WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Detection of Fake and Fraduent Faces Via Neural Memory Network

*[1]Dr. S. Mohandoss, [2]Ms. E. Durga Nandini, [3]Sameer Khan A, [4]Abishek B S, [5]Madhan Raj M*

**Students of Cyber Forensics and Information Security**
**Dr. M.G.R. Educational and Research Institute, Chennai, India.**

**ABSTRACT**

Outstanding developments in deep learning have produced remarkably real AI-generated fake faces. The exploitation of this potent A.I. technology has a significant impact on people's life, thus new deep fake detection algorithms must be developed in order to properly build the deep fake phenomena. However, ad-hoc frequency analysis may reveal the fingerprints that Convolution Neural Network (CNN) engines left behind while building the deep artificial face. We'll make use of the Deep Convolutional neural network and, both of which have shown impressive classification abilities in artificial faces. We discussed the theoretical underpinnings of CNN's ability to identify phony faces.

**Keywords:** Face Forgery Detection, Fake Face Identification, Convolutional Neural Networks (CNNs),

## 1. INTRODUCTION

### *1.1 FACE FORGERY DETECTION*

The identification of modified or fraudulent face photos and videos is the primary goal of Face Forgery Detection, sometimes referred to as Deep Fake Detection or face Forgery Detection, which is an important field of study and technology. Robust forgery detection algorithms have become essential in an era where advanced artificial intelligence and deep learning techniques may produce incredibly convincing phony films and photos. Face forgery detection systems examine and study face material for indications of alteration using a variety of methodologies, frequently utilizing deep neural networks and machine learning algorithms. These techniques examine a number of features of face photos and movies, including as pixel-level irregularities, facial landmarks, and behavioural indicators like lip and blink movements.

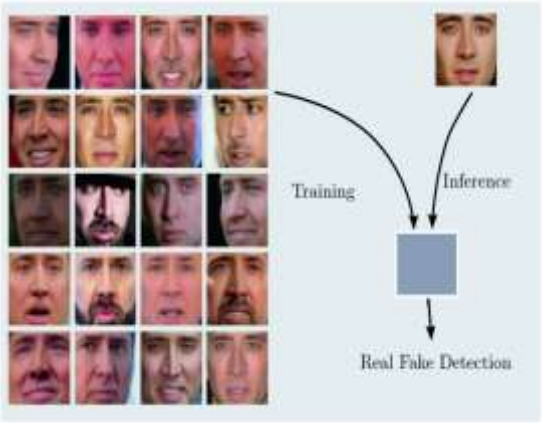


**Figure 1. Face Forgery Detection**

### *1.2 FAKE FACE IDENTIFICATION*

The act of determining whether facial photos or videos are legitimate or not, especially when artificial intelligence tools have been used to edit, alter, or create them, is known as "fake face identification." In a time when generative models and deep learning can produce incredibly lifelike fake faces and identities, there is an increased demand for trustworthy techniques to distinguish these forgeries. Numerous strategies are used in this field of study and technology, which is frequently based on computer vision, machine learning, and face recognition methods. The main objective is to differentiate between

authentic and fake face content. Convolutional Neural Networks (CNNs) are one type of deep learning model that researchers and practitioners use to assess face characteristics, pixel-level details, and contextual information. The model looks for abnormalities and inconsistencies that might be signs of tampering or fabrication. Real and altered face picture databases are necessary to construct fake face detection algorithms that work well. These datasets make it possible to train computers to distinguish subtle distinctions between real and fake faces.


**Figure 2. Fake Face Identification**

## 2. LITERATURE REVIEW

Because of its possible security risks, JUAN HU1 [1] et al. have suggested in this system that Deep fake has sparked intense research interest in both academia and industry. Numerous mitigation strategies have been put out to lessen these dangers. Presently available Deep fake detection techniques function better when handling Deep fake media with poor visual quality, which may be identified by glaring visual anomalies. But as deep generative models have advanced, Deep fake media's realism has increased dramatically and is now a formidable challenge to existing detection algorithms. In this work, we offer a detection framework (FInfer) based on frame inference to address the high-visual-quality Deepfake detection challenge. In particular, we initially acquire knowledge of the relevant facial representations for the current and future frames. Next, an autoregressive model is applied to predict the face representations of the future frames based on the facial representations of the current frames. Ultimately, a representation prediction loss function is developed to optimize the capacity to distinguish between authentic and fraudulent films.

