Abstract

The proliferation of DeepFakes—videos that replace one individual's facial identity with another—poses significant risks, such as the spread of misinformation, fraud, and privacy violations. This project proposes an advanced AI/ML framework leveraging Multi-task Cascaded Convolutional Networks (MTCNN), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models to detect face-swap-based DeepFake videos. By integrating spatial, temporal, and frequency-based analyses, the system ensures robust detection of DeepFake content. Comprehensive reports are generated to explain detected abnormalities, contributing to ongoing efforts in content verification and media authenticity.

KeyWords:

AI-based DeepFake detection, Face-swap manipulation, MTCNN, CNNs, RNNs, LSTMs, Hybrid models, Video forensics.

Introduction / Problem Statement

The advent of DeepFake technology showcases the power of AI but also introduces significant risks, including privacy breaches, fraud, and the dissemination of misinformation [1], [2]. Face-swap DeepFakes, in particular, are increasingly difficult to detect due to their high level of realism and the subtle manipulations they employ [3]. Existing detection methods often fall short when dealing with low-resolution inputs or complex transformations, highlighting a pressing need for more resilient and explainable detection frameworks [4], [5].

To address these challenges, this research proposes a **hybrid AI/ML-based system** that combines Convolutional Neural Networks (CNNs) for spatial feature extraction and Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, for temporal analysis [6], [7]. This hybrid approach enables the detection of both visual inconsistencies and motion artifacts that are common in DeepFake videos [8]. In addition, the system integrates Explainable AI (XAI) techniques to produce **technical interpretability reports**, enhancing transparency and user trust [9], [10]. The proposed framework is evaluated on multiple datasets to demonstrate its robustness against real-world DeepFake manipulations.

Project Category

Artificial Intelligence and Machine Learning with a focus on Media Forensics.

System Analysis

Identification of the Need

The increasing sophistication of DeepFake technology necessitates advanced detection systems. Current methods fail to consistently identify realistic face-swap