Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

DETECTION OF DEEPFAKE VIDEO AND IMAGE FORENSICS IN WARPING ARTIFACT USING DEEP LEARNING

Ms.Mahalakshmi Dheenan
Assistant Professor Department of Information Technology Panimalar Engineering College
mahalakshmi1607cs@gmail.com

Priyadharshini N
Department of Information Technology Panimalar Engineering College
priyadharshininr485@gmail.com

Priyadharshini E
Department of Information Technology Panimalar Engineering College
priyadharshinielumalai05@gmail.com

Shanya M
Department of Information Technology Panimalar Engineering College
shanyam341@gmail.com

ABSTRACT - Deepfakes are manipulated, excellent, lifelike films and photos that have become increasingly popular recently. These days, AI - generated videos and images of famous people, political leaders, and other public figures is circulating on the internet. It is very difficult to identify whether the visuals are real. Using deepfake visuals in movie stunt scenes is essential ensuring the safety of the actor. We came across AI generated news readers; these deepfakes are useful in one way or another. It also grew tremendously on the negative aspect. This presentation presents a detection of deepfake using Deep Learning algorithm. The motive of this project that leverages videos and images of political leaders and celebirities in order to detect data originality. The stages of the proposed approach are as follows : Gathering the video and image files of people from online sites, Pre-processing, Extracting features, Deep Learning Algorithms and Prediction

Keywords: EfficientNet-B0, Deepfake video detection, Deepfake image detection, Pixel deformations, Neural Network, Facial landmarks, Deep Learning Techniques, MySQL.

I. INTRODUCTION

A thorough analysis of DeepFake, a recent and well-known technique. It describes the principles, benefits, and dangers of DeepFake and DeepFake applications based on GANs. Even a person with no knowledge on deepfake can use the FaceApp or Face swap applications, which gives the modified content. It is very difficult for an individual to identify the truthfulness of the data.

Moreover, DeepFake detection models are also discussed. The majority of deep learning based detection techniques now in use lack the capacity to transfer and generalize, suggesting that multimedia forensics is still in its infancy. Numerous significant companies and professionals who are advancing applied approaches have expressed a great deal of interest. Other forms of security are necessary since maintaining data integrity still requires a lot of work. Experts also predict a fresh wave of DeepFake propaganda in AI vs AI conflicts in which no side has the upper hand. Several deepfake detection approaches have been presented in this effort on picture forensics of generic image alterations. This used noise due to erroneous geometry and light predictions, as well as colour mismatch in two eyes. Two fundamental fake-face detection networks were proposed by Deepfake, which leverage macroscopic features by taking advantage of the difference in 2D direction between the entire head and the restricted facial area. The colour space was changed to HSV because the training employs an RGB colour space distribution. Deepfake was then found by examining the statistical differences between the two colour spaces.

II. RELATED WORKS

Ahmad Neyaz Khan, Sani M. Abdullahi, Minoru Kuribayashi, AsadMalik[1] a research on deepfake detection was published as “DeepFake Detection for Human Face Images and Videos: A Survey”. The original content is either created or altered by deepfake, making it challenging to distinguish between actual and fake photos. Interest in DeepFake detection has increased dramatically research employing Deep Neural Networks (DNNs) to detect and categorize DeepFakes. Siwei Lyu , Xin Yang, Honggang Qi, Yuezun Li, , Pu Sun[4], “Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics”, DeepFakes, or artificial intelligence-generated face-swapping movies, are a growing issue that jeopardizes the reliability of internet information. DeepFake detection systems cannot be developed or evaluated without large-scale databases. We provide CelebDF, a brand-new, highly challenging DeepFake video dataset with 5, 639 top-notch celebrity DeepFakes created using an improved synthesis method. Nenghai Yu, Hui Xue, Honggu Liu, Weiming Zhang, Xiaodan Li, Yuan He, Wenbo Zhou, Yuefeng Chen[3], “Spatial-Phase Shallow Learning: Rethinking Face Forgery Detection in Frequency Domain”. In order to increase transferability, they provide a unique approach called SPSL, it fin distortions of facial counterfeiting by combining phase spectrum and spatial image. Dong Chen ,Hao Yang, Baining Guo, Lingzhi Li, Ting Zhang, Fang Wen , Jianmin Bao[6] , “Face X-ray for More General Face Forgery Detection”, is a proposed method for identifying different kinds of face forgeries, such as face swaps, deepfake films, and other facial manipulation techniques. The method is made to be strong and reliable in spotting anomalies in photos or films of altered faces, even when the human eye is not able to detect them.

