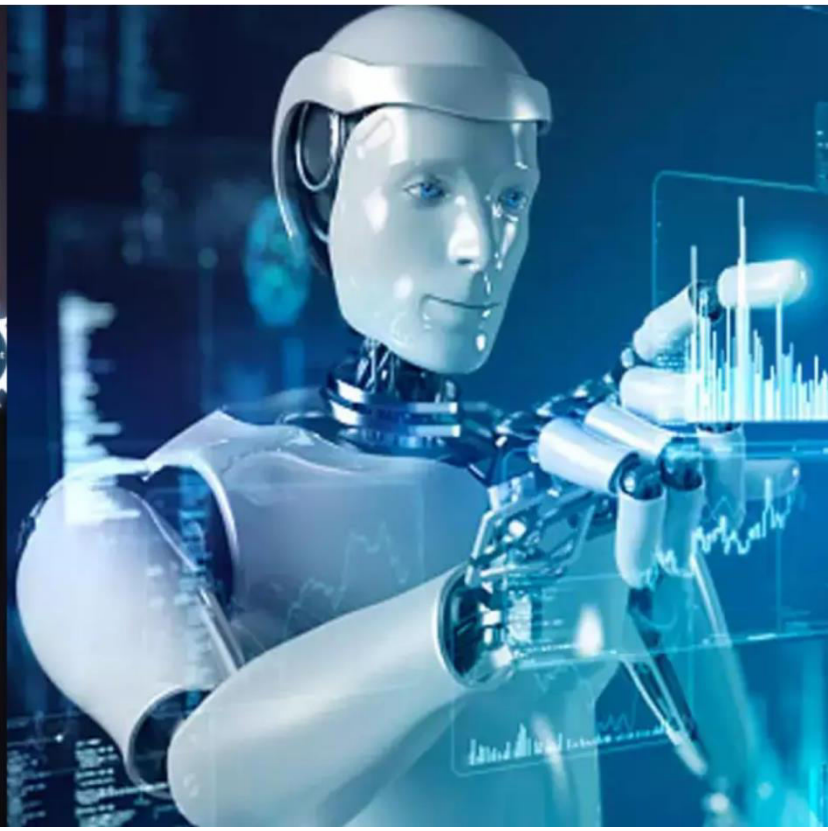
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 5, May 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505038

Detection of Manipulated Images using Deep Learning Techniques

Karanam Tarun Kumar, Vennavalli Kiran, Padala Ganesh Reddy, Ms.D.Nandhini

Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

Assistant Professor, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

ABSTRACT: Deepfake detection is targeted at recognizing and distinguishing between real images and those manipulated through advanced artificial intelligence technologies. This work adopted deep learning for the development of a strong detection system that can identify minute artifacts and inconsistencies from image manipulations. The system learns to distinguish between real and fake images by studying a large collection of examples. It uses advanced tools, such as CNNs and ResNets, to do this effectively. This will ensure that the model accurately classifies images in real time by adapting to various manipulation techniques and quality levels of images. Preprocessing, training the models, and evaluating based on metrics like accuracy, precision, and recall are some of the steps involved in the detection process. This paper keeps improving its methods to keep up with new deepfake tricks and better technology. The goal is to make the internet safer and more trustworthy by offering a reliable way to spot fake images or videos. This helps fight misinformation and makes it easier to trust what you see online.

KEYWORDS: deepfakes, detection, classification, neural networks.

I. INTRODUCTION

Deepfakes are videos, images, and audio generated with the help of AI that make a person seem to say or do something he or she never actually did. While originally developed for entertainment purposes, such as face swaps in movies, deepfakes are now being used for malicious purposes: to disseminate fake news, manipulate political outcomes, bully, and spread non-consensual adult content.

Experts catch deepfakes by zeroing in on little slip-ups—strange lighting, faces that twitch in odd ways, or voices that don't really fit. They also feed loads of real and fake clips into machine learning models, so those systems learn to spot what feels wrong. And yeah, some folks add digital watermarks to videos, just to make fakes stand out more. Watermarks or blockchain can help establish whether a video is original and unedited.

II. LITERATURE SURVEY

In recent years, deepfake generation and detection have seen rapid development with a marked trend toward robust segmentation, multiscale feature learning, and biometric-driven forensic analysis. It is in this context that one of the most critical recent developments in SAM is by Kirillov et al. (2023) [1], a universal segmentation framework for segmenting objects and facial regions with very high accuracy. The flexibility introduced by SAM can enable region-focused forensic analysis in deepfake detection pipelines, where detectors can put more emphasis on those areas that are more prone to manipulations, hence spotting subtle local artifacts with higher accuracy.

