iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** V **Month of publication:** May 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.52674

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue V May 2023- Available at www.ijraset.com*

# Deepfake: Creation and Detection using Deep Learning

Prof. Shashi Rekha G[1], Anusha D V[2], Muskan[3], Rakshith K Panchalingalu[4], Sahitya Modi[5]
[1]*Asst. Professor, Sapthagiri College of Engineering Bangalore, Karnataka-560057*
[2, 3, 4, 5]*Student, Sapthagiri College of Engineering, Bengaluru, Kamataka-560057*

***Abstract:*** ***The aim of this study is to develop customized photo-realistic talking head models, which refers to the creation of systems capable of generating believable video sequences that mimic the speech expressions and facial movements of a specific person. The authors propose a system that can produce talking head models using just a few photographs, a technique known as "few-shot learning," and with minimal training time. This system is capable of generating a plausible outcome using just one photograph, and additional photographs enhance the level of personalization. The authors present a system that can perform few-shot learning by conducting meta-learning on a vast collection of videos, which allows it to address the neural talking head models of new and unseen individuals as adversarial training problems with high-capacity generators and discriminators. The system can personalize both the generator and the discriminator's parameters based on each person, enabling training to be performed quickly with only a few images, despite the need to fine-tune millions of parameters.***
***Index Terms: deepfake, deep learning, few shot learning, one shot learning, convolutional neural networks.***

## I. INTRODUCTIOIN

Deep learning, which is also referred to as deepstructured learning, is a category of machine learning techniquesthat is based on artificial neural networks with representation learning. This type of learning can be done through supervised, semi-supervised, or unsupervised methods. Deep learning hasplayed a significant role in image processing by providingpowerful models that can automatically learn and extract features from images. One example of such models is Convolutional Neural Networks (CNNs), which can be used forvarious tasks including image classification, object detection, segmentation, super-resolution, and more. Compared to traditional image processing techniques, CNNs haveoutperformed and are currently considered state-of-the-art inmany tasks.

Deepfake is an example of a deep learning-powered application that has recently emerged. It refers to the use of artificial intelligence (AI) techniques to create manipulated videos or images that appear to be authentic but are actually synthetic. Thisis done by training deep neural networks to learn facial features, body movements, and speech patterns, and then using that information to generate new content that imitates them. While Deepfakes can be used for harmless entertainment, such as creating humorous videos, they can also be exploited for malicious purposes, such as spreading misinformation, propaganda, or creating fake news.

The technology has raised serious concerns about its potential to undermine the integrity of visual media, erode trust in public institutions, and threaten privacy. Deepfake technology is a controversial and disruptive technology with far-reaching impacts on society. It has been associated with several issues such as election biasing, cyberbullying, and the potential to manipulate public opinion. In this project, we propose an integrated system with –

The proposed integrated system includes a face forensics model that combines the conventional image forensic approach with the fake face image forensic approach a system where we can detect manipulated or altered media with convolutional approaches.

Deepfake creation involves training a deep neural network to generate highly realistic synthetic images and videos by learning from a large dataset of real images and videos. These techniques can be used to create highly convincing forgeries of people appearing to say or do things they never actually did, which can have serious implications for privacy, security, and democracy. On the other hand, deepfake detection aims to identify and distinguish real videos and images from fake ones using machine learning algorithms that can identify anomalies in visual and audio signals. This involves training a deep neural network to distinguish between authentic and manipulated media based on patterns in the data. Deepfake creation and detection are important topics in the field of AI and computer vision, as they have the potential to greatly impact various industries, such as entertainment, journalism, and politics. As deepfake technology advances, it becomes increasingly important to develop robust detection methods to prevent the spread of malicious and misleading content.

©IJRASET: All Rights are Reserved 

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue V May 2023- Available at www.ijraset.com*

## II. RELATED WORK

### *A. Generative Adversarial Networks (GANs)*

GANs (Generative Adversarial Networks) are a popular type of deep learning architecture used for generating realistic images and videos. GANs consist of two neural networks: a generator and a discriminator. The generator creates fake images or videos, and the discriminator tries to distinguish the fake ones from the real ones. The two networks are trained together, and as a result, the generator gets better at creating more realistic deepfakes overtime. The training process of a GAN involves multiple steps.First, the generator creates a batch of fake images, and the discriminator is trained on both the real and fake images to improve its ability to distinguish between the two. The generatoris then updated to try to create better fake images that can fool the discriminator. This process is repeated until the generator can create realistic images that are indistinguishable from real images. One popular GAN-based deepfake technology is StyleGAN, which was developed by researchers at Nvidia. StyleGAN can generate highly realistic images of human faces that are difficult to distinguish from real images.

