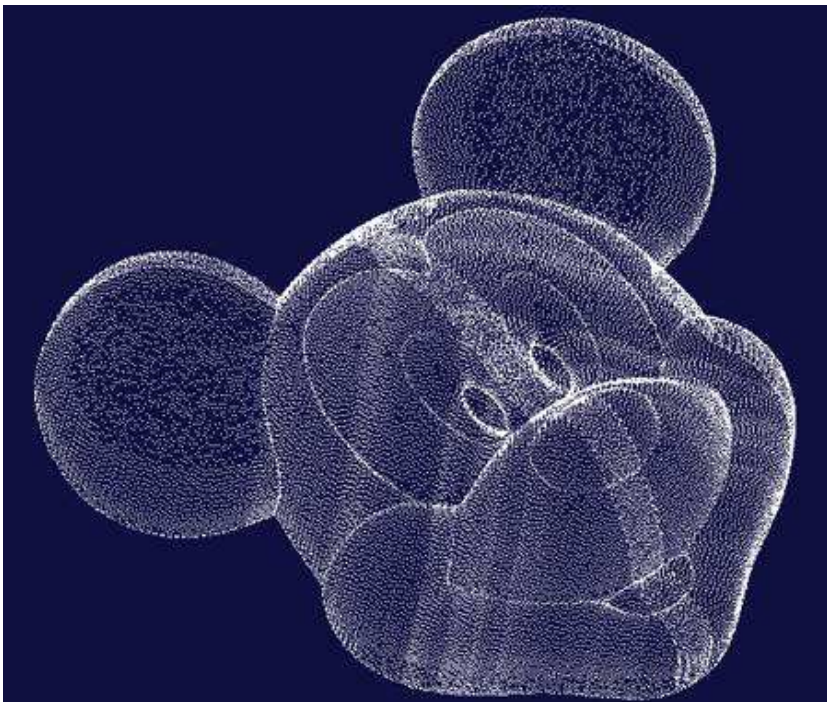


# 点云分类、分割、目标检测的概述

报告人：孙飞

## 一、什么是点云

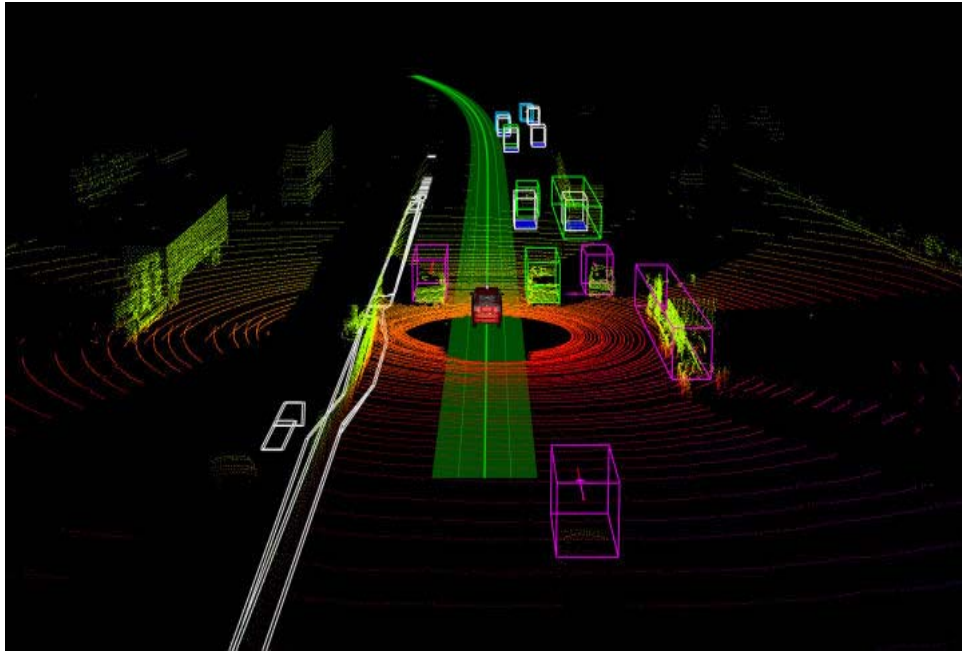


## 二、点云数据的优点



- 数据易采集
- 点云数据包含了三维信息，相对于2D图像而言，数据包含的信息量更丰富。

### 三、应用场景

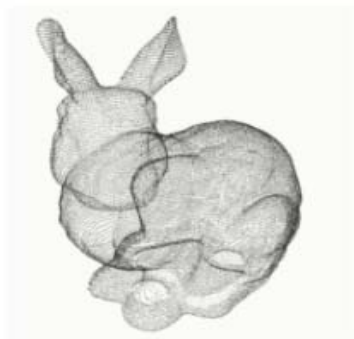


autonomous driving

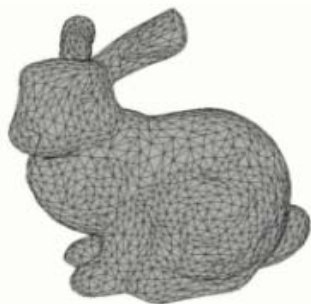


AR

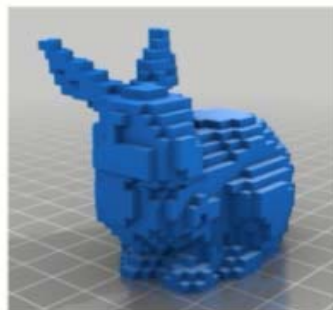
## 四、点云的常见处理方法



Point Cloud



Mesh



Volumetric



Multi-View  
Images  
RGB(D)

