International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)
ISSN(Print) : 2249-6831 ISSN(Online) : 2249-7943
Vol. 13, Issue 1, Jun 2023, 21-36

DEEP FAKE DETECTION USING DEEP LEARNING TECHNIQUE

ANKUSH GHOSH[1], SHASHIDHAR T[2] & AJITH PADYANA[3]

[1,2&3]*Department of Computer Science and Engineering Acharya Institute of Technology, Bangalore, India*

ABSTRACT

Deep fake refers to image or video that are fakes and depict events that never occurred. That is, manipulated digital media as image of a person which is replaced by another person's likeness. The development in the area of deep fakes are equal parts astonishing and concerning. This paper demonstrates a method for automatically and proficiently detecting face alteration in an image, with an emphasis on a current technique that is used to produce forged videos that are incredibly convincing. Videos typically do not lend themselves well to traditional visual forensics approaches because of the compression, which severely degrades the data. By examining several technologies and how they are used to detect deep fakes. Researchers in this discipline will benefit from this study since it will feature cutting-edge techniques for finding deep-fakes on social media. Due to the thorough description of the most recent techniques and dataset utilized in this field, it will also aid in comparison with earlier research.

KEYWORDS: Deep fake detection, deep learning, Deepfake & Social Media

Received: May 17, 2023**; Accepted:** Jun 01, 2023**; Published:** Jun 16, 2023; **Paper Id.:** IJCSEITRJUN20233

Original Article

INTRODUCTION

The concept of deepfake refers to image or video that are fakes and depict events that never occurred. That is, manipulated digital media as image of a person which is replaced by another person's likeness.

Unlike old methods of Photoshop, these deepfakes are created by deep neural network to be nearly indistinguishable from the real counterpart. The development in the area of deepfakes are equal parts astonishing and concerning. Nowadays, the dangers of false news are generally recognized, where more than 1 billion hours of video footageare consumed on social media per day [1]. So, the spread of fake footage raises increasing worries. In wrong hands, these technologies can be used to spread misinformation and undermine public trust almost like an identity theft situation, where anyone can get anyone to say anything and it's opposite. In order to give Barack Obama credit for saying things that he never said, a fictitious film was made about him in 2018. Joe Biden's footage of him sticking out his tongue during the US 2020 election has already been manipulated using deepfakes [2]. These harmful deepfake apps have the potential to significantly impact our society and contribute to the propagation of misleading information, especially on social media [3].

This project brings together the most recent research ondeepfake recognition in order to improve its use by forensic. Therefore, as we improve at creating deepfakes, we have to also improve at spotting them. As a result, there is a need to detect deepfakes quickly for digital forensics based on the image's mesocopic features. Recent investigations have shown that deepfake films and pictures are being widely circulated on social media. As a result, it is now more critical than ever to identify deep fake films and images [1].

To recognize and stop deepfake, multiple companies like Google and Facebook have started research projects. Deep learning algorithms like recurrent neural network (RNN), long short-term memory (LSTM) and even mixed model are being used for deepfake detection in this field. Recent studies have demonstrated that deep neural networks are efficient at identifying fraudulent information and rumors in social media posts. It is important to successfully detect fraudulent images that are difficult for humans to distinguish from the real ones. For these detections, the Deepfake model requires a vast amount of training data [4]. According to the current result, deep neural networks have achieved exceptional success in detecting false news and rumors in social media posts.

Social media has been used to spread a lot of deepfake films as technology has become more widely available. Deepfakes manipulate images and films of people using deep learning techniques so that humans are unable to tell them apart from the real thing [5]. Deepfake is becoming a more important problem in current life. Deepfake has been used to manufacture false informationand rumors for politicians by swapping the faces of well-knownHollywood stars' over obscene photographs and films. These harmful applications of deepfakes can havethe potential to significantly impact our society and contribute to the propagation of misleading information, especially on social media. It implements Deepfake detection to assist overcome the shortcomings of the present approach.

RELATED WORK

A brief on various DeepFake detection techniques is provided in this section which are categorized into many groups based on various deep learning and feature extractiontechniques [6]. This machine learning technique called deep learning is built on the same principle as a neural network.

A. Deepfake detection techniques based on Deep Learningtechniques

The deep learning architecture, which was influenced by artificial networks, employs several number of hidden layersof finite size to extract more advanced data from raw input data. The term "deep" in deep learning refers to the network's utilization of several hidden layers. The sophistication of the dataset is used to calculate the hidden layers. For more complicated data to efficiently provide the desired results, more hidden layers are needed [7]. Deepfake detection

depends on neural networks' strong capacity for learning andthe progressively larger sample set.

Using deep learning, in these disciplines gives more latest/accurate results compared to methods of machine learning. Deep learning has shown promising results in the identification of deepfakes. Deep learning is also has been effectively applied in a number of fields recently other than deepfake detection, such as natural language processing, computer vision and speech recognition. There are several techniques based on deep learning which are been proposed.

