Modeling and Simulation 建模与仿真, 2024, 13(3), 3653-3662
Published Online May 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.133333

Hans汉斯

基于改进PointNet++模型的毫米波点云语义分割

柳　荧，王　韬

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2024年4月22日；录用日期：2024年5月23日；发布日期：2024年5月31日

摘　要

毫米波成像技术在安检领域得到了普遍应用，研究基于毫米波点云图像的语义分割技术具有重要的意义。PointNet++采用了与任务无关的最远点采样(FPS)来逐步下采样点云，导致毫米波点云中为数不多的前景点信息丢失。因此，本文提出了一种基于自注意力机制的实例感知下采样点云语义分割网络。具体来说，本文结合自注意力机制实现面向任务的下采样策略来保留前景点，防止前景点信息的丢失。最后，由于毫米波点云图像中人体点云数量与前景点云数量极不平衡，改进使用Focal Loss作为语义分割损失函数以提升性能。实验结果表明，本文提出的语义分割模型相对于基准模型PointNet++在平均交并比(mIoU)方面有6.19%的提升，同时准确率有5.57%的提升。

关键词

PointNet++，毫米波点云图像，语义分割，自注意力机制

Semantic Segmentation of Millimeter Wave Point Clouds Based on Improved PointNet++

Ying Liu, Tao Wang

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 22nd, 2024; accepted: May. 23rd, 2024; published: May. 31st, 2024

Abstract

Millimeter-wave imaging technology has been widely applied in the security screening field, making the study of semantic segmentation techniques based on millimeter-wave point cloud images of pa-

文章引用: 柳荧, 王韬. 基于改进 PointNet++模型的毫米波点云语义分割[J]. 建模与仿真, 2024, 13(3): 3653-3662.
DOI: 10.12677/mos.2024.133333

ramount importance. PointNet++ employs a task-agnostic farthest point sampling (FPS) method for progressively downsampling the point cloud, resulting in the loss of sparse foreground information in the millimeter-wave point clouds. Consequently, this paper introduces an instance-aware downsampling point cloud semantic segmentation network based on the self-attention mechanism. Specifically, this work integrates the self-attention mechanism to implement a task-oriented downsampling strategy to preserve foreground points and prevent the loss of foreground information. Lastly, due to the significant imbalance between the number of human body points and foreground points in millimeter-wave point cloud images, an improved focal loss is utilized as the semantic segmentation loss function to enhance performance. Experimental results demonstrate that the semantic segmentation model proposed in this paper achieves a 6.19% improvement in mean Intersection over Union (mIoU) and a 5.57% increase in accuracy compared to the baseline model PointNet++.

Keywords

PointNet++, Millimeter-Wave Point Cloud, Semantic Segmentation, Self-Attention Mechanism

1. 引言

毫米波段(Millimeter-Wave, MMW)已广泛应用于人体安检系统[1] [2] [3]，特别是在机场、车站等公共场所，如何快速且准确地检测并识别人体隐匿物品成为了一项重要的技术挑战。

近期基于深度学习的目标检测算法在毫米波安检领域取得了显著的进展，陈国平[4]等人则在YOLOv3-Tiny 模型的基础上进行改进，引入注意力机制提升了毫米波图像中目标检测的准确性。而张格菲[5]等人提出了一种基于 YOLOv5 的毫米波图像目标检测方法，通过改进 GIOU_Loss 损失函数和网络结构优化，实现了对安检人员身上隐匿违禁品的高效识别和检测。程秋菊[6]等人提出了一种利用 Faster R-CNN 深度学习方法来检测人体上隐藏的危险物品的技术。通过将区域建议网络(RPN)与 VGG16 训练的卷积神经网络模型结合，并采用在线难例挖掘(OHEM)技术优化网络模型。这些算法主要使用毫米波全息图像投影后得到的二维灰度图作为网络输入，输出模型预测的边界框，用于定位毫米波图像中的隐匿物品，然而毫米波灰度图像信噪比和分辨率远低于光学图像，边界框中不仅包含检测目标，还有人体背景以及噪声，这不仅会影响目标识别的精度，也不利于安检人员的查看[7]。

近年来，点云数据成为三维空间数据处理的一个重要分支，尤其在毫米波安检领域展现出了独特的价值。相较于毫米波灰度图，点云数据直接提供了物体的三维空间信息，包括物体的形状、尺寸和在三维空间中的位置[8]。这种数据形式使得点云数据在处理具有复杂几何结构的对象时表现出了更高的效率和准确性。因此，对毫米波点云图像进行语义分割，可以更准确地定位隐匿物品，减少背景噪声和人体背景的干扰，便于安检人员查看。

