

硕 士 研 究 生 读 书 报 告



题目 基于补丁的3D形状深度表示

作者姓名 贺晨韬

作者学号 21851416

指导教师 赵斌

学科专业 软件工程

所在学院 软件学院

提交日期 二○18 年 12 月

A Patch-based Deep Representation of 3D Shapes

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor:Bin Zhao

By

Chentao He

Zhejiang University, P.R. China

2018

摘要

本文介绍了一种自适应的八叉树卷积神经网络，用于高效的三维形状的编码与解码。这种方法使用平面补片的方式对每个8分圆内的3D形状进行建模，通过这种基于补丁的方式提出一种用于编码和解码3D形状的自适应O-CNN编码器和解码器。相对于传统的框架，这种方法自适应O-CNN不仅降低了存储器和计算成本，而且提供了比现有3D-CNN方法更好的形状生成能力。最后从不同形状分析和生成任务的效率和有效性方面验证了自适应O-CNN，包括形状分类，3D自动编码，单个图像的形状预测以及噪声和不完整点云的形状完成。

关键词：补丁引导自适应八叉树，自适应八叉树卷积神经网络，形状重建，3D形状编码与解码

Abstract

This paper presents an adaptive octree convolution neural network for efficient coding and decoding of three-dimensional shapes. This method uses a planar patch to model the 3D shape in each 8-minute circle. Through this patch-based approach, an adaptive O-CNN encoder and decoder for encoding and decoding the 3D shape are proposed. Compared with traditional frameworks, this adaptive O-CNN not only reduces memory and computing costs, but also provides better shape generation capability than existing 3D-CNN methods. Finally, the efficiency and validity of different shape analysis and generation tasks are validated, including shape classification, 3D automatic coding, shape prediction of single image and shape completion of noise and incomplete point clouds.

Keywords：Patch-guided adaptive octree, Adaptive O-CNN, shape reconstruction, 3D encoder and decoder

1.研究背景

3D形状的分析和生成是计算机图形学中的2个基本任务。传统方法对于处理复杂图形的能力有限，并且会花费大量的时间和精力。着创建和捕获的3D数据的快速增长，借助于机器学习技术从大型3D数据集中学习形状空间并利用所学习的特征指导形状分析和生成已成为可能。最近，卷积神经网络的深度学习已经应用于3D形状分析和合成。

2.传统方法介绍

与基于网格的表示简单且规则的图像不同，由于来自实际应用的不同需求，3D形状具有各种表示。对于基于学习的形状生成任务，3D形状的表示起着至关重要的作用，其影响学习架构的设计和生成的形状的质量。

1.体素表示法[1][2]

体素表示法2D图像的自然扩展，非常适合现有的学习框架。但是它对存储器和计算能力有很高的需求，并且在实践中难以产生高分辨率输出。

2.多视图图像[3]

多视图图像已经被广泛应用于形状生成中，生成的多视图图像可以融合以重建完整的形状。但是如何选取合适的视图，强制执行不同视图的一致性以及形状遮挡是此表示的主要挑战。

3.点云表示[4]

最近随着PointNet的发展，点已经成为形状分析和生成的合适表示。但是，点的输出质量收到点数的限制，从点云中提取高质量的表面需要额外的处理。

4.多边形网格[5]

表面贴片或网格可以通过神经网络直接预测，该网络使模板网格变形或发现2D到3D的映射。但是，预定义的网格拓扑和模板网格的常规曲面细分会阻止生成高质量的结果，尤其是对于不规则和复杂的形状。

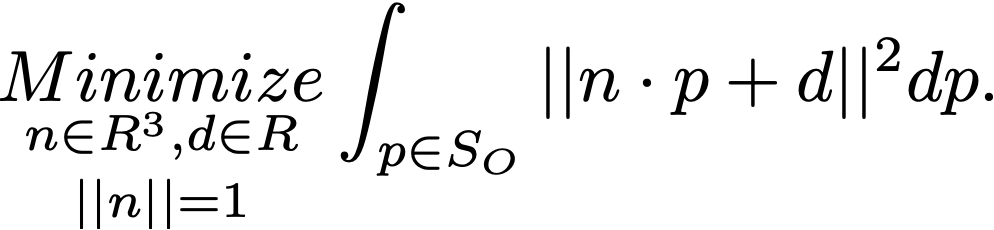
5.八叉树表示[6]

