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研究的意义**  从单张图像恢复三维形状是计算机视觉中的经典问题，由来已久。传统的有从阴影恢复形状与从纹理恢复形状的方法。以从阴影恢复形状的方法为例，早在博士论文[HORN--Psychol Comput Vis--1975]中就提出了从阴影恢复形状的方法。几十年来无数的方法研究者投入其中[Zhang--PAMI--1999]，针对这一问题在方法上做持续的提高与改进。  这一问题本身极具挑战性，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三维形状成像的先验知识。如何有效地用数学表达相关先验知识是解决该类方法的关键之一。  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--CVPR--2017]等一系列工作展示了深度学习方法在解决这一问题上的优势。它能够超越传统方法，不但能够对图像所示的可观测到的部分的形状进行推断，而且能够对不可见的部分进行推断，这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。  **1.2研究针对的主要问题**  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：  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中，以及其它深度学习为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--CVPR--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：[Fan--CVPR--2017]提出体素表达不但耗费较多存储空间，而且这种表达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，而相对应的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的深度网络所学习。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的深度神经网络。想要回归无序点集，就必须有一种能够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--2018]则需要针对每一类物体单独训练网络模型。  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，我们认为基于网格的表达具有更广泛的应用前景，因此，我们的研究专注于三维网格的表达方式。  我们首先要解决的问题就是怎样设计一个通用的端到端可训练的网络，使之能够同时学习从多种物体图像重建三维网格。其次，对于通用三维网格生成网络而言，在衡量输出形状与Ground Truth的我们无法事先建立形状中点到点的对应关系，现有可行方案中只能按照无序点集来处理，也就是使用Chamfer Distance Loss来衡量形状之间的差异。那么如何利用网格的局部拓扑关系来避免所生成的网格出现表面顶点杂乱达到生成表面连续视觉上美观的物体形状的目的，便是我们要解决的第二个问题。第三个方面，在解决之前两个问题的基础上，我们希望着手进一步提升三维网格生成网络的准确率和应用范围。  **2.研究内容****与技术路线**  2.1**端到端可训练的通用三维表面网格生成网络：**  overview_redraw要实现一个端到端可训练的通用网格生成网络的主要难点在于现有神经网络技术一般只能拟合连续函数，但是传统技术中三角化建立网格的过程难以被表达为一个连续可导的函数。而要使得网络能够同时学习多种物体的形状生成，意味着我们难以为Ground Truth形状的顶点建立有序的对应，所以我们提出的方案需要能够为无序点集建立网格而不能像[Dou--CVPR--2017]与[Pontes--ACCV--2018]两篇文章一样通过有序的点集对应将数据集中的网格连接关系迁移到输出中。  图 1 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。  图 2 *K*-neighbor PointNet 的内部结构  由此，我们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parameterization network实现拟合逼近从球面到输出表面的映射关系。而semantic network 实现从输入图像提取特征的过程。我们通过semantic network输出作为parameterization network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映射的过程。  更具体来说，我们所设计的parameterization network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。  而semantic network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parameterization network的结构。  这样的网络结构能够一定程度上实现我们的目标，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。  我们的研究内容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。  2.2**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我们发现在我们提出的网格生成网络输出的三维表面网格存在严重的表面自交问题。同时其它使用类似思路的通用三维网格生成网络[Groueix--CVPR--2018]与[Wang--ECCV--2018]也存在相同的问题。我们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 3 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式来解决这一问题。  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 4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。  2.3**带有混合变形的三维网格生成网络：**  针对新的三维点集的表达方式设计新的有足够表达能力的网络结构来实现从单张彩色图像生成三维形状的网络结构设计。网络结构设计主要解决的问题是如何使网络的输出能够表达一种空间分布。从Variational AutoEncoder借鉴的思想是通过预测一个从预定的分布到目标分布的映射来实现对任意连续分布的表达，在研究中所作的网络设计将会主要借鉴这一思想并探索不同的具体的网络结构设计。  三维点集所发生的变换实际是由一个单隐层的多层感知机来实现的，但值得注意的是这一多层感知机的参数是从图像的特征中提取出来的。这一单元结构的任务便是从图像特征中提取信息拟合形状映射网络所需要的参数。  图 6：利用参数化表达的 |
|  |

