arXiv:2304.00749v1 [cs.CV] 3 Apr 2023

小而强大：使用 U-Next 框架增强三维点云语义分割

Ziyin Zeng[1], Qingyong Hu[2],ZhongXie[1],JianZhou[3] 和 Yongyang Xu[1*]中国地质大学,
[2]牛津大学, [3]武汉大学{zengziyin, xiehzhong, yongyangxu}@cug.edu.cn,
qingyong.hu@cs.ox.ac.uk, jianzhou@w
hu.edu.cn

摘要

我们研究大规模三维点云的语义分割问题。近年来，研究重点主要集中在局部特征聚合、改进损失函数和采样策略上。然而，点云语义分割的基础框架在很大程度上被忽视，大多数现有方法默认采用 *U-Net* 架构。本文提出 *U-Next*，这是一种小巧而强大的点云语义分割框架。该框架的关键在于从语义相似的特征图中学习多尺度层次化表示。具体而言，我们以嵌套且密集排列的方式堆叠多个 *U-Net* L^1 *codecs*，以最大限度地缩小语义差距，同时跨尺度融合特征图，从而有效恢复细粒度细节。我们还设计了一种多层深度监督机制，以进一步平滑梯度传播并促进网络优化。在 *S3DIS*、*Toronto3D* 和 *SensatUrban* 这三个大规模基准测试上进行的大量实验表明，所提出的 *U-Next* 架构具有优越性和有效性。我们的 *U-Next* 架构在不同任务和基线方法上均展现出稳定且显著的性能提升，表明其有望成为未来研究的通用框架。

1. 引言

得益于三维采集设备的快速发展，近年来，三维点云分析日益受到关注 [21]。作为三维场景理解的基础步骤，点云语义分割旨在根据每个点的几何结构，为其分配唯一的语义标签。由于其在自动驾驶、机器人学和虚拟现实等领域有着广泛应用，这项任务已引起研究人员的广泛兴趣。

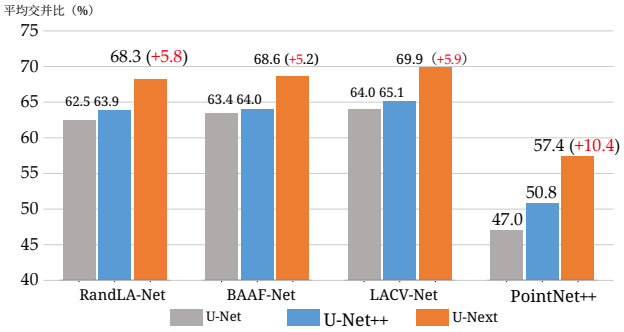


图1. 不同框架下各类算法在 S3DIS数据集（区域5）上的性能。我们的 U-Next 框架明显优于 U-Net 和 U-Net++。

领域 [4, 5, 52]。然而，要实现细粒度的语义场景解析仍极具挑战，因为采集到的点云天然无序、非规则且无结构，同时还存在点数量庞大、物体尺寸各异以及分布非均匀等问题。

近期，开创性研究 PointNet [45] 被提出，可直接从无结构的三维点云中学习，同时具备较高效率，并在多个下游任务中展现出令人鼓舞的性能。此后，针对点云分析的深度学习 [21] 得到了广泛探索。大量复杂的神经网络架构 [1, 17, 24, 28, 44, 46, 49, 63, 68, 74] 相继提出，使大规模三维点云基准 [2, 4, 10, 22, 26]上的性能持续提升。尽管取得了令人瞩目的进展，现有方法主要聚焦于表示学习，即为三维点云的下游任务学习有效表示，包括基于点 [27, 45–47, 81], 、基于投影 [6, 32], 、基于体素 [13, 17, 62], 以及基于点与体素 [40, 59]的技术。此外，一些方法深入探索了各种局部特征聚合机制 [27, 38, 74] 和高效的神经网络架构 [13, 17]。然而，在三维点云学习领域，鲜有研究致力于改进分割框架。大多数现有方法倾向于不加区分地采用广泛使用的 U-Net 架构，

*通讯作者

却忽视了其固有局限性，而这些局限性已在二维自然图像和医学图像领域得到充分记录。

作为应用最广泛的神经网络架构之一，U-Net [50] 基于编码器-解码器结构构建，其关键在于利用跳跃连接，将编码器和解码器中的中间潜在特征图结合起来。尽管这种架构看似对称且自然，但相同尺度的特征图在语义上实际上相差甚远（编码器中的浅层和低级特征 + 与解码器中的深层和语义特征），而且没有可靠的理论能够证明这种组合是最优的 [84]。另一方面，网络的最优深度也无法预先确定，因此需要手动调整架构。为此，一些方法 [8, 19, 19, 66, 72, 72] 开始明确改进 U-Net 框架，包括采用嵌套式、密集重设计跳跃路径的 U-Net++ [84] 、采用全尺度跳跃连接的 UNet 3+ [29] ，以及将 Transformer 用作编码器的 TransUNet [8] 等。尽管这些方法已被证明在医学图像分割中非常有效，但它们改进后的架构是否同样适用于三维点云分割，仍是一个悬而未决的问题，尤其是考虑到三维点云与图像之间的内在差异。

为回答这一问题，我们首先深入分析了 U-Net 及其变体，随后开展探索性实验，以验证现有架构能否顺利调整以适用于三维点云。然而，U-Net++架构在二维图像分割中表现出色，却未能显著提升三维点云的分割效果（如图1所示）。这可归因于三维点云无序且非规则的特性：在激进的上采样和下采样过程中会发生显著的信息损失，尤其是考虑到现有方法大多采用简单的下采样（随机采样或最远点采样）和上采样方案（最近邻插值和三线性插值）。此外，在整体网络架构框架中，特征维度连续缩减（由高到低）和扩展（由低到高）会引入大量噪声，这些噪声会逐层累积，并对分割准确率产生深远影响。在这种情况下，简单地聚合不同尺度的特征可能无助于提高分割准确率，反而会由于不同层级的特征差异较大而增加优化难度。

鉴于此，本文提出了一种通用、概念简洁且效果显著的点云语义分割架构，称为 U-Next。U-Next 的关键在于采用最基础的 U-Net L^1 子网络，如图 2-(B) 所示，该网络仅执行一次下采样和上采样操作，因此语义差距最小，且没有额外噪声，并将其用作构建 U-Next 架构的构建模块。

。也就是说，尽可能多地堆叠 U-Net L^1 编解码器，以学习不同尺度的点特征，并使局部特征图能够自由地横向、向上或向下流动，如图 2 所示。在框架中引入额外尺度有助于增强输入数据的表示能力，从而提高总体准确率。最后，我们在每个中间节点应用多层深度监督，以降低各组成 U-Net L^1 子网络的学习和优化难度。在包括 S3DIS [2], Toronto3D[58], 和 SensatUrban [25], 在内的三个大型公开基准上进行的实验表明，基于广泛使用的 RandLA-Net 网络，我们的框架可使 mIoU 分数提升 2.2% $\sim$ 10.1%。此外，我们的框架可以无缝集成到现有分割网络中，例如 PointNet++ [46]、BAAF-Net[49], 和 LACV-Net [78], ，从而提高分割准确率。值得注意的是，这些提升并未带来任何明显的额外计算开销，凸显了 U-Next 框架的通用性和有效性。总体而言，我们的主要贡献可概括如下：

- 我们深入分析了三维点云领域中 U-Net 架构的演变，发现 U-Net L^1 子网络适合作为精细粒度点云分割的组件。
- 我们提出了一种通用且有效的分割架构 U-Next，它通过堆叠多个基础U-NetL^1 编解码器并采用多级深度监督构建而成。
- 在三个大规模基准测试和至多四种基线方法上进行的大量实验，证明了 U-Next 架构的有效性和通用性。

2. 相关工作

2.1. 三维点云分割学习

为分析非结构化三维点云，早期研究通常采用手工设计的特征描述符 [33, 51],，而近期研究则主要基于数据驱动的深度神经网络架构。本文简要回顾相关的学习方法，包括基于投影、基于体素和基于点的方法。

鉴于三维点云具有无序和非结构化特性，直接将标准卷积神经网络应用于三维点云仍然颇具挑战。因此，研究者提出了基于投影的方法 [11, 35, 77] 和基于体素的方法 [18, 36, 42], 先通过投影或体素化步骤，将无结构点云转换为中间的规则二维网格图像或三维稠密体素，再利用表现出色的卷积神经网络从结构化表示中学习。尽管取得了令人期待的结果，这些技术的主要缺点在于信息

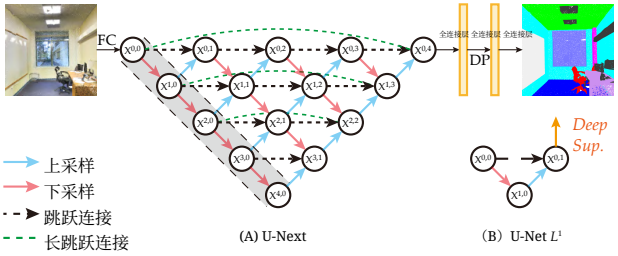


图 2. 所提出的 U-Next 和 U-Net L^1示意图。(A) 所提出的 U-Next 的详细架构。(B) 采用深度监督的 U-Net L^1 子网络。FC：全连接层；DP：Dropout；深度监督：深度监督。

在三维到二维投影或体素化步骤中造成的损失，尤其是在为大规模三维点云学习细粒度特征时。相比之下，基于点的方法旨在直接处理非规则点云，而不使用中间表示。开创性的 Point- [45]Net 被提出，通过利用共享多层感知机学习逐点特征表示。此后，越来越多的基于点的方法被提出，用于从无结构点云中学习 [16, 27, 46, 49, 55, 63,70, 74, 79, 81]。若干具有代表性的研究工作引入了循环神经网络 (RNNs) [30, 76],图神经网络 (GNNs) [34, 38, 68], 以及 Transformer 架构 [20, 23, 74], , 以提升在不同任务上的表现。

2.2. 多尺度特征融合

多尺度融合已得到广泛研究，并成功应用于二维图像处理 [7, 9, 53, 54, 56]。现有工作主要基于编码器-解码器框架，在编码器或解码器子网络中使用不同尺度或不同分辨率的特征图。通常使用求和或拼接操作来融合不同的特征图 [3, 43, 67, 82]。为有效分析由不同尺寸物体构成的大规模三维点云，已有少数研究也开始探索三维点云中的多尺度特征融合。开创性的 PointNet++ [46] 采用多尺度分组和抽象，从不同尺度聚合特征。BAAF-Net [49]提出的自适应融合模块用于自适应融合来自多分辨率特征图的细粒度表示。

2.3. 深度监督学习

深度监督最初用于解决深度神经网络训练中的梯度消失和收敛速度缓慢问题 [37, 57]。作为一种有效的训练技术，它也被应用于不同任务以提升性能 [12, 56, 64, 85]。Lee 等人 [37]证明，深度监督层能够提升隐藏层的学习能力，进一步促使中间层学习判别性特征，并使网络能够快速收敛和实现正则化。Dou 等人 [14]提出了一种深度监督范式，通过结合来自不同分辨率特征图的预测结果，解决网络训练中的优化问题。深度监督也可用于类似 U-Net 的架构，其中间特征图被视为子网络的输出。U-Net ++ [84]在 U-Net 架构中采用重新设计的跳跃连接，构建深度监督方案，以进一步提升特征图的质量。U-Net3+ [29] 通过分层学习聚合特征图来实现深度监督。

2.4. 改进 U-Net 架构

最初的 U-Net [50], 采用对称的编码器-解码器架构，在图像分割中得到广泛应用。具体而言，跳跃连接用于结合解码器的高级语义特征与编码器的低级细节。在这一架构的基础上，研究者进一步提出了许多改进的网络架构。Res-UNet [72] 和 Dense-UNet [19] 分别将 U-Net 的每个子模块替换为残差连接或密集连接。U-Net++ [84] 通过融合深度各异的多个 U-Net 子网络，重新设计跳跃连接路径以探索最佳深度；每个子网络均通过一系列嵌套密集跳跃路径相连。U-Net3+ [29] 提出了全尺度跳跃连接，以充分利用多尺度特征表示。TransUNet [8] 将 Transformer 引入 U-Net, 兼取两种模型的优势。这种方法不仅能够编码强大的全局上下文，还能利用低级 CNN 特征。

同样，本文也旨在从根本上改进用于三维点云分割的 U-Net 架构。然而，与现有技术不同，我们首先系统地分析现有架构的固有缺陷，然后提出 U-Next 架构：通过堆叠尽可能多的子网络，并尽量缩小编解码器语义差距，从而有效地学习非均匀、无结构的三维点云。

