arXiv:2304.03980v1 [cs.CV] 8 Apr 2023

面向激光雷达语义分割的持续学习：稀疏数据上的类别增量与由粗到细策略

Elena Camuffo 和 Simone Milani，信息工程系，帕多瓦大学，意大利{elena.camuffo,simone.milani}@dei.unipd.it

摘要

近年来，图像分类和分割领域的持续学习（*CL*）策略得到了广泛研究，研究者设计了多种创新方案来应对灾难性遗忘，例如知识蒸馏和自图像修复。然而，持续学习范式在点云上的应用仍未得到探索，尤其是采用能够捕捉激光雷达数据稀疏性及分布不均特征的架构时，仍有必要开展进一步研究。本文分析了应用于点云语义分割的类别增量学习问题，并比较了多种方法和最先进的架构。据我们所知，这是激光雷达点云语义分割领域类别增量持续学习的首个实例。我们将不同的持续学习策略适配于激光雷达点云并进行测试，涵盖经典的微调场景以及由粗到细学习范式。该框架在 *SemanticKITTI* 上采用两种不同的架构进行了评估 *[2, 16]*, , 所得结果与最先进的持续学习策略及标准离线学习相当。

1. 引言

近年来，配备自动视觉理解系统的自动驾驶发展，推动了目标识别和分割研究朝着更具通用性和可移植性的策略发展，即通过在线学习行为取代彼此独立的离线训练阶段。特别是，持续学习研究如何在成熟模型中融入新知识，并转向新的任务、标签或数据分布。事实上，大多数情况下都无法利用先前的训练数据，因为出于可移植性、隐私问题或高昂的标注过程等原因，这些数据通常不可用。因此，该过程通常通过保留先前知识，同时避免灾难性遗忘来完成。人们已经深入研究了不同的持续学习策略，用于分类 [11, 24, 29, 43] 和目-标检测 [23, 40, 45] 任务，而语义分割的相关示例则较少。其原因在于，相较于分类，语义分割是一项更困难的任务：分类旨在为每张图像提供一个单独的类别标签，而在分割中，已知类别和全新类别可能会同时出现在同一次采集（图像、点云等）中。事实上，处理这种异质性是针对语义分割考虑多种持续学习场景时所要解决的额外问题之一 [6, 13, 25, 36, 48]。迄今为止，持续学习通常应用于视觉任务中的图像。与点云数据相比，得益于其在空间中的规则分布、有限的内存需求和计算成本，图像更易于获取、标注和处理。相比之下，点云数据上的持续学习技术仍未得到充分探索，仅有少数尝试 [5, 8, 12, 26, 30], , 而类别增量持续学习的研究尚未受到关注。尽管将基于图像的网络架构简单地调整到点云上似乎并不困难，但激光雷达点云数据的稀疏性以及处理架构之间的显著差异，使得有必要重新开展相关研究。更确切地说，对于基于点的架构而言 [21, 41, 42],, 问题变得更加具有挑战性，因为原始点未经离散化便被直接处理（详见第 2节）。此外，点云数据集的固有特性——它们通过不同的方法并在不同的噪声条件下采集——可能会改变最终的准确率。

本文分析了点云语义信息分割中类别增量持续学习面临的挑战，并针对不同架构和划分方式调整了不同的持续学习解决方案。本研究的主要创新点如下。

- 这是类别增量持续学习首次应用于激光雷达语义分割，其中点稀疏性加剧了已知类别与全新类别共现所带来的问题。
- 它对缓解激光雷达点云中灾难性遗忘的不同技术进行了全面评估。

点云（从图像语义分割调整到具有新特性的领域）。

- 它同时考虑了通用的类别增量和由粗到细场景 [28, 34, 46, 50]。

本文结构如下：第2节概述近期最先进的研究成果，第3 节介绍主要的单头类别增量持续学习场景。第4 节分析不同的持续学习解决方案，概述不同的划分方式以及缓解灾难性遗忘的策略。最后，第5 节讨论采用不同技术获得的结果，并比较不同的数据集设置和持续学习策略；第6节给出结论。

2. 相关工作

激光雷达语义分割通常采用深度学习方法处理输入点，具体方式包括：离散化为体素、投影到二维图像，或直接处理 [3, 15]。基于体素的方法 [33, 47]通常用于结构良好的点云（例如，通过地面激光扫描采集的数据 [1]），并且可以轻松应用三维卷积。基于投影的方法 [10, 38, 51] 非常适用于激光雷达点云，因为这类点云以同心圆的形式采集，且可轻松投影到圆柱形 [51] 或球形 [38]表面。这两类方法在计算复杂度方面都很高效，但会将点分布近似为规则结构。直接处理数据的方法 [21, 41, 42] 避免了此类结构信息损失；在处理样本数量少且资源有限的情况时，这一点至关重要。近期研究采用四维卷积 [14] 或 Transformer [18] 来完成该任务，但这需要大量计算能力和存储空间；更轻量的方法则采用混合策略处理点，先提供逐体素预测，再用逐点标签进行细化 [52, 53]。此类模型已广泛应用于多种学习方法，包括类别均衡正则化 [4, 10] 知识蒸馏 [20] 和开放世界三维场景理解 [5]。此外，一些研究也在三维数据上采用了增量学习设置，应用于目标检测 [12], 地点识别 [26],遥感 [30] 和少样本学习 [8]。然而，尚无先前研究探索激光雷达语义分割中的类别增量持续学习。

类别增量持续学习（持续学习）在图像分类 [11, 24, 29, 43], 目标检测 [23, 40, 45] 任务领域引发了越来越多的研究兴趣。许多研究在顺序学习具有不同输入模式的样本时，发现了灾难性干扰问题。减少遗忘最常用的策略包括回放方法 [32, 43] ，即存储样本记忆以便在后续步骤中重放，以及非样本记忆方法；在这些方法中，参数隔离会为每个任务分配不同的模型参数，以避免可能的遗忘；正则化方法 [7, 24, 29, 48], 则通过增加额外的正则化损失来巩固先前数据中的知识。这类方法避免存储原始输入，优先保障隐私，并降低内存需求。

语义分割中的持续学习相关文献较少 [6, 13, 25, 36, 48]。研究主要聚焦于类别增量语义分割，即从新数据中学习新类别。 [36]将该问题形式化，并采用参数冻结、知识蒸馏等正则化方法加以解决。 [25] 将知识蒸馏与类别重要性加权方案相结合，以强调困难类别上的梯度。 [37]对潜在空间进行正则化，以改善类别条件特征的分离。Cermelli 等人研究了 [7] 中背景类别的分布偏移，并在 [6]中以弱监督方式解决这一问题。Maracani等人 [32] 提出使用生成网络或网络爬取图像来重放旧类别。

由粗到细语义分割在少数先前研究中有所探索，包括用于基于部件的正则化的方法[28, 50] ， 以及将粗粒度类别细化为细粒度类别的方法 [34, 46]。先前的方法还提出了不同的类别层级构建方式。 [44]根据语义信息估计先验； [4], 则根据网络误分类事后构建层级。一些研究也展示了将由粗到细方法用于持续学习的做法 [44, 49], ，但此前尚无文献探索其在点云数据上的应用。

