引文格式：路创军,魏博,郝娇娇,等. 顾及多高程多密度Transformer的点云地面滤波[J]. 地理空间信息, 2025, 23(10):68-72. doi:10. 3969/j.issn. 1672-4623.2025.10.015

顾及多高程多密度Transformer的点云地面滤波

路创军[1]，魏　博[1]，郝娇娇[1*]，王丽丽[2]，李海燕[1]，郑　超[1]
（1. 甘肃省水利水电勘测设计研究院有限责任公司，甘肃 兰州 730000；2. 西北师范大学物理与电子工程学院，甘肃 兰州 730070）

摘　要：地面滤波是处理ALS点云数据的基础步骤之一，Transformer机制为点云数据处理提供了新的解决方案，但现有方法大多通过堆砌Transformer提高地面滤波精度，未曾建立长距离点云特征依赖关系。为此，提出一种多层Point Transformer模型并将其应用于大范围ALS点云地面滤波，该模型结合了Transformer、反向残差MLP和卷积的多重优势，能够实现点云场景多层位置信息的捕获和轻量化计算。首先，提出融合了多层位置信息的多高程多密度Transformer，该模块分别依据高程和密度对点云场景进行分层，选取高程层和密度层的邻域信息构成多层特征，并分层计算注意力，实现点云高程和密度特征融合。其次，提出一种点云上下文位置编码算法，以准确描述点云上下文依赖关系。最后，在特征维度较小的输入输出层放置Transformer块，以减少内存占用，实现模型轻量化。实验引入大范围ALS点云数据集OpenGF充分验证多层Point Transformer的综合性能，结果表明，多层Point Transformer能够满足大范围ALS点云地面滤波需求，与现有方法相比，取得了更先进的地面滤波精度，并降低了模型复杂度。

关键词：点云地面滤波；Point Transformer；多高程多密度；注意力计算；OpenGF
中图分类号：P284　　**文献标志码：**B　　**文章编号：**1672-4623（2025）10-0068-05

Point Cloud Ground Filtering Based on Multi-elevation Multi-density Transformer

LU Chuangjun[1], WEI Bo[1], HAO Jiaojiao[1], WANG Lili[2], LI Haiyan[1], ZHENG Chao[1]
(1. Gansu Water Resources and Hydropower Survey Design and Research Institute Co., Ltd., Lanzhou 730000, China;
2. College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China)

Abstract: Ground filtering is one of the foundation steps for processing ALS point cloud data. The Transformer mechanism provides a solution for point cloud data processing, but most of the existing methods improve the ground filtering accuracy by stacking Transformers, without ever establishing the long-distance point cloud feature dependency. For this reason, we proposed a multi-layer Point Transformer model and applied it to large range ALS point cloud ground filtering, which combining the multifaceted advantages of Transformer, inverted residual MLP, and convolution, and realizing the capture of multilayered positional information of point cloud scene and lightweight computing. Firstly, we proposed a multi-elevation and multi-density Transformer fusing multi-layer position information, which could layer the point cloud scene based on elevation and density respectively, choose the neighboring information of elevation and density layers to constitute multi-layer features, and calculate the attention in layers to realize the fusion of elevation and density features of point cloud. Secondly, we proposed a point cloud context location coding algorithm to accurately describe the point cloud context dependencies. Finally, we placed a Transformer block in the input-output layer with smaller feature dimensions to reduce the memory occupation and achieve model lightweighting. Experiments used large-scale ALS point cloud dataset OpenGF to fully validate the comprehensive performance of multilayer Point Transformer. The results show that the multilayer Point Transformer can meet the demand of ground filtering for large-scale ALS point clouds, and achieve a more advanced ground filtering accuracy and reduce the model complexity compared to the existing methods.

