

文章主要内容

- 图像识别与自然语言处理领域相关工作总结

Transformers 模型的提出^[1]，Bert 模型的提出^[2]；Self-attention (自注意力) 机制在图像识别算法中的应用^[3]，Transformers 模型在图像识别领域中的应用^[4]

- Vision Transformer 模型的提出、相关细节与更多可能的形式

模型总体架构展示；数据预处理与向量化；模型传播过程；混合结构模型试验

- 对比分析模型的表现

与现有 CNN 方法对比；与超参数不相同的同结构模型对比

文章主要成果

- 结合图像识别与自然语言处理领域知识，通过使用全连接神经网络与自注意力机制对图像进行处理与识别，创新性地实现了可应用于图像识别的 Vision Transformer 模型
- 对比分析了 Vision Transformer 模型在不同数据集上的表现，实验结果表明该模型在大型数据集上的表现超过当时的 SOTA (具体任务中表现最好) 模型