

从像素到掩码：脱离分布的分割研究综述

WENJIE ZHAO, University of Texas at Dallas, USA

JIA LI, University of Texas at Dallas, USA

YUNHUI GUO, University of Texas at Dallas, USA

随着对人工智能安全问题的关注增加，分布外 (OoD) 检测和分割引起了越来越多的关注。传统的 OoD 检测方法能够识别 OoD 对象的存在，但缺乏空间定位能力，限制了它们在下游任务中的应用价值。OoD 分割通过在像素级粒度定位异常对象来解决这一限制。这一能力对于诸如自动驾驶等安全关键应用至关重要，其中感知模块不仅需要检测还需要精确分割 OoD 对象，以实现有针对性的控制措施并增强整体系统的鲁棒性。在本次综述中，我们将当前的 OoD 分割方法分为四类：(i) 测试时 OoD 分割，(ii) 用于监督训练的异常曝光，(iii) 基于重建的方法，(iv) 和利用强大模型的方法。我们系统性地回顾了自动驾驶场景中 OoD 分割的最新进展，识别了新出现的挑战，并讨论了未来有前途的研究方向。

CCS Concepts: • **Computing methodologies** → **Image segmentation; Scene anomaly detection; Vision for robotics; Computer vision.**

Additional Key Words and Phrases: Out-of-Distribution, Segmentation, Autonomous Driving

ACM Reference Format:

Wenjie Zhao, Jia Li, and Yunhui Guo. 2018. 从像素到掩码：脱离分布的分割研究综述. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 16 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 介绍

自动驾驶吸引了越来越多的关注，因为它有可能让人类摆脱驾驶。自动驾驶系统 (ADS) 通常包含多个模块，其中包括一个感知模块，该模块收集和处理环境数据以指导决策。深度神经网络在计算机视觉任务上的显著成功，使得将它们集成到 ADS 的感知模块中成为可能。虽然深度神经网络在标准视觉基准测试 [23] 中表现出显著成功，但将它们部署到自动驾驶系统中带来了比控制环境（如 ImageNet [41]）更大的挑战。ADS 在真实的环境中运行，经常遇到意料之外的情景。这些情景通常涉及训练过程中未出现的物体。

基础深度神经网络在遇到训练时未接触的物体时常表现出过度自信，导致将这些物体错误地归为已知类别 [3]。感知模块中的此类缺陷为自动驾驶系统引入了不可接受的安全风险。理想情况下，自动驾驶系统应在各种条件下表现出稳健性，并清晰识别其局限性，在遇到超出其能力范围的情景时提示驾驶员或采用更保守的策略 [39]。为实现这一目标，模型必须准确识别其无法可靠处理的输入。换句话说，它应有效检测训练期间未见分布的数据。OoD 检测的概念最初由 Hendrycks 在图像分类环境中提出，后来扩展到语义分割，这对于自动驾驶系统至关重要，因为它要求对像素级异常进行定位。[17]

Authors' Contact Information: Wenjie Zhao, wxz220013@utdallas.edu, University of Texas at Dallas, Richardson, Texas, USA; Jia Li, jxl220096@utdallas.edu, University of Texas at Dallas, Richardson, Texas, USA; Yunhui Guo, yunhui.guo@utdallas.edu, University of Texas at Dallas, Richardson, Texas, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Manuscript submitted to ACM

1

OoD 分割自然地 OoD 检测问题中演变而来。早期的研究将现有的 OoD 检测方法适配到分割任务中，并取得了可喜的成果。例如，Rottmann 等人 [40] 通过聚合多种离散测量提出了一种 OoD 分割方法。一些研究人员也尝试通过训练在异常暴露数据集上提高模型对 OoD 数据的敏感性 [1, 2, 4, 13, 32, 45]。还有一些人探索了基于重建的方法，试图通过将重建的图像与原始输入进行比较来识别异常 [6, 8, 15, 29, 46, 47, 53]。最近，利用强大的预训练模型进行 OoD 分割的方法越来越受到关注。尽管该领域的研究增长迅速，但对最近进展的全面系统回顾仍然缺失。为填补这一区缺口并促进未来的研究，我们对自动驾驶场景中的 OoD 分割方法进行了综述。

在这项调查中，我们首先在第 §2 节中描述了自动驾驶中的分布外分割问题设置。然后，我们在第 §3 节中研究数据集和指标。此外，我们在第 §4 节中回顾了自动驾驶中的分布外分割方法，这包括四个类别。最后，我们在第 §?? 节中回顾潜在挑战并概述自动驾驶分布外分割领域的未来方向。

2 问题设置/问题回顾

给定一个从分布内环境 $P_{in}(X, Y)$ 中抽取的带标签训练数据集，每个图像 $X \in \mathbb{R}^{H \times W \times C}$ 与一个像素级别的掩码 $Y \in \mathcal{C}_{in}^{H \times W}$ 配对，其中 $\mathcal{C}_{in} = \{1, \dots, K\}$

表示已知类别的集合。在该数据上训练的模型 f_θ 处理输入图像 X ，并为每个像素 $x \in X$ 生成一个 logits 向量 $z(x) \in \mathbb{R}^K$ 。然后将这些 logits 转换为类别概率。在闭集假设下，期望模型能够正确地将图像分割到已知标签空间中。然而，在实际部署中，系统可能会遇到来自未知分布 $P_{out}(X, Y)$ 的输入，其中包含的像素的真实类别属于满足 $\mathcal{C}_{in} \cap \mathcal{C}_{out} = \emptyset$ 的不相交集合 $\mathcal{C}_{out}$ 。

像素级 OoD 分割在单个像素的粒度上区分 ID 和 OoD 区域，生成一个二值掩码 $M \in \{0, 1\}^{H \times W}$ ，该掩码将每个像素标记为 ID (0) 或 OoD (1)：

$$M_{h,w} = \begin{cases} 0, & \text{if pixel } (h, w) \in X \text{ belongs to ID region,} \\ [4pt] 1, & \text{if pixel } (h, w) \in X \text{ belongs to OoD region.} \end{cases}$$

3 基准测试

在本节中，我们回顾了自动驾驶中广泛采用的 OoD 分割数据集。同时，我们也介绍了常用的评估指标。

3.1 数据集

LostAndFound 由 Pinggera 等人于 2016 年引入。它是第一个专门用于评估小型障碍物分割的数据集，这是自动驾驶安全的关键任务。数据集由 13 个具有挑战性的街道场景中拍摄的真实世界图像组成，包括 37 种不同类型的障碍物，其大小和材质各异。该数据集包含 2,104 张分辨率为 2048×1024 像素的图像，其中 1,036 张用于训练/验证，1,068 张用于测试。此数据集的注释

Manuscript submitted to ACM

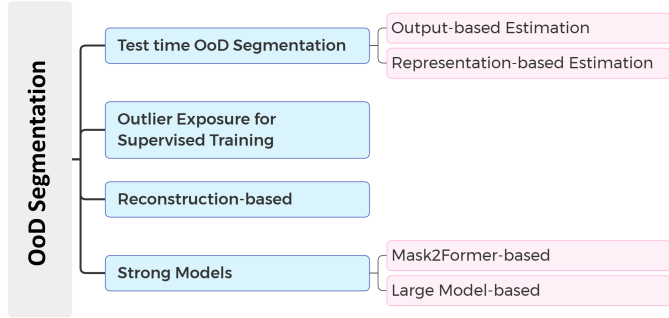


Fig. 1. 自动驾驶中 OoD 分割方法的分类法。

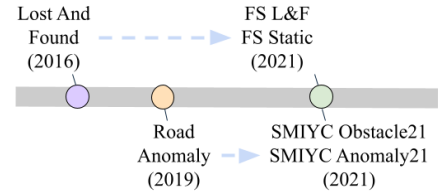


Fig. 2. 用于 OoD 分割的数据集的时间顺序发展。

Dataset	Resolution	Val / Test	Data Source	Evaluation Area
LostAndFound	2048 × 1024	1036 / 1068	Captured	Road-only
RoadAnomaly	1024 × 512	60 / -	Collected	Road-context
FS Static	2048 × 1024	30 / 1000	Synthetic	Road-context
FS L & F	2048 × 1024	100 / 275	Collected	Road-context
SMIYC Anomaly	2048 × 1024 or 1280 × 720	10 / 100	Collected	Road-context
SMIYC Obstacle	1920 × 1080	30 / 327	Collected and Captured	Road-only