### *B. Voice Cloning*

Voice cloning is a type of deepfake technology used to produce synthetic voice recordings that resemble a particular person's voice. To accomplish this, a deep learning model is trained usinga large dataset of audio recordings of the target person's voice to learn their unique vocal characteristics. The model can then create new speech in the target person's voice. The voice cloning process includes multiple stages. First, the target person's voice is recorded, and the recording is then segmented into smaller units, such as phonemes or syllables. The deep learning model is then trained on this segmented dataset using a technique called sequence-to-sequence learning. During training, the model learns to generate new speech based on the target person's voice characteristics. Once the model is trained, it can generate new speech in the target person's voice. To generate new speech, the model is provided with a text input, which it then converts into speech using the target person's voice characteristics.

### *C. Natural Language Processing (NLP)*

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI) that focuses on the interaction between computers and human language. NLP involves developing algorithms and models that enable computers to understand, interpret, and generate human language. NLP has numerous applications, including language translation, sentiment analysis, speech recognition, and text summarization. One of the main challenges of NLP is that human language is complex and ambiguous, making it difficult for computers to understand and interpret it accurately. NLP techniques involve the use of various mathematical models and algorithms, such as deep learning, statistical models, and rule-based approaches. Some of the popular techniques used in deepfake detection include - Sentiment analysis, This technique involves analyzing text data to determine the sentiment expressed in it. Sentiment analysis is often used to detect deepfake reviews, comments, and social media posts.

Named Entity Recognition (NER), This technique involvesidentifying and classifying named entities in text data, such as names of people, organizations, and locations. NLP techniques have been used in various deepfake detection approaches, such as detecting fake reviews, detecting fake news, and detecting fake social media posts. However, deepfake detection using NLP is still a challenging task, as deepfake techniques are becoming increasingly sophisticated, making it difficult for NLP models to detect them accurately.

## III. LITERATURE SURVEY

*1)* Contrast Enhancement (CE) forensic methods can be performed using relatively simple handcrafted features based on first-and second-order statistics, but these methods have encountered difficulties in detecting modern counter-forensic attacks. The experimental results indicate that the proposed method outperforms conventional forensic methods in terms of forgery-detection accuracy, especially in dealing with counter- forensic attacks.

*2)* With recent advances in deep learning, it is now possible to seamlessly generate manipulated images/videos in real-time using technologies like image morphing, Snap-Chat, Computer Generated Face Image (CGFI), Generative Adversarial Networks (GAN) and Face2Face. Here two types of manipulation is considered: Source-to-target, Self-reenactment.

*3)* To propose a system that processes a conceptually simple and general framework called MetaGAN for few-shot learning problems. MetaGAN: An Adversarial Approach to Few-Shot Learning presents MetaGAN as a general and flexible framework for few-shot learning also proposes MetaGAN, a simple and generic framework to boost the performance of few-shot learning models.

©IJRASET: All Rights are Reserved 

4) Data Augmentation Generative Adversarial Network (DAGAN)enables effective neural network training even in low- data target domains. The datasets used are The Omniglot Dataset. DAGAN and then evaluate its performance on low-data target domains using (a) standard stochastic-gradient neural network training, and (b) specific one-shot meta-learning methods.
5) StarGAN is an architecture which is not only able to synthesize novel expressions, but also changes other attributes of the face, such as age, hair colour or gender. The facial expression synthesistask, is trained on the RaFD dataset which has 8 binary labels for facial expressions, namely sad, neutral, angry, contemptuous, disgusted, surprised, fearful and happy.
6) This paper provided a thorough understanding of deepfake images and videos: how they are generated, how they can be detected, and how AI could be useful in their detection. The approach proposed an architecture that explores attention and feature maps to holistically investigate how well an AI can detect deepfake images.
7) In this paper, a unique approach has been developed to reveal AI- generated deepfake video together with powerful feature extraction & classification utilizing customized CNNs. The comparative analysis found that the proposed customized CNN outperform two existing methods like CNN and MLP-CNN.

