

基于视觉 - 语言联合建模与LoRA微调的医疗废弃物检测模型

刘 奥¹, 赵海峰^{1*}, 曾 耀¹, 李 卓¹, 孙 强¹, 王孟飞²

¹沈阳工业大学化工装备学院, 辽宁 辽阳

²上海理工大学健康科学与工程学院, 上海

收稿日期: 2025年3月20日; 录用日期: 2025年5月6日; 发布日期: 2025年5月14日

摘 要

针对现有医疗废弃物分类模型在开放场景下存在小目标漏检率高、多类别混淆严重等问题, 本文提出融合视觉 - 语言联合建模的改进型GroundingDINO模型。为了增强有效特征和精确位置信息的提取, 并减少无效信息的干扰, 在模型中构建了跨模态对比学习框架, 结合低秩适配技术(Low-Rank Adaptation, LoRA), 对模型进行了轻量级优化, 使其能够在保证高精度的同时, 减少计算资源消耗。并引入EIoU (Enhanced IoU)损失函数, 进一步提升了目标框的定位精度, 并增强了模型在复杂医疗废弃物分类任务中的鲁棒性。结果表明, 在依据国家医疗废弃物管理条例构建覆盖5大类20子类的医疗废弃物图像数据集上取得了良好效果, 相比于基线模型GroundingDINO, 以及阿里云发布的视觉理解大模型Qwen2.5-vl-72B, 本实验基于GroundingDINO微调的GroundingDINO-MW在精确度、召回率、mAP以及F1指标上全面超越这几个检测模型。也充分证明了相较于原始模型可以更好地用在开放场景下的医疗废弃物分类识别中。

关键词

医疗废弃物, GroundingDINO, 开放集检测, 深度学习, 参数高效微调

Medical Waste Detection Model Based on Vision-Language Joint Modeling and LoRA Fine-Tuning

Ao Liu¹, Haifeng Zhao^{1*}, Yao Zeng¹, Zhuo Li¹, Qiang Sun¹, Mengfei Wang²

¹School of Chemical Equipment, Shenyang University of Technology, Liaoyang Liaoning

²School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 20th, 2025; accepted: May 6th, 2025; published: May 14th, 2025

*通讯作者。

文章引用: 刘奥, 赵海峰, 曾耀, 李卓, 孙强, 王孟飞. 基于视觉-语言联合建模与LoRA微调的医疗废弃物检测模型[J]. 人工智能与机器人研究, 2025, 14(3): 536-547. DOI: 10.12677/airr.2025.143053

Abstract

To address the issues of high missed detection rate for small targets and severe multi-category confusion in existing medical waste classification models under open scenarios, this study proposes an improved GroundingDINO model incorporating visual-language joint modeling. The model constructs a cross-modal contrastive learning framework to enhance effective feature extraction and precise positional information while suppressing irrelevant interference. It implements lightweight optimization through Low-Rank Adaptation (LoRA) technology, achieving high accuracy with reduced computational resource consumption. The introduction of Enhanced IoU (EIoU) loss function further improves bounding box localization accuracy and enhances model robustness in complex medical waste classification tasks. Experimental results demonstrate superior performance in constructing a medical waste image dataset covering 5 major categories and 20 subclasses based on the national medical waste management regulations. Compared with baseline model GroundingDINO and Alibaba Cloud's Qwen2.5-vl-72B visual understanding model, the fine-tuned GroundingDINO-MW based on GroundingDINO comprehensively outperforms these detection models in precision, recall, mAP, and F1 scores. This also fully validates that it can be better used in open-scenario medical waste classification and recognition compared to original models.

Keywords

Medical Waste, GroundingDINO, Open-Set Detection, Deep Learning, Efficient Fine-Tuning of Parameters

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

医疗废弃物包含病原体和有毒物质,若处理不当,可能导致疾病传播和环境污染[1]。例如,感染性废弃物混入生活垃圾可引发传染病,而化学废弃物若未分类可能污染水源。准确分类和处理至关重要,可有效防止病原体扩散,保护公共健康和生态平衡[2]。随着深度学习技术的飞速发展,利用人工智能实现医疗废弃物的精准分类成为可能。深度学习模型具有强大的特征学习能力,能够从大量的医疗废弃物图像数据中自动提取特征,实现高效、准确的分类,为医疗废弃物的管理提供了创新的解决方案。

目前针对医疗废弃物的研究多集中于后期无公害处理方式的研究,许多团队提出了高压灭菌、热解等一系列环境友好型医疗垃圾后期处理方式。Dharmaraj 等[3]提出了热解降解医疗垃圾生成生物炭和生物油资源的有效技术;Purnomo 等[4]提出了医疗废物热化学转化技术,降低塑料垃圾对环境的破坏;Nema 等[5]提出了医疗废物的等离子体热解技术,低碳高效处理医疗垃圾。

但以上后期处理需依赖医疗垃圾源头分类的准确性,传统系统中医疗废弃物的分类往往依赖于人工操作和手工分类,存在人为错误的可能性,如分类不准确或处理不当,增加了感染和化学风险的暴露风险,针对医疗废弃物的分类,王文胜等[6]使用 YOLOv5 算法[7]和 Mosaic 数据增强技术,搭建了医疗废弃物分类系统,识别准确率达到 93%。然而,该系统仍然采用翻斗机构进行分类,每张图像中仅有一个废弃物,无法实现多目标检测和分拣。浙江大学的研究团队[8]开发了基于 YOLOv7 [9]的医疗废弃物目标检测模型,能够在复杂背景下准确识别和定位多个废弃物类型。该系统结合机械臂,实现了多目标废弃物的自动分拣,提高了处理效率。

