東南大學

毕业设计(论文)报告

题 目： 基于深度学习的三维重建和点云生成

学 号： 04216747

姓 名： 周烨凡

学 院： 信息科学与工程学院

专 业： 信息工程

指导教师： 杨绿溪

起止日期： 2020 年 1 月-2020 年 5 月

东南大学毕业（设计）论文独创性声明

本人声明所呈交的毕业（设计）论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得东南大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

论文作者签名：_____ 日期：_____ 年 _____ 月 _____ 日

东南大学毕业（设计）论文使用授权声明

东南大学有权保留本人所送交毕业（设计）论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文。本人电子文档的内容和纸质论文的内容相一致。除在保密期内的保密论文外，允许论文被查阅和借阅，可以公布（包括刊登）论文的全部或部分内容。论文的公布（包括刊登）授权东南大学教务处办理。

论文作者签名：_____ 导师签名：_____

日期：_____ 年 _____ 月 _____ 日 日期：_____ 年 _____ 月 _____ 日

摘　要

单视角三维物体形状重建 (Single-view 3D Reconstruction) 是三维视觉领域一直以来的核心问题之一。由于拟合非线性方程和学习模式的有效性，深度神经网络被期望在重建三维非线性形状的任务中表现出色。本课题探究了深度神经网络在三维重建任务中的有效性和内在机制，主要的工作和创新如下：

1. 针对基于自编码器结构的深度神经网络多次训练结果差异较大的问题，提出采用多任务训练对编码器进行初始化，以及引入残差层增加解码器网络深度这两种方法优化网络架构。在 ShapeNet 公开数据集上进行测试，取得了优于基于检索的非学习方法 Oracle Nearest Neighbor 的结果。

2. 为了衡量三维数据集的聚类程度，提出了基于 affinity propagation 和 silhouette score 的度量标准，该方法得到的量化结果与数据集降维后的可视化结果吻合。

3. 提出了影响三维重建任务中深度神经网络内在机制的主要因素是数据集的本质特征：当训练集中的三维形状集有相比于图片集更高的聚类程度的结构时，在这个数据集上训练的深度神经网络更有可能会执行识别机制而非重建机制，并定性和定量的证明了这一强相关性。

最后，本文分析了上述结论对三维重建任务中的数据收集和神经网络训练的指导意义。

关键词：三维重建，点云，深度学习，数据挖掘

ABSTRACT

Single-view 3D reconstruction is one of the elementary tasks in the field of 3D vision. Due to the effectiveness in learning and approximating non-linear function, deep neural networks are expected to perform well on the task of reconstructing 3D nonlinear shapes. This paper investigates the effectiveness and internal mechanism of deep neural network in 3D reconstruction task. The main contribution and innovation are summarized as follows:

1. For the problem that the variance of multiple training results of deep neural network based on autoencoder structure is too large, it proposes to leverage multi-task training to initiate the encoder and introduce residual layer to deepen the decoder in order to optimize network architecture. These methods are trained and tested on ShapeNet public dataset and they outperform the non-learning method based on retrieve, namely Oracle Nearest Neighbor.

2. Define novel way to measure clustering coefficient of 3D reconstruction dataset based on affinity propagation and silhouette score, the quantitative results of this metric is corresponding to qualitative results of visualization of dataset in low dimension.

3. It claims that the bias of internal mechanism of network is mainly affected by the intrinsic properties of dataset: when the training set of 3D shapes has a more clustered structure than images, the deep neural networks trained on this dataset become more likely to perform recognition than reconstruction, and it proves the strong correlation between the deep neural networks and dataset property both qualitatively and quantitatively.

Finally, this paper analyzes the significance of the above conclusions for data collection and neural network training.

