

文章编号: 2095-4980(2025)12-1269-09

基于 DeepSeek-LoRA 微调的列车运行环境风险目标识别方法

崔广炎^{1,2}, 李宇杰¹, 李熙¹, 沈忱¹, 李琨¹, 刘可¹

(1.北京市地铁运营有限公司, 北京 100044; 2.北京交通大学 交通运输学院, 北京 100044)

摘要: 由于地铁列车运行环境复杂, 光照变化、遮挡频繁、目标尺度差异大等问题, 现有模型难以有效识别列车运行中的风险目标, 如对向列车、行人、落石以及异常侵入物等。针对上述问题, 本文探索了视觉语言大模型在地铁列车运行风险目标识别中的可行性。基于列车运行数据, 首先构建了基于低秩自适应微调(LoRA)的风险目标数据集, 将标注数据重构为结构化自然语言描述。通过低秩自适应微调 DeepSeek-VL2 多模态大模型参数, 获得了针对列车运行环境风险目标识别的优化权重, 并采用自然语言问答评估模型识别精确度。实验表明, 仅通过自然语言指令即可实现列车运行环境风险目标识别, $F1$ 分数达到 80.5%, 满足地铁列车运行场景中对对向列车和行人等风险目标的检测精确度需求, 可有效降低列车碰撞风险, 且模型具有极强泛化能力, 可应对多种地铁列车运行场景。

关键词: 列车运行风险识别; 大模型微调; 目标识别; 人工智能

中图分类号: TP273

文献标志码: A

DOI: 10.11805/TKYDA2025248

A DeepSeek-LoRA fine-tuned approach for risk object detection in railway operational environments

CUI Guangyan^{1,2}, LI Yujie¹, LI Xi¹, SHEN Chen¹, LI Kun¹, LIU Ke¹

(1.Beijing Mass Transit Railway Operation Corp. Ltd., Beijing 100044, China;

2.School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Due to the complexities of train operation environments—such as frequent lighting variations, occlusions, and significant scale differences among targets—existing models struggle to effectively identify risk objects during train operations, including oncoming trains, pedestrians, falling rocks, and abnormal intrusions. To address these challenges, this study explores the feasibility of applying large visual-language models to risk object identification in train operation scenarios. Based on train operation data, this paper first constructs a risk object dataset using Low-Rank Adaptation(LoRA) fine-tuning, reformatting annotated data into structured natural language descriptions. By applying LoRA fine-tuning to the parameters of the DeepSeek-VL2 multimodal large model, optimized weights for risk object identification in train operation environments are obtained. The model's recognition accuracy is evaluated through natural language question-answering tasks. Experiments show that risk objects in train operation environments can be identified using only natural language instructions, achieving an $F1$ -score of 80.5%. This meets the accuracy requirements for detecting risk objects such as oncoming trains and pedestrians in subway operation scenarios, effectively reducing train collision risks. Moreover, the model exhibits strong generalization capability to adapt to diverse subway train operation scenarios.

收稿日期: 2025-08-12; 修回日期: 2025-11-01

基金项目: 北京市地铁运营有限公司科研资助项目

引用格式: 崔广炎,李宇杰,李熙,等. 基于DeepSeek-LoRA微调的列车运行环境风险目标识别方法[J]. 太赫兹科学与电子信息学报, 2025,23(12): 1269-1277. DOI:10.11805/TKYDA2025248.

Citation format: CUI Guangyan,LI Yujie,LI Xi,et al. A DeepSeek-LoRA fine-tuned approach for risk object detection in railway operational environments[J]. Journal of Terahertz Science and Electronic Information Technology, 2025,23(12):1269-1277. DOI:10.11805/TKYDA2025248.

Keywords: risk identification for train operation; model fine tuning; object recognition; artificial intelligence

随着我国各地地铁线路建设增多, 列车运行环境风险智能监控系统的需求也随之增大。为减少列车运行风险, 提升安全等级, 需对运行环境中的行人、落石等危险物体进行及时的检测并预警。当前检测模型主要基于卷积神经网络(Convolutional Neural Network, CNN)和 Transformer 架构^[1-3], 包括 Faster R-CNN^[4]、YOLO^[5-6]、DETR(Detection Transformer)^[7]等通用检测模型, 为实现列车运行环境中风险识别提供了基础与可能。

