

DOI:10.20033/j.1003-7241.(2026)10-0179-05

基于自相关性和模糊规则的风险源信息查询方法

蔡海滨, 肖建毅, 陈维汉

(南方电网数字企业科技(广东)有限公司, 广州 510000)

摘要:掌握风险源信息是各类生产生活中规避风险的重要手段,文中针对风险源信息的查询问题研究了基于自相关性分析和模糊关联规则的风险源信息查询方法。首先,基于图理论和信息边界初始化处理构建了风险源信息的数据路径,然后,基于适应度函数对风险源信息特征的自相关性进行分析,并利用候选集的隶属度划分进行基于模糊规则的风险源信息模糊集合查询。最后,通过实测数据集与传统检索方法进行对比实验。实验结果表明,该方法能够有效捕捉风险源数据隐含关联关系,大幅提升复杂场景下风险源信息查询的准确率与召回率,具备更强的容错性与适应性,可高效适配电力、化工等多行业风险源快速检索、风险溯源场景,为工业安全风险管控、智能风险预警提供可靠的技术支撑。文中所述方法对风险源信息查询的最高性能分别提升约 10.5% 和 22.5%,平均性能分别提升约 24.4% 和 27.9%,计算过程的标准差分别降低约 36.2% 和 45.8%,能够有效提高风险信息查询的准确性和效率。

关键词:风险源信息;自相关性;信息查询;模糊关联规则;风险管控;查询方法

中图分类号: TP311.13;TP391.3

文献标志码: A

文章编号: 1003-7241(2026)10-0179-05

Risk information source query method based on auto-correlation and fuzzy association rules

Cai Haibin, Xiao Jianyi, Chen Weihai

(CSG Digital Enterprise Technology (Guangdong) Co., Ltd., Guangzhou 510000, China)

Abstract: Mastering risk source information is an important means of avoiding risks in various production and daily life. This article studies a risk source information query method based on autocorrelation analysis and fuzzy association rules to address the problem of querying risk source information. First, data path for risk source information is constructed based on graph theory and initial information boundary. Then, autocorrelation of risk source information features is analyzed based on the fitness function, and a fuzzy rule-based fuzzy set query for risk source information is performed using the membership degree division of candidate sets. The experimental results show that the proposed method can improve the highest performance of risk source information query by about 10.5% and 22.5% and the average performance by about 24.4% and 27.9%. The standard deviation of the calculation process is reduced by about 36.2% and 45.8%, respectively, which can effectively improve the accuracy and efficiency of risk information query.

Keywords: risk source information; auto-correlation; information query; fuzzy association rules; risk management; query method

日常各领域的生产和生活过程中都存在着不同类型的风险,为了提高应对风险的能力,详细了解风险源信息是规避风险的有效手段之一。然而,由于风险源信息的种类繁多、结构繁杂,如何在大量的信息中获取有效的风险源信息是其中的关键。在这样的情况下,对风险源信息进行精准查询,获取有效的风险源信息成为各领域重点关注的问题。

针对上述信息查询问题已有较多的研究并取得了许多成果。例如,文献[1]基于历史数据信息进行分析,将历史数据进行预分区处理,得到历史数据的时空索引,对时空索引下的数据轨迹分布不均特性进行编码,在时空维度上建立历史数据的索引结构,根据该结构设定基于

Hbase 的存储模型,为该模型编写基于 Geohash 编码的信息查询分解算法,通过该算法分析信息的行为特征,进而得出信息查询结果。该方法存在的主要问题是信息查询数据量较小,难以应对大规模数据的情形。文献[2]建立信息查询的知识库,对知识库中的数据信息进行逐一搜索,将搜索得到的数据信息进行融合,分析数据融合的基础结构,对数据信息查询的范围进行标定,将融合的数据信息存储到临时表中,将信息查询的语句转换成为临时表能够读取的联合查询语句,在多表的近似范围内进行查询,多角度对信息查询结果进行精准定位,得出信息查询结果。该方法存在的主要问题是信息查询效率较低。文献[3]在大量信息中抽取样本数据信息,对样本数据信息

收稿日期:2024-11-14;录用日期:2024-12-23

基金项目:中国南方电网有限责任公司科技项目(080016KK52205)

作者简介:蔡海滨(1986—),男,硕士,工程师,研究方向为计算机软件。

引用本文:蔡海滨,肖建毅,陈维汉.基于自相关性和模糊规则的风险源信息查询方法[J].自动化技术与应用,2026,45(10):179-183.(Cai Haibin, Xiao Jianyi, Chen Weihai. Risk information source query method based on auto-correlation and fuzzy association rules[J]. Techniques of Automation and Applications, 2026, 45(10): 179-183.)

