SYNTHETIC MEDIA DETECTION

[1] Guide - Mrs. N. Swapna Goud (Assistant Professor)
[2] A. Manoj Kumar [3] G. Bhanu Prakash [4] K. Mithil Reddy

Department of CSE

ANURAG GROUP OF INSTITUTIONS

ABSTRACT

Deepfakes are images or videos that have been altered to feature someone else's face, similar to an advanced form of face-swapping. Although some deepfake videos are clearly doctored and inauthentic, the majority appear and sound convincingly real. Deepfake technology has also been used to aid in the spread of false news and other forms of deception. Deep fake detection is a method for detecting face tampering in videos that focuses on two recent techniques used to generate hyper realistic forged videos: Deepfake and Face2Face.

Traditional image forensics techniques are typically unsuitable for video due to the compression, which severely degrades the data. As a result, this paper takes a deep learning approach and presents two networks with a low number of layers to focus on image mesoscopic properties. We test those fast networks on an existing dataset as well as a dataset created from online videos.

INTRODUCTION

Today, the dangers of fake news are widely recognised, and in a world where more than 100 million hours of video content are watched daily on social media, the spread of falsified video raises increasing concerns. While image forgery detection has advanced significantly, digital video falsification detection remains a difficult task. Most methods used with images cannot be directly extended to videos, owing to the significant degradation of the frames after video compression.

Deepfake technology's rapid development has raised serious concerns about the authenticity and credibility of visual media. Deepfakes are computer-generated images or videos that have been altered to appear real. They are created with advanced machine learning algorithms that allow individuals to replace faces, manipulate voices, and change context, making traditional detection methods difficult to distinguish between real and fake media. Several deep learning techniques have been developed to detect deepfakes, with the Meso4 architecture being one of the most promising. As deepfakes become more sophisticated, the need for accurate and efficient detection methods has grown. Deepfakes are difficult to detect due to the sophistication of the algorithms used to create them. Recent advances in deep learning, on the other hand, have shown promising results in identifying and detecting deepfakes.

Meso4 is a lightweight neural network model that was created to detect deepfakes in images and videos. This

model detects the presence of deepfakes in a video or image by combining spatial and temporal features. A convolutional neural network (CNN) that can distinguish between real and fake media based on visual patterns is another popular approach to deepfake detection. These CNNs were trained on large datasets of both real and fake media and are extremely effective at detecting deepfakes. While deepfake detection techniques have yielded promising results, significant challenges remain. One of the primary challenges is a lack of large, diverse deepfake media datasets, which can limit the effectiveness of deep learning models. Furthermore, deepfakes are constantly evolving, and new algorithms are being developed to create more convincing fakes, necessitating the continuous development of new detection methods.

EXISTING SYSTEM

Deepfake Detection Challenge (DFDC)

The Deepfake Detection Challenge (DFDC) is a competition launched in 2019 by Facebook in collaboration with several academic institutions, including MIT, UC Berkeley, and the University of Oxford. The goal of the challenge is to develop and improve algorithms for detecting deepfake videos, which are videos that have been manipulated using artificial intelligence to make it appear as if someone said or did something they did not.The challenge is divided into two tracks: the unrestricted track, in which participants have access to the entire dataset, and the restricted track, in which participants have access to a smaller, more controlled subset of the dataset.

The challenge also includes a separate "audio deepfake" track, which focuses on detecting manipulated audio. The DFDC has attracted a wide range of participants, including academic researchers, industry professionals, and hobbyists, and has resulted in several state-of-the-art deepfake detection models. The challenge is ongoing, with new iterations and updates to the dataset planned for the future. The ultimate goal of the challenge is to improve the ability of automated systems to detect deepfake videos, which could help combat the spread of misinformation and protect the integrity of digital media.

XceptionNet

It is a deep learning convolutional neural network (CNN) architecture proposed by François Chollet in 2016. The architecture is inspired by Google's Inception architecture, but takes the concept of depthwise separable convolution to an extreme level. The main idea behind XceptionNet is to use depthwise separable convolutions in place of standard convolutions in order to reduce the number of trainable parameters and improve computational efficiency, while still maintaining high accuracy. The XceptionNet architecture consists of a series of modules, each of which contains several depthwise separable convolution layers, followed by a skip connection to preserve spatial information. The output of each module is fed into a global average pooling layer, followed by a fully connected layer and a softmax activation function to produce the final classification output.