近期, 基于 Transformer 架构的大语言模型, 如 Qwen^[8]、GPT^[9]、DeepSeek^[10]在自然语言处理领域展现出卓越的上下文理解与推理能力。通过结合视觉编码器, 如 Vision Transformer 等架构, 研究者开发出能同时处理图文信息的视觉语言模型(Visual-Language Models, VLMs)。最新发布的 DeepSeek-VL 系列大模型在多项视觉任务中表现突出, 为拓展至目标检测等复杂应用提供了可能。但由于列车运行环境恶劣, 经常出现不良光照、极端天气以及落石等异形障碍物, 使通用视觉语言大模型难以准确理解列车运行场景。首先是恶劣环境下数据干扰严重。列车运行环境中常存在强烈逆光、夜间照明不足、雨雪雾霾等恶劣天气条件, 导致图片质量下降, 噪声显著, 严重影响模型对风险目标的辨识能力^[11-12]。其次, 铁路沿线异形障碍物识别难度大。铁路沿线可能出现落石、倒树、非法侵入行人或大型动物等非常规风险目标, 这些目标形态多样, 出现频率低, 传统检测模型难以实现稳定检测^[13-14]。现有模型虽在专项检测任务中表现良好, 但本文旨在开发无需领域专用模块即可实现列车运行场景理解的风险目标识别模型。因此, 本文通过微调 DeepSeek-VL2-7B 模型, 构建新型列车运行专用模型 Rail-DeepSeek。DeepSeek-VL2 模型作为一种先进的大语言与视觉融合模型, 具有极强的表征学习和上下文理解能力, 特别适用于复杂环境下的语义感知与决策任务。本文基于自然语言指令完成风险目标识别, 通过多模态大模型深度融合视觉与文本信息, 提升风险目标检测精确度; 基于多模态大模型的高泛化性能, 实现对列车运行过程中罕见风险目标(如落石、非法侵入行人、小型障碍物等)的高精确度识别。利用大模型强大的小样本学习与语义推理能力, 解决传统方法在样本稀缺类别上识别精确度下降的问题, 为列车环境感知提供可靠保障。

本文基于现有列车数据集, 开发了专门的低秩自适应微调(LoRA)^[15]数据集, 数据被转换为适合视觉语言模型训练的自然语言指令格式。通过列车真实运行数据实验, 验证了方法的有效性。实验表明, 无需修改模型结构, 通用视觉语言模型(VLMs)即可实现精确的列车目标检测。

1 数据集构建

为使视觉语言大模型依据自然语言指令执行列车目标检测任务, 依据图 1 所示框架设计了数据集构建方案。此方案采用任务特定提示词引导模型定位与识别列车风险目标, 输入图像以占位符表示, 输出格式经标准化设计以确保检测结果一致性, 并简化坐标解析等后处理步骤。

数据集构建涵盖 3 个关键要素: 针对列车运营场景设计的自然语言提示词、模型训练所需的统一标注格式, 以及保留列车环境必需的空间与上下文信息。如图 1 所示, 该方法的核心在于通过自然语言指令引导多模态大模型完成列车检测任务, 并实现图像特征与文本语义的深度融合。针对典型列车场景图像, 输入提示被构造为: “你是一位视觉任务专家。这张铁路图像包含多个物体。请识别它们的位置, 并以 JSON 格式输出坐标。<image>”。在此框架下, 图像与文本的融合过程如下: 首先, 图像编码层基于图像特征提取器如 ViT 将输入图像编码为一组嵌入向量, 构成视觉特征序列; 同时, 文本指令通过模型的词嵌入层被转换为文本特征序列。随后, 在模型的 Transformer 解码器中, 通过交叉注意力机制, 文本序列中的每一个查询 token 都会主动与完整的视觉特征序列进行交互, 使模型在生成如“行人”、“信号灯”等类别词和坐标值时, 能够动态地聚焦于图像中最相关的视觉区域, 从而实现基于语义的视觉定位。简言之, 融合机制使模型能够真正理解文本指令的语义, 并在像素空间中找到对应的视觉实体。

模型的输出严格遵循预定义的 JSON 结构, 明确标注风险目标的类别与归一化边界框坐标 $[x_1, y_1, x_2, y_2]$ 。该输出格式将目标检测任务转化为结构化的语言生成任务, 不仅简化了坐标解析等后处理步骤, 其内置的置信度评分也直接满足了列车监控系统对检测结果可信度的使用需求。检测框架通过 2 个核心要素处理每个列车目标: 以 $[x_1, y_1, x_2, y_2]$ 格式定义的 2D 边界框坐标及检测目标的具体类别, 如信号灯、列车或障碍物。标准化预处理与数据集划分方案既可兼容现有大模型架构, 也可最大化利用 COCO 等标准图像数据集, 减少后续重复开发时间, 实现数据驱动型模型训练。