3. 问题定义

给定一个输入点云，即，一组 N 3D 点$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$，以及一组候选语义标签 $Y = \{y_1, y_2, \ldots, y_N\}$，语义分割的目标是为每个输入点$\mathbf{x}_i$ 关联一个语义标签 y_i。每个标签均属于给定的语义类别子集 C，其中还包含一个特殊类别 b （空类，对应于图像中的背景 ）。空类通常用于标识超出激光雷达精度范围或属于未知类别的物体。为清晰起见，我们将其称为背景，与图像中的用法一致，以避免与空体素和掩码点混淆。语义分割任务通常通过使用合适的深度学习模型 $\mathcal{M}$ 来完成： $X \mapsto Y_C$，该模型通常由特征提取器 $\mathcal{E}$ 和后接的解码模块 $\mathcal{D}$, $\mathcal{M} = \mathcal{D} \circ \mathcal{E}$ 组成。

在标准监督学习中，模型通过对固定且完整的数据集 $T \subset X \times Y$ 进行标准离线训练，一次性完成学习。

而在类别增量学习中，我们假设训练分多个步骤进行，并且每一步训练时只能使用部分训练数据 $k = 0, \ldots, K$。具体而言，我们从初始步骤 $k = 0$ 开始，此时只能使用涉及部分类别的训练数据

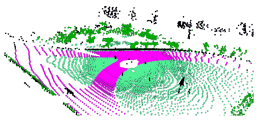
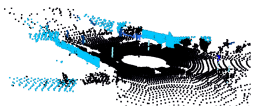
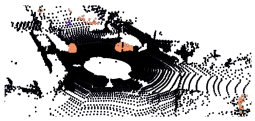
	训练子集 T_0	训练子集 T_1	训练子集 T_2	验证集
		SemanticKITTI 火车		SemanticKITTI 验证集
序列	D_0 ={01, 02, 03}	D_1 ={04, 05, 09, 10}	D_2 ={00, 06, 07}	D ={08}
已标注类别	C_0 = {道路、停车区、人行道、其他地面、植被，地形}	C_1 = {建筑物、围栏、树干，杆，交通标志}	C_2 = {自行车、摩托车、卡车, 其他车辆, 人, 骑自行车的人、摩托车骑手、汽车}	$C = \{C_0 \cup C_1 \cup C_2\}$
				
# 点云	6563	4623	4541	4071
# 点	355280	375140	121494	19130
% 已标注点	81.73%	19.21%	8.82%	95.53%

表1。SemanticKITTI [2] 经验中的划分。训练集类别和序列根据点频率和基数分为三部分，同时遵循 [25]。下划线表示在 Cityscapes 中不存在的类别 [9], 因此也不存在于原始划分中。

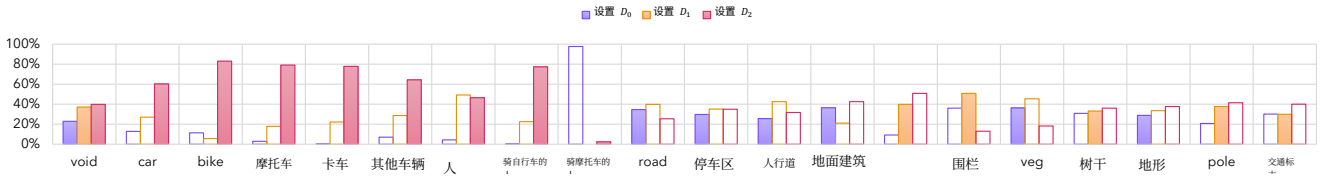


图1. 每个训练数据集划分 D_k中点云百分比的直方图。百分比按整个数据集 $D_0 \cup D_1 \cup D_2$中特定类别的点总数进行归一化 c。

，这些类别属于所有类别 $C_0 \subset C$ （我们假设 $^b \in C_0$）。我们将初始步骤训练得到的模型记为 $\mathcal{M}_0$：$X \mapsto Y_{C_0}$，$\mathcal{M}_0 = \mathcal{D}_0 \circ \mathcal{E}_0$。在一般步骤 k 中，会将一组新类别 C_k 加入截至该步骤所学到的类别集合 $C_{0\to(k-1)}$，从而扩展可学习类别集合 $C_{0\to k} = C_{0\to(k-1)} \cup C_k$ （我们假设 $C_{0\to(k-1)} \cap C_k = \emptyset$）。第 k 步训练完成后的模型为 $\mathcal{M}_k$：$X \mapsto Y_{C_{0\to k}}$，其中 $\mathcal{M}_k = \mathcal{D}_k \circ \mathcal{E}_k$。
一些方法在初始步骤之后仅训练解码器，同时保持编码器冻结。在我们的方法中，编码器和解码器都会更新，但只有最后一层会发生改变。这是因为在随机对象数据集上进行训练时，编码器通常保持冻结，并且第一步与后续步骤之间的数据比例向第一步倾斜。相反，在我们的划分方式中（以及一般而言，在自动驾驶数据集中），不同步骤之间的数据集基数、样本数量和类别频率是均衡的。

3.1. 场景

针对图像语义分割的持续学习，已有多种场景被提出 [35,36]。在每种场景中，原始数据集 T 被划分为 K 组，分别对应各个学习步骤；标签总集通常也被划分为 K 组。在学习步骤 k，训练过程仅使用元组 (T_k, C_k) 更新模型。

顺序：第 k 个经验中的样本带有所有 $\leq k$ 组的标签，而标签为 $> k$ 的点则完全不存在。每个学习步骤都包含一组独特的样本，其中的点属于当前步骤或先前步骤中见过的类别。

顺序掩码：第 k 个经验阶段中的样本带有第 k 个经验阶段的标签，而类别 $< k$ 的标签对应的点被标记为未标注，标签为 $> k$ 的点则完全不存在。在此设置中，每个学习步骤都包含一组独特的样本，这些样本中的点要么属于新类别，要么属于未标注类别；模型不会预测该类别，并且会在结果和训练过程中将其屏蔽。

不相交：第 k 个经验阶段中的样本带有第 k 个经验阶段的标签，而类别 $< k$ 的标签对应的点被标记为背景，标签为 $> k$ 的点则完全不存在。在每个学习步骤中，独特的训练样本集与顺序设置中的相同，但旧类别会归入一个共同标签（背景，而非像序列掩码场景中那样未标注），且该标签的分布会随每个步骤而变化。

重叠：在每个学习步骤中 k，所有数据集样本均可用，但只有当前步骤类别 C_k 被标注为正确标签，其余样本（$C_j \neq C_i$）均被标注为背景标签，无论它们属于经验 $j < k$ 还是 $j > k$。在此设置下，每个训练步骤包含