八叉树是最具代表性的稀疏体素表示，最近已经与卷积神经网络结合用于形状分析及生成高分辨率的形状。基于八叉树的生成网络通常预测八个卦限的占用概率，即完全占有，自由和边界，并利用标签边界对八叉树进行分割。通过递归地执行预测和分隔过程达到预定义的最大深度。在最好的状况下，非空的叶子结点表示预测的形状表面。在现有工作中，最优水平下的非空叶子结点可以看作是x、y和z方向上形状的均匀样本。作者认为实际上没有必要以这种统一的方式存储形状信息，因为一些八分圆内部的局部形状可以用简单的补丁（如平面补丁）来表示。因此，通过存储修补信息可以使生成的八叉树可以具有更紧凑和自适应表示。此外，所存储的补丁具有比使用八分圆的中心或一个角更高的阶近似精度。

基于上述观察，作者提出一种新的3D形状卷积神经网络，称为自适应八叉树神经网络。自适应O-CNN基于新颖的贴片引导的自适应八叉树形状表示，其根据估计的简单贴片的近似误差自适应地将八分圆分割成八分圆所包含的局部形状。自适应O-CNN的解码器预测八分圆的占用概率：空的，表面良好近似的和表面近似的;在每个级别的每个非空的八分圆处推断局部补丁，并且分割表面近似程度最差的卦限。它产生自适应八叉树，其在非空叶子八分圆处的估计局部斑块是预测形状的多尺度和自适应表示。除了解码器之外，我们还开发了一种用于自适应八叉树的高效3D编码器，并将其用于形状分类和3D自动编码器。

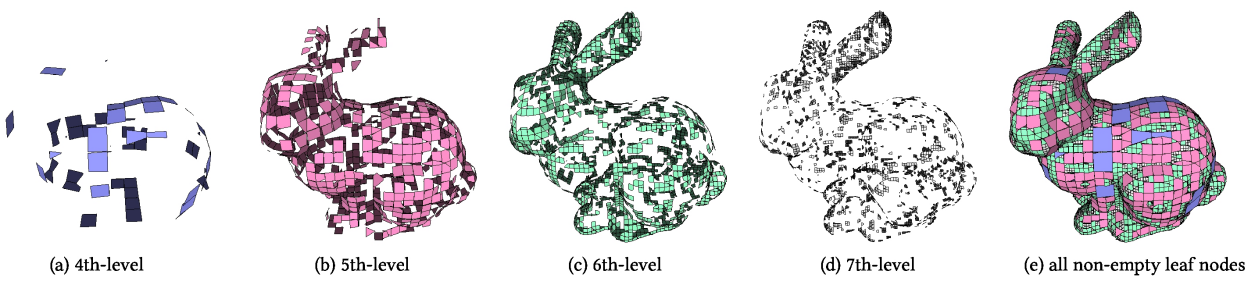
3.补丁引导自适应八叉树的方法

作者引入一个补丁引导分区策略来生成自适应八叉树。对于给定的曲面S，我们从其边界框开始并执行1到8的细分。对于其中的一个卦限O，将SO表示为受O的立方区域限制的S的局部表面。如果是SO不为空 ，我们将选择一个简单的曲面贴片补丁近似作为SO。在本文中，我们选择最简单的平面作为补丁来指导自适应八叉树构造。具有最小近似误差的最佳平面P满足以下要求：

