

**2022-2023学年第1学期期末考试（公选课）**

**论文**

**考试科目：**虚拟现实技术

**学院：**计算机学院

**专业：**计算机科学与技术

**班级：**2020211304

**学号：**2020211346

**姓名：**倪玮昊

**手机：**13965540430

**任课教师：**黄海

北京邮电大学

时间：2022年12月12日

**虚拟现实技术与人工智能**

倪玮昊

北京邮电大学计算机学院

**摘要:**

本文将阐述虚拟现实的概念、特征及应用领域，并对其中涉及的关键技术进行总结。首先，我们将对虚拟现实进行定义，并提取出它的主要特征。然后，我们将探讨虚拟现实在三维重建的应用。最后，我们将对虚拟现实涉及的关键技术进行总结，包括未来展望等方面。此外，本文还将着重介绍虚拟现实研究的最新进展，包括三维重建、增强现实等方面的技术、方法和算法。通过对这些方向的研究进展进行综述，我们将能够更好地了解虚拟现实技术的发展方向，并为今后的研究工作提供参考。

**关键字:**

虚拟现实应用领域，计算机图形学，感知与认知，人机交互，三维重建，人工智能，算法。

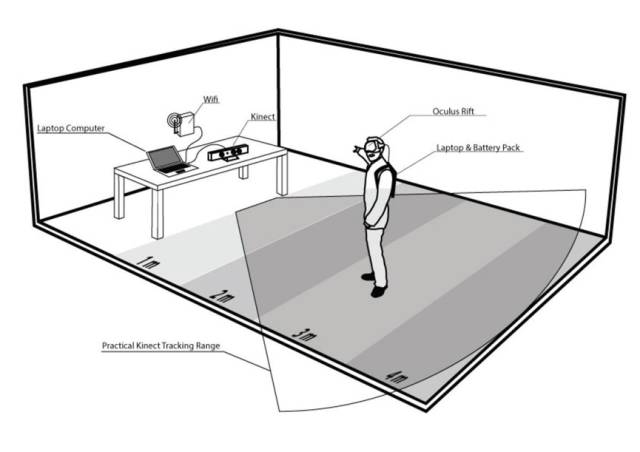
虚拟现实是一种以计算机为基础的技术，它通过电脑程序创建一个接近真实的、交互式的三维虚拟环境，以便用户能够体验到身临其境的感官信息，包括视觉、听觉、触觉等。虚拟现实的特征包括：可以创建一个接近真实的、交互式的三维虚拟环境，让用户能够体验到身临其境的感官信息；可以让用户通过操作特定的设备进入虚拟环境，并在其中进行探索、操作和交互；能够通过各种感官信息创造出一个接近真实的虚拟环境，使用户能够更好地感受到虚拟世界的真实感；能够为用户提供丰富的交互体验，如创建虚拟道具、工具和其他物品；可以被应用于多个领域，如娱乐、教育、医疗、军事等。总的来说，虚拟现实是一种由计算机创建的接近真实的三维虚拟环境，它能够让用户体验到身临其境的感官信息，并且具有丰富的交互性。它在许多领域都有着广泛的应用，并且有望在未来发挥更大的作用。

**技术种类:**

虚拟现实技术不是水面浮萍，而是多种先进技术的积累虚拟现实涉及到多种技术，包括视觉技术、声音技术、传感器技术、人机交互技术和计算机图形学技术。

**视觉技术:**

视觉技术是虚拟现实中最重要的技术之一。它包括头戴式显示器，例如虚拟现实眼镜，可以提供视觉冲击，带给人们更真实的感觉。而无论是什么设备，其追求和核心目标都是给用户创造一种沉浸感，这决定了用户在虚拟现实环境中的感官体验是否自然。为了要让用户在“全包围”的环境中感受到自然的沉浸感，就需要恰当地得到三维空间化的视觉感官。沉浸感的核心参数包括视场角和刷新率。视场角要足够大，以便让人感觉身处一个空间内，而不是看到一块屏幕；刷新率也要足够高，以便避免图像闪烁和抖动。



