2022、3、30组会发言

不规则的数据形式，由于立体平面，物体具有六自由度，因此搜索空间迅速变大。传统的二维空间目标检测算法不再适用。

这次我主要的主题是关于融合检测与transformer。

首先我介绍一下融合检测的基本方法，分为前中后融合三种方式，首先看后融合，是在决策方面上的融合，就是2D与3D检测独立，依靠2维平面上的识别框对3维空间的搜索范围进行限制。

第二种就是候选框层级上的检测，通过自适应池化可以将不同模态的数据获取的特征进行融合，后续利用融合的特征进行目标检测。

第三种就是最底层的数据融合，所造成的信息损失最少，但复杂程度与所需计算量最大，也是最近非常主流的方式，具体网络看右图就是将图片特征图对应位置的特征逐点分配到点云特征里。

然后就是特征提取方式，之前已经介绍了3D目标检测的主流的特征提取方式为3D卷积核特征提取与pointnet++特征提取，最近的趋势大家已经使用transformer进行特征提取，我最近读的几篇cvpr2022的文章基本上都是使用transformer。所以接下来我重点介绍transformer与数据融合的文章。

首先是transformer的基本结构，左下角分为编码器与解码器，编码器将输入的数据进行全局融合，解码器通过Q索引从整体融合的特征中再挑选所需要的特征融合后进行输出。

编码器概括起来就是给n个特征输入，会同样输出n个同维的输出，但是每个输出都会或多或少的具有其他n-1个输入的特征，具体权重有很多计算方式，最简单的就是向量点乘。最简单的思路这种特征提取方式可以插到特征图后，也可以插到pointnet++的中间层。因此我一开始使用这种特征提取方式与pointnet++进行结合。这也是

解码器就是通过设计特定的索引人为挑选出更倾向的特征，两者相互结合，所向无敌。

上面这篇是cvpr2021的点云补全，使用transformer获得残缺的中心点，再使用后续网络进行补全。

下面是cvpr2022的目标检测，使用transformer对点云进行特征提取，再使用获得的结果作为Q对图片进行特征提取。这两篇文章的关键都是Q的选择由之前的随机初始化变成了有意识的设计。

借鉴了EPNET,它主要是点在特征图对应位置的特征

未传入位置信息，所以64个点无法获得远近关系，只能依靠自身特征获取全局特征，容易造成特征的混淆，后续进行上采样展开中心点效果很差。

第一轮效果很好原因是之前网络已经学到了这批样本特定的位置关系，如果样本随机采样位置关系改变之后，网络无法做出调节。造成效果迅速下降