International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 04 | April - 2021 ISSN: 2582-3930

Research Paper on Deepfake Detection

SUBASH SHARMA
&
PRABHAKAR VARMA
KeraleeyaSamajam's Model College, Dombivli East, Mumbai, Maharashtra, India

ABSTRACT

Deepfake is a combination of fake and deeplearning technology. Deep learning is the function of artificialintelligence that can be used to create and detect deepfakes.
Deepfakes are created using generative adversarial networks, inwhich two machine learning models exit. One model trains on adataset and then creates the deepfakes, and the other model tries
to detect the deepfakes. The forger creates deepfakes until theother model can't detect the deepfakes. Deepfakes creating fake
videos, images, news, and terrorist events. When deepfake videos,and images increase on social media people will ignore to trustthe truth. Deepfakes are increasingly affecting individuals, communities, organizations, security, religions, and democracy.
This paper aims to investigate deepfake challenges, and to detectdeepfake videos by using eye blinking. Deepfake detections aremethods to detect real or deepfake images and videos on socialmedia. Deepfake detection techniques are needed original andfake images or video datasets to train the detection models. Inthis study, first discussed deepfake technology and its challenges,then identified available video datasets. Following, convolutionalneural networks to classify the eye states and long short-termmemory for sequence learning has been used. Furthermore, theeye aspect ratio was used to calculate the height and width ofopen and closed eyes and to detect the blinking intervals. Themodel trained on UADFV dataset to detect fake and real video byusing eye blinking and detects 18.4 eye blinks per minute on thereal videos and 4.28 eye blinks per minute on fake videos. Theoverall detection accuracy on real and fake videos was 93.23%and 98.30% respectively. In the future research and developmentneeds more scalable,accurate, reliable and cross-platformdeepfake detection techniques.

Keywords: Deepfake, deepfake detection, deep learning,detection techniques, eye blinking.

1. Introduction

Photos and videos are frequently used as evidence in policeinvestigations to resolve legal cases since they areconsidered to be reliable sources. However, sophisticated
technology increases the development of fake videos, andphotos that have potentially made these pieces of evidence
unreliable. Fake videos and images created by deepfaketechniques have been become a great public issue recently.
The authors in define the term deepfake as it is adeep learning-based method to create deepfake images orvideos by altering the face or full-body of a person in animage or video by the face or full-body of another person.
Deep learning is the arrangement ofalgorithms that canlearn the dataset and make intelligent decisions on their own.Generative Adversarial Networks (GANs)is the recentadvanced image and video manipulating tool to create highquality manipulated deepfake videos and images, and themedia increases the fastdistribution of these fake imagesand videos.

The GAN models were trained using a largenumber of images or videos, it can generate realistic faces orfull-body that can be seamlessly spliced into the originalvideo, and the generated video can lead to forgery of thesubject's identity in the video. Deepfake manipulationallows a user to replace the face or the full-body of a personin a video with the face or the full-body of another person,provided that enough images may be a large number ofimages are available of both persons; these videos are calleddeepfake videos. The authors instate that by usingthe merger of GANs and Convolutional NeuralNetworks (CNNs) can design quality deepfake that thedeepfakes notifying techniques can't detect them.

The existence of, open software mobile applicationsincreasing to everyone to generate fake videos and images. The smartphone availability, advancement of cameras,and social media popularity have made the editing, creation,anddissemination of images and videos more than ever.

This increases the tampering of videos and makes effectiveto propagate falsified information. To detect deepfakes,

various detection methods have been proposed afterdeepfakes were introduced. Deepfake detections aremethods to detect real and fake images or videos. Thedetection methods detect the deepfakes by eye blinking, eyeteach and facial texture, head poses, face warping artifacts,eye color, lip movements, audio speakers, reflections in theteeth, spatiotemporal features and capsule forensics.

In this study, investigate deepfakes, deepfake manipulationtools, available datasets, deepfake challenges, deepfake

detection challenges, and deepfake detection techniques,deepfake detection by using eye blinking. Finally, this studypresents eye blinking detection accuracy and overalldetection accuracy results.

2.1 Existing Systems

There has been an observabledevelopment in the field ofdeepfake video creation and detection in the commercialarena with the rise of software applications such as theFakeApp and the Face-Swap.

FaceApp is one such powerful face transformationapplication developed for usage in Android or IOS smartphones which is powered by Artificial Intelligence (AI)techniques of Neural Networks and Genetic Algorithms.

