arXiv:2301.03407v2 [cs.CV] 5 Apr 2023

论掩码级识别在异常值感知分割中的优势

ˇ ˇMatej Grci´c, Josip Sari´c, SinišaSegvi´c 萨格勒布大学，电气工程与计算学院，克罗地亚萨格勒布，Unska 3，邮编 10000{name.surname}@fer.hr

摘要

大多数密集识别方法都会针对每个特定像素分别做出判断。这些方法在常见的闭集设置中表现出色。然而，现实世界中的重要应用通常要求模型在存在异常值时仍具备强劲性能。我们表明，即使对于未经微调的基线模型，这一要求严苛的设置也能从掩码级预测中显著获益。此外，我们提出了一种密集识别不确定性的替代表述，能够有效减少语义边界处的假阳性响应。所提出的表述进一步提升了性能，超越了极强的基线，并在使用和不使用负样本数据训练的情况下，均达到异常值感知语义分割的最先进水平。我们的贡献也提升了近期一种全景设置下的性能。深入实验证实，我们的方法之所以有效，是因为它将像素级线索隐式聚合为掩码级预测结果。

1. 引言

深度学习的兴起彻底改变了计算机视觉领域 [34]。复杂而高效的深度网络提升了机器理解场景的能力 [20,61]。分割是一种非常重要的场景理解形式，应用于医学、农业、机器人学和汽车行业。在过去十年中，分割任务被建模为逐像素分类 [20, 44]。然而，这种方法假设相邻像素彼此独立，而这一假设在实践中并不成立。相邻像素通常因属于同一物体或场景部分而高度相关 [39]。尽管在设计和训练时错误地假设相邻像素彼此独立，所得模型在分布内场景中仍表现出有竞争力的泛化性能 [14, 15]。然而，由于对分类体系外场景部分的处理不足，其真实场景性能仍有很大的提升空间 [6, 11]。

一种近期提出的逐像素分类方法将定位与识别解耦 [17]。定位通过将像素分配给大量掩码来实现，每个掩码都经过训练，以捕获语义相关区域（例如道路或建筑物）。随后，将恢复出的语义区域作为整体进行分类。这种方法被称为掩码级识别 [16]。将定位与分类解耦，还能进一步使用同一模型进行语义分割、实例分割和全景分割。该共享架构在标准分割基准测试 [18, 39, 64] 上表现出色。

然而，先前的工作并未考虑基于掩码的方法在要求严苛的应用中的表现。因此，我们研究掩码级识别在迈向真实世界场景理解所面临的少数几个重大挑战中的价值——异常值感知语义分割 [7, 11, 29]和异常值感知全景分割 [29]。我们的实验表明，掩码级方法在这些挑战中表现出色。我们探究这种表现背后的原因，并提出改进，以支持这些重要应用。

掩码级识别具有若干有趣的特性。例如，掩码会被分类为 K 个已知类别或空类别，而掩码分配并非互斥的 [17]。与标准的逐像素方法相比，这为拒绝预测结果提供了更多机会。掩码级方法可以将掩码级不确定性传播至像素级。这不同于标准方法，后者必须为每个像素单独估计异常分数 [26]。显然，标准方法很容易忽略像素邻域中的局部相关性，而这似乎并不可取。就可扩展性而言，掩码级识别模型无需在输出分辨率上生成逐类别特征图。因此，设计者可以降低训练开销 [8]，并提高训练的灵活性。这些特性使掩码级识别成为一个极具吸引力的研究课题。

本文提出了以下贡献。我们指出，掩码级识别在异常值感知分割的标准基准上能够取得强劲的基线性能。我们的方法进一步利用了掩码级识别的特定偏置。

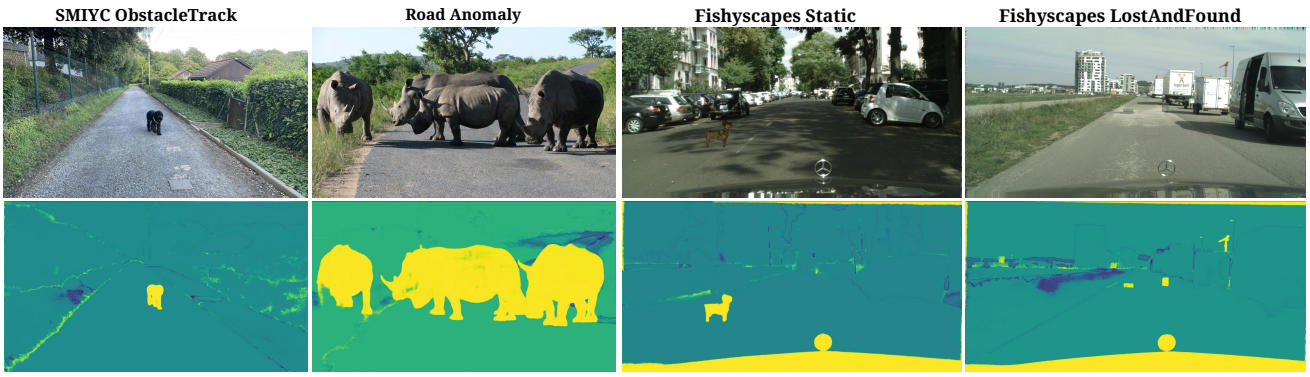


图 1. 采用所提出的掩码级方法进行异常值感知分割。图中展示了输入图像（上）和密集分布外（OOD）分数（下）。

将所提出的 EAM 异常值检测器与负监督相结合，在异常值感知语义分割和全景分割中取得了具有竞争力的结果。将所提出的方法与负监督相结合，还可以获得进一步改进。由此得到的模型在 Segment Me If You Can（SMIYC）基准的两个赛道以及经过适配的 MS COCO 上，将异常值感知分割的性能提升至最先进水平。

2. 相关工作

相关工作研究了掩码级识别任务（第 2.1节）以及存在异常值时的分割（第 2.2节）。

2.1. 自由形状区域的识别

早期的掩码级识别方法依赖类别无关的自底向上提议。它们沿着提议区域汇聚手工设计的 [9] 或卷积 [19, 25, 51] 特征，并通过对池化表示进行分类，做出掩码级决策。Mask-RCNN 通过在提议检测与掩码级分类之间共享特征，以及端到端训练所有参数，扩展了这一方法。近期，PointRend 提出仅通过选定的低不确定性预测结果反向传播损失 [33]。这使得 Mask-RCNN 的分辨率能够从 28×28 提升至 224×224，而对训练开销的影响可以忽略不计。最近，MaskFormer 摆脱了对自底向上提议的依赖，直接将像素分配给跨越任意图像区域的掩码 [17]。其核心组件是一个超网络 [24]，它为两个 1×1 卷积生成权重，将像素级嵌入转换为掩码分配分数，继而转换为语义图。这是首个在三种密集识别任务上取得有竞争力实验性能的架构：语义分割、实例分割和全景分割。Mask2Former [16] 通过引入一种特殊的注意力层，进一步改进了掩码超网络，促使模型逐步聚焦于特定掩码的前景像素。我们的工作探讨了 Mask2Former 在异常值感知分割和异常值感知全景分割场景中的性能。

