

基于大小模型结合的货车车牌识别算法与应用

林兵, 颜燕

联通(广东)产业互联网有限公司, 广东 广州 510000

摘要

针对公路大型货车治理中存在的多摄像头视角异构、目标检测任务复杂、传统算法泛化能力不足等问题, 本文提出一种基于视觉语言大模型 QWen-2.5-VL 与 YOLO11 目标检测模型、PaddleOCRv5 文字识别模型结合的多模型智能监控算法。该算法针对道路监控系统采集的车头、侧面、车尾多视角图像, 实现吊车检测、车牌清晰度判定、车牌识别及车轴可见性判断四项核心功能。通过构建多任务独立并行推理框架, 将视觉语言模型的语义理解能力与目标检测算法的精准定位能力分别应用于不同的任务场景, 在真实场景数据集上的实验结果表明, 算法在吊车检测任务中 mAP 达到 94%, 车牌清晰度判定 F1 值提升至 96%, 车轴可见性判断 F1 值达 94.3%。本文的研究成果为公路货运车辆智能化监管提供了高效、可靠的技术方案, 具有重要的工程应用价值。

关键词

视觉语言大模型; 大小模型结合; 公路货车治理; 目标检测; 车牌识别

中图分类号: U495

文献标志码: A

Truck License Plate Recognition Algorithm and Application Based on Large and Small Model Collaboration

Lin Bing, Yan Yan

Unicom (Guangdong) Industrial Internet Co., Ltd., Guangzhou 510000, China

Abstract

To address issues such as heterogeneous camera perspectives, complex object detection tasks, and insufficient generalization capabilities of traditional algorithms in the governance of large trucks on highways, this paper proposes a multi-model intelligent monitoring algorithm based on the deep integration of the visual language model QWen-2.5-VL, the YOLO11 object detection model, and the PaddleOCRv5 text recognition model. The algorithm targets multi-angle images collected by road monitoring systems, including front, side, and rear views, to achieve four core functions: crane detection, license plate clarity assessment, license plate recognition, and axle visibility judgment. By constructing a multi-task independent parallel processing framework, it applies the semantic understanding capability of the visual language model and the precise positioning ability of the object detection algorithm to different task scenarios respectively. Experimental results on real-world scenario datasets show that the algorithm achieves an mAP of 94% in

crane detection, an F1 score improvement to 96% in license plate clarity assessment, and an F1 score of 94.3% in axle visibility judgment. The research outcomes provide an efficient and reliable technical solution for the intelligent supervision of highway freight vehicles, with significant engineering application value.

Key words

Vision-Language Model, large and small model combination, highway truck management, object detection, license plate recognition

0 引言

随着我国公路货运量的持续增长,大型货车(尤其是超限超载车辆)已成为影响道路交通安全、破坏道路基础设施的重要因素。文献[1]中总结得出,驾驶道路货运车辆引发的事故多且危害大,事故数占比18.57%,平均死亡人数为0.48人/起。传统的人工监管模式存在效率低、覆盖面窄、主观性强等弊端,难以满足大规模、全天候的监管需求。因此,利用计算机视觉与人工智能技术构建智能化的道路监控系统,实现对货运车辆的自动化检测与识别,已成为公路交通治理领域的研究热点。

当前,基于传统视觉的车辆检测算法已取得一定进展^[10-16],但在复杂场景下仍面临诸多挑战:一是道路监控系统多采用多摄像头部署,车头、侧面、车尾图像视角差异大,目标特征分布不均,车辆样式各异;二是车辆检测任务涉及吊车识别、车牌清晰度判定、车轴可见性判断等多维度需求,传统单一模型难以兼顾所有任务;三是恶劣天气、光照变化、遮挡等因素导致图像质量下降,影响算法鲁棒性。近年来,视觉语言大模型(Vision-Language Model, VLM)的快速发展为解决上述问题提供了新的思路。VLM通过融合视觉信息与语言语义,能够实现更高级的图像理解与推理能力,但其在目标定位精度与实

