Standard Transformer Model

Multi-Head Attention

Scaled Dot-Product Attention