2.2. 存在异常值时的分割

野外环境中的识别涉及超出训练分类体系的测试区域。理想的模型应拒绝在此类像素上进行决策 [54]。这可以通过限制决策边界的形状 [1, 55]，或为分类器配备异常检测器 [27, 38] 来实现。可以通过对嵌入空间中与已学习类别中心的距离进行阈值处理来限制决策边界 [10, 55]。采用更强的分类器可以进一步提升效果 [58]。然而，如果未知样本恰好映射到与已知类别样本相同的特征上，许多此类方法注定会失败。这种现象被称为特征坍塌 [45]。

早期将判别式预测扩展至分布外检测的方法，主要基于预测置信度 [27]，输入扰动 [38]，密度估计 [48] 和贝叶斯不确定性 [47]。多项研究指出，语义异常 [53] 可能尤其难以检测 [32, 48, 56]。一种有前景的方法是，在执行判别任务的同时生成合成异常 [13, 22, 35, 63]。通过使用负样本训练数据来模拟异常，也取得了进一步的实证改进 [28, 42]。然而，由于可能与测试异常重叠，这种做法可能导致对性能的估计过于乐观。

由于在鲁棒场景理解中的重要应用，离群点检测在密集预测场景下尤为重要 [7, 11, 62]。然而，直接

移植图像级方法会遇到两种重要的失效模式。首先，这些方法往往无法准确定位内点背景前方的异常 [3]。其次，内点像素在语义边界处经常会出现高熵预测结果，因此这些方法容易产生假阳性 [52]。因此，大量研究提出了定制设计来缓解这些问题。

对于部分异常图像，可以通过在混合内容图像上进行训练来处理 [3,5,23,57]。通过元分类汇总证据 [52]，或对输入进行预处理 [38]，可以解决相邻像素之间的相关性问题。通过对预训练特征 [7] 或联合训练的特征 [29,37] 拟合生成模型的预测头，可以避免使用真实训练数据。另一类方法则使用与对抗噪声 [2] 对应的合成负样本，或联合训练的生成模型的样本 [21] 进行训练。最后，一些方法会检测输入与重合成场景之间的差异 [5,41,59,60]。

与以往所有工作不同，我们根据掩码级预测来制定离群点检测。与元分类方法 [12,52] 不同，我们的方法只需要一次学习过程，且不需要负样本数据。我们的方法与大多数以往方法正交，因此为未来研究提供了一个令人期待的基线。

3. 存在离群点时的掩码级识别

我们提出了一种新方法，将掩码级密集预测扩展至异常值感知分割。我们的方法可应用于许多近期基于掩码级识别的密集预测方法之上 [17,29,36]。我们通过集成掩码级异常分数，构造了一种新颖的密集分布外（OOD）分数。由于该方法聚合了图像区域内的像素级证据，并降低了对语义边界的敏感性，因此能够提升真实数据集上的异常值感知分割效果。

3.1. 基于掩码级识别的语义分割

掩码级分割方法将分类与定位解耦，并通过独立的预测头对二者进行建模 [17]。定位可通过概率分配（掩码） $\mathcal{S} = \{\mathbf{m}_i \,|\, i = 1, \dots, N\}$ 来表示，这些掩码能够捕捉语义相关区域。每个掩码 $\mathbf{m}_i$ 都是一个数组，其中包含对应像素的概率分配。我们可以将掩码合并为一个三维张量 $\mathbf{m}^{N\times H\times W}$。通过将标准密集特征 $\mathbf{E}$ 输入推断得到的投影 $\mathbf{w}_{\text{loc}}$ 并进行 Sigmoid 激活，即可恢复掩码：

$$\mathbf{m} = \sigma(\text{conv}_{1\times 1}(\mathbf{E}, \mathbf{w}_{\text{loc}})). \tag{1}$$

识别可通过推断 N 个掩码级类别分布来完成，类别包括 K 个已知类别和一个空类。我们将这些预测结果记为 $P_i(Y = k|\mathbf{x})$，$i \in 1..N$，$k \in 1..K+1$。接下来，考虑非空类别的概率，并将其排列成一个 $N \times K$ 矩阵 $\mathbf{w}_{\text{cls}}$。然后，可根据 $\mathbf{w}_{\text{cls}}$ 投影掩码，从而恢复闭集语义分割分数张量：

$$\mathbf{H}_{\text{closed}} = \text{conv}_{1\times 1}(\mathbf{m}, \mathbf{w}_{\text{cls}}). \tag{2}$$

请注意，该张量不包含分布，因为 $\sum_i m_i[r,c] \neq 1$ 和 $\sum_k \mathbf{w}_{\text{cls}}[i,k] \neq 1$。上述卷积可以解释为根据逐掩码分类器的加权集成对每个像素 (r,c) 进行分类，其中权重对应于稠密掩码分配：

$$\hat{y}[r,c] = \underset{k=1\dots K}{\text{argmax}} \sum_i \mathbf{m}_i[r,c] \cdot P_i(Y = k|\mathbf{x}) \,. \tag{3}$$

图 2 （左）显示，密集特征 $\mathbf{E}$ 以常规方式生成：将现成的骨干网络连接到带有跳跃连接的上采样解码器。主要创新在于一个称为掩码解码器的超网络，它接收潜在特征并推断图像级权重 $\mathbf{w}_{\text{loc}}$ 和 $\mathbf{w}_{\text{cls}}$。训练过程将掩码分配 $\mathbf{m}$ 和掩码级识别 $P_i(Y = k|\mathbf{x})$ 拟合到稠密标签。

3.2. 检测像素级预测中的离群点

密集分布外检测需要一个评分函数 $\mathbf{s}_{\text{ood}}$：$[0, 255]^{3\times H\times W} \to \mathcal{R}^{H\times W}$，将每个像素映射为相应的异常分数。随后，我们可以通过对异常分数 $\mathbf{s}_{\text{ood}}(\mathbf{x})$ 进行阈值处理来检测异常。将异常与闭集分割融合，即可恢复异常值感知分割。

若干标准基线根据像素级预测结果的不确定性来检测异常区域 [7,26]。预测的不确定性可以用最大分数 [27],熵 [12], 能量 [42] 等指标量化。我们将通过 PerPixel 基线评估这种方法：该基线移除了掩码解码器，并将其替换为标准的逐像素预测结果[44]。

也可以通过掩码级模型恢复像素级预测结果。训练过程促使掩码 $\mathbf{m}_i$ 专门用于捕捉特定的视觉概念。因此，可以定义一种像素级异常分数，将未被任何掩码覆盖的像素判定为异常：

$$\mathbf{s}^{\text{AM}}_{\text{ood}}(\mathbf{x})[r,c] = -\max_i \mathbf{m}_i[r,c] \tag{4}$$

AM 是“最大掩码异常”（Anomaly of the max-Mask）的缩写。因此，当所有掩码的置信度都较低时，我们应得到较高的异常分数。尽管这种方法优于逐像素基线，但仍远不够完善。图3 展示了 Fishyscapes L&F 验证集上内点和异常值的直方图，其依据是最大值 m_i 分数。左侧直方图显示，几乎