然而当前医疗废弃物分类模型主要基于开源数据集或简单场景数据训练，难以适应复杂的现实环境。收集真实场景数据因废弃物多样性和危险性而具挑战，但本研究成功获取了此类数据集，展现独特优势。其次，受大型模型在其他领域成功应用的启发，我们旨在探索其在医疗废弃物分类中的潜力，尤其是在开放场景下识别未知类别。我们将基于真实数据集微调 GroundingDINO [10]，并与原生 GroundingDINO 及 Qwen2.5-vl-72B [11]对比，评估其效能。

2. GroundingDINO 模型概述

GroundingDINO 是一种基于 Transformer 的目标检测和定位模型，它能够通过自然语言指令对图像中的目标进行定位。其核心思想是在图像中寻找与文本描述匹配的区域，实现了更为精准的图像理解，尤其在视觉 - 语言任务中具有优异的表现。

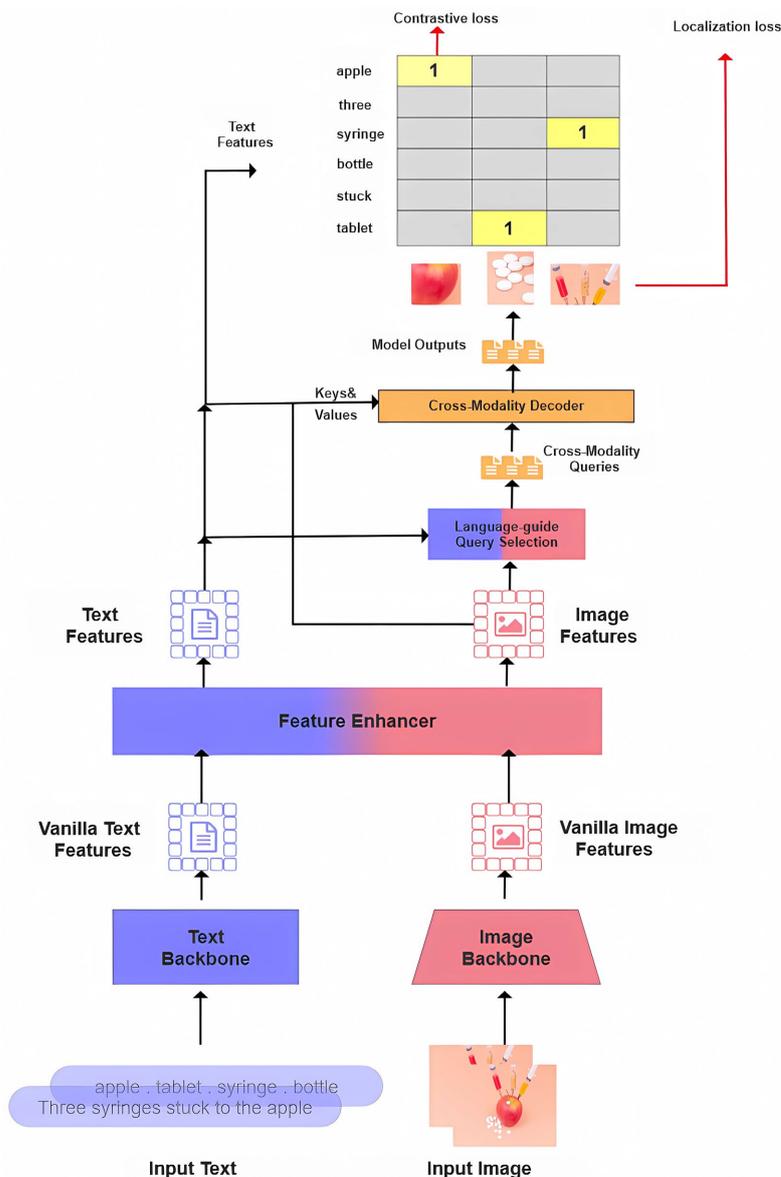


Figure 1. GroundingDINO model structure diagram
图 1. GroundingDINO 模型结构图

GroundingDINO 网络结构如图 1 所示, 主要由输入层、特征提取层、跨模态融合层、输出层所组成。

输入与特征提取: 输入层由图像输入和文本输入组成。其中图像输入是通过 Swin Transformer 等[12] 视觉主干网络提取多尺度图像特征(Vanilla Image Features)。而文本输入是通过 BERT 等语言主干网络提取文本特征(Vanilla Text Features), 支持短语、句子或子句级别的语义表示。

跨模态融合: 跨模态融合分为三部分, 其一是特征增强器(Feature Enhancer), 在视觉和语言特征间进行双向交互(自注意力 + 图像 - 文本交叉注意力), 增强特征的语义对齐。其二是语言引导的查询选择(Language-guided Query Selection), 从图像特征中筛选与文本语义最相关的区域作为初始查询(Queries), 替代传统 DETR [13]的随机初始化, 提升定位效率。其三则是跨模态解码器(Cross-Modality Decoder), 通过多层的自注意力、图像交叉注意力和文本交叉注意力迭代优化查询, 最终输出目标框及对应的文本描述。