3. 所提出的 U-Next

3.1. 重新审视类 U-Net 架构

类 U-Net 架构已广泛应用于现代分割模型中。如图3(A) 所示，该网络采用对称的编码器-解码器架构，跳跃连接在其中发挥着关键作用，将解码器子网络中深层且语义丰富的特征图与编码器子网络中浅层且细粒度的特征图整合起来。尽管这种架构在恢复语义分割和实例分割等密集预测任务的细粒度细节方面取得了令人鼓舞的成果，但仍存在局限。具体而言，它只在相同尺度上进行特征聚合，这种做法限制较多，且缺乏坚实的理论基础。此外，研究表明，聚合后的特征图在语义上存在差异 。

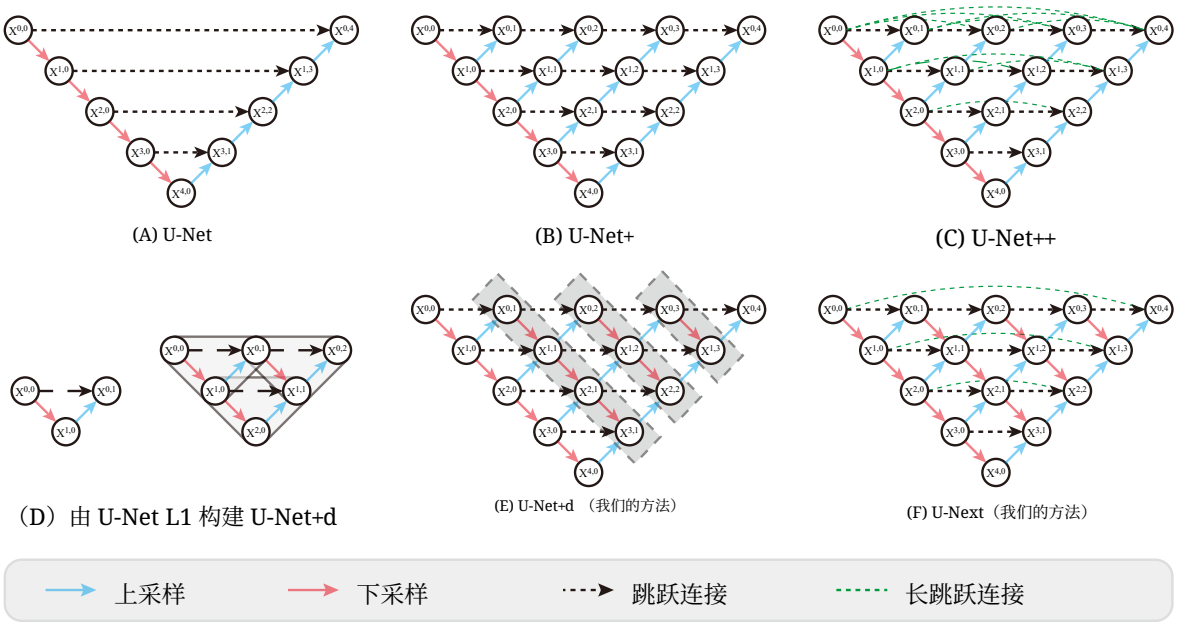


图3. U-Net 到 U-Next 的演进示意图。图中标出了各架构的主要贡献。（A）为原始 U-Net，（B）和（C）为+ 提出的 U-Net ++ 和 U-Net [84]，；（D）展示了如何从+^{d} U-Net L^1构建 U-Net+^{d}；（E）和（F）分别为提出的 U-Net+ 和 U-Next。U-Net ++ 比原始 U-Net 具有更多上采样分支，以实现更精细的粒度。U-Net+ 基于 U-Net 采用密集长跳跃路径，聚合分辨率相同的特征。U-Net+^{d} 具有更多下采样分支，并通过子网络反复融合多尺度特征图，以恢复细粒度细节。U-Next 基于 U-Net+^{d}，在每个水平层级中采用长跳跃连接，专注于基本特征。

割任务中取得了令人鼓舞的成果，但这种架构仍存在局限。具体而言，它只在相同尺度上进行特征聚合，这种做法限制较多，且缺乏坚实的理论基础。此外，研究表明，聚合后的特征图在语义上存在差异 [84]。**挑战**。近年来，已有少数先进架构，例如 U-Net+（图3（B））和 U-Net++ [84]（图3（C）），相继被提出。这些架构将不同层级的 U-Net 组装为具有深度监督的子网络，以有效利用多尺度特征图，并致力于缩小医学图像分割中融合特征图之间的语义差距。然而，将这些架构应用于三维点云仍然极具挑战性。**首先**，大多数现有架构从根本上看都可视为由深度各异的多个 U-Net 组成的集成。因此，更深层 U-Net 中融合特征之间的语义差距仍然显著，可能会妨碍网络学习和优化。**其次**，考虑到点云神经网络架构中经常采用激进的上采样和下采样，再加上三维点云无序且非结构化的特性，通过水平向以及由低到高特征融合所恢复的全分辨率特征图，对于细粒度语义分割而言可能仍不充分。

解决方案。鉴于下采样和上采样过程中不可避免的信息损失，以及随着 U-Net 深度增加，编解码器之间不断扩大的语义差距，我们决定重新审视 U-Net 架构，并提出一种专为点云语义分割设计的框架。受构建模块这一概念的启发，我们思考能否以基础 U-Net L^1 子网络（图3（D））为基础构建 U-Net 架构，因为 U-Net L^1 子网络只有一次下采样和上采样操作，因此语义差距天然较小。接下来，考虑到 U-Net L^1 子网络结构较浅且学习能力有限，我们采用一种直接的方法来增强其特征学习能力：以嵌套且密集排列的方式堆叠多个 U-Net L^1 编解码器。这样网络便能学习多尺度特征，最终提升性能。此外，为确保梯度反向传播顺畅，并降低训练过程中的优化难度，最好进一步引入专门的深度监督方案。总而言之，一种可行的解决方案是将基础 U-Net L1 编解码器作为框架的构建模块，从而同时最大限度地缩小语义差距并恢复细粒度特征图，进而实现高质量的三维语义分割。

3.2. U-Next：使用 U-Net L^1 子网络构建骨干架构

在构建最终的 U-Next 框架之前，我们首先提出了一种基于基础U-Net的层次集成架构 L^1 编解码器，如图3(E)所示。可以看出，该架构显式添加了不同层级解

码器节点之间的连接。与之前的 U-Net+架构相比，这使每个节点都能从其他并行且相邻（较低层和较高层）的节点接收信息，从而实现充分、全面的信息融合。这种架构也可以看作尽可能逐步堆叠多个 U-Net L^1 编解码器，因此编解码器语义差距自然最小。此外，我们为 U-Net+^{d} 架构增添了长跳跃连接，以构建最终的 U-Next 架构，如图3(F)所示。与 U-Net++ 框架中使用的嵌套式密集跳跃路径不同，我们的 U-Next 架构仅添加了类似于原始 U-Net 架构的简单稀疏跳跃连接。网络的详细说明总结如下。

构建 U-Net+^{d} 基于 U-Net L^1： 如图 3(E) 所示，U-Net+^{d} 中的每个节点都会接收来自其他平行且相邻（较低和较高）节点的信息，从而实现全面且充分的多尺度信息融合。具体而言，我们使用 $x^{i,\,j}$ 表示节点 $X^{i,\,j}$ 的输出（同一水平行中具有相同 i 的各节点，以及沿下采样方向位于同一斜线且具有相同 j 的各节点）。U-Net+^{d} 中的特征图堆叠 $x^{i,\,j}$ 可表示如下：

$$x^{i,j}=\begin{cases}\mathcal{C}[x^{i-1,j}], & j=0\\ \mathcal{C}[x^{i-1,j},\mathcal{U}(x^{i-1,j+1})], & j>0, i=0\\ \mathcal{C}[x^{i-1,j},\mathcal{U}(x^{i-1,\,j+1}),\mathcal{D}(x^{i-1,j-1})], & j>0, i>0\end{cases} \quad (1)$$

其中，$\mathcal{C}(\cdot)$ 是编码块（通常为特征提取模块），$\mathcal{D}(\cdot)$ 和 $\mathcal{U}(\cdot)$ 分别表示下采样操作和上采样操作，$[\cdot]$ 表示拼接操作。

基于 U-Net 构建 U-Next+^{d}： 我们通过引入长跳跃连接，基于 U-Net+^{d} 构建了 U-Next 架构。如图 3(F) 所示，较早的特征图通过 U-Next 的长跳跃连接传递，并与较后的特征图融合。具体而言，通过将第 i 行的第一个节点与节点 $X^{i,\,j}$ 拼接，计算加入额外长跳跃连接后的特征图；当且仅当节点 $X^{i,\,j}$ 是第 i 行的最后一个节点时，计算方式如下：

$$x^{i,j}=\begin{cases}\mathcal{C}[x^{i-1,j},\mathcal{U}(x^{i-1,j+1}),x^{i,0}], & i=0\\ \mathcal{C}[x^{i-1,j},\mathcal{U}(x^{i-1,\,j+1}),\mathcal{D}(x^{i-1,j-1}),x^{i,0}], & i>0\end{cases} \quad (2)$$

3.3. 多层深度监督学习

到目前为止，我们的框架中已按层次堆叠了多个 U-Net L^1 子网络。为了进一步从聚合特征图中学习层次化表征，我们将多层深度监督引入我们的

从优化和学习的角度来看，该框架

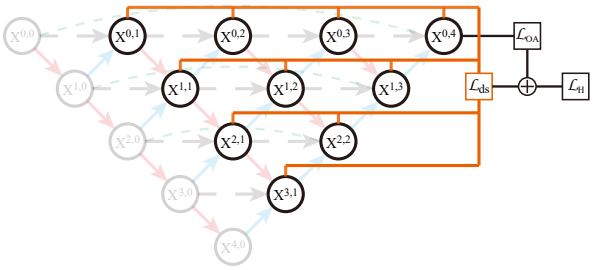


图 4. 深度监督示意图。所有子网络均受到监督。 $\mathcal{L}_{ds}$：多级深度监督损失； $\mathcal{L}_{OA}$：整体网络损失； $\mathcal{L}_H$：混合损失。

过程。如图所示4，与 U-Net++仅对生成的全分辨率特征图应用深度监督不同，我们的 U-Next 框架明确地对不同层级的每个解码器节点都应用了深度监督（即，对每个 U-Net L^1 子网络）。通过在框架的不同解码阶段引入真实标注监督信号，更有可能实现平滑的梯度传播和简便的网络优化。因此，所有组成 U-Net L^1 子网络都经过同步训练，以提升多尺度特征图的整体学习能力。

多层深度监督损失： 为实现多层深度监督，将每个 U-NetL^1 子网络中解码器的输出输入一个普通的 1×1 卷积层，使其与真实标签对齐。随后，我们采用广泛使用的交叉熵损失来计算每个 U-Net L^1 子网络的损失。最后，深度监督的总损失定义为所有 U-Net L^1 子网络损失的平均值，如下所示：

$$\mathcal{L}_{ds}(Y,P)=-\frac{1}{N}\sum_{i=0}^{L}\sum_{j=0}^{L-i}\sum_{c=0}^{C}y_c^i\log p_c^{i,j} \quad (3)$$

其中，y_c^i 是第 i行的真实标签向量，$p^{i,\,}{}_c{}^j$是节点 $X^{i,\,j}$的预测标签向量，C 是标签类别数，L 是架构层级数，N是解码器节点数。在多层深度监督中，所有子网络均参与训练，从而增强每个子网络对局部特征的学习能力。