KEY WORDS: 3D Reconstruction, Point Cloud, Deep Learning, Data mining

目　录

第一章　绪论

1.1 课题的背景和意义

基于单视角二维图片输入进行三维物体形状重建是三维视觉领域的核心问题之一，它能有效解决现实世界中二维数据丰富而三维数据稀少的问题，满足无人驾驶，智能建造，机器人领域的需求。点云数据具有易处理，易存储，易获得的特点，因此成为重建任务中常用的三维数据表示之一。

使用深度学习来进行单视角三维重建持续受到人们的关注。尽管很多的文章已经展示了创新的深度学习框架来提高三维重建任务中的最高水平[1–11]，但很少有文章来尝试探究这些任务的本质特性。不可否认的是，重建三维非常规形状的问题已经变成了一个新的机器学习范例, 并且从事者使用的处理数据和训练网络的方法与在常规数据上的学习使用的方法不同（比如，Adam[12] 的使用频率比 SGD 要高）。因此，神经网络在三维重建学习这个新范例上怎样进行学习？是否与传统的向量分类和回归问题不同？

最近, [13] 的作者针对上述问题提出了一个让人惊讶的观点。他们尝试性的展示了当前最先进的用于三维重建任务的深度神经网络更倾向于通过首先分类输入图片到一个特定的簇，然后生成对应簇的平均三维形状，以此来进行预测。支持这一观点的主要实验证据是这些深度神经网络的三维预测结果与纯粹基于聚类和基于检索的基准模型的三维预测结果效果相近。这是一个非常有趣的观察，因为它说明了对于三维重建的任务来说，深度神经网络倾向于记住平均形状并且将其与图片输入的语义联系起来，而不是使用几何方法来生成一个形状，比如通过融合细粒度的局部结构形成一个整体形状。如果这个观点是正确的，这说明对于三维重建任务来说，最先进的深度神经网络实际执行了一个记忆任务而不是一个泛化任务[14]。

本课题旨在优化三维重建任务中深度神经网络的架构，证实其有效性。同时进一步探究深度神经网络在三维重建任务中的内在机制，分析导致其偏倚于识别或重建机制的因素，这一研究将会帮助当前应用于三维重建的深度学习方法避免陷入识别机制的误区，同时也探讨了深度学习的本质问题，即如何帮助神经网络进行更好的泛化。

1.2 研究现状

1.2.1 相关模型

[5] 首次提出了以单张图片作为输入的三维点云重建任务中深度神经网络的基本架构, 即自编码器架构。编码器由卷积层和 ReLU 层构成，编码器的输入是一张图片和一个向量，向量用来模拟重建任务中的不确定性。解码器由全连接层构成，解码器输出点云的坐标，用 N × 3 的矩阵表示，N 为一个点云中三维点的数量。[22] 提出了以三维点云作为输入的三维点云重建任务中深度自编码器的架构，创造性的提出用折叠的思想来生成点云，并重建三维形状，根据该思想实现的解码器仅使用了全连接层解码器的 7% 的参数量，却在重建效果上超过当

时的基准模型。[2] 中评估了多个效果拔尖的单视角三维重建深度神经网路模型，并提出了基于识别和检索机制的多个非深度学习方法模型，且这些模型的重建性能超过了当前的深度学习网络模型，这使得深度学习在单视角三维重建中的有效性受到了置疑。

1.2.2 相关数据集

ShapeNet[27] 是一个注释丰富且规模较大的三维形状数据集，涵盖 55 个常见的类别，有大约 5 万个样本，每个样本内有一个三维模型和多张从不同视角渲染的该三维模型的图片，三维模型的数据格式为网格、体素，图片格式为 PNG，在进行单视角三维重建时会从多个视角图片中选择一个视角的图片，因此这种情况下每个样本内有一张图片与一个三维模型。该数据集的创立者在 ICCV 2017 举办了基于该数据集的单视角三维重建任务的竞赛，并将当时的基准成绩发表在 [28]。[2] 提供的数据集在 ShapeNet [27] 的基础上增加了点云数据格式，每个三维形状由 9000 多个点构成。共有 52430 个样本，涵盖 55 个类。

1.2.3 神经网络学习机制

深度神经网络是执行记忆还是泛化一直是现代机器学习中的主要问题。与我们将要介绍的类似，众所周知的猜测是优化过程是‘内容感知’的，并且取决于数据本身的属性 [2]。[2] 中还显示，训练期间的某些正则化技术可帮助深度神经网络泛化而不是记住任务。对于三维形状重建，[26] 表明神经网络倾向于记忆平均形状，而不是在几何意义上进行重建。确实，许多作品还显示了平均形状和识别信息在提高三维重建效果中的有效性 [10,20,13]。