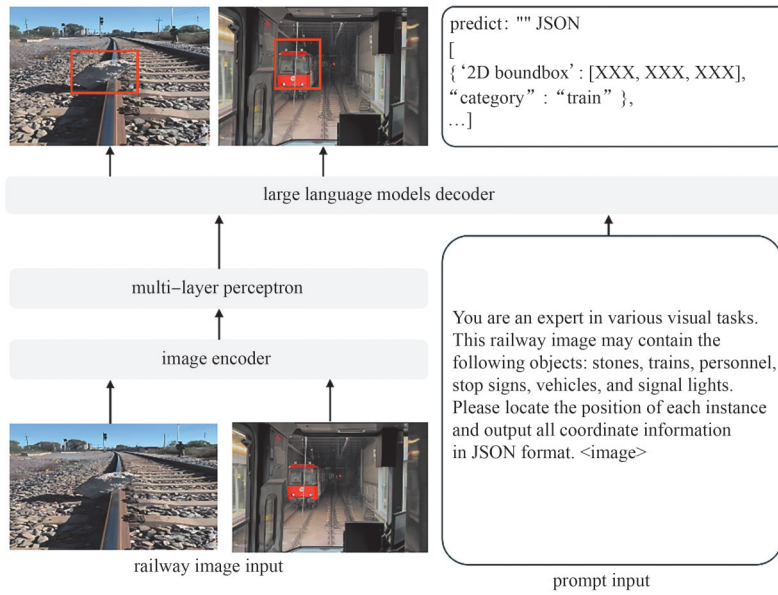


Fig.1 Model and dataset format diagram
图1 方法流程以及数据集格式示例

2 模型 LoRA 微调

为提升模型对列车风险目标的识别能力, 采用 LoRA 方法^[16]对通用大模型进行领域特异性微调。通用大模型虽广泛适用, 但缺乏列车场景专业知识, 对“列车行人”“信号灯”等术语理解有限。传统全参数微调方法计算成本高且易导致灾难性遗忘, 而 Adapter 等轻量方法又会在推理时引入延迟, 不适用实时性要求高的列车运行安全。因此本文设计了基于 LoRA 微调 and 模型知识蒸馏的高效大模型微调方法, 提升大模型在列车领域的风险目标识别能力。方法流程如图 2 所示, 分为预训练阶段和训练阶段。预训练阶段中, 通过参数量较大的教师模型, 基于知识蒸馏的方式指导学生模型学习列车领域知识; 训练阶段中, 基于 LoRA 微调方法, 利用图 1 中模型提示词实现大模型微调。

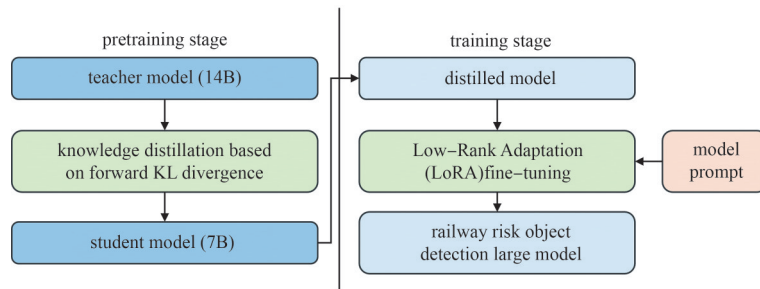


Fig.2 Flowchart of LoRA fine-tuning method
图2 低秩自适应微调方法流程图

首先通过知识蒸馏将大体量教师模型(本文采用 14B 教师模型)对地铁场景的认知迁移至 7B 学生模型。在知识蒸馏领域, 教师模型通常指复杂且性能强大的预训练模型, 负责提供知识指导; 学生模型是轻量化的待训练模型, 目标是通过学习教师模型的输出继承其核心知识, 从而实现模型压缩或性能提升。基于上述定义, 采用基于前向 KL(Kullback-Leibler)散度的方法对齐两者概率分布, 以提升学生模型对场景风险的辨识能力; 随后利用 LoRA 微调进一步强化模型对地铁信息的学习, 在保证效果的同时避免增加推理开销。

基于前向 KL 散度的知识蒸馏方法^[17]的轻量级特性不仅保持了跨模态特征学习的能力, 还为大规模模型部署提供了可行的压缩途径。通过优先考虑输出层对齐, 避免了传统知识蒸馏方法的巨大计算开销, 同时保留了教师模型预测分布中嵌入的基本知识, 从而为资源密集型多模式环境中的有效知识转移提供了一种实用的解决方案。前向 KL 散度知识蒸馏公式^[18]为:

$$\hat{q} = \arg \min \int_{-20}^{20} p(x) \ln^{(p(x)+q(x))} dx \tag{1}$$

式中： \hat{q} 表示KL散度的值，可以衡量教师模型和学生模型之间的差异； $\arg \min$ 操作表示取使该积分得到最小值时的 q 值； $p(x)$ 表示教师模型的概率分布； $q(x)$ 表示学生模型的概率分配。

通过最小化散度值 \hat{q} ，可使学生模型获得与教师模型相同的特征分布，从而使学生模型获得教师模型关于地铁风险辨识的知识。该公式确保学生模型接近教师的边界，同时避免计算密集型的特征对齐操作。最终知识蒸馏中教师模型和学生模型的KL散度拟合效果曲线如图3所示，其中纵坐标表示模型在不同变量下的概率，横坐标表示模型输入。可以看出，在不同的输入变量下，学生模型较好地模拟了教师模型的波峰位置，从而保证学生模型和教师模型在输入相同数据的情况下输出的概率分布基本一致，即输入数据相同，输出结果也相同。

LoRA提供了一种高效且轻量的模型训练解决方案。该方法冻结预训练模型的原始权重，仅在Transformer关键层的Query和Value投影矩阵中注入可训练的低秩矩阵，通过矩阵分解实现参数高效更新。这种设计使模型在微调阶段仅需优化极少量的参数，同时由于低秩约束，模型能够稳定保留预训练阶段学习到的通用知识，避免过拟合。此外，LoRA适配的权重在推理时可合并回原模型，不会增加任何计算负担，在真实场景的部署中更具优势。