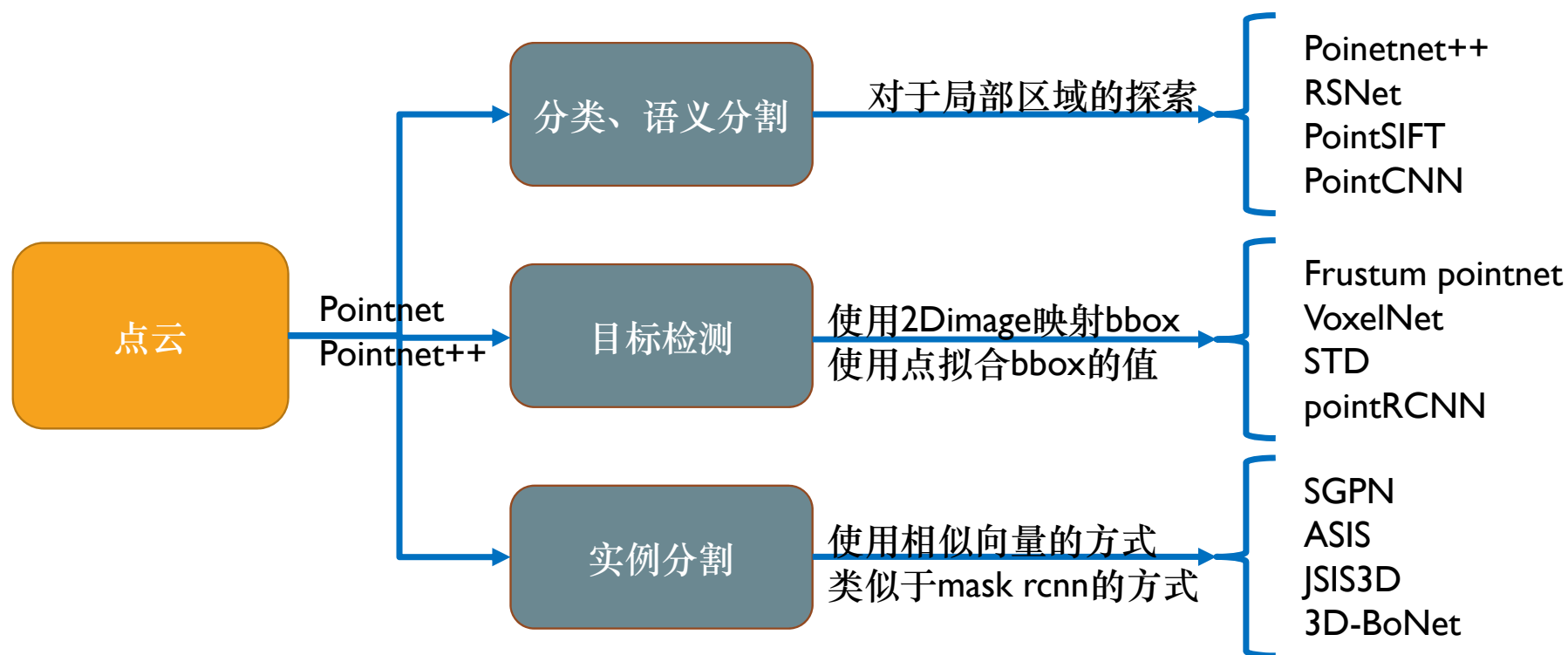
Point cloud: 点云非常接近原始传感器的数据集，数据稀疏，难处理。

Mesh: 连接点生成三角面片或多边形面片，多在计算机图形学中使用。

Volumetric: 将空间划分成三维网格，点映射到网格中，使用三维卷积

Multi-View: 多个图像二维卷积，需要选择拍摄角度等。

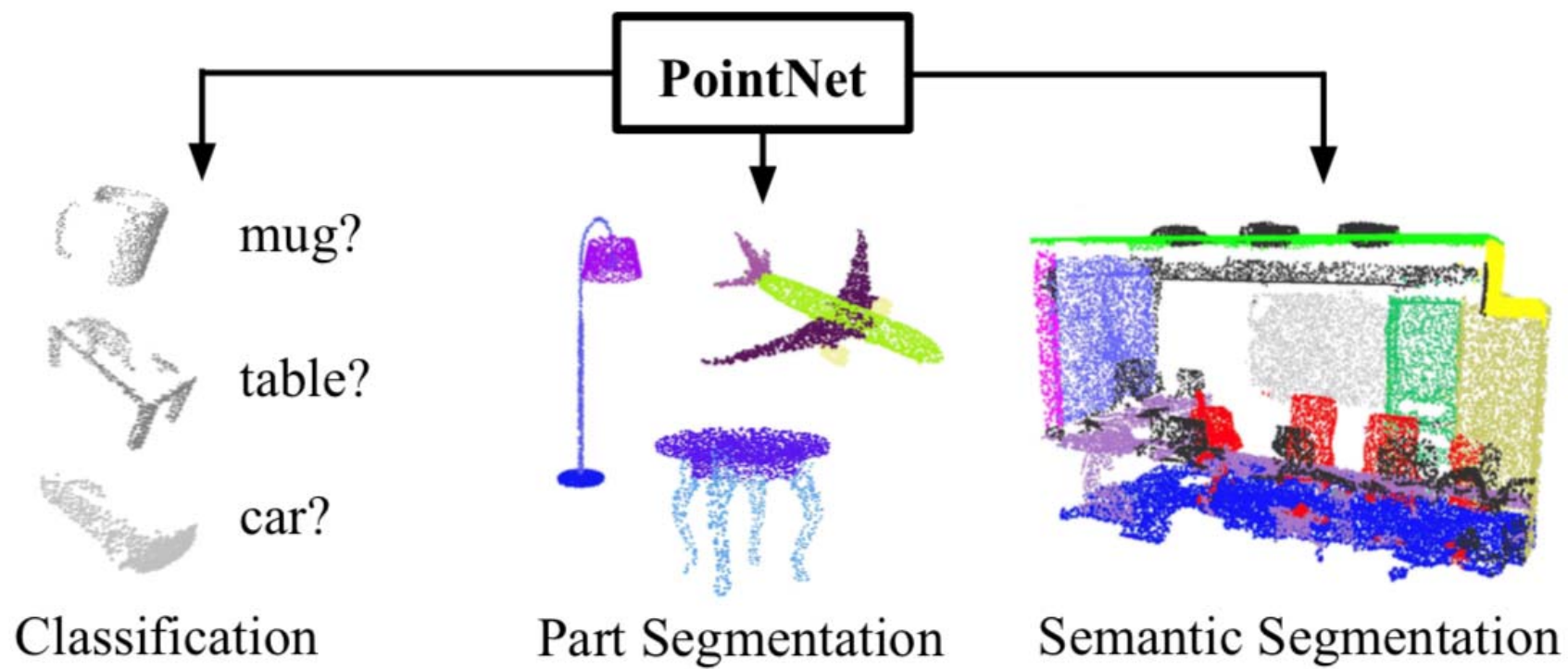