时性方面仍存在不足。

本文提出一种大小模型结合的智能监控算法,将视觉语言大模型QWen-2.5-VL与YOLO11目标检测模型、PaddleOCRv5文字识别模型相结合,充分发挥各模型的优势。具体而言,利用YOLO11实现对车辆关键部位的快速定位,通过PaddleOCRv5完成车牌区域的文字检测与识别,结合QWen-2.5-VL的语义理解能力实现车轴可见性判断等复杂推理任务。本文的主要贡献包括:(1)构建了多任务独立并行处理框架,实现多视角车辆图像的综合分析;(2)提出基于先验信息与PaddleOCR算法的车牌清晰度评估方法,有效解决低质量图像中的车牌识别难题;(3)设计了基于视觉语言模型的车轴可见性判断策略,提升复杂场景下的检测精度。

1 发展现状和特点

1.1 车辆目标检测算法

车辆目标检测是智能交通系统中的基础任务,传统算法如HOG+SVM、Haar特征+Adaboost等依赖手工特征设计,难以适应复杂场景。近年来,基于深度学习的目标检测算法取得突破性进展,其中YOLO系列算法以其高效性和实时性被广泛应用^[21-22]。文献[2]提出基于YOLOv8改

进的车辆检测算法，在精简后的 UA-DETRAC 数据集上在 mAP@0.5 达到 95.8%，但未针对多视角车辆图像进行优化。文献[3]采用 Faster R-CNN 实现车辆检测，精度提升至 99%，但模型参数量大，推理速度较慢。

1.2 车牌识别技术

车牌识别技术主要包括车牌定位、字符分割与字符识别三个步骤。传统方法如基于边缘检测的车牌定位、基于模板匹配的字符识别等易受图像质量及环境影响。在深度学习方法中，YOLO 系列算法常用于车牌定位，CRNN、Transformer 等模型则在字符识别任务中表现优异。PaddleOCR 作为一款开源的文字识别工具，集成了 DBNet 文字检测与 CRNN 文字识别模型，具有较高的准确率与速度。文献[4]提出基于 PaddleOCRv3 的车牌识别算法，在钢铁行业场景下对于清晰车牌的准确率达 99.6%，但未考虑车牌清晰度对识别结果的影响。文献[5]采用传统图像处理模块与模板匹配结合，达到了 93% 的识别准确率，但难以在道路复杂环境下稳定运行。

1.3 视觉语言大模型研究

视觉语言大模型通过预训练学习视觉与语言之间的关联，能够实现图像描述、视觉问答等复杂任务。QWen-2.5-VL 作为阿里达摩院推出的多模态模型，支持图像理解、文字识别、视觉推理等功能，具有较强的语义理解能力，但在交通领域的研究较少^[17-20]。文献[6]基于 QWen-VL 提出 GUI 自动导航方法 Qwen-VL-i VP，在构建的中文 GUI 导航任务场景上任务成功率较 QWen-VL 模型提高了 45.2%。文献

[7]提出基于视觉语言模型的工业缺陷检测多模态算法，将语言提示融入检测过程，提升了模型的泛化能力，达到 95.26% 的平均识别精确率，但难以应对强反光等复杂环境。

现有车辆检测与识别相关研究虽在各自任务上取得显著进展，但仍存在若干共性局限：其一，多数方法针对单一任务（如目标检测或车牌识别）独立设计，缺乏面向复杂交通监控场景的多任务机制，难以同时满足车辆定位、车牌识别、结构判断等多样化需求；其二，传统检测与 OCR 模型虽在特定子任务上性能优异，但普遍缺乏高层语义理解能力，无法有效完成如“车轴是否可见”等需结合上下文或图像质量判断的推理任务；其三，尽管视觉语言大模型（如 QWen-2.5-VL）具备较强的语义推理能力，但其本身并非为高精度目标定位或实时文字识别而设计，单独使用难以满足交通监控对效率与细粒度分析的双重要求。针对上述问题，本文提出一种大小模型结合的智能监控算法，通过合理分工与流程整合，将 YOLO11 用于快速车辆部件定位、PaddleOCRv5 用于车牌文字识别、QWen-2.5-VL 用于语义级推理判断，构建一个兼顾效率、精度与推理能力的多任务独立并行处理框架，从而提升复杂场景下车辆综合感知的实用性与鲁棒性。

2 算法总体框架

本文提出的大小模型结合智能监控算法总体框架如图 1 所示，主要包括图像预处理模块、多任务检测模块、视觉语言推理模块及结果整合模块四部分。多任务检测模块中各模型独立并行处理，算法总体流程：首先，对多摄像头采集的车头、侧

面、车尾图像进行预处理，包括图像增强、尺寸缩放等操作；其次，利用 YOLO11 模型对图像中的吊车及车辆关键部位（车牌）进行检测；通过 PaddleOCRv5 模型实现车牌区域的文字检测与识别，将结果与给定