XceptionNet has achieved state-of-the-art performance on a variety of image classification benchmarks, including the ImageNet dataset, and has been used as a backbone architecture in many other computer vision tasks such as object detection and segmentation.

PROPOSED SYSTEM

DEEPFAKE DETECTION WITH MESONET

We propose detecting forged face videos at the mesoscopic level of analysis. Indeed, image noise-based microscopic analyses cannot be used in a compressed video context because the image noise is severely degraded. At a higher semantic level, the human eye has difficulty distinguishing forged images, particularly when the image depicts a human face. As a result, we propose an intermediate approach based on a deep neural network with a few layers. Mesonet is a deep learning-based model that can be used for deepfake detection. Mesonet stands for multi-scale encoder for video-depth prediction, which is a type of neural network architecture that can be used for detecting deepfake videos.

The Mesonet model is designed to learn how to distinguish between real and fake videos by analyzing the differences in patterns and features between them. The model is trained using a large dataset of real and fake videos, and it learns to identify the specific characteristics that are unique to deepfake videos.One of the key advantages of Mesonet is that it uses a multi-scale approach, which means that it analyzes the video at different levels of detail. This allows it to detect deepfake videos that may have been created using different techniques or at different resolutions.

This network starts with four layers of successive convolutions and pooling, then moves on to a dense network with one hidden layer. Convolutional layers use ReLU activation functions to introduce nonlinearities and Batch Normalization to regularise their output and prevent the vanishing gradient effect, while fullyconnected layers use Dropout to regularise and improve their robustness.

METHODOLOGY

Image and video analysis is one of the most common methodologies used for deepfake detection. This methodology involves analyzing the input image or video to identify visual artifacts that are indicative of deepfakes, such as inconsistent lighting, unnatural facial movements, or inconsistent shadows. Image and video analysis can be performed manually by human experts or using automated tools, such as computer vision algorithms. In this article, we will discuss the various methodologies used for deepfake detection through image and video analysis.

1. **Facial Landmarks Detection**

   Facial landmarks detection is a popular technique used for deepfake detection. The methodology involves analyzing facial landmarks, such as the position of the eyes, nose, and mouth, to detect any inconsistencies in the movement and alignment of the facial features. The technique can be performed using deep neural networks or other computer vision algorithms. Facial landmarks detection is effective in identifying deepfakes that have been generated using a face-swapping technique.

2. Texture Analysis

Texture analysis is another technique used for deepfake detection. The methodology involves analyzing the texture of the image or video to detect any inconsistencies in the image. Deepfakes can often be identified by analyzing the texture of the image, as the texture of a deepfake is often different from that of a real image. Texture analysis can be performed using various algorithms, such as local binary patterns, wavelet transforms, and convolutional neural networks.

3. Lighting Analysis

Lighting analysis is a technique used for deepfake detection that involves analyzing the lighting in the image or video to detect any inconsistencies. Deepfakes often have inconsistencies in lighting due to the difficulty of simulating lighting conditions in a realistic way. Lighting analysis can be performed using various algorithms, such as histogram equalization, color space analysis, and Gaussian blur detection.

4. Motion Analysis

Motion analysis is another technique used for deepfake detection that involves analyzing the movement of the subject in the image or video to detect any inconsistencies. Deepfakes often have inconsistencies in motion due to the difficulty of simulating natural movement. Motion analysis can be performed using various algorithms, such as optical flow analysis, motion vectors analysis, and frame differencing.

5. Compression Analysis

Compression analysis is a technique used for deepfake detection that involves analyzing the compression artifacts in the image or video to detect any inconsistencies. Deepfakes often have inconsistencies in compression artifacts due to the difficulty of simulating the compression algorithm used in the original image or video. Compression analysis can be performed using various algorithms, such as Fourier transform, discrete cosine transform, and JPEG compression analysis.

EXPERMENTAL TOOLS

PyCharm

PyCharm is a Python integrated development environment (IDE) that provides a wide range of essential tools for Python developers. These tools are tightly knit together to create a working environment for productive Python, web, and data science development.

OpenCV

OpenCV is a massive open-source computer vision, machine learning, and image processing library. OpenCV is compatible with a wide range of programming languages, including Python, C++, and Java. It can analyze images and videos to recognize objects, faces, and even human handwriting. When it is combined with other libraries, such as NumPy, a highly optimized library for numerical operations, the number of weapons in your arsenal grows, as any operation that can be done in NumPy can be combined with OpenCV.

