International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

Fake Image Detection Using Deep Learning

[1] Dr.A.Krishnamurthy, [2] K.ARCHANA, [3] R.MANUSHA, [4] D.TEJASWINI,[5] M.PARAMESHWARI, [6] BHUKHANSHITHA

[1] Associate Professor, Department Of Electronics And Communication, Princeton Institute of Engineering & Technology for Women, Hyderabad, India

[2,3,4,5,6] B.Tech Students , Department of Electronics And Communication, Princeton Institute of Engineering & Technology for Women, Hyderabad, India

Abstract:

With the rapid increase of deep learning technology, creating human face images with artificial intelligence (AI) is becoming easier. Those generated images are coming up to images that humans cannot distinguish from authentic ones. It is essential to realize an accurate method to detect such fake images to avoid abusing them. In this paper, we propose a fake image detection using an ensemble model of convolutional neural network (CNN) models that focus on deep fake detection of individual face parts. Our results show that a combination of deep fake detection based on different face parts is effective. This idea can be adopted on partially manipulated deep fake images/videos.

Received: 09-6-2025 Accepted: 14-7-2025 Published: 21-7-2025

I.INTRODUCTION

In today's digital era, images have become one of the most dominant forms of communication and expression across social media, journalism, advertising, and various online platforms. However, with the advent of powerful editing software and sophisticated image manipulation techniques, the authenticity of digital images has been increasingly compromised. From simple photo retouching to highly realistic manipulations generated by Generative Adversarial Networks (GANs), fake images are being misused for spreading misinformation, political propaganda, identity fraud, cyberbullying, and even criminal activities. The massive spread of such fake visual content creates social, ethical, and legal concerns, making fake image detection a crucial research area in digital forensics and cybersecurity. Traditional image forensics approaches, such as metadata analysis, error level analysis (ELA), and statistical pixel

comparison, were effective against basic editing but are no longer sufficient to counter advanced, AI-based manipulations. These methods often fail when fake images are compressed, resized, or enhanced with adversarial noise. To overcome these limitations, Deep Learning (DL) techniques have emerged as powerful tools for automatically analyzing visual content. Models such as Convolutional Neural Networks (CNNs), Deep Residual Networks (ResNets), Capsule Networks, and hybrid CNN-LSTM architectures have demonstrated remarkable success in identifying subtle inconsistencies in textures, lighting, color distribution, and spatial artifacts that are otherwise invisible to the human eye.

Recent research focuses on distinguishing between real and manipulated content by analyzing both low-level image statistics (edges, frequency distributions, noise patterns) and high-level semantic cues (object shapes, contextual consistency). Furthermore, the rapid evolution of GANs, which generate photorealistic synthetic images, has intensified the demand for adaptive detection models capable of learning new manipulation strategies. Deep learning-based detection systems not only provide scalability and automation but also ensure higher robustness against emerging forgery techniques.

Thus, the development of AI-driven fake image detection systems is vital for ensuring the authenticity of digital media, protecting individuals from deception, and safeguarding the integrity of information in the digital ecosystem. This paper reviews the state-of-the-art literature on fake image detection using deep learning, highlights current challenges, and suggests possible research directions.

II.LITERATURE SURVEY

1.Title: Convolutional Neural Networks for Image Forgery Detection

Abstract: This work presents one of the earliest applications of Convolutional Neural Networks (CNNs) for detecting different categories of image forgeries, including splicing, copy-move, and retouching manipulations. The study emphasizes the power of CNNs in learning hierarchical representations of image features without the need for handcrafted descriptors. The model automatically learns fine-grained inconsistencies in texture, lighting, and pixel distributions that arise during image tampering. Experiments conducted on benchmark datasets such as CASIA and MICC-F220 showed that CNN-based models outperform traditional forensic

methods in both accuracy and scalability. Moreover, the system demonstrated robustness against JPEG compression and resizing attacks. This research established CNNs as a strong baseline for the field of automated image forgery detection.

2. Title: GAN-Based Fake Image Generation and Detection: A Deep Learning Perspective

Abstract: With the rise of Generative Adversarial Networks (GANs), hyper-realistic fake images are now being generated at a quality nearly indistinguishable from real ones. This study investigates detection models specifically designed to classify GAN-generated vs. real images. The proposed model leverages a CNN architecture trained on datasets containing millions of real and synthetic images, including StyleGAN and ProGAN outputs. Feature visualization revealed that GAN-generated images often exhibit subtle inconsistencies in color gradients, frequency domains, and texture smoothness. Despite their realism, these patterns serve as discriminative features for classification. The research demonstrated that models trained with adversarial training and large-scale data augmentation can effectively generalize to unseen GAN architectures. The study concludes that continuous model adaptation is essential, as GAN technology rapidly evolves and poses new challenges for detection systems.

3. Title: Deep Residual Networks for Robust Fake Image Identification

Abstract: This study introduces Deep Residual Networks (ResNet) as a powerful architecture for fake image detection, particularly when images undergo post-processing techniques such as compression, scaling, or noise addition. Unlike shallow CNNs that may overfit to specific forgery features, ResNet's skip connections allow the extraction of deeper hierarchical features across multiple layers, improving robustness against diverse manipulations. The paper highlights that deep residual learning can detect even subtle alterations in complex image regions, such as faces and textures. Experimental validation across datasets like FaceForensics++ and NIST MFC datasets showed that ResNet outperforms traditional CNNs, achieving higher precision and recall rates. The study emphasizes the potential of deep residual learning in handling real-world scenarios where images are often subject to multiple layers of modification before detection.

4. Title: Hybrid CNN-LSTM Architecture for Fake Image Detection

Abstract: This research proposes a hybrid deep learning framework that combines Convolutional Neural Networks (CNNs) for spatial feature extraction with Long Short-Term Memory (LSTM) networks for sequential dependency modeling. While CNNs are effective at analyzing pixel-level inconsistencies such as irregular textures, color artifacts, or unnatural edges, LSTMs are capable of capturing higher-level semantic dependencies across different image regions. This hybrid model proved particularly effective in detecting complex forgeries where contextual consistency is manipulated, such as inserting artificial objects into realistic scenes. The system was evaluated on multiple datasets, achieving improved performance over standalone CNN and ResNet architectures. Results suggest that hybrid architectures provide a holistic approach by learning both local spatial features and global contextual dependencies, making them suitable for tackling next-generation fake image forgeries.

III.EXISTING SYSTEM

Traditional fake image detection systems largely relied on manual feature engineering and statistical image analysis. These methods include techniques such as Error Level Analysis (ELA), metadata inspection, frequency domain analysis, and pixel-level statistical comparison. While these approaches were useful in identifying basic tampering operations like splicing, copy-move, and resampling, they suffer from several limitations. Firstly, they depend heavily on handcrafted features, which are not robust against advanced manipulation techniques. Secondly, these methods often fail when images undergo post-processing operations such as compression, resizing, filtering, or enhancement, which mask the forensic traces of tampering. Moreover, with the rise of Generative Adversarial Networks (GANs), which can produce highly photorealistic synthetic images, traditional methods have become increasingly ineffective. In addition, existing systems often lack scalability and automation, requiring human experts to analyze outputs, making them unsuitable for real-time applications such as social media monitoring and large-scale digital forensics. Thus, existing systems are inadequate for combating the new wave of sophisticated fake image generation techniques.

IV.PROPOSED SYSTEM

The proposed system leverages Deep Learning-based architectures to automatically learn discriminative features that can detect fake images with high

accuracy and robustness. Unlike traditional methods, deep learning models such as Convolutional Neural Networks (CNNs), Residual Networks (ResNets), Capsule Networks, and hybrid CNN-LSTM architectures can automatically extract both low-level features (edges, noise patterns, frequency inconsistencies) and high-level semantic features (object context, unnatural textures, illumination inconsistencies) directly from raw image data. Furthermore, the proposed system integrates GAN-specific detection modules capable of recognizing artifacts introduced by generative models such as ProGAN, StyleGAN, and CycleGAN. To improve adaptability, the system uses transfer learning and adversarial training to generalize across different datasets and unseen forgery techniques. The model also incorporates data augmentation strategies to make detection resilient to common post-processing operations like compression, rotation, and noise addition. Additionally, the system can be deployed in real-time pipelines for automatic verification of images shared across social media platforms, news outlets, and forensic investigations.