## 五、点云分类、分割、目标检测发展概述



# POINTNET论文阅读

报告人：孙飞





Pointnet完成的任务：分类、部分分割、语义分割



## 一、点云处理的三大难题

- 输入无序性的问题

a. 点云处理的最大问题，对于同一组点云，输入的次序不同，输出的结果必须相同，若共有 $N$ 个点，则其排列共有 $N!$ 个。

b. 对于二维图像的卷积，存在局部空间，但点云并不存在，导致无法进行卷积。

- 点和点之间的交互

点云中点并不是孤立存在的，邻近的局部点云构成了一个有用的集合。

- 旋转不变性问题

当点云旋转或移动时，不应改变点云及每个点的语义分类。

## 二、输入无序性问题

解决方法1: 将输入的点转为一个权威的点。 ✕

理由: 不存在这样一个权威序列。

解决方法2: 使用RNN进行点云的处理。 ✕

理由: RNN模型的训练也和其输入序列的位置相关。

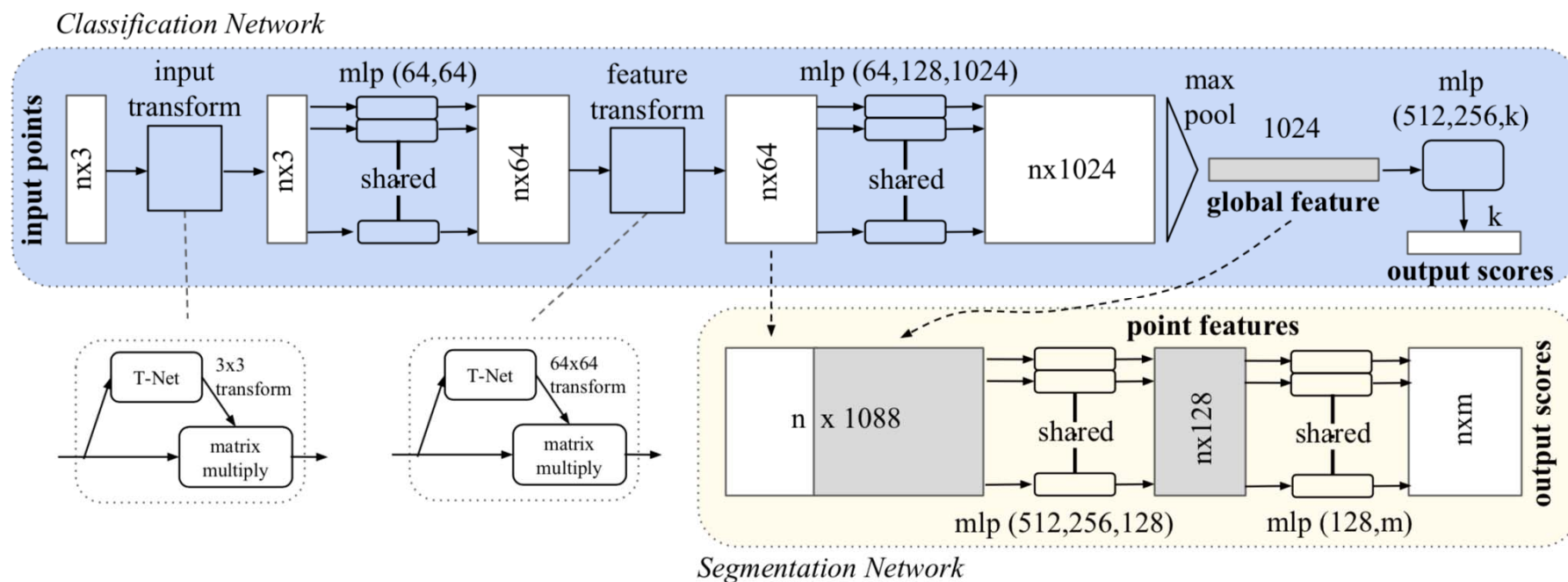
解决方法3: 使用对称结构。 ✓

例子:  $a + b + c + d = a + b + d + c$