Oliver Giudice, Luca Guarnera, and Sebastiano Battiato[7], “Fighting Deepfake By Exposing The Convolutional Traces On Images”, This technique looks for distinctive patterns linked to deepfake manipulation by examining convolutional traces. Their goal is to effectively distinguish real and altered photos by teaching a detection model to identify these patterns. D. Warde-Farley, S. Ozair, I. Goodfellow, Y. Bengio, J. Pouget-Abadie, B. Xu, M. Mirza , A. Courville [5] , ‘‘Generative adversarial nets, ”’A novel approach for dynamic model estimation has been introduced to instruct two models simultaneously: A model with discriminatory characteristics ascertains the probability that an instance comes from the training set of data and model that is generative depicts the data distribution.

J. Kannala, E.Rahtu and S.Tripathy[10] ‘‘ICface: Interpretable and controllable face reenactment using GANs,’’ This research presents an animator for faces in general that manipulates expressions and position of a face picture using manipulable signs of control, such as angles of the head posture and Action Unit ratings. M. N. Nobi, M. S. Rana, A. H. Sung, B. Murali [2],‘‘DeepFake detection: A systematic literature review’’, This study reviews 112 papers from 2018 to 2020 on Deepfake detection, analyzing techniques based on deep learning, traditional machine learning, statistics, and blockchain. The review categorizes these techniques into four groups: traditional machine learning, Deep learning, blockchain and statistics. It assesses the detection capabilities of each approach for diverse data sets, providing an updated overview of research efforts in this area.

S. Raj, S.K.Jha, S. Fernandes, E.Ortiz ,R. Ewetz, M. Salter, J.S.Pannu, I.Vintila[11] “Detecting deepfake videos using attribution based confidence metric”, The ResNet50 model is pretrained using VGGFace2 approach, and the methodology employed here is the attribution based confidence (ABC) measure. When training data is not available, deepfake movies are identified using the ABC metricL. Yin, I. Demir, and U.A. Ciftci[9], “Fakecatcher: Detection of synthetic portrait videos using biological signals”, Phony portrait videos pose a threat to privacy, law, and society. Deepfakes can lead to people believing fakes, sharing celebrity- pornography, and creating court-use evidence. The detector uses generative models, which yield realistic results, indicating that biological signals in portrait films are not geographically or temporally specific. Z. Zeng, Y. Yang, T. Huang, S. Wen, and W. Liu[12],‘‘Generating realistic videos from keyframes with concatenated GANs’’, suggested a concatenated architecture made up of many GANs, each of which is in charge of producing intermediate frames that relate to distinct temporal periods. To assess the method's performance and show that it can produce realistic films with a variety of content and motion patterns, we do in-depth tests on benchmark datasets. R. Zhang , O. Wang, S.-Y. Wang, A.A.Efro, A.Owens[8], “Cnn generated images are surprisingly easy to spot... for now”, The research investigates the creation of a universal detector that can differentiate between real images and CNN. A dataset of fictitious images from eleven CNN-based models is collected, showing that a typical image classifier can generalize well to other architectures and datasets.

III. EXISTING SYSTEM

Many have worked with detecting the deepfake video using many techniques like face wrapping artifacts, LSTM, RNN, a capsule network detect forged and manipulated videos and biological signs. They produced a limited resolution picture, some of the models performed beneficial with their given dataset but fail with real time data due to noise in training. Datasets consist of small number of images which may not perform very well on real time data.

