International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

DeepFake Detection Using Eye Blinking

Ruksa Gazi[1], Needhi Kore[2], Raj Jani[3], Manjyot Singh[4], Deepti Pawar[5]

[1,2,3,4]*Student, 4th Year Degree, Computer Engineering Department, Shah & Anchor Kutchhi Engineering College, Mumbai, India.*
[5]*Assistant Professor, Computer Engineering Department, Shah & Anchor Kutchhi Engineering College, Mumbai, India.*

Abstract— Deep Neural Networks are now being used to generate realistically-looking fake face videos. These videos can be used to violate someone's identity and privacy. In this paper, we defined a method to detect these machine-generated videos by using eye blinking features. Eye blinking is a psychological signal which is not well presented in machine generated fake videos.

Keywords— Deep Neural Network, Detecting Fake Video, Eye Blinking.

I. INTRODUCTION

As technologies are being improved, Social media are also growing in popularity, and video sharing platforms provide a convenient way to share, edit or propagate videos. Editing a video requires too much time, unlike digital images. To edit a single 10 second short video with 25 frames per second, it requires editing of 250 images which is a very time consuming task. Even if someone did alter a video, it is impossible that it would have some realistic characteristic, Which means it is easy to spot that the video is edited by someone.

But the situation was changed with recent development of generative deep neural networks, in particular, generative adversarial networks(GAN), which has led to development of tools that can generate realistic-looking fake videos called DEEPFAKE. In this method we expose deep fake videos by the lack of eye blinking in the synthesized video.

The opening and closing movement of the eyelid refers to blinking. The spontaneous blink, which is blinking unconsciously without any external stimuli and internal effort, is controlled by the premotor brain stem. Generally, for a human, there is an interval of 2-10 seconds between each blink but it can vary by individuals. The length of a typical blink is 0.1-0.4 seconds/blink. In normal videos, it is expected to observe spontaneous blink with before mentioned frequency and duration. This is not the case for GAN-created deepfake videos. Therefore using eye blinking features to expose deep fake videos is efficient.

This method uses Long-term Recurrent CNN(LRCN) which is a deep neural network that combines recursive neural network and CNN, to distinguish open and close states of the eye with the consideration of previous temporal knowledge.

II. RELATED WORKS

A. AI Generation of Fake Videos

At first thorough 3D graphic models were used to edit images or videos. Lately, the development of deep neural networks has led to the more advanced deep new algorithms, one of which is generative adversarial network(GAN), which consists of two networks: the generator network, and the discriminator network. The task that the generator network does is to produce images which cannot be distinguished from the training images. And the task that the discriminator network does is to spot the difference between the training images and the synthesized images generated by the generator network. The generator and discriminator network compete with one another. The generator network tries to produce images so that it can confuse the discriminator network, while the discriminator network tries to classify the synthesized images from the training images. The process of deepface algorithms is given in Figure 1.

B. Eye Blinking Detection

In many operations like fatigue detection and face spoof detection, detection of eye blinking has already been studied. Soukupova et al proposed a scalar quantity, it measures the ratio of the rectangular bounding box of an eye like eye openness degree in each frame.