Table 1. 自动驾驶中的代表性 OoD 分割数据集总结。每个数据集的特点包括其分辨率、验证/测试划分、数据来源和评估区域。

仅关注路面区域，包括障碍物和空闲路面空间的掩码。值得注意的是，建筑物和路边结构没有进行标注。

RoadAnomaly 由 Lis 等人于 2019 年提出。它旨在支持检测道路场景中的意外物体。它包括 60 张在线收集的真正世界图像，包含异常物体如动物、石头、轮胎、垃圾、罐头和建筑设备。这些图像具有 1280 × 720 像素的统一分辨率，并被调整为 1024 × 512 像素进行评估。注释涵盖了路面及其周围的环境。与仅专注于道路区域内的小型道路障碍物的 LostAndFound 数据集相比，RoadAnomaly 数据集涵盖了各种尺度的异常。此外，它将感兴趣区域扩展到了道路以外，包括路边的环境。

Fishyscapes 由 Blum 等人于 2021 年提出，用于异常分割。它由三个子集组成，分别是 FS Static、FS Web 和 FS LostFound。FS Static 提供了一个 30 幅图像的公共验证集和一个 1,000 幅图像的私有测试集，而 FS LostFound 提供了一个可访问的 100 幅图像的验证集和 275 幅图像的私有测试集。所有验证图像的分辨率为 2048 × 1024 像素。FS Static 和 FS Web 共享相同的构建方法。作者将来自 Pascal VOC 的异常对象（如动物和家居用品）覆盖在来自 Cityscapes 验证集的背景图像上。这些异常对象经过精心选择，以确保它们的语义标签和 Cityscapes 的标签空间不重合，从而代表真正的分布外内容。对象随机缩放和定位，哺乳动物更可能出现在下半部分，而鸟类或飞机更可能出现在上半部分。FS Web 的构建方式类似于 FS Static，但使用从互联网自动收集的异常对象。该数据集设计为可持续更新。然而，自 2022 年 8 月以来没有更多更新发布。为了进一步评估在真实世界图像上的性能，作者基于原始的 LostAndFound 数据集提出了 FS LostFound。此子集为路边区域添加了像素级别的注释，并过滤掉在 Cityscapes 中出现的物体，因为它们不应该属于分布外实例。为了减少冗余，作者对高度重复的序列进行了子采样，并排除了没有对象的图像。

SegmentMeIfYouCan (SMIYC) 是由 Chan 等人于 2021 年提出的，包括 RoadAnomaly21 和 RoadObstacle21。RoadAnomaly21 提供了一个可访问的验证集，包括 10 张图像，以及一个私人测试集，包括 100 张图像，图像分辨率为 1280 × 720 或者 2048 × 1024。RoadObstacle21 提供了一个可访问的验证集，包括 30 张图像，以及一个私人测试集，包括 327 张图像，分辨率为 1920 × 1080。RoadAnomaly21 专注于完整街道场景中的通用异常分割，并提供了一个包含像素级注释的 100 张图像评估集。该数据集通过去除低质量图像和标注错误，介绍了一种用于模糊区域的空集类别，并加入了 68 张新收集的图像，扩展了 RoadAnomaly 数据集 [29]。每张图像至少包含一个异常对象，且异常可能出现在跨越多种环境的场景中的任何地方。RoadObstacle21 专注于那些障碍物直接出现在路面上并对自动驾驶车辆构成直接威胁的安全关键场景。为了包容各种条件，这些图像捕捉了不同的道路类型、光照条件和天气场景。RoadAnomaly21 和 RoadObstacle21 都使用三种标注类别，即异常或障碍物、非异常或非障碍物以及空。

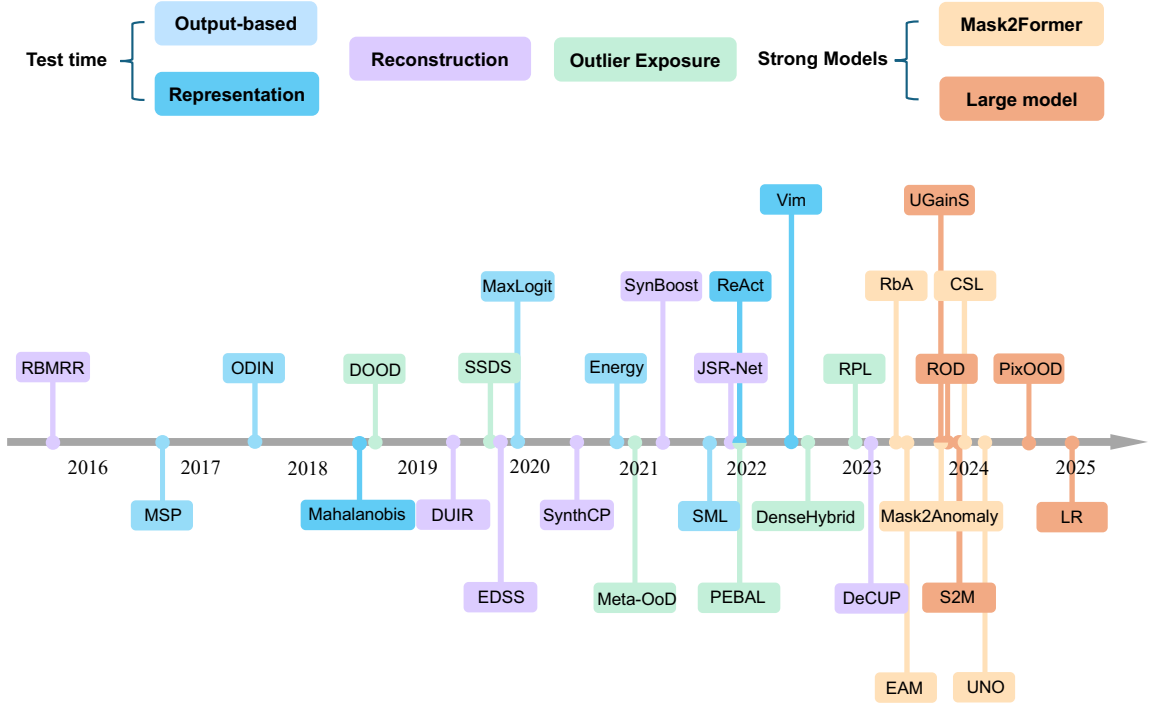


Fig. 3. 提到的方法的时间轴。

用于评估 OoD 分割的两个常用指标是 95% 真实阳性率 (True Positive Rate, TPR) 处的假阳性率 (False Positive Rate, FPR95) 和平均 F1 分数。

FPR95 度量的是当 OoD 像素的召回率达到 95 % 时, 在分布内像素上的误报率。它反映了在高召回率设置下, 一个模型在多大程度上避免将分布内像素误分类为 OoD。其定义为:

其中 $\hat{P}_{\text{out}}$ 表示阈值 δ 下预测的 OoD 像素。

为了补充像素级评估, 平均 F1 分数通过比较预测的和真实的 OoD 区域提供了一个组件级的视角。它强调整个 OoD 对象是否被成功检测。阈值 τ 处的 F1 分数计算方式为:

$$F1(\tau) = \frac{2 \cdot TP(\tau)}{2 \cdot TP(\tau) + FP(\tau) + FN(\tau)}. \quad (1)$$

然后在多个阈值上取平均以获得最终分数。

4 OoD 分割方法

在本节中, 我们根据每项工作的主要贡献将现有的 OoD 分割方法分为四组。

4.1 测试时的外部数据分割

其中一些方法最初是为图像级 OoD 检测开发的, 后来被扩展用于解决像素级分割任务。更具体地说, 图像分割是通过将类别标签分配给每一个像素来执行的。这些方法对 OoD 分割的适应是通过在像素级计算异常分数来实现的。这种像素级处理能够精准识别分布外区域, 这在自动驾驶中尤为有用。

4.1.1 基于输出的估计。基于输出的方法 [16, 17, 21, 28] 最初由 Hendrycks 和 Gimpel 引入，同时也提出了 OoD 检测问题的公式化 [17]。这些方法依赖于模型的输出来计算异常得分，然后用该得分来识别 OoD 输入。

Hendrycks 等人 [17] 使用模型输出的最大软概率 (MSP) 作为 OoD 分数。核心思想是，ID 输入应该具有较高的置信度分数，因为模型应该能够正确分类它们。对于 OoD 输入，模型应表现出较低的置信度，因为它无法正确分类。这种方法成功区分了 ID 和 OoD 输入，并作为基准。然而，它在大规模数据集上失败，因为由于存在许多相似元素，概率会被稀释。因此，对于某些 ID 像素，MSP 分数可能相对较低，但对于与任何分布内类别不相似的 OoD 区域则可能相对较高。

