

创新实践三第六组，期末

Hand PointNet 3D Hand Pose Estimation using Point Sets

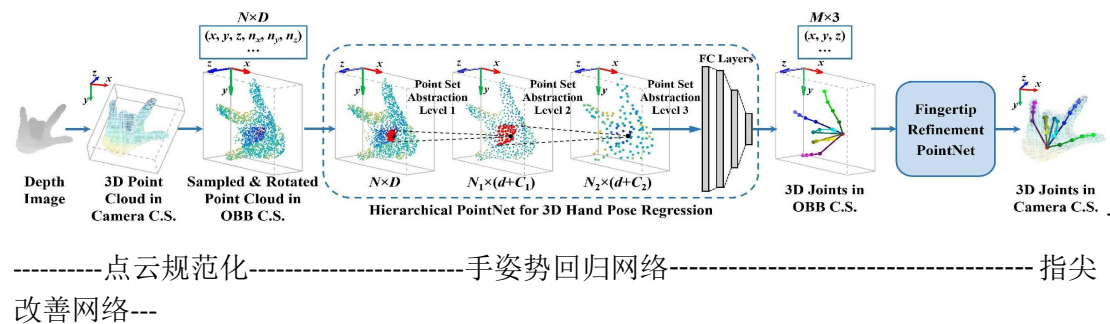
姓名：杨彪 学号：17051433

姓名：万宇航 学号：xxxxxx

1. 论文的理解，总结以及一些关键技术

CNN 与 3D 点云PointNet

- 1.cnn,以 2D 图像和 3D 体积作为输入
- 2.CNN 缺点
- 3.PointNet 直接以规范化的 3D 点云为输入
- 4.对比CNN 的优点
- 5.PointNet 的创新点，点云规范化，指尖改进网络



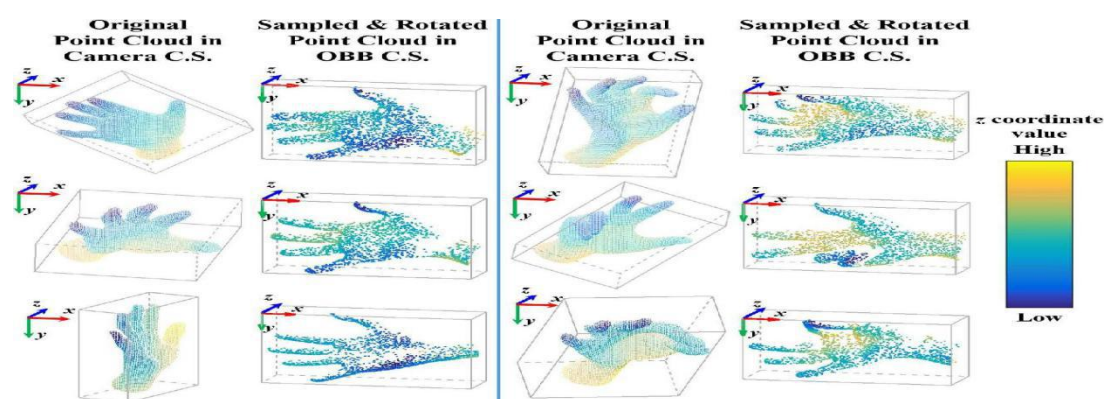
点云规范化

- 1.原因：由于 3D 手姿势的高维度，手部方向的大变化，手指的高自相似性和严重的自

遮挡，3D 手部姿势估计仍然存在准确性和鲁棒性问题。输出 3D 手姿势取决于输入手点云的方向，这使得网络难以以端到端方式进行训练

2. 步骤：分割的手部深度图像首先被转换为一组 3D 点；手的 3D 点云被下采样并在定向边界框中标准化，以使我们的方法对各种手部方向稳健。分层 PointNet 再将附有估计表面法线的归一化点的 3D 坐标作为输入

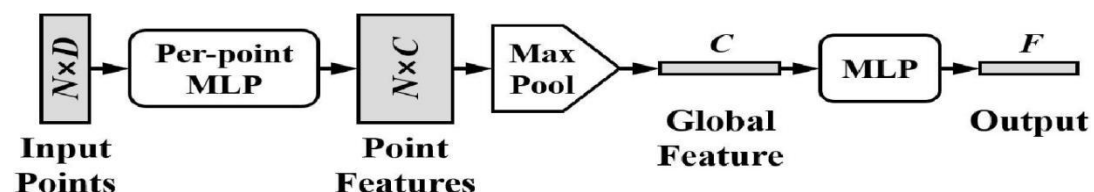
3. 结果：不再需用任何额外的网络来转换手点云，避免复杂网络，具有更一致的全局方向的归一化点云使得PointNet 更容易学习 3D 手部关节，提高基本PointNet 和分层 PointNet 的性能。



