,

=

Attention(Q,K,V)=softmax ().V

MultiHead(Q,K,V)=[].W˚

ReLU()=max(0,)