最大 Logit 值 (MaxLogit) 被提出用于利用 logit 值作为 OoD 分数 [16]。其动机在于 MSP 在 softmax 分布中的类似类别之间会产生竞争。具体来说，logit 值可能在类似的 ID 类别之间一直很高，例如不同类型的汽车，这导致每个类别的概率都很低。虽然 OoD 输入相对于 ID 可能具有较小的最大 logit 值，但最高 logit 值仍可能显著大于所有其他类别。因此，OoD 输入显示出比某些 ID 输入更高的置信分数。MaxLogit 避免了这种竞争，并使 ID 和 OoD 输入之间的比较成为可能。该方法的一个挑战是 logit 分布在不同类别之间有所不同，这降低了其性能。标准化最大 Logits (SML) [21] 被提出以通过在类别之间标准化最大 logit 来解决此问题，从而可以更准确地比较类别分布。

另一个重要进展是 ODIN [28]，它结合了温度缩放和小的输入扰动，以更好地分离分布内和分布外的样本。温度缩放通过使预测概率更柔和来校准 softmax 输出，而输入扰动则旨在有选择地增加分布内输入的 softmax 分数。由于更好的校准和对输入的敏感性，ODIN 有效地扩大了分布内和分布外样本之间的 softmax 分数差距。尽管这些基于输出的方法很有效，但它们仍然面临过度自信的问题。神经网络倾向于对即使是分布外的输入也产生高置信度的预测。这种过度自信来源于模型的 logit 输出。因此，分布外的输入可能仍会得到高置信度的分数，导致错误的识别。

Liu 等人提出了能量分数 [30]，它利用了所有类别的 logits。该分数通过测量输入的总能量来反映输入与训练分布的接近程度，并有助于缓解过度自信的影响。它提供了一个更为稳健的置信估计，并已被证明优于之前基于 softmax 的分数。

Method	Idea
MSP [17]	$s_{\text{MSP}}(x) = \max_{1 \leq c \leq K} \frac{e^{z_c(x)}}{\sum_{j=1}^K e^{z_j(x)}}$
MaxLogit [16]	$s_{\text{MaxLogit}}(x) = \max_{1 \leq c \leq K} z_c(x)$
SML [21]	$s_{\text{SML}}(x) = \text{Smooth}\left(\text{BoundarySuppress}\left(\max_{1 \leq c \leq K} \frac{z_c(x) - \mu_c}{\sigma_c}\right)\right)$
ODIN [28]	$x' = x - \epsilon \cdot \text{sign}\left(-\nabla_x \log\left(\max_{1 \leq c \leq K} \frac{\exp(z_c(x)/T)}{\sum_{j=1}^K \exp(z_j(x)/T)}\right)\right)$
Energy Score [30]	$s_{\text{Energy}}(x) = -T \log\left(\sum_{c=1}^K \exp\left(\frac{z_c(x)}{T}\right)\right)$

Table 2. 基于输出的 OoD 检测方法总结。 $z(x) \in \mathbb{R}^K$ 是输入 x 的 logits。对于 ODIN，我们提供获取预处理图像的公式。

这些基于输出的方法很容易集成到任何模型中，使它们在自动驾驶中的异类检测分割得到了广泛采用。此外，它们在操作时无需访问模型参数或训练数据，这在自动驾驶系统的感知模块由第三方提供商供应时特别有价值。然而，这些方法主要利用模型输出，可能不足以可靠地区分 ID 和异类样本。

4.1.2 基于表示的估计 与仅依赖于模型输出不同, 基于表示的方法探索神经网络的中间特征以检测异常分布样本。这一类别中的一种代表性方法是由 [26] 提出的 Mahalanobis 距离方法, 其有效性已在 [19] 的自动驾驶数据集上得到了验证。其理念是使用多变量高斯分布来对每一类的特征分布建模。在测试期间, 计算样本特征与最近类别分布之间的 Mahalanobis 距离。那些远离所有类别分布的样本更有可能是异常分布 (OoD)。这种方法利用了特征空间中包含的信息。通过分析特征如何偏离已知的训练分布, 它比单独依赖最终输出能够更准确地进行检测。然而, 这种方法需要访问训练集以计算特征偏差, 这在训练数据是私密的或不可访问时限制了其适用性。

其他方法不是直接从特征空间中提取测量值, 而是专注于修改特征空间以增强基于输出的 OoD 分数。例如, Sun 等人提出了 ReAct [44], 它通过设置阈值 c 来修正倒数第二层的激活。具体来说, 每个神经元的输出在 c 处被截断, 取神经元原始激活值和阈值之间的较小值。一个理想的 c 应足以保留 ID 数据的激活模式, 同时修正 OoD 数据的激活模式。通过截断过高的激活, 修正后的激活可以限制噪声的影响, 使整体激活模式更接近正常表现情况。这种方法有效地减少了神经网络中常见的过度自信问题。此外, ReAct 可以与最常用的 OoD 分数集成, 使其易于应用于自动驾驶的 OoD 分割。然而, 截断也会改变 ID 的激活, 并可能降低 ID 的准确性。在像自动驾驶这样的安全关键系统中, 即使是细微的下降也很重要。因此, ReAct 在部署之前需要额外的验证。

虽然基于表示的方法利用了特征空间中的丰富信息, 但它们并不能保证对所有输入进行最佳的 OoD 检测。一些 OoD 样本在特征空间中易于识别, 而另一些在 logit 空间中更容易识别。为了有效结合这两种视角, Wang 等人引入了 ViM [50]。它生成一个额外的虚拟 logit, 代表从特征残差中派生的合成 OoD 类。然后, 这个虚拟 logit 被缩放到与原始 logits 相匹配的幅度, 以确保兼容性。在推理过程中, 与此虚拟类对应的概率作为有效的 OoD 得分。这些方法利用了深层特征中嵌入的丰富区分信息, 且可以无需重新训练骨干网络进行部署。然而, 它们通常需要访问训练统计数据, 并涉及如阈值 c 等超参数, 这使得部署复杂化。

测试时间 OoD 分割方法的实验结果呈现在表 6 的第一部分。这些方法引入了各种 OoD 评分函数, 这些函数已被广泛采用, 现在成为大多数后续 OoD 分割技术的基础。

4.2 监督训练的异常值暴露

在既定假设下, 模型在训练过程中预期不会接触到 Out-of-Distribution (OoD) 的数据。然而, 训练是放大 ID 和 OoD 数据之间区别的直观方法。为了使 OoD 检测训练切实可行, Hendrycks 等人提出了异常暴露 (Outlier Exposure, OE) [18], 这包含一个与测试过程中可能出现的真实 OoD 数据不重叠的异常数据的辅助数据集。他们的损失函数旨在保持原始任务目标, 同时鼓励模型对异常样本赋予低置信度分数。异常暴露显著提高了各项任务上的 OoD 检测性能。图像级异常暴露的优秀表现表明类似的策略可能也有利于自动驾驶应用中的逐像素 OoD 分割。图 4 展示了用于改善 OoD 分割的 OE 过程的生成数据集。

Bevandic 等人将像素级的 OoD 检测视为一个二分类任务 [1]。我们将这种方法称为 DOOD。为了训练判别模型, 他们将外部样本引入到语义分割数据集中。他们选择一个道路驾驶数据集, 例如 Cityscapes, 作为 ID 数据, 这些图像中的每个像素都标记为 ID。然后, 他们将来自 ImageNet 的物体粘贴在 ID 图像上作为 OoD 像素。这个过程生成了一个综合的合成数据集, ID 和 OoD 区域共存于同一帧中, 使监督训练成为可能。使用这个合成数据集, 他们训练了一个专用的 OoD 判别模型, 以识别每个像素是属于 ID 还是 OoD 类别。实验结果表明, 这种监督训练使检测器能显著优于基于 softmax 的基线。然而, 专用的 OoD 分割模型仅关注检测 OoD 像素, 而忽视了分割任务。

在其早期的判别检测器之后, Bevandic 等人推出了一种统一的架构, 该架构在一次前向传递中输出语义分割结果和二值 OoD 映射 [2], 我们将这种方法称为 SSDS。他们在标准分割骨干上附加了一个轻

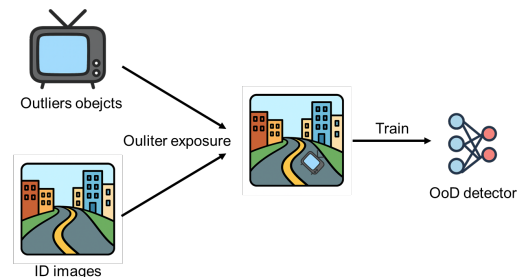


Fig. 4. 在监督训练中引入离群样本的动机。离群样本暴露通过将离群数据集中的物体粘贴到 ID 图像中来引入代理的 OoD 样本。然后可以使用生成的图像进行监督训练, 以改善 OoD 分割性能。

