Deepfake Detection with Deep Networks

Rakesh Omprakash

Abstract

Deepfake detection is growing in popularity as the number of fake videos increases as well as the difficulty for finding if the videos are real or not increases. My method tackles the problem by using face recognition software and predicting the authenticity of the videos. ResNeXt and Long Short Term Memory *(LSTM) are powerful and apt networks for this use case. The models are tested on various datasets and the hyper-parameters are fine-tuned to get the best results. A pre-processing algorithm has also been aligned with the model to detect the faces and crop only the faces. At the same time a limited number of frames of all the videos are chosen to ensure efficiency of the model.*

1. Introduction

Deepfake is basically a fake video which has never happened and causes reactions which would come if the video was real. It is created with the help of Deep learning. More specifically Generative Adversarial Networks (GANs) are used to create these unrealistic videos which looks very close to real videos because of the capability of the deep networks. Due to this there is a lack of trust in digital media and there is always an uncertainty as to what is real and what is not. In order to ensure whether the content we see is real or not and to ease the uncertainty in people's mind there has been research in the Deepfake detection methods. There are several Deepfake softwares like FaceSwap which makes videos of real people doing fictional activities. The proposed method is highly accurate in distinguishing between fake and real videos using ResNeXt CNN and LSTM. The pre-processing algorithm used with the network has a face recognition model to detect the faces in the videos and only the faces are cropped.

2. Methodology

Firstly, for the feature extraction a ResNeXt CNN which is already trained is used. This network is selected because it is better than the usual Residual Network (ResNet) as it needs lesser hyperparameters. One of the differences between them is cardinality which is the size of set of transformations. The ResNeXt consists of 50 layers and is of 32x4 dimensions. An apt learning rate is chosen for converging the gradient descent of the model.

Following this network a Long Short Term Memory (LSTM) network is used which predicts whether the videos are realistic or not. They are a type of Recurrent Neural Networks which learns order dependence in prediction problems involving sequence. The feature vectors from the ResNext network are passed to the LSTM network. One LSTM layer is used with 2048 hidden layers. It is used to compare the sequential frames of the video and performs temporal analysis. The Leaky Relu activation function is used as it deals with the dying ReLU problem and therefore increases the performance of Deep Neural Networks. In the model an adaptive average pooling layer with 1 as the output parameter is used. Hence a H x W image output size is achieved. A sequential layer is also used for running on the frames in a sequence. The confidence of the model at prediction is given by a SoftMax layer.

For the dataset to be used for training equal amount of realistic and unrealistic videos are used. The videos from datasets like Celeb-DF, Deepfake detection challenge (DFDC) and FaceForensic++(FF) are used. In these videos only the face is cropped and used in the model. All the frames of the cropped video are joined together. Frames without the face are not considered and are removed. A limit is placed on the number of frames. It is chosen considering the mean of total frames of all the videos. This also helps to increase the efficiency of the model. The limit is set to the first 150 frames of the video. The dataset is divided into 80% training and 20% testing.

The Adam optimizer is used and a learning rate of 0.00001 is used to get the global minimum of gradient descent. Also 0.001 weight decay is used. Cross entropy approach is used to find the loss. It gives smooth gradients, and hence better converging is achieved.

Once the videos are preprocessed, they are directly feed to the trained model for prediction whether real or fake.

3. Experimental Results

4000 videos in total used for training and testing the model. All the videos are pre-processed first and then used for training, testing and predicting new videos. The figures given in this paper shows a sample from the Celeb dataset(Figure 2), output after preprocessing(Figure 3), few outputs of the real videos after prediction(Figure 4) and few outputs of the fake videos after prediction(Figure 5).

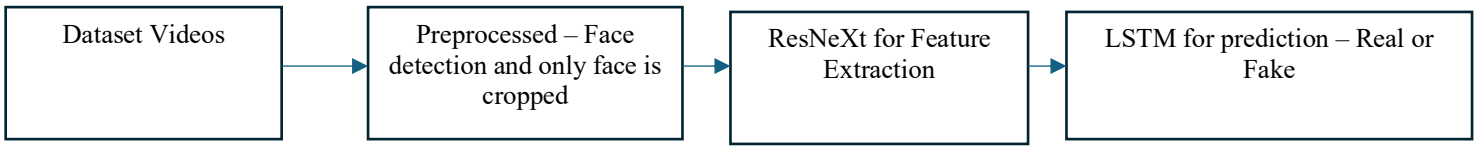
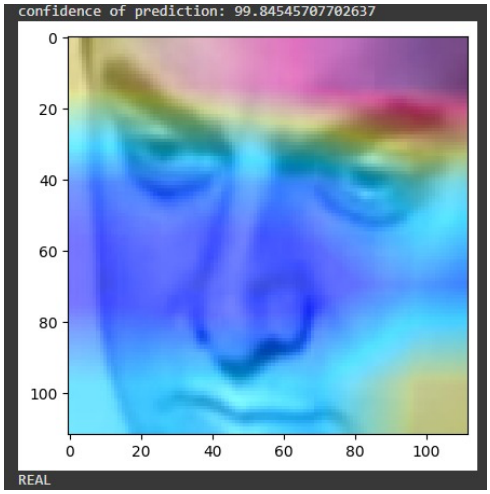
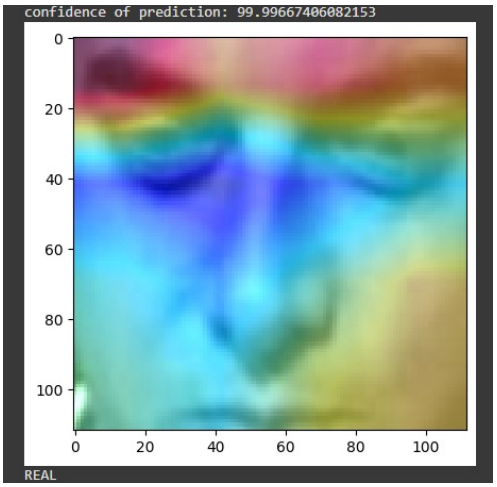


Figure 1: Methodology of the system

Figure 2: Sample from a video in Celeb_real dataset

Figure 3: Sample after preprocessing





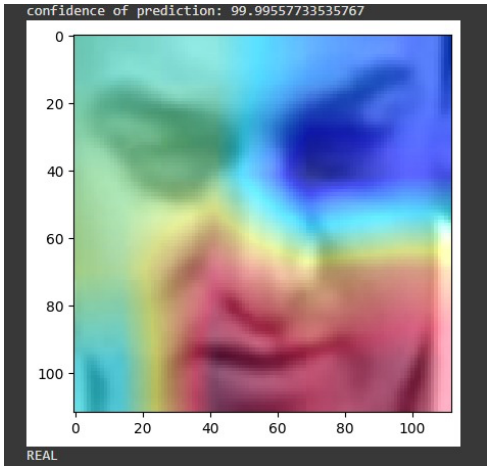
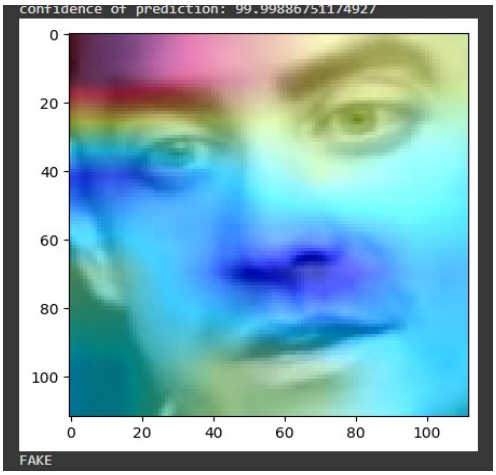


Figure 4: Outputs from the model when feed with real videos



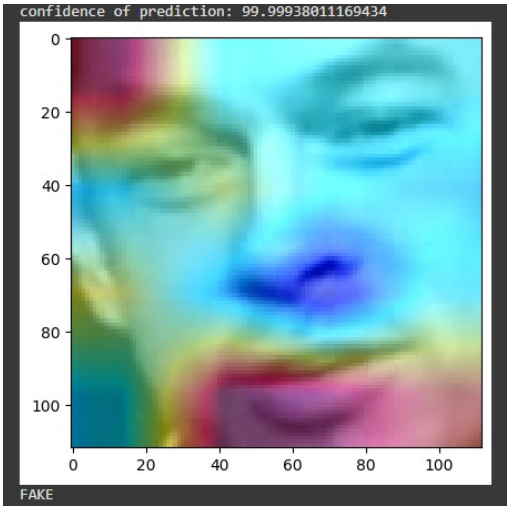
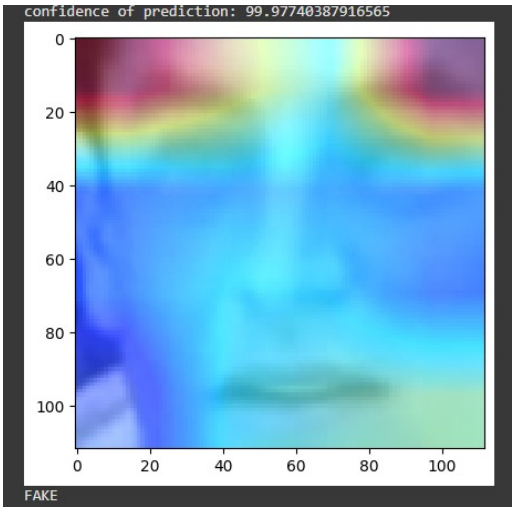


Figure 5: Outputs from model when feed with fake videos

4. Discussion and Conclusion

The proposed model predicts whether videos are real or not along with the prediction confidence of the model. The prediction accuracy is very high, and this system can be confidently used for Deepfake detection. However, this method only detects the face and finds the authenticity. This can be extended to be able to predict the entire content of the video and produce the result. Also, more data can be used to make the existing network robust.



References

[1] M. S. Rana, B. Murali and A. H. Sung, "Deepfake Detection Using Machine Learning Algorithms," 2021 10th International Congress on Advanced Applied Informatics (IIAI-AAI), Niigata, Japan, 2021, pp. 458-463, doi: 10.1109/IIAI-AAI53430.2021.00079.

[2] M. S. Rana, M. N. Nobi, B. Murali and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," in IEEE Access, vol. 10, pp. 25494-25513, 2022, doi: 10.1109/ACCESS.2022.3154404.

[3] S. R. B. R, P. Kumar Pareek, B. S and G. G, "Deepfake Video Detection System Using Deep Neural Networks," 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 2023, pp. 1-6, doi: 10.1109/ICICACS57338.2023.10099618.

[4] A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," in IEEE Access, vol. 10, pp. 18757-18775, 2022, doi: 10.1109/ACCESS.2022.3151186.

[5] M. L. Saini, A. Patnaik, Mahadev, D. C. Sati and R. Kumar, "Deepfake Detection System Using Deep Neural Networks," 2024 2nd International Conference on Computer, Communication and Control (IC4), Indore, India, 2024, pp. 1-5, doi: 10.1109/IC457434.2024.10486659.

[6] S. S. Chauhan, N. Jain, S. C. Pandey and A. Chabaque, "Deepfake Detection in Videos and Picture: Analysis of Deep Learning Models and Dataset," 2022 IEEE International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2022, pp. 1-5, doi: 10.1109/ICDSIS55133.2022.9915885.

[7] S. Guefrachi et al., "Deep learning based DeepFake video detection," 2023 International Conference on Smart Computing and Application (ICSCA), Hail, Saudi Arabia, 2023, pp. 1-8, doi: 10.1109/ICSCA57840.2023.10087584.