## IV. PROPOSED METHOD

In this section, we will discuss the proposed method in detail particularly the software architecture that has been developed for deepfake creation and detection.

### A. Software Architecture

Deepfake creation and detection involve a complex software architecture that combines several deep learning techniques. Here, we will discuss the software architecture for deepfake creation and detection using deep learning,specifically the Convolutional Neural Network (CNN), Haar Cascade algorithm, and encoder-decoder techniques.

#### 1) Deepfake Creation

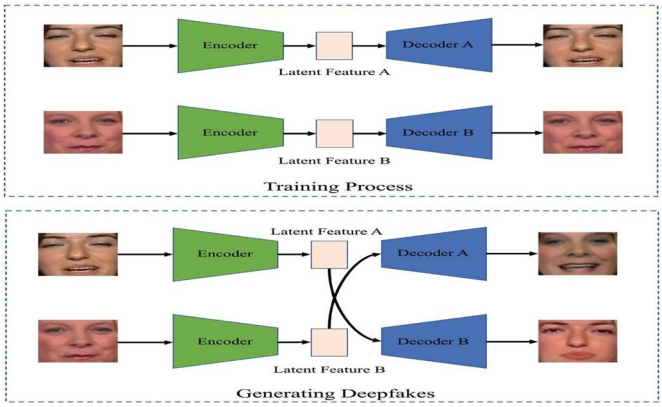



Fig. 1 An overview of the Deepfake creation using Encoder-Decoder technique.

Fig. 1 shows the overview of deepfake creation and has thefollowing steps:
Data Collection and Preparation: The first step involves collecting a large dataset of images and videos of the target person and preparing the data for training the deep neural network.
Face Detection and Alignment: In this step, the faces of the target person are detected and aligned to a common orientation to ensure consistent input to the deep neural network.
Feature Extraction: The next step involves extracting key facial features such as eye shape, lip movements, and skin texture, which are important for generating realistic deepfake images and videos.
Encoder-Decoder Architecture: The deep neural network is typically designed as an encoder-decoder architecture, which is capable of generating high-quality images and videos. The encoder network compresses the input data into a lower dimensional representation, while the decoder network generates the output image or video.
Training and Optimization: The deep neural network is trained on the prepared dataset, and the model parameters areoptimized to minimize the error between the generated deepfake image or video and the ground truth.
Post-Processing: Finally, the generated deepfake image or video is post-processed to improve its visual quality and realism, which may involve adding noise, adjusting brightnessand contrast, and applying filters.

*2) Deepfake Detection*

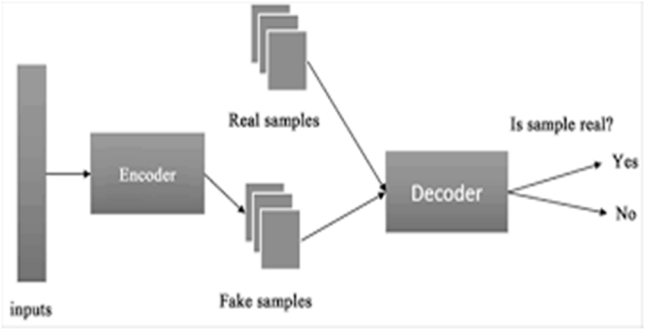



Fig. 2 An overview of the Deepfake detection using Encoder-Decoder technique.

