

文献分享： 点云的语义分割

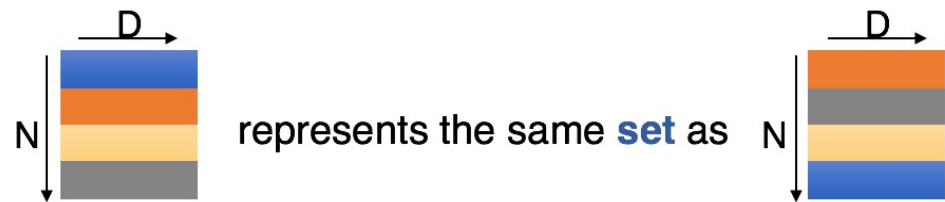
刘宗健
7.9

目录

- ❑ Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *CVPR*, 2017
- ❑ Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *NeurIPS*, 2017
- ❑ Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, Andrew Markham. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. *CVPR*, 2020

点云数据的特点

□ 无序性（置换不变性）



□ 交互性

□ 刚体变换不变性



1. Pointnet

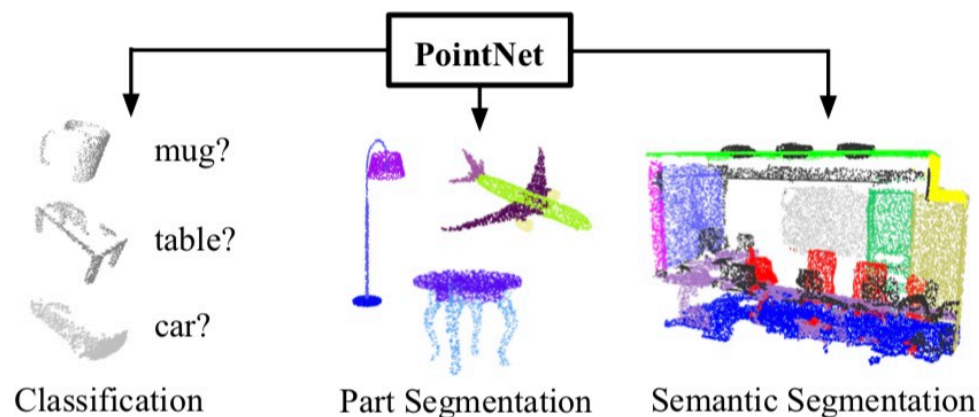
□ 主要贡献

设计了一个适用于**无序点**集合的深度网络框架；

展示了如何训练这样一个网络来实现3D形状分类、形状部分分割和场景语义分析任务；

理论和实践分别分析了该方法的稳定性和有效性；

显示了由网络中所选神经元计算出的3D特征，并为其性能提供了直观的解释。



1. Pointnet

□ 问题描述

输入为无序点集，集合 $\{P_i \mid i = 1, \dots, n\}$

P_i 是一个由坐标 (x, y, z) 加上额外的特征通道组成的向量

目标分类输出 k 个分类的 k 个分值

语义分割输出 $n*m$ 分数， n 个点云， m 个分类

1. Pointnet

□ 背景和思路

点云数据结构自身特性

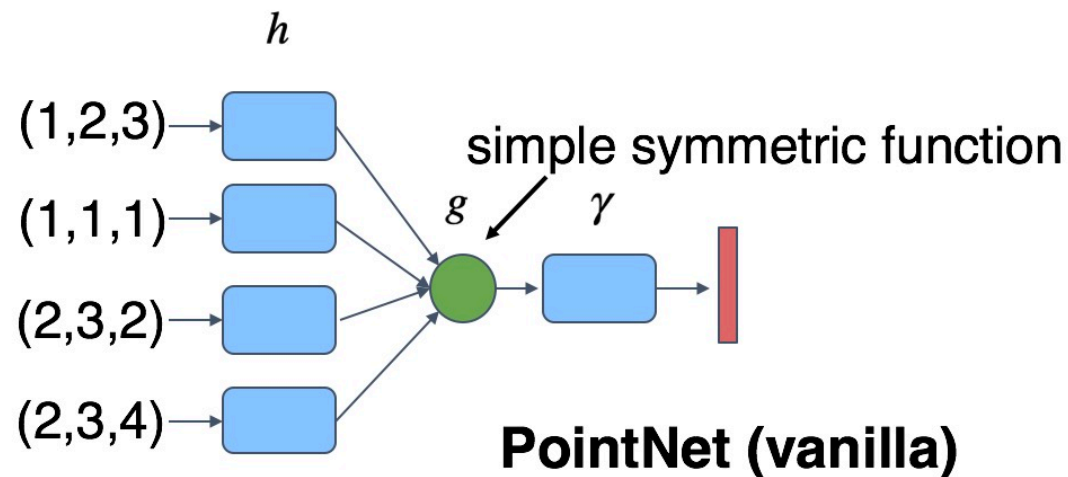
点云的处理手段（创新点）

加入变换来规范化数据

$$L_{reg} = \|I - AA^T\|_F^2,$$

