



A Survey Paper On Efficient Machine Learning Technique For Deep Fake Detection

Gursal Prabodhini S.¹, Prof. Gade S. A.²

ME Student, Department of Computer Engineering¹,

Head of Department, Department of Computer Engineering²

SND College of Engineering and Research Center, Yeola, Nashik, India

Abstract: Deep fakes are highly realistic 'fake' videos and they are entirely the creation of advanced technologies of recent decades including machine learning, artificial intelligence, and deep learning. Such videos, very similar to real videos, are obtained by training deep computer models on the abundant face database and gradually mapping one person's facial movements onto the other. The following questions arise when Deepfakes are often used; Political instabilities, fake terrorism cases, losing confidence in digital platforms among others. Hence, there is a need for society and advance technology to develop new techniques for deep fake detection and prevention. This essay looks critically at the situation of today.

Keywords- Deepfakes, Haar Cascades, MTKNN (Multi-task Kascaded Convolutional Networks), Convolutional Neural Network (CNN), Sigmoid Activation, Real-time Detection, Prevention Model, Steganalysis.

I. INTRODUCTION

Machine Learning Models for Deepfake Detection, Misinformation, and Biometric Safety: Introduction There's a need to build a machine learning system that On deepfake detected images or videos helps in mitigating the untruths being spread and protect individuals from such predicament Inappropriate shifts of media content.

1.1 Motivation

Widening Misinformation: With the advent of deepfake technology, has made people more apprehensive about the reality of fake news. With deepfakes technology Improving to the extent that the user's ability to tell a truth image and an altered one will be crucial in such a scenario.

Out of Sight Out of Mind: One ill purpose for using deepfakes is on individual impersonation where One's likeness is used out of context which can lead to identity fraud, slander or bullying and so is the need for working on ways to develop the mechanisms that would enable one to detect this abuse.

Societal Effects: Deepfake technology is capable of enhancing propaganda, swaying people's preferences, comprising electoral politics, and promoting civil disorder. Its even more important to recognise such deceptions as a social control mechanism for press and media pollution.

Evolution of Technology: The field of artificial intelligence and its associated technologies is changing at a tremendous speed. This presents a rare chance to apply the same technology advancement to help come up with better and faster methods of detection that will help in curtailing the growth of the deepfake menace.

Cross-Pollination of Disciplines: This type of project integrates computer vision and aspects of ethics and law as well as social science, thus is a thrilling interdisciplinary problem which is applicable to many areas.

1.2 Objective

Construct Efficient Discriminator Model: Create machine learning imaging based framework to test and classify still or motion picture content into real or deep fake versions with high efficacy.

Maintain Optimal Performance And Reliability: Ensure completion of the task with a desired degree of accuracy while minimizing the rate of the false positive and the false negative results in the detection of fake content which assure high level of trust in the results of the detection.

Employ Non-Homogeneous Datasets: Train and test the model on real and faked visible media in order to enhance generalization and robustness to a particular faking method or style.

Make Detection Available for Use in Real-Time: Provision the possibility for instant processing, so that users may upload content for deepfake detection where they do not have to wait.

Construct an Engaging Interface: Create an interface that encourages easy interaction – where users can upload files and get feedback clearly indicating whether or not the media is genuine.

II. LITERATURE SURVEY

[1] Rana, S., Gudla, C., Solaiman, M., & Sohan, F Proposed a hypothesis testing framework for classifying genuine and Deepfakes. Integrated Blockchain and teganalysis technologies for real-time Deepfake detection. Utilized masks and PRNU patterns to generate unique signatures.

[2] Md Shohel Andrew Sung proposed Deepfake Stack detects manipulated videos using deep ensemble learning technique. Combines DL models to create an improved classifier for Deepfake detection. Achieves 99.65% accuracy and AUROC of 1.0 in detecting Deepfake.

[3] Xin Yang, Yuezun Li and Siwei Lyu proposed method based on the observations that Deep Fakes are created by splicing synthesized face region into the original image.

[4] ISROEL MIRSKY proposed Generative deep learning algorithms have progressed to a point where it is difficult to tell the difference between what is real and what is fake.

[5] Kim, J., Brian Dolhansky, Joanna Bitton, Ben Pflaumal define Deepfakes are a recent off-the-shelf manipulation technique that allows anyone to swap two identities in a single video.

[6] Md. Shohel Rana Andrew H. Sung proposed Recent advances in technology have made the deep learning (DL) models available for use in a wide variety of novel applications; for example, generative adversarial network.

[7] Falko Matern Christian Riess Marc Stamminger proposed High quality face editing in videos is a growing concern and spreads distrust in video content. However, upon closer examination, relatively simple visual artifacts can be already quite effective in exposing such manipulations.

[8] Govind Mi Jiraphon Yenphraphaittal proposed The integrity of online video interactions is threatened by the widespread rise of AI-enabled high-quality deepfakes that are now deployable in real-time.

III. EXISTING SYSTEM

3.1 System Elements

Input Data: Video/picture Upload: Individuals upload content that is either video or picture and for which deep fakes have to be verified **Preprocessing:** Media might undergo preprocessing steps such as scaling, normalization, or even face alignment in the case of movies before feeding it data to the KNN model.

Input data Preprocessing Feature Extraction Video Frame Extraction: The input video for the extraction of an individual frame in order to check for deepfake features. **Face Detection:** Identify the faces of the individuals appearing in the input images or frames using Haar Cascades or MTKNN techniques to focus the analysis.

Normalization: Normalize the pixel values to some scale such as 0–1 or -1–1.

Data Augmentation: During training, apply transformations to the input images/videos, such as rotation, zooming, and flipping, to enhance model generalization.

3.2 Haar Cascades: A common method for identifying faces in pictures and videos, the Haar Cascade algorithm is frequently employed as a first step in the detection of deepfakes. A cascade of classifiers is used by the machine learning-based Haar Cascade technique to identify faces in an image. It distinguishes between face and non-facial regions by looking at aspects of the image, such as borders, corners, and texture variations.

3.3 Kascaded Convolutional Networks for Multiple Tasks:

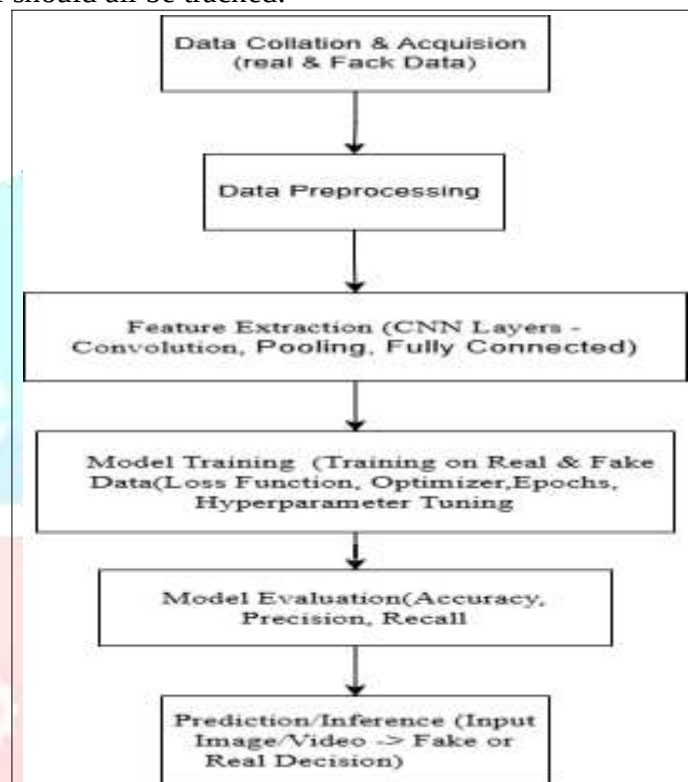
Multitask A framework known as Cascaded Convolutional Networks (MTCNN) was created to address the problems of face alignment and detection. Three stages of convolutional networks that could recognize faces and also locations of major landmarks like the mouth, nose, and eyes comprised the process. For combining

these two tasks, namely, alignment and recognition, the paper proposed multi-task learning as a method. Three stages of convolutional networks that could recognize faces and also locations of major landmarks like the mouth, nose, and eyes comprised the process. For combining these two tasks, namely, alignment and recognition, the paper proposed multi-task learning as a method.

IV. METHODOLOGY

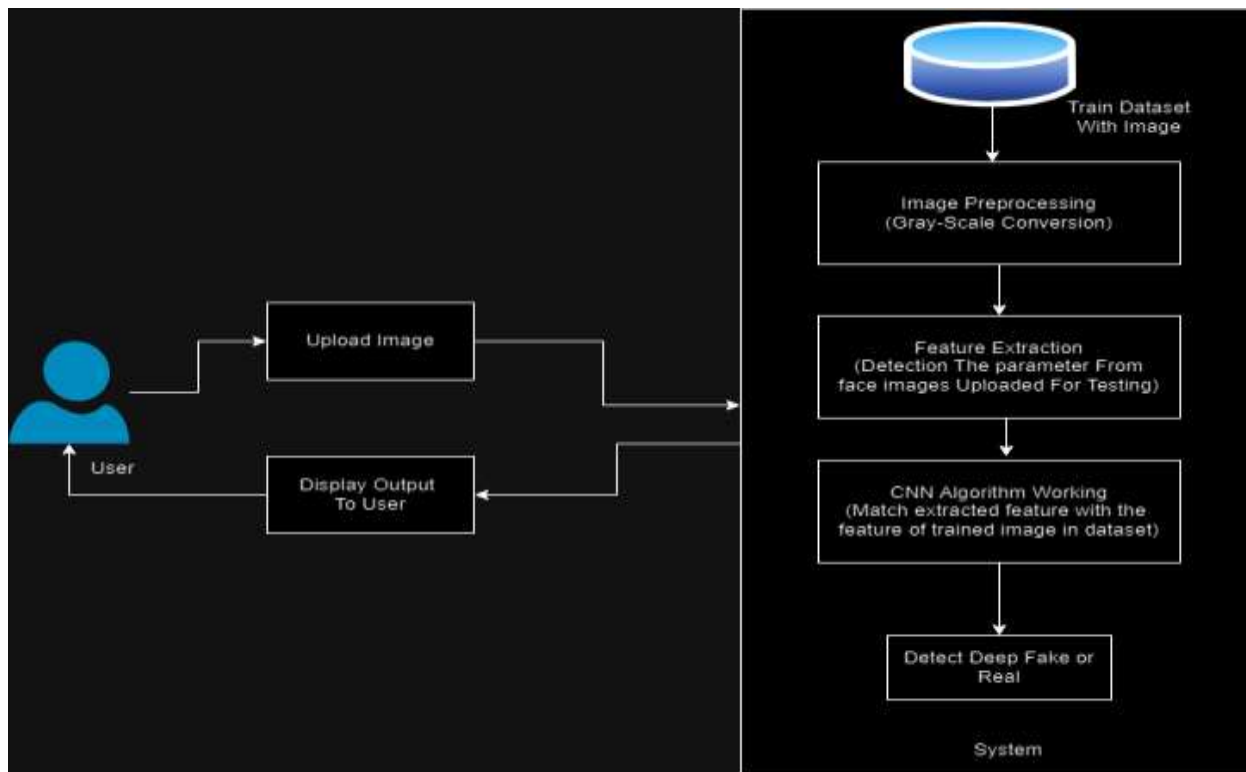
PROPOSE SYSTEM

Training the Model: Train the model using the training dataset. As the CNN model is trained, it learns to recognize relevant features in both authentic and deepfake content. Overfitting can be prevented by monitoring the validation loss and discontinuing training early when performance no longer improves. Loss Function Binary cross-entropy loss, commonly employed for deepfake detection tasks, ought to be applied to binary classification. Training accuracy, validation accuracy, loss, and other parameters like as F1-score, precision, and recall should all be tracked.



Propose System diagram

System Architecture:-



System Architecture Diagram

V. RESULT AND DISCUSSION

The method of recognizing faces automatically. Face recognition is the task of identifying or verifying a person's face in an image. The great difficulty is making sure this process is real-time, which is not the case for all biometric facial recognition software providers. The rapid advancement of deepfake technology poses significant threats to the integrity of digital media causing widespread misinformation, identity fraud, and degradation of faith towards the real content. It is difficult to spot deepfakes – realistic looking but ineffective media, altered by computer technology - because they usually come with real looking faces and movements. This scenario calls for the urgent need of effective detection mechanisms that can differentiate between real and fake media in all its forms. However, even with the available techniques, most of the current deepfake detection systems have issues such as poor generalization, low accuracy, and difficulties in learning new deepfake techniques. More over there are no easy to use solutions that the general public can use to be mindful about deepfakes and exposure to its variants. Therefore the issue has two components: a. Detection: It is extremely important that there is an efficient deepfake detection machine learning system that is reliable across a wide range of datasets, reduces false positive and negative rates and can evolve with new methods of deepfake generation. b. Awareness and Accessibility: Users need sophisticated tools that – while still detecting deepfakes – also mitigate the effects of such materials – by dealing with media literacy in a positive way. The objective of this project is to counter these issues by creating an all-inclusive deepfake detection system that incorporates modern techniques of machine learning without neglecting reach and user training, hence helping make the digital realm more tidy and educated.

VI. CONCLUSION

Here, in this paper, we review different approaches in Deepfake detection also using primarily deep learning approaches. As with other methods described above, these approaches have revealed very high performance in detecting such manipulations; none the less, their drawbacks have also been highlighted, especially in respecting different visual features. We therefore support the prospect of feature reduction in computer vision algorithms and, simultaneously, the investigation of these features as functional approaches to identifying new manipulations of facial identity or expressions. To this end, we have proposed a hypothesis for a real-time Deepfake detection and prevention model. However it should also be noted that for the time being this model is still theoretical and has not yet been implemented to be actually tested on any practical context. The development process has also brought up significant issues, such as the way the cost of using

this smart contract architecture on Ethereum affects its deployment, the way to link to large sets of pretrained images, and the time required to train deep learning models where the time can take days or even weeks. In our further research activities, we will pinpoint promising approaches on how to work with various types of detection and prevention schemes that can be used to counter digital media.

VII. REFERENCES

- [1] Rana, S., Gudla, C., Solaiman, M., & Sohan, F” Deepfakes - Reality Under Threat?”2024.
- [2] Md ShohelAndrew Sung” DeepfakeStack: A Deep Ensemble-based Learning Technique for Deepfake Detection”,2020.
- [3] Xin Yang? , Yuezun Li? and Siwei Lyu” EXPOSING DEEP FAKES USING INCONSISTENT HEAD POSES ”2019
- [4] Falko Matern Christian Riess Marc Stamminger “Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations”2019.
- [5] ISROEL MIRSKY, “The Creation and Detection of Deepfakes: A Survey”,2020.
- [7] Md. Shohel Rana Andrew H. Sung, “DeepfakeStack: A Deep Ensemble-based Learning Technique for Deepfake Detection”,2020.
- [8] Govind Mi Jiraphon Yenphraphaittal,” GOTCHA: A Challenge-Response System for Real-Time Deepfake Detection”2022.
- [9] FaceApp, [Online]. Available: <https://bit.ly/3Yw4TCG>, last accessed: 2023/11/2.
- [10] FaceApp, [Online]. Available: <https://bit.ly/40N5K3C>, last accessed: 2023/11/2.