DISADVANTAGES

- The majority of the times, deepfake news and propaganda are created as retaliation.
- When the fake picture gets viral, people believe the information initially and they keep sharing these content with others and makes them a targeted person.
- Privacy problems
- Datasets are very less to train with real time data

IV. PROPOSED SYSTEM

We present a generalized detection approach, in this work we develop an application that can be used by an individual, whenever they have doubt regarding the originality of the data. This model consists of a Home page where people will be given awareness about the deepfake. A registration form to be filled by the new user and these details are stored in MySQL workspace. After registration the user can login, Followed by a workspace area, where it consist of two sections one area to work with the deepfake video detection followed by the next section, which focus on detecting image. This proposed system helps us to identify the truthfulness of the data. Residual noise present in data is to be removed. EfficientNet B0 as a feature extractor enables the extraction of significant representations from pictures and movies. Next we go with finding the abnormal high-level features that are found in the image or video frame. To overcome the blur effects in image we utilized Image quality measurement features. If faces are not identified in the image then a note is displayed indicating that the image has no face. Gesture recognition is done to identify the conscious and unconscious movement of body parts like hand, shoulder. The body languages are taken for training. The frames are exposed to convolutional operation to identify the pixel changes. This method has two types of input: video and image. The video is split into frames and each frame is exposed to the above mentioned techniques. We also make use of facial artifacts in situations like face swap or face morph. The outcomes showed that every detection method works and the suggest network is more user friendly. A face swapped image is constructed online, and it is given as input, the model works and produces the correct results.

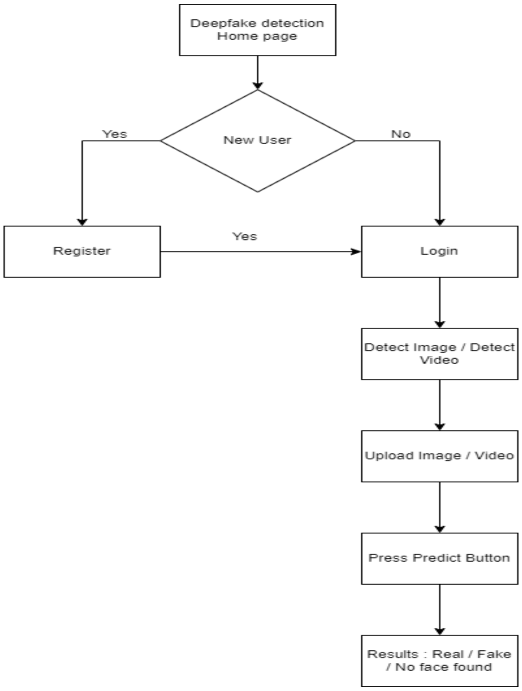

Deepfake detection Home page
New User
Yes
No
Register
Yes
Login
Detect Image / Detect Video
Upload Image / Video
Press Predict Button
Results : Real / Fake / No face found

Fig 1: Overview of proposed system

ADVANTAGES

- Single web application to detect deepfake video and images

- User friendly
- Attracts large number of target audience
- The residual noise present in the image or video frame are removed
- Real time images and videos can be given as input and the model performs well.
- Image or video with no face is notified
- Data is secured
- MySql database to store the user details for authentication
- Face swap images are identified in real time.