添加了正则化

$$f(x_1, x_2, \dots, x_n) = \gamma \circ g(h(x_1), \dots, h(x_n))$$

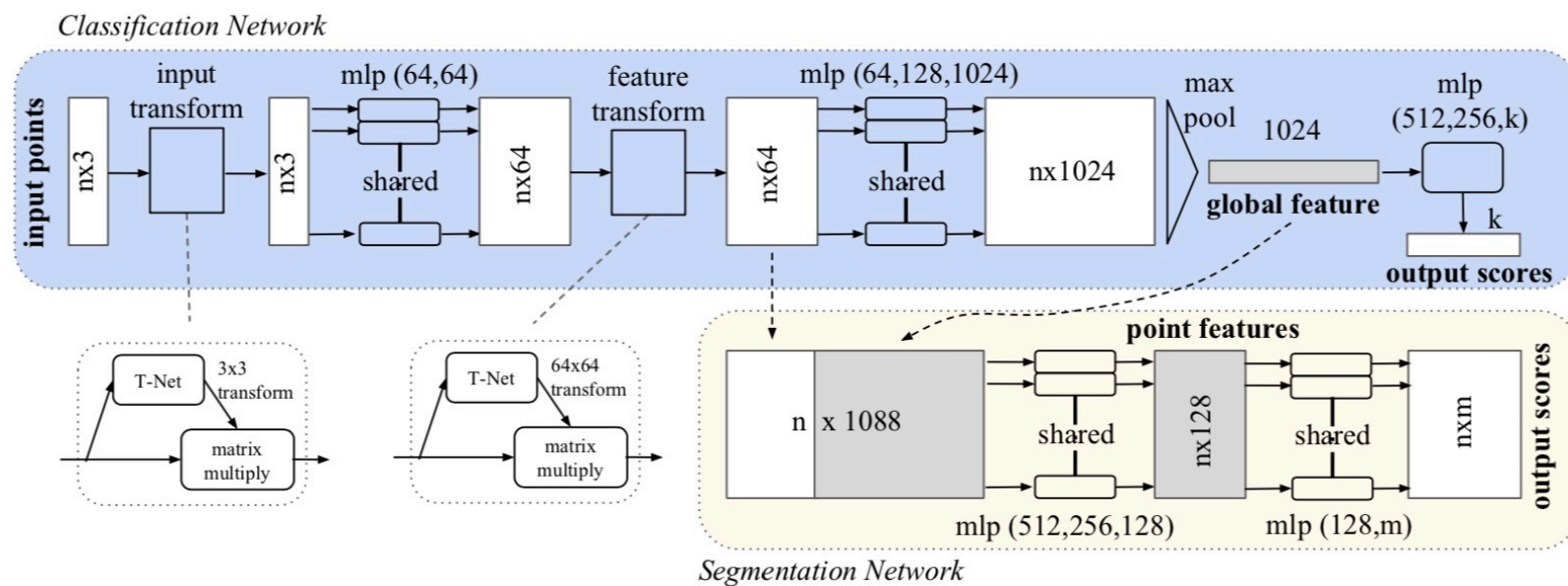


1. Pointnet

□ 网络结构

三个关键模块：

最大池化层聚集所有点信息的对称函数；一个局部和全局信息组合结构；两个对齐输入点和点特征的联合对齐网络



1. Pointnet

□ 实验（场景语义分割）

6个区域，271个房间。每一个扫描到的点都标注了13个类别标签（椅子、桌子、地板、墙等）中的一个

准备训练集，按空间分割点，将空间样本分割为 $1\text{m} \times 1\text{m}$ ；XYZ、RGB及归一化后的位置组成的9维向量

每个模块中随机采样4096个点进行训练

预测每个块中每个点的类别

	mean IoU	overall accuracy
Ours baseline	20.12	53.19
Ours PointNet	47.71	78.62



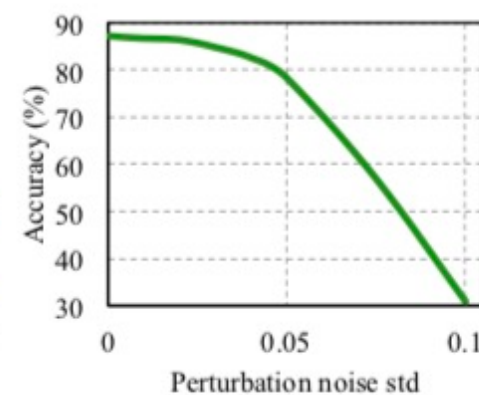
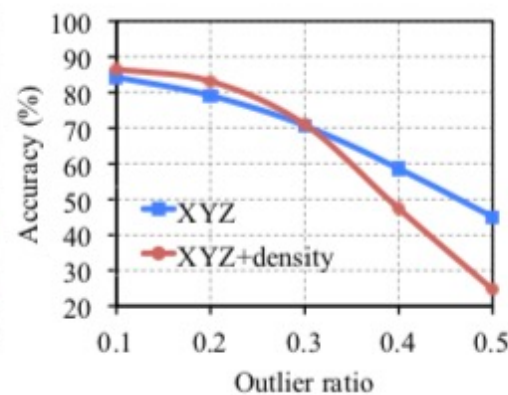
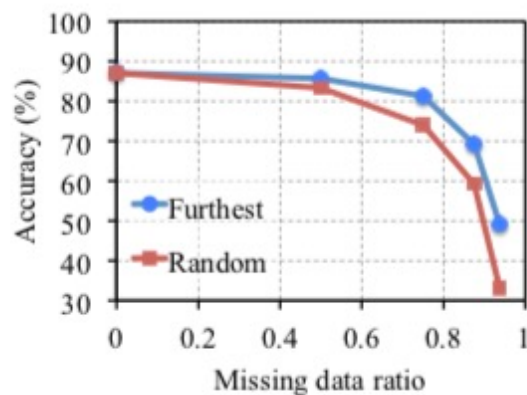
1. Pointnet

□ 对于无序数据的处理对比；对于对称函数选择对比；对于变换矩阵的有效性证明

	accuracy
MLP (unsorted input)	24.2
MLP (sorted input)	45.0
LSTM	78.5
Attention sum	83.0
Average pooling	83.8
Max pooling	87.1

Transform	accuracy
none	87.1
input (3x3)	87.9
feature (64x64)	86.9
feature (64x64) + reg.	87.4
both	89.2

