第十周我对原始网络的layer\_attention模块进行了消融实验，并将实验结果与no\_adjustment消融实验进行了对比。从最终的实验分析中，我发现网络去掉了层注意力机制之后，在性能上并没有多大的损失，在某些指标上甚至超越了原始性能（+0.1AP~0.2AP），这个结果说明了层注意力机制的加入并没有有效地提高性能，在任务对齐的场景下也没有发挥其期待的作用。

同时在运行上周改进过后的DistanceLoss时，发现在运行过程中出现了报错，这个报错导致显卡出现锁定的情况，一度影响了显卡上所有任务的训练。为了在debug过程中能够不影响其他任务的训练，我选择通过AUTODL云显卡平台对模型进行调试。经过几天的debug后，我发现了根本原因：在计算distance时，相关执行函数中没有对结果的最小值进行截断，这使得有些距离结果非常小，而这些很小的距离值在经过反向传播后，会将一部分梯度变为nan，这些nan会随着反向传播的过程弥漫到整个网络中去，使得第二次网络前向推理时得到的预测值全是nan，在计算分类loss时，其中的BCE函数需要保证类别在索引范围内，而nan的结果使得这个保证无法满足，最终出现了报错。最终我在距离计算函数中对最小值进行了截断限定，该报错消失。

下周的计划：

（1）对预测头做refinement，即预测两次，优化框质量

（2）在网络中预测iou指标，并用预测的iou替代原始metric中的iou，在定位loss中拉近iou和预测iou的分布，nms采用iou做分数排序。

（3）做原始论文stack\_convs的消融实验（删去堆叠卷积层对性能的影响）