arXiv:2601.00398v1 [cs.CV] 1 Jan 2026

RoLID-11K：用于路边垃圾小目标检测的行车记录仪数据集

TaoWu[1], * Qing Xu[1], * Xiangjian He[1], † Oakleigh Weekes[2] James Brown[2] Wenting Duan[2]1宁波诺丁汉大学 [2]林肯大学wduan@lincoln.ac.uk

摘要

路边垃圾带来了环境、安全和经济方面的挑战，但当前的监测依赖劳动密集型调查和公众举报，空间覆盖范围有限。现有用于垃圾检测的视觉数据集主要聚焦于街景静态图像、航空场景或水生环境，无法反映行车记录仪视频的独特特征：垃圾通常极其微小、分布稀疏，并嵌入杂乱的道路边缘背景中。我们推出了 *RoLID-11K*，这是首个用于行车记录仪路边垃圾检测的大规模数据集，包含超过 *1.1* 万张经过标注的图像，涵盖多样化的英国驾驶条件，并呈现出显著的长尾分布和小目标分布。我们对广泛的现代检测器进行了基准测试，从注重精度的变换器架构到实时*YOLO*模型，并分析了它们在这一挑战性任务中的优势与局限。结果表明，尽管 *CO-DETR* 及相关变换器在定位精度方面表现最佳，实时模型仍受到粗粒度特征层级的限制。*RoLID-11K* 为动态驾驶场景中的极端小目标检测确立了一个具有挑战性的基准，旨在支持可扩展、低成本的路边垃圾监测系统开发。数据集可在 https://github.com/xq141839/RoLID-11K。

1. 引言

道路沿线垃圾堆积带来了环境、安全和经济负担。英国各主管部门每年花费数亿英镑进行街道清洁[10]，，而路边碎屑会造成受污染径流、排水受阻，并危害路缘带野生动物[24]。然而，日常监测仍不一致，通常依赖人工检查和公众报告，这些方式在空间和时间覆盖范围上都有限。现有的商业垃圾检测工具，如 LitterCam 和 EnviroEye.AI，侧重于通过固定闭路电视／杆载摄像机抓拍从车辆中乱扔垃圾的人，而非检测或绘制道路路缘带垃圾堆积情况的分布图。此外，这些系统的部署和维护成本高昂，难以大规模采用并实现全面的道路网络覆盖。

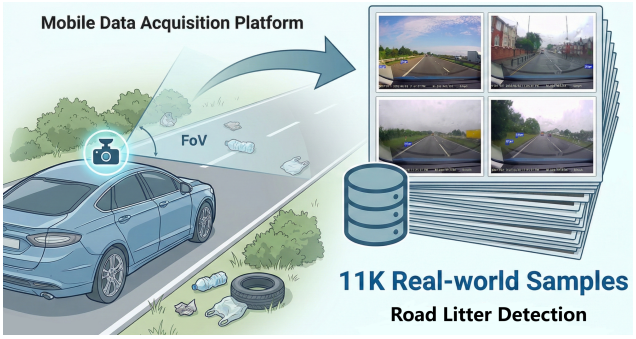

图 1. 所提出的 RoLID-11K 数据集概览。车载行车记录仪作为移动数据采集平台，在各种真实世界驾驶条件下拍摄路边垃圾。该数据集包含 1.1 万张标注图像，涵盖不同的天气、光照和道路环境。

相比之下，行车记录仪价格低廉、应用广泛，并持续拍摄前方道路场景。它们在私家车和商用车队中的普及，为利用已录制的视频进行被动式路边垃圾监测提供了切实可行的机会。然而，检测行驶车辆拍摄到的垃圾颇具挑战：目标通常较小、分布稀疏、类别极不平衡，且容易受到运动模糊、压缩以及杂乱道路背景的影响。现有的废弃物相关数据集，如 TACO [20], TrashNet [25], UAVVaste [15] 和 FloW [6],，并未反映行车记录仪特有的这些条件。

为弥补这一空白，我们推出了 RoLID-11K，这是一个道路垃圾检测数据集，包含超过 11,000 帧标注的行车记录仪画面，展现了具有显著长尾特征且小型目标占比很高的真实路边垃圾。我们对从注重准确率的 Transformer模型到实时 YOLO变体等多种现代目标检测器进行了基准测试，首次对行车记录仪视频中的垃圾检测进行了系统评估

*共同贡献。
†通讯作者。

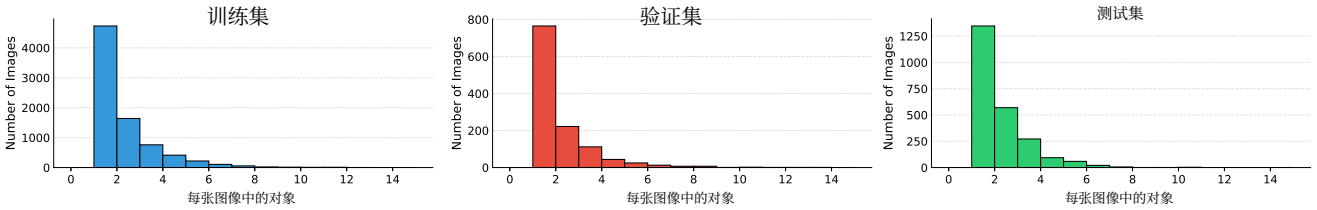

图 2. 目标分布　　我们 RoLID-11K 数据集中训练集、验证集和测试集划分中每张图像的目标数量。该分布呈现　　s 呈现长尾模式，反映了真实　世界道路垃圾场景，其中大多数图像包含 1–3 个物体。

