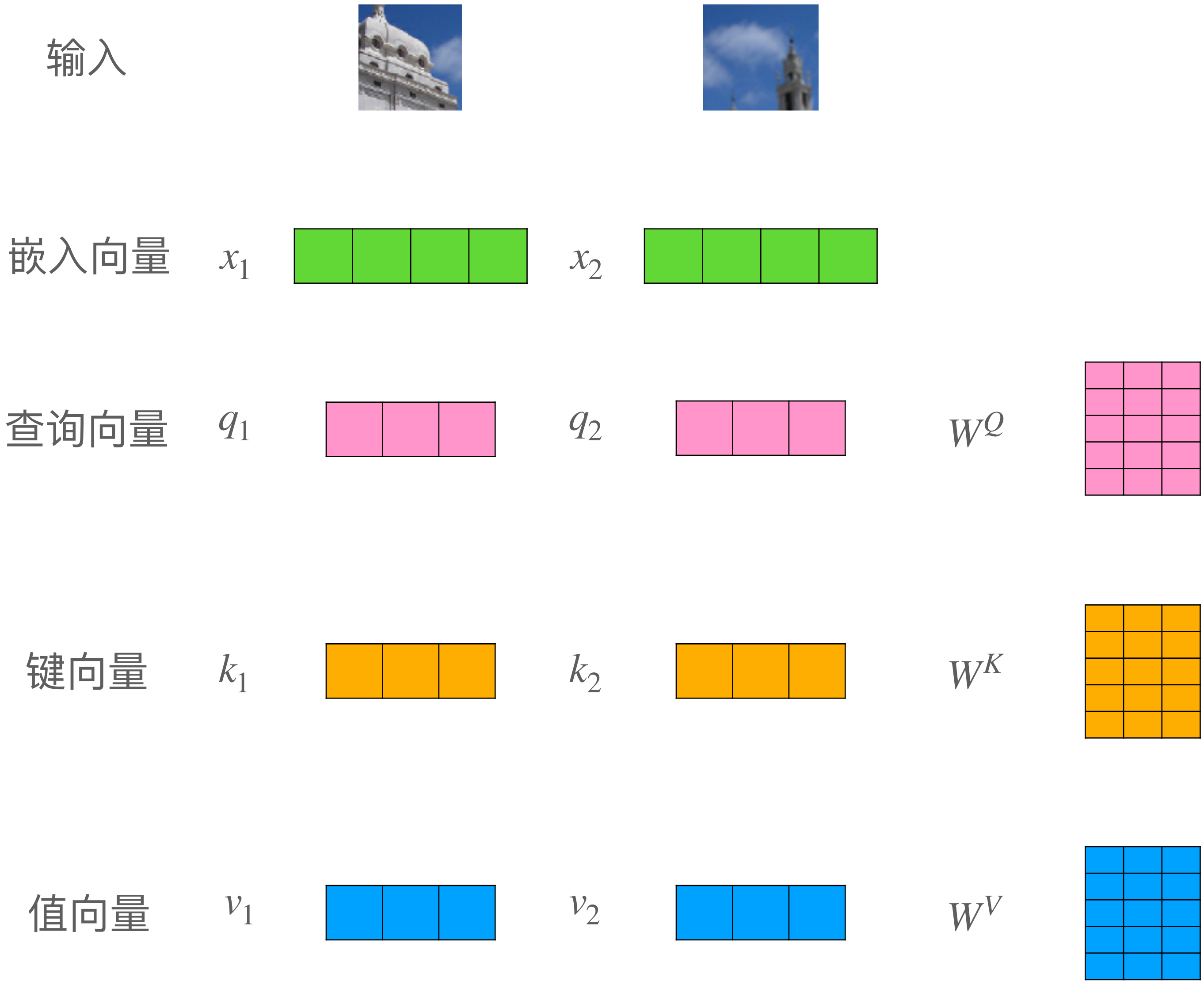


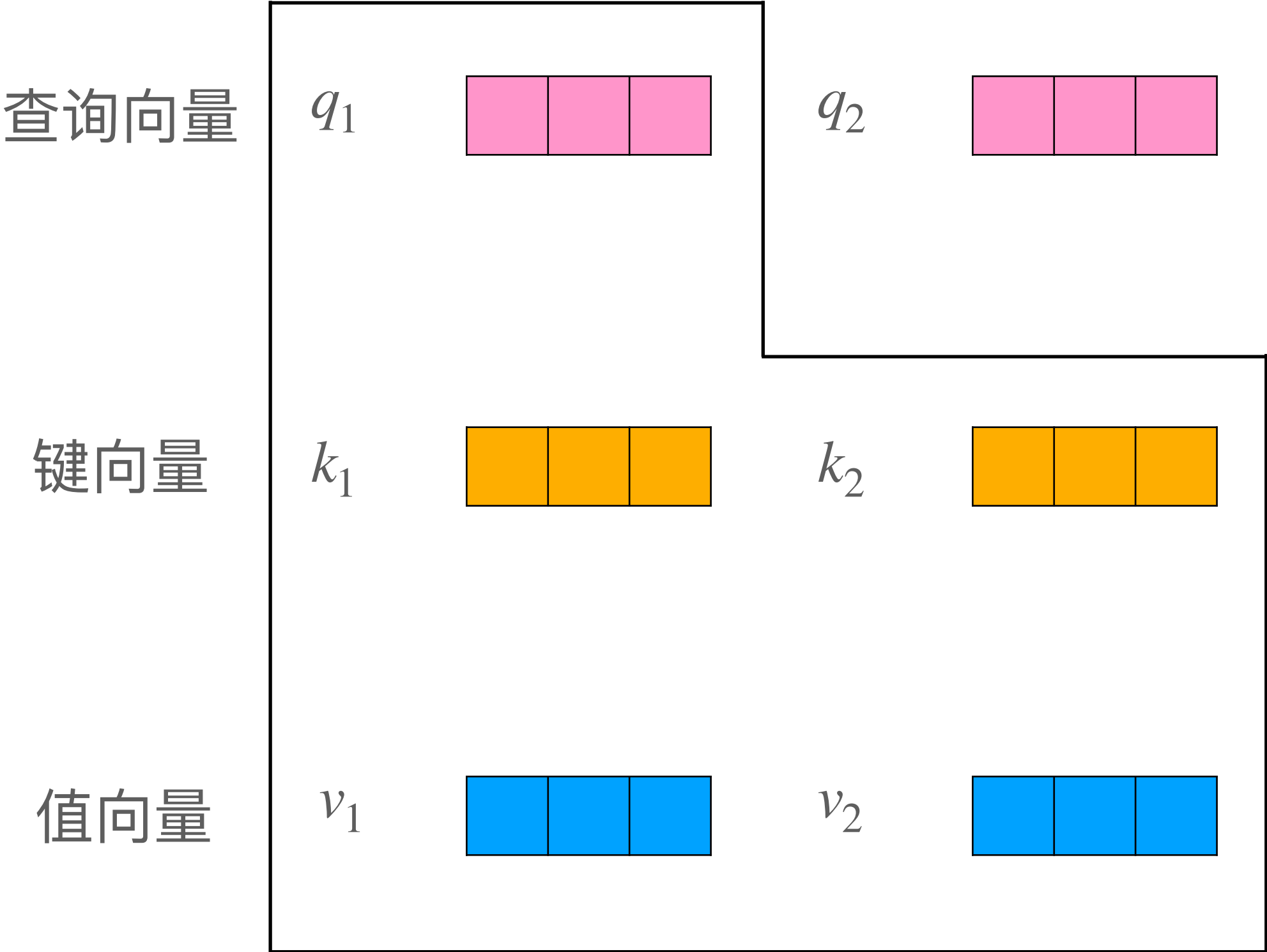
Self-Attention



- Self-attention (自注意力) 机制是注意力机制的改进，更擅长捕捉数据或特征的内部相关性
- 首先通过三个可训练的、随机初始化的权重矩阵将嵌入向量转化为三个低维向量，随后计算自注意力信息

Self-Attention

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \dots\dots\dots \text{式 3}$$



得分 (QK^T)	$q_1 \cdot k_1 = 112$	$q_1 \cdot k_2 = 96$
除以模长开平方 ($\frac{QK^T}{\sqrt{d_k}}$)	14	12
计算占比 ($Softmax$)	0.88	0.12
乘值 ($Softmax \times V$)	$0.88 \times v_1$	$0.12 \times v_2$
求和	$z_1 = 0.88 \times v_1 + 0.12 \times v_2$	