基于视觉AI的数据中心异常识别技术及应用研究

Research on Visual AI-Based Anomaly Detection Technologies and Applications in Data Centers

王清华，范纪明，朱昊然，秦真涛，李　鲲，陈一丁（中国联通上海分公司，上海 200082）
Wang Qinghua，Fan Jiming，Zhu Haoran，Qin Zhentao，Li Kun，Chen Yiding（China Unicom Shanghai Branch，Shanghai 200082，China）

摘　要：
针对数据中心人工巡检覆盖不足、复杂场景下堆物异常识别准确率偏低的问题，提出一种基于视觉AI的云边端协同异常识别方法。采用结构先验约束的生成式数据增强缓解样本不均衡，通过两阶段迁移学习提升边缘侧目标检出能力，并结合时空约束与多模态大模型开展云端语义研判。实际部署验证结果表明，该方法可将巡检频次提升10倍，异常响应时间由2~4 h缩短至5~10 min，识别准确率达93.6%，为数据中心极简智能运维提供技术支撑与应用验证。

关键词：
异常识别；云边端协同；数据中心；多模态语义理解；智能运维
doi：10.12045/j.issn.1007-3043.2026.10.006
文章编号：1007-3043（2026）10-0035-07
中图分类号：TP308
文献标识码：A
开放科学（资源服务）标识码（OSID）：

Abstract：
To address insufficient manual inspection coverage in data centers and the low accuracy of detecting improperly stacked objects in complex scenes，a visual AI-based anomaly detection method using cloud-edge-end collaboration is proposed. Generative data augmentation guided by structural prior constraints mitigates sample imbalance. Two-stage transfer learning improves object detection at the edge，while spatiotemporal constraints and a multimodal large model support cloud-based semantic assessment. Field deployment verification results show that the method increases inspection frequency by ten times，reduces anomaly response time from 2~4 hours to 5~10 minutes，and achieves a recognition accuracy of 93.6%. These results provide technical support and practical validation for streamlined intelligent operations and maintenance in data centers.

Keywords：
Anomaly detection；Cloud-edge-end collaboration；Data center；Multimodal semantic understanding；Intelligent operation and maintenance

引用格式：王清华，范纪明，朱昊然，等. 基于视觉AI的数据中心异常识别技术及应用研究[J]. 邮电设计技术，2026（10）：35-41.

1 概述

数据中心是支撑数字经济的核心信息基础设施，承载政务服务、企业算力与互联网业务等关键业务，国家及行业规范对其安全防护、运行监测与智能管控提出了明确要求。在实际运维中，数据中心重点区域常出现废旧蓄电池、施工余料、包装箱等物品短时占用或长期滞留的情况，一旦压缩通行宽度甚至阻塞消防通道，将直接威胁高可靠性连续运行场景的安全稳定。然而，当前行业普遍存在“重建设、轻运营”的现象，单园区通常部署数百乃至上千路监控摄像头，但通行安全管控仍主要依赖7×24 h视频采集、人工值守和事后回放的传统模式，难以满足高频次、全覆盖、快响应的运维要求。因此，借助计算机视觉与人工智能技术提升异常识别能力，已成为数据中心智能运维的重要发展方向。

针对通行空间异物检测问题，国内外研究主要分为背景建模、目标检测和语义理解3类技术路径。文献[1]系统梳理了视频监控中基于背景减法的静态与遗留目标检测方法，该类方法在受限场景下可通过前

收稿日期：2026-08-20

景提取识别停留物体，但对光照变化与背景动态更新敏感，在数据中心运维场景下误报率较高。为提升异物检测的准确性与场景适应性，Hu等人[2]采用改进YOLOv8构建铁路障碍物入侵检测与风险评估系统，通过多尺度特征融合提升了小目标与远距离障碍物的检出率；郑含博等人[3]针对输电线路隐患样本不足的问题，采用混合数据增强扩充训练集，并结合YOLO-2MCS模型开展检测。但上述方法仅能判定物体存在，难以结合通行区域约束与运维规则判定是否构成实质性阻塞，且异常样本的长尾分布限制了其跨场景适用性。近年来，相关研究将多模态视觉语言模型应用于施工安全识别、消防合规检查和复杂场景风险研判等任务，表明其在物体关系理解和规则符合性判断方面具有应用潜力[4-6]。但在生产场景中对海量视频帧全量推理仍存在计算开销大、分析效率低与结果不稳定等问题。