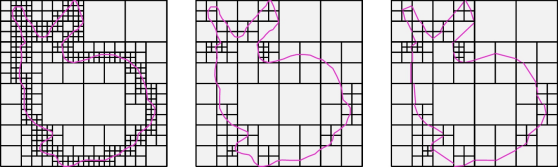


其中n为平面P的法向量，平面P的方程为。为了使法线方向与形状的方向一致，通过检查n与SO平均法线方向的夹角是否小于90度，如果小于90度，则将n与d均取反。我们将PO表示为受O的立方体限制的P的表面，局部斑块的质量δO由PO和SO的Hausdorff距离定义：

构造八叉树的规则为:对于任何非空的卦限O，如果它的深度小于最大深度且δO大于PO和SO的Hausdorff距离，则对它继续分割。



以Bunny为例，在第四层，第五层，第六层和第七层中分别对已经达到近似要求的八分圆使用平面来代替，剩下的继续分割，最终达到最大的深度。图中的不同颜色代表不同深度。



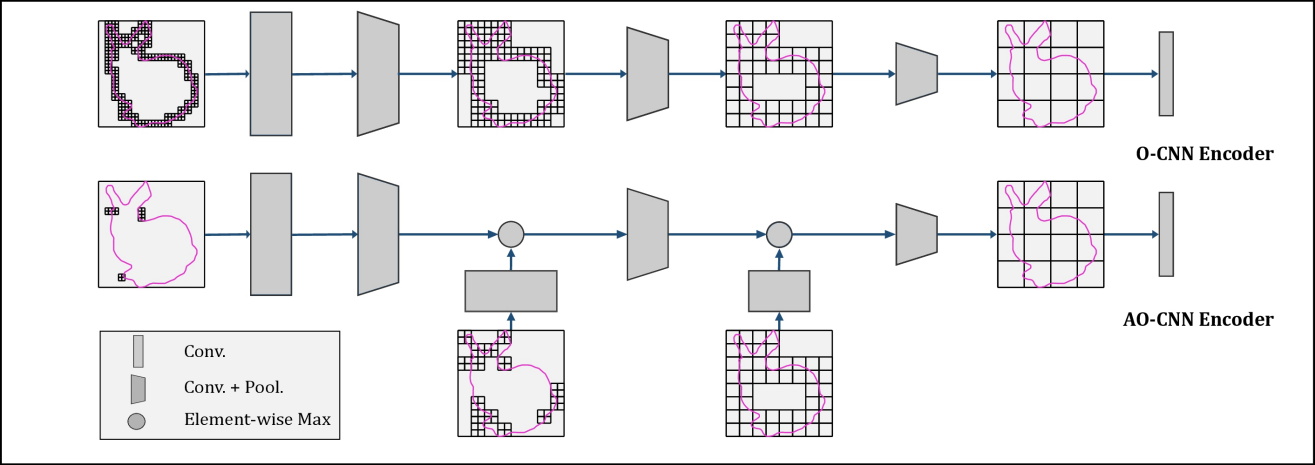
同样这种方法可以用在2D的形状中，如图所示。通过这种方法可以极大的减小四叉树所占用的空间。但是，由于近似的误差，相邻的非空叶子八分圆之间的局部斑块不是无缝连接的，即在八分圆的边界区域上存在间隙。为了填补这些空白可以使用表面重建[7]和多边形修复技术[8]。