□ 鲁棒性测试

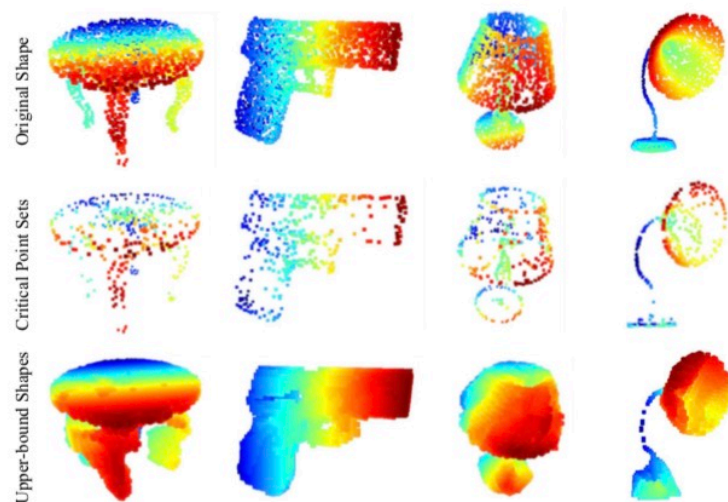


1. Pointnet

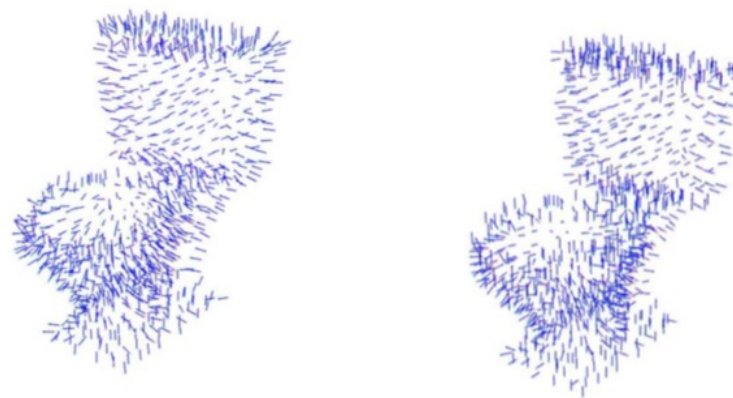
□ C_s, N_s 可视化（分类鲁棒实验）

C_s 最大合并特征，形状的骨架

N_s 给出与输入点云 S 相同的全局形状特征 $f(S)$ 的最有可能的点云



用余弦距离的绝对值来表示损失，分割网络可以训练预测点法线，证明学习到部分点的邻域决定的局部几何性质。



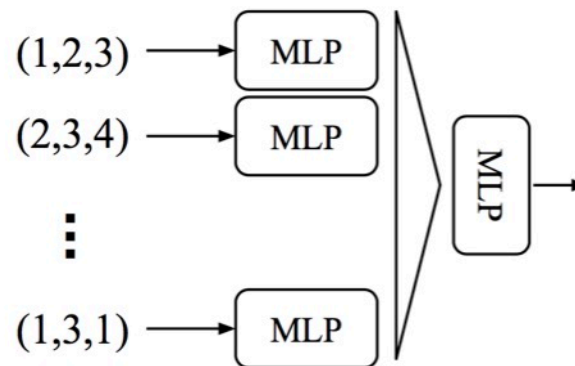
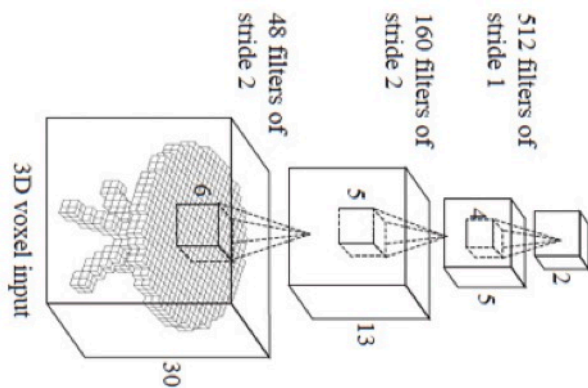
2. Pointnet++

□主要的改进方向

1. PointNet不能捕获局部结构
(复杂场景的泛化能力不足)

需要通过增加上下文来学习局部特征

2. 点集采样不均匀



2. Pointnet++

□思路:

分层处理

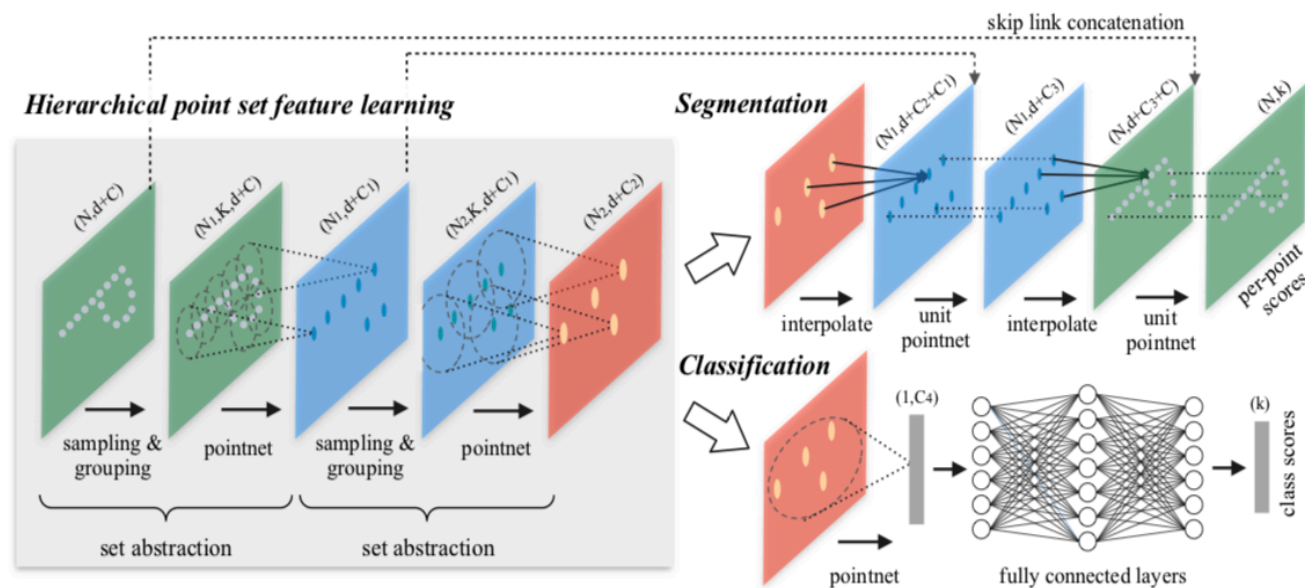
距离（邻域球）划分局域

FPS选点、均匀覆盖

局域特征提取用共享权重Pointnet

2. Pointnet++

□ 网络结构



采样层

分组层

(输入 $N \times (d + C)$ 的点集和大小为 $N' \times d$ 的一组质心的坐标；输出是大小为 $N' \times K \times (d + C)$ 的点集的组)

Pointnet层

(输入 $N' \times K \times (d + C)$ 的点 N' 个局部区域；相对质心编码学习，输出为 $N' \times (d + C')$)

2. Pointnet++

□ 非均匀密度下的特征学习

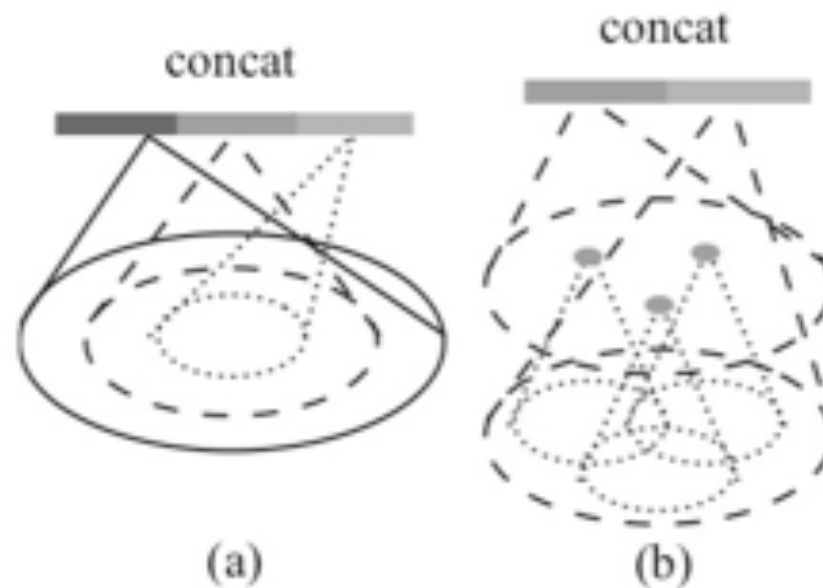
$\text{SA}(K, r, [l_1, \dots, l_d])$

多尺度分组 (MSG)

不同尺度的特征被连接起来以形成多尺度特征。

多分辨率分组 (MRG)

特征是两个矢量的连接。



2. Pointnet++

□ 点特征传播

从子采样点传播到原始点

方法：基于距离的插值

N_{l-1} 个点坐标处插 N_l 个点的特征值 f ;

N_{l-1} 个点内插特征与抽象层的跳连点特征连接;

连接好的特征通过 Pointnet 单元传递

$$f^{(j)}(x) = \frac{\sum_{i=1}^k w_i(x) f_i^{(j)}}{\sum_{i=1}^k w_i(x)} \quad \text{where} \quad w_i(x) = \frac{1}{d(x, x_i)^p}, \quad j = 1, \dots, C$$

2. Pointnet++

□ 实验

变化采样密度的鲁棒性

证明密度自适应层设计的有效性

MSG网络仅受到轻微影响并且

比较方法中实现最佳准确度

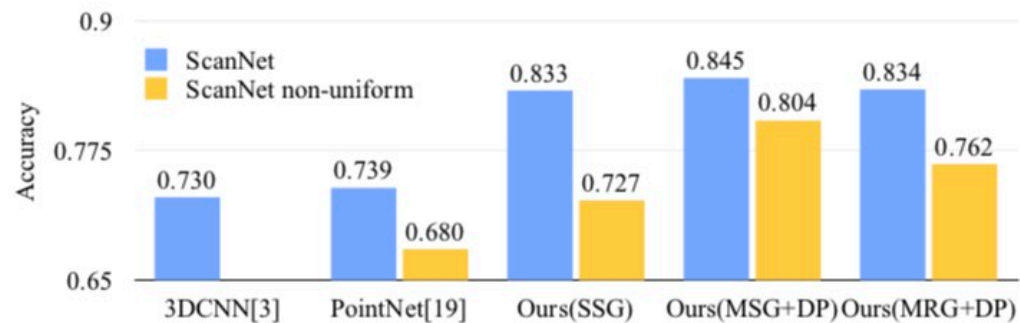
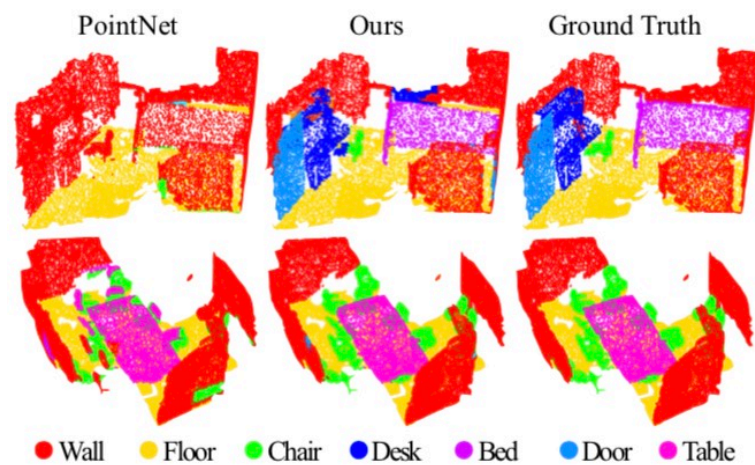


Figure 5: Scannet labeling accuracy.