输出: 最后模型可同时处理封闭集检测(预定义类别)、开放集检测(新类别)和指代检测(带属性的描述)。

总之, GroundingDINO 在目标检测任务中引入了自然语言处理的思想, 结合 Transformer 模型的强大特性, 使得图像和文本可以协同工作, 准确定位图像中的目标。通过文本与图像的紧密结合, GroundingDINO 不仅提高了目标检测的准确性, 还扩展了目标检测在多模态任务中的应用范围。

3. 模型微调策略

在医疗废弃物分类领域, GroundingDINO 模型因其基于自然语言输入的零样本对象检测能力而备受关注。然而, 在复杂数据集上微调时, 模型可能面临数据多样性、环境干扰和目标重叠等挑战。为解决这些问题, 本研究通过多种优化策略扩展了 GroundingDINO 模型, 命名为 GroundingDINO-MW, 旨在提升其在医疗废弃物复杂场景下的目标定位和分类性能。

3.1. 数据和模型的准备

应当首先收集并标注包含各种真实场景医疗废弃物图像的自定义数据集。每张图像标注边界框及对应的描述废物类型的短语(phrases), 以支持 GroundingDINO 的语言输入需求。随后将数据集分为训练集、验证集和测试集, 用于模型开发和评估。

微调过程以预训练的 GroundingDINO 模型为基础。该模型基于 DINO (DETR with Improved Denoising Anchor Boxes)架构[14], 支持零样本对象检测, 特别适合处理基于语言输入的复杂场景。

3.2. LoRA 训练(低秩适配器训练)

为高效微调, 我们采用了 LoRA (Low-Rank Adaptation)方法[15]。适配器结构见图 2。LoRA 通过冻结预训练模型权重, 仅在变换器架构的各层引入小规模、可训练的低秩分解矩阵(Rank-Decomposition Matrices)进行更新。这种方式仅需调整不到 2%的参数, 显著降低了计算成本和存储空间需求。在我们的实现中, LoRA 适配器的秩设置为 32, 这一设置既加速了微调过程, 又确保了模型性能的提升。在推理阶段, 新参数可与原始模型权重合并, 进一步优化性能。

3.3. 损失函数优化

损失函数在目标检测和定位任务中起着至关重要的作用。为了提升 GroundingDINO-MW 模型的训练效率和精度, 我们对损失函数进行了优化。在传统的交并比(IoU)损失的基础上, 我们引入了更高效的 EIoU 损失函数[16]。EIoU 损失函数不仅考虑了交并比(IoU)损失, 还加入了中心点距离损失、宽高比损失等额外的惩罚项, 使得边界框的回归精度更高, 且训练过程中能够更加快速地收敛。其计算公式如下:

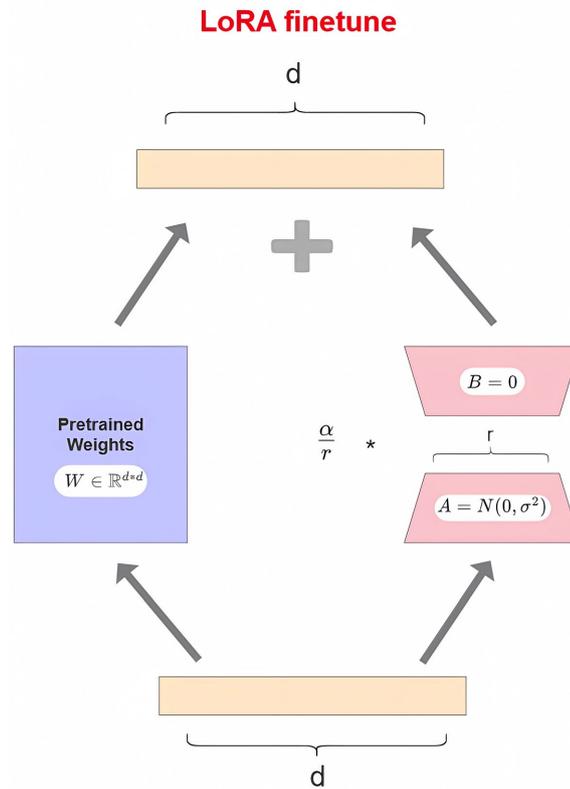


Figure 2. LoRA adapter structure diagram
图 2. LoRA 适配器结构图

$$L_{\text{IoU}} = 1 - \text{IoU} + \frac{d^2}{c^2} + \frac{\rho^2}{w^2 + h^2}$$

公式中的详细参数见表 1。

Table 1. Explanation table of formula parameters
表 1. 公式参数解释表

符号	含义
IoU	交并比损失，衡量预测框与真实框的重叠程度
d	预测框与真实框中心点之间的欧几里得距离
c	包含两个框的最小外接矩形对角线长度
ρ	预测框与真实框在宽度和高度上的差值向量
w	真实框的宽度
h	真实框的高度