1. 自适应八叉树

3D CNN主要包括编码器和解码器它们对于形状分类，形状生成和其他任务至关重要。在4.1和4.2中将会分别介绍编码器和解码器。

4.1 Adaptive O-CNN的3D编码器

OCNN编码器：OCNN的关键思想是把在非空卦限中的稀疏的表面信息存储在八叉树中。在八叉树的每一层中，每一个卦限都通过它的混合密钥[9]来识别。混洗后的密钥按升序排序，并存在连续的数组中。给定一个卦限的混洗密钥，我们可以立即计算其邻居的混洗密钥并检索相应的邻域信息，这对于实现有效的CNN操作是必不可少的。为了获得两个相邻层级之间的父子对应关系并标记出空的卦限，引入另一个Label数组来记录每个卦限的信息。在八叉树上定义的常见CNN操作（卷积和池化）类似于体积CNN操作。唯一的区别是基于八叉树的CNN操作遵循以下原则：“存在卦限的地方，就存在CNN计算”。最初，形状信号存在于叶子节点，然后在八叉树的每个级别，顺序地应用CNN操作。当CNN操作的步幅为1时，信号以未改变的分辨率处理，并保持在当前八叉树级别。当CNN操作的步幅大于1时，信号被压缩并沿着八叉树从底部流向顶部。图4上显示出了O-CNN的编码器结构。

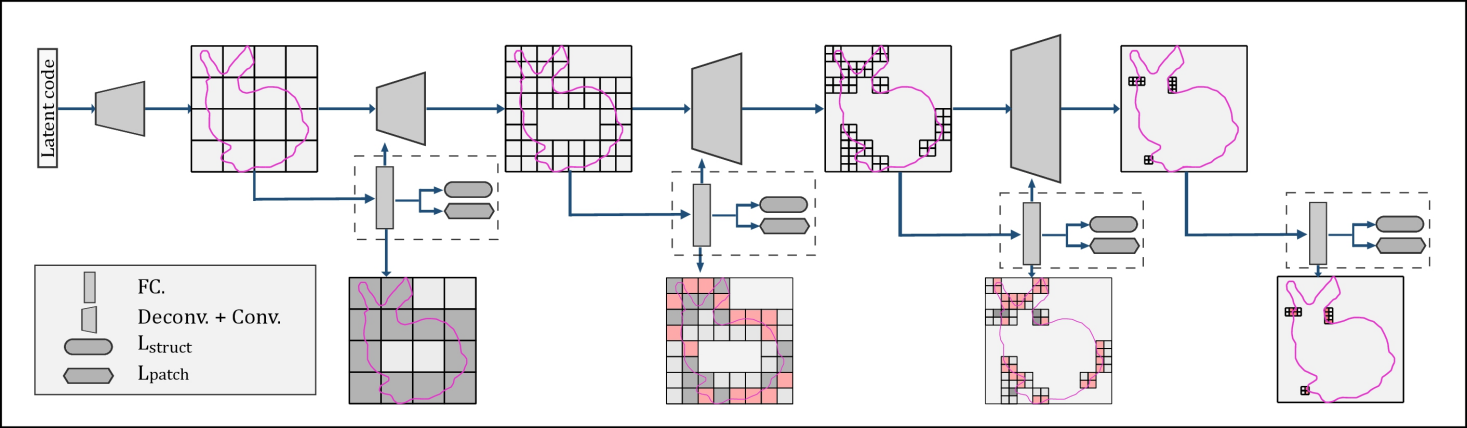


输入信号：与仅在叶子节点存储形状信息的OCNN不同，我们利用所有估计的局部平面作为输入信号。对于i层的任何一个节点O，它的局部平面为n·x+d=0,,我们设置一个4通道的信号：（n，）,=d - n·c,其中c是O平面的中心点。这里使用而不是d的原因是可以受到网格大小的限制，是一个相对的值。而d很大，因为d是从原点到平面的距离。对于一个空的卦限，它的输入是（0，0，0，0）；

自适应八叉树编码器：我们设计了一种新颖的网络结构，将自适应八叉树作为输入。在八叉树的每个级别上，我们将一系列卷积运算符和ReLU应用于此级别的所有八分圆上的特征，并且这些八分圆共享卷积内核。然后，通过合并将第l级的已处理特征下采样到 l-1级别，并通过逐元素最大化操作与 l-1级别的特征融合。这些新功能可以进一步处理，并与 l-2级别，l-3级别的融合，达到最粗糙的水平。在我们的实现中，最粗糙的级别设置为2，其中第二级的节点被强制设为已满，因此这些特征都具有相同的维度。图4-下说明了我们的自适应O-CNN 3D编码器架构。

4.2 Adaptive O-CNN的3D解码器

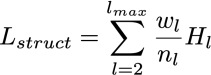
我们设计的一种解码器能够使用给定的隐藏代码生成八叉树，如图5所示。



在八叉树的每一层，我们通过训练神经网络来预测每一个卦限的近似状态，分别为空，表面良好近似的和表面较差近似的，并且对补丁的参数回归。表面较差近似的卦限将会被细分，并且将它们的特征进行反卷积后传递给子节点。表面良好近似的说明符合误差δ的要求，因此可以停止分裂并将它们留在叶子节点处。这种方法可以通过递归来达到最大深度。

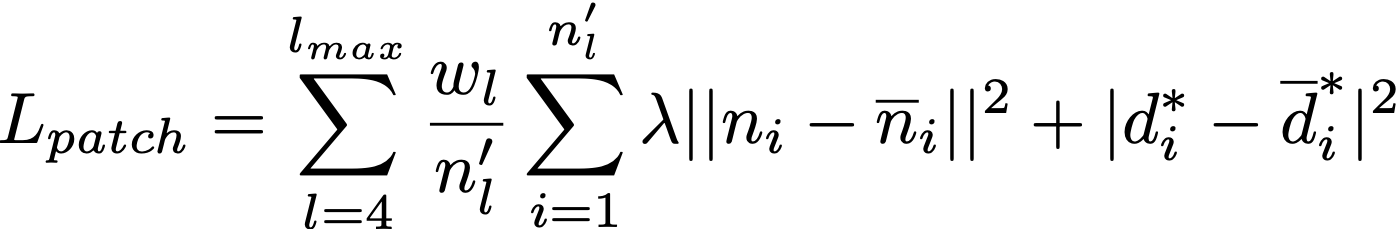
预测功能：用于预测的神经网络包含4个部分“FC+BN+ReLU+FC”。BN代表批标准化，FC表示全连接层。这个模块由八叉树中相同层级的节点共享。预测的输出包括补丁的近似状态和平面补丁参数（n，）。补丁近似状态能够指导八叉树的细分。

损失函数：损失函数包括2个部分，结构损失和补丁损失。

1.结构损失用来衡量预测的八叉树与真值的差异。由于每一个卦限都有3种类别，因此使用交叉熵来定义结构损失。令H为第l层的交叉熵，则结构损失为所有层的交叉熵加权和形成：。

其中代表第l层的节点个数，代表树的最大深度，代表每一层的权重。八叉树的最少层数为2，因此从l=2开始。同时设置w=1。

2.补丁损失用来衡量平面的参数和真实的值之间的平方差：



其中和表示需要预测的参数，和表示真值，表示在预测八叉树中第l层的叶子数量,λ设为0.2。在我们的实现中，当树的深度超过4时，使八叉树自适应，因此l从4开始。若出现真值中不存在的情况，则只计算结构损失而不计算补丁损失。

