MAJOR-1 PROJECT

Final Report on

DEEPFAKE VIDEO DETECTION

Submitted By:

Name	Roll No	Branch
Shahzeb Rizvi	R110216143	B.TECH CSE CCVT 4th YEAR
Satyam Gupta	R110216138	B.TECH CSE CCVT 4th YEAR
Satyansh	R110216140	B.TECH CSE CCVT 4th YEAR

Under the guidance of

Mr. Alind

Assistant Professor- Senior Scale

Department of Computer Science

School of Computer Science

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

Dehradun248007

2019-20

Approved By

(Mr. Alind)
Project Guide

(Dr. Deepshikha Bhargava)
Department Head

INDEX

FINAL Report (2019-20)

1. Project Title

DEEP-FAKE VIDEO DETECTION

2. Abstract

Deep-fake is a technique used for human image synthesis. To understand it simply, it is a technique used to map the face of another person onto the face of the person in the image or video. It is used to combine and superimpose existing images and videos onto source images or videos using the generative adversarial network(GANs). It is an amazing technology but it can also be used to harm someone. It is used in making fake pornographic videos of celebrities or revenge porn. Deep-fakes can also be used to create fake news and hoaxes. In this project we are going to propose and make a deep-fake detection system with the help of Convolutional Neural Networks (CNNs) self-trained using manually created data-sets, data-sets from Kaggle and various face-recognition data-sets. Data-set created by us are in general created by using gaussian blurring and affine wrapping of one's frontal face features to another face to simulate the deepfake images. DeepFake images can be made up to a limited size which results in resolution issues and blurring of the created image and in the general affine wrapping is used to map the features to the target image which can be detected due to its unique artifact. This model will help us to detect the presence of wrapping, illumination changes, resolution blur, etc. which are key features of a deepfake image.

Keywords: Deep-fakes, Deep Learning, CNNs, GANs.

3. Introduction:

Deep-fake is a portmanteau of deep learning and fake ("deep" and "fake"). The phrase "deep-fake" was coined in 2017.

The research related to deep-fakes lies predominantly within the field of computer vision. Computer vision is a subfield of computer science and related to artificial intelligence that focuses on digital image and video processing. The term deep-fakes originated around the end of 2017 from a Reddit user named "deep-fakes". He, shared deep-fakes they created of many celebrities involved with their bodies swapped onto actresses in pornographic videos, while non-pornographic content included many videos with actor Nicolas Cage's face swapped into various movies. In February 2018, deep-fakes was banned by Reddit for sharing involuntary pornography [4], including other social media platform like Twitter. In 2017, "Synthesizing Obama" program published modified video footage of former president Barack Obama. In that video the audio was synthesized to sound like him and facial expressions were re-enacted and synthesized to the face of Barack Obama [5]. Those facial expressions were done in real-time using a camera that does not capture depth, which makes it look realistic.

Convolutional auto-encoders [2] and generative adversarial network models can provide convincing deep-fakes [16], which are trained and run upon computers that are only accessible to highly-trained professional. Nowadays, it is very easy to make deep-fakes with help of some simple mobile apps like FaceApp and FakeApp. Snapchat is also a platform that uses many human face image syntheses to make age variant faces of the source image faces.

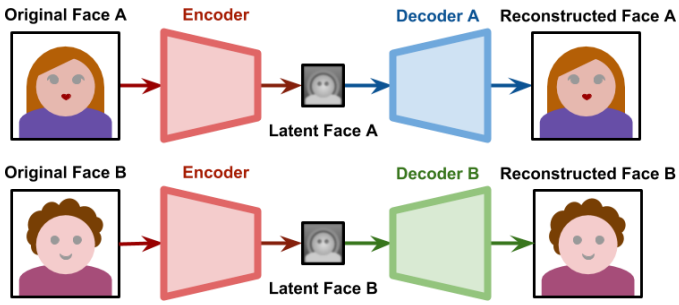


Figure 1: Different decoder is generated for different faces due to their features. Encoders are first use to make latent face and then decoder is applied on it for reconstruction. For DeepFake Person A's face is encoded and the decoded by Person B's face decoder.

Deep neural networks trained on face images are used to map facial expressions of the source image to the destination image. Source image is the image from which the features are extracted from and the destination image is the image from which the key facial points are extracted to map the features from source image to give a high sense of realism, lighting and illumination is also checked between the two images and manipulated accordingly.

In this paper, we would exploit key techniques used in deepfake to distinguishing it from the real image. Affine wrapping is one of the key techniques exploited in our project to detect deepfakes

but our training dataset also contains some cases of illumination changes, blurring, head posture, etc. for providing better results to detect DeepFake.

