第三部分目标：可以通过图像序列对餐厅相关的顾客动作进行预测，这种预测不仅仅只是预测出人未来一定时间的动作序列，还要和不同物品结合形成特定动作，这些特定动作可以编码。

目前的预测想法：motion prediction + 物品的分类，最后学出餐厅相关的顾客动作。

motion prediction就是通过对一定时长的动作序列分析，预测出未来一定时间的动作序列（当前又找到论文实现，效果至少能预测10s好像），物品分类就是图像分类，分类最多就盘子，碗，杯子（要是能拓展筷子，叉子，勺子）。

最后两个组合也是需要一个算法，最后才能输出特定动作。（这里算法目前没有想法）。

然后就是数据问题？从哪里获得数据来学习？图像分类没问题，我不知道motion prediction情况如何，我看看论文。

时间安排（初步）：

目前想的是中期报告肯定在我们回学校以后，而回学校以后才有可能部署实体，对于算法鲁棒性才能测试，所以我理想想法是可以在回学校前，第三部分能跑出结果，刚好你最近不是学习yolo嘛，既可以和控制部分结合，这个yolo也可以当作图像分类器。所以你的想法是什么？第三部分分工？还是我们先完善第一部分部署到实体的算法？