

Lecture14 Transformers and Self-Attention For Generative Models

卷积神经网络

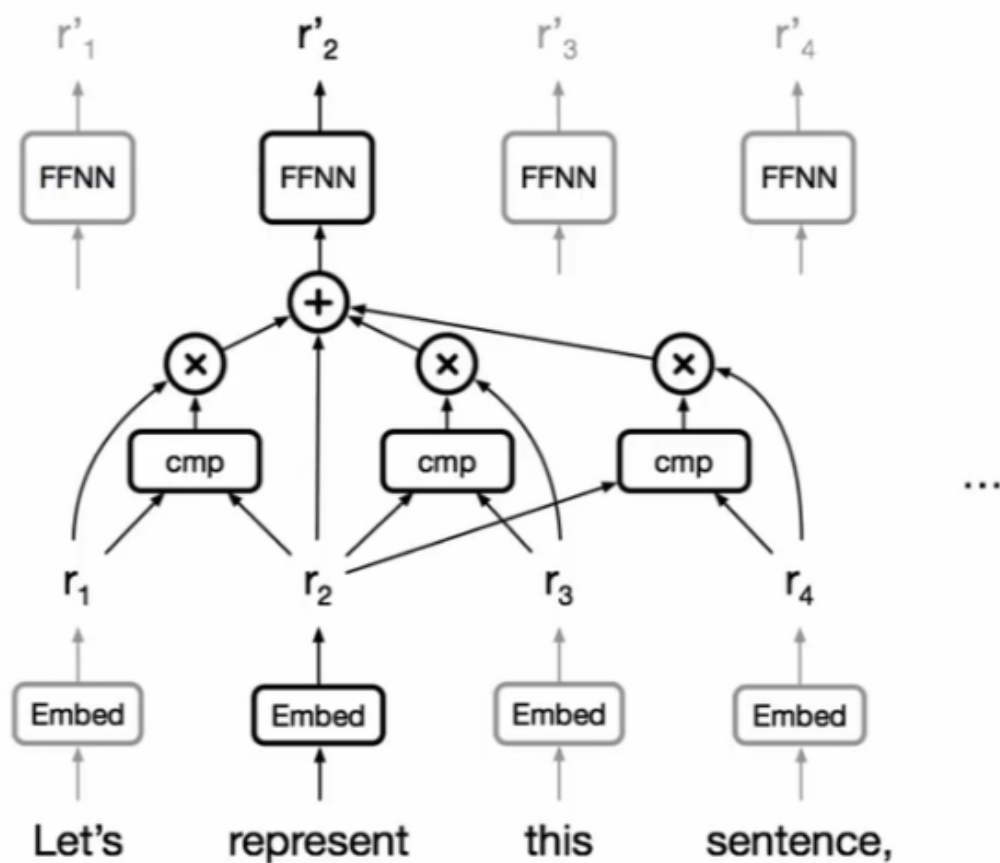
并行化(每层)很简单。

利用当地的依赖关系

位置之间的相互作用距离。

远程依赖需要多个层。

Self-Attention



Music generation using relative self-attention

传统的 RNN 模型需要将长序列嵌入到固定长度的向量中