综上，针对数据中心通道异物堆放场景中长尾样本稀缺、语义判别割裂以及全量大模型推理成本高等难点，本文提出结构先验引导的可控生成式数据增强策略，增强长尾场景下异物样本的覆盖度与多样性；在此基础上，构建融合轻量检测与多模态语义理解的大小模型协同机制，形成云边端架构下的智能异常识别方案，支撑数据中心从被动人工巡检向主动智能管控的范式转变。

2 基于场景结构先验的可控生成式数据增强策略

数据中心通道异物占用属于典型长尾运维场景，真实异常样本采集难度大、场景覆盖维度有限，样本分布不均衡已成为制约目标检测模型泛化能力的核心瓶颈[7]。传统几何变换、色彩调整类扩增方法仅能实现像素空间的低层次变换，无法生成全新场景与异物形态；通用生成式模型虽可构造新样本，但缺少机房空间规则约束，易产生位置越界、形态不符的无效数据。

针对上述问题，本文提出基于场景结构先验的可控生成式数据增强策略，以潜空间扩散模型为生成基座，将机房空间规则、环境特征与风险语义转化为结构化约束注入扩散流程，从场景分布拓展与类内多态增强2个维度实现长尾场景的精准覆盖。

2.1 潜空间扩散模型基础原理

潜空间扩散模型（Latent Diffusion Model，LDM）[8]通过高维像素空间到低维潜空间的降维映射，结合迭代式加噪—去噪机制实现高质量图像生成。模型前向过程为加噪阶段：原始图像经变分自编码器（Variational Auto-Encoder，VAE）映射得到潜特征 x_0，向其中逐步注入高斯噪声，经过 t 步加噪后的潜特征可通过闭式解直接计算[9]：

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \varepsilon_t \tag{1}$$

其中，$\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$ 为累积噪声权重系数，α_i 为第 i 步的噪声衰减系数；$\varepsilon_t \sim N(0,\boldsymbol{I})$ 为第 t 步注入的标准高斯噪声。逆向去噪阶段以DiT（Diffusion Transformer）为骨干网络，结合外部条件输入预测噪声并逐步还原潜特征，单步去噪的迭代更新公式为[10]：

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\, \varepsilon_\theta(x_t,t,c) \right) + \sigma_t z \tag{2}$$

其中，$\varepsilon_\theta(\cdot)$ 为DiT网络预测的噪声值；θ 为网络可训练参数；c 为文本语义、空间掩码等多模态条件输入；σ_t 为逆向扩散过程的后验标准差；$z \sim N(0,\boldsymbol{I})$ 为随机噪声变量。

潜空间扩散模型条件生成架构如图1所示，其中，条件注入层融合场景结构先验与背景基底图像，经多条件编码模块生成条件特征 c；扩散生成层以DiT为骨干网络，将加噪潜特征划分为图像块Token序列，通过自注意力完成全局特征建模，以自适应层归一化（adaptive Layer Normalization，adaLN）注入条件引导去噪过程；图像重建层通过VAE解码器将潜特征映射回像素空间，实现高质量生成[11]。

2.2 结构先验约束下的通道异物可控合成方法

尽管潜空间扩散模型[12]在生成效率与质量方面具有显著优势，但通用模型缺乏对数据中心通道场景的空间先验知识，生成的异物位置、尺度与场景关系易偏离运维规则。因此，本研究以真实机房无异常监控画面为基底，将场景结构先验转化为可控生成条件，通过场景解耦与语义引导2个维度的协同，实现跨场景泛化与类内形态增强。面向数据中心通道异物场景的可控合成样本生成流程如图2所示。

