- 文章概要
- 架构与细节
- 对比分析

归纳偏置差异

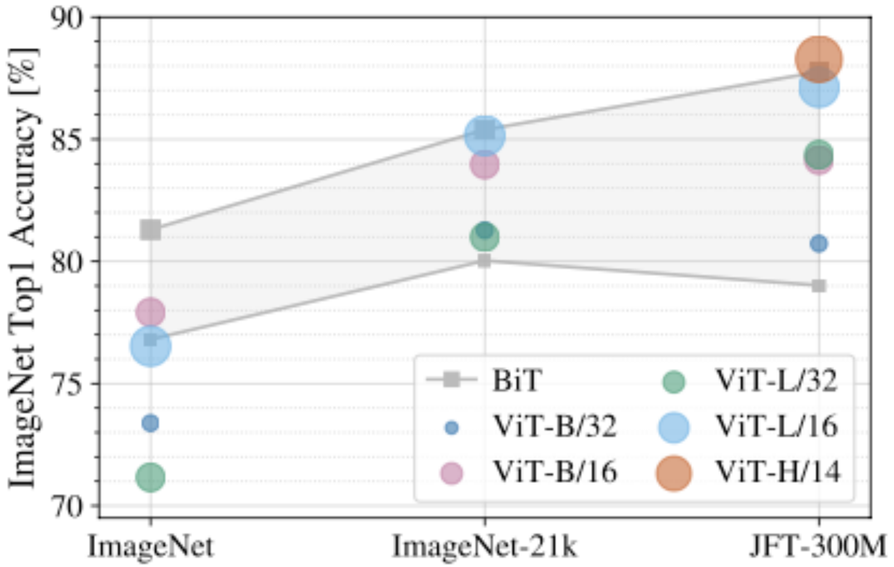

图 4　ResNets 与 Vision Transformer 对比

- CNN (卷积神经网络)

  CNN 架构本身具有两个归纳偏置：局部相关性，即临近的像素是相关的；权重共享，即图像的不同部分处理方式相同。具备归纳偏置的结构能够在数据较少时获得较高的表现

- Vision Transformer

  Vision Transformer 的架构基于自注意力机制，最小化了归纳偏置，因而在小数据集上的表现低于 CNN 模型；但在大型数据集上训练时，基于自注意力机制的架构在精确度与时间成本方面足够媲美甚至超越传统 CNN 模型