**声音技术:**

声音技术是另一个重要的技术。它包括声学设计，可以提供更真实的声音效果，例如环绕声和空间声音等。这些声音可以让人们更好地感知虚拟世界，并且可以带来更强烈的参与感。3D音效定位技术可以通过一组喇叭或耳机发出逼真的立体声效，定位出环绕用户身边不同位置的音源。HRTF技术用于换算人耳监测和区分空间声音来源的方法。支援声源定位的游戏将声音与游戏物件、人物或其他声音源结合，当这些声音与游戏中的位置改变时，芯片根据相对位置调整声波信号的发送。种音效定位技术，使得游戏中的声音不再只是单纯的背景音乐，而是可以根据玩家在游戏中的位置和动作，精确地模拟出声音来源的方向和距离，让玩家更加沉浸在游戏中。



**传感器技术:**

传感器技术是虚拟现实系统中重要的组成部分。它可以捕捉人体动作和运动，并将其转化为计算机可以理解的信息，例如指针或触摸等。这样，人们就可以通过手部动作或身体运动来控制虚拟现实系统，与虚拟世界互动。比如，人们进入虚拟世界不仅仅是想坐在那里，他们也希望能够在虚拟世界中到处走走看看。为了实现这一目标，一些公司正在研发万向跑步机等设备，但是，目前这种设备的体验并不太好。另一种方式是使用脚上的惯性传感器来实现在原地走，但是这种方法也存在一些问题。另外，有一些公司推出了全身VR套装，这种套装装备有各种传感器，能够让人感受到虚拟环境的变化，比如微风的吹拂和射击游戏中弹膛的冲击力。但是，目前已有的应用传感器的设备体验度都不高，在技术上还需要做出很多突破。



**人际交互技术:**

人机交互技术是虚拟现实系统中重要的组成部分。它可以改善人与虚拟世界的交互，让人们更容易控制虚拟现实系统，并且能够享受到更好的体验。随着人工智能语音识别技术的发展和商业落地，基于语音识别的交互技术已广泛应用于现实生活中。基于触控的交互技术实现了单点到多点触控的转变，用户可以使用双手进行单点触控，也可以通过识别不同的手势实现单击、双击等操作。基于动作识别的交互技术依赖于动作捕获系统获得的关键部位的位置进行计算、处理，实现用户与计算机之间的交互。基于眼动追踪的交互技术是利用传感器捕获、提取眼球特征信息，实现人类眼球的运动信息获取，从而实现用户与计算机的交互。除此以外还有基于视觉的交互技术，利用摄像机和计算机视觉技术实现用户与计算机的交互，例如MicrosoftKinect。还有基于触觉的交互技术，利用触觉传感器和振动马达实现用户与计算机的触觉交互，例如HaptoMime。这些交互技术都为用户提供了不同的体验和方式进行交互，逐渐改变了我们与计算机之间的交互方式。

**计算机图形学技术:**

计算机图形学技术是虚拟现实系统中另一个重要的技术。它用于生成虚拟世界中的图形，并让这些图形看起来更加逼真。计算机图形学技术包括三维建模、纹理映射、光照计算等。此外，计算机图形学也可以用来实现物体的运动、变形、碰撞等物理效果，为用户提供一种真实的触觉感受。此外，计算机图形学还可以用来实现虚拟现实中的图像合成、光线追踪、视频特效等技术，为用户提供更加逼真的视觉效果。