Itassists users to clickportrait of them having certain advancedfilters that act as invisible layers in the neural network with

input layer having the original image and the output layergenerating an edited image or photo.Fake-App is a desktopapplication program that allows ushiding certain features and modifying them in an image bymeans of some AI training method which later, can beoverlap with the photo of the face in video thus produceda deepfake digital video contented after classification of video

segments into individuals' image which have a watermark forits detection and recognition.

Certain web applications such as Face-SwapOnline alsoallow us to give our videos or images as inputs with AI and DL powered systems running in thebackground that allowus to modify the image content according to our wish through changing or applying filters and even changing resolution ordimensions of the images. Such web based applications use very powerful servers at datacenters all all over the world, in order to direct the beat traffic from user and also perform the images or videos manipulation in the backend.

Fig - 1: Depicts how different a processed image or video might get after using filters in the Face App desktop application.

2.1.1. Disadvantages

All of the above methodologies might require high-endspecifications of systems that some systems might only beable to work with. They also might require fast internet connection to work.

2.2. Proposed Solution

The solution or method we have put forward, focuses onthe usefulness of hashing and image processing facilities for
detection of deepfake video content. The work is majorlysplit into five modules which shows how the videos areaccepted at a web page, traverse through thelocalhost,processing of the videos will happen at the backend andresult is displayed over the webpage.

2.2.1. Advantages

The static web application we have developed works welleven with decent specifications of a system (such as in a
system with 4GB RAM and 32-bitOperating System).It even runs without internet connection as it is developed to worked on the Offline system.

3. Video Generation

For final deepfake video generation, autoencoders go withthe deep learning technique of CNN. As we know that deeplearning is a software mechanism that allows the computersystem to imitate the nervous system of humans byrecreating an artificial network of neuron functions just likethe network of neurons in our brain. This member ofMachine Learning (ML) is coined the term Deep Learning(DL) as it makes appropriate use deep neural networkmechanisms.

All the neural network related DL algorithms are createdvia interlinked layers which are input layer (first layer),output layer (last layer) and multiple hidden layers. Theselayers facilitate the automated learning concept of a DeepLearning software powered system without having previousinformation externally entered by the programmers.

The Convnets or Convolutional Neural Networks (CNNs)are the type of multilayered networks we are going to beusing here. The CNNs' structure is develop in such a methodthat it exactly recognizes an object's dimensions andfeature from a pictures or a videos. Hence CNNs are mostlyused over the unstructured data such as images or digitalvideos.

A CNN mainly completes generation process in 2 phaseswhere the first one implies the design of a non-continualsystematic model as output and the second phase is driventowards modifying the developed model using CNNmanipulation techniques.

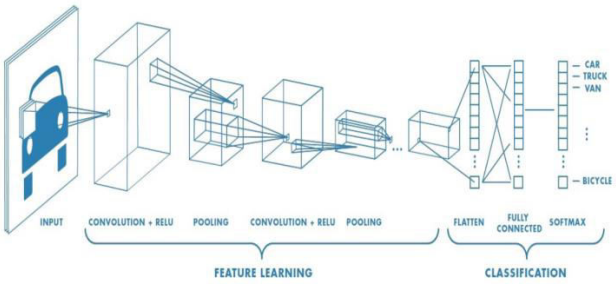

Fig -1: A CNN's structure hence can be depicted as in theabove figure.

The four major components of CNNs play an importantrole in the video generation process from each segment'simage modification.

The first component of convolution derives informationof the object in an image sequence and gets to know about various patterns specifically. The 2nd component of nonlinearity examine the huge concepts derived from complexity about the features of the images such as sharpness, edge modification and border recognize. Then the third component of pooling (also known as subsampling) provides a stage for the user to manipulate with the data presented originally in an image at a very deeper level with it's features.

The final component of classification then checks for the relation between information and if even after this level of processing, the image's

classification is very different from the original image data features, the image is yet again going to be processed using another set of layers of CNN.

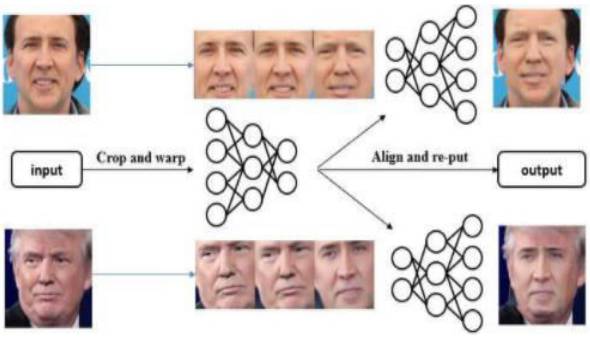

Fig -2: Here is how swapping of faces is done just by solving through all the possible face swaps being encoded as having same image information or features.

