

# 面向自动驾驶道路场景中异常案例的多模态数据 挖掘算法\*

王海<sup>1</sup>, 张桂荣<sup>1</sup>, 罗彤<sup>3</sup>, 邱梦<sup>2</sup>, 蔡英凤<sup>2</sup>, 陈龙<sup>2</sup>

(1. 江苏大学汽车与交通工程学院, 镇江 212013; 2. 江苏大学汽车工程研究院, 镇江 212013;  
3. 江苏理工学院, 常州 213001)

**[摘要]** 基于深度学习的视觉感知技术的发展有利于自动驾驶系统中环境感知技术的进步。然而, 对于自动驾驶场景中的异常案例, 目前的感知模型还存在一些问题。这是因为基于深度学习的感知模型的能力取决于训练数据集的分布。尤其是驾驶场景中的类别从未出现在训练集中, 感知系统也往往很脆弱。因此识别未知类别和极端场景仍然是自动驾驶感知技术安全性的挑战。本文从处理数据集的角度出发, 提出了一种新颖的多模态异常案例自动挖掘流程(corner case mining pipeline, CCMP)。为验证CCMP的有效性, 在Waymo开放数据集的基础上构建了异常案例子集“Waymo-Anomaly”, 该子集共有3 200个图像, 每个图像都将包含文本中定义的异常案例场景。并且基于私有数据集Waymo-Anomaly, 证明了CCMP针对异常案例场景挖掘的召回率可以达到91.7%。此外, 还通过实验验证了目标检测器在包含异常案例的数据集中针对长尾分布的有效性。最终, 希望从处理数据集的角度来提高自动驾驶感知模型在现实世界中的真实性。

**关键词:** 自动驾驶; 深度学习; 目标检测; 异常案例

## A Multi-modal Data Mining Algorithm for Corner Case of Automatic Driving Road Scene

Wang Hai<sup>1</sup>, Zhang Guirong<sup>1</sup>, Luo Tong<sup>3</sup>, Qiu Meng<sup>2</sup>, Cai Yingfeng<sup>2</sup> & Chen Long<sup>2</sup>

1. School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013;  
2. Institute of Automotive Engineering, Jiangsu University, Zhenjiang 212013;  
3. Jiangsu Institute of Technology, Changzhou 213001

**[Abstract]** The development of visual perception technology based on deep learning is beneficial for the advancement of environment perception technology in automatic driving systems. However, for corner cases of autonomous driving scenario, there are still some problems in the current perception model. This is because the ability of the perception model based on deep learning depends on the distribution of the training dataset. Especially when categories in the driving scene never appear in the training set, the perception system is often fragile. Therefore, identifying unknown categories and extreme scenarios remains a challenge for the safety of automatic driving perception technology. From the perspective of processing data sets, in this paper a novel multimodal automatic corner case mining process called "Corner Case Mining Pipeline (CCMP)" is proposed. In order to verify the effectiveness of "CCMP", the concern case subset "Waymo-Anomaly" on the basis of Waymo open datasets is established, with a total of 3 200 images, each of which will contain the corner case scene defined in the text. Then based on the private data set Waymo-Anomaly, it is proved that the recall rate of "CCMP" corner case mining can reach 91.7%. In addition, the effectiveness of object detectors targeting long-tailed distributions in datasets containing corner case is ex-

\* 国家自然科学基金(52225212, U20A20333, 52072160)和江苏省重点研发项目(BE2020083-2, BE2020083-3)资助。

原稿收到日期为2023年11月03日, 修改稿收到日期为2024年01月15日。

通信作者: 王海, 教授, 博士, E-mail: wanghai1019@163.com。

perimentally verified. Ultimately, the authenticity of the automatic driving perception model in the real world is expected to improve from the perspective of datasets processing.

**Keywords:** autonomous vehicles; deep learning; object detection; corner case

## 前言

自动驾驶技术的发展逐渐影响着我们的生活,而环境感知技术在高度自动驾驶系统中是一项极具挑战性的工作<sup>[1]</sup>。如果要获得广泛的社会认可,自动驾驶系统的感知能力必须更加稳定和可靠。感知技术通常负责检测自动驾驶系统中物体的语义和位置信息,甚至警告潜在的危险。感知系统是否对关键场景处理具有鲁棒性是衡量感知模型性能的重要标准。

近年来,基于深度学习的感知模型在自动驾驶领域得到了广泛应用<sup>[2]</sup>。这些模型主要依赖于大量的训练数据集,并通过不断迭代训练逐渐适应多个场景。有监督的深度学习方法通常只能检测在训练过程中已经遇到的类别。如果模型的推理过程遇到比例很小的类别,甚至是超过训练集类别分布的类别,则检测效果往往较差,这类特殊类别存在的情况被称为“异常案例”。由于异常案例的定义决定了此类物体在正常自动驾驶环境中出现的概率较低,因此异常案例的存在大大增加了驾驶环节中的潜在风险。总而言之,针对此类物体的检测对于可靠、安全的自动驾驶环境感知系统至关重要。

本文讨论的异常案例问题分为两大类:“像素级异常案例”和“对象级异常案例”。(1)当图像中的所有或大量像素点分布在预期测量范围之外时,就会出现像素级异常案例。这通常是异常照明条件或过度曝光图像的结果。图1(a)是像素级异常案例的一个例子,图中水滴部分模糊图像完全遮挡住了交通物体的视线。(2)对象级异常案例主要发生在推理过程中,当模型遇到训练过程中没有发生的实例时,这些未知物体被归类为新颖类别。如图1(b)所示,人力三轮车通常会被排除在自动驾驶训练数据集之外。因此对象级异常案例通常是指出现在自动驾驶场景中的任何新颖对象,如动物、各种形状的汽车和其他障碍物。这些对于环境感知系统的安全性至关重要,因为对于异常案例的检测是当前主流检测器的弱点。然而,为自动驾驶车辆的安全运行,自动驾驶感知系统必须对这些异常案例做出积极反馈。