2. Pointnet++

□ 邻域查询对比

kNN (k=16)	kNN (k=64)	radius (r=0.1)	radius (r=0.2)
89.3	90.3	89.1	90.7

□ 采样层FPS稳定度

Feature difference std.	Accuracy std.
0.021	0.0017

□ 时间空间复杂度

	PointNet (vanilla) 1	PointNet 1	Ours (SSG)	Ours (MSG)	Ours (MRG)
Model size (MB)	9.4	40	8.7	12	24
Forward time (ms)	11.6	25.3	82.4	163.2	87.0

□ Pointnet++ 主要缺陷

循环分组、切块预处理不适用

降采样时间代价高

点云密度不均匀

预处理切块	mIoU(%)	Time(s)
1m×1m	13.1	1322
2m×2m	15.9	2733
3m×3m	18.1	8977

点数			
时间 (s)			
方法	10^4	10^5	10^6
FPS	0.03	4.2	384
RS	0.001	0.001	0.013
IDIS	0.08	0.9	21

3. RandLA-Net

- 主要针对大场景点云语义分割，对方法进行剪枝，目标是高效轻量的神经结构。
- 舍弃繁琐预处理等步骤，采样方法精简为随机点采样，相应的为了弥补舍弃特征的缺陷，加入局部特征聚合模块，增加感受野。
- 可以单次处理100万个点，速度比现有方法快200倍。

3. RandLA-Net

□ 之前限制：

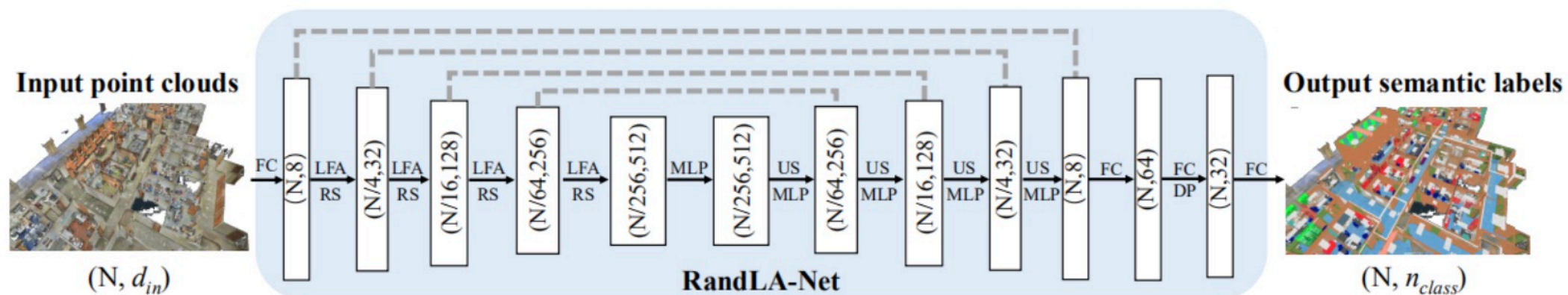
感受野不足，无法捕获每个点广泛的上下文信息

点采样方法效率低下

现有的特征学习通常依靠计算昂贵的核化或图构造

3. RandLA-Net

□ 基本架构



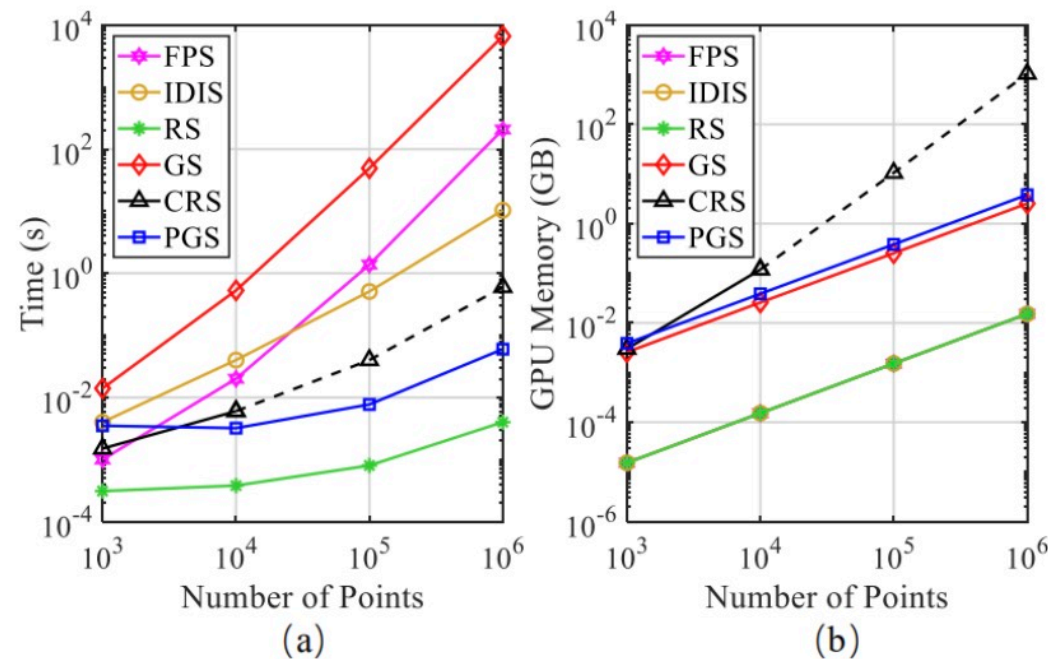
3. RandLA-Net

□ 采样

给定不同规模的点云，连续进行下采样，仅保留25%的点

小规模点云 ($\sim 10^3$)，所有采样方法有相似的时间和内存消耗

大规模点云 ($\sim 10^6$)，FPS/IDIS/GS/CRS/PGS要么费时要么费内存



3. RandLA-Net

□ 局部特征聚合模块

并行地应用于每个3D点，由三个神经元组成：

- 1) 局部空间编码 (LocSE)
- 2) attentive pooling
- 3) dilated residual block (扩张的残差块)

3. RandLA-Net

□ Local Spatial Encoding

明确嵌入相邻点的x-y-z坐标

KNN算法收集其相邻点→编码相对位置及距离等→相对位置 and 对应点特征串联成增强特征向量 $\hat{F}_i = \{\hat{f}_i^1 \dots \hat{f}_i^k \dots \hat{f}_i^K\}$

$$\mathbf{r}_i^k = MLP(p_i \oplus p_i^k \oplus (p_i - p_i^k) \oplus \|p_i - p_i^k\|)$$

相对位置起重要作用，了解局部几何形状

LocSE	mIoU(%)
(1) (p_i)	40.7
(2) (p_i^k)	41.1
(3) (p_i, p_i^k)	42.5
(4) $(p_i, p_i^k, \ p_i - p_i^k\)$	44.1
(5) $(p_i, p_i^k, p_i - p_i^k)$	48.8
(6) $(p_i, p_i^k, p_i - p_i^k, \ p_i - p_i^k\)$ (The Full Unit)	52.0

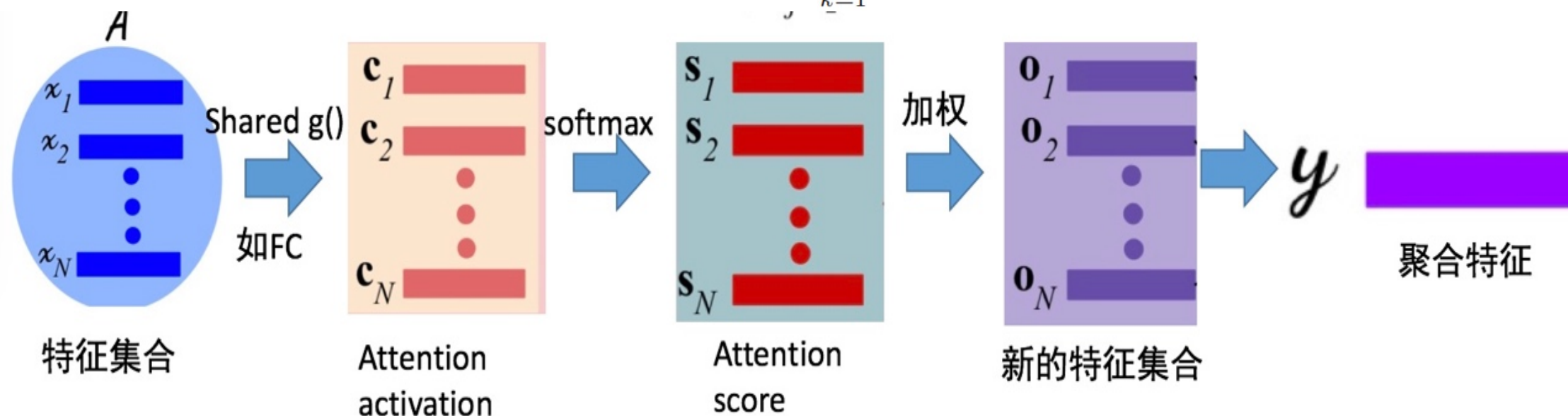
3. RandLA-Net

Attentive Pooling

汇总的临近点特征集合用注意力机制自动学习局域特征产生聚合特征

$$\mathbf{s}_i^k = g(\hat{\mathbf{f}}_i^k, \mathbf{W})$$

$$\tilde{\mathbf{f}}_i = \sum_{k=1}^K (\hat{\mathbf{f}}_i^k \cdot \mathbf{s}_i^k)$$



3. RandLA-Net

□ 可视化注意力分数

注意力权重在第一个编码层有较大的值，而后平滑稳定；

开始倾向于选择突出的关键特征，而后大幅下采样倾向于保留大多数点特征

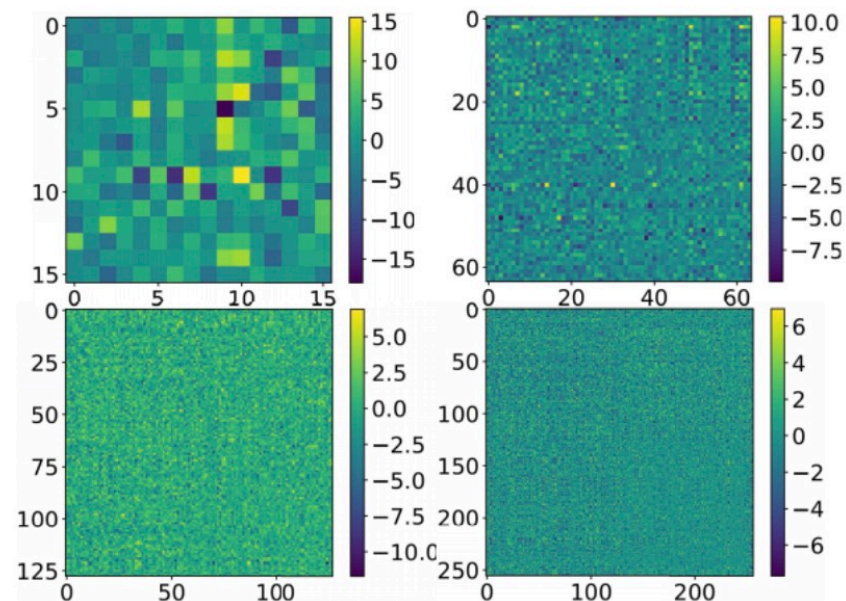


Figure 8. Visualization of the learned attention matrix in different layers. From top left to bottom right: 16×16 attention matrix, 64×64 attention matrix, 128×128 attention matrix, 256×256 attention matrix. The yellow color represents higher attention scores.

<https://blog.csdn.net/Dujing2019>

3. RandLA-Net

□ Dilated Residual Block

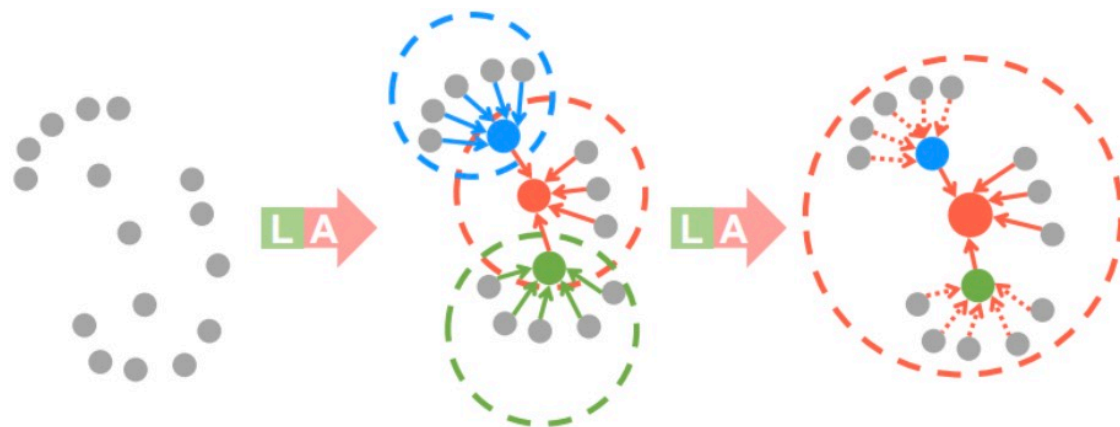
增加每个点的感受野

模块通过跳连接堆叠，作为扩张的残差块。

第一次观察到 K 个相邻点，第二次能最多 K^2 个相邻点

堆叠了两组标准残差块，效率和结果平衡

Dilated residual block	mIoU(%)
(1) one aggregation unit	41.9
(2) three aggregation units	48.7
(3) two aggregation units (The Standard Block)	52.0



3. RandLA-Net

□ 实验结果

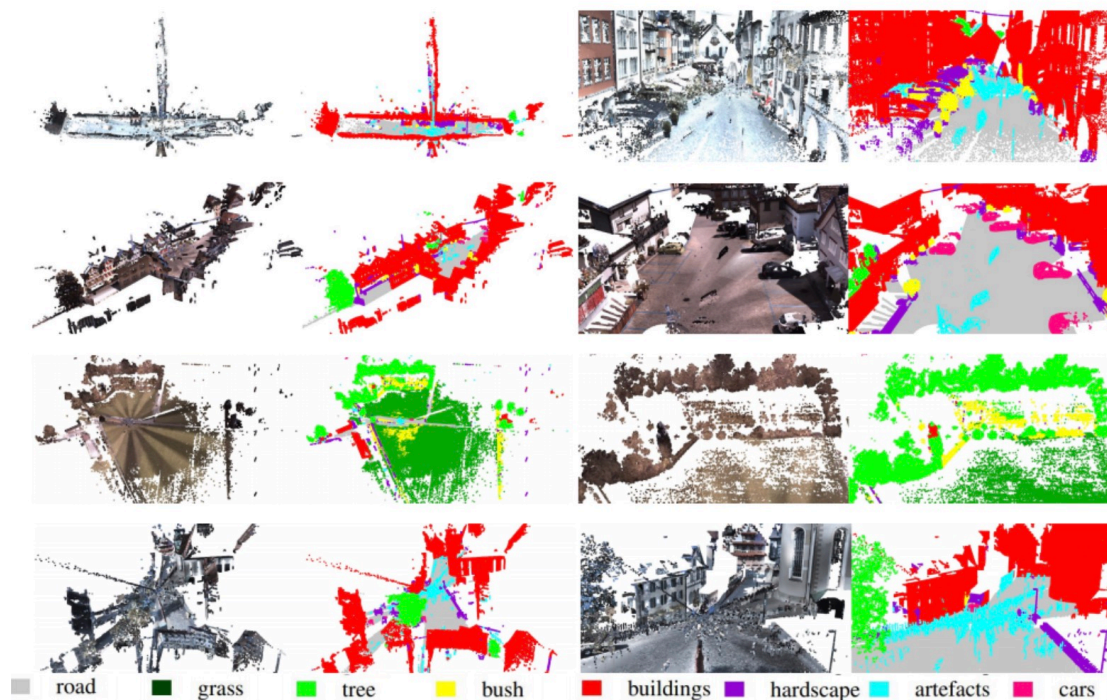
RandLA-Net花费最短的时间（每秒23帧）来推断大规模点云（最多 10^6 个点）的语义标签

	Total time (seconds)	Parameters (millions)	Maximum inference points (millions)
PointNet (Vanilla) [37]	192	0.8	0.49
PointNet++ (SSG) [38]	9831	0.97	0.98
PointCNN [29]	8142	11	0.05
SPG [23]	43584	0.25	-
KPConv [48]	717	14.9	0.54
RandLA-Net (Ours)	176	0.95	1.15

3. RandLA-Net

□ Semantic 3D

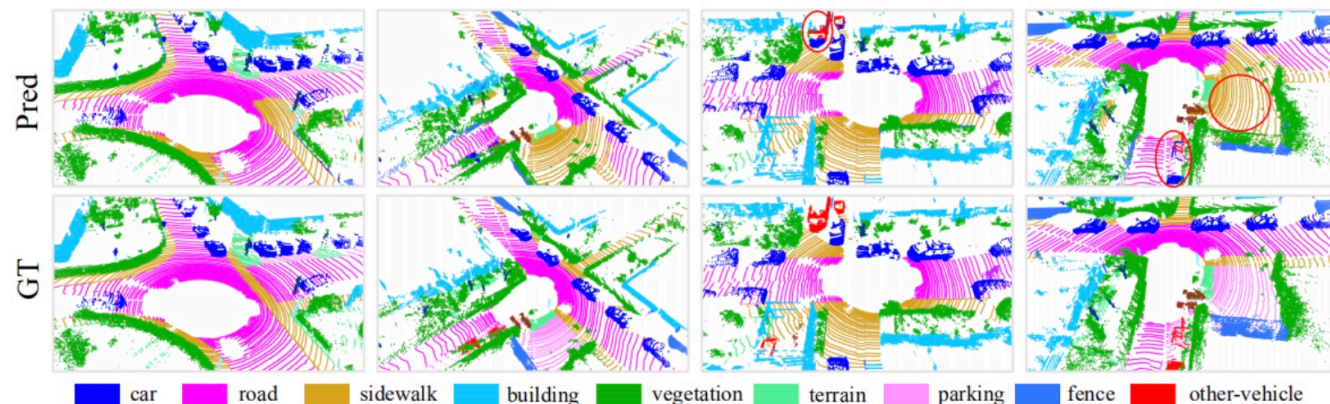
	mIoU (%)	OA (%)	man-made.	natural.	high veg.	low veg.	buildings	hard scape	scanning art.	cars
SnapNet_ [4]	59.1	88.6	82.0	77.3	79.7	22.9	91.1	18.4	37.3	64.4
SEGCloud [46]	61.3	88.1	83.9	66.0	86.0	40.5	91.1	30.9	27.5	64.3
RF_MSSF [47]	62.7	90.3	87.6	80.3	81.8	36.4	92.2	24.1	42.6	56.6
MSDeepVoxNet [40]	65.3	88.4	83.0	67.2	83.8	36.7	92.4	31.3	50.0	78.2
ShellNet [63]	69.3	93.2	96.3	90.4	83.9	41.0	94.2	34.7	43.9	70.2
GACNet [50]	70.8	91.9	86.4	77.7	88.5	60.6	94.2	37.3	43.5	77.8
SPG [23]	73.2	94.0	97.4	92.6	87.9	44.0	83.2	31.0	63.5	76.2
KPConv [48]	74.6	92.9	90.9	82.2	84.2	47.9	94.9	40.0	77.3	79.7
RandLA-Net (Ours)	76.0	94.4	96.5	92.0	85.1	50.3	95.0	41.1	68.2	79.4



3. RandLA-Net

□ Semantic kitti

Methods	Size	mIoU(%)	Params(M)	road	sidewalk	parking	other-ground	building	car	truck	bicycle	motorcycle	other-vehicle	vegetation	trunk	terrain	person	bicyclist	motorcyclist	fence	pole	traffic-sign
PointNet [37]	50K pts	14.6	3	61.6	35.7	15.8	1.4	41.4	46.3	0.1	1.3	0.3	0.8	31.0	4.6	17.6	0.2	0.2	0.0	12.9	2.4	3.7
SPG [23]		17.4	0.25	45.0	28.5	0.6	0.6	64.3	49.3	0.1	0.2	0.2	0.8	48.9	27.2	24.6	0.3	2.7	0.1	20.8	15.9	0.8
SPLATNet [43]		18.4	0.8	64.6	39.1	0.4	0.0	58.3	58.2	0.0	0.0	0.0	0.0	71.1	9.9	19.3	0.0	0.0	0.0	23.1	5.6	0.0
PointNet++ [38]		20.1	6	72.0	41.8	18.7	5.6	62.3	53.7	0.9	1.9	0.2	0.2	46.5	13.8	30.0	0.9	1.0	0.0	16.9	6.0	8.9
TangentConv [45]		40.9	0.4	83.9	63.9	33.4	15.4	83.4	90.8	15.2	2.7	16.5	12.1	79.5	49.3	58.1	23.0	28.4	8.1	49.0	35.8	28.5
SqueezeSeg [52]	64*2048 pixels	29.5	1	85.4	54.3	26.9	4.5	57.4	68.8	3.3	16.0	4.1	3.6	60.0	24.3	53.7	12.9	13.1	0.9	29.0	17.5	24.5
SqueezeSegV2 [53]		39.7	1	88.6	67.6	45.8	17.7	73.7	81.8	13.4	18.5	17.9	14.0	71.8	35.8	60.2	20.1	25.1	3.9	41.1	20.2	36.3
DarkNet21Seg [3]		47.4	25	91.4	74.0	57.0	26.4	81.9	85.4	18.6	26.2	26.5	15.6	77.6	48.4	63.6	31.8	33.6	4.0	52.3	36.0	50.0
DarkNet53Seg [3]		49.9	50	91.8	74.6	64.8	27.9	84.1	86.4	25.5	24.5	32.7	22.6	78.3	50.1	64.0	36.2	33.6	4.7	55.0	38.9	52.2
RandLA-Net (Ours)	50K pts	50.3	0.95	90.4	67.9	56.9	15.5	81.1	94.0	42.7	19.8	21.4	38.7	78.3	60.3	59.0	47.5	48.8	4.6	49.7	44.2	38.1



3. RandLA-Net

- 随机采样、局部特征聚合模块
- 多个基准上进行的广泛实验证明了方法的高效率和最新性能
- 证明了通过使用轻量级网络体系结构可以有效地分割大规模点云

谢谢聆听！