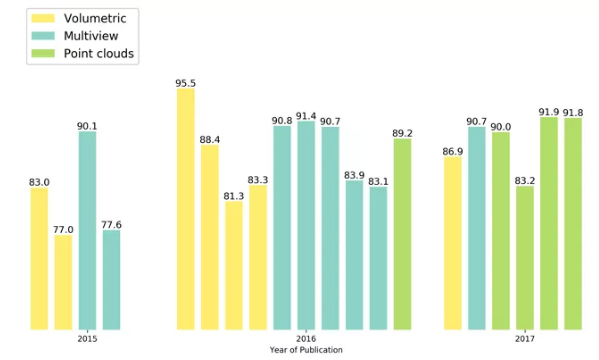
三维点云分类的深度学习方法研究

1. 三维点云分类技术概述

三维数据的表示方法：

1. point cloud：点云，也就是三维坐标系统中点的集合，这些点通常以x，y，z坐标来表示，并且一般用来表示物体的外表形状。当然，除了最基本的位置信息以外，也可以在点云中加入其他的信息，如点的色彩信息等。大多数的点云是由3D扫描设备获取的，如激光雷达，立体摄像机，深度相机等。
2. Mesh：网格，是由一组凸多边形顶点以及凸多边形表面组成的，也叫做非结构化网格。多边形网格是希望通过一种易于渲染的方式来表示三维物体模型。在三维可视化等方面有很大的作用。现在有很多种方法来将点云转换成多边形网格。
3. Voxel：体素，概念上类似于二维空间中的最小单位--像素，体素可以看作是是数字数据在三维空间分区中的最小单位，体素化是一种规格化的表示方法，在很多方面都有着重要的应用。
4. Multi-View Images：多视角图片，是通过不同视角的虚拟摄像机从物体模型中获取到的二维图像的集合。多视图通常需要使用比较多的图片来构建完整的三维模型，在固定图片数量的情况下，很容易受到物体自遮挡等因素的影响。

研究调查（2015-2017）中三种发表的文章，主要针对对点云进行深度学习的论文。 正如下图所示，展示了3D点云分类准确性出版（准确性，年份和数据类型），它总结了数据集上的最新准确性结果。 以及每种方法正在处理的数据的类型。 可以看到，在2015年，大多数方法都用于多视图数据（这是一种简短的说法 - 让我们拍摄3D模型的几张照片并使用2D方法处理它们），2016年更多的方法使用了体素表示的点云学习和2017年的基于点的方法有了大幅度的增长。

