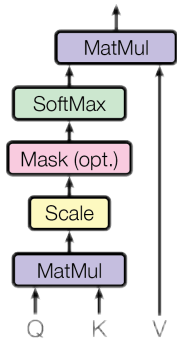Scaled Dot-Product Attention

Multi-Head Attention