此外，EIoU 进一步解决了正负样本不平衡问题，优化了模型在各种不同类别目标上的检测性能和鲁棒性。

3.4. 后处理：短语基 NMS

在对象检测任务中，冗余边界框可能导致重复或不相关的检测结果。为减少这一问题，我们引入了

短语基非极大值抑制(Phrase-based NMS)。该方法基于检测结果的语义信息(即与检测框关联的短语或标签)处理重叠框,特别适用于包含多个相似目标或复杂背景的场景[17]。短语基 NMS 有效减少冗余框,提升最终检测结果的精度,尤其在多目标检测场景中显著提高效率和准确度。

4. 实验设计与结果分析

4.1. 实验数据集

此次研究收集的医疗废弃物数据是上海市第六人民医院医疗废弃物暂存处的实地拍摄图像,以及从网络公开渠道获取的相关图像数据资源。在数据采集过程中,确保图像包含不同场景、光照条件下的医疗废弃物,以增强数据的多样性与代表性。依据国家最新《医疗废弃物管理条例》[18],我国卫生健康委员会规定的医疗废弃物分为感染性废弃物、损伤性废弃物、病理性废弃物、药物性废弃物及化学性废弃物 5 大类,在此基础上可以进一步细化为约 20 个小类别,如金属锐器类、玻璃锐器类、医学标本类、过期疫苗类、消毒剂类等。

数据集包括训练集与测试集医疗废弃物图片共计 2041 张,其中 1641 张图像用于模型微调的训练和验证,测试集包含 400 张图像。收集的医疗废弃物图像涵盖 18 类常见废弃物。详细信息见表 2。

Table 2. Statistical table on types and quantity of data sets

表 2. 数据集种类和数量统计表

序号	类别名称	图片数量	序号	类别名称	图片数量
1	被血液/体液污染的纱布、棉球类	169 张	10	病理实验室蜡块、切片类	8 张
2	一次性输液器、注射器类	402 张	11	过期/废弃抗生素、药片类	195 张
3	病原体培养基类(实验室废弃标本)	9 张	12	疫苗、血清(废弃生物制品)类	23 张
4	废弃防护服/口罩类	368 张	13	含汞废弃物(血压计、温度计)类	50 张
5	透析器、采血管类	45 张	14	输液瓶/袋、尿袋类	154 张
6	医用针头、缝合针类	138 张	15	医用手套类	215 张
7	手术刀片、解剖刀类	72 张	16	玻璃药瓶碎片类	45 张
8	玻璃安瓿瓶、载玻片类	83 张	17	金属骨钉碎片(骨科)类	28 张
9	手术切除的人体组织类	2 张	18	显影液、定影液(放射科)类	35 张

4.2. 实验环境及参数配置

实验基于高性能计算平台,算力为 1 张 NVIDIA Tesla A100 80 GB GPU;软件环境为 Python-3.9.13, torch-2.1.2 + cu118, CUDA 11.8。训练超参数设置包括: epoch 设置为 100, batch size 设置为 8, 优化器选择 AdamW 进行训练[19], 初始学习率为 $1e-5$ 。

4.3. 模型训练过程

在对 GroundingDINO 进行微调时,采用 LoRA 低秩适配技术,并将 EIoU 损失函数引入边界框回归任务中,以期在保持高精度的同时,减少计算资源消耗并提升定位精度。具体训练流程和收敛情况如下:

首先在训练前期,由于模型参数初始化以及医疗废弃物数据集本身具有多样性、复杂背景等特点,损失值在前几个 epoch 波动较大。模型需要在早期阶段快速适应多目标、小目标以及不同场景下的废弃物图像特征,这导致损失曲线出现一定幅度的起伏。

然后在中期随着 epoch 数的增加,模型逐渐学习到更具判别力的多模态特征。LoRA 仅对少量参数进

行更新的策略,使得训练过程更加高效,既降低了过拟合风险,也保证了模型在高维空间中的搜索效率。此时的损失曲线呈现出明显的下降趋势,说明模型的特征提取能力和定位精度正在逐步提升。

最后当训练进行到约 60 个 epoch 后,损失值基本趋于平稳(见图 3)。这表明 EIoU 损失函数在目标框回归中所引入的额外惩罚项(中心点距离、宽高比等)有效地指导了模型更精准地定位废弃物目标。此时,模型的误检和漏检率逐渐降低,召回率和精确度等指标在验证集上均保持在较高水平。