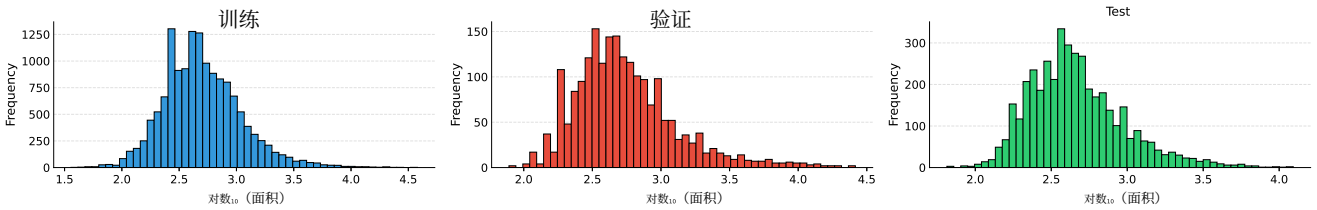

图 3. 物体面积直方图　　在数据集划分中的对数刻度　　. $\log_{10}(\text{Area}) \approx [2.4, 2.8]$ 附近的峰值表明，大多数垃圾物体占据的相对图像中仅占据很小区域，给小目标检测带来了挑战。

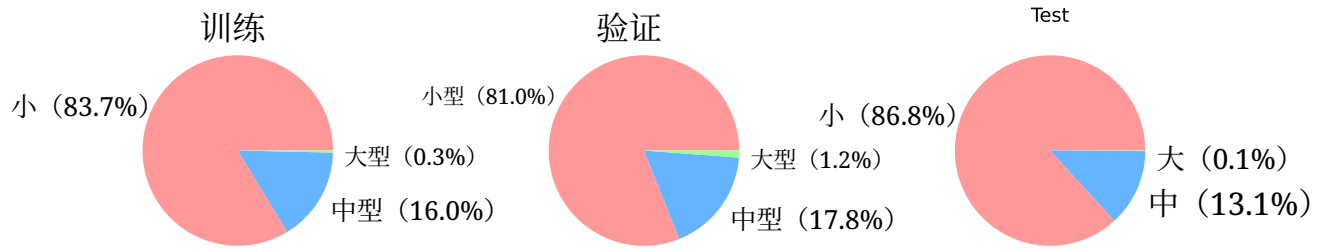

图 4. 按照 COCO 评估标准划分的小型、中型和大型目标尺寸分布。在训练（83.7%）、验证（81.0%）和测试（86.8%）数据集中，小型目标均占主导地位，凸显了道路垃圾场景中小目标检测的挑战。

。RoLID-11K 旨在支持可扩展、低成本的道路污染跟踪方法，并为小目标检测和基于行车记录仪的环境监测研究奠定基础。我们的主要贡献包括：(i) RoLID-11K，这是首个基于行车记录仪进行道路垃圾检测的大规模数据集；(ii) 对真实世界垃圾分布的分析，揭示了显著的长尾特征和小型目标特征；(iii) 对最先进检测器在准确率和实时性场景下进行全面基准测试；以及 (iv) 深入洞察基准测试性能，突出准确率与效率之间的权衡，以及行车记录仪场景下小目标检测所面临的挑战。

2. 相关工作

现有的垃圾和废弃物检测数据集涵盖街道层面、空中环境和水生环境，但没有一个针对通过前视行车记录仪拍摄的道路边缘场景。在街道层面，TACO [20] 提供了约 1.5k 张图像，包含多种类别垃圾标注，而 TrashNet [25] 提供小规模单目标分类数据。UAV-based datasets（如 UAVVaste [15] ）提供废弃物的低空航拍影像；若干水生环境数据集，例如 TrashCan [11]、TrashICRA [9] 和 SeaClear [28]，则聚焦于水下或水面场景中的海洋垃圾。PlastOPol [7]也利用众包的 Marine Debris Tracker 图像，聚焦于多种户外环境中的单类别“垃圾”检测；FloW [6]则针对内陆水域中的漂浮废弃物，提供纯图像子集和多模态图像—雷达子集。近期一项旨在整合这些不同资源的工作是 DetectWaste基准 [18]，它统一了多个数据集（包括扩展版TACO、UAVVaste、TrashCan 和 TrashNet 等）的标注，并修正了标签不一致问题。然而，所有现有数据集都与我们的应用场景有显著差异：图像由固定摄像机、手持设备、无人机或水下机器人拍摄，且垃圾通常占据画面中相当大的部分，或呈聚集斑块状。没有任何数据集提供大规模的