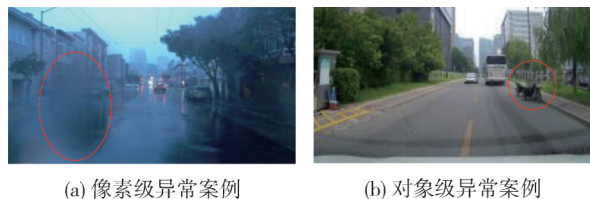


图1 自动驾驶场景中的异常案例

当前的针对自动驾驶道路场景中的异常案例解决方案可分为在线处理和离线处理。大多数基于视觉的在线异常检测方法都是基于置信度。例如依赖于开放数据集,引入未知类别分支,为训练数据集之外的未知类别提供预警<sup>[3-4]</sup>。文献[5]中旨在解决自动驾驶低光照场景下难以检测的问题。另外,在语义分割任务中,使用蒙特卡罗失活采样方法来测量每个像素的不确定性<sup>[6-7]</sup>。上述方法都是从在线的角度来优化感知模型结构,以增强模型对各种极端场景的泛化能力,然而在线处理方式会增加计算的复杂度。在离线方案中主要是使用重构的方法。该方法通常是设计一个生成网络来重建一个与原始图像的风格和类型一致的图像,但不包含异常案例目标的图像。通过将重建图像与实际图像进行比较,可以识别出特殊目标。在文献[8]和文献[9]中,使用图像分割编码器和解码器结构对输入图像进行重新合成,旨在依赖于解码结构和编码结构之间的感知损失,实现了异常图像的生成。然而,此类离线方案难以捕捉稀有类别的特征,尤其是像素级异常案例对于图像的重构会产生极大的干扰。

场景分布的普遍性和数据集中类别分布的平衡性直接影响模型的异常案例处理性能。因此,本文从数据集的角度提出了一种离线优化方法,以增强自动驾驶系统对异常案例的鲁棒性。当前训练数据集中针对异常案例的信息挖掘仍是一个繁琐的过程,因为在BDD100K<sup>[10]</sup>、Waymo<sup>[11]</sup>等大型自动驾驶数据集中有超过100 000张图像。并且异常案例的发生又与数据集中的长尾分布问题一致<sup>[12]</sup>,即异常案例发生的频率通常与长尾问题中的尾部数据相对应。因此,手工逐个搜索的效率极低。当然通过模拟生成的方法以达到人为扩充异常案例数据集可以为像素级

异常案例创建大量的模拟场景,如强光、微光或雾天<sup>[13-14]</sup>。然而,模拟算法获得的异常案例场景中像素分布与真实场景的像素分布之间仍然存在差距,导致模型在真实世界中的性能普遍低于模拟测试集数据。

本文中创新性地利用激光雷达数据和相机数据各自的优势,提出了一种在现有大型数据集中自动挖掘异常案例的流程CCMP。同时,参考Breitenstein对自动驾驶场景中异常案例的定义<sup>[15]</sup>,即只关注车辆可行驶范围内的物体。这是因为在可行驾驶区域之外或位于远处的物体对驾驶任务的规划和决策影响较小。另外,激光雷达和相机传感器数据各有优势,具体来说,激光雷达数据可以降低模型故障率,且相机数据的语义信息比激光雷达数据更强。CCMP将来自两种类型传感器的数据结合起来,进行异常案例数据挖掘。同时,利用CCMP在Waymo数据集的基础上构建了一个自动驾驶的异常案例子集“Waymo-Anomaly”,为基于监督深度学习的方法处理现实世界中的异常案例场景提供了研究基础。最后,由于异常案例的长尾性质,使用能够处理长尾问题的检测器<sup>[16-17]</sup>对合并的异常案例场景的训练集进行了重新训练,进一步研究并验证了现有方法在现实世界存在异常案例中解决长尾分布问题的有效性。

## 1 异常案例挖掘流程

### 1.1 流程概览

提出一种新颖的多模态异常案例挖掘算法,称

为CCMP。其采用多种模式作为输入,主要由激光雷达和相机传感器组成。这是由于激光雷达传感器的特性,其对周围物体的感知具有较高的召回率,而相机传感器比激光雷达传感器具有更多的语义信息。同时,本文明确定义了CCMP的异常案例挖掘对象,主要满足两个条件:(1)物体必须位于自动驾驶的潜在驾驶区域内;(2)目标不属于训练集类别,或者由于光线较低或较强而难以使用视觉方法进行检测的类别。为获得质量更可靠的异常案例提议框,该挖掘过程设计成两个主要阶段:第一个阶段是潜在提议框生成阶段,在该阶段,提出了一个自动化的初始提议框筛选过程,从大量初始数据中筛选出存在异常案例的场景;第二阶段是手动处理阶段,它进一步消除了第一阶段中生成的假阳性实例,对发现的实际异常案例的场景进行分类,并细化其边界框以获得更高的精度。

研究的主要工作是在第一阶段,即如何构建一个自动化的异常案例过滤及筛选过程,如图2所示。CCMP过程须具有相同时间戳的多模态输入,例如2D图像数据和3D点云数据。为满足第一个异常案例的定义,必须对当前场景的道路平面进行重建,以确定符合异常案例的目标是否出现在道路平面上,从而避免背景对象(如树木和建筑物)的干扰。具体地,首先使用语义分割模型对输入的2D图像数据进行可行驶区域分割,以获得图像中可行驶区域的掩码坐标。然后将同一时间戳的3D点云数据投影到2D图像域中,并获得其2D图像坐标。当点云的2D