车牌信息比对判定清晰度；Qwen-2.5-VL 模型独立基于图像语义完成车轴可见性判断；最后，整合各模块输出结果，生成最终的车辆监管辅助判定信息。

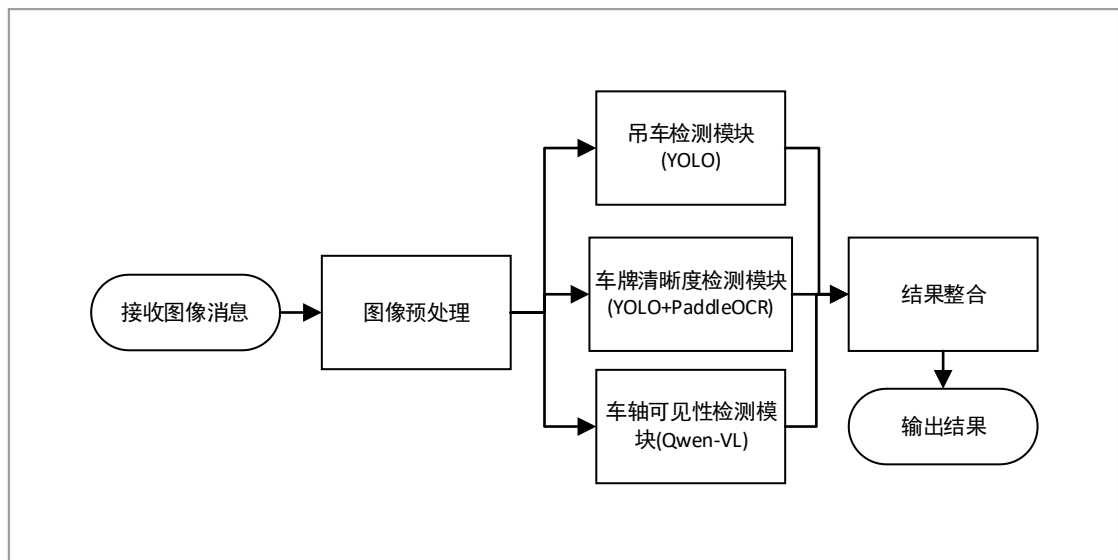


图1 大小模型结合智能监控算法总体框架

2.1 图像预处理模块

由于道路监控图像受光照、天气、拍摄角度等因素影响，质量差异较大，因此需要进行预处理以提升后续算法的鲁棒性。预处理模块主要包括以下操作：

- 图像增强：采用自适应直方图均衡化（CLAHE）增强图像对比度，减少光照不均对检测结果的影响；
- 尺寸缩放：调整图像尺寸以适应各模型的输入要求；
- 噪声去除：使用高斯滤波去除图像中的高斯噪声，采用中值滤波处理椒盐噪声。

2.2 多任务检测模块

多任务检测模块是算法的核心部分，负责实现吊车检测、车牌定位两项任务。该模块以 YOLO11、PaddleOCRv5 为基础模型，通过以下方式实现多任务检测：

- 吊车检测：针对车头图像，在 YOLO11 模型中添加吊车类别，通过标注真实场景中的吊车图像进行微调，实现对吊车的精准识别；
- 车牌定位：利用 YOLO11 检测图像中的车牌区域，输出车牌的边界框（bounding box）信息；
- 车牌识别：利用 PaddleOCRv5 检测识别裁剪后的车牌区域图像，输出车牌

字符信息。车牌覆盖黄色、黄绿混色、单行、多行等多种形态。

2.3 视觉语言推理模块

视觉语言推理模块基于 QWen-2.5-VL 模型实现，主要完成车轴可见性判断任务。该模块的工作流程如下：

- 车轴可见性判断：将侧面图像输入 QWen-2.5-VL 模型，通过提示“请判断图中车轴是否存在遮挡”，模型结合视觉信息与语义知识输出判断结果。

2.4 结果整合模块

结果整合模块负责整合多任务检测模块与视觉语言推理模块的输出结果，生成最终的车辆监管信息。该模块的主要功能包括：

- 结果筛选：对车牌识别结果进行筛选，保留置信度较高的车牌识别结果，并结合给定车牌信息定位指定车辆；
- 信息整合：将吊车检测结果、车轴可见性判断结果整合为结构化数据；
- 异常报警：当检测到侧轴遮挡、存在吊车车辆、车牌不清晰时，触发报警机制。