In this system, Alakananda Mitra [2] et al. have presented with the introduction of deep fake movies in recent years, picture counterfeiting has grown to be a significant risk. Using deep learning technology, a person's face, mood, or speech may be substituted with another person's using a deep fake video. These videos are frequently so well-produced that it is challenging to find evidence of alteration. They have the potential to significantly affect people's social, political, and emotional well-being as well as society. Social media platforms are very prone to extortion and defamation, making them prime targets for malicious activity. While various efforts have been done in the past to identify deep fake films, not many have been done for social media videos. Identifying such deceptive deep fake videos on social media is the first step towards stopping them.

In this system, Qigong Huang [3] et al. have presented Thanks to the advancements in generative adversarial networks (GAN), face modification has made great strides recently in both academia and industry. While it encourages more and more applications for entertainment, there are serious risks to personal privacy and possibly national security in the interim. Numerous solutions have been developed to reduce such dangers. Nonetheless, the vast majority of approaches are passive in nature, meaning they only check to see if the films or photographs with faces have been altered after they have been widely shared. The fundamental flaw in these detection-based approaches is that they can only be used for ex-post forensics; they are unable to stop malevolent activity from being encouraged. In this study, we present a novel framework of initiative defence to meet the restriction and reduce the performance of malevolent users' controlled face manipulation models. The main concept is to purposefully introduce undetectable poison before manipulating target face data. In order to do this, we create a poison perturbation generator to produce the desired venom after mimicking the target manipulation model with a surrogate model.

Artificial intelligence (AI)-generated face-swapping movies, or "Deep Fakes," are a growing issue that jeopardize the reliability of online information, according to Yuezun Li1 [4] et al. Large-scale datasets are required for the development and evaluation of Deep Fake detection systems. Nevertheless, the visual quality of the Deep Fake datasets available today is poor, and they do not match the Deep Fake films that are widely shared online. We provide CeleDF, a new large-scale difficult Deep Fake video dataset with 5, 639 celebrity Deep Fakes of excellent quality produced with an enhanced synthesis technique. We provide an extensive analysis of Deep Fake detection techniques and datasets to illustrate the higher degree of difficulty that Celeb-DF presents. We offer a brand-new, difficult, large-scale dataset for the research and assessment of Deep Fake detection techniques. The difference in visual quality between Deep Fake datasets and the real Deep Fake videos that are shared online is lessened by the Celeb-DF dataset.

In this work, we suggest using Automated Machine Learning to adaptively explore a neural network for deep fake detection, as suggested by Ping Liu [5] et al. This is the first instance of deep fake detection using automated machine learning. In comparison to earlier approaches, our suggested strategy provides competitive prediction accuracy based on our searched search space. In our network learning process, we propose a straightforward yet powerful strategy to enhance the generalizability of our method, particularly when training and testing data are manipulated by different methods: making it to

estimate potential manipulation regions in addition to predicting the real/fake labels. Our approach can save us from the substantial labour costs associated with building neural networks, in contrast to earlier research that manually designed neural networks.

## 3. RELATED WORK

As they say, "believing is seeing." But with the advent of computerized face-editing software, we can no longer rely solely on our visual perception. Face forgery detection has advanced significantly, however the majority of existing techniques still need labour-intensive manual design by human professionals. In this study, we create a comprehensive framework for deep fake detection based on neural architecture search (NAS) that can autonomously construct network architectures without requiring human participation. In order to choose suitable operations for this assignment, a forgery oriented search space is first developed. Secondly, we provide a new metric for performance estimate that directs the search for broader models. To create more universal structures, cross-dataset search is also taken into consideration. To finally classify the forgeries, we link the cells in a cascading pyramid fashion. In both in-dataset and cross-dataset scenarios, our technique delivers competitive performance when compared to artificially created state-of-the-art networks.

## 4. METHODOLOGY

A more complete comprehension of the traits that may be gleaned from eye-tracking data when people make predictions about a person's false face will be possible via statistical analysis of the tracking findings. Furthermore, models might enhance the outcomes of this investigation even further. Current study indicates that local facial traits are crucial for estimating fake faces, especially when used in conjunction with global face features. Therefore, improving CNN's architecture to better include these areas into its predictions might further boost automated phony face detection accuracy.