总的来说，虚拟现实涉及到多种关键技术，这些技术在不同方面都发挥着重要的作用。它们的发展为虚拟现实的普及奠定了基础，并为人们带来更好的体验。

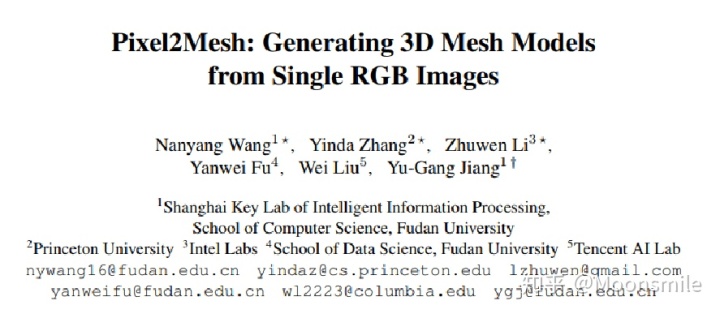
**三维重构:**

三维重建是指用数字图像处理技术从多个视角采集到的图像数据，通过计算机算法来重构出物体的三维形态，可以获得更加精细、准确的三维物体信息。这个技术在多个领域包括工业制造、医学影像、遥感等都有广泛应用。在工业制造领域，三维重建技术可以用于从多个视角采集产品的三维图像数据，然后通过计算机算法重构出产品的三维模型，方便制造商进行产品设计和检验。在医学影像领域，三维重建技术可以用于从CT或MRI等检查影像中重建出患者的器官或骨骼等三维模型，方便医生进行诊断和手术规划。在遥感领域，三维重建技术可以用于从卫星或航空器拍摄到的遥感影像中重建出地面地物的三维模型，方便地理信息系统等应用的开发。

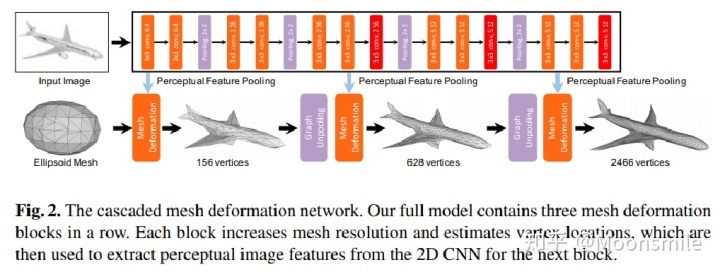
三维重建方法可以根据采集设备是否主动发射测量信号来分为两类：基于主动视觉和基于被动视觉。基于主动视觉的三维重建方法主要包括结构光法和激光扫描法。而基于被动视觉的三维重建方法则只使用摄像机采集三维场景得到其投影的二维图像，根据图像的纹理分布等信息恢复深度信息，进而实现三维重建。其中，双目视觉和多目视觉理论上可以精确恢复深度信息，但实际中，受拍摄条件的影响，精度无法得到保证。单目视觉只使用单一摄像机作为采集设备，具有低成本、易部署等优点，但其存在固有的问题：单张图像可能对应无数真实物理世界场景，故使用单目视觉方法从图像中估计深度进而实现三维重建的难度较大。

而在最近的研究中，研究者尝试将深度学习引入到三维重建中，并且取得了较为好的成果。

**用三角网格来做单张RGB图像的三维重建:**



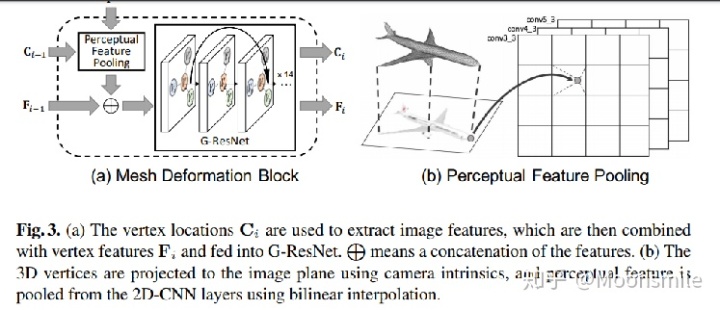
在这篇文章中，作者提出的方法不需要借助点云、深度或者其他更加信息丰富的数据，而是直接从单张彩色图片直接得到3Dmesh。



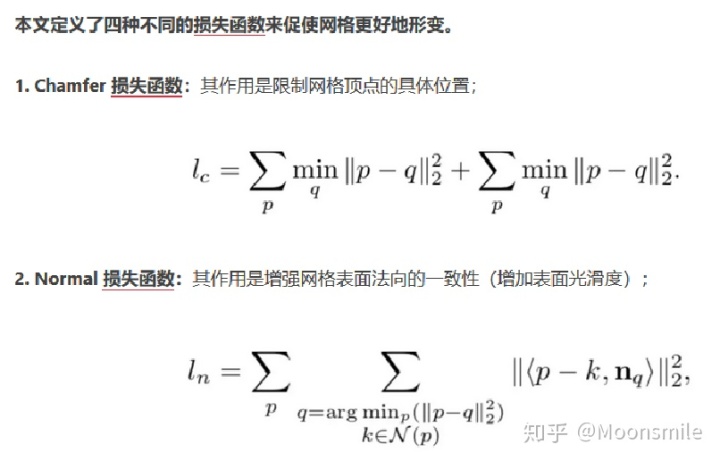
整个模型的主框架是给定一张输入图像：Inputimage，再将为任意的输入图像都初始化一个椭球体作为其初始三维形状：EllipsoidMesh。