Key words: point cloud ground filtering, Point Transformer, multi-elevation and multi-density, attention computing, OpenGF

地形测绘是水利工程的基础，其依赖机载激光雷达点云生成DEM[1-5]。目前，地面滤波方法从传统机器学习发展到深度学习，但现有方法缺乏长距离特征依赖，影响复杂地形分析[6-10]。Transformer通过自注意力机制获取远程信息，是生成长距离特征依赖的有效手段，该机制能够同时整合点云局部和全局信息，从而在复杂地形和大范围环境建模中提供更准确和全面的分析[11]。因此，一些基于Transformer的点云分割方法相继提出，Point Transformer[12]将“向量自注意力机制”和“减法关系”融合，实现点云全局特征聚合；Voxel Transformer[13]以体素为单位执行自我关注，但丢失了点的位置信息；Fast Point Transformer[14]是Point

项目来源：甘肃省水利科学试验研究及技术推广项目（GSKY2021-9）。
第一作者简介：路创军（1968—），本科，高级工程师，从事点云数据智能处理工作，E-mail：LU_6805_Chuangjun@163.com。
通信作者：郝娇娇（1997—），硕士，从事深度学习算法及点云数据智能处理工作，E-mail：Janice1_hao@163.com。

Transformer的改进版本，训练速度更快，模型复杂度更小。Stratified Transformer[15]则利用一种分层的采样策略，增强了点云长距离依赖关系，但只局限于密度层面。上述基于Transformer的点云分割方法都是在每次采样结束后加入Transformer实现特征融合，平均每个模型需要8~9个Transformer模块，导致模型计算复杂度较高。因此，如何减少基于Transformer模型的计算复杂度以及进一步建立点云长距离依赖关系，是需要解决的问题。

针对以上问题，提出一种适用于大范围ALS点云地面滤波的多层Point Transformer模型。首先，对于点云上下文利用不充分问题，提出一种基于点云分块的多高程多密度Transformer，通过密度分层捕获不同距离的点云邻域信息，实现目标点上下文由局部到多距离的延伸；在此基础上，提出一种点云上下文位置编码算法来描述点云空间依赖关系，将获取到的多密度点云邻域信息添加到自注意力学习中，实现多高程多密度的点云上下文信息建模，进而提高点云地面标签预测的准确率；最后，为解决模型反复堆叠Transformer模块造成的内存消耗过大问题，提出一种轻量级Transformer机制，将多高程多密度Transformer放置于维度较小的输入输出特征层，减少模型的计算复杂度和内存消耗，并引入编码层和反向残差MLP（inverted residual MLP）对特征空间进行拉伸，避免模型结构单一引起特征学习不充分。

1　研究方法

1.1　多层Point Transformer模型概述

多层Point Transformer模型框架如图1所示，输入信息包括 *XYZ* 坐标和相对位置坐标信息，整体框架采用了编码–解码器结构。

1.2　多高程多密度Transformer

多高程多密度Transformer模块由多高程多密度采样策略、自注意力机制[2]、线性层和MLP组成，若直接使用整幅点云数据计算自注意力，数万个点会产生高达 $O(N^2)$ 的内存消耗，其中 N 是输入点数。因此，为避免内存消耗过高，首先对场景进行分块后再计算自注意力；接着，为建立多层位置依赖，提出多高程多密度采样策略，通过高程分层和密度分层的方式获取点云上下文信息[15]。

1.2.1　点云分块

点云分块是将三维空间划分为多个立方体，查询点只考虑同一立方体内的邻域点，而不是像全局自注意力那样关注场景中所有点。立方体构建方法分为规则不重叠分块和随机重叠分块方法，规则不重叠分块是针对整体场景点云 $P=\{p_n|n=1,2,\ldots,N\}$，将其划分 m 个尺寸为 $h\times w\times d$ 的固定立方体。随机重叠分块则是根据场景中的点云数 N，估算该场景立方体个数 m，接着随机选取 m 个中心点，寻找每一个中心点的 $h\times w\times d$ 范围，作为当前中心点所在分块。2种点云分块方法如图2所示。

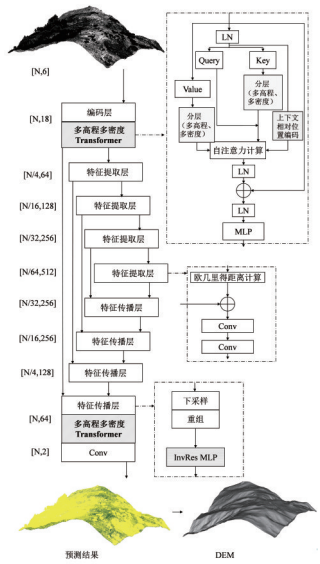
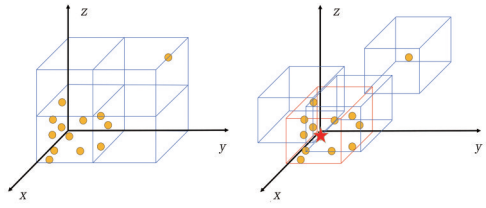


