INTERNATIONAL JOURNAL OF ADVANCED INNOVATIVE TECHNOLOGY IN ENGINEERING

Published by Global Advanced Research Publication House Journal

Home page: www.ijaite.co.in

# DeepFake Detection Using Inception-ResNet-v2

[1]Dr. A. S. Manekar, [2]Anand Wankhade, [3]Aman Surkar, [4]Pranjali Pande, [5]Kunal Nemade

[1,2,3,4,5]Department of Information Technology, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, Maharashtra, India

[1]asmssgmcoe@gmail.com, [2]anandwankhade0007@gmail.com, [3]assurkar@gmail.com, [4]pranjalipande25@gmail.com, [5]kunalnemDe06@gmail.com

**Article History**

Received on: 10 April 2022

Revised on: 25 April 2022

Accepted on: 29 May 2022

**Keywords:** Deepfake, Inception-Resnet-v2, CNN, Generator Adversarial Networks (GANs)

**e-ISSN:** 2455-6491

**Production and hosted by**
www.garph.org



## ABSTRACT

Control of pictures, recordings, and sounds utilizing face alter applications and web administrations have been being used, for quite a long time yet ongoing progress in deep learning have prompted AI-produced counterfeit pictures and recordings with traded faces, lip-adjusted sounds prominently known as Deepfakes. Deepfakes are created primarily using one of the following two methodologies: Autoencoders and Generator Adversarial Networks, both of which are based on pretrained deep neural networks. The level of authenticity accomplished by deep learning controlled deepfakes increments with expanding measures of information i.e., counterfeit pictures and recordings promptly accessible on the web at removal to prepare GANs. Deepfake algorithms make media leaving an uncovered edge of distinction between the true or unique source and the manufactured or deepfake objects. In this way, new systems and methods to distinguish through such deepfakes are the need of great importance. The methodology proposed based upon strong deep learning-based CNN designs to be specific, Inception-Resnet-V2 for recognizing the deepfakes. Our proposed approach not just surpasses the current methodologies regarding effectiveness and precision yet additionally offers the best concerning the given existence intricacy.

## 1. INTRODUCTION

The terminology of deepfakes comes from "deep-learning" in addition to "fakes''. A general term covers counterfeit recordings, pictures, sound, and different media incorporated utilizing AI-controlled deep learning procedures. Deepfakes ("deep learning" and "fake") are artificially generated media wherein an individual in a recorded video or image is replaced to look like someone else in a ditto way. Generally used Deep Learning methods to create deepfakes entail the use of generative neural network designs for training, such as Auto-encoders [21] or GANs (Generative Adversarial Networks) [20]. The source individual consequently imitates the objective individual and does activities or gives talks which can prompt deception publicity during political races influencing the fair appointive cycle, hampering the social picture of conspicuous characters or big-name slander and fake news. In 2018, an extremely brief video where previous US President Barack Obama is seen conveying a misleading message that won't ever say was released! [22]-[23]. It is plainly obvious that deepfakes can prompt a protected emergency, common and military agitation, cause strict and socio-political pressures between

fighting groups and nations and is a powerful threat to protection, security, and public respectability. This requires the steadily expanding requirements for confirming the respectability and legitimacy of computerized content and especially the visual ones.

The other side of deepfakes provides an amazingly useful application considering re-filming successions of motion picture films without a trace of the entertainer as occurred in the Fast and Furious series. Deepfakes were utilized to convey an extremely serious level of photorealism in the scenes. Deepfakes may also be used to provide sound to performers who are experiencing character voice confusion. For professionals, deepfakes can provide actual sound visuals with images of someone else lip-synchronized with the voice of someone else [24]. It is important to draw out a contrast between happy integrated utilizing picture control apparatuses like Adobe Photoshop 10.0 and Artificial Intelligence (AI) incorporated deepfakes. The groundworks of deepfakes lay on deep learning networks prepared and tried on significant measures of fake and genuine face pictures and recordings information to normally plan the facial elements, looks, and other face antiquities between the progenitor and the objective.

