Deepfake Detection System - Project Report

Team Members

- Sujal Singh (E22CSEU0770)

- Saatvika Singh (E22CSEU0805)

- Sahil Gupta (E22CSEU0756)

Introduction

Deepfake technology is expected to become a major challenge in cybersecurity, social media, and media forensics. AI-generated fake videos and images will likely increase risks such as misinformation, identity theft, and fraud. This project will focus on developing an advanced system to detect deepfakes using deep learning and computer vision techniques.

Objectives

- We will develop a highly accurate deepfake detection model.

- We will implement deep learning architectures like CNNs and Vision Transformers (ViTs).

- A real-time deepfake detection system will be created.

- The system will ensure high accuracy, precision, recall, and F1-score in detecting deepfakes.

Tools & Technologies to be Used

- Programming Language: Python

- Libraries: OpenCV, TensorFlow, PyTorch, Dlib, Mediapipe

- Frameworks: Keras, PyTorch

- Datasets: FaceForensics++, DFDC, Celeb-DF

- Deployment: Flask/Streamlit (Web), TensorFlow Lite (Mobile)

- Hardware: NVIDIA GPUs for training deep learning models

Methodology

Step 1: Data Collection & Preprocessing

- We will collect FaceForensics++, DFDC, and Celeb-DF datasets.

- Frames from videos will be extracted, and face detection and data augmentation techniques will be applied.

- Preprocessing will ensure uniformity in model training.

Step 2: Model Training

- CNN-based architectures like Xception and ResNet will be implemented.

- Vision Transformer (ViT) models will be explored for deepfake detection.

- Transfer Learning will be used to improve accuracy.

- Models will be trained with real vs. fake labels to classify deepfake content.

Step 3: Model Evaluation

- Performance will be measured using accuracy, precision, recall, F1-score, and AUC-ROC.

- The best model will be selected based on performance on benchmark datasets.

Step 4: Deployment

- A web-based tool will be built using Flask/Streamlit.

- OpenCV will be integrated for real-time video analysis.

- The model will be optimized for mobile using TensorFlow Lite.

Expected Results & Discussion

Expected Results

- The best model (Xception + Vision Transformer) is expected to achieve over 90% accuracy.

- The system will successfully detect deepfake artifacts like blurring and unnatural facial movements.

Challenges to be Addressed

- High-quality deepfakes may be difficult to detect.

- Real-time processing will require high computational power.

- Limited diverse datasets may impact model generalization.

Future Improvements

- We will integrate Explainable AI (XAI) for better interpretability.

- Blockchain technology will be explored for verifying digital media authenticity.

- Advanced GAN detection techniques will be implemented to counter adversarial attacks.

- The system will be optimized for mobile and edge devices.

Conclusion

This project will develop an advanced deepfake detection system that will classify real and fake images/videos with high accuracy. Future enhancements will focus on real-time performance, security, and explainability to ensure the system remains effective against evolving deepfake techniques.

References

1. T. Rössler et al., 'FaceForensics++: Learning to Detect Manipulated Facial Images,' IEEE CVPR, 2019.

2. H. Farid, 'Image Forgery Detection: A Survey,' IEEE Signal Processing Magazine, 2016.

3. J. Kietzmann et al., 'Deepfakes: Trick or Treat?,' Business Horizons, 2020.

4. Official dataset repositories: FaceForensics++, DFDC, Celeb-DF.