视觉注意力机制是人类视觉所特有的大脑信号处理机制。

而抑制其他无用信息。

其实注意力模型可以看作一种通用的思想，本身并不依赖于特定框架，这点需要注意。

Encoder部分的输入是一副图片，Decoder的输出是能够描述图片语义内容的一句描述语。一般而言，文本处理和语音识别的Encoder部分通常采用RNN模型，图像处理的Encoder一般采用CNN模型。

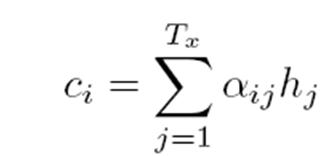
Soft Attention模型的基本原理

如果拿机器翻译来解释这个分心模型的Encoder-Decoder框架更好理解，比如输入的是英文句子：Tom chase Jerry，Encoder-Decoder框架逐步生成中文单词：“汤姆”，“追逐”，“杰瑞”。

在翻译“杰瑞”这个中文单词的时候，分心模型里面的每个英文单词对于翻译目标单词“杰瑞”贡献是相同的，很明显这里不太合理，显然“Jerry”对于翻译成“杰瑞”更重要，但是分心模型是无法体现这一点的，这就是为何说它没有引入注意力的原因。

没有引入注意力的模型在输入句子比较短的时候问题不大，但是如果输入句子比较长，此时所有语义完全通过一个中间语义向量来表示，单词自身的信息已经消失，可想而知会丢失很多细节信息，这也是为何要引入注意力模型的重要原因。

理解Attention模型的关键就是这里，即由固定的中间语义表示C换成了根据当前输出单词来调整成加入注意力模型的变化的Ci。



**g函数就是对构成元素加权求和**，也就是常常在论文里看到的下列公式：

每一时刻的信息是之前隐状态的加权求和

比如“汤姆”的时候，**你怎么知道AM模型所需要的输入句子单词注意力分配概率分布值呢？**就是说“汤姆”对应的概率分布：

即**通过函数F(hj,Hi)来获得目标单词Yi和每个输入单词对应的对齐可能性**

然后**函数F的输出经过Softmax进行归一化就得到了符合概率分布取值区间的注意力分配概率分布数值（这就得到了注意力权重）**。图5显示的是当输出单词为“汤姆”时刻对应的输入句子单词的对齐概率。绝大多数AM模型都是**采取上述的计算框架来计算注意力分配概率分布信息**，**区别只是在F的定义上可能有所不同**。