RESEARCH PAPER

Detecting Fake Images Generated by Gans Using Deep Convolutional Feature Analysis

Sethu Vinayagam T[1], Shanmugam C[2], Srinivasan V[3], Krishnamoorthy[4], Dr. V. Subedha[5], Dr. L. Jabasheela[6]

[1]Dept. of Computer Science and Engineering, Panimalar Engineering College, Chennai, India.
Email: sethuvinayagam70@gmail.com
[2]Dept. of Computer Science and Engineering, Panimalar Engineering College, Chennai, India.
Email: theshanmugam2520@gmail.com
[3]Dept. of Computer Science and Engineering, Panimalar Engineering College, Chennai, India.
Email: srinivasanv28022005@gmail.com
[4]Dept. of Computer Science and Engineering, Panimalar Engineering College, Chennai, India.
Email: krishnamoorthymuniyan@gmail.com
[5]Dept. of Computer Science and Engineering, Panimalar Engineering College, Chennai, India.
Email: subedha@gmail.com
[6]Dept. of Computer Science and Engineering, Panimalar Engineering College, Chennai, India.
Email: ljsheela@gmail.com

ABSTRACT

Background: The rapid development of GAN has made it easy to create very realistic fake images, including images generated by popular AI platforms such as Gemini, Grok, and ChatGPT. This has raised serious concerns about misinformation, digital investigations, and whether online images can be trusted.

Objective: This paper presents a deep learning–based method to detect fake images created using GANs and modern AI image-generation tools.

Methods: The approach uses Convolutional Neural Networks (CNNs) to automatically learn important spatial and frequency-based features that reveal small errors and hidden patterns left during image generation. By training the model on a wide range of real images and AI-generated images from different sources, the system becomes capable of identifying fake images produced by various GAN models and platforms.

Results: Experimental results show that this method achieves high accuracy and performs better than traditional techniques that depends on manually designed features, even when images are edited or compressed.

Keywords: GANs, Fake Image Detection, Deep Learning, CNN

How to cite this article: Sethu Vinayagam T, Shanmugam C, Srinivasan V, Krishnamoorthy, Subedha V, Jabasheela L. Detecting Fake Images Generated by Gans Using Deep Convolutional Feature Analysis. Int J Drug Deliv Technol. 2026;16(19s): 377. DOI: 10.25258/ijddt.16.19s.45

Source of support: Nil.

Conflict of interest: None

I. INTRODUCTION

The rapid growth of Generative Adversarial Networks (GANs) and modern generative AI systems has made it possible to create very realistic fake images that are often hard to tell apart from real ones. Advanced AI platforms such as ChatGPT with image generation, Gemini, and Grok can create images from text and also easily add, remove, or change objects in existing images using simple language commands. While these tools are very useful in areas like creative design, entertainment, advertising, and data creation, they also create serious problems. The easy modification of images increases the risk of fake content, misinformation, and misuse, especially on social media, in news reporting, politics, and security-related fields.These powerful image-generation tools are now being used in sensitive areas, which has increased concerns .

Older detection approaches that rely on manually designed features, fixed rules, or human checking do not scale well. Their accuracy drops further when images are compressed, resized. Because of these challenges, there is a clear need for automated and intelligent systems that can accurately identify whether an image is real or fake

Deep convolutional neural networks (CNNs) are well suited for this task because they can automatically learn meaningful patterns from images. The proposed method uses deep convolutional feature analysis to identify small spatial and frequency-based inconsistencies left behind during image creation or manipulation. This allows the system to accurately classify real and fake images. Experimental results show that the approach achieves high detection accuracy and works well across different GAN models and modern AI image-generation platforms. As a result, it helps improve digital trust, supports forensic investigations, and reduces the impact of AI-generated visual misinformation. The same approach can also be extended to detect deepfake images and content created using popular deepfake tools and AI-based image editing platforms.

