Vote3Deep: Fast object detection in 3D point clouds using efﬁcient convolutional neural networks,”

<http://pdfs.semanticscholar.org/be37/149628faa6e6a213bcadc609bb71dade4a4a.pdf>

摘要

本文提出了一种计算有效的方法来使用卷积神经网络（CNN）在三维点云中本地检测物体。具体来说，这是通过利用以特征为中心的投票方案来实现新颖的卷积层，其明确地利用输入中遇到的稀疏性。为此，我们研究了不同体系结构的准确性和速度之间的平衡，并且还建议在滤波器激活中使用L1惩罚来进一步鼓励中间表示中的稀疏性。就我们所知，这是第一个提出稀疏卷积层和L1正则化的有效的3D数据大规模处理的工作。我们证明了我们的方法在KITTI物体检测基准测试中的有效性，并显示只有三层的Vote3Deep模型在基于激光和激光视觉的方法中的表现优于以前的技术水平，边缘高达40％，同时保持高度在处理时间方面具有竞争力。

介绍

3D点云数据在移动机器人应用（如自动驾驶）中无处不在，其中有效和强大的目标检测对于规划和决策至关重要。 最近，通过使用卷积神经网络（CNN）（例如[1]，[2]，[3]，[4]），计算机视觉正在经历一个转变。 然而，处理3D点云的方法尚未经历类似的突破。 我们将这种缺乏进展归因于第三个空间维度引入的计算负担。 由此产生的输入和中间表示尺寸的增加使CNN从二维视觉应用转变为在大规模应用中不可行的点云本地3D感知。 因此，先前的方法倾向于首先将数据转换为2D表示，其中附近的特征不一定在物理3D空间中相邻 - 需要模型来恢复这些几何关系。

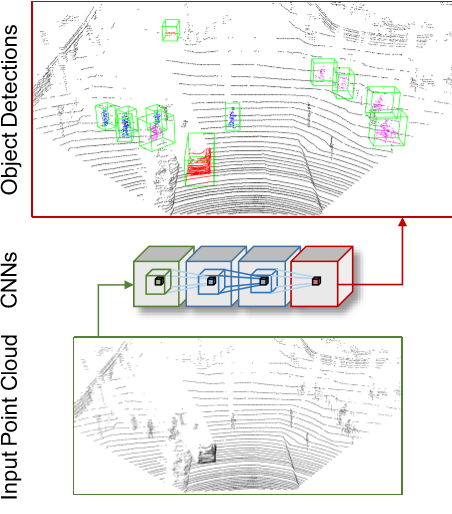
然而，与图像数据相比，移动机器人中遇到的典型点云在空间上是稀疏的，因为大部分区域是空闲的。 这个事实在[5]中被利用，其中作者提出Vote3D，一个以特征为中心的投票算法，利用这些点云内在的稀疏性。 计算成本仅与被占用的单元的数量成正比，而不是3D网格中的单元的总数。 证明了投票方案与稠密卷积运算的等价性，并通过将点云离散成三维网格并利用线性支持向量机（SVM）执行穷尽的3D滑动窗口检测来证明其有效性。 因此，[5]通过KITTI Vision基准测试套件[6]的目标检测任务，在检测点云中的车辆，行人和骑车人的性能和处理速度方面达到了先前的技术水平。

受到[5]的启发，我们建议利用以特征为中心的投票来建立有效的CNN，以便在3D中本地检测点云中的对象 - 也就是说，不需要将输入投影到较低维空间或约束搜索空间探测器（图1）。这使得我们的名为Vote3Deep的方法能够在测试时提供恒定时间评估的同时学习高容量的非线性模型，这与非参数方法相反。此外，为了在整个CNN堆栈中提高与稀疏输入相关的计算效益，我们通过在训练期间施加L1模型调节器来证明在中间层的输入中鼓励稀疏性的好处。就我们所知，这是第一个提出基于投票和L1正则化的稀疏卷积层来有效处理含CNN的全三维点云的第一个工作。本文的贡献尤其可以概括如下：

1）利用投票机制利用输入数据的固有稀疏性，构建有效的卷积层作为基于CNN的点云处理的基本构件;

2）使用整型线性单位和L1稀疏性惩罚来特别鼓励中间表示中的数据稀疏性，以便在整个CNN堆栈中利用稀疏卷积层。 我们展示了只有三层的Vote3Deep模型能够在流行的KITTI对象检测基准中考虑的所有类别的纯粹基于激光的方法中实现最先进的性能。 Vote3Deep模型在基于三维点云的物体检测方面的平均精度超过了先前的技术水平，达到了40％的余量，而在检测速度方面仅略微慢了一点。