PointNet [9]模型作为处理点云数据的先驱，通过学习每个点的空间编码并对所有点特征进行全局聚合，成功地将深度学习应用于点集数据处理。然而，PointNet 本身并不直接考虑点之间的局部结构，这限制了其在识别细粒度模式和处理复杂场景时的能力。为了克服这一限制，PointNet++模型被提出。PointNet++ [8]通过在嵌套分割的输入点集上递归应用 PointNet，引入了一种层次化的神经网络结构。这

使得模型能够度量空间中的距离，从而学习具有不同尺度的局部特征。此外，PointNet++还考虑了点集常常伴随的不同密度采样问题，通过提出的适应性特征学习层，有效地组合了多尺度特征，显著提升了处理点云数据的效率和鲁棒性。然而，尽管 PointNet++在处理非均匀采样密度的点集方面取得了显著进展，但在最远点下采样(FPS)算法中仍存在一些不足。Zhang [10]等人提出的 IA-SSD 模型指出，FPS 算法虽然能够在采样过程中尽可能覆盖整个点集，但在实际应用中，由于最远点采样的随机性，可能导致采样结果对原始数据的代表性不足，特别是在点云密度变化大时，可能会忽略重要的局部信息。为了解决这一问题，IA-SSD 中使用了一种面向任务的、基于学习的实例感知下采样策略，以更明确的保留对目标检测任务的更重要的前景点。在毫米波点云图像中，前景点(隐匿物品)数量远远低于背景点(噪声以及人体背景)数量，然而 PointNet++采用的最远点下采样策略会在逐层的特征提取过程中会丢失前景点信息，导致其分割性能下降，故本文以 PointNet++为基线模型提出了一种改进的基于自注意力机制的实例感知下采样毫米波点云语义分割方法。该方法能有效防止模型在对点云进行下采样时丢失前景点信息，提升 PointNet++语义分割性能。

2. 模型结构

2.1. 相关技术

2.1.1. PointNet++

如图 1 所示，PointNet++采用 Encoder-Decoder 架构，其编码层(Encoder)通过引入集合抽象模块(SA, Set Abstraction)来处理点云数据，该模块首先利用远点采样(FPS, Farthest Point Sampling)算法从输入的点云中选取一组关键点，这些关键点旨在覆盖点云的主要区域。接着模型根据这些关键点通过球查询将点云分组(Grouping)，每一组包含一个关键点及其球形区域内的点。对于每一组点，PointNet++使用一个小型的 PointNet 网络来提取该组内点的局部结构信息，最后通过最大池化操作，聚合该组内所有点的特征以提取点云局部特征同时实现模型的置换不变性。PointNet++通过多次重复上述步骤，实现了一个分层的特征学习编码器，使得模型能够捕获局部细节信息。

