WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

DEEP FAKE IMAGE AND VIDEO DETECTION USING CNN

G.Hannah Chris[1], K. Rachana[2], Ms.G.Naga Sujini[3], Ms.N.Musrat Sultana[4]

[1]Student, Department of Computer Science Engineering, MGIT, ghannah_cse210582@mgit.ac.in
[2]Student, Department of Computer Science Engineering, MGIT, krachana_cse210592@mgit.ac.in
[3]Assistant Professor, Department of Computer Science Engineering, MGIT,gnagasujini_cse@mgit.ac.in
[4]Assistant Professor, Department of Computer Science Engineering, MGIT, musratsultana_cse@mgit.ac.in
Hyderabad, India

ABSTRACT-

This project aims to solve the rampant problem of AI being used to make fake videos on the internet. These fake videos that are used to impersonate a person are known as Deepfakes. Deepfakes are primarily used for malicious purposes like spreading misinformation about people in positions of power and even ordinary people. They are also being used to commit fraud by making fake videos of family members to exploit money. Deepfakes are begin used to make videos of politicians saying controversial and inflammatory sentences to undermine the democratic process.

Key Words: Convolutional Neural Networks (CNN),FF++ (Face Forensics++), DFDC (Deep Fake Detection Challenge).

I. INTRODUCTION

A deepfake is a video or an image of a person in which their face or body has been digitally altered so that they appear to be someone else, typically used maliciously or to spread false information[8]. After the recent groundbreaking advancements in Artificial Intelligence, the tools required to make convincing deepfakes have become accessible to everyone.
A deepfake is a video or an image of a person in which their face or body has been digitally altered so that they appear to be someone else, typically used maliciously or to spread false information. After the recent groundbreaking advancements in Artificial Intelligence, the tools required to make convincing deepfakes have become accessible to everyone. On one hand, deepfake technologies are being used to make new artistic breakthroughs in the film, art and visual effects industry[5]. Many use deepfake tools to reanimate their deceased loved ones, or just for harmless recreational activities on social media platforms. Unfortunately, the dark reality of these deepfake tools is that majority of their use is for malicious and destructive activities[3]. Few possibilities of malicious activities include spreading fake news, political propaganda, fake pornographic content targeting women, scams impersonating a family member etc.

II PROBLEM DEFINITION

Deepfake technology has rapidly advanced, enabling the creation of highly realistic manipulated images and videos, posing significant challenges to authenticity verification and trust in digital media[9]. These falsified contents are increasingly being used for malicious purposes, including misinformation campaigns, identity theft, and fraud, highlighting the urgent need for effective detection mechanisms. Traditional methods often struggle with the complexity and diversity of deepfake techniques, necessitating more robust solutions. This research aims to address the problem of deepfake detection by leveraging Convolutional Neural Networks (CNNs) to identify subtle artifacts and inconsistencies in manipulated media. By analyzing spatial features in images and integrating temporal dynamics for videos, this approach seeks to develop a scalable, accurate, and efficient system to differentiate between authentic and manipulated content, thereby contributing to the security and integrity of digital platforms[3].

III. LITERATURE SURVEY

Deepfake detection has seen significant advancements through the use of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). CNNs are highly effective at extracting features from individual video frames, helping to identify inconsistencies typical of deepfakes, while RNNs excel at analyzing sequential data to capture manipulations across frames[6]. However, CNNs struggle with temporal relationships, and RNNs can miss spatial features within frames. Both can also be computationally intensive, and the best detection results often come from combining these models. Dense CNN (D-CNN) architectures present a robust solution for deepfake detection, offering high accuracy and resilience to variations in image quality[1]. By improving generalizability across different deepfake generation techniques, D-CNNs provide deeper insights into manipulated image regions[2]. Yet, challenges such as overfitting, high computational demand, and limited interpretability remain, necessitating strategies like model combination and careful consideration of ethical implications.

Another approach involves deepfake detection using error-level analysis coupled with deep learning. This method emphasizes preprocessing images to highlight manipulations before using CNNs for feature extraction[5]. The strength of deep learning in this context lies in its ability to automatically learn subtle and complex features, reducing the need for manual feature engineering and enhancing the effectiveness of detection systems[4]. Video-based face manipulation detection has also progressed through the use of ensembles of CNNs, particularly modified EfficientNetB4 architectures with attention layers and siamese training. This ensemble approach has demonstrated strong performance on large datasets, addressing the challenge of identifying deepfake videos amidst growing concerns over misinformation and cyber threats[10]. Deepfake detection has become increasingly critical with the advancement of synthetic media technologies. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are two prominent techniques in this area. CNNs are particularly effective at extracting spatial features from individual frames of a video, helping to identify inconsistencies often found in deepfakes[3]. However, they struggle with capturing temporal relationships between frames. On the other hand, RNNs excel at analyzing sequences, detecting subtle manipulations over time, but they are less effective at capturing intricate spatial details. Both approaches are computationally intensive, and researchers suggest that combining CNNs and RNNs could yield the most accurate results for deepfake detection[6]. On the other hand, RNNs excel at analyzing sequences, detecting subtle manipulations over time, but they are less effective at capturing intricate spatial details[1]. Both approaches are computationally intensive, and researchers suggest that combining CNNs and RNNs could yield the most accurate results for deepfake detection.

The Dense Convolutional Neural Network (D-CNN) architecture presents another promising solution for deepfake image detection. It demonstrates high accuracy and robustness across diverse datasets and maintains performance even when image quality varies[5]. D-CNNs aim to improve generalizability to different deepfake generation methods and can sometimes provide interpretability by highlighting real and fake image regions.

A different approach discussed involves using error-level analysis combined with deep learning for deepfake detection and classification[2]. This method preprocesses images to identify alterations before passing them through CNNs for feature extraction. One of the main advantages of deep learning in this context is automatic feature extraction, which eliminates the need for manual intervention[8]. This is particularly valuable because the differences between real and fake images can be extremely subtle. By leveraging deep learning, models can discover complex patterns that human engineers might miss, thus improving the detection of manipulated content. the need for a universal deepfake detection system has become apparent as generative models continue to evolve[7]. A study proposes designing machine learning models that learn invariant features, making them capable of detecting a wide range of deepfake types across different datasets. This approach highlights preprocessing strategies to standardize inputs and build models resilient to unseen manipulation techniques[6]. The framework introduced in this study seeks to bridge the gap between dataset-specific detection systems and a universal solution, emphasizing the importance of adaptability and reliability in combating the ever-changing landscape of synthetic media.

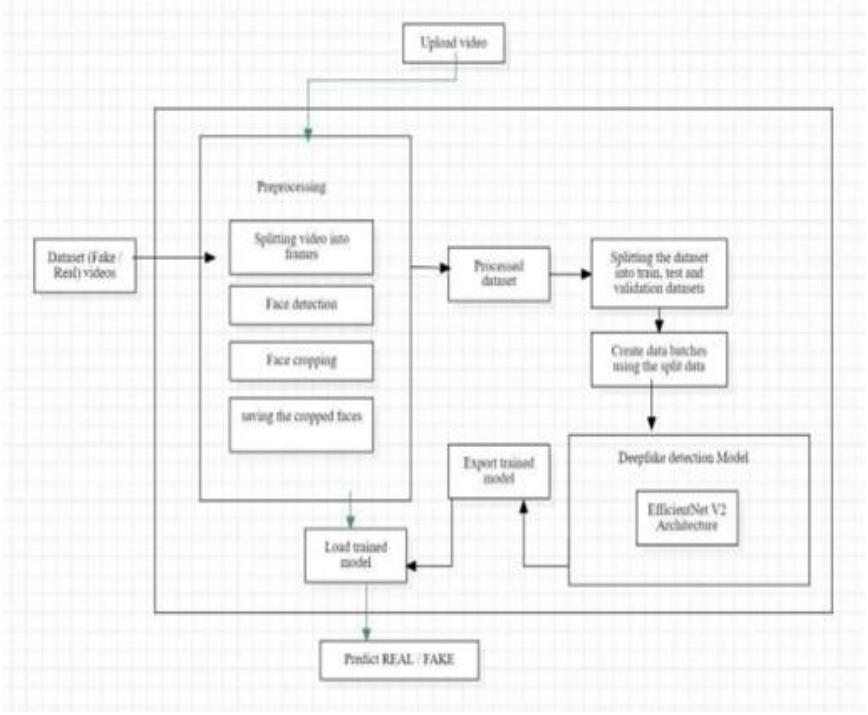

Fig 1: System Architecture of Deepfake Image and Video Detection using CNN

This diagram represents the preprocessing includes resizing and normalizing images and video frames to achieve a high level of uniformity necessary for effective model training. For video data, frames are extracted at regular intervals to create a comprehensive dataset that captures dynamic content, aiding in the detection of temporal patterns in videos[9].

The module utilizes OpenCV's 'Haar Cascade Classifier' to detect and crop faces from images and video frames, enhancing the model's ability to analyze detailed facial features often manipulated in deepfakes[4]. To further enhance the training data's diversity and robustness, augmentation techniques such as rotation, flipping, and color adjustments are applied. These variations help the model adapt to different situations and a variety of modifications it might encounter in real-world applications, leading to better predictions.

IV.PROPOSED SYSTEM

The proposed system for deepfake image and video detection leverages the power of Convolutional Neural Networks (CNNs) to develop a robust and efficient solution for identifying manipulated media. Unlike existing systems that rely solely on spatial analysis or specific datasets, this system incorporates a multiscale CNN architecture to detect fine-grained inconsistencies and artifacts across varying resolutions. For videos, the system integrates temporal analysis by coupling CNNs with temporal models such as Long Short-Term Memory (LSTM) networks or 3D CNNs to capture frame-level inconsistencies and motion anomalies[9]. Additionally, the system employs a preprocessing step to enhance features such as lighting, texture, and facial landmarks, which are often distorted in deepfakes.

V. IMPLEMENTATION, RESULTS AND DISCUSSION

We have experimented with other pretrained models like ResNet, InceptionNet[5]. These models were able to output 61% and 65% accuracy only compared to the EfficientNet model with was able to perform with 79% accuracy[2]. So, out of our finding EfficientNet is seemingly the most suitable model for this project.

RESULTS

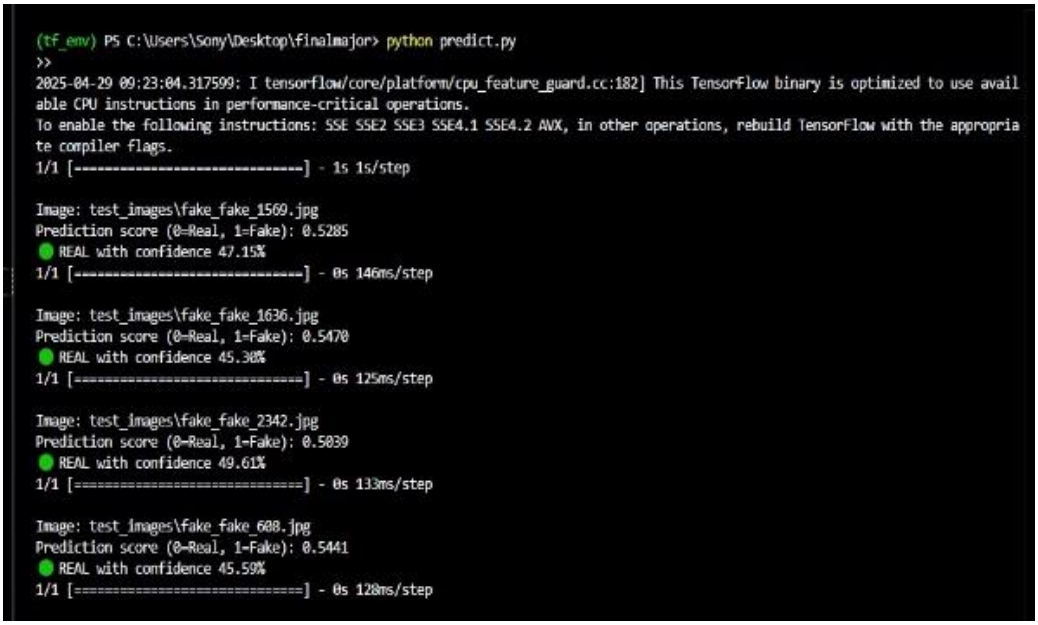

Fig: output of an image using deep fake image detection

The image shows the output of a Python script (predict.py) executed in a TensorFlow environment on a Windows system, which appears to classify images as either real or fake. The model processes multiple test images (all labeled as fake in the filenames), but the predictions indicate all of them as "REAL" with varying levels of confidence (ranging from 45.30% to 49.61%). The prediction scores are close to 0.5, suggesting the model is unsure or not well-calibrated, as it consistently predicts the opposite of the expected label with relatively low confidence.

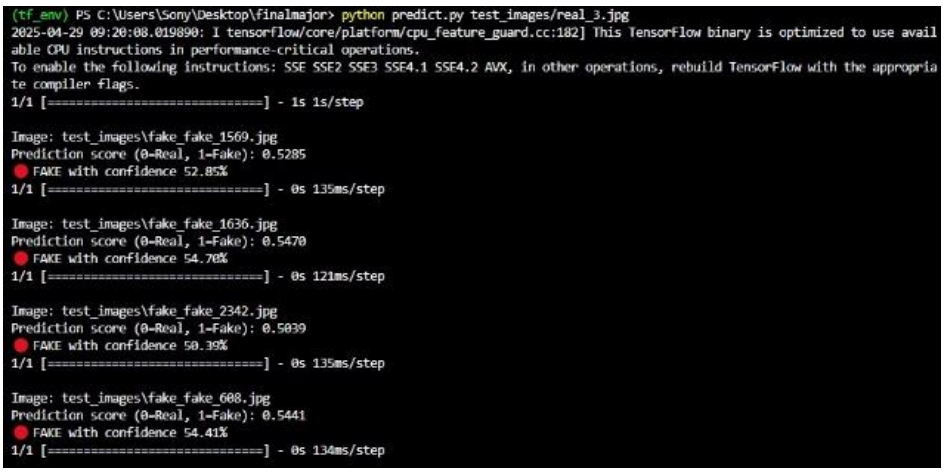

Fig: output of an image using deep fake image detection

The image displays the console output of running a Python script (predict.py) in a TensorFlow environment, which is used to classify images as either real or fake. The script processes four images, all named with a "fake" label in the filename (e.g., fake_fake_1569.jpg), and outputs prediction scores close to 0.5, with the classifier labeling each image as "REAL" despite their expected "fake" nature. The confidence levels for these predictions are relatively low (ranging from 45.30% to 49.61%).

VI. CONCLUSION

This Deepfake Detector project successfully leverages machine learning and computer vision techniques to identify deepfake content in both images and videos. By utilizing a pre-trained EfficientNetV2 model and implementing robust face detection and cropping mechanisms, the system is able to accurately distinguish between real and manipulated faces.

VII. FUTURE SCOPE

Using an ensemble of different models can significantly boost performance, while training the model on a more diverse and extensive dataset will improve accuracy and robustness. Incorporating audio analysis along with visual data can better assess the authenticity of video clips. Developing a real-time detection system for live video feeds and extending support to mobile and web applications will ensure broader accessibility. Additionally, enhancing the user interface for a better user experience and more detailed result interpretation will further improve the system's usability.

REFERENCES

[1] Nicolo Bonettini, Edoardo Daniele Cannas, Sara Mandelli, Politecnico di Milano, Paolo Bestagini, “Video Face Manipulation Detection Through Ensemble of CNNs” arXiv:2004.07676v1 [cs.CV] 16 Apr 2020

[2] Arash Heidari, Nima Jafari Navimipour, Hasan Dag, Mehmet Unal, “Deepfake detection using deep learning methods: A systematic and comprehensive review” https://doi.org/10.1002/widm.1520

[3] Aarti Karandikar, Vedita Deshpande, Sanjana Singh, Sayali Nagbhidkar, Saurabh Agrawal, “Deepfake Video Detection Using Convolutional Neural Network”, ISSN 22783091 https://doi.org/10.30534/ijatcse/2020/6292202 0

[4] Rafique, R., Gantassi, R., Amin, R. *et al.* Deep fake detection and classification using error-level analysis and deep learning. *Sci Rep* 13, 7422 (2023). https://doi.org/10.1038/s41598-023-34629-3

[5] Y. Patel et al., "An Improved Dense CNN Architecture for Deepfake Image Detection," in IEEE Access, vol. 11, pp. 22081-22095, 2023, doi: 10.1109/ACCESS.2023.3251417.

[6] S. E. VP, C. M. S and R. Dheepthi, "LLM-Enhanced Deepfake Detection: Dense CNN and Multi-Modal Fusion Framework for Precise Multimedia Authentication," 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 2024, pp. 1-6, doi: 10.1109/ADICS58448.2024.10533511.

[7] V. Kumar Pandey, S. Jain and S. K. Saritha, "Advanced IoT-Based Fire and Smoke Detection System leveraging Deep Learning and TinyML," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-10, doi: 10.1109/ICCCNT56998.2023.10307805.

[8] R. Singh, K. Ashwini, B. Chandu Priya and K. Pavan Kumar, "Deepfake Face Extraction and Detection Using MTCNN-Vision Transformers," 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 2024, pp. 01-08, doi: 10.1109/ICDCECE60827.2024.10549578.

[9] C. Yavuz, "A Multidisciplinary Look at History and Future of Deepfake With Gartner Hype Cycle," in IEEE Security & Privacy, vol. 22, no. 3, pp. 50-61, May-June 2024, doi: 10.1109/MSEC.2024.3380324.

[10] D. Garg and R. Gill, "Deepfake Generation and Detection - An Exploratory Study," 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Comuputer Engineering (UPCON), Gautam Buddha Nagar, India, 2023, pp. 888-893, doi: 10.1109/UPCON59197.2023.10434896