 万元/年。

综合测算显示,该 AI 系统年均总运营成本约为 44 万元,相较传统方案,仍年均节约 46 万元的净成本,如表 3 所示。

表 3 巡检方案成本对比

Table 3 Cost comparison of inspection schemes

项目	一次性投入/万元	年运营成本/万元	5 年总成本/万元	每年均成本/万元	年节省成本/万元	投资回收期
传统人工巡检	0	90	450	90	—	—
智能视觉巡检	95	25	220	44	46	约 2 年

进一步敏感性分析表明,在系统使用年限超过 2 年后,累计节约成本即可完全覆盖初始投资,投资回收期合理。综上所述,尽管智能系统的部署引入了新的技术复杂性与持续投入需求,但从全生命周期视角看,其综合经济效益具有显著优势,具备良好的实用推广价值。

4 结论

针对城轨机房智能巡检实际业务需求,本文构建云边协同智能巡检框架,并提出一种基于视觉大模型的特征提取与迁移学习能力,在各类关键电力设备检测任务中平均 mAP@50 达 92.3%,较 YoloV8 提升 2.8%。成本测算结果表明,本方案年均运营成本约 44 万元,相比传统巡检方案每年可节约成本约 46 万元,可为工程落地与规模化推广提供有效参考。

智能检测算法是城轨机房巡检自动化、智能化的核心基础,能够显著降低人工巡检的成本与资源消耗。未来研究工作将重点围绕以下两方面展开:一是持续

优化算法在极端光照、设备老化、遮挡干扰等复杂环境下的鲁棒性与适应性；二是推进“一机多场景”泛化应用，将本技术迁移至地铁隧道、车站站厅等多样化运维场景中，攻克光照剧烈变化、设备种类繁多、环境动态性强等挑战，全面检验方法的推广价值与稳定性，从而为构建覆盖轨道交通全场景的智能运维体系提供有效技术支撑。

参考文献

- [1] 高翔, 贺苗苗, 徐明宇, 等. 轨道交通中心级设备全周期管理研究与应用[J]. 设备管理与维修, 2024(14): 7-9.
- [2] 王坚强, 陈祥龙, 曾科智, 等. 铁路信号机房智能巡检系统设计[J]. 铁路计算机应用, 2023, 32(2): 34-39.
Wang Jianqiang, Chen Xianglong, Zeng Kezhi, et al. Intelligent inspection system of railway signal equipment room[J]. Railway Computer Application, 2023, 32(2): 34-39.
- [3] 苏秀梅. 基于计算机视觉的计算机机房设备安全巡检方法[J]. 长江信息通信, 2024, 37(10): 141-143.
Su Xiumei. Computer vision based security inspection method for computer room equipment[J]. Changjiang Information & Communications, 2024, 37(10): 141-143.
- [4] 张显鸿. 面向机房巡检的小目标检测算法研究与应用[D]. 武汉: 武汉工程大学, 2024.
Zhang Xianhong. Research and application of small object detection algorithm for data center inspection[D]. Wuhan: Wuhan Institute of Technology, 2024.
- [5] Jung S, Song A, Lee K, et al. Advanced building detection with faster R-CNN using elliptical bounding boxes for displacement handling[J]. Remote Sensing, 2025, 17(7): 1247.
- [6] Wang C Y, Yeh I H, Mark Liao H Y. YOLOv9: learning what you want to Learn using programmable gradient information[C]//Computer Vision – ECCV 2024. Cham: Springer, 2025: 1-21.
- [7] 国务院办公厅. 国务院办公厅关于加快场景培育和开放推动新场景大规模应用的实施意见[EB/OL]. [2025-11-30]. https://www.gov.cn/zhengce/zhengceku/202511/content_7047420.htm.
- [8] 夏威, 周廉钧, 徐琳玮. 基于视觉大模型的输电线入侵物单目相机检测方法及应用[J]. 产业创新研究, 2025(10): 4-8.
- [9] 代明睿, 李文浩, 史维峰, 等. 铁路行业视觉大模型构建及应用[J]. 中国铁路, 2025(1): 1-12.
Dai Mingrui, Li Wenhao, Shi Weifeng, et al. Development and application of a large vision model for railway industry[J]. China Railway, 2025(1): 1-12.
- [10] Wang W H, Dai J F, Chen Z, et al. InternImage: exploring large-scale vision foundation models with deformable convolutions[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, 2023: 14408-14419.
- [11] Zhu X Z, Hu H, Lin S, et al. Deformable ConvNets V2: more deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2020: 9300-9308.
- [12] Russell B C, Torralba A, Murphy K P, et al. LabelMe: a database and web-based tool for image annotation[J]. International Journal of Computer Vision, 2008, 77(1): 157-173.
- [13] Yaseen M. What is YOLOv8: an in-depth exploration of the internal features of the next-generation object detector [EB/OL]. 2025. <https://arxiv.org/abs/2408.15857>.
- [14] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 936-944.
- [15] Liu S T, Huang D, Wang Y H. Learning spatial fusion for single-shot object detection[EB/OL]. [2025-11-30]. <https://arxiv.org/abs/1911.09516>.

(编辑: 王艳菊)