Among the most accepted model in deep neural network model isconvolutional neural network (CNN) [8]. Similar to other deep neural network consisting of input, output and multiplehidden layers where the input is being passed through and while passing of input, convolutional operation is being applied on the input. The Convolution operation in this context denotes a matrix dot product [9]. Non-linearity activation functions like RELU is being implemented after getting the dot product, proceeded by further convolutional operations like maximum, minimum or average pooling on pooling layer which has a main objective to minimize the data's dimension.

Another neural network that can be used for feature extraction and learning in case of sequential data called as recurrent neural network (RNN) [10]. The ability to uncover temporal dynamic behavior is one benefit of RRN. By

including a recurrent hidden layer that captures correlations across several time scales, RNN can manage a temporal sequential data [11]. RNN can retain sequential data from earlier input in its internal memory which is helpful in fields where weight and bias of previous input is needed like natural language processing, machine translation and speech recognition. RNN are composed of numerous hidden layers containing its own individual bias and weight, much like in neural networks. The connections among layers are in a direct cycle graph which occur sequentially in RNN.

B. Generation of Deepfakes

Most of the time Generative adversarial network (GAN)which are a type of deep neural network, which is used to create Deepfakes [12].

GAN comprised up of 2 network networks: encoder and decoder, which is collectively called an auto-encoder. An encoder network and a decoder network are typically chainedtogether to form an auto-encoder. Depending on the output size, intended training duration, anticipated quality, and resource availability, their design might change. The parallel training of two networks is the central concept.

A neural network that has been trained to duplicate its input to its output is called an Auto-encoder [13]. Deepfakes can be produced at the maximum degree of automatic encoding, where data is compressed by an encoder as it is processed, such as image data. This compression's goals are to lessen computing complexity and diminish the impact of data noise.

On the other hand, the original picture can be restored or at least approximately by passing the image's compressed form using a decoder. Therefore, for this model to produce faces that appear realistic, a lot of data are needed.

To create a Deepfake that blends Image A and Image B. So,when both the images are being trained on an auto-encoder for different dataset. The auto-encoder is allowed to share weights while maintaining separate decoders. In this manner, Image A can be compressed using logic, considering many factors like illumination, position of face and facial features of Image A. However, this will be carried out in accordance with the logic of Image B when it is restored, which has the effects of overlaying Image B's distinctive style on Image A.

So, false sample is being created by encoder after getting therandom input data. And decoder acts as a simple binary classifier which accepts both genuine and false samples as input from encoder and uses SoftMax function on to separate genuine from false data. GAN's benefit is its ability to produce a small set of sample data which will have the sameattribute and feature as that of training data. Therefore, GAN can be used for swipe genuine face of a person with a fake face of another person [14].

Fig.1.1. Image A (source)

Fig.1.2. Image B (Target)

Fig.1.3.DeepFake (Reenactment)

There are numerous deepfake applications for deepfake production like FakeApp [15]. This technique uses auto-encoder to swap faces on videos which was being created bya reddit user [16]. The false video created by this technique are quite convincing and difficult for viewers to distinguish from genuine video as the encoder builds latent feature of human face and decoder extracting those features from the same images.

CycleGAN is a based on GAN model which uses cycle loss function to teach latent characteristics [17]. It takes the traits from one picture to create identical traits in a new image. In contrast to FakeApp, it uses an unsupervised approach capable of doing image-to-image conversion without the useof paired samples.

Another well-known deepfake method based on GAN isVGGFace [18]. Traditional GAN is been improved further byadding Adversarial loss layer and perceptual layer which improves VGGFace design even further. It can even capture eye movement aspect of facial photo for making the false pictures more convincing.

C. Detection of Deepfakes

Forensic techniques for images may be applied to find Deepfake videos. There are many methods of detecting deepfakes. Hand-crafted feature-based detection is one of thetechnique which heavily rely on identifiable tampering traces[19]. Some of the distinctive picture feature information is frequently left during generation and post-processing of the video. Statistical and frequency domain characteristics with noticeable variations in pictures, like optical flow, photo response non-uniformity (PRNU) and image quality evaluation, etc., are extracted to achieve detection [20]. Edge features were employed by Faten et al. [21] for detection, and knowledge about picture texture features was collected usinga variety of image feature point detectors. The features were then transmitted to an SVM model. The detection model's accuracy mostly on limited dataset UADFV is around 90%. According to Frank et al. [22] PRNU can successfully identify Deepfake films on limited datasets. The detection performance are just not very compelling when the PRNU feature-based detection model is evaluated on the GAN- generated picture dataset. Researchers must manually extract feature data from such detection algorithms using more effective facial image discrimination, then input the acquiredfeature information first train the classifier, and then obtain adetection model.

Lugstein et al. [23] suggested using PRNU to identify Deepfake film, and they had successful detection outcomes. Koopman et al. [24] took the initiative in suggesting toemploy PRNU to detect Deepfake films due to the huge difference in the generating process between Deepfake and actual videos. Each frame's face was cropped, divided into eight sections, and the mean PRNU value for each group wasdetermined. The trial findings revealed a considerable difference between the genuine video as well as the Deepfakevideo's average normalized correlation coefficient.

