Project Proposal

Student : Ziqiang Wang (ID : 40252657)

Background & Motivation

In recent years, transformer has achieved state-of-the-art performance on many natural language processing tasks. In computer vision, transformer-based models have also shown remarkable performance on various tasks. However, these models are large and computationally expensive, making them unsuitable for deployment on small edge devices such as cell phones. Knowledge distillation is a widely used method for compressing huge models by a teacher-student framework, which transfers knowledge from a teacher model to a small student model. Research on vision transformer distillation is very scarce. This project will explore this direction.

Objective & Methodology

The basic objective of this project is to develop an efficient and tiny image classification model by Knowledge Distillation (KD) for edge devices that can match the accuracy of a large transformer-based model while using fewer resources. As an advanced objective, this project will try to modify the traditional KD framework to further improve the effect. Specifically, the traditional KD trains a student model from a fixed teacher model thus the teacher is not aware of the student's specific need of the knowledge. This project try to optimize the teacher model during distillation in a parameter-efficient manner, to achieve better knowledge transfer.