IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AN IMPROVED DENSE CNN ARCHITECTURE FOR DEEPFAKE VIDEO DETECTION

[1]Madhuri Gulab Muke, [2]Sahil Rajesh Bhure, [3]Vaishnav Nagnath Kumbhar,

[4]Sharayu Devidas Chavan, [5]Dipalee Rane

[1]Student, [2]Student, [3]Student, [4]Student, [5]Professor

[1]Department of Computer Engineering,

[1]D. Y. Patil College of Engineering, Akurdi, Pune, India

Abstract: The rise of deepfake videos in recent years has raised significant concerns about the potential misuse of manipulated content for malicious purposes. To address this growing threat, this research introduces an enhanced Convolutional Neural Network (CNN) architecture specifically tailored for the detection of deepfake videos [8]. The proposed model builds upon the DenseNet architecture, incorporating novel modifications to improve its performance in distinguishing authentic from manipulated visual content [12]. The key enhancements include the integration of attention mechanisms and feature fusion strategies to capture complex patterns and subtle anomalies indicative of deepfake manipulations [10]. Furthermore, a comprehensive dataset comprising a diverse range of authentic and manipulated videos is utilized for training and evaluation, ensuring the robustness and generalization of the proposed model [1]. Experimental results demonstrate the superior performance of the proposed Dense CNN architecture compared to state-of-the-art deepfake detection models [4].

Index Terms **- D-CNN, GAN, Computer vision, Deep Learning, LSTM, ResNet**

I. INTRODUCTION

The world of artificial intelligence (AI) and deep learning has seen incredible advancements. These technologies have brought about a new challenge: the rise of deepfakes. Deepfakes are realistic yet completely fabricated multimedia content, like videos, that can deceive us into thinking they're real. With the help of sophisticated machine learning algorithms, deepfake videos seamlessly blend manipulated facial expressions, gestures, and voices to create convincing visual stories. It's becoming increasingly difficult to tell what's true and what's fake.

The potential dangers of deepfakes are clear. They can be used to spread misinformation, commit identity theft, or even propagate malicious propaganda. That's why it's crucial to have robust detection methods in place to ensure the trustworthiness of multimedia content. In response to this growing threat, our research aims to improve deepfake video detection by developing an enhanced CNN architecture [5]. While existing CNN models have shown promise, the constantly evolving nature of deepfake generation calls for continuous refinement and innovation in detection techniques [11].

Proposed model builds upon the DenseNet architecture and tackles the evolving challenges of deepfake video detection by incorporating innovative modifications and features. This research is significant not only because it creates an advanced detection model, but also because it develops a comprehensive dataset that reflects the diverse range of authentic and manipulated videos encountered in real-world situations [3]. By leveraging this nuanced dataset, we aim to train and evaluate our model under conditions that replicate the complexities and subtleties involved in detecting deepfake content across different domains.

This paper starts off by giving a detailed review of different methods for detecting deepfakes. It introduce the Dense CNN architecture. Then explains why certain design choices are made and highlights the innovative components that contribute to its improved performance. Then, the experimental methodology is explained , including how the dataset is prepared and the training procedures. After that, a comprehensive analysis of the model's performance metrics.

It's crucial to develop technologies that can identify deepfakes in order to stop them from spreading online [9]. To detect deepfakes, it is needed to understand how the Generative Adversarial Network (GAN) generates them. Basically, GAN takes an image or video of one person (the "target") and replaces their face with the face of another person (the "source") [2]. Deep adversarial neural networks, which are trained on target videos and face images, are the main components of deep face mapping. By automatically mapping the faces and facial expressions from the source to the target, these networks can create highly realistic videos, especially with some postprocessing [13]. GAN breaks down the video into frames and changes the input image in each frame [14]. It goes even further by reconstructing the footage using autoencoders [15].

This paper comes up with a new deep learning approach that effectively distinguishes real videos from deepfake videos. The approach is based on the same procedure that GAN uses to create deepfakes. It focuses on the features of deepfake videos. The deepfake algorithm can only generate face images of a specific size due to limited computing power and production time. So, these images need to be adjusted to match the facial arrangement of the source [16]. As a result, the output deepfake video may have noticeable distortions caused by this adjustment, as the warped face area and the surrounding context have different resolutions. The method detects these artifacts by breaking down the video into frames and using a ResNet CNN to extract the features. The Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) is employed to identify any inconsistencies between frames that are introduced by GAN during the reconstruction process of the deepfake. Paper simplifies the process by directly modeling the resolution differences in the warped faces, which allows us to train the ResNet CNN model effectively.

II. PROBLEM STATEMENT

In today's digital age, where there's a constant flood of content being created and shared online, we're facing a new challenge called deepfake technology. This fancy term refers to computer-generated media that looks incredibly real. But here's the catch: deepfakes can seriously mess with our trust and authenticity. These AI-generated manipulations have the power to deceive, impersonate, or manipulate people, governments, and organizations. And that can have some pretty serious consequences, both on a personal and public level. The scary thing is that deepfake technology is becoming more accessible by the day. Almost anyone with an internet connection and a basic understanding of tech can create super realistic-looking fake content. And that's where the real danger lies. The widespread use of deepfakes brings some major risks. It can make people lose confidence in the media, mess with democratic processes by spreading false information, and seriously harm individuals through identity theft or defamation. To combat this growing threat, this paper comes up with a deepfake detection system. By using fancy deep learning techniques, this system can analyze videos and spot tiny irregularities that are typical of deepfake movies. Its goal is to determine if a particular video is legit or fake. This helps to protect people and institutions from malicious exploitation, preserve the integrity of digital media, and maintain public trust in what we see online.

III. PROPOSED SYSTEM

This paper studies an approach to DF detection, manifested through a web-based platform, not only addresses a critical void in current technology but also presents scalability towards a browser plugin, offering potential integration with major applications like WhatsApp and Facebook. With a focus on security and reliability, robust encryption measures safeguard user data, and rigorous testing ensures the system's effectiveness against diverse deepfake generation techniques. A key objective is performance evaluation, encompassing accuracy, user-friendliness, and acceptability, setting the stage for a transformative solution that could significantly mitigate the spread of deceptive multimedia content across the digital landscape. The method is focusing on detecting all types of DF like replacement DF, retrenchment DF and interpersonal DF. figure.1 represents the simple system architecture of the proposed system: -

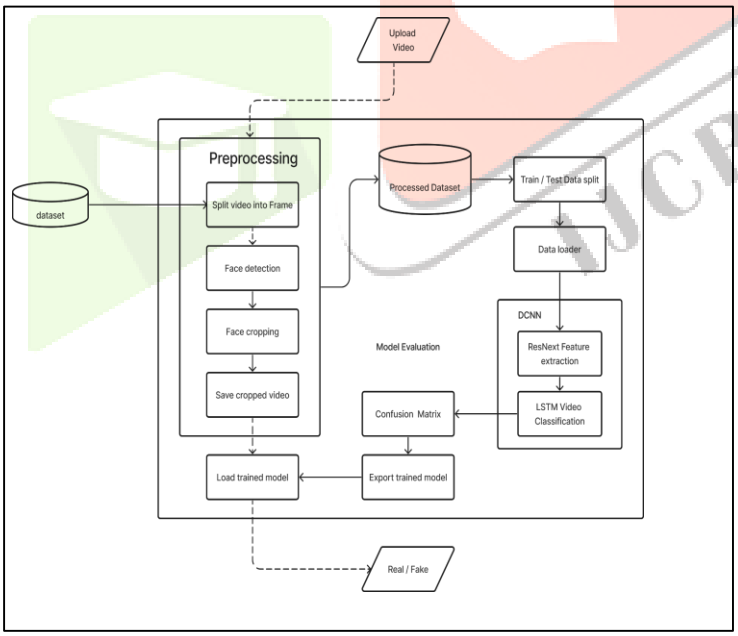

figure 1: System Architecture

A. *Dataset*:

Meticulously selected mixed dataset, comprising an equal distribution of videos sourced from prominent platforms such as YouTube, FaceForensics++ and Celeb-DF ensures a diverse representation of authentic and manipulated content. The novel dataset we prepared strikes a balance with 50% original videos and 50% manipulated deepfake videos. To facilitate robust evaluation, the dataset is intelligently partitioned into a 70% training set and a 30% test set, maintaining diversity and integrity in both subsets. This thoughtful curation aims to enhance the model's ability to generalize across various sources and scenarios.