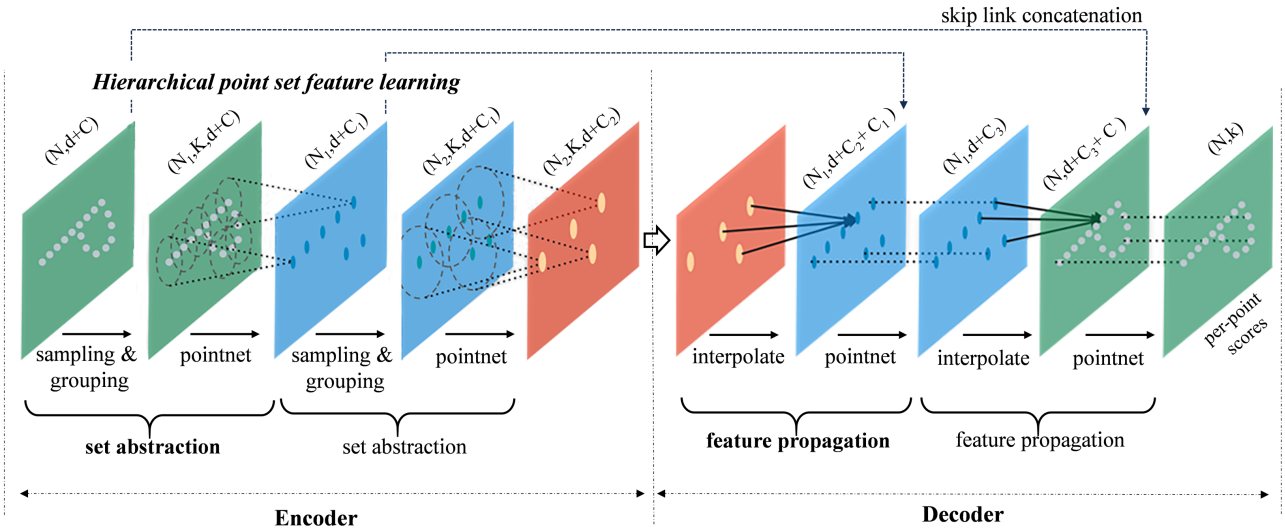

Figure 1. PointNet (++) network architecture for semantic segmentation
图 1. 面向语义分割任务的 PointNet (++)结构

PointNet++的解码层(Decoder)由与 SA 层相同数量的特征传播层(FP, Feature Propagation)组成。该过

程首先计算原始(或高密度)点云中每个点与下采样(或低密度)点集中点的距离，接着使用逆距离加权的方法对低密度点云中的特征进行加权平均，以此插值得到高密度点云中每个点的特征。这种加权平均通常依赖于 K 近邻算法来找到每个高密度点云在低密度点云中的最近邻点。此外，PointNet++还将这些上采样得到的特征与编码层中对应点云的特征进行拼接，进一步丰富每个点的特征表示。这种特征的上采样和融合过程将编码层学习到的高层特征映射回原始点云，从而使每个点都获得丰富的语义信息，进而实现语义分割任务。

2.1.2. 全局上下文网络(GCNet)

全局上下文网络(GCNet, Global Context Network)旨在捕捉和利用全局上下文信息以增强深度学习模型的性能，于 Cao [11]等人在 2019 年提出，该模块能在保持计算效率的同时，有效地整合全局上下文信息。

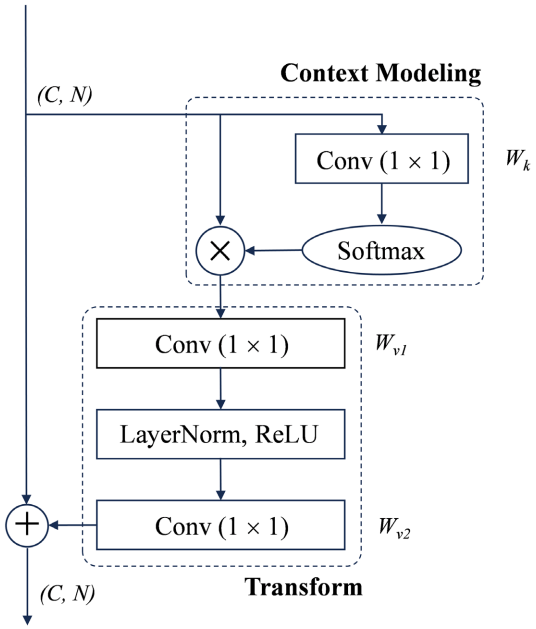

Figure 2. Global context network architecture
图 2. 全局上下文网络结构

GCNet 结合了非局部(Non-local)网络[12]和 Squeeze-and-Excitation (SE) [13]网络的特点，通过对全局上下文信息的有效利用，增强了网络对于重要特征的敏感性，并抑制了不重要特征的干扰。如图 2 所示，其首先使用全局平均池化来聚合全局上下文信息，该步骤可以捕获整个输入特征图的统计信息。然后，通过一系列变换(如 1 × 1 卷积)和激活函数处理聚合得到的全局信息，以学习特征间的复杂依赖关系。接着，利用学习到的全局上下文信息对原始特征进行重标定(即按元素乘法)，从而增强或抑制某些特征，使网络能够更加关注于那些对当前任务更为重要的信息。

2.1.3. 实例感知下采样

2021 年 Zhang 提出 IA-SDD 目标检测网络用于 3D 目标检测，该方法核心在于通过两种可学习的、面向任务的、实例感知的下采样策略，来层次化地选择属于感兴趣对象的前景点。具体来说，类别感知采样通过学习每个点的语义，实现了选择性下采样。其具体实现方式如图 3 所示，该策略通过在某一 SA 层提取特征后加入两个多层感知机(MLP, Multilayer Perceptron)组成的分类网络来利用潜在特征中的丰富语义信息。