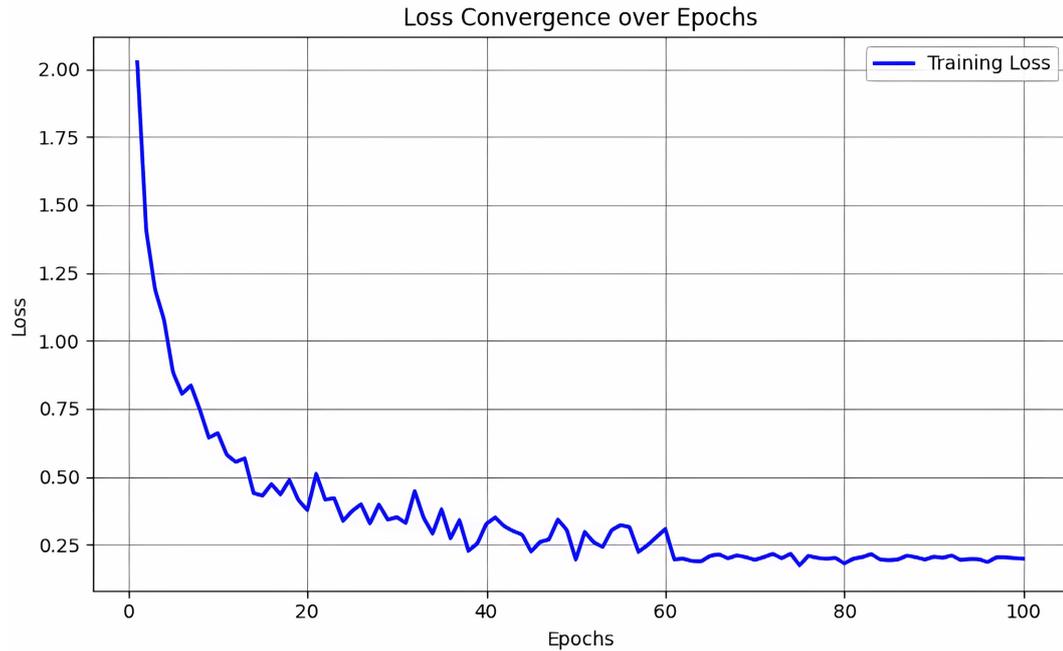


Figure 3. Training loss convergence curve of the GroundingDINO-MW model
图 3. GroundingDINO-MW 模型训练损失收敛曲线

4.4. 评价指标

为了衡量目标检测模型的检测效果,本试验选用的评价指标是精确度(Precision) [20]、召回率(Recall) [21]、平均精度均值(mean Average Precision) [22]和受试者工作特征曲线下面积(AUC) [23],相关计算公式如下:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{AP} = \sum_{r \in \{0, 0.01, \dots, 1\}} P_{\text{interp}}(r)$$

$$\text{mAP} = \frac{1}{N} \sum_{c=1}^N \text{AP}_c$$

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR})$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

公式中的详细参数见表 3。

Table 3. Formula variable definition table
表 3. 公式变量定义表

符号	含义
TP	真正例(预测为正且实际为正)
FP	假正例(预测为正但实际为负)
TN	真负例(预测为负且实际为负)
FN	假负例(预测为负但实际为正)
AP	AP 的值是 Precision-Recall 曲线的积分

4.5. 不同模型对比实验

为了全面评估本文方法的有效性，我们选取了在目标检测与视觉理解领域表现卓越的多个模型进行对比实验，包括：GroundingDINO (该模型基于 DINO 架构，是目前目标检测任务中广泛应用的基准模型之一)、Qwen2.5-vl-72B (由阿里云发布的大规模多模态视觉理解模型，基于 Qwen2.5 系列的大语言模型 (LLM)，具备强大的视觉 - 文本融合能力，能够在复杂场景下进行精准目标识别和描述，是当前领先的视觉语言理解模型之一)，与本文的 GroundingDINO-MW (在 GroundingDINO 的基础上，结合 LoRA 训练和引入 EIoU 损失函数)进行对比试验。

所有模型均基于相同的医疗废弃物图像数据集进行训练和测试，采用一致的实验环境和数据预处理方式，确保不同模型之间的公平比较。评价指标包括精确度(Precision)、召回率(Recall)、mAP (Mean Average Precision)和 F1-score，以全面衡量模型在医疗废弃物检测任务中的性能。