量级的异常检测头。这个额外的检测头是在一个异常数据集上进行训练的，使模型能够识别属于 OOMD 区域的像素。尽管报告的结果是有前景的，但多任务训练可能会降低分割精度。此外，OE 数据集中异常高的异常像素出现频率可能会导致异常频率偏差。这种偏差可能会阻碍在自动驾驶场景中的部署。

同样地，Chan 等人在 Meta-OoD [4] 中引入了一种熵最大化损失，应用于通过异常暴露生成的合成数据集。他们将一个元分类器作为后处理步骤，用于过滤掉误报的 OoD 样本。该元分类器使用基于从 softmax 概率导出的像素级不确定性度量计算的手工衡量标准来识别错误的 OoD 分割。实验结果表明，这种两阶段方法有效地降低了 OoD 检测错误。此外，作者还探讨了对原始语义分割任务进行重新训练的影响。他们的研究表明，重新训练导致分割准确性可测量但边际的下降。此外，meta-OoD 倾向于生成碎片化的异常掩码，其中许多误报的像素被过滤掉。它依赖于对整个网络进行重新训练以进行异常暴露，这效率低下，并且有偏向于异常值的预测的风险。

为了将异常检测与不确定性解耦并减少计算成本，Tian 等人提出了像素级能源倾斜弃权学习 (PEBAL) [45]。PEBAL 增加了一个像素级异常类，当一个像素不匹配任何内在类别时，模型可以选择该类。它联合优化了像素级弃权学习头和基于能量的模型。能源模型被训练用于给来自异常数据集的异常像素分配高能量，这分解了具有高不确定性的分布外像素。具有高能量的像素会受到更大的惩罚，因此像素级弃权学习更倾向于将它们分类为异常。此外，平滑损失鼓励邻近的异常像素保持一致，而稀疏损失减少了误报检测。由于仅微调了主干的最后一个模块，该方法对于自动驾驶系统来说是轻量级的。各种数据集上的实验表明，PEBAL 在减少误报的情况下改善了 OoD 分割，比 Meta-OoD 更有效。然而，重新训练过程牺牲了 ID 分割的准确性，这对于自动驾驶至关重要。

为了更好地利用判别方法和密度估计方法，DenseHybrid [13] 将这两种观点结合到一个混合异常检测器中。该模型在 OE 数据集上进行训练，其损失函数由三个目标组成：分类损失 L_{cls} 、未归一化数据似然损失 L_x 和数据集后验损失 L_d 。他们提出最大化在分布内像素的似然，并最小化分布外像素的似然。数据集后验预测 $P(d_{out}|x)$ 代表了一个像素属于分布外的概率。作者基于分布外数据集后验预测 $P(d_{out}|x)$ 和似然预测 $p(x)$ 定义了混合异常检测器 $s(x)$ 为：

$$s(x) \cong \ln P(d_{out}|x) - \ln \hat{p}(x).$$

然后，异常分数图与基于阈值的语义分割输出进行融合。DenseHybrid 在以较小的分布内分割精度损失的情况下，实现了更好的分布外分割。尽管如此，该方法仍然引入了轻微分布内分割性能下降，这可能会影响其在自动驾驶中的适用性。

一个关键问题是，OoD 检测器可能会在环境协变量变化（例如场景从城市变为农村环境）时错误地将高 OoD 分数分配给分布内的像素。Liu 等人提出了残差模式学习 (RPL) [32]，以充分减少再训练对分布内分割的影响，并能够在各种环境下进行泛化。他们引入了一个外部模块 RPL，附加到冻结的封闭集分割网络上作为 OoD 检测器。基础分割网络表示为：

$$\tilde{y} = f_{\phi_{seg}} \left(f_{\phi_{aspp}} \left(f_{\phi_{fcn}}(x) \right) \right).$$

RPL 接受中间特征作为输入，以学习异常的残差模式。输出与分割特征融合以在分布外区域引入高不确定性。修改后的网络表示为：

$$\hat{y} = f_{\phi_{seg}} \left(f_{\phi_{aspp}} \left(f_{\phi_{fcn}}(x) \right) + f_{\theta_{rpl}} \left(f_{\phi_{fcn}}(x) \right) \right).$$

。通过冻结分割模型来保证分布内分割的准确性。为了增强模型对环境与分布外对象之间关系的理解，作者们引入了上下文鲁棒对比学习（CoroCL），其目标是使同一分布的嵌入更加相似，不同分布的嵌入更加不同。在推理时，分布内分割结果从原始网络分支 $\hat{y}$ 获得，而分布外分割结果 $\hat{y}$ 由 RPL 模块计算。RPL [32] 在不牺牲分布内分割准确性的情况下展示了强大的分布外分割性能。然而，这两个输出图有时可能不一致。具体来说，异常图上检测到的分布外区域可能与分割图中的任何物体不对应，这可能会在自动驾驶系统的决策模块中引起混淆。

Method	Year	ID influence	Retrain backbone	Key idea
DOOD [1]	2018	-	☒	Discriminative binary classifier on synthetic ID+OoD mix
SSDS [2]	2019	是	☒	Extra detection head joint-trained with seg. network
Meta-OoD [4]	2021	是	☒	Entropy-max + meta-classifier to filter FP
PEBAL [45]	2022	是	☒	Energy-biased abstention; sparse + smooth losses
DenseHybrid [13]	2022	是	☒	Likelihood + posterior hybrid anomaly score
RPL [32]	2023	否	∧	External residual pattern head + CoroCL

Table 3. 基于 OE 的监督 OoD 分割方法总结。ID 影响列表明训练是否影响 ID 分割，而在重训练主干列中，☒、☒和 ∧ 分别表示微调大多数主干层、仅更新最后一个块以及保持主干冻结。

在 OE 的合成数据集上训练显著提高了 OoD 分割性能。然而，所使用的异常数据集在异常曝光过程中能否准确模拟现实驾驶场景中出现的 OoD 对象仍然存在疑问。此外，牺牲分布内分割的准确性在感知模块中可能是不可接受的。即使模型产生两个独立的输出，一个用于分布内分割，另一个用于 OoD 分割，这些输出之间的冲突可能会让自动驾驶系统中的决策模块感到困惑。

表 3 提供了上述基于 OE 的监督 OoD 分割方法的详细比较。基于异常曝光的监督训练的实验结果展示在表 6 的第二部分。虽然这些方法显著提高了性能，但利用异常数据集作为现实世界 OoD 数据的代理引入了一个关于 OoD 样本的强假设。在实践中，现实世界的 OoD 数据是不可预测的。引入 OE 数据集可能会限制其在实际场景中的泛化能力。

4.3 基于重建的 OoD 分割

一个良好训练的模型应该能够重构原始图像。重构图像与原始图像之间的差异可能表明道路障碍物的潜在区域。基于这个思路，以重构为基础的方法已经成为在自动驾驶场景中进行 OoD 分割的另一主流方法。这些方法通过将输入图像与重构版本进行比较来分割异常，重构版本表示期望的分布内外观，如图 5 所示。这两张图像之间的差异更可能对应异常区域。

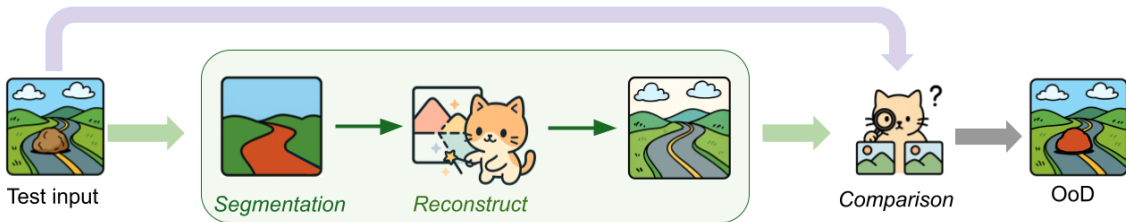


Fig. 5. 重建基础的 OoD 分割的动机。OoD 对象可能会被分割器忽略，从而在重建中被遗漏。对比原始图像和重建图像可以实现 OoD 分割。

Creusot 等人使用训练好的限制玻尔兹曼机在 [6] 中重建图像。我们将此工作称为 RBMRR。重建结果可能会遗漏背景中的小障碍物。例如，即使输入图像包含道路上的障碍物，模型可能只重建路面。重建图像与原始图像之间的差异用于进行异常分割。这里用于重建的输入包括归一化的图像块。重建向量 x' 计算为

$$x'_i = \text{Sigmoid}(x_i \cdot W + b_{\text{hid}}) \cdot W^T + b_{\text{vis}},$$

