

本文引用格式: 樊玉平, 孙子正, 李雅芹, 等. 融合双模态目标检测的作业人员安全管控系统[J]. 自动化与信息工程, 2026, 47(2): 55-61.

FAN Yuping, SUN Zizheng, LI Yaqin, et al. Design of a worker safety management and control system integrating dual-modality object detection[J]. Automation & Information Engineering, 2026, 47(2): 55-61.

融合双模态目标检测的作业人员安全管控系统

樊玉平 孙子正 李雅芹 赵海月

(南京南自信息技术有限公司, 江苏 南京 210000)

摘要: 为进一步保障新能源发电场站作业人员安全, 设计一套融合双模态目标检测的作业人员安全管控系统。该系统采用融合预处理后图像和目标分割图像的双模态目标检测方法对作业人员及其附属物进行目标检测, 并利用分割一切模型提高目标分割精度。实验结果表明, 该系统提升了作业人员及其附属物的检测准确率, 提高了作业人员行为分析识别的可靠性, 可满足电力行业作业人员安全管控的高标准需求。

关键词: 融合双模态; 目标检测; 安全管控; 分割一切模型; 新能源发电场站

中图分类号: TP391

文献标志码: A

文章编号: 1674-2605(2026)02-0007-07

DOI: 10.12475/aie.20260207

开放获取

Design of a Worker Safety Management and Control System Integrating Dual-modality Object Detection

FAN Yuping SUN Zizheng LI Yaqin ZHAO Haiyue

(Nanjing Nanzi Information Technology Co., Ltd., Nanjing 210000, China)

Abstract: To further ensure the safety of workers in new energy power plants, a worker safety control system integrating dual-modality object detection is designed. The system employs a dual-modality object detection method that combines preprocessed images and segmented object images to detect workers and their accessories, and utilizes the segment anything model to improve the accuracy of object segmentation. Experimental results show that the system improves the detection accuracy of workers and their accessories, enhances the reliability of worker behavior analysis and recognition, and meets the high standards of worker safety control in the electric power industry.

Keywords: dual-modality fusion; object detection; safety management and control; segment anything model; new energy power plants

0 引言

随着能源转型深入推进, 风电、光伏等新能源发电进入快速发展阶段。新能源发电场站通常位于偏远地区; 而运维管理中心一般设在邻近城市, 距离新能源发电场站较远, 这种布局模式给安全运维带来挑战。尽管已采取诸多管理措施, 如严格执行“两票三制”管理^[1]、加强安全培训、佩戴防护用品等, 但由于受作业环境复杂、作业人员流动性大等因素影响, 人员伤亡事故仍难以避免。

随着人工智能技术的迅猛发展, 基于深度学习的作业人员安全管控技术已成为电力行业安全领域的重要辅助手段^[2-9]。该技术依托现有视频监控或工业电视系统, 融合目标检测与行为分析算法, 通过对海量作业场景数据的训练与推理, 可实时识别安全帽未佩戴、倒地等异常行为, 并即时触发预警, 从而实现对作业场景的实时监测、风险预警与动态管控。文献[2]提出一种基于 OpenPose 的人体姿态识别技术, 通过深度神经网络对作业人员骨骼图进行行为分析, 可有

效识别作业人员危险行为并发出预警。文献[3]提出电网场景下轻量化的安全帽佩戴状态检测网络,可实现立体化作业人员行为检测,在降低计算量的同时,识别精度达到 63.4%。文献[4]针对关键岗位作业人员提出一种基于视频的行为分析方法,在光线好、距离近的条件下,作业人员行为识别精度达到 90%。尽管上述研究在作业人员行为分析识别精度上取得了阶段性成果,但将其迁移至新能源发电场站时仍面临诸多挑战,如场站环境复杂多变,设备与管线层叠布置,导致作业人员目标特征易被遮挡;摄像机通常安装在高处,成像距离远、图像分辨率受限,降低了目标检测的可靠性。因此,目标检测性能已成为制约后续行为分析识别准确率的“卡脖子”环节。