Convolutional Neural Network (CNN) model which we are going to build ourselves is going to train on the dataset and provide us results and insight on how much deepfakes can we would be able to detect with affine wrapping detection. We are creating training dataset for deepfake images by first two frontal images as source and destination. The source image is first scaled down to a random size with some additional gaussian blur and the features of the face is extracted. Facial landmarks are detected from the faces to create a transformmatrix to wrap the features to the destination image. Applying real deepfake algorithm would take a lot more time than our way of creating simulations of them.

In the end we would see the training results and demonstration of deepfake video detection of videos taken from various websites.

4. Literature Review:

Digital Media Forensics: The field of digital media forensics plans to create advancements for the computerized evaluation of the integrity of a picture or video. Both feature-based and CNN-based trustworthiness examination strategies have been investigated in the literature. For video-based digital forensics, most of the proposed arrangements attempt to identify computationally shabby controls, for example, dropped or copied edges or duplicate move controls. Procedures that recognize face-based controls incorporate strategies that recognize PC created faces from normal ones, for example, Conotter et al.[6] or on the other hand Rahmouni et al. [7]. In biometry, Raghavendra et al. [8] as of late proposed to recognize transformed countenances with two pertained profound CNNs and Zhou et al. [9] proposed discovery of two diverse face-swapping controls utilizing a two-stream system. Of unique enthusiasm to experts is another dataset by Rossler et al. [10], which has about a large portion of a million altered pictures that have been created with highlight-based face altering.

Face-based Video Manipulation Methods: A few face picture blend procedures utilizing profound learning have likewise been investigated as studied by Lu et al. [11]. Generative adversarial networks (GANs) [1] are utilized for maturing adjustments to faces, or to change face traits such as skin shading. The profound component introduction indicates noteworthy outcomes in modifying face qualities, for example, age, facial hair and mouth demeanours. Comparable aftereffects of characteristic interjections are accomplished by Lample et al. [12]. The vast majority of these profound learning-based picture union systems endure from low picture goals. Karras et al. [13] show the high-quality union of countenances, improving the picture quality utilizing dynamic GANs.

Recurrent Neural Networks: Long-ShortTerm Memory (LSTM) networks are a specific sort of Recurrent Neural Network (RNN), first acquainted by Hochreiter and Schmid Huber [14] with adapt long-term conditions in information arrangements. At the point when a profound learning design is outfitted with an LSTM joined with a CNN, it is ordinarily considered as "somewhere

down in space" and "somewhere down in time" individually, which can be viewed as two unmistakable framework modalities [15]. CNN's have made monstrous progress in visual acknowledgement undertakings, while LSTMs are generally utilized for long succession handling issues. In light of the innate properties (rich visual portrayal, long-term transient memory and end-to-end preparing) of a convolutional LSTM design, it has been altogether read for other PC vision assignments including successions (for example movement acknowledgement or human re-distinguishing proof in recordings) and has led to critical upgrades.

Deepfake Video Detection Using Recurrent Neural Networks: This article was written by David et al. [15] and published by IEEE in February 2019. Their system uses a convolutional neural network (CNN) to extract frame-level features. These features are then used to train a recurrent neural network (RNN) that learns to classify if a video has been subject to manipulation or not. They trained their systems on large datasets of videos and got an accuracy greater than 97% overall on the deepfake video detection.

Exposing DeepFake Videos by Detecting Face Wrapping Artifacts: This article was written by Y. Li et al. [17] and published in May 2019. Their system uses convolutional neural networks (CNNs) to expose deep-fakes with the help of self-generated negative images of the original image as the dataset. They trained their system with original image and a negative image which wasan image manipulation using image morphing methods like gaussian blur and face wrapping of the original source image. They used various pre-trained CNN models and fine-tuned them. VGG, ResNet50, ResNet101and ResNet152 models were compare on the basis of accuracy. They achieved an accuracy of 97% and above in all deepfake cases.

5. Problem Statement:

Deepfakes has become a common feature that can be used on a smart-phone by any person.So, it is essential to detect manipulations in videos and images to categorize fake and original source to reduce confusion and stop spreading of involuntary information. We also have to come up with an effective method to detect deep-fakes.

6 Objectives:

The objective of this project is:

- Propose a deep-fake detection method.
- We would be using Convolution Neural Network (CNN) model for fake video detection.
- Understanding the pipeline of DeepFake video creation and exploiting the techniques used in it.
- The result would help us to detect deep-fakes using model trained on exploited techniques like affine wrapping, illumination changes, head pose, resolution blur, etc.