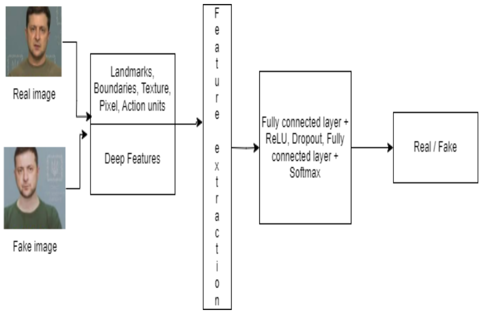

Real image
Fake image
Landmarks, Boundaries, Texture, Pixel, Action units
Deep Features
Feature extraction
Fully connected layer + ReLU, Dropout, Fully connected layer + Softmax
Real / Fake

Fig 2: Architecture diagram for deepfake image detection

V. METHODOLOGY

DATASETS

Using human-pose estimation, deepfake video detection is the suggested method. Since the publicly accessible datasets do not adhere to the recommended guidelines for facial landmarks, they are not suitable for use in this study. The only films of world leaders (presidents, vice presidents, etc.) used to train and test. These data have an excellent pixel ratio, are publicly available, and allow us to swiftly test them against deepfakes. To create a bespoke dataset, we thus manually downloaded movies from the internet. These films were labelled and annotated in accordance with the suggested specifications. To test the suggested theory, we'll need two types of videos: synthetic and original. The first versions were downloaded from the official site in order to ensure that the recordings had not been tampered with. The prerequisites are satisfied by the video download. The dataset samples must satisfy the criteria: file formats are MP4 for video, Face of an individual should be there in the video, images and videos should be taken with proper lighting on the face.

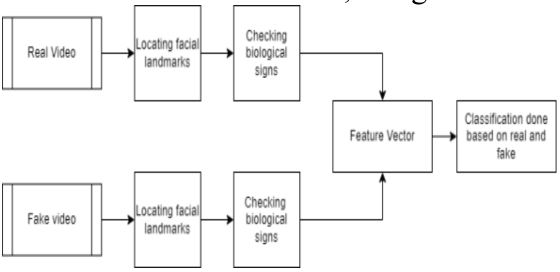

Real Video
Locating facial landmarks
Checking biological signs
Feature Vector
Classification done based on real and fake
Fake video
Locating facial landmarks
Checking biological signs

Fig 3: Video detection using facial artifact

PREPROCESSING

There could be extra pixels around the face in image or frame. The background of image or face could cause inconsistency in prediction. In order to make the model produce accurate results, the outliers are to be removed.

FEATURE EXTRACTION

Using body language analysis, DL model instructed in comprehension both temporal along with geographical data to handle fraudulent movies. The recommended approach will be an automatic framework which can determine the degree to which the provided posture change corresponds to the intended individual. The CNN is the most suitable option for capture spatial aspects of dataset in order to use the suggested strategy. It may also be used to understand graphical elements and the relationships of sequential data. As advanced CNNs, LSTMs are able to identify long-term relationships without gradient fading.

EXPERIMENTAL SETUP

The notion that body languages may detect and reveal deepfakes was tested by building a LSTM (many to one).

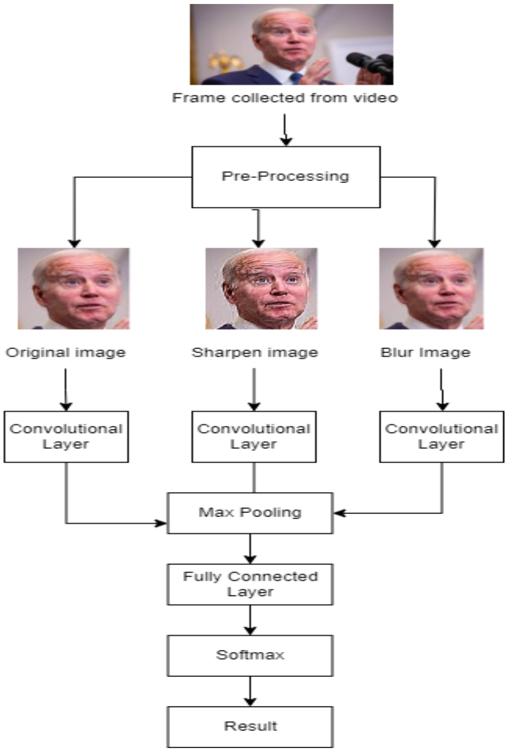


Fig 4: Exposing the frames to deep learning algorithms to find the result

GPU support and customizable DNN implementation are offered by PyTorch. Two different kinds of objects are required to set up both the training and the testing experiment: (a) a component for loading info, (b) an element for building a model. Ten percent was used for validation, ten percent for testing, and eighty percent was used for training the custom dataset. A specifically created Data Loader (based on the requirements of our unique dataset) was utilized and given to LSTM for training. The proposed model is designed to be a binary classification model with a layer of inputs consisting of body languages with 24 key points in the data. Additionally, it has a single output unit and a completely linked output layer.