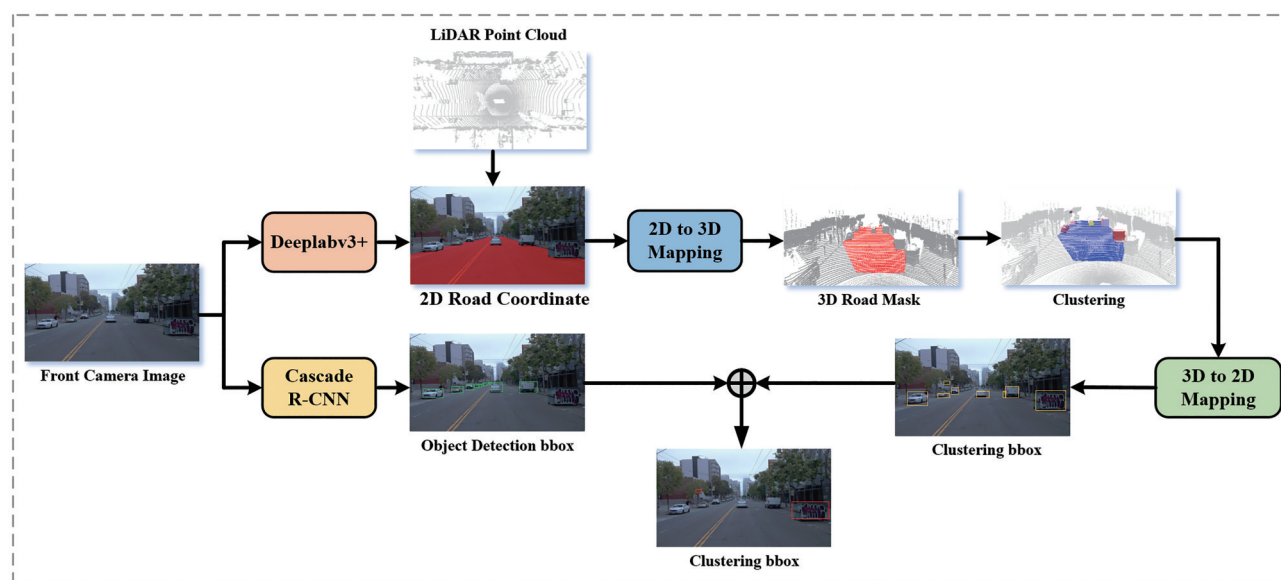


图2 CCMP异常案例数据自动化挖掘流程

图像坐标与可驾驶区域的掩码坐标重合时,将重合的点云2D图像坐标反投影到3D空间,且获得感兴趣区域的近似道路3D点云坐标。然后利用一致性采样方法优化道路上方的点云,并利用图形重建道路模型。为满足本文所提出的异常案例第二个定义并提高异常案例数据挖掘的召回率,最初的异常案例使用点云聚类生成的建议通过2D检测器的检测结果来过滤掉不符合条件的异常案例场景。

### 1.2 可行驶区域分割

由于并非当前场景中的所有对象都符合文中对异常案例的第一个定义,为减少非车辆路线上的对象对异常案例挖掘过程的干扰,首先,必须确定感知空间内的可行驶区域。由于使用二维图像数据,可以获得比稀疏激光雷达点云更加详细、更有纹理的语义级别信息。因此,对输入图像的可行驶区域进行语义分割。因为大多数多模态数据集,如Waymo Open Dataset<sup>[11]</sup>,不包括可行驶区域的像素级标签,Nusence也没有提供这样的标签。因此,在MMsegmentation框架<sup>[18]</sup>和BDD100K数据集<sup>[10]</sup>的基础上进行训练,以完成可行驶区域分割的任务。经过实验比较,选择了Resnet-101-D8<sup>[19]</sup>为特征提取网络的Deeplabv3+模型<sup>[20]</sup>。由于该模型是BDD100k官方模型中在可驾驶区域分割方面得分最高的模型,在测试数据集中的mIoU也高达54.41%。在对私人数据集中测试后,发现Deeplabv3+也是对陌生数据表现最为稳健的模型之一。图3展示了可驾驶区域分割为Deeplabv3+收集的真实世界图像数据的视觉影响。要在三维空间中对道路进行建模,必须将二维图像中的道路掩码投影到三维空间中。如图4所示,首先将点云数据反向投影回2D空间,且将反向投影区域与道路掩码区域重合的区域认定为3D道路点云数据。



图3 可行驶区域语义分割掩码

### 1.3 道路平面重构

可行驶区域分割提供了粗略的道路点云数据,但仍然存在问题。在图5的左上角,可以观察到,在将2D图像道路掩码反向投影到3D的过程中,由于

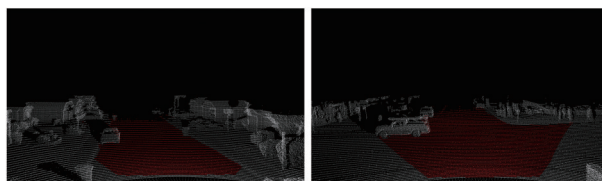


图4 可行驶区域分割结果投影

投影映射本身的缺陷,一些点云被错误地归类为道路点云。同时,在图5的左下方,由于道路和其他物体边界之间的分割模型不准确,路面、车辆或其他交通参与者之间的分割边界不清晰,导致道路上投影回3D空间中可驾驶区域的点云与对象的点云产生重叠现象,从而极大地影响道路平面的重建效果。因此,必须进一步细化粗糙的道路点云数据。