，其中 W 、 b_{hid} 和 b_{vis} 是限制玻尔兹曼机的参数。该工作验证了基于重建的 OoD 分割的可行性，并在纹理为主的合成数据集和高速公路场景的真实世界图像上展示了其有效性。然而，重建图像的质量较低限制了检测性能。基于图像块的重建也固有限制了模型的适应能力，导致在检测较大物体时表现不佳。

随着语义分割的发展，网络可以为输入图像生成精确的语义图。然而，道路上的障碍物可能会被分割网络忽略，这激发了 Lis 等人 [29] 从预测的语义图中重建原始图像。这种方法在文献中称为 DUIR。他们提出基于语义预测重新合成原始图像，其中分割错误会导致重新合成的图像与原始图像之间存在显著差异。为实现这一点，他们采用 pix2pixHD [51]，一种在分布内数据集上训练的条件 GAN [12]，以语义图作为条件生成图像。在他们的不一致网络中，使用预训练的 VGG 网络从原始图像和重新合成的图像中提取特征，而一个独立的 CNN 处理预测标签的 one-hot 表示。在特征金字塔的每个级别，这些特征被连接并通过 1×1 卷积。此外，计算并结合原始与重新合成图像之间特征的点对点相关性。所有融合的特征和相关性随后被输入到上卷积解码器中，以生成最终的不一致得分。

同样地，Xia 等人提出了一个统一的框架 SynthCP，通过一种合成和比较策略来解决语义分割中的失败和异常检测问题 [53]。一个语义到图像的条件生成对抗网络 [36] 在分布内数据上进行训练，以从预测的分割结果重建输入图像。然后，原始图像和合成图像都被输入到比较模块中，该模块计算从分割模型的最后一层在每个像素处提取的对应特征向量之间的余弦距离。为了克服分割模型中固有的误分类，作者使用最大 softmax 概率来优化比较结果 $\hat{c}_n^{(i)}$ ：

$$\hat{c}_n^{(i)} \leftarrow \hat{c}_n^{(i)} \cdot \mathbb{I}\{p^{(i)} \leq t\} + (1 - p^{(i)}) \cdot \mathbb{I}\{p^{(i)} > t\},$$

其中 $p^{(i)}$ 是第 i 个像素处的最大 softmax 概率， $t \in [0, 1]$ 是一个阈值并且 $\mathbb{I}$ 表示指示函数。这个后处理过滤掉了分割预测具有高置信度的像素。此方法依赖于原始和重建图像之间的数值差异。然而，它缺乏对语义区域的全面理解。

虽然原始图像和重构图像之间的差异通常表明是分布外 (OoD) 区域，但像素级比较可能不是衡量语义差异的有效方法。例如，两张亮度差异很大的图像可能会产生高像素级差异。然而，它们在语义层面上对人眼非常相似。在通过 GAN 生成的图像中会出现类似的问题，这种变化可能导致误报。Haldimann 等人通过提出一种基于 CNN 的不相似性检测器来估计原始和重构图像之间的语义不一致性，从而识别并解决了这一限制。我们将这项工作命名为 EDSS。他们首先使用 pix2pix 从预测的语义图重合成 RGB 图像，pix2pix 是一个在分布内数据集上训练的条件 GAN。为了在语义层面上测量差异，他们训练了一个不相似性检测器，用于比较原始图像和重构图像。它使用基于 VGG 的特征提取，并通过三元组损失进行优化，该损失对不匹配的图块赋予更高的不相似性，并对匹配的对赋予较低的值。损失被定义为

$$\mathcal{L}(D) = \lambda_D \mathbb{E}_{t_i} [\log D(p_i^+)] + \mathbb{E}_{t_i} [\log(1 - D(p_i^-))],$$

，其中 D 是不相似性检测器， p_i^+ 和 p_i^- 分别表示正负图块对。不相似性图中的更高异常分数意味着一个像素更可能属于一个分布外的物体。该方法利用超越简单像素级差异的语义信息。然而，它仍然局限于比较原始图像与重合成的图像。

为了更好地利用语义信息进行分布外分割，Vojir 等人提出了 JSR-Net [47]，其中包括一个分割耦合模块，用于融合来自分割图和重构误差的信息。作者没有使用 GAN，而是采用轻量级瓶颈解码器来重构道路区域。原始图像与重构图像的相似性由 SSIM [52] 衡量。重构模块经过训练，使道路区域的重构图像与原始图像之间的

相似性最大化，而异常区域的相似性最小化。分割耦合模块首先将语义分割概率与重构误差串联。然后使用两个卷积块处理该输入，以预测一个二元掩码，用于区分道路区域与异常区域。分割耦合模块使用二元交叉熵损失 $\mathcal{L}_{xent}$ 进行优化。最终的损失函数定义为

$$\mathcal{L} = \mathcal{L}_{xent} + 0.5\mathcal{L}_R,$$

，其中 $\mathcal{L}_R$ 是重构损失。这个方法减少了计算成本并提高了分布外分割性能。然而，它仍然在细小结构上遭受误判，并且对图像质量下降敏感。作者通过在 DaCUP [46] 中引入几个微妙的组件扩展了这项工作。为了更好地利用数据并捕捉道路表面的外观，作者注入了一个通过三元组损失优化的嵌入网络。此外，它设计了在嵌入空间中的基于距离的评分机制和一个修补模块，以减低误判率。这些组件提高了分布外分割性能。然而，整个流程复杂，包含多个模块。

基于重构的方法可能在分割结果嘈杂且碎片化时失效。为了更好地利用基于重构的方法和多样化的 OoD 分数，Biase 等人提出了 SynBoost [8]，这是一种新颖的框架，将这两个组件结合起来以产生更稳健的 OoD 分割。它将重构方法与不确定性测量相结合：softmax 熵 [10, 24]、softmax 差异 [40] 和感知差异 [9, 20]。它使用 CC-FPSE [31]，一个条件 GAN，从预测的语义分割图重构原始图像。softmax 熵和 softmax 差异是从分割网络的输出计算得到的。为了避免合成图像中的低层次纹理差异，例如汽车颜色的差异，作者计算相应像素 x 和 r 之间的感知差异如下所示：

$$V(x, r) = \sum_{i=1}^N \frac{1}{M_i} \|F^{(i)}(x) - F^{(i)}(r)\|_1,$$

，其中 $F^{(i)}$ 表示 VGG 网络的 i 层，包含 M_i 元素， N 是层的数量 [45]。这些输入然后被传递给一个差异模型，以预测分布外对象的掩码。该模型使用 VGG 提取原始和合成图像的特征，以及一个 CNN 处理语义预测和不确定性图。然后，来自输入图像、重构结果和分割输出的特征在每个层级串联起来，并通过 1×1 卷积进行比较。之后，得到的特征图将与不确定性图通过点相关融合。最后，一个解码器利用跨层次的融合特征以及语义图生成最终的 OoD 分割预测。

表格 4 提供了本节中讨论的基于重建的 OoD 分割流程的简要比较。表格 6 的第三部分展示了基于重建方法的实验结果。这些方法曾在一段时间内广受欢迎，但在社区中的关注度有所下降。

Method	Generator Model	Comparison Method	Feature Extraction	Gen. Training	Semantic Info in Comparison
RBMRR [6]	Restricted Boltzmann Machine	Pixel-wise diff.	N	P	N
DUIR [29]	pix2pixHD (cGAN)	CNN discrepancy net	Y	P	Y
SynthCP [53]	SPADE (cGAN)	Cosine + threshold	Y	P	N
EDSS [15]	pix2pix (cGAN)	Dissimilarity detector	Y	P	N
JSR-Net [47]	Bottleneck decoder	SSIM + coupling	N	J	Y
DaCUP [46]	Bottleneck decoder	SSIM + Embedding-space dist.	Y	J	Y
SynBoost [8]	CC-FPSE (cGAN)	Decoder blocks	Y	O	Y

Table 4. 重建基础的 OoD 分割流程比较。特征提取指示比较模块是否依赖于特征提取器。生成器训练总结了生成器的处理方式：O——来自其他工作的现成权重；P——先在 ID 数据上预训练然后冻结用于检测；J——与检测头一起优化。比较中的语义信息指定在比较过程中是否结合了语义线索。

4.4 强大的 OoD 分割模型

几乎所有先前讨论的方法都面临的一个显著挑战是，它们的输出通常是零散的，无法在语义层面进行解释。然而，自动驾驶更关心的是语义层面的理解，而不是某些像素属于分布内而另一些像素属于分布外。语义分割领域的发展带来了一些强大的模型，可以为图像提供精确的掩膜。利用这些强大的分割模型以不同方式精确分割分布外的对象已成为一种趋势。此外，CLIP [37] 统一了自然语言和图像的特征空间，这使得在文本信息的