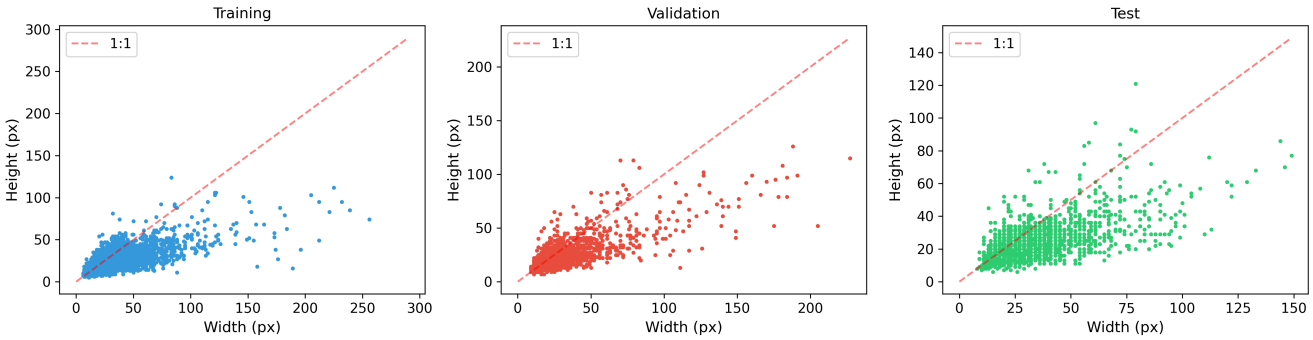

图5。数据集划分中边界框尺寸（宽度与高度）的散点图。虚线表示 1:1 的宽高比。训练集和验证集呈现出集中且相似的分布模式，而测试集则显示出更多样的形状变化和宽高比，为评估模型鲁棒性和泛化能力提供了具有挑战性的基准。

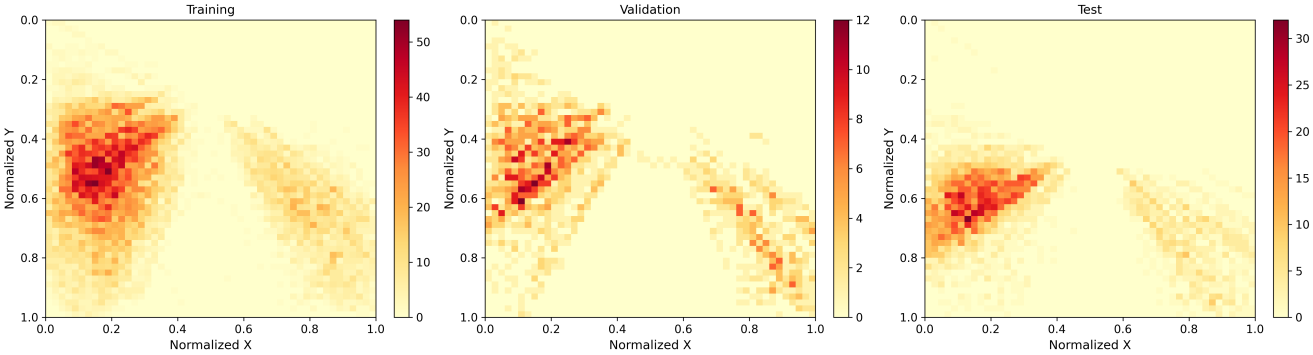

图6。归一化图像坐标中目标中心位置的空间分布。热力图显示，训练集、验证集和测试集中的分布模式一致，目标集中在左侧中央区域，对应于车载摄像头拍摄的道路侧方区域。这种分布反映了道路垃圾检测的真实数据采集场景。

道路边缘行车记录仪视频，其中垃圾以小而稀疏的目标形式出现在路旁，而这正是 RoLID-11K 的关注重点。

基于视觉的垃圾监测已在海洋环境、空中环境和河流环境中得到探索，所采用的方法涵盖多种深度目标检测器和分割器。在 TrashCan 等水下数据集上 [11] 以及相关海洋垃圾数据集中，基线实验通常使用 Mask R-CNN 和 Faster R-CNN 进行实例分割和检测，后续研究则 [16]将轻量级 YOLACT 与 Mask R-CNN 进行比较，以实现实时水下垃圾分割。Sea-Clear [28] 海洋垃圾数据集报告了使用 Faster R-CNN 和 YOLOv6 获得的基线结果，突出了跨不同地点和摄像机进行泛化的难度。对于水面和河流漂浮垃圾 [19], 基于YOLOv5的流水线通常用于视频流中的检测和跟踪。使用 UAVVaste 进行的空中垃圾检测 [15] 依赖于 YOLOv4 和 EfficientDet 等部署在嵌入式硬件上的单阶段检测器，在准确率与机载推理速度之间进行权衡。在更广泛的层面上，Detect-Waste [18] 以及近期的综述 [1] 表明，YOLO变体与 Faster R-CNN、RetinaNet 及相关架构共同主导着当前的垃圾检测系统，通常在扩展版TACO及类似数据集上进行训练。然而，这些方法假设目标尺度适中、视角相对静态，并具有特定领域背景（水面、空中俯视视角、室内分拣线）；而在行车记录仪视频中，垃圾会以极小且稀疏的目标形式出现在道路边缘附近，且场景具有高度动态性。这促使我们对当代检测器开展系统性基准评测，涵盖面向准确率的变换器（DINO [12],C O-DETR [27], DiffusionDet [4]）和实时YOLO模型 [5], , 并在新的 RoLID-11K 数据集上刻画它们在小目标、长尾分布和行车记录仪特有挑战相结合的条件下的表现。

3. 数据集

3.1. 数据采集与标注

RoLID-11K由2022年2月至

表1. 最先进的目标检测模型在我们的 RoLID-11K 数据集上的比较。

方法	发表文献	骨干网络	训练轮次	AP_{50}	$AP_{50:95}$	$AP^{small}_{50:95}$	$AP^{medium}_{50:95}$	$AP^{large}_{50:95}$
CO-DETR [27]	2023年国际计算机视觉大会	ResNet-50	50	**79.2**	**32.1**	**31.2**	**37.5**	**40.0**
DiffusionDet [4]	2023年国际计算机视觉大会	ResNet-50	50	67.0	24.5	24.3	26.7	9.6
DINO [2]	2023年国际学习表征会议	ResNet-50	50	78.5	31.5	30.9	36.1	11.2
RT-DETR [26]	2024年计算机视觉与模式识别会议	ResNet-50	50	73.9	28.9	28.3	32.1	18.5
DEIMv2 [12]	arXiv'25	ViT-Tiny	50	74.3	27.8	27.4	30.3	21.7

