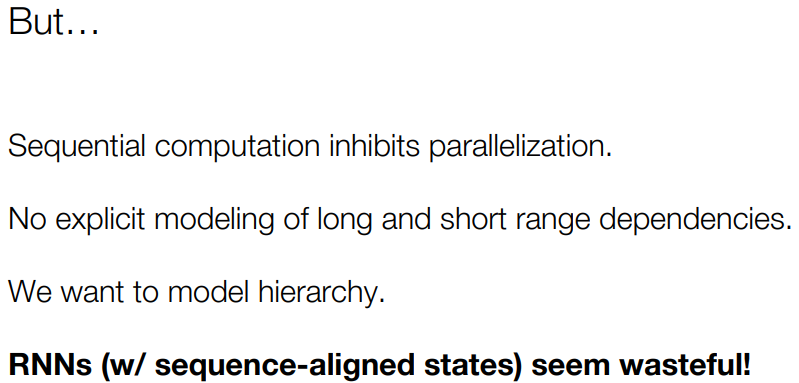
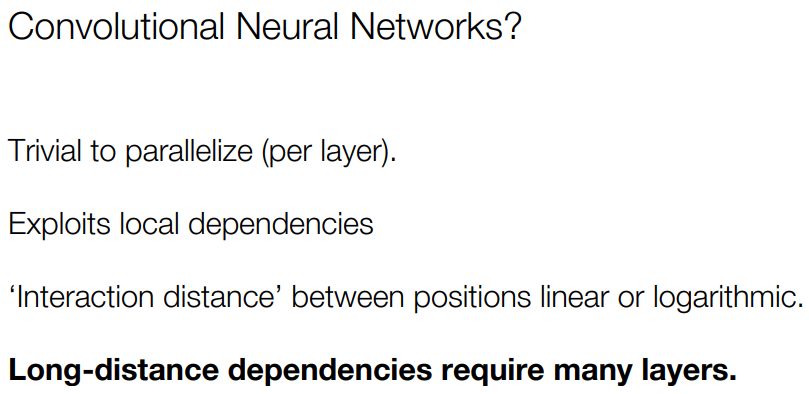
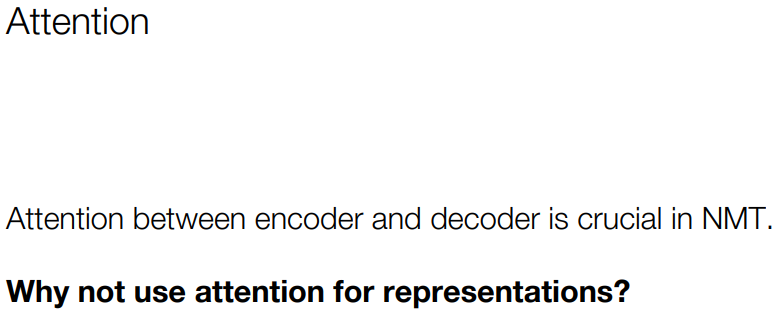
13강에서 많이 정리했으니 추가적인 내용만 더 적도록 하겠다.

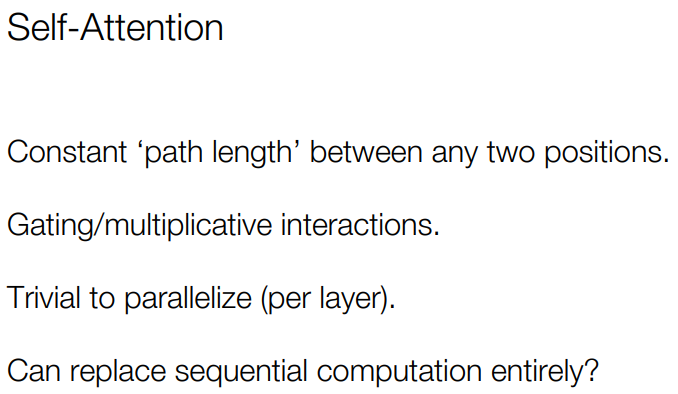


RNN의 단점

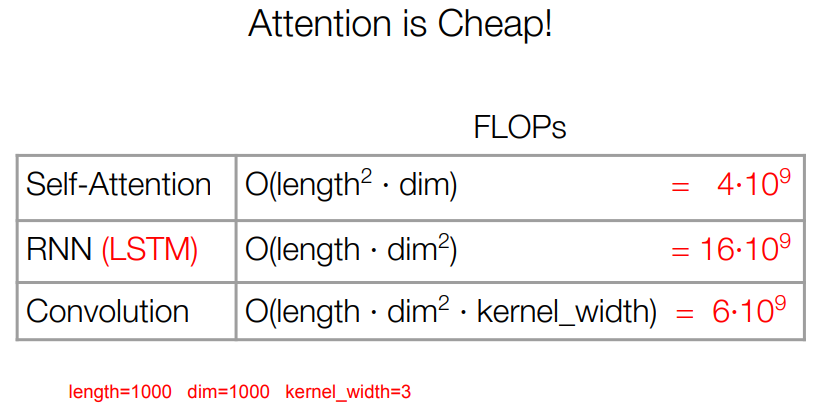


CNN의 단점





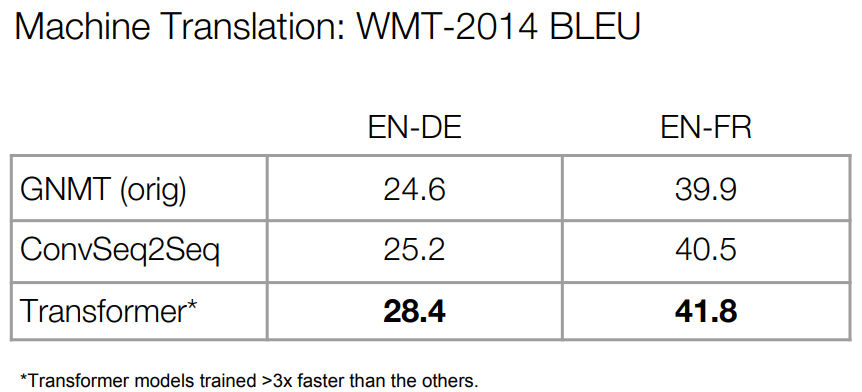
Self-Attention의 장점?



Dim > length 인 경우 더 효과적

Conv – 여러 필터가 적용되며 다른 목표로 값을 도출(하지만 지역적)

Attention – 기존에는 1개의 어텐션만 적용해 average -> 원하는 값 도출 힘듬 -> 따라서 여러 개의 어텐션을 적용하면 각 어텐션이 다른 목표로 값 도출(multi-head attention)



Residual Connection을 이용하면 positional info를 higher layer에 다른 정보와 함께 전달 가능

y = f(x) + x 로 x가 기존에 학습한 정보이다. 따라서 추가적인 정보만 학습하면 된다. => 학습 속도가 빨라질 것 => bottle neck in learning process 해결

f(x)는 나머지(residual) 학습이다. 또한 x가 그대로 skip 되는 거라 연산 증가는 없다.

여기까지 Text Generation (NMT) 였고

이번엔 Image & Music Generation 이다.

Relative Attention And Graphs