7 Methodology:

We will be following the steps given below to expose deep-fakes. The steps are:

a. Collection and Creation of Dataset.

b. Preprocessing of the data collected.

c. The model created and its architecture.

d. Training and testing of the proposed model.

e. Prediction of DeepFake Videos.

a. Collection and Creation of Dataset:

The data set of fake and real images is used to trained our CNN model.80% off the fake images are created by and 20% are from Kaggle. The simulation of deep fake is created by the following steps:

1. The source and destination image are selected.

2. Source image is blurred using gaussian blur of kernel (5,5).

3. The dlib library is used to extract face from the image.

4. 3D wrap image is created using scikit-spatial. Delaunay on destination feature points to find out the Delaunay points that would be used to make the affine matrices.

5. Bilinear interpolation is used to find the resultant warp image by using the source image face and finding the outer codes of the grid and Delaunay points of the both images.

6. Mask is created to blend the warp created to the destination image to make it look real.

7. Colors are corrected of the warp image.

8. Shrinking of the mask is done using erosion of kernel (10,10).

9. Output image is created using warped source face, destination face and mask created.

We have a dataset of 3125 images with a portion of 80-20 for training and testing sets for model respectively.

b. Preprocessing of the data collected:

1. The images are resized to the shape of 299x299 pixels size to train with a batch size of 10.

2. Keras library is used for the preprocessing method which us to create the training set and test set ready for model training.

c. The model used and their architecture:

The model we have created is an CNN model with following architecture and parameters:

1. 11 layers are used to create the deepfake detection system.

2. First 4 layers are convolution and max pooling layers with necessary parameters and relu activation function is used:

layer 1: Conv layer with 64 feature extractors of (4,4)

layer 2: Max pooling layer with pooling size (4,4)

layer 3: Conv layer with 32 feature extractors of (3,3)

layer 4: Max pooling layer with pooling size (3,3)

layer 5: Conv layer with 16 feature extractors of (3,3)

layer 6: Max pooling layer with pooling size (3,3)

3. Next layers are flattening layers and 3 fully connected layers with sigmoid activation function in the end.

4. The 3 fully connected layers have 40,10,1 unit respectively.

d. Training and testing of the proposed model:

Adam is used as the optimizer; Binary cross entropy is used for loss function and accuracy matrix in the compiling process of the model.

e. Prediction of deepfake videos:

Prediction and validation would be done using various deepfake videos collected from anonymous websites.

Steps taken to find if video is fake or real:

1. Loop the video per frame.

2. Predict the frame if it is real or not.

3. Count the number of fake predictions.

4. Calculate percentage of fake predictions.

8. Result:

Observed accuracy of our model after training was 84.26% after 3 epoch.

```
H:\Study Material\UPES\4th year\Major 1\code>python training_model_code.py
Using TensorFlow backend.
2019-12-09 16:28:27.200861: I tensorflow/core/platform/cpu_feature_guard.cc:142] Your CPU supports instructions that thi
s TensorFlow binary was not compiled to use: AVX2
Found 2501 images belonging to 2 classes.
Found 624 images belonging to 2 classes.
Epoch 1/3
3000/3000 [==============================] - 2398s 799ms/step - loss: 0.5350 - accuracy: 0.7011 - val_loss: 0.7761 - val
_accuracy: 0.7773
Epoch 2/3
3000/3000 [==============================] - 2309s 770ms/step - loss: 0.2904 - accuracy: 0.8697 - val_loss: 1.1189 - val
_accuracy: 0.7985
Epoch 3/3
3000/3000 [==============================] - 2348s 783ms/step - loss: 0.2091 - accuracy: 0.9086 - val_loss: 0.7375 - val
_accuracy: 0.8426
Saved model to disk
```

Figure 2: Result after training model.

We were able to validate our model by doing prediction on videos collected from various video uploading websites. Our model was able to detect 50% or more presence of affine warp in deepfake videos.

```
H:\Study Material\UPES\4th year\Major 1\code>python predict.py --vid "fake_videos/fake/_ (1).mp4"
Using TensorFlow backend.
2019-12-12 09:28:43.294722: I tensorflow/core/platform/cpu_feature_guard.cc:142] Your CPU supports instructions that thi
s TensorFlow binary was not compiled to use: AVX2
Total fake frames counted = 249
Total frames = 292

Percentage of fake frames counted = 85.27397260273972
```

Figure 3: Validation of model with fake videos collected from websites.

9. Conclusion:

After observing the results of our project:

- We were able to create a deep-fake detection model by finding the presence of affine warp in a given image.
- We got an accuracy of 84% during training our model.
- We were able to detect fake frames in a video and detect if it is fake or not.

Advancements that can be done:

- Create a more complex model that would need better system performance for training.
- Creating a model with better accuracy.
- Model should be able to detect deepfakes by using different exploiting techniques like head posture, resolution of video, lighting on the face, etc.