表 2. 我们的 RoLID-11K 数据集上实时目标检测模型的比较。

方法	发表文献	骨干网络	训练轮次	AP_{50}	$AP_{50:95}$	$AP^{small}_{50:95}$	$AP^{medium}_{50:95}$	$AP^{large}_{50:95}$
YOLOv8 [13]	Ultralytics'23	CSPDarknet	50	50.1	17.5	16.6	22.9	6.0
YOLOv9 [23]	欧洲计算机视觉会议'24	GELAN	50	50.8	17.1	16.0	23.5	4.0
YOLOv10 [22]	神经信息处理系统大会'24	CSPDarknet	50	49.7	17.4	16.3	23.2	5.1
YOLOv11 [14]	Ultralytics'24	C3K2	50	**52.1**	**18.3**	**17.2**	**24.6**	5.7
YOLOv12 [21]	NeurIPS'25	R-ELAN	50	51.6	17.7	16.9	23.3	**15.1**

2022年7月期间在英国林肯郡录制的4K行车记录仪视频构建而成，视频使用安装在车辆标准前向位置的WOLFBOX 4K/1080p行车记录仪拍摄。视频涵盖多种驾驶环境，包括乡村道路、郊区街道、双向车道和城市环境，真实呈现了各种路边垃圾场景。视频还涵盖了多种天气和光照条件，包括晴天、阴天、弱光和有阴影的环境，从而代表了丰富多样的真实驾驶场景。所有画面均以原始视频的帧率通过OpenCV的VideoCapture接口提取，并以JPEG图像格式保存，确保不会产生依赖内容的采样偏差。为减轻垃圾与背景之间显著的自然类别不平衡，移除了不含可见垃圾的画面。

尽管行车记录仪会采集4K超高清视频，但提取的画面均统一为 1920×1080 分辨率；缩小至1080p既能保持小型垃圾物体的可见性，又能减少存储和标注开销。在基准测试期间，还根据各检测器的输入要求进一步调整了图像尺寸。通过对画面中任何可见的车辆牌照和人脸进行模糊处理，所有图像均完成了匿名化。标注使用VGG图像标注工具 [8], 完成，类别仅设为"垃圾"，并在所有可见垃圾物品周围绘制边界框，包括植被中嵌入的极小物体或部分被遮挡的物体。最终，训练集中获得了14,645个已标注实例，验证集中有2,094个，测试集中有4,189个。

3.2. 数据集划分与统计

最终数据集包含 11,565 张图像，划分为 7,990 张训练图像、1,201 张验证图像和 2,374 张测试图像，即我们基准中使用的数据集划分。RoLID-11K 展现出若干

表 3. 模型复杂度与推理速度对比。

方法	图像尺寸	#参数(M)	浮点运算次数(G)	延迟(ms)
CO-DETR [27]	800×1333	64.5	267.5	6.0
DiffusionDet [4]	800×1333	72.3	192.8	24.3
DINO [2]	800×1333	47.5	274.0	6.6
RT-DETR [26]	640×640	32.0	103.4	2.8
DEIMv2 [12]	640×640	9.7	25.4	10.7
YOLOv8 [13]	640×640	3.0	8.1	0.8
YOLOv9 [23]	640×640	**2.0**	7.6	1.0
YOLOv10 [22]	640×640	2.3	6.5	**0.6**
YOLOv11 [14]	640×640	2.6	**6.3**	0.8
YOLOv12 [21]	640×640	2.6	6.3	0.9

目标检测的一些挑战性特征。每张图像中的对象数呈现明显的长尾分布，大多数图像包含一到三个实例（图 2）。对象尺寸极小：边界框面积的分布峰值约为 $\log_{10}$(面积)≈ 2.4–2.8，这意味着如图3所示，垃圾通常只占每个画面中的极小部分。根据 COCO 尺寸标准，在所有数据集划分中，超过 80% 的标注对象均被归类为小目标（图 4）。边界框长宽比分析进一步显示出较高的差异性，尤其是测试集中的对象形状多种多样，如图 5所示，这增加了实现稳健检测的难度。最后，图 6 展示了物体中心热力图，揭示出图像左下区域存在明显的空间偏差，这反映了英国靠左行驶的典型情况：行车记录仪主要拍摄道路左侧边缘。由于驾驶员行为（将物品丢向路缘）以及风力驱动的位移，垃圾也往往聚集在这一侧，使得

表4. 不同架构的目标检测方法中骨干网络的影响。

方法	骨干网络	AP_{50}	$AP_{50:95}$	$AP^{small}_{50:95}$	$AP^{medium}_{50:95}$	$AP^{large}_{50:95}$	#参数量(M)	浮点运算次数(G)	延迟（毫秒）
DEIMv2 [12]	HGNetv2	71.8	26.1	25.8	28.4	10.7	3.5	6.8	8.8
	ViT-Tiny	74.3	27.8	27.4	30.3	21.7	9.7	25.4	10.2
	ViT-Tiny+	74.2	27.3	26.5	30.6	10.5	18.0	51.9	10.4
YOLOv12 [21]	R-ELAN-N	51.6	17.7	16.9	23.3	15.1	2.6	6.3	0.9
	R-ELAN-S	55.7	20.3	19.2	26.1	13.3	9.2	21.2	1.1
	R-ELAN-M	56.7	20.8	19.9	27.2	5.6	20.1	67.1	1.48
	R-ELAN-L	56.8	21.0	19.9	27.6	18.1	26.3	88.5	1.98
	R-ELAN-X	55.5	20.6	19.4	26.8	14.6	59.0	198.5	2.73