1. 通过对输入点的 3D 坐标执行主成分分析（PCA）来确定 OBB 的方向
2. 矩阵，向量对点云进行变换

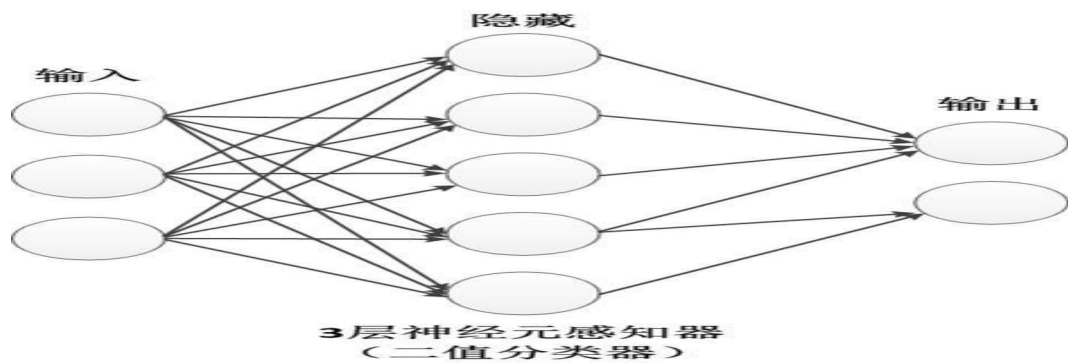
PointNet 网络

1. 基本的PointNet: 一种神经网络，它直接将一组点作为输入，并能够提取点云的判别特征。
2. 分层的PointNet: 用于手姿势回归网络，基本 PointNet 的主要限制是它无法以分层方式捕获点云的本地结构，分层的PointNet，由于其分层特征提取架构，它具有更好的泛化能力。



PointNet 的基本架构。网络直接取 N 个点作为输入。每个 D -dim 输入点通过MLP 映射到

C-Dim 特征。通过max-pooling 将每点特征向量聚合为全局特征向量。再通过MLP 使全局特征映射到F -dim 输出向量。



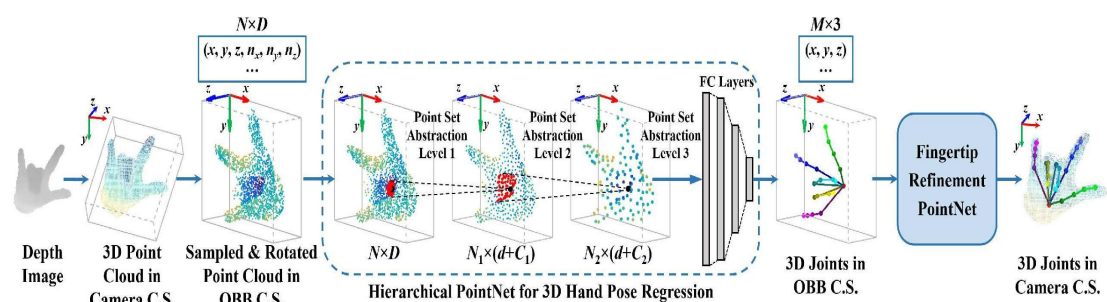
ML

P（多层感知器）神经网络是常见的ANN（即 Artificial Neuro Network 人工神经网络）算法，它由一个输入层，一个输出层和一个或多个隐藏层组成。

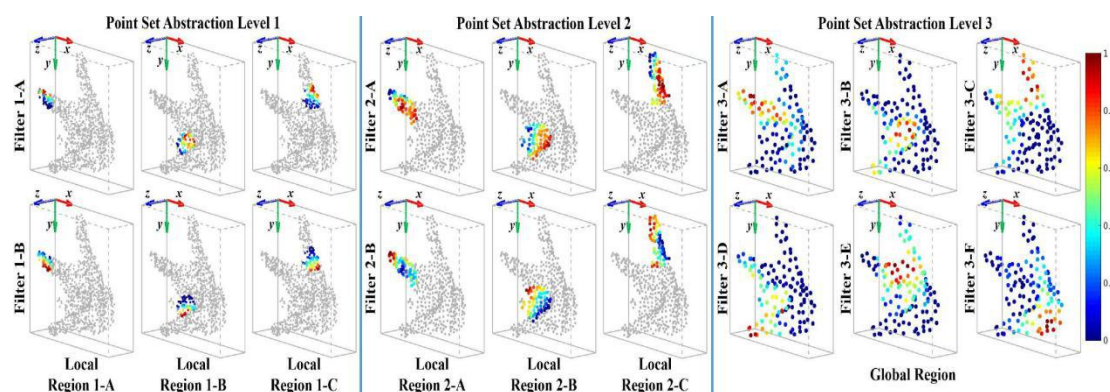
在 MLP 中的所有神经元都差不多，每个神经元都有几个输入（连接前一层）神经元和输出（连接后一层）神经元，该神经元会将相同值传递给与之相连的多个输出神经元，如下图所示。