混合损失： 如图4所示，在最终节点的输出（$X^{0,\,4}$）之后，使用全连接层预测所有输入点的语义标签。因此，我们也使用交叉熵损失，根据最终预测结果计算整体网络的损失。此外，该框架的最终损失函数是多级深度监督损失与整体网络损失相结合的混合损失，如下所示：

$$\mathcal{L}_H=\mathcal{L}_{ds}+\mathcal{L}_{oa} \quad (4)$$

表 1. S3DIS数据集上不同方法的定量结果 [2]（六折交叉验证）。

方法	OA	mIoU	天花板	地板	wall	beam	col.	窗户	door	桌子	椅子	sofa	书。	板	杂物。
PointNet [45]	78.6	47.6	88.0	88.7	69.3	42.4	23.1	47.5	51.6	54.1	42.0	9.6	38.2	29.4	35.2
DGCNN [69]	84.5	55.5	93.2	95.9	72.8	54.6	32.2	56.2	50.7	62.8	63.4	22.7	38.2	32.5	46.8
SPGraph [34]	86.4	62.1	89.9	95.1	76.4	62.8	47.1	55.3	68.4	73.5	69.2	63.2	45.9	8.7	52.9
DeepGCNs [38]	85.9	60.0	93.1	95.3	78.2	33.9	37.4	56.1	68.2	64.9	61.0	34.6	51.5	51.1	54.4
PointWeb [81]	87.3	66.7	93.5	94.2	80.8	52.4	41.3	64.9	68.1	71.4	67.1	50.3	62.7	62.2	58.5
ShellNet [80]	87.1	66.8	90.2	93.6	79.9	60.4	44.1	64.9	52.9	71.6	**84.7**	53.8	64.6	48.6	59.4
KPConv [63]	-	70.6	93.6	92.4	83.1	63.9	54.3	66.1	76.6	57.8	64.0	69.3	74.9	61.3	60.3
BAAF-Net [49]	88.9	72.2	93.3	96.8	81.6	61.9	49.5	65.4	73.3	72.0	83.7	67.5	64.3	67.0	62.4
BAF-LAC [55]	88.2	71.7	92.5	95.9	81.3	63.2	57.8	63.0	**79.9**	70.3	74.6	60.6	67.2	65.3	60.4
SCF-Net [16]	88.4	71.6	93.3	96.4	80.9	64.9	47.4	64.5	70.1	71.4	**81.6**	67.2	64.4	67.5	60.9
LACV-Net [78]	89.7	72.7	94.5	96.7	82.1	65.2	48.6	69.3	71.2	72.7	78.1	67.3	67.2	70.9	61.6
CBL [60]	89.6	73.1	94.1	94.2	**85.5**	50.4	**58.8**	**70.3**	78.3	**75.7**	75.0	**71.8**	**74.0**	60.0	62.4
Point Transformer [23]	90.2	73.5	94.3	**97.5**	84.7	55.6	58.1	66.1	78.2	77.6	74.1	67.3	71.2	65.7	**64.8**
PointNeXt [47]	**90.3**	**74.9**	-	-	-	-	-	-	-	-	-	-	-	-	-
RandLA-Net [27]	88.0	70.0	93.1	96.1	80.6	62.4	48.0	64.4	69.4	69.4	76.4	60.0	64.2	65.9	60.1
+ U-Next	89.5 **+1.5**	73.2 **+3.2**	93.6	96.9	84.2	**66.1**	54.6	67.6	75.5	73.6	74.5	62.9	66.2	**74.0**	61.7

4. 实验

4.1. 实验设置

为全面评估所提出的 U-Next 的有效性，我们在广泛使用的室内 S3DIS 数据集 [2]，、室外基于激光雷达的 Toronto3D 数据集 [58], 以及城市尺度的摄影测量 SensatUrban 数据集 [25]上开展实验。为了定量评估分割性能，我们采用各类别交并比（IoU）、平均交并比（mIoU）和总体准确率（OA）作为评估指标。我们还展示了 U-Next 框架在实例分割任务中的性能。不过，受篇幅限制，更多实验收录于附录中。

实现细节。鉴于基于点的 RandLA-Net 简洁且高效 [27], , 我们选择它作为基线模型，以评估 U-Next 架构的有效性。需要注意的是，我们的架构并不局限于 RandLA-Net，第4.3节对此进行了说明。具体而言，RandLA-Net 中使用的局部特征聚合模块被整合为 U-Next 的编码块，如公式 1 和公式 2 所示。在网络训练期间，输入网络的点数为 40960，批量大小设为 4，最大训练轮数设为 100。使用 Adam优化器 [31] ，并采用默认参数优化损失函数。我们遵循 RandLA-Net [27] ，将 K=16设为 K近邻中的参数，初始学习率设为 0.01。所有实验均在单个 NVIDIA GeForce RTX3090 图形处理器和 i7-12700K 中央处理器上进行。训练期间，我们采用简单的随机下采样来提高计算效率。

4.2. 语义分割结果

在 S3DIS 上进行评估。斯坦福大规模三维室内空间（S3DIS）[2] 数据集使用 Matterport 扫描仪采集，涵盖 3 栋建筑中的 6 个大型室内区域，共有 272 个房间。该数据集中的点被标注为 13 个语义类别，每个点表示为一个 6 维向量（坐标（xyz）和颜色（rgb））。我们采用广泛使用的六折交叉验证，全面评估分割性能。

表1 展示了所提出的方法与其他基线方法在 S3DIS 数据集上的定量结果。可以看出，采用所提出的 U-Next 架构的 RandLA-Net 表现出良好的分割性能，总体准确率为 89.5%，mIoU 分数为 73.2%；与基线方法 RandLA-Net 相比，其总体准确率和平均交并比分别提高了 1.5% 和 3.2%。还值得注意的是，除椅子类别略有波动外，所提出的架构持续提升了所有类别的分割性能。尤其是，我们的方法在墙壁等大平面类别以及板材、柱子和梁等小实例类别上的表现优于基线方法。这清楚地表明，堆叠更多子网络以捕获局部几何细节，能够有效提升分割性能。

尽管我们提出的方法与最先进的方法之间仍存在轻微的性能差距，但需要重申的是，本文的首要目标并非建立新的性能基准。相反，我们旨在提出一种简单灵活的语义分割框架，便于其他研究人员和从业者轻松实现并加以扩展。

在 Toronto3D 上进行评估。 Toronto3D 数据集 [58] 由车载移动激光扫描系统采集，覆盖了加拿大多伦多 Avenue Road 上约 1 km 的道路场景。该数据集划分为四个瓦片，每个瓦片约 250 m，每个点均标注为 8 个语义标签之一。训练期间可使用三维坐标和颜色属性。按照 [58], 区域 2 用作测试集，其余三个区域用作训练集。

我们的方法与其他基线方法在 Toronto3D 数据集上取得的定量结果如表 Ta- 所示。

表 2. 不同方法在 Toronto3D 数据集上的定量结果 [58] (区域 2).

RGB	方法	OA	mIoU
No	PointNet++ [46]	92.6	59.5
	DGCNN [69]	94.2	61.7
	MS-PCNN [41]	90.0	65.9
	KPConv [63]	95.4	69.1
	TGNet [39]	94.1	61.3
	MS-TGNet [58]	95.7	70.5
	LACV-Net [78]	95.8	78.5
	RandLA-Net [27]	93.0	77.7
	+ U-Next	**96.0 + 3.0**	**79.2 + 1.5**
Yes	ResDLPS-Net [15]	96.5	80.3
	BAF-LAC [55]	95.2	82.2
	BAAF-Net [49]	94.2	81.2
	LACV-Net [78]	97.4	82.7
	RandLA-Net [27]	94.4	81.8
	+ U-Next	**97.7 + 3.3**	**84.0 + 2.2**

ble 2. 鉴于多个基线方法在该数据集上未将颜色信息作为输入，因此，为了公平比较，我们报告了本方法在使用和不使用颜色信息时的性能。可以看出，所提出的 RandLA-Net（U-Next）无论是否使用颜色属性，均取得了最佳性能。具体而言，与基线方法 RandLA-Net 相比，所提出的方法的平均交并比分数分别提升了 1.5% 和 2.2%。

在 SensatUrban 上的评估。SensatUrban 数据集[25]是一个摄影测量点云数据集，覆盖城市景观中的 7.6 平方公里区域^2，包含近 30 亿个点，并以 13 个语义类别进行了密集标注。我们遵循官方划分来训练网络，并在在线测试服务器上评估性能。

表 3 展示了本方法与其他基线方法在 SensatUrban 数据集上取得的定量结果。可以看出，所提出的方法显著提升了平均交并比分数，从 52.7% 提高到 62.8%，并且几乎在所有类别上都有明显改进。我们还注意到，所提出的方法在具有挑战性的 铁路 类别上取得了 36.8% 的交并比分数，这主要是因为本方法的关键在于堆叠尽可能多的 U-Net L^1 编解码器，并采用多层深度监督，从而减少信息损失并缩小语义差距，进一步提升分割性能，尤其是对小物体的分割性能。这一点也可以从 墙 和人行道类别中得到验证。

4.3. 消融研究

在此，我们从以下四个方面对所提出的 U-Next 架构进行了广泛的消融研究。所有后续模型均在区域 1–4和6上训练，并在区域5S3DIS [2] 数据集上测试。

采用不同基线方法的 U-Next。为验证所提出的 U-Next 对不同基线模型的泛化能力，我们将四种具有代表性的方法：PointNet++ [46]、RandLA-Net [27]、、BAAF-Net [49] 和 LACV-Net [78], 整合到我们的 U-Next 架构中。为确保比较公平，尽管一些基线方法引入了额外的损失函数和架构改进，我们仍仅将其局部特征提取模块用作式（1）和式（2）中的编码块。桌子 4 展示了不同基线方法取得的定量结果。可以看出，将基线方法整合到我们的 U-Next 架构后，语义分割性能均得到提升（mIoU 分数平均提高 6.8%）。我们还注意到，PointNet++带来的提升十分显著（10.4%），这表明我们的 U-Next 架构能够显著提升较弱骨干网络的特征学习能力，促使网络迭代融合由高到低和由低到高子网络生成的多尺度局部特征图。

改变 U-Net 类架构。另一方面，为了进一步证明所提出的 U-Next 相较于现有架构的有效性，我们分别采用 U-Net[50], U-Net+、U-Net++ [84] 以及所提出的 U-Net+^{d}和 U-Next 作为点云语义分割框架。如表 4所示，与其他架构相比，所有基线方法在采用 U-Next 架构时均取得了更好的性能。我们还注意到，U-Net++ 在二维医学图像领域非常成功，但未能显著提升点云网络的语义分割性能。相比之下，通过堆叠更多 U-Net L^1子网络并利用多层深度监督，可以持续提升分割性能。

多层深度监督。 为进一步展示所设计的多层深度监督的有效性，我们进一步验证了所提出的深度监督的效果。根据桌子 4中的结果，与不采用深度监督的 U-Next 相比，采用多层深度监督的 U-Next 平均将 mIoU 分数提高了 1.7%。这进一步表明，堆叠 U-Net L^1 子网络与多层深度监督方案可能相辅相成，共同提升性能。

计算成本和模型复杂度。 为研究所提出的 U-Next 架构的复杂度和计算成本，我们在桌子 5 中报告了采用 RandLA-Net 的不同架构的参数数量和浮点运算次数（FLOPs）。不出所料，所提出的 U-Next 确实具有更多可训练参数和 FLOPS，但增幅并不显著，因为增加中间节点并不会带来大量可训练参数，而大多数可训练参数都产生于编码器-解码器架构的交汇层（这是由于中间层的维度最高）。考虑到

表3. SenastUrban数据集上不同方法的定量结果 [25]。

方法	OA	mIoU	地面	veg.	建筑物	wall	bri.	公园。	rail	交通。	街道。	car	英尺。	bike	水体
PointNet [45]	80.8	23.7	67.9	89.5	80.1	0.0	0.0	3.9	0.0	31.6	0.0	35.1	0.0	0.0	0.0
PointNet++ [46]	84.3	32.9	72.5	94.2	84.8	2.7	2.1	25.8	0.0	31.5	11.4	38.8	7.1	0.0	56.9
TagentConv [61]	76.9	33.3	71.5	91.4	75.9	35.2	0.0	45.3	0.0	26.7	19.2	67.6	0.0	0.0	0.0
SPGraph [34]	85.3	37.3	69.9	94.6	88.9	32.8	12.6	15.8	15.5	30.6	22.9	56.4	0.5	0.0	44.2
SparseConv [17]	88.7	42.7	74.1	97.9	94.2	63.3	7.5	24.2	0.0	30.1	34.0	74.4	0.0	0.0	54.8
KPConv [63]	93.2	57.6	87.1	**98.9**	95.3	**74.4**	28.7	41.4	0.0	55.9	**54.4**	**85.7**	40.4	0.0	**86.3**
BAF-LAC [55]	91.5	54.1	84.4	98.4	94.1	57.2	27.6	42.5	15.0	51.6	39.5	78.1	40.1	0.0	75.2
BAAF-Net [49]	92.0	57.3	84.2	98.3	94.0	55.2	48.9	57.7	20.0	57.3	39.3	79.3	40.7	0.0	70.1
LACV-Net [78]	**93.2**	61.3	**85.5**	98.4	**95.6**	61.9	**58.6**	**64.0**	28.5	62.8	45.4	81.9	42.4	**4.8**	67.7
RandLA-Net [27]	89.8	52.7	80.1	98.1	91.6	48.9	40.6	51.6	0.0	56.7	33.2	80.1	32.6	0.0	71.3
+ U-Next	93.0 + 3.2	**62.8 + 10.1**	85.2	98.6	95.0	68.2	53.6	60.4	**36.8**	**64.0**	48.9	84.9	**45.1**	0.0	76.2

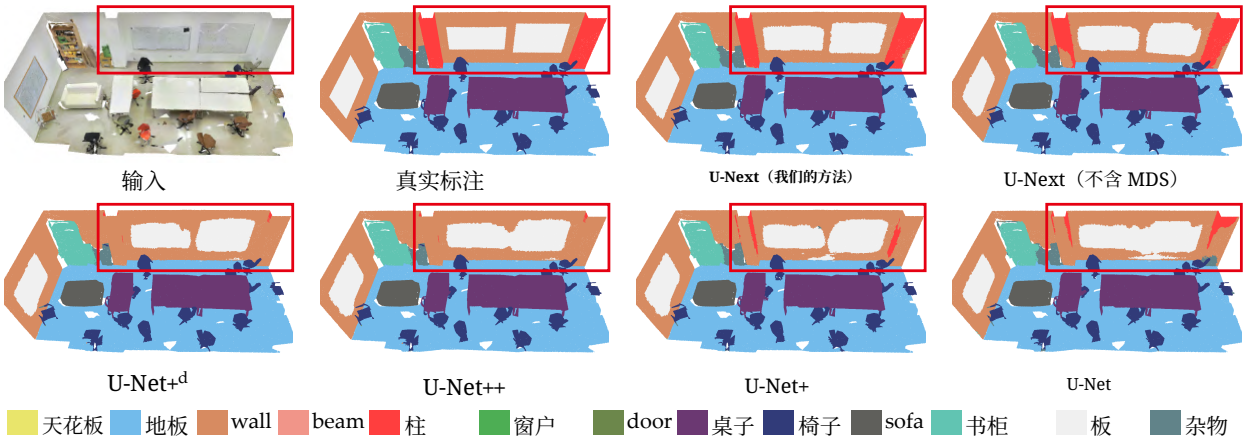


图5. 定性比较 了 RandLA-Net 在 S3DIS（区域 5）上采用不同架构的结果。MDS：多层深度监督 .