我们使用作为整个的损失函数。同时令,其中是第l层的网格大小，这可以通过tanh函数来完成。

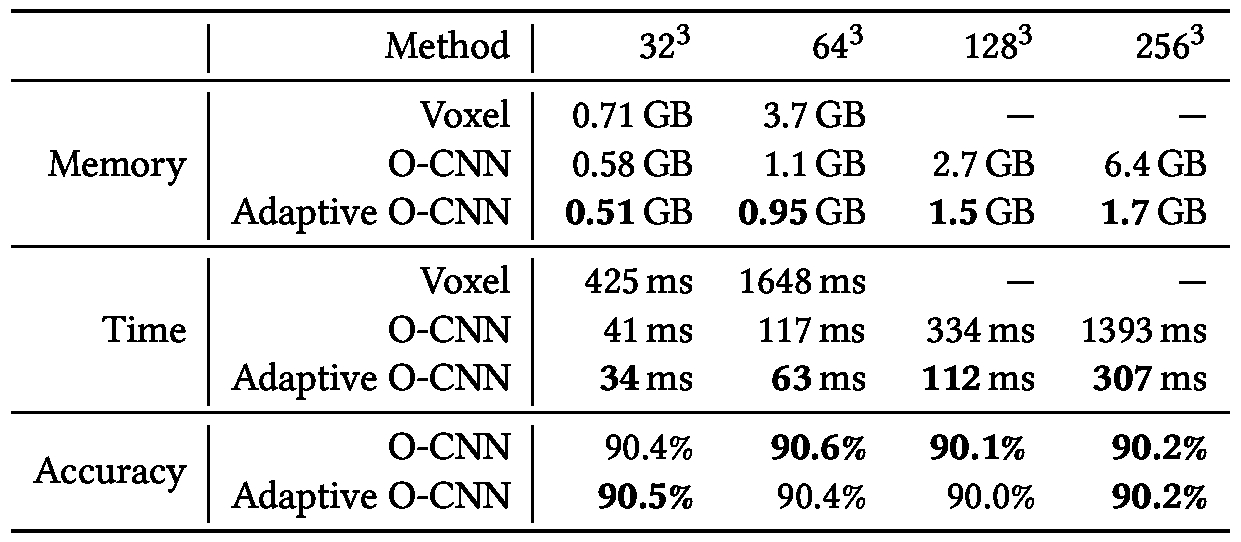
5 实验和比较

为了验证这种方法的性能，作者通过3个实验进行测试。3D形状分类，3D形状编码和3D形状预测。

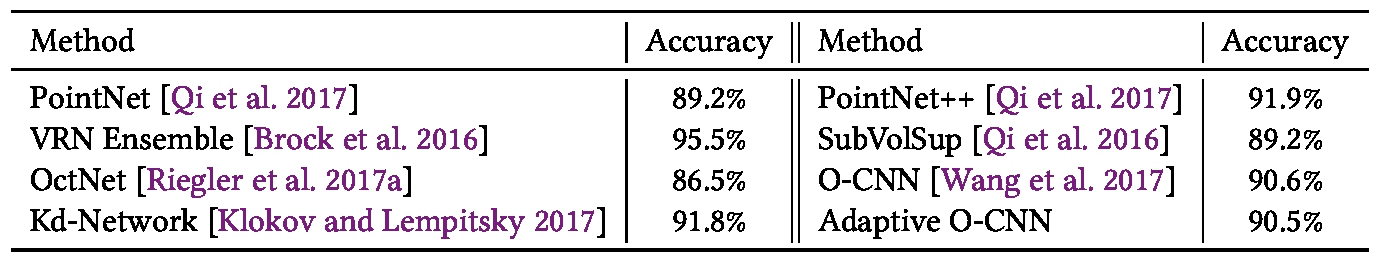
所有的测试都在Intel Core I7-6900K CPU（3.2GHz）和GeForce GTX Titan X GPU（12 GB内存）的台式计算机上完成。实验基于Caffe框架[10]。

5.1 3D形状分类

分类的数据集选择ModelNet40数据集[11]。分别使用32,64,128和256的分辨率进行测试。同时采用与O-CNN相同的配置，仅编码的格式不同。



这张表表示了O-CNN和AO-CNN在性能和准确率方面的差别。同时加入了基于体素的3D CNN的时间和存储器成本。

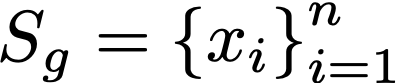
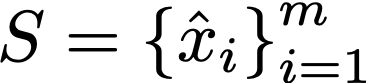