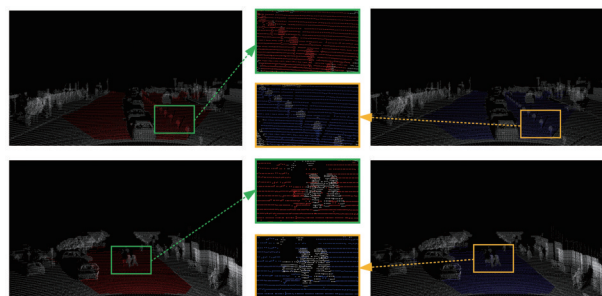


图5 道路异常点云可视化及优化后可视化

首先,为拟合图像地平线的范围,在整个点云数据范围中只保留当前车辆的前瞻范围,去除了图像地平线之外的点云信息,且转换了点云的视角。接下来,使用随机样本一致性方法RANSAC<sup>[21]</sup>来进一步拟合平面,以减少由不准确的图像分割和投影缺陷引起的道路重建干扰。属于平面方程的点云被定义为内部点,不属于平面方程式的点云是外部点。RANSAC一致性采样由3个主要部分组成:(1)将平面方程拟合到初始采样的15个点;(2)将所有点依次放入平面方程中,并根据设定的距离阈值进行判断(本文认为平面模型30cm以外的点云被视为非道路点云),如果该点在阈值内,则认为该点属于平面内点,否则为平面外点;(3)展开的局部点优化了平面模型方程。将迭代次数设置为500轮。在迭代范围内,点数最多的平面方程为实平面方程,平面方程下的内点为道路点云,外点为障碍物点云。图5的右侧显示了RANSAC处理的点云。可以发现,汽车边缘的点云由于二维到三维投影过程而被错误分类的问题得到了很好的解决。同时,设置30cm的距离阈值主要是为考虑到道路上的一些小物体目标

(如树枝和瓶子)对驾驶安全影响不大,但存在于车辆的可驾驶区域。

引入RANSAC算法主要解决了两个问题:(1)初始粗糙道路点云将道路与目标清晰地分开;(2)解决“小目标”对异常案例挖掘道路的干扰。除重新优化道路点云外,文中还消除了导致后续道路建模结果恶化的异常点云。具体地,计算了15个相邻点云的标准差。如果点云超过其邻居的平均标准偏差的阈值,则视为异常。在去除异常值后,远处的点云也得到了明显的优化。随着距离的增加,激光雷达数据变得越来越稀疏,如果它的标准偏差大于最近点的8倍,文中就排除了这个点。

接下来需要进一步地处理来确定道路上是否存在目标物体。在图6中,需要在优化后的道路点云上重建道路平面。使用一种简单有效的阿尔法形状估计方法,将点云重建为基于三角形网络的道路平面。该方法将凹形车体定义为被点云占据的区域,并通过阿尔法形状估计算法创建一个多边形来近似表示道路区域。多边形的顶点是点云数据,多边形的精细度由一个常数 $\alpha$ 决定。

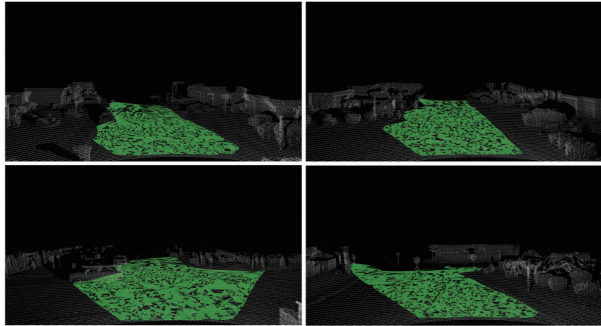


图6 道路平面重构可视化

$\alpha$ 值越大,多边形越接近凸包。当 $\alpha$ 参数值设置为零时,多元变量区域退化为原始点云数据。在实验部分,文中设置了各种 $\alpha$ 值,并比较了道路平面重建效果,最终将 $\alpha$ 值设置为30。接下来,将道路点云以外的点云映射到道路平面上,即将其他点云平铺到2D维度。点云的其余部分用道路上方的物体标记。

#### 1.4 异常案例挖掘初始化

在获得道路上的点云后,为提高异常案例挖掘的召回率,对道路上的点云进行聚类。最终,使用简单高效的DBSCAN聚类算法<sup>[22]</sup>初步筛选出可能符合异常案例的目标对象。DBSCAN是一种基于密度的聚类算法,通常假设类别可以通过样本紧密度来确

定。通过将密切相关的样本划分为单个对象,获得聚类个体。聚类的结果是将所有连接紧密的样本分成不同的样本。通过DBSCAN算法,可以获得不同的簇,其中每个簇由至少MinPts个样本点组成,且具有足够高的密度。其优点是不需要预先指定簇的个数,可以发现任意形状的簇,对噪声数据具有较好的鲁棒性。

潜在的目标初始化提议流程如算法1所示:在优化可驾驶区域的道路上设置点云 $P = \{p_1, p_2, \dots, p_m\}$ 。如果点簇 $P$ 与点簇 $p_j$ 之间的距离小于设置的阈值 $\alpha$ ,且小于阈值的样本个数大于设置的个数MinP,则定义 $p_j$ 为核心点:

$$N_\alpha(p_j) = \{p_i \in P \mid \text{dist}(p_i, p_j) \leq \alpha\} \quad (1)$$

$$|N_\alpha(p_j)| \geq \text{MinP} \quad (2)$$

#### 算法1

---

**Input:** Point cloud:  $P = \{p_1, p_2, \dots, p_m\}$ ;

---

**Parameter:**  $(\alpha, \text{MinP})$ .

---

**output:** Target division:  $C = \{C_1, C_2, \dots, C_k\}$

---

**process:**

---

1: Initialize core point collection:  $\Omega \neq \phi$

---

2: **for**  $J = 1, 2, \dots, m$  **do**

---

3: Determining the domain of point cloud:  $p_j; N_\alpha(p_j)$

---

4: **if**  $|N_\alpha(p_j)| \geq \text{MinP}$  **then**

---

5: Add point cloud to the core point:  $\Omega = \Omega \cup \{p_j\}$

---

6: **end if**

---

7: **end for**

---

8: Initialize target quantity:  $k = 0$

---

9: Initialize point cloud collection not accessed:  $\Gamma = P$

---

10: **while**  $\Omega \neq \phi$  **do**

---

11: Record that the point cloud collection:  $\Gamma_{\text{old}} = \Gamma$

---

12: Randomly select a core point  $O \in \Omega$ , Initialize Point Cloud Queue  $Q = \langle O \rangle$