$\max(a,b,c,d) = \max(a,d,b,c)$

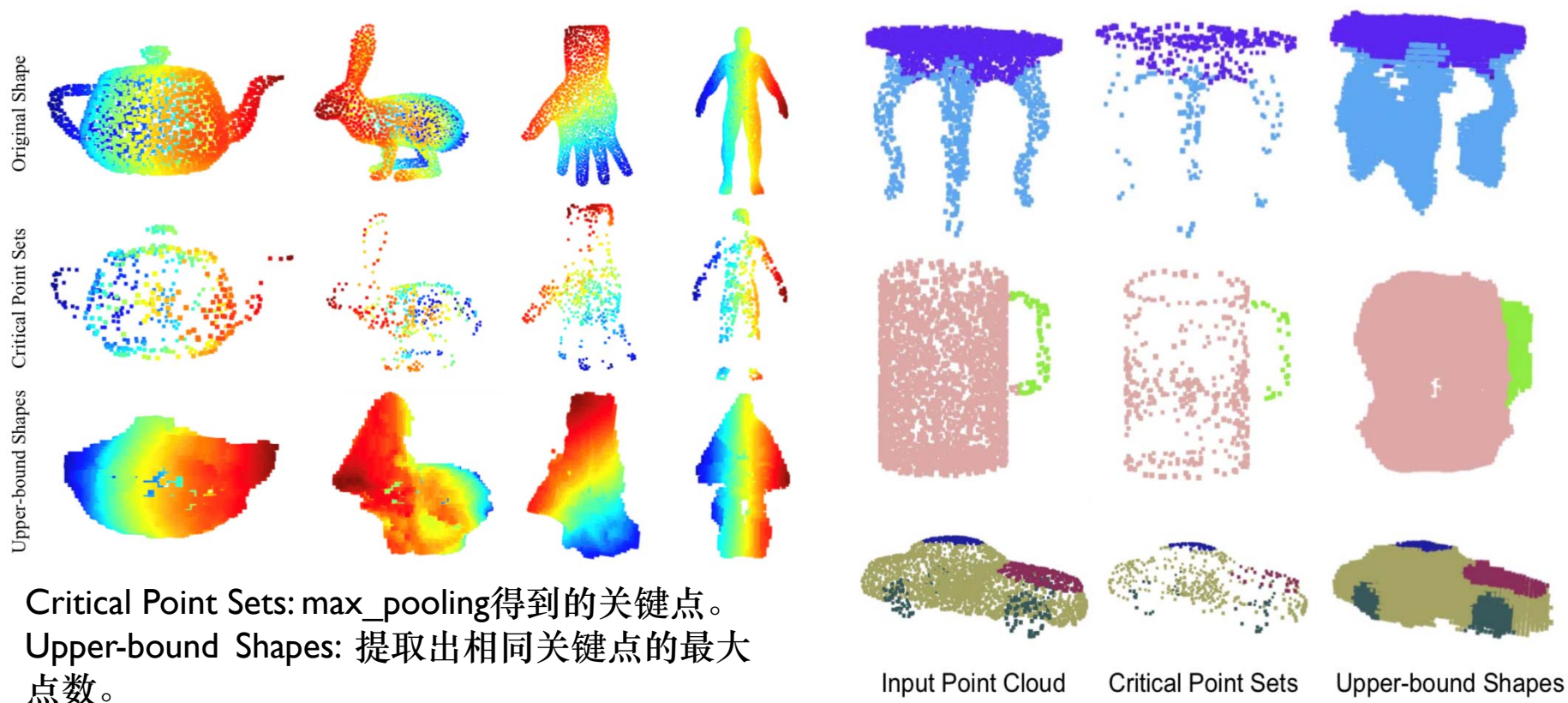
论文中处理无序性问题使用了最大池化，在对n个点进行卷积等操作后，生成 $n * 1024$ 维的矩阵，1024维的1024个整体特征，对于每个维度，求其最大值，解决了输入无序性的问题。即有 $\max(a,b,c,d,e,f,g) = \max(a,d,b,e,c,f,g) = \dots$

