IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 9, September 2023

DOI: 10.17148/IJARCCE.2023.12916

Deepfake Detection Using Xception and Mobilenets Deep Learning Mod

D. Rupasri[1], M. Kumaran[2], J. Lin Eby Chandra[3]

Student M.E (CSE), Jaya Engineering College, Chennai, India[1]

Prof, Department of CSE, Jaya Engineering College, Chennai, India[2]

Prof, Department of CSE, Jaya Engineering College, Chennai, India[3]

Abstract: The project “Deepfake Detection Using Xception and Mobilenets Deep Learning Models” is a web-based application for identifying deepfake media contents i.e., image and video using deep learning technologies. Deepfake can be simply defined as “an image or video of a person in which their face or body has been digitally altered so that they appear to be someone else”. It is a controversial technology with many wide-reaching issues impacting society, e.g., election biasing. The existing system is based on cross-domain fusion, which works on the basis of traditional spatial domain features. This method had utilized the publicly deepfake datasets, and the results show that the method is effective particularly on the Meso-4 Deepfake Database. But this system is only capable of analysing the spatial features, so we propose a system that can process both image and video input and performs both spatial and depth-wise analysis over the input data. The deep learning models Xception and Mobile Net are the two approaches used for classification tasks to detect deepfakes. We utilize training and evaluation datasets from Face Forensics++ comprising four datasets, Face swap, Face2Face, Deepfake, Neural Texture generated using four different and popular deepfake technologies. The input is analysed for both spatial and depth features which is made possible through Xception and Mobile nets that uses depth wise convolutions. It is capable of detecting almost all kind of deepfakes since we train our model with dataset that contains the data obtained from popular deepfake creation.

Keywords: Deep learning, Web, Database, Texture.

I. INTRODUCTION

Artificial intelligence is the capability of a machine to imitate intelligent human behaviour. Machine learning (ML) is a branch of AI that gives computers the ability to “learn” often from data without being explicitly programmed. Deep learning is a subfield of ML that uses algorithms called artificial neural networks (ANNs), which are inspired by the structure and function of the brain and are capable of self-learning. ANNs are trained to “learn” models and patterns rather than being explicitly told how to solve a problem. The building block of an ANN is called the perceptron, which is an algorithm inspired by the biological neuron (5). Although the perceptron was invented in 1957, ANNs remained inobscurity until just recently because they require extensive training, and the amount of training to get useful results exceeded the computer power and data sizes available.

To appreciate the recent increase in computing power, consider that in 2012 the Google Brain project had to use a custom-made computer that consumed 600 kW of electricity and cost around $5,000,000. By 2014, Stanford AI Lab was getting more computing power by using three off-the-shelf graphics processing unit (GPU)-accelerated servers that each cost around $33,000 and consumed just 4 kW of electricity. Today, you can buy a specialized Neural Compute Stick that delivers more than 100 gigaflops of computing performance for $80 So, where did this all originate? Deep learning first gained popularity in academic circles as machine learning researchers looked to expand the scope of machine learning using larger datasets and more computation times.

However, deep learning was viewed with scepticism in the AI community for the longest time due to its shortcomings, particularly the lack of large enough datasets for training and lower available processing power in computers. All of that changed in 2012. Two researchers from the University of Toronto made history at ImageNet – an annual competition where contestants develop computer vision software based on large datasets, by developing an algorithm with an error rate of 15.3%. This was a huge improvement over the benchmark.

In the business community, Amazon had quietly been working on refining its item-to item collaborative filtering technology that relied on massive behavioural and catalogue datasets to deliver recommendations in real time. In 2014, Amazon tasked a team in the personalization group to design a new recommendation algorithm for Prime Video based

DOI: 10.17148/IJARCCE.2023.12916

on a neural network called an autoencoder. From a business standpoint, Amazon's recommendation engine Deep learning networks learn by discovering intricate structures in the data they experience. By building computational models that are composed of multiple processing layers, the networks can create multiple levels of abstraction to represent the data. For example, a deep learning model known as a convolutional neural network can be trained using large numbers (as in millions) of images, such as those containing cats.

This type of neural network typically learns from the pixels contained in the images it acquires. It can classify groups ofpixels that are representative of a cat's features, with groups of features such as claws, ears, and eyes indicating the presence A convolutional neural network (CNN) is a type of artificial neural network used primarily for image recognition and processing, due to its ability to recognize patterns in images. It is a powerful tool but requires millions of labelled data points for training, that is a large number of categorized data for its training.