为此,本文设计一套融合双模态目标检测的作业人员安全管控系统(以下简称作业人员安全管控系统)。该系统采用融合预处理后图像和目标分割图像

的双模态目标检测方法对作业人员及其附属物进行目标检测,并利用分割一切模型(segment anything model, SAM)提高目标分割精度,进一步提高了作业人员及其附属物目标检测的准确率,为实时识别作业人员安全帽未佩戴、吸烟、玩手机、倒地等异常行为提供可靠输入,并在检测到异常行为时发出告警信息,从而实现新能源发电场站作业人员的安全管控。

1 系统总体设计

作业人员安全管控系统充分考虑了新能源发电企业“统一调度、分级管理”的生产管理模式,以及新能源发电场站“无人值守、少人维护”的实际运维需求,采用分层、分布式、开放式构架。该系统主要包括感知计算层、网络层、平台层和应用层,其架构如图 1 所示。

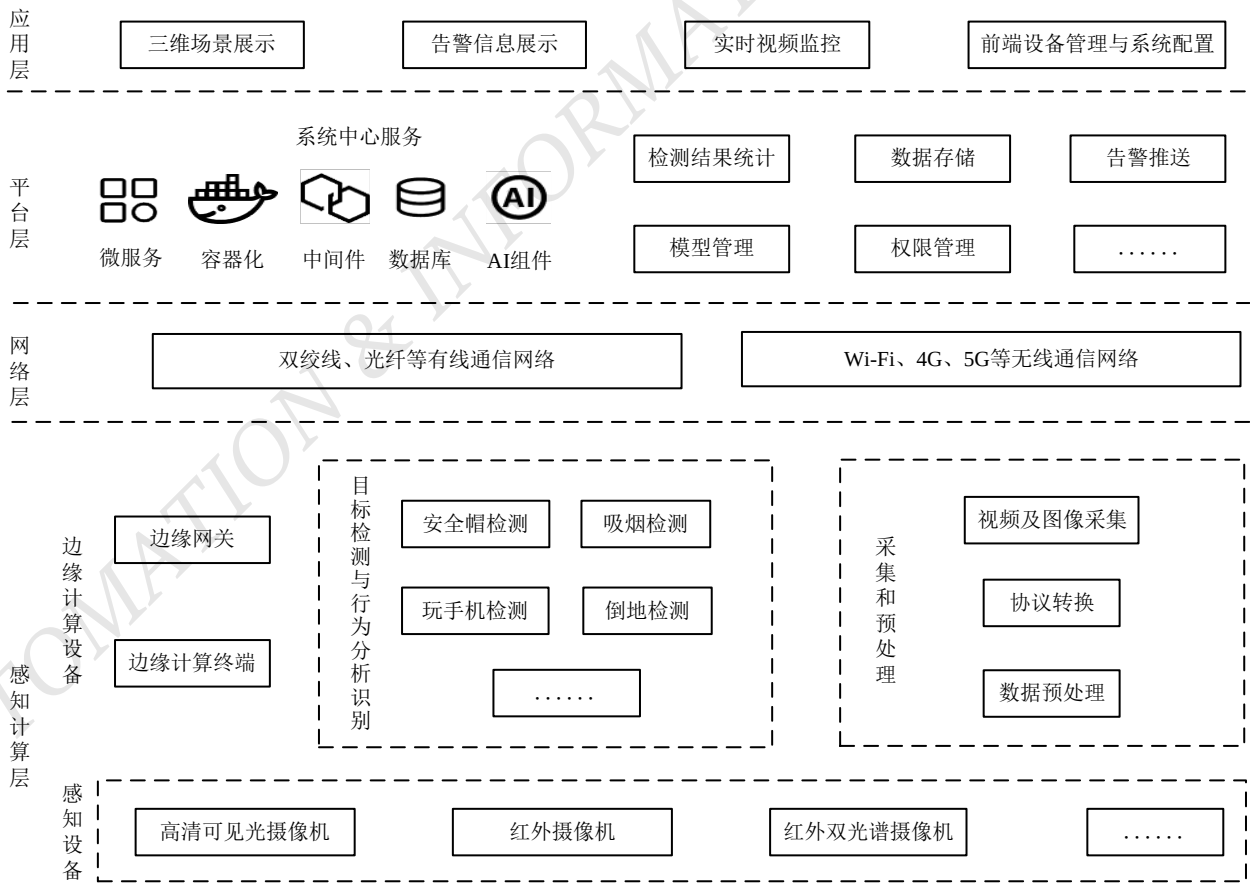


图 1 作业人员安全管控系统架构图

感知计算层是作业人员安全管控系统的数据基础层，用于对安全管控区域内作业人员的智能感知和边缘计算。通过在各生产区域、设备侧部署感知设备，如高清可见光摄像机、红外摄像机、红外双光谱摄像机等，采集作业场景的视频或图像数据；利用边缘计算设备，如边缘计算终端，完成作业人员目标检测与行为分析识别。在部署感知设备时，该系统依据“集约高效、因地制宜、充分利旧”的原则，与新能源发电场站现有监控系统无缝对接，即充分利用已有摄像机，并根据检测点位要求进行增补与调换，实现作业人员安全管控点位的全覆盖。边缘计算终端采用国产 GPU/NPU 芯片，内置视频及图像采集、协议转换、数据预处理等模块，兼容 PyTorch/TensorFlow 等主流框架，可在本地完成作业人员安全帽未佩戴、吸烟、玩手机、倒地等异常行为的目标检测与行为分析识别，识别结果通过边缘网关上传至网络层。