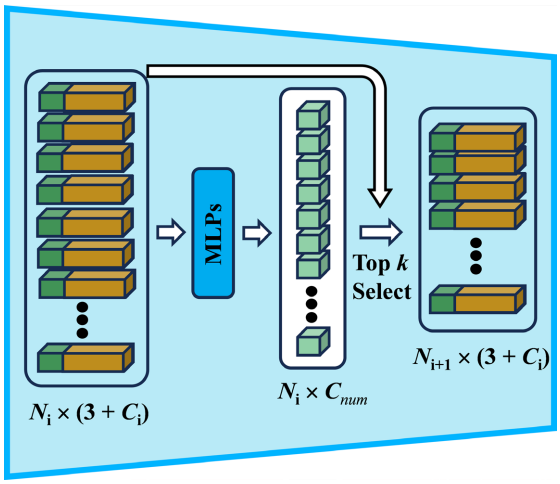

Figure 3. Instance-aware downsapling modle
图 3. 实例感知下采样模块

该方法的核心思想是，并不是所有的点对于目标检测任务都同等重要，尤其是前景点对于目标检测器来说，比背景点更为重要。但在每个点的语义预测时的方法较简单，可能会受到类别不平衡分布的影响。

2.2. 模型结构

本文采用的检测模型的总体结构框图如图 4 所示。我们将特征编码层分为两部分，一部分(SA0)保留了一组使用最远点采样的 SA 层来提取背景点(人体)以及前景点(隐匿物品)特征，另一部分(SA1)使用相同数量的基于 GCNet 模块的实例感知下采样的 SA 层专注于提取前景点特征。我们将上一个 SA 层提取的特征输入 GCNet 模块以整合全局上下文信息，输出的特征首先会经过 MLP 组成的分类网络来估计每个点的语义类别，从而实现选择性的降采样以保留更多的前景点，然后与对应 FP 层上采样后的特征拼接，以实现上下文信息的传递。

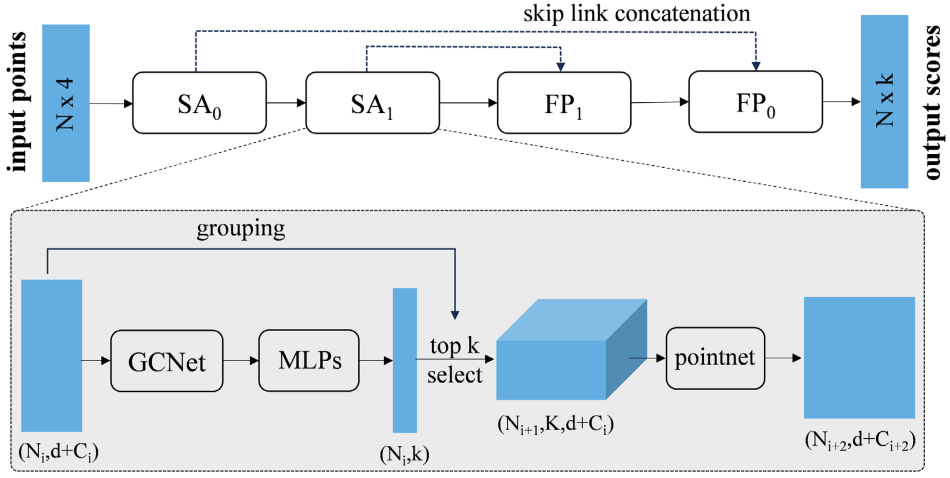

Figure 4. Instance-aware downsapling modle
图 4. PointNet++优化下采样策略后的网络结构

实例感知下采样策略能够有效识别并保留对于完成语义分割任务更为重要的前景点，这使得模型能够更加聚焦于毫米波点云图像的关键部分，如隐匿物品边缘或具有独特纹理的区域，这些区域对于区分

人体与隐匿物品至关重要。但在毫米波点云语义分割场景中，前景点人体背景点数量分布不平衡，单纯使用由两个 MLP 组成的下采样模块分类精度不高导致模型分割性能下降。因此，我们在实例感知下采样模块中集成了 GCNet 模块以增加算法复杂度，改进后不仅可以进一步提高实例感知下采样模块分类的准确性，以保留更多的前景点信息，同时，经过 GCNet 模块后的点云特征富含全局上下文信息，将其输入 FP 层，能进一步增强模型对场景的理解能力从而提升分割性能。