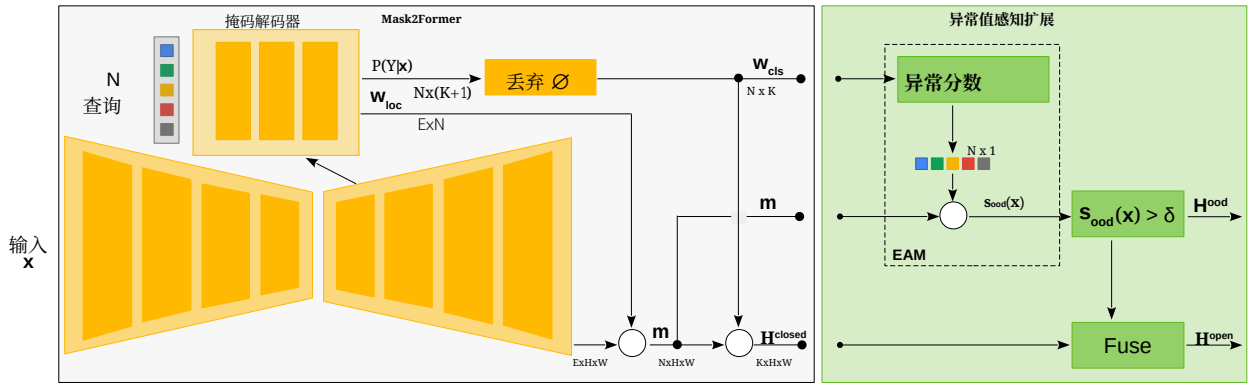


图2. 我们关注标准 M2F 模型（左侧）生成的三个张量 [16]：闭集分割 $\mathrm{H}_{\text{closed}}$ (K×H×W)、逐掩码稠密二元分配 m (N×H×W)，以及全图范围的掩码级类别分数 w_{cls} (N×K)。我们首先通过量化掩码级预测的不确定性 $_{\text{cls}}$，开始构建具备离群点感知能力的扩展（右侧）。如式 (7) 所示，我们根据稠密掩码分配 m，将逐掩码异常分数重新分配给各像素，从而恢复稠密异常图 $\mathrm{s}^{\text{EAM}}_{\text{OOD}}$ (H×W)。我们通过对 $\mathrm{s}^{\text{EAM}}_{\text{OOD}}$ 进行阈值处理并将其与 $\mathrm{H}_{\text{closed}}$融合，构建具备离群点感知能力的分割 H^{open}。注意，$\sum_{rc} \mathrm{m}_i[r,c] \neq 1$。

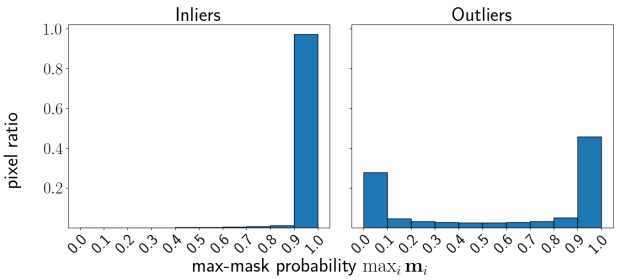


图3. Fishyscapes L&F 验证集上，内点和异常值像素中依据最大掩码概率计算的相对像素频率。

所有内点几乎都具有高置信度的掩码分配。另一方面，异常值分布呈现出明显的两极分化。左侧峰值很容易与内点区分开来，但右侧峰值则更具挑战性。这表明，像素级预测可能并非解决我们问题的最佳方案，因为许多真实的异常值像素都获得了高置信度的掩码分配。因此，我们考虑以掩码级不确定性为基础构建方法。

3.3. 检测掩码级预测中的离群点

我们首先考虑一种方法，该方法将稠密异常分数恢复为最强掩码的掩码级不确定性。如果选择最大 Softmax 作为不确定性度量，则可以将该分数表述为：

$$\mathbf{s}^{\text{AHM}}_{\text{ood}}(\mathbf{x})[r,c] = -\max_{k=1\ldots K} P_{\text{argmax}_i \mathbf{m}_i[r,c]}(Y=k|\mathbf{x}). \quad (5)$$

AHM 是“硬分配掩码异常分数”（Anomaly score of Hard-assigned Masks）的缩写。然而，这种方法完全忽略了主导掩码分配的不确定性。这显然并非最优方案，我们的实证结果也证实了这一直觉。因此——

我们着手结合像素级掩码分配和掩码级识别的不确定性。

我们接下来考虑闭集语义分割分数（3）。我们可以根据任意异常检测器来量化其不确定性。如果选择最大 logit 检测器 [26]，则得到如下结果：

$$\mathbf{s}^{\text{AEM}}_{\text{ood}}(\mathbf{x})[r,c] = -\max_{k=1\ldots K} \sum_i \mathbf{m}_i[r,c] \cdot P_i(Y=k|\mathbf{x}). \quad (6)$$

闭集语义分割分数可以视为逐掩码分类器的集成输出，其中掩码分配充当集成成员的权重。因此，我们将该分数表示为集成掩码级预测异常度（AEM）。

最后，我们考虑将异常检测器直接应用于掩码级分类分数。我们提出根据其掩码分配$\mathbf{m}$，聚合每个特定像素中得到的证据。该方法可以解释为对掩码级预测异常分数进行集成（EAM）。由于其与掩码级不确定性存在直接联系，该方法具有直观的合理性。如果根据最大类别概率量化掩码级不确定性，则可以得到 AEM 分数的下界（6）：

$$\begin{aligned}\mathbf{s}^{\text{EAM}}_{\text{ood}}(\mathbf{x})[r,c] &= \sum_i \mathbf{m}_i[r,c] \cdot (-\max_{k=1\ldots K} P_i(Y=k|\mathbf{x})) \\ &\leq -\max_{k=1\ldots K} \sum_i \mathbf{m}_i[r,c] \cdot P_i(Y=k|\mathbf{x}) \quad (7)\end{aligned}$$

图2 （右）展示了如何根据 M2F 输出计算 EAM 分数。

我们预期，这两种方法之间的差异在语义边界处最为明显。这里

相邻掩码通常会降低其像素分配置信度。在这种情况下，我们提出的 EAM 方法会正确地输出低于 AEM 的异常分数。图 4展示了 Fishyscapes L&F 中两个场景下 EAM 与 AEM 的评分差异。我们在大多数图像像素中观察到类似的行为。不过，所提出的 EAM 方法在语义边界处输出的异常分数明显更低。这有助于减少语义边界处内点像素上的假阳性检测。

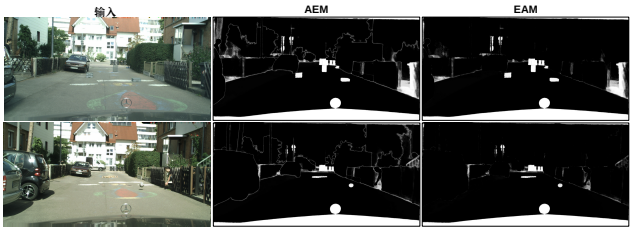


图 4. 像素级与掩码级分布外检测。掩码级分布外检测缓解了语义边界处假阳性的已知问题。请放大查看详情。

3.4. 利用负样本提升性能 data

使用负样本数据进行训练，是近期许多离群点检测方法的重要组成部分， [5, 12, 23, 28, 57]因为这有望解决特征坍塌问题。在密集预测中，这通常涉及将负样本数据粘贴到内点训练图像上， [3, 21]。现有实现要求在负样本像素上增加一个损失项， [12, 28, 42]。相反，我们的方法无需对模型或损失函数作任何改动。