网络层主要用于数据传输。根据新能源发电场站的网络部署情况及前端设备特性，网络层的传输方式包括由双绞线、光纤等组成的有线通信网络和由 Wi-Fi、4G、5G 等组成的无线网络。

平台层以系统中心服务为核心，通常部署在新能源公司的运维管理中心，可接入多个风电场、光伏发电厂的作业人员安全管控系统。系统中心服务是基于微服务开发的 B/S 架构应用服务软件，具有检测结果统计、数据存储、告警推送、模型管理及权限管理等功能。

应用层是面向终端用户的统一客户端，负责可视化呈现与系统配置。用户可通过该层查看新能源发电场站的三维场景、告警信息、实时视频，并对前端设备进行管理 with 系统配置。

2 系统流程

作业人员安全管控系统流程如图 2 所示。

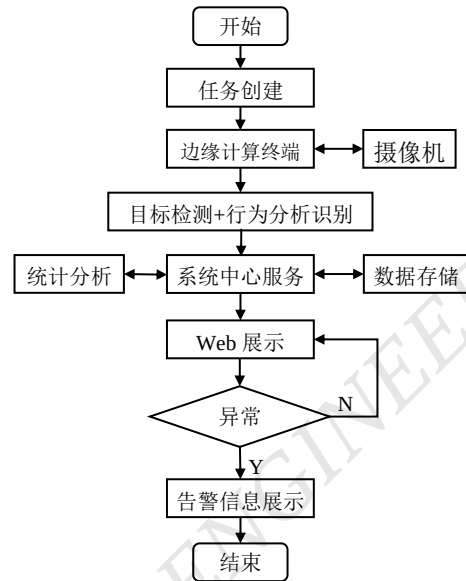


图 2 作业人员安全管控系统流程图

1) 通过客户端创建作业人员安全管控任务，即根据新能源发电场站的实际业务需求，为不同的安全管控区域配置相应的作业人员目标检测与异常行为分析识别任务（见表 1），设置任务执行时间、告警等级等参数，并通过 TCP 协议发送给边缘计算终端。

表 1 作业人员目标检测与异常行为分析识别任务

序号	安全管控区域	目标检测与异常行为分析识别任务
1	主变区域	倒地、吸烟、玩手机
2	开关室	安全帽未佩戴、倒地
3	继保室	安全帽未佩戴、倒地、吸烟、玩手机
4	GIS 室	安全帽未佩戴、倒地、吸烟、玩手机
5	配电室	安全帽未佩戴、倒地、吸烟
6	集控室	倒地、吸烟、玩手机
7	场站道路	安全帽未佩戴、吸烟、玩手机

2) 边缘计算终端接收到作业人员安全管控任务后，根据检测点位读取相应摄像机的视频或图像数据。

3) 利用融合双模态目标检测方法完成作业人员及其附属物目标检测，采用 YOLOv8 模型对检测到的目标进行行为分析识别，并将识别结果（检测点位、检测算法、检测截图等数据）通过 TCP 协议上传至系