Tkinter

Tkinter is a Python module for creating graphical user interfaces. Because it is simple and easy to use, it is one of the most commonly used modules for creating GUI applications in Python. You do not need to install the Tkinter module separately because it is included with Python. It provides the Tk GUI toolkit with an object-oriented interface.

TensorFlow

TensorFlow is an open-source, machine learning framework created by the Google Brain team. It is designed to make it easier to build and train machine learning models, especially deep neural networks, by providing a flexible architecture that can be used for a wide range of applications. At its core, TensorFlow is based on the concept of a computational graph. Users define a set of operations that should be performed on data, and TensorFlow creates a graph of those operations. This graph can then be optimized and executed efficiently on a variety of hardware, including CPUs, GPUs, and even custom-built ASICs.

ImageDataGenerator

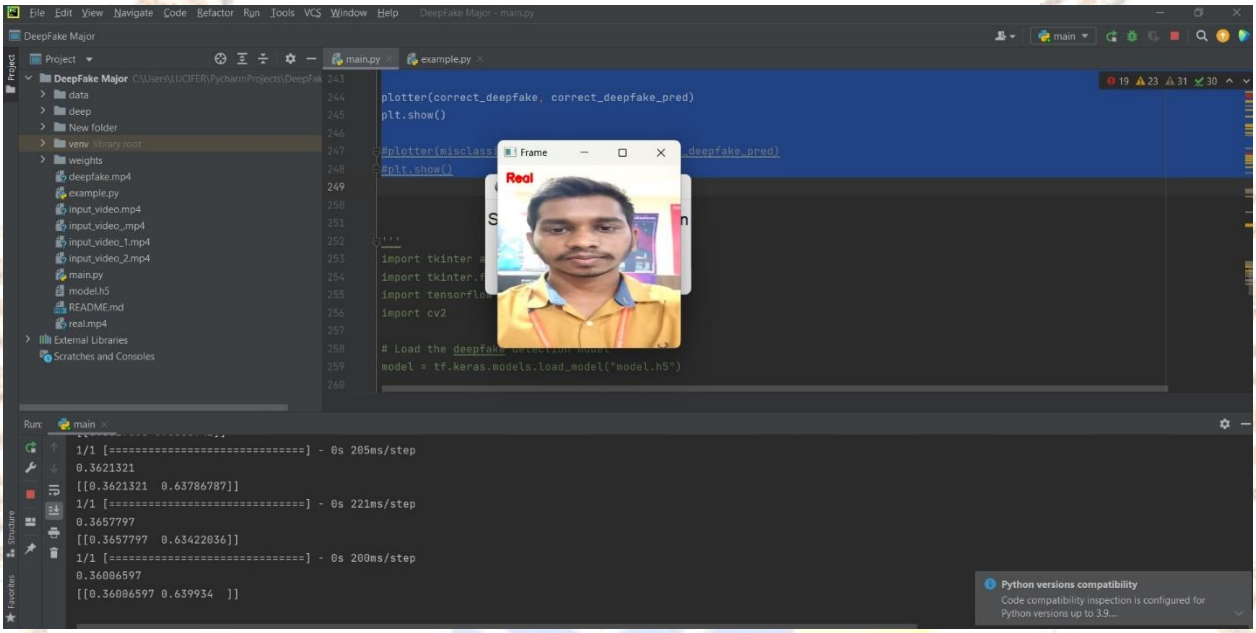
ImageDataGenerator generates augmented images on the fly by randomly transforming the input images. Rotation, shearing, zooming, flipping, and shifting are examples of transformations. It can also perform image normalisation, rescaling, and preprocessing operations on them. Using ImageDataGenerator entails specifying the image transformations desired and then fitting the generator to the training data. During training, the augmented images are generated and fed into the model. This allows the model to see many different versions of each training image, which can help it generalise to new data.

Adam

Adam (Adaptive Moment Estimation) is a well-known optimisation algorithm that is used in deep learning models to update the neural network weights during training. It is an adaptive learning rate optimisation algorithm that is an extension to stochastic gradient descent (SGD). Adam derives individual adaptive learning rates for various parameters from estimates of the gradient's first and second moments.