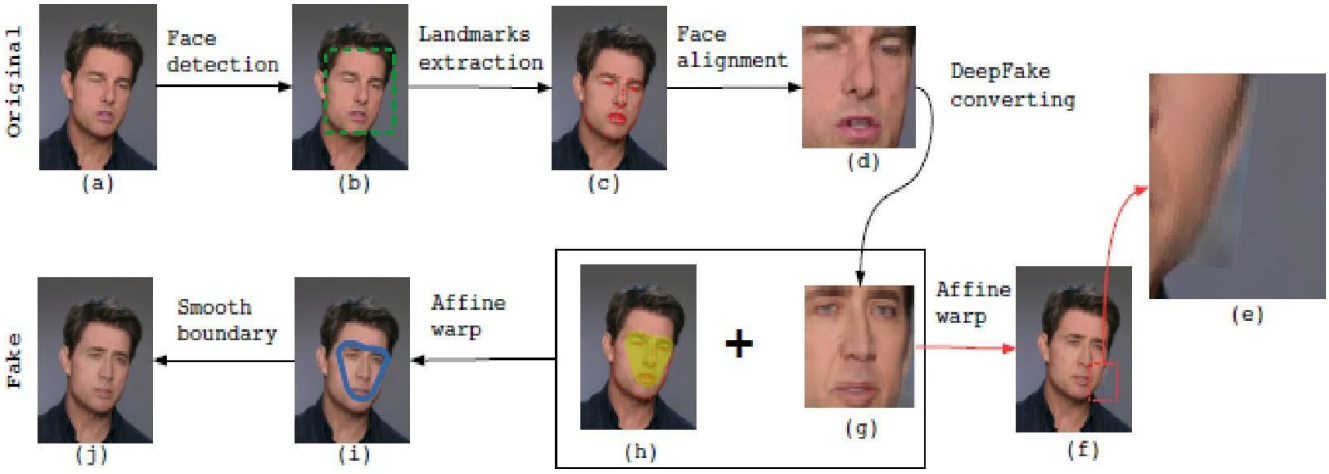

Figure 1: Deep fake Generation

III. METHOD

In this section, the method of exposing a deep fake video by detecting the eye blinking state is defined in detail. We use a CNN model named as LRCN which incorporates the relationship between consecutive frames, which means LRCN can memorize the long term dynamics. As eye blinking is a temporal process from opening to closing of eyelids.

A. Pre-processing

The process starts with converting the video in frame sequence. Generally 10 frames per seconds are derived. The next step is to extract the face area from the frames using a face detector. Then landmarks are extracted from each face region which are important features of the face such as eyes
, nose, cheeks and mouth.

Dlib- landmark's facial detector with pre-trained models. The dlib is employed to estimate the situation of 68 coordinates (x, y) that map the facial points on a person's face like figure 2.

Figure 2: 68-point facial landmarks

Then almost like an eyelid, rectangular region is extracted from the aligned face areas which is given as an input to the LRCN for dynamic open or close state prediction of eyes.The rectangle is generated by extracting bounding boxes of every eye's landmark points, then scaling the bounding box by 1.75 in vertical direction, and 1.25 in horizontal direction, to make sure that the eye region is included within the cropped rectangle region.The LRCN then takes a cropped eye region sequence for dynamic state prediction.

B. Long-term Recurrent CNNs

We use the Long-term Recurrent Convolutional Neural Network (LRCN) to store the temporary state of the eye. The LRCN model mainly consists of 3 parts: (1) feature extraction, (2) sequence learning and (3) state prediction. The input eye region is converted to discriminative features by the feature extraction module. It is implemented by Convolutional Neural Network (CNN) based on the VGG16 framework without the fc7 and fc8 layers.

C. LSTM-RNN

The output of feature extraction is fed to the sequence learning module, which is implemented with a recursive neural network (RNN) with Long Short Term Memory (LSTM) cells. The utilization of LSTM-RNN is to extend the capacity of memory of RNN models.

LSTMs are memory units that decide when to forget previous states and when and the way to update these states which are hidden states. A neural network with fully connected layers is employed which takes the output of every neuron as an input within the state prediction module. The state prediction module takes the output of LSTM and generates the sequence of 0s and 1s which is the probability of eye open or close state respectively.

D. Model Training

Two steps are needed to perform the training of the LRCN model. In the first step, we use labeled training data to train a CNN model which uses the VGG framework. These labeled training data consist of the eye region and their corresponding state according to open and close movement of eye. In the second step, the fully connected part of the network and LSTM-RNN are jointly trained using the back-propagation-through-time (BPPT) algorithm.

IV. CONCLUSION

As development in deep neural networks is increasing rapidly, which led to the generative adversarial network(GAN) that can generate realistically-looking fake face videos called deepfake. In this paper, we define a method that can expose these fake face videos by using the lack of eye blinking, which is a psychological signal that is not well presented in synthesized fake videos.

REFERENCES

[1] Yuezun Li, Ming-Ching Chang and Siwei Lyu, "In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking," University at Albany, State University of New York, USA

[2] T. Soukupova and J. Cech. Real-time eye blink detection using facial landmarks. In 21st Computer Vision Winter Workshop,pages 1–8, 2016.

[3] K. W. Kim, H. G. Hong, G. P. Nam, and K. R. Park. A study of deep cnn-based classification of open and closed eyes using a visible light camera sensor. Sensors, 17(7):1534, 201