Anas Puthawala, Jayaram Chandar

Real vs Fake Facial Image Detection

The project that we've decided on is utilizing Siamese Neural Networks and evaluate between two types of images:

1. Facial images generated by AI (i.e. using stable diffusion) and
2. Real faces

The primary goal is to have the network learn and be able to accurately differentiate between the facial images that are generated by AI and those that are real facial images.

Description of the Problem

Currently there are an exuberant number of deep fakes being circulated on the web. Many of these are expertly crafted which make it very difficult for anyone to be able to discern real from fake. In addition, according to Cybernews[1] the number of deep fakes videos circulating on the web doubles every six months, this leads to a dire need for an adequate detection system for common tasks like facial recognition. One of the benefits of a Siamese Network is that it's a few-shot learner, i.e. we can use this architecture in the case that we don't have much data for a complex task such as in our case where we are attempting to detect whether a face is real or fake.

The Data

We utilize a kaggle dataset (https://www.kaggle.com/datasets/ciplab/real-and-fake-face-detection) to obtain the images of real faces vs fake faces. This dataset is originally from Computational Intelligence and Photography Lab from Yonsei University. Here's a brief example of a real vs a fake face and the primary differences that we want the model to learn and be able to detect:

Figure 1.0 - Real v. Fake Face

In this case, we want our siamese network to pick up on the nose and be able to classify with a high certainty that it is indeed a fake image. This was an easy example, there are going to be tougher examples where the differentiator between a real vs a

fake image is not so prominent. There are a total of 2041 samples, 960 fake faces, 1081 real faces.

Methods

We will be starting off with analyzing the data and looking at the distribution of training data for any discernible biases. We are also going to create more samples through data augmentation through methods such as rotation, reflections, random colors and etc. Our first model would be to replicate the model architecture from [2] and then train. If we see that our evaluation metrics on our validation set are low, we improvise on the model to optimize it. We are also planning on creating a secondary CNN model to see how the siamese network compares. Our final model would be us, we are planning on trying to classify some of these images by eye. The final models will be the siamese-CNN network, vanilla CNN model and us.

Responsibilities

Anas will be responsible for:

- Data cleaning & augmentation pipeline
- CNN Network implementation
- Plotting loss curves & assessing overfitting, regularization techniques

Jayaram will be responsible for:

- Siamese Network Implementation
- Evaluation of the model (F1 score, ROC/AUC curve, precision, recall, accuracy)

References

[1] *Report: Number of Deepfakes Double Every Six Months | Cybernews*. https://cybernews.com/privacy/report-number-of-expert-crafted-video-deepfakes-double-every-six-months/.

[2] “Papers with Code - Siamese Neural Networks for One-Shot Image Recognition.” *Siamese Neural Networks for One-Shot Image Recognition | Papers With Code*, https://paperswithcode.com/paper/siamese-neural-networks-for-one-shot-image.