---

13:  $\Gamma = \Gamma / \{O\}$

---

14: **while**  $Q \neq \phi$  **do**

---

15: Take the first point cloud sample from queue  $Q$ :  $q$

---

16: **if**  $|N_\alpha(p_j)| \geq \text{MinP}$  **then**

---

17:  $\Delta = N_\alpha(q) \cap \Gamma$

---

18: Adding the point cloud samples in  $\Delta$  to the queue  $Q$

---

19:  $\Gamma = \Gamma / \Delta$

---

20: **end if**

---

21: **end while**

---

22:  $k = k + 1, C_k = \Gamma_{\text{old}} / \Gamma$

---

24: **end while**

---

式中:  $N_{\alpha}(p_j)$  是  $p_j$  的领域,  $|N_{\alpha}(p_j)|$  为点簇  $P$  的中心点, 所有核心点的集合为  $\Omega$ ;  $dist$  为  $p_i$  和  $p_j$  之间的欧氏距离。然后, 随机选择未遍历的核心点作为种子点, 利用欧氏距离判断点云是否属于同一物体。当遍历完  $\Omega$  中的所有核心点时, 循环结束。经过多次实验之后, 最终选择  $\alpha = 1.5$  和  $MinP = 35$  作为 DBSCAN 的参数。最后, 道路上的所有目标物(包括不属于原始训练集的障碍物)通过点云聚类生成 3D 框, 之后再将 3D 框映射到二维空间, 并将二维空间的目标物体提议框作为异常案例对象的初始提议框。

### 1.5 常规类别抑制

为满足自动驾驶场景中异常案例的第二个定义, 文中使用更强大的目标检测器来过滤掉非常案例对象。具体地, 使用 Cascade R-CNN 级联检测标头<sup>[23]</sup>, 并用 ConvNeXt-S<sup>[24]</sup> 替换原始特征提取网络, 在 BDD100K 等自动驾驶数据集上进行训练, 以便为每个可识别或易于识别的对象生成一组边界框。下一步是比较检测后生成的边界框和聚类投影的异常案例初始提议框, 并消除一些检测到的对象的 IOU 值超过阈值的提议框。值得一提的是, 文中并没有直接使用数据集的正值框与生成的异常案例的初始提议框进行比较, 因为通过这种方法无法挖掘出目标检测中的“困难案例”。这种困难案例是指上文所提及的因强光或弱光所引起的像素级异常案例。最后, 可以通过手动微调 and 标记类别来完成 CCMP 过程。

### 1.6 Waymo-Anomaly 的特性

最后, 为证明 CCMP 过程在异常案例数据挖掘中的有效性, 最终将其应用于 Waymo 数据集, 以自动生成异常案例子集, 称之为 Waymo-Anomaly (值得注意的是 CCMP 仍然适用于其他多模态数据集)。Waymo 数据集是由谷歌旗下自动驾驶公司 Waymo 收集的一套大规模自动驾驶数据集。该数据集旨在支持自动驾驶系统的研究和开发, 为算法训练和验证提供有用的数据。该数据集包含了来自多个传感器(如激光雷达、摄像头和雷达)的高分辨率感知数据, 以及相关的车辆信息和行为轨迹数据。这些数据涵盖了各种不同的行驶场景, 包括城市道路、高速公路和农村地区等。

除 Waymo 数据集定义的类别外, Waymo-Anomaly 还包括车辆、行人和骑自行车的人。在 CCMP 数据挖掘的过程中, 文中还定义了一些新

类别, 如障碍物、动物、交通设施和杂项。常用类表示数据集中已经定义的类别, 而新颖类表示通过 CCMP 挖掘过程发现的新类别。检测器很难检测到基于封闭世界数据集(公共类)的新类别, 而新类别的识别对于现实世界的驾驶安全非常重要。图 7 为 Waymo-Anomaly 的分类统计。可以发现数据集的角落案例有一个长尾, 新类只占 0.4%, 这对基于深度学习的感知网络来说仍然是一个挑战。在随后的实验部分中, 研究了长尾方法对 Waymo-Anomaly 数据集的适用性。

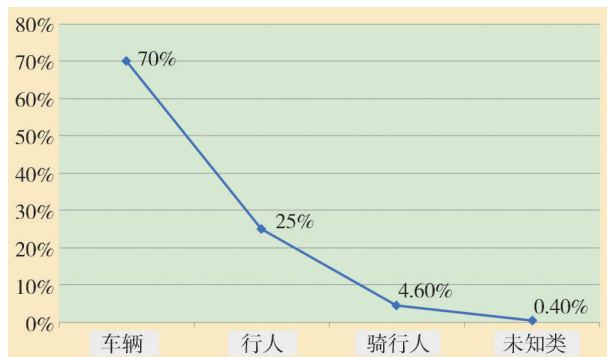


图7 Waymo-Anomaly 的类别统计

## 2 实验验证

### 2.1 实施细节

为验证“CCPM”的有效性, 基于 Waymo 大型多模态数据集构建了一个异常案例子集 Waymo-Anomaly。除原始数据集中的“已知”类别包括车辆、行人、骑自行车的人和标志外, 文中还新定义了一些超类, 如障碍物、动物、交通设施和杂项。去除符号类别的评价, 使实验结果更加简洁明了。最初使用“CCPM”筛选了 3 257 张带有异常案例的自动驾驶场景图片, 然后手动调整了总共 3 200 张图像。在“未知”类别部分, 添加了诸如“动物、障碍物和其他”之类的类别。同时, 基于 8:2 的比例将 3 200 张图像划分为一个训练集和一个验证集, 即 2 560 个训练集和 640 个验证集。具体来说, 3 200 张训练集图像不用于模型再训练, 而是在标准数据集训练完成后进行微调。640 个验证集和 Waymo 随机选择的 1 360 个验证集相结合, 用于验证未知的类对象。同时, 采用了 3 种性能更好的基线模型: 单级模型(如 YOLOX-L<sup>[25]</sup> 和 ATSS<sup>[26]</sup>)、两级模型(Cascade R-CNN 和 Sparse RCNN<sup>[27]</sup>) 和基于 Transformer 的检测模型