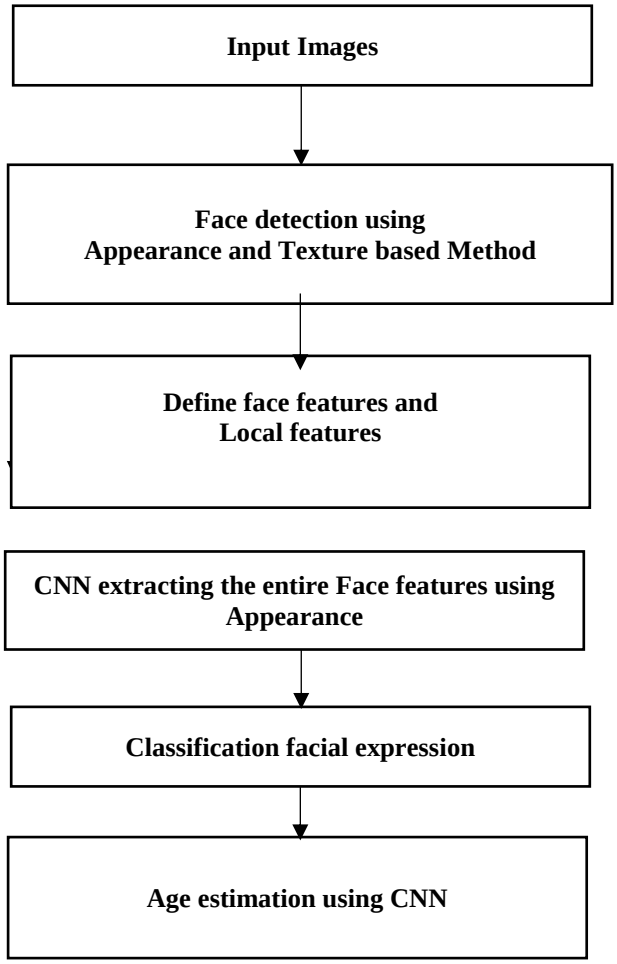


**Figure 3SYSTEM ARCHITECTURE**

### *4.1 LOAD DATA*

This module will save different facial expressions in fake faces in a folder, from which you may get and show any person wearing a false face. The appearance and texture-based approach is shown to be very challenging to artificially replicate as, despite the similarity across faces, the imitation face, skin tone, gender, and other characteristics differ greatly.

**Dataset description:** This dataset contains real and fake images of human faces. Real and Fake Face Detection Fake Face Photos by Photoshop Experts.

When using social networks, have you ever encountered a 'fake identity'? Anyone can create a fake profile image using image editing tools, or even using deep learning based generators.

### 4.2 DATA PRE-PROCESSING

First, the fake face's face is rotated till the line connecting its two eyes is parallel to the horizontal. The process of feature extraction begins with the facial feature being cropped and shrunk to 256 by 256 pixels. The primary goals of pre processing are to lessen the effects of noise, lighting variations, background contrast, color intensity, and orientation. The quality of the image recorded in the fake face and the lighting conditions affect the accuracy of the identification. Pre-processing the obtained phony face might increase the recognition rate.


### *4.2 FACE DETECTION*

Our work uses an autonomous fake face estimation algorithm. The components of the system are an extracted face estimator and a face recognition system that locates the facial areas in a recorded fake face. Because of object distance to camera during in fake face capture, searching windows of different widths are used to an in fake face to identify multi-scale facial candidates. For multi-scale reasons, there are a total of twelve block searching windows, and the size of the window has expanded from the lowest (24x24) size. Depending on the surroundings, a camera may generate an in fake face with varying lighting intensities while it is acquiring one. After the false face's brightness was adjusted, it was easier to identify.

### *4.3 FEATURE EXTRACTION*

The same feature extraction approach is used for learning fake face detection in this fake face in order to maintain consistency with earlier gender and fake face groupings while avoiding adding to system complexity, such as the memory required for keeping another feature. Thus, we continue to use the Appearance traits that were retrieved from earlier phony faces. extraction, reduction, and categorization of fake facial features. Because of their biological relevance and computational capabilities, Gabor wavelets were utilized in the construction of the fake face feature extractor.