统中心服务。

4) 系统中心服务接收到行为分析识别结果后, 进行统计分析与数据存储。

5) 在 Web 客户端实时展示作业人员位置, 若识别到作业人员的异常行为, 则弹出告警信息进行预警, 从而实现新能源发电场站作业人员的安全管控。

3 融合双模态目标检测方法

采用固定或可学习模板集的传统目标检测方法在目标特征(大小、外观、颜色)差异较大时, 有限的学习模板集难以充分捕捉目标特征变化, 导致目标检测准确率不足。本文采用融合双模态目标检测方法对作业人员及其附属物(安全帽、手机、香烟等)进行目标检测。融合双模态目标检测方法流程图如图 3 所示。

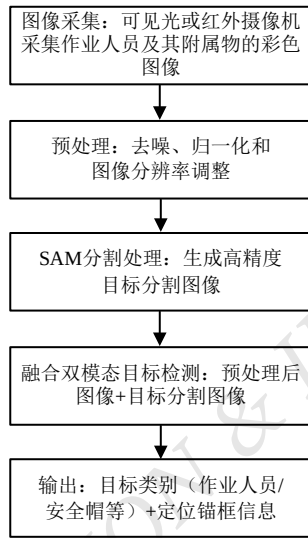


图 3 融合双模态目标检测方法流程图

融合双模态目标检测方法的具体步骤如下:

1) 图像采集, 利用可见光或红外摄像机等采集作业人员及其附属物的彩色图像;

2) 预处理, 采用基于小波变换的去噪方法, 抑制彩色图像中的光线、粉尘、遮挡等噪声影响; 采用最小-最大归一化方法, 将彩色图像的像素值缩放到 $[0, 1]$; 将图像分辨率统一调整为 640×640 像素, 生成预处理后图像;

3) SAM 分割处理, 通过 SAM 引导的自适应模

板生成机制生成高精度目标分割图像, 且分割掩码在目标分割图像中对所有显著的前景目标及其轮廓信息进行像素级标注;

4) 融合双模态目标检测, 将预处理后图像与其对应的目标分割图像共同输入 YOLOv8 模型, 并利用 SAM 提供的分割掩码进行目标检测;

5) 输出, 最终输出图像中所有待检测目标的类别信息(如作业人员、安全帽、手机等)和定位锚框信息(边界框位置坐标), 为后续作业人员的行为分析识别提供可靠的数据输入。

4 SAM 引导的自适应模板生成机制

SAM 引导的自适应模板生成机制以编码器-解码器网络为核心, 结合 SAM 分割结果的监督约束, 实现目标检测模板与图像内容的自适应匹配, 从而完成对目标图像的高精度分割, 具体步骤如下:

1) 在模型训练阶段, 为引导编码器网络学习并生成有效的目标检测模板, 采用监督约束策略: 将训练集中每幅图像预处理后输入 SAM, 获得像素级分割掩码, 以精准划分作业人员及其附属物前景目标区域与背景区域的边界; 将预处理后图像输入编码器网络, 输出对应的目标检测模板; 利用目标检测模板与 SAM 分割掩码的相似度约束, 迭代优化编码器网络参数; 利用余弦相似度函数衡量目标检测模板与 SAM 分割掩码的一致性, 编码器网络损失 J_E 的计算公式为

$$J_E = 1 - \cos(\varepsilon(I), G) \quad (1)$$

式中: J_E 越接近 0, 表明编码器网络生成的目标检测模板与 SAM 分割掩码的一致性越高; $\varepsilon(\cdot)$ 为编码器网络; I 为预处理后图像; $\varepsilon(I)$ 为编码器网络输出的输入图像目标检测模板; G 为训练集经 SAM 获得的分割掩码(像素级标注); $\cos(\cdot, \cdot)$ 为余弦相似度函数, 取值范围为 $[-1, 1]$;