(DDetr<sup>[28]</sup>)。对于评价标准,遵循 CoCo 数据集<sup>[29]</sup>的评价指标,而对于“未知”类别,只使用  $AR_{50}$  (平均召回率为 50% 时的精度值) 来评价模型对未知类别的鲁棒性。

此外,实验的整个模型代码来源于商汤的 MMDetection<sup>[30]</sup>, 并采用 Pytorch1. 12. 2 深度学习框架, 在 64 位 NVIDIA 和使用 8 个 GTX3090 GPUS 的 Ubuntu20. 04 上进行分布式数据并行训练和同步批处理归一化。训练时统一使用随机梯度下降(SGD)作为优化器,其权重衰减率为  $4 \times 10^{-4}$ , 动量为 0. 9。

## 2.2 实验结果

对 Waymo-Anomaly 进行微调后,“已知”类别有

所改善,如表 1 所示。通过检测结果的可视化,可以确定指标的改善主要是由于如图 8 所示的异常光照条件。在对角点情况进行增强训练后,该模型对曝光和低光照情况具有更强的鲁棒性。同时,可以发现,虽然多阶段模型对“已知”类别的检测精度普遍较高,但其对“未知”类别的检测性能不如单阶段网络。这可能是因为单阶段模型无须先区分前景,更密集的预测方法对于“未知”类别检测的召回率指标更有利<sup>[31]</sup>。因此,本文提出的 CCMP 异常案例挖掘过程适用于大型数据集,允许更有效的异常案例场景的过滤,且是建立基于深度学习模型的数据闭环的关键步骤。

表 1 各类检测器在 Waymo 和 Waymo-Anomaly 上的检测结果对比

%

算法	数据集	车辆(AP)	行人(AP)	骑行者(AP)	所有类别(AP)	未知类( $AR_{50}$ )
YOLOX	Waymo	58.2	43.8	34.0	41.3	
RetinaNet <sup>[32]</sup>		56.6	42.1	33.7	40.5	
Faster R-CNN <sup>[33]</sup>		59.5	44.9	36.1	42.0	
Cascade R-CNN		63.7	46.3	40.5	43.9	
Sparse RCNN		56.1	41.0	30.6	39.9	
DDetr		60.1	42.3	35.3	42.7	
YOLOX	Waymo-Anomaly	58.3	43.8	34.0	41.3	34.5
RetinaNet		56.9	42.1	33.6	40.5	30.1
Faster R-CNN		62.1	46.0	36.2	43.1	29.8
Cascade R-CNN		66.5	47.8	40.5	45.2	29.5
Sparse RCNN		56.8	41.3	30.6	40.0	28.1
DDetr		61.5	42.5	35.9	43.8	35.0

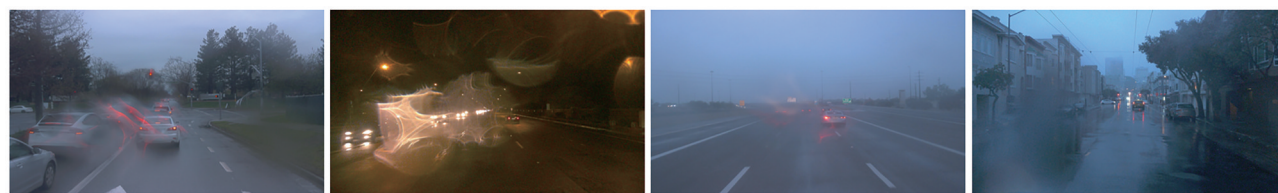


图 8 特殊光照场景的异常案例

## 2.3 验证长尾方法的适用性

本文中认为在异常案例场景中“未知”类别的数量通常是很小的。如图 7 所示, Waymo-Anomaly 中的“未知”类别仅占 0. 4%。因此,在实验部分,可以使用一些以前的方法来优化长尾数据,以检查他们的方法对现实世界的异常案例的适用性。主要从 Re-weight 的角度选择一些方法来解决长尾问题,并将 Waymo-Anomaly 中的“未知”类别添加到 Waymo 训练数据集中进行重新训练。Waymo 随机抽取的部分验证集和 Waymo-Anomaly 验证集共 2 000 张图像作为本次实验的验证集。观察表 2 未优化长尾数据

的结果( $r$  表示“未知”类别, $c$  表示普通类别, $f$  表示车人结合), 可以发现在 CCMP 过程中挖掘异常案例后,在“未知”类别参与再训练时仍然存在严重的漏检现象。

然而,通过进一步实验,发现引入针对长尾问题的方法可以更好地解决“未知”类别召回率低的问题。本文分别探索了均衡焦点损失(equalized focal loss, EFL)和动态类别抑制损失(adaptive class suppression loss, ACSL)对于单阶段检测器和多阶段检测器应对长尾问题的效能。其中,EFL是在传统的 Focal Loss 的基础上引入了类别权重平衡因子,且

考虑了不同对象的易分类程度。如此,稀有对象的权重会被放大,以便模型更关注这些稀有对象。ACSL则是自适应调整每个类别样本的抑制梯度,从而提高检测器对于罕见类别的敏感度。通过表2后3列的实验结果对比,可以发现EFL更适用于单阶段目标检测模型中的“尾类”检测,而ACSL更适用于两阶段检测算法。