相比之下，也有很多作品利用三维形状的连续潜在空间中的分布信息 [15,8,21,14,34,36,37] 来提高三维重建效果，这超出了基于识别的范围。值得注意的是，[33,6] 表明形状算术可以在三维形状的潜在空间中进行，从而排除了神经网络仅在此问题设置中执行识别的可能性（因为执行算术需要的不是均值信息离散簇的形状）。其他一些工作建议通过将每个形状分解为部分 [27,16,24] 或通过连续过程 [23,4] 来生成三维形状，这也超出了简单的识别任务。但是, [33,1,38] 的作者将基于三维重建的自动编码视为点云上无监督分类的基础。尽管这些作品中的输入数据是三维形状，而不是二维图像，但是形状信息有助于分类的事实似乎确实增强了这样的概念，即三维重建更着重于识别而不是重建。尽管如此，形状信息有助于识别的事实不能成为判断三维重建网络所学知识的主要理由。有意义的未来方向是研究分别用于无监督分类和受监督三维形状重建的神经网络学习之间的差异，因为这两个方向的主要目标并不完全相同。

注意三维重建问题可以被看作是一个更普适的分布学习 [17,19] 的特殊情况。但是，与分布学习中的理论工作如拓展核方法到回归分布不同，我们的工作集中于深度学习。即便如此，使人感兴趣的是能看到分布学习的传统工作把输出分布当做一个所有训练分布样本的连续线性组合直接处理，而不是使用两步法，预测簇索引后再预测平均分布。

1.3 本文研究内容

1.3.1 课题关键问题以及难点

首先，正式定义本课题的关键问题：一、优化深度神经网络框架，使其在单视角三维重建任务中超过基于识别机制的非深度学习方法，证明其有效性。二、讨论深度神经网络在单视角三维重建任务中执行的是识别机制还是重建机制。初步猜想是影响其在两者之间偏倚的因素为训练数据集的整体特征：聚类程度。因此本课题需要考虑以下几个难点：

1. 改进当前作为基准的神经网络架构以获得更好的重建性能。我们希望能对当前作为基准线的深度神经网络进行网络结构优化，以期望其在公认的标准数据集上能接近并超越基于识别机制的非深度学习方法。因此我们考虑借鉴图像领域成熟且有效的神经网络架构优化方法和训练技巧。

2. 定义神经网络的识别与重建这两种机制的数学表达，并用具体的实验结果来描述。

3. 设计并产生具有量化特征的三维重建数据集。为了探究神经网络训练集和网络的性能之间是否有强相关性，需要能定量的操控数据集的某些整体特征，如聚类系数。可以考虑的方式就是生成自定义的数据集，同时在生成过程中通过采样改变整体特征。为此可以考虑使用计算机图形学的相关软件来合成三维模型，并探究一些能进行不同三维形状之间插值的算法。

4. 定义衡量数据集指标的度量标准。在解决问题三后，需要构建一个度量标准来衡量数据集的聚类趋势，得到量化评分。

1.3.2 主要贡献

在这项工作中，我们定量和定性的展示了 [26] 的结论不是通用的，而且这是一个不合适的数据集收集和不合理的数据使用导致的复杂结果。我们建议使用数据挖掘中确立已久的理论测量一个数据集的“聚类趋势”，称为 affinity propagation [30] 和 silhouette score [29], 并且我们展示了当用于训练三维重建任务的数据集不具有聚类特征时，训练结束的神经网络可以学会更集中在重建而不是识别。更重要的是，我们展示了即使是现实数据集如 ShapeNet[3], 在其上进行训练的神经网络依然呈现出重建机制，且通过优化网络架构和使用训练技巧，其重建效果接近并超过了基于检索的非深度学习方法。我们关于识别与重建这两种机制的理解更丰富，将三维重建的内在机制与数据以及训练过程相关联。之更具体的是，我们展示了训练集中的三维形状之间应该比训练集中的图片间呈现出更多的差异，以此来避免产生一个基于识别的机器学习模型。这一结论很实用，因为它为三维数据集的收集以及三维重建训练提供了指导。

1.4 论文组织结构

本论文主要有六个章节，各章的内容安排如下:

第一章 绪论，简要介绍了单视角三维重建任务的研究背景和意义，并对现行的相关理论，模型和数据集进行简单概述与分析。介绍本论文的研究思路、关键问题以及主要贡献。

第二章 问题与定义，介绍了本论文探究的损失函数和算法的理论基础。给出了重建机制与识别机制的数学定义。

第三章 模型，详细介绍了基准模型，以及优化模型的架构，并给出了网络在 ShapeNet 大型数据集上训练的过程信息以及训练结果的分析比较。

第四章 机制分析，给出了机制探究的理论基础与实验设计，介绍了自定义数据集的制作，网络在自定义数据集上训练的过程信息以及训练结果的定性与定量分析。

第五章 总结与展望，分析并总结本课题研究成果的总体优缺点，并提出未来的研究改进方向。

第二章　问题与定义

2.1 问题描述

我们探究的问题是从单张图片重建一个三维形状。输入 I 是一张二维 RGB 图片。输出 S 由一个点云表示。我们在本课题中不考虑基于体素的体积表达。对于基于点云的表达，每个形状 S 是一个包含三维点的点集。一个神经网络在训练中通过减小特定损失函数定义的经验损失 l，以此来从输入图片 l 预测形状 S。

$$\min_{f} \sum_{i=0}^{n-1} l(f(I_i), S_i). \tag{2.1}$$

我们想通过优化 f，使其在训练数据集上训练后，在测试数据集上取得更低的平均损失。同时，我们还想学习 f 的特性来了解它执行的是一个重建任务还是识别任务。两个指标被用来测量重建结果 $\widehat{S} = f(I)$ 与标签点云 S 之间的差异，称为 Chamfer Distance 和 F-score。

Chamfer distance. Chamfer Distane 通过搜索另一个点集中最近点来测试一个点集到另一个点集的整体距离。

$$d_{CH}(S, \widehat{S}) = \frac{1}{|S|} \sum_{\mathbf{x} \in S} \min_{\widehat{\mathbf{x}} \in \widehat{S}} \|\mathbf{x} - \widehat{\mathbf{x}}\|_2 + \frac{1}{|\widehat{S}|} \sum_{\widehat{\mathbf{x}} \in \widehat{S}} \min_{\mathbf{x} \in S} \|\widehat{\mathbf{x}} - \mathbf{x}\|_2. \tag{2.2}$$

尽管 Chamfer Distance 是运算高效且方便的，它会被很小一部分异常点的强烈影响。所以我们也要如 [26] 建议的那样采用 F-score 来预测点集。

F-score. 另外一个衡量形状重建的方法是 F-score, 该指标是精确度和回忆度的调和平均数。在给定相应的标签 S，且在固定的距离阈值 d 内，重建点云 $\widehat{S}$ 的精确度 *Prec* 为定义为：

$$\mathit{Prec}(d, S, \widehat{S}) = \frac{1}{|\widehat{S}|} \sum_{r \in \widehat{S}} [\![\min_{s \in S} ||r - s|| < d]\!], \tag{2.3}$$

where $[\![\cdot]\!]$ is the Iverson bracket.

相似的，标签点云 S 对重建点云 $\widehat{S}$ 的回忆度 *Rec* 是：

$$\mathit{Rec}(d, S, \widehat{S}) = \frac{1}{|S|} \sum_{s \in S} [\![\min_{r \in \widehat{S}} ||s - r|| < d]\!]. \tag{2.4}$$

使用这两个量，如下计算 F-score：

$$F(d, S, \widehat{S}) = \frac{2 \times \mathit{Prec}(d, S, \widehat{S}) \times \mathit{Rec}(d, S, \widehat{S})}{\mathit{Prec}(d, S, \widehat{S}) + \mathit{Rec}(d, S, \widehat{S})}. \tag{2.5}$$

重建的准确度被精确度量化，用以测量重建的点集离标签点集有多近。重建的完整度被回忆度量化，用来测量标签点云多少部分被重建点云覆盖。所以，一个高的 F-score 显示了重建是准确且完整的 [12]。

