Deep Fake Detector Project Report

Introduction

The Deep Fake Detector project aims to develop a deep learning-based solution for detecting deep fake images. Deep fake images are synthesized media that depict events or situations that never actually happened, often created using advanced machine learning techniques. Detecting deep fake images is crucial for combating misinformation and ensuring the authenticity of visual content.

In this project, we explore two approaches for building a deep fake detection model: a Convolutional Neural Network (CNN) and transfer learning using the Xception architecture. The CNN model is trained from scratch, while the Xception model leverages pre-trained weights to improve performance.

Objective

The objective of this project is to develop a deep learning-based solution capable of accurately detecting deep fake images. With the proliferation of deep fake technology, there's a pressing need to mitigate the spread of misinformation and ensure the integrity of visual content. By creating robust models for identifying manipulated images, the project aims to contribute to the ongoing efforts to combat the negative consequences of deep fake technology.

Problem Statement

The rapid advancement of deep learning techniques has made it increasingly easy to create convincing deep fake images. These manipulated images can be used for malicious purposes, such as spreading false information, impersonating individuals, and manipulating public opinion. Detecting deep fake images poses significant challenges due to the sophistication of the forgery techniques and the continuous evolution of deep fake technology. Therefore, the project seeks to address the following key challenges:

1. Developing deep learning models capable of accurately distinguishing between real and fake images.
2. Ensuring the models are robust against various types of manipulations and adversarial attacks.
3. Providing a scalable and efficient solution for real-time detection of deep fake images.

Dataset

deepfake-and-real-images
www.kaggle.com/datasets/manjilkarki/deepfake-and-real-images/data

Methodology

The methodology for the Deep Fake Detection project involves the following steps:

1. Dataset Collection and Preprocessing:
    - Obtain a diverse dataset containing real and fake images
    - Preprocess the dataset to ensure uniformity in size, format, and quality. Apply data augmentation techniques to increase the diversity of the training data and improve model generalization.

2. Model Selection and Training:
    - Experiment with transfer learning using pre-trained models such as Xception, VGG, and ResNet to leverage existing knowledge and improve model performance.
    - Train multiple models on the dataset using appropriate loss functions, optimizers, and regularization techniques to minimize overfitting and enhance generalization.
3. Evaluation and Validation:
    - Evaluate the trained models on validation and test datasets to assess their performance in detecting deep fake images.
    - Measure key performance metrics such as accuracy.

Model Architecture

CNN Model

- The CNN model consists of several convolutional and pooling layers followed by fully connected layers.
- It is trained from scratch using the Adam optimizer and binary crossentropy loss function.

Xception Model

- The Xception model serves as the base architecture, with its pre-trained weights loaded.
- The base layers of the Xception model are frozen, and custom top layers are added for binary classification.
- Transfer learning is performed using the SGD optimizer with momentum and binary crossentropy loss function.

Training and Evaluation

Both models are trained on the training dataset and evaluated on the validation and test datasets. Training histories are visualized to assess model performance and convergence. The final test accuracies are reported for both models.

Training Results

- CNN Model:
    - Training Accuracy (Epoch 10/10): 85.47%
    - Validation Accuracy (Epoch 10/10): 82.38%
- Xception Model:
    - Training Accuracy(Epoch 3/3): 76.38%
    - Validation Accuracy: 70.00%

Test Results

- CNN Model Test Accuracy: 79.61%
- Xception Model Test Accuracy: 66.77%

Conclusion

The Deep Fake Detector project demonstrates the effectiveness of deep learning models in detecting deep fake images. Both the CNN and Xception models show promising results, with the CNN model outperforming the Xception model in terms of test accuracy. Further investigation and optimization may be required to improve the performance of the Xception model.