The proposed approach can also be extended to detect deepfake images and content created using popular deepfake tools and AI-based image editing platforms. Today, many people use AI tools to create or modify images and videos with just a simple text command. While this makes image creation fast and convenient, it also introduces serious concerns. Personal photos can be altered or misused without consent, leading to privacy risks. At the same time, people experience increased stress and uncertainty because it is becoming harder to tell whether an image online is real or fake, and they may fear being misrepresented or unfairly targeted.. Since AI-generated images look very real and are easy to make, the risk of fake news, identity misuse, and emotional Many modern deepfake tools use advanced GAN and diffusion models to change faces, expressions, and even identities, making fake images harder to detect. A CNN-based neural network model helps identify images generated by these tools by learning deep visual features and frequency-based patterns. The model can detect hidden errors and inconsistencies left during the image generation process, which are not easily visible to the human eye or traditional detection methods. Because of this proposed system improves the accuracy of fake images detection in real world situations.It can be effectively used for deepfake detection,identity verification ,social media content modification,and helping to increase the trust and security in ai generated visual content.

II. LITERATURE SURVEY

Hany Farid [13] says that fake images generated using GANs can be detected through machine learning and image forensics techniques, demonstrating the importance of deep learning in digital media authentication. The proposed approach is trained on a limited dataset of real and synthetic images, which leads to reduced performance when images generated by unseen GAN models are tested. This affects the system’s ability to generalize across different image generation sources.Yonghao Xu [14] says that with the advancement of deep learning, GAN-generated images can be classified in less time with high accuracy. This paper explores various deep learning models such as VGGNet, ResNet, DenseNet, and InceptionNet based on CNN architectures and concludes that deeper networks achieve superior detection performance. Optimization techniques such as normalization and data aug- mentation are applied. However, the paper provides lim- ited discussion on deployment and robustness against post- processing operations. Shiqing Ma [15] says that an advanced neural network architecture combining spatial-domain and frequency-domain features is proposed for detecting GAN- generated fake images. The system integrates CNN-based feature extraction with spectral analysis to identify image inconsistencies. Although the approach shows good accuracy, it struggles to recognize fake images generated by differ- ent GAN architectures and is sensitive to compression and resizing.Nguyen Thanh-Toan

[16] says that a hybrid model combining MobileNet for feature extraction and traditional classifiers for fake image detection is proposed. The model achieves high accuracy while reducing computational com- plexity. However, reliance on pre-trained networks limits the ability to capture GAN-specific artifacts introduced by newly emerging image generators.Zhihua Yang

[17] says that an EfficientNet-based deep learning framework is introduced for detecting GAN-generated fake images. In addition, k-fold cross-validation is used to validate the proposed approach. Experimental results show high accuracy in classifying real and fake images. However, the training time is high, and the computational cost is significant.Yuezun Li

[18] says that the quality of extracted features plays a crucial role in fake image detection. The study integrates deep learning features with handcrafted texture and noise features to improve classifica- tion accuracy. This work mainly focuses on specific GAN- form of the images or pictures which has been