分割性能和计算成本，所提出的 U-Next 架构在实践中具有意义且实用。

此处，我们还提供了 RandLA-Net 采用不同架构在 S3DIS 区域 5 上取得的定性比较结果。如图 5所示，所提出的 U-Next 能够平滑、准确地分割边界区域。同时，所提出的 U-Next 能清晰地分割柱子等多个物体，而 U-Net 或 U-Net++则不能，这++主要是因为 U-Net 和 U-Net

5. 结论

本文提出了一种类似 U-Net 的点云语义分割框架，称为 U-Next。U-Next 框架的关键在于堆叠尽可能多的 U-Net L^1 子网络，从而确保语义差距最小化，并获得更好的多尺度特征表示。我们还引入了多层深度监督机制，以促进梯度平滑传播，并在不同抽象层次提供额外指导。大量实验——

表 4. S3DIS（区域5）上采用不同架构的基线模型的语义分割结果（交并比：%）。MDS：多层深度监督。

架构	RandLA-Net	BAAF-Net	LACV-Net	PointNet++
U-Net	62.5	63.4	64.0	47.0
U-Net+	63.4 +0.9	62.9 -0.5	64.4 +0.4	49.1 +2.1
U-Net++	63.9 +1.4	64.0 +0.6	65.1 +1.1	50.8 +3.8
U-Net+^{d}	65.9 +3.4	65.6 +2.2	67。2 +3.2	54.9 +7.9
不使用 MDS 的 U-Next	66.7 +4.2	67.1 +3.7	68。0 +4.0	55.8 +8.8
U-Next（我们的方法）	68.3 +5.8	68.6 +5.2	69.9 +5.9	57.4 +10.4

表 5. 不同架构的 RandLA-Net 模型参数量（百万）和浮点运算次数（百万）。

架构	参数	浮点运算次数
U-Net	4.99	31.35
U-Net+	5.24	31.96
U-Net++	5.39	33.02
U-Net+^{d}	5.51	33.79
U-Next	5.61	34.51

在多个大规模基准和不同骨干网络上取得的实验结果证明了所提出框架的有效性和多功能性。未来，我们计划探索 U-Next 在不同模态数据和任务中的适用性，以进一步验证其在多样化应用中的泛化能力和潜力。

参考文献

[1] Sheng Ao、Yulan Guo、Qingyong Hu、Bo Yang、Andrew Markham 和 Zengping Chen。《You only train once: Learning general and distinctive 3D局部描述符》。IEEE模式分析与机器智能汇刊，2022年。1[2] Iro Armeni、Ozan Sener、Amir R. Zamir、Helen Jiang、Ioannis Brilakis、Martin Fischer 和 Silvio Savarese。《大规模室内空间的三维语义解析》。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2016 1, 2, 6, 7, 13, 17, 19。Vijay [3] Badrinarayanan、Alex Kendall 和 Roberto Cipolla。《Segnet：用于图像分割的深度卷积编码器-解码器架构》。IEEE模式分析与机器智能汇刊，第2481–2495页，2017年。3 [4] Jens Behley、Martin Garbade、Andres Milioto、Jan Quenzel、Sven Behnke、Cyrill Stachniss 和 Jurgen 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），2019年。1[5] Thomas Blanc、Mohamed El Beheiry、Cl´ement Caporal、Jean-Baptiste Masson 和 Bassam Hajj。《Genuage：在虚拟现实中可视化和分析多维单分子点云数据》。《自然·方法》，17，1102，2020年11月，1100–1。[6] A Boulch、B Le Saux 和 N Audebert。《使用深度分割网络进行非结构化点云语义标注》。载于 Eurographics 3D对象检索研讨会（3DOR），2017年。1[7] Zhaowei Cai、Quanfu Fan、Rogerio S Feris 和 Nuno Vasconcelos。《用于快速目标检测的统一多尺度深度卷积神经网络》。载于 欧洲计算机视觉会议（ECCV），2016年。3[8] Jieneng Chen、Yongyi Lu、Qihang Yu、Xiangde Luo、Ehsan Adeli、Yan Wang、Le Lu、Alan L Yuille 和 Yuyin Zhou。《Transunet：Transformer为医学图像分割构建强大的编码器》。arXiv:2102.04306，2021年。2, 3[9] Liang-Chieh Chen、George Papandreou、Iasonas Kokkinos、Kevin Murphy 和 Alan L Yuille。《DeepLab：结合深度卷积网络、空洞卷积和全连接条件随机场的语义图像分割》。IEEE模式分析与机器智能汇刊，第834–848页，2017年。3[10] Meida Chen、Qingyong Hu、Thomas Hugues、Andrew Feng、Yu Hou、Kyle McCullough 和 Lucio Soibelman。《STPLS3D：合成与真实航空摄影测量3D点云数据集》。2022 1。Xiaozhi Chen、[11] Huimin Ma、Ji Wan、Bo Li 和 Tian Xia。《面向自动驾驶的多视图3D目标检测网络》。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2017年。2[12] Yu Chen、Chunhua Shen、Xiu-Shen Wei、Lingqiao Liu 和 Jian Yang。《对抗式PoseNet：用于人体姿态估计的结构感知卷积网络》。载于 IEEE/CVF国际计算机视觉会议（ICCV），2017年。3[13] Christopher Choy、JunYoung Gwak 和 Silvio Savarese。《4D时空卷积网络：Minkowski卷积神经网络。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年。1[14] Qi Dou、Hao Chen、Yueming Jin、Lequan Yu、Jing Qin和Pheng-Ann Heng。用于从CT体积数据中进行肝脏自动分割的三维深度监督网络。发表于医学图像计算与计算机辅助干预国际会议，2016年。3[15] Jing Du、Guorong Cai、Zongyue Wang、Shangfeng Huang、Jinhe Su、Jos´e Marcato Junior、Julian Smit和Jonathan Li。ResDLPS-Net：用于大规模点云语义分割的联合残差-稠密优化。ISPRS摄影测量与遥感学报，2021年。7, 17[16] Siqi Fan、Qiulei Dong、Fenghua Zhu、Yisheng Lv、Peijun Ye和Fei-Yue Wang。SCF-Net：学习用于大规模点云分割的空间上下文特征。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2021年。3, 6[17] Benjamin Graham、Martin Engelcke和Laurens Van Der Maaten。使用子流形稀疏卷积网络进行三维语义分割。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2018年。1, 8, 17[18] Benjamin Graham、Martin Engelcke和Laurens Van Der Maaten。使用子流形稀疏卷积网络进行三维语义分割。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2018年。2[19] Steven Guan、Amir A Khan、Siddhartha Sikdar和Parag V Chitnis。用于二维稀疏光声层析成像伪影去除的全稠密UNet。IEEE生物医学与健康信息学期刊，第568–576页，2019年。2, 3[20] 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，2021年。3, 16[21] Yulan Guo、Hanyun Wang、Qingyong Hu、Hao Liu、Li Liu和Mohammed Bennamoun。三维点云深度学习：综述。IEEE模式分析与机器智能汇刊，43 12 4338 4364, 2020 1 ( ): – 。[22] Timo Hackel、N. Savinov、L. Ladicky、Jan D. Wegner、K. Schindler和M. Pollefeys。SEMANTIC3D.NET：一种新的大规模点云分类基准。发表于ISPRS摄影测量、遥感与空间信息科学年刊，2017 1。Zhao [23] Hengshuang、Jiang Li、Jia Jiaya、Torr Philip H.S.和Koltun Vladlen。Point Transformer。发表于IEEE/CVF国际计算机视觉会议（ICCV），2021年。3, 6, 16, 17[24] Qingyong Hu、Bo Yang、Guangchi Fang、Yulan Guo、Ale˘s Leonardis、Niki Trigoni和Andrew Markham。SQN：大规模三维点云的弱监督语义分割。发表于欧洲计算机视觉会议（ECCV），第600–619页。Springer，2022年。1[25] Qingyong Hu、Bo Yang、Sheikh Khalid、Wen Xiao、Niki Trigoni和Andrew Markham。迈向城市尺度三维点云语义分割：数据集、基准与挑战。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2021年。2, 6, 7, 8, 13, 19[26] Qingyong Hu、Bo Yang、Sheikh Khalid、Wen Xiao、Niki Trigoni和Andrew Markham。SensatUrban：学习

从城市尺度摄影测量点云中获取语义。国际计算机视觉杂志，130 343，2022 2 316 1 ( ): – 。[27] 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。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2020年。1, 3, 6, 7, 8, 16, 17, 19[28] Qingyong Hu、Bo Yang、Linhai Xie、Stefano Rosa、Yulan Guo、Zhihua Wang、Niki Trigoni和Andrew Markham。使用随机采样学习大规模点云的语义分割。IEEE模式分析与机器智能汇刊，2021年。1[29] Huimin Huang、Lanfen Lin、Ruofeng Tong、Hongjie Hu、Qiaowei Zhang、Yutaro Iwamoto、Xianhua Han、Yen-Wei Chen和Jian Wu。Unet 3+：用于医学图像分割的全尺度连接式UNet。发表于IEEE国际声学、语音与信号处理会议，2020 2, 3, 18 。[30] Qiangui Huang、Weiyue Wang和Ulrich Neumann。用于点云三维分割的循环切片网络。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2018 3 。[31] Diederik P Kingma和Jimmy Ba。Adam：一种随机优化方法。国际学习表征会议，2015年。6[32] Abhijit Kundu、Xiaoqi Yin、Alireza Fathi、David Ross、Brian Brewington、Thomas Funkhouser和Caroline Pantofaru。用于三维语义分割的虚拟多视图融合。发表于欧洲计算机视觉会议（ECCV），2020年。1[33] Loic Landrieu、Hugo Raguet、Bruno Vallet、Cl´ement Mallet和Martin Weinmann。一种用于对三维点云语义标注进行空间平滑的结构化正则化框架。ISPRS摄影测量与遥感学报，2017 2 。[34] Loic Landrieu和Martin Simonovsky。使用超点图进行大规模点云语义分割。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2018年。3, 6, 8[35] Alex H Lang、Sourabh Vora、Holger Caesar、Lubing Zhou、Jiong Yang和Oscar Beijbom。Pointpillars：用于点云目标检测的快速编码器。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2019 2 。[36] Truc Le和Ye Duan。Pointgrid：一种用于三维形状理解的深度网络。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2018年。2[37] Chen-Yu Lee、Saining Xie、Patrick Gallagher、Zhengyou Zhang和Zhuowen Tu。深度监督网络。发表于人工智能与统计学，2015年。3[38]Guohao Li、Matthias Muller、Ali Thabet和Bernard Ghanem。DeepGCNs：图卷积网络能否像卷积神经网络一样加深？发表于IEEE/CVF国际计算机视觉会议（ICCV），2019年。1, 3, 6[39] Ying Li、Lingfei Ma、Zilong Zhong、Dongpu Cao和Jonathan Li。TGNet：用于三维点云分割的几何图卷积神经网络。发表于IEEE地球科学与遥感汇刊，2020年。7, 17