左侧路缘的垃圾出现频率高于右侧。综上，这些特性使 RoLID-11K 成为一个具有挑战性的基准，可用于评估真实驾驶条件下的小目标检测。

4. 实验

4.1. 基准设计与依据

RoLID-11K 代表了一种极具挑战性的目标检测场景，其原因在于极高比例的超小型目标、明显的长尾实例分布，以及动态变化的行车记录仪视角。为在这些条件下有意义地评估性能，我们对两类互补的检测器进行了基准测试：

•以精度为导向的变换器架构 (CO-DETR [27], DiffusionDet [4], DINO [2],RT-DETR [26],和 DEIMv2 [12])，这类架构在 COCO 等数据集上以出色的定位精确率和小型目标敏感度而著称 [17]。

•实时架构 (YOLOv8 [13] - YOLOv12 [21])，广泛应用于汽车和边缘应用领域，因为这些领域对推理速度要求很高。

这种组合使我们能够评估准确性与可部署性之间的权衡，并确定哪些架构趋势（如基于 Transformer 的建模、密集预测头或实时卷积）最适用于基于行车记录仪的垃圾检测。我们纳入 RT-DETR 和 DEIMv2，作为兼顾高准确率与实时推理的现代尝试；同时纳入多个世代的 YOLO，以体现轻量级检测器在实时道路监控系统中的实际价值。这一选择涵盖了当代检测器的现有全貌（2021–2025年），确保我们的基准能够反映最先进水平。

4.2. 实现细节

我们在一台配备 Intel Xeon Silver 4216 中央处理器、256GB 内存和 NVIDIA H200 GPU（141GB 显存）的工作站上完成所有实验。为确保可比性和可复现性，模型均使用其框架提供的默认配置进行训练，Transformer 模型（CO-DETR、Diffu- sionDet、DINO）使用 MMDetection 实现 [3]，而 YOLO 系列模型、RT-DETR 和 DEIMv2 使用 Ultralytics 框架 [13]。对于基于 MMDetection 的检测器，输入分辨率设置为 800×1333，遵循标准 COCO 协议。对于 YOLO 系列、RT-DETR 和 DEIMv2，我们使用默认输入尺寸640×640。所有模型均使用在 COCO 上预训练的权重进行初始化 [17]，并在我们的训练集上微调 50 个训练轮次。推理延迟在批量大小为 1 的条件下，通过完整测试集进行测量，并以每个画面的平均运行时间（毫秒）表示。这些指标（如表3所示）支持对准确率与效率之间权衡的直接比较。

4.3. 评估指标

我们采用标准的 COCO 评估协议 [17]，全面评估模型在 RoLID-11K 数据集上的检测性能。主要指标包括平均精度（AP）。具体而言，我们报告 AP_{50}，，用于衡量 IoU 阈值为 0.5 时的检测准确率；还报告 $AP_{50:95}$，即在 IoU 阈值从 0.5 到 0.95、步长为 0.05 的范围内对 AP 取平均值。$AP_{50:95}$ 对定位质量的评估更为严格。此外，鉴于路边垃圾场景中小型目标十分常见，我们报告小型目标的 $AP^{small}_{50:95}$（面积 $< 32^2$ px^2）、中型目标的 $AP^{medium}_{50:95}$（32^2 ≤面积 $< 96^2$ px^2）以及: 大型目标的 $AP^{large}_{50\ 95}$（面积 $\geq 96^2$ px^2）。这些指标对于评估模型在数据集中固有的、具有挑战性的小目标检测任务上的表现尤为重要。为评估模型在实际部署中的计算效率，我们报告参数量（#Param）、浮点运算次数（FLOPs）以及以每张图像毫秒数衡量的推理延迟。

5. 结果与讨论

表 1和 2总结了以精度为导向的检测器和实时检测器的性能。在基于 Transformer 的模型中，CO-DETR 的 AP 最高$_{50:95}$ 这证实了其出色的定位能力，以及对 RoLID-11K 极端小目标分布特征的鲁棒性