2) 在推理阶段, 首先, 编码器网络接收预处理后图像, 并生成与该图像匹配的自适应动态目标检测模板; 然后, 通过解码器网络恢复动态目标检测模板

的特征信息；最后，将恢复后的动态目标检测模板与预处理后图像进行像素级乘法运算，形成加密输入图像并输入 YOLOv8 模型，加密公式为

$$T(I;S)=I \cdot S \quad (2)$$

式中： I 为预处理后图像， S 为生成的动态目标检测模板，二者通过像素级乘法运算可增强模板的前景目标特征，抑制背景干扰；

3) 本文采用 YOLOv8 作为目标检测模型，利用其无预先设定的锚框设计 (Anchor-Free) 和高效的 CSPDarkNet 骨干网络，使目标检测具备多尺度特征融合能力；其中，Anchor-Free 通过直接预测目标的中心点坐标和宽高，可自适应适配多尺度、不规则形态的目标，适合场景复杂、遮挡较多的新能源发电场站场景；CSPDarkNet 骨干网络通过跨阶段部分连接结构，将特征图分成两部分分别进行预处理、融合，既能满足边缘计算终端的实时性要求，又能有效区分作业人员与复杂背景，提高了目标检测准确率；

4) 针对新能源发电场站的实际应用场景，本文采用总损失函数对预训练的 YOLOv8 模型进行微调，使其适配动态目标检测模板引导的待检测目标前景与背景特征，进而输出优化后的目标类别与边界框信息；总损失函数 L_{total} 由目标分类损失 L_{cls} 、目标框回归损失 L_{box} 和目标置信度损失 L_{obj} 组成，计算公式为

$$L_{total} = \lambda_{cls} L_{cls} + \lambda_{box} L_{box} + \lambda_{obj} L_{obj} \quad (3)$$

式中： L_{cls} 采用交叉熵损失函数，用于优化作业人员、安全帽、手机等目标类别的分类精度； L_{box} 可同时优化边界框的位置、大小及长宽比； L_{obj} 用于区分前景目标区域与背景区域； $\lambda_{cls} = 1.0$ 、 $\lambda_{box} = 1.5$ 、 $\lambda_{obj} = 0.5$ 为损失权重系数，在新能源发电场站数据集上通过 5 折交叉验证确定，以平衡分类、回归及置信度预测的训练优先级；

5) 在目标检测模型训练时，YOLOv8 模型基于大量标注数据集进行学习，融合预处理后图像 (提供色彩、纹理细节) 与目标分割图像 (提供目标区域引导) 的优势，通过随机梯度下降算法进行权重参数 θ

的迭代优化，权重参数 θ 的更新公式为

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t) \quad (4)$$

式中： θ_t 为第 t 次迭代的权重参数； θ_{t+1} 为更新后的权重参数； η 为学习率，采用余弦退火策略动态调整，初始值设为 0.001，最小值可降至 0.000 1； $\nabla L(\theta_t)$ 为总损失函数 L_{total} 关于权重参数 θ_t 的梯度，负梯度方向为参数优化方向，以确保每次迭代均降低总损失。

5 实验验证

为验证本文提出的融合双模态目标检测方法的性能，选取 2 种主流目标检测方法 (Faster R-CNN、YOLOv8) 进行对比实验。实验环境：边缘计算终端采用国产 GPU，华为 Atlas 300I Pro 推理卡，输入图像分辨率为 640×640 像素。实验数据集源自某集团新能源发电场站采集并构建的 4 种作业人员安全管控图像，包括安全帽未佩戴、倒地、吸烟、玩手机图像各 2 000 幅，并按照 7:3 的比例划分为训练集和测试集。评估指标包含推理速度 (inference speed, FPS)、准确率 (accuracy) 和误检率 (false positive rate, FPR)，计算公式为

$$V_{FPS} = \frac{N}{T} \quad (5)$$

$$R_{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$R_{FPR} = \frac{FP}{FP + TN} \quad (7)$$

式中： N 为目标检测的图像总帧数； T 为消耗的总时间 (s)； TP 为输入图像中存在且被正确检测出的作业人员及其附属物样本数； TN 为输入图像中不存在且被正确检测出无的作业人员及其附属物样本数； FP 为输入图像中不存在但被检测出有的作业人员及其附属物样本数 (误检)； FN 为输入图像中存在但未被检测出的作业人员及其附属物样本数 (漏检)。