[40] Zhijian Liu、Haotian Tang、Yujun Lin 和 Song Han。Point- voxel CNN：用于高效三维深度学习。在神经信息处理系统进展（NeurIPS），2019年。1[41] Lingfei Ma、Ying Li、Jonathan Li、Weikai Tan、Yongtao Yu 和 Michael A Chapman。面向大规模环境中激光雷达点云的多尺度逐点卷积神经网络，用于三维目标分割。在IEEE智能交通系统汇刊，2021年。7、17[42] Daniel Maturana 和 Sebastian Scherer。Voxnet：一种用于实时目标识别的三维卷积神经网络。在IEEE/RSJ智能机器人与系统国际会议，2015 2 。[43] Alejandro Newell、Kaiyu Yang 和 Jia Deng。用于人体姿态估计的堆叠沙漏网络。在欧洲计算机视觉会议（ECCV），2016 3 。[44] Dong Nie、Rui Lan、Ling Wang 和 Xiaofeng Ren。用于点云分割多尺度处理的金字塔架构。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2022年。1[45] Charles Ruizhongtai Qi、Hao Su、Kaichun Mo 和 Leonidas J Guibas。PointNet：基于点集的深度学习，用于三维分类与分割。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2017年。1、3、6、8、16[46] Charles Ruizhongtai Qi、Li Yi、Hao Su 和 Leonidas J Guibas。PointNet++：在度量空间中基于点集进行深层次层级特征学习。在神经信息处理系统进展（NeurIPS）。1、2、3、7、8、17[47] Guocheng Qian、Yuchen Li、Houwen Peng、Jinjie Mai、Hasan Abed Al Kader Hammoud、Mohamed Elhoseiny 和 Bernard Ghanem。PointNeXt：重新审视PointNet++，采用改进的训练和扩展策略。神经信息处理系统进展（NeurIPS），2022年。1、6、16[48] Guocheng Qian、Xingdi Zhang、Abdullah Hamdi 和 Bernard Ghanem。通过图像预训练改进标准Transformer模型，以实现三维点云理解。arXiv预印本 arXiv:2208 12259, 2022 16 。。[49] Shi Qiu、Saeed Anwar 和 Nick Barnes。通过双边增强和自适应融合实现真实点云场景的语义分割。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2021年。1、2、3、6、7、8, 16, 17[50] Olaf Ronneberger、Philipp Fischer 和 Thomas Brox。U-Net：用于生物医学图像分割的卷积网络。在医学图像计算与计算机辅助干预国际会议，2015年。2、3、7、13、16、18[51] Radu Bogdan Rusu、Nico Blodow 和 Michael Beetz。用于三维配准的快速点特征直方图（FPFH）。在IEEE机器人与自动化国际会议，2009年。2[52] Radu Bogdan Rusu 和 Steve Cousins。三维时代已至：点云库（PCL）。在IEEE机器人与自动化国际会议。IEEE，2011年。1[53] Mohamed Samy、Karim Amer、Kareem Eissa、Mahmoud Shaker 和 Mohamed ElHelw。Nu-Net：用于语义分割的深度残差宽视场卷积神经网络

mentation。在IEEE/CVF计算机视觉与模式识别会议研讨会（CVPRW），2018 3。

[54] Shreyas Saxena 和 Jakob Verbeek。卷积神经网络织物。神经信息处理系统进展（NeurIPS），2016 3。Hui Shuai、Xiang Xu 和 Qingshan Liu。采用局部聚合分类器的反向注意力融合网络，用于三维点云语义分割。IEEE图像处理汇刊，2021 3, 6, 7, 8, 17。Ke Sun、Bin Xiao、Dong Liu 和 Jingdong Wang。用于人体姿态估计的深度高分辨率表征学习。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年。3[57] Christian Szegedy、Wei Liu、Yangqing Jia、Pierre Sermanet、Scott Reed、Dragomir Anguelov、Dumitru Erhan、Vincent Vanhoucke 和 Andrew Rabinovich。卷积网络的更深层次。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2015年。3[58] Weikai Tan、Nannan Qin、Lingfei Ma、Ying Li、Jing Du、Guorong Cai、Ke Yang 和 Jonathan Li。Toronto-3D：用于城市道路语义分割的大规模移动激光雷达数据集。在IEEE/CVF计算机视觉与模式识别会议研讨会（CVPRW），2020年。2、6、7、13、17、19[59] Haotian Tang、Zhijian Liu、Shengyu Zhao、Yujun Lin、Ji Lin、Hanrui Wang 和 Song Han。利用稀疏点体素卷积搜索高效三维架构。在欧洲计算机视觉会议（ECCV），2020年。1[60] Liyao Tang、Yibing Zhan、Zhe Chen、Baosheng Yu 和 Dacheng Tao。面向点云分割的对比边界学习。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2022年。6、16[61] Maxim Tatarchenko、Jaesik Park、Vladlen Koltun 和 Qian-Yi Zhou。用于三维密集预测的切线卷积。在IEEE/CVF计算机视觉与模式识别会议（CVPR），2018年。8[62] Lyne Tchapmi、Christopher Choy、Iro Armeni、JunYoung Gwak 和 Silvio Savarese。Segcloud：三维点云语义分割。在国际三维视觉会议，2017 1。Hugues Thomas、Charles R. Qi、Jean-Emmanuel Deschaud、Beatriz Marcotegui、Franc。ois Goulette 和 Leonidas Guibas。KPConv：面向点云的灵活可变形卷积。在IEEE/CVF国际计算机视觉会议（ICCV），2019年。1、3、6、7、8、16、17[64] Jonathan Tompson、Ross Goroshin、Arjun Jain、Yann Le-Cun 和 Christoph Bregler。利用卷积网络进行高效目标定位。在IEEE/CVF国际计算机视觉会议（ICCV），2015年。3[65] He Tong、Liu Yifan、Shen Chunhua、Wang Xinlong 和 Sun Changming。用于点云实例分割的实例感知嵌入。在欧洲计算机视觉会议（ECCV），2020 17。Haonan Wang、Peng Cao、Jiaqi Wang 和 Osmar R Zaiane。UCTransNet：从通道视角重新思考U-Net中的跳跃连接与Transformer。在AAAI人工智能会议论文集，2022 2。

[63] [66] [67] Jingdong Wang、Zhen Wei、Ting Zhang 和 Wenjun Zeng。深度融合网络。arXiv:1605.07716，2016年。3

[68] Lei Wang、Yuchun Huang、Yaolin Hou、Shenman Zhang 和 Jie Shan。用于点云语义分割的图注意力卷积。发表于IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年。1、3

[69] Yue Wang、Yongbin Sun、Ziwei Liu、Sanjay E Sarma、Michael M Bronstein 和 Justin M Solomon。用于点云学习的动态图卷积神经网络。《ACM图形学汇刊》，2019 6, 7, 16, 17。Chi-Chong Wong 和 Chi-Man Vong。面向关键道路物体和分布式上下文的高效室外三维点云语义分割。发表于 欧洲计算机视觉会议（ECCV），2020 3。Wenxuan Wu、Zhongang Qi 和 Li Fuxin。Pointconv：面向三维点云的深度卷积网络。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2019 17。Xiao Xiao、Shen Lian、Zhiming Luo 和 Shaozi Li。用于高质量视网膜血管分割的加权 Res-UNet。发表于国际医学与教育信息技术会议，2018年。2、3[73] Wang Xinlong、Liu Shu、Shen Xiaoyong、Shen Chunhua 和 Jia Jiaya。在点云中联合分割实例与语义。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年。17[74] Xu Yan、Chaoda Zheng、Zhen Li、Sheng Wang 和 Shuguang Cui。PointASNL：利用自适应采样和非局部神经网络进行鲁棒点云处理。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2020年。1、3、17[75] Bo Yang、Jianan Wang、Ronald Clark、Qingyong Hu、Sen Wang、Andrew Markham 和 Niki Trigoni。学习用于点云三维实例分割的目标边界框。发表于 神经信息处理系统进展（NeurIPS），2019年。16、17[76] Xiaoqing Ye、Jiamao Li、Hexiao Huang、Liang Du 和 Xi-aolin Zhang。结合上下文融合的三维循环神经网络，用于点云语义分割。发表于 欧洲计算机视觉会议（ECCV），2018 3。

[70] [71] [72] [77] Tan Yu、Jingjing Meng 和 Junsong Yuan。用于三维目标识别的多视图协调双线性网络。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2018年。2[78] Ziyin Zeng、Yongyang Xu、Zhong Xie、Wei Tang、Jie Wan 和 Weichao Wu。LACV-Net：通过局部自适应综合VLAD进行大规模点云场景语义分割。arXiv:2210.05870，2022年。2、6、7、8、16、17[79] Ziyin Zeng、Yongyang Xu、Zhong Xie、Wei Tang、Jie Wan 和 Weichao Wu。LEARD-Net：面向大规模点云场景的语义分割。《国际应用地球观测与地理信息杂志》，2022 3。

[80] Zhiyuan Zhang、Binh-Son Hua 和 Sai-Kit Yeung。Shell-net：利用同心壳层统计实现高效点云卷积神经网络。发表于 IEEE/CVF国际计算机视觉会议（ICCV），2019 6。

[81] Hengshuang Zhao、Li Jiang、Chi-Wing Fu 和 Jiaya Jia。PointWeb：增强点云处理的局部邻域特征。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年。1、3、6[82] Hengshuang Zhao、Jianping Shi、Xiaojuan Qi、Xiaogang

Wang 和 Jiaya Jia。金字塔场景解析网络。发表于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2017年。 3[83] ~~Liang Zhihao、Li~~ Zhihao、Xu Songcen、Tan Mingkui 和 Jia Kui。用于点云实例与语义联合分割的自预测。发表于 欧洲计算机视觉会议（ECCV），2020年。 17 [84] ~~Zongwei Zhou~~、Md Mahfuzur Rahman Siddiquee、Nima Tajbakhsh 和 Jianming Liang。Unet++：通过重新设计跳跃连接来利用图像分割中的多尺度特征。《IEEE医学影像汇刊》，第1856–1867页，2019年。 2，3，4，7，13，18[85] Qikui Zhu、Bo Du、Baris Turkbey、Peter L Choyke 和 Pingkun Yan。用于前列腺分割的深度监督卷积神经网络。发表于 国际神经网络联合会议，2017年。 3

附录

概述

本补充材料的组织结构如下：

- 章节A 介绍了网络架构。
- 章节B介绍了我们实验中使用的评估指标，并提供了各种基准和任务上的更多定量分割结果。此外，我们进行了消融研究，以分析所提出方法的有效性。
- 章节 C展示了更多语义分割结果的可视化内容。

A. 网络架构详情

主论文的图2 展示了 U-Next 的整体架构。本节将更详细地介绍所提出的架构。

首先，将维度为 $N \times d_{in}$ 的点云输入网络（为简洁起见，此处省略批次维度），其中 N 表示点云中的点数，d_{in}表示输入点的特征维度。对于 S3DIS [2],和 Toronto3D [58], 以及 SensatUrban [25] 数据集，每个点由其三维坐标和颜色信息表示（即 x-y-z-R-G-B）。我们使用一个单层多层感知机，其中包含八个 1×1 卷积核，并结合 ReLU 和批归一化来提取初始嵌入。

网络逐步将点云的分辨率从高到低降低：（$\frac{N}{4} \rightarrow \frac{N}{16} \rightarrow \frac{N}{64} \rightarrow \frac{N}{256} \rightarrow \frac{N}{512}$），而特征的通道维度则逐步增加：（$16 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512$），如表 6所示。请注意，点数和特征维度是每个编码块的输出；在同一层（相同的i）中，每个编码块的输出具有相同的点数和特征维度。在此过程中会生成不同分辨率的点云，所提出的 U-Next 利用多个子网络融合多尺度特征图。

最后，我们使用三个全连接层，其中包含一个丢弃层（$64 \rightarrow 32 \rightarrow$ (dp) $\rightarrow c$，丢弃率设为 0.5），并使用一个 Softmax层，将抽象的语义特征转换为点云中每个点的标签分类得分。语义分割结果由得分最高的标签决定。

U-Net变体的演进细节。如图6所示，我们进一步比较了U-Net变体的演进细节，包括U-Net [50], U-Net+、U-Net++ [84]、U-Net+^{d} 以及所提出的U-Next架构。显然，U-Net+ 和U-Net++ 本质上是由不同层级的U-Net集成而成，但共享编码层。类似地，所提出的U-Next也可以看作是U-Net的集成，但完全由U-Net L^1 子网络组成，这意味着U-Next中的每个节点都融合了水平、由低到高和由高到低的特征，同时语义差距也很小。此外，与U-Net++不同，我们的U-Next不使用密集跳跃连接，这主要是出于效率考虑（已在第B.3节中验证）。

二维与三维网络架构对比。如图 7所示，我们进一步直观地比较二维与三维网络的架构和操作。显然，输入的二维图像具有结构化且规则的特点，而三维点云则是离散且非规则的。因此，尽管二者的目标都是完成相同的分割任务，但仍存在以下差异：**(1) 采样**。二维图像通常使用带步幅的最大池化来降低分辨率，而三维点云通常使用随机采样或最远点采样进行下采样；**(2) 邻域定义**。考虑到三维点云分布不均且密度不均匀，通常使用K近邻和球形搜索来确定邻域；而二维图像中的邻域则可通过邻域像素轻松确定；**(3) 底层操作**。标准卷积神经网络是二维分割的基本操作，而共享多层感知机通常用作三维点云中的基本操作。总之，鉴于三维点云无序且非结构化的特性，加之邻域不稳定以及激进的下采样操作，在U-Net 架构中聚合来自不同节点的多尺度特征时应更加谨慎。受此启发，我们回到起点，尽可能多地堆叠U-Net L^1 编解码器，并在U-Next 框架中引入多级深度监督，以确保聚合特征之间的语义差距尽可能小，从而进一步提升语义分割任务的性能。

