Riphah International University, Faculty of computing

Deepfake vs Real Image Detection Project

Submitted By:

- Muhammad Sadeem Choudhary –
- SAP ID: 66385
- Muhammad Yousaf –
- SAP ID: 66160

Project Title

Deepfake vs Real Image Detection System

Problem Statement

Deepfake technology enables highly realistic manipulation of facial images. These deepfake images can be misused for misinformation, identity theft, political manipulation, and cybercrime. The increasing availability of AI tools makes it harder for humans to manually detect such manipulations. Thus, creating an automated system for distinguishing real from deepfake images has become essential.

Dataset Description

The dataset used is the 'Deepfake and Real Images' dataset available on Kaggle by Manjil Karki. It includes two folders: Real Images: Original face images without alteration.

Fake Images: Deepfake-generated images created using various GAN-based face manipulation techniques. The dataset contains high variability in lighting, expression, and resolution, making it ideal for training CNN models.

Sample Dataset Structure

Root Folder / ■■■ Fake / ■■■ Real / Each folder contains hundreds of face images in JPG format.

Project Objectives

Build a detection model to classify images as Real or Deepfake. Preprocess images using normalization, resizing, and augmentation. Train a CNN model (EfficientNet/Xception/custom CNN). Analyze image artifacts common in deepfake generation. Provide a classification interface for user uploaded images.

Proposed Methodology

1. Data Loading & Preprocessing: Resize all images to 224×224. Normalize pixel values. Apply augmentation (flip, rotation, noise) to improve generalization.

2. Model Building: Choose EfficientNet or Xception. Add dropout and dense layers for binary classification.
3. Training: - Use
4. 80% training, 20% validation. Train for 10–20 epochs depending on performance.
5. Evaluation: Metrics: accuracy, precision, recall, F1-score.
6. Deployment: - Use Gradio for real-time testing interface.

Expected Results or Outcomes

A trained CNN capable of above 75 - 85% accuracy. Clear visual explanation of classification probability. Ability to detect deepfake characteristics such as blending artifacts and unnatural textures. Deployment-ready interface for testing.