International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1294-1299 www.ijaem.net ISSN: 2395-5252

Forensic Verification of Fake Image and Fake Videos Detection

(Dr.Sunil Maggu , Adtiya Kumar
Assistant Professor in Department of Information Technology , Maharaja Agrasen Institute of Technology

Submitted: 01-06-2021 Revised: 14-06-2021 Accepted: 16-06-2021

ABSTRACT—Deep learning has been successfully applied to solve various complex problems ranging from big data analytics to computer vision and human-level control. Deep learning advances however have also been employed to create software that can cause threats to privacy, democracy and national security. One of those deep learning-powered applications recently emerged is "deepfake". Deepfake algorithms can create fake images and videos that humans cannot distinguish them from authentic ones. The proposal of technologies that can automatically detect and assess the integrity of digital visual media is therefore indispensable. This paper presents a survey of algorithms used to create deepfakes and, more importantly, methods proposed to detect deepfakes in the literature to date. We present extensive discussions on challenges, research trends and directions related to deepfake technologies. By reviewing the background of deepfakes and state- of-the-art deepfake detection methods, this study provides a comprehensive overview of deepfake techniques and facilitates the development of new and more robust methods to deal with the increasingly challenging deepfakes.

Keywords—Deepfake ,Deep learning,CNN ,Fake Image Detection,Auto Encoders.

I. INTRODUCTION

Deepfake (stemming from "deep learning" and "fake") is a technique that can superimpose face images of a target person to a video of a source person to create a video of the target person doing or saying things the source person does. Deep learning models such as autoencoders and generative adversarial networks have been applied widely in the computer vision domain to solve various problems . These models have also been used by deepfake algorithms to examine facial expressions and movements of a person and synthesize facial images of another person making analogous expressions and movements . Deepfake algorithms normally require a large amount of image and video data to train models to create photo-realistic images and videos. As public figures such as celebrities and politicians may have a large number of videos and images available online, they are initial targets of deepfakes. Deepfakes were used to swap faces of celebrities or politicians to bodies in porn images and videos.

The main contribution of this project as follow:

- o prevent fake messages and fake information about any person throough fake video and photo.

A. RESEARCH Deepfake Creation

Deepfakes have become popular due to the quality of tampered videos and also the easy-to-use ability of their applications to a wide range of users with various computer skills from professional to novice. These applications are mostly developed based on deep learning techniques. Deep learning is well known for its capability of representing complex and high-dimensional data. One variant of the deep networks with that capability is deep autoencoders, which have been widely applied for dimensionality reduction and image compression.In that method, the autoencoder extracts latent features of face images and the decoder is used to reconstruct the face images. To swap faces between source images and target images, there is a need of two encoder-decoder pairs where each pair is used to train on an image set, and the encoder's parameters are shared between two network pairs. In other words, two pairs have the same encoder network. This strategy enables the common encoder to find and learn the similarity between two sets of face images, which are relatively unchallenging because faces normally have similar features such as eyes, nose, mouth positions.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1294-1299 www.ijaem.net ISSN: 2395-5252

Tools	Links	Key features
Faceswap	https://github.com/deepfakes/faceswap	• Using two encoder-decoder pairs. • Parameters of the encoder are shared.
Faceswap-GAN	https://github.com/shaoanlu/faceswap-GAN	Adversarial loss and perceptual loss (VGGface) are added to the auto-encoder architecture.
DeepFaceLab	https://github.com/iperov/DeepFaceLab	• Expand from the Faceswap model with new models, e.g. H64, H128, LIAEF128, SAE [33]. • Support multiple face extraction modes, e.g. S3FD, MTCNN, dlib, or manual [33].
DFaker	https://github.com/dfaker/df	• DSSIM loss function [34] is used to reconstruct face. • Implemented based on Keras library.
DeepFake-tf	https://github.com/StromWine/DeepFake-tf	Similar to DFaker but implemented based on tensorflow.

