www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 02, February 2025, pp : 1157-1161

e-ISSN : 2583-1062
Impact Factor : 7.001

DEEP LEARNING FOR FAKE FACE DETECTION

Asst. Prof. P. S. Sontakke[*1], Rajat R Nimje[*2], Rohan. R. Nakade[*3]

[*1]Assistant Professor Computer Engineering Department, SRPCE, Nagpur, Maharashtra, India.

[*2,3]UG Student, Computer Engineering Department, SRPCE, Nagpur, Maharashtra, India.

DOI: https://www.doi.org/10.58257/IJPREMS38747

ABSTRACT

The proliferation of deepfake technology, which leverages advanced deep learning techniques to create highly realistic fake videos, poses significant threats to the integrity of digital media. This project aims to develop a robust system for detecting deepfake faces in video content, utilizing a combination of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The system is designed to accurately identify manipulated media by analyzing spatial and temporal inconsistencies in video frames. The detection model is trained on extensive datasets, including FaceForensics++, DeepFake Detection Challenge (DFDC), and Celeb-DF, ensuring high accuracy and generalizability. The implementation leverages Python for core algorithm development, with Flask serving as the web framework to create an intuitive user interface.

Keywords: Artificial Intelligence, Deep Learning, Deepfake, Digital Deception, Deepfake Detection, Convolutional Neural Networks (CNNs), Python, Flask , Javascript, Real-Time Analysis.

1. INTRODUCTION

Deepfake technology becomes increasingly accessible, the ability to detect these manipulated media files is crucial. Deepfake face detection involves developing algorithms and models capable of identifying signs of manipulation, distinguishing between authentic and synthetic content. This project aims to explore various methods of deepfake detection, analysing the underlying technologies and examining their effectiveness in real-world scenarios. The need for robust detection systems has never been more urgent. With the proliferation of social media and digital communication, the potential for deepfakes to mislead the public and influence opinion grows exponentially. By leveraging computer vision, machine learning, and data analysis, this project will contribute to the ongoing efforts to safeguard the integrity of digital media.

Deepfake face detection models are designed to identify and distinguish manipulated or synthetic images and videos from real ones. These models leverage advanced machine learning techniques, particularly deep learning, to detect subtle inconsistencies and artifacts that are often present in deepfakes, Deep learning models, such as Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) are used

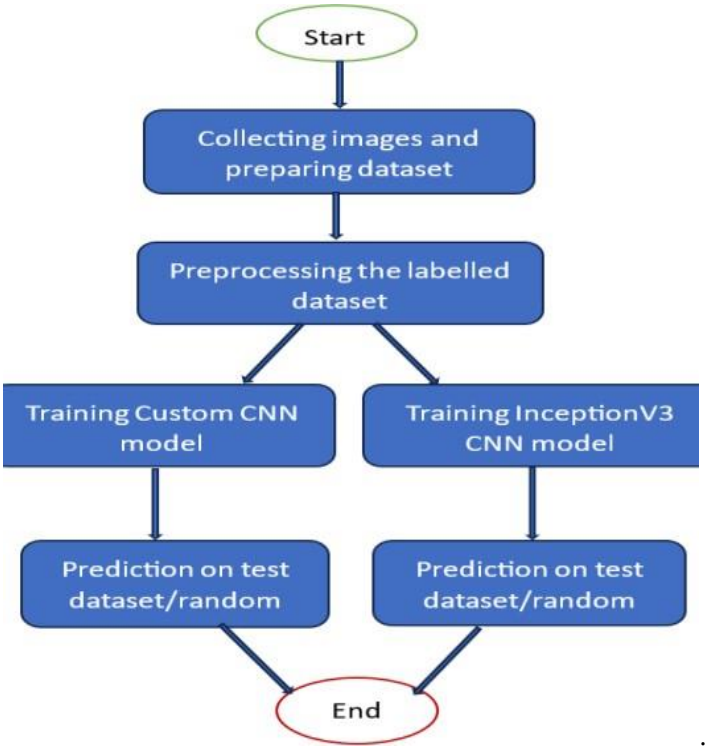

Fig 1: Architecture of System

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 02, February 2025, pp : 1157-1161

e-ISSN : 2583-1062
Impact Factor : 7.001

PROBLEM STATEMENT

With the rise of advanced AI and machine learning techniques, deepfake videos—synthetic media where a person's likeness is replaced with someone else's—have become increasingly realistic and difficult to detect. These deepfakes pose significant risks, including misinformation, fraud, and privacy violations. The challenge is to develop an AI/ML-based solution that can accurately detect deepfake videos by analyzing inconsistencies in facial features, expressions, movements, and audio-visual synchronization. The goal is to create a robust system that can authenticate videos and provide detailed reports on potential deepfake characteristics.

2. RELATED WORK

DeepFake Detection: This resource provides a comprehensive overview of various papers, benchmarks, and datasets specifically focused on deepfake detection. The models discussed, such as EfficientNetB4, Vision Transformer, and XceptionNet, are used to detect manipulated facial images. This work is highly relevant in the field of machine learning and computer vision.[2]

Multiclass AI-Generated Deepfake Face Detection Using Patch-Wise Deep Learning Model: This paper explores the use of Vision Transformers (ViTs) for detecting multiclass deepfake images. The study focuses on using patch-wise analysis to identify subtle artifacts and inconsistencies in deepfake images. The field of this work is deep learning, particularly in image processing and pattern recognition.[1]

Deepfake Face Detection Using Machine Learning: This research discusses the use of Long Short-Term Memory (LSTM) networks combined with Convolutional Neural Networks (CNNs) for spatial and temporal analysis of deepfake videos. The approach focuses on detecting artifacts and discrepancies in deepfake videos. This work is situated in the field of machine learning, with a specific focus on video analysis and temporal data.[5]

3. ARCHITECTURAL DESIGN

The system architecture for deepfake face detection involves several key components working together to accurately identify manipulated media. Initially, data collection and preprocessing are crucial steps, where a large dataset of real and deepfake videos or images is gathered and standardized. This involves extracting frames from videos, detecting faces, and aligning them to ensure uniformity. Feature extraction follows, typically using Convolutional Neural Networks (CNNs) to capture spatial hierarchies in images, and Vision Transformers (ViTs) to model long-range dependencies

The architecture is designed to leverage the strengths of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. Below is a detailed explanation of the model components:

Input Video Frame: The input to the model consists of frames extracted from a video. Each frame represents a single instance of a person's face, which will be analyzed for deepfake detection.

Convolutional Neural Network (CNN) Layer: The initial layer of the model is a Convolutional Neural Network (CNN), which is responsible for feature extraction. The CNN applies convolution operations to the input frame, capturing essential features such as edges, textures, and patterns. These features are represented as feature maps. The use of multiple CNN layers allows the model to extract increasingly complex features as the input data progresses through the network.

Feature Extraction: The CNN layers produce a set of feature maps that encapsulate critical information about the input frame's spatial characteristics. These feature maps highlight key facial attributes, including the shape and structure of the eyes, mouth, and other facial regions.

Pooling Layer: The feature maps are passed through a pooling layer, which performs dimensionality reduction. This layer retains the most important information while reducing the spatial dimensions of the feature maps. Pooling helps to improve computational efficiency and reduce the risk of overfitting.

Long Short-Term Memory (LSTM) Layer: To analyze the temporal dependencies across video frames, the model incorporates a Long Short-Term Memory (LSTM) network. The LSTM layer processes sequences of feature maps from multiple frames, capturing temporal information such as movements and expressions. This temporal analysis is crucial for detecting deepfakes, as it helps identify inconsistencies and unnatural transitions over time.

Fully-Connected (FC) Layer: The outputs from the LSTM layer are fed into a fully-connected layer. This layer integrates the extracted spatial and temporal features, facilitating the final classification decision. The fully-connected layer combines the learned features to provide a comprehensive representation of the input data.

Output Layer: The final layer of the model produces the output, indicating whether the analyzed video frame (or sequence of frames) is real or a deepfake. The output can be in the form of a probability score or a binary classification (real vs. fake). The output layer provides an interpretable result, aiding in the identification of deepfake content.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 02, February 2025, pp : 1157-1161

e-ISSN : 2583-1062
Impact Factor : 7.001

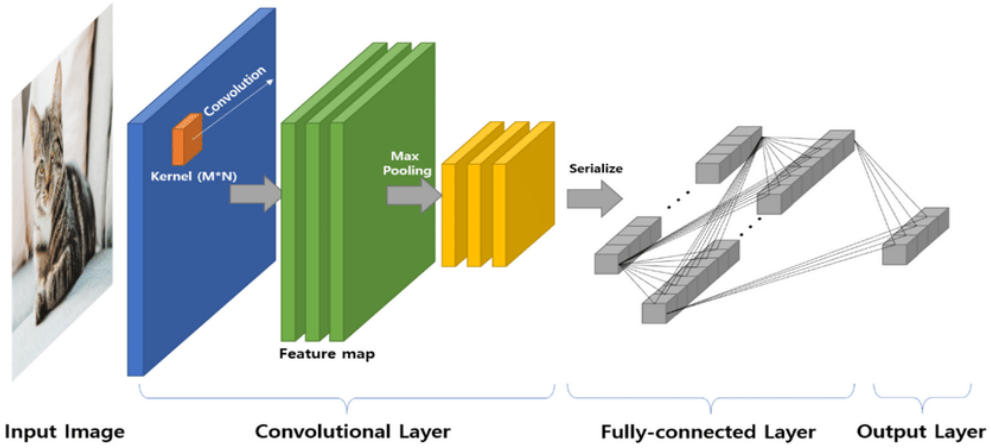

Fig 2: CNN for Image Classification

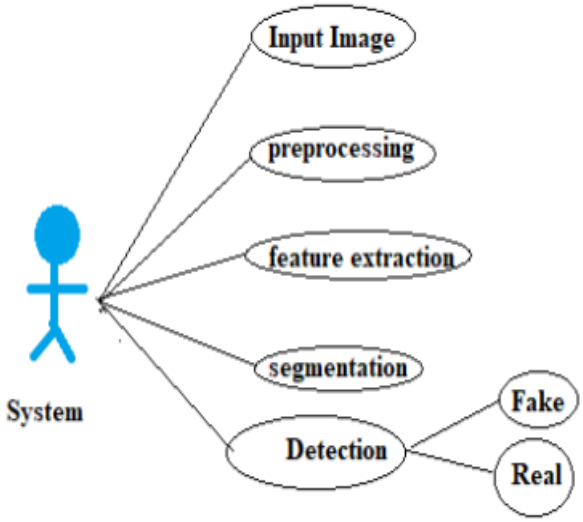

Fig 3: Functioning of model

TOOLS AND TECHNIQUE

Python is Interpreted− During execution, the interpreter handles Python. Your software doesn't have to be compiled before it is run. This is comparable to PHP and PERL.

The best thing about Python is that it's interactive; you can develop programs by just sitting at a Python prompt and interacting with the interpreter.

Python is compatible with the "Object-Oriented" programming approach, which encapsulates code in objects.

Python is a great language for programmers who are new to the field. Python is a great choice for new programmers because it gives them the freedom to create a wide variety of applications, from simple text editors to web browsers

4. RESULTS AND DISCUSSION

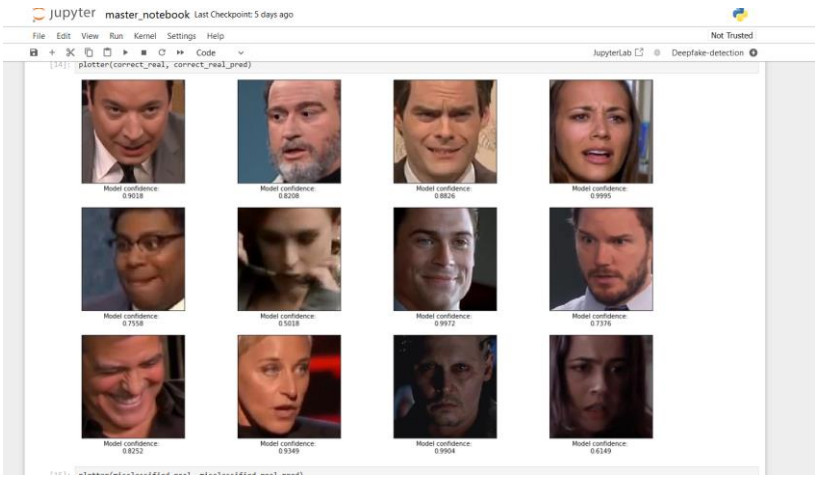

Fig 4: Frames authenticity confidence

The output of the deepfake face detection model provides a robust framework for assessing the authenticity of facial images. By combining quantitative scores, visual explanations, and detailed analyses, the model empowers users to make informed decisions regarding potential deepfakes. This comprehensive approach not only aids in detection but also fosters trust and transparency in the model's capabilities.

```
Creating Directory: .\train_sample_videos\abofeumbvv
Converting Video to Images...
Original Dimensions:  (1080, 1920, 3)
Scale Ratio:  0.33
Resized Dimensions:  (356, 633, 3)
Original Dimensions:  (1080, 1920, 3)
Scale Ratio:  0.33
Resized Dimensions:  (356, 633, 3)
Original Dimensions:  (1080, 1920, 3)
Scale Ratio:  0.33
Resized Dimensions:  (356, 633, 3)
Original Dimensions:  (1080, 1920, 3)
Scale Ratio:  0.33
Resized Dimensions:  (356, 633, 3)
Original Dimensions:  (1080, 1920, 3)
Scale Ratio:  0.33
Resized Dimensions:  (356, 633, 3)
Original Dimensions:  (1080, 1920, 3)
Scale Ratio:  0.33
Resized Dimensions:  (356, 633, 3)
Original Dimensions:  (1080, 1920, 3)
Scale Ratio:  0.33
Resized Dimensions:  (356, 633, 3)
Original Dimensions:  (1080, 1920, 3)
Scale Ratio:  0.33
Resized Dimensions:  (356, 633, 3)
Original Dimensions:  (1080, 1920, 3)
```

Fig 5: Video to image standardization

This descripes as a collection of images extracted from videos, each representing a single frame's facial region. Standardized image sizes, ensuring consistency across the dataset.Normalized pixel values, which help in speeding up the training process and improving model convergence.Optionally augmented data, enhancing the robustness of the model against variations.

Layer (type)	Output Shape	Param #
efficientnet-b0 (Functional)	(None, 1280)	4,049,564
dense (Dense)	(None, 512)	655,872
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 128)	65,664
dense_2 (Dense)	(None, 1)	129

```
Total params: 4,771,229 (18.20 MB)
Trainable params: 4,729,213 (18.04 MB)
Non-trainable params: 42,016 (164.12 KB)
```

Fig 6: Extracted frame is analyzed by model

Each extracted frame is analyzed to detect faces using algorithms like Haar Cascades or modern deep learning methods.

The Purpose is to Isolating the face region focuses the analysis on relevant features and reduces noise from the background.

5. CONCLUSION

In conclusion, the deepfake face detection system utilizing machine learning with LSTM represents a notable advancement in the fight against digital misinformation. By spatial feature extraction with long short-term memory networks for temporal analysis, the system offers a sophisticated approach to detecting deepfake content. Its ability to analyze both spatial and temporal aspects of images allows it to identify subtle inconsistencies and artifacts that might escape human detection. The system's modular design ensures adaptability and real-time processing capabilities, making it applicable across various platforms from social media to forensic analysis

6. REFERENCES

[1] William, Youssef, SherineSafwat, and Mohammed AM. Salem. Robust Image Forgery Detection Using Point Feature Analysis. 2024 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, 2024.

[2] Prof. P.S. Sontakke, Rajat Nimje , Rohan Nakade. "Fake Face Detection using Deep Learning" International Journal of Progressive Research in Engineering Management and Science- 09 Sept 2024.

[3] Mal´ık, Peter, Stefan Kri ˇ stof ˇ ´ık, and Krist´ına Knapova.´ Instance Segmentation Model Created from Three Semantic Segmentations of Mask, Boundary and Centroid Pixels Verified on GlaS Dataset. 2024 15th Conference on Computer Science and Information Systems (FedCSIS). IEEE, 2024.

[4] Al-Berry, M. N., et al. Directional Multi-Scale Stationary Wavelet-Based Representation for Human Action Classification. Handbook of Research on Machine Learning Innovations and Trends. IGI Global, 2023. 295-319.

[5] Muhammad, Ghulam, et al. Image forgery detection using steerable pyramid transform & local binary pattern.Machine Vision applications. 25(4)(2023), 985-995.

[6] Kuznetsov, Andrey, and Vladislav Myasnikov. A new copy-move forgery detection algorithm using image preprocessing procedure. Procedia engineering. 201(2022), 436-444.

[7] Li, Yuezun, et al. "Exposing DeepFake Videos by Detecting Face Warping Artifacts." IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022.

[8] Matern, Fabian, Christian Riess, and Matthias Stamminger. "Exploiting Visual Artifacts to Expose DeepFakes and Face Manipulations." IEEE Winter Conference on Applications of Computer Vision (WACV), 2021.

[9] Nguyen, Huy H., et al. "Deep Learning for Deepfake Detection: Analysis and Benchmarking." Advances in Neural Information Processing Systems (NeurIPS), 2020.