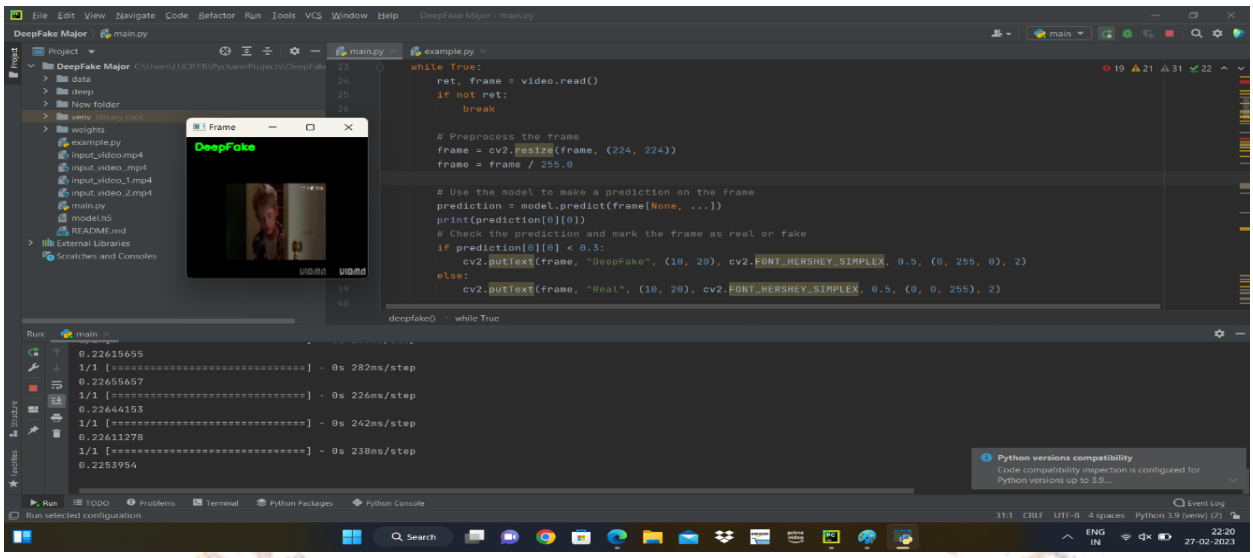
The algorithm keeps exponentially decaying moving averages of past gradients and squared gradients and uses them to adjust the learning rate for each weight based on their previous training behaviour. Adam is well-known for his performance in practise and has emerged as a popular choice for optimising deep neural networks. It is used in many popular deep learning libraries, including TensorFlow and Keras.

OUTPUT



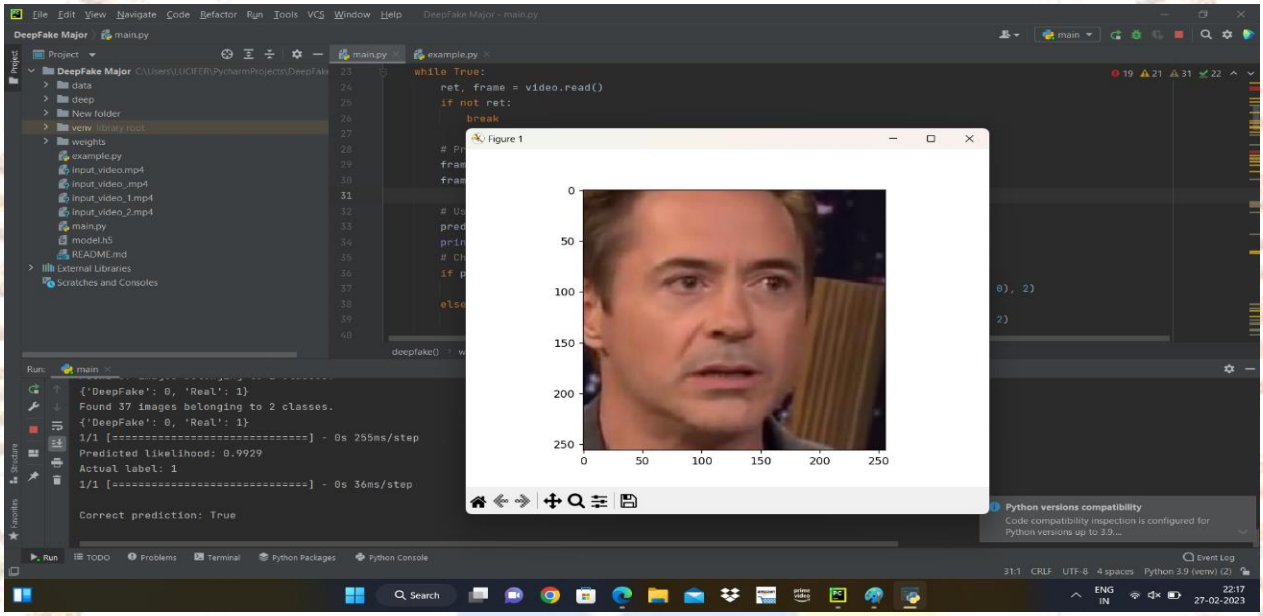
(Fig 7.1 Real Video)

Accurate detection of real videos. Our deepfake detection system utilizing Meso4 architecture has successfully identified this video as authentic, ensuring reliable results.