4. Deepfakevideoidentification

After studying about what are deepfake videos and going through the process of creating them, here comes the challenge of detecting them. Many technologies and techniques have come into picture after the deepfakes came into existence.

Just as the Deep Learning method of CNN is being used in the generation of deepfake videos, another Deep Learning. based technique of Recurrent Neural Networks (RNNs) can be used for identification or detection of deepfake videos.

The Deep Learning software and it's methods mainly use TensorFlow in order to put those concepts in practical usage through coding in python with it. But we are not going to use it here right now in our work.

Numerous programs and applications have alreadyproceed into locomotion using the RNN methods even though they are still undergoing rapid growth and development as lots of research is still under progressive conditions. The RNNs have facilitated the booming technology in present market which is, "AI in HR", i.e., Artificial Intelligence in Human Resource management.

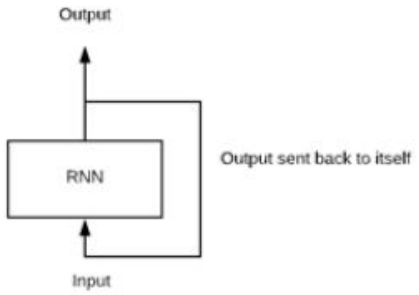

Fig -3: This shows how RNNs work with the concept of automation and improvisation.

5. Deepfake Challenges

Deepfakes are affecting the world since people around the world are using deepfakes for multiple reasons such as faces-wrapping, recreating pornographic videos with someone's face or body, and to create and disseminate fake news.

Deepfakes are more and more affecting democracy, privacy, security, religion and cultures of the people. Deepfakes are increasing from time to time, but there is no standard to evaluate deepfake detection techniques. The number of deepfake videos and images found online has nearly doubled since 2018. Massachusetts Institute of Technology (MIT) analyzed 126,000 news disseminated by 3,000,000 users for more than 10 years. Finally, they concluded that fake news spreads 1,500 people 6 times more rapidly than true news. Deepfakes creating fake news, images, videos, and terrorist events. Deepfake erodes people's trust in media and causes to social and financial fraud. Deepfake affects religions, organizations, politicians, artists, and voters. When deepfake videos and images increase in social media people will ignore to trust the truth.

The authors in analyze deepfakes that have the potential to harm individuals and societies. Using deepfakes to harm other people are yet to be largely seen including joke to embarrass a coworker, identity theft or even to spur violence, a porn video for someone's gratification and so on.

Also, deepfakes are used to fake terrorism events, blackmail, defame individuals, and to create political distress. Although nobody is safe from deepfakes, some people are more vulnerable than others. With minimum data and computing power, somebody can create a video the country leader saying something leading to civil conflict.
Deepfakes negatively affects targeted person, increase fake news and hate speech, create political tension, distress the public or create war. For example, a person can modify the contents of the video and people in a video to spread fake news, which may lead to war between nations; especially a country that contains diverse nations and nationalities.

6. Future scope for enhancement

We will be making use of the advanced technological concept of the "Blockchain" in the detection of deepfake videos.
Blockchain even though an emerging trend currently in the world of technology, is not new in application as it is just an integration of three main existing technological applications namely Peer-To-Peer Networks, Private Key Encryption and Software Programming. In a blockchain, the nodes in the network are interlinked having control over a decentralized database known as a "Ledger" which means that all the details of every transaction made by each of the node owner or user is made available to every other user who is a part of the chain.
Each individual block over a blockchain network is considered as a user and the block consists of user data or the transaction details, a unique hash for the current block, hash of the previous block and some program which enables the transaction of that user.
Even a slightest modification in the user data of a block changes the hash entirely and the whole blockchain gets disrupted.
This could be seen as an applicative enhancement for our project where we developed a hashing function for each video frame image (as a key). This hash, video frame data, the previous block hash and an action whenever the hash gets changed, forms a Blockchain for deepfake video detection purpose.
Etherium is one such most widely known and used Blockchain network whose range is all over the world where the software programs embedded into each block are known as Smart Contracts. This could be used in much more practical and simple manner in nearest future.

