**Convolutional Sequence to Sequence Learning**

**Jonas Gehring**

**Michael Auli**

**David Grangier**

**Denis Yarats**

**Yann N. Dauphin**

Facebook AI Research

**Abstract**

The prevalent approach to sequence to sequence learning maps an input sequence to a variable length output sequence via recurrent neural net- works. We introduce an architecture based entirely on convolutional neural networks.1 Com- pared to recurrent models, computations over all elements can be fully parallelized during training to better exploit the GPU hardware and optimization is easier since the number of non-linearities is fixed and independent of the input length. Our use of gated linear units eases gradient propaga- tion and we equip each decoder layer with a separate attention module. We outperform the accuracy of the deep LSTM setup of Wu et al. (2016) on both WMT’14 English-German and WMT’14 English-French translation at an order of magni- tude faster speed, both on GPU and CPU.

**摘要**

序列到序列学习的流行方法是通过递归神经网络将输入序列映射到可变长度的输出序列。我们引入了一个完全基于卷积神经网络的架构，与循环模型相比，其在训练期间可以完全并行化所有元素的计算，以更好地利用GPU硬件，并且由于非线性的数量是固定的且与输入长度独立，因此它的优化更容易。我们使用门控线性单元简化了梯度传播，为每个解码器层配备了一个单独的注意模块。我们的表现在设置的准确性上优于Wu等人的深LSTM。在WMT'14英语 - 德语和WMT'14英语 - 法语翻译任务上，GPU和CPU上都能够提速。