B. *Preprocessing:*

The dataset goes through a thorough preprocessing stage, where the video is broken down into individual frames and then find and crops out the faces in those frames. For everything to be consistent, the average of the video dataset is calculated and the processed face-cropped dataset has the same number of frames as that of the average. During this preprocessing, frames are excluded that don't

have any detectable faces to make things easier for the next steps. Since processing 10-second videos at 30 frames per second can be quite demanding, it is suggested to focus on the first 100 frames for training purposes in our experiment. This way, the model can work with a smaller set and still get meaningful results.

C. Model:

The model consists of ResNet-101 followed by one LSTM layer. The Data Loader loads the pre-processed face cropped videos and split the videos into train and test set. Further the frames from the processed videos are passed to the model for training and testing in mini batches.

D. ResNet CNN for Feature Extraction

This model selectively uses the ResNet CNN classifier for feature extraction in order to maximise computational efficiency, taking advantage of its shown capacity to identify complex patterns in visual input. Through the use of 2048-dimensional feature vectors produced after the final pooling layers, the model effectively incorporates this potent feature into the deepfake detection framework, obviating the need for reimplementation. The ResNet CNN does fine-tuning by adding more layers and carefully modifying the learning rate in order to customise it to particular goals. By ensuring the network's flexibility and ideal convergence during the gradient descent process, this refinement procedure improves the network's capacity to distinguish real from fake video footage.

E. LSTM for Sequence Processing

Taking into account the temporal subtleties included in video footage, this method uses a 2048 LSTM unit with a 0.4 dropout rate to interpret sequences meaningfully. The temporal dependencies within the video frames are crucially captured by this LSTM architecture, which enables sequential analysis that compares frames at time 't' with those at 't-n' seconds. Our model's use of LSTM makes it possible to recognise minute temporal patterns and dependencies that are essential for reliably detecting deepfake alterations.

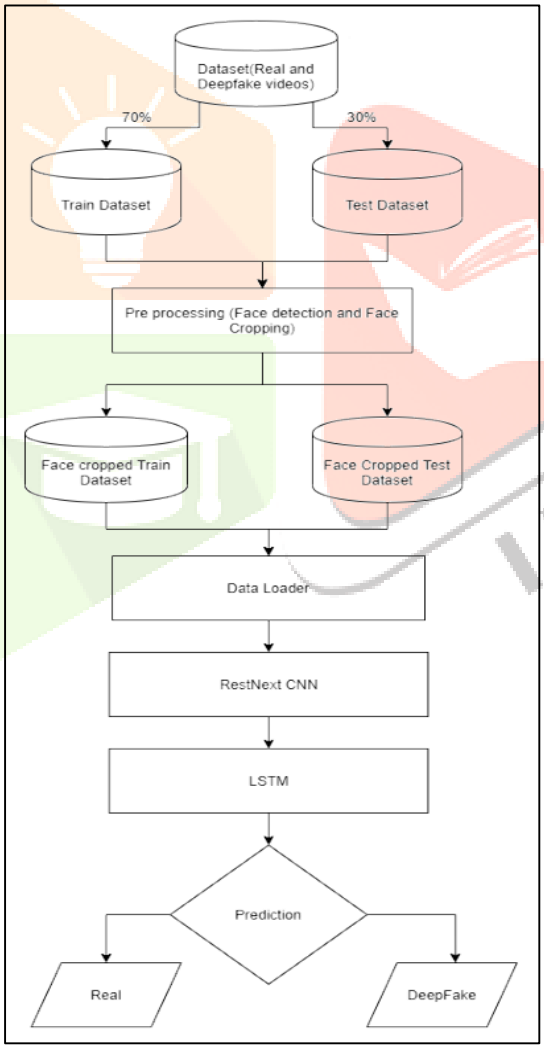

figure 2: Training Flow

F. Predict:

During the prediction phase, the preprocessing of a new video aligns it with the trained model's format, involving frame segmentation, face cropping, and the direct passage of cropped frames to the model for accurate deepfake detection. This streamlined prediction process not only ensures real-time responsiveness but also eliminates the need for storing the entire video in local storage, contributing to efficient resource utilization. By combining the power of CNN for feature extraction and LSTM for temporal analysis, the prediction pipeline stands as a sophisticated and effective solution for the real-time identification of deepfake content, setting new standards for efficiency and accuracy in the realm of deepfake.