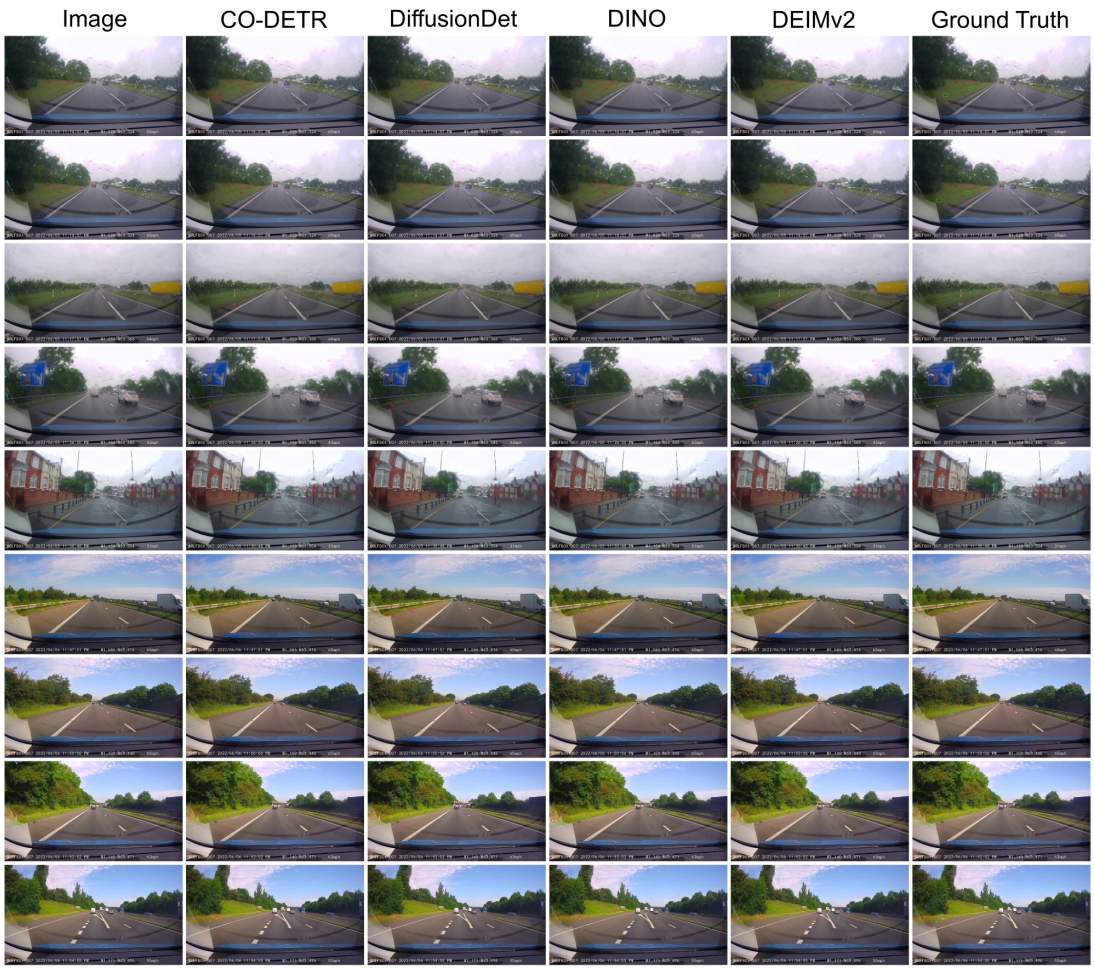

图7. 定性　　在 MMDetection 平台上，最先进检测器在 RoLID-11K 测试集上的对比。

。DINO 的表现也颇具竞争力，而 DiffusionDet 在该数据集上的表现较差，这表明其粗略的去噪调度难以检测嵌入杂乱背景中的微小目标。各模型的 AP_{50} 普遍高于 $AP_{50:95}$ ，反映出对仅占据几十个像素的目标进行精确定位极具挑战性。

实时检测器呈现出预期的速度与准确率权衡。YOLO模型（v8–v12）实现了亚毫秒级推理延迟，同时保持了具有竞争力的 AP_{50} 分数，但与变换器架构相比，其 $AP_{50:95}$ 表现明显落后。这一性能差距在 $AP^{small}_{50:95}$ 上最为显著，进一步表明轻量级检测头和较低的输入分辨率会限制对极小目标的细粒度定位。

表4 表明，骨干网络的选择对性能有显著影响。我们对 DEIMv2 和 YOLOv12 进行了消融研究，因为它们分别是各自模型系列中最新的代表，并且提供了模块化骨干网络，使架构对比更有意义。

对于 DEIMv2，将默认的 CNN骨干网络替换为 ViT-Tiny 后，$AP_{50:95}$ 有了明显提升。这与以下观察结果一致：基于Transformer的编码器能够保留长程上下文信息和精细的空间细节，这对于检测尺寸仅有几个像素的垃圾至关重要。对于 YOLOv12，骨干网络设计和预测头的改进带来了小幅提升，但在检测极小垃圾时，所有版本仍受限于输入分辨率和轻量级特征层级。这些结果表明，在 RoLID-11K 条件下，早期特征提取阶段的架构容量是实现小目标检测的关键因素。

图 7和 8展示了模型在具有挑战性的场景中的预测结果。精度导向型检测器能更可靠地捕捉较小且部分被遮挡的物体，而实时模型则经常漏检嵌在植被中或阴影区域内的物体。YOLO变体往往会产生更多假阴性，但在存在中等大小目标时，仍能保持稳定的检测结果。Transformer模型减少了假