进行复杂聚合查询,根据这一初步查询结果进行全局扫描,并得出聚合计算结果,进而进行两阶段的分层抽样,对数据信息的聚合特征进行分组,再次进行随机抽样,对该次抽样结果进行评估,建立自组织特征映射网络,将评估结果进行聚类分组,得到近似聚合的信息查询结果。该方法存在的主要问题是计算过程相对复杂且信息查询结果准确度较低。

考虑到上述现有研究中所提出的信息查询方法无法满足风险源信息查询的现实需求,为了进一步提高对风险源信息的查询效率,本文研究了一种基于模糊关联规则的风险源信息查询方法。该方法在对风险源信息路径构建和风险源信息自相关性分析的基础上,通过模糊关联规则将风险源信息的属性权重进行聚类,进而实现更高准确率的风险源信息查询效果。

1 自相关性和模糊关联规则

1.1 风险源信息数据路径

为了提高风险源信息的查询效率,首先需要对风险源信息的数据存储路径进行梳理,将风险源信息的网络状态构建成为数据空间,该基于图的数据空间^[4]的表达如式(1)所示:

$$\theta = \frac{G(N,A)}{C} \quad (1)$$

式中, θ 表示风险源信息网络状态空间, G 表示信息有向图, N 表示空间中所有信息节点的集合, A 表示有向图的弧集合, C 表示风险源信息内容集。

对上述图数据空间中的信息路径进行统计,得到路径集合如式(2)所示:

$$P = \sum S_0, (\sigma \in \theta) \quad (2)$$

式中, P 表示信息路径集合, S_0 表示信息在空间中的源节点, σ 表示信息在空间中的状态。

将统计得到的信息路径集合 P 进行初始化处理,对信息的边界初始化值进行求解^[5-6],如式(3)所示:

$$\varphi = \frac{P}{\varepsilon} \cdot \lambda, (\varphi_\sigma \in \theta) \quad (3)$$

式中, φ 表示信息边界初始化值, ε 表示信息路径的边界集合, λ 表示初始化系数, φ_σ 表示初始化后的信息状态。

将信息边界的起点和终点进行连接,将连接的地方命名为合并点。进而对合并点进行分析得到信息路径与合并点的连接度如式(4)所示:

$$a = \frac{2(c/\varphi)}{c + mc} \quad (4)$$

式中, a 表示连接度, c 表示合并点, m 表示平均户信息量。

将经过上述处理后风险源信息路径统计,汇总到查询信息库中,以为后续风险源的进一步分析和查询提供支持。

1.2 风险源信息自相关性

对上述汇总信息库中的风险源信息进行特征分析,以

便后续进行进一步的查询,采用模块化结构对风险源信息的特征自相关性进行分析,建立风险源信息的适应度函数如式(5)所示:

$$f = w \cdot \delta \quad (5)$$

式中, f 表示适应度函数的表达式, w 表示信息路径节点子集, δ 表示权重系数。

适应度函数能够凸显风险源信息的特征,在查询过程中需要匹配信息特征,为了提高查询效率,在匹配的过程中,风险源信息特征将会进行融合聚类^[7-8],对两个信息特征融合聚类的关联度进行表达如式(6)所示:

$$l = f(M_i, M_j) \quad (6)$$

式中, l 表示特征融合聚类关联度, i 和 j 分别表示两个发生融合聚类的风险源信息特征, M 表示聚类簇。

对两个聚类簇的查询节点进行均衡调度,得到风险源信息的特征负荷如式(7)所示:

$$F = \frac{x(u - l \cdot \sin a)}{v \cdot \cos a} \quad (7)$$

式中, F 表示信息特征负荷, x 表示信息特征聚类簇的中心点, u 表示信息聚类分布尺度特性, v 表示基函数。

由此得出风险源信息特征自相关性特征量如式(8)所示:

$$V_i = w_i \sum_{i=1} (H_i - F_i) + b_i \quad (8)$$

式中, V_i 表示风险源信息特征 i 的自相关性, H_i 表示负载均衡控制变量步长, b_i 表示自相关性特征量频移。对风险源信息的自相关性分析能够有效提高对风险信息的查询效率。