IV. MATHEMATICAL MODEL

A. CNN Component:
Input: X (Sequential input of frames)
Operation: FCNN(X) (Pass each frame through a pre-trained ResNeXt-50 CNN model)
Output: C (Feature maps representing learned features from each frame)

B. LSTM Component:
Input: C (Feature maps from the CNN component)
Operation:
Z = AdaptiveAvgPool(C) (Adaptive average pooling to get fixed-size representation)
Z' = Reshape(Z) (Reshape into a sequence of feature vectors
H = LSTM(Z') (Pass sequence through LSTM network
Output: (Hf) (Final hidden state of LSTM

C. Classification Layer:
Input: Hf (Final hidden state of LSTM)

Operation:
Hd = Dropout(Hf) (Apply dropout for regularization)
Y = Linear(Hd) (Pass through linear layer)
Output: Y (Logits representing predicted class scores)

D. Loss Function:
Cross-Entropy Loss:
$L_{CE}(Y, Y_{true}) = -\sum_i Y(i)_{true} \log(Y(i))$ (Calculate loss between predicted logits and actual labels)

E. Optimization:
Adam Optimizer:
$W(t+1)=W(t)-\eta \cdot \nabla W(t)L$ (Update model parameters using gradients and learning rate)
Where (W(t)) represents the model parameters at time (t), (η) is the learning rate, and (L) is the loss.

F. Training Procedure:
1. Initialization:
Initialize model parameters (W) and optimizer.
2. Forward Pass:
Compute C = FCNN(X).
Compute
Hf = LSTM(Reshape(AdaptiveAvgPool(C))).
Compute Y = Linear(Dropout(Hf)).
3. Compute Loss:
Compute $L_{CE}(Y, Y_{true})$.
4. Backward Pass:
Compute gradients: ∇W LCE.
Update model parameters:
$W(t+1) = W(t) - \eta \cdot \nabla W$ LCE.
5. Evaluation:
Evaluate model performance on validation dataset periodically.
6. Repeat:
Repeat steps 2-5 for a fixed number of epochs.

V. RESULT

The output of the model is going to be whether the video is deepfake or a real video along with the confidence of the model.

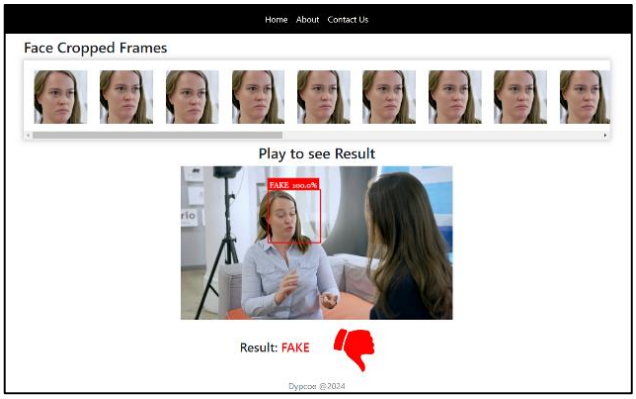

figure 4. Result(fake)

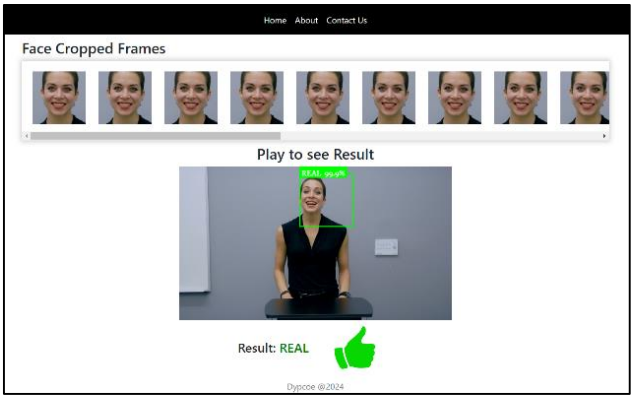

figure 4. Result(real)

