ImageCaption领域要解决的问题以及现有方法：

提高网络识别的准确性：attention。包括对图像不同部分的关注或者加入detection得到的attribute进一步引导generation。可以做的点：寻找新的cues以及它与网络的结合方式。Attention的计算用softmax的局限性

定义新的优化问题取代交叉熵：用增强学习，定义不同reward。可以做的点：定义新的reward，这个reward的baseline可以用怎样的网络学？

运用新的framework： memory network，review network，rnn-cnn+mlp

扩充词汇：数据集融合

基于语言结构

多任务（正在try的方案）