1.3 基于模糊关联规则的查询方法

在上述分析风险源信息特征自相关性特征分析的基础上,进一步将对应信息的属性权重进行提取。首先计算信息属性权重的聚点深度,如式(9)所示:

$$D = |d_i - d_{i-1}| \quad (9)$$

式中, D 表示聚点深度, d_{i-1} 和 d_i 分别表示风险源信息特征进行融合聚类前后的空间距离。

根据计算所得的聚点深度即可将风险源信息的属性权重提取出来,该属性权重提取过程的示意如图1所示。

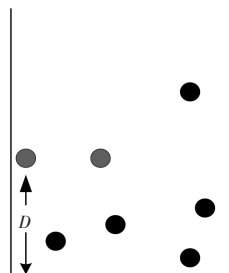


图1 属性权重提取示意图

Fig. 1 Schematic diagram of attribute weight extraction

在提取风险源信息属性权重之后,采用模糊关联规则对提取出来的信息属性权重进行聚类处理以提高相关信

息的查询效率。首先,添加风险源信息的候选项集,按照信息属性权重提取的时间进行隶属度划分如表 1 所示。

表 1 隶属度划分

Tab. 1 Classification of degree of subordination

提取时间	隶属度
t_1	1 B
t_2	(0.8, 0.2) B
t_3	(0.6, 0.4) B
t_4	(0.5, 0.5) B
t_5	(0.4, 0.6) B
t_6	(0.2, 0.8) B
t_7	0 B

根据表 1 中的隶属度计算候选项集的支持度,如式(10)所示:

$$\iota = \frac{k}{t} \tag{10}$$

式中, ι 表示候选项集支持度, k 表示项集的出现频率。

对支持度 ι 进行可信度定义,得到两个项集之间的支持度比值,如式(11)所示:

$$e = \frac{s(\iota_q \cap \iota_Q)}{s(\iota_q)} \tag{11}$$

式中, e 表示支持度比值, q 和 Q 分别表示两个候选项集, s 表示可信度。

基于式(6)中的候选集支持度比值,进一步构建基于模糊规则的模糊集合查询方法。基于上述候选集生成候选连贯模糊规则^[9-10]的及模糊信息查询过程如下。

1) 构建统计值的列联表,基于前述支持度比值 e 计算如式(12)所示项集模糊值。

$$M = \min \sum e \tag{12}$$

2) 对计算所得的模糊值 M 进行判定,大于查询所需的最小值则满足判定,生成频繁项集。

3) 生成模糊关联规则。根据生成的模糊关联规则对信息属性权重进行基于规则项集的聚类,完成基于模糊关联规则的信息属性权重聚类处理过程。

4) 根据聚类得到的风险源信息属性权重,进行风险源信息的查询和输出。每个风险源信息的属性权重如式(13)所示对应相关性语义:

$$W = (\prod z) \cdot \rho^{g-1}, \quad \rho \in (0,1) \tag{13}$$

式中, W 表示相关性语义, z 表示属性权重聚类值, ρ 表示语义衰减指数, g 表示语义属性量。

5) 根据风险源信息对应的相关性语义,将需要的查询语句关键词进行输入,即可匹配到对应的风险源信息,将风险源信息查询结果进行输出,完成风险源信息的查询过程。

2 实验结果

2.1 实验设置

为了验证文中所述基于自相关性和模糊规则的风险

源信息查询方法的有效性,文中采用模拟仿真数据的方法进行实验。为了保证实验结果贴近实践应用,实验所使用的数据采用真实风险源信息数据作为实验数据模拟样本。实验过程中所采用的数据量及风险源类型如表 2 所示。

表 2 模拟实验数据

Tab. 2 Simulated experimental data

风险源类型	模拟数据量/次
电子器械风险源	500
网络信息风险源	500
电气环境风险源	500
运营决策风险源	500
设备线路风险源	500
管线布置风险源	500

实验过程采用 RS-485 接口的 8 路模拟量输入采集模块装置将模拟数据采集至主控计算机,以模拟不同的样本数据源和数据的远程采集过程,该模拟量输入采集模块的主要技术参数如表 3 所示。

表 3 模拟量输入采集模块技术参数

Tab. 3 Technical parameters of analog input acquisition module

参数项目	技术参数
AI	8 路单端
电压采集	0~30 V
电流采集	4~20 mA
采集速度	1 kHz
DO	2 路单端
通讯接口	RS485
通讯协议	Modbus RTU
过压过流保护	45 V
功耗	25 mA
宽电压	DC 5~35 V