Fig. 2 shows the overview of deepfake detection process andhas the following steps:
Dataset Collection and Preparation: The first step involves collecting a large dataset of real and deepfake images and videos and preparing the data for training the deep neural network.
Feature Extraction: In this step, visual and audio features are extracted from the input data, which may include facial expressions, lip movements, and speech patterns.
CNN and Haar Cascade Algorithm: The deep neural network is typically designed as a CNN with the Haar Cascade algorithm for detecting faces in images or videos. The CNN is trained to distinguish between real and fake images or videos by identifying patterns in the extracted features.
Training and Optimization: The deep neural network is trained on the prepared dataset, and the model parameters are optimized to minimize the error between the predicted andactual labels.
Evaluation and Testing: The trained deep neural network is evaluated and tested on a separate dataset to assess its performance in detecting deepfakes.
Deepfake detection involves identifying and distinguishing real images and videos from fake ones. The Haar Cascadealgorithm is a popular technique for detecting faces in images and videos, which is often used as a preliminary step in deepfake detection. The Haar Cascade algorithm is a machine learning-based approach that uses a cascade of classifiers to detect faces in an image. It works by examining features suchas edges, corners, and texture variations in the image and using these to distinguish between face and non-face regions. Once a face is detected, the image or video can be further analyzed using deep learning techniques such as CNNs to detect any signs of manipulation or distortion. Overall, the software architecture for deepfake creation and detection using deep learning involves a combination of techniques such as CNNs, Haar Cascade algorithm, and encoder-decoder architecture. These techniques allow for the creation of highly realistic deepfakes and the detection of malicious and misleading content. In summary, the software architecture for deepfake creation and detection using deep learning involves a combination of different techniques, including encoder- decoder architectures, CNNs, and the Haar Cascade algorithm.

*B. Haar-Cascade Algorithm*

The Haar Cascade algorithm is a popular technique used in computer vision for object detection, and specifically for face detection. It was first proposed by Viola and Jones in 2001 and is a widely used technique for deepfake detection. The algorithm uses a machine learning approach to identify the presence of faces in an image or video. It works by breaking down the detection process into a series of simpler, localized feature classifiers. These classifiers are arranged in a cascade, where each stage of the cascade applies a more complex set of rules to the input data. To train the algorithm, a large dataset of positive and negative examples is used. The positive examples consist of images containing faces, while the negative examples are images without faces. The algorithm learns to distinguish between these examples by optimizing a set of parameters that define the feature classifiers used in the cascade. In deepfake detection, the Haar Cascade algorithm is used as a preliminary step to identify regions of an image or video that are likely to contain a face. Once these regions have been identified, more complex deep learning techniques, such as CNNs, can be usedto further analyze the image or video for signs of manipulation or distortion. Overall, the Haar Cascade algorithm is a powerful tool for deepfake detection that can quickly and accurately identify regions of an image or video that contain faces. Its ability to eliminate non-face regions early in the detection process helps to reduce computational requirements and speed up the detection process.

## V. CONCLUSION

In conclusion, deepfake creation and detection are emergingareas of research in the field of AI and computer vision. While deepfake technology has the potential to revolutionize various industries, it also poses serious threats to privacy, security, and democracy. Deep learning algorithms, such as convolutional neural networks (CNNs) and encoder-decoder techniques, are used to generate and detect deepfakes. CNNs can extract relevant features from images and videos to accurately classify them as authentic or manipulated. Encoder-decoder techniques, on the other hand, use a combination of convolutional and deconvolutional layers to generate synthetic media that is highly convincing. Haar Cascade algorithm, another machine learning algorithm used for deepfake detection, analyzes facial features such as eyes, nose, and mouth to detect anomalies and inconsistencies in facial expressions. While deepfake creation and detection are still evolving, it is clear that robust and effective detection methods are necessary to combat thepotential harms of deepfakes. As deepfake technology continues to advance, researchers and developers must prioritize developing reliable detection methods to prevent the spread of malicious and misleading content.