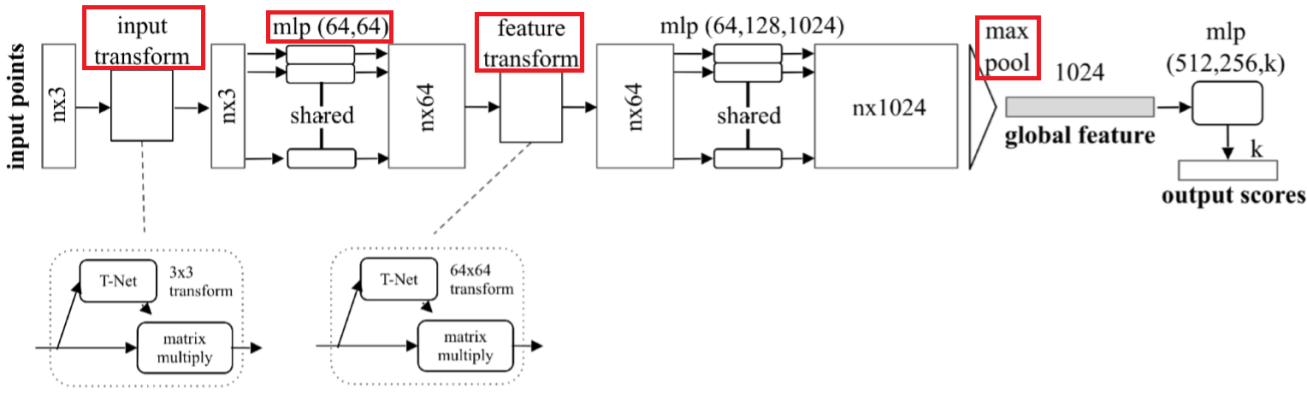


3D点云应用深度学习面临着一些挑战，主要有两个方面：

1. 首先在神经网络上面临的挑战：
2. 非结构化数据（无网格）：点云是分布在空间中的XYZ点。 没有结构化的网格来帮助CNN滤波器。
3. 不变性排列：点云本质上是一长串点（nx3矩阵，其中n是点数）。 在几何上，点的顺序不影响它在底层矩阵结构中的表示方式，例如， 相同的点云可以由两个完全不同的矩阵表示。
4. 点云数量上的变化：在图像中，像素的数量是一个给定的常数，取决于相机。 然而，点云的数量可能会有很大的变化，这取决于各种传感器。
5. 在点云数据方面的挑战：
6. 缺少数据：扫描的模型通常被遮挡，部分数据丢失。
7. 噪音：所有传感器都是嘈杂的。 有几种类型的噪声，包括点云扰动和异常值。 这意味着一个点有一定的概率位于它被采样的地方（扰动）附近的某一半径范围内，或者它可能出现在空间的任意位置（异常值）。
8. 旋转：一辆车向左转，同一辆车向右转，会有不同的点云代表同一辆车

基于上述问题，来自斯坦福大学的学者提出了PointNet，主要的流程如下：

1. **pointnet介绍**



Pointnet的优势体现：

1. 解决无序性问题：他们分别在每个点上训练了一个MLP（在点之间分享权重）。每个点被“投影”到一个1024维空间。然后，他们用点对称函数（max-pool）解决了点云顺序问题。

对称函数：函数的输出值不随输入变数的排列而改变，如c = a + b;改变a和b的顺序不影响c的值。

1. 解决旋转问题：利用他们称为T-net的“迷你网络”解决了旋转问题。

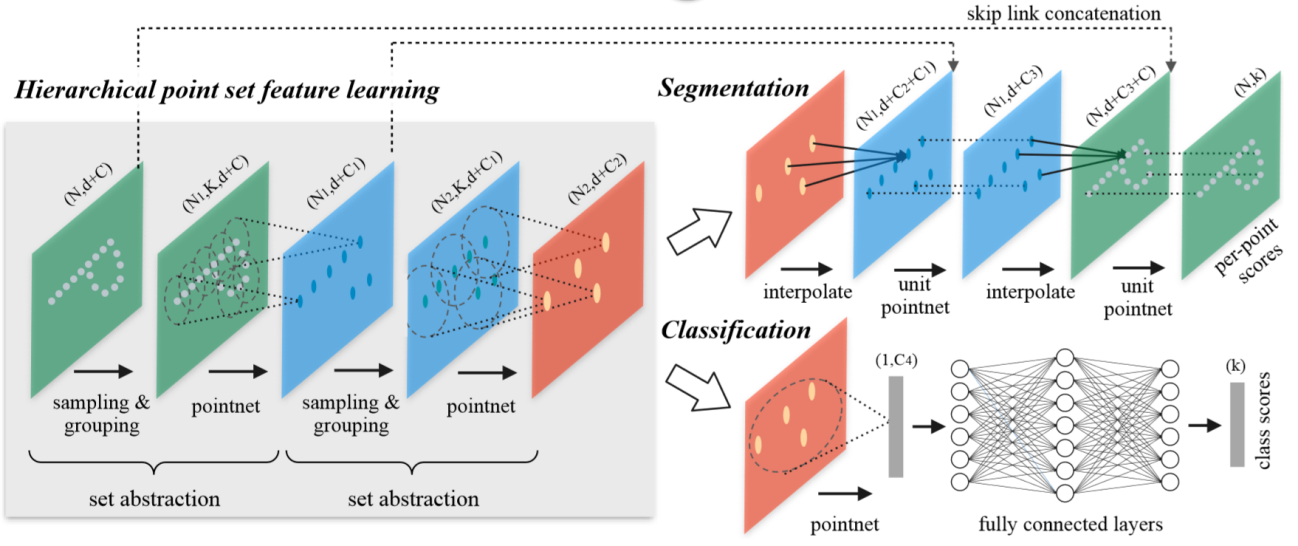
T-Net 是一个预测特征空间变换矩阵的子网络，它从输入数据中学习出与特征空间维度一致的变换矩阵，然后用这个变换矩阵与原始数据向乘，实现对输入特征空间的变换操作，使得后续的每一个点都与输入数据中的每一个点都有关系。通过这样的数据融合，实现对原始点云数据包含特征的逐级抽象。