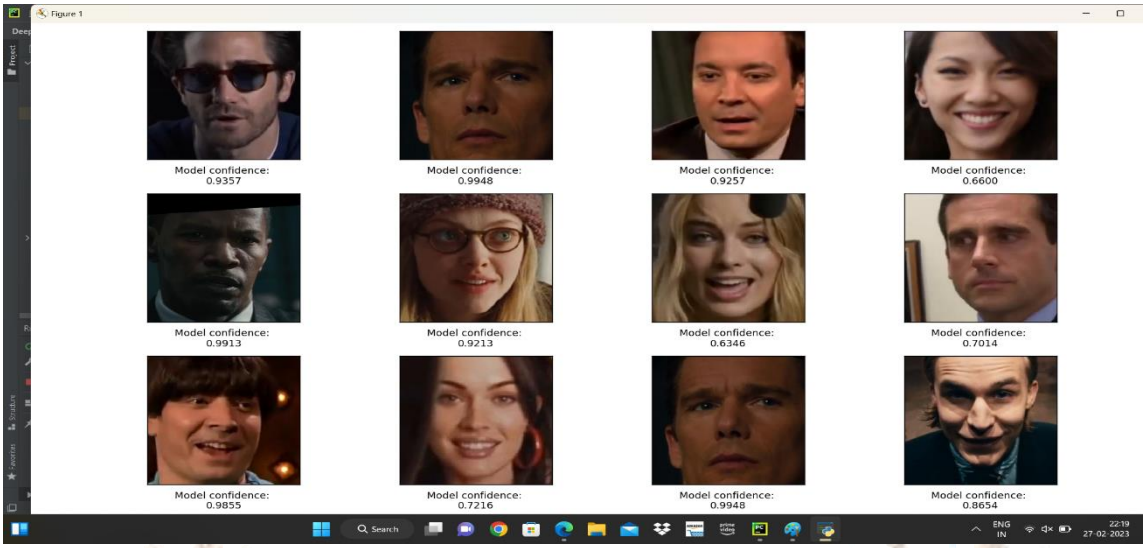
(Fig 7.2 Face swap video)

The above image shows the system detecting a face swap video and indicating it as a DeepFake.



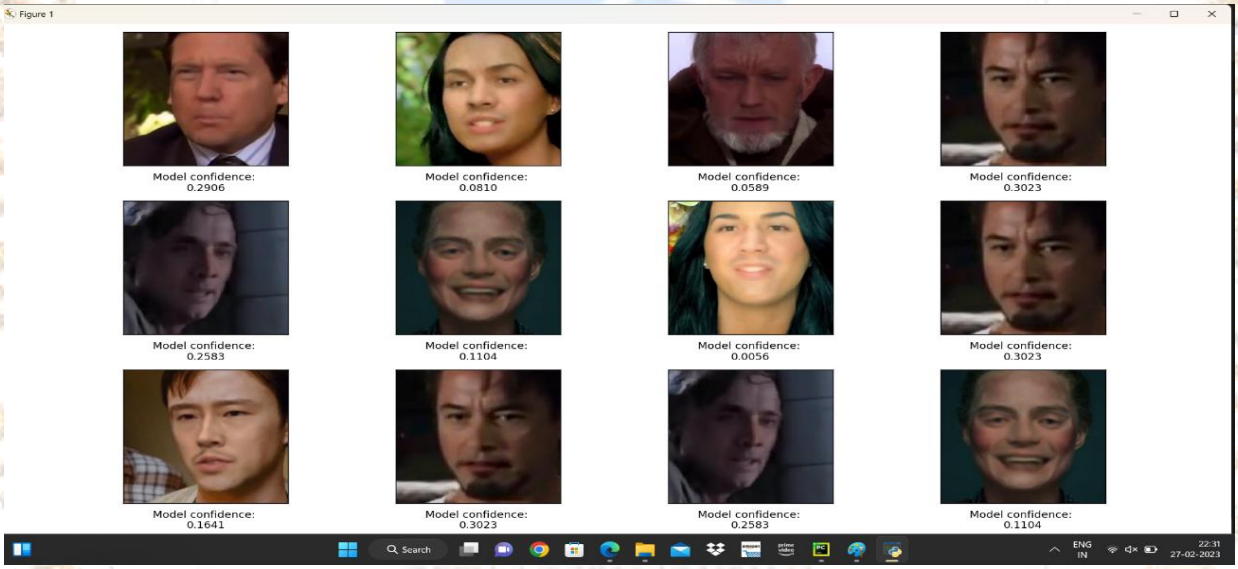
(Fig 7.3 Training Facial data)