3 核心算法设计

3.1 基于 YOLO11 的多任务目标检测

YOLO11 模型采用 C2fNeXt 模块作为主干网络，该模块通过引入深度可分离卷积与残差连接，在减少参数量的同时提升了特征提取能力。针对多任务检测需求，本文对 YOLO11 模型进行以下改进：

- 多任务损失函数设计：采用联合损

失函数，包括目标检测损失（CIoU 损失）、分类损失（交叉熵损失），通过加权求和实现多任务优化；

- 注意力机制引入：在特征融合阶段添加 CBAM 注意力模块，增强模型对关键区域（如车牌、车轴）的关注；

- 数据集增强：采用亮度调整、平移、缩放等数据增强方法，提升模型的泛化能力。

改进后的 YOLO11 模型在验证集上的实验结果表明，模型在吊车检测任务中 mAP 达到 94%，车牌检测准确率达 99.1%。

3.2 基于 PaddleOCRv5 的车牌清晰度判定

PaddleOCRv5 模型由 DBNet 文字检测子模型与 CRNN 文字识别子模型组成，本文对其进行以下优化以适应车牌识别任务，识别效果如图 2 所示：

- 检测子模型优化：针对车牌区域的狭长特性，调整 DBNet 模型的锚框尺寸，提升车牌区域的检测精度；

- 识别子模型微调：收集大量真实场景中的车牌图像，构建车牌识别数据集，对 CRNN 模型进行微调，并限定车牌字符集合，提升字符识别准确率；

- 后处理优化：设计基于车牌规则的后处理算法，对识别结果中的错误字符进行修正，如将“0”与“O”、“1”与“l”映射。

优化后的 PaddleOCRv5 模型在车牌识别任务中，准确率达到 98.7%，提高模型对中等清晰度及以上车牌数据的识别能力，如图 3 所示。

3.3 基于 QWen-2.5-VL 的视觉推理



图2 清晰货车车牌识别结果

QWen-2.5-VL 模型采用 Transformer 架构，支持图像与文本的联合输入。对于车轴可见性判定这类现实数据收集困难、目标形态多样的复杂推理任务，小模型难以有效处理，因此采用 QWen-2.5-VL 大模型实现：其强大的语义理解与跨模态推理能力能够有效识别遮挡、模糊等复杂场景下的车轴可见性，弥补传统检测模型在高层语义判断上的不足：

提示工程设计：针对不同任务设计特定的语言提示，如侧轴遮挡判定提示为



图3 较清晰货车车牌识别结果

“请评估图片中货车的可见性质是否无遮挡，能够完整地看到车辆侧面轮轴”。车轴可见性判断具体提示工程设计如表 1 所示：

表1 车轴可见性判断提示工程设计

提示词组成	具体内容
系统提示词 (System Prompt)	<p>“你是一个专业的货车图像质量评估助手，专门为道路安全处罚提供技术依据。请严格按照以下要求回复：</p> <p>【核心任务】</p> <p>1. 评估图片中货车的可见性质是否无遮挡，能够完整地看到车辆侧面轮轴</p> <p>2. 提供标准化的判断结果</p> <p>【判断标准】</p> <p>- 可见性要求：车辆侧面轮轴部分是否无遮挡，无论是哪个侧面，有一个侧面无遮挡即可</p> <p>- 保守原则：有任何疑虑都应标记为“无遮挡”</p> <p>【注意事项】</p> <p>- 详细描述影响判断的具体因素</p> <p>{format_instructions}</p> <p>”</p>
用户提示词 (User Prompt)	“请评估图片中货车的可见性质是否无遮挡，能够完整地看到车辆侧面轮轴”
多模态输入 (Image)	待检测 base64 侧面车轴图片
输出格式约束 (format_instructions)	<p>{visibility: int = Field(description="车辆轮轴部位的可见性程度:0=完全可见,无遮挡,1=有遮挡")</p> <p>remarks: str = Field(description="具体的遮挡情况、拍摄角度问题或其他影响判断的因素")}</p>