2.3. 损失函数

交叉熵损失(Cross-Entropy Loss)是分类任务中一种常用的损失函数。它用于衡量模型预测的概率分布与真实标签的概率分布之间的差异，目的是最小化这种差异，从而提高模型的预测准确性，对与二分类任务，其定义如公式(1)所示：

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{y}_i\right) + \left(1-y_i\right)\log\left(1-\hat{y}_i\right)\right] \tag{1}$$

其中， N 是样本总数， y 是第 i 个样本的真实标签(0 或 1)， $\hat{y}_i$ 是模型预测第 i 个样本为正类的概率。

Focal Loss[14]是一种在处理类别不平衡问题时特别有效的损失函数，尤其是在正负样本比例失衡极端的情况下，在传统的交叉熵损失中，大量的简单负样本(即容易分类的样本)可能会主导损失值，导致模型对少数正样本的学习不足。Focal Loss 通过重塑交叉熵损失，减少了容易分类样本的相对损失，使模型更加关注那些难以分类的样本，因为毫米波安检成像场景中，前景点数量远少于背景点数量，所以该损失函数非常适用于毫米波点云语义分割任务场景。Focal Loss 的定义如下：

$$L_{seg} = -\alpha_t\left(1-p_t\right)^{\gamma}\log\left(p_t\right) \tag{2}$$

$$\text{where } p_t = \begin{cases} p & \text{for forgound point} \\ 1-p & \text{otherwise} \end{cases}$$

其中， p_t 是模型对于每个类别的预测概率。如果样本被正确分类，则 p_t 即为模型输出的概率；参数 α_t 是用来平衡正负样本贡献的权重因子， γ 是一个调节参数，用于减少容易分类样本的损失贡献。随着 γ 的增大，对容易分类样本的惩罚也增大。在训练过程中，我们保持 $\alpha_t = 0.25$ 和 $\gamma = 2$ 的默认设置。

本文提出方法的损失函数如下列公式(3)所示，其由 Focal Loss 监督的语义分割损失以及由交叉熵损失监督的实例感知下采样的分类损失构成。

$$L_{total} = L_{seg} + L_{cls} \tag{3}$$

3. 实验结果与分析

3.1. 数据集

在毫米波点云图像语义分割领域仍缺少大规模公共数据集供研究人员使用，因此我们建立了一个毫米波三维点云图像的人体隐匿物品初等规模数据集，用于毫米波点云图像三维语义分割与目标检测研究。为模拟实际安检场景，本文数据集中准备了 7 件物品，包括钳子、扳手、尖嘴钳、锤子、枪、铁车模具和美工刀。选择 1~4 个物品放置于检测场景中模特身体的各个部位，经带宽 35 GHz 的毫米波成像系统扫描并重建其点云图像。参与数据采集的有 5 名模特，覆盖了不同的衣着、身高与体型。

支撑本文实验的数据规模约 2000 幅点云图像，且每幅点云图像至少有 1~4 个样本框，共包含 4366 个隐匿物品。其中每幅点云数量分布如图 5(a)所示，数据集中每幅点云图像总点数范围可从 21,509 延伸到 29,418。图 5(b)为数据集中每幅点云图像中前景点和背景点的数量分布，由图 5(c)可见，在整个数据集中前景点的数量仅占总数的 5.5%。