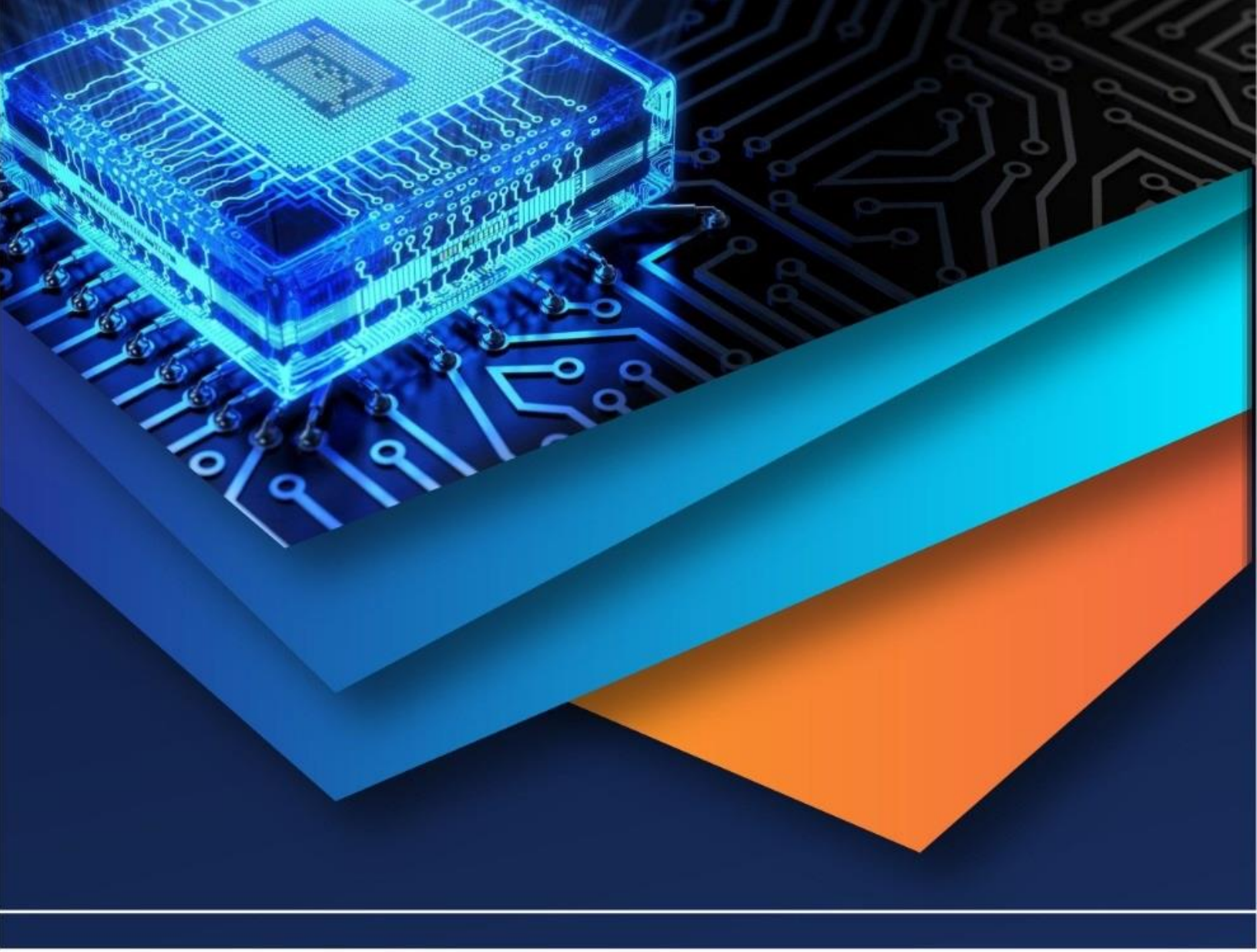
## VI. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who have provided their invaluable support and assistance during the course of this project. We would like to extend our heartfelt thanks to Dr. Kamalakshi Naganna, Professor and Head, Department of Computer Science and Engineering, Sapthagiri College of Engineering and our project guide Prof. Shashi Rekha G, Asst. Professor, Department of Computer Science and Engineering, Sapthagiri College of Engineering, who has provided constant guidance, support, and encouragement throughout the project. Their invaluable insights and suggestions have been instrumental in shaping this project. Finally, we are deeply grateful to our friends and family for their constant support and encouragement.