对比实验过程如下：

1) 完成实验准备工作, 包括搭建实验环境、准备数据集等, 确保本文融合双模态目标检测方法 with Faster R-CNN、YOLOv8 方法在相同的实验环境下进行测试;

2) 采用统一训练参数, 利用训练集对 3 种目标

检测方法所用模型进行训练, 直至所有模型均达到收敛状态;

3) 将训练后的 3 种方法所用模型部署在相同的实验环境下, 基于测试集开展性能测试, 记录测试结果并计算评估指标, 结果如表 2 所示。

表 2 3 种目标检测方法实验结果

目标检测方法	推理速度/(f/s)	安全帽未佩戴		倒地		吸烟		玩手机	
		准确率/%	误检率/%	准确率/%	误检率/%	准确率/%	误检率/%	准确率/%	误检率/%
Faster R-CNN	8.1	73.5	11.2	75.6	11.4	70.2	14.6	68.2	15.5
YOLOv8 (单模态可见光)	20.2	82.1	8.1	84.8	7.6	79.3	9.4	76.8	9.8
本文融合双模态	18.7	92.3	3.1	93.4	2.5	88.3	3.8	87.6	4.3

由表 2 可知: 本文融合双模态目标检测方法的推理速度较 Faster R-CNN 方法提升了 130.9%, 较 YOLOv8 方法下降了 7.4%, 该降幅处于可接受范围, 能够满足新能源发电场站实时监控的工程需求; 作业人员安全帽未佩戴、倒地、吸烟、玩手机 4 类场景的检测准确率较 YOLOv8 方法分别提升了 10.2、8.6、9.0、10.8 个百分点, 误检率降低了 5.0、5.1、5.6、5.5 个百分点, 证明了本文提出的融合双模态目标检测方法的有效性和性能优势。

6 系统应用

本系统已在福建省、内蒙古自治区、陕西省的多个风电场和光伏发电厂试点应用了一年, 取得了良好的试运行效果。其中, 作业人员安全帽未佩戴、倒地检测准确率达到 92% 以上; 吸烟、玩手机等检测准确率达到 87% 以上, 检测准确率较传统单模态检测方法提升明显。

7 结论

针对新能源发电场站复杂场景下的作业人员安全管控难题, 本文设计了融合双模态目标检测的作业

人员安全管控系统。该系统在目标检测环节融合预处理后图像与目标分割图像, 并引入 SAM 引导的自适

应模板生成机制来提高目标检测准确率。经实验与应用验证, 该系统优化了新能源发电场站作业人员安全管控效果, 提升了作业人员异常行为目标检测准确率, 能够切实保障新能源发电场站作业人员安全。

参考文献

[1] 全国高压电气安全标准化技术委员会. 电力安全工作规程 发电厂和变电站电气部分: GB 26860-2011[S]. 北京: 中国标准出版社, 2011.

[2] 王鸿, 邓元实, 常政威, 等. 基于深度学习的电力作业人员行为识别技术[J]. 四川电力技术, 2022, 45(3): 23-28.

[3] 胡戈飏, 林志驰, 郭政, 等. 面向电网施工人员识别的轻量化检测网络[J]. 沈阳工业大学学报, 2024, 46(3): 248-254

[4] 云旭, 宋焕生, 梁浩翔, 等. 基于深度学习的关键岗位人员行为分析系统[J]. 计算机工程与应用, 2021, 57(6): 225-231.

[5] 闫云凤, 陈汐, 金浩远, 等. 基于计算机视觉的电力作业人员行为分析研究现状与展望[J]. 高电压技术, 2024, 50(5): 1842-1854.

[6] 赵文聪. 基于 YOLOv5 算法的热电厂人员安全行为识别方法探究[J]. 黑龙江科学, 2024, 15(8): 98-101.

[7] 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.

[8] 李柯泉, 陈燕, 刘佳晨, 等. 基于深度学习的目标检测算法综述[J]. 计算机工程, 2022, 48(7): 1-12.

[9] 李琳琳, 孙海龙. 基于 YOLOv8 的林区分行人目标检测研究[J]. 森林工程, 2025, 41(1): 138-150.