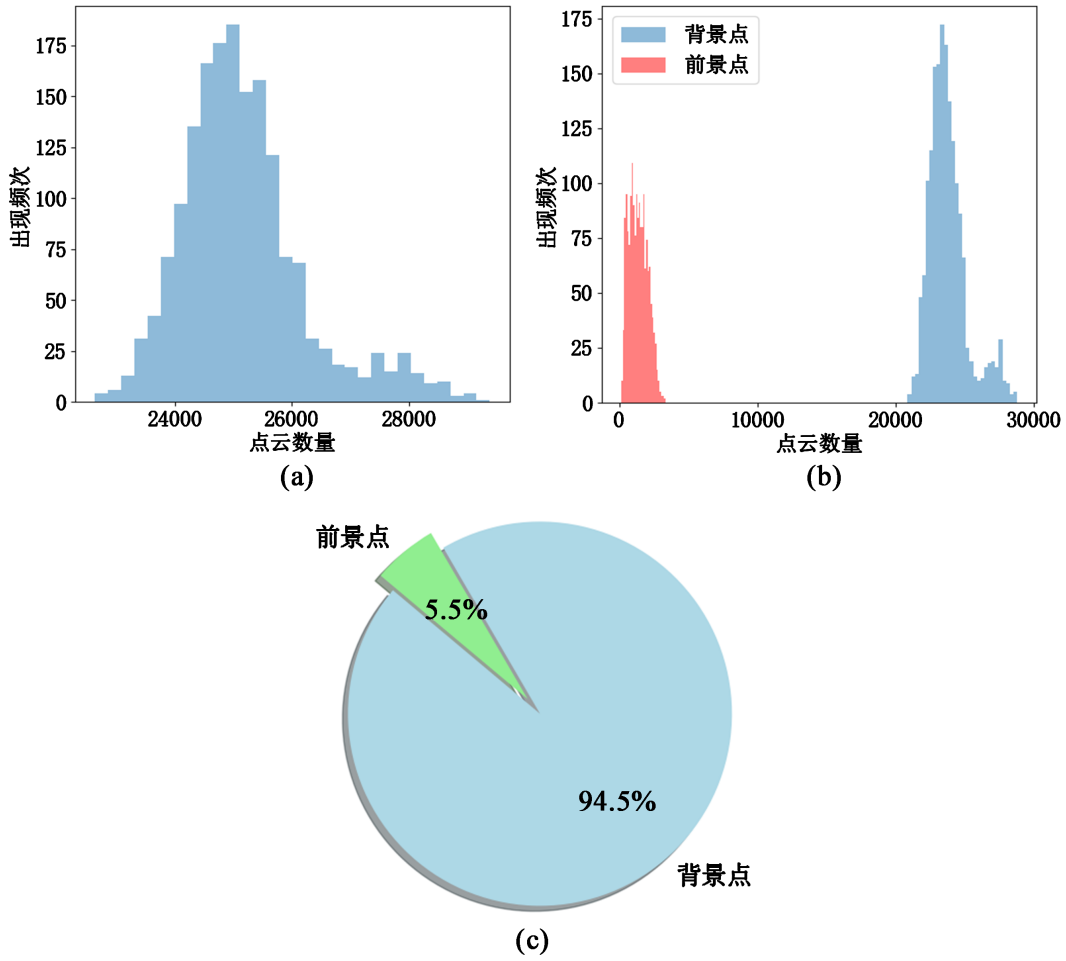

Figure 5. Dataset statistics
图 5. 数据集统计

3.2. 评价指标

本文为了验证该模型用于毫米波点云语义分割任务的性能，采用平均交并比(mIoU, Mean Iinteraction over Union)和分割准确率(Acc , Accuracy)作为模型的评价指标。

对于点云语义分割，IoU 定义为每个类别的预测点云与真实点云交集的大小除以它们的并集大小。

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

其中，TP (True Positive)表示正确预测为某个类别的点数，FP (False Positive)表示错误预测为某个类别的点数，FN (False Negative)表示实际为该类别但未被正确预测的点数。

mIoU 是所有类别上 IoU 的平均值，用于评估模型在所有类别上的整体性能。计算方式如下：

$$\mathrm{mIoU} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{IoU}_i \tag{5}$$

其中，N 是类别的总数，IoU_i 是第 i 个类别的 IoU 值。

准确率是另一种评价模型语义分割性能的直观指标，它计算的是所有被正确分类的点的比例。在点云语义分割中，准确率可以通过以下公式计算：

$$\mathrm{Acc}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{6}$$

其中，TN (True Negative)表示正确预测为非目标类别的点数。

3.3. 模型训练参数设置

本实验在 Ubuntu 18.04 平台上进行，GPU 型号为 RTX 4090，显存为 24,564 MiB，Python 版本为 3.6.13，深度学习框架为 Pytorch，版本为 1.8.0。本文在模型的训练过程中，以 7:3 的比例控制训练集和测试集的规模大小，在各网络梯度更新过程中，选用 Adam 优化器及 one-cycle 更新策略对基准和优化网络从头训练，最大学习率为 0.001，除法因子为 10，权值衰减为 0.01，动量为 0.9，Batchsize 为 4，训练时约占用 23.6 G 显存，共训练 40 个 Epoch。

本实验的模型输入 16,000 个点云，每个点云包含空间位置坐标及其强度，模型均使用了 4 个 SA 层作为特征编码层，分别使用 0.02、0.04、0.08、0.10 作为球查询半径(单位：米)，使用对应的下采样方法逐层将输入点云下采样到 4096、1024、512、256 个点。最后依次使用 4 个 FP 层对最后一个 SA 层的输入进行上采样。

