CYBER GYAN VIRTUAL INTERNSHIP PROGRAM

Centre for Development of Advanced Computing (CDAC), Noida

Submitted By:

Gufran Khan

Project Trainee, (June-July) 2025

TOPIC NAME

1.Develop application for detection of manipulated Facial Images using AI/ML & Detect Open source technologies.

PROBLEM STATEMENT

- In today's digital age, cybercriminals increasingly exploit advanced image and video manipulation techniques—such as deepfakes and facial tampering—for malicious purposes, including political misinformation, financial fraud, identity theft, and the creation of non-consensual explicit content (fake pornography). The ease of access to deep learning-based manipulation tools has significantly raised concerns over digital trust, media authenticity, and personal privacy.
- This project aims to design and develop an application capable of detecting fake or tampered facial images and videos, focusing on deepfake detection and facial manipulation analysis. The goal is to identify forgeries generated through AI techniques such as autoencoders, GANs (Generative Adversarial Networks), face-swapping algorithms, and other deep learning-based methods.
- In addition, this project will involve a comparative analysis of existing open-source tools and frameworks for face manipulation detection, evaluating them based on accuracy, speed, model architecture, dataset support, and usability.

TECHNOLOGY/TOOLS TO BE USED

- **Deep Learning & ML Frameworks:**
  - TensorFlow (with tf.keras) – model building
  - MobileNetV2 (Pretrained Model)
- **Web Application Development:**
  - Streamlit – building the interactive web UI
- **Image Processing & Preprocessing:**
  - OpenCV – image decoding and resizing
  - NumPy – numerical operations
  - matplotlib – plotting training metrics
- **Data Handling:**
  - ImageDataGenerator (from Keras) – data augmentation and preprocessing
- **Model Saving & Deployment:**
  - .h5 model format – saving trained model
  - app.py – deployment using Streamlit interface
  - predict.py – inference logic
  - train.py – training logic
- **Tools Used:**
  - Visual Studio Code (VS Code) – code development and debugging

ABOUT THE ATTACK/TOPIC/PROBLEM STATEMENT

- In the digital era, cybercriminals are increasingly leveraging advanced AI-based manipulation techniques to create forged multimedia content—commonly known as **deepfakes**. These manipulated images and videos often involve **face swapping, facial expression modification, and voice cloning**, and are generated using sophisticated deep learning methods such as **Generative Adversarial Networks (GANs)**. The accessibility of deepfake creation tools has significantly lowered the barrier for producing hyper-realistic fake content.
- Such manipulated media poses a **serious threat to information integrity**, public trust, and individual privacy. Deepfakes are commonly weaponized for malicious purposes, including:
- **Political manipulation** (e.g., fake speeches or controversial statements by public figures)
- **Financial fraud** (e.g., spoofed identities in scams)
- **Non-consensual pornography** (e.g., fake explicit content of individuals)
- **Misinformation and disinformation campaigns**
- Given the increasing realism and spread of these media, it becomes critically important to **develop robust detection mechanisms** capable of identifying even subtle signs of tampering. This project focuses on **face manipulation detection**, i.e., identifying whether an input facial image is genuine or artificially manipulated.
- We propose a deep learning–based solution that uses a **pretrained MobileNetV2 model** as the feature extractor, fine-tuned to distinguish between real and fake facial images. The application integrates a user-friendly web interface using **Streamlit**, enabling real-time detection and visualization of results. The ultimate goal is to **contribute to digital media authenticity**, help curb the misuse of AI-generated content, and foster trust in online visual communication.

WHAT ARE THE REASONS BEHIND THE PROBLEM(TELL ABOUT THE ISSUES WHY THIS PROBLEM/ATTACKS ARE HAPPENING)

- The rapid rise of deepfake technology is driven by advancements in deep learning, easy access to open-source tools, and the widespread availability of data. These developments have made it increasingly simple for individuals, even without advanced technical skills, to create hyper-realistic fake media content. The key reasons behind the rise of deepfake attacks include:
- **Technological Accessibility**: Sophisticated machine learning models such as GANs (Generative Adversarial Networks) are openly available, allowing anyone to generate manipulated content with minimal effort.
- **Social Media Amplification**: Platforms like Twitter, Facebook, and TikTok allow deepfakes to spread rapidly, reaching millions before verification mechanisms can respond.
- **Lack of Awareness and Detection Mechanisms**: The general public and many organizations still lack effective tools to identify manipulated media, making them easy targets for deception.
- **Malicious Intent and Motivation**: Cybercriminals, political actors, and malicious users exploit deepfakes for financial fraud, misinformation campaigns, identity theft, blackmail, and the creation of non-consensual explicit content.
- **Data Abundance**: Massive amounts of facial images and videos are available online, often without consent, making it easy to train deepfake models on real people's identities.
- These factors combined have created a perfect environment for the misuse of AI-driven manipulation tools, highlighting the urgent need for robust detection and mitigation strategies.

SUGGEST SOME POSSIBLE SOLUTIONS/COUNTERMEASURES

- To effectively counter deepfake threats, a multi-layered approach that combines technological, legal, and educational strategies is essential. Below are powerful and practical solutions::
- **AI-Powered Detection Systems**
  Deploy advanced deepfake detection models using deep learning architectures (e.g., CNNs, MobileNetV2) trained to identify subtle visual and audio artifacts not perceptible to the human eye.
- **Digital Watermarking & Media Authentication**
  Use cryptographic watermarks or blockchain-based verification systems to track the origin and integrity of audio-visual content, ensuring authenticity.
- **Regulatory Frameworks**
  Establish and enforce legal regulations to criminalize malicious deepfake creation and distribution, especially for misinformation, impersonation, or harassment.
- **Platform-Level Moderation**
  Social media platforms must implement automated content screening tools and flag suspected deepfakes before they go viral.
- **Public Awareness & Digital Literacy**
  Conduct awareness campaigns and integrate media literacy into education systems so users can critically evaluate and report manipulated content.

Github repo of my Project /project structure

- deepfake-detection/
- |
- ├── app.py # Streamlit frontend interface
- ├── train.py # Model training script
- ├── predict.py # Image prediction logic
- ├── model/
- | └── deepfake_model.h5 # Trained MobileNetV2 model
- ├── data/
- | ├── train/ # Training dataset
- | └── test/ # Testing dataset
- ├── utils/
- | └── preprocess.py # OpenCV + NumPy image preprocessing
- ├── requirements.txt # Python dependencies
- └── README.md # Project documentation
- GITHUB REPO LINK-- https://github.com/iGufrankhan/DeepFake_Face_Detection_Model

THANKYOU