Mirsky and Lee [2] created, in 2022, a detailed survey on both deepfake creation and detection techniques. In it, they discussed the dynamics of the adversaries, limitations of datasets, and detectors' vulnerability to unseen manipulation techniques. Their analysis has underlined the growth in sophistication of the generative models, along with the increased demand for detectors capable of generalizing beyond known forgery patterns. Gao et al. (2021) [4] proposed Res2Net, a multiscale feature extraction mechanism enabling more detailed analysis of both fine-grained inconsistencies and high-level structures—a necessary capability for forged facial content detection. Another relevant

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505038

architecture is the Point Transformer by Engel et al. (2021) [5], generalizing attention-based feature aggregation to tasks like 3D-aware and multimodal deepfake detection.

Earlier studies also provided the foundational understanding of deepfake forensics. Nguyen and Derakhshani (2020) [6] researched biometric-driven methods, focusing on micro-level facial cues such as eyebrow recognition. Their work indicated that deepfakes tend to poorly replicate subtle physiological patterns and, therefore, that biometric irregularities remain very informative indicators for detection. The same year, Hellsten et al. (2020) [7] studied the process of training GANs with limited data; their surprising finding was that good-quality synthesis is possible even in resource-scarce settings due to regularization and augmentation strategies. This complicates the life of detectors since attacks need fewer resources to yield plausible forgeries. Finally, Verdoliva (2020) [8] contributed an essential media-forensics perspective, thereby detailing multimodal detection strategies and emphasizing the importance of reliable benchmarks for evaluating detection robustness under real-world conditions.

III. EXISTING SYSTEM

The different existing image deepfake detection systems make use of deep learning models, visual artifact analysis, and biometric evaluations in combination to identify manipulated content. Many of the detection systems are based on convolutional neural networks that are trained on large datasets of genuine and fake images in order to learn subtle discrepancies. Take MesoNet, for example. It's a CNN-based model that zeroes in on mesoscopic features to catch forgeries. Some systems take a different route, using deep learning to spot weirdness in facial landmarks—think odd shapes in the eyes, nose, or mouth—that often pop up in deepfakes. Then there's temporal analysis, which checks for odd changes from frame to frame in videos, and spatial analysis, which hunts for strange pixel patterns in images. Both help boost how well these systems catch fakes.

On top of that, researchers use adversarial training. Basically, they keep making their models smarter by feeding them tougher and tougher deepfakes. When deepfakes get even trickier—like those that mess with both visuals and audio—people are turning to multimodal detection systems that look at both what you see and what you hear.

Even with all these advances, the truth is, current detection systems still have a hard time handling the sheer variety of deepfakes out there. They also struggle to keep up when new techniques come along. That's why ongoing research and fresh ideas matter if we want these tools to stay reliable and actually catch the fakes that slip through today.

DISADVANTAGES:

Image deepfake detection faces, despite its importance, numerous challenges. A big disadvantage is the continuous cat-and-mouse game that creators of deepfakes and detectors are playing: the better the methods for detection become, the more cunning the techniques of deepfake creation are, too, often managing to outsmart detection systems. This creates false positives, when real content is taken for fake, and vice versa—false negatives, when deepfakes have managed to pass through undetected. Then, most models struggle to generalize among different types of deepfakes, especially in cases with the newest algorithms or involving non-facial content. Running deepfake detection algorithms can be computationally expensive, especially in real-time use cases, which seriously limits their scalability and accessibility. Besides that, the usual pain point in machine learning—large and diverse datasets for model training—can be a challenge to overcome because biases in the data may result in inaccurate or unfair detection outcomes. Finally, deepfake detection as such cannot overcome the problems of misinformation and digital manipulation but has to be complemented by public awareness, education, and legislation if it is going to be effective.

IV. PROPOSED SYSTEM

The goal of the proposed system for image deepfake detection is to improve the accuracy and robustness of the system by combining various sophisticated detection methods. The new system blends Convolutional Neural Networks (CNNs) with Transformers, so it can pick up on both tiny details and the bigger picture. This makes it a lot better at spotting even subtle manipulations. It works by running two checks at once: one digs into pixel-level details to catch visual artifacts, and the other focuses on biometric clues in facial landmarks and expressions. This method resolves problems that face irregularities in facial geometry and unnatural movements which deepfakes, in general, cannot replicate with high accuracy.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505038

Moreover, the system will have temporal coherence analysis to enhance the detection power, even if the pictures are static, by finding very small signals that indicate the different parts of the image have been tampered with. On top of that, adversarial training will be implemented where the detection model will be continuously challenged with the latest and advanced deepfake generation techniques, thereby, the model will be able to fight off the new threats. In order to be scalable and generalize well to different types of deepfakes, the system will be exposed to a wide variety and large scale of deepfakes generated by different algorithms during training.