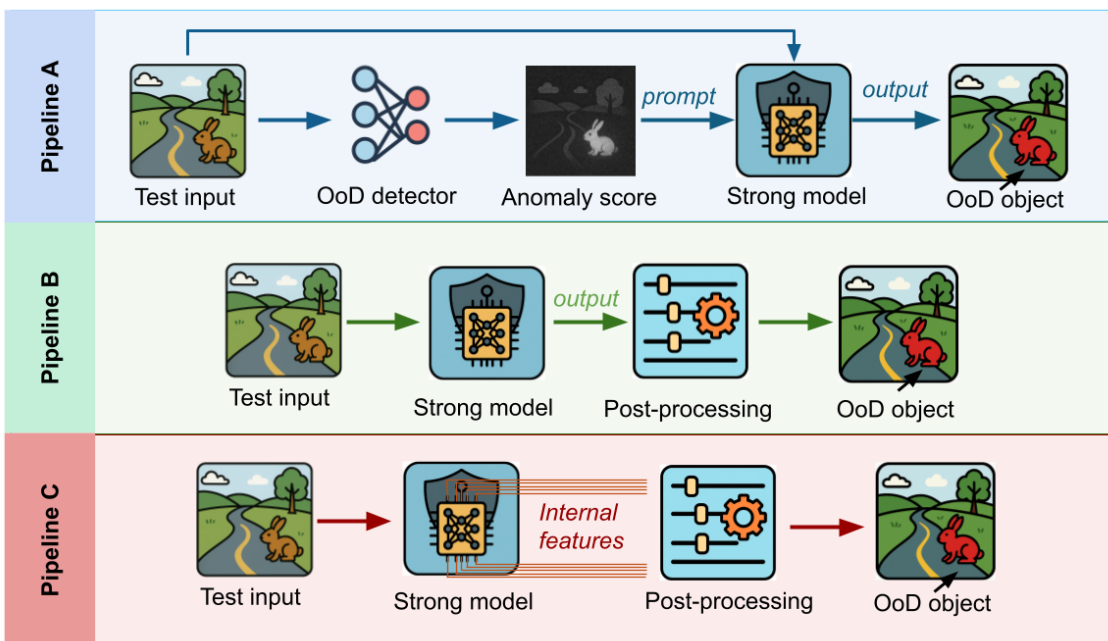


Fig. 6. 使用强大的预训练模型进行 OoD 分割的典型流程。流程 A 使用外部的 OoD 检测器，流程 B 使用最终模型输出，流程 C 使用中间特征

帮助下分割分布外的对象成为可能。在本节中，我们回顾了通过不同管道利用强大分割骨干或视觉语言模型的方法，如图 6 所示。

4.4.1 基于 Mask2Former 的 . Mask2Former [5] 是一个通用图像分割框架，专为使用遮罩注意力变换器解码器进行语义、实例和全景分割而设计。Mask2Anomaly [38] 基于 Mask2Former [5]，并将异常分割重新定义为一个遮罩分类任务，而不是逐像素分类。为了在对象层面上改善 OoD 分割，作者通过三个关键贡献扩展了 Mask2Former [5]，形成了 Mask2Anomaly 框架。首先，它引入了一种全球遮罩注意机制，帮助模型同时关注前景和背景。在训练期间，它使用来自 OE 的合成数据集进行对比学习，以鼓励模型最大化 ID 和 OoD 类之间的异常分数差距。最后，在推理期间应用了一个优化模块，基于全景分割结果过滤掉误报区域。Mask2Anomaly 显著降低了误报率，并提高了跨数据集的性能。基于 Mask2Former [5] 构建的复杂模块设计使 Mask2Anomaly 与高级分割模型的即插即用部署不兼容。

RbA [33] 由 Nayal 等人提出，用于解决逐像素 OoD 分割的局限性。RbA 没有修改 Mask2Former [5]，而是利用 Mask2Former 的输出，并引入一个评分函数来识别被所有已知类别拒绝的掩膜。具体来说，该方法聚合区域的类别概率和掩膜输出来估计每个像素的分类概率。为了测量被所有已知类别拒绝的掩膜，作者假设类别标签 $K + 1$ 代表分布外物体的区域。此异常值概率定义为：

$$p(y = K + 1 | x) = 1 - \sum_{k=1}^K p(y = k | x).$$

较高的异常值概率表明属于分布外物体的像素在所有已知类别中具有较低的预测概率。RbA 利用预训练的 Mask2Former 而无需在异常数据上进行训练就显示出强大的性能。一种轻量级微调变体仅优化少量模型参数，

Manuscript submitted to ACM

在不降低闭集性能的情况下超越了最新技术的性能。然而，Mask2Former 中的 $K + 1$ -th logit 最初被设计用于表示“无物体”类别，这可能无法准确对应异常区域，并将不确定性与异常检测相混淆。

EAM 类似于 RbA [14]，也认识到在 OoD 分割中利用掩码级识别模型的潜力。它提出的 EAM 异常检测器结合了负监督以取得有竞争力的结果。EAM 通过聚合多个掩码中的低置信度预测来估计不确定性。

为了解开不确定性预测和负面对象分类，Delic 等人提出了 UNO [7]，它整合了一个不确定性得分和一个负面得分。对于分布外分割场景，它将 Mask2Former 扩展到 $K + 2$ 类，并在异常数据集上进行微调，其中 $K + 1$ -th logit 代表负面对象， $K + 2$ -th logit 代表无对象类。具体而言，不确定性得分 $s_{Unc}(z)$ 被定义为已知 K 类中负的最大 softmax 概率：

$$s_{Unc}(z) := - \max_{k=1 \dots K} P(Y = k | z)$$

负目标得分 s_{NO} 被定义为：

$$s_{NO}(z) := P(Y = K + 1 | z)$$

UNO 得分被定义为这两个得分的总和。UNO 将负预测从不确定性预测中解耦，并在像素级和图像级基准测试中取得了强大的性能。然而，额外类别可能会对不确定性估计引入偏差。

张等人提出了类无关结构约束学习 (CSL) [56]，这是一种嵌入结构约束的插件框架。它提供了两种集成方案：一种是将知识从教师网络端到端蒸馏到 CSL 框架中，另一种是在推理过程中通过将区域提议与像素级预测融合来应用这些约束。此外，CSL 引入了一种掩码分割预处理方法，将语义掩码分解为独立的组件，从而减弱了类别偏差。CSL 改善了 OOD（出域）分割、零样本语义分割和领域适应的性能。然而，与更直接的方法相比，它显著增加了架构复杂性。

这些方法利用 Mask2Former 的掩码级预测来为分布外的对象生成精确的掩码。然而，Mask2Former 对新类别缺乏强大的泛化能力，并且难以准确分割其训练期间从未遇到过的对象。这一限制影响了所有基于 Mask2Former 的自动驾驶系统方法，因为新物体可能随时出现。

大型模型如 Segment Anything Model (SAM)、CLIP 和 DINO 系列为开放世界分割提供了强大的能力。SAM 在超过 1100 万张图像和 10 亿个掩码上进行了预训练，通过提示输入实现了零样本分割。CLIP 对齐了视觉和文本嵌入，使模型能够通过自然语言描述来推理新类别。DINO 系列作为特征提取的强大基础模型。这些模型为推进分布外分割提供了有希望的机会。

UGainS 利用点提示 [34] 的 SAM。它首先使用 RbA [33] 计算异常评分图。然后通过最远点采样策略进行点提示的采样。SAM 接收原始图像和这些点提示以生成分布外对象的掩码。尽管 SAM 在提示正确时能生成高度精确的掩码，但其性能受限于采样点的质量。如果异常评分图存在噪声，生成的提示可能次优并降低整体性能。值得注意的是，与使用 OoD 数据训练的 RbA 相比，UGainS 在路面异常数据集上的 FPR95 准确性表现更差。

为了提高对噪声异常图的鲁棒性，赵等人提出了 Score-to-Mask S2M [57]，该方法利用框提示而不是点提示。即使这些区域与真实物体不相似，点提示也可能受到假阳性结果的影响。S2M 可以构建在其他生成异常评分图的 OoD 检测器之上，例如 RPL [32] 或 PEBAL [45]。S2M 提示生成器处理异常评分图并为分布外物体生成框提示。它被实现为一个物体检测模块，并在一个异常曝光数据集上进行训练。SAM 是冻结的，只有提示生成器需要在训练期间更新。在推理过程中，将框提示和原始图像输入到 SAM 中，以生成 OoD 物体的精确掩码。S2M 在 SAM-B 上取得了强大的成果，证明了其对自动驾驶系统的有效性和潜力。然而，提示生成器可能会忽略极小的道路障碍物，而这些障碍物未被 OoD 检测器注意到。