Table : Summary of notable deepfake tools

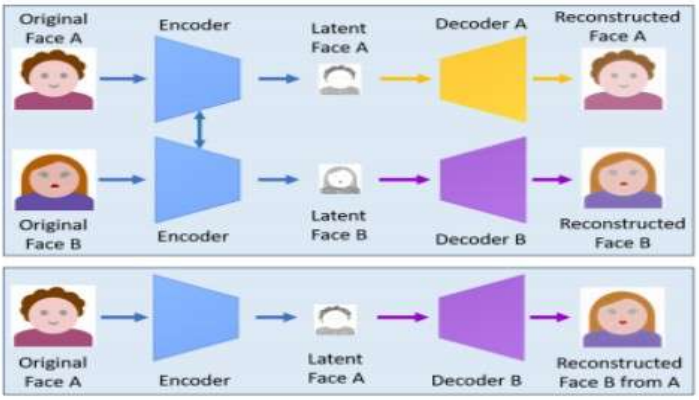

Fig. : A deepfake creation model using two encoder-decoder pairs. Two networks use the same encoder but different decoders for training process (top). An image of face A is encoded with the common encoder and decoded with decoder B to create a deepfake (bottom).

Deep fake Detection

Deepfake detection is normally deemed a binary classification problem where classifiers are used to classify between authentic videos and tampered ones. This kind of method requires a large database of real and fake videos to train classification models. The number of fake videos is increasingly available, but it is still limited in terms of setting a benchmark for validating various detection methods. To address this issue, Korshunov and Marcel produced a notable deepfake data set consisting of 620 videos based on the GAN model using the open source code Faceswap-GAN . Videos from the publicly available VidTIMIT database were used to generate low and high quality deepfake videos, which can effectively mimic the facial expressions, mouth movements, and eye blinking. These videos were then used to test various deepfake detection methods. Test results show that the popular face recognition systems based on VGG and Facenet are unable to detect deepfakes effectively. Other methods such as lip-syncing approaches and image quality metrics with support vector machine (SVM) produce very high error rate when applied to detect deepfake videos from this newly produced data set.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1294-1299 www.ijaem.net ISSN: 2395-5252

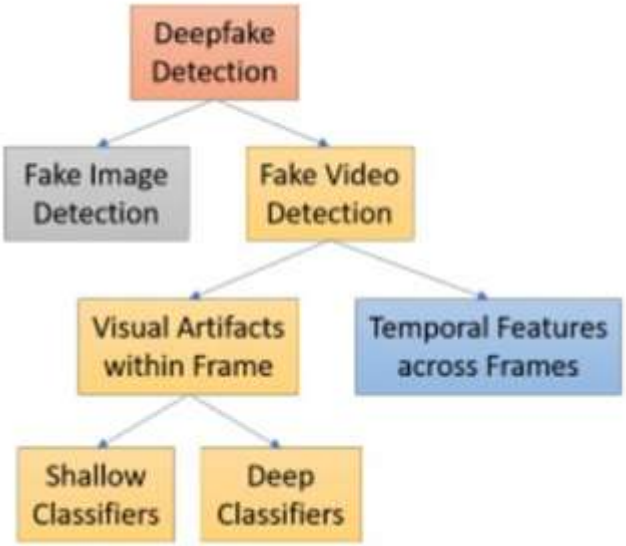

Fig : Categories of reviewed papers relevant to deepfake detection methods where we divide papers into two major groups, i.e. fake image detection and face video detection.

