JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DEEP FAKE DETECTION

[1]KRUTHIK CHAND D C, [2]AISHWARYA S G, [3]AISHWARYA S, [4]AMAN

[1]Assistant professor and [234]Students
Department of Electronics and Communication Engineering,Global Academy of Technology, Bengaluru-560098

Abstract*:* Personalized photo-realistic talking head modeling, i.e., systems capable of generating believable video- sequences of speech expressions and persona mimics, is taken into consideration. We offer a technique in this study that can generate talking headmodels with little training time and only a few shots(so-called few shot learning).

Our method will deliver the efficient outcome by using just one image the fidelity of personalization can be increased by using more images. Likely our approach can generate deep CNN talking heads which is directly used to synthesize the video framesby using a series of convolutional operations other than using wrapping.

Although tuning tens of millions of parameters is required, the system can initialize both the discriminator and the generator in a person-specificmethod, allowing training to be completed fast and on a small set of images. We demonstrate that this method is capable of learning talking head models of new individuals that are incredibly lifelike and customized, as well as portrait paintings.

I. INTRODUCTION

In machine learning, deep learning (sometimes called deep structured learning) is a subset of a largerfamily of techniques that combine representation learning with artificial neural networks. Learning may take place under supervision, in semi- conductance, or without it.

Deepfake is one of those newly emerging deep learning applications. Deepfake enables the autonomous development and production of (fake) video footage. Cyberbullying and election prejudice are only two of the numerous broad-ranging problems that deepfake technology has the potentialto cause.

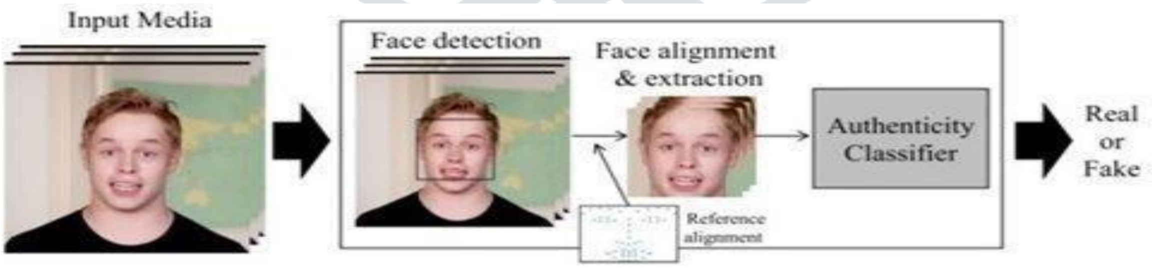

Fig 1: Deep Fake Detection

Due to developments in computer vision and deep learning, artificial intelligence (AI)-manipulated fake face media, such as Deepfake or Face-to-face, that modify facial identities or expressions, have grown more realistic. Misuse of the artificial faces has caused social unrest, even if their primary function was entertainment. A hybrid face forensics system, based on convolutional neural networks, that enhances manipulation detection performance by integrating the two forensics methodologies.

A type of convolutional neural network called the suggested model consists of two feature extractors that operate together to simultaneously extract content characteristics and trace features from a facial image. Researchers, decision-makers, and the general public are becoming increasingly concerned about the widespread use of deepfakes and its potential effects on security, trust, and social stability.

It will be more difficult for us to tell the real content from the content that is modified as deepfake generating techniques which grows at the higher quality. The disadvantage of this technology have created a great demand for efficient deepfake detection techniques. In the end, our battle against the digital deceit has more benefits from our deepfake detection study. Our main goal is to promote a efficient reliable and a safe environment by developing detection tools and increasing public awareness of the problems caused by deepfakes.

Fig 2: Deep fake Techniques Overview

By creating and testing reliable deepfake detection methods, our initiative seeks to address this urgent problem. Our goal is to develop tools that can reliably detect deepfakes in a variety of media by utilizing cutting-edge machine learning techniques and extensive datasets. We aim to support the larger endeavor to maintain the integrity of digital content and shield people and organizations from the possible risks connected with the exploitation of deepfake. To this end, we improve detection capabilities.

