DeepDetect : Detect Fake and AI-generated Images
Mid-Semester Project Report

Himang Chandra Garg Dasari Sai Harsh Nishil Agarwal Piyush Narula
Indraprastha Institute of Information Technology, Delhi
{himang22214, dasari22144, nishil22334, piyush22354}@iiitd.ac.in

1. Abstract

Our project seeks to address the critical issue of detecting deepfake images and AI-generated content, which have become increasingly sophisticated and challenging to distinguish from real media. The goal is to enhance the ability to discern genuine content from fabricated images, thereby reducing the potential for misinformation.

2. Introduction

2.1. Problem Statement

The recent rise in AI content, particularly deepfake images and videos, has raised serious concerns about the integrity of information online. This synthetic media is often indistinguishable from real content, hence making it very difficult for an average person to tell the truth—an aspect with large implications in public trust, privacy, and security.

3. Literature Survey

Detecting fake images with machine learning is a difficult and growing research field because image editing tools are becoming more accessible and user-friendly. Recent studies have focused on creating automated systems utilizing machine-learning methods in order to identify counterfeit images. This review of literature outlines the current research in this area and addresses the difficulties and future paths ahead.

3.1. Detecting Fake Images Using Machine Learning

This article delves into strategies for detecting manipulated images using both traditional image processing and machine learning approaches. The writers suggest a combination model that uses CNNs to extract features like color patterns, texture, and statistical characteristics from images. Next, these characteristics are inputted into a classifier that differentiates between authentic and altered images. The suggested model attains high accuracy and can be utilized in social media content moderation, news verification, and forensic investigations.

3.2. Detecting Deepfake Images Using Deep Learning Techniques and Explainable AI Methods

The effectiveness of deep learning models in differentiating real images from AI-generated ones is investigated in the paper "Deep Learning for Image Authentication: A Comparative Study on Real and AI-Generated Image Classification" by Gaye Ediboglu Bartos and Serel Akyol. The research is centered around two models: Residual Networks (ResNet) and Variational Autoencoders (VAEs). ResNet, renowned for its strong feature extraction abilities, reached an impressive 94% accuracy in distinguishing between real and synthetic images in the CIFAKE dataset, comprising 60,000 real and 60,000 AI-generated images. On the other hand, VAEs, taking an anomaly detection stance, had a lower accuracy of 71% because of their generative nature and less discriminative architecture. The paper highlights how crucial it is to tune hyperparameters, such as batch sizes and epochs, in order to improve model performance. The findings show that ResNet is better for image authentication tasks than VAEs, which had difficulty with the CIFAKE dataset's complexity. In summary, the study emphasizes the importance of choosing the right model and optimizing it for specific tasks to improve the detection of AI-generated images. It indicates that ResNet's discriminative capabilities make it the best option for these tasks.

4. Dataset

5. Methodology

6. Results and Analysis

7. Conclusion

8. References