The limit of deep learning methods to deal with complicated and tremendous volumes of information is taken advantage of in deepfakes age. Many information tests of fake pictures and recordings increment the photograph authenticity accomplished in the result deepfake. The famous Obama deepfake was delivered from a GAN which utilized 56 hours (around 2 and a half long periods) of test input recordings to duplicate the specific lip, head, and eye relics in the face. On account of picture alter applications like photoshop restricted facial controls are permitted attributable to the need for muddled altering instruments and space capability required. Making photo-realistic trades utilizing such applications is a complex and tedious interaction.

In the beginning, a deepfake video could have been effectively-recognized through natural eyes inferable from the peculiarity of pixel breakdown which leads to uncanny visual antiquities in the face, skin, and so forth, and goal irregularity in pictures and others. In any case, the new improvements in deep-learning-network advances and the free accessibility of gigantic measures of information produces deep fakes that can't be separated utilizing either direct human perception abilities or refined computer calculations. An example of Deepfaked Faces has been shown below in figure 1.


Figure 1: Real and Deepfaked Faces

## 2. RELATED WORK

Deep learning is slowly becoming part of everyone's lives without them even noticing. Its wide range of applications in both good and bad areas such as cinematography, politics, social media, AI, etc. is a major factor in easing our lives. It's been used to solve a variety of difficulties from large data to human controls, there is a slew of issues to deal with. controls. As amazing as this technology is and its application is, every coin has two sides and hence deepfakes application on the negative side is becoming more and more prominent by the clock. It's becoming a threat to national security, causing a major threat to security, and ultimately instigating people to disturb and cause societal harassment. Hence, it's imperative that we must develop software that is able to distinguish whether the content in front of us is fake or not. This is essentially what we are targeting to achieve and hence have referred to various research papers. Deep learning has been successfully used to address a variety of complicated issues ranging from big data analysis to computer vision to human-level control. Deep learning progress anyway has additionally been utilized to make programming that can pose threat to security, a majority rules system may malfunction and raise questions on public safety. Deepfake is now one of those deep learning applications. Deepfakes are created using techniques that lapse face photos of a target human onto a video of a source individual in order to create a video of the target individual performing or making the same statements as the source individual. Deepfake computations can create counterfeit images and recordings that are difficult to distinguish from genuine ones. As a result, breakthroughs that can naturally recognize and assess the legitimacy of sophisticated media that is visual are vital. This paper discusses a study of calculations used to make deepfakes and, all the more critically, techniques proposed to recognize deepfakes to this date. Manikin deepfakes incorporate recordings of a targeted individual (manikin) who is enlivened following the looks, eye, and head developments of another individual sitting before a camera. While some deepfakes can be made by the conventional Generative Adversarial Model [20], the new

normal basic instrument for the development of deepfakes in deep learning models such as autoencoders and generative ill-disposed networks, which have been widely used in the computer vision field. We discuss the problems, research trends, and directions associated with deepfake advancements. This study provides an extensive description of deepfake techniques and collaboration on the development of new and more powerful ways to handle the inexorably testing of deep fakes by examining the basis of deepfakes and cutting-edge deepfake discovery tactics [1].

With late advances in computer vision and illustrations, it is presently conceivable to create recordings with very reasonable manufactured faces, even progressively. Innumerable applications are conceivable, some of which raise a genuine caution, calling for dependable finders of fake recordings. Recognizing unique and controlled video can be a, as a matter-of-fact challenge for people and computers the same, particularly when the recordings are packed or have a low goal, as it frequently occurs in informal communities. Research on the discovery of face controls has been genuinely hampered by the absence of sufficient datasets. To this end, we present an original face control dataset of about a portion of 1,000,000 altered pictures (from over 1000 recordings). The controls have been produced with a cutting-edge face-altering approach. It surpasses all current video control datasets by basically a significant degree. Utilizing our new dataset, we present benchmarks for old-style picture legal errands, counting grouping, and division, taking into account recordings compacted at different quality levels. Likewise, we present a benchmark assessment for making undefined frauds with known ground truth; for example, with generative refinement models. [2]