Fake Image Detection

Face swapping has a number of compelling applications in video compositing, transfiguration in portraits, and especially in identity protection as it can replace faces in photographs by ones from a collection of stock images. However, it is also one of the techniques that cyber attackers employ to penetrate identification or authentication systems to gain illegitimate access. The use of deep learning such as CNN and GAN has made swapped face images more challenging for forensics models as it can preserve pose, facial expression and lighting of the photographs The analytic results show that this distance increases when the GAN is less accurate, and in this case, it is easier to detect deepfakes. In case of high-resolution image inputs, an extremely accurate GAN is required to generate fake images that are hard to detect.The first phase is a feature extractor based on the common fake feature network (CFFN) where the Siamese network architecture presented in is used. The CFFN encompasses several dense units with each unit including different numbers of dense blocks to improve the representative capability for the fake images. The number of dense units is three or five depending on the validation data being face or general images, and the number of channels in each unit is varied up to a few hundreds. Discriminative features between the fake and real images, i.e. pairwise information, are extracted through CFFN learning process. These features are then fed into the second phase, which is a small CNN concatenated to the last convolutional layer of CFFN to distinguish deceptive images from genuine. The proposed method is validated for both fake face and fake general image detection. On the one hand, the face data set is obtained from CelebA , containing 10,177 identities and 202,599 aligned face images of various poses and background clutter

Fake Videos Detection

Temporal Features across Video Frames: Video manipulation is carried out on a frame-by-frame basis so that low level artifacts produced by face manipulations are believed to further manifest themselves as temporal artifacts with inconsistencies across frames. A recurrent convolutional model (RCN) was proposed based on the integration of the convolutional network DenseNet and the gated recurrent unit cells to exploit temporal discrepancies across frames . The proposed method is tested on the FaceForensics++ data set, which includes 1,000 videos , and shows promising results.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1294-1299 www.ijaem.net ISSN: 2395-5252

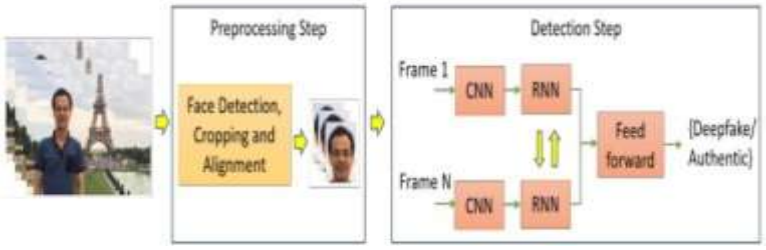

Fig : A two-step process for face manipulation detection where the preprocessing step aims to detect, crop and align faces on a sequence of frames and the second step distinguishes manipulated and authentic face images by combining convolutional neural network (CNN) and recurrent neural network (RNN) .

a) Selection: Highlight all author and affiliation lines.
b)important part of capsule network that deals with forgery detection .

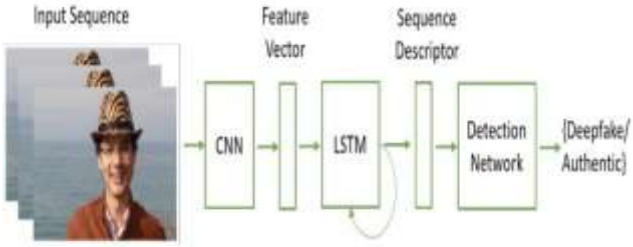

Fig : A deepfake detection method using convolutional neural network (CNN) and long short term memory (LSTM) to extract temporal features of a given video sequence, which are represented via the sequence descriptor. The detection network consisting of fully-connected layers is employed to take the sequence descriptor as input and calculate probabilities of the frame sequence belonging to either authentic or deepfake class

Visual Artifacts within Video :

This subsection investigates the other approach that normally decomposes videos into frames and explores visual artifacts within single frames to obtain discriminant features. These features are then distributed into either a deep or shallow classifier to differentiate between fake and authentic videos.

Deep classifiers :

Advantage of the proposed method is that it needs not to generate deepfake videos as negative examples before training the detection models. Instead, the negative examples are generated dynamically by extracting the face region of the original image and aligning it into multiple scales before applying Gaussian blur to a scaled image of random pick and warping back to the original image. This reduces a large amount of time and computational resources compared to other methods, which require deepfakes are generated in advance. A dynamic routing algorithm is deployed to route the outputs of the three capsules to the output capsules through a number of iterations to separate between fake and real images. The method is evaluated through four data sets covering a wide range of forged image and video attacks.