采用如图4所示的方法实现LoRA微调训练。对于原始大模型的权重矩阵 $W_0 \in \mathbb{R}^{d \times k}$ ，本文通过引入低秩分解 $\Delta W = BA$ 约束其更新，其中 $B \in \mathbb{R}^{d \times r}$ 且 $A \in \mathbb{R}^{r \times k}$ ，且秩 $r \ll \min(d, k)$ 。在训练过程中， W_0 被冻结且不接收任何梯度更新；仅更新 A 和 B ：

$$h = W_0 x + \Delta W x = W_0 x + B A x \quad (2)$$

式中： h 为输出； x 为输入； ΔW 为LORA微调引入的新权重矩阵增量；低秩矩阵 A 和 B 分别为降维矩阵和重构矩阵，其中降维矩阵 A 的作用是将输入的原始高维特征投影到一个低维子空间，重构矩阵 B 将经过矩阵 A 降维后的低维特征映射回原始的高维空间，从而使其与原始大模型权重维度对齐。该微调方法在保持检测精确度的同时，可显著降低内存需求并加快训练速度。部署阶段，将预训练模型权重与学习到的LoRA自适应参数集成，构建无需增加计算开销的风险目标识别系统。

3 风险目标识别

完成大模型微调后，还需对大模型输出结果进行后处理，进一步去除干扰目标和误检目标，获取最终识别结果并告警。为实现高置信度的风险目标识别，引入基于概率置信度阈值筛选与空间聚类分析^[19]相结合的后处理机制。

首先通过大模型给风险目标的置信度进行阈值过滤，剔除置信度过低的目标^[20]，其极有可能是虚假目标，会造成误检。对于模型输出的每个目标边界框 b_i 及其置信度分数 p_i ，设定阈值 k 进行初步筛选，仅保留置信度高于 k 的预测结果：

$$B_{\text{filter}} = \{b_i | p_i > k\} \quad (3)$$

式中 B_{filter} 为过滤后的边界框集合。

为进一步提高召回率并抑制冗余检测，采用非极大值抑制(Non-Maximum Suppression, NMS)算法，其目标函数为：

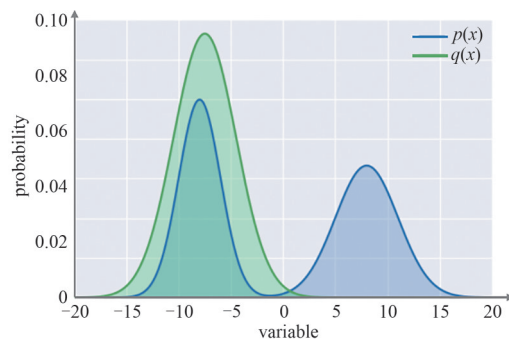


Fig.3 Forward KL divergence fitting curve

图3 前向KL散度拟合曲线

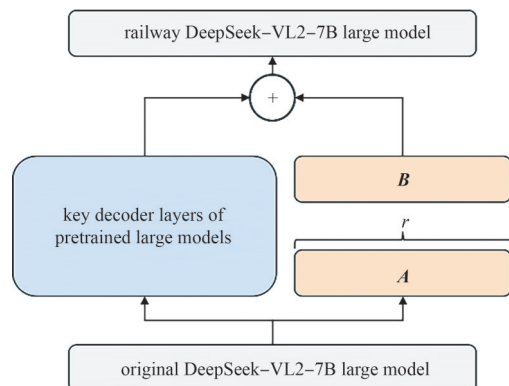


Fig.4 LoRA fine tuning

图4 LoRA微调示意图

$$s_i = \begin{cases} \text{keep}, & \text{IoU}(b_i, b_j) < \theta_{\text{NMS}}, j \neq i \\ \text{suppress} & \text{otherwise} \end{cases} \quad (4)$$

式中: $\text{IoU}(b_i, b_j)$ 表示两个边界框的交并比; θ_{NMS} 为预设重叠阈值。

此外, 地铁场景中真实风险目标, 如障碍物、侵入人员等通常具有空间聚集特性, 而误检往往呈随机分布。为区分二者, 采用基于密度的空间聚类(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)方法^[21]对检测目标进行聚类分析。

设检测目标集合为 $X = \{x_1, x_2, \dots, x_n\}$, 其中每个目标由其图像平面坐标 (x_i, y_i) 表示。DBSCAN 通过以下准则判断聚类归属:

$$N_\epsilon(x_i) = \{x_j \in X \mid \|x_i - x_j\| \leq \epsilon\} \quad (5)$$

式中: N_ϵ 为 DBSCAN 聚类函数; ϵ 为聚类阈值。若聚类结果 $N_\epsilon(x_i)$ 大于聚类阈值, 则目标 x_i 被标记为真实目标, 并以此扩展聚类簇, 最终得到聚类集合 $C = \{C_1, C_2, \dots, C_k\}$; 如聚类结果 $N_\epsilon(x_i)$ 小于聚类阈值, 则认为其中目标为误检, 予以剔除。经过上述后处理, 获取最终的列车运行环境风险目标结果, 并以此对列车运行系统进行告警, 保证列车运行安全。

