DeepFake Detection

Abhinav Madabhushi, Pierce Ohlmeyer-Dawson

1 Objective

1.1 Application

The goal of this Project is to predict whether a given image is AI-generated or Real. This has a lot of applications, especially in the advancement of AI. 0.1 percent of all images on the internet are currently AI-generated images, and there is current research going on that suggests that repeated use of these AI-generated images (synthetic data) can lead to unrecoverable model collapse in Generative Models (Ex: ChatGPT). To prevent the collapse of such Generative Models, it is important to be able to classify what images are Real and what images are AI-generated. This task is getting increasingly more complex by the day since AI-generated images are getting closer and closer to Real data as AI models get more advanced.

1.2 Dataset

The dataset used in this project is a Kaggle dataset that contains 60,000 real images from the CIFAR-10 dataset and 60,000 AI-generated synthetic images generated by the Stable Diffusion version 1.4, a text-to-image generation model. The images were equivalent to the images in the CIFAR-10 dataset, making it hard to distinguish as Real or Synthetic. The authors divided the images into training and testing datasets: 100,000 images for training and 20,000 images for testing.

2 Model

The plan for the model is the following:
1) Create a baseline binary classification CNN model with convolution layers, Batch normalization layers, a ReLU layer, a Pooling layer, a Fully connected layer, and finally a softmax layer for binary output.
2) Experiment with different pre-trained models like ResNet-18, ResNet-50, and EfficientNet-B0 to achieve better results.

3 Project Steps

The following are the project steps:
1) Load dataset and preprocess data to fit model requirements
2) Build a baseline CNN model using either PyTorch or Keras (Tensorflow)
3) Train and validate the model

4) Evaluate the model on test data
5) Tune Hyperparameters of the model to make the model better
6) Experiment with pre-trained models for improved performance

4 Project Distribution

Outline of a plan for dividing the work fairly. Example reference citation [1]

References

[1] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.