中国科学技术大学电子工程与信息科学系

研究生学位论文工作检查报告

论文题目 基于深度学习的从单张图像生成三

维形状的方法研究

学生姓名 胡思宇

学生学号 BA15006031

指导教师 吴枫、陈雪锦

所在院系 电子工程与信息科学系

学科专业 信息与通信工程

研究方向 计算机图形学、计算机视觉

填表日期 2019.02.27

中国科学技术大学电子工程与信息科学系

二零一一年九月制表

一．开题内容

|  |
| --- |
| 简要介绍学位论文开题确定的选题、研究内容以及研究方法与思路。  **1.选题背景**  **1.1研究意义**  从单张图像恢复三维形状是计算机视觉中的经典问题。该问题极具挑战性，其难度主要在于视觉系统中三维物体投影到二维形成图像的过程是不可逆过程。这使得该问题本身是欠定的，为了实现对该问题的求解往往需要利用关于视觉成像的先验知识。如何有用数学表达相关先验知识并有效的应用到形状的优化求解中是解决该问题的关键之一。  近年来，有越来越多的研究工作开始在深度学习的框架下对这一问题进行求解。使用深度学习方法的优势在于无须精心推导视觉先验的数学表达而是利用深度学习方法从大量的数据集中学习重建形状所需要的先验知识。[Choy--ECCV--2016; Fan--CVPR--2017; Tatarchenko--ICCV--2017]等一系列工作展示了深度学习方法在这一问题上的优势。它超越传统方法，不但对图像所示的可观测到的部分进行推断，而且对不可见的部分能进行合理补全。能达到这样的效果主要归功于深度神经网络能对大数据集进行统计学习从而提取先验知识的能力。  综上所述，本文所选择的基于深度学习的从单张图像生成三维形状的方法研究是目前解决单图像三维建模问题最具前景的研究方向之一具有重要意义。  **1.2相关工作**  利用深度学习从单张图像恢复三维形状的研究近年来得到了蓬勃的发展，从中可以归纳出若干不同的主流技术路线。这些并行发展的技术路线主要基于不同的三维形状表达方式，这使得它们有各自的优势也有各自急需解决的问题，以下分别介绍：  **基于体素：**从[Wu--CVPR--2015]开始，基于体素的形状表达方式成为了该问题，以及其它深度学习问题中对于三维形状最常见的表达方式。这种表达方式的一大优势在于可以自然地将在二维栅格（图象）上的卷积操作直观的扩展到三维中，相应的也可以将二维栅格应用中的网络设计经验、训练策略、损失函数等迁移扩展到三维中。以[Choy--ECCV--2016]为代表的一系列工作在体素表达下对从图像到三维形状的深度神经网络设计问题展开了研究。体素表达的主要缺陷在于它对于存储的消耗巨大，这限制了它所能表达的形状的精细程度，在一段时间里体素表达方式的相关工作大多只能输出的体素格来表达形状，这样的三维分辨率极大的限制了其形状表达能力。于是不少研究者尝试在体素表达的基础上利用八叉树这种数据结构来使有限的存储空间更多的被用于表达物体表面的形状而不是被浪费在填充实心物体内部。[Tatarchenko--ICCV--2017]率先将这一构想运用于从图像生成形状的网络中将输出的体素网格分辨率提高到。  **基于点集：**体素表达不但耗费较多存储空间，而且不利于网络自动的学习诸如旋转、平移之类的三维变换。因为这些变换在体素表达下需要置换体素格，而这种置换操作不是简单的连续可微的函数，而在三维点集的表达下，旋转、平移都是线性函数，更加易于被现有深度网络学习。因而[Fan--CVPR--2017]倡导通过网络直接回归无序点集的方式来实现从图像生成形状。想要回归无序点集，就必须有一种衡量两个无序点集的形状差异的可微的损失函数。[Fan--CVPR--2017]提出了使用Chamfer Distance Loss (CD) 通过双向统计两个无序点集中最近点对之间距离的方式来衡量两个无序点集之间的形状差异。基于点集表达的问题之一也源于这种loss的缺陷，由于Chamfer Distance Loss在驱动输出形状逼近Ground Truth时只考虑以最近邻的原则建立三维点之间的对应关系，它没有考虑常见物体形状表面连续的先验条件，这导致很多物体的输出形状表面点集杂乱，因而所表达的形状表面模糊有明显瑕疵。  **基于网格：**在现代工业系统中三维网格是对三维形状最常见的表达之一，它被广泛的运用于影视动画、工业设计、有限元分析等等领域中。相比于点集的表达，网格的表达不但通过表面顶点的坐标来记录形状而且包含了顶点的局部连接关系，依照这些连接关系可以将顶点连接成面片来逼近完整的连续表面。[Dou--CVPR--2017]与[Pontes--ACCV--2018]分别提出了不同的可以从图像输出三维网格的深度神经网络。其中[Dou--CVPR--2017]通过在人脸网格数据集中先统计学习出一个人脸网格的双线性模型，然后通过神经网络拟合线性系数的方法来实现从由神经网络从图像重建三维人脸的过程。 [Pontes--ACCV--2018]则将由图像重建三维网格的问题拆分为两个子问题由神经网络分别求解。 换言之，[Pontes--ACCV--2018]先通过一个网络学习从输入图像拟合线性插值系数然对数据集中的若干个模板形状进行差值，然后通过从图像拟合一个Free-Form-Deformation层来进一步对三维形状做变形丰富形状细节。[Dou--CVPR--2017]与[Pontes--ACCV--2018]的问题在于只能针对特定类型的物体进行模型的学习，其中[Dou--CVPR--2017]只专注于人脸形状，而[Pontes--ACCV--2018b]则需要针对每一类物体单独训练网络模型。  **2.研究内容**  综合考虑现有主流技术路线的发展，应用前景和各自的优劣，基于网格的的表达有更广泛的应用前景，因此，文本将专注基于三维网格表达的单图像形状生成网络。  