Deepfake is a strategy for human picture amalgamation in light of fake insight. Deepfake is utilized to combine and superimpose existing pictures and recordings onto source pictures or recordings utilizing AI techniques. Deep Fake location through Mobile Net, what's more, Xception looking fake recordings that can't be recognized by unaided eyes. They can be utilized to spread disdain addresses, make political trouble, extort somebody, and so forth. At present, Cryptographic marking of recordings from their source is done to actually take a look at the genuineness of recordings. Hashing of a video document into fingerprints (little line of text) is done and reconfirmed with the test video and accordingly confirmed regardless of whether the video is the one initially recorded. In any case, the issue with this procedure is that the fingerprints and hashing calculations are not accessible to commoners. In this paper, the proposed framework follows a recognition approach of Deepfake recordings utilizing Neural Networks. Double order of deepfakes was done utilizing a mix of Dense and Convolutional neural network layers. It was noticed that 91% precision was acquired in Adam and 88% was obtained in SGD (stochastic inclination drop) for straight out cross-entropy. In double cross-entropy, 90% exactness was found in Adam also, 86% exactness was seen in SGD while, 86% precision in Adam and 80% precision in SGD was gotten in mean square.[3]

The quick advancement in technology has now gotten to a place where it raises critical worries for the ramifications of society. At best, this prompts a deficiency of confidence in computerized content, yet might actually hurt by spreading misleading data or fake news. The authenticity of best-in-class picture controls is investigated in this research, as well as why it is so difficult to discern them, either naturally or by people. To normalize the assessment of recognition strategies, they have proposed a mechanized standard for face control location. Specifically, the standard depends on Face2Face [12], Face Swap [9]-[10]-[11], and Neural Textures [17]. The standard or benchmark is openly accessible and contains a test set as well as a data set of approximately 1.7 million controlled pictures. This dataset is over a request for a size bigger than tantamount, freely accessible, fraud datasets. In light of this information, we played out a careful examination of information-driven imitation locators. We show that the utilization of extra domain-specific information further develops falsification identification to uncommon exactness, indeed, even within the sight serious areas of strength for of, and obviously beats human eyewitnesses.[4]

The research describes an Inception module translation in convolutional neural networks that is akin to a middle ground between standard convolution and layers of distinct convolution activity (a layered convolution followed by a pointwise convolution). In this light, a layer's distinct convolution may be thought of as an Inception module with the most pinnacles possible. As a result of this perspective, we suggest a deep convolutional neural network engineering inspired by Inception, in which Inception modules are replaced with layers of different convolutions. We show that this architecture, Xception [17], outperforms Inception V3 [19] on the ImageNet dataset (for which Inception V3 was designed) and outperforms Inception V3 on a larger picture grouping dataset with 350 million images and 17,000 classes. Because the Xception engineering has the same number of boundaries as Inception

V3, the exhibition advantages are due to more efficient usage of model boundaries rather than an increased limit [5].

The detection of deepfake images is not only restricted to detection via ImageNet but there are several new methodologies too. One of such is detection using "photoplethysmography" [6]-[14]-[15]. The term photoplethysmography is made from 'photo' which means light and 'plethysmography' which is used to calculate changes in the volume of a body. This technique uses the change in blood flow in the target person's face via camera input and compares it with the real data of blood flow which it was already trained upon. It Chrom-PPG (PCA or ICA can also be used [16]) to process the signal and create a map out of it. Then this map is given as input to a CNN model for training.

## 3. PROPOSED METHODOLOGY

Dataset: The DFDC (Deepfake Detection Challenge) [7] is a Preview dataset that was used in our research. In comparison to previous datasets that include video counterfeiting, we found DFDC to be highly varied in gender, age, skin tone, and color of the persons in the videos. Participants could choose any background of their choice to film videos and incorporate a variety of head postures, poses, and situations, as well as visually different backgrounds. The preview dataset contains nearly 5,000 movies made using two separate facial alteration algorithms which include approximately 1,100 real and 4,100 fraudulent videos.