Furthermore, a simple-to-use interface will enable detecting in real-time, thus giving confidence scores and also pointing the manipulated areas, which will help non-experts as well as experts to make a decision. The multi-pronged strategy here is to develop a stable, flexible and user-centric tool for the problem of image deepfakes, and the tool should be able to detect the problem and fight against it efficiently.

V. SYSTEM ARCHITECTURE

Image deepfake detection system architecture is set up to combine several components that perform the task of detecting fake images with high precision independently. Here's how the architecture breaks down:

1. Data Preprocessing Module: This is where everything starts. The system gets the images ready—resizing, normalizing, and adding some variations so it can handle all kinds of image data. It also pulls out key features, like edges, textures, and color patterns.

2. Dual-Stream Detection Module: Here's the heart of the operation. There are two deep learning models running side by side. One's a CNN, digging into the tiny details and pixel-level changes. The other's a Transformer, looking at the bigger picture and making sense of the overall context.

3. Biometric and Temporal Analysis Module: This part checks facial landmarks and other biometric clues, comparing them to what's expected. It's also on the lookout for anything weird in the sequence of images—even if it's just a single frame—to catch the subtlest signs of tampering or deepfakes.

4. Adversarial Training Module: This is the system's defense mechanism. It keeps throwing the latest deepfake tricks at the models, forcing them to adapt and get better at spotting new types of forgeries.

5. User Interface Module: Through a user-friendly interface real-time detection results are delivered, areas suspected of manipulation are highlighted, and confidence scores are given thus enabling users to take action from the system's feedback.

The present framework is intended to be a reliable, flexible, and extendable solution for the problem of deepfakes, which can be intricate and resourceful in nature.

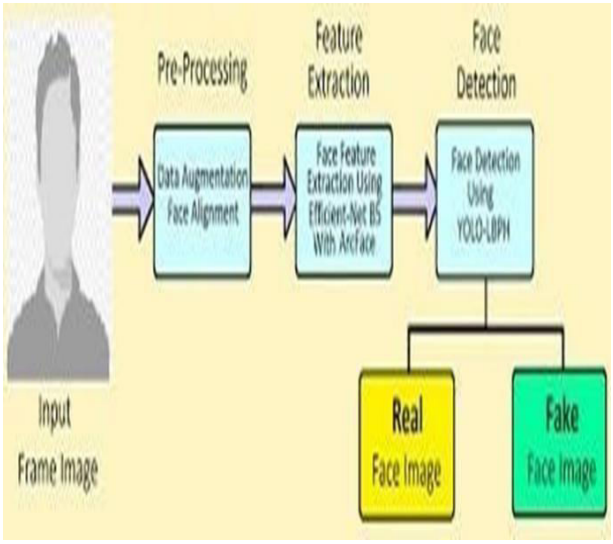


Fig.1. Real or Fake Face Checking Flow

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505038

Fig. 1 shows the system takes a face picture and first prepares it by adjusting and improving the image. Then it looks at important parts of the face. In the end, it checks the face and tells whether it is real or fake.

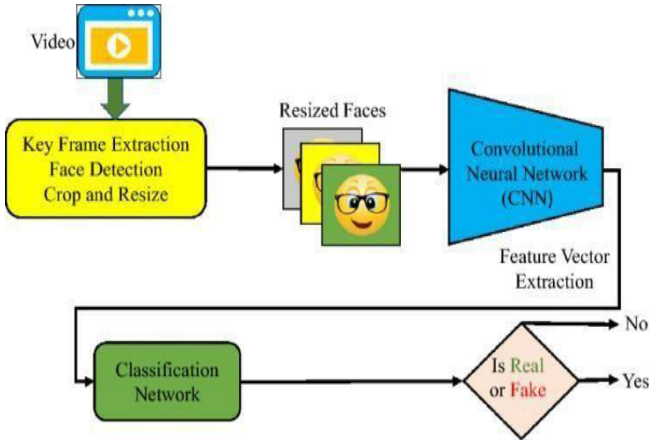


Fig.2. Real or Fake Video Frame Analysis

Fig. 2 shows the system takes a video, extracts key frames, and detects faces, which are then cropped and resized. These faces are passed through a CNN to extract important features. Finally, a classification network checks the features and decides whether the face is real or fake.