本文首先要研究的问题是如何实现通用的端到端可训练的网络，使之同时学习从多种物体图像重建三维网格形状。要实现一个端到端可训练的可处理多类物体的通用网格生成网络，一种思路是在现有有序回归网络的基础上通过回归网格顶点位置来实现。但是，这需要在多种物体的网格之间建立一致的有序对应关系，由于多类物体之间形状各异，需要复杂的数据处理才有可能实现。另一种思路是在现有无序点集回归网络[Fan--CVPR--2017]的基础上实现连续可导的网格建立过程。但是现有的三角化建立网格的过程难以被转化为连续可导函数。本文对此提出使用基于表面参数化的表达。  其次，在解决前一问题的基础上，本文将研究如何控制网格表面质量，以达到避免网格表面产生自交、尖刺等常见缺陷，使所生成的物体表面连续美观的目的。本文针对其中表面自交问题的控制提出使用循环正则的方法。  第三，在解决之前两个问题的基础上，如何进一步提升三维网格生成网络的所生成的三维形状的细节，提高准确率，是我们要研究的第三个问题。针对这个问题本文引入基于混合变形的形状表达方法。  图 1研究内容与技术路线对照图  **3.研究方法与技术路线**  针对前述的研究内容我们制定了如图 1所示的技术路线接下来我们将分别介绍：  **3.1基于表面参数化的端到端可训练的通用三维表面网格生成网络：**  C:\Users\SamHu\Desktop\overview_redraw.png  图 2 三维表面网格生成网络  受到计算机图形学中的经典问题，参数化问题的启发。我们发现有很多物体的表面可以建立与预定表面之间的一一映射关系，寻找这样的一一映射的过程便被称作求解参数化。而这样的映射函数本身是连续函数是可以由神经网络层来逼近的。  由此，本文提出了如所示的网络结构。在这个网络结构中，我们从输入图像提取特征来预测一个从球面到输出物体表面的映射关系。这个映射关系本身也是用网络来拟合逼近的。  因此总体而言我们的网络可以分为如图 1所示的两个部分，其中参数化网络实现拟合逼近从球面到输出表面的映射关系。而语义网络实现从输入图像提取特征的过程。该网络通过语义网络的输出作为参数化网络的网络参数的方式将两者联系起来，从而实现由网络预测复杂的表面参数化映射的过程。  更具体来说，本文所设计的参数化为了能够处理无序球面点集到目标的映射采用了如图 2所示的网络结构。这种结构是基于[Charles--CVPR--2017]所提出的方法扩展而来的。[Charles--CVPR--2017]主要是被应用于三维形状的分析，主要完成语义分割任务，它的特点是可以从无序点集中提取特征。我们将它更改为类似卷积的形式，即以共享的网络参数从每个点的K邻域中提取特征，并预测新的三维顶点位置，以实现对无序点集的变形。  图 3 *K*-neighbor PointNet 的内部结构  而语义网络的细节结构以卷积网络为主。是在[Fan--CVPR--2017]的hourglass网络结构的基础上调整不同输出层的宽度以匹配参数化网络来实现的。  这样的网络结构能够一定程度上实现本文的研究目标，即实现一个端到端可训练的通用三维网格生成网络，它能够以不同类的物体的图片-三维形状对为数据进行训练。  我们的研究主要集中于通过实验探寻这一网络结构的合理训练流程和配套的损失函数设计等。  3.2**通用三维网格生成网络中基于循环正则的表面防自交方法：**  bitmap现有的三维表面网格存在严重的表面自交问题。这样的问题不仅出现在本文所提出网格生成网络中，也出现在其它的使用类似思路的通用三维网格生成网络（[Groueix--CVPR--2018]与[Wang--ECCV--2018]）中。本文认为引发这个问题的原因与引发点集生成网络[Fan--CVPR--2017]中表面点集杂乱问题的原因是一样的，是由于使用了Chamfer Distance Loss作为损失函数来衡量输出形状与Ground Truth之间差异的度量。  图 4 输出网格的表面自交问题: (a) [Groueix--CVPR-2018]的结果。(b) [Wang--ECCV--2018]的结果  Chamfer Distance Loss 如公式 1所示，它在点集与之间双向搜索最近点作为匹配点对来统计衡量两者之间的形状差异，通过最小化这一损失函数可以驱动输出点集逼近Ground Truth的点集，但是以最近点为匹配对的原则很难找到全局最优的形状对应，再加上这一损失函数完全没有考虑物体表面的连续性质所以很容易使得点集表面杂乱而网格表面自交。在点集网络[Fan--CVPR--2017]的网络框架下我们没有找到合适的适合于深度学习的方法来解决这一问题。  公式 1 Chamfer Distance Loss  然而在通用三维网格生成网络中我们可以从另一个角度来解读这一问题的成因，因为在这样的网络框架中输出顶点集被表达为将预定表面映射到输出表面的过程。而发生了表面自交就意味着预定表面中存在的不同点在中的像点发生了重合。  公式 2表面自交的数学表达  **定义1（单射函数的定义）：**定义域为的函数f为单射如果满足：  **定理1（单射函数的判定定理）：**如果函数f的左逆函数存在则f为单射。换言之对于函数 ，存在函数,使得：  根据**定义1**我们可知只需要保证*f*为单射函数，则我们可以保证公式 2所表述的自交情况不会发生。根据**定理1**我们可知道只需要保证左逆函数的存在那么我们就可以使得*f*为单射。因此我们设计出如下正则项来尽量实现这一目标：  公式 3循环正则项  E:\WorkSpace\ParamNet\ParamNet\doc\GMP2019\img\net\bitmap.png如图 4所示，示意了循环正则项在网格生成网络中的作用。