

# Position Embedding

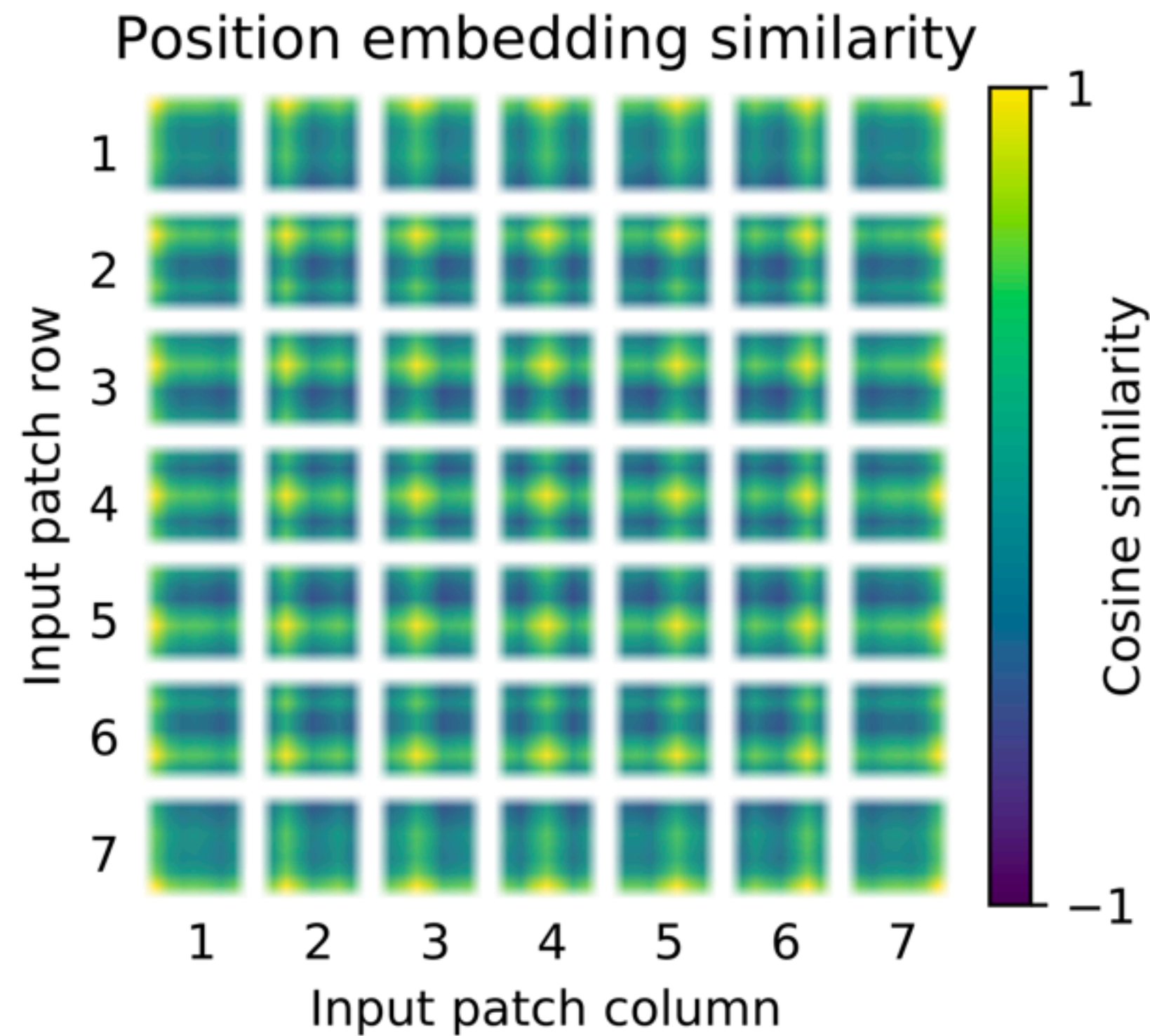


图 2 相似程度可视化

- 对训练后的 Position Embedding 两两间的相似程度进行可视化
- 可视化结果显示所有 Position Embedding 与其自身和近邻向量的相似性较高，可以证明模型能够学习到较准确的二维位置信息

