Jason Jordan: Implementation of vision transformer for assessment of image content

The purpose of this hands-on project is to classify images containing American football-related scenes. This is an evaluation of transformer models to determine suitability for use in the CSC 5542 class project, which seeks to analyze football player performance and mobility data to generate a scouting report to assist teams with game preparation and strategy modification. A vision transformer (ViT) was utilized to analyze football-related images, such as live-action video, game play still shots, stadium scenery, and/or play formation diagrams.

The ViT used in this project was the ‘Vit Large Patch16 384’ which was pre-trained on a dataset of over 14 million images for classification into one of the 1,000 ImageNet classes. A video link is provided below that shows the operation of the program, which creates a Streamlit application to classify a given image and allows a user to upload an additional image for classification.

The application was somewhat successful in that the first image was classified as a “football helmet” which is partially correct given that players in the scene were wearing helmets. However, this may not be suitable for the final research project as a more thorough scene description is desired instead of predicting a single class from the ImageNet classes, many of which are unrelated to American football. An additional image of play formation diagram was also uploaded in the video and the result was even worse, with the model classifying the image as an “analog clock.” This project demonstrates the ability to deploy a ViT on a user-end application, but a more powerful model, perhaps with additional fine-tuning, will be necessary for better scene description.

Video Link:

<https://umsystem.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=5bd5898a-6a13-4531-b11a-b28b00129987>

Colab notebook:

<https://colab.research.google.com/drive/1OxoevcOoYx5cksIoHuDnLI71K6XQRLde?usp=sharing>

README Instructions for ViT:

1. Save notebook as ViT\_app.py
2. Open terminal and change directory to file location using ‘cd C:\file\_location’
3. Run ‘streamlit run ViT\_app.py’ to start the application