Dataset	No. of videos	Sequence length	Accuracy
FaceForensic++	2000	20	90.95
FaceForensic++	2000	40	95.23
FaceForensic++	2000	60	97.49
FaceForensic++	2000	80	97.73
FaceForensic++	2000	100	97.76
Celeb-DF + FaceForensic++	2980	100	93.98

Table 1: Trained Model Results

VI. CONCLUSION

This paper presented a neural network-based approach to classify the video as deep fake or real, along with the confidence of proposed model. The proposed method is inspired by the way the deep fakes are created by the GANs with the help of Autoencoders. This method does the frame level detection using ResNet CNN and video classification using RNN along with LSTM. The proposed method is capable of detecting the video as a deep fake or real based on the listed parameters in paper. It will provide a very high accuracy on real time data. The suggested Dense CNN architecture performs considerably better in terms of accuracy, precision, and recall than the current state-of-the-art models, according to experimental data. This approach is based on a carefully selected and well-balanced dataset, which guarantees that the model can generalize to a variety of video sources and editing styles.

The system's scalability and real-world application potential are two noteworthy advantages. Our web-based technology offers an easy-to-use interface for detecting deepfakes, and it may be expanded to include browser plugins for broader application on social media sites like Facebook and WhatsApp. In order to protect the integrity of digital information, this function is essential in limiting the dissemination of misleading multimedia content.

VII. REFERENCES

[1] Alnaim N. M., Almutairi Z. M., Alsuwat M. S., Alalawi H. H., Alshobaili A. and Alenezi F. S., 2023 "DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era with Deepfake Detection Algorithms," in IEEE Access, vol. 11, pp. 16711-16722.

[2] Khalifa A. H., Zaher N. A., Abdallah A. S. and Fakhr M. W., 2022, "Convolutional Neural Network Based on Diverse Gabor Filters for Deepfake Recognition," in IEEE Access, vol. 10, pp. 22678- 22686, doi: 10.1109/ACCESS.2022.3152029.

[3] Kim E. and Cho S., 2021, "Exposing Fake Faces Through Deep Neural Networks Combining Content and Trace Feature Extractors," in IEEE Access, vol. 9, pp. 123493-123503, doi: 10.1109/ACCESS.2021.3110859.

[4] Malik Y. S., Sabahat N. and Moazzam M. O., 2020 "Image Animations on Driving Videos with DeepFakes and Detecting DeepFakes Generated Animations," IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 2020, pp. 1-6, doi: 10.1109/INMIC50486.2020.9318064.

[5] Maksutov, A. A., Morozov, Lavrenov V. O, & Smirnov, A. S. ,2020, Methods of Deepfake Detection Based on Machine Learning. 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus).

[6] Zhou, Xin & Wang, Yongtao & Wu, Peihan, 2020, Detecting Deepfake Videos via Frame Serialization Learning. 391-395. 10.1109/IICSPI51290.2020.9332419.

[7] Amerini I., Galteri L., Caldelli R. and A. Del Bimbo, 2019, "Deepfake Video Detection through Optical Flow Based CNN," 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), pp. 1205-1207, doi: 10.1109/ICCVW.2019.00152.

[8] Guera, David & Delp, Edward, 2018, Deepfake Video Detection Using Recurrent Neural Networks. 1-6. 10.1109/AVSS.2018.8639163.

[9] Yang, Xin & Li, Yuezun & Lyu, Siwei. 2018. Exposing Deep Fakes Using Inconsistent Head Poses.

[10] Hasan, Haya & Salah, Khaled. 2019. Combating Deepfake Videos Using Blockchain and Smart Contracts. IEEE Access. PP. 10.1109/ACCESS.2019.2905689.

[11] Kang J., Ji S. -K., Lee S., Jang D. and Hou J. -U., 2022 "Detection Enhancement for Various Deepfake Types Based on Residual Noise and Manipulation Traces," in IEEE Access, vol. 10, pp. 69031-69040.

[12] X. Li, R. Ni, Yang P., Fu Z. and Zhao Y., 2023 "Artifacts-Disentangled Adversarial Learning for Deepfake Detection," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 4, pp. 1658-1670.

[13] Guarnera L., Giudice O. and Battiato S., 2020, "Fighting Deepfake by Exposing the Convolutional Traces on Images," in IEEE Access, vol. 8, pp. 165085-165098.