将Vote3Deep应用于KITTI数据集中的一个看不见的点云，以及相应的图像作为参考。 CNN通过投票在3D中本地应用稀疏的卷积。 模型检测汽车（红色），行人（蓝色）和骑自行车者（洋红色），即使在远距离范围内，也可以按班级分配边界框（绿色）。 最好的颜色。



相关工作：

一些作品试图在3D点云数据的背景下应用CNN。基于CNN的方法[7]通过将点云投影到二维深度图中获得与KITTI用于汽车检测的性能相当的性能[5]，并且具有用于从地面点的高度的附加通道。他们的模型预测检测分数并回归到边界框。然而，对特定观点的投射丢弃了有价值的信息，这在特别有害的情况下，例如在拥挤的场景中。它还要求网络过滤器学习深度方面的本地依赖性，在3D表示中容易获得的信息可以通过稀疏卷积有效提取。

在[8]和[9]中用CNN处理从点云获得的密集3D占用网格。在最小单元尺寸为0.1m的情况下，[8]在GPU上报告6ms的速度，以32×32×32单元的网格尺寸对单个作物进行分类。类似地，在[9]中报告了着陆区检测的处理时间为5毫米每立方米。由于三维点云大于60m×60m×5m，会导致每帧处理时间为60×60×5×5×10-3 = 90s，不符合机器人应用中通常遇到的速度要求。

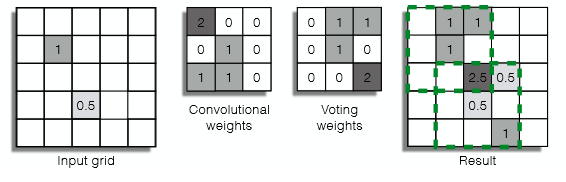
另外一种利用稀疏表示的方法可以在[10]和[11]中找到，其中稀疏卷积分别应用于相对较小的2D和3D作物。虽然卷积核仅适用于稀疏特征位置，但是所提出的算法仍然需要考虑具有零或恒定偏差值的相邻值，导致不必要的操作和存储器消耗。另一种执行稀疏卷积的方法在文献[12]中引入，他们利用“超自动面格子”，但只考虑相对较小的输入，而不是我们的工作。 CNN也被应用于生物医学图像分析中的密集三维数据（例如[13]，[14]，[15]）。残差网络的3D等价物[4]被用于[13]脑图像分割。文献[14]提出了一个两阶段的级联模型来检测脑微出血。 [15]中提出了三个CNN的组合。每个CNN处理一个不同的2D平面，三个流在最后一层连接起来。这些系统在相对较小的输入上运行，并且在GPU加速的情况下处理单一框架。

方法：

本节描述了卷积神经网络在预测从可变尺寸的稀疏3D输入网格中检测分数的应用。作为网络的输入，点云被分离成稀疏的3D网格，如[5]。对于包含非零点数的每个单元格，根据单元格中点的统计信息提取特征向量。特征向量保存二进制占用值，反射率值的平均值和方差以及三个形状因子。空格中的单元格不被存储，导致稀疏表示。我们采用[5]中的投票方案对这个本地3D表示进行稀疏卷积，然后是ReLU非线性，它返回一个新的稀疏3D表示。这个过程可以像传统CNN一样重复和堆叠，输出层预测检测分数。类似于[5]，CNN被应用于N个并行线程中N个不同角度方向的点云，以最小的增加计算时间来处理不同方向的对象。在3D空间中使用非最大抑制（NMS）修剪重复检测。 3D中的NMS能够更好地处理彼此背后的物体，因为3D边界框比2D投影少。基于三维空间中的边界框对于同一类的对象实例大小相似的前提，我们假定每个类都有一个固定大小的边界框，这消除了回归边界框大小的需要。我们根据训练集上第95百分位的地面实际边界框大小，为每个感兴趣的类选择三维边界框尺寸。

网络的接收域至少应与对象的边界框一样大，但不能过大，这会浪费计算时间。 因此，我们使用了几个类别特定的网络，可以在测试时间并行运行，每个网络根据对象类别具有不同的总接受范围。 原则上，可以使用单个网络来计算多个类别的检测分数; 为未来的工作留下了一个任务。

