Parameter Efficient Fine-Tuning of LLMs towards Logical Reasoning from Images

Gugan S Kathiresan - 002756523
Aditya Shanmugham - 002738073
Prabhleenkaur Bindra - 002781738
Maria Anson - 002931419

I. Abstract

The ability to fine-tune LLMs to different domains has enabled organizations and researchers to adapt their foundation models to their liking. In recent days, Parameter Efficient Fine-Tuning (PEFT) has made this possible on lower computational requirements compared to training LLMs on new domains from scratch. Fine-tuning LLMs have various applications, but a simple way of putting it would be, teaching your LLM to talk in the tone, vocabulary and structure. In this study, we attempt to extend these benefits to the image modality. Visual Question Answering is a popular domain that combines Computer Vision and NLP. We attempt to employ VQA models to identify the components of an image, and employ an LLM adapter to modify the results according to the logical reasoning of the user input question. For example, if a user needs to know whether a food item in an image is healthy or not, or whether a particular dish is vegetarian or not. Essentially, the proposal is to finetune an LLM that identifies the context and answers logically to the user's questions. We intend to explore various finetuning techniques and datasets to achieve this task.