Deep Fake Face Detection

[1]Mr. Shlok Samudare, UG Student SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India, shloksamudre8139@gmail.com

[2]Mr. Jaideep Kure, UG Student SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India, kurejaideep2003@gmail.com

[3]Mr. Sanket Shinde , UG Student SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India, shindesanket9272@gmail.com

[4]Mrs. Himanshi Agarwal, Asst. Professor SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India, hagrawal.sknsits@sinhgda.edu

Abstract: Detecting deep fake faces has become increasingly crucial with the rise of sophisticated AI-based manipulation techniques. In this paper, we propose a novel deep fake face detection framework that leverages advanced machine learning algorithms and neural network architectures. Our approach integrates multi-modal features extracted from both spatial and temporal domains, enabling robust detection of manipulated facial images and videos. Through extensive experimentation on diverse datasets, we demonstrate the effectiveness of our method in accurately identifying deep fake faces while maintaining high computational efficiency. Our framework holds promise in combating the proliferation of fake visual content across various online platforms, thereby safeguarding the integrity of digital media and preserving trust in online communication channels.

Keywords: ***Deepfake, Deep Learning, Convolution neural network, Audio-Visual Analysis, Digital Trust , Data model.***

I. INTRODUCTION

In an era marked by swift technological advancements, the surge in digital content, and a growing reliance on visual information, the advent of deepfake technology has sparked profound concerns. Deepfakes, powered by sophisticated AI algorithms, manipulate and fabricate multimedia content with remarkable realism, posing unprecedented challenges to privacy, security, and information integrity. Among the many dimensions of this transformative technology, the manipulation of facial features and identity, commonly termed "deepfake face synthesis," stands out as a significant concern. This paper explores the dynamic domain of deepfake face detection, an area receiving considerable attention due to the potential repercussions of these manipulated visual identities. Deepfake face detection involves various techniques and methodologies aimed at identifying synthetic or manipulated faces within digital content. As deepfake technology evolves and becomes more accessible, the imperative for robust and effective detection mechanisms becomes increasingly apparent. The ramifications of deepfake technology extend beyond mere entertainment or artistic expression. Malicious deployment of deepfakes holds the potential to undermine trust in visual media, disrupt political landscapes, tarnish reputations, and even facilitate cybercrimes. Thus, there is an urgent need to develop and refine methods capable of reliably discerning authentic from manipulated facial content. This paper seeks to offer a comprehensive understanding of the current state of deepfake face detection techniques, the challenges faced by researchers and practitioners, and emerging solutions promising in addressing these challenges. Through this exploration, we aim to contribute to the ongoing discourse on deepfake technology and its societal implications, as well as guide future research endeavours in fortifying our defences against the risks posed by synthetic faces in the digital domain. In a world where trust in digital media holds paramount importance, the ability to reliably detect and authenticate facial content remains an ongoing challenge. This paper underscores the critical need for continuous research, development, and vigilance in the realm of deepfake face detection, emphasizing the importance of multidisciplinary efforts to safeguard the integrity of visual information in our increasingly digitized society.

II. LITERATURE SURVEY

Author: Rohita Jagdale , et. al. [1] have proposed a novel algorithm NA-VSR for Super resolution. The algorithm initially reads the low-resolution video and converts it into frames. Then the median filter is used to remove unwanted noise from video. The pixel density of the image is increased by bicubic interpolation technique. Then Bicubic transformation and image enhancement is done for mainly resolution enhancement. After these steps the design metric is computed. It uses the output peak signal-to-noise ratio

 DOI : 10.35291/2454-9150.2024.0194

(PSNR) and structural similarity index method (SSIM) to determine the quality of image. Peak signal-to-noise ratio and structural similarity index method parameters are computed for NA-VSR and compared with previous methods. Peak signal to noise ratio (PSNR) of the proposed method is improved by 7.84 dB, 6.92 dB, and 7.42 dB as compared to bicubic, SRCNN, and ASDS respectively.

