Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 43 Issue: 07-2024
DOI: 10.5281/zenodo.12703636

CNN-BASED MODEL FOR DEEPFAKE ANALYSIS USING IMAGES AND VIDEOS

RATNESH KUMAR SHUKLA
Computer Science & Engineering, Shambhunath Institute of Engineering & Technology Prayagraj, Uttar Pradesh, India. Email: ratnesh.nitttr@gmail.com

ALOK SINGH SENGAR
Computer Science, School of Sciences, Noida, International University, Gautum Buddh Nagar, Uttar Pradesh, India. Email: aalok_iitr@live.com

VINAY MISHRA
School of Computer Application, Babu Banarasi Das University, Lucknow, Uttar Pradesh, India.
Email: mishravinay78@gmail.com

NUPA RAM CHAUHAN
Computer Science & Engineering, College of Computing Science & Information Technology, Moradabad, Uttar Pradesh, India. Email: nrcua80@gmail.com

SAURABH KUMAR
Computer Science, School of Sciences, Noida, International University, Gautum Buddh Nagar, Uttar Pradesh, India. Email: saurabhpathak.gkv@gmail.com

RAJESH KUMAR PATHAK
Computer Science & Engineering, GNIOT Greater Noida, Uttar Pradesh, India.
Email: rajeshkumar.pathak@gmail.com

Abstract

The convolutional neural network (CNN) is a frequently used and beneficial approach in many domains, such as computer vision, machine learning and natural language processing (NLP). CNN is used by deepfakes to modify people's photos and videos so that viewers are unable to tell the fake from the real one. Many investigations into the mechanics of deepfakes have been carried out in recent years and a variety of CNN-based methods have been introduced to identify deepfakes in photos or videos. In this research, CNN-based proposed model are implemented the deepfake analysis and face detection technologies. Furthermore, these are provided a comprehensive examination of different technologies and used in the identification of deepfake analysis. Researchers in this field will find this study useful as it covers the most recent state-of-the-art techniques for identifying deepfakes in photos or videos found in social media content. It will also facilitate comparison with previous research due to its comprehensive description of the most recent techniques employed in this field.

Keywords: CNN, Deep Learning, Facial Expression Recognition, Machine Learning and Generative Adversarial Network (GAN).

1. INTRODUCTION

Deepfake AI is an artificial intelligence approach that creates photo, audio and video hoaxes that look realistic. The term, which combines the terms fake and deep learning, describes both the resulting phoney content and the method itself. Deepfakes often use pre-existing source material to create new personas for characters. They also create entirely original videos, where actual people are depicted saying or acting in ways they have never done. It is used for preventing image manipulation to environments [1].

Deepfake artificial intelligence is a relatively new technique that started out as a way to manipulate images using programmes like Adobe Photoshop. By the mid-2010s, deep learning algorithms have advanced in sophistication thanks to a combination of low-cost computer power, big data sets, artificial intelligence, and machine learning technologies [2].

In 2014, Ian Goodfellow, a researcher at the University of Montreal, developed GAN, the technique at the core of deepfakes. In addition to sharing deepfake films of celebrities, an anonymous Reddit user going by the handle deepfakes also started posting GAN tools that allowed users to interchange faces in videos. These became widely shared on social media and the internet. Tech giants like Facebook, Google, and Microsoft invested in creating technologies to identify deepfakes as a result of the rising popularity of deepfake material. The technology is still developing and producing increasingly realistic deepfake photos and movies, even with the efforts of governments and tech corporations to tackle the deepfake detection problem [3].

The task of identifying phoney images or movies created with deep learning techniques is called DeepFake Detection. Deep fakes are produced by manipulating or replacing certain elements of an original video or image, like a person's face, using machine learning algorithms. Finding these kinds of alterations and differentiating them from authentic videos or images is the aim of deepfake detection. Current scenario image manipulation is big issues for the society [4].