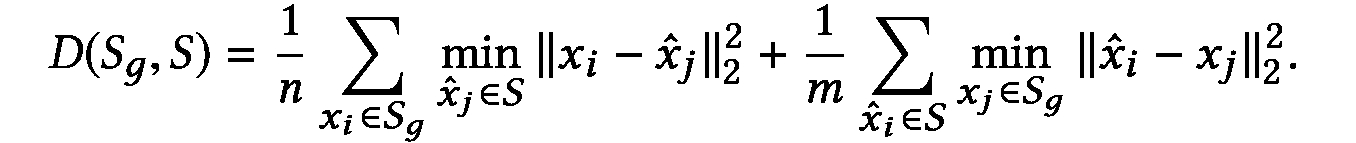
这张表显示不同方法在ModelNet40分类基准下，不同的学习方法的准确率。

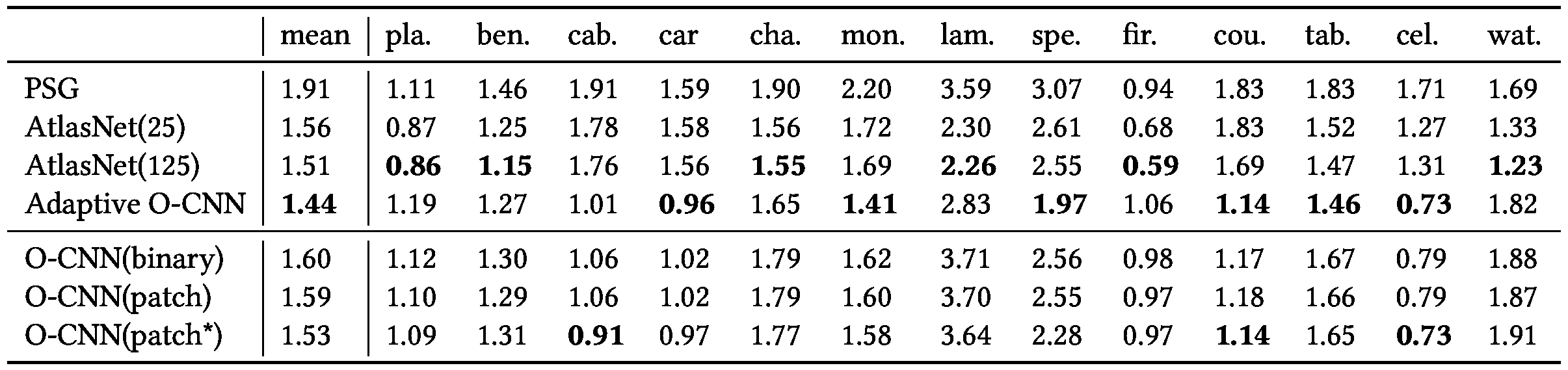
从实验结果看出，在所有分辨率下AO-CNN的准确率与O-CNN相当，且时间和存储的需求显著降低，尤其是在高分辨率下，效率达到了4倍以上。同时与现有学习方法相比，AO-CNN仍然具有较高的准确性。

但是，在128分辨率的情况下，AO-CNN的分辨率精度反而下降。作者通过对比从32到128的训练损失和测试损失得出结论：AO-CNN可能对ModelNet40造成了过拟合。这与O-CNN的情况相同。通过更多的数据可以减少过拟合。

5.2 3D自动编码

数据集采用ShapeNet Core v2 dataset[12]，八叉树最大深度为7，通过测量结果与真值之间的倒角距离作为质量指标。真值的点云表示为，预测的点云表示为，它们之间的倒角距离定义为：