我们建议将负样本像素的真实标注设为空洞区域类别。这会指示所有掩码避开负样本像素。这是合理的，因为空像素不属于任何感兴趣的类别。这种训练增加了空洞区域内容的多样性；如果掩码将任何空洞区域内容纳入其中，就会受到惩罚。

标准密集分类器 [14, 50] 无法使用标记为空的负样本进行训练。这是因为标准的逐像素交叉熵损失不会在空像素上计算。因此，我们的粘贴过程专用于掩码级识别。

图5展示了一个训练示例：输入图像裁剪块及其对应的真实标注二元标签。真实标注中没有任何标签涵盖粘贴的负样本像素。我们的实验表明，这种监督方式能够泛化到真实世界图像中的离群点检测。

4. 实验

我们的实验探讨了掩码级识别在离群点感知语义分割中的优势。我们考察语义分割（第4.1节）和全景分割（第4.2节）。

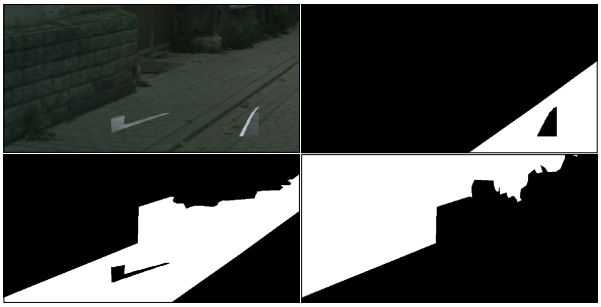


图5. 混合内容训练图像（左上）以及道路（右上）、人行道（左下）和建筑物（右下）类别的掩码分配真实标注。模型经过训练，能够将粘贴的两个负样本实例排除在所有掩码之外。

4.1. 道路驾驶图像的异常值感知分割

我们在两个标准基准上评估异常值感知分割的性能。Fishyscapes基准包含两个专注于城市道路驾驶的赛道 [7]。FS L&F赛道重新标注了 Lost and Found数据集的一个子集。FS Static赛道将异常物体粘贴到 Cityscapes 验证集图像中。SMIYC 基准（Seg- ment Me If You Can）包含两个赛道，涵盖了多种多样环境中的真实世界异常。异常赛道包含可能出现在图像任意位置的大型异常，而障碍物赛道则专注于路面上的小型异常。

我们根据平均精度（AP）和真正率为 95% 时的假阳性率（FPR_{95}）来衡量分布外检测的性能。我们使用带有 Swin-L [43] 骨干网络的 Mask2Former（M2F） [16] 。按照通常的做法，我们在两种设置下训练模型：使用和不使用负样本数据。不使用辅助数据的模型仅使用 Cityscapes 图像 [18]。由于域偏移较大，这可能会降低模型在 SMIYC 上的性能 [58]。使用负样本数据的模型首先在 Cityscapes 和 Mapillary Vistas 的图像上，使用 Cityscapes 分类体系进行训练 [49]。随后，我们使用混合内容图像对模型进行 2K 次迭代的微调。我们将 ADE20K [64] 实例作为负样本数据进行粘贴。除批量大小设为 18 外，我们使用标准超参数 [16]。耗时最长的实验在 3×A6000 上持续约 48 小时。

表1 比较了我们最佳方法（M2F-EAM）与 SMIYC 上相关工作的性能。这两个部分根据方法是否使用真实负样本数据进行训练来组织。我们的模型在不使用负样本数据训练的情况下，在两个赛道中都取得了较高的平均精度。较高的平均精度和相对较差的假阳性率$_{95}$ 得分表明，置信度极高的假阴性检测极少发生。对AUROC曲线的分析支持这一假设，因为我们取得了假阳性率$_{90}$ = 20%。

使用更多样化的封闭集图像进行训练，并通过负样本数据微调，可显著提升结果。此外，使用辅助数据训练的模型在 SMIYC 基准的所有指标上都达到了最先进的性能。假阳性率的大幅改善表明，使用负样本数据训练能够提升模型检测多样化异常的能力。

方法	Aux data	AnomalyTrack AP	AnomalyTrack 假阳性率$_{95}$	ObstacleTrack AP	ObstacleTrack 假阳性率$_{95}$
图像重合成 [41]	✗	52.3	**25.9**	37.7	**4.7**
道路修复 [40]	✗	-	-	54.1	47.1
JSRNet [59]	✗	33.6	43.9	28.1	28.9
最大 Softmax [27]	✗	28.0	72.1	15.7	16.6
蒙特卡洛 Dropout [31]	✗	28.9	69.5	4.9	50.3
ODIN [38]	✗	33.1	71.7	22.1	15.3
嵌入密度 [7]	✗	37.5	70.8	0.8	46.4
M2F-EAM（我们的方法）	✗	**76.3**	93.9	**66.9**	17.9
SynBoost [5]	✓	56.4	61.9	71.3	3.2
DenseHybrid [23]	✓	78.0	9.8	78.7	2.1
PEBAL [57]	✓	49.1	40.8	5.0	12.7
空类分类器 [7]	✓	36.6	63.5	10.4	41.5
最大熵 [12]	✓	85.5	15.0	85.1	0.8
M2F-EAM（我们的方法）	✓	**93.8**	**4.1**	**92.9**	**0.5**

表1。SMIYC上的异常值感知分割。我们的AP性能在两个类别中均优于此前的所有方法。

表 2 比较了我们的方法（M2F-EAM）与Fishyscapes基准上的相关工作 [6]。与之前一样，这两个部分根据是否使用真实负样本数据，将方法分别归类（底部为使用，顶部为不使用）。我们的方法在两个类别的FS Static上均取得了最佳性能，并在FS Lost and Found上取得了最佳AP性能。

方法	FS L&F AP	FS L&F FPR	FS Static AP	FS Static FPR	CS val mIoU
最大熵 [12]	15.0	85.1	0.8	77.9	9.7
图像重合成 [41]	5.7	48.1	29.6	27.1	81.4
最大 Softmax [27]	1.8	44.9	12.9	39.8	80.3
SML [30]	31.7	21.9	52.1	20.5	-
嵌入密度 [7]	4.3	47.2	62.1	17.4	80.3
NFlowJS [22]	39.4	9.0	52.1	15.4	77.4
SynDHybrid [23]	51.8	11.5	54.7	15.5	79.9
M2F-EAM（我们的方法）	9.4	41.5	**76.0**	**10.1**	83.5
SynBoost [5]	43.2	15.8	72.6	18.8	81.4
先验熵 [46]	34.3	47.4	31.3	84.6	70.5
分布外检测头 [4]	30.9	22.2	84.0	10.3	77.3
空类分类器 [7]	10.3	22.1	45.0	19.4	70.4
狄利克雷先验 [46]	34.3	47.4	84.6	30.0	70.5
DenseHybrid [23]	43.9	**6.2**	72.3	5.5	81.0
PEBAL [57]	44.2	7.6	92.4	1.7	-
M2F-EAM（我们的方法）	**63.5**	39.2	**93.6**	**1.2**	83.5

表2. Fishyscapes基准上的异常值感知分割。我们的方法在平均精度方面优于所有先前方法。

表3 在 Road Anomaly [41] 和 Fishyscapes [6]的验证子集上评估异常值感知分割。我们将掩码级方法与标准的像素级基线（PerPixel）及先前的工作进行比较。同样，底部部分的方法使用辅助负样本数据进行训练，而其他方法只接触内点。我们的两种掩码级方法均优于像素级基线和所有先前的方法。在这两种掩码级方法中，对异常分数进行集成（M2F-EAM）优于对集成结果计算异常分数（M2F-AEM）。