II. LSTM TECHNOLOGY

Throughout this paper, we shall refer to the LSTM architecture (Long Short-Term Memory). By conceptualizing variables from the study, it was possible to better comprehend how the timestamp affects the value of cryptocurrency coin. Transactions involving cryptocurrencies are used to execute prediction simulations under various scenario-designed simulations.

The LSTM model was employed in this study despite the fact that time series prediction has been the subject of extensive research. This is because the LSTM model estimated the timestamp data based on the root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient (R), as well as the possibility that it may be connected to either a linear or nonlinear attribute, or possibly both.

Even while it is obvious that factors like the open price at the start of the time frame and the closing price at the end of the time window have an effect on the bitcoin transaction recognized by the timestamp, it is unclear if such impacts may be classified as patterns.

III. LITERATURE REVIEW

Liwei Deng, Hongfei Suo and Dong created the approach in 2021. In this work As technology advances and society evolves, deep learning is becoming easier to operate. Many unscrupulous people are using deep learning technology to create fake pictures and fake videos that seriously endanger the stability of the country and society. Examples include faking politicians to make inappropriate statements, using face-swapping technology to spread false information, and creating fake videos to obtain money. In view of this social problem, based on the original fake face detection system, this paper proposes using a new network of EfficientNet-V2 to distinguish the authenticity of pictures and videos. Moreover, our method was used to deal with two current mainstream large-scale fake face datasets, and EfficientNet-V2 highlighted the superior performance of the new network by comparing the existing detection network with the actual training and testing results. Finally, based on improving the accuracy of the detection system in distinguishing real and fake faces, the actual pictures and videos are detected, and an excellent visualization effect

Yujiang Lu, Yaju Liu, Jianwei Fei and Zhihua Xia done a progress at 2021. that is Recent progress in deep learning, in particular the generative models, makes it easier to synthesize sophisticated forged faces in videos, leading to severe threats on social media about personal privacy and reputation. It is therefore highly necessary to develop forensics approaches to distinguish those forged videos from the authentic. Existing works are absorbed in exploring frame level cues but insufficient in leveraging affluent temporal information. Although some approaches identify forgeries from theperspective of motion inconsistency, there is so far not a promising spatiotemporal feature fusion strategy. Towards this end, we propose the Channel-Wise Spatiotemporal Aggregation (CWSA) module to fuse deep features of continuous video frames without any recurrent units. 15 Our approach starts by cropping the face region with some background remained, which transforms the learning objective from manipulations to the difference between pristine and manipulated pixels. A deep convolutional neural network (CNN) with skip connections that are conducive to the preservation of detection-helpful low-level features is then utilized to extract frame level features. The CWSA module finally makes thereal or fake decision by aggregating deep features of the frame sequence. Evaluation against a list of large facial video manipulation benchmarks has illustrated its effectiveness. On all three datasets, Face Forensics++, Celeb-DF, and Deepfake Detection Challenge Preview, the proposed approach outperforms the state of-the-art methods with

Nan Wu, Xin Jin, Qian Jiang created the approach in 2022. That is with the continuous development of deep learning

DOI: 10.17148/IJARCCE.2023.12916

techniques, it is now easy for anyone to swap faces in videos. Researchers find that the abuse of these techniques threatens cyberspace security; thus, face forgery detection is a popular research topic. However, current detection methods do not fully use the semantic features of deepfake videos. Most previous work has only divided the semantic features, the importance of which may be unequal, by experimental experience. To solve this problem, we propose a new framework, which is the multisegmented pathway network (MSPNN) for fake face detection. .is method comprehensively captures forged information from the dimensions of microscopic, mesoscopic, and macroscopic features. These three kinds of semantic information are given learnable weights. The artifacts of deepfake images are more difficult to observe in a compressed video. Therefore, pre-processing is proposed to detect low-quality deepfake videos, including multi-scale detail enhancement and channel information screening based on the compression principle. Centre loss and cross-entropy loss are combined to further reduce intra-class spacing. Experimental results show that MSPNN is superior to contrast methods, especially low- quality deepfake