4 实验验证

4.1 数据集准备

为评估本文方法有效性, 收集涵盖典型列车运行环境的数据, 包括公开数据集 COCO^[22] 的相关类别及 2 000 张含多种列车运行场景的图像。首先使用 LabelMe^[23] 工具, 标注轨道附近风险目标, 包括列车、行人等地铁场景常见的目标以及轨道线异物等风险目标, 并将标注转换为 COCO 格式以确保兼容性。数据集划分为 3 000 张训练图像和 1 000 张测试图像, 用于验证模型在真实条件下的检测性能。

该数据集可全面测试模型在复杂列车运行环境中识别风险目标的能力。由于铁路列车通常需要全天候运行, 经常受到雨雪雾等恶劣环境的影响, 同时夜晚低照度和复杂光线条件也会对自动驾驶列车行驶造成极大的影响。为全面测试列车运行环境, 本文数据集包含多种光照、天气和轨道配置等真实场景变化, 并且所有数据均为真实列车运行数据。如图 5 所示, 数据集中包含列车运行中常见的多种场景, 如地铁隧道、高架路以及常规轨道场景。环境方面, 该数据集包含了夜晚低照度和正常光线等多个天气条件, 并且本文还收集了大量轨道内异物图像数据, 模拟列车运行常见风险目标。



Fig.5 Sample images
图5 数据集示例图片

4.2 实验结果

为完成大模型微调的训练与测试, 利用 Llama Factory 框架^[24] 并采用全参数冻结的方式进行 LoRA 训练。所有训练均采用 4 张英伟达 RTX4090 显卡, 并采用 1 张英伟达 RTX3090 显卡进行推理测试。为控制训练变量, 每次训练轮数固定为 50 轮, 实际测试确定训练 50 轮时模型已完全收敛, 精确度基本保持稳定。

在列车数据集上评估不同秩配置的 LoRA 微调方法。针对列车运行安全应用, 采用精确率(Precision, P)、召回率(Recall, R)和 F1 分数作为关键指标, 并设定 0.5 的 IoU 阈值以契合实际运营需求。与传统检测模型不同, 多模态大模型直接输出边界框且不提供置信度分数, 故标准平均正确率指标不适用, 因此选用 F1 分数量化预测目标和真实目标的重合程度。

精确率(P)衡量模型目标误检程度, 精确率越高, 误检率越低:

$$P = T_p / (T_p + F_p) \quad (6)$$

式中: T_p 为真正例, 代表目标实际为真且被模型预测为真的样本; F_p 为假正例, 代表目标实际为假但被模型预测为真的样本。

召回率(R)衡量模型目标漏检程度, 召回率越高, 漏检率越低:

$$R = T_p / (T_p + F_N) \quad (7)$$

式中 F_N 为假负例, 代表目标实际为真但被模型预测为假的样本。

$F1$ 分数是精确率和召回率的调和平均数, 是一个综合指标, 用于平衡精确率和召回率:

$$F1 = 2PR / (P + R) \quad (8)$$

考虑模型计算量和计算效率, 训练服务器可实现秩为 8、16 和 32 的训练和测试。秩取 64 时显存占用过大, 本文服务器无法支撑训练。经全面测试, 本文方法在多种秩值配置下均取得优异结果, 准确率随秩值增大而提升。当秩值=32 时, 模型达到了最高性能指标: 精确率为 89.3%, 召回率为 74.5%, $F1$ 分数为 80.5%, 如表 1 所示。这表明在 LoRA 微调过程中增加秩值能显著提升检测精确度。

为进一步验证本文方法有效性, 测试不同大模型框架的风险目标识别精确度。本文选用现阶段主流的一些大模型框架进行微调测试, 包括阿里 Qwen2-VL 和智谱 AI。2 个网路微调训练配置与 DeepSeek-VL 相同, 均训练 50 轮, 在相同的测试集上进行测试, 且 LoRA 微调的秩均为 32。

表 2 为不同大模型的列车运行环境风险目标识别效果对比。

本文选用 DeepSeek-VL2-7B、DeepSeek-VL2-3B、阿里 Qwen2-VL-7B 和智谱 GLM-4V-9B 进行对比, 通过选用参数量接近的大模型, 可更好地体现不同模型架构对列车环境风险目标的辨识能力。同时 7B 模型更容易部署, 且计算速度适中, 便于后续列车边缘部署。测试不同参数量的 DeepSeek 模型及其他广泛使用的开源模型的识别效果, 所有模型均采用相同数据进行微调。最终在控制变量的情况下, 基于 DeepSeek-VL2-7B 的微调方法取得了最高的 $F1$ 分数, 比 Qwen2-VL-7B 和智谱 GLM 分别提高了 0.6 个百分点和 1.9 个百分点; 同时比网络结构相同, 参数量降低一半的 DeepSeek-VL2-3B 提高了 5.4 个百分点, 证明了 DeepSeek-VL2-7B 模型具有更强大的语义理解和风险识别能力。