## 二、输入无序性问题



- Pointnet网络结构：在前期进行多次卷积后，使用max\_pooling，得到了1024维的全局向量，代表了点云的整体特征。
- 直观来讲，1024维特征中，每个维度通过max操作得到了点云在某个方向下的特点，其中global\_feature即1024越大，得到的效果越好。
- 2d图像卷积: space-centric. 点云: shape-centric

## 二、输入无序性问题



### 三、点和点之间的交互（语义分割问题）

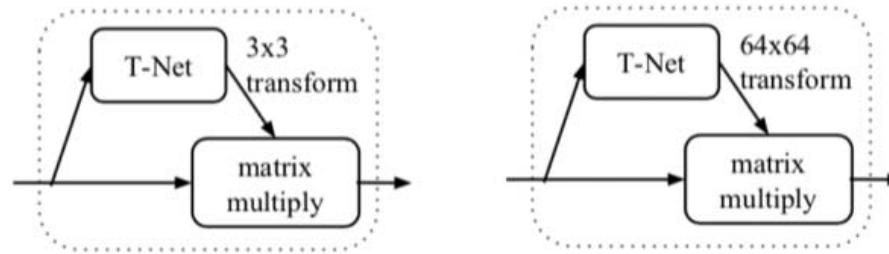
- 在分类问题中，此问题其实不是特别重要，在分类中，只要处理出全局特征，通过对全局特征进行mlp或者用svm分类等，即可得到最终答案，但是在part segmentation 以及 scene semantic parsing中，此问题就显得极其重要。
- 论文中进行处理时，通过将提取出的全局特征以及点特征结合，得到 $n * 1088$  ( $64 + 1024$ ) 的特征矩阵，然后在此矩阵的居处上，重新提取每个点的特征，得到包含局部及全局特征的点信息。



## 四、旋转不变性（仿射变换）

- 论文中使用两个仿射变换矩阵T-Net去简单模拟STN
- T-Net网络结构和主网络类似，其输出换为shape=[3,3]的矩阵
- 在Feature T-Net中添加正则损失  $L_{reg} = \|I - AA^T\|_F^2$

期望预测出的矩阵A为正交矩阵帮助训练[64,64]的矩阵



## 五、实验结果

	input	#views	accuracy avg. class	accuracy overall
SPH [11]	mesh	-	68.2	-
3DShapeNets [28]	volume	1	77.3	84.7
VoxNet [17]	volume	12	83.0	85.9
Subvolume [18]	volume	20	86.0	<b>89.2</b>
LFD [28]	image	10	75.5	-
MVCNN [23]	image	80	<b>90.1</b>	-
Ours baseline	point	-	72.6	77.4
Ours PointNet	point	1	86.2	<b>89.2</b>

Table 1. **Classification results on ModelNet40.** Our net achieves state-of-the-art among deep nets on 3D input.

