

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

 文章概要

 架构与细节

 对比分析