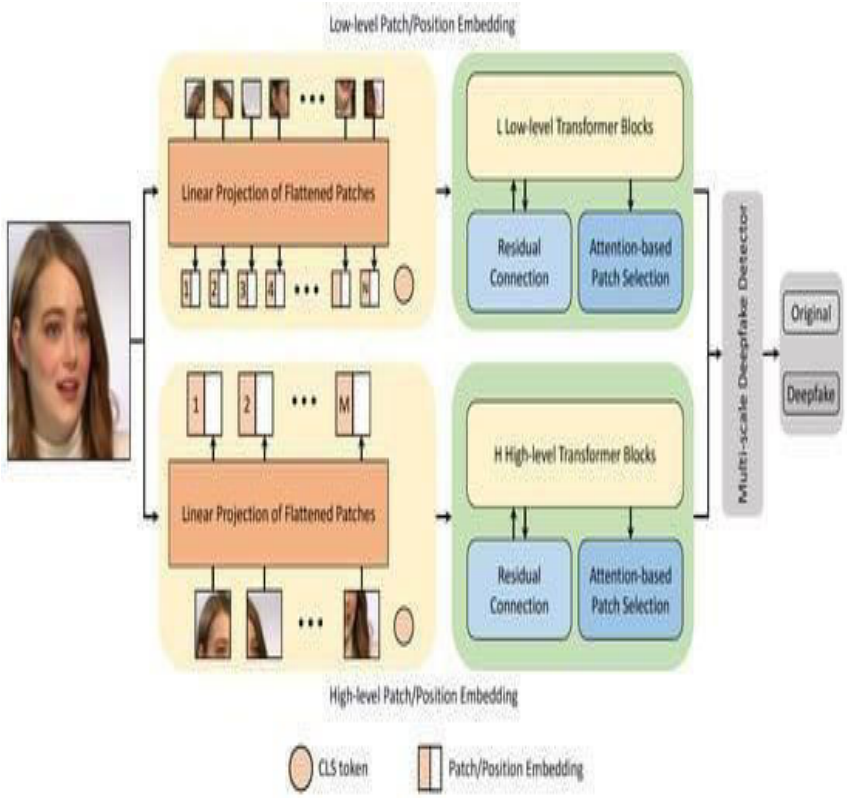


Fig.3. Sample Design

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505038

Fig. 3 shows how the input face image is divided into small patches and converted into embeddings. These embeddings pass through multiple transformer blocks that use attention to focus on important facial regions and features. Finally, the model analyzes the learned features to classify whether the image is original or a deepfake.

VI. RESULT AND DISCUSSION

Fig.4. HomePage

Fig. 4 shows the interface where we have give the image as input the homepage of the website.

Fig.5. Result

Fig. 5 shows the result of the input image as real image.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

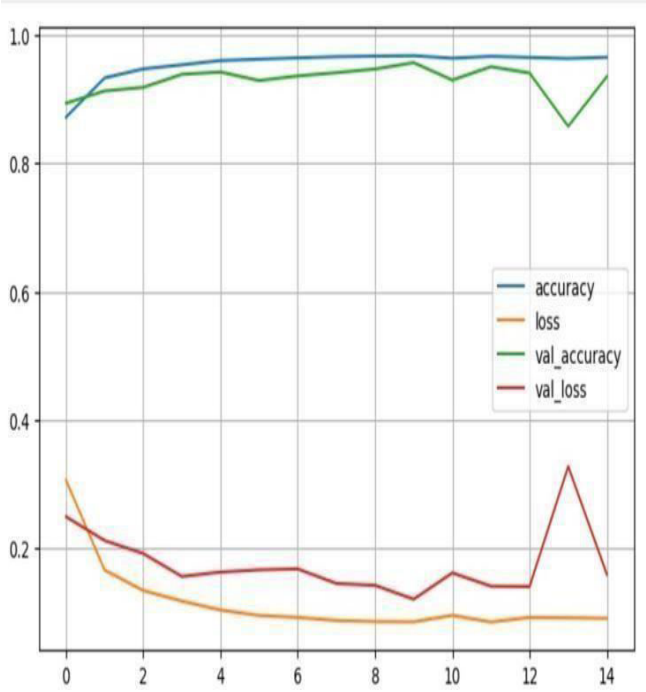
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505038

- The image which is to be tested must be uploaded by browsing local files.
- Then the uploaded image is classified by the model.
- The model has a final accuracy of 91.9%.

Performance Evaluation Metrics for Deepfake Detection Model



The graph shows how well the deepfake detection model works, breaking down its precision, recall, and F1-score for both real and fake images. The model scores high across the board, which means it doesn't just get one thing right—it's solid all around. In the end, these results make it clear: the system reliably tells deepfakes apart from real images.

VII. CONCLUSION

Spotting image deepfakes is now a big deal in the fight against all this slick, manipulated media online. Deepfake tools just keep getting sharper—and honestly, almost anyone can use them now. So, yeah, the pressure's on to come up with ways to spot fakes that actually work. We have some impressive stuff already, like deep learning models and clever face analysis tricks, but the problems haven't disappeared. Deepfake tech doesn't sit still. As soon as detection gets better, new fakes pop up that slip through the cracks. Plus, let's be real, there are still plenty of false alarms, missed fakes, and it takes a ton of computer muscle to catch everything in real time.

But cracking this deepfake problem isn't just about inventing the next smart tool. We need a bigger approach. People have to know what's out there, learn how to spot trouble, and have laws that actually keep up with the tech. No one group can handle this mess solo. It takes tech experts, lawmakers, teachers—the whole bunch—actually working together. At the end of the day, there's no single fix. We need strong detection tools, sure, but also smart social moves to stay ahead of the curve. That's our best shot at keeping digital media honest.

VIII. FUTURE SCOPE

The future of image deepfake detection is all about keeping up with the ever-evolving tricks behind fake images and making these detection tools common across different industries. Deepfake algorithms are getting smarter every day, so researchers are pushing to build tougher and more flexible detection models. The goal? Spot all kinds of deepfakes—even the ones cooked up by the latest AI tools like Generative Adversarial Networks (GANs) and Diffusion Models.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 5, May 2026

|DOI: 10.15680/IJIRSET.2026.1505038

Looking ahead, using systems that combine multiple types of information—not just visuals, but also audio, text, and context—seems like a smart move. Adding blockchain for content authentication is another promising idea, since it locks down image integrity in a way that's tough to tamper with. And with deepfakes spreading fast on social media and in live broadcasts, there's real pressure to come up with faster, more efficient algorithms that can handle huge amounts of data without losing accuracy.

REFERENCES

1. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," 2023, arXiv:2304.02643.
2. Y. Mirsky and W. Lee, "The creation and detection of deepfakes: A survey," ACM Computing Surveys, vol. 54, no. 1, pp. 1–41, January 2022.
3. J. Hui, "How Deep Learning Fakes Videos (Deepfake) and How to Detect it," Accessed: Jan. 4, 2021. [Online]. Available: https://medium.com/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-it-c0b50fbf7cb9.
4. M. Masood, M. Nawaz, K. M. Malik, A. Javed, and A. Irtaza, "Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward," arXiv preprint, arXiv:2103.00484, 2021.
5. N. Engel, V. Belagiannis, and K. Dietmayer, "Point transformer," IEEE Access, vol. 9, pp. 134826–134840, 2021.
6. S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2Net: A new multiscale backbone architecture," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 2, pp. 652–662, Feb. 2021.
7. H. M. Nguyen and R. Derakhshani, "Eyebrow recognition for identifying deepfake videos," in Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, Sep. 2020, pp. 1–5.
8. M. Karras, M. Hellsten, J. Laine, S. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," arXiv preprint, arXiv:2006.06676, 2020.
9. L. Verdoliva, "Media forensics and DeepFakes: An overview," IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, August 2020.
10. O. Fried, A. Tewari, M. Zollhöfer, A. Finkelstein, E. Shechtman, D. B. Goldman, K. Genova, Z. Jin, C. Theobalt, and M. Agrawala, "Text-based editing of talking-head video," ACM Transactions on Graphics, vol. 38, no. 4, pp. 1–14, August 2019.
11. X. Zhang, S. Karaman, and S.-F. Chang, "Detecting and simulating artifacts in GAN fake images," in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Dec. 2019, pp. 1–6.
12. A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," arXiv preprint, arXiv:1809.11096, 2018.
13. G. Patrini, F. Cavalli, and H. Ajder, "The state of deepfakes: Reality under attack," Deeptrace B.V., Amsterdam, The Netherlands, Annu. Rep. v.2.3., 2018.
14. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1800–1807.
15. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Advances in Neural Information Processing Systems, vol. 27, 2014, pp. 1–9.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details