表 1 实验结果(单位:%)
Table1 Experiment result(unit:%)

r	P	R	$F1$
8	78.6	71.3	74.7
16	87.4	73.6	79.9
32	89.3	74.5	80.5

表 2 对比实验结果

Table2 Comparative experimental results

model	$F1$ -score/%	memory/GB	time-consuming/ms
DeepSeek-VL2-7B	80.5	14	325
DeepSeek-VL2-3B	75.1	6	241
Qwen2-VL-7B	79.9	14	347
GLM-4V-9B	78.6	16	393

在计算效率方面, 参数量较小的 DeepSeek-VL2-3B 模型展现出显著优势, 其单张图像平均处理耗时仅为 241 ms, 这对于需要快速响应的列车安全预警系统是一个重要的基准性能。本文采用的 DeepSeek-VL2-7B 模型在精确度与速度之间取得了最佳平衡, 其计算耗时为 325 ms, 虽略低于 3B 模型, 但仍能满足地铁场景下对于“近实时”处理的需求。如, 在列车进站、通过固定监控点位等非极端瞬时决策场景中, 该速度足以对潜在风险目标进行有效识别与预警。作为对比, 参数量相近的 Qwen2-VL-7B 和智谱 GLM-4V-9B 模型耗时分别为 347 ms 和 393 ms, 表明本文采用的 DeepSeek-VL2-7B 模型在架构上具有更高的计算效率。

显存占用是决定算法能否在边缘计算设备上部署的关键因素。DeepSeek-VL2-3B 模型仅需 6 GB 显存, 这意味着它可以轻松部署在 NVIDIA GeForce RTX 3060/4060 等消费级显卡上, 极大地降低了硬件采购和部署成本。DeepSeek-VL2-7B 模型在微调与推理过程中的显存占用为 14 GB, 这要求使用 RTX 4090 或 NVIDIA Tesla T4、L4 等专业级边缘计算显卡。

综合上述应用能力分析, 本文提出的基于 DeepSeek-VL2-7B 的微调方法, 在取得最高识别精确度(80.5%)的同时, 保持了可接受的推理延迟和合理的显存需求。表明该模型具有较强的应用部署能力, 能满足列车风险辨

识任务需求, 保障列车安全运行。

本文对预测结果进行了可视化展示, 如图 6 所示。基于 DeepSeek-VL2-7B 的微调模型能有效识别列车场景中的各类障碍物, 包括行人、石块、列车和信号灯小目标等。同时, 本文还测试了夜晚低照度、雨天和雾天等多种恶劣环境下识别效果, 基于 DeepSeek-VL2 微调的列车运行环境风险目标识别方法均取得了极佳的检测效果, 证明了该方法可满足多种环境下列车风险目标识别需求, 有力保障列车运行安全。