In 2020 Aarti Karandikar is also worked on Improving With the advent of new technological enhancements in artificial intelligence, new sophisticated AI techniques are used to create fake videos. Such videos can pose a great threat to the society in various social and political ways and can be used for malicious purposes. These fake videos are called deepfakes. Deepfakes refer to manipulated videos, or other digital representations produced by sophisticated artificial intelligence, that yield fabricated images and sounds that appear to be real. A deep-learning system can produce a persuasive counterfeit by studying photographs and videos of a target person from multiple angles, and then mimicking its behaviour and speech patterns. Detecting these videos is a massive problem because of the increasing developments in more realistic deepfake creation technologies emerging every now and then. The paper aims to solve this problem by proposing a model that analyses the frames of the videos using deep learning approach to detect inconsistencies in facial features, compression rate and discrepancies introduced in the videos while creating them. The model uses a convolutional neural network along with transfer learning to train the model that can catch these instilled errors in the deepfakes. The neural network is trained on these discrepancies induced during deepfake creation around the face. It uses a dataset called “Celeb- DF: A New Dataset for Deepfake Forensics” to train the model. The paper further discusses methods that can be used, in detail, to improve learning by this model.

IV. SYSTEM DESIGN

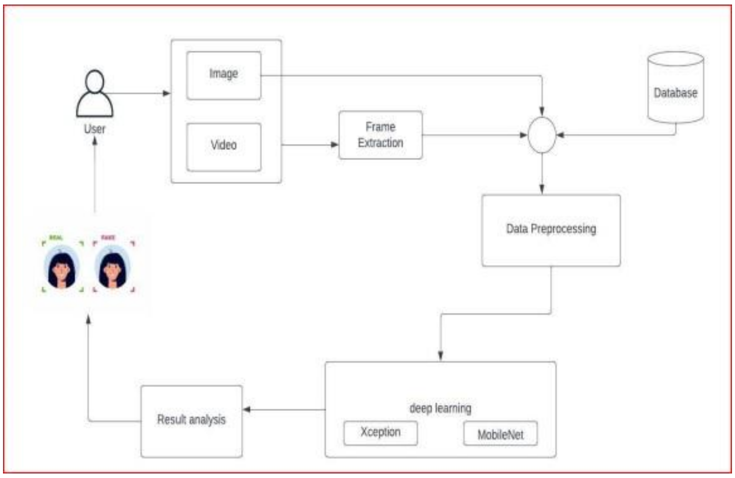


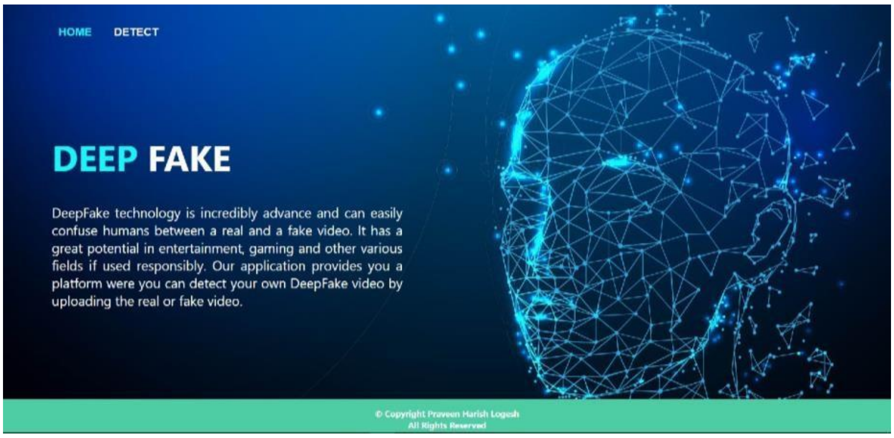
Fig 1: System Architecture

In this article we proposed a method that deals with deep learning technology for the detection of fake content (image and video). The aim of our project is to provide an application that can identify deepfakes by using deep learning algorithms. The proposed system “Deepfake Detection Using Xception and Mobile Nets Deep Learning Models” is designed to identify the deepfake image and video content.

Here we use depth wise separable deep convolutional neural networks Xception and Mobile nets that deals with both spatial (width and height) and depth dimensions. For training and evaluation, we consider dataset from Face Forensics++ comprising four datasets generated using four different and popular deepfake technologies (Deepfake, Face2Face, Face Swap, Neural Textures).

The video files converted into frames and the face area in the image/frames is the focus, and the pre-processing module needs to capture the face in the image as model input.