如图所示，通用网格生成网络会将预定表面的顶点(红点)通过拟合映射函数*f*来映射到输出形状表面的顶点(绿点)，而我们的循环正则方法需要同时拟合一个逆向映射将输出表面的顶点又映射回到预定表面的顶点（蓝点）。循环正则通过最小化循环映射后的顶点距离（蓝点与红点的距离）来使得函数*f*在Chamfer Distance Loss的引导下逼近Ground Truth 的同时尽量保证存在一个左逆映射，从而能够保证*f*的单射性质。  图 5循环正则示意图  严格来说似乎我们需要在协同优化中将循环正则项优化到逼近0才有可能保证*f*是单射的，然而根据经验可知在神经网络的训练中是几乎不可能收敛到这一状态的。于是我们通过**命题1**来说明即使不能将循环正则优化到0，最小化该循环正则项依然有利于保证*f*的单射性质，只不过*f*所对应的左逆函数并不是刚好是我们逆解码器所拟合的函数。在**命题1**所例举的条件下需要在的基础上再复合一个最近邻函数才能得到*f*的左逆函数。  **命题1**对于欧式空间的子集与，如果对于给定函数 ，如果存在函数,使得：  那么*f*是一个单射函数。  **命题1**可以很容易的通过构造法得到证明，我们可以构造一个最近邻函数:它对任意欧式空间中的顶点找到在集合中的最近邻作为像点：  那么根据**命题1**的条件可以  所以是*f*的左逆函数，因此*f*是单射函数。  综上所述，我们提出的循环正则方法来源于基本的单射判定定理，数学形式简单通用，理应能够帮助现有的通用三维网格生成网络减少输出形状的表面自交，我们将通过实验验证这一方法在嵌入到现有网络中的实际使用效果。  **3.3基于混合变形的通用三维网格生成网络：**  混合变形（blend shape）方法是图形学中的常用方法，它常见于三维游戏中的人物编辑， |
| **参考文献**  [Charles--CVPR--2017] Su, H, Kaichun, M, and Guibas, LJ PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *CVPR* pages 77–85, 2017  [Choy--ECCV--2016] Xu, Danfei, Gwak, Jun Young, Chen, Kevin, and Savarese, Silvio 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction. In *ECCV* pages 628–644, 2016  [Dou--CVPR--2017] Shah, SK and Kakadiaris, IA End-to-End 3D Face Reconstruction with Deep Neural Networks. In *CVPR* pages 1503–1512, 2017  [Fan--CVPR--2017] Su, Hao and Guibas, Leonidas J A Point Set Generation Network for 3D Object Reconstruction from a Single Image. In *CVPR* pages 605–613, 2017  [Groueix--CVPR--2018] Fisher, Matthew, Kim, Vladimir G, Russell, Bryan C, and Aubry, Mathieu A Papier-Mâché Approach to Learning 3D Surface Generation. In *CVPR*, 2018  [Pontes--ACCV--2018a] Kong, Chen, Sridharan, Sridha, Lucey, Simon, Eriksson, Anders P, and Fookes, Clinton B Image2Mesh: A learning framework for single image 3D reconstruction. In *ACCV*, 2018  [Pontes--ACCV--2018b] Kong, Chen, Sridharan, Sridha, Lucey, Simon, Eriksson, Anders P, and Fookes, Clinton B Image2Mesh: A learning framework for single image 3D reconstruction. In *ACCV*, 2018  [Tatarchenko--ICCV--2017] Dosovitskiy, A and Brox, T Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs. In *ICCV* pages 2107–2115, 2017  [Wang--2018] Zhang, Yinda, Li, Zhuwen, Fu, Yanwei, Liu, Wei, and Jiang, Yu-Gang Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images. , 2018  [Wu--CVPR--2015] Song, S, Khosla, A, Yu, Fisher, Zhang, Linguang, Tang, Xiaoou, and Xiao, J 3D ShapeNets: A deep representation for volumetric shapes. In *CVPR* pages 1912–1920, 2015 |

