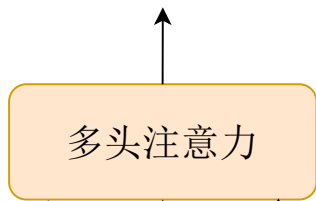


Q K V

多头注意力机制

$$(H_i W_i) \times C_i$$



Q K V

空间缩减注意力机制

$$\frac{H_i W_i}{R_i^2} \times C_i$$

$$(H_i W_i) \times C_i$$