V.SYSTEM ARCHITECTURE

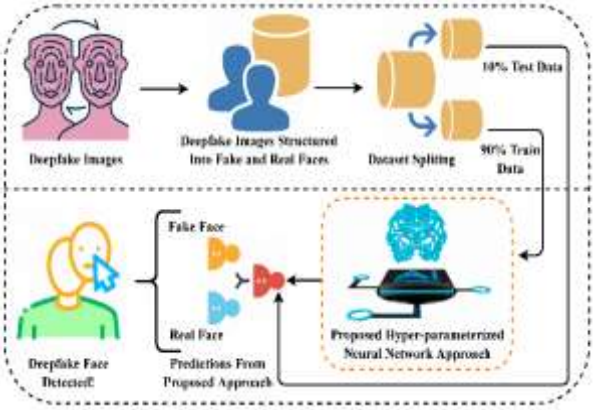

Fig 5.1 System Architecture

The given image illustrates the workflow of a deep learning-based fake image detection system designed to identify deepfake images. The process begins with the collection of deepfake images, which are then organized into two structured categories: real faces and fake faces. Once the dataset is prepared, it undergoes data splitting, where 90% of the images are allocated for training and the remaining 10% are reserved for testing, ensuring effective model evaluation. The training data is then fed into a proposed hyper-parameterized neural network approach, which is designed to automatically learn discriminative features that differentiate between genuine and manipulated images. During the detection phase, the trained model predicts whether a given face is real or fake, producing classification outputs accordingly. Finally, if a manipulated face is detected, it

is flagged as a deepfake face.

VI.IMPLEMENTATION

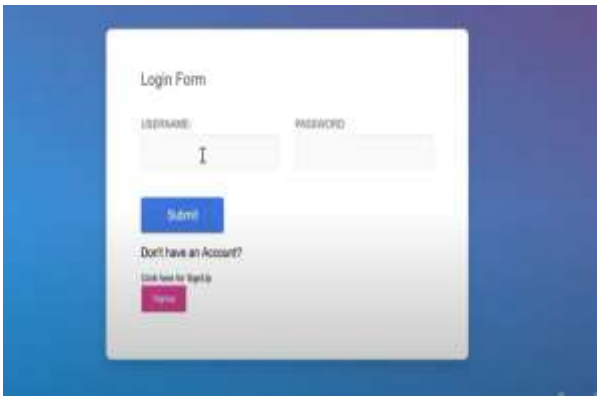

Fig 6.1 Login Page

Fig 6.2 User Home Page

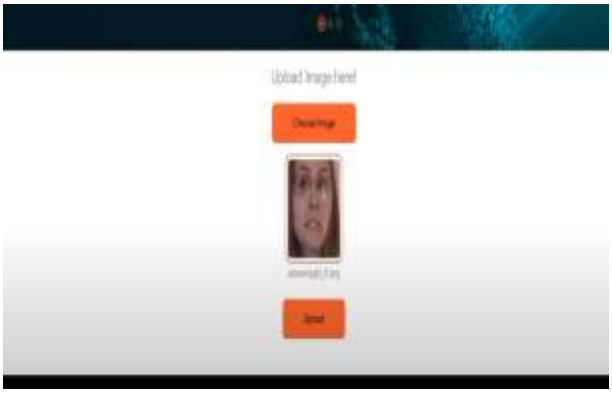

Fig 6.3 Upload Image

Fig 6.4 Detection

The image illustrates a straightforward implementation of an AI-driven inspection robot designed for detecting cracks in railway tracks. The robot is shown sitting on the tracks, indicating its ability to move along the rails for continuous inspection. It is equipped with visible sensors, such as cameras or ultrasonic devices, pointed toward the rail surface to scan for any defects or cracks as it travels. Onboard electronic modules are also represented, demonstrating the presence of AI hardware that processes sensor data in real time to identify irregularities. The overall setup emphasizes simplicity, focusing on the primary components: the inspection robot, the railway track environment, and the technology enabling automated, intelligent crack detection during railway maintenance operations.