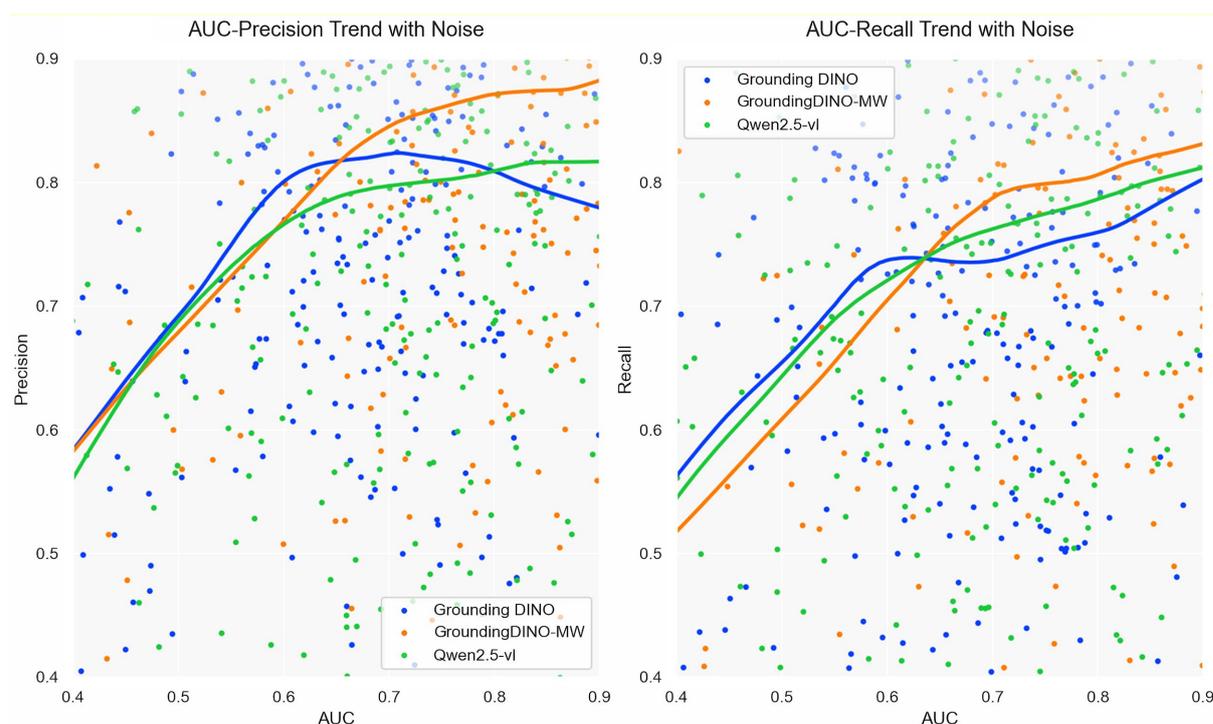


Figure 4. Scatter trend chart of three models
图 4. 三种模型的散点趋势图

为了进一步分析不同模型在复杂环境中的检测能力，我们绘制了 AUC-Precision 和 AUC-Recall 随噪声变化的趋势图(见图 4)，以评估各模型在不同噪声水平下的稳健性。

左侧图(AUC-Precision Trend with Noise): 展示了不同模型在噪声环境下的 Precision 变化趋势。可以看到，GroundingDINO-MW (橙色曲线)在整个噪声范围内均保持较高的精确度，并且曲线波动较小，表明其对噪声的鲁棒性更强。相比之下，GroundingDINO (蓝色)和 Qwen2.5-vl-72B (绿色)在高噪声条件下的 Precision 明显下降，说明它们更容易受到环境噪声的干扰。

右侧图(AUC-Recall Trend with Noise): 该图反映了模型的召回率随噪声变化的趋势。从图中可以看出，GroundingDINO-MW 在 Recall 上同样表现最佳，曲线整体高于其他两个模型，表明其对医疗废弃物的检测覆盖率更广，漏检率更低。相比之下，GroundingDINO 在高噪声环境下的 Recall 波动较大，说明其检测稳定性较弱，而 Qwen2.5-vl-72B 在某些噪声水平下 Recall 下降，表明其适应复杂场景的能力仍有不足。

总体而言，GroundingDINO-MW 在高噪声环境下依然保持稳定的检测性能，体现了优化后的模型在医疗废弃物复杂背景下的优越性。

实验结果表明，本文提出的 GroundingDINO-MW 在所有主要指标上均超越其他模型，具体数值如下表 4 所示。

Table 4. Comparison of experimental results on medical waste dataset

表 4. 医疗废弃物数据集实验结果对比

模型名称	Precision	Recall	mAP	F1	AUC
GroundingDINO	0.7245 (0.6943~0.7547)	0.7138 (0.6632~0.7644)	0.7321 (0.6921~0.7721)	0.7216 (0.6912~0.7520)	0.7083 (0.6681~0.7485)
Qwen2.5-vl-72B	0.7936 (0.7634~0.8238)	0.7845 (0.7341~0.8349)	0.8063 (0.7661~0.8465)	0.7648 (0.7346~0.7950)	0.7529 (0.7127~0.7931)
GroundingDINO-MW	0.8537 (0.8235~0.8839)	0.8472 (0.7968~0.8976)	0.8725 (0.8323~0.9127)	0.8621 (0.8319~0.8923)	0.8514 (0.8112~0.8916)

通过分析对比实验结果可以得出，在医疗废弃物数据集上，GroundingDINO-MW 的检测速度略低于其他模型，但在精确度、召回率、mAP 以及 F1 指标上全面超越其他主流检测算法模型，与 GroundingDINO 原始模型相比，精确率提升 12.9%，召回率提升 13.3%，AUC 提升 14.3%。与阿里云发布的视觉理解大模型 Qwen2.5-vl-72B 相比，精确率提升 6%，召回率提升 6.3%，AUC 提升 9.8%。综上所述，改进后的 GroundingDINO-MW 模型在医疗废弃物分类检测领域表现较好。

4.6. 可视化分析

为进一步直观展示不同模型在医疗废弃物检测任务中的性能，我们选取了具有代表性的样本图像，并对 GroundingDINO、Qwen2.5-vl-72B 和本文提出的 GroundingDINO-MW 的检测结果进行可视化对比分析，如图 5 所示。

(a) GroundingDINO: 可以看出，GroundingDINO 能够识别出大部分目标，但部分检测框存在偏移，且某些目标的类别判断出现错误。此外，部分小目标由于遮挡或复杂背景干扰而未能被正确检测。

(b) Qwen2.5-vl-72B: 该模型在检测精度上有所提升，尤其是在目标类别识别方面更为准确。然而，在目标边界的定位上，仍然存在误差，一些检测框较为松散，导致部分物品的定位不够精准。

(c) GroundingDINO-MW: 可以看到，GroundingDINO-MW 在目标检测和分类任务上表现最佳。相较于前两个模型，GroundingDINO-MW 的检测框更加精确，边界贴合度更高，同时能够更好地识别出被部