3.4. 实验结果

本文提出的模型与 PointNet++对测试集中的部分毫米波点云图像进行语义分割后的可视化结果如图 6

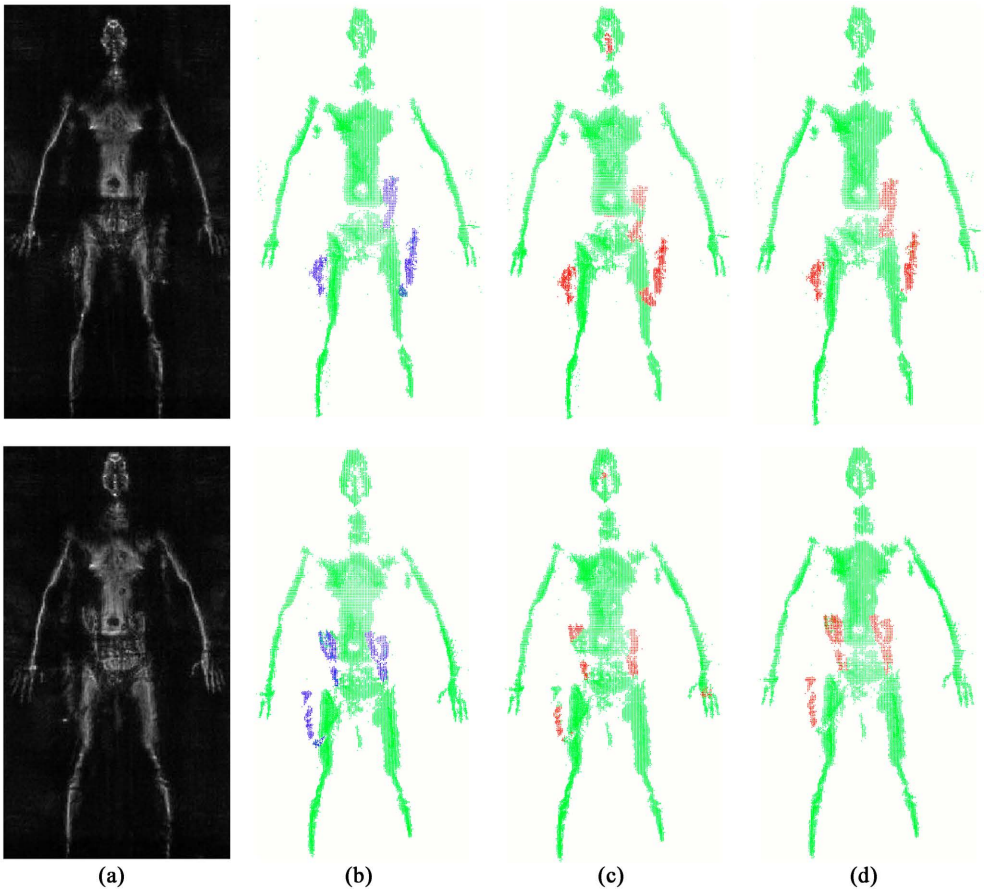

Figure 6. Semantic segmentation visualization results
图 6. 语义分割可视化结果

所示。图 6(a)为毫米波点云图像正投影得到的灰度图，图 6(b)为真实的标注数据，图 6(c)为 PointNet++模型预测结果，图 6(d)为本文所提方法的预测结果。

为了验证基于自注意力机制的实例感知下采样能有效的保留前景点，我们在测试集对比了基于不同降采样方式的模型经过每个 SA 层后的前景点召回率，PointNet++原模型 4 个 SA 层均使用了最远点采样，改进模型中实例感知下采样应用于最后两个 SA 层，前两个 SA 层的采样方式仍为最远点采样，如表 1 所示，本文所提方法进一步提升了前景点的召回率，在编码层的逐步采样过程中，有效地保留了前景点信息。

Table 1. System resulting data of standard experiment
表 1. 与原模型在测试集中各 SA 层前景点召回率对比

采样层	SA_1	SA_2	SA_3	SA_4
采样点数	4096	1024	512	256
PointNet ++	27.46%	26.73%	28.66%	27.97%
PointNet (++) with cls-aware	27.46%	26.73%	87.62%	82.43%
Proposed method	27.46%	26.73%	94.36%	86.43%