7. Conclusion

Deep learning can be used as a deepfake creation, anddetection methods. Deepfake creates forged images orvideos that persons cannot differentiate from real images orvideos. Deepfakes are created using generative adversarialnetworks, in which two machine learning models exit. Onemodel trains on a dataset and the other model tries to detectthe deepfakes. The forger creates fakes until the other modelcan't detect the forgery. Deepfakes creating fake news,videos, images, and terrorism events that can cause social,and financial fraud. It is increasing affects religions,organizations, individuals and communities', culture,security, and democracy. When deepfake videos and imagesincrease on social media people will ignore to trust the truth.In this study, the accessible datasets, deepfake creation software,deepfake challenges, fake video noticing techniques anddetect fake videos by using eye blinking were considered.
Also, the detection models trained on the datasets and thetotal and the eye-blink detection accuracy results werecomputed. Deepfake detection is a method to detect real andfake images or videos. In this study, the CNN to extractframe feature and to classify the eye states, and LSTM fortemporal sequence analysis have been used. Also, the eyeaspect ratio, used for eye blinking rate classification and theCNN and eye aspect ratio detect the eye blinking intervals.
The detection models have been trained on UADFVpublicly available real and fake videos. The deepfakedetection methods detect the deepfakes by eye blinking. Inthe examination, the eye blinking noticing precision result onreal videos is 91.59% and eye blinking noticing

precisionon fake videos 90.27%. Furthermore, the overall detectionaccuracy results on real videos is 93.23% and the overallnoticing accuracy on fake videos is 98.30%. In the eyeblinking detection, when the person moves his/her headquickly and when the eye focus on the area below them theeyelids cover the eye and the eye detected as blink or close
this affects the accuracy of the model.Now a day deepfake creation tools can create fake videos bymimic facial expressions of the person exactly so that it isbecome difficult to detect deepfakes by using facialexpressions like eye blinking, and lip-movement. Therefore,both image and video deepfake detection techniques areneeded performance improvement, evaluation standards, andparameters.
Future work will focus on evaluating different detectionmethods by using real and manipulated datasets. Due to theadvancement of technology full-body deepfakes arereleased. The continuous advancements of the face and full-body deepfakes development will be difficult to detect bythe existing detection techniques. So, deepfake datasets andcross-platform detection techniques need to be developed inthe future. Furthermore, due to the high computational cost,most detection techniques are unfit for mobile applications.
This needs efficient, reliable and robust mobile detectors todetect deepfakes in widely used mobile devices. Moreover,will improve deepfake detection by integrating deepfakedetection and object detection algorithms.

8. References

[1] https://en.wikipedia.org/wiki/Deepfake.
[2] https://www.guru99.com/deep-learning-tutorial.html.
[3] https://www.guru99.com/convnet-tensorflow-imageclassification.html.
[4] https://www.malavida.com/en/soft/fake-app/#gref.
[5] https://faceswaponline.com/.
[6] https://www.faceapp.com/.
[7] https://www.djangoproject.com/.
[8]https://docs.python.org/3/library /argparse.html.
[9] J. A. Marwan Albahar, Deepfakes Threats and CountermeasuresSystematic Review, JTAIT, vol. 97, no. 22, pp. 3242-3250, 30November 2019.
[10] D. D. Zahid Akhtar, A Comparative Evaluation of Local FeatureDescriptors for DeepFakes Detection, 07 November 2019.
[11] R. V.-R. J. F. A. M. J. O.-G. Ruben Tolosana, DeepFakes and BeyondA Survey of Face Manipulation and Fake Detection, pp. 1-15, 2020.
[12] G. Shao, What 'deepfakes' are and how they may be dangerous, 13October 2019. [Online]. Available:https://deepfakedetectionchallenge.ai/.
[13] B. P. W. W. J. D. Xinsheng Xuan, On the Generalization ofGenerative Adversarial Networks Image Forensics, pp. 1-8, 2019.
[14] S. M. Pavel Korshunov, Vulnerability assessment and detection ofDeepfake videos, pp. 1-6, 2019.
[15] A. M. R. Z. G. Marissa Koopman, Detection of Deepfake VideoManipulation, in Proceedings of the 20th Irish Machine Vision andImage Processing Conference, Belfast, 2018.
[16]M.A.R.S. Subash Sharma, Deep Learningfor Deepfakes Creation and Detection, pp. 1-16, 2021.
[17] R. H. B. P. N. B. C. C. F. Brian Dolhansky, The Deepfake Detection
Challenge (DFDC) Preview Dataset, Deepfake Detection Challenge,
pp. 1-4, 2019.
[18] M. V. R. S. Prabhakar Kumar, Detecting Face2Face Facial Reenactment
in Videos, pp. 1-9, 2020.