The proposed method yields the best performance compared to its competing methods in all of these data sets. This shows the potential of the capsule network in building a general detection system that can work effectively for various forged image and video attacks.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 June 2021, pp: 1294-1299 www.ijaem.net ISSN: 2395-5252

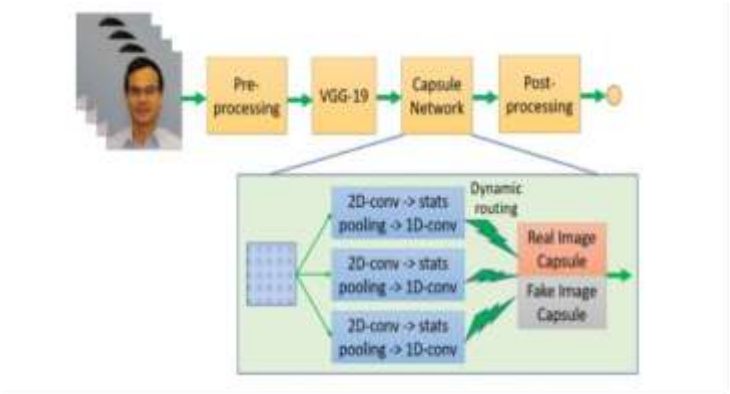

Fig : Capsule network takes features obtained from the VGG-19 network to distinguish fake images or videos from the real ones (top). The pre-processing step detects face region and scales it to the size of 128x128 before VGG-19 is used to extract latent features for the capsule network, which comprises three primary capsules and two output capsules, one for real and one for fake images (bottom). The statistical pooling constitutes an

Training

Two sets of training images are required. The first set only has samples of the original face that will be replaced, which can be extracted from the target video that will be manipulated. This first set of images can be further extended with images from other sources for more realistic results. The second set of images contains the desired face that will be swapped in the target video. To ease the training process of the autoencoders, the easiest face swap would have both the original face and target face under similar viewing and illumination conditions.

II. RESULTS

It is not unusual to find deepfake videos where the manipulation is only present in a small portion of the video (i.e. the target face only appears briefly on the video, hence the deepfake manipulation is short in time). To account for this, for every video in the training, validation and test splits, we extract continuous subsequences of fixed frame length that serve as the input of our system. In Table I present the performance of our system in terms of detection accuracy using sub-sequences of length N = 20, 40, 80 frames. These frame sequences are extracted sequentially (without frame skips) from each video. The entire pipeline is trained end-to- end until we reach a 10-epoch loss plateau in the validation set.

Model	Training acc. (%)	Validation acc. (%)	Test acc. (%)
Conv-LSTM, 20 frames	99.5	96.9	96.7
Conv-LSTM, 40 frames	99.3	97.1	97.1
Conv-LSTM, 80 frames	99.7	97.2	97.1

Table . Classification results of our dataset splits using subsequences with different lengths

As we can observe in our results, with less than 2 seconds of video (40 frames for videos sampled at 24 frames per second) our system can accurately predict if the fragment being analyzed comes from a deepfake video or not with an accuracy greater than 97%

III. CONCLUSION

In this paper we have presented a temporal-aware system to automatically detect deepfake videos. Our experimental results using a large collection of manipulated videos have shown that using a simple convolutional LSTM structure

we can accurately predict if a video has been subject to manipulation or not with as few as 2 seconds of video data. We believe that our work offers a powerful first line of defense to spot fake media created using the tools described in the paper. We show how our system can achieve competitive results in this task while using a simple pipeline architecture. In future work, we plan to explore how to increase the robustness of our system against manipulated videos using unseen techniques during training.