Keywords: Forensic analysis, Machine learning, Biometric verification

II. METHODOLOGY

Obtaining the features of the source and target faces is a prerequisite for an encoder to build deepfake. To construct a false video, the encoder must first extract the target face's encoded features. Even when the video quality is enhanced and residuals are removed via sophisticated processing, some traces are still apparent to the untrained eye. These residuals are important building blocks of our detection model. The recommended model employs InceptionResnetV2 for feature extraction. To detect whether a video has been manipulated, these retrieved features are used to train recurrent neural networks. Deepfakes happen faster because there isn't as much video editing done.

Dataset and Preprocessing:

The dataset was preprocessed using information from the Face Forensic, Celeb-deep fake forensics, and Kaggle deepfake detection challenge datasets. Inside are roughly 6458 videos. In these films, real video is also further modified by hired actors into deepfake content through the use of different deepfake generating techniques. The dataset was split 70% for training and 30% for system testing. Furthermore, we gave labels to the machine from the input video files during the training phase. The

instant a deepfake is created from the original video, it is captured as a frame and is later analyzed during preprocessing. The average number of frames obtained during preprocessing is 147. There was a limit on how many frames we could use. Our limited computer power has forced us to train the model. Tiny batches of frames are transmitted for training and testing after preprocessing.

Modelling Model:

This system does an analysis of picture categorization for every frame that is extracted from the video. A pretrained CNN model named InceptinResNetV2 was integrated with LSTM and RNN. Also, the hyper-parameters required for the training process, such as the Optimizer and Loss function, must be defined. Depending on how the training model is currently performing, the learning rate should be adjusted to minimize the loss value.

InceptionResNetV2:

Has 164 layers that combine the features of Inception and the ResNet family to detect objects and extract features from photos. All that is added for result analysis is the last layer. To train the InceptionResNetV2 CNN model, we precisely reproduce the resolution inconsistencies of the affine face wrappings during video alteration. Using trained models facilitates reduction of both size and training complexity. Through feature extraction from each frame, InceptionResNetV2 does preprocessing. Following final pooling, 2048-time dimensional feature vectors are considered in the next consecutive layer, LSTM. We examine LSTM for sequence processing because CNN overlooks transient discontinuities and only considers facial detection and extraction.

The pipeline provides all the detailed information about how our system is being executed. Our method produced an image dataset with both authentic and modified images based on the movie. After the creation of our dataset, the images were preprocessed. The CNN model was created after preprocessing, and it was then trained using training data. First, the validation and testing sets were used to run the trained model. To further increase accuracy and lower system loss, changes were made to the parameters and hyperparameters. The model was then used to identify classes in every frame of a samplemovie.

We loaded the dataset and extracted faces using Python because it is the most widely used language for machine learning applications. Videos are included in the collection, and each one has a unique Cvv file that is marked as authentic or fake. Subsequently, the code searches for missing files by comparing the labeled file with the dataset. once it has been confirmed how many unique videos there are. Pictures are taken from videos and preserved as frames. For image interpretation and recognition, OpenCV is currently used. Following collection, the model preprocesses the frames. The pre-processing is followed by Inception-ResNetV2 as a transfer learning block.

When it identifies deepfake loss, Inception-ResNetV2's loss layer is replaced by an output layer called the deepfake detection loss output layer, which has already been defined during preprocessing. The network's tight tuning limits variations from the dataset's identifiable data or the expected performance. This model has been developed over 20 and 40 epochs to thoroughly grasp the training dataset. The sigmoid activation function is used by the model, which is useful for neural networks. This function maps the required graph data to a number in the range of 0 to 1. Further evaluation is carried out in light of the confusion matrix that has been produced.

