IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204124 |

Deep Fake Detection Using Neural Network

Sneha N R, Smitha C S, Sinchana T C, Sinchana G R, Shruthi K R

Department of Computer Science and Engineering, Malnad College of Engineering, Hassan, India

Department of Computer Science and Engineering, Malnad College of Engineering, Hassan, India

Department of Computer Science and Engineering, Malnad College of Engineering, Hassan, India

Department of Computer Science and Engineering, Malnad College of Engineering, Hassan, India

Assistant professor, Department of Computer Science and Engineering, Malnad College of Engineering, Hassan, India

ABSTRACT: Deepfake is a technique for human image synthesis based on artificial intelligence. Deepfake is used to merge and superimpose existing images onto source images using machine learning techniques. They are realistic looking fake images that cannot be distinguished by naked eyes. In this project the proposed system follows a detection approach of Deepfake images using Neural Networks. Deepfake Guard presents a novel convolutional neural network-based solution for detecting deep fake videos, a prevalent form of manipulated media posing serious threats to society. Currently, Cryptographic signing of images from its source is done to check the authenticity of images.

KEYWORDS: CNN, manipulated media, authenticity of images.

I. INTRODUCTION

The aim of this machine learning project is to detect the realistic looking fake images. Deepfake creation is a technique for human image synthesis based on neural network tools like GAN (Generative Adversarial Network) or Auto Encoders etc. These tools super impose target images onto source images using a neural network and create a realistic looking deep fake images. These deep-fake image are so real that it becomes impossible to spot difference by the naked eyes. In this work, we describe a new deep learning-based method that can effectively distinguish AI- generated fake images from real images. We are using the limitation of the deep fake creation tools as a powerful way to distinguish between the pristine and deep fake images.

A. Objective

- Discovering the distorted truth of the deep fakes.
- Reduce the Abuses and misleading of the common people on the world wide web.
- It distinguishes and classifies the images whether it is real or fake
- It concerns for personal privacy.

II. LITERATURE SURVEY

[1] Yuval Nirkin, Lior Wolf, Yosi Keller, and Tal Hassner; "Deep Fake Detection Based on Discrepancies Between Faces and their Context"-27 August 2020 In this work, they propose a novel detection cue which utilizes the commonalities of all recent face identity manipulation methods. It is complementary to conventional real/fake classifiers and can be used alongside them. This is in contrast to artifact detection methods, which are susceptible to the constant progress in the visual quality of generated images. The drawbacks of this paper are, Faces can exhibit a wide range of expressions, lighting conditions, and poses, making it challenging to establish a consistent context. Deep Fakes may exploit these variabilities to better blend with the context, making detection more difficult. Overcoming this approach would require a much broader integration of the new identity into the image, making our contribution hard to circumvent without additional technological breakthroughs.

[2] Nicholas Carlini, Hany Farid; "Evading Deepfake-Image Detectors with White and Black Box Attacks"-2020 To the extent that synthesized or manipulated content is used for nefarious purposes, the problem of detecting this content is inherently adversarial. We argue, therefore, that forensic classifiers need to build an adversarial model into their defences The drawbacks of this paper are, Limited Transferability: Adversarial examples crafted to fool one deepfake-

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204124 |

image detector might not generalize well to other detectors. The lack of transferability can limit the effectiveness of white-box attacks across different models. Model Updates: White-box attacks may become less effective if the target deepfake-image detector regularly updates its model parameters or employs mechanisms to detect and mitigate adversarial.

[3] Luca Guarnera, Oliver Giudice, Sebastiano Battiato; "Deep Fake Detection by Analysing Convolutional Traces"-2020 In this work, they focus on the analysis of Deepfakes of human faces with the objective of creating a new detection method able to detect a forensics trace hidden in images: a sort of finger-print left in the image generation process. The proposed technique, by means of an Expectation Maximization (EM) algorithm, extracts a set of local features specifically addressed to model the underlying convolutional generative process. The drawbacks of this paper are, the final result of the study to counter the Deepfake phenomenon was the creation of a new detection method based on features extracted through the EM algorithm. The underlying fingerprint has been proven to be Un-effective to discriminate between images generated by recent GANs architectures specifically devoted to generate realistic people's face. Some more works will be devoted to investigate the role of the kernel dimensions. Also the possibility to extend such methodology to image analysis and/or evaluate the robustness with respect to standard image editing (e.g. photometric and compression).

