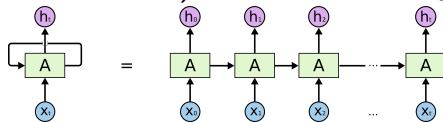


Self-Attention Motivation (1/2)

- Sequences of arbitrary length n need to be processed.
- RNNs make the (time-unrolled) network as deep as n .



- CNNs allow to trade kernel size k and depth for a target “receptive field”:

