Proposal - DeepDetect : Using CV and CNNs to Detect Fake Images and Prevent Misinformation

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

1. Motivation

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.

1.1. Why this project?

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.

1.2. How did you think about this?

We visualized this project after many problems of AI-generated content in work and personal experience and no website or app formally available is solving the challenge effectively. This realization motivated us to create our own system—one that could effectively detect and reduce the risks associated with AI-generated content(specifically images).

2. Related work

2.1. Detecting Fake Images Using Machine Learning

This article utilizes methods such as Convolutional Neural Networks (CNNs) for pattern recognition, feature extraction techniques for identifying statistical properties and textures, and conventional image forensics methods like digital watermarking and error level analysis.

2.2. Deep Learning for Image Authentication: AComparative Study on Real and AI-GeneratedImage Classification

The paper investigates deepfake detection using neural networks like Xception, NAS-Net, MobileNet, and VGG16 with a focus on transfer learning. It employs a Kaggle dataset of real and fake faces and presents a hybrid model combining VGG16 and CNN layers, achieving 94% accuracy. The study emphasizes hyperparameter tuning and k-fold cross-validation, making a notable contribution to cybersecurity and deepfake detection.

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

The project aimed to create a deepfake detection system by pre-processing a large image dataset and splitting it into training, validation, and test sets. Data augmentation enhanced model generalization, while pre-trained models like InceptionResNetV2 and DenseNet201 were fine-tuned and evaluated. The best model's predictions were interpreted using the LIME algorithm for Explainable AI.

3. Timeline

- Phase 1: Find datasets
- Phase 2: Data preprocessing
- Phase 3: Model building
- Phase 4: Model refinement
- Phase 5: Model optimization
- Phase 6: Model testing
- Phase 7: Result analysis

4. Final Outcome

4.1. What are you expecting from this project?

Our first main goal is to establish a high level of accuracy in the model when it can distinguish between real and AI-generated pictures even in the case of deep fakes. We hope to use deep learning and machine learning techniques to find patterns and irregularities in this synthetic content that partition it from real content. This model, therefore, should have very high precision and recall and hence should be dependable. Additionally, we plan to implement a basic front-end interface that will make this detection system accessible, addressing the current lack of formal tools or platforms dedicated to this purpose.

4.2. What do you want to contribute to this idea?

We aim to develop a system for detecting deepfakes and AI-generated images. We will focus on machine learning techniques to accurately identify and analyze synthetic content. This project addresses the current lack of formal solutions for deepfake detection and aims to provide a reliable tool for combating misinformation.

5. Individual Tasks

- Himang Chandra Garg: Model Assembly and training, Model Evaluation and Optimisation, Result and Analysis
- Dasari Sai Harsh: Literature Review, Model Evaluation and Optimisation, Model Selection and Refinement
- Nishil Agarwal: Data Preprocessing, Literature review, Model Assembly and training
- Piyush Narula: Data Preprocessing, Model Assembly and training, Model Selection and Refinement