Deepfakes is the name of a particular manipulation technique that was popularised through internet forums, but it has also come to be used interchangeably with face replacement based on deep learning [5].

2. RELATED WORKS

As the following technologies are created and improved, creating deepfakes is becoming simpler, more accurate, and more common. Understanding the reasoning behind each decision that may be explained in human words is made possible by the use of traditional machine learning methods. Because there is a greater understanding of the data and processes, these methodologies are appropriate for the Deepfake area. Furthermore, it is far easier to adjust model designs and adjust hyper-parameters [6].

2.1 Generative Adversarial Network (GAN)

All deepfake content is created utilising GAN neural network technology, which combines discriminator and generator techniques. Using GANs, generative models are automatically trained to create realistic-looking synthetic faces in photos or videos by treating the unsupervised problem as supervised. Some machine learning techniques aim to highlight specific anomalies in these GAN-generated fictitious pictures or films [7].

2.2 Convolutional Neural Networks (CNNs)

Patterns in visual input are analysed by convolutional neural networks. CNNs are employed in motion tracking and facial recognition applications [8].

2.3 Autoencoders

Neural network technology known as autoencoders recognises pertinent characteristics of a target, such as body language and facial emotions, and then superimposes these characteristics on the original footage [9].

2.4 High-Performance Computing

Deepfakes require a substantial amount of processing power, which is provided by high-performance computing [10].

2.5 Natural Language Programming (NLP)

Deepfake audio is produced with the use of natural language processing. NLP algorithms use a target's voice characteristics to analyse and then create original text based on those characteristics [11].

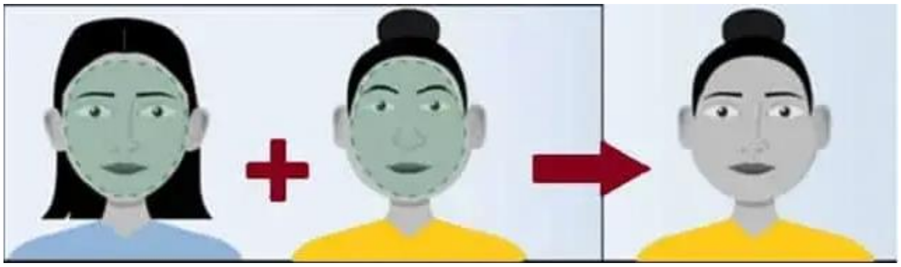


Figure 1: Combination of 2 faces using faceswap.

2.6 FaceSwap

The word deepfake is comes from combining the terms deep learning" and fake, which both mean not true. Deepfakes are synthetic media that alter a person's face in an image or video to make it appear like someone else.

It's referred to as FaceSwap in short. Deepfake Technology uses essential methods from artificial intelligence and machine learning to create and modify visual and aural content that has a high potential for deception. The primary function of FaceSwap or Deepfake technology is to swap faces or alter facial emotions [12].

In figure 1 autoencoder is used to extract inactive features from face images, and the decoder is used to reconstruct the face images. This technique is known as face swapping. The image below shows how to manipulate facial expressions with faceswap. Even though it sounds like a fun activity, there are benefits that make this technology extremely difficult to overcome. Nevertheless, there are two sides to this technology, and it is blamed for negative consequences that don't mesh well with the modern environment [13].

2.7 Face2Face

Face2face connection provides the context for the majority of human communication and is fundamental to human sociality and evolution. Investigating the intricacies that comprise in-person communication necessitates a multidisciplinary and multilevel approach, providing insight into our interactions with other animals from several angles [14].

Figure 2: Face2face interaction with 2 persons.

In the figure 2, the communication between two person are provides an overview of the processes that were looked at in this position. Here are some instances of interaction processes that cannot be thoroughly investigated in solitary persons [15]. There are specific physical requirements for these processes to function, such as shared spaces and the use of specific effectors; there are specific cognitive requirements as well, such as processes that have been studied separately but that have a specific role to play or may interact differently [16]. Regarding the performance issue with machine learning-based Deepfake techniques, it has been noted that these methods can effectively detect Deepfakes.