Shoeb 等人提出了一个框架，利用来自 SAM 解码器的特性和似然比 (LR) 来预测分布外 (OoD) 对象 [43]。该方法无需 OoD 检测器，但在像素级 OoD 分割上表现不佳。为了探索 OoD 分割中的人机交互，Shoeb 等人 [42] 在他们的流程中引入了文本查询，我们将其缩写为 ROD。他们使用 RbA 分数来获取单帧中的 OoD 对象掩

码。然后，他们使用轻量级对象跟踪器链接后续帧中的相似片段。最后，截取的 OoD 路障补丁和用户提供的文本查询都通过 CLIP 嵌入，用于根据用户输入检索相关的 OoD 对象。该方法的一个基本挑战是用户可能并不总是知道如何描述未见过的对象，因为任何新实体都可能出现在路上。

Method	Powerful model	OoD detector	Fine-tuning	Pipeline	OoD info source	Key idea
Mask2Anomaly [38]	Mask2Former	^	☒	B	High-entropy background masks	Mask-classification reformulation; global mask attention
RbA [33]	Mask2Former	^	^	B	Rejected-by-all class prob.	Aggregate class probabilities to score masks without retraining
EAM [14]	Mask2Former	^	☒	B	Low-confidence masks + neg. sup.	Uncertainty aggregation across masks with negative supervision
UNO [7]	Mask2Former	^	☒	B	Uncertainty + neg. object logit	Decouple uncertainty and anomaly via UNO score
CSL [56]	Mask2Former	^	☒or ^	C	Teacher masks / region proposals	Structural constraints / mask splitting to reduce bias
UGainS [34]	SAM	☒	^	A	RbA anomaly map	Point-prompt from anomaly map
S2M [57]	SAM	☒	^	A	RPL/PEBAL anomaly map	Box-prompt generator for noise filtering
LR [43]	SAM	^	^	C	SAM decoder features	Likelihood-ratio on SAM features
ROD [42]	CLIP	☒	^	B	RbA masks + CLIP text embed.	Retrieve OoD objects via user text/ semantic similarity
PixOOD [49]	DINOv2	^	^	B	DINOv2 features	Probabilistic modelling in 2-D projected space

Table 5. 利用强大模型的 OoD 分割方法概述。Fine-tuning 列指出强大模型是经过微调 (☒) 还是被冻结 (^)，而 OoD 信息来源列则说明异常线索的来源。

在 OE 数据集上进行训练可以增强 OoD 分割，但可能引入某些偏差。PixOOD [49] 提出了一个用于 OoD 分割的框架，避免了在异常数据上进行训练。它使用冻结的 DINOv2 [35] 提取特征，这些特征被投影到二维空间。对于每一个类别，PixOOD 建模了分布内和分布外的似然性，并定义了 OoD 分数。相比于 GROOM [48]，PixOOD 通过在推理时放大特征图来支持像素级分割。这项工作表明，使用通用视觉特征可以实现 OoD 分割。这种方法是通用的，因为它消除了异常数据上进行训练的需要。

表 5 总结了利用强大的预训练模型的 Out-of-Distribution 分割方法，详细说明了它们的主干网络、流水线类型、微调策略、异常提示来源和核心思想。尽管这些方法已经展示了令人印象深刻的性能，但参数数量的增加也导致了更高的计算成本和能源消耗。还需要进一步的研究来评估在车载系统有限的计算资源下部署如此大型模型的可行性。此外，它们的推理时间应在实时应用场景中进行评估。

在 OoD 分割方面的最新进展使得在像素级别准确分割 OoD 对象成为实际可行的。这一成就可以为自动驾驶系统中的感知模块提供安全指导，使其能够在必要时更加谨慎地操作或请求人工接管。然而，仍然存在一些挑战，需要在未来研究中进行探索。

建立大规模、现实的像素级别 OoD 分割基准。现有研究主要在小型数据集上评估方法，这不能充分反映真实驾驶场景。与其他领域的大型基准相比，数千张图像相对有限，不足以评估多样化的方法。设计分割方法以在多样化的环境条件下保持 OoD 分割的性能。在实际中，语义偏移 (OoD 对象) 通常与协变量偏移 (如雨雪等噪声) 同时发生。这些协变量偏移进一步增加了 OoD 分割的难度。最近的工作已经开始利用视觉基础模型来处理这些输入 [22]。一些其他人尝试通过测试时自适应方法来解决这个问题 [11, 25, 27, 54, 55]。然而，这些方法主要关注于图像级别的检测，而不是在恶劣环境下的逐像素分割。为了实现带有协变量偏移输入的 OoD 分割，还需要更多的努力。

利用现成的视觉基础模型进行像素级的 OoD 分割。PixOOD [49] 展示了使用现成的视觉基础模型进行 OoD 分割的有效性。这种方法显示出巨大潜力，尤其是在视觉基础模型逐步强大的快速进展下。

在任务特定的未知类 (OoD) 分割中加入自然语言。这是将自然语言整合到 OoD 分割中。尽管描述一个未见的对象可能显得自相矛盾，但我们可以描述分布内环境。例如，城市驾驶与越野驾驶意味着不同的驾驶环境，而个人车辆与警车追捕反映了不同的情境。通过提供这样的基于语言的先验，我们可以为任务特定环境定制 OoD 分割模型，并在不同设置中提高其性能。

Category	Method	SMIYC-Anomaly		SMIYC-Obstacle	
		FPR ↓	mean F1 ↑	FPR ↓	mean F1 ↑
Test-time	MSP [17]	72.02	5.37	16.6	6.25
	SML [21]	39.5	12.20	36.8	3.00
	ODIN [28]	71.68	5.15	15.28	9.37
	Mahalanobis [26]	86.99	2.68	13.08	4.70
Outlier Exposure	RPL [32]	7.18	30.16	0.09	56.69
	PEBAL [45]	40.82	14.48	12.68	5.54
	Meta-OoD [4]	15.00	28.72	0.75	48.51
	DenseHybrid [13]	9.81	31.08	0.24	50.72
Reconstruction	DUIR [29]	25.93	12.51	4.70	8.38
	JSR-Net [47]	43.85	13.66	28.86	11.02
	DaCUP [46]	–	–	1.13	46.01
	SynBoost [8]	61.86	9.99	3.15	37.57
Powerful Models	RbA [33]	11.60	46.80	0.50	60.90
	EAM [14]	4.09	60.86	0.52	75.58
	Mask2Anomaly [38]	14.60	48.60	0.20	69.80
	PixOOD [49]	54.33	19.82	0.30	50.82
	UNO [7]	2.00	–	0.16	77.65
	CSL [56]	7.16	50.39	0.67	51.02
	LR [43]	–	–	0.20	78.40
	S2M [57]	1.04	60.4	0.02	64.96

Table 6. OoD 分割方法在 SMIYC-Anomaly 和 SMIYC-Obstacle 上的比较。较低的 FPR 和较高的平均 F1 更好。

References

- [1] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. 2018. Discriminative out-of-distribution detection for semantic segmentation. *arXiv preprint arXiv:1808.07703* (2018).
- [2] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. 2019. Simultaneous semantic segmentation and outlier detection in presence of domain shift. In *Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10–13, 2019, Proceedings 41*. Springer, 33–47.
- [3] Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. 2019. Fishyscapes: A benchmark for safe semantic segmentation in autonomous driving. In *proceedings of the IEEE/CVF international conference on computer vision workshops*. 0–0.
- [4] Robin Chan, Matthias Rottmann, and Hanno Gottschalk. 2021. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*. 5128–5137.
- [5] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. 2022. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1290–1299.
- [6] Clement Creusot and Asim Munawar. 2015. Real-time small obstacle detection on highways using compressive RBM road reconstruction. In *2015 IEEE Intelligent Vehicles Symposium (IV)*. 162–167. doi:10.1109/IVS.2015.7225680
- [7] Anja Delić, Matej Grcić, and Siniša Šegvić. 2024. Outlier detection by ensembling uncertainty with negative objectness. *arXiv preprint arXiv:2402.15374* (2024).
- [8] Giancarlo Di Biase, Hermann Blum, Roland Siegwart, and Cesar Cadena. 2021. Pixel-wise anomaly detection in complex driving scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16918–16927.
- [9] Alexey Dosovitskiy and Thomas Brox. 2016. Generating images with perceptual similarity metrics based on deep networks. *Advances in neural information processing systems* 29 (2016).
- [10] Yarin Gal et al. 2016. Uncertainty in deep learning. (2016).
- [11] Zhengqing Gao, Xu-Yao Zhang, and Cheng-Lin Liu. 2024. Unified Entropy Optimization for Open-Set Test-Time Adaptation. arXiv:2404.06065 [cs.CV] <https://arxiv.org/abs/2404.06065>