2.2.1 场景基底与空间约束

采集数据中心主机房通道、消防疏散通道等重点区域的无异常监控画面，构建标准化背景库，覆盖不同时段光照、拍摄角度与机柜布局的典型场景。对每张背景图，以结构化文本描述定义异物的目标放置区域，将通道的空间布局与通行规则转化为空间先验约

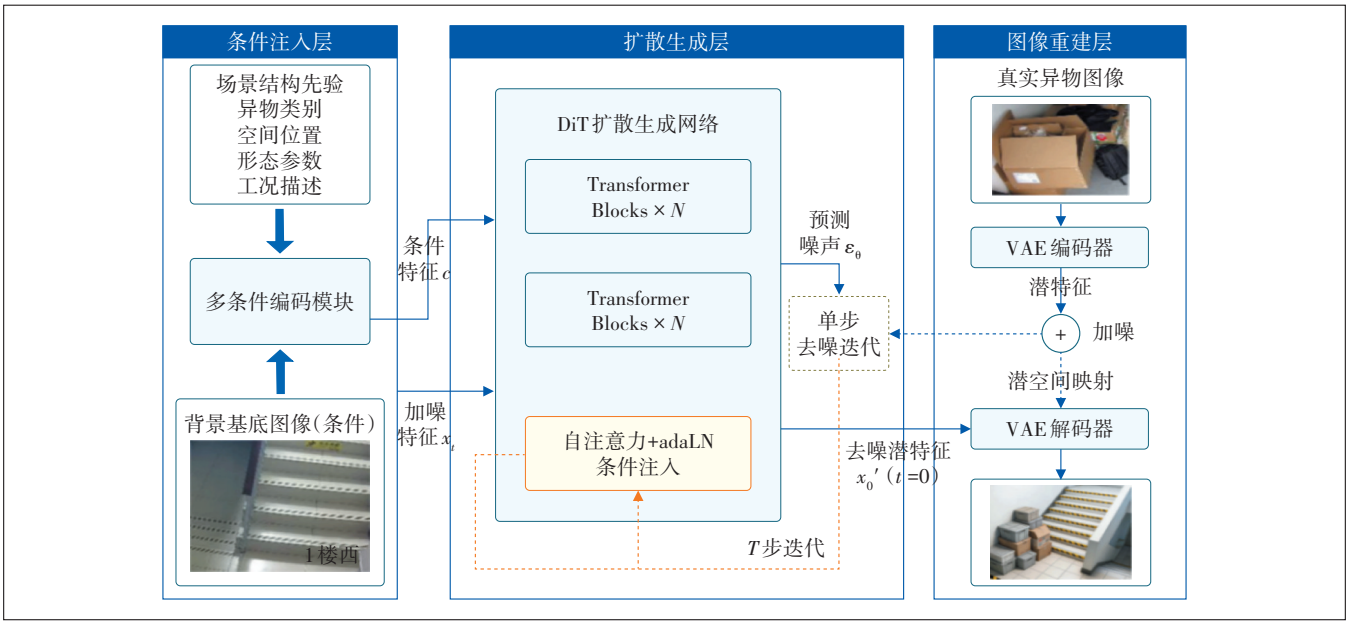


图1　潜空间扩散模型条件生成架构

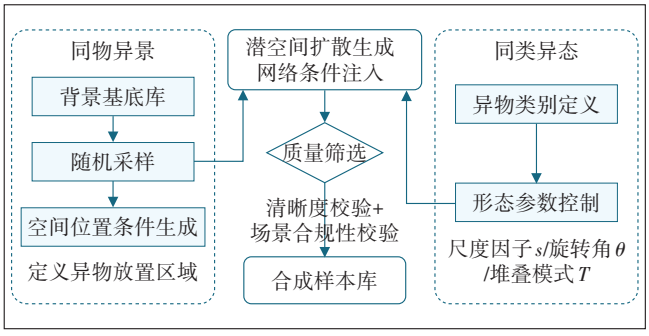


图2　面向数据中心通道异物场景的可控合成样本生成流程

束。生成过程中强制将异物约束在指定通行区域内，其可被形式化表示为：

$$x_{\text{syn}} = G\left(x \middle| c_{\text{bg}}, c_{\text{spatial}}\right),\ \text{s.t.}\,\text{loc}\left(x_{\text{syn}}\right) \in \Omega_{\text{valid}} \tag{3}$$