There are many hybrid techniques in conjunction with standard deepfake detection algorithms to efficiently identifyfalse photos [25]. GoogleNet is also been utilized with face classification stream to train model on manipulated and real photos [26]. Following that, aspects from patch triplet streamare analyzed by utilizing low level camera features, local noise residual and steganalysis feature extractor. Results indicates that the method can learn both false as well as genuine visuals.

For the purpose of identifying bogus GAN films, Tariq et al. [27] presented neural network-based techniques. This technology makes use of pre-processing methods to assess a picture's statistical components and improve the identification of artificially created unreal faces.

There are strategies that employs paired learning for the identification of deep fake images was described. Method starts by generating a false picture using GAN. The difference in data between false picture and genuine image is then captured using a pairwise-learning model on the well-known network (CFFN) produced by GAN [28]. The assessment findings demonstrate that this strategy can improve upon thedrawbacks of the current time's false picture detectors.

To address the issue of high-quality false video detection, Hu et al. [29] presented a framework known as Flnfer (frame inference-based detection framework). It estimates losses that are been created to maximize the capacity to distinguish between false and genuine films by forecasting facial representation of the next upcoming frame from the given facial representation of present frame.

For the purpose of identifying fraudulent images produced by GANs, Nhu et al. [30] also offer another method according to CNN. In order to obtain facial characteristics from facerecognition networks, the model uses a deep learning network. The facial characteristics are then adjusted to make them suitable for both genuine and false picture recognition. Using this method, the validation dataset's data produces results that are satisfactory.

The concept of Spatial Pyramid Pooling layer was being introduced by Li et al. [31] and it was using ResNet-50.

Networks named as Meso-4 and MesoInception-4 are produced by Afchar et al. [32] Image's mesoscopic characteristics are used to conduct deep fake detection. A promising result is being achieved using the author's own deepfake dataset.

Xception network was been utilized Rossler et al. [33] on theFF++ dataset to identify deepfake films.

Therefore, even though the forensics system based on conventional picture characteristics is quite developed and capable of producing strong detection results, it produces subpar Deepfake films when dealing with certain. The original visual characteristics would be lost when the fake material is processed in the presence of noise, blurring and compression. The identification rate will decrease since it is simple to circumvent the detection model built using these characteristics.

PROBLEM AND UNRESOLVED ISSUES

Many new sorts of created images and videos might not be recognized by the existing learning models as a result of quick growth of new GAN models. Thus, the aforementioneddifficulties might demonstrate the huge need for developing reliable and robust learning models to identify false photos and videos.

There is a significant amount of false photos and films are produced daily as a result of widespread accessibility of internet, tools and apps for their creation and spreading [34].Deepfake pictures and films provide a significant problem foracademic scholars who examine and analyze them. Scalability is a problem for the learning techniques now in use. The absence of good datasets is one of the biggest problems the researchers are experiencing. To put it another way, face alteration is detected by deepfake algorithms usingfragmented data sets [35]. But, using these methods on big datasets can provide in unacceptably bad outcomes.

Future research must concentrate on developing models that are more reliable and robust so that they may be used with any sizeable, strong dataset. It is certain that deep learning models frequently necessitate big datasets for the training the model in order to yield decent results. But unfortunately, many datasets are either not publicly available or requires approval from different website and social media sources.

METHODOLOGY

This segment reflects on setup and processes. The method called deepfake tries to swap out a target's face in any media with another person's face. The parallel training of two auto encoders is the central concept. Depending on the output size, intended training duration, anticipated quality, and resource availability, their design might change. An encoder network and a decoder network are typically chained together to forman auto-encoder.

In the project, Meso-4 model is being used on Deepfake dataset. A neural network that has been trained to duplicate its input to its output is called an Auto-encoder. Deepfakes can be generated at highest level auto-encoder work like, when the data are Data is compressed by an encoder when it is processed, like picture data. This compression's goals are to lessen computing complexity and diminish the impact of data distortion. On the other hand, by running the picture's decoder through its compressed form, the actual picture can be roughly recreated.

A. Dataset

The Deepfake dataset was used to train and evaluate the models. Deepfake dataset was created instead of creating fake entirely from scratch, which might restrict the quantity and variety of the false data that is then feed to the Mesonet. It is chosen to extract face image from deepfake videos that already exist. Approximately 175 preexisting videos were taken from well-known deepfake platforms to create the deepfake dataset. It is explained that specific frames with faces are being retrieved from deepfake films. It should also be noted that comparable method is being utilized to extract authentic visual information from authentic video sources, such as movies and television broadcasts. It is also explained that the Data is stratified to evenly divide the different clarity levels and face angles between the real and false datasets [36]. Deepfake dataset is consisting 7104 images which belongs to2 classes. Randomly 80% of the dataset has been chosen for training and 20% for validation.