采用输入采集模块可将模拟所得的实验数据上传至主控计算机,实验过程中所使用的主控计算机参数配置如表 4 所示。

表 4 主控计算机参数配置

Tab. 4 Parameter configuration of main control computer

项目	配置参数
操作系统	Windows 10
节点服务器	CentOS
硬盘	24 TB
内存	256 GB
主频	2.7 GHz
处理器	Intel i7 12700

上述风险源数据模拟设备和模拟量输入采集模块和主控计算机搭建出本次模拟仿真实验环境如图 2 所示。

在上述实验设置的基础上,为了便于对风险源信息进行查询和结果展示。文中进一步开发了如图3所示的风险源信息查询结果可视化软件系统,该风险源信息查询软件能够将风险源信息进行排列显示,通过下拉键可以进一步获取详细的风险源信息。

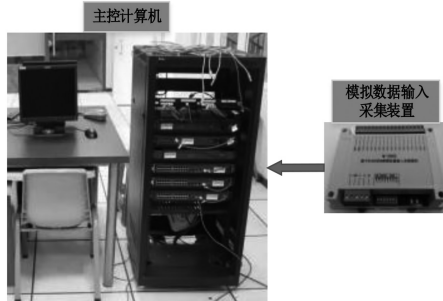


图2 实验环境及主要装置

Fig. 2 Experimental environment and main equipment



图3 风险源信息查询系统

Fig. 3 Risk source information query system

2.2 结果分析

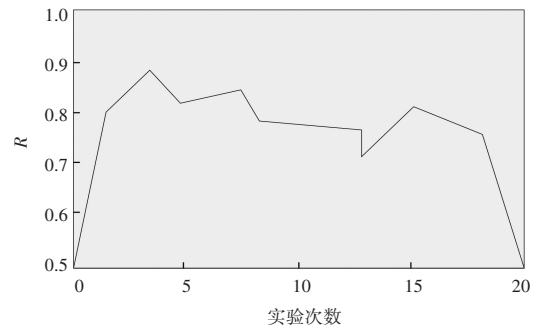
为了更加直观地体现出本研究所提方法的有效性,将本次模拟仿真实验的结果评价指标设置为如式(14)所示,查询得到风险源信息与标定的目标信息之间的相关性系数^[11-13]:

$$R_{(X,Y)} = \frac{C_{(X,Y)}}{\sqrt{D_{(X)}} \sqrt{D_{(Y)}}} \quad (14)$$

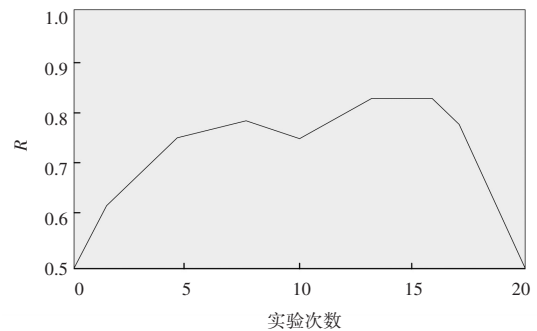
式中, X 和 Y 分别表示风险源信息查询结果和目标风险源信息, $R_{(X,Y)}$ 表示两者之间的相关性系数, $C_{(X,Y)}$ 表示两者之间的协方差, $D_{(X)}$ 表示信息 X 的方差, $D_{(Y)}$ 表示信息 Y 的方差。上述计算结果直接反映了风险源信息查询结果的优劣程度。相关性系数表达了风险源信息查询结果与目标查询结果的重合度^[14]。计算所得的相关性系数值越高,说明所得到的风险源信息查询结果越贴近真实的目标结果。

为了验证文中所述方法的优越性,分别采用文献[1]和文献[2]所提出的信息查询方法作为对照组进行风险源信息查询实验。同时,为了保证实验结果的可靠性,在采用不同的方法进行风险信息查询时,不同方法所采用的风险源数据量保持一致。此外,为了减少实验误差,实验过程中对于每个类别的风险源信息的查询实验分别进行