其中，c_{bg}为背景条件，c_{spatial}为包含位置与边界描述的空间先验条件，Ω_{valid}为合法生成区域。

2.2.2 语义引导与形态控制

面向废旧蓄电池、施工余料、包装箱、工具推车等高频通道占用异物，在文本条件中对异物的类别属性、几何尺寸和堆叠方式进行差异化参数控制。设形态控制参数向量$\phi = [s, \varphi, \tau]$，其中$s$为异物尺度因子，$\varphi$为平面旋转角度，$\tau$为堆叠模式标识。通过参数空间的随机采样实现跨形态多样化生成：

$$\tilde{x}_{\text{syn}} = G\left(x \middle| c_{\text{bg}}, c_{\text{spatial}}, \phi\right),\ \phi \sim P(\Phi) \tag{4}$$

其中，$P(\Phi)$为形态参数的先验分布，覆盖真实运维场景中常见的异物呈现方式。同时，在条件中融入地面反光、局部遮挡等机房典型工况参数，进一步扩充边缘场景样本，使同一类别异物呈现不同形态，实现“同类异态”的类内多态增强。

2.2.3 质量筛选与数据集构建

生成样本经清晰度校验与场景合规性校验2轮筛选，剔除模糊、扭曲或空间关系不合理的低质量样本。合格样本与真实采集样本按比例混合，形成最终训练数据集。

3 基于两阶段迁移学习的改进YOLO检测模型

第2章通过结构先验引导的可控生成式数据增强，补足了通道异物跨场景的样本覆盖。在此基础上，采用“场景预适配—难例精调”两阶段迁移训练策略，使检测模型在边缘侧以低延迟实现高召回率的候选提取，语义理解与通行影响判定则交由后续多模态大模型完成。

3.1 候选发现任务建模

通道异物检测建模为二分类目标检测，仅区分“异物”与“背景”，不区分子类别。在实际运维场景中，通道异物除纸箱、线缆盘、施工余料等常见类型外，还可来源于临时堆放的备品备件、外来施工机具及非常规堆放物，其类别范围具有高度开放性，难以通过预定义类别集合覆盖完备。因此，检测器聚焦候选发现，以高召回为优化目标，提取画面中偏离正常

背景的疑似异常区域，其输出形式化表示为[13]：

$$f_\theta(x)=\left\{\left(b_i,c_i,s_i\right)\right\}_{i=1}^{k} \tag{5}$$

其中，b_i为边界框坐标，c_i为类别标签，s_i为置信度，k为候选框数量。选用YOLOv11s作为基座模型[13]，在轻量推理条件下，其多尺度特征融合架构兼顾了检出能力与边缘侧低延迟部署需求。

3.2 面向高召回适配的渐进式迁移训练

候选发现环节的漏检代价远高于误报，这是因为漏检使风险目标穿透检测层，无法进入后续的语义研判与处置流程；而误报仅增加少量冗余推理开销，可由云端语义环节统一过滤。针对上述问题，本文设计“场景预适配—难例精调”两阶段迁移学习范式。

第1阶段：场景预适配。以COCO预训练权重[14]初始化模型，利用生成的合成样本进行全参数微调，每批次以固定比例混入少量真实样本，使模型学习通道异物在各类机房场景下的外观分布与空间关系。该阶段采用标准检测组合损失[15-17]，如式(6)所示。

$$\mathcal{L}_1=\mathcal{L}_{\text{box}}+\mathcal{L}_{\text{cls}}+\mathcal{L}_{\text{DFL}} \tag{6}$$

其中，$\mathcal{L}_{\text{box}}$为CIoU边界框回归损失，$\mathcal{L}_{\text{cls}}$为二元交叉熵分类损失，$\mathcal{L}_{\text{DFL}}$为分布焦点损失。