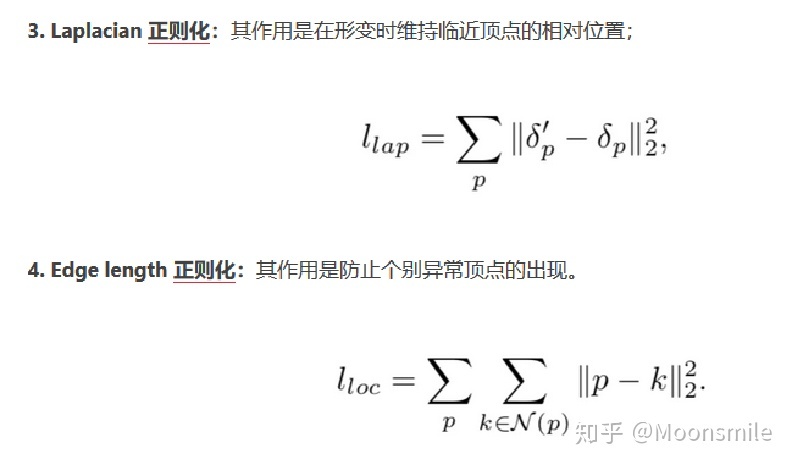
整个网络可以大概分成上下两个部分，上面部分负责用全卷积神经网络提取输入图像的特征，下面部分负责用图卷积神经网络来表示三维mesh，并对三维mesh不断进行形变，目标是得到最终的输出。期中表示三维顶点坐标，P表示图像特征，F表示三维顶点特征，而perceptualfeaturepooling层负责根据三维顶点坐标C(i-1)去图像特征P中提取对应的信息。

以上提取到的各个顶点特征再与上一时刻的顶点特征F(i-1)做融合，作为G-ResNet的输入。G-ResNet(graph-basedResNet)产生的输出又做为meshdeformableblock的输出，得到新的三维坐标C(i)和三维顶点特征F(i)。



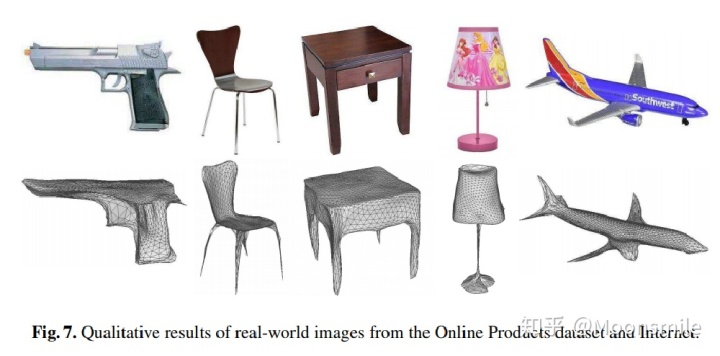
同时这篇文章定义了四种loss来约束网格更好的形变：

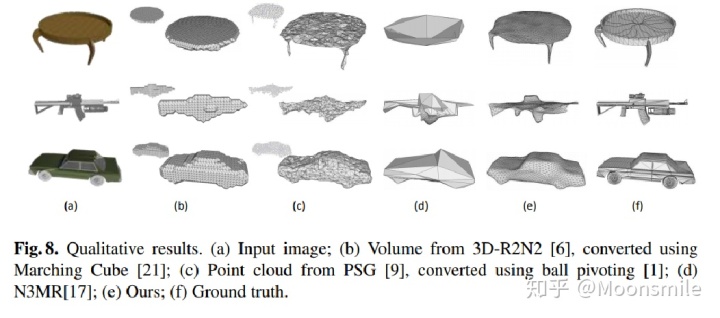




总得来说该文章实现用端到端的神经网络实现了从单张彩色图直接生成用mesh表示的物体三维信息。并且采用图卷积神经网络来表示3Dmesh信息，利用从输入图像提到的特征逐渐对椭圆尽心变形从而产生正确的几何形状。为了让整个形变的过程更加稳定，文章还采用coarse-to-fine从粗粒度到细粒度的方式。文章为生成的mesh设计了几种不同的损失函数来让整个模型生成的效果更加好。

