Interpretable Vision Transformers: A Multimodal Framework for Visual and Textual Explanation of Image Classification

Abstract

This project aims to develop an explainable AI system that combines Vision Transformers (ViTs) language models to interpret images. The ViT is expected to classify classes while using attention rollout to identify most influential regions in its predictions. These attention maps will be overlaid on the original images and passed to a vision-language model, which will generate explanations describing the significance of the highlighted regions. This multimodal framework enables both visual and textual insight into model decisions. The project will deliver a codebase, visualization outputs with technical reports.

Academic Project Requirements:

1) Required No. of student(s) for academic project: 1

2) Name of course with branch/discipline: B.E./B.Tech. Mechanical Engineering

3) Academic Project duration:

(a) Total academic project duration: 8 Weeks
(b) Student's presence at IPR for academic project work: 3 Full working Days per week

Email to: tripathi@ipr.res.in[Guide's e-mail address] and
project_me@ipr.res.in [Academic Project Coordinator's e-mail address]

Phone Number: 079 -4108 [Guide's phone number]