在车轴可见性判断任务中，以QWen-2.5-VL输出结果中visibility的值作为判断车轴可见性的依据，准确率达到94.3%。

4 实验结果与分析

4.1 数据集构建

为验证算法的有效性，本文构建了一个真实场景下的公路货车多视角图像数据集，数据集包含以下内容：

- 数据来源：采集自某高速公路监控系统的车头、侧面、车尾图像，包含清晰/模糊数据，共计10000张；
- 标注信息：对图像中的吊车、车牌进行标注；
- 数据集划分：按照7:2:1的比例划分为训练集、验证集与测试集

4.2 评价指标

本文采用以下评价指标对算法性能进行评估：

- 目标检测任务：采用平均精度均值(mAP)作为评价指标；
- 分类任务（车牌清晰度判定、车轴可见性判断）：采用F1值作为评价指标。

4.3 实验结果

本文算法与现有单模态算法的对比实验结果如表2所示。从表2中可以看出，本文算法在各项任务中的性能均优于对比算法：在大型货车车牌清晰度判定任务中，将基于本文微调的PaddleOCRv5模型检测得到的车牌图像车牌字符识别结果与给定车牌字符信息计算莱文斯坦距离，结合车牌规则及置信度阈值完成车牌清晰度判定，F1值达到96%，较PaddleOCRv5模型提升18.8%；在车轴可见性判断任务中，F1值达到94.3%，较传统分类算法提升3.8%；在吊车检测任务中，mAP达到94%。

表2 算法性能对比

算法	吊车检测mAP(%)	车牌清晰判定F1值(%)	车轴可见性判断F1值(%)
YOLO11	92.9	-	-
PaddleOCRv5	-	77.2	-
ResNet50	-	-	90.5
本文算法	94	96	94.3

4.4 计算资源分析

基于NVIDIA A100 (80GB HBM2e)显卡环境计算模型的计算复杂度和计算开销，结果如表3所示，符合现场运行环境实时性需求与算力资源需求：

4.5 结果分析

实验结果表明，本文提出的大小模型结合算法能够有效提升公路货车监控任务的性能，主要原因如下：

- 大小模型结合优势：将视觉语言大

表3 模型计算复杂度与开销分析

模型名称	输入图像尺寸	参数量(M)	计算量(G FLOPs)	单帧推理时间(ms)
YOLO11s	704*1280	9.43	47.41	6.60
PaddleOCRv5-det	640*640	2.69	4.83	10.34
PaddleOCRv5-rec	48*224	6.78	0.87	4.89
Qwen2.5-VL-32B-Instruct-AWQ	512*512	32000	1200	800

模型的语义理解能力与传统视觉算法的精准定位能力相结合，弥补了单一模型的不足；

● 任务独立并行处理：通过多任务检测模块与视觉语言推理模块的独立并行工作，实现了多视角图像的多功能分析；

● 模型优化策略：对各模型进行针对性优化，提升了模型在特定任务上的性能。

同时，实验中也发现算法存在一些不足，如在极端恶劣天气（如暴雨、大雾）下，图像质量严重下降，导致车牌识别准确率降低。未来将进一步研究基于图像恢复的预处理算法，提升算法在恶劣天气下的鲁棒性。

5 结论与展望

本文提出一种基于视觉语言大模型与传统视觉算法深度融合的大小模型结合智能监控算法，针对公路货车治理中的多视角图像检测任务，实现了吊车检测、车牌清晰度判定、车牌识别及车轴可见性判断四项核心功能。通过构建多任务独立并行处理框架，算法充分发挥了各模型的优势，在真实场景数据集上的实验结果表明，算法综合性能优于现有单模态算法。

未来的研究方向主要包括以下几个方面：一是进一步优化大小模型结合策略，

提升模型的语义理解能力与检测精度；二是研究基于少样本学习的模型微调方法，减少对大规模标注数据的依赖；三是将算法部署到边缘计算设备上，实现实时、高效的公路货车监控任务。

本文的研究成果为公路货运车辆智能化监管提供了新的技术思路，使得执法人员人工阅读和审核车牌图片的工作量降低95%以上，大幅提高了工作效率，具有重要的工程应用价值。



参考文献：

- [1] 高轶男, 巩建强. 我国道路交通事故特征及成因分析[J]. 安全与环境学报, 2023, 23(11): 4013-4023.
- [2] 吴林, 曹雯. YOLO-SCDI: 基于改进YOLOv8的车辆检测算法[J]. 电子与信息学报, 2025, 47(10): 3967-3978.
- [3] 陈鑫鑫, 郭晓雯, 刘洪霞. 基于深度学习的车辆检测算法研究[J]. 江苏通信, 2024, 40(02): 114-118.

