堂上练习-三维视觉基础知识和技术

\*请填写如下个人信息，并注意不要删除文档中的试题描述。

姓名： 郭思齐 学号： 3019244140 班级： 计科1班

1. 典型的三维数据表达有哪些？请简述它们各有的优点和应用场景。（10分）

一、三维点云：

点云表示保留了三维空间中原始的几何信息，不进行离散化。点云是某个坐标系下的点的数据集。含了丰富的信息，包括三维坐标 X，Y，Z、颜色、分类值、强度值、时间等等。可以将现实世界原子化，通过高精度的点云数据可以还原现实世界。

应用场景：大规模三维重建，自动驾驶领域，消费电子领域，机器人技术，医疗领域（目前用于医学领域的点云方法：医学点云的特点、获取途径以及数据转换方法；医学点云分割中的传统算法和深度学习算法；医学点云的配准任务定义、意义，以及基于有/无特征的配准方法。

二、深度图像：

三维图像是在二维彩色图像的基础上又多了一个维度，即深度（Depth，D），可用一个很直观的公式表示为：

应用场景：三维重建、姿态估计、目标检测、人体感应、图像分割、SLAM、前景提取、目标跟踪。

三、网格：

网格通常由三角形、四边形或者其它的简单凸多边形组成，这样可以简化渲染过程。但是，网格也可以包括带有空洞的普通多边形组成的物体。

应用场景：建模、渲染、3D 打印

四、体素网格：

概念上类似二维空间的最小单位——像素，像素用在二维电脑图像的视频数据上。体积像素一如其名，是数字数据于三维空间分区上的最小单位。如同像素，体素本身并不含有空间中位置的数据(即它们的座标)，然而却可以从它们相对于其他体素的位置来推敲，意即它们在构成单一张体积视频的数据结构中的位置。

应用场景：三维成像、科学数据与医学视频等领域

1. 根据载体来划分，三维激光扫描获取数据的方式常分为哪几类？（10分）

LiDAR获取数据的方式主要分为三大类：星载、机载和地面。

**星载：**LiDAR采用卫星平台，运行轨道高、观测视野广，基本可以测量到地球的每一个角落，为三维控制点和数字高程模型的获取提供了新的途径，有些星载激光雷达还具有观察整个天体的能力。

**机载：**主要借助无人机（UAV/UAS）进行大规模的点云数据采集。

**地面**分为三小种：地上三维激光扫描；车载MMS；手持激光扫描。

1. 使用三维激光扫描获取数据需要哪几方面工程系统的配合？请描述这些系统之间如何配合工作。（15分）

使用三维激光扫描获取数据分由四个系统Laser、GPS、IMU、Computer工程系统共同配合完成。

Laser（LiDAR单位）对地面进行扫描，

Global Positioning System 追踪飞行器x、y、z的空间位置

Inertial Measurement Unit（IMU）惯性测量，追踪飞行器的位置

Computer计算机记录数据

1. 描述Spatial Transformer Network的作用。（15分）

如果网络能够对经过平移、旋转、缩放及裁剪等操作的图片得到与未经变换前相同的检测结果，这个网络具有空间变换不变性，其包括平移、旋转、缩放及裁剪不变性。

Spatial Transformer Networks提出的空间网络变换层，具有平移不变性、旋转不变性及缩放不变性等强大的性能。这个网络可以加在现有的卷积网络中，提高分类的准确性。

总的来说，Spatial Transformer Network使得神经网络自适应学习物体的旋转、缩放、平移。

1. 描述空洞卷积的特点以及相对于普通卷积的优点。（15分）

空洞卷积引入另一个卷积层的参数被称为扩张率。这定义了卷积核中的值之间的间距。直观上，空洞卷积通过在卷积核部分之间插入空间让卷积核膨胀。这个增加的参数 l（空洞率）表明了我们想要将卷积核放宽到多大。

空洞卷积的两大**优势**在于：1.特征图相同情况下，空洞卷积可以得到更大的感受野，从而获得更加密集的数据。

2.特征图相同情况下，更大的感受野可以提高在目标检测和语义分割的任务中的小物体识别分割的的效果。

1. 描述PointNet++的网络结构特点，以及为什么PointNet++比PointNet能取得更好的性能。（15分）

PointNet++网络结构特点：

与图像分割网络U-Net类似，对于相同的点个数的层，采用了skip connect的方式将Encoder里面的特征拼接到了Decoder的特征后。

而网络模块由Set Abstraction提取全局特征、逐步上采样、分割组成

一般的PointNet缺失多尺度特征。

但是PointNet++中主要借鉴了CNN的多层感受野的思想。CNN通过分层不断地使用卷积核扫描图像上的像素并做内积，使得越到后面的特征图感受野越大，同时每个像素包含的信息也越多。而PointNet++就是仿照了这样的结构，引入Set Abstraction。

1. 设想以下应用场景：

“在户外较大范围的开放场景中，你需要利用机器人对场景中的垃圾进行清理。这需要机器人能对场景中的物件进行精细的理解。确认物件属于垃圾，再进行垃圾的抓取和清理。机器人配备机械臂，具有良好的物件抓取性能。但是，受限于成本，机器人只配备基本的RGB摄像机，以及低功耗的深度摄像机，意味着场景的信息采集的时候有缺失。”

基于上述场景，利用已学习的三维视觉技术，描述如何设计机器人的垃圾识别算法。(20分)

首先，因为“机器人只配备基本的RGB摄像机，以及低功耗的深度摄像机”，我们需要选择合理的数据获取方式。并选择合适的三维数据表达方式。比如可以以深度图像、点云的方式表达三维数据、如机器人实时位置以及周围物体位置信息。

在对垃圾进行识别的时候、可能需要用到一些列的图像处理算法，比如利用SIFT算法、提取特征点。以及前面的题目中提及的 Spatial Transformer Networks。

在不同尺度空间的图像下检测出具有方向信息的局部极点，具有尺度、方向、大小三个特征。

在后续也可以用到点云的PointNet++等等算法进行处理。