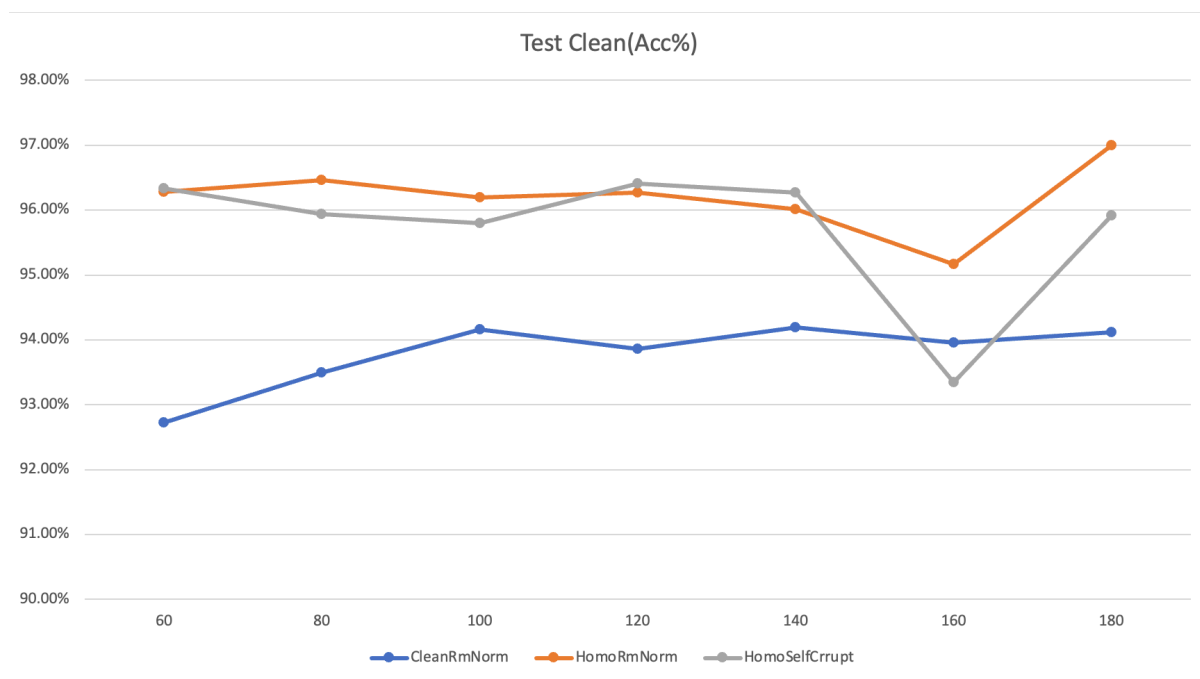


CleanRmNorm: 使用clean数据训练, MLP层中去除LayerNorm

HomoRmNorm: 使用mixture数据(干扰语音为非关键词语音)与clean数据计算HomoLoss, 使用clean数据计算CrossEntropy Loss

HomoSelfCrrupt: 使用mixture数据(干扰语音为关键词语音)与clean数据计算HomoLoss, 使用clean数据计算CrossEntropy Loss

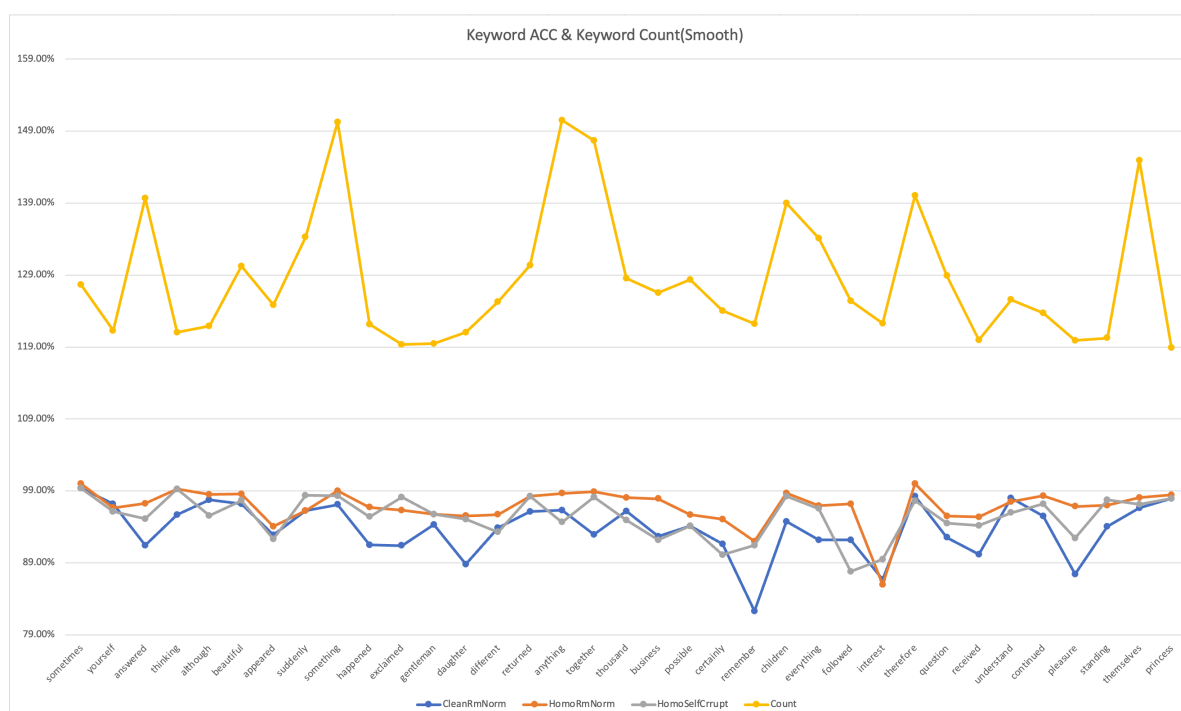
=====Clean测试集的结果=====



上图为三个模型在clean测试集上的结果, 纵轴为Accuracy, 横轴为训练轮数Epoch。

### 结论:

1. HOMO RmNorm与HOMO SelfCrrupt, 整体均优于CleanRmNorm。
2. 在Epoch160处, HOMO RmNorm与HOMO SelfCrrupt都有性能的突然下降, 可能是训练不稳定导致的。



上图为分别为测试集中的每个关键词统计的准确率。图中最上层的黄色折线为每个关键词在训练数据中出现的次数(为了方便在一个图中展示，次数做了Softmax+常数，值越大在训练集中出现的次数越多)

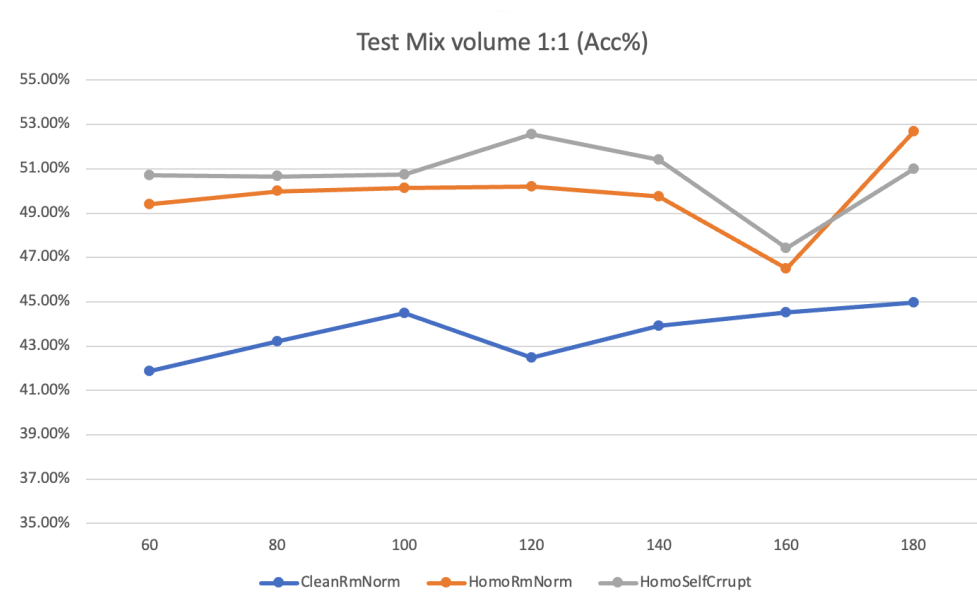
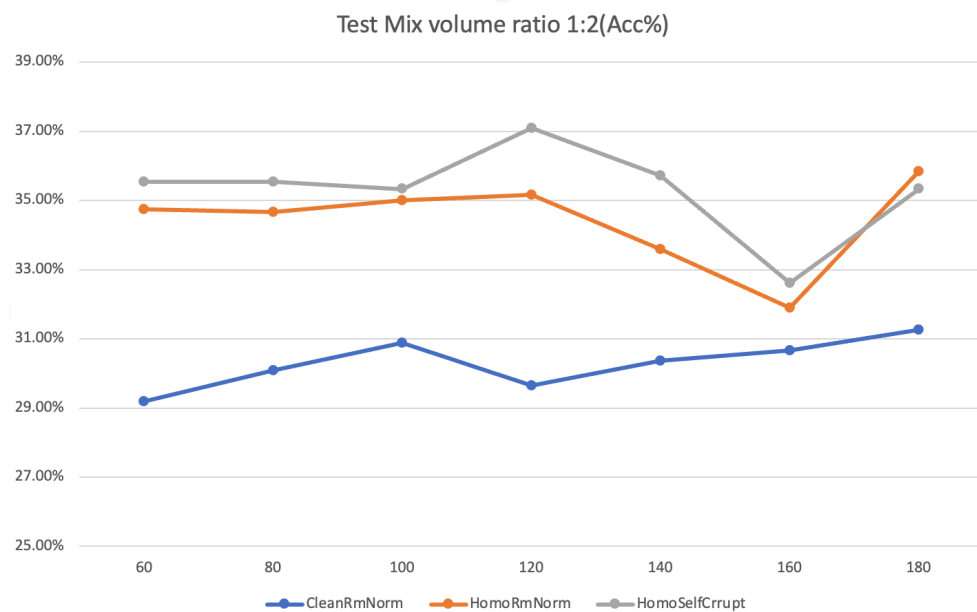
### 结论:

1. CleanRmNorm识别不好的词大部分为训练集中出现次数相对较少的词，可以对应看黄线的波谷与蓝线的波谷
2. 两个Homo模型（橙色线，灰色线）与CleanRmNorm（蓝色的线）之间Gap比较大的地方也大多出现在黄线波谷的位置。

### 可能的原因:

关键词之间存在竞争，比如为了让A更准确不得不把B的某些Pattern丢掉，因此当A的训练数据充分而B的训练数据不充分时，B的Pattern丢失的较为严重，而HomoLoss强制可加，所以会尽可能多的留下有用的Pattern，因此可以挽救一些因为数据量不足导致的性能下降(有待进一步验证)。

=====Mixture测试集的结果=====



以上两个图为在Mixture测试集上的测试结果。其中关键词语音与干扰语音音量比值为分别为1:2与1:1

1. HomoRmNorm与HomoSelfCrrupt均好于CleanRmNorm，性能差异最大处绝对差值可达10%。
2. HomoSelfCrrupt略强于HomoRmNorm，原因可能是HomoLoss希望留下所有的Pattern，HomoSelfCrrupt的训练数据中，只有关键词的Pattern出现，因此关键词相关的Pattern保留的较为完整

3. 下图为在Mixture测试集(干扰语音为非关键词语音)上不同CNN层激活值的比例，由于该测试集中有非关键词语音，这些语音为HomoSelfCrrupt没有见过的语音，因此HomoSelfCrrupt在大部分CNN层的激活值均多于HomoRmNorm，尤其是在最后一层，HomoSelfCrrupt放出了更多的Pattern而HomoRmNorm过滤了更多的Pattern (**HomoRmNorm把垃圾Pattern的激活值置0了?**)。

不同层CNN激活值比例

