Vision Transformers

Mansoor Rezghi*

Abstract

In recent years, transformers have played an important role in advancements in artificial intelligence, especially in the field of NLP. Recently, transformer-based methods, known as vision transformers, have been introduced, aiming to provide a transformerbased architecture for computer vision tasks. These methods treat images as sequences of patches and attempt to integrate information across different layers based on the similarity of patches, which is obtained through the attention mechanism. Recent results have demonstrated the effectiveness of these methods in important applications such as image classification and segmentation. Despite these applications, challenges such as the large number of parameters and the need for very large datasets to demonstrate performance remain. In this talk, we will introduce the fundamental concepts of the attention mechanism and transformer and discuss several key approaches in this field, highlighting their strengths and weaknesses.

*Department of Computer Science, Tarbiat Modares University, Tehran, Iran, `rezghi@modares.ac.ir`