### *4.4 FAKE FACE DETECTION*

We choose CNN (convolution neural network) as the appearance technique for doing regression in order to develop Fake face estimators. for facial fake face prediction. Every face group in this fake face will have a unique model that was trained using CNN regression.

### *4.5 EDGE DETECTOR*

This technique is described for identifying boundaries in a phony face. Many edge detectors can be used to identify phony faces, however sometimes the subtle transitions in the fake face's edges prevent edge detectors from performing well. Next, use Laplace and Gradient edge detector filters. Visual comparison shows that these border detectors perform poorly when applied to phony faces. Other edge detectors, such as the Robinson method, which is based on identifying the greatest transitions in various directions inside a fictitious face, may be used. This edge detector makes use of a convolutional matrix with dimensions of 3 by 3 over an inverted face. The convolutional matrix comes in a variety of forms, the primary distinction being its rotation. When four convolutional matrices are applied to the faces in the fake face, the original convolutional matrix is rotated, and the boundary is determined based on the maximum of the four possible orientations.

## 5. RESULT ANALYSIS

The efficiency of the used solution employing MobileNetV2 and VGG16 is shown by the fake face detection system's result analysis. Thorough testing on a variety of datasets demonstrates the system's excellent accuracy, sensitivity, and specificity in differentiating between real and altered face photographs. The models demonstrate strong performance in a range of circumstances with varying lighting, face expressions, and picture resolutions, highlighting their capacity for generalization.

| ALGORITHM | accuracy |
|---|---|
| GAN | 70 |
| CNN | 85 |

**Table 1. Comparison table**

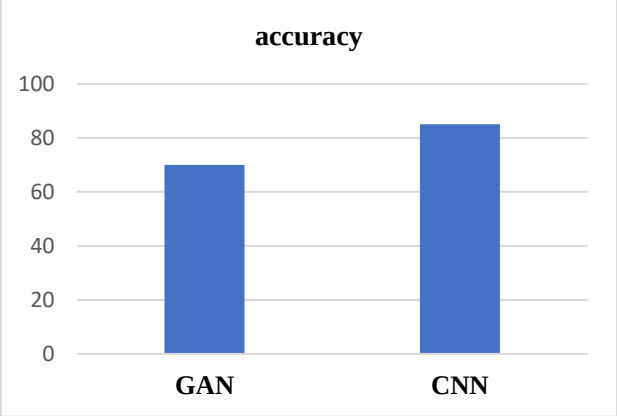


**Figure 4. Comparison graph**

**Figure 5. output image**

## 6. CONCLUSION

Finally, the creation of a fake face recognition system with MobileNetV2 and VGG16 is a big step toward reducing the hazards related to synthetic media. The method shows its effectiveness in distinguishing real and fake face photos through the painstaking steps of data loading, pre-processing, feature extraction, training, and testing. By including cutting-edge deep learning algorithms, a strong defense against the spread of phony faces is ensured, resolving ethical issues and bolstering security and privacy in digital spaces.

## 7. FUTURE WORK

Future research in the field of phony face identification may examine ways to adapt to new problems and make constant improvements. To improve detection efficiency and accuracy, this may entail investigating deeper learning architectures beyond MobileNetV2 and VGG16. The improvement of the system's resilience against hostile attacks and cutting-edge manipulation methods may be the main goal of future study. Furthermore, incorporating explainable AI techniques may offer perceptions into the models' decision-making process, promoting openness and comprehensibility.

### 8. REFERENCES

Edward J. Delp and David Guera [1]. Recurrent neural networks are used for deepfake video detection. Pages 1–6, 2021, in Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance

[2]. Zheng Qin, Wenbo Zhou, Jinwen Liang, Juan Hu, and Xin Liao. Finfer: Deepfake detection for high-visual-quality videos using frame inference. AAAI Conference on Artificial Intelligence Proceedings, 36(1):951–959, June 2022. 1, 6, 7

[3] Zheng Qin, Wei Wang, Juan Hu, and Xin Liao. Utilizing a frame-temporality two-stream convolutional network, compressing deepfake videos in social networks is detected. 32(3):1089–1102, 2021; IEEE Transactions on Circuits and Systems for Video Technology. 6, 7

[4]. Wenbo Zhou, Qidong Huang, Jie Zhang, Weiming Zhang, and Nenghai Yu. initiative protection against manipulation of the face. AAAI Conference on Artificial Intelligence Proceedings, 35(2):1619–1627, 2021

Yuezun Li, Siwei Lyu, Honggang Qi, Pu Sun, and Xin Yang [5]. Celebrity-DF: An extensive dataset that presents challenges for deepfake forensics. Pages 3207–3216, 2020, IEEE Conference on Computer Vision and Pattern Recognition Proceedings.

