Project Proposal

Hugo Garrido-Lestache, Vamsi Sudersanam, Sydney Balboni

Objective

Our goal is to create a model which can detect Ai-generate images vs real images.

Research question

CHAT-GPT : What are the limitations and challenges in accurately differentiating between AI-generated and real images, particularly considering advancements in AI technology?

HUGO : Is ai generated content currently to the point which it cannot be differentiated to human generated content?

Hypothesis

HUGO: Ai-generated images are not realistic enough to be able undetected when compared to real images. Because of this it is possible to create a model which is capable of correctly classifying real vs fake images with a high accuracy (95%).

CHAT-GPT: Hypothesis: AI-generated images possess detectable deviations from real images, enabling the development of a classification model capable of accurately distinguishing between the two categories. This model, employing various techniques, can achieve a classification accuracy significantly higher than random chance, reflecting inherent differences in image characteristics.

Initial Ideas

For model we could create a CNN, as it handles spatial data well.

We could do a bootstrap on our testing results to see how different the results we have is different from 50% or more to see if the model has learnt the ability to classify these images and is not just lucky.

Further expansion could be to test our model on newer image generation model such as DALL-E.

Further-further expansion could be to test our model on our own GAN to see if we can fool our own model.

Dataset

- CIFAKE: Real and AI-Generated Synthetic Images - https://www.kaggle.com/datasets/birdy654/cifake-real-and-ai-generated-synthetic-images/data
    - Low quality (1kb)
    - High number of images (120,000)

- AI Generated Images vs Real Images – https://www.kaggle.com/datasets/cashbowman/ai-generated-images-vs-real-images
    - High Quality (~600 kb)
    - Low number of images (1,000)