二．学位论文工作完成情况

|  |
| --- |
| 简介已完成和待完成的学位论文工作，若存在与开题内容不同之处，应给予详细说明。   1. **已完成的工作**   **1.1 基于表面参数化的端到端可训练的通用三维表面网格生成网络：**  本文所提出的基于表面参数化C:\Users\SamHu\Desktop\res.png的方法无论从视觉还是数值上都超过了当时最优秀的[Fan--CVPR--2017]  图 6本文与[Fan—CVPR--2017]在视觉效果上的比较  C:\Users\SamHu\Desktop\abl.png  图 7本文方法的各模块功能对比实验  **2.待完成的工作**  **C:\Users\SamHu\Desktop\real.png**  图 8本文方法运用于真实图像（基于渲染图像所训练的网络模型）  1.2**通用三维网格生成网络中基于循环正则的表面防自交方法：**  已经实现将本文所提出的循环正则的方法应用于现有的通用三维网格生成网络，并验证了效果。 |

三．已取得的与论文研究内容相关的成果

|  |
| --- |
| 已发表或被接收发表的文章目录或其它相关研究成果。  Hu, Siyu, Xuejin Chen, and Xin Tong. "Point sets joint registration and co-segmentation." *The Visual Computer* (2018): 1-13.  “Preventing Self-intersection with Cycle Regularization in Mesh Reconstruction Networks” 投稿到[*Computer Aided Geometric Design*](http://www.baidu.com/link?url=F1XUD5K-p0kOOQTzb1jG2nmdVvlPtJeBGK6XYwUrARVmWEyd2bic2Xmdch4pDEcB3GDeQpkwL2VRAlc7N5mdRjW4VZNF5v3kBx7-kjvuY9_&wd=&eqid=eb6c681600014f24000000035c77cc4a) 目前状态accepted with revision  “Blended AtlasNet for Mesh Generation and Co-segmentation” |

四．下一步工作进度安排

|  |
| --- |
| 至论文答辩前的学位论文工作安排  4月1日前完成””的修改稿。  4月30日前完成所提方法和现有方法在语义多样性方面的的对比实验 。  5月15日前完成论文撰写准备投稿，同时撰写毕业论文 。  5月底前完成毕业论文的撰写 |

导师签名： 研究生本人签名：

年 月 日