Pointnet的局限性：

1. Pointnet的基本思想是对输入点云中的每一个点学习其对应的空间编码，之后再利用所有点的特征得到一个全局的点云特征。这里欠缺了对局部特征的提取及处理，比如说点云空间中临近点一般都具有相近的特征，同属于一个物体空间中的点的概率也很大，就好比二维图像中，同一个物体的像素值都相近一样。
2. 再者现实场景中的点云往往是疏密不同的，而Pointnet是基于均匀采样的点云进行训练的，导致了其在实际场景点云中的准确率下降。
3. **Pointnet++介绍**

针对Pointnet的局限性，提出了Pointnet++的框架。主要在两个方面有了优化处理：

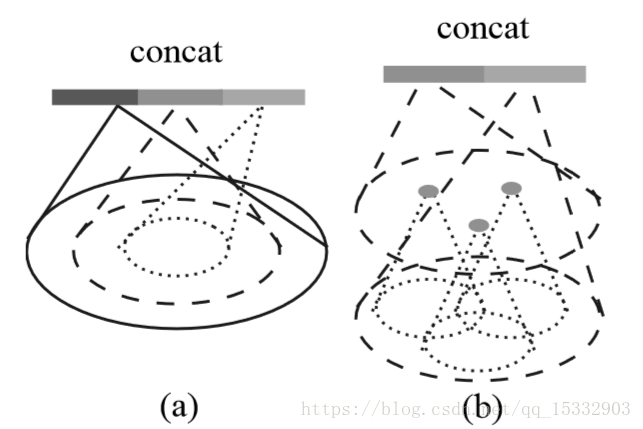
1. 局部特征提取：改进特征提取方法：pointnet++使用了分层抽取特征的思想，把每一次叫做set abstraction。分为三部分：采样层、分组层、特征提取层。首先来看采样层，为了从稠密的点云中抽取出一些相对较为重要的中心点，采用FPS（farthest point sampling）最远点采样法，这些点并不一定具有语义信息。当然也可以随机采样；然后是分组层，在上一层提取出的中心点的某个范围内寻找最近个k近邻点组成patch；特征提取层是将这k个点通过小型pointnet网络进行卷积和pooling得到的特征作为此中心点的特征，再送入下一个分层继续。这样每一层得到的中心点都是上一层中心点的子集，并且随着层数加深，中心点的个数越来越少，但是每一个中心点包含的信息越来越多。



1. 密度不均匀化：

方案一：多尺度分组: 如上图左所示，比较直观的思想是，在每一个分组层都通过多个尺度来确定每一个中心点的邻域范围，并经过pointnet提取特征之后将多个特征联合起来，得到一个多尺度的新特征。

 方案二：多分辨率分组:   很明显，通过上述做法，对于每一个中心点都需要多个patch的选取与卷积，计算开销很大，所以pointnet++提出了多分辨率分组法解决这个问题。如上图右所示，类似的，新特征通过两部分连接起来。左边特征向量是通过一个set abstraction后得到的，右边特征向量是直接对当前patch中所有点进行pointnet卷积得到。并且，当点云密度不均时，可以通过判断当前patch的密度对左右两个特征向量给予不同权重。例如，当patch中密度很小，左边向量得到的信息就没有对所有patch中点提取的特征可信度更高，于是将右特征向量的权重提高。以此达到减少计算量的同时解决密度问题。