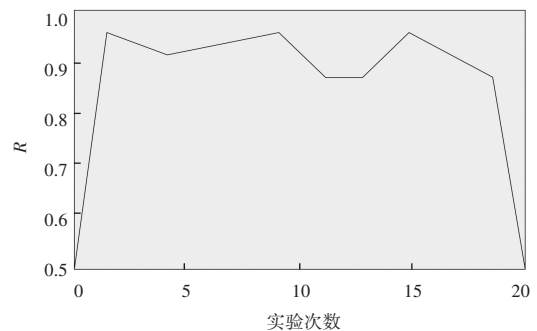
20轮次,最终结果取多次实验结果的平均值。不同方法各轮风险源信息查询结果对比如图4所示。由图4可知,相比于文献[1]和文献[2]中方法,文中所提方法在各轮实验中均具有较高的相关性系数 R ,最大值约为0.9637,而两种对比方法所得到的相关性系数最大值分别为0.8723和0.7864,文中方法相对于两种对比方法的性能相对提高约10.5%和22.5%。不同方法在20次风险源信息查询实验中的统计性能对比如表5所示,其中文中方法的平均相关性系数为0.9366,两种对比方法的平均相关性系数分别为0.7528和0.7324,文中方法相对于文献[1]和文献[2]的方法分别提高约24.4%和27.9%,此外,文中方法在20次实验中的 R 标准差为0.0807,相对于文献[1]和文献[2]的方法分别降低约36.2%和45.8%。上述实验结果表明文中所述风险源信息查询方法具有高的计算准确度,查询结果更接近真实的风险源信息,同时计算标准差较小,表明所述方法具有更好的计算稳定性,能够为风险源信息的查询提供更加高效的支持^[15]。



(a) 文献[1]方法



(b) 文献[2]方法



(c) 文献[3]方法

图4 不同方法查询结果对比

Fig. 4 Comparison of query results from different methods

表 5 不同方法性能对比
Tab. 5 Performance comparison of different methods

不同方法	指标 R		
	最大值	平均值	标准差
文献[1]方法	0.872 3	0.752 8	0.126 4
文献[2]方法	0.786 4	0.732 4	0.148 9
本文方法	0.963 7	0.936 6	0.080 7

3 结论

文中针对风险源信息的查询问题,研究了基于自相关性分析和模糊关联规则的风险源信息查询方法。所述方法首先构建了基于图的风险源信息的数据路径,然后对风险源信息进行自相关性分析,并在此基础上构建了基于模糊规则的模糊集合查询方法用于风险源信息的查询。实验结果表明,相对于对比方法文中所述方法对风险源信息查询的最高性能分别提升约 10.5% 和 22.5%,平均性能分别提升约 24.4% 和 27.9%,计算过程的标准差分别降低约 36.2% 和 45.8%,表明所述方法能够在有效提高风险信息查询准确性的同时保持更好的计算稳定性,从而能够进一步提高风险源信息的查询效率。后续研究中可进一步对风险源信息的聚类效果进行改进,以进一步优化文中所述方法的性能。

参考文献

[1] 李征宇,赵卓峰. 基于轨迹大数据时空分布的索引与查询方法[J]. 南

京航空航天大学学报,2022,54(3):528-536.
[2] 刘宏新,张一鸣,解勇涛,等. 联合收获机知识库数据多表联合查询方法研究[J]. 农业机械学报,2023,54(5):150-162.
[3] 房俊,赵博,左昌麒. 基于两阶段分层抽样的近似聚合查询方法[J]. 数据采集与处理,2022,37(5):1049-1058.
[4] 赵鑫,万英格,刘英博. 一种时序数据模式演化的跟踪与查询方法[J]. 计算机研究与发展,2022,59(9):1869-1886.
[5] 李朝奎,王露瑶,周新邵,等. 基于 HBase 的矢量空间数据存储与查询方法及其应用[J]. 地理科学,2022,42(7):1146-1154.
[6] 保富,马文,高宇豆,等. 基于逻辑回归模型的电力用户停电敏感度预测[J]. 微型电脑应用,2022,38(7):67-68,72.
[7] 陶洋,祝小钧,杨柳. 基于皮尔逊相关系数和信息熵的多传感器数据融合[J]. 小型微型计算机系统,2023,44(5):1075-1080.
[8] 何馥彤,李存斌. 基于 Logistic 回归的韧性城市电网信息物理融合系统风险源辨识[J]. 黑龙江电力,2024,46(2):138-145.
[9] 杨星星,冯依虎. 基于时间有效期的模糊关联规则优化算法[J]. 通化师范学院学报,2025,46(10):25-33.
[10] 刘世安,鲁海涛,张志芹. 基于模糊关联规则的深基坑支护结构变形量估计研究[J]. 砖瓦,2023(7):52-54.
[11] 张根柱. 基于关联规则挖掘算法的模糊多属性决策方法研究[D]. 哈尔滨:哈尔滨理工大学,2023.
[12] 陈瑜,沈颖洁. 计及 KNN 算法的医疗信息查询系统研究[J]. 微型电脑应用,2025,41(6):40-43,48.
[13] 倪乙鹏,徐久晟,舒德亮,等. 基于 Python 的地震监测中心站运维信息查询软件设计与开发[J]. 地震地磁观测与研究,2025,46(3):150-155.
[14] 蒋超伟,杨拯,弋立平,等. 基于 WebGL 技术的电力安全生产风险监测系统[J]. 自动化技术与应用,2025,44(1):57-60,136.
[15] 吴金珠,王天师. 基于动态贝叶斯网络的作业现场风险精细化控制方法[J]. 自动化技术与应用,2024,43(8):87-90,153.