图1　多层Point Transformer模型

注：InvResMLP是反向残差MLP，LN（Layer Normalization）是归一化层，MLP是多层感知机，Conv是卷积层。



a 规则不重叠分块　　b 随机重叠分块

图2　点云分块示意图

1.2.2　多高程多密度采样策略

由于只关注查询点的局部信息会导致有效接收域缩小，无法捕获到远距离上下文特征，从而拉低分割精度。虽然扩大立方体尺寸有利于对远距离上下文特征捕捉，但内存消耗也会随着立方体尺寸的增大而增大，为在节约内存成本的前提下有效捕捉远距离上下文特征，提出多高程多密度采样策略。

1）多高程采样。针对任意立方体，首先通过高程值检索出最大最小高程 $Z_{\max}$ 和 $Z_{\min}$。接着，设置高程参数 φ（单位：m）将场景分为 $M_{\text{alti}}=(Z_{\max}-Z_{\min})/\varphi$ 个高程层。最后，将同一高程层中的点作为检查点 p_i

的邻近点集 S_i，并将该点集特征赋予 p_i。为避免高程信息缺失，高程参数 φ 不宜设置过小，考虑到大规模室外点云存在显著高程差异，本文以立方体尺寸为依据，将立方体内点云分为5层，每层高程参数 φ 设置为10 m。

2）多密度采样。多高程采样只关注同一高程层点的上下文关系，未扩散到更远的高程关系，为弥补该问题，提出一种多密度采样策略与其相结合。多密度采样将每个点云立方体分为 M_{des} 个密度级，实现不同密度级下的注意力计算。首先以点云立方体为单位，将原始点云设为第一密度层，点数为 n_1。接着，利用最远点采样方法下采样点云，获取稀释的密度层，其中第 l 个密度层的点数 $n_l = n_1/2l$，$l \in (2, M_{\text{des}})$。最后，在不同的密度层下，检索当前距离当前点 p_i 最近的 k 个点，形成邻近点集 S_i。

为适应不同密度不同场景的数据，通过改变分级数 M_{des} 设置了3个采样级别，分别为 M_{des} =2、M_{des} =3和 M_{des} =4。M_{des} 越大，则可以获得更长距离的点云依赖，从而构建更稳健的上下文交互关系。

1.2.3 自注意力计算

多高程多密度采样完成后，在每个立方体中独立执行自注意力机制。与传统自注意力机制不同，本文提出点云上下文位置编码，对输入点集特征及其邻域特征重新编码，增加了位置信息的利用率。假设 n^t 表示为第 t 个立方体内点的数量，那么在第 l 层（高程层或任意密度层）的上下文相对位置编码（Contextual Relative Position Encoding）$\text{cRPE}_l^{n_l^t \times k \times c^e}$ 计算公式为：

$$\text{cRPE}_l^{n_l^t \times k \times c^e} = \text{pos}(\sigma(f_{\text{text}}^{n_l^t \times k \times c}) - \varphi(f_{\text{in}}^{n_l^t \times c})) \tag{1}$$

式中，$f_{\text{in}}^{n_l^t \times c}$ 为第 l 个密度层点集输入的特征；c 为输入特征通道数；c^e 为上下文相对位置编码输出的特征通道数；$f_{\text{text}}^{n_l^t \times k \times c} = f_{\text{in}}^{n_l^t \times c} \bullet \text{idx}_l^{n_l^t \times k}$ 为第 l 个密度层点集对应的上下文特征矩阵；k 为邻接的上下文点数；$\text{idx}_l^{n_l^t \times k}$ 为第 l 个密度层点集上下文的位置索引；$\bullet$ 为特征映射；σ 和 φ 为特征变换，这里特指线性层；pos为位置编码函数，它由线性层和RELU函数构成。