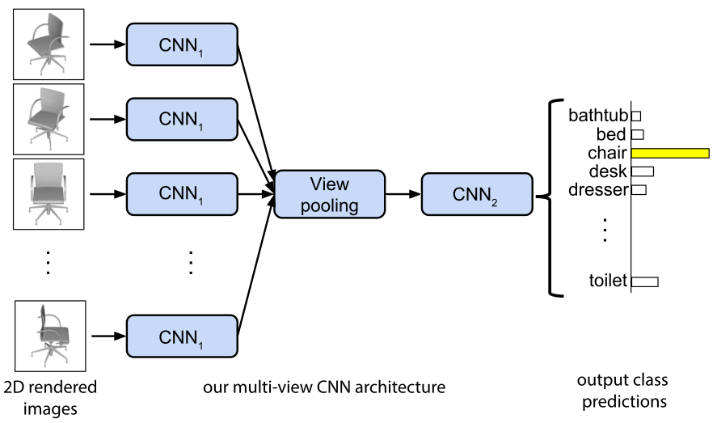


1. **其他的相关技术**

针对点云的非结构化特性，有很多可以利用的网络结构来进行处理。如下：

1. SONet：pointnet++中试图解决中心点邻域内点的排序问题，但没有考虑到选择中心点时就涉及到排序的问题，如果每次中心点都不一样的话，即使能够解决邻域的排序问题，但是这些邻域压根可能就互不相同。SONet在试图解决这个问题。
2. pointCNN : 在pointnet++中我们希望不论点云在怎样的坐标系下呈现，网络都能正确的识别出。这个问题可以通过STN（spacial transform netw）来解决。pointCNNd 核心思想X变换主要包含两部分，特征提取和X矩阵训练。解决点云排序问题，从代码来看，更确切地说，是K邻域的排序问题。提取特征的时候，小心翼翼，邻域维度上不敢轻举妄动，因为诸如卷积操作等都会因排序的变换而使结果发生变化，因此在这个维度上，pointnet用了maxpooling，本文作者没有进行任何操作，就静静地让邻域的点，各自升高了维度，然后听侯差遣。但是在训练X变换矩阵的时候，就放开手脚、大开大合了，为了得到K\*K的矩阵，不管物理意义，进行各种维度变化，可以说无所顾忌了。
3. kd-tree network：一种解决点云无序状态的结构，使用着名的Kd树在点云中创建一定的顺序结构的点云。一旦点云被结构化，他们就会学习树中每个节点的权重（代表沿特定轴的细分）。每个坐标轴在单个树层级上共享权重如下图中的所有绿色都具有共享权重，因为它们将数据沿x维度细分。测试了随机和确定性的空间细分，并说明了随机版本效果最好。但同时也说出了一些缺点。对旋转（因为它改变树结构）和噪声（如果它改变树结构）敏感。对于每个输入点云数据，都需要上采样，下采样或训练一个新模型。
4. **MVCNN介绍**

除了基于点云的深度学习算法还有有一种基于多视图像的三维数据的深度学习算法MVCNN。



如图所示，同一个3D形状的 每一张视角图像 各自独立地经过第一段的CNN1卷积网络，在一个叫做View-pooling层进行“聚合”。之后，再送入剩下的CNN2卷积网络。

整张网络第一部分的所有分支，共享相同的 CNN1里的参数。

在View-pooling层中，我们逐元素取最大值操作，另一种是求平均值操作，但在我们的实验中，这并不有效。 这个View-pooling层，可以放在网络中的任何位置。经过我们的实验，这一层最好放在最后的卷积层（Conv5），以最优化的执行分类与检索的任务。

View-pooling优点类似于max-pooling layer与maxout layer，不同点在于进行max操作时的维度不同，这里，所谓的“维度不同”，我看了好一会儿，我想应该是指，这里的max操作是“纵向的”，即在12个视角图像中，同一个位置的地方，进行max操作。而一般意义上所说的max-pooling，通常是指在一个2×22×2的领域像素单元内，取max像素值保留的操作。

存在的问题：

1. 在多视角的2D图像表示3D特征时，哪一张视角图像是最重要的，或者说哪一张图像包含了主要的特征信息？多少涨视角图像可以满足给定的精度要求？可以在程序运行中提取选择有价值的视角图像吗？
2. 该方法的实验都是在“虚拟的”3D形状上的，对于真实世界的3D物体（或者视频、而不仅仅是3D多边形网格模型），的特征描述算子提取效果还需进一步的研究？