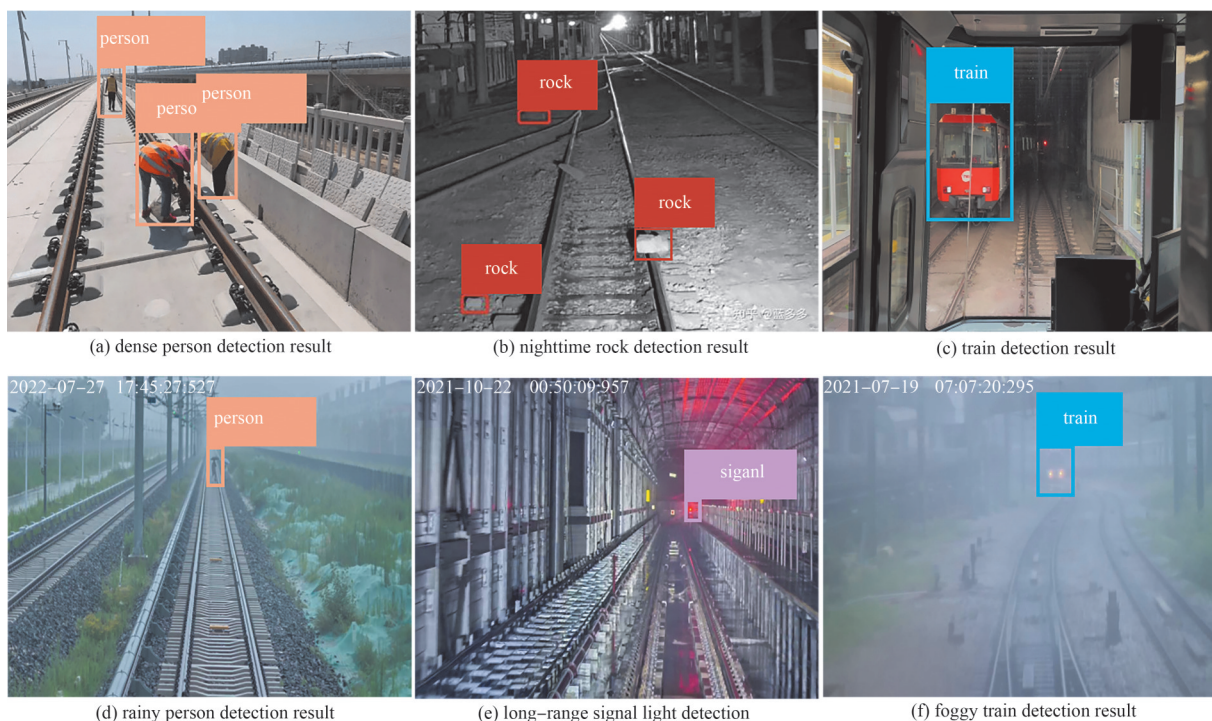


Fig.6 Experiment results
图6 实验结果

5 结论

本文提出基于 DeepSeek-VL2-LoRA 微调方法的列车运行环境风险目标识别方法, 通过微调多模态大模型参数, 使大模型学习列车领域相关知识, 从而实现列车风险目标的辨识。实验结果表明, 通过 LoRA 微调通用大模型的方法, 无需修改大模型架构即可有效执行列车领域识别任务。微调后模型表现出强大的泛化能力, 在识别列车、信号灯和行人等任务中取得了极高的精确度。这些结果映证了多模态大模型在列车运行监控与智能化领域中的潜力, 为开发融合视觉感知与语言提示词的智能列车运行监控方案提供了技术基础。在未来, 计划开发针对列车环境的大模型实时定位方法, 利用大模型的长时间记忆能力理解列车运行场景, 从而实现长距离下高精度定位。此类改进将拓展大模型在列车运营中的实际应用价值, 在保持自然语言交互优势的同时实现更全面的场景理解。

参考文献:

- [1] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing System. Red Hook, NY, USA: NIPS, 2017:6000-6010.
- [2] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149. DOI:10.1109/TPAMI.2016.2577031.
- [3] CAO Zhiwei, QIN Yong, JIA Limin, et al. Railway intrusion detection based on machine vision: a survey, challenges, and perspectives[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(7):6427-6448. DOI:10.1109/TITS2024.3412170.
- [4] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards realtime object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149. DOI:10.1109/TPAMI.2016.2577031.
- [5] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: attention-centric real-time object detectors[J/OL]. arXiv:2502.12524, 2025. DOI:10.48550/arXiv.2502.12524.