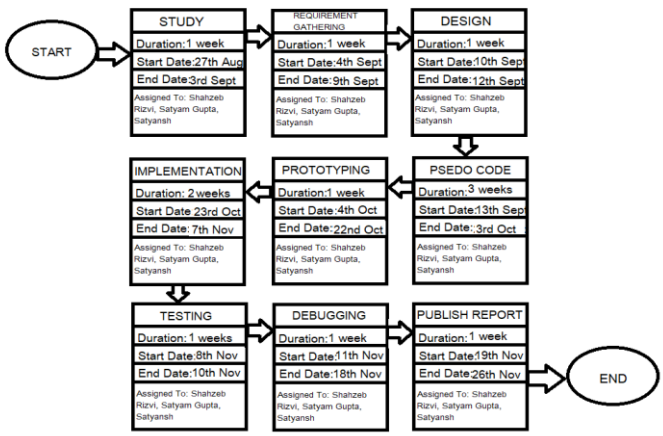
10. System Requirements (Software/Hardware):

- Hardware Interface:
  - 64 bits processor architecture supported by windows.
  - Minimum RAM requirement for proper functioning is 4 GB.

- Required input as well as output devices.
- Sufficient Graphic card for image processing.

- Software Interface:
  - This system will be developed in the Python 3.6.8 programming language.
  - Python IDLE.
  - Keras, OpenCV, Pillow, TensorFlow, dlib, NumPy, Scikit-learn and additional pip libraries for image processing.

11. Schedule (PERT Chart):



12. References:

[1] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face ageing with conditional generative adversarial networks. arXiv:1702.01983, Feb. 2017.

[2] A. Tewari et al. Mofa: Model-based deep convolutional face auto-encoder for unsupervised monocular reconstruction. Proceedings of the IEEE International Conference on Computer Vision Workshops, pages 1274-1283, Oct. 2017. Venice, Italy.

[3] J. Thies et al. Face2Face: Real-time face capture and reenactment of RGB videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pages 2387–2395, June 2016. Las Vegas, NV.

[4] What are deepfakes& why the future of porn is terrifying? https://www.highsnobiety.com/p/what-are-deepfakes-ai-porn/. (Accessed on08/27/2019).

[5] The Outline: Experts fear face-swapping tech could start an international showdown. https://theoutline.com/post/3179/deepfake-videos-are-freaking-experts-out?zd=1&zi=hbmf4svs. (Accessed on 08/27/2019).

[6] V. Conotter, E. Bodnari, G. Boato, and H. Farid. Physiologically-based detection of computer generated faces in the video. Proceedings of the IEEE International Conference on Image Processing pages 248–252, Oct. 2014. Paris, France.

[7] N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen. Distinguishing computer graphics from natural images using convolution neural networks. Proceedings of the IEEE Workshop on Information Forensics and Security pages 1–6, Dec. 2017. Rennes, France.

[8] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch. Transferable deep-CNN features for detecting digital and print-scanned morphed face images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops pages 1822–1830, July 2017. Honolulu, HI.

[9] P. Zhou et al. Two-stream neural networks for tampered face detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops pages 1831–1839, July 2017. Honolulu, HI.

[10] A. R¨ossler et al. Faceforensics: A large-scale video dataset for forgery detection in human faces. arXiv:1803.09179, Mar. 2018.

[11] Z. Lu, Z. Li, J. Cao, R. He, and Z. Sun. Recent progress of face image synthesis. arXiv:1706.04717, June 2017.

[12] G. Lample et al. Fader networks: Manipulating images by sliding attributes. Advances in Neural Information Processing Systems pages 5967–5976, Dec. 2017. Long Beach, CA.

[13] T. Karras, T. Aila, S. Laine, and J. Lehtinen. The progressive growing of gans for improved quality, stability, and variation. arXiv:1710.10196, Oct. 2017.

[14] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735–1780, Nov. 1997.

[15] David Guetta and Edward J. Delp. Deepfake Video Detection Using Recurrent Neural Networks. Video and Image Processing Laboratory (VIPER), Purdue University. Publisher: IEEE. Date of Conference: 27-30 Nov. 2018. Date Added to IEEE Xplore: 14 February 2019.

[16] I. Goodfellow et al. Generative adversarial nets. Adavances in Neural Information Processing Systems, pages 2672-2680, Dec. 2014. Montréal, Canada.

[17] Yuezun Li and SiweiLyu. Exposing Deepfake Videos By Detecting Face Wraping Artifacts. Computer Science Department University at Albany, State University of New York. Publisher: IEEE. Date of Conference: 12-17 May 2019. Date Added to IEEE Xplore: 17 April 2019.