2.8 Neural Textures

Video coding with an adaptive upsampling approach was aided by neural texture transfer. This scheme determines whether or not to down-sample a frame adaptively. The down-sampled frames are reconstituted in the decoder by examining their correlations with the non-down-sampled frames using multi-scale neural texture transfer [17].

3. METHODOLOGIES

3.1 Face Manipulation Methods

In the field of digital media forensics, the detection of altered facial photos and videos is becoming more and more significant. The emergence of sophisticated techniques for face synthesis and manipulation has led to the creation of novel forms of artificial facial representations, prompting serious concerns regarding their usage on social media [18].

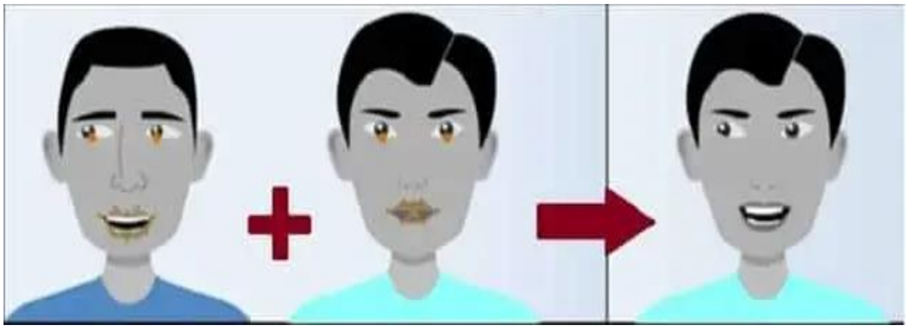


Figure 3: Combination of 2 faces using manipulation approach.

In the above figure 3 Deepfake's core strategies are to trick viewers by manipulating people's faces. There are various methods for achieving that. To trick the users, most approaches alter specific facial features, such the colour of the eyes or an ear band, among other things. The altered area can only be identified or detected by such one-part procedures [19]. Therefore, it is essential to identify and locate modified regions in face image manipulations. In order to further enhance the binary classification (real face vs. false face), the trained attention maps both visualise the altered regions and highlight the informative regions. To make it possible for us to research modified face localization and detection [20].

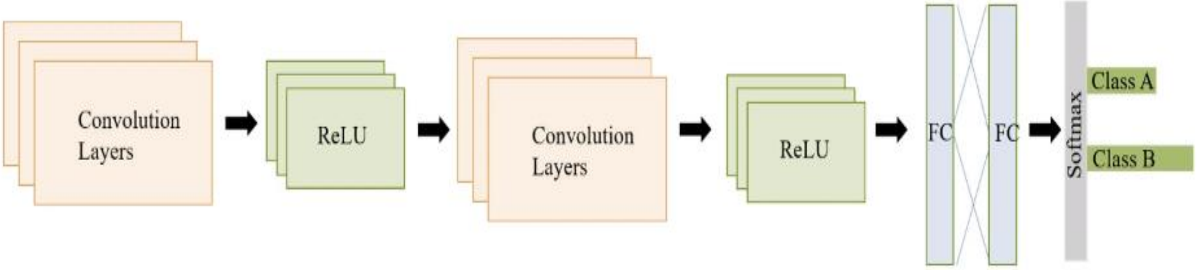


Figure 4: Flow chart of CNN Architecture.

3.2 Multimedia Forensics

One area of forensics that collaborates closely with computer forensics is multimedia forensics. As though The area of forensics known as computer forensics is concerned with using investigative methods to remove and preserve evidence from digital devices that are turned over to legal authorities [21]. On the other hand, multimedia forensics focuses on the analysis of the evidence that has been extracted. The analysis comprises a scientific assessment of the digital evidence in order to preserve its integrity, identify its source, and verify its authenticity [22].