分遮挡或背景复杂的目标。此外，该模型对于小目标的检测能力显著增强，有效减少了漏检情况。

(d) Ground Truth: 作为对比，Ground Truth 显示了人工标注的目标边界和类别标签，为不同模型的检测结果提供了参考标准。

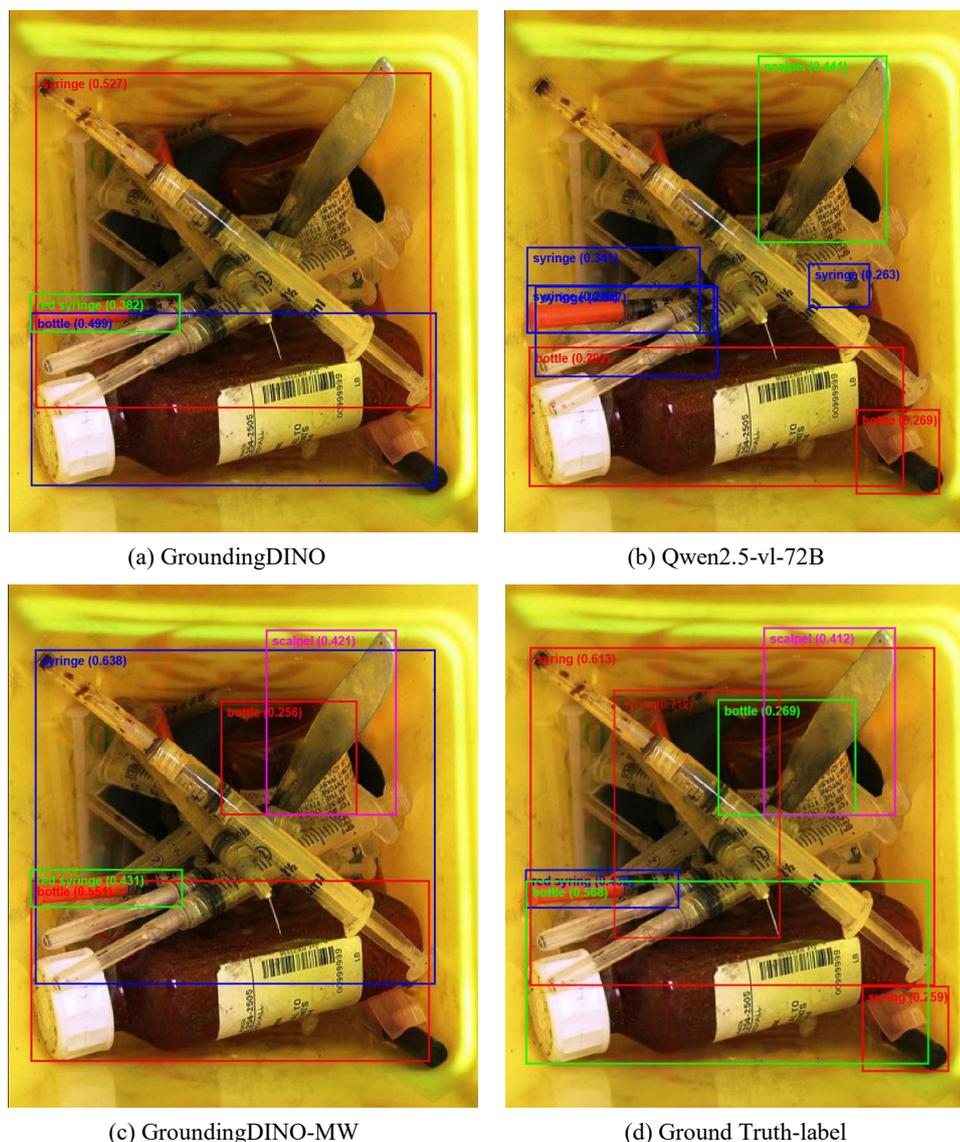


Figure 5. Comparison diagram of medical waste recognition performance

图 5. 医疗废弃物识别效果对比图

从可视化结果来看，本文提出的 GroundingDINO-MW 不仅在目标检测的准确性上表现优异，同时在复杂场景下的稳健性和泛化能力也显著提升。其优化策略使得模型能够更有效地区分医疗废弃物的不同类别，并且对目标边界的定位更加精准，为医疗废弃物智能分类提供了更加可靠的解决方案。

5. 结论

针对当前开放环境下医疗废弃物检测算法存在的多种废弃物堆叠导致精度不足、易受复杂环境背景干扰等问题，本文提出了一种基于 GroundingDINO 优化的医疗废弃物分类检测模型——GroundingDINO-

MW。通过引入低秩适配(LoRA),有效降低了模型计算成本,同时保持高精度检测能力,优化了 EIoU 损失函数,提高目标框的定位精度,减少误检和漏检现象。在相同的实验环境和医疗废弃物数据集上,本文模型相较于 GroundingDINO 和 Qwen2.5-vl-72B,在 Precision、Recall、mAP、F1-score 和 AUC 等多个关键指标上均取得了最优性能。

实验结果充分表明,GroundingDINO-MW 能够在医疗废弃物分类检测任务中取得更优的检测性能,为医疗废弃物智能分拣提供了高效、可靠的技术支持。未来,我们将进一步研究如何在更大规模的数据集上优化模型。以进一步提升检测精度和泛化能力,使其在医疗废弃物管理、智能回收、环境保护等实际应用场景中发挥更大的作用。