第2阶段：难例精调。以预适配阶段的收敛权重为初始化，利用真实采集样本对预适配模型进行精调。该阶段引入难例挖掘机制，将预适配模型在真实验证集上的低置信度检测、漏检目标及高置信度误报归入难例集合H，对难例施加额外梯度加权，其损失函数为：

$$\mathcal{L}_2=\mathcal{L}_{\text{box}}+\mathcal{L}_{\text{cls}}+\mathcal{L}_{\text{DFL}}+\lambda_h\sum_{(x,y)\in H}w_h(x,y)\times\mathcal{L}(f_\theta(x),y) \tag{7}$$

其中，λ_h为难例加权系数，$w_h(x,y)$为样本级权重。通过提升难例的梯度贡献，驱动模型聚焦边缘场景与易混淆目标的判别。训练评估采用召回倾向的加权指标替代标准mAP：

$$\mathcal{F}=0.4R+0.4\text{mAP}_{50}+0.2\text{mAP}_{50:95} \tag{8}$$

其中，R为召回率，mAP_{50}与$\text{mAP}_{50:95}$分别为IoU阈值0.5与0.5:0.95下的平均精度。召回率与粗定位精度各赋0.4的权重，引导模型优先保障召回，降低漏检率。

4 多模态语义研判与协同调度

边缘侧检测器输出的候选框仅标识了疑似异物的位置，无法区分固定设施与临时堆物，也难以量化其对通行空间的实际影响。本文引入多模态大模型[18]，将候选框位置与结构化运维规则共同编码为推理条件，联合进行语义理解与风险判定，并通过动态优先级调度协调多路视频流与有限算力资源。

4.1 基于结构化规则约束的风险语义研判

通道异物的属性判别与通行影响评估涉及空间关系推理、运维规则匹配及场景语义理解等多维度信息的融合，传统规则引擎难以覆盖开放环境下的多样化判定需求。本文将上述判别任务统一于多模态大模型的推理框架，通过结构化条件注入实现规则约束下的风险判定，整体研判流程如图3所示。

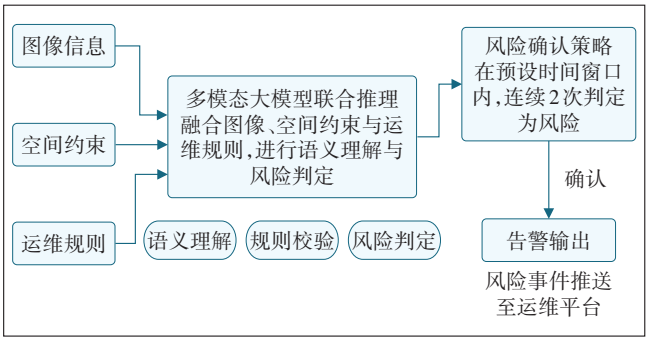


图3　多模态语义研判流程

将当前帧图像I_t、候选框坐标b_i、通行空间约束集S及结构化运维规则K一并输入多模态大模型，由模型联合完成语义推理与风险判定：

$$\mathcal{R}_i=\mathcal{M}(I_t,b_i,S,K) \tag{9}$$

其中，S包含通行区域边界、关键路径走向及最小通行净宽等空间约束；K覆盖固定设施排除、临时作业豁免及应急通道不可占用等判别规则。模型输出结构化判定结果，包含阻塞判定q_i、风险等级g_i、对象类型及固定设施标识f_i。最终通行风险判定为：

$$z_i=\mathbf{1}\left[q_i=1\wedge f_i=0\wedge g_i\in\{\text{medium},\text{high}\}\right] \tag{10}$$

仅当模型判定存在阻塞、对象为非固定设施且风险等级达到中高时，系统确认构成通行风险；固定设施与低风险场景不触发告警。

单帧判定易受光影变化或人员临时经过等因素干扰。为此，引入时序确认机制，同一摄像头在预设时间窗口内连续2次判定为风险，方触发告警，以滤除临时堆放、人员短时聚集及检测抖动造成的误报。

4.2 动态优先级调度与反馈闭环

实际部署中，数百路摄像头并行运行边缘检测，候选框的产生速率远高于多模态大模型的推理吞吐。

先全量检测再集中研判的传统串行方式，易使高风险候选因排队延迟错过处置窗口[19]。为此，本文设计动态优先级调度机制，将检测与研判解耦，大模型在并发能力范围内按优先级异步消费队列中的候选。