[4] Chih-Chung Hsu , Yi-Xiu Zhuang and Chia-Yen Lee; "Deep Fake Image Detection Based on Pairwise Learning"-2019-2020 In this paper, a fake feature network-based pairwise learning is proposed to detect the fake face and general images generated by the state-of-the-art GANs. The proposed CFFN can be used to learn the middle- and high-level and discriminative fake features by aggregating the cross-layer feature representations. The proposed pairwise learning strategy enables the fake feature learning, which allows the trained fake image detector to have the ability to detect the fake image generated by a new GAN, even it was not included in the training phase. The experimental results demonstrated that the proposed method outperformed other state-of-the-art methods in terms of precision and recall rate. The fake video detection is also an important issue, so they will extend the proposed method to fake video detection also, incorporating the object detection and Siamese.

[5]Recent advances in technology have made the deep learning (DL) models avail2able for use in a wide variety of novel applications; for example, generative adversarial network (GAN) models are capable of producing hyper-realistic images, speech, and even videos, such as the so-called "Deepfake" produced by GANs with manipulated audio and/or video clips, which are so realistic as to be indistinguishable from the real ones in human perception. Aside from innovative and legitimate applications, there are numerous nefarious or unlawful ways to use such counterfeit contents in propaganda, political campaigns, cybercrimes, extortion, etc. To meet the challenges posed by Deepfake multimedia, we propose a deep ensemble learning technique called Deepfake Stack for detecting such manipulated videos. The proposed technique combines a series of DL based state-of-art classification models and creates an improved composite classifier. Based on our experiments, it is shown that Deepfake Stack outperforms other classifiers by achieving an accuracy of 99.65.

III. DESIGN

The architecture of diagram is given below this shows the pre-processing phase which includes resizing of images, removal of noise and normalization.

Pre-Processing: For classifying the deep fake image pre-processing is needed which enhancing the image for further processing. The three stages of pre-processing namely, resizing of image, removal of noise and normalization.

- Resize image: In the data set, all the images were in various sizes and the processing of various size data could not provide accurate result.
- Removal of noise: In order to improve the efficiency in the classification of deep fake images, the noise was removed from raw input face image.
- Normalization**:** For the enhancement, the contrast of image was used by using normalization. It was carried out based on pixel intensity value.

Pipeline of Implementation for Deepfake Detection: The pipeline gives all the detailed information about the implementation of our system. A dataset of real and manipulated images was created in our system. After creating dataset, the images in the dataset were pre processed, it ensures uniformity in image size and format. Augment the data to enhance model generalization. The CNN model was created after the pre processing and then training was done on the training sets. Trains the selected model using the prepared dataset, adjusting hyperparameters as needed. After training, the model was run on the testing and validation sets. It assess the model's performance on a dedicated test

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204124 |

dataset to measure its effectiveness in real-world scenarios. Further, tuning the parameters and hyper parameters was done to increase accuracy and decrease loss of the system. Lastly, running the model on sample image and detecting classes for each frame of the image was done. Throughout the implementation process, iterative refinement and adaptation based on performance feedback and advancements in deep fake generation techniques are crucial to building an effective and reliable detection system.

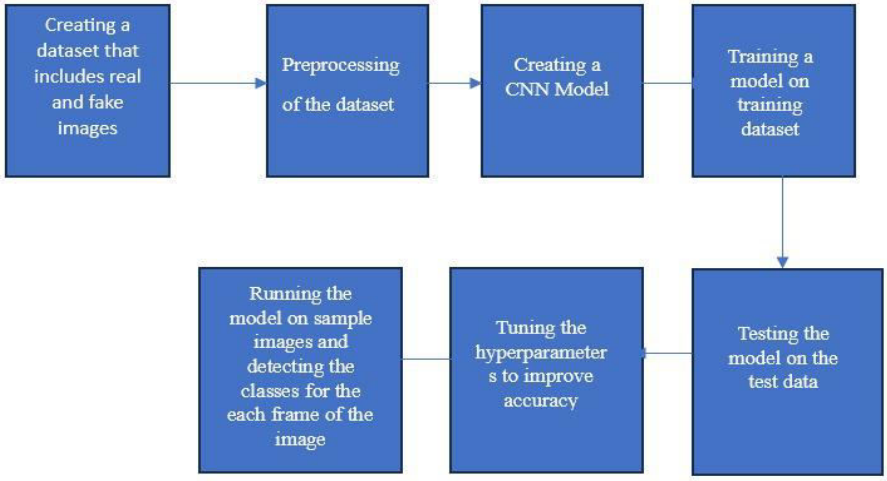

Figure 1. Block diagram

IV. IMPLEMENTATION

Neural Networks consist of network of neurons which are the computational units. The first layer is the input layer, there can be one or more hidden layers and last layer is the output layer .Neurons of one layer is connected to neurons of next layer through weighted connections. These connections allows information to flow through the network. Neurons consist of a number and an activation function. Activation functions are the non-linear functions used which determine the output of the neuron. The commonly used activation functions are ReLU (Rectified Linear Unit), tanh, sigmoid, etc. The connection between layers have weights present and every layer has a bias. Backpropagation in neural networks adjusts these weights and biases according to the label of the training data. Thus, the values of weights and bias of the real and deepfake manipulated frames are different. Similar properties in images cause similar neurons to fire and thus they have similar values of weights and biases.

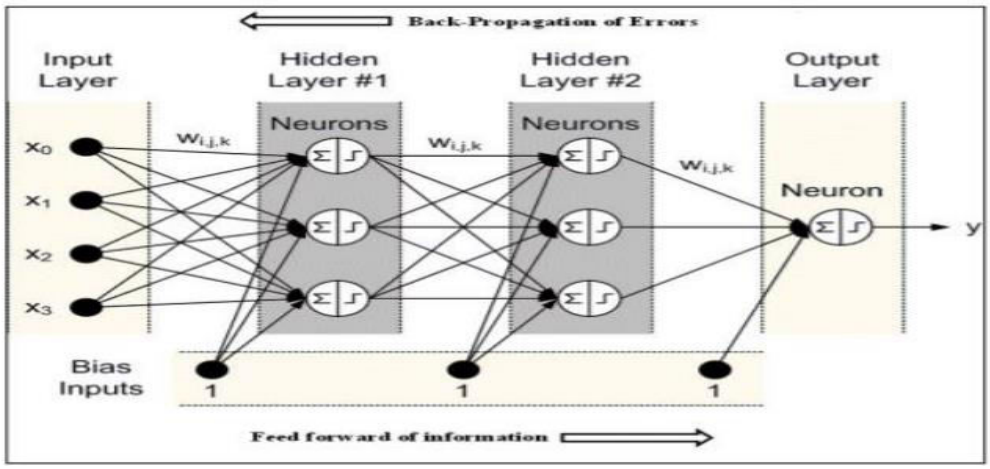

Figure 2. Architecture Diagram

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204124 |

V. RESULT

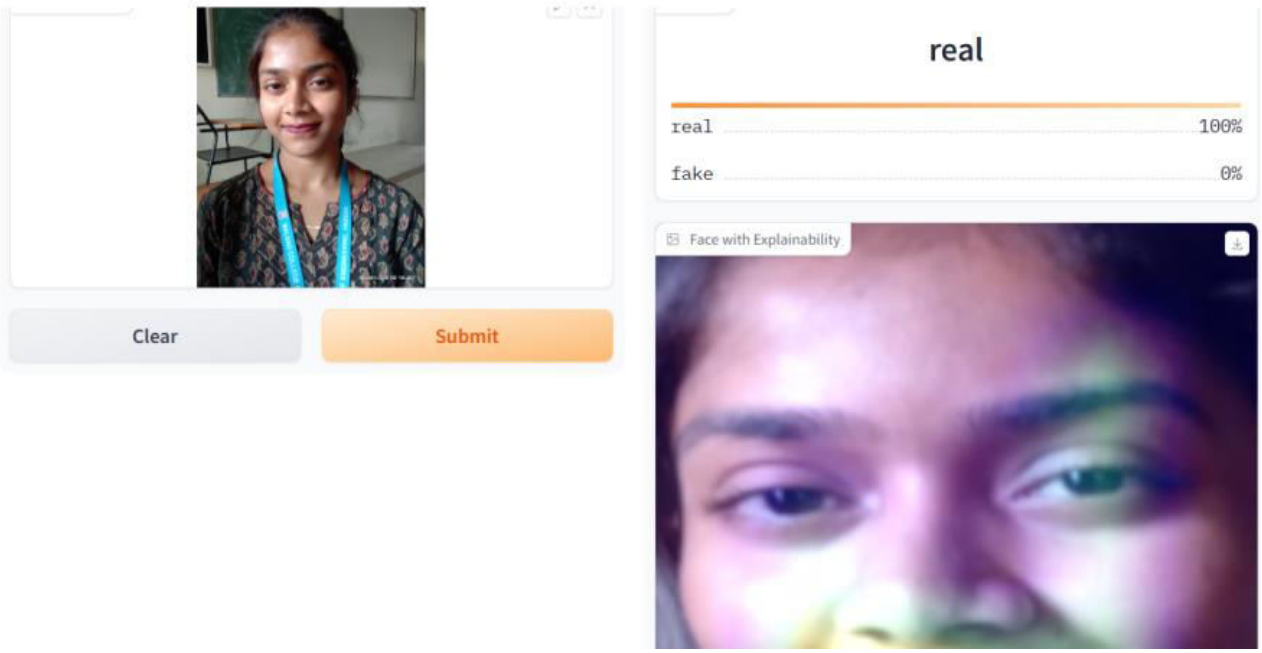

Figure 3: Real Image Detection

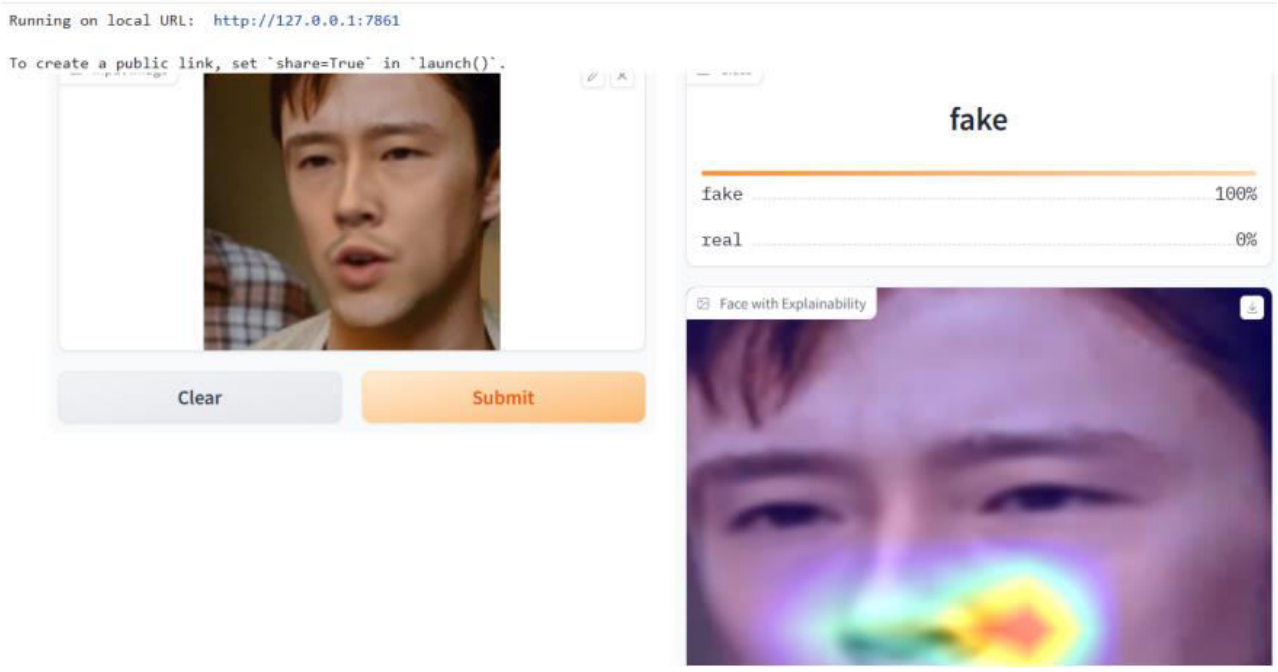

Figure 4: Fake Image Detection

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204124 |

VI. CONCLUSION

Different combinations of hyperparameters with respect to Neural Networks can be used and hyperparameter tuning can be done for the purpose of studying deepfakes. The outputs of those algorithm models can be analyzed and compared, so that Deepfakes can be combated in the most efficient way. Modern technologies like Blockchain can be used for immutable storage in order to preserve the originality of videos.

REFERENCES

[1] Yuval Nirkin, Lior Wolf, Yosi Keller, and Tal Hassner. Deepfake detection based on discrepancies between faces and their context. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10):6111–6121, 2021.
[2] Nicholas Carlini and Hany Farid. Evading deepfake-image detectors with whiteand black-box attacks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pages 658–659, 2020.
[3] Luca Guarnera, Oliver Giudice, and Sebastiano Battiato. Deepfake detection by analysing convolutional traces. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pages 666–667, 2020.
[4] Chih-Chung Hsu, Yi-Xiu Zhuang, and Chia-Yen Lee. Deep fake image detection based on pairwise learning. Applied Sciences, 10(1):370, 2020
[5] Md Shohel Rana and Andrew H Sung. Deepfakestack: A deep ensemble-based learning technique for deepfake detection. In 2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom), pages 70–75. IEEE, 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details