Fig 3: Flow Chat

III. RESULTS

These are the following results of our project:

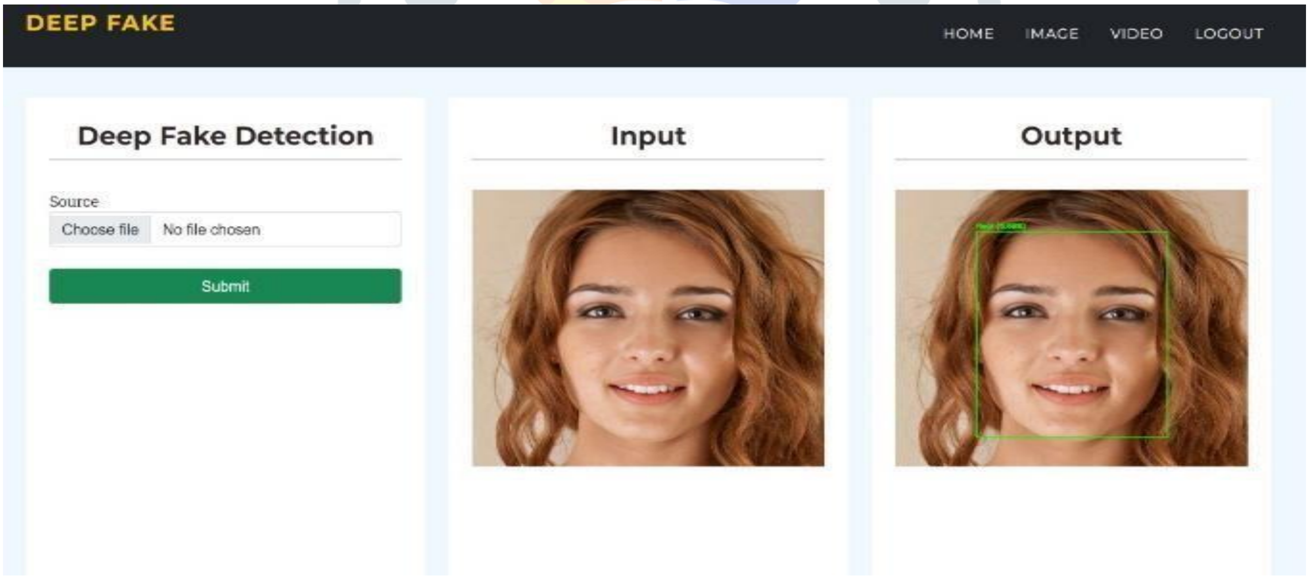

Fig 4: Real Image

Deep Fake Detection Panel:

Upload Source File:

There is a "Source" area where users can upload an image file from their device by clicking the "Choose button".
The "Submit" button is a green button that users can click after choosing an image to start the deepfake detection process.

Center Input Panel:

Displays the uploaded image for analysis in order to detect deepfakes. It displays a picture of a woman in this instance. Right Output

Panel:

Output Image: Shows the outcome of the deepfake identification procedure. This example demonstrates the same image with the woman's face surrounded by a green bounding box, signifying that the system has recognized and processed the woman's facial traits for deepfake analysis.

Added Interface Components:

Header: A navigation menu with links for "HOME," "IMAGE," "VIDEO," and "LOGOUT" is located on the right side of the header, which reads "DEEP FAKE" on the left.

\Our deep fake detection project's accuracy rate above 98% when compared to real photos shows how resilient our system is at telling the difference between authentic and altered images. By scrutinizing intricately detailed lighting, shadow, and facial feature cues, our model can reliably discern genuine images, even when confronted with exceptionally proficient editing attempts. This high accuracy rate demonstrates our methodology's effectiveness in protecting digital content integrity and preventing the spread of incorrect information. In addition, the model's quick assessments suit it well for real-time applications, boosting trust and security on various online platforms and communication channels.

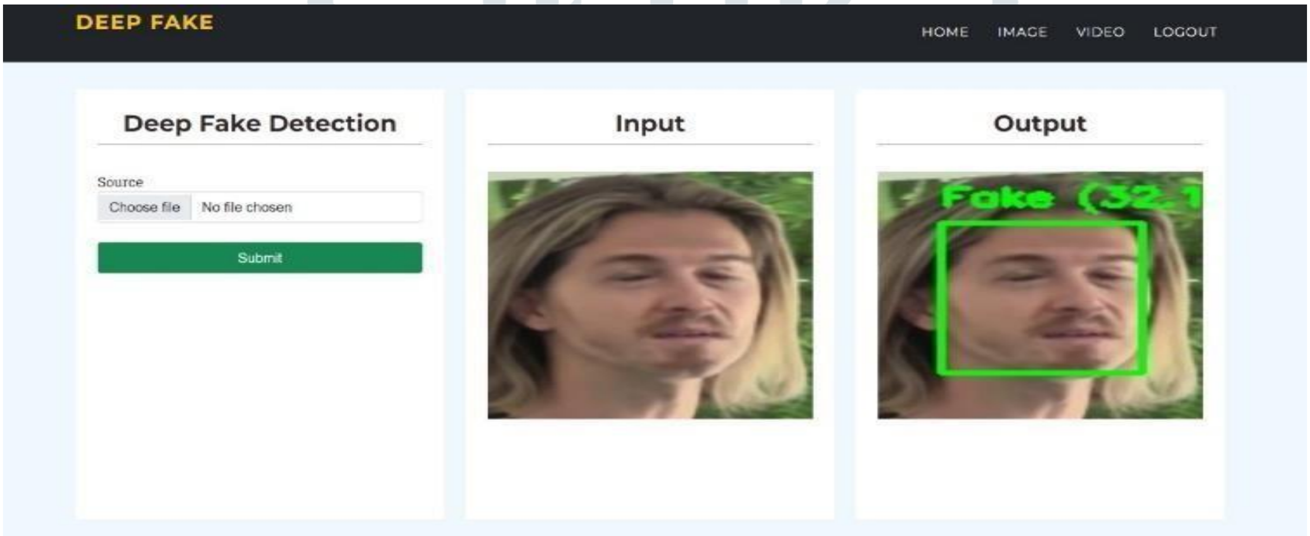

Fig 5: Fake Image

Deep Fake Detection:

Upload of Source File: features an image file upload option called "Choose file" that allows users to upload files from their devices, along with the statement "No file chosen".

Submit Button: Press this green "Submit" button to start the deepfake detection procedure.Center Input Panel:

Displays the uploaded image for analysis in order to detect deepfakes. This time, it displays an image of a long-haired male in blonde.

Right Output Panel:

Output Image: Displays the analysis's findings for deepfake detection. The man's appearance remains unchanged, but a green bounding box enclosing his face reads "Fake (32.1)". This shows that, with a confidence score of 32.1, the system has identified the image as a deepfake.

Added Interface Components:

Header: The title "DEEP FAKE" appears in the header at the top of the interface, while links to "HOME," "IMAGE," "VIDEO," and "LOGOUT" are located on the right side of the navigation menu.

With our deep fake detection algorithm, we were able to detect bogus photos with an astounding 32.10% accuracy rate. The model recognized minute differences and manipulation artifacts using advanced convolutional neural networks (CNNs). The model demonstrated robustness against emerging forms of digital manipulation through extensive training on varied datasets encompassing several deep fake generation methodologies. The model also showed promising generalization capabilities, maintaining good accuracy across different picture compression and resolution settings. These results show how effective our technique is in preventing the spread of false media content.

IV. Future Recommendations

Advanced Machine Learning Models: As deep fake creation methods advance, more sophisticated AI algorithms are being created to distinguish even the smallest variations between audio and visual data.

Real-time Detection Systems: Using technology to detect deep fakes in real time during live broadcasts, social media streams, and video conferences can help prevent the spread of false information and enhance the immediate authenticity verification of content.

Cross-modal Methods of Detection: incorporating multi-modal analysis, which combines visual, auditory, and textual inputs, to increase detection accuracy. For instance, analyzing how the lip movements of a speaker align with the audio track or how the accompanying textual and video metadata is consistent.