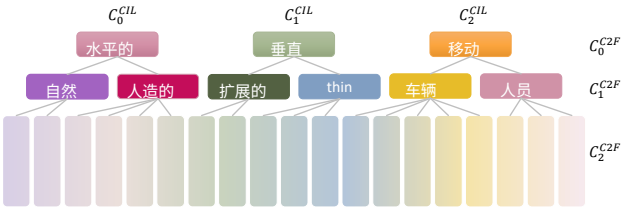


图 2。SemanticKITTI **C2F** 划分中的类别层次结构。每个粗类别通过将在**CIL** 划分中属于同一学习步骤的类别进行分组构成。序列划分 D_k 按完全相同的顺序进行考虑。

至少有一个点属于集合 C_k中某个类别的所有点云，其中仅对该集合中的类别进行标注，其余部分设为背景。本文报告了这一场景以供比较，并假设数据永远不能在不同经验之间共享。

由粗到细：在每个学习步骤 k中，所有点均已标注，但其类别相对于上一步得到了细化。在步骤 k 中，粗标签 C_{k-1} 被细类别中的标签替代 C_k，$|C_k| > |C_{k-1}|$。每个粗类别都对应于其唯一的一组细类别。

4. 方法

我们将**RandLA-Net** [21], 这一最具代表性的基于点的架构之一作为参考，以便从不同于图像的视角来构建问题（即，使用多层感知机代替卷积设置）。RandLA-Net是一种高效的轻量级基于点的网络，由基于多层感知机的编码器-解码器结构组成，在内存和计算方面实现了极高的效率。此外，我们还评估了**Cylin-der3D** [52, 53] 基于体素的架构以作比较。尽管如此，该框架可以应用于任何用于点云语义分割的架构之上。

SemanticKITTI [2, 16] 被选为参考数据集，因为它是自动驾驶领域中最受欢迎的激光雷达语义分割基准之一。Se-manticKITTI 包含 43,552个经过密集标注的激光雷达扫描，其中 19, 130 个用于训练，4,071 个用于验证（我们将其用于测试，正如所有竞争性工作所做的那样，因为测试标签尚未公开）。

4.1. 数据集划分

我们对数据集进行划分，以定义持续学习的增量式经验。由于尚无相关先前研究，我们提出了SemanticKITTI 的两种划分方式：一种依据不同序列采集中的类别基数和频率，另一种基于由粗到细的语义划分。我们将前者称为**CIL**，因为它基于对 [25]，的实际划分；将后者称为**C2F**。

4.1.1 语义划分（CIL）

表1展示了 SemanticKITTI 的第一种划分方式，该方式依据不同序列采集中的类别基数和频率。这种分组基于 [25] 中提出的 Cityscapes [9] 的实际划分。Cityscapes 是一个用于图像语义分割的自动驾驶数据集，其类别集与 SemanticKITTI 相似。具体而言，在 T_0 中，天空被停车区和其他地面取代；在 T_1 中，交通信号灯和墙被树干取代；在 T_2 中，火车、公交车和骑行者则分别被摩托车骑手、自行车骑手和其他车辆取代。这些类别在表1中以带下划线标示。序列划分旨在使每组中已标注类别的数量最大化，同时保持语义一致性，并平衡每个训练子集中的点数和采集次数。图1展示了各组中已标注点的百分比；针对每个类别，该比例按特定类别的基数进行归一化 c。

4.1.2 由粗到细划分（C2F）

图 2 展示了SemanticKITTI 的第二种划分方式，其中通过将标签拆分为三个学习步骤来进行分组，并对标签进行聚类以保证语义一致性。第一组粗粒度集合体现了 **CIL** 的语义：每个 C_k^{CIL}中的类别构成一个粗类别 C_0^{C2F}。C_1^{C2F}是通过将粗粒度类别进一步划分为各包含 2 个中等粒度类别的集合而得到的；最后，细粒度集合 $C_2^{C2F} = C$，即，包含 SemanticKITTI 的所有类别。请注意，该划分方式与 **CIL** 设置共享序列的数据集划分方式，该设置已在表 1 中报告。

4.2. 持续学习策略

在标准训练中，背景 点通常会从学习过程中屏蔽掉，并使用交叉熵目标函数优化模型权重。

$$\mathcal{L}_{ce} = -\frac{1}{|T_k|} \sum_{\mathbf{X}_n \in T_k} \sum_{c \in C_{k-1}} \mathbf{Y}_n[c] \cdot \log(\mathcal{M}_k(\mathbf{X}_n)[c]) \quad (1)$$

其中，$\mathbf{Y}_n[c]$ 是独热编码真实标签，而$\mathcal{M}_k(\mathbf{X}_n)[c]$ 是类别 c 的预测分数。在增量学习设置中，我们假设在每个增量训练步骤 k 中，只有新类别 C_k 的样本可用。为将这些新信息纳入模型，权重从先前模型 $\mathcal{M}_{k-1}$ ($k \geq 1$) 初始化，然后通过优化标准目标函数 $\mathcal{L}_{ce}(\mathcal{M}_k; C_{0\to k}, T_k)$，利用当前训练划分 T_k中的数据学习新类别集 $C_{0\to k}$。然而，仅针对新类别进行朴素微调会导致灾难性遗忘，因此，将过去类别（无标签）的点作为 背景纳入后，我们可以通过自我预测旧标签来恢复先前的知识。为此，我们考虑两种策略：知识蒸馏（KD） [29]和背景

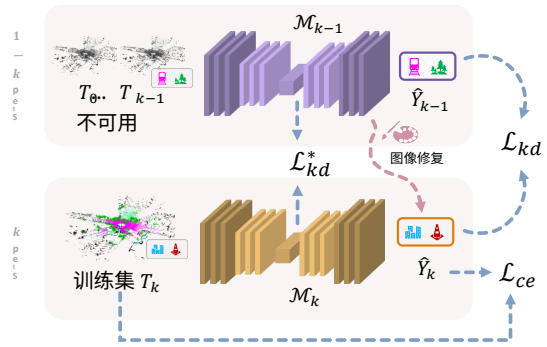


图 3. 所采用的持续学习策略概述。在当前学习步骤（k）中，只有训练集 T_k 和先前模型 $\mathcal{M}_{k-1}$ 可用。

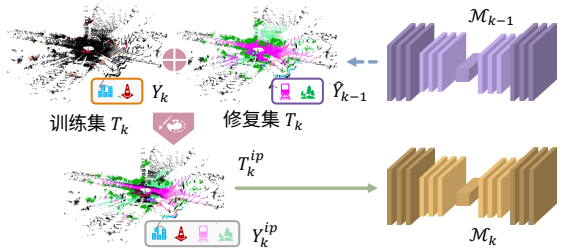


图 4. 背景自修复。旧类别C_{k-1} 的标签由先前模型 $\mathcal{M}_{k-1}$ 预测，并修复到当前训练集 T_k的背景 标签上。

自修复 [32]。图 3 直观总结了这些技术。

4.2.1 知识蒸馏

知识蒸馏（KD）是将知识从一个模型迁移到另一个模型（通常规模较小）的过程 [19,20]。在我们的场景中，我们通过应用知识蒸馏，在学习新类别的同时回忆过去类别，并处理背景偏移现象。旧标签由前一学习步骤的模型恢复 $\mathcal{M}_{k-1}$。KD 被建模为一个额外的目标函数 $\mathcal{L}_{kd}$，并由超参数λ调节，从而得到模型 $\mathcal{M}_k$的最终目标函数：

$$\mathcal{L} = \mathcal{L}_{ce}(y_k, \hat{y}_k) + \lambda \cdot \mathcal{L}_{kd}(\hat{y}_{k-1}, \hat{y}_k) \tag{2}$$