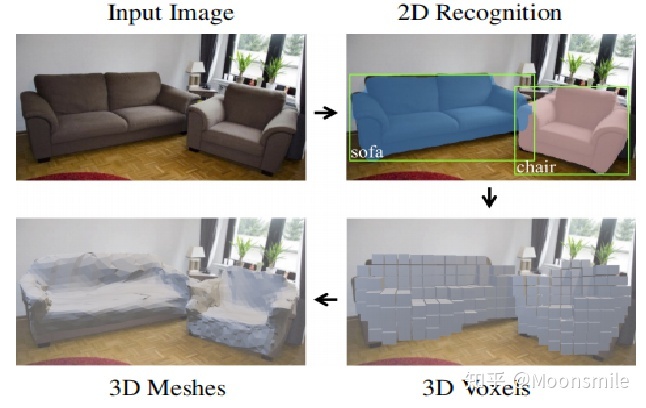
文章的核心思路就是给用一个椭球作为任意物体的初始形状，然后逐渐将这个形状变成目标物体。最终获得了如下结果





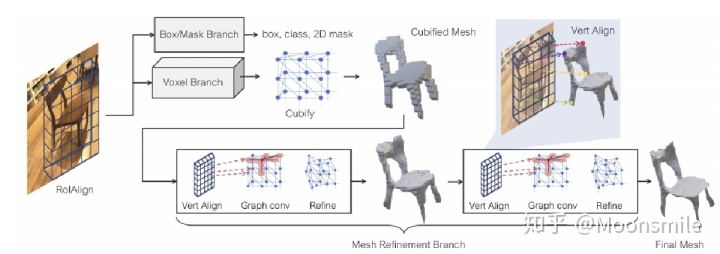
**基于现实图片的物体检测系统:**

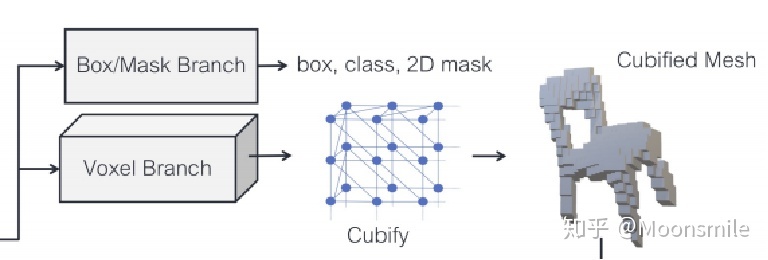
本篇文章提出了一种基于现实图片的物体检测系统，同时为每个检测物体生成三角网格给出完整三维形状。文中的系统mesh-rcnn是基于mask-rcnn的增强网络，添加了一个网格预测分支，通过先预测转化为物体的粗体素分布并转化为三角形网格表示，然后通过一系列的图卷积神经网络改进网格的边角输出具有不同拓扑结构的网格。



该模型通过输入一个吐血图像，检测图像中的所有对象，并输出所有对象的类别标签，边界框、分割掩码以及三维三角形网格。

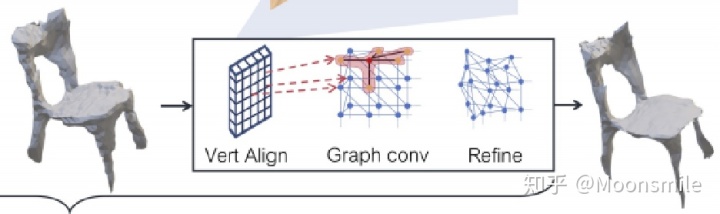
模型主框架基于mask-rcnn，使用一个额外的网格预测器来获得三维形状，其中包括体素预测分支和网格细化分支。先由体素预测分支通过预选框对应的RoIAlign预测物体的粗体素分布，并将粗体素转化为初始的三角形网格，然后网格细化分支使用作用在网格顶点上的图卷积层调整这个初始网格的定点位置。总框架图如下所示：