一个稀疏的2D示例输入没有偏见的投票程序的说明。 投票权重是通过沿每个维度卷积卷积权重获得的。 而标准卷积在输入中的每个位置应用滤波器，而只需要在每个非零位置应用等价的投票程序来计算相同的结果。 Vote3Deep不是具有单一特征的2D网格，而是将投票程序应用于具有多个特征映射的3D输入。 对于一个完整的数学证明，读者可以参考[5]。 最好的颜色。



1. 通过投票进行稀疏卷积

当在离散点云上进行密集的三维卷积时，大部分计算时间被浪费，因为大部分运算是乘以零。 与形成基于图像的CNN的基础的二维卷积相比，额外的第三空间维度使得该过程在计算上更加昂贵。 使用有意义的计算仅在3D特征非零的情况下发生的见解，[5]引入以特征为中心的投票方案。 该算法的基础是让每个非零输入特征向量按照过滤器的接受域来定义一组由投票权重加权的投票到其在输出层中的周围单元。 通过沿每个空间维度卷积卷积滤波器内核来获得投票权重。 最后的卷积结果是通过累积投票的每个单元格中的投票（图2）。

这个程序可以正式说明如下。 在不失一般性的情况下，假设我们在网络层c中有一个具有奇数核维数的三维卷积滤波器，在一个输入特征上运算，滤波器的权重由wc∈R（2I + 1）×（2J + 1）× （2K + 1）。然后，输入网格hc-1∈RL×M×N，位置（l，m，n）的卷积结果由下式给出：

其中bc是应用于网格中所有单元的偏差值。

这个操作需要被应用到输入网格中的所有L×M×N个位置，以进行规则的密集卷积。 与此相反，给定所有非零单元的单元索引集合Φ= {（l，m，n）‖hc-1 l，m，n 6 = 0}，卷积可以被重写为 以特征为中心的投票操作，每个输入单元格投票表决增加相邻单元格位置的值根据：

表决输出通过ReLU非线性传递，丢弃了下一小节所述的非正向特征。 至关重要的是，偏见被限制为非正面的，因为单一的正面偏见会返回一个输出网格，其中几乎每个细胞都被一个特征向量所占据，因此消除了稀疏性。 因此，偏置bc只需要被添加到每个非空的输出单元。 使用这种稀疏的投票方案，只需要将滤波器应用于输入网格中的占用单元，而不是在整个网格上进行卷积。 该算法在[5]中有更详细的描述，其中包括形式证明，即以特征为中心的投票相当于穷举卷积。

1. 用ReLU保持稀疏性

在所有层执行快速投票的能力是基于对每个单独层的输入的稀疏性的假设。虽然输入点云是稀疏的，非空单元的区域是由每个连续的卷积层扩大，大约相应的滤波器的接收场大小。因此，选择一个非线性激活函数是非常重要的，这有助于保持每个卷积层输入的稀疏性。这是通过在稀疏卷积层之后应用[16]中提出的整型线性单元（ReLU）来实现的。 ReLU激活可以被写为：

hc = max（0，zc）（3）

其中zc是通过稀疏卷积计算的层c中的ReLU非线性的输入，并且hc是输出，表示隐藏的激活在随后的稀疏中间表示中。在这种情况下，只有具有大于零值的要素才能在下一个稀疏卷积图层中投票。除了使网络能够学习非线性函数逼近并因此增加其表示容量之外，通过丢弃有助于保持中间表示中的稀疏性的负面特征值，ReLU有效地执行阈值操作。最后，与其他非线性相比，ReLU的另一个优点是计算速度快。

Xunlian：

由于使用了固定尺寸的边界框，网络可以直接在正面和负面实例的三维作物上进行训练，其尺寸等于体系结构指定的接受场大小。 在固定的训练时期之后，通过定期进行硬性负面挖掘来获得负面训练的例子。 类特定的网络是二元分类器，我们选择一个线性铰链损失的训练，由于其最大的保证金属性。

表1中的“D型”体系结构的示意图。层c的输入x（绿色）和中间表示hc（蓝色）是稀疏3D网格，其中每个占据空间位置保持特征向量（立方体）。 具有滤波器权重wc的稀疏卷积在3D中本地执行以计算预测（红色）。

。 表1中的“D型”体系结构的示意图。层c的输入x（绿色）和中间表示hc（蓝色）是稀疏3D网格，其中每个占据空间位置保持特征向量（立方体）。 具有滤波器权重wc的稀疏卷积在3D中本地执行以计算预测（红色）。 最好的颜色。

最好的颜色