EVALUATION METRIC

The effectiveness of the model is evaluated using metrics. Conversely, overfitting denotes a model with poor generalization performance, while underfitting denotes a model that fails to converge. It is possible that insufficient training data led to this situation. This project's produced dataset is limited, and a high degree of confidence might cause overfitting. Video editing employs frame-by-frame editing a framework wherein temporal aberrations with abnormalities between the frames are anticipated to further manifest as low-level face manipulation faults.

EFFICIENTNET – B0

Utilizing EfficientNet B0 as a feature extractor enables the extraction of significant representations from pictures and movies. A classifier can then be trained with these attributes to differentiate between real and deepfake video. EfficientNet B0 models that have already been trained can be improved using a dataset that includes real and deepfake photos and videos.

By fine-tuning the network's parameters to better fit the current detection goal, it increases the network's capacity to distinguish between actual and fraudulent content. Multiple instances of EfficientNet B0, each trained with different settings or on different subsets of data, can be combined using ensemble methods.

This method lowers the chance of overfitting and uses a variety of models to increase detection accuracy.

In adversarial training, the model is trained with deepfake content that has been produced both intentionally and unintentionally. By exposing the model to various forms of manipulation, it can learn to detect subtle inconsistencies indicative of deepfake content.

The robustness of the deepfake detection system can be increased by integrating EfficientNet B0 with other methods including metadata analysis, audio-visual synchronization checks, and optical flow analysis.

As a result of EfficientNet B0's computational efficiency, photos and videos uploaded can be quickly analyzed and used in real-time deepfake detection systems.

VI. OUTPUT SCREENSHOTS

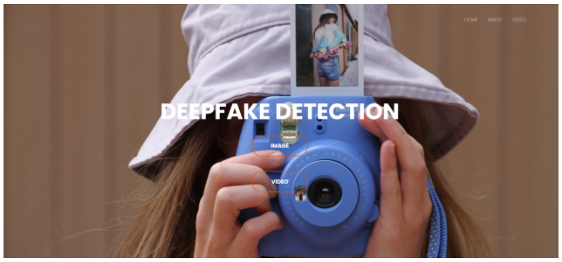
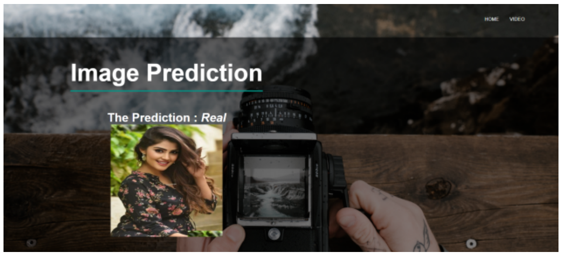
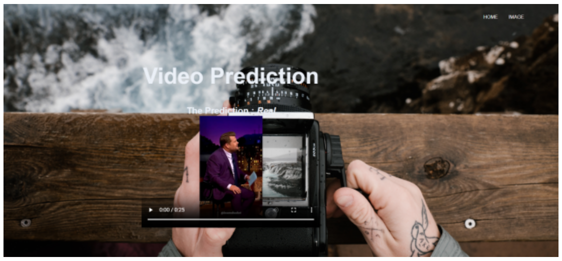