其中， y_k 是步骤 k的真实标签，$\hat{}y_k$ 是步骤 k 的预测结果，$\hat{}y_{k-1}$ 是步骤 $k-1$的预测标签。在步骤 $k=0$ 中不应用 KD，因为在该阶段我们缺乏关于过去类别的任何先验知识。

输出级蒸馏 ($\mathcal{L}_{kd}$) 通常被建模为额外的交叉熵 [17]，，其中过去类别的标签来自先前的模型预测结果：

$$\mathcal{L}_{kd} = -\frac{1}{|T_k|} \sum_{\mathbf{x}_n \in T_k} \sum_{c \in C_{k-1} \cup C_k} \mathcal{M}_{k-1}(\mathbf{X}_n)[c] \cdot \log(\mathcal{M}_k(\mathbf{X}_n)[c]) \tag{3}$$

其中， $\mathcal{M}_k(\mathbf{X}_n)[c]$ 和 $\mathcal{M}_{k-1}(\mathbf{X}_n)[c]$ 分别是使用当前模型和先前模型获得的类别 c预测分数。知识蒸馏可以通过屏蔽每个模型 $\mathcal{M}_k$ 在类别 C_k 上的预测结果，或合并所有未知类别 [7]的预测结果来建模。另一种表述方式是同时更新先前模型，以优化预测结果 [48]。对于由粗到细的持续学习，在应用交叉熵损失之前，会对同一粗类别内各类别的预测结果求和。

特征级蒸馏（$\mathcal{L}^*_{kd}$）通常被建模为当前模型和先前模型的特征之间的 L_p （L_2 或 L_1）范数：

$$\mathcal{L}^*_{kd} = \frac{||\mathcal{E}_{k-1}(\mathbf{X}_n) - \mathcal{E}_k(\mathbf{X}_n)||_p}{|T_k|} \tag{4}$$

其中， $\mathcal{E}_k(\mathbf{X}_n)[c]$ 和 $\mathcal{E}_{k-1}(\mathbf{X}_n)[c]$ 分别是当前模型和先前模型的特征。采用这种设计的动机在于，该策略能够学习特征图而非分类图，而距离损失则能对这类嵌入进行对齐。还可以结合特征层级和输出层级的知识蒸馏，以同时发挥两者的作用。

4.2.2 背景自修复

背景自修复是一种简单的自我教学机制，可为当前背景样本提供伪标签， [32], 减少背景偏移，同时带来类似于知识蒸馏的正则化效应。在每个步骤 k 使用训练集 T_k时，我们获取每个真实标注图中的背景点 Y_k 并使用前一个模型的相应预测结果为其标注 $\mathcal{M}_{k-1}$（图 4）。更正式地说，我们将每个原始标签图Y_k 在步骤 $k>0$ 可用的修复版本 Y_k^{ip}：

$$Y_k^{ip}[i] = \begin{cases} Y_k[i] & \text{if} \quad Y[i] \in C_k \\ \rho \cdot \hat{Y}_{k-1}[i] & \text{otherwise} \end{cases} \tag{5}$$

其中 $(X_k,\ Y_k) \in T_k$。 $[i]$ 表示 T_k 中的第 i个样本， ρ定义了修复规则。 ρ 设计如下：

$$\rho = \begin{cases} 1 & \text{if} \quad \hat{\mathcal{Y}}^*_k[i] - \hat{\mathcal{Y}}^{**}_k[i] > \tau_1 \quad \text{and} \quad \hat{\mathcal{Y}}^*_k[i] > \tau_2 \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

$\hat{}\mathcal{Y}^*_k[i]$ 和$\hat{}\mathcal{Y}^{**}_k[i]$ 表示步骤 k中Softmax 预测结果 $\mathcal{Y}_k$ 里的第一大值和第二大值。在步骤$k=0$ 中，标签不会进行图像修复，因为在该阶段我们还没有任何先验

Cylinder3D									RandLA-Net								
	步骤 0	步骤 1			第 2 步					步骤 0	步骤 1			步骤 2			
方法	$mIoU_0$	$mIoU_0$	$mIoU_1$	$mIoU_{0,1}$	$mIoU_0$	$mIoU_1$	$mIoU_2$	$mIoU_{0,1,2}$	方法	$mIoU_0$	$mIoU_0$	$mIoU_1$	$mIoU_{0,1}$	$mIoU_0$	$mIoU_1$	$mIoU_2$	$mIoU_{0,1,2}$
基线†	—	—	—	—	—	—	—	54.1	基线†	—	—	—	—	—	—	—	47.2
顺序	55.4	52.7	46.3	**50.1**	52.8	50.6	32.7	**44.9**	顺序	49.0	57.0	39.0	**48.8**	58.0	48.4	28.2	42.9
顺序掩码	55.4	0.0	77.4	20.4	0.0	0.0	25.7	10.8	顺序掩码	49.0	0.0	48.5	22.0	0.0	0.0	31.4	13.2
不相交	55.4	0.0	63.3	20.0	0.0	0.0	25.2	10.6	不相交	49.0	0.0	37.9	18.0	0.0	0.0	26.1	11.0
输出知识蒸馏 [29]	55.4	45.6	41.1	43.5	47.4	32.0	24.5	33.7	输出知识蒸馏 [29]	49.0	56.2	38.0	47.9	58.5	49.1	31.0	**44.4**
自修复 [32]	55.4	51.2	47.1	49.4	46.2	36.8	26.9	35.6	自修复 [32]	49.0	56.1	34.9	46.5	57.5	43.2	22.6	39.0

表 2. **Cylinder3D** 和**RandLA-Net (CIL)**的结果。**粗体** 表示最佳结果，下划线表示次佳结果。†: 使用我们的配置重新训练的基线使用我们的配置。