generated datasets and does not consider prompt-based image editing such as object insertion or removal.Bin Li [19] says that image processing techniques such as edge detection and statistical analysis combined with machine learning classifiers like SVM can be used for detecting synthetic images. The process involves preprocessing followed by feature extraction and classification. The method performs well on early GAN-generated images but fails to detect highly realistic images produced by modern generative AI systems.Xiaodan Zhang
[20] says that accuracy and consistency in fake image classifi-cation are improved using advanced CNN architectures such as MobileNetV3 and ShuffleNetV2. Feature selection techniques are applied to extract the most relevant features. However, the approach does not address image noise suppression and is not evaluated on images modified using prompt-based editing.Tamal Kumar Kundu [21] says that recent advance-ments in machine learning and deep learning techniques for fake image detection are reviewed, highlighting challenges related to dataset diversity and model generalization. The study emphasizes that rapid evolution of GAN architectures limits the robustness of existing detection models.Rabia Asghar
[22] says that machine learning and deep learning models are reviewed for image authenticity verification. The study highlights the importance of transfer learning and discusses challenges such as dataset availability and scalability. It also identifies a lack of meta-learning approaches in current fake image detection research.Srishti Srivastava [23] says that a mobile-based fake image detection system using pre-trained CNN models is proposed. The system demonstrates practical feasibility but heavily depends on image quality and controlled acquisition conditions, limiting its applicability in real-world environments.Sami H. Ismael [24] says that segmentation-based feature extraction followed by machine learning clas-sification can be used for image classification tasks. Texture and shape features are extracted and classified using different algorithms. The complexity of visual patterns makes accurate fake image detection challenging.Changhun Jung [25] says that a W-Net–based CNN architecture is proposed for dis-tinguishing real and synthetic images. The study compares original images with GAN-generated images and reports high similarity. However, dataset imbalance affects classification performance and limits generalization.Hao Dang [26] says that neural network–based classification models are useful for decision-support systems. The approach mainly focuses on improving classification accuracy, but it does not properly handle the detection of manipulated or AI-generated images in real-world situations. According to Verdoliva[29], multimedia forensics is very important for identifying edited images and images created using GANs. The study highlights the use of deep learning–based forensic features to tell the difference between real and fake images. However, this method is sensitive to image compression and does not adapt well to newer image-generation models. Frank et al. [30] and their team explain that GANs leave different patterns in images because of their process, particularly in the frequency domain. These patterns can be used to identify fake images. Their CNN-based approach performs effectively and achieves high accuracy when tested on images generated by various GAN models.

PROPOSED SYSTEM

A. *Overview*

B. *E. Feature Extraction*

The rapid advancement of Generative Adversarial Networks (GANs) and modern AI image-generation tools has made it possible to create highly realistic fake images that are often difficult to distinguish from real ones. These technologies are helpful in areas such as creative design, entertainment, cartoons, VFX, advertising, and creating training data. However, they also bring serious challenges, including the spread of false information, difficulty in verifying digital evidence, and a loss of trust in images and media.

Earlier, fake image detection mainly depended on man-ual checking or traditional image processing methods. These methods take more time and can make mistakes, especially when the images are highly realistic. To overcome this prob-lem, deep learning–based image analysis is used. By applying Convolutional Neural Networks (CNNs), fake images can be detected more accurately and in much less time compared to traditional approaches.

C. *Dependencies Software Requirements:*

- *Python*
    - *Anaconda Navigator Framework used:*
- *Numpy*
- *Pandas*

- *Matplotlib*
- *TensorFlow*
- *Keras*
- *OpenCv*

D. Data Collection

To identify fake images, the system needs a dataset that contains both real images and images created using GANs. These images are gathered from open and trusted public datasets, as well as from popular AI image-generation tools.

The dataset has fake images produced by different GAN models so the system does not learn patterns from only one type of image. Using clear high-quality, and different images is very important because it helps the model understand real world differences and improves its ability to correctly detect fake images or generated images.

E. Pre-processing

Before training the model, the images are prepared so they are easier to study and learn from. This process involves resizing and arranging all images to a uniform size, normalizing pixel values to a common range, and removing unnecessary noise to improve image quality.

These steps help the model focus on important features rather than being distracted by differences in image size or lighting conditions. Pre-processing also speeds up the training process and improves the overall accuracy of the system.

Instead of manually selecting image features, theCNN automatically learns them directly from the images. In the layers, it identifies basic elements such as edges, shapes, and texture.In the other deeper layers, it learns more detailed patterns and small errors that often appear in images created by GANs.so that it classify the images based outcome of all the layers in cnn.

These learned features help the system clearly tell the difference between real and fake images, even when the differences are very hard for humans to see.

F. Database

All the images are taken from Kaggle datasets and other offline real-time sources. Along with the images, the extracted features and their labels real or fake are stored in an organized database. This database is used during both the training and testing stages of the model.

It also helps in checking results, finding mistakes, and improving the system by showing which types of images are more difficult for the model to identify correctly.

G. CNN-Based Classification

A Convolutional Neural Network (CNN) is used as the main model to classifyimages in this system. CNNs work verywell with images because theycan automaticallylearn useful from patterns directly from imagedata. By training the CNN on both real and fake images, the system learns to detect hidden patterns left during image creation. This reduces the need for manual inspection and makes the detection process faster and more accurate.