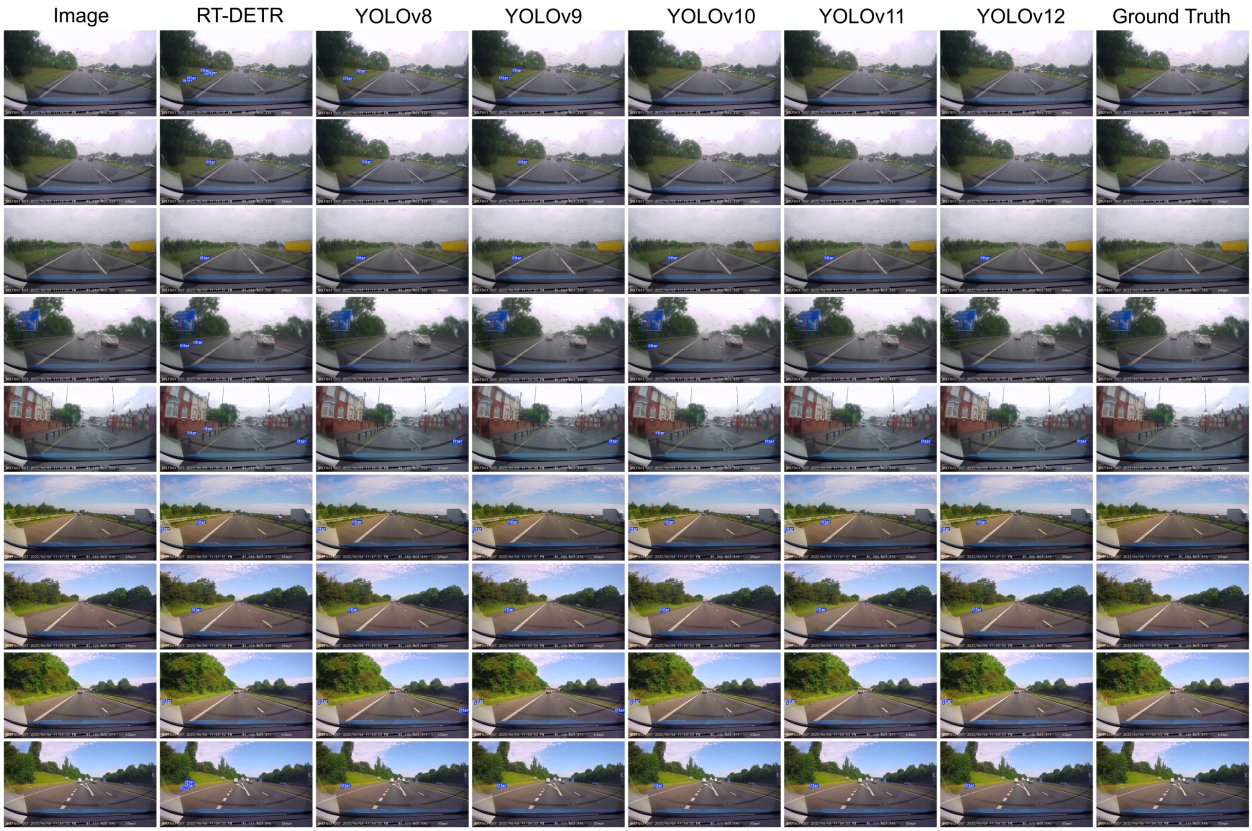

图8。Q 使用 Ultralytics 平台对 RoLID-11K 测试集上的最先进检测器进行定性比较 .

阴性，但偶尔会在纹理复杂的路边区域产生假阳性，这反映了行车记录仪图像中典型的杂乱背景。这些示例凸显了平衡精确率与召回率的难度，尤其是在物体视觉上不明显且位置倾向于靠近图像边界时。

总体而言，我们的结果表明，在所有评估模型中，CO-DETR 的整体 $AP_{50:95}$ 最高，这说明基于变换器的密集分配机制能够为极小且分布稀疏的垃圾实例提供最可靠的定位。然而，尽管以准确率为导向的变换器检测器表现最佳，其计算成本却限制了它们在低功耗平台上的实时部署。相比之下，YOLO模型推理速度极快，但难以捕捉实现稳定小目标检测所需的精细空间细节。这种矛盾凸显了针对极端小目标场景专门设计架构的必要性，例如结合高分辨率特征路径与高效推理机制。RoLID-11K 基准清晰地揭示了这些局限，并为未来开发能够有效用于实时道路监测的模型奠定了基础。

6. 结论

我们推出了 RoLID-11K，这是首个基于行车记录仪视频进行道路垃圾检测的大规模数据集，记录了真实世界监测所面临的挑战：目标极小、分布稀疏，并且在空间上偏向道路边缘。通过对当代检测器进行基准评测，我们发现，目前以准确率为导向的变换器架构在定位性能方面表现最强；而实时YOLO模型尽管速度快，却难以捕捉检测垃圾大小目标所需的精细空间细节。这些发现凸显了针对动态驾驶环境中的极端小目标检测专门设计架构的必要性。RoLID-11K 为可部署的环境监测模型奠定了基础，我们希望它能推动用于跟踪道路污染的低成本系统开发。

致谢

本研究部分由甬江科技创新项目（2022A-097-G）、浙江省交通运输厅一般研究与开发项目（2024039）以及国家自然科学基金项目资助

(62476037).