Pre-processing: The videos used in the dataset are pre-processed. Image frames are extracted from it and saved to a new folder. To capture a video, we first need to make a Video-Capture object in Python using the Open-CV module. The frame rate is then set to 0 and extracted up to the frame needed. We extract the image frames and save them to a new folder after renaming it on the frame-Id and finally establishing a dataset of image frames extracted from the videos. Furthermore, these frames are fed into a program that employs MTCNN [6] (a face classification neural network) to classify the faces. These classified faces are cropped and saved into a new folder. Extracting frames from the video, recognizing faces from these individual frames, and storing face regions as pictures are the three processes of the pre-processing module (as illustrated in Figure 2).

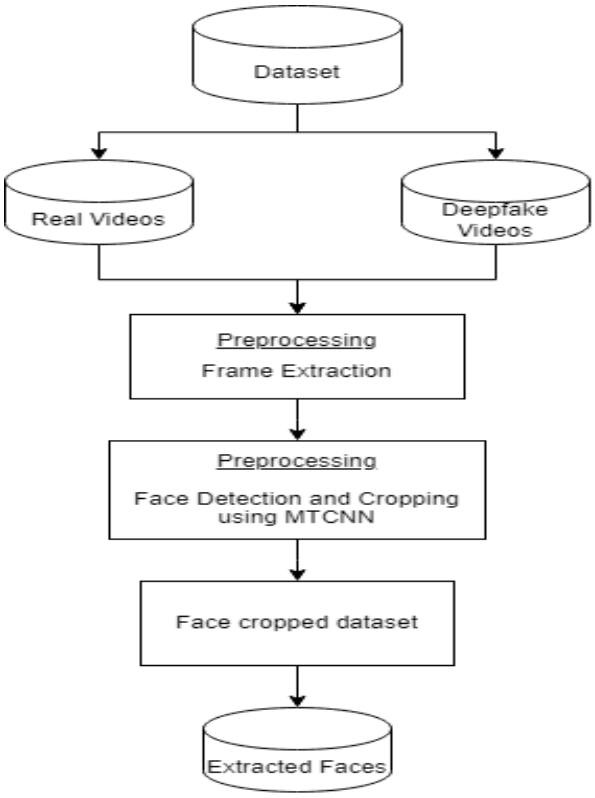


Figure 2: Pre-processing of dataset (Schematic diagram face detection)

### A. *Inception-Resnet-V2*

Inception and Resnet models play had a critical impact in image recognition progress as of late, with exhibited great execution at nearly low computational expenses. Inception- ResNet-v2 design is made using the Inception Algorithm architecture, with the residual relations. Inception-Resnet-V2 is a CNN model that depends on the group of Inception designs, with residual relations. The architecture has 164 layers, pretrained on pictures of the enormous ImageNet data set. The overall architecture and compressed view of the Inception-Resnet-v2 network have been depicted in Fig. 3.2.

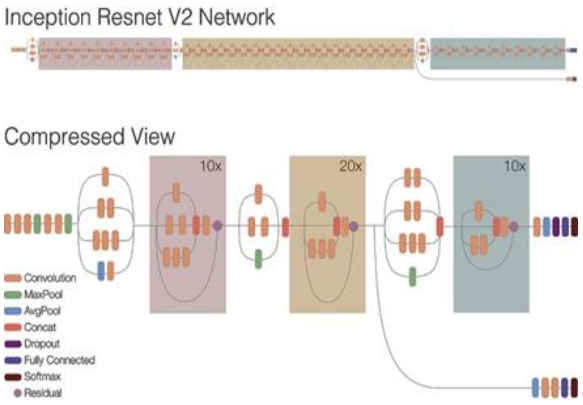


Figure 3: Architecture of Inception-Resnet-V2 (Google-AI-Blog)

### *B. Transfer Learning*

This paper proposes the use of a transfer learning model instead of basic CNN as they are more accurate. We performed transfer learning on the Inception-Resnet-v2 network or simply added a model trained on one task upon another trained model, which would optimize and allow rapid progress when modeling the second task. We added a flattened layer and a dense layer that predicts the output for two classes – deepfake or real. The network has been trained upon input images (faces) of shape (229, 229, 3), pre-processed from the set of dataset videos. The model was trained for 20 epochs, which yielded a good accuracy.