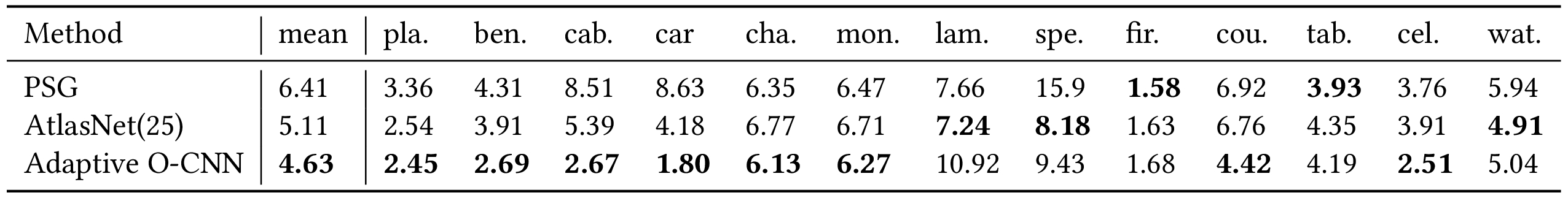


这张表显示了3D自动编码器实验的统计结果。

实验结果：与分别预测25和125网格补丁的两种类型的AtlasNets[13]相比，自适应O-CNN自动编码器平均可以达到最佳质量。并且两者表现均好于PSG[14]。

5.3 3D形状预测

数据集仍然使用ShapeNet Core v2 dataset。



实验结果：AO-CNN在13个类别的8个中实现了最佳性能，特别是对于具有较大平坦区域的物体。

6.结论

作者提出了一种用于3D编码和解码的新型自适应O-CNN，它的编码器和解码器利用了补丁引导的八叉树的优良特性：紧凑性，适应性和高质量近似。作者展示了AO-CNN的高效率，并且证明了它优于其他的方法。但是这种方法目前存在两个限制：

1. AO-CNN构成的八叉树之间的相邻补丁不是无缝的，解决的方法是通过其他表面修复或网格重建等技术。同时，作者观察到大多数接缝可以通过捕捉相邻贴片的附近顶点来缝合。因此作者准备开发一种后处理方法来缝合间隙。
2. AO-CNN中的平面补丁不能适应弯曲特征，如车轮等。作者认为可能可以添加非平面的补丁如二次曲面等。或者选用其他的拟合质量度量来引导八叉树的构造。

参考文献

1.Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 2015. 3D ShapeNets: A deep representation for volumetric shape modeling. In Computer Vision and Pattern Recognition (CVPR).

2.Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, and Joshua B. Tenen- baum. 2016. Learning a probabilistic latent space of object shapes via 3D generative- adversarial modeling. In Neural Information Processing Systems (NIPS).

3.H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller. 2015. Multi-view convolutional neural networks for 3D shape recognition. In International Conference on Computer Vision (ICCV).

4.Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. 2017. PointNet: Deep learning on point sets for 3D classification and segmentation. In Computer Vision and Pattern Recognition (CVPR).

5.Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. 2018. Pixel2Mesh: Generating 3D mesh models from single RGB images. In Computer Vision and Pattern Recognition (CVPR).

6.Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. 2017. O-CNN: Octree-based convolutional neural networks for 3D shape analysis. ACM Trans. Graph. (SIGGRAPH) 36, 4 (2017), 72:1–72:11.

7.Simon Fuhrmann and Michael Goesele. 2014. Floating scale surface reconstruction. ACM Trans. Graph. (SIGGRAPH) 33, 4 (2014), 46:1–46:11.

8.Tao Ju. 2004. Robust repair of polygonal models. ACM Trans. Graph. (SIGGRAPH) 23, 3 (2004), 888–895.

9.Jane Wilhelms and Allen Van Gelder. 1992. Octrees for faster isosurface generation. ACM Trans. Graph. 11, 3 (1992), 201–227.

10.Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: convolutional archi- tecture for fast feature embedding. In ACM Multimedia (ACMMM). 675–678.

11.Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 2015. 3D ShapeNets: A deep representation for volumetric shape modeling. In Computer Vision and Pattern Recognition (CVPR).

12.Angel X. Chang, Thomas Funkhouser, and etal. 2015. ShapeNet: an information-rich 3D model repository. arXiv:1512.03012 [cs.GR]. (2015).

13.Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C. Russell, and Mathieu Aubry. 2018. AtlasNet: A Papier-Mâché approach to learning 3D surface generation. In Computer Vision and Pattern Recognition (CVPR).

14.Hao Su, Haoqiang Fan, and Leonidas Guibas. 2017. A point set generation network for 3D object reconstruction from a single image. In Computer Vision and Pattern Recognition (CVPR).