### 2.4 可视化验证

在本文提出的异常案例的多模态数据挖掘过程

中,首先利用二维图像中更丰富的语义信息对可行驶区域进行分割,然后将二维图像的道路坐标反向投影到三维空间中。下一步是使用一致性采样算法对道路点云进行优化,剔除可能对道路重建造成较大干扰的点云。在道路重建完成后,基于密度分布的聚类方法可以只处理正上方区域以上的点云,然后将结果与强大检测器的结果进行比较,以研究潜在的异常案例。接下来,对基于Waymo数据集的异常案例挖掘过程的几个关键步骤进行可视化以验证挖掘数据的准确性和可行性。

图9为基于Waymo数据集挖掘异常案例的可视化效果图。第1列为原始图像的异常案例目标,第2列为同一时间戳对应的点云图像,第3列为优化后的点云聚类后的道路重建可视化和未知目标。这些对象满足本文提出的异常案例两条定义,符合异常案例的目标只处于当前车辆潜在可行驶区域内。

表2 长尾问题的优化方案及结果 %

Model	$AP_r$	$AP_c$	$AP_f$	$mAP$
RetinaNet	13.3	33.6	58.6	42.3
Faster R-CNN	12.1	36.0	60.1	43.9
RetinaNet+EKF	20.3	34.9	58.8	44.1
Faster R-CNN + EFL	16.1	36.7	60.1	44.2
Faster R-CNN + ACSL	18.8	37.2	60.9	44.7

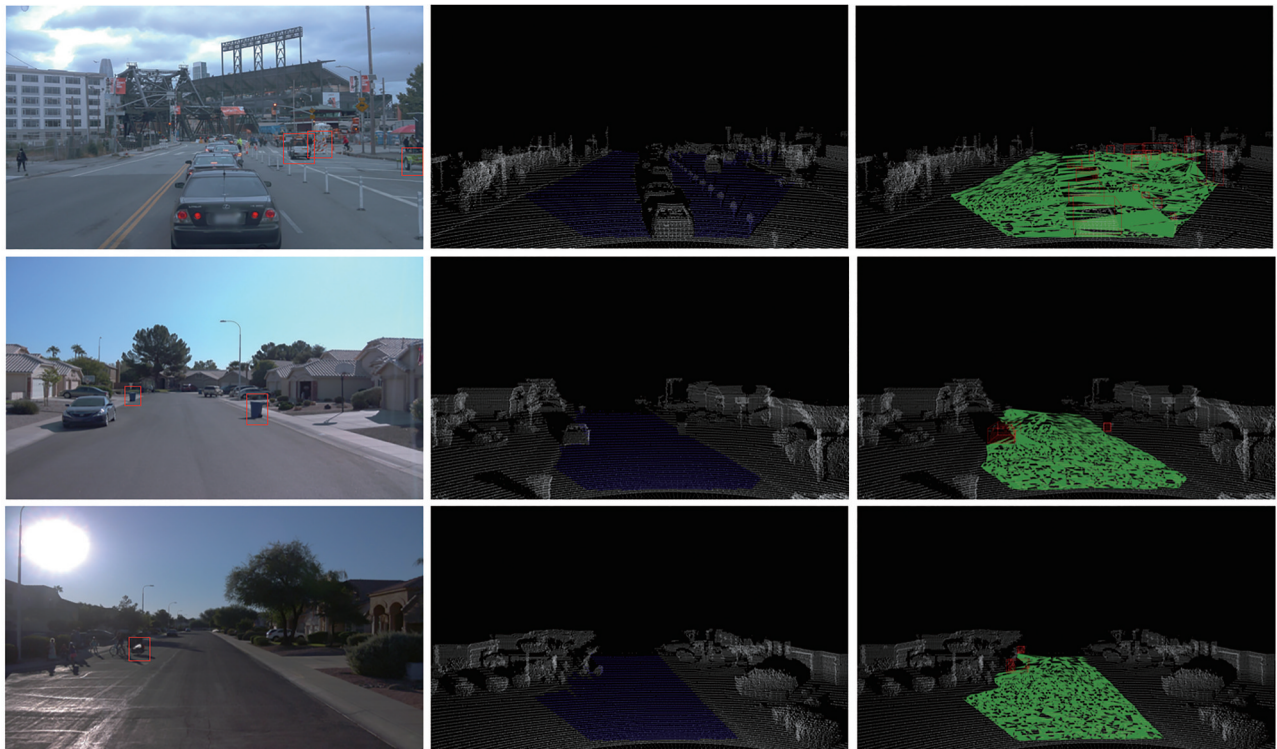


图9 点云聚类 and 道路重建渲染

图10显示了Waymo- Anomaly数据集被归类为存在异常案例场景的可视化。第1行是原始图像,第2行是在正常数据集中训练的检测器的检测结果。此时,检测器只能响应常规训练集中的已存在的对象。第3行是经过CCMP数据挖掘后投射回2D

维度的检测帧。通过比较和计算两排检测帧之间的IOU,可以消除常规目标,提取符合异常案例的目标,提取后的可视化如图10的最后1行所示。可以发现CCMP除可以挖掘常规类别之外的对象,例如手推车、婴儿车和其他对象。同时,它还可以挖掘当

前检测器不足以处理像素级异常案例的场景,例如图 10 中的阴雨和低光场景。由于照明或天气条件对图像产生了像素级损坏,导致常规 2D 检测器很难进行目标的准确捕捉,在这种情况下成功挖掘的主

要原因是文中使用在线检测结果与点云聚类结果进行比较。如果使用原始图像的真值框和点云聚类生成的提议框进行甄别,这种情况下的像素级的异常案例就很容易被忽略。



图 10 CCMP挖掘的异常案例效果图

### 3 结论

提出一种自动化异常案例挖掘流程 CCMP,用于解决自动驾驶现实世界存在的异常案例场景。CCMP 的输入是对多模态传感器数据,因为视觉传感器具有更丰富的语义信息,而激光雷达传感器可以减少漏检的可能性。在现实世界中,异常案例存在的场景往往属于长尾分布中的尾部数据,而基于深度学习的异常案例感知模型的性能决定了感知系统的安全性。目前,基于数据集解决此类问题的标准方法是收集大量数据,手动选择检测器性能不足的场景,然后进行多次迭代训练。但是,手工选择异常案例的场景是低效的。CCMP 可以自动挖掘角落案例场景,这对于大型公共或私有数据集非常有效。最后,本文使用 CCMP 方法构建 Waymo 子集 Waymo-Anomaly,并在训练异常案例的情况下进一步研究探索当前检测方法对长尾问题的有效性。