IV .RESULT AND ANALYSIS

This section describes how the experiments were configured to identify images produced by GANs. Real and fake images produced by various GAN models were included in the experiments' public and private dataset, which was gathered from multiple sources such as Kaggal and real-time Deep learning tools like TensorFlow and Keras, along with Python, were used in the system's construction.

A computer running Windows 11 (64-bit), an Intel Core i5 (12th generation) processor, and 6 GB of RAM was used for all experiments. The CNN model underwent 55 training cycles, or epochs, with a batch size of 50. To help the model learn, the images were cleaned, resized, and normalized before training. The accuracy and ability of the finished model to recognize phony images produced by various GAN techniques were evaluated.

THE FOLLOWING FIG 4.1 SHOWS THE MODEL TRAINING:

THE FOLLOWING FIG 4.2SHOWS THE GRAPH OBTAINED AT THE END OF MODEL TRAINING:

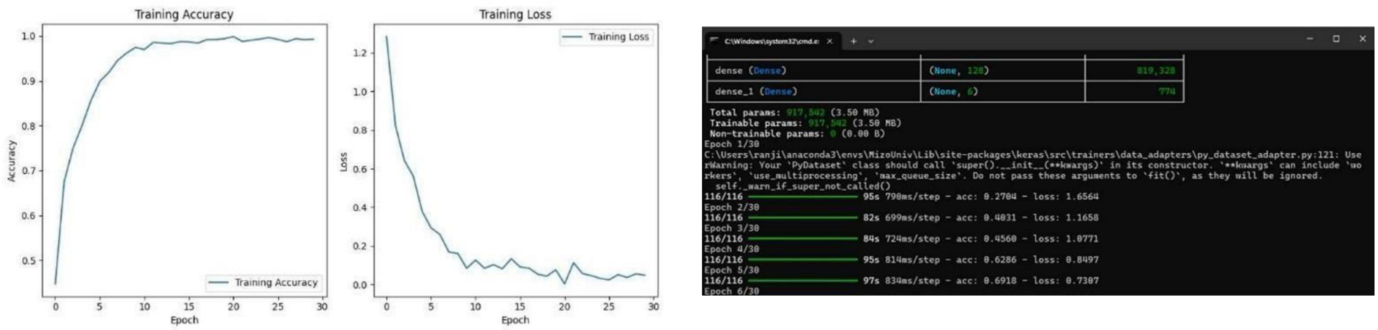

FIGURE 4.3SHOWS THE RESULT OF DETECTING WHETHER THE GIVEN IMAGE IS AI-GENERATED OR REAL.

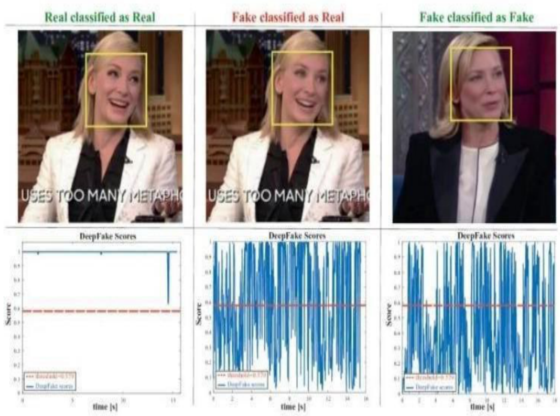

FIGURE 4.4SHOWS THE WORKING MECHANISM.

