

Transformer Encoder

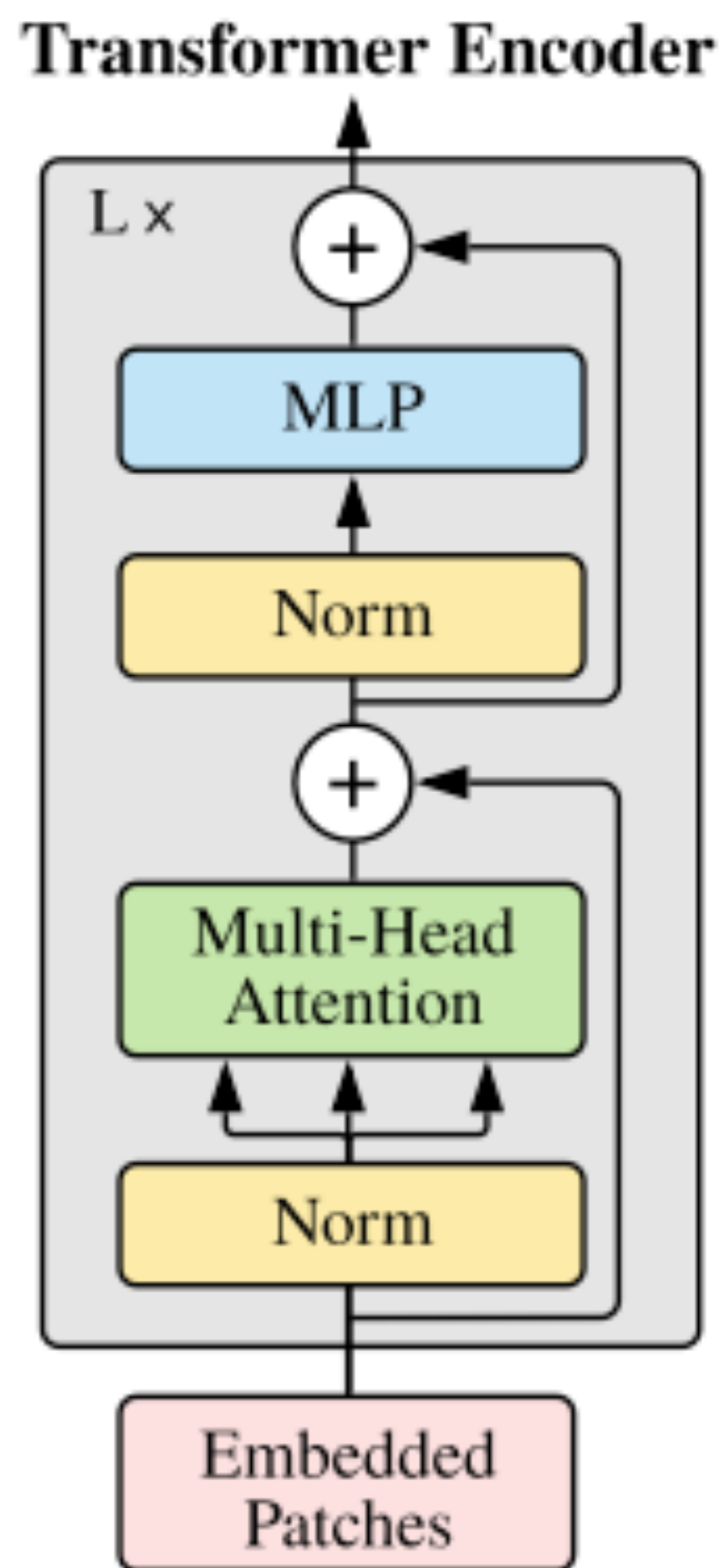
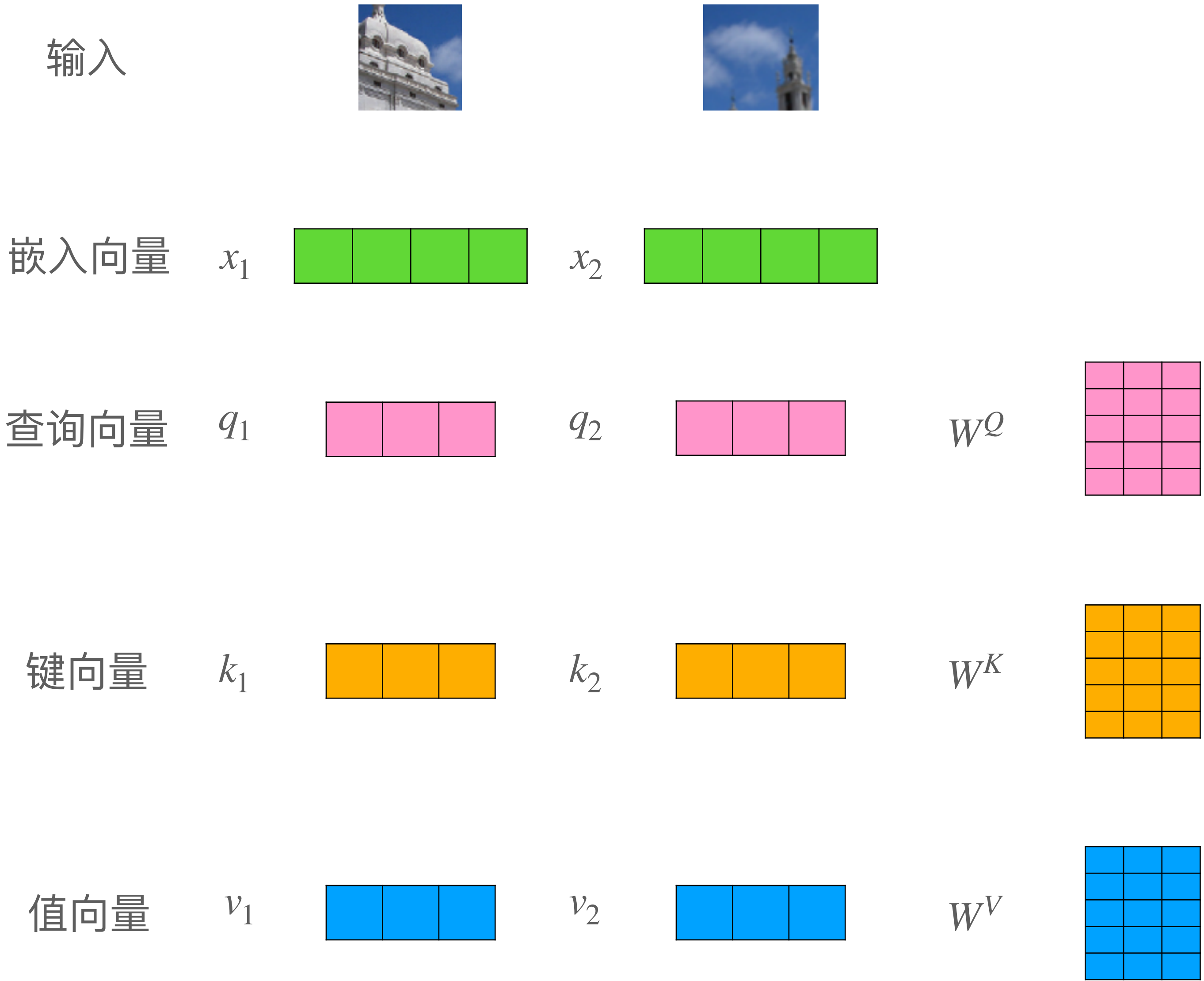


图 3 Encoder 架构

- Norm
对经过转化所得的向量输入进行 Layer Normalization (针对单个样本特征的标准化的)
- Multi-Head Attention
利用多头自注意力机制，并行地计算 Patches 之间的注意力信息并拼接
- MLP
通过 2 层全连接神经网络，实现对输入向量先放大后缩小投射，输出与原输入维度相同的向量

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