- [4] 许晨杰,金睿哲,韩帅峰,等.基于计算机视觉的钢铁行业车牌识别算法[J].信息技术与标准化,2022,(11):68-72.
- [5] 童宇行.基于边缘检测的车牌识别系统[J].信息记录材料,2025,26(02):184-186+207.
- [6] 庄凤云.基于视觉语言大模型的可进化GUI自动导航方法研究[J].江苏通信,2024,40(05):93-97.
- [7] 董禹彤,侯惠芳,龚明明,等.视觉语言大模型驱动的多模态工业缺陷检测与智能决策[J].计算机工程与应用,2026,62(08):176-188.
- [8] Zhang S, Chi C, Yao Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9759-9768.
- [9] Ma Y, Li Z, Zhang X Y, et al. PaddleOCR: A practical ultra-lightweight OCR system[C]// Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 4181 - 4183.
- [10] MEESAD P, THUMTHONG W. Advanced deep learning techniques for automated license plate recognition[J]. Scientific Reports, 2025, 15(1): 24967.
- [11] SATYA B, MANONGGA D, HENDRY, et al. Optimized YOLOv8 for Automatic License Plate Recognition on Resource Constrained Devices[J]. Engineering, Technology & Applied Science Research, 2025, 15(2): 21976-21981.
- [12] WANG X, WANG M, GUO H, et al. License plate recognition system for complex scenarios based on improved YOLOv5s and LPRNet[J]. Scientific Reports, 2025, 15(1): 34741.
- [13] MANE V, POKE A, RAIBAGKAR C, et al. Automatic number plate detection and recognition using YOLO world based convolutional neural networks[J]. Computers and Electrical Engineering, 2024, 120(Part A): 109668.
- [14] ZHANG X, YU S. A Lightweight License Plate Recognition Method Based on YOLOv8[J]. Electronics, 2025, 14(17): 3482.
- [15] GANESH M A, SARAVANA PERUMAAL S, GOMATHI SANKAR S M. Automated license plate authentication framework using multi-view vehicle images[J]. Journal of Intelligent & Fuzzy Systems, 2025, 48(5): 645-671.
- [16] NASCIMENTO V, LIMA G E, RIBEIRO R O, et al. Toward advancing license plate super-resolution in real-world scenarios: a dataset and benchmark[J]. Journal of the Brazilian Computer Society, 2025, 31(1): 435-449.
- [17] TIJINK M, LEVENDEEV S, NIEUWENHUIS E, et al. Vision on the Move: Automated Hazardous Material Plate Detection in Freight Transport [C]//Computer Analysis of Images and Patterns (CAIP 2025). Cham: Springer Nature Switzerland, 2025.
- [18] TIAN X Y, GU J R, LI B L, et al. DriveVLM: the convergence of autonomous driving and large vision-language models[C]//AGRAWAL P, KROEMER O, BURGARD W (Eds.). Proceedings of The 8th Conference on Robot Learning. Proceedings of Machine Learning Research, Vol. 270. [S.l.]: PMLR, 2025: 4698-4726.
- [19] GU Z P, ZHU B K, ZHU G B, et al. AnomalyGPT: detecting industrial anomalies using large vision-language models[C]// Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto: AAAI Press, 2024: 215.
- [20] GUAN W, LAN J, CAO J, et al. EMIT:

- enhancing MLLMs for industrial anomaly detection via difficulty-aware GRPO[EB/OL]. (2025-07-29) [2026-02-12]. <https://arxiv.org/abs/2507.21619>.
- [21] 张霞,乔环宇,曹峰.基于改进YOLOv8的高分辨率遥感图像目标检测算法[J].大数据,2025,11(05):152-169.
- [22] 彭森,张珂珂,杨一群等.基于改进YOLOv8的矿井复杂环境小目标检测算法[J].大数据,2025,11(06):166-181.

作者照片 林兵(1978-),男,硕士,联通(广东)产业互联网有限公司,高级工程师,公司副总经理,主要研究方向为人工智能、视频处理、云计算、网络安全。

收稿日期:XXXX-XX-XX

通信作者:

基金项目:作者照片 颜燕(1979-),女,硕士,联通(广东)产业互联网有限公司,高级工程师,团队长,主要研究方向为人工智能、视频监控、云计算、网络安全、视频安全。

Foundation Items: