



中国科学技术大学  
University of Science and Technology of China

# PointNet++

中国科学技术大学

2023 年 12 月 17 日



- 1 介绍
- 2 PointNet
- 3 PointNet++



## 1 介绍

## 2 PointNet

## 3 PointNet++

点云是一种重要的几何数据结构。由于其不规则的格式，大多数研究人员将这些数据转换为规则的 3D 体素网格或图像集合。然而，这使得数据不必要地庞大，并造成问题。研究人员设计了一种新型的神经网络 PointNet，它直接以点云作为输入，考虑了输入点的排列不变性。PointNet 提供了一个统一的架构，用于对象分类，物体分割，场景分割。

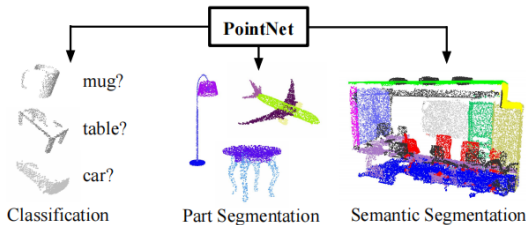


图: 1

研究者设计了一个深度学习框架，直接使用无序点集作为输入。点云表示为一组 3D 点  $\{P_i | i = 1, \dots, n\}$ ，其中每个点  $P_i$  是其  $(x, y, z)$  坐标加上诸如颜色、法线等的额外特征通道的向量。

对于对象分类任务，Pointnet 为所有  $k$  个候选类输出  $k$  个分数。对于语义分割，输入可以是用于部分区域分割的单个对象，或者是用于对象区域分割的 3D 场景。

Pointnet 将为  $n$  个点中的每个点生成  $m$  个语义类别分数，共输出  $n \times m$  个分数。

### 3D 计算机视觉

**Volumetric CNNs:** 将三维卷积神经网络应用于体素化 3D 视觉学习。然而，由于 3D 数据的稀疏性和三维卷积的计算成本，体素表示受到其分辨率的限制。FPNN 和 Vote3D 提出了处理稀疏性问题的特殊方法，然而，处理非常大的点云仍然是一个挑战。

**Multiview CNNs:** 将三维点云渲染成二维图像，然后应用二维卷积网络对其进行分类。通过设计良好的图像学习网络，这种方法在形状分类和检索任务上取得了主要的性能。然而，将它们扩展到场景分割或其他 3D 任务，如点云补全是非常困难的。

**Feature-based DNNs:** 将三维数据转换为一个向量，然后利用全连接网络对形状进行分类。这样提取到特征的表示能力有很大的约束，因为将无序的点云转化成了有序的向量。

无序性，与图像中的像素阵列或体素网格中的体素阵列不同，点云是一组没有特定顺序的点，输入为  $N$  个点的网络需要对  $N!$  保持输出不变性。交互性，点云中的这些点来自于具有相同距离度量的欧几里得空间。这意味着点与点之间不是孤立的，相邻的点形成了一个相互关联的子集，这意味着模型需要从附近的点捕获局部结构的语义特征。仿射性，点云是一个完整的仿射对象，即对整个点云进行仿射变换如旋转和平移后仍和原始点云语义特征保持一致。



1 介绍

2 PointNet

3 PointNet++



应用点对称函数来近似点集上的一般函数，通过多层感知器网络近似  $h$ ，通过单变量函数和最大池函数的组合近似  $g$

$$f(\{x_1, \dots, x_n\}) \approx g(h(x_1), \dots, h(x_n))$$

$$\forall \epsilon > 0, \exists h, \gamma, \quad st. \left| f(S) - \gamma \left( \max_{x_i \in S} \{h(x_i)\} \right) \right| < \epsilon$$



在计算全局点云特征向量之后，通过将全局特征与每个点特征拼接起来反馈给每个点。然后我们基于拼接的点特征提取新的逐点特征，其中包含了局部信息和全局信息。

通过一个 T-net 预测仿射变换矩阵，并将直接将此变换应用于输入点云，其中 T-net 只由简单的全连接层组成。这一思想也可以进一步扩展到特征空间的对齐。我们可以对逐点特征应用另一个对齐网络，使用同样的方法预测特征仿射变换矩阵，以对齐来自不同输入点云的特征。为了将特征变换矩阵约束为接近正交矩阵，将正则化项添加到训练损失中。

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



PointNet 只能捕捉点云的全局特征，不能捕获欧氏空间点集的局部结构特征，从而限制了其识别细粒度模式的能力和对复杂场景的泛化能力。



- 1 介绍
- 2 PointNet
- 3 PointNet++

使用分层特征学习代替 PointNet 中使用单个最大池化聚合点云特征。具体来说，每一个抽象层由一个采样层，一个分组层和一个学习层组成。

采样层，给定输入特征点集，使用迭代最远点采样（FPS）来选择特征点集的子集，与随机采样相比，在相同的质心数下，该方法对整个点集的覆盖率更高。

分组层，使用球查询搜索到质心点的特征距离在半径内的所有点（在实现中设置上限  $K$ ），与 kNN 相比，球查询的局部邻域保证了固定的区域尺度，从而使局部区域特征在空间上更具泛化性，且利于学习层中的 PointNet 提取特征。

学习层，使用 PointNet 学习逐点特征，局部区域中的点的坐标首先被转换成相对于质心点的局部坐标系，通过 PointNet 得到该局部区域的全局特征作为下一个采样层的输入。

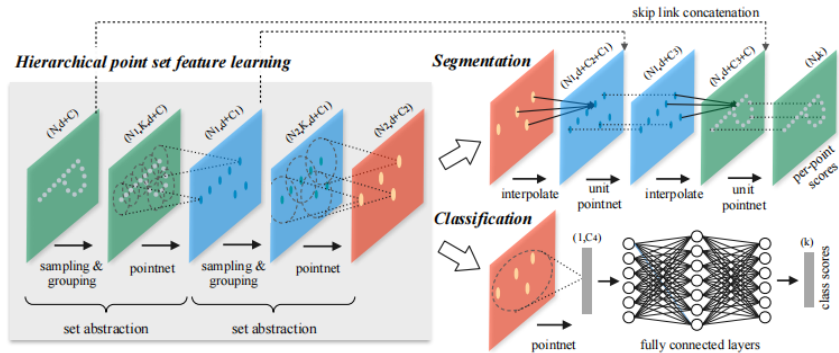


图: 2

多尺度分组 (MSG), 如图 (a) 所示, 在每一个抽象层中使用具有不同尺度的分组层, 然后根据 PointNets 提取每个尺度的特征。不同尺度的特征被连接以形成多尺度特征。

多分辨率分组 (MRG), 由于上面的 MSG 计算开销较大, 因为它在每个质心点的大规模邻域中运行 PointNet。通常由于在底层时质心点的数量相当大, 因此时间成本是高昂的。MRG 的层特征是两个向量的拼接。一个向量 (图中左侧) 是通过使用单尺度分组 (SSG) 汇总每个子区域的特征而获得的。另一个矢量 (右) 是通过使用单个 PointNet 直接处理局部区域中的所有原始点获得的特征。当局部区域的密度低时, 左侧向量比右侧向量更不可靠, 因为在计算左侧向量时子区域点云稀疏导致采样不足。在这种情况下, 右侧向量的权重应该更高。相反, 当局部区域的密度高时, 左侧矢量提供了更精细的信息, 此时应该提高左侧向量的权重。



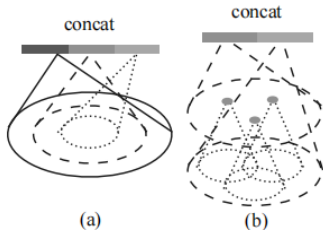


图: 3



使用基于特征距离的插值传播方法逆向生成前一个抽象层所有点的近邻特征，再与前一个抽象层的分组学习特征拼接为逐点分割特征，使用 PointNet 的部分仿射架构进行特征学习，进行下一轮特征传播，直到传播至原始点云。在插值中，使用  $k$  个最近邻的反距离加权平均特征。

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



谢谢!