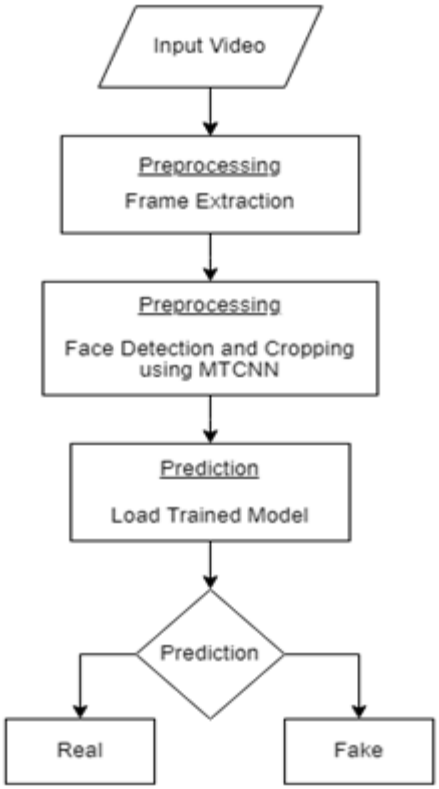


Figure 4: Workflow of DeepFake Prediction

For the prediction, a GUI was designed which takes a video file as input and passes it for pre-processing. The pre-processed images are stored in a new folder created during the workflow. These pre-processed images are given to a trained model which predicts whether the images are real or not. The workflow is depicted in Figure 4.

## 4. RESULT ANALYSIS

The results are based upon the implementations of Python as a programming language, OpenCV [13] for image processing, MTCNN for face detection, Keras for implementation of neural networks, NumPy for scientific calculations, and Matplotlib for graphical analysis. The Graphic card used for computation is Nvidia GeForce GTX 1650.

Before training a model, the parameters specified for training the model such as selection of loss functions, optimizer, and activation functions can have a huge impact on producing faster and optimum results. These different components of a neural network thus play a very crucial role in effectively training a model and that too in an efficient way in order to produce more accurate results. For instance, SGD (stochastic gradient descent) worked much faster than other algorithms but the results weren't optimized at all. However, Adam (adaptive moment estimation) produces much better results than SGD but is computationally expensive (It's still preferred for practical purposes).

Here, using Transfer Learning with Inception Resnet V2, we compared the testing and validation accuracy of different models made up of different activation functions such as ReLU, Leaky ReLU, and Sigmoid on different numbers of epochs. The results of the above experiment are tabulated in table 1.

Table 1: Comparing accuracy obtained from different activation functions over a different number of epochs

| Activation Function | Accuracy on 5 epochs | Accuracy on 10 epochs | Accuracy on 15 epochs | Accuracy on 20 epochs |
|---|---|---|---|---|
| Sigmoid | 67.37 | 69.78 | 73.8 | 77.45 |
| ReLU | 67.9 | 70.23 | 74.7 | 78.20 |
| Leaky ReLU | 69.1 | 72.36 | 76.28 | 80.12 |

After using Transfer Learning through Inception Resnet V2, our results show an increase in accuracy ranging from 67.37 % to 80.12 % with increasing epochs along with different optimizers and loss functions.

## 5. CONCLUSION

This paper presented a neural network approach to classify whether the given video is a deep fake video or not. We propose the implementation of Transfer Learning through the Inception-Resnet-V2 model. We experimentally found out that different parameters such as optimizers, loss functions, activation functions, and a number of epochs, before training a model indeed have an effect on accuracy predictions. Using transfer learning and applying different activation functions, we found out that Leaky ReLU gives even better results for the identification of deepfakes and thus is more suitable and reliable.

Furthermore, created software that is successfully able to produce desired results with high accuracy.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.


## FUNDING SUPPORT

The author declares that they have no funding support for this study.


## REFERENCES


[1] Deep Learning for Deepfakes Creation and Detection: A Survey Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Cuong M. Nguyen, Dung Nguyen, Duc Thanh Nguyen, Saeid Nahavandi, Fellow, IEEE.

[2] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M. (2018). FaceForensics: A large-scale video dataset for forgery detection in human faces. arXiv preprint arXiv:1803.09179, 24th March 2018.