方法	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	**mIoU** ↑	**Std Dev.** σ ↓
基线†	90.7	1.6	14.6	53.9	37.1	30.5	54.2	0.0	90.9	35.4	75.1	2.2	84.0	46.3	84.9	47.1	69.7	51.7	27.6	47.2	30.1
顺序	89.3	0.0	12.5	59.4	33.0	8.0	22.3	0.0	90.5	26.0	73.9	0.8	84.2	47.8	85.1	51.9	71.6	49.3	8.7	42.9	33.0
输出知识蒸馏 [29]	88.6	0.0	11.6	52.6	32.2	10.2	52.5	0.0	90.4	28.5	73.9	0.0	84.6	47.0	85.0	55.9	73.3	51.4	6.7	**44.4**	**32.8**
输出无监督知识蒸馏 [7]	88.0	0.0	7.3	27.2	17.8	0.9	16.9	0.0	90.3	29.8	74.3	2.5	83.3	47.4	84.8	50.8	72.8	44.0	1.3	38.9	34.3
输出交叉知识蒸馏 [48]	86.5	0.0	11.6	59.3	21.0	2.5	45.2	0.0	90.8	32.3	75.1	0.2	83.0	44.3	85.0	46.0	72.6	46.5	8.9	42.7	33.0
特征知识蒸馏(L_2)	88.4	0.0	13.2	57.0	16.4	9.0	35.1	0.0	90.3	22.4	73.9	0.5	82.5	44.1	84.8	52.6	72.8	49.3	0.1	41.7	33.6
特征知识蒸馏(L_1)	87.5	0.0	10.9	36.1	35.9	0.0	39.9	0.0	90.4	32.1	74.2	0.1	82.4	40.7	85.7	53.4	75.2	46.6	1.1	41.7	33.4
两者知识蒸馏 [36]	85.3	0.0	10.2	42.9	20.2	0.0	33.9	0.0	90.0	24.8	72.9	0.4	82.7	42.8	84.1	43.7	72.0	44.5	0.0	39.5	33.3
自修复 [32]	86.3	0.0	7.3	54.5	5.3	0.1	27.0	0.0	90.4	25.0	74.1	1.6	82.9	42.3	82.7	45.3	71.1	45.5	0.0	39.0	34.4

表 3. **RandLA-Net (CIL)**的逐类别平均交并比结果。**粗体** 表示最佳结果，下划线表示次佳结果。†: 使用我们的配置重新训练的基线。

过去类别的知识。在执行背景修复时，每个集合 $T_k^{ip} \subset X \times Y_{C_{0\to k}}$ $(k>0)$ 都包含图像修复后的 T_k 的所有样本。

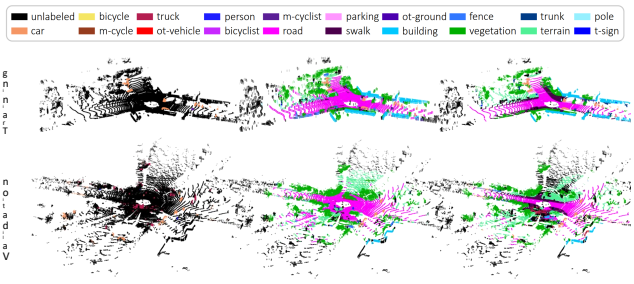
5. 实验结果

对于实验评估，我们使用 RandLA-Net [21]（一种广泛使用的轻量级基于点的架构），并将结果与 Cylinder3D [52, 53]（一种基于体素并采用逐点预测细化的先进模型）进行比较。请注意，该方法可泛化到任何其他用于点云的分割架构。我们使用 Se-manticKITTI [2, 16] 作为参考数据集，并根据 **CIL** 和**C2F** 划分，按照第 4.1节所述，将类别和数据划分到各个学习步骤中。模型性能通过平均交并比（mIoU）进行衡量 [31]。下文中，我们使用记号 $mIoU_k$ 表示在类别集合 C_k上计算得到的 mIoU。

我们使用 Adam优化器及各自的标准学习率策略、动量和权重衰减来训练 RandLA-Net 和 Cylinder3D。两个模型的初始学习率分别设为 0.01 和 0.001。增量训练过程中，学习率逐步降低，直至学习——率按照幂为0.95的多项式衰减规则递减。需要注意的是，在步骤 k 中，学习率会初始化为步骤 $k-1$ 的最后一个学习率，以更好地保留先前的权重。我们将批次大小设为 3，并让每个学习步骤 k训练 $2\times|C_k|$个周期，周期数与待学习的新类别数量成正比。我们使用 PyTorch [39] 实现模型，并在配备 CUDA 11.6 的 NVIDIA GeForce RTX 3090Ti 图形处理器上训练这两个模型。使用 RandLA-Net 训练每个学习步骤约需 5 小时，使用 Cylinder3D 则约需 30 小时。代码见：https://github.com/LTTM/CL-PCSS。在以下章节中，我们将讨论使用 **CIL** 和**C2F** 划分方式得到的结果。基线结果是通过采用上述配置，从头开始在包含所有类别的整个数据集上训练模型得到的。

5.1. CIL

我们按照**CIL**配置开展了第一组实验。首先，我们设计了标准微调方法、知识蒸馏和自修复方法，以比较 Cylinder3D 与 RandLA-Net。结果见表2。总体而言，所得性能反映了其表现



不相交　　自修复　　顺序

图 5. 使用**CIL** 和**Cylinder3D**得到的一些定性结果。

在相应的基于图像的设置中。如预期所示，不相交场景在所有学习步骤中的表现都最差，深受背景偏移现象的影响。顺序掩码策略略微提升了结果，同时在纳入新类别时再次表现出灾难性遗忘的影响。

总体而言，尽管 Cylinder3D 在非增量场景中的表现更好，但在持续学习框架下，其性能仍逊于 RandLA-Net。基线结果显示，顺序设置下性能大幅下降，平均交并比损失了9.8%（微调的最佳情况）。此外，架构规模和体素处理使训练时间延长了 6 倍。知识蒸馏能够部分恢复先前类别的知识，同时在学习新类别时保持稳定。相比之下，自修复的表现更好，也更能兼顾新类别。请注意，平均交并比$_0$ 在各个步骤中都能通过这两种方法有效恢复，而平均交并比$_1$ 则几乎无法恢复。这在一定程度上是因为步骤 $k=0$ 中的标签对应的是具有大量点的稳健类别，因此，采用通用的灾难性遗忘缓解方法便足以恢复知识。相反，步骤 $k=1$ 中已标注点的数量大幅减少，这反映出从先前训练中获得良好恢复效果的难度不断增加。自修复很少为旧类别生成伪标签，未能保留预测结果的不确定性，并对新类别的学习造成不利影响。引入修复规则（表 5）带来了小幅改进，避免将严重错误的预测当作真实标签用于后续步骤。这样一来，只有在预测准确时，先前知识才会得到少量恢复，并以比知识蒸馏更强的方式融入当前模型。最后，这两种方法的表现都逊于顺序设置。图5。展示了自修复策略的定性结果。

另一方面，RandLA-Net 非常适用于持续学习设置；在序列设置中，相较于基线，平均交并比仅下降了 4.7%。令人惊讶的是，通过蒸馏恢复知识可达到 44.4% 的平均交并比，使序列设置的平均交并比提升 1.5%，与基线的差距仅为 2.8%。这同样是由每个划分的标签频率所致：

