

$$\begin{aligned}
& \text{multihead} \left(\mathbf{q}, \{ \mathbf{k}_0, \mathbf{k}_1, \dots \}, \{ \mathbf{v}_0, \mathbf{v}_1, \dots \} \right) \\
&= \sum_i \text{attention} \left(\tilde{\mathbf{T}}_i \mathbf{q}, \{ \mathbf{T}_i \mathbf{k}_0, \mathbf{T}_i \mathbf{k}_1, \dots \}, \{ \mathbf{W}_i \mathbf{v}_0, \mathbf{W}_i \mathbf{v}_1, \dots \} \right)
\end{aligned}$$