节点连接流水线。为了进一步说明我们如何聚合不同节点的特征，我们使用伪代码展示 U-Next 架构中的节点连接流水线，如算法1所示。基本上，如图 8 (A-C) 所示，带有 $j = 0$ 的节点仅接收前一个编码层的输出作为输入；带有$i = 0$ ($j > 0$)的节点接收两个输入，且这两个输入均来自同一个 U-Net L^1 子网络；带有 $i > 0$ 和 $j > 0$ 的节点则接收前一个编码层的输出作为输入，同时也接收来自同一个 U-Net L^1 子网络的两个输入。与 U-Net++的密集跳跃连接不同，U-Next 在 U-Net+^{d}的基础上引入了额外的长跳跃连接，因为密集跳跃连接并未带来明显的性能提升（已在第B.3节中验证）。图8 (D-E)进一步说明了额外长跳跃连接中由 $x^{i,\ j}$ 表示的特征图堆叠，当且仅当节点$X^{i,\ j}$ 是第 i行中的最后一个节点时才会进行堆叠。

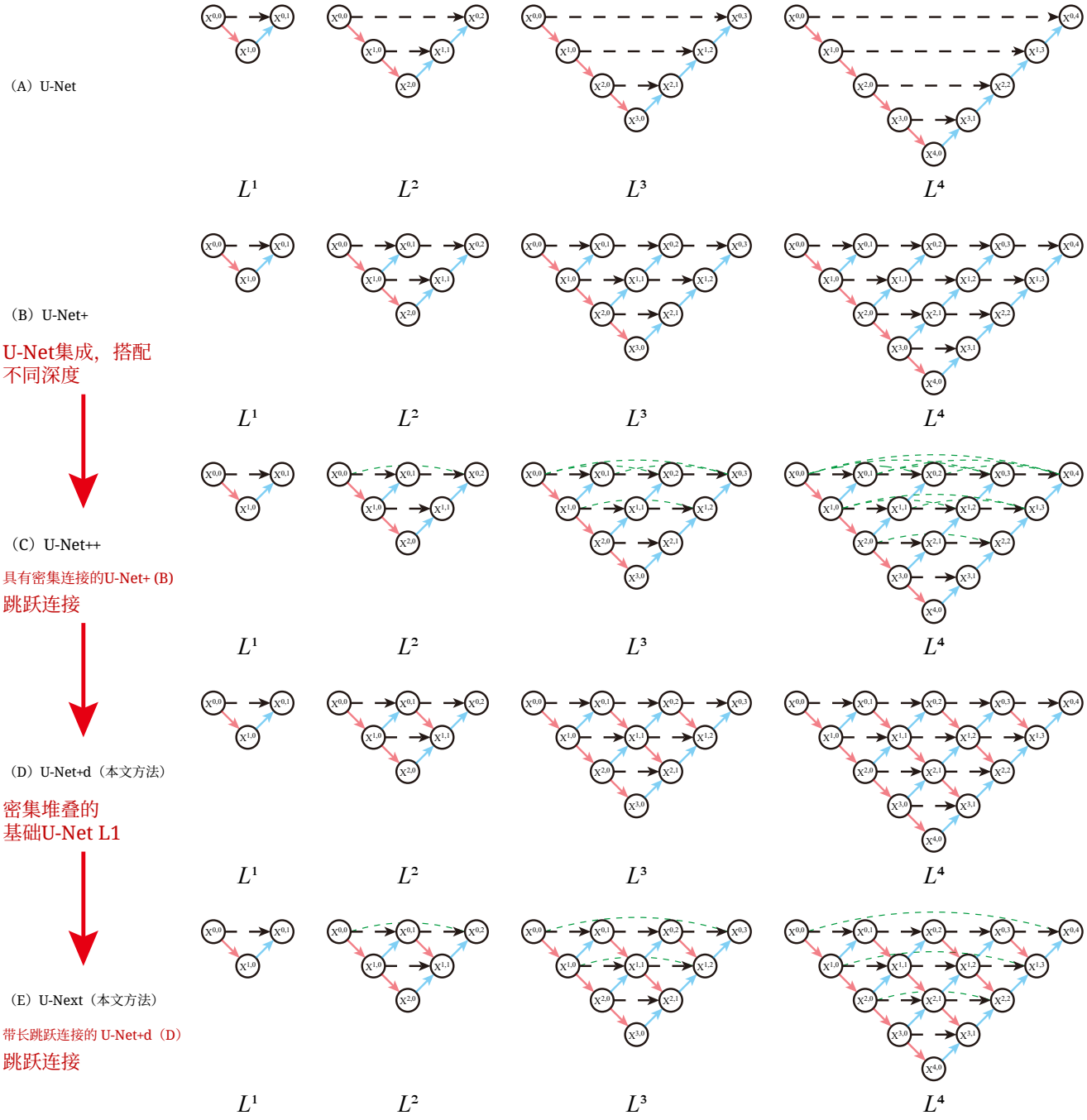


图6. U-Net变体的详细层级演变。

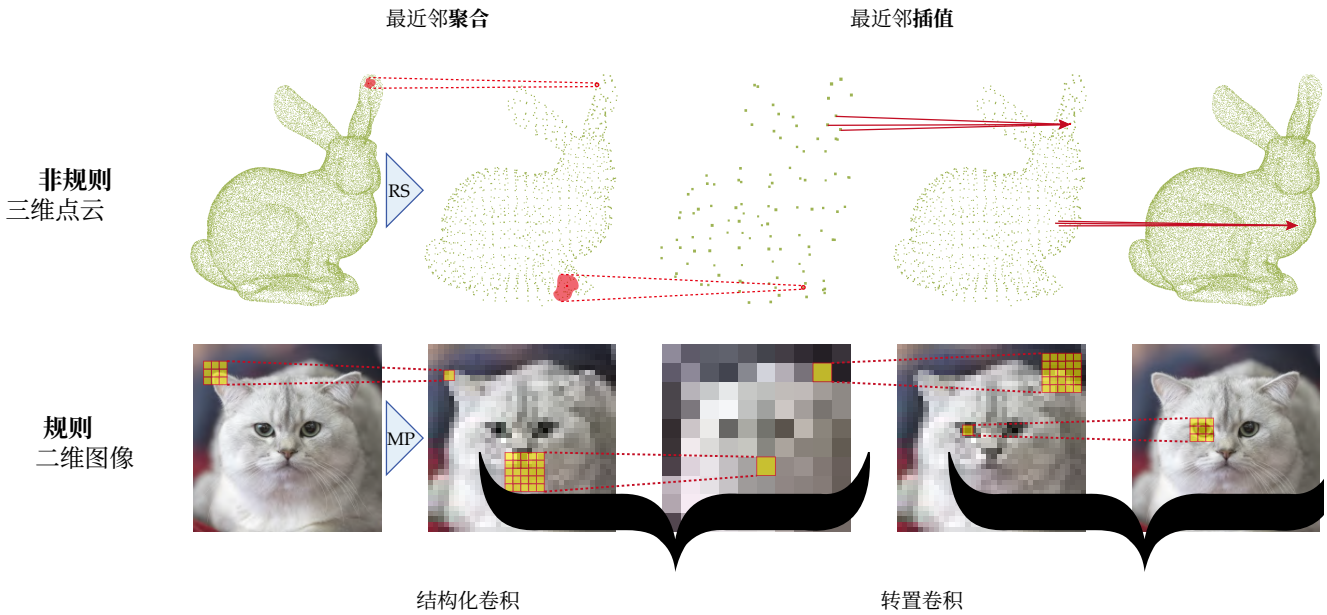


图7. 非规则三维点云处理与规则二维图像处理示意图。RS：随机采样；MP：最大池化。通常，使用最大池化对规则二维图像进行下采样，在此过程中，使用结构化卷积神经网络提取局部特征图；使用随机采样或最远点采样对非规则三维点云进行下采样，在此过程中，使用非结构化的最近邻聚合提取局部特征图。随后，通常使用结构化转置卷积神经网络和像素重排来恢复二维图像的高分辨率信息，而三维点云通常采用不均匀的最近邻插值。

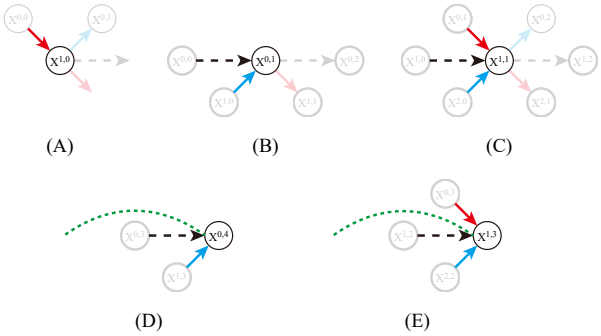
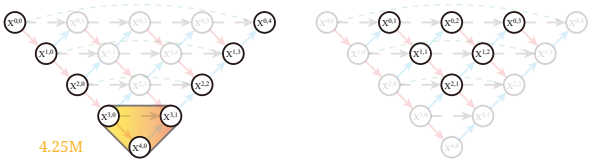


图8. U-Net+^{d} 和 U-Next 中不同类型节点的节点连接流水线。每种节点类型的输入均有明确标示。U-Net+^{d} 仅在（A-C）中展示，而 U-Next 在 U-Net+^{d} 的基础上还包含（D-E）。



额外参数可忽略不计

图9。U-Next 架构中参数量（百万）的示意图。(A) 展示了 U-Next 中“U-Net”骨干网络的参数量。(B) 展示了 U-Next 中相较于“U-Net”骨干网络新增节点的参数量。

B. 实验结果

B.1. 评估指标

为定量分析所提出架构的性能，采用总体准确率（OA）、各类别交并比（IoUs）和平均交并比（mIoU）作为评估指标，如下所示：

$$\mathrm{OA} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{TP_i}}{\mathrm{N}} \tag{5}$$

$$\mathrm{IoU_i} = \frac{\mathrm{TP_i}}{\mathrm{TP_i} + \mathrm{FP_i} + \mathrm{FN_i}} \tag{6}$$

$$\mathrm{mIoU} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{IoU_i}}{\mathrm{n}} \tag{7}$$

其中，TP 表示真正例样本的数量，FP 表示假正例样本的数量，FN 表示假负例样本的数量， i表示 i_{th} 语义类别， n 表示语义类别的数量，N 表示总点数。

B.2. 补充定量结果

S3DIS Area5 上的定量实验。 我们还报告了 U-Nextf 框架的定量结果

算法1： 节点连接流水线

输入： 初始嵌入

输出： 恢复的原始分辨率特征图

```
对于 j 在 range(N) 执行
    如果 j==0则
        对于 i 在 range(N − j) 执行
            x = 编码(x);
            列表 [j] ← x;
            x =下采样(x);
        结束循环
    结束条件判断
    否则如果j>0则
        对于 i 范围(N − j) 执行
            x0 = 列表 [j − 1][i];
            x1 = 上采样(列表 [j − 1][i + 1]);
            如果 i>0则
                x2 = 下采样(x);
            结束如果
            如果 i+j==N-1则
                // 每一层的最后一个节点
                x3 =列表 [0][i];
            结束如果
            x =拼接以上所有内容;
            x = 编码(x);
            列表 [j] ← x;
        结束循环
    结束条件判断
    // 切换到下一个序列;
结束循环
```

表 6. 我们的 U-Next 中编码块的点数和特征维度输出。同一层中每个节点的输出大小相同。

层	点	维度
L_0	$\frac{N}{4}$	16
L_1	$\frac{N}{16}$	64
L_2	$\frac{N}{64}$	128
L_3	$\frac{N}{256}$	256
L_4	$\frac{N}{512}$	512

以 RandLA-Net [27]和 LACV-Net [78] 作为基线方法，在表 7 中对 S3DIS 区域 5 进行了评估。从结果可以看出，将原始 U-Net [50] 替换为我们的 U-Next 后，整体分割性能显著提升。我们还注意到，采用 LACV-Net 的 U-Next 几乎达到了最佳性能，仅略逊于 PointNeXt-XL [47]，；后者采用了许多训练技巧，包括数据增强和最远点采样，并进行了复杂的预处理/后处理。还值得注意的是