自注意力由输入特征 $f_{\text{in}}^{n_l^t \times c}$ 和 $\text{cRPE}_l^{n_l^t \times k \times c^e}$ 计算得到。首先，将 $f_{\text{in}}^{n_l^t \times c}$ 线性变换得到query、key和value；然后将query和key的上下文特征相减，得到相对位置差值特征信息后，添加上下文相对位置编码，并利用soft max()[16]函数归一化，得到第 l 个密度层的自注意力特征矩阵，将value的上下文特征与上下文相对位置编码相加，与注意力矩阵完成加权聚合运算得到所有密度层的聚合特征；最后把聚合特征和输入特征进行残差链接，得到输出的残差特征矩阵，完成自注意力计算。

1.3 反向残差MLP

PointNet++的分割架构仅由4个特征提取块组成，每个特征提取块中使用3层MLP完成特征拟合。PointNeXt发现，添加更多的特征提取块或者使用更多的MLP层都不会使得准确度提高，反而会导致吞吐量[17]显著下降，造成特征梯度消失和模型过拟合，而反向残差MLP可以通过残差连接和反向瓶颈设计（inverted bottleneck design）显著提高分割性能。因此，多层Point Transformer引入反向残差MLP实现点云特征拟合。反向残差MLP的示意图如图3所示。

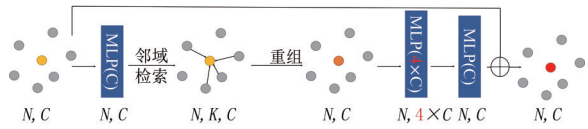


图3 反向残差MLP（点的颜色越深代表上下文信息越丰富）

注：N 是输入点数，C 是特征通道数，K 是邻域检索点数。

2 实验结果与分析

2.1 实验设置

2.1.1 OpenGF数据集

OpenGF发布于2021年4月15日[18]，是一个大范围地面滤波数据集，训练数据建立在全球4个不同国家的开放ALS点云上，覆盖面积超过47.7 km²，拥有5.42亿个精细标记的地面和非地面点。此外，OpenGF还提供了2个测试数据（Test1，Test2），测试场景中包含多种土地覆盖类型，这对于点云地面滤波算法来说具有一定挑战性。Test1覆盖面积约6.6 km²，包括村庄、城市和山区，点密度为6.97 m²。Test2是一个面积约1.1 km²的都市区，包含多种大型建筑物和少量植被，点密度为5.65 m²。需要注意的是，Test2原始点云中包含大量异常点，实验过程已去除异常点。

2.1.2 参数设置

实验设置初始学习率为0.7，权重衰减为 10^{-4}，batch size为20，使用Adam优化器来训练多层Point Transformer模型，立方体点数为8 192；特征提取层使用最远点采样方法进行下采样，下采样点数矩阵表示为[20 48,512,128,32]，训练epoch设置为30，Transformer深度为1。另外，多层Point Transformer模型缩

写为MLPoint Transformer（Multi-Layer Point Transformer），根据 M_{des} 由小到大分别表示为 S 、 M 、 L 。

2.2　点云地面滤波结果分析

为了验证多层 Point Transformer 点云地面滤波性能，采用可视化分析和定量分析来比较不同方法的点云地面滤波结果，从而验证方法的优越性和不足。

2.2.1　可视化分析

图 4 展示了 OpenGF 数据集 Test1 和 Test2 点云地面滤波的可视化结果，对比方法包括 PointNet++、Point Transformer 和 MLPoint Transformer。通过对比可知，MLPoint Transformer 能够更准确的预测地面点云，且 DEM 更真实，相比其他方法，MLPoint Transformer 误判范围明显变小，误差点较少（如黄色方框中所示），这是因为多高程多密度 Transformer 中的采样策略能够获取更大接收域上的上下文信息，而其他方法则局限于小部分范围。此外，MLPoint Transformer 对非地点的预测更准确（如绿色方框中所示），大型建筑物完整度更高，这说明随机分块算法不会导致组间独立，相反，由于随机分块学习到了更多上下文依赖关系，能够精准表达点云特征，从而降低误判几率。

