Detection of Deepfake Video Using Residual Neural Network and Long Short-Term Memory

Article *in* INTERNATIONAL JOURNAL OF NEXT-GENERATION COMPUTING · February 2023

DOI: 10.47164/ijngc.v14i1.1046

Aarti Karandikar
Shri Ramdeobaba Kamla Nehru Engineering College

Yogesh Thakare
Shri Ramdeobaba College of Engineering and Management

Roshan Kumar Sah
Katihar Engineering College,katihar

Detection of Deepfake Video Using Residual Neural Network and Long Short-Term Memory

A M Karandikar, Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur.

Y N Thakare, Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur.

O Sah, R K Sah, S Nafde and S Kumar, Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur.

The appearance of web-based media has implied genuine and anecdotal stories introduced in such a comparative manner that it can now and then be hard to differentiate the two. Similarly, manipulation of real photos, audios or videos with the help of Artificial Intelligence techniques is done such that it is difficult to distinguish between the real and fake thus called Deepfake. It can happen to big celebrities, politicians, and to layman as well for some malicious purpose. Consequently, this procedure can end up being very threat to human culture subsequently expected to identify it appropriately. This paper intends to tackle this issue by proposing a model that uses Residual Neural Network (ResNet50) and Long Short-term Memory (LSTM) to detect video as fake or real. This approach tries to find flaws in the fake data left behind while its creation using neural based techniques like generative adversarial networks (GAN).

Keywords: GAN, Deep Learning, Neural Network, Memory.

1. INTRODUCTION

Advancement of Deep Learning(Pan, Sun, Wang, Zhang, and Sinnott, 2020) Techniques and accessibility of good Internet Connection empowers the mass creation synthetic videos that closely resemble real videos known as deepfake videos. Deepfake technology is impacting the human beings and society with decimating results prone to be seen if not controlled conveniently. Rather using the artificial intelligence technology for betterment of the people, some are putting their mind in developing technology that enabled fake contents to influence the people in our society. Forestalling such fake news and rumours in the general public become incomprehensible particularly when the offenders are deliberately and attempting to stigmatize the personality by making the counterfeit contents. Therefore, there is need to detect this deepfake images, audios or videos. Prior to create procedure or detecting deepfake videos, it is needed to know how it is made and understand what flaws it made during its creation so it can be used to detect deepfake videos. Deepfakes are generally made using a neural network-based architecture. Nowadays, generative adversarial networks (GANs) prove to be one of the most effective models to do this type of task. This involves two neural networks that is Generator and Discriminator which are placed in contest with one another. Initially the Discriminator is trained with real images. After this Generator try to create its own data which is then sent to Discriminator to classify it as fake or real. The loss is then calculated and back propagation is done so that Generator will learn from this loss and this process is continued till Generator is able to fool Discriminator. GAN was presented in 2014 and its advanced designs can create images that seems to be real and one can't perceive if it's genuine or not. There can be various approach to spot the flaws, irregularities and inconsistencies made during the creation of deepfake (DF) videos. The techniques can be Phoneme-Viseme Mismatches (Agarwal, Farid, Fried, and Agrawala, 2020), Appearance and Behaviour (Li, Yang, Sun, Qi, and Lyu, 2020), Eye Blinking (Li, Chang, and Lyu, 2018), Learning based Features, Facial Artifact technique(Goyani and Patel, 2017). The most sophis-

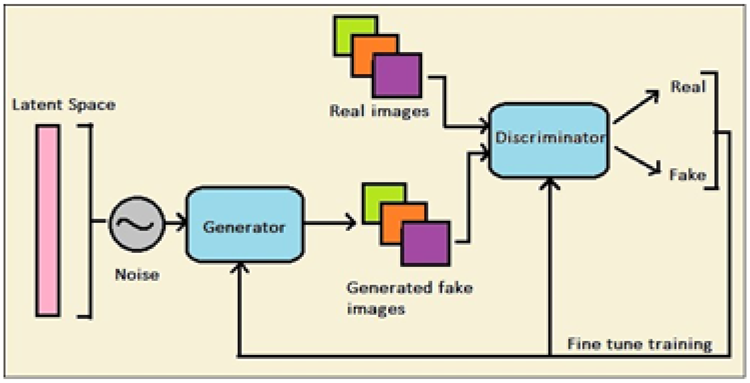


Figure 1. Deepfake Generation using GAN (Karandikar et al., 2020).