Classification: ModelNet40



## 五、实验结果

	mean	aero	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
# shapes		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
Wu [27]	-	63.2	-	-	-	73.5	-	-	-	74.4	-	-	-	-	-	-	74.8
Yi [29]	81.4	81.0	78.4	77.7	<b>75.7</b>	87.6	61.9	<b>92.0</b>	85.4	<b>82.5</b>	<b>95.7</b>	<b>70.6</b>	91.9	<b>85.9</b>	53.1	69.8	75.3
3DCNN	79.4	75.1	72.8	73.3	70.0	87.2	63.5	88.4	79.6	74.4	93.9	58.7	91.8	76.4	51.2	65.3	77.1
Ours	<b>83.7</b>	<b>83.4</b>	<b>78.7</b>	<b>82.5</b>	74.9	<b>89.6</b>	<b>73.0</b>	91.5	<b>85.9</b>	80.8	95.3	65.2	<b>93.0</b>	81.2	<b>57.9</b>	<b>72.8</b>	<b>80.6</b>

Table 2. **Segmentation results on ShapeNet part dataset.** Metric is mIoU(%) on points. We compare with two traditional methods [27] and [29] and a 3D fully convolutional network baseline proposed by us. Our PointNet method achieved the state-of-the-art in mIoU.

### Part Segmentation: ShapeNet

	mean IoU	overall accuracy
Ours baseline	20.12	53.19
Ours PointNet	<b>47.71</b>	<b>78.62</b>

Table 3. **Results on semantic segmentation in scenes.** Metric is average IoU over 13 classes (structural and furniture elements plus clutter) and classification accuracy calculated on points.