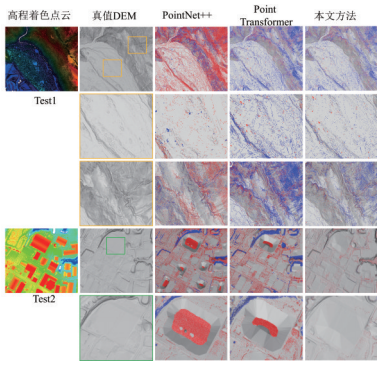


图 4　OpenGF 点云地面滤波可视化结果对比

注：红色区域为多余检测的地面点映射，蓝色区域为未检测到的地面点映射。

2.2.2　定量分析

定量对比选择了 6 种现有方法，包括 PointNet++、DGCNN、RandLA-Net[19]、SeqGP、Point Transformer[12]和 Stratified Transformer[15]，精度评价指标包括 IoU（Intersection over Union）、*OA*（Overall Accuracy）和 RMSE（Root Mean Square Error）[18]，IoU 是正确预测点数与所在类别交并点数（真值点数和多余预测点数之和）的比值，其中 IoU_{NG} 和 IoU_{G} 分别是非地面点和地面点的 IoU，*OA* 统计了数据中正确分类点所占比例，RMSE 则是评价预测 DEM 与真值 DEM 高程之间均方根误差，值越小代表预测误差越小。对大范围 ALS 点云场景来说，多高程多密度分层查询点信息能够更精确地描述点云的上下文关系。MLPoint Transformer-L 的各项精度评价指标优于 S 和 M 型模型，这是因为随着密度层增多，模型学习到更远距离点云特征，进而检索到更多点云上下文依赖关系，特征学习也更充分。在 Test2 数据上，对比其他方法，MLPoint Transformer 的 *OA* 值明显提高，最高可达 96.88%，这证明 MLPoint Transformer 对含有大型地物场景的地面点预测能力较好，且地形细节保留更完整。

2.3　消融实验

通过消融实验来验证 MLPoint Transformer 中每个模块的有效性，分别对模型中的随机重叠分块、多高程多密度采样、反向残差 MLP 和数据增强模块进行消融分析，添加各模块后 Test1 的 IoU_{G} 如表 1 所示，Test2 的精度变化与 Test1 类似。其中 ID 为 1~5 时，模型依次添加一个组件，ID 为 6~8 时，则从最终模型中删除一个模块来进行双重验证。需要说明的是，消融实验中的基础框架是 PointNet++（使用规则的不重叠点云分块方法）。

1）随机重叠划分。在表 2 中，比较 ID4 和 ID5 可以看到，采用随机重叠划分使 IoU_{G} 精度提高了 8.26%。而单独摘除随机重叠划分后（ID6），精度下降 8.08%，可见随机重叠点云划分能够极大提高点云地面滤波，比起传统的规则分块，随机方式能够捕捉到更多变的点云分布特征。

2）多高程多密度采样。在表 1 中，比较 ID3 和 ID4，可以注意到添加多密度注意力模块使得 IoU_{G} 精度提高了 6.78%。而单独摘出多高程多密度采样模块后（ID8），精度下降 15.25%，这说明该模块极大提高了点云地面滤波精度。多高程多密度采样增加了点云上下文依赖关系，且由于密度分层，使得无论点云近距离关系还是远距离关系，都能准确表达。

表 1　消融实验（每列最大值以粗体显示）

ID	随机重叠分块	多高程多密度采样	反向残差 MLP	数据增强	IoU_{G}/%
1					68.86
2				√	76.30
3			√	√	83.73
4		√	√	√	90.51
5	√	√	√	√	**98.77**
6		√	√	√	90.69
7	√	√		√	93.43
8	√		√	√	83.52

3）反向残差 MLP。在表 1 中，比较 ID2 和 ID3 可以看到，反向残差 MLP 模块使得 IoU_{G} 精度提高了 7.43%。而摘除反向残差 MLP 模块后（ID7），精度下降 5.34%，说明该模块分割性能优于 MLP，能够更准确的拟合点云特征。

4）数据增强。数据增强[20]在训练基于 Transformer

的网络时起着重要作用，在MLPoint Transformer中也是如此，表1中加入数据增强模块后，IoU_C提高了7.44%。为进一步评估数据增强的效果，单独加入各项数据增强方法，比较IoU_C的增量。数据增强方式分别为旋转、降低颜色、抖动和缩放，实验结果证明，数据增强可以有效提高点云地面滤波精度，并且缩放的数据增强方式效果最好，结果如表2所示。