DOI: 10.17148/IJARCCE.2023.12916

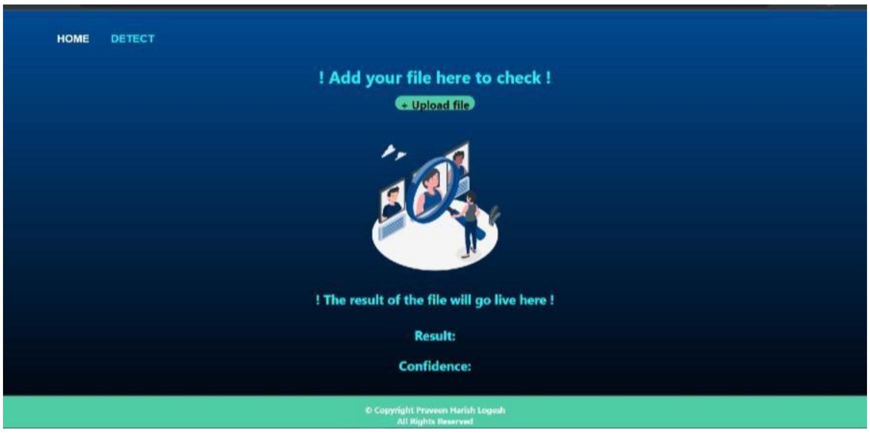


DEEP FAKE

DeepFake technology is incredibly advance and can easily confuse humans between a real and a fake video. It has a great potential in entertainment, gaming and other various fields if used responsibly. Our application provides you a platform were you can detect your own DeepFake video by uploading the real or fake video.

© Copyright Praveen Harish Logesh
All Rights Reserved

Fig 2: Home Page



! Add your file here to check !

! The result of the file will go live here !

Result:

Confidence:

© Copyright Praveen Harish Logesh
All Rights Reserved

Fig 3: Classification Page

The pre-processing module itself consists of three steps as intercepting frames from the video, detecting faces from these individual frames, and saving face areas as images and they are sent for classification The system architecture explains the detailed workflow of the application. The application is capable of classifying both the image and video files so the users who tend to use this application has to choose the type of input they are going to enter. If the input data is in the form of image, then they are directly sent for data pre- processing and the followed by the deep learning models. But if the input is in form of videos, then there will be an additional process before the data pre-processing called the frame extraction, where the frames that suit with the face are captured and then sent for data pre-processing. Then they enter the depth-wise convolution deep learning models where the actual classification process occurs. The models provide a result that contains percentage of likeliness of “real vs fake”, from those values final results over the input data is produced

V. CONCLUSION

Our goal of creating a system that can be used to Detect Deepfake using data has been achieved. The version of this template is V2. Most of the formatting instructions in this document have been compiled by Causal Productions from the IEEE LaTeX style files. Causal Productions offers both A4 templates and US Letter templates for LaTeX and Microsoft Word. The LaTeX templates depend on the official IEEEtran.cls and IEEEtran.bst files, whereas the Microsoft Word templates are self-contained.

Our prediction models' accuracy would rise as a result. Other deep learning models could also be researched to see what accuracy level they produce. Additionally, using this in mobile applications will facilitate quick data access for end users.

DOI: 10.17148/IJARCCE.2023.12916

REFERENCES

[1]. Liwei Deng , Hongfei Suo and Dongjie Li, "Deepfake Video Detection Based on EfficientNet-V2 Network", Hindawi Computational Intelligence and Neuroscience, Volume 2022, Article ID 3441549, April 2022.
[2]. Yujiang Lu, Yaju Liu ,Jianwei Fei and Zhihua Xia, "Channel-Wise Spatiotemporal Aggregation Technology for Face Video Forensics", Hindawi Security and Communication Networks, Volume 2021, Article ID 5524930, August 2021.
[3]. Nan Wu, Xin Jin, Qian Jiang, "Multisemantic Path Neural Network for Deepfake Detection", Hindawi Securityand Communication Networks, Volume 2022, Article ID 4976848, October 2022.
[4]. Aarti Karandikar, "Deepfake Video Detection Using Convolutional Neural Network", International Journal of Advanced Trends in Computer Science and Engineering, Volume 9 No.2, March -April 2020
[5]. Abhijit Jadhav, Abhishek Patange, "Deepfake Video Detection using Neural Networks", IJSRD - International Journal for Scientific Research & Development| Vol. 8, Issue 1, 2020 | ISSN (online): 2321-0613.
[6]. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprintarXiv:1704.04861.
[7]. Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2face: Real-time face captureand reenactment of rgb videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2387-2395).
[8]. Thies, J., Zollhöfer, M., & Nießner, M. (2019). Deferred neural rendering: Image synthesis using neural textures. ACM Transactions on Graphics (TOG), 38(4), 1-12.
[9]. Gardiner, N. (2019). Facial re-enactment, speech synthesis and the rise of the Deepfake.