手姿势回归网络

分层的PointNet:由 L 个点集抽象级组成。在第 1 级中，选择 N 个点作为局部区域的质心;质心点的最近邻居被归为一个局部区域;应用基本PointNet 提取每个局部区域的C-dim特征;带有 d-dim 坐标的N 个质心点和 C-dim 特征被输入下一级，重复下去。在最后一级，通过使用基本PointNet 从该级别的整个输入点抽象出全局点云特征。



在三级分层PointNet 中，将三个集合抽象级别的不同区域中的点对不同过滤器的敏感性可视化。

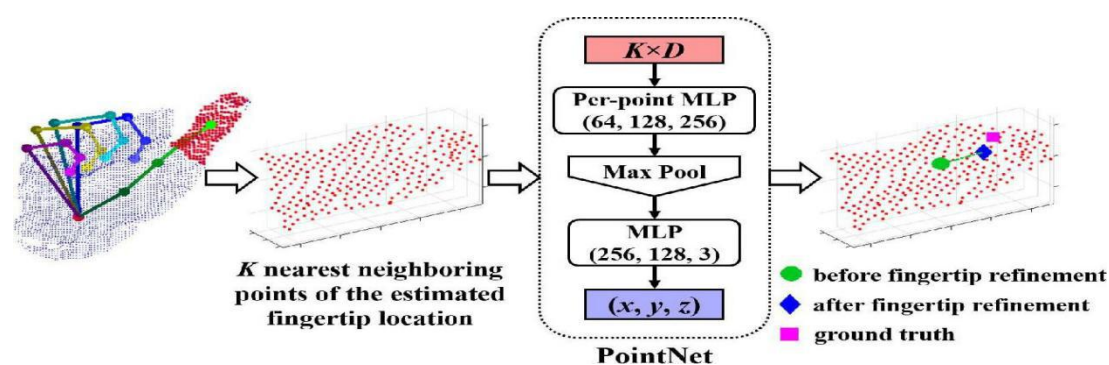


在前两个级别中的每一个，每列对应于相同的局部区域，并且每行对应于相同的滤波器。为了便于说明，仅显示三个局部区域中的点对前两个级别中的两个滤波器的敏感度;在第三级，显示点对六个过滤器的敏感度。高灵敏度的点以红色显示，低灵敏度的点以蓝色显示。不属于本地区域的点以灰色显示。

在前两个级别，过滤器从当地区域提取特征，不同的过滤器夸大了当地区域的不同局部结构。在最高级别，过滤器从该级别的整个输入点提取特征，并且不同的过滤器夸大手的不同部分，例如不同的手指，手掌等。利用这种分层架构，网络可以捕获从本地到全局的手的结构。

指尖改善网络

- 1.使用基本的PointNet 重新确定指尖位置，将估计的指尖位置的相邻点作为输入来回归重新确定的指尖位置。改进网络可以进一步利用原始点云中的细节，并回归更准确的指尖位置。
- 2.原因：第一，与其他关节相比，指尖位置的估计误差通常相对较大。第二，拉直手指的指尖位置通常很容易被确定，因为即使当 k 相对较大时估计位置偏离地面真实位置到某种程度，指尖的最近邻点也不会发生很大变化。



指尖弯曲，则不进行指尖改善，直接用姿势回归的结果

只针对矫直手指重新指尖，首先通过使用关节位置计算关节角度来检查每个手指是否弯曲或伸直。对于拉直手指，我们找到原始点云中指尖位置的 K 个最近邻点以及点数的上限，以确保实时性能。然后在 OBB 中对 K 个最近的相邻点进行归一化。基本的 PointNet 将这些归一化点作为输入并输出重新定义的指尖 3D 位置。在训练阶段，使用地面实际关节位置来计算关节角度；对于最近邻搜索中使用的指定位置，我们将半径内的 3D 随机偏移添加到地面实况指尖位置，以使指尖改进网络对于不准确的指尖估计更加鲁棒。在测试阶段，使用手姿势回归网络估计的关节位置来计算关节角度并搜索最近的相邻点。

实验

在三个公共手势数据集上评估论文提出的方法：NYU [38]，MSRA [33]和 ICVL [34]。

- 1.对比，与之前的方法对比验证论文的优越
- 2.自比较，设置变量，定量，在同一方法下进行对比，以确定合适的参数
- 3.得出实验结果

结论

提出了一种新方法，直接将点云作为网络输入来回归 3D 手部关节位置。为了处理手部全局方向的变化，我们将 OBB 中的手点云标准化为更一致的全局方向。然后将归一化的点云馈送到用于手姿势回归的分层 PointNet 中。指尖位置由基本的 PointNet 进一步确定。三个公共手势数据集的实验结果表明，这种方法实现了三维手势估计的优越性能。

论文代码复现步骤

阅读代码并了解相应的复现建议然后配置环境，复现