## REFERENCES

[1] Jee-Young Sun, Seung-Wook Kim, Sang-Won Lee, Sung-JeaKo," A novel contrastenhancement forensics based on convolutional neural networks", 01 April 2018.
[2] Andreas R¨ossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner, "FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces" 24 March 2018.
[3] Ruixiang Zhang, Tong Che, Zoubin Grahahramani, Yoshua Bengio, Yangqiu Song, "MetaGAN: An Adversarial Approach to Few-Shot Learning" 3 December 2018.
[4] Antreas Antoniou, Amos Storkey, Harrison Edwards "Data Augmentation GenerativeAdversarial Networks" 8 March 2018.
[5] Albert Pumarola, Antonio Agudo, Aleix M. Martinez, Alberto Sanfeliu, Francesc Moreno-Noguer, "GANimation: Anatomically aware Facial Animation from a SingleImage" 28 August 2018.
[6] Samuel Henrique Silva, Mazal Bethany, Alexis Megan Votto, Ian Henry Scarff, Nicole Beebe, Peyman Najafirad, "Deepfakes Forensics Analysis: An explainable hierarchical" Forensix Science International Synergy 4 (2022) 100217.
[7] Usha Kosarkar, Gopal SarkarKar, Shilpa Gedam, "Revealing and classification of deepfakes videos's images using a customized convolutional neural network model" International Conference on Machine Learning and Data Engineering (2023).
[8] A. Bromme, C. Busch, A. Dantcheva, C. Rathgeb, A.Uhl, "Fake Face Detection Methods: Can they be generalized?" 28 Sep t 2018 .
[9] Ammar Elhassan, Mohammad Al-Fawareh, Mousa Tayseer Jafar, Mohammad Ababneh, Shifaa Tayseer Jafar, "DFT-MF: Enhanced deepfake detection using mouthmovement and transfer learning" SoftwareX 19 (2022) 101115.
[10] Vurimi Veera Venkata Naga Sai Vamsi, Sukanya S Shet, Sodum Sai Mohan Reddy, Sharon S Rose, Sona R Shetty, S Sathvika, Supriya M S, Sahana P Shankar, "Deepfakedetection in deep media forensics" Global Transaction Proceedings 3 (2022) 74-79.
[11] A. K. Jain, A. Ross, and S. Pankanti, ''Biometrics: A tool for information security,'' IEEE Trans. Inf. Forensics Security, vol. 1, no. 2, pp. 125–143, Jun. 2006
[12] A. K. Jain, A. Ross, and S. Prabhakar, ''An introduction to biometric recognition,'' IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 1, pp. 4–20, Jan. 2004.
[13] DeepFake Github Repository. Accessed: Jun. 14, 2019. [Online]. Available: https://github.com/deepfakes/faceswap
[14] Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niessner, ''Face2Face: Real-time face capture and reenactment of RGB videos,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2387–2395.
[15] P. Pérez, M. Gangnet, and A. Blake, ''Poisson image editing,'' ACM Trans. Graph., vol. 22, no. 3, pp. 313–318, Jul. 2003.
[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.
[17] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large- scale image recognition,'' in Proc. 3rd Int. Conf. Learn. Represent. (ICLR), 2015.
[18] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.
[19] F. Chollet, ''Xception: Deep learning with depthwise separable convolutions,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1251– 1258.
[20] M. D. Zeiler and R. Fergus, ''Visualizing and understanding convolutional networks,'' in Proc. Eur. Conf. Comput. Vis. (ECCV). Cham, Switzerland: Springer, 2014, pp. 818–833.
[21] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, ''FaceForensics: A large-scale video dataset for forgery detection in human faces,'' 2018, arXiv:1803.09179. [Online]. Available: http://arxiv.org/abs/1803.09179

[22] A. Khodabakhsh, R. Ramachandra, K. Raja, P. Wasnik, and C. Busch, ''Fakeface detection methods: Can they be generalized?'' in Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG), Sep. 2018, pp. 1–6.
[23] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, ''FaceForensics++: Learning to detect manipulated facial images,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 1–11.
[24] A. C. Popescu and H. Farid, ''Exposing digital forgeries by detecting traces ofresampling,'' IEEE Trans. Signal Process., vol. 53, no. 2, pp. 758–767, Feb. 2005.
[25] M. Kirchner, ''Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue,'' in Proc. 10th ACM Workshop Multimedia Secur. (MM Sec), 2008, pp. 11–20.
[26] N. Dalgaard, C. Mosquera, and F. Perez-Gonzalez, ''On the role of differentiation for resampling detection,'' in Proc. IEEE Int. Conf. Image Process., Sep. 2010, pp. 1753–1756.
[27] X. Feng, I. J. Cox, and G. Doërr, ''Normalized energy density-based forensicdetection of resampled images,'' IEEE Trans. Multimedia, vol. 14, no. 3, pp. 536–545, Jun. 2012.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)