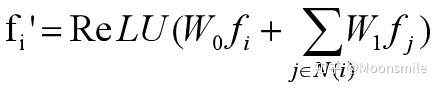
在细支中，Box/Mask分支与mask-rcnn中的两个分支一样，体素预测分支类似于mask-rcnn中的mask分支，使用一个小的全卷积网络来保持输入特征和体素占用预测概率之间的对应关系，并输出用G个通道生成G\*G的特征图，为每个位置提供一列体素占用率分数。最后，将体素占用概率转化为二值化体素占用，并将每个被占用的体素替换为具有8个顶点、18个边和12个面的立方体三角形网格，然后合并相邻占用体元之间的共享顶点和边，消除共享内面，就可以形成一个拓扑结构依赖于体素预测的密集网格。

网格细化分支是输入一个三角形网格，然后通过三个步骤：顶点对齐、图卷积和顶点细化，获得更精细的网格结构。网络的每一层都为网格的每个顶点维护一个三维坐标以及特征向量。



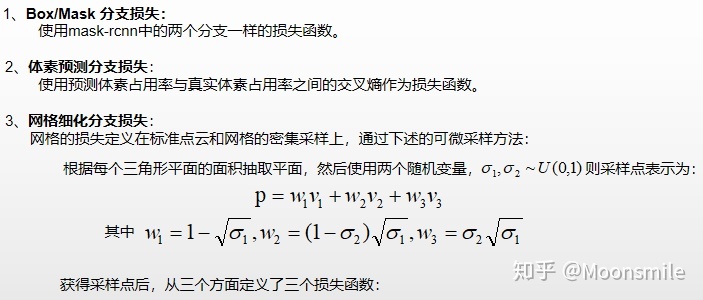
顶点对齐是利用摄像机的内在矩阵将每个顶点的三维坐标投影到图像平面上，然后根据获取的RoIAlign在每个投影的顶点位置上计算一个双线性插值图像特征来作为对应顶点的图像特征。这一步的目的是为了获得顶点位置对应的图像特征。