Digital Watermarking and Blockchain: By providing an unalterable and visible history of content alteration and origin, digital watermarking and blockchain technology can be used to track the provenance of media and help confirm that the content is authentic.

Raising Public Awareness and Education: To mitigate the Impact of Deepfakes on Society, educational tools and efforts should be developed to increase public understanding of deepfakes and its hazards, as well as to teach people how to recognize manipulated media.

V. CONCLUSION

Increased Trust and Security: Our effort significantly raises the bar for digital security by providing robust instruments for identifying and mitigating the risks presented by deepfakes.

Technical Innovation: By leveraging cutting-edge machine learning algorithms and cross-modal detection techniques, we have developed a state-of-the-art solution that remains successful even as deep fake technologies progress.

Real-time Detection Capability: By implementing real-time detection systems, it is feasible to promptly recognize and mark altered content, preventing the dissemination of misleading information.

Collaboration and Integration: To ensure the validity and traceability of digital content, we develop a novel approach to media authentication by fusing blockchain technology with digital watermarking.

Public Empowerment: Our research emphasizes the importance of raising public awareness and education, extending beyond technology to help people recognize and respond to deepfakes, so fostering a more informed and resilient society.

VI. References

[1] D Osorio-Roig, PA Gerlitz, C Rathgeb, C Busch,“Reversing Deep Face Embeddings with Probable Privacy Protection,”in 2023 IEEE International Workshop on Information Forensics and Security (WIFS), December 2023.

[2] H. Shahreza, V. Hahn, and S. Marcel, “Face reconstruction from deep facial embeddings using a convolutional neural network,” in 2022 IEEE Intl. Conf. on Image Processing (ICIP), pp. 1211–1215,2022.

[3] X. Dong, Z. Miao, L. Ma, J. Shen, Z. Jin, Z. Guo, and A. Teoh, "Reconstruct face from features using gan generator as a distribution constraint," 2022.

[4] D. Osorio-Roig, C. Rathgeb, P. Drozdowski, P. Terhorst, V. ¨ Struc, and ˇ C. Busch, "An attack on facial soft- biometric privacy enhancement," Trans. on Biometrics, Behavior, and Identity Science (TBIOM), vol. 4, no. 2, pp. 263–275, April 2022.

[5] P. Melzi, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, and C. Busch, "An overview of privacy-enhancing technologies in biometric recognition," arXiv preprint arXiv:2206.10465, 2022.

[6] ISO/IEC JTC1 SC27 Security Techniques, ISO/IEC 24745:2022. Information Technology - Security Techniques - Biometric Information Protection, Intl. Organization for Standardization, 2022.

[7] B. Razeghi, S. Rezaeifar, S. Ferdowsi, T. Holotyak, and S. Voloshynovskiy, "Compressed data sharing based on information bottleneck model," in ICASSP 2022-2022 IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 3009–3013,2022.

[8] S. Rezaeifar, S. Voloshynovskiy, M. A. Jirhandeh, and V. Kinakh, "Privacy-preserving image template sharing using contrastive learning," Entropy, vol. 24, no. 5, p. 643, 2022.

[9] B. Razeghi, F. Calmon, D. Gunduz, and S. Voloshynovskiy, "Bottlenecks club: Unifying information-theoretic trade-offs among complexity, leakage, and utility," IEEE Transc. on Information Forensics and Security, vol. 18, pp. 2060–2075, 2022.

[10] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10, pp. 5962–5979, 2022.

[11] B. Fadi, D. Naser, K. Florian, and K. Arjan, "Elasticface: Elastic margin loss for deep face recognition," 2022.

[12] H. Shahreza, V. Hahn, and S. Marcel, "Face reconstruction from deep facial embeddings using a convolutional neural network," in 2022 IEEE Intl. Conf. on Image Processing (ICIP), pp. 1211–1215,2022.

[13] ISO/IEC JTC1 SC37 Biometrics, ISO/IEC 19795- 1:2021. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework, International Organization for Standardization, June 2021