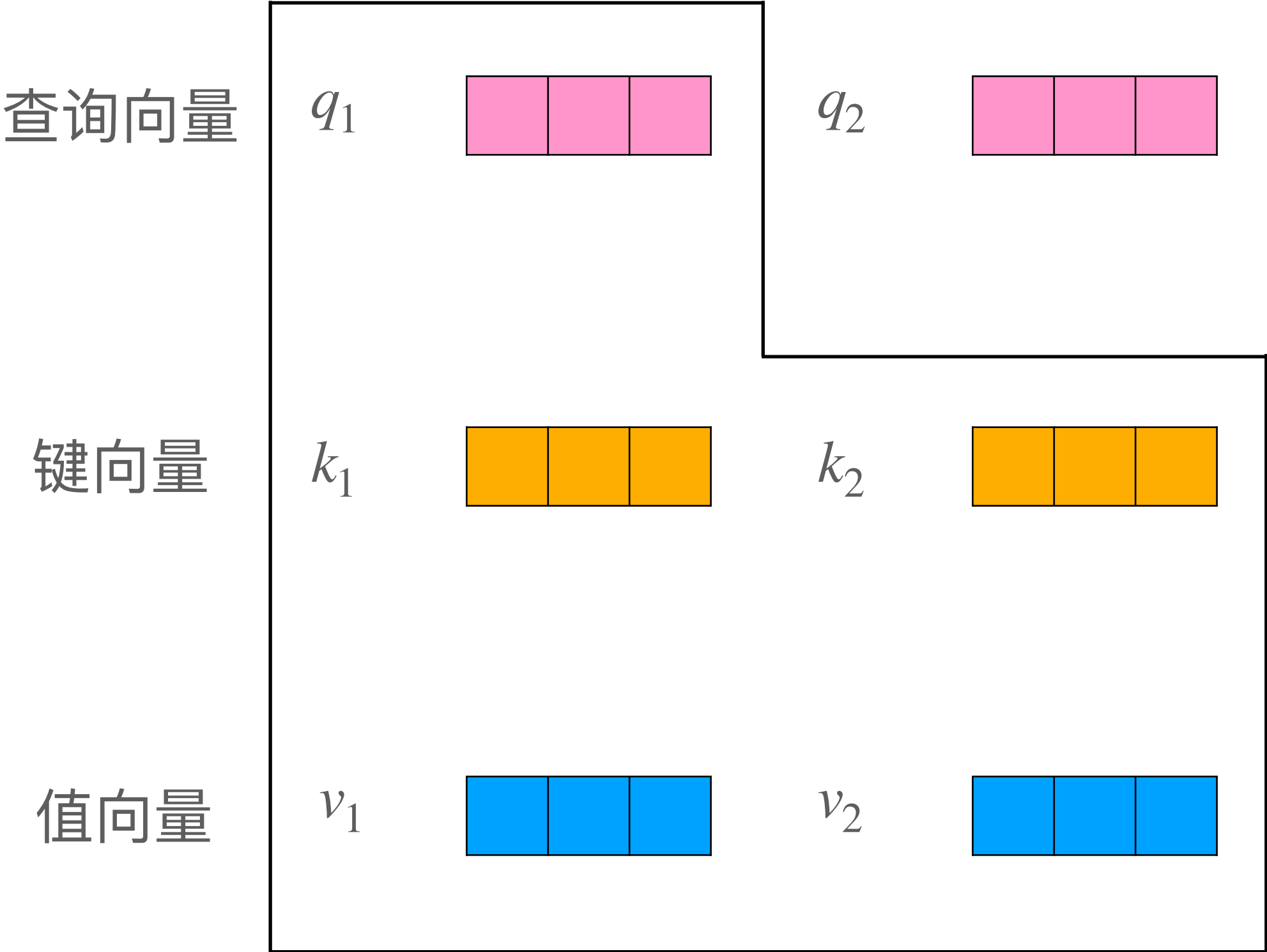


Self-Attention

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

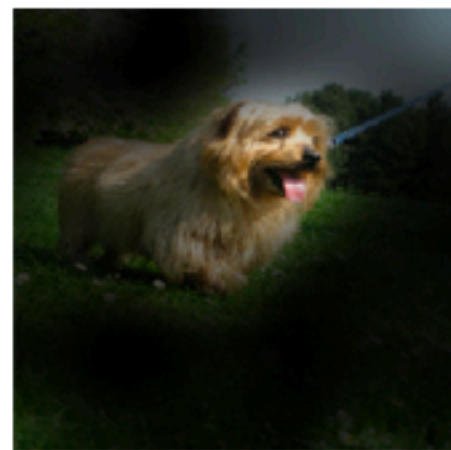
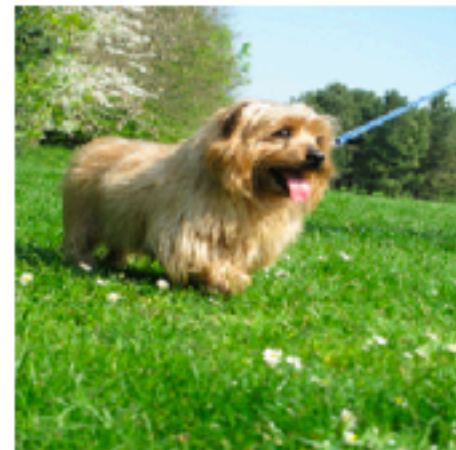


得分 (QK^T)	$q_1 \cdot k_1 = 112$	$q_1 \cdot k_2 = 96$
除以模长开平方 ($\frac{QK^T}{\sqrt{d_k}}$)	14	12
计算占比 ($Softmax$)	0.88	0.12
乘值 ($Softmax \times V$)	$0.88 \times v_1$	$0.12 \times v_2$
求和	$z_1 = 0.88 \times v_1 + 0.12 \times v_2$	

Self-Attention

原始输入

注意力



- 注意力可视化结果显示，模型倾向于将注意力集中于原始输入图像中、与分类信息相关性较高的区域

图 4 注意力可视化