作者简介：

樊玉平，女，1983年生，硕士研究生，工程师，主要研究方向：电力系统安防。E-mail: 307272712@qq.com

孙子正，男，1998年生，硕士研究生，工程师，主要研究方向：视觉人工智能。

李雅芹，女，1981年生，本科，高级工程师，主要研究方向：电力系统安防。

赵海月，女，1990年生，硕士研究生，工程师，主要研究方向：电力系统安防。

“人工智能赋能的机器视觉与智能检测”专题征稿启事

人工智能、机器视觉与智能检测技术的深度融合，正加速推动制造业数字化、网络化、智能化升级，是自动化、信息工程、智能制造与高端检测领域的重要前沿研究方向。随着深度学习、多模态感知、大模型、边缘智能、智能传感与自主决策等技术的快速发展，机器视觉已由传统图像采集与目标识别，逐步迈向复杂场景理解、缺陷精准检测、质量智能评估、过程状态感知、健康监测预警与检测决策优化等更高层次任务，在智能制造、先进装备、电子信息、医疗健康等领域展现出广阔的应用前景。当前，面向复杂工况、非结构化对象和高精度检测需求，人工智能赋能的机器视觉与智能检测仍面临诸多挑战，如多源异构数据融合困难、复杂环境下视觉感知鲁棒性不足等，严重制约相关技术的深入应用与产业化发展。与此同时，面向检测任务的智能算法与工业软件系统配套能力薄弱，难以保障复杂场景下高可靠、高实时的检测需求；相关前瞻性技术和基础共性技术研究布局不足，创新链与产业链深度融合不够，进一步制约了我国该领域技术的高端化、智能化发展。

为推动相关理论创新、关键技术突破与应用落地，本刊组织“人工智能赋能的机器视觉与智能检测”专题。该专题聚焦于人工智能、机器视觉、智能感知、智能检测与工程应用等方向，围绕视觉感知、图像理解、智能识别、缺陷检测、质量评估、多模态融合、在线检测、装备智能化及交叉应用等主题，特邀本刊编委刘桂雄教授、邓耀华教授、周松斌研究员、曹东教授、黄国健教授、叶廷东教授组织专题，通过多篇高水平论文，深度报道该领域发展方向，为机器视觉与智能检测产业战略性新兴产业发展提供科技支撑。诚邀相关领域专家学者围绕理论方法、关键技术、系统研发与工程应用踊跃投稿，共同推动该领域的学术进步与产业发展。

一、专刊投稿范围（包括但不限于）

- 1、机器视觉感知与图像智能分析
- 2、人工智能驱动的目标检测、识别与分割
- 3、面向复杂场景的缺陷检测与质量智能评估
- 4、多模态感知与视觉—传感信息融合检测
- 5、小样本、弱监督、自监督视觉检测方法
- 6、大模型与视觉基础模型在智能检测中的应用
- 7、工业视觉检测与智能制造质量控制
- 8、智能检测装备与在线检测系统

- 9、三维视觉、精密测量与高端检测技术
- 10、边缘智能与实时视觉检测技术
- 11、异常识别、故障诊断与预测维护
- 12、面向医疗健康的智能视觉检测与医学图像分析
- 13、面向中医药领域的智能识别、品质评价与检测技术
- 14、面向电子信息、机电装备与精密器件的视觉检测
- 15、检测数据建模、可信人工智能与可解释检测方法
- 16、机器视觉检测系统集成、工程应用与典型案例

二、投稿要求：请参考本刊官方网站 <http://www.giim.ac.cn/scipub/postdoc/> 或扫描右边本刊微信公众号的二维码获取“投稿须知”。

三、投稿截止时间：2026年6月30日

四、出版时间：2026年第4期（2026年8月28日）

五、投稿方式

- 1、登录 <http://gzxg.cbpt.cnki.net/WKE/WebPublication/index.aspx?mid=gzxg> 网上投稿或 aie@giim.ac.cn 邮箱投稿。
- 2、投稿时请务必注明“人工智能赋能的机器视觉与智能检测”专题投稿。

六、编辑部联系方式

邮箱：aie@giim.ac.cn；电话：(020)87687999