方法	步骤 0	步骤 1			步骤 2			
	mIoU$_0$	mIoU$_0$	mIoU$_1$	mIoU$_{0,1}$	mIoU$_0$	mIoU$_1$	mIoU$_2$	mIoU$_{0,1,2}$
基线†	—	—	—	—	—	—	—	47.2
重叠	60.2	0.0	51.6	**23.4**	0.0	0.0	32.2	**13.6**
不相交	49.0	0.0	37.9	18.0	0.0	0.0	26.1	11.0
输出知识蒸馏 [29]	49.0	56.2	38.0	47.9	58.5	49.1	31.0	**44.4**
输出无监督知识蒸馏 [7]	49.0	56.1	35.3	46.6	59.1	45.4	19.8	38.9
输出交叉知识蒸馏 [48]	49.0	56.0	35.4	46.7	59.3	45.7	28.3	42.7
特征知识蒸馏（L$_2$）	49.0	55.9	40.7	**49.0**	57.5	45.7	27.4	41.7
特征知识蒸馏 (L$_1$)	49.0	55.5	36.4	46.8	59.6	44.8	26.3	41.7
两者知识蒸馏 [36]	49.0	56.5	33.9	46.2	57.4	42.7	24.1	39.5

表 4。**RandLA-Net(CIL)**。**粗体** 表示最佳结果，下划线表示次佳结果。†：基线使用我们的配置重新训练。

τ_1	τ_2	步骤 0	步骤 1			步骤 2			
		mIoU$_0$	mIoU$_0$	mIoU$_1$	mIoU$_{0,1}$	mIoU$_0$	mIoU$_1$	mIoU$_2$	mIoU$_{0,1,2}$
0	**0**	55.4	50.7	39.6	45.8	47.7	42.8	19.4	34.5
0	**0.7**	55.4	48.4	37.9	43.6	44.0	38.7	25.0	34.6
0.2	**0**	55.4	55.8	40.5	48.9	45.2	35.9	28.0	35.5
0.2	**0.7**	55.4	51.2	47.1	**49.4**	46.2	36.8	26.9	**35.6**

表5. 关于使用**Cylinder3D** 进行自修复的消融实验。 τ_1， τ_2 是图像修复的参数。**粗体** 表示最佳结果。

方法	步骤 0	步骤 1		步骤 2		
	mIoU$_0$	mIoU$_0$	mIoU$_1$	mIoU$_0$	mIoU$_1$	mIoU$_2$
基线†	—	—	—	—	—	47.2
微调	86.7	74.6	**74.6**	51.0	51.0	**47.1**
输出知识蒸馏	86.7	52.9	52.9	48.8	48.8	45.1
特征知识蒸馏（L$_1$）	86.7	57.2	57.2	47.1	47.1	42.8

表6。**RandLA-Net** 在**SemanticKITTI (C2F)**上。**粗体**表示最佳结果。†：使用我们的配置重新训练的基线。

早期步骤包含大量已标注样本，而后期步骤仅包含少量已标注样本。当仅使用 $\mathcal{L}_{ce}$ 进行训练时，无论采用顺序微调还是自修复，网络都会受到这种不平衡的影响。

为此，我们使用 RandLA-Net 围绕知识蒸馏开展了大量实验，同时考察了蒸馏输出预测和中间特征两种情况。表 4 报告了各学习步骤的总体平均交并比结果，而表 3 报告了最后一个学习步骤的各类别结果。两种蒸馏方法都持续提升了总体结果。总体表现最佳的是标准的输出层级蒸馏，但特征层级的知识蒸馏表现也相当不错。不过，总体而言，各类别结果表明，增量设置会加剧高频类别与低频类别之间的性能差距

方法	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	**mIoU** ↑	**Std Dev.** σ ↓
基线†	90.7	1.6	14.6	53.9	37.1	30.5	54.2	0.0	90.9	35.4	75.1	2.2	84.0	46.3	84.9	47.1	69.7	51.7	27.6	47.2	30.1
微调	89.7	6.2	17.4	47.0	24.8	28.3	65.5	0.0	88.8	30.7	73.0	0.6	83.6	45.8	87.2	55.4	75.1	50.8	24.4	**47.1**	**30.7**
输出知识蒸馏	90.3	0.2	16.6	53.2	30.0	14.8	42.1	0.0	90.7	26.9	74.9	0.1	84.5	43.9	86.5	56.8	73.9	52.0	21.1	45.1	31.9
特征知识蒸馏(L_1)	89.2	0.0	12.5	65.7	29.6	3.6	41.4	0.0	91.2	37.9	74.9	0.3	84.8	45.4	86.2	50.1	72.9	48.7	18.2	44.9	32.7

表 7. 使用**C2F** 和**RandLA-Net**得出的各类别平均交并比结果。**粗体** 表示最佳结果。†：使用我们的配置重新训练的基线。

step	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	vegetation	terrain	building	fence	trunk	pole	traffic-sign	**mIoU** ↑ [31]	**Std Dev.** σ ↓	**PA** ↑ [22]	**PP** ↑ [27]
0	85.1								95.6						79.5					86.7	8.2	93.0	92.6
1	89.1					43.4			92.2			89.8			81.7		51.3			74.6	21.5	80.4	89.6
2	89.7	6.2	17.4	47.0	24.8	28.3	65.5	0.0	88.8	30.7	73.0	0.6	87.2	75.1	83.6	45.8	55.4	50.8	24.4	47.1	30.7	56.2	67.7

表8. 每类 Co　　粗到细结果，标准微调。不同颜色对应不同的粗粒度划分。

。例如，类别 汽车 始终取得良好结果，平均交并比约为 90%；而类别 自行车 （属于同一分区）在基线上的平均交并比仅为 1.6%，在增量设置下则从未被正确分类。同样，交通标志在基线上的平均交并比为 27.6%，但应用增量学习后，其性能大幅下降；在自修复方法中，该类别从未被正确分类。事实上，基线模型取得了最低的标准差。

最后值得一提的是，与不相交设置相比，重叠设置的结果（见表 4）：在整个数据集上火车，正如预期的那样带来了提升（最终平均交并比提高 2.6% ），但仍然受到灾难性遗忘的影响。

5.2. 由粗到细

粗到细划分的结果见表 7。对粗粒度类别进行微调，所得结果与标准基线训练相当（47.1% 对比 47.2%平均交并比），尽管类别间的标准差较大。请注意，此处的微调是在表 1 所示的拆分序列上进行的，而一般的由粗到细方法依赖于重叠方法。与 **CIL** 不同，直接应用知识蒸馏在此会导致性能不佳，无论是在输出层级还是中间特征空间应用，平均交并比分别下降 2.0% 和 2.2% 。因此，要取得显著提升，此设置需要采用更复杂的损失函数。表 8 报告了 **C2F** 微调在平均交并比、点准确率（PA） [22] 和点精确率（PP） [27]方面的结果。第一步的结果准确率较高且标准差较低；随着学习步骤增加，标准差也随之增大（截至步骤 $k = 1$，类别分布较为均衡）。这一结果表明，增加中间步骤可能有助于提升最终性能，而这通常需要进一步细化。