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。  **已完成的工作**  相比于开题时，在近期的研究中，技术方案发生了一些改动。做这样的改动的原因是因为像ShapeNet这样的公共数据集本身所包含的物体并不是可参数化的流形表面模型，而是一般的CAD模型。它们并不能直接参数化到圆盘来生成目标形状的参数化结果，也就无法直接使用均方误差来对网络输出与目标形状进行比较。  为了像之前的工作一样支持无序点集的相互比较，仿照之前的工作使用了Chamfer距离来比较输出点集与目标点集的差异。但是，实验表明，在原有网络结构的基础上使用Chamfer距离作为损失函数来进行网络的训练，网络无法收敛到能够输出有意义的结果的程度。  为此，在研究中对方案作了进一步改动。整体的网络结构更改为如所示的结构，它的主要改动是在参数化网络中使用“k-n point net” 的基础网络结构来实现从参数域到目标表面的映射。与之前使用全连接网络拟合参数化的结构不同，它对输入的形状的变形不但考虑了输入点的全局坐标而且考虑了邻域的形状。  在经过这样的修改以后，网络便能够输出有意义的形状。  图 7：修改后的利用参数化表达的网络结构：  其中通过语义网络预测参数化网络的网络参数来预测有效的映射将参数域映射到目标形状  这其中的“k-n point net”这一基本单元的的具体结构如所示，它主要借鉴PointNet [Qi et al. CVPR 2017]的网络结构，实现了直接以无序的点集作为输入，对点集形状做变形的操作。相较于PointNet [Qi et al. CVPR 2017]，本文所提出这一新结构以k 邻域内的点集而不是以全部的点集作为输入为每个点预测一个位移量。它既考虑每个点的全局位置信息也考虑它的邻域形状信息。  图 8展示了这一网络结构逐步对形状进行变形并最终生成目标形状的过程。它表明网络能够从输入图片逐步学习对输入点集的变形方法，使得输出形状逐步接近目标形状。  图 8：修改后的利用参数化表达的网络结构中的基本单元  图 9：参数化网络的中间输出结果：这一结果展示了网络能够学习模仿人逐步对形状进行变形并最终得到目标形状  表格 1：所提方法与PSGN[Fan et al. CVPR 2017]方法的视觉效果对比  从表格 1所展示的结果来看，在某些个例上所提方法是能够还原更多细节的，比如第一行中飞机的机尾。以及第三行示例中具有较高的底盘的汽车。  但是如所示，所提方法与PSGN [Fan et al. CVPR 2017]方法在ShapeNet 数据集的55类物体上的数值质量进行对比，实验结果并不理想，从目前的结果来看，数值上来比较，所提方法实际仅仅在三类物体上略超越之前的方法，整体反而比之前的方法要差。  **待完成的工作**  如前所述，所提方法从数值评价上还没有超越之前的方法，可能的原因是：对现有网络的训练方法还不够合理，网络超参数的调整仍在实验中。  开题时所计划的设计更好的损失函数这一目标还没有完成，现有的损失函数并没有充分利用网格结构，如中第一行所示的结果所产生的网格有较多的翻转和自交，所产生的网格质量较差。在接下来我将尝试使用对偶Chamfer Distance，也就是既衡量输出顶点与目标点集的差异又衡量网格的面心与目标点集的差异从而期望使网络所学习到的映射更加平滑，也能够使得网络具有更好的泛化能力。  http://www.latex2png.com/output/latex_433c0e0729a65c1f587c01bebaf9c76d.png  公式 4 对偶Chamfer距离  表格 2：所提方法与PSGN[Fan et al. CVPR 2017]方法的数值质量对比  **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  [HORN--Psychol Comput Vis--1975] Obtaining Shape from Shading Information. In *PSYCHOL COMPUT VIS*, 1975  [Pontes--ACCV--2018] 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--CVPR--2017] Dosovitskiy, A and Brox, T Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs. In *CVPR* pages 2107–2115, 2017  [Wang--ECCV--2018] Zhang, Yinda, Li, Zhuwen, Fu, Yanwei, Liu, Wei, and Jiang, Yu-Gang Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images. In *ECCV*, 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  [Zhang--PAMI--1999] Tsai, Ping-Sing, Cryer, JE, and Shah, M Shape-from-shading: a survey. In *PAMI* pages 690–706, 1999 |

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。  Point Set Joint Registration and Co-segmentation 投稿到SCI期刊 The Visual Computer，目前状态是accept after minor revisions |

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15日前完成分子数据集训练与测试的实验，并分析实验结果。  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