Deepfake Video Detection: A Time-Distributed Approach

Baran Deniz Korkmaz

August 2020

This paper includes the introduction of some of the deepfake creation and detection techniques selected by the authors. Afterwards, the authors propose their own technique.

1 Introduction

DeepFake Creation

1. One technique is used to substitute the face of a targeted person by the face of the source person in a video. Initially, this technique was released by a Reddit user to create manipulated pornographic videos of famous actors. Later, it was developed as a user-friendly application so that it could be accessed by anyone easily (aka FaceSwap, FaceApp).

2. DeepFake uses the concept of generative adversarial net- works (GANs), in which two deep learning models com- pete. One model gets trained on real data and tries to create forgeries; meanwhile, the other strives to detect the forgery. The forger keeps on creating better and better fakes until the other model is unable to detect the forgery. To gener- ate DeepFake, one has to cumulate the aligned faces of two discrete individuals X and Y , then auto-encoders E_x and E_y are trained to regenerate the faces from the dataset of the images of X and Y, respectively, as shown in **Figure 1**. The magic lies in sharing the weights of the encoding part of the two auto-encoders, but their decoders are separated. After the training, any image comprising a face of X can be encoded through this shared encoder but decoded with the decoder of E_y .

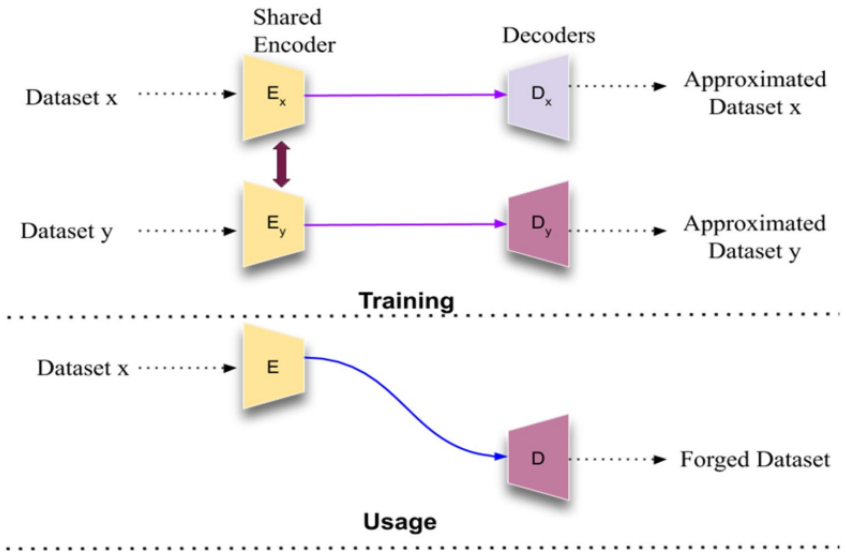

Figure 1: A DeepFake creation model using two encoder–decoder pairs

In this paper, first and foremost, all the related works formerly present have been discussed.

2 Related Work

In this section, various state-of-the-art face forging creation and detection techniques have been discussed by the authors articulately.

Some Early Methods in Deepfake Creation

1. Dale et al. [9] introduced a technique to manipulate faces in a video with the help of a 3D multilinear model, in which they warp the source to the target face and retime the source so that the target performance is matched.

2. Garrido et al. [10] used a novel image matching metric that merges the appearance and motion to choose candidate frames from the source video, and the user's identity is preserved as the face transfer uses a 2D warping technique.

3. Face2Face technique by Thies et al. [11] in which face re-enactment is performed by taking dense photometric consistency measures while tracking the facial expressions of the source as well as targeted video.

Modern Methods in Deepfake Creation

1. Modern methods leverage the use of deep learning techniques, especially GANs, to create forget videos. (By Goodfellow et al. [13])

2. FaceSwap [14] is another technique that uses CNN to catch the appearance of the target entity from an amorphous collection of photos.

Remarkable Datasets

1. Rössler et al. [15] intro- duced a novel face manipulation dataset having nearly half a million edited images (from over 1000 videos)

2. Rössler et al. introduced FaceForensics++ [16], which is an extension of the previ- ously introduced FaceForensic dataset.

3. The Deep Fake Detection Challenge dataset [17, 18] has been made publicly available, which consists of 1,19,146 videos

Deepfake Detection Techniques

1. Marra et al. [19] showed that each GAN leaves its specific fingerprints in the images it generates, this can play an important role while detecting forgeries.

2. Mittal et al. [20] introduced an algorithm which uses multiple threshold approach (B-edge) for efficient edge detection.

3. Afchar et al. [21] presented two networks, both having a low number of layers so that they could focus on the mesoscopic properties of images, which made their for- gery detection technique fast and reliable

4. Image segmen- tation has been found useful for various purposes [22–24].

5. In addition, specific techniques utilized in the medical domain and watermarking [25–27] could be used as a ref- erence for forgery detection.