模型	Road Anomaly AP	Road Anomaly FPR	FS L&F AP	FS L&F FPR	FS Static AP	FS Static FPR
最大软最大概率 [27]	15.7	71.4	4.6	40.6	19.1	24.0
机器学习 [26]	19.0	70.5	14.6	42.2	38.6	18.3
NFlowJS [21]	-	-	40.2	18.7	34.4	11.2
SML [30]	25.8	49.7	36.6	14.5	48.7	16.8
SynthCP [60]	24.9	64.7	6.5	46.0	23.2	34.0
密度 [7]	-	-	4.1	22.3	-	-
PerPixel	49.3	31.0	2.5	56.7	11.5	34.8
M2F-AEM	**66.9**	15.3	51.2	28.0	86.2	3.5
M2F-EAM	**66.7**	**13.4**	**52.0**	**20.5**	**87.3**	**2.1**
SynBoost [5]	38.2	64.8	60.6	31.0	66.4	25.6
能量 [42]	19.5	70.2	16.1	41.8	41.7	17.8
PEBAL [57]	45.1	44.6	58.8	4.8	92.1	1.5
DenseHybrid [23]	-	-	63.8	6.1	60.0	4.9
M2F-EAM	**69.4**	**7.7**	**81.5**	**4.2**	**96.0**	**0.3**

表3. 我们的掩码级方法（M2F-EAM、M2F-AEM）与像素级基线（PerPixel）及以往在 RoadAnomaly 和 Fishyscapes 验证集上的工作的比较。

4.2. MS COCO 上的异常值感知全景分割

掩码级离群检测也可应用于全景分割。我们采用近期相关工作中的最具挑战性的设置 [29]：训练期间将 COCO 中 20% 的物体类重新标记为空像素。这些类别包括餐桌、香蕉、自行车、蛋糕、水槽、猫、键盘和熊。推理时，模型必须将这些类别中的所有像素分类为专门的异常物体类。异常值感知性能根据标准指标全景质量、分割质量和识别质量进行衡量。与先前工作 [29] 一样，我们的模型采用 ResNet-50 骨干网络。

掩码级训练促使所有掩码避免涵盖空像素。我们的异常检测器能够感知由此产生的掩码分配缺失。因此，我们的监督强度与空像素抑制 [29] 非常相似。推理时，我们通过对掩码级异常分数进行阈值处理，恢复稠密异常图。我们在一个留出的验证图像上校准阈值，使离群点检测的真正率达到 95%。我们将每个异常像素分配给其最合适的掩码，并保留所有像素数超过 200 的掩码来构成实例。

表4 将我们的方法与 EOPSN 论文中的几种方法进行了比较 [29]。尽管所需的监督少得多，我们的表现仍优于所有先前工作。值得注意的是，我们的方法可以轻松适用于异常背景类。

图 6展示了 COCO 验证集中三个场景的定性结果。各行依次为：输入图像、真实标注、两种

方法	已知			未知		
	PQ	SQ	RQ	PQ	SQ	RQ
空背景	37.7	76.3	46.6	4.0	71.1	5.7
空无知	37.2	76.3	45.9	3.7	71.8	5.2
空抑制	37.5	75.9	46.1	7.2	**75.3**	9.6
空训练	36.9	76.4	45.5	7.8	73.4	10.7
EOPSN [29]	37.4	76.2	46.2	11.3	73.8	15.3
Open-M2F-AEM	**43.5**	**82.0**	**52.2**	11.3	73.3	15.3
Open-M2F-EAM	**43.5**	**82.0**	**52.2**	**13.2**	73.4	**18.0**

表4. COCO 上的异常值感知全景分割。我们将 20% 的物体类重新标记为未知空类别 [29]。无论是在已知类别还是未知类别上，我们的表现都优于其他方法。

种来自 [29] 的结果，以及最后我们的结果。这些结果清楚地展示了我们的方法在离群点感知全景分割方面较以往最先进方法的改进。

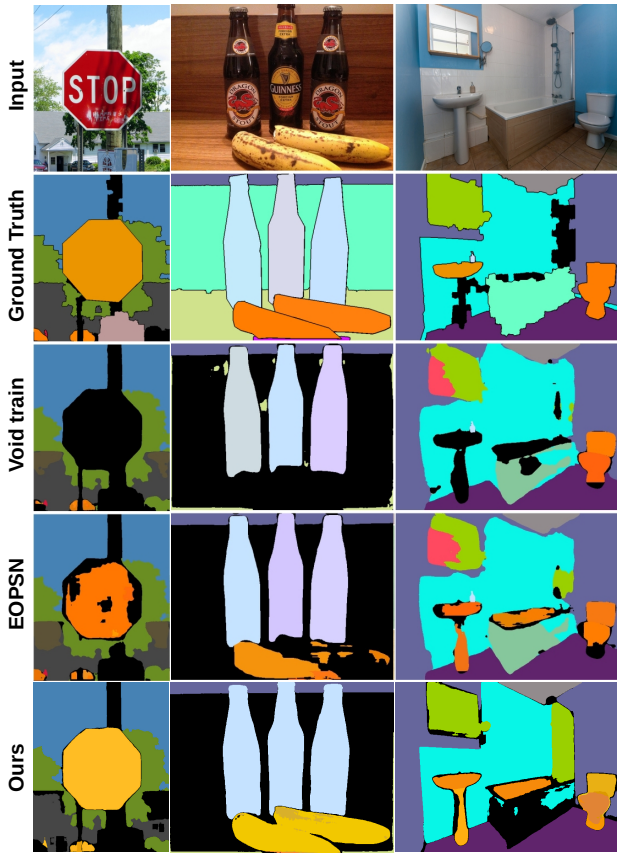


图6. 使用 M2F-EAM 进行异常值感知全景分割。停车标志、香蕉、马桶和水槽被视为未知物体类别 [29]。我们检测出所有未知类别，并区分出部分实例。

最后请注意，全景掩码级模型也可用于标准的异常值感知语义分割。事实上，在表3所列的 6 项指标中，全景模型有 3 项优于其语义分割对应模型。

5. 消融实验

我们对分布外分数的选择（第 5.1 节）、骨干网络（第 5.2 节）、掩码数量（第5.3 节）以及负样本数据来源（第 5.4 节）进行了消融研究。

5.1. 分布外分数的影响

表 5 考察了可用于我们方法的多种分布外检测器。这五个部分分别考察了逐像素基线，以及前文提到的 M2F-AM、M2F-AHM、M2F-AEM 和 M2F-EAM。我们注意到，掩码分数集成和掩码分数本身都不是分布。因此，在最后四个部分中，我们不考虑概率异常检测器。相反，我们只考虑直接取硬最大值（这与最大 Softmax 有关）或能量分数（对数求和指数）。这两种选择的表现相当，因此我们在提交给 SMIYC 的结果中选择使用更简单的硬最大值。与之前一样，我们观察到 M2F-EAM 略优于 M2F-AEM，而逐像素离群点检测的表现较差，这与先前研究的结果一致 [7, 26]。此外，我们观察到，基于集成的方法优于更简单的对应方法 M2F-AM 和 M2F-AHM。

