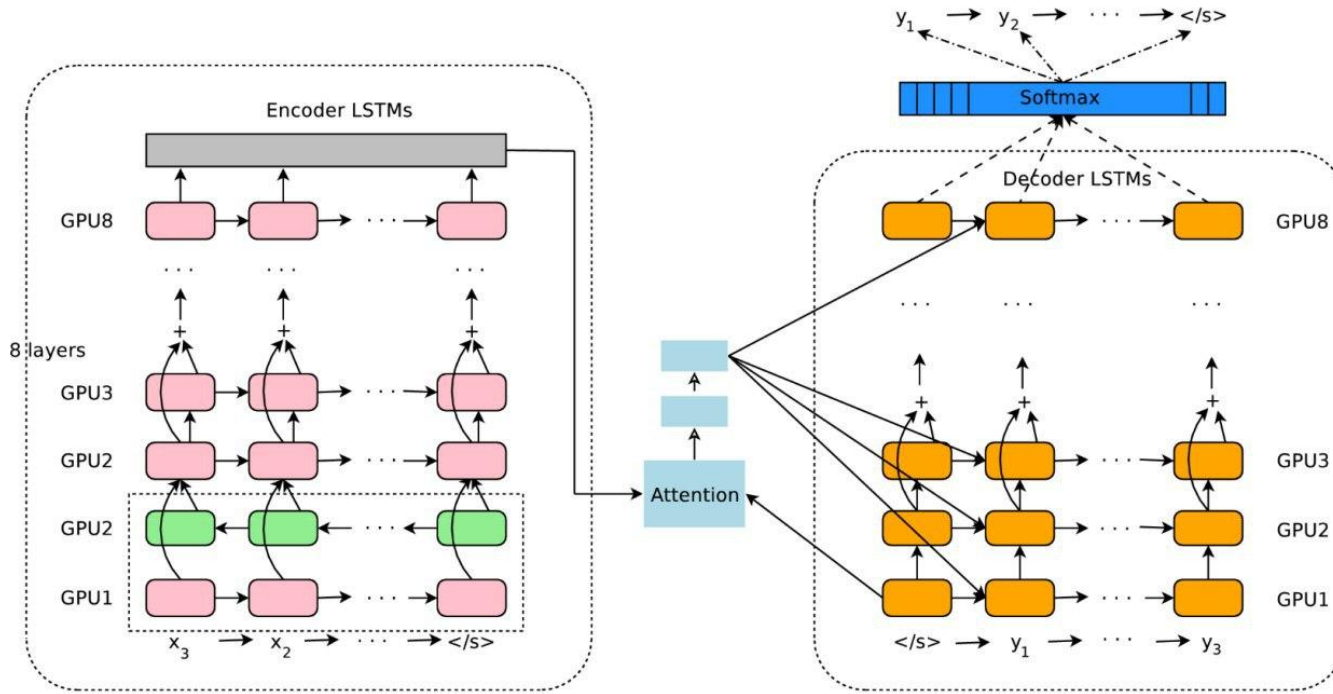


Encoder / Decoder Architectures

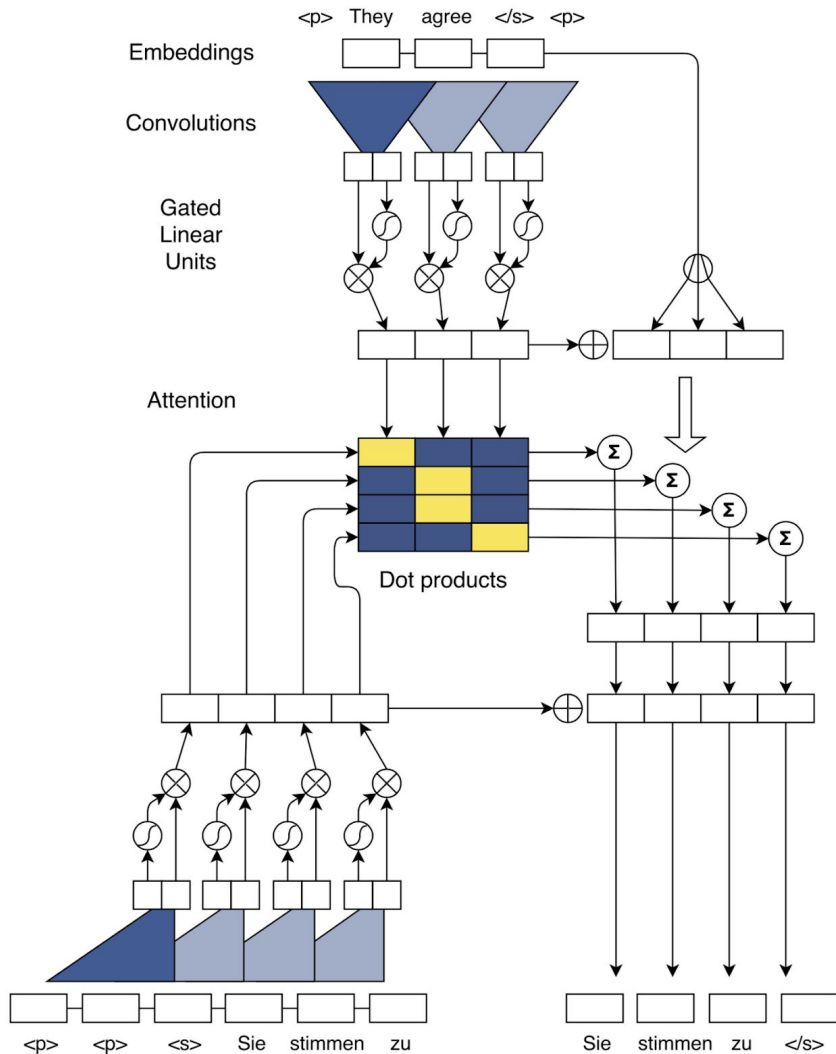
● Google's Neural Machine Translation System

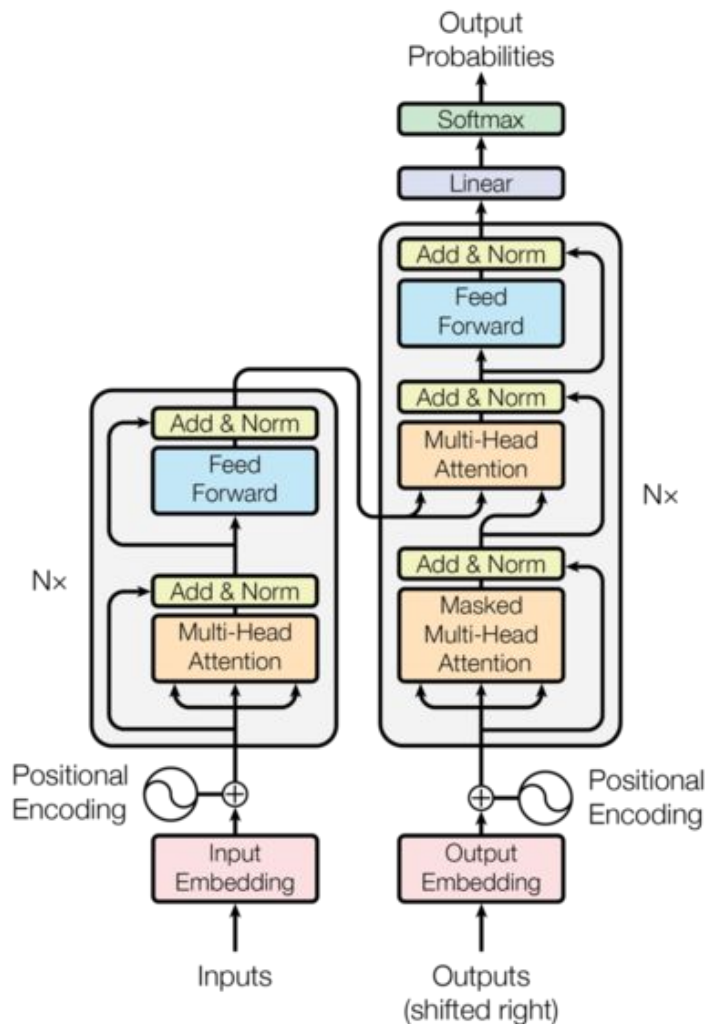


- Model is deep
- Bi-directional encoder for first layer
- Parallelism

Convolutional Sequence to Sequence Learning

- 9x times faster than RNN
- CNNs can outperform RNNs on big benchmarks
- Multi-step attention





- Attention Is All You Need

- Outperforms all previous SotA in quality and speed of training
- Encoder is highly parallel