2022、6、19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，通过自适应池化或者注意力机制可以将不同模态的数据获取的特征统一到相同尺寸直接进行相加或者concat融合，这种方式强行将不同后续利用融合的特征进行目标检测。

第三种就是**最底层的点级数据融合**，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点级方式，分为两种主流方式，第一种是将点云的特征融合到图片上，即将图片变成深度图。然后用2Dcnn处理RGB-D数据。然而，这种激光雷达到相机的投影引入了严重的几何失真(近处与远处的点重叠在一起)，这使得它对几何导向的任务效果不好，如3D物体识别。

将图片像素值投影到点云虽然保留了空间信息，但是这些点级融合方法几乎不能在面向语义的任务上工作，如BEV语义分割。这是因为相机到激光雷达的投影在语义上是有损的(见图1b)：对于一个典型的32束激光雷达扫描仪，只有5%的像素点将与激光雷达点相匹配，而其他所有像素点都将被删除。

这两种是输入之上的融合，有的方法将输入先转换为特征图，再将特征图按点级进行一一融合。**要是整体的不同模态特征图就是点级的融合，要是出来ROI范围内再取出其他的特征就是候选框层级的融合。**

**候选框和点级是主流的融合方式！**



首先这是一篇今年五月份的**多数据点级层面特征层级的融合**，首先将点云与图片的特征投影到俯视图（BEV）上，再将它们直接concat然后使用3D卷积核进行融合。

然后这两篇今年CVPR2022的工作，是利用transformer对不同数据进行融合，使用一种特征图作为Q索引，提取另一种特征图的特征，然后将特征concat。

这两个都是在特征图层面上的融合，得到的融合特征再放入检测网络进行物体边界的回归。

下面这篇三月份的是候选框层面上的融合。它利用候选框的中心点作为参考点，在几个不同特征的特征图对应位置提取特征进行concat并融合，**这里的融合是几何位置对应获取特征。（微波雷达点云取最近的k个点，点云与图片双线性差值）**得到的特征作为Q送入BERT的检测器（实际上Q就是每个候选框的特征）。

传统的voxelnet二阶段候选框层面的融合也是几何位置获取，每个点获取周围最近的k个体素特征进行平均池化作为候选框的特征。

因此我对候选框层面的特征使用transformer融合，分为三个阶段。第一个阶段将候选框中的关键点统一使用transformer进行融合，它们每个点可以考虑内部其他关键点的特征。

第二个阶段把每个关键点提取周围体素特征的池化操作改成transformer提取，可以使点根据周围体素的远近动态提取。

第三个阶段原来每个候选框中有216个关键点，我觉得太多，以中心为一个关键点，然后将周围48个体素的特征首先进行充分融合，再使用中心点的几何位置对这48个体素的特征进行动态选择。

我二阶段先使用几何位置缩减搜索范围，再使用transformer灵活挑选。比起以往一阶段的特征层面的特征图所有特征都要进行transformer融合减少了计算量。又摒弃了原来单纯使用双线性差值或者池化的固定，可以自主学习挑选什么特征。

未来工作：

在特征融合层面进行融合，单特征提取使用新方式进行提取。多特征融合使用transformer

检测器可以选择DERT等新检测器

二阶段融合方式也可以进行创新

损失函数改变