2022、5、16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使用transformer进行融合检测。

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。

即主要分为两个方向：首先使用一种模态产生候选框，再利用得到的候选框提取在其中另一种模态的数据对候选框进行进一步优化，这要求提取特征的能力要强，因为另一模态的特征特别依赖于后选框的建立；另一种是特征图层面的，将不同模态的特征图进行融合后再产生检测框。这就要求它们具有非常好的对应关系。而trans可以自适应的找到它们的对应关系且融合特征的能力非常强大。

接下来是三篇融合的文章，这是关于多视角融合的，将两个视角分别产生特征图，使用trans将它们融合起来。具体来说，先使用3D卷积核提取特征，然后将其分别投影至前视图与俯视平面形成2维特征图，再使用俯视图作为Q提取对应位置的前视图特征，将他们融合起来形成融合特征。

第二三篇是使用trans对图片与雷达数据进行融合，均是分别提取两种类型的特征，再将雷达数据作为k获取图片的特征再与自身雷达点的特征融合起来。这不需要传统方式（之前讲的epnet）的投影变换矩阵，网络自适应的进行选取。

还有我之前做得一些实验，就是不同阶段的特征进行融合，我选取voxel的网络在其二阶段精调时加上了候选框中的原始点云的特征，效果还不错。可以看到第一张图是未融合时对于所有点的建模，第二张图是融合后对17000个抽样点的建模，即使所需的点是原来的五分之一，可使用的信息减少了，但最终效果差不多，甚至有些要更好。

因此综合来看使用trans进行特征融合已成为主流方式，下一步我将尝试将其三个融合起来。