6. 结论

本文研究了激光雷达语义分割中的类别增量持续学习问题。我们以语义一致的方式对 SemanticKITTI 进行了正式划分，并评估了针对不同防止灾难性遗忘技术的持续学习策略。对 RandLA-Net 和 Cylinder3D 性能的比较表明，前者（基于点且轻量）更适合类别增量设置。在我们提出的 SemanticKITTI 划分方案上进行的实验，证明了持续学习策略在缓解灾难性遗忘方面的有效性，即使是在稀疏数据上也是如此。总体而言，该问题仍需进一步改进，因为其性能低于标准单步训练所达到的水平。未来的工作将通过在先前经验与当前经验之间引入类别平衡约束，以及针对激光雷达点云专门设计的几何约束，来扩展实验框架。

致谢

本研究部分获得欧盟通过 NextGenerationEU 的意大利国家复苏与韧性计划（NRRP）提供的支持；该计划属于“未来电信”合作项目（PE0000001——“RESTART”计划）。

参考文献

[1] Iro Armeni、Ozan Sener、Amir R Zamir、Helen Jiang、Ioannis Brilakis、Martin Fischer 和 Silvio Savarese。大规模室内空间的3D语义解析。载于 CVPR，第1534–1543页，2016年。2[2] J. Behley、M. Garbade、A. Milioto、J. Quenzel、S. Behnke、C. Stachniss 和 J. Gall。SemanticKITTI：用于激光雷达序列语义场景理解的数据集。载于 IEEE/CVF 国际计算机视觉会议（ICCV）论文集，2019年。1、3、4、6[3] Elena Camuffo、Daniele Mari 和 Simone Milani。点云学习算法的最新进展：更新综述。Sensors，22(4):1357，2022年。2[4] Elena Camuffo、Umberto Michieli 和 Simone Milani。从错误中学习：点云分割中的自正则化层次语义表示。arXiv预印本 arXiv:2301.11145，2023年。2[5] Jun Cen、Peng Yun、Shiwei Zhang、Junhao Cai、Di Luan、Michael Yu Wang、Ming Liu 和 Mingqian Tang。面向激光雷达点云的开放世界语义分割。arXiv预印本 arXiv:2207.01452，2022年。1、2[6] Fabio Cermelli、Dario Fontanel、Antonio Tavera、Marco Ciccone 和 Barbara Caputo。基于图像标签的语义分割增量学习。载于 CVPR，2022 年 6 月。1 、 2[7] Fabio Cermelli 、Massimiliano Mancini、Samuel Rota Bulo、Elisa Ricci 和 Barbara Caputo。为语义分割中的增量学习建模背景。载于 IEEE计算机视觉与模式识别会议（CVPR）论文集，第9233–9242页，2020年。2、5、6、7[8] Townim Chowdhury、Ali Cheraghian、Sameera Ramasinghe、Sahar Ahmadi、Morteza Saberi 和 Shafin Rahman。面向3D点云物体的少样本类别增量学习。ECCV，第204–220页，2022年10月。1、2[9] 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。载于 IEEE计算机视觉与模式识别会议（CVPR）论文集，2016年6月。3、4[10] Tiago Cortinhal、George Tzelepis 和 Eren Erdal Aksoy。SalsaNext：面向自动驾驶的快速、不确定性感知激光雷达点云语义分割，2020年。2[11] Matthias De Lange、Rahaf Aljundi、Marc Masana、Sarah Parisot、Xu Jia、Aleš Leonardis、Gregory Slabaugh 和 Tinne Tuytelaars。持续学习综述：在分类任务中克服遗忘。IEEE模式分析与机器智能汇刊，44(7):3366–3385，2022年。1、2[12] Jiahua Dong、Yang Cong、Gan Sun、Bingtao Ma 和 Lichen Wang。I3dol：在不发生灾难性遗忘的情况下进行3D物体增量学习。载于 AAAI 人工智能会议，2020年。1、2[13] Arthur Douillard、Yifu Chen、Arnaud Dapogny 和 Matthieu Cord。PLOP：持续语义分割中的无遗忘学习。载于 CVPR，2021年。1、2

[14] Hehe Fan、Xin Yu、Yuhang Ding、Yi Yang 和 Mohan Kankanhalli。PSTNet：点云序列上的点云时空卷积。ICLR，2021年。2[15] Biao Gao、Yancheng Pan、Chengkun Li、Sibo Geng 和 Huijing Zhao。我们是否迫切需要用于语义分割的3D激光雷达数据？数据集与方法综述。IEEE智能交通系统汇刊，2021年。2[16] A. Geiger、P. Lenz 和 R. Urtasun。我们是否已准备好迎接自动驾驶？KITTI视觉基准套件。载于 IEEE计算机视觉与模式识别会议（CVPR）论文集，页码

3354 3361, 2012 1, 4, 6 – 。

[17] I. J. Good。理性决策。《英国皇家统计学会期刊》，B辑（方法论），14(1):107–114，1952年。5[18] Meng-Hao Guo、Jun-Xiong Cai、Zheng-Ning Liu、Tai-Jiang Mu、Ralph R. Martin 和 Shi-Min Hu。PCT：点云变换器。《计算视觉媒体》，7(2):187–199，2021年4月。2[19] Geoffrey Hinton、Oriol Vinyals 和 Jeff Dean。蒸馏神经网络中的知识。WACV，2015年。5[20] Yuenan Hou、Xinge Zhu、Yuexin Ma、Chen Change Loy 和 Yikang Li。用于激光雷达语义分割的点到体素知识蒸馏。载于 IEEE计算机视觉与模式识别会议论文集，页码 8479–8488，2022年。2, 5[21] Qingyong Hu、Bo Yang、Linhai Xie、Stefano Rosa、Yulan Guo、Zhihua Wang、Niki Trigoni 和 Andrew Markham。RandLA-Net：大规模点云的高效语义分割。IEEE计算机视觉与模式识别会议论文集，2020年。1, 2, 4, 6[22] Matthew Johnson-Roberson、Charlie Barto、Rounak Mehta、Sharath Nittur Sridhar、Karl Rosaen 和 Ram Vasudevan。矩阵中的驾驶：虚拟世界能否取代现实世界任务中的人工标注？ICRA，页码 746–753，2016年。8[23] KJ Joseph、Salman Khan、Fahad Shahbaz Khan 和 Vineeth N Balasubramanian。迈向开放世界目标检测。载于 IEEE计算机视觉与模式识别会议（CVPR）论文集，页码 5830 5840，2021年 1, 2 – 。[24] James Kirkpatrick、Razvan Pascanu、Neil Rabinowitz、Joel Veness、Guillaume Desjardins、Andrei A. Rusu、Kieran Milan、John Quan、Tiago Ramalho、Agnieszka Grabska-Barwinska、Demis Hassabis、Claudia Clopath、Dharshan Kumaran 和 Raia Hadsell。克服神经网络中的灾难性遗忘。《美国国家科学院院刊》，114(13):3521–3526，2017年。1, 2[25] Marvin Klingner、Andreas B¨ar、Philipp Donn 和 Tim Fin-gscheidt。用于语义分割的类别增量学习：既不重用旧数据也不重用旧标签。载于 2020年IEEE第23届智能交通系统国际会议（ITSC），页码 1–8。IEEE，2020年。1, 2, 3, 4[26] Joshua Knights、Peyman Moghadam、Milad Ramezani、Sridha Sridharan 和 Clinton Fookes。InCloud：用于点云地点识别的增量学习。载于 2022年IEEE/RSJ智能机器人与系统国际会议（IROS），页码 8559–8566。IEEE，2022年。1, 2[27] Alex Lang、Sourabh Vora、Holger Caesar、Lubing Zhou、Jiong Yang 和 Oscar Beijbom。PointPillars：用于从点云中检测目标的快速编码器。IEEE