表2 数据增强方法评估

数据增强类型	–	抖动	旋转	颜色降维	缩放	全部
IoU_C/%	89.50	90.43	91.90	90.41	94.12	98.77
Δ		+0.93	+2.40	+0.91	+4.62	+9.27

2.4 模型复杂度对比

为评估轻量级多密度Point Transformer模型的计算成本，本节对比了不同模型的参数量（Parameters）和计算复杂度（floating point operations per second，FLOPs）[4]。参数量是指模型中需要学习的参数数量，用来评估模型的大小。计算复杂度是指模型进行一次前向传播时所需要的浮点数操作数量，通常评估模型的计算量。

表3 模型复杂度对比

模型名称	参数量（*M*）	计算复杂度（*G*）
Point Transformer	4.53	5.57
Fast Point Transformer	3.79	4.52
Stratified Transformer	1.93	2.87
MLPoint Transformer–S	1.32	2.21
MLPoint Transformer–M	1.56	2.47
MLPoint Transformer–L	1.89	2.81

表3展示了不同模型的复杂度评测指标（参数量和计算复杂度），其中对比模型均是基于Transformer的，分别是Point Transformer、Fast Point Transformer[4]和Stratified Transformer[5]。从表格中可以看出，相对于其他Transformer模型，多层Point Transformer模型整体具有更少的参数量和更低的计算复杂度，MLPoint Transformer–S的2项指标远低于其他模型，这表明该模型在处理点云数据时具有较高的效率和实用性，在资源受限的情况下（例如移动设备和嵌入式系统）也能得到应用。受密度层数影响，MLPoint Transformer–L的计算复杂度略高于MLPoint Transformer–S和MLPoint Transformer–M，但地面预测结果精度也更高，因此这部分计算量是必要存在的。

3 结 语

本文提出一种多层Point Transformer方法并将其应用于大范围ALS点云地面滤波。该方法针对点云上下文信息利用不充分问题，提出多高程多密度Transformer获取多层点云上下文特征，通过高程分层和密度分层，选取上下文关键点的方式扩大感受野，建立了更加完善的特征采样体系，提高了点云地面滤波性能；为应对点云的不规则性，提出一种多密度上下文位置编码方法，该方法在多个密度层上检索目标点的邻域点，并使用线性层赋予邻域点到目标点相对距离的可学习性，进一步提高了Transformer在点云地面滤波上的适用性；针对基于Transformer模型计算复杂度高的问题，设计了一个轻量级Transformer机制，该机制仅在下采样前和上采样流程结束后放置Transformer，其余部分使用卷积和反向残差MLP，在减少计算量的同时也获取了更准确的全局上下文信息。

可视化和定量化结果表明，多层Point Transformer能够有效处理点云地面滤波，并且取得了更高的地面滤波精度，消融实验证明了各模块的有效性，模型复杂度对比分析则表明以多高程多密度Transformer为主体的多层Point Transformer模型计算复杂度更低，参数量更少。

参考文献

[1] 陈斐然,李浩,李语旻,等. 基于改进坡度的自适应数学形态学点云滤波[J]. 地理空间信息,2017,15(9):28-31

[2] Zhang K, Chen S C, Whitman D, et al. A Progressive Morphological Filter for Removing Nonground Measurements from Airborne LiDAR Data[J]. IEEE Transactions on Geoscience and Remote Sensing,2003,41(4):872-882

[3] Błaszczak-bąk W, Janowski A, Kamiński W, et al. Application of the Msplit Method for Filtering Airborne Laser Scanning Data-sets to Estimate Digital Terrain Models[J]. International Journal of Remote Sensing, 2015, 36(9): 2 421-2 437

[4] Zhang W, Qi J, Wan P, et al. An Easy-to-use Airborne LiDAR Data Filtering Method Based on Cloth Simulation[J]. Remote Sensing,2016,8(6):501

[5] Wallach H M. Conditional Random Fields: An Introduction[J]. University of Pennsylvania CIS Technical Report MS-CIS-04-21,2004,24(10):33-42

[6] Qi C R, Su H, Mo K, et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:652-660

[7] Wu W, Qi Z, Fuxin L. PointConv: Deep Convolutional Networks on 3D Point Clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019

[8] Yan X, Zheng C, Li Z, et al. Point Asnl: Robust Point

（下转第83页）