表7. 在 S3DIS（区域*5*）上与最先进方法的定量比较。

year	方法	OA	mAcc	mIoU
2017	PointNet [45]	-	49.0	41.1
2018	DGCNN [69]	83.6	-	47.9
2019	KPConv [63]	-	72.8	67.1
2021	BAAF-Net [49]	88.9	73.1	65.4
2021	PCT [20]	-	67.7	61.3
2021	Point Transformer [23]	90.8	76.5	70.4
2022	CBL [60]	90.6	75.2	69.4
2022	PViT [48]	-	75.2	69.6
2022	PointNeXt-L [47]	90.0	75.3	69.0
2022	PointNeXt-XL [47]	90.6	76.8	70.5
2020	RandLA-Net [27]	87.2	71.4	62.5
Ours	RandLA-Net +U-Next	89.5 +2.3	76.2 +4.8	68.3 +5.8
2022	LACV-Net [78]	89.3	73.6	65.9
Ours	LACV-Net +U-Next	90.1 +1.8	77.1 +3.5	69.9 +4.0

PointNeXt-XL 拥有 4160 万个参数，是我们的 U-Next（LACV-Net）的七倍多。

Toronto3D 上的详细定量结果。表 8 展示了我们的 U-Next（RandLA-Net）在 Toronto3D 区域 2 上取得的详细定量结果。可以看出，我们的 U-Next 框架能够稳步提升 RandLA-Net 在所有类别上的分割性能。特别是，在道路标线和杆柱等局部几何信息模糊的类别上，我们的方法优于基线方法。这主要是因为我们的框架堆叠了更多子网络，以捕获充足的局部细节。

ScanNet v2 上的定量结果。ScanNet v2 数据集由通过 RGB-D 帧重建的 1,513 个室内房间扫描组成。该数据集划分为 1,201 个训练场景和 312 个验证场景，所有场景均标注了 20 个语义类别。在本研究中，我们评估了所提出的 U-Next（RandLA-Net）在 ScanNet v2 验证集上的性能，并在表9中报告了定量结果。实验结果表明，用我们的 U-Next 替换原始 U-Net [50] 后，两个基线网络的整体分割性能都能得到显著提升。具体而言，与基线 RandLA-Net 和 LACV-Net 相比，所提出的方法的平均交并比分别提高了 6.2% 和 5.7%。值得注意的是，U-Next（LACV-Net）在 ScanNet v2 数据集验证集上达到了最先进水平，证明了我们所提出的 U-Next 框架的有效性。

实例分割的定量实验。 为考察我们提出的 U-Next 框架在语义分割任务之外的多功能性和泛化能力，我们进一步针对实例分割任务开展了实验。具体而言，我们使用 U-Next 架构重新实现了 3D-BoNet [75]，并在 S3DIS 数据集的区域5上评估其性能。我们报告了

表8. 不同方法在 Toronto3D 数据集（区域 *2*）上的详细定量结果。

RGB	方法	OA	mIoU	Road	道路标线	自然地物	建筑物	公用设施线	Pole	Car	围栏
No	PointNet++ [46]	92.6	59.5	92.9	0.0	86.1	82.2	60.9	62.8	76.4	14.4
	DGCNN [69]	94.2	61.7	93.9	0.0	91.3	80.4	62.4	62.3	88.3	15.8
	MS-PCNN [41]	90.0	65.9	93.8	3.8	93.5	82.6	67.8	71.9	91.1	22.5
	KPConv [63]	95.4	69.1	94.6	0.1	96.1	91.5	87.7	**81.6**	85.7	15.7
	TGNet [39]	94.1	61.3	93.5	0.0	90.8	81.6	65.3	62.9	88.7	7.9
	MS-TGNet [58]	95.7	70.5	94.4	17.2	95.7	88.8	76.0	73.9	94.2	23.6
	LACV-Net [78]	95.8	78.5	94.8	42.7	96.7	91.4	**88.2**	79.6	93.9	**40.6**
	RandLA-Net [27]	93.0	77.7	94.6	42.6	96.9	93.0	86.5	78.1	92.9	37.1
	+ U-Next	**96.0** +3.0	**79.2** +1.5	**95.1**	**44.8**	**97.2**	**93.5**	87.6	80.5	**94.3**	40.8
Yes	ResDLPS-Net [15]	96.5	80.3	95.8	59.8	96.1	90.9	86.8	79.9	89.4	43.3
	BAF-LAC [55]	95.2	82.2	96.6	64.7	96.4	92.8	86.1	83.9	93.7	43.5
	BAAF-Net [49]	94.2	81.2	96.8	67.3	96.8	92.2	86.8	82.3	93.1	34.0
	LACV-Net [78]	97.4	82.7	97.1	66.9	97.3	93.0	87.3	83.4	93.4	43.1
	RandLA-Net [27]	94.4	81.8	96.7	64.2	96.9	94.2	88.0	77.8	93.4	42.9
	+ U-Next	**97.7** +3.3	**84.0** +2.2	**97.3**	**68.6**	**97.7**	**95.2**	**88.4**	**86.1**	**95.1**	**43.2**

桌子 9. 与 ScanNetv2 上最先进水平方法的定量比较。

方法	验证集 平均交并比 (%)
PointNet++ [46]	53.5
PointConv [71]	61.0
PointASNL [74]	63.5
KPConv [63]	69.2
SparseConvNet [17]	69.3
PointTransformer [23]	70.6
RandLA-Net [27]	63.5
RandLA-Net +U-Next	69.7 +6.2
LACV-Net [78]	66.1
LACV-Net +U-Next	71.8 +5.7

在表中使用平均类别精确率（mPrec）和平均类别召回率（mRec）作为评估指标的定量结果10。结果表明，除了语义分割之外，我们的 U-Next 架构还能够显著提升实例分割性能。具体而言，与基线方法 3D-BoNet 相比，所提出的方法的 mPrec 和 mRec 分数分别提高了 5.6% 和 9.4%。这一发现表明，我们的 U-Next 框架能够很好地泛化到不同任务，并凸显了其在各类计算机视觉任务中的潜力。

可训练参数。 主论文的表 4 报告了 U-Next 的可训练参数。这里进一步详细列出了架构不同层中的参数，如图 9所示。可以看出，最深层的 U-Net L^1 子网络占据的参数数量最多，约占总参数的 75.76%；与“U-Net”骨干网络相比，U-Next 中新增节点的参数仅占一小部分（约 10.57%）。因此，与原始 U-Net 相比，所提出的 U-Next 并未引入过多额外参数。

B.3. 额外消融研究

本节从以下四个方面对所提出的 U-Next 进行消融研究。所有消融

表10. S3DIS（区域*5*）上**实例分割** 结果（%）的定量比较。*表示我们的实现。

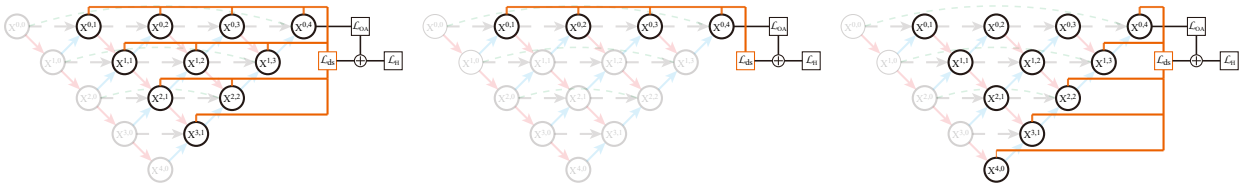
架构	平均精确率 (%)	mRec (%)
ASIS [73]	55.3	42.4
3D-BoNet [75]	57.5	40.2
PN++ [83]	60.1	47.2
InsEmb [65]	61.3	48.5
3D-BoNet*	55.2	39.1
3D-BoNet + U-Next	60.8 +5.6	48。5 +9.4

桌子11. 不同级别 U-Next 在 S3DIS（区域*5*）上的语义分割结果（%）和参数（百万）。

架构	OA (%)	mIoU (%)	参数 （百万）
U-Next L^4	89.5	68.3	5.61
U-Next L^3	89.1	66.1	1.43
U-Next L^2	86.8	60.6	0.31
U-Next L^1	81.0	49.6	0.05

网络均在区域1至4和区域6上训练，并在 S3DIS [2] 数据集的区域5上测试。RandLA-Net [27] 用作主要基线模型。

不同层级 U-Next 的消融研究。在这里，我们比较了不同层级 U-Next 的性能和参数数量。减少 U-Next 的层数可以显著缩短推理时间，但分割性能也会下降。因此，U-Next 的层级应在计算成本和准确性之间取得平衡。如桌子 11所示，U-Next L^3 将内存占用（参数数量）减少了 74.6%，而平均交并比（mIoU）分数仅下降了 2.1%。更激进地降低层级可以进一步减少内存占用，但代价是准确性显著下降。另一方面，我们还对不同层级的 U-Net 进行了类似实验，如桌子12所示。可以看出，层级越高，我们的 U-Next 相较于 U-Net 框架的性能提升越大，这主要是因为随着层级增加，堆叠的子网络更多，也意味着特征融合更加充分。



(A) 多层深度监督　　(B) 全分辨率深度监督　　(C) 横向深度监督

图10. p的示意图　　提出的多层深度监督、提出的全分辨率深度监督，以及后期　　ral
[提出的深度监督　　29].

表12. S3DIS（区域*5*）上不同层级U-Net的语义分割结果（%）和参数（百万）。

架构	OA (%)	平均交并比 (%)	参数 (百万)
U-Net L^4	87.2	62.5	4.99
U-Net L^3	86.1	61.2	1.24
U-Net L^2	84.7	58.8	0.30
U-Net L^1	80.7	49.1	0.05

表 13. 不同架构在 S3DIS（区域*5*）上的语义分割结果（%）和参数（百万）。

架构	OA (%)	mIoU (%)	参数 (百万)
宽型 U-Net	87.4	63.2	5.64
宽型 U-Net++	87.8	64.9	6.09
U-Next	89.5	68.3	5.61

表14. U-Next 在 S3DIS（区域*5*）上采用不同跳跃连接时的语义分割结果（%）和参数（百万）。conn.：连接。

架构	OA (%)	mIoU (%)	参数 (百万)
无跳跃连接	89.2	67.4	5.51
密集跳跃连接	89.4	67.9	5.66
长跳跃连接	89.5	68.3	5.61

宽型 U-Net 和 U-Net 的消融研究++。 与先前的 U-Net [50]和 U-Net++ [84]框架相比，所提出的 U-Next 确实引入了额外的参数。为验证 U-Next 的性能提升并非源于额外参数，而是源于网络架构固有的优势，我们进行了以下实验。我们通过增加编码器和解码器的特征输出维度，人为增加 U-Net 和 U-Net++ 的参数量。具体而言，特征的通道维度按以下方式增加：（$16+2 \rightarrow 64+4 \rightarrow 128+8 \rightarrow 256+16 \rightarrow 512+32$）。表13 展示了定量分析结果。显然，尽管引入额外参数能够提升分割性能，但所提出的 U-Next 的分割性能仍显著优于宽型 U-Net 和 U-Net++，这验证了所提出的 U-Next 的性能提升主要得益于其网络架构的优势。

长跳跃连接的消融研究。所提出的 U-Next 采用了长跳跃连接设计。值得注意的是

表 15. U-Next 在 S3DIS（区域*5*）上采用不同深度监督的语义分割结果（%）。DS.：深度监督。

架构	OA (%)	mIoU (%)
无深度监督。	89.0	66.7
全分辨率深度监督。	89.2	67.5
侧向深度监督。	89.2	67.7
多级深度监督。	89.5	68.3

U-Net++利用密集跳跃连接来连接同一层级的所有节点。为验证我们的长跳跃连接设计的有效性（以及为何不选择 U-Net++ 的设计），我们开展了这项消融研究。在本实验中，我们将密集长跳跃连接集成到 U-Next 中。如桌子 14 所示，可以观察到，尽管采用密集跳跃连接的网络比不使用长跳跃连接的网络性能更优，但其性能与使用长跳跃连接的网络相当。这可能是因为每个子网络都能充分学习不同尺度的局部特征，因此密集跳跃连接并非十分必要，反而会增加过拟合的风险。尤其是，密集跳跃连接会引入额外的参数和计算开销，导致冗余计算。因此，我们最终采用了基础 U-Net 中使用的标准长跳跃连接。

多层深度监督的消融研究。所提出的 U-Next 从优化和学习的角度，在每个解码器节点中引入了多层深度监督。我们注意到，U-Net++ [84]仅对生成的全分辨率特征图应用深度监督，而 U-Net3+ [29]仅对侧向节点应用深度监督。为了进一步验证多层深度监督的有效性，我们进行了这项消融研究（三者的比较如图10所示）。在本实验中，我们将 U-Net++中使用的全分辨率深度监督，以及 U-Net3+中使用的横向深度监督整合到 U-Next 中。表15 展示了定量分析结果。从结果可以看出，与未使用深度监督的网络相比，全分辨率和横向深度监督的性能更优，但不及多层深度监督。这主要是因为所提出的多层深度监督会训练每个 U-Net L^1 子网络，且各自独立；

而全分辨率深度监督或横向深度监督只在原始分辨率尺度或最右侧节点上进行训练，因此受监督子网络的数量有限。

C. 结果可视化

在本节中，我们直观比较基线方法 RandLA-Net [27]与我们提出的 U-Next 在所用三个大规模基准数据集上取得的语义分割结果。在每个可视化结果中，红框标出了 RandLA-Net 分割错误或边界分割不明显的部分。