Manuscript submitted to ACM

- [12] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Networks. *arXiv:1406.2661* [stat.ML] <https://arxiv.org/abs/1406.2661>
- [13] Matej Grcić, Petra Bevandić, and Siniša Šegvić. 2022. Densehybrid: Hybrid anomaly detection for dense open-set recognition. In *European Conference on Computer Vision*. Springer, 500–517.
- [14] Matej Grcić, Josip Šarić, and Siniša Šegvić. 2023. On advantages of mask-level recognition for outlier-aware segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2937–2947.
- [15] David Haldimann, Hermann Blum, Roland Siegwart, and Cesar Cadena. 2019. This is not what i imagined: Error detection for semantic segmentation through visual dissimilarity. *arXiv preprint arXiv:1909.00676* (2019).
- [16] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joe Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. 2019. Scaling out-of-distribution detection for real-world settings. *arXiv preprint arXiv:1911.11132* (2019).
- [17] Dan Hendrycks and Kevin Gimpel. 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136* (2016).
- [18] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. 2018. Deep anomaly detection with outlier exposure. *arXiv preprint arXiv:1812.04606* (2018).
- [19] Jens Henriksson, Christian Berger, Stig Ursing, and Markus Borg. 2023. Evaluation of out-of-distribution detection performance on autonomous driving datasets. In *2023 IEEE International Conference on Artificial Intelligence Testing (AITest)*. IEEE, 74–81.
- [20] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*. Springer, 694–711.
- [21] Sanghun Jung, Jungsoo Lee, Daehoon Gwak, Sungha Choi, and Jaegul Choo. 2021. Standardized max logits: A simple yet effective approach for identifying unexpected road obstacles in urban-scene segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 15425–15434.
- [22] Mert Keser, Halil Ibrahim Orhan, Niki Amini-Naieni, Gesina Schwalbe, Alois Knoll, and Matthias Rottmann. 2025. Benchmarking Vision Foundation Models for Input Monitoring in Autonomous Driving. *arXiv preprint arXiv:2501.08083* (2025).
- [23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger (Eds.), Vol. 25. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [24] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems* 30 (2017).
- [25] Jungsoo Lee, Debasmit Das, Jaegul Choo, and Sungha Choi. 2023. Towards open-set test-time adaptation utilizing the wisdom of crowds in entropy minimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 16380–16389.
- [26] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. 2018. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems* 31 (2018).
- [27] Yushu Li, Xun Xu, Yongyi Su, and Kui Jia. 2023. On the robustness of open-world test-time training: Self-training with dynamic prototype expansion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 11836–11846.
- [28] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. 2017. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690* (2017).
- [29] Krzysztof Lis, Krishna Nakka, Pascal Fua, and Mathieu Salzmann. 2019. Detecting the unexpected via image resynthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2152–2161.
- [30] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. 2020. Energy-based out-of-distribution detection. *Advances in neural information processing systems* 33 (2020), 21464–21475.
- [31] Xihui Liu, Guojun Yin, Jing Shao, Xiaogang Wang, et al. 2019. Learning to predict layout-to-image conditional convolutions for semantic image synthesis. *Advances in neural information processing systems* 32 (2019).
- [32] Yuyuan Liu, Choubao Ding, Yu Tian, Guansong Pang, Vasileios Belagiannis, Ian Reid, and Gustavo Carneiro. 2023. Residual pattern learning for pixel-wise out-of-distribution detection in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1151–1161.
- [33] Nazir Nayal, Misra Yavuz, Joao F Henriques, and Fatma Güney. 2023. Rba: Segmenting unknown regions rejected by all. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 711–722.
- [34] Alexey Nekrasov, Alexander Hermans, Lars Kuhnert, and Bastian Leibe. 2023. Ugains: Uncertainty guided anomaly instance segmentation. In *DAGM German Conference on Pattern Recognition*. Springer, 50–66.
- [35] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. 2024. DINOv2: Learning Robust Visual Features without Supervision. *arXiv:2304.07193* [cs.CV] <https://arxiv.org/abs/2304.07193>
- [36] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 2019. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2337–2346.

- [37] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. Pmlr, 8748–8763.
- [38] Shyam Nandan Rai, Fabio Cermelli, Dario Fontanel, Carlo Masone, and Barbara Caputo. 2023. Unmasking anomalies in road-scene segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4037–4046.
- [39] David Rios Insua, William N Caballero, and Roi Naveiro. 2022. Managing driving modes in automated driving systems. *Transportation science* 56, 5 (2022), 1259–1278.
- [40] Matthias Rottmann, Pascal Colling, Thomas Paul Hack, Robin Chan, Fabian Hüger, Peter Schlicht, and Hanno Gottschalk. 2020. Prediction error meta classification in semantic segmentation: Detection via aggregated dispersion measures of softmax probabilities. In *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–9.
- [41] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115 (2015), 211–252.
- [42] Youssef Shoeb, Robin Chan, Gesina Schwalbe, Azarm Nowzad, Fatma Güney, and Hanno Gottschalk. 2024. Have We Ever Encountered This Before? Retrieving Out-of-Distribution Road Obstacles from Driving Scenes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 7396–7406.
- [43] Youssef Shoeb, Nazir Nayal, Azarm Nowzad, Fatma Güney, and Hanno Gottschalk. 2025. Segment-Level Road Obstacle Detection Using Visual Foundation Model Priors and Likelihood Ratios. arXiv:2412.05707 [cs.CV] <https://arxiv.org/abs/2412.05707>
- [44] Yiyu Sun, Chuan Guo, and Yixuan Li. 2021. React: Out-of-distribution detection with rectified activations. *Advances in neural information processing systems* 34 (2021), 144–157.
- [45] Yu Tian, Yuyuan Liu, Guansong Pang, Fengbei Liu, Yuanhong Chen, and Gustavo Carneiro. 2022. Pixel-wise energy-biased abstention learning for anomaly segmentation on complex urban driving scenes. In *European Conference on Computer Vision*. Springer, 246–263.
- [46] Tomáš Vojtíš and Jiří Matas. 2023. Image-consistent detection of road anomalies as unpredictable patches. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 5491–5500.
- [47] Tomas Vojir, Tomáš Šipka, Rahaf Aljundi, Nikolay Chumerin, Daniel Olmeda Reino, and Jiri Matas. 2021. Road anomaly detection by partial image reconstruction with segmentation coupling. In *Proceedings of the IEEE/CVF international conference on computer vision*. 15651–15660.
- [48] Tomáš Vojtíš, Jan Šochman, Rahaf Aljundi, and Jiří Matas. 2023. Calibrated out-of-distribution detection with a generic representation. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. IEEE, 4509–4518.
- [49] Tomáš Vojtíš, Jan Šochman, and Jiří Matas. 2024. PixOOD: Pixel-level out-of-distribution detection. In *European Conference on Computer Vision*. Springer, 93–109.
- [50] Haoqi Wang, Zhizhong Li, Litong Feng, and Wayne Zhang. 2022. Vim: Out-of-distribution with virtual-logit matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4921–4930.
- [51] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 2018. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 8798–8807.
- [52] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. 2003. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Vol. 2. Ieee, 1398–1402.
- [53] Yingda Xia, Yi Zhang, Fengze Liu, Wei Shen, and Alan L Yuille. 2020. Synthesize then compare: Detecting failures and anomalies for semantic segmentation. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I* 16. Springer, 145–161.
- [54] Yongcan Yu, Lijun Sheng, Ran He, and Jian Liang. 2024. STAMP: Outlier-Aware Test-Time Adaptation with Stable Memory Replay. In *European Conference on Computer Vision*. Springer, 375–392.
- [55] Longhui Yuan, Binhui Xie, and Shuang Li. 2023. Robust test-time adaptation in dynamic scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15922–15932.
- [56] Hao Zhang, Fang Li, Lu Qi, Ming-Hsuan Yang, and Narendra Ahuja. 2024. Csl: Class-agnostic structure-constrained learning for segmentation including the unseen. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 7078–7086.
- [57] Wenjie Zhao, Jia Li, Xin Dong, Yu Xiang, and Yunhui Guo. 2024. Segment Every Out-of-Distribution Object. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3910–3920.