Deepfake Detection

R. TEJASWINI[1], K. RAKESH[2], M. VEERA MANI KANTA[3], M. MAHESH REDDY[4], M. TEJAGGNA[5]
[1] *Assistant Professor, Department of Electronics and Communication Engineering, Vasireddy Venkatadri Institute of Technology, India.*
[2, 3, 4, 5] *UG Student, Department of Electronics and Communication Engineering, Vasireddy Venkatadri Institute of Technology, India.*

Abstract- ***The expeditious progress in facial image generation and exploitation has now come to a point where it raises serious concerns to the social and political society. This leads to the creation of fake information and new which ultimately results in loss of trust in digital content. We have developed a detection model using convolution neural network (CNN) for face detection and Recurrent neural network (RNN) for video classification. Even though this technology is remarkable it leads to social and political concerns. So far, with the help of released tools for the generation of deep fake videos have been widely used to create fake celebrity videos or revenge porn and fake political speeches, etc. Governmental entities are already looking into the issue of these fake videos which are likely to create political tensions. so, it is essential to have a tool for detecting these fake videos. (We need AI to fight an AI)***

Indexed Terms- Convolution Neural Networks (CNN), LSTM, RNN.

I. INTRODUCTION

[Font: Times New Roman, Size:10] Manipulation of facial images has now become universal, and in the digital community this is one of the most demanding topics. DeepFakes has proved that deep learning and machine learning techniques can be used to manipulate a person's video by substituting their faces with another face of a different person. The facial expression manipulation and facial identity manipulation are the two major facial manipulation methods. The most popular methods for facial manipulation are Face2Face and DeepFakes which are methods in facial expression manipulation and identity manipulation respectively, these techniques generate highly realistic manipulation of faces in photographs. we show that we can accurately and reliably detect these deepfakes videos better than human detection. We take advantage of the recent advances in deep learning, especially the ability to extract powerful image features using convolutional neural networks (CNNs) and recurrent neural network (LSTM) for forgery detection. We develop the detection technique by training the neural network in an efficient and optimal way.

II. PREVIOUS METHODS

These are few detection methods used earlier to predict a given video whether it is a fake or real video. In all these detection methods we used techniques like geometry of face and the dynamics of the mouth shape according to the speaking word

1. Deepfake Detection Using Biological Signals: based on the facial expressions and using a method called the Photoplethysmography (PPG) cells.
2. Deepfake Detection Using Phoneme-Viseme Mismatches: based on the fact that the dynamics of the mouth shape, are sometimes different or inconsistent with the spoken phoneme
3. Forensic Technique Using Facial Movements: This model tracks facial expression and movements of a single video provided as input. This detection uses support vector model (SVM)
4. Recurrent Convolutional Strategy: The Recurrent Convolutional Strategy uses recurrent convolutional models (RCM) for detecting face manipulation in videos.

For detection from learned features, we evaluate five network architectures known from the literature to solve the classification task:

1. Cozzolino is CNN-based network using this to extract the important feature of the facial images like distance between the lips and the dynamics of mouth

2. Bayar and Stamm have also created a convolution neural network which uses a constrained convolutional layer which is then given to another two convolutional followed by 2 max pooling and finally taken by three fully connected layers.
3. Rahmouni have also adopted a different Convolution neural network architectures which has a global pooling layer. This CNN can be used to computes different kind of statistical measurements like mean, variance, maximum and minimum.
4. Inspired by InceptionNet, MesoInception-4 has been created which is a CNN-based network to detect face munipulations in given image. This network has two inception modules and two classic convolution layers together with a completed max-pooling layers. Further, there are two fully-connected layers.
5. XceptionNet is a traditional CNN trained on Inspried by ImageNet, XceptionNet has been created which is a traditional CNN based on separable convolutions. By interchanging the final fully connected layer with two outputs, the other layers are initialized with the ImageNet weights. This network predicts the deepfake image by taking the fact of skin color, blurness of the image in various sections, pixel segmentation.

Compression	Raw	HQ	LQ
[13] XceptionNet Full Image	82.01	74.78	70.52
[26] Steg. Features + SVM	97.63	70.97	55.98
[16] Cozzolino *et al.*	98.57	78.45	58.69
[9] Bayar and Stamm	98.74	82.97	66.84
[49] Rahmouni *et al.*	97.03	79.08	61.18
[4] MesoNet	95.23	83.10	70.47
[13] XceptionNet	**99.26**	**95.73**	**81.00**

III. PROPOSED METHOD

- DATASET

The data is gathered from different datasets available like face Forensic++, deepfakes detection challenge dataset from Kaggle, celebrity deepfakes videos. This data is pre-processed as face cropped videos by detecting a face in an image

- PRE-PROCESSING THE DATA

There are 6 CNNs out of them 3 CNNs are used for facial binary classification and 3 CNNs for calibration, this is formulated as multiclass classification of discretized displacement pattern. In these CNNs, without specific explanation, we follow AlexNet to apply the ReLU nonlinearity function after the pooling layer and fully connected layer

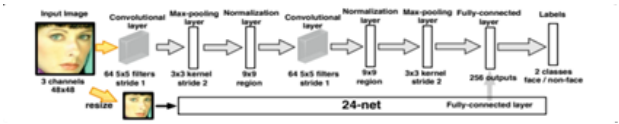


Fig. 1 – 48net face detection and extraction

Using this network, we can detect the person's face and crop the frame accordingly. This will process the data to have only face-cropped videos. Using various tools to extract face from an image like OpenCV, PILLOW library.

- CNN FOR FEATURE EXTRACTION

In reference with IEEE Signal Processing Society Camera Model Identification Challenge, using the ImageNet pre-trained model we output a deep representation of every frame by the InceptionV3 with fully connected layer. The final feature vectors after the final pooling layer is been used as the input for sequential LSTM.

- LSTM FOR SEQUENCE PROCESSING.

Let us take an image frames' sequence of CNN feature vector as input and a 2-node neural network having the probabilities of the sequences which is a part of a given deepfake video. we'd like to handle a significant challenge that is the design of a model in a consequential manner which can recursively process a sequence. To resolve this issue, By the usage of a 2048 wide LSTM unit, which is the expected result. Especially, in the period of training, a sequence of ImageNet feature vectors is given to out LSTM model which is succeed by a 512 fully connected layer. Lastly, we compute the probabilities of each frame using a softmax layer which gives the result of sequence being either pristine or deepfake.

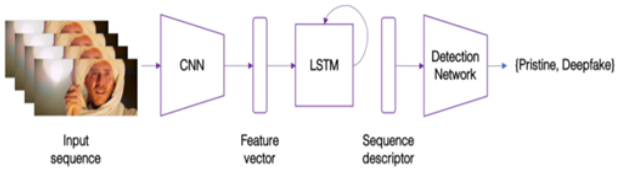


Fig. 2 – Layout of the detection model

Using processed data, we can train the model for deepfake detection in the detection model, we use ImageNet and LSTM networks for Feature extraction and video classification respectively. The trained model which is XceptionNet can be loaded using tools such as Torch and Network models to predict the given video as Real/Fake. We use XceptionNet as our trained model to predict the deepfakes

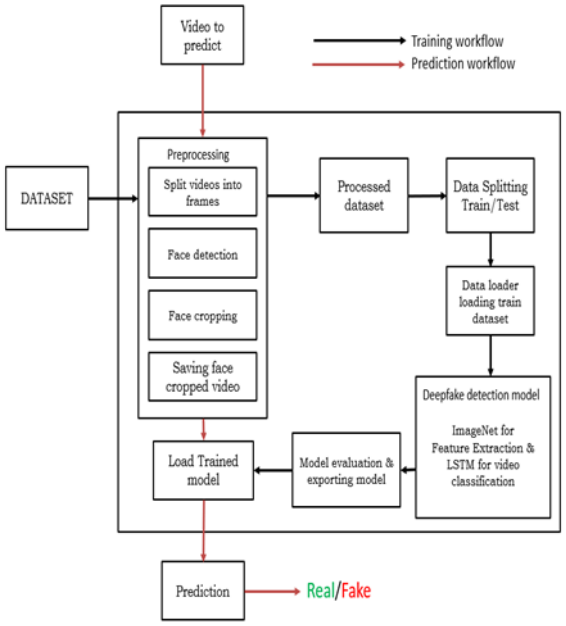


Fig. 2 – Process of the detection model system with Training workflow and Prediction workflow

Acuuracy measurement for different deepfake detection methods

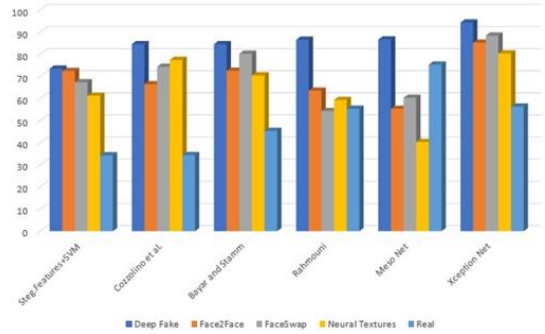
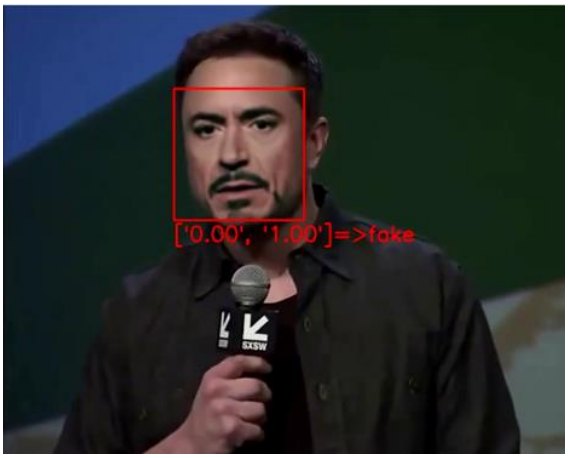


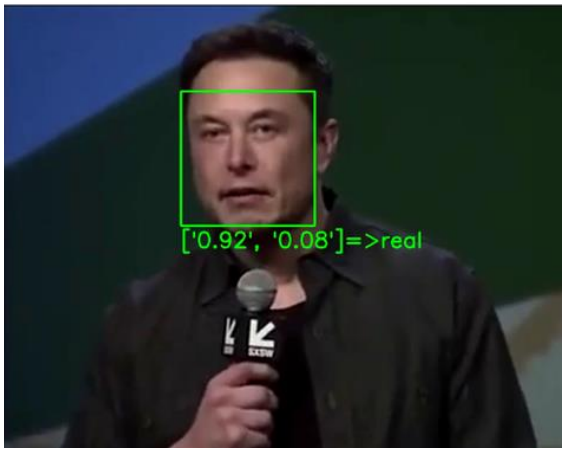
Fig. 3 – Accuracy comparision for different detection models

Here is the comparision of the various deepfakes detetion methods for the accuracy of detection. The comaparision is made with different deepfakes generation methids like DeepFake, Face2Face, FaceSwap, Neural Textures, etc. from the table we can clearly see that XceptionNet is predicting with much better accuracy than the other method.

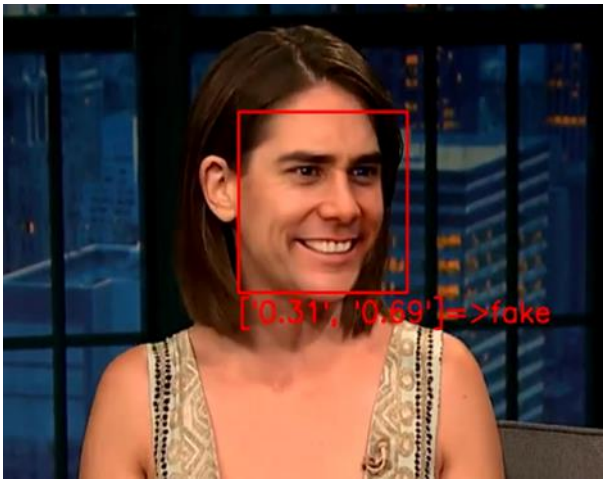
IV. RESULTS

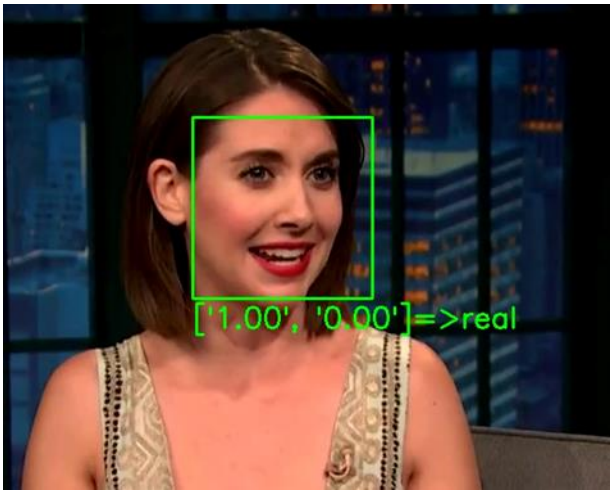
Our work attempts to give the advanced tool for the defense of spotting fake media created using advance deep learning algorithms. We show how our system can achieve competitive results in this task while using simple architecture. These are few deepfakes videos taken from youtube given along with their link, we have taken these videos, processed them, and tested them to predict the deepfake along with probability. Our model predicts the output for every frame and combines them to get resultant videos as follows



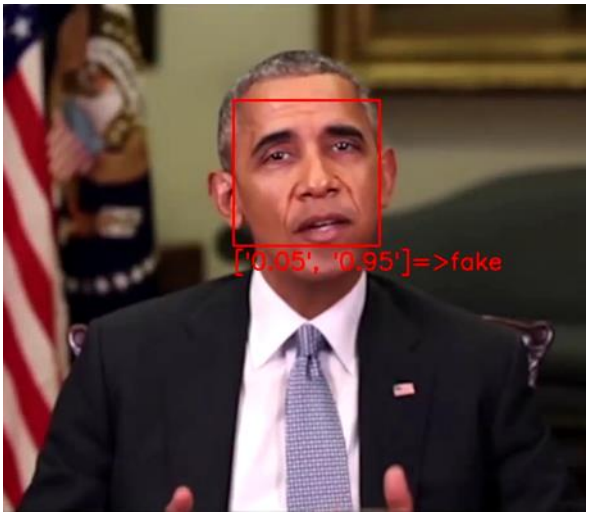


Reference from YouTube; link to above video click here.

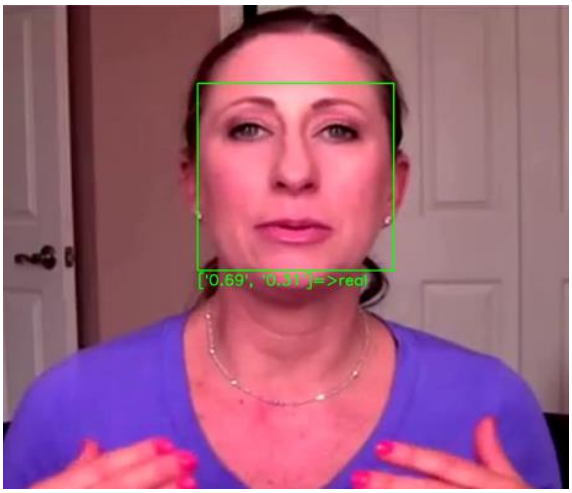




Reference from YouTube; Links to above videos: Original and Fake



These are the fake video generated by deepfake methods which our model has detected perfectly



These are real video taken randomly from YouTube and our model has detected perfectly

CONCLUSION

While the present advanced facial image manipulation methods such as DeepFakse, Face2Face, FaceSwap, etc; exhibit visually impresive results, we show that these fake videos can be detected by the properly trained forgery detectors. By a few different learning-based approaches we can solve the issue of detection in low-quality video. In this paper we focus on the influence of compression on the detectability of state-of-the-art manipulation methods, proposing a standardized benchmark for follow-up work. All image data, trained models, as well as our benchmark, are publicly available and are already used by other researchers. In particular, transfer learning is of high interest in the forensic community. By the increase in various new manipulation techniques, there is a need to develop certain methods which can detect deepfakes with little to no training data. We hope that the dataset and benchmark become a stepping stone for future research in the field of digital media forensics, and in particular with a focus on facial forgeries.

ACKNOWLEDGMENT

We have taken a lot of effort into this project. However, completing this project would not have been possible without the support and guidance of our faculty of Electronics and Communication Engineering department at Vasireddy Venkatadri Institute of Technology and our fellow students. We would like to extend our sincere thanks to all of them. We are highly indebted to Ms. R. Tejaswini (Ass. Professor) for her guidance and supervision. We would like to thank her for providing the necessary information and resources for this project.

We would like to express our gratitude towards our parents & our friends for their kind co-operation and encouragement which help us a lot in completing this project. Our thanks and appreciations also go to our colleague in developing the project. Thank you to all the people who have willingly helped us out with their abilities.

REFERENCES

[1] FaceForensics detection paper: FaceForensics++: Learning to Detect Manipulated Facial Images – 2019 paper from IEEEexplore.

[2] Irene Amerini, Lamberto Ballan, Roberto Caldelli, Alberto Del Bimbo, and Giuseppe Serra A SIFT-based forensic method for copy-move attack detection and transformation recovery IEEE Transactions on Information Forensics and Security, Mar. 2011.

[3] P. Bestagini et al. Local tampering detection in video sequences. IEEE International Workshop on Multimedia Signal Processing, pages 488–493, Sept. 2013. Pula, Italy

[4] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. In IEEE International Conference on Image Processing, 2017.

[5] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes. In IEEE Computer Vision and Pattern Recognition, 2017

[6] Francois Chollet. Xception: Deep Learning with Depthwise Separable Convolutions. In IEEE Conference on Computer Vision and Pattern Recognition, 2017

[7] Paul Upchurch, Jacob Gardner, Geoff Pleiss, Robert Pless, Noah Snavely, Kavita Bala, and Kilian Weinberger. Deep feature interpolation

for image content changes. In IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[8] Luca D'Amiano, Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. A PatchMatch-based Dense-field Algorithm for Video Copy-Move Detection and Localization. IEEE Transactions on Circuits and Systems for Video Technology, in press, 2018

[9] Yuezun Li, Ming-Ching Chang, and Siwei Lyu. In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking. In IEEE WIFS, 2018

[10] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. FaceVR: Real-Time Gaze-Aware Facial Reenactment in Virtual Reality. ACM Transactions on Graphics (TOG), 2018.

[11] Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. Mesonet: a compact facial video forgery detection network. arXiv preprint arXiv:1809.00888, 2018.

[12] Justus Thies, Michael Zollhofer, and Matthias Nießner. Deferred neural rendering: Image synthesis using neural textures. ACM Transactions on Graphics 2019 (TOG), 2019.