B. Proposed method

The proposed architecture identifies forged/altered image of face by establishing a technique at mesoscopic level of analysis. Image noise-based microscopic investigations can't be used in compressed formats since the picture noise is severely damaged. Human eye fails to differentiate fake pictures at a higher semantic level, especially when the image represents a face. As a result, deploying a deep neural network with a minimal number of layers in an intermediate method. This began with more sophisticated structures and gradually reduced them till reaching the final result, which is more effective. With a negligible degree of depiction and an unexpectedly small number of parameters, this design earnedthe best classification score in the test. It is built on high- performing image classification networks with convolutionaland pooling layers. Dense network for feature extraction andcategorization. The network begins with 4 layers of pooling and repeated convolutions, then moves on to a dense network with one hidden layer. The fully-connected layers employ Dropout to regularize and increase their resilience, while ReLU activation functions in the convolutional layers produce non-linearities, and batch normalization regulates itsoutput to avoid the vanishing gradient effect. Along with L2regularization in its dense layers. It leads the weights to decaytowards zero, but not completely zero (weight decay). As a result, it is very helpful when condensing the model and prevent the risk of overfitting in a model.

C. Meso-4

A CNN called Meso-4 has 4 convolutional blocks and 1 fullylinked hidden layer. Convolutional block always include a convolutional layer and a max pooling layer. In Mesonet, these convolutional also includes a batch normalization layer.

The convolutional layer needed to set the size and number offilters/kernel. Each filter/kernel represents a distinct image feature like horizontal line, etc. This filter is used to an image during convolution to determine how closely certain areas of the image match the filter. Calculating the dot product of the filter/kernel with each image region's filter size for each color channel is how it is done. As a result, the filter or kernel recognizes the presence and placement of particular visual elements like horizontal and vertical lines.

Batch normalization is applied following the convolutional layer. The neural network can be made faster, more effective, and more stable by using batch normalization. By adjusting the input to every network layer, it lowers the dependency between the characteristic for any specific layer and thedispersion of input for such subsequent layer. Internal covariate shift is the term for this dependency, which has a destabilizing influence on the learning process.

The last layer of convolution block is pooling layer. It is in the pooling layer where the dimensionality of the data is decreased significantly, which increases the speed upcomputation.

D. Mesonet uses max pooling for the layer, which means the region of pixel value is reduced to that region’s max value. *System setup*

Models are created utilizing Python 3.6 programming language and with Keras 2.1.5 APIs and various Python libraries. The tests were performed on a Windows 10 machinewith the specified hardware setup: AMD Radeon (TM) R7 M360 GPU with 4GB GDDR3 900 MHz memory and Intel® CoreTM i5-6200 U CPU @ 2.30 GHz Dual Core processor with 8.0 GB (RAM).

E. System setup

Models are created utilizing Python 3.6 programming language and with Keras 2.1.5 APIs and various Python libraries. The tests were performed on a Windows 10 machinewith the specified hardware setup: AMD Radeon (TM) R7 M360 GPU with 4GB GDDR3 900 MHz memory and Intel® CoreTM i5-6200 U CPU @ 2.30 GHz Dual Core processor with 8.0 GB (RAM).

F. System setup

Models are created utilizing Python 3.6 programming language and with Keras 2.1.5 APIs and various Python libraries. The tests were performed on a Windows 10 machinewith the specified hardware setup: AMD Radeon (TM) R7 M360 GPU with 4GB GDDR3 900 MHz memory and Intel® CoreTM i5-6200 U CPU @ 2.30 GHz Dual Core processor with 8.0 GB (RAM).

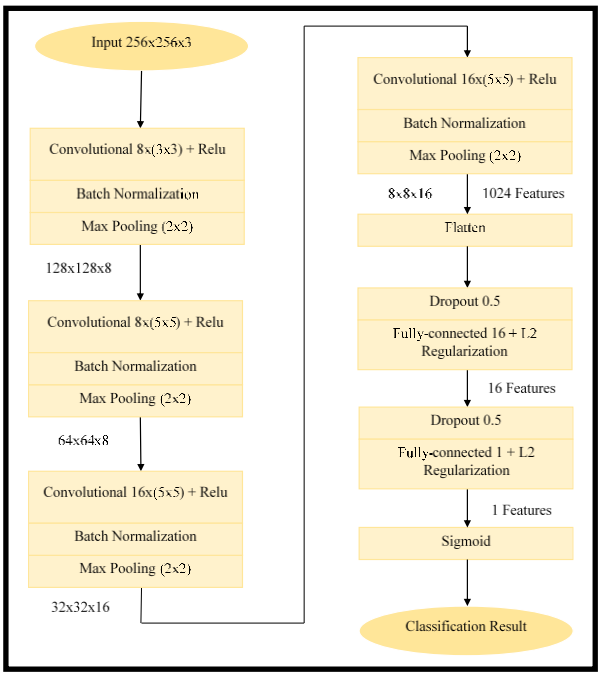

Figure: 1.4. Meso-4 Architecture

RESULT

The trained network's classification accuracy score on the Deepfake dataset is 88.8%.

1) Accuracy:

Accuracy is a statistic that expresses the proportion of accurate predictions.

Accuracy Score = (True Positive + True Negative) / (TruePositive + True Negative + False Positive + False Negative)

Table I. Error Matrix

True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative(FN)
3745	2564	281	514

unbalanced data since it can't discriminate between differentsorts of mistakes (false positives and false negatives)

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

Precision, recall and F1 score for the trained network are shown in Table II for the Deepfake dataset.