3.3 Forensic Analysis Datasets

Forensic scientists and researchers can use the open-source datasets and databases provided by Centre for Statistics Application in Forensic Evidence (CSAFE) in their labs. By consulting our databases and datasets, they can also increase the statistical rigour of their evidence analysis methods [23].

Academics now find CNN more popular, and it has inspired them to persevere through challenging issues when they had previously given up. In order to address a variety of issues in many study domains, including deep fake detection, researchers have recently developed a number of CNN designs. As seen in figure 4, the general design of CNN often consists of multiple layers stacked on top of one another. A feature extraction module made up of convolutional layers to learn features and pooling layers to reduce image dimensionality makes up CNN's design. Second, it has a module for classifying images that includes a fully connected (FC) layer.

While deep learning has demonstrated impressive results in detecting deepfakes, the quality of deepfakes has been rising. For the purpose of effectively identifying phoney films and photos, the existing deep learning techniques must therefore also be improved. Furthermore, there is currently no reliable way to determine which architecture is best suited for deepfake detection or how many layers are required for deep learning algorithms. In order to increase social media platforms' ability to handle the widespread effects of deepfakes and lessen their consequences, another field of research is the integration of deepfake detection techniques.

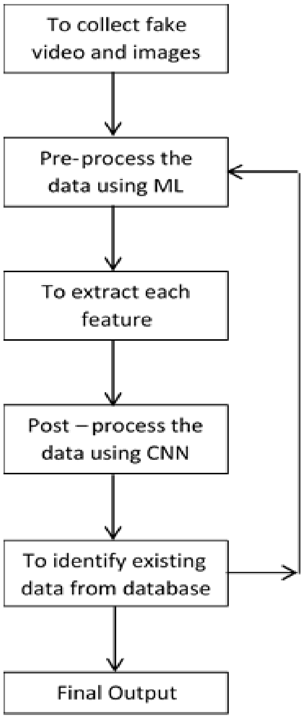

To collect fake video and images
Pre-process the data using ML
To extract each feature
Post – process the data using CNN
To identify existing data from database
Final Output

Figure 5: Data flow diagram of proposed model.

4. PROPOSED MODEL

Figure 5 is showing deepfake facial recognition system using machine learning process to identify the original image from fake video and images. The actual process of extracting data from the video and photos is designed using the proposed model. We meticulously explored reliable databases in order to locate potentially related photographs. The following criteria were met by the papers we selected:

1. The entities that were or needed to be retrieved were specified in the methods or results section.
2. At least one entity was automatically extracted, along with the assessment results for this type of entity.

4.1 Steps belong to collect the real images from Video and Images

In this process, we are applying following steps for correct result.

Step 1: The first step to collect the fake image and videos. Then goto step 2.

Step 2: After step 1, we are applying ML pre-processing and improves the quality of video and images. Then goto step 3.

Step 3: In step 3, we are applying feature extraction technique to extract the features of the video and images. Then goto step 4.

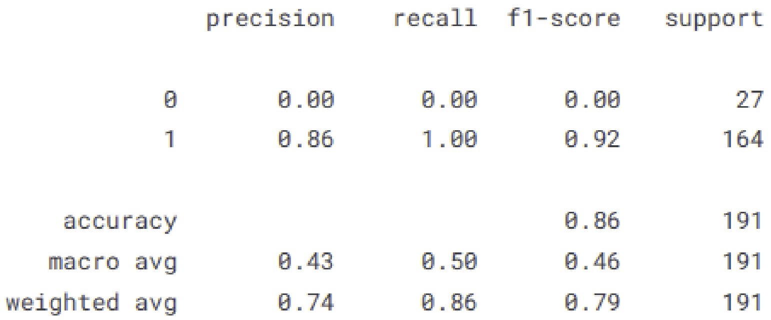
Step 4: In step 4, we are applying post- processing method of deep learning using CNN. Then goto step 5.