Author: Siwei Lyu,[2] has surveyed various challenges and also discussed research opportunities in the field of Deepfakes. One critical disadvantage of the current Deep Fake generation methods is that they cannot produce good details such as skin and facial hairs. This is due to the loss of information in the encoding step of generation. Head puppetry involves copying the source person's head and upper shoulder part and then pasting it on the target person's body , so that target appears to behave in a similar way as that of the source. The second method is facing swapping which swaps only the face of the source person with that of the target. It also keeps the facial expressions unchanged. The third method is Lip syncing which is used to create a falsified video by only manipulating the lip region so that the target appears to speak something that she/he does not speak in reality. The detection methods are formulated as frame level binary classification problems. Out of the three widely used detection methods, the first category considers inconsistencies exhibited in the physical/physiological aspects in the Deepfake videos. The second algorithm makes use of the signal-level artifacts. Data driven is the last category of Detection in this, it directly employs multiple types of DNNs trained on genuine and Fake videos but captures only explicit artifacts. It also sheds some light on the limitations of these methods such as quality of

Author: Digvijay Yadav, et. al. [3] have elaborated the working of the deepfake techniques along with how it can swap faces with high precision. The Generative Adversarial Neural Networks (GANs) contain two neural networks, the first is generator and other is discriminator. Generator neural networks create the fake images from the given data set. On the other hand, discriminator neural networks evaluate the images which are synthesized by the generator and check its authenticity. Deepfake are harmful because of cases like individual character defamation and assassination, spreading fake news, threat to law enforcement agencies. For detection of Deepfakes blinking of eyes can be considered as a feature. The limitations for making Deepfakes are the requirement of large datasets, training and swapping is time consuming, similar faces and skin tones of people, etc. Deepfake video detection can be done using recurrent neural networks. CNN is best known for its visual recognition and if it is combined with LSTM, it can easily detect changes in the frames and then this information is used for detecting the Deep Fakes. The paper suggests that Meso-4 and Mesoinception-4 architectures are capable of detecting the Deepfake video with the accuracy of 95% to 98% on Face2Face dataset. Irene Amerind, et. al.

4] have proposed a system to exploit possible interframe dissimilarities using the optical flow technique. CNN classifiers make use of this clue as a feature to learn. The optical flow fields calculated on two consecutive frames for an original video and the corresponding Deepfake one is pictured and it can be noticed that the motion vectors around the chin in the real sequence are more vociferous in comparison with those of the altered video that appear much smoother. This is used as a clue to help neural networks learn properly. Face Forensics++ dataset was used, in that 720 videos were used for 3(training, 3000 videos for validation, and 3000 videos for testing. The uniqueness of this paper is the consideration of inter-frame dissimilarities, unlike other techniques which rely only on intraframe inconsistencies and how to overcome them using the optical flow-based CNN method

III. METHODS OF FAKE FACE DETECTION

Real-Time Detection:

We will prioritize real-time detection, allowing the system to identify deepfake content as it is being streamed or uploaded to digital platforms. This is critical for mitigating the rapid dissemination of potentially harmful deepfake media.

Large-Scale Dataset Augmentation:

To improve the robustness of our deepfake face detection system, we will create and curate a comprehensive dataset, incorporating a wide range of ethnicities, ages, and facial variations. The use of generative models will be explored to augment this dataset, simulating diverse deepfake scenarios. Continuous Learning and Adaptation: The system will be designed to learn and adapt to evolving deepfake generation techniques. Regular updates and fine-tuning of the detection models will ensure that the system remains effective in the face of new deepfake methods.

User-Friendly Interface:

The system will be developed with a user-friendly interface to encourage widespread adoption. This includes providing users with easy access to the deepfake detection service on various digital platforms

Convolution: Convolutional Neural Networks (CNNs) utilize convolution operations to extract features from input images. This process entails applying a filter, also known as a kernel, to the input image by sliding it across the image and computing the element-wise multiplication between the filter and the overlapping regions of the image. Mathematically, this operation is represented as the convolution of the input image with the filter and is commonly implemented using techniques like matrix multiplication or Fourier transforms.

Activation Functions: Activation functions introduce non-linearity into the CNN model, enabling it to capture complex patterns in the data. ReLU (Rectified Linear Unit), sigmoid, and tanh are among the common activation functions. These functions incorporate mathematical operations such as exponentiation and division to transform the input data into a non-linear output.