方法	异常检测器	FS L&F	FS Static
PerPixel	熵 [28]	2.9	12.7
	KL散度 [22]	4.1	16.4
	能量 [42]	2.4	11.3
	最大 Softmax [27]	1.8	8.9
M2F-AM	最大分数	30.9	30.2
M2F-AHM	最大分数	3.5	44.4
M2F-AEM	能量	51.1	86.6
	最大分数	51.2	86.2
M2F-EAM	能量	48.5	69.3
	最大分数	**52.0**	**87.3**

表 5. 可集成到我们的方法中的异常检测器验证。能量分数（对数求和指数）的表现与采用硬最大值相近。同样，M2F-EAM 优于 M2F-AEM，而这两种掩码级方法均优于 M2F-AM、M2F-AHM 和逐像素基线。

5.2. 骨干网络的影响

表6 比较了采用不同骨干网络的逐像素模型和掩码分类模型的分布外检测性能。我们考察了两种卷积骨干网络：ResNet-50 和更先进的 ConvNeXt-L。我们还考察了基于变换器的骨干网络 Swin-L。此外，我们展示了采用 ResNet-50 骨干网络的 DeepLabV3+模型的结果。我们的逐像素基线与 DLv3+的表现相近，而 Mask2Former 的表现优于这两种方法。基于 Swin-L 的 M2F 模型表现出色，这表明

大容量和Transformer架构可能对基于掩码的异常值感知分割至关重要。

骨干网络	模型	FS L&F AP	FS L&F FPR	FS Static AP	FS Static FPR	CS val mIoU
ResNet-50	DLv3+	3.5	45.0	-	-	77.8
	PerPixel	1.3	64.0	9.0	42.9	79.6
	M2F-EAM	20.8	22.7	36.7	23.8	79.4
ConvNeXt-L	M2F-EAM	31.5	28.6	76.3	6.3	82.6
Swin-L	PerPixel	2.5	56.7	11.5	34.8	83.2
	M2F-EAM	**52.0**	**20.5**	**87.3**	**2.1**	**83.5**

表6. Fishyscapes 验证集上采用不同骨干网络的若干模型的比较。掩码级模型优于其逐像素对应模型，这是我们工作的一个重要结论。

5.3. 掩码数量的影响

表7 探讨了掩码数量 N 对闭集识别和离群点检测的重要性。我们考察了掩码数量等于类别数量（N=19）的情况，以及另外两种掩码数量更多的设定（N=50, 100）。这些实验表明，N 对离群点检测性能有很大影响，尽管掩码数量较多对两项任务都有帮助。

掩码数量	FS L&F AP	FS L&F 假阳性率$_{95}$	FS Static AP	FS Static 假阳性率$_{95}$	CS val mIoU
19	33.5	18.7	72.5	6.8	82.8
50	47.9	24.7	69.7	4.8	83.1
100	**52.0**	**20.5**	**87.3**	**2.1**	**83.5**

表7. 掩码数量对 M2F-EAM 的离群点检测和闭集分割的影响。丰富的掩码集提升了对离群点的鲁棒性。

5.4. 负样本数据源的影响

表8 验证了 Road Anomaly 和 Fishyscapes 验证子集上不同负样本数据来源的效果。第一行展示了未使用负样本数据训练时的结果。第二行对应于将从批次中其他图像随机选取的方形图块粘贴到当前图像上。第三行对应于将仅使用内点图像训练的归一化流模型生成的图块粘贴到图像上。最后一行对应于粘贴来自 ADE20K、按照其真实标注掩码裁剪的实例。结果表明，在 Fishyscapes 上，粘贴 ADE20K 实例的效果优于所有其他方法；在 Road Anomaly 上，该方法取得了最佳假阳性率，并获得了相当的平均精度分数。因此，我们选择这种方案作为使用负样本数据训练时的默认设置。

负样本	道路异常 AP	道路异常 FPR	FS L&F AP	FS L&F FPR	FS Static AP	FS Static FPR
无负样本	66.7	13.4	52.0	20.5	87.3	2.1
内点补丁	**69.7**	8.8	77.0	10.1	95.8	0.7
生成样本	68.9	8.4	80.6	4.5	91.9	0.9
ADE20K 实例	69.4	**7.7**	**81.5**	**4.2**	**96.0**	**0.3**

表8. 各类负样本数据的验证。广泛的负样本数据集优于其他替代方案。

6. 结论

在存在离群点的情况下保持稳健性能，是许多令人振奋的场景理解应用的重要前提。以往大多数密集预测方法都基于像素级分布外检测，因此未能考虑相邻像素之间的相关性。我们通过将分布外检测的对象从像素转为区域，来解决这一研究问题。由此得到的掩码级预测汇总了像素级证据，从而提高了相应异常分数的统计效力。我们还表明，在针对特定掩码进行决策集成之前执行分布外检测尤其有益。通过向空洞区域内容中注入负样本数据，我们进一步提升了性能。最后，我们将基于掩码的模型扩展到存在离群点时的全景推理。实验表明，掩码级离群检测大幅优于像素级方法，并且在不使用真实负样本数据训练的方法中，达到了最先进的 AP 性能。此外，尽管所需监督少于以往工作，它也超越了此前离群点感知全景分割的最先进水平。所提出的掩码级离群点感知分割方法适用于任何基于判别式识别分数的异常检测器，并可与许多以往方法结合。未来值得探索的方向包括使用合成负样本进行学习，以及对掩码范围描述子的概率密度进行建模。论文发表后将公开源代码。

7. 局限性

尽管我们的方法取得了极具竞争力的 AP分数，但如果异常值对象与某个已知类别相似，FPR95性能可能会很差。不过，如实验所示，使用负样本训练数据可以有效缓解这一问题。

致谢

本研究得到克罗地亚科学基金会 IP-2020-02-5851 ADEPT 项目、NVIDIA 学术硬件资助计划以及欧洲区域发展基金 KK.01.1.1.01.0009 DATACROSS 和 KK.01.2.1.02.0119 A-Unit 项目的支持。