6. Nguyen et al. [28] introduced a cap- sule network that could detect various kinds of attacks with very few parameters than the traditional CNN with similar performance. Later, it became crucial to localize the manip- ulated regions and detect manipulated face images.

7. Attention mechanism was proposed by Dang et al. [29] to improve the feature maps and to highlight the informative region to further improve the binary classification

Using CNNs for object localization/detection has been a customary practice. More often than not, these networks are felicitous in feature extraction from images. Some of the most widely used state-of-the-art CNNs are XceptionNet [31], InceptionV3 [32], SeNet [33], and ResNet [34] which could be used as a feature-extractor/backbone network.

3 Proposed Method

The authors proposed an architecture which leverages the use of spatio-temporal features to detect the DeepFake videos.

There is no hard and fast rule about the size of the input image as such. One can use a larger input size but at the expense of more computational power.

DeepFake detection problem has been framed as a binary classification problem using a CNN model wrapped in a time-distributed layer followed by a long short-term memory (LSTM) layer whose output is fed into dense layers as shown in **Figure 2**. The authors aim to detect forged videos which may get unnoticed by the naked human eye but the model being spatio-temporal will be able to capture those minuscule details.

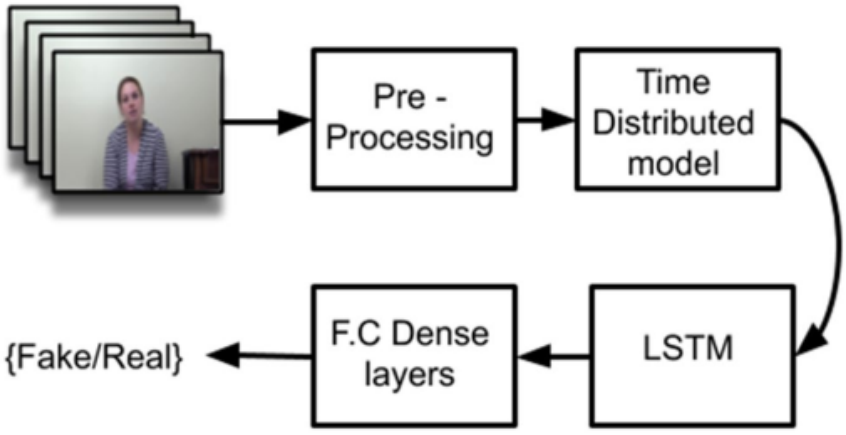

Figure 2: Basic structure of the proposed method

4 Dataset Pre-Processing

The entire DFDC dataset (about 470 GB of data) has been used for the experimentation.

DeepFake techniques manipulate mostly the face and regions around the face. Therefore, the face as the region of interest has been considered and extracted from the video as shown in **Figure 3**. Video manipulation has been car- ried out on a frame-by-frame basis by retrieving the faces so that low-level artifacts produced by face manipulation further manifest themselves as temporal artifacts with inconsistencies across frames.

Due to the humongous amount of data, there was a need for a face extractor which was fast as well as accurate. MobileNet- SSD [36, 37] provided a righteous trade-off between the two. A 35% extra margin around the faces has been added so that the distortions in that region could be detected.

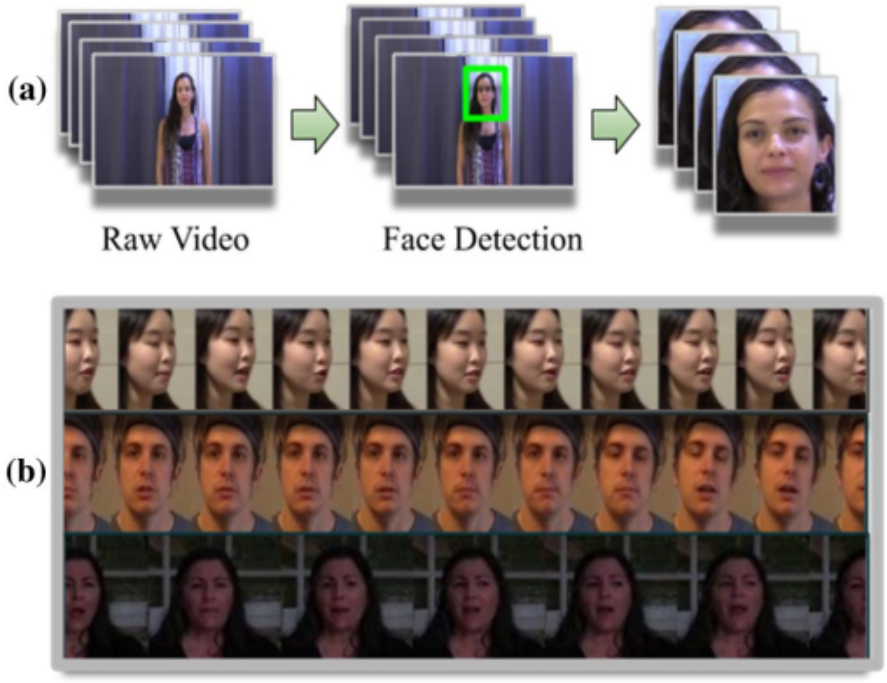

Figure 3: a Pre-processing pipeline, b example faces in DFDC dataset

5 Source

- DeepFake Video Detection: A Time-Distributed Approach; Amritpal Singh, Amanpreet Singh Saimbhi, Navjot Singh, Mamta Mittal