(上接第 108 页)

5 结论

使用人工对大规模自动化物流输送线的产品进行批次信息采集,工作量大,速度慢,与上位机对接困难。采用机器视觉方法进行此项工作,则大幅度降低成本、提高效率和准确率^[16-17]。利用卷积神经网络自动提取特征的特点,对图像进行特征提取和识别,可以避免手工提取特征的不完整性和主观性,使得图像识别系统的构建变得更加简单、高效及智能。

本系统在自动化物流输送线上的应用,具有识别速度快、精确度高等特点,极大提高智能物流仓储信息管理系统效率,减少物流设备停滞时间,减少人工干预,保障现场工作人员安全,是实现现代智能物流管理的必要手段。

参考文献

[1] 尹洪岩. 大数据图像识别技术对智慧物流影响研究[J]. 中国储运,2024(4):192.
[2] 李平. 基于深度学习的图像识别技术研究[J]. 信息与电脑,2025,37(20):60-62.
[3] 高永超. 图像自动识别技术在物流自动化中的研究与应用现状分析[J]. 物流科技,2025,48(14):85-88.
[4] 金倩. 基于数字孪生的智能车间生产物流分拣系统虚拟调试研究与

应用[D]. 北京:机械科学研究总院,2025.
[5] 陈丰琦. 基于机器人的脐橙自动分拣系统研究[D]. 昆明:云南财经大学,2025.
[6] 杨劲松. 基于深度学习的多物体分拣系统研究[D]. 长春:吉林大学,2022.
[7] 田昊. 基于机器视觉的生产线智能分拣系统的研究与设计[D]. 锦州:辽宁工业大学,2022.
[8] 崔向东,王生飞,李峰,等. 基于机器视觉的条盒缺包检测设计[J]. 自动化技术与应用,2026,45(3):133-137.
[9] 范忠明. 基于神经网络图像识别技术的放顶煤煤矸自动识别方法[J]. 自动化技术与应用,2024,43(10):39-42.
[10] 徐卉清,王禾扬,史宇坤,等. 基于迁移学习的卷积神经网络在化石图像分类中的应用初探[J]. 古生物学报,2025,64(2):250-260.
[11] 刘昊,杨燕婷,余国庆,等. 基于卷积神经网络的树叶识别技术研究[J]. 商丘职业技术学院学报,2026,25(1):66-73.
[12] 张杨华. 基于卷积神经网络的天空图像蔚蓝程度分类方法[J]. 电脑知识与技术,2026,22(2):29-31.
[13] 夏文生,夏晓婧. 基于卷积神经网络的遥感图像分类综述[J]. 机电工程技术,2025,54(21):1-8.
[14] 高翠玲. 图像分析算法在无人果园巡检系统中的应用[J]. 农机化研究,2024,46(11):214-218.
[15] 郑永奇,廖明辉. 基于卷积神经网络的图像分类准确性提升研究[J]. 无线互联科技,2025,22(21):87-90.
[16] 盖文东,刘雅琨,荆刚,等. 基于机器视觉的物料自动分拣实验平台设计[J]. 实验室科学,2023,26(6):53-57.
[17] 吕宜光,郝明,孙凯明,等. 机器视觉技术在薄膜胶缺陷检测中的应用[J]. 自动化技术与应用,2025,44(5):131-134.