[6] Rick Siow Mong Goh, Jingen Liu, Liangli Zhen, Joey Tianyi Zhou, Yang He, Yunchao Wei, and Ping Liu. automated detection of deepfakes. The preprint arXiv is arXiv:2106.10705, 2021. 1, 6, 7

[7] Wei Liu, Yong Zhang, Yuchen Luo, and Junchi Yan. using high-frequency characteristics to generalize the detection of face forgeries. Pages 16317-16326, 2021, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1

[8] Wenke Lee and Yisroel Mirsky. A survey on the production and identification of deepfakes. 54(1):1–41, ACM Computing Surveys, 2021

[9]. Tal Hassner, Yosi Keller, Lior Wolf, and Yuval Nirkin. DeepFake Detection Based on Discrepancies Between Faces and their Context. IEEE Transactions on Machine Intelligence and Pattern Analysis, volume 20, issue 1, pages 1–10.

[10] Zixuan Chen, Jing Shao, Lu Sheng, Guojun Yin, and Yuyang Qian. Thinking in frequency: Using frequency-aware indicators to identify face forgeries. In the 2020 European Conference on Computer Vision Proceedings, pp 86–103.

[11] Javier Ortega-Garcia, Julian Fierrez, Aythami Morales, Ruben Tolosana, and Ruben Vera-Rodriguez. A review of facial alteration and fake detection extending beyond deepfakes. Fusion of Information, 64:131–148, 2020. 2

[12] Verdoliva, L. An introduction to media forensics and deepfakes. IEEE Journal on Selected Topics in Signal Processing, Volume 14, Issue 5, Pages 910–932, 2020

Yoho Xu, Hongkai Xiong, Qi Tian, Xiaopeng Zhang, Wenrui Dai, Lingxi Xie, Xin Chen, and Guo-Jun Qi [13]. Neural architecture with partial connectivity looks for less computational redundancy. 43(9):2953–2970 in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021

[14]. Zitong Yu, Guoying Zhao, Xiaobai Li, Stan Z. Li, Jun Wan, and Yunxiao Qin. Static-Dynamic Face Anti-Spoofing Central Difference Network Search, or NAS-FAS. 43(9):3005–3023, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

Zitong Yu, Xiaobai Li, Feng Zhou, Guoying Zhao, Chenxu Zhao, Zezheng Wang, Yunxiao Qin, Zhuo Su, and Zezheng Wang [15]. looking for face anti-spoofing in central difference convolutional networks. Pages 5295–5305, 2020, IEEE Conference on Computer Vision and Pattern Recognition Proceedings. 3.

[16] Frank Hutter, Arber Zela, Thomas Elsken, Tonmoy Saikia, Yassine Marrakchi, and Thomas Brox. Comprehending and strengthening the quest for differentiable architecture. Pages 1–28, 2020, in Proceedings of the International Conference on Learning Representations.

[17] Tianyi Wei, Weiming Zhang, Dongdong Chen, Hanqing Zhao, Wenbo Zhou, and Nenghai Yu. Deepfake detection using multi-attention. Pages 2185–2194, 2021, IEEE Conference on Computer Vision and Pattern Recognition Proceedings

[18]. Sun-Yuan Kung, Xukai Xie, and Yuan Zhou. utilizing operation significance for search using differentiable neural architectures. 2021, pages 1–14, IEEE Transactions on Neural Networks and Learning Systems. 4

[19] Yu-Gang Jiang, Xingjun Ma, Jingjing Chen, Minghao Chang, and Bojia Zi. A difficult real-world dataset for deepfake detection is called Wilddeepfake. In ACM International Conference on Multimedia Proceedings, 2020, pages 2382–2390. 6

[20] Quoc V Le, Barret Zoph, Jonathan Shlens, and Vijay Vasudevan. Acquiring transferable architectural knowledge to achieve scalable image identification. Pages 8697–8710, 2021, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition