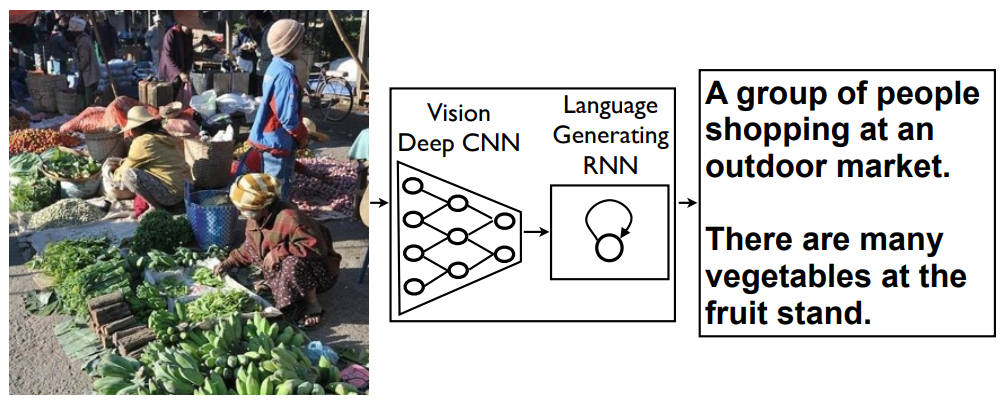
**Show and Tell**

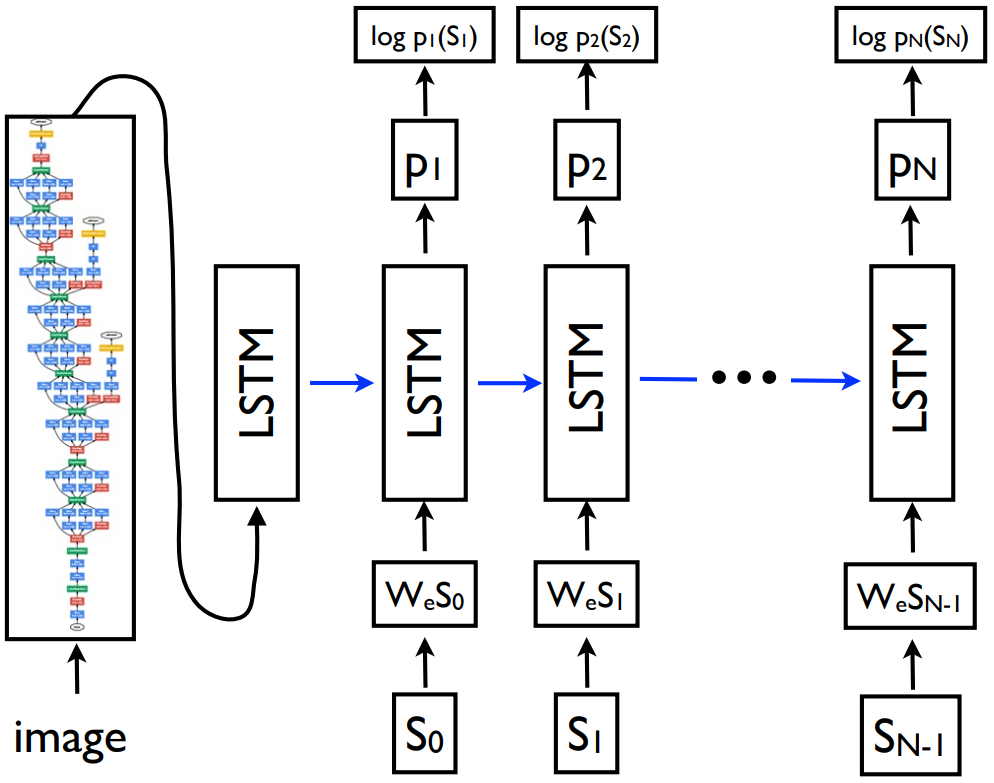
**1基本思想**



模型采用编码器-解码器结构，其中编码器采用CNN，获取图像的特征映射，并将其输出作为解码器的输入，解码器采用RNN，序列化生成正确的图像描述中各单词的概率。

模型训练目的是最大化生成给定图像的正确描述的概率。

**2 模型结构**



**变量解释**

卷积神经网络的特征映射维度是，将其在宽高维度上做均值处理，得到

表示词嵌入，嵌入向量的

表示起始单词，是某特定图片的正确描述

**3 损失函数**

对于图片和其描述句子，损失函数为

对LSTM全部参数，CNN网络最顶层，词嵌入进行优化，最小化各图像-描述对的损失函数之和