 DOI : 10.35291/2454-9150.2024.0194

Loss Functions: During the training phase of CNNs, loss functions quantify the disparity between the predicted outputs of the model and the actual ground truth labels. Cross-entropy loss and mean squared error are common loss functions utilized in classification tasks. These loss functions entail mathematical operations such as logarithms and summation of errors.

IV. OVERVIEW OF REMOVAL DEEP FAKES

Deepfakes are a form of synthetic media that involve using sophisticated machine learning techniques, particularly deep learning algorithms, to manipulate or generate highly convincing multimedia content, such as videos or images, often featuring human faces. Deepfake technology can seamlessly replace one person's face with another's or superimpose expressions and actions onto an individual, creating content that can be incredibly challenging to distinguish from authentic media. 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. All the images were resized as 256 × 256 and it was used for further processing. For resizing the input image down sampling and up sampling methods were employed .Eventually, as the disease progresses, the lesions enlarge and form reddish-brown spots on the leaves. A common symptom of bacterial infection is leaf spots or fruit spots. Unlike fungal spots, these are often contained by veins on the leaf.

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 by using Kalman filter. Generally, it is a recursive mathematical model and it consists of two different processes; the prediction process and the update process.

Figures :

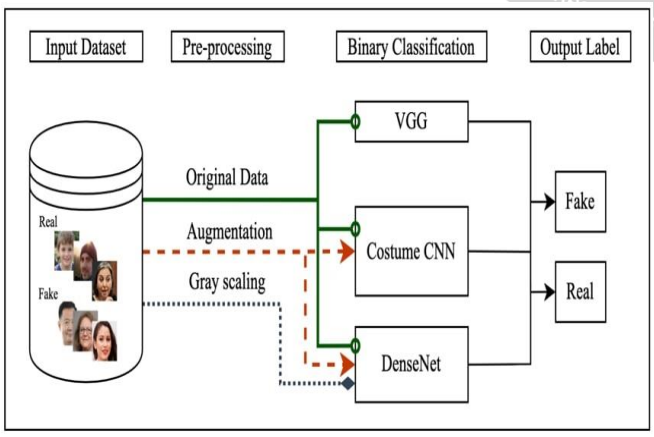

Fig 4.1 System Architecture

Normalization

For the enhancement, the contrast of image was used by using normalization. It was carried out based on pixel intensity value. The normalization process of this proposed work had used RGB pixel compensation method. It was based on the adaptive illumination of compensation dependent on the black pixel with histogram equalization.

Fig 4.2 fake face detection

The theory involves analysing facial geometry and landmarks to detect any deviations from the natural structure of a human face. Facial landmarks like the position of eyes, nose, and mouth are crucial in identifying warping or misalignment in deepfake images. Deepfake detectors examine the texture of the facial skin to detect irregularities. These irregularities can include variations in skin tone, reflections, and inconsistencies in skin texture that may be introduced during the deepfake generation.

Fig 4.3 REAL OR FAKE

So, these are our observation on how to classify the various fake Images and how to be catch that whether it is fake or not .

V. DISCUSSION

Furthermore, the proposed system will engage in collaborations with organizations and institutions involved in media integrity, cybersecurity, and digital forensics.

Deepfake generation techniques are continually evolving, becoming more sophisticated and difficult to detect. With the advent of novel generative models and techniques like GANs(Generative Adversarial Networks), it has become increasingly challenging to differentiate between real and Duplicate faces. This complexity underscores the need for deepfake detection mechanisms to continuously adapt and make it more perfect. Deepfake face detection is a dynamic and critical field with far-reaching implications. This discussion underscores the need for ongoing research and threats. It is only through concerted efforts that we can hope to navigate the complex challenges posed by deepfakes and

 DOI : 10.35291/2454-9150.2024.0194

safeguard the trustworthiness of digital media in our interconnected world. Development , multidisciplinary collaboration, and a proactive stance in the face of deepfake technology's potential .

