**Review on Face Forensics++: Learning to Detect Manipulated Facial Images**

In this paper, the authors explore the reality of manipulations of modern day images and videos and how difficult it is to detect them, whether automatically or by people. They proposed an automated benchmark for the detection of facial manipulation to standardize the evaluation of detection methods. The benchmark is based, in general, on the following methods that are found in the artificial intelligence sector: Deep-Fakes, Face2Face, Face Swap and Neural Textures. This methods are regarded as prominent members of random compression rate and size facial manipulations.

The benchmark is available to the public2 and includes a hidden test set as well as a database of more than 1:8 million images were manipulated. This dataset is over an order of magnitude as it is larger than comparable, publicly available, forgery datasets.

They carried out a thorough analysis of computer-driven forgery detectors based on this data. In their research it was shown that, even in the presence of strong compression, the use of additional domain-specific knowledge improves forgery detection to unprecedented accuracy and clearly outperforms human observers.

An image-based approach called Video Rewrite was used to automatically create a new video of a person with generated mouth movements. This is credited as an advanced real-time facial reenactment system, which is capable of altering facial movements in commodity video streams, e.g., videos from the internet. For this research, the National Institute of Standards and Technology released the most extensive dataset for generic image manipulations comprising of about 50, 000 forged images and around 500 forged videos. The methods that were used for this idea are defined below:

* Face2Face is a facial reenactment system that transfers the expressions of a source video to a target video while maintaining the identity of the target person. The authors generated the reenactment video outputs by transferring the source expression parameters of each frame to the target video.
* In contrast, face swapping method replaces the face in the target video with the face in the source video. In case of facial reenactment, the expressions of the source videos are transferred to the target video while retaining the identity of the target person.
* To ensure high quality face swapping, specific video pairs were selected with similar large faces with the same gender of the persons in the similar video frame rates.
* For deep-fakes, they crop and align the images, and use a face detector. The trained source face encoder and decoder are then applied to the target face to create a fake image. The auto encoder output is then mixed with the rest of the image by editing Poisson photo
* For Neural Textures, the original video data is used to learn the target person's neural texture, including a network of rendering. In combination with an adversarial loss, this is trained with a photometric reconstruction loss.

They created production videos of different quality levels, similar to the video processing of many social networks, to create a realistic environment for manipulated images. They conduct a manual screening of the clips which resulted in 1,000 video sequences containing 509, 914 images to ensure a high quality video range and to prevent videos with face recognition. The dataset was divided into 3 sections: fixed training (720), validation (140) and test set (140).

Because the raw videos are rarely found on the Web, the authors used the H.264 codec to compress the images, which is commonly used by social networks and websites for video sharing.

To ensure adequate video quality, they only downloaded videos that offer a resolution of 480p or higher. The accuracy depended on the quality of the video, resulting in a decreasing accuracy rate of 68.69% on average on raw videos, 66.57% on high quality, and 58.73% on low quality videos. They used the YouTube-8 m database to collect videos with similar word tags as well as clips that they received from the YouTube search interface and many other social Medias.

For more clarification, a survey was conducted among computer science students to set the benchmark for fake detection. The ration received for real and fake image was 50:50 and only a certain amount of time was given to determine whether the image was real or fake. The survey included photographs of different quality, showing that human accuracy decreased with photographs of lesser quality. It was inferred from these findings that humans did not detect when using Face2Face and Neural Textures. So it's safe to say these two methods make it the most difficult for the human eye to identify.

A conservative cropping method was used for the method of forgery detection when the photo was tracked by a scale factor of 1.3 around the center of the face. This is better than other naïve approaches that include the whole picture as opposed to this method where only the face was used. Methods in are used as well as the method of computer-generated vs. natural image detection and face detection. However, the classification using XceptionNet exceeds all other techniques of forgery detection. So we can conclude that when it comes to detecting fake images, XceptionNet delivers the best result. As it reached an accuracy of 96.36%.

This paper mainly focused on the compression influence on the detectability of modern day methods of manipulation, proposing a standardized benchmark for follow-up work. All image data, trained models and the benchmark methods are available to the public and are already being used by other researchers. For this project, transfer learning in particular is of great interest to the forensic community. When new methods of deception emerge by the day, it is necessary to develop methods that can identify fakes with little or no training data.