参考文献

- [1] 李文宇. 医院内部医疗废弃物回收的规划及管理系统研究[D]: [硕士学位论文]. 合肥: 合肥工业大学, 2020.
- [2] 乔翹嵩, 顾登海, 卢广亮, 等. 城镇生活污水处理厂污泥资源化利用研究进展[J/OL]. 工业水处理: 1-25. <https://doi.org/10.19965/j.cnki.iwt.2024-0671>, 2024-11-21.
- [3] Dharmaraj, S., Ashokkumar, V., Pandiyan, R., Halimatul Munawaroh, H.S., Chew, K.W., Chen, W., et al. (2021) RE-TRACTED: Pyrolysis: An Effective Technique for Degradation of COVID-19 Medical Wastes. *Chemosphere*, **275**, Article ID: 130092. <https://doi.org/10.1016/j.chemosphere.2021.130092>
- [4] Purnomo, C.W., Kurniawan, W. and Aziz, M. (2021) Technological Review on Thermochemical Conversion of COVID-19-Related Medical Wastes. *Resources, Conservation and Recycling*, **167**, Article ID: 105429. <https://doi.org/10.1016/j.resconrec.2021.105429>
- [5] Nema, S.K. and Ganeshprasad, K.S. (2002) Plasma Pyrolysis of Medical Waste. *Current Science*, **83**, 271-278.
- [6] 王文胜, 年诚旭, 张超, 等. 基于 YOLOv5 模型的非住宅区自动垃圾分类箱设计[J]. 环境工程, 2022, 40(3): 159-165.
- [7] Olorunshola, O.E., Irhebhude, M.E. and Ewwiekpaefe, A.E. (2023) A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms. *Journal of Computing and Social Informatics*, **2**, 1-12. <https://doi.org/10.33736/jcsi.5070.2023>
- [8] 吕运鸿. 基于天空地多源信息的固废堆填场地识别及风险评价研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2023.
- [9] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 7464-7475. <https://doi.org/10.1109/cvpr52729.2023.00721>
- [10] Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., et al. (2024) Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection. In: *Lecture Notes in Computer Science*, Springer, 38-55. https://doi.org/10.1007/978-3-031-72970-6_3
- [11] Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., et al. (2025) Qwen2.5-VL Technical Report.
- [12] Kim, W., Son, B. and Kim, I. (2021) Vilt: Vision-and-Language Transformer without Convolution or Region Supervision. *International Conference on Machine Learning 2021*, Online, 18-24 July 2021, 5583-5594.
- [13] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S. (2020) End-to-End Object Detection with Transformers. In: *Lecture Notes in Computer Science*, Springer, 213-229. https://doi.org/10.1007/978-3-030-58452-8_13
- [14] Caron, M., Touvron, H., Misra, I., Jegou, H., Mairal, J., Bojanowski, P., et al. (2021) Emerging Properties in Self-Supervised Vision Transformers. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 11-17 October 2021, 9650-9660. <https://doi.org/10.1109/iccv48922.2021.00951>
- [15] Sheng, T.J., Islam, M.S., Misran, N., Baharuddin, M.H., Arshad, H., Islam, M.R., et al. (2020) An Internet of Things Based Smart Waste Management System Using Lora and Tensor Flow Deep Learning Model. *IEEE Access*, **8**, 148793-148811. <https://doi.org/10.1109/access.2020.3016255>
- [16] Zhang, Y., Ren, W., Zhang, Z., Jia, Z., Wang, L. and Tan, T. (2022) Focal and Efficient IOU Loss for Accurate Bounding Box Regression. *Neurocomputing*, **506**, 146-157. <https://doi.org/10.1016/j.neucom.2022.07.042>
- [17] 雷建云, 邹金林, 夏梦, 等. 基于改进 YOLOv5s 的可回收垃圾检测方法[J]. 武汉纺织大学学报, 2023, 36(3): 56-64.
- [18] 生态环境部, 国家卫生健康委. 关于印发医疗废物分类目录(2021年版)的通知[EB/OL]. https://www.gov.cn/zhengce/zhengceku/2021-12/02/content_5655394.htm, 2021-11-25.
- [19] 蔡玉芳, 王涵, 李琦, 等. 联合自然梯度和 AdamW 算法的 RSF 图像分割模型[J]. 仪器仪表学报, 2023, 44(3): 261-270.

-
- [20] Gupta, S., Agrawal, A., Gopalakrishnan, K. and Narayanan, P. (2015) Deep Learning with Limited Numerical Precision. *International Conference on Machine Learning*, Lille, 6-11 July 2015, 1737-1746.
- [21] 曹靖城, 张继东, 史国杰. 一种使用边缘增强技术提高相似图片检索召回率的方法[J]. 电信科学, 2021, 37(1): 76-84.
- [22] Li, K., Huang, Z., Cheng, Y. and Lee, C. (2014) A Maximal Figure-of-Merit Learning Approach to Maximizing Mean Average Precision with Deep Neural Network Based Classifiers. 2014 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 4-9 May 2014, 4503-4507. <https://doi.org/10.1109/icassp.2014.6854454>
- [23] 杨智勇, 许倩倩, 何源, 等. 半监督 AUC 优化的 Boosting 算法及理论[J]. 计算机学报, 2022, 45(8): 1598-1617.