- [6] HUANG Xiaohua,ZHU Jiaohao,HUO Ying. SSA-YOLO:an improved YOLO for hot-rolled strip steel surface defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2024,73:5040017.1-17. DOI:10.1109/TIM.2024.3488136.
- [7] JRONDI Z,MOUSSAID A,HADI M Y. Exploring end-to-end object detection with transformers versus YOLOv8 for enhanced citrus fruit detection within trees[J]. Systems and Soft Computing, 2024,6(1):200103. DOI:10.1016/j.sasc.2024.200103.
- [8] YANG An,YANG Baosong,ZHANG Beechen,et al. Qwen2.5 technical report[J/OL]. arXiv:2412.15115, 2024. DOI:10.48550/arXiv.2412.15115.
- [9] ACHIAM J,ADLER S,AGARWAL S,et al. Gpt-4 technical report[J/OL]. arXiv: 2303.08774, 2023. DOI: 10.48550/arXiv.2303.08774.
- [10] GUO Daya,YANG Dejian,ZHANG Haowei,et al. Deepseek-r1:incentivizing reasoning capability in llms via reinforcement learning[J/OL] ArXiv:2501.12948, 2025. DOI:10.48550/arXiv. 2303.08774.
- [11] ZHU Ruoxi,TU Zhengzhong,LIU Jiaming,et al. Mwformer:multi-weather image restoration using degradation-aware transformers[J]. IEEE Transactions on Image Processing, 2024(33):6790-6805. DOI:10.1109/TIP.2024.350185.
- [12] BAI Jiesong, YIN Yuhao, HE Qiyuan, et al. Retinexmamba: Retinex-based mamba for low-light image enhancement[C]// International Conference on Neural Information Processing. Singapore:Springer Nature Singapore, 2024:427-442.
- [13] 梁雅楠,刘昌瑞,石雪涛. 面向列车自主运行的边缘智能系统[J]. 太赫兹科学与电子信息学报, 2024,22(8):893-900. (LIANG Yanan,LIU Changrui,SHI Xuetao. Edge intelligent system for autonomous train operation[J]. Journal of Terahertz Science and Electronic Information Technology, 2024,22(8):893-900.) DOI:10.11805/TKYDA2023071.
- [14] 徐建喜,魏思雨,李宗平. 基于 5G 的列车云边缘协同计算设计与优化[J]. 太赫兹科学与电子信息学报, 2024,22(11):1199-1208. (XU Jianxi,WEI Siyu,LI Zongping. 5G-based design and optimization of cloud-edge-train collaborative computing[J]. Journal of Terahertz Science and Electronic Information Technology, 2024,22(11):1199-1208.) DOI:10.11805/TKYDA2023049.
- [15] 钱铭,洪涛,张更新. 基于低轨卫星物联网的差分 LoRa 体制性能分析[J]. 太赫兹科学与电子信息学报, 2025,23(5):434-445. (QIAN Ming,HONG Tao,ZHANG Gengxin. Performance analysis of differential LoRa scheme for LEO satellite-based Internet of Things[J]. Journal of Terahertz Science and Electronic Information Technology, 2025,23(5):434-445.) DOI:10.11805/TKYDA 2024170.
- [16] ZHU Ruidong,JIANG Ziyue,ZHANG Zhi,et al. Cannikin:no lagger of SLO in concurrent multiple LoRA LLM serving[J]. IEEE Transactions on Parallel and Distributed Systems, 2025,36(9):1972-1984. DOI:10.1109/TPDS.2025.3590014.
- [17] 张程,曹京旭,吕劲昕,等. 自然语言视角下基于凸包的知识蒸馏可解释性研究[J]. 计算机科学与探索, 2025,19(8):2174-2187. (ZHANG Cheng,CAO Jingxu,LYU Jinxin,et al. Study on interpretability of knowledge distillation based on convex hulls from perspective of natural language[J]. Journal of Frontiers of Computer Science and Technology, 2025,19(8):2174-2187.) DOI: 10.3778/j.issn.1673-9418.2407077.
- [18] 陈志翔,谢志鹏. 基于大语言模型的事件因果数据增广方法[J/OL]. 计算机科学, <https://link.cnki.net/urlid/50.1075.TP.20250910.1556.004>.
- [19] 王阳,肖顺平. 基于注意力机制的三维点云目标识别[J]. 太赫兹科学与电子信息学报, 2024,22(7):730-740. (WANG Yang, XIAO Shunping. Attention mechanism based 3D point cloud target recognition[J]. Journal of Terahertz Science and Electronic Information Technology, 2024,22(7):730-740.) DOI:10.11805/TKYDA2022101.
- [20] HAN G X,LIM S N. Few-shot object detection with foundation models[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle,Washington,USA:IEEE, 2024:28608-28618. DOI:10.1109/CVPR52733.2024.02703.
- [21] XING Zheng, ZHAO Weibing. Block-diagonal guided DBSCAN clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2024,36(11):5709-5722. DOI:10.1109/TKDE.2024.3401075.
- [22] LIN T Y,MAIRE M,BELONGIE S,et al. Microsoft COCO:common objects in context[J]. Lecture Notes in Computer Science, 2014,8693(1):740-755. DOI:10.1007/978-3-319-10602-1_48.
- [23] MUSLEH A,ALRYALAT S A,QASEM A. Image annotation software for artificial intelligence applications[J]. High Yield Medical Reviews, 2023,1(2):1-5. DOI:10.59707/hymrXHMX8234.
- [24] ZHENG Yaowei,ZHANG Richong,ZHANG Junhao,et al. Llamafactory:unified efficient fine-tuning of 100+ language models[J/OL]. arXiv:2403.13372, 2024. DOI:10.48550/arXiv.2403.13372.

作者简介:

崔广炎(1994-), 男, 博士, 工程师, 主要研究方向为智能列车运行系统 .email:subwaybj2025@126.com.

李宇杰(1983-), 男, 博士, 正高级工程师, 主要研究方向为智能列车运行系统.

李熙(1981-), 男, 博士, 正高级工程师, 主要研究方向为智能列车运行系统.

沈忱(1973-), 男, 学士, 高级工程师, 主要研究方向为智能列车运行系统.

李琨(1996-), 男, 硕士, 工程师, 主要研究方向为智能列车运行系统.

刘可(1994-), 男, 学士, 工程师, 主要研究方向为智能列车运行系统.

(上接第 1249 页)

- [17] 艾渤,何丹萍,钟章队. 高性能射线跟踪技术在智能交通领域的应用[J]. 太赫兹科学与电子信息学报, 2023,21(4):530-538. (AI Bo, HE Danping, ZHONG Zhangdui. Application of high-performance ray tracing technology in intelligent transportation[J]. Journal of Terahertz Science and Electronic Information Technology, 2023,21(4):530-538.) DOI:10.11805/TKYDA2023031.

作者简介:

谢克绪(1981-), 男, 硕士, 高级工程师, 主要研究方向为铁路通信 .email:11074726@chnenergy.com.cn.

孙斌(1981-), 男, 硕士, 助理研究员, 主要研究方向为宽带移动通信与专用移动通信.

王丽鑫(2002-), 女, 在读硕士研究生, 主要研究方向为宽带移动通信与专用移动通信.

冯源(1998-), 男, 在读博士研究生, 主要研究方向为算力网络、无线通信安全.

刘腾(2001-), 男, 在读博士研究生, 主要研究方向为人工智能安全.

丁建文(1980-), 男, 博士, 研究员, 主要研究方向为宽带移动通信与专用移动通信.