Table II. Performance Matrix

Network	Precision	Recall	F1-Score
Meso-4	0.93	0.87	0.90

B. Area under curve (AUC)

The AUC, which assesses a classifier's capacity to distinguish between classifications, is a summary of the ROC curve. The AUC demonstrates the model's capability to distinguish between positive and negative classifications. The better will be the model's performance, the higher the AUC value.

AUC is a composite measure of efficiency that considers all possible categorizations levels. It refers to a likelihood curve that contrasts the TPR and FPR at various threshold values to distinguish the "noise" from the "signal". The model managed to obtain a 96.2% score on Deepfake dataset as shown in Fig.1.5. Face manipulation is well-known for its risks. A network architecture has been provided to identify such forgeries quickly and at a cheap computational cost. The approach has a precision of 93% and recall of 87% forDeepfake for real-world internet dispersion settings, according to an experiment.

For comparison, few of the deepfake detection techniques arebeing selected based on AUC as shown in Table 1.6.

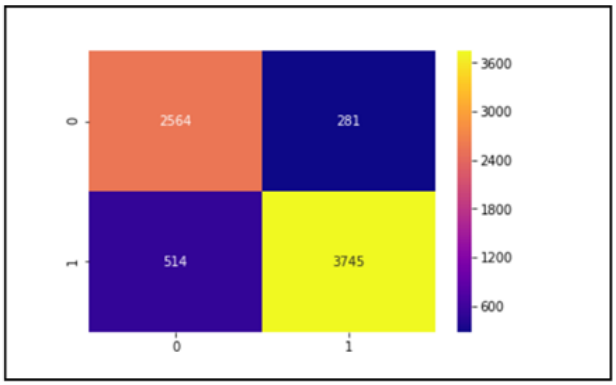

Figure: 1.5. Confusion Matrix

2) Precision:

Precision is an easy method to attain accuracy is to make onepositive prediction and ensure it is correct. This would be unsuccessful since the classifier would only keep one of the good examples.

Precision = (True Positive) / (True Positive + False Positive)

3) Recall:

The recall is a metric used to measure how accurately a modelcan find samples were positive. The bigger the recall, the more positive samples were discovered. Recall is calculated as the proportion of Positive samples that were correctly classified as Positive to all Positive samples.

Recall = (True Positive) / (True Positive + False Negative)

4) F1 score:

F1 score is simply counts the number of correct predictions generated by a machine learning model. As you've seen, accuracy is a poor statistic to employ when dealing with

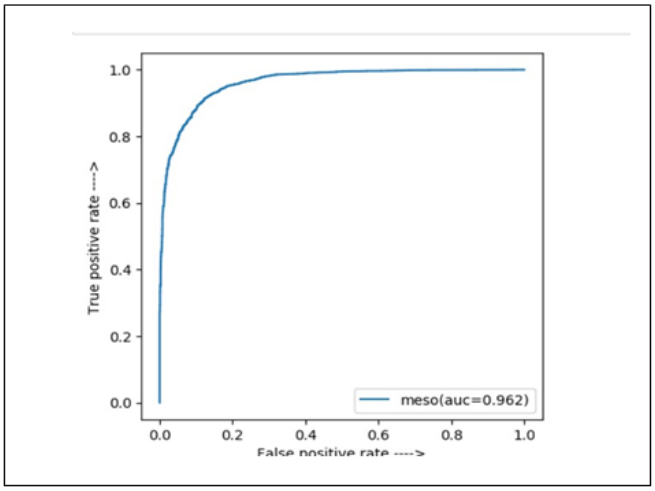

Figure: 1.6. Area under the curve

Face manipulation is well-known for its risks. A network architecture has been provided to identify such forgeriesquickly and at a cheap computational cost. The approach hasa precision of 93% and recall of 87% for Deepfake for real- world internet dispersion settings, according to an experiment.

For comparison, few of the deepfake detection techniques arebeing selected based on AUC as shown in Table III.

Table III: **Area under Curve Performance**

Model	AUC (%)
Meso4	96.2
MesoInception	80.69
Xception	79.57
FWA	71.88
Multi-task	68.25
EfficientNet-B5 [37]	74.9

5) Proposed model:

In comparison with the Meso-4 architecture developed by Afchar et al. [32] uses CNN made up of 4 convolutional blocks. While the proposed model is also based on Meso4 architecture along with L2 regularization in its dense layers. L2 regularization leads the weights to decay towards zero, but not completely zero (weight decay). As a result, it is very helpful when condensing the model and prevent the risk of overfitting in a model. This strategy inhibits learning a more sophisticated model and thus preventing overfitting.

The result showed that Afchar et al.'s model achieved a training accuracy of 95.4% and Validation accuracy of 79.6%. On the other hand, proposed model achieved a training accuracy of 96.2% and validation accuracy of 83.9%. These results indicates that the proposed model outperforms Afchar et al.'s model in terms of accuracy.

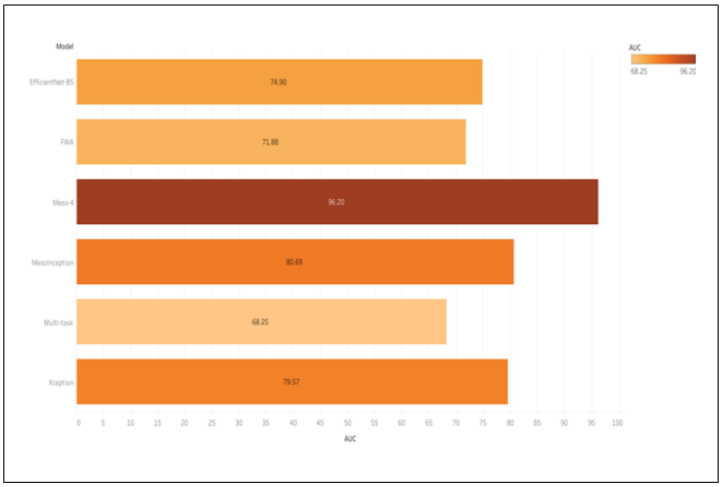

Figure: 1.7.Area Under Curve Performance

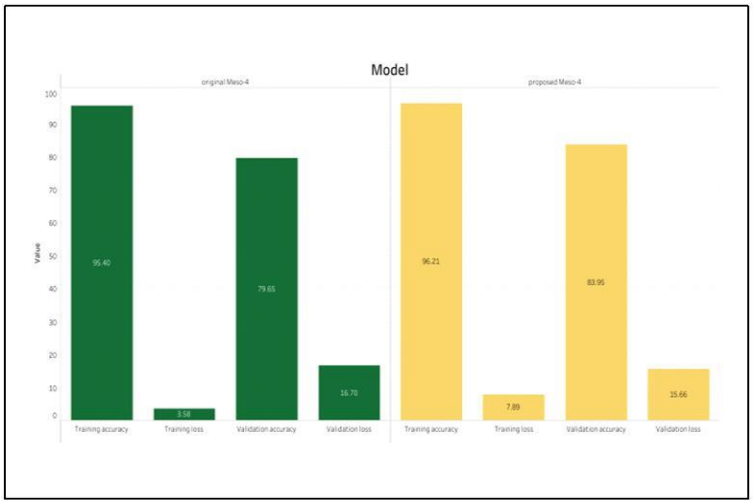

Figure: 1.8.Meso-4 Comparison

1. *MesoInception4:*

It is unable to apply microscopic investigations based on picture noise. The model is very similar to proposed Meso4 model, but the first 2 layers of the network is being replacedby extended convolution to process multiscale data [32].

Xception:

This model uses visual features to generate image sequencesfrom RGB data which is quite good at classifying images. In order to create a picture classification of the main class of a picture, XceptionNet focuses on high level aspects rather than subtler features. The network replaces the fullyconnected layer with 2 outputs which is based on separated convolutions with feedback network. It is to improve its performance and requires more domain-specific information[38].

FWA:

This algorithm fit to varied image resolution, different processing steps and will provide result in form of certain unique artifacts. These can be used for detecting of forged images [39].

1) Multi-task:

This method employs the concept of shared weights in auto- encoder to find tampered area and identify false images. Minimizing loss of each component, that communicate the data among each other and performance can be enhanced [40].

Therefore, from the above analysis performed, it is clear that Mesonet (Meso-4) algorithm is best suited for detecting the Deepfakes with the obtained maximum accuracy of training data as 88.8%. The presented method is able to detect the Deepfake image by taking the facial features.

Table IV: Meso-4 Comparison

Model	Accuracy and Loss (%)			
	Train accuracy	*Validation accuracy*	*Training loss*	*Validation loss*
Meso-4	95.40	79.65	03.58	16.70
Proposed model	96.21	83.95	07.89	15.66

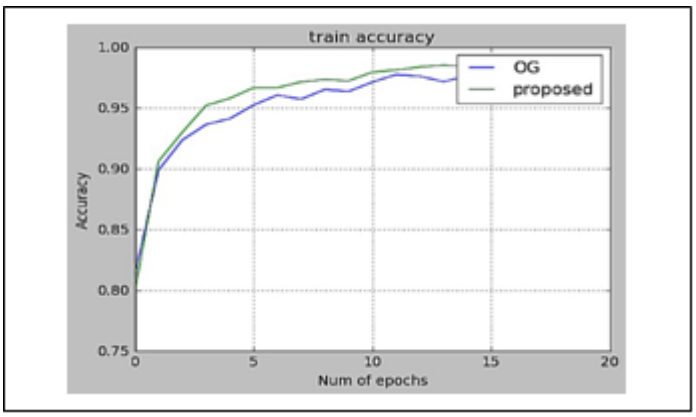

Figure: 1.9.Training Accuracy Comparison

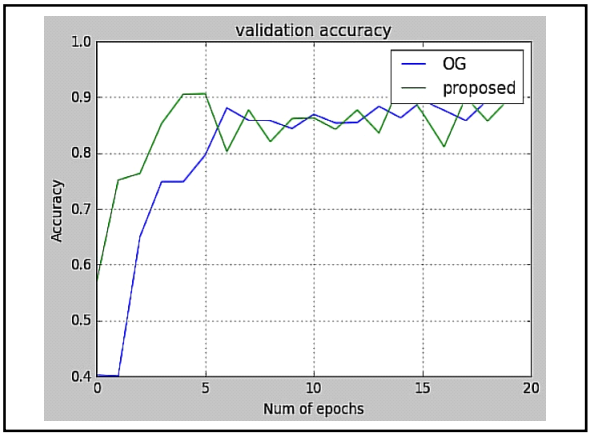

Figure: 1.10.Validation Accuracy Comparison

CONCLUSIONS

Face manipulation is well-known for its risks. This is all dueto extensive accessibility of photographs and films in various websites, social media material, making Deepfake grown in popularity. It is especially crucial now since social networking sites make it simple for users to spread and circulate such false information. And since, deepfake-making tools are much more widely available. Numerous deep learning-based approaches have lately been put out to deal with problem like effectively identify phoney photos and films.

A thorough explanation of various tools, models and effectiveness of present day deepfake approaches is been provided. Touching some current programs and equipment that are been frequently utilized production of false photos and films. Covering the present issues and given some suggestions for next deep learning researcher working on these detection issues.

Recently, the complexity and accuracy of deepfakes has been raising despite of learning method's impressive results for their detection. There seems to be no obvious way to determine the proper/fixed architectural structure or layer count for detection model using the existing learning approaches.

It is being figured that, among other things, that the eyes andlips plays a crucial part in identifying tampered faces. In order to construct deeper networks that are more effective and efficient, it is expected that more technologies will develop in the future. Field of research needs to be improved to betterequip social media platform to deal with ubiquitous effects of false and mitigate their effects by integrating deepfake detection technologies. In order to properly recognize false films and photos, the present deep learning techniques must also be improved.

REFERENCES

1. *M. Westerlund, "The Emergence of Deepfake Technology: A Review," Technology Innovation Management Review, vol. 9, pp. 40-53, 2019.*
2. *a. C. A. Vaccari, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News," Social Media + Society, no. 6, pp. 1-13, 19 February 2020.*
3. *a. K. S. Kwok, "Deepfake: A Social Construction of Technology Perspective. Current Issues in Tourism," pp. 1-5, 2020.*
4. *Piva, "An Overview on Image Forensics," International Scholarly Research Notices, vol. 2013, p. 22, 2013.*
5. *S. H. a. S. S. J. Kim, "Classifying Genuine Face images from Disguised Face Images," 2019 IEEE International Conference on Big Data (Big Data), pp. 6248-6250, 2019.*
6. *O. M. A. A. L. a. A. S. S. A. A. Maksutov, "Methods of Deepfake Detection Based on Machine Learning," 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), pp. 408-411, 2020.*
7. *H. Farid, "Image forgery detection," IEEE Signal Processing Magazine, vol. 26, no. 2, pp. 16-25, 2009.*
8. *O. W. R. Z. A. O. A. A. E. Sheng-Yu Wang, "CNN-Generated Images Are Surprisingly Easy to Spot... for Now," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, no. 8695-8704, pp. 13-19, 2020.*
9. *G. a. Y. B. a. A. Courville, "Deep Learning," no. 2, 2016.*
10. *Y. S. P. a. F. P. Bengio, "Learning Long-Term Dependencies with Gradient Descent Is Difficult," IEEE Transactions on Neural Networks, no. 5, pp. 157-166, 1994.*

11. *L. Elman, "Finding Structure in Time," Cognitive Science, vol. 14, no. 2, pp. 179-211, 1990.*
12. *Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets," Neural Information Processing Systems, no. 27, p. 2672–2680, 2014.*
13. *T. M. M. S. C. A. F. J. H. B. A. K. R.-C. B. S. M. Lakshmanan Nataraj, "Detecting GAN generated Fake Images using Co-occurrence Matrices," Electronic Imaging, pp. 532-1-532-7, 2019.*
14. *M. a. O. S. Mirza, "Conditional Generative Adversarial Nets," 2014.*
15. *"FakeApp 2.2.0.," [Online]. Available: https://www.malavida.com/en/soft/fakeapp.*
16. *"Faceswap: Deepfakes Software for All.," [Online]. Available: https://github.com/deepfakes/faceswap.*
17. *"CycleGAN," [Online]. Available: https://junyanz.github.io/CycleGAN/.*
18. *"Keras-VGGFace: VGGFace Implementation with Keras Framework.," [Online]. Available: https://github.com/rcmalli/keras-vggface.*
19. *G. L. C. R. D. B. A. Amerini, " Deepfake video detection through optical flow based cnn," IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Korea, pp. 27-28, 2019.*
20. *F. J. G. M. Lukáš, "Detecting digital image forgeries using sensor pattern noise," Security, Steganography, and Watermarking of Multimedia Contents VIII, San Jose, CA, USA, vol. 6072, p. 362–372, 2006.*
21. *Kharbat, T. Elamsy, A. Mahmoud and R. Abdullah, "Image feature detectors for deepfake video detection," e 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, pp. 1-4, 2019.*
22. *Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa and T. Holz, "Leveraging frequency analysis for deep fake image recognition," International Conference on Machine Learning (PMLR), Virtual, p. 3247–3258, 2020.*
23. *a. B. S. a. B. G. a. U. A. Lugstein, "PRNU-based Deepfake Detection," 2021 ACM Workshop on Information Hiding and Multimedia Security, Virtual, p. 7–12, 2021.*
24. *Koopman, A. Rodriguez and Z. Geradts, "Detection of deepfake video manipulation," 20th Irish machine vision and image processing conference (IMVIP), Ulster University, Ulster, Northern Ireland, p. 29–31, 2018.*
25. *J. L. a. T. X. Liu, "Task-Oriented GAN for PolSAR Image Classification and Clustering.," IEEE Transactions on Neural Networks and Learning Systems, no. 30, pp. 2707-2719, 2019.*
26. *V. V. I. S. S. J. a. W. Z. Szegedy, "Rethinking the Inception Architecture for Computer Vision," IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, pp. 2818-2826, 2016.*
27. *S. L. S. K. H. S. Y. a. W. S. Tariq, "Detecting Both Machine and Human Created Fake Face Images in the Wild," 2nd International Workshop on Multimedia Privacy and Security, Toronto, pp. 81-87, 2018.*
28. *C.-C. Z. Y.-X. a. L. C.-Y. Hsu, "Deep Fake Image Detection Based on Pairwise Learning," Applied Sciences, no. 10, p. 370, 2020.*
29. *Hu, X. Liao, J. Liang, W. Zhou and Z. Qin, "FInfer: Frame Inference-based Deepfake Detection for High-Visual-Quality Videos," IEEE Trans. Pattern Anal. Mach. Intell, pp. 1-9, 2022.*
30. *N.-T. N. I.-S. a. K. S.-H. Do, "Forensics Face Detection from GANS Using Convolutional Neural Network," ISITC, 2018.*
31. *Y. Li, X. Yang, P. Sun, H. Qi and S. Lyu, " Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics,"*

IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, p. 3207–3216, 2020.

32. *Afchar, V. Nozick, J. Yamagishi and I. Echizen, "Mesonet: A compact facial video forgery detection network," 2018 IEEE international workshop on information forensics and security (WIFS), Hong Kong, China, p. 1–7, 2018.*
33. *Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," IEEE International Conference on Computer Vision, Seoul, Korea, pp. 1-11, 2019.*
34. *"Photo tampering throughout history," 2012. [Online]. Available: http://www.fourandsix.com/photo-tampering-history/.*
35. *T. N. C. N. D. N. D. a. N. S. Nguyen, "Deep Learning for Deepfakes Creation and Detection," vol. 1, 2019.*
36. *V. N. J. Y. I. E. Darius Afchar, "MesoNet: a Compact Facial Video Forgery Detection Network," 2018 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1-7, 2018.*
37. *P. &. Egorov, "EfficientNets for DeepFake Detection: Comparison of Pretrained Models," 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), 26 01 2021.*
38. *Chollet, "Xception: Deep Learning With Depthwise Separable Convolutions," IEEE Conference on Computer Vision, 2017.*
39. *R. M. S. F. Matern, "Exploiting visual artifacts to expose deepfakes and face manipulations," 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 83-92, 2019.*
40. *F. J. Y. I. E. H. H. Nguyen, "Multi-task learning for detecting and segmenting manipulated facial images and videos".*
41. *W. T. Y. a. G. F. X. Chang, "DeepFake Face Image Detection based on Improved VGG Convolutional Neural Network," 2020 39th Chinese Control Conference (CCC), pp. 7252-7256, 2020.*
42. *S. C. S. E. a. M. S. K. T. Van Lanh, "A survey on digital camera image forensic methods," Proceedings of the IEEE International Conference onMultimedia and Expo (ICME '07), pp. 16-19, 2007.*
43. *J. K. V. M. P. a. F. K. S. Suratkar, "Employing Transfer-Learning based CNN architectures to Enhance the Generalizability of Deepfake Detection," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-9, 2020.*
44. *R. S. Khalaf and A. Varol, "Digital Forensics: Focusing on Image Forensics," 2019 7th International Symposium on Digital Forensics and Security (ISDFS), 2019.*
45. *Farid, "Digital doctoring: how to tell the real from the fake," Significance, vol. 3, no. 4, pp. 162-166, 2006.*
46. *C.-Y. L. Y.-X. Z. Chih-Chung Hsu, "Learning to Detect Fake Face Images in the Wild," IEEE International Symposium on Computer, Consumer and Control (IS3C), pp. 388-391, 2018.*
47. *L. B. T. S. a. H. J. Li, "Detection of Deep Network Generated Images Using Disparities in Color Components," 2018.*
48. *P. H. X. M. V. a. D. L. Zhou, "Two-Stream Neural Networks for Tampered Face Detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, pp. 1831-1839, 2017.*