12 recent approaches in various forms such  
as neural machine translation [ 25, 34], image captioning [ 28] and speech recognition [ 39]. They  
allow modeling dependencies without regard to their distance in the input or output sequences, which  
makes them more expressive than their counterpart recurrent and convolutional layers.  
The authors introduce the Transformer: a new network architecture based solely on attention mechanisms, which eliminates the need for recurrent or convolutional layers. This model improves upon existing sequence transduction models by utilizing attention mechanisms, which model dependencies without regard to distance in the input or output sequences. The authors demonstrate the effectiveness of the Transformer through experiments on two different machine translation tasks, achieving state-of-the-art results with significantly less training time and computational costs. They also show that the Transformer can be applied to other tasks, such as English constituency parsing, with promising results. The paper also provides proper attribution to the authors and their contributions to the research.