2.2 识别或重建

在这个部分，我们正式定义本文中研究的两种学习范式，被称为识别（*recognition*）和重建（*reconstruction*）

定义 1 *(识别) 一个基于识别的神经网络 f 用两步预测形状重建。神经网络重建方程可以写成：*

$$\widehat{S} = f(I) = f_1(f_2(I)), \tag{2.6}$$

方程中的 $f_2(\cdot)$ 将输入图片映射到一个标量索引，并且 $f_1(f_2(I))$ 将这个标量索引映射到索引 $f_2(I)$ 对应的特定簇的平均形状。

定义 2 *(重建) 一个基于重建的神经网络直接进行三维重建。即*

$$\widehat{S} = f(I), \tag{2.7}$$

并且重建不会明显的使用图片簇的任何信息。因为当前最流行的单视角三维重建经常使用一个编码器-解码器结构，所以一个相似的概念是由编码器获得的码字不会形成簇。

主要问题

在这篇论文中，我们研究神经网络执行识别机制还是重建机制。我们展示了向两者任一偏倚的趋势由数据集特性决定。

注意 [26] 的主要结论是单视角三维重建中的深度神经网络主要执行识别工作。换句话说，神经网络的方程更接近于定义1而不是定义2。

2.3 数据集的聚类趋势

在这个部分，我们定义用来测量数据集聚类趋势的指标。具体来说，我们使用 silhouette score 来

参考文献

[1] Li C L, Zaheer M, Zhang Y, et al. Point cloud gan[J]. arXiv preprint arXiv:1810.05795, 2018..

[2] Park J J, Florence P, Straub J, et al. Deepsdf: Learning continuous signed distance functions for shape representation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019. 165–174.

[3] Fan H, Su H, Guibas L J. A point set generation network for 3d object reconstruction from a single image. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. 605–613.

[4] Tatarchenko M, Dosovitskiy A, Brox T. Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs. Proceedings of the IEEE International Conference on Computer Vision, 2017. 2088–2096.

[5] Groueix T, Fisher M, Kim V G, et al. A papier-mâché approach to learning 3d surface generation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. 216–224.

[6] Yang Y, Feng C, Shen Y, et al. Foldingnet: Point cloud auto-encoder via deep grid deformation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018. 206–215.

[7] Wang N, Zhang Y, Li Z, et al. Pixel2mesh: Generating 3d mesh models from single rgb images. Proceedings of the European Conference on Computer Vision (ECCV), 2018. 52–67.

[8] Sun X, Wu J, Zhang X, et al. Pix3d: Dataset and methods for single-image 3d shape modeling. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018. 2974–2983.

[9] Tulsiani S, Zhou T, Efros A A, et al. Multi-view supervision for single-view reconstruction via differentiable ray consistency. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. 2626–2634.

[10] Wu J, Wang Y, Xue T, et al. Marrnet: 3d shape reconstruction via 2.5 d sketches. Advances in neural information processing systems, 2017. 540–550.

[11] Yan X, Yang J, Yumer E, et al. Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision. Advances in neural information processing systems, 2016. 1696–1704.

[12] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014..

[13] Tatarchenko M, Richter S R, Ranftl R, et al. What do single-view 3d reconstruction networks learn? Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019. 3405–3414.

[14] Arpit D, Jastrzębski S, Ballas N, et al. A closer look at memorization in deep networks. Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017. 233–242.

附录 A LaTeX 实验

技术实验结果在这里写

附录 B　MATLAB 实验

技术实验结果在这里写

致　　谢

这次的毕业论文设计总结是在我的指导老师 xxx 老师亲切关怀和悉心指导下完成的。从毕业设计选题到设计完成，x 老师给予了我耐心指导与细心关怀，有了莫老师耐心指导与细心关怀我才不会在设计的过程中迷失方向，失去前进动力。x 老师有严肃的科学态度，严谨的治学精神和精益求精的工作作风，这些都是我所需要学习的，感谢 x 老师给予了我这样一个学习机会，谢谢!

感谢与我并肩作战的舍友与同学们，感谢关心我支持我的朋友们，感谢学校领导、老师们，感谢你们给予我的帮助与关怀; 感谢肇庆学院，特别感谢计算机科学与软件学院四年来为我提供的良好学习环境，谢谢!