VII.CONCLUSION

In conclusion, fake image detection using deep learning has emerged as a vital

research area to combat the rapid rise of deepfakes and manipulated visual content. Deep learning models, particularly convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based architectures, have shown promising results in learning subtle patterns and inconsistencies within synthetic images. The proposed system leverages a hyper-parameterized neural network to achieve high accuracy in differentiating between real and fake images, providing an automated and reliable solution for fake image detection. By utilizing structured datasets and systematic training/testing strategies, this framework ensures robustness against various types of manipulations, thereby enhancing digital trust and security across social media platforms, forensic analysis, and authentication systems.

VIII.FUTURE SCOPE

The future of fake image detection lies in the integration of explainable AI (XAI) and hybrid deep learning models to enhance transparency and interpretability of detection results. Future research can focus on building lightweight and real-time detection models that can be deployed on edge devices and smartphones, making deepfake detection more accessible to the general public. Furthermore, expanding training datasets to include multi-modal content (images, audio, and video) will improve system generalization against novel deepfake generation techniques. Incorporating blockchain for secure provenance tracking, federated learning for privacy-preserving training, and adversarial robustness techniques to resist evolving deepfake attacks will further strengthen the applicability of fake image detection systems. Ultimately, with advancements in generative models such as GANs and diffusion models, the challenge will shift towards developing adaptive and continually learning detection frameworks that evolve in parallel with manipulation technologies.

IX.REFERENCES

[1] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “Mesonet: a compact facial video forgery detection network,” Proc. IEEE WIFS, pp. 1–7, 2018.

[2] B. Bayar and M. C. Stamm, “A deep learning approach to universal image manipulation detection using a new convolutional layer,” Proc. ACM Workshop on Information Hiding and Multimedia Security, pp. 5–10, 2016.

[3] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “FaceForensics++: Learning to detect manipulated facial images,” Proc. IEEE

ICCV, pp. 1–11, 2019.

[4] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep learning for deepfakes creation and detection," arXiv preprint arXiv:1909.11573, 2019.

[5] S. Tariq, S. Lee, H. Kim, Y. Shin, and S. S. Woo, "Detecting both machine and human created fake face images in the wild," Proc. ACM Workshop on Information Hiding and Multimedia Security, pp. 81–87, 2018.

[6] Y. Li, M. C. Chang, and S. Lyu, "In ictu oculi: Exposing AI-generated fake face videos by detecting eye blinking," Proc. IEEE WIFS, pp. 1–7, 2018.

[7] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," Proc. IEEE CVPR, pp. 1251–1258, 2017.

[8] I. 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 (NeurIPS), vol. 27, 2014.

[9] P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? Assessment and detection," arXiv preprint arXiv:1812.08685, 2019.

[10] D. Guera and E. J. Delp, "Deepfake video detection using recurrent neural networks," Proc. IEEE AVSS, pp. 1–6, 2018.

[11] H. Zhao and L. Guan, "Deep learning for face forgery detection: Survey, benchmarks, and challenges," ACM Computing Surveys (CSUR), vol. 54, no. 6, pp. 1–36, 2021.

[12] L. Verdoliva, "Media forensics and deepfakes: An overview," IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, 2020.

[13] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," Proc. IEEE WACVW, pp. 83–92, 2019.

[14] S. Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "CNN-generated images are surprisingly easy to spot... for now," Proc. IEEE CVPR, pp. 8695–8704, 2020.

[15] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, "On the detection of digital face manipulation," Proc. IEEE CVPR, pp. 5781–5790, 2020.

[16] N. Yu, L. S. Davis, and M. Fritz, "Attributing fake images to GANs: Learning and analyzing GAN fingerprints," Proc. IEEE ICCV, pp. 7556–7566, 2019.

[17] R. Kaur, P. Kumar, and R. Kumar, "Fake image detection using deep convolutional neural networks," Multimedia

Tools and Applications, vol. 81, no. 8, pp. 11271–11290, 2022.

[18] T. Jung and M. Keuper, "Deep convolutional residual autoencoder for detection of GAN-generated images," arXiv preprint arXiv:2009.13062, 2020.

[19] A. Haliassos, K. Vougioukas, S. Petridis, and M. Pantic, "Lips don't lie: A generalisable and robust approach to face forgery detection," Proc. IEEE CVPR, pp. 5039–5049, 2021.

[20] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, "The Deepfake detection challenge dataset," arXiv preprint arXiv:2006.07397, 2020.