S3DIS：图 11定性展示了 S3DIS [2] 数据集上的可视化对比。从结果可以清楚地看出，我们对板材、柱子、门和书柜的分割效果优于 RandLA-Net。此外，所提出的 U-Next 能够更平滑、准确地分割物体边界。这验证了反复融合来自语义差距极小的子网络的信息，有助于重建细粒度感知，从而缓解三维点云中不规则且不均匀的采样与聚合问题。

Toronto3D：图 12 定性展示了以 [58] 数据集为基础的 Toronto3D 可视化对比，初始输入采用 rgb 。从结果中可以观察到，我们对汽车、自然地物和围栏的分割效果显著优于 RandLA-Net。尤其是在人行横道的语义分割方面，与表现出色的 RandLA-Net 相比，所提出的 U-Next 分割更加准确，边界也更加清晰。

SensatUrban：图13 定性展示了 SensatUrban [25] 数据集上的可视化对比。由于在线测试集的真实标注尚未公开，为了更好地展示可视化效果，我们展示了验证集上的结果。可以观察到，与几乎无法分割铁路的 RandLA-Net 相比，所提出的 U-Next 能够清晰地将铁路与地面区分开来。同时，该方法在大型停车场和小型道路上的表现也有显著提升。

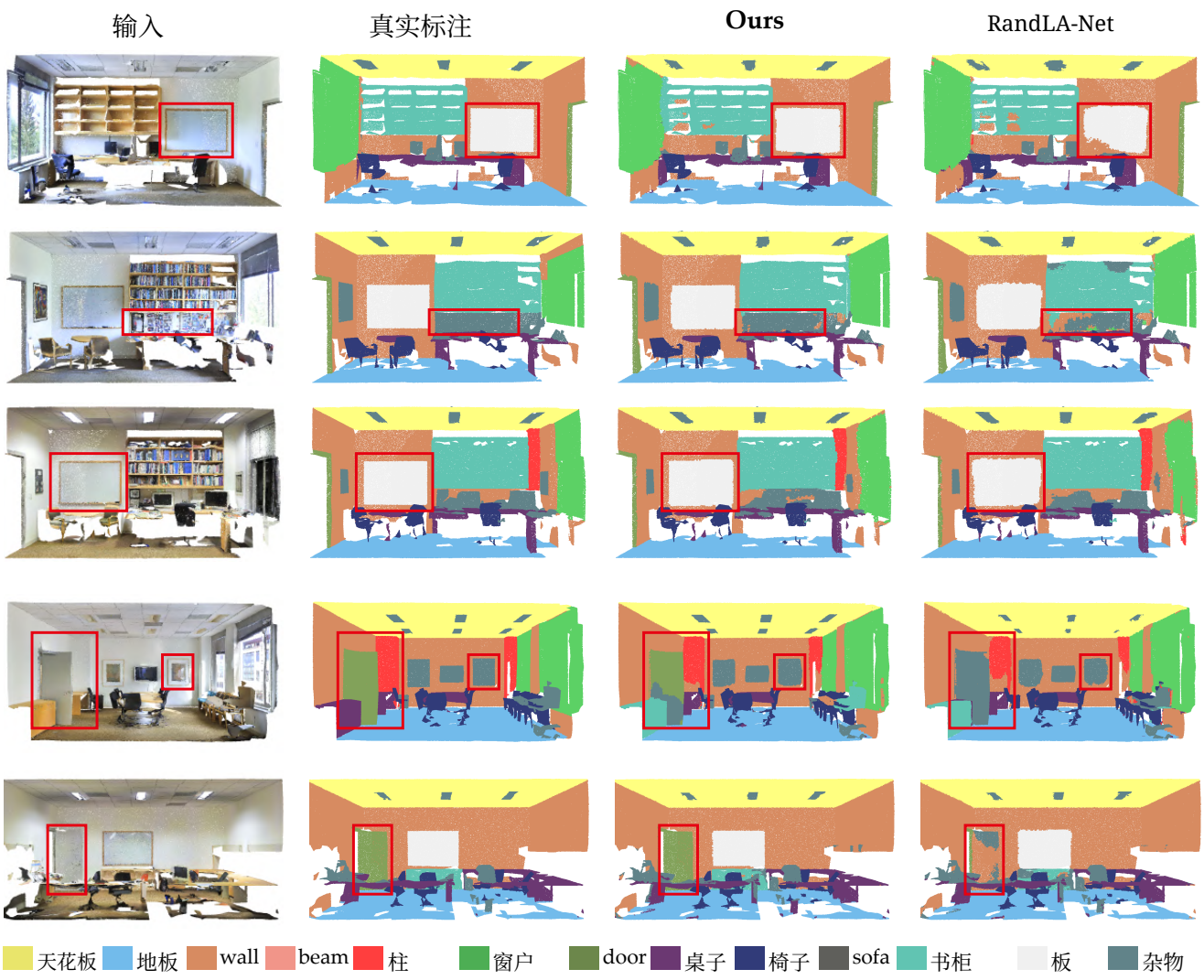


图11. S3DIS数据集上的语义分割结果可视化对比。

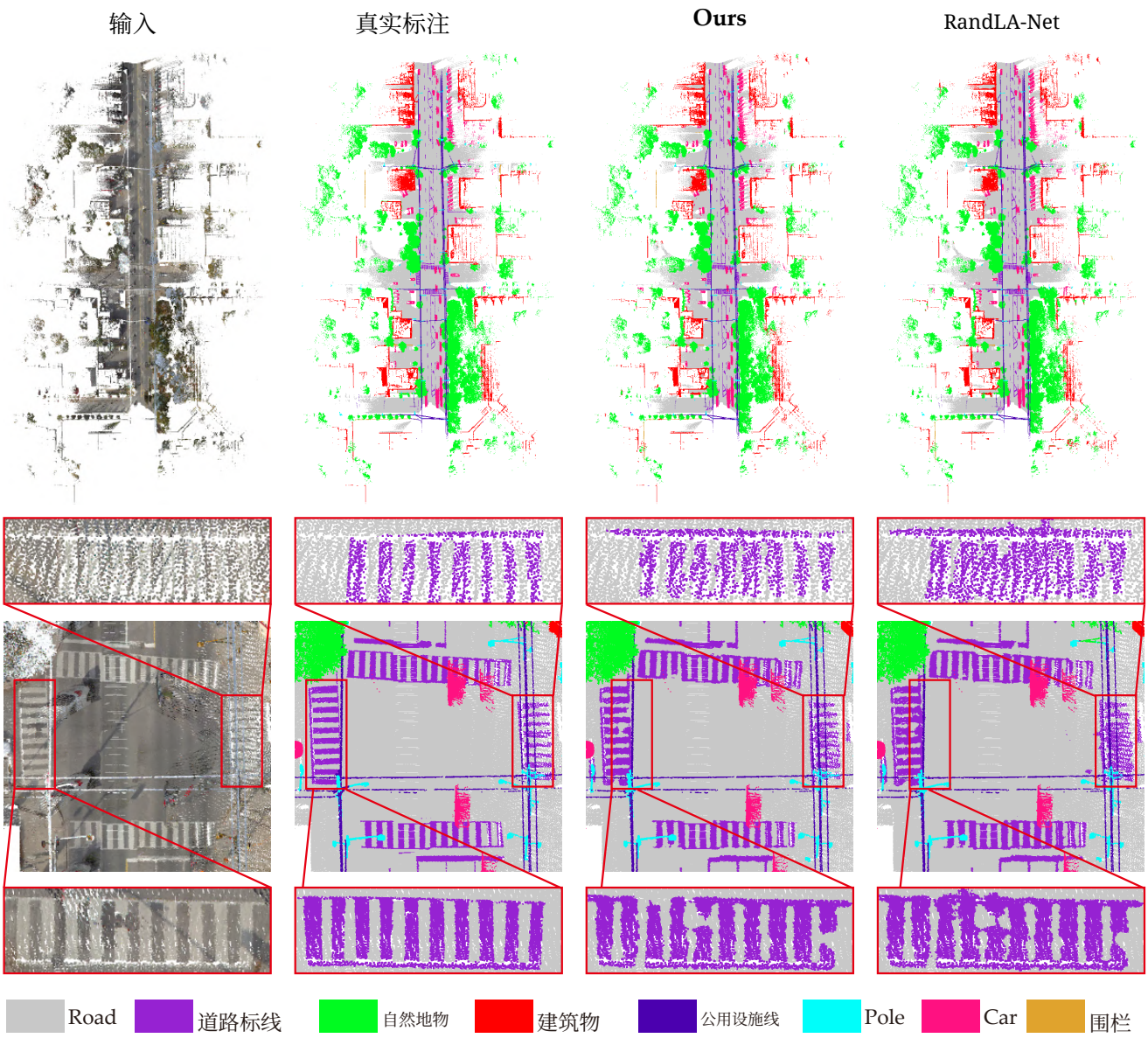


图 12. Toronto3D 数据集上的语义分割结果视觉对比。

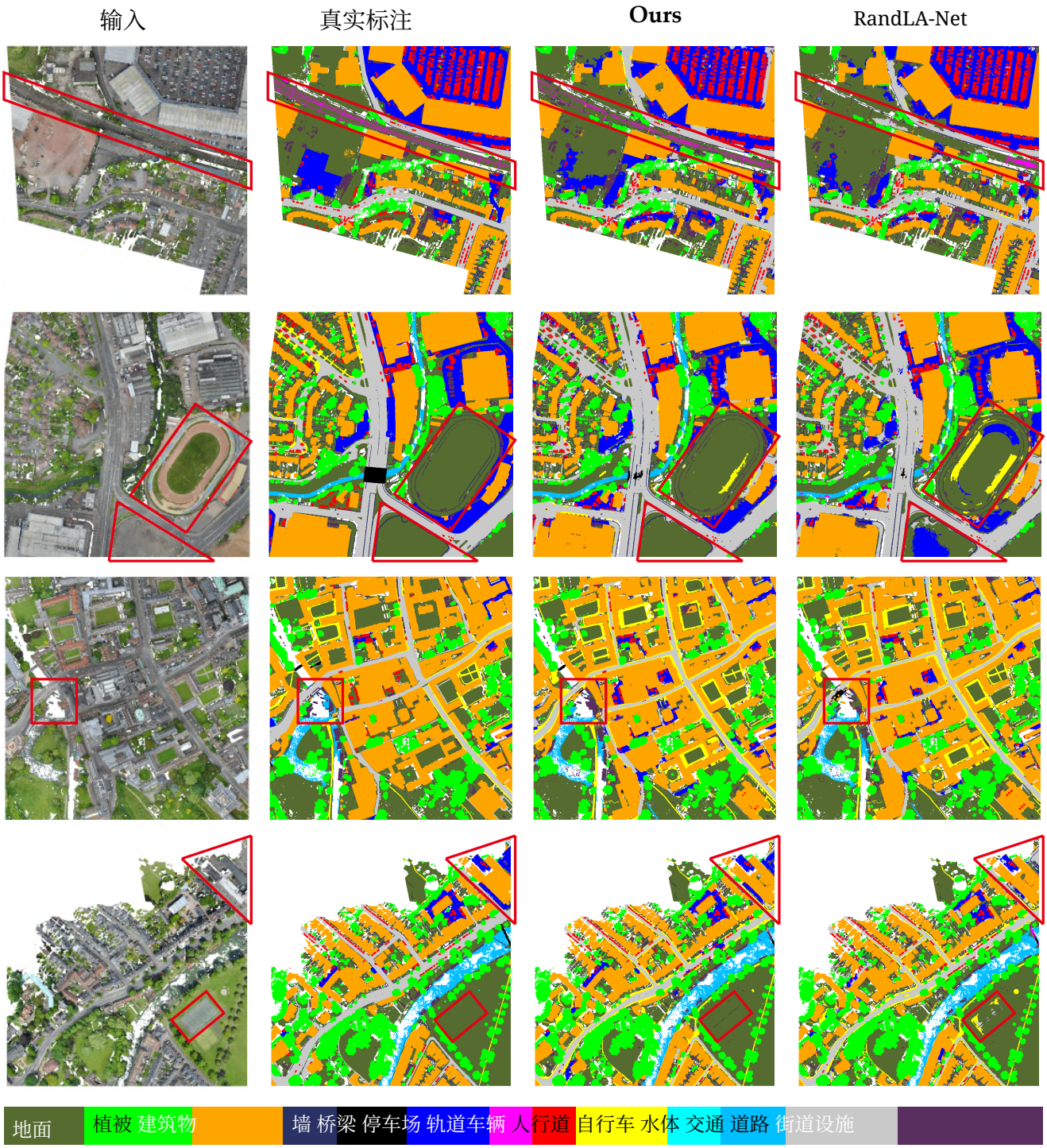


图 13. SensatUrban 数据集上的语义分割结果可视化对比。