# 2021-11-07 会议记录

一、 工作汇报

1. 鲍威

**本周工作**：

Transformer相关论文阅读+制作PPT；实验：Transformer+SAR舰船检测；多光谱数据集文章复现。

**下周工作**：

尝试将Transformer+BNN结合解决匹配问题；多光谱数据集+多模态目标检测代码。

1. 张雅琴

**本周工作**：

变化环境光以及human\_posture数据集的实验。包括，网络的修正、数据正确性检验，数据增强等技巧尝试；多模态成像文献以及非视域成像文献阅读。

**下周工作**：

项目；多模态成像文献阅读+PPT制作；继续调代码。

1. 谭佳

**本周及下周工作**：低截获波形期刊文章+海上目标项目。

1. 钱李鑫

**本周工作**：

Transforemer 分割文章+代码+复现；遥感数据集+代码复现；海上目标本子。

**下周工作**：

其他遥感数据集数据集上进行测试；分割文章阅读+复现。

1. 李欣荣

**本周及下周工作**：修改文章文字+图像；阅读匹配相关的文章。

1. 黄舒心

**本周及下周工作**：多模态综述文章阅读；复现GAN相关代码。

1. 刘雅诗

**本周及下周工作**：跨模态生成的论文阅读；复现GAN相关代码。

二、Paper reading：

本周会议的题目：Vision Transformer--可以捕获全局信息。

问题1：Position embedding 的作用是啥？

答： Transformer Encoder 对输入patch没有位置感，在模型中增添Position embedding，可以学习patch之间的位置关系。

问题2：为什么sin-cos函数可以实现Position embedding？

答： 从图1列举的0-15的二进制表示可以发现，最后一列的数值每隔一个数值就发生变化，倒数第二列的数值每隔两个数才发生变化，以此类推。使用二进制值会浪费空间，因此使用有周期性的可以对应等效交替位的正弦或者余弦函数来替代。

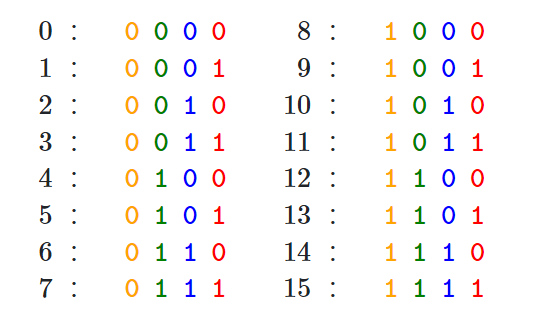


图1 数字0-15的二进制表示

问题3：图2中的公式是什么？

答：t是输出词的位置，pt是相应的编码，d是编码的维度。随着k的增大，即i的增大，wk的取值在减小，也就是频率沿着向量维度d递减。这个函数和二进制编码数的思想类似。

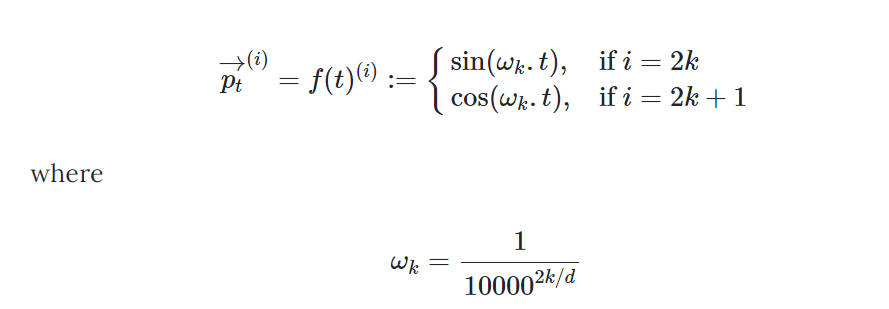


图2 产生输出向量的函数

问题4. Position embedding是直接加到patch embedding 上吗？

答：是的。

问题5：假如图像分割的个数、以及每个patch的像素一样，那么对于同样大小的图像得到的位置编码图是否一致？

答：是的。

问题6：Layer Normalization 与batch Normalization的区别?

答：Batch Normalization 的处理对象是对一批样本， Layer Normalization 的处理对象是单个样本。Batch Normalization 是对这批样本的同一维度特征做归一化， Layer Normalization 是对这单个样本的所有维度特征做归一化。

问题7：Swin的性能比不上CNN(AP: 59.3 vs. 60.3)的结果，在否是在两个模型参数量一致的情况下进行比较的？

答：是的。用的Tiny和resnet50进行比较的。

问题8：重构模型的输入是由70%遮挡的图像吗？

回答：是的。

问题9：MAE中的mask是指遮挡图像？还是说是注意力机制中，只计算相邻词向量的那种mask？

回答：就是指遮挡图像。

问题10：MAE的重构网络有使用repeating吗？如果有的话，模型在训练过程还是可以看到整个图像的真实场景的。

答案：不确定。需要结合代码来看。

问题11：MAE中重构图像为啥会有灰色色块？？

答：本身是没有的，此处是为了让读者更好的看到遮挡处的恢复效果。

问题12：patch embedding是网络学习的还是固定的？

答案：是学习出来的。

问题13：Attention Map 可视化的图是怎么做出来的，感觉经过embedding和flatten后维度不一样了？

答：做这个图的时候是对维度做，而不用维度做。

记录人：张雅琴