6 Referenced Sources

DeepFake Creation

9 - Dale K, Sunkavalli K, Johnson MK, Vlasic D, Matusik W, Pfister H. Video face replacement. In: Proceedings of the 2011 SIGGRAPH Asia conference. 2011. p. 1–10

10 - Garrido P, Valgaerts L, Rehmsen O, Thormahlen T, Perez P, Theobalt C. Automatic face reenactment. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 4217–24.

- **Face Enactment** - [11] - Thies J, Zollhofer M, Stamminger M, Theobalt C, Nießner M. Face- 2face: Real-time face capture and reenactment of rgb

videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 2387–95.

Modern Deepfake Creation Techniques

- **GANs** - [13] - Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Advances in neural information processing systems. 2014. p. 2672–80.

- **FaceSwap** - [14] - Korshunova I, Shi W, Dambre J, Theis L. Fast face-swap using con- volutional neural networks. In: Proceedings of the IEEE interna- tional conference on computer vision. 2017. p. 3677–85.

Remarkable Datasets

15 - Rössler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M. Faceforensics: a large-scale video dataset for forgery detection in human faces. arXiv preprint. arXiv:1803.09179 (2018).

16 - Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M. Faceforensics++: learning to detect manipulated facial images. In: Proceedings of the IEEE international conference on computer vision. 2019. p. 1–11.

17 - `https://www.kaggle.com/c/deepfake-detection-challenge/data`

18 - Dolhansky B, Howes R, Pflaum B, Baram N, Ferrer CC. The Deep- fake Detection Challenge (DFDC) preview dataset. arXiv preprint. arXiv:1910.08854 (2019).

Deepfake Detection Techniques

- **GAN Fingerprints in Images** - [19] - Marra F, Gragnaniello D, Verdoliva L, Poggi G. Do gans leave artifi- cial fingerprints? In: 2019 IEEE conference on multimedia informa- tion processing and retrieval (MIPR). IEEE; 2019. p. 506–11.

20 - Mittal M, Verma A, Kaur I, Kaur B, Sharma M, Goyal LM, Roy S, Kim TH. An efficient edge detection approach to provide better edge connectivity for image analysis. IEEE Access. 2019;13(7):33240–55

21 - Afchar D, Nozick V, Yamagishi J, Echizen I. Mesonet: a compact facial video forgery detection network. In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE; 2018. p. 1–7.

22 - Yu CM, Chang CT, Ti YW. Detecting Deepfake-forged contents with separable convolutional neural network and image segmenta- tion. arXiv preprint. arXiv:1912.12184 (2019)

- **Image Segmentation** - [23] - Mittal M, Goyal LM, Kaur S, Kaur I, Verma A, Hemanth DJ. Deep learning based enhanced tumor segmentation approach for MR brain images. Appl Soft Comput. 2019;1(78):346–54

- **Image Segmentation** - [24] - Mittal M, Arora M, Pandey T, Goyal LM. Image segmentation using deep learning techniques in medical images. In: Advancement of machine intelligence in interactive medical image analysis. Singa- pore: Springer; 2020. p. 41–63

- **Image Segmentation** - [25] - Mittal A, Kumar D, Mittal M, Saba T, Abunadi I, Rehman A, Roy S. Detecting pneumonia using convolutions and dynamic capsule routing for chest X-ray images. Sensors. 2020;20(4):1068

- **Watermarks** - [26] - Goyal LM, Mittal M, Kaushik R, Verma A, Kaur I, Roy S, Kim T-H. Improved ECG watermarking technique using curvelet transform. Sensors. 2020;20:2941

- **Watermarks** - [27] - Mittal M, Kaushik R, Verma A, Kaur I, Goyal LM, Roy S, Kim TH. Image watermarking in curvelet domain using edge surface blocks. Symmetry. 2020;12(5):822.

- **Capsule Networks** - [28] - Nguyen HH, Yamagishi J, Echizen I. Capsule-forensics: using capsule networks to detect forged images and videos. In: ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; p. 2307–11

- **Attention Mechanism** - [29] - Stehouwer J, Dang H, Liu F, Liu X, Jain A. On the detection of digital face manipulation. arXiv preprint. arXiv:1910.01717 (2019)