The above image shows likelihood of the prediction and labelling of the image as real or deepfake are checked to see if they are correct.



(Fig 7.4 Analyzing images of real people faces)

The image sequence containing faces of persons has been analyzed and the result indicates that it is not a deepfake, but a real.



(Fig 7.5 Analyzing images of Face swap)

The image sequence containing faces of persons has been analyzed and the result indicates that it is a deepfake.

CONCLUSION

The dangers of video face tampering are well known these days. We propose a network architecture that could detect such forgeries efficiently and at a low computational cost. In addition, we provide access to a dataset devoted to the Deepfake approach, a popular but under-documented topic in our opinion. The ability to generate a solution to a given problem without prior theoretical study is a fundamental aspect of deep learning. However, understanding the origin of this solution is critical in order to evaluate its qualities and limitations, which is why we spent so much time visualising the layers and filters of our network.

Using a deep learning approach, the Meso4 provides a promising solution for detecting deepfake videos. The proposed network architectures detect deepfakes and face2face videos with high accuracy rates, which are commonly used for tampering and manipulation. In addition, the project provides a valuable dataset dedicated to deepfake videos, which can be used for further research and development in this area. The efforts to visualise their networks' layers and filters have shed light on the critical role of the eyes and mouth in detecting deepfakes, which can aid in the development of more effective and efficient deep learning models for this task. Overall, the Meso4 project is a significant step forward in improving the detection of deepfake videos, which is critical in ensuring the authenticity and integrity of content.

REFERENCE

[1]. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1-11).

[2]. Agarwal, S., Farid, H., Guo, Y., He, W., & Li, H. (2020). Detecting deepfake videos from audio-visual discrepancies using temporal consistency. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11488-11497).

[3]. Gafni, R., & Koppel, M. (2018). Deep learning approaches for detecting facial manipulations in images and videos: A review. arXiv preprint arXiv:1811.00656.

[4]. Dang-Nguyen, D. T., Pasquini, C., Conotter, V., & Boato, G. (2019). Deepfake video detection using recurrent neural networks.

[5]. Zhou, X., Han, X., Morariu, V. I., & Davis, L. S. (2018). Two-stream neural networks for tampered face detection.

[6]. Wang, C., Zhang, H., Zhang, X., Liu, J., & Wang, J. (2020). Xception-based deep forgery detection. Information Sciences, 514, 562-574.

[7]. Zhang, J., Wu, J., & Shen, C. (2020). FvDetect: Detecting fake videos generated by deep fake techniques. Information Sciences, 514, 614-627.

[8]. Yang, D., Huang, X., Wang, Y., & Zhao, Q. (2020). Deepfake detection based on dual path neural networks. Signal Processing: Image Communication, 87, 115997.

[9]. Qian, Y., Wang, X., & Li, S. (2019). Recurrent neural networks with attention mechanism for detecting deepfake videos.

[10]. Albiero, V., Gragnaniello, D., Verdoliva, L., & Poggi, G. (2020). Semi-supervised detection of face manipulation in videos.

[11]. Li, Y., Li, X., Li, H., & Li, B. (2020). Improved recurrent neural network-based deepfake detection by fusing different features. IEEE Access, 8, 117258-117270.

[12]. Huang, Y., Wu, Q., Wang, J., & Huang, J. (2020). Real-time deepfake detection method based on improved convolutional neural network. Journal of Real-Time Image Processing, 17(2), 339-350.

[13]. Qian, Y., Yu, X., Xie, Y., & Chen, X. (2020). Deepfake Detection using Multi-scale Convolutional Neural Network with Attention Mechanism. IEEE Access.

[14]. Zeng, Y., Qin, X., Qian, C., Yang, J., & Wen, F. (2020). Detection of Deepfake Video Manipulation Based on Convolutional Neural Networks. Journal of Intelligent & Fuzzy Systems.