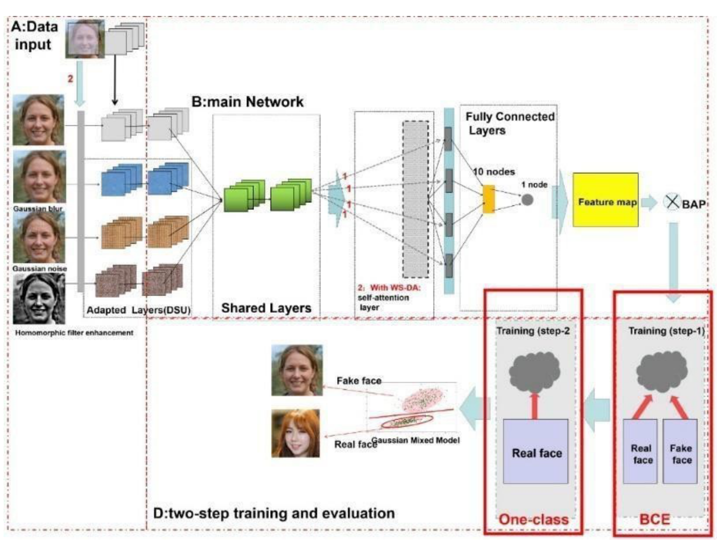

FIGURE 4.5PRESENTS THE FINAL OUTPUT OF THE SYSTEM

CONCLUSION

The rise of AI-generated images and videos has made it really hard to trust what we see online anymore. Thankfully, new research shows we can fight back using deep learning—especially Convolutional Neural Networks (CNNs). When we pair these networks with some smart image preprocessing steps, they pick up on the tiny “fingerprints” that GANs leave behind little clues the human eye usually misses completely.

It can actually be used in real places like digital forensics, social media moderation, or cybersecurity teams. No more slow manual checks we can verify authenticity quickly and at huge scale. Still, it’s not a one-and-done fix. As GANs get smarter and more realistic, our detectors have to keep improving too. And beyond the tech, we need to use these tools responsibly—with transparency, fair rules, and a real focus on rebuilding trust online.

A. Future Work

Learning Over Time

By assisting the model in repeatedly learning from fresh data, it can be made better. The system will remain accurate and helpful if it is updated on a regular basis because new AI image-making techniques are emerging rapidly.

Working Across Different Platforms

The model should perform well on images produced by a variety of AI tools in the future. The results remain accurate regardless of the image's source, including those created with text prompts and contemporary image-generation techniques.

Handling Image Changes

It is possible to strengthen the system so that it remains functional even if the images are altered. This includes alterations that frequently occur when photos are shared online, such as low quality, noise addition, cropping, or resizing.

Fast And Real Time Use

Increasing the model's speed will enable real-time image verification. This can facilitate the rapid detection of AI-generated images by digital investigation tools, content checking systems, and social.

REFERENCES

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, "Generative Adversarial Networks", in Proceedings of the International Conference on Neural Infor-mation Processing Systems (NIPS), 2014.

[2] Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, Cambridge, MA, USA, 2016.

[3] Christian Szegedy, Wei Liu, Yangqing Jia et al., "Going deeper with convolutions", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[4] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition", in International Conference on Learning Representations (ICLR), 2015.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, "Deep residual learning for image recognition", in IEEE CVPR, 2016.

[6] Gao Huang, Zhuang Liu, Laurens van der Maaten and Kilian Q. Weinberger, "Densely connected convolutional networks", in IEEE CVPR, 2017.

[7] Ming-Yu Liu, Thomas Breuel and Jan Kautz, "Unsuper- vised image-to-image translation networks", in Advances in Neural Information Processing Systems (NeurIPS), 2017.

[8] Alec Radford, Luke Metz and Soumith Chintala, "Un-supervised representation learning with deep convolutional generative adversarial networks", in ICLR, 2016.

[9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei
A. Efros, "Image-to-image translation with conditional adversar- ial networks", in IEEE CVPR, 2017.

[10] Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei
A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks", in IEEE ICCV, 2017.

[11] Siwei Lyu, "Deepfake detection: Current challenges and next directions", in IEEE International Conference on Multimedia Expo (ICME), 2020.

[12] Yuezun Li, Ming-Ching Chang and Siwei Lyu, "In Ictu Oculi: Exposing AI generated fake images by detecting eye blinking", in IEEE International Workshop on Information Forensics and Security (WIFS), 2018.