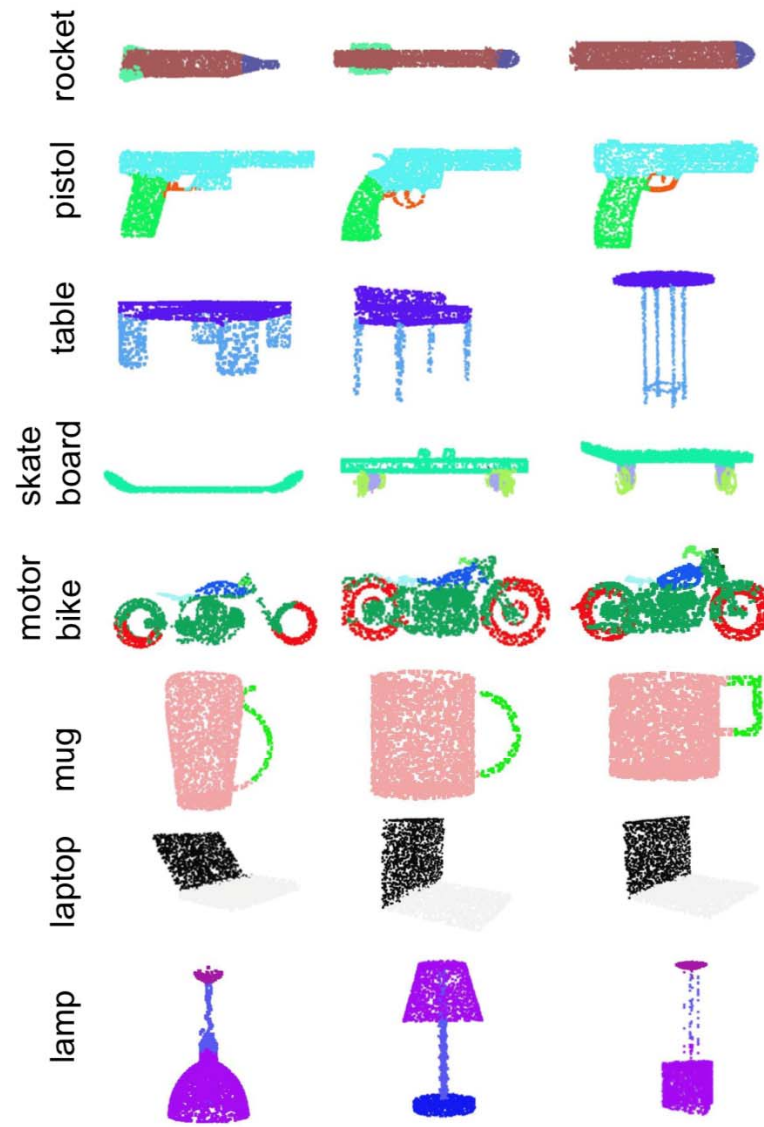
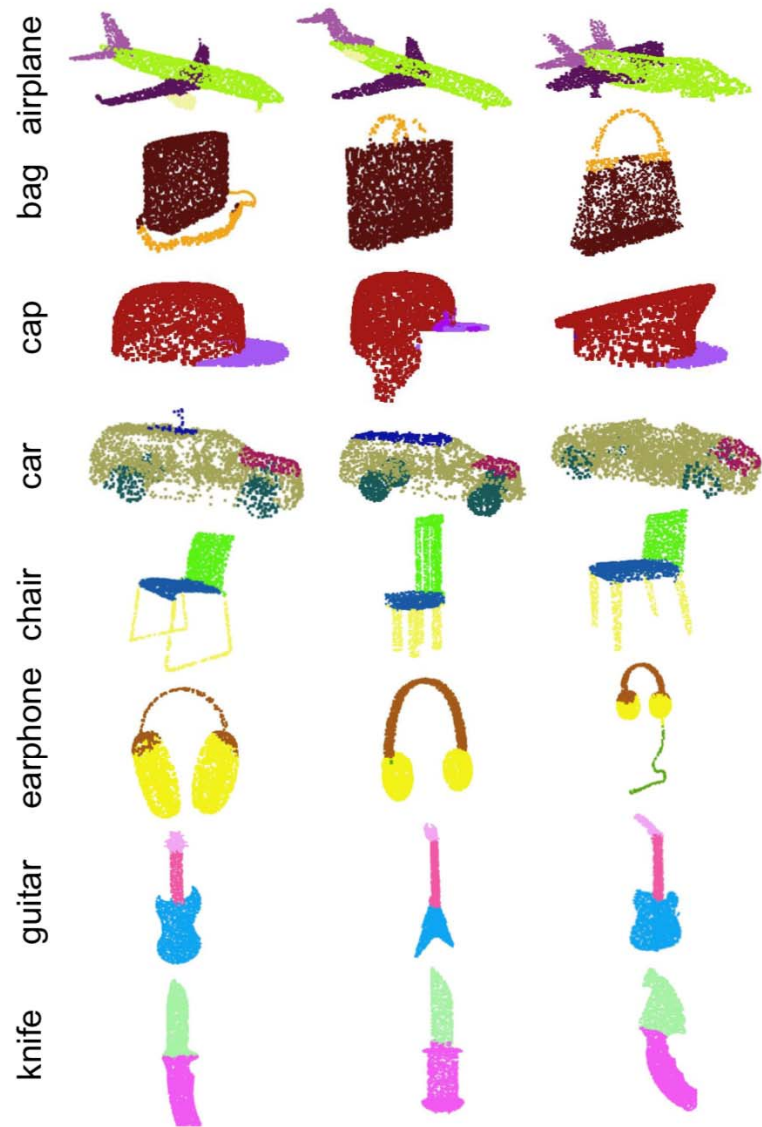
Semantic  
Segmentation:  
S3DIS

## 五、实验结果

	#params	FLOPs/sample
PointNet (vanilla)	0.8M	148M
PointNet	3.5M	440M
Subvolume [18]	16.6M	3633M
MVCNN [23]	60.0M	62057M

#params: 网络参数量, 模型大小的评估

FLOPs: 浮点数运算次数, 即需要多少的运算量



Pointnet部分可视化结果

## 六、总结

- 第一个可以直接处理点云的网络。
- 可以进行分类、部分分割、语义分割多重任务。
- 实验结果超出当前最好成绩或和当前最好成绩相当。

谢 谢 观 看！