ticated technique can be learning based technique which will learn the features of the data and find differences in features of real and fake data. So, we are using this learning-based approach to train our model. Our method learns the features of both real and fake videos. It discovers the abnormalities that is available in the fake videos while its creation. First, dataset is preprocessed to get the face of the personalities at frame level. We are interested in face as it is mainly the face that is changed for malicious purpose. After this the features are extracted using CNN transfer learning model. Here, Residual Neural Network (ResNet50) is used to extract features from face and then Long Short-Term Memory (LSTM) is used for sequence processing of frames in a video. At last layer Softmax is used to classify video as genuine or fake and Confusion Matrix is used for metrics evaluation(Pulver and Lyu, 2017).

2. LITERATURE SURVEY

There are various techniques that has been proposed to detect deepfake videos. This portion examines such strategies and their limitations that empowers to go for other methods. In published paper (Karandikar et al., 2020), initially face is extracted and then aligned which is then passed through classifier (VGG-16) to classify video as fake or real. Here only CNN model is used to train the model but to know the sequence of frames in the videos it is required to use recurrent neural network. This is the limitation that is solved by our model. Also, they have used dataset from one source only which is not enough to train model sophisticatedly. In research paper (Agarwal et al., 2020), deepfake (DF) videos are detected by Phoneme-Viseme Mismatches technique. Here, inconsistencies in the movement of the mouth with the words is observed to detect DF videos. Experiment suggest that this isn't the situation in most of the DF videos. It should be needed to detect other features as well which is efficiently done by our model. Published paper (Li et al., 2018) is detecting deepfake (DF) videos by detecting Eye Blinking. It is found that the eye blinking signal is not much accurate or not even found in fake videos. Their technique uses the absence of flickering as a hint for detection. However only one parameter is not enough to detect DF videos efficiently and certain other parameters should be taken into consideration to detect DF videos. Our strategy can consider all other parameters contained in face to learn flaws in a better way. In research paper (Nguyen, Yamagishi, and Echizen, 2019), to detect synthesized images and videos, they have used capsule networks. Here, random noise is utilized in the training phase due to which their model is performing better for their dataset only but may not perform well on actual time data. Our technique is working well on noiseless and instantaneous datasets as well.

3. METHODOLOGY

This paper proposes a method in which first video frame is provided as input for training and after training, output is saved which is then utilized during prediction on new video and finally give result as fake or real. Our system has two separate architecture for model training and prediction. The architecture of model training and prediction is shown in figure below:

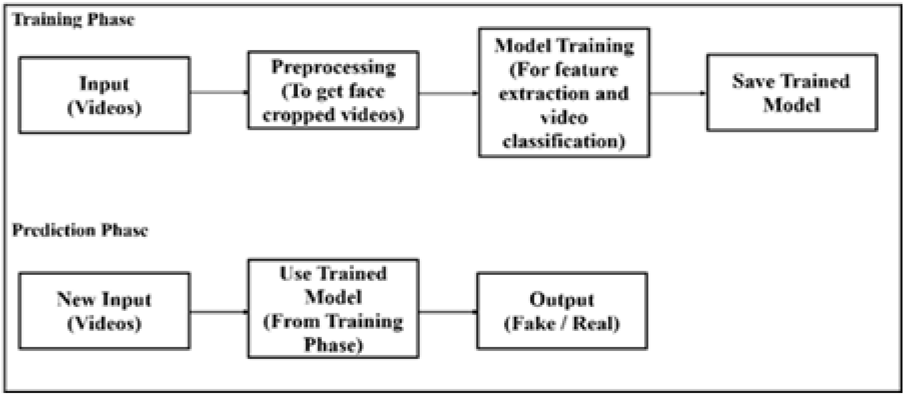

Training Phase
Input (Videos)
Preprocessing (To get face cropped videos)
Model Training (For feature extraction and video classification)
Save Trained Model
Prediction Phase
New Input (Videos)
Use Trained Model (From Training Phase)
Output (Fake / Real)

Figure 2. System Architecture

3.1 Dataset:

Dataset has been downloaded from two repositories and total of 1603 videos from two sources are used as shown in Table 1. below:

Dataset Name	Contents	No. of Videos	Total No. of Videos
Deepfake Detection Challenge Dataset (DFDC)	Real	100	400
	Fake	300	
Celeb-DF-v1	Celeb-real	158	1203
	Celeb-synthesis	795	
	YouTube-real	250	
Grand Total:			1603

3.2 Preprocessing:

In preprocessing phase, the aim is to get the face of the person at frame level. The face is focussed as in a large portion of the deepfake videos predominantly face is swapped for malevolent reason(He, Zhang, Ren, and Sun, 2016). Dataset preprocessing includes four steps: Split Video into frames, Face detection, Cropping face Creating and saving new Face Cropped Videos. Dataset's mean is first calculated to keep up the consistency in the number of frames. After this the new dataset of processed cropped face is made which contains the frames equivalent to the mean. During this preprocessing phase, the frames that lack face are also ignored.

3.3 Model Architecture

First to extract features from the preprocessed videos it is needed to use one of the best Convolutional Neural Network (CNN). We have tried with different CNN models like ResNext50, Inception Net, and ResNet50 and found that ResNet50 is showing better performance than other CNN model therefore we have used ResNet50 as a transfer learning CNN model for extracting features of the preprocessed videos at frame level. The extracted features are then needed to classify as fake or real, but before its classification it is needed that there should be a network in between which should handle sequence dependency of frames in a video. For this recurrent neural network (RNN) is mainly used. Long Short-Term Memory (LSTM) is one of the types of RNN. We are using LSTM because enormous structures can be effectively trained using it. At last, we are using Softmax to classify video as fake or real.

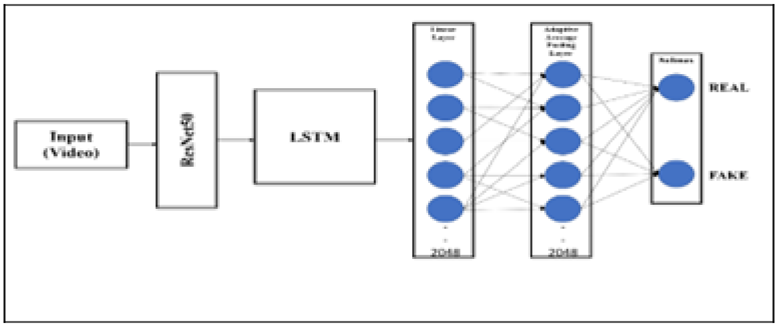
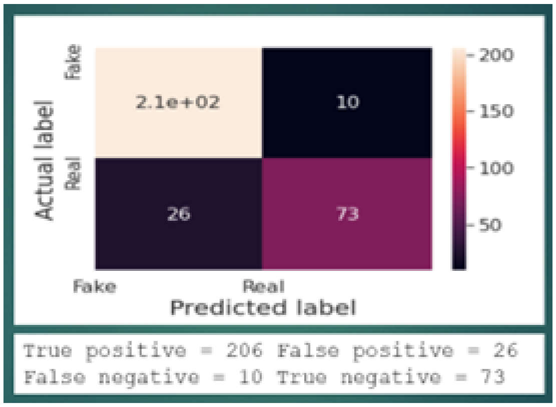


Figure 3. Model Architecture

In the Prediction phase, a video which is needed to predict as fake or real goes through a preprocessing phase so that it becomes compatible with trained model. Then it is sent for prediction using trained model.

3.4 Evaluation Metrics

For evaluating the model, Confusion Matrix(Salmon, Kleynhans, Schwegmann, and Olivier, 2015) has been used. It is a mainly used to evaluate the classification model's performance. This matrix contrasts the genuine target values and those anticipated by the model. The confusion matrix after validating our model is shown in Figure 4.



```
True positive = 206 False positive = 26
False negative = 10 True negative = 73
```

Figure 4. Confusion Matrix

For hypertuning the model certain hypermeters are used which are listed below: Rectified Linear Unit (ReLU): "It is an activation function which can overcomes the vanishing gradient problem, allowing models to learn faster and perform better." Adam Optimizer: "Adam is a replacement optimization algorithm for stochastic gradient descent for training deep learning models" This optimizer is efficient for complex datasets where large number of parameters are required. It is also memory efficient comparing to others. Dropout: "Dropout is a regularization technique to avoid overfitting problem and thus increasing the generalizing power." In our model, Dropout value of 40Learning Rates: "It determines the step size at each iteration while moving toward a minimum of a loss function." It is needed to choose right learning rate as too small learning rate make model to not learn at all and too large learning rate can lead to exploding gradient problem. We have used learning rate as 0.001 in our model based on experimental observation. Number of Epoch and batch size: "Number of Epoch defines the number times that the learning algorithm will work through the entire training dataset and batch size defines the number of samples to work through before updating the internal model parameters." We have used 100 epochs with 314 batch size for training data and 79 batch size for validation data.

4. RESULT

Loss and accuracy that we got in training and validation set is shown below: In Training set, loss was found to be 0.107631and accuracy was found to be 95.38Loss was 0.503155 and Accuracy was 88.57

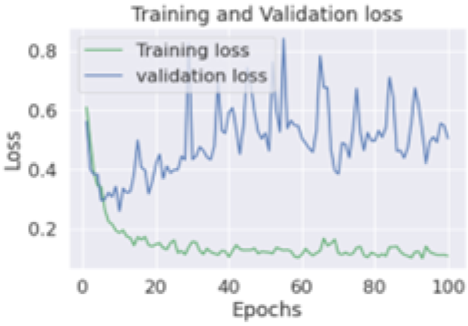


Figure 5. Training and Validation Loss

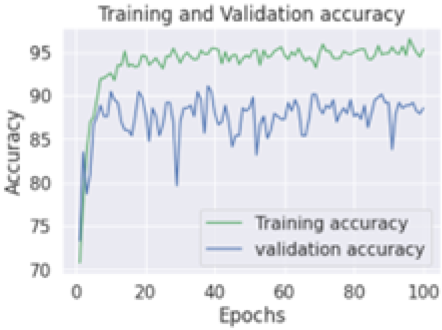


Figure 6. Training and Validation Accuracy

The outputs showing classification of video (real or fake) with its confidence are shown in Figure 7 and Figure 8.

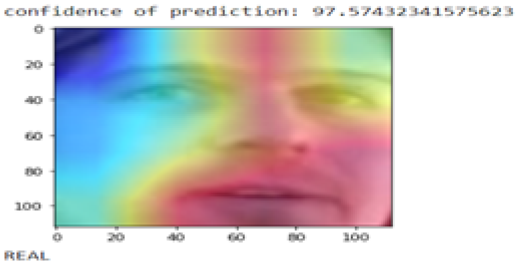


Figure 7. Confidence of Prediction for Real Video

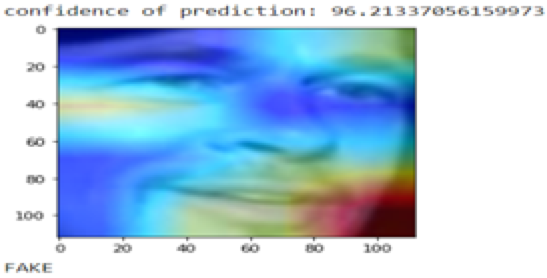


Figure 8. Confidence of Prediction for Fake Video

5. CONCLUSIONS

We have proposed a learning-based model that can classify the video as deepfake or real with good training and validation accuracy. The model is also tuned using hyperparameter based on experimental observation that proves to enhance the accuracy and performance of the model. The proposed technique is using one of the best CNN models that is ResNet CNN for feature extraction and video classification using LSTM which is helpful in sequence processing of frames in a video. For the actual time data also this model is giving high accuracy.

References

Agarwal, S., Farid, H., Fried, O., and Agrawala, M. 2020. Detecting deep-fake videos from phoneme-viseme mismatches. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.*

Goyani, M. M. and Patel, N. M. 2017. Judgmental feature based facial expression recognition and fer datasets-a comprehensive study. *International Journal of Next-Generation Computing*, 62–81.

He, K., Zhang, X., Ren, S., and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 770–778.

Karandikar, A., Deshpande, V., Singh, S., Nagbhidkar, S., and Agrawal, S. 2020. Deepfake video detection using convolutional neural network. *International Journal of Advanced Trends in Computer Science and Engineering 9,* 2, 1311–1315.

Li, Y., Chang, M.-C., and Lyu, S. 2018. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In *2018 IEEE International workshop on information forensics and security (WIFS).* IEEE, 1–7.

Li, Y., Yang, X., Sun, P., Qi, H., and Lyu, S. 2020. Celeb-df: A large-scale challenging dataset for deepfake forensics. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 3207–3216.

Nguyen, H. H., Yamagishi, J., and Echizen, I. 2019. Capsule-forensics: Using capsule networks to detect forged images and videos. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 2307–2311.

Pan, D., Sun, L., Wang, R., Zhang, X., and Sinnott, R. O. 2020. Deepfake detection through deep learning. In *2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT).* IEEE, 134–143.

Pulver, A. and Lyu, S. 2017. Lstm with working memory. In *2017 International Joint Conference on Neural Networks (IJCNN).* IEEE, 845–851.

Salmon, B. P., Kleynhans, W., Schwegmann, C. P., and Olivier, J. C. 2015. Proper comparison among methods using a confusion matrix. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).* IEEE, 3057–3060.

Ms. A. M. Karandikar is working as an assistant professor in computer science and engineering department at Shri Ramdeobaba Collge of Engineering and Management, Nagpur.
E-mail:karandikara@rknec.edu

Mr. Yogesh Thakare received his B.E. degree in electronics and telecommunication engineering from the Amravati University, Amravati, India, in 2009, and the M.Tech. in electronic system and communication engineering from Government College of Engineering, Amravati, India, in 2013. He is pursuing his Ph.D from Sant Gadge Baba Amravati University, Amravati. In 2017, he joined Shri Ramdeobaba College of Engineering and Management, Nagpur, as an Assistant Professor, in department of electronics engineering. His current research interests include digital system design, VLSI system design, memory circuits design and its analysis. Mr. Thakare is an associate member of the Institution of Engineers (India) (IEI) and a Life Member of the Indian Society for Technical Education (ISTE).
E-mail: thakareyn@rknec.edu

Mr. Omprakash Sah (BE, Computer Science and Engineering), SRCOEM. Omprakash's research interests center on ML, AI, DL projects and interested in making full end to end projects, And the research was carried out under the observation of Prof. Aarti Karandikar.
E-mail:saho@rknec.edu

Mr. Roshan Kumar sah has been a student at Department of Computer Science and Engineering at Shri Ramdeobaba College of Engineering and Management, Nagpur. He received his bachelor degree in the batch of 2021. This is his final semester group project and inspired from watching many deepfake videos on internet. Roshan currently works as Software Engineer.
E-mail:saho@rknec.edu

Mr. Sushil Kumar (BE, Computer Science and Engineering), SRCOEM. Sushil's reserach interests center on the deepfake videos, inspired from deepfake videos on internet which are being misused for awful reasons. And the research was carried out under the observation of Prof. Aarti Karandikar.
E-mail:saho@rknec.edu

Mr. Sarang Nafde , a final year student of Computer Science and Engineering at RCOEM, developed his skills in Machine learning and Deep learning by self-learning through courses. Deepfake videos were surging, and that piqued his interest, which ultimately led to the extensive research and execution of this project. The research was carried out under the guidance of Prof. Aarti Karandikar.
E-mail:saho@rknec.edu