参考文献

[1] Abhijit Bendale 和 Terrance E. Boult。《迈向开放集深度网络》。载于 IEEE计算机视觉与模式识别会议（CVPR），2016年。2[2] Victor Besnier、Andrei Bursuc、David Picard 和 Alexandre Briot。《触发失效：通过学习局部对抗攻击进行分布外检测，应用于语义分割》。载于 ICCV，第15681–15690 页，2021年。3[3] Petra Bevandic、Ivan Kreso、Marin Orsic 和 Sinisa Segvic。《域偏移情况下的语义分割与离群点检测同步进行》。载于 GCPR，2019年。3, 5[4] Petra Bevandic、Ivan Kreso、Marin Orsic 和 Sinisa Segvic。《基于使用带噪声负样本图像进行训练的密集开放集识别》。《图像与视觉计算》，124:104490，2022年。6[5] Giancarlo Di Biase、Hermann Blum、Roland Siegwart 和 C´esar Cadena。《复杂驾驶场景中的像素级异常检测》。载于计算机视觉与模式识别（CVPR），2021年。3, 5, 6[6] Hermann Blum、Paul-Edouard Sarlin、Juan I. Nieto、Roland Siegwart 和 Cesar Cadena。《Fishyscapes：自动驾驶中安全语义分割的基准》。载于2019年 IEEE/CVF国际计算机视觉大会研讨会，第 2403–2412 页。IEEE，2019年。1, 6[7] Hermann Blum、Paul-Edouard Sarlin、Juan Nieto、Roland Siegwart 和 Cesar Cadena。《Fishyscapes基准：衡量语义分割中的盲点》。《国际计算机视觉期刊》，2021年。1, 2, 3, 5, 6, 7[8] Samuel Rota Bul`o、Lorenzo Porzi 和 Peter Kontschieder。《用于内存优化训练深度神经网络的原位激活批归一化》。载于2018年 IEEE计算机视觉与模式识别会议（CVPR 2018），美国犹他州盐湖城，2018年6月18–22日，第 5639–5647 页。计算机视觉基金会 / IEEE计算机学会，2018年。1[9] Jo˜ao Carreira、Rui Caseiro、Jorge P. Batista 和 Cristian Sminchisescu。《采用二阶池化的语义分割》。载于 ECCV 2012——第12届欧洲计算机视觉会议。施普林格，2012年。2[10] Jun Cen、Peng Yun、Junhao Cai、Michael Yu Wang 和 Ming Liu。《面向开放世界语义分割的深度度量学习》。载于2021年 IEEE/CVF国际计算机视觉大会（ICCV 2021）。IEEE，2021年。2[11] Robin Chan、Krzysztof Lis、Svenja Uhlemeyer、Hermann Blum、Sina Honari、Roland Siegwart、Pascal Fua、Mathieu Salzmann 和 Matthias Rottmann。《Segmentmeifyoucan：异常分割基准》。载于 NeurIPS数据集与基准 1，2021年。1, 2[12] Robin Chan、Matthias Rottmann 和 Hanno Gottschalk。《语义分割中用于分布外检测的熵最大化与元分类》。载于国际计算机视觉会议（ICCV），2021年。3, 5, 6[13] Guangyao Chen、Peixi Peng、Xiangqian Wang 和 Yonghong Tian。《用于开放集识别的对抗互易点学习》。《IEEE模式分析与机器智能汇刊》，2022年。2[14] Liang-Chieh Chen、Yukun Zhu、George Papandreou、Florian Schroff 和 Hartwig Adam。《带空洞卷积的编码器-解码器用于语义图像分割的可分离卷积。收录于 ECCV，2018年。1, 5[15] Bowen Cheng、Maxwell D. Collins、Yukun Zhu、Ting Liu、Thomas S. Huang、Hartwig Adam 和 Liang-Chieh Chen。Panoptic-deeplab：一种简单、强大且快速的自底向上全景分割基线。收录于 IEEE/CVF计算机视觉与模式识别会议，2020年。1[16] Bowen Cheng、Ishan Misra、Alexander G. Schwing、Alexan- der Kirillov 和 Rohit Girdhar。用于通用图像分割的掩码注意力掩码变换器。收录于 IEEE/CVF计算机视觉与模式识别会议，2022年计算机视觉与模式识别会议，第1280–1289页。IEEE，2022年。1, 2, 4, 5[17] Bowen Cheng、Alexander G. Schwing 和 Alexander Kirillov。逐像素分类并非语义分割所需的一切。收录于2021年神经信息处理系统大会，2021年。1, 2, 3[18] Marius Cordts、Mohamed Omran、Sebastian Ramos、Timo Rehfeld、Markus Enzweiler、Rodrigo Benenson、Uwe Franke、Stefan Roth 和 Bernt Schiele。用于语义城市场景理解的 Cityscapes数据集。收录于2016年 IEEE计算机视觉与模式识别会议，2016年计算机视觉与模式识别会议，拉斯维加斯，内华达州，美国，2016年6月27日至30日，第3213–3223页。IEEE 计算机学会，2016年。1, 5[19] 戴继锋、Kaiming He 和孙剑。用于目标与背景联合分割的卷积特征掩码。收录于 CVPR，第3992–4000页，2015年。2[20] Cl´ement Farabet、Camille Couprie、Laurent Najman 和 Yann LeCun。学习用于场景标注的层次特征。《IEEE模式分析与机器智能汇刊》，2013年。1[21] Matej Grcic、Petra Bevandic、Zoran Kalafatic 和 Sinisa Segvic。通过基于合成负样本数据的鲁棒学习进行密集异常检测。CoRR，abs/2112.12833，2021年。3, 5, 6[22] Matej Grcic、Petra Bevandic 和 Sinisa Segvic。利用 Real NVP 生成的合成离群样本进行密集开放集识别。收录于第16届国际计算机视觉、成像与计算机图形理论及应用联合会议论文集。SCITEPRESS，2021年。2, 6, 7[23] Matej Grcic、Petra Bevandic 和 Sinisa Segvic。Densehy-brid：面向密集开放集识别的混合异常检测。收录于 ECCV 2022——第17届欧洲会议。施普林格，2022年。3, 5, 6[24] David Ha、Andrew M. Dai 和 Quoc V. Le。超网络。收录于 ICLR，2017年。2[25] Bharath Hariharan、Pablo Andr´es Arbel´aez、Ross B. Gir-shick 和 Jitendra Malik。同步检测与分割。收录于 ECCV 2014——第13届欧洲会议。施普林格，2014年。2[26] Dan Hendrycks、Steven Basart、Mantas Mazeika、Andy Zou、Joseph Kwon、Mohammadreza Mostajabi、Jacob Steinhardt 和 Dawn Song。面向真实世界场景扩展分布外检测。收录于国际机器学习会议，2022年国际机器学习会议。机器学习研究论文集，2022年。1, 3, 4, 6, 7[27] Dan Hendrycks 和 Kevin Gimpel。用于检测神经网络中误分类样本和分布外样本的基线。收录于 ICLR，2019年。2, 3, 6, 7[28] Dan Hendrycks、Mantas Mazeika 和 Thomas G. Dietterich。利用离群值暴露进行深度异常检测。收录于 ICLR，2019年。2, 5, 7