图卷积用于沿着网格边缘传播顶点信息，公式定义如下：

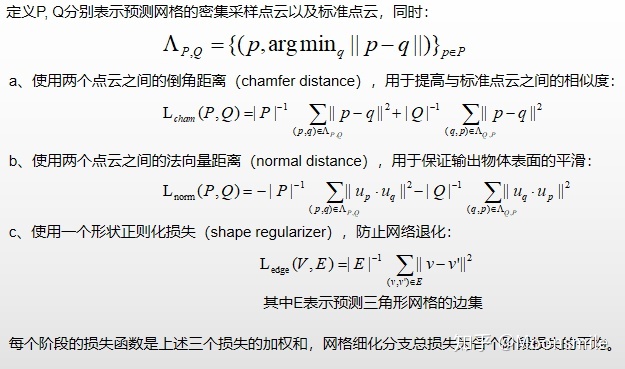


其中N(i)表示顶点i的邻点集合，使用多个图卷积层在局部网格区域上聚合信息。

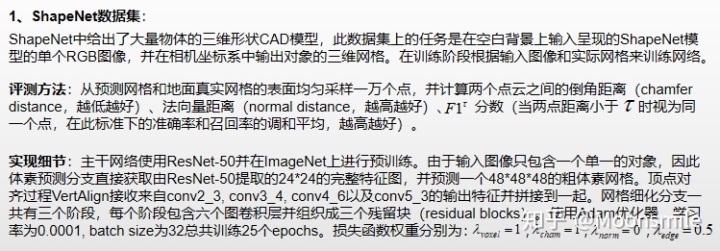
模型损失函数



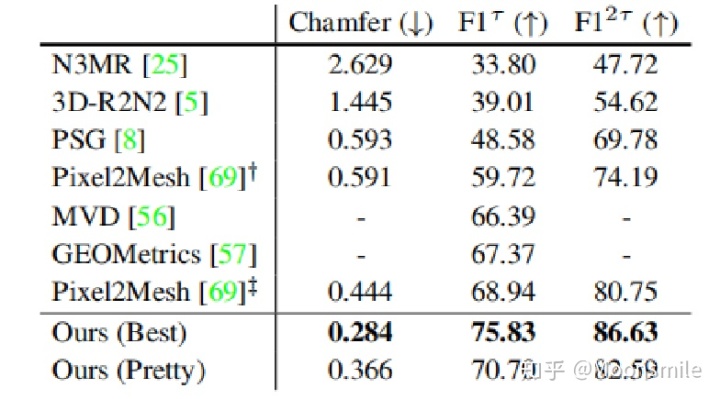
网格细化损失（从三个方面定义了三个损失函数）



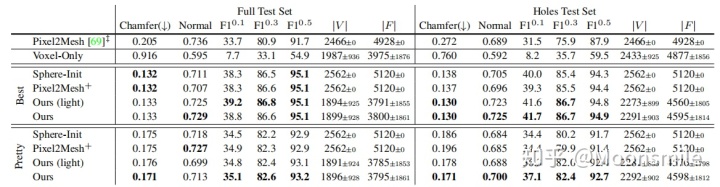
论文在两个数据集上验证模型：在ShapeNet数据集上对网格预测器进行了基准测试与最先进的方法进行比较并且对模型中的各个模块进行单独分析；在Pix3D数据集上测试完整MeshR-Cnn模型在复杂背景下的物体[三维网格](https://www.zhihu.com/search?q=三维网格&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra={"sourceType":"article","sourceId":"79628068"})预测结果。



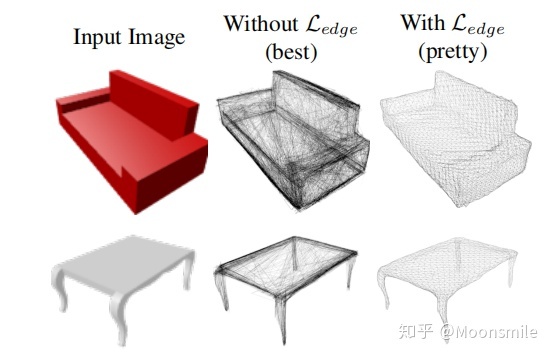
在ShapeNet数据集：MeshR-Cnn与其他的模型比较结果如图下：



其中Ours(Best)表示去掉形状正则化损失后的结果，在后面的实验中可以发现，去掉形状正则化损失后尽管在标准度量上有好的表现，但是在视觉层面上生成的网格并不如加上后的结果（Ours(Pretty)）。



表格中比较了模型的完整版本以及不同去除模块版本的表现，其中FullTestSet表示在完整测试集上的表现，HolesTestSet表示在打孔对象测试集中的表现；Voxel-Only表示不适用网格细化分支；Best和Perry分别表示不使用形状正则化损失和使用形状正则化损失；Ours(light)表示在网格细化分支中使用较轻量的非残差架构。

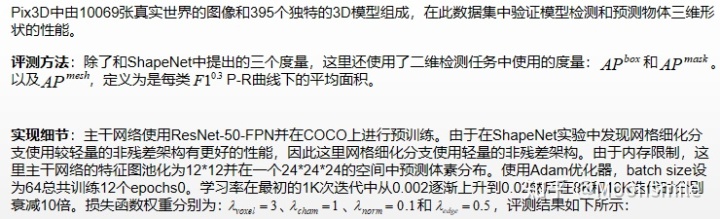


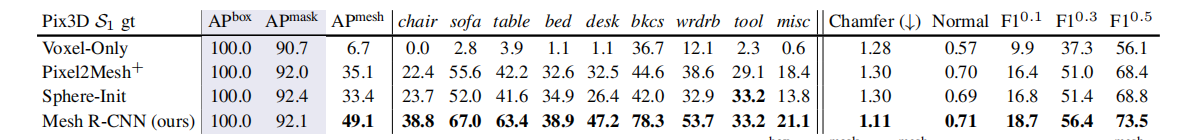
尽管不使用边长正则化器的训练结果在标准度量中有更好的表现，但是会产生退化的预测网格，会导致输出的网格出现许多重叠的面。



对比Pixel2Mesh模型，Pixel2Mesh模型的输出结果是从一个标准椭圆变形得到的，不能正确地建模有孔的物体。相反，MeshR-Cnn可以对任意拓扑结构的物体进行建模。