Researchers have made significant strides in developing detection techniques, ranging from traditional methods based on facial landmarks and image forensics to sophisticated machine learning and deep learning models. However, despite these advancements, several challenges persist. The rapid evolution of deep fake technology poses a constant challenge to detection systems, requiring continuous adaptation and innovation. Additionally, the scarcity of large-scale labelled datasets hampers the development and evaluation of robust detection algorithms. Ethical considerations, such as privacy implications and the potential for censorship, further complicate the landscape of deep fake detection. Moreover, the democratization of synthetic media tools raises concerns about the democratization of disinformation and its potential impact on society. Addressing these challenges requires a concerted effort from researchers, policymakers, and industry stakeholders to develop effective detection methods, establish ethical guidelines, and promote media literacy among the public.

VI. CONCLUSION

In conclusion, the proliferation of deepfake technology presents significant challenges to the integrity of visual media and poses threats to privacy, security, and societal stability. The manipulation of facial features and identities, encapsulated within the realm of deepfake face synthesis, underscores the urgency for robust detection mechanisms. As deepfake technology advances and becomes more accessible, the need for effective detection techniques becomes paramount to mitigate its potentially harmful impacts. This paper has provided a comprehensive overview of the landscape of deepfake face detection, exploring various techniques and methodologies employed in identifying synthetic or manipulated faces within digital content. Despite the challenges faced by researchers and practitioners, emerging solutions offer promise in addressing these issues and fortifying defenses against the proliferation of deepfake content. The implications of deepfake technology extend far beyond mere entertainment, with potential consequences including erosion of trust in media, political disruption, reputational damage, and facilitation of cybercrimes. Thus, continued research, development, and vigilance in the realm of deepfake face detection are imperative to safeguard the integrity of visual information in our digitalized society. As we navigate an era where trust in digital media is of paramount importance, the ability to reliably detect and authenticate facial content remains an ongoing challenge. Multidisciplinary efforts are essential to stay ahead of the evolving threats posed by deepfake technology and to ensure the preservation of trust and integrity in our increasingly digital world. By remaining vigilant and proactive in the development and deployment of detection mechanisms, we can mitigate the risks associated with deepfake manipulation and uphold the integrity of visual information for the benefit of society as a whole.

VII. REFERENCES

[1] Face App. Accessed: Jan. 4, 2021. [Online].

Available: https://www.faceapp.com/

[2] Fake App. Accessed: Jan. 4, 2021. [Online].

Available: https://www.fakeapp.org/

[3] G. Oberoi. Exploring Deep Fakes. Accessed: Jan. 4, 2021. [Online].

Available: https://goberoi.com/exploringdeepfakes-20c9947c22d9

[4] J. Hui. How Deep Learning Fakes Videos (Deepfake) and How to Detect it. Accessed: Jan. 4, 2021. [Online].

Available: https://medium.com/how-deep-learningfakes-videos-deepfakes-and-how-to-detectit-c0b50fbf7cb9

[5] [Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," in Proc. IEEE/CVF Conf. Compute. Vis. Pattern Recognise. (CVPR) Workshops, 2019, pp. 46–52. [Online]. Available: https://openaccess.thecvf.com/content_CVPRW_2019/html/Media_Forensics/Li_Exposing_DeepFake_Videos_By_Detecting_Face_Warping_Artifacts_CVPRW_2019_paper.html.

[6] G. Patrini, F. Cavalli, and H. Ajder, "The state of deepfakes :Reality under attack," Deep trace B.V., Amsterdam, The Netherlands, Annu. Rep. v.2.3., 2018. [Online].

Available:https://s3.euwest2.amazonaws.com/rep2018/2018-the-stateof-deepfakes.pdf

[7] M. S. Rana et al.: Deepfake Detection: Systematic Literature Review

[8] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niessner,"Face2Face: Real-time face capture and reenactment of RGB videos ,"in Proc. IEEE Conf. Compute. Vis. Pattern Recognise. (CVPR), Las Vegas,NV, USA, Jun. 2016, pp. 2387–2395, doi: 10.1109/CVPR.2016.262.

[9] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Oct. 2017, pp. 2242–2251

 DOI : 10.35291/2454-9150.2024.0194