# Transformer Encoder

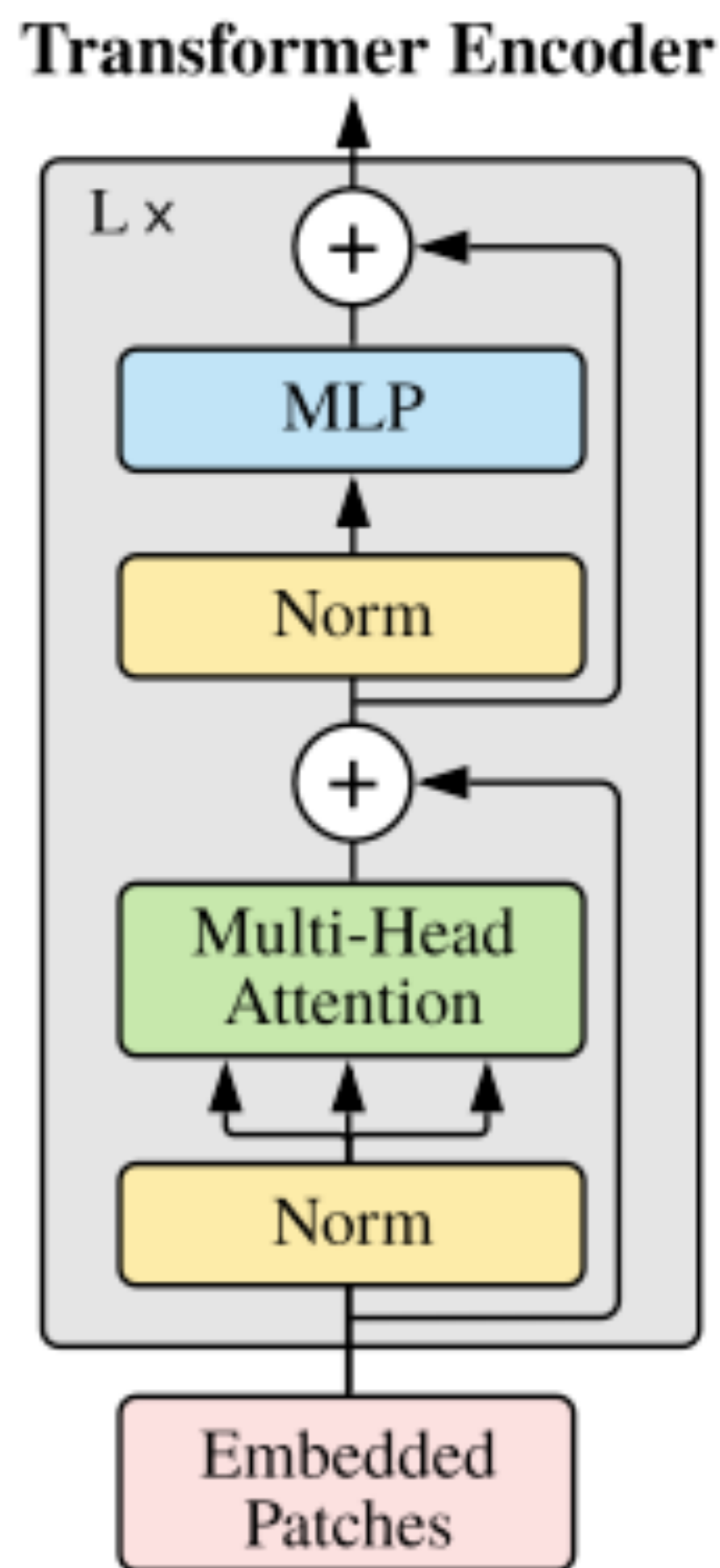


图 3 Encoder 架构

- 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