CNN Model for Deepfake Face Identification

Dhruvi Sachin Shah (106122038)
Sanjana Gummuluru (106122106)

Introduction and Background

The original research paper, titled “Deep Fake Face Detection using Convolutional Neural Networks” by Alben Richards MJ et al. addresses the growing challenge of detecting deep fake images generated by AI-driven tools like GANs. The authors developed a CNN-based model to identify deep fakes by analysing visual features from images.
Despite the success of their approach, there remains a need for more robust models capable of handling increasingly sophisticated deep fake technologies.

Approach

The project adopts a data-driven approach, wherein the CNN model is composed of 5 convolutional layers followed by 3 dense layers and a final fully connected layer. A balanced dataset of 1,00,000 images is utilized, split into training and validation sets for performance monitoring, which is evaluated using accuracy, precision, recall and F1-score.

Project Objectives and Modifications

This project aims to enhance the detection capabilities of the original CNN model. The key objectives are to:

- Improve the accuracy of manipulation detection
- Reduce the false positive rate by refining the feature extraction process
- Adapt the model to recognise new types of manipulations that have since been popularised

Methodology

The data from the original paper, consisting of 50k real and 50k deep fake images from the Flickr dataset will be used.
The enhanced model will build upon the existing CNN architecture by introducing:

- Additional Convolutional Layers: to capture more complex patterns and deeper features
- Advanced Data Augmentation: to improve generalisation ability
- Additional datasets: to improve the model’s robustness against newer deep fake techniques

References

Alben Richards MJ, et al. "Deep Fake Face Detection using Convolutional Neural Networks." IEEE, 2020.