[29] Jaedong Hwang、Seoung Wug Oh、Joon-Young Lee 和 Bo-hyung Han。基于样本的开放集全景分割网络。收录于 IEEE计算机视觉与模式识别会议，CVPR 2021，第1175–1184页。计算机视觉基金会 / IEEE，2021年。1, 3, 6, 7[30] Sanghun Jung、Jungsoo Lee、Daehoon Gwak、Sungha Choi 和 Jaegul Choo。标准化最大逻辑值：一种简单而有效的城市场景分割中意外道路障碍物识别方法。收录于国际计算机视觉会议，ICCV，2021年。6[31] Alex Kendall 和 Yarin Gal。计算机视觉的贝叶斯深度学习需要哪些不确定性？收录于神经信息处理系统会议，2017年。6[32] Polina Kirichenko、Pavel Izmailov 和 Andrew Gordon Wil-son。归一化流为何无法检测分布外数据。收录于 NeurIPS，2020年。2[33] Alexander Kirillov、Yuxin Wu、Kaiming He 和 Ross B. Girshick。PointRend：作为渲染的图像分割。收录于 CVPR，第9796–9805页，2020年。2[34] Alex Krizhevsky、Ilya Sutskever 和 Geoffrey E. Hinton。使用深度卷积神经网络进行 ImageNet 分类。收录于《神经信息处理系统进展》第25卷：第26届神经信息处理系统年会（2012）。会议论文集，会议于2012年12月3日至6日在太浩湖、内华达州、美国举行，第1106–1114页，2012年。1[35] Kimin Lee、Honglak Lee、Kibok Lee 和 Jinwoo Shin。训练置信度校准分类器以检测分布外样本。收录于第6届国际学习表征会议，ICLR，2018年。2[36] Feng Li、Hao Zhang、Shilong Liu、Lei Zhang、Lionel M Ni、Heung-Yeung Shum 等。Mask DINO：迈向用于目标检测与分割的统一基于Transformer的框架。arXiv预印本 arXiv:2206.02777，2022年。3[37] Chen Liang、Wenguan Wang、Jiaxu Miao 和 Yi Yang。GMMseg：基于高斯混合的生成式语义分割模型。收录于 NeurIPS，2022年。3[38] Shiyu Liang、Yixuan Li 和 R. Srikant。提高神经网络中分布外图像检测的可靠性。收录于 ICLR，2018年。2, 3, 6[39] Tsung-Yi Lin、Michael Maire、Serge J. Belongie、James Hays、Pietro Perona、Deva Ramanan、Piotr Doll´ar 和 C. Lawrence Zitnick。Microsoft COCO：上下文中的常见物体。收录于 ECCV 2014——第13届欧洲会议。施普林格，2014年。1[40] Krzysztof Lis、Sina Honari、Pascal Fua 和 Mathieu Salz-mann。通过擦除道路障碍物来检测它们。CoRR，abs/2012.13633，2020年。6[41] Krzysztof Lis、Krishna Kanth Nakka、Pascal Fua 和 Math-ieu Salzmann。通过图像重合成检测意外情况。收录于国际计算机视觉会议，ICCV，2019年。3, 6[42] Weitang Liu、Xiaoyun Wang、John D. Owens 和 Yixuan Li。基于能量的分布外检测。收录于 NeurIPS 2020，2020年。2, 3, 5, 6, 7[43] Ze Liu、Yutong Lin、Yue Cao、Han Hu、Yixuan Wei、Zheng Zhang、Stephen Lin 和 Baining Guo。Swin Transformer：

采用移位窗口的层次化视觉变换器。在2021年IEEE/CVF国际计算机视觉会议（ICCV）上，页码 – 。IEEE，。Jonathan Long、Evan Shelhamer 和 Trevor Darrell。用于语义分割的全卷积网络。载于 IEEE计算机视觉与模式识别会议，CVPR 2015。IEEE计算机学会，2015年。1，3[45] Thomas Lucas、Konstantin Shmelkov、Karteek Alahari、Cordelia Schmid 和 Jakob Verbeek。生成模型的自适应密度估计。《神经信息处理系统进展》，32, 2019 2 。Andrey Malinin [46]和 Mark J. F. Gales。通过先验网络进行预测不确定性估计。载于神经信息处理系统年度会议，2018年。6[47] Jishnu Mukhoti 和 Yarin Gal。评估用于语义分割的贝叶斯深度学习方法。CoRR，abs/1811.12709，2018年。2[48] Eric T. Nalisnick、Akihiro Matsukawa、Yee Whye Teh、Dilan G¨or¨ur 和 Balaji Lakshminarayanan。深度生成模型知道自己不知道什么吗？载于第七届国际学习表征会议（ICLR），2019年。2[49] Gerhard Neuhold、Tobias Ollmann、Samuel Rota Bulo 和 Peter Kontschieder。用于街景语义理解的 Mapillary Vistas 数据集。载于 IEEE国际计算机视觉会议论文集，页码4990–4999，2017年。5[50] Marin Orsic 和 Sinisa Segvic。采用金字塔融合的高效语义分割。《模式识别》，110，107611，2021年 5：。[51] Pedro Oliveira Pinheiro、Tsung-Yi Lin、Ronan Collobert 和 Piotr Doll´ar。学习优化目标分割区域。载于ECCV，计算机科学讲义第9905卷，页码75–91，2016年。2[52]Matthias Rottmann、Pascal Colling、Thomas-Paul Hack、Robin Chan、Fabian H¨uger、Peter Schlicht 和 Hanno Gottschalk。语义分割中的预测误差元分类：通过softmax概率的聚合离散度量进行检测。载于IJCNN，页码1–9，2020年。3[53]Lukas Ruff、Jacob R. Kauffmann、Robert A. Vander-meulen、Gr´egoire Montavon、Wojciech Samek、Marius Kloft、Thomas G. Dietterich 和 Klaus-Robert M¨uller。深度与浅层异常检测的统一综述。《IEEE会议录》，109(5):756–795，2021年。2[54] Walter J. Scheirer、Anderson de Rezende Rocha、Archana Sapkota 和 Terrance E. Boult。迈向开放集识别。《IEEE模式分析与机器智能汇刊》，35(7):1757–1772，2013年。2[55] Walter J. Scheirer、Lalit P. Jain 和 Terrance E. Boult。开放集识别的概率模型。《IEEE模式分析与机器智能汇刊》，36(11):2317–2324，2014年。2[56] Joan Serr`a、David Alvarez、Vicenc¸ G´omez、Olga Slizovskaia、Jos´e F. N´u˜nez 和 Jordi Luque。基于似然的生成模型中的输入复杂度与分布外检测。载于 ICLR，2020年。2[57] Yu Tian、Yuyuan Liu、Guansong Pang、Fengbei Liu、Yuanhong Chen 和 Gustavo Carneiro。用于复杂城市驾驶场景异常分割的像素级能量偏置弃权学习。CoRR，abs/2111.12264，2021年。3，5，6

[58] Sagar Vaze、Kai Han、Andrea Vedaldi 和 Andrew Zisser-man。开放集识别：优秀的闭集分类器就是你所需要的一切。载于第十届国际学习表征会议（ICLR 2022），2022年。2，5 ˇ[59] Tomas Vojir、Tom´aˇs Sipka、Rahaf Aljundi、Nikolay Chumerin、Daniel Olmeda Reino 和 Jiri Matas。通过结合分割的部分图像重建进行道路异常检测。载于国际计算机视觉会议（ICCV），2021年。3，6[60] Yingda Xia、Yi Zhang、Fengze Liu、Wei Shen 和 Alan L. Yuille。先合成再比较：检测语义分割中的失效与异常。载于 ECCV 2020——第16届欧洲会议。施普林格，2020年。3，6[61] Jure Zbontar 和 Yann LeCun。通过训练卷积神经网络比较图像块来进行立体匹配。《机器学习研究杂志》，17:65:1–65:32，2016年。1[62] Oliver Zendel、Katrin Honauer、Markus Murschitz、Daniel Steininger 和 Gustavo Fern´andez Dom´ınguez。Wilddash——创建风险感知基准。载于 ECCV，页码407–421，2018年。2[63] Hongjie Zhang、Ang Li、Jie Guo 和 Yanwen Guo。用于开放集识别的混合模型。载于第16届欧洲计算机视觉会议（ECCV），2020年。2[64] Bolei Zhou、Hang Zhao、Xavier Puig、Tete Xiao、Sanja Fi-dler、Adela Barriuso 和 Antonio Torralba。通过 ADE20K数据集实现对场景的语义理解。《国际计算机视觉杂志》，127(3):302–321，2019年。1，5