[3] Anuj Badale, Chaitanya Darekar, Lionel Castelino, Joanne Gomes, "Deepfake Video Detection Using Neural Network ", INTERNATIONAL JOURNAL OF ENGINEERING rESEARCH & TECHNOLOGY (IJERT) IJERTCONV9IS03075 22-02- 2021.

[4] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies,Matthias Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images" in arXiv:1901.08971.

[5] Francois Chollet Google, Inc., "Xception: Deep Learning with Depth Wise Separable Convolutions", 2017 IEEE Conference on Computer Vision and Pattern Recognition 4 Apr 2017.

[6] U. A. Ciftci, I. Demir, and L. Yin. FakeCatcher: Detection of synthetic portrait videos using biological signals. IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI), 2020.

[7] Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks - Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Senior Member, IEEE, and Yu Qiao, Senior Member, IEEE.

[8] The DeepFake Detection Challenge (DFDC) Dataset - Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, Cristian Canton Ferrer.

[9] Deepfakes. https://github:com/deepfakes/faceswap. Accessed: 2020-03-16.

[10] Faceswap. https://github:com/MarekKowalski/FaceSwap. Accessed: 2020-03-16.

[11] Faceswap-gan. https://github:com/shaoanlu/faceswap-GAN. Accessed: 2020-03-16.

[12] TY - BOOK AU - Thies, Justus AU - Zollhöfer, Michael AU - Stamminger, Marc AU - Theobalt, Christian AU - Nießner, Matthias PY - 2020/07/29 SP - T1 - Face2Face: Real-time Face Capture and Reenactment of RGB Videos ER.

[13] G. Bradski. The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.

[14] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Algorithmic principles of remote ppg. IEEE Transactions on Biomedical Engineering, 64(7):1479–1491, 2017.

[15] W. Wang, S. Stuijk, and G. de Haan. Living-skin classification via remote-ppg. IEEE Transactions on Biomedical Engineering, 64(12):2781–2792, 2017.

[16] Robust pulse-rate from chrominance-based rPPG Gerard de Haan and Vincent Jeanne.

[17] Xception: Deep Learning with Depthwise Separable Convolutions Franc¸ois Chollet

[18] J. Thies, M. Zollh¨ofer, and M. Nießner. Deferred neural rendering: Image synthesis using neural textures. ACM Trans. Graph., 38(4), July 2019.

[19] Rethinking the Inception Architecture for Computer Vision Christian Szegedy Google Inc. szegedy@google.com Vincent Vanhoucke vanhoucke@google.com Sergey Ioffe sioffe@google.com Jonathon Shlens shlens@google.com Zbigniew Wojna University College London zbigniewwojna@gmail.com.

[20] "Deep Fakes" using Generative Adversarial Networks (GAN) Tianxiang Shen UCSD La Jolla, USA tis038@eng.ucsd.edu Ruixian Liu UCSD La Jolla, USA rul188@eng.ucsd.edu Ju Bai UCSD La Jolla, USA jub010@eng.ucsd.edu Zheng Li UCSD La Jolla, USA zhl153@eng.ucsd.edu

[21] OC-FakeDect: Classifying Deepfakes Using One-class Variational Autoencoder Hasam Khalid Computer Science and Engineering Department Sungkyunkwan University, South Korea hasam.khalid@g.skku.edu Simon S. Woo Computer Science and Engineering Department Sungkyunkwan University, South Korea swoo@g.skku.edu

[22] Protecting World Leaders Against Deep Fakes Shruti Agarwal and Hany Farid University of California, Berkeley Berkeley CA, USA {shruti agarwal, hfarid}@berkeley.edu Yuming Gu, Mingming He, Koki Nagano, and Hao Li University of Southern California Los Angeles CA, USA {ygu, he}@ict.usc.edu, koki.nagano0219@gmail.com, hao@hao-li.com

[23] https://ars.electronica.art/center/en/obama-deep-fake/ How Deep Are the Fakes? Focusing on Audio Deepfake: A Survey ZAHRA KHANJANI, GABRIELLE WATSON, and VANDANA P. JANEJA, University of Maryland Baltimore County, Information System department, USA