调度优先级综合考虑候选置信度、摄像头场景权重与大模型当前的负载[20]，定义为：

$$P_i = \alpha \times s_i + \beta \times w_k - \gamma \times L_t \quad (11)$$

其中，s_i为候选置信度，w_k为第k路摄像头的场景权重（核心通道取1.0，普通通道取0.6，边缘区域取0.3），L_t为大模型当前负载率。调度器在最大并发数约束下按优先级提交，超限候选排队等待。

5 实验验证与工程应用

本章从实验评估与工程部署2个维度验证所提方法。实验部分基于自建数据集验证两阶段训练各环节的性能贡献；工程部分依托华东地区某大型数据中心，验证系统在生产环境下的运行效果。

5.1 实验设置

数据集共包含2 000张图像，由真实采集图像与合成图像2部分构成。其中，真实采集图像共200张（通道正常通行背景图像120张，含临时堆物的异常图像80张），合成图像共1 800张。合成图像以Seedream 4.5为生成基座，将真实背景与堆物主体作为输入，基于第2章所述的场景先验约束引导文本注入扩散过程生成，并经清晰度与场景合规性校验后保留。数据集按8∶1∶1划分为训练集、验证集与测试集。部分合成样本与真实样本对比如图4所示。

检测模型选用YOLOv11s，训练流程按第3章所述分2个阶段执行。第1阶段以COCO预训练权重初始化，在合成样本上全参数微调150轮，初始学习率为1×10^{-3}，余弦退火衰减至1×10^{-5}，批大小为16；第2阶段在280张真实样本与难例回流样本上精调50轮，初始学习率为1×10^{-4}，余弦退火衰减至1×10^{-6}。训练平台配置NVIDIA GeForce RTX 3090 GPU，推理部署于搭载Intel Xeon Gold系列处理器的Linux服务器。评估指标采用精确率P、召回率R、mAP@0.5及mAP@0.5∶0.95。



图4　真实样本与合成样本对比示例

5.2 检测性能评估

为验证两阶段迁移训练各环节的性能贡献，设计消融实验并与同等量级主流检测模型进行对比，从定量与定性2个方面开展评估，其性能对比与真实场景中的检测效果分别如表1和图5所示。

表1　检测性能对比

方法	精确率	召回率	mAP@0.5	mAP@0.5∶0.95
COCO预训练（基线）	0.585	0.392	0.443	0.209
直接微调（合成样本）	0.812	0.584	0.635	0.398
本文两阶段训练	0.926	0.730	0.797	0.511
YOLOv8s*	0.873	0.662	0.704	0.452
YOLOv10s*	0.901	0.695	0.741	0.483

注：*表示采用与本文所提的两阶段召回优先训练策略与数据集，仅更换检测模型。



图5　验证集真实场景中检测效果

由表1可知，与COCO预训练基线相比，直接微调方案的召回率由39.2%提升至58.4%，精确率与mAP@0.5分别提升22.7和19.2个百分点，验证了场景化训练的必要性。两阶段训练在此基础上进一步提升性能，召回率达73%，精确率为92.6%，mAP@0.5升至0.797，其中第1阶段通过合成样本扩展长尾场景特征覆盖；

第2阶段经难例精调强化边缘场景判别能力。在同等训练策略下，YOLOv8s与YOLOv10s均获得明显增益，YOLOv11s各项指标最优，验证了该训练策略的有效性及YOLOv11s架构的场景适配优势。

由图5可知，模型在纯净背景下无冗余误检，可准确识别包装箱、工具推车等异物，在弱光、局部遮挡等工况下仍具备稳定检出能力。

5.3 工程部署与运行效果

本文方法已在华东地区某大型数据中心落地部署，按功能区域分级布设286路监控点位，覆盖5栋机房楼宇，连续稳定运行3个月。系统采用云边端三层协同架构（见图6）：端侧通过RTSP协议接入摄像头视频流，按5 min间隔抓取通道图像；边侧运行YOLOv11s推理服务，完成候选发现与优先级调度；云侧基于昇腾910B部署Qwen3-VL-235B多模态大模型，承载语义研判与告警管理。