Step 5: In step 5, After completing step 4 the match these input data from our registered database, If data is match then goto step 6 unless until goto step 2.

In figure 5 is representing the proposed model of the recognition of facial images from video and images. The most widely utilised deep neural network model is CNN. Similar to neural networks, CNN comprises one or more hidden layers in addition to its input and output layers. In CNN, the input values are subjected to a convolution mathematical process by the hidden layers after they have read the inputs from the first layer. In this case, convolution denotes a dot product or matrix multiplication. CNN uses a nonlinearity activation function, such as the Rectified Linear Unit (RELU), after matrix multiplication. Additional convolutions, such as pooling layers, are then used. By using functions like average or maximum pooling to compute the outputs, pooling layers primarily aim to reduce the dimensionality of the data.

5. EXPERIMENTAL RESULTS

Data extraction and analysis are related to internal validity. The current project required extensive data extraction and processing labour. After we reached a consensus on the comparison outcomes, we applied the cross-checking method to the gathered data and obtained the final data. However, there could still be mistakes in the data collection and processing that we did. To prevent any unforeseen errors, we think the original authors could double-check the results that were published.



```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        27
           1       0.86      1.00      0.92       164

    accuracy                           0.86       191
   macro avg       0.43      0.50      0.46       191
weighted avg       0.74      0.86      0.79       191
```

Figure 6: Output result of the precision, recall, f1-score and support.

Table 1: Flowchart of Accuracy, Precision, Recall, F1-Score and Support Value.

	Precision	**Recall**	**F1-Score**	**Support**
0	0.00	0.00	0.00	27
1	0.86	1.00	0.92	164
Accuracy			0.86	191
Macro avg	0.43	0.50	0.46	191
Weighted avg	0.74	0.86	0.79	191

In figure 6 and table 1 shows the output of the deepfake analysis of the proposed model. The precision value of the output result is calculating maximum 86%, recall value is finding 100% and f1-score is finding 92%.

This is especially crucial today that people may readily disseminate and share these kinds of fake contents on social media platforms and have easier access to the tools needed to create deepfakes.

Many domains have shown great interest in deep learning techniques. Due to its excellent feature extraction capabilities, which make use of matrix multiplication and other linear algebraic concepts to find patterns in images, CNN is widely employed in object recognition and image classification applications. However, the activities involved in using CNN need a significant amount of processing power.

6. CONCLUSION

Because of their feature extraction and selection process, which enables them to learn or extract features directly from the input, deep learning models are widely used in computer vision. Deepfake has become more popular due to the easy access to images and videos in social media content.

To detect Deepfake, deep learning-based methods are commonly utilised. Deep learning models (mainly CNN models) make up a significant percentage of all the models. The precision value of the output result is calculating maximum 86%, recall value is finding 100% and f1-score is finding 92%. Moreover, one could argue that overall, deep learning models outperform non-deep learning models.

References

1) R. Shukla, A. K. Tiwari and A. Mishra, “Face Recognition using LBPH and CNN,” *Journal of Recent Advances of Computer Science and Communications,* vol. 17, no. 5, pp. 48-58, 2024.

2) A. Singh, and A. K. Tiwari, “Machine learning-based approach for prediction of ion channels and their subclasses,” *Journal of Cellular Biochemistry*, vol. 124, no. 1, pp. 72-88, 2023.

3) R. Kumar Shukla and A. Kumar Tiwari, “Comparative analysis of machine learning based approaches for face detection and recognition” *Journal of Information Technology Management*, vol. 13, no. 1, pp. 1-21, 2021.

4) S. Kumar, S. Rani, A. Jain, C. Verma, M. S. Raboaca, M. S., Z. Illés and B.C. Neagu, “Face spoofing, age, gender and facial expression recognition using advance neural network architecture-based biometric system,” Sensors, vol. 22, no. 14, 2022.