计算机视觉与模式识别会议（CVPR）论文集，页码 12689–12697，2019年6月。8[28] Junnan Li、Pan Zhou、Caiming Xiong 和 Steven CH Hoi。无监督表征的原型对比学习。载于 ICLR，2021年。2[29] Zhizhong Li 和 Derek Hoiem。学会遗忘。IEEE模式分析与机器智能汇刊，40:2935–2947，2016年。1, 2, 4, 6, 7[30] Yaping Lin、George Vosselman、Yanpeng Cao 和 Michael Ying Yang。用于语义ALS点云分割的主动增量学习。ISPRS摄影测量与遥感学报，169:73–92，2020年。1, 2[31] Jonathan Long、Evan Shelhamer 和 Trevor Darrell。用于语义分割的全卷积网络。IEEE计算机视觉与模式识别会议（CVPR）论文集，页码 3431–3440，2015年。6, 8[32] Andrea Maracani、Umberto Michieli、Marco Toldo 和 Pietro Zanuttigh。Recall：语义分割中的基于重放的持续学习。载于 IEEE/CVF国际计算机视觉会议论文集，页码 7026 7035，2021年 2, 5, 6 – 。[33] Daniel Maturana 和 Sebastian Scherer。Voxnet：用于实时目标识别的3D卷积神经网络。载于 2015年IEEE/RSJ智能机器人与系统国际会议（IROS），页码 922–928，2015年。2[34] Mazen Mel、Umberto Michieli 和 Pietro Zanuttigh。由粗到细语义分割的增量学习与多任务学习策略。《Technologies》，8(1):1，2020年。2[35] Umberto Michieli、Marco Toldo 和 Pietro Zanuttigh。第8章——语义分割中的域适应与持续学习。载于 E.R. Davies 和 Matthew A. Turk 编，《计算机视觉中的高级方法与深度学习》，计算机视觉与模式识别，页码 275–303。Academic Press，2022年。3[36] Umberto Michieli 和 Pietro Zanuttigh。语义分割的增量学习技术。载于国际计算机视觉会议（ICCV）计算机视觉中源知识迁移与适应研讨会（TASK-CV），2019年。1, 2, 3, 6, 7[37] Umberto Michieli 和 Pietro Zanuttigh。通过排斥-吸引稀疏且解耦的潜在表征实现持续语义分割。载于 IEEE计算机视觉与模式识别会议（CVPR）论文集，页码 1114–1124，2021年。2[38] Andres Milioto、Ignacio Vizzo、Jens Behley 和 Cyrill Stach-niss。RangeNet++：快速且准确的激光雷达语义分割。载于 IROS，页码 4213–4220。IEEE，2019年。2[39] Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Rai-son、Alykhan Tejani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala。PyTorch：一种命令式、高性能深度学习库。载于《神经信息处理系统进展》第32卷，页码 8024–8035。Curran Associates, Inc.，2019年。6

[40] Can Peng、Kun Zhao 和 Brian C. Lovell。Faster ilod：一种基于 Faster R-CNN 的目标检测器增量学习方法。《Pattern Recognition Letters》，140：109–115，2020年。1, 2[41] Charles R Qi、Hao Su、Kaichun Mo 和 Leonidas J Guibas。Pointnet：用于三维分类和分割的点集深度学习。IEEE计算机视觉与模式识别会议（CVPR）论文集，2016年。1, 2[42] Charles R Qi、Li Yi、Hao Su 和 Leonidas J Guibas。Point-net++：度量空间中点集的深度层次特征学习。NeurIPS，2017年。1, 2[43] Sylvestre-Alvise Rebuffi、Alexander Kolesnikov、Georg Sperl 和 Christoph Lampert。iCaRL：增量分类器与表示学习。第5533–5542页，07 2017年。1, 2[44] Donald Shenaj、Francesco Barbato、Umberto Michieli 和 Pietro Zanuttigh。语义分割中的持续由粗到细域适应。《Image and Vision Computing》，121：104426，2022年。2[45] Konstantin Shmelkov、Cordelia Schmid 和 Karteek Alahari。无需灾难性遗忘的目标检测器增量学习。载于2017年IEEE国际计算机视觉会议（ICCV），第3420–3429页，2017年。1, 2[46] Otilia Stretcu、Emmanouil Antonios Platanios、Tom Mitchell 和 Barnab´as P´oczos。用于分类的由粗到细课程学习。载于ICLRW，2020年。2[47] Lyne P. Tchapmi、Christopher Bongsoo Choy、Iro Armeni、JunYoung Gwak 和 Silvio Savarese。Segcloud：三维点云语义分割。2017年国际三维视觉会议（3DV），第537–547页，2017年。2[48] Marco Toldo、Umberto Michieli 和 Pietro Zanuttigh。风格学习：跨任务和域的持续语义分割。2022年。1, 2, 5, 6, 7[49] Xiang Xiang、Yuwen Tan、Qian Wan、Jing Ma、Alan Yuille 和 Gregory D Hager。由粗到细的增量少样本学习。载于《Computer Vision–ECCV 2022：第17届欧洲会议》，特拉维夫，以色列，2022年10月23日至27日，论文集，第三十一部分，第205–222页。Springer，2022年。2[50] Boyu Yang、Fang Wan、Chang Liu、Bohao Li、Xiangyang Ji 和 Qixiang Ye。用于少样本语义分割的基于部件的语义变换。《IEEE神经网络与学习系统汇刊》，2021年。2[51] Yang Zhang、Zixiang Zhou、Philip David、Xiangyu Yue、Zerong Xi、Boqing Gong 和 Hassan Foroosh。PolarNet：一种改进的在线激光雷达点云语义分割网格表示。载于IEEE计算机视觉与模式识别会议（CVPR）论文集，第9601–9610页，2020年。2[52] Xinge Zhu、Hui Zhou、Tai Wang、Fangzhou Hong、Wei Li、Yuexin Ma、Hongsheng Li、Ruigang Yang 和 Dahua Lin。用于基于激光雷达感知的柱状与非对称三维卷积网络。《IEEE Transactions on Pattern Analysis and Machine Intelligence》，2021年。2, 4, 6[53] Xinge Zhu、Hui Zhou、Tai Wang、Fangzhou Hong、Yuexin Ma、Wei Li、Hongsheng Li 和 Dahua Lin。用于激光雷达分割的柱状与非对称三维卷积网络。载于IEEE计算机视觉与模式识别会议，第9939–9948页，2021年。2, 4, 6