为了进一步验证本文提出的模型对毫米波点云的分割性能，实验中本文对比了 PointNet++、基于实例感知下采样的 PointNet++以及结合自注意力机制的实例感知下，采样 PointNet++模型在不同评价指标下的性能结果如表 2 所示。

Table 2. System resulting data of standard experiment
表 2. 在测试集中不同评价指标下的性能结果

模型	mIoU (%)	Acc (%)
PointNet (++)	73.98%	84.66%
PointNet (++) with cls-aware	71.34%	78.98%
Proposed	80.17%	90.23%

对比基于实例感知下采样的 PointNet++与基准模型可以发现，单纯引入实例感知下采样因为会在该任务场景受到类别分布不平衡的影响使得性能相较于基准模型有所下降，然而在实例感知下采样中集成 GCNet 模块后提升了上下文信息的获取，提升了实例感知下采样模块分类性能的同时，其模型分割性能相较于基准模型得到提升，在平均交并比方面提升 6.19%，准确率提升 5.57%。

4. 结论

本文通过在实例感知下采样中集成自注意力模块，优化了传统点云在下采样过程中容易丢失重要前景点信息的问题，从而保留了更多对实现高精度语义分割任务至关重要的前景点信息，此外结合自注意力模块不仅进一步提高了实例感知下采样模块的分类性能，还丰富了模型对全局上下文信息的捕获能力，显著提高了毫米波点云图像语义分割的性能。与基准模型 PointNet++相比，提出的模型在平均交并比和准确率方面分别提升了 6.19%和 5.57%，这证明了在毫米波点云语义分割任务中保留前景点信息、增强模型上下文信息的捕获能力能有效提升模型的性能。

参考文献

[1] Sheen, D.M., McMakin, D.L. and Hall, T.E. (2001) Three-Dimensional Millimeter-Wave Imaging for Concealed

Weapon Detection. *IEEE Transactions on Microwave Theory and Techniques*, **49**, 1581-1592. https://doi.org/10.1109/22.942570

[2] Appleby, R. and Anderton, R.N. (2007) Millimeter-Wave and Submillimeter-Wave Imaging for Security and Surveillance. *Proceedings of the IEEE*, **95**, 1683-1690. https://doi.org/10.1109/JPROC.2007.898832

[3] Wang, Z., Chang, T. and Cui, H.-L. (2019) Review of Active Millimeter Wave Imaging Techniques for Personnel Security Screening. *IEEE Access*, **7**, 148336-148350. https://doi.org/10.1109/ACCESS.2019.2946736

[4] 陈国平, 彭之玲, 黄超意, 等. 基于改进 YOLOv3-Tiny 的毫米波图像目标检测[J]. 电子测量技术, 2021(21): 44.

[5] 张格菲, 李春宇, 刘金坤, 等. 基于 YOLOv5 的毫米波图像目标检测方法研究[J]. 宇航计测技术, 2021, 41(5): 5. https://doi.org/10.12060/j.issn.1000-7202.2021.05.09

[6] 程秋菊, 陈国平, 王璐, 等. 基于卷积神经网络的毫米波图像目标检测[J]. 科学技术与工程, 2020, 20(13): 6. https://doi.org/CNKI:SUN:KXJS.0.2020-13-032

[7] 丁俊华, 袁明辉. 基于双分支多尺度融合网络的毫米波 SAR 图像多目标语义分割方法[J]. 光电工程, 2023, 50(12): 75-86. https://doi.org/10.12086/oee.2023.230242

[8] Qi, C.R., Yi, L., Su, H., *et al*. (2017) Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *Advances in Neural Information Processing Systems*, **30**, 5105-5114.

[9] Qi, C.R., Su, H., Mo, K., *et al*. (2017) Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Venice, 21-26 July 2017, 652-660.

[10] Zhang, Y., Hu, Q., Xu, G., *et al*. (2022) Not All Points Are Equal: Learning Highly Efficient Point-Based Detectors for 3D Lidar Point Clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 18953-18962. https://doi.org/10.1109/CVPR52688.2022.01838

[11] Cao, Y., Xu, J., Lin, S., *et al*. (2020) Global Context Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 6881-6895. https://doi.org/10.1109/TPAMI.2020.3047209

[12] Wang, X., Girshick, R., Gupta, A., *et al*. (2018) Non-Local Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7794-7803. https://doi.org/10.1109/CVPR.2018.00813

[13] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745

[14] Lin, T.Y., Goyal, P., Girshick, R., *et al*. (2017) Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/ICCV.2017.324