[13] Hany Farid, "Image forgery detection and digital image forensics", in IEEE Signal Processing Magazine, Vol. 26, No. 2, pp. 16–25, March 2009.

[14] Yonghao Xu, Bo Zhao, Zhen Liu and Fei Wu, "De-tecting GAN-generated fake images using deep convolutional neural networks", in IEEE Access, Vol. 8, pp. 134848–134863, 2020.

[15] Shiqing Ma, Ruize Wang, Xiaokang Yang and Wenjun Zhang, "Detecting GAN-generated fake images using spatial and frequency domain analysis", in IEEE Transactions on Information Forensics and Security, Vol. 15, pp. 2495–2509, 2020.

[16] Nguyen Thanh-Toan, Nguyen Thanh-Binh and Le Hoang Son, "A lightweight deep learning approach for fake image detection using MobileNet", in Journal of Information Security and Applications, Vol. 58, 2021.

[17] Zhihua Yang, Jing Zhang and Wei Wang, "EfficientNet- based deep learning framework for GAN generated image detection", in IEEE Access, Vol. 9, 2021.

[18] Yuezun Li and Siwei Lyu, "Exposing deepfake images using inconsistent noise patterns", in IEEE Transactions on Information Forensics and Security, Vol. 14, No. 8, August 2019.

[19] Bin Li, Yong Wang and Hui Chen, "Detection of synthetic images using image processing and machine learning techniques", in International Journal of Computer Vision and Image Processing, Vol. 10, No. 3, 2020.

[20] Xiaodan Zhang, Jianming Zhang and Qiang Liu, "Lightweight CNN architectures for fake image detection", in IEEE Access, Vol. 10, 2022.

[21] Tamal Kumar Kundu, Dinesh Kumar Anguraj and Debnath Bhattacharyya, "A review on fake image detection using machine learning and deep learning techniques", in Artificial Intelligence Review, Vol. 56, 2023.

[22] Rabia Asghar, Sanjay Kumar and Arslan Shaukat, "A review on image authenticity verification using machine

[23] Alec Radford, Luke Metz and Soumith Chintala, "Un- supervised representation learning with deep convolutional generative adversarial networks", in ICLR, 2016.

[24] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros, "Image-to-image translation with conditional adversar- ial networks", in IEEE CVPR, 2017.

[25] Jun-Yan Zhu, Taesung Park, Phillip Isola and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks", in IEEE ICCV, 2017.

[26] Siwei Lyu, "Deepfake detection: Current challenges and next directions", in IEEE International Conference on Multimedia Expo (ICME), 2020.

[27] Yuezun Li, Ming-Ching Chang and Siwei Lyu, "In Ictu Oculi: Exposing AI generated fake images by detecting eye blinking", in IEEE International Workshop on Information Forensics and Security (WIFS), 2018.

[28] Hany Farid, "Image forgery detection and digital image forensics", in IEEE Signal Processing Magazine, Vol. 26, No. 2, pp. 16–25, March 2009.

[29] Yonghao Xu, Bo Zhao, Zhen Liu and Fei Wu, "De-tecting GAN-generated fake images using deep convolutional neural networks", in IEEE Access, Vol. 8, pp. 134848–134863, 2020.

[30] Shiqing Ma, Ruize Wang, Xiaokang Yang and Wenjun Zhang, "Detecting GAN-generated fake images using spatial and frequency domain analysis", in IEEE Transactions on Information Forensics and Security, Vol. 15, pp. 2495–2509, 2020.

[31] Nguyen Thanh-Toan, Nguyen Thanh-Binh and Le Hoang Son, "A lightweight deep learning approach for fake image detection using MobileNet", in Journal of Information Security and Applications, Vol. 58, 2021.

[32] Zhihua Yang, Jing Zhang and Wei Wang, "EfficientNet-based deep learning framework for GAN-generated image detection", in IEEE Access, Vol. 9, 2021.

[33] Yuezun Li and Siwei Lyu, "Exposing deepfake images using inconsistent noise patterns", in IEEE Transactions on Information Forensics and Security, Vol. 14, No. 8, August 2019.