### 参考文献

- [1] 范丽丽,赵宏伟,赵浩宇,等.基于深度卷积神经网络的目标检测研究综述[J].光学精密工程,2020,28:1152-1164.  
FAN Lili, ZHAO Hongwei, ZHAO Haoyu, et al. Survey of target detection based on deep convolutional neural networks[J]. Optics and Precision Engineering, 2020, 28: 1152-1164.
- [2] 王海,蔡柏湘,蔡英凤,等.基于语义分割网络的路面积水与湿滑区域检测[J].汽车工程,2021,43(4):485-491.  
WANG Hai, CAI Baixiang, CAI Yingfeng, et al. Detection of water-covered and wet areas on road pavement based on semantic segmentation network [J]. Automotive Engineering, 2021, 43 (4): 485-491.
- [3] WONG K, WANG S, REN M, et al. Identifying unknown instances for autonomous driving[C]. Conference on Robot Learning. PMLR, 2020: 384-393.
- [4] DU X, WANG Z, CAI M, et al. VOS: learning what you don't know by virtual outlier synthesis [J]. arXiv preprint arXiv: 2202.01197, 2022.
- [5] WANG H, CHEN Y, CAI Y, et al. SFNet-N: an improved SF-Net algorithm for semantic segmentation of low-light autonomous driving road scenes [J]. IEEE Transactions on Intelligent Trans-

- portation Systems, 2022, 23(11): 21405–21417.
- [6] KENDALL A, BADRINARAYANAN V, CIPOLLA R. Bayesian SegNet: model uncertainty in deep convolutional encoder–decoder architectures for scene understanding[J]. arXiv preprint arXiv: 1511.02680, 2015.
- [7] GRCIC M, BEVANDIC P, KALAFATIC Z, et al. Dense anomaly detection by robust learning on synthetic negative data[J]. arXiv preprint arXiv:2112.12833, 2021.
- [8] OHGUSHI T, Horiguchi K, YAMANAKA M. Road obstacle detection method based on an autoencoder with semantic segmentation[C]. Proceedings of the Asian Conference on Computer Vision, 2020.
- [9] VOJIR T, ŠIPKA T, ALJUNDI R, et al. Road anomaly detection by partial image reconstruction with segmentation coupling[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 15651–15660.
- [10] YU F, CHEN H, WANG X, et al. BDD100K: a diverse driving dataset for heterogeneous multitask learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2636–2645.
- [11] SUN P, KRETZSCHMAR H, DOTIWALLA X, et al. Scalability in perception for autonomous driving: Waymo open dataset[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2446–2454.
- [12] ZHANG Y, KANG B, HOOI B, et al. Deep long-tailed learning: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
- [13] CHAN R, ROTTMANN M, GOTTSCHALK H. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 5128–5137.
- [14] BAUR C, WIESTLER B, ALBARQOUNI S, et al. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images[C]. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4. Springer International Publishing, 2019: 161–169.
- [15] BREITENSTEIN J, TERMÖHLEN J A, LIPINSKI D, et al. Corner cases for visual perception in automated driving: some guidance on detection approaches[J]. arXiv preprint arXiv: 2102.05897, 2021.
- [16] LI B, YAO Y, TAN J, et al. Equalized focal loss for dense long-tailed object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 6990–6999.
- [17] WANG T, ZHU Y, ZHAO C, et al. Adaptive class suppression loss for long-tail object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 3103–3112.
- [18] CONTRIBUTORS M M S. MMSegmentation: openMMLab semantic segmentation toolbox and benchmark [EB/OL]. <http://github.com/open-mmlab/mms Segmentation>, 2020.
- [19] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016: 630–645.
- [20] WANG C, DU P, WU H, et al. A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net[J]. Computers and Electronics in Agriculture, 2021, 189: 106373.
- [21] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381–395.
- [22] MOORE T, STOUCHE D. A generalized extended Kalman filter implementation for the robot operating system[C]. Intelligent Autonomous Systems 13: Proceedings of the 13th International Conference IAS-13. Springer International Publishing, 2016: 335–348.
- [23] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154–6162.
- [24] WOO S, DEBNATH S, HU R, et al. Convnext v2: co-designing and scaling convnets with masked autoencoders[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 16133–16142.
- [25] GE Z, LIU S, WANG F, et al. YOLOX: exceeding YOLO series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.
- [26] BIFFI L J, MITSHITA E, LIESENBERG V, et al. ATSS deep learning-based approach to detect apple fruits[J]. Remote Sensing, 2020, 13(1): 54.
- [27] SUN P, ZHANG R, JIANG Y, et al. Sparse R-CNN: end-to-end object detection with learnable proposals[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14454–14463.
- [28] GUPTA A, NARAYAN S, JOSEPH K J, et al. OW-DETR: open-world detection transformer[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 9235–9244.
- [29] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740–755.
- [30] CHEN K, WANG J, PANG J, et al. MMDetection: open mmlab detection toolbox and bench mark[J]. arXiv preprint arXiv: 1906.07155, 2019.
- [31] LI K, CHEN K, WANG H, et al. CODA: a real-world road corner case dataset for object detection in autonomous driving[C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 406–423.
- [32] ZHANG H, CHANG H, MA B, et al. Cascade RetinaNet: maintaining consistency for single-stage object detection[J]. arXiv preprint arXiv:1907.06881, 2019.
- [33] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440–1448.