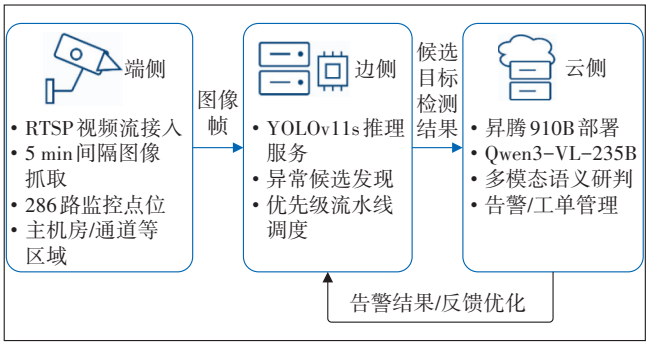


图6　云边端协同系统架构示意

运行期间，系统有效识别多类典型通道阻塞事件，典型告警输出样例如图7所示，涵盖废旧蓄电池箱体长期占用应急通道入口、施工余料与包装箱堆放于设备搬运路径、机柜走道杂物未及时清理等场景。经多模态研判确认为中、高风险的事件自动同步至基础设施智慧运营平台，运维人员可在线查看现场抓拍画面、风险等级与点位位置，完成告警接单、现场处置、结果回单的全流程闭环操作；平台配套短信、语音电话双通道推送，异常发生后第一时间触达运维责任人，有效缩短响应链路。

系统核心运行指标对比如表2所示。结果表明，所提大小模型协同架构结合动态优先级调度机制，可将大模型单轮调用量压缩约85%，巡检频次提升至8~10轮/h，异常事件响应时间由小时级缩短至分钟级。运行期间系统共触发告警328次，经现场核验，真实通道阻塞事件307起，告警准确率达93.6%；误报14起，主要为楼梯间金属扶手、通道闸机等固定设施因光影偏移、视角变化被误判为临时异物；漏报7起，主要由堆物颜色与地面相近、个别非常规物料未纳入训练集导致边缘检测器漏检。系统持续运行可有效遏制异物堆放、消防通道堵塞等违规行为，为数据中心安全管控与消防合规提供了自动化保障手段。


IDC/B1电梯厅1新增气瓶和警示牌靠近通道左侧墙边，部分占用通道导致行人需绕行


DC/1F卸货区新增多人及大量堆叠纸箱、手推车占据通道中央及两侧，导致行人无法正常直行


西10F走廊中新增工具箱和木板位于通道右侧靠墙，部分占用通道导致行人需侧身通过


数据中心/A2-1C-C10 1F货梯厅走廊新增堆叠在托盘上的白色箱子位于通道中央，占据大部分宽度并影响通行


局/西1F电梯口新增设备架及线缆导致严重堵塞


房/A2西大门出现纸箱、工程器械/工具箱、木制托盘及其他杂物堆积在通道上，占据通道宽度约35%，可能影响通行

图7　典型通道阻塞告警输出示例

表2　系统运行指标对比

指标	全帧轮询方案	本文方案
单轮调用次数	≈300次	≈45次
单轮巡检时间	≈42.5 min	≈5.6 min
巡检频次	1~2轮/h	8~10轮/h
异常响应时间	2~4 h	5~10 min
人力成本	6人三班制	2人复核制

6 结论

针对数据中心通道异物占用检测中异常样本稀缺、语义判别能力不足、大模型全量推理成本偏高的问题，构建了“数据增强—边缘检测—云端研判—协同调度”的全链路智慧运营体系。首先，提出基于场景结构先验的可控生成式数据增强策略，将有限真实采集样本拓展为跨场景、跨形态的高质量数据集，有效缓解长尾运维场景的样本分布失衡问题；其次，设计场景预适配与难例精调两阶段迁移训练范式，缩小生成样本与真实画面的特征分布偏差，使轻量化检测模型在有限标注条件下实际生产场景的召回率超95%；最后，构建融合结构化运维知识的多模态语义研判机制，实现规则约束下的自主风险研判。

