**Review on Protecting World Leaders Against Deep Fakes**

Recent advances in deep learning have made creating sophisticated and compelling fake videos significantly easier. Those so-called deep fakes pose a significant threat to our democratism, national security, and culture. We describe a forensic technique that models facial expressions and movements that typify an individual's speaking pattern to contend with this growing threat. Although not visually apparent, the nature of how deep-fake videos are created often violates these correlations and can, therefore, be used to authenticate them.

While convincing manipulations of digital images and videos have been shown for several decades by the use of visual effects, recent advances in deep learning have led to a dramatic increase in the realism of fake content and the accessibility in which it can be created. But, since AI-synthesized content is a relatively new phenomenon, there is a shortage of forensic techniques for detecting deep fakes in particular.

One such example is that in the first generation of fake face-swapping, people did not blink or blink at the anticipated frequency, as depicted in the clever observation that was found. This error was because usually, the data used to synthesize faces did not represent the human with their closed eyes. In some cases, the next generation of synthesis methods has incorporated a pause in the processes, soon after this forensic method was made public, so that this is less successful now. D head is calculated from characteristics around the whole face in the central facial region only.

In a large number of deep fakes of different U.S. leaders such as Hillary Clinton, Barack Obama, Bernie Sanders, Donald Trump, and Elizabeth Warren, they showed how powerful this strategy is. This approach is resilient to laundry, unlike previous methods, because it relies on relatively rough measures that are not easily destroyed.

All videos were downloaded manually from YouTube, where the POI mainly faces the camera. Five examples of a second clip from the original, fake deep lip sync, comedy image, deep artificial voice, and deep fake puppet master are shown from top to bottom.

One hundred ninety models were iteratively trained with 1 to 190 features to select the best features for classification. One hundred eighty-nine models were educated on the second iteration with two functions, the first being defined on the first iteration. The accuracy of the test depended on the number of functionalities for the primary 29 processes. The AUC is almost 0.95 on average, with only 13 characteristics.

After including 30 features, accuracy slowly begins to decrease. The authors assumed that those characteristics are most important due to the nature of lip-sync fakes, which only affect the mouth, and facial swap, marionet, and comedy impersonators are not in a position to capture subtle mouth motion. The accuracy of 190 characteristics and 29 characteristics listed in Figure 4 is compared in the lower half of Table 2. The next measure of the robustness is the washing, the span of the extracted video clip, and the speaking sense of those 29 functions.

Videos have different contexts so that the necessary features have not been captured in their original model. Despite the improvement, they saw that the precision is not as high as before, indicating that POI specific models can be trained, and the existing features with stable and POI-specific features expanded. They compared their techniques to FaceForensics++'s CNN-based approach in which many models were trained to detect three kinds of facial manipulations, including facial swap deep fakes.

A model with a total of 190 features was trained for each POI. The correlations between facial expressions and head motion can be used both for a person and for videos of others. In which, compression, video clip length, and the context in which the individual talks have been checked for the effectiveness of this technique. Their methodology is robust against compression, as opposed to current pixel-based detection methods with a collection of a broader range and more effective videos or developed POI and context-related models in a wide range of contexts.

**Review on MesoNet: a Compact Facial Video Forgery Detection Network**

The paper shows an excellent rate of identification for Deepfake with more than 98% and Face2Face with 95%. Virtual images and videos have become ubiquitous virtual artifacts over the past decades as devices, and social networks became popularised. Followed by this enormous use of digital imagery, techniques for changing content have increased, for example, using editing software such as Photoshop. The field of digital forensic images is dedicated to the detection of image falsifications to regulate their circulation. The analysis of compressed image objects often provides useful insights on image manipulation.

In the context of over 100 million hours of video content on social networks every day, the spread of falsified videos is now widely recognized. The risk of falsifying news raises more and more concerns. Digital video forgery detection remains a challenging task while considerably improving the detection of forgery pictures. Besides, most picture approaches can not be applied directly to videos mainly because the frame after video compression is heavily degraded.

Deepfake is a technique aimed at replacing a specific person's face with another person's face in a picture. A new method called the car encoder typically defines the encoder network and a decoder network chaining. The encoder aims to reduce the size of the data by encoding several variables from the input layer.

Deep Fake images are generated by collecting aligned faces of two distinct A and B persons and then training an EA-auto encoder to reconstruct A faces from the face images of A datasets and an EB-auto-encoder for reconstructing the faces of B from the facial image dataset of B.

The experiment duration has a range from two seconds to three minutes and has a minimum resolution of 854 X 480 pixels. A standard value of H.264 codec but with different compression levels, which puts us in real conditions of analysis. In addition to the Deepfake dataset, they have examined whether the proposed architecture could be used to detect other facial forgeries. As a right candidate, the FaceForensics dataset contains over a thousand forged videos using the Face2Face approach, and their original dataset. This dataset is already broken down into a training, validation, and test set. In addition to extending the use of the proposed architecture to another classification task, one advantage received from the FaceForensics set is that it provides less lost accuracy compressed videos that enabled them to evaluate their model's robustness with different levels of compression.

Of more than a thousand, only 300 videos were used for training. The 150 fake videos and the original of the study series were used for sample evaluation. They denoted X as the set of inputs and Y as the set of outputs, the random variable pair that takes values in X and Y, and f as the projection function of the chosen classifier that takes values in X to the range of operation A. The function selected is to minimize the error E= E with l= 1. All networks were developed using the Keras 2.1.5 package with Python 3.5.

Network weight optimization is achieved using ADAM with default parameters with successive batches of 75 images of size 256 X 256 X 3. The authors didn't expect a higher score as there are some facial images extracted from the dataset with a very low resolution. At the strong video compression level, they observed a noticeable deterioration of the scores. The paper presenting the FaceForensics dataset used in their experiments showed better performance of classification using the state-of-the-art Xception face classification network.

Nonetheless, with the configuration provided, they only managed to fine-tune Xception to achieve a compression level 0 at 96.1 percent and with a score of 93.5 percent at compression level 23. An average of the network projection over the video is a natural way to do so. In theory, there is no justification for a gain in scores or a confidence interval indicator since frames of the same video are strongly correlated with each other.