ISSN: 2366-1313

DEEP LEARNING BASED DEEPFAKE CREATION AND IDENTIFICATION

[1]R. VEERA REDDY, [2]S. AKHILA

[1,2]Assistant Professor, Dept of CSE, Megha institute of engineering and Technology for women, Ghatkesar (T.S)

Abstract: *Deep learning has been used in a wide range of applications, such as wearable vision, natural language processing, and image detection. Advances in deep knowledge of algorithms in image detection and manipulation have led to the introduction of DeepFax. Deepfakes use deep learning algorithms to create fake images that are sometimes very difficult to distinguish from real ones. With increasing concerns about privacy and security, many techniques have emerged to detect deepfake images. This article explores the use of deep learning to generate and detect deepfakes. This article also suggests using deep domain photo enhancement techniques to enhance the magnificence of the created deepfakes.*

Keywords*: Res-Next Convolution neural network, Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Computer vision.*

I. INTRODUCTION

Machine vision is advancing day after day in different fields from basic image detection software to automotive and robotics [1], one of the applications that stem from machine vision is Deepfake. Deepfake is a technique that uses deep learning algorithms to create fake images usually by swapping a person's face from a source image into another person's face in a target image, with a resulting fake image that is sometimes hard to detect. The underlying mechanism for deepfake creation is using deep learning encoders and decoders, which have been used extensively in the machine vision domain [2]. the encoders work by extracting all the features in an image and then decoders are used to generate the fake image. deepfake methods need a large number of images and videos to train the deep learning models, this used to be a hard task but in the time that we live in you could easily find a large dataset of images on social media, this wide availability of data has led to the development of more complicated deepfake techniques, many of deepfake algorithms are designed using Tensorflow [3]. TensorFlow is an open-source software library for numerical computation using data-flow graphs. It was originally developed by Google to be used internally in it research and development of machine learning and deep neural networks, but the system is

ISSN: 2366-1313

general enough to be applicable in a wide variety of other domains as well and it became very popular for machine learning applications after it was made available publicly and free to use.

upload the video. The video will be processed by the model and the output will be rendered back to the user with the classification of the video as deepfake or real and confidence of the model.

While some deepfakes can be created by traditional visual effects or computer-graphics approaches, the recent common underlying mechanism for deepfake creation is deep learning models such as autoencoders and generative adversarial networks (GANs), which have been applied widely in the computer vision domain [2]. These models are used to examine facial expressions and movements of a person and synthesize facial images of another person making analogous expressions and movements. Deepfake methods 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 first deepfake video emerged in 2017 where face of a celebrity was swapped to the face of a porn actor. It is threatening to world security when deepfake methods can be employed to create videos of world leaders with fake speeches for falsification purposes [3]. Deepfakes therefore can be abused to cause political or religion tensions between countries, to fool public and affect results in election campaigns, or create chaos in financial markets by creating fake news. It can be even used to generate fake satellite images of the Earth to contain objects that do not really exist to confuse military analysts, e.g., creating a fake bridge across a river although there is no such a bridge in reality. This can mislead a troop who have been guided to cross the bridge in a battle.

MOTIVATION

The increasing sophistication of mobile camera technology and the ever-growing reach of social media and media sharing portals have made the creation and propagation of digital videos more convenient than ever before. Deep learning has given rise to technologies that would have been thought impossible only a handful of years ago. Modern generative models are one example of these, capable of synthesizing hyper

ISSN: 2366-1313

realistic images, speech, music, and even video. These models have found use in a wide variety of applications, including making the world more accessible through text-to-speech, and helping generate training data for medical imaging.

Like any trans-formative technology, this has created new challenges. So-called "deep fakes" produced by deep generative models that can manipulate video and audio clips. Since their first appearance in late 2017, many open-source deep fake generation methods and tools have emerged now, leading to a growing number of synthesized media clips. While many are likely intended to be humorous, others could be harmful to individuals and society. Until recently, the number of fake videos and their degrees of realism has been increasing due to availability of the editing tools, the high demand on domain expertise.

II. LITERATURE SURVEY

Face Warping Artifacts [4] used the approach to detect artifacts by com-paring the generated face areas and their surrounding regions with a dedicated Convolutional Neural Network model. In this work there were two-fold of Face Artifacts.

Their method is based on the observations that current deepfake algorithm can only generate images of limited resolutions, which are then needed to be fur-ther transformed to match the faces to be replaced in the source video. Their method has not considered the temporal analysis of the frames.

Detection by Eye Blinking [5] describes a new method for detecting the deep-fakes by the eye blinking as a crucial parameter leading to classification of the videos as deepfake or pristine. The Long-term Recurrent Convolution Network (LRCN) was used for temporal analysis of the cropped frames of eye blinking. As today the deepfake generation algorithms have become so powerful that lack of eye blinking cannot be the only clue for detection of the deepfakes. There must be certain other parameters must be considered for the detection of deep-fakes like teeth enchantment, wrinkles on faces, wrong placement of eyebrows etc.

Capsule networks to detect forged images and videos [6] uses a method that uses a capsule network to detect forged, manipulated images and videos in different scenarios, like replay attack detection and computer-generated video detection.

In their method, they have used random

ISSN: 2366-1313

noise in the training phase which is not a good option. Still the model performed beneficial in their dataset but may fail on real time data due to noise in training. Our method is proposed to be trained on noiseless and real time datasets.

Recurrent Neural Network [7] (RNN) for deepfake detection used the approach of using RNN for sequential processing of the frames along with Image-Net pre-trained model. Their process used the HOHO dataset consisting of just 600 videos.

Their dataset consists small number of videos and same type of videos, which may not perform very well on the real time data. We will be training out model on large number of Real-time data.

Synthetic Portrait Videos using Biological Signals approach extract bio-logical signals from facial regions on pristine and deepfake portrait video pairs. Applied transformations to compute the spatial coherence and temporal consistency, capture the signal characteristics in feature vector and photoplethysmography (PPG) maps, and further train a probabilistic Support Vector Machine (SVM) and a Convolutional Neural Network (CNN). Then, the average of authenticity probabilities is used to classify whether the video is a deepfake or a pristine.

Fake Catcher detects fake content with high accuracy, independent of the generator, content, resolution, and quality of the video. Due to lack of discriminator leading to the loss in their findings to preserve biological signals, formulating a differentiable loss function that follows the proposed signal processing steps is not straight forward process.

There have been existing survey papers about creating and detecting deepfakes, presented in[8]. For example, Mirsky and Lee focused on reenactment approaches (i.e., to change a target's expression, mouth, pose, gaze or body), and replacement approaches (i.e., to replace a target's face by swap or transfer methods). Verdoliva separated detection approaches into conventional methods (e.g., blind methods without using any external data for training, one-class sensor-based and model-based methods, and supervised methods with handcrafted features) and deep learning-based approaches (e.g., CNN models). Tolosana et al. [9 categorized both creation and detection methods based on the way deepfakes are created, including entire face synthesis, identity swap, attribute manipulation, and expression swap. On the other hand, we carry out the survey with a different perspective and taxonomy. We categorize the

ISSN: 2366-1313

deepfake detection methods based on the data type, i.e., images or videos, as presented in Fig. 1. With fake image detection methods, we focus on the features that are used, i.e., whether they are handcrafted features or deep features. With fake video detection methods, two main subcategories are identified based on whether the method uses temporal features across frames or visual artifacts within a video frame. We also discuss extensively the challenges, research trends and directions on deepfake detection and multimedia forensics problems.

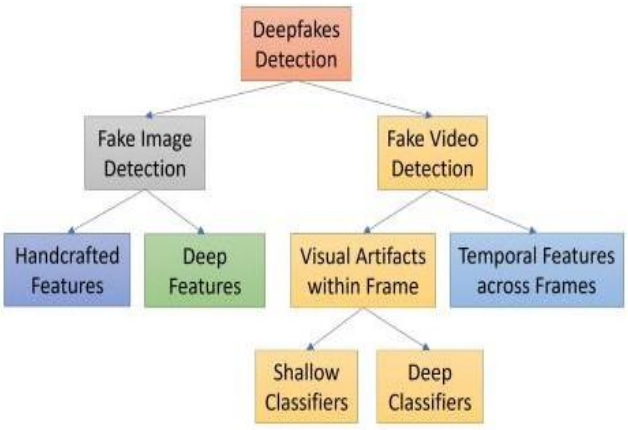

Fig.1 Categories of reviewed papers relevant to deepfake detection methods where we divide papers into two major groups, i.e., fake image detection and face video detection.

III. PROPOSED SYSTEM

This document lays out a project plan for the development of Deepfake video detection using neural network.The intended readers of this document are current and future developers working on Deepfake video detection using neural network and the sponsors of the project. The plan will include, but is not restricted to, a summary of the system functionality, the scope of the project from the perspective of the "Deepfake video detection" team (me and my mentors), use case diagram, Data flow diagram,activity diagram, functional and non- functional requirements, project risks and how those risks will be mitigated, the process by which we will develop the project, and metrics and measurements that will be recorded throughout the project.

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 [33–

ISSN: 2366-1313

35]. The first attempt of deepfake creation was FakeApp, developed by a Reddit user using autoencoder-decoder pairing structure. 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. Fig. 3 shows a deepfake creation process where the feature set of face A is connected with the decoder B to reconstruct face B from the original face A. This approach is applied in several works such as DeepFaceLab, DFaker, DeepFaketf (tensorflow-based deepfakes).

By adding adversarial loss and perceptual loss implemented in VGGFace to the encoder-decoder architecture, an improved version of deepfakes based on the generative adversarial network, i.e., faceswap-GAN, was proposed in [10]. The VGGFace perceptual loss is added to make eye movements to be more realistic and consistent with input faces and help to smooth out artifacts in segmentation mask, leading to higher quality output videos. This model facilitates the creation of outputs with 64x64, 128x128, and 256x256 resolutions. In addition, the multi-task convolutional neural network (CNN) from the FaceNet implementation is used to make face detection more stable and face alignment more reliable. The CycleGAN [60] is utilized for generative network implementation in this model.

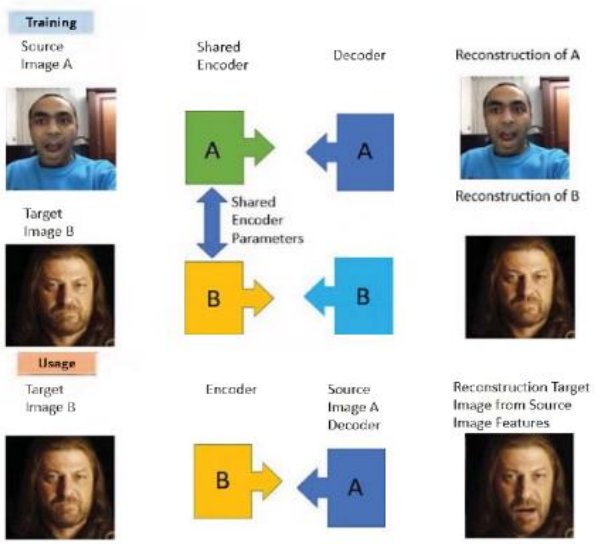

Fig.2 : Deepfake Creation Layout

Creating deepfake videos

To detect the deepfake videos it is very important to understand the creation process of the deepfake. Majority of the tools including the GAN and autoencoders takes a source image and

ISSN: 2366-1313

target video as input. These tools split the video into frames, detect the face in the video and replace the source face with target face on each frame. Then the replaced frames are then combined using different pre-trained models. These models also enhance the quality of video my removing the left-over traces by the deepfake creation model. Which result in creation of a deepfake looks realistic in nature. We have also used the same approach to detect the deepfakes. Deepfakes created using the pretrained neural networks models are very realistic that it is almost impossible to spot the difference by the naked eyes. But in reality, the deepfakes creation tools leaves some of the traces or artifacts in the video which may not be noticeable by the naked eyes. The motive of this paper to identify these unnoticeable traces and distinguishable artifacts of these videos and classified it as deepfake or real video.

MesoNet CNN

Mesonet is a neural network designed specifically to detect deepfakes. Deepfakes videos are usually found all around social media, the nature of videos on social media platform like Instagram is that they are low quality compressed video so microscopic analysis based on image noise is not possible and MesoNet takes that into consideration, also detecting deepfake on at a higher semantic level is hard as even humans sometimes struggle with detecting deepfakes [21], therefore MesoNet relies on an intermediate approach using a deep neural network with small amount of layers. This network begins with a pattern of four layers of successive convolutions and pooling Fig. 3, and is followed by a dense network with one hidden layer. convolution and pooling are used to extract features of an image, it common pattern to use a convolution layer followed by a pooling layer as the convolution layer extract the features and the pooling layer creates a down sampled version of the feature map. To improve generalization, the convolutional layers use ReLU activation functions that introduce non-linearities and Batch Normalization to regularize the output, and the fullyconnected layers use Dropout to regularize and improve their robustness.

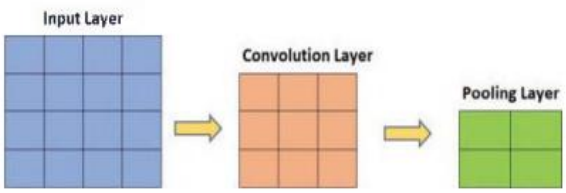

Fig.3 Convolution and Pooling Layers Pattern used in CNNs to extract feature maps

ISSN: 2366-1313

IV. CONCLUSION

In this paper deepfake creation and detection was explored as well as the integration of deep learning image enhancement method to increase the quality of deepfakes created. From our experiment, the use of image enhancement methods like DFDNet gave higher-quality look deepfake images, which give them a more genuine look unlike the typically low quality associated with fake images on the internet. Deepfake detection in the case of face-swapping deepfakes, they are hard to detect when the person is facing straight towards to camera, when the person looks to his side, imperfections in the deepfake generated image can be seen. We believe that to further improve the performance of deepfake detectors, the focus should be on using datasets of difficult conditions like this. Our future work will explore generating deepfakes with reduced imperfections and higher quality using image enhancement methods to try and make them harder to detect for deepfake detection methods.

REFERENCES

[1] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies,MatthiasNießner, “FaceForensics++: Learning to Detect Manipulated Facial Images” in arXiv:1901.08971.

[2] Deepfake detection challenge dataset:https://www.kaggle.com/c/deepfak e-detection-challenge/data Accessed on 26 March, 2020

[3] Yuezun Li , Xin Yang , Pu Sun , Honggang Qi and SiweiLyu “Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics” in arXiv:1909.12962

[4] Deepfake Video of Mark Zuckerberg Goes Viral on Eve of House A.I. Hearing :https://fortune.com/2019/06/12 /d eepfake-mark-zuckerberg/ Accessed on 26 March, 2020

[5] 10 deepfake examples that terrified and amusedtheinternet:https://www.creative blo q.com/features/deepfake-examples Accessed on 26 March, 2020

[6] TensorFlow:https://www.tensorfl ow.or g/ (Accessed on 26 March, 2020)

[7] Keras: https://keras.io/ (Accessed on 26 March, 2020)

[8] PyTorch:https://pytorch.org/ (Accessed on 26 March, 2020)

[9] N Srivani, Prasadu Peddi (2021),

ISSN: 2366-1313

Face Assessment Learned From Existing Images In Order To Classify The Gender Of The Images Based On Improved Face Recognition, (TURCOMAT), Vol 12, issue 6, pp: 5724-5735

[10] J. Thies et al. Face2Face: Real-time face capture and reenactment of rgb videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2387–2395, June 2016. Las Vegas, NV.

[11] Prasadu Peddi (2019), "Data Pull out and facts unearthing in biological Databases", International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.

[12] FaceSwap:https://faceswaponline.com/ (Accessed on 26 March, 2020)

[13] Deepfakes, Revenge Porn, And The Impact On Women :

https://www.forbes.com/sites/chenxiwang/ 2019/11/01/deepfakes-revenge-porn-an d- the-impact-on-women/