本文方法已在华东地区某大型数据中心稳定运行3个月，大模型调用量压缩85%，巡检频次提升10倍，异常响应时间由2~4 h缩短至5~10 min，异常识别准确率超93%，综合人力成本降低75%，验证了其在大规模生产环境下的长期运行有效性。后续工作将围绕跨机房场景的泛化验证、大模型的轻量化部署，以及多源物理安全感知数据的融合分析展开，进一步拓展所提方法在数据中心综合安全管控中的应用深度。

参考文献：

[1] Cuevas C, Martínez R, García N. Detection of stationary foreground objects[J]. Computer Vision and Image Understanding, 2016, 152: 41-57.

[2] Hu T Y, Gao F, Zhou F Q. Railway obstacle intrusion detection and risk assessment based on MSIA-YOLOv8 and DALNet[J]. Expert Systems with Applications, 2026, 299(Part C): 130132.

[3] 郑含博，胡思佳，梁炎燊，等. 基于YOLO-2MCS的输电线路走廊隐患目标检测方法[J]. 电工技术学报，2024，39(13)：4164-4175.

[4] 安思齐，蔡昂林，马子程，等. 集成多模态大模型的施工安全隐患识别[J]. 中国安全科学学报，2025，35(9)：185-192.

[5] 董磊，吴福居，史健勇，等. 基于大语言模型的施工安全多模态知识图谱的构建与应用[J]. 计算机工程与应用，2025，61(9)：325-333.

[6] Hussain R, Lee D, Abbas M S, et al. Vision-language model-based intelligent assistant for onsite construction safety inspection[J]. Automation in Construction, 2026, 182: 106728.

[7] 许俊，程序，胡孝俊，等. 基于人工智能算法的数据中心机房气流组织温度预测研究[J]. 邮电设计技术，2022(12)：52-56.

[8] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 10674-10685.

[9] Peebles W, Xie S N. Scalable diffusion models with transformers[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023: 4172-4182.

[10] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020: 6840-6851.

[11] Esser P, Kulal S, Blattmann A, et al. Scaling rectified flow transformers for high-resolution image synthesis[C]//Proceedings of the 41st International Conference on Machine Learning. Vienna: PMLR, 2024: 12606-12633.

[12] Seedream T, Chen Y P, Gao Y, et al. Toward next-generation multimodal image generation[EB/OL]. [2026-01-20]. https://arxiv.org/abs/2509.20427.

[13] 杨勇，万超伦，马建友，等. 大模型数据扩增技术在输电线路智能监控中的应用研究[J]. 邮电设计技术，2024(9)：31-37.

[14] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context [C]//Computer Vision - ECCV 2014. Cham: Springer, 2014: 740-755.

[15] 许灵军，蔡庆宇，王常玲，等. 5G智能风电巡检无人机方案研究及应用[J]. 邮电设计技术，2025(2)：71-76.

[16] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[17] 刘林，杨志敏，黄强，等. 面向电力通信配电屏数字化的人工智能识别关键方法[J]. 电信科学，2025，41(4)：176-190.

[18] Bai S, Cai Y X, Chen R Z, et al. Qwen3-vl technical report[EB/OL]. [2026-01-20]. https://arxiv.org/abs/2511.21631.

[19] 董紫淼，孔令楠，徐亚楠，等. 基于异构算力边端协同的智慧巡检组网部署分析[J]. 邮电设计技术，2026(4)：16-21.

[20] Wang X F, Han Y W, Leung V C M, et al. Convergence of edge computing and deep learning: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2020, 22(2): 869-904.

作者简介：

王清华，工程师，博士，主要从事网络运营智能化技术研究工作；范纪明，高级工程师，硕士，主要从事网络IT系统架构设计工作；朱昊然，工程师，博士，主要从事数字孪生与仿真工作；秦真涛，高级工程师，硕士，主要从事算力基础设施规划与运维管理工作；李鲲，工程师，学士，主要从事网络智能运维技术研发工作；陈一丁，工程师，学士，主要从事网络IT系统架构设计工作。