Why self-attention?

①每一层的计算复杂度

假设输入序列长度为 n，序列维度为 d，k 为卷积操作卷积核的尺寸。则对于 self-attention 来说，