参考文献

[1] Juan Carlos Arbeláez-Estrada、Paola Vallejo、Jose Aguilar、Marta Silvia Tabares-Betancur、David Ríos-Zapata、Santiago Ruiz-Arenas 和 Elizabeth Rendón-Vélez。《自动分拣系统中废弃物识别的系统性文献综述》。《Recycling》，8(6)，2023。3[2] Mathilde Caron、Hugo Touvron、Ishan Misra、Hervé Jégou、Julien Mairal、Piotr Bojanowski 和 Armand Joulin。自监督视觉Transformer中的新兴特性。载于《IEEE/CVF国际计算机视觉会议论文集》，第9650–9660页，2021。4，5[3] Kai Chen、Jiaqi Wang、Jiangmiao Pang、Yuhang Cao、Yu Xiong、Xiaoxiao Li、Shuyang Sun、Wansen Feng、Ziwei Liu、Jiarui Xu等。MMDetection：OpenMMLab检测工具箱与基准。arXiv预印本 arXiv:1906.07155，2019。5[4] Shoufa Chen、Peize Sun、Yibing Song 和 Ping Luo。DiffusionDet：用于目标检测的扩散模型。载于《IEEE/CVF国际计算机视觉会议论文集》，第19830–19843页，2023。3，4，5[5] Tianheng Cheng、Lin Song、Yixiao Ge、Wenyu Liu、Xinggang Wang 和 Ying Shan。YOLO-World：实时开放词汇目标检测。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第16901–16911页，2024。3[6] Yuwei Cheng、Jiannan Zhu、Mengxin Jiang、Jie Fu、Chang-song Pang、Peidong Wang、Kris Sankaran、Olawale Onabola、Yimin Liu、Dianbo Liu 和 Yoshua Bengio。FloW：内陆水域漂浮废弃物检测数据集与基准。载于《IEEE/CVF国际计算机视觉会议论文集》，第10953–10962页，2021。1，2[7] Manuel Córdova、Allan Pinto、Christina Carrozzo Helle-vik、Saleh Abdel Afou Alaliyat、Ibrahim A. Hameed、Helio Pedrini 和 Ricardo da Silva Torres。深度学习垃圾检测：一项比较研究。《Sensors》，22(2)：548，2022。2[8] Abhishek Dutta 和 Andrew Zisserman。用于图像、音频和视频的VIA标注软件。第27届ACM国际多媒体会议论文集，第2276–2279页，2019。4[9] Michael S. Fulton、Jungseok Hong、Md Jahidul Islam 和 Junaed Sattar。利用深度视觉检测模型进行海洋垃圾机器人检测。载于《IEEE机器人与自动化国际会议（ICRA）论文集》，第5752–5758页。IEEE，2019。2[10] 英国政府。《英格兰垃圾治理战略》。技术报告 PB 14461，英国环境、食品与农村事务部，2017。1[11] Jungseok Hong、Michael S. Fulton 和 Junaed Sattar。Trashcan：一个面向海洋垃圾视觉检测的语义分割数据集。arXiv预印本 arXiv:2007 08097，2020 2，3 . .[12] Shihua Huang、Yongjie Hou、Longfei Liu、Xuanlong Yu 和 Xi Shen。实时目标检测遇上DINOv3。arXiv预印本 arXiv:2509.20787，2025。3，4，5[13] Glenn Jocher、Jing Qiu 和 Ayush Chaurasia。Ultralytics YOLOv8，2023。4，5[14] Glenn Jocher、Jing Qiu 和 Ayush Chaurasia。Ultralytics YOLOv11，2023。4[15]Marek Kraft、Mateusz Piechocki、Bartosz Ptak 和 Krzysztof Walas。利用无人机采集的低空航空图像，在空中环境中自主式地进行机载视觉垃圾检测。《遥感》，13 965，2021 5 1，2，3 ( ): 。[16] Zheng Yong Lim 等。实时实例分割：检测作为塑料来源的水下垃圾。《海洋科学与工程杂志》，11(8):1532，2023。3[17] Tsung-Yi Lin、Michael Maire、Serge Belongie、James Hays、Pietro Perona、Deva Ramanan、Piotr Dollár 和 C Lawrence Zitnick。Microsoft COCO：上下文中的常见物体。载于《欧洲计算机视觉会议》，第740–755页。Springer，2014。5[18] Sylwia Majchrowska、Agnieszka Mikołajczyk 和 Jacek Nowak。用于废弃物检测的综合数据集。《废物管理》，138:279–290，2022。2，3[19] Maiyatat Nunkhaw 和 Hitoshi Miyamoto。使用深度学习技术分析河流漂浮废弃物图像。《水》，16(10)，2024。3[20] Pedro F. Proença 和 Pedro Simões。TACO：用于垃圾检测的上下文垃圾标注。arXiv预印本 arXiv:2003.06975，2020。1，2[21] Yunjie Tian、Qixiang Ye 和 David Doermann。YOLOv12：以注意力为中心的实时目标检测器。载于第三十九届神经信息处理系统年会，2025。4，5[22] Ao Wang、Hui Chen、Lihao Liu、Kai Chen、Zijia Lin、Jungong Han 等。YOLOv10：实时端到端目标检测。《神经信息处理系统进展》，37:107984–108011，2024。4[23] Chien-Yao Wang、I-Hau Yeh 和 Hong-Yuan Mark Liao。YOLOv9：利用可编程梯度信息学习所需内容。载于《欧洲计算机视觉会议》，第1–21页。Springer，2024。4[24] Ting-Wei Wu、Hua Zhang、Wei Peng、Fan Lü 和 Pin-Jing He。卷积神经网络在智能废弃物识别与回收中的应用：综述。《资源、保护与回收》，190:106813，2023。1[25] Mindy Yang 和 Gary Thung。TrashNet。https://github.com/garythung/trashnet，2016。1，2[26] Yian Zhao、Wenyu Lv、Shangliang Xu、Jinman Wei、Guanzhong Wang、Qingqing Dang、Yi Liu 和 Jie Chen。DETR在实时目标检测方面胜过YOLO。载于IEEE/CVF计算机视觉与模式识别会议论文集，第16965–16974页，2024。4，5[27] Zhuofan Zong、Guanglu Song 和 Yu Liu。采用协作式混合分配训练的DETR。载于IEEE/CVF国际计算机视觉会议论文集，第6748–6758页，2023。3，4，5[28] Antun Đuraš、Aikaterini Ilioudi、Bram J. Wolf、Ivana Palunko 和 Bart De Schutter。用于检测和分割浅水域水下海洋垃圾的数据集。《科学数据》，2024。2，3