Fig 5: Home page with login and registration



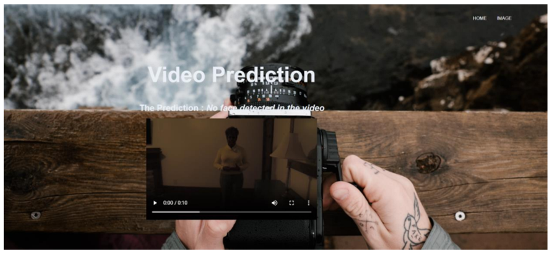
The Prediction : *Real*

Fig 6: Image Prediction



The Prediction : *Real*

Fig 7: Video Prediction



The Prediction : *No face detected in the video*

Fig 8: No face detected

VII . CONCLUSION

In summary, using Deep Learning techniques to detect the truthfulness of the data has priceless insights. Deepfake produces manipulated photos or videos that are indistinguishable from authentic ones. Two machine learning models exit generative adversarial networks, which are used to build deepfakes. While one model learns on a single dataset, a second model looks for deepfakes. Up until the other model is unable to recognize the fake, the forger produces fakes. Deepfakes are used to produce fictitious photos, videos, news stories, and terrorist incidents that can lead to financial and social fraud. It is having an increasing impact on democracy, security, culture, and religions as well as organizations, people, and communities. When the number of deepfake photos and videos on social media rises, people will stop believing the real thing. Therefore, cross-platform detection methods and deepfake datasets need to be created in the future. To identify deepfakes in commonly used mobile devices, effective, dependable, and durable mobile detectors are required. Additionally, by combining object detection and deepfake detection methods, would enhance deepfake detection. This project gives clarity to the users about the data.

REFERENCE

[1] AsadMalik, Minoru Kuribayashi, Sani M. Abdullahi, Ahmad Neyaz Khan, “DeepFake Detection for Human Face Images and Videos: A Survey”,2022

[2] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, ‘‘DeepFake detection: A systematic literature review,’’ IEEE Access, vol. 10, pp. 25494–25513, 2022.

[3] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, Siwei Lyu, “Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics”, in Handbook of Digital Face Manipulation and Detection. Cham, Switzerland: Springer, 2022, pp. 71–96.

[4] Honggu Liu, Xiaodan Li, Wenbo Zhou, Yuefeng Chen, Yuan He, Hui Xue, Weiming Zhang, Nenghai Yu, “Spatial-Phase Shallow Learning: Rethinking Face Forgery Detection in Frequency Domain”,2021 10.1109 / CVPR 46437.2021.00083

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ‘‘Generative adversarial nets,’’ in Proc. Adv. Neural Inf. Process. Syst., 2021, pp. 2672–2680.

[6] Dong Chen , Baining Guo, Ting Zhang, Fang Wen , Lingzhi Li, Jianmin Bao, Hao Yang, “Face X-ray for More General Face Forgery Detection”,2020, 10.1109/CVPR 42600.2020.00505.

[7] Sebastiano Battiato , Luca Guarnera ,Oliver Giudice[5], “Fighting Deepfake By Exposing TheConvolutional Traces On Images”,2020, 10.1109 / ACCESS .2020.3023037

[8] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, “Cnn generated images are surprisingly easy to spot... for now,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 8695–8704.

[9] U. A. Ciftci, I. Demir, and L. Yin, “Fakecatcher: Detection of synthetic portrait videos using biological signals,” IEEE transactions on pattern analysis and machine intelligence, 2020.

[10] S. Tripathy, J. Kannala, and E. Rahtu, ‘‘ICface: Interpretable and controllable face reenactment using GANs,’’ in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Mar. 2020, pp. 3385–3394.

[11] S. Fernandes, S. Raj, R. Ewetz, J. S. Pannu, S. K. Jha, E. Ortiz, I. Vintila, and M. Salter, “Detecting deepfake videos using attribution based confidence metric,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 308309.

[12] S. Wen, W. Liu, Y. Yang, T. Huang, and Z. Zeng, ‘‘Generating realistic videos from keyframes with concatenated GANs,’’ IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 8, pp. 2337–2348, Aug. 2019.