5) A. K. Tiwari and R. K. Shukla, “Machine learning approaches for face identification feed forward algorithms,” *In Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE),* March 2019.

6) N. El Abbadi, A. M. Hassan and M.M. AL-Nwany, “Blind fake image detection,” *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no. 4, 180, 2013.

7) R. K. Shukla, V. Prakash and S. Pandey, “A Perspective on Internet of Things: Challenges & Applications” *In 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*, pp. 184-189, December 2020, IEEE.

8) Arvind Kumar Tiwari and Rajeev Srivastava, "A survey of computational intelligence techniques in protein function prediction," *International journal of proteomics,* 2014.

9) A. Gelfert, “Fake news: A definition,” *Informal logic*, vol. 38, no. 1, pp. 84-117, 2018.

10) R. K. Shukla and A. K Tiwari, “Masked Face Recognition Using MobileNet V2 with Transfer Learning,” Computer *Systems Science & Engineering*, vol. 45, no. 1, 2023.

11) P. Singhal, P.K. Srivastava, A. K. Tiwari and R.K. Shukla, “A Survey: Approaches to facial detection and recognition with machine learning techniques,” In Proceedings of Second Doctoral Symposium on Computational Intelligence: DoSCI 2021 (pp. 103-125, 2021, Springer Singapore

12) R. K. Shukla, A. K. Tiwari and V. Verma, “Identification of with face mask and without face mask using face recognition model,” *In 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART),* pp. 462-467, December 2021, IEEE.

13) R. K. Shukla and A. K Tiwari, “A Machine Learning Approaches on Face Detection and Recognition,” *Solid State Technology*, vol. 63, no. 5, pp. 7619-7627, 2020.

14) D. M. E. D. M. Hussein, “Analyzing scientific papers based on sentiment analysis,” *Syst. Dep. Fac. Comput*, no. June, 2016.

15) R. K. Shukla, A. K. Tiwari and A. K. Jha, “An Efficient Approach of Face Detection and Prediction of Drowsiness Using SVM,” *Mathematical Problems in Engineering*, 2023.

16) R. K. Shukla and A. K Tiwari, “Machine Learning approaches for Face Identification Feed Forward Algorithms,” *In Proceedings of 2nd International Conference on Advanced Computing and Software Engineering*, vol. 10, 2019.

17) A. Jain, A. Gupta, A. S. Sengar, R. K. Shukla, and A. Jain, “Application of Deep Learning for Image Sequence Classification,” *In 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, pp. 280-284, December 2021, IEEE.

18) S. Sengar, A. Bhola, R. K. Shukla and A. Gupta, “A Review on Phishing Websites Revealing through Machine Learning,” *In 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, pp. 330-335, December 2021, IEEE.

19) G. Grolleau, T. Lakhal, and N. Mzoughi, “An introduction to the Economics of Fake Degrees,” *Journal of Economic Issues*, vol. 42, no. 3, pp. 673-693, 2008.

20) A. K. Tiwari, “Introduction to machine learning,” *In Ubiquitous Machine Learning and Its Applications*, pp. 1-14, 2017, IGI Global.

21) Kumar, R. Sushil, A. K. Tiwari and D. S. Satwaliya, “Feature extraction and elimination using machine learning algorithm for breast cancer biological datasets,” *International Journal of Advanced Science and Technology*, vol. 28, no. 20, pp. 425-435,2019.

22) S. B. Naeem, R. Bhatti and A. Khan, “An exploration of how fake news is taking over social media and putting public health at risk,” *Health Information & Libraries Journal*, vol. 38, no. 2, pp. 143-149, 2021.

23) V. K. Tiwari, A. Kumar and A. Kumar. “Enhancing ice slurry generation by using inclined cavity for subzero cold thermal energy storage: Simulation, experiment and performance analysis,” *Energy*, vol. 183, pp. 398-414, 2019.