Pix3D数据集如下





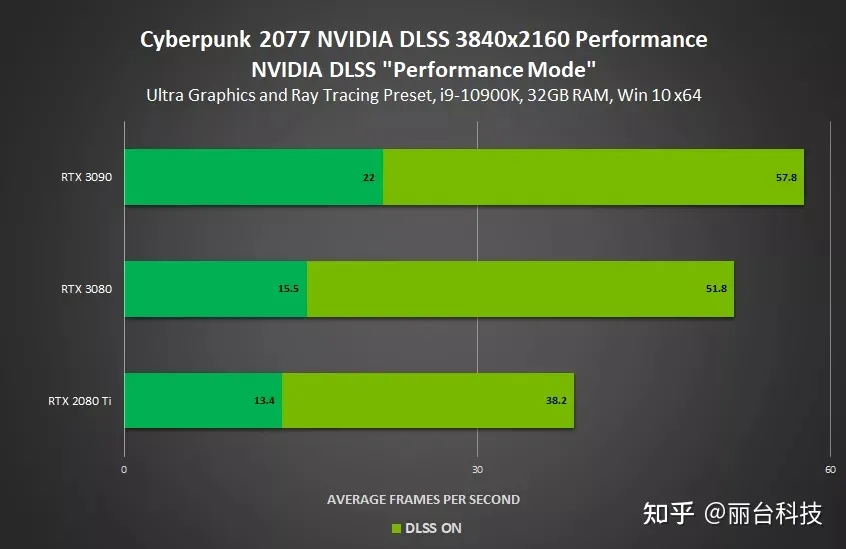
该文章借鉴了maskrcnn框架，提出了一种由粗到细调整的思想，并使用图卷积神经网络和多种损失来进行训练，实现了三维重建的目标。最终得到可视化结果可视化结果





**总结:**

综上，可以看到人工智能技术对于三维重建有着巨大的影响，而人工智能技术对于虚拟现实的影响远远不止于此，人工智能和虚拟现实可以结合在一起，以实现更加逼真的虚拟世界。例如，人工智能可以用来创建更加真实的虚拟人物，使他们能够更加真实地与用户交互。此外，人工智能还可以用来模拟真实世界的物理现象，使虚拟现实更加逼真。



总之，人工智能和虚拟现实的结合可以为用户带来更加真实和丰富的体验。这对我这名来自计算机专业的学生也很有启发，作为AI浪潮下成长的学子，我深深的明白我所学的知识可以对虚拟现实技术产生一些微小的贡献，比如英伟达知名技术DLSS，DLSS使用深度学习算法训练大量的图像数据，然后根据游戏画面的实际情况选择合适的渲染方案，从而提高游戏的图像质量。相比传统的图像抗锯齿技术，DLSS可以在保持画面质量的同时降低计算复杂度，提高游戏的运行速度。这项技术让很多电脑用一个较差的显卡实现了更高的帧数，我认为这就是一个很好的点，可能我在光学方面一窍不通，但是我可以在程序软件方面优化虚拟现实技术，让这项技术有更好的未来，我认这就是我未来努力的方向。

**参考文献:**

[1]啦啦啦是十三吖. VR交互——传感器[N]. 知乎专栏, 2022

[2]二圈妹. 从计算机图形学到虚拟现实(一)[N]. 知乎专栏, 2022

[3]张心欣. 计算机图形学: 我们的成就与挑战(一)[N]. 知乎专栏, 2021

[4]MoonSmile. 基于深度学习的视觉三维重建研究总结[N]. 知乎专栏, 2021

[5]饶毅. 虚拟现实：视觉盛宴背后的技术革命[N]. 知乎专栏, 2016

[6]辜磊. 用虚拟现实（VR）的方式听音乐[N]. 知乎专栏, 2021

[7]Nanyang Wang Yinda Zhan, Zhuwen Li Yanwei Fu, Wei Liu, Yu-Gang Jiang.Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images[J]. arXiv, 2018, 1804.01654.

[8]Georgia Gkioxari, Jitendra Malik, Justin Johnson. Mesh R-CNN[J]. arXiv, 2019, 1906.02739.